NOTTINGHAM TRENT UNIVERSITY

Proteomic Analysis of Biomarkers Associated with Immunotherapy in Murine Tumour Models

Thesis Submitted by

Baharak Vafadar-Isfahani

To

The Nottingham Trent University
In requirement for the degree of Doctor of Philosophy

December 2009

Acknowledgements

Firstly I would like to thank my Director of Studies, Professor Bob Rees, for giving me the opportunity to carry out this piece of research and for their guidance during this time. I woul like to specially say a BIG thank you to Dr Balwir Matharoo-Ball, my supervisor, whose supervision, guidance and support for the last 3.5 years was invaluable. I also whish to thank Dr Murrium Ahmad for all her help with the extensive animal work. I wish to thank Dr Selman Ali and Mr Rob Davy for their help and assistance with the animal work. I would also like to thank Dr Graham Ball, Mr Christophe Lemetre and Dr Lee Lancashire for their help for the bioinformatic side of my project. Thanks must also go to Ms Clare Coveney and Mr Steve Reeder for their help and assictant in the lab. I would also like to thank everyone, who I have worked with in labs 008 and 009 for their help and friendship during my time at NTU. I would like to specialy thank Ms Stephanie Laversin for helping me with the RT-PCR and Dr Susan Gill for her support in the lab. I greatly acknowledge the John and Lucille van Geest Foundation for funding this research.

Finally, I would like to thank my family; firstly my mum and dad for all their support, words of wisdom and encouragement, my brother and sisters for support and entertainment and last but not least my fiancé Farhad for all the support. I cannot begin to explain how much the love and support you have all given to me over the past 3.5 years has meant to me, there is absolutely no way I could have gone through all of the ups and downs if you weren't all behind me, for this I am ever grateful to you all.

Contents

Figures

Tables

Abbreviations

1D-PAGE	1 Dimensional Polyacrylamide Gel Electrophoresis
2D-PAGE	2 Dimensional Polyacrylamide Gel Electrophoresis
ACN	Acetonitrile
AI	Artificial Intelligence
ANNs	Artificial Neural Networks
APC	Antigen Presenting Cell
BCG	Bacilli Calmette-Guerin
BP	Back Propogation
BP	Back-Propagation
CEA	Carcinoembryonic Antigen
CHCA	α-Cyano-4-hydroxycinnamic acid
CSF	Cerebrospinal Fluid
CTL	Cytotoxic T Lymphocytes
CV	Coefficient of Variation
Da	Daltons
DC	Dendritic Cell
DIGE	Differential In Gel Electrophoresis
DISC-HSV	Disabled Infectious Single Cycle-Herpes Simplex Virus
DTCH	Delayed-Type Cutaneous Hypersensitivity
ECM	Extracellular Matrix
ESI	Electro Spray Ionisation
ETP	End Time Point
FAB	Fast Atom Bombardment
FPA	Forward Phase Array
FTP	First Time Point
GEM	Genetically Engineered Mice
GFs	Growth Factors
GM-CSF	Granulocyte-Macrophage Colony-Stimulating Factor
HSP	Heat Shock Protein
HTP	High Throughput

ICAT	Isotope-Coded Affinity Tagging
iTRAQ	Isobaric Tags for Relative and Absolute Quantitation
kDa	Kilodaltons
LCM	Laser Capture Microdissection
LMW	Low Molecular Weight
m/z	Mass-to-charge ratio
MALDI	Matrix-Assisted Laser Desorption Ionisation
MHC	Major Histocompatibility Complex
MLP	Multi-Layer Perceptron
MS	Mass Spectrometry
MSI	Microsatellite Instability
NIR	Normalised Intensity Ratio
PCA	Principal Components Analysis
PCR	Polymerase Chain Reaction
PDVB	Polystyrenedivinylbenzene
PFU	Plaque Forming Unit
pI	Isoelectric Point
PSA	Prostate Specific Antigen
PSD	Post Source Decay
PVDF	Polyvinylidene Fluoride
QC	Quality Control
RPA	Reverse Phase Array
RP-SPE	Reversed Phase-Solid Phase Extraction
SELDI	Surface-Enhanced Laser Desorption Ionisation
SNP	Single-Nucleotide Polymorphism
SPA	Sinapinic Acid
SVM	Support Vector Machine
T+V	Trypsin and Versine
TAA	Tumour Associated Antigens
TB	Tumour Bearer
TFA	Tri-Fluoro Acetic Acid
TNM	Tumor Node Metastasis

TOF	Time-Of-Flight
TP	Time-Point

Abstract

Emergence of proteomics and high-throughput technologies has allowed the identification of protein expression patterns of disease that potentially hold clinical importance in predictive medicine. The analysis of complex data generated by these technologies incorporates the use of computer algorithms for data mining and identification of important protein biomarkers. Such candidate biomarkers can potentially be used for diagnosis, prognosis and monitoring a variety of diseases as well as the prediction of therapy response.

Mass spectrometry has been used widely, for the discovery and quantitation of disease associated biomarkers using a variety of samples such as serum and tissue. In particular, matrix assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-TOF MS) has been used to generate proteomic profiles or "fingerprints" from serum to distinguish patients at different clinical stages of disease. Currently, early stage disease is difficult to diagnose in most cancers as current cancer markers have limited sensitivity and specificity. In advanced stage metastatic disease, treatment options are limited, although it is recognised that some patients may benefit from immunotherapy and in particular vaccine therapy.

The use of animal models is critical to evaluate the efficacy of immunotherapies and to investigate tumour immunity in general and the mechanisms involved in tumour progression. These models provide an *in vivo* environment which cannot be reproduced *in vitro,* which results in more accurate and reliable information on the host response to immunotherapy and the mechanisms involved.

The research presented in this thesis has introduced the use of MALDI-TOF MS proteome profiling and bioinformatic analysis, to detect candidate biomarkers of tumour progression and responce to immunotherapy in a CT26 murine model of colorectal carcinoma. Proteomic profiles from serum and tissue were generated by MALDI-TOF MS followed by artificial neural networks (ANNs) analysis of the complex data. The methods used in this study for sample preparation and analysis demonstrates that good quality proteomic data from serum and tissue can be obtained, and that it is possible to generate discriminatory protein profiles that correlate with clinical outcomes. In the first instance, using the CT26 progression model, serum and tissue samples were collected at four time-points from tumour-bearer and control mice, providing the opportunity to assess the tumour proteome

changes with in a time-course from tumour initiation and through different stages of growth. Through the analysis of serum and tissue it is possible to classify samples based on their stage of tumour growth and the discriminatory patterns may reveal novel pathways associated with tumour progression. In addition, this study employed two separate mouse models of colon carcinoma immunotherapy (CT26 tumour model), to investigate biomarkers that are associated with therapy response. Using either disabled infectious single cycle-herpes simplex virus (DISC-HSV) or dendritic cell-based vaccination therapy with CTLA-4 and blockade of VEGFR-2 immunotherapy, up to 70% of the treated tumours tend to regress after receiving the immunotherapy (tumour regressors). Therefore, these models of immunotherapy were used to screen and evaluate serum protein and peptide biomarkers for the detection of progressors from regressors by using MALDI-MS coupled with an ANN algorithm. Comprehensive clean-up methods were conducted on the sera prior to MALDI analysis to reduce the complexity of the specimens. A panel of 4 biomarkers associated with response to DISC-HSV therapy was identified and successfully validated using non-mass spectrometry techniques. Furtheremore, discriminatory patterns corresponding to different stages of tumour progression and immunotherapy were identified in the mouse model with DC-based immunotherapy. Moreover, potential markers associated with response to therapy were proposed using this model. The work presented demonstrates a proof-of-principle that the different types of information that can be obtained from animal models can be expanded and applied to human studies.

Chapter 1 – Introduction

1.1 Cancer and carcinogenesis

Cancer is a collection of diseases, involving dynamic changes in the genome that stimulate uncontrolled cell proliferation and invasion of cells from the primary site, to other organs (metastasis), which results in secondary tumours. The process of cancer development and formation is known as carcinogenesis. Cancers originating from epithelial cells, mesoderm cells and glandular tissue are termed carcinoma, sarcoma and adenocarcinoma respectively. The ability of cancer cells to invade other organs is what makes them lethal because this results in tissue instability and disrupted organ function. Disrupted cellular regulation in cancer is commonly caused by genetic damage to two broad groups of genes named proto-oncogenes and tumour suppressor genes (TSGs), which are responsible for cell growth or death. The net number of cells in tissue is tightly controlled by equilibration between cell proliferation, cell death and cell differentiation. These physiological events are governed by cell cycle machinery. A normal cell cycle consists of 4 steps (G1, S phase, G2 and M) which are coordinated and regulated by a number of proteins and pathways (e.g. cyclins and cyclin-dependent kinases). At the end of each cell cycle event, checkpoints ensure the maintenance of the integrity of DNA. If any alteration or damage in DNA occurs, the process is halted and the damage is either repaired or the cell will be programmed to die (apoptosis). These molecular apoptotic machinery components act as sensors of DNA damage, signal transducers and effectors (executing the cell death) and determine whether the cell proliferates, becomes quiescent or differentiates (Lundberg and Weinberg 1998). Mutations in the genes involved in the production of the apoptotic machinery result in damaged cells continuing to proliferate and potentially adopting a malignant phenotype (Bertman 2000).

Cancer is a generic term for a group of heterogeneous diseases that can potentially occur in almost every type of tissue and each cancer type has unique characteristics which depend on the tissue of origin (Pedraza-Farina 2006). For example, pancreatic tumours are highly aggressive, while prostate tumours are more organ-confined. In addition, the tumour cell aggression varies from benign to highly malignant. Despite diversity in the properties of different tumours, six fundamental alterations in the normal cell physiology are required for

transformation to a malignant form. These essential cellular alterations during carcinogenesis have been illustrated by Hanahan and Weinberg and are known as 'hallmarks of cancer' (Hanahan and Weinberg 2000): 1) self-sufficiency in growth signalling, 2) insensitivity to antigrowth signals, 3) evasion of apoptosis, 4) limitless replicative potential, 5) tissue invasion and metastasis and 6) sustained angiogenesis (capacity to form new blood vessels). Although these properties are required by malignant cells and arise through genetic mutations, they can occur at different stages of tumour progression.

1.1.1 Oncogenes and tumour suppressor genes

Cancer is a genetic disease and, as previously mentioned, mutations in the proto-oncogenes and TSGs results in carcinogenesis (Weinberg 1994). Proto-oncogenes (*e.g. ras*) and TSGs (*e.g. APC*) are essential functional genes in the normal cells that produce proteins involved in regulation of cell proliferation, apoptosis or DNA damage repair pathways. Genetic alterations in these genes results in gain or loss of functionality and disruption of cell proliferation.

1.1.1.1 Gain of function- mutation in proto-oncogenes and conversion to oncogenes

By definition, an oncogene is a gene which is abnormally expressed or mutated in (genomic alteration) transformed cells. Normal cellular genes that potentially can give rise to oncogenes are known as proto-oncogenes and the process of transformation is associated with oncogene activation (Diamandis 1997). There are four main mechanisms that transform proto-oncogenes into the corresponding carcinogenic oncogene. These include gene amplification, chromosomal translocation, point mutations and chromosome rearrangement and the activation of proto-oncogenes leads to an unrestrained progression of the cell cycle and cell growth (Rieger 2004). Oncogenes are highly conserved genes which are mainly involved in cell division and differentiation pathways as well as cell signaling networks. The oncogene products can locate to the cell surface (acting as ligands or growth factors), the cell membrane (acting as receptors and signal transducers), the cytoplasm (acting as messenger proteins) and the nucleus (involving in transcription). Mutations in any of these genes can result in uncontrolled cell proliferation and independent growth potential. For example, in normal conditions, cells produce growth factors (GFs) to stimulate the proliferation of other cells; mutation in the GF gene can

enable the cell to produce excessive amounts of growth factors that stimulates its growth independent of other cells. For example, platelet-derived growth factor (PDGF) and tumour growth factor-α (TGF-α) are produced by sarcomas (Hanahan and Weinburg 2000). Oncogenes are dominant genes and only 1 allele needs to acquire a point mutation to cause a gain-of-function (activating) mutation in cancer (Fearon and Dang 1999). A number of oncogenes have been identified in humans and some of these are listed in table 1-1.

Gene	Gene product function	Mechanism of activation	Tumour associations
erb B1	Growth factor receptor	Amplification	Mammary carcinoma, glioblastoma
erb B2	Cell surface growth factor receptor	Amplification	Mammary, ovarian, stomach
Raf	Cytoplasmic serine/threonine kinase	Rearrangement	Stomach
H-Ras	GDP/GTP binding	Point mutation	Bladder
K-ras	GDP/GTP binding	Point mutation	Lung, colon
N-ras	GDP/GTP binding	Point mutation	Leukaemia
bcl 2	Cytoplasmic perhaps mitochondrial	Chromosomal translocation	Follicular and undifferentiated lymphomas
Myc	Nuclear transcription Factor	Amplification, chromosomal translocation	Lymphomas, carcinomas
Hst	Growth factor	Rearrangement	Stomach
sis	Growth factor	Over expression	Prostate, lung
cdk4	Cyclin dependent kinase	Amplification, point mutation	Sarcoma, familial melanoma

Table 1-1. Examples of human oncogenes and their associated carcinomas.

One of the well charecterised oncogenes is *ras*, which produces a protein (RAS), important in several signal transduction pathways involved in cell growth. Around half of the colon carcinomas, 90% of the pancreatic carcinomas (Goodsell 1999) and 30% of lung carcinomas and leukemias acquire a *ras* mutation (Diamandis 1997). Normally, RAS protein is responsible for delivering signals from receptors situated on the surface of cell (e.g. GF and G-protein receptors). These signals transmitted through various proteins to ultimately participate in DNA synthesis, cytoskeletal organisation and lipid metabolism. In the natural state, RAS binds to GDP and is inactivated or "switched off". A signal from the

cell surface receptors such as guanine nucleotide exchange factors (GEFs) or G-proteins is transmitted to RAS, resulting in the ejection of GDP which is replaced by GTP, now the RAS is switched on. Binding of GTP causes a change in protein conformation which results in interaction of RAS with downstream signaling molecules that eventually results in transcription factor activation and cell proliferation. The RAS protein subsequently switches itself off by hydrolysing GTP to GDP, assisted by GTPase activation. Point mutations in the *ras* gene affect GTPase activation and the RAS protein is activated for longer and therefore leads to initiation of cancer.

1.1.1.2 Loss of function-mutation in tumour suppressor genes can be oncogenic

The normal version of TSGs encoded proteins, act directly or indirectly as an inhibitor of progressive cell growth. The TSGs become oncogenic as a result of mutations that eliminate their function in cell growth control. Alteration in two copies of TSG is required for the loss of function of the gene and cells can normally proliferate when only one copy of the TSG is altered (Weinberg 1994). Several mechanisms can result in loss of function of TSGs that includes: cytogenic aberrations, genetic mutations, polymorphisms, loss of heterozygosity (LOH) and methylation (Wang *et al.* 2008, Zingde 2001). Some examples of TSGs are presented in table 1-2. Among the TSG, the p53 and RB1 are the best characterised.

The p53 gene is responsible for cell cycle regulation and functions as the checkpoint of the cell cycle. A point mutation in the p53 gene has been shown to be present in approximately 50% of human tumours (Bueter *et al.* 2006). The p53 gene is activated by DNA damage, oncogene activation, aberrant growth signals and cell stress (*e.g.* hypoxia and nucleotide depletion). Moreover, activation of p53 affects cell cycle arrest, apoptosis, DNA repair and inhibition of angiogenesis (Vogelstein and Kinzler 2004). Mutation in p53, either by deletion or loss of function of p53 protein, disrupts the normal function of p53 in the cell, leading to accumulation of DNA damage and inhibition of apoptosis and finally to the development of cancer.

Tumour suppressor gene	Gene production function	Associated human tumour	Associated cancer syndrome
*p*53	Transcriptional regulator/growth/ arrest/apoptosis	Sarcoma, breast, brain	Li-Fraumeni
BRCA1	Transcriptional regulator/DNA repair	Breast, ovarian	Familial breast cancer
BRCA2	Transcriptional regulator/DNA repair	Breast, ovarian	Familial breast cancer
LKB1	Serine/threonine kinase	Colorectal, breast	Peutz-Jeghers
PMS1	Mismatch repair	Colorectal	HNPCC
PMS2	Mismatch repair	Colorectal	HNPCC
RB1	Transcriptional regulation of cell cycle	Retinoblastoma, osteosarcoma	Retinoblastoma
APC	Binds/regulates β-catenin activity	Colon	Familial adenomatous polyposis
PTEN	Dual specificity phosphatise	Breast, prostate, glioblastoma	Cowden syndrome, BZS,LDD
DPC4	TGF-β signal transducer	Colon, pancreatic	Juvenile polyposis

Table 1-2. Examples of human cancer suppressor genes and their associated carcinomas.

1.1.2 Cancer a multistep disease

Transformation of a normal cell to an aggressive malignant form is a multistep process in which the end result is the accumulation of several genetic and epigenetic alterations. Evidence suggests that a single mutation in oncogenes and/or TSGs is not sufficient for carcinogenesis initiation. The genetic basis (acquired mutations) and the multistep development of colon cancer is well defined by Fearon and colleagues (Fearon *et al.* 1990) and is presented in figure 1.1. Briefly, mutation of the tumour suppressor gene *APC* allows the growth of polyps in the colon, subsequently a somatic mutation in the *k-ras* oncogene activates the gene and a more advanced, benign polyp may then eventually gain mutations in its *DCC* and *p53* tumour suppressor genes causing the uncontrolled growth of the colon carcinoma cell.

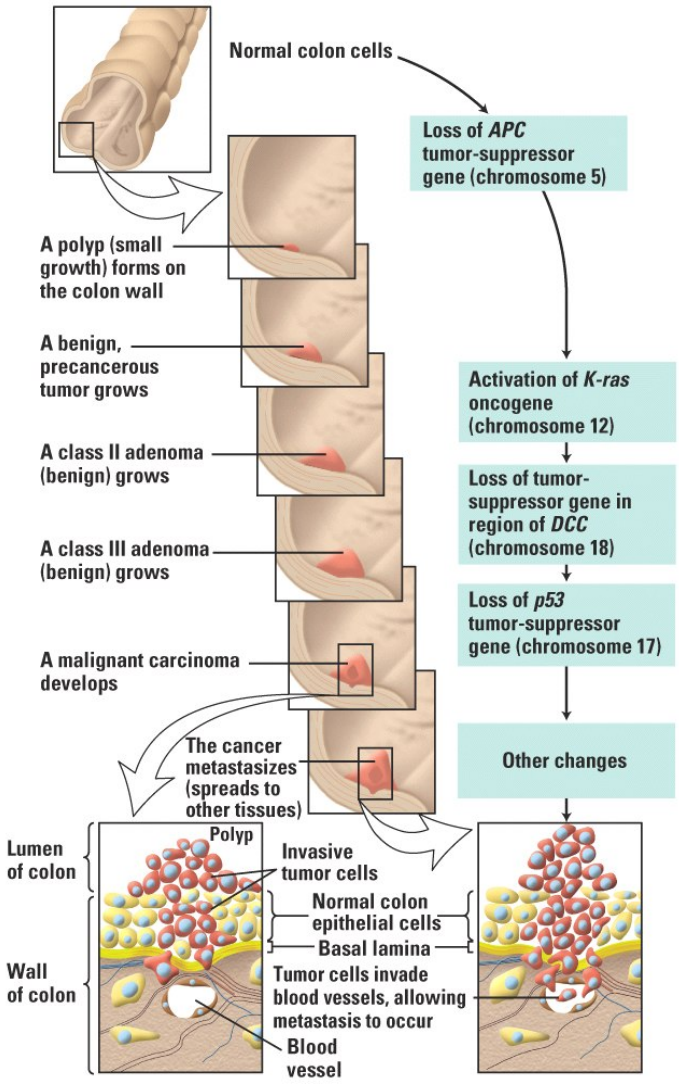


Figure 1-1. Genetic pathways of colorectal tumourogenesis. A mutation in tumour suppressor gene *APC* initiates growth of small polyps and sequential mutation in *k-ras* and *DCC* results in loss of *p53* functionality and formation of highly aggressive tumours that start to metastasise to other organs. The diagram is adapted from Lodish *et al.* 2004.

1.1.3 Cancer treatment

There are a number of therapies currently available for the treatment of cancer, including surgery, radiotherapy, chemotherapy, hormone therapy and immunotherapy. The choice of treatment usually depends on several factors including the type and stage of the cancer. They have been used in different ways for various type of cancer. However, despite recent improvements to enhance the efficiency of these treatments they still have their limitations. For example, surgery is effective in cancer patients where the tumour is localised and there is no metastasis to the surrounding tissues. Chemotherapy and radiotherapy are directed toward rapidly dividing cells and therefore can affect normal as well as the cancerous cells and significant toxicity can occur. Immunotherapy has several advantages over other types of treatment since it has low toxicity because it is highly specific and can be optimised to target only the cancer cells. However, the use of immunotherapy is not widespread and several approaches have only recently reached clinical trials and therefore improvements are required to fully develop the technology. Treatments such as hormone therapy are only effective on certain types of cancer.

Despite substantial recent effort in development and improvement for different cancer therapy approaches, to date no therapy has been shown to prolong the survival of patients with advanced disease. Therefore, diagnosis of cancer in early stages may significantly improve treatment and survival because treatments are more likely to be effective in the early stages (Wagner *et al.* 2005).

Recently there has been growing evidence supporting the idea that, cancer is a stem cell disease and the cancer stem cell (CSC) population within the solid tumour may have implications for diagnosis and treatment of cancer (Dalerba *et al.* 2007). According to the CSC theory, malignancies initiate from a small subpopulation of cancer cells (CSC) that have acquired tumour-related features, but on the other hand have maintained their self-renewal and pluripotency (Ricci-Vitiani *et al.* 2009). Cancer stem cells (CSCs) initially were identified in hematologic malignancies and recently have been identified in a number of solid tumours such as breast, colon, head and neck, brain and pancreatic cancer (Dalerba *et al.* 2007). If tumour formation is driven by these cells, then the goal of therapy would be to identify and target this population (Clarke *et al.* 2006). Isolation of CSCs from nontumourigenic tissues can be performed by flow cytometry, based on their distinctive surface markers. Identification and characterisation of CSCs would contribute to the

development of novel cancer therapies that target the CSCs (Kai *et al.* 2009, Ricci-Vitiani *et al.* 2009).

1.1.4 Immunotherapy and cancer treatment

Immunotherapy, as a treatment for cancer, was initiated in the 1890s, when William Coley administered live bacteria, capable of producing an acute streptococcal infection in the host that was followed by the regression of the cancer in some patients (Kim *et al.* 2002). The immune system of the host is capable of initiating an immune response toward tumour cells, which is the basic principle that is employed in the development of cancer vaccines. The identification of tumour antigens has allowed us, for the first time, to tailor therapies to specific molecular targets expressed on tumour cells. However, most tumours are considered to be "weak" antigens and consequently, the immune response against them is not strong. Furthermore, recent investigations in the field of cancer immunotherapy are to make the tumour cells more antigenic.

1.1.4.1 Tumour vaccines

1. Whole-cell vaccines

Prior to the identification of tumour-associated antigens (TAAs), many investigators believed that the best source of antigen was the tumour itself. The advantages of this approach to activate anti-tumour immunity is that all the tumour antigens regardless of being specific or shared, are presented to the immune system. However, the downside of this technique is that tumour antigens that are present in a low abundance might be 'under-presented' to the immune system (de Gruijl *et al.* 2008). In addition, whole-cell vaccines are often costly to produce. Use of irradiated whole tumour cell combined with adjuvants such as bacilli Calmette-Guerin (BCG) has been examined in colon and renal cancer therapy and is one of the early developed whole-cell vaccines. This strategy for tumour vaccination was able to generate a tumour-cell-specific immune response which was measured by delayed-type cutaneous hypersensitivity (DTCH) but showed insignificant effect on patient survival. More recent approaches to whole-tumour cell vaccines focus on the genetic modification of the tumour cell to enhance antigen-presentation by altering the microenvironmental conditions at the site of the tumour. Several clinical trails of GM-CSF modified tumour cell vaccines have been performed and shown to be successful in

pancreatic and renal cell carcinoma. The benefit of genetically modified whole tumour cells is that cross-presentation of antigen to T cells by antigen presenting cells does not require the tumour cells to be MHC matched with the recipient. Clinical trails in melanoma and renal cancer were carried out using genetically modified tumour cells engineered to express cytokines such as IL-2, IFN-γ or GM-CSF; however, only 20-30% of the patients showed a DTCH that is not significant (Parmiani *et al.* 2000). The production of whole-cell vaccines for cancer should be strictly controlled as they can evolve over multi passages with the potential loss of immunogenicity.

2. Heat-shock-protein-based vaccines

Cancer vaccines using heat shock proteins (HSPs) from tumour cells have been developed. These proteins are known to be involved in the folding of newly synthesised proteins and also in the transport of peptides and delivery into the MHC class I presentation pathway. Moreover, HSP-peptide complexes have also been used as cancer vaccines. Vaccination of rodents with these HSP-peptide complexes have resulted in powerful immune responses against peptides bound to HSP but not HSP itself (Srivastava and Udono 1994). As with all types of vaccination, these have advantages and disadvantages. These complexes have the ability to elicit a specific CTL response in mice of any haplotype. However, this approach requires patient-specific vaccines to be developed which can prove costly and difficult.

3. Peptide vaccines

Recently many TAAs have been identified and, when bound to MHC molecules, these TAA peptides are recognised by T cells (Parmiani *et al.* 2002). The efficiency of active immunisation or vaccination with TAA peptides has been tested in melanoma patients where 10-30% of the patients exhibited partial to complete regression. However, this response did not correlate with a detectable T cell specific antitumour response when patient T cells were evaluated with *ex vivo* assays. In murine models, most of the studies have concentrated on administrating single TAA epitopes with a variety of adjuvants. For virally induced tumours, prophalytic vaccination with synthetic peptides recognised by T cells was effective. However, for non-virally induced tumours this strategy was less effective.

4. Dendritic-cell vaccines

Dendritic cells (DCs) as antigen presenting cells have a crucial function of presenting antigens, including TAAs to naïve T cells in the lymph nodes. At the tumour site, DCs can internalise and process TAAs and then take them to the draining lymph nodes where they present with high efficiency, peptide-MHC complexes to T cell (Dalla and Lotze 2000). Hence DCs loaded with antigen act as potential activators of antitumour response. Mayordomo *et al.* (1995) showed that the mice injected with antigen loaded DCs were protected against subsequent challenge with the same tumour and even when given as a vaccine, in a therapeutic setting, was effective. It should be noted that the route of administration is also very important to the eventual outcome of therapy. Murine bone marrow derived DCs loaded with peptide when given subcutaneously had greater antitumour activity when compared to intravenously injected DCs as they homed to T cell areas of the draining lymph nodes whereas the intravenous injected DCs homed to the spleen (Eggert *et al.* 1999). However when similar studies carried out in cancer patients, the induced immunity was to be independent of the route of administration.

5. DNA vaccines

The DNA vaccine is an antigen-encoding gene on a bacterial plasmid and can elicit an antigen specific immune response. The usual route of administration involves direct inoculation of expression plasmids, which result in the induction of long lasting immune responses against the expressed antigen. The DNA vaccine administration is performed intramuscularly, intranasally, intramucosally or into the dermis by utilising gene gun. In a study carried out by Fynan *et al.* (1995) six routes of inoculation of naked DNA were compared for efficacy. Intramuscular injection of DNA generated the most potent responses whereas DNA coated gold particles using gene gun approach significantly lowered doses of DNA. DNA vaccines have several advantages over peptide vaccines, they are simple and cheap to produce and also they have the ability to generate long lasting immune responses.

6. Recombinant viral and bacterial vaccines

In recent years, the potency of recombinant viral and bacterial vectors has been examined in animal models and clinical trials as gene delivery systems. Several research groups have attempted to deliver the genes of IL-2 and GM-CSF to the tumour site via different methods such as using viral vectors. Many viruses have been used to construct recombinant

vaccines. The advantage of using this approach includes the induction of both humoural and cell mediated immunity. However, the disadvantages include the possibility of immunosuppression, conversion to virulence and the oncogenic potential. Therefore as the viral vectors have a high efficiency of delivering gene, they have to be safe and have low intrinsic immunogenicity such that the administration could be carried out to boost relevant specific immune response. To date, a number of recombinant viruses have been used successfully in animal models and proven useful as vectors and safe to use, these include pox viruses, retroviruses, adenoviruses and adeno-associated viruses (AAV).

Retroviruses are small RNA viruses that can replicate through a DNA intermediate and they can only infect actively proliferating cells (El-Aneed 2004). Retrovirus carriers are developed by replacing the viral proteins (*gag, pol* and *env*) with therapeutic substitutes. Recently a study describing tumour regression was achieved by use of encapsulated retrovirus with IL-12 and HSV-tk genes. Moreover, herpes simplex virus (HSV) vectors have been used successfully to deliver cytokines such as GM-CSF to the tumour site and provoke an immune response in animals with subcutaneous tumours and prolong life span of the treated animals. The CT26 mouse model of colon carcinoma has been extensively used in laboratories to study many aspects of cancer biology and therapeutics. This tumor model was used by a number of research groups to investigate the mechanisms involved in colorectal carcinoma metastasis to organs such as liver and lung. In addition, it has been widely used to examine the efficiency of novel cancer vaccines. The CT26 mouse model is a well established model in our laboratory for the past few years which was firstly used to assess a novel immunogene therapy using DISC-HSV as a vector which encodes mGM-CSF cytokine gene. Results of the direct injection of DISC-HSV into the established CT26 colon tumours were remarkable and up to 70% of the animals who received the vaccine showed complete regression of the tumour (tumour regressors) leaving the rest of the animals with progressive tumours (tumor progressors) (Ahmad *et al.* 2005).

1.2 Biomarker identification for cancer

A biomarker is defined as a biochemical molecule with specific characteristics that can be measured and then be used as an indicator of the health status, disease initiation, progression or response to treatment (Chatterjee and Zetter 2005, Holch *et al.* 2005). Significant efforts have been made to identify biomarkers in various types of diseases, in

particular cardiovascular disease, neurodegenerative disease and cancer. Different types of molecules (*e.g.* DNA, mRNA, metabolites and proteins) and biological processes (*e.g.* apoptosis, angiogenesis and proliferation) can potentially serve as biomarkers (Kulasingam and Diamandis 2008). Although biomarkers are used as indicators of biological status in both normal and pathological conditions, here cancer biomarkers are discussed. A potential cancer biomarker can be produced from the tumour itself or from other tissues as a response to the presence of the tumour and is therefore detectable in the tissue itself and in various biofluids (Holch *et al.* 2005). Defining specific and sensitive cancer biomarkers would contribute significantly to the clinical management of cancer patients in various ways. These include better understanding of the physiology and pathology of cancers, early detection of cancer, efficient disease screening, and development of novel and/or better cancer treatments and therapy outcome prediction which could result in patient tailored treatment. Cancer biomarkers can be divided into four main groups (Kulasingam and Diamandis 2008, Ludwig and Weinstein 2005) which include:

1) **Risk assessment and screening biomarkers:** This group of biomarkers have implications for the identification of individuals that are more likely to develop a specific type of cancer. Detection of patients that are at higher risk of cancer which will improve prevention and management of the disease.
2) **Diagnostic biomarkers:** Used in early detection and diagnosis of malignant disease and a good example of this type of biomarker is the detection of Bence-Jones protein in urine of myeloma patients.
3) **Prognostic biomarkers**: These types of markers are used to evaluate the course of the disease and probability of disease recurrence or recovery. Markers can potentially facilitate a clinician's decision of how aggressively to treat the disease (Swayers 2008). For example, Oncotype DX (Genomic Health) is an assay that is currently used to detect the likelihood of breast cancer relapse after surgical removal of tumour. The assay generates gene expression signatures (from 12 genes) that can be used to detect patients that can benefit from systematic therapy after surgical removal of the primary tumour, to eliminate any remaining tumour cells (Hornberger *et al*. 2005). Moreover, serum levels of human chorionic gonadotropin and α-fetoprotein is used to classify testicular teratoma patients with higher survival rate (Kulasingam and Diamandis 2008).

4) **Predictive biomarkers:** This type of biomarker is used to predict a patients' response to therapy prior to administration of therapy and assessment of therapeutic efficiency and toxicity. These markers have promise for developing patient tailored therapies and that can classify patients into responders and non-responders (predict therapy response).

In 1958, a combination of tumour size or depth (T), lympho node spread (N) and presence or absence of metastasis (M), known as TNM staging, was introduced as a system to classify cancer patients based on the progression of disease (Ludwig and Weinstein 2005). The TNM staging system is still used to stage cancer and is the basis for diagnosis, progression, predicting survival and deciding the choice of treatment for cancer patients. Furthermore, histological evaluation of tissue is then used to make a definitive diagnosis of the cancer (Chatterjee and Zetter 2005). The ability to subdivide tumour stages into subsets with distinct biological characterisations and distinctive genetic profiles has enabled the classification of patients that are more likely to respond to a specific therapy.

An ideal tumour marker would be a protein or protein fragment that is specific to cancer (high specificity), can be detected in majority of the caner patients (high sensitivity), can be easily measured by use of non-invasive tests (Schulte *et al.* 2005) and is cost-effective. Although detection of genetic changes in tumour tissue, such as mutation, amplification and translocation, may be potential tumour biomarkers, invasive procedures such as biopsy of the tumour may be required for their identification. Several techniques have been utilised to detect tumour markers without using invasive sampling by biopsy and/or surgery. For instance, analysis of circulating tumour cells, detection of mutations via analysis of circulating DNA, proteomic profiling of biofluids, assessment of tumour specific or tumour-associated antibodies and molecular evaluation of tumours by *in situ* imaging are among common techniques (Sawyers 2008). Advances in various areas of biology over the past decades, has significantly enhanced knowledge of the mechanisms involved in cancer initiation and progression. However, the number of biomarkers accepted into the clinic is not significant when compared to the level of scientific effort in the field.

Cancer biomarker discovery was initiated with the search for single blood biomarkers specific for one disease. However, considering the complexity of the biological mechanisms involved in cancer initiation and progression, it is unlikely that a single protein will reliably provide high specificity and sensitivity. The first tumour marker was reported

in 1848, which was the light chain of immunoglobulin, present in the urine of 75% of the myeloma cancer patients (Kulasingam and Diamandis 2008). The first recognised cancer biomarker was carcinoembryonic antigen (CEA), reported in 1965 by Dr Joseph Gold (Gold and Freedman 1965) which was present in the blood of colon cancer patients. The CEA is only expressed significantly in the fetal tissue and in very low amounts in intestinal adult cells. Cancer biomarkers that are currently in clinical use and their utility are shown in table 1-3.

Biomarker	Cancer type	Clinical application	Specificity
CA125	Ovarian, fallopian tube, uterine, cervical	Monitoring therapy, screening	Moderate
CA15-3	Breast	Monitoring therapy	Poor
CA19-9	Gastro, pancreatic, stomach	Monitoring therapy	Poor
CA27-29	Breast	Disease monitoring	
CEA	Colorectal, pancreatic, breast, medullary thyroid, lung	Disease monitoring, prognosis, monitoring therapy	Low
Epidermal growth factor receptor	Colon, non-small cell lung cancer	Selection of therapy	Low
Her2/Neu	Breast, ovarian	Disease monitoring, selection of therapy	Moderate
Human chrionic gonadotropin-β	Testicular, ovarian	Staging, diagnosis, monitoring therapy	Low
PSA	Prostate	Screening, disease monitoring	Moderate
Calcitonin	Medullary thyroid carcinoma	Diagnosis, monitoring therapy	
Lactate dehydrogenase	Germ cell	Diagnosis, monitoring therapy	
Thyroglobulin	Thyroid	Disease monitoring	Poor
α-fetoprotein	Non-seminomatous testicular n, Hepatocellular	Diagnosis, staging, detecting recurrence, monitoring therapy	Moderate

Table 1-3. Examples of human cancer serum biomarkers in clinic use.

These findings encouraged researchers to search for novel cancer biomarkers and develop blood based tests for diagnosis of different types of cancer; Prostate specific antigen (PSA) for prostate cancer, CA 15-3 for breast cancer and CA 19-9 for pancreatic and colorectal cancer are examples. Although these early defined cancer biomarkers were encouraging, further investigations revealed a number of problems. For example, PSA is a prostate

tumour marker and a marker of disease recurrence, but lacks specificity due to high concentrations in both benign and malignant disease which results in false positives. CA 15-3 has only 23% sensitivity and 69% specificity for the diagnosis of breast cancer (Jinong *et al.* 2002). Furthermore, CA 125 for ovarian cancer can also be elevated in pre and post menopausal women as well as individuals with ovarian cancer (Kong *et al*. 2006). All the above mentioned biomarkers have been approved by the US Food and Drug Administration, but are recommended to be used for screening therapy response and recurrence rather than diagnosis.

The second phase for biomarker identification was to identify mutations that are involved in carcinogenesis. As genetic changes are initiators of carcinogenesis, detection of the molecular changes benefits the early detection of cancer or the risk of disease initiation. The "signature" changes of genetic material occur prior to the symptoms of the disease and can be identified by a number of methods (discussed later). The next generation of biomarker research developed following the introduction of high-throughput technologies which includes genomic and proteomic based biomarkers. There are a number of gene-based biomarkers currently being used in the clinical management of cancer. For example, the presence of the amplified *ERBB2* gene (also known as *HER2* or *NEU*) and the expression of oestrogen receptor gene in breast cancer patients are used to determine which therapy to administer. In the case of leukaemia, the presence of the Philadelphia chromosome or the *PML-RARA* translocation can be used to prescribe therapy with *trans*-retinoic acid or imatinib mesylate.

Despite the breakthroughs in technologies used for identification of biomarkers, very few serum markers have been accepted into routine clinical practice. An approach that incorporates different technologies may prove to be key to identifying biomarkers that have real clinical significance. The lack of biomarkers with desirable specificity and sensitivity is believed to reflect tumour heterogeneity, which is due to both epidemiological and pathological variations (Petricoin *et al.* 2006). Therefore, identification of "sets" of tumour markers and tumour associated markers (genetic, proteomic or both) holds promise for the future. Defining panels of biomarkers may allow an increase in specificity and sensitivity which may be more applicable for general clinical use.

1.2.1 Current trends in cancer biomarker identification

In the past decades, a number of strategies have been adapted to detect cancer biomarkers. These strategies may provide novel areas for cancer screening, diagnosis and treatment. The ultimate goal of cancer biomarker discovery is to define a panel of markers with high specificity and sensitivity for different types of cancer. The emergence of new genomic and proteomic based technologies such as mass spectrometry based profiling and identification, gene and protein arrays, high-resolution 2D electrophoresis and high-throughput gene sequencing has encouraged researchers to identify and validate novel cancer biomarkers (Nicolette and Miller 2003). The following discussion focuses on commonly used approaches for the identification of cancer biomarker, outlining their limitations.

1.2.2 Genetic and molecular signatures

A number of commonly used genetic and molecular signatures used for biomarker discovery are discussed below.

1.2.2.1 Microsatellite instability (MSI)

Microsatellites are repeated sequences of DNA (less than six bases), the length of which may vary in individuals. The appearance of abnormally long or short microsatellites in an individual's DNA is referred to as MSI which is caused by inefficiency in the DNA repair process. They proved to be significant in identification of individuals that have a higher chance of developing cancer. The presence of MSI has been thought to increase the risk of colon, endometrium, breast and ovarian cancer.

1.2.2.2 Hypermethylation

An increase in the epigenetic methylation of cytosine and adenosine residues of DNA is known as hypermethylation. A number of tumour genes have increased levels of methylation specific genes, which makes them good targets for tumour therapy. A number of studies reported that using methylation inhibitors, the growth of the tumour can be suppressed. Some of the well know hypermethylated tumour genes are *BRCA1* in breast, *VHL* in renal, *APC* in colon and *PRB* in retinoblastoma cancers.

1.2.2.3 Single-nucleotide polymorphisms

Single-nucleotide polymorphisms (SNPs) are alterations in single nucleotides of the DNA sequence, which occurs in at least 1% of the population (Negm *et al.* 2002). Considering the frequency of SNP occurrence in the DNA (approximately are every 100-300 bases), the majority of the SNPs are synonymous and do not affect the function of the gene and its encoded protein (Martin and Nelson 2001). However, non-synonymous SNP change the function and/or structure of the encoded protein and may predispose individuals to cancer or other disease. For example, novel SNPs in the gene encoding the regulatory subunit 3 of the protein phosphatase 1 have been associated with tumour formation in cancer patients (Srinivas *et al.* 2001). Detecting these SNPs as candidate cancer biomarkers is challenging, but maybe useful in screening, diagnosis, prediction and designing future therapies for cancer.

1.2.3 Omics technologies

The commercial techniques for detection and screening of altered protein and genes in various diseases are often expensive, time-consuming and labour intensive. There is a need for the development of methodologies which are automated, cost-effective and high-throughput. 'Omics' is the general term, referring to a group of high-throughput techniques which are used in the field of biology for various research purposes, these include: genomics, transcriptomics, proteomics and metabolomics. These technologies offer the ability of simultaneous detection of a number of proteins/genes from a complex mixture, in the form of protein/gene expression profiles. They have been extensively used to provide a better understanding of molecules and molecular pathways. Each of these four technologies is distinct and presents a different perspective on the processes underlying disease initiation and progression as well as on ways of predicting, preventing, or treating disease.

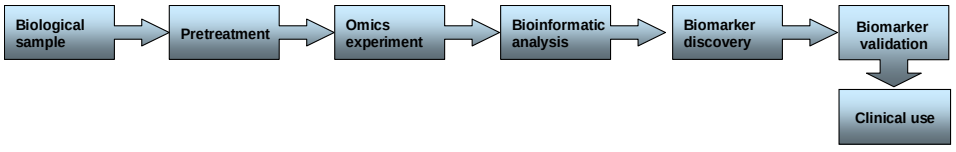


Figure 1-2. General workflow for cancer biomarker identification via omics technologies.

By definition the term 'genomics' was coined by Tom Roderick and it refers to high throughput analysis of the whole genome, regions of genome or multiple genes. The study of gene expression profiles is known as 'transcriptomics'. A surface (often glass) is spotted with DNA fragments or oligonucleotides (up to 200000 spots/cm^2) that represent specific gene coding regions. The interaction of the array with the RNA or DNA extract reveals the alterations of these materials in the sample.

1.2.4 Proteomics

The proteome refers to the whole protein concept of cell, tissue or organism. The term 'proteomics' was first coined in 1994 by Mark Wilkins and co-workers and is the study of complete protein component or the proteome of the cell (Wu *et al.* 2002). Proteomics is the complementary step for genomics, however, because genes are the producers of proteins, genomic studies facilitate proteomic studies. Although the initial interpretation of proteome imply that there is a one-to-one relation between the genome and proteome, it soon becomes apparent that the proteome is far more complex that the genome. This complication is due to several biological facts that are briefly discussed here. The estimated number of human genes is approximately 45000, however the approximate number of human proteins is 1.5 million (Alaoui-Jamali *et al.* 2006). Moreover, analysing the global gene expression at the messenger RNA (mRNA) level (transcriptomics) revealed that there are around 250000 variants of RNA which does not correspond to the number of human proteins. Therefore the RNA expression levels do not necessarily correlate to protein expression levels. The large increase in protein diversity is due to post-translational modifications (*i.e.* phosphorylation and glycosylation) and protein splicing that occurs on the initial proteins (Martin *et al.* 2001). Therefore the complexity of the proteome is partly due to production of variety of proteins from one gene. Moreover, the genomic sequence does not provide information on protein interactions and localisations and even if a gene eventually translates to protein. The advantage of proteomic over genomic evaluation is that the identified proteins represent the "end" products of a biological system whereas detecting a gene requires further investigations on its functionality after translation; proteomic studies are highly complex but more specific. Despite the great potential of proteomic research in biomedical studies, there are difficulties which are related to the dynamic properties of the proteome that adds another level of complexity for proteome

analysis. Quantity and type of proteins produced by a cell may vary at different time periods to react to its environmental changes or developmental programs which results in variation in protein abundance (Heck and Krijgsveld 2004). Moreover, protein expression is cell-type dependent and differs between different type of cells and this expression can further change as a result of disease initiation which is an additional factor to proteome complexity.

There are three main areas related to proteomic studies: protein expression profiling, functional and structural proteomics (Pastwa *et al.* 2006). Using "expression" proteomics, the aim is to identify proteins that are differentially displayed in a certain biological condition, *e.g.*, cancer *versus* normal. This approach provides the identity of proteins that may then become clinically significant as disease biomarkers which can be utilised in diagnosis, prognosis and prediction of response to therapy. Protein expression profiling is the most commonly used strategy in clinical proteomics. Functional proteomics investigates how proteins interact with each other and/or with DNA and RNA. Proteins can be further modified by post-translational processes, which results in various proteins from one gene, each protein can participate or mediate up to 5-10 interactions in a cell (Pastwa *et al.* 2006). This approach provides insights into how proteins interact and function in complex biological pathways that determine cell-cell interactions. This may define protein function in biological pathways involved in disease progression and potential targets for drugs in treatment. Structural proteomics provides information on the tertiary structure of proteins as well as functional structure and may facilitate the production and design of drugs that are more selective and specific.

The general trend in analysis of a complex biochemical system, such as proteins of importance in a disease state, involves the identification of their molecular weight (MW), amino acid sequences, determination of any existing post-translational modifications and finally their localisation and function in a given system. The methodologies and instrumentation used in proteomics aim to define the properties of proteins and are categorised into methods used to define protein expression, function or structure. Both gel-based and non-gel-based (*e.g.* chromatography) techniques coupled with a variety of mass spectrometry instrumentation are generally used for protein expression profiling. Protein arrays, protein chips and yeast two-hybrid systems are common methods used in functional proteomic studies and determination of posttranslational modifications. Finally, X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and *de novo* structure

prediction are strategies used in structural proteomics. In addition, bioinformatics play a crucial role in design and/or analysis of data when most of the mentioned techniques are applied.

1.3 Proteomics and its applications in cancer research

The potency of proteomic technologies has been exploited in the field of oncology. Although cancer is initially caused by DNA changes, detection of the abnormalities at the DNA and RNA levels do not reveal the complete picture of mechanisms involved in cancer. The genomic abnormalities in neoplasia, result in changes at the protein level, which include altered protein expression, function, modification, localisation and ultimately disrupted cell signalling pathways. In addition, cancer patients often show altered phenotypes compared to healthy individuals, which is a reflection of the changes in the proteome.

The advent of novel proteomic strategies, specifically high throughput methods, to detect cancer biomarkers has allowed significant progress to be made in the field of cancer biomarker identification. The development and application of novel technologies have enabled researchers to detect proteome differences between cancer patients and healthy individuals, utilisation of variety of biological samples ranging from biofluids, such as serum, plasma, cerebrospinal fluid (CSF) and urine, tumour biopsies and cell lines. Furthermore, other "exotic" body fluids with minimal availability, such as bronchoalveolar lavage fluid (BALF), breath condensation and breast aspirant have been used in proteomic studies (Wattiez and Falmage 2005). One of the remaining challenges in the proteomic based studies in cancer is the collection of high quality patient-derived material, where degradation of the proteome is minimal.

In cancer, it is reasonable to expect biomarkers to be present in tissue interstitial fluid and blood since biomarkers that originate from the tumour itself or its microenvironment can be released from the tumour cells into the blood. For example, alteration in the basement membranes corresponding to the presence of prostate cancer results in release of PSA into blood (Simpson *et al.* 2008). Serum and plasma are easily accessible and carry blood specific proteins as well as protein and peptides derived from a variety of tissues. According to "peptidome hypothesis" (figure1-3), protein fragments (resulting from the proteolytic cascades in the cell) and peptides are shed from normal and cancerous cells into

the tissue microenvironment. These diffuse into the bloodstream where some may combine with carrier proteins such as albumin. Therefore, analysis of the serum/plasma proteome may reveal protein/peptide fingerprints that might be associated with the presence of cancer and/or be specific to the disease. A number of these protein/peptides may also be present in urine.

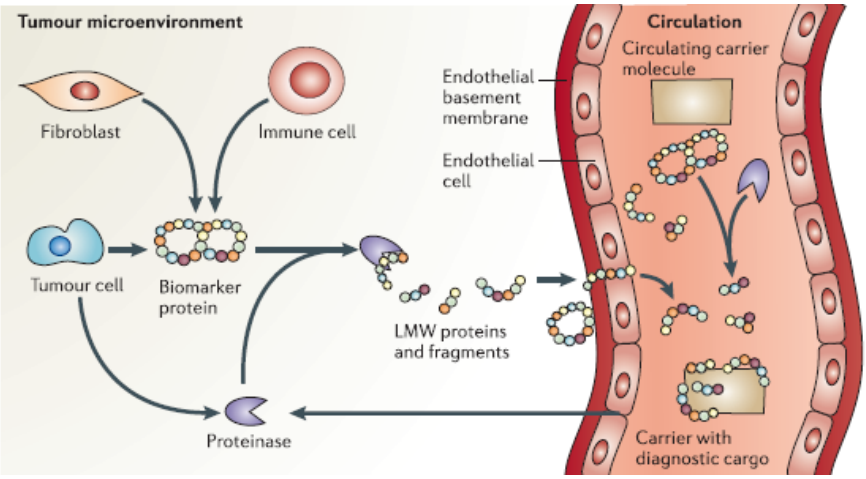


Figure 1-3. The peptidome hypothesis. Circulating protein and peptides are shed from the tumour cell and other cell types in the tissue microenvironment and then cleaved by proteases and enter the blood or other surrounding biofluids.

Serum and plasma proteomic profiling studies are possibly diverse and challenging, due to the dynamic range of protein and peptides present. There can be ten orders of magnitude difference between the most abundant serum protein and the low abundant ones. For example, the concentration of albumin in human serum is normally 40 mg/mL whereas proteins such as cytokines are present at approximately 5 ng/mL or even pg/mL concentrations (Schult *et al.* 2005). Fractionation methods have been applied to overcome these difficulties and some of these technologies are discussed in the following sections. Regardless of the technique used to identify biomarkers, proteomic studies consist of three main steps: 1) sample preparation, 2) protein separation, and 3) protein identification.

1.3.1 Two-Dimensional Polyacrylamide Gel Electrophoresis (2D PAGE)

For the last 25 years, 2D PAGE has been utilised for separating thousands of different proteins (up to 5000) in a single analysis according to two independent physicochemical

properties: isoelectric point (pI) and molecular weight. Proteins are firstly separated (first dimension) based on their pI called isoelectric focusing (IEF) and then in the second dimension using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) according to their molecular weight. The first dimensional IEF is performed by applying a pH gradient strip and then an electric potential is applied across the gel, making one end more positive than the other. At all pHs other than their pI, proteins will be charged and are drawn towards the opposite electrical charge. The proteins will arrive at the point in the gradient equal to its pI where the net charge of protein is zero. After IEF separation, the gel strips are immediately equilibrated in a buffer containing SDS, transferred and separated using SDS PAGE. SDS is an anionic detergent with capability of binding to proteins and the resulting complex is soluble in water and can migrate in the polyacrylamide gel. The incorporation of SDS to the gel masks the charge of the protein and therefore protein migration on the gel is just based on the molecular weight only.

2D gel electrophoresis is generally capable of separating proteins with pI range of 3.5-10 and molecular weight of 6-300 kDa (Ashcroft 2003). For the last two decades, 2-D PAGE has been the technique of choice for analysing the protein content of human samples and over this time there have been many improvements applied to increase in resolving power. These improvements include the use of immobilised pH gradient (IPG) strips (increases the resolution and the reproducibility), the introduction of new reducing agents and new surfactants to improve sample solubilisation (Cristea *et al.* 2009). Once the gel is run, it can be stained by techniques such as silver, coomassie blue and fluorescent staining. Following separation of proteins by 2D gels, each protein spot can be excised from the gel and digested into peptides. Various types of proteases (e.g. trypsin, chymotrypsin, endoproteinase Glu C, endoproteinase Lys C and endoproteinase Asp N) can be used however trypsin in the most widely used (Wu *et al.* 2002). The tryptic peptides from the gel can be characterised by mass spectrometry using peptide mass fingerprinting (PMF). The combination of 2D PAGE and mass spectrometry has been used to study protein expression differences between cancer patients and healthy individuals. The methodology described above was used recently to analyse serum proteome of patients with nasopharyngeal carcinoma to identify biomarkers associated with carcinogenesis and lymph node metastasis (Liao *et al.* 2008).

The new variant of 2D PAGE is 2D difference gel electrophoresis (DIGE). The principle of DIGE technology is based on covalent labeling of proteins with different fluorescent dyes.

Cyanine (Cy2, Cy3 or Cy5) and Alexa dyes are commonly used materials for DIGE analysis. Briefly, the first sample (*e.g.* cancer serum) is labeled with Cy3 and the second sample (*e.g.* healthy serum) is labeled with Cy5. In addition, the two samples are mixed with the equal concentrations and then labeled with Cy5; this would be used to assess the variation between different runs. The three labeled samples are then run as a normal 2D process and then the gel is scanned with the free wavelengths and this will generate an image of the gel in which quantitative measurement of content of each gel spot is possible. The intensity of Cy3 or Cy5 in each spot represents the protein concentration and shows differential expression of that protein in each sample (Pastwa *et al.* 2006).

Despite the advances, the use of 2D electrophoresis in conjunction with mass spectrometry and its application in proteomic studies is subject of debate due to a number of limitations. Difficulties in automation of the procedure, low-throughput, difficulty in detection of low abundant proteins in the presence of high abundant proteins such as albumin and immunoglobulin in serum samples, gel-to-gel variation, requirement of high volume of sample are just to name few (Matharoo-Ball *et al.* 2007, Wu *et al.* 2002). The issue of presence of high abundant proteins in serum and plasma samples was partly addressed by the use of immunoaffinity depletion columns prior to 2D electrophoresis. A recent study utilised combination of immunoaffinity depletion of high abundant proteins from serum, 2D DIGE analysis and mass spectrometry analysis, to identify 13 human serum markers of progression of prostate cancer (Byrne *et al.* 2009). The expression of these 13 proteins in the serum were significantly changed between the healthy and cancer patients and 2 of the proteins, pigment epithelium-derived factor (PEDF) and zinc-α-glycoprotein (ZAG) were further validated in tissues and a different cohort of samples.

1.3.2 Protein microarray

In recent years, microarray technologies have been applied to genomics as well as proteomics. Protein microarrays allow the systematic analysis of several thousands of proteins simultaneously and determine the posttranslational modifications of the proteins analysed, their levels in biological samples and their selective interaction with other proteins and antibodies. The technology is shown to be high throughput for screening yeast proteome molecular interactions (Zhu *et al.* 2001) and there are two general trends to design protein arrays which are forward phase arrays (FPAs) and reverse phase arrays (RPAs). The most commonly used technique is FPAs, where antibodies are arrayed and are

probed with the sample; RPAs are where proteins of the sample are arrayed and then probed with the antibody. These have been used in relatively few studies. The limitations with the use of RPAs are due to instability of the structure of the proteins and their physiochemical diversities toward their interacting partners on the array (Kolch *et al.* 2005). Hudelist *et al.* (2004) employed a FPA (using 378 monoclonal antibodies) to assess the protein expression of human breast tissue compared with healthy breast tissue. Overexpression of a number of proteins including casein kinase I, p53, annexin XI, CDC25C, eIF-4E and mitogen-activated protein kinase (MAPK) was shown and these were then validated by immunohistochemistry analysis (Hudelist *et al.* 2004). The use of RPAs has been commonly used for studying phosphoproteome (Conrad *et al*. 2005). Currently the main problems in using this approach, is the lack of available antibodies and low sensitivity. Also, post-translational modifications are not captured using recombinant proteins or antibodies that do not distinctly recognize specific conformational forms of a protein (Misek *et al.* 2004, Pastwa *et al.* 2007). Finally, there is also a problem with cross-reactivity of the proteins with affinity agents (Baak *et al.* 2003).

1.3.3 Mass spectrometry

Mass spectrometry (MS) is an analytical technique which provides information on the mass of the molecules and their relative abundances. In addition, using tandem mass spectrometry it is possible to obtain partial amino acid sequence of the protein of interest. Currently, there are two main strategies for protein identification by mass spectrometry: top-down approach and bottom-up approach (Chen 2008). The top-down approach involves direct gas phase fragmentation of intact proteins, followed by mass spectrometry analysis whereas bottom-up approach involves enzymatic or chemical cleavage of a complex mixture followed by mass spectrometry analysis of the produced peptides. Briefly, a mass spectrometer converts the sample to gas-phase ions, separates them based on their mass-to-charge (*m/z*) ratio and detects them qualitatively or quantitatively and produces a mass spectrum. Regardless of the type of instrument or the type of analysis, a mass spectrometer (figure 1-4) consists of three main separated compartments: the ion source, which produces ions from the sample; the mass analyser, that resolves ions based on their *m/z* ratio; and the detector, which detects the resolved ions and the data processing system will produces the mass spectrum.

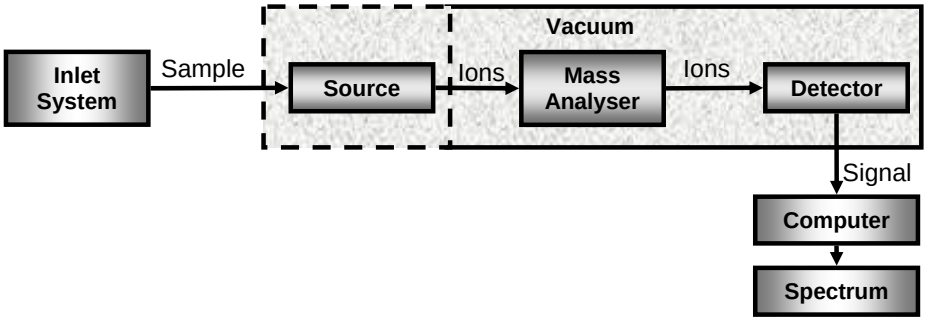


Figure 1-4. Basic diagram of a mass spectrometer. Ionisation of the sample occurs in the ion source. The source can be under vacuum or atmospheric pressure. The generated ions are separated based on their mass-to-charge (*m*/z) ratio by the mass analyser and detected ions produce a mass spectrum.

There are various types of techniques that can be used for sample ionisation, ion mass analysis and detection of ions which will be discussed.

1.3.3.1 Ionisation

The ion source of a mass spectrometer is where the ionisation of the analyte occurs and this part can be under atmospheric or vacuum pressure. Ions are formed from a neutral molecule in the gas-phase by various techniques. In the process of ionisation, the internal energy transferred during the process and the ability of the analyte to ionise is important. There are two types of ionisation; firstly techniques that only produce ions of molecular species which is referred to as soft ionisation methods and secondly, techniques that cause extensive fragmentation and are very energetic (Hoffmann and Stroobant 2007). One way of ionising a molecule is to add or remove electrons from/to the molecule. This process can be initiated using techniques such as electron ionisation (EI), bombarding the sample with atoms or ions (Fast Bombardment Ionization, FAB), or by photons (Laser desorption/Ionization, LDI). Techniques such as plasma desorption ionisation introduced by Macfarlane and Torgerson (Macfarlane and Torgerson 1976) and fast atom bombardment (FAB), introduced by Barber and colleagues in 1981 are probably the first methods for ionisation for large biomolecules. The use of radioactive material (*i.e.* ^{252}Cf) in plasma desorption ionisation was the limitation and therefore it is not commonly used in laboratory practices. Despite the diverse techniques of ionisation only few of these have been used for analysis of biomolecules. These include FAB, ESI and MALDI. Although FAB ionisation is a relatively soft ionisation technique it has now been replaced by

electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) methods. Detailed principles of ESI and MALDI will be discussed later in this chapter.

1.3.3.2 Mass analyser

Once the sample is ionised in the source of the mass spectrometer, the ions are separated in the mass analyser according to their *m/z* ratio. The most important features of a mass analyser are mass accuracy, mass resolution, sensitivity, mass range limits and signal-to-noise (S/N) ratio. The different types of mass analyser include: quadrupole, ion trap, time-of-flight (TOF) and Fourier transform ion cyclotron resonance (FT-ICR) mass analysers. The most commonly used mass analysers are the ion trap and TOF mass analysers. The quadrupole consists of 4 circular rods, set parallel to each other. Ions are separated in a quadrupole based on the stability of their trajectories in the oscillating electric fields that are applied to the rods.

1.3.3.3 Ion detection

Several types of detectors currently exist, and the choice of detector depends on the analytical applications of instrument and the design of the mass spectrometer. The detection of ions is based on their charge, mass or velocity. Photographic plate, Faraday cup, electron-optical and electron multiplier are the commonly used detectors.

1.4 MALDI-MS

The term MALDI was introduced simultaneously by Franz Hillenkamp and Michael Karas and Tanaka *et al.* in 1988 (Karas and Hillenkamp 1988, Tanaka *et al.* 1988) and rapidly became one of the widely used techniques for analysis of non-volatile, high-molecular weight compounds. Analysing a variety of biomolecules such as proteins, peptides, oligonucleotides, oligosaccharides, carbohydrates and lipids and large synthetic polymers is possible by use of MALDI (Knochenmuss and Zenobi 2003). The general principal of MALDI revolves around co-crystallisation of analyte with matrix which is usually a low mass, weak organic acid that is capable of absorbing laser at a specific wavelength. The choice of matrix used for MALDI analysis depends on the nature of analyte and the laser type used in the MALDI instrument. Common matrices for various types of analytes are described in Table 1-4.

Matrix	Analyte
2,4,6-trihydroxyacetophenone (THAP)	Oligonucleotide
2,5-dihydroxybenzoic acid (DHB)	Protein digests and proteins, oligosaccharides
2,6-dihydroxyacetophenone (DHBP)	Protein and peptides
4-hydroxy-3,5-dimethoxycinnamic acid (Sinapinic acid, SA)	Large polypeptides and proteins
α-cyano-4-hydroxycinnamic acid (CHCA)	Peptide and protein digests

Table 1-4. MALDI matrices and their applications.

The mixture of matrix and analyte are then spotted onto the MALDI target plate and the solvents are evaporated in the air which causes the formation of matrix crystals. The ratio of mixing the matrix with analyte does vary but a ratio of 1 to 1 is often used. A well functioning matrix will incorporate the analyte in the crystals in a homogenous distribution. The target plate is then placed into the source of the mass spectrometer and irradiated with a laser beam. Lasers at wavelengths ranging from ultraviolet (UV) to infra-red (IR) can be used but UV nitrogen lasers (337 nm) are the most common. The homogenous incorporation of analyte and matrix in conjunction with matrix's capability to absorb the laser light energy, results in the formation of a gas-phase plume containing analyte molecules, which then become ionised via gas-phase proton transfer reactions. The MALDI process is illustrated in figure 1-5.

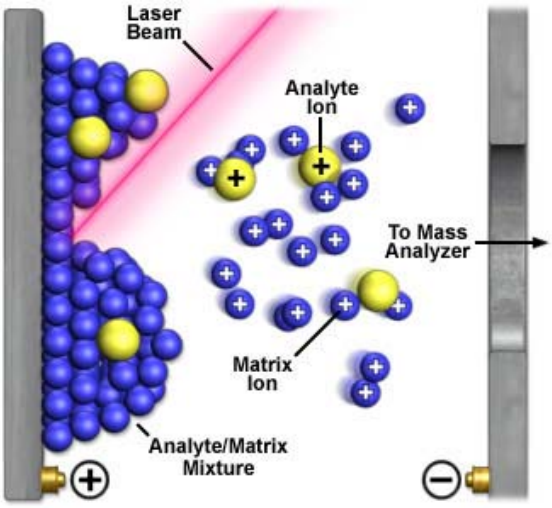


Figure 1-5. MALDI ionisation process. Laser irradiation to the analyte/matrix mixture leads to generation of gas phase ions which then are accelerated into the mass analyser.

The mechanism of ion formation in MALDI is poorly understood but several processes have been proposed in the literature. The ionisation mechanisms in MALDI are generally divided into two categories which are 'primary' and 'secondary' ionisations. The primary ionisation is referred to initial ions, generated from the natural molecules in the sample and they are usually matrix-derived species. The secondary ionisation is the process that ions are produced by ion-molecule reactions (matrix-matrix and matrix-analyte interactions) in the expanding plume. Some of the proposed primary ionisation mechanisms include multiphoton ionisation, energy pooling and multicenter models, excited-state proton transfer and thermal ionisation. The most acceptable mechanism is the multiphoton ionisations in which upon absorption of laser beam photons by matrix, the matrix becomes electrically excited. The excess energy from the matrix is then transferred to the analyte (*e.g.* protein or peptide) and in the gas phase, each molecule picks a single proton and therefore they become singly charged. The predominant generated ions in MALDI are singly charged and generations of multiple charged ions are very rare.

The generated ions are then accelerated into the mass analyser due to an electrical field. The mass analyser in a MALDI instrument is usually a time-of–flight (TOF) device. There are 2 types of TOF analysers typically used in MALDI mass spectrometers, a linear TOF (figure 1-6) and a reflectron TOF (figure 1-7). The principal of mass analysis in a TOF analyser is based on the fact that accelerated ions from the source to the flight tube have

different *m/z* values, same energy but different velocities. Therefore, while the ions are accelerated with the same potential at a fixed point and a fixed initial time and drift, the ions will separate according to their mass to charge ratios; higher *m/z* in comparison to lower *m/z* travel the drift region in a longer time. In the linear TOF, the detector is situated at the end of the flight tube (figure 1-6). The main drawback of linear TOF analysers is their poor mass resolution as a consequence of space (size of the space that the ions are formed), time (length of the ion formation) and kinetic energy (difference in the kinetic energy between the ions of same mass) distribution of the ions with the same *m/z* value.

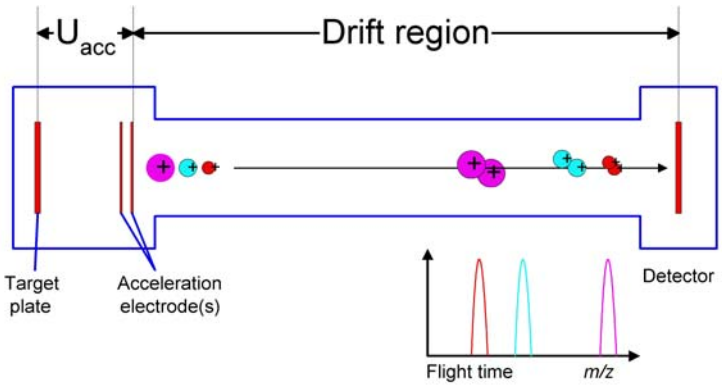


Figure 1-6. Schematic of a linear TOF mass analyser in MALDI. Separation of the ions occurs based on the fact that ions that are accelerated into the draft region have same energy but different *m/z* and velocity. Therefore, ions with higher *m/z* travel the flight tube slower that the lighter ions and they will hit the detector at the end of the flight tube.

Higher resolution can be achieved in TOF mass analyser by increasing the flight time of the ions. Reduction of acceleration voltage or increasing the length of the drift region would lead to higher resolution which is the basic principal of reflectron TOF mass analyser. The samples with the same mass can be accelerated into the flight tube with different kinetic energies or have different initial kinetic energies (known as kinetic energy distribution). In the reflectron TOF as shown in figure 1-7, there is an ion mirror which will turn the ions to the second detector.

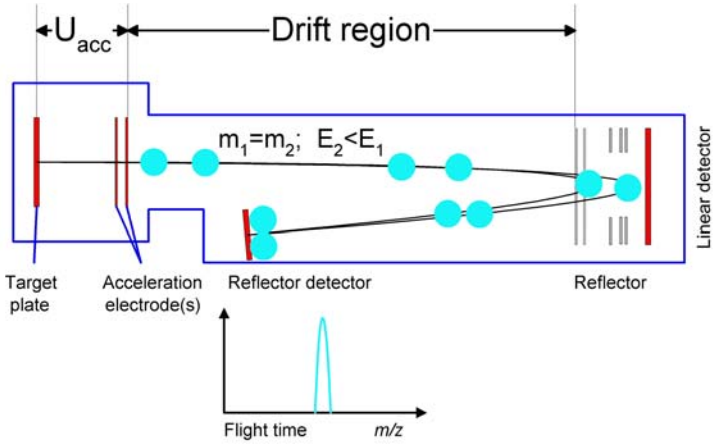


Figure 1-7. Schematic of a reflectron TOF mass analyser in MALDI.

Ions that have higher energy can continue flying and reach the detector. The benefit of using reflectron TOF mass analyser is generation of spectra with higher resolution. However, the mass range that can be analysed in compromised in comparison to linear TOF mass analyser. The reflectron mode is often used for peptide analysis.

1.5 SELDI

The surface-enhanced laser desorption/ionisation time-of-flight mass spectrometry (SELDI-TOF MS) is an array based method introduced by Hutchens and Yip (1993). This novel analytical high throughput technology is currently used in the development of a ProteinChip system, Ciphergen Biosystems Inc (Freemont, CA, USA). The principle of this approach is based on capturing a specific group of proteins on a solid-phase protein chip chromatographic surface. Although the sample preparation of SELDI is different to MALDI, the two techniques are similar in the ionisation and mass analysis procedures. The captured proteins on the chromatographic surface are co-crystallised with matrix and the laser beam ionises the sample in the same manner that occurs in MALDI and their molecular weight is then measured by TOF MS.

The SELDI-TOF MS consists of three major components: protein chip array, mass analyser and software for data analysis. Each chip consists of 8 or 16 spots with a specific chromatographic surface. There are number of different protein chips types that can be used in SELDI based analysis and are shown in figure 1-8. Different types of samples such as body fluids and cell extracts can be analysed using various SELDI chips. The different

chips are designed to retain proteins with specific or general physiochemical properties from a complex mixture.

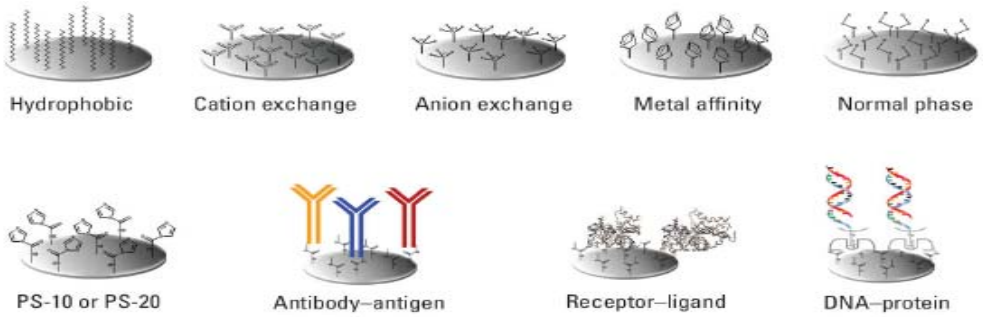


Figure 1-8. The variety of common protein chips used in SELDI sample preparation. The top row shows the chemically modified chips which capture group of proteins with similar physical properties. The bottom row shows the biochemically modified chips that retain the protein of interest.

The chips are either chemically or biochemically treated (Bischoff *et al.* 2004). The chemically treated surfaces include hydrophobic, hydrophilic, anionic and cationic. These will separate a whole class of proteins with a specific physicochemical property. This will facilitate the analysis of complex biological specimens such as biofluids that consist of an immense number of proteins with different properties. For instance, the hydrophobic chips are designed to capture all the hydrophobic proteins in a complex mixture, the unbound proteins are then removed by sequential washes, and after application of the matrix, the retained hydrophobic proteins on the surface are analysed using mass spectrometry.

The biochemical chips are treated with an antibody or affinity reagents and unlike the chemical chips, binds specifically to the protein of interest or the functional class of proteins. The biochemical treated chips are often custom made and are used for investigations of protein-protein, protein-DNA, antigen-antibody and receptor-ligand interactions. Once the proteins of interest are retained on the surface, a series of washes are carried out to remove the unbound proteins and contamination. This matrix is then applied on the chip and the retained proteins are then analysed by a linear-TOF mass spectrometer in a similar way to MALDI analysis. In recent years SELDI has been used for recognition and biomarker identification in the field of cancer, through the analysis of body fluids such as plasma, serum and urine as well as tissue samples.

There are number of limitations to this technique. The main limitation of SELDI technology is that the individual proteins cannot be identified. For protein identification, it

is required to perform conventional chromatography techniques to purify the protein of interest. Poor resolution and the sensitivity for proteins with molecular weights greater than 30 kDa is another limitation. Although SELDI has been classed as one of the high throughput techniques, the maximum number of samples that can be analysed in a day would be up to 100 (including control standards) whereas MALDI target plates can hold 384 samples. In addition tryptic digestion of the retained proteins on the chip is not possible; therefore peptide fingerprints of retained proteins cannot be obtained. Also, due to the difficulty in protein quantification the expression levels of identified protein peaks is not possible. Despite these limitations the SELDI ProteinChips' strengths lie in the systematic identification and characterization of proteins for diagnostic and prognostic markers in tissues and body fluids and the speed at which potential targets for therapy can be identified is increased (Seibert *et al.* 2004). Improvements in the instrumentation and bioinformatics methods associated with SELDI will be needed.

1.6 ESI MS/MS

Electrospray ionisation (ESI) is used for analysis of large biomolecules such as protein and peptides. This technique of ionisation was first reported in 1968 by Dole and significantly evolved over the years to become one of the key mass spectrometry techniques in proteomics. The mechanism of ESI is illustrated in figure 1-9. The source of ESI mass spectrometer is under atmospheric pressure. Typically, an analyte is dissolved in an appropriate volatile solvent and can be introduced to the source by two means, from a syringe pump or as the eluent flow from liquid chromatography (LC). The analyte is then passed through a narrow metal coated capillary under a strong electrical field. The potential difference between the capillary and the core (counter-electrode) produces the electrical field, which is normally held around 3-6 kV (can be positive or negative potential). The electrical field at the end of the capillary tip induces an accumulation of charge (ions with same polarity) at the liquid surface of the droplet emerging from the end of the capillary and the formation of a Taylor cone. Moreover, a fine aerosol containing charged droplets is released from the Taylor core. Evaporation of the solvent contained in the droplets occurs which leads to a reduction of the droplet size and an increase in their charge per unit volume. The solvent evaporation continues until the surface tension can no longer sustain the charge and the droplet becomes unstable (the Rayleigh limit), then the droplet explodes

into smaller droplets (the Coulombic explosion). The same procedure is repeated on the smaller droplets until individually charged 'naked' analyte ions are produced and extracted to the mass analyser system.

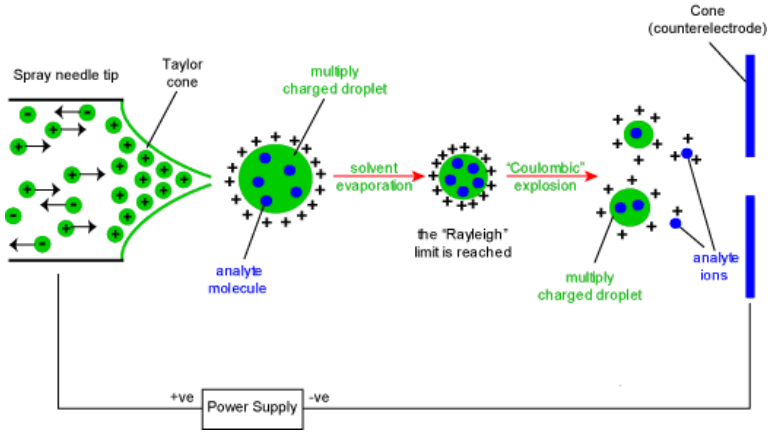


Figure 1-9. A schematic diagram of ESI method.

The mass analyser of the ESI mass spectrometer used for the experimentation discussed in this thesis was a quadrupole ion trap.

The ESI mass spectrometry is beneficial in several ways as it can be used in conjunction with chromatographic and electrophoretic techniques. In addition, ESI generates multiply charged ions of the type $[M+nH]^{n+}$ as well as the singly charged $[M+H]^{+}$ species and therefore analysis of analytes with high molecular mass is possible by quadrupole mass analysers (Ashcroft 2003).

1.7 Quantitative mass spectrometry in proteomics

The major aim of proteomics in cancer biomarker detection is to investigate global proteomic expression changes between healthy and disease, using a variety of complex sample types such as biofluids and tissue via high throughput techniques such as mass spectrometry. Since biomarker detection is often based on quantity differences among samples, accurate measurement of protein in complex samples is crucial, although particularly difficult and challenging. In the field of molecular biology, quantitative measurement of protein expression relies on use of antibodies. However, commercial antibody based molecular techniques (*i.e.* western blotting and ELISA) are considered more

semiquantitative due to quality and binding affinity of antibodies (Ong *et al*. 2002). The potential of current antibody based molecular techniques to develop high throughput assays is limited due to production cost and quality and specificity of the antibodies. Advance in proteomic technologies and investigations toward adaptation of current tools for quantification of proteins provides better insight and more accurate measurement of potential biomarkers in complex samples. Gel-based and non-gel-based strategies are the main two fundamental approaches used in quantitative proteomics (Lilley *et al.* 2002). Application of 2D gel electrophoresis is probably the first proteomic based technique, potential of analysing global protein expression in complex samples and simultaneous quantification of all proteins in the gel. However, cross-comparison of protein expression changes requires running a gel for each sample and therefore introduces variation. However, introduction of 2D-DIGE gel electrophoresis provides the ability of protein expression profiling of two set of samples on one gel via the use of internal standards. Despite the power of 2D gels, limited dynamic range and limited detection of low abundant proteins and reproducibility are the main drawbacks of this technique. Over the last few years alternative strategies for accurate measurement of protein abundance in complex mixtures have been introduced, mostly mass spectrometry based (MS-based) which potentially provides information that is missing from gel-based techniques. The MS-based strategies for protein quantification fall into two main categories: 1) label-free approaches which rely on measurement of ion signal intensity of protein and peptides or peak frequency count, 2) labelling approach that employs isotope labelling of samples prior to MS analysis (Choi *et al.* 2002, Lui *et al.* 2004). Some of the MS-based quantitative approaches and their applications are discussed here.

The label-free strategies are commonly used for quantification of proteins in LC-MS experiments. However, as the technique is based on peptide ion intensity count or by spectral count, the accuracy of quantification is often compromised due to chromatography conditions and ionisation efficiency (Heck and Krijgsveld 2004). The ion intensity count strategy presumes that there is a linear relationship between the peak height/area and the abundance of a protein or peptide in a specific sample and by counting the ions of a selected protein, the relative abundance of it can be determined. The spectral count approach relies simply by counting the number of MS/MS obtained from a selected proteolytic peptide ion. In both of these approaches this data can be obtained for a protein in different samples and then quantitatively compared between the groups. Despite relative

cost effectiveness of the label-free strategies to the isotope labelling techniques, there are several disadvantages to the use of label-free techniques. Ionisation of a specific protein can vary in different environments and slight variation in the chromatography conditions can affect the time of the elution of peptides from the chromatography column, both resulting in irreproducible spectra and peptide separation. In addition, the absolute intensity of a peak corresponding to the selected protein is determined by its efficiency of ionisation (this is associated to the physiochemical properties of protein/peptide) and suppression effects of other proteins present in the complex sample mixture on the selected protein. However, some of these limitations can be overcome by strict sample preparation and analysis procedures and over the past years several investigations were carried out by various research groups to improve the label-free proteomic quantification strategies. Known concentrations of a certain protein have been spiked into the complex sample mixture as an internal control to insure the reproducibility and reliability of the peak counts or ion intensity counts. Huang and coworkers utilised LC-MS label-free proteomic quantification approach to detect protein biomarkers associated with melanoma metastasis from paraffin-embedded tissues. After protein extraction from primary and metastatic paraffin-embedded melanoma tissues, they used LC-MS/MS to generate proteomic profiles and used an intensity based algorithm to detect the differentially expressed proteins. This study identified 120 significant proteins that were significantly different between the primary and metastatic melanoma tissues (Huang *et al.* 2009). Chelius and Bondarenko determined that their quantification method based on peak areas can be reproducibly used as long as the measured and expected protein ratio differs by less than 20% (Heck and Krijgsveld 2004).

Applications of various stable isotope labeling strategies have shown great promise for quantification of proteins present in a complex mixture by the use of mass spectrometry. The main sources of variation in MS-based studies are associated with sample preparation procedures and variations in ionisation efficiency in MS analysis. These errors can be significantly reduced by the use of internal standards. A desired internal standard has similar physical and chemical properties to the analyte (Heck and Krijgsveld 2004). Therefore, in order to quantify a peptide using MS, it is optimal to use the same peptide that is labeled with stable isotope which leads to a mass shift in the mass spectrum. Differentially labeled samples (usually one is labeled with a heavy isotope and one with light isotope) are combined and analyzed together. The differences in the peak intensities of

the isotope pairs accurately reflect differences in the abundance of the corresponding proteins (Sechi and Oda 2003).

Isotope-Coded Affinity Tagging (ICAT) uses mass spectrometry for protein separation and different isotope tags for distinguishing populations of proteins. ICAT reagents label the thiol group of cysteine residues contained within the protein during the alkylation step of sample preparation (Gygi *et al.* 1999). These labelled proteins then go on to be enzymatically or chemically digested, or separated using a gel based system prior to digestion. The prepared sample is then subjected to tandem mass spectrometry (MS/MS) allowing sequence identification and accurate quantification of the proteins contained within complex mixtures (Shiio and Aebersold 2006). In addition to ICAT , other stable isotope coding techniques applied to quantitative proteomics have been reported including a more recent approach that is analogous to ICAT termed Isobaric Tags for Relative and Absolute Quantitation (iTRAQ). iTRAQ reagents consist of a set of amine reactive isobaric tags that derivatize peptides at the N-terminus and the lysine side chains and allow for the simultaneous identification and quantitiation of up to four different samples. Even though iTRAQ is a relatively new technique it is gaining in popularity compared to ICAT because of the advantages it offers over ICAT. A comparative study of these two techniques along with 2D gel based systems is given in a review by Wu *et al* (Wu *et al.* 2006).

1.8 Bioinformatic approaches in clinical proteomics

The emergence of high throughput 'omics' technologies has led to a surge of extremely complex data sets, which posed a new challenge for detection of clinically relevant biomarkers. Interrogation of the high throughput genomic and proteomic data sets became impossible by manual means and therefore the use of computational algorithms suitable for large-scale data handling and analysis became a necessity (Krutchinsky *et al.* 2001). Development and applications of bioinformatics facilitated the data management and mining of high throughput technologies. Bioinformatics aims to create databases, algorithms, computational and statistical techniques, and theory to solve practical problems arising from the management and analysis of biological data. Although computational data analysis algorithms are applicable for both genomic and proteomic data analysis, here the focus would be on the methodologies applied in MS data analysis.

The clinical proteomics studies that aim to identify cancer biomarkers use large cohort of samples for each of the control and case groups. This is particularly important to ensure the generality and specificity of the identified biomarkers for the specific case study. Most proteomic studies employ some type of sample fractionation prior to MS analysis and then the resulting data is subjected to pattern recognition algorithm which generates the potential set of biomarkers. The strength of the results from these types of studies depends on a number of critical elements which includes careful study design, strict protocols for sample processing and data acquisition procedures and conservative approach to data mining. Automation and the use of robotics in high throughput proteomic instruments such as MS, enables processing of large numbers of samples in a short period of time and generation of large complex data sets. Several methods are available and have been applied to MS data to detect the discriminatory futures between cases and controls. However, selection of an appropriate method to analyse the MS data is crucial as different algorithms are not superior to one or another and they come with their own advantages and biases. Here, some of the strategies that are commonly implemented for MS based data analysis for biomarker identification are discussed. The general workflow of MS based proteomic for bioinformatics identification is shown in figure 10 and is illustrated in the diagram, the process of data analysis fall into two main stages, data pre-processing which reduces the dimensionality of the data set and allows the detection and locating spectral peaks, and data mining which is the statistical analysis of the data and ultimate biomarker detection.

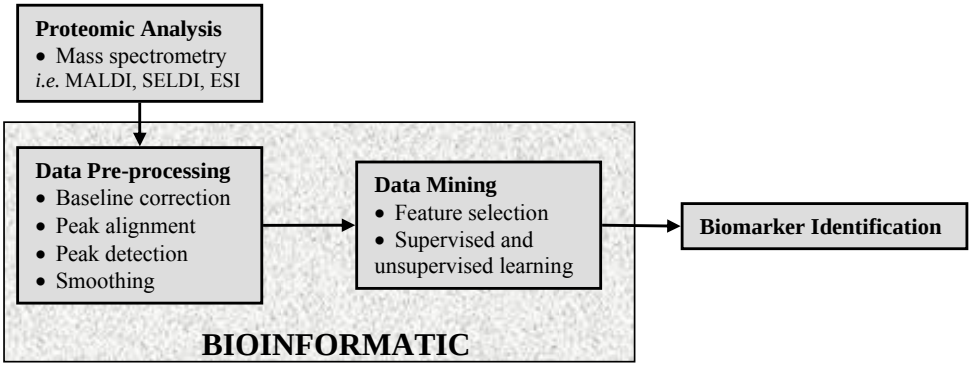


Figure 1-10. General framework of mass spectrometry data analysis.

The use and suitability of any of the pre-processing and data mining algorithms depend experimental design, the biological and the type of methodology used in sample preparation for MS analysis.

1.8.1 Data pre-processing

The aim of data pre-processing is to generate a more manageable data set for downstream data mining (Hilario *et al.* 2006). A number of strategies are available for data pre-processing that include spectrum alignment, smoothing, baseline correction, normalisation and peak detection (Yu *et al*. 2006). Selecting appropriate methods for data pre-processing is key to obtaining optimal results from the data mining algorithm. A number of investigators prefer to employ minimum or no pre-processing (using the raw *m/z* and relative intensities from MS analysis), whereas others extract a short peak list using peak detection algorithms which can then be processed for data mining. Excessive MS data pre-processing may create bias however, due to high number of samples and utilisation of highly sensitive instrumentations with high resolution, generally only mild data manipulation is required to create manageable data sets. Some of the strategies for MS data pre-processing are illustrated above (figure 1-10).

1.8.1.1 Smoothing

The raw spectrum from a typical MALDI analysis contains thousands of data points (*i.e. m/z* values) and the large scale of spectra in high throughput experiments and data handling is very challenging. Smoothing is one of the methodologies used to reduce the dimensionality of the data as well as removing noise. The strategy which is commonly used is by binning the data set to a specific mass range with a defined median intensity (Matharoo-Ball *et al.* 2007). This produces a more manageable data set to be proceeding to the data mining step.

1.8.1.2 Peak alignment

Inaccurate calibration of the MS instrument results in intersample shifts of the location of the peaks (*m/z* values) and inaccurate mass measurement. Moreover, the peak shift is often observed even when internal calibration is used due to experimental and instrumental complexities (Yu *et al.* 2005). This is because the relation between the mass error and the *m/z* is not linear. The accuracy of a peak in MALDI-MS analysis is reported to be between

0.15% and 0.3% (Lin *et al.* 2005). Generally, the MS data mining for biomarker discovery compares large groups of samples and therefore accurate mass of each spectral peak is crucial. It is absolutely essential to correct any misalignment as the data mining algorithms classify samples based on the location of the *m/z* values. Several algorithms have been developed to overcome this challenging problem (Torgrip *et al*. 2003). Wong and coworkers developed an algorithm that generate a reference spectrum and aligns all the spectra to this reference spectrum (Wong *et al.* 2005). Randolph and Yasui applied coarse scale-specific peaks (extracted by multiscale wavelet decomposition) to align MALDI data (Randolph and Yasui 2006). A number of groups use a set of known peaks which are common among the samples and use these set of peaks to align the spectra across the samples (Lin *et al.* 2005).

1.8.1.3 Baseline correction and baseline subtraction

Presence of noise in MALDI or SELDI spectra is inevitable and generally varies between the spectra. The source of noise in MS spectra is often chemical and electronic. Noisy spectra do not represent relatively accurate intensity values and show shifted baseline (base line is an offset of the intensities of masses). Therefore, it is essential to remove the baseline to flatten the baseline of spectrum. For instance, Wu and colleagues used a local linear regression algorithm to calculate the intensity of the background and eliminate it from the spectra (Wu *et al.* 2003). Determination of the noise threshold in spectra and subtraction of it from the data is another approach used by some groups (Fung *et al*. 2005). Although this method is generally used for MS data pre-processing, it is highly specific to the instrument in use.

1.8.1.4 Data normalisation

The normalisation step is applied to reduce the variation among the samples that are associated with operator and instruments errors (Marcuson *et al*. 1982). Elimination of these experimental variations from the data prior to data mining reduces the chance of detecting differences that are due to artifacts and not relevant biological differences. Data normalisation is applicable for samples analysed in one experiment and between runs and therefore makes independent data sets comparable.

1.8.1.5 Peak detection

The peak detection algorithms help to reduce the size of the spectral data by converting the raw mass spectra list to a more manageable list of monoisotopic mass and intensities. Coombes and coworkers proposed a method that generated a preliminary peak list and then from this list the candidate peaks of importance were excluded (Coombes *et al.* 2003). The disadvantage of this technique is that some of the lower abundant peaks with low intensities can be considered as noise and removed from further analysis.

1.8.2 Data mining

Once MS data have been pre-processed, they are submitted to a data mining algorithm. It is crucial to note that it is not necessary to use all the pre-processing procedures and undertaking these procedures needs to be selective. Various computational algorithms have been developed to interrogate the data in faster and efficient manner for detection of potential biomarkers in MS-based proteomic studies. Although different biomaterials are analysed in these studies, in principal similar data mining algorithms can be applied for data analysis. The data mining systems fall into two main categories: supervised and unsupervised learning approaches. The supervised learning approaches requires the class labels (*i.e*. cancer and control), meaning that they classify the data while the outcome is known whereas unsupervised learning systems cluster data and do not require class label (Fung *et al*. 2005). Classification and regression trees, neural networks, genetic algorithms and support vector machines (SVMs) are a number of supervised learning systems. Examples of unsupervised learning systems include *k*-means clustering, principal components analysis (PCA) and hierarchal clustering. A selection of both supervised and unsupervised methods will be briefly discussed in the following section.

1.8.2.1 Cluster analysis

Cluster analysis is an unsupervised learning technique, used for visualisation of data distribution. Cluster analysis is a general term for a group of classification algorithms that assign individual samples into groups (or clusters) that the members of that group have similarities to each other and to the new sample assigned (Fung *et al.* 2005). Therefore, samples that are more similar are closer together. The similarities among the samples are based on the distance between the samples that can be calculated in various ways based on

the algorithm used to classify the samples. The analysis starts with the two samples of the population that have the most similarity (*i.e.* shortest distance between them) and the sample arrangement is continued iteratively until all the samples are completely arranged. The results of the cluster analysis are commonly depicted as a dendrogram.

1.8.2.2 Principal component analysis

Principal components analysis (PCA) is a commonly used unsupervised learning technique for classification of data that can significantly reduce the dimensionality of data (*i.e.* by reducing the number of input variables) to make it more manageable. PCA analysis defines structure in the relationships between the variables, thereby having the ability to classify them (White *et al.* 2004). The reduction of data dimension by PCA is achieved by creating principal components (linear combinations), which accounts for the majority of the variability in the data, allowing better separation of data using few variables. In the first instance, the first principal component is created that explains the data better than any other principal component and subsequently other principal components are added to enhance the data explanation. Once the directions of the principal components are plotted, the values of the individual samples can be expressed as linear summations of the original data multiplied by the coefficient that best describes the principal components, these new values are known as Eigenvalues and each sample will have a score for each principal component (Fung *et al.* 2005). The PCA has been used in a number of studies for example it was utilised by Lancashire and co-workers to classify different species of *Neisseria* based on their SELDI proteomic profiles (Lancashire *et al.* 2005).

1.8.2.3 Artificial neural networks

Artificial neural networks (ANNs) are a form of artificial intelligence (AI) and are an example of computer-based algorithms used in this study for analysis and modelling complex proteomic data to identify discriminatory patterns. ANNs have been used in a variety of applications including modeling, classification, pattern recognition, and multivariate analysis (Krogh 2008; Matharoo-Ball *et al*. 2007). There are advantages to ANNs that makes them attractive to be used in analysis of complex proteomic data where there is considerable degree of variations (Fung *et al.* 2005). Complex systems, contain complicated interactions which are non-linear and difficult to interpret; however ANNs are

non-linear statistical data modeling tools and capable of dealing with this issue (Mian *et al.* 2003). Furthermore, ANNs can cope with data containing high levels of background noise and the results of ANNs can be generalised and provide a "real-world" solution (Lancashire *et al.* 2005, Matharoo-Ball *et al.* 2007). These characteristics make ANNs suitable to be used for analysis of mass spectrometry data. The structure of ANNs is based on the biological neurons and are designed to mimic the behaviour of neurons in the human brain therefore are capable of being trained and subsequently recognise complex systems (Agatonvic-Kustrin and Beresford 2000). ANNs can be created computationally by applying algorithms and adjusting parameters that stimulates the way that neurons in human brain process information. They gain the pattern recognition ability through a process called learning or training which resembles the process of learning in human. ANNs learn by experiencing with appropriate examples and minimising the errors in classification and recognition of patterns (Chen *et al.* 2004).

There are several types of ANNs however; one commonly used algorithm is the back-propagation (BP) algorithm applied to the multi-layer perceptron (MLP) architecture. ANNs are made up of units known as nodes that can be arranged in several layers (hence named MLP) that are: an input layer, a hidden layer and an output layer and each of these layers and nodes are connected with a series of weighted links, see figure 1.11 (Agatonvic-Kustrin and Beresford 2000). The input layer receives and represents the independent variables which could be the *m/z* value and intensity from a set of mass spectral profile data. The hidden layer is the processing element of the ANNs and it is not connected to the external environment and it just mathematically processes the information that it received from the input layer. The output layer represents the results.

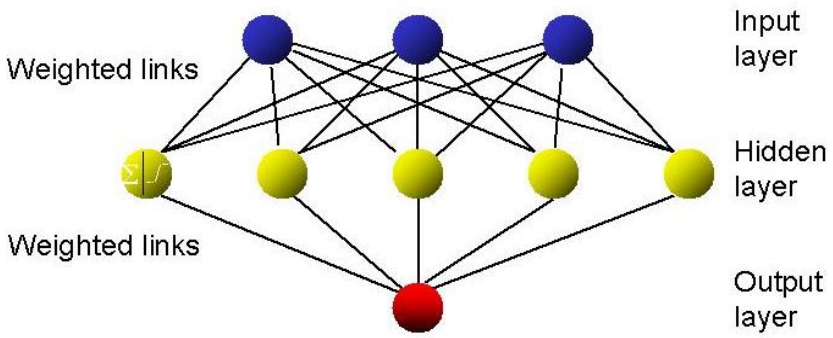


Figure 1-11. Structure of a multi-layer perceptron artificial neural network. Three layers and the weighted links are shown in which each node in each layer is connected to the nodes in the next layer via the weighted links. Adopted from Ball *et al.* 2002.

Training or learning process in ANNs is similar to the human learning process which involves adjustment of synaptic connections between the neuron cells. The process of learning in ANNs can be supervised (where input with known output are used for training) or unsupervised (where inputs with unknown output is used for training) however in this study BP algorithms (form of supervised learning method) have been used. The ANNs are trained by iterative adjustment of the weights between the layers to minimise the error in correct classification of data. The principle of supervised network with BP learning rule is shown in figure 1-12.

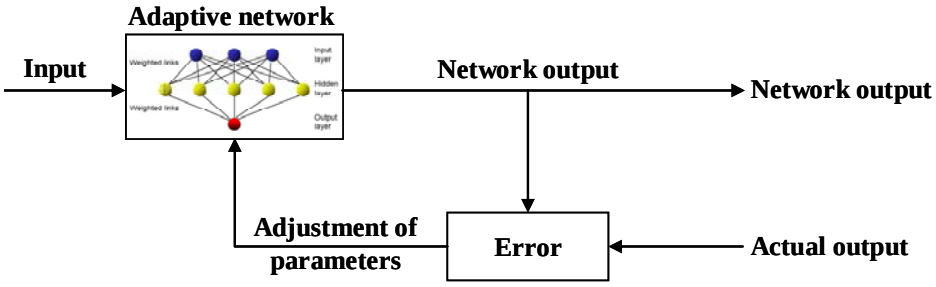


Figure 1-12. Supervised artificial neural network with backpropagation algorithm learning rule. Input is fed forward through the network as examples to train the model, in hidden layer the class of the input is calculated and put through the output layer. The network predicted class of output is compared to the actual output class and an error is calculated. The error is fed back through the network and the backpropagation of the error to the network improves the weights and eventually the output of the network.

In the first step of ANN training process with BP learning rule, information from the inputs (*i.e.* the *m/z* value and intensity from a set of mass spectral profile data of cancer or control

samples) is fed forward through the network (forward step). In this step, inputs are used as examples to train the network and output (*i.e.* classification of the sample, whether is a cancer or control) is calculated in the hidden layer and presented to the output layer (networks output). This is then compared to the actual output (the actual class of that particular sample) and an error is calculated. In the second step, the difference between the network and actual output is calculated and the weights are updated in proportion to the error calculated. The weights between the different layers are then modified and the training process is repeated until the minimum error is achieved or the outcome prediction of the network failure to improve after a certain amount of training cycles named epochs (Lancashire *et al.* 2005, Agatonvic-Kustrin and Beresford 2000).

In order to analyse a dataset using ANNs, the data input for the ANN analysis is often split into training, test and validation sets. The training set is used to train the model and the output prediction is validated using the test set. Once the training is completed, the accuracy and efficiency of the model is further tested using input cases that are completely unseen by the model (validation set) and an output is calculated based upon the new data (blind dataset). The accuracy of neural network to classify the blind dataset indicates whether the model can be generalised on unseen independent data. Obtaining a generalised model is often problematic. One solution to this problem which has been applied in this study is by using multiple models with different unseen datasets (random sample cross validation) a more generalised model may be obtained. Random sample cross validation randomly selects different samples sets (inputs) to be used in the process of training and testing of number of models. ANNs as mentioned before can be used for pattern recognition and classification of mass spectrometry data in which thousands of inputs (*i.e.* the *m/z* and their corresponding intensities) are presented to the neural network system. However, for the purpose of biomarker discovery and detection of patterns which are discriminatory, it is important to determine the importance and influence of each of the individual inputs on the created model. This is possible via a parameterisation process. Parameterisation is a process that reduces the complexity of the dataset and increases the performance and accuracy of the models. This process allows the identification of the most important inputs and elimination of inputs that are of little or no importance. Parameterisation can be achieved by performing a weightings analysis, sensitivity analysis or by stepwise approaches (Lancashire *et al.* 2005). ANNs work by weighting the links from the inputs to the outputs; the stronger the weight leading from a particular output, the

greater the influence it has on the model. A stepwise approach was used in this study which allows the determination of the minimum subset of ions that are required to predict a particular outcome and correctly assign individuals to their relevant groups. In addition, utilising this approach, the interaction between the ions may also be identified. Briefly, this analysis is based on training a number of models where in the first instant, each input is used as a single input and an error is determined for each model. The input that gives the lowest error is selected and put with all of the remaining inputs sequentially in a number of sub models once more. This process continues until addition of further inputs does not improve the model's performance.

Despite the advantages of ANN in analysis of various datasets including mass spectrometry data, there are two major limitations to ANNs; firstly, they do not explain how they reach the conclusions and therefore are often described as a "black box" and secondly ANNs are prone to overparameterisation where the network is left to train for too long and the problem of overtraining or overfitting has been emerged therefore the results cannot be generalised (Krogh 2008). This issue is often addressed by use of random sample cross validation. Nevertheless, neural networks have been successfully applied to many interesting areas of biomedical science for analysis of complex data obtained from proteomic analysis (Li *et al.* 2004). The ANNs have been utilised to model recurrence post radical prostatectomy (Mattfeldt *et al.* 1999) and therapy response prediction (Michael *et al.* 2005). In addition, ANNs have been successfully applied on SELDI and MALDI data for detection of brain, melanoma and breast cancer biomarkers (Matharoo-Ball *et al.* 2007, Mian *et al.* 2005, Ball *et al.* 2002, Liu *et al.* 2004, Lancashire *et al.* 2003, Lancashire *et. al.* 2005) and bacterial species classifications (Lancashire *et al.* 2005, Schmid *et al.* 2005, Iversen *et al.* 2006).

1.9 Personalised medicine and proteomics

The aim of personalised medicine is to use biomarkers to select the most effective treatment for each individual patient, which improves the clinical management of the disease. Diagnosis and staging of the solid tumours in the current clinical practise is through physical examination, imaging and pathological evaluation. The appropriate treatment is prescribed based on these morphological assessments, although patients that were classified

at the same stage often respond differently to the same treatment. This is probably due to the heterogenousity of cancer and unique "molecular qualities" of the individuals. The current strategies for diagnosis of cancer are often inadequate and provide limited information on the molecular profiles of individuals that may affect the therapy results. Molecular based diagnosis techniques such as immunohistochemistry, *in situ* hybridisation and chromosomal analysis have been developed for better assessment of the individual patient tumours however they are time consuming, low throughput and have limited specificity and sensitivity (Geho *et al.* 2006). The field of personalised medicine has been significantly progressed by emerging novel high throughput tools in genomics and proteomics, offering the ability of molecular profiling in a broader range. Such technologies are typically classed as multiple biomarker discovery approaches that divert the idea of single biomarker of disease to a new perspective. These novel technologies provide genotype and phenotype information on the nature of the tumour, genetic background of the patient and how each individual's body interact with the presence of the disease. The genotype and phenotype information are often in the format of gene and protein expression profiles. This will allow clinicians to decide on a treatment based on both the tumour type and susceptibility of individuals to the selected treatment. As genomic and proteomic profiling do not present a complete molecular information individually, it is essential to use them hand-in-hand (Petricoin *et al.* 2004). There are number of trials based on gene microarray to identify gene signatures in breast and prostate cancer to stratify individuals for therapy. Although gene profiling is promising technique for detection of individual molecular profiles, the DNA code often does not reflect the protein product due to post-translational modifications such as phosphorylation and glycosylation. There are a number of ways proteomics can be of use in the personalized management of cancer. Investigating the global changes in the protein expression patterns reveals all the post translational modifications and as mentioned previously these make up the functional components that could be used as markers for diagnosis/prognosis or as novel targets for drug delivery.

1.10 Aims and objectives

The large scale, high-throughput study of proteins in various patient samples by MALDI mass spectrometry to generate proteomic profiles, has experienced an explosive growth in the last decade. In addition, development and application of computational algorithms has

aided the identification of patterns within the proteomic profiles that are potentially useful in diagnosis and prognosis of cancer patients and prescription of suitable treatment regimens with high efficiency. Despite the wider use of MALDI mass spectrometry in conjunction with bioinformatic analysis in the field of cancer biomarker discovery and generation of valuable results, the technology has not been sufficiently validated for clinical use.

The aim of this research was to address the issues related to sample preparation strategies and reproducibility of the MALDI mass spectrometry and the robotic sample handler which was used for sample processing. Over the past few years, reproducibility of MALDI mass spectrometry and sample preparation strategies have been the subject of some criticisms. Therefore, the aims of the first part of this study was to optimise and standardise the protocols for future sample preparation and processing which involves reverse-phase chromatography tips followed by MALDI-MS analysis. Generation of standard operating procedures are important to obtain consistency of results over a period of time. Furthermore, the reproducibility of instruments in use will be assessed within and between runs using QC samples.

Following optimisation of protocols, these strategies would be applied to identify serum proteomic patterns that may correlate to tumour progression in CT26 tumour bearer mice. By identifying biomarkers that may correspond to tumour progression, these markers can be indicative of tumour initiation and may applicable for early diagnosis of cancer.

The research will further investigate whether it is possible to use the principles of MALDI mass spectrometry and artificial neural networks as tools to identify patterns that correlate to immunotherapy (*i.e.* DISC-GM-CSF immunotherapy) outcome in a CT26 mouse model. As research into personalised medicine seeks to identify patterns that can indicate certain therapy outcomes, human patient material is sometimes limited and difficult to obtain in sufficient quantity and numbers to make the findings statistically significant. Therefore, initially this study aims to identify whether serum samples can be used for proteomic studies to identify biomarkers correlating with therapy response. In addition, validation of any candidate biomarkers that were discovered by proteomic analysis would be carried out using non-MS based methodologies. Success in validation of discoveries from the proteomic studies would be valuable as the majority of critisms in the field of biomarker discovery using mass spectrometry is directed towards failure of validation.

The final part of this research will use serum samples from CT26 mouse model where a different immunotherapy (*i.e.* DC-based vaccination) was used and serum samples were collected pre and post therapy. Serum proteomic patterns will be obtained using MALDI mass spectrometry and will then be analysed using ANNs to determine candidate biomarkers that differentiate different time-points and therapy responder and non-responders. The results of this study can be compared with the previous immunotherapy model as well as independently for discovery of biomarkers that may associate with therapy responses.

Chapter 2 - Materials & Methods

2.1 Materials

2.1.1 Reagents

All the reagents were stored in accordance to manufactures instructions and were used before the expiry date.

Sample preparation reagents	**Company**
Acrylamide	Genflow
Ammonium persulphate	National Diagnosis
Benzamidine	Sigma-Aldrich
Bovine serum albumin (BSA)	Sigma-Aldrich
Coomasie Blue	PhiBio Fisons
Deoxycholic acid	Sigma-Aldrich
Dithiothreitol	Apollo Scientific
Ethanol	Sigma-Aldrich
Glycine	Fischer Scientific
IGEPAL CA-630	Sigma-Aldrich
Marvel	Premier International Foods
Methanol	Sigma-Aldrich
Octyl glucopyranoside	Apollo Scientific
Phenylmethanesulphonylfluoride (PMSF)	Sigma-Aldrich
Protein assay BCA solution	Sigma-Aldrich
Protein assay copper(II) sulfate pentahydrate 4% solution	Sigma-Aldrich
Sodium Azide	Sigma-Aldrich
Sodium Dodecyl Sulphate	Sigma-Aldrich
Sodium ortho-vanadate	Sigma-Aldrich
Sodium Tetraborate	Sigma-Aldrich
Trizma Base	Sigma-Aldrich
Tween 20	Sigma-Aldrich

Urea	Sigma-Aldrich
Water	Sigma-Aldrich

Tissue culture reagents	**Company**
1640 RPMI	Bio Whittaker Europe
DMEM	Bio Whittaker Europe
DPBS	Bio Whittaker Europe
Ethanol	BDH
Foetal Calf Serum	Bio Whittaker Europe
Trypan blue	Sigma-Aldrich
Trypsin	Gibco
Versene	Gibco

Proteomics reagents	**Company**
Aceton HPLC grade	Fisher Scientific
Acetonitrile HPLC grade	Fisher Scientific
CHCA (α-Cyano-4-hydroxycinnamic acid)	Laser Bio Labs
Dichloromethane HPLC grade	Fisher Scientific
Formic acid	Fisher Scientific
Hexane HPLC grade	Fisher Scientific
Methanol HPLC grade	Fisher Scientific
Peptide calibrant Mix 4	Laser Bio Labs
Propanaol HPLC grade	Fisher Scientific
Protein calibrant Mix 2	Laser Bio Labs
Sinapinic Acid (3,5 Dimethoxy 4-hydroxycinnamic acid)	Laser Bio Labs
Trifluoroacetic acid	Fisher Scientific
Trypsin Gold	Promega
Water HPLC grade	Fisher Scientific

2.1.2 Buffers

Buffers were prepared as indicated below:

Name	Composition
0.5M Tris HCl Buffer (pH 6.8) (stacking gel buffer) (1-D PAGE) Stored at room temperature	6g Trizma base 0.4g SDS make up to 100ml with option 4 H_2O adjust pH to 6.8 with HCl
1.5M Tris HCl Buffer (pH 8.8) (resolving gel buffer) (1-D PAGE) Stored at room temperature	18.16g Trizma base 0.4g SDS make up to 100ml with option 4 H_2O adjust pH to 8.8 with HCl
10x Running Buffer (1-D PAGE) Stored at room temperature	0.25M Trizma base 2M Glycine 1% (w/v) SDS
1x TBS Buffer (Western blotting) Stored at room temperature	2.422g Tris 29.22g NaCl make up to 100ml with option 4 H_2O adjust pH to 7.5
1x TBST Buffer (Western blotting) Stored at room temperature	2.422g Tris 29.22g NaCl 500µl Tween 20 make up to 100ml with option 4 H_2O adjust pH to 7.5
Bjerrum & Schafer-Nielsen Transfer Buffer (Western blotting) Stored at 4°C	5.82gr Tris (48mM) 2.93g Glycine (39mM) 200ml Methanol make up to 100ml with option 4 H_2O
Blocking Buffer (Western blotting) Prepared fresh	5g Marvel 100ml TBS buffer 50µl Tween 20
Modified RIPA buffer (protein extraction) Prepared fresh	5ml RIPA buffer 50µl IGEPAL CA-630 25mg Deoxycholic acid 50µl Urea (10%) 10µl Benzamidine (500mM) 5µl PMSF (100mM) 25µl Sodium ortho-vanadate (200mM) 5µl Sodium azide (1mM)

RIPA Buffer (protein extraction) Stored at 4°C	4.38g NaCl (150mM) 3.03g Tris (50mM) 0.93g EDTA (5mM) make up to 500ml with option 4 H_2O
Sample Reducing Buffer Stored at room temperature	2.5ml 0.5M Tris HCl buffer (pH 6.8) 400mg SDS 2ml Glycerol 200mg DTT A few grains of bromophenol blue make up to 20ml with option 4 H_2O

2.1.3 General laboratory consumables

Plastic and non-plastic Ware	**Company**
0.5ml microtube	Sarstedt
1.5ml microtube	Sarstedt
10ml pipette	Sarstedt
10ml syringe	BD Biosciences
12 – well tissue culture (TC) plate	Sarstedt
1ml cryovials	TPP
25ml pipette	Sarstedt
300µl glass vials	Chromacol
30ml universal tube	Sarstedt
384 – plate	Sarstedt
384 MALDI target plate	Shimadzu
	Bruker Daltonic
50ml centrifuge tube	Sarstedt
5ml pipette	Sarstedt
7ml Bijou tube	Sarstedt
96 – well plate	Sarstedt
C_{18} Solid Phase Extraction Colums	Phenomenex
C_{18} ZipTip	Millipore
C_4 ZipTip	Millipore
ELISA plate	Sarstedt
FACS tube	Elkay
Glass slides	Menzel GmBH

Pasteur Pipette	SLS
PVDF membrane	GH Healthcare
T175 TC flask	Sarstedt
T25 TC flask	Sarstedt
T75 TC flask	Sarstedt
High Performance Chemiluminescence film	Amersham

2.1.4 Equipment

Hardware	**Model, *Company***
-80°C freezer	Ultima II, *Revco*
	U570 Premium, *New Brunswick Scientific*
Absorbance plate reader	Model 680, *Biorad*
Centrifuge	Mistral 1000, *MSE*
	Mistral 2000R, *MSE*
	Falcon 6/300, *MSE*
Cryostat	CM 1900, *Leica*
Cryostore	Cryo 200, *Forma Scientific*
Liquid chromatography (LC)	Dionex UltiMate 3000 nanoflow, *Dionex*
Xograph Imaging System	Compact X4,
WB	
WB	
Electrophoresis gel tank	*GeneFlow*
Incubators	CO_2 water jacketed incubator, *Forma Scientific*
Mass Spectrometers	Ultraflex III TOF/TOF, *Bruker Daltonic Inc*
	Axima CFR+, *Kratos*
	LTQ, *ThermoFisher Scientific*
Microcentrifuge	Microcentaur, *MSE*
	Mikro 22R, *Hettich Zentrifugen*
Microscope and Camera	Model PIM, *World Precision Instruments*
Power supply for electrophoresis	Consort E122, *GeneFlow*
Robotic sample processer	XCISE, *Proteome Systems/Shimadzu*

Safety cabinet	Microflow biological safety cabinet, *Walker*
Sonicators	Precision Ultrasonic Cleaning, *Ultawave*
	Ultrasonic Cleaner, *VWR*
Vortex	Whirlimixer, *Fisher Brand*
Water bath	Y14, Grant

2.1.5 Softwares

Software Product	**Company**
BioTools	Bruker Daltonics
ClinProTools	Bruker Daltonics
FlexAnalysis	Bruker Daltonics
FlexControl	Bruker Daltonics
Kompact Launchpad v2.4.1	Kratos Analytical Ltd
Mascot	Matrix Science Ltd
SpecAlign v2.4	Cartwright Group
SEQUEST	ThermoFisher Scientific
Statistica v6.1	StatSoft Inc

2.1.6 Antibodies

Specificity	**Conjugated**	**Host**	**Company**
Chicken IgY	HRP	Goat	Immune Systems Ltd
Mouse HBB	None	Goat	Santa Cruz Biotechnology Inc
Mouse HPX	None	Chicken	Immune Systems Ltd
Mouse SAA1	None	Goat	R&D System
Mouse SAP	None	Rabbit	Abcam
Precision Strep Tactin	HRP		BioRad
Rabbit Immunoglobulins	HRP	Swine	Dako
Rabbit Immunoglobulins	Biotinylated	Goat	Dako

2.1.7 Kits

Product	Company
Amersham ECL Western blotting detection reagent	GE Healthcare
Bicinchoninic acid (BCA) protein Assay	Sigma-Aldrich
Mouse serum amyloid A ELISA	Immunology Consultant Laboratory Inc
Mouse serum amyloid P ELISA	Immunology Consultant Laboratory Inc
Mouse serum hemopexin ELISA	Immunology Consultant Laboratory Inc
VECTASTAIN® *Elite*® ABC	Vector Laboratories Inc

2.1.8 Company addresses

Company	Address
Abcam Inc	Cambridge, USA
Acros	Loughborough, UK
BD	Cowley, UK
BDH	Leicester, UK
Beckman Coulter	High Wycombe, Bucks, UK
Bio Whittaker Europe	Wokingham, UK
Biorad	Hemel Hempstead, UK
Bruker Daltonics Inc	Bremen, Germany
Cambrex	Nottingham, UK
Chromacol	Welwyn Garden City, Herts, UK
Ciphergen	Guildford, Surrey, UK
Dako UK Ltd	Ely, UK
Dionex Corporation	Surrey, UK
Elkay	Basingstoke, UK
Fischer Scientific	Loughborough, UK
Forma Scientific (Thermo)	Basingstoke, UK
GeneFlow	Fradley, Staffs, UK
Gibco	Paisley, UK
Greiner Bio-One	Gloucestershire, UK
Hettich Zentrifugen	Tuttlingen, Germany

Immunology Consultant Laboratory	Newberg, USA
Kratos Analytical Ltd	Manchester, UK
Laser Bio Labs	Cedex, France
Leica	Milton Keynes, UK
Menzel GmBH	Braunschweig, Germany
MSE	London, UK
New Brunswick Scientific	Edison, New Jersey, USA
Proteome Systems	North Ryde, New South Wales, Australia
Revco	Asheville, North Carolina, USA
Santa Cruz Biotechnology Inc	Santa Cruz, USA
Sarstedt	Leicester, Leicestershire, UK
Scientific Laboratory Supplies (SLS)	Nottingham, Notts, UK
Shimadzu	Milton Keynes, UK
Sigma-Aldrich	Gillingham, Surrey, UK
StatSoft Inc	Bedford, UK
Stratec	Birkenfeld, Germany
ThermoFisher Scientific	Loughborough, UK
TPP	Switzerland
Ultrawave	Cardiff, UK
Vector Laboratories Inc	Burlingame, USA
VWR	Poole, Dorset, UK
Walker	Massachusetts, USA
Ward Systems Group, Inc	Frederick, Maryland, USA
World Precision Instruments	Stevenage, Herts, UK

2.2 Methods

2.2.1 Animals, cell lines and DISC virus

2.2.1.1 Animals

Male and female BALB/c mice ($H2L^d$) were purchased from Harlan Olac (Oxon, UK) and bred at the Nottingham Trent University animal house in accordance with the Home Office Codes of Practice for the housing and care of animals.

2.2.1.1 Cell lines

The CT26 cell line used in this study were purchased from two sources: 1) The CT26 cell line purchased from American Type Culture Collection (ATCC) and was maintained by serial *in vitro* passage in RPMI 1640 tissue culture medium supplemented with 5ml L-glutamine and 10% FCS. 2) The CT26 cell line provided by Prof. I. Hart, Imperial Cancer Research Fund and was maintained by serial *in vitro* passage in DMEM tissue culture medium supplemented with 10% FCS. The cell were remained at 37ºC supplemented with 5% CO_2 humidity chamber.

Once the cultured cells were 70-80% confluent, they were maintained by harvesting and splitting them in tissue culture flasks. The CT26 are adherent cell lines and their harvesting from the tissue culture flask was by using a combination of trypsin and versine (T+V). Firstly, the culture media was disposed to waste and the cells were washed with sterile PBS three times. According to the volume of the tissue culture flask, 0.5-2ml of T+V was added to the cells and incubated at 37°C for 5-10 minutes. During the incubation period, cells were detached by tapping the flask few times which also avoids cell clamping. Following the process of detaching cells by T+V, 5-10ml of media was added to the flask and the cell suspension was re-cultured in a new flask with appropriate volume of fresh medium. If the CT26 cells were prepared for *in vivo* injections, they were counted and then the appropriate numbers of cells were resuspended in FCS-free medium.

2.2.1.2 DISC virus

The DISC-mGM-CSF-HSV was constructed as described by Ali *et al.* (Ali *et al.* 2000a). The DISC virus was propagated using CR1 cells (African green monkey kidney cells). The CR1 cells were cultured with DISC virus (0.1pfu/cell) and incubated at 37ºC. The cells

were scraped from the culture flask and followed by sonication in a water bath (4-5ºC) for 1 minute.

The virus titration was carried out by initial growth of CR1 cells in a 96 well plate ($4.5x10^5$/100µl/well) overnight. This was followed by adding the virus to the plate at different concentrations ranging from 10^{-1} to 10^{-10}, and the plate was incubated at 37ºC for 3 days. The number of infected cells were determined under the microscope using $TCID_{50}$ computer software the titer of the virus was determined.

2.2.2 Animal models

Two different CT26 colorectal carcinoma murine models were used out in the present project. The first one was a CT26 progression study and the second one was the CT26 tumour therapy model. Balb/c mice (female and male) were purchased from Harlan Olac, Oxon, UK. The animals were maintained and bred in accordance with Home Office Codes of Practice for the housing and care of animals.

2.2.2.1 CT26 progression model

Prior to start of the experiment, male Balb/c mice, aged 6-8 weeks (n = 60), were randomly assigned to one of the two groups: tumour bearer (TB) (n = 45) and control (n = 15). The CT26 tumours were harvested, counted, resuspended in FCS-free RPMI 1640 tissue culture medium and then used to inject the TB group of mice. Prior to injection of the cells, the site of the injection was shaved and subcutaneous (s.c.) CT26 tumour were induced by injection of 8 x 10^4 cells into the right hand side flank of the mice. The tumours were allowed to develop and the tumour size was measured at 2-3 day interval. The tumour was allowed to grow to reach a maximum 1 cm^2 in size.

2.2.2.2 CT26 therapy model

CT26 retrospective DISC immunotherapy

A series of CT26 immunotherapy mouse model experiments with DISC-HSV were carried out previously in our laboratory by Dr Murrium Ahmad during the period 2002-2004 where the procedures were exactly the same to those described below. The CT26 cell line used in these experiments were from those provided by Prof. I. Hart, Imperial Cancer Research Fund. Blood samples were collected from each of these experiments at the end of the

therapy outcome which resulted in *n* = 50 regressors (mice that responded to therapy) and *n* = 50 progressor animals that were therapy resistant. The samples from these models in the present thesis were referred to as retrospective models.

CT26 prospective DISC immunotherapy models

Two independent experiments were carried out in this project and both experiments were conducted in a similar manner and in accordance to strict procedures to reduce any bias. The first experiment was conducted in February 2007 and the second one in January 2008. The CT26 cell line used in 2007 experiment was provided by provided by Prof. I. Hart, Imperial Cancer Research Fund and in the 2008 experiment both provided by Prof. I. Hart, Imperial Cancer Research Fund provided cells and also a second one purchased from American Type Culture Collection (ATCC) CT26 cell lines were used. The experiment conducted in 2007/2008 was referred to as the prospective model in this thesis.

Prior to start of the experiment, Balb/c mice, aged 6-8 weeks, were randomly assigned to one of the three groups: DISC therapy (injected with CT26 and receiving DISC therapy), TB and control (received no CT26 tumour injection and no therapy). A week later, following the growth of the CT26 cell line *in vitro*, trypsinised cells were harvested, counted and resuspended in serum free DMEM media. A suspension of 8 x 10^4 cells in serum free DMEM media was inoculated s.c. into the right flank of the mice (to both DISC therapy and TB groups). After tumours developed to approximately 0.04-0.36 cm^2 surface area (approximately in 7-10 days), tumour therapy was initiated by intra-tumour (i.t.) injection of 2.5 x 10^7 PFU of DISC-mGM-CSF (this was applied only to DISC therapy group). The DISC-HSV-mGM-CSF used for therapy, was constructed as previously described (Ali *et al.* 2000). Treatment was repeated 2-3 days later and animals were monitored to assess the response to therapy at 3-4 day interval. Mice that responded to the therapy (regressors) were characterized by reduction of tumour size which resulted in complete rejection of tumour. However, failure to DISC therapy resulted in aggressive growth of tumours (progressors). The number of animals used in each experiment is shown in table 2-1.

	DISC therapy		TB		Control	
	Male	Female	Male	Female	Male	Female
Experiment 1 (February 2007)	20	20	5	5	5	5
Experiment 2 (January 2008)	40	-	20	-	11	-

Table 2-1. Table showing the number of animals used in each category for the two CT26 murine DISC immunotherapy model.

2.2.2.3 Immunotherapy of CT26 tumour-bearer mice with dendritic cell vaccine in combination with blockade of vascular endothelial growth factor receptor 2 and CTLA-4

This study was conducted by our collaborators in Denmark, University of Copenhagen ad details of the therapy procedures were published by this group (Pedersen *et al.* 2006). A total number of 37 samples were received from this group that included serum samples from 21 regressors, 6 progressors and 10 tumour-bearer mice.

2.2.3 Mouse serum and tissue collection

Serum samples from all the animal models (experiment 1 conducted at February 2007, retrospective, prospective DISC immunotherapy, CT26 progression model and Denmark immunotherapy model) and tissue samples from experiment 1 conducted at February 2007 and prospective DISC-HSV immunotherapy models and CT26 progression model were collected as follow:

2.2.3.1 Blood Sample Collection

The tail bleeding procedure was carried out in accordance with the Home Office regulations. Blood samples were taken from the naïve, tumour bearer, progressor and regressor mice from all experiments. In order to locally anaesthesise the tails, EMLA cream was applied and animals were kept at 37°C for approximately 15 minutes. Blood was collected by nicking the tip of the tail and approximately 200µl of blood was collected from each mouse in 1.5ml eppendorfs. Collected blood was clotted on the bench for 30 minutes and serum was harvested after spinning the blood at 3500 rpm for 10 minutes. Repetitive freeze and thaw of serum samples was prevented by aliquoting serum samples into required volumes and then immediately frozen at -80°C.

2.2.3.2 Surgical excision of tissues

The animals were culled using S1 procedure and tumour, spleen, liver, heart, intestine, lung and kidneys were removed from the animal, placed into the cryovials and snap-frozen in liquid nitrogen and immediately stored at -80°C. The spleen, liver and the tumour were weighed prior to snap freezing the samples.

2.2.4 Tissue homogenisation and protein extraction from tissue

For total protein extracting from frozen tissues, the tissue was ground under liquid nitrogen in a Class II safety cabinet using a mortar and pestle to a powder. The powder then transferred to a 1.5 ml Eppendorf tube to which 0.5-1ml of RIPA buffer was added directly to the powder for protein extraction from the tissue by vortexing the tube followed by two 30 sec sonication cycles and finally the samples were stored on ice for 30 minutes. The samples were then centrifuged at 14,000 rpm for 20 minutes at 4°C. The supernatant was separated and stored at -80°C until use. Prior to protein assay, 3 cycles of freeze/thaw were carried out.

2.2.5 Protein assays

The BCA 96-well plate protein assay was used to determine the protein concentration of all of the samples (tissue and serum) used in this study. The assay was performed in duplicates for individual samples. A BSA protein standard curve was used with concentrations of 0, 0.2, 0.4, 0.6, 0.8 and 1 mg/ml in water. For serum samples, they were diluted 1:2100 and tissue lysates 1:200 in water. 25µl of different sample and BSA standards were transferred to individual wells of a 96-well plate (flat bottom) and 200µl of BCA working reagent was added to each well and left to incubate for 30 minutes at 37°C. The absorbance was determined using a spectrophotometer at 562nm. The BSA standard curve was calculated and produced and sample protein concentrations were calculated from this using Microsoft Excel.

2.2.6 Mass spectrometry

2.2.6.1 Xcise automated sample processing system

Sample processing and preparation for mass spectrometry was preformed using the Xcise robotic system to ensure consistency of sample processing. Moreover, to reduce the risk of

bias correlated to the position of the samples on the target plate, the sample position on the 96 well plate of the Xcise and the MALDI target plate was randomly selected using a Microsoft Excel computer program. Subsequently, samples (serum or tissue lysates) were thawed on ice prior to analysis. Serum samples were diluted 1:20 in 0.1% TFA and protein concentration of tissue lysates were equalised to 1 mg/ml in 0.1% TFA. Typically, 25 µl of the diluted sample was used for the analysis and placed into a 96 well plate according to the randomized order. For protein/peptide analysis, samples were initially fractionated by C_{18} ZipTip reversed-phase chromatography according to manufacturer's instructions. The samples were bound to the C_{18} ZipTip with 15 cycles of binding, followed by two washes in 0.1% TFA, where the washes were discarded. The samples were eluted off of the ZipTip in 8 µl of 80% ACN/0.1% TFA. This was followed by spotting 1 µl of the eluted sample with 1 µl of matrix solution containing 10 mg of sinapinic acid (SA) in 1 ml of 50% ACN in 0.1% TFA on the MALDI target plate, which was then allowed to air dry at room temperature. The remaining eluted sample (7 µl), was used for peptide analysis which was mixed with 16.6 µl of ammonium bicarbonate and 7.6 µl of water. This was transferred to an eppendorf and 0.7 µl of mass spectrometry grade trypsin gold (100 µg/100 µl) was added manually. Samples were incubated at 37°C overnight and the reaction was stopped by adding 1 µl of 1% TFA. Samples were cleaned by use of C_{18} ZipTip and 1 µl of the eluted sample was spotted on the MALDI target plate with 1 µl of matrix solution of 10mg α-cyano-4-hydroxycinnamic acid (CHCA) in 1 ml of 50% ACN/0.1%TFA and allowed to air dry at room temperature. For analysis of the tryptic peptides with ESI MS/MS, sample preparation was as described, except that 105 µl of 0.1%TFA or 0.1% formic acid was added to the eluted samples by Xcise robotic system.

2.2.6.2 MALDI mass spectrometry

The majority of the MALDI mass spectrometry analysis in this study has been preformed using the Shimadzu instrument however, Bruker Daltonic MALDI instrument was utilised for the analysis of the CT26 progression model. Although the methodology used for sample preparation and analysis of these two instruments were identical, there were slight differences in calibration strategies.

2.2.6.2.1 Shimadzu MALDI instrument

The serum sand tissue samples were analysed using an Axima CFR^{+} mass spectrometer operated in linear or reflectron modes. To calibrate the instrument close external calibration was performed in which for every 4 sample spots on the MALDI target plate, 1 calibration spot was used; this was used to ensures the mass accuracy of the instrument. An appropriate calibration mixture in accordance to the MALDI operation mode (*i.e.* linear or reflectron) was made-up fresh for each experiment. The two operation modes are described bellow:

1. *Linear MALDI-TOF MS.* For the analysis of the proteins, the instrument was operated in linear mode mass range of 1000-30000 Da. The calibration mixture used in this mode was the calibration mix 2, that contained cytochrome C (horse heart) *m/z* 12361.12, myoglobin, (horse) *m/z* 16181.06, trypsinogen *m/z* 23981.98 and insulin beta chain *m/z* 3494.65 (5 µl of 5mM). The mass spectral data was collected in 'raster mode' and the raw data was exported as ASCII text files and processed for bioinformatic analysis.

2. *Reflectron MALDI-TOF MS.* Peptides were analyzed in reflectron mode mass range of 800-3500 Da and peptide calibration mix 4 was used for calibration of the instrument which was based on monoisotopic masses and contained bradykinin fragment 1-5 *m/z* 573.31, angiotensin II *m/z* 1046.54, neurotensin *m/z* 1672.91, ACTH clip (18-39) *m/z* 2465.19 and insulin B-chain oxidized *m/z* 3494.65. The mass spectral data was obtained using 'auto quality' in which the instrument performs a scan over each spot to find the hot spots and then the mass spectral data is collected from these areas. The raw data was then exported as ASCII text files and processed for bioinformatic analysis.

2.2.6.2.2 Bruker Daltonic MALDI instrument

The serum and tissue samples were analysed using an Ultraflex III TOF/TOF mass spectrometer operated in reflectron mode. To calibrate the instrument internal calibration was performed in which trypsin autolysis peaks in the tryptic digest peptide samples were used for calibration. Tryptic peptides were analyzed in reflectron mode mass range of 800-3500 Da.

2.2.6.3 ESI-MS/MS

For identification of the predictive peptides, ESI-MS/MS was carried out using a ThermoFinnigan LTQ linear ion-trap mass spectrometer attached to a Dionex UltiMate 3000 nanoflow liquid chromatography system after using identical sample preparation to that described above for MALDI-MS analysis. Samples were de-salted and pre-concentrated online using a precolumn at a flow rate of 30 µL/min (LC Packings, C_{18}-PepMap, 100 Å, 3µm particle size, 300 µm ID x 5 mm, Dionex Ltd, UK) and then switched to an analytical column (LC Packings, C_{18}-PepMap, 100 Å, 3µm particle size, 300 µm ID x 5 mm, Dionex Ltd, UK) at a flow rate of 180 nL/min connected to the nanospray interface of the mass spectrometer. Following analysis, an automated database searching of fragment ion spectra was carried out using the SEQUEST algorithm to allow peptide identification.

2.2.7 Bioinformatic analysis

2.2.7.1 Data pre-processing

Once the data was obtained by the MALDI, the profiles of each sample were exported from the mass spectrometer as ASCII text files. The data was then smoothed and bucketed to values by taking median intensities of 1 Dalton. Furthermore, the data was baseline corrected using a rolling baseline determination method that used the noise around non-peak regions. Once the data had been smoothed and base line corrected, peaks were aligned using Specalign software. In addition, the data was transported into STATISTICA7 software to be analyzed by ANNs.

2.2.7.2 Artificial neural networks (ANNs) stepwise approach

Prior to ANNs analysis, one group of samples were coded as 1, and the other group of samples were coded as 2. The ANNs were trained with 1 hidden layer and 2 hidden nodes at a learning rate of 0.1 and a momentum of 0.5. The models were trained using random sample cross validation, where the samples were randomly split into three groups; 60% for the training set, 20% for the test set and 20% for the validation set. The models were trained using random sample cross validation, as described above, using the training set, and the network error with regards to predictive performance was monitored with the test set, which was unseen during training. Once this error failed to improve for a pre-determined number of training events, training was terminated, and the model validated on the blind data set. This process was repeated 50 times; so that each sample was treated as

truly blind a number of times, enabling confidence intervals to be calculated for the network predictions on blind data. Initially, each variable from the dataset was used as an individual input in a network. These models were then trained over 50 randomly selected subsets and the network predictions and mean squared error values for these predictions were calculated for each model with regards to the separate validation set. The inputs were ranked in ascending order based on the mean squared error values for test data and the model input which performed with the lowest error was selected for inclusion into the subsequent step. Thus, approximately 1 million models were trained and tested at each step of model development. Next, each of the remaining inputs were then sequentially added to the previous best input, creating *n*-1 models each containing two inputs. Training was repeated and performance evaluated. The model which showed the best capabilities to model the data was then selected and the process repeated, creating *n*-2 models each containing three inputs. This process was repeated until no significant improvement was gained from the addition of further inputs resulting in a final model containing the proteomic pattern which most accurately predicted between the two outcomes.

2.2.7.3 Cluster analysis of quality control (QC) samples

The spectra obtained from QC samples were visually checked as described before and for the cluster analysis of protein spectra, *m/z* values 6000-7000 Da and 9000-10000 Da were used and for the peptides from 800 to 3500 Da. The approved spectra were imported as ASCII files and the data was smoothed to reduce the dimensionality using SpecAlign software which was available online. The cluster analysis was performed using Statistica 7.0 software. In contrast to principal component analysis, cluster analysis does not reduce the number of characters but it will reduce the number of the subjects by placing them in groups. If the different runs are seen to mix together well it means that the instrument and protocols are working properly, but if one run is not mixing with the other runs it means some thing is wrong with the instrumentation or the protocols and the run was rejected.

2.2.8 RP-SPE fractionation

This experiment was carried out in collaboration with Dr Robert Layfield and Dr David Tooth at University of Nottingham. Briefly, serum fractionation was performed using SPE large pore (1000 Å) polystyrenedivinylbenzene (PDVB) 25mg resin (International Sorbent Technologies, mid-Glamorgan, UK). MS grade mobile phase (Riedel de Haën, Sigma) was

applied by vacuum, to condition the column using 70% (v/v) ACN (acetonitrile)/0.1% (v/v) TFA (trifluroacetate) and equilibrated with 0.1% TFA. Serum samples (100 µL, 7 mg of total protein) were then applied and washed through with 0.1% (v/v) TFA. Bound proteins were eluted using a stepwise 5−100% (v/v) ACN gradient. Eluates were dried using SpeedVac, re-suspended in loading buffer, and denatured prior to SDS-PAGE analysis. Samples were analysed using precast 4−12% acrylamide bis-tris SDS-PAGE gels (Invitrogen, Paisley, UK). Gels were stained with Coomassie blue.

2.2.9 1-Dimensional Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (1-D SDS PAGE)

A 12.5% resolving gel was made up by placing 8.76 ml 30% (w/v) Acrylamide (0.8% (w/v) Bis-Acrylamide stock solution (37.5:1)), 5.25 ml 1.5M Tris HCl buffer (pH 8.8), 7 ml water, 210µl of 0.1% ammonium persulphate (100mg in 1ml water) and 21µl TEMED, this was enough for 4 gels. 6ml of this mix was then placed in each of the four protein gel casts and 100µl of butanol was placed on top to keep the gel even and to stop dehydration while it set. Once set, the butanol was removed using filter paper and the appropriate combs were put on top of the casts. A 4% stacking gel was made up by adding 933µl of Acrylamide, 1.75 ml of 0.5M Tris HCl buffer (pH 6.8), 4.32 ml water, 70µl ammonium persulphate and 7µl TEMED and then 2ml of this mix was placed on top of the resolving gel in each cast. Once set, the combs were removed and the gels and casts were placed in to a tank containing 1 x running buffer.

The samples were prepared by mixing 40µg of sample with approximately 5µl of reducing sample buffer and denaturing the proteins for 5 minutes at 95°C.

The gels were completely covered in running buffer in the tanks and the samples were loaded into the wells, 5µl of a ProtoMetrics protein ladder was added in one well into each gel. Once loaded, the gels were run at 100V until the dye front reached the bottom of the stacking gel, approximately 30-40 minutes, the voltage was then increased to 150V until the dye front reached 1cm from the bottom of the gel, approximately 1 hour. The gels were stained with coomasie blue for 30-60 minutes and the excess stain was removed by placing the gels in destain (10% acetic acid and 30% methanol in HPLC grade water) over night.

2.2.10 Gel spot tryptic digestion

The gel spot was excised manually from the gel using a scalpel. Each gel piece was placed into an eppendorf with 500µl distilled water on a shaker. This was followed by distaining with 500µl of 50%ACN/40mM NH_4HCO_3 for 10 minutes on shaker. The solution was discarded and the distaining step was repeated once again. The gel piece was then dehydrated by adding 200µl of 100%ACN for 15 minutes while shaking. In addition, 15.2µl of distilled water, 33.2µl of NH_4HCO_3 (100 mM) and 2.3µl of 0.5µg/µl trypsin gold was added to the gel piece and this was incubated at 37°C overnight. The trypsinisation was stopped by adding 2µl of 1%TFA and samples were C_{18} ZipTipped (using Xcise robotic system) and prepared for ESI or MALDI mass spectrometry analysis as described previously.

2.2.11 Western Blot analysis

Serum samples were separated on a 12.5% SDS-PAGE (previously described). Gels were equilibrated in chilled transfer buffer for 5 minutes while shaking. The transfer buffer was discarded and equilibration step was repeated once with fresh buffer. Filter papers (8 pieces for each gel) were cut into required size (the size is based on the dimentions of the gels). One PVDF membrane per each gel was pre-wetted according to manufacturer's guide, briefly this was done by placing the PVDF membrane in 100% methanol (10 seconds), washed by water (5 minutes) and transfer buffer for 15-30 minutes. The transfer units were assembled as follow: 4 pieces of filter paper (remove the air bubbles in-between the filter papers), the PVDF membrane, the SDS gel and 4 pieces of filter paper on top. The SDS gel was transferred onto the PVDF membrane for 1 hour at 13 Volt. The membrane was stained with Ponceau S to ensure the complete transfer and then removed by placing them in water for 20 minutes. The PVDF membrane was washed twice with TBS each for 5 minutes. Membranes were blocked with a 5% milk solution containing 0.05% Tween 20 in TBS overnight followed by incubation with primary antibody at 4˚C overnight. Primary antibody was made in TBST with an appropriate dilution and incubated at 4˚C overnight. The primary antibodies used in this study included rabbit anti-mouse serum amyloid P (1 in 10000), chicken anti-mouse hemopexin (1 in 20000), goat anti-mouse haemoglobin (HBB 1:200) and goat anti-mouse serum amyloid A-1 (1 in 2000). The primary antibody was discarded and membrane was washed in TBST for 5 minutes and this was repeated once

more. The membranes were incubated with appropriate peroxidase-conjugated secondary antibody (made in TBST) at 1:1000 (SAA1, SAP and HBB) and 1:200000 (HPX) respectively for 2 hours at room temperature. This was followed by 4 washes with TBST, each 15 minutes to get cleaner background. To detect the membrane on the autoradiography film, the Amersham ECL Western blotting detection reagent was used. Equal volumes of solution 1 and 2 of this kit was mixed together prior to developing and was added to the protein side of the membrane, sufficient to cover the membrane surface. Membrane was incubated at room temperature for precisely 1 minute. The excess solution was removed and membrane was placed in a film cassette and an autoradiography film was placed on top of the membrane for the required time which varies for each experiment. Imaging and quantitation using densitometry were performed on Fuji image station as well as developing the film with a developer. Densitometric analysis was preformed using AIDA software.

2.2.12 Immunohistochemistry (IHC)

The CT26 colorectal tumours from progressor and TB animals collected from the prospective experiment were used for IHC staining. Frozen tissues were embedded directly onto the cork using OCT and were frozen in isopantane and liquid nitrogen. A piece of tumour tissue on cork was fixed directly onto the metal chuck using a small drop of OCT and freezing spray and then the chuck was placed onto the specimen holder of the cryostat. The tissues were then sectioned using a Leica 1900 cryostat while the specimen holder was cooled down to -20ºC and the cryochamber cooled down to -20ºC prior to use. The blade and the anti-roll plate were wiped with a small amount of acetone to prevent static and the specimen holder was set to cut 7µm sections and sections were transferred onto microscopic glass slides. The slides were air dried, and fixed with acetone for 15 minutes and stored at -80°C. Immunohistochemistry was performed using the ABC kit. After leaving the tissues in 0.5-1ml of hydrogen peroxide (0.03%) for 5 minutes, hydrogen peroxide was tipped of and the sections were blocked with diluted goat serum (10% goat serum in PBS) for 15 minutes. The primary antibody (rabbit polyclonal anti-mouse SAP, 1:100 in 5% goat serum/PBS) was added to the sections (approximately 50µl per tissue) and this was left at 4°C overnight. The excess antibody from the slide was washed by rinsing PBS and this was followed by incubation for 30 minutes with diluted biotinylated

secondary antibody (polyclonal goat anti-rabbit immunoglobulins/biotinylated, 1:1000 in 5% goat serum/PBS) solution and then 1ml ABC reagent was added for 30 minutes at room temperature. The slide was rinsed with PBS and 1ml of DAB reagent (for visualization) was added to the slides for 10 minutes at room temperature. The slide was then rinsed with PBS and this was followed by leaving the slides in distiled H_2O for 5 minutes. The slides were dried at room temperature which was followed by Harris staining (placing the slides in Harris staining for 8-10 seconds). The DAB and Harris staining was then alcohol fixed by sequentially placing the slides in 70% ethanol (1 minute), 100% ethanol (1 minute), 100% ethanol (2 minutes), 100% xylene (I minute) and 100% xylene (2 minutes). The slide was air dried in the hood followed by mounting the slide with a cover-slip using few drops of premount. Once the slide was dried, then it was examined using a microscope.

2.2.13 ELISA assay

Absolute quantification of mouse serum amyloid P (SAP), mouse serum amyloid A (SAA) and mouse hemopexin (HPX) in serum samples were determined using ELISA.

2.2.13.1 SAP ELISA assay

Commercially available sandwich ELISA for SAP was purchased from Immunology Consultant Laboratory (Immunology Consultant Laboratory, Inc., US). The assay was performed according to manufacturer's protocols. In the typical assay procedure, all incubations were performed at room temperature. Serum samples were diluted in s 1:50000. This assay was carried out on samples from the retrospective experiment and validated using the prospective samples. Each calibrator or serum samples (100 µl) was added to the wells and incubated for 1 hour. After that the wells were aspirated and washed 5 times with washing buffer (600 µL), enzyme-antibody conjugate (100 µL) was added to the wells and incubated for 1 hour. The wells were aspirated and washed again, and then TMB substrate solution (100 µl) was added to each well. After 10 minutes of incubation (enzyme reaction), stop solution (100 µl) was added, and the absorbance at 450 nm was measured with a microplate reader system. Experiments were performed in duplicate.

2.2.13.2 SAA ELISA assay

Commercially available sandwich ELISA for SAA was purchased from Immunology Consultant Laboratory (Immunology Consultant Laboratory, Inc., US). The assay was performed according to manufacturer's protocols. In the typical assay procedure, all incubations were performed at room temperature. Serum samples were diluted in s 1:20000. This assay was carried out on samples from the retrospective experiment and validated using the prospective samples. Each calibrator or serum samples (100 µl) was added to the wells and incubated for 1 hour. After the wells were aspirated and washed 5 times with washing buffer (600 µL), enzyme-antibody conjugate (100 µL) was added to the wells and incubated for 1 hour. The wells were aspirated and washed again, and then TMB substrate solution (100 µl) was added to each well. After 10 minutes of incubation (enzyme reaction), stop solution (100 µl) was added, and the absorbance at 450 nm was measured with a microplate reader system. Experiments were performed in duplicate.

2.2.13.3 HPX ELISA assay

Commercially available sandwich ELISA for HPX was purchased from Immunology Consultant Laboratory (Immunology Consultant Laboratory, Inc., US). The assay was performed according to manufacturer's protocols. In the typical assay procedure, all incubations were performed at room temperature. Serum samples were diluted 1 in 50000. This assay was carried out on samples from the retrospective experiment and validated using the prospective samples. Each calibrator or serum samples (100 µl) was added to the wells and incubated for 1 hour. After the wells were aspirated and washed 5 times with washing buffer (600 µL), enzyme-antibody conjugate (100 µL) was added to the wells and incubated for 1 hour. The wells were aspirated and washed again, and then TMB substrate solution (100 µl) was added to each well. After 10 minutes of incubation (enzyme reaction), stop solution (100 µl) was added, and the absorbance at 450 nm was measured with a microplate reader system. Experiments were performed in duplicate.

2.2.14 RNA isolation and Real-Time Polymerase Chain Reaction (RT-PCR)

Total RNA was isolated from the CT26 cell lines and the different tissues using RNA STAT-60™ according to the manufacturer's protocol. For extracting RNA from frozen tissues, the tissue was ground under liquid nitrogen in a Class II safety cabinet using a

mortar and pestle to a powder. The powder then transferred to a 1.5 ml Eppendorf tube to which 1ml of RNA STAT-60™ was added and stored at room temperature (RT) for 5 minutes. For extracting RNA from the adherent CT26 cells, 1ml RNA STAT-60™ was directly added to the culture vessel to lyse them. The homogenate was then transferred to a 1.5 ml Eppendorf tube and stored at room temperature (RT) for 5 minutes. Chloroform (0.2 ml) was added to the homogenate in the hood and the tube was shaken vigorously by hand for 60 seconds and left at RT for 3 minutes. The samples were then centrifuged at 14,000 rpm for 15 minutes at 4°C. The colourless upper aqueous phase which contained the RNA was transferred to a fresh 1.5 ml Eppendorf tube, mixed with 0.5 ml of isopropanol and the tube stored at RT for 8 minutes. The samples were then again centrifuged at 12,000 rpm for 10 minutes at 4°C. The supernatant was discarded and the white pellet was washed with 1ml 75% ethanol. The RNA pellet was dried in the fume hood and re-suspended in double-distilled water (ddH_2O) by passing pipetting up and down a few times. The concentration and purity of the isolated RNA was measured using a UV spectrophotometer. Samples were stored at -80°C until use.

Total RNA was reverse transcribed into cDNA using M-MLV RT and random primers (Promega, UK) following manufacturer's instructions. Briefly, in a 0.5 ml Eppendorf tube, 2 µg of total RNA (or all the RNA solution if less than 2 µg available) was mixed with 0.5 µl of random primer solution and ddH_2O was added to make the final volume to 10 µl. The tube was heated in a thermal block at 70°C for 5 minutes and immediately cooled on ice followed by a brief spin in a microcentifuge. The following mix of components was added to the tube at RT: 5 µl of 5X M-MLV RT buffer (Promega, UK), 1 µl of dNTP (12.5 mM), 0.7 µl of RNasin® ribonuclease inhibitor (Promega, UK), 1 µl of M-MLV RT and 7.3 µl of ddH_2O. The tube was gently mixed and incubated in a waterbath at 37°C for 60 minutes. The tube was then immediately cooled on ice, spun briefly and the reverse transcription reaction was stopped by heating to 95°C for 5 min. Finally, cDNA was stored at -20°C until use.

Real-time quantitative PCR: Primer sequences for HBB, SAA-1, SAP, HPX and the house keeping genes glyceraldehyde-3-phosphate dehydrogenase (Gapdh) and hypoxanthine guanine phosphoribosyl transferase 1 (Hprt1) were either taken from literature when available or designed to cover two different exons in the sequence to eliminate amplification of trace amounts of genomic DNA in the cDNA samples where possible. Also all primers were designed to generate PCR products under 250 base pairs in size to

optimise the quantitative RT-PCR. Primer design was carried out with the assistance of the Primer-BLAST program (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). The primers were supplied by Eurofins MWG Operon (UK). The name, the sequences and the temperature of annealing (Ta) of all the primers used in this study are listed in table 2-2.

Gene	Accession number	Forward Primer	Reverse Primer	Ta
GAPDH	NM_008084.2	TGACGTGCCGCCTGGAGAAA	AGTGTAGCCCAAGATGCCCTTCAG	60
HPRT1	NM_013556.2	GCTTGCTGGTGAAAAGGACCTCTCGAAG	CCCTGAAGTACTCATTATAGTCAAGGGCAT	60
HBB	XM_903245.1	TCAGAAACAGACATCATGGTGC	TAGACAATAGCAGAAAAGGGGC	60
HPX	NM_017371.2	ATCTCAGCGAGGTGGAAGAA	AACCACTTGCGGTTACCTTG	60
SAA1	NM_009117.3	CCCAGGAGACACCAGGATG	TCATGTCAGTGTAGGCTCGC	60
SAP	NM_011318.2	AGCCTTTTGTCAGACAGACCTC	TGTCTCTGCCCTTGACACTG	60

Table 2-2. The name, the sequences and the temperature of annealing (Ta) of the primers.

The Real-time quantitative PCR reactions were performed using the Rotor-Gene 6000 (Corbett Research, UK) with iQ™ SYBR® Green Supermix. The samples were run in duplicates. Thermocycling for each reaction was done in a final volume of 12.5 µl containing 0.5 µl of template or standard, 6.25 µl of SYBR® Green Supermix, 0.5 µl of gene-specific forward primer, 0.5 µl of gene-specific reverse primer and 4.75 µl of ddH_2O. In each experiment, 4 no template controls for each individual mastermix were included to ensure there was no contamination and also to indicate the degree of amplification due to primer dimers. The cycling conditions were as follows: melting step at 95°C for 3 minutes, (denaturation at 95°C for 30 seconds, annealing at primer-specific Ta for 30 seconds, extension at 72°C for 30 seconds) x 40 cycles (house keeping genes) or x 45 cycles (genes of interest). Finally, a melting (dissociation) curve was acquired by slowly ramping the temperature from Ta to 95°C by 1°C increment. The fluorescence of each sample was measured at the end of extension step and at the end of each cycle in the case of the dissociation curve. The C_t value, which corresponds to the number of cycles at which the reaction crosses a threshold value (fluorescence exceeds the background level), was calculated by the software to give the standard curve. Following completion of the quantitative RT-PCR reaction, the C_t value for each sample was recorded. Assuming that primer efficiencies are similar for the genes studied, the standard curve can be omitted from the assay, allowing more samples to be screened. The comparative C_t method was therefore

used for the quantification of the results. Following completion of the quantitative RT-PCR that did not include a standard curve, the analysis of the results was carried out using the comparative C_t method, also known as the $2^{-[delta][delta]Ct}$ method, where $[delta][delta]C_t=[delta]C_{t,sample}-[delta]C_{t,\ reference}$. Here, $[delta]C_{t,sample}$ is the Ct value for any sample normalised to the housekeeping gene and $[delta]C_{t,\ reference}$ is the C_t value for the calibrator also normalised to the housekeeping gene.

The result is given as a relative gene expression level of the gene of interest in the tissues or cell lines. Calculations were performed with Microsoft® Office Excel 2003 (Microsoft Inc., Redwood, Wash.).

Chapter 3 – Development, Optimisation and Evaluation of Methodology for MALDI Serum Proteome Profiling

3.1 Introduction

3.1.1 The debate over reproducibility of serum proteomic profiling by MALDI/SELDI MS for cancer biomarker discovery

The recent emergence of high throughput technologies such as MALDI and SELDI mass spectrometry opened new horizons in clinical proteomics for rapid biomarker discovery by serum proteome profiling and produced exciting and encouraging results (Matharoo-Ball *et al.* 2008, Munro *et al*. 2006, Ricolleau *et al*. 2006, Taguchi *et al*. 2007). The MS-based studies aimed to compare serum proteomic patterns of healthy and cancerous individuals to detect discriminatory diagnostic signatures that can be adapted for clinical use. However, shortly after publication of the early investigations, concerns were raised regarding the reproducibility of the results as well as the technical and experimental design. In addition, the ability of MALDI/SELDI to detect low abundant proteins/peptides as tumour biomarkers have been questioned (Diamandis 2006, Ransohoff 2005, Diamandis and van der Merwe 2005).

In 2002 Petricoin *et al*. was the first group, to demonstrate serum proteomic profiling by SELDI as a new approach for diagnosis of ovarian cancer. They reported a panel of anonymous markers that correctly classified ovarian cancer patients from healthy individuals based on the SELDI serum proteomic patterns with 100% sensitivity, a specificity of 95% and a positive predictive value of 94% (Petricoin *et al*. 2002). The classification algorithm used by Petricoin *et al*. to analyse the SELDI proteomic patterns was generated using 50 healthy and 50 ovarian tumour samples and subsequently tested the model on a blind set of samples which was able to classify 50/50 of the ovarian patients, 47/50 of the healthy correctly. In addition, the classification algorithm was tested on 16 benign patient samples, all of which were classified separately from healthy or cancer

patient samples. Since this encouraging publication, numerous investigators have applied similar MS-based technology to identify biomarkers for other cancers such as prostate, colon, breast, bladder, head and neck and melanoma (Simpkins *et al*. 2005, Hortin 2006), mostly reporting high sensitivity and specificity for classification. However, shortly after these publications the technology was the subject of criticism when attempts to reproduce and validate some of the previous results failed (Li *et al*. 2005, Karsan *et al*. 2005, Baggerly *et al*. 2004, Sorace and Zhan 2003). The most well-known study that was subjected to several criticisms was SELDI serum profiling of ovarian cancer study by Petricoin and colleagues (Petricoin *et al.* 2002). There are three datasets available online from the SELDI serum profiling of ovarian cancer samples analysed by Petricoin and coworker which are referred as datasets 1-3 (Baggerly *et al.* 2004). The first dataset is the initial ovarian cancer paper published by Petricoin *et al.* using Ciphergen H4 ProteinChip arrays to obtain SELDI profiles from serum samples and 216baseline subtracted spectra were used for the analysis which were divided into training cancer, training healthy, test cancer, test healthy and test healthy and benign disease (Petricoin *et al.* 2002). The second data set used the same 216 samples from the first dataset and SELDI profiles were obtained using Ciphergen WCX2 ProteinChip array and spectra were baseline subtracted. The final dataset available from Petricoin *et al.* used new set of samples, prepared robotically on Ciphergen WCX2 ProteinChip array and no baseline subtraction was carried out on the spectra. The ovarian cancer data provided by Petricoin and coworkers was reassessed by Baggerly *et al.* to determine the reproducibility of the technology. Baggerly *et al.* (2004) compared the two ovarian datasets (dataset 1 and dataset 2) obtained from the same set of samples and made the assumption that although there were discriminatory differences between the cancer and control subjects within each data set, these differences were not the same between the two experiments. They suggested that pre-analytical and analytical variations may introduce variations within and between experiments and that these artifacts were then be picked up as differences between the groups by the classification algorithm while they were not truly related to the disease. Parallel with Baggerly *et al.* unsuccessful attempts to reproduce the ovarian results of Petricoin and coworkers, a study by Sorace and Zhan on the third ovarian data set also failed to find reproducible diagnostic patterns for ovarian cancer (Sorace and Zhan 2003). Both these and similar studies suggest that inconsistency in experimental design and preanalytical procedures introduce artifacts within and between experiments that

might account for failure to reproduce and validate some of the MS-based proteomic profiling studies.

A more recent study by Li *et al.* assessed the validation and reproducibility of three previously identified biomarkers (three biomarkers termed BC1, BC2 and BC3) for breast cancer by SELDI serum proteomic profiling (Li *et al.* 2005). They used an independent set of sera from breast and healthy patients collected in a different institute from the original study samples. They were able to reproduce two (*i.e.* BC2 and BC3) out of the three biomarkers however, they identified the proteins corresponding to the two protein peaks as fragments of complement C3a which is one of the most abundant serum proteins produced by liver. In addition, attempts to identify discriminatory peaks in serum in different biomarker discovery experiments revealed that these discriminatory peaks were mostly fragments of abundant proteins and therefore not specific enough to be considered as disease biomarkers originating from the tumour (Diamandis 2006). Furthermore, a study by Honda *et al.* reported a panel of four biomarkers that classified plasma protein SELDI profiles of pancreatic cancer patients from healthy individuals with a sensitivity of 97.2% and a specificity of 94.4% (Honda *et al.* 2005). They attempted to improve their study by validating their findings on an independent set of samples and succeeded in validating their classification algorithm with a sensitivity of 90.9% and a specificity of 91.1%, although the small numbers of samples used in the validation set restricted the generalisation of the findings and therefore no significant conclusion could be drawn.

The above studies are representative of experiments carried out by various research groups, raising critical questions regarding the reproducibility of the technique. The evidence that pre-analytical and analytical variability has a significant impact on protein/peptide profiles cannot be denied and shows the susceptibility of this technology to production of artifacts which manipulates the final results. The unsuccessful validation studies prompted researchers to improve the shortcomings of MS-based serum proteomic profiling methods by applying standardised and carefully designed protocols. A number of recently published reports have addressed various aspects of the reproducibility, pre-analytical variability and analytical issues in MS-based proteomics in order to avoid previous shortcomings and facilitate the progress in the field (Matharoo-Ball *et al.* 2007, Tiss *et al.* 2007, Callesen *et al.* 2008, Peng *et al.* 2009, Semmes *et al.* 2005).

3.1.2 Effects of preanalytical, analytical and postanalytical variability in MS-based serum proteomic profiling: How to overcome the issues

The preanalytical step refers to all the procedures performed prior to analysis of the specimen by mass spectrometer. The analytical phase includes the mass spectrometry analysis and the postanalytical phase refers to the bioinformatic tools used for data mining (Rai and Vitzthum 2006).

Little effort has been made to investigate the reproducibility of the spectra generated by MALDI-TOF-MS instrumentation which has a higher resolution when compared to SELDI-TOF-MS (Coombes *et al.* 2005). In the proposed study, MALDI-MS was used as the method of choice (due to its high-throughput and high sensitivity) for the analysis of the mouse and QC serum peptides and proteins. In addition, the initial results of reproducibility studies on the MALDI-MS instrument and effects of automated sample handling on both protein and peptide patterns acquired from naïve mice and QC serum is presented. Preliminary results of novel method development for enrichment of low molecular weight (LMW) peptides are also reported which uses C_4 ZipTip in conjugation with C_{18} ZipTip fractionation prior to tryptic digestion of human serum. Standardisation and optimisation of these protocols is essential before further experiments are preformed on limited material.

3.1.3 Aims and objectives

Detailed objectives for this part of the study were as follow:

- To programme the Xcise robotic liquide handler to perform the required sample preparation strategies involving the use of C_{18} ZipTips.
- To optimise the optimum dilution for serum prior C_{18} ZipTip clean-up.
- To investigate the effects of C_{18} ZipTips on removal of high molecular mass proteins from serum.
- To investigate the use of C_4 ZipTip in conjunction with C_{18} ZipTip for serum sample preparation.
- To investigate the reproducibility of Xcise robotic system and MALDI mass spectrometry within and between runs, using QC sampl

3.2 Results

3.2.1 Optimised sample volume and dilution for Xcise robotic system sample processing

An automated system for sample preparation prior to MALDI-MS analysis was utilised in this study. The robotic system was programmed to carry out a previously established manual sample preparation and processing protocol described by Matharoo-Ball and colleagues (Matharoo-Ball *et al.* 2007). Briefly, a 1:10 dilution of serum samples in 0.1% TFA were subjected to C_{18} ZipTip clean-up and then spotted on the MALDI plate for protein profiling. In addition, tryptic digestion was performed and samples were analysed by MALDI for peptide profiling. Moreover, in this study protein profiles of 1:10 dilution of serum without C_{18} ZipTip clean-up were also examined. Initially, the effects of sample dilution and volume used on the robotic system were examined and optimised and the workflow is presented in figure 3-1.

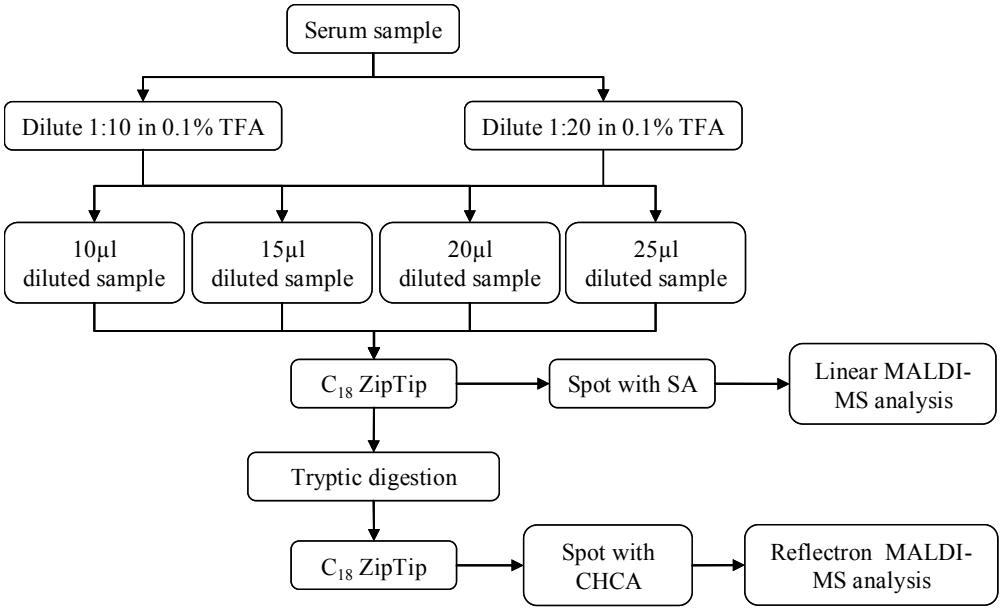


Figure 3-1. Workflow of sample preparation.

Serum was diluted 1:10 and 1:20 with 0.1% TFA and different volumes (10µl, 15µl, 20µl and 25µl) of each were used on the robotic system which either directly spotted onto the

sample plate or subjected the sample to C_{18} ZipTip clean-up. This was followed by tryptic digestion of serum and peptide analysis. Figure 3-2 shows the results of different serum dilution and the protein spectra obtained from peaks that were prominent for in the 1:20 dilution of the sample compared to the 1:10 dilution of serum. This demonstrated that both dilutions showed similar profiles that could be reproduced. Therefore, the 1:20 dilution of serum was used in future experiments.

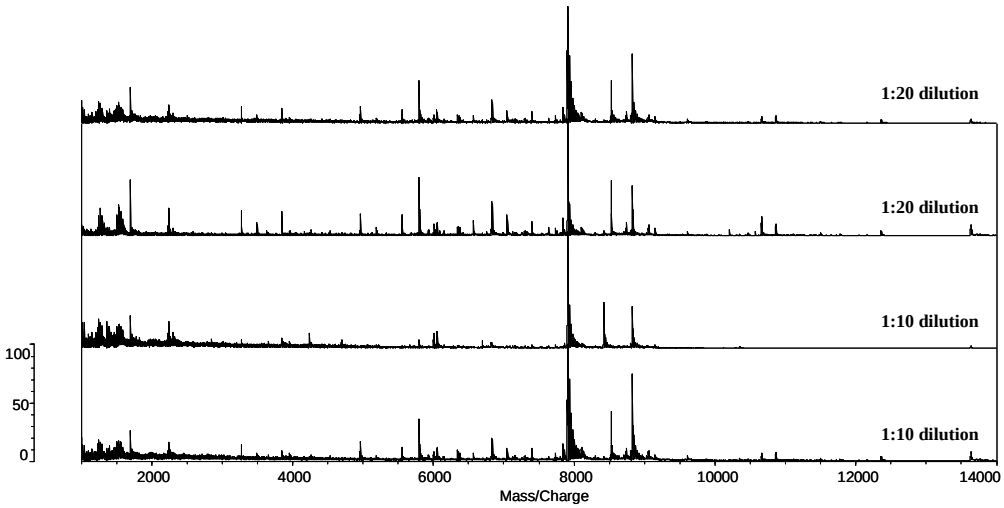


Figure 3-2. Dilution of serum samples with subsequent C_{18} ZipTip clean-up. Replicate MALDI-TOF analysis of 1:10 and 1:20 dilution of same sample after C_{18} ZipTip obtained from MALDI linear mood. The 1:20 dilution of sample showed higher number of peaks and better intensity compare to 1:10 dilution of sample.

Use of different starting volumes of diluted serum processed by the Xcise robotic system was also examined. The results of the 1:20 dilution of serum after C_{18} ZipTip clean-up using different starting volumes (10 µl, 15 µl, 20 µl and 25 µl) are presented in figure 3-3 and the regions with significant differences are highlighted. Use of different starting volume on the Xcise robotic system appeared to affect the quality of both protein and peptide MALDI spectra. The MALDI-TOF analysis of serum protein profiles after C_{18} ZipTip clean-up using volumes of 20 µl and 25 µl were shown to give the best protein profiles with higher numbers of peaks and of higher intensities (figure 3-3A). The same samples were subjected to tryptic digestion and peptide profiles were obtained by MALDI-MS and the results are presented in figure 3-3B. As shown in the figure 3-3B, using an initial volume of 25 µl results in superior spectra (figure 3-2B). Spectra obtained from the

using 25 µl show higher number of peaks with higher intensities in the lower mass range (1000-1200 Da) of spectra as highlighted, compare to the use of other volumes. These low mass peaks as shown in figure 3-3B, are not present or present in lower intensities in the spectra obtained from 10 µl, 15 µl and 20 µl samples. In conclusion, a significant number of peaks are present in both protein and peptide spectra when 25 µl of sample were utilised for processing, which are not apparent using other volumes and therefore this volume was optimum and was applied in all future experiments.

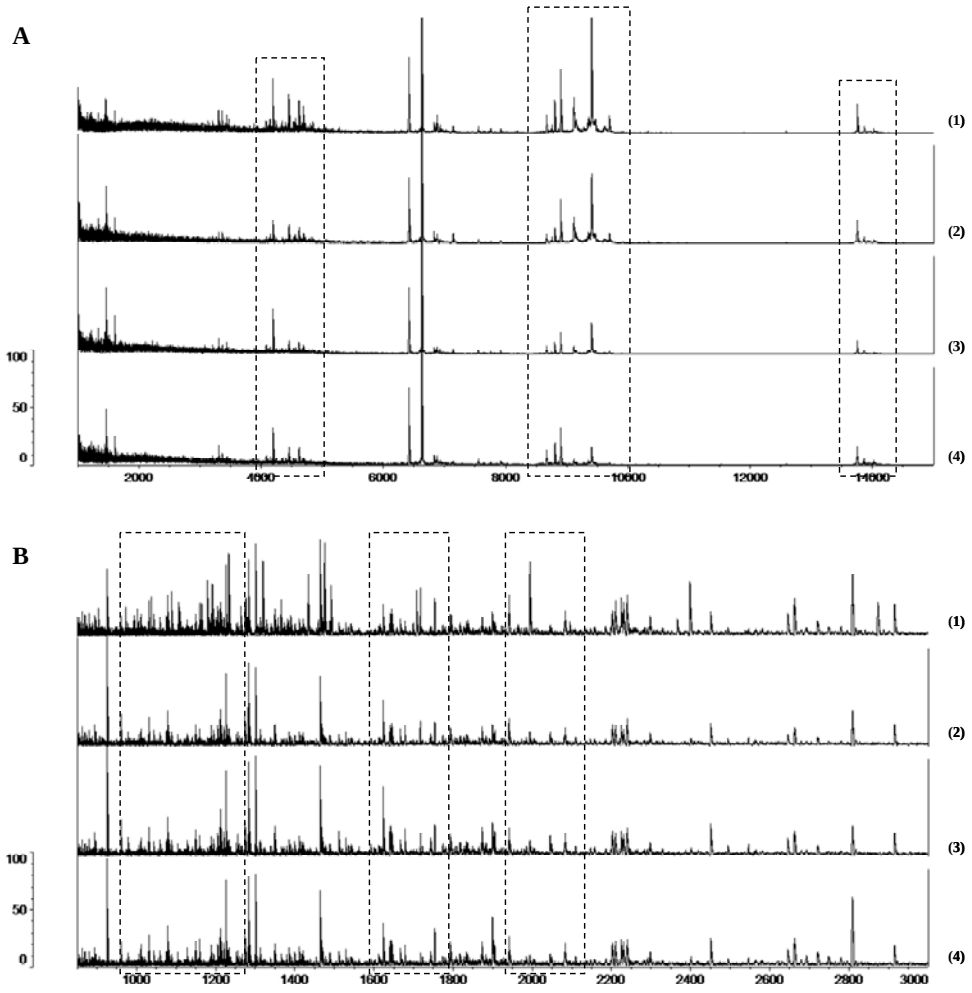


Figure 3-3. Impacts of different sample volumes used on the Xcise system, on the quality of serum protein (A) and tryptic peptide (B) MALDI spectra. A serum sample was diluted 1:20 in 0.1% TFA and 25µl **(1)**, 20µl **(2)**, 15µl **(3)** and 10µl **(4)** of this was used as the initial volume of sample on the Xcise robotic system prior to MALDI-TOF analysis. Figure A shows serum protein spectra after C_{18} ZipTip clean-up corresponding to each of the initial volumes used on Xcise and figure B is the MALDI-TOF tryptic peptide profiles of the same samples. The best spectra for both protein and tryptic peptides were obtained from the initial sample volume of 25µl which showed spectra with higher number of peaks and better intensities in comparison to other initial sample volumes and some of these differences are highlighted.

3.2.2 Evaluation of C_{18} ZipTip solid phase extraction for MS serum profiling

As C_{18} ZipTip chromatography is the main method used in this study and as part of standardization of the protocols in use, we examined the efficiency of this column in removal of high mass proteins such as albumin from serum samples. In order to examine this, the 1:20 dilution of mouse serum was spotted with SA before C_{18} ZipTip clean-up on the MALDI plate. Then the same aliquot of sample was put through C_{18} ZipTip clean-up and the eluted material from the column was spotted with SA on the MALDI plate. In addition, the remnant serum sample which was left behind after C_{18} ZipTip binding was spotted with SA on the MALDI plate. The samples were analysed with MALDI-TOF operated in linear mode and representative spectra are shown in figure 3-4. The mass value for albumin is approximately 66 kDa and as it can be seen in figure 3-4, spectra obtained from 1:20 dilution without any clean-up contains a high amount of albumin however, significant removal of albumin after use of C_{18} ZipTips was achieved. There are also more peaks with higher resolution with low molecular weights (LMW) observed which demonstrates that with the use of C_{18} ZipTips we can to some extent enrich the LMW components of serum samples.

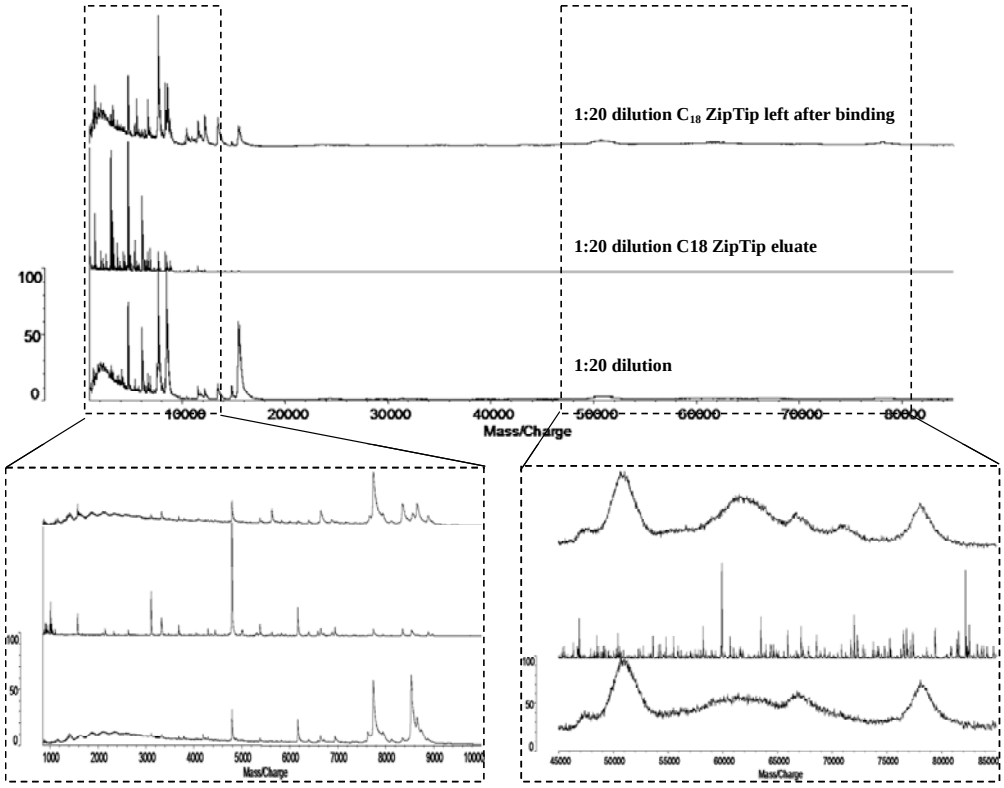


Figure 3-4. Effects of C $_{18}$ ZipTip on removal of serum albumin. Figure A represents MALDI-MS protein profile of 1 in 20 dilution of serum sample spotted with SA and analysd in linear mode. Figure B represents the same aliquot of sample after C $_{18}$ ZipTip spotted with SA and analysed in linear mode in which significant removal of albumin is observed. Figure C represents spectra from remain of sample after C $_{18}$ ZipTip spotted on the plate with SA.

3.2.3 Within run reproducibility of protein and peptide MALDI spectra profiles of mouse serum

Reproducibility studies were preformed on replicate aliquots of the same serum, which were spotted and analysed by MALDI-TOF before and after C_{18} ZipTip chromatography clean-up (protein profiling) and after tryptic digestion (peptide profiling). This study allows the assessment of spot-to-spot variation within a single MALDI run for the same aliquot of mouse serum sample. The analytical procedures were carried out using a robotic system, followed by MALDI-TOF analysis of proteins in linear mode and peptide analysis in reflectron mode. Representative spectra of protein profiles without C_{18} ZipTip clean-up (*n* =

8), after C_{18} ZipTip clean-up (n = 8) and tryptic peptide (n = 6) analysis are show in figure 3-5. Visual inspection of both protein and tryptic peptide spectra indicated a satisfactory spectra quality and reproducible for both protein replicates.

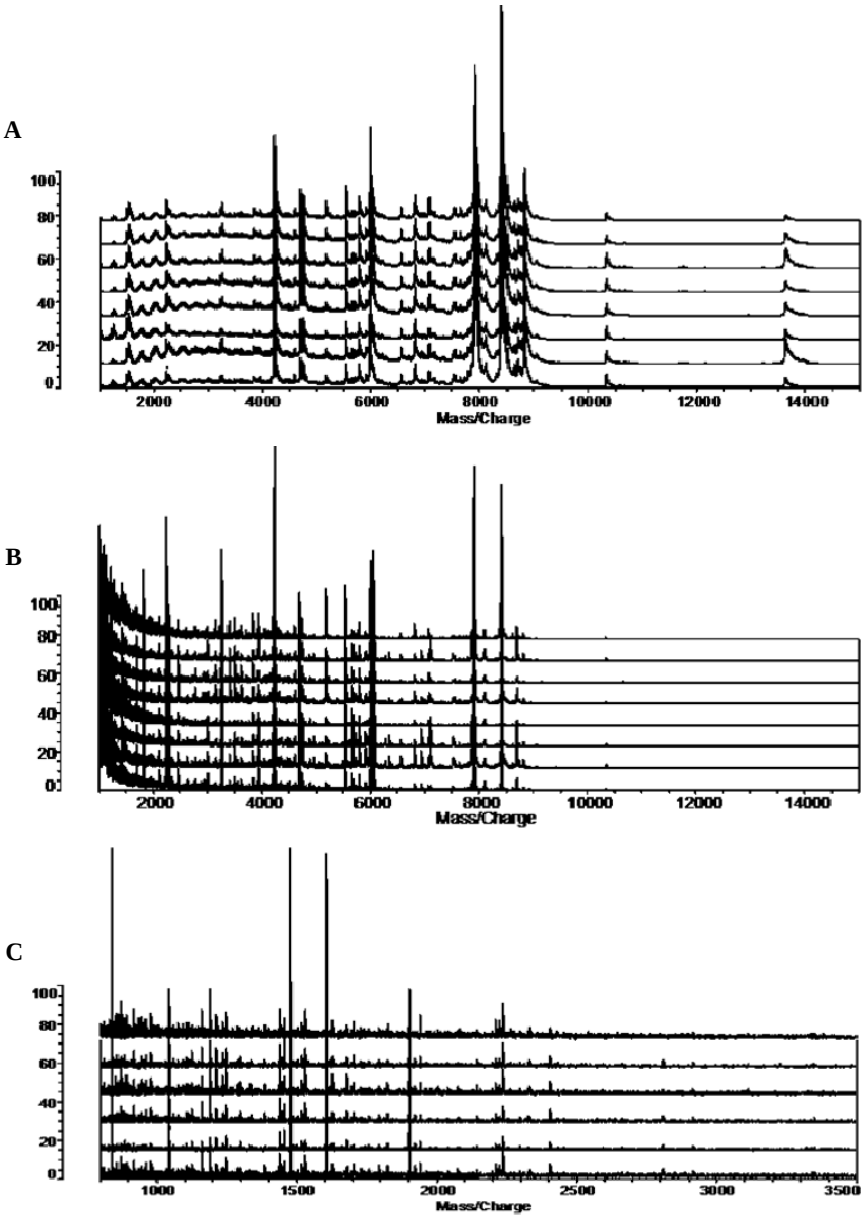


Figure 3-5. Replicate MALDI-TOF analysis of serum protein and tryptic peptides processed by Xcise robotic system. Replicate MALDI-TOF analysis of aliquots of the same sample diluted 1:20 in 0.1% TFA and analysed in linear mode **(A)** without C_{18} ZipTip clean-up and **(B)** after C_{18} ZipTip. **(C)** Replicates of tryptic peptide profiles following digestion and MALDI-TOF analysis in reflectron mode.

As can be seen from table 3-1, the mass accuracy both in linear and reflectron mode is very high with CVs for linear mode showing less than 0.06% and in reflectron mode showing CVs less than 0.05%. For each of these three types of profiling experiments, 5 peaks that were constantly present in the spectra were identified and used to assess the reproducibility. The peaks range from a low to high mass and for the protein profiles these 5 peaks were selected within the range of 1000-15000 Da and 800-3500 Da for the tryptic peptides. The mean and CV for the *m/z* values and normalised intensity ratios were calculated and shown in table 1-1A for the serum protein profiles without C_{18} ZipTip (n = 8), table 1-1B for serum protein profiles after C_{18} ZipTip clean-up (n = 8) and table 1-1C for digested serum profiles (n = 6). The CVs for serum protein profiles without C_{18} ZipTip were between 15-25.3%, after C_{18} ZipTip ranged from 13.6-44.5% and for the serum tryptic digested peptides ranged from 16.2-33.3%. The mass accuracy and resolution increased for the peptide profiles because they can be measured in the reflectron mode by a TOF mass analyser however, range of CV is wider for proteins. Although, these results are promising but it is essential to determine the reproducibility of peaks between runs which will be carried out in future experiments.

A

Linear mode	Mass	Normalized intensity ratio	Mass	Normalized intensity ratio	Mass	Normalized intensity ratio	Mass	Normalized intensity ratio	Mass	Normalized intensity ratio
	2244.8	0.62	4245.7	2.68	5557.4	1.13	8426.2	6.88	10377.3	0.12
	2245.5	0.91	4241.8	3.77	5557.6	1.80	8428.3	8.36	10386.0	0.18
	2243.8	0.95	4242.2	3.08	5555.3	1.56	8428.6	8.32	10371.2	0.12
	2243.8	0.99	4246.6	3.36	5559.3	1.51	8428.9	7.63	10364.3	0.09
	2242.6	0.97	4246.7	3.28	5562.3	1.30	8429.1	7.62	10371.8	0.12
	2244.0	1.28	4243.9	4.32	5556.3	1.60	8428.3	10.17	10370.3	0.16
	2246.7	0.97	4247.6	4.02	5559.1	2.14	8429.1	10.34	10369.7	0.20
	2244.3	1.11	4246.6	3.49	5557.5	1.67	8428.3	9.78	10370.5	0.16
Mean	**2244.44**	**0.98**	**4245.14**	**3.50**	**5558.10**	**1.59**	**8428.35**	**8.64**	**10372.64**	**0.14**
CV%	**0.06**	**19.04**	**0.05**	**15.02**	**0.04**	**19.26**	**0.01**	**15.10**	**0.06**	**25.33**

B

Linear mode	Mass	Normalized intensity ratio	Mass	Normalized intensity ratio	Mass	Normalized intensity ratio	Mass	Normalized intensity ratio	Mass	Normalized intensity ratio
	2244.1	2.46	4244.5	3.32	6004.0	0.96	7915.8	2.29	8424.8	1.22
	2241.8	2.37	4243.6	3.75	6004.4	1.03	7916.4	2.21	8425.4	1.67
	2243.6	2.23	4245.2	2.38	6005.5	0.58	7917.0	0.86	8426.1	0.89
	2243.9	2.59	4245.1	4.07	6005.3	1.05	7916.9	1.29	8426.1	1.44
	2245.0	1.87	4246.1	2.55	6007.0	0.69	7917.6	1.15	8427.8	1.25
	2243.5	1.99	4245.0	2.33	6005.8	0.65	7917.4	0.72	8427.0	1.02
	2241.9	2.39	4243.7	2.95	6004.3	0.75	7916.9	2.23	8425.9	1.03
	2240.0	1.76	4241.2	2.80	6002.2	1.14	7914.6	1.08	8423.0	2.24
Mean	**2242.96**	**2.21**	**4244.31**	**3.02**	**6004.81**	**0.86**	**7916.57**	**1.48**	**8425.75**	**1.35**
CV%	**0.07**	**13.60**	**0.04**	**21.28**	**0.02**	**24.79**	**0.01**	**44.54**	**0.02**	**32.61**

C

Reflectron mode	Mass	Normalized intensity ratio	Mass	Normalized intensity ratio	Mass	Normalized intensity ratio	Mass	Normalized intensity ratio	Mass	Normalized intensity ratio
	1048.8	0.64	1194.0	0.49	1610.3	1.40	2241.0	0.38	2812.3	0.11
	1048.8	0.45	1194.0	0.41	1610.2	2.47	2240.9	0.33	2810.8	0.13
	1048.9	0.75	1194.0	0.59	1610.4	2.05	2241.0	0.33	2812.6	0.15
	1048.7	0.88	1193.9	0.75	1610.3	2.22	2240.9	0.43	2808.8	0.12
	1048.8	0.56	1194.0	0.51	1610.3	1.45	2240.9	0.29	2810.3	0.11
	1048.8	0.49	1194.0	0.43	1610.3	0.93	2241.0	0.29	2812.4	0.07
Mean	**1048.79**	**0.63**	**1193.98**	**0.53**	**1610.31**	**1.75**	**2240.92**	**0.34**	**2811.18**	**0.11**
CV%	**0.004**	**25.69**	**0.003**	**23.66**	**0.002**	**33.25**	**0.002**	**16.15**	**0.05**	**23.90**

Table 3-1. Reproducibility data for serum protein and tryptic peptides. Masses (*m/z)* and intensities (after normalisation) acquired by MALDI-TOF-MS in linear mode for proteins and in reflectron mode for the digested peptides are shown as mean values and their respective CV. The CV for five selected protein peaks with normalized intensities in each of the regions between the range of *m/z* 1000-15000 Da was less than 25.3% **(A)** and 44.5% **(B)** for the protein profiles without C_{18} ZipTip and after C_{18} ZipTip respectively. The tryptic peptides lead to CV of normalised intensities between 16.2-33.3% in the regions between the range of *m/z* 800-3500 Da **(C)**.

3.2.4 Method development for enrichment of low molecular weight peptides using C_4 and C_{18} ZipTips

We examined the efficiency of a new sample preparation protocol and compared it with the established protocol that employs the use of C_{18} ZipTip. Briefly, the proposed protocol involved the use of C_4 ZipTips in conjugation with C_{18} ZipTips clean-up that aimed to remove the higher molecular weight compounds and enrich the LMW peptides and/or proteins. Serum was passed through the C_4 ZipTips and the eluate was then passed through C_{18} ZipTips. This was followed by spotting the eluate with SA and on the MALDI target plate and MALDI analysis in linear mode. Tryptic digestion was carried out on the eluate overnight and after C_{18} ZipTips clean-up the samples were spotted with CHCA on the MALDI target plate for peptide profiling in reflectron mode. The use of C_4 ZipTips instead of the C_{18} ZipTips was also investigated in which serum was C_4 ZipTipped and then digested which was followed by a C_4 ZipTip clean-up and MALDI analysis. Representative spectra of the different sample preparation prior to MALDI analysis are shown in figure 3-6.

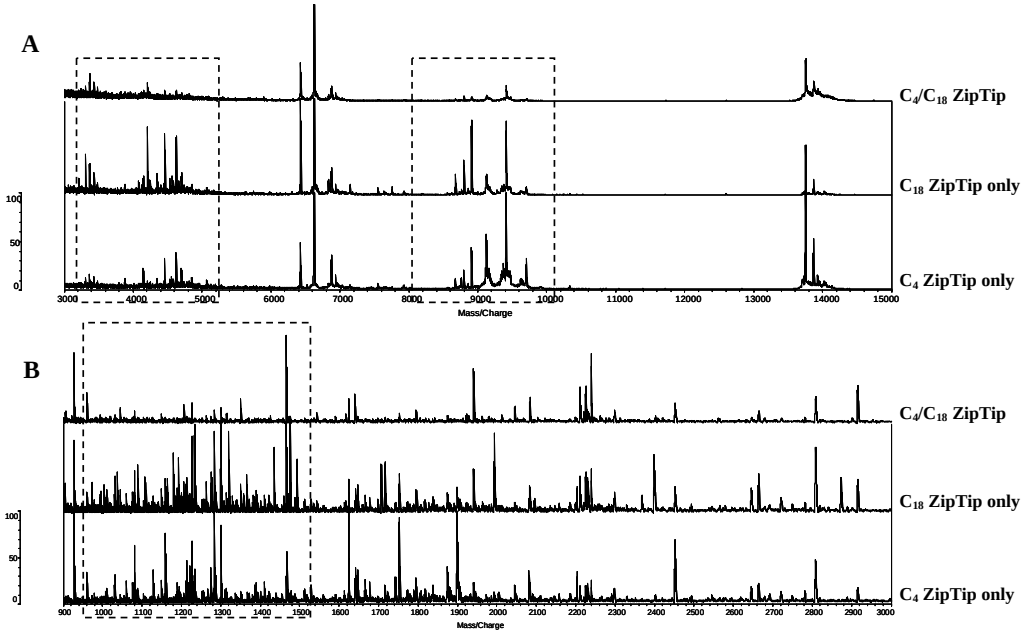


Figure 3-6. Comparison of protein and tryptic peptide MALDI spectra with different sample preparation protocols. (A) Representative MALDI-TOF spectra of same serum proteins after either C_4, C_{18} or C_4 followed by C_{18} ZipTip clean-up, acquired in linear mode. **(B)** Representative spectra of tryptic peptide profiles following digestion and MALDI-TOF analysis in reflectron mode.

Comparison of the 3 types of serum preparation protocol produced slightly different results. As seen in figures 3-6 A and B above. The protein profiles of serum sample after C_4 or C_{18} and combined C_4/C_{18} ZipTip clean-up show that there are more low mass range peaks (figure 3-6A) with higher intensity and better resolution for C_{18} ZipTip clean-up compared to the other 2 methods. Similarly for the peptides if we home in on the area between 800-1500 Da (figure 3-6B) there appear to be greater number of peaks and of higher intensity in the samples post C_{18} ZipTip compared to C_4 ZipTip and combined C_4/C_{18} ZipTip. Therefore; for this reason the protocol involving C_{18} ZipTip for both protein and peptide processing was chosen for future experiments.

3.2.5 Between run reproducibility of protein and peptide MALDI spectra profiles of QC samples

We assessed the performance and reproducibility of both instruments (MALDI-MS and Xcise robotic system) used in our study by including QC samples in each run, to ensure that the protocols that we used for sample preparation and data acquisition were robust. The QC samples were diluted and processed on Xcise as described before and all samples were analyzed in duplicate. The spectra were checked visually and cluster analysis was performed on the spectra that had the acceptance criteria. The spectra below are representative of an acceptable QC spectrum for proteins without any clean-up procedure, proteins after ZipTipping and the peptide profile. The following data from cluster analysis will explain the procedure better.

Proteins without C_{18} ZipTip clean-up: The protein profiles of QC samples spotted directly with SA on the plate were analyzed in linear mode on a daily basis and after checking the profiles they were smoothed and cluster analysis was performed. This was repeated on 8 different days and they were named run 1 to 8 and figure 3-6 shows one representative spectrum from each day. As shown, there is a good reproducibility in the spectra obtained from different runs on individual days and the cluster analysis (figure 3-6) show good integration of spectra from different runs. We selected mass range of 6-7 kDa and 9-10 kDa to be used for cluster analysis.

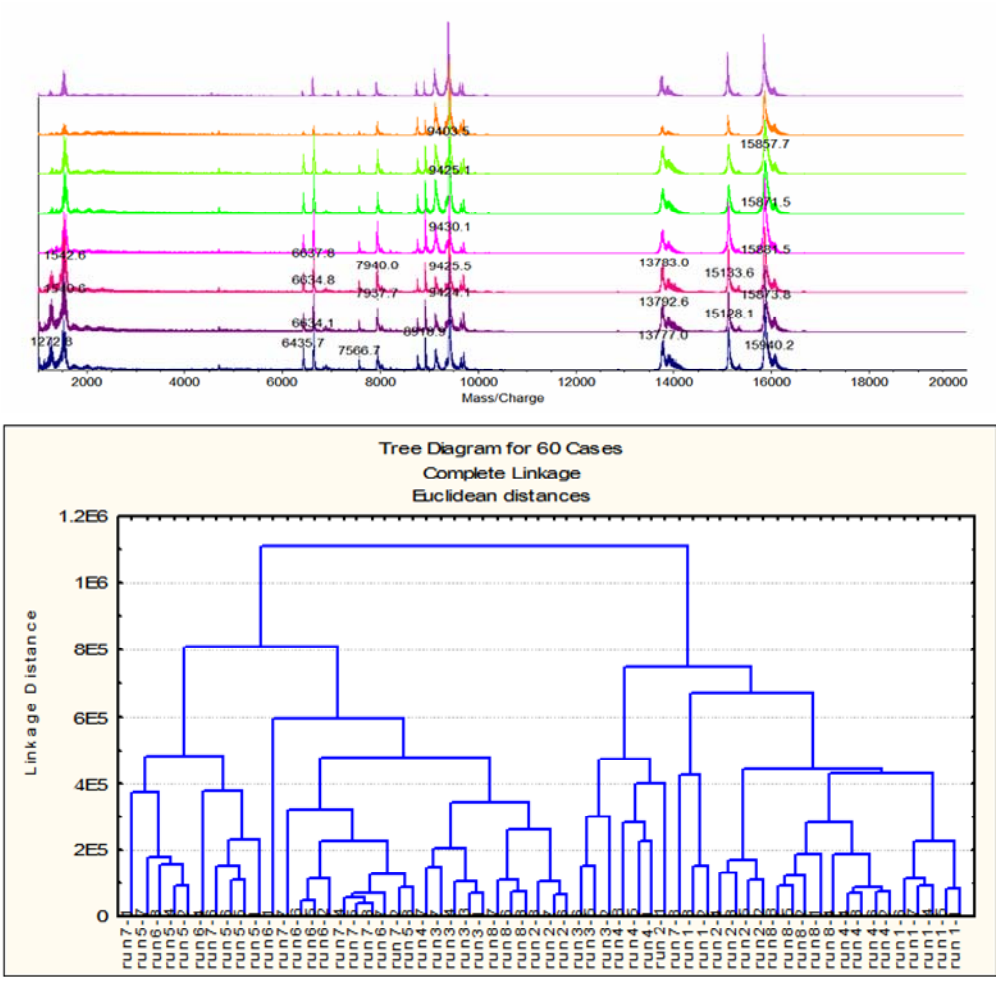


Figure 3-7. The figure on top represents the reproducibility of QC samples without C_{18} ZipTip clean-up, run in linear mode. Each spectrum is a representative of a QC sample run on different days. The bottom figure represent the cluster analysis for the same samples which shows that samples are similar to each other irrespective to day that they were run.

Protein data after C_{18} ZipTip clean-up: Figure 3-7 shows the 5 runs collected from QC samples on separate days and the result of the cluster analysis is also shown in figure 5. As can be seen from the spectra data reproducible peaks can be observed. The cluster analysis shows how similar each separate runs are together. In the cluster analysis all the runs are mixing together which demonstrates that these runs although have been carried out in separate days, there is not much variation between them and are similar to each other. For the cluster analysis of these samples also *m/z* of 6-8 and 9-10 kDa was used.

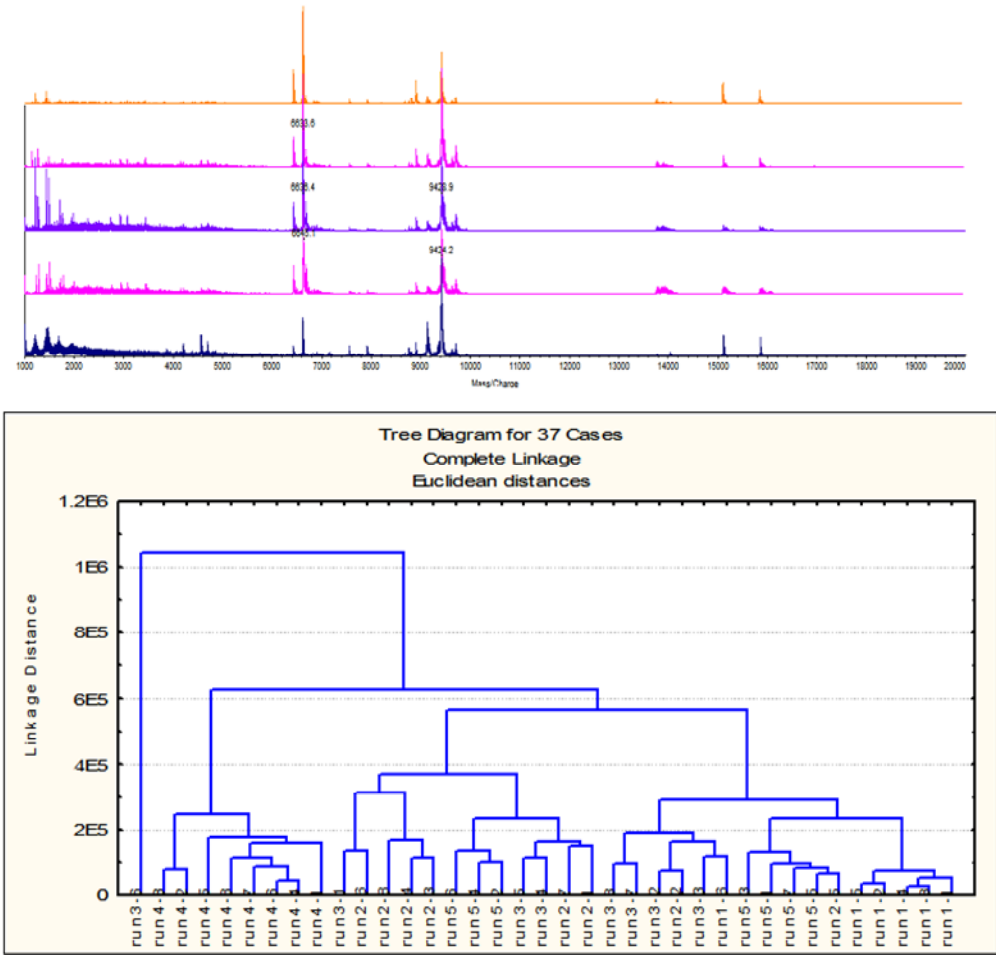


Figure 3-8. The top figure represents the reproducibility of QC samples after C_{18} ZipTip clean-up, run in linear mode. Each spectrum is a representative of a QC sample run on different days. The bottom figures represent the cluster analysis for the same samples which shows that samples are similar to each other irrespective to the day that they were run.

Peptide data after C_{18} fractionation: The peptides were run on 4 different days and a representative QC peptide profiles and the relevant cluster analysis is shown in figure 3-8. Our previous reproducibility studies showed that, CVs of peptide profiles were slightly higher that the protein profiles which can be due to the fact that these samples go through the cleaning procedure twice using C_{18} ZipTips and this may increase the variability between the samples. This also has an impact on the cluster analysis of the tryptic QC peptides. The QC peptides did not cluster as well as the two other datasets shown. The problem lies in the fact that although the selected spectra have similar patterns, a significant

difference in the intensity of the peaks were observed which might be correlated to the use of different laser power in different runs, experimental conditions, difference in quality of the matrix and its crystallisation.

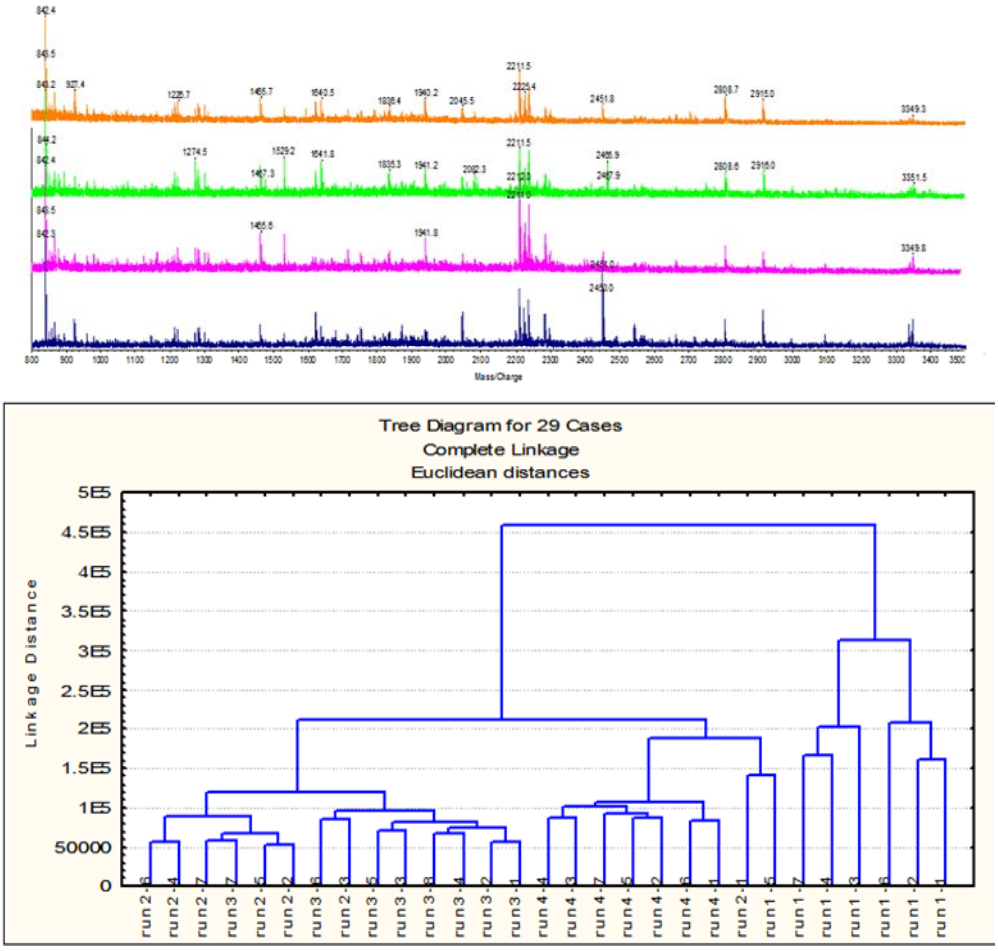


Figure 3-9. The top figure represents the reproducibility of QC samples after tryptic digest, run in reflectron mode. Each spectrum is a representative of a QC sample run on different days. The bottom figures represent the cluster analysis for the same samples which shows that there is a difference between the samples run in different days however the pattern of the spectra is similar with considerable differences in their intensities.

3.4 Discussion

The objective of the investigation described in this chapter was the development of a high throughput method for the analysis of LMW serum peptides. In order to achieve the sensitivity and reproducibility required, multi-stage clean-up procedures were developed for both native and tryptic peptides. This section details the various methods explored, their efficiency and optimisation.

Despite recent investigations on proteomic profiling of biological fluids and tissue aiming to identify candidate tumour markers, few have been approved and utilised in clinical practice (Taylor *et al.* 2006). The main concerns in the use of this technology are due to debates over the reproducibility and reliability of generated data (de Noo *et al.* 2005). Proteomic expression profiling by MALDI-MS is clearly high throughput with high mass accuracy (Dekker *et al*. 2005), but the stability of proteomic patterns over multiple experiments remains a question. To qualify the serum proteome profiling technique as a future diagnostic test, use of simple and robust techniques is more desirable than the use of more complex multi-step tools. The use of one step fractionation protocols such as various reversed-phase magnetic beads and pre-packed ZipTips have been proved to reduce serum complexity. Although number of studies has focused on standardisation of sample preparation protocols for MALDI, limited number of investigations has been reported regarding the reproducibility and rigorousness of the technique (de Noo *et al*. 2005 & Tiss *at al*. 2007). Within this study, we used a systematic automated sample preparation protocol (to minimise the impacts of preanalytical procedures on the final outcomes) for the high-throughput MALDI analysis of serum proteome and the technical reproducibility of this approach was examined at various levels of analysis. An established protocol utilising pre-packed C_{18} ZipTips (Matharoo-Ball *et al.* 2007) have been applied for serum fractionation using Xcise robotic sample handling system and we compared this with the use of C_4 and C_{18} ZipTips in conjunction for serum sample fractionation prior to MALDI analysis of serum samples.

Optimisation experiments showed Analysis show that utilisation of 1:20 dilution of serum in 0.1% TFA followed by C_{18} ZipTips fractionation prior to MALDI analysis obtained rich spectra containing high number lower molecular weight peaks and showed much details in spectra which then preceded for tryptic digestion and peptide profiling. Therefore this

protocol was adapted for further studies. In addition, use of C_{18} ZipTip clean-up on the 1:20 dilution of serum significantly reduced the concentration of albumin in serum.

Moreover, more detailed investigations regarding the robustness of the instrumentation used in the proposed protocol were carried out by assessing the reproducibility of final result which is the MALDI spectra. Batch-to-batch variation in the C_{18} ZipTips, the robotic sample processing system and the MALDI instrument are potential factors that affect reproducibility. The reproducibility of automated sample preparation and spotting and MALDI measurement in this study was assessed and the CVs for NIR mouse serum MALDI spectra without C_{18} ZipTip were 15-25% (mean CVs value of 18.75%), after C_{18} ZipTip clean-up was 13-44% (mean CVs value of 18.75%) and for tryptic peptides was 23-33% (mean CVs value of 24.53%). Reproducibility of (mean CVs value of 18.75%) MALDI investigated The CVs of the human serum MALDI-TOF-MS profiles using C_8 magnetic bead sample preparation in recent study (de Noo *et al* 2005) was reported to range between 17% and 26% which is similar to our observation despite the fact that they used different sample preparation method. Our results are promising and show acceptable levels of reproducibility from one run although further analyses on the reproducibility of the same samples over multiple experiments will be carried out in the future.

Chapter 4 – Serum and Tissue Proteomic Profiling in the Murine CT26 Colon Carcinoma Progression Model

4.1 Introduction

4.1.1 Tumour progression and survival

Cancer cells must obtain the ability to generate their own mitogenic signals in order to resist to exogenous growth inhibitory signals. This allows cancer cells to avoid apoptosis, infinitely proliferate, and acquire vasculature and to invade and metastasize in order to survive and progress (Eccles 2005, Rieger 2004). Normal cells move from a quiescent into a proliferative state, growth signals are transmitted into the cell via transmembrane receptors. Oncogenes are capable of producing growth signals, similar to signals produced by the normal cell which disrupts mechanisms involved in normal behavior of cells within the tissue (Hanahash and Weinberg 2000). During the course of carcinogenesis, growth factor receptors are over expressed and cancer cells to become hyper-responsive to levels of growth factors that would not normally trigger proliferation (Fedi *et al.* 1997). In normal cells, extracellular matrix (ECM) receptors (integrins) link cells to the ECM and transducer signals into the cytoplasm that influence activities such as cell motility, resistance to apoptosis and entrance into the cell cycle. However, cancer cells acquire the ability to alter the expression of the ECM receptors (integrins), favoring receptors that transmit growth factors into the cytoplasm from the ECM. Soluble growth factor inhibitors and immobilised inhibitors are molecules that are embedded in the ECM and associated with receptors on the surfaces of nearby cells acting as an anti-proliferative signal that functions to maintain normal cellular quiescence and tissue homeostasis.

Since proteins are involved in structural changes and catalysis in tissue, and because they are the most commonly targeted molecules for therapeutics, a proteome analysis was performed on the CT26 murine colorectal solid tumour. In this study, the whole tumour proteome at three different stages (7, 14 and 20 days post tumour implantation) was analysed using MALDI-MS and C_{18} ZipTip fractionation. The MS data was interpreted by ANN analysis to generate a panel of ions, most discriminatory between the different stages of CT26 tumour progression. The candidate biomarkers identified in early stages (both serum and tissue studies) may have the potential to serve as early diagnostic biomarkers and the late stage candidate biomarkers may provide a better understanding of the mechanisms involved in tumour progression or act as therapeutic putative targets.

Since proteins are involved in structural changes and catalysis in tissue, and because they are the most commonly targeted molecules for therapeutics, a proteome analysis was performed on the CT26 murine colorectal solid tumour. In this study, the whole tumour proteome at three different stages (7, 14 and 20 days post tumour implantation) was analysed using MALDI-MS and C_{18} ZipTip fractionation. The MS data was interpreted by ANN analysis to generate a panel of ions, most discriminatory between the different stages of CT26 tumour progression. The candidate biomarkers identified in early stages (both serum and tissue studies) may have the potential to serve as early diagnostic biomarkers and the late stage candidate biomarkers may provide a better understanding of the mechanisms involved in tumour progression or act as therapeutic putative targets.

Since proteins are involved in structural changes and catalysis in tissue, and because they are the most commonly targeted molecules for therapeutics, a proteome analysis was performed on the CT26 murine colorectal solid tumour. In this study, the whole tumour proteome at three different stages (7, 14 and 20 days post tumour implantation) was analysed using MALDI-MS and C_{18} ZipTip fractionation. The MS data was interpreted by ANN analysis to generate a panel of ions, most discriminatory between the different stages of CT26 tumour progression. The candidate biomarkers identified in early stages (both serum and tissue studies) may have the potential to serve as early diagnostic biomarkers and the late stage candidate biomarkers may provide a better understanding of the mechanisms involved in tumour progression or act as therapeutic putative targets.

Since proteins are involved in structural changes and catalysis in tissue, and because they are the most commonly targeted molecules for therapeutics, a proteome analysis was performed on the CT26 murine colorectal solid tumour. In this study, the whole tumour proteome at three different stages (7, 14 and 20 days post tumour implantation) was analysed using MALDI-MS and C_{18} ZipTip fractionation. The MS data was interpreted by ANN analysis to generate a panel of ions, most discriminatory between the different stages of CT26 tumour progression. The candidate biomarkers identified in early stages (both serum and tissue studies) may have the potential to serve as early diagnostic biomarkers and the late stage candidate biomarkers may provide a better understanding of the mechanisms involved in tumour progression or act as therapeutic putative targets.

Since proteins are involved in structural changes and catalysis in tissue, and because they are the most commonly targeted molecules for therapeutics, a proteome analysis was

performed on the CT26 murine colorectal solid tumour. In this study, the whole tumour proteome at three different stages (7, 14 and 20 days post tumour implantation) was analysed using MALDI-MS and C_{18} ZipTip fractionation. The MS data was interpreted by ANN analysis to generate a panel of ions, most discriminatory between the different stages of CT26 tumour progression. The candidate biomarkers identified in early stages (both serum and tissue studies) may have the potential to serve as early diagnostic biomarkers and the late stage candidate biomarkers may provide a better understanding of the mechanisms involved in tumour progression or act as therapeutic putative targets.

Since proteins are involved in structural changes and catalysis in tissue, and because they are the most commonly targeted molecules for therapeutics, a proteome analysis was performed on the CT26 murine colorectal solid tumour. In this study, the whole tumour proteome at three different stages (7, 14 and 20 days post tumour implantation) was analysed using MALDI-MS and C_{18} ZipTip fractionation. The MS data was interpreted by ANN analysis to generate a panel of ions, most discriminatory between the different stages of CT26 tumour progression. The candidate biomarkers identified in early stages (both serum and tissue studies) may have the potential to serve as early diagnostic biomarkers and the late stage candidate biomarkers may provide a better understanding of the mechanisms involved in tumour progression or act as therapeutic putative targets.

Since proteins are involved in structural changes and catalysis in tissue, and because they are the most commonly targeted molecules for therapeutics, a proteome analysis was performed on the CT26 murine colorectal solid tumour. In this study, the whole tumour proteome at three different stages (7, 14 and 20 days post tumour implantation) was analysed using MALDI-MS and C_{18} ZipTip fractionation. The MS data was interpreted by ANN analysis to generate a panel of ions, most discriminatory between the different stages of CT26 tumour progression. The candidate biomarkers identified in early stages (both serum and tissue studies) may have the potential to serve as early diagnostic biomarkers and the late stage candidate biomarkers may provide a better understanding of the mechanisms involved in tumour progression or act as therapeutic putative targets.

Since proteins are involved in structural changes and catalysis in tissue, and because they are the most commonly targeted molecules for therapeutics, a proteome analysis was performed on the CT26 murine colorectal solid tumour. In this study, the whole tumour proteome at three different stages (7, 14 and 20 days post tumour implantation) was analysed using MALDI-MS and C_{18} ZipTip fractionation. The MS data was interpreted by ANN analysis to generate a panel of ions, most discriminatory between the different stages of CT26 tumour progression. The candidate biomarkers identified in early stages (both serum and tissue studies) may have the potential to serve as early diagnostic biomarkers and the late stage candidate biomarkers may provide a better understanding of the mechanisms involved in tumour progression or act as therapeutic putative targets.

Since proteins are involved in structural changes and catalysis in tissue, and because they are the most commonly targeted molecules for therapeutics, a proteome analysis was performed on the CT26 murine colorectal solid tumour. In this study, the whole tumour proteome at three different stages (7, 14 and 20 days post tumour implantation) was

analysed using MALDI-MS and C_{18} ZipTip fractionation. The MS data was interpreted by ANN analysis to generate a panel of ions, most discriminatory between the different stages of CT26 tumour progression. The candidate biomarkers identified in early stages (both serum and tissue studies) may have the potential to serve as early diagnostic biomarkers and the late stage candidate biomarkers may provide a better understanding of the mechanisms involved in tumour progression or act as therapeutic putative targets.

Since proteins are involved in structural changes and catalysis in tissue, and because they are the most commonly targeted molecules for therapeutics, a proteome analysis was performed on the CT26 murine colorectal solid tumour. In this study, the whole tumour proteome at three different stages (7, 14 and 20 days post tumour implantation) was analysed using MALDI-MS and C_{18} ZipTip fractionation. The MS data was interpreted by ANN analysis to generate a panel of ions, most discriminatory between the different stages of CT26 tumour progression. The candidate biomarkers identified in early stages (both serum and tissue studies) may have the potential to serve as early diagnostic biomarkers and the late stage candidate biomarkers may provide a better understanding of the mechanisms involved in tumour progression or act as therapeutic putative targets.

4.1.2 Aims and objectives

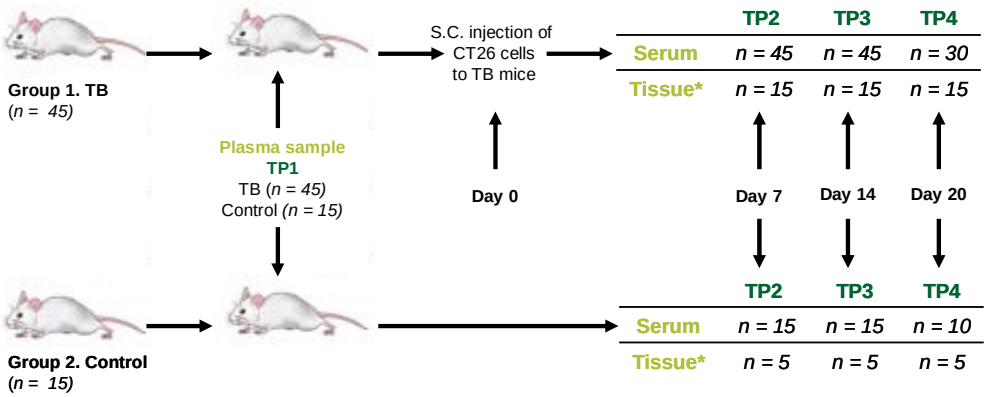
Detailed aims for this part of the study were:

- To collect serum samples from individual CT26 tumour-bearing mice in four different time points (*i.e.* naïve satus and 7, 14 and 21 days post tumour implantation).
- To obtain serum peptide profiles using MALDI-MS followed by ANN analysis to investigate whether is possible to obtain a panel of markers that are discriminatory between the different time-points of tumour progression.
- To identify the proteins associated with the candidate panel of biomarkers.

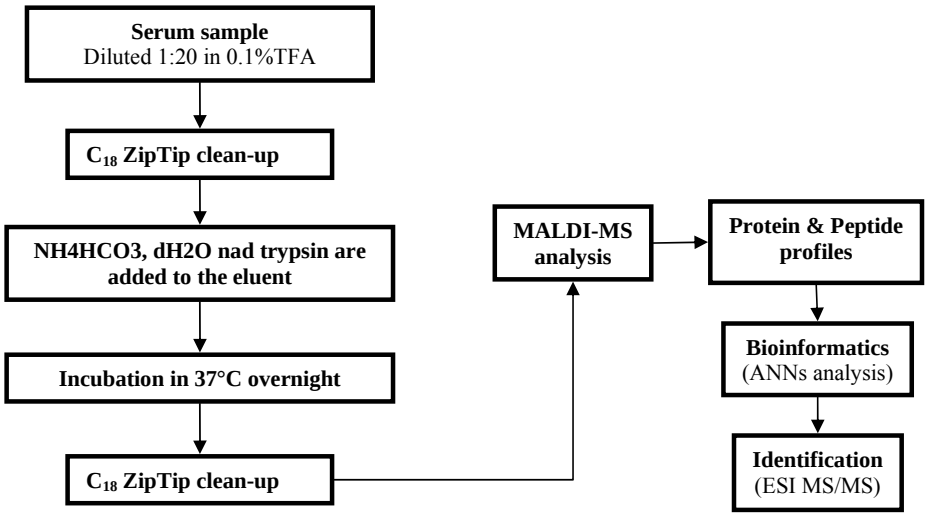
- Validate the proteomic based results using independent methods.

4.2 Summary of methods

A) CT26 progression murine model and sample collection



B) Sample preparation for proteomic analysis



4.3 Results

4.3.1 Peptide analysis of serum and tissue from the CT26 progression model by MALDI-MS

Serum and tissue samples from the CT26 progression model were subjected to MS analysis. Using MALDI-MS profiling, the serum and tumour samples were subjected to C_{18} ZipTip clean-up followed by tryptic digestion and a second C_{18} ZipTip chromatography; in addition, peptide profiles at different time-points were compared and analysed by ANNs. The proteomic analysis by MALDI is divided into serum and tissue analysis; the details are described in the following sections.

4.3.1.1 Time-Dependent Tryptic Peptide Analysis of Serum from the CT26 Progression Model.

Serum samples from 15 control (n = 15 TP0, n = 15 TP1, n = 10 TP2 and n = 5 TP3) and 45 tumour-bearer (n = 45 TP0, n = 45 TP1, n = 30 TP2 and n = 15 TP3) mice were subjected to MS analysis. The sample preparation procedure utilising C_{18} ZipTip was described previously and the tryptic peptide serum profiles were generated by MALDI-TOF MS (Bruker Daltonic, Bremen, Germany), operated in reflectron mode. Visual assessment of the MALDI spectra was necessary to determine the quality of the produced spectra and the reproducibility of sample preparation procedures. In addition, all the acquired MALDI spectra were visually assessed according to strict criteria (described in chapter 3) prior to statistical analysis for biomarker discovery to ensure the high quality of the data. Overall, there appeared to be visual spectral differences between the tryptic peptide proteomic profiles of serum sample from different time-points, and especially noticeable variation in the TP3 proteomic profiles in comparison to the other time-points. The methodology used here was relatively high-throughput and very high resolution spectra were obtained, therefore it is not possible to analyse multiple samples for each experiment by visual assessment of the spectra. Most of the spectral variation can be visualised when the spectra is “zoomed in” and here only the overall differences are shown. Moreover, the visual spectral differences highlighted in figure 4-1 show peaks that have higher abundance and therefore it is relatively easy to detect differences between the groups. In each spectrum, there were a number of peaks with lower intensities which may be significant

and possibly discriminatory between the groups, but it was not possible to assess these visually with any degree of accuracy. Therefore, these complex data sets from the CT26 progression model were subjected to a competitive algorithm (*i.e.* ANN) analysis to identify differences.

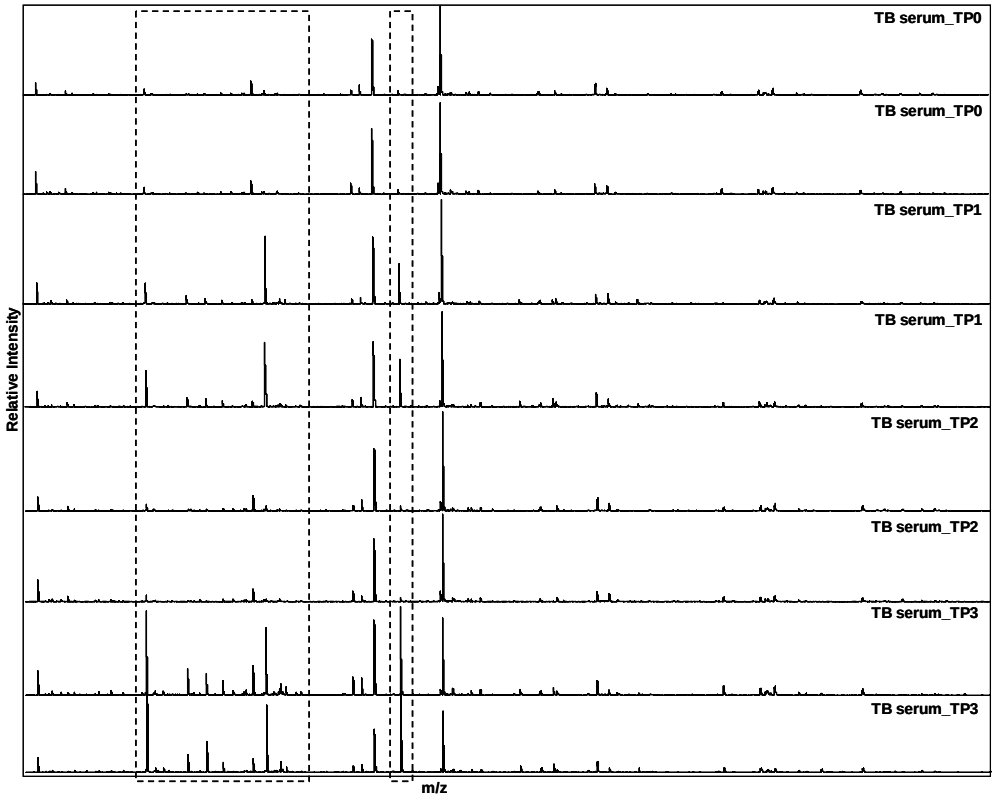


Figure 4-1. Figure showing tryptic peptide MALDI spectra obtained from serum samples from TP0, TP1, TP2 and TP3 tumour-bearer mice. The X axis represents the *m/z* ratio and the Y axis represents relative intensity. Representative MALDI-TOF mass spectra of TP0, TP1, TP2 and TP3 tumour-bearer mice serum tryptic digest mass range of 800-3000 Da, acquired in reflectron mode. Serum samples were spotted with CHCA on the MALDI steel plate and allowed to air dry. Peptide peaks in the regions highlighted in the figure appears to be up or down regulated when a four-way comparison was carried out on the different time-points.

Comment [A1]: Baharak show the data for the control serum as well for each time point to show that there are no changes in the serum proteome.

4.3.1.2 Time-dependent tryptic peptide analysis of tissue from the CT26 progression model.

Liver and spleen tissue samples from 15 controls (n = 5 TP1, n = 5 TP2 and n = 5 TP3 for each tissue) and CT26 tumour, liver and spleen tissue samples from 45 tumour-bearers (n = 15 TP1, n = 15 TP2, n = 15 TP3 for each tissue) were subjected to MS analysis. The tumour tissue as described was ground to a fine powder under liquid nitrogen and buffers added for protein extraction. The tissue homogenate was then subjected to C_{18} ZipTip separation as described previously and the tryptic peptide serum profiles were generated by MALDI-MS. Visual spectral differences between the different tissue types were observed and a representative tryptic peptide MALDI spectrum for each of the tissue samples obtained from the tumour-bearer animals is presented in figure 4-2. Interestingly, the tryptic peptide spectrum of liver tissue is very different compared to spleen and CT26 tumour tissue spectra. Spectra obtained from the tumour and spleen tissues show similarities specially in the region of 2200-4000 Da as highlighted in figure 4-2 however, the main visual differences between these tissues was observed in lower mass ranges (1300-2200 Da).

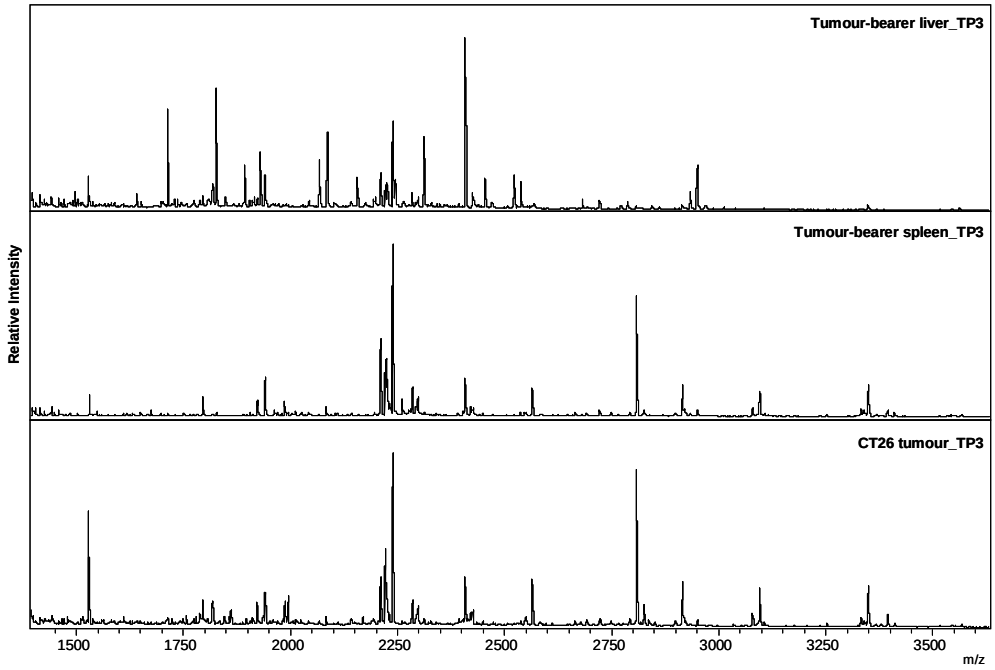


Figure 4-2. Figure showing tryptic peptide MALDI spectra obtained from liver, spleen and CT26 tumour samples from TP1, TP2 and TP3 of tumour-bearer mice. The X axis represents the *m/z* ratio and the Y axis represents relative intensity. Representative MALDI-TOF tryptic digest mass spectra of liver, spleen and CT26 tumour of tumour-bearer mice, mass range of 1400-3600 Da was acquired in reflectron mode. Tissue samples were spotted with CHCA on the MALDI steel plate and allowed to air dry. The spectra acquired from different tissue samples show significant different patterns. Intriguing observations is similarity in the above spectra between the spleen and tumour but are visually very different when compared to the liver.

4.3.2 Discovery of discriminatory serum biomarkers of tumour progression using ANN modeling.

To define changes in the serum proteome over a time course of tumour progression, we charted the plasma proteome of CT26 tumour-bearing mice at 4 different time points. In order to identify candidate biomarkers associating with tumour progression, initially we compared serum samples collected from 15 tumour-bearer mice at 4 different time points by generating various ANN models. Three models were developed to identify candidate biomarkers of tumour progression; the first model was based on serum TP0 and TP1 serum samples from the tumour-bearer mice, in which the TP0 serum was collected from 15 tumour free (naïve) animals and the TP1 serum samples were collected 7 days post tumour implantation. The second model was based on TP1 and TP2 (collected 14 days post tumour implantation) serum samples and the third model was TP2 and TP3 (collected 21 days post tumour implantation) serum samples.

The MALDI serum tryptic peptide profiles (mass range of 1000-3500 Da) of tumour-bearer mice from four different time points (n = 15 in each group) were subjected to ANN modeling using 12267 inputs for each sample (each input corresponds to intensity at a specific m/z value). Initially each input was used as a single input to train 50 models and in each model the total number of cases were randomly split into training (60%), testing (20%) and validation (20%) sets (random cross validation), after which the best performing model was selected for further analysis and all the remaining inputs were added sequentially to the first input to train the model. This procedure was continued until no further improvement in the model was observed.

The first stepwise analysis was carried out on the TP0 and TP1 serum samples from the tumour-bearer mice, using a total of 6 steps. The best performance was achieved at step 5 and the results are shown in table 4-1 which shows the median accuracies and mean squared errors of training, test and validation for each best performed model in each of the 5 steps. The highest accuracy of prediction that distinguished TP0 from TP1 sera was observed at step 5 (highlighted in red in table 4-1), with peaks with m/z values of 3181, 2659.8, 2600.8, 3187 and 1797 with an accuracy of 92%, sensitivity 93.75% and specificity of 100%.

Comment [A2]: Baharak what about the control ANNs was this done because if was not it needs doing to show there are no difference in the ptoteome

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training Error	Test error	Validation Error	Presence of ion
1	3181	67%	67%	67%	0.22	0.22	0.24	↑TP0
2	2659.8	78%	75%	75%	0.17	0.16	0.19	↑TP1
3	2600.8	92%	83%	83%	0.08	0.11	0.14	↑TP1
4	3187	100%	100%	83%	0.05	0.07	0.10	↑TP0
5	1797	100%	100%	**92%**	0.04	**0.05**	0.10	↑TP0
6	1474.8	100%	100%	83%	0.04	0.05	0.11	–

Table 4-1. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum tryptic peptide profiling of TP0 (naïve) and TP1 tumour-bearer mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in step 5, highlighted in red, with median accuracy of 92% and mean squared test error value of 0.05 and further addition of ions did not affect the accuracy of model.

Some of the MALDI tryptic peptide spectra differences between the TP0 and TP1 serum profiles for the top ions identified by ANN analysis are shown in figure 4-3. The *m/z* value of 1797 and 3181 show higher intensities in TP0 serum samples compare to TP1 serum samples whereas, ion with value of 2659.8 showed considerable higher intensity in TP1 samples.

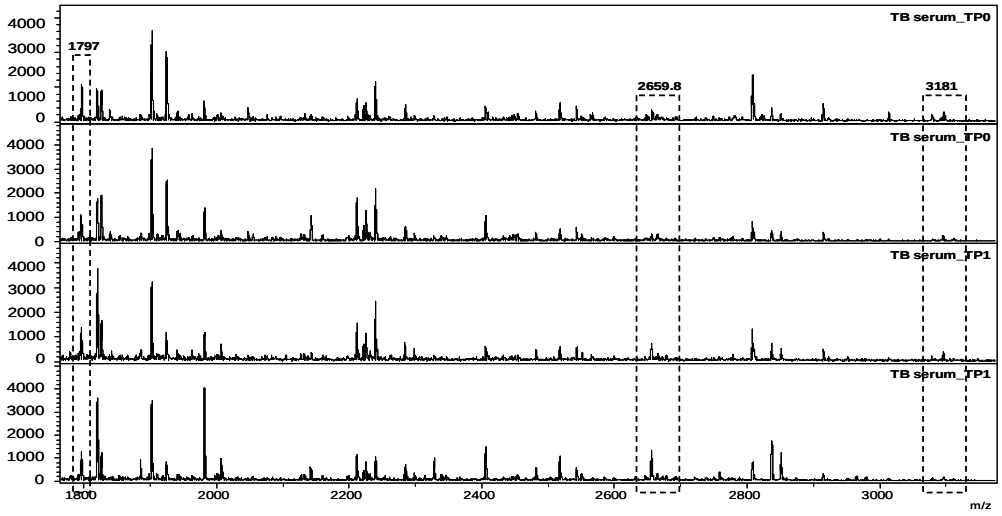


Figure 4-3. Figure showing visual spectral differences for some of the top discriminatory ions between TP0 and TP1 serum samples from tumour-bearer mice based on MALDI tryptic peptide profiles and ANNs analysis. The 1797 and 3181 discriminatory ions show higher intensity in TP0 samples whereas a peak with *m/z* of 2659.8 shows significant higher intensity in TP1 serum samples.

Furthermore, the potency of ANNs correctly classifying TP0 from TP1 based on the 5 ion model was further examined and is shown as a population chart in Figure 4-4. The ratios below 1.5 were assigned to the TP0 group and ratios above 1.5 were classified as TP1 (figure 4-4). Based on the 5 ion model, all the TP0 and TP1 samples were correctly classified, demonstrating that this model was able to stratify individual mice correctly.

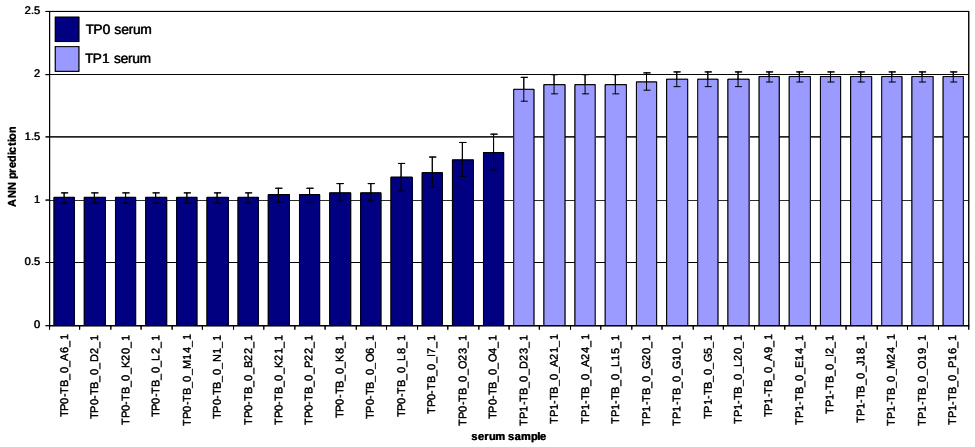


Figure 4-4. Predictive capability of ANNs to recognise tryptic peptide profiles based on 5 ion ANNs model. The dark bars indicate TP0 samples, and the light bars indicate TP1 samples. A predictive value below 1.5 indicates a TP10 sample, while a prediction greater than 1.5 indicates a TP1 sample.

The second step for identification of the candidate progression biomarkers was to compare TP1 and TP2 serum samples. The MALDI spectra of serum tryptic peptides of TP1 and TP2 mice were visually checked (according to the criteria previously described) for acceptance of the spectra. A total of 15 TP1 and 15 TP2 tryptic peptide profiles were accepted and used for bioinformatic analysis. Stepwise analysis was performed on these profiles to assess the ability of ANNs to classify tryptic peptide serum profiles of TP1 and TP2 mice serum. A total of 4 steps were carried out and the results of the stepwise analysis are illustrated in table 4-2. The best ANN prediction was achieved with 4 peptide peaks (*m/z* 2838.4, 3441.4, 2733.6 and 1197.8) that discriminated between the TP1 and TP2 serum samples with an accuracy of 98%, sensitivity of 86.7% and specificity of 100% and the mean test error of 0.07. Comparison between the discriminatory ions from the TP1/TP2 model and panel of ions from the TP0/TP1 model show no similarity between the panels of biomarkers associated with each of these models.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training Error	Test error	Validation Error	Presence of ion
1	2838.4	72%	67%	67%	0.20	0.20	0.23	↑TP1
2	3441.4	89%	100%	83%	0.12	0.11	0.16	↑TP2
3	2733.6	89%	100%	88%	0.10	0.09	0.14	↑TP2
4	1197.8	94%	100%	**98%**	0.06	**0.07**	0.12	↑TP1

Table 4-2. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum tryptic peptides of TP1 and TP2 mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in the fourth step, highlighted in red, with median accuracy of 98% and mean squared value of 0.07 and further addition of ions did not affect the accuracy of model.

The MALDI tryptic peptide spectra differences between the TP1 and TP2 profiles for the top 4 ions identified by ANN analysis are shown in figure 4-5. The *m/z* value of 2838.4, represents the first discriminatory ion (figure 4-5), is present in the TP1 and absent in the TP2 serum spectra. In addition, similar scenario is true for the ion with *m/z* value of 1197.8 which is presented in figure 4-5. The second discriminatory ion was *m/z* values of 3441.4 which show slightly higher intensity in the TP2 spectra.

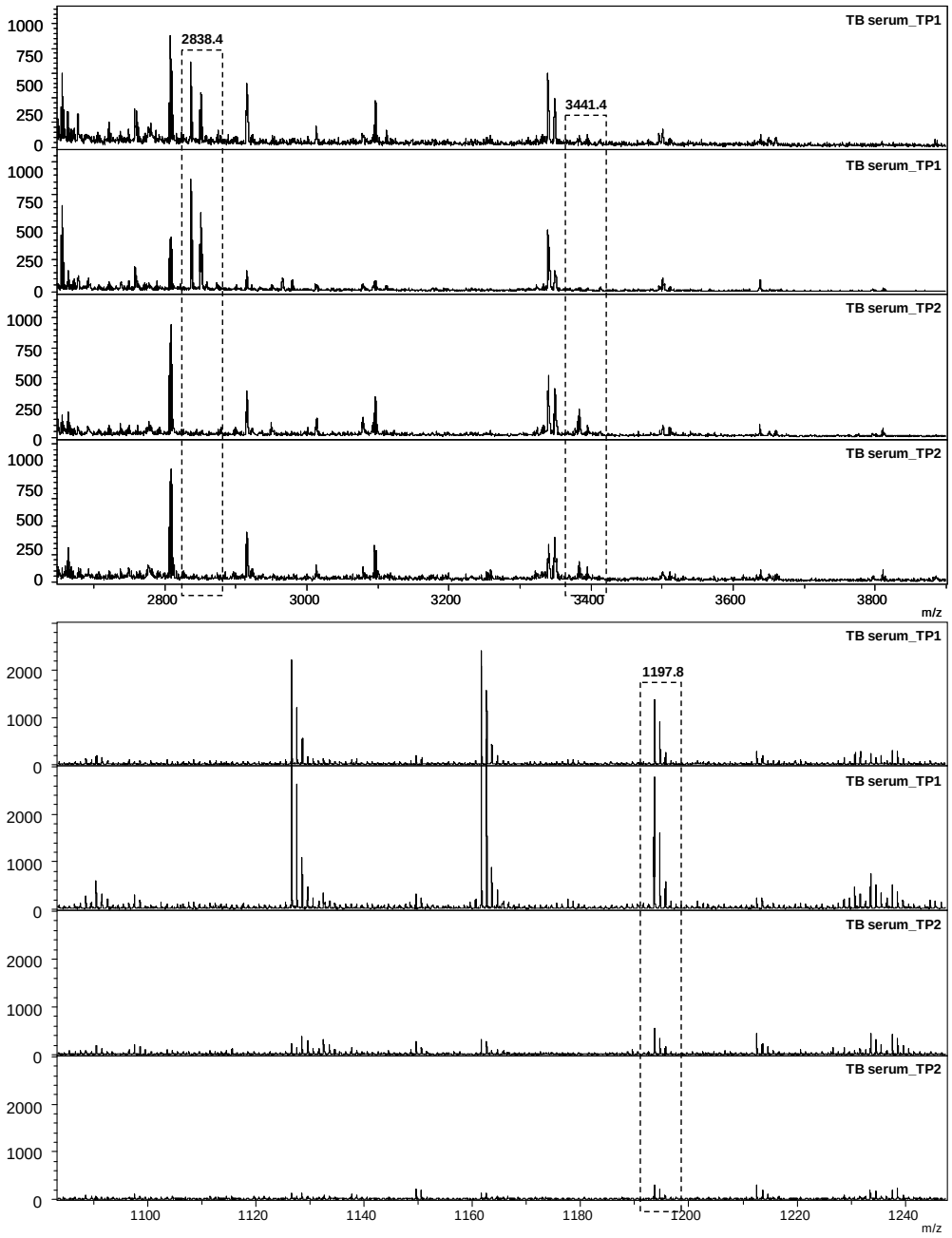


Figure 4-5. Figure showing representative visual spectral differences for the top discriminatory ion between TP1 and TP2 based on MALDI tryptic peptide profiles and ANNs analysis. The discriminatory ions for the TP1 and TP2 were 233.4 and 1197.3 which is present in the spectra corresponding to TP1 and absent in the spectra obtained from the TP2. The ion with *m/z* of 3441.4 was present with higher intensity in TP2 compare to TP1 spectra.

The ANNs ability to classify TP1 and TP2 was examined by generating a population chart and the results are shown in figure 4-6. ANN analysis classified 100% of the TP2 progression samples correctly whereas two TP1 serum samples were misclassified.

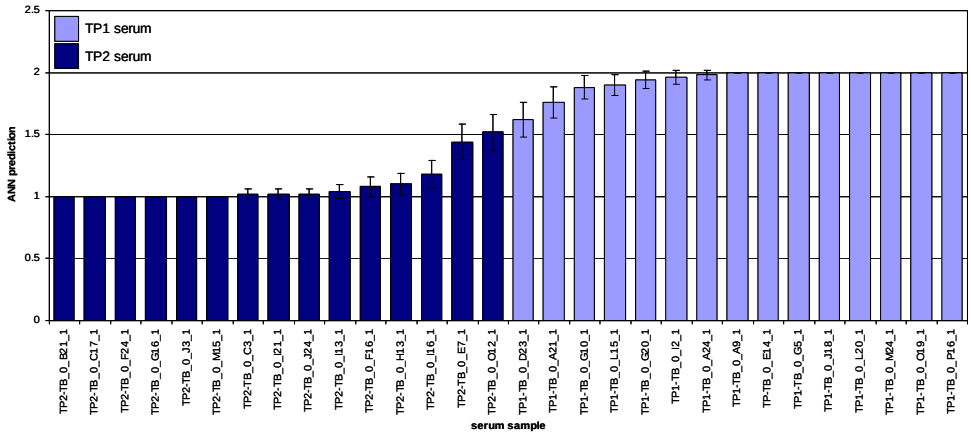


Figure 4-6. Predictive capability of ANNs to recognise tryptic peptide profiles based on 4 ion ANNs model. The dark bars indicate TP2 samples, and the light bars indicate TP1 samples. A predictive value below 1.5 indicates a TP2 sample, while a prediction greater than 1.5 indicates a TP1 sample.

The final step for identification of the candidate progression biomarkers was to compare TP2 and TP3 serum samples. The back propagation algorithm was used to identify patterns in the data and 18 samples were randomly selected for training, 6 for the test set and 6 for the blind dataset for each model, with 50 models run. A total of 9 steps was performed on the TP2/TP3 serum samples and the ANNs correctly classified the serum samples originating from either TP2 or TP3 tumour-bearer mice with an accuracy of 100% and a sensitivity and specificity of 98% and 99%, respectively and the mean test error of 0.04 which was reached at step 3 and table 4-3 shows the step wise results. Addition of ions after step 3 did not improve the models performance and resulted in over fitting of data.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training error	Test error	Validation error	Presence of ion
1	2837.4	72%	83%	67%	0.19	0.17	0.22	↑TP3
2	3125.8	83%	92%	83%	0.12	0.10	0.14	↑TP2
3	2916.4	100%	100%	**100%**	0.05	**0.04**	0.09	↑TP3
4	2985.8	100%	100%	92%	0.04	0.03	0.08	–
5	1413.2	100%	100%	100%	0.03	0.04	0.08	–
6	3011.2	100%	100%	100%	0.03	0.04	0.07	–
7	1594.6	100%	100%	100%	0.03	0.04	0.09	–
8	2769.4	100%	100%	100%	0.03	0.05	0.10	–
9	1890.4	100%	100%	83%	0.04	0.05	0.10	–

Table 4-3. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum tryptic peptides of TP2 and TP3 mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in step 3, highlighted in red, with median accuracy of 100% and mean squared test error value of 0.04 and further addition of ions did not affect the accuracy of model.

Figure 4-7 shows the response curve of the top three ions identified by ANN in the TP2/TP3 model. The response curve shows that ion *m/z* 2837.4 has higher intensity in TP3 samples (figure 4-7 A) whereas the *m/z* 3125.8 ion appears to have significantly higher intensity in the TP2 sample (figure 4-7 B). In addition, the *m/z* 2916.4 shows considerable higher intensity in the TP3 samples (figure 4-7 C). These differences can be seen in the spectra corresponding to TP2 and TP3 samples and representative spectra are shown in figure 4-7 D and the differences are highlighted.

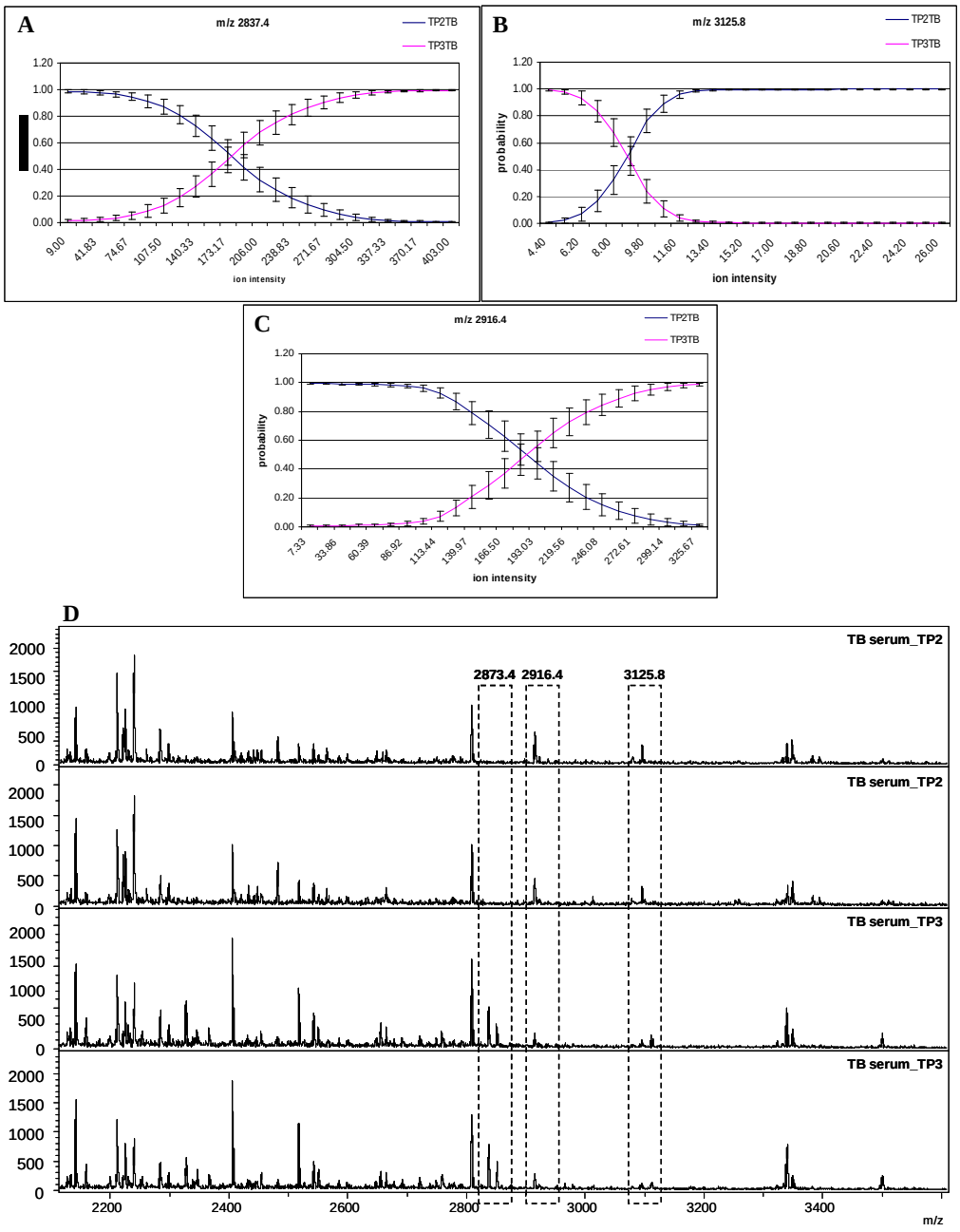


Figure 4-7. Response graphs generated singly for each of the 2 peptide biomarker ions used in the TP2/TP3 tumour model and the visual tryptic peptide MALDI spectral differences for the top three discriminatory ions. (A) Response graph for *m/z* 2837.4 indicating the relationship between intensity of the ion and the probability of predicted TP2 or TP3. **(B)** Response graph for *m/z* 3125.8 indicating the relationship between intensity of the ion and the probability of predicted TP2 or TP3. **(C)** Response graph for *m/z* 2916.4 indicating the relationship between intensity of the ion and the probability of predicted TP2 or TP3. **(D)** Representative MALDI spectra indicating discriminatory ions identified by the three ion ANN model of TP2/TP3.

Figure 4-8 shows the population distribution of the predicted outputs for all 30 samples. Samples originating from TP2 are highlighted in dark and TP3 are highlighted in lighter colour. The figure shows that all of the serum samples were correctly classified based on the 3 ANNs predicted ions.

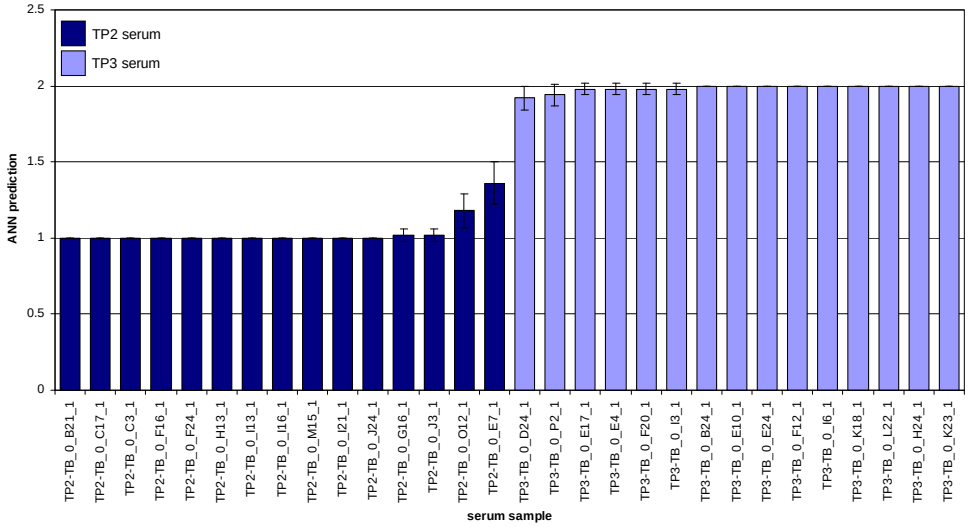


Figure 4-8. Predictive capability of ANNs to recognise tryptic peptide profiles based on 3 ion ANNs model. The dark bars indicate TP2 samples, and the light bars indicate TP3 samples. A predictive value below 1.5 indicates a TP2 sample, while a prediction greater than 1.5 indicates a TP3 sample.

As mentioned above at each time point of the experimental design also included naïve (control) mice to ensure that the markers were truly reflective of tumour burden and not the age of the animal. ANNs analysis was however, not possible due to small samples ($n = 5$) at each time-point.

4.3.3 Time-course analysis of CT26 tumour proteome changes during tumour progression using ANN modeling.

The goal of the study was to identify candidate tumour, spleen and liver tissue biomarkers that may associate with CT26 tumour progression. An experimental Balb/c murine progression model was established by CT26 tumour implantation and this was used to investigate tissue proteome changes in tumour-bearing mice over three different time-points. Tissue specimens were collected from 15 mice at three different time points: 7 (TP1), 14 (TP2) and 21 (TP3) days post tumour implantation. Tumour tissues were

homogenised and total protein extraction was performed as described in chapter 2. To identify candidate tumour tissue biomarkers, tumour proteome extracts were analysed using MALDI-MS in conjunction with ANN analysis. Following tumour protein extraction, the total protein concentrations for all tissues were equalised to 1mg/ml in 0.1%TFA. Tissue samples were passed through C_{18} ZipTip columns to remove high mass proteins. Tryptic digestion was carried overnight, followed by a second C_{18} ZipTip clean-up procedure. The final eluate was spotted on to a MALDI a plate with 10 mg/ml CHCA and tryptic peptide profiles were generated by MALDI-MS, operated in reflectron mode. The tryptic peptide profiles were visually checked and spectra with poor quality were eliminated from the study; the remaining spectra were subjected to proceed for further bioinformatic analysis, using ANN modeling.

The MALDI-MS tumour tissue tryptic peptide profiles (mass range of 1-4 kDa) of 9 TP1, 15 TP2 and 15 TP3 samples were used for ANN modeling. A stepwise approach was utilised to identify the minimum number of ions capable of assigning samples correctly to their respective groups. Prior to the ANN analysis, the data was smoothed, aligned and baseline corrected as described in materials and methods (chapter 2). As each sample was spotted on the MALDI plate in duplicate, one spot for each sample was randomly selected to be used in ANNs modeling. Three different ANN modes were generated to identify candidate biomarkers which are discriminatory of tumours at different time-points of tumour progression; the first model was a TP1 tissues versus TP2 tissues, the second model was TP2 tissues versus TP3 tissues and finally TP1 tissues versus TP3 tissues model were generated. In order to generate ANN models from MALDI tryptic peptide profiles of various time-point tumour samples, the stepwise approach was utilised which involves training a number of models using 14841 inputs for each sample (each input corresponds to intensity at a specific *m/z* value). Initially each input was used as single input to train 50 models and for each model, the total number of cases were randomly split into training (60%), test (20%) and validation (20%) sets; this allows random cross validation analysis to be performed. After training 50 models, the best performing model was selected for further analysis and all the remaining inputs were added sequentially to the first input to train the model. This procedure was continued until no further improvement in the model was observed.

The first model was TP1/TP2 in which the tryptic peptide profiles from 7 days (n = 9) and 14 days (n = 15) CT26 tumours were used to generate the ANN model and a total of 4 steps were carried out. The results of the 4 steps are shown in table 4-4 that presents the median accuracies and mean squared errors of training, test and validation for each best performed model in each of the 4 steps. The highest accuracy that predicted TP1 tumours from TP2 tumours was generated in step 2 (highlighted in red in table 4-4, using two peaks with *m/z* values of 1517 and 3548.8 with an accuracy of 80%, a sensitivity of 66.7% and specificity of 86.7%. The addition of further ions did not improve the prediction of the model and therefore the first 2 ions were considered biomarkers that can best predict TP1 tumours from TP2 tumours based on MALDI tryptic peptide profiles.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training Error	Test error	Validation Error	Presence of ion
1	1517	64%	80%	60%	0.21	0.20	0.26	↑TP2
2	3548.8	86%	100%	**80%**	0.10	**0.08**	0.17	↑TP1
3	3985.6	100%	100%	80%	0.05	0.06	0.11	–
4	3607.2	100%	100%	80%	0.05	0.06	0.11	–

Table 4-4. The table represents the data obtained from the stepwise analysis of MALDI data generated from CT26 tumour tissue tryptic peptide profiling of 7 (TP1) and 14 (TP2) day tumours. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in step 2, highlighted in red, with median accuracy of 80% and mean squared value of 0.08 and further addition of ions did not affect the accuracy of model.

Response curves were generated for each of the ions in the 2 ion model by presenting the model with varying values within the range of those found in the data and maintaining the other ions as the mean value. In this way the responses of ions were investigated singly for *m/z* values of 1517 and 3548.8 (figures 4-9 A & B). These analyses indicated that there is an increase in the intensity of ion *m/z* 1517 in the tumour tissue samples of TP2 in comparison to TP1 samples. As the intensity of this ion increases, the tumour tissue is more likely to belong to TP2 and may be a candidate biomarker, associated with tumour progression between day 7 to 14 post tumour implantation. In contrast to this, the ion with *m/z* value of 3548.8 shows a higher intensity in TP1 tumours and a decrease in TP2 tumour samples. Intensity differences associated with the top 2 discriminatory ion for TP1 and TP2 tumour tissues can be visualised in the MALDI tryptic peptide profiles and are shown in figure 4-9 C and D.

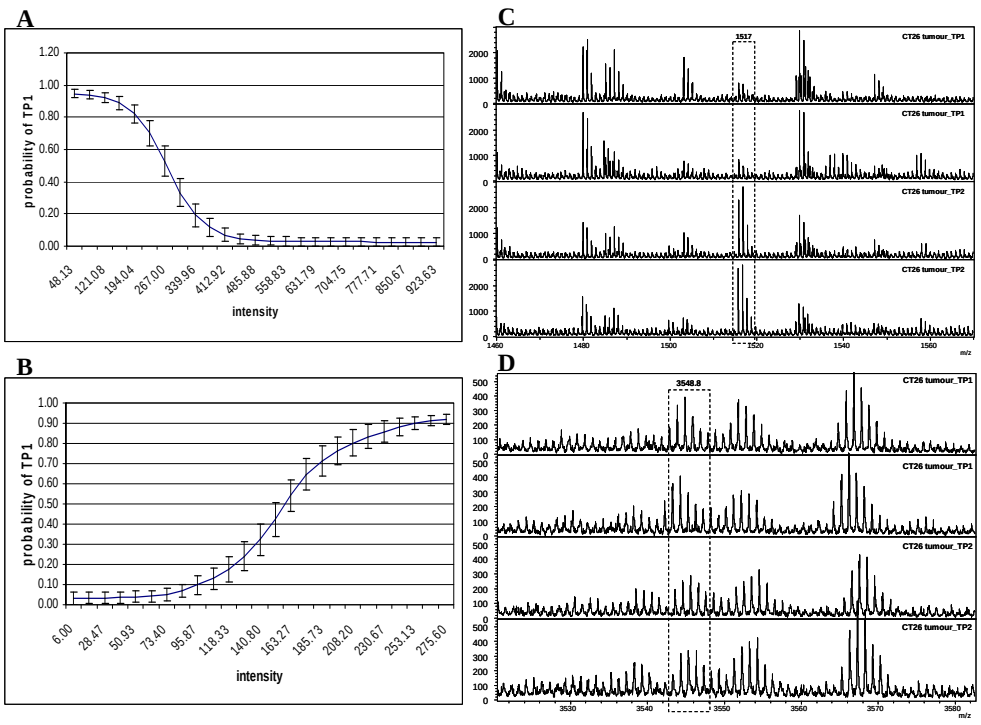


Figure 4-9. Response graphs generated singly for each of the 2 peptide biomarker ions used in the TP1/TP2 tumour model and the visual tryptic peptide MALDI spectral differences for the top two discriminatory ions. (A) Response graph for *m/z* 1517 indicating the relationship between intensity of the ion and the probability of predicted TP1. Increase in the intensity of this ion has a negative influence in the probability of samples to be a TP1 tumour. **(B)** Response graph for *m/z* 3548.8 indicating the relationship between intensity of the ion and the probability of predicted TP1. Increase in the intensity of this ion has a positive influence in the probability of samples to be a TP1 tumour. **(C)** Representative MALDI spectra indicating the significant increase in the intensity of *m/z* 1517 in the TP2 tumour tissue samples. **(D)** MALDI spectra indicating higher intensity of *m/z* 3548.8 in the TP1 tumour tissue samples.

The potency of ANNs to correctly predict tryptic peptide profiles based on 2 predictive ions obtained from TP1 and TP2 tumour tissue modeling was further examined by generating a population chart. The chart represents cases classified based on the 2 ion ANNs modeling of tryptic peptide MALDI profiles and the results are illustrated in figure 4-10. The ratios below 1.5 were assigned as TP1 tumours whilst a ratio above 1.5 was used to classify TP2 tumours. Based on the panel of biomarkers, three TP1 and one TP2 tumours were misclassified.

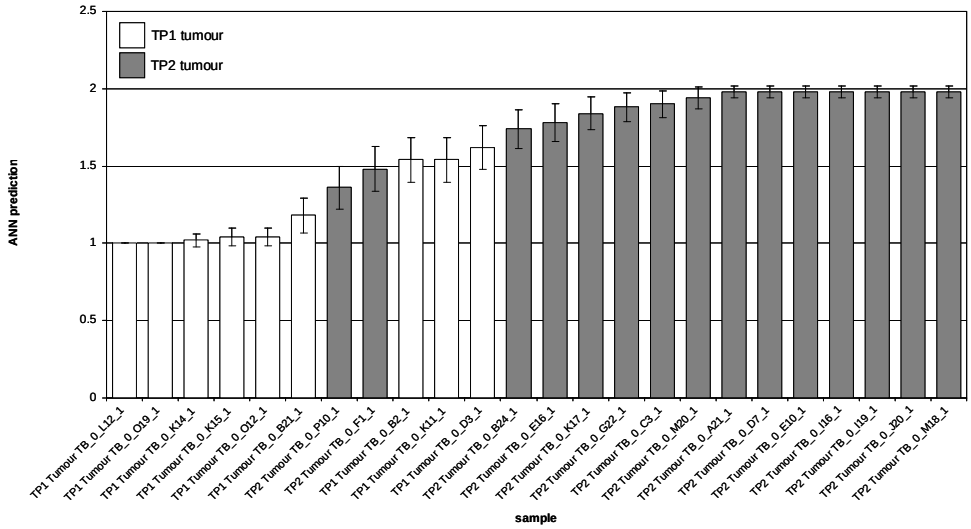


Figure 4-10. Predictive capability of ANNs to recognise tryptic peptide profiles based on a 2 ion ANN model. The white bars indicate TP1 tumours, and the gray bars indicate TP2 tumours. A predictive value below 1.5 indicates a TP1 tumour, while a prediction greater than 1.5 indicates a TP2 tumour.

Once the discriminatory candidate biomarkers between tumour tissue from TP1 and TP2 had been identified, the second model for TP2/TP3 was generated in which the tryptic peptide profiles from 14 days (n = 15) and 20 days (n = 15) CT26 tumours were used to train the ANN models and a total of 3 steps were carried out as described previously. The results of the 3 steps are shown in table 4-6 that presents the median accuracies and mean squared errors of training, test and validation for each best performing model in each of the 3 steps. In step 2, the highest accuracy for the prediction of TP2 tumours form TP3 tumours was reached (highlighted in red in table 4-5). The top two peaks, identified by ANN analysis were *m/z* 1333.6 and 1013.6 which were able to discriminate TP2 and TP3 tumour tissues with an accuracy of 83%, sensitivity of 76.7% and specificity of 87.5%. The addition of further ions did not improve the prediction of the model and therefore the first 2 ions were considered biomarkers that can best predict TP2 tumours from TP3 tumours based on MALDI tryptic peptide profiles. None of these markers were common with the top discriminatory ions from the previously described TP1/TP2 model which indicates that during tumour progression, different biomarkers may be more dominant and stronger and therefore better predictors of tumour progress.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training Error	Test error	Validation Error	Presence of ion
1	1333.6	74%	83%	67%	0.18	0.15	0.23	↑TP3
2	1013.6	84%	83%	**83%**	0.13	**0.13**	0.19	↑TP2
3	2727.2	84%	83%	75%	0.12	0.11	0.18	–

Table 4-5. The table represents the data obtained from the stepwise analysis of MALDI data generated from CT26 tumour tissue tryptic peptide profiling of 14 (TP2) and 21 (TP3) day tumours. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in step 2, highlighted in red, with median accuracy of 83% and mean squared value of 0.13 and further addition of ions did not affect the accuracy of model.

Response curves were generated for each of the ions in the 2 ion of TP2/TP3 model and are illustrated in figure 4-11. These analyses indicated that as the intensity of the ion *m/z* 1333.6 increases; the probability of a sample associated with TP3 is higher whereas, the intensity of the ion with mass value of 1013.6 is higher in the TP2 tumours MALDI tryptic peptide profiles.

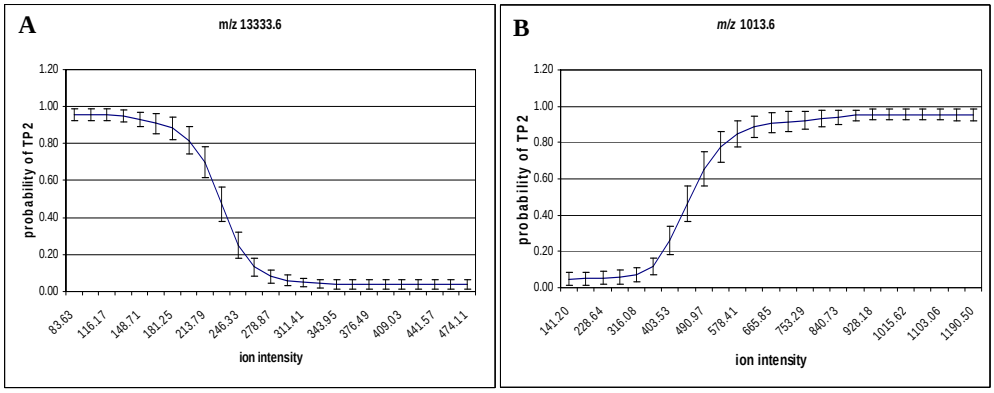


Figure 4-11. Response graphs generated singly for each of the 2 peptide biomarker ions used in the model indicating the relationship between intensity for an ion of a given *m/z* value and the probability of predicted TP2 tumour tissue. (A) The single ion *m/z* 1333.6 had a weak influence at lower intensities and a stronger influence at higher intensities with respect to TP2 tumours. **(B)** The single ion *m/z* 1013.6 had a strong positive influence at lower intensities and a weaker influence at higher intensities with respect to TP2 tumours.

The discriminate ability of the model was tested on individual TP2 and TP3 tumour tissue samples based on their associated tryptic peptide MALDI profiles with the top 2 predicted ions was examined by generating the population charts, presented in figure 4-12. The ratio below 1.5 is representative of a TP2 tumour and ratio above 1.5 is representative of a TP3 tumour. Based on the 2 ion model, two of the TP2 and TP3 tumours were misclassified. It

appear that there might be slight differences between the TP2 and TP3 tryptic peptide MALDI profiles of this study, but for same samples was no distinct separation between the TP2 and TP3 sample groups in the population charts.

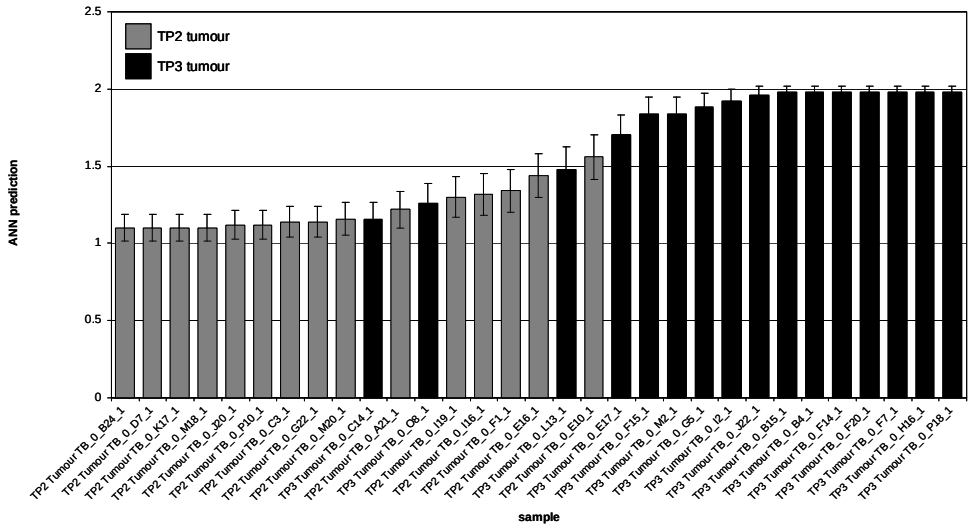


Figure 4-12. Predictive capability of ANNs to recognise tryptic peptide profiles based on a 2 ion ANNs model. The gray bars indicate TP2 tumours, and the black bars indicate TP3 tumours. A predictive value below 1.5 indicates a TP2 tumour, while a prediction greater than 1.5 indicates a TP3 tumour.

The final ANN modeling attempt was to compare the early stage CT26 tumour tissues and the late stage tumours. The final model was generated for the TP1/TP3 tumour tissues in which the tryptic peptide profiles from day 7 (n = 9) and day 20 (n = 15) CT26 tumours were used to train the ANN models. A total of 10 steps were carried out (as described previously) and the results are demonstrated in table 4-5. A single ion model had a median predictive performance of 100% and mean squared value of 0.06 for the test error. The top discriminatory peaks, identified by ANN analysis was *m/z* 3907.4 and no further improvement in the model was achieved by the further addition of ions.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training Error	Test error	Validation Error	Presence of ion
1	3907.4	93%	100%	**100%**	0.07	**0.06**	0.08	↑TP3
2	1334.8	100%	100%	100%	0.03	0.03	0.07	–
3	1613.8	100%	100%	100%	0.03	0.03	0.06	–
4	1327.8	100%	100%	100%	0.03	0.03	0.05	–
5	1161	100%	100%	100%	0.04	0.03	0.09	–
6	1030.8	100%	100%	100%	0.02	0.03	0.07	–
7	1132.4	100%	100%	100%	0.03	0.04	0.07	–
8	1516	100%	100%	100%	0.03	0.04	0.09	–
9	1055	100%	100%	100%	0.03	0.04	0.11	–
10	2971.8	100%	100%	100%	0.04	0.04	0.08	–

Table 4-6. The table represents the data obtained from the stepwise analysis of MALDI data generated from CT26 tumour tissue tryptic peptide profiling of 7 (TP1) and 21 (TP3) day tumours. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in step 1, highlighted in red, with median accuracy of 100% and mean squared value of 0.06 and further addition of ions did not affect the accuracy of model.

The response curves for the 1 ion model was generated and this showed that an increase in the intensity of ion *m/z* 3907.4 increases, the probability of tumour to be a TP3 is higher (figure 4-13 A). The visual differences for the ion *m/z* 3907.4 is presented in figure 4-10 B showing that this peak is present in the TP3 tumour tissues and absent in the TP1 tumour tissues.

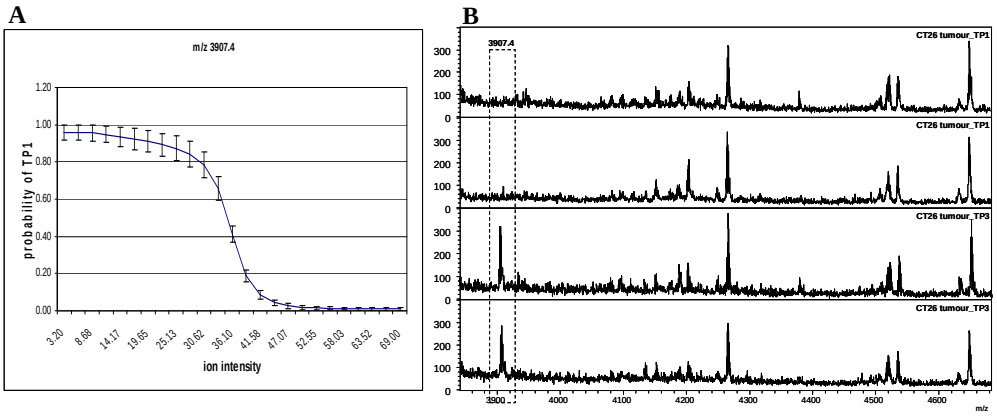


Figure 4-13. R**esponse graphs generated singly for the 1 peptide biomarker ions used in the TP1/TP3 tumour model and the visual tryptic peptide MALDI spectral differences for the top two discriminatory ions. (A)** Response graph for *m/z* 3907.4 indicating the relationship between intensity of the ion and the probability of predicted TP1. Increase in the intensity of this ion has a negative influence in the probability of samples to be a TP1 tumour. **(B)** MALDI spectra indicating the **s**ignificant increase in the intensity of *m/z* 3907.4 in the TP3 tumour tissue samples.

A population chart for discrimination between TP1 and TP3 tumour tissue samples shows the correct classification of all cases (figure 4-14). Comparison between the two extreme cases of CT26 tumour time-points (*i.e.* TP1 and TP2) shows better classification of the individuals by ANN analysis and a distinction between the two groups.

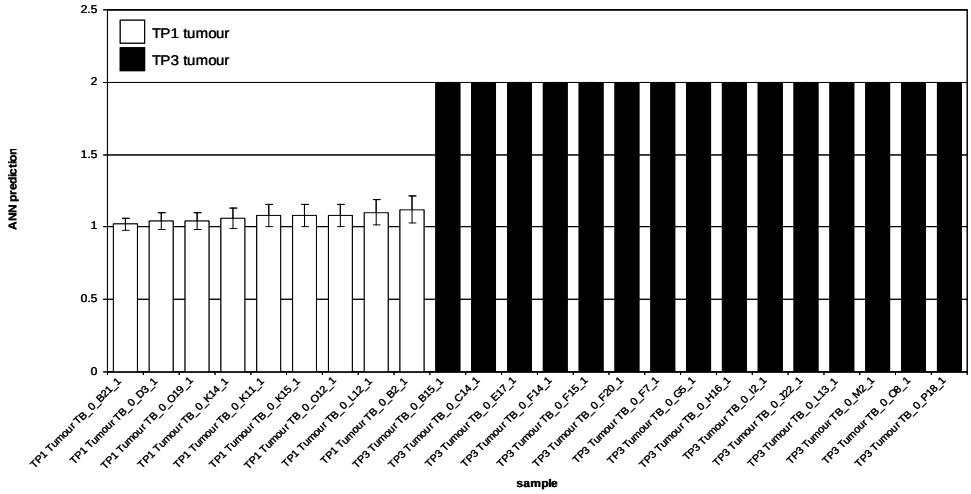


Figure 4-14. Predictive capability of ANNs to recognise tryptic peptide profiles based on a 1 ion ANN model. The white bars indicate TP1 tumours, and the black bars indicate TP3 tumours. A predictive value below 1.5 indicates a TP1 tumour, while a prediction greater than 1.5 indicates a TP3 tumour.

During the time course of tumour progression, spleen and liver tissues were also collected from the tumour-bearer and control mice. The samples were homogenised and protein extraction was carried out. Tryptic peptide profiles were obtained using MALDI-MS analysis. However, ANN analysis and modeling of these samples would not be presented in this thesis due to the limited time. Visual spectral differences can be seen when different time-points are compared although these differences may not be of clinical importance and sophisticated computational algorithms such as ANNs may reveal markers that are associated with tumour progression in liver and spleen.

4.4 Discussion

The tumour microenvironment consists of tumour and non-tumour cells, extracellular matrix components (*e.g.* collagens and elastins) and soluble components such as cytokines, chemokines and growth factors. Quantity of these components is likely to change as a result of the presence of tumour. The presence of several different biological products in the tumour microenvironment facilitates tumour progression. In addition, the non-tumour cells (*e.g.* fibroblasts, endothelial cells and immune cells), present in the tumour microenvironment, are involved in tumour progression through tumour-stroma interactions. Current therapeutics mainly target interactions and pathways involved in tumour progression; however, studying the tumour as it progresses *in vivo*, and using high throughput "omic" technologies may reveal new additional targets for therapy and early diagnosis of cancer.

High-throughput approaches that enable the profiling of large numbers of samples has allowed the complexity of cancer to be investigated. It is widely thought that a combination of markers is likely to be more sensitive and specific for patient diagnosis and predicting prognosis and response to treatment than a single biomarker (Bertucci *et al.* 2006). Application of the novel peptide methodology developed by Matharoo-Ball and colleagues (Matharoo-Ball *et al.* 2007), enabled visual differences in the region 800-3500 Da of the peptide profiles in melanoma samples compared with controls. This methodology proved to have good reproducibility and potential to be adapted for proteomic analysis of mouse serum samples as shown in chapter 3. The present study was a proteome analysis of serum and tumour tissue samples collected from 4 stages in the progression of the CT26 mouse colorectal model. The 4 stage investigation of the serum proteome change in this study is unique as the 4 different time-point samples were collected from the same individuals and 15 tumour-bearer animals at each time-point were used in this study. Our results show that ANN algorithms in conjunction with MALDI peptide profiling can discriminate between the serum protein expression patters from different time points of tumour progression. Initially MALDI tryptic peptide profiles of serum samples from TP0 (naïve mice) and TP1 (7 days post tumour implantation) were compared. Here, ANNs analysis identified 5 discriminatory ions capable of classifying 92% of the TP0 from TP1 and these peptide biomarker ions may indicate early biomarkers for cancer in the CT26 colorectal mouse model as the naïve mice were compared to an early tumour initiation where the size of the

tumour was an average size of 0.2 mm. Possible markers of CT26 tumour progression were investigated via ANN modelling of TP1 *versus* TP2 (serum samples collected 7 and 14 days post tumour implantation respectively) and TP2 versus TP3 (serum samples collected 14 and 21 days post tumour implantation) serum samples. The ANN analysis of TP1/TP2 and TP2/TP3 revealed a 4 ion model with accuracy of 98% and a 3 ion model with accuracy of 100% respectively. These biomarkers may serve as potential progression serum biomarkers that can be used as therapeutic targets. Comparison between the panel of biomarkers from the TP0/TP1, TP1/TP2 and TP2/TP3 models did not show any common ions except the ion with *m/z* of 2837.4±1 was common between the TP1/TP2 and TP2/TP3 panel of biomarkers. In both models this ion was the first discriminatory ion between the associated groups and could predict 67% of the samples correctly. This marker could be a late stage tumour progression marker as it is not of important in the early stage model (TP0/TP1). As the other panel of biomarkers appear to be unique to each model, it can be concluded that these panel of biomarkers may be specific to each time point, although further investigations and identification of proteins associated to each ion is required to validate these invesitgations. The significance of these markers is beyond the limitation of the present study as they have not been identified. The technology illustrates the necessity of peptide sequencing and the advantages of a combination of MS techniques to reliably and unambiguously establish the identity of predictive biomarkers in cancer studies. In addition, attempts to identify some of the proteins corresponding to the discriminatory ions using MALDI MS/MS failed to identify the proteins with high level of confidence. This seems to be due to the complexity of samples. In some cases, there are several overlapping peptides around the ion of interest with close masses which can undergo fragmentation with the ion of interest during the MS/MS procedure and presence of their fragments in the MS/MS spectra of ion of interest will interfere in the database search and obtaining reliable identities. Further fractionation of samples utilising techniques such as LC prior to MALDI MS/MS analysis would significantly reduce the complexity of samples and separates the overlapping proteins. Limited numbers of studies have investigated serum proteome change in mouse tumour models in a time-dependent manner during the progression of tumour. Sandoval *et al.* used mouse model of human high-risk neuroblastoma and collected serum samples at three time-points (2, 4 and 6 weeks post tumour implantation) during the tumour progression. They used 2D gel electrophoresis and identified 6 proteins (5 up-regulated and

1 down-regulated) that were uniquely expressed in the tumour-bearing mice compared to control individuals. Furthermore, the expression of the up-regulated proteins (α_1-acid glycoprotein, α_1-antitrypsin, α_2-microglobin, serum amyloid P and serum amyloid A) shown to increase as the tumour progressed. All the 5 up-regulated proteins represent acute-phase proteins that represent common serum proteins. The advantage of our study to the use of 2D gel electrophoresis based studies is firstly the ability of analysing a high number of samples for proteomic analysis (n = 15 CT26 tumour-bearing mice at each time point) which leads to more generalised results that are more representative of the different groups (*i.e.* serum samples from different time-points). In addition, the use of a simple, high throughput, automated sample preparation reduced the bias due to sample handling in this study. However, further studies could reveal the identity of proteins that may allow new biomarkers to be identified or the biological pathways involved in the disease progression to be further understood and may be applicable to human cancers.

Utilising the same strategy for analysing the serum samples, CT26 tumour tissue samples were also used to identify tissue biomarkers of progression. The advantage of analysing tissue proteins over serum proteome is that any identified candidate biomarkers from the tissue analysis may have the potential to be used as a direct tumour target for therapy. However, biomarkers discovery by serum proteomic analysis reveals markers that are secreted from either tumour or other tissues (*i.e.* liver and spleen) as a response to the presence of tumour. In this study CT26 murine colorectal tumour tissues were collected at three different time-points TP1 (7 days post tumour implantation), TP2 (14 days post tumour implantation) and TP3 (20 days post tumour implantation). The average tumour size for TP1, TP2 and TP3 tumours were 0.2, 0.5 and 1 mm respectively. In this study, identification of potential tumour tissue biomarkers of progression was investigated in a time-dependent manner in a CT26 colorectal tumour-bearing mouse model. High throughput MALDI-MS proteomic profiling and ANN algorithms for data processing and interpretation was utilised to detect candidate biomarkers from different tumours, collected at 3 time-points. ANN analysis of tryptic peptide profiles from day 7 (TP1) and day 14 (TP2) tumours produced 2 ions that were based on these ions, 80% of the TP1 and TP2 tumours were correctly classified. The ANNs modelling of day 14 (TP2) and day 20 (TP3) tumour tissues resulted in a 2 ion model that accurately classified 83% of the samples. In general, although the classification accuracy for both of the models was relatively high,

there was no significant clear-cut difference between the time-points (based on the population charts) and as these proteins have not been identified, suggestive of their possible involvement in tumour progression process needs to be investigated. Tumours from early stage (TP1) and late stage (TP3) were also compared and based on one ion ANN model (*m/z* value of 3907.4) all the TP1 and TP3 samples were correctly classified. This ion was present in the TP3 tryptic peptide spectra and absent in the TP1 spectra. Moreover, the ions identified in each model were exclusive to the corresponding model and no similar ions were observed. However, the relatively high accuracy of ANN analysis for classification of different time-points is promising and further studies may disclose possible tumour tissue markers. ANN analysis shows sample outliers that may contain distinct tumour features and represent subtypes of tumour. To our knowledge, only one study by Culp and colleagues (Culp *et al.* 2006) carried out a similar approach to ourselves, where they analysed tumour tissue proteome changes in a time-dependent manner in the B16-F10 mouse melanoma model. Tumour proteome changes in tumour tissue collected at different time-points were analysed by 2D gel electrophoresis to obtain a protein map and they identified several proteins that were up or down regulated over the course of tumour progression. The identified proteins were all involved in pathways occurring in tumour progression such as molecules involved in protein folding, cell cycle regulation, RNA processing and angiogenesis, many of which belong to high concentration proteins. It is estimated that number of expressed proteins in mammalian cells or tissues is in the order of 10^6 in magnitude. This is far exceeds the dynamic range of 2D gels and therefore for detection of low abundant proteins more sensitive approaches are required (Gygi *et al.* 2000). The methodology used in this study of CT26 progression model was the use of C_{18} ZipTip clean-up prior to MALDI-MS profiling which remove the high abundant proteins as already demonstrated in Chapter 2. Therefore, by de-convoluting the sample via C_{18} ZipTips it is more plausible to detect lower abundant proteins although further investigations are required to identify the proposed biomarkers and to verify our findings. However, our methodology for analysis of tissue lysates could benefit from additional fractionation to reveal a complete map of the proteins in the tumour tissues obtained from each time-point. A study by Whiteaker *et al.* (Whiteaker *et al.* 2007) demonstrated the use of fractionation methods to enhance the number of protein detection. The sudy by Whiteaker *et al.* used tumour and normal tissues from a conditional HER2/Neu-driven

mouse model of cancer and identified approximately 700 proteins by LC-MS/MS. This was followed by statistical approach that detected differentially expressed proteins in the tumour tissues (*i.e.* proteins that were under or over expressed compared to the normal tissue based on the LC-MS/MS data) and the discriminatory proteins were then validated using antibody-based and multiple reaction monitoring-mass spectrometry. A number of proteins including fibrinogen-γ, Osteopontin, plastin-2 and kappa-casein were identified that were over expressed in the breast tumour tissues.

In conclusion, our data shows that an integrated approach to tryptic peptide profiling of serum and tumour tissue samples, with MALDI-MS combined with ANNs, leads to identification of panel of biomarkers capable of classifying different samples collected at different time-points during tumour progression. However, further investigations are required to identify the corresponding proteins to the identified discriminatory ions, followed by studies on the relevance of the proteins in the tumour progression process. The identified candidate biomarkers have the potential as prognostic markers or can be used as targets of therapy. In addition, better understanding of mechanisms involved in tumour progression can be achieved by these biomarkers. The use of this methodology can be refined and extended to be used for human tissue specimens in order to detect tumour biomarkers.

Chapter 5 – Serum and Tissue Proteomic Profiling in the Murine CT26 Colon Carcinoma Model of DISC-mGM-CSF-HSV Immunotherapy

5.1 Introduction

The syngeneic CT26 mouse model of colon carcinoma has been extensively used to study aspects of cancer biology, therapeutics and to investigate the mechanisms involved in colorectal carcinoma metastasis to organs such as liver (Wen *et al.* 2007) and lungs (Heinrich *et al*. 2006). In addition, it has been widely used to examine the efficiency of novel cancer vaccines (Schanzer *et al*. 1997, Suh *et al.* 1999 and Levy *et al.* 2006). The CT26 mouse model is established in our laboratory and was used to assess a novel immunogene therapy using DISC-HSV as a vector which encoding the gene for mGM-CSF (Rees *et al.* 2002). Direct injection of DISC-HSV into the established CT26 colon tumours induced regression in up to 70% of the mice receiving the vaccine; complete regression of the tumour (tumour regressors) was observed with the remaining animals leaving progressive tumours (tumor progressors) (Ali *et al.* 2002). The use of DISC-HSV infected whole cell vaccine in CT26 tumour bearer was successful and this model was used to study the molecular/immunological mechanisms involved in the regression/progression following DISC immunotherapy (Ahmad *et al.* 2005). This resulted in the identification and isolation of a CD3+ effector T cell population from the spleen of the regressor animals that were cytotoxic for tumour cells and essential for tumour regression. A suppressed or very low CTL activity was observed in progressor animals co-incidential with the production of high levels of IL-10 and low expression of IFN-gamma. In addition, it was shown that depletion of CD3+CD4+ T cells resulted in an increase in production of IFN-gamma and a drop in IL-10 production and finally restoration of CTL activity. To advance our understanding of mechanisms involved in tumour progression/regression, this model was used to investigate protein biomarkers in serum and tissue which associate with regression or progression of CT26 tumour following DISC-HSV-mGM-CSF therapy.

5.1.1 DISC-HSV in cancer immunotherapy of murine CT26 colorectal carcinoma

The application of viral vectors, to delivering a range of immunomodulatory genes to autologous and allogeneic tumour cell lines has been investigated. Genetically engineered cells expressing immunomodulatory genes, such as cytokines, tumour antigens and costimulatory molecules, have been used in cell based vaccines for cancer therapy. For example, vaccination with transfected tumour cell lines encoding interlukin-4 (IL-4), tumour necrosis factor-alpha (TNF-α), granulocyte-macrophage colony-stimulating factor **(**GM-CSF), IL-6, IL-7 and interferon-gamma (IFN-γ) show enhanced antitumour immunity in mice (Ali *et al*. 2000a). Moreover, viral vectors have been used for direct delivery of immunomodulatory genes to solid tumours and antigen presenting cells (APCs). Genetically modified adenovirus and herpes simplex virus (HSV) have been used for the purpose of cancer therapy and show significant efficacy and safety in humans and mice. Disabled infectious single-cycle herpes simplex virus (DISC-HSV) is a genetically modified HSV, which is attenuated and was developed initially as a vaccine candidate against genital herpes infection (Dilloo *et al.* 1997). Additional studies inferred DISC-HSV to be a suitable vector for cancer immunotherapy. DISC-HSV is a virus lacking the glycoprotein H (*gH*) that is essential for production of infectious progeny hence, deletion of the *gH* gene restricts virus replication to one cycle. This property prevents the spread of DISC-HSV from cell to cell and although the virus can proliferate after infecting the cell, its progeny are non-infectious (Assudani *et al.* 2005). In addition, DISC-HSV is capable of infecting a broad range of dividing (mitotic) and non-dividing (non-mitotic) cell lines including human carcinoma cell lines (*e.g.* breast and prostate), murine tumour cell lines (colorectal and renal carcinoma) and primary human leukaemia and neuroblastomes with high efficiency (Ali *et al*. 2000, Dilloo *et al.* 1997). DISC-HSV contains a relatively small genome (approximately 150 kb), making it suitable for insertion of genetic material of 30-50 kb size without compromising virus replication (Latchman 2001). DISC-HSV can infect both human and murine cancerous cell line and tissue, making it an attractive vector for delivering immunomodulatory genes to solid tumours. A number of DISC-HSV vectors have been developed to deliver cytokines such as IL-2 and GM-CSF and the co-stimulatory molecule B-7.1 to the tumour site (Rees *et al.* 2002). Experimental studies demonstrate that the majority of the infected cells die from necrosis while a small percentage (approximately

5-10%) of cells die by apoptosis; following infection the expression of the therapeutic gene delivered to cells by DISC-HSV lasts up to 72 hours. Moreover, the DISC-HSV empty vector has the ability to enhance the immune response against tumours *in vivo*, acting as an immune adjuvant (Rees *et al.* 2002).

The efficacy of genetically engineered DISC-HSV for cancer immunotherapy has been investigated using a number strategies in mouse RENCA (murine renal cancer cells) and CT26 (murine colorectal carcinoma) tumour models (Ali *et al.* 2002, Ali *et al.* 2000 and Todryk *et al.* 1999). Initially, RENCA cells were infected *ex vivo* with DISC-HSV virus encoding for either IL-2 or mGM-CSF. Mice received two prophylactic immunisations prior to tumour challenge. Both DISC-HSV-IL-2 and mGM-CSF showed similar enhanced immunity to tumour challenge as demonstrated by a reduced tumour incidence and growth rate. Subsequently, the DISC-mGM-CSF-HSV infected RENCA tumour cells were used as a whole cell vaccine to treat established RENCA tumours, resulting in a reduced or delayed tumour growth up to 9 weeks post vaccination (Ali *et al.* 2000). Further investigations verified that the antitumour activity in these models required antigen-specific cytotoxic T lymphocytes and T helper lymphocytes. When DISC-mGM-CSF-HSV was directly injected intra-tumourally into established CT26 murine colorectal carcinoma tumours, complete tumour regression occurred in up to 70% of the treated animals (Ali *et al.* 2002 and Ahmad *et al.* 2005) while same animals also regressing their tumours following injection of the empty DISC-HSV vector or DISC-HSV carrying an irrelevant gene (Assudani *et al.* 2005).

5.1.2 Aims and objectives

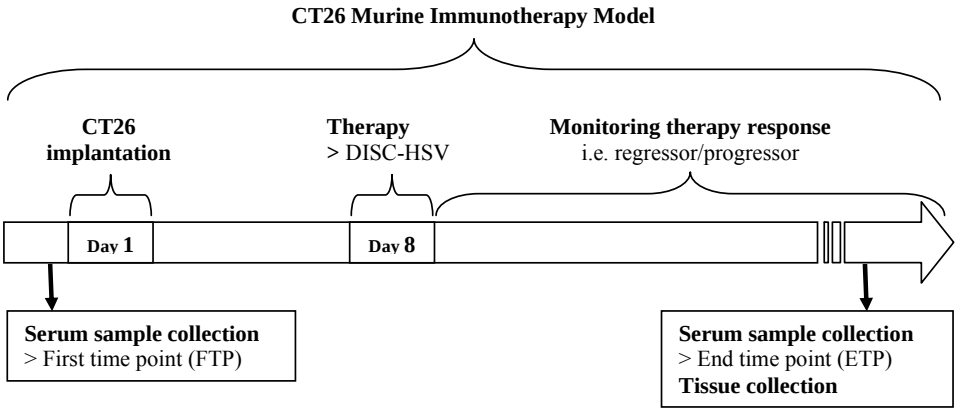
The main aims are as follow:

- To use a prospective cohort of samples collected at the end of therapy outcome as a discoveriy set to obtain MALDI-MS profiles. To interrogate these profiles using ANNs to investigate whether is possible to obtain panel of markers that are discriminatory between the responder, non-responder and naïve mice and may associate with outcome of therapy.
- To identify the proteins associated with the candidate panel of biomarkers.

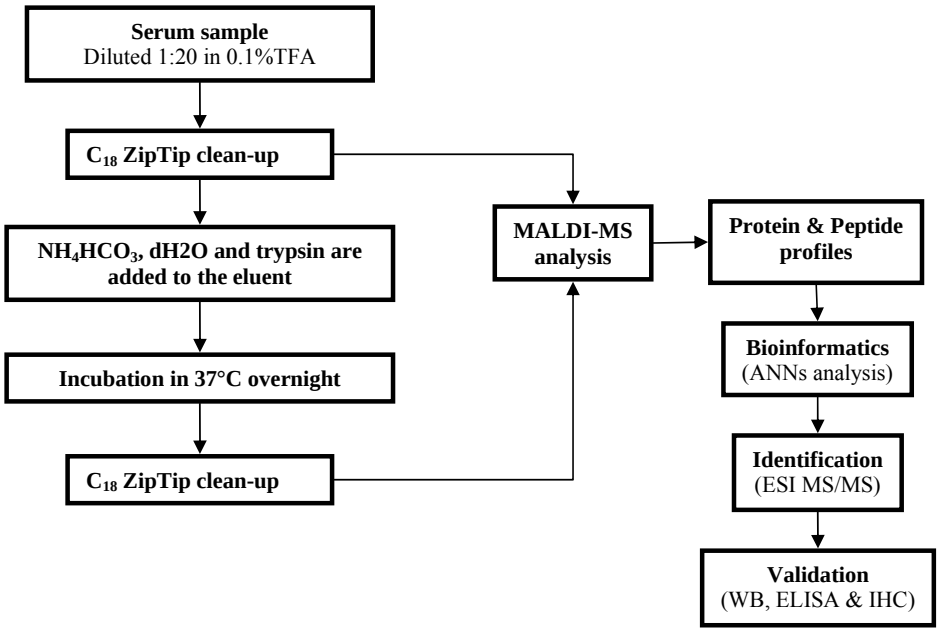
- To collect prospective set of samples from the CT26 colorectal cancer model of DISC-HSV therapy model as a validation set both prior to therapy and at the end of therapy outcome.
- To validate the proteomic based results using non-MS based methods (*i.e.* Western blotting, ELISA, IHC and qRT-PCR) in the discovery set as well as a prospective set of samples.

5.2 Summary of methods

A) CT26 Murine immunotherapy model and sample collection



B) Sample preparation for proteomic analysis



5.3 Results

5.3.1 Protein and peptide analysis of serum from the retrospective model by MALDI-MS.

Serum samples (from the retrospective model) from 14 regressors, 21 progressors and 35 control mice were subjected to MS analysis. Using MALDI-TOF MS serum profiling, the serum protein (after C_{18} ZipTip clean-up) and peptide profiles of DISC therapy responder mice (regressors), DISC therapy non-responder mice (progressors) and control healthy mice (naïve) with no therapy or tumour were compared. The reproducibility of the sample preparation for MALDI-MS analysis and both MALDI and Xcise instruments were previously shown in chapter 3 and reported by Matharoo-Ball and colleagues (Matharoo-Ball *et al.* 2007). Representative serum protein and tryptic peptide MALDI profiles are shown in figure 5-1 A and B respectively. The protein serum spectra after C_{18} ZipTip clean-up was generated by MALDI-MS, operated in linear mode and the tryptic peptide serum profiles were generated by MALDI-MS, operated in reflectron mode. Several visual differences were observed between the protein serum spectra after C_{18} ZipTip clean-up and the related tryptic peptides and a number of of these are highlighted in figure 5-1. For example, visual spectral differences can be observed in the protein profile between the three different groups (figure 5-1A), in the mass range of 3200-3800 Da. Peaks present in this range showed higher intensity in the spectra obtained from the progressors when compared with regressor and controls which showed similarities. In addition, the relative intensity of peak having an *m/z* value of 6400 Da appears to gradually decrease from the control to regressors and then progressors. An example of visual spectral differences for the tryptic peptide profiles can be observed in the mass range of 1750-2000 Da (figure 5-1B). The spectra in this region showed considerable similarities between the control and regressors whereas the intensity of the peaks was significantly higher in the spectra obtained from progressor mice.

Visual assessment of the MALDI spectra was necessary to determine the quality of the produced spectra and the reproducibility of sample preparation procedures. In addition, all the acquired MALDI spectra were visually assessed according to strict criteria (described in chapter 3) prior to statistical analysis for biomarker discovery to insure the high quality of the data. Over all, there appeared to be slight visual spectral differences between the

proteomic profiles of regressor and control mice in the protein profile after C_{18} ZipTip and tryptic peptide profiles. However, there was more noticeable variation in the progressor proteomic profiles in comparison to regressor and controls.

The methodology used here was relatively high-throughput and therefore it is not possible to analyse multiple samples for each experiment by visual assessment of the spectra. Moreover, the visual spectral differences highlighted in figure 5-1 show peaks that have higher abundance and therefore it is easy to detect differences between the groups. In each spectrum, there were a number of peaks with lower intensities which may be significant and possibly discriminatory between the groups, but it was not possible to assess these visually with any degree of accuracy. Therefore, these complex data sets from the retrospective model were subjected to competitive algorithm (*i.e.* ANN) analysis to identify differences. Here, serum samples collected from the retrospective model was used for ANNs modeling. Although serum samples from 14 regressors, 21 progressors and 35 control mice was used for MALDI proteomic profiling, the number of samples used to generate the ANN models may be lower in each analysis. This is due to the visual spectral checks prior to the ANNs modeling which eliminates spectra with poor quality.

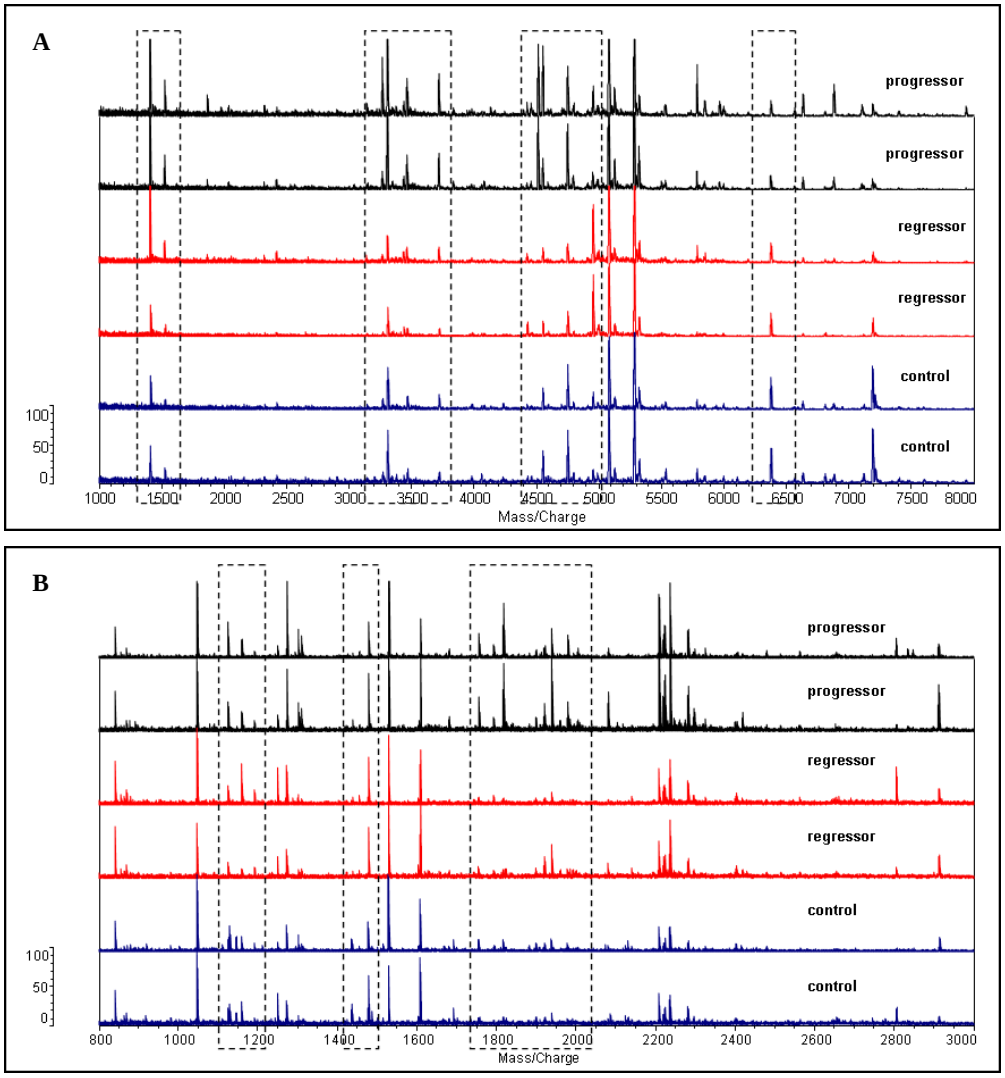


Figure 5-1. Figure showing protein and tryptic peptide MALDI spectra obtained from control, regressor and progressor mice. The X axis represents the *m/z* ratio and the Y axis represents relative intensity. **A)** Representative MALDI-TOF mass spectra of regressor, progressor and control mouse serum proteins after C_{18} ZipTip clean-up mass range of 1000-8000 Da, acquired in linear mode. Serum samples were spotted with SA on the MALDI steel plate and allowed to air dry. Protein peaks in the regions highlighted in the figure appears to be up or down regulated visually, when a three-way comparison was carried out on the progressor, regressor and control mice. **B**) Representative MALDI-TOF mass spectra of regressor, progressor and control mouse serum tryptic digest mass range of 800-3000 Da, acquired in reflectron mode. Serum samples were spotted with CHCA on the MALDI steel plate and allowed to air dry. Peptide peaks in the regions highlighted in the figure appears to be up or down regulated when a three-way comparison was carried out on the progressor, regressor and control mice.

5.3.2 Discovery of discriminatory biomarkers using ANN modeling for the regressor/progressor retrospective model.

The goal of the study was to identify biomarkers that may predict the outcome of DISC-mGM-CSF-HSV therapy and/or may associated with tumour progression/regression. To identify serum biomarkers we used MALDI-MS profiling in conjunction with ANN analysis. Initially, we examined the capability of ANNs to identify and correctly classify regressors from progressors. Hence, a stepwise approach was utilized which produced the minimum number of ions with the potential of detecting regressors from progressors based on the MALDI protein/peptide profiles.

The biomarker discovery experiments were carried out on serum samples which were collected from the retrospective CT26 immunotherapy model (samples collected in 2004) which included 35 serum samples, 21 progressors and 14 regressors. The protein and tryptic peptide profiles were generated via MALDI-MS following C_{18} ZipTip chromatography and processed and analysed as described previously. The generated protein and tryptic peptide profiles were visually checked and spectra with poor quality were eliminated from the study; the remaining spectra were subjected to proceed for further bioinformatic analysis, using ANN modeling. The results of these findings were then validated using serum samples from a prospective CT26 immunotherapy model (sample collection carried out in 2008).

5.3.2.1 Discriminatory serum protein biomarkers identified in the regressor/progressor retrospective model

The MALDI-MS serum protein profiles following C_{18} ZipTip (mass range of 1-20 kDa) of 17 progressors and 12 regressors were used for ANN modeling. A stepwise approach was utilised to identify the minimum number of ions capable of assigning samples correctly to their respective groups. Prior to the ANN analysis, the data was smoothed, aligned and baseline corrected as described in materials and methods (chapter 2). As each sample was spotted on the MALDI plate in duplicate, one spot for each sample was randomly selected to be used in ANNs modeling. In order to detect a fingerprint based on the MALDI spectra that assigns samples to their respective groups, a stepwise approach was utilised and regressors were labeled as 1 and the progressors were labeled as 2 prior to analysis. The stepwise approach is based on training a number of different ANN models and the general

principles of stepwise approach have been described previously in detail in chapter 3. In order to generate ANN models from MALDI protein profiles of regressor and progressor mice, the stepwise approach involves training a number of models using 19000 inputs for each sample (each input corresponds to intensity at a specific *m/z* value). Initially each input is used as single input to train 50 models and for each model a total of 29 cases (n = 17 progressors and n = 12 regressors) were randomly split into 17 samples for training (60%), 6 for testing (20%) and 6 as a validation set (20%); this allows random cross validation analysis to be performed. After training 50 models, the best performing model was selected for further analysis and all the remaining inputs were added sequentially to the first input to train the model. This procedure was continued until no further improvement in the model was observed. In the case of modeling the protein profiles after C_{18} ZipTip clean-up from the regressor and progressor mice, a total of 5 steps were carried out. The results of the 5 steps are shown in table 5-1 that presents the median accuracies and mean squared errors of training, test and validation for each best performed model in each of the 5 steps. The highest accuracy that predicted progressors from regressors was generated in step 2 (highlighted in red in table 5-1), using two peaks with *m/z* values of 4804 and 11053 with an accuracy of 86%, a sensitivity of 100% and specificity of 94%. The addition of further ions did not improve the prediction of the model and therefore the first 2 ions were considered biomarkers that can best predict regressors from progressors based on MALDI protein profiles.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training error	Test error	Validation Error
1	4804	84%	86%	80%	0.29	0.16	0.2
2	11053	88%	94%	**86%**	0.27	**0.11**	0.21
3	7203	91%	93%	82%	0.19	0.05	0.19
4	9791	90%	92%	84%	0.16	0.08	0.19
5	5790	88%	91%	81%	0.21	0.14	0.28

Table 5-1. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum protein profiling after C_{18} ZipTip clean-up of regressor and progressor mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in step 2, highlighted in red, with median accuracy of 86% and mean squared value of 0.11 and further addition of ions did not affect the accuracy of model.

The MALDI protein spectra differences between the regressor and progressor profiles for the top two *m/z* identified by ANN analysis are shown in figure 5-2. The *m/z* value of 4804 is present in the progressor profiles but absent in the regressors (figure 5-2A) and based on this ion ANNs can accurately classify 80% of the progressor and regressor samples. However, the mass value of 11053 (figure 5-2B) is present in the regressor protein profiles (after C_{18} ZipTip clean-up) but cannot be detected in the progressor profiles. Addition of this ion to the first discriminatory ion, improved the progressor/regressor classification by 6%, achieving the accuracy of 86%.

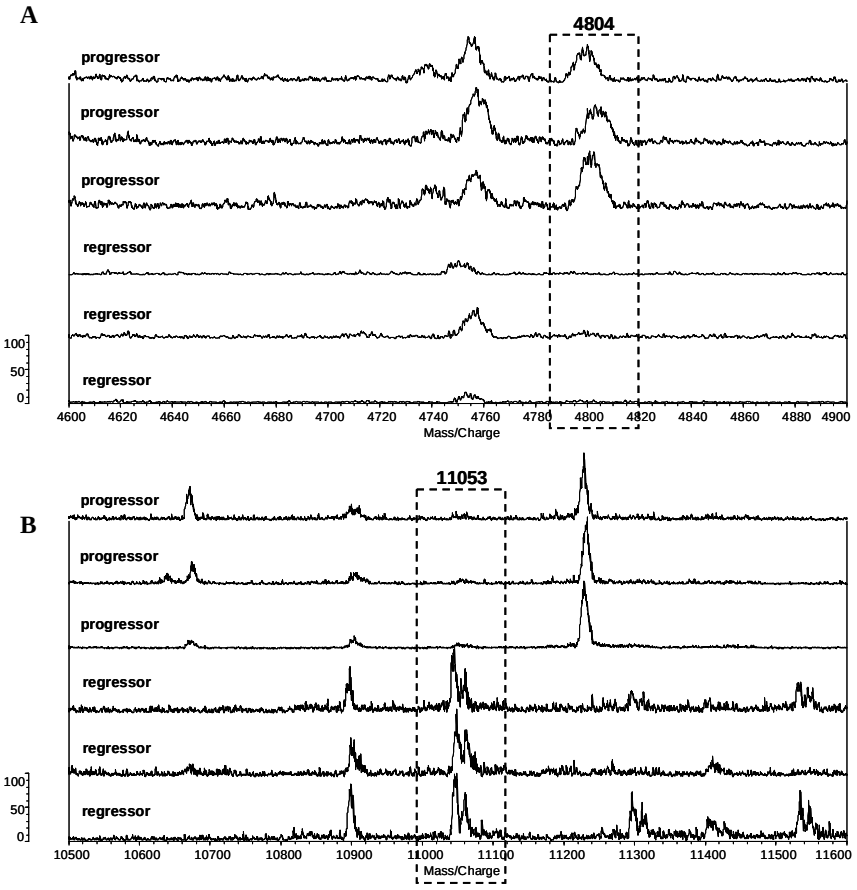


Figure 5-2. Figure showing visual spectral differences for the top two discriminatory ions between regressor and progressors based on MALDI protein profiles after C_{18} ZipTip and ANNs analysis. A) The first discriminatory ion between the regressor and progressor serum MALDI protein profiles was 4804 which is present in the progressor profiles and not in the regressor protein MALDI profiles. **B)** The second discriminatory ion for the progressor and regressor serum protein profiles was 11053 which is present in the regressors protein profiles and not in the progressor profiles.

In order to assess how individual samples were classified by the 2 ion ANNs model, a population chart was generated. Figure 5-3 shows the population distribution of the predicted outputs for all 29 mice using *m/z* values of 4804 and 11053. Samples from progressor animals are highlighted in red (assigned as number 2) and samples from regressor animals are shown in green (assigned as number 1). The ratios below 1.5 were assigned as regressors whilst a ratio above 1.5 was used to classify progressors. The figure shows that all the regressors were correctly classified whereas one of the progressors (mouse number 112) was misclassified as regressor.

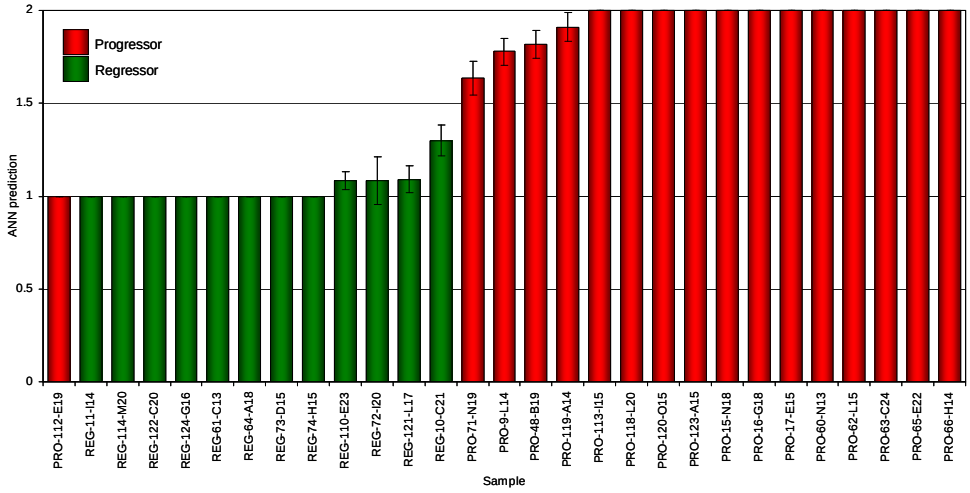


Figure 5-3. Predictive capability of ANNs to recognise protein profiles after C_{18} ZipTip clean-up based on a 2 ion ANNs model. The red bars indicate progressor samples, and the green bars indicate regressor samples. A predictive value below 1.5 indicates a regressor sample, while a prediction greater than 1.5 indicates a progressor sample.

5.3.2.2 Discriminatory serum peptide biomarkers identified in the regressor/progressor retrospective model

In order to identify discriminatory ions between the regressor and progressors based on MALDI tryptic peptide profiles (mass range 800-3500 Da), the same ANN modeling procedure (outlined above) was preformed (27 cases, n = 10 regressors and n = 17 progressors). The ANN model was trained as described previously using 2667 inputs and the stepwise approach was applied using 5 ions; the results are illustrated in table 5-2. A combination of four peaks (*m/z* 1312, *m/z* 1132, *m/z* 1192 and *m/z* 2133) was identified as

being discriminatory between the two groups with the lowest test error. The median accuracy for this analysis was 86%, with a sensitivity 90% and specificity 81%.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training error	Test error	Validation Error
1	1312	81%	81%	78%	0.34	0.25	0.26
2	1132	84%	87%	80%	0.26	0.16	0.25
3	1192	89%	92%	90%	0.18	0.08	0.07
4	2133	89%	91%	**86%**	0.2	**0.08**	0.07
5	2808	88%	92%	83%	0.21	0.13	0.24

Table 5-2. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum tryptic peptides of regressor and progressor mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in step 4, highlighted in red, with median accuracy of 86% and mean squared value of 0.08 and further addition of ions did not affect the accuracy of model.

The MALDI tryptic peptide spectra differences between the regressor and progressor profiles for the top 4 ions identified by ANN analysis are shown in figure 5-4. The *m/z* value of 1312, which represents the first discriminatory ion (figure 5-4C), showed higher relative intensity in the regressor spectra when compared to the progressors. This ion was able to discriminate 78% of the regressors from the progressor. The second and third discriminatory ions were *m/z* values of 1132 and 1192 and the visual spectral differences for these two peaks are shown in figure 5-4A. The peak with a mass value of 1132 was present in the tryptic peptide profiles of progressors and absent in the regressor spectra, whereas the 1192 peak show slightly higher intensity in the regressor spectra. The final *m/z* value of the panel of discriminatory ions was a peak with a mass value of 2133 (figure 5-4B), which was detectable in the tryptic peptide MALDI spectra of progressor mice and absent in the regressor profiles.

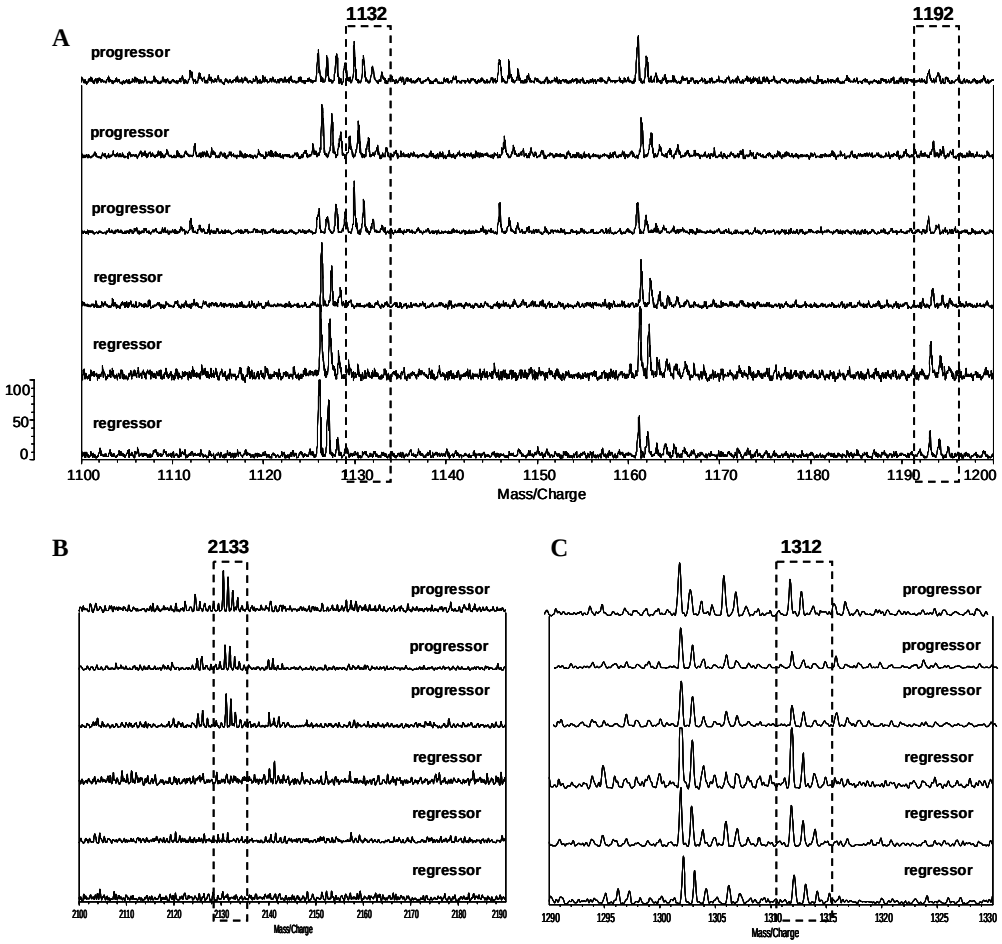


Figure 5-4. Figure showing visual spectral differences for the top 4 discriminatory ions between regressor and progressors based on MALDI tryptic peptide profiles and ANNs analysis. A) The *m/z* values of 1132 and 1192 were two of the panel of discriminatory ions for regressor and progressors based on the tryptic peptide profiles. The 1132 peak is present in the progressor profiles and not in the regressor profiles however the relative intensity of the 1192 peak appears to be higher in the regressor tryptic peptide profiles. **B)** The last discriminatory ion between the regressor and progressors was 2133 which is present in the progressors and not in the regressor profiles. **C)** The first discriminatory ion for the regressor and progressors was 1312 which shows a significant higher relative intensity in the regressor spectra when compared to the progressors.

The potency of ANN analysis to correctly predict tryptic peptide profiled based on the 4 predictive ions obtained from regressor and progressor modeling was further examined by generating a population chart. The chart represents cases were classified based on the 4 ion ANNs modeling of tryptic peptide MALDI profiles and the results are illustrated in figure 5-5. The ratios below 1.5 were assigned as regressors whilst a ratio above 1.5 was used to classify progressors. Based on the panel of biomarkers, only one regressor and two progressors were misclassified.

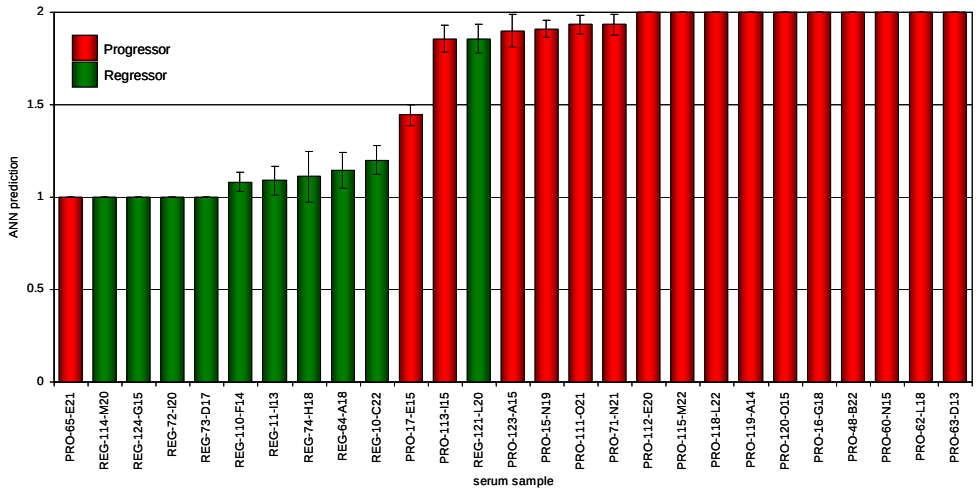


Figure 5-5. Predictive capability of ANNs to recognise tryptic peptide profiles based on a 4 ion ANNs model. The red bars indicate progressor samples, and the green bars indicate regressor samples. A predictive value below 1.5 indicates a regressor sample, while a prediction greater than 1.5 indicates a progressor sample.

5.3.3 Biomarker identification by ANN analysis discriminating controls from progressors.

Once the discriminatory biomarkers between the regressors and progressors had been identified, it was important to identify biomarkers that potentially were involved in tumour progression by comparing proteomic data from healthy mice with the progressors. Serum samples from 21 progressors and 35 control mice were used to generate MALDI proteomic profiles (after C_{18} ZipTip clean-up) and tryptic peptide profiles. Following MALDI analysis and visual evaluation of the spectra and elimination of poor quality spectra, the stepwise approach was used to identify candidate biomarkers between progressor and control healthy mice.

5.3.3.1 Discriminatory serum protein biomarkers identified in the progressor/control retrospective model

The MALDI serum protein profiles (mass range of 1-20 kDa) of 17 progressors and 29 control (naïve) mice were proceed for ANN modeling using 19000 inputs for each sample (each input corresponds to intensity at a specific *m/z* value). Initially each input is used as a single input to train 50 models and in each model the 46 cases (n = 17 progressors and n = 29 controls) were randomly split into 27 samples for training (60%), 9 for testing (20%) and 10 as validation set (20%) (random cross validation). After which the best performing model is selected for further analysis and all the remaining inputs were added sequentially to the first input to train the model. This procedure was continued until no further improvement in the model was observed. The stepwise analysis was stopped at the end of step 3, were no significant improvement was achieved by addition of new inputs and the results shown in table 5-3 presents the median accuracies and mean squared errors of training, test and validation for each best performed model in each of the 3 steps. The highest accuracy of prediction that distinguished progressors form controls was generated in step 2 (highlighted in red in table 5-3), using two peaks with *m/z* values of 4515 and 18318 with an accuracy of 83%, sensitivity 100% and specificity of 94%.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training Error	Test error	Validation Error
1	4515	83%	86%	80%	0.28	0.13	0.24
2	18318	85%	88%	**83%**	0.28	**0.17**	0.24
3	5076	85%	88%	83%	0.26	0.19	0.23

Table 5-3. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum protein profiling after C_{18} ZipTip clean-up of control (naïve) and progressor mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in step 2, highlighted in red, with median accuracy of 83% and mean squared value of 0.17 and further addition of ions did not affect the accuracy of model.

5.3.3.2 Discriminatory serum tryptic peptide biomarkers identified in the progressor/control retrospective model

The MALDI spectra of serum tryptic peptides of regressors and control mice were visually checked (according to the criteria previously described) for acceptance of the spectra. A total of 26 control and 17 progressor tryptic peptide profiles were accepted and used for

bioinformatic analysis. Stepwise analysis was performed on these profiles to assess the ability of ANNs to classify tryptic peptide serum profiles of progressors and control mice. A total of 5 steps were carried out and the results of the stepwise analysis are illustrated in table 5-4. The best ANN prediction with one peptide peak (*m/z* 1130) that discriminated between the controls and progressors with an accuracy of 85%, sensitivity of 88% and specificity of 96%.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training Error	Test error	Validation Error
1	1130	86%	87%	**85%**	0.26	**0.21**	0.24
2	1956	89%	90%	87%	0.23	0.12	0.20
3	3149	87%	89%	86%	0.2	0.09	0.13
4	3082	87%	89%	85%	0.16	0.08	0.13
5	2432	94%	94%	87%	0.15	0.07	0.2

Table 5-4. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum tryptic peptides of control (naïve) and progressor mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in the first step, highlighted in red, with median accuracy of 85% and mean squared value of 0.21 and further addition of ions did not affect the accuracy of model.

The MALDI tryptic peptide spectra differences between the control and progressor profiles for the top ions identified by ANN analysis are shown in figure 5-6. The *m/z* value of 1130 which was (the first discriminatory ion) was is present in the spectra obtained from progressors while this peak cannot be detected in the spectra of control mice. This ion predicted progressors from the controls with an accuracy of 85%.

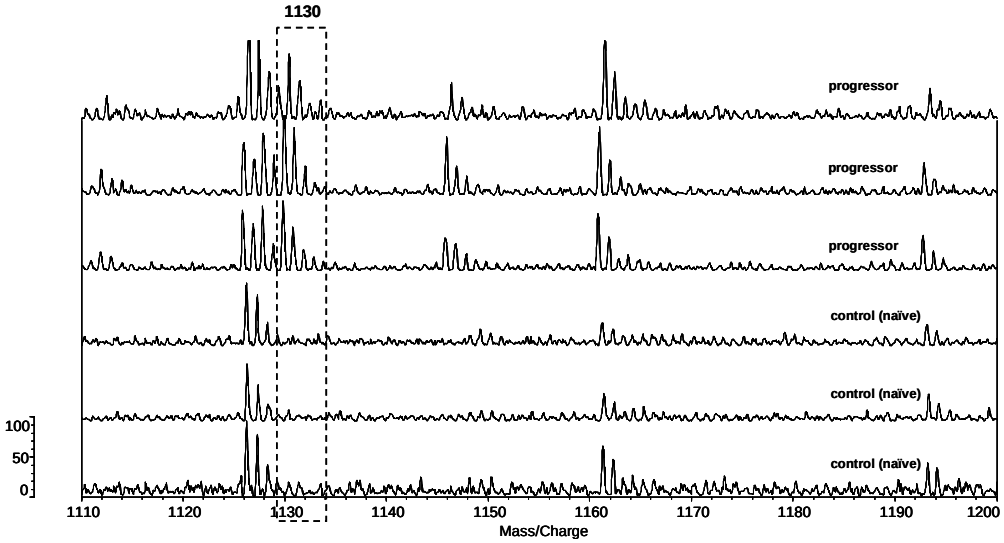


Figure 5-6. Figure showing visual spectral differences for the top discriminatory ion between control (naïve) and progressors based on MALDI tryptic peptide profiles and ANNs analysis. The discriminatory ion for the control and progressors was 1130 which is present in the spectra corresponding to progressors and absent in the spectra obtained from the controls.

Furthermore, the potency of ANNs correctly classifying controls from progressors based on the one ion was further examined and shown in population charts for protein and peptides. The ratios below 1.5 were assigned to the control group and ratios above 1.5 were classified as progressors (figure 5-7). Two progressors and one control were misclassified using based on one ion. The chart shows that one control and 2 progressors were misclassified based on the one discriminatory ion.

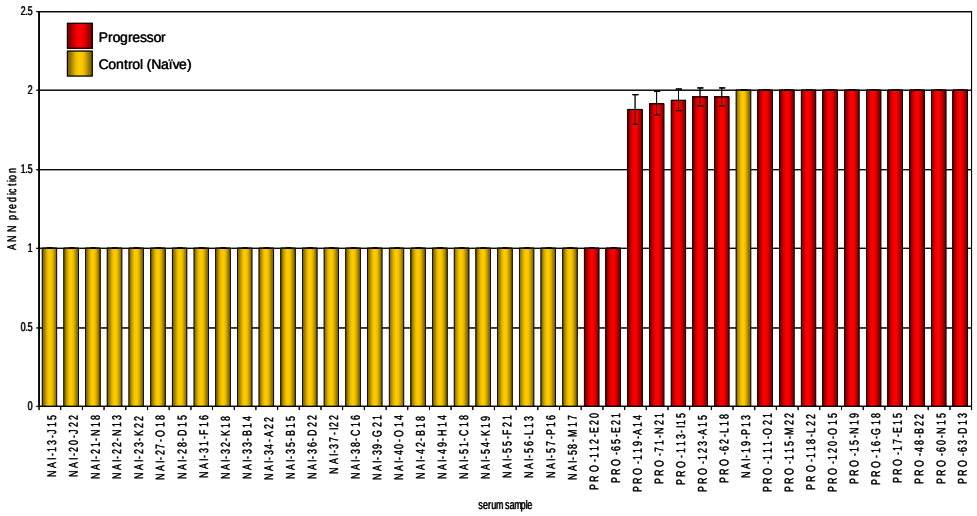


Figure 5-7. Predictive capability of ANNs to recognise tryptic peptide profiles based on 1 ion ANNs model. The red bars indicate progressor samples, and the yellow bars indicate control (naïve) samples. A predictive value below 1.5 indicates a control sample, while a prediction greater than 1.5 indicates a progressor sample.

5.3.4 Biomarker Identification by ANN analysis discriminating controls from regressors.

Once the spectra of control and regressor tryptic peptides were checked visually, 26 control and 10 regressor spectra that met the criteria were subjected to stepwise analysis (as described above). The capability of ANN analysis to classify naive and regressor samples was low with just over 50% of the data being correctly classified.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training error	Test error	Validation error
1	1611	56%	57%	56%	0.42	0.33	0.43
2	1458	56%	56%	52%	0.44	0.29	0.41
3	1287	50%	48%	47%	0.48	0.28	0.41

Table 5-5. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum tryptic peptides of control (naïve) and regressor mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The model failed to classify regressors from the progressors and no significant discriminatory ion was identified by ANNs analysis.

The ANNs ability to classify control and progressors was examined by generating a population chart and the results are shown in figure 5-8. ANN analysis classified 100% of the naive samples correctly whereas all the regressor samples were misclassified. These results indicate that the regressor peptide profiles could not be discriminated from healthy controls, whereas progressor profiles were significantly different from controls. The analysis suggests that there are no or insignificant differences between the MALDI tryptic peptide profiles of control and regressor animals with the inference that regressor mice reverted to a normal proteomic phenotype.

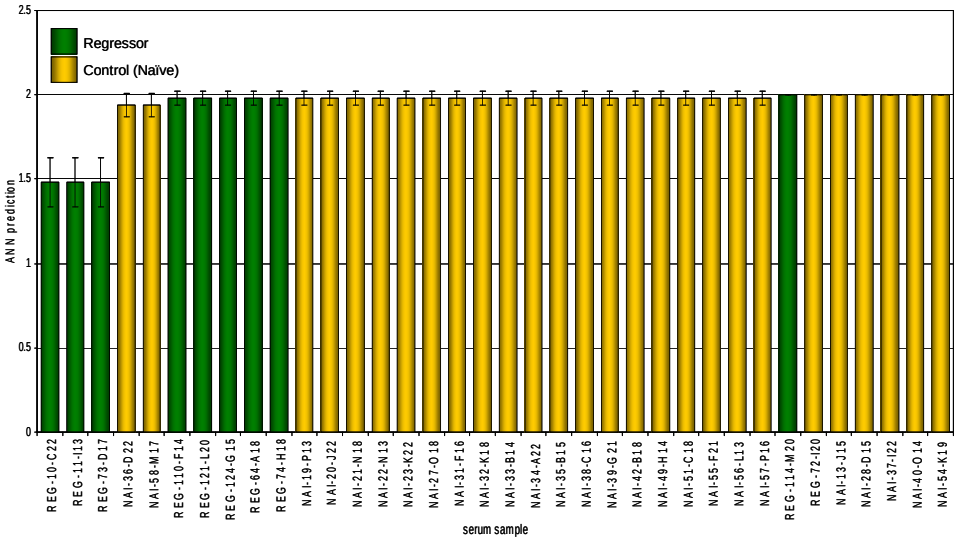


Figure 5-8. Predictive capability of ANNs trained with the top 1 discriminatory ion. The black bars represent regressors and the white bars are naives. Sample with predictive values below 1.5 are regressors and those with values above 1.5 are naives.

5.3.5 Identification of predicted biomarkers that discriminated progressors from regressors, using ESI MS/MS.

The four top ions from tryptic digest predicted by ANN analysis were identified by LC-ESI-QIT MS/MS and the results are summarised in table 5-6. The samples for ESI analysis were prepared using the same method used for MALDI-MS analysis; pooled serum samples were then run on the ESI to identify the ions.

m/z	Test Perf.	Protein	Sequence	Score	pI
1312	78%	Haemoglobin β-2 subunit	VNPDEVGGEALGR	126.1	8.22
1132	80%	Serum amyloid A-1 protein precursor	EAFQEFFGR	251.2	6.61
1192	90%	Hemopexin precursor	NPITSVDAAFR	178.1	7.63
2133	86%	Serum amyloid P component precursor	APPSIVLGQEQDNYGGGFQR	196.3	5.95

Table 5-6. Protein identification by LC-ESI-QIT MS/MS using ANNs predicted tryptic peptide ions for discrimination of progressors from regressors.

The top 4 tryptic peptide peaks predicted by ANNs (*m/z* 1312, *m/z* 1132, *m/z* 1192 and *m/z* 2133) were analysed by ESI MS/MS to obtain sequence data (figure 5-9). The doubly charged peptide from peak *m/z* 1312 was *m/z* 656.5 and the amino acid sequence was identified as VNPDEVGGEALGR which corresponds to part of the haemoglobin β-2 subunit. Sequence analysis of doubly charged peaks of mass values of *m/z* 1132 with amino acid sequence of EAFQEFFGR verified that this protein was serum amyloid A-1 (SAA-1). Peaks with *m/z* 2133 and *m/z* 1192 were identified as fragments of serum amyloid P (SAP) and hemopexin (HPX) respectively; the amino acid sequence corresponding to SAP was APPSIVLGQEQDNYGGGFQR and for HPX was NPITSVDAAFR.

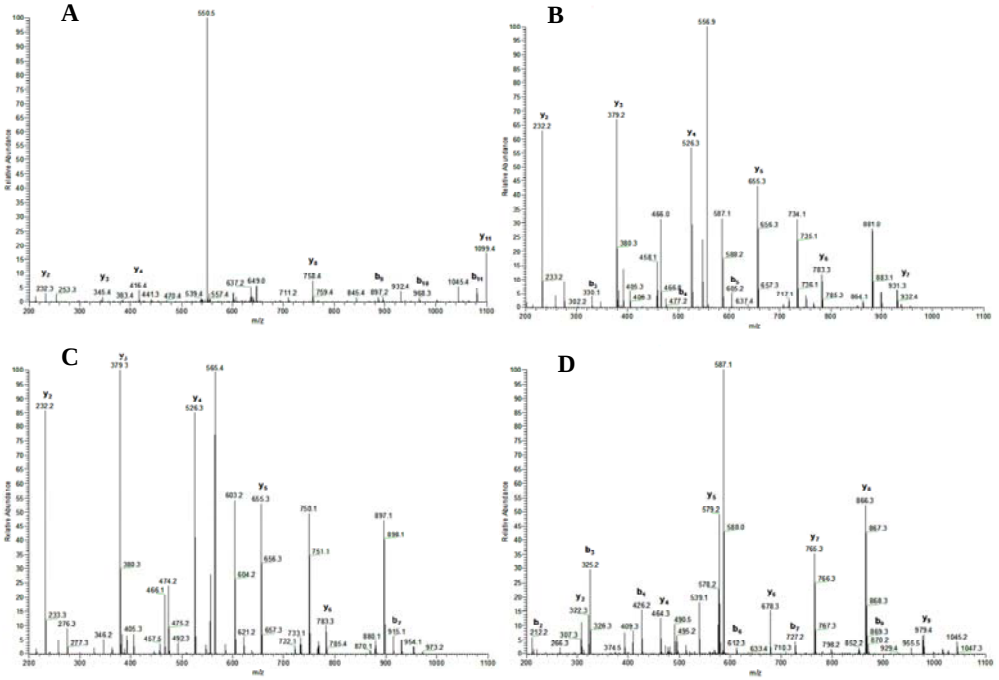


Figure 5-9. Sequence analysis of serum tryptic peptide by LC-ESI-QIT MS/MS for the biomarkers derived from the ANNs modeling of progressors and regressors. The ion with *m/z* of **A)** 1312 was derived from haemoglobin β-2 and amino acid sequence of VNPDEVGGEALGR, **B)** 1132 was derived from serum amyloid A-1 (SAA1) and amino acid sequence of EAFQEFFGR, **C)** 1192 was derived from hemopexin (HPX) with amino acid sequence of NPITSVDAAFR and **D)** 2133 was derived from serum amyloid P (SAP) with amino acid sequence of APPSIVLGQEQDNYGGGFQR.

5.3.6 Identification of predicted biomarkers that discriminated progressors from controls, using ESI MS/MS.

The discriminatory peak from the ANN model of progressors versus controls was *m/z* 1130 and this was targeted in both pooled naïve and progressor serum by LC-ESI-QIT MS/MS (table 5-7). The sample for ESI analysis was prepared as described previously and the amino acid sequence of the doubly charged peptide from peak *m/z* 1130 was *m/z* 565.5 and identified as WFWDFART, a sequence of the hemopexin protein.

m/z	Test Perf.	Test Error	Protein	Sequence	pI
1312	86%	0.21	Hemopexin precursor	WFWDFART	7.63

Table 5-7. Protein identification by LC-ESI-QIT MS/MS using ANNs predicted tryptic peptide ions for discrimination of progressors from regressors.

5.3.7 Analysis of RP-SPE serum fractions.

In order to identify proteins that were specific to the progressors (failure of immunotherapy), we compared the serum protein expression of progressors with serum of tumour bearer (TB) animals. The hypothesis was that proteins that are differentially expressed might be correlated to failure of therapy and not only to the tumour expression. Serum samples from 4 TB and 4 progressors from the prospective mice model were prefractionated using RP-SPE columns and by sequential elution, 10 fractions for each sample were collected. The first fraction eluted with 20% ACN did not contain any bands meaning that no protein was eluting from the RP-SPE column with the lowest concentration of ACN. The equivalent ACN concentration fractions were analysed by SDS-PAGE and stained with Coomassie blue and the results are presented in figure 5-11 (the fraction eluted with 20%ACN is not shown in this figure). Visual analysis of the gels reveals 19 pronounced differences between the TB and progressors and these were excised from the gel and tryptically digested. The peptide mass fingerprints of the gel band digests were generated by MALDI-TOF MS analysis. A representative SDS-PAGE containing all the 10 fractions collected from one serum sample is presented in figure 5-10. In addition, figure 5-10 show the location, identity number and approximate mass of the excised bands that were differently expressed between the equivalent ACN concentration fractions of TB and progressor samples. The list of the identified proteins from band 1-19 (shown in figure 5-10) is represented in table 5-8.

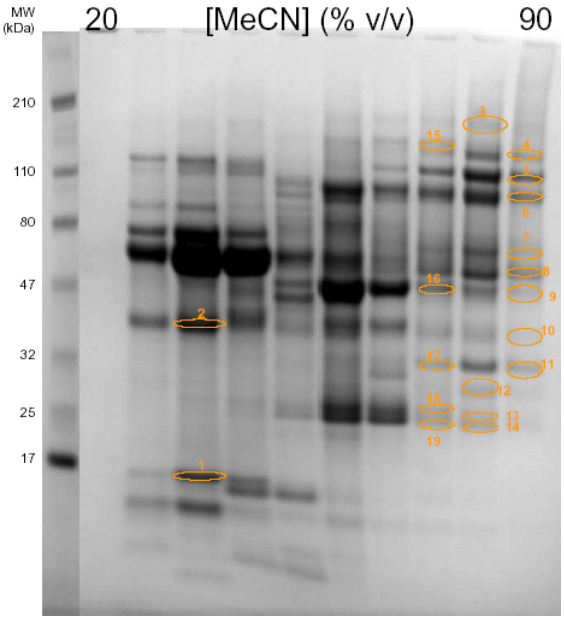


Figure 5-10. Representative SDS-PAGE containing all the 10 fractions collected from progressor serum sample. The figure show the location, identity number and approximate mass of the excised bands that were differently expressed between the equivalent ACN concentration fractions of TB and progressor samples.

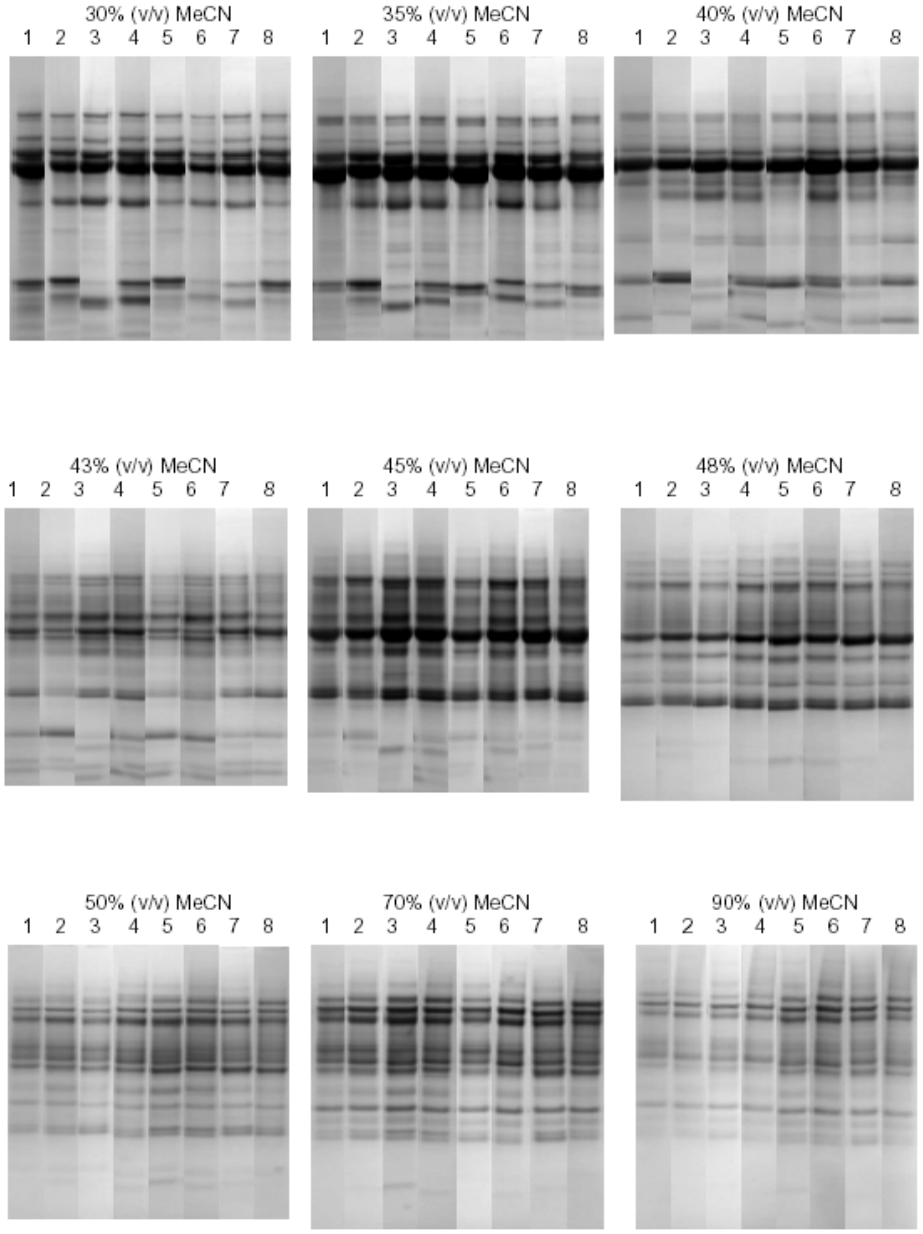


Figure 5-11. Comassie blue stained image of the SDS-PAGE showing the 30%, 35%, 40%, 43%, 45%, 48%, 50%, 70% and 90% ACN wash fractions from RP-SPE analysis of serum samples from tumour bearer (TB) and progressor mice. Numbers 1-4 are TB and 5-8 are progressors.

Gel band No.	Protein name
1	hemoglobin alpha chains
	hemoglobin beta major chain
2	haptoglobin precursor
	Adult male stomach cDNA, RIKEN full-length enriched library
	Bone marrow macrophage cDNA, RIKEN full-length enriched library
3	Apob protein
	Pregnancy zone protein
	Fibronectin precursor
4	Pregnancy zone protein
	Murinoglobulin
	complement C3 precursor
	Trio protein
5	Pregnancy zone protein
	Golgin subfamily A member 3 (Golgin-160)
6	complement C3 precursor
	Pregnancy zone protein
	Adult male liver tumor cDNA, RIKEN full-length enriched library
7	Serine
	complement C3 precursor
	Activated spleen cDNA
8	complement C3 precursor
	Serine
	Krt2-5 protein
9	Alpha-1-antitrypsin 1-6 precursor
	Serpina1a protein
	alpha-1 proteinase inhibitor
	Immunoglobulin gamma1 heavy chain
	Ig gamma-2b chain precursor
	Ig gamma-3 chain C region
	alpha-1-antichymotrypsin-like protein
10	haptoglobin precursor
11	Pregnancy zone protein
	complement C3 precursor
	Hydrocephalus-inducing protein
12	Pregnancy zone protein
13	Serum amyloid P-component
14	ANTIBODY LIGHT CHAIN FAB (FRAGMENT)
	fabe8a fab, chain L
	idiotypic fab 730.1.4 (igg1) of virus neutralizing antibody, chain L
	Ig kappa chain V-C region (Fab fragment with N9 neuraminidase-NC41), chain L
	antibody desire-1 fab cleaved by papain, chain L
15	Murinoglobulin 1
16	Serpina1a protein
17	complement C3 precursor
18	Serum amyloid P-component
19	Hemopexin precursor

Table 5-8. Peptide mass fingerprint of differentially expressed proteins between serum samples from progressor and tumour bearer mice by MALDI-TOF MS.

Many of the proteins identified were immune response related, for example, complement C and apolipoprotein. However, SAP, HPX and HBB were also identified as proteins differentially expressed between tumour-bearer and progressor serum samples and were identified discriminatory biomarkers between regressor and progressor sera by ANN modeling. These results further suggest that there is a correlation between these three proteins (SAP, HPX and HBB) and failure of immunotherapy in the progressor animals.

5.3.8 Validation of serum biomarkers corresponding to tumour progression and regression, using Western blot analysis.

The results of both MALDI serum profiling and RP-SPE fractionation identified SAP, HPX, SAA1 and HBB as important discriminatory biomarkers and therefore further validation studies using immunoassays were conducted for these proteins. A unique aspect of our validation study is the use of both serum samples from prospective and retrospective CT26 immunotherapy model. Using samples collected form the retrospective murine model (conducted in 2004) which was used for biomarker discovery and serum samples from the prospective model (conducted in 2008). The three proteins (SAP, SAA-1, HPX and HBB) identified by ANN analysis of the regressor versus progressor model were further analysed by semi-quantitative Western blotting analysis. In the first instant, western blot analysis was carried out on the serum samples of regressor (n = 7), progressor (n = 17) and control/naïve (n = 16) mice from the retrospective murine model used for biomarker discovery. Serum protein (10mg/ml) from regressor, progressor and control mice were separated by SDS-PAGE and transferred to PVDF membranes, and probed with the appropriate antibodies. Western blotting using HBB antibody and ECL reagent revealed a major band of approximately 16 kDa which is the expected mass for mouse HBB. Western blots were preformed on all the retrospective samples and a representative result is shown in figure 5-12A. In the case of haemoglobin, a band was observed at 16 kDa (figure 5-12A), however no differential pattern of expression for this protein was observed despite it being the first top discriminatory protein identified by ANN analysis, distinguishing progressor from regressors serum. Western analysis for SAA-1 revealed bands around 15 kDa as shown in figure 5-12B. There was a significant ($P < 0.05$, $P < 0.01$) up-regulation of SAA-1 protein expression in the progressor animals in comparison to regressor and control animals. Western blot analysis of HPX was also carried out on the same samples

and revealed a band at 50 kDa, which is the expected mass of HPX; a representative result is shown in figure 5-12C with a noticeable increase in HPX band intensity in the progressor samples although the difference is less apparent than shown for SAA-1. Western blotting for SAP revealed a band of approximately 25 kDa, as shown in figure 5-12D. There was a significant ($P < 0.05$, $P < 0.01$) up-regulation SAP protein expression in the progressor animals in comparison to regressor and control mice.

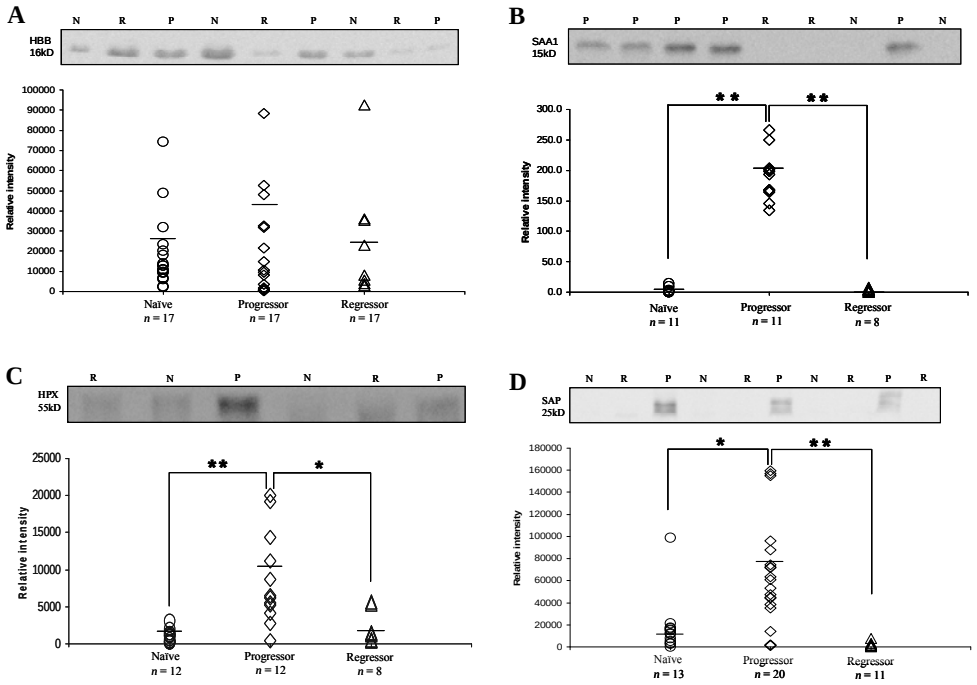


Figure 5-12. Western blot analysis of HBB, HPX, SAA-1 and SAP in mouse serum sample. A) SAP was detected at 25kDa and it was highly expressed in the serum samples from animals with a progressive tumour compared to naïve and regressors. **B**) There was an increase in the expression of HPX in the progressor animals compared to naive and regressors. **C**) Western blot of HBB in serum samples which shows variation in expression between the three groups of animals. **D)** There was a significant increase in the expression of SAA1 in the progressor animals compared to naive and regressors. Relative intensity: A QC sample was run on every single western blot experiment and the density of each sample was normalised to the density of the QC sample.

Validation of SAP using western blotting was carried out on the prospective serum samples (n = 6 progressors and n = 5 regressors) from the CT26 murine model of immunotherapy. Two sets of serum sample was collected from individual mice: first time point (FTP) collected prior to tumour implantation and end time point (ETP) collected at the end of the

tumour immunotherapy experiment when tumour regression or progression was confirmed. The FTP and ETP serum samples from individual animals were used to assess SAP concentration in the serum by western blotting. Densitometry results of SAP blots are presented in figure 5-13 and show a significant increase of SAP in the ETP of all progressor animals. There were no significant differences between the FTP and ETP for individual regressor mouse.

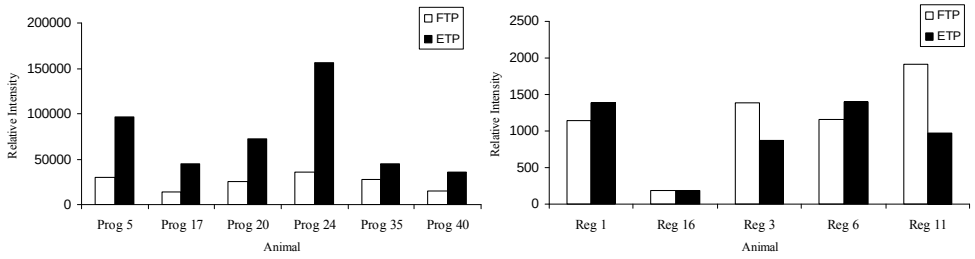


Figure 5-13. Western blot analysis of SAP on serum samples obtained from regressor/progressor animals in the prospective CT26 immunotherapy model. Serum samples from each animal were taken in the naive state (FTP) and also after success or failure of therapy (ETP). Results of SAP western blotting showed that there was a significant increase in the protein levels in the ETP compared to the FTP in animals that rejected the tumour after immunotherapy. However, the levels remained constant in the animals that responded to therapy.

These results are supported by the previously proposed suggestion that serum proteome of regressor and control (naïve) animals are similar although they are very different to proteome of mice with progressive tumours. However, it would be interesting to follow the changes in the SAP pattern throughout the period of immunotherapy experiment by collecting serum samples in different time points.

5.3.9 Quantification of SAP, SAA and HPX using ELISA and immunohistochemistry.

SAP was further quantitated using serum from individual animals from the prospective model (figure 5-14A). There was no significant difference in the concentration of SAP in the FTP serum samples of progressor and regressor animals; however SAP levels were significantly increased after failure of therapy but the concentration remained similar in the serum samples of animals that responded to therapy. The level of SAP increased by an average of 2.3 fold in the ETP compared to FTP of the progressor animals. SAP concentration was assessed at the terminal stage of therapy in the retrospective model

(figure 5-14B). While the concentration of SAP in control and regressor animals was similar, there was a significant increase in SAP in progressor serum samples, which validated the findings from proteomic based studies.

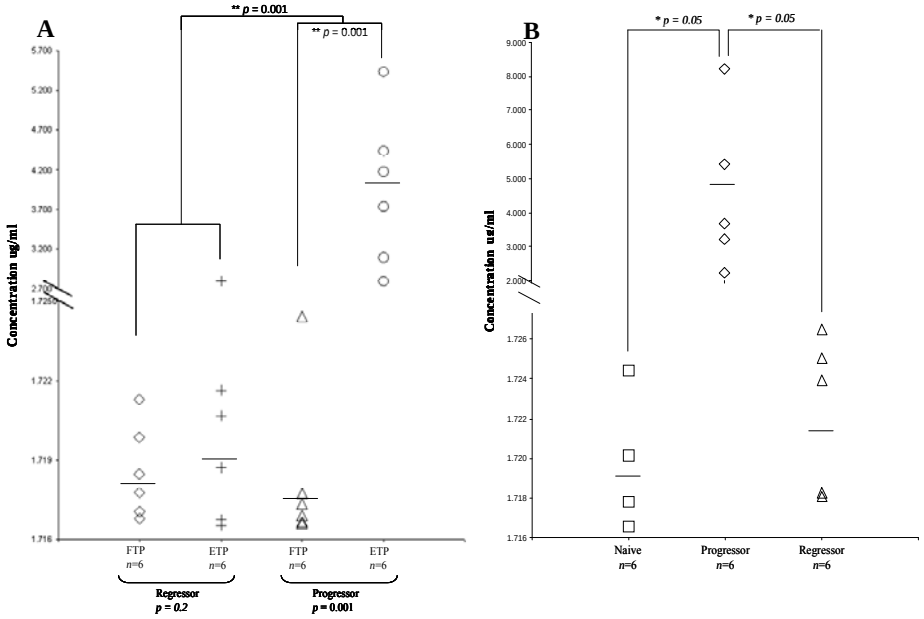


Figure 5-14. ELISA assay on SAP. A) Serum samples from each individual mouse (prospective model) from its healthy status (FTP) and after failure or response to therapy (ETP) were also assessed for SAP concentration using ELISA. There was no significant difference between the healthy statuses (FTP) of animals that did not respond to therapy in comparison to therapy responders. However, SAP levels were significantly higher ($p = 0.05$) for ETP in therapy non-responders compared with FTP of the same animal. The levels of SAP however, remained at similar levels for both time points in therapy responders. **B)** The serum sample from retrospective animal model were quantified for SAP which resulted in a significant increase ($p = 0.001$) in the SAP in the progressor animals when compared to regressor and control animals.

Since the differential expression of SAA-1 was clearly confirmed in the western analysis we further analysed the SAA protein levels in sera via ELISA. We assessed SAA concentration in the serum samples collected from the prospective study (figure 5-15A). The SAA levels in sera was quantified in the individual animals in the naive state (FTP) for a baseline level as well as at the end of tumour progression or regression (ETP) following DISC-HSV therapy (figure 5-15A). There was no significant difference in the concentration of SAA in the FTP serum samples of progressor and regressor animals. However, SAA levels were significantly ($p < 0.001$) increased in the ETP sample after failure of therapy whilst their concentration remained similar in the serum samples of animals that responded to therapy. We were further able to show that the levels of SAA increased by an average of

11.7 fold respectively in the ETP compared to FTP of the progressor animals. We further assessed SAA concentration in the serum samples collected at the end of the retrospective study (figure 5-15B). While the concentration of SAA in the control and regressor animals was similar, there was a drastic increase in the progressors which also confirmed and validated our finding from the proteomic based studies as well as western blotting.

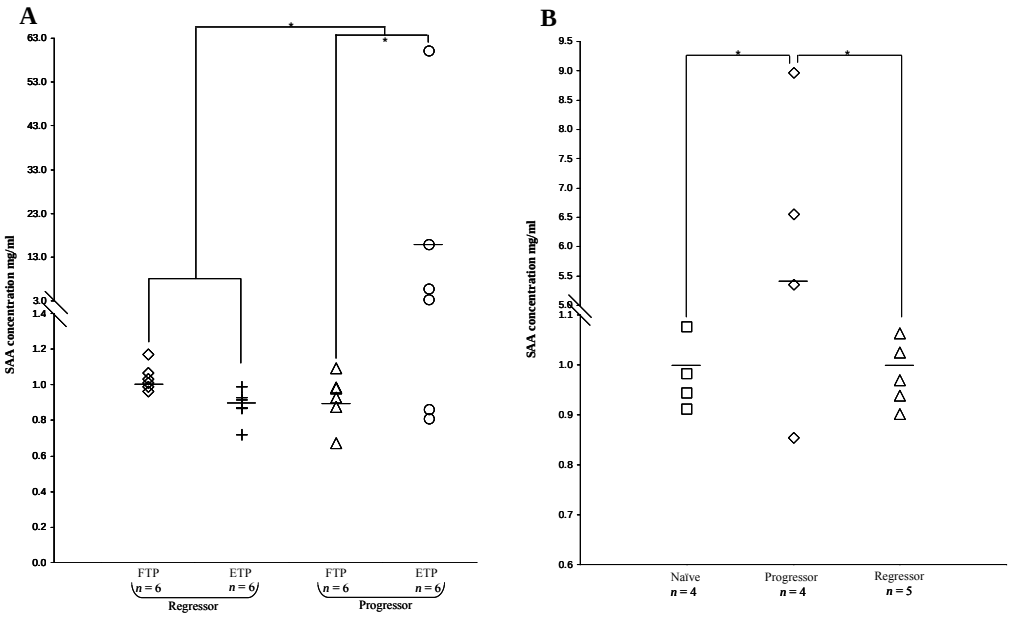


Figure 5-15. ELISA assay on SAA. A) Serum samples from each individual mouse (prospective model) from its healthy status (FTP) and after failure or response to therapy (ETP) were also assessed for SAA concentration using ELISA. There was no significant difference between the healthy statuses (FTP) of animals that did not respond to therapy in comparison to therapy responders. However, SAA levels were significantly higher (p = 0.05) for ETP in therapy non-responders compared with FTP of the same animal. The levels of SAA however, remained at similar levels for both time points in therapy responders. **B)** The serum sample from retrospective animal model were quantified for SAA which resulted in a significant increase (p = 0.001) in the SAA in the progressor animals when compared to regressor and control animals.

Finally, the quantification of serum HPX in the serum samples from FTP and ETP of regressor and progressor animals (prospective model) showed similar results to SAP and SAA (figure 5-16A). While the HPX levels in both FTP and ETP of the regressor animals remain the same, there was an average of 14.4 fold increases in the ETP samples of progressors compare to FTP progressor samples. However, HPX levels in FTP progressors show similarity to FTP and ETP regressor samples. Moreover, the HPX concentration in the serum samples collected at the end of the retrospective study from the naïve, regressor and progressor (Figure 5-16B). The concentraction of HPX in the serum sample of progressor animals was significantly higher compared to regressor and naïve animals.

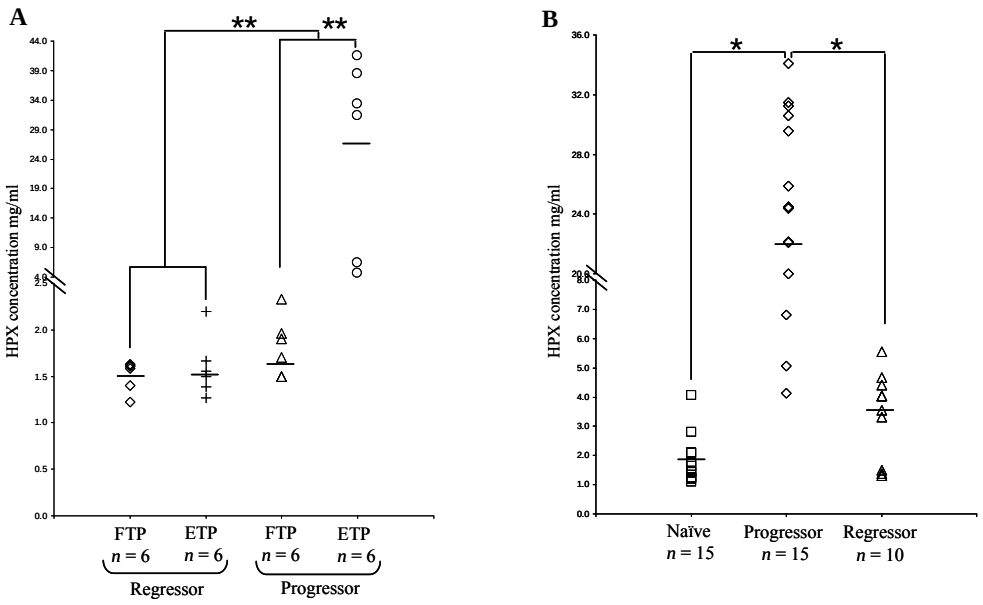


Figure 5-16. ELISA assay on HPX. A) Serum samples from each individual mouse (prospective model) from its healthy status (FTP) and after failure or response to therapy (ETP) were also assessed for HPX concentration using ELISA. There was no significant difference between the healthy statuses (FTP) of animals that did not respond to therapy in comparison to therapy responders. However, HPX levels were significantly higher ($p = 0.001$) for ETP in therapy non-responders compared with FTP of the same animal. The levels of HPX however, remained at similar levels for both time points in therapy responders. **B)** The serum sample from retrospective animal model were quantified for HPX which resulted in a significant increase ($p = 0.05$) in the HPX in the progressor animals when compared to regressor and control animals.

We examined whether SAP was present in tumour tissue by immunohistochemistry (IHC). Tumour tissue sections (CT26 colorectal carcinoma tumours) form progressor were from the prospective experiment were immunostained with polyclonal rabbit anti mouse SAP antibody (figure 5-17B1-5), or with biotinylated polyclonal goat anti-rabbit immunoglobulin (figure 5-17C). The H&E stain for the tumour tissue is shown in figure 5-17A. Low to intense SAP protein expression was evident in the cytoplasm of the CT26 tumour cells from the progressor (figure 5-17B1-5) and a summary of the intensity scores of the tissue staining for SAP is summarised in table 5-9.

.

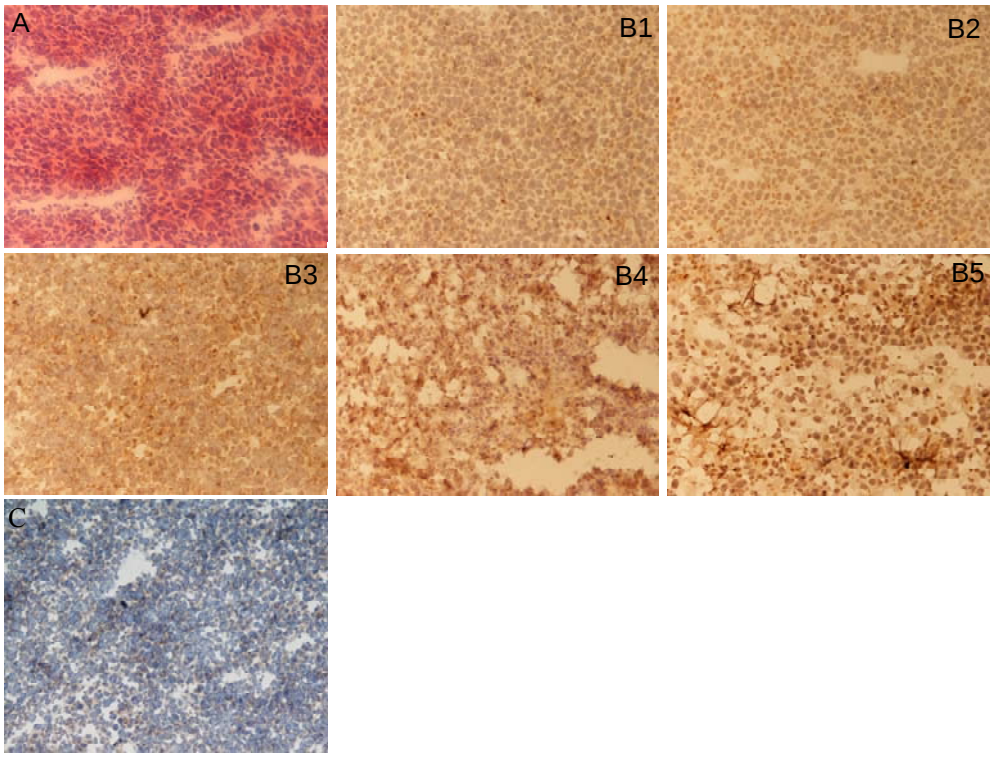


Figure 5-17. Immunohistochemistry demonstrating serum amyloid P (SAP) protein expression in CT26 colorectal tumour tissue. Tissue sections were immunostained with polyclonal rabbit anti mouse SAP antibody. The reddish-brown staining represents a positive SAP protein signal, counterstaining was shown as blue-purple. **A)** H&E stain of the tissu, **B1-B5)** Demonstrate intense SAP protein expression is evident in the cytoplasm of the tumour cells. **B1** is representative of a tissue with 1-20% positive cells therefore a score of 1 is given, **B2** = score 2 (21-40% positive cells), **B3** score 3 (41-60% positivity of cells), **B4** score 4 (61-80% positive cells) and **B5** score 5 (81-100% cells positive) for SAP protein expression. It is can be seen from the table that 16/21 tissues from progressor mice showed 20 – 60% (scores of 3 and 4) protein staining for SAP. **C)** Negative control, stained with secondary antibody only.

score	Percentage cells showing positive staining	No. tissues
0	0%	1
1	1-20%	1
2	21-40%	5
3	41-60%	11
4	61-80%	2
5	81-100%	1

Table 5-9. Summary of the IHC expression scores for SAP in tumour tissues.

5.3.10 Validation of serum biomarkers corresponding to tumour progression and regression in the CT26 tumour and liver using RT-PCR analysis

Due to positive staining for the protein expression of SAP and SAA-1 in the tumour tissues of progressor and TB animals we investigated further whether the transcript mRNA for all four discriminatory biomarkers was also present in tumour and liver using quantitative RT-PCR. Interestingly none of the tumour tissues either from TB or progressors showed a positive mRNA for any of the four biomarkers investigated. A graphic representation of the mRNA transcript levels in the liver for HBB, SAA-1, HPX and SAP are shown in figure 5-18. As can be seen from the graphs there is a significant increase in mRNA levels of SAP, SAA-1 and HPX in the liver of TB and progressor animals ($p < 0.05$, $p < 0.01$) respectively compared with both naïve and regressors. Whilst HBB mRNA expression does not show a clear pattern of expression in liver, in the spleen (Table 5-10) the site of protein production there is a significant increase of the mRNA ($p < 0.05$) in both the TB and progressors compared with naïve animals but there was not a significant difference between TB and progressors versus regressors. The other three proteins showed negligible expression in the spleen. The CT26 cell line used to induce tumour growth was also analysed as well as the kidney which demonstrated little or no mRNA presence for the four proteins (HBB, SAA-1, HPX and SAP).

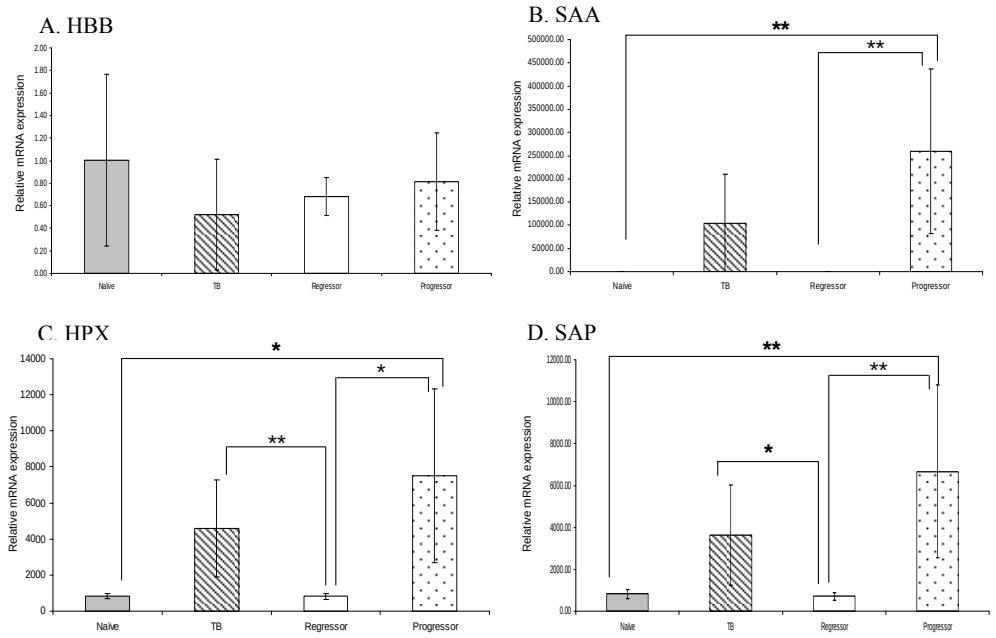


Figure 5-18. RT-PCR demonstrating expression of HBB (A), SAA (B), HPX (C) and SAP (D) genes in liver of the CT26 mouse model. RNA was extracted from the liver of mice prior to CT26 tumour implantation (Naïve), without therapy (Tumour bearer, TB) and those which received DISC-HSV immunotherapy and responded to the treatment (Regressors) or those that were therapy resistant and where the tumour progressed (Progressors). As can be seen there was a significant increase in the mRNA transcripts of SAA, HPX and SAP genes in the progressors and TB compared to both naïve and regressor animals. ∗ p = <0.05 and ∗∗ p = < 0.001

Gene	Tissue	Mice	Relative mRNA expression*
mHBB	spleen	naïve	23.17
		tumour bearer	109.06
		regressor	54.47
		progressor	109.07
	tumour	tumour bearer	0.14
		progressor	0.05
	CT26 cell line		0.00
	kidney		0.35
	liver	naïve	1*
mSAA1	spleen	naïve	0.00
		tumour bearer	0.00
		regressor	0.00
		progressor	0.00
	tumour	tumour bearer	0.00
		progressor	0.00
	CT26 cell line		0.00
	kidney		0.00
	liver	naïve	1000*
mHPX	spleen	naïve	0.17
		tumour bearer	0.03
		regressor	0.09
		progressor	0.05
	tumour	tumour bearer	0.06
		progressor	0.06
	CT26 cell line		0.02
	kidney		0.22
	liver	naïve	1000*
mSAP	spleen	naïve	0.12
		tumour bearer	0.00
		regressor	0.02
		progressor	0.00
	tumour	tumour bearer	0.00
		progressor	0.00
	CT26 cell line		0.00
	kidney		0.45
	liver	naïve	1000*

Table 5-10. RT-PCR demonstrating relative mRNA expression of HBB (A), SAA (B), HPX (C) and SAP (D) genes in spleen, tumour, CT26 cell line, kidney and liver of mice.

4.5 Discussion

Immunotherapy is rapidly developing as a potential treatment option for cancer, and previous studies demonstrated that the DISC-HSV vector encoding the mGM-CSF cytokine gene when administrated to mice with CT26 colorectal carcinoma (intra-tumour injection) induced tumour regression in a significant proportion of animals (Rees *et al*. 2002, Ali *et al*. 2002 and Ahmad *et al*. 2005). The model was subsequently used to study the immune mechanisms involved in tumour regression and to investigate pathways involve in tumour immunotherapy escape (Ahmad *et al*. 2005). The potency of DISC-HSV to infect human prostate cancer cell-lines and induce expression of reporter and therapeutic cytokine gene GM-SCF in a xenograft tumor model has also been studied (Parkinson *et al*. 2003). This demonstrated the ability of DISC-HSV to infect prostate cancer cells and express GM-CSF at significant levels. In the present study we demonstrate the ability of MALDI-MS serum profiling in conjunction with ANNs modeling to discover biomarkers that are associated with response or failure of immunotherapy in a CT26 colorectal murine immunotherapy model. We have used MALDI-MS and ANN algorithms with a comprehensive sample preparation procedure prior to mass spectrometry analysis with the aim of differentiating between tumour regressor and progressor mice by analysing the serum proteome and identifying distinguishing biomarkers.

The discovery of biomarkers that predict therapeutic response is of increasing importance in the development and application of appropriate therapies for cancer. In this study we identified and confirmed serum peptide biomarkers (SAA-1, SAP and HPX) that associated with failure of immunotherapy and disease progression in a colorectal murine model, demonstrating the ability of MALDI-TOF-MS in conjunction with ANN bioinformatic modelling to discover biomarkers. We used “retrospective” CT26 colorectal cancer cohort samples, collected at the end of the study period for the discovery/identification of the peptides reflecting the outcome of therapy and in an independent “prospective” experiment used regressor/progressor/naïve/tumour bearer samples for the biological and immuno-validation. Additionally for this sample set we were able to collect the blood prior to and following tumour implantation and therapy.

In the discovery phase of the study, “retrospective” samples collected under strict standard operating protocols at the end of therapy trial were used to derive cumulative MALDI-MS protein and tryptic peptide data which were subsequently used to generate ANN models

which compared regressor versus progressors, progressor versus healthy controls and regressors versus controls. For protein analysis, two ions (*m/z* 4804 and *m/z* 11053) were identified to serve as final discriminating biomarkers with an accuracy of 86% and sensitivity 100% and specificity of 94%. Subsequently regressor/progressor peptide analysis demonstrated an 86% accuracy (sensitivity of 90% and specificity of 81%) with a panel of 4 predicted biomarkers. In the case of progressors versus healthy controls, 1 tryptic peptide peak classified samples with an accuracy of 85% (sensitivity of 88% and specificity of 96%), however, it was not possible to discern regressors from healthy controls, suggesting that following successful therapy the serum proteome reverted to a "healthy" status. Four predictive ions from the regressor/progressor analysis were identified as peptides derived from haemoglobin beta-2 subunit (HBB), serum amyloid A-1 (SAA-1), hemopexin (HPX) and serum amyloid P (SAP) proteins, while the tryptic peptide ion discriminating progressors from healthy controls was also identified as a component of HPX. Three out of the identified proteins (SAA-1, SAP and HPX) have been previously described as acute phase reactants in mice (Pepys *et al.* 1997).

Analysis by independent techniques, RT-PCR and immunoassays, were used to validate the respective expression of these biomarkers. The results suggest a role for SAA-1 and SAP as biomarkers related to tumour progression following immunotherapy, a premise supported by a recent study in a xenograft model of Balb/c mice, which reported that differential glycosylation patterns of acute phase proteins (APP) correlates with disease progression in breast and colon cancers (Diamandis 2006). In the present study, a mass value of 1132 was identified as SAA-1 with sequence of EAFQEFFGR which is identical to the reported biomarker peptide identified in the plasma of mice bearing tumours established from human gastric cancer cell lines (Juan *et al.* 2004). Elevated levels of SAA have been reported in the plasma/serum of mice bearing intestinal tumours (Hung *et al.* 2006) and neuroblastomas (Sandoval *et al.* 2007), as well as in sera of patients with various types of human cancers including nasopharyngeal (Cho 2007), prostate (Kaneti *et al.* 1984), pancreatic (Müller *et al.* 1997).

Serum amyloid A (SAA-1) is a member of the acute phase protein family which has been recently reported, using in situ hybridisation and IHC on paraffin tissue sections from 26 colon cancer patients revealed barely detected SAA mRNA expression in normal looking colonic epithelium whilst expression was increased gradually as epithelial cells progressed through dysplasia to neoplasia (Gutfeld *et al.* 2006). Deeply invading colon carcinoma cells

showed the highest levels of SAA, although in most cases it is believed that the elevated levels of SAA-1 in the blood of cancer patients is derived from the liver (Gutfeld *et al.* 2006). Cocco *et al*., (2009) reported that SAA was highly expressed in USPC; it is actively secreted in vitro and high concentrations of SAA are present in the serum of USPC patients. Moreover; serum SAA levels in USPC patients clinically diagnosed with early-stage disease are predictive of more advanced stage disease at the time of comprehensive surgical staging. Alternatively mediators derived from the tumour may stimulate SAA-1 synthesis in the liver resulting in high SAA-1 blood levels (Ghezzi *et al.* 1993). We were unable to detect SAP mRNA transcripts in CT26 tumour tissue suggesting that the tumour/tumour microenvironment produces mediators that increase SAA-1 production in the liver. RT-PCR revealed a 1200 and 5500 fold increase in SAA-1 expression in the liver of progressor mice compared with naïve mice or mice responding to DISC immunotherapy respectively. Further to this, SAA-1 mRNA levels were increased 2.5 times in the liver of therapy resistant mice compared to tumour bearer mice (not receiving therapy). The positive protein expression of both SAA-1 and SAP in tumour cells is intriguing although the mechanism whereby they are taken up by tumour cells remains to be established. However; the increase in liver production could be due to an increase in the levels of IFN-γ in the progressor animals which our group have previously reported (Ahmad *et al.* 2005); it is known that IFN-γ, TNF-α, IL-1, IL-6 and other proinflammatory mediators induce the upregulation of acute phase proteins including SAA transcription (Steel *et al.* 1994). Up-regulation of proinflammatory cytokines and acute phase proteins in colonic mucosa has also been reported for individuals with inflammatory bowel disease (Niederau *et al.* 1997, Keshavarzian *et al.* 1999), who are at very high risk of developing colon cancer (Bachwich *et al.* 1994); epidemiologic observations also suggest that chronic inflammation predisposes to colorectal cancer (Rhodes and Campbell 2002, Farrell and Peppercorn 2002). Thus, the observation by Hao *et al*., (2005) of a down-regulation of PPAR-γ and up-regulation of IL-8 and SAA1 in the normal mucosa of individuals with a family history of sporadic colon cancer and individuals with inflammatory bowel disease is perhaps indicative of the involvement of common pathways leading to colon carcinogenesis in these two groups.

Although the biological importance of SAA is not well understood, previous reports have suggested multiple important biologic functions relevant to the mechanism of tumor cell invasion and metastasis. SAA plays a role in transportation of cholesterol to liver, recruitment of immune cells to the site of inflammation and the induction of extracellular

matrix degrading enzymes; it is possible therefore that SAA may influence tumor cell invasion through re-modelling of the extracellular matrix through the induction of matrix metalloproteinases (Migita *et al.* 1998, Hara *et al.* 2004). SAA may also play a role in p53-induced apoptosis (Polyak *et al.* 1997), the mechanism of which is not yet known. There has been a renewed interest in the concept of inflammation-associated tumorigenesis (Balkwill and Mantovani 2001, Pikarsky *et al.* 2004, Clevers 2004). It has been suggested that local inflammation within the microenvironment of malignant tissue induces migration and tissue infiltration of inflammatory cells (Badolato *et al.* 1994, Xu *et al.* 1995), associating with the production of pro-inflammatory cytokines: tumour necrosis factor-α, interleukin-1ß, and interleukin-8 (Patel *et al.* 1998, Furlaneto *et al.* 2000).

SAP is a glycoprotein that belongs to the APP family and it was thought to be uniquely present in the mouse serum (Duan *et al.* 2004). However, SAP has also been identified in atherosclerotic lesions and as a component of human amyloid deposits in the brains of Alzheimer's patients (Kolstoe *et al.* 2009, Tennent *et al.* 1995, Nishiyama *et al.* 1996) and potentially a role in protecting individuals from autoimmune disorders (Soma *et al.* 2001). Serum amyloid P is involved in the clearance of DNA by binding to it when present in the extracellular environment released from cells undergoing necrosis or apoptosis which has an impact on protecting individuals from autoimmune disorders (Kravitz *et al.* 2005). As with SAA1, cytokines such as IL-1 and IL-6 stimulate SAP production in mouse hepatoma cell lines *in vitro,* and it has been recently reported that the expression of SAP and its homologue protein CRP are regulated by IL-1 and TNF-α (Zahedi and Whitehead 1993).

In this study HPX was identified from *m/z* of 1192 with a sequence identity of NPITSVDAAFR a protein that is produced by the liver as was verified by the RT-PCR. Hemopexin is a protein produced by liver and part of APP family. It binds haem with the highest affinity among known proteins (Tolosano and Altruda 2002). The levels of HPX in blood reflect the amount of haem present and it participates in iron metabolism, protection against oxidative damage and is induced during inflammation. HPX, haptoglobin and transferrin are the most abundant protein complex in plasma after albumin, immunoglobulin and plasma proteases (Delanghe and Anglois 2001). The production of HPX is shown to be slightly increased as a response to IL-6 but not IL-1, and using MALDI-MS serum screening, it has been detected as a biomarker of human breast cancer patients with leptomeningeal metastasis (Rompp *et al.* 2007). It is interesting to note that the haem-hemopexin complex stimulates growth and proliferation of T-lymphocytes at the

site of inflammation, injury and infection (Smith *et al.* 1997), which suggests its involvement in inflammation-associated tumorigenesis, where chronic inflammation is proposed as a pre-requisite in the aetiology of neoplasia (Schottenfeld and Beebe-Dimmer 2006, Cousses and Werb 2002).

The sequence of the tryptic peptides were also identified by MS/MS analysis for the mass values of 1312 was related to haemoglobin β-2 chain subunit (HBB) and was VNPDEVGGEALGR. Haemoglobin is oxygen transporting molecules and there was an increase in its concentration during tumour progression due to angiogenesis around the tumour in proteomic analysis of melanoma tumour tissues obtained from B16-F10 murine melanoma model (Culp *et al.* 2006) however, our serum proteomic profiling of mouse with CT26 tumour after DISC-HSV therapy revealed that the average intensity of the peak identified as haemoglobin β-2 chain in regressor mice is higher compared to profiles of progressor mice. This interesting result was further suported by a recent study published by Komita and colleagues (Komita *et al.* 2008). This group reported that a $CD8^+$ T-cell response against haemoglobin β-2 prevents solid tumour growth. They used a vaccine pulsed with haemoglobin β-2 subunit to vaccinate Balb/c mice prior to tumour challenge with sarcoma, breast and CT26 colon carcinoma. After the tumour challenge, there was a significant decrease in the growth of tumour in the animals receiving the vaccination prior to the tumour challenge. This vaccination induced complete regression of sarcoma tumour in the animals vaccinated prior to tumour challenge. Their *in situ* imaging suggest that the haemoglobin β-2 peptide pulsed vaccine limit or destabilise tumour-associated vascular structures by promoting immunity against haemoglobin β-2 positive vascular pericytes. In another recent study, SELDI analysis of the sera obtained from patients with ovarian cancer demonstrated that haemoglobin alpha and beta chain are potential biomarkers for ovarian cancer and the levels of haemoglobin was significantly different in patients with ovarian cancer from healthy specimens (Woong-Shick *et al.* 2005) however the patients in this study did not receive any therapy. In concurrence to our findings, Obermair and coworkers measured the serum haemoglobin levels of patients with advance uterine cervix carcinoma before administration of radiotherapy and showed that patients with serum haemoglobin levels > 12 g/dL have better survival rate compared with patients with haemoglobin levels lower than 12 g/dL (Obermair *et al.* 1998) and several other reports also confirmed in other types of cancer and therapy (Wagner *et al.* 2000). The reasons of this phenomenon are unclear; however, secretion of IL-1, TNF-α and IFN-gamma thought to have an impact on

appearance of tumour anemia by haemolysis. Therefore, the presence of aggressive tumour colonies in patients with low levels of haemoglobin may indirectly reduce the survival rate after administration of the therapy.

Although SAA, SAP and HPX are not strictly "cancer-related" biomarkers, from the results shown here in an experimental model, these proteins hold potential value for the assessment of patients harbouring colorectal cancer as well as in the potential monitoring of therapeutic results. This premise is also inferred from a recent publication by Malle *et al.* (Malle *et al.* 2009) who suggested a critical role for SAA as a marker for monitoring disease outcome and survival prediction. There is overwhelming evidence that elevation of SAA occurs during cancer relapse (Parle-McDermott *et al.* 2000, Hao *et al.* 2005, Nishie *et al.* 2001, Kosari *et al.* 2005, Vreughenhil *et al.* 1999, Liang *et al.* 2007, Kovacevic *et al.* 2008). Future studies are required to reveal the underlying mechanisms of how SAA could promote tumour development and accelerate tumour progression and metastasis. The results presented in this report demonstrate the utility of MALDI-TOF-MS serum (proteome) profiling and ANN analysis as a powerful tool for cancer biomarker discovery.

Chapter 6 – A Time Course Evaluation of Proteomic Changes in CT26 Immunotherapy with dendritic cell vaccine in combination with blockade of VEGFR-2 and CTLA-4.

6.1 Introduction

6.1.1 Treatment of CT26 tumour with dendritic cell vaccine in combination with blockade of VEGFR-2 and CTLA-4.

Dendritic cells (DCs) represent the most potent antigen presenting cell (APC) type that are capable of producing immunomodulatory molecules (*i.e.* chemokines and cytokines) and stimulating the induction of cytotoxic lymphocytes (CTL) against tumours (Onji and Akbar 2004, Muehlbauer and Schwartzentruber 2003). In addition, they can act as their own adjuvant, a further advantage when they are utilised as cancer vaccines. Several murine studies demonstrate the rejection of established tumours using DC-based vaccines. Moreover, *in-vitro* production of large quantities of functional autologous DCs from the patient (*i.e.* DCs derived from monocyte or from $CD34^{+}$) was shown to be pretty straightforward, in the presence of GM-CSF and IL-4 in the culture media (Siena *et al.* 1995). Therefore, these unique characteristics of DCs led to great interest in the utilisation of DCs as therapeutic agents, and development of new cancer vaccines for various human cancers. Several strategies have been demonstrated for loading DCs (*ex-vivo*) with tumour derived antigens in the form of peptides, protein, tumour cell lysates or RNA/DNA , where the DCs become capable of efficiently presenting tumour associated antigen (TAA) to T cells (Dermime *et al.* 2002, Moyer *et al.* 2006). Transfection with viral vectors expressing tumour antigen, and fusing with whole tumour cells, are additional demonstrated strategies for loading DCs with antigens and generation of CTLs (Muehlbauer and Schwartzentruber 2003). The efficancy and safety of several experimentally designed DC-based vaccines, generated by loading different antigens and different antigen-loading techniques, have been further examined in multiple phase I and II human clinical trials in attempts to treat different type of cancers. The clinical trials carried out using DC-based vaccinations

includes melanoma, multiple myeloma, prostate cancer, non-Hodgkin lymphoma, renal-cell carcinoma, colorectal, liver, lung, bladder and breast cancer (Dermime *et al.* 2002, Saha *et al.* 2003). In most studies, only half or less of the patients undergoing the vaccination exhibit immune responses against the vaccinating antigen. No significant results have been obtained from phase III clinical trials on the use of DC-based vaccinations, and the results from these studies are often inconclusive.

In order to overcome the issues with the DC-based vaccinations, novel strategies have been explored, and different strategies have been developed, which include genetic alteration of DCs for production of cytokines, and combining DC vaccination with other therapies, to increase the efficacy of vaccination. The effects of DC-based vaccines can be suppressed due to secretion of immunosuppressive agents such as IL-6, IL-10 and vascular endothelial growth factor (VEGF) by tumour tissue, or because of low immunogenicity of the tumour epitope (Saha *et al.* 2003, Pedersen *et al.* 2006). These issues were addressed by Pedersen and coworkers, who investigated the efficiency of DC vaccination in combination with blockade of vascular endothelial growth factor receptor 2 (VEGFR-2) and CTL- associated antigen 4 (CTLA-4), for immunotherapy of established CT26 colorectal carcinoma tumours in murine (Pedersen *et al.* 2006). The DCs were pulsed with AH1 peptide, derived from MuL gp70 protein that is expressed by CT26 cells. The combination of this DC-based vaccination with blockade of VEGFR-2 and CTLA-4, resulted in rejection of CT26 tumours in 80% of the treated mice.

Angiogenesis plays an important role in the growth, invasion and metastasis of most solid tumours (Ferrara 2000). Among molecules that stimulate angiogenesis, VEGF and its receptor VEGF-2 are one of the key mediators of physiological and pathological vessel development. Abnormal production of VEGF by several types of tumour has previously been reported, and it has been proposed that blocking the VEGF/VEGFR-2 pathway may inhibit tumour growth (Taraboletti *et al.* 2005, Pedersen *et al.* 2006). In addition, VEGF could negatively affect the immune system's response to the growth of tumour, by decreasing the ability of DCs to differentiate to functional forms; as a result of VEGF binding to a receptor (*i.e.* VEGFR-1/FLT1 receptor) of $CD34^{+}$ bone marrow progenitor cells, as well as inhibiting T-cell development (Li *et. al*. 2006). Several studies investigated the use of VEGFR-2 blockade in conjunction with an independent immunotherapy which showed an enhancement in the efficacy of treatment (Pedersen *et al.* 2006, Li *et. al*. 2006).

CTLA-4 is a member of the immunoglobulin superfamily, which is expressed on the surface of helper T-cells and transmits an inhibitory signal to T-cells. CTLA-4 is similar to the T-cell costimulatory protein CD28, and both molecules bind to CD80 on APCs. CTLA-4 transmits an inhibitory signal to T cells, whereas CD28 transmits a stimulatory signal. Intracellular CTLA-4 is also found in regulatory T-cells and may be important to their function. T-cell activation through the T-cell receptor and CD28 leads to increased expression of CTLA-4, an inhibitory receptor for B7 molecules (i.e. CD80) (Pedersen and Ronchese 2007). It has been demonstrated that injection of anti CTLA-4 antibody, blocks the CTLA-4/CD80 interaction and enhances the antitumour immunity response, and therefore has been used in a combination therapy with other vaccination strategies to enhance the therapeutic T-cell immunity against weak immunogenic tumours (Pedersen *et al.* 2006). Moreover, human anti CTLA-4 monoclonal antibody has been entered in phase III clinical trials of melanoma and renal carcinoma (Bashyam 2007).

6.1.2 Immunotherapy of CT26 tumour with DISC-mGM-CSF vaccine.

The application of DISC-mGM-CSF vaccination in treatment of established murine CT26 colorectal tumour has been previously described (chapter 5). Briefly, direct intratumour injection of DISC-mGM-CSF-HSV into CT26 tumours has been shown to induce complete tumour rejection in upto 70% of the treated animals (regressors), while the remaining mice had progressively growing tumours (progressors) (Ahmad *et. al.* 2005). Ahmed and coworkers proposed several molecular regulatory pathways that were involved in response to therapy.

6.1.3 Importance of assessment of immune response and patient classification prior therapy administration

Emergence of novel and improved targeted therapies, including various immunotherapies, complementary to classic chemotherapy and radiotherapies, proved promising and in some cases led to improved patient survival. The efficacy of proposed immunotherapy strategies is explored initially in animal models, and the most promising strategies are ultimately tested in human clinical trials. However, in addition to development of novel therapy protocols, it is important to develop assays that qualitatively and quantitatively assess the immune response after administration that potentially can be ultimately applied for

prediction of therapy response before treatment. In this regard, it is essential to identify biomarkers associated with therapy response and this will greatly accelerate progress toward novel diagnostic and predictive tools to track early disease and tailor treatments to specific patients (Alaoui-Jamali and Xu 2006). Identification of candidate biomarkers of therapy response prediction is possible by investigating biomolecules that are involved in cellular pathways that are integral to cell function, survival, proliferation, receptor expression and immune response, but biomarkers are not necessarily tumour-derived. Several approaches explore molecular strategies for evaluation of immune responses after therapy and some of these include identification of T-cell epitopes restricted to major MHC Class I and Class II haplotypes (Moingeon 2001). Emergence of high-throughput technologies such as proteomics and genomics enables screening of thousands of proteins and genes, and detection of differences between various samples (*i.e.* serum) obtained from patients. Animal models of immunotherapy (where reasonable proportions of therapy responder and non-responder mice are achieved) provide a unique opportunity to investigate the genomic or proteomic signatures that are associated to therapy responses. In this study, MALDI-MS proteome profiling of serum samples was utilised to generate proteomic fingerprints from different stages of immunotherapy in the CT26 mouse model of colorectal carcinoma. Peptide profiling of serum by MALDI-MS in conjunction with ANN analysis has been utilised to identify differentially expressed proteins/peptides over the course of disease initiation, treatment, and therapy outcome. The ANN analysis generates a panel of biomarkers that are discriminatory between therapy responders and non-responders at different stages of treatment. These candidate biomarkers may be associated with failure/response to therapy, and potentially used for classification of responders from non-responders to a specific therapy. The strategy used for biomarker discovery proved to be promising, and capable of discrimination between responders and non-responders in the DISC-HSV immunotherapy model (chapter 5), where biomarkers corresponding to therapy response were identified and validated. Here, we used two independent murine models of immunotherapy (DISC-HSV and dendritic vaccination) for CT26 colorectal carcinoma, to investigate proteome changes in a time-dependent manner throughout the immunotherapy process.

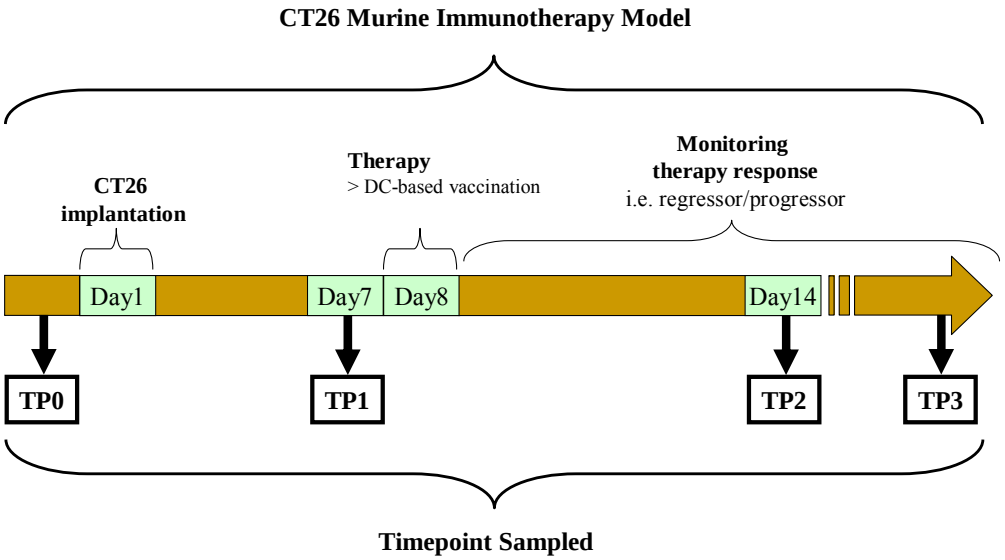
6.1.4 Aims and objectives

Detailed objectives for this part of the study are as follow:

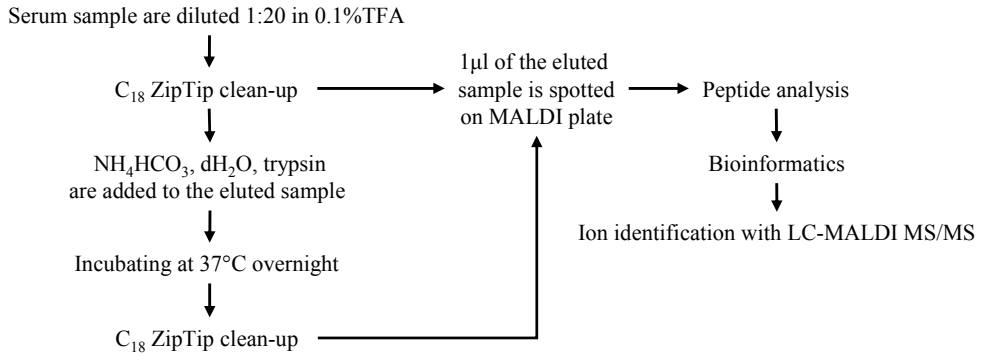
- Serum samples from a CT26 colorectal cancer model undergoing a DC-based immunotherapy was used in this stuidy in which serum samples were collected from individual animals at the naïve status, 7 day post tumour implantation, 7 days post imuunotherapy administration and once failure or response of therapy had beein confirmed.
- To obtain peptide profiles using MALDI-MS followed by ANN analysis to investigate whether is possible to obtain panel of markers that are discriminatory between the responder, non-responder and naïve mice in different time-points and may associate with outcome of therapy.
- To identify the proteins associated with the candidate panel of biomarkers.
- Validate the proteomic based results using independent methods.

6.2 Summary of methods

1) CT26 Murine Immunotherapy Model



2) Sample Preparation for Proteomic Analysis



6.3 Results

6.3.1 Time-course analysis of mouse serum proteome change following CT26 tumour therapy with dendritic cell in combination with VEGFR-2 and CTLA-4 blockades using MALDI-MS.

Using a MALDI-based protein profiling methodology, serum proteome from mice were analysed at four different time-points that included: serum samples collected pre and post tumour implantation and pre and post immunotherapy administration. The C_{18} ZipTip chromatography columns were used for deconvoluting serum samples and concentration of detectable low abundance protein/peptides by MALDI-MS. The tryptic peptide profiles of serum samples were generated using MALDI-MS and the profiles were analysed by ANN modeling. Serum samples were collected from four different time-points during a complete CT26 tumour immunotherapy with dendritic cell in combination with VEGFR-2 and CTLA-4 blockades. The first time-point (TP0) serum samples were collected at the naïve stage (one day before tumour implantation), where animals were healthy (n = 37). The second time-point (TP1) serum samples were collected 7 days post tumour implantation (n = 37). Subsequent to CT26 tumour implantation, animals were divided into two groups: tumour-bearers (n = 10) and those receiving therapy (n = 27). Following immunotherapy administration (except tumour bearer mice), serum samples were collected 7 (TP2, n = 10 tumour bearer mice and n = 27 therapy mice) and 14 (TP3, n = 10 tumour bearer mice and n = 27 therapy mice) days post treatment. From the total number of 27 animals receiving the DC-based vaccination, 21 mice responded to therapy and tumours was completely disappeared and were termed regressors. The remaining 6 mice that failed to respond to therapy were termed progressors, where tumour continued to grow aggressively.

Serum samples were subjected to C_{18} ZipTip chromatography and tryptic digestion, followed by MALDI-MS profiling of 800-3500 Da mass range in reflectron mode. Reproducible visual differences can be observed in the MALDI-TOF spectra obtained by the analysis of tryptic peptides derived from different time-point serum samples. Figure 6-1 represents some of these visual spectral differences, in the tryptic peptide spectra of TP2 and TP3 of progressor mice, although others were only detected through the use of ANNs.

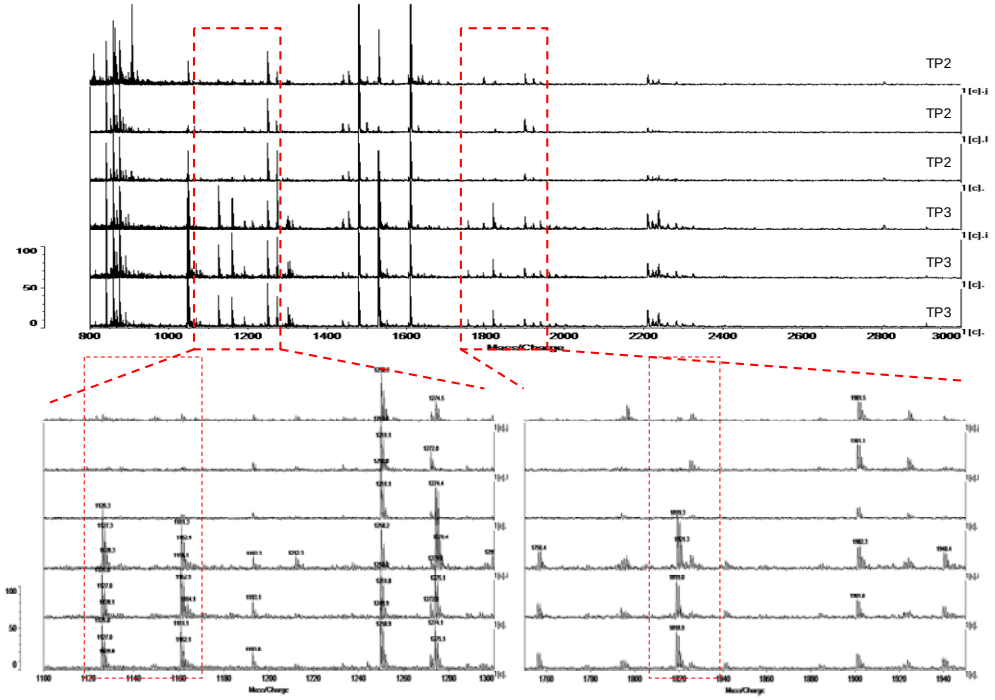


Figure 6-1. Figure showing tryptic peptide MALDI spectra obtained from TP2 and TP3 progressor mice. The X axis represents the *m/z* ratio and Y axis relative intensity. Representative MALDI-TOF mass spectra of TP2 and TP3 progressor mouse serum tryptic digest mass range of 1000-3000 Da, acquired in reflectron mode. Serum samples were spotted with CHCA on the MALDI steel target plate and allow to air dry. Peptide peaks in the region highlighted in the figure appear to be up regulated in the TP3 progressor mice in compression to TP2 progressor mouse serum.

Once the MALDI-MS tryptic peptide profiles were generated, it was necessary to visually assess the data and exclude spectra with poor quality from further analysis. In addition, the data were subjected to ANN analysis to detect discriminatory patterns between different sample groups.

6.3.2 Discovery of discriminatory candidate biomarkers for prediction of therapy response by ANN modeling.

The goal of the study as a discovery experiment was to identify candidate biomarkers that may predict the outcome of DC-based vaccination with blockade of VEGFR-2 and CTLA-4 therapy. To identify serum biomarkers we used MALDI-MS profiling in conjunction with ANN analysis. Samples were classified into two groups: progressor/tumour-bearers and regressors. The analysis aimed to compare the serum tryptic peptide patterns from each of the four time-points from the regressor mice to the associated time-point of progressor/tumour-bearer group. This may reveal discriminatory proteomic patterns between the regressor and progressor/tumour-bearers that may associate with therapy response at a specific time-point. In addition, this analysis may indicate the time point that is the most discriminatory between the two groups and therefore the best stage to assess patients for prediction of therapy outcome. Four separate ANN models were developed that included: TP0 of regressors versus TP0 of progressors/tumour-bearers, TP1 of regressors versus TP1 of progressors/tumour-bearers, TP2 of regressors versus TP2 of progressors/tumour-bearers and TP3 of regressors versus TP3 of progressors/tumour-bearers. The following sections discuss the results of the ANN models.

6.3.2.1 Discriminatory serum tryptic peptide biomarkers identified in the TP0 regressors vs. TP0 progressors/tumour-bearers model

Initially, the ANN modeling was carried out by comparing TP0 of progressor/tumour-bearer versus TP0 of regressors. Although the TP0 indicates the naïve status of the animals, this model was generated to investigate whether ANN modeling of tryptic peptide MALDI profiles is capable of revealing difference in the serum proteome of regressor and progressors that may be involved in failure or success in therapy. Hence, a stepwise approach was utilised which produced the minimum number of ions with the potential of assigning samples correctly into their respective groups based on the MALDI–MS peptide profiles (mass range of 800-3500 Da). A total number of 16 regressor samples from TP0 and 14 progressor/tumour-bearer samples (n = 4 progressors and n = 10 tumour-bearers) were used in the stepwise analysis. The stepwise approach is based on training a number of different ANN models and the general principles of stepwise approach have been described previously in details in chapter 3. In order to generate ANN models from MALDI tryptic

peptide profiles of TP0 of regressor, progressor/tumour-bearer mice, the stepwise approach involves training a number of models using 2700 inputs for each sample (each input corresponds to intensity at a specific *m/z* value). Initially each input is used as single input to train 50 models and for each model a total of 30 cases (n = 16 TP0 regressors and n = 14 progressors/tumour-bearers) were randomly split into 18 samples for training (60%), 6 for testing (20%) and 6 as a validation set (20%); this allows random cross validation analysis to be performed. After training 50 models, the best performing model was selected for further analysis and all the remaining inputs were added sequentially to the first input to train the model. This procedure was continued until no further improvement in the model was observed. In the case of modeling the tryptic peptide profiles from the TP0 of regressor, progressor/tumour-bearer mice, total of 8 steps were carried out. The results of the 8 steps are shown in table 6-1 that presents the median accuracies and mean squared errors of training, test and validation for each best performed model in each of the 8 steps. The highest accuracy that predicted progressors from regressors was generated in step 2 (highlighted in red in table 6-1), using 2 peaks with *m/z* values of 2750 and 2877 with an accuracy of 83%, a sensitivity of 85.7% and specificity of 75%. The addition of further ions did not improve the prediction of the model and therefore the top 2ions were considered biomarkers that can best predict TP0 regressors from TP0 progressors/tumour-bearers based on MALDI tryptic peptide profiles.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training error	Test Error	Validation Error
1	2750	72%	67%	67%	0.20	0.20	0.23
2	2877	83%	83%	**83%**	0.14	**0.14**	0.17
3	1577	89%	100%	83%	0.08	0.08	0.13
4	3047	94%	100%	100%	0.07	0.06	0.11
5	1912	94%	100%	83%	0.07	0.06	0.10
6	1033	92%	100%	83%	0.07	0.05	0.10
7	877	94%	100%	83%	0.04	0.04	0.10
8	858	94%	100%	83%	0.04	0.04	0.09

Table 6-1. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum tryptic peptides of TP0 regressors and TP0 progressor/tumour-bearer mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in step 2 highlighted in red, with median accuracy of 83% and mean squared value of 0.14 and further addition of ions did not affect the accuracy of model.

In order to assess how individual samples were classified by the 2 ion ANNs model, a population chart was generated. Figure 6-2 shows the population distribution of the predicted outputs for all 30 mice using the top 2 ANN model *m/z* values. Samples from TP0 progressor/tumour-bearer animals are highlighted in red (assigned as number 2) and samples from TP0 regressor animals are shown in green (assigned as number 1). The ratios below 1.5 were assigned as TP0 progressors/tumour-bearer whilst a ratio above 1.5 was used to classify TP0 regressors. The figure shows that 2 TP0 progressors/tumour-bearer and 4 TP0 regressors were incorrectly classified.

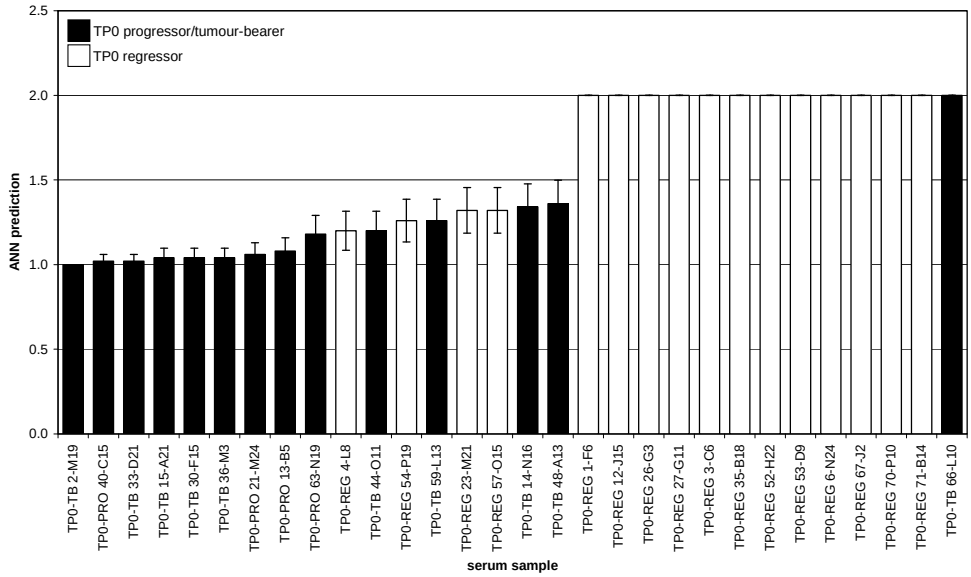


Figure 6-2. Predictive capability of ANNs to recognise MALDI serum tryptic peptide profiles based on a 2 ion ANNs model. The black bars indicate TP0 progressor/tumour-bearer samples, and the white bars indicate TP0 regressor samples. A predictive value below 1.5 indicates a TP1 progressor/tumour-bearer sample, while a prediction greater than 1.5 indicates a TP0 regressor sample.

Although majority of the sample have been correctly assigned to their associated groups, considerable number of samples have been misclassified and the 2 ion model has a low specificity (75%) which increases the chance of false positives.

6.3.2.2 Discriminatory serum tryptic peptide biomarkers identified in the TP1 regressors vs. TP1 progressors/tumour-bearers model

The second ANN modeling was carried out on the serum tryptic peptide profiles of TP1 regressors and TP1 progressors/tumour-bearers. Considering that time-point 1 is 7 days post CT26 tumour implantation, the hypothesis for the modeling here is to investigate whether the proteome of an animal that fails to respond to therapy is different to a therapy responder at this time-point. In addition, it may reveal that if the host reaction to the presence of tumour in early days plays a role in failure or response to therapy and whether these reactions are reflected in the serum and detected by MALDI profiling of serum. The ANN model was generated using 30 serum spectra (n = 17 TP1 regressors and n = 13 progressors/tumour-bearers) and stepwise analysis was carried out as described before for a total number of 7 steps and the results are presented in table 6-2. The highest accuracy that predicted TP1 progressors/tumour-bearers form TP1 regressors was obtained in the second step (highlighted in red in table 6-2), using a peaks with *m/z* values of 2404 and 1373 with an accuracy of 90%, a sensitivity of 84%, specificity of 100% and mean error of 0.09. The addition of further ions did not improve the prediction of the model and although an accuracy of 100% is obtained at step 7 this is more likely to be due to a phenomenon termed over-fitting of data. Therefore the top 2 ions were considered to be the biomarker that can best predict TP1 regressors from TP1 progressors/tumour-bearers based on MALDI tryptic peptide profiles.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training error	Test error	Validation Error
1	2404	83%	83%	83%	0.16	0.17	0.17
2	1373	89%	100%	**90%**	0.09	**0.09**	0.11
3	2987	89%	100%	83%	0.08	0.05	0.11
4	1155	89%	100%	83%	0.08	0.06	0.11
5	1997	92%	100%	83%	0.08	0.05	0.11
6	1057	94%	100%	83%	0.07	0.05	0.13
7	1705	100%	100%	100%	0.03	0.05	0.07

Table 6-2. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum tryptic peptides of TP1 regressors and TP1 progressor/tumour-bearer mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in step 2, highlighted in red, with median accuracy of 90% and mean squared value of 0.09 and further addition of ions did not affect the accuracy of model.

The top discriminatory ion between TP1 regressors and TP1 progressor/tumour-bearer serum tryptic peptide fingerprints are presented in figure 6-3. The intensity of *m/z* value of 2404 in slightly higher in the tryptic peptide profiles of TP1 progressors and tumour-bearer animals in comparison to TP1 of the regressors.

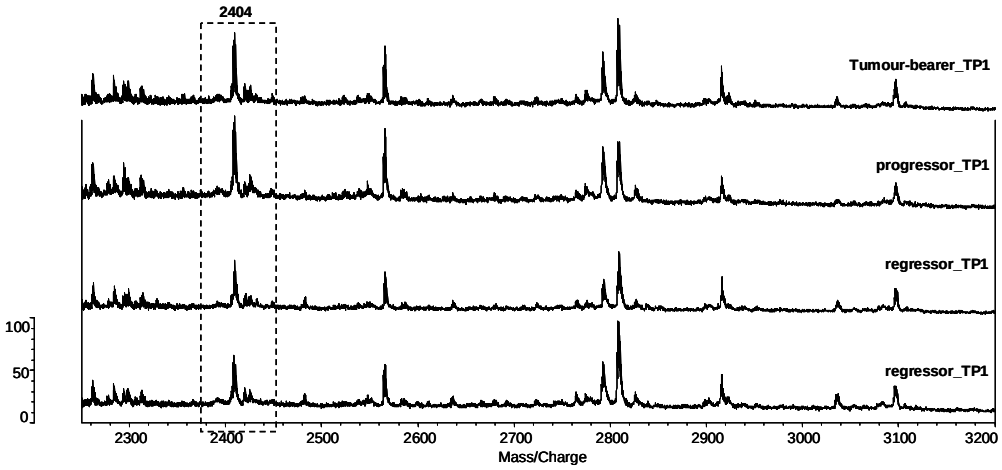


Figure 6-3. Figure showing visual spectral differences for the top discriminatory ion between TP1 regressor and TP1 progressors/tumour-bearers based on MALDI tryptic peptide profiles and ANNs analysis. The *m/z* value of 2404 was the top ion of the panel of two discriminatory ions for TP1 regressor and TP1 progressors/tumour-bearers based on the tryptic peptide profiles. The 2404 peak has slightly a higher intensity in the TP1 progressor/tumour-bearer profiles and less relative intensity in the TP1 regressor profiles.

The classification capacity of the model was further tested on individual animals by examining the population chart as shown in figure 6-4. The ratios below 1.5 were assigned to the TP1 progressors/tumour-bearers and ratios above 1.5 were classified as TP1 regressors. Based on the 2 ion model, all the TP1 regressor samples were correctly classified however, 2 of the TP1 progressors/tumour-bearers samples were misclassified.

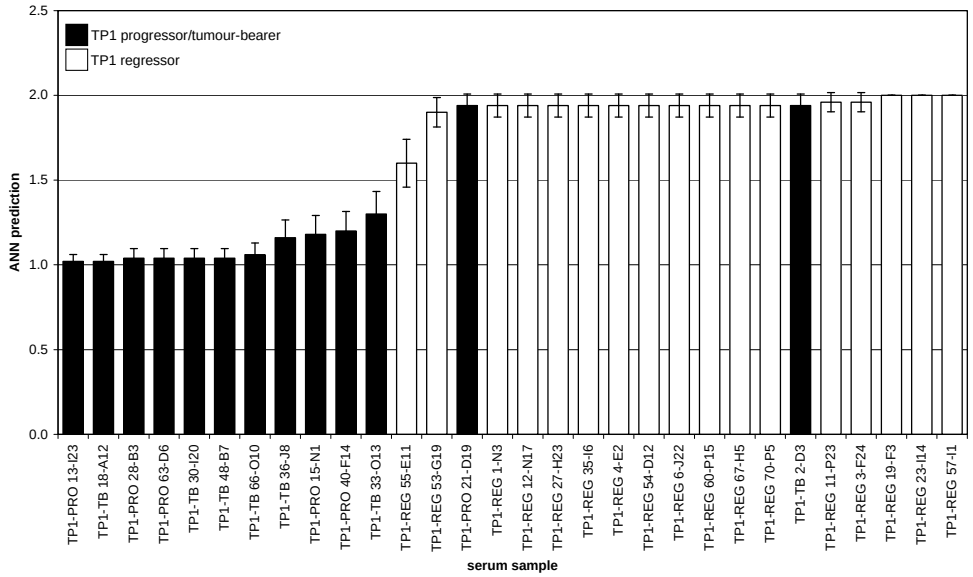


Figure 6-4. Predictive capability of ANNs to recognise MALDI serum tryptic peptide profiles based on a 2 ion ANNs model. The black bars indicate TP1 progressor/tumour-bearer samples, and the white bars indicate TP1 regressor samples. A predictive value below 1.5 indicates a TP1 progressor/tumour-bearer sample, while a prediction greater than 1.5 indicates a TP1 regressor sample.

This model shows that at the TP1 (7 days post tumour implantation) serum proteome of regressor and progressor mice are more defined and differentiated. In addition, these markers can be either therapy response predictors or early markers of tumour initiation; therefore more investigations are required to test this hypothesis.

6.3.2.3 Discriminatory serum tryptic peptide biomarkers identified in the TP2 regressors vs. TP2 progressors/tumour-bearers model

The third step for identification of the candidate therapy response biomarkers was to compare TP2 of regressors and progressors/tumour-bearers serum samples. Time-point 2 is 7 days after administration of the DC vaccination for treatment of the CT26 tumours and possible differences between the therapy responders and non-responders has been investigated at this stage. The MALDI spectra of serum tryptic peptides of TP2 progressor, tumour-bearer and progressor mice were visually checked (according to the criteria previously described) for acceptance of the spectra. A total of 18 TP2 regressors and 13 TP2 progressors/tumour-bearers (n = 6 progressors and n = 7 tumour-bearers) tryptic peptide profiles were accepted and used for bioinformatic analysis. Stepwise analysis was

performed on these profiles to assess the ability of ANNs to classify tryptic peptide serum profiles of TP2 regressor and progressor/tumour-bearer mice serum. At step 3 of stepwise analysis an accuracy of 100% was achieved and therefore no additional steps were carried out and the stepwise analysis results are illustrated in table 6-3. The best ANN prediction was achieved with 3 peptide peaks (*m/z* 1795, 1816 and 963) that discriminated between the TP1 regressor and progressor/tumour-bearer mice serum samples with an accuracy of 100%, sensitivity of 100% and specificity of 100% and the mean test error of 0.07.

Input	***m/z***	**Training Perf.**	**Test Perf.**	**Validation Perf.**	**Training Error**	**Test error**	**Validation Error**
1	1795	79%	83%	67%	0.18	0.16	0.22
2	1816	89%	100%	83%	0.11	0.10	0.14
3	963	95%	100%	**100%**	0.07	**0.07**	0.09

Table 6-3. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum tryptic peptides of TP2 regressor and TP2 progressor/tumour-bearer mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in the third step, highlighted in red, with median accuracy of 100% and mean squared value of 0.07 and further addition of ions did not affect the accuracy of model.

The MALDI tryptic peptide spectra differences between the TP2 regressor and progressor/tumour-bearer profiles for the top 2 ions identified by ANN analysis are shown in figure 6-5. Both the *m/z* values of 1795 and 1816 are present in the TP2 regressor profiles and not detected in the TP2 progressor/tumour-bearer peptide profiles. At this stage which is 7 days after therapy administration, more obvious differences are detected between the profiles of responder and non-responder mice.

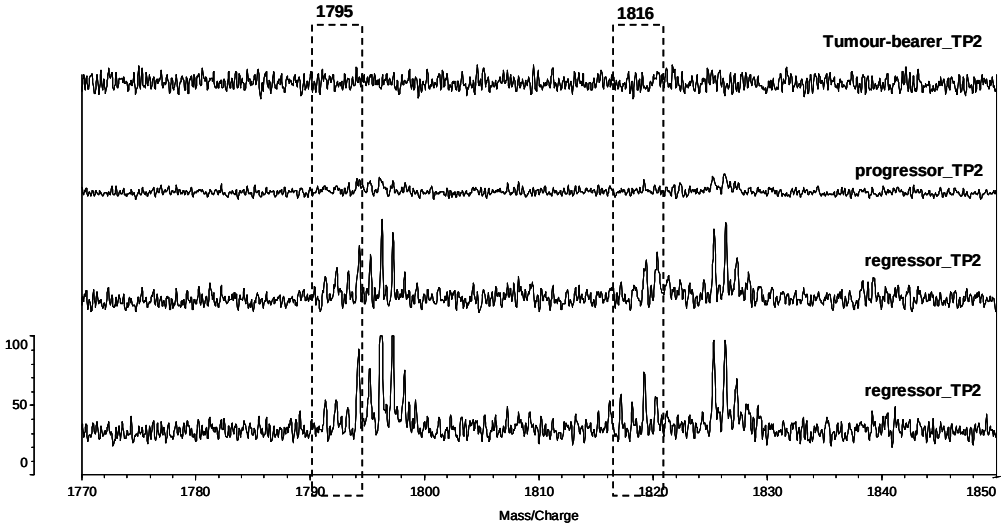


Figure 6-5. Figure showing visual spectral differences for the top discriminatory ions between TP2 regressor and TP2 progressors/tumour-bearers based on MALDI tryptic peptide profiles and ANNs analysis. The *m/z* values of 1795 and 1816 were two of the panel of discriminatory ions for regressor and progressors based on the tryptic peptide profiles. Both of these peaks are present in the regressors and absent in the progressor and tumour-bearer profiles.

The ANN's ability to classify TP2 regressors and TP2 regressor and progressor/tumour-bearer was examined by generating a population chart and the results are shown in figure 6-6. ANN analysis classified 100% of the samples correctly for both TP2 regressor and TP2 progressor/tumour-bearer mice. These results indicate that the discrimination of serum samples from responders and nonresponders mice at this stage is the best and the proteome of the two groups are highly different from each other.

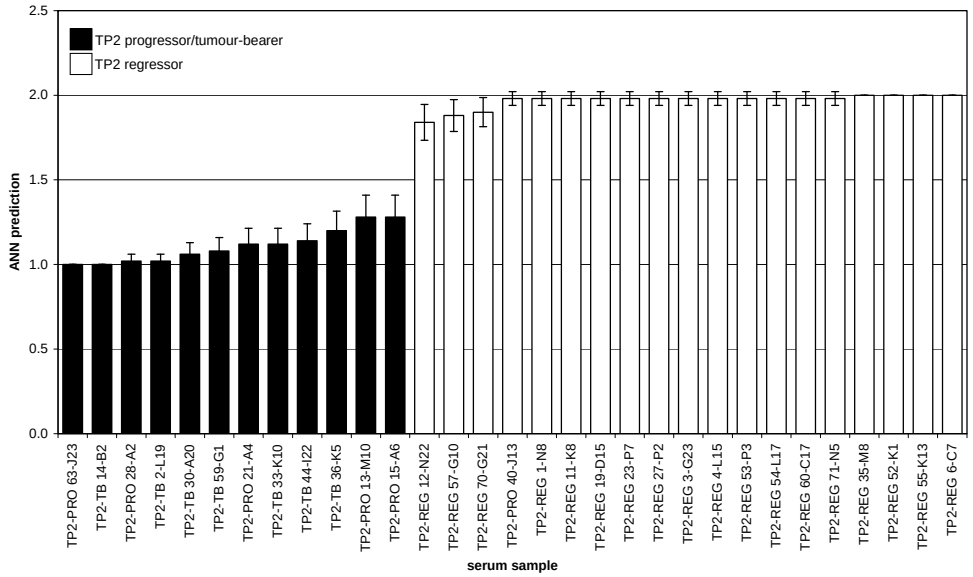


Figure 6-6. Predictive capability of ANNs to recognise MALDI serum tryptic peptide profiles based on a 3 ion ANNs model. The black bars indicate TP2 progressor/tumour-bearer samples, and the white bars indicate TP2 regressor samples. A predictive value below 1.5 indicates a TP2 progressor/tumour-bearer sample, while a prediction greater than 1.5 indicates a TP2 regressor sample.

6.3.2.4 Discriminatory serum tryptic peptide biomarkers identified in the TP3 regressors vs. TP3 progressors/tumour-bearers model

The TP3 serum samples have been collected at the end of the immunotherapy procedure, where the failure or response to therapy was confirmed. After visual inspection 11 TP3 regressors and 16 TP3 progressor/tumour-bearer (n = 6 progressors and n = 10 tumour-bearer) spectra were subjected to stepwise analysis (as described previously). Eight steps were carried out and the best accuracy (100%) was obtained at step 4. The model sensitivity and specificity were 100%. The results demonstrate that the tryptic peptide profiles from the TP3 regressor and progressors are the most different compared to the other time points and the top 4 ions are selected as the best predictors.

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training Error	Test error	Validation Error
1	855	81%	80%	83%	0.17	0.17	0.18
2	3013	88%	80%	88%	0.09	0.09	0.14
3	2631	100%	100%	92%	0.05	0.06	0.10
4	963	100%	100%	**100%**	0.04	**0.05**	0.07
5	942	100%	100%	92%	0.04	0.04	0.09
6	2764	100%	100%	100%	0.04	0.04	0.09
7	1496	100%	100%	100%	0.04	0.04	0.07
8	3387	100%	100%	83%	0.04	0.04	0.09

Table 6-4. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum tryptic peptides of TP3 regressor and TP3 progressor/tumour-bearer mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. The highest accuracy was achieved in the third step, highlighted in red, with median accuracy of 92% and mean squared value of 0.05 and further addition of ions did not affect the accuracy of model.

A population chart was generated to test the predictive 3 ion model from TP3 regressors and TP3 progressors/tumour-bearers for which the results are shown in figure 6-7. ANN analysis classified 100% of the TP3 progressor/tumour-bearer samples correctly whereas 2 of the TP3 regressor samples were misclassified. These results show that the 3 ion ANN model generated for this stage of the experiment is highly specific for classification of progressors from the regressors since most of the progressors have been correctly assigned during the 50 models and therefore have no or small variation indicated in figure 6-7 with lower error bars for progressor/tumour-bearer samples. However, the regressors have been generally classified correctly but they have higher error bars which indicate that this panel of markers is less specific and sensitive for detection of regressors from progressors.

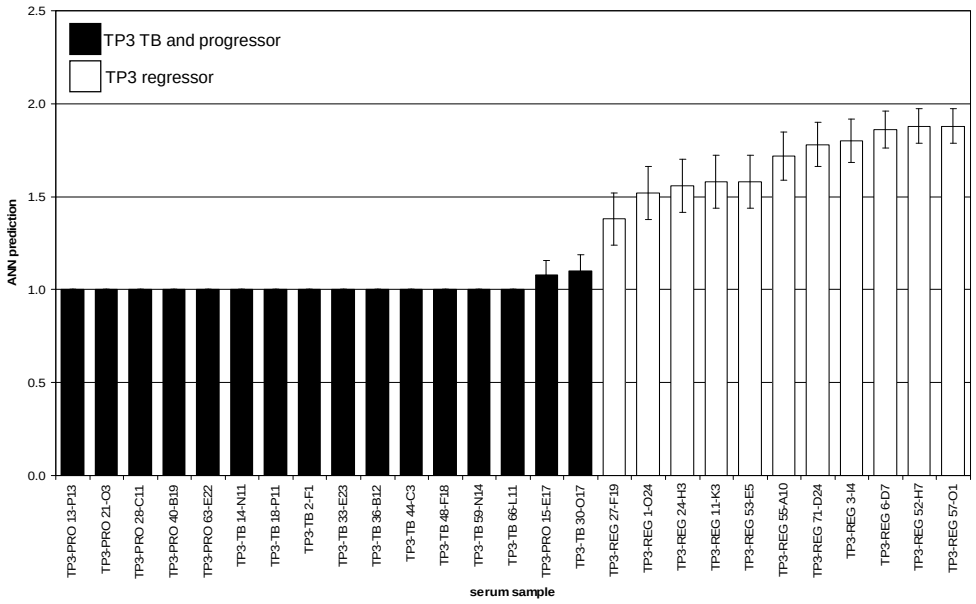


Figure 6-7. Predictive capability of ANNs to recognise MALDI serum tryptic peptide profiles based on a 4 ion ANNs model. The black bars indicate TP3 progressor/tumour-bearer samples, and the white bars indicate TP3 regressor samples. A predictive value below 1.5 indicates a TP3 progressor/tumour-bearer sample, while a prediction greater than 1.5 indicates a TP3 regressor sample.

6.3.3 Time-course analysis of proteome change and identification of biomarkers of regression by ANN modeling.

In order to define the proteome change over a time-course of tumour initiation, therapy administration and response, serum samples from therapy responder mice (regressors) at 4 different time point were subjected to MALDI-MS tryptic peptide profiling, followed by interrogation of MS data by ANN modeling. This analysis investigates discriminatory markers in regressor animals at different stages of disease initiation and response to therapy. Through comparing these results to the same analysis but in progressor animals, the specificity of the markers for each stage of tumour initiation, immunotherapy and response to therapy for the regressors and progressors may be revealed. Sample processing prior to MALDI analysis and ANN modeling parameterisation was carried out as described before. The MALDI serum tryptic peptide profiles (mass range of 800-3500 Da) of regressor mice from four different time points (n = 16 TP0, n = 17 TP1, n = 18 TP2 and n = 11 TP3) were processed for ANN modeling using 2700 inputs for each sample (each

input corresponds to intensity at a specific *m/z* value). The ANN models generated for this analysis were: TP0 versus TP1, TP1 versus TP2 and TP2 versus TP3 of regressor mice.

Results of stepwise analysis for each of the three generated ANN model is presented in table 6-5. The stepwise analysis was carried out on total of 2, 2 and 3 steps for the TP0/TP1, TP1/TP2 and TP2/TP3 models respectively. Table 6-5 presents the median accuracies and mean squared errors of training, test and validation for each best performed model in each of the three models. Stepwise analysis on the serum tryptic peptide profiles of TP0 and TP1 regressors resulted in a 2 ion model with accuracy of 77%, sensitivity of 75% and specificity of 88.2% (table 6-5 A). Similarly, stepwise analysis on the serum tryptic peptide profiles of TP1 and TP2 regressors also resulted in a 2 ion model with accuracy of 83%, sensitivity of 77.8% and specificity of 84.4% (table 6-5 B). Finally, ANN modeling of tryptic peptide serum samples from TP2/TP3 regressors resulted in a 3 ion model with accuracy of 85%, sensitivity of 100% and specificity of 90% (table 6-5 C). None of the *m/z* values except *m/z* 2878, predicted by ANN modeling in this section, have been shown in the ANN generated models in the previous section and therefore these ions may be uniquely associated with prediction of regression in a CT26 colorectal cancer immunotherapy model.

A

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training Error	Test error	Validation Error
1	1901	70%	73%	64%	0.45	0.40	0.49
2	1435	83%	84%	**77%**	0.35	**0.34**	0.43

B

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training Error	Test error	Validation Error
1	1276	80%	81%	80%	0.40	0.36	0.43
2	1985	83%	87%	**83%**	0.40	**0.26**	0.38

C

Input	*m/z*	Training Perf.	Test Perf.	Validation Perf.	Training Error	Test error	Validation Error
1	1127	70%	75%	66%	0.42	0.40	0.46
2	2878	88%	85%	79%	0.29	0.28	0.36
3	2298	88%	93%	**85%**	0.23	**0.16**	0.21

Table 6-5. The table represents the data obtained from the stepwise analysis of MALDI data generated from serum tryptic peptides of TP0, TP1, TP2 and TP3 regressor mice. The table shows a summary of the median accuracies and the mean squared error for the training, test and validation data sets as each input is added to the model. **A)** Results of the TP0/TP1 regressor model where the highest accuracy was achieved in the second step, highlighted in red, with median accuracy of 77% and mean squared error value of 0.35. **B)** Results of the TP1/TP2 regressor model where the highest accuracy was achieved in the second step, highlighted in red, with median accuracy of 83% and mean squared error value of 0.26. **C)** Results of the TP2/TP3 regressor model where the highest accuracy was achieved in the third step, highlighted in red, with median accuracy of 85% and mean squared error value of 0.16.

The MALDI tryptic peptide spectral differences between the different time-point serum samples of regressor profiles for some of the ANN identified ions are shown in figure 6-8. The *m/z* value of 1435 from the TP0/TP1 model is higher in the TP1 regressors compared to TP0 regressor. The *m/z* values of 1276 and 1985 from the TP1/TP2 regressor model are both significantly higher in the TP2 tryptic peptide profiles. Finally, the peak with *m/z* of 1127, from the TP2/TP3 regressor model is present in TP2 and TP3 and absent in TP0 and TP1, however is significantly higher in TP3 regressors.

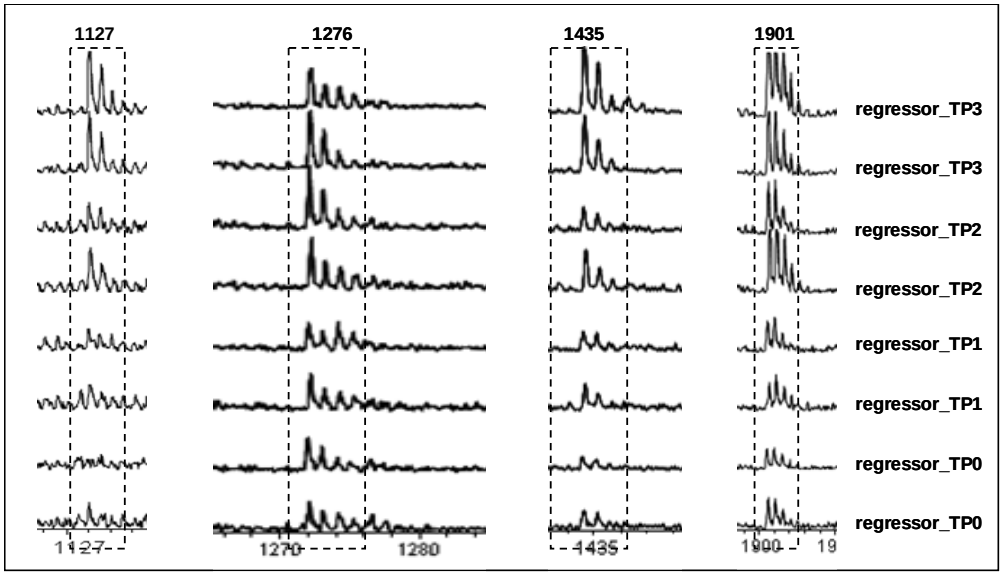
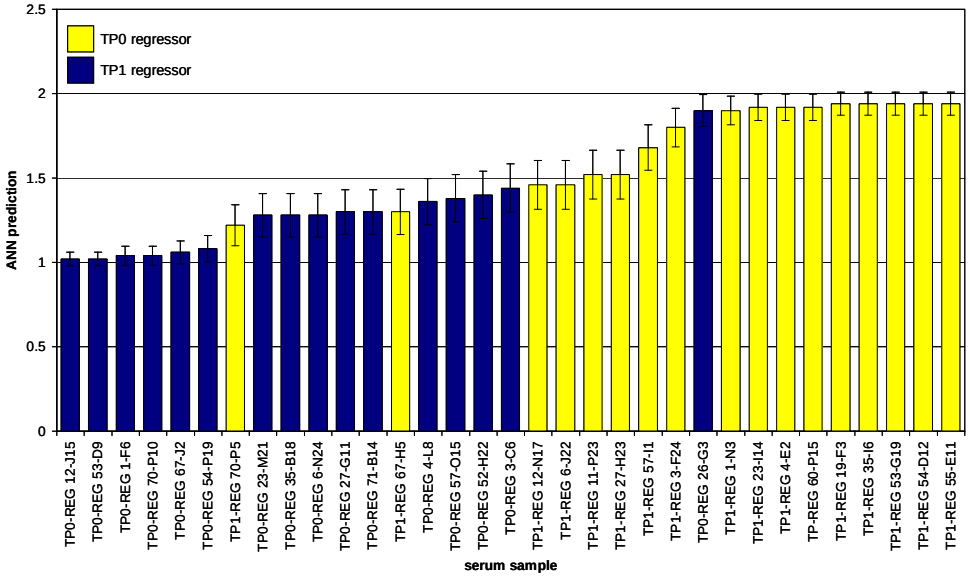


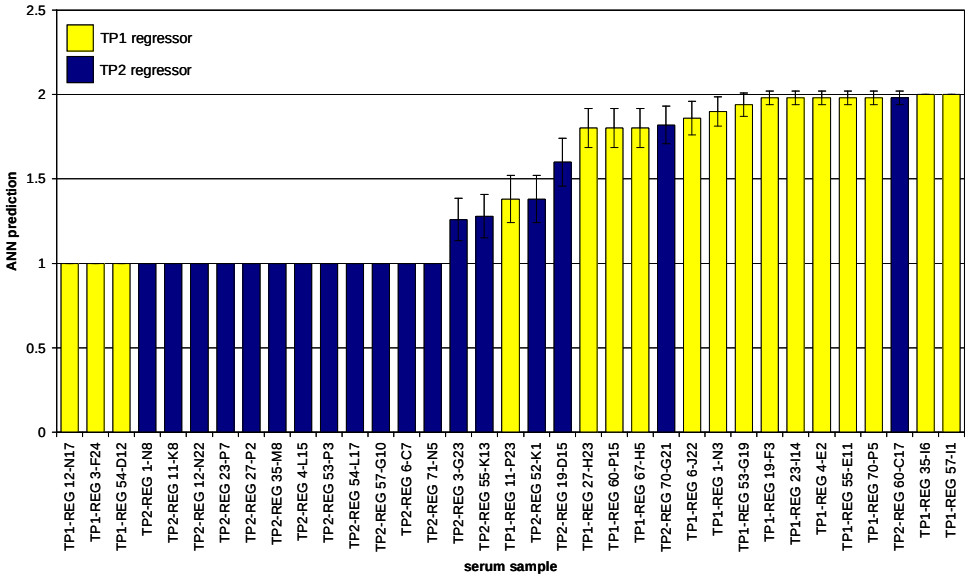
Figure 6-8. Figure showing some of the visual spectral differences for the top discriminatory ions from the TP0/TP1, TP1/TP2 and TP2/TP3 regressor's models, based on MALDI tryptic peptide profiles and ANNs analysis. The *m/z* value of 1435 from the TP0/TP1 model is higher in the TP1 regressors compare from TP0 regressor. The *m/z* values of 1276 and 1985 from the TP1/TP2 regressor model are both significantly higher in the TP2 tryptic peptide profiles. Finally, the peak with *m/z* of 1127, from the TP2/TP3 regressor model is present in TP2 and TP3 and absent in TP0 and TP1, however is significantly higher in TP3 regressors.

The individual regressor samples for each time point were plotted as population charts to examine the predictive capability of the ANN models (figures 6-9A-C). The first ANN analysis was a 2 ion model of TP0/TP1 regressors that correctly classify 77% of the population (figure 6-9 A). The accuracy of this model is not significantly high, indicator of less discrimination between the TP0 and TP1 tryptic peptide serum samples of regressor mice. Moreover, figure 6-9 B is the population prediction of TP1/TP2 regressor model, using the top 2 discriminatory tryptic peptide ions. The dark blue bars represent TP2 and the yellow bars are representative of TP1. Sample with predictive values of 1.5 or below are TP2 and those with values above 1.5 are TP1. In this model 2 TP2 regressors and 3 TP1 regressor samples were misclassified. Finally, the last analysis was the TP2/TP3 regressors and the population prediction using the top 3 discriminatory tryptic peptide ions are shown in figure 6-9 C. The dark blue bars represent TP2 and the yellow bars are representative of TP3. Sample with predictive values of 1.5 or below are TP2 and those with values above 1.5 are TP3.

A


TP0 regressor
TP1 regressor
ANN prediction
2.5
2
1.5
1
0.5
0
TP0-REG 12-J15
TP0-REG 53-D9
TP0-REG 1-F6
TP0-REG 70-P10
TP0-REG 67-J2
TP0-REG 54-P19
TP1-REG 70-P5
TP0-REG 23-M21
TP0-REG 35-B18
TP0-REG 6-N24
TP0-REG 27-G11
TP0-REG 71-B14
TP1-REG 67-H5
TP0-REG 4-L8
TP0-REG 57-O15
TP0-REG 52-H22
TP0-REG 3-C6
TP1-REG 12-N17
TP1-REG 6-J22
TP1-REG 11-P23
TP1-REG 27-H23
TP1-REG 57-I1
TP1-REG 3-F24
TP0-REG 26-G3
TP1-REG 1-N3
TP1-REG 23-I14
TP1-REG 4-E2
TP-REG 60-P15
TP1-REG 19-F3
TP1-REG 35-I6
TP1-REG 53-G19
TP1-REG 54-D12
TP1-REG 55-E11
serum sample

B


TP1 regressor
TP2 regressor
ANN prediction
2.5
2
1.5
1
0.5
0
TP1-REG 12-N17
TP1-REG 3-F24
TP1-REG 54-D12
TP2-REG 1-N8
TP2-REG 11-K8
TP2-REG 12-N22
TP2-REG 23-P7
TP2-REG 27-P2
TP2-REG 35-M8
TP2-REG 4-L15
TP2-REG 53-P3
TP2-REG 54-L17
TP2-REG 57-G10
TP2-REG 6-C7
TP2-REG 71-N5
TP2-REG 3-G23
TP2-REG 55-K13
TP1-REG 11-P23
TP2-REG 52-K1
TP2-REG 19-D15
TP1-REG 27-H23
TP1-REG 60-P15
TP1-REG 67-H5
TP2-REG 70-G21
TP1-REG 6-J22
TP1-REG 1-N3
TP1-REG 53-G19
TP1-REG 19-F3
TP1-REG 23-I14
TP1-REG 4-E2
TP1-REG 55-E11
TP1-REG 70-P5
TP2-REG 60-C17
TP1-REG 35-I6
TP1-REG 57-I1
serum sample

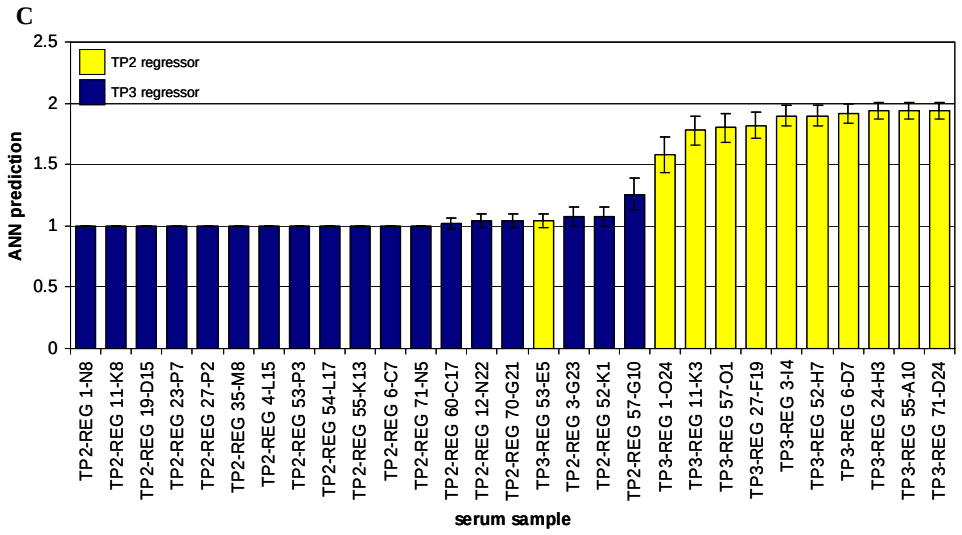


Figure 6-9. Predictive capability of ANNs to recognise MALDI serum tryptic peptide profiles based on top discriminatory ions of TP0/TP1 **(A)**, TP1/TP2 **(B)** and TP2/TP3 **(C)** regressors ANN models. **A)** The dark blue bars indicate TP0 regressor samples, and the yellow bars indicate TP1 regressor samples. A predictive value below 1.5 indicates a TP0 regressor sample, while a prediction greater than 1.5 indicates a TP1 regressor sample. **B)** The dark blue bars indicate TP2 regressor samples, and the yellow bars indicate TP1 regressor samples. A predictive value below 1.5 indicates a TP2 regressor sample, while a prediction greater than 1.5 indicates a TP1 regressor sample. **C)** The dark blue bars indicate TP2 regressor samples, and the yellow bars indicate TP3 regressor samples. A predictive value below 1.5 indicates a TP2 regressor sample, while a prediction greater than 1.5 indicates a TP3 regressor sample.

The TP2/TP3 regressor model is the best model in which only one sample was misclassified. These results indicate that during regression or response to the DC-based vaccination therapy in CT26 murine model, the serum proteomic patterns of TP2 (7 days post therapy) and the TP3 (14 days post therapy, complete regression of tumour) are highly different and therefore the discriminatory biomarkers identified in TP2/TP3 model might be associated with response to therapy. However, more investigations are required to confirm this hypothesis.

6.3.4 Assessment of the specificity of candidate biomarkers associated with regression

In the previous section (6.3.3) potential biomarkers which may associate with tumour regression after therapy were identified. However, in order to determine the specificity of these markers for tumour regression after immunotherapy, the panel of predicted markers for each time-point model was applied to data from progressor and tumour-bearer mice. Failure of biomarkers from the three models (TP0/TP1, TP1/TP2 and TP2/TP3) to classify progressor and tumour-bearer will indicate the specificity of these markers for classification of different time-points of regressor mice.

A two ion model was able to classify 77% of the TP0 and TP1 regressors correctly. In the fist instance, the data from TP0 and TP1 progressor and tumour-bearer mice was used as blind data and presented to this model to assess the ability of the model to correctly classify TP0 progressor and tumour-bearer from the TP1 progressor and tumour-bearer and the results are presented in figure 6-10. The dark blue, yellow, light blue and orange bars are corresponded to TP0 regressor, TP1 regressor, TP0 progressor/tumour-bearer and TP1 progressor/tumour-bearer mice respectively.

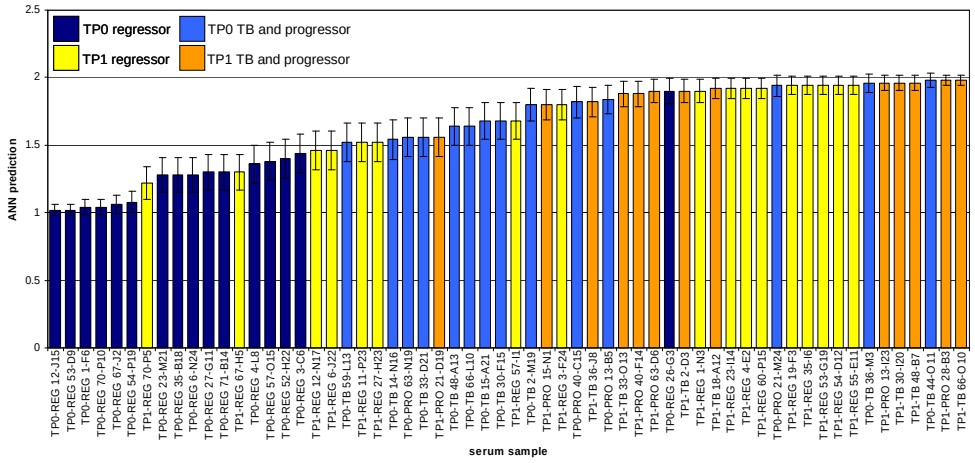


Figure 6-10. Validation of TP0/TP1 regressor two ions ANN model using TP0 and TP1 progressor/tumour-bearer samples. The dark blue, yellow, light blue and orange bars are corresponded to TP0 regressor, TP1 regressor, TP0 progressor/tumour-bearer and TP1 progressor/tumour-bearer mice respectively. Based on the two discriminatory ions of TP0/TP1 regressor model, all the TP0 and TP1 progressor/tumour-bearer samples are classified as TP1 regressor.

As shown in figure 6-10, although all the TP1 progressor/tumour-bearer samples have been correctly classified along with the TP1 regressors, all the TP0 progressor/tumour-bearer are misclassified. All the TP0 progressor/tumour-bearer samples are classified as TP1 samples which indicate that the proteome of the TP0 progressor/tumour-bearer may be more similar to the TP1. In addition, correct classification of TP1 from both groups suggest that these marker may be associated with the presence of the tumour as the TP1 samples are 7 days post CT26 tumour implantation.

Accordingly, the TP1/TP2 regressor model (a two ion model) was tested on the TP1 and TP2 progressor/tumour-bearer data and the classification of the samples are presented in figure 6-11. All the TP2 progressor/tumour-bearer samples are correctly classified however, majority of the TP1 progressor/tumour-bearer samples (9 out of 13) are misclassified based on the two-ion ANN model of regressors. As most of the TP2 samples from both groups are classified correctly based on the two discriminatory ions (*m/z* 1276 and 1985), it is possible that these markers are more therapy related as the TP2 is after therapy administration however, further investigations are required to confirm this hypothesis.

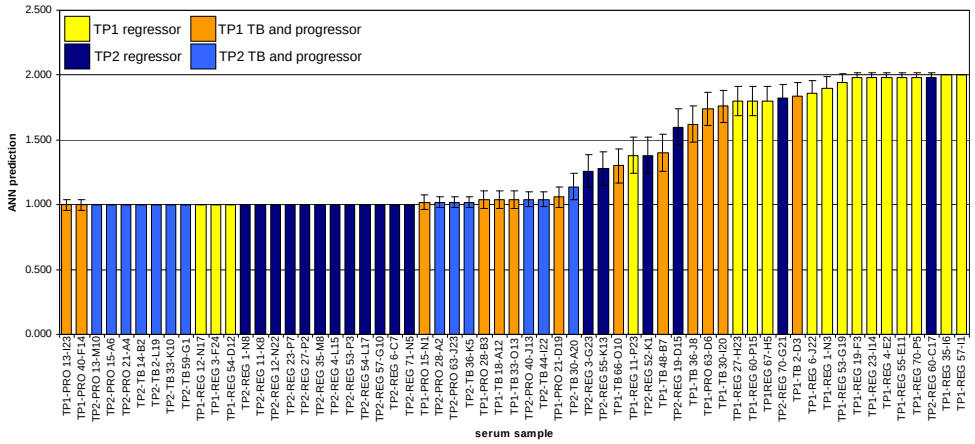


Figure 6-11. Validation of TP1/TP2 regressor two ions ANN model using TP1 and TP2 progressor/tumour-bearer samples. The yellow, dark blue, orange and light blue bars are corresponded to TP1 regressor, TP2 regressor, TP1 progressor/tumour-bearer and TP2 progressor/tumour-bearer mice respectively. Based on the two discriminatory ions of TP1/TP2 regressor model, all the TP2 progressor/tumour-bearer samples are correctly classified however, majority of the TP1 progressor/tumour-bearer samples are misclassified.

Finally, the TP2/TP3 regressor model (a three-ion model) was tested using the data from TP2 and TP3 of progressor/tumour-bearer mice and the results are shown in figure 6-12. In this chart, the dark blue and yellow bars indicate TP2 and TP3 regressors respectively. The light blue and orange bars are corresponded to TP2 regressor and TP3 regressor mice respectively.

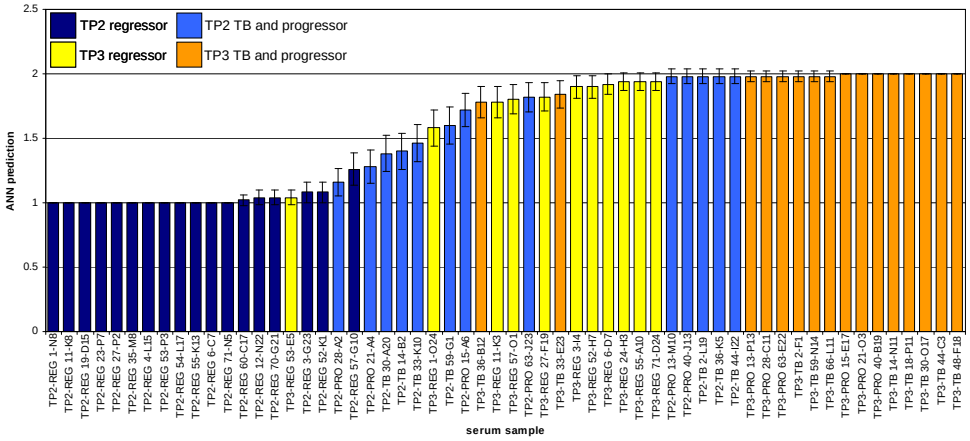


Figure 6-12. Validation of TP2/TP3 regressor two ions ANN model using TP2 and TP3 progressor/tumour-bearer samples. The dark blue, yellow, light blue and orange bars are corresponded to TP2 regressor, TP3 regressor, TP2 progressor/tumour-bearer and TP3 progressor/tumour-bearer mice respectively. Based on the three discriminatory ions of TP2/TP3 regressor model, the model failed to classify TP2 and TP3 progressor/tumour-bearer from each other.

The three discriminatory ions from the TP2/TP3 regressor model (*m/z* values of 1127, 2878 and 2298), fail to classify the TP2 and TP3 progressor/tumour-bearer samples suggesting these ions may be more specific for classification of the regressor animals.

6.3.5 Identification of predicted biomarkers from the DC-based immunotherapy of CT26 murine model.

The top ions from the tryptic digest predicted by ANN analysis of all models were identified by LC-MALDI MS/MS and the results are summerised in table 6-7. The LC-MALDI MS/MS analysis was carried out by our collaborators in Bruker Daltonics (Bremen, Germany). The samples for LC-MALDI testing were prepared using the same methodology used for MALDI-MS analysis except after tryptic digestion, serum samples were fractionated (384 fraction for each sample) with nano-C_{18} columns and each fraction was spotted on the MALDI plate and the top discriminatory ions were identified.

In some cases, there is more than one possible protein identity for each mass value. This is possibly due to the binning of the spectra data points to 1 Dalton prior to ANN analysis were the *m/z* value is a rounded number (no decimal places). Some proteins may have close *m/z* values that are different in decimals and therefore here for some *m/z* values there are more than one protein identity that the true identity requires further analysis and validation. The other issue of having different identities for one *m/z* value can be correlated to tryptic digestion which produces fragments that have similar sequences such as the peak with *m/z* value of 963 in the table 7-6. There are number of different proteins that have the same sequence of amino acids.

m/z	Protein	Sequence	ANN model
855	Identification not acquired	Identification not acquired	TP3 regressor / TP3 progressor and TB
963	preprocomplement component C3	QPLTITVR	TP2 regressor / TP2 progressor and TB
	Serine (or cysteine) peptidase inhibitor, clade A, member 1c [Mus musculus]	VINDFVEK	TP3 regressor / TP3 progressor and TB
	Serpina1b protein [Mus musculus]	VINDFVEK	
	alpha-1 protease inhibitor 2 [Mus musculus]	VINDFVEK	
	alpha-1 antitrypsin precursor	VINDFVEK	
1127	Hemoglobin subunit beta (Hemoglobin beta chain) (Beta-globin)	LHVDPQDFR	TP2/TP3 regressor
1276	Kininogen-1 precursor [Contains: Kininogen-1 heavy chain; Bradykinin; Kininogen-1 light chain]	MNEETASLLLR	TP1/TP2 regressor
1373	prepro complement component C3 [Bos taurus]	LEEDVLPEMGIK	TP1 regressor / TP1 progressor and TB
	serine (or cysteine) proteinase inhibitor, clade C (antithrombin), member 1 [Mus musculus]	LEEDVLPEMGIK	
1435	PREDICTED: hypothetical protein [Ornithorhynchus anatinus]	LVAGVASALAHKYH	TP0/TP1 regressor
	inter alpha-trypsin inhibitor, heavy chain 4 [Mus musculus]	VVAGVATALAHKYH	
1795	inter alpha-trypsin inhibitor, heavy chain 4 [Mus musculus]	DIVWEPPVEPDNTKR	TP2 regressor / TP2 progressor and TB
1816	inter alpha-trypsin inhibitor, heavy chain 4 [Mus musculus]	TYFPHFDVSHGSAQVK	TP2 regressor / TP2 progressor and TB
1901	TAGL-alpha [Mus musculus]	ENPTTFMGHYLHEVAR	TP0/TP1 regressor
1985	histidine-rich glycoprotein [Mus musculus]	YFDSFGDLSSASAIMGNAK	TP1/TP2 regressor
	Complement C4 precursor	YFDSFGDLSSASAIMGNAK	
2298	Identification not acquired	Identification not acquired	TP2/TP3 regressor
2404	apolipoprotein C-II [Mus musculus]	DQSPASHEIATNLGDFAISLYR	TP1 regressor / TP1 progressor and TB
2631	Identification not acquired	Identification not acquired	TP3 regressor / TP3 progressor and TB
2750	Identification not acquired	Identification not acquired	TP0 regressor / TP0 progressor and TB
2877	Zinc-alpha-2-glycoprotein precursor (Zn-alpha-2-glycoprotein) (Zn-alpha-2-GP)	LQEHLKPYAVDLQDQINTQTQEMK	TP0 regressor / TP0 progressor and TB
2878	Zinc-alpha-2-glycoprotein precursor (Zn-alpha-2-glycoprotein) (Zn-alpha-2-GP)	LQEHLKPYAVDLQDQINTQTQEMK	TP2/TP3 regressor
3013	Zinc-alpha-2-glycoprotein precursor (Zn-alpha-2-glycoprotein) (Zn-alpha-2-GP)	WYDVQSVVPHPGSRPDSLEDDLILFK	TP3 regressor / TP3 progressor and TB

Table 6-6. Protein identification by LC-MALDI MS/MS using ANNs predicted tryptic peptide ions.

6.4 Discussion

The discovery of biomarkers that predict therapeutic response is of increasing importance in the development and administration of appropriate therapies for cancer. The serum biomarker identified to date in mass spectrometry based proteomic studies have revealed that patient populations can be determined by the up or down regulation of inflammatory proteins and so the diagnosis of a patient could be determined by the presence of molecules that are not tumour cell specific. In this study we used serum samples collected from the CT26 murine model of immunotherapy where we are eliminating the differences such as genetic background, diet and age from our experiments. Moreover, using this model it is possible to collect serum samples in different time-points of the experiment. We have used MALDI-MS serum proteomic profiling to ascertain whether it would be possible to distinguish between different time-points of pre/post tumour implantation and post therapy serum samples by serum proteome analysis. The detection of proteomic changes in serum of these mice will allow us to predict the response of the mice to the immunotherapy or identification of markers which are correlated to tumour regression or progression.

We report novel panels of multiparametirc proteins identified through proteomic phenomic fingerprinting and ANNs using sera collected from 4 different time-points during the course of pre/post tumour implantation and therapy administration. The finding of biomarkers is of fundamental relevance for detection of early tumour initiation, therapy response and the evolutions associated with therapy and tumour growth. In order to study serum proteome changes we used the CT26 mouse model of DC-based immunotherapy. MALDI-MS proteomic profiling coupled with ANN analysis was used to investigate proteins altered in expression in sera of therapy responder and non-responder mice. Proteome changes were monitored from day 0 (TP0), after tumour implantation (TP1), after therapy administration (TP2) and at the end of experiment where animals respond or fail to respond to the therapy (TP3). Initially, matching time-points of regressor and progressor/tumour bearer mice were compared together to investigate whether there is a difference in the serum proteome between the responder and non-responders. We observed that its possible to distinguish between the regressor and progressors in all the time-points by ANN modeling and 83%-100% correct classification was achieved in different models. The smallest differences were between the TP0 regressor/ TP0 progressor and tumour-bearer profiles were 83% of the samples were correctly classified although this indicates differences between the serum

profiles of regressor and progressor mice. The best predictions (100%) were achieved for the late time-points (TP2 and TP3) which are after therapy, indicating the different status in animals that respond or fail to respond to therapy. Following this success, proteomic profile changes in the regressor animals (therapy responders) was investigated by modeling the different time-points from the regressor animals only. The results indicate that there are differences between the time-points and these markers can be candidates that may associate with therapy response in the mouse model. Moreover, the specificity of these panels was determined from the ANN model when progressor/tumour-bearer samples were used as a totally blind set and non of the ANN models were able to classify the progressor/tumour-bearer samples correctly, which indicates that the regressor biomarkers are more specific for classification of the regressors. Numbers of the discriminatory peaks were identified using LC MALDI MS/MS (Bruker Daltonics, Bremen, Germany), all of which have been recognised as biologically relevant to the presence of tumour. Down- or up-regulation of several proteins, in the different time-points was shown which could prove to be good candidates to prognosis the evolution of tumours and response to therapy. Although some of these protein markers in serum or plasma individually have been tentatively linked with presence of tumour but not directly produced by them, their combination as a panel of protein biomarkers has not previously been associated with therapy response.

Previously (chapter 5), HBB was shown to be discriminatory between the regressor and progressor mice in DISC-HSV therapy model. In the present study HBB is also the first discriminatory peak for the TP2/TP3 regressor model indicating the potential of this protein to be associated with tumour regression and therapy response. The peak intensity of HBB is increased from the TP0 to TP3 in the regressor mice undergo DC-based vaccination. Komita and coworkers shown that a $CD8^+$ T-Cell response against HBB, prevents solid tumor growth after establishment of CMS4 sarcoma (Komita *et al.* 2008). Moreover, they suggested that vaccines limit or destabilise tumor-associated vascular structures, potentially by promoting immunity against HBB vascular pericytes. Although we were unable to validate our finding for HBB in both immunotherapy models by Western Blotting (WB), as future work WB should be carried out after samples are cleaned by C_{18} ZipTip which may show clear cut differences between the different time-points or groups of animals.

Kininogen is another interesting protein that previously has been associated with induction of endothelial cell apoptosis and inhibition of angiogenesis (Zhang *et al.* 2000, Zhang *et al.*

2002). This was the first discriminatory ion between the TP1 and TP2 regressors. The intensity of this peak is also increases in the regressor animals from TP0 to TP3.

Other proteins identified from different models includes complement proteins, apolipoproteins and zinc-alpha-2-glycoprotein which have been reported previously in number of accessions in mouse tumour models that use serum proteome profiling for biomarker identification. Although these proteins may not specific to tumour, a panel of these markers that are tumour associated can be used for screening of tumour progression or therapy response. However, more investigations are required to confirm this suggestion and it would be interesting to use different sample preparation methods to even more reduce the complexity of serum samples from different time-points. LC followed by MALDI analysis of sera from different time-points may reveal a more complete picture of proteins so that the differences can be detected much easily.

In future experiments we believe it is going to be important that the biomarkers we have identified are validated by MS-independent assays. In addition, it is important to identify all the associated proteins to our panel of peaks using possibly different fractionation methods prior to MS/MS to reduce the complexity of serum for accurate targeting of the peaks. We will also explore the value of these markers in tumour conditions. One of the key limitations in biomarker discovery studies is the sample size. Given the level of replication it becomes necessary to reduce the number of components within the model in order to overcome dimensionality issues such as the ability of models to produce a generalised solution. Replication and generalisation issues in this study were overcome by the application of extensive random sample cross validation techniques. In this study three data sets were utilised training data, test data and validation data. The membership of the training and selection data are randomly reshuffled over 50 cross validations. By stopping training on the predictive error of the test data maximum generality is achieved and overtraining prevented. This is verified by performance on validation data. This double validation provides a more realistic representation of model performance than using a single validation data set.

In conclusion we have identified a panel of candidate biomarkers which could be used for the diagnosis of early tumours or may be associated with immunotherapy response in a CT26 immunotherapy model. We have provided proof of principle evidence for the use of robust and standardised automated protocols for sample preparation and MALDI-MS

analysis in conjunction with ANN modelling for the study of serum samples collected from different time-points during tumour growth and therapy with high sensitivity and specificity.

Chapter 7 – Discussion and further studies

Proteomic technologies have been successfully applied in a number of medical areas including infectious, cardiovascular and neurological diseases. Accumulation of DNA changes in genes is broadly accepted as a source of cancer initiation and development that leads to protein expression changes. Hence, investigating changes in protein expression in addition to gene expression could potentially result in the discovery of new cancer biomarkers of clinical utility. The research presented in this thesis has introduced the use of MALDI-MS proteome profiling and bioinformatic analysis, in detection of candidate biomarkers of tumour progression and response to immunotherapy in a CT26 murine model of colorectal carcinoma. Proteomic profiles from serum and tissue were generated by MALDI-MS followed by ANN analysis of the complex data. Candidate biomarkers were identified by ESI-MS/MS and in some cases validated using independent assays.

7.1 Clinical proteomics: Applications in cancer research

Diseases manifest themselves through alterations in the normal phenotype that is typically a reflection of a series of changes in the genome and proteome. Analysis of either the genome or proteome with emerging high throughput technologies can potentially reveal alterations which are associated with disease initiation and progression. Clinical proteomics refers to the applications of current proteomic technologies for clinical investigations which may affect both clinical practice and care by identification of novel biomarkers (Sahab *et al.* 2007, Matharoo-Ball *et al.* 2007). Clinical proteomics aims to quantify and evaluate the proteome changes that are associated with disease initiation, progression or treatment and discover candidate diagnostic, prognostic and/or therapy predictor biomarkers (Fung *et al.* 2005). The field of clinical proteomics to date has focused mainly on detection of disease biomarkers and especially cancer associated biomarkers. Identification of potential candidate cancer biomarkers is possible through utilising a variety of biological body fluids such as serum, plasma, urine and cerebrospinal fluid as well as tissue and cell lines. The research presented in this thesis has used both mouse serum and tumour tissue samples and several different analytical tools have been applied for the detection of biomarkers.

Proteomic profiling of a variety of complex biological samples by matrix assisted laser desorption ionisation-mass spectrometry (MALDI-MS) is one of the commonly used analytical technologies in the field of cancer proteomics in the last 10 years that potentially could yield in identification of multiple cancer biomarkers (Matharoo-Ball *et al.* 2007). Proteomic patterns represent a new paradigm that can potentially be used in clinical diagnosis independent of identity of the discriminatory peaks; however discovering the identity of discriminatory ions of clinical importance, can potentially facilitate and form the basis of developing independent serum antibody-based assays that can be easily used in clinical practice (Petricoin *et al.* 2004). MALDI-MS profiling of serum, plasma, tissue and cell lysates in conjunction with computational artificial intelligence systems has been successfully used in identification of cancer associated proteins. Presence of salts, lipids and high abundant proteins in complex mixtures can be problematic when MALDI-MS or other mass spectrometry instruments used for proteomic profiling (Tiss *et al.* 2007). The problem relies on the competition between the different protein and salts for ionisation which results in a phenomenon known as ion suppression that affects the number of detected ions as well as reducing the sensitivity of detection and suppressing the signal from lower abundant proteins (Callesen *et al.* 2009). In addition, salt and lipids introduce chemical noise to the spectrum where lower signals from molecules can be lost in the noise. In general, most proteins in complex mixtures such as plasma that can be detected by common mass spectrometric instruments have concentration higher that 1µmol/L; however, the tumour associated molecules that are released by the tumour in the blood stream is generally believed to be present in extremely low concentrations (Annesly 2003, Omenn 2006). These limitations can be enhanced by applying a number of techniques to increase sensitivity and quality of proteomic profiling by reducing the dynamic range and complexity of samples by utilizing different pre-fractionation techniques prior to MALDI-MS analysis. A number of common pre-fractionating techniques used by a variety of researchers include solid phase extractions (using different chemical properties of proteins), chromatography, fractionation or capture on magnetic beads, precipitation with organic solvents, gel filtration and gel electrophoresis (Callesen *et al.* 2009, Gilar *et al.* 2001, Hortin 2006, Jin *et al.* 2005). The use of MALDI-MS with different sample fractionation methods may prove to be more beneficial in biomarker discovery studies if these methods can be shown to be reproducible as well as high-throughput. There have been strong

debates about MALDI-MS reproducibility, however use of standard operation procedures (SOPs) that are carefully designed in conjunction with clear and continuous communication between the clinical staff that acquire clinical samples from the patients, with the technical staff performing the analytical procedures, will help to reduce errors due to pre-analytical procedures and sample handling. The present study used C_{18} ZipTip chromatography for sample preparation prior to MALDI analysis to remove high molecular weight proteins such as albumin, remove salts and separate hydrophobic proteins for analysis. Serum sample preparation using C_{18} ZipTip have proven to be the technique with good performance that results in higher number of peaks in the spectra with good signal-to-noise and significant reproducibility over different experiments which has been also shown by several other reports (Matharoo-Ball *et al.* 2007, Tiss *et al.* 2007). In addition, use of quality control samples and constancy in the study design and operational procedures in this study has shown to provide clinically relevant results. The benefit of MALDI-MS in biomarker discovery to date cannot be confirmed as few groups managed to validate their results using multiple patient groups, between different research centres and independent techniques. However, the technology still needs to be explored in further detail and at a larger scale, before established as a robust tool in clinical practice.

7.2 The potential use of mouse models in cancer proteomic analysis and biomarker identification

Animal models played a critical role for pre-clinically evaluating and studying many aspects of cancer biology (*i.e.* tumour immunotherapy, drug efficiency testing and tumour physiology studies) by providing an *in-vivo* milieu that in many respects reflects human disease and complements *in-vitro* studies using patient material. Tissue, serum and plasma are easily accessible for bimolecular investigations and studying the emerging proteomic-based strategies and their applications in oncology.

Direct analysis of serum from cancer patients has been extensively studied to reveal new cancer biomarkers; however, the complexity and heterogeneity of human samples remains the major challenge and data interpretation is often difficult due to physiological differences among patients (Faca *et al.* 2008). Genetic backgrounds, age, gender and the presence of chronic illnesses are some of the factors that may affect biomarker discovery experiments. These factors influence the serum proteome concept and any detected

candidate biomarkers may by related to these factors rather than cancer itself (Juan *et al.* 2004). These factors can be controlled to some extent by using inbred animal models where aged, gender and genetic background is to a large extent constant and where environmental conditions and diet can be controlled. Most biomarker studies are based on the comparison of healthy and disease patient groups and it is difficult to obtain clinical materials from the same individuals prior to cancer development and in some cases prior to therapy. This is particularly important in studies that aim to identify biomarkers that predict therapy outcome. Most clinical cancer therapy trials collect samples from patient's pre and post treatment, but it is not possible to obtain pre-cancer samples from individual patients. Despite the use of animal tumour models for research, there are only relatively few reports on the use of these for biomarker discovery.

The animal models used in this study involved a CT26 tumour progression model and two separate CT26 immunotherapy models. All the animals used were kept in controlled conditions and were aged and gender matched. In addition, higher numbers of mice is used, compare to similar published proteomic studies, to provide increased confidence and generalisation of the results. The sample size in the design of clinical proteomics that aim for discovery of biomarker is of highly importance and low number of samples in experiments may reduce the confidence on the final results (Cairns *et al.* 2009). The CT26 mouse models used provided a platform to study proteome changes during tumour growth and identified biomarkers that may be involved in tumour rejection and progression. However, unlike human cancer, the CT26 cell line grows rapidly following implantation into mice; moreover, CT26 cells are implanted subcutaneously which is not the site of origin of the tumour. The use of genetically engineered mice (GEM) has increasingly become popular while recently sophisticated GEM models have been developed that are capable of mimicking the histological and biological behavior of human cancers (Gutmann *et al.*2006, Faca *et al.* 2008). An interesting study by Faca *et al.* described the use of a GEM model of human pancreatic cancer and identified a panel of five proteins from serum proteomic analysis of GEM mice for diagnosis of early stage cancer. Furthermore, this panel of five proteins was capable of discriminating human serum samples from pancreatic cancer patients from healthy controls. In addition, immunodeficient mice have been used to establish tumours from human tumour cell lines in order to detect tumour biomarkers in the serum (Juan *et al.* 2004, van den Bemd *et al*. 2006).

7.3 Discovery of cancer biomarkers through serum proteomic profiling

Serum is readily and in large quantities obtained from patients and as a result serum-based studies are widespread. It is reasonable to expect biomarkers in interstitial tissue fluid and blood, originating from the tumour itself or its microenvironment. There are five different mechanisms during cancer initiation and progression which leads to the elevation of proteins in biological fluids such as serum and plasma (Kulasingam & Diamandis 2008). These mechanisms includes: gene over expression, increased protein secretion and shedding, tumour metastasis and invasion, angiogenesis and distortion of tissue structure. Over expression of human epididymal secretory protein 4 (HE4) gene in patients with ovarian cancer results in elevation of HE4 protein in the serum which can be quantified via immunoassays and therefore could potentially serve as a marker of ovarian cancer. Invasion and metastasis of normal tissue by tumour cells may permit release of molecules to the interstitial fluid which subsequently enters the lymphatic system and ultimately the blood (Kulasingam & Diamandis 2008). For example, in patients with prostate cancer, although transcription of the PSA gene in the prostatic epithelial cells is down regulated, alteration and disruption of the anatomic barriers such as lymphatic capillaries between the prostatic columnar epithelial cells and glandular lumen allows the release of abnormal amount of PSA into the blood stream (Simpson *et al.* 2008). In healthy males, such anatomic barriers (*i.e.* basement membrane, stromal layer, blood vessel walls and lymphatic capillaries) control the amount of secreted PSA from the prostatic columnar epithelial cells to the glandular lumen (Kulasingam & Diamandis 2008). Moreover, abnormal secretion and shedding of membrane bound proteins by the tumour cells may be detectable in biological fluids and a good example of such biomarkers is HER2 which is involved in cell growth and differentiation.

For MS-based proteomic studies, serum is often described as a "dual-edged sword" as it contains high concentration of a wide range of proteins; however, high abundance proteins may mask the identification of low abundance ones and cancer ion suppression (Veenstra *et al.* 2005). In addition to small molecules such as amino acids, lipids, salts and sugars, human serum contains 60-80 mg/ml proteins and in rodents the protein concentration of serum is 40-60 mg/ml (Ziad *et al.* 2007). Proteins such as albumin and IgG, present in normal serum, account for >95% of the total protein content (Anderson & Anderson 2002) and approximately 12g of albumin is synthesised on a daily basis, constituting 50% (w/w)

of the total serum protein (McFarlane *et al.* 2000). Considering the broad dynamic range and complexity of serum proteome, sample treatment using a variety of analytical techniques (*i.e.* fractionation, chromatography and antibody-based immunoaffinity subtraction) is required prior to biomarker identification. However, the drawback of high abundant protein removal from serum is the possible elimination of low abundant small protein/peptides that are attached to the high abundant carrier proteins such as albumin. The proteomic profiling studies presented in this thesis have utilised pre-packed ZipTip chromatography columns for sample preparation prior to MALDI-MS analysis. This simple procedure of sample preparation for both tissue and serum samples have revealed MALDI proteomic profiles that can be used to identify the presence of certain proteins that classify different groups. A number of proteomic studies have been published where the authors have used serum to identify proteins of possible clinical significance.

7.3.1 Biomarkers associated with tumour initiation and progression

Using the CT26 progression model, serum and tissue samples were collected at four time-points from tumour-bearing and control mice which provided the opportunity to assess the tumour proteome changes in a time-course from tumour initiation and at different stages of growth. The present study, using MALDI-MS and ANN analysis resulted in the identification of panels of candidate biomarkers capable of correctly classifying a high proportion of mice at different stages of tumour progression, and revealing important prognostic information. Classification of TP0 (no tumour) from the TP1 (7 days post tumour implantation) was possible using a panel of 5 ions based on their proteomic profiling. In addition, TP1 *versus* TP2 (14 days post tumour implantation) and TP2 *versus* TP3 (21 days post tumour implantation) classification based on the MALDI-MS proteomic profiles was possible using panels of 4 and 3 ions respectively. The discriminatory panel for each of these stages of tumour progression was unique to each model, except one ion (*m/z* 2837) which was present on the panel of markers corresponding to the TP1vsTP2 and TP2vsTP3 model. The *m/z* values in the panel from these models have not been presented in the discriminatory markers from the immunotherapy models and therefore may only be associated with tumour progression but these can be only confirmed once the identity of the proteins related to the ions are revealed. This aspect of the study still requires further

identification of the nature and characteristics of the biomarkers. Validation studies with samples from an independent model would strengthen the results and alternative assays should be used for validation of results. These biomarkers may be indicative of tumour progression and therefore could indicate important molecular species which could have relevance for human cancer. The use of MALDI-MS and ANNs represents an appropriate technology for biomarker assessment and identification, relative to tumour status and progression.

7.3.2 Biomarkers associated with response to immunotherapy

The protein and peptide biomarkers of biological samples from a CT26 murine colon cancer model that associate with response or failure to immunotherapy with DISC-HSV or DC-based vaccination therapy with CTLA-4 and blockade of VEGFR-2 was investigated. In both immunotherapy models, upto 70-80% of the treated mice responded to therapy (termed regressors) while the remaining animals showed progressive growth of their tumours; hence, it was possible to analyse differences between the responder and non-responder animals which is an additional unique feature of this study. In the DISC-HSV immunotherapy model, serum samples from tumour-free and after failure or response to therapy from same individuals were used to determine whether it is possible to classify responders from non-responders using our strategy for serum proteomic profiling. In the DC-based vaccination model samples were collected from four different time-points over the period of tumour implantation, immunotherapy and response or failure to therapy. To the best of my knowledge this is the first study that uses animal immunotherapy models to identify biomarkers that may be associated with therapy outcome. The long term impact of identification of novel therapeutic cancer makers is critical to providing a means for patient stratification and a step closer to implementing personalised medicine. The data obtained in murine immunotherapy may indicate pathways, important in response to therapy, that may be translated to human or at least indicate important principles associated with the response. The first immunotherapy model investigated for the biomarker identification was the DISC therapy. In this study we identified and confirmed serum peptide biomarkers (SAA-1, SAP , HBB and HPX) that associate with failure of immunotherapy and progression in a colorectal murine model, demonstrating the ability of MALDI-MS in conjunction with ANN modeling to discover biomarkers. The uniqueness of our study was that we used a

retrospective CT26 colorectal cancer set of samples collected in 2004 at the end of the study period only for the discovery/identification of the peptide markers and in a separate independent prospective (2008) regressor/progressor/naïve/tumour bearer CT26 colorectal cancer set of samples collected 4 years later for the biological and immuno-validation of the biomarkers. Additionally with this cohort of samples we were able to collect the blood prior to any tumour/therapy in the healthy state of the animals and then following tumour/therapy outcome. This scenario provides the ideal experimental design to elucidate how the markers change in the same animal from a healthy state and after tumour/therapy outcome. In the case of regressor/progressor peptide analysis we were able to demonstrate 86% accuracy (sensitivity of 90% and specificity of 81%) with a panel of 4 ANN predicted biomarkers. In the case of progressors versus healthy controls 1 tryptic peptide peak identified by ANN analysis was able to predict the two groups with accuracy 85%, sensitivity of 88% and specificity of 96%. The ANNs however, failed to classify regressors from healthy controls, suggesting that the serum proteome of mice responding to therapy is similar to healthy individuals. Furthermore, we were able to identify by ESI MS/MS all four biomarker ions from the regressor/progressor ANN analysis as (1) ion *m/z* 1312 as a fragment of haemoglobin beta-2 subunit (HBB), (2) ion *m/z* 1132 as a peptide of serum amyloid A-1 (SAA-1), (3) ion 1192 as a fragment of hemopexin (HPX) and finally (4) *m/z* 2133 as serum amyloid P (SAP) protein fragment. The one tryptic peptide ion from ANN analysis discriminating progressors from healthy controls was also identified as a component of HPX. The findings of this study were successfully validated using western blotting, IHC, ELISA and RT-PCR. Further investigations are required to define possible mechanism(s) and the involvement of these markers in failure or response to immunotherapy of CT26 tumours in mice.

In the second immunotherapy model, where mice received DC-based vaccination combined with anti CTLA4 therapy, serum samples were collected at four different time-points during the course of tumour initiation and immunotherapy with DC-based vaccine. In this study as proof-of-principle using MALDI-MS and ANN analysis we were able to detect a panel of discriminatory biomarkers between the time-point that were identified with ESI MS/MS that may be involved in tumour progression of therapy response. However, further studies are required to validate these findings.

7.4 Discovery of cancer biomarkers through tissue proteome profiling

A number of proteomic studies have been published using cancer cell lines to identify proteins of possible clinical significance (Koike *et al.* 2005). Despite the widespread use of cell lines in the field of cancer proteomics to identify proteins with possible clinical significance, cell lines do not provide an exact model of a tumour, as the tumour microenvironment invokes changes upon the tumour genome and proteome and the molecules expressed and secreted by the host as well as tumour cells. The development of assays for detection and treatment of tumours may start with investigating molecules and mechanisms of tumour progression. The studies presented here utilised a CT26 murine model of progression and show that proteins extracted from CT26 tumour tissue could be used to classify tumours from 3 different time-point of tumour progression. This study provide a "proof-of-principle" for using MALDI-MS in conjunction with ANN analysis to successfully identify discriminatory pattern between CT26 tumour tissue extracts of different stage of growth. Future studies would necessitate the use of larger cohorts of samples and methodologies to identify discriminatory markers. Pre-clinical evaluation of biomarkers in tumour models forms a useful and essential pre-requisite for studies of human disease, where identification of discriminatory proteins (biomarkers) with clinical significant would have clinical utility. Thus, the translation of these methods to study human tumour tissues to detect stage specific candidate biomarkers involved in tumour progression, metastasis and/or predicting patient response to therapy is now possible.

The method of tissue analysis used in this study involved homogenisation of whole tumour tissue and extraction of the total protein. As heterogeneity is one of the most important characteristics of cancer, manifested through diverse genetic and proteomic expression pathways, it has been stated that the use of whole tumour tissue extracts may not reflect accurately biomarker changes during tumour progression (Wang *et al.* 2009). Human tumour cell heterogeneity and the presence of subpopulations within the tumours will be reflected through patient (Heppner 1984). However, the use of technologies such as laser capture microdissection (LCM) combined with MALDI-MS profiling may increase the chance of identifying relevant proteins involved in tumour progression. Moreover, whole protein extraction from tissue does not allow determination of protein localisation within the tissue. Protein localisation is often possible using techniques such as immunohistochemistry, however previous knowledge of proteins of interest and the

Comment [A3]: Baharak this senstence is not complete I am not sure what this reflects? Can u have a look
Cheers
Bal

availability of specific antibodies to the protein is required. Intact tissue profiling using MALDI-MS allows molecular weight specific maps or images of tissue sections to be obtained rapidly at high resolution and sensitivity. This method is still in the early stages of development but improvements in sample preparation, instrumentation and data analysis to increase sensitivity and specificity of this approach has made this possible (Caldwell & Caprioli 2005). So far this method has been applied to the profiling of glioblastoma, prostate and colon cancers, resulting in the discovery of numerous disease specific biomarkers and their spatial localization within the tumours (Stoeckli et al. 2001).

7.5 Current technology challenges: independent validation of candidate biomarkers

The discovery of biomarkers that predict therapeutic response is of increasing importance in the development and administration of appropriate therapies for cancer. The peptidome hypothesis states that many proteins and peptides are shed from the tumour cells into the tissue microenvironment and enter the circulatory systems therefore; direct screening of the serum proteome may provide protein signatures indicative of a specific state of disease or phenotype with high sensitivity and specificity (Petricoin *et al.* 2006). Serum is the most desirable and widely used biological fluid for this goal due to the presence of high concentrations of proteins and because it is easily obtainable. Therefore, the use of high throughput technologies such as MALDI-MS offers the opportunity for precise identification of proteins and peptides that cannot be resolved using methodologies such as 2D gel electrophoresis (Petricoin & Liotta 2004). The ability of MALDI serum profiling combined with several different computer based algorithms for data analysis to classify and analyze different states of disease has been confirmed by a variety of groups however, important concerns were raised over the reproducibility and reliability of MS based proteomic profiling techniques. Experimental variations occurring in MALDI-MS experiments are well-known and are discussed by others and they addressed a number of important factors that influence serum proteome profiles. Individual and environmental differences (i.e. age, sex, general physical health, stress, diet, race and etc) and preanalytical conditions (i.e. sample handling and preparation, sample processing protocols etc) are two main factors that have significant impacts on the MALDI profiles (Baggerly *et al.* 2004 and Villanueva *et al.* 2005). Therefore, use of standardized and well designed

approaches for proteome profiling plays an important role in reduction of bias from MS based results and findings. Moreover, identification of the sources of variation which may initiate from the instrument in use, sample preparation or/and from individual differences and their influence on the proteome profiles may direct us to more accurate and relevant conclusions. Validation and reproduction of results by independent laboratories under highly strict conditions for SELDI and MALDI has been shown previously (Enqwegen *et al.* 2007). In our laboratory a standard workflow for serum proteome profiling by MALDI-MS in conjunction with ANNs algorithms has been designed in which most of the issues and bias over MALDI-MS proteome profiling is taken into consideration. In this study we used serum samples collected from the CT26 murine model of immunotherapy where we are eliminating the differences such as genetic background, diet and age from our experiments. We have used two different strategies, one gel based and the other one non-gel based, in conjunction with MALDI-MS to ascertain whether it would be possible to distinguish between tumour regressor and tumour progressor mice by serum proteome analysis. The detection of proteomic changes in serum of these mice will allow us to predict the response of the mice to the immunotherapy or identification of markers which are correlated to tumour regression or progression. Protein biomarkers in the serum hold great promise for disease detection and classification but often these biomarkers cannot be detected easily in the serum samples by mass spectrometry due to the presence of high abundance proteins such as albumin in the blood. The low molecular range of the serum proteome contains shed proteins and protein fragments originated from physiologic and pathologic events taking place in all tissues. As the larger protein molecules cannot enter the blood circulation, the low molecular weight region of serum is an attractive specimen to detect biomarkers associated with a specific disease (Geho *et al.* 2006) and recently novel methods have been applied by the researchers to eliminate the high molecular weight proteins present in the serum and analyse the low molecular weight proteins by mass spectrometry techniques (Tirumalai *et al.* 2003 and Chertov *et al.* 2005). Use of C_{18} ZipTips for fractionation and concentration of serum samples prior to MALDI-MS analysis has been examined previously by a number of investigators also efficiency of these columns for removal of serum albumin has been shown in previous chapters of this study. The number of unique proteins and peptides that can be identified from complex mixtures such as serum depends on the fractionation procedures therefore we used C_{18} ZipTip to

reduce the complexity of the samples as well as concentrating and de-salting them. The C_{18} ZipTip are chromatographic columns that hydrophobic proteins can bind to, and they have a preferable mass range that can bind to these columns letting behind proteins such as albumin in fact, 98% loss of protein for C_{18} ZipTips have been reported (Matharoo-Ball *et al.* 2007b). As sample handling and processing holds an important role in results of the MALDI-MS analysis, we utilised a robotics liquid sample handling system which ensures confidence and reliability in the instrument for reproducible sample processing. The results presented in this report demonstrate the use of MALDI-MS serum proteome profiling in conjunction with ANNs analysis for cancer biomarker identification and using robust analytical protocols can provide reproducible and meaningful data. We further validated three selected proteins (SAP, HPX and haemoglobin) by semi-quantitative Western blotting analysis. The resulting data for these proteins using an independent assay show the same general trends found by mass spectrometry analysis. Additionally, the identified potential biomarkers were further validated using sera obtained from an independent prospective experiment; moreover, SAP and SAA were quantified using ELISA. The findings of this study suggest that a panel of serum biomarkers identified by MALDI-MS in this experimented model of immunotherapy along with the clinical parameters may be used to predict treatment outcome. In addition, our findings demonstrate that a panel of APPs might serve as biomarkers for detection of therapy response. The mRNA levels of SAP, HPX and SAA expression was also validated using RT-PCR which shows increased levels of expression in progressors compared to naïve and regressors which confirmed the proteomic and WB findings. In addition, our findings demonstrate that a panel of APPs might serve as biomarkers for detection of therapy response.

7.6 Bioinformatics

The application of ANNs to identify patterns correlating with clinical parameters allows us to gain further understanding of the biological diversity of different cancers. The emergence of subtypes of disease also makes it difficult to determine the prognosis of patients and decide upon the most appropriate treatment. Classifying tumours into distinct groups according to their protein profiles may provide additional information of diagnostic and prognostic benefit to the patient. ANN analysis also identified sample outliers that could represent new subtypes of the cancers. The stepwise analytical approaches used for

the studies presented here has resulted in the identification of multiple markers associated with, for example, the response to therapy. This highlights the fact that using a simplistic, single marker approach to describe very complex and heterogenous diseases is unrealistic and that the use of multi-marker models will allow much more accurate conclusions to be drawn. Also, it has been revealed that predictive patterns can be identified to classify a high proportion of samples using different bioinformatics approaches. However, these approaches identify different groups of ions that could be used for prediction and so biomarker identification using these types of approaches is dependent upon which method is used for the analysis of the data. This present study has shown that ANNs can be used to determine predictive markers of therapy in CT26 immunotherapy models and markers of tumour progression, but it has also been used in studies involving prostate and melanoma cancer.

Sample handling and processing, instrument noise and data analysis all contribute to the challenges of reproducibility in any proteomics experiment (White et al. 2004). Variability in sample handling and the mass spectrometer causes a baseline across the spectra as well as 'noise' and variability in the amount of protein bound onto the target also causes a fluctuation in the intensity scale of the spectra. Variablilty in the spectra has the largest effect on small peaks as the baseline and instrument noise can be as large as the peak itself, yet it is possible that these small peaks contain much of the biological information; hence the standardisation of sample and data processing procedures should be of utmost importance when embarking upon biomarker discovery studies (Rodland 2004).

7.7 Cancer diagnosis and treatment and personalised medicine

The clinical and pathological biomarkers that are currently used poorly predict early disease development and response to treatment. The aim of this study and others was to attempt to identify biomarkers that can improve upon markers currently in use.

Majority of the biomarkers generated by MALDI MS have been identified and these have studies have mainly concentrated on biomarkers present in serum and appear to be isoforms of ubiquitous proteins that occur as a result of secondary tumour effects, for example altered enzymatic activity; the specificity of these isoforms requires further investigation (Engwegen *et al.* 2006). Like normal cells, most cancer cells use multiple intracellular signaling pathways to ensure the maintenance of functions that are critical to their survival.

Thus, cellular pathways that are integral to cell function, survival, proliferation, and receptor expression are potential targets for therapeutic intervention but are not necessarily tumour cell specific, one example being the epidermal growth factor receptor signaling pathway. Molecules that mediate the production of angiogenic and invasion factors that allows tumor growth and metastasis, such as the vascular endothelial growth factor and downstream events that result in cellular apoptosis represent additional potential targeting pathways (Ajani and Allgood 2005).

Standard diagnostic methods, including tissue histopathology are being replaced or complimented by the use of molecular diagnosis, which can identify proteins and their posttranslational modifications that occur in disease conditions, and hence greatly accelerate progress toward novel diagnostic and predictive tools to track early disease and tailor treatments to specific patients (personalised medicine) (Alaoui-Jamali and Xu 2006). In the clinic, new patients could be tested (either using serum or by isolating protein from a biopsy) by SELDI or MALDI analysis to generate protein or peptide fingerprints. The profiles of patients could then be compared to control or 'baseline' profiles from normal subjects in order to determine the diagnosis, or prognosis or to predict response to a particular treatment. Continued investigations may result in the development of proteomic profiling databases through which a patient could be matched with protein profiles relevant to the disease and potential benefits of treatment. In this way, clinicians would be able to recommend combinations of molecularly targeted agents and therapies on the basis of an individual patient's proteomic profile (Ajani and Allgood 2005). Samples compared to profiles in an existing database would have to be applicable to the general population, taking into consideration factors such as age, sex, ethnicity and nutritional status. The proteins identified to date in MALDI MS based proteomic studies have revealed that patient populations can be determined by the up or down regulation of inflammatory proteins and so the diagnosis of a patient could be determined by the presence of molecules that are not tumour cell specific.

In conclusion, the MALDI approach utilised in the present research has the potential to be used as a clinical tool in the future for the diagnosis and prognosis of patients. The MALDI proteomic profiling strategy has the capability of producing reliable results that can be validated as shown here to be true results. However, further developmental work is required before this can become a reality, including the type of samples that could be analysed, the

bioinformatics approach used to analyse the data as well as the method used to generate the spectra. The work presented has shown that different types of samples can be used to derive clinically relevant information from a proteomic profile. Different types of bioinformatics approaches can be used to analyse the data and to classify samples with the same accuracy; this is likely to be achieved using different ions/biomarkers to do this. It has also been shown that different sample preparation methods can be utilised, each generating a different set of prominent ions; despite this, however, the patterns within these spectra allow the classification of samples with a similar degree of accuracy. Further work should determine the best method to use for sample preparation, proteome profiling and bioinformatics analysis, using the same sample set derived from a large cohort of samples. To date proteome profiling and analysis of samples to establish reproducibility in several centres has not been carried out. There is a need to validate proteomic data using a multi-centre approach.

References

Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal. 2000 Jun;22(5):717-27.

Ahmad M, Rees RC, McArdle SE, Li G, Mian S, Entwisle C, Loudon P, Ali SA. Regulation of CTL responses to MHC-restricted class I peptide of the gp70 tumour antigen by splenic parenchymal CD4+ T cells in mice failing immunotherapy with DISC-mGM-CSF. Int J Cancer. 2005;115(6):951-9.

Ahmed A, Tollefsbol T. Telomeres, telomerase, and telomerase inhibition: clinical implications for cancer. J Am Geriatr Soc 2003; 51 (1):116-122.

Ajani J, Allgood V. Molecular mechanisms in cancer: what should clinicians know? Semin Oncol 2005; 32 (6 Suppl 8):2-4.

Alaoui-Jamali MA, Xu YJ. Proteomic technology for biomarker profiling in cancer: an update. J Zhejiang Univ Sci B 2006; 7 (6):411-420.

Al-Hajj M, Becker MW, Wicha M et al. Therapeutic implications of cancer stem cells. Curr Opin Genet Dev 2004; 14 (1):43-47.

Ali SA, Lynam J, McLean CS, Entwisle C, Loudon P, Rojas JM, McArdle SE, Li G, Mian S, Rees RC. Tumor regression induced by intratumor therapy with a disabled infectious single cycle (DISC) herpes simplex virus (HSV) vector, DISC/HSV/murine granulocyte-macrophage colony-stimulating factor, correlates with antigen-specific adaptive immunity. J Immunol. 2002; 168(7): 3512-3519.

Ali SA, McLean CS, Boursnell ME, Martin G, Holmes CL, Reeder S, Entwisle C, Blakeley DM, Shields JG, Todryk S, Vile R, Robins RA, Rees RC. Preclinical evaluation of "whole" cell vaccines for prophylaxis and therapy using a disabled infectious single cycle-herpes simplex virus vector to transduce cytokine genes. Cancer Res. 2000a;15;60(6):1663-70.

Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics. 2002 Nov;1(11):845-67.

Annesley TM. Ion suppression in mass spectrometry. Clin Chem. 2003 Jul;49(7):1041-4.

Ashcroft AE. Protein and peptide identification: the rôle of mass spectrometry in proteomics. Nat Prod Rep. 2003;20(2):202-15.

Assudani DP, Ahmad M, Li G, Rees RC, Ali SA. Immunotherapeutic potential of DISC-HSV and OX40L in cancer. Cancer Immunol Immunother. 2006;55(1):104-11.

Baak JP, Path FR, Hermsen MA, Meijer G, Schmidt J, Janssen EA. Genomics and proteomics in cancer. Eur J Cancer. 2003 Jun;39(9):1199-215.

Bachwich DR, Lichtenstein GR, Traber PG. Cancer in inflammatory bowel disease. Med Clin North Am. 1994 Nov;78(6):1399-412.

Badolato R, Wang JM, Murphy WJ, Lloyd AR, Michiel DF, Bausserman LL, Kelvin DJ, Oppenheim JJ. Serum amyloid A is a chemoattractant: induction of migration, adhesion, and tissue infiltration of monocytes and polymorphonuclear leukocytes. J Exp Med. 1994;180(1):203-9.

Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics. 2004;20(5):777-85.

Balkwill F, Mantovani A (2001) Inflammation and cancer: back to Virchow? The Lancet. 2001; 357:539–545.

Ball G, Mian S, Holding F, Allibone RO, Lowe J, Ali S, Li G, McCardle S, Ellis IO, Creaser C, Rees RC. An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. Bioinformatics. 2002; 18(3): 395-404.

Bashyam H. CTLA-4: From conflict to clinic. J Exp Med. 2007 Jun 11;204(6):1243.

Bertman JS. The molecular biology of cancer. Mol Aspects Med. 2000;21(6):167-223.

Bertucci F, Birnbaum D, Goncalves A. Proteomics of breast cancer: principles and potential clinical applications. Mol Cell Proteomics. 2006 Oct;5(10):1772-86. Epub 2006 May 29.

Bhatia M, Wang JC, Kapp U et al. Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. Proc Natl Acad Sci U S A 1997; 94 (10):5320-5325.

Bischoff R, Luider TM. Methodological advances in the discovery of protein and peptide disease markers. J Chromatogr B Analyt Technol Biomed Life Sci 2004; 803 (1):27-40.

Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med 1997; 3 (7):730-737.

Bruce WR, Van Der Gaag H. A Quantitative Assay For The Number Of Murine Lymphoma Cells Capable Of Proliferation In Vivo. Nature 1963; 199:79-80.

Bueter M, Gasser M, Schramm N, Lebedeva T, Tocco G, Gerstlauer C, Grimm M, Nichiporuk E, Thalheimer A, Thiede A, Meyer D, Benichou G, Waaga-Gasser AM.T-cell response to p53 tumor-associated antigen in patients with colorectal carcinoma. Int J Oncol. 2006;28(2):431-8.

Byrne JC, Downes MR, O'Donoghue N, O'Keane C, O'Neill A, Fan Y, Fitzpatrick JM, Dunn M, Watson RW. 2D-DIGE as a strategy to identify serum markers for the progression of prostate cancer. J Proteome Res. 2009 Feb;8(2):942-57.

Cairns DA, Barrett JH, Billingham LJ, Stanley AJ, Xinarianos G, Field JK, Johnson PJ, Selby PJ, Banks RE. Sample size determination in clinical proteomic profiling experiments using mass spectrometry for class comparison. Proteomics. 2009;9(1):74-86.

Callesen AK, Christensen R, Madsen JS, Vach W, Zapico E, Cold S, Jørgensen PE, Mogensen O, Kruse TA, Jensen ON. 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.

Callesen AK, Madsen JS, Vach W, Kruse TA, Mogensen O, Jensen ON. Serum protein profiling by solid phase extraction and mass spectrometry: a future diagnostics tool? Proteomics. 2009;9(6):1428-41.

Chatterjee SK, Zetter BR. Cancer biomarkers: knowing the present and predicting the future. Future Oncol. 2005;1(1):37-50.

Cho WC. Contribution of oncoproteomics to cancer biomarker discovery. Mol Cancer. 2007;6:25.

Choi W, Song SW, Zhang W. Understanding cancer through proteomics. Technol Cancer Res Treat. 2002 Aug;1(4):221-30.

Christofori G, Semb H. The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. Trends Biochem Sci 1999; 24 (2):73-76.

Clarke MF, Dick JE, Dirks PB, Eaves CJ, Jamieson CH, Jones DL, Visvader J, Weissman IL, Wahl GM. Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells. Cancer Res. 2006;66(19):9339-44.

Clevers H. At the crossroads of inflammation and cancer. Cell 2004; 118:671–674.

Cocco E, Bellone S, El-Sahwi K, Cargnelutti M, Casagrande F, Buza N, Tavassoli FA, Siegel ER, Visintin I, Ratner E, Silasi DA, Azodi M, Schwartz PE, Rutherford TJ, Pecorelli S,Santin AD. Serum amyloid A (SAA): a novel biomarker for uterine serous papillary cancer. Br J Cancer. 2009 Jul 21;101(2):335-41.

Conrad DH, Goyette J, Thomas PS. Proteomics as a method for early detection of cancer: a review of proteomics, exhaled breath condensate, and lung cancer screening. J Gen Intern Med. 2008 Jan;23 Suppl 1:78-84.

Coombes KR, Fritsche HA, Jr., Clarke C et al. Quality control and peak finding for proteomics data collected from nipple aspirate fluid by surface-enhanced laser desorption and ionization. Clin Chem 2003; 49 (10):1615-1623.

Coombes KR, Tsavachidis S, Morris JS, Baggerly KA, Hung MC, Kuerer HM. Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. Proteomics. 2005 Nov;5(16):4107-17.

Coussens, L. M. and Werb, Z. Inflammation and cancer. Nature 2002; 420: 860–867.

Culp WD, Neal R, Massey R, Egevad L, Pisa P, Garland D. Proteomic analysis of tumor establishment and growth in the B16-F10 mouse melanoma model. J Proteome Res. 2006 Jun;5(6):1332-43.

Dalerba P, Cho RW, Clarke MF. Cancer stem cells: models and concepts. Annu Rev Med. 2007;58:267-84.

Dalla RM and Lotze MT. The dendritic cell and human cancer vaccine. Curr Opin Immuno. 2000;12(5):583-8.

de Gruijl TD, van den Eertwegh AJ, Pinedo HM, Scheper RJ. Whole-cell cancer vaccination: from autologous to allogeneic tumor- and dendritic cell-based vaccines. Cancer Immunol Immunother. 2008;57(10):1569-77.

de Noo ME, Tollenaar RA, Ozalp A, Kuppen PJ, Bladergroen MR, Eilers PH, Deelder AM. Reliability of human serum protein profiles generated with C8 magnetic beads assisted MALDI-TOF mass spectrometry. Anal Chem. 2005;77(22):7232-41.

Dekker LJ, Boogerd W, Stockhammer G, Dalebout JC, Siccama I, Zheng P, Bonfrer JM, Verschuuren JJ, Jenster G, Verbeek MM, Luider TM, Smitt PA. MALDI-TOF mass spectrometry analysis of cerebrospinal fluid tryptic peptide profiles to diagnose leptomeningeal metastases in patients with breast cancer. Mol Cell Proteomics. 2005; 4(9):1341-9. Epub 2005 Jun 21.

Dermime S, Armstrong A, Hawkins RE, Stern PL. Cancer vaccines and immunotherapy. Br Med Bull. 2002;62:149-62.

Diamandis EP, van der Merwe DE. Plasma protein profiling by mass spectrometry for cancer diagnosis: opportunities and limitations. Clin Cancer Res. 2005;11(3):963-5.

Diamandis EP. Clinical applications of tumor suppressor genes and oncogenes in cancer. Clin Chim Acta. 1997;257(2):157-80.

Diamandis EP. Serum proteomic profiling by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry for cancer diagnosis: next steps. Cancer Res. 2006; 66(11): 5540-5541.

Dilloo D, Rill D, Entwistle C, Boursnell M, Zhong W, Holden W, Holladay M, Inglis S, Brenner M. A novel herpes vector for the high-efficiency transduction of normal and malignant human hematopoietic cells. Blood. 1997;89(1):119-27.

Duan X, Yarmush DM, Berthiaume F, Jayaraman A, Yarmush ML. A mouse serum two-dimensional gel map: application to profiling burn injury and infection. Electrophoresis. 2004;25(17):3055-65.

Eccles SA. Targeting key steps in metastatic tumour progression. Current opinion in genetics & development 2005; 15:77-86.

Eggert AA, Schreurs MW, Boerman OC, Oyen WJ, de Boer AJ, Punt CJ, Figdor CG, Adema GJ. Biodistribution and vaccine efficiency of murine dendritic cells are dependent on the route of administration. Cancer Res. 1999;59(14):3340-5.

El-Aneed A. Current strategies in cancer gene therapy. Eur J Pharmacol. 2004;498(1-3):1-8.

Engwegen JY, Gast MC, Schellens JH et al. Clinical proteomics: searching for better tumour markers with SELDI-TOF mass spectrometry. Trends Pharmacol Sci 2006; 27 (5):251-259.

Faca VM, Song KS, Wang H, Zhang Q, Krasnoselsky AL, Newcomb LF, Plentz RR, Gurumurthy S, Redston MS, Pitteri SJ, Pereira-Faca SR, Ireton RC, Katayama H, Glukhova V, Phanstiel D, Brenner DE, Anderson MA, Misek D, Scholler N, Urban ND, Barnett MJ, Edelstein C, Goodman GE, Thornquist MD, McIntosh MW, DePinho RA, Bardeesy N, Hanash SM. A mouse to human search for plasma proteome changes associated with pancreatic tumor development. PLoS Med. 2008 Jun 10;5(6):e123.

Farrell RJ, Peppercorn MA. Ulcerative colitis. Lancet 2002; 359:331–40.

Fearon ER, Dang CV.Cancer genetics: tumor suppressor meets oncogene. Curr Biol. 1999 Jan 28;9(2):R62-5.

Fedi P, Tronick SR, Aaronson SA. Growth Factors. In: JF Holland; RC Bast; DL Morton et al., editors Cancer Medicine. Baltimore, MD: Williams and Wilkins; 1997; p. 41-64.

Ferrara N. VEGF: an update on biological and therapeutic aspects. Curr Opin Biotechnol. 2000 Dec;11(6):617-24.

Fung ET, Weinberger SR, Gavin E, Zhang F. Bioinformatics approaches in clinical proteomics. Expert Rev Proteomics. 2005;2(6):847-62.

Furlaneto CJ, Campa A. A novel function of serum amyloid A: a potent stimulus for the release of tumor necrosis factor-alpha, interleukin-1beta, and interleukin-8 by human blood neutrophil. Biochem Biophys Res Commun. 2000; 268:405–408.

Fynan EJ, Webster RG, Fuller DH, Haynes JR, Santoro JC, Robinson HL. DNA vaccines: A novel approach to immunization. Int J Immunopharmacol. 1995;17(2):97-83.

Geho DH, Liotta LA, Petricoin EF, Zhao W, Araujo RP. The amplified peptidome: the new treasure chest of candidate biomarkers. Curr Opin Chem Biol. 2006; 10(1): 50-55.

Ghezzi P, Bertini R, Bianchi M, Erroi A, Mengozzi M, Delgado R, Giavazzi R, et al. Serum amyloid A induction in tumorbearing mice: evidence for a tumor-derived mediator. Int J Immunopathol Pharmacol 1993; 6:169–186.

Gilar M, Bouvier ES, Compton BJ. Advances in sample preparation in electromigration, chromatographic and mass spectrometric separation methods. J Chromatogr A. 2001 Feb 16;909(2):111-35.

Gold P, Freedman SO. Demonstration of Tumour-specific Antigens in Human Colonic Carcinomata by Immunological Tolerance and Absorption Techniques. J. exp. Med. 1965; 121, 439.

Green DR, Reed JC. Mitochondria and apoptosis. Science 1998; 281 (5381):1309-1312.

Gutfeld O, Prus D, Ackerman Z, Dishon S, Linke RP, Levin M, Urieli-Shoval S. Expression of serum amyloid A, in normal, dysplastic, and neoplastic human colonic mucosa: implication for a role in colonic tumorigenesis. J Histochem Cytochem. 2006;54(1):63-73.

Gutmann DH, Baker SJ, Giovannini M, Garbow J, Weiss W. Mouse models of human cancer consortium symposium on nervous system tumors. Cancer Res. 2003 Jun 1;63(11):3001-4.

Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc Natl Acad Sci U S A. 2000 Aug 15;97(17):9390-5.

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 1999;17(10):994-9.

Hanahan D, Folkman J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. Cell 1996; 86 (3):353-364.

Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000 7;100(1):57-70.

Hao CY, Moore DH, Wong P, Bennington JL, Lee NM, Chen LC. Alteration of gene expression in macroscopically normal colonic mucosa from individuals with a family history of sporadic colon cancer. Clin Cancer Res. 2005; 11(4):1400-7.

Hara O, Murphy REP, Whitehead AS, FitzGerald O, and Bresnihan B. Local expression of the serum amyloid A and formyl peptide receptor-like 1 genes in synovial tissue is associated with matrix metalloproteinase production in patients with inflammatory arthritis. Arthritis Rheum. 2004; 50:1788–1799.

Harlozinska A. Progress in molecular mechanisms of tumor metastasis and angiogenesis. Anticancer Res 2005; 25 (5):3327-3333.

Harris CC. p53 tumor suppressor gene: from the basic research laboratory to the clinic--an abridged historical perspective. Carcinogenesis 1996; 17 (6):1187-1198.

Heck AJ, Krijgsveld J. Mass spectrometry-based quantitative proteomics. Expert Rev Proteomics. 2004 Oct;1(3):317-26.

Heinrich S, Jochum W, Graf R and Clavien P. Portal vein ligation and partial hepatectomy differentially influence growth of intrahepatic metastasis and liver regeneration in mice?. Journal of Hepatology, 2006;45(1): 35-42.

Hilario M, Kalousis A, Pellegrini C, Müller M. Processing and classification of protein mass spectra. Mass Spectrom Rev. 2006 May-Jun;25(3):409-49.

Hoffman and Stroobant. Mass spectrometry: Principles and applications. 2007

Holyoake T, Jiang X, Eaves C et al. Isolation of a highly quiescent subpopulation of primitive leukemic cells in chronic myeloid leukemia. Blood 1999; 94 (6):2056-2064.

Honda K, Hayashida Y, Umaki T, Okusaka T, Kosuge T, Kikuchi S, Endo M, Tsuchida A, Aoki T, Itoi T, Moriyasu F, Hirohashi S, Yamada T. Possible detection of pancreatic cancer by plasma protein profiling. Cancer Res. 2005 Nov 15;65(22):10613-22.

Hope KJ, Jin L, Dick JE. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. Nat Immunol 2004; 5 (7):738-743.

Hornberger J, Cosler LE, Lyman GH. Economic analysis of targeting chemotherapy using a 21-gene RT-PCR assay in lymph-node-negative, estrogen-receptor-positive, early-stage breast cancer. Am J Manag Care. 2005 May;11(5):313-24.

Hortin GL. The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. Clin Chem. 2006 Jul;52(7):1223-37. Epub 2006 Apr 27.

Huang SK, Darfler MM, Nicholl MB, You J, Bemis KG, Tegeler TJ, Wang M, Wery JP, Chong KK, Nguyen L, Scolyer RA, Hoon DS. LC/MS-based quantitative proteomic analysis of paraffin-embedded archival melanomas reveals potential proteomic biomarkers associated with metastasis. PLoS One. 2009;4(2):e4430. Epub 2009 Feb 16.

Hudelist G, Pacher-Zavisin M, Singer CF, Holper T, Kubista E, Schreiber M, Manavi M, Bilban M, Czerwenka K. Use of high-throughput protein array for profiling of differentially expressed proteins in normal and malignant breast tissue.

Hung KE, Kho AT, Sarracino D, Richard LG, Krastins B, Forrester S, Haab BB, Kohane IS, Kucherlapati R. Mass spectrometry-based study of the plasma proteome in a mouse intestinal tumor model. J Proteome Res. 2006; 5(8): 1866-1878.

Hutchens TW, Yip TT. New desorption strategies for the mass spectrometric analysis of macromolecules. Rapid Communications in Mass Spectrometry 1993; 7:576-580.

Iversen C, Lancashire L, Waddington, Michael F, Stephen, Ball G. Identification of Enterobacter sakazakiifrom closely related species: The use of Artificial Neural Networks in the analysis of biochemical and 16S rDNA data. 2006.

Jinong Li, Zhen Zhang, Jason Rosenzweig, Young Y. Wang, and Daniel W. Chan. Proteomics and Bioinformatics Approaches for Identification of Serum Biomarkers to Detect Breast Cancer. Clin Chem 2002 48: 1296-1304.

Juan HF, Chen JH, Hsu WT, Huang SC, Chen ST, Yi-Chung Lin J, Chang YW, Chiang CY, Wen LL, Chan DC, Liu YC, Chen YJ. Identification of tumor-associated plasma biomarkers using proteomic techniques: from mouse to human. Proteomics 2004; 4(9): 2766-2775.

Kaneti J, Winikoff Y, Zimlichman S, Shainkin-Kestenbaum R. Importance of serum amyloid A (SAA) level in monitoring disease activity and response to therapy in patients with prostate cancer. Urol Res. 1984; 12(5): 239-241.

Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem. 1988 Oct 15;60(20):2299-301.

Karsan A, Blonder J, Law J, Yaquian E, Lucas DA, Conrads TP, Veenstra T. Proteomic analysis of lipid microdomains from lipopolysaccharide-activated human endothelial cells. J Proteome Res. 2005 Mar-Apr;4(2):349-57.

Keshavarzian A, Fusunyan RD, Jacyno M, Winship D, MacDermott RP, Sanderson IR. Increased interleukin-8 (IL-8) in rectal dialysate from patients with ulcerative colitis: evidence for a biological role for IL-8 in inflammation of the colon. Am J Gastroenterol 1999; 94:704–12.

Knochenmuss R, Zenobi R. MALDI ionization: the role of in-plume processes. Chem Rev. 2003;103(2):441-52.

Kolch W, Mischak H, Pitt AR. The molecular make-up of a tumour: proteomics in cancer research. Clin Sci (Lond). 2005 May;108(5):369-83.

Kolstoe SE, Ridha BH, Bellotti V, Wang N, Robinson CV, Crutch SJ, Keir G, Kukkastenvehmas R, Gallimore JR, Hutchinson WL, Hawkins PN, Wood SP, Rossor MN, Pepys MB. Molecular dissection of Alzheimer's disease neuropathology by depletion of serum amyloid P component. Proc Natl Acad Sci U S A. 2009;106(18):7619-23.

Komita H, Zhao X, Taylor JL, Sparvero LJ, Amoscato AA, Alber S, Watkins SC, Pardee AD, Wesa AK, Storkus WJ. CD8+ T-cell responses against hemoglobin-beta prevent solid tumor growth. Cancer Res. 2008 Oct 1;68(19):8076-84.

Kong B, Huang S, Wang W, Ma D, Qu X, Jiang J, Yang X, Zhang Y, Wang B, Cui B, Yang Q. Arsenic trioxide induces apoptosis in cisplatin-sensitive and -resistant ovarian cancer cell lines. Int J Gynecol Cancer. 2005 Sep-Oct;15(5):872-7.

Kosari F, Parker A. S, Kube D M, Lohse C M, Leibovich B C, Blute M L, Cheville J C and Vasmatzis G. Clear cell renal cell carcinoma: gene analyses identify a potential signature for tumor aggressiveness. Clin. Cancer Res. 2005; 11(14): 5128–5139.

Kovacevic A, Hammer A, Stadelmeyer E, Windischhofer W, Sundl, M, Ray A, Schweighofer N, Friedl G, Windhager R, Sattle W and Malle E. Expression of serum amyloid A transcripts in human bone tissues, differentiated osteoblast-like stem cells and human osteosarcoma cell lines. J. Cell. Biochem. 2008; 103: 994–1004.

Kravitz MS, Pitashny M, Shoenfeld Y. Protective molecules--C-reactive protein (CRP), serum amyloid P (SAP), pentraxin3 (PTX3), mannose-binding lectin (MBL), and apolipoprotein A1 (Apo A1), and their autoantibodies: prevalence and clinical significance in autoimmunity. J Clin Immunol. 2005 Nov;25(6):582-91.

Krogh A.What are artificial neural networks? Nat Biotechnol. 2008 Feb;26(2):195-7.

Krutchinsky AN, Ayed A, Donald LJ, Ens W, Duckworth HW, Standing KG. Studies of noncovalent complexes in an electrospray ionization/time-of-flight mass spectrometer. Methods Mol Biol. 2000;146:239-49.

Kulasingam V, Diamandis EP. Tissue culture-based breast cancer biomarker discovery platform. Int J Cancer. 2008 ;123(9):2007-12.

Lancashire L, Schmid O, Shah H et al. Classification of bacterial species from proteomic data using combinatorial approaches incorporating artificial neural networks, cluster analysis and principal components analysis. Bioinformatics 2005; 21 (10):2191-2199.

Lancashire LJ, Mian S, Rees RC, Ball GR. Preliminary Artificial Neural Network Analysis of SELDI Mass Spectrometry Data For The Classification Of Melanoma Tissue. 17th European Simulation Multiconference 2003; 131-135.

Latchman DS. Gene delivery and gene therapy with herpes simplex virus-based vectors. Gene. 2001;264(1):1-9.

Levy Y, Symons H, Chowdhury W, Rodriguez R and Fuchs E. Adoptive immunotherapy plus cryotherapy in murine model of metastatic colon cancer. Journal of Clinical Oncology, 2006; 24(18S): 2520.

Li J, Hong M, Pan T. linical significance of VEGF-C and VEGFR-3 expression in non-small cell lung cancer. J Huazhong Univ Sci Technolog Med Sci. 2006;26(5):587-90.

Li J, Steen H, and Gygi SP. Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents: the yeast salinity stress response. Mol. Cell. Proteomics. 2003;2, 1198 – 1204.

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

Li L, Tang H, Wu Z, Gong J, Gruidl M, Zou J, Tockman M, Clark RA. Data mining techniques for cancer detection using serum proteomic profiling. Artif Intell Med. 2004;32(2):71-83.

Liang L, Qu L and Ding Y. Protein and mRNA characterization in human colorectal carcinoma cell lines with different metastatic potentials. Cancer Invest. 2007; 25: 427–434.

Liao Q, Zhao L, Chen X, Deng Y, Ding Y. Serum proteome analysis for profiling protein markers associated with carcinogenesis and lymph node metastasis in nasopharyngeal carcinoma. Clin Exp Metastasis. 2008;25(4):465-76. Epub 2008 Mar 21.

Lilley KS, Razzaq A, Dupree P.Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. Curr Opin Chem Biol. 2002 Feb;6(1):46-50.

Lin SM, Haney RP, Campa MJ, Fitzgerald MC, Patz EF. Characterising phase variations in MALDI-TOF data and correcting them by peak alignment. Cancer Inform. 2005;1:32-40.

Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer. 2005 Nov;5(11):845-56.

Lundberg AS and Weinberg RA. Functional inactivation of the retinoblastoma protein requires sequential modification by at least two distinct cyclin-cdk complexes. Mol Cell Biol. 1998;18(2):753-61.

Macfarlane RD and Torgerson DF. Ion-Temperature Measurements of Fission-Fragment Tracks in CsBr Films. Phys. Rev. Lett. 1976;36:486-488.

Malle E, Sodin-Semrl S, Kovacevic A. Serum amyloid A: an acute-phase protein involved in tumour pathogenesis. Cell Mol Life Sci. 2009 66, 9-26.

Marcuson R, Burbeck SL, Emond RL, Latter GI, Aberth W. Normalization and reproducibility of mass profiles in the detection of individual differences from urine. Clin Chem. 1982 Jun;28(6):1346-8.

Martin DB, Nelson PS, From genomics to proteomics: techniques and applications in cancer research, Trends in Cell Biology, 2001;11(11):60-5.

Martin DB, Nelson PS. From genomics to proteomics: techniques and applications in cancer research. Trends Cell Biol. 2001;11(11):S60-5.

Matharoo-Ball B, Hughes C, Lancashire L, Tooth D, Ball G, Creaser C, Elgasim M, Rees R, Layfield R and Atiomo W. Characterization of biomarker in polycystic syndrome (PCOS)using multiple distinct proteomic platforms. J Proteome Res 2007; 6 (8): 3321-3328.

Matharoo-Ball B, Miles AK, Creaser CS, Ball G, Rees R. Serum biomarker profiling in cancer studies: a question of standardisation? Vet Comp Oncol. 2008 Dec;6(4):224-47.

Matharoo-Ball B, Ratcliffe L, Lancashire L, Ugurel Selma, Miles AK, Weston DJ, Rees R, Schadendorf D, Ball G and Creaser CS. Diagnostic biomarkers differentiating metastatic melanoma patients from healthy controls identified by an integrated MALDI-TOF mass spectrometry/bioinformatic approach. Proteomics Clin. Appl. 2007; 1(6): 605-620.

Mattfeldt T, Kestler HA, Hautmann R, Gottfried HW. Prediction of prostatic cancer progression after radical prostatectomy using artificial neural networks: a feasibility study. BJU Int. 1999 Aug;84(3):316-23.

Mayordomo JI, Rivera F, Díaz-Puente MT, Lianes P, Colomer R, López-Brea M, López E, Paz-Ares L, Hitt R, García-Ribas I, et al. Improving treatment of chemotherapy-induced

neutropenic fever by administration of colony-stimulating factors. J Natl Cancer Inst. 1995;87(11):803-8.

Mian S, Ball G, Hornbuckle J, Holding F, Carmichael J, Ellis I, Ali S, Li G, McArdle S, Creaser C, Rees R. A prototype methodology combining surface-enhanced laser desorption/ionization protein chip technology and artificial neural network algorithms to predict the chemoresponsiveness of breast cancer cell lines exposed to Paclitaxel and Doxorubicin under in vitro conditions. Proteomics 2003; 3(9): 1725-1737.

Mian S, Ugurel S, Parkinson E, Schlenzka I, Dryden I, Lancashire L, Ball G, Creaser C, Rees R, Schadendorf D. Serum proteomic fingerprinting discriminates between clinical stages and predicts disease progression in melanoma patients. J Clin Oncol. 2005; 23(22): 5088-5093.

Michael A, Ball G, Quatan N, Wushishi F, Russell N, Whelan J, Chakraborty P, Leader D, Whelan M, Pandha H. Delayed disease progression after allogeneic cell vaccination in hormone-resistant prostate cancer and correlation with immunologic variables. Clin Cancer Res. 2005;11(12):4469-78.

Migita, K., Kawabe, Y., Tominaga, M., Origuchi, T., Aoyagi, T. and Eguchi, K. Serum amyloid A protein induces production of matrix metalloproteinases by human synovial fibroblasts. Lab. Invest. 1998; 78:535–539.

Misek DE, Imafuku Y, Hanash SM. Application of proteomic technologies to tumor analysis. Pharmacogenomics. 2004 Dec;5(8):1129-37. Review.

Moingeon P. Cancer vaccines. Vaccine. 2001 Jan 8;19(11-12):1305-26.

Moyer JS, Maine G, Mulé JJ. Early vaccination with tumor-lysate-pulsed dendritic cells after allogeneic bone marrow transplantation has antitumor effects. Biol Blood Marrow Transplant. 2006 Oct;12(10):1010-9.

Muehlbauer PM, Schwartzentruber DJ. Cancer vaccines. Semin Oncol Nurs. 2003 Aug;19(3):206-16.

Müller TF, Trösch F, Ebel H, Grüssner RW, Feiber H, Göke B, Greger B, Lange H. Pancreas-specific protein (PASP), serum amyloid A (SAA), and neopterin (NEOP) in the diagnosis of rejection after simultaneous pancreas and kidney transplantation. Transpl Int. 1997; 10(3): 185-191.

Negm RS, Verma M, Srivastava S. The promise of biomarkers in cancer screening and detection. Trends Mol Med. 2002 Jun;8(6):288-93.

Nicolette CA, Miller GA. The identification of clinically relevant markers and therapeutic targets. Drug Discov Today. 2003 Jan 1;8(1):31-8.

Niederau C, Backmerhoff F, Schumacher B. Inflammatory mediators and acute phase proteins in patients with Crohn's disease and ulcerative colitis. Hepatogastroenterology 1997; 44:90–107.

Nishie A, Masuda K, Otsubo M, Migita T, Tsuneyoshi M, Kohno K, Shuin T, Naito S, Ono M and Kuwano M.(2001) High expression of the Cap43 gene in infiltrating macrophages of human renal cell carcinomas. Clin. Cancer Res. 2001; 7:2145–2151.

Nishiyama E, Iwamoto N, Kimura M, Arai H. Serum amyloid P component level in Alzheimer's disease. Dementia. 1996 Sep-Oct;7(5):256-9.

Noursadeghi M, Bickerstaff MC, Gallimore JR, Herbert J, Cohen J and Pepys MB. Role of serum amyloid P component in bacterial infection: protection of the host or protection of pathogen. Proc Natl Acad Sci U S A 2000; 97(26): 1458-1459.

Obermair A, Handisurya A, Kaider A, Sevelda P, Kölbl H, Gitsch G. The relationship of pretreatment serum hemoglobin level to the survival of epithelial ovarian carcinoma patients: a prospective review. Cancer. 1998;83(4):726-31.

Omenn GS. Strategies for plasma proteomic profiling of cancers.Proteomics. 2006 Oct;6(20):5662-73.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002 May;1(5):376-86.

Onji M, Akbar SM. On dendritic cell-based therapy for cancers. J Zhejiang Univ Sci B. 2005 Jan;6(1):1-3.

Parkinson RJ, Mian S, Bishop MC, Gray T, Li G, McArdle SE, Ali S, Rees RC. Disabled infectious single cycle herpes simplex virus (DISC-HSV) is a candidate vector system for gene delivery/expression of GM-CSF in human prostate cancer therapy. Prostate. 2003 Jun 15;56(1):65-73.

Parle-McDermott A., McWilliam, P, Tighe, O, Dunican, D and Croke, D T. Serial analysis of gene expression identifies putative metastasis-associated transcripts in colon tumour cell lines. Br. J. Cancer 2000; 83: 725–728.

Parmiani G, Sensi M, Castelli C, Rivoltini L, Anichini A. T-cell response to unique and shared antigens and vaccination of cancer patients.Cancer Immun. 2002;2:6.

Parmiani G, Rodolfo M, Melani C. Immunological gene therapy with ex vivo gene-modified tumor cells: a critique and a reappraisal. Hum Gene Ther. 2000;11(9):1269-75.

Pastwa E, Somiari SB , Czyz M , Somiari RI. Proteomics in human cancer research. PROTEOMICS - Clin Apps. 2006;1(1):Pages 4 – 17.

Patel H, Fellowes R, Coade S, Woo P. Human serum amyloid A has cytokine-like properties. Scan J Immunol. 1998; 48:410–418.

Pedersen AE, Buus S, Claesson MH. Treatment of transplanted CT26 tumour with dendritic cell vaccine in combination with blockade of vascular endothelial growth factor receptor 2 and CTLA-4. Cancer Lett. 2006;235(2):229-38.

Pedraza-Fariña LG. Mechanisms of oncogenic cooperation in cancer initiation and metastasis. Yale J Biol Med. 2006;79(3-4):95-103.

Peng J, Stanley AJ, Cairns D, Selby PJ, Banks RE. Using the protein chip interface with quadrupole time-of-flight mass spectrometry to directly identify peaks in SELDI profiles--initial evaluation using low molecular weight serum peaks. Proteomics. 2009;9(2):492-8.

Pepys MB, Baltz M, Gomer K, Davies AJS and Doenhoff M. Serum amyloid P-component is an acute-phase reactant in the mouse. Nature 1979; 278: 259 – 261.

Petricoin EF and Liotta LA. Proteomic approaches in cancer risk and response assessment. Trends Mol Med 2004; 10 (2): 59-64.

Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC and Liotta LA. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002; 359(9306): 572-577.

Petricoin EF, Belluco C, Araujo RP, Liotta LA. The blood peptidome: a higher dimension of information content for cancer biomarker discovery. Nat Rev Cancer. 2006;6(12):961-7.

Pikarsky E, Porat RM, Stein I, Abramovitch R, Amit S, Kasem S, Gutkovich-Pyest E, et al. NF-kB functions as a tumor promoter in inflammation-associated cancer. Nature 2004; 431:461–466.

Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W. And Vogelstein, B. A model for p53-induced apoptosis. Nature 1997; 389: 300–305.

Rai AJ, Vitzthum F. Effects of preanalytical variables on peptide and protein measurements in human serum and plasma: implications for clinical proteomics. Expert Rev Proteomics. 2006;3(4):409-26.

Randolph TW, Yasui Y.Multiscale processing of mass spectrometry data. Biometrics. 2006 Jun;62(2):589-97.

Ransohoff DF. Lessons from controversy: ovarian cancer screening and serum proteomics. J Natl Cancer Inst. 2005; 97(4): 315-319.

Rees RC, McArdle S, Mian S, Li G, Ahmad M, Parkinson R, Ali SA. Disabled infectious single cycle-herpes simplex virus (DISC-HSV) as a vector for immunogene therapy of cancer. Curr Opin Mol Ther. 2002;4(1):49-53.

Rhodes JM, Campbell BJ. Inflammation and colorectal cancer: IBD-associated and sporadic cancer compared. Trends Mol Med 2002; 8:10–6.

Ricci-Vitiani L, Fabrizi E, Palio E, De Maria R. Colon cancer stem cells. J Mol Med. 2009; 87(11):1097-104.

Ricolleau G, Charbonnel C, Lodé L, Loussouarn D, Joalland MP, Bogumil R, Jourdain S, Minvielle S, Campone M, Déporte-Fety R, Campion L, Jézéquel P. 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 Mar;6(6):1963-75.

Rieger PT. The biology of cancer genetics. Semin Oncol Nurs 2004; 20 (3):145-154.

Römpp A, Dekker L, Taban I, Jenster G, Boogerd W, Bonfrer H, Spengler B, Heeren R, Smitt PS, Luider TM. Identification of leptomeningeal metastasis-related proteins in cerebrospinal fluid of patients with breast cancer by a combination of MALDI-TOF, MALDI-FTICR and nanoLC-FTICR MS. Proteomics. 2007; 7(3): 474-481.

Saha A, Chatterjee SK, Foon KA, Primus FJ, Bhattacharya-Chatterjee M. Murine dendritic cells pulsed with an anti-idiotype antibody induce antigen-specific protective antitumor immunity. Cancer Res. 2003 Jun 1;63(11):2844-54.

Sandoval JA, Turner KE, Hoelz DJ, Rescorla FJ, Hickey RJ and Malkas LH. Serum protein profiling to identify high-risk neuroblastoma: preclinical relevance of blood-based biomarkers. J Surg Res. 2007;142(2):268-74.

Sawyers CL. The cancer biomarker problem. Nature. 2008;452(7187):548-52.

Schanzer JM, Fichtner I, Baeuerle PA, Kufer P. Antitumor activity of a dual cytokine/single-chain antibody fusion protein for simultaneous delivery of GM-CSF and IL-2 to Ep-CAM expressing tumor cells. J Immunother 1997; : 477-488.

Schottenfeld, D. and Beebe-Dimmer, J. Chronic inflammation: a common and important factor in the pathogenesis of neoplasia. CA Cancer J. Clin. 2006; 56: 69–83.

Schulte I, Tammen H, Selle H, Schulz-Knappe P. Peptides in body fluids and tissues as markers of disease. Expert Rev Mol Diagn. 2005 Mar;5(2):145-57.

Sechi S, Oda Y. Quantitative proteomics using mass spectrometry. Curr Opin Chem Biol. 2003 Feb;7(1):70-7.

Seibert V, Wiesner A, Buschmann T et al. Surface-enhanced laser desorption ionization time-of-flight mass spectrometry (SELDI TOF-MS) and ProteinChip technology in proteomics research. Pathol Res Pract 2004; 200 (2):83-94.

Semmes OJ, Feng Z, Adam BL, Banez LL, Bigbee WL, Campos D, Cazares LH, Chan DW, Grizzle WE, Izbicka E, Kagan J, Malik G, McLerran D, Moul JW, Partin A, Prasanna P, Rosenzweig J, Sokoll LJ, Srivastava S, Srivastava S, Thompson I, Welsh MJ, White N, Winget M, Yasui Y, Zhang Z, Zhu L. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. Clin Chem. 2005;51(1):102-12.

Shay JW, Bacchetti S. A survey of telomerase activity in human cancer. Eur J Cancer 1997; 33 (5):787-791.

Shay JW, Wright WE. Senescence and immortalization: role of telomeres and telomerase. Carcinogenesis 2005; 26 (5):867-874.

Shiio Y, Aebersold R. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. Nat Protoc 2006;1(1):139-45.

Simpkins F, Czechowicz JA, Liotta L, Kohn EC. SELDI-TOF mass spectrometry for cancer biomarker discovery and serum proteomic diagnostics. Pharmacogenomics. 2005 Sep;6(6):647-53.

Smith A, Eskew JD, Borza CM, Pendrak M, Hunt RC. Role of heme-hemopexin in human T-lymphocyte proliferation. Exp Cell Res. 1997; 232(2): 246-254.

Soma M, Tamaoki T, Kawano H, Ito S, Sakamoto M, Okada Y, Ozaki Y, Kanba S, Hamada Y, Ishihara T, Maeda S. Mice Lacking Serum Amyloid P Component Do Not Necessarily Develop Severe Autoimmune Disease. Biochemical and Biophysical Research Communications 2001;286:200–205.

Sorace JM, Zhan M. A data review and re-assessment of ovarian cancer serum proteomic profiling. BMC Bioinformatics. 2003;4:24.

Srinivas PR, Kramer BS, Srivastava S. Trends in biomarker research for cancer detection. Lancet Oncol. 2001 Nov;2(11):698-704.

Srivastava PK, Udono H.Heat shock protein-peptide complexes in cancer immunotherapy. Curr Opin Immunol. 1994;6(5):728-32.

Steel DM, Whitehead AS. The major acute phase reactants: C-reactive protein serum amyloid P component and serum amyloid A protein. Immunol Today 1994; 15:81–8

Suh KW, Piantadosi S, Yazdi HA, Pardoll DM, Brem H, Choti MA. Treatment of liver metastases from colon carcinoma with autologous tumor vaccine expressing granulocyte-macrophage colony-stimulating factor. J Surg Oncol. 1999;72(4):218-24.

Taguchi F, Solomon B, Gregorc V, Roder H, Gray R, Kasahara K, Nishio M, Brahmer J, Spreafico A, Ludovini V, Massion PP, Dziadziuszko R, Schiller J, Grigorieva J, Tsypin M,Hunsucker SW, Caprioli R, Duncan MW, Hirsch FR, Bunn PA Jr, Carbone DP. Mass spectrometry to classify non-small-cell lung cancer patients for clinical outcome after treatment with epidermal growth factor receptor tyrosine kinase inhibitors: a multicohort cross-institutional study. J Natl Cancer Inst. 2007 Jun 6;99(11):838-46.

Taraboletti G, Micheletti G, Dossi R, Borsotti P, Martinelli M, Fiordaliso F, Ryan AJ, Giavazzi R. Potential antagonism of tubulin-binding anticancer agents in combination therapies. Clin Cancer Res. 2005 Apr 1;11(7):2720-6.

Taylor BS, Varambally S, Chinnaiyan AM. A systems approach to model metastatic progression. Cancer Res. 2006 Jun 1;66(11):5537-9.

Tennent GA, Lovat LB, Pepys MB. Serum amyloid P component prevents proteolysis of the amyloid fibrils of Alzheimer disease and systemic amyloidosis. Proc Natl Acad Sci U S A. 1995 May 9;92(10):4299-303.

Tiss A, Smith C, Camuzeaux S, Kabir M, Gayther S, Menon U, Waterfield M, Timms J, Jacobs I, Cramer R. Serum peptide profiling using MALDI mass spectrometry: avoiding the pitfalls of coated magnetic beads using well-established ZipTip technology. Proteomics. 2007;7 Suppl 1:77-89.

Todryk S, McLean C, Ali S, Entwistle C, Boursnell M, Rees R, Vile R. Disabled infectious single-cycle herpes simplex virus as an oncolytic vector for immunotherapy of colorectal cancer. Hum Gene Ther. 1999 Nov 20;10(17):2757-68.

Tolosano E, Altruda F. Hemopexin: structure, function, and regulation. DNA Cell Biol. 2002; 21(4): 297-306.

Torgrip RJO, Aberg M, Karlberg B, Jacobsson SP. Peak alignment using reduced set mapping.Journal of Chemometrics. 2003;17:573–582.

van den Bemd GJ, Krijgsveld J, Luider TM, van Rijswijk AL, Demmers JA, Jenster G. Mass spectrometric identification of human prostate cancer-derived proteins in serum of xenograft-bearing mice. Mol Cell Proteomics. 2006;5(10):1830-9.

Vapnik V, Vashist A. A new learning paradigm: learning using privileged information. Neural Netw. 2009;22(5-6):544-57.

Veenstra TD, Prieto DA, Conrads TP. Proteomic patterns for early cancer detection. Drug Discov Today. 2004;9(20):889-97.

Vogelstein B, Kinzler KW.Cancer genes and the pathways they control. Nat Med. 2004;10(8):789-99.

Volpert OV, Dameron KM, Bouck N. Sequential development of an angiogenic phenotype by human fibroblasts progressing to tumorigenicity. Oncogene 1997; 14 (12):1495-1502.

Vreugdenhil A C, Dentener MA, Snoek A M, Greve JW and Buurman WA. Lipopolysaccharide binding protein and serum amyloid A secretion by human intestinal epithelial cells during the acute phase response. J. Immunol. 1999; 163:2792–2798.

Wagner E, Culmsee C, Boeckle S. Targeting of polyplexes: toward synthetic virus vector systems. Adv Genet. 2005; 53:333-54.

Wagner W, Hermann R, Hartlapp J, Esser E, Christoph B, Müller MK, Krech R, Koch O. Prognostic value of hemoglobin concentrations in patients with advanced head and neck cancer treated with combined radio-chemotherapy and surgery. Strahlenther Onkol. 2000; 176(2): 73-80.

Wang JC, Dick JE. Cancer stem cells: lessons from leukemia. Trends Cell Biol 2005; 15 (9):494-501.

Wang X, Zbou C, Qiu G, Fan J, Tang H, Peng Z. Screening of new tumor suppressor genes in sporadic colorectal cancer patients. Hepatogastroenterology. 2008 Nov-Dec;55(88):2039-44.

Wattiez R, Falmage P. Proteomics of bronchoalveolar lavage fluid. J Chromatogr B.2005;815:169-78. 80

Weinberg RA. Oncogenes and tumor suppressor genes. CA Cancer J Clin 1994; 44 (3):160-170.

Weinberg RB. Identification of functional domains in the plasma apolipoproteins by analysis of inter-species sequence variability. J Lipid Res. 1994;35(12):2212-22.

Wen JB, Nie B and Jainq B. Establishment and comparison of liver metastasis models of two colorectal carcinoma cell lines in mice. Wen JB, Nie B and Jainq B. Nan Fang Yi Ke Da Xue Xue Bao 2007; 27 (7): 1044-1046.

Werb Z. ECM and cell surface proteolysis: regulating cellular ecology. Cell 1997; 91 (4):439-442.

White CN, Chan DW, Zhang Z. Bioinformatics strategies for proteomic profiling. Clin Biochem 2004; 37 (7):636-641.

Whiteaker JR, Zhang H, Zhao L, Wang P, Kelly-Spratt KS, Ivey RG, Piening BD, Feng LC, Kasarda E, Gurley KE, Eng JK, Chodosh LA, Kemp CJ, McIntosh MW, Paulovich AG. Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. J Proteome Res. 2007;6(10):3875-6.

Wong JW, Cagney G, Cartwright HM. SpecAlign--processing and alignment of mass spectra datasets. Bioinformatics. 2005;21(9):2088-90.

Woong-Shick A, Sung-Pil P, Su-Mi B, Joon-Mo L, Sung-Eun N, Gye-Hyun N, Young-Lae C, Ho-Sun C, Heung-Jae J, Chong-Kook K, Young-Wan K, Byoung-Don H, Hyun-Sun J. Identification of hemoglobin-alpha and -beta subunits as potential serum biomarkers for the diagnosis and prognosis of ovarian cancer. Cancer Sci. 2005; 96(3): 197-201.

Wu CH, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu ZZ, Ledley RS, Lewis KC, Mewes HW, Orcutt BC, Suzek BE, Tsugita A, Vinayaka CR, Yeh LS, Zhang J, Barker WC. The Protein Information Resource: an integrated public resource of functional annotation of proteins. Nucleic Acids Res. 2002 Jan 1;30(1):35-7.

Wu WW, Wang G, Baek SJ, Shen RF. Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. J Proteome Res 2006;5(3):651-8.

Wyllie AH, Kerr JF, Currie AR. Cell death: the significance of apoptosis. Int Rev Cytol 1980; 68:251-306.

Xu L, Badolato R, Murphy WJ, Longo DL, Anver M, Hale S, Oppenheim JJ, Wang JM. A novel biologic function of serum amyloid A. Induction of T lymphocyte migration and adhesion. J Immunol. 1995; 155:1184–1190.

Yu W, Wu B, Lin N, Stone K, Williams K, Zhao H. Detecting and aligning peaks in mass spectrometry data with applications to MALDI. Comput Biol Chem. 2006;30(1):27-38.

Zahedi K and Whitehead AS. Regulation of mouse serum amyloid P gene expression by cytokines in vitro. Biochim Biophys Acta 1993; 1176(1-2): 162-168.

Zhang JC, Claffey K, Sakthivel R, Darzynkiewicz Z, Shaw DE, Leal J, Wang YC, Lu FM, McCrae KR. Two-chain high molecular weight kininogen induces endothelial cell apoptosis and inhibits angiogenesis: partial activity within domain 5. FASEB J. 2000 Dec;14(15):2589-600.

Zhang JC, Qi X, Juarez J, Plunkett M, Donaté F, Sakthivel R, Mazar AP, McCrae KR. Inhibition of angiogenesis by two-chain high molecular weight kininogen (HKa) and kininogen-derived polypeptides. Can J Physiol Pharmacol. 2002 Feb;80(2):85-90.

Zhang M, Rosen JM. Stem cells in the etiology and treatment of cancer. Curr Opin Genet Dev 2006a; 16 (1):60-64.

Zhang X, Leung SM, Morris CR et al. Evaluation of a novel, integrated approach using functionalized magnetic beads, bench-top MALDI-TOF-MS with prestructured sample supports, and pattern recognition software for profiling potential biomarkers in human plasma. J Biomol Tech 2004; 15 (3):167-175.

Zhang Z, Bast RC, Jr., Yu Y et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Res 2004b; 64 (16):5882-5890.

Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M, Snyder M. Global analysis of protein activities using proteome chips. Science. 2001 Sep 14;293(5537):2101-5.

Communications resulting from study

Publications:

Baharak Vafadar-Isfahani, Stephanie Laversin, Murrium Ahmad, Graham Ball, Clare Coveney, Christophe Lemetre, Amanda Miles, Gerhard van Schalkwyk, Robert Rees and Balwir Matharoo-Ball1. Serum biomarkers which correlate with failure to respond to immunotherapy and tumor progression in a murine colorectal cancer model. *Proteomics Clin. Appl.* 2010, 4, 1–15.

Christopher Medway, Hui Shi, James Bullock, Holly Black, Kristelle Brown, Baharak Vafadar-isfahani, Balwir Matharoo-Ball, Graham Ball, Robert Rees, Noor Kalsheker, Kevin Morgan: Using In silico LD clumping and meta-analysis of genomewide datasets as a complementary tool to investigate and validate new candidate biomarkers in Alzheimer's disease. *Int J Mol Epidemiol Genet.* 2010;1(2):134-144.

Abstracts:

Identification of proteomic phonemic fingerprint associated with regression or progression after immunotherapy in a murine colorectal cancer model. The 5th Joint BSPR/EBI Proteomics Meeting 2008, Cambridge, UK. Delivered 20 minute short talk.

Application of proteomics and bioinformatics to identify discriminatory patterns in murine colorectal cancer model. The 7th East Midlands Proteomic Workshop 2008, Nottingham, UK. Poster

Murine serum cancer biomarkers to predict response to immunotherapy. The 4th Joint BSPR/EBI Proteomics Meeting 2007, Cambridge, UK. Poster

Murine serum cancer biomarkers to predict response to immunotherapy. The 6th East Midlands Proteomic Workshop 2007, Nottingham, UK. Delivered 20 minute short talk.