PLASMA PROTEIN PROFILING FOR BLADDER CANCER BIOMARKER DISCOVERY USING UPLC-HDMS^{E} LABEL-FREE QUANTITATION

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Amirmansoor Hakimi

Department of Cancer Studies and Molecular Medicine

University of Leicester

2013

Dedicated to:

My beloved wife, Maryam

Abstract

PLASMA PROTEIN PROFILING FOR BLADDER CANCER BIOMARKER DISCOVERY USING UPLC-HDMS^E LABEL-FREE QUANTITATION

Amirmansoor Hakimi

In the UK, bladder cancer is the 4th most common cancer in men and 11th most common in women. In 2010, just over 10,000 new cases were diagnosed and 4,900 deaths were recorded. At their first diagnosis, the majority of bladder cancer patients (75-85%) present with non-muscle invasive disease. In 50-70% of these patients the tumour will recur and in 10-20% of them it will progress to muscle invasive disease.

Mass spectrometry based proteomics has been chosen for clinical biomarker discovery due to its ability to perform qualitative and quantitative protein profiling on clinical samples. In total 90 plasma samples were used in this study in two groups of disease and control. An optimised and evaluated UPLC-IMS-DIA-MS^E label-free quantitation method was used for plasma protein profiling. To our knowledge, this is the first report investigating the biomarkers of bladder cancer incorporating label-free quantitation and UPLC-IMS-DIA-MS^E methodology. To assess expression level of proteins of samples in different groups a plan consisting of four data processing packages was used. Each of the packages uses different statistical means by which to identify proteins and/or compare expression levels alteration.

Optimisation of the methodology helped in the thorough investigation of the plasma proteome with coverage of up to five orders of magnitude of plasma protein concentration dynamic range. In total, 11 proteins were found as possible markers of diagnosis for bladder cancer. Four of these candidates (afamin, alpha 1-B-glycoprotein, apolipoprotein-A1 and haptoglobin) were previously reported to be urinary markers of bladder cancer. CRP was overexpressed when plasma samples from patients with low grade-Ta tumours were compared to every other sample and may be used as a diagnostic marker. Similarly, afamin and haptoglobin were overexpressed in plasma samples from patients with high grade-high stage tumours when compared to samples from patients with high grade-low stage disease.

Acknowledgements

It is a great pleasure to acknowledge with thanks the assistance I have received from the following individuals during my time at the department of Cancer Studies and Molecular Medicine, University of Leicester.

First of all, I would like to thank my supervisors Dr George DD Jones and Dr Don JL Jones for the deft ways in which they challenged and supported me throughout the years of this work. I wish to extend my gratitude to Dr Don JL Jones, who was not only my supervisor during this journey but also a helpful source in every aspect of my life in the UK. It was my honour to be his first PhD student and have the privilege to work under his supervision. His technical advices were critical in my overcoming the dilemmas that originated mainly from my inexperience. I appreciate all his contributions of time and ideas to make my PhD experience productive and stimulating. He showed me how fascinating the mass spectrometry field is. I would also like to offer my appreciation to Dr George DD Jones for his guidance to clarify ambiguous issues from time to time, sharing with me interesting and creative ideas. I will never forget the day I went to his office, asked about a PhD position and he pleasantly said “why don’t we arrange a meeting to discuss this further”.

I would specially like to thank all the Mass Spectrometry lab members at the RKCSB for being always prepared to extend their help to me in a pleasant and friendly manner. Furthermore, thanks to the Chemoprevention group members and people at the Biocenter who had kept me company, both scientifically and non-scientifically. I have to give special thanks to Dr Leonie Norris and Dr Emma Horner-Glister for their “always

available" kind helps. Their assistance especially toward the end of this work is highly appreciated.

I would also like to acknowledge our collaborator in the University of Torino, Turin, Italy who provided me precious samples for this work. I am grateful to Felix A Achana who helped me finding what I was looking for in my data with his excellent knowledge in biomedical statistics. I also gratefully acknowledge the HOPE Against Cancer charity for partially funding this project.

I am also taking this opportunity to thank my family who have been a source of encouragement and inspiration to me throughout my life. My special thanks to my parents for being so supportive.

My sincere thank goes to whom I am greatly indebted, my mother and father-in-law for their invaluable support through the years of my graduate studies in the UK without which none of these would be achieved. Your support is greatly appreciated.

Last but not least, immense appreciation to the most important person in my life, my wife, Maryam. You have supported me in the hardest times and believed in me even when I did not believe in myself. Your tireless effort enabled me to take the time necessary to complete this work. Your faithful and everlasting support, wholehearted encouragement, endless patience and overwhelmingly unconditional love, are so appreciated. This would not have been possible without you, **Thank you**.

Table of content

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Abbreviations

2D-PAGE	2 Dimension Polyacrylamide Gel Electrophoresis
A1BG	Alpha-1 Beta Glycoprotein
ACN	Acetonitrile
ADH	Alcohol Denohydrogenase
Apo-A1	Apolipoprotein A1
ASM	Auxiliary Solvent manager
BCA	Bicinchoninic Acid
BCG	Bacille Calmette Guerin
BPI	Base Peak Intensity
BSM	Binary Solvent Manager
BTA	Bladder Tumour Antigen
CID	Collision Induced Dissociation
CIS	Carcinoma *In Situ*
CRP	C-Reactive Protein
CSF	Cerebrospinal Fluid
DC	Direct Current
DIA	Data Independent Acquisition
DIGE	Difference Gel Electrophoresis
DTT	DL-Dithiothreitol
ECL	Enhanced Chemiluminescence
EGFR	Epidermal Growth Factor Receptor
ELISA	Enzyme Linked Immunosorbent Assay
EMRT	Exact Mass Retention Time
ESI	Electro Spray Ionisation
FA	Formic Acid

FDA	Food and Drug Administration
FDR	False Discovery Rate
FGF	Fibroblast Growth Factor
FGFR3	Fibroblast Growth Factor Receptor 3
GFP	Glu-Fibrinopeptide B
GO	Gene Ontology
HPLC	High Performance Liquid Chromatography
HUPO	Human Proteome Organization
IAA	Iodoacetamide
IARC	International Agency for Research on Cancer
IEF	Iso Electric Focusing
IMS	Ion Mobility Separation
IT	Immunotherapy
ITI-H3	Inter alpha Trypsin Inhibitor Heavy chain 3
iTRAQ	isobaric Tag for Relative and Absolute Quantitation
LC	Liquid Chromatography
Leu Enk	Leucine Enkephalin
LOD	Limit of Detection
MALDI	Matrix Assisted Laser Desorption Ionisation
MAPED	Microwave Assisted Protein Enzymatic Digestion
MAPK	Mitogen-Activated Protein Kinase
MCM5	Mini-Chromosome Maintenance protein 5
MID	Muscle Invasive Disease
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
MS^E	Mass Spectrometry Elevated

Nano-LC	Nano Liquid Chromatography
NMID	Non-Muscle Invasive Disease
NMP-22	Nuclear Matrix Protein-22
NMSC	Non-Melanoma Skin Cancer
PAH	Polycyclic Aromatic Hydrocarbons
PDT	Photodynamic Therapy
PLGS	ProteinLynx Global Server
ppm	Parts Per Million
PSA	Prostate Specific Antigen
PTM	Post Translational Modification
QToF	Quadrupole Time of Flight
RF	Radio Frequency
RT	Retention Time
SDS	Sodium Dodecyl Sulfate
SEER	Surveillance, Epidemiology and End Results
SELDI	Surface Enhanced Laser Desorption Ionisation
SM	Sample Manager
SRM	Selected Reaction Monitoring
TIC	Total Ion Count
TIS	Tumour *In Situ*
TURBT	Transurethral Resection of Bladder Tumour
UPLC	Ultra Performance Liquid Chromatography
VDBP	Vitamin-D Binding Protein
VEGF	Vascular Epidermal Growth Factor
WB	Western Blot
ZAG	Zinc Alpha-2 Glycoprotein

Chapter One

INTRODUCTION

1 Introduction

1.1 Cancer

In the latest report published by the World Health Organization (WHO) cancer is the second most common cause of death worldwide, responsible for 21% of all mortalities. Cardiovascular disease with 48% is the most common reason for death, whilst chronic respiratory disease with 12% and diabetes with 3.5% are the third and fourth, respectively (World Health Organisation, 2013). It has been reported that one in every three people will develop cancer throughout their lifetime with the highest number of incidents in the developed countries (Figure 1.1) (*Cancer Research UK.*2013). This high level of risk and burden drives the need to tackle this 21st century problem.

Death from infectious disease, which was previously the main cause of death worldwide, has been drastically reduced due to a number of initiatives (sanitation, vaccinations and public health controls). However, partly as a consequence of this reduction, populations are living longer which in itself is leading to a significant increase in the proportions developing cancer in later life (*Cancer Research UK.*2013). Globally, over 12.5 million new cases of cancer were diagnosed in 2008 and more than 7.5 million deaths from the disease were recorded (accounting for ~60% mortality) in the same year. It has been estimated that the worldwide annual incidence will rise to more than 22 million newly diagnosed cancer sufferers by 2030 (World Health Organisation, 2013).

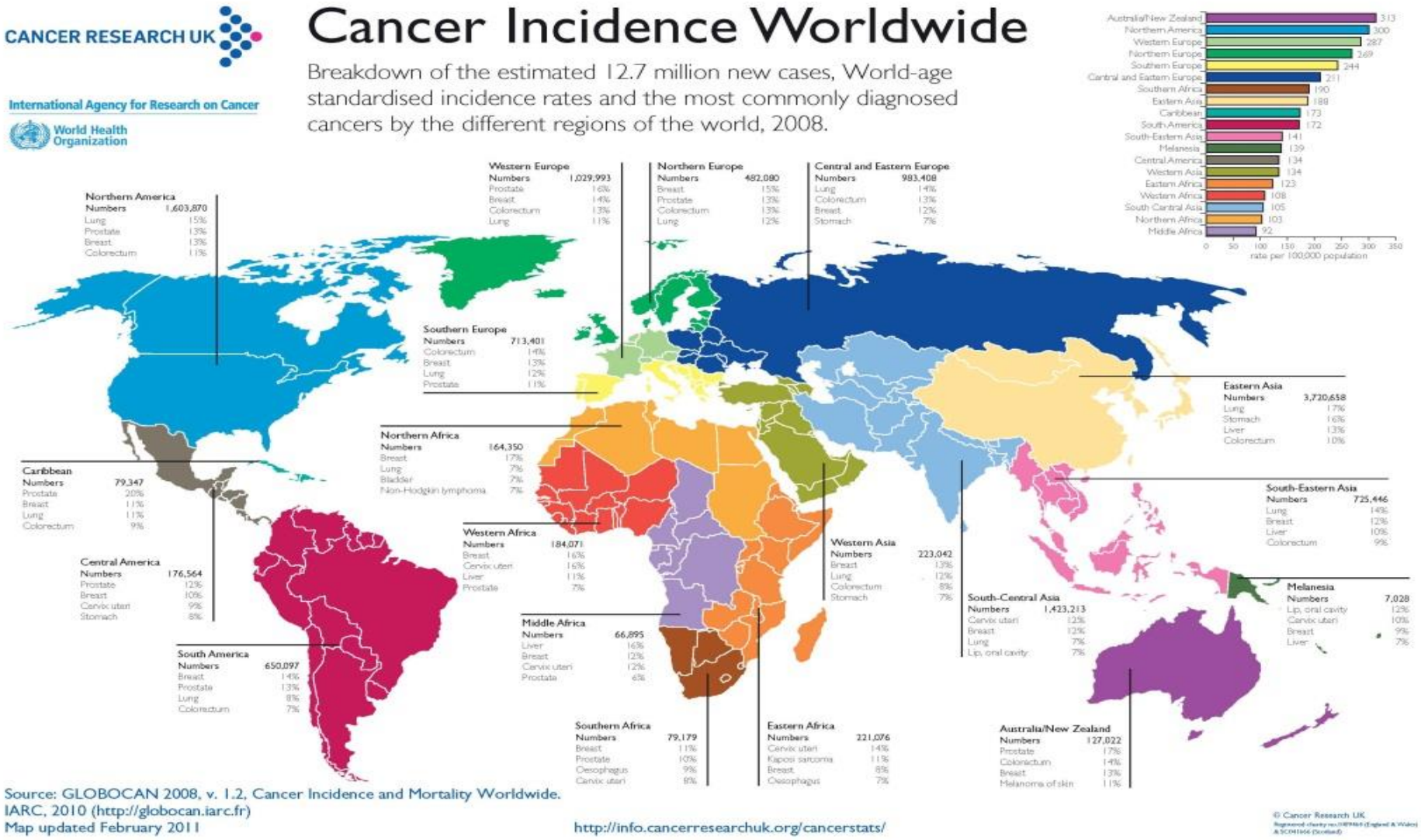

Figure 1.1 Worldwide cancer incidence (*Cancer Research UK*.2013)

1.2 Bladder cancer

1.2.1 Background

Bladder cancer is the 9th most common cancer worldwide (excluding non-melanoma skin cancers (NMSC)) with over 350,000 new cases diagnosed every year (*International Agency for Research on Cancer.*2012). Developed countries in Western Europe and North America have the highest rate of bladder cancer due to their industrialisation and high number of cigarette smokers. Africa and some Middle Eastern countries also have high rate as a consequence of Schistosomiasis being endemic (*International Agency for Research on Cancer.*2012). In the United States it is estimated that approximately 20% of the 70,000 new cases of bladder cancer diagnosed in 2010 (*ca.* 14,500) resulted in death (*Surveillance Epidemiology and End Results.*2010). In the UK, just over 10,000 new cases of cancer were diagnosed in 2010 and more than 4,900 deaths from the disease were recorded in the same year. Although the mortality rate in the UK is very high (~50%), the incidence rate is still lower than many neighbouring European countries (*Cancer Research UK.*2013). In the UK, bladder cancer is the 4th most common cancer in men and 11th most common in women. Men are more likely to develop bladder cancer than women, with the ratio of 5:2 (*Cancer Research UK.*2013); this gender discrepancy possibly being due to higher occupational exposure to carcinogens (e.g. aromatic amines in dyes) which is considered one of the main risk factors for bladder cancer (Kogevinas *et al.*, 2003). Bladder cancer is most common in the elderly (>50) (Jemal *et al.*, 2009) and is rare in people under the age of 50. Since early 1970s the one year relative survival rate has increased by 10%, whilst the 5 year survival rate has increased by almost 25% in men and 5% in women (the higher rate in men being due to earlier diagnosis (Sant *et al.*, 2009). The same pattern also applies for the 10 year survival rate

in both men and women. The increase in the 5 and 10 years survival rates (men and women overall), that has also been seen in the other types of cancer (*Cancer Research UK.*2013), could be due to the improvements in detection and subsequent cancer treatment i.e. improved modern radiotherapy techniques, superior adjuvant therapies (such as hormone-therapy) and more effective chemotherapy drugs (Lerner *et al.*, 2008).

1.2.2 Classification

A. Disease type

The most common type of cancer in urinary system is Transitional Cell Carcinoma (TCC) which accounts for almost 90% of cases in the UK (*Cancer Research UK.*2013). Typically bladder cancer arises as a consequence of the exposure of normal bladder epithelium to carcinogens present in the urine; consequently the site of tumour origin is the bladder's inner epithelial lining. Based on the extent of penetration of cancer cells into the layers of bladder tissue, bladder cancer has been clinically classified into three main groups: non-muscle invasive (previously known as superficial disease), muscle invasive and metastatic (Waters, 1996).

1) Non-muscle invasive disease

Non-muscle invasive disease (NMID) is where the tumour is present in the mucosa (stage Ta and CIS) or submucosa (stage T1) of the bladder wall (Figure 1.2).

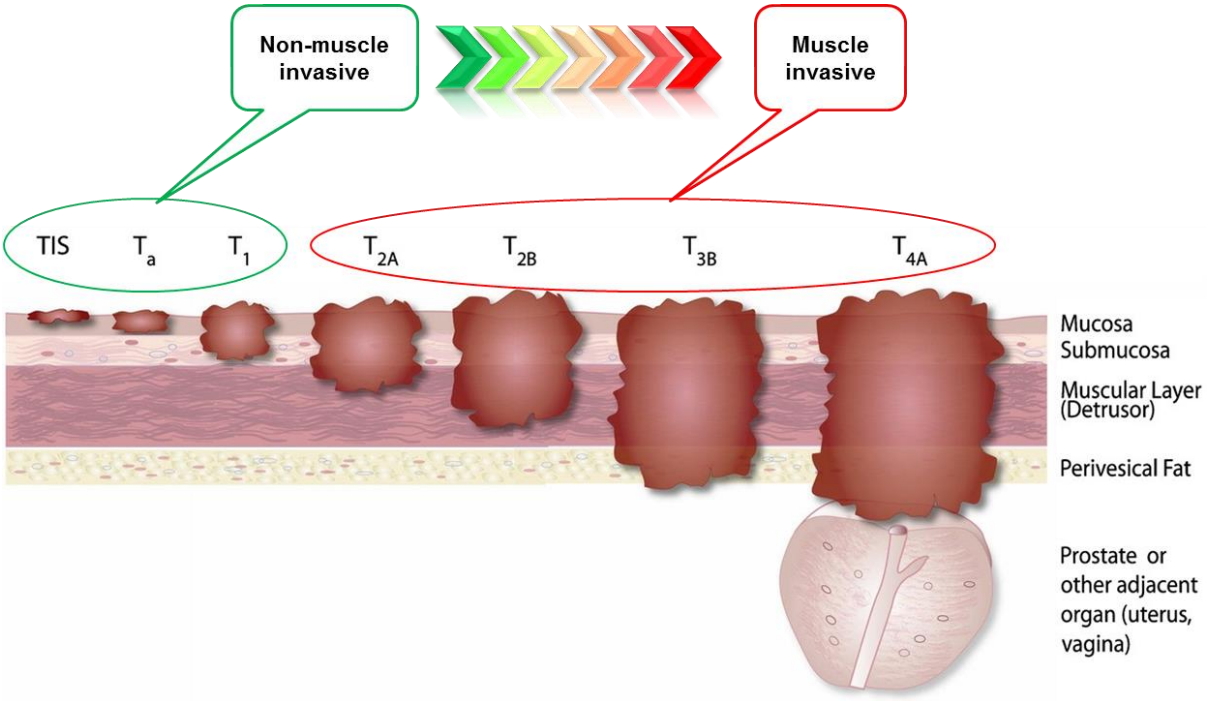

Figure 1.2 Bladder cancer disease type and tumour stages. Tumour staging is based on tumour cell's penetration into the layers of bladder wall (adopted from Matta *et al.*, 2005).

In terms of presentation, most bladder cancer patients (75-85%) at their first diagnosis present with NMID which pathologists classify as either Ta or T1 (accounting for 70% and 20% of patients with NMID, respectively) (Black *et al.*, 2006); the remaining 10% present with Carcinoma *in situ* (CIS) (Kirkali *et al.*, 2005). Melicow (1952) was the first to introduce the term CIS, also called tumour *in situ* (TIS) (MELICOW, 1952). CIS is a flat tumour in which urothelial cells proliferate in a disordered manner with high tendency to progress to the muscle invasive form of the disease. NMID tumours are mostly (~75%) low grade tumours (grades 1) with low recurrence rates and low rates of progression (top route in Figure 1.3); this group effectively represents the patients that achieve immediate disease free status post-treatment.

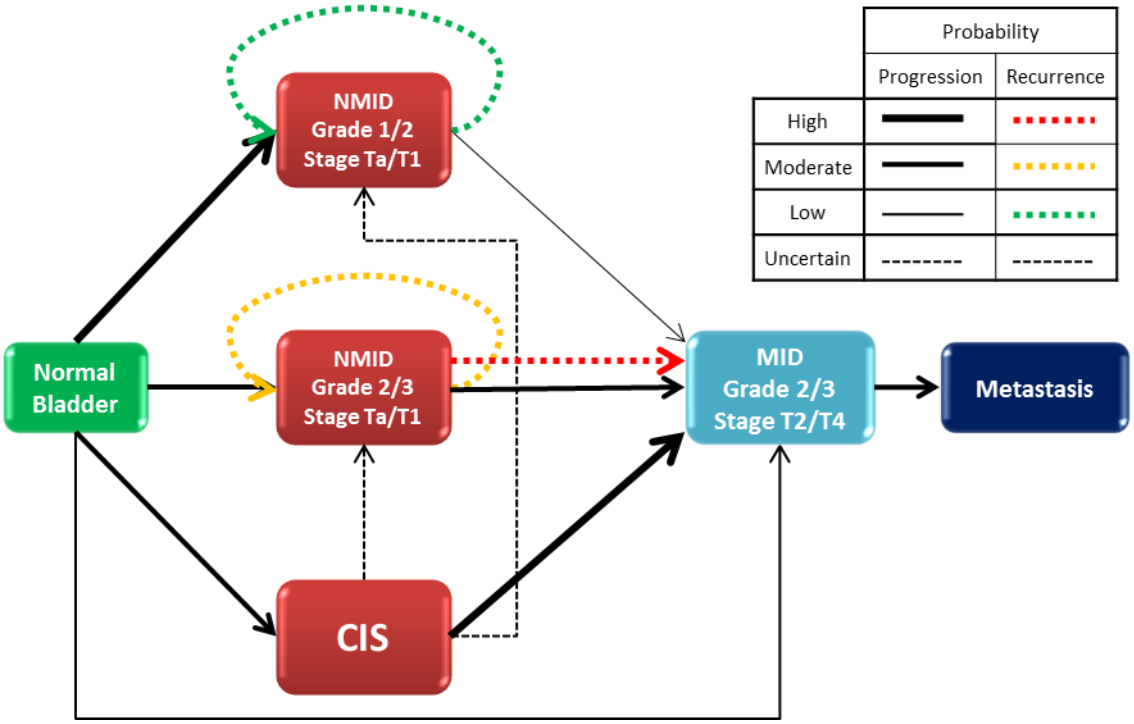

Figure 1.3 Different types of bladder cancer and the progression of the disease. Probability of the disease progression is shown by thickness of the arrows and recurrence rates by different colours.

The other 25% of NMID (the middle route in Figure 1.3) have high grade tumours (grades 2 and 3) with higher rates of recurrence (60-70%) post treatment with approximately 20-30% of recurring tumours progressing to muscle invasive disease, with consequently high rates of metastasis (~80%) and poor prognosis (Grignon, 2009).

2) Muscle invasive disease (MID)

MID is the more aggressive type of bladder cancer in which cancer cells have invaded deeper into the layers of bladder wall including the muscular and perivesical fat layers (Figure 1.2). Approximately 15-25% of bladder cancer patients are diagnosed with muscle invasive disease (stages T2-T4) at their first presentation (Black *et al.*, 2006). In MID patients the tumour is usually classified as high grade (grades 2 and 3). For stages T2 and T3 the cancerous cells are still retained in the bladder wall whilst for in T4 they

have invaded nearby organs such as prostate, uterus and vagina (T4a) and pelvic wall and abdominal wall (T4b) (Waters, 1996).

3) Metastasis

Metastasis is the most advanced type of the disease with a very low survival rate. Metastatic disease is characterised by the formation of secondary tumours established at a distance from the site of the primary tumour. Typically, in metastatic disease tumour cells invade the lymph nodes (N1-N3, depending on the distance from the original tumour) or develop a new tumour in distant organs i.e. typically prostate in men and ovary in women (M1).

Other routes for the development and progression of bladder cancer have been suggested (Figure 1.3), such as direct conversion of normal urothelial cells to high grade MID or progression from CIS to low and high grade NMID, but none of these routes are yet confirmed.

B. Grading

Tumours grades are indicators of malignancy level representing a morphological variation in the distribution of dysplasia (Waters, 1996). Bergkvist and co-workers were the first group to introduce a grading system for bladder cancer tumours in 1965 (Bergkvist *et al.*, 1965). It was then developed by Mostofi (1973) and used by WHO as a common reference (Mostofi, 1973). Based on the 1973 system, tumours are classified into three grades called grade 1, 2 and 3 with the least degree of cellular anaplasia grouped into grade 1 and the highest level into grade 3. This system was then converted to a two-group grading system of low and high grade (Montironi & Lopez-Beltran, 2005) which is known as the WHO-2004 grading system. In both of the systems

with increasing grade, from grade 1 to 3 in WHO-1973 or low grade to high grade in WHO-2004, the degree of nuclear crowding, alteration in cell polarity, irregularity in cell size, nuclear pleomorphism and alteration in chromatin pattern are more apparent. A comparison of 1973 and 2004 grading systems is shown in Table 1.1.

<table>
<tr><th>Description</th><th>WHO 1973</th><th>WHO 2004</th></tr>
<tr><td>Non-cancerous cells</td><td>Papilloma</td><td>Urothelial Papilloma</td></tr>
<tr><td rowspan="2">Slow growing and unlikely to spread, Well differentiated</td><td rowspan="2">Grade 1</td><td>PUNLMP</td></tr>
<tr><td rowspan="2">Low Grade</td></tr>
<tr><td>Grow slowly, Poorly differentiated</td><td>Grade 2</td></tr>
<tr><td>Grow fast, High recurrence tendency</td><td>Grade 3</td><td>High Grade</td></tr>
</table>

Table 1.1 Comparison of 1973 and 2004 bladder cancer tumour grading systems. PUNLMP stands for Papillary Urothelial Neoplasm of Low Malignant Potential.

C. Staging

Staging a tumour, regardless of organ, was first introduced by Pierre Denoix in 1950. Three main characteristics were defined for this purpose, (*i*) **t**umour size and if it invades to the neighbour tissues **(T)**, (*ii*) lymph **n**ode involvement and how advanced it is **(N)** and finally (*iii*) if the tumour progressed to neighbouring or distant organs and developed **m**etastasis there **(M)**. This method is called **TNM** staging of cancer and was recognized and developed by International Union Against Cancer (UICC, derived from French translation of the organization name) and is globally used to date. Figure 1.4 shows TNM classification in detail for bladder cancer.

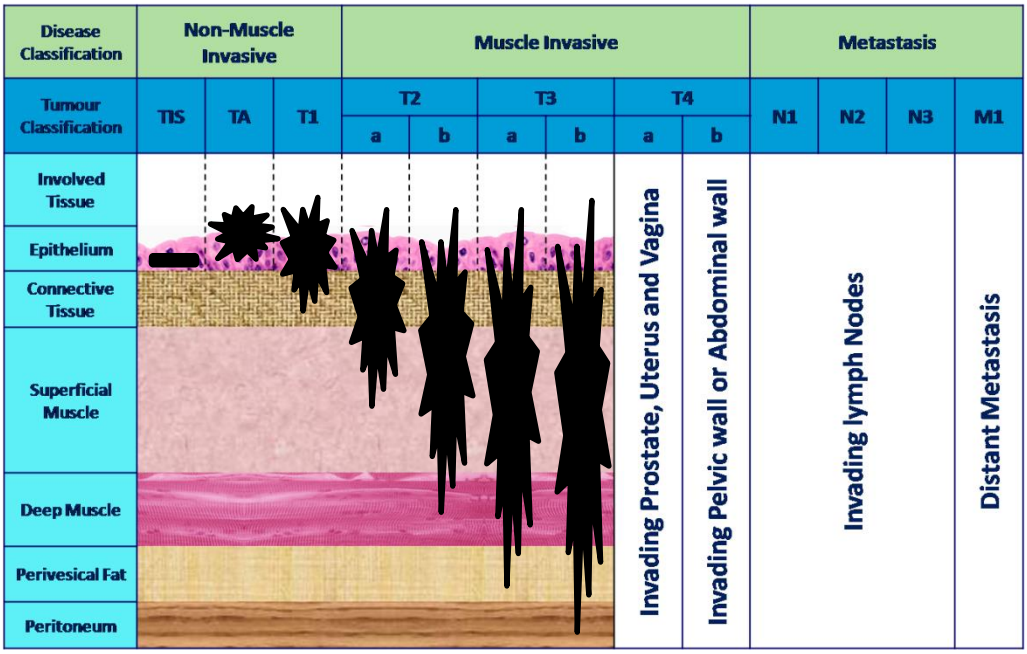

Figure 1.4 TNM classification of malignant tumours (7th edition, 2009)

A. **T** for existence of solid tumour and based on level of invasion into the bladder wall:

- **a.** **TIS** Tumour *in situ* also Carcinoma *in situ* (CIS)
- **b.** **Ta** for non-invasive papillary carcinoma cases
- **c.** **T1** for tumours that invade into sub-epithelial connective tissue
- **d.** **T2** when tumour invades muscle:
  - **i.** **T2a** if tumour invades superficial muscle (inner half)
  - **ii.** **T2b** if tumour invades deep muscle (outer half)
- **e.** **T3** Tumour invades perivesical tissue:
  - **i.** **T3a** microscopic invasion
  - **ii.** **T3b** macroscopic invasion (extravesical mass)
- **f.** **T4** when tumour invades any of the surrounding organs:
  - **i.** **T4a** for if the tumour invades prostate, uterus or vagina
  - **ii.** **T4b** for if the tumour invades pelvic wall or abdominal wall

B. **N** for when tumour invades lymph nodes in the region:
    a. **N1** if metastasis is developed in a single lymph node with the maximum size of 2 cm
    b. **N2** for metastasis in a single lymph node with a size of > 2 cm but less than 5 cm or if multiple lymph nodes are involved but none of them are more than 5 cm in size
    c. **N3** when metastasis develops in a lymph node which is bigger than 5 cm in size
C. **M** for the tumours with metastasis in distant organs:
    a. **M1** for developing metastases in distant organs

There is an alternative, less complicated method of staging called the Surveillance, Epidemiology and End Results (SEER) program which group tumours based on their location:

A. ***In Situ*** refers to a tumour that is in the layer of cells in which it originates.
B. **Localized** is where tumour cells have not spread to other organs and the cancer only develops in the organ that it originated in.
C. **Regional** refers to when cancerous cells spread out to surrounding lymph nodes or tissues and organs.
D. **Distant** is used when the cancer has progressed into distant organ or lymph nodes.
E. **Unknown** is used where there is not sufficient information for staging.

1.2.3 Causes and Risk factors

Environmental factors are believed to have carcinogenic characteristics as a consequence of their enforcing chemical metabolic changes within the body. Binding of an organic compound to DNA which leads to DNA adduct formation and ultimately mutation, is one of the most common and best known mechanisms of carcinogenesis. Polycyclic Aromatic Hydrocarbons (PAH), a common and well-studied class of carcinogens, are produced mainly in heavy industries such as oil refinery, aluminium factories and are also present in cigarette smoke. There may be more than 100 potential carcinogens within cigarette smoke which enter the systemic circulation after absorption by the lungs (Hecht, 1999). These chemicals require metabolic activation to mediate their carcinogenic effects. Following their absorption, they pass through the kidneys and reach to the bladder where they are in direct contact with the bladder urothelial cells. Carcinogens are able to enter the cells where their metabolic activation leads to formation of DNA adducts (La & Swenberg, 1996). Altered/mutated DNA may lead to uncontrolled proliferation resulting in a mass of tumour cells with mutated DNA (Weinberg, 2006).

a) Smoking

As for many other types of cancer, smoking is a risk factor for bladder cancer and indeed it is considered the primary risk factor for bladder cancer in Western Europe. Approximately 60% of men and 30% of women with bladder cancer are cigarette smokers (*Cancer Research UK.*2013). Thompson *et al*. (1987) showed that tumours are in higher grades in bladder cancer patients who smoked cigarettes compared to age-matched non-smoker cases of bladder cancer (Thompson *et al.*, 1987). In a study carried out by European researchers (Bjerregaard *et al.*, 2006) it has been shown that the risk of

developing bladder cancer increases among adults which have been exposed to the tobacco smoke environmentally during their childhood (second hand smoke from parents). In another study Samanic *et al*. (2006) found higher levels of carcinogens for black tobacco compared to blond tobacco (Samanic *et al.*, 2006). They reported that black tobacco smokers (northern Europe and North America) have a higher risk of developing bladder cancer (2-3 fold more so) compared to the blond tobacco smokers in southern European countries and South America. Increased risk of developing cancer (lung, bladder and kidney cancer) is reported in adults where their mother smoked at the time of pregnancy (*in utero*) and during breast feeding (Hemminki & Chen, 2006). Smokers that were diagnosed with bladder cancer also have the risk of developing secondary cancers (mainly lung cancer) (Ray *et al.*, 2010). Although cessation of smoking cigarette has been proved to reduce the risk of bladder cancer, Lerner *et al*. (2010) showed that the risk of developing bladder cancer remains higher when compared with patients who were non-smokers (Lerner *et al.*, 2010).

b) Occupational exposure

A strong risk factor for bladder cancer is industrial exposure to carcinogens. Higher risks of bladder cancer amongst dye industrial workers were reported for the first time by Rehn in 1895. Exposure through occupation is considered a cause for developing bladder cancer in 5 to 25% of cases. Moreover, in developed countries the risk factor is approximately 5-10% whilst in developing countries where the occupations which utilise carcinogens are more prevalent and occupational exposure is less rigorously regulated, the risk factors are much greater (Kogevinas *et al.*, 2003). A significant increase in the risk of developing bladder cancer was found amongst workers of aluminium industry which is due to the release of PAH from the aluminium production

cycle (Guzzo *et al.*, 2008). PAH may account for more than 4% of bladder cancer cases among men in Western Europe (Kogevinas *et al.*, 2003).

c) Hair dyes

Increased incidence of bladder cancer in hairdressers was investigated by international agency for research on cancer (*International Agency for Research on Cancer.*2012). Their report highlighted that hair dyes which contain aromatic amines are the primary reason for bladder cancer in this industry. This was then confirmed by a number of subsequent epidemiological investigations (Reulen *et al.*, 2008; Harling *et al.*, 2010). On the other hand, a Swedish group of researchers suggested that there is no link between the hairdressers who used hair dyes and bladder cancer (Czene *et al.*, 2003). In another study carried out by Klesh *et al.* (2008) the effects of hair dyes for personal users (with less frequent exposure to the hair dyes compared to hairdressers) was investigated (Kelsh *et al.*, 2008). They were unable to find any relationship between hair dyes and bladder cancer. Consequently, the authors reported the exposure time as a key factor in increasing the risk of developing bladder cancer.

d) Schistosomiasis

Schistosoma Hematobium is one of three subtypes of Schistosome belonging to a group of parasites which infect the human body. During its life cycle Schistosoma Hematobium infects the human host by directly penetrating the skin and enters the human circulatory system. The parasite can then pass through the kidney before reaching the urinary bladder where it can reside in the bladder wall causing inflammation. Inflammatory cells such as macrophage and neutrophils have the ability to initiate tumorigenesis by inducing mutagenic DNA damage (Weitzman & Stossel, 1981)

and by activation of polycyclic aromatic hydrocarbon and DNA adduct formation (described earlier in this Chapter). Schistosomiasis is correlated with squamous cell carcinoma of the bladder. Egypt, due to the lack of an effective health system, has the highest number of cases of Schistosomiasis in the world with 27 patients per 100,000 populations (*Cancer Research UK.*2013). Mostafa *et al*., (1999) studied the link between other types of Schistosomia and cancer and found association between Schistosomia Japonicum and liver and colorectal cancer (Mostafa *et al.*, 1999).

e) Secondary cancers

The risk of developing bladder cancer as a secondary cancer is increasing because of two reasons: firstly, the improvements in the cancer treatment of patients lead to higher survival rates and consequently longer periods of treatment, and secondly, the anatomical location of the urinary bladder in the abdomen. Patients with a different type of cancer i.e. prostate, ovary or cervix have a greater chance of developing secondary bladder cancer due to the anatomical location of the primary cancer site (adjacent to bladder) and the exposure of the bladder to radiotherapy as part of their primary cancer's treatment.

f) Other factors

Drinking more than 10 cups of coffee per day can slightly increase the risk of developing bladder cancer among the patients who never smoked (Sala *et al.*, 2000). The same relationship has been investigated by (Villanueva *et al.*, 2006) amongst smokers. Their results showed that the smokers who drank 5 cups of coffee per day have higher risk of developing bladder cancer. In another study carried out in the United States, Jiang *et al.* (2008) found that the risk of developing bladder cancer could

be slightly reduced among the people who drank a reasonable amount of water every day (Jiang *et al.*, 2008). Conversely, chlorine which is used for disinfecting water has been shown to increase the risk of developing bladder cancer (Villanueva *et al.*, 2006).

1.2.4 Diagnosis and Treatment

Haematuria, the presence of blood in voided urine, is one of the first symptoms of bladder cancer. It is painless and although sometimes it changes the urine colour to bright red, visible by the patient, urinary analysis to check for the presence of blood cells in the urine is the standard method of diagnosis. There are other symptoms such as changes in frequency and urgency of urination and also pain at the time of urine voiding. The final decision of whether a patient has bladder cancer, or whether any the above mentioned symptoms are due to an alternative etiology such as infection, is made by a clinician following cystoscopy. Cystoscopy is a procedure in which a camera is passed through the urethra into the bladder to visualise the actual tumour. Transurethral Resection of Bladder Tumour (TURBT) is the surgical procedure to remove cancer tumour tissue for histopathological analysis. Patients with NMID are primarily managed/treated by TURBT followed by chemotherapy (Sylvester *et al.*, 2006). Based on the clinical and pathological assessment, NMID patients may receive further chemotherapy or immunotherapy. For the patients with MID there are two options, either cystectomy or organ-preserving treatment. In case of non-metastatic MID (when the tumour is in stages T2-T4a, N0, M0) the gold standard choice in many clinics (especially in the US) is radical cystectomy (Parekh *et al.*, 2006) which usually entails the complete removal of the bladder with urinary diversion. An alternative treatment method for MID patients by which the bladder can be preserved, is radiotherapy (RT) with or without chemotherapy (CT) (James *et al.*, 2012). Either one of these methods

(surgery vs. radiotherapy) has advantages and disadvantages for patients. Radical cystectomy, as a major operative procedure; it has a significant risk of mortality and may adversely affect the quality of life post-operation. RT has none of the cystectomy's disadvantages with similar cure rates (Kotwal *et al.*, 2008), but for up to 50% of patients radiotherapy fails leading to tumour recurrence that ultimately requires follow-up (salvage) radical cystectomy (Quilty *et al.*, 1986). Chemotherapy is used for the bladder cancer patients with metastatic tumours (Raghavan, 2003). Although chemotherapy doubles the survival rate in metastatic patients (Loehrer PJ *et al.*, 1992) ultimately more than 80% of the patients die of the disease.

1.2.5 Molecular and clinical markers

A. Pathogenesis

Clinical and histopathological research on bladder cancer has led to the discovery of two divergent molecular pathways for the disease. A third possibility has also recently arisen, which is still under investigation. The first and the most frequent (70-80%) pathway is the presence of hyperplasia in the normal bladder cells (van Rhijn *et al.*, 2004). Mutation of the Fibroblast Growth Factor Receptor 3 (FGFR3) transforms the hyperplasia to NMID with tumour cells of low grade (G1/G2) and stage (Ta/T1). These tumour cells are able to progress and form muscle invasive disease (Figure 1.5) or a non-muscle invasive high grade tumour (G2/G3) (Knowles, 2007); for the latter the further occurrence of a mutation in the p53 gene of these cells can cause them to progress to MID (van Rhijn *et al.*, 2004).

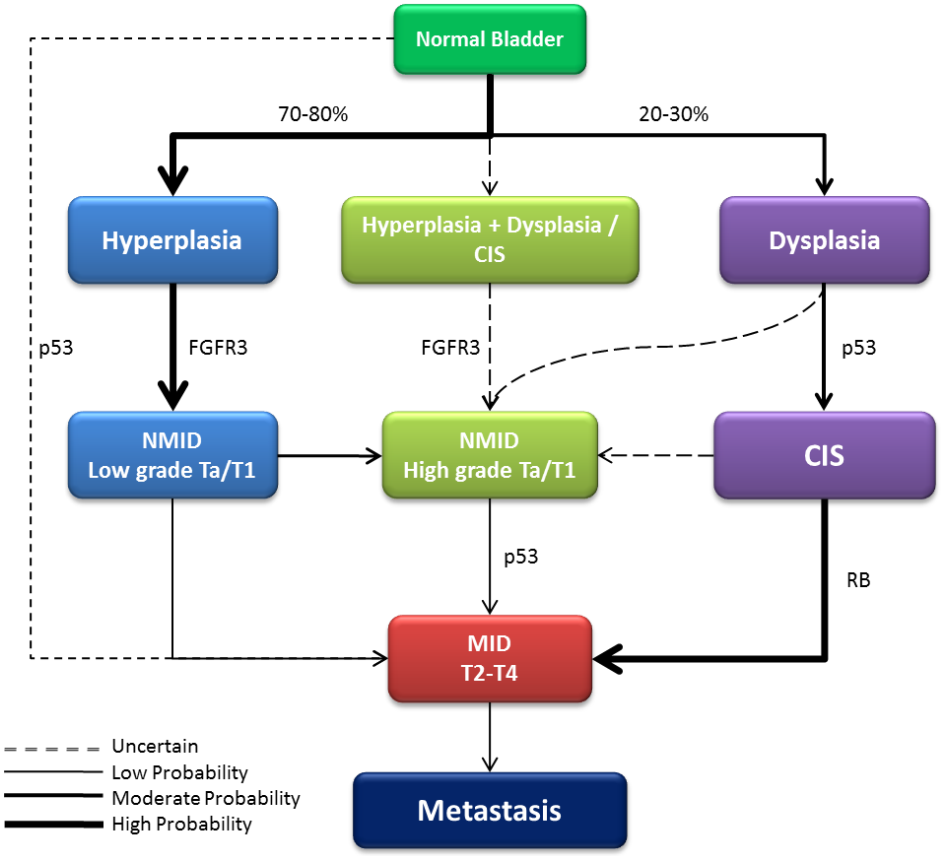

Figure 1.5 Pathogenesis of bladder cancer. The arrow thickness shows the frequency of the occurrence. The thicker the line, the higher the probability. Dashed lines are uncertain routes that are still under investigation.

The second pathway which occurs in 20 to 30% of the cases is abnormal growth of the bladder wall cells (dysplasia). Mutation in p53 genes leads to horizontal growth that forms CIS which has high tendency to progress to the advanced forms of the disease (Theodorescu, 2003). A second mutation in p53 and also in pRB genes converts the CIS cells to the MID with high risk of local/distance metastasis (Knowles, 2007). Researchers are still investigating whether there is a direct link between dysplasia and NMID/high grade tumours or whether dysplasia converts to CIS and then a second mutation develops high grade NMID (Theodorescu, 2003). A third possible route is formation of a complex from hyperplasia, dysplasia and CIS. Mutation in FGFR3 of this complex leads to the development of high grade NMID (van Rhijn *et al.*, 2004). Direct transformation of

normal bladder cells to muscle invasive form of the disease due to the mutation in p53 genes of the healthy bladder wall cells is another route, though this pathway is still an uncertain route (Theodorescu, 2003).

B. Molecular Markers

There is much evidence to imply that both genetic and epigenetic changes are involved in bladder cancer development, and that chromosomal aberration, activation of oncogenes and inactivation of tumour suppressor genes are the main outcomes responsible for transformation. In the papillary low grade tumours of bladder cancer, the Mitogen-Activated Protein Kinase (MAPK) signalling pathway is activated due to the mutation of H-Ras gene (20%) and mutation or over-expression of FGFR3 (80%)(Figure 1.6).

FGFR3, a member of FGF family of proteins, plays key roles in the cell growth regulation, differentiation and angiogenesis. Tyrosine kinase activity of the receptor can be activated as a consequence of receptor dimerisation due to the binding of FGFR3 to a FGF ligand. It has been considered as a marker of progression (van Rhijn *et al.*, 2001) due to the significant association between its mutation and papillary low grade tumours. Rhijn and co-workers compared the rate of bladder cancer recurrence between mutated FGFR3 and its wild type in a single year and found that when FGFR3 is mutated the chance of developing recurred bladder cancer is 80% less compared to its wild type. So, the higher the grade and the stage, the lower the mutation frequency in FGFR3. This has been substantiated by other research groups (Zieger *et al.*, 2005; Lindgren *et al.*, 2006). Mutation analysis on different grades and stages of bladder tumours was also carried out. Junker *et al.* (2008) showed that with increasing tumour

stage the frequency of mutation in FGFR3 decreases. In 69% of Ta tumours, 38% of T1 and a very low percentage of T2 tumours the tumours had their FGFR3 mutated (Junker *et al.*, 2008). The same pattern was seen in grade progression. FGFR3 mutation was reported in 72% of the grade 1 tumours, 56% of grade 2 and 4% of grade 3 tumours (Bakkar *et al.*, 2003).

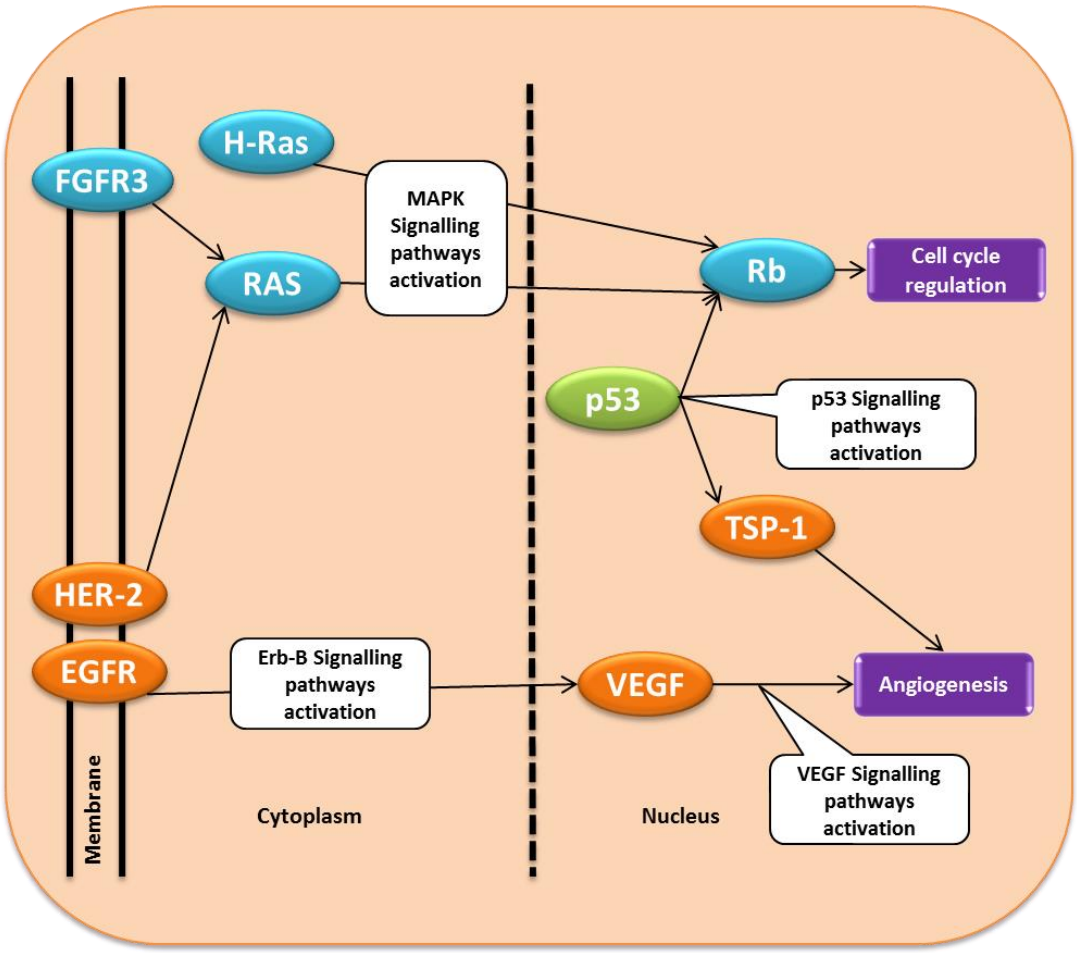

Figure 1.6 Molecular markers of bladder cancer

p53 plays fundamental roles in the maintenance and repair of DNA, regulation of cell cycle and apoptosis; consequently, it's been termed the "guardian of the genome" (Lane, 1992). There are correlations between over-expression of p53 and high grade and stage of the tumour in bladder cancer patients. Mutation in *p53* gene was seen only in 3% of the early stage tumours (Ta) while 65% of the higher grade tumours

(advanced stages) had mutant forms of *p53* (Lokeshwar *et al.*, 2011). High levels of p53 are also detectable in the urine sediments of bladder cancer patients (Schlichtholz *et al.*, 2004).

H-Ras is an oncogene involved in the regulation of the cell cycle. Similar to FGFR3, an up-regulated form of H-Ras activates Ras-MAPK pathway with effects on the cell cycle regulation ultimately driven through activation of pRB (Mitra & Cote, 2009).

EGFR, also known as a marker of progression, is a member of trans-membrane family of proteins playing key roles in the cell growth, motility and differentiation. Overexpression of EGFR activates ErbB and VEGF signalling pathways and is seen in the muscle invasive form of the disease (Kramer *et al.*, 2007). A link between mutant *EGFR* and high stage/grade bladder tumours was also reported (Naik *et al.*, 2011).

HER-2 (Human EGFR-related 2) is a member of EGFR family of proteins. There is disagreement between researchers about the relation between HER-2 and bladder cancer. Mellon *et al.* (1996) could not find any link between HER-2 and progression of the tumour while Kruger *et al*. (2002) showed over-expression of HER-2 in high grade/stage bladder tumours (Mellon *et al.*, 1996; Kruger *et al.*, 2002).

pRB plays key roles in the regulation of cell cycle and apoptosis. Similar to p53, in the higher grade and stages of bladder tumours, pRB is overexpressed which leads to uncontrolled cell growth (Hitchings *et al.*, 2004). It has been seen in both non-muscle invasive and muscle invasive form of the disease (Mitra & Cote, 2009). Loss of RB expression is also a poor prognostic factor in bladder cancer patients.

VEGF is over-expressed in almost all types of cancer including bladder tumours. It is an important protein in angiogenesis and plays a critical role in tumour growth and disease progression. Increased level of VEGF was reported in plasma and urine samples of bladder cancer patients (Zaravinos *et al.*, 2012). Also as an indicator of high grade/stage of the tumour, it is a good marker to predict the metastatic progression of the disease (Garcia-Closas *et al.*, 2007).

Ki-67 is known as marker of proliferation in the nucleus. Depending on how advanced the tumour, the expression of the protein increases. This makes Ki-67 a good marker for prediction of progression (Margulis *et al.*, 2009). Alteration in the expression level of Ki-67 can be measured in urine and also blood sample of bladder cancer patients.

Other possible markers also have effects on the tumorigenesis of bladder cancer such as e-cadherin, matrix metalloproteinase and thrombospondin-1. These markers have a complementary role in the procedure of disease progression i.e. e-cadherin is a marker of metastasis. SIP1 protein, a regulator of epithelial mesenchymal transition (EMT) program and a proposed marker of tumour cell aggressiveness and micro-metastases, has been shown to be of independent prognostic value in bladder cancer (Sayan *et al.*, 2009).

C. Clinical markers

There are a few commercially available urinary tests that are currently used in medical diagnostic laboratories for the detection of bladder cancer.

Hemastix, is a test for the detection of haematuria, detecting the levels of blood (haemoglobin) in the urine. Although presence of blood in the urine does not always indicate bladder cancer, it is a sign for follow up.

NMP–22 Bladder Check. Increased level of Nuclear Matrix Protein-22 (NMP-22) in the urine is a marker for bladder tumours that can be measured by the NMP-22 Bladder Check test. Nuclear matrices are fibre networks playing a structural role in the nucleus and can be found in urine as a consequence of cell death. Combination of NMP-22 Bladder Check and cytology increases the sensitivity of detection of bladder cancer (Hwang *et al.*, 2011).

Bladder Tumour Antigen (BTA) test. When the tumour cells invade the bladder wall, proteolysis occurs which can be detected. The membrane proteins (including complement factor H) are released into the urine and are detectable by BTA test (Raitanen *et al.*, 2001).

ImmunoCyt™ is an immunocytofluorescence technique that uses antibodies labelled with fluorescent markers. Its combination with cytology improves cytology's sensitivity at detecting tumour cells in the urine of patients.

UroVysion™ is a fluorescence *in situ* hybridisation (FISH) assay in urine sample of patients with haematuria. Aneuploidy for chromosomes 3, 7, 17 and loss of the 9p21 can be detected using this assay. UroVysion™ is useful in initial diagnosis of bladder cancer and also in monitoring for tumour recurrence.

MCM5. Mini-chromosome Maintenance protein 5 (MCM5) is involved in chromosome replication. Korkolopoulou *et al.* (2005) showed that MCM-5 is over-expressed in high grade/stage of the bladder tumours making it a marker for muscle invasive disease (Korkolopoulou *et al.*, 2005).

1.2.6 Current Clinical Needs

Presently, all of the patients presenting at the clinic with the non-muscle invasive form of the disease are treated the same, although not all of them will have recurrence in the future. In other words, whilst the treatment is adequate for those patients that do not recur, it is inadequate for those patients that do recur. Hence identifying patients with high chance of recurrence, and the possibility of considering/planning a more aggressive treatment for them, may be beneficial. Accurate prediction of progression would ultimately give clinicians the ability to predict future outcome of the disease for the patients at the time of diagnosis and help them to apply a different plan of treatment to individual patients. Currently, a prognosis for bladder cancer can be predicted using stage, grade and invasion of the tumour cells into the local or distant organs. However, prediction of recurrence (if any) and progression to metastatic disease is limited. A number of molecular markers have been reported for clinical prediction of disease progression. However, none of these markers has yet been approved by regulatory authorities as predictive biomarkers for bladder cancer. For the clinicians to be able to monitor and follow the progress of treatment (post TURBT, radical cystectomy, radiotherapy, chemotherapy and to evaluate the response to BCG treatment) reliable biomarkers are needed.

1.3 Proteomics

1.3.1 Background

Proteomics is a post-genomic technology and was a natural step after the success of the sequencing of the human genome (Figure 1.7). The name proteomics derives from proteome, a mixture of **prote**in and gen**ome** first introduced by Mark Wilkins in 1994 (Wasinger *et al.*, 1995).

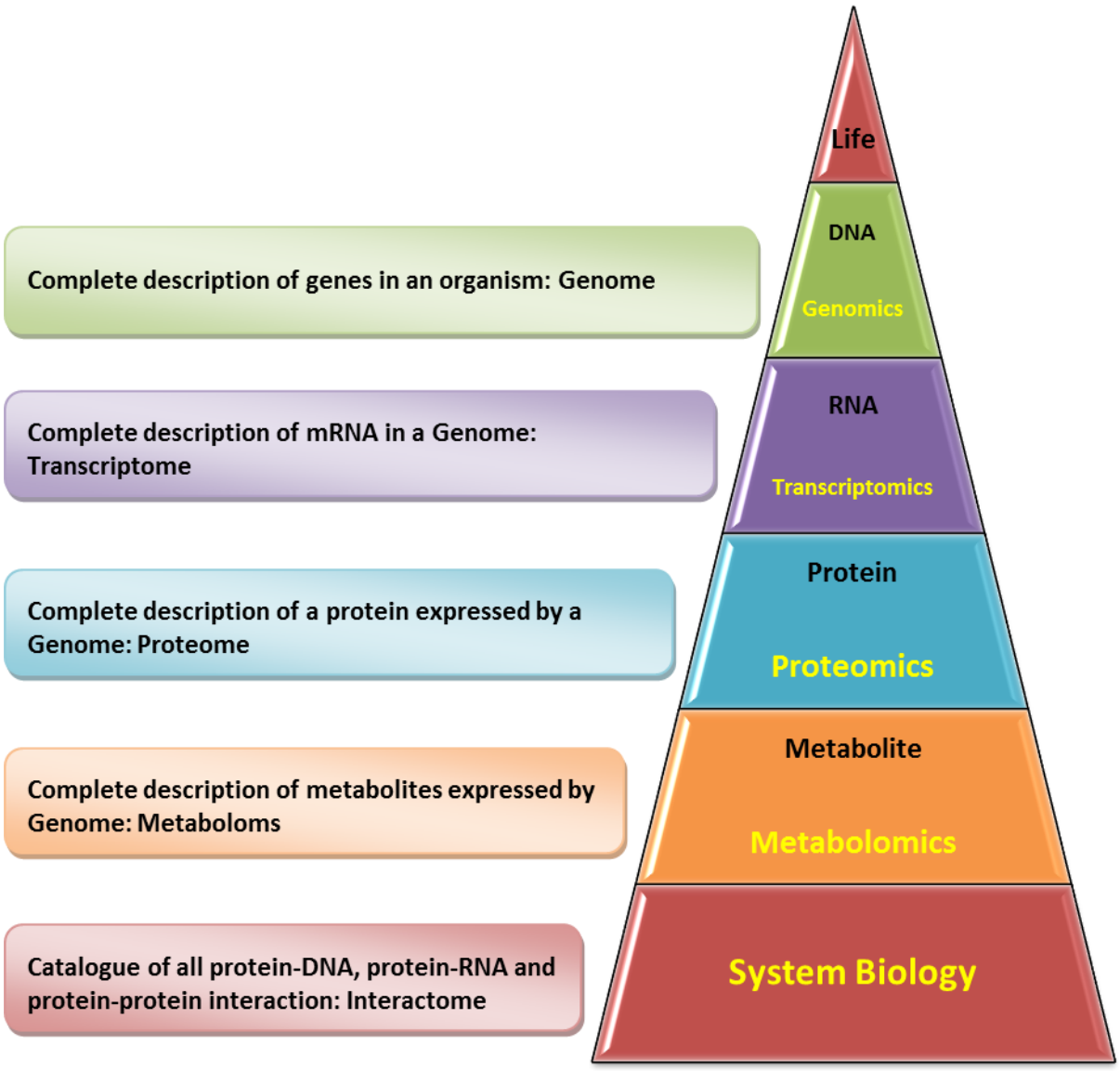

Figure 1.7 Different levels of molecular features of the cell

The proteome is dynamic and every single alteration, whether it is induced genetically or environmentally, results in an alteration of expression levels of proteins. The main aim of proteomics is to study dynamic changes in protein expression in an organism, a tissue or at a cellular level. Consequently three fundamental goals of researches in proteomics are: (*i*) identification of proteins extracted from a complex sample such as plasma and (*ii*) measuring the expression level of the identified proteins in that sample. The other (*iii*) goal for researchers in the proteomics field is characterisation of identified proteins from which structural information, location of protein and post translational modification (PTM) can be obtained. Due to its sensitivity and robustness, proteomics has been chosen for clinical biomarker discovery. Clinical proteomics is defined as: qualitative and quantitative protein profiling study of proteins existing in a clinical tissue sample or body fluids (plasma, urine or CSF) (Banks & Selby, 2003). Comparison of the protein profile obtained from patients with and without disease helps identify differences in the protein expression between the two groups that ultimately relate to the disease phenotype and consequently could be putative disease biomarkers. Studies have shown that early cancer detection can have a profound effect on patient outcome (Skates & Iliopoulos, 2004; Baker *et al.*, 2002). The diagnostic specificity can be enhanced when a panel of biomarkers is used. Obviously, for accurate detection and analysis of biomarkers qualitatively and quantitatively, a high throughput and reliable technology is needed. Recent advances in proteomics technologies such as state of the art mass spectrometry (MS) led to significant improvements in biomarker discovery.

1.3.2 Bottom-up proteomics

There are three main stages in the proteomic studies: sample preparation, sample analysis, and data analysis, all of which impact on the identification and quantification of proteins. One of the most common protocols for bottom-up proteomics (identifying a protein from its digested peptides is called bottom-up proteomics (Figure 1.8)) is the extraction of proteins from a sample so as to deplete high abundant proteins. Extraction of proteins is followed by enzymatic digestion (i.e. trypsin) of the remaining proteins to create tryptic peptides. Finally the tryptic peptides are separated on HPLC and the m/z (called mass to charge ratio which is a unit used in mass spectrometry equal to molecular mass of an ion divided by total number of charges it carries) for precursors and dissociated products determined on MS. The final stage (data analysis), is the searching databases for the matched sequences of the peptides, so identifying the proteins.

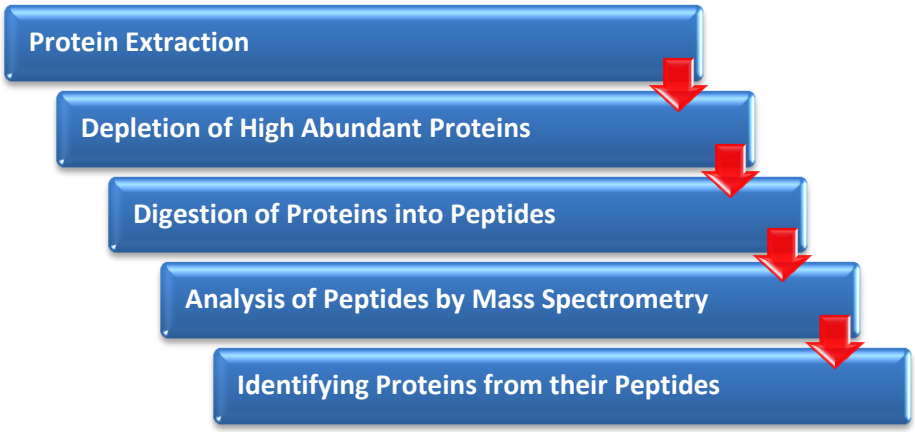

Figure 1.8 Bottom-up proteomics approach

The quantification of proteins in every sample allows differences to be established. Differentially expressed proteins can be committed to verification as candidate

biomarkers. Verification is the next step after discovery and confirms the initial findings in a larger sample cohort.

1.3.3 Proteomics workflow

1.3.3.1 Sample preparation

Prior to introducing samples into MS, the sample should be free of compounds that are known to interfere with the mass spectrometric process. Many different methods have been introduced for protein extraction before MS analysis such as 1D and 2D-PAGE, IEF, HPLC and affinity chromatography (Brewis & Brennan, 2010). 2D-PAGE (2 Dimension Poly Acryl-amide Gel Electrophoresis) is the most commonly used method in proteomics research. O'Farrell and Klose were the first researchers used this method for separation of proteins (Klose, 1975; O'Farrell, 1975). It is a combination of two different methods of separation. Firstly proteins are separated based on their iso-electric point and secondly the separated proteins are subject to electrophoresis on a poly acrylamide gel for separation based on their molecular weight. Protein spots are then cut out from the gel and analysed by MS after tryptically digesting of proteins to peptides. This has been a widely used method for many years, thus the protocols are widely established and consequently is easy to carry out to experienced researchers. 2D-PAGE is also a relatively low cost assay. However, they are not reproducible especially when comparing two samples (although chemical labelling methods have improved this aspect). Another disadvantage of 2D-PAGE is the incompatibility of some of the hydrophobic proteins with the first dimension of separation in this method (Issaq & Veenstra, 2008).

1.3.3.1.1 Depletion of high abundant proteins

The first step in the preparation of plasma samples for proteomics studies is depletion of high abundant proteins. Plasma is an extremely complex fluid. However, it is particularly useful for biomarker discovery as it acts as an archive of protein alteration for tissues. Any changes in the expression of proteins due to the abnormal activities in the tissues results in the secretion of putative biomarkers into the plasma and are consequently clinically invaluable for providing information about pathological processes. A major difficulty in using plasma is its large dynamic range of protein concentration (Hortin & Sviridov, 2010). At the Human Proteome Organization (HUPO) international meeting in 2002, Anderson and Anderson suggested that the dynamic range of protein in plasma spans over 12 orders of magnitude (Anderson & Anderson, 2002). Schiess and his co-workers in a more recent paper published Andersons' findings in more detail (Figure 1.9) (Schiess *et al.*, 2009).

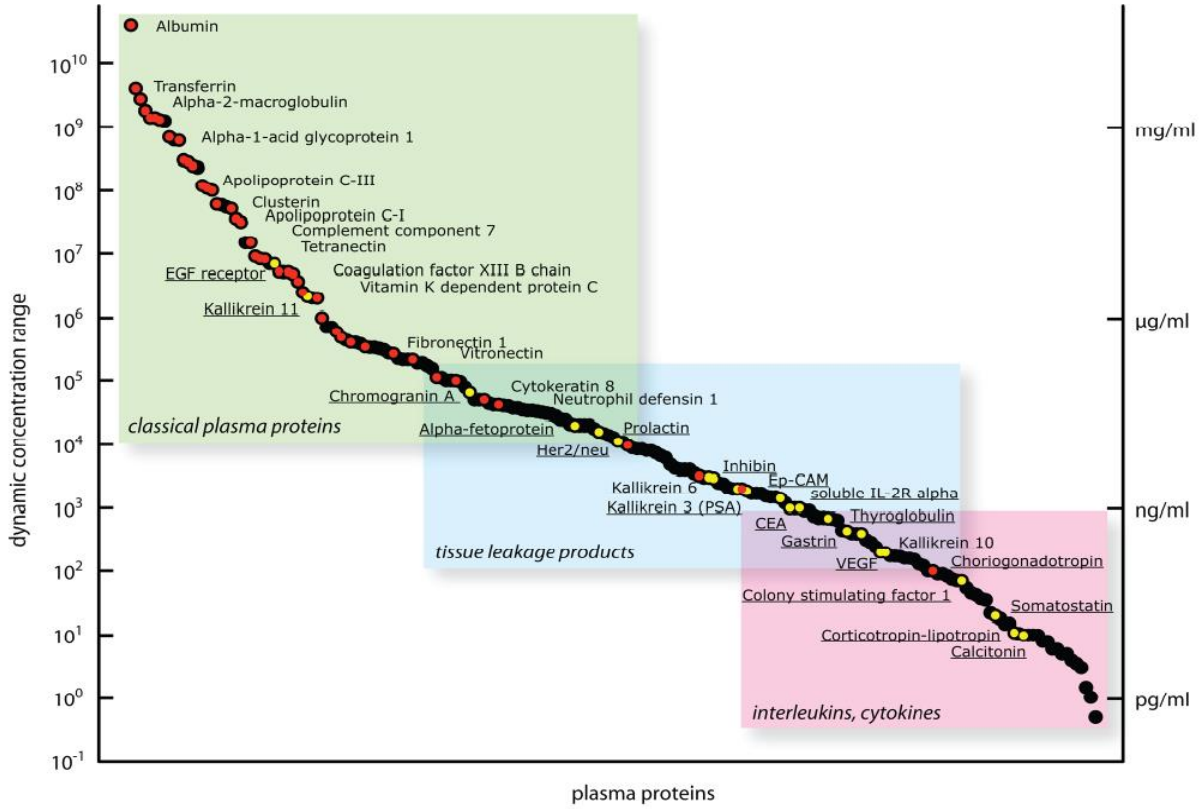

Figure 1.9 Plasma protein dynamic range adopted from Schiess *et al.*, 2009. Red dots indicate proteins that were identified by the HUPO plasma proteome initiative and yellow dots represent currently utilized biomarkers.

Albumin as the most abundant protein constitutes more than half of total plasma protein with concentration of between 40 to 60mg/mL. However, disease-related biomarkers are believed to be presented in plasma at low copy numbers and very low concentration such as PSA at sub-ng/mL. Overall 99% of plasma protein content is derived from only 22 proteins (Figure 1.10).

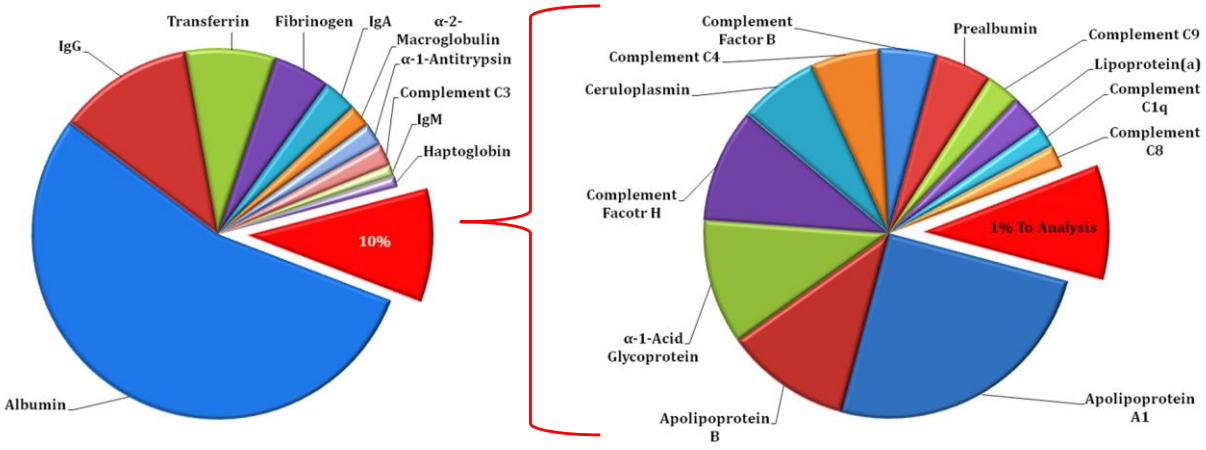

Figure 1.10 Plasma protein content. 99% of total plasma protein consists of only 22 proteins.

By depleting these common proteins, the dynamic range of proteins can be reduced by two orders of magnitude and thus increasing the ability to identify the low abundance proteins. Currently the in-spectrum dynamic range of a typical mass spectrometric platform is not compatible with the dynamic range of protein concentration found in plasma (even after depletion) and as a consequence limits the ability of mass spectrometer to detect low abundance proteins (Tu *et al.*, 2010). Although adding a separation step (such as HPLC) could potentially enhance the detection by 1-2 orders of magnitude, it is still far from what is required for total plasma proteome coverage.

Depletion strategies

Several groups of researchers have tried different methods to deplete high abundant proteins from plasma samples. Currently, many kits are commercially available for the depletion of high abundant proteins. Test-kit protocols for the depletion of six (Bjorhall *et al.*, 2005; Echan *et al.*, 2005), twelve (Hu *et al.*, 2006) and fourteen (Tu *et al.*, 2010) high abundance proteins using immunoaffinity columns have been described. Although depletion of high abundance proteins is successful, some proteins that remain post

depletion could still mask the low abundance proteins. Sigma-Aldrich's SEPPRO IgY-14 (Gong *et al.*, 2006) and ProteoPrep 20 (Millioni *et al.*, 2011) are able to deplete 14 and 20 of the high abundant proteins of plasma respectively. These two methods were assessed for the sample preparation part of this project.

1. ProteoPrep 20 immunoaffinity spin column

Depletion of 20 of high abundant proteins using ProteoPrep 20 plasma immunodepletion kit (Sigma Aldrich, Poole, UK) offers the greatest depletion potential of all immunoaffinity kits. Depletion of one, six, twelve and twenty of high abundant proteins prior to analyse by mass spectrometry has been compared by Roche *et al.* (2009). They suggested ProteoPrep 20 as the best method of depletion in terms of depletion efficiency (Roche *et al.*, 2009). It depletes up to 99% of the top 20 of high abundant proteins in plasma (Table 1.2) in a two-step procedure using spin columns with a concentration step in between.

Albumin	α-2-Macroglobulin	Apolipoprotein A1	Complement C4
IgGs	IgMs	Apolipoprotein A2	Complement C1-q
Transferrin	α-1-Antitrypsin	Apolipoprotein B	IgDs
Fibrinogene	Complement C3	Acid-1-Glycoprotein	Prealbumin
IgAs	Haptoglobin	Ceruloplasmin	Plasminogen

Table 1.2 The top 20 high abundant proteins depleted by ProteoPrep 20

However, whilst the column has high selectivity for certain abundant proteins, there are protein-protein interactions (and possibly non-specific antibody interactions) that result in the non-desired retention of proteins (Koutroukides *et al.*, 2011; Yadav *et al.*, 2011). Moreover, the depletome has been recently characterised and shown to contain proteins previously imperceptible in the analysis of depleted plasma samples (Koutroukides *et al.*, 2011). The procedure of depletion is also time consuming (Figure 1.11) and expensive but the main advantage of the method is the efficient

reduction of abundant proteins which consequently enables detection of more low abundance proteins post depletion.

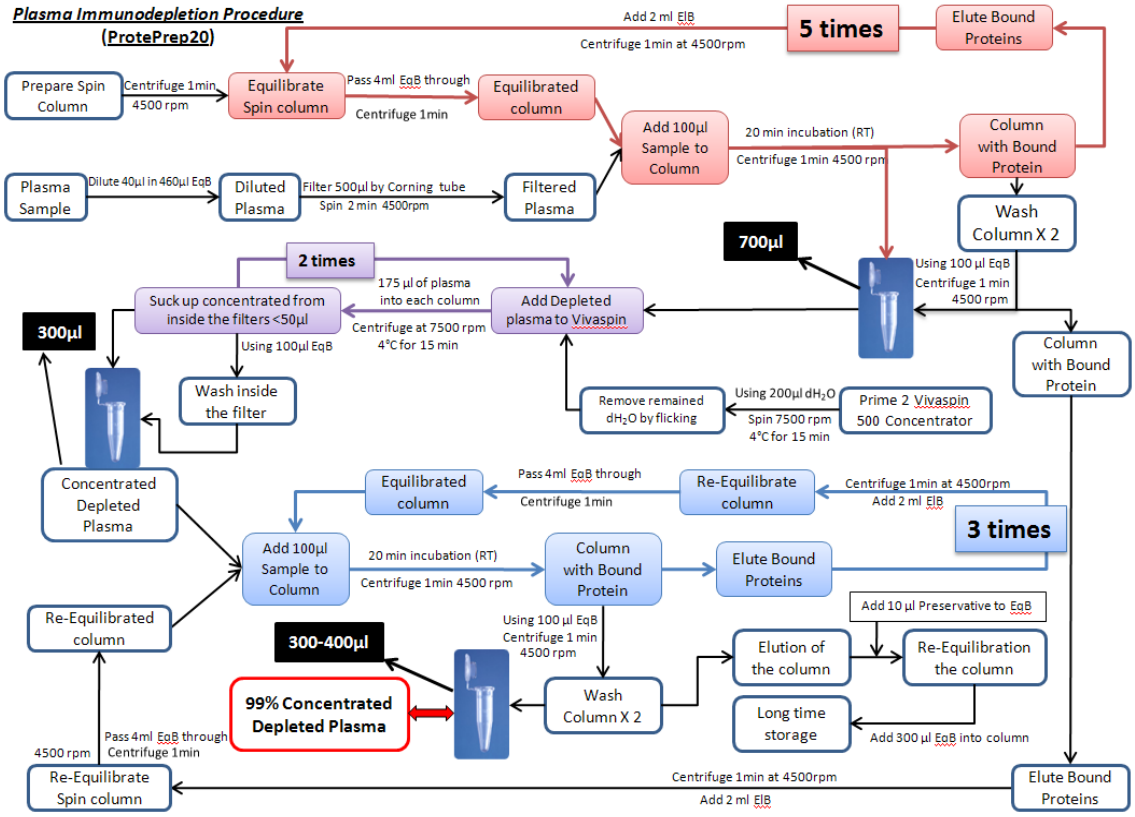

Figure 1.11 ProteoPrep 20 immunoaffinity spin column depletion procedure

An extensive comparison of the above mentioned depletion method and enrichment of the proteins using Proteominer kit has been undertaken and accepted for publication recently by our group. The article shows that both methods have advantages and disadvantages but they are both highly reproducible and complementary.

2. SEPPRO IgY-14 immunoaffinity LC column

SEPPRO IgY-14 LC-2 (Sigma Aldrich, Poole, UK) benefits from the same chemical strategy as the ProteoPrep 20 kit but the antibodies that bind to the top 14 high

abundant plasma proteins (Table 1.3) are packed in a liquid chromatography (LC) column.

Albumin	IgG	Complement C3	α2-Macroglobulin	Apolipoproteins A-I
Transferrin	IgA	Haptoglobin	Apolipoprotein B	Apolipoproteins A-II
Fibrinogen	IgM	α1-Antitrypsin	α1-Acid Glycoprotein	

Table 1.3 Top 14 of high abundant proteins depleted by SEPPRO IgY-14

The diluted plasma sample passed through the column in a mobile phase; while the high abundant proteins bind to the antibodies in the column, the low abundant proteins pass through the column (Figure 1.12).

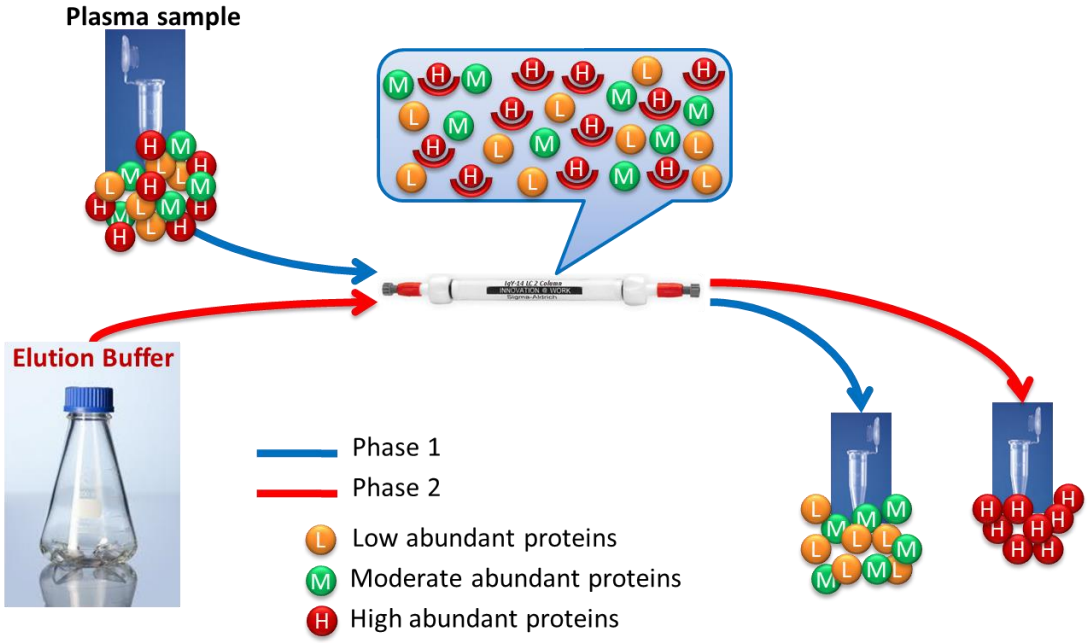

Figure 1.12 Schematic view of the SEPPRO immunodepletion column. Using the first phase (phase 1: blue) low and moderate abundant protein of plasma passes through the column while high abundant proteins bind to the antibodies. Elution buffer elutes them from the column in phase 2 (red).

Since the column is assembled on a HPLC instrument the flow can be passed through a UV lamp; and the absorbance reading recorded at 280nm (Figure 1.13).

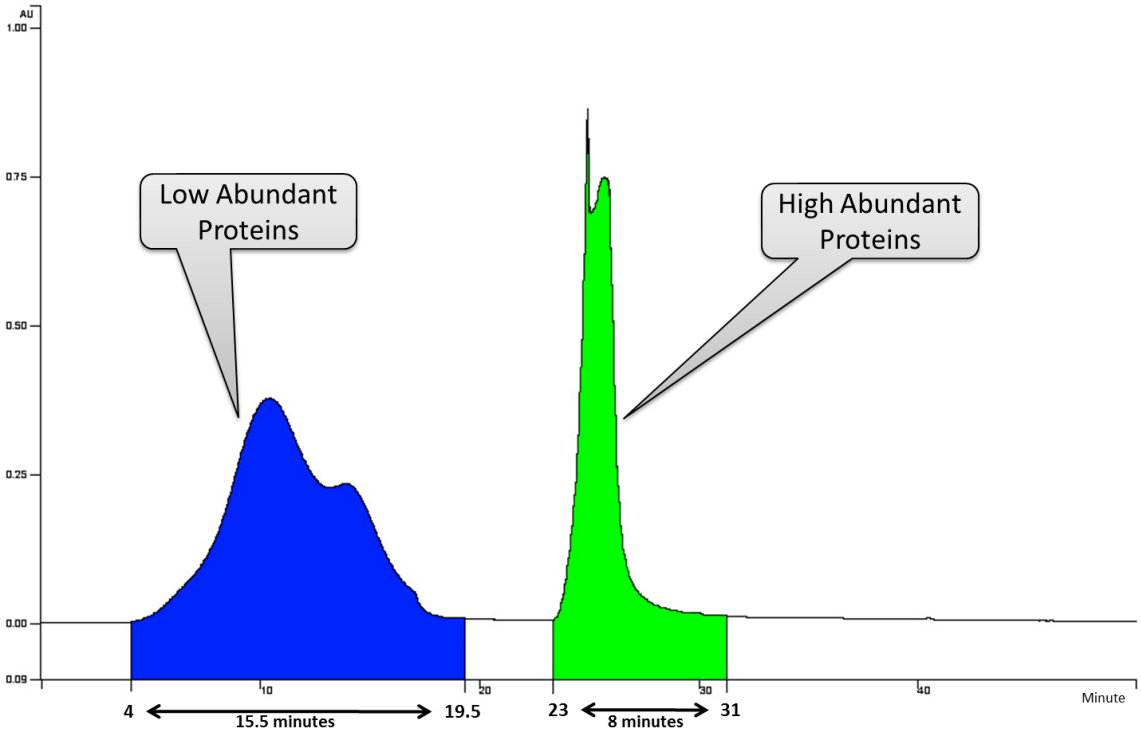

Figure 1.13 SEPPRO IgY-14 chromatogram. Fraction A contains low abundant proteins and B contains high abundant proteins

The first peak contains the low abundant fraction of proteins in the sample. Subsequently, stripping buffer elutes bound proteins from the column followed by neutralisation of the column using neutralisation buffer; dilution buffer then re-conditions the column ready for the next sample.

The SEPPRO IgY-14 LC column was used for the depletion of high abundant proteins for plasma samples in this study. Although the depletion efficiency is less (95% compare to 99%), it depletes the proteins in less than an hour which is a major advantage when considering the total number of samples to be analysed in this study (see Chapter 4, section 4.2.1).

1.3.3.1.2 Total protein concentration measurement

A number of different methods have been introduced to assess the concentration of protein in a sample such as the Lowry assay (LOWRY *et al.*, 1951), the Bicinchoninic Acid (BCA) assay (Smith *et al.*, 1985) and the Bradford assay (Bradford, 1976), with the last two methods being the most common ones amongst researchers in the field. Measuring protein concentration in a sample enables evaluation of the protein depletion procedure (described above). It is also a mandatory step before digesting the proteins to allow the correct ratio of trypsin to protein to be established.

A. Bradford protein assay

The Bradford assay, was named after its developer Marion M. Bradford in 1976. It is based on the binding of the Coomassie G-250 dye to protein that converts the dye colour from brown/red with absorbance of ~400nm to a blue colour with optimum absorbance at 595nm (Bradford, 1976). The more protein in the sample, the darker the blue colour and the higher the 595nm absorbance. Concentration of the protein in the sample can be calculated using the equation derived from the standard curve plot for the assay. The Bradford assay is fast and simple to carry out, with high sensitivity and reproducibility but it is not compatible with detergents used in mass spectrometry analysis (*Labome.*2013).

B. BCA protein assay

The BCA protein assay is an alternative method for assessment of protein concentration in bio-fluids. It is also called the Smith assay due to its introduction by Paul K. Smith for the first time in 1985 (Smith *et al.*, 1985). It is a preferred method as it is compatible with detergents; for example, TWEEN20 and other detergents which are generally used in biochemical and molecular biology experiments. Two major chemical

reactions lead to a change of colour from green to purple. Firstly, peptide bonds (4 or more) in a protein reduce Cu^{2+} ions from cupric sulfate to Cu^{+}. The reduced copper then binds to two Bicinchoninic Acid molecules leading to the formation of a purple coloured product. Similar to the Bradford assay, protein concentration is calculated based on an equation derived from a standard curve plotted for the assay. For the purpose of this project BCA assay was chosen due to its compatibility with detergents used in proteomics.

1.3.3.1.3 Digestion

Bottom-up proteomics requires digestion of proteins to peptides by proteases which make them more amenable to mass spectrometric analysis. There is a wide variety of commercially available protease enzymes such as pepsin, trypsin and chymotrypsin. Trypsin, a member of serine protease family, is the most common protease used in mass spectrometric analysis due to its high specificity and rare miss-cleavages. It recognises target amino acids and in particular cleaves the peptide bonds at C-terminal Arginine (R) and Lysine (K) residues except when they are followed by Proline (P) (Olsen *et al.*, 2004). In the global protocol of tryptic digestion (explained in detail in Chapter 2, section 2.2.5), DL-Dithiothreitol (DTT) hydrolyses the -S-S- bonds of the proteins to –SH at pH 8.0 (maintained by 50mM ammonium bicarbonate). Then alkylation with Iodoacetamide (IAA) stabilises the protein derivatives by preventing reformation of disulfide bonds. A fixed amount of trypsin (1µg trypsin for every 50µg protein in this experiment) is then added to digest proteins into peptides. Optimum digestion time is 16 hours (12-20h) at 37°C and the reaction is terminated by acidification which denatures the trypsin. There are disadvantages with tryptic digestion such as long incubation time and also high pH and high temperature which leads to unwanted

reactions. Long incubation time (16h) makes the assay one of the most time consuming steps of proteomics workflow. Applying microwave energy, called MAPED (Microwave Assisted Protein Enzymatic Digestion), to speed up the digestion of proteins in-solution (Pramanik *et al.*, 2002) or in-gel (Sun *et al.*, 2006) are alternatives methods to overcome this issue. Lopez-Ferrer *et al.* (2008) achieved a significant decrease in the reaction time by increasing the trypsin:protein ratio under high pressure (Lopez-Ferrer *et al.*, 2008). Regardless of the time, high concentration of the enzyme used results in the production of unwanted peptides via autolysis reactions (the trypsin used in proteomics studies is specifically modified to reduce the chance of autolysis). These peptides could lead to intense peaks in the spectrum of mass spectrometry which ultimately have suppressing effects on the other peaks, leading to a subsequent loss of sensitivity. Ultrasound has been also used in a modified digestion protocol which reduced incubation time to 15–30 seconds (Lopez-Ferrer *et al.*, 2005). Increasing the temperature of incubation was also investigated (Lopez-Ferrer *et al.*, 2008) during which digestion reaction was faster, but the high temperature used had a detrimental effect due to the denaturing the enzyme's active sites. Effects of pH 8.0 and 37°C on proteins have been shown by Ren *et al.* (2009) to include: (*i*) Deamidation induction which is the conversion of asparagine residues in a peptide into (1) Aspartic acid plus (2) iso-Aspartic acid, with a corresponding +0.98Da mass shift, and (3) Succinimide with a -17.03Da mass shift. (*ii*) The second effect of high pH and temperature was N-Terminus Glutamine cyclisation that is loss of ammonia from N-terminal Glutamine residue which in turn reduces peptide mass by 17.03Da (Ren *et al.*, 2009). Dick *et al.* (2007) showed 55% cyclisation when incubation time was increased to 24 hours (Dick *et al.*, 2007).

1.3.3.2 Sample analysis

Due to the high complexity of the plasma sample, a fractionation step post-depletion, but before introducing the sample into mass spectrometer, further improves the dynamic range of detection. The most common method of fractionation is separation of the peptides by a High Performance Liquid Chromatography (HPLC). Combination of HPLC and MS allows the compounds in the sample with the same chemical characteristics (hence the same retention time) to separate before entering into the mass spectrometer.

1.3.3.2.1 Nano-UPLC

UPLC (Ultra Performance Liquid Chromatography) systems are the latest version of HPLC systems capable of working under higher pressure (up to 10,000psi). Different flow rates can be used on the UPLC systems from 5µl/min down to 0.1µl/min which in case of the latter the system is called nano-UPLC. The configuration of liquid chromatography systems is broadly similar. The sample being picked up from a vial is carried over within the mobile phase under the pressure created by pumps. The mobile phase (the sample buffer or high aqueous solution) then transfers the sample to a trap column for removing salt in the sample (if any) and chromatographically focusing the sample. The sample then transfers to the analytical column where hydrophobic peptides bind to the stationary phase while hydrophilic ones pass through the column towards mass spectrometer. By increasing the organic composition of the mobile phase (a gradient of 3 to 85%), peptides bound to the column are eluting from the hydrophobic compound in the column (stationary phase). The more hydrophobic a peptide is, the higher the composition of organic solution required for elution and consequently, the later they elute from the column. Depending on the complexity of

the sample and the type of separation column, different gradients are evaluated to find the best gradient pattern for optimum separation.

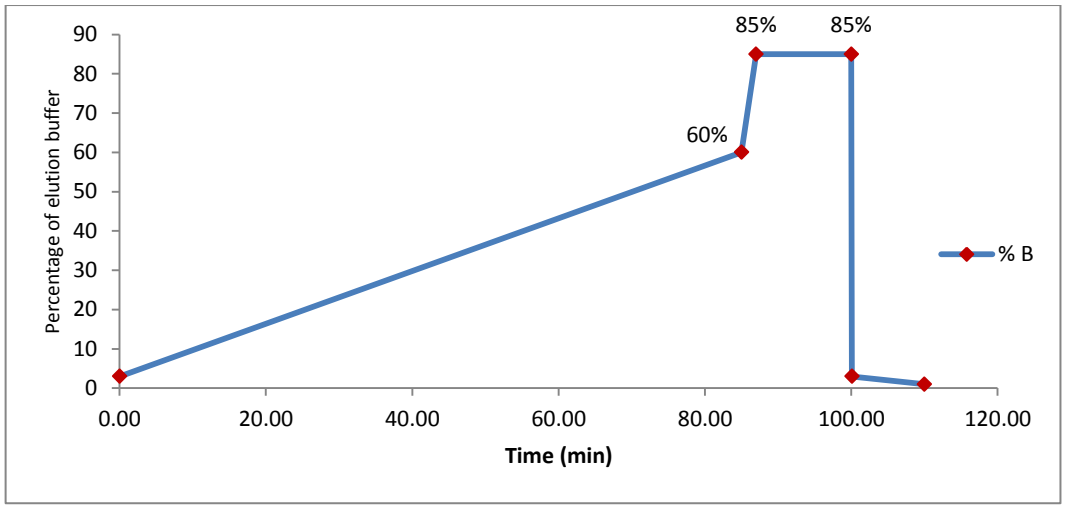

Figure 1.14 A typical gradient on the UPLC system. The graph shows the percentage of the elution buffer (mobile phase B) over 110 minutes of total run time.

A typical gradient starts at very low percentage (1-3%) of mobile phase B (elution buffer) for a maximum of 3 minutes (Figure 1.14). The percentage of mobile phase B increases linearly up to 60% over 80 minutes during which most of the peptides in the sample are eluted. A quick ramp of B from 60% to 85% over 2 minutes followed by 13 minutes isocratic at 85% B elutes all the remaining peptides from the column. Flushing of the column with 97% of mobile phase A over 10 minutes equilibrates and reconditions the column for the next sample.

NanoACQUITY UPLC

The liquid chromatography system used for the purpose of this study was a nanoAcquity-UPLC system from Waters (Waters Ltd, Hertfordshire, UK). It has a binary solvent manager (BSM) where it controls the mobile phase A and B compositions, an auxiliary solvent manager (ASM) to control the LockSpray and a sample manager (SM), a temperature controlled chamber to keep the sample vials at 6˚C while waiting for the

automatic injection. The system benefits from a 2cm trapping column packed with 5µm beads connected to a 25cm analytical column for the separation of peptides.

1.3.3.2.2 Mass spectrometry

Mass spectrometry has become the backbone of proteomics studies. A mass spectrometer is a powerful instrument capable of identification and quantification of known and unknown compounds. A mass spectrometer measures the mass to charge ratio (m/z) of the ionised molecules in the gas phase. For mass spectrometers to be able to analyse and measure the mass of a molecule, the molecule should be converted into gas phase ions. In electrospray, to achieve that, high voltage electricity (3-5kV) is applied to a spray emitting from the nano-spray emitter which contains the analyte molecules. The molecules simultaneously ionise and enter the gas phase as a consequence of the high voltage and nebulisation. Once ionised the ions are transmitted into the high vacuum regions of the mass spectrometer using a series of electrical fields which control the passage of ions to the detector. The detector converts the signal response of the ion into digital information which is displayed as a peak in a mass spectrum. Mass spectrometer separates ions based on their mass to charge (m/z) where mass is the atomic mass unit and the charge is the charge of one electron in absolute value. Based on intensity, the peak with highest intensity is assigned 100% and called base peak and the rest of the peaks are measured relative to that. Mass spectrometers have hugely improved since it first introduced by Thomson in 1897 (Thomson, 1899). Sensitivity and specificity of the machine has reached to femto- and attomole levels with mass accuracy of less than 1ppm. (Mass accuracy is the difference between theoretical mass calculated based on the chemical formula and measured mass obtained by MS, with the unit of part per million (ppm)).

Mass spectrometer

All mass spectrometers are made of three main parts: source, analyser and detector. Depending on the application or the molecule to be analysed, different types of sources and analysers are used. The coupling of two or more of the same or different, types of analyser has been used to yield more sensitivity and selectivity. Analyser and detector of all mass spectrometry machines are under vacuum regardless of the type of the analysers. The vacuum condition throughout the MS helps ions to be transmitted through the instrument unhindered by interactions with air molecules.

A. Source

More than 15 different types of ionisation (source) have been reviewed (Hoffmann & Stroobant, 2006). Based on the sample to be analysed, different types of sources can be used. One of the most frequently used method of ionisation especially in proteomics area is Electrospray Ionisation (ESI) briefly described here.

Electrospray Ionisation (ESI)

ESI, introduced by Dole and his co-workers in 1968, is a soft ionisation technique, in which ions are generated in the gas phase without extensive fragmentation. Yamashita and Fenn were the first researchers who coupled ESI with mass spectrometry in 1984 and received the Nobel Prize for it (Fenn *et al.*, 1989; Yamashita & Fenn, 1984). ESI is a spraying technique under the effect of an electric field that produces positively or negatively charged ions based on the field polarity setting. Ion formation can be assisted by adding formic acid or acetic acid to help protonation (positive charge) or ammonia solvent for de-protonation (negative charge). To ionise molecules within a sample using ESI, it should be dissolved in a solvent and transferred to a potential

gradient through a needle at a low flow rate (~1-5µl/min in normal sprayer and ~200nl/min in nano-spray systems). By applying a potential difference between the needle and counter-electrode, a high electric field (generally between 3 to 5 kV) is obtained. When the sample (within mobile phase) reaches the end of the needle, under the effect of electric field, it makes a cone shape of accumulated charge ions which is called a Taylor cone (Figure 1.15). When the puling force of the counter electrode exceeds the surface tension of the end of the Taylor cone a droplet is separated into the ESI chamber. As droplets shrink through nitrogen gas nebulisation aided desolvation, charge accumulation occurs on the surface of the droplet to a point where charge repulsion meets columbic repulsion and the droplets break down. This procedure continues until the analyte molecule becomes free of solvent and remains with the attached charge (Wilm, 2011).

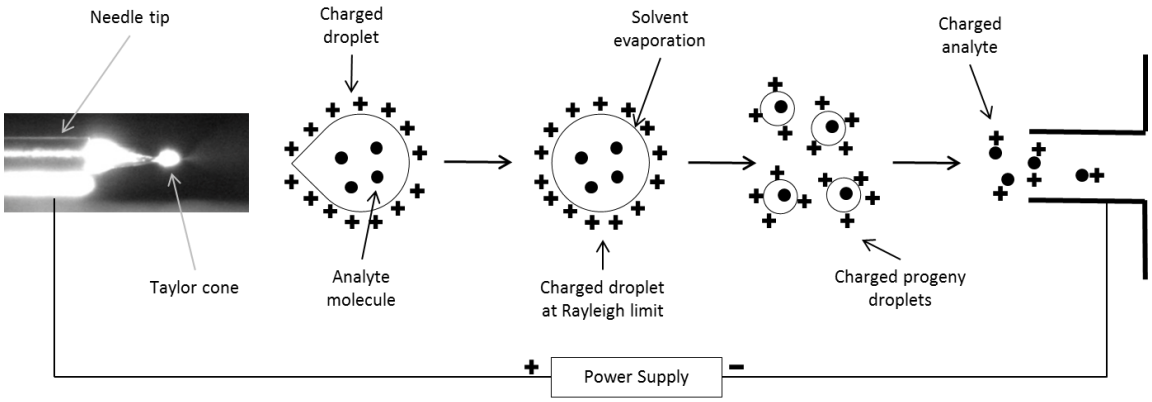

Figure 1.15 Electrospray ionisation process

These ions then pass through the skimmer towards a semi vacuum area in which neutral molecules are pumped out and the charged ions are transferred into the analyser under vacuum pressure. Another consequence of shrinkage is increasing the ion's energy by accumulation of multiple charges, which makes ESI preferred ionisation method for collision induced dissociation (CID) in tandem MS.

B. Analyser

The analyser controls the transmission of ions through the MS to the detector. The principle roles involve movement of ions through the mass spectrometer, separation of ions and maintaining focussing of ions to reduce ion loss as a consequence of transmission. To have more sensitivity and be able to use mass spectrometer for multiple experiments, different analysers are combined in a single machine such as triple quadrupole (using three quadrupole) or Q-ToF with a quadrupole and a time of flight analyser. Quadrupole and time-of-flight, two of the most common analysers used in mass spectrometry systems, are described briefly here.

a) Quadrupole

Quadrupole is made up with four rods parallel to each other with opposite electric fields. Paul and Steinwedel were the first researchers who introduced quadrupole in 1953 (Paul & H. Steinwedel, 1953). A fixed direct current (DC) and radio frequency (RF) between rods separates the masses and only the ions with a specific mass are able to pass through towards the detector. Assembling three quadrupoles continually, also called triple-quadrupole, improves the selectivity and specificity. In triple quadrupoles, the first and the third quadrupole are filtering and transferring ions while the middle one which has RF only is the fragmentation area, also called collision cell. The first part selectively chooses the ions of interest and accelerates the ions through a second quadrupole which contains an inert gas (collision gas typically Argon) which causes dissociation of the ion. This fragmentation is called Collision Induced Dissociation (CID). Dissociated ions then transferred to the third quadrupole where they are guided towards the detector and measured.

b) TOF

Stephen was the first scientists to suggest the idea of time-of-flight in 1946 (Stephens, 1946). Time-of-flight separates ions based on the velocity of the ion in a flight tube. As ions enter the flight tube, they are typically transmitted by a pusher region which accelerates the ions through the flight tube. The flight of ions is under effect of an electric field inside the tube with a detector at the end of it. Ion's mass to charge is calculated based on the time it takes for every ion to pass through flight tube to reach the detector. The time taken to traverse the flight tube is inversely proportional to its m/z. So ions with a larger m/z take longer than ions with a smaller m/z, but the relationship is linear over several orders of magnitude (Cotter, 1994). Time-of-flights analysers are made in linear and reflectron mode. In reflectron fashion, ions travel further more with the aid of lenses and pushers leading to a better separation and consequently, increasing resolution without losing sensitivity.

C. Detector

Molecules convert into ions and selected ions pass through analyser and hit to detector. Detector generates electric current from every single ion proportional to its abundance. Electric current then converts to a signal showing a peak in mass spectrum which contains two types of information about the ion, mass to charge and the intensity of the ion.

Synapt G2 HDMS

Synapt G2 HDMS, (Synapt Generation 2 High Definition Mass Spectrometry) from Waters (Waters Ltd, Hertfordshire, UK) is one of the latest technologies in the field of mass spectrometry. Peptides that are eluted off the analytical column on the nanoACQUITY UPLC are introduced into the Synapt G2 HDMS through the nano-ESI source (Figure 1.16) where they are converted into the gas phase. An ion guide maintains the flow of the ions to a quadrupole where a single or a series of ions can be selected for subsequent transmission. Triwave is the next section that adds another level of separation to the analysis by separating the ions based on their shape, size and mobility. The Triwave consists of three ion guides including the trap (creates focused packets of ions), the ion mobility separation (separates ion based on their mobility) and the transfer ion guide (maintains transfer of the mobility separated ions to the QuanToF). The principles of the ion mobility are discussed in more detail below. QuanTof is a detection system that allows ions to travel further with the aid of a dual stage reflectron and a mirror which results in higher resolution separation.

When the mass spectrometer is set on the MS^E mode, Triwave and QuanTof are functioning as analyser separating ions before reaching the detector. When the system is set on the $HDMS^E$ mode the middle ion guide in the Triwave is functioning as an ion mobility separation part (Giles *et al.*, 2011).

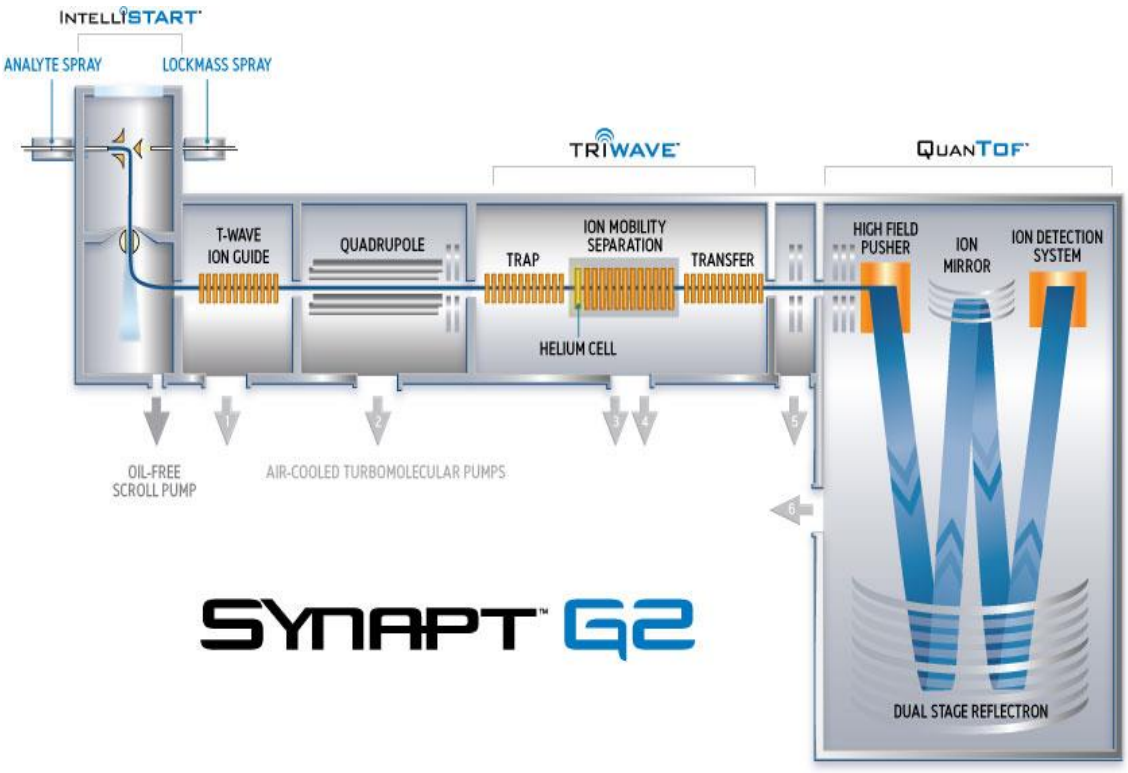

Figure 1.16 Schematic view of a Synapt G2 HDMS

The Synapt G2 HDMS has a data independent acquisition mode called MS^E. Every 2 seconds in the MS^E method, a full MS scan of the ions at low collision energy (4-6eV) is carried out followed by a full MS scan at higher collision energies (~27-50eV). This results in a spectrum being recorded of the intact precursor ions in the first spectrum followed by a spectrum containing the dissociated product ions.

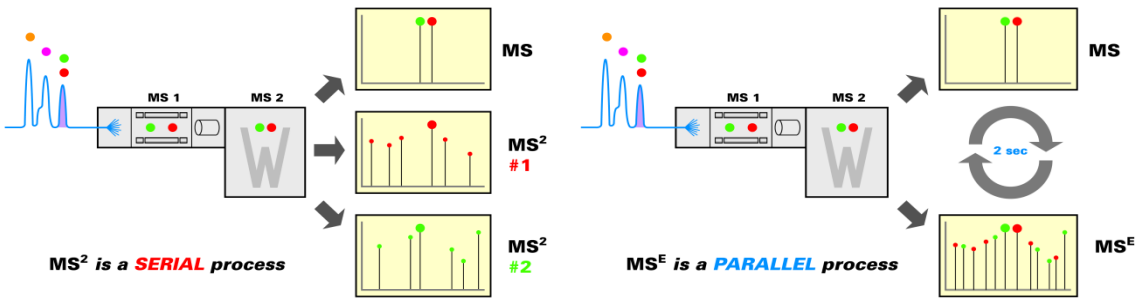

Figure 1.17 Comparison of MS/MS and MS^E

MS^E, which is a data independent acquisition, is able to perform MS and MS/MS in parallel (Figure 1.17) through which the MS mode provides high selectivity and accuracy for quantitative profiling and high mass resolution for identification in the MS/MS mode. An alternative strategy is called Data Dependant Acquisition (DDA).In comparison to MS^E, DDA is consuming serial process where not all the ions are sampled which is one of the key disadvantages of the method. This is due to the way it functions as only the most intense peptides are selected and MS/MS is applied to them while the peptide of interest could be less intense or it can be eluted off the column at the time of MS/MS. Another issue with the MS/MS is its inability to perform quantitative analysis on proteins without the use of chemical labels. Using MS^E/$HDMS^E$ methods utilise label free quantification by the addition of a known concentration of an internal standard (such as alcohol dehydrogenase (ADH)).

Ion Mobility MS

Ion mobility mass spectrometry adds another level of separation to the sample analysis. It improves separation of the ions with similar m/z and retention time but different in size and shape, based on their drift time. A series of transient voltages applied to sequential electrodes establishes a travelling wave field along which ions can travel. A flow of nitrogen gas from the opposite direction (towards source) helps in the separation of the ions (Giles *et al.*, 2011). Small molecules will travel slower than larger molecules which are unable to overcome the potential apex of the wave and so the time to reach the end of the mobility separator is longer. The time taken for ions to travel through a mobility separator is known as the drift time (Figure 1.18).

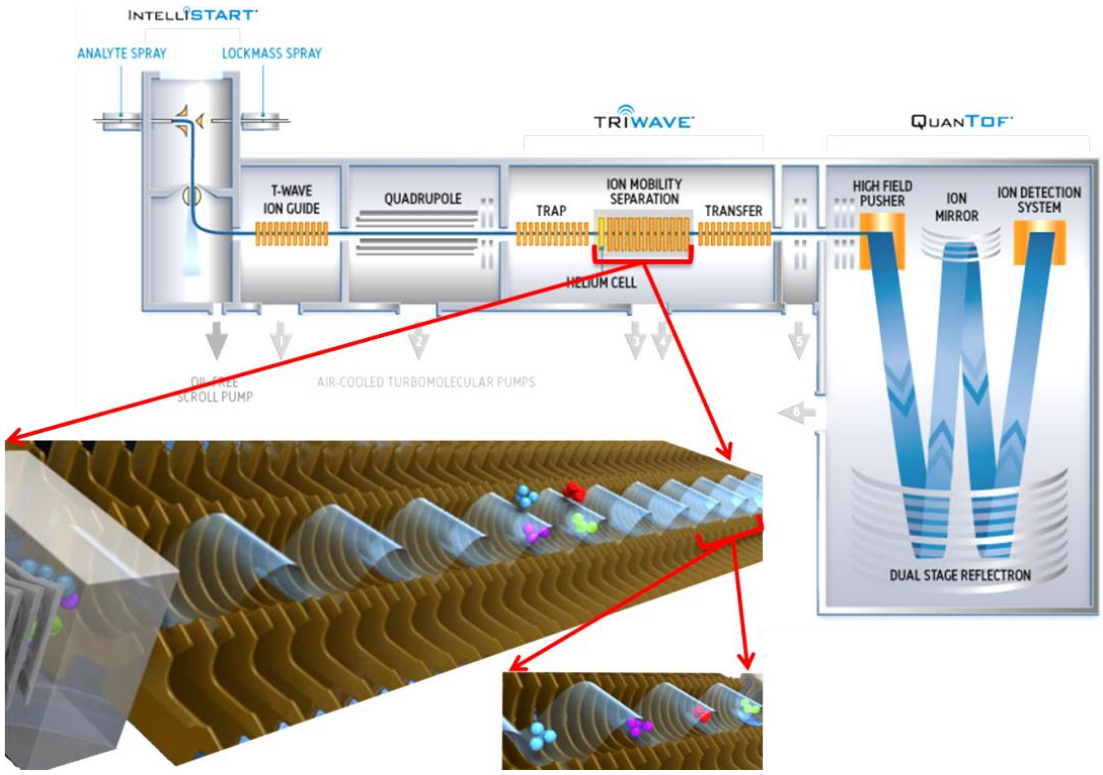

Figure 1.18 Ion mobility separation in TRIWAVE

Triwave benefits from three ion guides that allow ion packeting, ion mobility separation and transfer of the mobility separated ions to the ToF region. The first ion guide (called trap) accumulates the ions while the previously trapped ions are being mobility separated. This accumulation is performed with the aid of a potential barrier applied at the gate (placed at the end of the trap ion guide). When the previous set of ions are separated, the gate opens (by removing the potential barrier) for the accumulated ions to transfer (in packets) into the second ion guide called ion mobility separation (IMS) ion guide. In the IMS ion guide, ions are separated based on their mobility with the aid of travelling waves of the electric field and nitrogen gas from the opposite direction. To maintain the travelling wave of electric field, opposite phases of radiofrequency (RF) voltage are applied to adjacent rings (to provide radial ion confinement) followed by direct current (DC) voltage which is superimposed on the

rings in a repeated pattern helping the ions to transfer through the nitrogen gas. Separation of the ions depends on their ability to keep up with the travelling waves of electric field with low mobility species being overtaken by the waves more frequently than high mobility species. The helium cell at the entrance of the IMS minimises scatter of ions when they transfer from a low pressure to a high pressure area (the pressure in the IMS ion ring is about 3mbar compare to 0.5mbar throughout the system (Giles *et al.*, 2011)). Once ions have passed through the mobility separator they are fragmented in the transfer ion guide (the third stacked ring ion guide in the Triwave). Fragmentation of the mobility separated ions is the main advantage of this setting. Generally, product ions are assigned to the precursor ions by aligning their retention time whereas in this setting the precursor and the product ions will also have the same ion mobility and retention time. Thus, assignment of the fragmented ions to the precursor ions will be based on their retention time and drift time which leads to a much greater confidence in the protein identification (Valentine *et al.*, 2011).

MS^E and $HDMS^E$

As described earlier in this chapter, in DDA (data dependant acquisition) under-sampling of low abundant peptides (could be putative biomarkers) due to the speed of performing MS/MS on the preselected ions is one of the main disadvantages of the DDA method (Liebler, 2004).

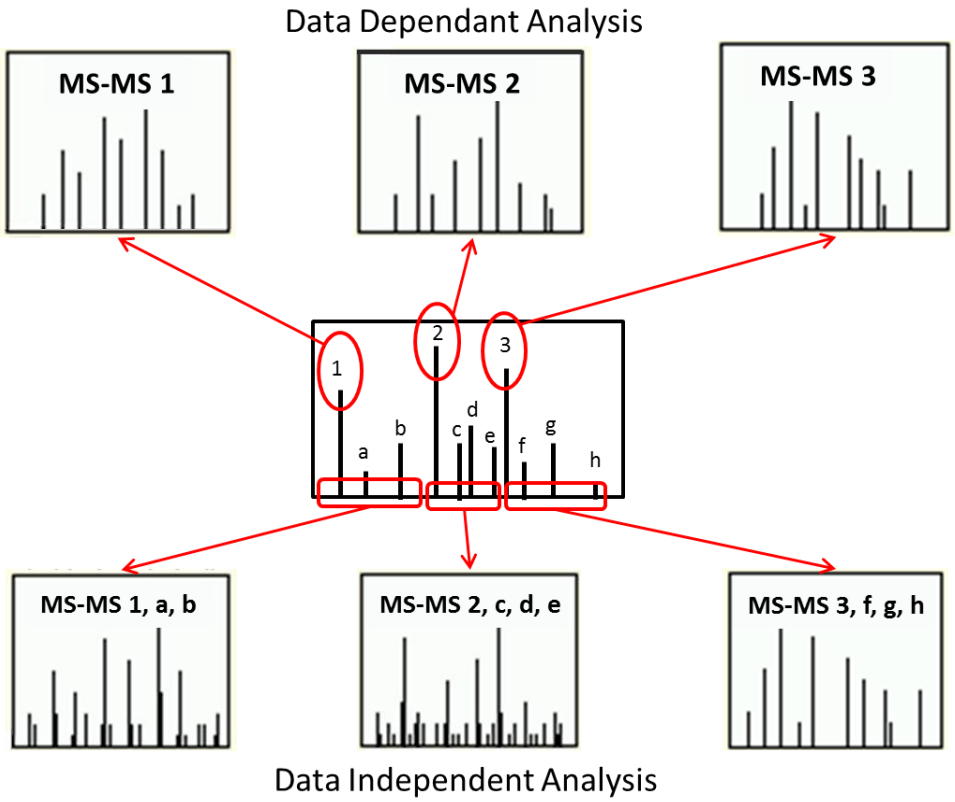

Figure 1.19 Comparison of DDA vs. DIA (Liebler, 2004)

Figure 1.19 compares the two methods of DDA and DIA. Data dependant analysis chooses 3 of the most abundant peptides and fragments them whereas in the data independent analysis (MS^E) sequential windows of m/z are chosen to be fragmented. This different method of analysis results in less chance of missing the low abundant peptides and consequently fragmentation of them. As mentioned above, the Triwave, enables separation of the ions based on their shape and size (mobility). When the instrument is set to analyse the ions based on their mobility, the method of acquiring data is called High Definition Mass Spectrometry or $HDMS^E$. In the $HDMS^E$ method chromatographically separated peptides are ionised by electrospray at the source and transferred to the ion mobility section of the Triwave. Fragmentation takes place post mobility separation and the fragmented ions hit the detector after travelling through the ToF. So all of the information regarding to the parent and product ions of every

detectable component is collected and the masses are measured. Aligning of the retention time and drift time results in the complete cataloging of the peptides in the sample (Giles *et al.*, 2011).

1.3.3.2.3 Label-free quantitation method

To be able to perform expression analysis in a biomarker discovery study, identified proteins need to be quantified. Many quantitative analyses in the past have relied heavily upon isotopic labelling of peptides for MS/MS analysis. Different methods of labeling such as Stable Isotope Labeling by Amino acids in Cell culture (SILAC) (Ong *et al.*, 2002), Isotope-Coded Affinity Tag (iCAT) (Gygi *et al.*, 2002) and isobaric Tags for Relative and Absolute Quantification (iTRAQ) (Ross *et al.*, 2004) are among the most common methods used for quantitation studies. The need to have highly concentrated samples for isotopically labelling quantification and also the addition of another complex step into the sample preparation protocol (addition of variability) are the two main disadvantages of these methods of quantitation. To overcome those issues, a label-free quantitation method was introduced by Silva *et al*, (2006), also called Hi-3 Quantitation (Silva *et al.*, 2006). In the Hi-3 quantitation, the software chooses three of the most intense signal responses of each peptide and calculates the average of them. The same calculation applies to the internal standard spiked in. Since concentration of internal standard in known absolute quantity is calculated based on the following equation:

$$\textit{Absolute quantity} = \frac{\textit{Average of MS signal responce of three most intense peptide of each protein}}{\textit{Average of MS signal responce of three most intense peptide of internal standard}}$$

Equation 1.1 Calculation of absolute quantity for each protein

1.3.3.3 Data analysis

1.3.3.3.1 Protein identification and quantification

Proteinlynx Global Server (PLGS) is a data management informatics platform that can be used to process mass spectral data obtained on the Synapt G2 HDMS. It has two main sections called Identity^{E} and Expression^{E}. Identification and quantification of the proteins are done by Identity^{E} and alteration in the expression of the proteins in different groups is checked by Expression^{E}. Each one of the two sections is split into sub-sections. The user's role is mainly limited to choosing the appropriate database, establishing the correct thresholds for processing and selecting a suitable false discovery rate (FDR). In the Identity^{E} section of PLGS, it uses a novel algorithm to process raw data obtained by mass spectrometer using different properties of ions i.e. retention time, precursor/product ions intensity and accurate mass. After the first analysis, PLGS generates a list of all precursor and product ions. De-convolution of these ions in the list is based on another algorithm called Apex3D which creates an exact mass and retention time (EMRT) table in low and elevated energy. This table contains precursor and product ion masses for each peptide which is ready to be searched against the non-redundant protein databank (Uniprot). The database was appended with alcohol dehydrogenase for label-free quantitation. PLGS also creates a decoy database in which the protein amino acid sequences are reversed or randomised and concatenated to the original database. This is used for calculation of FDR using the following equation:

$$FDR = \frac{FP}{(FP + TP)}$$

Equation 1.2 Calculation of FDR. In the above equation, FP is the number of proteins identified from decoy database and TP is the number of proteins identified from the original database. FDR is expressed as percentage.

In the above equation, FP is the number of proteins identified from decoy database and TP is the number of proteins identified from the original database. The data passes through the search algorithm (Figure 1.20) three times after which it creates the final list of proteins and peptides (Li *et al.*, 2009).

PLGS performs many different processes on the data including filtering, sorting, smoothing and centering of the data, subtracting the background noise and removing the isotops under two main sections: Data processing and databank searching (Levin *et al.*, 2007; Li *et al.*, 2009).

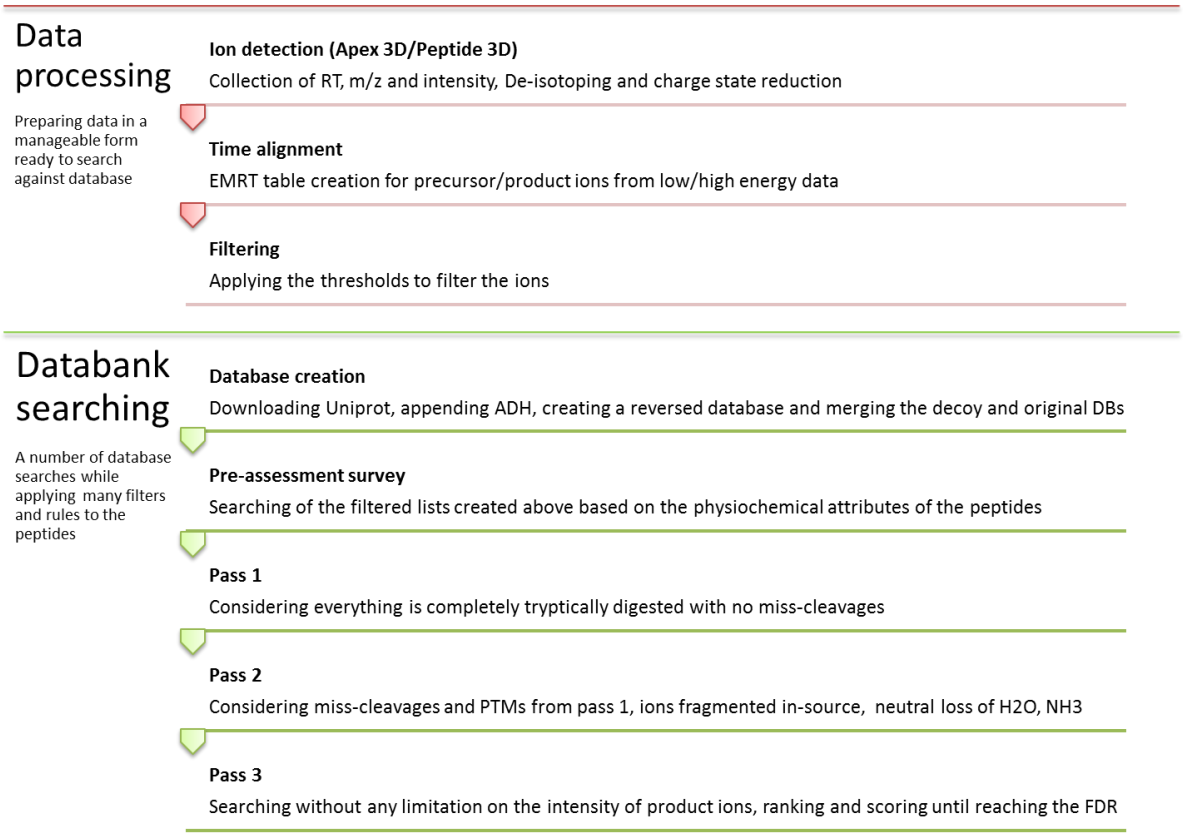

Figure 1.20 Identity^{E} overview in PLGS

When the protein IDs are established PLGS carries out label-free quantification based on the Hi-3 quantitation method (Silva *et al.*, 2006) followed by the creation of a list of identified proteins with a number of relevant parameters which pertain to each of the proteins including the absolute amount. Quantitation of the total amount of protein in the sample (performed in the scouting run) enables adjusting the injection volume of the sample to load optimum amount of protein on the column for maximum sensitivity. Loading less than the optimum amount of protein results in the identification of fewer proteins, whereas loading of higher amount protein results in saturation of the column and mass spectrometer detector. Column saturation leads to peak broadening, and hence a decrease in the retention time and less number of detected proteins (Dorschel, 2008).

1.3.3.3.2 Expression analysis

To assess the expression levels of the proteins of samples in different groups a plan consisting of four data handling packages was prepared (Figure 1.21). Each of the packages uses different statistical means by which to identify proteins and/or compare expression levels alteration. Comparisons of these approaches and identifying common candidate proteins within different groups of samples allow greater confidence in this initial phase.

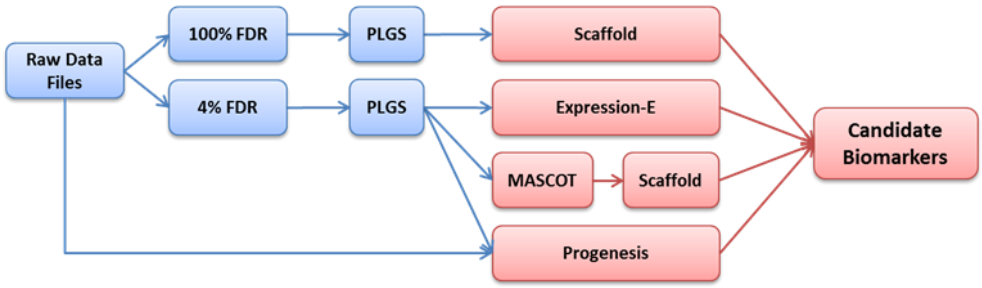

Figure 1.21 Data analysis plan. The pattern followed for identification, quantification and expression analysis

The output for each one of the packages is an excel sheet containing the proteins with/without alteration in the expression profile. Comparison of those sheets and finding proteins revealed in all four packages results in a list of proteins (candidate biomarkers) that will be verified with other methods.

1. **Expression^{E}**

For every protein, Expression^{E} collects all the matching peptides from sample A and sample B and compares the intensities of the peptides to give relative quantification of the protein. Using the intensities of peptides belonging to a known amount of ADH spiked in the samples, the program then uses auto normalisation method to normalise the datasets. Protein lists are then filtered to show only the proteins that present in all

three replicate injections of every sample and then PLGS creates the output table (Figure 1.22).

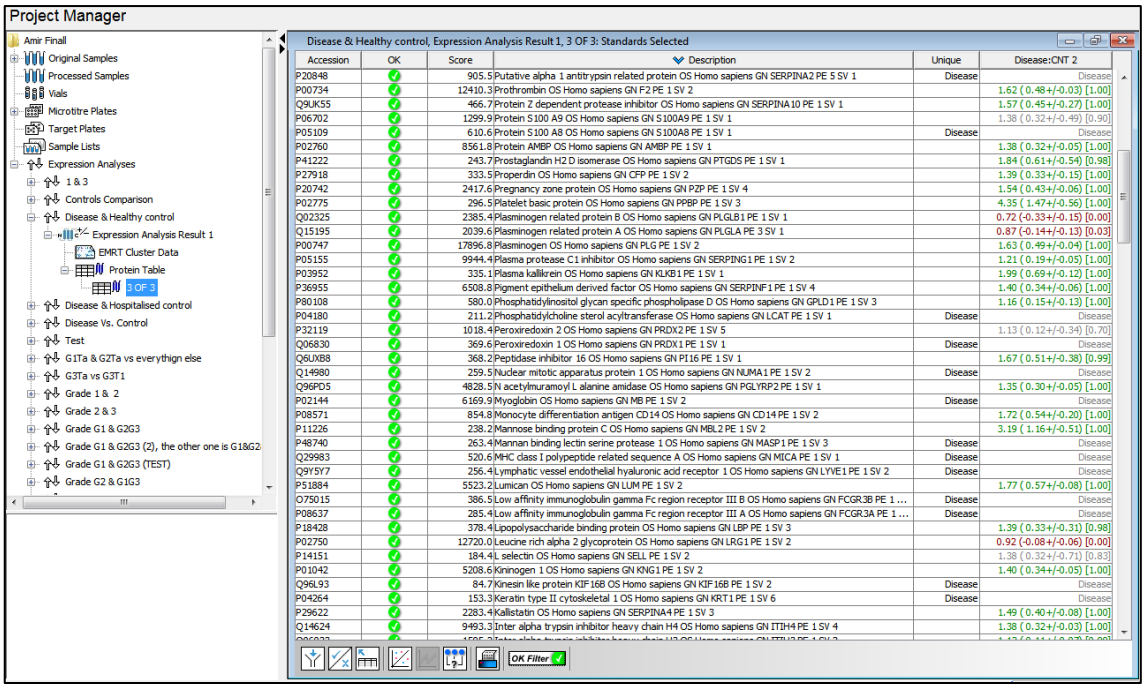

Figure 1.22 An output table of the Expression[E]. The far right column shows the alteration in the expression profiles of the proteins. Greens are up-regulated (p-Value ≥ 0.95), reds are down-regulated (p-Value ≤ 0.05) and the grey ones did not have significant changes in their expression levels.

The output table has the protein names, accession numbers, a column to show if the protein is unique in one group and the ratio of the proteins between the selected groups. The probability of the ratio shows the significance of the protein up-regulation (shown in green) and down-regulation (shown in red). The proteins that do not have significant change between the groups are shown in grey.

2. Progenesis LC-MS

Progenesis LC-MS is a data analysing package from Nonlinear Dynamics Ltd. It benefits from a peak modelling algorithm (*Nonlinear Dynamics.*2013) in which peak identification is followed by creation of a model that contains position and quantification

information. To be able to compare between sample groups Progenesis LC-MS aligns and combines data to compensate for the variations between the runs for greater reliability and better reproducibility. It then creates a dataset containing information about all peaks in all samples that contains no missing values (unique to Progenesis). A number of stages for alignment, filtering and statistical analyses increase the confidence and sensitivity of the procedure (Figure 1.23).

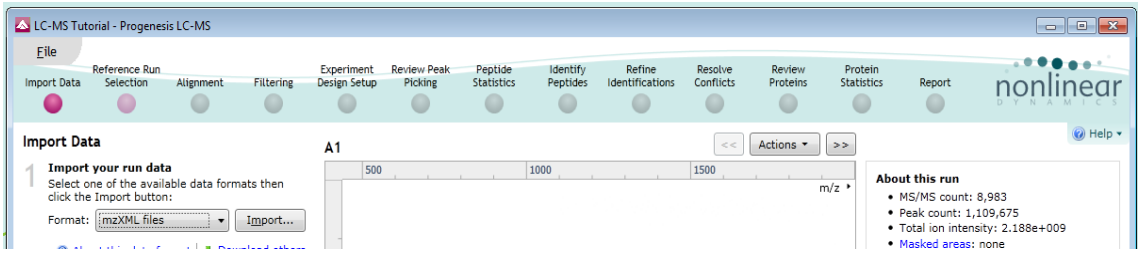

Figure 1.23 A screen shot from the front panel of the Progenesis LC-MS software. The whole procedure consists of 13 steps from uploading of the raw data files to creation of a report.

When the raw data files are imported, the software aligns and filters the data based on the reference run chose by the user (one of the runs with the best chromatogram in case of peak separation and expected start and end retention times). The next stage is the experiment design where raw data files from different sample group (e.g. patients and controls) are grouped. The protein identities identified by PLGS exported before are imported into Progenesis LC-MS. At this stage, data can be subjected to several statistical tests including PCA and/or ANOVA. Results can then be filtered based on different parameters such as fold change and p-value and the final protein list with all of the significant alterations is created.

3. Scaffold

Scaffold is another data analysis package provided by Proteome Software Inc. It uses X! Tandem (a search engine for protein identification from MS data (Craig & Beavis, 2004)

that comes bundled with Scaffold) to increase the confidence levels of protein identification (unique to Scaffold). Scaffold searches the raw data files that are already searched using another search engine (i.e. PLGS or MASCOT) with X! Tandem and checks the identification agreement between the two search engines results in higher confidence in the resultant identification (Searle *et al.*, 2008). Similar to Progenesis LC-MS, scaffold is capable of importing raw data files from different instrument using specific plug-ins. Data processing with Scaffold helps in validation, interpretation and organising of the data. Scaffold also enables graphical comparison of the protein expressions between different sample groups and across the experiment. Statistical algorithms (ANOVA/t-test) used by the software during the validation increase the existing probability of that protein in the samples. The first step is to export the data from PLGS in a zipped file created by the relevant plug-in into Scaffold. The only adjustment that needs to be done in PLGS before processing of the samples is to change the percentage of FDR in the workflow from 4% to 100%. This leads to reporting of everything in the sample without any statistical analysis or filtering. The output list generated after processing data with 100% FDR normally has between 800 to 900 proteins for plasma. Scaffold will apply filters and performs statistical analysis on the list later. The same database used in PLGS is used in Scaffold (Figure 1.24).

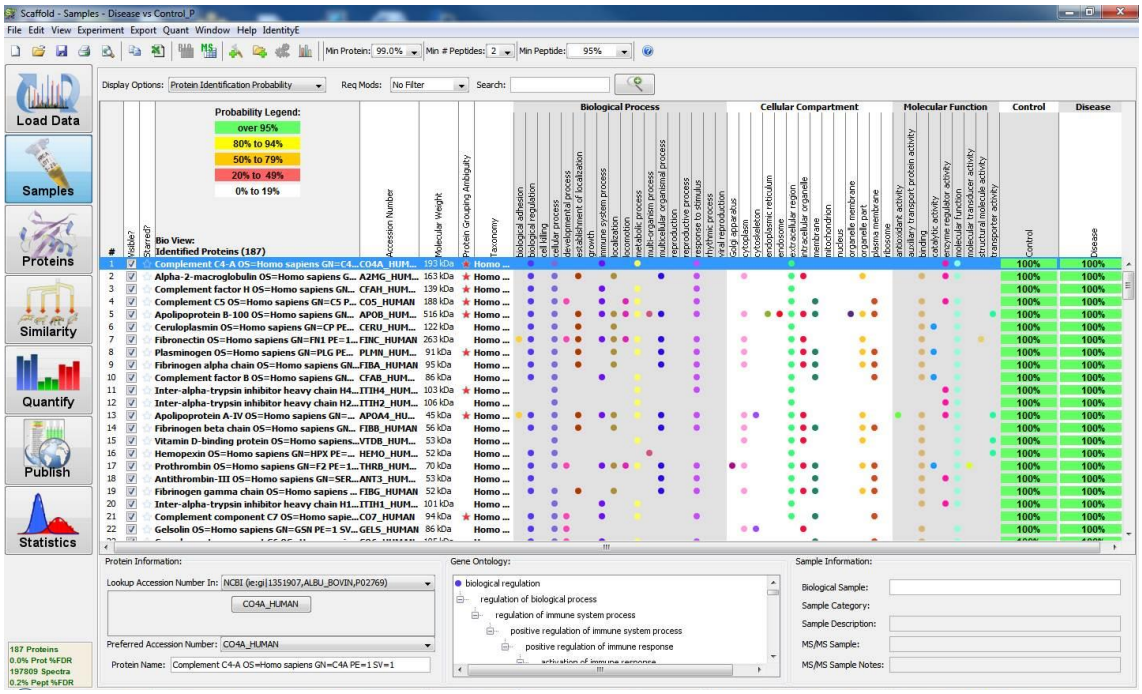

Figure 1.24 The protein list created by Scaffold. It shows protein details (such as protein name, accession number and …) and identification probability in different groups of the experiment design.

Analysis of the uploaded data results in a list of proteins with a number of displaying options. Choosing of the options such as protein identification probability and number of unique peptides re-sorts the protein list so that comparison of proteins can be made using different metrics. Scaffold also allows Gene Ontology (GO) on the samples and shows GO information about every protein in the final list.

4. MASCOT/Scaffold

MASCOT, the search engine provided by Matrix Science, is another data analysis program used for the confident identification of proteins. Unlike other search engines, MASCOT integrates a variety of searching methods for protein identification (unique to MASCOT) including PMF (Peptide Mass Fingerprint, where peptide mass values are the only experimental data (Thiede *et al.*, 2005), Sequence Query (where it uses a combination of amino acid sequence/composition information and peptide mass data

(Mann & Wilm, 1994) and MS/MS ion search (where it uses un-interpreted data from a peptide (Reinders & Sickmann, 2009). Coupling MASCOT data with Scaffold increases the probability of correct expression analysis. PLGS creates a file (.pkl format), following processing of every raw data file, that can be imported into MASCOT and searched against the database. The list of identified proteins can then be exported from MASCOT and imported to Scaffold for protein quantification.

1.3.3.4 Clinical biomarker verification

Verification of the candidate biomarkers found by different expression analysis is another step in the biomarkers discovery pathways (as described earlier in this chapter). All of the candidates must be verified to check and confirm the alteration in their expression analysis. Verification is typically done in a larger number of samples. Two of the most common methods of verifications currently used to confirm the alteration in the expression level of the candidate proteins are Western Blot and SRM (Selected Reaction Monitoring) method.

A. SRM

SRM method is performed on a triple quadrupole mass spectrometer to identify and quantify the candidate protein in a single sample.

SRM is one of the most sensitive methods in MS. Its sensitivity is achieved by statically detecting a specific product ion derived from a precursor. Combined with retention time data, a very high level of specificity is achieved by elimination of chemical Noise. A schematic view of a Waters Xevo TQ mass spectrometer is shown in Figure 1.25.

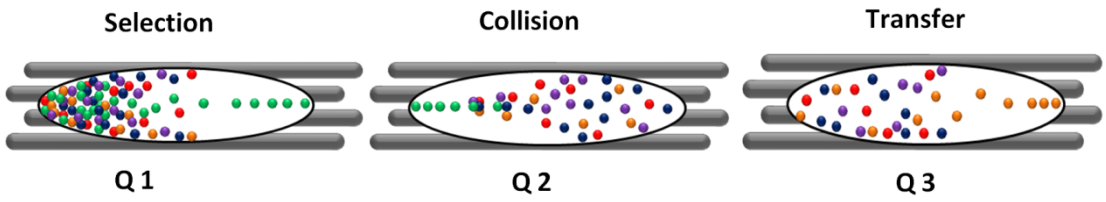

Figure 1.25 Scheme of a SRM experiment on a triple quadrupole mass spectrometer. Precursor Ions are selected in Q 1, collisionally dissociated in Q 2 and a specific fragment are transferred through Q 3 to the detector.

The first quadrupole selects the peptide of interest and transfers it to the second quadrupole where the peptide fragments under the effect of collision gas and an increased acceleration voltage. Fragmented ions are then transferred to the third

quadrupole in which specific fragments are selectively chosen (sequence specific transition ions) to pass through towards the detector.

B. Western Blot

Western blotting is a molecular technique widely used in proteomic studies. Specific antibody-antigen interaction results in detecting the proteins of interest in a complex sample. It was first introduced by Towbin *et al.* in 1979 and well optimised since then (Towbin *et al.*, 1979). Western blot is based on immobilising proteins on a membrane, the protein is then probed with the antibody and then the probed antibody is detected. The western blot protocol is mostly sample dependant but the principles are the same. Briefly, the proteins in the prepared sample are separated by gel electrophoresis. They are then transferred onto a membrane and the areas of the membrane with no protein on are blocked to prevent non-specific binding. Primary antibody binds to the proteins of interest on the membrane and then secondary antibody binds to the primary one. There are different methods for detection of the secondary antibody with the most common one as conjugating an enzyme into the secondary antibody and then using a fluorophore to detect the enzyme. Using specific software, the intensity of the signal detected for each protein is calculated.

To verify the findings of the putative biomarkers in the discovery phase where high resolution mass spectrometry was used, western blots were employed to check that the pattern of differential expression of the putative biomarkers in the sample matched the discovery phase.

1.4 Advantages of using Proteomics for Discovery of Bladder Cancer Biomarkers

Proteomic technology and especially mass spectrometry has become one of the most powerful techniques for the identification and measurement of proteins. It is due to the significant improvements in the separation and detection on LC-MS platforms leading to increases in the dynamic range, much lower levels of sensitivity and improved specificity. Exploiting these improvements enables scientists to attempt biomarker discovery with more powerful tools. Mass spectrometers enable the accurate determination of intact mass and sequence confirmation to an accuracy of less than 2ppm. Additionally, measurement of proteins can now be quantified to low attomole levels or even as low as one copy of protein per cell. Additionally, column chemistries which have much increased peak capacity for the separation of peptides are now used for the introduction of peptides to the mass spectrometer. The coupling of a high performance liquid chromatography with mass spectrometry is the most common method of analysing complex samples and has enabled researchers to profile the protein pattern of biological samples. It is likely that disease biomarkers are among the low abundant proteins of any proteome, thus requiring high levels of sensitivity which maybe beyond the scope of many analytical techniques. A key advantage of mass spectrometry in this case is the dynamic range of the system, which is 4 to 5 orders of magnitude. Whilst plasma has a dynamic range in excess of 10 orders of magnitude, implementing pre-analytical steps such as fractionation and sample preparation steps will enable the detection of moderate to low abundance proteins using high resolution LC-MS. With regards to bladder cancer, using mass spectrometry will enable the discovery of biomarkers which may ultimately help in the development

of diagnostic, prognostic and predictive methods for early and advanced disease management.

1.4.1 Biomarker discovery

Biomarker, as Oxford English Dictionary defines, is “a naturally occurring molecule, gene or characteristic by which a particular pathological or physiological process, disease, etc. can be identified” (*Oxford Dictionary.*2013). Many other definitions for biomarker are proposed in the literature. One of the first was suggested by Sir Richard Doll (1951) when he was researching the relationship between smoking and lung cancer. He defined biomarker as: “the linkage between measurable events and disease”. Biomarkers are increasingly playing key roles in clinical research studies. A biomarker, also called molecular marker, is a molecule which alters as a consequence of a change in a disease prior and/or post treatment. They are also useful for predictive purpose by monitoring the body's response to a treatment. Studies have shown that early detection of cancer can have a profound effect on patient outcome with regards to success in treating cancer (Baker *et al.*, 2002; Skates & Iliopoulos, 2004). Early detection can be enhanced by having a panel of biomarkers which can potentially improve diagnostic specificity. The complete route from discovery of biomarkers through to the development of an assay for diagnosis of the disease consists of many different stages. The main stages are shown in (Figure 1.26) (Rifai *et al.*, 2006).

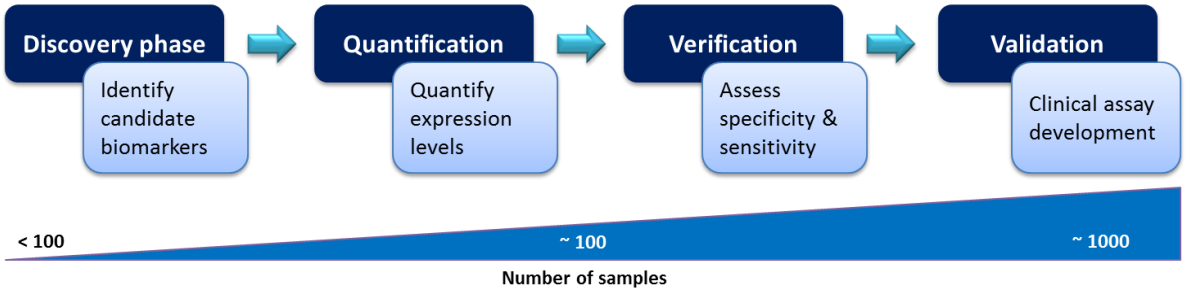

Figure 1.26 The biomarker discovery pipeline

In the proposed study the first two stages (discovery phase and quantification) of the biomarker discovery pipeline and part of the stage 3 (verification) were performed.

1.4.2 Previous research on bladder cancer using proteomics

There are a number of publications that report investigations of proteomics for the diagnosis and prognosis of bladder cancer in the literature using different types of samples such as plasma, urine, tissue and tumour cells. Different techniques of proteomics are used for bladder cancer investigations as of:

MALDI is a method of ionisation in which the peptides that are bound to a matrix become ionised under the effect of laser beam. Its combination with a fractionation method enables urinary protein profiling (Fiedler *et al.*, 2007). In a recent study published by a group of researchers in Germany, (Oezdemir *et al.*, 2012) they were able to stratify tumours into different grades of 79 patients with bladder cancer accurately based on their tissue protein profiling. Profiling of the proteins in the serum of more than 200 samples from bladder cancer and healthy controls using MALDI-TOF mass spectrometry, Schwamborn *et al.* (2009) were able to discriminate patients and healthy samples with high sensitivity and specificity (Schwamborn *et al.*, 2009). MALDI also has been used for early detection of muscle invasive disease by comparison of the protein profiles from patients in very early stages of the disease and high grade/stage

tumours. Significant over-expression of Choroideremia-like protein was reported by Li J *et al.* (2008) as a marker for early detection of advanced disease (Li *et al.*, 2008).

SELDI (Surface Enhanced Laser Desorption Ionisation) similar to MALDI, a laser beam ionises the proteins which are bound to the surface of a chip. Since it is one of the simplest methods, it needs very low volume of sample and is high throughput (evaluated by (Schaub *et al.*, 2004)); it has been used in many studies for biomarker discovery. Langbein (2006) compared 2D-PAGE with SELDI-TOF mass spectrometry and found the latter more reliable (Langbein *et al.*, 2006). SELDI has been used (Munro *et al.*, 2006) for highly accurate discrimination of bladder cancer samples from healthy controls. Using machine learning methods (Tree-Analysis) and SELDI, Liu *et al.* (2004) were able to discriminate the bladder cancer patient samples from healthy controls (Liu *et al.*, 2005).

2DE-MS is the separation of proteins in two dimensions (as a fractionation step) and then the separated proteins are introduced into the mass spectrometer. Using this method Li *et al.* (2011) reported overexpression of Apolipoprotein A1 in bladder cancer patient samples when compared to controls and called it as a potential biomarker for bladder cancer (Li *et al.*, 2011).

LC-MS is the most commonly used method for protein profiling and biomarker discovery in the proteomics field. Using liquid chromatography as a separation step before analysing the samples in mass spectrometer enhances the level of detection and increases the dynamic range of identified proteins. In a recent publication (Linden *et al.*, 2012) 29 proteins were found up-regulated in the urine samples of bladder cancer patients when comparing their profile with healthy controls. The main

methodology utilised LC-MS. Another study conducted by a group of researchers in Taiwan (Tyan *et al.*, 2011) found over-expression of 13 proteins in urinary protein profiles of bladder cancer samples when compared with healthy controls. They identified these up-regulated proteins using nano-LC-MS, three of which were confirmed by western blot. Niu *et al.* (2011) used tissue samples from patients with invasive tumours and normal bladder wall tissue from healthy controls to profile their protein patterns using LC-MS (Niu *et al.*, 2011). They found 488 proteins unique in tumour cells and 493 proteins unique in healthy cells. Although there are lots of publications with regards to bladder cancer biomarker discovery using liquid chromatography mass spectrometry, up to now no one has used label free profiling of plasma proteins for identification and quantification of bladder cancer biomarkers. Additionally there is no FDA approved protein marker for the clinical management of bladder cancer that was identified using proteomic strategies.

1.4.3 Urine or plasma?

Profiling of the proteomic pattern of body fluids can lead to the development of a (or a panel of) novel biomarker(s) which may enable the early detection of cancer. Having a disease biomarker identifiable in the biological fluids such as plasma and urine and measurable by a relatively easy method has become the main goal of the researchers in the clinical proteomics field (Hu *et al.*, 2006). With regards to bladder cancer, there is no clear agreement about which sample (plasma or urine) should be used for proteomic analysis of bladder cancer (Bischoff & Clark, 2009). Some research groups suggest that urine is the best source for investigation of bladder tumours due to its direct contact with the tumour in bladder wall hence the possibility of finding cancer related protein secreted into the urine is higher (Shao *et al.*, 2011; Pejcic *et al.*, 2010). In

contrast, researchers using plasma samples for proteomic study of bladder cancer argue that the high number of proteins present in plasma compared to urine increases the chance of detecting plausible proteins. These proteins may also be present in higher concentrations in the circulatory system (if not excreted) than markers that are quickly voided from the bladder via urine. Consequently plasma is arguably the most advantageous source for biomarker discovery in the field of proteomics (Farrah *et al.*, 2011; Omenn, 2006).

Although urine samples are non-invasively obtainable without volume limitation, there are still a number of disadvantages in the usage of urine compare to the plasma in proteomic profiling of patients with bladder cancer.

Urinary proteomics disadvantages:

Variability in the protein content, there is huge biological variability between urine samples of different donors compared to plasma samples. Many factors can alter the protein content of the urine sample from patient to patient such as the progress of the disease, kidney disease, patient's diet and hydration (Rai, 2010). On average plasma protein concentration in healthy donors is about 60mg/mL whereas in urine of the same cohort, protein concentration is in the µg/mL region which makes urine a more problematic sample to deal with. This lower concentration of proteins in the urine can be overcome by concentrating the sample using concentrator filters that can contribute to preanalytical variability and the loss of some proteins.

The first pass urine sample collected in the morning is much more concentrated than the second and the third pass urine samples which results in different protein patterns between the samples (Reinders & Sickmann, 2009).

Presence of infectious and inflammatory proteins originating from the urinary tract due to severe conditions also creates variability between the urine samples. To overcome this issue it has been suggested to discard the first 5-10mL of the urine pass and use the rest for the analysis which in turn adds another variable to the analysis (Rai, 2010).

Thus in consideration of the reasons mentioned above, and also the availability of plasma samples at the time of this investigation, plasma samples from patients with bladder cancer were used for the purpose of this project.

1.5 Aims and objectives

There is a paucity of biomarkers for the clinical management of bladder cancer. This thesis is testing the hypothesis that novel biomarkers for the diagnosis and prognosis can be attained via proteomic strategies. In particular, this current study aims to characterise plasma proteins using a label free proteomic platform which couples ultra performance liquid chromatography, and mass spectrometry and novel informatics strategies. Prior to testing clinical samples, the methodology needs to be assessed for reproducibility and suitability for analysing valuable clinical samples supplied by colleagues from the University of Torino, Turin, Italy.

The main aims of this project were:

- To undertake the development of label-free UPLC-DIA-MS on a Waters Synapt G2 HDMS platform.
- To evaluate and optimise sample preparation using immunodepletion of highly abundant proteins with the developed label-free UPLC-DIA-MS method.
- To use these methods to analyse a set of plasma samples obtained from collaborators at the University of Torino. The experimental design will enable a series of questions to be asked that can help discover plasma based biomarkers of bladder cancer for different clinical scenarios.

Chapter Two

MATERIALS AND METHODS

2 Materials and methods

2.1 Materials and instruments:

All chemicals, reagents and immunodepletion columns were bought from Sigma-Aldrich (Poole, UK). All solvents for nanoACQUITY UPLC were purchased from Fisher-Scientific Inc. (Loughborough, UK). A Varian ProStar HPLC system (AuroSampler 410, UV detection 325 and the pump 230) was used for immunodepletion. Pierce BCA protein assay kit was purchased from Thermo Scientific (IL, USA). A Fluostar Optima spectrophotometer from BMG-Labtech (Offenburg, Germany) was used for measuring the absorbances. Trypsin gold (proteomics grade) was bought from Promega (Madison, USA). Spin concentrator filters were purchased from Agilent Technologies Inc. (Berkshire, UK). Tryptic peptide standards (Enolase MassPrep digest standard and alcohol dehydrogenase MassPrep digest standard) and analytical columns were bought from Waters Corp. (Herts, UK).

2.2 Methods:

2.2.1 Sample preparation

2.2.1.1 Control Plasma sample

The human blood samples were collected in EDTA blood sample tubes from a healthy donor who had provided informed consent. The blood samples were centrifuged at 1500 x g for 20 minutes at 4°C. The plasma layer was separated from the buffy layer and red blood cells, and stored at -80°C for further use.

2.2.2 High abundant protein depletion

2.2.2.1 ProteoPrep 20

ProteoPrep 20 plasma Immunodepletion kit (Prot20S) was used for depletion of the top 20 proteins from plasma sample.

Procedure:

1. **Buffer preparation:**

Both of the buffers (equilibration and elution) were diluted to 1X (10X original) using HPLC grade water to the total volume of 5ml and 15ml per sample for elution buffer and equilibration buffer, respectively.

2. **Column Equilibration (room temperature)**

The bottom plug was removed from the spin column, the upper cap was loosened and the spin column was placed in a 2ml collection tube. The column was centrifuged for 60 seconds at 4500rpm. The upper screw cap was replaced by a Luer-lock cap and the spin column was placed on a 15ml universal tube for waste collection. Using a Luer-lock syringe, 4ml of 1X equilibration buffer was slowly passed through the spin column

by attaching it to the Luer-lock cap. The remaining buffer was removed from the column by centrifugation of the spin column on a 2ml collection tube for 30 seconds at 4500rpm after replacing the Luer-lock cap/syringe by the screw cap. The collected buffer was discarded and the dried spin column was placed on a new collection tube.

3. Plasma depletion

40µl of on-ice thawed plasma sample was diluted to 500µl using 1X equilibration buffer. Diluted sample was then filtered using Corning SpinX centrifuge tube filters at 4500rpm for 2 minutes. 100µl of diluted plasma (maximum capacity of the column) was added into the depletion column and incubated for 20 minutes at room temperature (should be repeated 5 times for 500µl of diluted plasma). Depleted plasma was centrifuged at 4500rpm for 60 seconds and the flow through saved in a new collection tube. For the column to be ready for next round of depletion, high abundance proteins bound to the column should be eluted off by following the elution steps (stage 4 of this section) and the column should be re-equilibrated by following the equilibration steps (stage 5 of this section). The remaining diluted sample (400µl) was incubated in the depletion column (100µl each time) followed by elution and re-equilibration of the column after every incubation (same as the first 100µl). After final depletion, the column was washed twice with 100µl of 1X equilibration buffer followed by 60 seconds of spinning at 4500rpm. At this stage 95% of the top 20 of high abundant proteins of plasma are depleted.

4. Elution of Bound Proteins:

The spin column screw cap was replaced by a Luer-lock cap and the spin column was placed on a 15ml universal tube for waste collection. Using a Luer-lock syringe, 2ml of

1X elution buffer was gently passed through the spin column (no more than one drop per second). The remaining buffer in the column was removed by centrifugation into a 2ml collection tube for 30 seconds at 4500rpm after replacing the Luer-lock cap and syringe with the screw cap. The collected buffer was discarded and the spin column was placed on a new collection tube.

5. **Re-equilibration of the column:**

The Luer-lock cap was replaced with the screw cap on a new 2ml collection tube. All of the equilibration steps (above) were repeated to re-equilibrate the column for the next depletion round.

6. **Concentrating of multiple depletion:**

Vivaspin-500 centrifugal concentrators (2 per sample) were primed by adding 400µl of HPLC grade water into them (200µl each). Concentrator tubes were centrifuged at 7500rpm and 4°C for 15 minutes and the remaining water was removed from inside the tube by flicking it a couple of times. 150µl of depleted plasma was added into each Vivaspin concentrator and centrifuged at 7500rpm for 30 minutes at 4°C. After spinning, the sample volume is reduced to 40-50µl which should be removed from inside the filter using a gel loading tip. Proteins stuck onto the filter or around the top of the tube were washed by pipetting 100µl of equilibration buffer up and down in the filter. This wash was also pooled with the concentrates.

7. **Final depletion:**

To the re-equilibrated depletion column, 100µl of depleted-concentrated plasma was added and incubated for 20 minutes at room temperature. The column was spun down

at 4500rpm for 60 seconds and flow through was saved in a collection tube. Elution and re-equilibration steps were performed post incubation. Same incubation procedure was carried out for the rest of depleted-concentrated plasma (100µl each time). After the final depletion, the depletion column was washed twice with 100µl of 1X equilibration buffer following by a 60 seconds of spinning at 4500rpm. At the end of this stage 99% of the top 20 high abundant proteins of plasma sample is depleted.

2.2.2.2 SEPPRO IgY-14 LC2

Procedure:

1. **Buffer preparation**

All three buffers were diluted from 10X to 1X using HPLC grade water and degassed for 15 minutes using an ultrasonic water bath. Storage buffer was made by making a solution of 0.02% sodium azide in 100ml of the dilution buffer.

2. **Sample preparation**

A 50µl aliquot of plasma sample (thawed on ice) was diluted 4-fold (250µl final volume) using 1X dilution buffer. Particulate materials which could clog the column were filtered using Corning SpinX centrifuge tube filters over the centrifuge at 9000rpm for 1.5 minute.

3. **HPLC preparation**

All the lines were purged for 5 minutes at 2ml/min. System back pressure was checked before connecting the column using 100% dilution buffer (buffer A) at 2 ml/min for 5 minutes. Maximum pressure for the column to avoid damage is 350psi which is equal to 23.81atm (the unit used in the HPLC system). The column introduces 3-4 atm

pressure to the system so the system pressure should not be more than 20atm at this stage. After the pressure check, the column was attached and equilibrated using 100% buffer A at 2ml/min for 20 minutes to obtain a flat baseline in the UV readout. In the meantime, a gradient based on the following table was created (Table 2.1).

	Cycle	Time (min)	Dilution Buffer (%)	Elution Buffer (%)	Neutralization Buffer (%)	Flow Rate (ml/min)	Maximum Pressure (atm)
1	Injection	0	100	0	0	0.2	23
2	Depletion	17.00	100	0	0	0.2	23
3	Depletion	17.01	100	0	0	1.5	23
4	Elution	22.00	0	100	0	1.5	23
5	Elution	22.01	0	100	0	1.5	23
6	Neutralisation	36.00	0	0	100	1.5	23
7	Neutralisation	36.01	0	0	100	1.5	23
8	Re-equilibration	42.00	100	0	0	1.5	23
9	Re-equilibration	42.01	100	0	0	1.5	23
	End of the run	50					

Table 2.1 The gradient used for immunodepletion with SEPPRO IgY-14 LC-2 column

As a blank run, 200µl of dilution buffer was injected using the above gradient to check if the column is clean and ready for injection of the sample.

4. High abundance protein depletion

The ProStar 410 autoSampler (part of the Varian HPLC system) was set to inject 227µl of diluted filtered plasma sample based on the micro litre pick-up injection method. The volume of 227µl was calculated based on the loop size (500µl), needle size (15µl) and tubing size (250µl) installed on the autoSampler. The first fraction (low abundance proteins) collected between 4 and 20 minutes in a 15ml falcon tube placed on ice with the total volume of 5-7ml. The collected fraction was stored at -20°C straight away for further analysis. The rest of the gradient (elution of high abundant proteins and

neutralisation of the column) was directed to waste. Another blank run using 200µl of dilution buffer was injected at the end of the runs to eliminate carryover.

Both of the immunoaffinity methods explained above have advantages and disadvantages summarised in Table 2.2.

	ProteoPrep 20	SEPPRO IgY-14
Average number of proteins identified post-depletion	90-120	120-170
Required time for depletion of a single plasma sample	7 hours	50 minutes
Cost per sample	~£80-200	~£40
Yield of depletion	99%	95%
Number of proteins to be depleted	20	14

Table 2.2 Advantage and disadvantages of the two immunoaffinity methods for depletion of high abundant proteins

Although ProteoPrep 20 has advantages (depletion of up to 99% of top 20 of high abundant proteins compared to 95% depletion yield of top 14 proteins in SEPPRO IgY-14), it is time consuming, labour intensive and costly. Also it results in lower number of proteins identified post-depletion. Hence the proteomics pipeline has been designed based on the usage of SEPPRO IgY-14 for immunodepletion of high abundant proteins.

2.2.3 Concentration and buffer exchange

A. Acetone precipitation (post ProteoPrep 20 depletion)

A solution of 50mM ammonium bicarbonate was prepared by adding 10mg of ammonium bicarbonate into 2.5ml of HPLC grade water. Plasma sample depleted from high abundance proteins using ProteoPrep 20 column was transferred into an acetone-safe tube capable of holding 5 times the sample volume. A volume of ice-cold (-20°C) 100% acetone equal to 4 times the sample volume was added to the tube. Samples were briefly vortexed and incubated at -20°C for 60 minutes. After incubation proteins were spun down at 15000g for 30min at 4°C and the supernatant removed carefully. Extra care was made not to dislodge the protein pellet. The precipitation procedure was repeated again using the above mentioned steps. At the end of second round of precipitation, the pellets were air-dried by keeping the tube lid open at room temperature for no more than 30 minutes. Dried pellets were re-constituted in 250µl of 50mM ammonium bicarbonate. The incompatible buffer has now exchanged and the sample is ready for further analysis.

B. On-filter buffer exchange (post SEPRRO depletion)

Plasma samples depleted with the SEPPRO column were concentrated down to 350µl (5-7ml original volume) using 4.5ml Agilent concentrating tubes with 5K MWCO filters by centrifuging at 6000rpm for 20 minutes. As an alternative to acetone precipitation, the Tween20 in the samples was exchanged with 50mM ammonium bicarbonate on the same filters that were used for the concentration. The concentrated samples were washed by adding 4ml of 50mM ammonium bicarbonate to the filter followed by centrifugation at 6000rpm for 20 minutes. This wash step was repeated 4 times to

ensure all of the interfering substances are washed away from the sample leaving the depleted sample in 50mM ammonium bicarbonate. Although the procedure is time consuming compared to acetone precipitation, it overcomes the disadvantages of the acetone precipitation procedure such as the difficulties in the reconstitution of the pellet back into the solution which result in the loss proteins and also the peptide modification in presence of acetone residuals in the protein pellet (Simpson & Beynon, 2010).

2.2.4 Protein quantitation assay (BCA assay)

Procedure:

A. Standards preparation:

BSA standard stock solution (1ml ampoule of 2mg/ml) was aliquoted into 10 eppendorf tubes of 100µl each. One aliquot was used for making up the concentrations and the rest were stored in -20°C. Seven different concentrations of the standard were prepared based on the table below (Table 2.3) and HPLC grade water was used as a blank.

Vial	Volume of dH_2O (µl)	Volume of BSA (µl)	Final concentration (µg/µl)
A	0	30	2
B	12.5	37.5	1.5
C	32.5	32.5	1
D	17.5	17.5 of B	0.75
E	32.5	32.5 of C	0.5
F	32.5	32.5 of E	0.25
G	32.5	32.5 of F	0.125
H	40	0	Blank

Table 2.3 Different concentration to be prepared for standard curve preparation

B. Working reagent preparation:

The volume of working reagent (WR) to be prepared was calculated based on the following equation:

$$\boldsymbol{Total\ volume\ of\ WR\ required\ (ml) = (8 + number\ of\ samples) \times 3 \times 0.2}$$

Equation 2.1 Equation for calculation of working reagent for BCA assay

In the above equation, 8 is the number of standards (7 standards and a blank), number 3 is for measuring every sample 3 times (in triplicate) and 0.2 is the volume of working reagent needed for each sample containing well. After the total volume of working reagent was calculated, it was prepared by adding 50 parts of reagent A into 1 part of reagent B followed by a brief mixing over vortex.

C. Procedure:

First three columns of the 96-well plate were used for the standard. 10µl of the standards and the blank were added into each well in triplicate. Samples were also added into the wells in triplicate (Figure 2.1).

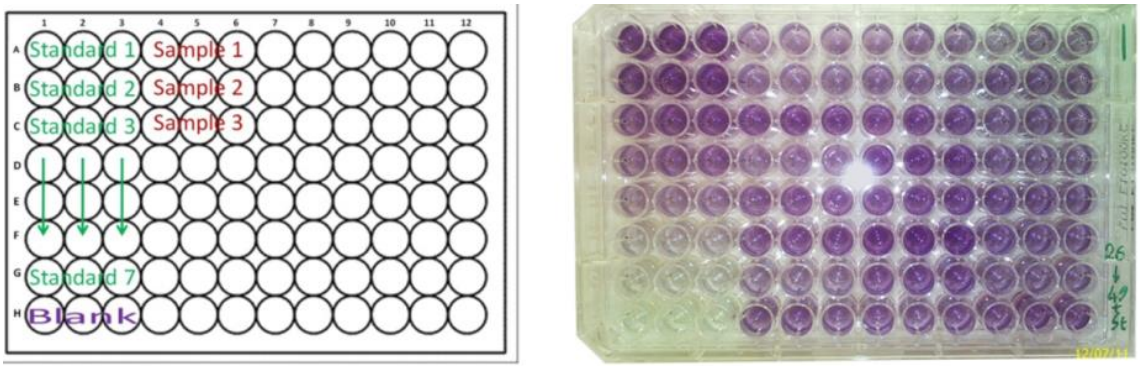

Figure 2.1 A 96-well plate used for BCA assay. Schematic lay out of the plate (left) and purple colour representing the protein concentration in each well post incubation (right) are shown.

To each well, 200µl of the prepared working reagent was added and mixed for 30 seconds. The plate was covered and incubated at 37°C for 30 minutes. The plate was then cooled down to room temperature before measuring the absorbances at 595nm on a Fluostar Optima spectrophotometer. Average measurement of every triplicate was calculated and the blank was subtracted from all other measurements (including standards and samples) (Table 2.4).

Concentration	Abs.1	Abs.2	Abs.3	Average	Final absorbance
2	0.753	0.729	0.724	0.735	**0.520**
1.5	0.603	0.592	0.621	0.605	**0.390**
1	0.431	0.478	0.496	0.469	**0.253**
0.75	0.428	0.418	0.438	0.428	**0.213**
0.5	0.360	0.345	0.361	0.355	**0.140**
0.25	0.296	0.278	0.296	0.290	**0.074**
0.125	0.254	0.247	0.274	0.258	**0.043**
0	0.223	0.210	0.214	0.216 ==>	**Blank**

Table 2.4 Absorbances of the standards and the blank used for BCA assay standard curve creation

A standard curve was prepared by plotting the average measurements for each standard vs. its concentration (µg/µl) in Excel (Figure 2.2). Using the equation determined by the standard curve, protein concentration of the samples was calculated.

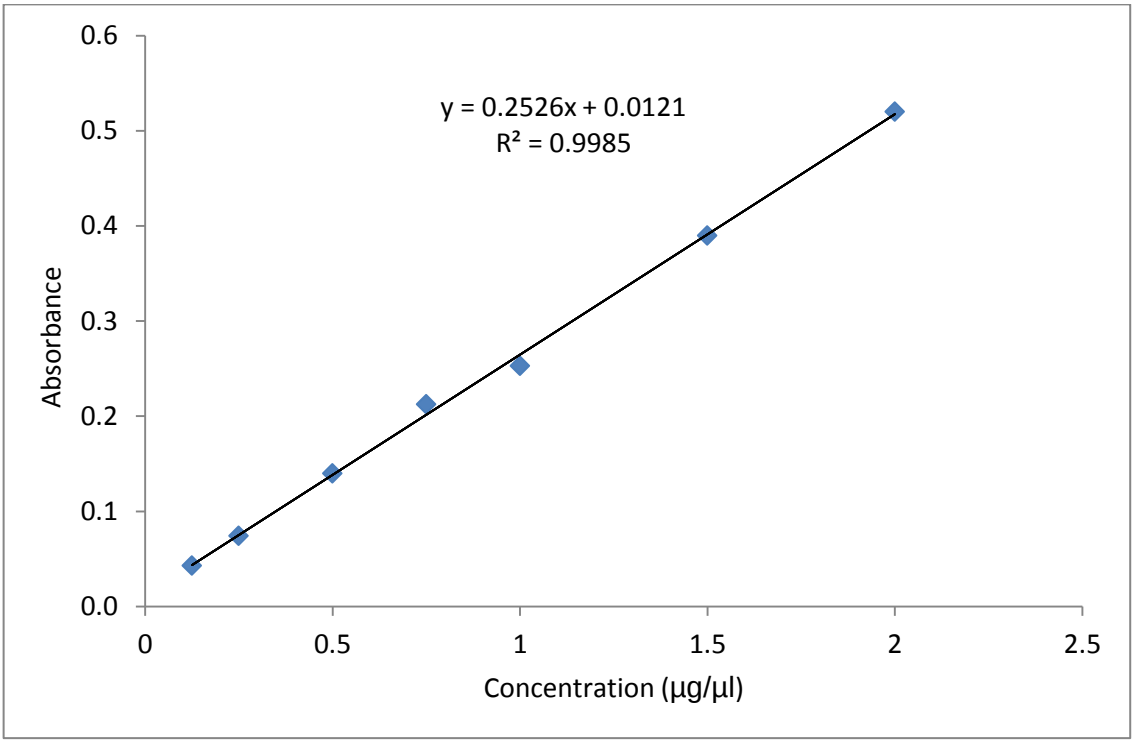

Figure 2.2 A standard curve plot of the absorbances vs. concentration. The equation was used for calculation of the protein concentration in the samples.

In the equation, y is the absorbance of a sample and x is the concentration of the protein in the sample. This equation will be used to calculate protein concentration of samples in the discovery phase of the study.

2.2.5 Digestion

A. Buffer Preparation:

50mM ammonium bicarbonate was prepared in HPLC grade H_2O and 100mM DTT solution, 200mM IAA solution and 1 µg/µl trypsin solution were prepared in 50mM ammonium bicarbonate.

B. Procedure:

For every 100µg of protein in the sample (protein concentration assessed with the BCA protein assay), 5µl of 100mM DTT was added to the samples followed by a brief vortex and incubation at 60°C for 15 minutes. After incubation, 5µl of 200mM IAA was added to samples and incubated at room temperature for 30 minutes in the dark. To digest the proteins, 2µl of 1µg/µl trypsin was added to the samples (1:50, trypsin:protein ratio) and incubated at 37°C overnight (12-18 hours). After the incubation, 2µl of concentrated formic acid was added to the samples to stop the digestion following by a brief vortex and re-incubation at 37°C for 20 minutes. The sample tubes were centrifuged at 12000rpm for 10 minutes. The supernatant containing peptides was carefully collected and transferred into a new 1.5ml tube for further analysis.

2.2.6 Sample analysis

2.2.6.1 Instrument preparation

A Waters Synapt G2 mass spectrometer (Waters Corp., Herts, UK) coupled to a nanoACQUITY UPLC was used for separation and analysis of tryptic peptides. A trapping column with 180µm inner diameter and 20mm long packed with 5µm Symmetry C_{18} material connected to an analytical column with 75µm inner diameter and 250mm long packed with 1.7µm BEH130 C_{18} material were used for peptide separation. Formic acid (0.1%) in LCMS grade water was used for mobile phase A and 0.1% formic acid in acetonitrile for mobile phase B. The pre-set gradient was performed at the flow rate of 0.3µl/min and the column temperature was 40°C. To be able to quantify the absolute amount of every identified protein an internal standard, Yeast alcohol dehydrogenase (ADH) at the concentration of 50fmol/µl was added to the samples at the ratio of 1:1. ADH is a protein extracted from yeast, and thus absent in human databases. For the mass spectrometer to deliver reliable results a series of checks should be done before injection of any sample to the machine. These checks are semi-automated and reports indicating pass or fail are generated. Prior to the analysis of a large batch of samples the detector voltage, calibration, sensitivity and chromatographic reproducibility should be checked. A specific protein standard was used for every one of these checks. Leucine Enkephalin (Leu Enk) with mass of 555.2771 Da was used to check the sensitivity of detector (detector voltage). The detector check ensures that the mass spectrometer (especially the ToF) has adequate nominal sensitivity and that the voltage of the detector is appropriately optimised for maximal signal. After the detector check, the system was calibrated using Glu-Fibrinopeptide β (GFP) at the concentration of 500fmol/µl with m/z of 785.84265. The

calibration ensures that the recording of an ion is accurately determined by MS across the mass range of m/z 50-2000. GFP was also used for LockMass solution at a flow rate of 0.5µl/min as a reference signal. Following a successful calibration the nanoACQUITY UPLC was checked to assess chromatographic reproducibility. For this purpose, enolase at the concentration of 25fmol/µl was used based on a 20min gradient. This standard is commercially available as tryptically digested form of enolase. The resultant chromatograms (TIC and BPI) were inspected to identify a number of expected peaks (Figure 2.3). Total ion count (TIC) chromatogram shows the sum intensity of all of the peaks at a single time point whereas base peak intensity (BPI) chromatogram shows the most intense peak at a single time point.

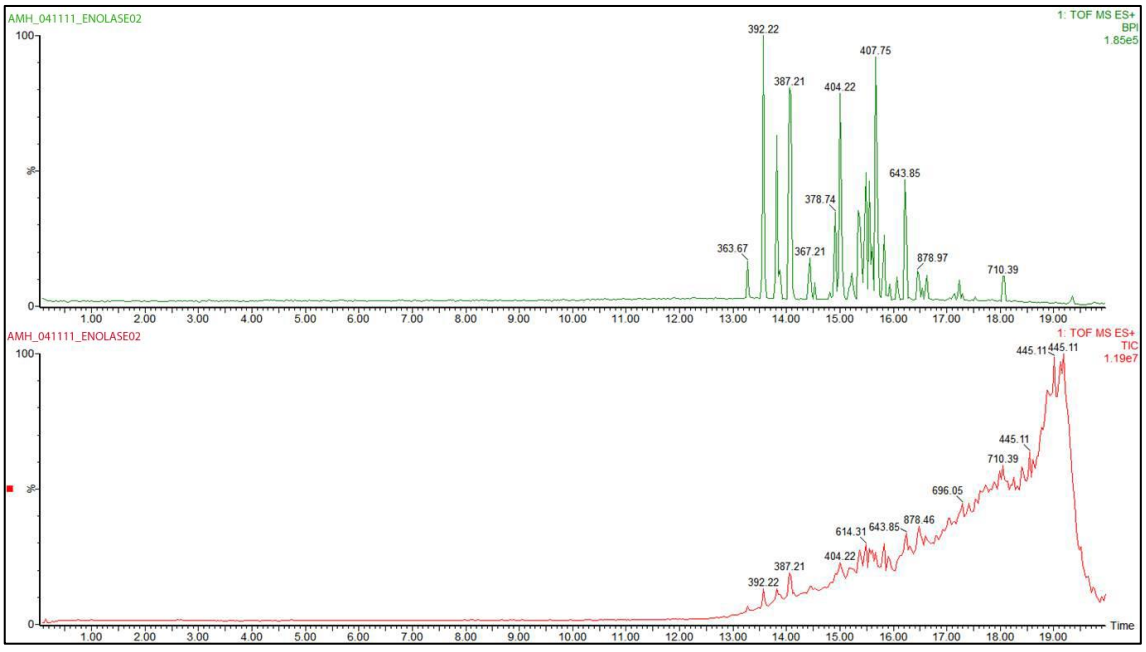

Figure 2.3 A chromatogram of an enolase run. A TIC chromatogram (bottom) shows the sum intensity of all detected peaks in a single point and a BPI chromatogram (top) shows the most intense peak in a single time point

To asses this further, five tryptic peptide ions of enolase with known m/z (643.9, 644.9, 708.9, 789.9 and 878.5) were extracted from the chromatogram (Figure 2.4).

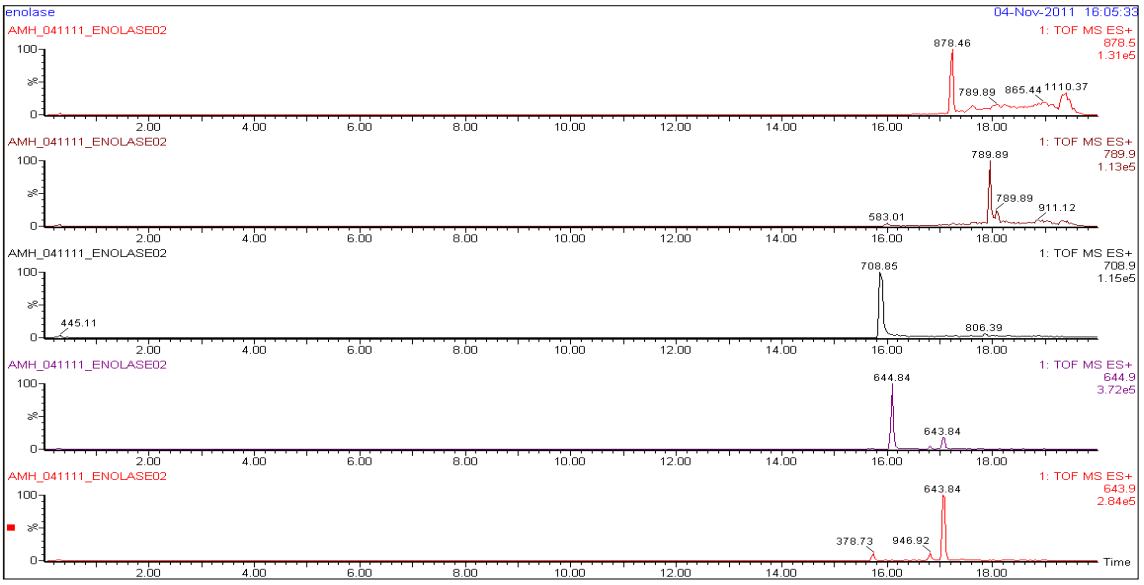

Figure 2.4 Extracted ion chromatograms for the 5 ions (643.9, 644.9, 708.9, 789.9, and 878.5) of enolase. Peak width, retention time and intensity are the parameters to check from the extracted ion chromatograms.

The peaks were assessed for their retention time, peak width and peak intensity (typically in e^5 range). Finally, the raw data file was processed by PLGS to check the coverage of the enolase. PLGS coverage of 40% or more demonstrates (Figure 2.5) that the instrument is at its optimum performance and ready to start the sample run.

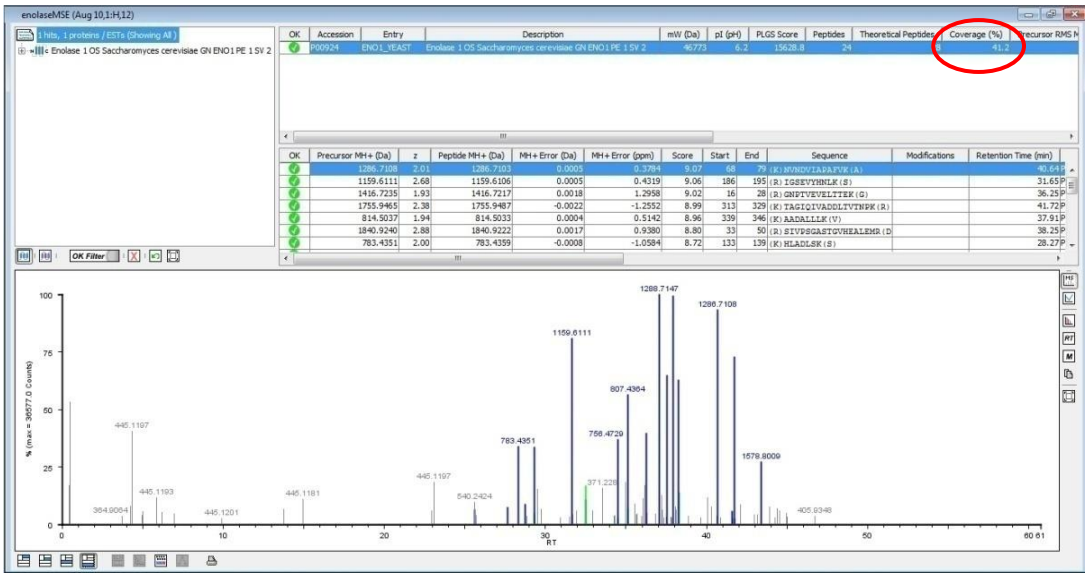

Figure 2.5 PLGS results of processing the enolase raw data file. Coverage of ≥ %40 shows that the instrument is ready for analysis of the samples.

2.2.6.2 Mass spectrometry analysis

Chromatographically separated peptides were introduced into a Waters Synapt G2 mass spectrometer (Waters Corp., Herts, UK) through the nano-spray emitter stage with a PicoTip (New Objective, MA, USA). The source temperature was set to 70°C and the capillary voltage to 3.0kV. The low collision energy was set to 4V to obtain the intact peptide information and elevated energy was set to a ramp between 15-40V for peptide dissociation. MassLynx 4.1 was used to acquire the data. A sample list (Figure 2.6) was created on MassLynx where every detail regarding a sample was inserted from sample ID to the MS method (settings by which MS acquire data), the gradient to be used for analysis of the samples, sample location in the sample manager part of the nanoACQUITY UPLC and the injection volume.

Spectrum Chromatogram Map Edit ▾ Samples ▾

	File Name	File Text	MS File	Inlet File	Bottle	Inject Vo...
1	AMH_12032011_S1a	Sample 1 a			2:A,1	2.000
2	AMH_12032011_S1b	Sample 1 b			2:A,1	2.000
3	AMH_12032011_S1c	Sample 1 c			2:A,1	2.000
4	AMH_12032011_Wash 1	wash with 1% Formic Acid			2:B,1	5.000
5	AMH_12032011_Wash 2	wash with 0.1% Formic Acid			2:B,2	5.000
6	AMH_12032011_Wash 3	wash with 0.1% Formic Acid			2:B,2	5.000
7	AMH_12032011_Enolase1	Enolase			2:C,2	2.000
8	AMH_12032011_S2a	Sample 2 a			2:A,2	2.000
9	AMH_12032011_S2b	Sample 2 b			2:A,2	2.000
10	AMH_12032011_S2c	Sample 2 c			2:A,2	2.000
11	AMH_12032011_Wash 4	wash with 1% Formic Acid			2:B,1	5.000
12	AMH_12032011_Wash 5	wash with 0.1% Formic Acid			2:B,2	5.000
13	AMH_12032011_Wash 6	wash with 0.1% Formic Acid			2:B,2	5.000
14	AMH_12032011_Enolase2	Enolase			2:C,2	2.000

Figure 2.6 A typical sample list created using MassLynx

2.2.7 Data analysis

2.2.7.1 Protein identification and quantification

Raw data derived from mass spectrometer were processed using ProteinLynx Global Server 2.5 (PLGS) (Waters Corp., Herts, UK). Database, workflow and processing parameters for PLGS were set as:

Database: the human reviewed canonical sequence data in FASTA format was downloaded from the Uniprot on July 2011 and used in the processing workflow (Figure 2.7).

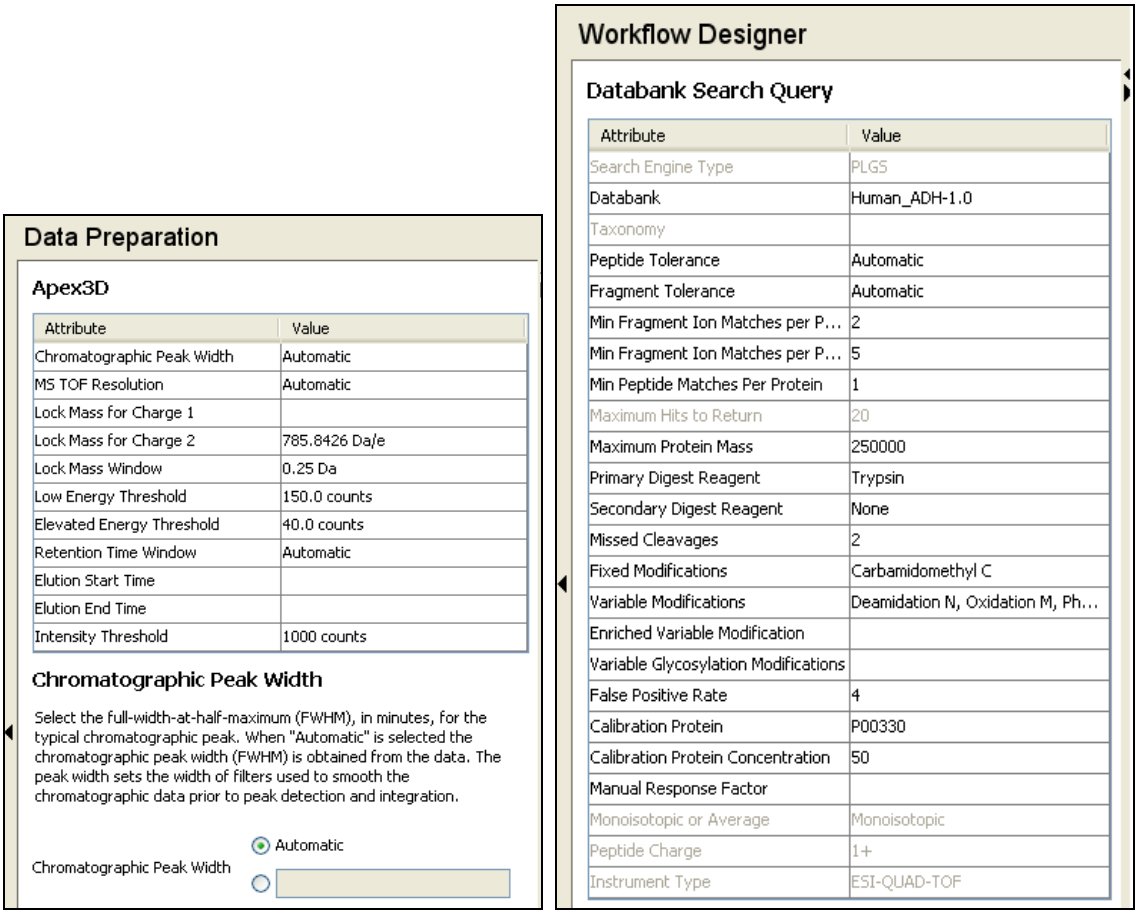

Figure 2.7 Processing parameters used for data preparation (left) and parameters used to design workflow (right) for databank searching in PLGS. In the workflow designer, minimum number of fragment ion matches required for a peptide and a protein to remain under consideration was chosen as 2 and 5, respectively. Minimum number of peptide matches per protein was set to 1.

In the workflow design, trypsin was chosen as the enzyme used for digestion with maximum 2 missed cleavages. It was assumed that every cysteine is modified at the time of alkylation using IAA hence carbamidomethyl C was selected as the fixed modification. Oxidation of methionine (M), phosphorylation of tyrosine (Y), threonine (T) and serine (S) and deamidation of asparagine (N) were selected as variable modifications. The digest protein used as internal standard and its concentration were also set in the workflow designer. Each sample raw data file was processed based on the set processing parameters and workflow. ADH was used as internal standard and its sequence was downloaded and appended into the human database. PLGS uses the concentration of internal standard in order to quantify the amount of all proteins within the sample and creates a list of identified proteins with the details relevant to each protein (Figure 2.8).

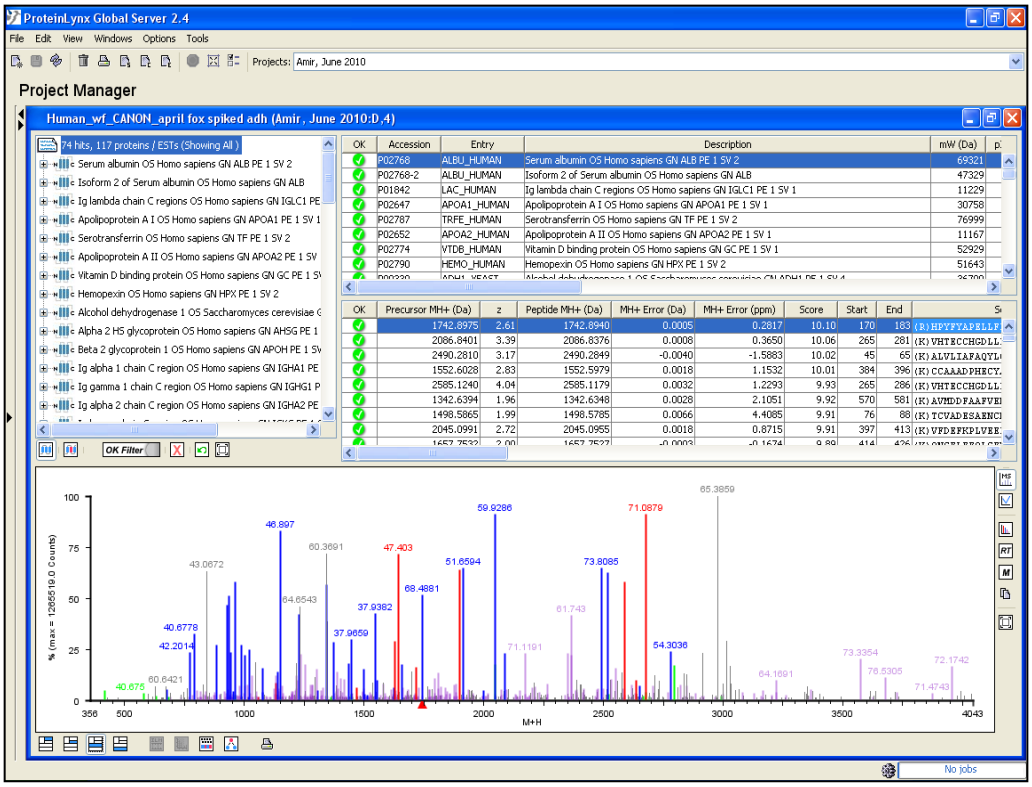

Figure 2.8 Results from processed data of an analysed sample. Total number of proteins, details about every protein and a chromatogram showing retention times are provided.

Hit is the total number of identified protein (protein groups) whereas **Protein** is the number of protein and related iso-forms. The number of hits is reported throughout this project. The top right hand side table (Figure 2.8) contains all the details about every identified protein were exported into excel sheets where further assessments were done. The exported Table data shows the information regarding identified proteins; including absolute amount of each protein in nano-grams (Figure 2.9).

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Entry	Description	mW (Da)	pI (pH)	Peptides	Theoretical Peptides	Coverage (%)	Precursor RMS Mass Error (ppm)	Products	Digest Peptides	Modified Peptides	Products RMS Mass Error (ppm)	Products RMS RT Error (min)	Amount (fmol)	Amount (ngrams)
2	ALBU_HUMAN	Serum albumin OS Homo sapiens GN ALB PE 1 SV 2	69321	5.8601	131	55	80.4598	2.1768	2226	75	20	5.1854	0.012051	1289.955	86.1603
3	ALBU_HUMAN	Isoform 2 of Serum albumin OS Homo sapiens GN ALB	47329	5.9225	78	36	77.2182	2.3427	1419	46	12	5.2136	0.012369	11.3919	0.5395
4	LAC_HUMAN	Ig lambda chain C regions OS Homo sapiens GN IGLC1 PE 1 SV 1	11229	7.0734	6	7	82.8571	1.46	90	6	0	4.9228	0.011501	52.0356	0.5847
5	APOA1_HUMAN	Apolipoprotein A I OS Homo sapiens GN APOA1 PE 1 SV 1	30758	5.4309	61	32	74.1573	1.8895	713	38	7	6.0606	0.012003	267.9127	8.2458
6	TRFE_HUMAN	Serotransferrin OS Homo sapiens GN TF PE 1 SV 2	76999	6.7509	69	76	72.7794	2.1807	986	54	5	6.3411	0.011107	267.6895	20.6255
7	APOA2_HUMAN	Apolipoprotein A II OS Homo sapiens GN APOA2 PE 1 SV 1	11167	6.6396	20	11	74	2.2285	274	11	1	5.8389	0.013238	168.0204	1.8776
8	VTDB_HUMAN	Vitamin D binding protein OS Homo sapiens GN GC PE 1 SV 1	52929	5.2381	39	45	69.4093	2.0376	474	30	1	6.1235	0.012222	178.0877	9.4322
9	HEMO_HUMAN	Hemopexin OS Homo sapiens GN HPX PE 1 SV 2	51643	6.5654	29	34	58.4416	2.0374	346	21	2	6.0323	0.012419	307.6004	15.8957
10	ADH1_YEAST	Alcohol dehydrogenase 1 OS Saccharomyces cerevisiae GN ADH1 PE 1 SV 4	36799	6.2734	21	25	45.1149	1.6405	221	17	3	6.1851	0.010374	50	1.8412
11	FETUA_HUMAN	Alpha 2 HS glycoprotein OS Homo sapiens GN AHSG PE 1 SV 1	39299	5.3311	18	18	57.4932	6.0195	254	14	2	7.0138	0.01441	108.7959	4.2784
12	APOH_HUMAN	Beta 2 glycoprotein 1 OS Homo sapiens GN APOH PE 1 SV 3	38272	7.8679	16	22	55.0725	1.5557	206	16	1	6.0041	0.01384	83.7676	3.2082
13	IGHA1_HUMAN	Ig alpha 1 chain C region OS Homo sapiens GN IGHA1 PE 1 SV 2	37630	6.056	14	23	45.8924	1.9013	208	13	2	6.7941	0.012292	114.8981	4.3265
14	IGHG1_HUMAN	Ig gamma 1 chain C region OS Homo sapiens GN IGHG1 PE 1 SV 1	36083	8.1782	19	21	57.8788	1.6744	251	16	1	7.2405	0.013425	106.5199	3.846
15	IGHA2_HUMAN	Ig alpha 2 chain C region OS Homo sapiens GN IGHA2 PE 1 SV 3	36503	5.6603	14	22	58.8235	7.0424	190	13	2	7.6976	0.012056	37.9997	1.388
16	IGKC_HUMAN	Ig kappa chain C region OS Homo sapiens GN IGKC PE 1 SV 1	11601	5.5001	6	7	80.1887	1.9995	79	6	1	7.6142	0.012219	58.198	0.6756
17	A1BG_HUMAN	Alpha 1B glycoprotein OS Homo sapiens GN A1BG PE 1 SV 3	54238	5.4948	22	29	58.3838	2.598	243	19	2	7.1518	0.011618	59.2795	3.2173
18	HBB_HUMAN	Hemoglobin subunit beta OS Homo sapiens GN HBB PE 1 SV 2	15988	6.8795	11	13	80.9524	2.0664	97	11	0	7.9784	0.010878	12.3413	0.1974
19	FIBG_HUMAN	Fibrinogen gamma chain OS Homo sapiens GN FGG PE 1 SV 3	51478	5.2392	22	37	56.5121	1.6182	231	21	4	7.4057	0.011876	0	0
20	HPT_HUMAN	Haptoglobin OS Homo sapiens GN HP PE 1 SV 1	45176	6.1199	23	28	53.4483	1.4363	230	21	1	6.6182	0.012774	42.8189	1.9356
21	IGHG3_HUMAN	Ig gamma 3 chain C region OS Homo sapiens GN IGHG3 PE 1 SV 2	41260	7.7875	16	24	38.992	1.4958	181	13	1	7.3116	0.013136	15.2379	0.6291
22	IGHG4_HUMAN	Ig gamma 4 chain C region OS Homo sapiens GN IGHG4 PE 1 SV 1	35917	7.1098	16	20	46.1774	1.9713	192	13	2	7.3137	0.011902	9.0829	0.3264
23	A2MG_HUMAN	Alpha 2 macroglobulin OS Homo sapiens GN A2M PE 1 SV 2	163188	5.9678	71	95	56.1058	1.7071	687	66	11	7.6105	0.012345	68.2723	11.1483
24	KNG1_HUMAN	Isoform LMW of Kininogen 1 OS Homo sapiens GN KNG1	47852	6.2593	23	45	63.9344	1.981	188	22	3	6.2843	0.013874	27.6777	1.3253
25	FIBA_HUMAN	Fibrinogen alpha chain OS Homo sapiens GN FGA PE 1 SV 2	94314	5.6116	32	59	34.7575	5.2677	317	27	2	7.258	0.011889	50.4829	4.7945
26	CFAH_HUMAN	Complement factor H OS Homo sapiens GN CFH PE 1 SV 4	139004	6.1747	54	96	48.7409	1.6141	408	52	7	7.8411	0.013116	40.4721	5.6295
27	AMBP_HUMAN	Protein AMBP OS Homo sapiens GN AMBP PE 1 SV 1	38973	5.8731	11	26	39.2045	1.324	83	11	1	7.1984	0.013376	34.0776	1.329
28	APOC2_HUMAN	Apolipoprotein C II OS Homo sapiens GN APOC2 PE 1 SV 1	11276	4.4366	5	7	52.4753	2.4401	48	5	0	7.7107	0.013366	8.5553	0.0965
29	IGHG2_HUMAN	Ig gamma 2 chain C region OS Homo sapiens GN IGHG2 PE 1 SV 2	35877	7.4359	16	20	36.8098	2.7422	172	11	1	6.9442	0.011797	78.2617	2.8097
30	FIBB_HUMAN	Fibrinogen beta chain OS Homo sapiens GN FGB PE 1 SV 2	55892	8.2502	24	42	57.2301	2.0094	235	23	4	6.8555	0.013213	34.0269	1.9031
31	AACT_HUMAN	Alpha 1 antichymotrypsin OS Homo sapiens GN SERPINA3 PE 1 SV 2	47620	5.1808	20	31	45.8629	2.6588	158	17	0	7.5327	0.012645	41.8912	1.9962
32	ITIH4_HUMAN	Inter alpha trypsin inhibitor heavy chain H4 OS Homo sapiens GN ITIH4 PE 1 SV	103293	6.5259	36	64	52.9032	2.8497	302	36	4	8.1694	0.012195	41.0127	4.239
33	ANGT_HUMAN	Angiotensinogen OS Homo sapiens GN AGT PE 1 SV 1	53120	5.8471	8	25	24.3299	2.2974	78	8	0	6.9126	0.010032	20.1169	1.0693
34	HRG_HUMAN	Histidine rich glycoprotein OS Homo sapiens GN HRG PE 1 SV 1	59540	7.0957	14	33	25.3333	2.7158	134	13	1	6.5568	0.010646	43.7582	2.607
35	IGHM_HUMAN	Ig mu chain C region OS Homo sapiens GN IGHM PE 1 SV 3	49275	6.3312	22	38	43.8053	1.8061	182	19	2	7.8822	0.012938	30.5059	1.5041
36	ANT3_HUMAN	Antithrombin III OS Homo sapiens GN SERPINC1 PE 1 SV 1	52568	6.3036	27	37	60.7759	2.188	209	27	5	7.6008	0.011509	53.557	2.8172
37	APOA4_HUMAN	Apolipoprotein A IV OS Homo sapiens GN APOA4 PE 1 SV 3	45371	5.1129	23	42	54.798	4.7231	171	21	1	7.9069	0.012216	22.8524	1.0375
38	CFAH_HUMAN	Isoform FHL 1 of Complement factor H OS Homo sapiens GN CFH	51000	6.7189	25	38	58.7973	1.2736	170	24	3	7.2463	0.013011	0	0
39	KNG1_HUMAN	Kininogen 1 OS Homo sapiens GN KNG1 PE 1 SV 2	71912	6.3395	23	61	42.236	2.0005	178	22	4	6.4987	0.01421	6.4628	0.465
40	CFAB_HUMAN	Complement factor B OS Homo sapiens GN CFB PE 1 SV 2	85478	6.6555	24	64	40.5759	4.4495	187	23	0	7.0618	0.01093	30.3811	2.5986

Figure 2.9 Exported list of proteins identified by PLGS. It shows all the details about every single protein including the absolute amount of each protein.

Total amount of protein loaded on the column was calculated from the exported table which then used to adjust injection volume of the samples so that the same amount of protein can be injected for every sample.

Chapter Three

Method Development

3 Method development

3.1 Introduction

3.1.1 Development and optimisation of novel pipeline for biomarker discovery

The use of proteomics in biomarker discovery requires a reliable workflow which ensures that the quantitative measurement of a particular analyte demonstrates suitable reproducibility (Callesen *et al.*, 2008; Duncan, 2012). There are three main stages in a proteomics workflow (Reinders & Sickmann, 2009) including: sample preparation, sample analysis and data analysis (Figure 3.1).

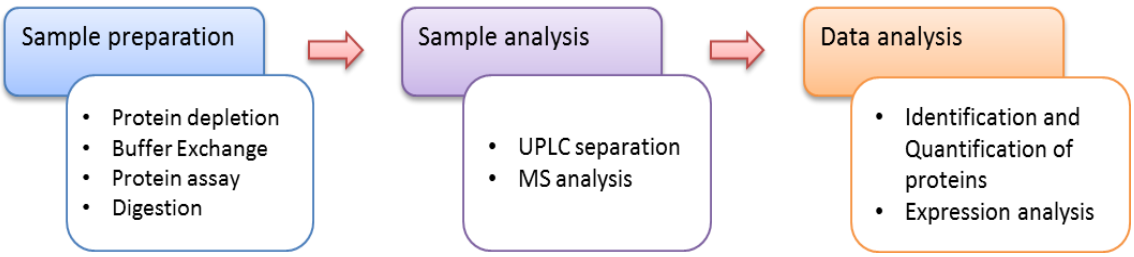

Figure 3.1 Proteomics workflow

Several different proteomics workflows for analysis of clinical samples are used by researchers in the field which mostly depends on the instrument availability. In the search for biomarkers of bladder cancer, SELDI was one of the first methods used for discrimination of bladder cancer from healthy controls (Liu *et al.*, 2005; Munro *et al.*, 2006); this was then superseded with MALDI (Schwamborn *et al.*, 2009; Oezdemir *et al.*, 2012) for increased sensitivity and specificity. One of the most popular workflows used for bladder cancer biomarker discovery is protein separation using 2D gel electrophoresis prior to MS analysis (Li *et al.*, 2011). However, the use of liquid chromatography instead of electrophoresis enables reproducible profiling of all proteins in a sample (Niu *et al.*, 2011; Linden *et al.*, 2012). Although the usage of nano-LC

systems (flow rate of 0.1 to 3 µl/min) coupled to mass spectrometry has become a common method used in many of the recent publications in proteomics biomarker discovery, to date, no publication has reported the use of nano-LC coupled to ion mobility incorporated mass spectrometry for this purpose. Using this orthogonal combination of analytical techniques, another level of separation is added to the sample analysis which can provide a greater confidence in the protein identification (Valentine *et al.*, 2011). The MS analysis method used in the proposed workflow is called $HDMS^E$ (also DIA) that has advantages over normal MS/MS or DDA methods (see Chapter 1, section 1.3.3.2). The other novelty about the proposed workflow is utilising label-free quantification method to assess absolute amount of proteins in a sample. The method, called Hi-3 Quantitation (Silva *et al.*, 2006), has advantages over other labelling methods such as iTRAQ (Ross *et al.*, 2004) and SILAC (Ong *et al.*, 2002) (see Chapter 1, section 1.3.3.2).

In the sample preparation section of the proposed workflow (Figure 3.2), two different immunodepletion kits (ProteoPrep 20 spin column for depletion of the top 20 high abundant proteins and SEPRRO IgY-14 LC-2 column for depletion of top 14 high abundant proteins) were evaluated (see Chapter 2, section 2.2.2). Also for the buffer exchange, two different protocols were investigated. In one experiment acetone is used for protein precipitation followed by replacing of the supernatant with a MS compatible buffer (Kay *et al.*, 2008). In the other method, the buffer is exchanged after multiple washes of the proteins using a MS compatible buffer on a filter (see Chapter 2, section 2.2.3).

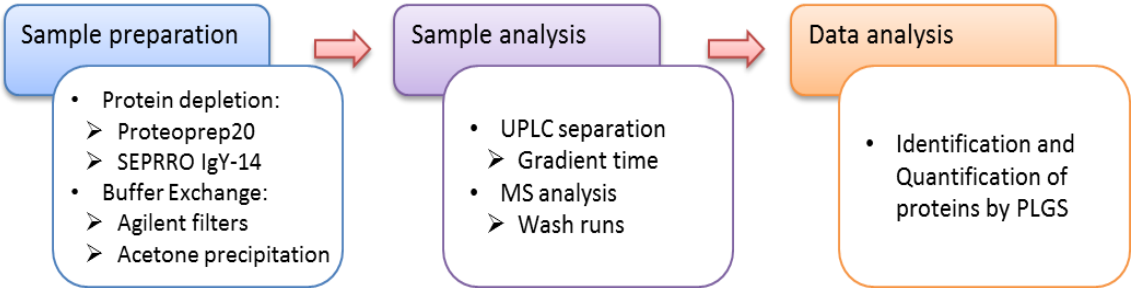

Figure 3.2 The proteomics workflow. Different procedures for every step of the workflow were investigated.

In the sample analysis, different variables were investigated such as the use of formic acid to wash the column between the runs and various gradient times. The best processing parameters for the PLGS to process the raw data files were also evaluated.

Using a plasma sample obtained from a healthy donor, most of these variables were evaluated. The ultimate goal of the evaluation was to establish a workflow for the biomarker discovery phase of the study with optimised and evaluated assays in each stage of the procedure.

3.1.2 Systematic assessment of reproducibility of the optimised workflows

Proteomics analyses of samples, particularly with mass spectrometry as the choice detection system, necessitate considerable sample work-up which ultimately increases the chance of introducing variability. It was the aim of this experiment to see whether the proposed workflow is technically reproducible. To do so, a single plasma sample was obtained from a healthy volunteer and divided into three equal aliquots (to have technical replicate, as discussed by Duncan, to be used in the experiment (Duncan, 2012). Aliquots of the sample were injected in triplicate on day 1, day 2 and day 3 (Figure 3.3).

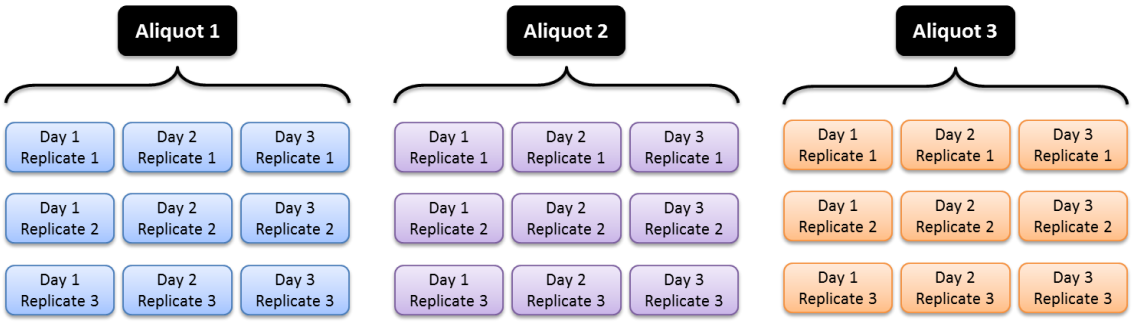

Figure 3.3 Study design for the assessment of reproducibility. Three aliquots of a plasma sample were injected in triplicates over three days.

3.1.3 Pilot study: assessment of clinical samples

Every biomarker discovery study has four different stages (discovery phase, quantification, verification and validation (Pavlou *et al.*, 2013), as described in Chapter 1. Due to the limited amount of material available (clinical samples) for biomarker discovery (Luque-Garcia & Neubert, 2007), it is best to use a previously optimised and evaluated protocol to minimise the risk of sample loss. In the proposed workflow, the sample preparation and sample analysis sections of the protocol were evaluated and optimised using plasma samples from healthy volunteers. However, plasma from patients who have disease may exhibit molecular differences that profoundly compromise the analysis (Anderson, 2010). Furthermore, there is a paucity of data regarding the inherent effect of disease on the molecular constitution of plasma. Consequently, a further evaluation step was carried out before performing the discovery phase. Thus, 18 plasma samples from patients with bladder cancer were randomly selected. Also, a further 3 plasma samples from control group (age/sex matched) were selected to enable evaluation of expression analysis.

So the aim of this chapter was to develop and evaluate a proteomics workflow using nanoACQUITY UPLC coupled to Synapt G2 HDMS to identify and quantify proteins in plasma samples using label-free methodology; also to check the workflow's reproducibility using healthy control samples. Incorporation of plasma samples from bladder cancer patients into the workflow enables a check of reliability and ultimately the ability to carry out expression analysis. The workflow will then be used for the discovery phase of the project.

3.2 Materials and methods

3.2.1 Materials and auxiliary equipment

All buffers, reagents and columns were purchased from either Sigma-Aldrich (Poole, UK) at HPLC grade or from Fisher Scientific (Loughborough, UK) at Optima LC-MS grade. A Varian ProStar HPLC system (AutoSampler 410, UV detector 325 and the pump 230) was used for the immunodepletion step with the SEPPRO IgY-14 LC column. Standards (Enolase MassPrep digest standard and Alcohol dehydrogenase MassPrep digest standard) and analytical columns were bought from Waters Corp. (Herts, UK).

3.2.2 Protocol for systematic assessment of reproducibility of the workflow

The human blood sample was collected in EDTA blood sample tubes from healthy donor. The blood samples were centrifuged at 1500 x g for 20 minutes at 4°C. The plasma layer was separated from the buffy layer and red blood cells, and stored at -80°C.

Procedure

The top 20 high abundant proteins in plasma were depleted using ProteoPrep 20 immunodepletion spin columns followed by buffer exchange using acetone precipitation method (see Chapter 2 section 2.2.3). Protein concentration was assessed using Pierce BCA protein assay followed by digestion and concentration using vacuum evaporator/freeze drier. Dried pellets were re-constituted in 10 µl of 0.1% formic acid + 10 µl of ADH (50 fmol/µl) prior to injection and separation of the peptides using a nanoACQUITY UPLC and analysis on a Waters Synapt G2 HDMS. LC and MS conditions are described in Table 3.1 for sample analysis.

LC Condition		MS Condition	
Trap column: 5µm C_{18} in 180µm, 20mm long		Ionisation mode: ESI positive	Capillary voltage: 3.0 kV
Analytical column: 1.7µm BEH130 C_{18} in 75µm, 250mm long		Acquisition range: m/z 50-2000	Cone voltage: 30 V
Mobile Phase A: 0.1% FA in LCMS water, Mobile Phase B: 0.1% FA in ACN		Source temp: 70°C	LockMass flow rate: 0.5µL/min
Column Temp.: 40°C	Flow rate: 0.3µl/min	Collision energy: 4eV	Elevated energy: 15-40eV
Gradient: 0-85min, 1-40% B # 85-87min, 40-85% B # 87-100 min, 85% B # 100-110 min, 3% B			

Table 3.1 LC and MS conditions used for sample analysis

Instrument preparation steps were performed (see Chapter 2, section 2.2.7.1) before analysis of the samples. Sample 1 was analysed in triplicate in MS^E mode (without the use of ion mobility) on the day 1 of the experiment. The procedure was repeated for sample 2 on the 2nd day and sample 3 on the 3rd day of the experiment. Raw data files were processed using PLGS 2.4 (see Chapter 2, section 2.2.8). To further minimise false positive identifications, protein identification was only accepted if they were found in all three injections of the triplicate.

3.2.3 Protocol for the pilot study to assess the pipeline using clinical samples

The first batch of plasma samples received from collaborators in the University of Torino (Turin, Italy) consisted of 60 plasma samples including 40 bladder cancer samples and 20 non-cancerous hospitalised samples to be used as controls (age and sex matched). An aliquot of 50µl from each sample was transferred into a 1.5ml LoBind Eppendorf tube and the remaining volume was stored at -80°C for future analysis. High abundance protein depletion using SEPPRO IgY-14 immunoaffinity column, concentration, buffer exchange and digestion (see Chapter 2, section 2.2) were performed on every sample individually (Figure 3.4).

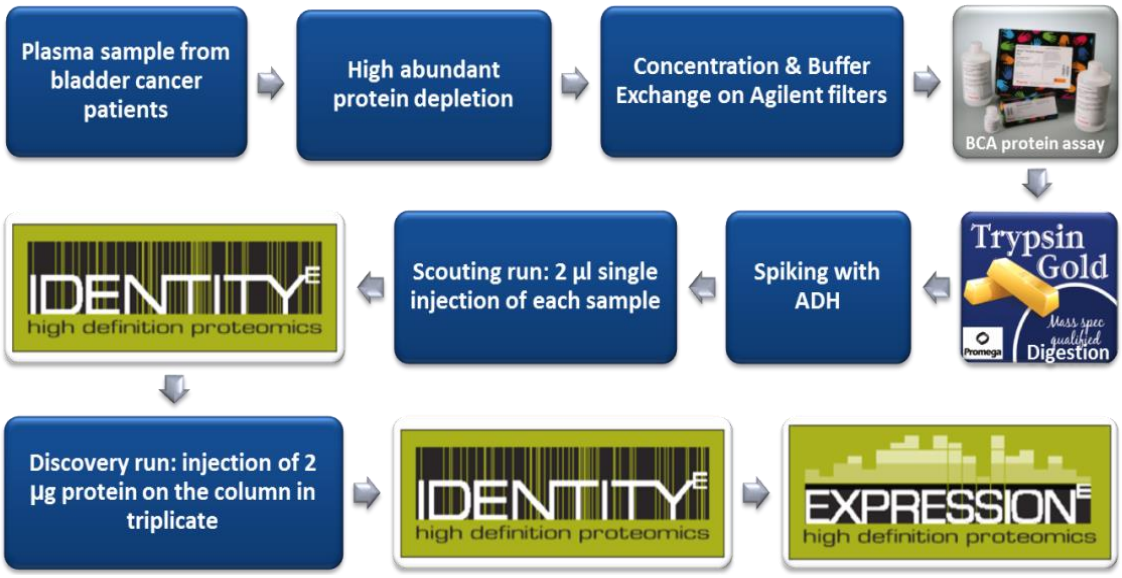

Figure 3.4 Workflow for mass spectrometric analysis of bladder cancer samples

Before spiking the samples with internal standard, 18 patient samples (in 3 different tumour grades) and 3 control samples were randomly selected. Every three bladder cancer samples with the same grade and stage were grouped and pooled (15µl each) into a new 1.5ml LoBind Eppendorf tube. Three non-cancerous hospitalised control samples were also pooled and used as a control. The pooling procedure resulted in six disease samples and one control sample; details of which are shown in Table 3.2.

Pooled Sample ID	P1G1Ta			P4G2Ta			P6G2T1			P7G3T1			P10G3T2			P13G3T3			P14CNT1		
Sample number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Patient Sample ID	L1-4	L1-7	L1-11	L1-5	L1-16	L1-10	L1-6	L1-2	L1-19	L1-3	L1-18	L1-20	H1-31	H1-23	H1-34	H1-29	H1-26	H1-39	CNT1-44	CNT1-48	CNT1-53
Stage (T)	A			a			1			1			2			4	3b		**********		
Grade (G)	1			2			2			3			3			3					

Table 3.2 Pooling details of selected samples. Every 3 samples with the same grade and stage (21 sample were used in total) were pooled. It resulted in 7 pooled samples to be used to check the method.

All 7 pooled samples were spiked with ADH (50fmol/µl) as an internal standard (50:50). Pre-analysis instrument checks (see Chapter 2, section 2.2.7.1) were carried out. For an initial scouting run, 2µl of every pooled sample was injected to assess the total protein

content of each sample. Raw data files obtained by MassLynx 4.1 were processed using PLGS 2.4 for protein identification and quantification. Total amount of protein loaded on the column was calculated for each sample by summing up the amount of individual proteins. Based on the total protein amount obtained from the initial scouting run, the injection volumes for the discovery phase were adjusted to load 2µg of protein on the column for every sample (manufacturer's recommendation). Every pooled sample was injected in triplicate and the chromatographically separated peptides were introduced to the mass spectrometer through a nano-spray ESI using a 110 minutes gradient. The column was washed and re-conditioned by 3 injections of a 20 minute gradient (1 x 1% FA and 2 x 0.1% FA) between each triplicate injection. Raw data files obtained by MassLynx were processed with PLGS. Identification and quantification of the proteins were performed with Identity^{E} and alteration in the expression level of the proteins in each group was assessed with Expression^{E}.

3.3 Results

3.3.1 Development and optimisation of the pipeline

3.3.1.1 Sample preparation

3.3.1.1.1 Investigating the effects of using LoBind tubes

LoBind eppendorf tubes have been shown to have high protein recovery (Turcotte & Raines, 2008; Grou *et al.*, 2008). Those tubes were used throughout the sample preparation in parallel to the use of normal tubes to check the protein recovery of the LoBind tubes in mass spectrometry analysis. This experiment was performed using two aliquots of the same plasma sample.

The raw data files of the samples prepared in the LoBind tubes and the ones prepared in the normal tubes were processed using PLGS. The amount of identified proteins in the sample prepared in normal tubes was compared to the one prepared in LoBind tubes and plotted in Figure 3.5.

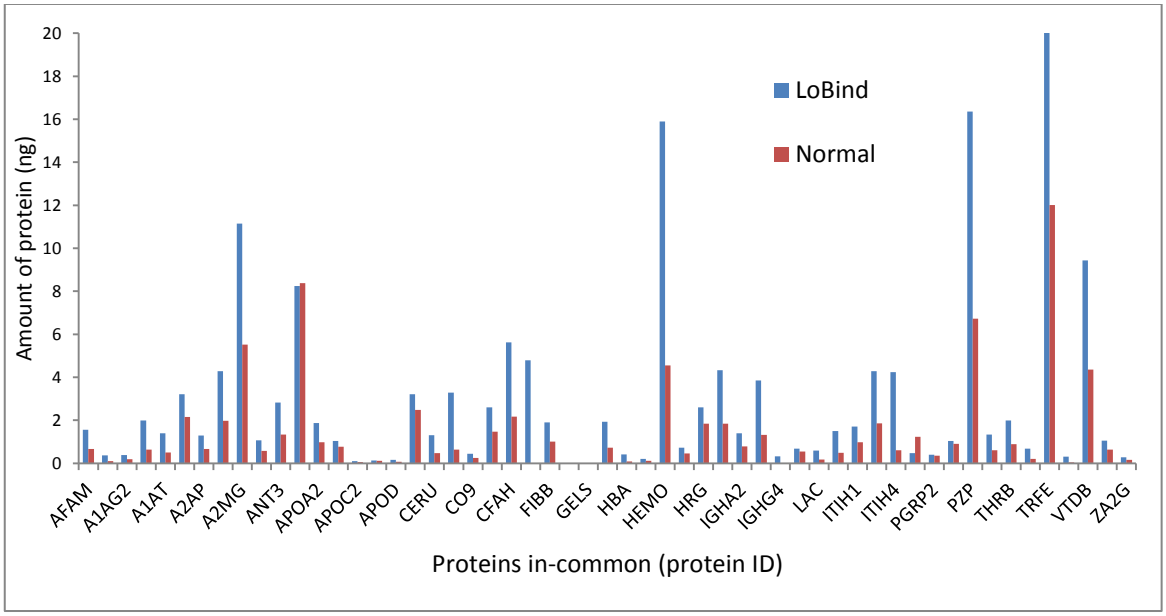

Figure 3.5 Comparison of the amounts of identified proteins in normal and LoBind tubes

The number of peptides was also calculated and plotted (Figure 3.6). The plot clearly shows higher number of peptides per protein for the representative proteins.

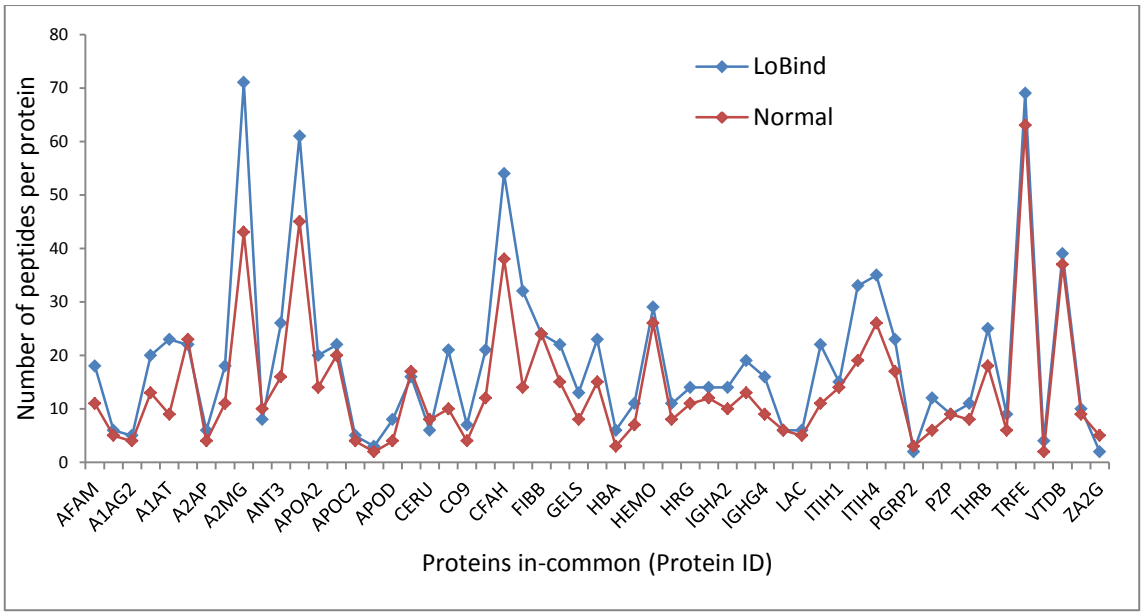

Figure 3.6 Comparison of the number of peptides per proteins between normal and LoBind tubes

On average 68 proteins were identified using the LoBind tubes and 60 proteins using normal tubes. The amounts of the identified proteins were different (Figure 3.7).

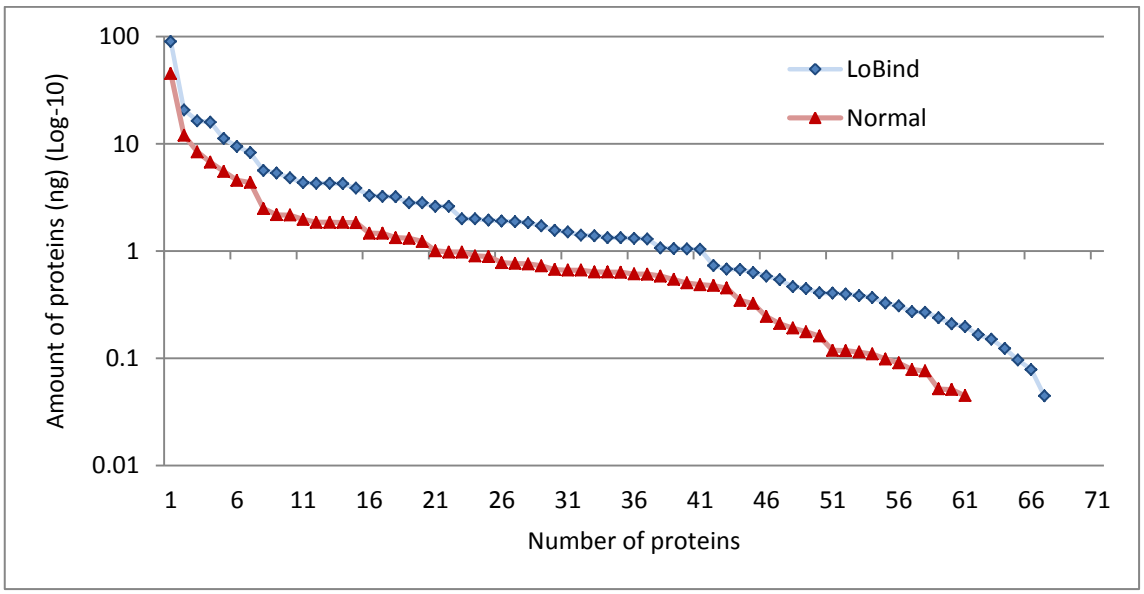

Figure 3.7 Absolute amount of every protein identified in the samples prepared in different tubes

Figure 3.7 shows that the amount of proteins identified in the sample prepared in the LoBind eppendorf tubes are higher compared to the one prepared in normal tubes. Since the amount of proteins is the key aspect of the analysis in the proteomics studies, we decided to use protein LoBind eppendorf tubes throughout the whole experiment in the discovery phase of this project.

3.3.1.2 Sample analysis

3.3.1.2.1 Wash and carry over

A wash run which was performed after a sample run resulted in the identification of more than 30 proteins. These are the proteins bound to the column from a previously run sample. Due to the complexity of plasma not all of the proteins eluted off the column even at the isocratic wash phase of the analytical run. These proteins can easily be carried over and elute with the next sample's proteins leading to false identification and quantification. To overcome this issue, two different methods to wash the column were tried: (*i*) addition of wash cycles to the end of the gradient and (*ii*) performing separate wash runs after each sample.

In the first instance, two wash cycles (10min each) using 0.1% formic acid was added to the end of the 110min normal gradient used for the sample analysis (Figure 3.8). The total run time was increased to 130min.

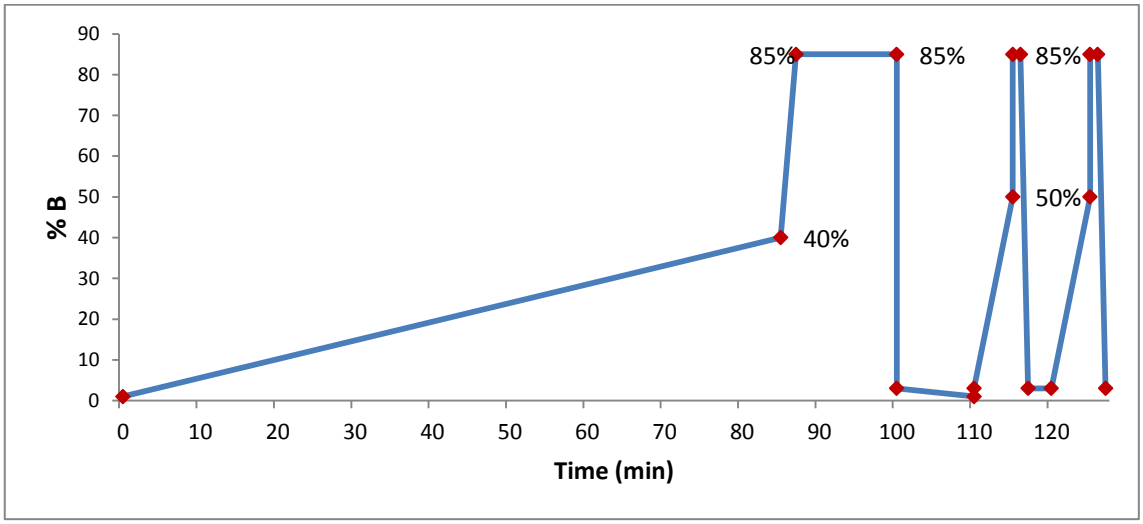

Figure 3.8 Composition of mobile phase B in the 110min gradient plus two wash cycles of 10min duration

In another experiment, the effects of carrying out three short sequential ramps of mobile phase B (3min ramps) were investigated. The percentage of mobile phase B

was increased up to 95%. The three additional ramps add an extra 9 minutes to the end of the normal 110min gradient (Figure 3.9).

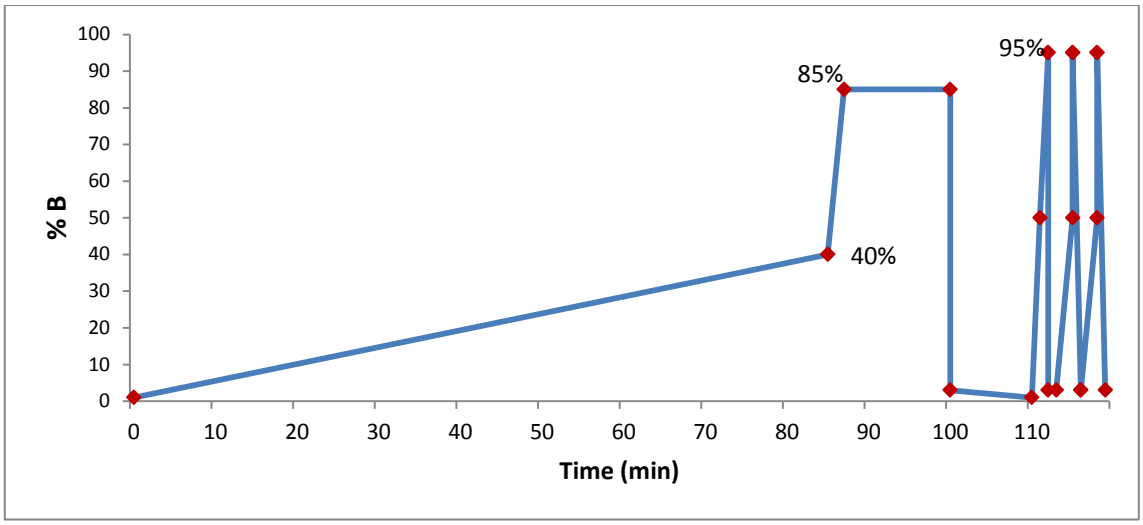

Figure 3.9 Composition of mobile phase B in the 110min gradient plus three wash cycles of 3min with higher percentage of B

An alternative wash method involved running a separate injection run after each sample injection. A 20min gradient was used to run wash cycles. In total, the column was washed with four wash runs injecting 1% formic acid for the first run and 0.1% formic acid for the other subsequent injections.

The raw data files of the wash runs carried out after the sample analysis were processed using PLGS. PLGS identified a significant number of proteins in the first wash post-sample analysis illustrating significant carry over. In the first wash run after a 120min gradient, 19 proteins were seen and in the other three washes 10, 7 and 2 proteins were identified, respectively (Figure 3.10). When the first wash run after the sample was analysed with the 130min gradient processed 21 proteins were seen. The number of identified proteins in the subsequent three washes decreased to 12, 6 and 2 proteins, respectively.

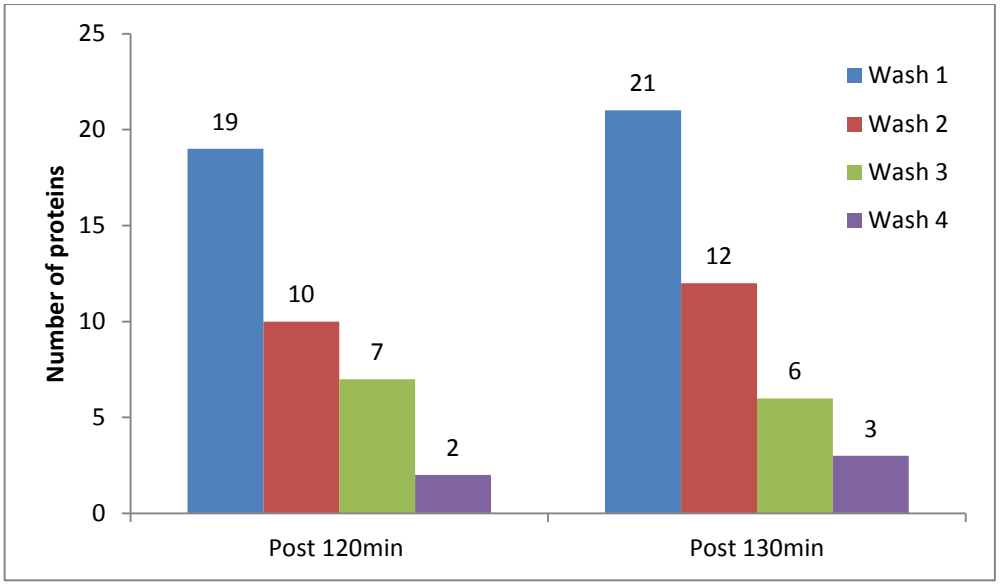

Figure 3.10 The number of proteins identified from the 4 wash runs after the sample analysed by 120 and 130 minutes gradients

Where wash runs were ran separately after the 110min gradient, the first wash contained 27 proteins, the second wash 11, the third 3 and the fourth wash 3 proteins (Figure 3.11).

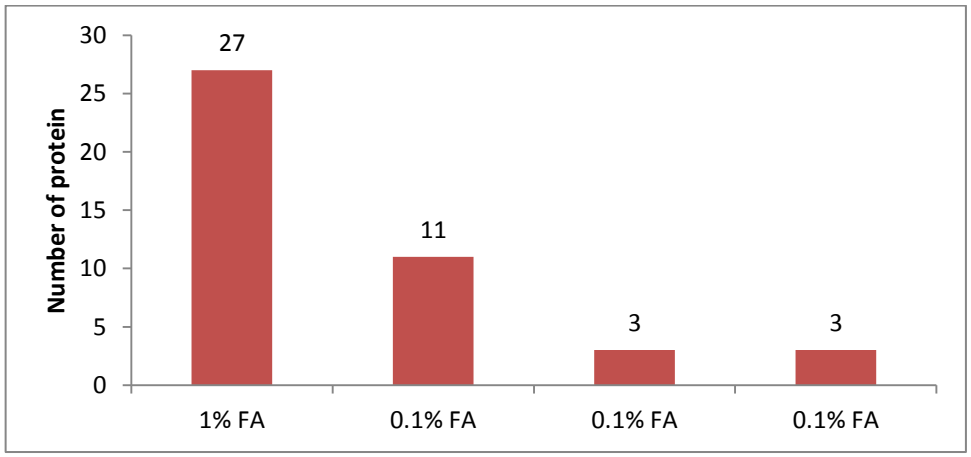

Figure 3.11 The number of proteins identified in 4 wash runs performed after the sample analysed by 110min gradient

A significant decrease in the number of proteins was seen when the wash run was repeated four times. Identification of 27 proteins in the first wash after the sample run showed the need to wash the column prior to injection of another sample. The

number of proteins decreased to less than half when the second wash performed and the third wash showed only 3 proteins. Equal number of proteins in the last two washes suggested that there is no need to do the fourth wash and the column can be considered clean and ready for the next sample injection. Although the number of proteins in the first wash after the 110min gradient was higher than the washes post 120min and 130min (27 compared to 19 and 21) the decrease in the number of proteins in the subsequent washes was greater. The column became clean after the third wash whereas in the other two methods four washes were needed to clean the column. Also performing 3 washes after the 110min gradient adds an hour to the total analysis time per sample (170 minutes) while 4 washes for the other methods (80 minutes) plus the longer gradient of the sample analysis (120min or 130min) increase the total analysis time per sample between 20 to 30 minutes. This extra time is significant when considering the number of samples to be. Consequently, 3 wash runs were incorporated into the standard workflow for any clinical study and were performed after every sample analysed with a 110 minutes gradient time (1% formic acid for the first and 0.1% formic acid for the other two).

3.3.1.2.2 Different run time (inlet method)

Different gradient times were checked to choose the best for the workflow. A compromise between run time and maximal number of proteins identified was assessed. A variety of gradient times from 50 minutes to 240 minutes total run time were investigated (Table 3.3).

% B	Gradient Time (min)						
	50min	60min	70min	80min	90min	110min	240min
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40	40	50	60	70	80	85	210
80	45	55	65	75	85	87	215
85	47	57	67	77	87	100	218
3	47.50	57.50	67.50	77.50	87.50	100.50	225
1	50	60	70	80	90	110	240

Table 3.3 Different gradient times. Evaluation of gradient times between 50min to 240 min.

The gradient shape was the same for all runs, just the time between changes in the gradient were altered. The majority of peptides elute between 0 and 40% of mobile phase B (Figure 3.12).

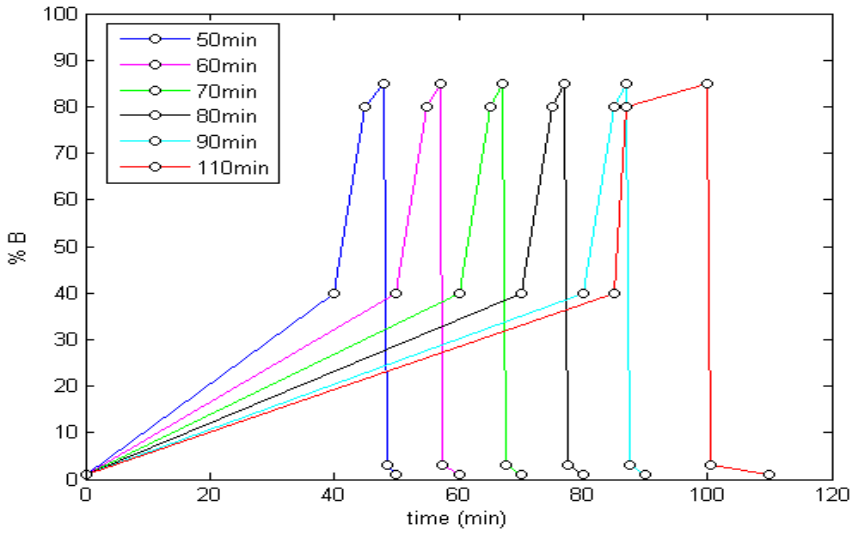

Figure 3.12 Composition of mobile phase B for the different run times

Next, a single sample was used to investigate seven different gradient times from 50min to 240min. The raw data files obtained by MassLynx were processed using PLGS and between 101 and 129 proteins were identified (Figure 3.13).

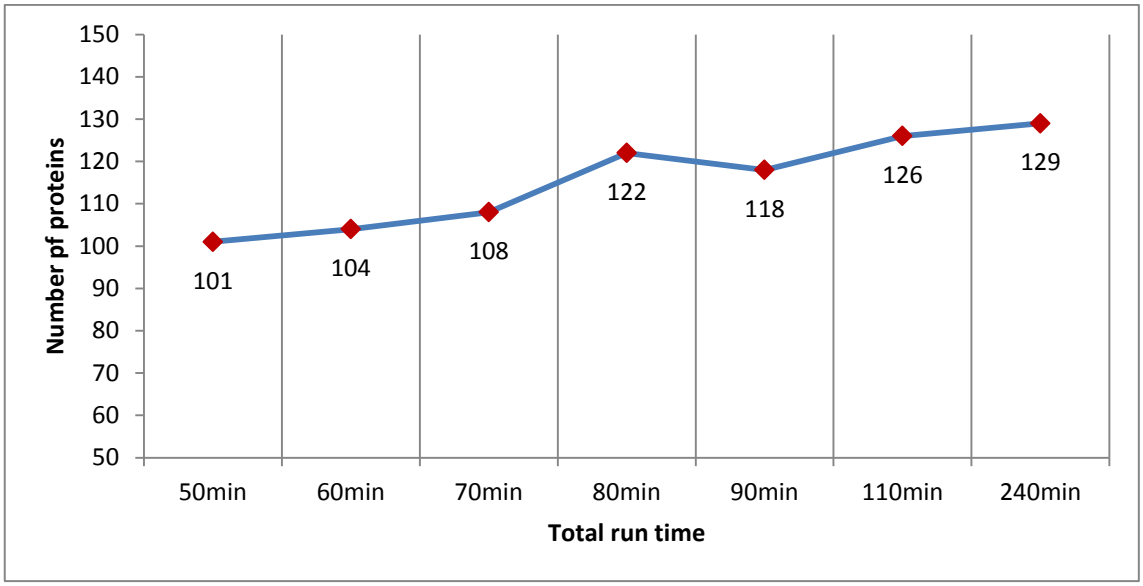

Figure 3.13 Total number of proteins identified in the different gradient times

Figure 3.13 shows that by increasing the gradient time from 50 to 240min the number of proteins identified increased by approximately 30%. Since a 240min gradient makes the total run time more than double and results in the identification of only 3 more proteins compared to 110min, we decided to use the 110min gradient for all subsequent mass spectrometric analysis in this project.

3.3.1.2.3 Curve evaluation

One of the key parameters in the gradient is the curve which is related to the rate of change of the solvent. There are 11 different curves predefined in the nanoACQUITY UPLC (Figure 3.14).

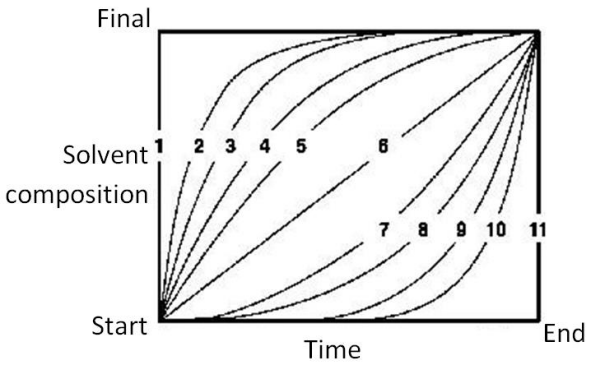

Figure 3.14 Different predefined curves in the nanoACQUITY UPLC. The change in the proportion of the solvents follows opposite pattern in the concave (2 to 5) and convex (7 to 10) curves. Curve 6 is a linear pattern.

Three different curves (number 6, 8 and 10) were chosen to investigate which one promotes the best peptide separation. The linear one (curve number 6), curve 8 and number 10 were investigated. A single plasma sample depleted from high abundant proteins was used for this experiment.

The total number of proteins identified by PLGS for the samples analysed by all three gradients with different curves was calculated and plotted (Figure 3.15).

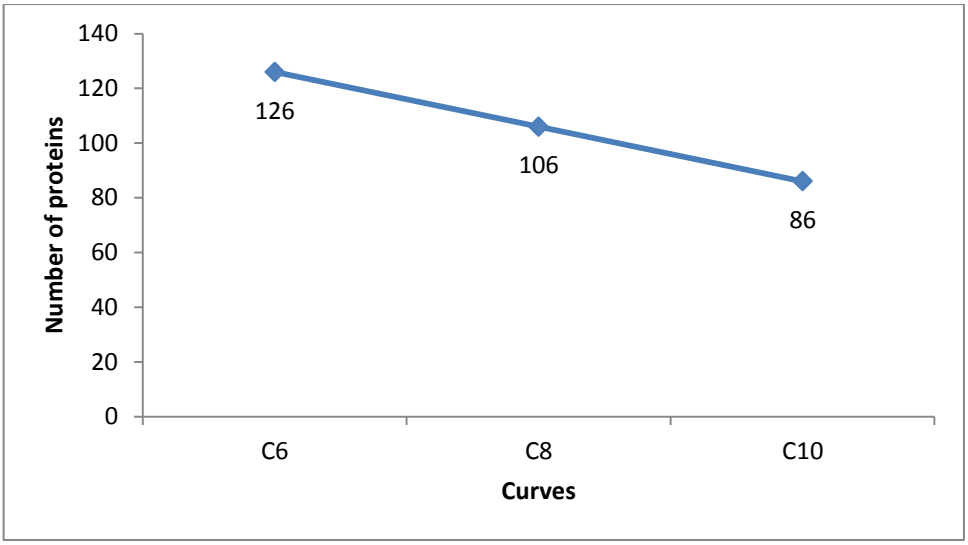

Figure 3.15 Comparison of the LC-gradients with different curves

In total, 126 proteins were identified by the first gradient where the curve was set to 6 meaning the proportion of the solvents changed in a linear fashion. In the samples analysed with the convex curves (8 and 10), the number of proteins decreased to 106 and 86, respectively. Although the change in the proportion of the solvents in the curves 8 and 10 is slower than the linear one, the number of identified proteins was still less. Since the number of proteins is a crucial aspect of the project we decided to use the linear one (curve 6) for the discovery phase of the project.

3.3.1.3 Data analysis

3.3.1.3.1 Establishment of optimum threshold

The parameters to be used for data processing in PLGS, low, high and intensity thresholds should be specified to exclude low intensity ions and minimise noise. Using these thresholds should mean that only the ions with intensities that exceed the threshold are reported. To be able to find the optimum threshold to use in the data preparation section of the PLGS, different mixtures of thresholds were tried for low energy, high energy and intensity threshold. Using a single sample analysed by MS, many different combinations of thresholds were used to find the one resulting in the highest number of proteins. Table 3.4 shows 9 representative searches.

	1	2	3	4	5	6	7	8	9
Low Energy Threshold	75	100	100	150	150	150	150	250	350
Elevated Energy Threshold	10	10	20	10	30	20	20	20	20
Intensity Threshold	500	500	500	750	750	500	1000	750	750

Table 3.4 Different combinations of the threshold in nine representative searches

After processing a sample with a variety of thresholds, nine of them were chosen as representative showing the effect of different thresholds on the number of identified proteins. Between 141 and 162 proteins were identified by PLGS using nine threshold combinations (Figure 3.16). The combination of 350 for the low energy threshold, 20 for the elevated energy threshold and 750 for the intensity threshold resulted in the highest number of proteins.

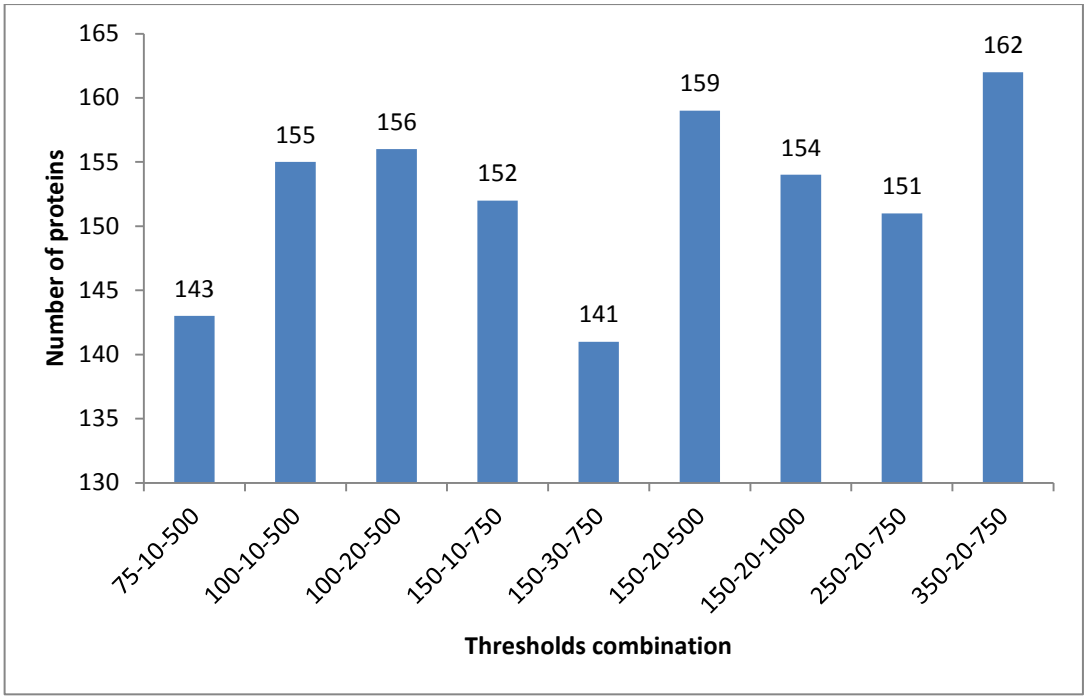

Figure 3.16 Total number of proteins identified when processing a sample with different combinations of thresholds

3.3.2 Systematic assessment of reproducibility in plasma proteomics workflow

A comparison of the number of confident protein assignments was carried out using label-free MS^E method. The names of identified proteins in replicate injections of each sample were compared. The proteins identified in at least two replicates of the 3 were selected and merged to form a new list for each day's experiment. Then, protein lists of all 3 days of each sample were merged and duplicates were removed. In total, 110, 103 and 103 proteins were identified in samples 1, 2 and 3 respectively (Table 3.5).

	Days 1	Day 2	Day 3	Merged lists
Sample 1	104	98	93	110
Sample 2	95	91	90	103
Sample 3	93	92	89	103

Table 3.5 Total number of proteins identified in each sample over 3 days

Total number of identified proteins in each replicate was averaged and plotted (Figure 3.17).

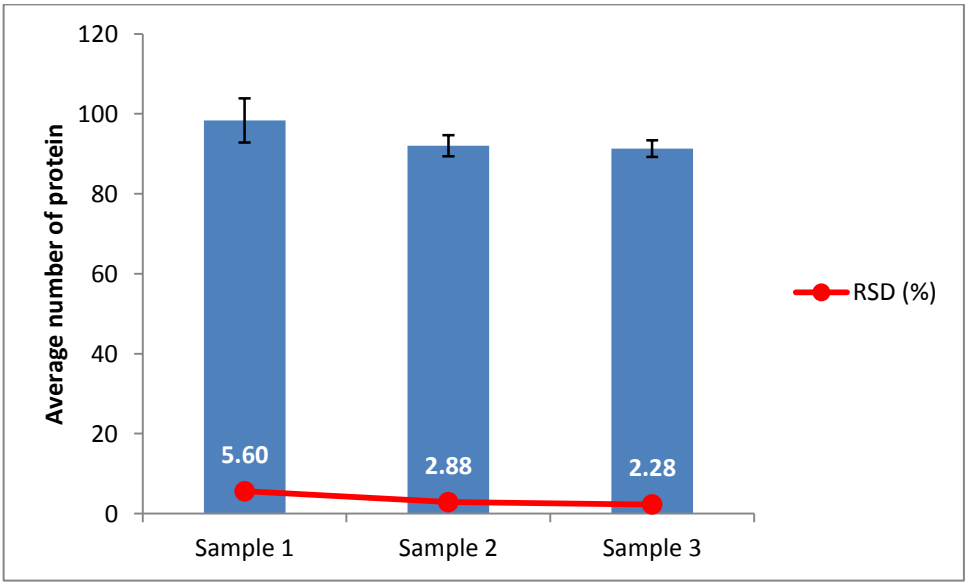

Figure 3.17 Average number of identified protein. Low levels of RSD value shows high degree of precision.

On average, 98, 92 and 91 proteins were identified in sample 1, 2 and 3 respectively over 3 days. The RSD value of below 6% for all proteins demonstrates the reproducibility of the experiment and in particular the high degree of precision. To analyse the data further, 18 representative proteins (Table 3.6) were randomly selected from different plasma concentration ranges (low, mid and high).

Accession	Description
P02765	Alpha 2 HS glycoprotein
P02647	Apolipoprotein A I
P02652	Apolipoprotein A II
P06727	Apolipoprotein A IV
P02649	Apolipoprotein E
P04003	C4b binding protein alpha chain
P10909	Clusterin
P00742	Coagulation factor X
P09871	Complement C1s subcomponent
P08603	Complement factor H
P02671	Fibrinogen alpha chain
P02790	Hemopexin
Q14520	Hyaluronan binding protein 2
Q96PD5	N acetylmuramoyl L alanine amidase
P02760	Protein AMBP
P00734	Prothrombin
P02768	Serum albumin
P04004	Vitronectin

Table 3.6 18 representative proteins randomly selected for reproducibility assessment

The average amount of these 18 proteins in 3 samples over 3 days was calculated. The RSD value calculated for each of the 18 proteins fell within 20% (Figure 3.18).

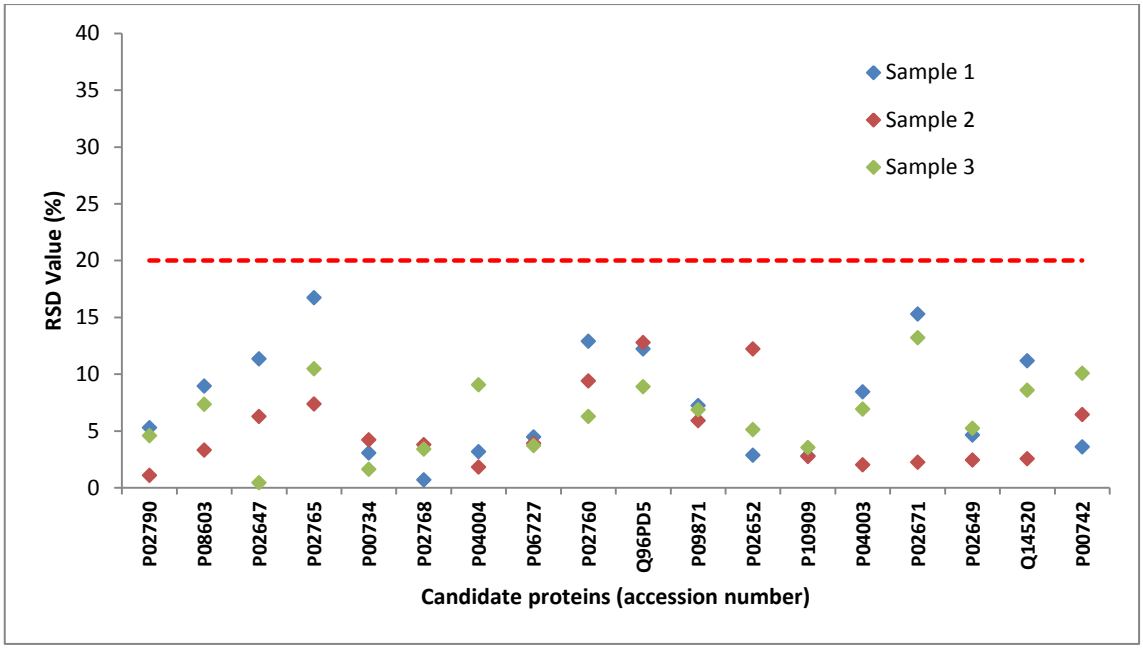

Figure 3.18 RSD value for the averaged amount of 18 representative proteins within three samples

In an attempt to delineate the technical variability we assessed the variability on a single day and compared the analysis of a three identically prepared samples. Individual inter-day variability for three samples is shown in Figure 3.19 for 18 randomly chosen proteins. The data shows that good levels of precision for both instrument platform and sample processing are obtained as shown in Figure 3.19 and Figure 3.20. Figure 3.19, in particular shows the amount of protein reported from each sample; the error bars indicate the low variance.

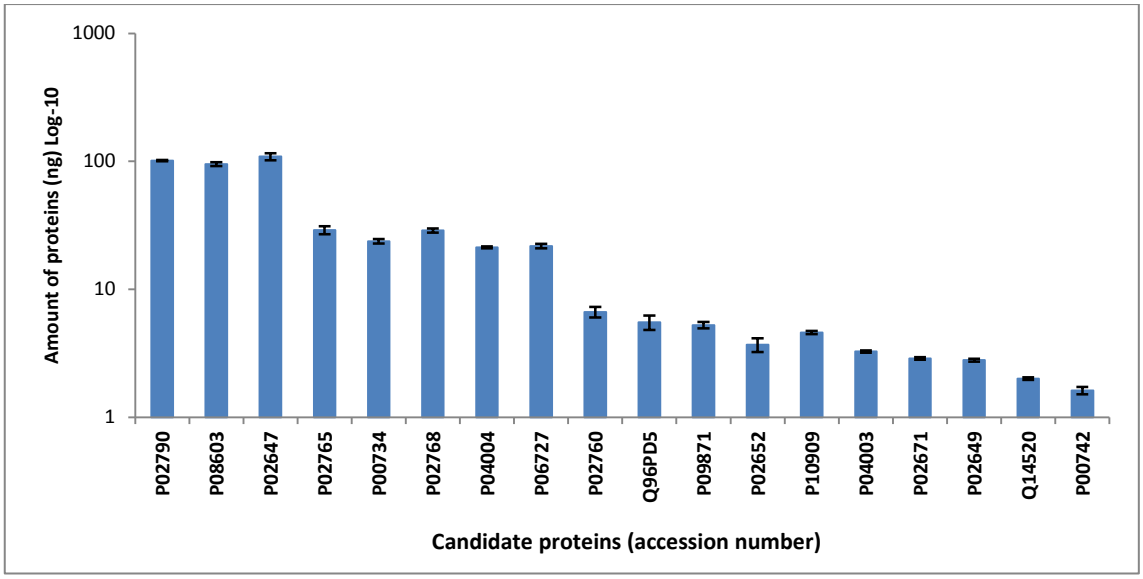

Figure 3.19 Inter-day variation of protein amount for 18 representative proteins

This data is further explored in a graph of 18 proteins and their respective RSDs. Here it can be clearly seen that the recorded RSDs for technical variability lay under 10% (Figure 3.20) in more than 90% of the samples.

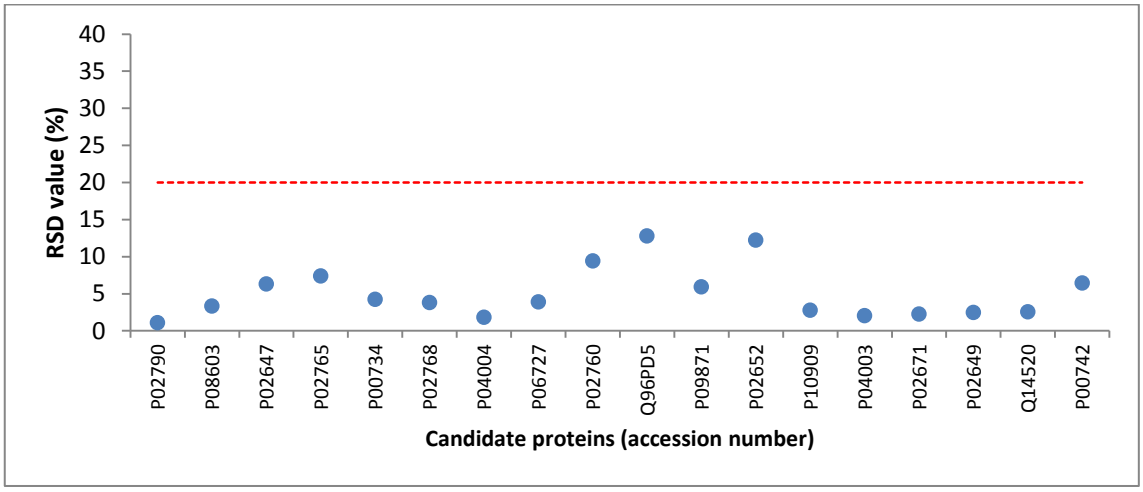

Figure 3.20 RSD value for the amount of 18 proteins in a sample within three days (inter-day variation). The red line represents acceptable value for relative standard deviation.

3.3.3 Pilot study to assess the pipeline using clinical samples

3.3.3.1 Protein Identification and Quantitation

Total number of proteins in each replicate of every sample (triplicate run) was averaged and plotted (Figure 3.21). Between, 93 to 146 proteins (mean average of triplicate) were identified in the samples and control and the standard deviation calculated. Error bars for the averages shows high levels of reproducibility as measured by the number of proteins identified in each run.

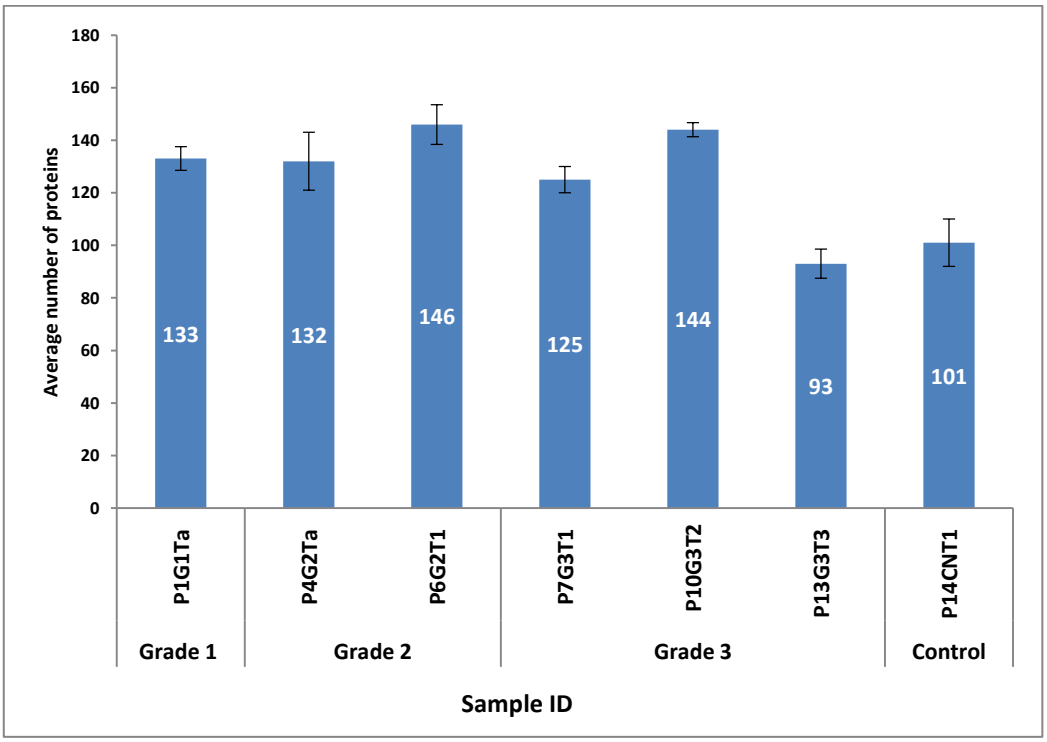

Figure 3.21 Average number of proteins identified in different grades and the control sample

The dynamic range of proteins quantified in a single sample was between 4 to 5 orders of magnitude showing its extension over 5 orders of magnitude (Figure 3.22).

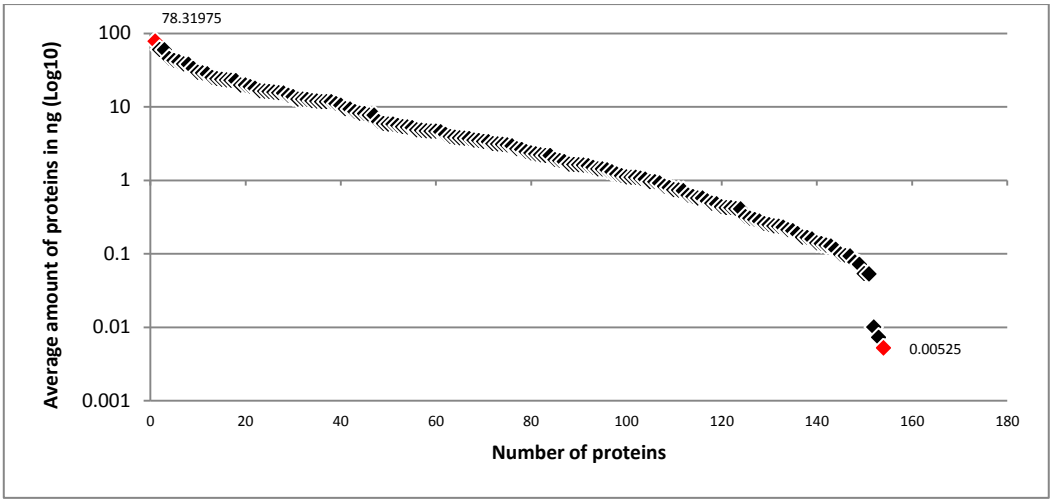

Figure 3.22 Dynamic range of the identified proteins in a sample obtained from a patient with a bladder malignancy

All 18 lists of proteins (triplicates of 6 runs) identified in patient samples were merged and duplicate proteins were removed. The same procedure was done for proteins identified in control samples. The merging of the protein lists resulted in identification of 239 proteins in patient group and 121 proteins in control group (Figure 3.23).

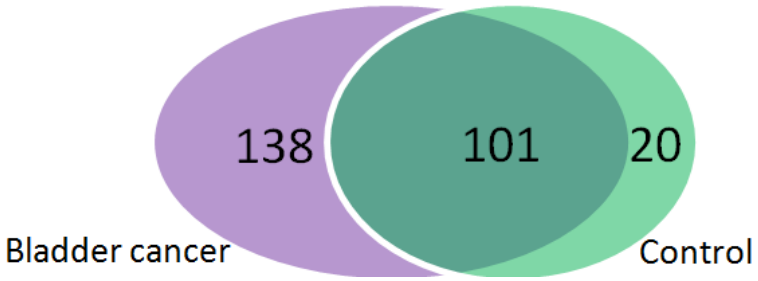

Figure 3.23 Venn diagram showing the number of proteins identified (merged lists) in bladder cancer and control

3.3.3.2 Expression Analysis

To validate the developed pipeline, samples were assessed via grade/stage progression. This type of comparison is not utilised clinically, but could demonstrate biological or pathological differences. By comparison of the quantified proteins in each group using Expression^{E} analysis, considerable changes in the protein profile of the samples with the tumour grade were observed. Some of the up/down regulated proteins can be picked up for further analysis as potential biomarkers. Based on the disease conditions used for classification and pooling of the samples, following expression analysis were carried out:

a) Disease vs. Control

By comparison of the pooled patient samples and control using Expression^{E}, alteration in the expression levels of 111 proteins was observed. 108 proteins were up regulated and 3 proteins were down regulated. Also 82 proteins were identified only in disease samples and only one protein was unique in control sample (uniqueness is defined as present in only one group of samples and absent in the other(s)).

b) Grade comparison

Differences in the level of proteins between different grades are shown in Figure 3.24 and Figure 3.25 representing down regulation in 7 proteins and up regulation in 11 proteins, respectively.

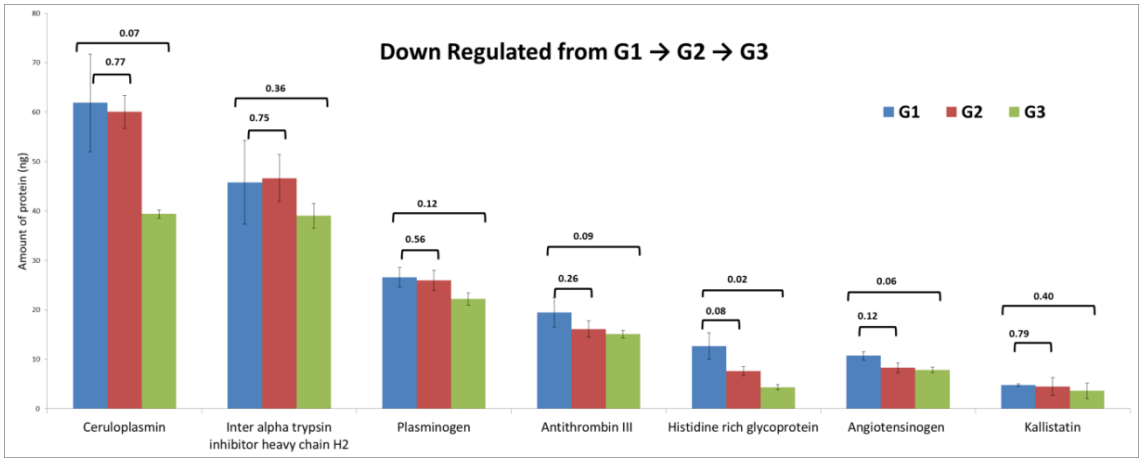

Figure 3.24 Down-regulated proteins and their correlation with the progression of the disease. The difference between the levels of expression is not statistically significant when comparing grades 1 and 2 but is more significant between grades 1 and 3.

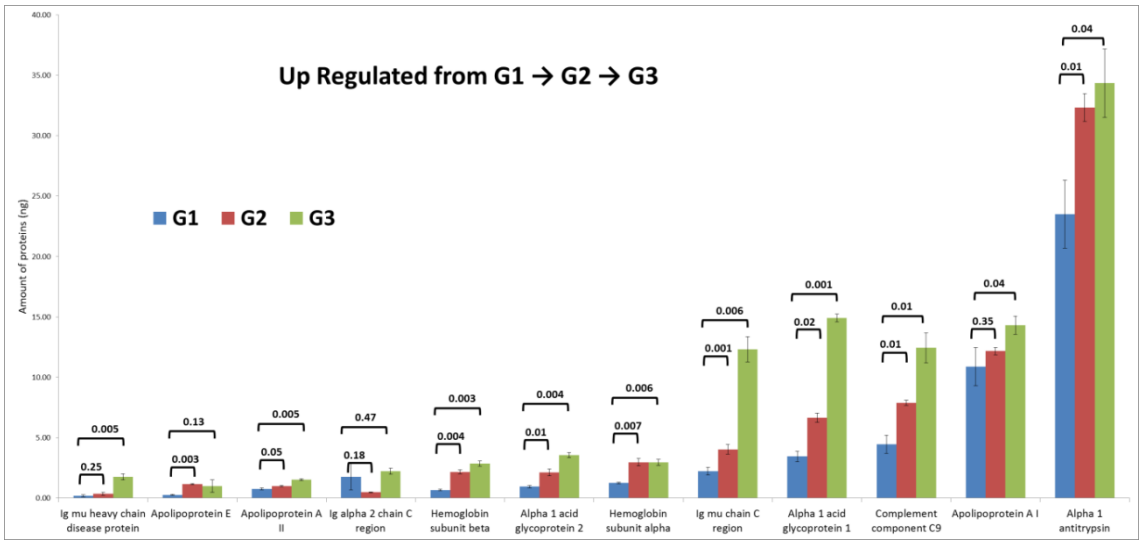

Figure 3.25 Up-regulated proteins and their correlation with the progression of the disease. Statistically significant difference was observed among the 3 grades.

c) Stage comparison

Due to the lack of samples in the different stages of all grades, expression analysis for the stage progression was performed only on the grade 3 tumours. Progression of tumour stage from T1 to T2 to T3 was investigated in this analysis. An increase in the expression level of 95 proteins and decrease in the expression level of only 1 protein was observed when tumours with stage T1 compared to stage T2 tumours. For

progression of the tumour from T2 to T3, up regulation was seen from stage 2 to 3 in only 1 protein and 82 proteins were down regulated (Table 3.7).

Stage	G3T1 → G3T2	G3T2 → G3T3	G3T1 → G3T2 → G3T3
Up-regulation	95	1	0
Down-regulation	1	82	3

Table 3.7 Expression analysis on the stage of the tumour progression. Expression analysis was performed on the samples with high grade tumour (grade 3) to investigate the changes in the protein profile of the sample when tumour stage progresses.

When comparing all 3 stages, in the progression of the disease from stage 1 to 2 to 3, three proteins showed down regulation whereas no protein was up regulated.

3.4 Discussion

Application of proteomics workflows using state of the art instrumentation (high resolution liquid chromatography and high resolution mass spectrometry) in the biomarker discovery route (Pavlou *et al.*, 2013) helps in the discovery of candidate biomarkers. Similar to every other strategy, there are limitations in a plasma proteomics workflow (Rifai *et al.*, 2006) (i.e. limited dynamic range of the instrument versus high dynamic range of proteins in plasma (Anderson, 2010) which needs to be addressed before performing any clinical study. A limitation in clinical sample volume also highlights the risk of wasting sample when using a non-optimised protocol. There are three main stages in a proteomics workflow (Reinders & Sickmann, 2009) including: sample preparation, sample analysis and data analysis. It was the aim of this experiment to assess, evaluate and optimise all 3 stages of the proposed workflow (Figure 3.26) for the study.

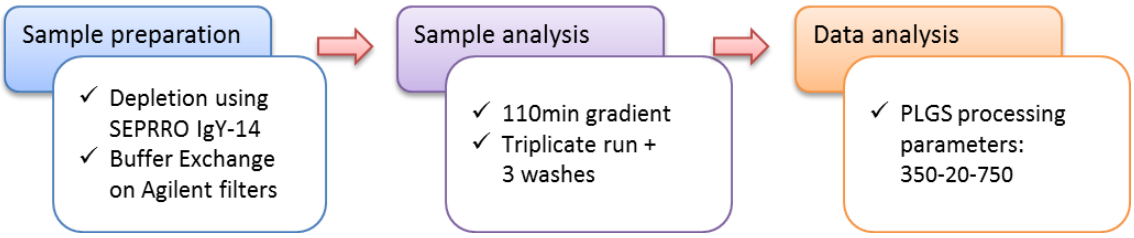

Figure 3.26 The optimised and evaluated protocol to use for the biomarker discovery phase

In the sample preparation, SEPRRO IgY-14 LC2 column was chosen for the immunodepletion of high abundance proteins from the plasma samples of controls and patients with bladder cancer. Agilent 5K MWCO filter was selected to concentrate the depleted samples followed by exchanging of the mass spectrometry incompatible buffer. In the sample analysis section, 110 minutes gradient was selected as the best for peptide separation. Injection of every sample in triplicate following by three wash

runs (20min each) using 1% and 0.1% formic acid was chosen as the best study design for the analysis leading to minimum sample carryover (contamination). In the data analysis, the best thresholds to be used in PLGS for identification of the proteins were selected as 350, 20 and 750 for the low energy threshold, high energy threshold and intensity threshold, respectively. The use of proteomics in biomarker discovery requires a reliable workflow which ensures that the quantitative measurement of a particular analyte demonstrates suitable reproducibility (Tabb *et al.*, 2010). Once optimised the reproducibility of the workflow was assessed to see whether the proposed protocol is reproducible using label-free LC-MS^E. The use of LC-MS^E in the workflow was clearly highly advantageous in terms of both capturing sufficient data to maximise the number of proteins identified but also to provide accurate quantitative data which is consequently amenable to comparative biomarker studies (Berg *et al.*, 2006). It has shown that protein profiling of plasma samples using LC-MS^E and label-free quantification (the proposed workflow for the discovery phase of the project) was precise and highly reproducible. Finally, for proteomics to make significant progress into plasma proteomics the reproducible measurement of low abundance proteins with clinically relevant mechanistic roles in disease aetiology is required. The work described within this chapter demonstrates that the analytical dynamic range is close to 5 orders of magnitude, which while less than the actual dynamic range of human plasma, is better than previously assessed workflows (Hortin & Sviridov, 2010). In the proposed workflow, using an immunodepletion strategy was clearly successful in detecting low abundant proteins such as interleukin 15 receptor subunit alpha and cell cycle checkpoint control protein RAD9A in plasma. These two proteins play integral roles in cell proliferation, cell cycle arrest and repair of DNA damage, (Fehniger &

Caligiuri, 2001; Broustas & Lieberman, 2012). Other examples of biologically interesting proteins with roles in cancer and other disease as ascertained from the literature were observed using this optimised workflow (Table 3.8).

Protein name	Disease	Reference range
Choliesterase	Gastric cancer (Gu *et al.*, 2005)	3-7 mg/L (Brock & Brock, 1990)
Biotinidase	Breast cancer (Kang *et al.*, 2010)	490 ng/ml (Farrah *et al.*, 2011)
Kallikerin	Lung cancer (Chee *et al.*, 2008)	50 mg/L (Hortin & Sviridov, 2010)
Hepatocyte growth factor activator	Activation in response to tissue injury (Miyazawa, 2010)	240 ng/ml (Farrah *et al.*, 2011)
Apolipoprotein C IV	Liver disease (Kim *et al.*, 2008)	1-4 mg/L (Kotite *et al.*, 2003)
Carboxypeptidase B2	****	4-15 mg/L (Willemse & Hendriks, 2006)

Table 3.8 The reference ranges of potential biomarkers identified by the methods. The associated diseases and normal plasma levels demonstrate that the proposed workflow is able to identify clinically relevant proteins.

These proteins are relevant to disease mechanisms and so can (and do) act as clinically relevant biomarkers. To this end, the proposed workflow was optimised, evaluated and its reproducibility was checked. Plasma samples used for all of the above mentioned steps were obtained from healthy donors. Overall, it has been shown that there are differences between the plasma protein content of a healthy person compared to a diseased one (Anderson, 2010). To assess the compatibility of the workflow using clinical samples a pilot study was performed using a representative number of plasma samples from bladder cancer patients. The dynamic range of the identified proteins includes 78.31ng (on column) for α-2 Macroglobulin down to 0.005ng (on column) for Plasminogen. This large dynamic range coverage, illustrates the efficiency of the depletion, the analytical (LC-MS) performance and the protocol compatibility with both patient and control samples. Having checked all of the identified proteins across the entire experiment, some clinically interesting proteins were found particularly for their role in cell cycle, cancer and inflammation processes.

Most of these proteins (Table 3.9) are previously reported in the literature for their cellular function and their role in cancer.

Protein Name	Found in	Amount (ng)	Function
Ras related protein Rab 4A	G 1 &3	0.55	Transportation
Anthrax toxin receptor like	G 1	0.48	Angiogenesis (Bradley & Young, 2003)
Citron Rho interacting kinase	G 2	5.82	Cytokinesis
Bardet Biedl syndrome 7	G 2	0.87	Leads to ciliopathy diseases
F box only protein 30	G 1	0.84	Ubiquitination
Spindlin-1	G 1	0.13	Tumourigenesis (Gao *et al.*, 2005)
Checkpoint protein HUS1	Control	3.20	DNA damage response
Ubiquitin D	Control	2.7	Ubiquitination
Tyrosyl-DNA phosphodiesterase 2	G 1	0.2	DNA repair enzyme (Cortes Ledesma *et al.*, 2009)
Kit ligand	G 2	0.05	Role in breast/ovarian cancer (Treff *et al.*, 2004)
Receptor binding cancer antigen expressed in SiSo cells	G 2	0.41	Apoptosis (Sonoda *et al.*, 2010)
Hemogen	G 3	0.03	hematopoietic development and neoplasms (Yang *et al.*, 2001)

Table 3.9 Low abundance proteins detected in different grades of bladder cancer patients. Most of these proteins are reported in the literature as candidate biomarkers due to their role in cancer.

The discrepancy between total number of identified proteins in the bladder cancer samples compared to the controls could be due to the un-equal number of samples in both groups which will lead to some variability (Rundle *et al.*, 2012). This was addressed in the discovery phase (Chapter 4) when more equivalent numbers of disease and control samples were used in the experiment. In the expression analysis, most of the down regulated proteins belong to a large group of protease inhibitors called Seprin which has control over coagulation and inflammation processes (Gatto *et al.*, 2013). Alteration in the expression level of most of these proteins is reported by other researchers as markers of carcinogenicity i.e. decreased level of Antithrombin-III in prostate cancer (Hong *et al.*, 2010) and lung cancer (Unsal *et al.*, 2004). Apolipoprotein-E up regulation has been reported in gastric cancer (Sakashita *et al.*, 2008) and ovarian

cancer (Chen *et al.*, 2005). Increased level of Alpha-1 acid glycoprotein in various types of cancer has also been reported by other researchers (Bruno *et al.*, 2003; Matsuura & Nakazawa, 1985).

In summary, in this chapter it has been shown that the optimised workflow is suitable for discovering biomarkers of bladder cancer.

Chapter Four

DISCOVERY PHASE

4 The discovery of biomarkers of bladder cancer

4.1 Introduction

The ultimate goal of using proteomic strategies in biomarker discovery is to identify disease biomarkers to be used in the clinic. Although many proteins have been reported as proteomics-based potential biomarkers for different diseases (mainly cancer) in more than 10,000 publications (Anderson, 2010), to date, only two have successfully received FDA approval to be used in the clinic: VeriStrat and OVA1 (*U.S. Food and Drug Administration.*2013). VeriStrat is used for managing the treatment of patients with lung cancer. OVA1 measures five different proteins (CA 125, transthyretin, ApoA1, beta-2 microglobulin and transferrin) in the blood sample of patients already diagnosed with ovarian cancer and helps in the evaluation of the tumour prior to surgery. This clearly demonstrates the failure to translate the promise of proteomics in biomarker discovery into the clinical tools which are much needed (Diamandis, 2010; Diamandis, 2012). This failure is due to a number of specific reasons:

A. Poor experimental design that failed to take into account the more complex clinical situation,

B. Analytical technology which had insufficient dynamic range to measure proteins in clinical samples,

C. Inelegant methods for protein quantification,

D. Narrow and often irrational choice of informatic approaches to identify and quantify proteins found in highly complex samples such as plasma.

The addressing of these issues would achieve a notable step forward towards the ultimate goal of clinical proteomics. In this study an attempt has been made to overcome the issues by performing the discovery phase (identification of candidate

biomarkers), quantification (quantifying the expression level) and verification (assessment of specificity and sensitivity) stages of the biomarker discovery workflow (see Chapter 1, section 1.4.1) using plasma samples obtained from patients with bladder neoplasms and disease free controls.

To perform the analysis based on a clinically accepted experimental design (to address A; see above) a rationale was developed in conjunction with a bladder cancer specialist. Using the experimental design, discrimination of the samples between and within the groups was possible, and ultimately enables the addressing of four clinically relevant questions:

1. Are we able to discriminate disease (plasma samples from patients with bladder cancer) from healthy control samples?
2. Are we able to discriminate disease from control (healthy and hospitalised plasma samples) samples?
3. Are we able to discriminate low grade Ta tumours from high grade Ta, high grade T1 and muscle invasive tumours (low risk vs. high risk)?
4. Are we able to discriminate early (TaT1-G3) and advanced stages (T2T3-G3) of high grade tumours?

The above mentioned questions will add useful information to the patients' profile which ultimately helps clinicians throughout the patient's journey from their first presentation at the clinic until treatment. By discovering markers of bladder cancer when comparing diseases to healthy controls (question 1 above) we will look to identify diagnostic markers. By including non-bladder cancer diseases into the control group (question 2) we look to exclude clinical markers that are not cancer related (i.e.

urinary system inflammation or haematuria). This subtraction of disease related markers from bladder cancer group may enable the discovery of more specific markers for bladder cancer. Low grade Ta tumours have very low risk of recurrence and/or progression compared to other stages and grades of the disease. Low grade Ta tumours are currently treated with TURBT followed by regular cystoscopy/cytology checks for up to 5 years and then discharge whereas other sub-groups of the disease receive more aggressive treatment for longer period of time. Assessing the protein profile of low grade Ta tumours and comparing it to other disease samples (question 3) we may be able to find the difference in molecular signature of low grade Ta compared to high grade Ta and MID. Currently patients with early stage high grade tumours (Ta/T1, grade 3) are receiving TURBT followed by BCG treatment with cystoscopy/cytology follow ups every 3-6 months and have a high mortality rate. Conversely, for patients with advanced stage tumour (T2/T3, grade 3), neo-adjuvant chemotherapy with/without follow-on radical cystectomy or radiotherapy is used for treatment. By looking at early vs. advanced stages of high grade tumours (question 4), and having identified likely markers of risk (question 3) we may be able to discover prognostic markers of progression for the Ta/T1 group of patients. Knowing this may enable clinicians to target more aggressive treatment towards patients with high risk Ta/T1 stage tumour. Being able to answer each one of the above mentioned questions will be a successful step towards the discovery of biomarkers of bladder cancer and ultimately would enable clinicians to prepare a more personalised treatment plan.

A specific UPLC-HDMS^{E} label free quantitation method (to address B), using data independent acquisition, for protein profiling of the samples was designed, developed

and optimised using plasma samples obtained from healthy volunteers. The optimised method was further evaluated using a batch of randomly selected plasma samples from bladder cancer patients to assess the compatibility of the methods with disease samples (discussed in Chapter 3).

Accurate quantitation of the identified proteins is always a crucial stage in the biomarker discovery. The quantitation method (to address C) used in the workflow, (label free Hi-3, (Silva *et al.*, 2006), quantifies the absolute amount of every single identified protein in relation to the internal standard (of a known concentration) added prior to the analysis. The UPLC-HDMS^E label-free quantitation strategy has not been used before for protein profiling of plasma samples obtained from bladder cancer patients.

The choosing of the right bioinformatics package (to address D) for interpretation of data is crucial as the nature of the package used can lead to significantly different results (News in Proteomics Research. 2013). In this Chapter, an investigation into the discovery of biomarkers using a combination of informatics approaches will be attempted to overcome the bias introduced by using a single informatics package. To assess expression levels of proteins of samples in different groups, a plan consisting of four data processing packages was used (see Chapter 4, section 4.2.4). Each of the packages uses different statistical means by which to identify proteins and/or compare expression levels alteration. Comparisons of these approaches and identifying common candidate proteins within different groups of samples allows for greater confidence in this initial phase. Alteration in expression of the candidates can then be verified using western blot (discussed in Chapter 5).

4.2 Methods

4.2.1 Sampling information

The second batch of samples received from collaborators in the University of Torino (Turin, Italy) consisted of 44 plasma samples including 20 healthy controls and 24 bladder cancer samples. The total number of plasma samples used in the study was 104 including 64 bladder cancer samples and 40 control samples in two groups consisting of 20 healthy controls and 20 hospitalised non-cancerous controls. All of the samples were age (average age was 66 years old) and sex matched with sufficient clinical metadata including stage, grade, diagnosis date, type of treatment and the date of recurrence and/or cystectomy (if any) for the bladder cancer samples (Appendix D). The diseased samples were grouped into 3 groups based on the grade and stage of the disease and controls into 2 groups (Table 4.1).

Grade	G1	G2	G3	CNT1	CNT2
Ta	18	7	6	20	20
T1		6	6		
T2		1	15		
T3			4		
T4			1		
Total	18	14	32	20	20

Table 4.1 Total number of samples in different groups of disease and control. G stands for grade of the tumour, T is the tumour stage, CNT1 is the hospitalised controls and CNT2 is healthy controls.

Due to the long analysis time (7 hours per sample), samples were pooled into groups of three within the same group. Due to the unequal number of samples in every group and lack of samples in others, the sample in grade 2, stage 2 (G2T2) was not used. Also, the sample in grade 3, stage 4 (G3T4) and one sample from grade 3, stage 2 (G3T2) was grouped with the other 4 samples of grade 3, stage 3 (G3T3) to have 6 samples in

that group (Table 4.2, left). Only 18 samples from each control group were used for the pooling. So in total, 90 samples were selected for pooling which resulted in 30 pooled samples (Table 4.2, right).

Grade	G1	G2	G3	CNT1	CNT2
Ta	18	6	6	18	18
T1		6	6		
T2		0	6		
T3			6		
T4			0		
Total	18	12	24	18	18

Classification	Un-pooled Sample	Pooled sample
Grade 1	18	6
Grade 2	12	4
Grade 3	24	8
CNT1	18	6
CNT2	18	6
Total	90	30

Table 4.2 Pooling information. Not all of the 104 samples were used for pooling (left table). The total number of samples used was 90 resulted in 30 pooled samples.

In total, 30 pooled samples were used in the analysis (Table 4.3). As described earlier in Chapter 3, the optimised protocol for sample preparation (see Chapter 3) was applied to every single sample until the end of the digestion step where every 3 samples were pooled and the protocol continued.

	Sample ID	Sample number	T	G
1	P1G1Ta	7	a	1
		64	a	1
		65	a	1
2	P2G1Ta	4	a	1
		66	a	1
		67	a	1
3	P3G1Ta	70	a	1
		75	a	1
		72	a	1
4	P4G1Ta	17	a	1
		73	a	1
		74	a	1
5	P5G1Ta	71	a	1
		77	a	1
		78	a	1
6	P6G1Ta	12	a	1
		11	a	1
		82	a	1
7	P7G2Ta	9	a	2
		5	a	2
		8	a	2
8	P8G2Ta	1	a	2
		10	a	2
		16	a	2
9	P9G2T1	2	1	2
		6	1	2
		13	1	2
10	P10G2T1	19	1	2
		79	1	2
		81	1	2
11	P11G3Ta	15	a	3
		61	a	3
		69	a	3
12	P12G3Ta	63	a	3
		68	a	3
		62	a	3
13	P13G3T1	3	1	3
		18	1	3
		20	1	3
14	P14G3T1	76	1	3
		80	1	3
		83	1	3
15	P15G3T2	23	2	3
		25	2	3
		40	2	3

	Sample ID	Sample number	T	G
16	P16G3T2	36	2	3
		30	2	3
		21	2	3
17	P29G3T3	26	3	3
		80	3b	3
		39	3b	3
18	P30G3T3	29	4	3
		24	2a	2
		32	2b	3
19	P17CNT1	41	-----	-----
		49	-----	-----
		47	-----	-----
20	P18CNT1	46	-----	-----
		50	-----	-----
		44	-----	-----
21	P19CNT1	45	-----	-----
		42	-----	-----
		58	-----	-----
22	P20CNT1	57	-----	-----
		56	-----	-----
		52	-----	-----
23	P21CNT1	54	-----	-----
		48	-----	-----
		55	-----	-----
24	P22CNT1	51	-----	-----
		43	-----	-----
		53	-----	-----
25	P23CNT2	86	-----	-----
		99	-----	-----
		88	-----	-----
26	P24CNT2	89	-----	-----
		90	-----	-----
		102	-----	-----
27	P25CNT2	92	-----	-----
		87	-----	-----
		94	-----	-----
28	P26CNT2	97	-----	-----
		96	-----	-----
		95	-----	-----
29	P27CNT2	98	-----	-----
		91	-----	-----
		100	-----	-----
30	P28CNT2	101	-----	-----
		93	-----	-----
		103	-----	-----

Table 4.3 Pooling information with details. 30 pooled samples were selected randomly from the same grade and stages.

4.2.2 Quality Controls

Quality controls (QC) are the main checkpoints when analysing a large number of samples in analytical chemistry. These are particularly important in studies with low sample throughput where a large number of samples are to be analysed. The need to have regular checkpoints (quality controls) throughout the workflow to assess the

instrument's performance is essential to ensure the data produced is meaningful (Cairns, 2011). In this project, a pooled sample of 5µl aliquot from 15 samples (including bladder cancer patients and controls) was prepared. This QC was used at the start of the sample list, after every 7 samples and at the end of the list (Figure 4.1). After every QC sample, the sample list was paused to investigate the QC's chromatogram and also to process its raw data files with PLGS. If the chromatogram and the total number of proteins/peptides identified by PLGS were similar to the first QC sample (ran at the start of the sample list) then the rest of the samples in the sample list were analysed.

File Name	File Text	MS file	Inlet file	Location	Injection volume
AMH_131211_QC_01	QC-1	110min_HDMSE_new3	110min_partialloop	1:F,8	2
AMH_wash_001	1% FA	20min_HDMSe	20min_partialloop	2:A,8	3.5
AMH_wash_002	0.1% FA	20min_HDMSe	20min_partialloop	2:A,7	3.5
AMH_wash_003	0.1% FA	20min_HDMSe	20min_partialloop	2:A,7	3.5
AMH_Final_131211_P1G1Ta_01	P1G1Ta 1	110min_HDMSE_new3	110min_partialloop	1:A,1	4
AMH_Final_131211_P1G1Ta_02	P1G1Ta 2	110min_HDMSE_new3	110min_partialloop	1:A,1	4
AMH_Final_131211_P1G1Ta_03	P1G1Ta 3	110min_HDMSE_new3	110min_partialloop	1:A,1	4
AMH_wash_04	1% FA	20min_HDMSe	20min_partialloop	2:A,8	3.5
AMH_wash_05	0.1% FA	20min_HDMSe	20min_partialloop	2:A,7	3.5
AMH_wash_06	0.1% FA	20min_HDMSe	20min_partialloop	2:A,7	3.5
AMH_Final_131211_P2G1Ta_01	P2G1Ta 1	110min_HDMSE_new3	110min_partialloop	1:A,2	3.9
AMH_Final_131211_P2G1Ta_02	P2G1Ta 2	110min_HDMSE_new3	110min_partialloop	1:A,2	3.9
AMH_Final_131211_P2G1Ta_03	P2G1Ta 3	110min_HDMSE_new3	110min_partialloop	1:A,2	3.9
AMH_wash_07	1% FA	20min_HDMSe	20min_partialloop	2:A,8	3.5
AMH_wash_08	0.1% FA	20min_HDMSe	20min_partialloop	2:A,7	3.5
AMH_wash_09	0.1% FA	20min_HDMSe	20min_partialloop	2:A,7	3.5
AMH_Final_131211_P7G2Ta_01	P7G1Ta 1	110min_HDMSE_new3	110min_partialloop	1:A,7	4.7
AMH_Final_131211_P7G2Ta_02	P7G1Ta 2	110min_HDMSE_new3	110min_partialloop	1:A,7	4.7
AMH_Final_131211_P7G2Ta_03	P7G1Ta 3	110min_HDMSE_new3	110min_partialloop	1:A,7	4.7
AMH_wash_22	1% FA	20min_HDMSe	20min_partialloop	2:A,8	3.5
AMH_wash_23	0.1% FA	20min_HDMSe	20min_partialloop	2:A,7	3.5
AMH_wash_24	0.1% FA	20min_HDMSe	20min_partialloop	2:A,7	3.5
AMH_131211_QC_02	QC-2	110min_HDMSE_new3	110min_partialloop	1:F,8	2
AMH_wash_25	1% FA	20min_HDMSe	20min_partialloop	2:A,8	3.5
AMH_wash_26	0.1% FA	20min_HDMSe	20min_partialloop	2:A,7	3.5
AMH_wash_27	0.1% FA	20min_HDMSe	20min_partialloop	2:A,7	3.5

Figure 4.1 A sample list created on MassLynx. QC samples were injected at the start and after every 7 sample injections.

4.2.3 Proteomics workflow

The optimised and evaluated protocol (see Chapter 3) was used to prepare and analyse the sample and process the data (Figure 4.2).

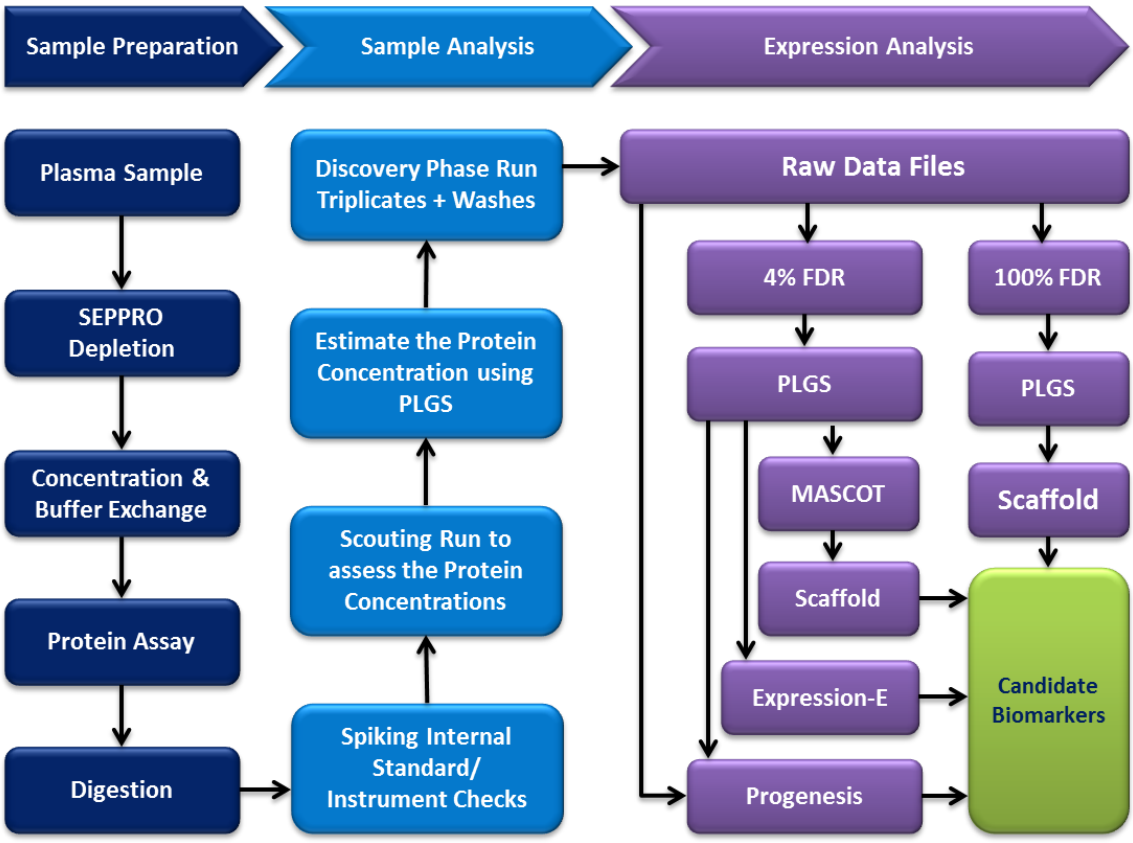

Figure 4.2 A schematic view of the protocol developed for the study. It includes sample preparation, sample analysis and expression analysis which lead to finding a list of candidate biomarkers.

4.2.4 Expression analysis

Expression analysis was performed on the identified and quantified proteins. Using installed plug-ins, protein lists were exported in the correct format for Progenesis LC-MS, Scaffold and MASCOT.

4.2.4.1 Expression^{E}

Identification of proteins was carried out using Identity^{E} and subsequent proteins were placed in different groups of samples. Groups of samples were then compared in a

number of different ways following experimental design of the study. Generated protein tables were normalised using ADH, the internal standard that was added to the samples for the purpose of quantitation. The protein list created by Expression^{E} was filtered to show the proteins that are presented in all 3 replicate injections of a sample. Proteins showing significant increase in the expression in one group compared to the other were reported as up-regulated proteins (p-value ≥ 0.95). Proteins with a significant decrease in the expression (p-value ≤ 0.05) were labelled as down-regulated proteins and every other protein with non-significant change (0.05 > p-value > 0.95) were reported as no change in the expression of that proteins between the groups. The final filtered protein list of every group was exported to excel sheets (*.csv files) for further analysis.

4.2.4.2 Progenesis LC-MS

The raw data files of the samples were imported into Progenesis LC-MS for expression analysis. Within Progenesis LC-MS, a raw data file that represents a typical profile in terms of peptide retention time (R*t*) and m/z is used as a reference run. This raw data file is used to establish a calibration for all other samples to be compared with. The reference run was chosen and every other peak was aligned based on the reference run. The data was filtered based on the charge state and only features with 2 to 5 charges were included. The same comparisons between groups (experimental design) that were chosen in Expression^{E} (see below, section 4.3.3) were applied for Progenesis LC-MS analysis and the protein identification data files (exported before using the Progenesis LC-MS plug-in) were imported. Another level of filtering was applied to the identified peptides where the ones with only one hit were excluded (the same as

PLGS). Using the peptide identification lists, protein lists were generated by the software and a third filtering step was applied where the proteins with unacceptable p-values were excluded. Final protein lists were exported into excel sheets, sorted and prepared for the final comparison.

4.2.4.3 Scaffold

The MS raw data files were processed using PLGS with a false discovery rate of 100% and imported into Scaffold. A database similar to the one used in PLGS was created and saved for Scaffold. The design of the experiment (group comparisons) was the same as previously chosen for the other two packages. The data was analysed and the generated protein lists were filtered based on accepted p-values and fold changes in some instances. The final protein lists were exported to excel and sorted for further analysis.

4.2.4.4 MASCOT/Scaffold

The output data processed using PLGS were imported into MASCOT and searched against the database. All of the search parameters were set the same as the ones in PLGS. Uniprot Human (reviewed) was chosen for the database, Carbamidomethyl C for fixed modification and Oxidation M, Deamidation N and Phospho STY selected for variable modification. Every single sample was searched in MASCOT and the identified protein lists were exported to Scaffold for quantitation. The same procedure as described above (see above, section 4.2.4.3) was performed for quantification and expression analysis of the proteins identified by MASCOT. The final protein lists, were exported into excel sheets for final comparison.

4.2.5 Unique proteins

To be able to select the proteins uniquely present in disease or control group, the list of identified proteins (using either of PLGS and MASCOT) was compared. This comparison resulted in a selection of proteins identified only in one of the groups (disease or control). The unique proteins were stringently filtered so that a particular protein had to be present in every one of the three replicate injections of a sample. Also, the unique proteins of every group were again compared to find the proteins present in all of the disease samples and all the control samples. Unique proteins in the disease samples identified by PLGS were compared to the unique proteins identified by MASCOT and the ones present in both of the methods were chosen as the refined final unique protein list for disease and control samples (Figure 4.3).

In addition to this analysis, the data was coded before being handed over to a biomedical statistician who produced a list of common protein in replicates. Those lists were then exported into excel sheets for further analysis.

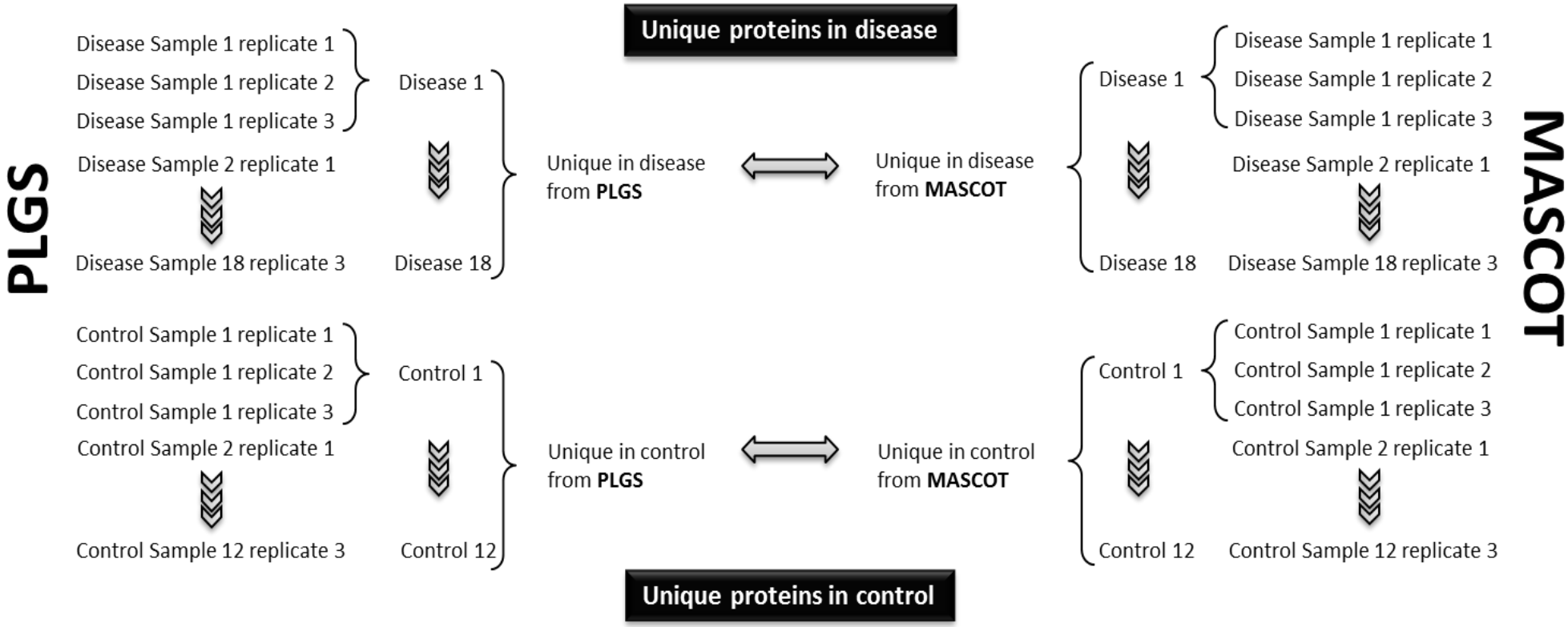

Figure 4.3 Schematic plan for the selection of unique proteins. Unique proteins were selected as follows: the replicates of every injection were compared to find the proteins presented in all 3 replicates. This resulted in 18 lists of unique proteins from each disease sample and 12 in control samples. By comparing lists separately (18 and 12) the common proteins found only in either disease and control was selected. Two lists were generated of protein identifications for disease and for control. The same procedure was repeated for the proteins identified by MASCOT. Then two amalgamated lists of unique proteins for disease and control found in both MASCOT and PLGS searches were generated.

4.3 Results

4.3.1 Quality controls

The samples were analysed in two different batches and consequently two sets of quality controls were used. A QC sample was injected five times for each sample set. Chromatograms of the five injections were compared to establish the reproducibility of the instrument (Figure 4.4, left). Intensity and retention time of a randomly selected ion (m/z 610.8) was compared between five injections of the quality control (Figure 4.4, right).

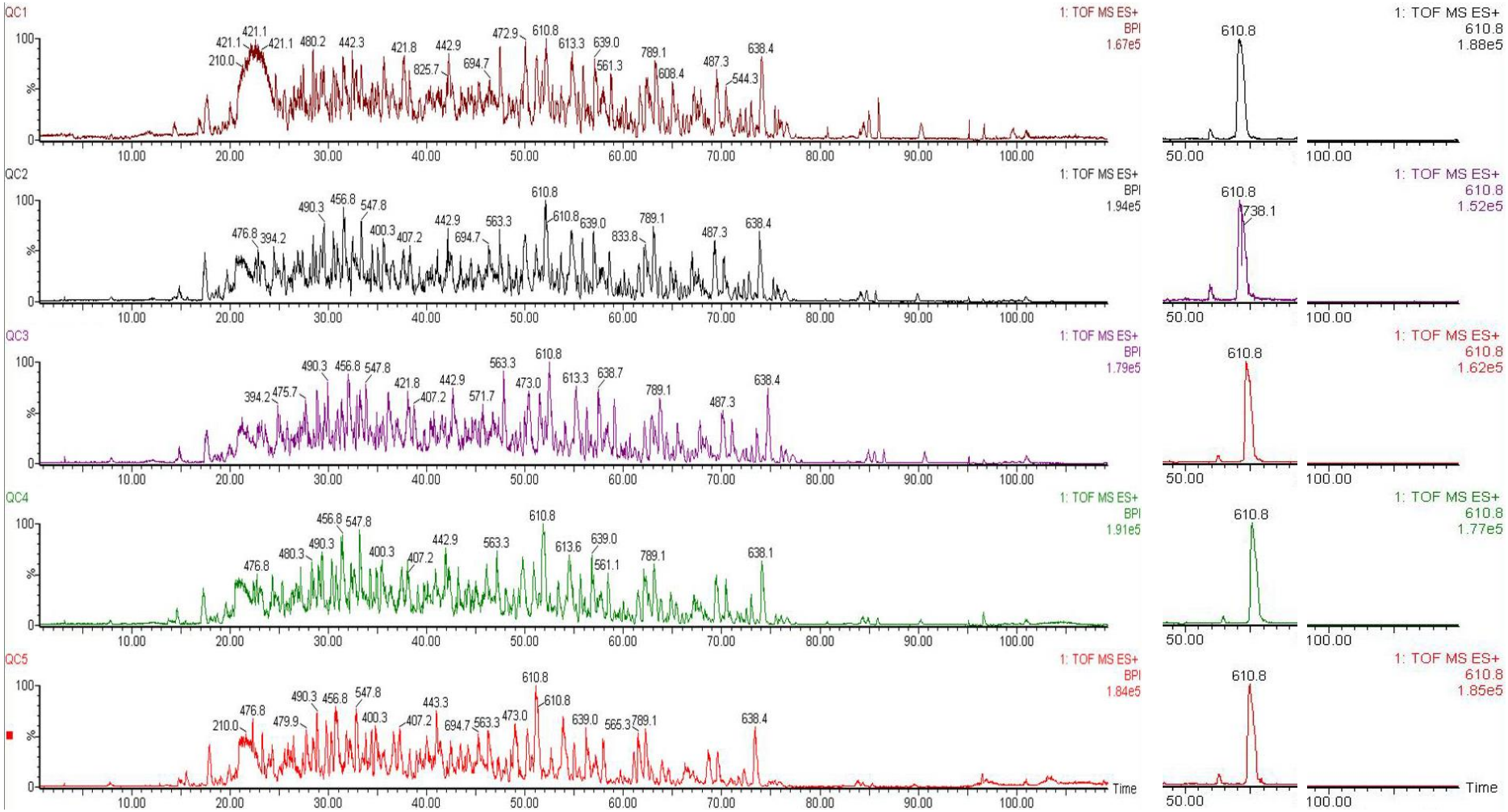

Figure 4.4 Quality control chromatograms. Chromatograms of 5 injections of a QC sample (left) repeatedly run throughout the study. Chromatograms were assessed for retention time and intensity consistency. A representative chromatogram (with m/z 610.8, right) reveals the time and intensity alignment.

Assessment of chromatographic separation of the peptides in the QC samples showed good reproducibility (similar chromatograms and intensities (on average 1.7e6)) throughout the study. The retention time of peptides began at about 20 minutes and ended at about 75 minutes for all of the injections. As a typical example, the ion with m/z 610.8 (randomly selected) eluted at the same time in all five of the QC injections (retention time at 55min ± 8 seconds) with the intensity of 1.73e5 ± 9%. Low levels of variations between the five QC injections, showed high levels of consistency in the instrument performance. Also, raw data files of the QC samples were processed using the thresholds evaluated in the method development (Chapter3) (350-20-750 for the low energy, high energy and intensity threshold, respectively). When raw data files of the QC samples were processed with PLGS, between 126 to 134 proteins in one set and between 122 to 130 proteins in the other set were identified (Figure 4.5). On average 129 proteins (RSD = 2.81%) were identified in the first batch and 126 proteins (RSD = 2.64%) in the second batch. The low level of RSD (<3%) demonstrates high levels of consistency between the two batches of quality controls.

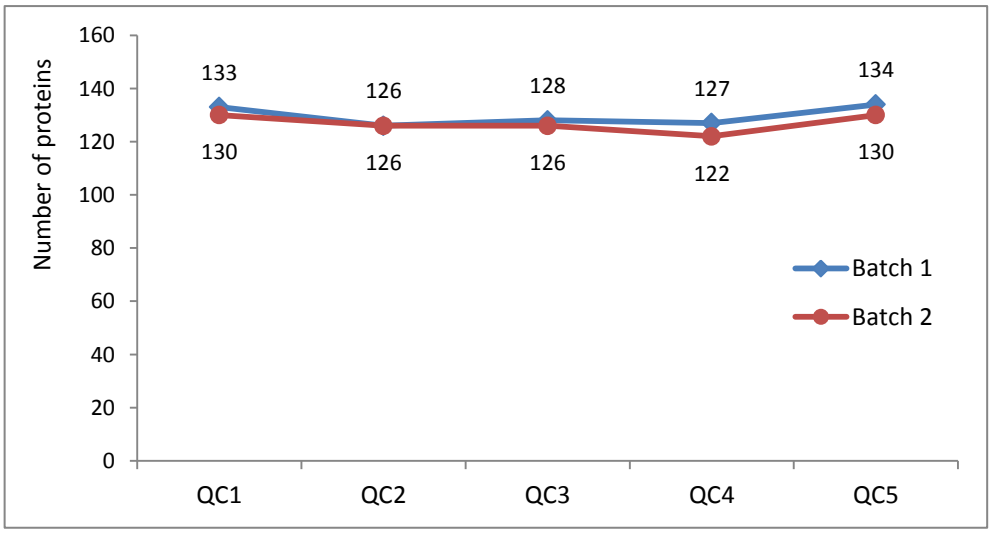

Figure 4.5 Total number of proteins identified in two different batches of the Quality Control (QC) runs

The protein lists generated by PLGS were compared and resulted in 104 proteins common between 5 injections of the first QC batch of samples and 100 proteins common between the second batch of QC samples. The same pattern was seen when analysing the number of peptides (Figure 4.6) where on average 2931 peptides (RSD = 4.41%) were identified.

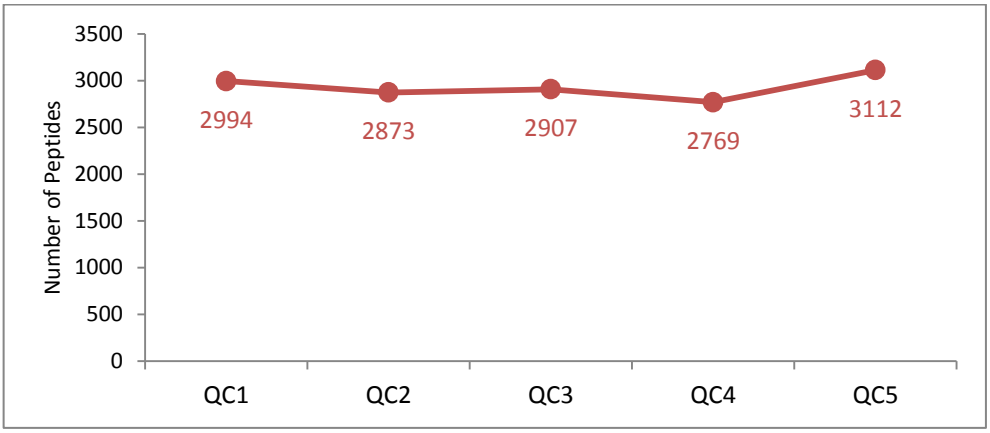

Figure 4.6 Total number of peptides identified in a QC run

To further assess the accuracy and reproducibility of the discovery phase of the project, 20 proteins were randomly selected from the list of in-common proteins. The number of peptides identified in every protein in 5 injections was averaged and plotted in Figure 4.7.

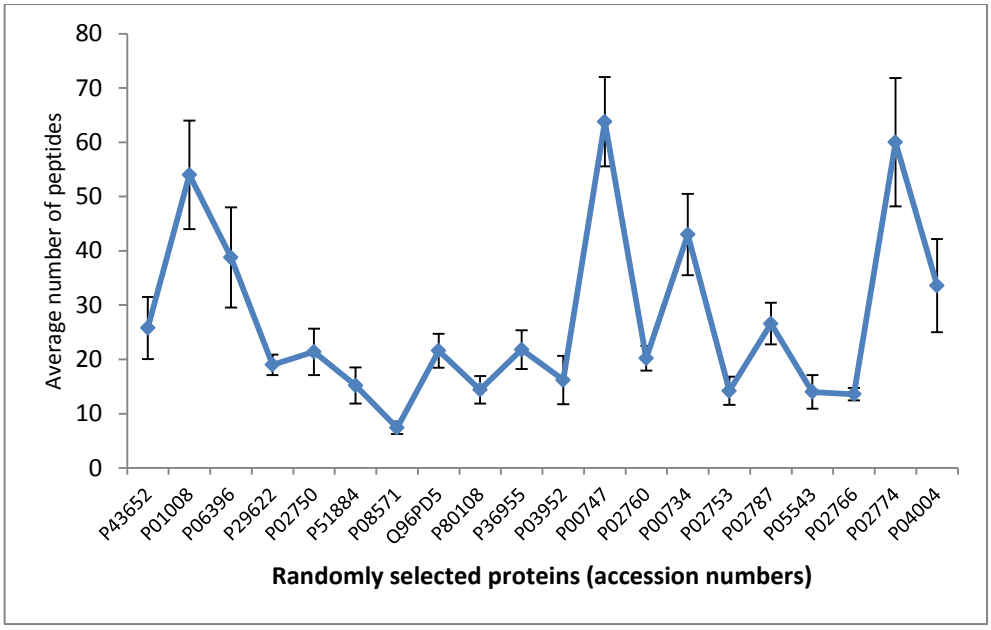

Figure 4.7 Average number of peptides identified in 20 of the randomly selected proteins in five QC runs

These results demonstrate that the study was carried out with a measurable consistency in instrument performance. Moreover, variation observed between samples would be due to the biological variation that is due to the disease pathogenesis rather than technical variation.

4.3.2 Total number of proteins

After adjustment of the injection volume based on the scouting runs samples were analysed and processed using PLGS. On average between 117 and 151 proteins were identified in the samples (Figure 4.8).

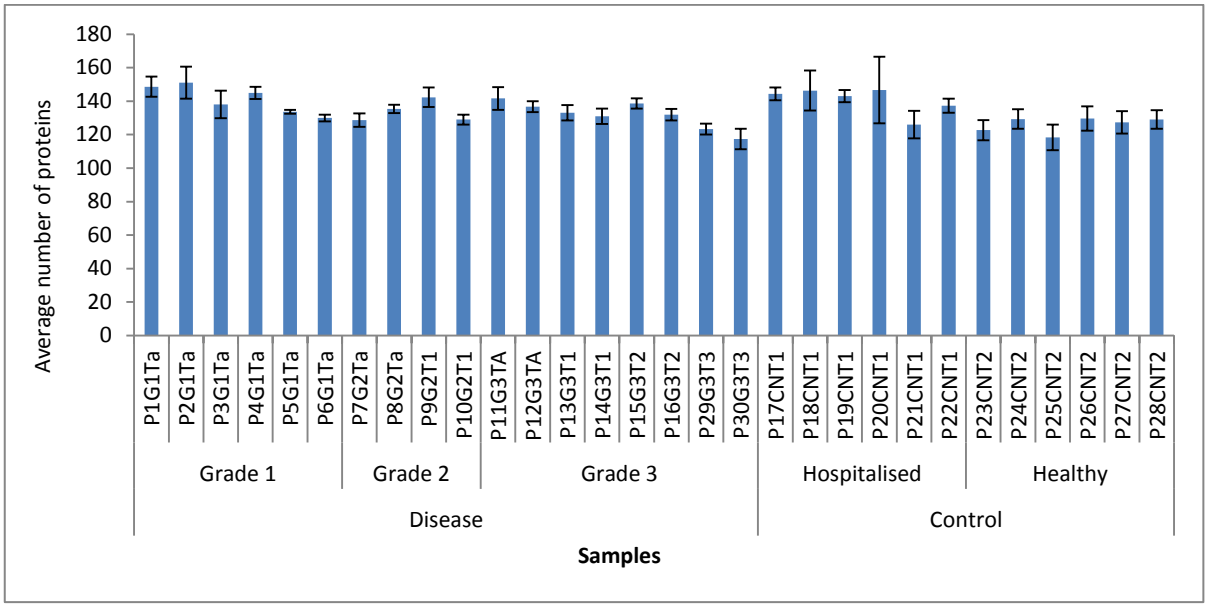

Figure 4.8 Total number of proteins identified in 18 disease and 12 control pooled samples

Total amount of protein in every injection of triplicate was calculated and averaged (Figure 4.9). The aim of injecting between 2 and 2.5µg protein on the column was achieved in almost all of the sample but for sample P15G3T2. Although the total amount of protein in that sample was about 1.5µg, the average number of identified protein was 139 which was within the range of other samples with higher concentration.

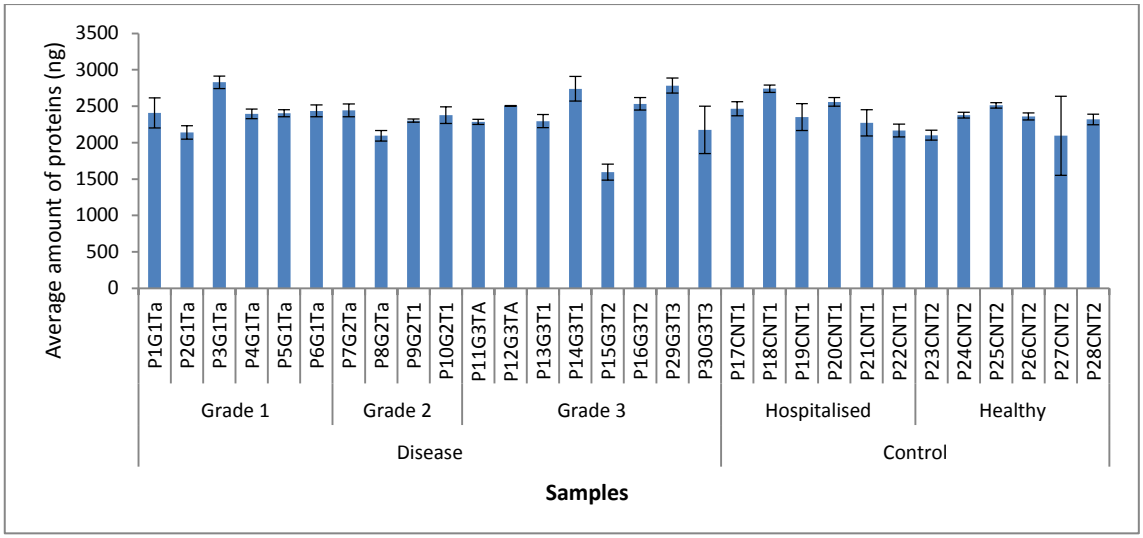

Figure 4.9 Total amount of proteins identified across all 30 pooled sample

Relative standard deviation for the above two graphs (total number of proteins and total amounts of proteins) was calculated and plotted in the Figure 4.10 below.

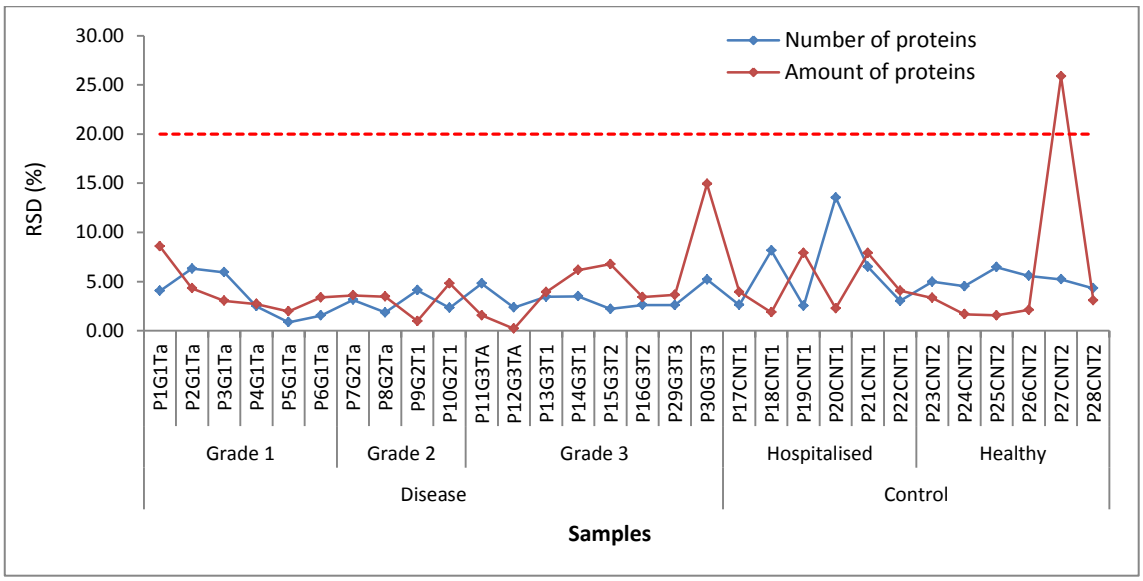

Figure 4.10 RSD value for total number of proteins and total amount of proteins

With the single exception of sample P27CNT2 (with RSD value of higher than the acceptable value (25%)), every other sample had RSD values of less than 15%. This again demonstrates that the study is reproducible and the identified proteins are reliable.

4.3.3 Expression analysis

Expression analysis was performed on the samples in both groups of disease and control and also subgroups of each based on Figure 4.11 (as discussed earlier in this Chapter).

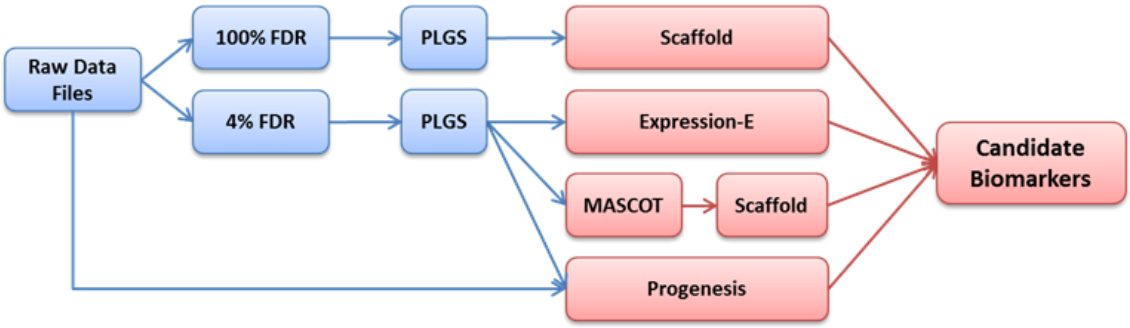

Figure 4.11 Expression analysis pattern

Four different comparisons (Figure 4.12) for expression analysis were made between the sample groups.

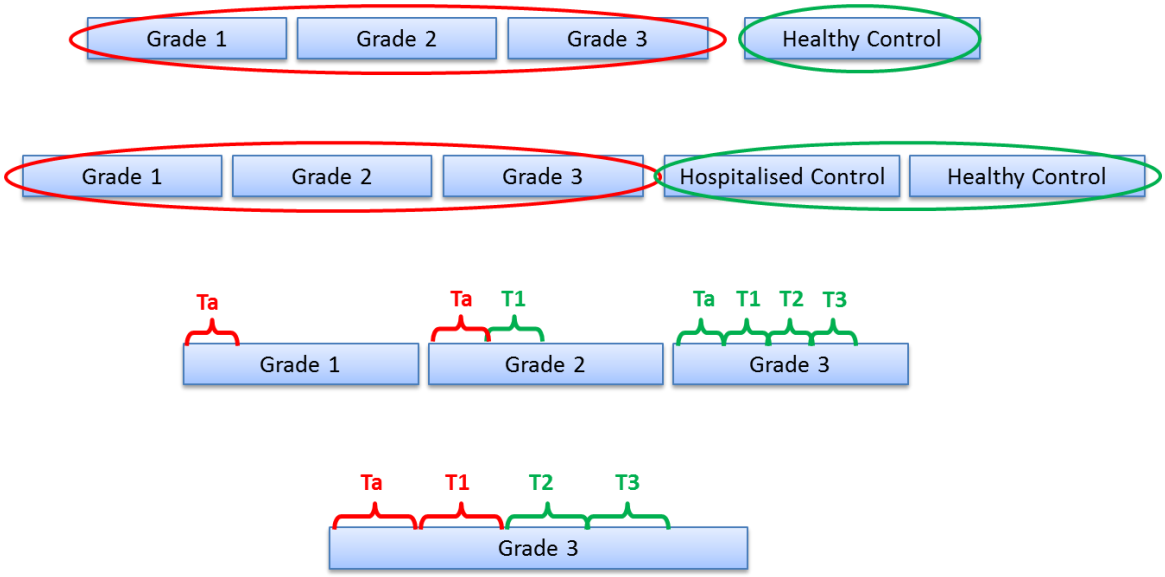

Figure 4.12 Four different comparisons for expression analysis between the sample groups based on the clinical questions. All of the reds were compared to the greens.

The rationale for the selection of the comparisons was primarily based on asking 4 clinically relevant questions: namely (**1**) by using this method are we able to find any

difference between disease and healthy control samples? (**2**) Are we able to exclude non-cancer related markers from bladder cancer samples by comparing disease to all of the controls? (**3**) Are we able to discriminate low grade Ta tumours from high grade Ta, high grade T1 and muscle invasive tumours? And finally (**4**) are we able to discriminate early (TaT1-G3) and advanced stages (T2T3-G3) of high grade tumours? This rationale was developed in conjunction with a bladder oncologist. All four data analysis packages were used to apply expression analysis for each comparison. Every two groups of data from the samples to be compared are coloured in red and green.

4.3.3.1 Bladder Cancer vs. healthy controls

Over-expression in 115 proteins of the bladder cancer samples was seen by **Expression^{E}** (Table 4.4). In the same comparison, 19 proteins were down-regulated and 31 proteins were reported with no significant changes in expression. **Progenesis LC-MS** reported up-regulation in 8 and down-regulation in 7 proteins. Expression analysis on the disease and healthy control samples using **Scaffold** showed over-expression in all of the identified proteins, 153. Hence no down-regulation/no changes were observed. **MASCOT/Scaffold** expression analysis resulted in 66 up-regulated and 6 down-regulated proteins.

	Disease > Control	Un-changed	Disease < Control
Expression^{E}	115	31	19
Progenesis LC-MS	8	-	7
PLGS-Scaffold	153	30	0
MASCOT-Scaffold	66	119	6

Table 4.4 Number of proteins with an alteration in their expression analysis identified using different package when comparing disease vs. healthy control. Complete list is provided in appendix A, tables A-3 & A-4.

When all 4 lists of up-regulated proteins were compared 5 proteins were common to all 4 lists and 35 proteins were common to at least 3 lists (Table 4.5).

Picked up in all 4 packages	Picked up in 3 packages	
Alpha 1B glycoprotein	Afamin	Haptoglobin related protein
Apolipoprotein A I	Alpha 1 acid glycoprotein 2	Hemoglobin subunit alpha
Coagulation factor XII	Alpha 1 antitrypsin	Hemoglobin subunit delta
Fibronectin	Alpha 2 macroglobulin	Ig alpha 1 chain C region
Hemoglobin subunit beta	Apolipoprotein B 100	Ig gamma 4 chain C region
	Biotinidase	Ig heavy chain V III region GAL
	Carboxypeptidase B2	Ig lambda chain V III region LOI
	Ceruloplasmin	Inter alpha trypsin inhibitor heavy chain H2
	Clusterin	Inter alpha trypsin inhibitor heavy chain H4
	Properdin	Complement factor H related protein 2
	Complement C5	Plasma protease C1 inhibitor
	Serotransferrin	Complement C1q subcomponent subunit A
	Complement factor B	Retinol binding protein 4
	Complement factor H	Complement component C8 gamma chain
	Plasma kallikrein	Serum amyloid P component
	Extracellular matrix protein 1	Vitamin K dependent protein S
	Fibrinogen gamma chain	Zinc alpha 2 glycoprotein
	Gelsolin	

Table 4.5 Up-regulation in the protein expression seen by 4 and 3 of the packages when comparing disease vs. healthy control

4.3.3.2 Bladder cancer vs. healthy and hospitalised controls

Expression^{E}: By comparing all of the disease samples to all of the control samples using Expression^{E} 115 proteins were up-regulated and 25 proteins down-regulated in the disease group (Table 4.6). 55 proteins were also reported to have no significant change in their expression. **Progenesis LC-MS:** Comparing the same groups of samples in Progenesis LC-MS resulted in a list of 24 proteins showing up-regulation and only one protein was down-regulated. When **Scaffold** was used for expression analysis on the disease vs. control samples, 95 proteins were up-regulated and 2 proteins were down-regulated. Scaffold also generated a list of 90 proteins with no significant change in their expression. Using Scaffold for expression analysis of the proteins identified by

MASCOT resulted in over-expression in a list of 64 proteins, down regulation of 11 proteins and no significant changes in the expression of 124 proteins.

	Disease > Control	Un-changed	Disease < Control
Expression[E]	115	55	25
Progenesis LC-MS	24	-	1
PLGS-Scaffold	95	90	2
MASCOT-Scaffold	64	124	11

Table 4.6 Number of proteins with an alteration in their expression analysis identified using different package when comparing disease vs. control. Complete list is provided in appendix A, tables A-1 & A-2.

The proteins that were present in the same expression pattern in all four packages were identified and extracted. Up-regulation in the protein expression of 11 proteins was seen by all four packages and 15 proteins were also seen in 3 of the packages (Table 4.7).

Picked up in 4 packages	Picked up in 3 packages
Afamin	Apolipoprotein B 100
Alpha 1 acid glycoprotein 1	Attractin
Alpha 1 antitrypsin	Beta 2 glycoprotein 1
Alpha 1B glycoprotein	Complement C1q subcomponent subunit A & B
Apolipoprotein A I	Complement C3
Complement component C7	Complement component C8 gamma chain
Complement factor H related protein 1	Haptoglobin related protein
Haptoglobin	Hemoglobin subunit alpha & delta
Hemoglobin subunit beta	Inter alpha trypsin inhibitor heavy chain H4
Serotransferrin	Kininogen 1
Serum albumin	Mannose binding protein C
	Plasma kallikrein
	Properdin
	Tetranectin
	Vitamin D binding protein

Table 4.7 Up-regulation in the protein expression seen by 4 and 3 of the packages when comparing disease vs. Control

4.3.3.3 Comparing low grade Ta tumours to all other samples

This model of comparison (Table 4.8) was chosen for discrimination of samples from patients with lower risk of progression and recurrence from the ones with higher risk.

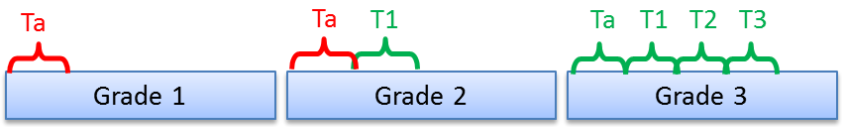

Figure 4.13 Comparison of the low grade Ta tumours to every other sample

Using **Expression**[E], 157 proteins were over-expressed in non-muscle invasive samples while only 4 proteins were down-regulated and 15 proteins showed no significant change (Table 4.8). **Progenesis LC-MS** showed almost the same pattern where 90 proteins were up-regulated and 9 proteins down-regulated. In contrast, expression analysis by either **Scaffold** or **MASCOT** showed over-expression in fewer proteins and no changes was reported for 139 and 86 proteins in Scaffold and MASCOT respectively.

	TaG1-TaG2 > all	Un-changed	TaG1-TaG2 < all
Expression[E]	157	15	4
Progenesis LC-MS	90	-	9
PLGS-Scaffold	7	139	20
MASCOT-Scaffold	16	86	27

Table 4.8 Number of proteins with an alteration in their expression analysis identified using different package when comparing low grade Ta with every other sample. Complete list is provided in appendix A, tables A-5 & A-6.

Comparing the protein lists generated by every one of the above mentioned packages, up-regulation in 4 proteins was seen by 3 packages (Table 4.9).

Picked up in 3 packages	
C reactive protein	Ig lambda 2 chain C regions
Properdin	Hemoglobin subunit alpha

Table 4.9 Up-regulation in the protein expression seen by 3 of the packages when comparing low grade Ta with every other sample

4.3.3.4 Comparing TaT1 to T2T3 in grade 3

Comparison of the disease samples where the patients have tumours in the early stages to the ones with advanced stage tumours may enable clinicians to target more aggressive treatment plans towards NMID high grade diseases (Figure 4.14).

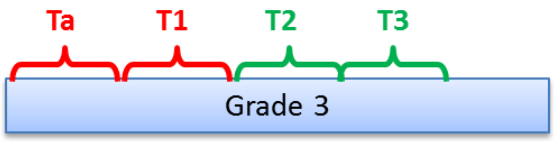

Figure 4.14 Comparison of stage TaT1 to T2T3 in the samples with tumour in grade 3

Samples in the stage TaT1 and T2T3 were compared and **Expression**[E] generated a list of 17 over-expressed proteins and 118 proteins down-regulated proteins (Table 4.10). 19 proteins were also showed no significant changes between the groups. **Progenesis LC-MS**, on the other hand reported up-regulation in 8 proteins and down-regulation in only 1 protein. Expression analysis by **Scaffold** showed 23 and 19 proteins up and down-regulated, respectively. Expression level of 121 proteins was not changed. 48 up-regulated, 24 down-regulated and 83 proteins with no change were seen by **MASCOT/Scaffold**.

	TaT1 > T2T3	Un-changed	TaT1 < T2T3
Expression[E]	17	19	118
Progenesis LC-MS	8	---	1
PLGS-Scaffold	23	121	19
MASCOT-Scaffold	48	83	24

Table 4.10 Number of proteins with an alteration in their expression analysis identified using different package when comparing TaT1 to T2T3 in grade 3. Complete list is provided in appendix A, tables A-7 & A-8.

3 proteins were common to 3 packages and 24 proteins were seen in at least 2 packages (Table 4.11).

Picked up by 3 packages		Picked up by 2 packages
Haptoglobin	Afamin	Complement component C6
Complement factor H related protein 2	Alpha 2 macroglobulin	Complement component C7
Ig gamma 1 chain C region	Apolipoprotein B 100	Complement factor B
	Beta 2 microglobulin	Corticosteroid binding globulin
	Biotinidase	Hyaluronan binding protein 2
	Cholinesterase	Monocyte differentiation antigen CD14
	Plasma kallikrein	Coagulation factor IX
	Coagulation factor X	Serum paraoxonase arylesterase 1
	Tetranectin	Complement C1q subcomponent subunit A
	Transthyretin	Complement C1q subcomponent subunit B
	Properdin	Complement C1q subcomponent subunit C
	Complement C2	Vitamin K dependent protein S

Table 4.11 Up-regulation in the protein expression seen by 3 and 2 of the packages when comparing TaT1 to T2T3 in grade 3

4.3.4 Unique proteins

Each of the subsequent four models of comparison between the sample groups generated unique proteins. The number of unique proteins for each comparison is shown in Table 4.12.

	Number of unique proteins		
	Group 1	vs.	Group 2
Disease vs. Control	6		7
Disease vs. Healthy Control	3		14
TaG1-TaG2 vs. Everything else	13		0
TaG3 - T1G3 vs. T2G3 - T3G3	0		0

Table 4.12 Number of proteins uniquely identified by PLGS and MASCOT

4.4 Discussion

Approximately, 70% of bladder cancer patients at diagnosis have early stage non-muscle invasive disease (Orenes-Pinero *et al.*, 2010) with high rate of recurrence (50-70%) post treatment. Around 20-30% of the recurred tumours progress to muscle invasive disease with high rate of metastasis (~80%) and poor prognosis (Grignon, 2009). The high rate of recurrence and the current treatment strategies for patients demonstrates the necessity for biomarkers of diagnosis and progression of bladder cancer. This ultimately will give clinicians the ability to predict future outcomes of the disease for the patients at the time of diagnosis and potentially enables them to provide a personalised treatment plan for different groups.

Proteomics profiling of human samples is becoming an accepted technique among researchers in the field. Recent advances in the instrumentation technologies (such as, incorporating ion mobility into the mass spectrometer (Giles *et al.*, 2011)) enable accurate and quantitative investigation of samples that ultimately result in a better understanding of biological variations between and within the groups of clinical. Whilst these advances have advantages, their complication can also be a potential source of error (Mann, 2009). To avoid such a situation, QC samples were used in the analysis for successful assessment of consistency of (*i*) retention time of the whole digested sample, (*ii*) retention time of a single ion, (*iii*) intensity, (*iv*) peak shape and (*v*) peak width of the extracted ion and (*vi*) total number of proteins in every run. Alignment of QC samples is indicative of consistent MS performance and consequently confidence in the data is achieved.

To our knowledge, this is the first report investigating the biomarkers of bladder cancer incorporating label-free quantitation and UPLC-IMS-DIA-MS^E method. The panels of overexpressed proteins in bladder cancer samples were mostly in accordance with the current literature.

Afamin, also called alpha-albumin, is the fourth member of the albumin family (albumin, alpha fetoprotein and vitamin D binding protein) discovered by Blanger and Lichenstein in 1994 (Belanger *et al.*, 1994; Lichenstein *et al.*, 1994). Afamin is a plasma protein also known as vitamin E binding protein (Jerkovic *et al.*, 2005) due to its role in vitamin E transportation when there is lack of lipoproteins (Voegele *et al.*, 2002). Significant down-regulation of afamin has been reported in hepatocellular carcinoma (Wu *et al.*, 2000). Similar results has been observed in another study (Jackson *et al.*, 2007) where serum levels of afamin in 57 ovarian cancer patients were significantly decreased when compared to 39 healthy controls. Dieplinger and his co-workers also confirmed this decrease in a larger cohort of 181 ovarian cancer patients and 177 healthy controls (Dieplinger *et al.*, 2009). Unlike ovarian cancer, afamin concentration has been up-regulated (Jeong *et al.*, 2008) in proteomics analysis of plasma in patient with squamous cell carcinoma of cervix. In this study, we found statistically significant up-regulation of afamin in plasma samples of patients with bladder cancer compared to the control group which is in agreement with another study performed on the urine samples obtained from bladder cancer patients (Chen *et al.*, 2012). They used liquid chromatography followed by SRM-MS methods to analyse urine samples of 76 bladder cancer to 57 controls. Over expression of afamin in urine and plasma samples of

patients with bladder cancer points out the potential of this protein as a candidate biomarker for bladder cancer.

Alpha-1-B glycoprotein (A1BG) is a plasma protein with average concentration of 22 mg/dl in normal adult plasma. Although its complete sequence of amino acid was determined in 1986 (Ishioka *et al.*, 1986), to date no known biological function is found for it. A1BG is over expressed in follicular fluid samples from patient with polycystic ovary syndrome when the samples were analysed using both nano-LC MS/MS and MALDI-TOF/MS and confirmed by western blot analysis (Kim *et al.*, 2013). Serum level of A1BG was reported to increase in patients with cervical cancer (Abdul-Rahman *et al.*, 2007; Jeong *et al.*, 2008) and breast cancer (Zeng *et al.*, 2011). Increased level of A1BG was also seen in pancreatic cancer tissue when analysed by DIGE (difference gel electrophoresis) followed by MS/MS and confirmed by western blot (Tian *et al.*, 2008). A1BG is also reported as an inflammatory marker in both serum (Baum *et al.*, 2008) and urine (Goo *et al.*, 2010) samples. In the current project we found significant up-regulation in the level of A1BG in plasma samples of bladder cancer patients compared to controls. The overexpression was also seen in comparison of disease vs. healthy controls samples. In another study using urine samples of bladder cancer patients Kreunin *et al.* (2007) found similar alteration in the expression of A1BG using nano-LC MS/MS in a relatively small study (5 bladder cancer samples vs. 3 noncancerous samples and 2 healthy controls) (Kreunin *et al.*, 2007). Again finding over-expression in the levels of Alpha 1 B Glycoprotein in both plasma samples (our study) and urine samples (Kreunin *et al.*, 2007) substantiates a possible link between this protein and bladder cancer.

Apolipoprotein A-1 (Apo-A1) forms about 2/3 of high density protein (HDL) in plasma. It is synthesised in the liver and intestines. The main function of Apo-A1 is in the reverse transport of cholesterol to the liver where it can be excreted (Grundy & Vega, 1990). Apo-A1 is known as a negative marker for inflammation due to its down-regulation during inflammation (Li *et al.*, 2011; Lei *et al.*, 2012). Alteration in the expression level of Apo-A1 in relation to cancer has been investigated by different researchers. Apo-A1 is mostly reported as a positive marker for cancer except in pancreatic cancer (Ehmann *et al.*, 2007). Using SELDI-TOF mass spectrometry methods Ehmann was able to report a decrease (2-fold) in the expression levels of Apo-A1 in serum samples of 96 patients with pancreatic cancer. In contrast, overexpression in plasma level of Apo-A1 has been reported in breast cancer (Melvin *et al.*, 2012). In another study, a group of researchers in Taiwan (Tung *et al.*, 2013) reported overexpression of Apo-A1 in patients with betel nut chewing disease and its correlation with oral squamous cell carcinoma. Using MALDI-TOF mass spectrometry, Giusti *et al.* (2008) published their findings on the up-regulation of Apo-A1 in patients with thyroid cancer (Giusti *et al.*, 2008). Their results were confirmed by western blot analysis. Analysis of urine samples obtained from patients with bladder cancer in comparison to control samples using MALDI-TOF mass spectrometry, reported Apo-A1 as a positive marker by two independent groups of researchers (Li *et al.*, 2011; Lei *et al.*, 2012). Their results were similar to another research lab where they analysed urine samples of 23 bladder cancer patients compared to 14 controls and found an increase in the level of Apo-A1 (Chen *et al.*, 2010). Two years after the first results, Chen *et al.* (2012) were able to confirm their findings using LC-SRM/MS methods on a larger group of samples (76 bladder cancer vs. 57 controls) (Chen *et al.*, 2012). In the current study,

we showed over expression in the level of Apo-A1 in plasma samples of bladder cancer cohort compared to the healthy controls. Similar alterations in the expression level was also seen when different stages of high grade bladder tumour were investigated.

Haptoglobin is a plasma glycoprotein synthesised in the liver. It binds to free haemoglobin and transports it to the liver to recycle its iron and amino acids in the liver (Tabassum *et al.*, 2012). Its concentration in plasma is increased as a result of inflammation (thus known as an inflammation marker) (Dobryszycka, 1997). Several research groups have reported up-regulation of Haptoglobin in serum samples of patients with different types of cancer compared to controls. Tsai *et al.* compared 45 lung cancer samples to 26 healthy controls using MALDI-TOF mass spectrometry method (Tsai *et al.*, 2011). Other groups also published similar results for small cell lung cancer using immunoassay methods (Shah *et al.*, 2010; Kossowska *et al.*, 2005). In another study performed on 38 ovarian cancers vs. 8 controls using MALDI TOF/MS methods, Ahmed *et al.* showed an increase in the serum level of Haptoglobin (Ahmed *et al.*, 2004). Similar results have been published for breast cancer (Tabassum *et al.*, 2012), hepatocellular carcinoma (Ang *et al.*, 2006) and prostate cancer (Fujimura *et al.*, 2008). Renal injury also leads to overexpression of haptoglobin in the patient's urine sample (Vanhoutte *et al.*, 2007). Using urine samples and iTRAQ LC-SRM/MS method Chen *et al.* showed an association between haptoglobin and bladder cancer (Chen *et al.*, 2012). In the current study we have shown similar correlation between bladder cancer and controls but within plasma samples. These results are also comparable to other findings that used similar methods of analysis (Kreunin *et al.*, 2007).

Fibronectin is a structural glycoprotein playing key roles in migration, cell adhesion and differentiation (Pankov & Yamada, 2002). It is widely present in plasma, cell and extracellular tissue matrix and is known as an abundant soluble protein in plasma. Fibronectin has a critical role in the procedure of wound healing where, along with fibrin, it forms a blood clot that stops bleeding (Grinnell, 1984). Overexpression of fibronectin has been found in non-small cell lung carcinoma (Han *et al.*, 2006). Using 60 plasma samples from patients with renal cell carcinoma (localised = 40 and metastatic = 20) and comparing them with 50 non-malignant renal diseases as control, an increase in the expression of fibronectin was reported (Hegele *et al.*, 2004). The role of fibronectin in bladder cancer has also been investigated. Laufer *et al.* (1999) found an increase in the expression of fibronectin in the urine samples obtained from patients with non-muscle invasive disease post TURBT (Laufer *et al.*, 1999). In another study carried out on 68 urine samples of bladder cancer patients (confirmed by transurethral resection), 10 noncancerous urologic disease samples and 45 healthy controls, the sensitivity of bladder tumour fibronectin was higher (75% vs 55%) in comparison to cytology (Menendez *et al.*, 2005). A similar investigation has been performed on a relatively larger cohort (bladder cancer = 100, noncancerous urological disorders = 93 and 47 healthy controls) (Eissa *et al.*, 2010) and found similar results to Menendez and co-workers (82% for bladder tumour fibronectin vs. 59% for urine cytology). Our study showed overexpression of fibronectin in plasma samples obtained from bladder cancer patients when compared to healthy controls.

Inter alpha trypsin inhibitor heavy chain 3 (ITI-H3) is a member of the large family of inter-alpha trypsin inhibitor proteins. They were first discovered in serum and urine

samples due to their inhibitory activity against trypsin (Zhuo & Kimata, 2008). Functioning as protease inhibitors, they are formed from a light chain and up to four heavy chains. The link between ITI-H3 and cancer has been studied by a number of researchers. Using plasma samples from mouse models of gastric cancer and iTRAQ LC-MS/MS methods, Chong *et al.* showed an increase in the levels of ITI-H3 compared to controls (Chong *et al.*, 2010). In a study similar to this, 3 serum samples from lung cancer patients were compared to 3 healthy control samples (Heo *et al.*, 2007). Nano-LC coupled to ESI-MS/MS was used for separation and analysis of the samples where they found overexpression of ITI-H3 in patient's samples. Using urine samples from bladder cancer patients and MRM-MS methods, overexpression in a list of 63 proteins including inter alpha trypsin inhibitor heavy chain were observed (Chen *et al.*, 2012). We have found up-regulation in ITI-H3 in plasma samples of bladder cancer patients compared to healthy controls.

Plasma Kallikrein (also called Fletcher factor) is a glycoprotein primarily synthesised and secreted by the liver with roles in blood coagulation, fibrinolysis and inflammation. As a serine protease it helps in the formation of kininogen peptides that plays a crucial role in the regulation of blood pressure and activation of inflammation (Bhoola *et al.*, 1992). Using similar strategies as used in this study (nano-LC coupled to ESI-MS/MS) an increase in the expression of Kallikrein was observed when protein profiles of three serum samples from patients with lung cancer were compared to three controls (Heo *et al.*, 2007). The alteration in the expression was confirmed by western blot analysis. In the current study, where we compared the protein profiles of a larger cohort of

samples (18 diseases vs. 12 controls), similar increase in the expression of plasma kallikrein was observed.

C-reactive protein (CRP) is a plasma protein synthesised primarily in the liver in response to stimulants, such as interleukin 6 (Erlinger *et al.*, 2004). Its concentration is greatly increased in response to inflammation for which CRP is known as a marker of inflammation (Allin & Nordestgaard, 2011). It has been reported as the first positive acute-phase protein in the acute-phase response (Gabay & Kushner, 1999) during which the synthesis of several plasma proteins increases (positive) or decreases (negative) in response to stimulants (i.e. IL6) originating from an inflammation, infection or tissue damage (Allin & Nordestgaard, 2011). The risk of developing cancer is higher in inflammatory sites (Coussens & Werb, 2002) and cancer related inflammation has been suggested as the seventh hallmark of cancer (Colotta *et al.*, 2009). In an epidemiologic study carried out in Greece including 496 cancer cases and 996 controls the mean CRP level was 4.1 mg/L in cancer cases compared to 2.6 mg/L in controls (Trichopoulos *et al.*, 2006). They found a stronger association between higher levels of CRP and cancers of kidney, lung, bladder and liver compared to other types of cancer. In a similar study (Erlinger *et al.*, 2004), increased levels of CRP was observed in colorectal cancer cases whereas a year after that Zhang *et al.* reported no association between colorectal cancer and increased level of CRP (Zhang *et al.*, 2005). The latter group also reported no association between elevated levels of CRP and breast cancer cases in a relatively large cohort (892 cancer cases) (Zhang *et al.*, 2007). The expression level of preoperative CRP in relation to prognosis and survival rate has also been investigated. Woo *et al.* (2012) reported that the high levels of preoperative CRP could be used as an indication for

recurrence in gastric cancer patients (Woo *et al.*, 2012). In a similar study no correlation between the levels of preoperative CRP and prognosis was found in breast cancer patients (Ravishankaran & Karunanithi, 2011). It has been shown that increased level of CRP pre-operation can be used as a predictive marker of poor survival rate (Steffens *et al.*, 2012) and prognosis of advanced metastasis (Ito *et al.*, 2012) in patients with renal cell cancer. Similar circumstances have been reported in muscle invasive bladder cancer patients treated with chemotherapy (Yoshida *et al.*, 2008). In our study, CRP overexpression has been observed when low grade Ta tumours were compared to every other sample across the study.

Kininogen 1 is one of the main protein members of the blood coagulation system. There are two types of kininogen namely high molecular weight kininogen (HMWK) produced in the liver and has important roles in pathophysiological processes i.e. coagulation and inflammation (Bhoola *et al.*, 1992). Low molecular weight kininogen (LMWK) is the other type which is synthesised locally by different tissues and secreted by tissue kalikrein (Bryant & Shariat-Madar, 2009). Diverse expression of kininogen has been observed in different cancers. Down-regulation of kininogen in plasma was observed in the patients with cervical cancer (Abdul-Rahman *et al.*, 2007) and also in urine sample of patients with ovarian cancer (Abdullah-Soheimi *et al.*, 2010). In contrast, kininogen was overexpressed in serum samples of patients with breast cancer (Schaub *et al.*, 2009) and gastric cancer (Liu *et al.*, 2012) using similar strategies (MALDI TOF/MS) for sample analysis. Our findings in this project are in agreement with the latter groups due to the up-regulation of kininogen in disease samples compared to controls.

Vitamin D binding protein (VDBP), also known as group-specific component (Gc-globulin), is a vitamin D transporter in circulation system where it binds to vitamin D and transports it to the target tissue (Gomme & Bertolini, 2004). The primary circulating form of vitamin D is 25-hydroxyvitamin D (25(OH)D) which activates to 1-25-dihydroxyvitamin D that has been shown to play inhibitory roles in proliferation, angiogenesis and metastasis (Holick, 2005). There is evidence of VDBP being activated in response to inflammation in urine sample (Mirkovic *et al.*, 2013). It has also a contributory role to inflammatory bowel disease pathogenesis (Eloranta *et al.*, 2011). Expression levels of VDBP have been shown to be inversely correlated with the risk of pancreatic cancer (Weinstein *et al.*, 2012). In an epidemiologic study on 250 bladder cancer cases (male, smoker), Mondul *et al.* (2010) reported an inverse correlation between the risk of developing bladder cancer and levels of 25(OH)D over 20 years (Mondul *et al.*, 2010). Two years after that, the same group repeated similar analysis by incorporating women and non-smoker cases into the study (Mondul *et al.*, 2012) in which they found no evidence of association between the risk and 25(OH)D. Overexpression in the levels of vitamin D binding protein was observed in bladder cancer samples compared to controls in our study which is not in agreement with the epidemiologic studies on bladder cancer.

Zinc alpha 2 glycoprotein (ZAG) is a 41kDa glycoprotein secreted by a variety of epithelia into body fluids. ZAG is also called lipid mobilising factor due to its involvement in fatty acid depletion from adipose tissue (Bing *et al.*, 2004). The expression level of ZAG has been shown to decrease in tumour tissues of patients with hepatocellular carcinoma (Huang *et al.*, 2012) whereas it's increased in plasma samples

of pancreatic cancer patients (Pan *et al.*, 2011). Up-regulation in the expression level ZAG has also been reported in serum samples of patients with cervical cancer (Abdul-Rahman *et al.*, 2007) and prostate cancer (Hale *et al.*, 2001). Using 2DiGE and MALDI TOF mass spectrometry Hassan *et al.* reported elevated expression of ZAG in prostate cancer seminal fluid (Hassan *et al.*, 2008). A similar pattern was observed in urine samples of patients with bladder cancer (Irmak *et al.*, 2005). Irmak *et al.* were also able to show that the expression of ZAG was much higher in the tumour cells at the bladder tumour invasion front. In this study, we observed overexpression of ZAG in bladder cancer patients compared to controls. Also a similar expression pattern was observed when plasma samples of early stage Ta and T1 were compared that showed higher levels in the Ta samples.

The eleven proteins discussed above were chosen as key representative proteins because of their role in cancer and particularly in bladder cancer (Table 4.13).

	Protein name	Role in cancer	Role in bladder cancer	Our findings
1	Afamin	↑ and ↓	↑ in urine	↑ in plasma ❶
2	Alpha1-B Glycoprotein	↑	↑ in urine	↑ in plasma ❶
3	Apolipoprotein A1	↑ and ↓	↑ in urine	↑ in plasma ❶
4	Haptoglobin	↑	↑ in urine	↑ in plasma ❶
5	Fibronectin	↑	↑ in urine	↑ in plasma ❶
6	Inter alpha trypsin inhibitor H3	↑	↑ in urine	↑ in plasma ❷
7	Zinc alpha2 Glycoprotein	↑ and ↓	↑ in urine	↑ in plasma ❷
8	Plasma Kallikrein	↑	unknown	↑ in plasma ❷
9	Kininogen 1	↑ and ↓	unknown	↑ in plasma ❷
10	C-reactive protein	↑	↑ post CHT	↑ in plasma ❷
11	Vitamin D binding protein	↑ and ↓	unknown	↑ in plasma ❷

Table 4.13 The 11 representative proteins selected by 4 (❶) and 3 (❷) expression analysis packages. ↑ shows up-regulation and ↓ shows down-regulation in the protein expression. CHT is chemotherapy and where no literature was found it is unknown

Reviewing the available current literature to identify a link between all of the over-expressed proteins identified (approximately 500) in patient samples (tables 6, 8, 10, 12 and 14) is a task that is beyond the scope of this project. However, more advanced informatics using DAVID analysis or pathway analysis would of helped triage the selection of proteins. These approaches can be considered for future work.

To summarise, although none of the molecular markers of bladder cancer (see Chapter 1, section 1.2.5) were identified in this study, the ones found here are reflected in the literature. Optimisation of the methodology helped in the thorough investigation of the plasma proteome (although the low abundant proteins are still not reached)

leading to a coverage of up to five orders of magnitude of plasma dynamic range. Most of the markers stated above are not tumour specific but maybe used collectively to diagnose bladder cancer. In total, 11 proteins were found (Table 4.7) as possible markers of diagnosis for bladder cancer. Four of the markers namely, afamin, alpha 1-B glycoprotein, apolipoprotein A1 and haptoglobin were found in plasma and are also reported to be urinary markers of bladder cancer. CRP was found overexpressed in MID high grade tumours when compared to NMID high grade ones. These results could be used prognostically and therapy could be tailored accordingly.

Similar to every other clinical study using proteomic strategies, there were limitations to this study as well. The limitation of the instrument's dynamic range coverage, which is not more than 5 orders of magnitude, compared to the plasma proteome dynamic range (of up to 12 orders of magnitude) inhibits identification of low abundant proteins, where most of the biomarkers are known to be. The other limitation was variety of disease plasma samples. Having equal number of samples in every group and sub-groups of the experimental design increases the reliability of the findings of the study. In this study an attempt has been made to group the samples according to grade and stage but still, unequal numbers in the sub-groups could be considered as a limitation to the study.

Chapter Five

VERIFICATION

5 Verification of candidate biomarkers

5.1 Introduction

The next step in the biomarker discovery workflow (figure 1.26, Chapter 1) is to verify the candidate biomarkers identified in the discovery phase (Rifai *et al.*, 2006). Although more than 1,200 proteins has been identified and reported as candidate biomarkers of different type of cancers, on average only 3 proteins over 2 years have received FDA approval to be used as disease biomarkers (Anderson, 2010). To date, a large number of proteins have been discovered and reported as proteomics-based candidate biomarkers, especially in cancer research, but only two of them (VeriStrat and OVA1) have been tested in clinical studies and ultimately received FDA approval to be used as a clinical test. The reason for this is the laborious and complicated route towards validation and approval of a candidate biomarker to be used in the clinic (Gutman & Kessler, 2006). A variety of methods are being used for verification of candidate biomarkers, namely immunoassays (i.e. ELISA, western blot and immunohistochemistry) and mass spectrometry methods (i.e. SRM). While both of these methods have advantages and disadvantages, selection of the method to be used mainly depends on time restriction and instrumentation availability. In immunoassays, highly specific antibodies, which is not always commercially available or very costly, are required for every single candidate biomarker. Verification of candidate biomarkers using SRM assay has higher sensitivity and specificity while development of the SRM method is time consuming. Western blot, as one of the most popular methods of verification in proteomics studies (Salasznyk *et al.*, 2005; Suzuki *et al.*, 2011; Yap *et al.*, 2011), was selected to confirm alteration in the expression level of proteins between the groups. Western blot was selected because of two reasons: (*i*) to

use a non-mass spectrometry method for confirmation of the results found using a mass spectrometry method and (*ii*) the time needed to develop and evaluate the SRM method was beyond the scope of this project. So the main aim of this section of the project was to validate/verify the results obtained from one of the experiments (for this we chose the 'unique proteins' identified in section 4.2.5) using western blot analysis.

5.2 Materials and methods

5.2.1 Materials and reagents

All reagents and buffers were bought from Sigma-Aldrich (Poole, UK) or Fisher-Scientific Inc. (Loughborough, UK) unless otherwise stated. Concentrated buffers (gel preparation buffers at 4X and running and transfer buffers at 10X) and ECL (Enhanced Chemiluminescence) detection reagents were bought from Geneflow Ltd. (Lichfield, Staffs, UK). Anti-rabbit polyclonal antibody H-25 (Santa Cruz Biotechnology Inc., CA, USA) was used to detect plasma kallikrein, anti-mouse monoclonal antibody D10/4 (Santa Cruz Biotechnology Inc.) was used to detect complement factor D, anti-mouse monoclonal antibody A-11 (Santa Cruz Biotechnology Inc.) was used to detect fibronectin, anti-mouse monoclonal antibody 13B9 (Santa Cruz Biotechnology Inc.) was used to detect coagulation factor IX, anti-mouse antibody 12C8 (Enzo life sciences Ltd., Exeter, UK) was used to detect apolipoprotein A-1 and anti-rabbit antibody Anti-ITIH3 (Sigma-Aldrich) was used to detect inter alpha trypsin inhibitor heavy chain H3.

5.2.2 Candidate biomarkers selection

The unique proteins identified only in bladder cancer samples, deduced as illustrated in Figure 5.1, were selected for verification experiments based on the certainty of the proteins uniqueness and availability of the antibodies locally; for detailed descriptions see Chapter 4, section 4.2.5.

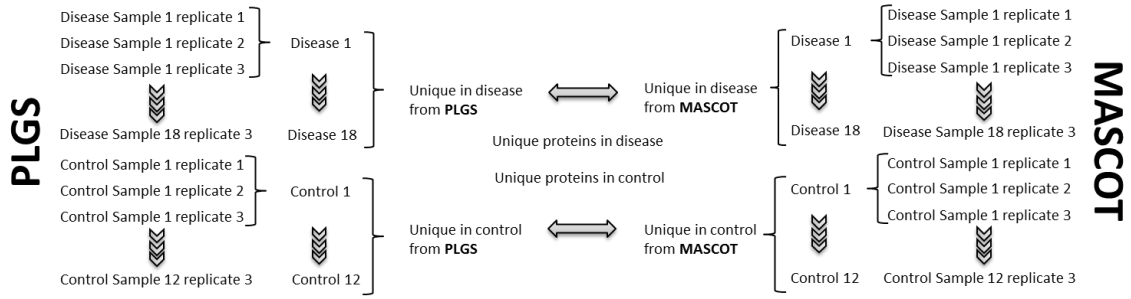

Figure 5.1 Schematic plan designed for selection of the unique proteins. For detailed description see Chapter 4, figure 4.3.

Six proteins (Table 5.1) were uniquely identified in the diseased samples. These proteins were used for western blot verification.

Unique proteins in disease samples			
1	Apolipoprotein A I	**4**	Fibronectin
2	Coagulation factor IX	**5**	Inter alpha trypsin inhibitor heavy chain H3
3	Complement factor D	**6**	Plasma kallikrein

Table 5.1 Unique proteins identified using both PLGS and MASCOT in bladder cancer samples

5.2.3 Sample information

Twelve representative samples (un-depleted/un-pooled) including 7 'disease' plasma samples from bladder cancer patients and 5 plasma samples from healthy control group were selected randomly for the verification experiment. A panel of 6 commercially available antibodies were used for the six candidate proteins (Table 5.1).

5.2.4 Gel preparation

A 10% gel for protein separation and a 5% gel for loading of the proteins were prepared based on the recipe in Table 5.2.

Order of adding	Substance (ml)	10% resolving gel (ml)	5% stacking gel (ml)
1	dH2O	8	6.9
2	Buffer	5	1.25
3	Acrylamide (30%)	6.7	1.7
4	AMPS (10%)	0.2	0.1
5	TEMED	0.015	0.015
Total volume		**~20 ml**	**~10 ml**

Table 5.2 Recipe for preparation of gel

The 10% gel was gently mixed and added to occupy approximately three quarters of the total plate volume and left to set for approximately 30min. The 5% gel was then added on top of the 10% gel; the comb was inserted and left to set for about an hour.

5.2.5 Sample running

Concentration of proteins in the samples was assessed using BCA protein assay which enabled loading of equal amount of protein (20µg) in the wells. Proteins were denatured at 100˚C for 5 minutes followed by cooling down on ice. 10µl of each sample was mixed with 10µl of sample loading buffer (Laemmli buffer, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue and 0.125 M Tris HCl) and 20 µl loaded per lane with two markers in either side of the samples (rainbow marker to check the transfer efficiency and secondary antibody marker for the molecular weights). The gel electrophoresis was run in running buffer (25mM Tris, 250m Glycine, 0.1% SDS) at a constant voltage of 120V (constant amp of 40mA) until the dye front was about 10mm from the end of the gel (approximately 60 minutes).

5.2.6 Transferring

The transfer cassette was assembled based on Figure 5.2 and proteins were transferred to the membrane at 100V power for 90 minutes in transfer buffer (47mM Tris, 37mM Glycine, 20% MeOH, 0.04% SDS).

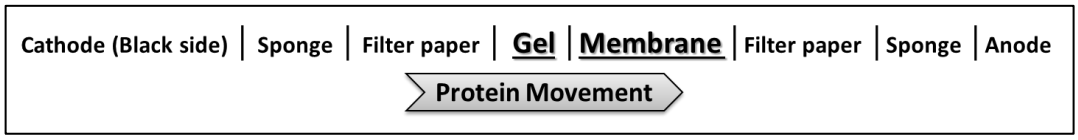

Figure 5.2 Transfer sandwich assembly

The transfer efficiency was assessed using Ponceaus S stain followed by washing the stain off the membrane with wash buffer (PBS + 0.1% TWEEN20) (3x5min) on a shaker.

5.2.7 Detection

Blocking buffer (2.5g milk powder into 50ml of PBS + 0.1% TWEEN20) was used to inhibit non-specific protein binding by incubation of the membrane in blocking buffer overnight at 4˚C. The excess buffer was discarded and the membrane was washed for 5min using the wash buffer. Primary antibody was diluted (1:1000) in 3% milk powder solution, added to the membrane and the membrane was incubated for 90min at room temperature on a shaker. The antibody solution was discarded and the membrane was washed for 20min (1x10min and 2x5min) on the shaker. Secondary antibody was diluted (1:20000) in 3% milk powder, added to the membrane and incubated for 1 hour at room temperature. The excess antibody solution was removed and the membrane was washed for 15min (1x10min and 1x5min) on the shaker. The TWEEN in the wash buffer was washed away using dH2O for 5min.

5.2.8 Analysis

The membrane was then incubated in ECL reagent for 5 minutes at room temperature. The ECL reagent was removed and the membrane was incubated in detection reagent for 2min. The excess reagent was removed; the membrane was wrapped in cling film and placed into an AutoRad cassette. In dark room the membrane was placed on a film for between 30 seconds to 1 minute before developing in the automated developer.

5.3 Results

Six candidate proteins were selected for the verification experiment using 12 undepleted plasma samples. The first five wells from the left (after the rainbow marker) were used for the control plasma samples and the next seven wells for the bladder cancer plasma samples followed by a positive control and a molecular weight marker. Apolipoprotein A1 gel showed similar level of binding for diseases and control samples at ~24kDa (Figure 5.3).

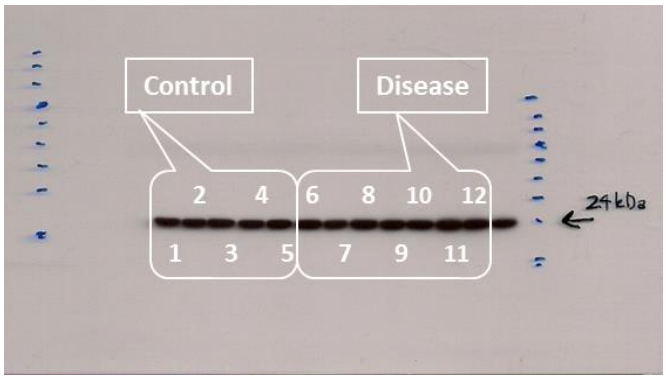

Figure 5.3 Protein bands for Apolipoprotein A1. Anti-mouse Apolipoprotein A1 antibody shows a band at ~24kDa.

Fibronectin with molecular weight of 250kDa showed a band at the relevant area (Figure 5.4). Despite possible overexpression visible when comparing diseases to controls, statistical analysis post densitometry showed no significant differences between them.

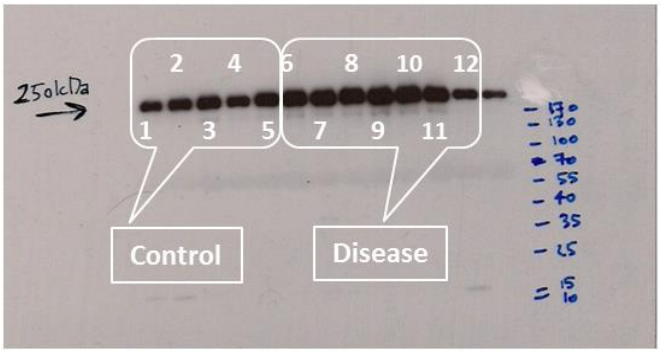

Figure 5.4 Protein bands for Fibronectin. Anti-mouse Fibronectin antibody shows a band at ~250kDa.

Human complement factor D (also known as Adipsin) also showed a weak band at 24kDa (Figure 5.5) but again similar level of expression in both diseases and control samples.

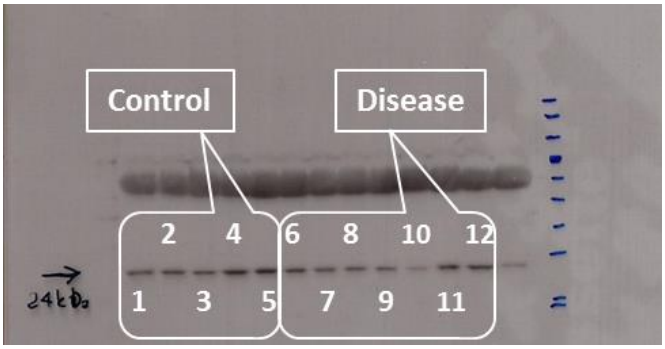

Figure 5.5 Protein bands for Complement factor D. Anti-mouse Complement factor D antibody shows a band at 24kDa.

Despite many attempts and due to some unpredictable/unknown complications we were unable to get results for the other three protein candidate markers used for the verification analysis.

Densitometry analysis of the developed films is shown in Figure 5.6 for the three candidate biomarkers when disease samples were compared to controls.

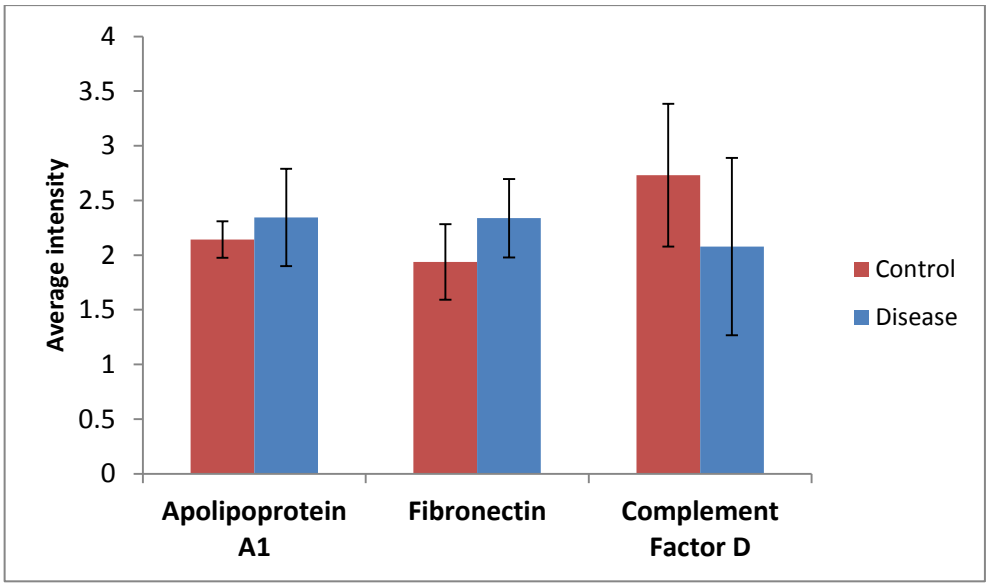

Figure 5.6 Average intensities of three candidate proteins in disease and control samples

Statistical analysis of the values for the protein bands demonstrated no significant differences in the expression of any of the three proteins (Table 5.3).

Candidate protein name	p-Value
Apolipoprotein A1	0.300
Fibronectin	0.084
Complement factor D	0.155

Table 5.3 p-Values for comparison of protein intensities when disease samples were compared to controls

5.4 Discussion

The number of potential biomarkers identified from discovery experiments is consistently increasing which, in time, necessitates validation and verification of these candidate markers (Anderson, 2010). The alteration in the expression level of putative biomarkers should be verified in a smaller number of samples and then the verified candidates assessed in clinical trials involving a large sample cohort of disease and control samples (Rifai *et al.*, 2006). Western blot is one of the most developed and used immunoassays methods used for validation and verification of the candidate biomarkers identified in the discovery phase of proteomics studies (Salasznyk *et al.*, 2005; Suzuki *et al.*, 2011). In the present study, a number of proteins were identified in the discovery phase of the project by investigating 30 pooled plasma samples using a developed and optimized proteomics workflow, including label-free quantitation of plasma proteins using UPLC-HDMS^E strategies. Incorporating a rigorous data analysis pattern using four different data analysis packages into the workflow (see Chapter 4, section 4.2.4) enabled us to find potential plasma biomarkers of bladder cancer. Also, by using an intense statistical model of analysis (see Chapter 4, section 4.2.5) we were able to extract a number of proteins identified uniquely in one group of samples, i.e. diseases or controls. Due to insufficient time to analyse all of the unique proteins in the different groups, only the proteins identified uniquely in all replicates of a sample by two different data analysis packages were selected for the verification experiment using western blot. In this study, we were not able to validate a similar pattern for candidate markers as what we found in the discovery phase of the project. This discrepancy could be accounted for as follows:

For the western blot analysis, un-pooled, undepleted plasma samples were used for the verification experiments whereas the ones used in the discovery phase of the study were pools of three individual samples. It was rationalised that the ideal disease biomarker should be assayable in the authentic clinical sample and consequently un-pooled/un-depleted samples were used (Rifai *et al.*, 2006).

The mechanism of protein identification in PLGS and western blot is different. In PLGS the peptide sequence is searched against a database to find a matched sequence (Li *et al.*, 2009) whereas in western blot, binding of an antibody to the protein of interest transferred onto a membrane produces a band visible on a developed film (Moore, 2009).

Also, there are differences in antibody specificities (Kurien *et al.*, 2011) which lead to binding/unbinding of a protein epitope to the recognition site of an antibody.

Western blot assay is not a true quantitation method as densitometry has limitations. Three main sources of variability can be introduced to the experiment are: (i) the technique to be used for digitising the blots, (ii) any modification in the protein of interest and (iii) the algorithm used to quantify a given band (Gassmann *et al.*, 2009).

The procedure of normalisation in western blot analysis is totally different to the expression analysis used in this project (See Chapter 4, section 4.2.4).

It is feasible that the lack of support for the discovery phase from the verification phase was due to the limitations of WB which is probably more suited to a qualitative assay.

Targeted proteomics application using selected reaction monitoring (SRM) methods has recently become the favourite method of verification in proteomics field (Maiolica *et al.*, 2012; Picotti & Aebersold, 2012) and may be a preferable method over western blot analysis for this project. This can be considered as a future direction to the current project.

Chapter Six

CONCLUSIONS AND FUTURE WORKS

6 Conclusions and future works

Bladder cancer, due to its high incidence (~10,000 new cases in 2010, CRUK), high recurrence and high care costs is one of the most expensive cancers to treat (Botteman *et al.*, 2003; Sangar *et al.*, 2005). To date a significant number of proteomics biomarkers for clinical diagnosis and prognosis of the disease have been investigated but none of them has yet been approved by regulatory authorities as a biomarker for bladder cancer.

Mass spectrometry-based proteomics has developed considerably since the first use of proteomic profiling using mass spectrometry as a diagnostic tool (Petricoin *et al.*, 2002). It has been used for the discovery of cancer biomarkers of breast cancer (Cohen *et al.*, 2013), ovarian cancer (Dieplinger *et al.*, 2009), prostate cancer (Pin *et al.*, 2013) and bladder cancer (Li *et al.*, 2011). In the latter, urine samples are typically used for proteomics approaches. Urine is commonly used as it is in contact with bladder tissue, is non-invasive and is easy to obtain. However, urine is a dynamic volume, it is full of ion suppressing molecules which can interfere with mass spectrometric analysis, it has a large dynamic range in the concentration of molecules and it requires significant pre-analytical workup.

This thesis sought to use plasma as an alternative strategy to evaluate the presence of biomarkers. The primary rationale was plasma-based markers will be potentially longer lived within the systemic circulation. Thus, observation of biomarkers may be achievable due to effective enriching of biomarker concentrations in blood whereas in urine, biomarkers will be immediately excreted and removed from the body. To date,

plasma has not been chosen for proteomics-based biomarker discovery of bladder cancer.

There are a number of steps which introduce limitations to proteomics biomarker discovery workflow (Diamandis, 2010; Diamandis, 2012). An attempt to improve or overcome each of these limitations has been attempted within this thesis.

Firstly, poor experimental study design is a key pre-analytical limitation in proteomics-based clinical biomarker discovery (Diamandis, 2010). In this study a clinically relevant experimental design was established. It was designed to answer four clinically relevant questions post-analysis (see Chapter 4, section 4.1). Using this design we were able to confidently identify overexpression in 11 proteins (candidate biomarkers) that can be used for discrimination of disease and control samples. Four of the candidate biomarkers have already been reported by other researchers to be overexpressed in urine samples of bladder cancer patients. This correlation between urine and plasma is partial vindication for the experimental design as well as the use of plasma. Using the proposed experimental study design to compare between and within the groups of samples, we were able to provide suitable candidates for further biomarker verification/validation studies.

The next challenge in the proteomics workflow is incompatibility of the plasma proteome dynamic range, which spans over 12 orders of magnitude (Anderson & Anderson, 2002), with the typical mass spectrometry detection dynamic range (3 to 5 orders of magnitude). Addition of a fractionation step to deplete up to 95% of high abundant proteins (Chapter 2) renders the plasma proteins' dynamic range more accessible and consequently enables assessment of low abundant proteins, including

disease biomarkers (Schiess *et al.*, 2009). The immunodepletion method (SEPPRO IgY-14) used in the protocol was clearly demonstrated to reliably decrease the plasma proteome dynamic range enabling us to access moderate to low abundant proteins with roles in cancer and other diseases (see Chapter 3, table 3.8).

When it comes to sample analysis, the workflow had to be optimised, evaluated and tested using healthy control samples before using diseased samples, due to the preciousness of the clinical samples (Luque-Garcia & Neubert, 2007). In this study a novel pipeline for proteomics biomarker discovery using mass spectrometry was developed and optimised (Chapter 3). Use of nano-UPLC coupled to ion mobility incorporated into mass spectrometry (Synapt G2 HDMS) for proteomic analysis of plasma samples from bladder cancer patients has not been reported in the literature before. Analysis of the raw data files using a label–free quantitation method (Silva *et al.*, 2006) enabled absolute quantification of proteins in samples that ultimately helps in a more accurate expression analysis.

The ability to produce reliable and reproducible results is another challenge of using mass spectrometry in biomarker discovery. A reliable workflow for proteomics-based biomarker discovery needs to demonstrate suitable reproducibility (Tabb *et al.*, 2010). Using aliquots of a healthy plasma sample, we were able to show that protein profiling of plasma samples using the proposed workflow was both highly precise and highly technically reproducible. It was also demonstrated that by using the optimised workflow, nearly 5 orders of magnitude of the plasma proteome dynamic range can be covered and assessed. This is still less than the actual dynamic range of human plasma

but is better than previously assessed workflows (Hortin & Sviridov, 2010) and allows penetration of the low abundance protein range (see Chapter 1, figure 1.9).

Due to the differences between the plasma protein content of a healthy person and a diseased person (Anderson, 2010), the optimised workflow was further assessed using plasma samples from bladder cancer patients. The amount of identified proteins covered up to 5 orders of magnitude of plasma proteome including 78.31ng (on column) for α-2 Macroglobulin down to 0.005ng (on column) for Plasminogen. This large dynamic range successfully demonstrated the protocol's compatibility for both patient and control samples.

The specific UPLC-HDMS^E label-free quantitation method used in this project uses a data independent acquisition method (DIA) which has many advantages over the traditional analysis method (DDA) (see Chapter 1, figure 1.19). The method also benefits from ion mobility separation which adds another level of separation to the sample's analysis. Reproducible identification and quantification of proteins over a wide range of concentrations using UPLC-IMS-DIA-MS^E method successfully demonstrates the capability of the method to be used in any proteomics based biomarker discovery project. To our knowledge, this is the first study to use this method for proteomics-based biomarker discovery of bladder cancer.

Another limitation in proteomics based biomarker discovery is the narrow and often irrational choice of informatics approaches to identify and quantify proteins found in highly complex samples such as plasma (Diamandis, 2012; News in Proteomics Research. 2013). In the present study in order to overcome the possible bias introduced through using a single informatics package, a combination of different approaches was

assessed. A plan consisting of four data processing packages each of which uses different statistical means for protein identification and quantification was used. Comparisons of these approaches and identifying common candidate proteins within different groups of samples resulted in high confidence with respect to the discovery of moderate to low abundant proteins, most of which play crucial role in inflammation and cancer.

A huge number of proteins were identified either overexpressed or down regulated when different sub-groups of disease samples were compared. The main aim of the study was to follow the patient's journey from presentation at the clinic through to diagnosis, treatment and management of the disease. When bladder cancer samples were compared to healthy controls, apolipoprotein A1 and afamin along with a number of other proteins were found overexpressed in disease sample. Afamin has been reported to have down regulated in hepatocellular carcinoma (Wu *et al.*, 2000) and ovarian cancer (Jackson *et al.*, 2007) and overexpressed in squamous cell carcinoma of cervix (Jeong *et al.*, 2008). Up-regulation of afamin has been shown in urine samples of bladder cancer patients (Chen *et al.*, 2012) and here we demonstrated similar expression in plasma samples. Similarly, apolipoprotein A1 has been shown to have down-regulated in pancreatic cancer (Ehmann *et al.*, 2007) and overexpressed in breast cancer (Melvin *et al.*, 2012) and thyroid cancer (Giusti *et al.*, 2008). Up-regulation of apolipoprotein A1 in urine samples of bladder cancer patients has been reported by various research groups (Li *et al.*, 2011; Chen *et al.*, 2012; Lei *et al.*, 2012; Chen *et al.*, 2010) which is in agreement with the findings in this study using plasma samples. When non-bladder cancer hospitalised sample were added into the control group and compared

to the disease group, afamin and apolipoprotein A1 were confidently picked up by all four packages which confirmed the previous results (question 1), that they are bladder cancer related markers. The observation of overexpression of the above mentioned proteins (apolipoprotein A1 and afamin) in both comparisons (disease vs. healthy controls & disease vs. all controls) and similar reports by other researchers, afamin and apolipoprotein A1 may be counted as potential biomarkers of bladder cancer.

Most bladder cancer patients (80%) are diagnosed with the low grade non-muscle invasive form of the disease at their first presentation at the clinic. The tumour cells in approximately 60% of these patients are still in the mucosa layer and have not invaded into the basement membrane of bladder wall. Low grade Ta tumours have the lowest rate of recurrence and progression and their molecular signature is different from all other stages and grades of the tumour. C-reactive protein (CRP) was found to be overexpressed when low grade Ta tumours were compared to every other group of samples in the study. CRP is known as marker of inflammation (Allin & Nordestgaard, 2011) and has been suggested as to be a positive marker for various cancers including lung, bladder (Trichopoulos *et al.*, 2006) and colorectal cancer (Erlinger *et al.*, 2004). CRP may be used as the marker which enables the discrimination of low grade Ta tumours.

The current protocol for treating early stage high grade (TaT1-G3) tumours is intravesical BCG followed by cystoscopy/cytology every 3 to 6 months which leads to high rate of mortality. Neo-adjuvant chemotherapy followed by radical cystectomy is the current treatment for advanced stage high grade (T2T3-G3) tumours. Knowing in advance the likelihood for future outcome of patients with high risk TaT1 stage tumours will help clinicians target these individuals with more aggressive therapies

thereby reducing future risk of disease. In this study, afamin and haptoglubin were identified overexpressed in T2T3 stage tumours when compared to early stage ones. This finding is potentially beneficial to enable more informative management of therapy which will ultimately benefits patients.

Future directions

Although proteomics-based biomarker discovery using mass spectrometry has significantly improved since its first use (Petricoin *et al.*, 2002), it is still a fairly new technique and will require further improvements to make greater impact into biomarker discovery. A major challenge in the overall technique is the inherent complexity of the samples. It has been demonstrated that fractionation of depleted samples results in increased depth of analysis based on total number of identified proteins (Zhang *et al.*, 2011). Two dimensional high pH low pH reversed phase HPLC fractionation method has been shown to be an attractive fractionation strategy (compared to SDS-PAGE and SCX) when using plasma samples (Cao *et al.*, 2012). Consequently, the addition of a high pH reversed phase fractionation step into the optimised workflow may lead to identification of a greater number of proteins.

The plasma samples used in this study were provided by our collaborators in the University of Torino, Turin, Italy. Due to the nature of their research (epidemiological studies) and also the need to follow up the disease post-treatment, the research group in Turin were able to collect urine samples from the patients already donated their blood for the study. Urine sample collection has already been started and is going on at the moment. Using the new sample set it would be possible to investigate urinary markers of bladder cancer. This would ultimately enable investigation of urine and plasma samples of the same patient and assess the link (if any) between the two samples and bladder cancer.

Due to the lack of plasma samples representative of certain patient groups of the proposed experimental design (see Chapter 4, section 4.2.4) it was not possible to

perform the best comparison between and within the groups. i.e. comparing the stages was only possible among the high grade tumours (G3). Having access to equal number of samples in every different stage and grade (more samples), the ideal experimental design can be planned that enables a deeper investigation of the plasma proteome.

The proteins identified as putative biomarkers must be verified in order to assess the alteration in their expression levels reported by expression analysis. As an alternative to western blot, discussed in Chapter 5, Selected Reaction Monitoring (SRM) can be performed to verify the alteration in protein expression of candidate biomarkers. The SRM protocol needs a huge amount of time for optimisation, evaluation and system adjustments hence is time consuming (this being the reason it was not selected for the verification step of this project). This would be an interesting research venture for the future; it would permit the further evaluation of the candidate biomarkers and ultimately preparation for the clinical studies, as the final stage in the biomarker discovery route.

7 Appendices

7.1 Appendix A

List of Up/Down-regulated proteins identified by different expression analysis packages

Table A-1, Up-regulated proteins in comparison of disease and control

Progenesis
Serum albumin
Complement component C7
Vitamin D binding protein
Serotransferrin
Haptoglobin
Kininogen 1
Afamin
Alpha 1 antitrypsin
Inter alpha trypsin inhibitor heavy chain H3
Apolipoprotein A I
Beta 2 glycoprotein 1
Alpha 1B glycoprotein
Clusterin
Carboxypeptidase N catalytic chain
Alpha 1 acid glycoprotein 1
Hemoglobin subunit beta
Complement factor H related protein 1
CD5 antigen like
Corticosteroid binding globulin
Apolipoprotein E
Tetranectin
Biotinidase
Beta 2 microglobulin
Insulin like growth factor binding protein 3
MASCOT-Scaffold
Immunoglobulin lambda-like polypeptide 5
Complement component C7
Myosin-XVIIIb
Fibrinogen gamma chain
Serum paraoxonase/arylesterase 1

Coagulation factor V
Inter-alpha-trypsin inhibitor heavy chain H4
Heparin cofactor 2
N-acetylmuramoyl-L-alanine amidase
Inter-alpha-trypsin inhibitor heavy chain H2
Hemoglobin subunit alpha
Plasma protease C1 inhibitor
Immunoglobulin J chain
Properdin
Fibulin-1
Ig lambda-2 chain C regions
Carboxypeptidase
Complement C1q subcomponent subunit B
Ig alpha-1 chain C region
Ankyrin repeat domain-containing protein 32
C4b-binding protein beta chain
Clusterin
Retinol-binding protein 4
von Willebrand factor
E3 SUMO-protein ligase RanBP2
Vasorin
Hyaluronan-binding protein 2
IgGFc-binding protein
Haptoglobin
Ig kappa chain C region
Myoglobin
Complement component C8 gamma chain
Complement C5
Haptoglobin-related protein

Hemoglobin subunit beta
Carboxypeptidase N subunit 2
Attractin
Keratin, type II cytoskeletal 1
Fibrinogen beta chain
Vitamin K-dependent protein S
Complement C1r subcomponent-like protein
Mannose-binding protein C
Basic complement C4
Carbonic anhydrase 1
Complement C3
Alpha-1-acid glycoprotein 2
Apolipoprotein A-II
Alpha-2-macroglobulin
Apolipoprotein B-100
Q7Z5P9-R
Ig mu chain C region
Alpha-1B-glycoprotein
Fibrinogen alpha chain
C4b-binding protein alpha chain
Complement C1q subcomponent subunit A
Hemoglobin subunit delta
Complement factor H-related protein 2
Apolipoprotein A-I
Plasma kallikrein
Alpha-1-acid glycoprotein 1
Alpha-1-antitrypsin
Afamin
Serotransferrin
Serum albumin
Scaffold

Mannose-binding protein C
Apolipoprotein D
Phosphatidylcholine-sterol acyltransferase
Complement C1q subcomponent subunit A
Complement factor H-related protein 1
Insulin-like growth factor-binding protein 3
Serum amyloid A-4 protein
Protein S100-A9
Plasminogen-related protein B
Ig kappa chain V-II region GM607 (Fragment)
Serotransferrin
Alpha-1-antitrypsin
Tripartite motif-containing protein 42
Hemoglobin subunit beta
Fibulin-1
Afamin
Apolipoprotein A-II
Prostaglandin-H2 D-isomerase
Complement C4-A
Apolipoprotein A-I
Ig mu heavy chain disease protein
Serum paraoxonase/lactonase 3
Haptoglobin
C-reactive protein
Alpha-1B-glycoprotein
Ig gamma-3 chain C region
Hemoglobin subunit delta
Lipopolysaccharide-binding protein
Hemoglobin subunit alpha
Alpha-1-acid glycoprotein 1
Haptoglobin-related protein
Transthyretin
Plasma kallikrein
Serum amyloid P-component
Alpha-1-acid glycoprotein 2
Glutathione peroxidase 3

Protein Z-dependent protease inhibitor
von Willebrand factor
Beta-2-glycoprotein 1
Complement component C8 gamma chain
Complement C1q subcomponent subunit C
Properdin
Ig lambda-6 chain C region
Biotinidase
Attractin
Hemoglobin subunit epsilon
Serine/threonine-protein kinase Nek9
Ig lambda chain V-I region HA
Apolipoprotein B-100
Ig gamma-1 chain C region
Complement C3
Complement C1q subcomponent subunit B
Ig kappa chain C region
Sex hormone-binding globulin
Galectin-3-binding protein
Ig heavy chain V-III region TEI
Fibronectin
Hemopexin
Inter-alpha-trypsin inhibitor heavy chain H1
Kininogen-1
Pigment epithelium-derived factor
Complement component C8 beta chain
Zinc-alpha-2-glycoprotein
Complement component C8 alpha chain
Histidine-rich glycoprotein
Coagulation factor X
Ig alpha-2 chain C region
Reticulocalbin-1
Ceruloplasmin
Plasminogen
Complement factor B
Inter-alpha-trypsin inhibitor

heavy chain H4
Apolipoprotein A-IV
Prothrombin
Antithrombin-III
Complement component C7
Complement component C9
Complement factor I
Vitronectin
Alpha-2-HS-glycoprotein
Protein AMBP
Tetranectin
Ig gamma-4 chain C region
Vitamin K-dependent protein C
Hepatocyte growth factor activator
Transforming growth factor-beta-induced protein ig-h3
Vitamin D-binding protein
Alpha-1-antichymotrypsin
Complement C2
Leucine-rich alpha-2-glycoprotein
Angiotensinogen
Lumican
Carboxypeptidase B2
Lymphatic vessel endothelial hyaluronic acid receptor 1
Serum albumin
ExpressionE
Serum albumin
Apolipoprotein A I
Ig heavy chain V III region WAS
Ig heavy chain V III region POM
Ig heavy chain V III region TUR
Alpha 1 antitrypsin
Mannose binding protein C
Tyrosine protein kinase transmembrane receptor ROR1
Afamin
Actin cytoplasmic 2
Alpha 1B glycoprotein
Serotransferrin

Complement C1q subcomponent subunit A
Actin cytoplasmic 1
Apolipoprotein A IV
Ig lambda chain V III region LOI
Sex hormone binding globulin
Platelet basic protein
Ig alpha 2 chain C region
Beta 2 glycoprotein 1
Serum amyloid P component
Complement factor H related protein 2
Hemopexin
Ig lambda 7 chain C region
Apolipoprotein A II
Alpha 1 antichymotrypsin
Lipopolysaccharide binding protein
Antithrombin III
Ig alpha 1 chain C region
Alpha 2 HS glycoprotein
Ig delta chain C region
Histidine rich glycoprotein
Complement C1q subcomponent subunit B
Vitamin D binding protein
Properdin
Ig lambda 1 chain C regions
Plasma kallikrein
Ceruloplasmin
Inter alpha trypsin inhibitor heavy chain H1
Transthyretin
Complement C1q subcomponent subunit C
Haptoglobin related protein
Heparin cofactor 2
Lumican
Ig lambda 6 chain C region
Inter alpha trypsin inhibitor heavy chain H2
Alpha 1 acid glycoprotein 1

Prothrombin
Complement factor H
Retinol binding protein 4
Haptoglobin
Plasminogen
Hemoglobin subunit alpha
Ig lambda 2 chain C regions
Complement C3
Complement factor H related protein 1
Pregnancy zone protein
Corticosteroid binding globulin
Inter alpha trypsin inhibitor heavy chain H4
Complement component C8 gamma chain
Attractin
Complement component C7
Hemoglobin subunit beta
Immunoglobulin lambda like polypeptide 5
Protein AMBP
C reactive protein
N acetylmuramoyl L alanine amidase
Glutathione peroxidase 3
Complement component C8 beta chain
Monocyte differentiation antigen CD14
Complement C4 B
Coagulation factor XII
Zinc alpha 2 glycoprotein
Alpha 2 antiplasmin
Complement factor H related protein 3
Kininogen 1
Complement C4 A
Pigment epithelium derived factor
Ig lambda 3 chain C regions
Complement factor I
Vitronectin
Complement component C9

Clusterin
Gelsolin
Inter alpha trypsin inhibitor heavy chain H3
Coagulation factor X
Fibronectin
Leucine rich alpha 2 glycoprotein
Complement C1s subcomponent
Complement factor B
Kallistatin
Thyroxine binding globulin
Complement component C8 alpha chain
Complement C5
Angiotensinogen
Ig mu chain C region
Plasma protease C1 inhibitor
Apolipoprotein B 100
Fibrinogen gamma chain
Complement C2
Complement component C6
Alpha 2 macroglobulin
Prostaglandin H2 D isomerase
Ig kappa chain V III region GOL
Peptidase inhibitor 16
Ig heavy chain V III region GAL
Hemoglobin subunit delta
Putative zinc alpha 2 glycoprotein like 1
Tetranectin
Ig kappa chain V III region Ti
Beta Ala His dipeptidase
Protein Z dependent protease inhibitor
Ig kappa chain V I region EU
Coagulation factor IX
Phosphatidylinositol glycan specific phospholipase D

Table A-2, Down-regulated proteins in comparison of disease and control

Progenesis
Ig lambda 2 chain C regions
Scaffold
Apolipoprotein C-III
Flavin reductase
MASCOT-Scaffold
Ig heavy chain V-III region GAL
Q8WXI7-R
Galectin-3-binding protein
Ig lambda chain V-III region LOI
Peroxiredoxin-2
Ig heavy chain V-III region BRO
Apolipoprotein C-III;
Apolipoprotein C-II
Low-density lipoprotein receptor-related protein 12
Ig kappa chain V-III region SIE
Kinesin-like protein KIF21A
ExpressionE
Complement C1r subcomponent like protein
Alpha 1 acid glycoprotein 2
Hemoglobin subunit epsilon
Carboxypeptidase N subunit 2

Carbonic anhydrase 2
Hepatocyte growth factor like protein
Ig kappa chain V II region RPMI 6410
Ig kappa chain V II region Cum
Fibrinogen alpha chain
Ig gamma 4 chain C region
Plasminogen related protein A
C4b binding protein beta chain
Beta 2 microglobulin
C4b binding protein alpha chain
CD5 antigen like
Insulin like growth factor binding protein complex acid labile subunit
Ig gamma 1 chain C region
Ig gamma 3 chain C region
Apolipoprotein E
Ig gamma 2 chain C region
Ig kappa chain C region
Peroxiredoxin 2
Peroxiredoxin 1
Ig lambda chain V I region HA
Flavin reductase NADPH

Table A-3: Up-regulated proteins when comparing disease vs. healthy control

Progenesis
Apolipoprotein A I
Serum albumin
Fibronectin
Alpha 1B glycoprotein
Hemoglobin subunit beta
Apolipoprotein A II
Coagulation factor XII
Cystatin C
MASCOT-Scaffold
Biotinidase
Dystonin
Inter-alpha-trypsin inhibitor

heavy chain H2
Gelsolin
Carboxypeptidase B2
Properdin
Peroxiredoxin-2
Dynein heavy chain 17, axonemal
Immunoglobulin lambda-like polypeptide 5
C-reactive protein
Ig lambda chain V-III region LOI
Alpha-1-acid glycoprotein 2
Ig kappa chain V-I region EU

Complement factor H
Clusterin
Ig heavy chain V-III region GAL
Carboxypeptidase N subunit 2
Coagulation factor XII
Zinc-alpha-2-glycoprotein
Vitamin K-dependent protein S
Keratin, type II cytoskeletal 1
Ceruloplasmin; AltName
Ig lambda-2 chain C regions
IgGFc-binding protein

C4b-binding protein beta chain
Complement C1r subcomponent-like protein
Complement C1q subcomponent subunit A
Retinol-binding protein 4
Complement C5
Plasma protease C1 inhibitor
Serum amyloid P-component
Ig gamma-4 chain C region
Complement factor B
Alpha-1-acid glycoprotein 1
Ig alpha-1 chain C region
Hemoglobin subunit delta
Complement component C8 gamma chain
Ig mu chain C region
Cystatin-C
Q7Z5P9-R
Leucine-rich alpha-2-glycoprotein
Alpha-1B-glycoprotein
von Willebrand factor
Complement factor H-related protein 2
Hemoglobin subunit alpha
Extracellular matrix protein 1
Fibrinogen gamma chain
Inter-alpha-trypsin inhibitor heavy chain H4
Alpha-1-antitrypsin
Apolipoprotein B-100
Alpha-2-macroglobulin
Fibrinogen beta chain
C4b-binding protein alpha chain
Attractin
Haptoglobin-related protein
Complement C4-B
Plasma kallikrein
Fibrinogen alpha chain
Hemoglobin subunit beta
Apolipoprotein A-II
Complement C3
Apolipoprotein A-I

Serotransferrin
Fibronectin
Afamin
Serum albumin
ExpressionE
Carbonic anhydrase 1
Apolipoprotein A I
Afamin
Tetranectin
Vitronectin
Ig lambda chain V III region LOI
Ceruloplasmin
N acetylmuramoyl L alanine amidase
Ig alpha 2 chain C region
Ig alpha 1 chain C region
Ig gamma 4 chain C region
Monocyte differentiation antigen CD14
Ig gamma 1 chain C region
Gelsolin
Kininogen 1
Complement factor B
Complement component C7
Ig kappa chain C region
Complement C5
Coagulation factor XII
Plasminogen
Coagulation factor X
Coagulation factor IX
Alpha 2 macroglobulin
Complement factor H related protein 1
Haptoglobin related protein
Haptoglobin
Prothrombin
Ficolin 3
Histidine rich glycoprotein
Angiotensinogen
Alpha 1 antichymotrypsin
Alpha 1 antitrypsin
Antithrombin III 1
Hemopexin
Inter alpha trypsin inhibitor heavy chain H1

Complement C2
Serotransferrin
Inter alpha trypsin inhibitor heavy chain H2
Mannose binding protein C
Platelet basic protein
Vitamin D binding protein
Transthyretin
Alpha 2 HS glycoprotein
Protein AMBP
Complement C4 B
Complement C4 A
Retinol binding protein 4
Fibronectin
Beta 2 glycoprotein 1
Complement component C9
Complement C1q subcomponent subunit C
Clusterin
Complement C1q subcomponent subunit B
Complement C1q subcomponent subunit A
Serum amyloid P component
Apolipoprotein B 100
C reactive protein
Hemoglobin subunit beta
Heparin cofactor 2
Thyroxine binding globulin
Zinc alpha 2 glycoprotein
Ig kappa chain V III region Ti
Alpha 2 antiplasmin
Alpha 1 acid glycoprotein 2
Complement factor H related protein 2
Hemoglobin subunit alpha
Ig lambda 3 chain C regions
Ig lambda 2 chain C regions
Ig lambda 1 chain C regions
Biotinidase
Pigment epithelium derived factor
Glutathione peroxidase 3
Corticosteroid binding globulin
Sex hormone binding globulin

Complement factor H
Inter alpha trypsin inhibitor heavy chain H4
Plasma kallikrein
Ig heavy chain V III region GAL
Ig heavy chain V III region WAS
Ig heavy chain V III region POM
Ig heavy chain V III region TIL
Ig heavy chain V III region VH26
Alpha 1B glycoprotein
Hemoglobin subunit delta
Serum paraoxonase arylesterase 1
Ig kappa chain V III region GOL
Apolipoprotein A IV
Complement component C8 gamma chain
Cholinesterase
Complement component C8 beta chain
Complement component C8 alpha chain
Complement C1s subcomponent
Properdin
Lumican
Protein Z dependent protease inhibitor
Ig lambda chain V I region HA
Extracellular matrix protein 1
Complement factor I
Phosphatidylinositol glycan specific phospholipase D
Plasma protease C1 inhibitor
Complement component C6
Kallistatin
Pregnancy zone protein
Fibrinogen gamma chain
Carboxypeptidase B2
Inter alpha trypsin inhibitor heavy chain H3
Peptidase inhibitor 16
Hemoglobin subunit gamma 2
Lipopolysaccharide binding protein
Prostaglandin H2 D isomerase
Pyruvate dehydrogenase E1 component subunit beta mitochondrial
Carboxypeptidase N catalytic chain
Vitamin K dependent protein S
Hemoglobin subunit gamma 1
Scaffold
Complement C3
Complement C1q subcomponent subunit A
Actin, cytoplasmic 1
Alpha-1B-glycoprotein
Vitamin K-dependent protein C
Ig mu heavy chain disease protein
Hemoglobin subunit epsilon
Protein Z-dependent protease inhibitor
Biotinidase
von Willebrand factor
Transthyretin
Glutathione peroxidase 3
Ig alpha-2 chain C region
Haptoglobin
Reticulocalbin-1
Cholinesterase
Tripartite motif-containing protein 42
Ig gamma-3 chain C region
Fibulin-1
Plasma kallikrein
Complement component C8 gamma chain
C-reactive protein
Beta-2-glycoprotein 1
Carbonic anhydrase 2
Coagulation factor V
Carbonic anhydrase 1
Kininogen-1
Tetranectin
Properdin
Fetuin-B
Protein S100-A9
Corticosteroid-binding globulin
Plasminogen-related protein B
Alpha-1-acid glycoprotein 1
Hepatocyte growth factor activator
Inter-alpha-trypsin inhibitor heavy chain H1
Heparin cofactor 2
Kallistatin
Coagulation factor X
Galectin-3-binding protein
Serum amyloid P-component
Ficolin-3
Hemoglobin subunit delta
Prothrombin
Ig kappa chain C region
Lipopolysaccharide-binding protein
Histidine-rich glycoprotein
Apolipoprotein A-IV
Antithrombin-III
Complement C2
Pigment epithelium-derived factor
Complement C1q subcomponent subunit C
Attractin
Sex hormone-binding globulin
Angiotensinogen
Serum paraoxonase/arylesterase 1
Alpha-1-acid glycoprotein 2
Hemopexin
Complement component C7
Complement component C8 beta chain
Complement component C8 alpha chain
Ig gamma-1 chain C region

Retinol-binding protein 4
Plasminogen
Complement factor I
Alpha-2-antiplasmin 3
Alpha-2-HS-glycoprotein
Lumican
Ig heavy chain V-III region TEI
Complement C5
Ceruloplasmin
N-acetylmuramoyl-L-alanine amidase
Coagulation factor XII
Thyroxine-binding globulin
Vitamin D-binding protein
Gelsolin
Protein AMBP
Phosphatidylinositol-glycan-specific phospholipase D
Coagulation factor IX
Ig gamma-4 chain C region
Transforming growth factor-beta-induced protein ig-h3
Apolipoprotein B-100
Complement factor B
Carboxypeptidase N catalytic chain
Immunoglobulin J chain
Inter-alpha-trypsin inhibitor heavy chain H4
Complement C1s subcomponent
Hyaluronan-binding protein 2
Alpha-2-macroglobulin
Inter-alpha-trypsin inhibitor heavy chain H2
Complement component C9
Insulin-like growth factor-binding protein complex acid labile subunit
Complement C1r subcomponent

Complement C1q subcomponent subunit B
CD5 antigen-like
Complement factor H
Alpha-1-antichymotrypsin
Vitronectin
Carboxypeptidase B2
Complement factor D
Complement component C6
Ig mu chain C region
Monocyte differentiation antigen CD14
Plasma protease C1 inhibitor
Zinc-alpha-2-glycoprotein
Extracellular matrix protein 1
Immunoglobulin lambda-like polypeptide 5
Pregnancy zone protein
Ig heavy chain V-III region VH26
Complement factor H-related protein 2
Hemoglobin subunit beta
Clusterin
Prostaglandin-H2 D-isomerase
Leucine-rich alpha-2-glycoprotein
Complement C1r subcomponent-like protein
Ig kappa chain V-III region SIE
Inter-alpha-trypsin inhibitor heavy chain H3
Fibrinogen gamma chain
Carboxypeptidase N subunit 2
Beta-Ala-His dipeptidase
Fibrinogen alpha chain
Selenoprotein P
Ig heavy chain V-III region GAL

Alcohol dehydrogenase 1
Ig alpha-1 chain C region
C4b-binding protein beta chain
Fibrinogen beta chain
Vitamin K-dependent protein S
Ig gamma-2 chain C region
Apolipoprotein E
Serine/threonine-protein kinase Nek9
Apolipoprotein A-I
C4b-binding protein alpha chain
Beta-2-microglobulin
Ig heavy chain V-III region BRO
Hepatocyte growth factor-like protein
Ig lambda chain V-I region HA
Cystatin-C
Ig lambda chain V-III region LOI
Hemoglobin subunit alpha
Haptoglobin-related protein
Platelet basic protein
Serotransferrin
Complement factor H-related protein 1
Afamin
Serum albumin
Complement C4-A
Mannose-binding protein C
Fibronectin
Alpha-1-antitrypsin
Peroxiredoxin-2
Lymphatic vessel endothelial hyaluronic acid receptor 1
Peptidase inhibitor 16

Table A-4, Down-regulated proteins when comparing disease vs. healthy control

Progenesis
Fibrinogen gamma chain
Putative alpha 1 antitrypsin related protein
Fibrinogen beta chain
Insulin like growth factor binding protein complex acid labile subunit
Fibrinogen alpha chain
Ig alpha 1 chain C region
Ceruloplasmin
MASCOT-Scaffold
Hepatocyte growth factor activator
Low-density lipoprotein receptor-related protein 12
Apolipoprotein C-III
Kinesin-like protein KIF21A
Apolipoprotein C-II

Galectin-3-binding protein
ExpressionE
Cystatin C
Plasminogen related protein A
Complement C1r subcomponent like protein
Apolipoprotein E
Apolipoprotein C III
Plasminogen related protein B
C4b binding protein alpha chain
Ig mu chain C region
Ig gamma 3 chain C region
Ig gamma 2 chain C region
Hepatocyte growth factor like protein
C4b binding protein beta

chain
Leucine rich alpha 2 glycoprotein
Insulin like growth factor binding protein complex acid labile subunit
Ig lambda 7 chain C region
Ig mu heavy chain disease protein
Beta 2 microglobulin
Fibrinogen beta chain
Fibrinogen alpha chain
Scaffold
None

Table A-5, Up-regulated proteins when comparing low grade Ta vs. all other disease

ExpressionE
Apolipoprotein A II
Complement C3
Ig kappa chain V I region EU
Ig kappa chain V IV region Len
MASCOT-Scaffold
Sex hormone-binding globulin
Attractin
Complement C1q subcomponent subunit A
Complement C1q subcomponent subunit B
Complement component C9
Extracellular matrix protein 1
Ig kappa chain C region
Immunoglobulin lambda-like polypeptide 5
Inter-alpha-trypsin inhibitor heavy chain H3
Protein AMBP
Complement C1r subcomponent

Glutathione peroxidase 3
Ig lambda-2 chain C regions
IgGFc-binding protein
Phosphatidylinositol-glycan-specific phospholipase D
Platelet basic protein
PLGS-Scaffold
Sex hormone-binding globulin
Adiponectin
Ig kappa chain V-IV region JI
Phosphatidylcholine-sterol acyltransferase
Platelet basic protein
Prostaglandin-H2 D-isomerase
Protein S100-A9
Progenesis
Sex hormone binding globulin
Attractin
Complement C1q subcomponent subunit A
Complement C1q

subcomponent subunit B
Complement component C9
Extracellular matrix protein 1
Ig kappa chain C region
Immunoglobulin lambda like polypeptide 5
Inter alpha trypsin inhibitor heavy chain H3
Protein AMBP
Adiponectin
Afamin
Alpha 1B glycoprotein
Alpha 2 antiplasmin
Antithrombin III
Apolipoprotein A IV
Beta 2 glycoprotein 1
Beta 2 microglobulin
Biotinidase
BolA like protein 2
C4b binding protein alpha chain
Carbonic anhydrase 1

Carboxypeptidase B2
Carboxypeptidase N catalytic chain
Carboxypeptidase N subunit 2
CD44 antigen
Ceruloplasmin
Coagulation factor IX
Coagulation factor X
Coagulation factor XII
Complement C1q subcomponent subunit C
Complement C1r subcomponent like protein
Complement C5
Complement component C6
Complement component C7
Complement component C8 alpha chain
Complement component C8 beta chain
Complement component C8 gamma chain
Complement factor B
Complement factor D
Complement factor H
Complement factor H related protein 1
Complement factor H related protein 3
Complement factor I
Corticosteroid binding

globulin
Cystatin C
Extracellular superoxide dismutase Cu Zn
Fetuin B
Fibronectin
Fibulin 1
Hemoglobin subunit epsilon
Hyaluronan binding protein 2
Ig alpha 1 chain C region
Ig gamma 2 chain C region
Ig heavy chain V III region TEI
Ig kappa chain V III region HAH
Ig kappa chain V III region VH Fragment
Ig kappa chain V III region WOL
Ig lambda chain V III region LOI
Inter alpha trypsin inhibitor heavy chain H1
Inter alpha trypsin inhibitor heavy chain H2
Inter alpha trypsin inhibitor heavy chain H4
Lipopolysaccharide binding protein
Monocyte differentiation antigen CD14
N acetylmuramoyl L alanine amidase

Nuclear receptor interacting protein 3
Peptidase inhibitor 16
Pigment epithelium derived factor
Plasma kallikrein
Plasma protease C1 inhibitor
Plasminogen
Plasminogen related protein A
Plasminogen related protein B
Pregnancy zone protein
Properdin
Putative zinc alpha 2 glycoprotein like 1
Retinol binding protein 4
Ribonuclease inhibitor
Selenoprotein P
Serum amyloid P component
Sulfhydryl oxidase 1
Tetranectin
Thyroxine binding globulin
Transthyretin
Tumor necrosis factor ligand superfamily member 18
UAP56 interacting factor
Vasorin
Vitamin D binding protein
von Willebrand factor
Zinc alpha 2 glycoprotein

Table A-6, Down-regulated proteins when comparing low grade Ta vs. all other disease

Expression
C reactive protein
Ig lambda 2 chain C regions
Hemoglobin subunit alpha
Properdin
Alpha 2 antiplasmin
Alpha 2 macroglobulin
Apolipoprotein B 100
C4b binding protein alpha chain

Ceruloplasmin
Coagulation factor IX
Complement C1s subcomponent
Complement C5
Complement component C6
Complement component C8 beta chain

Complement factor B
Hemoglobin subunit delta
Ig alpha 1 chain C region
Ig mu chain C region
Inter alpha trypsin inhibitor heavy chain H2
Inter alpha trypsin inhibitor heavy chain H4
Serotransferrin

Serum albumin
Thyroxine binding globulin
Transthyretin
Zinc alpha 2 glycoprotein
Actin cytoplasmic 1
Actin cytoplasmic 2
Adiponectin
Afamin
Alpha 1 acid glycoprotein 1
Alpha 1 acid glycoprotein 2
Alpha 1 antichymotrypsin
Alpha 1 antitrypsin
Alpha 1B glycoprotein
Alpha 2 HS glycoprotein
Angiotensinogen
Antithrombin III
Apolipoprotein A I
Apolipoprotein A IV
Apolipoprotein E
Attractin
Beta 2 glycoprotein 1
Beta 2 microglobulin
Biotinidase
C4b binding protein beta chain
Carbonic anhydrase 1
Carboxypeptidase B2
Carboxypeptidase N catalytic chain
Carboxypeptidase N subunit 2
Cholinesterase
Clusterin
Coagulation factor X
Coagulation factor XII
Complement C1q subcomponent subunit A
Complement C1q subcomponent subunit B
Complement C1q subcomponent subunit C
Complement C1r subcomponent like protein
Complement C1r subcomponent
Complement C2

Complement C4 A
Complement C4 B
Complement component C7
Complement component C8 alpha chain
Complement component C8 gamma chain
Complement component C9
Complement factor D
Complement factor H
Complement factor H related protein 1
Complement factor H related protein 2
Complement factor H related protein 3
Complement factor I
Corticosteroid binding globulin
Cystatin C
Extracellular matrix protein 1
Fibrinogen alpha chain
Fibrinogen beta chain
Fibrinogen gamma chain
Fibronectin
Fibulin 1
Ficolin 3
Galectin 3 binding protein
Gelsolin
Glutathione peroxidase 3
Haptoglobin
Haptoglobin related protein
Hemoglobin subunit beta
Hemoglobin subunit epsilon
Hemoglobin subunit gamma 1
Hemoglobin subunit gamma 2
Hemopexin
Heparin cofactor 2
Hepatocyte growth factor activator
Hepatocyte growth factor like protein
Histidine rich glycoprotein
Hyaluronan binding protein 2
Ig alpha 2 chain C region

Ig delta chain C region
Ig gamma 1 chain C region
Ig gamma 2 chain C region
Ig gamma 3 chain C region
Ig gamma 4 chain C region
Ig heavy chain V III region BRO
Ig heavy chain V III region BUT
Ig heavy chain V III region TEI
Ig heavy chain V III region VH26
Ig heavy chain V III region WEA
Ig kappa chain C region
Ig kappa chain V III region GOL
Ig kappa chain V III region SIE
Ig kappa chain V III region Ti
Ig kappa chain V III region WOL
Ig lambda 1 chain C regions
Ig lambda 3 chain C regions
Ig lambda 6 chain C region
Ig lambda 7 chain C region
Ig lambda chain V III region LOI
Ig mu heavy chain disease protein
Immunoglobulin J chain
Immunoglobulin lambda like polypeptide 5
Insulin like growth factor binding protein complex acid labile subunit
Inter alpha trypsin inhibitor heavy chain H1
Inter alpha trypsin inhibitor heavy chain H3
Kallistatin
Kininogen 1
Leucine rich alpha 2 glycoprotein
Lipopolysaccharide binding protein
Lumican
Lymphatic vessel endothelial hyaluronic acid receptor 1

Mannose binding protein C
Monocyte differentiation antigen CD14
N acetylmuramoyl L alanine amidase
Peptidase inhibitor 16
Peroxiredoxin 1
Peroxiredoxin 2
Phosphatidylcholine sterol acyltransferase
Phosphatidylinositol glycan specific phospholipase D
Pigment epithelium derived factor
Plasma kallikrein
Plasma protease C1 inhibitor
Plasminogen
Plasminogen related protein A
Plasminogen related protein B
Platelet basic protein
Pregnancy zone protein
Protein AMBP
Protein Z dependent protease inhibitor
Prothrombin
Putative zinc alpha 2 glycoprotein like 1
Retinol binding protein 4
Serum amyloid P component
Serum paraoxonase arylesterase 1
Sex hormone binding globulin
Tetranectin
Vitamin D binding protein
Vitamin K dependent protein C
Vitamin K dependent protein S

Vitronectin
MASCOT-Scaffold
Hemoglobin subunit alpha
Properdin
Alpha-2-antiplasmin
Alpha-2-macroglobulin
Apolipoprotein B-100
C4b-binding protein alpha chain
Ceruloplasmin
Coagulation factor IX
Complement C1s subcomponent
Complement C5
Complement component C6
Complement component C8 beta chain
Complement factor B
Hemoglobin subunit delta
Ig alpha-1 chain C region
Ig mu chain C region
Inter-alpha-trypsin inhibitor heavy chain H2
Inter-alpha-trypsin inhibitor heavy chain H4
Serotransferrin
Serum albumin
Thyroxine-binding globulin
Transthyretin
Zinc-alpha-2-glycoprotein
CD5 antigen-like
Complement C3
Keratin, type II cytoskeletal 1
Myoglobin
PLGS-Scaffold
C-reactive protein
Ig lambda-2 chain C regions

Hemoglobin subunit alpha
Properdin
Apolipoprotein C-II
Apolipoprotein C-III
CD5 antigen-like
Complement factor H-related protein 1
Fetuin-B
Fibrinogen beta chain
Ig mu heavy chain disease protein
L-selectin
Monocyte differentiation antigen CD14
Protein S100-A8
Protein Z-dependent protease inhibitor
Putative alpha-1-antitrypsin-related protein
Transforming growth factor-beta-induced protein ig-h3
Vitamin K-dependent protein C
Progenesis
C reactive protein
Ig lambda 2 chain C regions
Apolipoprotein C II
Ig kappa chain V IV region Fragment
Insulin like growth factor binding protein 3
Low affinity immunoglobulin gamma Fc region receptor III B
Myoglobin
Prostaglandin H2 D isomerase

Table A-7, Up-regulated proteins when comparing TaT1 vs. T2T3 in grade 3 samples

Expression
Complement C5
Plasminogen
Prothrombin
Complement C1s

subcomponent
Carboxypeptidase N catalytic chain
Inter alpha trypsin inhibitor heavy chain H1
Inter alpha trypsin inhibitor

heavy chain H2
Pigment epithelium derived factor
N acetylmuramoyl L alanine amidase
Insulin like growth factor

binding protein complex acid labile subunit
Alpha 2 antiplasmin
Serum amyloid P component
Lumican
Ficolin 3
Gelsolin
Complement C4 B
Complement C4 A
MASCOT-Scaffold
Galectin-3-binding protein
Ig lambda chain V-III region SH
Hemoglobin subunit beta
Plasma protease C1 inhibitor
Ig kappa chain V-III region SIE
Ig gamma-2 chain C region
Inter-alpha-trypsin inhibitor heavy chain H4
Apolipoprotein A-II
Ig mu chain C region
Leucine-rich alpha-2-glycoprotein
Alpha-1-antichymotrypsin
Ig heavy chain V-III region TIL
Ig kappa chain V-I region EU
Ig kappa chain C regio.n

Fibronectin
Ig alpha-2 chain C region
Myoglobin
C-reactive protein
Alpha-1-acid glycoprotein 1
Alpha-1-acid glycoprotein 2
Serotransferrin
Alpha-1-antitrypsin
Complement C3
Haptoglobin
PLGS-Scaffold
Vitamin K-dependent protein C
Attractin
Transforming growth factor-beta-induced protein ig-h3
Tetranectin
Beta-Ala-His dipeptidase
Peptidase inhibitor 16
Complement C1q subcomponent subunit B
Haptoglobin-related protein
L-selectin
Cholinesterase
Sex hormone-binding globulin
Peroxiredoxin-2
Beta-2-microglobulin

Properdin
Afamin
Complement C1q subcomponent subunit C
Fetuin-B
Protein Z-dependent protease inhibitor
Ficolin-3
Carboxypeptidase N catalytic chain
Corticosteroid-binding globulin
Complement C1q subcomponent subunit A
Cystatin-C
Progenesis
Inter alpha trypsin inhibitor heavy chain H2
Carboxypeptidase N catalytic chain
Biotinidase
Protein Z dependent protease inhibitor
Lipopolysaccharide binding protein
Lymphatic vessel endothelial hyaluronic acid receptor 1
Insulin like growth factor binding protein 3
Sex hormone binding globulin

Table A-8, Down-regulated proteins when comparing TaT1 vs. T2T3 in grade 3 samples

Expression
Complement factor I
Serum paraoxonase arylesterase 1
Complement factor D
Ig kappa chain V II region RPMI 6410
C reactive protein
Ig gamma 3 chain C region
Ig gamma 2 chain C region

Ig gamma 1 chain C region
Pregnancy zone protein
Plasminogen related protein A
Kininogen 1
Plasma kallikrein
Ig kappa chain C region
Complement factor B
Cystatin C
Carboxypeptidase N

subunit 2
Galectin 3 binding protein
Ig mu heavy chain disease protein
Coagulation factor XII
Hemoglobin subunit epsilon
Coagulation factor X
Coagulation factor IX
Complement C3

Alpha 2 macroglobulin
Ig lambda 7 chain C region
Hyaluronan binding protein 2
Alpha 1B glycoprotein
Complement C1r subcomponent like protein
Protein Z dependent protease inhibitor
Apolipoprotein A IV
Haptoglobin related protein
Ig heavy chain V III region TEI
Haptoglobin
Complement C1r subcomponent
Attractin
Complement C2
Alpha 1 antichymotrypsin
Ig heavy chain V III region BRO
Ig heavy chain V III region TIL
Ig heavy chain V III region VH26
Complement component C8 alpha chain
Beta Ala His dipeptidase
Complement factor H related protein 2
Alpha 1 antitrypsin
Antithrombin III
Hemoglobin subunit delta
Cholinesterase
Ig lambda 6 chain C region
Ig kappa chain V II region GM607 Fragment
Hemopexin
Extracellular matrix protein 1
Ig lambda 3 chain C regions
Ig lambda 2 chain C regions
Ig lambda 1 chain C regions
Carboxypeptidase B2
Serotransferrin
Complement component C7

Platelet basic protein
Vitamin D binding protein
Serum albumin
Transthyretin OS Homo sapiens GN TTR PE 1 SV 1
Alpha 2 HS glycoprotein
Alpha 1 acid glycoprotein 1
Protein AMBP
Plasma protease C1 inhibitor
Apolipoprotein B 100
Retinol binding protein 4
Fibronectin
Leucine rich alpha 2 glycoprotein
Heparin cofactor 2
Beta 2 glycoprotein 1
Complement component C9
Complement C1q subcomponent subunit C
Complement C1q subcomponent subunit B
Complement C1q subcomponent subunit A
Afamin OS Homo sapiens GN AFM PE 1 SV 1
Putative zinc alpha 2 glycoprotein like 1
Biotinidase
Ig lambda chain V III region LOI
Fibrinogen gamma chain
Fibrinogen beta chain
Fibrinogen alpha chain
Vitamin K dependent protein S
Immunoglobulin lambda like polypeptide 5
Immunoglobulin J chain
Hemoglobin subunit beta
Ig kappa chain V III region WOL
Ig kappa chain V III region SIE
Complement factor H related protein 3
Tetranectin

Ig kappa chain V II region TEW
Apolipoprotein A II
Ig kappa chain V II region Cum
Complement component C6
Hemoglobin subunit gamma 2
Apolipoprotein E
Hemoglobin subunit gamma 1
Apolipoprotein A I
C4b binding protein alpha chain
C4b binding protein beta chain
Alpha 1 acid glycoprotein 2
Clusterin
Properdin
Inter alpha trypsin inhibitor heavy chain H3
Corticosteroid binding globulin
Complement factor H related protein 1
Beta 2 microglobulin
Monocyte differentiation antigen CD14
Hemoglobin subunit alpha
Complement factor H
Inter alpha trypsin inhibitor heavy chain H4
Sex hormone binding globulin
Ig alpha 2 chain C region
Ig alpha 1 chain C region
Ig mu chain C region
Zinc alpha 2 glycoprotein
Selenoprotein P
Ig gamma 4 chain C region
MASCOT-Scaffold
Ig gamma-1 chain C region
Transthyretin
Complement component C7
Corticosteroid-binding globulin

Coagulation factor V
Complement C1q subcomponent subunit A
Ceruloplasmin
Apolipoprotein B-100
Complement factor H-related protein 2
Complement C1s subcomponent
Complement component C8 beta chain
Thyroxine-binding globulin
Carbonic anhydrase 1
Kallistatin
Lymphatic vessel endothelial hyaluronic acid receptor
Microtubule-actin cross-inking factor 1
Coagulation factor X
Beta-2-microglobulin
Complement component C6
Pigment epithelium-derived factor
Complement C2
IgGFc-binding protein
N-acetylmuramoyl-L-alanine amidase
Complement C1q subcomponent subunit C
Complement factor B
Monocyte differentiation

antigen CD14
Biotinidase
Hyaluronan-binding protein 2
Complement C4-B
Gelsolin
Peptidase inhibitor 16
Plasminogen
Coagulation factor IX
Serum paraoxonase/arylesterase 1
Properdin
Afamin
Cholinesterase
Complement C1q subcomponent subunit B
Vitamin K-dependent protein C
Serum amyloid P-component
Carboxypeptidase N catalytic chain
Alpha-2-macroglobulin
Peroxiredoxin-2
Tetranectin; Short=TN
Phosphatidylinositol-glycan-specific phospholipase D
Plasma kallikrein
Insulin-like growth factor-binding protein complex

Vitamin K-dependent protein S
PLGS-Scaffold
Ig gamma-1 chain C region
Ig gamma-3 chain C region
Mannose-binding protein C
Serotransferrin
Alpha-1-acid glycoprotein 2
Immunoglobulin lambda-like polypeptide 5
Apolipoprotein A-II
Ig kappa chain V-III region SIE
Platelet basic protein
Hemoglobin subunit epsilon
Apolipoprotein C-III
Alpha-1-acid glycoprotein 1
Ig alpha-2 chain C region
Haptoglobin
Ig kappa chain C region
C-reactive protein
Ig heavy chain V-III region GAL
Ig kappa chain V-II region GM607 (Fragment)
Complement factor H-related protein 1
Progenesis
Haptoglobin

7.2 Appendix B

List of the unique proteins identified by PLGS and MASCOT

Table 1, Unique proteins in comparison of disease and control

Unique proteins found in disease samples
Complement factor D
Apolipoprotein A I
Coagulation factor IX
Plasma kallikrein
Inter alpha trypsin inhibitor heavy chain H3
Fibronectin

Unique proteins found in control samples
Hemoglobin subunit beta
Immunoglobulin J chain
Complement C1q subcomponent subunit B
Zinc alpha 2 glycoprotein
Ig gamma 4 chain C region
Haptoglobin
Complement C1r subcomponent

Table 2, Unique proteins when comparing disease vs. healthy control

Unique proteins found in disease samples
Apolipoprotein A I
Coagulation factor IX
Fibronectin

Unique proteins found in healthy control samples
Beta 2 microglobulin
Hemoglobin subunit beta
Immunoglobulin J chain
C reactive protein
Serum amyloid P component
Complement C1q subcomponent subunit B
Zinc alpha 2 glycoprotein
Ig gamma 4 chain C region
Ig alpha 2 chain C region
Ig mu heavy chain disease protein
Haptoglobin
Alpha 1 antitrypsin
Vitamin K dependent protein S
Complement C1r subcomponent

Table 3, Unique proteins when comparing TaG1-TaG2 vs. all other disease

Unique proteins found in TaG1-TaG2 samples
Tetranectin
Serum amyloid P component

Unique proteins found in other disease
None

Glutathione peroxidase 3
Complement C1q subcomponent subunit B
Ficolin 3
Zinc alpha 2 glycoprotein
Ig gamma 4 chain C region
Extracellular matrix protein 1
Biotinidase
Cholinesterase
Serotransferrin
Complement C1r subcomponent
Complement C4 A

7.3 Appendix C

List of publication and presentations in scientific meetings

Publications:

- **Hakimi, A.**, Auluck, J., Jones, G.D.D., Ng, L.L., Jones D.J.L. (**2013**). Assessment of reproducibility in depletion and enrichment workflows for plasma proteomics using label-free quantitative data-independent liquid chromatography mass spectrometry. Manuscript re-submitted to Proteomics following revision.

- Daly, E.C, Ng, L.L., **Hakimi, A.**, Willingale R., Jones D.J.L. (**2013**). Qualitative and quantitative characterisation of plasma proteins when incorporating travelling wave ion mobility into a liquid chromatogeraphy mass spectrometry workflow for biomarker discovery: The use of product ion quantitation as an alternative data analysis tool for label free quantitation. Manuscript submitted to Analytical Chemistry.

Presentations:

- **Hakimi, A.**, Ng, L.L., Jones, G.D.D., Jones, D.J.L., (**2012**). Poster presentation at the 11th East midlands proteomics workshop, Loughborough University, UK.

- **Hakimi, A.**, Jones, G.D.D., Ng, L.L, Jones D.J.L. (**2012**). Poster presentation at the 11th annual world congress of HUPO (Human Proteome Organisation), Hynes convention centre, Boston, MA, USA.

- **Hakimi, A.**, Jones, G.D.D., Jones, D.J.L., (**2011**). Poster presentation at the HOPE researchers evening, Charles Frears Campus, Leicester, UK.

- **Hakimi, A.**, Daly, C. E., Jones, G.D.D., Jones, D.J.L., (**2011**). Poster presentation at the 32nd BMSS annual meeting, Cardiff, UK.

- **Hakimi, A.**, Daly, C. E., Ng, L.L., Jones, G.D.D., Jones, D.J.L., (**2011**). Poster presentation at the 10th East midlands proteomics workshop, Loughborough University, UK.

- **Hakimi, A.**, Jones, G.D.D., Jones, D.J.L., (**2010**). Poster presentation at the 31st BMSS annual meeting, Cardiff, UK.

7.4 Appendix D

Detailed information for 104 plasma samples

		Italian ID			Pathology						Follow-up	
	Leicester ID	Tube ID	Sample ID	Code	T	G	AGE	date diagnosis	TREATMENT	DIAGNOSIS OF CONTROLS	RECURRENCE Date	CISTECTOMY Date
1	L1-1	34	1189	LOW_9	a	2	74	23/11/2007	xxx	Disease samples	xxx	xxx
2	L1-2	31	1175	LOW_8	1	2	64	29/09/2007	xxx		xxx	xxx
3	L1-3	30	1171	LOW_7	1	3	60	11/09/2007	BCG		xxx	xxx
4	L1-4	29	1166	LOW_6	a	1	68	29/06/2007	xxx		xxx	xxx
5	L1-5	28	1158	LOW_5	a	2	63	05/06/2001	CISPLATIN+GEMCITABINE		xxx	21/06/2007
6	L1-6	25	1137	LOW_4	1	2	56	06/03/2007	xxx		19/10/2008	xxx
7	L1-7	20	1112	LOW_3	a	1	54	15/01/2007	xxx		xxx	xxx
8	L1-8	55	1438	LOW_20	a	2	71	08/04/2010	xxx		xxx	xxx
9	L1-9	15	1084	LOW_2	a	2	67	13/12/2006	BCG		15/01/2008	xxx
10	L1-10	54	1433	LOW_19	a	2	71	03/03/2010	xxx		xxx	xxx
11	L1-11	53	1429	LOW_18	a	1	67	23/03/2010	xxx		xxx	xxx
12	L1-12	52	1426	LOW_17	a	1	74	16/03/2010	xxx		xxx	xxx
13	L1-13	51	1425	LOW_16	1	2	69	17/03/2010	xxx		xxx	xxx
14	L1-14	47	1417	LOW_15	a	2	63	09/03/2010	xxx		xxx	xxx
15	L1-15	43	1267	LOW_14	a	3	73	14/07/2008	BCG		21/05/2009	22/05/2009
16	L1-16	42	1253	LOW_13	a	2	68	28/07/2008	xxx		xxx	xxx
17	L1-17	41	1227	LOW_12	a	1	70	09/04/2008	xxx		xxx	xxx
18	L1-18	37	1212	LOW_11	1	3	61	25/02/2008	BCG		07/04/2008	xxx
19	L1-19	35	1196	LOW_10	1	2	65	10/01/2008	xxx		xxx	xxx
20	L1-20	1	1001	LOW_1	1	3	58	05/07/2006	BCG		23/08/2006	23/08/2006
21	H1-21	19	1111	HIGH_9	2	3	73	22/01/2007	BCG		xxx	15/01/2007
22	H1-22	18	1101	HIGH_8	X	2	57	18/12/2006	xxx		xxx	xxx

23	H1-23	14	1069	HIGH_7	2	3	70	12/12/2006	xxx		xxx	12/02/2007
24	H1-24	11	1053	HIGH_6	2a	2	72	19/10/2006	xxx		xxx	25/10/2006
25	H1-25	8	1028	HIGH_5	2	3	67	29/05/2006	xxx		xxx	29/08/2006
26	H1-26	7	1027	HIGH_4	3b	3	69	04/08/2006	xxx		xxx	08/12/2006
27	H1-27	6	1022	HIGH_3	2	3	46	02/08/2006	CISPLATIN+GEMCITABINE		08/08/2006	04/12/2006
28	H1-28	60	735	HIGH_20	2	3	56	16/04/2003	CISPLATIN+GEMCITABINE		xxx	21/05/2003
29	H1-29	5	1021	HIGH_2	4	3	60	03/08/2006	GEMCITABINE		xxx	xxx
30	H1-30	59	716	HIGH_19	2	3	73	12/02/2003	xxx		xxx	24/03/2003
31	H1-31	58	405	HIGH_18	2	3	60	05/09/2000	CAPECITABINE		xxx	08/09/2000
32	H1-32	57	370	HIGH_17	2b	3	64	20/07/2000	GEMCITABINE		xxx	20/07/2000
33	H1-33	56	362	HIGH_16	2	3	65	16/02/2000	xxx		xxx	27/03/2000
34	H1-34	45	1292	HIGH_15	2	3	65	08/10/2008	BCG		xxx	28/01/2009
35	H1-35	44	1291	HIGH_14	2	3	73	22/10/2008	xxx		xxx	xxx
36	H1-36	40	1218	HIGH_13	2	3	61	12/03/2008	xxx		16/05/2008	xxx
37	H1-37	39	1217	HIGH_12	2	3	74	04/03/2008	xxx		02/06/2008	xxx
38	H1-38	26	1139	HIGH_11	2	3	72	15/03/2007	MITOMYCIN		xxx	xxx
39	H1-39	22	1126	HIGH_10	3b	3	66	13/02/2007	GEMCITABINE		xxx	xxx
40	H1-40	4	1014	HIGH_1	2	3	65	20/07/2006	xxx		xxx	18/07/2006
41	CNT1-41	21	1122	CNT_9	**Hospitalised controls**		69	13/03/2007	xxx	Hypercholesterolemia	xxx	xxx
42	CNT1-42	17	1100	CNT_8			72	19/12/2006	xxx	Chronic sinusitis	xxx	xxx
43	CNT1-43	16	1092	CNT_7			64	11/12/2006	xxx	Chronic sinusitis	xxx	xxx
44	CNT1-44	13	1068	CNT_6			58	28/11/2006	xxx	Benign nasal polyps	xxx	xxx
45	CNT1-45	12	1054	CNT_5			72	24/10/2006	xxx	Otitis	xxx	xxx
46	CNT1-46	10	1043	CNT_4			58	04/10/2006	xxx	Benign nasal polyps	xxx	xxx
47	CNT1-47	9	1036	CNT_3			68	08/09/2006	xxx	Benign prostatic hyperplasia		xxx
48	CNT1-48	50	1423	CNT_20			75	09/03/2010	xxx	Chronic obstructive pulmonary disease		xxx
49	CNT1-49	3	1010	CNT_2			65	12/07/2006	xxx	Benign nasal polyps	xxx	xxx
50	CNT1-50	49	1422	CNT_19			73	08/03/2010	xxx	Sciatica	xxx	xxx
51	CNT1-51	48	1418	CNT_18			49	02/03/2010	xxx	Cystitis	xxx	xxx

52	CNT1-52	46	1415	CNT_17			65	01/03/2010	xxx	Prostatic Hyperplasia	xxx	xxx
53	CNT1-53	38	1214	CNT_16			73	12/11/2007	xxx	Chronic obstructive pulmonary disease		xxx
54	CNT1-54	36	1200	CNT_15			59	21/01/2008	xxx	Cystitis	xxx	xxx
55	CNT1-55	33	1187	CNT_14			70	05/12/2002	xxx	Aortic aneurysm	xxx	xxx
56	CNT1-56	32	1181	CNT_13			56	05/11/2007	xxx	Cystitis	xxx	xxx
57	CNT1-57	27	1156	CNT_12			60	18/04/2007	BCG	xxx	xxx	xxx
58	CNT1-58	24	1135	CNT_11			69	13/03/2007	xxx	xxx	xxx	xxx
59	CNT1-59	23	1128	CNT_10			67	16/02/2007	CISPLATIN+GEMCITABINE	xxx	xxx	xxx
60	CNT1-60	2	1002	CNT_1			50	04/07/2006	xxx	Benign nasal polyps	xxx	xxx
Second batch of samples												
61	L2-61	1	1057	G3Ta	a	3	74	07/11/2006	xxx	**Disease samples**	01/08/2008	xxx
62	L2-62	2	1096	G3Ta	a	3	73	22/01/2007	BCG		xxx	xxx
63	L2-63	3	1116	G3Ta	a	3	69	12/03/2007	BCG		xxx	xxx
64	L2-64	4	1138	G1Ta	a	1	66	17/04/2007	Mitomycin (local)		xxx	xxx
65	L2-65	5	1142	G1Ta	a	1	61	30/04/2007	Mitomycin (local)		05/03/2008	xxx
66	L2-66	6	1169	G1Ta	a	1	73	16/07/2007	BCG		xxx	xxx
67	L2-67	7	1170	G1Ta	a	1	59	17/09/2007	Mitomycin (local)		xxx	xxx
68	L2-68	8	1172	G3Ta	a	3	72	19/09/2007	BCG		xxx	xxx
69	L2-69	9	1173	G3Ta	a	3	74	24/09/2007	BCG		xxx	xxx
70	L2-70	10	1191	G1Ta	a	1	61	04/12/2007	Mitomycin (local)		21/10/2008	xxx
71	L2-71	11	1198	G1Ta	a	1	75	16/01/2008	xxx		xxx	xxx
72	L2-72	12	1203	G1Ta	a	1	74	04/02/2008	Mitomycin (local)		29/09/2009	xxx
73	L2-73	13	1231	G1Ta	a	1	51	28/04/2008	BCG + Mitomycin		29/07/2008	21/05/2009
74	L2-74	14	1259	G1Ta	a	1	62	01/07/2008	xxx		xxx	xxx
75	L2-75	15	1324	G1Ta	a	1	72	10/02/2009	xxx		xxx	xxx
76	L2-76	16	1345	G3T1	1	3	72	04/06/2009	BCG		xxx	xxx
77	L2-77	17	1365	G1Ta	a	1	73	20/07/2009	xxx		xxx	xxx
78	L2-78	18	1368	G1Ta	a	1	69	28/07/2009	xxx		xxx	xxx

79	L2-79	19	1522	G2T1	1	2	72	04/11/2010	xxx		xxx	xxx
80	H2-80	20	1532	G3T3	3	3	61	26/11/2010	Cisplatin + Gemcitabin (systemic) + Radiotherapy		xxx	28/11/2010
81	L2-81	21	1539	G3T1	1	3	64	20/12/2010	radiotherapy		xxx	25/07/2011
82	L2-82	22	1552	G2T1	1	2	60	02/02/2011	xxx		xxx	xxx
83	L2-83	23	1560	G1Ta	a	1	65	05/03/2011	xxx		xxx	xxx
84	L2-84	24	1580	G3T1	1	3	74	02/05/2011	Cisplatin + Gemcitabin (systemic)		xxx	xxx
85	CNT2-85	25	1229	CNT_1	**Healthy Controls**		67					
86	CNT2-86	26	1238	CNT_2			74					
87	CNT2-87	27	1241	CNT_3			68					
88	CNT2-88	28	1255	CNT_4			68					
89	CNT2-89	29	1298	CNT_5			66					
90	CNT2-90	30	1299	CNT_6			65					
91	CNT2-91	31	1318	CNT_7			74					
92	CNT2-92	32	1319	CNT_8			68					
93	CNT2-93	33	1341	CNT_9			60					
94	CNT2-94	34	1342	CNT_10			63					
95	CNT2-95	35	1343	CNT_11			59					
96	CNT2-96	36	1492	CNT_12			70					
97	CNT2-97	37	1493	CNT_13			63					
98	CNT2-98	38	1495	CNT_14			70					
99	CNT2-99	39	1511	CNT_15			73					
100	CNT2-100	40	1520	CNT_16			65					
101	CNT2-101	41	1521	CNT_17			72					
102	CNT2-102	42	1569	CNT_18			72					
103	CNT2-103	43	1604	CNT_19			70					
104	CNT2-104	44	1283	CNT_20			75					

8 References

*Cancer Research UK.*2013 [online]. Available at: www.cancerresearchuk.org [accessed 10/02 2013].

*Labome.*2013 [online]. Available at: http://www.labome.com/method/Protein-Quantitation.html2013].

News in Proteomics Research. 2013 [online]. Available at: http://proteomicsnews.blogspot.co.uk/ [accessed 2013].

*Nonlinear Dynamics.*2013 [online]. Available at: http://www.nonlinear.com/2013].

*Oxford Dictionary.*2013 [online]. Available at: http://oxforddictionaries.com/ [accessed 11/01 2013].

*U.S. Food and Drug Administration.*2013 [online]. Available at: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm182057.htm [accessed 15/02 2013].

*International Agency for Research on Cancer.*2012 [online]. Available at: http://www.iarc.fr/ [accessed 10/02 2012].

*Surveillance Epidemiology and End Results.*2010 [online]. Available at: http://seer.cancer.gov/ [accessed 07/05 2010].

Abdullah-Soheimi, S.S., Lim, B.K., Hashim, O.H., Shuib, A.S., 2010. Patients with ovarian carcinoma excrete different altered levels of urine CD59, kininogen-1 and fragments of inter-alpha-trypsin inhibitor heavy chain H4 and albumin. *Proteome Science.* **8,** 58-5956-8-58.

Abdul-Rahman, P.S., Lim, B.K., Hashim, O.H., 2007. Expression of high-abundance proteins in sera of patients with endometrial and cervical cancers: analysis using 2-DE with silver staining and lectin detection methods. *Electrophoresis.* **28,** 1989-1996.

Ahmed, N., Barker, G., Oliva, K.T., Hoffmann, P., Riley, C., Reeve, S., Smith, A.I., Kemp, B.E., Quinn, M.A., Rice, G.E., 2004. Proteomic-based identification of haptoglobin-1 precursor as a novel circulating biomarker of ovarian cancer. *British Journal of Cancer.* **91,** 129-140.

Allin, K.H. & Nordestgaard, B.G., 2011. Elevated C-reactive protein in the diagnosis, prognosis, and cause of cancer. *Critical Reviews in Clinical Laboratory Sciences.* **48,** 155-170.

Anderson, N.L., 2010. The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. *Clinical Chemistry.* **56,** 177-185.

Anderson, N.L. & Anderson, N.G., 2002. The human plasma proteome: history, character, and diagnostic prospects. *Molecular & Cellular Proteomics : MCP.* **1,** 845-867.

Ang, I.L., Poon, T.C., Lai, P.B., Chan, A.T., Ngai, S.M., Hui, A.Y., Johnson, P.J., Sung, J.J., 2006. Study of serum haptoglobin and its glycoforms in the diagnosis of hepatocellular carcinoma: a glycoproteomic approach. *Journal of Proteome Research.* **5,** 2691-2700.

Baker, S.G., Kramer, B.S., Srivastava, S., 2002. Markers for early detection of cancer: statistical guidelines for nested case-control studies. *BMC Medical Research Methodology.* **2,** 4.

Bakkar, A.A., Wallerand, H., Radvanyi, F., Lahaye, J.B., Pissard, S., Lecerf, L., Kouyoumdjian, J.C., Abbou, C.C., Pairon, J.C., Jaurand, M.C., Thiery, J.P., Chopin, D.K., de Medina, S.G., 2003. FGFR3 and TP53 gene mutations define two distinct pathways in urothelial cell carcinoma of the bladder. *Cancer Research.* **63,** 8108-8112.

Banks, R. & Selby, P., 2003. Clinical proteomics--insights into pathologies and benefits for patients. *Lancet.* **362,** 415-416.

Baum, A., Pohl, M., Kreusch, S., Cumme, G.A., Ditze, G., Misselwitz, J., Kiehntopf, M., Udby, L., Meier-Hellmann, A., Rhode, H., 2008. Searching biomarker candidates in serum using multidimensional native chromatography. II Method evaluation with Alport syndrome and severe inflammation. *Journal of Chromatography.B, Analytical Technologies in the Biomedical and Life Sciences.* **876,** 31-40.

Belanger, L., Roy, S., Allard, D., 1994. New albumin gene 3' adjacent to the alpha 1-fetoprotein locus. *The Journal of Biological Chemistry.* **269,** 5481-5484.

Berg, M., Parbel, A., Pettersen, H., Fenyo, D., Bjorkesten, L., 2006. Reproducibility of LC-MS-based protein identification. *Journal of Experimental Botany.* **57,** 1509-1514.

Bergkvist, A., Ljungqvist, A., Moberger, G., 1965. Classification of bladder tumours based on the cellular pattern. Preliminary report of a clinical-pathological study of 300 cases with a minimum follow-up of eight years. *Acta Chirurgica Scandinavica.* **130,** 371-378.

Bhoola, K.D., Figueroa, C.D., Worthy, K., 1992. Bioregulation of kinins: kallikreins, kininogens, and kininases. *Pharmacological Reviews.* **44,** 1-80.

Bing, C., Bao, Y., Jenkins, J., Sanders, P., Manieri, M., Cinti, S., Tisdale, M.J., Trayhurn, P., 2004. Zinc-alpha2-glycoprotein, a lipid mobilizing factor, is expressed in adipocytes and is up-regulated in mice with cancer cachexia. *Proceedings of the National Academy of Sciences of the United States of America.* **101,** 2500-2505.

Bischoff, C.J. & Clark, P.E., 2009. Bladder cancer. *Current Opinion in Oncology.* **21,** 272-277.

Bjerregaard, B.K., Raaschou-Nielsen, O., Sorensen, M., Frederiksen, K., Christensen, J., Tjonneland, A., Overvad, K., Chapelon, F.C., Nagel, G., Chang-Claude, J., Bergmann, M.M., Boeing, H., Trichopoulos, D., Trichopoulou, A., Oikonomou, E., Berrino, F., Palli, D., Tumino, R., Vineis, P., Panico, S., Peeters, P.H., Bueno-de-Mesquita, H.B., Kiemeney, L., Gram, I.T., Braaten, T., Lund, E., Gonzalez, C.A., Berglund, G., Allen, N., Roddam, A., Bingham, S., Riboli, E., 2006. Tobacco smoke and bladder cancer--in the European Prospective Investigation into Cancer and Nutrition. *International Journal of Cancer.Journal International Du Cancer.* **119,** 2412-2416.

Bjorhall, K., Miliotis, T., Davidsson, P., 2005. Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics.* **5,** 307-317.

Black, P.C., Brown, G.A., Dinney, C.P., 2006. Molecular markers of urothelial cancer and their use in the monitoring of superficial urothelial cancer. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology.* **24,** 5528-5535.

Botteman, M.F., Pashos, C.L., Redaelli, A., Laskin, B., Hauser, R., 2003. The health economics of bladder cancer: a comprehensive review of the published literature. *PharmacoEconomics.* **21,** 1315-1330.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry.* **72,** 248-254.

Bradley, K.A. & Young, J.A., 2003. Anthrax toxin receptor proteins. *Biochemical Pharmacology.* **65,** 309-314.

Brewis, I.A. & Brennan, P., 2010. Proteomics technologies for the global identification and quantification of proteins. *Advances in Protein Chemistry and Structural Biology.* **80,** 1-44.

Brock, A. & Brock, V., 1990. Plasma cholinesterase activity in a healthy population group with no occupational exposure to known cholinesterase inhibitors: relative influence of some factors related to normal inter- and intra-individual variations. *Scandinavian Journal of Clinical and Laboratory Investigation.* **50,** 401-408.

Broustas, C.G. & Lieberman, H.B., 2012. Contributions of Rad9 to tumorigenesis. *Journal of Cellular Biochemistry.* **113,** 742-751.

Bruno, R., Olivares, R., Berille, J., Chaikin, P., Vivier, N., Hammershaimb, L., Rhodes, G.R., Rigas, J.R., 2003. Alpha-1-acid glycoprotein as an independent predictor for treatment effects and a prognostic factor of survival in patients with non-small cell lung cancer treated with docetaxel. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research.* **9,** 1077-1082.

Bryant, J.W. & Shariat-Madar, Z., 2009. Human plasma kallikrein-kinin system: physiological and biochemical parameters. *Cardiovascular & Hematological Agents in Medicinal Chemistry.* **7,** 234-250.

Cairns, D.A., 2011. Statistical issues in quality control of proteomic analyses: good experimental design and planning. *Proteomics.* **11,** 1037-1048.

Callesen, A.K., Vach, W., Jorgensen, P.E., Cold, S., Mogensen, O., Kruse, T.A., Jensen, O.N., Madsen, J.S., 2008. Reproducibility of mass spectrometry based protein profiles for diagnosis of breast cancer across clinical studies: a systematic review. *Journal of Proteome Research.* **7,** 1395-1402.

Cao, Z., Tang, H.Y., Wang, H., Liu, Q., Speicher, D.W., 2012. Systematic Comparison of Fractionation Methods for In-depth Analysis of Plasma Proteomes. *Journal of Proteome Research.*

Chee, J., Naran, A., Misso, N.L., Thompson, P.J., Bhoola, K.D., 2008. Expression of tissue and plasma kallikreins and kinin B1 and B2 receptors in lung cancer. *Biological Chemistry.* **389,** 1225-1233.

Chen, Y.C., Pohl, G., Wang, T.L., Morin, P.J., Risberg, B., Kristensen, G.B., Yu, A., Davidson, B., Shih, I., 2005. Apolipoprotein E is required for cell proliferation and survival in ovarian cancer. *Cancer Research.* **65,** 331-337.

Chen, Y.T., Chen, C.L., Chen, H.W., Chung, T., Wu, C.C., Chen, C.D., Hsu, C.W., Chen, M.C., Tsui, K.H., Chang, P.L., Chang, Y.S., Yu, J.S., 2010. Discovery of novel bladder cancer biomarkers by comparative urine proteomics using iTRAQ technology. *Journal of Proteome Research.* **9,** 5803-5815.

Chen, Y.T., Chen, H.W., Domanski, D., Smith, D.S., Liang, K.H., Wu, C.C., Chen, C.L., Chung, T., Chen, M.C., Chang, Y.S., Parker, C.E., Borchers, C.H., Yu, J.S., 2012. Multiplexed quantification of 63 proteins in human urine by multiple reaction monitoring-based mass spectrometry for discovery of potential bladder cancer biomarkers. *Journal of Proteomics.* **75,** 3529-3545.

Chong, P.K., Lee, H., Zhou, J., Liu, S.C., Loh, M.C., Wang, T.T., Chan, S.P., Smoot, D.T., Ashktorab, H., So, J.B., Lim, K.H., Yeoh, K.G., Lim, Y.P., 2010. ITIH3 is a potential biomarker for early detection of gastric cancer. *Journal of Proteome Research.* **9,** 3671-3679.

Cohen, A., Wang, E., Chisholm, K.A., Kostyleva, R., O'Connor-McCourt, M., Pinto, D.M., 2013. A mass spectrometry-based plasma protein panel targeting the tumor microenvironment in patients with breast cancer. *Journal of Proteomics.* **81,** 135-147.

Colotta, F., Allavena, P., Sica, A., Garlanda, C., Mantovani, A., 2009. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis.* **30,** 1073-1081.

Cortes Ledesma, F., El Khamisy, S.F., Zuma, M.C., Osborn, K., Caldecott, K.W., 2009. A human 5'-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. *Nature.* **461,** 674-678.

Cotter, R.J., 1994. *Time-of-flight mass spectrometry.*

Coussens, L.M. & Werb, Z., 2002. Inflammation and cancer. *Nature.* **420,** 860-867.

Craig, R. & Beavis, R.C., 2004. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics (Oxford, England).* **20,** 1466-1467.

Czene, K., Tiikkaja, S., Hemminki, K., 2003. Cancer risks in hairdressers: assessment of carcinogenicity of hair dyes and gels. *International Journal of Cancer.Journal International Du Cancer.* **105,** 108-112.

Diamandis, E.P., 2012. The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? *BMC Medicine.* **10,** 87-7015-10-87.

Diamandis, E.P., 2010. Cancer biomarkers: can we turn recent failures into success? *Journal of the National Cancer Institute.* **102,** 1462-1467.

Dick, L.W.,Jr, Kim, C., Qiu, D., Cheng, K.C., 2007. Determination of the origin of the N-terminal pyro-glutamate variation in monoclonal antibodies using model peptides. *Biotechnology and Bioengineering.* **97,** 544-553.

Dieplinger, H., Ankerst, D.P., Burges, A., Lenhard, M., Lingenhel, A., Fineder, L., Buchner, H., Stieber, P., 2009. Afamin and apolipoprotein A-IV: novel protein markers for ovarian cancer. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* **18,** 1127-1133.

Dobryszycka, W., 1997. Biological functions of haptoglobin--new pieces to an old puzzle. *European Journal of Clinical Chemistry and Clinical Biochemistry : Journal of the Forum of European Clinical Chemistry Societies.* **35,** 647-654.

Dorschel, C., 2008. Effects of sample loading on protein identification.

Duncan, M.W., 2012. Good mass spectrometry and its place in good science. *Journal of Mass Spectrometry : JMS.* **47,** 795-809.

Echan, L.A., Tang, H.Y., Ali-Khan, N., Lee, K., Speicher, D.W., 2005. Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics.* **5,** 3292-3303.

Ehmann, M., Felix, K., Hartmann, D., Schnolzer, M., Nees, M., Vorderwulbecke, S., Bogumil, R., Buchler, M.W., Friess, H., 2007. Identification of potential markers for the detection of pancreatic cancer through comparative serum protein expression profiling. *Pancreas.* **34,** 205-214.

Eissa, S., Zohny, S.F., Zekri, A.R., El-Zayat, T.M., Maher, A.M., 2010. Diagnostic value of fibronectin and mutant p53 in the urine of patients with bladder cancer: impact on clinicopathological features and disease recurrence. *Medical Oncology (Northwood, London, England).* **27,** 1286-1294.

Eloranta, J.J., Wenger, C., Mwinyi, J., Hiller, C., Gubler, C., Vavricka, S.R., Fried, M., Kullak-Ublick, G.A., Swiss IBD Cohort Study Group, 2011. Association of a common vitamin D-binding protein polymorphism with inflammatory bowel disease. *Pharmacogenetics and Genomics.* **21,** 559-564.

Erlinger, T.P., Platz, E.A., Rifai, N., Helzlsouer, K.J., 2004. C-reactive protein and the risk of incident colorectal cancer. *JAMA : The Journal of the American Medical Association.* **291,** 585-590.

Farrah, T., Deutsch, E.W., Omenn, G.S., Campbell, D.S., Sun, Z., Bletz, J.A., Mallick, P., Katz, J.E., Malmstrom, J., Ossola, R., Watts, J.D., Lin, B., Zhang, H., Moritz, R.L., Aebersold, R., 2011. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Molecular & Cellular Proteomics : MCP.* **10,** M110.006353.

Fehniger, T.A. & Caligiuri, M.A., 2001. Interleukin 15: biology and relevance to human disease. *Blood.* **97,** 14-32.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M., 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, N.Y.).* **246,** 64-71.

Fiedler, G.M., Baumann, S., Leichtle, A., Oltmann, A., Kase, J., Thiery, J., Ceglarek, U., 2007. Standardized peptidome profiling of human urine by magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clinical Chemistry.* **53,** 421-428.

Fujimura, T., Shinohara, Y., Tissot, B., Pang, P.C., Kurogochi, M., Saito, S., Arai, Y., Sadilek, M., Murayama, K., Dell, A., Nishimura, S., Hakomori, S.I., 2008. Glycosylation status of haptoglobin in sera of patients with prostate cancer vs. benign prostate disease or normal subjects. *International Journal of Cancer.Journal International Du Cancer.* **122,** 39-49.

Gabay, C. & Kushner, I., 1999. Acute-phase proteins and other systemic responses to inflammation. *The New England Journal of Medicine.* **340,** 448-454.

Gao, Y., Yue, W., Zhang, P., Li, L., Xie, X., Yuan, H., Chen, L., Liu, D., Yan, F., Pei, X., 2005. Spindlin1, a novel nuclear protein with a role in the transformation of NIH3T3 cells. *Biochemical and Biophysical Research Communications.* **335,** 343-350.

Garcia-Closas, M., Malats, N., Real, F.X., Yeager, M., Welch, R., Silverman, D., Kogevinas, M., Dosemeci, M., Figueroa, J., Chatterjee, N., Tardon, A., Serra, C., Carrato, A., Garcia-Closas, R., Murta-Nascimento, C., Rothman, N., Chanock, S.J., 2007. Large-scale evaluation of candidate genes

identifies associations between VEGF polymorphisms and bladder cancer risk. *PLoS Genetics.* **3,** e29.

Gassmann, M., Grenacher, B., Rohde, B., Vogel, J., 2009. Quantifying Western blots: pitfalls of densitometry. *Electrophoresis.* **30,** 1845-1855.

Gatto, M., Iaccarino, L., Ghirardello, A., Bassi, N., Pontisso, P., Punzi, L., Shoenfeld, Y., Doria, A., 2013. Serpins, Immunity and Autoimmunity: Old Molecules, New Functions. *Clinical Reviews in Allergy & Immunology.*

Giles, K., Williams, J.P., Campuzano, I., 2011. Enhancements in travelling wave ion mobility resolution. *Rapid Communications in Mass Spectrometry : RCM.* **25,** 1559-1566.

Giusti, L., Iacconi, P., Ciregia, F., Giannaccini, G., Donatini, G.L., Basolo, F., Miccoli, P., Pinchera, A., Lucacchini, A., 2008. Fine-needle aspiration of thyroid nodules: proteomic analysis to identify cancer biomarkers. *Journal of Proteome Research.* **7,** 4079-4088.

Gomme, P.T. & Bertolini, J., 2004. Therapeutic potential of vitamin D-binding protein. *Trends in Biotechnology.* **22,** 340-345.

Gong, Y., Li, X., Yang, B., Ying, W., Li, D., Zhang, Y., Dai, S., Cai, Y., Wang, J., He, F., Qian, X., 2006. Different immunoaffinity fractionation strategies to characterize the human plasma proteome. *Journal of Proteome Research.* **5,** 1379-1387.

Goo, Y.A., Tsai, Y.S., Liu, A.Y., Goodlett, D.R., Yang, C.C., 2010. Urinary proteomics evaluation in interstitial cystitis/painful bladder syndrome: a pilot study. *International Braz j Urol : Official Journal of the Brazilian Society of Urology.* **36,** 464-78; discussion 478-9, 479.

Grignon, D.J., 2009. The current classification of urothelial neoplasms. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc.* **22 Suppl 2,** S60-9.

Grinnell, F., 1984. Fibronectin and wound healing. *Journal of Cellular Biochemistry.* **26,** 107-116.

Grou, C.P., Carvalho, A.F., Pinto, M.P., Wiese, S., Piechura, H., Meyer, H.E., Warscheid, B., Sa-Miranda, C., Azevedo, J.E., 2008. Members of the E2D (UbcH5) family mediate the ubiquitination of the conserved cysteine of Pex5p, the peroxisomal import receptor. *The Journal of Biological Chemistry.* **283,** 14190-14197.

Grundy, S.M. & Vega, G.L., 1990. Role of apolipoprotein levels in clinical practice. *Archives of Internal Medicine.* **150,** 1579-1582.

Gu, S.Z., Zhao, X.H., Quan, P., Li, S.B., Pan, B.R., 2005. Alterations of serum cholinesterase in patients with gastric cancer. *World Journal of Gastroenterology : WJG.* **11,** 4604-4606.

Gutman, S. & Kessler, L.G., 2006. The US Food and Drug Administration perspective on cancer biomarker development. *Nature Reviews.Cancer.* **6,** 565-571.

Guzzo, T.J., Bivalacqua, T.J., Schoenberg, M.P., 2008. Bladder cancer and the aluminium industry: a review. *BJU International.* **102,** 1058-1060.

Gygi, S.P., Rist, B., Griffin, T.J., Eng, J., Aebersold, R., 2002. Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *Journal of Proteome Research.* **1,** 47-54.

Hale, L.P., Price, D.T., Sanchez, L.M., Demark-Wahnefried, W., Madden, J.F., 2001. Zinc alpha-2-glycoprotein is expressed by malignant prostatic epithelium and may serve as a potential serum marker for prostate cancer. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research.* **7,** 846-853.

Han, S., Khuri, F.R., Roman, J., 2006. Fibronectin stimulates non-small cell lung carcinoma cell growth through activation of Akt/mammalian target of rapamycin/S6 kinase and inactivation of LKB1/AMP-activated protein kinase signal pathways. *Cancer Research.* **66,** 315-323.

Harling, M., Schablon, A., Schedlbauer, G., Dulon, M., Nienhaus, A., 2010. Bladder cancer among hairdressers: a meta-analysis. *Occupational and Environmental Medicine.* **67,** 351-358.

Hassan, M.I., Waheed, A., Yadav, S., Singh, T.P., Ahmad, F., 2008. Zinc alpha 2-glycoprotein: a multidisciplinary protein. *Molecular Cancer Research : MCR.* **6,** 892-906.

Hecht, S.S., 1999. Tobacco smoke carcinogens and lung cancer. *Journal of the National Cancer Institute.* **91,** 1194-1210.

Hegele, A., Heidenreich, A., Kropf, J., von Knobloch, R., Varga, Z., Hofmann, R., Olbert, P., 2004. Plasma levels of cellular fibronectin in patients with localized and metastatic renal cell carcinoma. *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine.* **25,** 111-116.

Hemminki, K. & Chen, B., 2006. Parental lung cancer as predictor of cancer risks in offspring: clues about multiple routes of harmful influence? *International Journal of Cancer.Journal International Du Cancer.* **118,** 744-748.

Heo, S.H., Lee, S.J., Ryoo, H.M., Park, J.Y., Cho, J.Y., 2007. Identification of putative serum glycoprotein biomarkers for human lung adenocarcinoma by multilectin affinity chromatography and LC-MS/MS. *Proteomics.* **7,** 4292-4302.

Hitchings, A.W., Kumar, M., Jordan, S., Nargund, V., Martin, J., Berney, D.M., 2004. Prediction of progression in pTa and pT1 bladder carcinomas with p53, p16 and pRb. *British Journal of Cancer.* **91,** 552-557.

Hoffmann, E.D. & Stroobant, V., 2006. *Mass Spectrometry Principles and Aplications.*

Holick, M.F., 2005. Vitamin D: important for prevention of osteoporosis, cardiovascular heart disease, type 1 diabetes, autoimmune diseases, and some cancers. *Southern Medical Journal.* **98,** 1024-1027.

Hong, S.K., Ko, D.W., Park, J., Kim, I.S., Doo, S.H., Yoon, C.Y., Park, H., Lee, W.K., Kim, D.S., Jeong, S.J., Byun, S.S., Lee, S.E., 2010. Alteration of Antithrombin III and D-dimer Levels in Clinically Localized Prostate Cancer. *Korean Journal of Urology.* **51,** 25-29.

Hortin, G.L. & Sviridov, D., 2010. The dynamic range problem in the analysis of the plasma proteome. *Journal of Proteomics.* **73,** 629-636.

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

Huang, Y., Li, L.Z., Zhang, C.Z., Yi, C., Liu, L.L., Zhou, X., Xie, G.B., Cai, M.Y., Li, Y., Yun, J.P., 2012. Decreased expression of zinc-alpha2-glycoprotein in hepatocellular carcinoma associates with poor prognosis. *Journal of Translational Medicine.* **10,** 106-5876-10-106.

Hwang, E.C., Choi, H.S., Jung, S.I., Kwon, D.D., Park, K., Ryu, S.B., 2011. Use of the NMP22 BladderChek test in the diagnosis and follow-up of urothelial cancer: a cross-sectional study. *Urology.* **77,** 154-159.

Irmak, S., Tilki, D., Heukeshoven, J., Oliveira-Ferrer, L., Friedrich, M., Huland, H., Ergun, S., 2005. Stage-dependent increase of orosomucoid and zinc-alpha2-glycoprotein in urinary bladder cancer. *Proteomics.* **5,** 4296-4304.

Ishioka, N., Takahashi, N., Putnam, F.W., 1986. Amino acid sequence of human plasma alpha 1B-glycoprotein: homology to the immunoglobulin supergene family. *Proceedings of the National Academy of Sciences of the United States of America.* **83,** 2363-2367.

Issaq, H. & Veenstra, T., 2008. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. *BioTechniques.* **44,** 697-8, 700.

Ito, H., Shioi, K., Murakami, T., Takizawa, A., Sano, F., Kawahara, T., Mizuno, N., Makiyama, K., Nakaigawa, N., Kishida, T., Miura, T., Kubota, Y., Yao, M., 2012. C-reactive protein in patients with advanced metastatic renal cell carcinoma: usefulness in identifying patients most likely to benefit from initial nephrectomy. *BMC Cancer.* **12,** 337-2407-12-337.

Jackson, D., Craven, R.A., Hutson, R.C., Graze, I., Lueth, P., Tonge, R.P., Hartley, J.L., Nickson, J.A., Rayner, S.J., Johnston, C., Dieplinger, B., Hubalek, M., Wilkinson, N., Perren, T.J., Kehoe, S., Hall, G.D., Daxenbichler, G., Dieplinger, H., Selby, P.J., Banks, R.E., 2007. Proteomic profiling identifies afamin as a potential biomarker for ovarian cancer. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research.* **13,** 7370-7379.

James, N.D., Hussain, S.A., Hall, E., Jenkins, P., Tremlett, J., Rawlings, C., Crundwell, M., Sizer, B., Sreenivasan, T., Hendron, C., Lewis, R., Waters, R., Huddart, R.A., BC2001 Investigators, 2012. Radiotherapy with or without chemotherapy in muscle-invasive bladder cancer. *The New England Journal of Medicine.* **366,** 1477-1488.

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Thun, M.J., 2009. Cancer Statistics, 2009. **59,** 225-249.

Jeong, D.H., Kim, H.K., Prince, A.E., Lee, D.S., Kim, Y.N., Han, J., Kim, K.T., 2008. Plasma proteomic analysis of patients with squamous cell carcinoma of the uterine cervix. *Journal of Gynecologic Oncology.* **19,** 173-180.

Jerkovic, L., Voegele, A.F., Chwatal, S., Kronenberg, F., Radcliffe, C.M., Wormald, M.R., Lobentanz, E.M., Ezeh, B., Eller, P., Dejori, N., Dieplinger, B., Lottspeich, F., Sattler, W., Uhr, M., Mechtler, K., Dwek, R.A., Rudd, P.M., Baier, G., Dieplinger, H., 2005. Afamin is a novel human vitamin E-binding glycoprotein characterization and in vitro expression. *Journal of Proteome Research.* **4,** 889-899.

Jiang, X., Castelao, J.E., Groshen, S., Cortessis, V.K., Shibata, D.K., Conti, D.V., Gago-Dominguez, M., 2008. Water intake and bladder cancer risk in Los Angeles County. *International Journal of Cancer.Journal International Du Cancer.* **123,** 1649-1656.

Junker, K., van Oers, J.M., Zwarthoff, E.C., Kania, I., Schubert, J., Hartmann, A., 2008. Fibroblast growth factor receptor 3 mutations in bladder tumors correlate with low frequency of chromosome alterations. *Neoplasia (New York, N.Y.).* **10,** 1-7.

Kang, U.B., Ahn, Y., Lee, J.W., Kim, Y.H., Kim, J., Yu, M.H., Noh, D.Y., Lee, C., 2010. Differential profiling of breast cancer plasma proteome by isotope-coded affinity tagging method reveals biotinidase as a breast cancer biomarker. *BMC Cancer.* **10,** 114.

Kay, R., Barton, C., Ratcliffe, L., Matharoo-Ball, B., Brown, P., Roberts, J., Teale, P., Creaser, C., 2008. Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis. *Rapid Communications in Mass Spectrometry : RCM.* **22,** 3255-3260.

Kelsh, M.A., Alexander, D.D., Kalmes, R.M., Buffler, P.A., 2008. Personal use of hair dyes and risk of bladder cancer: a meta-analysis of epidemiologic data. *Cancer Causes & Control : CCC.* **19,** 549-558.

Kim, E., Li, K., Lieu, C., Tong, S., Kawai, S., Fukutomi, T., Zhou, Y., Wands, J., Li, J., 2008. Expression of apolipoprotein C-IV is regulated by Ku antigen/peroxisome proliferator-activated receptor gamma complex and correlates with liver steatosis. *Journal of Hepatology.* **49,** 787-798.

Kim, Y.S., Gu, B.H., Choi, B.C., Kim, M.S., Song, S., Yun, J.H., Chung, M.K., Choi, C.H., Baek, K.H., 2013. Apolipoprotein A-IV as a novel gene associated with polycystic ovary syndrome. *International Journal of Molecular Medicine.* **31,** 707-716.

Kirkali, Z., Chan, T., Manoharan, M., Algaba, F., Busch, C., Cheng, L., Kiemeney, L., Kriegmair, M., Montironi, R., Murphy, W.M., Sesterhenn, I.A., Tachibana, M., Weider, J., 2005. Bladder cancer: epidemiology, staging and grading, and diagnosis. *Urology.* **66,** 4-34.

Klose, J., 1975. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik.* **26,** 231-243.

Knowles, M.A., 2007. Role of FGFR3 in urothelial cell carcinoma: biomarker and potential therapeutic target. *World Journal of Urology.* **25,** 581-593.

Kogevinas, M., 't Mannetje, A., Cordier, S., Ranft, U., Gonzalez, C.A., Vineis, P., Chang-Claude, J., Lynge, E., Wahrendorf, J., Tzonou, A., Jockel, K.H., Serra, C., Porru, S., Hours, M., Greiser, E., Boffetta, P., 2003. Occupation and bladder cancer among men in Western Europe. *Cancer Causes & Control : CCC.* **14,** 907-914.

Korkolopoulou, P., Givalos, N., Saetta, A., Goudopoulou, A., Gakiopoulou, H., Thymara, I., Thomas-Tsagli, E., Patsouris, E., 2005. Minichromosome maintenance proteins 2 and 5 expression in muscle-invasive urothelial cancer: a multivariate survival study including proliferation markers and cell cycle regulators. *Human Pathology.* **36,** 899-907.

Kossowska, B., Ferens-Sieczkowska, M., Gancarz, R., Passowicz-Muszynska, E., Jankowska, R., 2005. Fucosylation of serum glycoproteins in lung cancer patients. *Clinical Chemistry and Laboratory Medicine : CCLM / FESCC.* **43,** 361-369.

Kotite, L., Zhang, L.H., Yu, Z., Burlingame, A.L., Havel, R.J., 2003. Human apoC-IV: isolation, characterization, and immunochemical quantification in plasma and plasma lipoproteins. *Journal of Lipid Research.* **44,** 1387-1394.

Kotwal, S., Choudhury, A., Johnston, C., Paul, A.B., Whelan, P., Kiltie, A.E., 2008. Similar treatment outcomes for radical cystectomy and radical radiotherapy in invasive bladder cancer treated at a United Kingdom specialist treatment center. *International Journal of Radiation Oncology, Biology, Physics.* **70,** 456-463.

Koutroukides, T.A., Guest, P.C., Leweke, F.M., Bailey, D.M., Rahmoune, H., Bahn, S., Martins-de-Souza, D., 2011. Characterization of the human serum depletome by label-free shotgun proteomics. *Journal of Separation Science.* **34,** 1621-1626.

Kramer, C., Klasmeyer, K., Bojar, H., Schulz, W.A., Ackermann, R., Grimm, M.O., 2007. Heparin-binding epidermal growth factor-like growth factor isoforms and epidermal growth factor receptor/ErbB1 expression in bladder cancer and their relation to clinical outcome. *Cancer.* **109,** 2016-2024.

Kreunin, P., Zhao, J., Rosser, C., Urquidi, V., Lubman, D.M., Goodison, S., 2007. Bladder cancer associated glycoprotein signatures revealed by urinary proteomic profiling. *Journal of Proteome Research.* **6,** 2631-2639.

Kruger, S., Weitsch, G., Buttner, H., Matthiensen, A., Bohmer, T., Marquardt, T., Sayk, F., Feller, A.C., Bohle, A., 2002. HER2 overexpression in muscle-invasive urothelial carcinoma of the bladder: prognostic implications. *International Journal of Cancer.Journal International Du Cancer.* **102,** 514-518.

Kurien, B.T., Dorri, Y., Dillon, S., Dsouza, A., Scofield, R.H., 2011. An overview of Western blotting for determining antibody specificities for immunohistochemistry. *Methods in Molecular Biology (Clifton, N.J.).* **717,** 55-67.

La, D.K. & Swenberg, J.A., 1996. DNA adducts: biological markers of exposure and potential applications to risk assessment. *Mutation Research.* **365,** 129-146.

Lane, D.P., 1992. Cancer. p53, guardian of the genome. *Nature.* **358,** 15-16.

Langbein, S., Lehmann, J., Harder, A., Steidler, A., Michel, M.S., Alken, P., Badawi, J.K., 2006. Protein profiling of bladder cancer using the 2D-PAGE and SELDI-TOF-MS technique. *Technology in Cancer Research & Treatment.* **5,** 67-72.

Laufer, M., Kaver, I., Sela, B., Matzkin, H., 1999. Elevated urinary fibronectin levels after transurethral resection of bladder tumour: a possible role in patients failing therapy with bacillus Calmette-Guerin. *BJU International.* **84,** 428-432.

Lei, T., Zhao, X., Jin, S., Meng, Q., Zhou, H., Zhang, M., 2012. Discovery of Potential Bladder Cancer Biomarkers by Comparative Urine Proteomics and Analysis. *Clinical Genitourinary Cancer.*

Lerner, S.P., Grossman, H.B., Messing, E.M., Kibel, A.S., Stephenson, A., Gee, J.R., O'Donnell, M.A., Reid, R.D., Kamat, A.M., Parnes, H.L., House, M.G., 2010. BCAN Think Tank session 3: Prevention of bladder cancer. *Urologic Oncology.* **28,** 338-342.

Lerner, S.P., Schoenberg, M.P. and Sternberg, C.N., eds, 2008. *Treatment and Managment of Bladder Cancer.*

Levin, Y., Schwarz, E., Wang, L., Leweke, F.M., Bahn, S., 2007. Label-free LC-MS/MS quantitative proteomics for large-scale biomarker discovery in complex samples. *Journal of Separation Science.* **30,** 2198-2203.

Li, G.Z., Vissers, J.P., Silva, J.C., Golick, D., Gorenstein, M.V., Geromanos, S.J., 2009. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics.* **9,** 1696-1719.

Li, H., Li, C., Wu, H., Zhang, T., Wang, J., Wang, S., Chang, J., 2011. Identification of Apo-A1 as a biomarker for early diagnosis of bladder transitional cell carcinoma. *Proteome Science.* **9,** 21.

Li, J., Abraham, S., Cheng, L., Witzmann, F.A., Koch, M., Xie, J., Rahman, M., Mohammed, S.I., 2008. Proteomic-based approach for biomarkers discovery in early detection of invasive urothelial carcinoma. *Proteomics.Clinical Applications.* **2,** 78-89.

Lichenstein, H.S., Lyons, D.E., Wurfel, M.M., Johnson, D.A., McGinley, M.D., Leidli, J.C., Trollinger, D.B., Mayer, J.P., Wright, S.D., Zukowski, M.M., 1994. Afamin is a new member of the albumin, alpha-fetoprotein, and vitamin D-binding protein gene family. *The Journal of Biological Chemistry.* **269,** 18149-18154.

Liebler, D.C., 2004. Shotgun mass spec goes independent. *Nature Methods.* **1,** 16-17.

Linden, M., Lind, S.B., Mayrhofer, C., Segersten, U., Wester, K., Lyutvinskiy, Y., Zubarev, R., Malmstrom, P.U., Pettersson, U., 2012. Proteomic analysis of urinary biomarker candidates for nonmuscle invasive bladder cancer. *Proteomics.* **12,** 135-144.

Lindgren, D., Liedberg, F., Andersson, A., Chebil, G., Gudjonsson, S., Borg, A., Mansson, W., Fioretos, T., Hoglund, M., 2006. Molecular characterization of early-stage bladder carcinomas by expression profiles, FGFR3 mutation status, and loss of 9q. *Oncogene.* **25,** 2685-2696.

Liu, W., Guan, M., Wu, D., Zhang, Y., Wu, Z., Xu, M., Lu, Y., 2005. Using tree analysis pattern and SELDI-TOF-MS to discriminate transitional cell carcinoma of the bladder cancer from noncancer patients. *European Urology.* **47,** 456-462.

Liu, W., Liu, B., Cai, Q., Li, J., Chen, X., Zhu, Z., 2012. Proteomic identification of serum biomarkers for gastric cancer using multi-dimensional liquid chromatography and 2D differential gel electrophoresis. *Clinica Chimica Acta; International Journal of Clinical Chemistry.* **413,** 1098-1106.

Loehrer PJ, S., Einhorn, L.H., Elson, P.J., Crawford, E.D., Kuebler, P., Tannock, I., Raghavan, D., Stuart-Harris, R., Sarosdy, M.F., Lowe, B.A., 1992. A randomized comparison of cisplatin alone or in combination with methotrexate, vinblastine, and doxorubicin in patients with metastatic urothelial carcinoma: a cooperative group study. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology.* **10,** 1066-1073.

Lokeshwar, V.B., Merseburger, A.S. and Hautmann, S.H., eds, 2011. *Bladder Tumors, Molecular Aspects and Clinical Management.* New York: Springer.

Lopez-Ferrer, D., Capelo, J.L., Vazquez, J., 2005. Ultra fast trypsin digestion of proteins by high intensity focused ultrasound. *Journal of Proteome Research.* **4,** 1569-1574.

Lopez-Ferrer, D., Petritis, K., Hixson, K.K., Heibeck, T.H., Moore, R.J., Belov, M.E., Camp, D.G.,2nd, Smith, R.D., 2008. Application of pressurized solvents for ultrafast trypsin hydrolysis in proteomics: proteomics on the fly. *Journal of Proteome Research.* **7,** 3276-3281.

LOWRY, O.H., ROSEBROUGH, N.J., FARR, A.L., RANDALL, R.J., 1951. Protein measurement with the Folin phenol reagent. *The Journal of Biological Chemistry.* **193,** 265-275.

Luque-Garcia, J.L. & Neubert, T.A., 2007. Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. *Journal of Chromatography.A.* **1153,** 259-276.

Maiolica, A., Junger, M.A., Ezkurdia, I., Aebersold, R., 2012. Targeted proteome investigation via selected reaction monitoring mass spectrometry. *Journal of Proteomics.* **75,** 3495-3513.

Mann, M., 2009. Comparative analysis to guide quality improvements in proteomics. *Nature Methods.* **6,** 717-719.

Mann, M. & Wilm, M., 1994. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical Chemistry.* **66,** 4390-4399.

Margulis, V., Lotan, Y., Karakiewicz, P.I., Fradet, Y., Ashfaq, R., Capitanio, U., Montorsi, F., Bastian, P.J., Nielsen, M.E., Muller, S.C., Rigaud, J., Heukamp, L.C., Netto, G., Lerner, S.P., Sagalowsky, A.I., Shariat, S.F., 2009. Multi-institutional validation of the predictive value of Ki-67 labeling index in patients with urinary bladder cancer. *Journal of the National Cancer Institute.* **101,** 114-119.

Matsuura, H. & Nakazawa, S., 1985. Prognostic significance of serum alpha 1-acid glycoprotein in patients with glioblastoma multiforme: a preliminary communication. *Journal of Neurology, Neurosurgery, and Psychiatry.* **48,** 835-837.

Matta, E.J., Kenney, A.J., Barré, G.M., Vanlangendonck Jr, R.M., 2005. Intradiverticular Bladder Carcinoma. **25,** 1397-1403.

MELICOW, M.M., 1952. Histological study of vesical urothelium intervening between gross neoplasms in total cystectomy. *The Journal of Urology.* **68,** 261-279.

Mellon, J.K., Lunec, J., Wright, C., Horne, C.H., Kelly, P., Neal, D.E., 1996. C-erbB-2 in bladder cancer: molecular biology, correlation with epidermal growth factor receptors and prognostic value. *The Journal of Urology.* **155,** 321-326.

Melvin, J.C., Seth, D., Holmberg, L., Garmo, H., Hammar, N., Jungner, I., Walldius, G., Lambe, M., Wigertz, A., Van Hemelrijck, M., 2012. Lipid profiles and risk of breast and ovarian cancer in the Swedish AMORIS study. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* **21,** 1381-1384.

Menendez, V., Fernandez-Suarez, A., Galan, J.A., Perez, M., Garcia-Lopez, F., 2005. Diagnosis of bladder cancer by analysis of urinary fibronectin. *Urology.* **65,** 284-289.

Millioni, R., Tolin, S., Puricelli, L., Sbrignadello, S., Fadini, G.P., Tessari, P., Arrigoni, G., 2011. High abundance proteins depletion vs low abundance proteins enrichment: comparison of methods to reduce the plasma proteome complexity. *PloS One.* **6,** e19603.

Mirkovic, K., Doorenbos, C.R., Dam, W.A., Lambers Heerspink, H.J., Slagman, M.C., Nauta, F.L., Kramer, A.B., Gansevoort, R.T., van den Born, J., Navis, G., de Borst, M.H., 2013. Urinary vitamin d binding protein: a potential novel marker of renal interstitial inflammation and fibrosis. *PloS One.* **8,** e55887.

Mitra, A.P. & Cote, R.J., 2009. Molecular pathogenesis and diagnostics of bladder cancer. *Annual Review of Pathology.* **4,** 251-285.

Miyazawa, K., 2010. Hepatocyte growth factor activator (HGFA): A serine protease that links tissue injury to activation of hepatocyte growth factor. *FEBS Journal.* **277,** 2208-2214.

Mondul, A.M., Weinstein, S.J., Mannisto, S., Snyder, K., Horst, R.L., Virtamo, J., Albanes, D., 2010. Serum vitamin D and risk of bladder cancer. *Cancer Research.* **70,** 9218-9223.

Mondul, A.M., Weinstein, S.J., Virtamo, J., Albanes, D., 2012. Influence of vitamin D binding protein on the association between circulating vitamin D and risk of bladder cancer. *British Journal of Cancer.* **107,** 1589-1594.

Montironi, R. & Lopez-Beltran, A., 2005. The 2004 WHO classification of bladder tumors: a summary and commentary. *International Journal of Surgical Pathology.* **13,** 143-153.

Moore, C., 2009. *Introduction to western blotting.* [e-book]. UK: AbD serotec-MorphoSys UK Ltd. Available from : http://static.abdserotec.com/Lit-pdfs/Brochures1/westernblotbook.pdf.

Mostafa, M.H., Sheweita, S.A., O'Connor, P.J., 1999. Relationship between schistosomiasis and bladder cancer. *Clinical Microbiology Reviews.* **12,** 97-111.

Mostofi, F., 1973. *Histological Typing of Urinary Bladder Tumours.* Geneva: World Health Organisation.

Munro, N.P., Cairns, D.A., Clarke, P., Rogers, M., Stanley, A.J., Barrett, J.H., Harnden, P., Thompson, D., Eardley, I., Banks, R.E., Knowles, M.A., 2006. Urinary biomarker profiling in transitional cell carcinoma. *International Journal of Cancer.Journal International Du Cancer.* **119,** 2642-2650.

Naik, D.S., Sharma, S., Ray, A., Hedau, S., 2011. Epidermal growth factor receptor expression in urinary bladder cancer. *Indian Journal of Urology : IJU : Journal of the Urological Society of India.* **27,** 208-214.

Niu, H.T., Dong, Z., Jiang, G., Xu, T., Liu, Y.Q., Cao, Y.W., Zhao, J., Wang, X.S., 2011. Proteomics research on muscle-invasive bladder transitional cell carcinoma. *Cancer Cell International.* **11,** 17.

Oezdemir, R.F., Gaisa, N.T., Lindemann-Docter, K., Gostek, S., Weiskirchen, R., Ahrens, M., Schwamborn, K., Stephan, C., Pfister, D., Heidenreich, A., Knuechel, R., Henkel, C., 2012. Proteomic tissue profiling for the improvement of grading of noninvasive papillary urothelial neoplasia. *Clinical Biochemistry.* **45,** 7-11.

O'Farrell, P.H., 1975. High resolution two-dimensional electrophoresis of proteins. *The Journal of Biological Chemistry.* **250,** 4007-4021.

Olsen, J.V., Ong, S.E., Mann, M., 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Molecular & Cellular Proteomics : MCP.* **3,** 608-614.

Omenn, G.S., 2006. Strategies for plasma proteomic profiling of cancers. *Proteomics.* **6,** 5662-5673.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., Mann, M., 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics : MCP.* **1,** 376-386.

Orenes-Pinero, E., Barderas, R., Rico, D., Casal, J.I., Gonzalez-Pisano, D., Navajo, J., Algaba, F., Piulats, J.M., Sanchez-Carbayo, M., 2010. Serum and tissue profiling in bladder cancer combining protein and tissue arrays. *Journal of Proteome Research.* **9,** 164-173.

Pan, S., Chen, R., Crispin, D.A., May, D., Stevens, T., McIntosh, M.W., Bronner, M.P., Ziogas, A., Anton-Culver, H., Brentnall, T.A., 2011. Protein alterations associated with pancreatic cancer and chronic pancreatitis found in human plasma using global quantitative proteomics profiling. *Journal of Proteome Research.* **10,** 2359-2376.

Pankov, R. & Yamada, K.M., 2002. Fibronectin at a glance. *Journal of Cell Science.* **115,** 3861-3863.

Parekh, D.J., Bochner, B.H., Dalbagni, G., 2006. Superficial and muscle-invasive bladder cancer: principles of management for outcomes assessments. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology.* **24,** 5519-5527.

Paul, W. & H. Steinwedel, 1953. Ein neues Massenspektrometer ohne magnetfield . **8,** 488.

Pavlou, M.P., Diamandis, E.P., Blasutig, I.M., 2013. The long journey of cancer biomarkers from the bench to the clinic. *Clinical Chemistry.* **59,** 147-157.

Pejcic, M., Stojnev, S., Stefanovic, V., 2010. Urinary proteomics--a tool for biomarker discovery. *Renal Failure.* **32,** 259-268.

Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C., Liotta, L.A., 2002. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet.* **359,** 572-577.

Picotti, P. & Aebersold, R., 2012. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods.* **9,** 555-566.

Pin, E., Fredolini, C., Petricoin, E.F.,3rd, 2013. The role of proteomics in prostate cancer research: Biomarker discovery and validation. *Clinical Biochemistry.* **46,** 524-538.

Pramanik, B.N., Mirza, U.A., Ing, Y.H., Liu, Y.H., Bartner, P.L., Weber, P.C., Bose, A.K., 2002. Microwave-enhanced enzyme reaction for protein mapping by mass spectrometry: a new approach to protein digestion in minutes. *Protein Science : A Publication of the Protein Society.* **11,** 2676-2687.

Quilty, P.M., Duncan, W., Chisholm, G.D., Fowler, J.W., Hargreave, T.B., Newsam, J.E., Tolley, D.A., 1986. Results of surgery following radical radiotherapy for invasive bladder cancer. *British Journal of Urology.* **58,** 396-405.

Raghavan, D., 2003. Progress in the chemotherapy of metastatic cancer of the urinary tract. *Cancer.* **97,** 2050-2055.

Rai, A.J., ed, 2010. *The Urinary Proteome.* Springer.

Raitanen, M.P., Marttila, T., Nurmi, M., Ala-Opas, M., Nieminen, P., Aine, R., Tammela, T.L., Finnbladder Group, 2001. Human complement factor H related protein test for monitoring bladder cancer. *The Journal of Urology.* **165,** 374-377.

Ravishankaran, P. & Karunanithi, R., 2011. Clinical significance of preoperative serum interleukin-6 and C-reactive protein level in breast cancer patients. *World Journal of Surgical Oncology.* **9,** 18-7819-9-18.

Ray, G., Henson, D.E., Schwartz, A.M., 2010. Cigarette Smoking as a Cause of Cancers Other Than Lung Cancer: An Exploratory Study Using the SEER Program. *Chest.*

Reinders, J. and Sickmann, A., eds, 2009. *Proteomics.* Springer.

Ren, D., Pipes, G.D., Liu, D., Shih, L.Y., Nichols, A.C., Treuheit, M.J., Brems, D.N., Bondarenko, P.V., 2009. An improved trypsin digestion method minimizes digestion-induced modifications on proteins. *Analytical Biochemistry.* **392,** 12-21.

Reulen, R.C., Kellen, E., Buntinx, F., Brinkman, M., Zeegers, M.P., 2008. A meta-analysis on the association between bladder cancer and occupation. *Scandinavian Journal of Urology and Nephrology.Supplementum.* **(218),** 64-78.

Rifai, N., Gillette, M.A., Carr, S.A., 2006. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnology.* **24,** 971-983.

Roche, S., Tiers, L., Provansal, M., Seveno, M., Piva, M.T., Jouin, P., Lehmann, S., 2009. Depletion of one, six, twelve or twenty major blood proteins before proteomic analysis: the more the better? *Journal of Proteomics.* **72,** 945-951.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., Pappin, D.J., 2004. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics : MCP.* **3,** 1154-1169.

Rundle, A., Ahsan, H., Vineis, P., 2012. Better cancer biomarker discovery through better study design. *European Journal of Clinical Investigation.* **42,** 1350-1359.

Sakashita, K., Tanaka, F., Zhang, X., Mimori, K., Kamohara, Y., Inoue, H., Sawada, T., Hirakawa, K., Mori, M., 2008. Clinical significance of ApoE expression in human gastric cancer. *Oncology Reports.* **20,** 1313-1319.

Sala, M., Cordier, S., Chang-Claude, J., Donato, F., Escolar-Pujolar, A., Fernandez, F., Gonzalez, C.A., Greiser, E., Jockel, K.H., Lynge, E., Mannetje, A., Pohlabeln, H., Porru, S., Serra, C., Tzonou, A., Vineis, P., Wahrendorf, J., Boffetta, P., Kogevina, M., 2000. Coffee consumption and bladder cancer in nonsmokers: a pooled analysis of case-control studies in European countries. *Cancer Causes & Control : CCC.* **11,** 925-931.

Salasznyk, R.M., Westcott, A.M., Klees, R.F., Ward, D.F., Xiang, Z., Vandenberg, S., Bennett, K., Plopper, G.E., 2005. Comparing the protein expression profiles of human mesenchymal stem cells and human osteoblasts using gene ontologies. *Stem Cells and Development.* **14,** 354-366.

Samanic, C., Kogevinas, M., Dosemeci, M., Malats, N., Real, F.X., Garcia-Closas, M., Serra, C., Carrato, A., Garcia-Closas, R., Sala, M., Lloreta, J., Tardon, A., Rothman, N., Silverman, D.T., 2006. Smoking and bladder cancer in Spain: effects of tobacco type, timing, environmental tobacco smoke, and gender. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* **15,** 1348-1354.

Sangar, V.K., Ragavan, N., Matanhelia, S.S., Watson, M.W., Blades, R.A., 2005. The economic consequences of prostate and bladder cancer in the UK. *BJU International.* **95,** 59-63.

Sant, M., Allemani, C., Santaquilani, M., Knijn, A., Marchesi, F., Capocaccia, R., EUROCARE Working Group, 2009. EUROCARE-4. Survival of cancer patients diagnosed in 1995-1999. Results and commentary. *European Journal of Cancer (Oxford, England : 1990).* **45,** 931-991.

Sayan, A.E., Griffiths, T.R., Pal, R., Browne, G.J., Ruddick, A., Yagci, T., Edwards, R., Mayer, N.J., Qazi, H., Goyal, S., Fernandez, S., Straatman, K., Jones, G.D., Bowman, K.J., Colquhoun, A., Mellon, J.K.,

Kriajevska, M., Tulchinsky, E., 2009. SIP1 protein protects cells from DNA damage-induced apoptosis and has independent prognostic value in bladder cancer. *Proceedings of the National Academy of Sciences of the United States of America.* **106,** 14884-14889.

Schaub, N.P., Jones, K.J., Nyalwidhe, J.O., Cazares, L.H., Karbassi, I.D., Semmes, O.J., Feliberti, E.C., Perry, R.R., Drake, R.R., 2009. Serum proteomic biomarker discovery reflective of stage and obesity in breast cancer patients. *Journal of the American College of Surgeons.* **208,** 970-8; discussion 978-80.

Schaub, S., Wilkins, J., Weiler, T., Sangster, K., Rush, D., Nickerson, P., 2004. Urine protein profiling with surface-enhanced laser-desorption/ionization time-of-flight mass spectrometry. *Kidney International.* **65,** 323-332.

Schiess, R., Wollscheid, B., Aebersold, R., 2009. Targeted proteomic strategy for clinical biomarker discovery. *Molecular Oncology.* **3,** 33-44.

Schlichtholz, B., Presler, M., Matuszewski, M., 2004. Clinical implications of p53 mutation analysis in bladder cancer tissue and urine sediment by functional assay in yeast. *Carcinogenesis.* **25,** 2319-2323.

Schwamborn, K., Krieg, R.C., Grosse, J., Reulen, N., Weiskirchen, R., Knuechel, R., Jakse, G., Henkel, C., 2009. Serum proteomic profiling in patients with bladder cancer. *European Urology.* **56,** 989-996.

Searle, B.C., Turner, M., Nesvizhskii, A.I., 2008. Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies. *Journal of Proteome Research.* **7,** 245-253.

Shah, A., Singh, H., Sachdev, V., Lee, J., Yotsukura, S., Salgia, R., Bharti, A., 2010. Differential Serum Level of Specific Haptoglobin Isoforms in Small Cell Lung Cancer. *Current Proteomics.* **7,** 49-65.

Shao, C., Wang, Y., Gao, Y., 2011. Applications of urinary proteomics in biomarker discovery. *Science China.Life Sciences.* **54,** 409-417.

Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P., Geromanos, S.J., 2006. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Molecular & Cellular Proteomics : MCP.* **5,** 144-156.

Simpson, D.M. & Beynon, R.J., 2010. Acetone precipitation of proteins and the modification of peptides. *Journal of Proteome Research.* **9,** 444-450.

Skates, S. & Iliopoulos, O., 2004. Molecular markers for early detection of renal carcinoma: investigative approach. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research.* **10,** 6296S-301S.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., Klenk, D.C., 1985. Measurement of protein using bicinchoninic acid. *Analytical Biochemistry.* **150,** 76-85.

Sonoda, K., Miyamoto, S., Nakashima, M., Wake, N., 2010. Receptor-binding cancer antigen expressed on SiSo cells induces apoptosis via ectodomain shedding. *Experimental Cell Research.* **316,** 1795-1803.

Steffens, S., Kohler, A., Rudolph, R., Eggers, H., Seidel, C., Janssen, M., Wegener, G., Schrader, M., Kuczyk, M.A., Schrader, A.J., 2012. Validation of CRP as prognostic marker for renal cell carcinoma in a large series of patients. *BMC Cancer.* **12,** 399-2407-12-399.

Stephens, W.E., 1946. A Pulsed Mass Spectrometer with Time Dispersion. Proceedings of the American Physical Society. Cambridge, MA., USA. . **69,** 691.

Sun, W., Gao, S., Wang, L., Chen, Y., Wu, S., Wang, X., Zheng, D., Gao, Y., 2006. Microwave-assisted protein preparation and enzymatic digestion in proteomics. *Molecular & Cellular Proteomics : MCP.* **5,** 769-776.

Suzuki, S., Yamayoshi, Y., Nishimuta, A., Tanigawara, Y., 2011. S100A10 protein expression is associated with oxaliplatin sensitivity in human colorectal cancer cells. *Proteome Science.* **9,** 76-5956-9-76.

Sylvester, R.J., van der Meijden, A.P., Oosterlinck, W., Witjes, J.A., Bouffioux, C., Denis, L., Newling, D.W., Kurth, K., 2006. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *European Urology.* **49,** 466-5; discussion 475-7.

Tabassum, U., Reddy, O., Mukherjee, G., 2012. Elevated serum haptoglobin is associated with clinical outcome in triple-negative breast cancer patients. *Asian Pacific Journal of Cancer Prevention : APJCP.* **13,** 4541-4544.

Tabb, D.L., Vega-Montoto, L., Rudnick, P.A., Variyath, A.M., Ham, A.J., Bunk, D.M., Kilpatrick, L.E., Billheimer, D.D., Blackman, R.K., Cardasis, H.L., Carr, S.A., Clauser, K.R., Jaffe, J.D., Kowalski, K.A., Neubert, T.A., Regnier, F.E., Schilling, B., Tegeler, T.J., Wang, M., Wang, P., Whiteaker, J.R., Zimmerman, L.J., Fisher, S.J., Gibson, B.W., Kinsinger, C.R., Mesri, M., Rodriguez, H., Stein, S.E., Tempst, P., Paulovich, A.G., Liebler, D.C., Spiegelman, C., 2010. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *Journal of Proteome Research.* **9,** 761-776.

Theodorescu, D., 2003. Molecular pathogenesis of urothelial bladder cancer. *Histology and Histopathology.* **18,** 259-274.

Thiede, B., Hohenwarter, W., Krah, A., Mattow, J., Schmid, M., Schmidt, F., Jungblut, P.R., 2005. Peptide mass fingerprinting. *Methods (San Diego, Calif.).* **35,** 237-247.

Thompson, I.M., Peek, M., Rodriguez, F.R., 1987. The impact of cigarette smoking on stage, grade and number of recurrences of transitional cell carcinoma of the bladder. *The Journal of Urology.* **137,** 401-403.

Thomson, J.J., 1899. On the Masses of the Ions in Gases at Low Pressures. **48,** 547-567.

Tian, M., Cui, Y.Z., Song, G.H., Zong, M.J., Zhou, X.Y., Chen, Y., Han, J.X., 2008. Proteomic analysis identifies MMP-9, DJ-1 and A1BG as overexpressed proteins in pancreatic juice from pancreatic ductal adenocarcinoma patients. *BMC Cancer.* **8,** 241-2407-8-241.

Towbin, H., Staehelin, T., Gordon, J., 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America.* **76,** 4350-4354.

Treff, N.R., Dement, G.A., Adair, J.E., Britt, R.L., Nie, R., Shima, J.E., Taylor, W.E., Reeves, R., 2004. Human KIT ligand promoter is positively regulated by HMGA1 in breast and ovarian cancer cells. *Oncogene.* **23,** 8557-8562.

Trichopoulos, D., Psaltopoulou, T., Orfanos, P., Trichopoulou, A., Boffetta, P., 2006. Plasma C-reactive protein and risk of cancer: a prospective study from Greece. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* **15,** 381-384.

Tsai, H.Y., Boonyapranai, K., Sriyam, S., Yu, C.J., Wu, S.W., Khoo, K.H., Phutrakul, S., Chen, S.T., 2011. Glycoproteomics analysis to identify a glycoform on haptoglobin associated with lung cancer. *Proteomics.* **11,** 2162-2170.

Tu, C., Rudnick, P.A., Martinez, M.Y., Cheek, K.L., Stein, S.E., Slebos, R.J., Liebler, D.C., 2010. Depletion of abundant plasma proteins and limitations of plasma proteomics. *Journal of Proteome Research.* **9,** 4982-4991.

Tung, C.L., Lin, S.T., Chou, H.C., Chen, Y.W., Lin, H.C., Tung, C.L., Huang, K.J., Chen, Y.J., Lee, Y.R., Chan, H.L., 2013. Proteomics-based identification of plasma biomarkers in oral squamous cell carcinoma. *Journal of Pharmaceutical and Biomedical Analysis.* **75,** 7-17.

Turcotte, R.F. & Raines, R.T., 2008. Interaction of onconase with the human ribonuclease inhibitor protein. *Biochemical and Biophysical Research Communications.* **377,** 512-514.

Tyan, Y.C., Yang, M.H., Chen, S.C., Jong, S.B., Chen, W.C., Yang, Y.H., Chung, T.W., Liao, P.C., 2011. Urinary protein profiling by liquid chromatography/tandem mass spectrometry: ADAM28 is overexpressed in bladder transitional cell carcinoma. *Rapid Communications in Mass Spectrometry : RCM.* **25,** 2851-2862.

Unsal, E., Atalay, F., Atikcan, S., Yilmaz, A., 2004. Prognostic significance of hemostatic parameters in patients with lung cancer. *Respiratory Medicine.* **98,** 93-98.

Valentine, S.J., Ewing, M.A., Dilger, J.M., Glover, M.S., Geromanos, S., Hughes, C., Clemmer, D.E., 2011. Using ion mobility data to improve peptide identification: intrinsic amino acid size parameters. *Journal of Proteome Research.* **10,** 2318-2329.

van Rhijn, B.W., Lurkin, I., Radvanyi, F., Kirkels, W.J., van der Kwast, T.H., Zwarthoff, E.C., 2001. The fibroblast growth factor receptor 3 (FGFR3) mutation is a strong indicator of superficial bladder cancer with low recurrence rate. *Cancer Research.* **61,** 1265-1268.

van Rhijn, B.W., van der Kwast, T.H., Vis, A.N., Kirkels, W.J., Boeve, E.R., Jobsis, A.C., Zwarthoff, E.C., 2004. FGFR3 and P53 characterize alternative genetic pathways in the pathogenesis of urothelial cell carcinoma. *Cancer Research.* **64,** 1911-1914.

Vanhoutte, K.J., Laarakkers, C., Marchiori, E., Pickkers, P., Wetzels, J.F., Willems, J.L., van den Heuvel, L.P., Russel, F.G., Masereeuw, R., 2007. Biomarker discovery with SELDI-TOF MS in human urine associated with early renal injury: evaluation with computational analytical tools. *Nephrology, Dialysis, Transplantation : Official Publication of the European Dialysis and Transplant Association - European Renal Association.* **22,** 2932-2943.

Villanueva, C.M., Cantor, K.P., King, W.D., Jaakkola, J.J., Cordier, S., Lynch, C.F., Porru, S., Kogevinas, M., 2006. Total and specific fluid consumption as determinants of bladder cancer risk. *International Journal of Cancer.Journal International Du Cancer.* **118,** 2040-2047.

Voegele, A.F., Jerkovic, L., Wellenzohn, B., Eller, P., Kronenberg, F., Liedl, K.R., Dieplinger, H., 2002. Characterization of the vitamin E-binding properties of human plasma afamin. *Biochemistry.* **41,** 14532-14538.

Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L., Humphery-Smith, I., 1995. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis.* **16,** 1090-1094.

Waters, W.B., 1996. Invasive bladder cancer--where do we go from here? *The Journal of Urology.* **155,** 1910-1911.

Weinberg, R.A., 2006. *The Biology of Cancer.*

Weinstein, S.J., Stolzenberg-Solomon, R.Z., Kopp, W., Rager, H., Virtamo, J., Albanes, D., 2012. Impact of circulating vitamin D binding protein levels on the association between 25-hydroxyvitamin D and pancreatic cancer risk: a nested case-control study. *Cancer Research.* **72,** 1190-1198.

Weitzman, S.A. & Stossel, T.P., 1981. Mutation caused by human phagocytes. *Science (New York, N.Y.).* **212,** 546-547.

Willemse, J.L. & Hendriks, D.F., 2006. Measurement of procarboxypeptidase U (TAFI) in human plasma: a laboratory challenge. *Clinical Chemistry.* **52,** 30-36.

Wilm, M., 2011. Principles of electrospray ionization. *Molecular & Cellular Proteomics : MCP.* **10,** M111.009407.

Woo, Y., Hyung, W.J., Obama, K., Kim, H.I., Pak, K.H., Son, T., Noh, S.H., 2012. Elevated high-sensitivity C-reactive protein, a marker of advanced stage gastric cancer and postgastrectomy disease recurrence. *Journal of Surgical Oncology.* **105,** 405-409.

World Health Organisation, 2013. *World Health Statistics.* Switzerland: World Health Organisation.

Wu, G.X., Lin, Y.M., Zhou, T.H., Gao, H., Pei, G., 2000. Significant down-regulation of alpha-albumin in human hepatoma and its implication. *Cancer Letters.* **160,** 229-236.

Yadav, A.K., Bhardwaj, G., Basak, T., Kumar, D., Ahmad, S., Priyadarshini, R., Singh, A.K., Dash, D., Sengupta, S., 2011. A systematic analysis of eluted fraction of plasma post immunoaffinity depletion: implications in biomarker discovery. *PloS One.* **6,** e24442.

Yamashita, M. & Fenn, J.B., 1984. Electrospray ion source. Another variation on the free-jet theme. **88,** 4451-4459.

Yang, L.V., Nicholson, R.H., Kaplan, J., Galy, A., Li, L., 2001. Hemogen is a novel nuclear factor specifically expressed in mouse hematopoietic development and its human homologue EDAG maps to chromosome 9q22, a region containing breakpoints of hematological neoplasms. *Mechanisms of Development.* **104,** 105-111.

Yap, J., Foo, C.F., Lee, M.Y., Stanton, P.G., Dimitriadis, E., 2011. Proteomic analysis identifies interleukin 11 regulated plasma membrane proteins in human endometrial epithelial cells in vitro. *Reproductive Biology and Endocrinology : RB&E.* **9,** 73-7827-9-73.

Yoshida, S., Saito, K., Koga, F., Yokoyama, M., Kageyama, Y., Masuda, H., Kobayashi, T., Kawakami, S., Kihara, K., 2008. C-reactive protein level predicts prognosis in patients with muscle-invasive bladder cancer treated with chemoradiotherapy. *BJU International.* **101,** 978-981.

Zaravinos, A., Volanis, D., Lambrou, G.I., Delakas, D., Spandidos, D.A., 2012. Role of the angiogenic components, VEGFA, FGF2, OPN and RHOC, in urothelial cell carcinoma of the urinary bladder. *Oncology Reports.* **28,** 1159-1166.

Zeng, Z., Hincapie, M., Pitteri, S.J., Hanash, S., Schalkwijk, J., Hogan, J.M., Wang, H., Hancock, W.S., 2011. A proteomics platform combining depletion, multi-lectin affinity chromatography (M-LAC), and isoelectric focusing to study the breast cancer proteome. *Analytical Chemistry.* **83,** 4845-4854.

Zhang, Q., Faca, V., Hanash, S., 2011. Mining the plasma proteome for disease applications across seven logs of protein abundance. *Journal of Proteome Research.* **10,** 46-50.

Zhang, S.M., Buring, J.E., Lee, I.M., Cook, N.R., Ridker, P.M., 2005. C-reactive protein levels are not associated with increased risk for colorectal cancer in women. *Annals of Internal Medicine.* **142,** 425-432.

Zhang, S.M., Lin, J., Cook, N.R., Lee, I.M., Manson, J.E., Buring, J.E., Ridker, P.M., 2007. C-reactive protein and risk of breast cancer. *Journal of the National Cancer Institute.* **99,** 890-894.

Zhuo, L. & Kimata, K., 2008. Structure and function of inter-alpha-trypsin inhibitor heavy chains. *Connective Tissue Research.* **49,** 311-320.

Zieger, K., Dyrskjot, L., Wiuf, C., Jensen, J.L., Andersen, C.L., Jensen, K.M., Orntoft, T.F., 2005. Role of activating fibroblast growth factor receptor 3 mutations in the development of bladder tumors. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research.* **11,** 7709-7719.