The Use of Proteomics Techniques to Identify Potential Markers of Early Stage Colorectal Cancer

By

Donna May O'Neil

A thesis submitted to the University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Cancer Sciences

College of Medical and Dental Sciences

The University of Birmingham

Febuary 2011

Abstract

Colorectal cancer (CRC) is the third most common cancer and the second most common cause of cancer related deaths in the UK. Research has shown that the five year survival rate for patients if diagnosed at an early stage is 83% however, only 11% of cases are diagnosed at this stage. The aim of this study was to use proteomic approaches to investigate secreted proteins from colorectal cancer cell lines to identify candidate biomarkers for early stage diagnosis. Microvesicles (MVs) are a mixed population of vesicles that are released by a wide range of cells and are thought to play a role in tumour development and progression. Stable Isotope Labeling of Amino Acids in Culture (SILAC) was used to investigate the relative abundance of proteins secreted in MVs released by two cell lines that are used as a model of early tumour progression. This study identified 86 potential candidates that demonstrated increased release and six of these proteins (AGR2, OLFM4, SBP1, HSP90α, HSP90β and CEACAM5) were selected for further investigation by Western blot analysis. These proteins show potential as markers of early stage CRC and would be suitable for further validation in patient serum samples.

Dedicated to:

My grandparents,

Leonard Albert Holmes and Dorothy May Holmes

I hope you'd be proud

My husband, John Daniel O'Neil

Thanks for your constant support and love

Acknowledgements

Firstly, I would like to acknowledge the College of Medical and Dental Sciences for giving me the opportunity to carry out this program of study part-time whilst employed at the School of Cancer Sciences and the School of Clinical and Experimental Medicine. I would also like to thank Dr Ashley Martin and Prof Philip Johnson for their supervision. Big thanks to Team Proteomics: Elisabeth Phillips, Neil Shimwell and Douglas Ward for all their help, support and motivational cups of tea throughout my studies. I would also like to thank Angela Taylor, Pushpa Patel, Suzy Eldershaw and Myriam Chimen for their support and encouragement whilst I was writing my thesis.

I would like to acknowledge my Mum and Dad, Denise and Charles Broadbent and all my family (Holmes, Pepper, Jones and Preedy) for their love, encouragement and support in everything I do. X

My biggest thanks go to my husband John O'Neil. I am forever grateful for your love, support and understanding, you have been a great help professionally and emotionally throughout my studies. Without you, I would've been lost many a time. Thank you for always believing in me. Love you always. X

Contents

List of Figures

Chapter 1 : Introduction

Chapter 2 : Materials and Methods

Chapter 3 : Investigating the secretome from RG/C2 and PRG/C2 cells to identify possible biomarkers for the early detection of colorectal cancer using proteomic techniques

List of Tables

Chapter 1 : Introduction

Chapter 2 : Materials and Methods

Chapter 3 : Investigating the secretome from RG/C2 and PRG/C2 cells to identify possible biomarkers for the early detection of colorectal cancer using proteomic techniques

Chapter 4 : Optimisation of Conditions Required for Quantitative Proteomic Profiling of MVs Released by RG/C2 and PRG/C2 Cell Line

Chapter 5 : Quantitative analysis of proteins secreted in MV's from CM derived from colorectal cancer cell lines using Stable Isotope Labelling of Amino acids in Culture (SILAC)

Chapter 6 : Validation of selected proteins identified in the SILAC profile of MV released by colorectal cancer cell lines during early tumour progression

List of Supplementary Material (see enclosed CD)

Chapter 3 : Investigating the secretome from RG/C2 and PRG/C2 cells to identify possible biomarkers for the early detection of colorectal cancer using proteomic techniques

Supplementary Table 3a: Proteins identified in MV's from CM derived from RG/C2 cells ($6x10^6$ cells) as outlined Figure 3.6.

Supplementary Table 3b: Proteins identified in MV's from CM derived from RG/C2 cells ($1.2x10^7$ cells) as outlined in Figure 3.7.

Supplementary Table 3: Proteins identified in MV's from CM derived from RG/C2 cells ($2.4x10^7$ cells) as outlined in Figure 3.8.

Chapter 5 : Quantitative analysis of proteins secreted in MV's from CM derived from colorectal cancer cell lines using Stable Isotope Labeling of Amino acids in Culture SILAC)

Supplementary Table 5a: Initial analysis of the data generated from the first SILAC experiment using the multiple thresholds filter parameters optimized in table 4.2.

Supplementary Table 5b: Comparison of the TPP results for all 4 SILAC experiments.

Supplementary Table 5c: Proteins identified in these SILAC surveys that were previously reported in proteomic profiles of CM and MVs derived from various cancer cell lines.

Supplementary Table 5d: Raw data from the TPP analysis for experiment C13 ARG A.

Supplementary Table 5e: Raw data from the TPP analysis for experiment C13 ARG B.

Supplementary Table 5f: Raw data from the TPP analysis for experiment C13 LYS A.

Supplementary Table 5g: Raw data from the TPP analysis for experiment C13 LYS B.

Chapter 6 : Validation of selected proteins identified in the SILAC profile of MV released by colorectal cancer cell lines during early tumour progression

Supplementary Figure 6e: Amino acid sequence of AGR2 and its identified peptides.

Supplementary Figure 6f: Amino acid sequence of OLFM4 and its identified peptides.

Supplementary Figure 6g: Amino acid sequence of SBP1 and its identified peptides.

Supplementary Figure 6h: Amino acid sequence of HSP90α and its identified peptides.

Supplementary Figure 6i: Amino acid sequence of HSP90β and its identified peptides.

Abbreviations

2D DIGE - 2D Differential In-Gel Electrophoresis

2D PAGE – Two-Dimensional Polyacrylamide Gel Electrophoresis

AGR2 – Anterior Gradient Homolog 2

AQUA – Absolute Quantification

ASAPR – Automated Statistical Analysis of Protein Abundance Ratios

BSA – Bovine Serum Albumin

CBB G250 - Coomassie Brilliant Blue G250

CEACAM5 (CEA) – Carcinoembryonic Antigen –Related Cell Adhesion Molecule 5

CID - Collision Induced Dissociation

CM – Conditioned media

CRC – Colorectal Cancer

DMEM – Dulbecco's Modified Eagle's Medium

DMSO – Dimethyl Sulphoxide

DTT – Dithiothreitol

EMMPRIN – Extracelleular Matrix Metalloproteinase Inducer

ESI – Electrospray Ionisation

FAP – Familial Adenomatous Polyposis

FCS – Foetal Calf Serum

FOBT - Faecal Occult Blood Test

GO – Gene Ontology

HNPCC – Hereditary Non-Polyposic Colon Cancer

HPLC – High Performance Liquid Chromatography

HSP90 – Heat Shock Protein 90

ICPL – Isotope-Coded Protein Labels

IEF – Isoelectric Focusing

IPG – Immobilised pH Gradients

iTRAQ – Isobaric Tags for Relative and Absolute Quantification

LC – Liquid Chromatography

m/z – mass to charge ratio

MALDI – Matrix Assisted Laser Desorption Ionisation

MRM – Multiple Reaction Monitoring

MV – Microvesicles

OLFM4 – Olfactomedin 4

PBS – Phosphate Buffered Saline

REM CM – Remaining Conditioned media

RP – Reverse Phase

SBP1 – Selenium Binding Protein 1

SCX – Strong Cation Exchange

SELDI – Surface Enhanced Laser Desorption Ionisation

SF – Serum Free

SILAC – Stable Isotope Labelling of Amino Acids in Culture

TACSTD1 – Tumour Associated Calcium Signal Transducer 1

TOF – Time-of-Flight

Chapter 1

Introduction

Chapter 1 : Introduction

Section 1.1 : Colorectal Cancer and Currently used Biomarkers

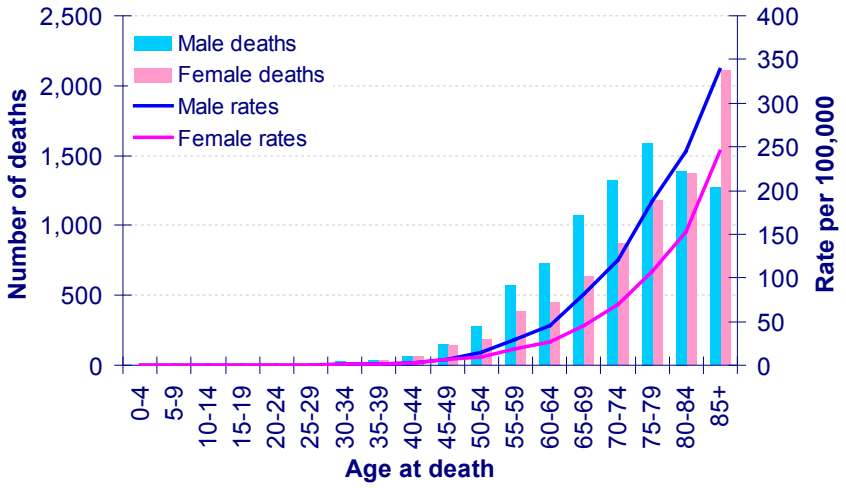
1.1.1. Introduction

Colorectal cancer (CRC) is the third most common cancer in the UK following breast and lung cancer with 38,608 newly diagnosed cases in 2007 (21,014 in men and 17,594 in women) (Cancer Research UK). In 2008, there were 16,254 deaths related to colorectal cancer (8,758 in men and 7,501 in women) making it the second most common cause of cancer related deaths in the UK (see Figure 1.1). Mortality rates increase with age with 69% of the total deaths due to colorectal cancer occurring in patients >70 years of age.

Figure 1.2 outlines the commonly used Dukes' staging which is used for colorectal cancer tumour staging. Dukes' stage A is early stage cancer with the tumour present in the innermost lining of the colon. Dukes' stage B is when the tumour has started to grow into the muscle layer. When the tumour has spread to the lymph nodes, this is known as Dukes' stage C and advanced stage cancer (Dukes' stage D) is when tumour cells have metastasized to distant organs such as the liver. The survival rate is dependent upon the stage of CRC at diagnosis with the 5 year survival rate being 83, 64, 38 and 3% for Dukes stage A, B, C and D respectively. If diagnosed early (Dukes stage A) the prognosis looks promising however, very few cases are diagnosed at an early stage with only 11% of cases diagnosed in 2006 at this stage (Figure 1.2e). Studies have shown that detection of CRC at an earlier stage via mass screening and intervention can reduce the risk of death from CRC, highlighting the need for biomarkers for early stage diagnosis (Hardcastle et al., 1996, Kronborg et al., 1996).

The symptoms of CRC include rectal bleeding and blood in stools, persistent diarrhea, the presence of a lump in the right side of the abdomen or rectum, weight loss, anaemia

(a)

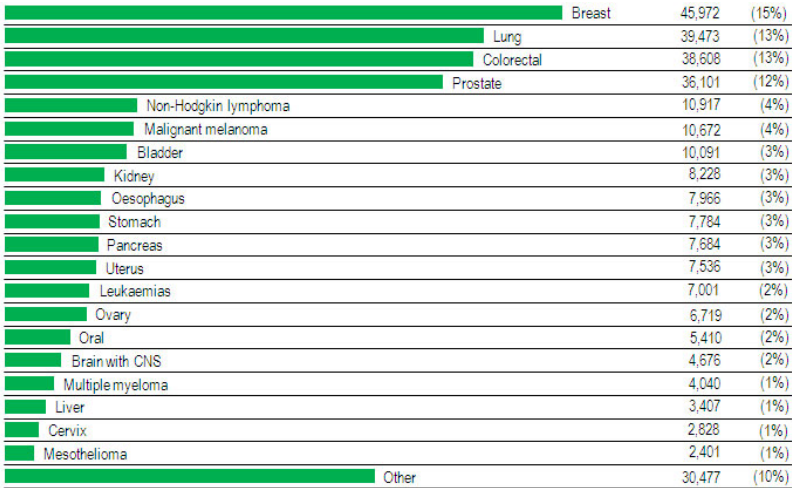


(b)

Lung	35,261	(22%)
Colorectal	16,259	(10%)
Breast	12,116	(8%)
Prostate	10,168	(6%)
Pancreas	7,781	(5%)
Oesophagus	7,606	(5%)
Stomach	5,178	(3%)
Bladder	5,002	(3%)
Non-Hodgkin lymphoma	4,438	(3%)
Ovary	4,373	(3%)
All leukaemias	4,367	(3%)
Kidney	3,848	(2%)
Brain with CNS	3,674	(2%)
Liver	3,390	(2%)
Multiple myeloma	2,660	(2%)
Mesothelioma	2,156	(1%)
Malignant melanoma	2,067	(1%)
Oral	1,822	(1%)
Uterus	1,741	(1%)
Bone & connective tissue	1,037	(1%)
Other	21,779	(14%)

UK mortality 2008: The 20 most common causes of death from cancer

(c)

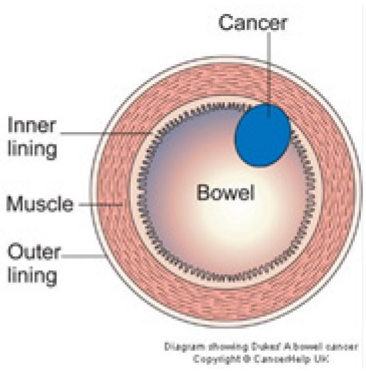


UK incidence 2007: the 20 most commonly diagnosed cancers in persons

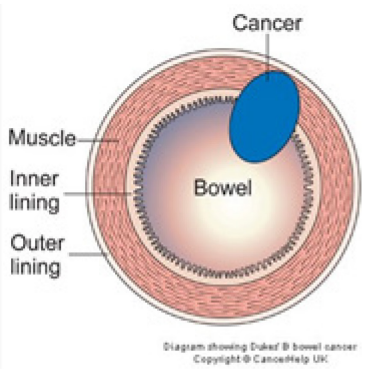
Figure 1. 1 : Cancer Research UK Colorectal Cancer Statistics.

(a) Number of deaths and age-specific mortality rates, colorectal cancer, by sex, UK, 2006; (b) The 20 most common causes of death from cancer in 2008; (c) The incidence of the 20 most common cancers in the UK in 2007. Statistics taken from the Cancer Research UK website http://info.cancerresearchuk.org/cancerstats.

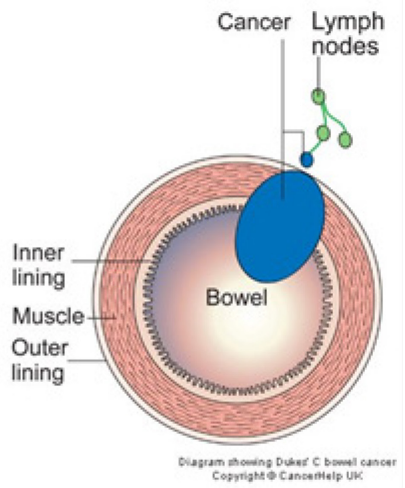
(a) Dukes' Stage A



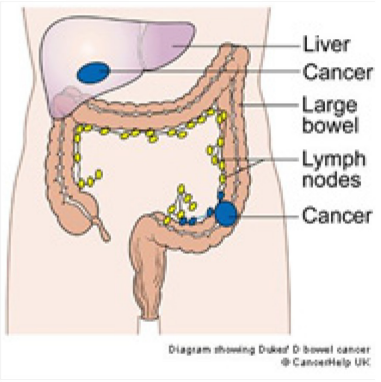
(b) Dukes' Stage B



(c) Dukes' Stage C



(d) Dukes' Stage D



(e)

Dukes' Stage	UIJCC / AJCC	Description #	Approximate frequency at diagnosis *	Appromimate 5 year survival *
A	I	Tumours invading submucosa or muscularis propria	11%	83%
B	II	Tumours invading through muscularis propria	35%	64%
C	III	Invasion of tumour to lymph nodes	26%	38%
D	IV	Metastasis to distant sites	29%	3%

Figure 1. 2 : Dukes' Staging of CRC.

Figure to show tumour staging according to the Dukes' system. The stages can also be referred to as stage I-IV according to the UICC / AJCC system (UICC – International Union Against Cancer, AJCC – American Joint Committee on Cancer). (a) Dukes' Stage A or Stage I – tumour is located in the innermost lining of the colon or rectum only or may have just started to grow in the muscle layer. (b) Dukes' Stage B or Stage II – the tumour has started to grow in to the muscle layer of the colon. (c) Dukes' Stage C or Stage III – tumour cells have spread to at least one lymph node in close proximity to the colon. (d) Dukes' Stage D or Stage IV – tumour cells have metastasised to distant organs such as the liver. (e) Table to summarise the stages of CRC, the approximate frequency and 5 year relative survival rate if diagnosed at each stage. * - taken from http://info.cancerresearchuk.org/cancerstats Diagrams a-d taken from http://www.cancerhelp.org.uk/type/bowel-cancer/treatment/dukes-stages-ofbowelcancer.

causing tiredness and breathlessness and pains in the abdomen caused by a bowel obstruction leading to constipation and vomiting (Cancer Research UK). The risk factors associated with CRC include age, obesity, lack of exercise, diet (low in fibre, fruit, vegetables, folate and high in red or processed meat), smoking, high alcohol consumption and genetic predisposition (Cancer Research UK).

CRC results from the cumulative effects of sequential genetic alterations in oncogenes, tumour suppressor genes and DNA repair genes. In sporadic CRC, these alterations are acquired after exposure to exogenous and endogenous carcinogens. In certain conditions such as Familial Adenomatous Polyposis (FAP) and Hereditary Non-Polyposis Colon Cancer (HNPCC) genetic alterations can lead to a predisposition to malignancy.

1.1.2. Currently used Biomarkers for Colorectal Cancer

Biomarker is a term often referring to an indicator of a biological state i.e. biological process, pathogenic process or pharmacological response to therapeutic intervention. A biomarker for early detection for CRC would require high sensitivity and specificity therefore producing a low number of false positives to prevent subjecting patients to unnecessary colonoscopies. Other useful biomarkers that could reduce the mortality rate of CRC include prognostic markers to predict the outcome of patient's survival and markers to predict response to drugs or treatment. As well as other research approaches, genomic and proteomic advances have contributed to our understanding of pathways that control growth, differentiation and death of cells. DNA microarray analysis and proteomic methods such as 2DE and mass spectrometry-based methods are commonly used to evaluate the expression profiles of genes and proteins in cells, tissues and body fluids (Faca et al., 2007, S. Y. Kim and Hahn, 2007). Biomarker discovery will be further discussed in Section 1.4.

1.1.2.1. Faecal Markers

Faecal Occult Blood Test (FOBT) is the least expensive and invasive screening method for CRC and detects faecal haemoglobin enzymatically or immunologically. Enzymatic FOBT (the guaiac test) measures the peroxidase-like activity of haemoglobin that originates from any source (i.e. both colorectal and upper gastrointestinal bleeding). Immunological FOBT uses antibodies to detect the globin moiety in haemoglobin. FOBT screening has been shown to reduce the incidence of advanced stage CRC and related mortality by 20% and 16% respectively (C. S. Huang et al., 2005a) and has been recommended as a screening test for early stage CRC in patients >50 yrs old alongside flexible sigmoidoscopy, colonoscopy and double contrast barium enema (Duffy et al., 2003).

The advantages of FOBT screening include: ability to examine the entire colorectal tract, non-invasive, inexpensive, requires no patient preparation and sample collection can be carried out in the privacy of the patient's home. However, the FOBT test does have its limitations demonstrating low sensitivity and specificity for both adenomas and carcinomas. Also, the ingestion of certain foods (red meats, fruits and vegetables) and certain medication (non-steroidal anti-inflammatory drugs) can yield false-positive results with the guaiac-based assay (limits don't apply to the immunochemical FOBT assay). The test also relies on multiple stool samples analysed annually to increase chances of detecting intermittent bleeding.

Various genes and epigenetic markers can also be used as markers of CRC in patients stool samples. Colonocytes (colon cells) that are shed into the faecal stream can provide material for the detection of genes and epigenetic markers in faeces. They are shed continuously but occur more frequently in CRC compared to normal colon. Faecal colonocytes are assessed by analysing DNA mutations in target genes such as *K-ras, p53* and

APC by analysing epigenetic markers such as microsatelite instability (MSI) or by measuring un-fragmented long form DNA (L-DNA).

After a review of the literature, Haug and Brenner showed that DNA marker panels detected CRC with a specificity of >95% and a sensitivity 60-90% (Haug and Brenner, 2005). In comparison with FOBT, DNA marker panels identified more cases of invasive CRCs and neither test detected the majority of neoplastic lesions but the DNA panel showed a higher sensitivity than FOBT without reducing specificity (Albaugh et al., 1992). However, faecal DNA markers exhibit a low sensitivity in asymptomatic patients and the assay is relatively expensive and complex, so it is unlikely to replace FOBT for general screening.

1.1.2.2. Serum or Blood Markers

1.1.2.2.1. CEA

Carcinoembryonic antigen (CEA) was first described by Gold and Freeman as an antigen detected in both foetal colon and colon adenocarcinoma and appeared to be absent in healthy adult colon (Gold and Freedman, 1965a). It was later shown to be present in certain healthy tissues but was 60-fold higher in tumours compared with non-malignant tissue (Gold and Freedman, 1965b). Thomson *et al* then described increased concentrations of CEA in the serum of 35 out of 36 patients with CRC (Thomson et al., 1969). High levels of CEA were not found in 'normal' patients, pregnant women, patients with non-gastrointestinal cancers or patients with miscellaneous benign gastrointestinal diseases. These results were not confirmed in other studies but nevertheless prompted the widespread use of CEA as a marker of CRC.

CEA can be detected in biopsies but is usually measured in serum. A high CEA level is associated with cancer progression and levels are expected to fall following surgery (Duffy, 2001). However in the absence of cancer, high levels of CEA can be observed in other

conditions such as hepatitis, inflammatory bowel disease, pancreatitis and obstructive pulmonary disease. Other factors that affect CEA levels include:

Tumour Stage. One study investigated the reoccurrence rates in 348 CRC patients and demonstrated that the proportion of these patients with an elevated CEA (>2.5µg/L) were 28%, 45%, 75% and 84% for Dukes' stage A, B, C and D respectively. These values are greatly reduced when the cut-off is increased to >5µg/L with 3, 25, 45 and 65% for Dukes' stage A, B, C and D respectively (Wanebo et al., 1978).

Tumour Grade. Serum concentrations of CEA tend to be higher in patients with well differentiated tumour compared to those with poorly differentiated tumours (Goslin et al., 1981).

Liver Status. Liver is the primary site for the degradation/removal of CEA and benign liver diseases impair liver function and thus the clearance of CEA. Therefore CEA can be increased in patients' serum with non-malignant liver disease (Begent, 1984, Thomas et al., 1990).

Tumour Site within Colon. Patients with tumours on the left side of the colon tend to have higher CEA levels than those on the right (Slater et al., 1979, Wanebo et al., 1978).

Presence or Absence of Bowel Obstruction. Bowel obstruction can lead to increased CEA levels in patients with colorectal malignancy (Sugarbaker and Wilson, 1976).

Smoking. Smoking appears to double the serum levels of CEA (Rognum, 1986).

CEA doesn't provide sufficient sensitivity and reliability for the detection of early stage cancer. Using a cut-off of 2.5ng/ml, it has been reported that CEA can detect early stage CRC (Dukes' A, B) with a low sensitivity of 30-40% and a specificity of 87% (Duffy et al., 2003). Therefore, the NIH does not recommend CEA is used in screening programmes for the

detection early stage of CRC. Instead, the potential value of the CEA test is as a prognostic marker once CRC has been diagnosed (with higher levels of CEA being indicative of greater disease severity and poor prognosis) (Cancer Research UK).

CEA is mainly used to monitor patients following curative resection for primary cancer (Duffy et al., 2003). Studies have shown that intensive follow up including regular serum CEA determinations results in a modest but statistically significant improvement in outcome compared with a minimal follow up strategy. Using CEA during follow-up is relatively inexpensive compared to other follow up strategies such as radiology and endoscopy and also causes minimal inconvenience for the patients.

CEA can also be used for the early detection of metastatic disease in candidate patients for liver resection (Duffy et al., 2003). Liver metastasis usually occurs within 5 years of initial diagnosis but curative resection is only suitable in ~25% of cases. In these patients, serum CEA should be determined every 2-3 months for at least the first 3 years after diagnosis. Two studies have demonstrated that serial CEA determination is useful in detecting liver metastasis and an increased CEA could detect liver metastases with a sensitivity of 94% and 100% and a specificity of 96% (Arnaud et al., 1980, Pietra et al., 1998).

CEA can also be used to monitor patient's response to chemotherapy. Patients who exhibit a decrease in serum CEA whilst on chemotherapy have been shown to exhibit an improved survival rate compared to those whose CEA levels failed to decrease ((no-authors-listed), 1996, Begent, 1984, Sugarbaker and Wilson, 1976). An increase in CEA in patients receiving chemotherapy predicts progressive disease.

1.1.2.2.2. CA19-9

Glycoprotein CA19-9 is the second most investigated gastrointestinal tumour marker and is the best marker available for diagnosing pancreatic adenocarcinoma however, it

demonstrates a lower sensitivity than CEA for CRC and is also less useful for monitoring patients once diagnosis has been made (Duffy, 1998). Other glycosylated proteins have been studied including CA 50, CA195, CA 242, CA M26, CA M25, CA M43 and CA 72-4 but have been shown to not be useful detection markers for CRC (Hundt et al., 2007).

1.1.2.3. Tissue-Based Markers for CRC

Thymidylate Synthase (TS) is involved in the conversion of deoxyuridinemonophosphate (dUMP) to deoxythymidine monophosphate (dTMP). It provides a source of thymidylate for DNA synthesis and has been investigated as a prognostic and therapy predictive marker of CRC. Protein levels of the TS enzyme is the target of many currently used chemotherapeutic agents such as 5-fluorauracil (5-FU). Only 20% of patients respond to 5-FU and a predictive marker that could indicate whether a patient is likely to respond preventing patients suffering from unnecessary side effects. Several studies have investigated the levels of TS in patients being treated with 5-FU, although the results of theses appear to be conflicting (Edler et al., 2002, Prall et al., 2007, Soong et al., 2008). Therefore, TS is not currently recommended for routine prognosis determination or predicting response to therapy.

1.1.2.4. Other Potential Protein Biomarkers for CRC

There are many proteins identified using a variety of mass spectrometry and antibody-based proteomic studies that have been reported as potential markers for colorectal cancer. In many of these studies, the candidate biomarkers are compared to CEA and although it is not an ideal marker it is the most commonly used biomarker for CRC (see Section 1.1.2.2.1). Table 1.1 shows a few examples other protein biomarkers for CRC.

Protein	Investigation	Ref
TIMP-1	Blood marker shown to be higher in patients with colon and rectal cancers.	1, 2
Five serum marker panel	Five proteins (Spondin-2, DcR3, Trail-R2, RegIV, MIC 1) shown to be elevated in CRC patients compared to normal controls and benign tumours.	3
NNMT and PSME2	Shown to be elevated in CRC tissue samples compared to matched normal tissue. Validated in serum of CRC patients.	4, 5
CRMP2	Elevated in CRC patient sera compared to normal controls. CRMP-2 alone showed increased sensitivity (61%) but decreased specificity (65%) compared to CEA (sensitivity 4% and 87% specificity if used alone). When used in conjunction with CEA, CRMP-2 showed	6
Six SELDI peak panel	six peaks corresponding to four proteins, apolipoprotein C1, complement C3a-desArg, a1-antitrypsin and transferrin. Used to classify CRC patients with 95% sensitivity and 91% specificity compared to CEA alone (54% sensitivity and 93% specificity).	7
a-Defensins 1,2 and 3 (HNP-1, HNP-2 and HNP-3	Two separate SELDI-TOF studies identified these proteins to be elevated in CRC pateint sera compared to normal controls. HNP1-3 showed a sensitivity of 69% and a specificity of 100% in one of these small studies.	8, 9
MIF	Significantly increased in sera from CRC patients. Demonstrated increased sensitivy for early stage CRC compared to CEA (47.3% and 29.5% respectively).	10
Prolactin	Could be used to distinguish CRC from normal healthy controls with a sensitivity of 77% and a specificity of 89%.	11

Table 1. 1: Other Potential Protein Biomarkers for CRC.

Table to show a few examples of other potential protein biomarkers for CRC reported in the literature. References: 1 - (Holten-Andersen et al., 1999), 2 - (Sorensen et al., 2008), 3 - (diaDexus-Inc., 2007), 4 - (Roessler et al., 2006), 5 - (Roessler et al., 2005), 6 - (C. C. Wu et al., 2008), 7 - (Ward et al., 2006c), 8 - (Habermann et al., 2006), 9 - (Melle et al., 2005), 10 - (H. Lee et al., 2008), 11 - (Soroush et al., 2004).

Section 1.2 : Mass Spectrometry-based Proteomics

1.2.1. Introduction

Proteomics is the study of peptides and proteins within a biological system i.e. the proteome. Traditionally, Edman degradation has been used to sequence proteins by chemically cleaving amino acids one at a time from the amino terminus (Edman, 1949). The technique is not always completely successful and requires a high level of expertise (Steen and Mann, 2004). More recently, advances in mass spectrometry (MS) has allowed the development of methods which are more sensitive and are not hindered by modifications or by a blocked N-terminus, as is the case with Edman degradation. MS-based proteomics is now the most commonly used technology for identifying proteins in a single spot or gel band slice from a simple experiment and is increasingly being used to identify as many proteins in a single experiment as possible. This is known as shotgun proteomics.

In order to identify proteins, they can be enzymatically cleaved using proteases and the resulting peptides are analysed by MS. Figure 1.3 outlines the general workflow of a typical shotgun proteomics experiment. Proteins are commonly digested using trypsin which cleaves proteins at the C-terminal side of arginine and lysine residues. Other proteases used include Lys-C and Glu-C which cleave at the C-terminus of lysine and glutamate residues respectively and Asp-N which cleaves at the N-terminus of aspartate residues. Digesting proteins to produce peptides overcomes problems that protein analysis presents such as solubility and instability under the conditions required for MS. However, sequence coverage is sacrificed and is not sufficient for complete protein characterisation but can be sufficient for unambiguous protein identification.

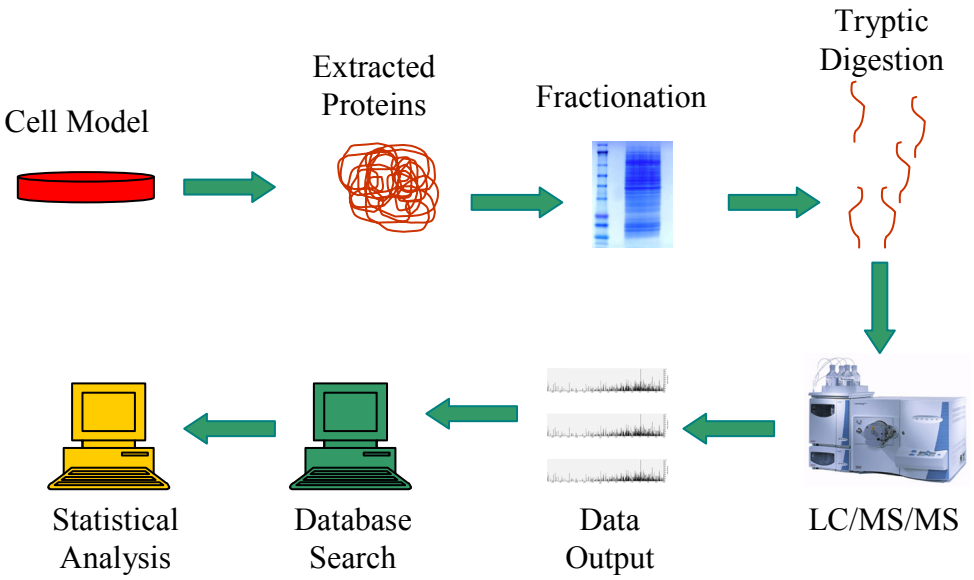


Figure 1. 3 : Tandem MS/MS based shotgun proteomics.

Diagram to show the typical workflow of a tandem MS/MS based proteomics study. Proteins are extracted from cell lines, tissues or body fluids and are then fractionated using various methods e.g. 1D or 2D gel electrophoresis. The proteins are digested in-gel or in solution with a proteolytic enzyme (typically trypsin). These peptides are then analysed by LC MS/MS to generate hundreds or thousands of spectra. This is then followed by the data analysis involving database searching and statistical approaches.

Typically, the peptides generated from a protein digest are not directly injected into a mass spectrometer but are first loaded onto a high-performance liquid chromatography (HPLC) column. Often, these columns contain a reversed-phase stationary phase that resolves peptides based on hydrophobicity and can be coupled directly or indirectly to the mass spectrometer. Peptides are eluted from the column using an increasing organic solvent gradient with hydrophilic peptides eluting at the start of the gradient.

HPLC columns provide one dimension of separation which may not be sufficient for some complex samples and an additional level of separation prior to analysis may be required. Protein fractionation can be achieved by one or two dimensional (1D or 2D) gel electrophoresis followed by in-gel digestion of protein bands or spots. This also has the added advantage of providing additional information about the mass of the protein that the digested peptides originated from adding confidence that the correct protein assignment has been made by a database search (see Section 1.3).

Multidimensional Protein Identification Technology (MudPIT) allows the separation of peptides by charge and relative hydrophobicity using strong cation exchange (SCX) and reverse phase (RP) liquid chromatography coupled to MS (Liu et al., 2002a). This has been shown to identify thousands of peptides in a sample, enhances the protein coverage and allows the identification of low-abundance and membrane proteins that are often difficult to detect using gel based methods (Kislinger et al., 2005). MudPIT has been utilised in cancer secretomics (Mbeunkui et al., 2006, Sardana et al., 2007) and to investigate the low molecular weight (LMW) proteome (Hu et al., 2006). The study of the LMW proteome has identified cytokines, chemokines, peptide hormones and proteolytic fragments of larger proteins and is often referred to as peptidomics. Petricoin et al have demonstrated that these protein fragments or peptides can be detected using accurate instruments such as the Orbitrap mass

spectrometers with or without trypsin digestion (Petricoin et al., 2004). BRCA2, a cancer-related low abundant protein, is found in the serum as fragments bound to albumin highlighting the importance of the LMW proteome in biomarker discovery (Lowenthal et al., 2005).

1.2.2. Tandem Mass Spectrometry

Mass spectrometers produce mass spectra that plots mass-to-charge (m/z) ratio on the X-axis against relative abundance on the Y-axis regardless of the instrument type or ionisation mode (Rifai et al., 2006). Proteins are enzymatically digested to generate peptides and the sequence of these peptides can be determined using tandem mass spectrometry (MS/MS) and the appropriate database search algorithm. Parent ions are fragmented by collision with a gas (such as nitrogen or helium) to produce daughter fragment ions. The masses and abundances of these product ions are recorded and used to search against *in silico* theoretical digests of proteins in a database of known protein sequences. The identified peptides are then assigned various correlation and scoring factors based upon how close the obtained spectra matches the theoretical spectra.

For ion trap instruments (see section 1.2.3 for description of types of instruments), tryptic peptides are often doubly charged and designated $(M + 2H)^{2+}$ therefore a peptide with a mass of 1232.55 would generate a peak at 617.28 (1232.55 + (2 x 1.0073))/2). Higher charge states are seen in peptides with more than 15 amino acids. The charge state can be calculated from the isotope cluster of peaks that can be found with each signal. These peaks are separated by 1.0073 and are caused by ^{12}C atoms being replaced with a ^{13}C isotope atom which occurs at a ~1% probability rate. Therefore a peak with a mass of 617.28 with an isotope peak at 618.28 must have a charge of 1 (Steen and Mann, 2004). Evaluating the isotope peaks can indicate the resolution of the mass spectrometer. Ion trap instruments are

unable to efficiently separate the isotope peaks of doubly charged species however, the FTICR-MS demonstrates the highest resolution for separating isotopic peaks.

1.2.2.1. Fragmentation Methods

In order to identify peptides, their ions need to be fragmented by the mass spectrometer to acquire MS/MS spectra that can be searched against a database using a search algorithm. This can be achieved by isolating the desired ions (precursor or parent ion) and fragmenting them further by collision with an inert gas such as nitrogen, argon or helium to generate the product or daughter ions. This is known as collision-induced dissociation (CID). The masses of these ions are determined and the fragmentation pattern (the MS/MS spectra) is used to determine the amino acid sequence. Other precursor fragmentation methods include: Electron-Capture Dissociation (ECD) which involves gas phase ions being fragmented by introduction of low energy electrons and is commonly used in fourier transform-ion cyclotron resonance instruments (FT-ICR) instruments only; and Electron-Transfer Dissociation (ETD) which induces fragmentation of cations (e.g. peptides or proteins) by the transfer of electrons. ETD involves protonated peptides being fragmented by electron transfer from a radical anion (such as fluoranthene) to cleave N-C bonds of the peptide backbone to produce c- and z- type ions. ETD allows the analysis of larger peptides and post translational modifications (PTMs) and can be used in ion trap instruments.

'Bottom-up' proteomics describes the process of digesting proteins and analysing the peptide mixture by MS. The studies carried out in this thesis used CID to fragment the peptide ions by cleavage of the backbone amide bonds to produce b- and y- type ions. The precursor ions are fragmented at the amide bonds between amino acids and the product ions that retain the charge at the N-terminus are called b-ions and those that retain the charge at the C-

terminus are known as y-ions (Figure 1.4). This fragmentation method was traditionally used for smaller molecules but is now used for larger intact molecules.

'Top-Down' proteomics refers to identifying proteins by ionising the large intact molecule rather than the smaller fragments generated from a proteolytic digest. Each of these fragments differ by the mass of one amino acid and the observed mass differences can be used to predict the amino acid sequence. This is often complicated by incomplete MS/MS fragmentation patterns and accurate data analysis is dependant on the quality of the spectra as well as the mass accuracy and resolution of the instrument used. The obtained MS/MS spectra may not be complete but may be sufficient to accurately match the fragmentation pattern with a theoretical pattern in a database. This is carried out using an algorithm to search databases for the closest match for example Sequest and Mascot (Eng, 1994, Perkins et al., 1999). See Section 1.3 for description of data analysis methods for shotgun proteomics.

1.2.3. Quantitative Proteomics

Many studies require quantification of peptides and proteins as well as their identification. For quantitative proteomics, relative abundances of proteins can be studied using labeling methods such as stable isotope labeling by amino acids in culture (SILAC), isotope-coded protein labels (ICAT) and isobaric tags for relative and absolute quantization (iTRAQ) or label free methods such as absolute quantification (AQUA) and multiple reaction monitoring (MRM). Peak intensity is affected by experimental factors (such as incomplete protein digestion and peptide solubility) and mass spectrometry specific factors (such as ionisation efficiency and ion suppression). These factors are all reproducible so the peak height of the corresponding peaks of the light and heavy peptides can be used to determine relative abundance of proteins between two samples.

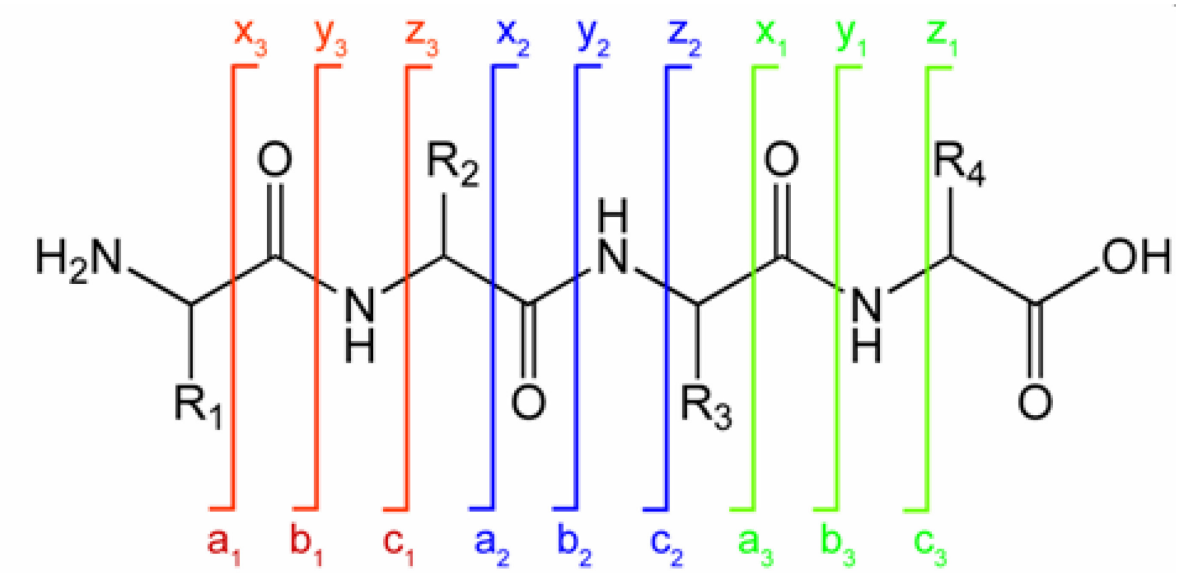


Figure 1. 4: Ion Fragmentation of Peptides.

Diagram to show the nomenclature of the fragmentation of amide bonds between amino acids known as the Roepstorff-Fohlman-Bieman nomenclature exhibited using CID (Biemann, 1988, Roepstorff and Fohlman, 1984). (R1, R2, R3 and R4 represent the amino acid R groups).

1.2.3.1. Stable Isotope Labeling of Amino Acids in Culture (SILAC)

SILAC is a relative quantification method used for the proteomic analysis of cell lines to investigate the expression levels of proteins between two or more samples. Figure 1.5 outlines the basic workflow of a proteomic experiment using SILAC.

The technique was first described by Oda et al in 1999 for labeling simple salts and modified for labeling of amino acids by Chen et al in 2000 (Chen et al., 2000, Oda et al., 1999). The method was then adapted to analyse complex protein mixtures by Ong et al in 2002 (Ong et al., 2002). Two cell lines are cultured in media depleted of arginine or lysine which is then supplemented with 'heavy' or 'light' isotopically labeled amino acids (e.g $[^{13}C_6]$ or $[^{12}C_6]$ labeled lysine or arginine) and dialysed foetal bovine serum (FBS) to ensure there is no contamination from free amino acids or peptides. They are cultured for a minimum of six population doublings to ensure >98% labeling efficiency.

Samples are harvested from both cell lines (cell lysates, conditioned media or subcellular fractions) and mixed in a 1:1 ratio at a very early stage of the experiment such that both samples are treated identically in subsequent steps ensuring that the abundance ratios are maintained throughout and are not affected by experimental errors. The sample is then prepared for MS/MS analysis.

Stable isotope labeling using SILAC offers several advantages (Ong et al., 2002):

- Does not require any changes to routine tissue culture methods already used in the laboratory.
- Does not require peptide labeling steps or protein purification steps after harvesting the proteins. This is important when your starting material for the cell models is limited. The number of manipulations are minimised to reduce loss of sample.

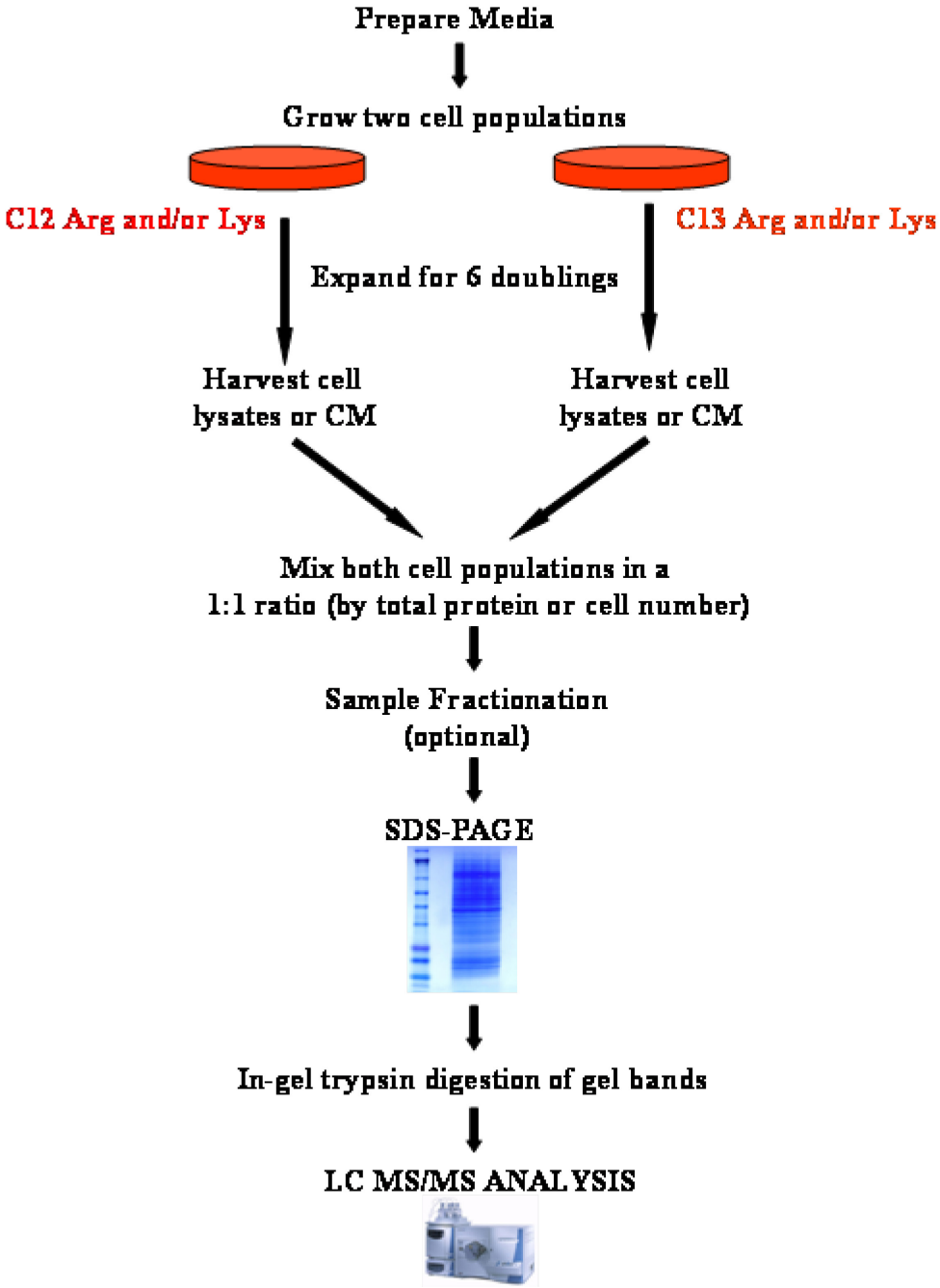


Figure 1. 5 : Stable Isotope Labeling by Amino acids in Culture (SILAC)

Diagram to demonstrate the experimental overview of using SILAC for relative quantification of proteins in cell based models as first described by Ong *et al* in 2002 (Ong et al., 2002).

- The labeling strategy is universal throughout the protein sequence therefore several peptides can be used to calculate the relative abundance.
- Allows the quantification of most proteins as the method does not rely on labeling of single residues that occur infrequently as other previously reported labeling strategies. Arg and Lys residues should be present in all peptides resulting from a trypsin digest except the C-terminal peptide (unless it contains an Arg or Lys residue).

1.2.3.2. ^{18}O labeling

^{18}O labeling was developed for use in comparative shotgun proteomic strategies by Yao et al (Yao et al., 2001). Most shotgun strategies involve digesting proteins to generate peptides and ^{18}O labeling combines both the digestion and labeling steps. It involves digesting two sample pools in parallel; one digested using $H_2{}^{16}O$ (light) and the other with $H_2{}^{18}O$ (heavy). Two heavy oxygen molecules are incorporated into the C-terminus of each peptide generated resulting in a mass difference of four. These two pools are mixed in a 1:1 ratio and analysed by mass spectrometry. As the peptides are labeled at the C-terminus, the y-ions in the MS/MS spectrum appear as labeled or unlabeled doublets with the b-ions remaining unlabeled. The ^{18}O labeling strategy has also been used in a post-digestion protocol (Bantscheff et al., 2004).

Due to the small mass differences observed between the heavy and light labeled peptides in ^{18}O labeling, this strategy was originally thought only to be possible using high resolution and high mass accuracy instruments such as FTICR-MS. However, this labeling strategy has been shown to be possible using low resolution instruments such as ion traps (Heller, 2003).

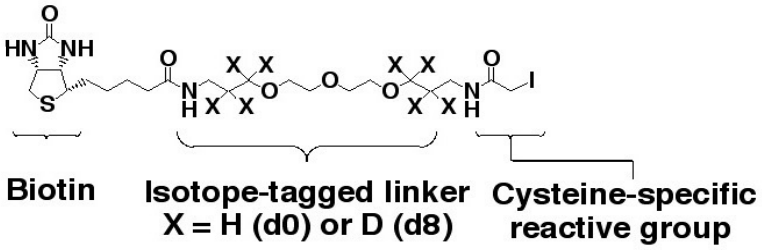
^{18}O labeling is simple and does not rely on the presence of specific amino acids as is the case with other labeling techniques however, it cannot be used to compare more than two samples and incomplete labeling can occur (Ye et al., 2009).

1.2.3.3. Isotope-Coded Affinity Tag (ICAT)

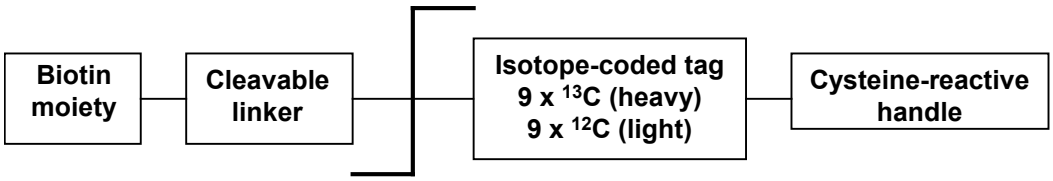
The ICAT reagent consists of an iodoacetamide residue, a linker region containing the heavy or light label and a biotin residue to allow enrichment of the labeled peptides using affinity chromatography (Figure 1.6a). Samples are incubated with either the 'light' or 'heavy' label which binds to the thiol groups of cysteine resides via the iodoaceatmide region of the ICAT molecule. The samples can then be mixed and digested followed by enrichment of the cysteine-containing peptides using affinity chromatography by binding of the biotin label to avidin coated stationary phase.

Originally the linker was tagged with H or D_2 (Figure 1.6a) however, the D_2 labeling affected the hydrophobicity causing the heavy labeled peptide to elute later. More recently, an ICAT reagent with a cleavable biotin molecule has been utilised that uses a [^{12}C] and [^{13}C] labeled linker (Figure 1.6b) (Qu et al., 2006). The 'light' and 'heavy' peptides are then analysed by LC-MS/MS to determine the relative abundance of the peptides using the intensities or the area under the respective peaks. ICAT demonstrates high specificity and sensitivity and is applicable to samples of all origins such as cells, tissues and fluids. ICAT labeling also decreases sample complexity and simplifying data interpretation but does not allow the comparison of proteins that do not contain cysteine residues. Also, data analysis requires high quality spectra with high search scores as often only a single cysteine-containing peptide for each protein is isolated. However, the absolute requirement for a cysteine residue in the peptide sequence does permit greater stringency in the database search.

(a)



(b)



(c)

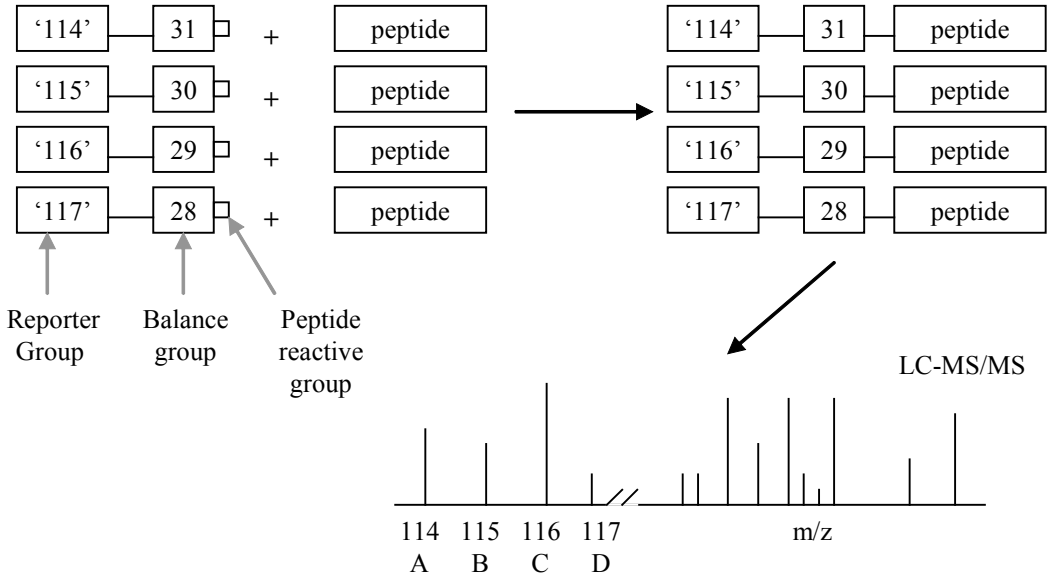


Figure 1. 6 : Examples of Chemical Tags used for Quantitative Proteomics

(a) The structure of the Isotope-Coded Affinity Tag (ICAT) Diagram taken (Kraj, 2008). (b) The structure of cleavable ICAT (cICAT). Diagram taken (Qu et al., 2006). (c) Isobaric Tags for Relative and Absolute Quantification (iTRAQ). Diagram taken (Kraj, 2008).

1.2.3.4. Isobaric Tags for Relative and Absolute Quantification (iTRAQ)

iTRAQ is a method that uses tags of the same molecular weight (isobaric) that produce different ions during fragmentation. It allows the simultaneous comparison of four or eight samples and consists of a reporter group, a balance group and a peptide reactive group. The masses of the reporter groups differ but with the addition of the balance group keeps the total mass of the tag constant (Figure 1.6c). The differences between the tags are seen after fragmentation of the labeled peptides during MS/MS with characteristic peaks at 114, 115, 116, and 117 m/z from the reporter groups. The peak intensity of these ions is used to determine the relative abundance of the proteins in the samples.

iTRAQ is not suitable for ion-trap analysers as they have a lower mass exclusion limit during fragmentation which excludes the characteristic peaks being present in the spectrum.

1.2.3.5. Absolute quantification (AQUA) and Multiple Reaction Monitoring (MRM)

AQUA can be achieved using an internal standard of a known quantity of a stable isotope-labeled standard proteotypic peptide allowing a comparison of the mass spectrometric signals of standard labeled peptide and corresponding endogenous peptide in a digest. This method utilises multiple reaction monitoring/stable isotope dilution MS (MRM/SID-MS) using triple quadrapole mass spectrometers which monitors for a specific intact peptide and its various fragments based on retention times, peptide mass and fragment masses to eliminate ambiguities in peptide identification and quantification. Coupling the MRM method with protein enrichment by immunodepletion or peptide enrichment by antibody capture may allow the measurement of plasma/serum proteins in a high throughput manner for biomarker validation (Bantscheff et al., 2007, Z. Hu, Hood, L. and Tian, Q., 2007).

MRM is a high throughput method that offers greater sensitivity and accuracy than methods used in the discovery phase and has the potential to target 30-100 candidate markers within the same assay (Rifai et al., 2006). When the appropriate standards are used, MRM can provide absolute quantification as well as relative quantification and demonstrates a high specificity for the sample. A fragment of interest is isolated in the first quadrapole and fragmented by collision with a neutral gas in a collision cell (the second quadrapole) and the mass of the resulting product ion is analysed in the third quadrapole. Multiple ion-product transitions can be monitored in a single assay when coupled to a liquid chromatographic elution method.

SID-MS can increase the reliability of the quantification by comparing the signal from an endogenous species with a known concentration of an exogenous stable isotope species. This method has been successfully used to measure human growth hormone and prostate specific antigen in serum or plasma (Barnidge et al., 2004, S. L. Wu et al., 2002) and also been used to quantify the levels of 53 high-abundance proteins in plasma (L. Anderson and Hunter, 2006). To detect lower abundance proteins, depletion of the high abundance proteins is required. For those proteins that are at too low a level to detect by MRM, stable isotope standards and capture by anti-peptide antibodies (SISCAPA) can also be used to capture isotopically labeled and unlabeled target peptides to improve sensitivity and specificity (N. L. Anderson et al., 2004). Labeled and unlabeled peptides both bind to the antibodies therefore quantification is still possible.

1.2.4. Mass Spectrometers

Two of the main components of a mass spectrometer are the ionisation source and the mass analyser.

1.2.4.1. Ionisation Sources

Within the ionisation source, the sample is ionised and desorbed into the gas phase using various techniques. Ionisation is the conversion of a molecule to an ion by the addition of or the removal of electrons to become positively or negatively charged. Desorption is the process of releasing ions from liquid phase to gas phase.

Some ionisation techniques can cause extensive fragmentation to produce fragment ions (ions derived from the fragmentation of a parent ion) and others are softer and only produce molecular ions (ions derived from a neutral molecule by loss or gain of an electron). Two common ionisation methods used in proteomics are electrospray ionisation (ESI) (Fenn, 1988) and matrix-assisted laser desorption ionisation (MALDI) (Karas, 1987). Though they can both be used to analyse large proteins, they have been mainly used for peptide analysis from proteolytic digests of proteins. The ability to directly link the mass spectrometer with a separation method such as HPLC has allowed the development of high-throughput methods for the analysis of biological molecules.

For liquid phase ion sources, such as electrospray ionisation (ESI) and atmospheric pressure chemical ionisation (APCI), the sample is in solution and is sprayed as droplets in to the mass spectrometer. For solid state ion sources, such as matrix-assisted laser desorption laser ionisation (MALDI) the sample is incorporated into a solid or viscous fluid matrix which is then irradiated by energetic particles that desorbs the ions close to the surface of the matrix. An electric field is then applied to extract the ions and focus them towards a mass analyzer.

1.2.4.1.1. Electrospray Ionisation (ESI)

ESI works by applying a strong electric field in a vacuum to a liquid passing through a capillary tube. This induces an accumulation of charged ions at the liquid surface at the end of the capillary tube which in turn causes the formation of highly charged droplets. When a

liquid passes through a thin capillary to which a voltage is applied, a cloud of small droplets (or a spray) is formed. The liquid leaves the capillary in a cone shape (known as a Taylor cone) and charged ions accumulate on the surface of the droplets.

These ions can be negatively or positively charged but in the case of peptides and proteins, these ions typically carry a positive charge due to additives such as formic acid which are added to the solvents to act as proton donors. The sample is then desorped of any solute before entering the mass analyser. Desorption occurs either by passing the sample through a heated capillary or through a curtain of drying gas (such as nitrogen) to evaporate the solvent from the sample (also known as desolvation). This process occurs under atmospheric pressure however, the mass analyser is maintained under a vacuum. It is this pressure differential that draws the ions into the mass analyser. The ions are then separated according to their mass by a mass analyser (Figure 1.7a).

ESI is particularly suited to the study of large biological molecules (such as proteins) due to its ability to produce multiply charged species. The number of charges on a molecule is dependant on factors such as size and the number of basic residues present. Peptides commonly carry a single (+1), double (+2) or triple (+3) charge. As trypsin is the most common protease used to digest proteins in proteomic studies, peptides are typically +1 and +2 charged due to the presence of basic Arg or Lys residues at the C-terminus.

Typically in an ion source, only 1% of the introduced sample is ionised. This is mainly influenced by the initial diameter of the droplets formed in the ESI needle and the flow rate. The ionisation yield (i.e. the sensitivity) can be increased by using a nanoelectrospray needle which has a diameter of 1-20μm. Nanoelectrospray requires lower voltages than ESI (as the delivery needle is usually positioned much closer to the transfer capillary) and uses lower

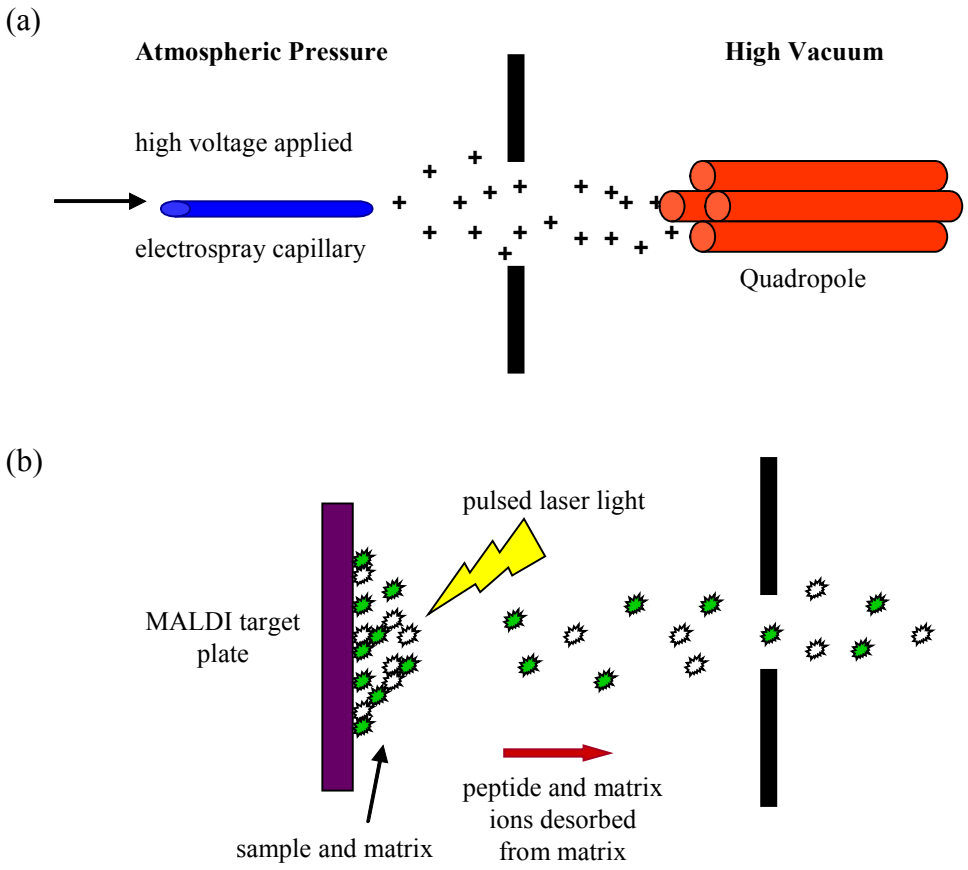


Figure 1. 7 : Ionistion Methods used in Proteomics

Diagram to demonstrate the basic principles of the two most commonly used ionisation methods used in proteomics. (a) Electrospray Ionisation (ESI). The sample in solution is passed through a capillary to which a high positive potential is applied resulting in positively charged droplets to leave the capillary. These droplets then travel towards the mass spectrometer under atmospheric pressure. The solvent within the droplets evaporate and the desolvated ions are drawn towards the low pressure of the mass analyser. (b) Matrix Assisted Laser Desorption Ionisation (MALDI). The sample is co-crystallised with an excess of matrix. Short pulsed laser shots are focused on the sample spot causing the sample/matrix mixture to become volatile and fly towards the vacuum of the mass analyser. (Diagram modified from (Veenstra, 2006)).

volumes of samples due to the decreased flow rate. The droplets that are formed are much smaller increasing the ionisation efficiency.

1.2.4.1.2. Matrix-Assisted Laser Desorption Laser Ionisation (MALDI)

MALDI produces ions by irradiating a solid mixture of sample dissolved in an organic matrix compound with a pulsed laser beam (usually UV or IR) (Figure 1.7b). This results in the sample molecules being both ionised and desorbed from the solid mixture. The solid mixture is usually prepared by mixing a sample with a saturated solution of matrix prepared in solvents such as water, acetonitrile, acetone or tetrahydrofuran. The matrix is a small conjugated organic molecule that is able to absorb energy in the UV region. Typically used matrices are α-cyano-4-hydroxycinnamic acid (CHCA), 2,5-dihydroxybenzoic acid (DHB) and 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) and the choice of matrix used is dependant upon the type of molecule being analysed. The matrix must be capable of dissolving in the same solvent as the sample, be stable in a vacuum, able to absorb the laser wavelength used, be able to co-crystalise with the sample and promote ionisation.

Microliter amounts of matrix are spotted onto a MALDI target plate (can be pre-mixed with sample or spotted after the sample) and dried to allow the formation of a crystal lattice structure incorporating the sample molecule. The plate is then inserted into the source where the laser is applied to the plate. As with ESI, the mass analyser is maintained at low pressure and the source at a higher pressure therefore the pressure differential draws the ions into the mass analyser. MALDI sources can be maintained at atmospheric pressure allowing them to be coupled to mass analysers such as ion traps, and quadrapole time-of-flight (TOF) analysers (see Section 1.2.4.2).

MALDI can produce both positive and negative ions but as with ESI, positive ions are typically generated during peptide and protein analyses. Positive ions are formed by the

acceptance of a proton from the matrix solvent. Most samples accept a single proton therefore predominantly generating ions with a (+1) charge and large mass-to-charge (m/z) ratios. Therefore, MALDI is commonly coupled to mass analysers with large m/z ranges such as TOF, TOF-TOF and QqTOF instruments which are capable of tandem mass spectrometry.

1.2.4.2. Mass Analysers

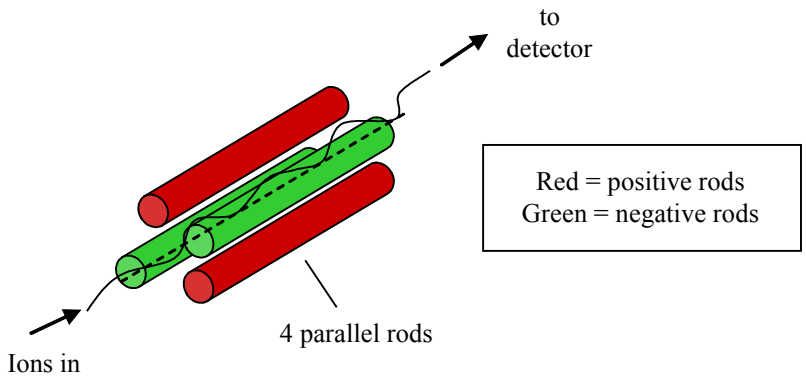
The mass analyser guides the gas-phase ions to the detector where the m/z ratio is determined. The most common types of mass analysers used in proteomics are time-of-flight (TOF), quadrapole, ion trap and ion cyclotron resonance mass analysers. Three important characteristics of a mass analyser are:

- *Upper mass limit*. The highest m/z ratio that can be measured.
- *Transmission*. The ratio between the number of ions reaching the detector and the number of ions produced in the source.
- *Resolution*. This refers to the instrument's ability to produce distinct signals from two ions with a small mass difference.

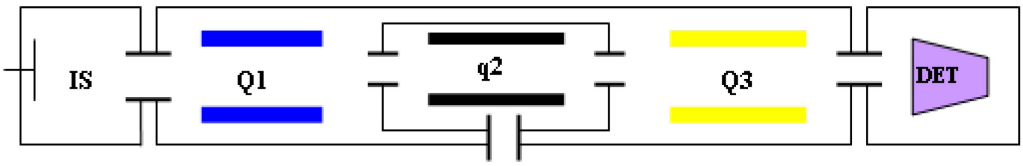
1.2.4.2.1. Quadrapole Mass Spectrometer

A quadrapole is made up of four metal rods arranged in parallel to which a voltage and radiofrequency (rf) is applied to guide and manipulate ions through the mass analyser (Figure 1.8a). The voltage can be altered to allow ions with specific m/z ratios to pass through. Single quadrapole instruments are not commonly used in proteomics as they are unable to perform tandem MS however, triple quadrapole instruments are capable of performing MS/MS analysis when a collision cell is placed between two quadrapoles (Figure 1.8b).

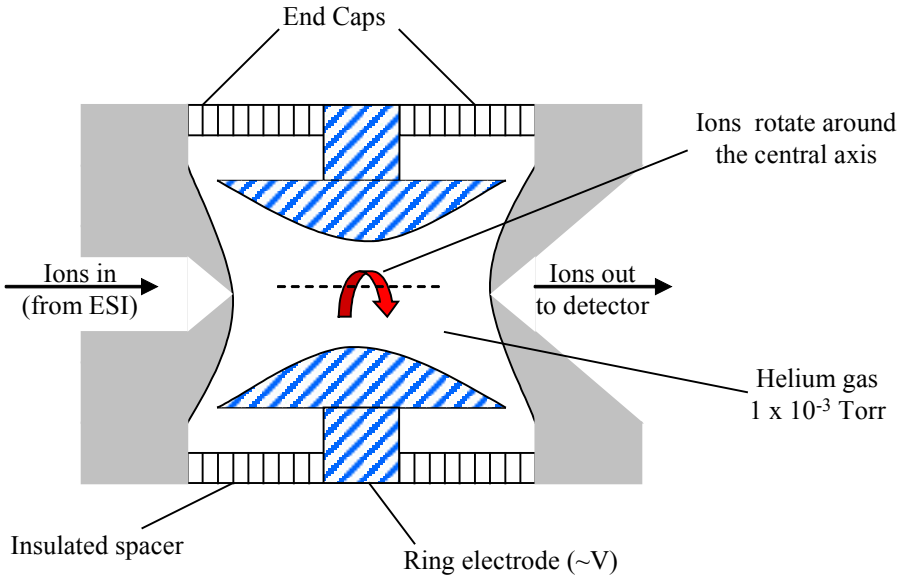
(a)


to
detector
Red = positive rods
Green = negative rods
4 parallel rods
Ions in

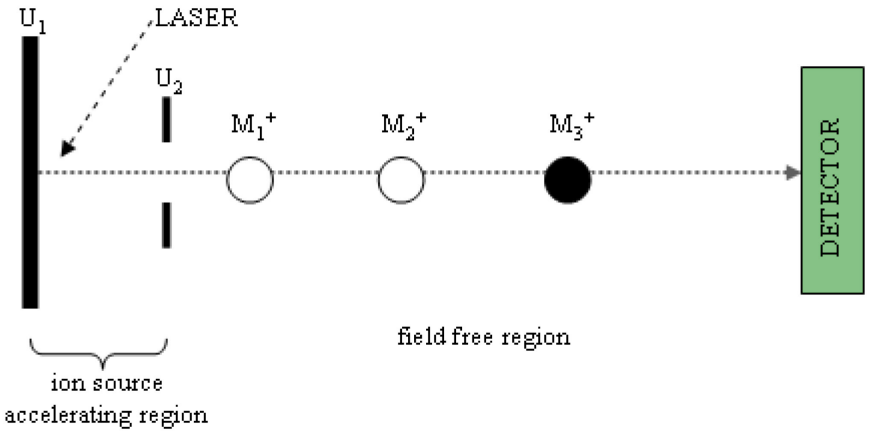
(b)


IS
Q1
q2
Q3
DET

(c)


End Caps
Ions rotate around
the central axis
Ions in
(from ESI)
Ions out
to detector
Helium gas
1 x 10-3 Torr
Insulated spacer
Ring electrode (~V)

(d)



(e)

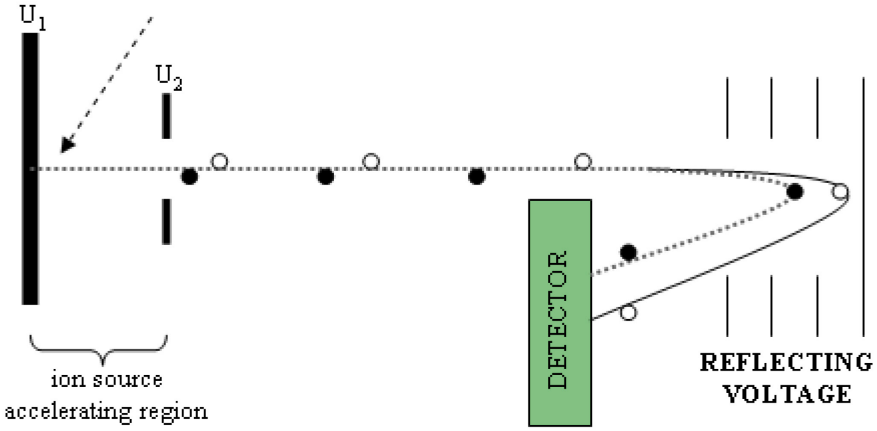


Figure 1. 8 : Mass Analysers

(a) Diagram of a linear quadrapole, Diagram taken from (Burlingame, 2005). (b) Schematic a triple quadrapole mass analyser. IS = Ion Source; Q1 = quadrapole 1; q2 = quadrapole 2; Q3 = quadrapole 3; DET = detector, Diagram taken from (Agnieszka and Silberring, 2008). (c) Ion Trap Mass Analyser. Schematic diagram to illustrate the basic principles of an ion trap mass analyser. Diagram taken from (Burlingame, 2005). (d) Schematic diagram to illustrate the basic principles of linear Time-of-Flight mass analyser (TOF) and reflectron (e) TOF mass analyser. Diagram taken (Kraj, 2008).

Triple quadrapole mass spectrometers used in conjunction with ESI are one of the most commonly used mass analysers in proteomics as they are able to switch between two different scan modes. In the first mode (or full scan mode), ions with a broad range of m/z ratios are allowed to pass through the first quadrapole. In this mode, the ions are then free to pass through the remaining two quadrapoles to be measured by the detector. In the second mode (or MS/MS mode), the first quadrapole only allows ions in a narrow m/z ratio range to pass through to the second quadrapole where they are subjected to fragmentation by collision with an inert gas. The fragmented ions then pass through the third quadrapole to be measured by the detector.

1.2.4.2.2. Ion-Trap Mass Spectrometer

Ion trap mass spectrometers became commonly used for proteomic studies when new developments were made in the ability to manipulate trapped ions. Stafford et al developed the mass instability mode which allowed ions with different m/z values to be 'trapped' over a period of time and ejected sequentially into a detector (Stafford, 1984). The resolution of the instrument was improved by using helium to cluster ions of the same m/z value in the ion trap and eject them faster and more efficiently than diffuse clouds of ions (Figure 1.8c).

Ion traps cycle through rounds of collecting, storing and ejecting ions from the trap and during which it can isolate and fragment peptide ions from complex mixtures and is therefore well suited for proteomic analyses. The ion trap selects and traps specific ions and then adjusts its trapping voltage to eject all other ions from the trap. The applied voltages are then increased to cause the energy of the remaining ions to increase. This in turn causes the ions to collide with helium and fragment. These fragmented ions (daughter ions) are trapped and ejected from the ion trap sequentially according to their m/z value.

Ion trap mass spectrometers are often thought of as 'work horses' of global proteomic studies and are commonly used to analyse complex mixtures of peptides. During these studies the ion traps work in a data-dependant MS/MS mode i.e. each full MS scan is followed by a specific number of MS/MS scans. The full scan identifies the m/z ratio of all the ions in the trap at that time and selects the pre-defined number of the most abundant ions for further fragmentation. Some instruments are also able to trap daughter fragments and subject them to further fragmentation (MS/MS/MS or MS^n). This is not commonly used to identify proteins in complex mixtures but is useful for identifying peptides with post-translational modifications such as phosphorylated peptides.

1.2.4.2.3. Time-of-Flight Mass Spectrometer

Time-of-Flight (TOF) mass spectrometers measure the m/z ratio of ions based on the time taken to move through the mass analyser to the detector. The time taken is proportional to the m/z ratio with smaller ions reaching the detector faster than larger ones (Figure 1.8d).

1.2.4.2.4 Matrix-Assisted Laser Desorption Time-of-Flight (MALDI-TOF)

MALDI-TOF instruments are relatively simple to operate and robust making them extremely useful for generating peptide fingerprints for individual proteins. MALDI-TOF/TOF instruments are also able to perform tandem MS by having two TOF analysers separated by a collision cell. Peptide ions are generated in the source and accelerated through the first TOF tube and become dissociated by colliding with an inert gas (such as air or nitrogen) in the collision cell. The ions are then fragmented and the daughter ions are accelerated through the second TOF tube to the detector. This allows proteins to be identified through peptide fingerprinting and confirmed by MS/MS analysis of specific peptides.

1.2.4.2.5. Quadrapole Time-of-Flight Mass Spectrometer (QqTOF)

The quadrapole time-of-flight mass spectrometer (QqTOF) combines the ion selectivity and sensitivity of a quadrapole with the high mass accuracy and resolution of a TOF instrument (Shevchenko et al., 2000). This instrument has the potential to perform de novo sequencing and can utilise either ESI or MALDI ionization methods to analyse samples (Loboda et al., 2000).

QqTOF instruments usually comprise of two quadrapoles, the first acts as an rf only quadrapole that provides collisional dampening, the second acts as a standard quadrapole which is followed by a reflecting TOF mass analyser (Figure 1.8e). MS measurements are taken in rf-only mode where all ions are allowed to pass through to the TOF tube resulting in high resolution and high mass accuracy spectra. For MS/MS spectra, only specified ions are allowed to pass through to the collision cell where they are fragmented after colliding with a neutral gas. The resulting fragments are highly charged with energy and need to be collisionally cooled before being accelerated and focused towards the field-free drift space of the TOF analyser where they are separated according to their m/z ratio and detected.

Due to its high resolving power and its ability to perform tandem MS, the QqTOF is often used in quantitative studies using isotopic internal standards and isotopic labeling strategies that require measuring small mass differences e.g ^{18}O labeling (see Chapter 1.2.3.2).

1.2.4.2.6. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR)

The fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS) was developed over thirty years ago by Comisarow and Marshall and is increasingly being used to generate high-quality proteomics data (Comisarow and Marshall, 1974). The FTICR-MS consists of an ion trap-like analyser within a high strength magnetic field. The magnetic field causes the ions trapped within the analyser to resonate at their cyclotron frequency. An

electric field is applied which is at or near the cyclotron frequency of the trapped ions which in turn are excited into a larger orbit pattern. The cyclotron frequency is measured by detector plates on opposite sides of the trap and uses fourier transformation to convert the frequencies to m/z ratios. Adjusting the applied electric field can be used to dissociate the ions or eject them from the trap.

These analysers have demonstrated the highest resolution, mass accuracy and sensitivity for peptide and protein analysis compared to all other mass analysers available and whilst they are relatively expensive, they are fast becoming more commercially available.

Section 1.3 : Data analysis for Shotgun Proteomic Approaches for Biomarker Discovery

1.3.1. Introduction to Data Analysis for Proteomics

The analysis of proteomic data is complicated due to the effect of multiple factors such as the diversity of methods and technologies available, the effects of biological artefacts, loss of data during data handling and the complexity of biological systems. The analysis of large datasets produced by shotgun proteomic approaches requires automated programs or algorithms to search and filter the data as manual data analysis is not feasible. Bioinformatics provides a key support for building proteomic knowledge databases. The management, analysis and interpretation of large amounts of raw and processed data require various bioinformatic technologies e.g. databases, sequence comparison, predictive and statistical models.

For biomarker discovery, multiple samples such as healthy vs disease or pre- or post-therapy are compared to produce patterns of protein expression and modifications. Software is used to identify peaks and the relative changes involving the normalisation of the raw data from consecutive runs against internal or external standards. Proteomic techniques produce large amounts of data to provide information on biological systems. The cost of proteomic experiments can often lead to poor experiment design including low number of replicates and insufficient controls leading to poor reproducibility (Brusic et al., 2007).

1.3.2. Factors that Affect Proteomic Profiling Experiments

1.3.2.1. Individual Variation

Genetic studies have identified potential variations associated with more than 1000 hereditary diseases and cancer. Large-scale studies of identical twins have failed to establish a

dominant role of direct genetic factors in uterine carcinoma and inherited breast cancer with only a small proportion of patients with major mutations in genes linked with breast cancer such as BRCA1, BRCA2 and TP53 (T. M. Mack et al., 2002, Thomsen et al., 2006). Subtle genetic changes in gene factors or their binding factors can lead to altered regulation of the cancer associated gene (Yan and Zhou, 2004).

Individual diversity amongst clonal populations of cells can lead to phenotypic and behavioural variation. One study showed that melanoma cells grown in 3-dimensional cultures exhibited distinct expression profiles and an increased number of differentially expressed genes compared to the same cells grown in 1-dimensional cultures (Ghosh et al., 2005).

Individual variation can also arise from experimental factors such as sample collection and pipetting or dilution errors. Differences in collection, handling, storage and profiling techniques used, complicates comparisons between studies. The techniques involved require standardisation to overcome these issues (Schrohl et al., 2008). Individual variation adds another level of complexity to the analysis of proteomic data making it difficult to distinguish between individual and disease specific variation which may involve sophisticated statistics and machine learning (Larranaga et al., 2006).

1.3.2.2. The immune system and disease

The host's response to a disease could be a potential source of useful biomarkers. The immune system plays a complex role during cancer development and progression involving interactions between the immune cells and its mediators with all cells and components of the tumour microenvironment. The innate immune response involves microbial recognition and clearance, inflammation and cell death which when stimulated, initiates the production of pro-inflammatory cytokines. This in turn activates the adaptive immune system.

Immune surveillance is among the most important defenses against cancer. Antigen processing and presentation include antigen degradation by proteases, transport to the site where they bind to MHC molecules, stabilization of MHC molecules and the display of MHC/peptide complexes on the cell surface. Cancer escape from immune recognition involves mechanisms such as the down-regulation of human leukocyte antigen expression, up-regulation of oncogenes and down-regulation of tumour suppressor genes. Cancer cells can also affect immune function in dendritic cells, macrophages and T-cells and inhibit immune recognition (Brusic et al., 2007). For example, differentiation and maturation of myeloid dendritic cells is blocked by tumour environmental factors such as vascular endothelial growth factor (VEGF) (Carmeliet and Jain, 2000, Kryczek et al., 2005). These immune responses reduce the effectiveness of the host's anti-tumour activity.

Proteins that do not originate from the tumour itself but from the host's immune response due to the presence of a tumour could also play an important role in biomarker discovery.

1.3.2.3. Biological Change

The dynamics of change of biological systems brings another level of complexity to the interpretation and understanding of proteomic data. Diagnostics, therapy and monitoring will increasingly depend upon comparisons with large datasets from an extensive range of prior studies. Modeling the data will require sophisticated methods for analysis (Brusic et al., 2007).

1.3.3. Data Analysis Methods

A single tandem MS/MS experiment generates hundreds of thousands of spectra for interpretation and analysis. This involves assigning peptide sequences to the MS/MS spectra,

identifying the proteins from which they originated from and potentially quantifying the peptides and proteins. The data analysis of tandem MS/MS experiments is a complex computational and statistical challenge (Nesvizhskii et al., 2007).

1.3.3.1. Assigning MS/MS spectra to peptide sequences

This is the first and most critical step in any large-scale proteomics experiment. This can be achieved using any one of the following three methods:

1.3.3.1.1. Database Searching

Database searching software such as Sequest or Mascot compares the acquired MS/MS spectra to theoretical peptide fragments from databases of known protein sequences downloaded from online sources such as the National Centre for Biotechnology Information (NCBI) (National Center for Biotechnology Information) and International Protein Index (IPI) (IPI - International Protein Index). Restrictions are set by user-specified criteria such as mass tolerance, enzyme cleavage and post translational modifications (Nesvizhskii et al., 2007). The search generates a list of MS/MS spectra that has been assigned to a peptide sequence accompanied by a correlation score to indicate the degree of similarity between the acquired and the theoretical fragment ion spectra. These correlation scores can be arbitrary values such as Xcorr (used with Sequest) or converted to a statistical measure such as the expectation value (*E*value, used with Mascot). The *E*value refers to the expected number or peptides with scores equal to or better than observed scores under the assumption that peptides are matching the experimental spectrum by random chance.

Database search engines will make assignments regardless of the quality of the acquired mass spectra therefore a selection or filtering process (manual or automated) is required to have confidence in the reliability of the assignments. Statistical validation of the assignments also adds confidence that the correct peptide and protein sequences have been

identified. The need for guidelines for the use of database search engines and the publication of proteomic data was first recognized, reported and developed in 2005 in *Molecular and Cellular Proteomics* (Bradshaw, 2005) and was extended to the publication of clinical proteomic studies in 2008 (Celis, 2008).

The Sequest algorithm utilises a technique called autocorrelation to determine the overlap between the theoretical spectra contained within a database and the observed spectra generated from a proteomic experiment. This overlap is measured by many scoring criteria but the main scores are Xcorr (measure of the correlation between theoretical and observed spectra) and Delta Cn (the difference in Xcorr scores between the top candidate and the next best matching peptide sequence). Sequest is quite robust and is ideal for low signal-to-noise spectra and is also useful for low-resolution data.

The Mascot search algorithm utilises probability-based matching. The theoretical spectra in the database are matched to the observed spectra in a top-down fashion by starting with the most intense b and y ions. The probability of these matches being random is calculated and assigned an identification score.

Every single amino acid substitution in a peptide will produce a different mass (with the exception of Leucine/Isoleucine) so even proteins with low sequence coverage (i.e percentage of protein sequence represented by identified peptide sequences) have the potential to identify different isoforms of the same protein. However, this can prove to be difficult especially for identifying alternatively spliced variants or post translational modifications where the affected peptide is not identified. Many different isoforms of proteins contain identical peptide sequences and distinguishing between these can be difficult. The ideal database would be non-redundant with each gene and all its spliced variants represented. NCBI databases can be large databases, containing many protein entries which claim to be

spliced variants or isoforms but are all actually the same protein sequence i.e. a high degree of redundancy. IPI databases are more suited to proteomics experiments as they contain reported protein sequences, translated cDNA sequences and also information on predicted genes (Steen and Mann, 2004).

1.3.3.1.3 Spectral Libraries

Another method used to assign peptide sequences to acquired MS/MS spectra is spectral matching. Spectral libraries are compiled from a large collection of experimentally observed MS data of correctly identified peptides. The acquired spectra are then searched against this database to identify the closest match. Spectral matching tools include SpectraST, X!P3 and Biblispec. Compared to database searching, this method is quicker and exhibits decreased error rates and increased sensitivity however is not suitable for unknown or novel proteins in a sample, i.e. peptides cannot be correctly identified if they are not represented in the spectral library (Nesvizhskii et al., 2007).

1.3.3.1.3. De Novo Sequencing

This is a method that directly reads the amino acid sequence from the fragment ion spectrum. This was originally performed manually but now software tools are available to assist in the process such as the PEAKS software (PEAKS software). The method allows sequencing of peptides not represented in a database, peptide sequences not previously observed and sequences containing polymorphisms or modifications. *De novo* sequencing is mainly used for species with limited genome information or for identifying modifications but requires high quality MS/MS spectrum and is computationally intensive.

1.3.3.1.4. Hybrid Approaches

These approaches combine elements of database searching and *de novo* sequencing approaches (Mann and Wilm, 1994). It starts with inference of short sequence tags (partial

sequences) from MS/MS spectra followed by an error-tolerant database search to allow one or more mismatches between acquired spectra and those in the database. By limiting the search space to those only containing the sequence tag extracted from the spectrum, the search time is significantly reduced. This approach has proved to be useful for identifying post-translational modifications.

1.3.3.1.5. Estimation of False Positive Rates

There are many cases of the incorrect peptide sequences being assigned to MS/MS spectra in a database search due to factors such as low quality spectra, the presence of contaminants, simultaneous fragmentation of peptides with similar m/z values and an incorrect determination of charge state or peptide mass (Nesvizhskii et al., 2007). Manual inspection of MS/MS spectra and database search results will give confidence that the correct peptide sequence has been assigned. However, this requires a degree of experience and isn't viable for large datasets as it is incredibly time consuming and subjective depending upon the users expertise. The data can be filtered using cut-off values for the threshold parameters used by the database search software but these need to be determined for all experiments as the error rates will vary from one experiment to the next. A high through-put method is required to statistically measure the confidence of peptide assignments and estimate the error rates.

Two of the approaches used to calculate false positive rates are:

(1) *Target-Decoy Searching*.

This approach involves searching the acquired MS/MS spectra against a decoy-target database and optimising the cut-off score for each dataset. The target-decoy database contains a randomised or reversed (decoy) sequence of every protein represented in the database as well as the correct sequences. This approach assumes that matches to decoy sequences and the correct sequences follow the same distribution pattern. The false-positive rate can be

determined by calculating the percentage of peptides assigned to decoy sequences. The peptide assignments can then be filtered using cut-off scores optimised to maintain a low false-positive rate. This approach is relatively simple but does double the database search time (Elias and Gygi, 2007).

(2) *Empirical-Bayes Approach.*

This method has been used by Keller et al to develop software known as Peptide Prophet (A. Keller et al., 2002a). It is reported to be a robust and accurate statistical model to assess the validity of peptide identifications made by MS/MS and database searches. Peptide sequences assigned to spectra are evaluated with respect to all other assignments in the dataset, including some incorrect assignments. Database search scores and the number of tryptic termini of the assigned peptides is employed using machine learning techniques to distinguish correctly from incorrectly assigned peptides in the dataset. The algorithm then computes a probability score for each peptide assignment.

Experimental factors to consider when deciding if the correct peptide sequence has been assigned to the acquired MS/MS spectra include the mass accuracy of the instrument used, separation of observed peaks (retention time is important for quantification experiments) and the presence of modifications (may have been utilised in purification methods).

1.3.3.2. Assigning Protein Identifications to Peptide Assignments

The aim of most shotgun proteomic experiments is not to identify peptides but to identify the proteins present in a sample prior to enzyme digestion which can be challenging for large datasets. Many correctly identified peptides tend to group to a small number of proteins and many incorrectly assigned peptides with high search scores tend to group to a large number of proteins. This can make the identification of low abundance proteins difficult.

Also, shared peptide sequences in multiple protein entries in a database due to the presence of homologous proteins, alternative spicing variants and the presence of redundant protein entries in the database can add complications to the analysis (Nesvizhskii et al., 2007). It can be difficult in shotgun proteomic studies to differentiate between different protein isoforms.

Protein identifications are determined by the identified peptide sequences from the acquired MS/MS spectra and grouping these into proteins. How protein identifications are reported and the confidence levels of peptide/protein assignments used differ from one study to the next. Some studies report all proteins identified with at least one unique peptide using defined filter criteria, however, these results often have a high false positive rate. Commonly, only proteins with multiple peptides identified are reported for increased confidence the protein is a true identification. Other studies report a representative protein among isoforms and homologs. The most common method is to use algorithms to determine the smallest number of proteins that can account for all the observed peptides with single-hit peptides carrying a low confidence score. This provides a consistent and concise method of reporting the results of proteomic experiments.

The identified peptides and proteins are scored using differing criteria (depending on software used) to enable the user to determine how well the observed spectra matched spectra from the database. Multiple peptide matches from a single protein increases the confidence of that protein being present in a sample however these peptides need to be acceptable matches with high scores. Also, only peptides that are ‘full’ tryptic peptides (i.e peptides with an arginine or lysine residue at its C-terminal end and preceding the N-terminal residue) should be counted. For a protein assignment using only a single peptide, a very high score using the chosen criteria is required and is only really possible if the spectra is of high quality with high signal-to-noise ratio using instruments with high mass accuracy.

Proteomic experiments producing small datasets, i.e. where only a few proteins need to be identified from single gel bands or spots, it is possible to manually inspect all spectra to determine the quality of the data and analyse the protein identities responsible for differential staining observed between two samples. The analysis of larger datasets involving hundreds or even thousands of proteins from a single experiment is far more challenging where manual inspection of all spectra is time consuming. These large datasets tend to have quite high error rates without appropriate filtering of the raw data. The number of incorrectly identified peptides and proteins (false positives) can be decreased by increasing the threshold of the scoring criteria excluding a large proportion of the assignments leading to a risk of missing true peptides (false negatives).

Several strategies have been developed to overcome these problems. Multiple rounds of tandem MS have been used to determine false positive and false negative rates using various threshold criteria for peptide identification (A. Keller et al., 2002b). The target-decoy approach can also be used for protein identification as well as peptide identifications as described in Section 1.3.3.1.5. Nesvizhskii et al have also used a method known as Protein Prophet by which the peptide identification scores were plotted against their frequency (Nesvizhskii et al., 2003). This produces two Gaussian-shaped distributions, one for low scores which is caused by random matches and one for high scores which is caused by true matches. These distributions can be used to determine a probability score for correct identification. All peptide identifications are then grouped into protein identifications and a protein probability score is determined. The individual peptide probability scores are increased if there are other peptides found for the same protein or decreased if no other peptides are assigned to the same protein to produce an adjusted probability score. It allows the data to be filtered according to protein probability score or false positive rates. The

advantage of this method is that it allows a standard method of analysis using multiple file formats from various instruments across many laboratories allowing comparisons between studies.

The absence of a protein in a database search from a large-scale proteomics experiment is not proof that the protein isn't present. Peptides from the protein may have co-eluted with higher abundances peptides and therefore may not have been selected for further fragmentation and identification. This situation is more problematic because of the phenomenon known as ion suppression where an abundant ion suppresses the ionisation of low abundance ions that are co-eluting. Peptides derived from the most abundant proteins can cause significant ion suppression of peptides from the lower abundance proteins masking the signals of ions with similar retention times. Therefore the abundant proteins can limit the amount of sample that can be used in a single LC-MS/MS run (Rifai et al., 2006).

1.3.4. Data Analysis for Quantitative Proteomics

1.3.4.1. Chemical and Stable Isotope Labelling

Chemical tags such as ICAT and iTRAQ and metabolic labeling strategies such as SILAC introduce a mass difference to the MS spectra allowing the relative abundance to be calculated using the relative intensity or area under the corresponding peaks. In the case of calculating the relative abundance, the label should not affect the chemistry of the peptide and therefore the 'heavy' and 'light' peptides should elute within the same time window also adding confidence that both peaks are results of the same peptide sequence.

1.3.4.2. Spectral Counting

Two samples are analysed in separate LC MS/MS runs using the same data acquisition protocol. A separate list of proteins is generated for each sample and the lists are compared to

identify differentially expressed proteins. Protein abundance is estimated from the number of MS/MS spectra identified corresponding to each protein normalised to account for protein length or expected number of tryptic peptides. One variation of this method is to also consider peak intensity but this is difficult to apply to low abundance proteins with only one or two peptide identifications. Spectral counting is not used very often and less accurate than isotope labeling but has been utilised in several studies (Ishihama et al., 2005, Old et al., 2005, Zybailov et al., 2005).

1.3.4.3. Spectral Feature Analysis

This is a label-free strategy that does not require identification of peptide sequences prior to quantification. Biological samples are analysed in separate LC MS/MS runs and computational tools are used to correlate the spectral features across the runs with minimal information from the MS/MS spectra (Radulovic et al., 2004, H. Zhang et al., 2005a). This method allows high through-put analysis and could be applied to profiling multiple biological samples required for biomarker discovery studies. However there are disadvantages, the computational complexity is greatly increased due to the number of spurious spectra, the background noise and the requirement for increased stringency in the robustness and reproducibility (Listgarten and Emili, 2005, Qian et al., 2006). This approach often requires a targeted workflow to verify the presence of features and abundance (L. Anderson and Hunter, 2006).

Section 1.4 : Cancer and Biomarker Discovery

1.4.1. Introduction

Cancer is a complex disease resulting from multiple changes and mutations accumulating in the cell. Histological classification distinguishes the five major types of cancer, i.e. carcinoma, leukaemia, lymphoma, myeloma and sarcoma and various subtypes. The molecular signature of a tumour varies with its subtype, shape, size, host characteristics and effects of the therapies. Biomarker discovery can encompass all aspects of tumour biology. It was proposed by Hanahan and Weinberg that normal cells need to acquire a number of the following six properties for the malignant process to begin: 1) uncontrolled growth; 2) insensitivity to anti-growth signals; 3) evasion of apoptosis; 4) unlimited potential to replicate; 5) ability to form and sustain new blood vessels (angiogenesis) and 6) invasion and metastasis to distant sites (Hanahan and Weinberg, 2000).

Proteomic and genomic studies could contribute to the understanding of the mechanism by which each of the six hallmarks of cancer impacts upon individual cancers. Full understanding of the pathways involved in normal and disease states and their response to various therapies is essential to determine effective treatments across a variety of cancers.

Carcinogenesis is initiated by alterations in the genome of a target cell caused by many factors including genetic, environmental, dietary and immune factors. These alterations include single point mutations, deletions, amplifications, translocations, chromosomal rearrangements, DNA methylation and other factors causing activation of oncogenes or inactivation of tumour suppressor genes (Martinkova et al., 2009). Genes shown to be associated with tumour progression are involved in a variety of cellular processes and disrupt signalling pathways and cellular communication. Many of these processes are mediated by

growth factors, hormones and cytokines via their receptors. Studies have also shown that as well as abnormalities in cancer related genes/proteins, the epigenetic regulation of genes (mediated by modifications to the DNA and/or histones without alteration to the nucleotide sequence of the gene) is also critical in carcinogenesis. Such modifications include DNA methylation and covalent modifications such as acetylation, methylation, phosphorylation and ubiquitination of specific amino acid residues of the N-termini of the core histones (Smith et al., 2007). It has been shown that cancer is accompanied by a genome wide histone hypoacetylation and histone deacetylase inhibitors have been shown to exhibit anti tumour activity (Hake et al., 2004, Pan et al., 2007).

Cancer was once thought to be caused by alterations in a small number of genes known as oncogenes and tumour suppressor genes and the identifications of these alone would lead to ways of preventing and treating cancer. More recently, studies have shown that tumours are far more complex and heterogeneous and involve alterations in multiple pathways and factors. Genes can be significantly amplified in a tumour without altering the RNA levels of that gene therefore the amplification is not represented at the protein level and vice versa.

Currently, studies are able to identify large numbers of differentially expressed genes using gene microarray techniques which does not reflect the changes in protein expression or proteins arising from alternative splicing and post-translational modifications. Proteomics combined with enrichment/ fractionation protocols, has the potential to identify these spliced variants and modified proteins that cannot be analysed by gene microarray analysis.

1.4.2. Cancer Proteomics

Many advances in research have contributed to our understanding of the molecular basis of cancer but increased knowledge of the processes involved in tumour progression and

development will lead to improved strategies for early diagnosis and therapy. Proteomics involves large-scale identification and characterisation of protein expression levels under specific conditions in defined samples. Proteomic techniques can be utilised in various aspects of cancer research including: profiling of cell lines, tumour tissue, subcellular components and biological fluids; biomarker discovery for diagnosis and early detection; identifying novel therapeutic targets and identifying markers for monitoring therapeutic strategies. Changes in the molecular mechanisms that occur within the tumour cell can be disease causing and therefore could be targets for therapies or may be a host response that is a consequence of the disease.

Clinical proteomics most commonly involves the following studies:

- Comparing healthy versus cancer tissue or biological fluids (using controls that eliminate markers of other diseases or other factors that may interfere)
- Comparison of protein expression levels pre- and post-treatment
- Comparison of tissues at various stages of disease

These studies usually result in lists of differentially expressed proteins generated by semi-quantitative analysis of relative abundance. Many candidates at this discovery level may be false positives due to factors such as the low abundance of potential biomarkers in complex proteomes, instrument sensitivity and detection limits all of which can potentially yield artificial differences. Therefore in order to have greater confidence that a particular protein is present or differentially expressed in your biological model, great care must be taken at the data analysis stage to eliminate as many false positives as possible without losing any true identifications. The biological model also needs to be robust to ensure that any interesting results can be validated in replicate experiments and using other methods.

1.4.3. Biomarker Discovery

As described in Section 1.1.2, a biomarker can be described as a measurable indicator of a biological state and usually refers to the presence of and/or stage of disease (Rifai et al., 2006). The discovery of new cancer biomarkers could improve the early detection of the disease, increase accurate prognosis, predicting patient response to treatment and be useful for monitoring patients during treatment. These markers could originate from the tumour itself or come from the host's response to the tumour and could be detectable in body fluids such as plasma/serum, cerebrospinal fluid and urine However, it is worth noting that the final concentration of biomarkers in the bloodstream is considerably lower compared to interstitial fluids due to the dilution of the proteins in the total blood volume. Proteins are released into the interstitial fluid surrounding the tissue which then passes through the lymph vessels before being released into the bloodstream but blood tests for biomarkers would often be the most applicable option in a clinical setting.

There is a vast amount of cancer research that focuses on improving current treatments and discovering novel treatments for advanced disease as the majority of patients are diagnosed at an advanced stage. Although the survival rates for all colorectal cancer cases have improved over the last 20yrs, the prognosis for patients diagnosed with advanced stage disease still remains relatively poor (Etzioni et al., 2003). In general, the survival rates for patients with early stage cancers at diagnosis are relatively good and considerably improved in comparison with late stage cancers. Therefore reliable biomarkers for the detection of early stage cancer would be vital to ensure patients receive treatment as soon as possible.

Any screen for early detection must be able to accurately distinguish between healthy and cancer patients with low false-negative (proportion of cancer patients that test negative) and false-positive rates (proportion of healthy patients that test positive). Ideally the screen

should also distinguish lesions that require treatment and those that are not aggressive and do not require treatment.

The development of a screen for the early detection of cancer involves the discovery, development and evaluation of potential biomarker. It was suggested by Pepe et al that there should be five phases of biomarker discovery (Pepe et al., 2001). These phases include: (1) pre-clinical exploratory studies to identify potential candidates; (2) assay development and validation; (3) retrospective studies to investigate the assay in patient samples prior to diagnosis; (4) prospective screening studies in asymptomatic patients including follow-up and (5) cancer control studies to investigate the impact of the assay on morbidity and mortality rates (Etzioni et al., 2003, Pepe et al., 2001).

1.4.3.1. What makes a good biomarker?

Biomarkers need to exhibit high sensitivity and specificity and exhibit a low false positive rate for a particular disease state. They also need to be robust, i.e. can be easily identified in samples collected and processed in a controlled environment such as a research lab as well as in a clinical setting where sample collection and processing is much more variable (Lescuyer et al., 2007). Proteomics based biomarker studies in serum/plasma samples need to be designed to identify proteins with real potential as biomarkers before proceeding to costly, time-consuming clinical trials. Such studies should include a discovery and a validation phase.

Many single cancer markers currently used for cancer screens exhibit low sensitivity and specificity and using a panel of multiple markers specific for a cancer type could increase sensitivity and specificity of these screens (S. Hu et al., 2006). Many proteins identified in biomarker discovery studies are acute phase proteins that are common amongst different cancer types. These are relatively abundant and overexpressed in other types of diseases such

as autoimmune disease. Many groups believe that a single biomarker is unlikely to be sufficient as a diagnostic tool. For example, an ovarian cancer study identified four proteins (leptin, prolactin, osteopontin and insulin-like-growth factor-II) whose expression levels were measured in a simple blood test and was able to distinguish between disease free and early ovarian cancer with a specificity and sensitivity of 95% when used in combination. However, each of the markers showed very poor specificity and sensitivity when used separately (Kumar et al., 2006).

New proteomic technologies are being used to attempt to improve these studies to identify low abundance proteins that are cancer type specific. Table 1.2 shows examples of potential biomarkers identified in various cancer types using proteomic methods.

1.4.4. Potential Sources of Biomarkers

1.4.4.1. Body Fluids

1.4.4.1.1 Blood

Cancer initiation and progression does not just involve the tumour cells themselves but also other processes that occur within the tumour microenvironment (Radisky and Bissell, 2004). Blood collection is a relatively non-invasive procedure and it is accepted that tissues release proteins into the surrounding biological fluids. Biomarkers released by both the tumour and the micrenvironment (by secretion or leakage) could be expected to be present in the interstitial fluid and blood therefore it is expected that proteomics could have an important role to play in blood based biomarker discovery. The pattern or the amount of the proteins released may be altered in disease states and these disease-associated markers could be measured in body fluids such as blood. This hypothesis has proved to be true as demonstrated by the discovery of Prostate Specific Antigen (PSA) as a marker of prostate cancer

Cancer Type	Techniques	Identified Markers	Ref
Monitoring anti-cancer therapy			
Breast	SELDI-TOF MS (tissues)	High levels of cytosolic ubiquitin and low levels of ferritin light chain associated with good prognosis during therapy.	1
	MALDI-TOF MS (cell lines)	Elevated 14-3-3σ associated with drug resistance in MCF7 / AdVp3000 cells.	2
Hepatocellular carcinoma	2-DE and MS/MS (tissue)	Elevated levels of chaperone proteins such as Hsp27, Hsp70 and glucose-regulated protein 78 associated with a stressful cancerous microenvironment to enhance tumour growth and metastasis	3
Lymphoma	2-DE (tissue)	Hsp27 over-expression and down-regulation of thioredoxin peroxidise 2 and protein disulphide isomerise associated with decreased survival times in B cell chronic lymphocytic leukemia	4
Nasopharyngeal carcinoma	SELDI-TOF MS (serum)	2 isoforms of serum amyloid A protein useful for monitoring relapse	5
Diagnosis			
Bladder	2-DE and MS/MS (tissue)	annexin V, heat shock protein 27, lactate dehydrogenase, annexin I, galectin-1, lysophospholipase, 15-hydroxyprostaglandin	6
Breast	ICAT and MS/MS (NAF)	Alpha2-HS-glycoprotein,lipophilin B, beta-globin, hemopexin, vitamin D-binding protein precursor	7
Colorectal	Immunoblotting and tissue microarray (tissue)	ANXA3, BMP4, LCN2, SPARC, MMP7, MMP11	8
	2-DE and MS/MS (tissue)	secretagogin	9
Oesophageal	2D-DIGE (tissue)	periplakin	10
Hepatocellular carcinoma	2-DE and MALDI-TOF/TOF (serum)	Pro-apolipoprotein, alpha2-HS glycoprotein, apolipoprotein A-IV precursor, PRO1708/PRO2044 (C-terminal fragment of albumin), leucine-rich alpha2-glycoprotein, alpha1-antitrypsin	11
	2D-DIGE	14-3-3γ	12
Lung	2-DE and MALDI-TOF/TOF (serum)	Immunoglobulin lamda chain, transthyretin monomer, haptoglobulin-alpha 2, 2 isoforms of serum amyloid protein, fragment of apolipoprotein A-1	13
Nasopharyngeal carcinoma	2-DE and image analysis (tissue)	Ceruloplasmin (confirmed in serum by ELISA)	14
Ovarian	MALDI-TOF MS (serum)	Haptoblobulin precursor, transferrin precursor (confirmed by western blotting)	15
Pancreatic	2-DE and MS/MS (serum)	Fibrinogen γ	16

Table 1. 2 : Potential Biomarkers Identified in MS-based Proteomic Studies.

Table to show examples of proteomic studies investigating markers for monitoring cancer therapy and diagnosis of cancer. References: 1 - (Granger et al., 2005), 2 - (Huang et al., 2005b), 3 - (Petricoin et al., 2004), 4 - (Lowenthal et al., 2005), 5 - (Wang et al., 2005), 6 - (Doustjalali et al., 2006), 7 - (Ahmed et al., 2005), 8 - (Bloomston et al., 2006), 9 - (Fields and Chevlen, 2006), 10 - (Wang et al., 1981), 11 - (Amacher, 1998), 12 - (Antman et al., 1996), 13 - (Hsich et al., 1996), 14 - (Simpson et al., 2008), 15 - (Radisky and Bissell, 2004), 16 - (Latterich et al., 2008).

(Wang et al., 1981), liver transaminases for liver cell destruction (Amacher, 1998) and Troponin I and T for acute myocardial infarction (Antman et al., 1996).

Plasma samples can be obtained from blood if collected in the presence of anti coagulants (EDTA, sodium citrate or heparin) and centrifuged to remove blood cells (S. Hu et al., 2006). Serum samples are obtained from blood collected in the absence of anti-coagulants by centrifugation following blood clotting. Blood collection protocols can differ between clinics which may result in differences in protein profiles between studies. The timing between the blood collection and serum/plasma preparation and how the samples are stored may affect the protein concentration and its proteomic profile due to proteolysis of proteins by proteases in the circulation (Latterich et al., 2008). Therefore, there is a requirement for a uniform method for blood collection, processing and storage for proteomic studies (Schrohl et al., 2008).

One challenge of using serum/plasma for biomarker discovery is due to the dynamic range of protein abundances present e.g. albumin is present at ~ 40mg/ml and cytokines can be present at levels as low as ~$5x10^{-9}$ mg/ml (Simpson et al., 2008). There are 22 proteins (including albumins, immunoglobulins, alpha-1-antitrypsin and fibrinogen) that account for 99% of the protein content of serum/plasma making it difficult to detect and identify less abundant proteins therefore highlighting the need for fractionation. One strategy that has been employed is to deplete the abundant proteins using a variety of approaches including immunological and/or bio-specific depletion (Bjorhall et al., 2005). Immunoaffinity depletion columns are commercially available from (Agilent, Palo Alto, CA, USA, Beckman Coulter Fullerton, USA and Sigma, USA). These techniques are able to remove ~99% of these proteins however, their remaining concentrations can still remain high enough to pose a problem e.g. albumin can still be detected at ~$5x10^{-3}$ mg/ml which is 10^4 and 10^6 fold higher

concentration than CEA and IL-6 respectively. Therefore additional strategies are still required to detect lower abundance proteins by MS. There is also the risk that these low abundance proteins could be removed along with the most abundant proteins as these are thought to act as carriers for smaller molecules as shown by the loss of some physiologically important cytokines when albumin was removed (Granger et al., 2005). This can be overcome by disrupting the binding between the LMW proteome and the carrier proteins by adding 5% acetonitrile (final concentration) to serum.

One study has addressed the problem of the high dynamic range in blood using orthogonal three-dimensional protein separations (Wang et al., 2005). Immunodepletion was followed by three dimensions of separation based on charge, hydrophobicity and mass however such approaches are be difficult to perform reproducibly.

Another approach that could be used to fractionate serum/plasma samples is to use a targeted strategy to investigate subproteomes such as glycosylated proteins. Many proteins in serum/plasma are glycosylated and their glycosylation state can be linked to cancer and other diseases (S. Hu et al., 2006). Lectin-affinity columns (LAC) can be used to enrich the glycoproteome prior to MS analysis as demonstrated by Bunkenborg et al with the identification of 86 N-linked glycosylation sites in 77 proteins (Bunkenborg et al., 2004). The glycosylation state of proteins is known to change in cells during cancer development therefore serum glycomics is an emerging approach in biomarker discovery. Many studies investigate glycosylated proteins using antibody based immunochemical tests in various cancers: ovarian (CA125); breast (CA27.29, CA15-3); pancreatic, gastric and colorectal (CA19-9).

1.4.4.1.2. Urine

Urine is a relatively non-invasive biological sample to obtain from patients and a potential source of biomarkers in some types of cancer. However, there are many issues to overcome when using urine for proteomic studies. Typically, urine has a low protein concentration, contains high levels of salts and interfering compounds and also demonstrates a high level of variation between patients suffering with the same disease and between samples collected from the same patient (Domon and Aebersold, 2006).

Urine is most commonly used in biomarker studies for bladder and renal cancer however, has also shown potential in prostate, lung, ovarian and colorectal cancer (Tantipaiboonwong et al., 2005, Theodorescu et al., 2005, Ward et al., 2008a, Ye et al., 2006). Sample preparation can include desalting, dialysis or precipitation using solvents such as acetone, acetic acid or acetonitrile (Hu et al., 2006). Some studies have also used immunodepletion of the abundant proteins to identify the low abundant proteins or used pre-fractionation methods such as molecular weight cutoff columns. Two-dimensional electrophoresis has also been used to investigate the urinary protein excretion patterns in healthy men compared with men with kidney disease (Lafitte et al., 2002).

One study has also shown the presence of exosomes in urine that are a rich source for biomarker discovery. Pisitkun et al identified 295 proteins in urinary exosomes many of which are known to be associated with renal and systemic diseases (Pisitkun et al., 2004).

1.4.4.1.3 Other Body Fluids

Cerebospinal fluid (CSF) has been used to investigate markers of neurodegenerative and CNS disorders. Typically, it has a very low protein concentration and the most abundant proteins originate from blood plasma masking the less abundant proteins in proteomic studies (Hu et al., 2006). CSF has been used in proteomic studies to investigate markers for

neurological diseases such as Creutzfeldt-Jakob disease (CJD) and Amyotrophic lateral sclerosis (ALS) (Davidsson et al., 2002, Ranganathan et al., 2005).

Saliva has also been used in biomarker discovery due to the ease and non-invasive sample collection with very little cost. Proteins found within saliva can be used as markers of oral and systemic disease but has been used to investigate markers of breast cancer as well as oral cancer (Fingleton et al., 2004, Streckfus et al., 2000).

Bronchoalveolar lavage fluid (BALF) contains secreted pulmonary proteins and is used to investigate the components of the epithelial lining to study lung disorders (such as sarcoidosis andiopathic pulmonary fibrosis) and the molecular structure of the lung (Rottoli et al., 2005).

Synovial fluid (SF) can be found in the cavities of synovial joints and reduces friction between cartilage and other tissues and is often used to investigate the cause of acute arthritis. Biomarkers in SF for rheumatoid arthritis (RA) and osteoarthritis (OA) have been studied using proteomic methods (De Ceuninck and Berenbaum, 2009, Wilson et al., 2009).

Nipple aspirate fluid (NAF) contains cells and proteins and can be collected non-invasively. It can be used to study the presence of malignant cells and markers for breast cancer and has been the focus of many proteomic studies (Alexander et al., 2004, Sauter et al., 2005, Varnum et al., 2003).

1.4.4.2. Tumour Tissue

Many studies have used tumour tissues to identify candidate biomarkers in various cancer types including HCC, pancreatic, gastric and lung (Lee et al., 2005, Li et al., 2004a, Li et al., 2006, Li et al., 2004b, Nishigaki et al., 2005, Shen et al., 2004, Yokoyama et al., 2004) . As there are numerous challenges when using blood for biomarker discovery, other studies are concentrating on tissue biopsies under the hypothesis that potential markers are at their

highest concentrations in the tumour microenvironment compared to the bloodstream. Laser capture microdissection (LCM) reduces heterogeneity by allowing the selection of specific cell types from tumour tissue (Emmert-Buck et al., 1996). This method has been used alongside DIGE to identify potential tumour markers in pancreatic adenocarcinomas (Sitek et al., 2005).

1.4.4.3. Cell-based Models

Although tumour cell lines do not represent the heterogeneity of cells within a tumour, they are commonly used to investigate protein changes during various stages of carcinogenesis such as epithelial-mesenchyml transition (EMT) in TGF-β treated lung cancer cells (Honda et al., 2005), p53-induced apoptosis in colorectal cancer cell lines (Gu et al., 2004) and the differences between androgen-dependant and androgen-independent prostate cancer cells (Lin et al., 2005).

Another strategy for biomarker discovery is to investigate secreted proteins from cell lines. Such proteins include growth factors, extracellular degrading enzymes, cell motility factors, cytokines and bioactive molecules involved in essential processes such as differentiation, invasion, metastasis, angiogenesis and regulation of cell-to-cell or cell-to-extracellular matrix interactions (Xue et al., 2008). It is possible that many of these proteins can enter and be detected in body fluids such as blood and urine.

Proteomic methods have been used to investigate novel secreted proteins or proteins with differential patterns of secretion in colon, prostate, breast and lung cancers (Kulasingam and Diamandis, 2007, Pereira-Faca et al., 2007, Sardana et al., 2007, Volmer et al., 2005). These methods include gel-based techniques such as 2DE / 2D-DIGE (Huang et al., 2006b, Khwaja et al., 2006, Volmer et al., 2005), gel-free methods (Kislinger et al., 2005, Mbeunkui et al., 2006, Sardana et al., 2007), SELDI-TOF MS (Xue et al., 2008) and quantitative

methods such as ICAT, iTRAQ and SILAC (Chenau et al., 2009, Gronborg et al., 2006, Martin et al., 2004).

One example of a cancer secretome study identifying a potential novel biomarker used a gel-based method to investigate the secretome of two nasopharyngeal carcinoma (NPC) cell lines (NPC-TW02 and NPC-TW04) (C. C. Wu et al., 2005). They identified three metastasis-related proteins (fibronectin, Mac-2 binding protein and plasminogen activator inhibitor I) with increased secretion in NPC which were then shown to be elevated in serum from NPC patients compared to healthy controls using an ELISA based assay. A study investigating novel biomarkers in renal cell carcinoma (RCC) identified pro-matrix metalloproteinase-7 (pro-MMP7) as a candidate marker with elevated levels in serum compared to healthy controls (Sarkissian et al., 2008). Another study investigated proteins secreted in lung cancer primary cultures and identified a panel of 4 proteins (CD98, fascin, polymeric immunoglobulin receptor/secretory component and 14-3-3ŋ) that demonstrated a higher sensitivity and specificity in plasma compared to single markers (Xiao et al., 2005). All these studies demonstrate the importance of investigating the secretome as a source of candidate biomarkers in cancer research.

Secretome analyses have not just been used for biomarker discovery but have also been used to investigate molecular mechanisms. One study used proteomics to analyse the secretome from two UV-induced fibrosarcoma cell lines to investigate tumour formation mechanisms and identified 28 differentially secreted proteins to suggest possible mechanisms involved (Shi et al., 2007). Secretome analysis has also been used to understand gene function. Khwaja et al used 2-DE to investigate the secretome of p53-null tumour cells in the presence and absence of wt-p53 in order to characterise the role that tumour suppressor p53 plays within the tumour microenvironment (Khwaja et al., 2006).

1.4.4.4. The future of MS-based cancer screening assays

Currently used biomarkers are assayed using immunoassays which target single biomarkers. Similar assays could be optimised for other biomarkers but those that could detect multiple biomarkers simultaneously using MS or antibody array methods could prove to be effective assay for cancer diagnosis. This can be achieved using an electrospray LC-MS/MS multiple reaction monitoring (MRM) assay to detect and quantify tryptic peptides from selected proteins (see section 1.2.3.5 for description). Anderson and Hunter demonstrated that 12 abundant proteins in serum including albumin (40mg/ml) and fibronectin (300µg/ml) and L-selectin (670ng/ml) could be detected in a single assay (L. Anderson and Hunter, 2006).

1.4.5. Biomarker Discovery and commonly used Proteomic Techniques

Once suitable samples have been collected, proteomic approaches to biomarker discovery include: 1) sample preparation which may include sample enrichment strategies such as phosphoproteins, glycosylated proteins or sub-cellular fractionation; and 2) protein separation strategies such as 1D or 2D gel electrophoresis, a gel free approach using 2D HPLC or a combination of these techniques. The advantages and disadvantages of commonly used proteomic techniques used for biomarker discovery are discussed below.

1.4.5.1. Two-dimensional polyacrylamide gel electrophoresis (2D PAGE)

Two-dimensional polyacrylamide gel electrophoresis (2D PAGE) or two-dimensional electrphoresis (2DE) was first described by O'Farrell et al in 1975 and was the first technique used to profile protein expression patterns in cancer (O'Farrell, 1975). It was first used to analyse human plasma proteins by Anderson and Anderson 1977 to separate 300 distinct protein spots upon staining (L. Anderson and Anderson, 1977).

Proteins are separated by charge in the first dimension using isoelectric focussing (IEF). Proteins will have a positive, negative or neutral charge depending upon the pH of the environment they are in. Every protein has a pH at which they carry a neutral overall net charge known as its isoelectric point (pI). Proteins are positively charged at a pH below its pI value and negatively charged at a pH above its pI value. IEF involves placing proteins in a medium with a pH gradient and subjecting the medium to an electric field. The proteins will then start to migrate through the pH gradient towards the electrode of the opposite charge. During the migration, the proteins will gain or lose protons and its net charge and mobility will decrease therefore slowing down the proteins migration. The protein will stop migrating once it reaches the pH which is equal to its pI (the pH at which the protein is uncharged). If the proteins then diffuse either side of this region, they will lose or gain a proton and migrate back to its pI region. In this way proteins are focused to the region of the pH gradient corresponding to their pI value and can remain there until the electric field is removed.

The proteins are then separated in the second dimension based on their molecular weight (MW) by gel electrophoresis. As proteins migrate through the pores in the gel, they become coated in dodecyl sulphate proportional to their MW and migrate with a uniform charge-to-mass ratio. The percentage of the acrylamide in the gel determines the pore size of the gel, the higher the percentage of the gel, the smaller the pore size. Gels are available as single-percentage gels or gradient gels. Single-percentage gels give good resolution of samples within a narrow MW range whereas gradient gels allow the analysis of proteins over a wide MW range in a single experiment. The percentage of acrylamide increases from the top to bottom with proteins migrating from larger pores to smaller ones. Gradient gels are more popular for complex mixtures of proteins as they allow large and small proteins to be resolved simultaneously however they do not match the resolution of single percentage gels.

Traditionally, complex protein samples (such as whole cell lysates, tissues or biological fluids) were solubilised and separated by 2DE and the proteins were stained using a silver staining protocol. The approach is limited as only the most abundant proteins are identified but did prove useful in a number of studies. IEF was originally carried out in polyacrylamide gel rods formed in glass or plastic tubes containing ampholytes that form pH gradient in an electric field. The development of immobilised pH gradients (IPGs) increased the potential utility of 2D PAGE (Bjellqvist et al., 1982). IPGs enabled stable, reproducible pH gradients to be formed capable of focusing acidic and basic proteins on a single gel (Issaq and Veenstra, 2008). The carrier ampholytes are attached to acrylamide molecules and cast into gels to form a fixed pH gradient. Fixing the gradient prevents drift in the gel and allows them to be cast in an efficient and reproducible manner. The pH range is immobilised within an acrylamide matrix and allows the production of gels that cover defined pH ranges allowing proteins with similar pI values to be separated with high resolution (Hoving et al., 2002, Langen et al., 2000).

It has also been well documented that 2DE does have its difficulties and limitations (Rabilloud, 2002, Roessler et al., 2006). These include:

- Dynamic range of proteins in samples can make it difficult for good separation of all proteins.
- Moderate reproducibility.
- Suboptimal resolution of proteins.
- Difficulty of spot matching between different gels. Gels must be run in duplicate or triplicate and imaging software can be used to identify differences although even the most advanced software requires manual intervention to assist the correct matching.

- Spot overlapping (i.e. presence of multiple proteins in single spots on smaller gels). If differences in staining intensity are observed in a spot between two gels, identification of the protein responsible for the difference would be difficult to identify due to the presence of multiple proteins. This could be overcome using larger gels to resolve the proteins further.
- Some proteins are difficult to resolve on a 2D gel such as very basic proteins, small proteins and hydrophobic proteins such as membrane proteins.
- Proteins exhibit a decreased solubility at their pI and therefore can prove difficult to resolve on a 2D gel.

More recently, 2D differential in-gel electrophoresis (2D DIGE) has been used to separate two distinct protein mixtures each labeled with different fluorescent dyes on the same 2D gel to quantify differences in protein expression (Kondo, 2008, Maurya et al., 2007). It was developed to attempt to overcome spot alignment problems (Unlu et al., 1997). It has been used to identify and quantify the differences in protein expression between oesophageal carcinoma cells and normal epithelial cells (Zhou et al., 2002).

Each protein mixture is covalently labeled with a cyanine fluorescent dye (Cy3 and Cy5) with different excitation and emission wavelengths (550nm and 650nm respectively). The samples are mixed in equal amounts of total protein and separated together using a single IPG strip and 2D gel. Identical proteins in both samples should migrate to the same region of the 2D gel at the same speed therefore there is no need for complex data analysis to match the spots. The gel is scanned at the specific excitation wavelengths of each dye using a fluorescent imager. These are merged and the fluorescent intensity is used to determine the relative abundance of all protein spots which can then be excised for identification. 2D DIGE allows the direct comparison of two samples without complicated spot matching between

samples in the same experiment and uses less 2D gels compared to 2D PAGE. However, the fluorescent dyes makes 2D DIGE are more costly than 2D PAGE.

The cyanine dyes first used in 2D DIGE involves labeling lysine residues and the average protein has approximately 30 lysine residues per protein which causes a large mass shift in the proteins which can compromise their solubility. Also, smears in the 2D gel can occur if the labeling efficiency is less than 100%. A new saturation labeling strategy uses the next generation of dyes which label cysteine residues, these occur less frequently than lysine residues resulting in a much smaller mass shift. Compared with lysine residue labeling, 100% labeling of cysteine residues is more easily achieved and is suitable for investigating low abundance proteins or when starting material is limited. DIGE labeling is rapid, sensitive (can detect <12amol using saturation labeling) and allows the detection of post-translational modifications and alternative splicing products but doesn't efficiently resolve large proteins or integral membrane proteins (Minden, 2007). 2D-DIGE has been used to compare cell lysates from oesophageal cancer and normal cells to identify 58 up-regulated spots and 107 down-regulated spots (Zhou et al., 2002). It has also been to used to profile 64 ovarian tissues with the aim of classifying them into four subtypes (Bengtsson et al., 2007).

Despite theses improvements to 2D gel techniques, it still remains a low throughput approach that requires a relatively large amount of protein which is a problem when sample is limited which can be the case with clinical tissue samples.

1.4.5.2. Surface Enhanced Laser Desorption Ionisation (SELDI)

SELDI pre-fractionates samples based on capturing different types of proteins on chips coated with various chromatographic surfaces such as hydrophilic, hydrophobic or ionic surfaces and can also be coated with antibodies or ligands. It uses MALDI technology to measure low molecular mass proteins/peptide fragments (<20kDa) and larger proteins.

SELDI is often used to compare the peptidome or LMW proteome profile resulting from *in vivo* proteolytic cleavage of plasma or tumour derived proteins. The intensities of the polypeptide peaks with differing m/z ratios between two or more populations (e.g healthy vs cancer) can be compared. Petricoin *et al* demonstrated the use of SELDI to compare serum from ovarian cancer patients and healthy individuals (Petricoin et al., 2002). Algorithms were used to compare peak patterns and identified five m/z peaks that could discriminate cancer patients from controls with high sensitivity and specificity. The proteins responsible for these peaks were subsequently identified and found to originate from high-abundance host response serum proteins rather than tumour derived proteins.

There are many SELDI based biomarker discovery studies reporting peak profiles that improve on the sensitivity and specificity of the currently used biomarkers. These potential biomarkers show potential and require validation in larger studies before they can be transferred to the clinic. One such multi-centre study identified five m/z peaks originating from four proteins to be differentially expressed in ovarian cancer. These peaks were shown to improve the detection of early stage ovarian cancer compared to CA-125 alone. It was hypothesized that this test could be used in conjunction with the measurement of CA-125 could detect early stage ovarian cancer (Kozak KR, 2005).

SELDI has been used to study body fluids in many types of cancer such as gastric, hepatocellular carcinoma (HCC), colorectal cancer and pancreatic (Honda et al., 2005, Mohri et al., 2009, Ward et al., 2006a, Ward et al., 2006b, Ward et al., 2008b, Ward et al., 2006c). Studies using SELDI require a lot of thought into the experimental design to eliminate data analysis problems and to ensure no biases are introduced due to blood collection, processing and storage (Diamandis, 2004, Ransohoff, 2005). One limitation of these surveys is the low mass accuracy of the mass spectrometer used making it difficult to positively identify

peptides or proteins responsible for the differential peak patterns. This is essential to investigate the biological relevance of these differences and for these assays to be moved to a clinical setting. Emerging technologies will improve upon peptide profiling and although these studies have not yet lead to the development of clinical assays, SELDI based techniques may still prove to be a useful tool for biomarker discovery.

1.4.5.3. Profiling Arrays

Protein profiling arrays are a sensitive high-throughput method that requires relatively small amounts of sample. Antibody arrays are the most common and have been used to detect proteins in serum. There are two main types: (i) direct labeling and (ii) dual antibody sandwich arrays (S. Hu et al., 2006). The direct labeling method involves labeling proteins in a complex sample with a tag followed by incubation with a corresponding antibody to detect those proteins. A two-colour rolling-circle amplification (RCA) method allows two samples to be compared by labeling with different tags such as biotin and digoxigenin to investigate relative abundances. A sandwich array allows the capture of proteins in a complex sample using a panel of antibodies. Antibody arrays have been used to profile serum proteins in prostate, lung, pancreatic, ovarian and liver cancers (Gao et al., 2005, Miller et al., 2003, Z. Sun et al., 2004). However, protein profiling arrays are not yet as comprehensive as those investigating the transcriptome and require new candidate targets for efficient unbiased profiling (Rifai et al., 2006).

1.4.5.4. Other Methods

More recently, MALDI imaging allows the detection of proteins in situ using tissue sections. The large number of mass spectra is evaluated using hierichial clustering to classify and interpret complex human tissue (Deininger et al., 2008).

Serum and tissue samples from cancer patients can be used in other ways to discover tumour-related peptides and proteins that could be used as markers. As well as tumour secreted proteins, autoantibodies to tumour–related antigens can be found in the circulation. Western blotting techniques utilising 2-DE gels of lung adenocarcinoma cell lines have been used to identify autoantibodies in patients' serum (Brichory et al., 2001). The immunoreactive spots are excised and the proteins are identified by MS. The assay identified annexin I and annexin II autoantibodies to be present in the serum from lung cancer patients however, these results are yet to be extended for clinical evaluation.

A similar approach replaces 2-DE with a combined two-dimensional liquid chromatography separation and protein microarrays. The proteins from prostate cancer cell line LNCaP cell lysates were separated in to 1760 fractions and each of these fractions were then spotted onto microarray slides and incubated with patient serum. The patterns of immunoreactive spots for cancer vs normal were plotted on a heat map and upon comparison were able to distinguish between organ-confined and metastatic prostate cancer (Bouwman et al., 2003).

Section 1.5 : Microvesicles

1.5.1. Introduction

Communication between cells was thought to be mediated only by neurotransmitters and hormones (secreted or cell surface) but electron microscopy showed the presence of small discrete vesicles in the intercellular space (Beaudoin and Grondin, 1991). It has since been shown that these vesicles may play various specific roles (Lakkaraju and Rodriguez-Boulan, 2008). Microvesicle (MV) is a term used to refer to a mixed population of vesicles released by cells whose composition varies depending upon cell type and status or condition.

There are two types of vesicles and processes used to release them from cells: (i) exocytosis of multivesicular bodies (MVBs) with the release of exosomes (<100nm) and (ii) the budding of small vesicles from the plasma membrane referred to as shedding vesicles (100-1000nm).

Exosomes were first identified as vesicular structures segregated within a membrane-bound organelle known as MVB's by electron microscopy and were thought to be involved in late stage maturation of endosomes to lysosomes. The membranes of theses MVB's were later shown to fuse with the plasma membrane of the cell and release its contents into the extracellular space. It was observed that these exosomes were released as a result of specific processes and in response to specific intracellular signals (Fevrier and Raposo, 2004, Lakkaraju and Rodriguez-Boulan, 2008, Schorey and Bhatnagar, 2008).

Media collected from 'healthy' exponentially growing cell lines and analysed by FACS shows the presence of small particles often referred to as cell debris. Electron microscopy has confirmed that this cell debris is enriched in shedding vesicles (Beaudoin and

Grondin, 1991). Shedding vesicles are more heterogenous in shape and larger than exosomes and have been shown to be released from various cell types under different conditions.

1.5.2. MV Secretion from Cells

MV secretion has been shown in a wide range of cell types *in vivo* and *in vitro* such as epithelial, fibroblast, haematopoetic, immune, placental and tumour cells (Chaput et al., 2004, Pap et al., 2008, Valenti et al., 2006, Valenti et al., 2007, VanWijk et al., 2003, Wieckowski and Whiteside, 2006). Cells constitutively release MVs and release increases in rapidly growing cells compared to slow growing cells and in response to a stimulus (e.g. cell activation, hypoxia, irradiation, oxidative injury, exposure to shearing stress and exposure to proteins from an activated complement cascade). Many studies have investigated MV release in specific cell types during a particular process e.g. fibroblasts under mechanical stress (T. L. Lee et al., 1993), platelets and macrophages during coagulation (Sims et al., 1988) and extracellular matrix (ECM) digestion with increased cell motility for tumour and endothelial cells (Ginestra et al., 1998, Gutwein et al., 2003, Taraboletti et al., 2002).

Many names are given to MVs depending on the cell type they have been secreted from including microparticles (MP) from platelets (George et al., 1982), ectosomes from neutrophils and firoblast (Hess et al., 1999), shedding vesicles from tumour cells, exovesicles for neurotransmitter vesicles and MVs released during tissue development from basolateral membrane of cells are known as argosomes (Greco et al., 2001).

MVs are released into the extracellular space and are either broken down within minutes at the site of discharge or others are able to travel some distance via body fluids such as blood. Vesicular structures have also been detected in other body fluids such as urine and cerebrospinal fluids. They are normal constituents of blood plasma and are secreted by leukocytes, endothelium, platelets and erythrocytes. In normal serum, the abundance of MVs

are platelet (80%), endothelial (10%) and leukocyte (10%) derived (Caby et al., 2005, George et al., 1982). The number of MVs in the peripheral blood increases during cell injury, inflammation, thrombosis and platelet activation (Barry and FitzGerald, 1999, Barry et al., 1998, Horstman et al., 2004, Hugel et al., 2005, VanWijk et al., 2003). MVs have also been shown to be released by neoplastic cells and to circulate at higher levels in the peripheral blood of cancer patients (Ginestra et al., 1998, H. K. Kim et al., 2003, Shedden et al., 2003). MVs have the potential to be a good source of diagnostic markers for a variety of conditions.

1.5.3. MV Composition

MVs contain a variety of molecules both enclosed within the vesicle and in the membrane. The size and composition of these vesicles is dependent upon the cell they originated from and their condition (Coleman et al., 2001, Distler et al., 2005). Shedding vesicles have been shown to contain ubiquitous proteins such as β1 integrins which have been identified on the cell surface of many vesicles from various cell types (Del Conde et al., 2005, Dolo et al., 1998). It has been reported that shedding vesicles from tumour cells are enriched with metalloproteinases and other proteolytic enzymes important for digestion of ECM which is important for inflammation and tumour growth (Gasser et al., 2003, Giusti et al., 2008, Gutwein et al., 2003, Millimaggi et al., 2006, Mochizuki and Okada, 2007). MP's are enriched with integrins, the plasma membrane glycoproteins GP1b and GP11b-111a, P-selectin, all important for coagulation (Del Conde et al., 2005, Ginestra et al., 1998, Heijnen et al., 1999). Macrophage derived vesicles are enriched with PSGL1 which is important for binding with platelets (Del Conde et al., 2005). They have also been shown to contain mRNA (Baj-Krzyworzeka et al., 2006, Ratajczak et al., 2006b), infectious particles such as HIV virus or prions (Fevrier et al., 2005, Pelchen-Matthews et al., 2004, Robertson et al., 2006, Rozmyslowicz et al., 2003) and even organelles such as mitochondria (Spees et al., 2006).

The membrane of MVs consists of typical markers (proteins and lipids) of the cell line they were secreted from. In the case of exosomes, all proteins analysed to date originate from the cytosol and plasma membrane. They do not contain nuclear, mitochondrial, golgi or ER proteins suggesting an endocytic origin of these vesicles but they do not contain ATPases or lysosomal proteases indicating that their formation through MVBs deviate from the lysosomal pathway (S. Keller et al., 2006, VanWijk et al., 2003). Exosomes secreted from antigen presenting cells such as B-lymphocytes and dendritic cells have membranes enriched in MHC-I, MHC-II and in co-stimulatory molecules.

Tumour cell MVs are rich in pre-apoptotic molecules such as Fas-ligand (FasL) or TRAIL (TNF related apoptosis inducing ligand) (Andreola et al., 2002, Huber et al., 2005, Valenti et al., 2006). Some proteins are commonly found in MVs from various cell types such as chaperone protein HSP70 and tetraspanin proteins (Aoki et al., 2007, Fritzsching et al., 2002, Hegmans et al., 2004).

MVs have also been found to contain nucleic acids (mRNA, miRNA and DNA) (Unlu et al., 1997, Zhou et al., 2002) and infectious particles such as HIV or prions (Kramer et al., 2005, Ratajczak et al., 2006b). MicroRNAs are known to control gene expression by regulating mRNA turnover and the transfer of RNA between cells has been shown to occur in vesicles from progenitor cells of differentiation and tumour cells (Baj-Krzyworzeka et al., 2006, Deregibus et al., 2007, Hunter et al., 2008).

1.5.4. MV Formation

There are two distinct processes involving vesicle production and release. Exosomes are released via exocytosis when MVBs leave the lysosomal pathway and fuse with the plasma membrane. The sorting process of direct particular proteins into vesicles within the MVB is not very well understood.

Larger vesicles (100-1000nm) are released by reverse budding. Multiple pathways may lead to vesicle formation but the following two appear to be the most relevant.

(1) *The elevation of intracellular Ca^{2+} concentration.*

Vesicle shedding has been shown to increase in resting cells upon elevation of intracellular $Ca^{2+}$with small structures observed protruding from the plasma membrane using time lapse confocal microscopy (Bianco et al., 2005, Pizzirani et al., 2007). Increasing intracellular Ca^{2+} alters phospholipid distribution in the plasma membrane. The phospholipids found on the inner membrane are phosphatidylserine (PS) and phosphatidylethanolamie (PE), the balance of these are maintained by the enzymes scramblase, floppase and translocase. Increasing cytoplasmic Ca^{2+} inhibits translocase and activates scramblase resulting in PS and PE not being returned to the inner side of the membrane. MVs display a large amount of PS on their outer surface of their membrane allowing the binding of AnnexinV. MV formation has been shown to be regulated in a Ca^{2+} dependant manner in monocytes, platelets, red blood cells, endothelial cells, T-cells and in mast cells (Piccin et al., 2007, Redman and Sargent, 2007, Thery et al., 2001).

(2) *The re-organisation of the cytoskelen*

The reorganisation of the cytoskeleton leads to the detachment of the plasma membrane from the cortical actin. This can be initiated by many processes. Phosphatidylinositol 4,5-biphosphate (PIP_2) plays an important role in the attachment of the membrane to the cortical cytoskeleton. In platelets, incubation with PIP_2 inhibited MV formation and decreasing PIP2 increases MV formation (Flaumenhaft, 2006). Calpain and gelsolin are cytosolic proteases which are activated by Ca^{2+} ions. Calpain cleaves talin and activin and gelsoin (specific to platelets) cleaves actin capping proteins. Both of these lead to the disruption of the cortical cytoskeleton allowing budding to occur (Redman and Sargent,

2007, Valenti et al., 2007). During the early stages of apoptosis, MV formation is associated with cortical myosin-actin movement. It starts with GTP-bound Rho proteins activating Rho-associated kinase 1 (ROCK-1) which in turn activates (by phosphorylation) myosin light chain kinase. This allows exchange between myosin and actin leading to detachment of the cytoskeleton from the plasma membrane and MV release (Coleman et al., 2001, Distler et al., 2005).

The formation of shedding vesicles starts with the budding of small cytoplasmic protrusions which then detach by fission from the plasma membrane (Cocucci et al., 2007, Dolo et al., 2000, T. Kobayashi et al., 1984). Accumulation of proteins in the membrane and within the vesicle has been found in these buds using confocal and electron microscopy immunocytochemistry (Cocucci et al., 2007, Schorey and Bhatnagar, 2008). How proteins are sorted in the proximity of the site of vesicle release is still largely unknown but studies with inhibitors of cholesterol synthesis have suggested a role for cholesterol rich micro domains of the plasma membranes known as lipid rafts which has also been suggested to be involved in exosome biogenesis (Del Conde et al., 2005, Pilzer et al., 2005, Stein and Luzio, 1991). The process of vesicle shedding involves a small portion of the plasma membrane being removed from the cell surface. When this occurs at a low constitutive rate, the missing membranes are replaced by the constitutive traffic of membranes rich in vesicle specific molecules (Morris and Homann, 2001). When the rate of vesicles shedding is increased (upon stimulation), the replacement of the membranes lost from the cell membrane occurs by a process known as regulated exocytosis of non-secretory vesicles (which is not yet fully understood) (Chieregatti and Meldolesi, 2005).

Many vesicles do not remain intact in the extracellular space once secreted from the plasma membrane but release their contents immediately. In cancer, this is required for cancer

development. For example, secretion of metalloproteinases along with EMMPRIN by tumour cells in MVs is essential for ECM digestion to increase their mobility and allow them to invade neighbouring tissue and enter the bloodstream (Sidhu et al., 2004). The breakdown of vesicles upon shedding also appears to be important for the release of signalling molecules e.g. IL1-β, caspase1, caspase 3 and cathespin D from dendritic cells and macrophages (Cocucci et al., 2009).

1.5.5. Microvesicle Function

Information transfer by MV secretion plays a role in many physiological and pathological processes e.g. immune modulation, inflammation blood coagulation, cancer and during pregnancy. Mechanisms by which MVs influence the biology of target cells include:

1. Stimulation of other cells by surface expressed ligands acting as signalling complexes
2. Transfer of cell surface receptors. MP's can transfer CD41 antigen from platelets to endothelial cells (VanWijk et al., 2003). MP's are also able to transfer platelet adhesion molecules to tumour cells to increase both the adhesion of lung cancer cells to the endothelium. (Janowska-Wieczorek et al., 2005).
3. The delivery of proteins, lipids and mRNA to target cells. In mouse models, MVs derived from murine embryonic stem cells contribute to epigenetic reprogramming of target cells. In serum free culture, MVs significantly enhanced the survival and improved expansion of murine hematopoietic stem/progenitor cells. This occurred by upregulating the expression of markers for early pluripotent (Oct4, Nanog, and Rex-1) and early hematopoeitic stem

cells (HoxB4, CATA2) and also inducing the phosphorylation of MAPKp42/44 and Ser/Thr kinase AKT. MVs from ES cells also express Wnt-3 protein and are highly enriched in mRNA for several pluripotent transcription factors (Ratajczak et al., 2006a).

4. The transfer of infectious particles. HIV receptors may be transferred between cells rendering those cells susceptible to HIV infection (M. Mack et al., 2000, Rozmyslowicz et al., 2003).
5. Delivery of organelles e.g. mitochondria. Cells defective in aerobic respiration have been shown to receive mitochondria and mitochondrial DNA transferred from other cells via MVs (Spees et al., 2006).

1.5.5.1. Tumour cell-derived MVs and their role in Tumour Progression

MVs are thought to play an important role in tumour development and progession and the number of circulating MVs in peripheral blood has been shown to be increased in cancer patients correlating to increased metastasis and a poor prognosis (Ginestra et al., 1998). The tumour microenvironment is rich in vesicles from proliferating cells, macrophages and neutrophils comprising proteases on the surface of vesicles involved in the digestion of intercellular matrix (matrix and ADAM metalloproteinases and cathespin D) and proteins within the vesicles involved in stimulating migration and metastasis of tumour cells (Ginestra et al., 1998, Giusti et al., 2008, Gutwein et al., 2003, Kim et al., 2003, Millimaggi et al., 2006, Mochizuki and Okada, 2007). Angiogenesis is then stimulated by matrix digestion and tissue factor (TF) generation which is upregulated by the proangiogenic factor vascular endothelial growth factor (VEGF) (Bluff et al., 2008, Janowska-Wieczorek et al., 2005, Taraboletti et al., 2002).

MVs expressing TF on their surface are a strong inducer of VEGF secretion and stimulators of angiogenesis (Osterud, 2003). Tumour-derived MVs stimulate the secretion of several pro-angiogenic factors by stromal fibroblasts, chemoattract and increase the proliferation of endothelial cells to promote angiogenesis (Janowska-Wieczorek et al., 2005). The formation of new blood vessels is crucial for the migration of metastatic cells and malignant cell survival (Bluff et al., 2008). The transfer of mRNA and miRNA via vesicles shed from tumour cells (glioblastoma) to endothelial cells have been shown to have a pro-angiogenic effect (Skog et al., 2008). Shedding vesicles can also contribute to tumour cell growth by inducing Fas-dependent apoptosis of activated lymphocytes (Albanese et al., 1998, Kim et al., 2005).

In a lung cancer model, human and murine lung cancer cell lines secrete MVs in response to non-apoptotic doses of hypoxia, irradiation and chemotherapy (Albanese et al., 1998). They also activated the phosphorylation of MAPKp42/44 and AKT in these stromal cells. They also induced expression of LIF (leukaemia inhibitory factor), IL-11, VEGF and MMP9 in bone marrow and lung derived fibroblasts.

Evidence suggests that MV from activated tumour cells can modulate the function of tumour infiltrating lymphocytes e.g. MVs from melanoma cells may express Fas ligand on their surface which becomes biologically active upon binding to CD95 on T-lymphocytes and then induce apoptosis. This could be a mechanism for cancer cells to elude tumour infiltrating T lymphocytes (Zwaal and Schroit, 1997). Conversely, MVs secreted by tumour cells may have a beneficial effect by increasing the survival of tumour infiltrating monocytes by interacting with them to alter their immunophenotype (transfer of CCR6 and CD44v7/8 to monocytes can prevent apoptosis) (Baj-Krzyworzeka et al., 2006).

The evidence described above highlights that MV secretion is essential in cancer progression therefore they could be a potential source of clinically useful biomarkers in the detection and diagnosis of various cancers.

Section 1.6 : Aims and Objectives

As previously described, early detection can greatly reduce the risk of death due to CRC and patients have a much greater 5-year survival rate if diagnosed at an early stage. Unfortunately, the currently used biomarkers are insufficient for wide scale mass screening to detect the presence of early stage CRC. Previous studies in our group have focused on biomarker discovery in urine and serum from patients with CRC using SELDI and MALDI techniques (Ward et al., 2008a, Ward et al., 2006c) as well as studies to investigate other cancer types such as hepatocellular carcinomas (HCC) and gastric cancers (Mohri et al., 2009, Ward et al., 2006a).

The aim of the work outlined in this thesis is to identify novel candidates for the early detection of CRC using a cell model. Proteomic methods will be used to identify differentially secreted proteins in conditioned media (CM) from two cell lines that will be used as model for early tumour progression. Initially, two-dimensional electrophoresis will be used to analyse proteins secreted by the RG/C2 and PRG/C2 cell lines and this will be followed by the use of quantitative methods such as stable isotope labeling of amino acids in culture (SILAC) to investigate the relative abundance of the secreted proteins identified.

Chapter 2

Materials and Methods

Chapter 2 : Materials and Methods

2.1 Cell Culture

2.1.1. Basic Media

DMEM (41965-039): Purchased from GIBCO Ltd (Invitrogen Corporation). Supplied as a liquid (1x) in 500ml sterile bottles, supplemented with L-glutamine and adjusted to pH 7.4. Media was stored at 4ºC.

2.1.2. Other sterile solutions and supplements

Penicillin/Streptomycin (15140-163): Purchased from GIBCO Ltd. Supplied as a liquid (10,000 units/ml penicillin and 10,000µg/ml streptomycin) and stored at -20ºC. 5ml was added to a 500ml of DMEM.

Fetal Calf Serum (FCS – 1009-148): Purchased from GIBCO Ltd. Supplied as a liquid in 500ml bottles and stored at -20ºC.

Insulin (I0516): Purchased from SIGMA-ALDRICH. Supplied as a 10mg/ml liquid and stored at 4ºC. 380µl was added to 500ml DMEM.

Hydrocortisone (H6909): Purchased from SIGMA-ALDRICH. Supplied as a 50µM liquid and stored at -20ºC in 680µl aliquots. 1 aliquot was defrosted and added to 500ml DMEM.

Trypsin, 0.25% (1X) with EDTA (25200-056): Purchased from GIBCO Ltd. Contains 2.5g/L of Trypsin (1:250) and 0.38g/L of EDTA•4Na in Hanks' Balanced Salt Solution without $CaCl_2$, $MgCl_2 \cdot 6H_2O$, and $MgSO_4 \cdot 7H_2O$, contains phenol red.

10% DMSO / FCS solution: DMSO (D4540 99.5%) was purchased from SIGMA-ALDRICH. 10% DMSO / FCS solution was prepared by adding 2ml of DMSO to 18ml of FCS.

2.1.3. Maintenance of Cell Lines

The media was prepared by adding the following to 500ml of DMEM: 5ml penicillin/streptomycin (final concentration - 100 units/ml penicillin and 100µg/ml streptomycin): 120ml of FCS (final concentration – 20%); 380µl of insulin (final concentration – 7.6µg/ml) and 680µl of hydrocortisone (final concentration – 68nM). This will be referred to as supplemented DMEM. Cells were grown to ~80% confluency in tissue culture coated plastic T75 flasks at 37ºC supplied with 5% CO_2. All solutions were incubated at 37ºC for at least 20mins. Cells were washed with sterile PBS and treated with 2ml trypsin until cells had detached from the flask. 8ml of supplemented DMEM was added to the flask and 5ml of cell suspension was added to fresh flasks along with another 5ml of supplemented DMEM and incubated at 37ºC in 5% CO_2.

2.1.4. Cryopreservation of Cell lines

Cells were grown to ~80% confluency in tissue culture coated plastic T75 flasks at 37ºC in 5% CO_2. All solutions were incubated at 37ºC for at least 20mins. Cells were washed with sterile PBS and treated with 2ml tissue culture grade trypsin until cells had detached from the flask. 8ml of supplemented DMEM was added and then transferred to a 50ml tube. The cells were centrifuged for 5mins at 1,000rpm at 4ºC. The cell pellet was re-suspended in PBS and the centrifugation step was repeated. The cell pellet was then re-suspended in 1ml of 10% DMSO/FCS solution per T75 flask and transferred to 1ml cryovials to be frozen at -20ºC

overnight in a 'Mr Frosty' freezing container (Nalgene) for a freezing the cells at -1ºC/min. Frozen aliquots of cells were then stored in the gas phase of liquid nitrogen at -140ºC.

2.1.5. Recovery of cryopreserved cell lines

All solutions were incubated at 37ºC for at least 20mins. 1ml aliquots of cells were defrosted at 37ºC and then added to 5mls of supplemented DMEM and centrifuged at 1,000rpm for 5mins. The cell pellet was re-suspended in 5mls supplemented DMEM and transferred to a T25 flask and incubated at 37ºC in 5% CO_2 until 80% confluent.

2.1.6. Cell Lines

RG/C2 and PRG/C2: Both cell lines were provided by Professor C Paraskeva from the University of Bristol. The RG/C2 is a non-tumorigenic cell line that was derived from a sporadic colonic adenoma (1-2cm in diameter) isolated from a 59 year old female (Paraskeva et al., 1989). The PRG/C2 cell line was isolated from the RG/C2 cell line following *in vitro* transformation by treatment with gamma radiation and represents a transformed adenoma (unpublished data, C. Paraskeva). Together, these cell lines will be used as a model of early tumour progression in one clonal cell.

2.2 Harvesting MVs from Conditioned Media (CM)

Cells were grown to 80% confluency and then washed three times with serum free (SF) DMEM. The cells are then incubated for 24hrs in 6mls SF DMEM (per T75 flask) at 37 ºC in 5% CO_2. The CM was transferred to a 50ml tube and centrifuged twice at 1,500rpm for 10mins to remove cell debris. The CM was then centrifuged at 150,000g for 1h at 4ºC (rotor and centrifuge pre-chilled at 4ºC, with brake off). The supernatant was collected and labeled remaining conditioned media (REM CM). The barely visible pellet was re-suspended in PBS and a protein concentration was determined using BCA protein assay. The proteins from the

REM CM fraction were extracted by acetone precipitation as follows. Five times volume of acetone (pre-chilled at -20ºC) was added to the REM CM and incubated at -20ºC for 16 hrs. It was then centrifuged at 12,000xg for 20mins at 4ºC. The supernatant was discarded and the pellet was re-suspended in 500µl of water. 2.5mls of acetone was added and then incubated at -20ºC at least 1 hr and centrifuged at 12,000xg for 20mins at 4ºC. The supernatant was discarded and the pellet was re-suspended in PBS for protein concentration determination using BCA protein assay.

2.3 Sucrose Density Gradient Ultracentrifigation

Sucrose Gradient: A sucrose stock solution (2.6M) was prepared; 44.5g of sucrose was added to 50mls of 20mM HEPES and warmed in a beaker of hot water for 10-15mins. This stock solution was used to prepare 8 different concentrations of sucrose in 20mM HEPES as described in Table 2.1.

Concentration (M)	Dilution Factor	Volume of 2.6M sucrose stock solution (ml)	Volume of 20mM HEPES (ml)
2.00	1 : 1.30	7.70	2.30
1.75	1 : 4.80	6.70	3.30
1.50	1 : 1.73	5.80	4.20
1.25	1 : 2.08	4.80	5.20
1.00	1 : 2.60	3.80	6.20
0.75	1 : 3.46	2.90	7.10
0.50	1 : 5.20	1.90	8.10
0.25	1 : 10.40	0.96	9.04
0.00	-	0.00	10.00

Table 2. 1 : Sucrose Density Gradient Ultracentrifugation Fractions

Table to show the composition and concentration of each sucrose fraction utilised in the Sucrose Density Gradient Ultracentrifugation protocol

An MV fraction was prepared form CM from RG/C2 cells as described in Section 2.2 (10 x T75 flasks). The MVs were re-suspended in 4ml of a 2.6M sucrose stock solution. 1ml of each solution was carefully layered on top of the MV fraction starting with the densest (2.60M) and ending with the 20mM HEPES only solution. This was centrifuged at 100,000xg for 15hrs (no brake). Each sucrose layer was collected, diluted in PBS and centrifuged at 150,000xg for 1hr (no brake). The pellet from each fraction was re-suspended in 40µl of 1 x SDS-SB and stored at -20ºC.

2.4 Harvesting Cell Lysates and Subcellular Fractionation

Lysis buffer: 20mM Tris-HCl pH 7.4, 10mM $MgCl_2$, 10mM $CaCl_2$, 1 protease inhibitor cocktail Tablet per 50ml.

Cells were grown to 80% confluency and then washed three times with SF DMEM. The cells are then incubated for 24hrs in 6mls SF DMEM (per T75 flask) at 37 ºC in 5% CO_2. The cells were washed three times with ice-cold PBS and then lysed in 1ml of ice cold hypotonic lysis buffer (per T75 flask) for 20mins at 4°C. The cells were then scraped into a 1.5ml tube and sonicated in a sonicating water bath for 5mins twice (allowing lysate to cool on ice between sonications). The lysates were then centrifuged at 2,000rpm for 5mins at 4°C. The supernatant was transferred to a fresh tube (A) and the pellet was re-suspended in 1ml lysis buffer. The pellet was washed twice and final pellet (nuclear fraction) was stored in lysis buffer at -20ºC. Supernatant (A) collected in previous step was centrifuged at 13,000rpm for 30mins at 4°C. The supernatant was transferred to a fresh tube (B) and the pellet was washed twice. The pellet (heavy membrane fraction) was stored in lysis buffer at -20ºC. Supernatant B (cytosolic/light membrane fraction) was acetone precipitated and stored in lysis buffer at -20ºC. The protein concentration for all fractions was determined by BCA protein assay.

2.5 BCA Protein Assay

BCA protein assay kits: Purchased from Pierce (Thermo Scientific) as two reagents (A and B). Both were stored at room temperature. Reagent A (cat no 23223). Reagent B (cat no 23224).

For 1 x 96-wellplate: 25mls of reagent A was mixed with 0.5mls reagent B to make a working reagent (WR). 25µl of each sample was added to a well of a 96-wellplate. Then add 200µl of WR to each well and incubate at 37ºC for 30 minutes. The plate was allowed to cool to room temperature then the absorbance was measured at 562nm using a UV spectrophotometer.

2.6 Two-Dimensional Electrophoresis (2-DE)

2.6.1. Sample Preparation

Rehydration Buffer: 8M urea, 50mM DTT, 4% CHAPS, 0.2% carrier ampholytes, 0.0002% bromophenol blue. Stored at -20ºC.

2D Quant Kit (80-6483-56): Purchased from GE Healthcare and stored at 4ºC. Components: Precipitant (renders proteins insoluble); Co-precipitant (contains reagents that co-precipitate with proteins and enhances their removal from solution); Wash buffer (used to remove non-protein contaminants from the protein precipitate); Wash additive (contains a reagent that promotes rapid and complete re-suspension of the sample proteins).

2D Clean Up Kit (80-6484-51): Purchased from GE Healthcare and stored at 4ºC. Components: Precipitant (renders proteins insoluble); Co-precipitant (contains reagents that co-precipitate with proteins and enhances their removal from solution); Copper solution: (precipitated protein is re-suspended with this solution); Colour reagent A: (this solution is mixed with colour reagent B to prepare the colour reagent used to measure unbound copper

ion); Colour reagent B (this solution is mixed with colour reagent A to prepare the colour reagent used to measure unbound copper ion); Bovine serum albumin standard solution: (this solution is used to prepare a standard curve).

All samples were collected in rehydration buffer and a protein assay was carried out using the 2D Quant kit as manufacturer's instructions. 500µl of precipitant was added to each sample (3µl) and BSA standards (0, 10, 20, 30, 40, 50µg), mixed and incubated for 2-3mins at room temperature. 500µl of co-precipitant was added to each tube and mixed by inverting tube. The tubes were centrifuged at 10,000 x g for 5mins and the supernatant was removed immediately. The pellet was briefly centrifuged again and the residual supernatant was removed. 100µl of copper solution and 400µl of de-ionised water was added and vortexed to dissolve pellet. 1ml of working colour reagent (prepared by mixing 1 part reagent B with100 parts reagent A) was added and then incubated at room temperature for 15-20mins. The absorbance of each sample was read at 480nm using water as a reference sample and a standard curve was generated using the BSA standard samples to determine concentration of test samples.

Once protein concentration was determined the samples were de-salted using the 2D Clean Up kit as manufacturer's instructions. 100µg (1-100µl) of sample was added to 300µl of precipitant, vortexed and incubated at 4º for 15mins. 300µl of co-precipitant was added and then centrifuged at 12,000 x g for 5mins. The supernatant was carefully removed, the pellet was centrifuged again briefly and the residual supernatant was removed. Without disturbing the pellet, 40µl of co-precipitant was layered on top of pellet and then incubated at 4ºC for 5mins. The pellet was the centrifuged again for 5mins and the supernatant was removed. Without disturbing the pellet, 25µl of de-ionised water was layered on top of pellet and vortexed for 5-10secs to disperse pellet (not dissolve). 1ml of wash buffer (pre-chilled at -

20ºC) followed by 5µl of wash additive was added, vortexed and incubated at -20ºC for 30mins (vortexing every 10mins). The tubes were centrifuged at 12,000 x g for 5mins and the supernatant was removed. The pellet was then air dried for 5mins and resuspended in an appropriate volume of rehydration buffer.

2.6.2. Isoelectric Focusing

Running buffer - 10x Tris/Glycine/SDS Buffer (161-0732): Purchased from BIO-RAD. Diluted 1:10 with water and stored at room temperature.

IPG STRIPS –_Ready Strip IPG strips 11cm pH 3-10NL 12pack (163-2016): Purchased from BIO-RAD and stored at -20ºC.

Equilibration buffer: as rehydration buffer with either DTT or iodoacetamide

2D Gels - Criterion Precast Gel 12.5% Tris-HCl, 1.0mm. IPG+1 well comb, 11cm (345-0102): Purchased from BIO-RAD and stored at 4ºC.

200µl of sample (~200µg) was added into 1 lane of the IEF focusing tray in one continuous line. The protective cover of the IPG strip was removed and laid gel side down over the sample with the '+' over the anode. 1ml of mineral oil was then used to cover the strip and sample to prevent drying out. The Isoelectric focusing was then carried out using the PROTEAN IEF CELL instrument at 20ºC with a maximum current of 50µA/IPG strip for 12hrs. For 11cm strip the following program was used: start voltage 0V, end voltage 8,000V, volt hours 20-35,000 V-hr, Ramp – rapid. The IPG strip was then removed from the IEF focussing tray carefully using forceps and the gel was placed gel side up into a clean, dry disposable rehydration/equilibration tray and incubated at room temperature with equilibration buffer/DTT (100mg of DTT in 10ml rehydration buffer) for 15mins. Remove the equilibration/DTT, overlay strip with equilibration buffer/iodoacetamide (200mg of iodoacetamide in 10ml rehydration buffer) and incubate for 15mins at room temperature. The

strip was then loaded onto a Criterion pre-cast gel for electrophoresis (200V constant, starting current: 90-120mA/gel, final current: 35-55mA/gel for 55mins). Following electrophoresis, the gel was stained with CBB G250.

2.7 Separation of Proteins by SDS-PAGE

SDS-SB: Made up at a 5X concentration (0.3125M Tris-HCl pH 6.8, 10% SDS, 50% glycerol, 0.01% bromophenol blue) and stored in 500µl aliquots at -20ºC. This stock solution was diluted 1:5 and DTT was added immediately prior to use (7.7mg to 1ml) to give a final concentration of 50mM.

MOPS running buffer: 50mM MOPS, 50mM Tris base, 0.1% SDS, 0.1mM EDTA (pH 7.7)

TBST: 150mM NaCl, 100mM Tris pH 7.5, 0.1% TWEEN 20.

NuPAGE Transfer Buffer: (Made fresh each time). 25mM Bicine, 25mM Bis-Tris, 1mM EDTA, 20% methanol.

NuPAGE Novex 4-12%Bis-Tris gels 1.0mm x 10 wells (NP0321BOX): Purchased from Invitrogen Corporation in packs of 10 and stored at 4ºC.

SeeBlue® Plus2 prestained markers (LC5925): Purchased from Invitrogen Corporation. 5µl was used in one well per gel.

5X SDS-SB was added to samples (1 part 5X SDS-SB and 4 parts sample) and heated at 95ºC for 5mins and centrifuged at 13,000rpm for 1min. 25µl of sample and markers were loaded into each well and the gel was run for approximately 50mins at a constant 200V 100-125mA/gel (start); 60-80mA/gel (end).

2.8 Western Blotting

ECL western blotting detection reagents (RPN2106): Purchased from GE Healthcare.

PVDF Immobilon-P 0.45um (IPHV00010): Purchased from Millipore.

The proteins on the gel were transferred to a PVDF membrane (pre-washed with methanol and water) for 1hr at a constant 30V. The PVDF membrane was then blocked in 5% milk powder in TBST for 1hr at room temperature with agitation. The primary antibody was diluted to the desired concentration in 5% milk powder/TBST and the PVDF membrane was incubated in the primary antibody for 16hrs at 4ºC. The membrane was then washed with TBST for 1hr (changing TBST every 10-15mins) and incubated with the secondary antibody (diluted to the desired concentration in 5% milk powder in TBST) for 1hr at room temperature. The wash step with TBST was repeated and the membrane was developed using ECL detection reagents and exposure to X-ray film.

2.8.1. Stripping Blots

Restore™ Western Blot Stripping Buffer (21059): Purchased from Pierce ready to use and stored at room temperature.

The blot was first washed in TBST as previously described and then incubated in 25mls *Restore™ Western Blot Stripping Buffer* for 10-15mins at room temperature. The TBST wash step was repeated and blot was ready to block and re-probe with primary antibody.

2.8.2. Antibodies

(a)

Antibody	Supplier	Working Dilution	Immunogen
Mouse monoclonal [AC-15] to beta Actin (ab6276)	AbCam	1:40,000	DDDIAALVIDNGSGL (aa 1-14 of *X.Laevis)*
AGR2 monoclonal antibody (M03), clone 1C3 (H00010551-M03)	Abnova (Tebu-Bio)	1µg/ml (1:1000)	Raised against a full length recombinant AGR2
Rabbit polyclonal to Annexin A1 (ab33061)	AbCam	1.25µg/ml (1:1600)	synthetic peptide derived from aa 12-61 of human Annexin A1
Mouse monoclonal [Col-1] CEA (ab17254)	AbCam	1:200	human CEA
EMMPRIN (N-19) (sc-9752)	Santa Cruz	1:1000	N-terminus human EMMPRIN
Rabbit monoclonal [E144] to EpCAM (TACSTD1) - (ab32392)	AbCam	1:2,500	C-terminus of human EpCAM
Rabbit polyclonal to Hsc70 (ab1427)	AbCam	1µg/ml (1:500)	not specified
Rabbit Anti-HSP90α (Hsp86) Polyclonal antibody (SPS-771)	Stressgen Bioreagents	1:2000	synthetic peptide PEETQTQDQPM(C) from 2-12 of mouse Hsp86
Mouse Anti-Hsp90ß Monoclonal antibody (SPA-842)	Stressgen Bioreagents	1µg/ml (1:950)	Immunogen is purified recombinant human Hsp90ß
Anti-OLFM4 (Olfactomedin 4) polyclonal antibody (60B898)	Antagene	1µg/ml	N-terminal human OLFM4
Anti- Human Selenium Binding Protein (M061 3)	MBL	1µg/ml (1:1000)	hybridoma - fusion of mouse myeloma cell SP2/0 with Balb/c mouse splenocyte immunised with GST-full length human SBP protein.

.

(b)

Antibody	Supplier	Working Dilution
Goat anti-mouse IgG-HRP (sc-2005)	Santa Cruz	1:5000
Goat anti-rabbit IgG-HRP (sc-2004)	Santa Cruz	1:5000
Goat anti-mouse IgM-HRP (sc-2064)	Santa Cruz	1:5000
Donkey anti-goat IgG-HRP (sc-2020)	Santa Cruz	1:5000

Table 2. 2 : Antibodies.

(a) Details of all primary antibodies used throughout this study. (b) Details of all secondary antibodies used throughout this study.

2.8.3. Densitometry

Western blots were analysed for differences in protein expression using the GS-800 calibrated densitometer and the Quantity One software.

2.9 Staining of Protein Gels

2.9.1. Coomassie Staining

Coomassie Brilliant Blue (CBB) G250 Stain: 0.08% CBB G250, 1.6% orthophosphoric acid, 8% ammonium sulphate, 20% methanol. The CBB G250 was purchased from Sigma and stored at room temperature. Staining solution was made up in the following order: 8ml orthophosphoric acid in 384ml distilled water and 40g ammonium sulphate was added. 0.5g CBB G250 was dissolved in 10ml distilled water and 8mls was added to previous solution. Then 100ml methanol was added whilst stirring (final volume 500ml).

Destain Solution: 1% acetic acid in distilled water.

The coomassie stain was prepared as described above. Protein gels were stained at room temperature with coomassie for 16hrs and destained for 6-8hrs (changing the destain solution every 30-60mins) or until the background looked clear (Shevchenko et al., 1996). Gels were stored at 4ºC.

2.9.2. Silver Staining

All solutions were made fresh each time and stored at room temperature.

Fixing Solution: methanol : acetic acid : distilled water (45 : 5 : 45).

Sensitising Solution: 0.02% sodium thiosulphate (0.1g $Na_2S_2O_3$ in 500ml distilled water).

Silver Stain: 0.1% silver nitrate (0.5g $AgNO_3$ in 500ml distilled water -store in the dark).

Developing Solution: 0.04% formaldehyde, 2% sodium carbonate (10g Na_2CO_3, 540µl formalin 37% in 500ml distilled water).

Stop Solution: 1% acetic acid in distilled water.

All steps were carried out at room temperature on a shaker. The gel was fixed in the fixing solution for 30mins and washed in distilled water for 20-60mins to remove acid (change water 2-3 times). The gel was then incubated in the sensitising solution and washed twice with distilled water for 1min each. The gel was then incubated with the silver stain for 20-40mins and washed twice with distilled water for 1min each. The gel was incubated in the developing solution until the gel had developed enough (0.5-4mins) and the stop solution was added (Shevchenko et al., 1996). The gels were stored at 4ºC.

2.10 In-Gel Digestion

Acetonitrile (SO-2856-B025): Purchased from LGC Promochem and stored at room temperature.

Water for HPLC (23 595 328): Purchased from VWR BDH PROLABO and stored at room temperature.

Formic Acid (F/1850/PB08): Purchased from Fisher Scientific and stored at room temperature.

Sequencing Grade Modified Trypsin (V5111): Purchased from Promega and stored at -20ºC.

DTT (Diothiothreitol ultrapure- 15397): Purchased from USB Corporation and stored at -20ºC.

Iodoacetamide (I6125): Purchased from Sigma and stored at 4ºC.

All tubes were rinsed with 0.1% formic acid/50% acetonitrile before use (unless sterile tubes from an unopened bag were used). The gel was placed on a clean glass plate and bands were excised from the gel (smallest size possible) and placed in the clean tubes.

The following step was for silver stained gels only. The bands were washed with 300µl of water for 15mins to remove the acetic acid. The supernatant was removed and 200µl of 15mM potassium ferricyanide/50mM sodium thiosulphate was added and incubated at room temperature on a shaker until the silver was removed (>1hr). The supernatant was removed and the bands were washed with 300µl of water for 15mins.

The in-gel digestion protocol is the same for both silver and coomassie stained gels following destain. The bands were washed twice with 200µl of 50% acetonitrile/50mM ammonium bicarbonate for 45mins each at 37°C. The supernatant was removed and the bands were dried using a vacuum evaporator for 15mins. 50µl of 50mM DTT (made up in 10%

acetonitrile/50mM ammonium bicarbonate) and incubated at 56°C for 1hr to reduce the disulphide bonds involved in protein folding. The supernatant was removed and 50µl of 100mM iodoacetamide (made up in 10% acetonitrile/50mM ammonium bicarbonate) was added to amidomethylate the cysteine residues to prevent re-folding of the protein. The bands were then incubated at room temperature in the dark for 30mins. The supernatant was removed and the bands were washed three times with 10% acetonitrile/50mM ammonium bicarbonate for 15mins each at room temperature with agitation. The bands were dried using a vacuum evaporator for 2-3 hrs. The bands were then rehydrated in 20µl of 12.5µg/ml sequencing grade modified trypsin in 10% acetonitrile/50mM ammonium bicarbonate for 1hr at room temperature. 30µl of 10% acetonitrile/50mM ammonium bicarbonate was added and the bands were incubated at 37°C for 16hrs. The supernatant was collected and 30µl of 3% formic acid in water was added and incubated at 37°C for 1hr. The supernatant was collected and added to previous supernatant. This step was repeated for 30min and supernatant was collected (a final volume of 70-90µl). The sample was stored at -20ºC until ready for mass spectrometry.

2.11 Mass Spectrometry

Figure 2.1 shows a schematic of the HPLC and LCQ Deca XP plus setup.

BioBasic *C18 Reverse Phase Column 150x0.18 mm – particle size 5µm (72105-150265*): Purchased from Thermo Electron Corporation.

µ-Precolumn™ cartridge (C18, 5µm, 100Å, 30µm ID x 5mm) P/N 16045: Purchased from LC Packings (Dionex)

Acetonitrile (SO-2856-B025): Purchased from LGC Promochem and stored at room temperature.

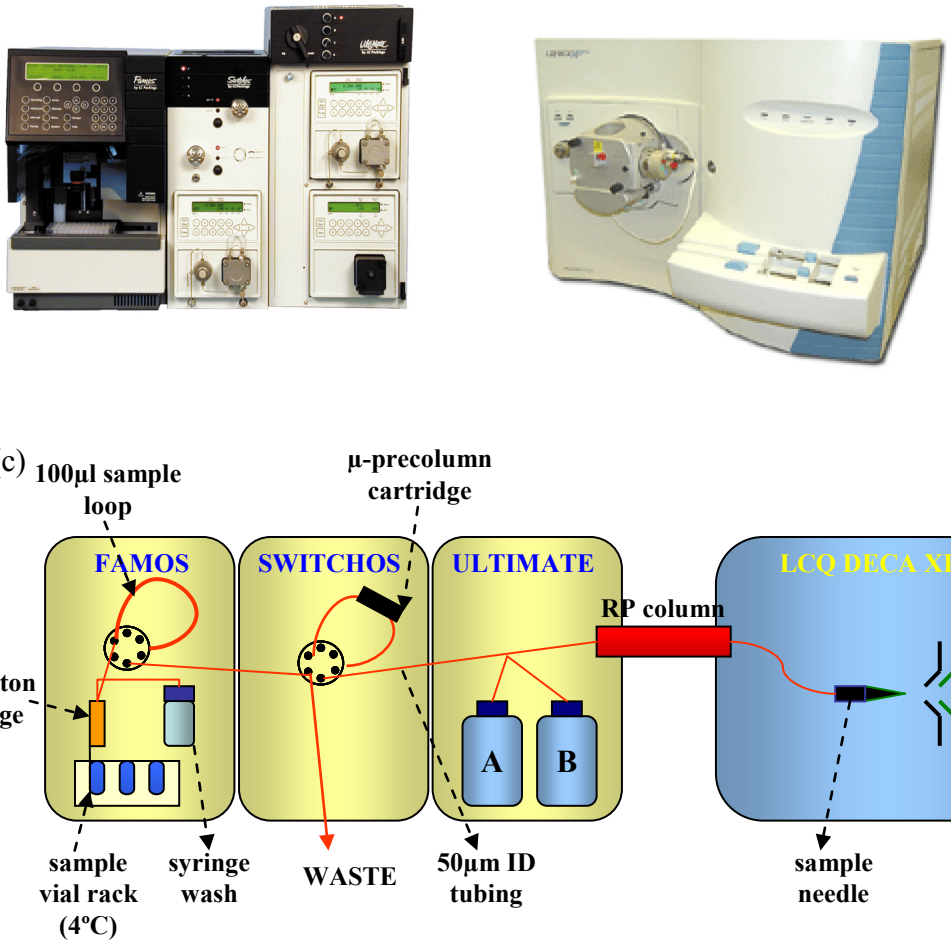


Figure 2. 1 : LC MS/MS set up used in this study.

(a) Photo of the LC Packings (Dionex) Ultimate HPLC system. (b) Photo of the LCQ Deca XP ion-trap mass spectrometer from Thermo Finnigan. (c) Schematic diagram of the setup of the HPLC coupled to the LC MS/MS. Samples (up to 100µl) are drawn from the sample vials by syringe into the sample loop. The sample is then passed through a C18 µ-precolumn cartridge at a flow rate of 30µl/min before then being loaded onto the reverse phase (RP) column. The C18 µ-precolumn cartridge cleans the sample and protects RP column from blockages. The peptides are then eluted from the RP column with an increasing acetonitrile gradient at a flow rate of 1µl/min and sprayed into the ion tube where they are analysed by the mass spectrometer.

Water for HPLC (23 595 328): Purchased from VWR BDH PROLABO and stored at room temperature.

Formic Acid (F/1850/PB08): Purchased from Fisher Scientific and stored at room temperature.

Solution A: 5% acetonitrile/0.1% formic cid in HPLC grade water.

Solution B: 95% acetonitrile/0.1% formic acid in HPLC grade water.

2.11.1. Calibration

*Caffeine (C-6035)***:** Purchased from SIGMA-ALDRICH. 1mg/ml solution of caffeine in 100% methanol was prepared and stored at 4ºC.

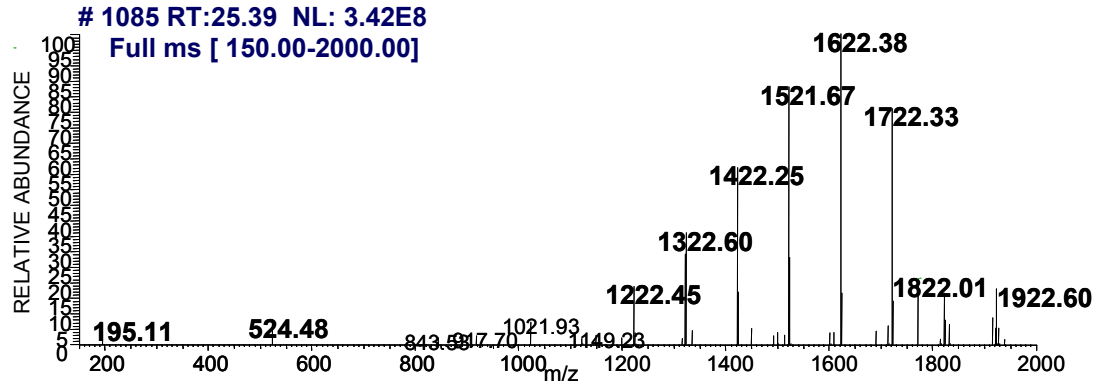
MRFA: Supplied in the accessory kit from manufacturers (Thermo Electron Corporation) and has an average molecular weight of 523.6u. 1ml stock solution of 5nmol/µl in 50:50 methanol/water was prepared by dissolving 3mg of the MRFA in 1ml 50:50 methanol:water and mixing thoroughly. Stored at 4ºC.

ULTRAMARK 1621 (L16698): Purchased from Alfa Aesar (distributor was Lancaster). A 10ml stock of 0.1% ultramark was prepared by mixing 10µl with 10ml of acetonitrile using a syringe.

*ESI CALIBRATION SOLUTION***:** 5ml of the calibration solution was prepared by adding 100µl of the caffeine stock solution, 2.5ml of the stock solution of ultramark 1621 and 50µl of glacial acetic acid to 2.34ml of 50:50 methanol:water. Solution was mixed thoroughly and stored at 4ºC.

The calibration solution was sprayed directly into the mass spectrometer at 2µl/min and the spectra were observed for the presence of the caffeine, MRFA and ultramark peaks and then the mass spectrometer was calibrated. Figure 2.2a shows an example of the typical spectra obtained during calibration.

(a)



(b)

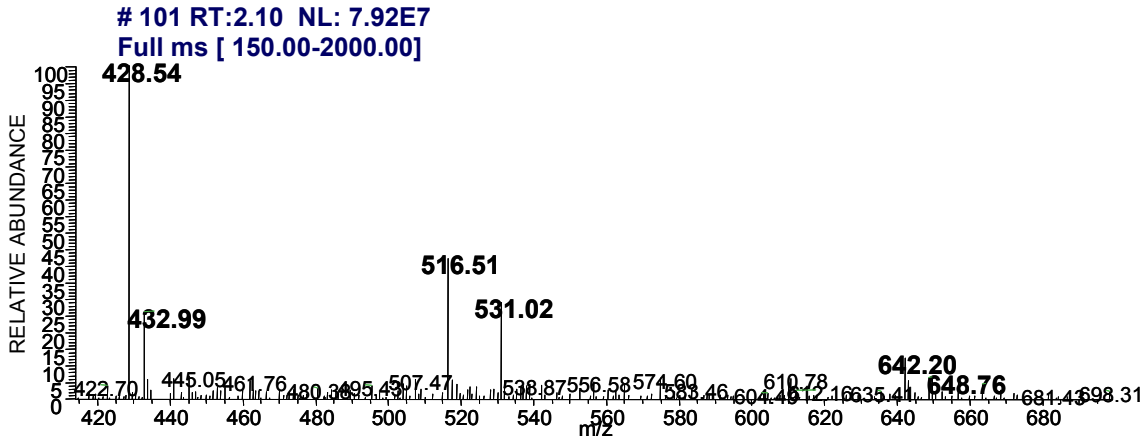


Figure 2. 2 : Calibration and Tuning Spectra.

(a) Spectra of the peaks (in bold) found during calibration: Caffeine: 195.11; MRFA: 524.48; Ultramark 1621: 122.45, 1322.60, 1422.25, 1521.67, 1622.38, 1722.23, 1822.01, 1922.60. (b) Spectra of the peaks found during tuning (in bold). The doubly and triply charged species of the peptides are more commonly seen: [Val^{5}] Angiotensin I: 642.20 (+2), 428.54 (+3); [Asn^{1}, Val^{5}] Angiotensin II: 516.51 (+2); Bradykinin: 531.02 (+2); Angiotensin: 648.2 (+2), 433.11 (+3).

2.11.2. Tuning using Peptide Standards

[Val^5] Angiotensin I (A-9402): Purchased from SIGMA-ALDRICH. Mass of 1282.5 (+1), 642.2 (+2), 428.5 (+3).

[Asn^1, Val^5] Angiotensin II (A-6402): Purchased from SIGMA-ALDRICH. Mass of 1031.2 (+1), 516.5 (+2), 344.7 (+3).

Bradykinin (B-3259): Purchased from SIGMA-ALDRICH. Mass of 1060.2 (+1), 531.1 (+2), 354.3 (+3).

Angiotensin (A-9650): Purhased from SIGMA-ALDRICH. Mass of 1296.5 (+1), 648.2 (+2), 433.11 (+3).

1mg of each of the above peptide standards were reconstituted in 1ml of solution A (5% acetonitrile, 0.1% formic acid). The stock solution of tuning standards was made by adding 10µl of each peptide to 960µl of solution A. This was diluted 1:10 to make the working stock for tuning. This solution was sprayed directly into the mass spectrometer at 2µl/min and the spectra were observed for the presence of the above peaks. The most abundant peak was chosen for tuning i.e. 428.54 which corresponds to triply charged species of Angiotensin I (see Figure 2.2b).

2.11.3. Testing sensitivity using a BSA standard

A 1mg/ml (15.15nmoles/ml) stock of BSA (in 25mM ammonium bicarbonate) was prepared. 50µl of 100mM DTT was added to 1ml of the BSA stock and then incubated at 56ºC for 1hr. 50µl of 200mM iodoacetamide was added and then incubated at room temperature for 30mins in the dark. Another 50µl of 100mM DTT was added to inactivate the excess iodoacetamide (interferes with trypsin digestion of proteins) followed by 10µg of sequencing grade modified trypsin (in 25mM ammonium bicarbonate) was added. The BSA

was incubated at 37ºC for 12-16hrs. The digested BSA was diluted to in solvent A to 312.5fmoles/ml.

80µl of the diluted digested BSA (25fmoles) was then analysed by LC MS/MS using the acetonitrile gradient outlined in Figure 2.3. The raw file was searched against a bovine database, created from an nr database downloaded from the NCBI website (ftp://ftp.ncbi.nih.gov/blast/db/FASTA/). If the LC MS/MS was working optimally, the results would show >25% coverage of BSA when the Xcorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50) filter was used.

2.12 SILAC

2.12.1. Preparing Media

2.12.1.1. For the C13 Arg SILAC experiments

L-Arginine:HCL (U-13C6) CLM-2265-01: Purchased from Cambridge Isotope Laboratories, Inc. Supplied as 100mg of lyophilised powder. Stored at room temperature.

L-Lysine Biotechnology Performance certified (L9037-25G): Purchased from SIGMA-ALDRICH. Stored at room temperature.

L-Arginine, reagent grade (A-5006): Purchased from SIGMA-ALDRICH. Stored at room temperature.

DMEM with 4.5g/L D-glucose and 25mM HEPES. Without L-glutamine, L-arginine and L-lysine (KI-05060 cat no 06-1055-52-1A): Custom made by Geneflow supplied in 10 x 500ml bottles. Stored at 4ºC.

Fetal Bovine Serum dialysed (F-0392): Purchased from SIGMA-ALDRICH and supplied in 500ml bottles. Stored at 4ºC.

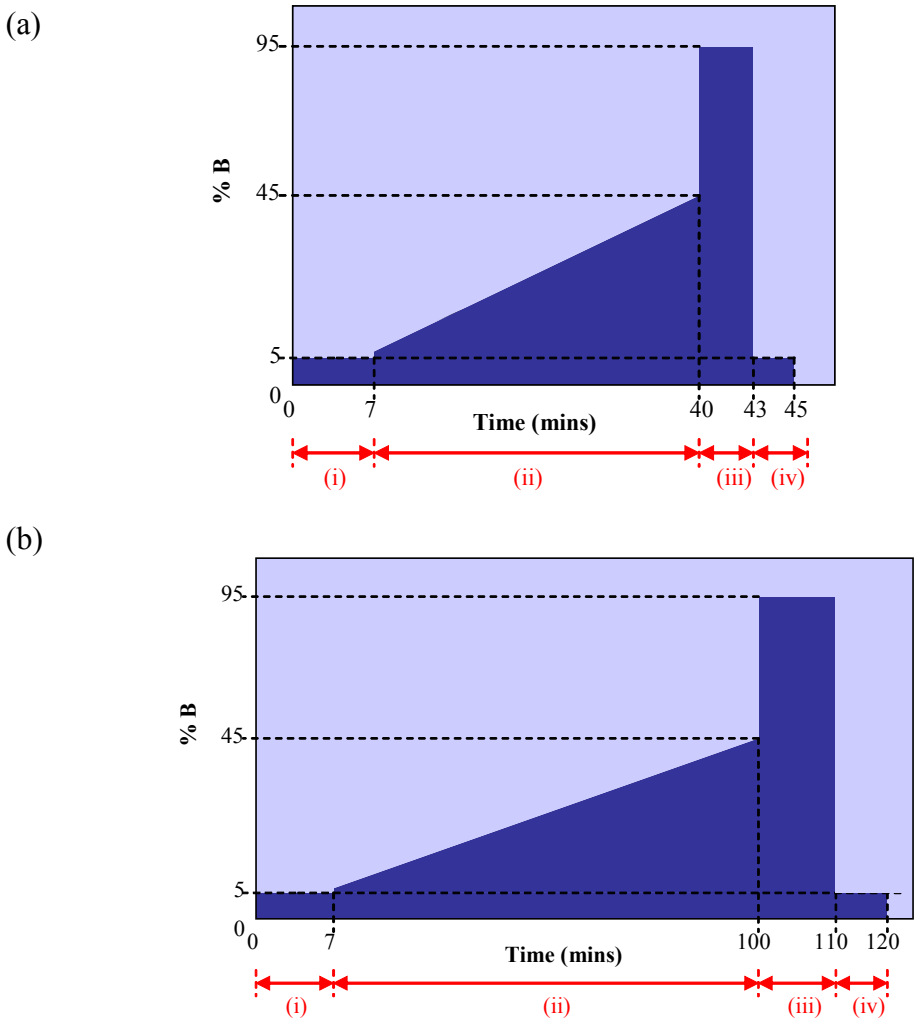


Figure 2. 3 : Acetonitrile gradients used in this Study.

The sample is loaded onto the RP column (i) and at 7mins, the acetonitrile gradient increases from 5% to 45% solvent B over 40 or 93mins (ii). The gradient is then sharply increased to 95% to clean the column for 10mins (iii) at which time it returns to 5% solvent B for 10mins to equilibrate column before the next sample is loaded (iv). (a) Diagram to show the increasing acetonitrile gradient used when analysing a BSA standard. (b) Diagram to show the increasing acetonitrile gradient used when analysing SILAC samples.

The media was prepared as follows:

For the 'heavy' SILAC media, 42mg of L-arginine:HCL, 73mg of L-lysine and 290g of L-glutamine were all added to separate tubes of 5ml of DMEM and mixed thoroughly. All 3 tubes were added to a 500ml bottle of DMEM along with 5ml of penicillin/streptomycin, 380µl insulin and 680µl of hydrocortisone. This was then divided into 12 x 40ml aliquots and stored at 4ºC. The final concentrations of amino acids were: 84mg/L L-arginine, 146mg/L L-lysine and 580mg/L glutamine. Immediately before use, 10ml of dialysed FBS was added and the media was filter sterilised using a 0.45µm syringe filter. The media for the 'light' SILAC was prepared as described above, replacing the L-arginine:HCl with L-arginine.

2.12.1.2. For the C13 Lys SILAC experiments

SILAC™ Protein Identification (ID) and Quantitation kit (SM1002) – Purchased from Invitrogen. Contents include: DMEM (2 x 1000ml); fetal bovine serum, dialysed (2 x 100ml); L-glutamine (100x) liquid (20ml); L-lysine HCl (100mg); L-arginine (2 x 100mg) and [U-$^{13}C_6$]-L-lysine (*Lys) (100mg). All components were stored at 4ºC except for glutamine and FBS which was stored at -20ºC. Additional dialysed FBS (100ml) was purchased from Invitrogen (26400-036).

The media for the C13 Lys SILAC experiments were prepared as per manufacturers (Invitrogen) instructions using the reagents supplied with the kit. 100mg of L-lysine HCl, [U-$^{13}C_6$]-L-lysine (*Lys), 100mg L-arginine (x2) were added to separate 1ml tubes of unsupplemented DMEM and mixed thoroughly. 200ml of DMEM was removed from each of the 2 x 1000ml bottles and stored at 4ºC. For the 'light' SILAC media, add the 1ml aliquot L-lysine HCl and one of the 1ml aliquots of L-arginine to 1 bottle of DMEM. For the 'heavy' SILAC' media, add the 1ml aliquot of [U-$^{13}C_6$]-L-lysine and the other 1ml aliquot of L-arginine. To each bottle add 10ml of 100x L-glutamine and 10ml of penicillin/streptomycin.

Each bottle was divided into 40 ml aliquots and stored at 4ºC. Immediately before use, 10ml of dialysed FBS (for a final concentration of 20%) was added and the media was filter sterilised using a 0.45µm syringe filter.

2.12.2. Confirming labelling efficiency

Each cell line was grown in either 'heavy' or 'light' SILAC media for 6 doublings and the serum free CM (using serum free media containing the appropriate 'heavy' and 'light' amino acids) was collected to confirm the labelling efficiency. The CM was acetone precipitated as previously described and 20ug was loaded onto a 4-12% Bis-Tris gel. Five bands were excised from each lane, in-gel digestion was performed and the samples were analysed by LC MS/MS. The raw data was searched against a human IPI database and the percentage of 'heavy' and 'light' Arg or Lys was calculated.

2.12.3. LC-MS/MS Analysis

The RG/C2 cell line was grown in media containing [$^{12}C_6$]-L-arginine and [$^{12}C_6$]-L-lysine (i.e. light Arg and Lys) and the PRG/C2 cell line was grown in media containing either [$^{13}C_6$]-L-arginine and [$^{12}C_6$]-L-lysine or [$^{12}C_6$]-L-arginine and [$^{13}C_6$]-L-lysine (heavy Arg or heavy Lys) as described in Section 2.12.1. MVs were prepared from the CM as described in section 2.2. The MV fractions from the RG/C2 and either the heavy Arg PRG/C2 or heavy Lys PRG/C2 were mixed in a 1:1 ratio (normalised by cell number).

The proteins in the MV sample were then separated by 1D electrophoresis on a 4-12% Bis-Tris gel. The whole lane was then cut into 40 equal slices and in-gel digestion was performed as previously described. The samples were then analysed by LC MS/MS using the acetonitrile gradient outlined in Figure 2.3b. The raw files were searched against a human IPI

database and the statistical analysis and relative abundance ratios were calculated using the TransProteomic Pipeline (TPP) (Section 2.13).

2.13 Data Analysis

2.13.1. Bioworks Browser v3.1

Bioworks Browser v3.1 was supplied with LCQ Deca XP plus from Thermo Finnigan and uses SEQUEST browser. The software uses the raw files from the LC MS/MS analysis and generates dta files for each spectra which are in turn searched against a database of known proteins. The peaks detected in the full scans and their accompanying MS^2 scans are used by the software to determine a peptide sequence match followed by a protein match for the peptides found.

Dta generation parameters: Tolerance and limits for Dta generation –

precursor mass - 1.40

group scan - 20 (The number of intermediate MS/MS scans between two MS/MS scans that have the same precursor *m/z* value).

minimum group count - 1 (The minimum number of MS/MS scans with the same precursor *m/z* value).

minimum ion count – 15 (The minimum number of ions that must be present in the MS/MS spectrum of a peptide before the spectrum is included in the results of a search.

Dta search parameters: Tolerance and limits for Dta search –

Peptide - 1.50

fragment ions - 0

number results score - 250

ion series calculated in Dta search A, B and Y ions

Modifications:

Carboxyamidomethylation on Cys residues 57.02

For SILAC experiments only: 6.00 on Lys or Arg residues

Databases:

IPI databases were downloaded from http://www.ebi.ac.uk/IPI/IPIhelp.html.

NCBI human databases were generated from the non-redundant (nr) database downloaded from ftp://ftp.ncbi.nih.gov/blast/db/FASTA/ .

The search results were exported to excel for further analysis and reporting. When using the IPI database, the protein names associated with accession numbers were not always shown in the exported excel spreadsheet. Single protein names could be found on the IPI website directly. This was also carried out in batches (up to 200 proteins at a time) using the DBfetch utility http://www.ebi.ac.uk/cgi-bin/dbfetch . A list of IPI accession numbers was generated in notepad and sorted in alphabetical order. A list of the associated protein names was retrieved in fasta format by uploading the notepad list of accession numbers on the DBfetch website.

2.13.2. Trans Proteomic Pipeline (TPP)

The TPP software is freely available online from the Institute for Systems Biology website (http://tools.proteomecenter.org/windows.php) (A. Keller et al., 2005, A. Keller et al., 2002a, X. J. Li et al., 2003, Nesvizhskii et al., 2003). Figure 2.4 outlines the basic steps in using TPP for the analysis of large proteomic datasets. The software uses Cygwin, a Linux-like environment for Windows. The MS/MS spectra were searched against a database of known proteins using Bioworks Browser (Section 2.13.1). The first step in the TPP analysis

involved converting the output files from SEQUEST to html and mzXML files as follows: (type in **bold** are the commands typed into the Cygwin terminal)

The raw file was placed in the same folder as the dta and out files and these folders were placed in the wwwroot folder. When you right click on your folder, the command prompt box will open.

out2summary *.out>my file.html .This converted the out files to 1 html file.

(myfile refers to name of results folder)

t2x.exe rawfilename.raw c . This converted the raw file to mzXML.

(rawfilename refers to the name of the raw file)

Sequest2XML myfile.html –Psequest.params . This converted the html file to xml file.

The 3 files generated in steps 2-4 were moved to one folder for all 40 samples in one SILAC experiment for combined analysis using PEPTIDE PROPHET™ and PROTEIN PROPHET™.

PEPTIDE PROPHET™

For C13 Arg SILAC: xinteract –X-nR,6.0 –A-lR *.xml

For C13 Lys SILAC: xinteract –X-nK,6.0 –A-lK *.xml

PROTEIN PROPHET™

For C13 Arg and C13 Lys SILAC: **runprophet –OXA interact.xml**

A web address is given in the Cygwin terminal to access the results.

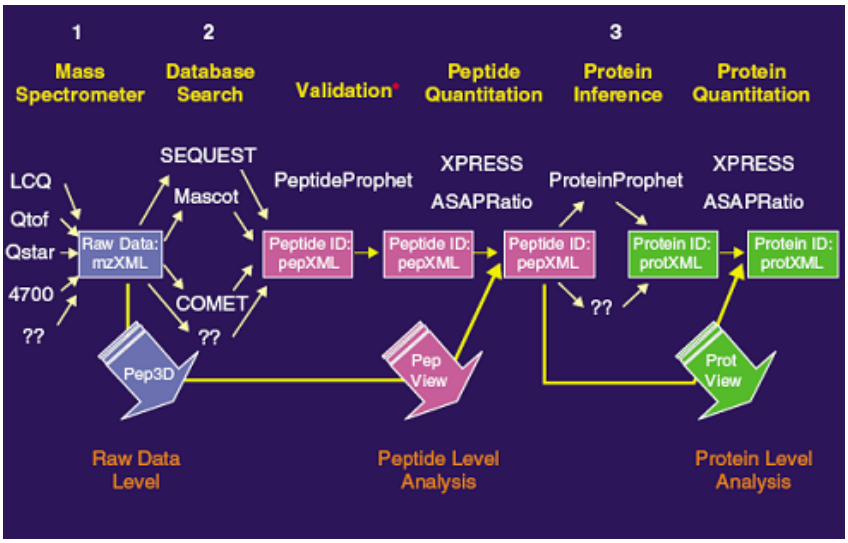


Figure 2. 4 : The TransProteomic Pipeline (TPP).

The above diagram outlines the basic steps in analysing large proteomic datasets using the TPP software. Stage 1 involves converting output files (both dta and out files) to an XML format. Stage 2 involves validating the peptide sequence assigned to each acquired spectra using Peptide Prophet (includes quantification using Xpress or ASAPRatio) and calculates a probability score that the correct assignment has been made. Stage 3 involves validating the protein assignment to the peptide sequences. This stage also involves calculating a probability that the correct protein assignment has been made and also calculates an Xpress or ASAPRatio for that protein based upon those calculated for each individual peptide assigned to that protein in Peptide Prophet. Figure taken from (A. Keller et al., 2005).

2.13.3. FatiGO+

Before looking at gene ontology and functional annotation of proteins, the IPI accession numbers were converted to entrez gene ID symbols. This was carried out using an online resource known as REFDIC http://refdic.rcai.riken.jp/welcome.cgi (Hijikata et al., 2007). A list of IPI accession numbers was generated in notepad. On the REFDIC website, choose tools then X-REF converter in toolbar. Options chosen:

Species – homosapiens

Type of input – IPI accession

Type of output – Entrez gene ID

Then, the list generated in notepad was uploaded then submit. This list of gene symbols was then used to look at functional annotation and gene ontology.

The GO terms for these proteins were then generated using an online resource known as FatiGO+ http://fatigo.bioinfo.cnio.es/bioinfo/ (Al-Shahrour et al., 2007, Al-Shahrour et al., 2006). A list of gene symbols was generated in notepad and uploaded on the FatiGO website. Then all the functions required were highlighted (i.e. GO terms biological processes, cellular component and molecular function; KEGG pathways and BIOCARTA pathways) and submit.

Chapter 3

RESULTS

Investigating the secretome from RG/C2 and PRG/C2 cells to identify possible biomarkers for the early detection of colorectal cancer using proteomic techniques

Chapter 3 : Investigating the secretome from RG/C2 and PRG/C2 cells to identify potential biomarkers for the early detection of colorectal cancer using proteomic techniques.

3.1 Introduction

Proteomic techniques have been used to identify potential candidate cancer biomarkers from tumour tissues, serum, plasma and urine in a range of cancer types (W. C. Cho, 2007, Kozak et al., 2003, Ward et al., 2008a, Ward et al., 2006c). An increasing number of proteomic studies are now utilising cell-based models to identify secreted proteins as a source of potential biomarkers under the hypothesis that proteins released by cancer cell lines could escape into the bloodstream (Gronborg et al., 2006, Kulasingam and Diamandis, 2007, Looi et al., 2008, Mbeunkui et al., 2006, Mbeunkui et al., 2007, C. C. Wu et al., 2008, Zhong et al., 2008). The aim of this study was to identify differentially secreted proteins from the colorectal cancer cell lines RG/C2 and PRG/C2 in an attempt to identify potential markers of early tumour progression using proteomic techniques (see section 2.1.6 for description of cell lines).

3.2 Analysis of Conditioned Media using 2D Gel Electrophoresis

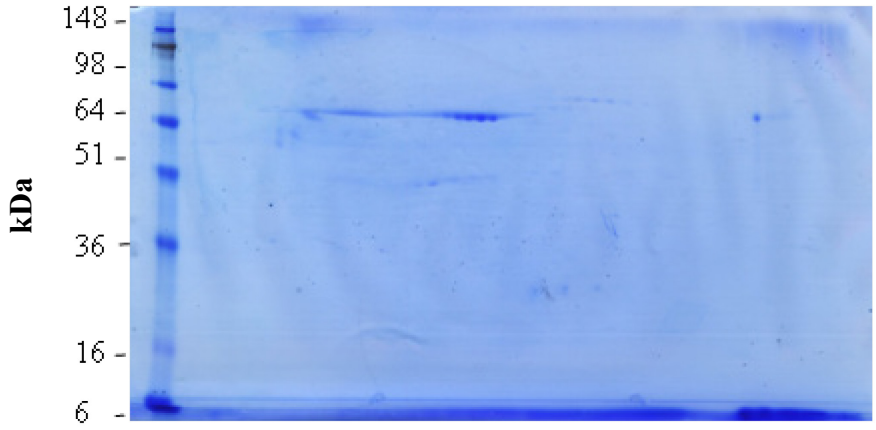
The starting point for this study was to use two-dimensional gel electrophoresis (2-DE) to investigate differentially secreted proteins from colorectal cancer cell lines. 2-DE is a proteomic technique commonly used to identify differences between two samples (e.g. cancer vs normal) and has been used in a variety of studies using cell lines (Al-Ghoul et al., 2008, Cecconi et al., 2007, Looi et al., 2008, Rondepierre et al., 2009), conditioned media (CM) (Kulasingam and Diamandis, 2007, Mbeunkui et al., 2006, Mbeunkui et al., 2007, Roessler et al., 2006, Sardana et al., 2008, Sato et al., 2001, C. C. Wu et al., 2008), tumour tissue

(Friedman et al., 2004, Iwadate, 2008), serum ((no-authors-listed), 1996, Ahmed et al., 2004, He et al., 2008, Looi et al., 2008) and plasma (S. Y. Cho et al., 2008, Darde et al., 2007). CM was collected from both the RG/C2 and PRG/C2 cell lines and the proteins were acetone precipitated and re-suspended in PBS (see Section 2.2). The protein concentration was determined by BCA protein assay and 125µg of protein from both cell lines was then acetone precipitated again and re-suspended in 100µl of 2D Re-hydration buffer. The proteins in both samples were then separated by 2-DE (see Section 2.6). Figure 3.1 shows the 2-D gels of the CM from the RG/C2 and the PRG/C2 cell lines. Both gels showed very few spots indicating that not enough protein was loaded onto the gels due to either not enough starting material or loss of protein during two rounds of acetone precipitation.

CM from both cell lines were collected again (as previously described), the proteins were acetone precipitated and re-suspended in 2D re-hydration buffer. The protein concentration was determined using the 2D Quant Kit (Pierce) (see Section 2.6) and 160µg of protein from both cell lines were separated by 2-DE. Figure 3.2 shows the 2-D gels of the CM from the RG/C2 and the PRG/C2 cell lines. Both gels still did not show the presence of many protein spots and the gels also exhibited horizontal and vertical streaking. This could be due to one the following:

- *Incomplete or excessive isoelectric focusing*. Incomplete focusing does not allow proteins the opportunity to form discrete spots. Also, isoelectric focusing of IPG strips of varying pH ranges at the same time can lead to the same problem. If a sample is more conductive, it will draw most of the current and decrease the focusing rate for the other strips.

(a)



(b)

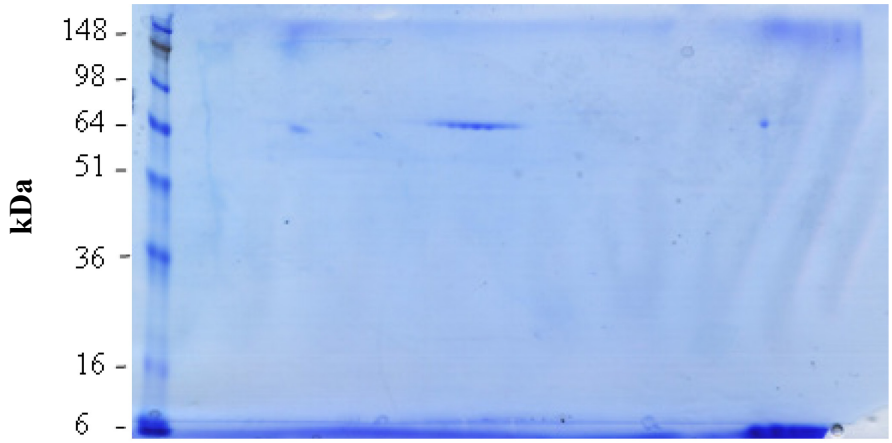


Figure 3. 1 : Initial 2D-PAGE Analysis of CM collected from the RG/C2 and PRG/C2 cell lines.

CM from $8.4x10^5$ cells for both cell lines were acetone precipitated and re-suspended in PBS and the protein concentration was determined using BCA protein assay. The proteins were then acetone precipitated again for re-suspension in re-hydration buffer and the proteins were separated by 2D gel electrophoresis. (a) CM derived from the RG/C2 cell line (125µg). (b) CM derived from the PRG/C2 cell line (125µg).

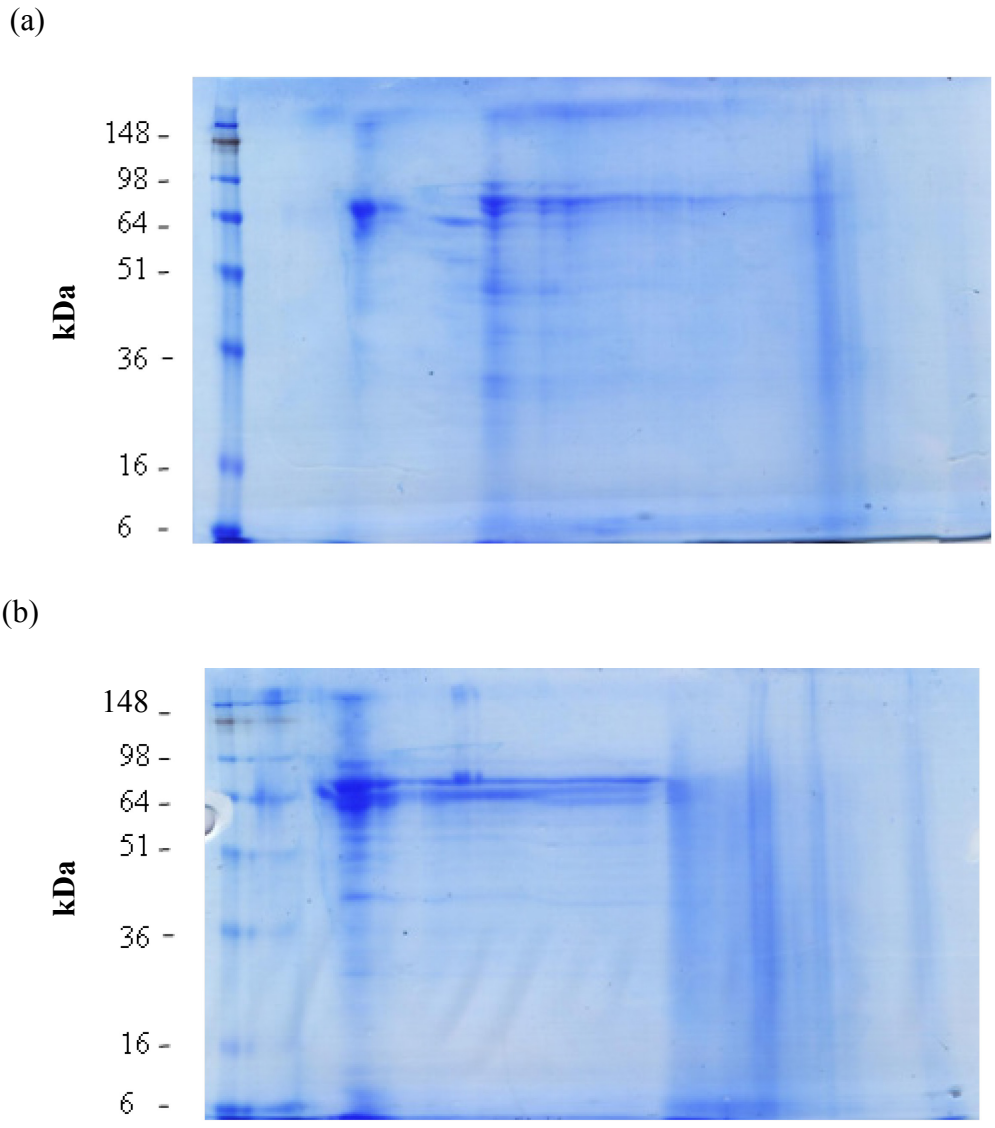


Figure 3. 2 : 2D-PAGE Analysis of CM collected from the RG/C2 and PRG/C2 cell lines.

CM from both cell lines was acetone precipitated and re-suspended in re-hydration buffer. Protein concentration was determined using the 2D Quant kit (Pierce) and the proteins were separated by 2D electrophoresis. (a) CM derived from the RG/C2 cell line (160µg). (b) CM derived from the PRG/C2 cell line (160µg).

- *Protein Overloading*. This causes aggregation of proteins resulting in precipitation at their isoelectric point. This is prevented by ensuring the protein load is within the guidelines set by the manufacturer. The recommended load for the IPG strips used in this study is 100-200µg and only 160µg of protein was used suggesting that protein overloading was not a problem in this case.
- *Problems with Protein Preparation*. Contaminants such as salt, detergents, nucleic acids and lipids can alter a protein isoelectric point therefore altering its focusing on an IPG strip. As these samples were only acetone precipitated prior to IEF, it's possible they may contain a few of these contaminants.
- *Protein Oxidation and Disulphide Formation*. Disulphide bond formation prior to 2DE can occur randomly intra- and inter molecularly to create various protein aggregates altering their isoelectric points and accumulation of protein multimers. Protein refolding and oxidative cross linking can be prevented by adding iodoacetamide to the equilibration buffer to IEF to block the cysteine sulfhydryl groups. As the equilibration buffer used in this buffer already contains iodoacetamide, this is unlikely to be the cause of streaking in these gels.
- *Poor Protein Solubilisation*. This can be prevented by adding urea, NP-40, CHAPS, DTT or ASB-14 to the samples. The re-hydration buffer used in this study contains urea, DTT and CHAPS therefore poor protein solubilisation is unlikely to be the cause of streaking in these gels.
- *Poor placement of IPG strips on the 2nd dimension gel.*

After considering the possible reasons for streaking in the 2D gels shown in Figure 3.2, the most likely cause appeared to be problems with protein preparation due to the

presence of contaminants. The experiment was then repeated including a sample cleanup step using the 2D Cleanup Kit (Pierce).

CM was collected from the RG/C2 and PRG/C2 cell lines as previously described. The proteins were acetone precipitated and re-suspended in 2D Re-hydration buffer. The protein concentration was determined using the 2D Quant Kit (Pierce) (as described in Section 2.6) and 200µg of protein from both cell lines was prepared for 2DE using the 2D Cleanup Kit (Pierce) (as described in Section 2.6). The proteins in both samples were then separated by 2DE and Figure 3.3 shows the 2D gels for both the RG/C2 and the PRG/C2 cell lines carried out in biological triplicate.

Figure 3.3 shows that the 2D gels were more successful with an increased number of spots observed demonstrating very similar protein spot patterns between the replicate experiments. Examination of the 2D gels from the biological replicates from the same cell line shows very similar spot patterns with differences in the staining intensities although the sample preparation in both experiments were identical. Figure 3.3 also shows very little difference in spot patterns between the RG/C2 and PRG/C2 gels in all three experiments. Each of the 2D gels also shows an area of the gel of heavy CBB G250 staining (circled) representing an area of high abundance proteins that do not resolve very well using 2-DE. These results suggested that the sample requires some pre-fractionation steps prior to 2-DE.

As described Section 1.4.4.1, 2-DE does have its limitations as demonstrated by the gels in Figure 3.3 with only moderate reproducibility and suboptimal separation of proteins. Taking into account these limitations and difficulties and after examination of the 2D gels in Figure 3.3, it was evident that due to the complexity of the sample, pre-fractionation of the CM was required to identify differences in the secreted proteins from these cell lines.

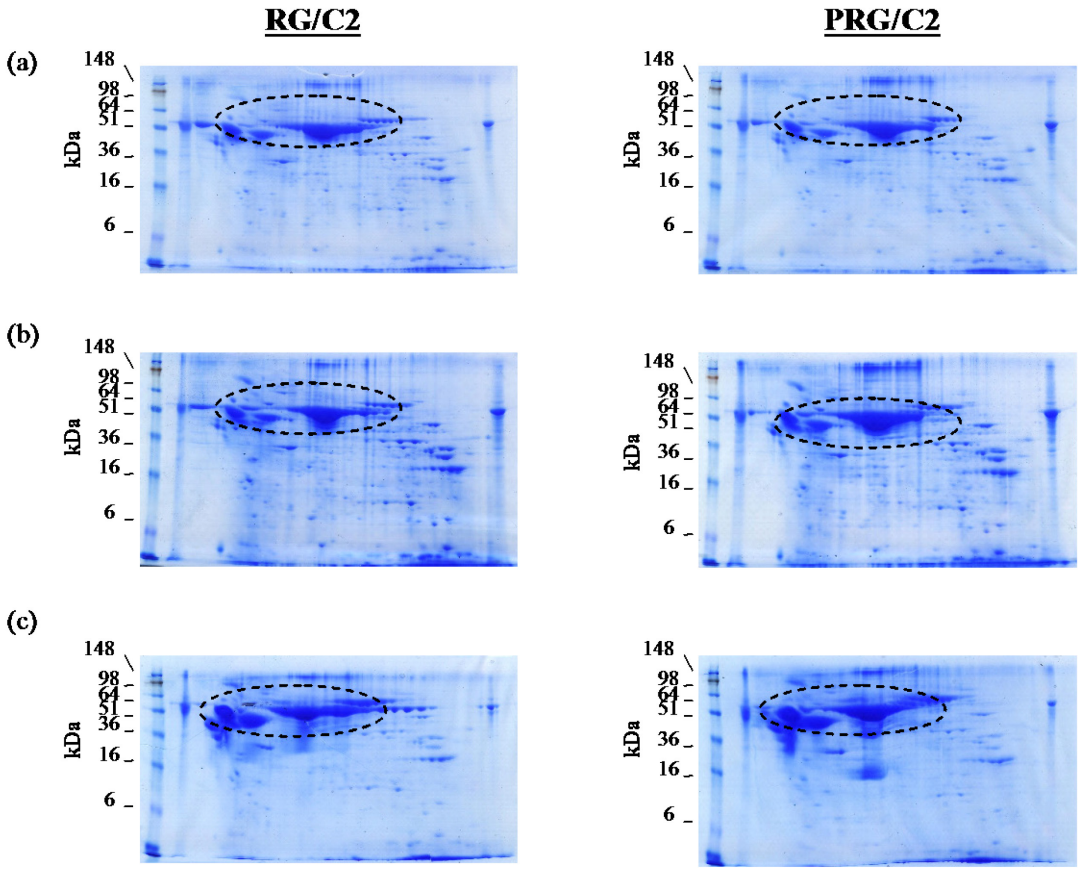


Figure 3. 3 : 2D-PAGE Analysis of CM collected from the RG/C2 and PRG/C2 cell lines following sample cleanup using the 2D Cleanup Kit.

CM from both cell lines was collected and acetone precipitated followed by re-suspension in re-hydration buffer. The protein concentration was determined using the 2D Quant kit (Pierce) and then cleaned using the 2D Cleanup kit (Pierce). 200µg of protein from both cell lines were separated by 2D electrophoresis. (a), (b) and (c) represent biological triplicates. (Dashed circle indicates area of gel with an abundance of proteins that do not resolve very well).

3.3 Analysis of Microvesicles Isolated from Conditioned Media

As described in Section 1.6, the importance of proteins being released in vesicles has been described in a variety of species and physiological conditions and is thought to be crucial in development (Ellis et al., 1989, Muraoka et al., 1993), platelet activation (Polette et al., 1997) and in the immune system (Klein et al., 2002, Sun and Hemler, 2001, Yurchenko et al., 2005). Vesicle shedding has been demonstrated in normal cells under certain physiological conditions (Klein et al., 2002, Polette et al., 1997, Yurchenko et al., 2002) but occurs at much higher rates in tumour cells (Tang and Hemler, 2004, Yurchenko et al., 2001). It's thought to play an important role in tumour progression and malignant transformation by shedding tumour surface antigens from plasma membranes; protecting the tumour from immune responses (Fiucci et al., 2002, P. Liu et al., 2002b, Razani and Lisanti, 2001, Tang and Hemler, 2004); important in angiogenesis (Basset et al., 1990) and plays an important role in tumour-stroma interactions (Guo et al., 2000). MVs consist of two types of vesicles: small exosome-like vesicles (formed in MVBs and secreted after fusion with the plasma membrane) and larger membrane derived vesicles often termed shedding vesicles. As the importance of MV release by tumour cells has already been shown, they could potentially be a useful source of proteins that could serve as biomarkers of early tumour progression.

Extracellular Matrix MetalloProteinase Inducer (EMMPRIN) or CD147/Basigin was first identified as a factor expressed on the surface of tumour cells responsible for inducing MMP synthesis in neighbouring fibroblasts (Ellis et al., 1989, Guo et al., 1997, Kataoka et al., 1993, Nabeshima et al., 1991). It was first isolated and purified from plasma membranes as a 58kDa glycoprotein and was originally designated tumour collagenase stimulating factor (TCSF) (Biswas et al., 1995). The cellular localisation of EMMPRIN at the cell surface was

confirmed by immunofluorescent staining but could also be observed in intracellular vesicles (Caudroy et al., 1999).

In various cell lines EMMPRIN stimulates the production of MMP-1, MMP-2, MMP-3 and MT-MMPs (membrane types MMPs) but has no effect on their inhibitors TIMP-1 or TIMP-2 (Ellis et al., 1989, Guo et al., 1997, Kanekura et al., 2002, Kataoka et al., 1993, Polette et al., 1997, Taylor et al., 2002). EMMPRIN induced expression of MMPs is thought to be cell specific with not all MMPs stimulated at the same time (Kanekura et al., 2002, Toole, 2003). EMMPRIN has also shown increased expression in various human carcinoma cells compared to normal cells (Guo et al., 1997, R. Li et al., 2001, Lim et al., 1998, Taylor et al., 2002) and expression levels can be correlated with MMP expression levels in stromal fibroblasts from various tumours (Caudroy et al., 1999, Thorns et al., 2002).

It was originally thought that tumour derived EMMPRIN acted on fibroblasts only by cell-to-cell contacts but evidence shows that native soluble or recombinant EMMPRIN in tumour cell derived CM induces MMP production/secretion from fibroblasts (Ronnov-Jessen et al., 1996, Sidhu et al., 2004). Taylor *et al* showed that full length EMMPRIN was the factor in CM from the breast cancer cell line MDA MB 231 responsible for inducing MMP2 secretion from fibroblasts which was mediated through the activation of the phospholipase A_2/5-lipoxygenase pathway (Taylor et al., 2002). Sidhu *et al* later demonstrated that EMMPRIN was released from lung carcinoma cell lines in MVs and could be used as a way for epithelial cells to communicate with distant cells as well as neighbouring cells (Sidhu et al., 2004). These MVs were short-lived, releasing there contents within four hours when incubated at 37ºC.

As EMMPRIN is known to be secreted in MVs and play an important role in tumourigenesis, it was an ideal protein to use as a positive control to optimise the techniques

used to investigate differentially secreted proteins in MVs from RG/C2 and PRG/C2 cell lines.

CM was collected from both cells lines (serum starved for 24hrs), the MV and REM CM fractions were collected and Western blotting analysis was performed. The Western blot shown in Figure 3.4a confirms the presence of EMMPRIN in the total CM and the MV fraction collected from both cell lines but with very little EMMPRIN observed in the REM CM fraction. These results demonstrate that the majority (if not all) of the EMMPRIN observed in CM is secreted in MVs and is in agreement with the observations made by Sidhu *et al* in lung carcinoma cell lines (Sidhu et al., 2004). The Western blot for EMMPRIN identified a band at 58kDa indicating that full length EMMPRIN is released in the MV fraction and not as a cleaved product from the plasma membrane, also in agreement with previous reports (Sidhu et al., 2004).

The cell lysates from both cell lines were also harvested and nuclear, heavy membrane and cytosolic/light membrane fractions were prepared as described in Section 2.2. Western blot analysis showed EMMPRIN was present in the heavy membrane fraction and very little in the nuclear or cytosolic/light membrane fraction (Figure 3.4b) confirming that EMMPRIN is a membrane-bound protein in these cells as published in many other cell lines (Guo et al., 2000). There was also an increased expression of EMMPRIN released in the heavy membrane of the PRG/C2 cells compared to the RG/C2 cells suggesting it may play a role in early tumour progression.

The MV preparation protocol described by Sidhu *et al* isolates a mixed population of both the small exosomes (50-100nm) and the larger shedding vesicles (100nm-1μm). An additional fractionation step to isolate the two subtypes using sucrose density gradient

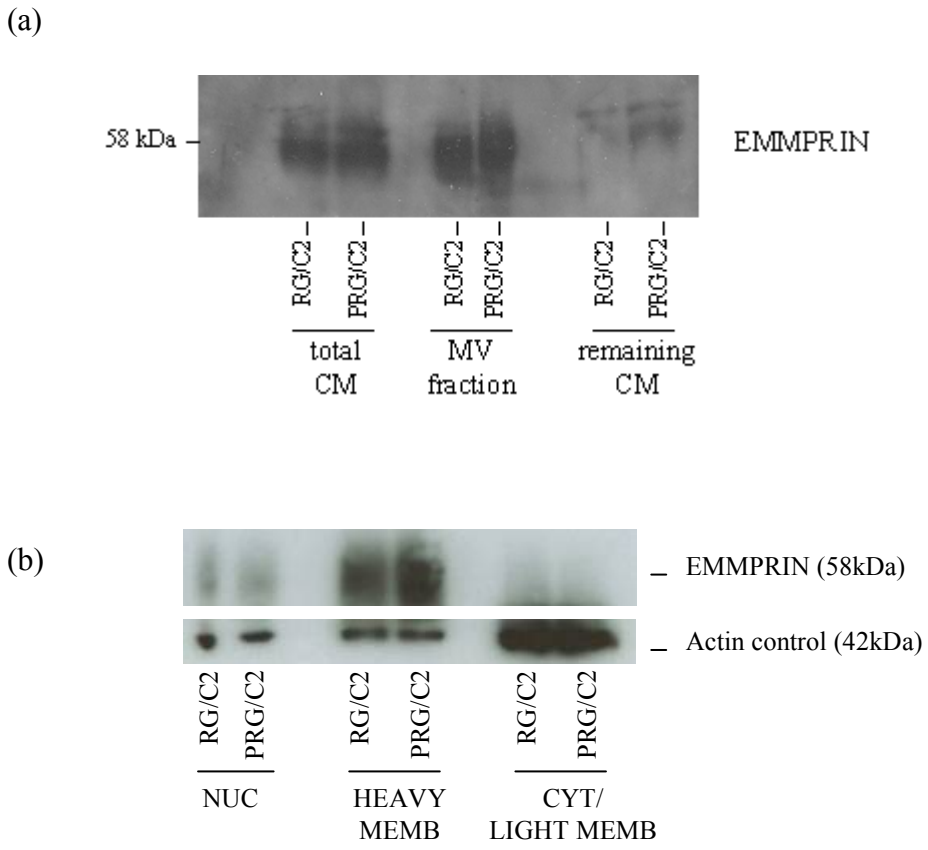


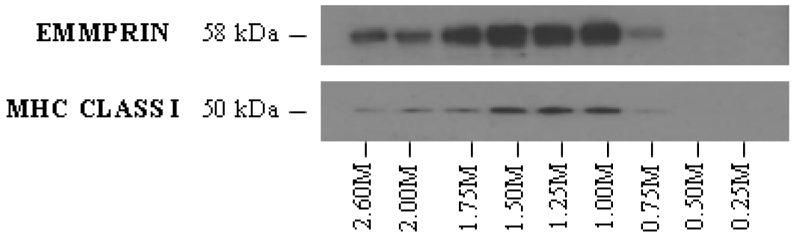
Figure 3. 4 : Investigating the presence of EMMPRIN in RG/C2 and PRG/C2 cells and its release into CM.

The presence of full length EMMPRIN was confirmed by Western blotting in total CM, the MV fraction and the REM CM fraction prepared from RG/C2 and PRG/C2 cells ($3x10^6$ cells). The samples were normalised using cell number. (b) The cell lysate was harvested from both cell lines and fractionated to identify where EMMPRIN is located in the cell. 20µg of each fraction was blotted for EMMPRIN and actin (loading control).

ultracentrifugation could be used to enrich vesicle sub species. In an attempt to investigate whether EMMPRIN is released by the RG/C2 cell line in exosomes or shedding vesicles, MVs were collected and enrichment for both types using the sucrose density gradient ultracentrifugation method (described in Section 2.3) was performed. The proteins in 5µl of each fraction were separated on 4-12% tris-glcine gels and the presence of EMMPRIN and MHC class 1 (an exosome marker) was investigated by Western blot analysis (Figure 3.5a). The remaining 30µl were also run on a 4-12% tris-glcine gel and the proteins were stained with CBB G250 (Figure 3.5b).

Western blot analysis showed the presence of EMMPRIN in the sucrose fractions between 2.60-0.75M and absent in fractions 0.50M and 0.25M with EMMPRIN enriched in the four fractions between 1.75-1.00M. MHC class 1 was also present in the sucrose fractions between 2.60-0.75M and enriched in the three fractions between 1.50-1.00M. These results could indicate that EMMPRIN is present in the same species of vesicle as MHC class 1 is released in (i.e. exosomes). However, EMMPRIN was present in seven of the nine fractions suggesting this method couldn't determine the subspecies of vesicle EMMPRIN was present in. This could be because it is present in both subtypes or could be very abundant leading to it's detection in most of the fractions. The CBB G250 stained gel shown in Figure 3.5b shows that similar protein staining patterns bands in multiple fractions indicating that the sucrose density gradient technique didn't successfully separate the vesicle subtypes. Also, the Western blots show bands for EMMPRIN and MHC class I in the majority of the fractions with the most intense coomassie staining i.e. 1.75M – 1.25M. This suggests that the more intense bands for both proteins are a result of more total protein being present in that fraction. As the sucrose density gradient method didn't appear to separate the vesicle subtypes very well, this technique was not investigated any further in this study.

(a)



(b)

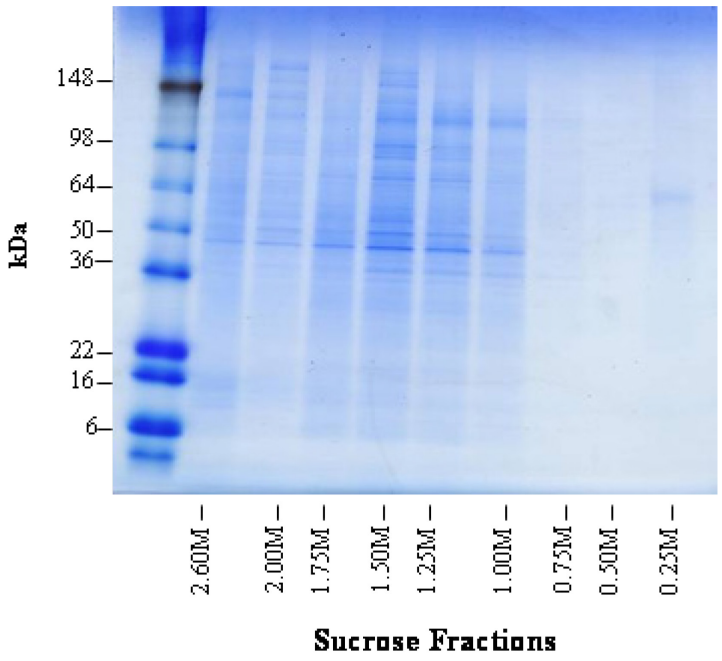


Figure 3. 5 : Enrichment of MVs prepared from CM derived from RG/C2 cells using Sucrose Density Gradient Ultracentrifugation.

MV's were prepared from CM derived from RG/C2 cells ($2x10^8$ cells) and the vesicles were then separated using a sucrose density gradient protocol to attempt to enrich the 2 sub-types of MVs. (a) Presence of EMMPRIN and MHC Class I in each fraction (5µl) was confirmed by Western blot analysis. (b) The remaining 35µl of each fraction was run on a 4-12% tris – glycine gel and stained with CBB G250.

Next, the proteins in the MV fractions from the RG/C2 cell line were separated by 1D gel electrophoresis and proteomic techniques were used to detect EMMPRIN. CM was collected from RG/C2 cells (6 x 10^6) and an MV fraction was prepared. The proteins in the total CM, MV and REM CM fractions were separated on a 4-12% tris-glycine gel twice. One gel was stained with CBB G250 and the other was used to confirm the presence of EMMPRIN by Western blot analysis. The whole MV lane of the CBB G250 stained gel was cut into 10 equal size bands and the proteins in each band were digested with trypsin (Figure 3.6). Each sample was then analysed by LC MS/MS using an acetonitrile gradient shown in Figure 2.3a and the acquired spectra was searched against a human database (generated from an nr database downloaded from www.ncbi.nlm.nih.gov). The results were filtered using the Xcorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50) parameter. There were 47 proteins identified (with at least 2 first hit peptides) however, no EMMPRIN peptides could be detected in any of the samples even after decreasing the filter parameters to Xcorr = 1.00 (see Supplementary Table 3a).

There are two possible reasons why EMMPRIN could be detected by Western blotting and not by LC MS/MS: 1) Western blotting is more sensitive than LC MS/MS and there was an insufficient amount of EMMPRIN present in the sample; or 2) The sample was far too complex to detect the lesser abundance proteins. The MS spectrum was collected in a data-dependant manner and only the five highest intensity ions are picked for further fragmentation (MS/MS). Therefore if the sample is too complex, the lesser abundant peaks are never picked for further fragmentation and identified.

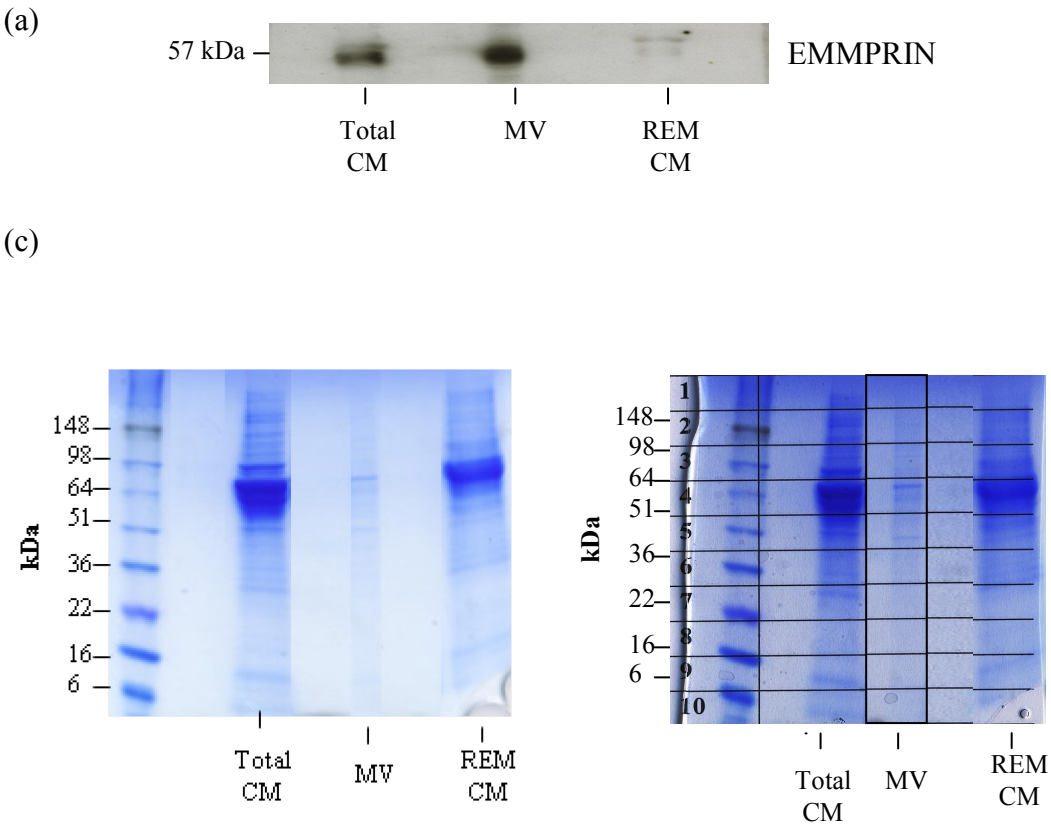


Figure 3. 6 : Preparation of an MV fraction from CM derived from RG/C2 cells ($6x10^6$ cells) to attempt to identify EMMPRIN using proteomic techniques.

Proteomic techniques were used to confirm the presence of EMMPRIN in total CM, the MV and the remaining CM fractions derived from RG/C2 cells. (a) Western blot analysis for the presence of EMMPRIN in all fractions ($3x10^6$ cells). (b) The proteins in each fraction ($6x10^6$ cells) were separated on a 4-12% tris-glycine gel and stained with CBB G250. (The unlabelled lanes were samples from a cell line that's not relevant to this study). (c) The MV fraction was cut into 10 equal size bands and in-gel digestion was performed for each band.

In this case, an insufficient amount of EMMPRIN seems the most likely reason why it could not be detected using proteomic techniques. There were very few intensely stained and discrete bands in the MV lane shown in Figure 3.6b indicating that more protein identifications could be made by starting with more protein. Also, there were proteins identified in the areas of the gel with very little staining (i.e. bands 7-10) indicating that the CBB G250 is not as sensitive as the LC-MS.

The previous experiment was repeated using twice the starting material in an attempt to identify EMMPRIN using LC/MS/MS methods. MVs collected from 1.2 x 10^7 RG/C2 cells (twice the starting material previously used) and the proteins were again separated on a 4-12% tris-glycine gel and stained with CBB G250 (Figure 3.7a). The presence of EMMPRIN was confirmed by Western blot analysis (data not shown) and the region of the MV lane between 40-70kDa was cut into 4 bands for in-gel digestion (Figure 3.7b). Each band was analysed by LC MS/MS and the acquired spectra were searched against the same human database used previously. The CBB G250 stained gel shown in Figure 3.7a shows an increased number of discrete protein bands with much greater staining intensity observed throughout the whole MV lane compared to the previous experiment (Figure 3.6). There were 194 proteins identified (see Supplementary Table 3b) in the 4 bands (with at least 2 first hit peptides) using the single threshold filter parameter of Xcorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50). Again, there were no EMMPRIN peptides observed in this region of the gel despite its presence being confirmed by Western blot analysis even when the filter parameters were reduced to Xcorr = 1.00 (data not shown). Figure 3.7a also shows that the proteins in the MV lane resolved well with discrete bands observed throughout the lane. Whereas, in the total CM and the REM CM lanes, there was diffuse staining throughout the whole lane with a large intensely stained band at ~64kDa not observed in the MV lane. These results demonstrate that

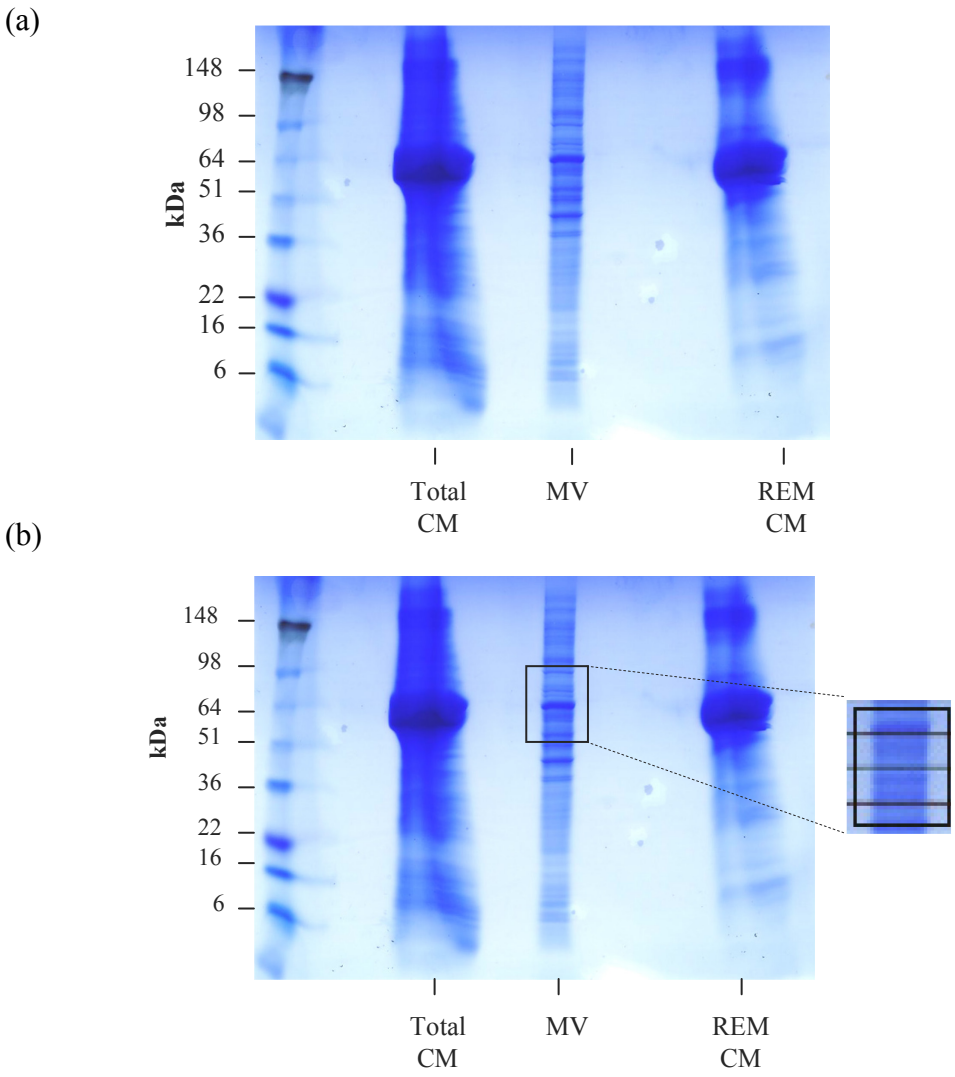


Figure 3. 7 : Investigating the presence of EMMPRIN in MVs from RG/C2 cells ($1.2x10^7$ cells) using proteomic techniques.

Proteomic techniques were used to confirm the presence of EMMPRIN in total CM, the MV and the remaining CM fractions derived from RG/C2 cells. (a) The proteins from the total CM, the MV and the remaining CM fractions from RG/C2 cells ($1.2x10^7$ cells) were separated on a 4-12% tris-glycine gel and the proteins were stained with CBB G250. (b) The region between 40-70kDa of the MV lane was cut into 4 bands and in-gel digestion was performed.

this large region of the most abundant proteins observed in the total CM fraction was isolated and retained in the REM CM fraction during MV preparation. These results demonstrate that the MV preparation protocol is an ideal method for fractionating the CM prior to 1D gel electrophoresis as the removal of these proteins should improve the identification of the lower abundance proteins.

This experiment was repeated (as described above) one more time in an attempt to successfully identify EMMPRIN peptides using a MV fraction prepared from $2.4x10^7$ RG/C2 cells. The region of the MV lane between 40-70kDa was cut into 6 bands and digested with trypsin for peptide analysis by LC MS/MS (Figure 3.8). The results generated 163 proteins (with at least 2 first hit peptides) using the single threshold filter parameter of Xcorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50) (see Supplementary Table 3c). EMMPRIN peptides still could not be identified even after reducing the filter parameters to Xcorr = 1.00 (data not shown). The gel shown in Figure 3.8a shows very intense CBB G250 staining throughout the whole gel indicating it has been overloaded with protein. The sample would need further fractionation either prior to separation by SDS-PAGE or peptide fractionation following in-gel digestion using a strong cation exchange column. Although more protein was used in this experiment compared to the previous experiment (Figure 3.7) which in theory should lead to more protein identifications as an increased amount of the lower abundance proteins are present, only 163 proteins were identified compared to the 194 proteins previously identified in the same region of the gel. These results demonstrate that more total protein does not necessarily lead to more protein identifications. This could be due to either experimental error when the bands were excised or most likely due to the gel being overloaded with protein. The most abundant proteins are no longer seen as discrete bands but become much more diffuse

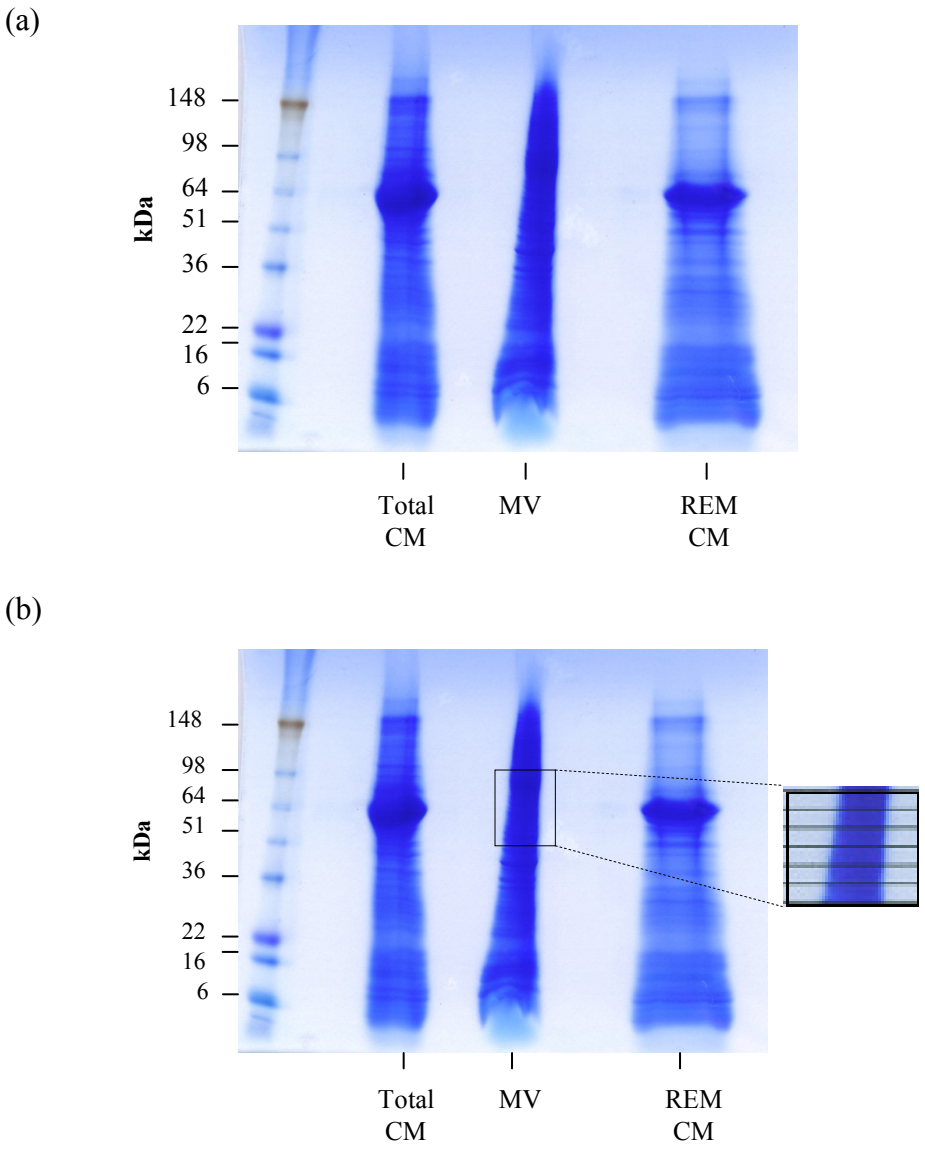


Figure 3. 8 : Investigating the presence of EMMPRIN in MVs from RG/C2 cells ($2.4x10^7$ cells) using proteomic techniques.

The proteins from the total CM, the MV and the remaining CM fractions derived from RG/C2 cells ($2.4x10^7$ cells) were separated on a 4-12% tris-glycine gel and the proteins were stained with CBB G250. (b) The region between 40-70kDa of the MV lane was cut into 6 equal size bands and in-gel digestion was performed.

and can be identified in multiple bands and the abundant proteins prevent the identification of the lower abundant ones.

One example is the protein KIAA0051 (gi40789059), Table 3.1 shows the peptides identified from this protein in the experiments outlined in Figures 3.7 and 3.8. Table 3.1a show that there were 19 peptides identified in 1 out of the 4 protein bands analysed in Figure 3.7 and there were 15, 10 and 10 proteins identified respectively in 3 of the 6 bands analysed in Figure 3.8. These results demonstrate that increasing the amount of protein does increase the number of peptides however, multiple peptides in only one band is sufficient to have confidence that the protein is present. The presence of same peptides in multiple bands only prevents the identification of other peptides that are present i.e. suppression of the lower abundant peptides by the high abundance ones preventing them from being isolated for further fragmentation and therefore being identified. This illustrates the importance of optimising the protein pre-fractionation methods before LC MS/MS analysis to allow sufficient separation of proteins.

3.4 Conclusion

The aim of this chapter was to use 2-DE to investigate differentially secreted proteins from the colorectal cancer cell lines RG/C2 and PRG/C2. CM was collected from both cell lines and the proteins were separated on a 2D gel. The gels shown in Figure 3.3 demonstrated the complexity of CM collected from these cell lines and also highlighted some of the difficulties with 2-D gel technology. The gels exhibited moderate reproducibility between biological replicates and an area of the gel showed poor separation of proteins therefore highlighting the need for pre-fractionation of the sample to improve separation of proteins.

(a)

Scan(s)	Sequence	MH+	Charge	XC	Score Delta Cn	Accession Sp	RSp	Peptides (Hits) Ions
Figure 3.3 sample 1								
KIAA0051					110.24	**40789059.0**		11 (11 0 0 0 0)
512 - 518	-.LQQTYAALNSK.-	1237.39	2	3.33	0.51	1054.7	1	17/20
560 - 567	-.TC*LDNLASK.-	1022.13	2	2.15	0.20	453.2	2	12/16
722	-.DHINDIIK.-	968.09	2	2.52	0.23	1179.6	1	12/14
748 - 755	-.TILLNTK.-	802.98	1	1.51	0.11	186.7	1	8/12
824	-.LTELGTVDPK.-	1073.22	2	3.23	0.35	2045.4	1	16/18
1505 - 1508	-.LPYDVTPEQALAHEEVK.-	1940.14	3	4.83	0.56	1736.6	1	32/64
1689 - 1694	-.IIGNLLYYR.-	1125.35	2	3.20	0.35	830.1	1	15/16
2122 - 2128	-.FQPGETLTEILETPATSEQEAEHQR.-	2843.01	3	2.52	0.25	429.3	1	29/96
2181	-.MFLGDNAHLSIINEYLSQSYQK.-	2572.88	2	3.85	0.43	1008.2	1	20/42
2186	-.MFLGDNAHLSIINEYLSQSYQK.-	2572.88	3	3.14	0.32	826.9	1	27/84
2200	-.GVLLEIEDLQVNQFK.-	1746.00	2	4.50	0.55	1978.9	1	20/28

(b)

Scan(s)	Sequence	MH+	Charge	XC	Score Delta Cn	Accession Sp	RSp	Peptides (Hits) Ions
Figure 3.4 sample 1								
KIAA0051					158.22	**40789059.0**		16 (15 1 0 0 0)
243 - 250	-.MLQHAASNK.-	1000.16	2	2.03	0.21	876.1	2	12/16
244	-.MLQHAASNK.-	1000.16	2	2.23	0.32	843.9	2	14/16
405	-.LQQTYAALNSK.-	1237.39	2	2.19	0.34	940.7	1	17/20
480	-.DHINDIIK.-	968.09	2	2.39	0.25	732.3	1	11/14
526	-.TILLNTK.-	802.98	1	1.53	0.11	174.8	5	8/12
640	-.FDVPGDENAEMDAR.-	1566.63	2	2.28	0.21	489.2	1	17/26
806 - 812	-.ATFYGEQVDYYK.-	1484.59	2	3.03	0.55	784.6	1	16/22
837 - 842	-.ITLQDVVSHSK.-	1227.39	2	2.56	0.18	1181.6	1	16/20
876	-.VDFTEEEINNMK.-	1469.60	2	2.45	0.32	818.5	1	15/22
938	-.RLIVDVIR.-	984.22	2	2.37	0.11	689.7	1	12/14
1018 - 1028	-.NKEQLSDMMMINK.-	1582.89	2	4.44	0.21	1787.7	1	19/24
1058 - 1064	-.YQELINDIAR.-	1235.37	2	2.14	0.28	1261.3	1	15/18
1299	-.FFQTAC*DVPELQDK.-	1698.85	2	2.48	0.38	542.2	1	15/26
1564 - 1570	-.FLSAIVSSVDK.-	1166.35	2	2.95	0.38	1430.5	1	17/20
1772 - 1779	-.IIGNLLYYR.-	1125.35	2	2.94	0.30	787.8	1	14/16
2097	-.EQLWLANEGLITR.-	1543.75	2	2.23	0.33	711.1	1	15/24

(b) - contd

Scan(s)	Sequence	MH+	Charge	XC	Score Delta Cn	Accession Sp	RSp	Peptides (Hits) Ions
Figure 3.4 sample 2								
KIAA0051					100.21	**40789059.0**		10 (10 0 0 0 0)
689	-.ALQSPALGLR.-	1026.22	2	2.35	0.22	886.0	2	15/18
789	-.VDFTEEEINNMK.-	1469.60	2	2.78	0.31	1127.8	1	16/22
808 - 812	-.RIPADTFAALK.-	1203.42	2	2.11	0.04	672.2	2	13/20
819 - 826	-.GLQQQNSDWYLK.-	1480.61	2	2.79	0.27	1131.1	1	16/22
843 - 850	-.TLQALQIPAAK.-	1154.39	2	2.64	0.38	1415.5	1	16/20
904	-.YQELINDIAR.-	1235.37	2	2.17	0.37	925.4	1	13/18
1115 - 1119	-.IFYPETTDIYDR.-	1533.66	2	2.20	0.30	515.1	1	15/22
1564	-.YGIQMPAFSK.-	1142.35	2	2.70	0.51	760.8	1	14/18
1898 - 1904	-.LGLAPQIQDLYGK.-	1416.65	2	2.89	0.48	582.7	1	19/24
2069 - 2090	-.EQLWLANEGLITR.-	1543.75	2	4.24	0.41	1594.0	1	21/24
Figure 3.4 sample 3								
KIAA0051					100.18	**40789059.0**		10 (10 0 0 0 0)
736 - 743	-.ALESGDVNTVWK.-	1319.45	2	2.36	0.38	1019.3	1	16/22
819	-.RIPADTFAALK.-	1203.42	2	2.27	0.13	1024.3	1	15/20
824	-.VDFTEEEINNMK.-	1469.60	2	2.99	0.35	1394.5	1	17/22
846 - 857	-.RIPADTFAALK.-	1203.42	2	2.15	0.24	996.9	1	14/20
848 - 854	-.GLQQQNSDWYLK.-	1480.61	2	3.16	0.31	1575.6	1	17/22
891 - 896	-.TLQALQIPAAK.-	1154.39	2	2.79	0.37	1404.7	1	16/20
1212 - 1226	-.IFYPETTDIYDR.-	1533.66	2	2.15	0.39	455.5	1	15/22
1583 - 1600	-.YGIQMPAFSK.-	1142.35	2	2.63	0.41	712.9	1	13/18
1899	-.LGLAPQIQDLYGK.-	1416.65	2	3.01	0.42	350.0	1	14/24
2090 - 2097	-.EQLWLANEGLITR.-	1543.75	2	3.51	0.47	1907.9	1	21/24

Table 3. 1 : The identification of KIAA0051 (gi40789059) in the experiments outlined in Figures 3.7 and 3.8.

KIAA0051 is an example of a protein identified in just 1 band out of the 4 analysed in Figure 3.7 and in 3 bands out of the 6 identified in Figure 3.8. Table (a) shows the 19 peptides identified in 1 sample from Figure 3.7. Table (b) shows the 15, 10 and 10 peptides identified in 3 separate protein bands from Figure 3.8.

As previously described, MV release from cells occurs at much higher rates in tumour cells compared to normal cells and are thought to play an important role in tumour progression and malignant transformation. It has been shown in many studies that MVs can be isolated from CM derived from cancer cell lines. Therefore, MVs were isolated from CM from both the RG/C2 and PRG/C2 cell lines in an attempt to pre-fractionate the CM further using EMMPRIN as a positive control for MV preparation. In both the RG/C2 and PRG/C2 cells, Western blot analysis confirmed the presence of full length EMMPRIN in the total CM and the MV fraction but not in the REM CM fraction indicating that all the EMMPRIN released by these cells is released via vesicle shedding (Figure 3.4a).

Western blot analysis of the subcellular fractions of the cell lysates collected from both cell lines also confirmed that in these cell lines EMMPRIN is a membrane bound protein with an increase in the membrane fraction of the PRG/C2 cell line. Separation of the two main sub-types of MV using sucrose density gradient ultracentrifugation failed to distinguish conclusively whether EMMPRIN was released in the smaller exosome like vesicles or those shed from the plasma membrane although logic would dictate that EMMPRIN is released in those shed from the plasma membrane as evidence shows it is a membrane bound protein.

The next step was to use proteomic techniques to further study the importance of EMMPRIN release in MVs during early stage tumour progression. After three attempts, EMMPRIN peptides could not be identified in MVs from the RG/C2 cell line despite confirmation of it's presence by Western blotting and double checking EMMPRIN's presence in the database used. In Figure 3.7, the area of the MV lane between 40-70kDa was excised in four bands and analysed by LC MS/MS. There were 194 proteins (with at least 2 first hit peptides) identified in this region of the gel but failed to identify EMMPRIN. These results

suggest that there may be other very interesting proteins released in MVs during early stage colorectal cancer that could be equally as important as EMMPRIN.

The remainder of the gel shown in Figure 3.7 was cut into 12 equal size bands, the proteins were digested with trypsin and the tryptic peptides were analysed by LC MS/MS. The combined results were filtered using XCorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50) and identified a total of 638 proteins (with at least two first hit peptides) (results not shown). These results indicate that quantitative proteomic techniques could be used to investigate the differential secretion of proteins in MV from the RG/C2 and PRG/C2 cell lines to identify possible markers of early tumour progression in colorectal cancer.

Chapter 4

RESULTS

Optimisation of Conditions Required for Quantitative Proteomic Profiling of MVs Released by RG/C2 and PRG/C2 Cell Lines

Chapter 4 : Optimisation of Conditions Required for Quantitative Proteomic Profiling of MVs Released by RG/C2 and PRG/C2 Cell Lines.

4.1 Introduction

As discussed in the Section 1.1, there are currently two non-invasive biomarkers for CRC in use, FOBT and CEA. Both have limited uses and show low sensitivity and specificity for detecting early stage colorectal cancer (H. J. Kim et al., 2008). The FOBT test can be affected by the patient's diet to yield false-positive results and studies have shown the test to be unreliable. Serum CEA can only detect 30-40% of early stage colorectal cancers (Dukes' stage A and B) and its main use has been as a prognostic marker following surgery in patients with elevated pre-operative CEA. There are many proteomic studies investigating potential new biomarkers in serum for early tumour progression in colorectal cancer for use in screening programs but as of yet there are no successful candidates that have been transferred to the clinic.

The majority of these studies involve investigating possible markers in clinical samples such as serum, plasma and urine. With the large number of proteins identified in MVs from RG/C2 described in Chapter 3, it may be possible to identify small changes in these proteins in a cell model of early tumour progression that could be detected in the bloodstream. Studies have shown that MVs can be detected in clinical samples and using a quantitative shotgun proteomics approach to identify the proteins within MVs released by cells could reveal new potential biomarkers.

The aim of the work described in this chapter was to test the sensitivity of the LCQ Deca XP_{PLUS} instrument, and to optimise the search and filter parameters required for

analysing large scale shotgun proteomic datasets. This optimisation is essential to ensure that the number of false peptide and protein identifications is kept to a minimum whilst minimising the loss of true protein identifications (see Section 1.3 for a review of data analysis methods).

4.2 Demonstrating the Sensitivity of the LCQ Deca XP_{PLUS}

Firstly, the sensitivity of the LCQ Deca XP_{PLUS} was investigated by analysing increasing amounts of BSA peptides to determine the detection limits of the instrument. BSA was digested with trypsin in solution (Section 2.11.3) and diluted to a concentration of 1000fmoles/80µl. This stock solution was then used to perform a further nine 2-fold serial dilutions (Table 4.1a) and 80µl of each dilution was analysed by LC MS/MS using a 45min acetonitrile gradient (Figure 2.3a). The raw files were searched against a bovine database generated from an NR database downloaded from the NCBI website and Table 4.1a shows the number of BSA peptides identified and the percentage coverage of the total protein sequence represented by these peptides for each dilution using the Xcorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50) filter parameters. The highest percentage coverage of BSA was observed in the 1000, 500, and 250fmole samples (22.90%) with 104, 86 and 80 total peptides identified respectively with 14 unique peptides identified in each. There is no change in percentage coverage because the same peptides are identified in all three samples with some identified multiple times due to their high abundance at the higher concentrations. Multiple identifications of a single peptide sequence in a sample increase the confidence that the sequence is present especially if both the (+2) and (+3) charged species are identified. Therefore, the percentage coverage of a protein identified can be used as a good indicator of confidence that a protein is present but the number of total and unique peptides should also be considered.

(a)

Number of fmoles in 80µl	ng	Dilution factor	Total Peptides	Unique peptides	Percent Coverage
1000.00	66.00	Stock	104	14	22.90
500.00	33.00	1 : 2	86	14	22.90
250.00	16.50	1 : 4	80	14	22.90
125.00	8.25	1 : 8	46	13	17.13
62.50	4.13	1 : 16	35	12	19.28
31.25	2.06	1 : 32	31	10	17.63
15.63	1.03	1 : 64	19	7	13.84
7.80	0.52	1 : 128	11	6	10.05
3.90	0.26	1 : 256	3	2	3.95
1.90	0.13	1 : 512	0	0	0.00

(b)

m/z	547.6		582.5		653.6		682.6		740.6	
	KVPQVSTPTLVEVSR (+2)		LVNELTEFAK (+2)		HLVDEPQNLIK (+2)		RHPYFYAPELLYYANK (+2)		LGEYGFQNALIVR (+2)	
No of fmoles	RT (mins)	NL	RT (mins)	NL	RT (mins)	NL	RT (mins)	NL	RT (mins)	NL
1000	28.66	1.54E+08	33.23	2.67E+08	27.53	3.00E+08	36.57	1.63E+08	37.5	5.66E+08
500	28.92	1.21E+08	33.33	1.41E+08	27.52	1.47E+08	36.91	5.89E+07	37.59	1.66E+08
250	28.96	5.11E+07	33.36	5.69E+07	27.41	6.46E+07	37.07	3.53E+07	37.9	1.20E+08
125	28.66	2.43E+07	33.12	3.04E+07	27.17	3.51E+07	36.95	1.34E+07	37.47	7.46E+07
62.5	28.59	1.80E+07	33.17	1.35E+07	27.17	1.40E+07	36.74	5.44E+06	37.42	4.24E+07
31.25	28.9	4.11E+06	33.35	7.26E+06	27.32	6.67E+06	37.5	7.90E+06	37.8	8.99E+06
15.63	30.7	3.30E+06	34.9	4.16E+06	29.12	3.58E+06	-	-	39.39	1.00E+07
7.8	30.5	2.08E+06	35	1.92E+06	29	2.28E+06	-	-	40	3.66E+06
3.9	-	-	-	-	29.82	2.26E+06	-	-	-	-
1.9	-	-	-	-	-	-	-	-	-	-

Table 4. 1 : Determination of the sensitivity of the LCQ Deca XP_{PLUS} using BSA digested with trypsin in solution.

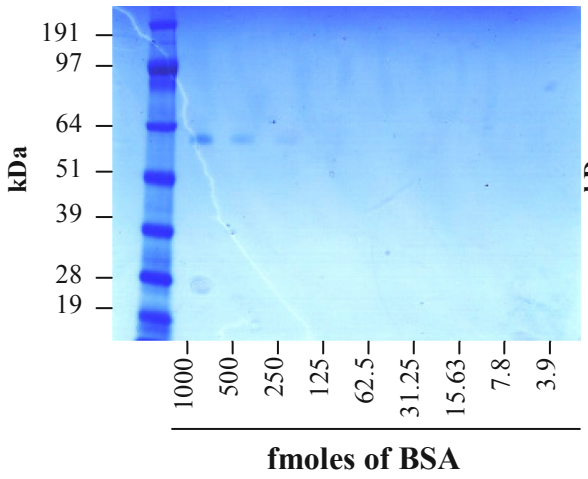
(a) Table to show the amount of BSA in fmoles and ng present in 80µl of each 2-fold serial dilution prepared from a 1000fmole/80µl stock solution and the number of peptides and percentage coverage observed in each sample (using the Xcorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50) filter parameters. (b) Table to show the retention time (RT), the peak intensity (NL) and the m/z ratio of the 5 most commonly identified peptides for each dilution.

The number of peptides identified and the percentage coverage then decreases with each serial dilution with no peptides identified in the lowest dilution (1.90fmole). The results in Table 4.1a indicate that the detection limit for BSA peptides was ~4 fmoles with only a single peptide observed. Table 4.1b shows the observed retention time, the peak intensity (NL), the m/z ratio and the sequence of the five peptides most commonly observed in the BSA peptide standards routinely used as indicators of optimal sensitivity in our lab. Each peptide has very similar retention times regardless of the amount of BSA with the NL decreasing as the amount of BSA decreases as expected.

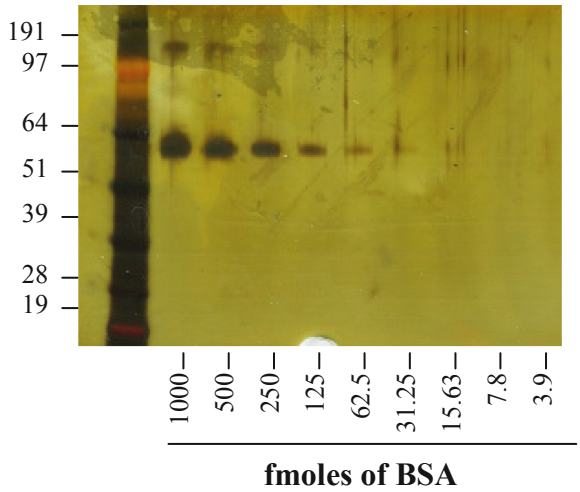
4.3 Investigating the Sensitivity of CBB G250 and Silver Staining Protocols

Next, the sensitivity of both the CBB G250 and the MS compatible silver staining protocols for proteins separated by 1D gel electrophoresis was determined by investigating the detection limits of both stains. A stock solution of 100fmoles/µl BSA was prepared and used to prepare eight further 2-fold serial dilutions (Figure 4.1a). 10µl of each sample (1000-3.91fmoles) was loaded onto a 4-12% bis-tris gel and stained using either the CBB G250 or the silver staining protocols described in Section 2.9.2. The table in Figure 4.1c shows the concentration, the number of fmoles and amount of BSA (in ng) present in 10µl of each dilution. Figure 4.1a and b shows the detection limits of both CBB G250 and the silver stain was 250fmoles (16.50ng) and ~31fmoles (~2ng) respectively. Winkler *et al* have since published a study comparing the detection limits of two CBB G250 staining protocols and fourteen different silver staining protocols (including the protocol used in this study) (Shevchenko et al., 1996, Winkler et al., 2007). The detection limit of their CBB G250 staining protocol for BSA was 125fmoles which was more sensitive than the 250fmole detection limit determined in this study. There is a slight difference in the composition of the two stains used which could explain these differences in sensitivity. Winkler *et al* also

(a)



(b)



(c)

Conc (fmoles/µl)	fmoles in 10µl	ng	CBB G250			Silver		
			Total peptides	Unique peptides	Percentage coverage	Total peptides	Unique peptides	Percentage coverage
100	1000	66	7	6	9.23	8	6	9.88
50	500	33	5	5	7.08	3	3	5.44
25	250	16.5	1	1	1.81	1	1	1.65
12.5	125	8.25	0	-	-	0	-	-
6.25	62.5	4.13	0	-	-	0	-	-
3.13	31.25	2.06	0	-	-	0	-	-
1.56	15.63	1.03	0	-	-	0	-	-
0.78	7.82	0.52	0	-	-	0	-	-
0.39	3.91	0.26	0	-	-	0	-	-

Figure 4. 1 : CBB G250 versus Silver staining?

A stock concentration of BSA (100fmoles/µl) was prepared and used to make a further eight 2-fold serial dilutions. 10µl of each concentration of BSA was then run on a 4-12% Bis- Tris gel to test the sensitivity of both the CBB G250 (a) and silver staining protocols (b). Each BSA band was excised and digested with trypsin for LC MS/MS analysis. Table (c) shows the number of fmoles and the amount (ng) in 10µl of each concentration along with the number of peptides and the percentage coverage of BSA.

determined the detection limit of BSA using the Shevchenko silver staining protocol to be 31fmoles in agreement the results shown here.

For both the CBB G250 and silver stained gels, the region of the gel where BSA runs to (~66kDa) (whether a band was present or not) for each concentration was excised and in-gel digestion was performed. Each sample was analysed by LC MS/MS on a 45min acetonitrile gradient to investigate the number of peptides and the sequence coverage of BSA at each concentration. The raw files were searched against a bovine database and the results were filtered using Xcorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50) and Table 4.1c shows the number of peptides identified and sequence coverage of BSA at each concentration for both stains.

Both stains showed very similar limits of detection, no BSA peptides were identified in samples containing <125fmoles with 1, ~4, and ~7 peptides identified in the 250, 500 and 1000fmoles samples respectively. These results suggest there are no differences in yields when using the in-gel digestion protocol for CBB G250 or silver stained bands. The results also indicate that an intensely stained band on a silver stained gel does not guarantee the protein being identified when analysed by LC MS/MS using the LCQ Deca XP_{PLUS}. However, this is not the case with CBB G250 stained bands. A band on a CBB G250 stained gel stands a good chance of a peptide from that protein being identified with a single peptide identified in the 250fmoles sample. These results suggest that for this study, the CBB G250 stain would be more efficient for a SILAC survey as both samples will be mixed 1:1 prior to 1D gel electrophoresis and the differences will be observed at the LC-MS level rather than by comparison of bands on a gel.

The results in Figure 4.1c also shows a poor recovery of BSA peptides from the in-gel digestion with the detection limit at 250fmoles compared to the digestion in solution results

shown in Table 4.1a where the detection limit was ~4fmoles. This is due to the loss of protein and peptides during the in-gel digestion protocol suggesting that digestion in solution would be preferential to the in-gel digestion protocol due to the significant increase in sample recovery. If the sample is very simple, clean and only contains a single protein (like the samples analysed in Figure 4.1) then digestion in solution would be the method of choice but most samples are complex mixtures of proteins that have been prepared from a biological sample and require a separation step prior to LC MS/MS. These samples also require a clean up step to remove contaminants such as salts and detergents and the proteins also need to be present in buffers compatible with conditions required for trypsin digestion and analysis by MS. Experience within the lab has shown this can be time consuming and each type of experiment requires some optimisation to clean up the sample. Separation of proteins by ID gel electrophoresis can overcome both these problems by cleaning the sample and using a very simple well established protein separation protocol requires no other optimisation other than investigating the amount of starting material required.

4.4 Optimising the Parameters used for Filtering Database Search Results

Analysis of small datasets such as those from experiments to investigate interacting proteins with a target (e.g. immunoprecipitations) is relatively simple. The parameters used to filter the database search results can be lowered to identify proteins present in a test sample compared to an appropriate control. Peptide sequence assignments to MS/MS spectra can be manually inspected to ensure they are correct but the analysis of large datasets is much more complex. Manual inspection of each peptide assignment is time consuming, requires a certain amount of experience and can be subjective. Optimised filter parameters are required to minimise the loss of true peptide and protein identifications and ensuring the number of false

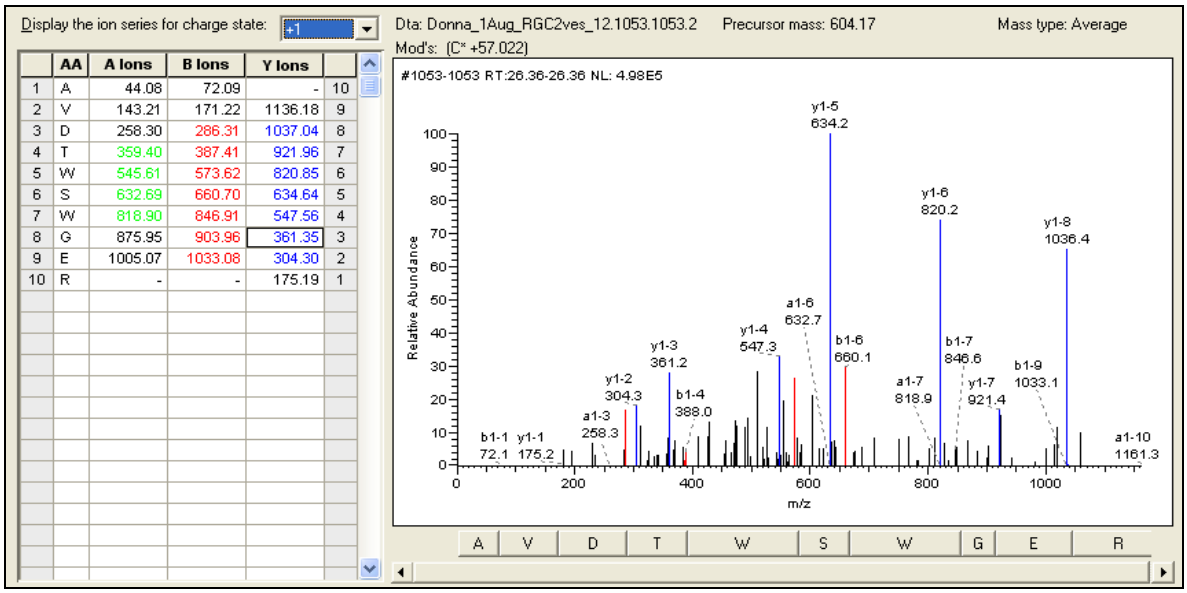
identifications are kept to a minimum. There is no 'gold standard' for which filter parameters should be used.

A database search result from a sample described in Figure 3.7 was used to investigate the optimal filter parameters to be used in this study. Figure 4.2 shows examples of 'good hits' and 'poor hits' when spectra are manually inspected. The process of distinguishing between 'good' and 'poor' hits is subjective but the MS/MS spectra shown here demonstrate clear differences between them, the peaks shown in green, red and blue (representing the a, b and y ions respectively) are those identified in the acquired MS/MS spectrum that matches those in the database. As seen in Figure 4.2a, a 'good hit' identifies a large percentage of the fragment ion peaks from theoretical MS/MS spectra in the database. In particular, the y ions are important as they contain the C-terminal residue which should be an Arg or Lys when digested with trypsin. In the case of 'poor hits', only a small number of peaks in the acquired MS/MS data matches to those in the database for that sequence.

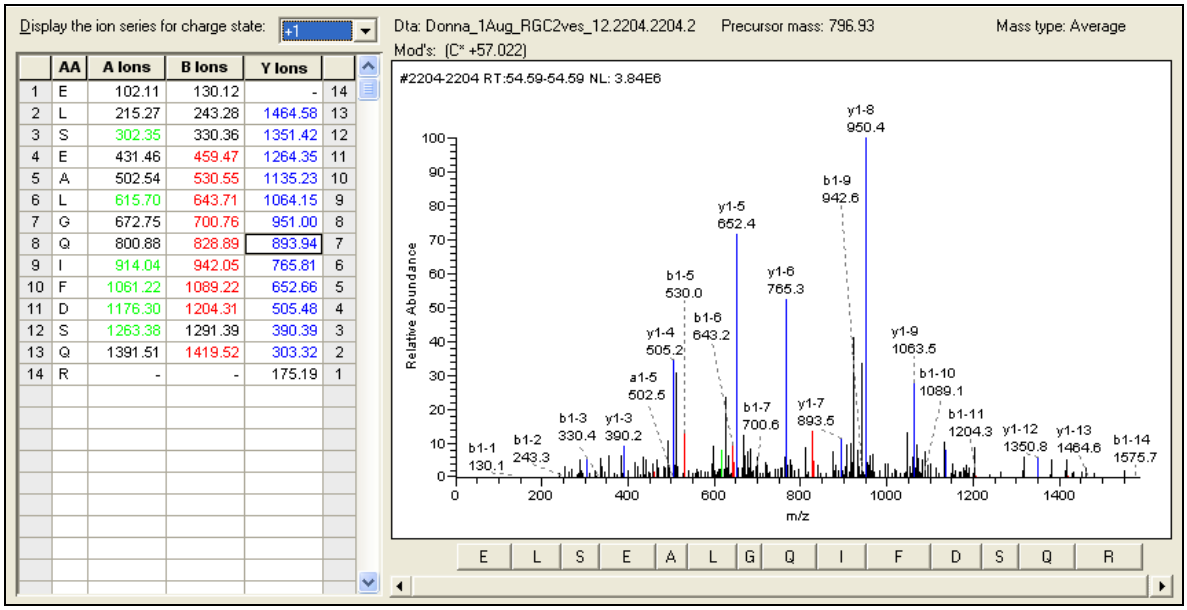
For all experiments described so far, the parameters used to filter the results of a database search, i.e. Xcorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50), are those that were routinely used in many proteomic studies and recommended by the manufacturer (Thermo Electron). Recommendations for filter parameter values that 'constitute a good hit' have been suggested by Novatia (http://www.enovatia.com). Table 4.2 shows the comparison of four different filter parameters (see Figure legend for description of parameters). The comparison of FILTERS I and II identified 31 and 36 proteins respectively. FILTER II identified all 31 proteins identified in FILTER I with an additional 5 proteins not identified using the latter parameters. Manual inspection of the spectra showed that the peptides of 4 of these proteins (in red) were 'good hits' and the peptides identified in the remaining protein were 'poor hits'. These results indicate that FILTER I was too strict but FILTER II was too

(a) gi|483474|emb|CAA55699.1| 90K [Homo sapiens

Reference					Score			Peptides
Scan(s)	Sequence	MH+	Charge	XC	Delta Cn	Sp	RSp	Ions
gi|483474					20.14			2 (2 0 0 0 0)
1053	-.AVDTWSWGER.-	1207.3	2	1.71	0.4	853.5	1	13/18
2204	-.ELSEALGQIFDSQR.-	1593.7	2	2.73	0.35	1068.9	1	17/26



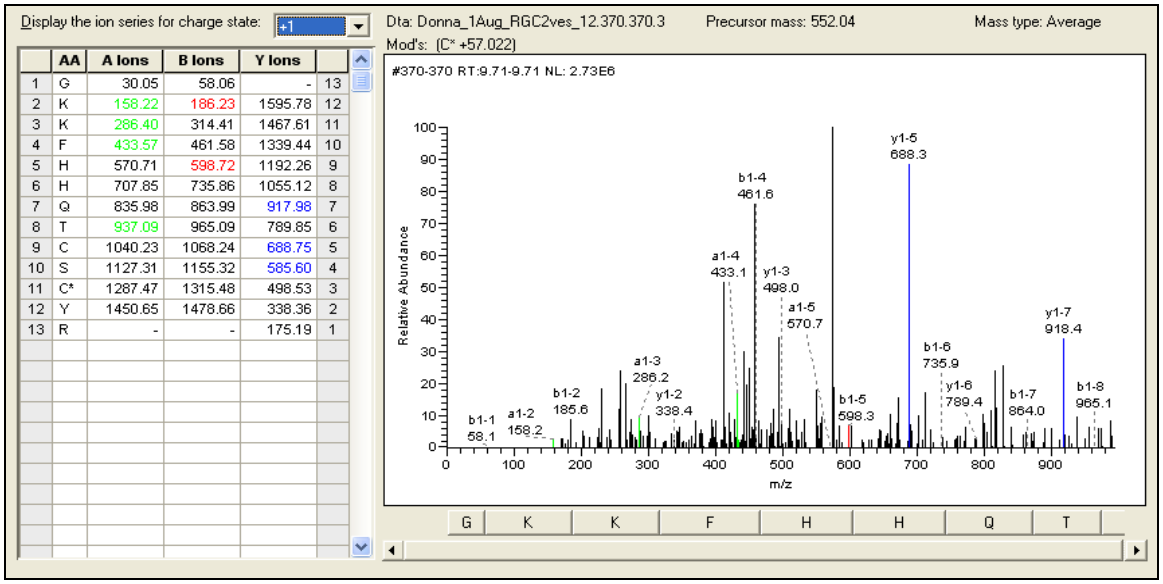
	AA	A Ions	B Ions	Y Ions	
1	A	44.08	72.09	-	10
2	V	143.21	171.22	1136.18	9
3	D	258.30	286.31	1037.04	8
4	T	359.40	387.41	921.96	7
5	W	545.61	573.62	820.85	6
6	S	632.69	660.70	634.64	5
7	W	818.90	846.91	547.56	4
8	G	875.95	903.96	361.35	3
9	E	1005.07	1033.08	304.30	2
10	R	-	-	175.19	1



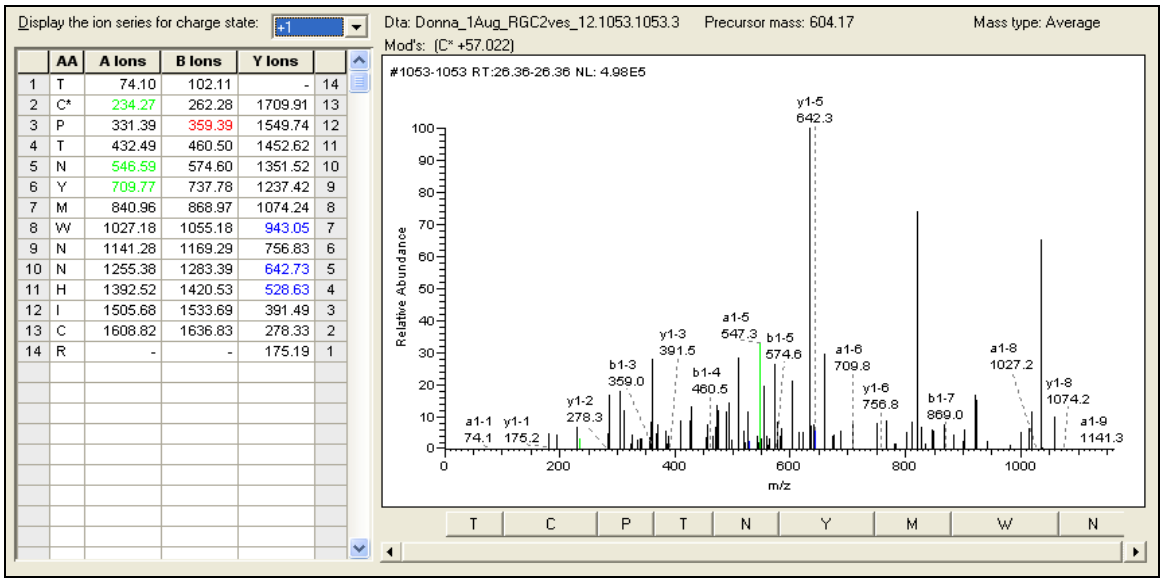
	AA	A Ions	B Ions	Y Ions	
1	E	102.11	130.12	-	14
2	L	215.27	243.28	1464.58	13
3	S	302.35	330.36	1351.42	12
4	E	431.46	459.47	1264.35	11
5	A	502.54	530.55	1135.23	10
6	L	615.70	643.71	1064.15	9
7	G	672.75	700.76	951.00	8
8	Q	800.88	828.89	893.94	7
9	I	914.04	942.05	765.81	6
10	F	1061.22	1089.22	652.66	5
11	D	1176.30	1204.31	505.48	4
12	S	1263.38	1291.39	390.39	3
13	Q	1391.51	1419.52	303.32	2
14	R	-	-	175.19	1

(b) gi|49456749|emb|CAG46695.1| VEGFC [Homo sapiens]

Reference					Score			Peptides
Scan(s)	Sequence	MH+	Charge	XC	Delta Cn	Sp	RSp	Ions
gi|49456749					20.47			2 (2 0 0 0 0)
370	.GKKFHHQTCSC*YR.	1652.9	3	1.64	0.17	337.9	5	15/48
1053	.TC*PTNYMWNNHICR.	1811	3	1.17	0.14	266.6	5	16/52



	AA	A Ions	B Ions	Y Ions	
1	G	30.05	58.06	-	13
2	K	158.22	186.23	1595.78	12
3	K	286.40	314.41	1467.61	11
4	F	433.57	461.58	1339.44	10
5	H	570.71	598.72	1192.26	9
6	H	707.85	735.86	1055.12	8
7	Q	835.98	863.99	917.98	7
8	T	937.09	965.09	789.85	6
9	C	1040.23	1068.24	688.75	5
10	S	1127.31	1155.32	585.60	4
11	C*	1287.47	1315.48	498.53	3
12	Y	1450.65	1478.66	338.36	2
13	R	-	-	175.19	1



	AA	A Ions	B Ions	Y Ions	
1	T	74.10	102.11	-	14
2	C*	234.27	262.28	1709.91	13
3	P	331.39	359.39	1549.74	12
4	T	432.49	460.50	1452.62	11
5	N	546.59	574.60	1351.52	10
6	Y	709.77	737.78	1237.42	9
7	M	840.96	868.97	1074.24	8
8	W	1027.18	1055.18	943.05	7
9	N	1141.28	1169.29	756.83	6
10	N	1255.38	1283.39	642.73	5
11	H	1392.52	1420.53	528.63	4
12	I	1505.68	1533.69	391.49	3
13	C	1608.82	1636.83	278.33	2
14	R	-	-	175.19	1

Figure 4. 2 : Examples of MS/MS spectra that match the theoretical spectra in the database and are considered to be 'good hits' and 'poor hits'.

Figure to show examples of spectra acquired from a gel slice in the MV preparation described in Figure 3.7 that were used to optimize the filter parameters used for data analysis. Fig 4.2(a) shows examples ms/ms data that matches the database very well and are considered to be 'good hits'. Fig 4.2(b) shows examples of ms/ms data that doesn't match the database very well and are considered to be 'poor hits'. The a, b and y ions highlighted in green, red and blue respectively are those ions observed in the acquired ms/ms spectra which matched with those in the theoretical fragmentation pattern for that peptide sequence in the database used. The y– ion series is considered to be the most important ion series used to determine the amino acid sequence as it identifies the C-terminal Arg or Lys present (if digested with trypsin). The tables show the score values used by Bioworks Browser for both the protein and the peptides.

	FILTER I		FILTER II		FILTER III		FILTER IV	
Reference	Score	Peptides	Score	Peptides	Score	Peptides	Score	Peptides
gi|9622850	30.1	3	30.1	3	30.1	3	30.1	3
gi|904214	20.1	2	20.1	2	20.1	2	20.1	2
gi|7328175	20.2	2	20.2	2	20.2	2	20.2	2
gi|7243131	20.1	2	30.1	3	30.1	3	30.1	3
gi|66361423	20.2	2	20.2	2	20.2	2	20.2	2
gi|6470150	200.4	20	258.4	25	246.4	23	258.4	25
gi|6424942	20.4	2	22.4	2	22.4	2	22.4	3
gi|62914009	70.2	7	116.2	11	106.2	10	106.2	10
gi|62898357	20.1	2	20.1	2	20.1	2	20.1	2
gi|62897167	90.2	9	90.2	9	90.2	9	90.2	9
gi|5572747	50.1	5	50.1	5	50.1	5	50.1	5
gi|556514	20.3	2	30.3	3	30.3	3	40.3	4
gi|52545896			20.2	2	20.2	2	20.2	2
gi|49456749			20.5	2	20.5	2		
gi|48762682	30.2	3	30.2	3	20.2	2	30.2	3
gi|48425720	70.4	7	90.4	9	80.4	8	90.4	9
gi|483474	20.3	2	20.1	2	20.1	2	20.1	2
gi|48257068			20.3	2	20.3	2	20.3	2
gi|47115169			20.1	2			20.1	2
gi|40789059	100.2	10	126.2	12	126.2	12	110.2	11
gi|3891633	30.2	3	30.2	3	30.2	3	30.2	3
gi|37674412	70.2	7	70.2	7	80.2	8	80.2	8
gi|3273383	20.4	2	20.4	2			20.4	2
gi|32172416	324.2	31	372.2	37	368.2	36	332.2	33
gi|30923274	110.2	11	130.2	13	130.2	13	130.2	13
gi|29126831	120.2	12	130.2	13	130.2	13	130.2	13
gi|2351110	20.2	2	20.2	2	20.2	2	20.2	2
gi|16878197	30.2	3	38.2	3	30.2	3	28.2	2
gi|14250593	50.2	5	70.2	7	60.2	6	70.2	7
gi|136378	20.2	2	20.2	2	20.2	2	20.2	2
gi|12653989			20.3	2	20.3	2	20.3	2
gi|12225000	30.2	3	40.2	4	40.2	4	40.2	4
gi|11414998	20.1	2	20.1	2	20.1	2	20.1	2
gi|10716563	30.2	3	30.2	3	30.2	3	30.2	3
gi|105155	30.2	3	30.2	3	30.2	3	30.2	3
gi|10437873	20.1	2	20.1	2	20.1	2	20.1	2

Table 4. 2 : Various filter parameters were changed to optimise those suitable for the analysis of large datasets using Bioworks Browser v3.1.

Table to show which filter parameters were modified to optimise those to be used for data analysis of large datasets. Data collected from one of the samples from the vesicle collection outlined in Figure 3.7 was used for the optimisation. Only proteins with 2 first hit peptides were included in the analysis. The proteins in blue writing had peptides that looked like 'poor hits' and the proteins in red writing had peptides that looked like 'good hits'. FILTER I – Single Threshold Xcorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50) 31 proteins identified; FILTER II – Default Multiple Threshold (Match = 4, Delta Cn = 0.085, Xcorr = 1.00, Sp = 500, RSp = 5, Percent ions = 30%) 36 proteins identified; FILTER III - Novatia Multiple Threshold (Match = 4, Delta Cn = 0.100, Xcorr = 1.50, Sp = 200, RSp = 5, Percent ions = 70%) 34 proteins identified; FILTER IV – Modified Multiple Threshold (Match = 4, Delta Cn = 0.100, Xcorr = 1.00, Sp = 350, RSp = 5, Percent ions = 50%) 35 proteins. The proteins found using all 4 filter parameters are highlighted in bold. The proteins highlighted in green are proteins identified with additional first hit peptides identified in proteins using FILTER IV compared to FILTER I.

relaxed. Next, FILTERS II and III were compared and 36 and 34 proteins identified respectively. FILTER II identified all 34 proteins identified using FILTER III with an additional 2 proteins, manual inspection confirmed that the peptides identified in both of these proteins were 'good hits' indicating that FILTER III is too strict. FILTER II was then modified to create a set of filter parameters that would be a compromise between FILTERS II and III.

The results were then analysed using FILTER IV and compared to FILTERS I, II and III with 35, 31, 36 and 34 proteins identified respectively. The 35 proteins identified using FILTER IV all the proteins with 'good hit' peptides from the other three parameters and didn't identify the 1 protein with peptides that were 'poor hits' when inspected manually. The proteins highlighted in green in Table 4.2 also identified additional peptides using FILTER IV in proteins identified using FILTER I. These results indicate that the modified parameters set in FILTER IV were optimal for analysing the large datasets that would be generated by a proteomic profile of MVs, i.e. generate very low number of false positive identifications whilst keeping the loss of true identifications to a minimum.

4.5 Source of Database

In a review by Nesvizhskii and Aebersold, the benefits and drawbacks of using databases from several sources were discussed (Nesvizhskii and Aebersold, 2005). As discussed in Section 1.3, the correct peptide sequence assignment to MS/MS spectra is limited to those found within the database therefore the experiment should determine the database required. The comparisons in their review suggested that a human IPI or a human database generated from the NCBI nr database would be most suitable for the large dataset that would be generated in the quantitative shotgun proteomic approach planned for this study. The search results from the previous chapter identified numerous proteins entries with vague names (e.g.

unknown, hypothetical protein) which require further manual investigation to decipher the actual protein present.

The NCBI database also contains a large number of non-redundant protein entries, i.e. many entries from the same protein entered by different research groups and proteins with small amino acid sequence changes making it very difficult to determine the exact protein present. Deciphering the identity of each unknown protein or whether the wild type or a variant is present in the sample would be time consuming when analysing large datasets.

In an attempt to overcome these problems, a raw data file generated in Figure 3.7 that illustrated the problems discussed above was used to compare search results using an NCBI and an IPI database. Table 4.3 shows that there were 43 proteins (with at least 2 first 'hit' peptides) identified from the search using the human IPIv3.14 database and compared to the 31 proteins identified in the NCBI database. There were 171 first hit peptides identified in the 31 proteins identified in the NCBI search and 306 first hit peptides in the 43 proteins identified in the IPI search. The search results also identified 45 and 76 proteins in the NCBI and IPI search results respectively with only a single peptide assignment. The results show that more protein assignments were made in the IPI search results compared to the NCBI results with all peptides in 'multiple peptide hit' proteins from the NCBI results represented in the IPI results. Table 4.3 also shows that some peptides identified in 'single peptide hit' proteins in the NCBI search results corresponded to multiple peptide hit proteins in the IPI search results, partly accounting for the increase in protein identifications found using the IPI database. The IPI results show that peptides were not assigned to proteins with unspecified names such as hypothetical or unknown proteins as observed in the NCBI results (proteins in bold) and these peptides were assigned to protein entries that doesn't require further investigation or BLAST searches to determine the protein they may have originated from.

	NCBI acc	Protein	no of pep	IPI acc	Protein	no of pep
1	32172416.0	MYO1D_HUMAN Myosin Id (all but 1 peptide found in ipi00329719)	31	IPI00329719	ISOFORM 1 OF MYOSIN-ID.	48
2	6470150.0	BiP protein [Homo sapiens] (all 20 peptides assigned to ipi00003362)	20	IPI00003362	HSPA5 PROTEIN.	31
				IPI00258514	HSPA5 PROTEIN	5
3	29126831.0	OLFM4 protein [Homo sapiens]	12	IPI00022255	OLFACTOMEDIN-4.	15
4	30923274.0	GFPT1_HUMAN Glucosamine--fructose-6-phosphate aminotran (8 peptides found in ipi00217952 and 3 peptides found in ipi00216159)	11	IPI00217952	ISOFORM 1 OF GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE [ISOMERIZING] 1.	18
				IPI00216159	GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE [ISOMERIZING] 2.	5
5	40789059.0	**KIAA0051 [Homo sapiens] all 10 peptides found in IPI00009342**	10	IPI00009342	RAS GTPASE-ACTIVATING-LIKE PROTEIN IQGAP1.	18
6	62897167.0	threonyl-tRNA synthetase variant [Homo sapiens]	9	IPI00329633	THREONYL-TRNA SYNTHETASE, CYTOPLASMIC.	9
7	48425720.0	B Chain B, Structure Of Human Transferrin Receptor-	7	IPI00022462	TRANSFERRIN RECEPTOR PROTEIN 1.	12
	136378.0	TFR1_HUMAN Transferrin receptor protein 1 (TfR1)	2			
8	62914009.0	HSPCA protein [Homo sapiens] 5 pep found in 19100031523 and 2 pep in ipi00334775	7	IPI00334775	HEAT SHOCK PROTEIN HSP-90 BETA	13
	2351110.0	heat shock protein beta [Homo sapiens] (both pep found ipi00334775)	2	IPI00031523	PUTATIVE HEAT SHOCK PROTEIN HSP 90-ALPHA A2.	6
	10437873.0	**unnamed protein product [Homo sapiens]**	2			
9	37674412.0	**unknown [Homo sapiens]**	7	IPI00009904	PROTEIN DISULFIDE-ISOMERASE A4.	12
10	14250593.0	Calpain 1, large subunit [Homo sapiens]	5	IPI00011285	CALPAIN-1 CATALYTIC SUBUNIT.	9
11	5572747.0	peptidylarginine deiminase type II [Homo sapien	5	IPI00294187	PROTEIN-ARGININE DEIMINASE TYPE-2.	8

	NCBI acc	Protein	no of pep		Protein	no of pep
12	3891633.0	A Chain A, Hsp90 N-Terminal Domain Bound To Adp-Mg (2 pep found in ipi00382470 and 1 pep found in ipi00031523)	3	IPI00382470	HEAT SHOCK PROTEIN 90KDA ALPHA (CYTOSOLIC), CLASS A MEMBER 1 ISOFORM 1.	4
13	105155.0	A33639 6-phosphofructokinase (EC 2.7.1.11), hepatic - human (2 pep found in ipi00220617 and 1 pep found in ipi00009790)	3	IPI00220617 (EC 2.7.1.11)	6-PHOSPHOFRUCTOKINASE, LIVER TYPE	6
				IPI00009790	6-PHOSPHOFRUCTOKINASE TYPE C.	2
14	10716563.0	calnexin precursor [Homo sapiens]	3	IPI00020984	CDNA FLJ55574, HIGHLY SIMILAR TO CALNEXIN.	8
15	12225000.0	**hypothetical protein [Homo sapiens]**	3	IPI00018931	VACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN 35.	12
	9622850.0	vacuolar sorting protein 35 [Homo sapiens]	3			
16	48762682.0	Na+/K+ -ATPase alpha 1 subunit isoform b p	3	IPI00006482	ISOFORM LONG OF SODIUM/POTASSIUM-TRANSPORTING ATPASE SUBUNIT ALPHA-1.	5
17	16878197.0	VIL1 protein [Homo sapiens]	3	IPI00218852	VILLIN-1.	5
	62898357.0	villin 1 variant [Homo sapiens]	2			
18	556514.0	acylamino acid-releasing enzyme [Homo sapiens]	2	IPI00337741	ACYLAMINO-ACID-RELEASING ENZYME.	6
19	6424942.0	ALG-2 interacting protein 1 [Homo sapiens]	2	IPI00246058	PDCD6IP PROTEIN.	6
	7243131.0	**KIAA1375 protein [Homo sapiens] 1 pepfound inipi00246058 and the other not identified at all**	2			
20	66361423.0	F Chain F, Crystal Structure Of Human 17-Beta-Hydroxy	2	IPI00019912	PEROXISOMAL MULTIFUNCTIONAL ENZYME TYPE 2.	2
21	48257068.0	HSPA8 protein [Homo sapiens]	2	IPI00003865	ISOFORM 1 OF HEAT SHOCK COGNATE 71 KDA PROTEIN.	3
22	7328175.0	**hypothetical protein [Homo sapiens] (1 pep found in ipi00479359 and the other peptide not found in any**	2	IPI00479359	EZRIN	2
23	11414998.0	NADPH-cytochrome P-450 reductase [Homo sapiens]	2	IPI00470467	NADPH--CYTOCHROME P450 REDUCTASE.	3
24	904214.0	prolyl endopeptidase [Homo sapiens]	2	IPI00008164	PROLYL ENDOPEPTIDASE.	4

	NCBI acc	Protein	no of pep		Protein	no of pep
25	3273383.0	TRAP1 [Homo sapiens]	2	IPI00030275	HEAT SHOCK PROTEIN 75 KDA, MITOCHONDRIAL.	2
26				IPI00017855	ACONITATE HYDRATASE, MITOCHONDRIAL.	6
27				IPI00020672	ISOFORM 1 OF DIPEPTIDYL-PEPTIDASE 3.	5
28	2135488.0	A41543 integrin alpha-6 chain precursor, splice form B - h	1	IPI00010697	ISOFORM ALPHA-6X1X2B OF INTEGRIN ALPHA-6.	5
29				IPI00005614	ISOFORM LONG OF SPECTRIN BETA CHAIN, BRAIN 1.	5
30	20143967.0	kinesin family member 23 isoform 1	1	IPI00291579	ISOFORM 1 OF KINESIN-LIKE PROTEIN	4
31				IPI00289334	ISOFORM 1 OF FILAMIN-B.	2
32	975290.0	phosphofructokinase	1	IPI00219585	ISOFORM 2 OF 6-PHOSPHOFRUCTOKINASE, MUSCLE TYPE.	3
33				IPI00010418	MYOSIN-IC.	2
34				IPI00645078	UBIQUITIN-LIKE MODIFIER-ACTIVATING ENZYME 1.	3
35				IPI00021439	ACTIN, CYTOPLASMIC 1.	2
36	42560541.0	CD97_HUMAN CD97 antigen precursor	1	IPI00299412	ISOFORM 2 OF CD97 ANTIGEN.	2
37	1304314.0	pyrroline 5-carboxylate synthetase	1	IPI00008982	ISOFORM LONG OF DELTA-1-PYRROLINE-5-CARBOXYLATE SYNTHETASE.	2
38				IPI00220644	ISOFORM M1 OF PYRUVATE KINASE ISOZYMES M1/M2.	2
39				IPI00003021	SODIUM/POTASSIUM-TRANSPORTING ATPASE SUBUNIT α-2.	2

Table 4. 3 : Which Database? IPI vs NCBI.

Table to show the results of database searching using a human IPI database (IPIv3.14) and a human database generated from an NCBI nr database and filtered using Xcorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50). The table lists all the proteins identified in the NCBI search results with at least 2 'first hit' peptides alongside the equivalent protein identification in the IPI search results. Those highlighted in red are proteins with similar names to other proteins identified in the same sample. All peptide sequences from these proteins were checked to determine whether they could have originated from the same protein. The names in bold are entries in the NCBI database search that would require further investigation to determine the protein identification (Sample taken from Figure 3.7).

One problem highlighted in the NCBI search results was the identification of multiple isoforms of the same protein in the same sample e.g. 2 Transferrin proteins gi|48425720 and gi|136378 highlighted in red in Table 4.3. The peptides identified in both of theses proteins were manually inspected, Figure 4.3 shows the peptides assigned to these proteins and their correlation scores (Figure 4.3a) and the location of the identified peptides in the full protein sequences (Figure 4.3b).

The peptide sequences highlighted in green and orange correspond to the peptides identified in gi|48425720 and gi|136378 respectively. The peptide sequence LTHDVELNLDYER, underlined and highlighted in orange in Figure 4.3b, was identified in gi|136378. A near identical sequence could also be found in gi|48425720 i.e. LTHDVELNLDYEE (underlined and highlighted in grey) with only a difference in the C-terminal residue. As all 7 peptide sequences could be found in gi|136378, this could suggest that only this protein is present in the sample but further investigation would be required to determine whether this is true. Database searching using the IPI database does not appear to completely solve this problem as there are still peptides identified that could belong in more than one protein entry but all the proteins identified had names that didn't require further analysis to determine the identity of that protein. Therefore, the IPI database would be more suitable for the analysis of large datasets as it requires less manual interpretation.

4.6 False-Positive Rate Determination

Large datasets generated during shotgun proteomic approaches require additional statistical analyses to give some indication that the correct peptide and protein sequence has been assigned to MS/MS spectra. One such analysis is the calculation of the false-positive rate or False Discovery Rate (FDR) determination as described by Elias *et al* (Elias et al., 2005).

(a)

Reference					Score			Peptides (Hits)
Scan(s)	Sequence	MH+	Charge	XC	Delta Cn	Sp	RSp	Ions
gi|48425720					70.38			7 (7 0 0 0 0)
1040 - 1048	-.SGVGTALLLK.-	959.17	2	2.54	0.25	668.5	1	15/18
1113	-.AFTYINLDK.-	1085.24	2	2.09	0.32	496.7	2	11/16
1324 - 1328	-.SSGLPNIPVQTISR.-	1469.67	2	2.59	0.10	367.8	1	14/26
1354 - 1360	-.SSGLPNIPVQTISR.-	1469.67	2	2.01	0.18	516.4	1	15/26
1474 - 1484	-.VEYHFLSPYVSPK.-	1566.78	2	3.46	0.43	641.5	1	15/24
2426 - 2432	-.VSASPLLYTLIEK.-	1434.70	2	3.70	0.53	1035.6	1	19/24
2564	-.ILNIFGVIK.-	1017.29	2	3.20	0.24	728.1	1	15/16
gi|136378					20.15			2 (2 0 0 0 0)
818	-.LTHDVELNLDYER.-	1617.74	2	3.08	0.36	1880.4	1	19/24
1895 - 1901	-.YNSQLLSFVR.-	1227.40	2	2.13	0.32	1094.1	1	15/18

(b)

>gi|48425720|pdb|1SUV|B Chain B, Structure Of Human Transferrin Receptor-Transferrin Complex 639 aa

LYWDDLKRKLSEKLDSTDFTSTIKLLNENSYVPREAGSQKDENLALYVENEFREFKLSKVWRDQHFVKIQ
VKDSAQNSVIIVDKNGRLVYLVENPGGYVAYSKAATVTGKLVHANFGTKKDFEDLYTPVNGSIVIVRAGK
ITFAEKVANAESLNAIGVLIYMDQTKFPIVNAELSFFGHAHLGTGDPYTPGFPSFNHTQFPPSRSSGLPN
IPVQTISRAAAEKLFGNMEGDCPSDWKTDSTCRMVTSESKNVKLTVSNVLKEIKILNIFGVIKGFVEPDH
YVVVGAQRDAWGPGAAKSGVGTALLLKLAQMFSDMVLKDGFQPSRSIIFASWSAGDFGSVGATEWLEGYL
SSLHLKAFTYINLDKAVLGTSNFKVSASPLLYTLIEKTMQNVKHPVTGQFLYQDSNWASKVEKLTLDNAA
FPFLAYSGIPAVSFCFCEDTDYPYLGTTMDTYKELIERIPELNKVARAAAEVAGQFVIKLTHDVELNLDY
EEYNSQLLSFVRDLNQYRADIKEMGLSLQWLYSARGDFFRATSRLTTDFGNAEKTDRFVMKKLNDRVMRV
EYHFLSPYVSPKESPFRHVFWGSGSHTLPALLENLKLRKQNNGAFNETLFRNQLALATWTIQGAANALSG
DVWDIDNEF

>gi|136378|sp|P02786|TFR1_HUMAN Transferrin receptor protein 1 (TfR1) (TR) (TfR) (Trfr) (CD71 antigen) (T9) (p90) 760aa

MMDQARSAFSNLFGGEPLSYTRFSLARQVDGDNSHVEMKLAVDEEENADNNTKANVTKPKRCSGSICYGT
IAVIVFFLIGFMIGYLGYCKGVEPKTECERLAGTESPVREEPGEDFPAARRLYWDDLKRKLSEKLDSTDF
TSTIKLLNENSYVPREAGSQKDENLALYVENQFREFKLSKVWRDQHFVKIQVKDSAQNSVIIVDKNGRLV
YLVENPGGYVAYSKAATVTGKLVHANFGTKKDFEDLYTPVNGSIVIVRAGKITFAEKVANAESLNAIGVL
IYMDQTKFPIVNAELSFFGHAHLGTGDPYTPGFPSFNHTQFPPSRSSGLPNIPVQTISRAAAEKLFGNME
GDCPSDWKTDSTCRMVTSESKNVKLTVSNVLKEIKILNIFGVIKGFVEPDHYVVVGAQRDAWGPGAAKSG
VGTALLLKLAQMFSDMVLKDGFQPSRSIIFASWSAGDFGSVGATEWLEGYLSSLHLKAFTYINLDKAVLG
TSNFKVSASPLLYTLIEKTMQNVKHPVTGQFLYQDSNWASKVEKLTLDNAAFPFLAYSGIPAVSFCFCED
TDYPYLGTTMDTYKELIERIPELNKVARAAAEVAGQFVIKLTHDVELNLDYERYNSQLLSFVRDLNQYRA
DIKEMGLSLQWLYSARGDFFRATSRLTTDFGNAEKTDRFVMKKLNDRVMRVEYHFLSPYVSPKESPFRHV
FWGSGSHTLPALLENLKLRKQNNGAFNETLFRNQLALATWTIQGAANALSGDVWDIDNEF

Figure 4. 3 : Manual Inspection of the peptides from the proteins gi|48425720|pdb|1SUV|B Chain B, Structure Of Human Transferrin Receptor-Transferrin Complex and gi|136378|sp|P02786|TFR1_HUMAN Transferrin receptor protein 1 (TfR1) identified in the NCBI database search results.

(a) Table to show the peptides assigned to each protein with their corresponding correlation scores. (b) Shows the location of the identified peptide sequences in the full length protein sequence. The peptide sequences highlighted in green and orange correspond to the peptides identified in gi|48425720 and gi|136378 respectively. The two underlined peptide sequences (one in grey and the other in orange) are identical apart from the C-terminal residue (in bold).

This is achieved by searching the raw files against a database that contains randomised sequences of all the proteins present in that database as well as the non-randomised sequences. This is known as target-decoy database searching and is described in Section 1.3.3). In this study, this method was tested on a single raw file acquired in Figure 3.7. The protein sequences in the human IPIv3.14 database were randomised using a perl script downloaded from the matrix science website (www.matrixscience.com/help/decoy) and the raw file was searched against this database (containing the correct and random protein sequences). The results were filtered using the four different filter parameters shown in Table 4.2 and the results are shown in Table 4.4. The results show the number of MS/MS spectra assigned to random protein sequences and the percentage of the total protein and peptide assignments made to random entries for proteins with at least 2 'first hit' peptides and proteins with at least 1 'first hit' peptide. A 0% false positive rate was calculated for all 4 filter parameters at the peptide and protein level when only proteins with at least 2 'first hit' peptides were considered. When proteins with at least 1 'first hit' peptide were also considered, the false positive rate increased to 43, 48, 54 and 46% for filters I, II, III and IV respectively at the protein level and 12, 23, 19 and 20% respectively at the peptide level. The results demonstrate that quite a large margin of error is introduced when proteins with only 1 'first hit' peptide is included in the analysis but failed to show any differences in false positive rates between the filter parameters used.

Next, all 16 raw files generated from Figure 3.7 were searched against the same randomised IPI database and the false positive rates were calculated as a percentage and the FDR (Elias et al., 2005) for each of the 4 filter parameters (see Table 4.5). Firstly, Table 4.5 shows that calculating the false positive rate as a percentage or the FDR yields the same result if the percentage is represented as a decimal number. Table 4.5 also shows that calculating the

Filter Parameters	Proteins with at least 2 first hit peptides				Proteins with at least 1 first hit peptides			
	no of proteins	no of peptides	% of false positives		no of proteins	no of peptides	% of false positives	
	IPI / random	IPI / random	PROTEIN	PEPTIDE	IPI / random	IPI / random	PROTEIN	PEPTIDE
Filter I	43 / 0	320 / 0	0%	0%	105 / 46	382 / 46	43.81	12.04
Filter II	49 / 0	345 / 0	0%	0%	282 / 137	578 / 137	48.58	23.70
Filter III	49 / 0	341 / 0	0%	0%	194 / 106	535 / 106	54.64	19.8
Filter IV	49 / 0	342 / 0	0%	0%	228 / 106	521 / 106	46.49	20.35

Table 4. 4 : False positive determination of a single sample following database searching using a decoy database.

A raw file collected in Figure 3.7 was searched against the randomised IPIv3.14 database to determine the false positive rate using the four previously discussed filter parameters. The Table shows the number of 'true' proteins and the 'false' or random proteins identified. The false positive rate is expressed as the percentage of the total protein identifications that were random entries i.e. (number of random hits / number of true hits) x 100. FILTER I –Single Threshold Xcorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50); FILTER II – Default Multiple Threshold (Match = 4, Delta Cn = 0.085, Xcorr = 1.00, Sp = 500, RSp = 5, Percent ions = 30%); FILTER III - Novatia Multiple Threshold (Match = 4, Delta Cn = 0.100, Xcorr = 1.50, Sp = 200, RSp = 5, Percent ions = 70%); FILTER IV – Modified Multiple Threshold (Match = 4, Delta Cn = 0.100, Xcorr = 1.00, Sp = 350, RSp = 5, Percent ions = 50%).

false positive rate using the decoy database method is useful for large datasets but not single gel bands as random proteins were identified in a relatively small number of samples in all 4 filter parameters. One way of identifying which filter parameters are useful for a particular dataset is to adjust them to yield a certain false positive rate.

The results in Table 4.5 show an FDR of 0.0081, 0.0132, 0.0058 and 0.0086 at the protein level and 0.0036, 0.0046, 0.0019 and 0.0032 at the peptide level for filters I, II, III and IV respectively. The lowest FDR was achieved using filter III but as shown in Table 4.2, many 'good hits' were missed using this parameter. As previously discussed, the results in Table 4.2 indicated that Filter IV was the optimal set of filter parameters for analysing proteomic datasets. Using these filter parameters on multiple samples from a single experiment generated an FDR of 0.0086 (false positive rate <1%) at the protein level which was comparable to those previously used on a routine basis in the lab (filter I FDR = 0.0081). Using these modified multiple threshold filter parameters, more protein and peptide hits were generated compared to Filter I without dramatically increasing the false positive rate. These results therefore confirm that the modified filter parameters are optimal for analysing a large proteomic dataset.

(a)

Sample no	FILTER I - Proteins with at least 2 first hit peptides							
	NUMBER OF PROTEINS		NUMBER OF PEPTIDES		% of false positives		FDR	
	IPI	random	IPI	random	PROT	PEP	PROT	PEP
1	9	0	100	0	0.00	0.00	0.0000	0.0000
2	33	0	323	0	0.00	0.00	0.0000	0.0000
3	49	1	364	2	2.04	0.55	0.0200	0.0055
4	36	0	411	0	0.00	0.00	0.0000	0.0000
5	49	0	354	0	0.00	0.00	0.0000	0.0000
6	45	4	276	8	8.89	2.90	0.0816	0.0282
7	44	1	294	2	2.27	0.68	0.0222	0.0068
8	53	0	337	0	0.00	0.00	0.0000	0.0000
9	51	0	268	0	0.00	0.00	0.0000	0.0000
10	53	0	337	0	0.00	0.00	0.0000	0.0000
11	84	0	404	0	0.00	0.00	0.0000	0.0000
12	81	0	306	0	0.00	0.00	0.0000	0.0000
13	76	0	325	0	0.00	0.00	0.0000	0.0000
14	95	1	221	5	1.05	2.26	0.0104	0.0221
15	53	0	215	0	0.00	0.00	0.0000	0.0000
16	42	0	159	0	0.00	0.00	0.0000	0.0000
total	853	7	4694	17	0.82	0.36	0.0081	0.0036

(b)

Sample no	FILTER II - Proteins with at least 2 first hit peptides							
	NUMBER OF PROTEINS		NUMBER OF PEPTIDES		% of false positives		FDR	
	IPI	random	IPI	random	PROT	PEP	PROT	PEP
1	14	0	121	0	0.00	0.00	0.0000	0.0000
2	38	0	383	0	0.00	0.00	0.0000	0.0000
3	56	1	432	2	1.79	0.46	0.0175	0.0046
4	43	0	486	0	0.00	0.00	0.0000	0.0000
5	59	0	426	0	0.00	0.00	0.0000	0.0000
6	47	2	380	4	4.26	1.05	0.0408	0.0104
7	45	2	376	4	4.44	1.06	0.0426	0.0105
8	56	0	470	0	0.00	0.00	0.0000	0.0000
9	51	0	292	0	0.00	0.00	0.0000	0.0000
10	57	1	362	2	1.75	0.55	0.0172	0.0055
11	93	0	464	0	0.00	0.00	0.0000	0.0000
12	90	0	360	0	0.00	0.00	0.0000	0.0000
13	85	1	369	2	1.18	0.54	0.0116	0.0054
14	53	2	258	6	3.77	2.33	0.0364	0.0227
15	60	1	251	2	1.67	0.80	0.0164	0.0079
16	49	2	185	4	4.08	2.16	0.0392	0.0212
total	896	12	5615	26	1.34	0.46	0.0132	0.0046

(c)

Sample no	FILTER III - Proteins with at least 2 first hit peptides							
	NUMBER OF PROTEINS		NUMBER OF PEPTIDES		% of false positives		FDR	
	IPI	random	IPI	random	PROT	PEP	PROT	PEP
1	12	0	112	0	0.00	0.00	0.0000	0.0000
2	36	0	375	0	0.00	0.00	0.0000	0.0000
3	54	0	419	0	0.00	0.00	0.0000	0.0000
4	40	0	465	0	0.00	0.00	0.0000	0.0000
5	53	0	399	0	0.00	0.00	0.0000	0.0000
6	46	1	288	2	2.17	0.69	0.0213	0.0069
7	42	0	314	0	0.00	0.00	0.0000	0.0000
8	54	0	359	0	0.00	0.00	0.0000	0.0000
9	49	1	268	2	2.04	0.75	0.0200	0.0074
10	55	1	353	2	1.82	0.57	0.0179	0.0056
11	89	0	437	0	0.00	0.00	0.0000	0.0000
12	87	0	345	0	0.00	0.00	0.0000	0.0000
13	85	1	362	2	1.18	0.55	0.0116	0.0055
14	49	0	239	0	0.00	0.00	0.0000	0.0000
15	59	0	240	0	0.00	0.00	0.0000	0.0000
16	49	1	180	2	2.04	1.11	0.0200	0.0110
total	859	5	5155	10	0.58	0.19	0.0058	0.0019

(d)

Sample no	FILTER IV- Proteins with at least 2 first hit peptides							
	NUMBER OF PROTEINS		NUMBER OF PEPTIDES		% of false positives		FDR	
	IPI	random	IPI	random	PROT	PEP	PROT	PEP
1	14	0	116	0	0.00	0.00	0.0000	0.0000
2	38	0	381	0	0.00	0.00	0.0000	0.0000
3	58	2	433	2	3.45	0.46	0.0333	0.0046
4	41	0	474	0	0.00	0.00	0.0000	0.0000
5	56	0	398	0	0.00	0.00	0.0000	0.0000
6	48	1	298	2	2.08	0.67	0.0204	0.0067
7	42	0	318	0	0.00	0.00	0.0000	0.0000
8	56	0	371	0	0.00	0.00	0.0000	0.0000
9	51	1	281	2	1.96	0.71	0.0192	0.0071
10	56	1	351	2	1.79	0.57	0.0175	0.0057
11	88	0	447	0	0.00	0.00	0.0000	0.0000
12	129	0	364	0	0.00	0.00	0.0000	0.0000
13	84	1	363	2	1.19	0.55	0.0118	0.0055
14	52	1	252	4	1.92	1.59	0.0189	0.0156
15	59	0	242	0	0.00	0.00	0.0000	0.0000
16	49	1	184	3	2.04	1.63	0.0200	0.0160
total	921	8	5273	17	0.87	0.32	0.0086	0.0032

Table 4. 5 : False positive rate determination of all samples in a single experiment following database searching using a decoy database.

All the raw files collected in Figure 3.7 were searched against an IPI_RANDOM database to determine the false positive rate using the four previously discussed filter parameters. The Table shows the number of 'true' proteins and the 'false' or random proteins identified. The false positive rate is expressed as both the percentage of 'false' hits (as described in Table 4.4) and as FDR. FILTER I –Single Threshold Xcorr vs charge state (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50); FILTER II – Default Multiple Threshold (Match = 4, Delta Cn = 0.085, Xcorr = 1.00, Sp = 500, RSp = 5, Percent ions = 30%); FILTER III - Novatia Multiple Threshold (Match = 4, Delta Cn = 0.100, Xcorr = 1.50, Sp = 200, RSp = 5, Percent ions = 70%); FILTER IV – Modified Multiple Threshold (Match = 4, Delta Cn = 0.100, Xcorr = 1.00, Sp = 350, RSp = 5, Percent ions = 50%). FDR = False Discovery Rate (FP/(FP+TP)). FP = False positive. TP = True Positive.

4.6 Conclusions

The aim of this chapter was to analyse the sensitivity of the LCQ Deca XP_{PLUS} and to optimise the search conditions and filter parameters to be used for the analysis of large proteomic datasets. Firstly, the instrument sensitivity was tested by digesting BSA with trypsin in solution and investigating the number of BSA peptides identified and percentage coverage with increasing concentrations. Table 4.1 indicated the sensitivity of the LCQ Deca XP_{PLUS} for BSA as ~4fmoles (0.26ng). This sensitivity was decreased when the BSA was digested with trypsin in gel slices following protein separation by electrophoresis with peptides identified by MS in bands >250fmoles (16.5ng), these results were the same for both CBB G250 and silver stained gels.

The sensitivity of both the CBB G250 and silver staining protocols was also investigated. Although the silver staining protocol was more sensitive, detection of a band on a gel did not guarantee protein identification when LC MS/MS analysis was performed and CBB G250 staining was chosen for the SILAC survey to be carried out as a high level of sensitivity at the protein staining level is not required. Also, there is a loss of sample recovery when performing in-gel digestion compared to digestion in solution however when analyzing complex mixtures of proteins, separation using gel electrophoresis is still favourable due to the added benefits of contaminant removal prior LC MS/MS analysis.

Next, the choice of database used for searching the raw files against was investigated. Table 4.3 indicated that the IPI database would be more suitable for analysing large datasets generated from shotgun proteomic experiments compared to a human NCBI database as less manual interpretation was required. Finally, the optimal filter parameters for analysing the search results were investigated. The results in Table 4.2 indicated that the modified multiple threshold filter parameters (Match = 4, Delta Cn = 0.100, Xcorr = 1.00, Sp = 350, RSp = 5,

Percent ions = 50%) was optimal to keep the number of false positive identifications to a minimum whilst ensuring that few true identifications were missed. Xcorr is a score resulting from a cross correlation analysis of the top 500 identified amino acid sequences and is a measure of the similarity of the obtained spectra with the theoretical spectra. Delta cn is a measure of the difference in Xcorr values for the top candidate and the given candidate. The Sp value is a preliminary score given to a predicted amino acid sequence which takes into account the number of predicted fragment ions that match the observed ions and their abundance. The RSp is the ranking of the Sp score among the candidate amino acid sequences. The percent ions parameter is the percentage of the fragment ions of the candidate sequence that matches the total number of the theoretical fragment ions. The Match parameter stipulates that each candidate must match at least four of the threshold parameters. This was confirmed in the results of the decoy database search results shown in Table 4.5. These filter parameters identified more proteins and peptides when compared to the Xcorr vs charge state filter parameters previously used on a routine basis in our lab (+/-1: 1.50, +/-2: 2.00, +/-3: 2.50) without increasing the false positive rate (<1%). The conditions optimised in this chapter were then ready to analyse the large proteomic dataset that would be generated by the SILAC analysis of MVs from the RG/C2 and PRG/C2 cell lines.

Chapter 5

RESULTS

Quantitative analysis of proteins secreted in MV's from CM derived from colorectal cancer cell lines using Stable Isotope Labelling of Amino acids in Culture (SILAC)

Chapter 5 : Quantitative analysis of proteins secreted in MV's from CM derived from colorectal cancer cell lines using Stable Isotope Labeling of Amino acids in Culture (SILAC)

5.1 Introduction

SILAC is a quantitative proteomic method used to investigate the relative abundance of proteins in two or more samples by the metabolic incorporation of stable isotopically labeled amino acids (see Section 1.2.3.1).

The aim of this chapter was to use SILAC to investigate the relative abundance of proteins in MVs released by the RG/C2 and PRG/C2 cell lines to identify differentially secreted proteins that could have the potential to be used as markers of early tumour progression. Any of the essential amino acids can be used but Arg and Lys are more commonly used. Trypsin digestion cleaves proteins on the C-terminal side of basic residues therefore every tryptic peptide generated should contain a C-terminal Arg or Lys (except for the C-terminal peptide). For this study, the RG/C2 cell line was grown in media containing [$^{13}C_6$] labeled Arg ('heavy') and [$^{12}C_6$] labeled Lys ('light') or [$^{13}C_6$] labeled Lys ('heavy') and [$^{12}C_6$] labeled Arg ('light'). The PRG/C2 cell line was grown in [$^{12}C_6$] labeled Lys ('light') and [$^{12}C_6$] labeled Arg ('light') in both cases as described in Section 2.12. The labeling efficiency was investigated after six rounds of cell division (approx 5 weeks) by harvesting a cell lysate from both cell lines and separating the proteins by 1D gel electrophoresis in separate lanes (Figure 5.1). A single gel slice was excised from both lanes, in-gel digestion was performed and the tryptic peptides were analysed by LC MS/MS using a 120min acetonitrile gradient (Figure 2.3).

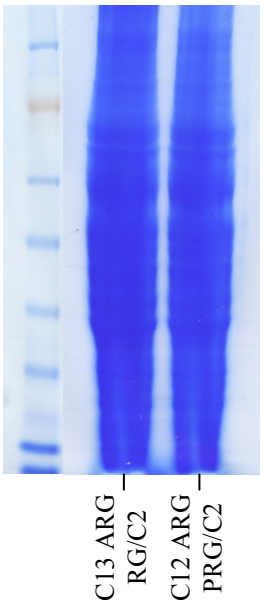


Figure 5. 1 : SILAC Labeling Efficiency.

RG/C2 cells were grown in media containing 'heavy' Arg and PRG/C2 cells were grown in media containing 'light' Arg for six rounds of cell divisions (5 weeks). The labeling efficiency was investigated by preparing a cell lysate and 25µg of the protein sample was separated on a 4-12% Bis-Tris gel and stained with CBBG250. A single gel slice was excised from both lanes for each experiment, in-gel digestion was performed and the tryptic peptides were analysed by LC MS/MS to calculate the percentage of proteins with each label.

The CBB G250 stained gel shown in Figure 5.1 was overloaded with protein but was sufficient for confirming the labeling efficiency of the cell lines grown in SILAC media. The results were searched against the human IPIv3.14 database and the results were filtered using the multiple thresholds parameters optimised in Chapter 4. There were a total of 48 proteins identified in the gel slice from the 'heavy' Arg labeled cells with 45 of these containing Arg residues. There were a total of 366 peptides identified in the 48 proteins, 166 of these contained Arg residues and 160 of those peptides contained 'heavy' labeled Arg residues which were not represented by the identical 'light' labeled peptide. Therefore the labeling efficiency was calculated as approx 97% (100 / 166 x 161 = 96.99). All the peptides identified in the gel slice from the PRG/C2 cell lysate contained 'light' labeled Arg residues (data not shown).

CM was then collected from the RG/C2 cell line grown in 'heavy' Arg media and the PRG/C2 cell line 'light' Arg media ($2.5x10^7$ cells each) and MV fractions were prepared as previously described. This was repeated for a total of four SILAC experiments with biological duplicate MV fractions prepared from CM collected from the RG/C2 cell line grown in either 'heavy' Arg or Lys and the PRG/C2 cell line grown in 'light' Arg and Lys. These four experiments are referred to as C13 ARG-A, C13 ARG-B, C13 LYS-A and C13 LYS-B throughout this study. The MV fractions for both the RG/C2 and PRG/C2 cell lines in each experiment were mixed in a 1:1 ratio (normalised by cell number) and the proteins were separated by 1D gel electrophoresis over 4cm of a 4-12% tris-bis gel (Figure 5.2a). The whole lane was then excised into 40 equal sized bands and in-gel digestion was performed on each. Each sample was numbered 1 → 40 (Figure 5.2b) and analysed by LC MS/MS using a 2hr acetonitrile gradient (as shown in Figure 2.3b).

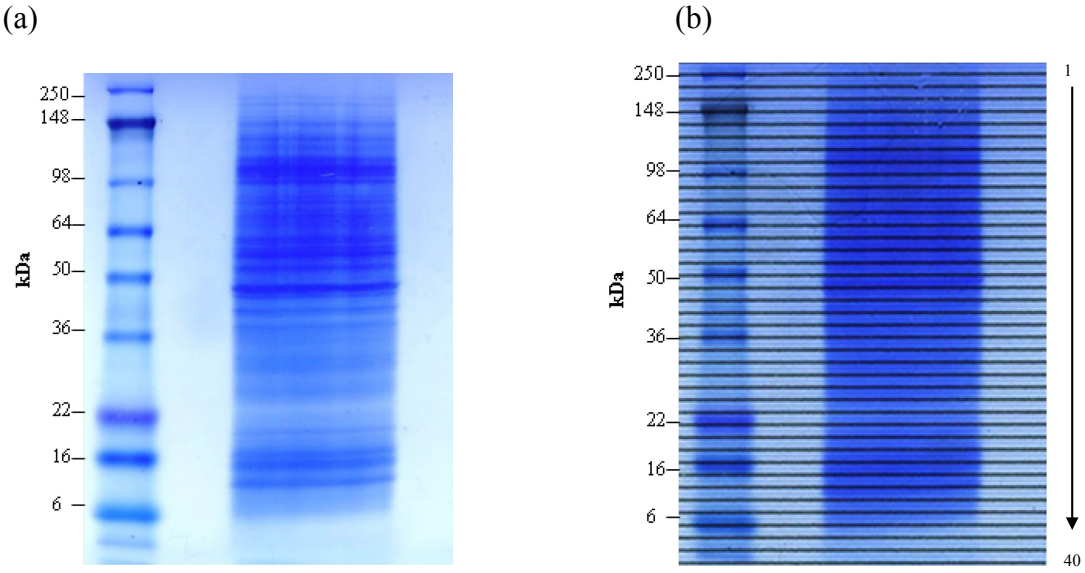


Figure 5. 2 : First collection of MV's from SILAC labeled RG/C2 and PRG/C2 cells.

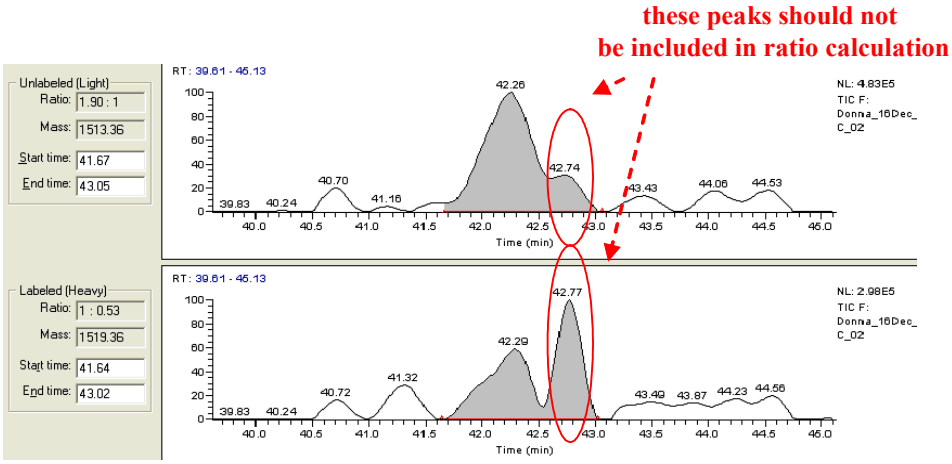
MV's prepared from CM derived from RG/C2 with 'heavy' labeled Arg residues and PRG/C2 with 'light' labeled Arg residues were mixed in a 1:1 ratio ($2.5x10^7$ cells each). (a) The proteins were then separated over 4 lanes of a 4-12% Tris-Glycine gel (b) and the whole lane was cut into 40 equal sized bands.

The raw files collected from experiment C13 ARG-A were then searched against the human IPI v3.14 database. The search results were filtered using the modified multiple threshold parameters optimised in Chapter 4 and only proteins with two unique peptides were included in the analysis. Supplementary Table 5a shows there were 485 proteins identified in total with the false positive rate calculated as 0.89% and a FDR of 0.0089. Only 34.7% of the total proteins identified were represented in both the 'light' and 'heavy' samples i.e. 168 out of 485. 128 of these proteins demonstrated an increase (≥ 1.5-fold) in secretion via MVs in the PRG/C2 cell line and none of the proteins showed a decrease (≤ 1.5-fold) in the PRG/C2 cell line. The false positive rate calculated for this dataset was identical to that calculated in the test dataset used in Table 4.5 when using the modified multiple thresholds filter parameters optimised in Chapter 4.

5.2 Using the Xpress Function to calculate an Abundance Ratio

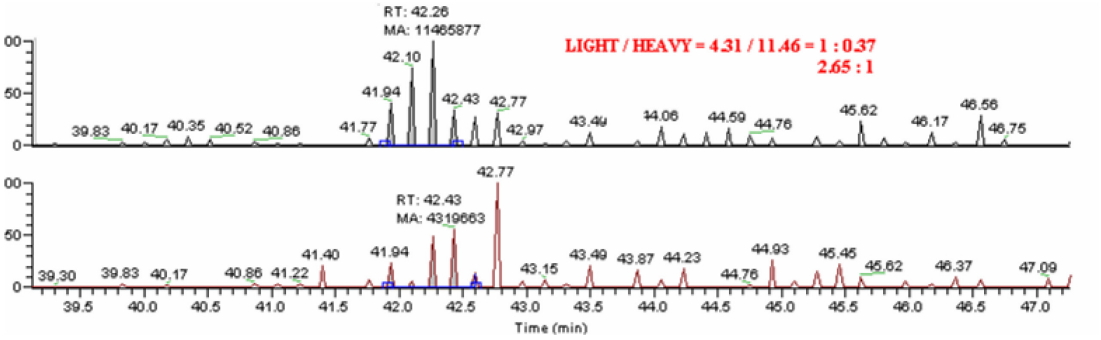
Supplementary Table 5a shows the relative abundance ratio as calculated using Xpress (light to heavy ratio) by Bioworks Browser. The software identifies the co-eluting 'light' and 'heavy' peaks and uses the area under the peak to calculate the relative abundance. In order to determine how accurate the Xpress function calculated the abundance ratio, a small number of the peaks used to calculate these ratios were manually inspected. Further investigation showed that many ratios were calculated using peaks that were incorrectly picked and Figure 5.3 shows an example of one of these peaks (peptide sequence IQGLTVEQAEAVVR identified from Dynein heavy chain IPI00456969 in sample 2). Xpress has chosen two peaks to calculate the area under the peak for both the light and the heavy peptides although manual inspection showed that only the first peak in both cases should be used. Xcalibur can be used to manually calculate the ratio by highlighting the 'light' and 'heavy' peaks and labeling the peaks using the area under the peak label function and Figure 5.3b shows the manual

(a)



Reference Scan(s)	Sequence	MH+	Charge	XC	Score Delta Cn	Sp	RSp	Peptides (Hits) Ions	XPRESS
IPI00456969					786.24			82 (75 3 0 2 2)	
1429 - 1435	R.IQGLTVEQAEAVVR.L	1513.72	2	4.12	0.54	1526.6	1	21/26	
1444 - 1462	R.IQGLTVEQAEAVVR#.L	1519.72	2	3.72	0.34	1594.2	1	20/26	1 : 0.53

(b)



(c)

(d)

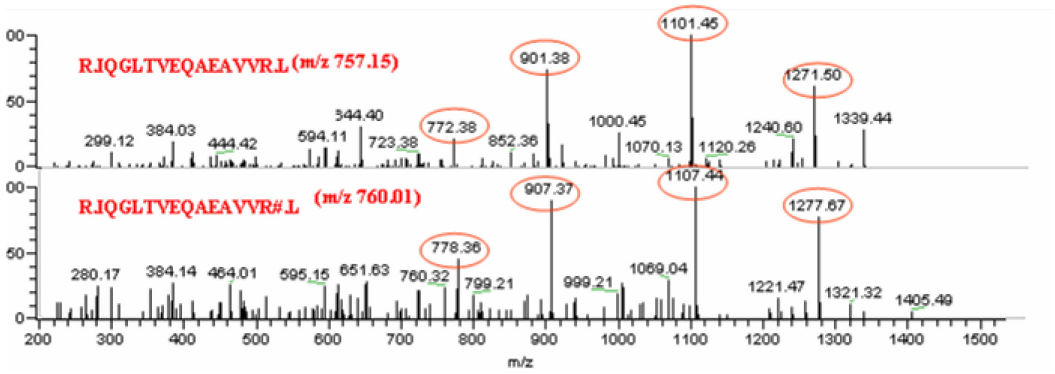


Figure 5. 3 : Example of an incorrectly calculated Xpress value.

Diagram to show an example of how the Xpress function in Bioworks can miscalculate the abundance ratio, highlighting the need to manually check all ratios. (a) Peak picking by Xpress used to calculate an abundance ratio. (b) Manual calculation of the ratio for the peptide sequence IQGLTVEQAEAVVR. (c) Spectrum to show the 'light' (m/z 757.15) and 'heavy' (m/z 760.01) peaks co-eluting at ~42mins. (d) MS/MS spectra for both the 'light' and 'heavy'. The MS/MS peaks circled in red represent the identical peaks in both the 'light' and 'heavy' spectra with a difference of 6Da indicating the fragments containing the heavy Arg residue.

calculation of the light/heavy ratio for the peak shown in Figure 5.3a.

The table in Figure 5.3a shows that Xpress has calculated the ratio as 1:0.53 but the ratio was calculated to be 1:0.37 when manually inspected therefore demonstrating a greater relative abundance ratio. These results suggest an increased amount of the 'light' peptide therefore demonstrating a 2.7-fold increase in that protein secreted by the PRG/C2 cell line compared to the 1.88 fold increase calculated by Xpress. Figure 5.3c shows the 'light' (m/z 757.15) and 'heavy' (m/z 760.01) peaks co-eluting at ~42mins and Figure 5.3d shows the MS/MS data for both peaks confirming that both ions have the same peptide sequence. Heavy Arg and Lys residues have a mass increase of 6 but the mass spectrometer measures the m/z ratio and not the mass. Therefore the observed mass difference is dependant upon the charge state of the peptide. The example shown in Figure 5.3 is the (+2) charged species of the peptide therefore giving a mass difference of 3 between the 'light' and 'heavy' peaks. The peak fragments highlighted by red circles in Figure 5.3d indicate those containing the labeled Arg residue (i.e the Y ion series) with a mass difference of 6 in the MS/MS spectra.

The analysis shown above demonstrated differences between the manual ratio calculation and the ratio calculated by Xpress for the peptide shown in Figure 5.3, other peptides with Xpress ratios were also manually inspected. Table 5.1 compares these ratios for a small number of proteins. The protein named IPI00024067.2 (sim to Clathrin heavy chain) has 10 peptide sequences where both the 'light' and 'heavy' peptides were identified. The Xpress ratios for these peptides range from 1:0.25 to 1:0.74 (average ratio of 1:0.45) whereas those manually calculated demonstrate a relatively narrow range of values of 1:0.17 to 1:0.39 (average ratio of 1:0.28). Not only does the manual calculation of the relative abundance demonstrate a greater difference in the light/heavy ratio but the ratios calculated for peptides from the same protein appear to show a greater consistency.

Protein name	Peptide sequence	Xpress ratio	manual ratio
sim to Clathrin heavy chain IPI:IPI00024067.2	-.ALEHFTDLYDIKR.- -.ALEHFTDLYDIKR#.-	1 / 0.45	1 / 0.17
	-.NNLAGAEELFAR.- -.NNLAGAEELFAR#.-	1 / 0.63	1 / 0.34
	-.TSIDAYDNFDNISLAQR.- -.TSIDAYDNFDNISLAQR#.-	1 / 0.40	1 / 0.24
	-.LAELEEFINGPNNAHIQQVGDR.- -.LAELEEFINGPNNAHIQQVGDR#.-	1 / 0.47	1 / 0.21
	-.ISGETIFVTAPHEATAGIIGVNR.- -.ISGETIFVTAPHEATAGIIGVNR#.-	1 / 0.28	1 / 0.22
	-.WLLLTGISAQQNR.- -.WLLLTGISAQQNR#.-	1 / 0.25	1 / 0.21
	-.LLLPWLEAR.- -.LLLPWLEAR#.-	1 / 0.42	1 / 0.34
	-.KFDVNTSAVQVLIEHIGNLDR.- -.KFDVNTSAVQVLIEHIGNLDR#.-	1 / 0.48	1 / 0.29
	-.LTDQLPLIIVC*DR.- -.LTDQLPLIIVC*DR#.-	1 / 0.74	1 / 0.39
	-.VGYTPDWIFLLR.- -.VGYTPDWIFLLR#.-	1 / 0.39	1 / 0.35
Ras GTPase-activating-like protein IQGAP1 IPI:IPI00009342.1	-.IFYPETTDIYDR.- -.IFYPETTDIYDR#.-	1 / 0.60	1 / 0.67
	-.IIGNLLYYR.- -.IIGNLLYYR#.-	0.29 / 1	1 / 0.38
	-.LEGVLAEVAQHYQDTLIR.- -.LEGVLAEVAQHYQDTLIR#.-	1 / 0.35	1 / 0.30
splice isoform 4 of Filamin B IPI:IPI00382698.1	-.EATTDFTVDSRPLTQVGGDHIK.- -.EATTDFTVDSR#PLTQVGGDHIK.-	1 / 0.97	1 / 0.89
	-.VLFASQEIPASPFR.- -.VLFASQEIPASPFR#.-	1 / 0.72	1 / 0.57
splice isoform long of Spectrin beta chain, brain 1 IPI:IPI00005614.5	-.LLEVLSGER#.- -.LLEVLSGER.-	1 / 0.94	1 / 0.85
	-.ITDLYTDLR#.- -.ITDLYTDLR.-	0.42 / 1	1 / 0.94
splice isoform 1 of Solute carrier family 12 member 2 IPI:IPI00022649.3	-.EGAQYLMQAAGLGR.- -.EGAQYLMQAAGLGR#.-	1 / 0.43	1 / 0.53
splice isoform long of Na/K transporting ATPase alpha chain 1 IPI:IPI00006482	-.GVGIISEGNETVEDIAAR.- -.GVGIISEGNETVEDIAAR#.-	1 0.48	1 / 0.38

Table 5. 1 : Comparison of Xpress ratio and manually calculated ratio.

Table to show all proteins identified in sample 5 from experiment C13 Arg-A with both light and heavy peaks identified. Also shows the comparison of abundance ratios calculated by Xpress compared to those calculated manually.

These results suggest that manual calculation of all the ratios for peptides with matching SILAC pairs would be more accurate than the Xpress function used by Bioworks Browser. However, this would be a very time-consuming task to perform not just for one experiment but for replicate experiments also. For a large dataset such as this, software that could calculate the ratio in an accurate and automated fashion is required to enable the analysis to be less labour intensive with reliable results.

5.3 Data Analysis using the Trans Proteomic Pipeline (TPP)

The Trans Proteomic Pipeline (TPP) is a free to download data analysis software package (http://tools.proteomecenter.org/windows.php) developed by the Institute for Systems Biology (Seattle) for the analysis of large proteomic datasets which has been utilised in a number of proteomic studies (Desiere et al., 2005, Han et al., 2001, A. Keller et al., 2002a, X. J. Li et al., 2003, Nesvizhskii and Aebersold, 2004, Nesvizhskii et al., 2003, von Haller et al., 2003). TPP involves a series of perl scripts which first converts all MS/MS files (generated from the spectra) to one XML file, it then calculates a probability score for peptide and protein assignments using Peptide Prophet™ and Protein Prophet™ respectively (A. Keller et al., 2002b, Nesvizhskii et al., 2003). As well as calculating an Xpress ratio, it also uses a program called Automated Statistical Analysis of Protein Abundance Ratios (ASAPRatio) to quantify both peptides and proteins and has been reported to be more accurate than Xpress (X. J. Li et al., 2003). TPP also automatically predicts the false positive rate of the dataset to enable the user to set a minimum protein probability threshold based upon the reported error rates (A. Keller et al., 2002a). Figure 2.4 shows an overview of the processes involved in TPP data analysis. As it was previously shown, the manual validation of the abundance ratios calculated using the Xpress function in the Bioworks Browser software

is time consuming for large datasets therefore TPP was used to analyse the data of the SILAC experiments described in this study.

The proteins in the MV fractions for the remaining three SILAC experiments were also separated by 1D electrophoresis, the whole lane was excised as 40 equal sized bands, digested with trypsin and analysed by LC MS/MS as described above. The raw files for each experiment were searched against a human IPI database (v3.22). Then using TPP, the raw files were converted to a mzXML file and all the .out and .dta files created by the database search for each individual experiment were converted to a single XML file. The data was then in the correct format for analysis by Peptide Prophet™ followed then by Protein Prophet™ analysis and ASAPRatio was used to calculate the abundance ratios at the peptide and protein level respectively. This analysis was performed for each of the four SILAC experiments and the results were compared manually in an excel spreadsheet format (Supplementary Table 5b).

Supplementary Tables 5d, 5e, 5f and 5g show the TPP results for experiments C13 ARG-A, C13 ARG-B, C13 LYS-A and C13 LYS-B respectively. Table 5.2 outlines the results of the comparison of the TPP results using a threshold of protein probability >0.90 and >0.70. The error rates calculated by TPP are very similar for all four SILAC experiments when using protein probability >0.90 with 0.013, 0.012, 0.011 and 0.007 for C13 ARG A, C13 ARG B, C13 LYS A and C13 LYS B respectively. These values increase when the threshold was decreased to protein probability >0.70 with 0.042, 0.042, 0.023 and 0.017 for C13 ARG A, C13 ARG B, C13 LYS A and C13 LYS B respectively (see Supplementary Figure 5a for the sensitivity and error rates observed for each of the four datasets). All proteins identified by the TPP analysis using the protein probability >0.70 threshold (including proteins with only a single peptide identified) for all 4 experiments were used for

the comparison. Table 5.2 shows there were a total of 1396 proteins identified with 136, 252, 561 and 1396 proteins identified in at least 4, 3, 2 and 1 experiment. It also shows the number of proteins with multiple and single peptides, proteins with a >1.5-fold change as calculated by the ASAPRatio and the number of proteins with increased and decreased secretion in MVs by the PRG/C2 cell line for each of the four experiments. A direct comparison of the manually calculated ratios with the ASAPRatio was not carried out as the software utilises extra functions to calculate the relative abundance ratio that would be very difficult to perform manually such as background subtraction and standard deviation calculation to determine the reliability of the ratio calculated.

Supplementary Table 5b shows the results of this comparison including the name and IPI accession number of all the proteins identified, ASAPRatios and the protein probabilities for each individual experiment and also the average protein probability and ASAPRatio over the four SILAC experiments. The proteins identified in multiple experiments have a higher protein probability than those identified in a single experiment adding greater confidence that the proteins identified in multiple experiments are present in MVs released by these cell lines. The proteins identified in 4, 3, 2 and 1 experiment demonstrate protein probabilities >0.92, 0.88, 0.76 and 0.70 respectively. Table 5.2 also shows the percentage of total proteins identified in all 4 experiments with calculated ASAPRatios. There were 49% and 65% of proteins identified in the Arg experiments A and B. The percentages were higher in Lys experiments A and B with 83% and 84% of the total proteins with calculated ASAPRatios. These results show that Lys should be the amino acid of choice for SILAC experiments using only a single labeled amino acid residue with at least 80% of all proteins containing peptides suitable for relative abundance calculation. Supplementary Table 5h shows the average protein probability, the average ASAPRatio and the individual ASAPRatios calculated in each

	>0.90 PROTEIN PROBABILITY				>0.70 PROTEIN PROBABILITY			
	C13ARG		C13LYS		C13ARG		C13LYS	
	A	B	A	B	A	B	A	B
Error rate (FDR)	0.013	0.012	0.011	0.007	0.042	0.042	0.023	0.017
Total proteins identified	829	542	348	364	977	643	369	379
Proteins with single peptides	210	139	114	153	311	137	133	149
Proteins with multiple peptides:	619	403	234	211	666	505	236	230
-with ASAPRatios #	415	231	195	184	435	249	199	191
->1.5 fold change #	290	143	160	160	305	157	159	160
-Increased release in PRG/C2 #	250	122	157	153	259	132	157	153
-Decreased release in PRG/C2 #	40	21	3	7	46	25	4	7
% of proteins with multiple peptides with ASAPRatios	67.04	57.32	83.33	87.20	65.30	49.30	84.32	83.04
Total number of proteins*	1198				1396			
Proteins found in all 4 experiments*	108				136			
Proteins found in 3 experiments*	89				116			
Proteins found in 2 experiments*	314				315			
Proteins found in 1 experiment*	1198				829			

Table 5. 2 : Comparison of the results from the TPP analysis of the 4 SILAC experiments using protein probability values of >0.90 and >0.70.

Table to show the false positive rates, number of proteins identified the number of proteins with ASAPRatios and the percentage of the total number of proteins identified with peptides used to calculate abundance ratios. * - includes proteins with single peptides. # - proteins with multiple peptides only.

experiment for the 1396 proteins identified in all four experiments using protein probability >0.70 threshold.

5.4 Data Analysis using Babelomics

Data analysis of genomic-profiling experiments (i.e. micoarray gene expression experiments) often involves interpretation of up-regulated and down-regulated genes to identify pathways or a particular biological function which has shown significant differences between two biological conditions. Babelomics (http://www.babelomics.org/) is an online set of tools used to perform this type of data interpretation required for microarray gene expression profiles (Al-Shahrour et al., 2006). FatiGO+ (http://www.fatigo.org/) is one of these tools used to compare the GO (Gene Ontology) terms associated with two lists of genes (Al-Shahrour et al., 2007). The Gene Ontology (GO) project (http://www.geneontology.org/) is a collaborative effort to address the need for consistent descriptions of gene products in different databases (Ashburner et al., 2000). They have developed a database that describes gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. The ontologies are structured so they can be viewed at different levels of complexity and gene products might be associated with one or more processes, cellular components or molecular function. Cellular component refers to an anatomical structure (e.g. rough endoplasmic reticulum or nucleus) or a gene product group (e.g. ribosome, proteasome or a protein dimer). Biological process is described by the GO project as a series of events accomplished by one or more ordered assemblies of molecular functions and is not equivalent to a pathway. Molecular function refers to activities, such as catalytic or binding activities, that occur at the molecular level and generally corresponds to activities that can be performed by individual gene products, but some

activities are performed by assembled complexes of gene products. Another function of FatiGO+ is to associate gene products with KEGG pathways (http://www.genome.jp/kegg/). KEGG (Kyoto Encyclopaedia of Genes and Genomes) is a database that associates gene products with biological pathways (Kanehisa et al., 2004).

The references described above were all used to interpret the proteomic data described in this chapter. The SILAC data generated in this study was analysed in the same way that microarray data is analysed using the following: Babelomics was used to investigate the biological function, the KEGG pathways and cellular components associated with the proteins identified in this study. Firstly, the protein accession numbers were converted to gene symbols using REFDIC (http://refdic.rcai.riken.jp/welcome.cgi) (Hijikata et al., 2007) and those found in at least 2 out of 4 experiments were then analysed using FATIGO+ as described in Section 2.13.3 (see Figure 5.4 for results). The table shown in Figure 5.4a shows the biological processes with ≥5 proteins identified in this survey associated with that particular process. The top three processes were apoptosis, regulation of cell death and glucose metabolic processes with 38, 32 and 21 proteins respectively. The table shown in Figure 5.4a shows the gene symbols associated with each of the biological processes. The bar chart in Figure 5.4b shows the KEGG pathways involving ≥7 of the identified proteins with ribosomal pathways and regulation of the actin cytoskeleton associated 30 and 28 proteins respectively. The table in Figure 5.4b shows the gene symbols associated with each of the KEGG pathways. Finally, the pie chart in Figure 5.4c shows the number of proteins found in each of the cellular components. These results show that the proteins found in these MVs originate from various compartments within the cell including the nucleus, cytosol, mitochondria and the cytoskeleton.

(a) Biological Processes (GO Level 8)

Biological process	Genes list	NºGenes
apoptosis	PPP2R1A PDIA3 HSPA1A TIAF1 CYCS VDAC1 YWHAG ACTN4 HSPD1 CIB1 DAD1 HSPE1 ACTN1 MIF ARHGDIA CFL1 EEF1E1 TUBB2C CLIC1 EIF5A VCP PDCD6IP PRDX2 YWHAH ANXA1 HSP90B1 RTN4 ALB CSE1L NME1 TPT1 PHB HSPA5 TUBB GSTP1 PDCD6 CALR ANXA5	38
regulation of programmed cell death	PPP2R1A PDIA3 HSPA1A TIAF1 CYCS YWHAG ACTN4 HSPD1 HSPE1 ACTN1 MIF ARHGDIA CFL1 EEF1E1 TUBB2C CLIC1 VCP PRDX2 YWHAH ANXA1 HSP90B1 RTN4 ALB NME1 TPT1 PHB HSPA5 TUBB GSTP1 PDCD6 CALR ANXA5	32
glucose metabolic process	PGK1 PFKP PGM1 PGAM1 PFKL PCK2 LDHB FBP1 PKM2 MDH1 ALDOA TALDO1 UGDH ENO1 MDH2 TPI1 UGP2 GAPDH GPI ENO3 LDHA	21
hexose catabolic process	PGK1 PFKP PGM1 PGAM1 PFKL LDHB PKM2 MDH1 ALDOA TALDO1 ENO1 MDH2 TPI1 GAPDH GPI ENO3 LDHA	17
regulation of transcription, DNA-dependent	STAT3 RPL6 SOD2 CARHSP1 PFN1 TTLL3 XRCC6 CTNND1 MYO6 VPS25 RUVBL2 RAN ENO1 NME1 RAB1A CALR SUB1	17
proteolysis involved in cellular protein catabolic process	PSMB5 PSMB4 PSMB8 PSMB1 PSMB3 PSMA2 PSMA7 PSMA1 PSMB2 PEPD VCP UCHL3 PSMA5 PSMA4 PSMC6 PSMA6 PSMA3	17
protein targeting	PDIA3 YWHAZ YWHAG HSPD1 YWHAQ SSR3 SDCBP AP1M2 RAN YWHAB CSE1L RHOA KPNB1 CALR COPB1	15
actin polymerization and/or depolymerization	SPTBN1 DSTN ARF6 CAPZA2 ARPC3 TTLL3 ARPC5L VIL1 ARPC2 CAPZB ARPC5 CAPZA1 SPTAN1 DIAPH1	14
regulation of actin filament length	SPTBN1 DSTN ARF6 CAPZA2 ARPC3 ARPC5L VIL1 ARPC2 CAPZB ARPC5 CAPZA1 SPTAN1	12
purine ribonucleotide biosynthetic process	ATP5A1 PAICS ATP5C1 ATP5O ATP5J2 ATP5B ATP5F1 NME1 ATP5L	9
proton transport	ATP5A1 ATP5C1 ATP5O ATP5J2 ATP4A ATP5B ATP5F1 ATP5L ATP1A1	9
neuron differentiation	PLXNB2 YWHAG MYO6 HPRT1 YWHAH RTN4 CNP MTPN NAPA	9
purine nucleoside triphosphate biosynthetic process	ATP5A1 ATP5C1 ATP5O ATP5J2 ATP5B ATP5F1 NME1 ATP5L	8
ribonucleoside triphosphate biosynthetic process	ATP5A1 ATP5C1 ATP5O ATP5J2 ATP5B ATP5F1 NME1 ATP5L	8
purine ribonucleoside triphosphate metabolic process	ATP5A1 ATP5C1 ATP5O ATP5J2 ATP5B ATP5F1 NME1 ATP5L	8
tRNA aminoacylation	TARS EPRS FARSLA CARS RARS NARS DARS	7
Rho protein signal transduction	ARF6 ARHGDIA CFL1 TAX1BP3 PTPLAD1 ARHGDIB RHOA	7
ER to Golgi vesicle-mediated transport	RAB2 ARF6 ARF1 RAB1A ARF4 NAPA	6
mitosis	YWHAG TTN KIF23 YWHAH RAN PAFAH1B1	6
microtubule-based movement	TUBB2C KIF23 PAFAH1B1 DST DYNC1H1 TUBB	6
tricarboxylic acid cycle	SUCLG2 IDH1 MDH1 ACO1 MDH2 FH	6
epidermis development	FABP5 KRT9 SFN KRT10 DSP KRT2	6
hexose biosynthetic process	PGM1 PCK2 FBP1 TPI1 GPI	5
transcription from RNA polymerase II promoter	STAT3 SOD2 PFN1 MYO6 SUB1	5
chromatin assembly or disassembly	H1F0 H2AFJ H2AFY HIST1H2BM NAP1L4	5
arginine metabolic process	ASL ASS1 CAD DDAH1 CLIC1	5
actin filament-based movement	MYH9 MYH14 MYO6 MYL6 MYO1A	5
cell projection biogenesis	SDCBP CDC42 ITGA6 CAPZB VCL	5
potassium ion transport	ATP1B1 ATP4A SLC12A2 ATP1A1 ATP1B3	5
intra-Golgi vesicle-mediated transport	COPE ARCN1 COPG COPB1 NAPA	5
nuclear import	PDIA3 RAN CSE1L RHOA KPNB1	5
sodium ion transport	NDUFA9 ATP1B1 SLC12A2 ATP1A1 ATP1B3	5

(b) KEGG pathways

KEGG pathways	Genes list	NºGenes
Ribosome	RPL19 RPS8 RPL38 RPL6 RPS5 RPS6 RPS3 RPS7 RPS15A RPS27L RPS19 RPL30 RPL18A RPS25 RPL18 RPS4X RPL11 RPL7 RPS17 RPL12 RPS20 RPL24 RPS9 RPL27 RPS10 RPSAP15 RPS13 RPLP2 RPS16 RPS24	30
Regulation of actin cytoskeleton	ITGB1 MYH9 ARPC1B ACTN4 MYH14 TTN IQGAP1 VIL2 ACTN1 PFN1 CFL1 ARPC3 ROCK2 TTLL3 ITGA2 ARPC5L ACTC1 IQGAP3 GNG12 CDC42 ITGA6 ARPC2 ACTB RHOA ITGB4 ARPC5 VCL DIAPH1	28
Focal adhesion	ITGB1 ACTN4 TTN ACTN1 FLNB ROCK2 MYL6 ITGA2 ACTC1 CDC42 ITGA6 ACTB RHOA ITGB4 TLN1 VCL DIAPH1	17
Glycolysis / Gluconeogenesis	PGK1 PFKP PGM1 PGAM1 PFKL LDHB FBP1 ALDH2 PKM2 ALDOA ENO1 TPI1 GAPDH GPI ALDH1B1 ENO3 LDHA	17
Proteasome	PSMB5 PSMB4 PSMB1 PSMB3 PSMD3 PSMA2 PSMC3 PSMA7 PSMA1 PSMB2 PSMA5 PSMA4 PSMC6 PSMA6 PSMC4 PSMA3	16
Oxidative phosphorylation	CYC1 ATP5A1 NDUFA9 UQCRC2 COX5A COX7A2 PPA1 ATP5C1 ATP5O ATP5J2 COX2 ATP4A ATP5B ATP5F1 UQCRB ATP5L	16
Tight junction	PPP2R1A MYH9 ACTN4 MYH14 PPP2CB ACTN1 CLDN3 MYL6 ACTC1 CDC42 ACTB CLDN9 RHOA SPTAN1 TJP2	15
Pathogenic Escherichia coli infection - EPEC	ITGB1 YWHAZ YWHAQ VIL2 TUBB2C ROCK2 ARPC5L ACTC1 CDC42 ACTB RHOA TUBB ARPC5 KRT18	14
Leukocyte transendothelial migration	ITGB1 ACTN4 VIL2 ACTN1 ROCK2 CLDN3 CTNND1 MYL6 ACTC1 CDC42 ACTB CLDN9 RHOA VCL	14
Pathogenic Escherichia coli infection - EHEC	ITGB1 YWHAZ YWHAQ VIL2 TUBB2C ROCK2 ARPC5L ACTC1 CDC42 ACTB RHOA TUBB ARPC5 KRT18	14
Pyruvate metabolism	ME1 PCK2 LDHB ALDH2 PKM2 MDH1 MDH2 ALDH1B1 LDHA GRHPR ACAT1	11
Cell Communication	KRT9 KRT8 ACTC1 ITGA6 ACTB ITGB4 LMNA KRT10 KRT19 KRT2 KRT18	11
Antigen processing and presentation	PDIA3 HSPA1A HSP90AB1 HSP90AA1 HSPA8 B2M PSME1 CANX HSPA5 CALR	10
Carbon fixation	PGK1 ME1 FBP1 PKM2 MDH1 ALDOA MDH2 TPI1 TKT	9
Arginine and proline metabolism	ASL CKMT1A ASS1 EPRS ALDH2 ALDH1B1 MAOA RARS LAP3	9
Adherens junction	ACTN4 IQGAP1 ACTN1 CTNND1 ACTC1 CDC42 ACTB RHOA VCL	9
Purine metabolism	PAICS GDA AK2 PKM2 HPRT1 ATIC NME1 APRT ENTPD6	9
Pentose phosphate pathway	PFKP PGM1 PFKL FBP1 ALDOA TALDO1 TKT GPI	8
Insulin signaling pathway	PFKP RPS6 PFKL PCK2 FBP1 CALM1 PKM2 FASN	8
Fatty acid metabolism	HADHA ALDH2 HADHB HADH ACSL5 ALDH1B1 ACAA2 ACAT1	8
Fructose and mannose metabolism	PFKP HSD17B12 PFKL FBP1 ALDOA PHPT1 TPI1 SORD	8
Valine, leucine and isoleucine degradation	HADHA ALDH2 HADHB HADH ALDH1B1 ACAA2 TRIM23 ACAT1	8
Calcium signaling pathway	VDAC1 SLC25A5 TTN CALM1 GNA14 TAX1BP3 VDAC3 SLC25A4	8
Citrate cycle (TCA cycle)	PCK2 SUCLG2 IDH1 MDH1 ACO1 MDH2 FH	7
PPAR signaling pathway	ME1 PCK2 FABP5 APOA1 ACSL5 FABP1 DBI	7
Tryptophan metabolism	HADHA ALDH2 CAT HADH ALDH1B1 MAOA ACAT1	7
Cell cycle	YWHAZ YWHAG YWHAQ YWHAH SFN YWHAB YWHAE	7
Aminoacyl-tRNA biosynthesis	TARS EPRS FARSLA CARS RARS NARS DARS	7
Propanoate metabolism	HADHA SUCLG2 LDHB ALDH2 ALDH1B1 LDHA ACAT1	7

(c) Cellular Components (GO level 8)

Cellular Components	Genes list	NºGenes
nucleus	PPP2R1A HSPA1A MYH9 STAT3 PSMB4 YWHAZ TIAF1 TAGLN2 PTBP1 RPS6 CYCS CMPK RPS7 ACTN4 PSMB3 CIB1 HSPA8 ANP32B TTN ABHD14B PTGES3 RPS19 MVP SFRS3 PSMC3 TSN PFN1 SDCBP CFL1 SH3BGRL HINT1 LGALS3 EEF1E1 PSMA1 FER1L3 CLIC1 EIF5A TTLL3 XRCC6 CTNND1 KIF23 PSMB2 HNRPH1 MYO6 TAX1BP3 VCP RBMX H1F0 VPS25 C11orf54 ACTB HNRPAB RUVBL2 RAN H2AFJ RTN4 HNRPK H2AFY CSE1L ENO1 PREP CSTB NME1 PAFAH1B1 KPNB1 PHB TRIM23 ANXA11 HSPA5 PARK7 RPSAP15 PSMC6 HIST1H2BM LMNA PDCD6 SH3BGRL2 FKBP4 G3BP1 TJP2 NAP1L4 SUB1 UBE2N HNRPA2B1 PSMC4 PSMA3 S100P CSRP1 RACGAP1 SYNCRIP HNRPF	90
cytosol	PPP2R1A PSMB5 RPS8 PFKP MYH9 RPL6 RPS5 PSMB4 ME1 CCT7 LGALS4 RPS6 CYCS RPS3 PSMB8 RPS7 PSMB1 PGAM1 PSMB3 PFKL HSP90AA1 CCT3 RPS15A RPS19 PSMD3 RPL18A CCT2 RPS25 PSMA2 IDH1 VIL2 PSMC3 RPL18 PSME1 JUP PSMA7 RPL7 PKM2 PSMA1 VBP1 MDH1 CCT8 CCT5 PSMB2 RPS17 VCP PDCD6IP RPL12 RPS20 PHPT1 ACTB RUVBL2 HSP90B1 ENO1 PSMA5 PSMA4 CYB5R3 HBA1 CCT4 RPS9 RPS10 ENO3 LDHA AKR7A3 HSPA5 TLN1 RPSAP15 PSMC6 CALR RPS13 COPB1 RPLP2 RPS16 PSMA6 RPS24 PSMC4 ARL2 PSMA3	78
cytoskeleton	PPP2R1A MYO1C MYH9 SPTBN1 TIAF1 ARPC1B DSTN ACTR2 ACTN4 MYH14 CCT3 TTN PTK9 IQGAP1 VIL2 ACTN1 KRT9 SLC9A3R1 TBCA ARHGDIA PFN1 SDCBP CAPZA2 JUP CFL1 HINT1 KRT8 ARPC3 PLS1 TUBB2C FLNB ROCK2 TTLL3 CTNND1 KIF23 WDR1 MYO6 MYL6 ARPC5L ACTC1 TPM3 CAP1 VIL1 ANXA1 ARPC2 ACTB CAPZB ARHGDIB RHOA ACTR3 TPT1 PAFAH1B1 DST DYNC1H1 TLN1 TUBB ARPC5 LMNA CAPZA1 MYO1D VCL SPTAN1 KRT20 KRT10 LASP1 KRT19 DSP KRT2 MYO1A RACGAP1 KRT18 DIAPH1	72
mitochondrion	PPP2R1A HINT2 C1QBP HSPA1A CYC1 ATP5A1 CLTC YWHAZ HADHA CYCS VDAC1 HSP90AB1 NDUFA9 SOD2 HSPD1 CKMT1A PCK2 SUCLG2 SLC25A5 ALDH2 HSPE1 CYB5B LYPLA1 HADHB UQCRC2 AK2 ETFA COX5A COX7A2 BRP44 PKM2 ATP5C1 TUFM ATP5O PRDX5 SLC25A17 ATP5J2 COX2 CAT MDH2 HADH ATP5B CYB5R3 ATP5F1 UQCRB ACSL5 ALDH1B1 VDAC3 MAOA ATP5L FH IMMT ECH1 ACAA2 PHB SLC25A4 ACAT1 LAP3 DSP PRDX4 MGST1	61
organelle membrane	CYC1 ATP5A1 CLTC RPN2 HADHA TAGLN2 CYCS COPE VDAC1 NDUFA9 SOD2 DAD1 SLC25A5 SSR3 CYB5B SSR4 UQCRC2 AK2 COX5A COX7A2 ARCN1 AP1M2 MYO6 ATP5C1 ATP5O SLC25A17 JPH4 ATP5J2 COX2 HSP90B1 RAN RTN4 CAT CSE1L HADH ATP5B CYB5R3 ATP5F1 UQCRB ACSL5 VDAC3 IMMT COPG KPNB1 ACAA2 PHB SLC25A4 HSPA5 LMNA COPB1 DDOST ACAT1 AP1S1 MGST1	54
endoplasmic reticulum	HYOU1 PDIA3 ITGB1 HSPA1A PGRMC1 DNAJC14 RPN2 RAB2 ERO1L HSD17B12 CIB1 DAD1 MGST3 SSR3 SDCBP PDIA4 SSR4 ERP29 CANX NCSTN VCP P4HB PPIB TMEM4 HSP90B1 PDIA6 DPEP1 RTN4 TXNDC12 CYB5R3 PTGFRN PRKCSH RAB1A CDIPT HSPA5 TMCO1 CALR DDOST NAPA MGST1 SYNCRIP PTP4A1 GANAB	43
organelle envelope	CYC1 ATP5A1 HADHA TAGLN2 CYCS VDAC1 NDUFA9 SOD2 SLC25A5 CYB5B HADHB UQCRC2 AK2 COX5A COX7A2 FER1L3 CLIC1 MYO6 ATP5C1 ATP5O SLC25A17 ATP5J2 COX2 RAN RTN4 CSE1L HADH ATP5B CYB5R3 ATP5F1 UQCRB ACSL5 VDAC3 PAFAH1B1 IMMT KPNB1 ACAA2 PHB ANXA11 SLC25A4 LMNA ACAT1 MGST1	43
integral to plasma membrane	ITGB1 MYH9 PGRMC1 ATP8B1 TFRC SLC25A5 ATP1B1 CEACAM5 SDCBP ABCC3 GALNT8 CLDN3 NCSTN ITGA2 SLC25A17 TAX1BP3 SLC1A5 ITGA6 TMEM4 GPR109B LNPEP ATP4A VDAC3 SLC12A2 ITGB4 PHB SLC25A4 CD97 RPSAP15 GPA33 SLC1A4 CD81 ATP1A1 TSPAN8 TJP2 CD9 LAMP1 ATP1B3	38
ribosome	RPL19 RPS8 RPL38 RPL6 RPS5 RPS6 RPS3 RPS7 RPS15A RPS27L RPS19 RPL30 RPL18A RPS25 RPL18 RPS4X RPL11 RPL7 RPS17 RPL12 RPS20 GCN1L1 RPL24 RPS9 RPL27 RPS10 RPSAP15 RPS13 RPLP2 RPS16 RPS24	31
Golgi apparatus	ARL1 RNPEP COPE CORO7 RAB2 GDI2 ARCN1 RAB11A RAB10 MYO6 NCSTN JPH4 RAB1A COPG COPB1 AP1S1 NAPA	17

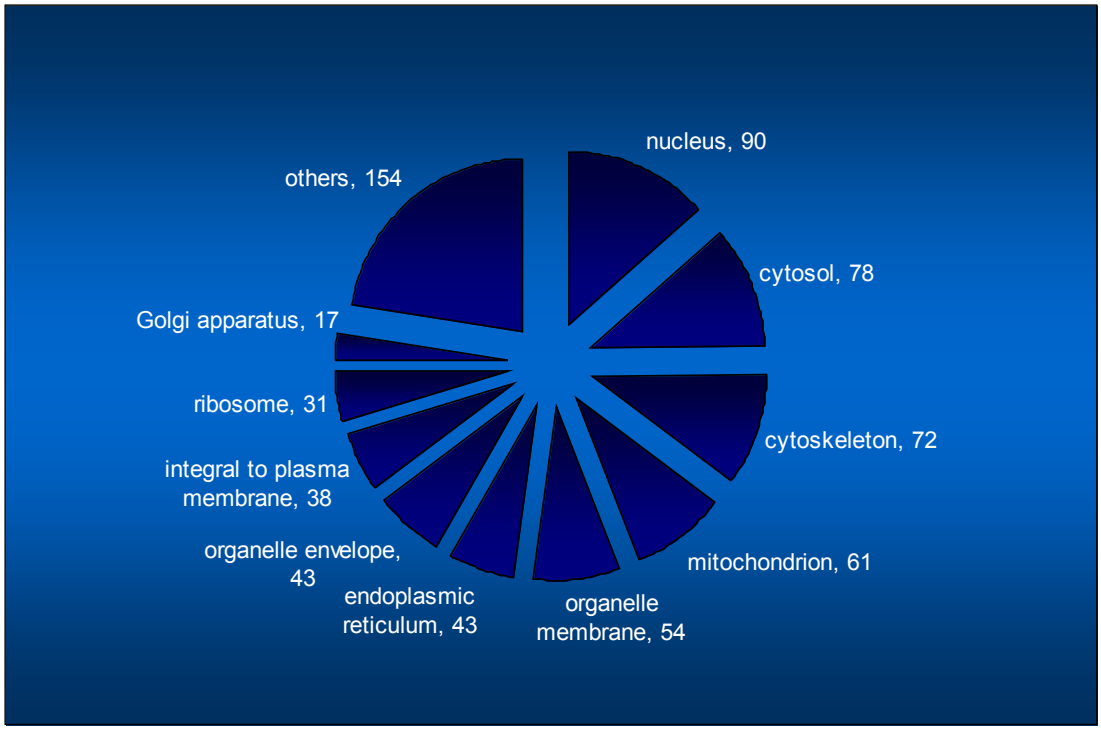


Figure 5. 4 : Analysis of all proteins found in at least 2 experiments using software commonly used in microarray analysis, Babelomics.

The accession numbers for the proteins identified in this survey that were found in at least 2 experiments were converted to gene names using online software called REFDIC. The gene names were then used to investigate which biological processes, cellular components and pathways these proteins are involved in Babelomics. (a) A bar chart to show the biological processes with ≥5 proteins identified in this survey (GO level 8). The table identifies the gene symbols of the proteins involved in each biological process. (b) A bar chart to show the KEGG pathways with ≥7 proteins identified in this survey. The table identifies the gene symbols of the proteins involved in each pathway. (c) Table and pie chart to show the number of proteins that are localized to various cellular compartments (GO level 8). The table identifies the gene symbols of the proteins found in each cellular component.

5.5 Potentially Interesting Proteins for Further Investigation

Next, the combined TPP results from the four SILAC experiments were compared to identify proteins with consistent ASAPRatios in multiple experiments that could be used to validate both the presence of the proteins in MVs and the relative abundance ratios. Table 5.3 shows 86 proteins identified in at least 2 out 4 experiments with an average ASAPRatio ≥2.00 and an average protein probability ≥0.90 also including the ASAPRatio calculated in each individual experiment. This list was used to identify proteins that were differentially released in MV by the RG/C2 and PRG/C2 cell lines that would be suitable to validate the SILAC results. The biological processes, KEGG pathways and cellular components associated with these proteins were investigated using FATIGO+.

The bar chart and table shown in Figure 5.5a shows the top two biological processes were also apoptosis and regulation of cell death with the same 12 proteins identified in both. The bar chart and table shown in Figure 5.5b show the top two KEGG pathways identified were ribosomal pathways and antigen processing and presentation with 8 and 7 associated proteins respectively. The table and pie chart shown in Figure 5.5c demonstrate these proteins are associated with a variety of cellular components with the top three compartments being mitochondria, the cytosol and the endoplasmic reticulum with 16, 15 and 15 associated proteins respectively.

PROTEIN	IPI Accession	PROT PROB	AVERAGE ASAPR	C13 A	Arg B	C13 A	Lys B
GUANINE NUCLEOTIDE-BINDING PROTEIN G(I)/G(S)/G(O) GAMMA-12 SUBUNIT PRECURSOR	IPI00221232.9	0.97	500.83	-1	-1	2.66	999
ATP SYNTHASE SUBUNIT BETA, MITOCHONDRIAL PRECURSOR	IPI00303476.1	1.00	334.49	1.86	2.6	-1	999
ANNEXIN A5	IPI00329801.12	1.00	334.45	2.73	1.63	999	-1
SH3 DOMAIN-BINDING GLUTAMIC ACID-RICH-LIKE PROTEIN 2	IPI00412272.2	1.00	251.45	2.72	1.85	2.21	999
TRANSITIONAL ENDOPLASMIC RETICULUM ATPASE	IPI00022774.3	1.00	251.30	0.88	999	2.3	3
CYTOCHROME C OXIDASE SUBUNIT 2	IPI00017510.3	0.99	251.25	2.22	1.73	2.03	999
RAB GDP DISSOCIATION INHIBITOR BETA	IPI00031461.1	1.00	251.19	2.18	999	2.38	1.18
PROLYL ENDOPEPTIDASE	IPI00747355.1	1.00	251.04	1.86	999	1.16	2.13
ATP SYNTHASE SUBUNIT ALPHA, MITOCHONDRIAL PRECURSOR	IPI00440493.2	1.00	250.44	1.71	2.04	999	-1
KERATIN, TYPE I CYTOSKELETAL 10	IPI00009865.1	1.00	5.19	2.53	2.36	10.69	-1
KERATIN, TYPE I CYTOSKELETAL 9	IPI00019359.3	1.00	4.30	1.32	-1	2.25	9.33
60S RIBOSOMAL PROTEIN L30	IPI00219156.7	1.00	3.43	3.4	4.43	2.47	-1
AGR2	IPI00007427.2	0.99	3.40	2.07	3.19	3.27	5.08
HEAT SHOCK PROTEIN HSP 90-ALPHA	IPI00784295.2	1.00	3.08	2.55	5.11	2.23	2.43
KERATIN, TYPE II CYTOSKELETAL 2 EPIDERMAL	IPI00021304.1	1.00	3.07	3.13	6.1	2.17	0.88
CARCINOEMBRYONIC ANTIGEN-RELATED CELL ADHESION MOLECULE 5 PRECURSOR	IPI00027486.3	1.00	2.74	3.88	1.6	-1	-1
KERATIN, TYPE II CYTOSKELETAL 1	IPI00220327.3	1.00	2.67	1.62	1.82	3.36	3.87
PROLIFERATION-INDUCING GENE 21 OR RACK1	IPI00784408.1	0.99	2.64	-1	-1	2.74	2.54
DIPEPTIDASE 1 PRECURSOR	IPI00059476.5	1.00	2.63	2.19	3.07	-1	-1
COMPLEMENT COMPONENT 1 Q SUBCOMPONENT-BINDING PROTEIN, MITOCHONDRIAL PRECURSOR	IPI00014230.1	1.00	2.60	2.16	1.32	3.01	3.92
TRANSMEMBRANE EMP24 DOMAIN-CONTAINING PROTEIN 10 PRECURSOR	IPI00028055.4	0.99	2.59	2.87	-1	2.15	2.75
HSPA5 PROTEIN	IPI00003362.2	1.00	2.49	2.51	1.57	2.91	2.95
PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A	IPI00419585.9	1.00	2.47	2.94	2.91	1.91	2.13
ISOFORM 1 OF WD REPEAT PROTEIN 1	IPI00746165.2	1.00	2.46	2.41	-1	1.9	3.07
T-COMPLEX PROTEIN 1 SUBUNIT ZETA	IPI00027626.3	0.99	2.46	-1	-1	2.23	2.68
ATP SYNTHASE SUBUNIT G, MITOCHONDRIAL	IPI00027448.3	0.96	2.45	2	-1	3.23	2.12
GLYCOGEN PHOSPHORYLASE, BRAIN FORM	IPI00004358.4	1.00	2.44	1.98	2.32	2.11	3.35
CALRETICULIN PRECURSOR	IPI00020599.1	0.98	2.43	2.12	-1	2.85	2.33
PEROXIREDOXIN-2	IPI00027350.3	1.00	2.36	3.56	1.66	2.21	2.02

PROTEIN	IPI Accession	PROT PROB	AVERAGE ASAPR	C13 A	Arg B	C13 A	Lys B
HEAT SHOCK PROTEIN HSP 90-BETA	IPI00414676.6	1.00	2.31	2.05	1.63	2.92	2.65
MICROSOMAL GLUTATHIONE S-TRANSFERASE 1	IPI00021805.1	1.00	2.28	1.78	2.12	2.88	2.35
ATP SYNTHASE O SUBUNIT, MITOCHONDRIAL PRECURSOR	IPI00007611.1	1.00	2.27	2.09	1.9	2.3	2.8
SIMILAR TO PEPTIDYLPROLYL ISOMERASE A ISOFORM 1	IPI00741973.1	0.92	2.20	1.28	-1	2.39	2.93
40S RIBOSOMAL PROTEIN S15A	IPI00221091.9	1.00	2.19	2.54	-1	1.43	2.59
60S RIBOSOMAL PROTEIN L7	IPI00030179.3	0.95	2.17	1.61	2.43	3.49	1.15
ISOFORM LONG OF SODIUM/POTASSIUM-TRANSPORTING ATPASE ALPHA-1 CHAIN PRECURSOR	IPI00006482.1	1.00	2.15	2.29	2.83	1.71	1.77
PEPTIDYLPROLYL ISOMERASE B PRECURSOR	IPI00646304.4	1.00	2.13	0.86	2.94	2.29	2.43
THIOREDOXIN	IPI00216298.6	1.00	2.11	-1	-1	2.09	2.12
ADP-RIBOSYLATION FACTOR 6	IPI00215920.8	1.00	2.08	-1	2.24	1.95	2.04
STRESS-70 PROTEIN, MITOCHONDRIAL PRECURSOR / MORTALIN 2	IPI00007765.5	0.99	2.03	2.46	-1	1.24	2.39
GLUCOSE-6-PHOSPHATE ISOMERASE	IPI00027497.5	1.00	1.98	2.4	0.98	2.4	2.12
MALATE DEHYDROGENASE, MITOCHONDRIAL PRECURSOR	IPI00291006.1	1.00	1.96	2.47	1.05	-1	2.37
RAS-RELATED PROTEIN RAB-7	IPI00016342.1	1.00	1.94	2.38	0.18	2.43	2.75
MALATE DEHYDROGENASE, CYTOPLASMIC	IPI00291005.8	1.00	1.90	2.48	0.57	1.92	2.63
PROFILIN-1	IPI00216691.5	1.00	1.89	2.85	0.89	2	1.83
SELENIUM BINDING PROTEIN 1	IPI00745729.1	1.00	1.86	1.9	0.69	2.44	2.41
DELTA(3,5)-DELTA(2,4)-DIENOYL-COA ISOMERASE, MITOCHONDRIAL PRECURSOR	IPI00011416.2	0.98	999.00	999	999		-1
UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L3	IPI00011250.3	1.00	999.00	999	-1		999
CYTIDYLATE KINASE	IPI00514049.1	0.99	500.87	999	2.74		-1
ISOFORM 1 OF 40S RIBOSOMAL PROTEIN S24	IPI00029750.1	0.98	500.52	-1	999	2.03	
SIMILAR TO ACTIN, ALPHA 2, SMOOTH MUSCLE, AORTA	IPI00737449.1	0.95	4.34	1.68		7.44	3.9
SIGNAL RECOGNITION PARTICLE 9 KDA PROTEIN	IPI00642816.2	1.00	3.81	2.9		4.71	-1
PROTEASOME SUBUNIT BETA TYPE 2	IPI00028006.1	1.00	3.75	3.81	3.68		-1
IMPORTIN BETA-1 SUBUNIT	IPI00001639.2	1.00	3.73	3.47	3.99		-1
HEBP2 PROTEIN (FRAGMENT)	IPI00644697.2	1.00	3.05	4.09	2	-1	
10 KDA HEAT SHOCK PROTEIN, MITOCHONDRIAL	IPI00220362.5	1.00	2.87	2.34	-1	3.39	
60S RIBOSOMAL PROTEIN L18A	IPI00026202.1	0.92	2.84	2.85	2.83	-1	
ENDOPLASMIC RETICULUM PROTEIN ERP29 PRECURSOR	IPI00024911.1	0.98	2.65	-1		3.28	2.02
CALNEXIN PRECURSOR	IPI00020984.1	1.00	2.39	2.29	1.4	3.47	
FRUCTOSE-BISPHOSPHATE ALDOLASE A	IPI00465439.5	1.00	2.39		-1	2.43	2.34

PROTEIN	IPI Accession	PROT PROB	AVERAGE ASAPR	C13 A	Arg B	C13 A	Lys B
40S RIBOSOMAL PROTEIN S13	IPI00221089.5	1.00	2.34	2.12	-1	2.55	
CYTOCHROME C	IPI00465315.6	1.00	2.32	2.59	-1	2.04	
60S RIBOSOMAL PROTEIN L38	IPI00215790.6	1.00	2.27	2.44	-1	2.09	
PROTEIN DISULFIDE-ISOMERASE A4 PRECURSOR	IPI00009904.1	1.00	2.20	2.32	2.52	1.77	
UDP-GLUCOSE 6-DEHYDROGENASE	IPI00031420.3	1.00	2.20	2.12		2.38	2.09
GLUCOSIDASE 2 SUBUNIT BETA PRECURSOR	IPI00026154.2	0.94	2.11	1.27		2.26	2.79
150 KDA OXYGEN-REGULATED PROTEIN PRECURSOR	IPI00000877.1	1.00	500.68	2.36	999		
DOLICHYL-DIPHOSPHOOLIGOSACCHARIDE-PROTEIN GLYCOSYLTRANSFERASE PRECURSOR	IPI00297084.7	1.00	500.57	999	2.14		
SODIUM/POTASSIUM-TRANSPORTING ATPASE SUBUNIT BETA-3	IPI00008167.1	1.00	500.57	999	2.14		
RIBOSOMAL PROTEIN S10	IPI00478810	1.00	500.55	999	2.1		
17 KDA PROTEIN	IPI00147581.4	1.00	32.62			38.09	27.15
HEMOGLOBIN SUBUNIT ALPHA	IPI00410714.5	0.94	5.39	3.11		7.67	
ISOFORM 1 OF SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION 3	IPI00784414.1	0.97	4.14			3.08	5.19
ISOFORM 1 OF SERUM ALBUMIN PRECURSOR	IPI00745872.2	1.00	3.91	5.75	2.06		
OLFACTOMEDIN 4 PRECURSOR	IPI00022255.1	1.00	3.81	2.77	4.84		
55 KDA PROTEIN	IPI00657680.1	1.00	3.51			2.94	4.07
RAS-RELATED PROTEIN RAB-25	IPI00027993.2	1.00	3.30			2.42	4.18
ADP/ATP TRANSLOCASE 1	IPI00022891.3	1.00	2.98			2.82	3.14
HEAT SHOCK 70 KDA PROTEIN 1	IPI00304925.4	1.00	2.71			2.29	3.12
ALPHA-ACTININ-1	IPI00013508.5	0.99	2.61			3.15	2.06
ALPHA-SOLUBLE NSF ATTACHMENT PROTEIN	IPI00009253.2	0.99	2.59	2.34			2.83
CREATINE KINASE, UBIQUITOUS MITOCHONDRIAL PRECURSOR	IPI00658109.1	1.00	2.56	3.04	2.07		
UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 7	IPI00554701.2	1.00	2.41	2.57		2.25	
T-COMPLEX PROTEIN 1 SUBUNIT EPSILON	IPI00010720.1	1.00	2.36	2	2.72		
60 KDA HEAT SHOCK PROTEIN, MITOCHONDRIAL PRECURSOR	IPI00784154.1	1.00	2.33	2.52	2.14		
ISOFORM ALPHA-6X1X2B OF INTEGRIN ALPHA-6 PRECURSOR	IPI00010697.1	1.00	2.32	2.25	2.39		

Table 5. 3 : Potentially interesting proteins suitable for validating the results of the SILAC proteomic profile of MV's secreted by RG/C2 and PRG/C2 cell lines.

Table to show a list of 86 proteins that were found in at least 2 SILAC experiments with an average ASAPRatio >2.00. All the proteins in the list have an average protein probability score >0.90. Grey cell = protein not identified in experiment. ASAPRatio of -1 = protein identified in experiment but ratio couldn't be calculated.

(a) Biological Processes (GO level 8)

Biological Function	Genes list	NºGenes
apoptosis	PDIA3 HSPA1A PRDX2 HSPA5 CYCS ALB CALR HSPD1 ANXA5 HSPE1 ACTN1 VCP	12
regulation of programmed cell death	PDIA3 HSPA1A PRDX2 HSPA5 CYCS ALB CALR HSPD1 ANXA5 HSPE1 ACTN1 VCP	12
proton transport	ATP5L ATP5A1 ATP1A1 ATP5B ATP5O	5
glucose metabolic process	ALDOA UGDH MDH1 MDH2 GPI	5
protein targeting	PDIA3 KPNB1 CALR HSPD1	4
purine nucleoside triphosphate biosynthetic process	ATP5L ATP5A1 ATP5B ATP5O	4
hexose catabolic process	ALDOA MDH1 MDH2 GPI	4
ribonucleoside triphosphate biosynthetic process	ATP5L ATP5A1 ATP5B ATP5O	4
purine ribonucleotide biosynthetic process	ATP5L ATP5A1 ATP5B ATP5O	4
purine ribonucleoside triphosphate metabolic process	ATP5L ATP5A1 ATP5B ATP5O	4
regulation of transcription, DNA-dependent	PFN1 STAT3 CALR	3
proteolysis involved in cellular protein catabolic process	UCHL3 PSMB2 VCP	3
epidermis development	KRT9 KRT10 KRT2	3
transcription from RNA polymerase II promoter	PFN1 STAT3	2
ER to Golgi vesicle-mediated transport	ARF6 NAPA	2
cortical actin cytoskeleton organization and biogenesis	CALR ARF6	2
tricarboxylic acid cycle	MDH1 MDH2	2
malate metabolic process	MDH1 MDH2	2
potassium ion transport	ATP1A1 ATP1B3	2
nuclear import	PDIA3 KPNB1	2
sodium ion transport	ATP1A1 ATP1B3	2

(b) KEGG pathways

KEGG Pathway	Genes list	NºGenes
Ribosome	RPS10 RPL38 RPL7 RPS13 RPS15A RPL30 RPS24 RPL18A	8
Antigen processing and presentation	PDIA3 HSPA1A HSPA5 HSP90AB1 CALR HSP90AA1 CANX	7
Oxidative phosphorylation	ATP5L ATP5A1 COX2 ATP5B ATP5O	5
Regulation of actin cytoskeleton	ITGA6 PFN1 GNG12 ACTN1	4
Cell Communication	KRT9 ITGA6 KRT10 KRT2	4
Carbon fixation	ALDOA MDH1 MDH2	3
MAPK signaling pathway	HSPA1A HSPA5 GNG12	3
Prion disease	HSPA5 HSPD1	2
Reductive carboxylate cycle (CO2 fixation)	MDH1 MDH2	2
Cholera - Infection	PDIA4 ARF6	2
Glyoxylate and dicarboxylate metabolism	MDH1 MDH2	2
Pyruvate metabolism	MDH1 MDH2	2
Citrate cycle (TCA cycle)	MDH1 MDH2	2
Pyrimidine metabolism	CMPK ENTPD6	2
Glycolysis / Gluconeogenesis	ALDOA GPI	2
Starch and sucrose metabolism	UGDH GPI	2
Focal adhesion	ITGA6 ACTN1	2
Pentose phosphate pathway	ALDOA GPI	2

(c) Cellular Components (GO level 8)

Cellular Component	Genes list	NºGenes
mitochondrion	ATP5L C1QBP HSPA1A ATP5A1 ECH1 SLC25A4 COX2 CYCS HSP90AB1 HSPD1 CKMT1A MDH2 ATP5B ATP5O MGST1 HSPE1	16
cytosol	RPS10 HSPA5 RPL7 CYCS CALR RPS13 MDH1 HSP90AA1 CCT5 RPS15A PSMB2 HBA1 RPS24 RPL18A VCP	15
endoplasmic reticulum	HYOU1 PDIA3 HSPA1A PPIB PDIA4 HSPA5 DPEP1 CALR DDOST ERP29 CANX NAPA MGST1 PRKCSH VCP	15
nucleus	PFN1 HSPA1A KPNB1 STAT3 HSPA5 CYCS CMPK SH3BGRL2 PSMB2 PREP VCP	11
organelle membrane	ATP5A1 KPNB1 SLC25A4 HSPA5 COX2 CYCS DDOST ATP5B ATP5O MGST1	10
ribosome	RPS10 RPL38 RPL7 RPS13 RPS15A RPL30 RPS24 RPL18A	8
organelle envelope	ATP5A1 KPNB1 SLC25A4 COX2 CYCS ATP5B ATP5O MGST1	8
cytoskeleton	KRT9 PFN1 KRT10 WDR1 KRT2 ACTN1	6
integral to plasma membrane	ITGA6 CEACAM5 SLC25A4 GPR109B ATP1A1 ATP1B3	6

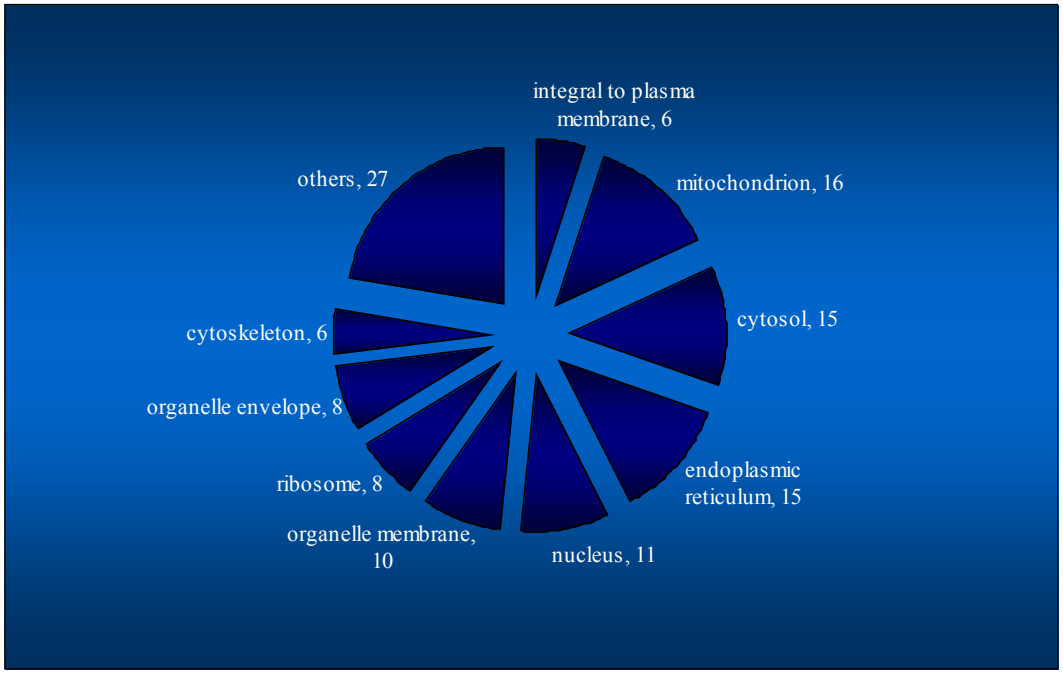


Figure 5. 5 : Analysis potentially interesting proteins with an average ASAPRatio ≥2.00 using software commonly used in microarray analysis, Babelomics.

The accession numbers for the potentially interesting proteins identified that could be candidates for validating the SILAC data were converted to gene names using online software called REFDIC. The gene names were then used to investigate which biological processes, cellular components and pathways these proteins are involved in Babelomics. (a) A table to show the biological processes with ≥2 proteins identified in this survey. The Table identifies the gene symbols of the proteins involved in each biological process. (b) A table to show the KEGG pathways with ≥2 proteins identified in this survey. The Table identifies the gene symbols of the proteins involved in each pathway. (c) Pie chart to show the number of proteins that are localized to various cellular compartments. The Table identifies the gene symbols of the proteins found in each cellular component.

5.6 Comparison with Previously Reported Secreted Protein Studies

Many proteomic profiling surveys of secreted proteins from various cancer cell lines have already been published but not many have compared the proteins secreted during early stage tumour progression. A literature search was performed in Pubmed for proteomic surveys to build a list of identified secreted proteins from various cancer cell lines to compare with all the proteins identified in this survey. Eight proteomic surveys of secreted proteins from pancreatic, breast, prostate, melanoma and osteosarcoma cell lines and one survey of proteins secreted in MVs from the HT29 cell line (human colorectal cancer cell line) were all used for this comparison. The venn diagram shown in Figure 5.6 shows the comparison of proteins in the first eight proteomic surveys (A) with those identified in MVs from the HT29 cell line (B) and the proteins identified in at least 2 out of 4 experiments in this study (C) (see Figure 5.6 for references). Supplementary Table 5c shows the 243 proteins out 564 proteins identified in this survey also previously been identified in the proteomic surveys outlined in A and B. Supplementary Table 5h shows the 101 proteins identified in both this survey and those identified in MVs from HT29 (Choi et al., 2007) and Supplementary Table 5i lists the proteins that were unique to this study.

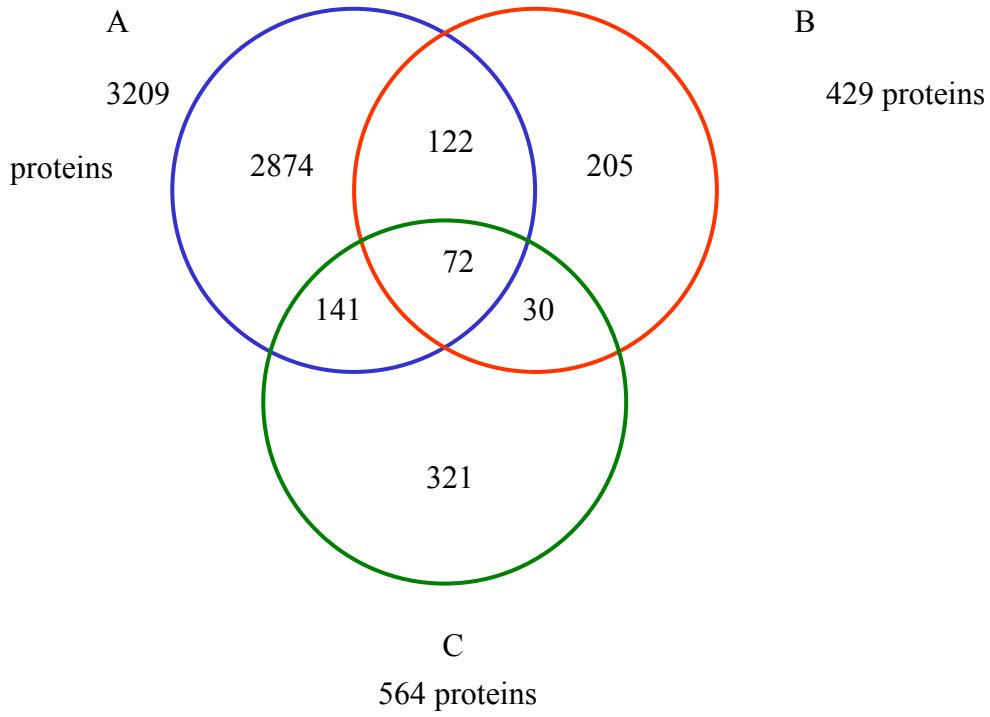


Figure 5. 6 : Comparison of proteins identified in RG/C2 and PRG/C2 SILAC proteomic profile with other similar studies to identify secreted proteins in CM derived from various cell lines.

The data generated in this SILAC study of MV's was compared to previously reported proteomic profiles of CM: A) Proteins identified in other secreted proteins proteomic surverys (Gronborg et al., 2006, Kulasingam and Diamandis, 2007, Mbeunkui et al., 2006, Mbeunkui et al., 2007, Mears et al., 2004, Ratajczak et al., 2006b, Sardana et al., 2008, C. C. Wu et al., 2008); B) Proteins identified in proteomic study of MV's derived from HT29 cell line (Choi et al., 2007); C) Proteins identified in MV's derived from RG/C2 and PRG/C2 cell lines generated in this study.

5.7 Conclusions

The aim of this chapter was to identify proteins released in MV by colorectal cancer cells and to also identify differentially secreted proteins which could have the potential to act as markers of early tumour progression using SILAC. The RG/C2 cell line was cultured in media containing either 'heavy' Arg or Lys and the PRG/C2 cell line was cultured in 'light' Arg and Lys for approximately 6 doublings. Figure 5.1 shows that the labeling efficiency of the technique was approx 97%. The MV fractions collected from both cell lines in experiment C13 ARG-A were then mixed 1:1 and the proteins were separated by 1D gel electrophoresis followed by in-gel digestion and LC MS/MS analysis. The results were analysed and filtered using the multiple thresholds filter parameters optimised in Chapter 4 using Bioworks Browser. Supplementary Table 5a shows there were 485 proteins with at least 2 unique peptides identified with an FDR of 0.00089 which was very similar to the test dataset used in Chapter 4. The Xpress function in Bioworks Browser calculates the relative abundance ratio from the area under the co-eluting 'light' and 'heavy' peaks. Figure 5.3 and Table 5.1 demonstrate that Xpress can sometimes yield misleading ratio values by using the wrong peaks to calculate the area under the peak. The ratios can be calculated manually but this is time consuming and requires a level of expertise and familiarity with the software highlighting the need for accurate and automated software to analyse the SILAC data. TPP was used to analyse all four SILAC experiments and the table in Figure 5.4a shows that a similar FDR could be achieved using a threshold of protein probability ≥0.90. The results from all four experiments were then compared using a threshold of protein probability ≥0.70. Table 5.2 shows there were 1396 proteins identified in total with 136, 252, 561 and 1396 proteins identified in 4, 3, 2 and 1 experiment and 86 proteins were found in multiple experiments with a consistent calculated ASAPRatio that would be potentially interesting to

investigate further (Table 5.3). The proteins identified in at least 2 out of 4 experiments (561 proteins) were compared to other proteomic surveys of secreted proteins from various cancer cell lines and one survey of MV from another colorectal cancer cell line, HT29. There were 101 proteins previously reported to be secreted in MV by cancer cell lines, these are listed in Supplementary Table 5h. In total there were 243 proteins previously reported and 321 proteins previously unreported (Supplementary Table 5i).

In order to investigate whether the proteins secreted in MV were involved in a particular pathway or have a common biological function, the IPI accession numbers were converted to gene symbols using REFDIC which were then analysed using FATIGO+. This analysis showed the two top biological processes were apoptosis and regulation of cell death for both the list of all proteins identified in at least two experiments and the list of potentially interesting proteins (Figures 5.5 and 5.6 respectively). The development of tumours is normally avoided by various mechanisms that trigger apoptosis in response to cellular stresses. Many cancer cells favour inhibition or inactivation of various key components of apoptosis to avoid cell death and allow proliferation. Many anti-apoptotic alterations in human tumour cells have previously been shown (Meijerink et al., 1998, Rampino et al., 1997, Reed, 1998, Wallace-Brodeur and Lowe, 1999, Yonish-Rouach et al., 1991) and the results shown here indicates a number of proteins identified in this survey are involved in apoptosis and the regulation of cell death to ensure the survival of not just the cancer cells themselves but also the other types of cells found within the tumour microenevironment.

The FATIGO+ analysis also showed that the proteins identified in this survey originate from a variety of cellular components. It also showed that the top two KEGG pathways for all proteins identified in multiple experiments were regulation of actin cytoskeleton and ribososmal pathways. The top two KEGG pathways associated with the list

of potentially interesting proteins was antigen processing and presentation and also ribosomal pathways. All these processes and pathways have been reported to play roles in tumour progression.

Cell migration and invasion are important steps in metastasis of cancer cells to distant sites to occur. They are triggered by a number of chemoattractants binding to cell surface receptors to stimulate intracellular signalling pathways that regulate the re-organisation of the actin cytoskeleton. Many proteins involved in this process have been linked invasive and metastatic cancer cells (Olson and Sahai, 2009, Yamaguchi and Condeelis, 2007).

Cancer cells are undergoing uncontrolled growth and proliferation and requires increased production of ribosomes for protein translation, this could explain the elevation of ribosomal proteins in this study . Whether this effect is a side effect of cancer cell transformation or directly involved in tumourigenesis to cause transformation is poorly understood but mutations in individual ribosomal protein genes have been partially correlated with an increased risk of cancer (Dai and Lu, 2008, Warner and McIntosh, 2009).

Cancer cells are transformed normal cells that are undergoing uncontrolled growth and proliferation and evidence has shown that the immune response to tumour antigens is very similar to those of 'self' antigens. With increasing knowledge of the steps involved in tumour progression and the features that distinguish cancer cells from normal cells there is evidence evidence to show how the immune system reacts to the changes occurring within the cell (Hijikata et al., 2007).

The results described in this chapter indicate that elevated levels of proteins involved in the regulation of the actin cytoskeleton, apoptosis and regulation of cell death and ribosomal proteins are secreted in MVs from colorectal cancer cell lines could play a role in early tumour progression.

Chapter 6

RESULTS

Validation of selected proteins identified in the SILAC profile of MV released by colorectal cancer cell lines during early tumour progression

Chapter 6 : Validation of selected proteins identified in the SILAC profile of MV released by colorectal cancer cell lines during early tumour progression

6.1 Introduction

In Chapter 5, SILAC was used to investigate proteins that were differentially secreted in MVs from a colorectal cancer cell model of early tumour progression to identify proteins that could be potential markers for early stage CRC. Table 5.3 lists 86 proteins that were shown to be differentially secreted in MV's released form the RG/C2 and PRG/C2 cell lines. Six proteins were selected from this list to validate the SILAC results using Western blotting: Anterior Gradient Homolog 2 (AGR2); Olfactomedin 4 (OLFM4); Selenium-Binding protein 1 (SBP1); Heat Shock protein 90 alpha (HSP90α), Heat Shock protein 90 beta (HSP90β) and Carcinoembryonic antigen-related cell adhesion 5 (CEACAM5). There were also three proteins selected from supplementary Table 5b with an average ASAPRatio of 1.00 to use as control validations. These were Tumour Associated Calcium Signal Transducer 1 precursor (TACSTD1) (also known as Epithelial Cell Adhesion molecule, EpCAM), Isoform 1 of Heat Shock Cognate 71 kDa protein (HSC70) and Annexin A1. The criteria for the selection of proteins were as follows: proteins identified in multiple experiments with relatively consistent ASAPRatios; proteins with an average protein probability score greater than 0.90; proteins with multiple peptides identified in each of the experiments; validation proteins with an average ASAPRatio >1.50 (except control proteins that required an average ASAPRatio ~1.00); and proteins with readily available antibodies were selected for validation by Western blotting. Tables 6.1a and 6.1b summarises the average protein probability, average ASAPRatio and the ASAPRatios calculated in each experiment for all nine proteins described

(a)

Protein Name	Accession Number	Average		C13 ARG		C13 LYS	
		Prot		ASAPRatio		ASAPRatio	
		Prob	ASAPRatio	A	B	A	B
Carcinoenbryonic Antigen-related Cell Adhesion molecule 5 precursor CEACAM5)	IPI00027486	1.00	2.74	3.88	1.6		
Anterior Gradient homolog 2 (AGR2)	IPI00007427	0.99	3.40	2.07	3.19	3.27	5.08
Olfactomedin 4 precursor (OLFM4)	IPI00022255	1.00	3.81	2.77	4.84		
Selenium Binding Protein 1 (SBP1)	IPI00745729	1.00	1.86	1.9	0.69	2.44	2.41
Heat Shock Protein HSP 90-alpha (HSP90α)	IPI00784295	1.00	3.08	2.55	5.11	2.23	2.43
Heat Shock Protein HSP 90-beta (HSP90β)	IPI00414676	1.00	2.31	2.05	1.63	2.92	2.65

(b)

Protein Name	Accession Number	Average		C13 ARG		C13 LYS	
		Prot		ASAPRatio		ASAPRatio	
		Prob	ASAPRatio	A	B	A	B
Tumour-Associated Calcium Signal Transducer 1 precursor (TACSTD1)	IPI00296215	1.00	1.44	1.85		1.54	0.94
Isoform 1 of Heat Shock Cognate 71 kDa Protein (HSC70)	IPI0000386	1.00	1.29	1.23	1.17	1.23	1.54
Annexin A1	IPI00218918	1.00	1.18		0.96	1.55	1.04

Table 6. 1 : Proteins chosen to validate ASAPRatios calculated by TPP.

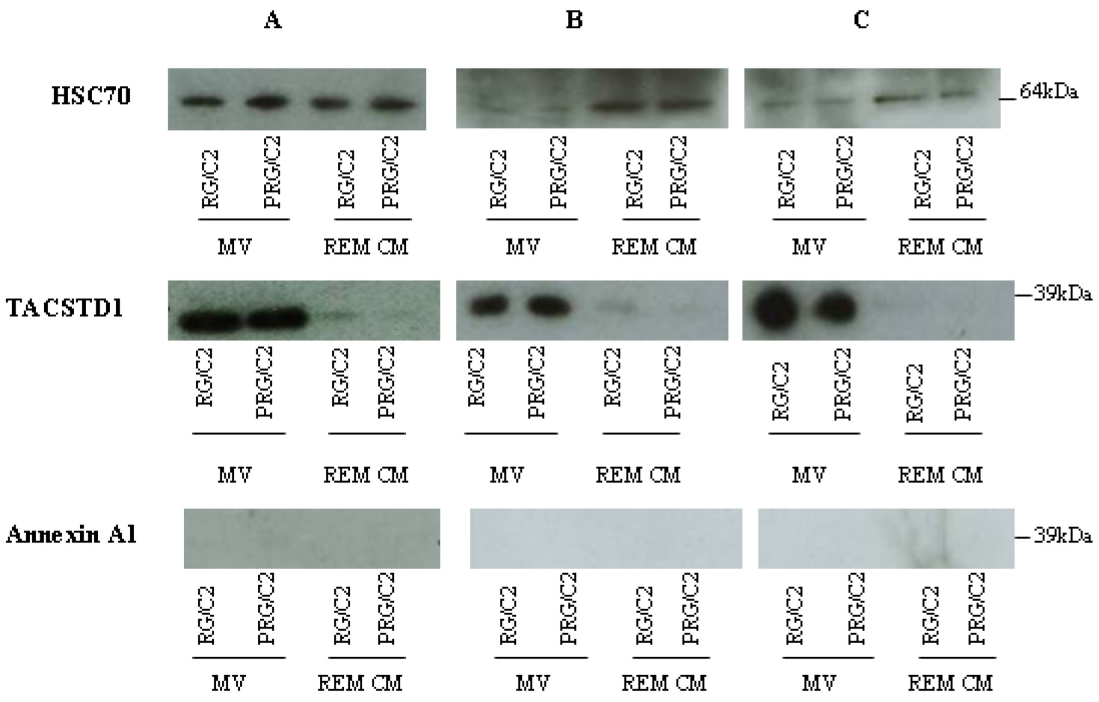
Table 6.1 shows the average protein ratio, average protein probability and ASAPRatios for each of the 4 SILAC experiments. Table (a) shows 6 proteins with an increased secretion in MV's from PRG/C2 compared to RG/C2. Table (b) shows 3 proteins identified with a 1:1 ratio. The grey boxes indicate the protein was identified but ASAPRatio could not be calculated. The black boxes indicate the protein was not identified.

above. These proteins were all used to validate the results of the SILAC experiments by Western blotting. CM was collected from cells grown under identical conditions to the SILAC experiment using unlabeled media supplemented with dialysed FBS. MV and REM CM fractions were prepared as previously described and Western blotting was used to confirm the presence of all eight proteins in an attempt to confirm the differential secretion of the six proteins mentioned above. MV fractions were prepared in biological triplicate and Western blot analysis for each of the proteins was performed for each experiment using the antibodies outlined in Table 2.2a. For each Western blot, 20μg of the total protein for both the MV and REM CM fractions was loaded onto the protein gel.

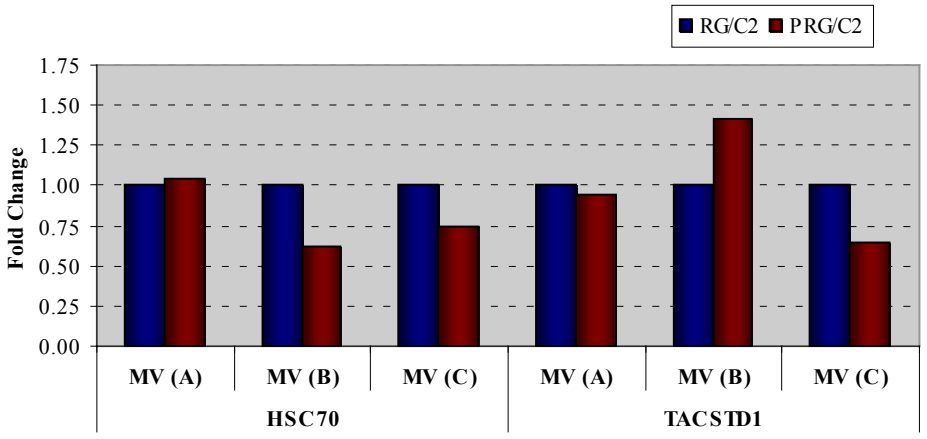
6.2 Tumour Associated Calcium Signal Transducer 1 precursor (TACSTD1), Isoform 1 of Heat Shock Cognate 71 kDa protein (HSC70) and Annexin A1

Figure 6.1 shows the Western blots for the three proteins chosen as controls in MV and REM CM fractions collected from the RG/C2 and PRG/2 cell lines. Supplementary Table 6a shows the TPP results for these proteins. HSC70 was detected by MS in all four SILAC experiments with ASAPRatios of 1.23, 1.17, 1.23 and 1.54 (average 1.29) with an average protein probability score of 1.00 (see Table 6.1b). Also, as shown in Supplementary Figure 6a, the peptides identified in HSC70 spanned the whole amino acid sequence suggesting that full length protein was secreted in MVs. Figure 6.1a shows that HSC70 (64kDa) could be detected to varying degrees in the MV and REM CM fractions in all of the triplicate experiments. The densitometry results shown in Figure 6.1b demonstrate a 1.04, 0.62 and 0.75 fold increase in MV's from the PRG/C2 cell line with an average of 0.80 compared to an ASAPRatio of 1.29.

(a)



(b)



Adj volume	2.70	2.80	4.82	3.00	5.53	4.14	9.38	8.85	4.54	6.46	8.97	5.76
Fold change	1.00	1.04	1.00	0.62	1.00	0.75	1.00	0.94	1.00	1.42	1.00	0.64

(c)

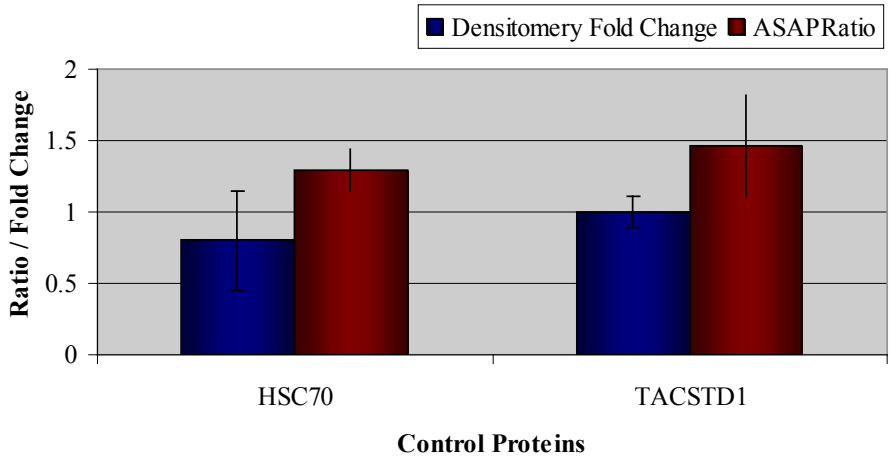


Figure 6. 1 : Confirmation of the presence of 3 proteins with an ASAPRatio of 1.00 using Western Blot analysis.

Western blot analysis was used to confirm the presence of three proteins with an ASAPRatio of 1.00 in MV's and the remaining CM (after MV removal) from RG/C2 and PRG/C2 cells ($5x10^6$ cells) grown in identical conditions to those used in the SILAC experiments. Fig (a) shows the Western blot results for TACSTD1, HSC70, Annexin A1. Fig (b) shows the densitometry results of the Western blots (shown in a) using a GS-800 calibrated densitometer (BIO-RAD) with fold change in PRG/C2 represented as a bar chart (A, B and C represent biological replicates). (c) Graph to show the combined results of the ASAPRatios from the SILAC experiments compared to the densitometry results from the Western blot validation experiments for the control proteins.

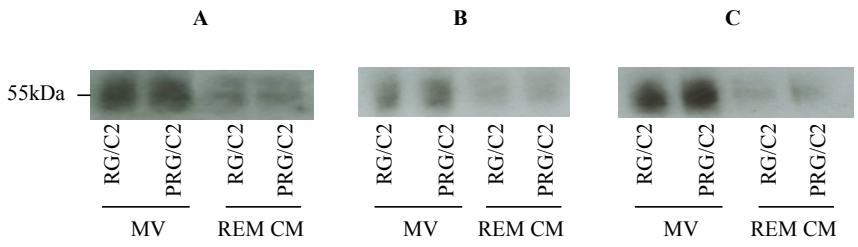
TACSTD1 was detected by MS in all four SILAC experiments with ASAPRatios calculated in three of them of 1.85, 1.54 and 0.94 (average 1.44) and an average protein probability score of 1.00 (see Table 6.1b). Also, as shown in Supplementary Figure 6b, the peptides identified in TACSTD1 spanned the whole amino acid sequence suggesting that full length protein was secreted in MVs. Figure 6.1a also shows that a single band for TACSTD1 (39kDa) was detected in the MV fractions of both cell lines in each of the biological triplicates with no bands detected in the REM CM fraction. This suggests that TACSTD1 is primarily found in MV's from both cell lines. The densitometry results also show a 1.42 a fold increase in the PRG/C2 in one of the experiments and a decrease in secretion in 2 experiments (0.94 and 0.64 fold). This gives an average fold change in band density of 1.00 compared to an average ASAPRatio of 1.44 as shown in Figure 6.1c.

Annexin A1 was detected by MS in all four SILAC experiments with ASAPRatios calculated for three experiments as 0.96, 1.55 and 1.04 (average 1.18) and the average protein probability score of 1.00 (see Table 6.1b). Also, as shown in Supplementary Figure 6c, the peptides identified in Annexin A1 spanned the whole amino acid sequence suggesting that full length protein was secreted in MVs. Figure 6.1a shows that the third control protein, Annexin A1 (39kDa) could not be detected by Western blot analysis in any of the biological triplicate experiments rabbit polyclonal AbCam antibody (ab33061) described in Table 2.2.

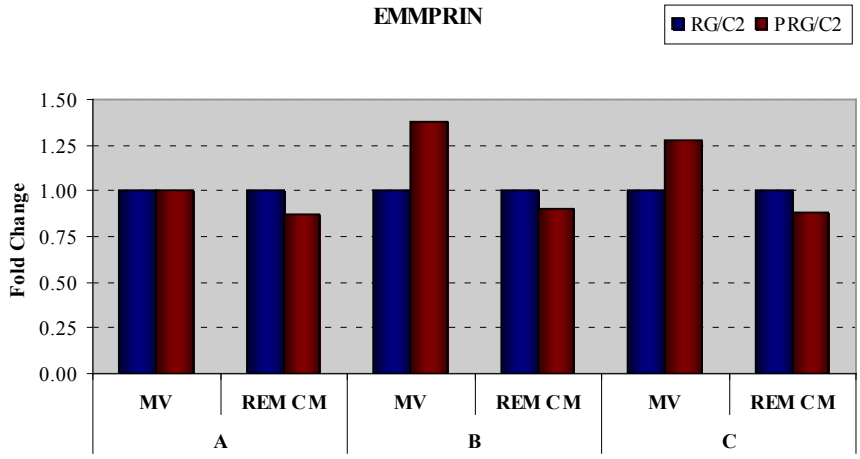
6.3 Extracellular Matrix Metalloproteinase Inducer (EMMPRIN)

As described in Chapter 3, EMMPRIN have been shown to be secreted in MV's and is thought to play an important role in tumorigenesis which was used as a positive control for MV preparation. Therefore, Western blot analysis for EMMPRIN was performed for each of the biological triplicate experiments prepared for this Chapter. Figure 6.2 shows that full length EMMPRIN (55kDa) could be detected in both the MV and REM CM fractions

(a)



(b)



Adj volume	29.24	29.34	13.95	12.16	3.68	5.06	1.77	1.60	25.19	32.23	4.47	3.95
Fold change	1.00	1.00	1.00	0.87	1.00	1.37	1.00	0.90	1.00	1.28	1.00	0.88

Figure 6. 2 : Confirmation of the presence of EMMPRIN in MV's prepared to validate the SILAC results.

Western blot analysis was used to confirm the presence of EMMPRIN in MV's and the remaining CM (after MV preparation) from RG/C2 and PRG/C2 ($5x10^6$ cells) grown in identical conditions used in the SILAC experiments. Fig (a) shows the Western blots. Fig (b) shows the densitometry results of the blots using a GS-800 calibrated densitometer (BIO-RAD). A, B and C represent biological replicates.

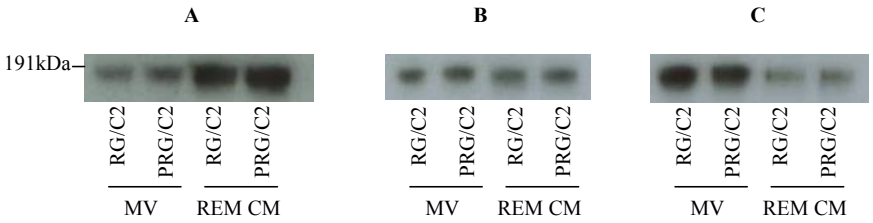
although the bands are more intense in the MV fractions in each of the triplicate experiments. The densitometry results shown in Figure 6.2b show there was a 1.00, 1.37 and 1.28 fold increase in the MV fraction in the PRG/C2 cell line with an average increase of 1.22. The Western blot analysis for EMMPRIN demonstrates its presence in the MVs prepared in this study but could not be detected by MS in the SILAC experiments. This could be because it is present a too low a level to be detected in these samples without further fractionation.

6.4 Carcinoembryonic antigen-related cell adhesion 5 (CEACAM5)

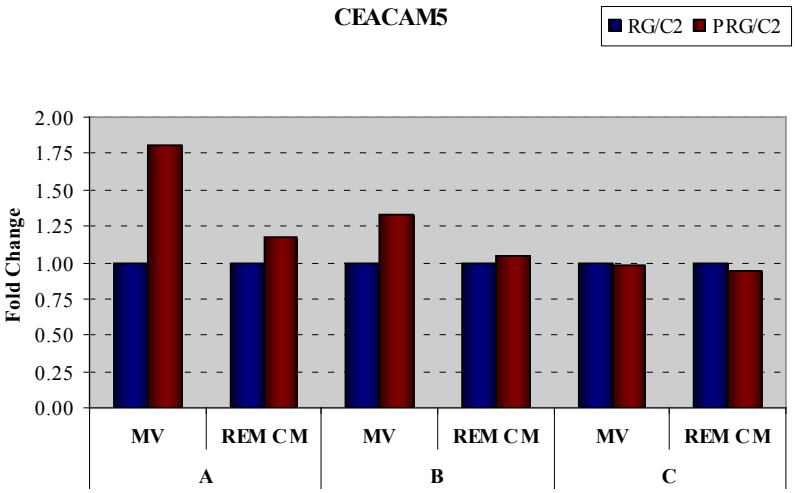
CEA is the protein product for the Carcinoembryonic antigen-related cell adhesion 5 (CEACAM5 or CD66e) gene. CEA is a ~200kDa highly glycosylated protein that was first discovered in tissue extracts from CRC patients by Gold and Freedman in 1965 who then later discovered it's presence in serum from CRC patients (Gold and Freedman, 1965a, Thomson et al., 1969). However, CEA has since been shown to be overexpressed in other cancer types such as pancreatic, liver, lung, breast, ovarian, bladder and prostate (Chan and Stanners, 2007). CEA already has limited uses as a biomarker for CRC in the clinic (see Section 1.1). The function of CEA is not yet fully understood but it has been demonstrated that it can act as homophilic and heterophilic cell adhesion molecules when expressed on the tumour cell surface (Benchimol et al., 1989). It has also been suggested that CEA could play an important role in protecting the colonic mucosa from microbial invasion (Hammarstrom, 1999).

CEACAM5 was detected by MS in all four SILAC experiments with ASAPRatios calcluted in two of them as 1.66 and 3.88 (average 2.74) and an average protein probability score of 1.00 (see Table 6.1a). Although the ASAPRatios were not similar in both experiments, they both showed an increase in MVs released by the PRG/C2 cell line. Also, CEA (the protein product of the CEACAM5) is a currently used biomarker for CRC and it was therefore encouraging to detect it by MS in this study. Supplementary Figure 6d shows

(a)



(b)



Adj volume	2.11	3.82	13.00	15.03	3.92	5.20	5.02	5.22	16.39	15.99	3.80	3.60
Fold change	1.00	1.81	1.00	1.18	1.00	1.33	1.00	1.04	1.00	0.98	1.00	0.95

Figure 6. 3 : Confirmation of the presence of CEACAM5 in MV's prepared to validate the SILAC results.

Western blot analysis was used to confirm the presence of CEACAM5 in MV's and the remaining CM (after MV preparation) from RG/C2 and PRG/C2 ($5x10^6$ cells) grown in identical conditions used in the SILAC experiments. Fig (a) shows the Western blots. Fig (b) shows the densitometry results of the blots using a GS-800 calibrated densitometer (BIO-RAD). A, B and C represent biological replicates.

that the peptides identified spanned the whole amino acid sequence suggesting that full length CEA was secreted in MVs. Supplementary Table 6b shows the results of the TPP analysis for CEACAM5. The Western blot analysis shown in Figure 6.3a demonstrated the presence of full length CEA (~200kDa) in both the MV and REM CM fractions in each of the triplicate experiments. The densitometry results shown in Figure 6.3b demonstrate a 1.81, 1.33 and a 0.94 fold increase in CEA release in the MV fraction from the PRG/C2 cell line with an average fold increase of 1.37.

6.5 Anterior Gradient Homolog 2 protein (AGR2)

Anterior Gradient Homolog 2 protein (AGR2) (also known as hAG-2), is the 20kDa human orthologue of the Xenopus laevis AGR protein XAG-2 in which it plays a role in development (Thompson and Weigel, 1998). Its function in humans is unclear however, it has been shown to attenuate p53 activation when over-expressed in UV-damaged cancer cells leading to the suggestion it plays a role in survival (Pohler et al., 2004). AGR2 has also been shown to enhance adhesion and metastatic potential in rat mammary cells when over-expressed (Liu et al., 2005). It has also recently been suggested to be essential in the production of intestinal mucus (Park et al., 2009). AGR2 expression has been detected in breast, prostate and lung cancers and also has been suggested as a potential marker for the presence of circulating tumour cells in the blood of patients with metastatic disease (Fletcher et al., 2003, Smirnov et al., 2005, Thompson and Weigel, 1998, Zhang et al., 2005b, Zhu et al., 2007). AGR2 was also identified as a secreted protein in CM collected from prostate and breast cancer cell lines in the proteomic studies referred to in Supplementary Table 5c (Kulasingam and Diamandis, 2007, Sardana et al., 2008).

AGR2 was detected by MS in all four SILAC experiments with ASAPRatios of 2.07, 3.19, 3.27 and 5.08 (average 3.40) and an average protein probability score of 0.99 (see Table

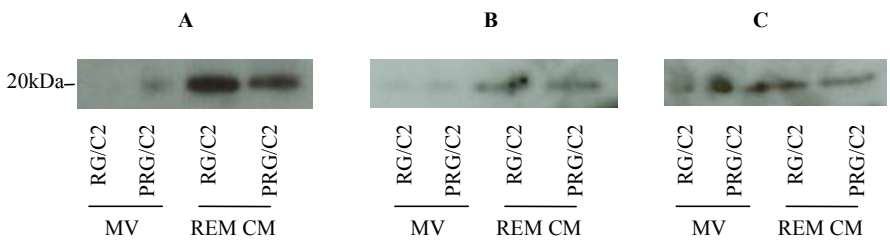
6.1a). Also, as shown in Supplementary Figure 6e, the peptides identified in AGR2 spanned the whole amino acid sequence suggesting that full length protein was secreted in MVs. The Western blot analysis shown in Figure 6.4a demonstrated the presence of full length AGR2 (~20kDa) in both the MV and REM CM fractions in each of the triplicate experiments although it was predominantly present in the REM CM fraction. The densitometry results shown in Figure 6.4b demonstrate a 2.04, 5.45 and a 2.44 fold increase in AGR2 release in the MV fraction from the PRG/C2 cell line with an average fold increase of 3.31. The fold change observed in the densitometry results are lower than those calculated in the SILAC experiments however, an increase ≥2.00 is observed in all experiments and the average ratios are similar. Therefore AGR2 shows potential as a protein that is differentially secreted in MVs in early tumour progression.

6.4 Olfactomedin 4 (OLFM4)

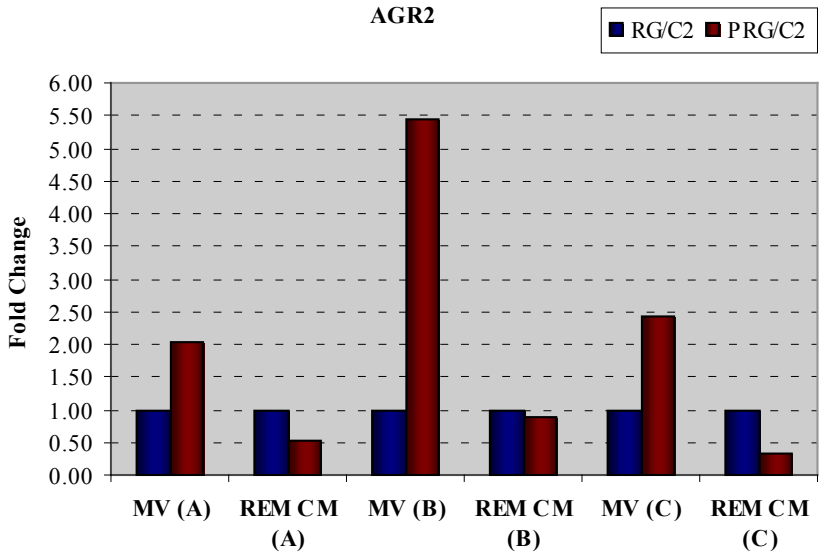
Olfactomedin 4 (OLFM4) (also known as hCG-1 and GW112) is 57kDa glycosylated protein which was originally cloned from human myeloblasts and found to be selectively expressed in inflamed colonic epithelium (Zhang et al., 2002). Compared to normal tissue samples, OLFM4 mRNA has been shown to be elevated 90%, 69% and 85% in colon, breast and lung cancer tissues respectively (Koshida et al., 2007). The function of OLFM4 is still unknown. However, several studies have suggested possible roles in proliferation and apoptosis and also as a marker of colon adenocarcinomas with cancer stem cell properties which are thought to act as tumour-initiating cells (Kobayashi et al., 2007, van der Flier et al., 2009, Zhang et al., 2004).

OLFM4 was detected by MS in two out of the four SILAC experiments with ASAPRatios of 2.77 and 4.84 (average 3.81) and an average protein probability score of 1.00 (see Table 6.1a). Also, as shown in Supplementary Figure 6f, the peptides identified in

(a)



(b)



Adj volume	0.65	1.33	9.83	5.11	0.04	0.22	2.97	2.66	2.42	5.89	7.75	2.63
Fold change	1.00	2.04	1.00	0.52	1.00	5.45	1.00	0.90	1.00	2.44	1.00	0.34

Figure 6. 4 : Confirmation of the presence of AGR2 in MV's prepared to validate the SILAC results.

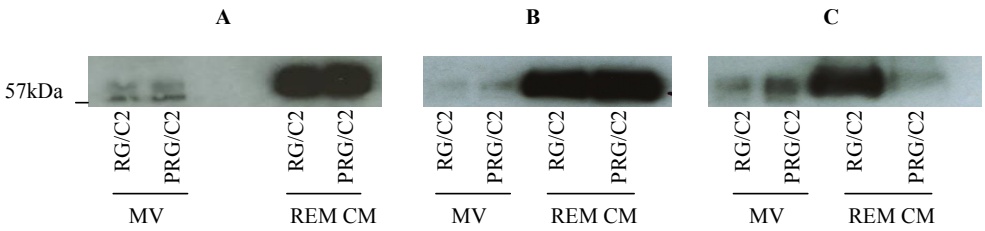
Western blot analysis was used to confirm the presence of AGR2 in MV's and the remaining CM (after MV preparation) from RG/C2 and PRG/C2 ($5x10^6$ cells) grown in identical conditions used in the SILAC experiments. Fig (a) shows the Western blots. Fig (b) shows the densitometry results of the blots using a GS-800 calibrated densitometer (BIO-RAD). A, B and C represent biological replicates.

OLFM4 spanned the whole amino acid sequence suggesting that full length protein was secreted in MVs. The Western blot analysis shown in Figure 6.5a demonstrated the presence of OLFM4 (~57kDa) in both the MV and REM CM fractions in each of the triplicate experiments although it was predominantly present in the REM CM fraction. The densitometry results shown in Figure 6.5b demonstrate a 1.67, 3.15 and a 1.96 fold increase in OLFM4 release in the MV fraction from the PRG/C2 cell line with an average fold increase of 2.26. The ASAPRatios calculated for OLFM4 in the SILAC experiments showed a high degree of variability with the ratio in the second experiment being double that calculated in the first experiment. The densitometry results from the Western blot analysis also showed a high degree of variability in the extent of the increase. However, the ratios in both cases are all increased and therefore OLFM4 warrants further investigation.

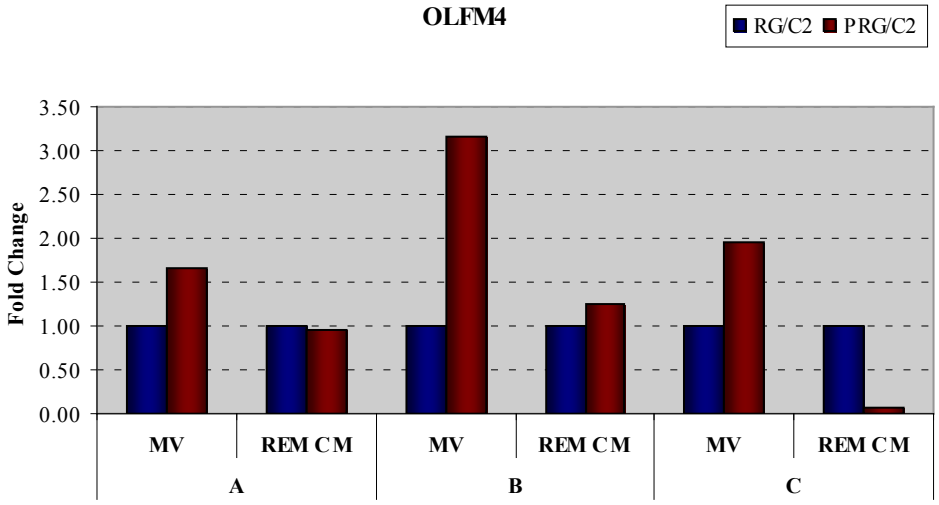
6.6 Selenium-binding Protein 1 (SBP1)

SBP1 is a 55kDa protein that belongs to the selenium-binding protein family and is the protein product of the SELENBP1 gene. Selenium has been suggested to play a role in preventing cancer and selenium deficiency has been linked to cancer (Russo et al., 1997). Decreased protein levels of SBP1 have been demonstrated in thyroid, CRC and ovarian cancer tumour tissue compared to normal tissue (Brown et al., 2006, Huang et al., 2006a, H. Kim et al., 2006). It has also been demonstrated that CRC patients exhibiting low levels of SELENBP1 expression had significantly lower overall survival rates and that this suppression of SELENBP1 activity is a late event in colorectal carcinogenesis and could contribute to rapid progression of CRC (Kim et al., 2006). SBP1 was also identified as a secreted protein in CM collected from a prostate cancer cell line in the proteomic studies referred to in Supplementary Table 5c (Sardana et al., 2008).

(a)



(b)

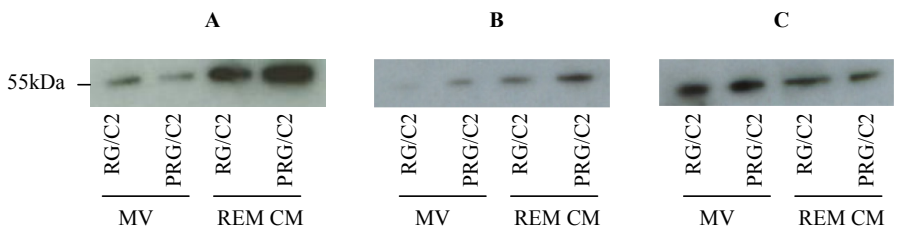


Adj volume	0.69	1.16	86.58	83.57	2.70	8.51	116.72	145.57	2.31	4.52	37.08	2.23
Fold change	1.00	1.67	1.00	0.97	1.00	3.15	1.00	1.25	1.00	1.96	1.00	0.06

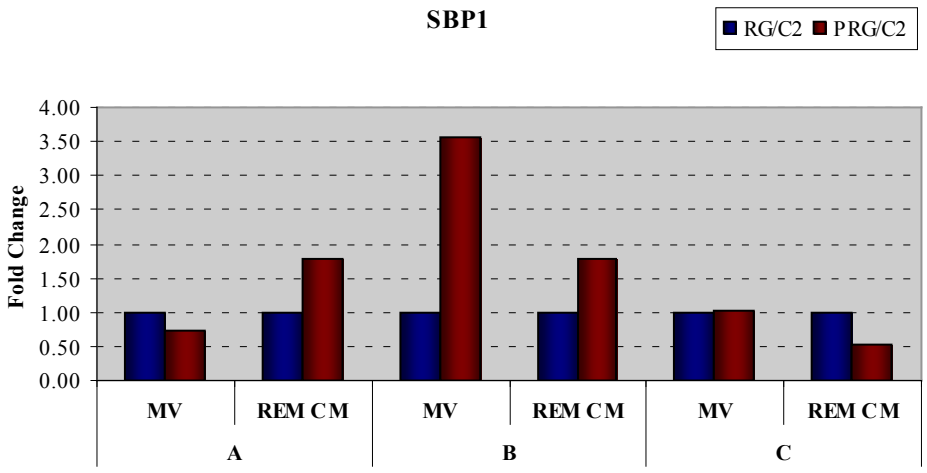
Figure 6. 5 : Confirmation of the presence of OLFM4 in MV's prepared to validate the SILAC results.

Western blot analysis was used to confirm the presence of OLFM4 in MV's and the remaining CM (after MV preparation) from RG/C2 and PRG/C2 ($5x10^6$ cells) grown in identical conditions used in the SILAC experiments. Fig (a) shows the Western blots. Fig (b) shows the densitometry results of the blots using a GS-800 calibrated densitometer (BIO-RAD). A, B and C represent biological replicates.

(a)



(b)



Adj volume	1.45	1.05	8.98	16.09	0.66	2.33	4.93	8.72	14.49	14.63	14.83	7.94
Fold change	1.00	0.73	1.00	1.79	1.00	3.56	1.00	1.77	1.00	1.01	1.00	0.54

Figure 6. 6 : Confirmation of the presence of SBP1 in MV's prepared to validate the SILAC results.

Western blot analysis was used to confirm the presence of SBP1 in MV's and the remaining CM (after MV preparation) from RG/C2 and PRG/C2 ($5x10^6$ cells) grown in identical conditions used in the SILAC experiments. Fig (a) shows the Western blots. Fig (b) shows the densitometry results of the blots using a GS-800 calibrated densitometer (BIO-RAD). A, B and C represent biological replicates.

SBP1 was detected by MS in all four SILAC experiments with ASAPRatios of 1.90, 0.69, 2.44 and 2.41 (average 1.86) and an average protein probability score of 1.00 (see Table 6.1a). Also, as shown in Supplementary Figure 6g, the peptides identified in SBP1 spanned the whole amino acid sequence suggesting that full length protein was secreted in MVs. The Western blot analysis shown in Figure 6.6a demonstrated the presence of SBP1 (~55kDa) in both the MV and REM CM fractions in each of the triplicate experiments and was predominantly present in different fractions in each. The densitometry results shown in Figure 6.6b demonstrate a 0.73, 3.56 and a 1.01 fold increase in SBP1 release in the MV fraction from the PRG/C2 cell line with an average fold increase of 1.77.

The ASAPRatios calculated for SBP1 in the SILAC experiments showed a high degree of variability in all four experiments ranging from a decrease in one experiment and increase in fold change in three experiments.The densitometry results for the Western blots also showed variability a decrease in one experiment and an increase and no change in the following two experiments. Due to the high degree of variability, SBP1 may not warrant further investigation.

6.6 Heat Shock Protein 90α and Heat Shock Protein 90β (HSP90α and HSP90β)

Heat shock proteins (HSPs) were originally identified because of their increased expression in response to cellular stress such as elevated temperature, heavy metals and oxidative stress. They are organised into families according to the molecular weight, such as HSP100, HSP90, HSP70, HSP60 etc. Some HSPs act as molecular chaperones which facilitate re-folding of proteins when they become damaged or target them for degradation if the damage is irreparable (Powers and Workman, 2007). HSP90 is an abundant molecular chaperone and accounts for 1-2% of total protein content. It functions to ensure the correct conformation, activity, intra-cellular localisation and proteolytic turnover of various proteins

involved in cell growth, differentiation and survival (Powers and Workman, 2006, Whitesell and Lindquist, 2005).

HSP90 is currently being investigated as a therapeutic target for the treatment of cancer as it is essential for the stability and function of many oncogenic client proteins (Powers and Workman, 2007). These client proteins include: receptor tyrosine kinases such as ErbB2/HER-2 and KIT; signalling molecules such as AKT and MET; transcription factors such as HIF-1α and ERα-receptors; and others proteins involved in cancer such as BCR-ABL, Bcl-2 and MMP2 (Sharp and Workman, 2006). HSP90 inhibition results in degradation of client proteins causing a down-regulation in their function affecting many aspects of the malignant phenotype therefore, HSP90 inhibitors could potentially be used in the treatment of cancer and are currently being investigated in various clinical trials (Powers and Workman, 2007).

HSP90α was also identified as a secreted protein in CM collected from various cancer cell lines such as breast, prostate, ovarian and bladder as well as MVs from the colorectal cancer cell line HT29 in the proteomic studies referred to in Supplementary Table 5c (Choi et al., 2007, Kulasingam and Diamandis, 2007, Sardana et al., 2008, Wu et al., 2008). HSP90β was also identified as a secreted protein in CM collected from a leukaemia cell line well as MVs from the colorectal cancer cell line HT29 in the proteomic studies referred to in Supplementary Table 5c (Choi et al., 2007, Wu et al., 2008).

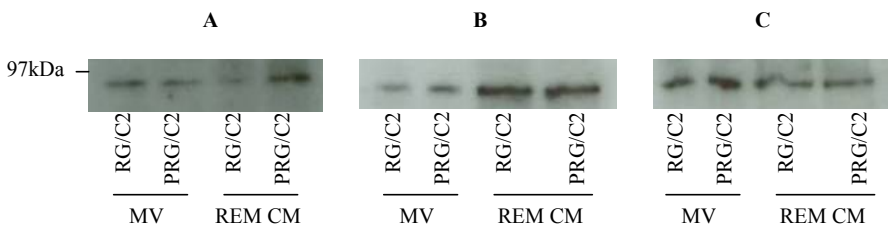
There are currently five isoforms of HSP90 described. The two major isoforms are HSP90α and HSP90β which share 85% sequence homology at the protein level and are both cytoplasmic. The other three isoforms are GRP94, TRAP1 and HSP90N.

HSP90α was detected by MS in all four SILAC experiments with ASAPRatios of 2.50, 5.11, 2.23 and 2.43 (average 3.08) and an average protein probability score of 1.00 (see

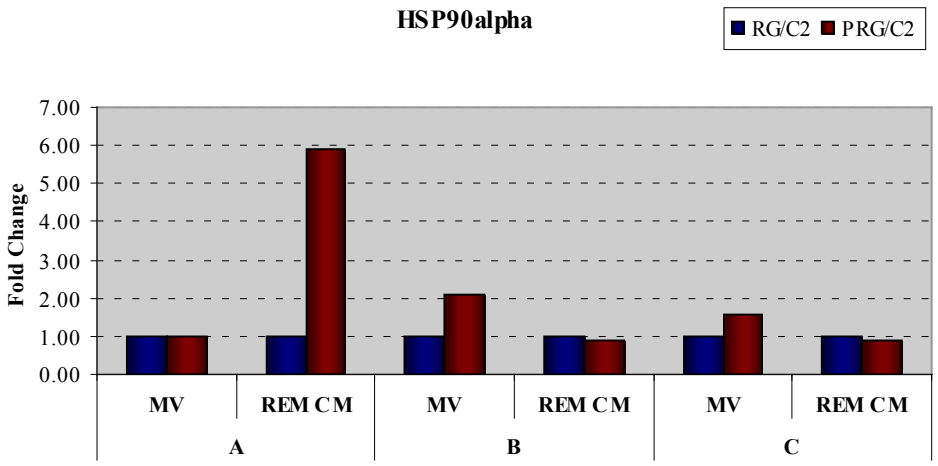
Table 6.1a). Also, as shown in Supplementary Figure 6h, the peptides identified in HSP90α spanned the whole amino acid sequence suggesting that full length protein was secreted in MVs. The Western blot analysis shown in Figure 6.7a demonstrated the presence of HSP90α (~97kDa) in both the MV and REM CM fractions in each of the triplicate experiments. The densitometry results shown in Figure 6.7b demonstrate a 0.98, 2.09 and a 1.58 fold increase in HSP90α release in the MV fraction from the PRG/C2 cell line with an average fold increase of 1.55. The ASAPRatios observed for HSP90α in the SILAC experiments were fairly consistent with three out four experiments showing an abundance ratio between 2.23 and 2.50. However, the densitometry results for the Western Blots showed a higher degree of variability with a smaller increase in two experiments and no change in the third. Although the great potential shown by the ASAPRatios in the SILAC experiments was not reflected in the Western blot analysis, HSP90α could warrant further investigation.

HSP90β was detected by MS in all four SILAC experiments with ASAPRatios of 2.05, 1.63, 2.92 and 2.65 (average 2.31) and an average protein probability score of 1.00 (see Table 6.1a). Also, as shown in Supplementary Figure 6i, the peptides identified in HSP90β spanned the whole amino acid sequence suggesting that full length protein was secreted in MVs. The Western blot analysis shown in Figure 6.8a demonstrated the presence of HSP90β (~97kDa) in both the MV and REM CM fractions in each of the triplicate experiments. The densitometry results shown in Figure 6.8b demonstrate a 1.51, 1.04 and a 0.97 fold increase in HSP90β release in the MV fraction from the PRG/C2 cell line with an average fold increase of 1.17.

(a)



(b)

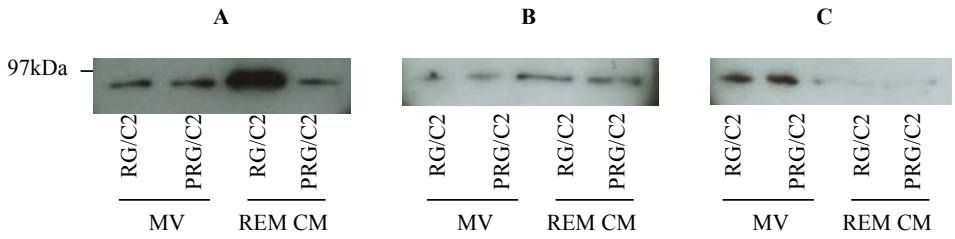


Adj volume	1.43	1.41	1.30	7.69	0.64	1.35	7.17	6.42	3.35	5.29	6.38	5.71
Fold change	1.00	0.98	1.00	5.92	1.00	2.09	1.00	0.90	1.00	1.58	1.00	0.90

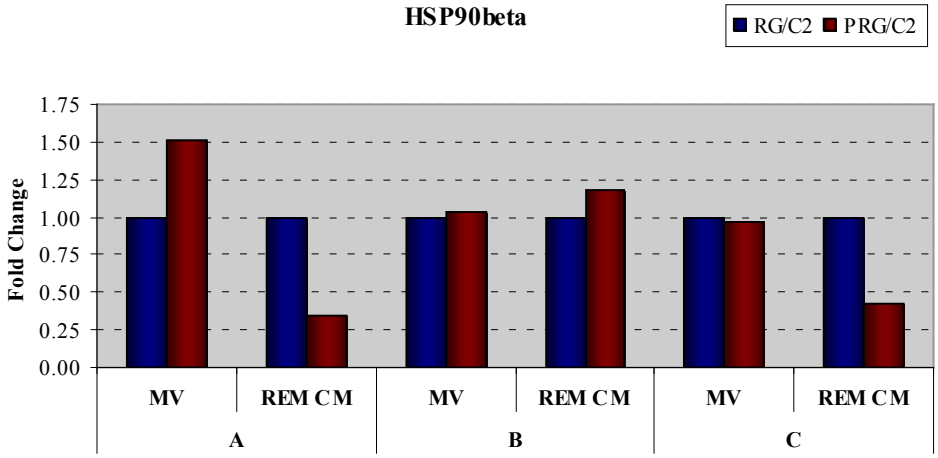
Figure 6. 7 : Confirmation of the presence of HSP90α in MV's prepared to validate the SILAC results.

Western blot analysis was used to confirm the presence of HSP90α in MV's and the remaining CM (after MV preparation) from RG/C2 and PRG/C2 ($5x10^6$ cells) grown in identical conditions used in the SILAC experiments. Fig (a) shows the Western blots. Fig (b) shows the densitometry results of the blots using a GS-800 calibrated densitometer (BIO-RAD). A, B and C represent biological replicates.

(a)



(b)



Adj volume	8.24	12.42	27.59	9.56	4.10	4.25	8.71	10.24	8.92	8.62	2.26	0.97
Fold change	1.00	1.51	1.00	0.35	1.00	1.04	1.00	1.18	1.00	0.97	1.00	0.43

Figure 6. 8 : Confirmation of the presence of HSP90β in MV's prepared to validate the SILAC results.

Western blot analysis was used to confirm the presence of HSP90β in MV's and the remaining CM (after MV preparation) from RG/C2 and PRG/C2 ($5x10^6$ cells) grown in identical conditions used in the SILAC experiments. Fig (a) shows the Western blots. Fig (b) shows the densitometry results of the blots using a GS-800 calibrated densitometer (BIO-RAD). A, B and C represent biological replicates.

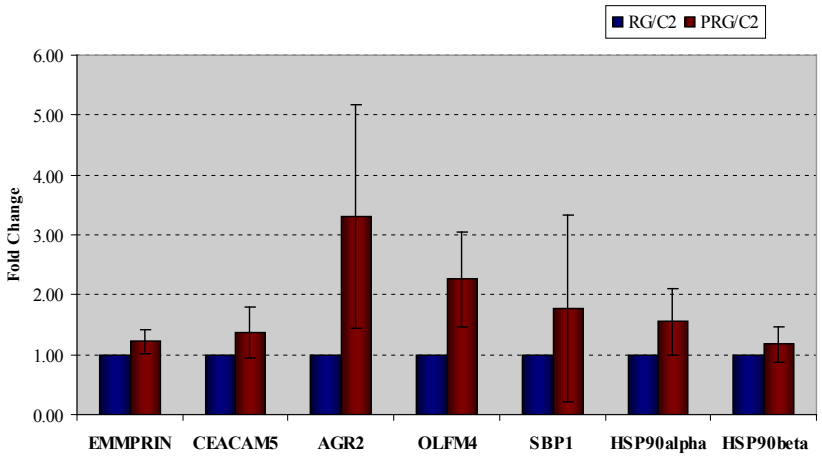
6.7 Conclusions

The aim of this Chapter was to validate the results of the SILAC experiments described in Chapter 5 using Western blot analysis. The Western blot results shown here for the proteins found to be elevated in the SILAC experiments demonstrate that they are not exclusively found in the MV fractions. As previously described in Section 1.8, the contents of MVs can be released shortly after secretion from the cell as demonstrated by Sidhu et al (Sidhu et al., 2004). The MV fractions collected in this study were collected after 24hr incubation. Figure 6.2a shows that EMMPRIN was predominantly present in the MV however all other proteins were found in both MV and REM CM fractions. This could be due to the incubation time for the MV collection used in these experiments.

The Western blot results shown in this chapter also demonstrate the variability between biological replicate experiments. All the proteins investigated (apart from Annexin A2) were detected by Western blotting confirming their presence thereby supporting the MS results from Chapter 5. However, the density of the bands and the abundance ratio between the RG/C2 and PRG/C2 cell lines was variable between biological replicate experiments.

Figure 6.9a demonstrates the average fold change in the Western blot experiments (as calculated by densitometry) and the variation between the triplicate experiments. This variation between replicate experiments was also observed in the SILAC data described in Chapter 5. Figure 6.9b demonstrates the average fold change (ASAPRatio) observed in the SILAC experiments compared to the densitometry results. The ASAPRatio values for each of the proteins were higher than the fold change values calculated from the densitometry results with an average ASAPRatio <1.50 for each of the six proteins. The densitometry results from the Western blot analysis demonstrate a >1.50 fold increase in AGR2, OLFM4, SBP1 and HSP90α in the PRGC2 cell line however, HSP90β and CEACAM5 demonstrated a <1.50 fold

(a)



(b)

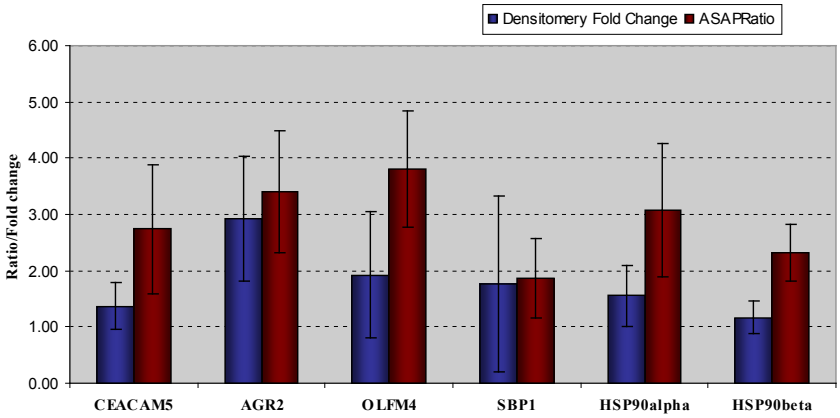


Figure 6. 9 : Summary of results from the Western blot validation experiments and the comparison with the ASAPRatio from the SILAC experiments.

(a) Bar chart to show the average fold change as calculated from the densitometry results from the Western blot validation experiments for the 6 proteins chosen to validate the SILAC results. (b) Bar chart to show the average ASAPRatio (calculated from all four SILAC experiments) compared to the average fold change calculated from the densitometry of the Western blot validation experiments.

increase. Considering the results of the SILAC experiments and the validation Western blot analysis, the proteins that showed the most potential that would warrant further investigation are AGR2, OLFM4, HSP90α and HSP90β.

Chapter 7

Discussion

Chapter 7 : Discussion

7.1 Conclusions and Discussion

Colorectal cancer (CRC) is the third most common cancer and the second most common cause of cancer related deaths in the UK (see Section 1.1). As shown in Figure 1.2e, the five-year survival rate for CRC if diagnosed at an early stage (i.e Dukes' stage A) is 83% which is much higher than the survival rates if diagnosed at Dukes' stage D (3%). However, only 11% of all CRC cases in the UK are diagnosed at Dukes' stage A. Unfortunately, none of the currently used biomarkers are sufficient for routine screening for the early detection of CRC. The aim of this study was to identify potential biomarkers for the early detection of CRC using proteomic techniques. A cell model of early tumour progression (i.e. RG/C2 and PRG/C2) was used to investigate changes in the secreted protein profile.

Chapter 3 describes how differentially secreted proteins in CM from both of these cell lines were initially investigated using 2DE. The proteins in CM from both cell lines were separated by 2D gel electrophoresis. Figure 3.3 illustrates the complexity of the secreted protein profile for both cell lines, especially one region of the gel that exhibited intense CBB G250 staining representing abundant proteins that demonstrated poor separation. This observation was reproducible in biological triplicate experiments and indicated that pre fractionation prior to gel electrophoresis would be required.

The CM was then pre-fractionated by isolating MV from the CM. MV release is thought to play an important role in tumour progression and is thought to be a potential source of clinically useful biomarkers in the detection and diagnosis of cancer. As described in Section 1.5, MVs consist of two main sub-types, larger shedding vesicles and smaller exosomes. Shedding vesicles have been shown to contain membrane proteins which are

difficult to separate using 2DE. Also, the 2D gels shown Figure 3.3 demonstrate a poor reproducibility between biological replicates so therefore MVs were isolated from CM from both cell lines and separated using 1D electrophoresis. EMMPRIN is a protein that has previously been shown to be released in MVs by Sidhu et al and was used as a marker for MV isolation in this study. MVs were isolated from the CM from the RG/C2 and PRGC2 using the method described by Sidhu et al and Western blot analysis confirmed the presence of EMMPRIN predominantly in the MV fraction and also showed an elevated secretion by the PRG/C2 cell line (Sidhu et al., 2004). However as described in Chapter 3, EMMPRIN could not be identified using MS based proteomics methods although its presence could be confirmed in multiple experiments. This could be due to the presence of higher abundance proteins that prevent the identification of EMMPRIN peptides. The CBB G250 stained 1D gel shown in Figure 3.7 demonstrated that the MV fraction contained distinct bands throughout the length of gel and not any regions of the gel with large diffuse bands of proteins that may mask the identification of other proteins (as was the case with the 2D gels shown in Figure 3.3).

Based on the results shown in Chapter 3, the next aim was to isolate MV fractions from CM from both the RG/C2 and PRG/C2 cell lines and identify differentially secreted proteins. This would be achieved by separating the proteins by 1D electrophoresis, digesting them with trypsin and analysing the peptides by MS. In Chapter 4, the sensitivity of the LCQ Deca XP^{PLUS} was investigated and the data analysis methods were optimized. The sensitivity of the instrument was investigated by analysing differing amounts of digested BSA. Table 4.1 shows that BSA peptides could be detected at ~4 fmoles when the sample was digested in solution. This sensitivity was decreased (~250 fmoles) when the samples were first run on a 1D gel and in-gel digestion was performed. Also, the sensitivities of using a CBB G250 or a

silver staining protocol were compared. Figure 4.1a showed that the silver staining protocol was more sensitive with the lowest amount of BSA detected as 31.25 fmoles compared to 250 fmoles for the CBB G250 protocol. However as shown in Figure 4.1b, BSA peptides could only be detected by MS in >250 fmoles therefore the CBB G250 staining protocol was used in this study as the sensitivity provided by the silver staining would not be required because the differences in secretion would be detected by mass spectrometry.

Next, the data analysis methods were optimized such as choice of database and data filter parameters. As described in Chapter 4, the human IPI database proved to be more useful for analysing large shotgun proteomic datasets compared to an NCBI database as it required less manual interpretation. Also, the optimal filter parameters that would be suitable for this type of database were determined as the following: Match = 4; Delta Cn = 0.100; Xcorr = 1.00; Sp = 350; RSp = 5 and Percent ions = 50%. In the same dataset, these parameters were shown to identify more proteins compared to those that were routinely used in our lab without increasing the false positive rate.

The aim of Chapter 5 was to isolate MVs from both the RG/C2 and PRG/C2 cell line and identify differentially secreted proteins using SILAC. The results from the initial SILAC experiment was analysed using the filter parameters optimized in Chapter 4 and the Xpress function in Bioworks Browser was used to calculate the abundance ratio of peptides between both experiments. Manual interpretation of the Xpress ratio and the peaks used by the software to calculate the ratios found that it could yield misleading results (see Figure 5.3 and Table 5.1). Alternative software that could be used to determine the abundance ratios more reliably and would not require time consuming manual interpretation needed to be used. The software utilized in this study was TPP which was used to analyse four separate SILAC experiments. The results for each were filtered using a threshold of protein probability >0.70

which yielded FDR values of 0.042-0.017 (4.2 – 1.7% false positive rate). All four experiments were then compared and 86 proteins were identified in multiple experiments and demonstrated similar ASAPRatios (abundance ratio) in each experiment (listed in Table 5.2).

The aim of the final Chapter of this study was to validate the SILAC results described in Chapter 5 using Western blot analysis. Six proteins were chosen from the listed proteins in Table 5.2, these were: CEACAM5; AGR2; OLFM4; SBP1; HSP90α and HSP90β. MV fractions were prepared from both cell lines cultured under identical conditions to the SILAC experiments in biological triplicate and Western blot analysis was performed for each of these proteins alongside three proteins that were shown to have an average ASAPRatio ~1.00 (i.e. HSC70, TACSTD1 and Annexin A1). The Western blots were able to confirm the presence of HSC70, TACSTD1, the positive control protein, EMMPRIN and all six validation proteins (see Figures 6.1-6.8). However, Annexin A1 could not be detected by Western blot analysis using the AbCam antibody described in Table 2.2a. Densitometry performed on all the Western blots confirmed an increase in secretion of the validation proteins by the PRG/C2 cell line although these increases were not as pronounced as the ASAPRatios calculated by the TPP software from the SILAC experiments (Figure 6.9b).

The results of the Western blot analysis demonstrate that the proteins chosen for validation of the SILAC experiments do show an increase secretion in MVs from the PRG/C2 cell line. Therefore, these proteins show potential as candidate markers of early tumour progression however, further work would be required to show that they are clinically useful.

7.2 Future Work

7.2.1 Further Validation

As described in Chapter 6, many of the proteins selected for validation were not predominantly detected in the MV fraction. Sidhu et al showed that MV can be short lived once released from the cells therefore releasing their contents into the surrounding media (Sidhu et al., 2004). This could explain why these proteins could also be detected in the REM CM fractions. Further time-course experiments would be required to determine whether these proteins would be predominantly detected in the MV fraction if shorter incubation times were used. Also, it would be interesting to investigate whether these incubation times are different for each protein and whether they are secreted in different MVs with different roles. This would also involve fractionating the MV fraction to isolate the two main sub types (i.e. exosomes and larger shedding vesicles) to determine which type of MV these proteins are secreted in. However the aim of this thesis was to discover potential secreted biomarkers for the detection of early stage diagnosis of CRC and the origin of these proteins is irrelevant for biomarker discovery but would be important for understanding the biology of these proteins and their importance in CRC. Also, as shown in Figure 3.7, there are areas of intense CBB G250 staining in the REM CM which could be investigated further as a potential source of markers.

Table 5.2 shows 86 proteins that were selected from the SILAC results because they were all detected in multiple experiments with ASAPRatios calculated in multiple experiments and an ASAPRatio >1.50. Apart from those proteins investigated in Chapter 6, these have not yet been investigated further or validated using Western blot analysis. For some of these proteins, validation by Western blot analysis may not be possible as antibodies are not widely available. This could be resolved by producing antibodies in-house.

Alternatively, these results could be validated using another quantitative method by repeating these experiments using methods such as iTRAQ (see Section 1.2.4.4) using newer technology instruments such as the Thermo linear ion trap/Fourier transform ion cyclotron resonance LTQ/FT, Orbitrap mass spectrometers or the Bruker Maxis TOF/TOF to improve mass accuracy and resolution.

7.2.2 Validation in Serum/Plasma

Once these proteins have been validated further, they need to be investigated in serum/plasma samples to demonstrate quantifiable differences in healthy vs disease. This could involve performing Western blot analysis in pooled serum/plasma samples from healthy, early stage CRC and also in advanced stage disease to investigate whether the changes are stage specific. Enzyme-linked immunosorbent assays (ELISA) could also be used to compare the presence of specific proteins in panels of serum from healthy control patients and patients with early stage CRC. ELISA assays involve detecting a protein of interest in a sample using a specific antibody immobilized to a solid support. A second detection antibody covalently linked to an enzyme is then incubated with the sample. Then after several washes, the assay is developed using an enzyme substrate that is linked to a reporter molecule that can be used to produce a signal that can be measured. There are ELISA kits commercially available for CEACAM5 and HSP90α however, ELISA assays would need to be developed and optimized for AGR2, OLFM2, SBP1 and HSP90β.

Once each of the single proteins have been validated, they could then be assessed for their use as a single biomarker or as part of a panel of biomarkers which could potentially increase the sensitivity and specificity. Candidate biomarker validation can be impeded by the cost of reagents such as the development of high-quality protein antibody assays which

may not be available for novel candidates. A high level of confidence in the candidates is required before spending too much time and money on developing an assay.

One method to overcome this hurdle is Selected Reaction Monitoring or Multiple Reaction Monitoring/Stable Isotope Dilution MS (MRM/SID-MS) using ion trap and triple quadropole mass spectrometers (Kitteringham et al., 2009). Selected Reaction Monitoring (SRM) can be accomplished using triple quadropole mass spectrometers for quantitative proteomics. The first and third quadropoles act as filters to specifically select a pre-defined narrow m/z range for the precursor and the fragment ion. The second quadropole acts as a collision cell to induce fragmentation of the parent ion. These precursor/fragment ion pairs are referred to as transitions and are monitored over time to investigate the retention time and signal intensity of specific transitions. This mass selection at two levels offers high selectivity (Lange et al., 2008).

SRM can be used to monitor a pre-defined set of proteins in a complex mixture. Previous knowledge is required to establish the optimal transitions to use. Firstly, the target proteins to be monitored needs to be determined. Secondly, peptides that produce good responses and uniquely identify the target protein need to be identified. These are referred to as proteotypic peptides or PTPs (Mallick et al., 2007). These assays take time and effort to be established but once set up, they can be used on various samples. This could be one method used to set up an assay to monitor a single or a panel of multiple biomarkers in a variety of samples (Lange et al., 2008).

One of the main challenges to developing a SRM assays using triple quadropole mass spectrometers is finding the best peptides that are sequence specific or unique to that protein (i.e. proteotypic) and give sufficient signal intensity without using protein/peptide standards. Peptide selection relies upon previous data or computational prediction and candidate lists

can be long with the transition list being even longer. Thermo Scientific have developed intelligent SRM (iSRM) for use with their triple quadropole instruments (Thermo Scientific TSQ Vantage) which is able to quantify and simultaneously confirm the identity of target peptides. iSRM involves data-dependant SRM acquisition that monitors primary and secondary transitions for peptide sequence confirmation which is only triggered when the signal intensities of the primary transitions reach a threshold for a certain number of cycles. The target protein list generated from discovery phase experiments is used to generate a preliminary list of peptides and their transitions to develop an iSRM method. This method is then used to analyse a control sample containing these peptides and only those peptides exhibiting good responses are retained in the method whilst those that do not show a good response are removed from the method. This ensures that the cycle time is only being used to analyse peptides that will produce high quality results.

MRM/SID-MS is a high throughput method that offers greater sensitivity and accuracy than methods used in the discovery phase and has the potential to target 30-100 candidate markers within the same assay (Rifai et al., 2006). When the appropriate standards are used, MRM can provide absolute quantification as well as relative quantification and demonstrates a high specificity for the analyte. An ion fragment of interest is isolated in the first quad and fragmented by collision with a neutral gas in a collision cell (the second quad) and the mass of the resulting product ion is analysed in the third quadropole. Multiple ion-product transitions can be monitored in a single assay when coupled to a liquid chromatographic elution method. SID-MS can increase the reliability of the quantification by comparing the signal from an endogenous species with a known concentration of an exogenous stable isotope species. This method has been successfully used to measure human growth hormone and prostate specific antigen in serum or plasma (Barnidge et al., 2004, S. L.

Wu et al., 2002) and also been used to quantify the levels of 53 high-abundance proteins in plasma (L. Anderson and Hunter, 2006). To detect lower abundance proteins, depletion of the high abundance proteins is required as described in Chapter 1. For those proteins that are at too low a level to detect by MRM, stable isotope standards and capture by anti-peptide antibodies (SISCAPA) can also be used to capture isotopically labeled and unlabelled target peptides to improve sensitivity and specificity (N. L. Anderson et al., 2004). Labeled and unlabeled peptides both bind to the antibodies therefore quantification is still possible.

7.2.3 Proteomic Profiling of Whole Cell line

Many cancer studies have compared datasets from similar studies using the same profiling technique but very few cancer studies have integrated the data from multiple data sets using different profiling techniques. Gene expression patterns could be compared with proteomic profiles to investigate any correlations. These are not necessarily associated with protein levels not reflected in the RNA level (S. Hanash, 2003). Comparing the data at the protein and RNA level in the same system could improve the understanding of tumour function. One such study investigated the gene expression levels in 60 human cancer cell lines (NCI-60) and also measured the expression levels of 52 cancer-related proteins in the same cell lines (Nishizuka et al., 2003). The study identified two markers that could be useful to distinguish colon from ovarian adenocarcinomas and also found that the level of structural proteins were highly correlated with the levels of the corresponding mRNAs. The same correlation was not observed for non-structural proteins. For this study, a global profile of the proteome for both the RG/C2 and the PRG/C2 cell line could be investigated using quantitative methods and compared with the gene expression profile for both cell lines. This would give a much wider view of the processes and the changes that occur during early

tumour progression in colorectal cancer and would provide an alternative approach to investigating the secretome.

References

(no-authors-listed) (1996) Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. Adopted on May 17, 1996 by the American Society of Clinical Oncology. **J Clin Oncol**, 14: (10): 2843-2877.

Agnieszka, K. and Silberring, J. (eds.) (2008) **Proteomics: Introduction to Methods and Applications.**: John Wiley & Sons.

Ahmed, N., Barker, G., Oliva, K.T., et al. (2004) Proteomic-based identification of haptoglobin-1 precursor as a novel circulating biomarker of ovarian cancer. **Br J Cancer**, 91: (1): 129-140.

Ahmed, N., Oliva, K.T., Barker, G., et al. (2005) Proteomic tracking of serum protein isoforms as screening biomarkers of ovarian cancer. **Proteomics**, 5: (17): 4625-4636.

Al-Ghoul, M., Bruck, T.B., Lauer-Fields, J.L., et al. (2008) Comparative proteomic analysis of matched primary and metastatic melanoma cell lines. **J Proteome Res**, 7: (9): 4107-4118.

Al-Shahrour, F., Minguez, P., Tarraga, J., et al. (2007) FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. **Nucleic Acids Res**, 35: (Web Server issue): W91-96.

Al-Shahrour, F., Minguez, P., Tarraga, J., et al. (2006) BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. **Nucleic Acids Res**, 34: (Web Server issue): W472-476.

Albanese, J., Meterissian, S., Kontogiannea, M., et al. (1998) Biologically active Fas antigen and its cognate ligand are expressed on plasma membrane-derived extracellular vesicles. **Blood**, 91: (10): 3862-3874.

Albaugh, G.P., Iyengar, V., Lohani, A., et al. (1992) Isolation of exfoliated colonic epithelial cells, a novel, non-invasive approach to the study of cellular markers. **Int J Cancer**, 52: (3): 347-350.

Alexander, H., Stegner, A.L., Wagner-Mann, C., et al. (2004) Proteomic analysis to identify breast cancer biomarkers in nipple aspirate fluid. **Clin Cancer Res**, 10: (22): 7500-7510.

Amacher, D.E. (1998) Serum transaminase elevations as indicators of hepatic injury following the administration of drugs. **Regul Toxicol Pharmacol**, 27: (2): 119-130.

Anderson, L. and Anderson, N.G. (1977) High resolution two-dimensional electrophoresis of human plasma proteins. **Proc Natl Acad Sci U S A**, 74: (12): 5421-5425.

Anderson, L. and Hunter, C.L. (2006) Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. **Mol Cell Proteomics**, 5: (4): 573-588.

Anderson, N.L., Anderson, N.G., Haines, L.R., et al. (2004) Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). **J Proteome Res**, 3: (2): 235-244.

Andreola, G., Rivoltini, L., Castelli, C., et al. (2002) Induction of lymphocyte apoptosis by tumor cell secretion of FasL-bearing microvesicles. **J Exp Med**, 195: (10): 1303-1316.

Antman, E.M., Tanasijevic, M.J., Thompson, B., et al. (1996) Cardiac-specific troponin I levels to predict the risk of mortality in patients with acute coronary syndromes. **N Engl J Med**, 335: (18): 1342-1349.

Aoki, N., Jin-no, S., Nakagawa, Y., et al. (2007) Identification and characterization of microvesicles secreted by 3T3-L1 adipocytes: redox- and hormone-dependent induction of

milk fat globule-epidermal growth factor 8-associated microvesicles. **Endocrinology**, 148: (8): 3850-3862.

Ardoin, S.P., Shanahan, J.C. and Pisetsky, D.S. (2007) The role of microparticles in inflammation and thrombosis. **Scand J Immunol**, 66: (2-3): 159-165.

Arnaud, J.P., Koehl, C. and Adloff, M. (1980) Carcinoembryonic antigen (CEA) in diagnosis and prognosis of colorectal carcinoma. **Dis Colon Rectum**, 23: (3): 141-144.

Ashburner, M., Ball, C.A., Blake, J.A., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. **Nat Genet**, 25: (1): 25-29.

Baj-Krzyworzeka, M., Szatanek, R., Weglarczyk, K., et al. (2006) Tumour-derived microvesicles carry several surface determinants and mRNA of tumour cells and transfer some of these determinants to monocytes. **Cancer Immunol Immunother**, 55: (7): 808-818.

Bantscheff, M., Dumpelfeld, B. and Kuster, B. (2004) Femtomol sensitivity post-digest (18)O labeling for relative quantification of differential protein complex composition. **Rapid Commun Mass Spectrom**, 18: (8): 869-876.

Bantscheff, M., Schirle, M., Sweetman, G., et al. (2007) Quantitative mass spectrometry in proteomics: a critical review. **Anal Bioanal Chem**, 389: (4): 1017-1031.

Barnidge, D.R., Goodmanson, M.K., Klee, G.G., et al. (2004) Absolute quantification of the model biomarker prostate-specific antigen in serum by LC-Ms/MS using protein cleavage and isotope dilution mass spectrometry. **J Proteome Res**, 3: (3): 644-652.

Barry, O.P. and FitzGerald, G.A. (1999) Mechanisms of cellular activation by platelet microparticles. **Thromb Haemost**, 82: (2): 794-800.

Barry, O.P., Pratico, D., Savani, R.C., et al. (1998) Modulation of monocyte-endothelial cell interactions by platelet microparticles. **J Clin Invest**, 102: (1): 136-144.

Basset, P., Bellocq, J.P., Wolf, C., et al. (1990) A novel metalloproteinase gene specifically expressed in stromal cells of breast carcinomas. **Nature**, 348: (6303): 699-704.

Beaudoin, A.R. and Grondin, G. (1991) Shedding of vesicular material from the cell surface of eukaryotic cells: different cellular phenomena. **Biochim Biophys Acta**, 1071: (3): 203-219.

Begent, R.H. (1984) The value of carcinoembryonic antigen measurement in clinical practice. **Ann Clin Biochem**, 21 (Pt 4): 231-238.

Benchimol, S., Fuks, A., Jothy, S., et al. (1989) Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule. **Cell**, 57: (2): 327-334.

Bengtsson, S., Krogh, M., Szigyarto, C.A., et al. (2007) Large-scale proteomics analysis of human ovarian cancer for biomarkers. **J Proteome Res**, 6: (4): 1440-1450.

Berckmans, R.J., Neiuwland, R., Boing, A.N., et al. (2001) Cell-derived microparticles circulate in healthy humans and support low grade thrombin generation. **Thromb Haemost**, 85: (4): 639-646.

Bianco, F., Pravettoni, E., Colombo, A., et al. (2005) Astrocyte-derived ATP induces vesicle shedding and IL-1 beta release from microglia. **J Immunol**, 174: (11): 7268-7277.

Biemann, K. (1988) Contributions of mass spectrometry to peptide and protein structure. **Biomed Environ Mass Spectrom**, 16: (1-12): 99-111.

Biswas, C., Zhang, Y., DeCastro, R., et al. (1995) The human tumor cell-derived collagenase stimulatory factor (renamed EMMPRIN) is a member of the immunoglobulin superfamily. **Cancer Res**, 55: (2): 434-439.

Bjellqvist, B., Ek, K., Righetti, P.G., et al. (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. **J Biochem Biophys Methods**, 6: (4): 317-339.

Bjorhall, K., Miliotis, T. and Davidsson, P. (2005) Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. **Proteomics**, 5: (1): 307-317.

Bloomston, M., Zhou, J.X., Rosemurgy, A.S., et al. (2006) Fibrinogen gamma overexpression in pancreatic cancer identified by large-scale proteomic analysis of serum samples. **Cancer Res**, 66: (5): 2592-2599.

Bluff, J.E., Brown, N.J., Reed, M.W., et al. (2008) Tissue factor, angiogenesis and tumour progression. **Breast Cancer Res**, 10: (2): 204.

Bouwman, K., Qiu, J., Zhou, H., et al. (2003) Microarrays of tumor cell derived proteins uncover a distinct pattern of prostate cancer serum immunoreactivity. **Proteomics**, 3: (11): 2200-2207.

Bradshaw, R.A. (2005) Revised draft guidelines for proteomic data publication. **Mol Cell Proteomics**, 4: (9): 1223-1225.

Brichory, F.M., Misek, D.E., Yim, A.M., et al. (2001) An immune response manifested by the common occurrence of annexins I and II autoantibodies and high circulating levels of IL-6 in lung cancer. **Proc Natl Acad Sci U S A**, 98: (17): 9824-9829.

Brown, L.M., Helmke, S.M., Hunsucker, S.W., et al. (2006) Quantitative and qualitative differences in protein expression between papillary thyroid carcinoma and normal thyroid tissue. **Mol Carcinog**, 45: (8): 613-626.

Brusic, V., Marina, O., Wu, C.J., et al. (2007) Proteome informatics for cancer research: from molecules to clinic. **Proteomics**, 7: (6): 976-991.

Bunkenborg, J., Pilch, B.J., Podtelejnikov, A.V., et al. (2004) Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. **Proteomics**, 4: (2): 454-465.

Burlingame, A.L. (ed.) (2005) **Methods in Enzymology: Biological Mass Spectrometry.**: Elsevier Academic Press.

Caby, M.P., Lankar, D., Vincendeau-Scherrer, C., et al. (2005) Exosomal-like vesicles are present in human blood plasma. **Int Immunol**, 17: (7): 879-887.

Cancer Research UK [online]. http://info.cancerresearchuk.org/cancerstats/index.html [Accessed October 2009]

Carmeliet, P. and Jain, R.K. (2000) Angiogenesis in cancer and other diseases. **Nature**, 407: (6801): 249-257.

Castaman, G., Yu-Feng, L. and Rodeghiero, F. (1996) A bleeding disorder characterised by isolated deficiency of platelet microvesicle generation. **Lancet**, 347: (9002): 700-701.

Caudroy, S., Polette, M., Tournier, J.M., et al. (1999) Expression of the extracellular matrix metalloproteinase inducer (EMMPRIN) and the matrix metalloproteinase-2 in bronchopulmonary and breast lesions. **J Histochem Cytochem**, 47: (12): 1575-1580.

Cecconi, D., Donadelli, M., Rinalducci, S., et al. (2007) Proteomic analysis of pancreatic endocrine tumor cell lines treated with the histone deacetylase inhibitor trichostatin A. **Proteomics**, 7: (10): 1644-1653.

Celis, J.E., Carr, S.A. and Bradshaw, R.A. (2008) New Guidelines for Clinical Proteomics Manuscripts. **Mol Cell Proteomics**, 7: 2071-2072.

Chan, C.H. and Stanners, C.P. (2007) Recent advances in the tumour biology of the GPI-anchored carcinoembryonic antigen family members CEACAM5 and CEACAM6. **Curr Oncol**, 14: (2): 70-73.

Chaput, N., Taieb, J., Schartz, N.E., et al. (2004) Exosome-based immunotherapy. **Cancer Immunol Immunother**, 53: (3): 234-239.

Chen, X., Smith, L.M. and Bradbury, E.M. (2000) Site-specific mass tagging with stable isotopes in proteins for accurate and efficient protein identification. **Anal Chem**, 72: (6): 1134-1143.

Chenau, J., Michelland, S., de Fraipont, F., et al. (2009) The cell line secretome, a suitable tool for investigating proteins released in vivo by tumors: application to the study of p53-modulated proteins secreted in lung cancer cells. **J Proteome Res**, 8: (10): 4579-4591.

Chieregatti, E. and Meldolesi, J. (2005) Regulated exocytosis: new organelles for non-secretory purposes. **Nat Rev Mol Cell Biol**, 6: (2): 181-187.

Cho, S.Y., Lee, E.Y., Kim, H.Y., et al. (2008) Protein profiling of human plasma samples by two-dimensional electrophoresis. **Methods Mol Biol**, 428: 57-75.

Cho, W.C. (2007) Contribution of oncoproteomics to cancer biomarker discovery. **Mol Cancer**, 6: 25.

Choi, D.S., Lee, J.M., Park, G.W., et al. (2007) Proteomic analysis of microvesicles derived from human colorectal cancer cells. **J Proteome Res**, 6: (12): 4646-4655.

Cocucci, E., Racchetti, G. and Meldolesi, J. (2009) Shedding microvesicles: artefacts no more. **Trends Cell Biol**, 19: (2): 43-51.

Cocucci, E., Racchetti, G., Podini, P., et al. (2007) Enlargeosome traffic: exocytosis triggered by various signals is followed by endocytosis, membrane shedding or both. **Traffic**, 8: (6): 742-757.

Coleman, M.L., Sahai, E.A., Yeo, M., et al. (2001) Membrane blebbing during apoptosis results from caspase-mediated activation of ROCK I. **Nat Cell Biol**, 3: (4): 339-345.

Comisarow, I.V. and Marshall, A.G. (1974) Fourier transform ion cyclotron resonance spectroscopy. **Chem Phys Lett**, 25: 282-283.

Dai, M.S. and Lu, H. (2008) Crosstalk between c-Myc and ribosome in ribosomal biogenesis and cancer. **J Cell Biochem**, 105: (3): 670-677.

Darde, V.M., Barderas, M.G. and Vivanco, F. (2007) Depletion of high-abundance proteins in plasma by immunoaffinity subtraction for two-dimensional difference gel electrophoresis analysis. **Methods Mol Biol**, 357: 351-364.

Davidsson, P., Sjogren, M., Andreasen, N., et al. (2002) Studies of the pathophysiological mechanisms in frontotemporal dementia by proteome analysis of CSF proteins. **Brain Res Mol Brain Res**, 109: (1-2): 128-133.

De Ceuninck, F. and Berenbaum, F. (2009) Proteomics: addressing the challenges of osteoarthritis. **Drug Discov Today**, 14: (13-14): 661-667.

DeCastro, R., Zhang, Y., Guo, H., et al. (1996) Human keratinocytes express EMMPRIN, an extracellular matrix metalloproteinase inducer. **J Invest Dermatol**, 106: (6): 1260-1265.

Deininger, S.O., Ebert, M.P., Futterer, A., et al. (2008) MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers. **J Proteome Res**, 7: (12): 5230-5236.

Del Conde, I., Shrimpton, C.N., Thiagarajan, P., et al. (2005) Tissue-factor-bearing microvesicles arise from lipid rafts and fuse with activated platelets to initiate coagulation. **Blood**, 106: (5): 1604-1611.

Deregibus, M.C., Cantaluppi, V., Calogero, R., et al. (2007) Endothelial progenitor cell derived microvesicles activate an angiogenic program in endothelial cells by a horizontal transfer of mRNA. **Blood**, 110: (7): 2440-2448.

Desiere, F., Deutsch, E.W., Nesvizhskii, A.I., et al. (2005) Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. **Genome Biol**, 6: (1): R9.

diaDexus-Inc. (2007) **Clinical application of diaDexus diagnostics in colorectal cancer** [online]. http://www.diadexus.com/products/research/oncology_diagnostics.php [Accessed

Diamandis, E.P. (2004) Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. **J Natl Cancer Inst**, 96: (5): 353-356.

Diamant, M., Nieuwland, R., Pablo, R.F., et al. (2002) Elevated numbers of tissue-factor exposing microparticles correlate with components of the metabolic syndrome in uncomplicated type 2 diabetes mellitus. **Circulation**, 106: (19): 2442-2447.

Distler, J.H., Pisetsky, D.S., Huber, L.C., et al. (2005) Microparticles as regulators of inflammation: novel players of cellular crosstalk in the rheumatic diseases. **Arthritis Rheum**, 52: (11): 3337-3348.

Dolo, V., Ginestra, A., Cassara, D., et al. (1998) Selective localization of matrix metalloproteinase 9, beta1 integrins, and human lymphocyte antigen class I molecules on membrane vesicles shed by 8701-BC breast carcinoma cells. **Cancer Res**, 58: (19): 4468-4474.

Dolo, V., Li, R., Dillinger, M., et al. (2000) Enrichment and localization of ganglioside G(D3) and caveolin-1 in shed tumor cell membrane vesicles. **Biochim Biophys Acta**, 1486: (2-3): 265-274.

Domon, B. and Aebersold, R. (2006) Mass spectrometry and protein analysis. **Science**, 312: (5771): 212-217.

Doustjalali, S.R., Yusof, R., Govindasamy, G.K., et al. (2006) Patients with nasopharyngeal carcinoma demonstrate enhanced serum and tissue ceruloplasmin expression. **J Med Invest**, 53: (1-2): 20-28.

Duffy, M.J. (1998) CA 19-9 as a marker for gastrointestinal cancers: a review. **Ann Clin Biochem**, 35 (Pt 3): 364-370.

Duffy, M.J., van Dalen, A., Haglund, C., et al. (2003) Clinical utility of biochemical markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines. **Eur J Cancer**, 39: (6): 718-727.

Duffy, M.J. (2001) Carcinoembryonic antigen as a marker for colorectal cancer: is it clinically useful? **Clin Chem** 47(4):624-30.

Edler, D., Glimelius, B., Hallstrom, M., et al. (2002) Thymidylate synthase expression in colorectal cancer: a prognostic and predictive marker of benefit from adjuvant fluorouracil-based chemotherapy. **J Clin Oncol**, 20: (7): 1721-1728.

Edman, P. (1949) A method for the determination of amino acid sequence in peptides. **Arch Biochem**, 22: (3): 475.

Elias, J.E. and Gygi, S.P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. **Nat Methods**, 4: (3): 207-214.

Elias, J.E., Haas, W., Faherty, B.K., et al. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. **Nat Methods**, 2: (9): 667-675.

Ellis, S.M., Nabeshima, K. and Biswas, C. (1989) Monoclonal antibody preparation and purification of a tumor cell collagenase-stimulatory factor. **Cancer Res**, 49: (12): 3385-3391.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., et al. (1996) Laser capture microdissection. **Science**, 274: (5289): 998-1001.

Eng, J.K., McCormack, A.L., and Yates, III J.R (1994) An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. **J Am Soc Mass Spectrom**, 5: 976–989.

Etzioni, R., Urban, N., Ramsey, S., et al. (2003) The case for early detection. **Nat Rev Cancer**, 3: (4): 243-252.

Faca, V., Krasnoselsky, A. and Hanash, S. (2007) Innovative proteomic approaches for cancer biomarker discovery. **Biotechniques**, 43: (3): 279, 281-273, 285.

Fenn, J.B., Mann, M., Meng, C.K., and Wong, S.F (1988) Electrspray ionisation for mass spectrometry of large biomolecules. **Science**, 246: 64-71.

Fevrier, B. and Raposo, G. (2004) Exosomes: endosomal-derived vesicles shipping extracellular messages. **Curr Opin Cell Biol**, 16: (4): 415-421.

Fevrier, B., Vilette, D., Laude, H., et al. (2005) Exosomes: a bubble ride for prions? **Traffic**, 6: (1): 10-17.

Fields, M.M. and Chevlen, E. (2006) Ovarian cancer screening: a look at the evidence. **Clin J Oncol Nurs**, 10: (1): 77-81.

Fingleton, B., Menon, R., Carter, K.J., et al. (2004) Proteinase activity in human and murine saliva as a biomarker for proteinase inhibitor efficacy. **Clin Cancer Res**, 10: (23): 7865-7874.

Fiucci, G., Ravid, D., Reich, R., et al. (2002) Caveolin-1 inhibits anchorage-independent growth, anoikis and invasiveness in MCF-7 human breast cancer cells. **Oncogene**, 21: (15): 2365-2375.

Flaumenhaft, R. (2006) Formation and fate of platelet microparticles. **Blood Cells Mol Dis**, 36: (2): 182-187.

Fletcher, G.C., Patel, S., Tyson, K., et al. (2003) hAG-2 and hAG-3, human homologues of genes involved in differentiation, are associated with oestrogen receptor-positive breast tumours and interact with metastasis gene C4.4a and dystroglycan. **Br J Cancer**, 88: (4): 579-585.

Friedman, D.B., Hill, S., Keller, J.W., et al. (2004) Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. **Proteomics**, 4: (3): 793-811.

Fritzsching, B., Schwer, B., Kartenbeck, J., et al. (2002) Release and intercellular transfer of cell surface CD81 via microparticles. **J Immunol**, 169: (10): 5531-5537.

Gao, W.M., Kuick, R., Orchekowski, R.P., et al. (2005) Distinctive serum protein profiles involving abundant proteins in lung cancer patients based upon antibody microarray analysis. **BMC Cancer**, 5: 110.

Gasser, O., Hess, C., Miot, S., et al. (2003) Characterisation and properties of ectosomes released by human polymorphonuclear neutrophils. **Exp Cell Res**, 285: (2): 243-257.

Gasser, O. and Schifferli, J.A. (2004) Activated polymorphonuclear neutrophils disseminate anti-inflammatory microparticles by ectocytosis. **Blood**, 104: (8): 2543-2548.

Gasser, O. and Schifferli, J.A. (2005) Microparticles released by human neutrophils adhere to erythrocytes in the presence of complement. **Exp Cell Res**, 307: (2): 381-387.

Gawaz, M., Neumann, F.J., Ott, I., et al. (1996) Platelet function in acute myocardial infarction treated with direct angioplasty. **Circulation**, 93: (2): 229-237.

Geiser, T., Sturzenegger, M., Genewein, U., et al. (1998) Mechanisms of cerebrovascular events as assessed by procoagulant activity, cerebral microemboli, and platelet microparticles in patients with prosthetic heart valves. **Stroke**, 29: (9): 1770-1777.

George, J.N., Thoi, L.L., McManus, L.M., et al. (1982) Isolation of human platelet membrane microparticles from plasma and serum. **Blood**, 60: (4): 834-840.

Ghosh, S., Spagnoli, G.C., Martin, I., et al. (2005) Three-dimensional culture of melanoma cells profoundly affects gene expression profile: a high density oligonucleotide array study. **J Cell Physiol**, 204: (2): 522-531.

Ginestra, A., La Placa, M.D., Saladino, F., et al. (1998) The amount and proteolytic content of vesicles shed by human cancer cell lines correlates with their in vitro invasiveness. **Anticancer Res**, 18: (5A): 3433-3437.

Giusti, I., D'Ascenzo, S., Millimaggi, D., et al. (2008) Cathepsin B mediates the pH-dependent proinvasive activity of tumor-shed microvesicles. **Neoplasia**, 10: (5): 481-488.

Goichot, B., Grunebaum, L., Desprez, D., et al. (2006) Circulating procoagulant microparticles in obesity. **Diabetes Metab**, 32: (1): 82-85.

Gold, P. and Freedman, S.O. (1965a) Demonstration of Tumor-Specific Antigens in Human Colonic Carcinomata by Immunological Tolerance and Absorption Techniques. **J Exp Med**, 121: 439-462.

Gold, P. and Freedman, S.O. (1965b) Specific carcinoembryonic antigens of the human digestive system. **J Exp Med**, 122: (3): 467-481.

Granger, J., Siddiqui, J., Copeland, S., et al. (2005) Albumin depletion of human plasma also removes low abundance proteins including the cytokines. **Proteomics**, 5: (18): 4713-4718.

Greco, V., Hannus, M. and Eaton, S. (2001) Argosomes: a potential vehicle for the spread of morphogens through epithelia. **Cell**, 106: (5): 633-645.

Gronborg, M., Kristiansen, T.Z., Iwahori, A., et al. (2006) Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. **Mol Cell Proteomics**, 5: (1): 157-171.

Gu, S., Liu, Z., Pan, S., et al. (2004) Global investigation of p53-induced apoptosis through quantitative proteomic profiling using comparative amino acid-coded tagging. **Mol Cell Proteomics**, 3: (10): 998-1008.

Guo, H., Li, R., Zucker, S., et al. (2000) EMMPRIN (CD147), an inducer of matrix metalloproteinase synthesis, also binds interstitial collagenase to the tumor cell surface. **Cancer Res**, 60: (4): 888-891.

Guo, H., Zucker, S., Gordon, M.K., et al. (1997) Stimulation of matrix metalloproteinase production by recombinant extracellular matrix metalloproteinase inducer from transfected Chinese hamster ovary cells. **J Biol Chem**, 272: (1): 24-27.

Gutwein, P., Mechtersheimer, S., Riedle, S., et al. (2003) ADAM10-mediated cleavage of L1 adhesion molecule at the cell surface and in released membrane vesicles. **Faseb J**, 17: (2): 292-294.

Habermann, J.K., Roblick, U.J., Luke, B.T., et al. (2006) Increased serum levels of complement C3a anaphylatoxin indicate the presence of colorectal tumors. **Gastroenterology**, 131: (4): 1020-1029; quiz 1284.

Hake, S.B., Xiao, A. and Allis, C.D. (2004) Linking the epigenetic 'language' of covalent histone modifications to cancer. **Br J Cancer**, 90: (4): 761-769.

Hammarstrom, S. (1999) The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. **Semin Cancer bIOL**, 9: (2): 67-81.

Han, D.K., Eng, J., Zhou, H., et al. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. **Nat Biotechnol**, 19: (10): 946-951.

Hanahan, D. and Weinberg, R.A. (2000) The hallmarks of cancer. **Cell**, 100: (1): 57-70.

Hanash, S. (2003) Disease proteomics. **Nature**, 422: (6928): 226-232.

Hanash, S.M., Madoz-Gurpide, J. and Misek, D.E. (2002) Identification of novel targets for cancer therapy using expression proteomics. **Leukemia**, 16: (4): 478-485.

Hardcastle, J.D., Chamberlain, J.O., Robinson, M.H., et al. (1996) Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. **Lancet**, 348: (9040): 1472-1477.

Haug, U. and Brenner, H. (2005) New stool tests for colorectal cancer screening: a systematic review focusing on performance characteristics and practicalness. **Int J Cancer**, 117: (2): 169-176.

He, Q.Y., Zhu, R., Lei, T., et al. (2008) Toward the proteomic identification of biomarkers for the prediction of HBV related hepatocellular carcinoma. **J Cell Biochem**, 103: (3): 740-752.

Hegmans, J.P., Bard, M.P., Hemmes, A., et al. (2004) Proteomic analysis of exosomes secreted by human mesothelioma cells. **Am J Pathol**, 164: (5): 1807-1815.

Heijnen, H.F., Schiel, A.E., Fijnheer, R., et al. (1999) Activated platelets release two types of membrane vesicles: microvesicles by surface shedding and exosomes derived from exocytosis of multivesicular bodies and alpha-granules. **Blood**, 94: (11): 3791-3799.

Heller, M., Mattou, H., Menzel, C., and Yao, X. (2003) Trypsin catalyzed 16O - 18O exchange for comparitive proteommics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF and ESI-ion trap mass spectrometers. **J Am Soc Mass Spectrom**, 14: 704-718.

Hess, C., Sadallah, S., Hefti, A., et al. (1999) Ectosomes released by human neutrophils are specialized functional units. **J Immunol**, 163: (8): 4564-4573.

Hijikata, A., Kitamura, H., Kimura, Y., et al. (2007) Construction of an open-access database that integrates cross-reference information from the transcriptome and proteome of immune cells. **Bioinformatics**, 23: (21): 2934-2941.

Holten-Andersen, M.N., Murphy, G., Nielsen, H.J., et al. (1999) Quantitation of TIMP-1 in plasma of healthy blood donors and patients with advanced cancer. **Br J Cancer**, 80: (3-4): 495-503.

Honda, K., Hayashida, Y., Umaki, T., et al. (2005) Possible detection of pancreatic cancer by plasma protein profiling. **Cancer Res**, 65: (22): 10613-10622.

Horstman, L.L., Jy, W., Jimenez, J.J., et al. (2004) New horizons in the analysis of circulating cell-derived microparticles. **Keio J Med**, 53: (4): 210-230.

Hoving, S., Gerrits, B., Voshol, H., et al. (2002) Preparative two-dimensional gel electrophoresis at alkaline pH using narrow range immobilized pH gradients. **Proteomics**, 2: (2): 127-134.

Hsich, G., Kenney, K., Gibbs, C.J., et al. (1996) The 14-3-3 brain protein in cerebrospinal fluid as a marker for transmissible spongiform encephalopathies. **N Engl J Med**, 335: (13): 924-930.

Hu, S., Loo, J.A. and Wong, D.T. (2006) Human body fluid proteome analysis. **Proteomics**, 6: (23): 6326-6353.

Hu, Z., Hood, L. and Tian, Q. (2007) Quantitative proteomic approaches for biomarker discovery. **Proteomics Clin. Appl**, 1: 1036–1041.

Huang, C.S., Lal, S.K. and Farraye, F.A. (2005a) Colorectal cancer screening in average risk individuals. **Cancer Causes Control**, 16: (2): 171-188.

Huang, H.L., Stasyk, T., Morandell, S., et al. (2005b) Enrichment of low-abundant serum proteins by albumin/immunoglobulin G immunoaffinity depletion under partly denaturing conditions. **Electrophoresis**, 26: (14): 2843-2849.

Huang, K.C., Park, D.C., Ng, S.K., et al. (2006a) Selenium binding protein 1 in ovarian cancer. **Int J Cancer**, 118: (10): 2433-2440.

Huang, L.J., Chen, S.X., Huang, Y., et al. (2006b) Proteomics-based identification of secreted protein dihydrodiol dehydrogenase as a novel serum markers of non-small cell lung cancer. **Lung Cancer**, 54: (1): 87-94.

Huber, V., Fais, S., Iero, M., et al. (2005) Human colorectal cancer cells induce T-cell death through release of proapoptotic microvesicles: role in immune escape. **Gastroenterology**, 128: (7): 1796-1804.

Hugel, B., Martinez, M.C., Kunzelmann, C., et al. (2005) Membrane microparticles: two sides of the coin. **Physiology (Bethesda)**, 20: 22-27.

Hundt, S., Haug, U. and Brenner, H. (2007) Blood markers for early detection of colorectal cancer: a systematic review. **Cancer Epidemiol Biomarkers Prev**, 16: (10): 1935-1953.

Hunter, M.P., Ismail, N., Zhang, X., et al. (2008) Detection of microRNA expression in human peripheral blood microvesicles. **PLoS One**, 3: (11): e3694.

IPI - International Protein Index [online]. http://www.ebi.ac.uk/IPI/ [Accessed

Ishihama, Y., Oda, Y., Tabata, T., et al. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. **Mol Cell Proteomics**, 4: (9): 1265-1272.

Issaq, H. and Veenstra, T. (2008) Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. **Biotechniques**, 44: (5): 697-698, 700.

Iwadate, Y. (2008) Clinical proteomics in cancer research-promises and limitations of current two-dimensional gel electrophoresis. **Curr Med Chem**, 15: (23): 2393-2400.

Janowska-Wieczorek, A., Wysoczynski, M., Kijowski, J., et al. (2005) Microvesicles derived from activated platelets induce metastasis and angiogenesis in lung cancer. **Int J Cancer**, 113: (5): 752-760.

Kanehisa, M., Goto, S., Kawashima, S., et al. (2004) The KEGG resource for deciphering the genome. **Nucleic Acids Res**, 32: (Database issue): D277-280.

Kanekura, T., Chen, X. and Kanzaki, T. (2002) Basigin (CD147) is expressed on melanoma cells and induces tumor cell invasion by stimulating production of matrix metalloproteinases by fibroblasts. **Int J Cancer**, 99: (4): 520-528.

Karas, M., Bachmann, D., Bahr, U., and Hillenkampf, F. (1987) **INT J Mass Spectrom Ion Processes**, (78): 53-68.

Kasinrerk, W., Fiebiger, E., Stefanova, I., et al. (1992) Human leukocyte activation antigen M6, a member of the Ig superfamily, is the species homologue of rat OX-47, mouse basigin, and chicken HT7 molecule. **J Immunol**, 149: (3): 847-854.

Kataoka, H., DeCastro, R., Zucker, S., et al. (1993) Tumor cell-derived collagenase-stimulatory factor increases expression of interstitial collagenase, stromelysin, and 72-kDa gelatinase. **Cancer Res**, 53: (13): 3154-3158.

Keller, A., Eng, J., Zhang, N., et al. (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. **Mol Syst Biol**, 1: 2005 0017.

Keller, A., Nesvizhskii, A.I., Kolker, E., et al. (2002a) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. **Anal Chem**, 74: (20): 5383-5392.

Keller, A., Purvine, S., Nesvizhskii, A.I., et al. (2002b) Experimental protein mixture for validating tandem mass spectral analysis. **Omics**, 6: (2): 207-212.

Keller, S., Sanderson, M.P., Stoeck, A., et al. (2006) Exosomes: from biogenesis and secretion to biological function. **Immunol Lett**, 107: (2): 102-108.

Khwaja, F.W., Svoboda, P., Reed, M., et al. (2006) Proteomic identification of the wt-p53-regulated tumor cell secretome. **Oncogene**, 25: (58): 7650-7661.

Kim, H., Kang, H.J., You, K.T., et al. (2006) Suppression of human selenium-binding protein 1 is a late event in colorectal carcinogenesis and is associated with poor survival. **Proteomics**, 6: (11): 3466-3476.

Kim, H.J., Yu, M.H., Kim, H., et al. (2008) Noninvasive molecular biomarkers for the detection of colorectal cancer. **BMB Rep**, 41: (10): 685-692.

Kim, H.K., Song, K.S., Park, Y.S., et al. (2003) Elevated levels of circulating platelet microparticles, VEGF, IL-6 and RANTES in patients with gastric cancer: possible role of a metastasis predictor. **Eur J Cancer**, 39: (2): 184-191.

Kim, J.W., Wieckowski, E., Taylor, D.D., et al. (2005) Fas ligand-positive membranous vesicles isolated from sera of patients with oral cancer induce apoptosis of activated T lymphocytes. **Clin Cancer Res**, 11: (3): 1010-1020.

Kim, S.Y. and Hahn, W.C. (2007) Cancer genomics: integrating form and function. **Carcinogenesis**, 28: (7): 1387-1392.

Kirmiz, C., Li, B., An, H.J., et al. (2007) A serum glycomics approach to breast cancer biomarkers. **Mol Cell Proteomics**, 6: (1): 43-55.

Kislinger, T., Gramolini, A.O., MacLennan, D.H., et al. (2005) Multidimensional protein identification technology (MudPIT): technical overview of a profiling method optimized for the comprehensive proteomic investigation of normal and diseased heart tissue. **J Am Soc Mass Spectrom**, 16: (8): 1207-1220.

Kitteringham, N.R., Jenkins, R.E., Lane, C.S., et al. (2009) Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. **J Chromatogr B Analyt Technol Biomed Life Sci**, 877: (13): 1229-1239.

Klein, C.A., Seidl, S., Petat-Dutter, K., et al. (2002) Combined transcriptome and genome analysis of single micrometastatic cells. **Nat Biotechnol**, 20: (4): 387-392.

Kobayashi, D., Koshida, S., Moriai, R., et al. (2007) Olfactomedin 4 promotes S-phase transition in proliferation of pancreatic cancer cells. **Cancer Sci**, 98: (3): 334-340.

Kobayashi, T., Okamoto, H., Yamada, J., et al. (1984) Vesiculation of platelet plasma membranes. Dilauroylglycerophosphocholine-induced shedding of a platelet plasma

membrane fraction enriched in acetylcholinesterase activity. **Biochim Biophys Acta**, 778: (1): 210-218.

Koga, H., Sugiyama, S., Kugiyama, K., et al. (2006) Elevated levels of remnant lipoproteins are associated with plasma platelet microparticles in patients with type-2 diabetes mellitus without obstructive coronary artery disease. **Eur Heart J**, 27: (7): 817-823.

Kondo, T. (2008) Tissue proteomics for cancer biomarker development: laser microdissection and 2D-DIGE. **BMB Rep**, 41: (9): 626-634.

Koppler, B., Cohen, C., Schlondorff, D., et al. (2006) Differential mechanisms of microparticle transfer toB cells and monocytes: anti-inflammatory propertiesof microparticles. **Eur J Immunol**, 36: (3): 648-660.

Koshida, S., Kobayashi, D., Moriai, R., et al. (2007) Specific overexpression of OLFM4(GW112/HGC-1) mRNA in colon, breast and lung cancer tissues detected using quantitative analysis. **Cancer Sci**, 98: (3): 315-320.

Kozak, K.R., Amneus, M.W., Pusey, S.M., et al. (2003) Identification of biomarkers for ovarian cancer using strong anion-exchange ProteinChips: potential use in diagnosis and prognosis. **Proc Natl Acad Sci U S A**, 100: (21): 12343-12348.

Kozak KR, S.F., Whitelegge JP, Faull K, Reddy S, Farias-Eisner R. (2005) Characterization of serum biomarkers for detection of early stage ovarian cancer. . **Proteomics**, 5: (17): 4589-4596.

Kraj, A.a.S., J. (ed.) (2008) Proteomics: Introduction to Methods and Applications: John Wiley & Sons.

Kramer, B., Pelchen-Matthews, A., Deneka, M., et al. (2005) HIV interaction with endosomes in macrophages and dendritic cells. **Blood Cells Mol Dis**, 35: (2): 136-142.

Kronborg, O., Fenger, C., Olsen, J., et al. (1996) Randomised study of screening for colorectal cancer with faecal-occult-blood test. **Lancet**, 348: (9040): 1467-1471.

Kryczek, I., Lange, A., Mottram, P., et al. (2005) CXCL12 and vascular endothelial growth factor synergistically induce neoangiogenesis in human ovarian cancers. **Cancer Res**, 65: (2): 465-472.

Kulasingam, V. and Diamandis, E.P. (2007) Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets. **Mol Cell Proteomics**, 6: (11): 1997-2011.

Kumar, S., Mohan, A. and Guleria, R. (2006) Biomarkers in cancer screening, research and detection: present and future: a review. **Biomarkers**, 11: (5): 385-405.

Lafitte D, D.B., Andersen S, Vazi A, Dupuy P, Jensen ON, Berland Y, Verdier JM. (2002) Optimized preparation of urine samples for two-dimensional electrophoresis and initial application to patient samples. . **Clin Biochem**, 35: (8): 581-589.

Lakkaraju, A. and Rodriguez-Boulan, E. (2008) Itinerant exosomes: emerging roles in cell and tissue polarity. **Trends Cell Biol**, 18: (5): 199-209.

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

Langen, H., Takacs, B., Evers, S., et al. (2000) Two-dimensional map of the proteome of Haemophilus influenzae. **Electrophoresis**, 21: (2): 411-429.

Larranaga, P., Calvo, B., Santana, R., et al. (2006) Machine learning in bioinformatics. **Brief Bioinform**, 7: (1): 86-112.

Latterich, M., Abramovitz, M. and Leyland-Jones, B. (2008) Proteomics: new technologies and clinical applications. **Eur J Cancer**, 44: (18): 2737-2741.

Lee, C.H., Lum, J.H., Cheung, B.P., et al. (2005) Identification of the heterogeneous nuclear ribonucleoprotein A2/B1 as the antigen for the gastrointestinal cancer specific monoclonal antibody MG7. **Proteomics**, 5: (4): 1160-1166.

Lee, H., Rhee, H., Kang, H.J., et al. (2008) Macrophage migration inhibitory factor may be used as an early diagnostic marker in colorectal carcinomas. **Am J Clin Pathol**, 129: (5): 772-779.

Lee, T.L., Lin, Y.C., Mochitate, K., et al. (1993) Stress-relaxation of fibroblasts in collagen matrices triggers ectocytosis of plasma membrane vesicles containing actin, annexins II and VI, and beta 1 integrin receptors. **J Cell Sci**, 105 (Pt 1): 167-177.

Leroyer, A.S., Tedgui, A. and Boulanger, C.M. (2008) Microparticles and type 2 diabetes. **Diabetes Metab**, 34 Suppl 1: S27-32.

Lescuyer, P., Hochstrasser, D. and Rabilloud, T. (2007) How shall we use the proteomics toolbox for biomarker discovery? **J Proteome Res**, 6: (9): 3371-3376.

Li, C., Hong, Y., Tan, Y.X., et al. (2004a) Accurate qualitative and quantitative proteomic analysis of clinical hepatocellular carcinoma using laser capture microdissection coupled with isotope-coded affinity tag and two-dimensional liquid chromatography mass spectrometry. **Mol Cell Proteomics**, 3: (4): 399-409.

Li, C., Xiao, Z., Chen, Z., et al. (2006) Proteome analysis of human lung squamous carcinoma. **Proteomics**, 6: (2): 547-558.

Li, L.S., Kim, H., Rhee, H., et al. (2004b) Proteomic analysis distinguishes basaloid carcinoma as a distinct subtype of nonsmall cell lung carcinoma. **Proteomics**, 4: (11): 3394-3400.

Li, R., Huang, L., Guo, H., et al. (2001) Basigin (murine EMMPRIN) stimulates matrix metalloproteinase production by fibroblasts. **J Cell Physiol**, 186: (3): 371-379.

Li, X.J., Zhang, H., Ranish, J.A., et al. (2003) Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. **Anal Chem**, 75: (23): 6648-6657.

Lim, M., Martinez, T., Jablons, D., et al. (1998) Tumor-derived EMMPRIN (extracellular matrix metalloproteinase inducer) stimulates collagenase transcription through MAPK p38. **FEBS Lett**, 441: (1): 88-92.

Lin, B., White, J.T., Lu, W., et al. (2005) Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: a systems approach to disease. **Cancer Res**, 65: (8): 3081-3091.

Listgarten, J. and Emili, A. (2005) Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. **Mol Cell Proteomics**, 4: (4): 419-434.

Liu, D., Rudland, P.S., Sibson, D.R., et al. (2005) Human homologue of cement gland protein, a novel metastasis inducer associated with breast carcinomas. **Cancer Res**, 65: (9): 3796-3805.

Liu, H., Lin, D. and Yates, J.R., 3rd (2002a) Multidimensional separations for protein/peptide analysis in the post-genomic era. **Biotechniques**, 32: (4): 898, 900, 902 passim.

Liu, P., Rudick, M. and Anderson, R.G. (2002b) Multiple functions of caveolin-1. **J Biol Chem**, 277: (44): 41295-41298.

Loboda, A.V., Krutchinsky, A.N., Bromirski, M., et al. (2000) A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. **Rapid Commun Mass Spectrom**, 14: (12): 1047-1057.

Looi, K.S., Nakayasu, E.S., Diaz, R.A., et al. (2008) Using proteomic approach to identify tumor-associated antigens as markers in hepatocellular carcinoma. **J Proteome Res**, 7: (9): 4004-4012.

Losche, W., Scholz, T., Temmler, U., et al. (2004) Platelet-derived microvesicles transfer tissue factor to monocytes but not to neutrophils. **Platelets**, 15: (2): 109-115.

Lowenthal, M.S., Mehta, A.I., Frogale, K., et al. (2005) Analysis of albumin-associated peptides and proteins from ovarian cancer patients. **Clin Chem**, 51: (10): 1933-1945.

Mack, M., Kleinschmidt, A., Bruhl, H., et al. (2000) Transfer of the chemokine receptor CCR5 between cells by membrane-derived microparticles: a mechanism for cellular human immunodeficiency virus 1 infection. **Nat Med**, 6: (7): 769-775.

Mack, T.M., Hamilton, A.S., Press, M.F., et al. (2002) Heritable breast cancer in twins. **Br J Cancer**, 87: (3): 294-300.

Major, T.C., Liang, L., Lu, X., et al. (2002) Extracellular matrix metalloproteinase inducer (EMMPRIN) is induced upon monocyte differentiation and is expressed in human atheroma. **Arterioscler Thromb Vasc Biol**, 22: (7): 1200-1207.

Mallat, Z., Benamer, H., Hugel, B., et al. (2000) Elevated levels of shed membrane microparticles with procoagulant potential in the peripheral circulating blood of patients with acute coronary syndromes. **Circulation**, 101: (8): 841-843.

Mallick, P., Schirle, M., Chen, S.S., et al. (2007) Computational prediction of proteotypic peptides for quantitative proteomics. **Nat Biotechnol**, 25: (1): 125-131.

Mann, M. and Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. **Anal Chem**, 66: (24): 4390-4399.

Martin, D.B., Gifford, D.R., Wright, M.E., et al. (2004) Quantitative proteomic analysis of proteins released by neoplastic prostate epithelium. **Cancer Res**, 64: (1): 347-355.

Martinkova, J., Gadher, S.J., Hajduch, M., et al. (2009) Challenges in cancer research and multifaceted approaches for cancer biomarker quest. **FEBS Lett**, 583: (11): 1772-1784.

Maurya, P., Meleady, P., Dowling, P., et al. (2007) Proteomic approaches for serum biomarker discovery in cancer. **Anticancer Res**, 27: (3A): 1247-1255.

Mbeunkui, F., Fodstad, O. and Pannell, L.K. (2006) Secretory protein enrichment and analysis: an optimized approach applied on cancer cell lines using 2D LC-MS/MS. **J Proteome Res**, 5: (4): 899-906.

Mbeunkui, F., Metge, B.J., Shevde, L.A., et al. (2007) Identification of differentially secreted biomarkers using LC-MS/MS in isogenic cell lines representing a progression of breast cancer. **J Proteome Res**, 6: (8): 2993-3002.

Mears, R., Craven, R.A., Hanrahan, S., et al. (2004) Proteomic analysis of melanoma-derived exosomes by two-dimensional polyacrylamide gel electrophoresis and mass spectrometry. **Proteomics**, 4: (12): 4019-4031.

Meijerink, J.P., Mensink, E.J., Wang, K., et al. (1998) Hematopoietic malignancies demonstrate loss-of-function mutations of BAX. **Blood**, 91: (8): 2991-2997.

Melle, C., Ernst, G., Schimmel, B., et al. (2005) Discovery and identification of alpha-defensins as low abundant, tumor-derived serum markers in colorectal cancer. **Gastroenterology**, 129: (1): 66-73.

Miller, J.C., Zhou, H., Kwekel, J., et al. (2003) Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. **Proteomics**, 3: (1): 56-63.

Millimaggi, D., Festuccia, C., Angelucci, A., et al. (2006) Osteoblast-conditioned media stimulate membrane vesicle shedding in prostate cancer cells. **Int J Oncol**, 28: (4): 909-914.

Minden, J. (2007) Comparative proteomics and difference gel electrophoresis. **Biotechniques**, 43: (6): 739, 741, 743 passim.

Mochizuki, S. and Okada, Y. (2007) ADAMs in cancer cell proliferation and progression. **Cancer Sci**, 98: (5): 621-628.

Mohri, Y., Mohri, T., Wei, W., et al. (2009) Identification of macrophage migration inhibitory factor and human neutrophil peptides 1-3 as potential biomarkers for gastric cancer. **Br J Cancer**, 101: (2): 295-302.

Morris, C.E. and Homann, U. (2001) Cell surface area regulation and membrane tension. **J Membr Biol**, 179: (2): 79-102.

Muraoka, K., Nabeshima, K., Murayama, T., et al. (1993) Enhanced expression of a tumor-cell-derived collagenase-stimulatory factor in urothelial carcinoma: its usefulness as a tumor marker for bladder cancers. **Int J Cancer**, 55: (1): 19-26.

Nabeshima, K., Lane, W.S. and Biswas, C. (1991) Partial sequencing and characterization of the tumor cell-derived collagenase stimulatory factor. **Arch Biochem Biophys**, 285: (1): 90-96.

National Center for Biotechnology Information [online]. http://www.ncbi.nlm.nih.gov/ [Accessed

Nesvizhskii, A.I. and Aebersold, R. (2004) Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. **Drug Discov Today**, 9: (4): 173-181.

Nesvizhskii, A.I. and Aebersold, R. (2005) Interpretation of shotgun proteomic data: the protein inference problem. **Mol Cell Proteomics**, 4: (10): 1419-1440.

Nesvizhskii, A.I., Keller, A., Kolker, E., et al. (2003) A statistical model for identifying proteins by tandem mass spectrometry. **Anal Chem**, 75: (17): 4646-4658.

Nesvizhskii, A.I., Vitek, O. and Aebersold, R. (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. **Nat Methods**, 4: (10): 787-797.

Nishigaki, R., Osaki, M., Hiratsuka, M., et al. (2005) Proteomic identification of differentially-expressed genes in human gastric carcinomas. **Proteomics**, 5: (12): 3205-3213.

Nishizuka, S., Charboneau, L., Young, L., et al. (2003) Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. **Proc Natl Acad Sci U S A**, 100: (24): 14229-14234.

Noguchi, Y., Sato, T., Hirata, M., et al. (2003) Identification and characterization of extracellular matrix metalloproteinase inducer in human endometrium during the menstrual cycle in vivo and in vitro. **J Clin Endocrinol Metab**, 88: (12): 6063-6072.

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. **J Biol Chem**, 250: (10): 4007-4021.

Oda, Y., Huang, K., Cross, F.R., et al. (1999) Accurate quantitation of protein expression and site-specific phosphorylation. **Proc Natl Acad Sci U S A**, 96: (12): 6591-6596.

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., et al. (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. **Mol Cell Proteomics**, 4: (10): 1487-1502.

Olson, M.F. and Sahai, E. (2009) The actin cytoskeleton in cancer cell motility. **Clin Exp Metastasis**, 26: (4): 273-287.

Ong, S.E., Blagoev, B., Kratchmarova, I., et al. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. **Mol Cell Proteomics**, 1: (5): 376-386.

Osterud, B. (2003) The role of platelets in decrypting monocyte tissue factor. **Dis Mon**, 49: (1): 7-13.

Pan, L.N., Lu, J. and Huang, B. (2007) HDAC inhibitors: a potential new category of anti-tumor agents. **Cell Mol Immunol**, 4: (5): 337-343.

Pap, E., Pallinger, E., Falus, A., et al. (2008) T lymphocytes are targets for platelet- and trophoblast-derived microvesicles during pregnancy. **Placenta**, 29: (9): 826-832.

Paraskeva C, Finerty S, Mountford RA, Powell SC. Specific cytogenetic abnormalities in two new human colorectal adenoma-derived epithelial cell lines. Cancer Res. 1989 Mar 1;49(5):1282-6.

Park, S.W., Zhen, G., Verhaeghe, C., et al. (2009) The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. **Proc Natl Acad Sci U S A**, 106: (17): 6950-6955.

PEAKS software [online]. http://www.bioinformaticssolutions.com/products/peaks/index.php [Accessed

Pelchen-Matthews, A., Raposo, G. and Marsh, M. (2004) Endosomes, exosomes and Trojan viruses. **Trends Microbiol**, 12: (7): 310-316.

Pepe, M.S., Etzioni, R., Feng, Z., et al. (2001) Phases of biomarker development for early detection of cancer. **J Natl Cancer Inst**, 93: (14): 1054-1061.

Pereira-Faca, S.R., Kuick, R., Puravs, E., et al. (2007) Identification of 14-3-3 theta as an antigen that induces a humoral response in lung cancer. **Cancer Res**, 67: (24): 12000-12006.

Perkins, D.N., Pappin, D.J., Creasy, D.M., et al. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. **Electrophoresis**, 20: (18): 3551-3567.

Petricoin, E.F., Ardekani, A.M., Hitt, B.A., et al. (2002) Use of proteomic patterns in serum to identify ovarian cancer. **Lancet**, 359: (9306): 572-577.

Petricoin, E.F., Ornstein, D.K. and Liotta, L.A. (2004) Clinical proteomics: Applications for prostate cancer biomarker discovery and detection. **Urol Oncol**, 22: (4): 322-328.

Piccin, A., Murphy, W.G. and Smith, O.P. (2007) Circulating microparticles: pathophysiology and clinical implications. **Blood Rev**, 21: (3): 157-171.

Pietra, N., Sarli, L., Costi, R., et al. (1998) Role of follow-up in management of local recurrences of colorectal cancer: a prospective, randomized study. **Dis Colon Rectum**, 41: (9): 1127-1133.

Pilzer, D., Gasser, O., Moskovich, O., et al. (2005) Emission of membrane vesicles: roles in complement resistance, immunity and cancer. **Springer Semin Immunopathol**, 27: (3): 375-387.

Pisitkun, T., Shen, R.F. and Knepper, M.A. (2004) Identification and proteomic profiling of exosomes in human urine. **Proc Natl Acad Sci U S A**, 101: (36): 13368-13373.

Pizzirani, C., Ferrari, D., Chiozzi, P., et al. (2007) Stimulation of P2 receptors causes release of IL-1beta-loaded microvesicles from human dendritic cells. **Blood**, 109: (9): 3856-3864.

Pohler, E., Craig, A.L., Cotton, J., et al. (2004) The Barrett's antigen anterior gradient-2 silences the p53 transcriptional response to DNA damage. **Mol Cell Proteomics**, 3: (6): 534-547.

Polette, M., Gilles, C., Marchand, V., et al. (1997) Tumor collagenase stimulatory factor (TCSF) expression and localization in human lung and breast cancers. **J Histochem Cytochem**, 45: (5): 703-709.

Powers, M.V. and Workman, P. (2006) Targeting of multiple signalling pathways by heat shock protein 90 molecular chaperone inhibitors. **Endocr Relat Cancer**, 13 Suppl 1: S125-135.

Powers, M.V. and Workman, P. (2007) Inhibitors of the heat shock response: biology and pharmacology. **FEBS Lett**, 581: (19): 3758-3769.

Prall, F., Ostwald, C., Schiffmann, L., et al. (2007) Do thymidylate synthase gene promoter polymorphism and the C/G single nucleotide polymorphism predict effectiveness of adjuvant 5-fluorouracil-based chemotherapy in stage III colonic adenocarcinoma? **Oncol Rep**, 18: (1): 203-209.

Preston, R.A., Jy, W., Jimenez, J.J., et al. (2003) Effects of severe hypertension on endothelial and platelet microparticles. **Hypertension**, 41: (2): 211-217.

Qian, W.J., Jacobs, J.M., Liu, T., et al. (2006) Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications. **Mol Cell Proteomics**, 5: (10): 1727-1744.

Qiu, Y., Patwa, T.H., Xu, L., et al. (2008) Plasma glycoprotein profiling for colorectal cancer biomarker identification by lectin glycoarray and lectin blot. **J Proteome Res**, 7: (4): 1693-1703.

Qu, J., Jusko, W.J. and Straubinger, R.M. (2006) Utility of cleavable isotope-coded affinity-tagged reagents for quantification of low-copy proteins induced by methylprednisolone using liquid chromatography/tandem mass spectrometry. **Anal Chem**, 78: (13): 4543-4552.

Rabilloud, T. (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. **Proteomics**, 2: (1): 3-10.

Radisky, D.C. and Bissell, M.J. (2004) Cancer. Respect thy neighbor! **Science**, 303: (5659): 775-777.

Radulovic, D., Jelveh, S., Ryu, S., et al. (2004) Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. **Mol Cell Proteomics**, 3: (10): 984-997.

Rampino, N., Yamamoto, H., Ionov, Y., et al. (1997) Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. **Science**, 275: (5302): 967-969.

Ranganathan, S., Williams, E., Ganchev, P., et al. (2005) Proteomic profiling of cerebrospinal fluid identifies biomarkers for amyotrophic lateral sclerosis. **J Neurochem**, 95: (5): 1461-1471.

Ransohoff, D.F. (2005) Bias as a threat to the validity of cancer molecular-marker research. **Nat Rev Cancer**, 5: (2): 142-149.

Ratajczak, J., Miekus, K., Kucia, M., et al. (2006a) Embryonic stem cell-derived microvesicles reprogram hematopoietic progenitors: evidence for horizontal transfer of mRNA and protein delivery. **Leukemia**, 20: (5): 847-856.

Ratajczak, J., Wysoczynski, M., Hayek, F., et al. (2006b) Membrane-derived microvesicles: important and underappreciated mediators of cell-to-cell communication. **Leukemia**, 20: (9): 1487-1495.

Razani, B. and Lisanti, M.P. (2001) Caveolin-deficient mice: insights into caveolar function human disease. **J Clin Invest**, 108: (11): 1553-1561.

Redman, C.W. and Sargent, I.L. (2007) Microparticles and immunomodulation in pregnancy and pre-eclampsia. **J Reprod Immunol**, 76: (1-2): 61-67.

Reed, J.C. (1998) Bcl-2 family proteins. **Oncogene**, 17: (25): 3225-3236.

Rifai, N., Gillette, M.A. and Carr, S.A. (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. **Nat Biotechnol**, 24: (8): 971-983.

Robertson, C., Booth, S.A., Beniac, D.R., et al. (2006) Cellular prion protein is released on exosomes from activated platelets. **Blood**, 107: (10): 3907-3911.

Roepstorff, P. and Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. **Biomed Mass Spectrom**, 11: (11): 601.

Roessler, M., Rollinger, W., Mantovani-Endl, L., et al. (2006) Identification of PSME3 as a novel serum tumor marker for colorectal cancer by combining two-dimensional polyacrylamide gel electrophoresis with a strictly mass spectrometry-based approach for data analysis. **Mol Cell Proteomics**, 5: (11): 2092-2101.

Roessler, M., Rollinger, W., Palme, S., et al. (2005) Identification of nicotinamide N-methyltransferase as a novel serum tumor marker for colorectal cancer. **Clin Cancer Res**, 11: (18): 6550-6557.

Rondepierre, F., Bouchon, B., Papon, J., et al. (2009) Proteomic studies of B16 lines: involvement of annexin A1 in melanoma dissemination. **Biochim Biophys Acta**, 1794: (1): 61-69.

Ronnov-Jessen, L., Petersen, O.W. and Bissell, M.J. (1996) Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. **Physiol Rev**, 76: (1): 69-125.

Rottoli, P., Magi, B., Perari, M.G., et al. (2005) Cytokine profile and proteome analysis in bronchoalveolar lavage of patients with sarcoidosis, pulmonary fibrosis associated with systemic sclerosis and idiopathic pulmonary fibrosis. **Proteomics**, 5: (5): 1423-1430.

Rozmyslowicz, T., Majka, M., Kijowski, J., et al. (2003) Platelet- and megakaryocyte-derived microparticles transfer CXCR4 receptor to CXCR4-null cells and make them susceptible to infection by X4-HIV. **Aids**, 17: (1): 33-42.

Russo, M.W., Murray, S.C., Wurzelmann, J.I., et al. (1997) Plasma selenium levels and the risk of colorectal adenomas. **Nutr Cancer**, 28: (2): 125-129.

Sardana, G., Jung, K., Stephan, C., et al. (2008) Proteomic analysis of conditioned media from the PC3, LNCaP, and 22Rv1 prostate cancer cell lines: discovery and validation of candidate prostate cancer biomarkers. **J Proteome Res**, 7: (8): 3329-3338.

Sardana, G., Marshall, J. and Diamandis, E.P. (2007) Discovery of candidate tumor markers for prostate cancer via proteomic analysis of cell culture-conditioned medium. **Clin Chem**, 53: (3): 429-437.

Sarkissian, G., Fergelot, P., Lamy, P.J., et al. (2008) Identification of pro-MMP-7 as a serum marker for renal cell carcinoma by use of proteomic analysis. **Clin Chem**, 54: (3): 574-581.

Sato, K., Sasaki, K., Akiyama, Y., et al. (2001) Mass spectrometric high-throughput analysis of serum-free conditioned medium from cancer cell lines. **Cancer Lett**, 170: (2): 153-159.

Sauter, E.R., Shan, S., Hewett, J.E., et al. (2005) Proteomic analysis of nipple aspirate fluid using SELDI-TOF-MS. **Int J Cancer**, 114: (5): 791-796.

Schorey, J.S. and Bhatnagar, S. (2008) Exosome function: from tumor immunology to pathogen biology. **Traffic**, 9: (6): 871-881.

Schrohl, A.S., Wurtz, S., Kohn, E., et al. (2008) Banking of biological fluids for studies of disease-associated protein biomarkers. **Mol Cell Proteomics**, 7: (10): 2061-2066.

Sharp, S. and Workman, P. (2006) Inhibitors of the HSP90 molecular chaperone: current status. **Adv Cancer Res**, 95: 323-348.

Shedden, K., Xie, X.T., Chandaroy, P., et al. (2003) Expulsion of small molecules in vesicles shed by cancer cells: association with gene expression and chemosensitivity profiles. **Cancer Res**, 63: (15): 4331-4337.

Shen, J., Person, M.D., Zhu, J., et al. (2004) Protein expression profiles in pancreatic adenocarcinoma compared with normal pancreatic tissue and tissue affected by pancreatitis as detected by two-dimensional gel electrophoresis and mass spectrometry. **Cancer Res**, 64: (24): 9018-9026.

Shevchenko, A., Loboda, A., Shevchenko, A., et al. (2000) MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. **Anal Chem**, 72: (9): 2132-2141.

Shevchenko, A., Wilm, M., Vorm, O., et al. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. **Anal Chem**, 68: (5): 850-858.

Shi, Y., Elmets, C.A., Smith, J.W., et al. (2007) Quantitative proteomes and in vivo secretomes of progressive and regressive UV-induced fibrosarcoma tumor cells: mimicking tumor microenvironment using a dermis-based cell-trapped system linked to tissue chamber. **Proteomics**, 7: (24): 4589-4600.

Sidhu, S.S., Mengistab, A.T., Tauscher, A.N., et al. (2004) The microvesicle as a vehicle for EMMPRIN in tumor-stromal interactions. **Oncogene**, 23: (4): 956-963.

Simpson, R.J., Bernhard, O.K., Greening, D.W., et al. (2008) Proteomics-driven cancer biomarker discovery: looking to the future. **Curr Opin Chem Biol**, 12: (1): 72-77.

Sims, P.J., Faioni, E.M., Wiedmer, T., et al. (1988) Complement proteins C5b-9 cause release of membrane vesicles from the platelet surface that are enriched in the membrane receptor for coagulation factor Va and express prothrombinase activity. **J Biol Chem**, 263: (34): 18205-18212.

Sims, P.J., Wiedmer, T., Esmon, C.T., et al. (1989) Assembly of the platelet prothrombinase complex is linked to vesiculation of the platelet plasma membrane. Studies in Scott syndrome: an isolated defect in platelet procoagulant activity. **J Biol Chem**, 264: (29): 17049-17057.

Singh, N., Gemmell, C.H., Daly, P.A., et al. (1995) Elevated platelet-derived microparticle levels during unstable angina. **Can J Cardiol**, 11: (11): 1015-1021.

Sitek, B., Luttges, J., Marcus, K., et al. (2005) Application of fluorescence difference gel electrophoresis saturation labelling for the analysis of microdissected precursor lesions of pancreatic ductal adenocarcinoma. **Proteomics**, 5: (10): 2665-2679.

Skog, J., Wurdinger, T., van Rijn, S., et al. (2008) Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. **Nat Cell Biol**, 10: (12): 1470-1476.

Smirnov, D.A., Zweitzig, D.R., Foulk, B.W., et al. (2005) Global gene expression profiling of circulating tumor cells. **Cancer Res**, 65: (12): 4993-4997.

Smith, L.T., Otterson, G.A. and Plass, C. (2007) Unraveling the epigenetic code of cancer for therapy. **Trends Genet**, 23: (9): 449-456.

Soong, R., Shah, N., Salto-Tellez, M., et al. (2008) Prognostic significance of thymidylate synthase, dihydropyrimidine dehydrogenase and thymidine phosphorylase protein expression in colorectal cancer patients treated with or without 5-fluorouracil-based chemotherapy. **Ann Oncol**, 19: (5): 915-919.

Sorensen, N.M., Schrohl, A.S., Jensen, V., et al. (2008) Comparative studies of tissue inhibitor of metalloproteinases-1 in plasma, serum and tumour tissue extracts from patients with primary colorectal cancer. **Scand J Gastroenterol**, 43: (2): 186-191.

Soroush, A.R., Zadeh, H.M., Moemeni, M., et al. (2004) Plasma prolactin in patients with colorectal cancer. **BMC Cancer**, 4: 97.

Spees, J.L., Olson, S.D., Whitney, M.J., et al. (2006) Mitochondrial transfer between cells can rescue aerobic respiration. **Proc Natl Acad Sci U S A**, 103: (5): 1283-1288.

Stafford, G.C., Kelley, P.E., Syka, J.E.P., Reynolds, W.E., and Todd, J.F.J. (1984) Recent improvements in and analytical applications of advanced ion-trap technology. **INT J Mass Spectrom Ion Processes**, 60: (85-98): 85-98.

Steen, H. and Mann, M. (2004) The ABC's (and XYZ's) of peptide sequencing. **Nat Rev Mol Cell Biol**, 5: (9): 699-711.

Stein, J.M. and Luzio, J.P. (1991) Ectocytosis caused by sublytic autologous complement attack on human neutrophils. The sorting of endogenous plasma-membrane proteins and lipids into shed vesicles. **Biochem J**, 274 (Pt 2): 381-386.

Streckfus, C., Bigler, L., Dellinger, T., et al. (2000) The presence of soluble c-erbB-2 in saliva and serum among women with breast carcinoma: a preliminary study. **Clin Cancer Res**, 6: (6): 2363-2370.

Sugarbaker, P.H. and Wilson, R.E. (1976) Using celioscopy to determine stages of intra-abdominal malignant neoplasms. **Arch Surg**, 111: (1): 41-44.

Sun, J. and Hemler, M.E. (2001) Regulation of MMP-1 and MMP-2 production through CD147/extracellular matrix metalloproteinase inducer interactions. **Cancer Res**, 61: (5): 2276-2281.

Sun, Z., Fu, X., Zhang, L., et al. (2004) A protein chip system for parallel analysis of multi-tumor markers and its application in cancer detection. **Anticancer Res**, 24: (2C): 1159-1165.

Tang, W. and Hemler, M.E. (2004) Caveolin-1 regulates matrix metalloproteinases-1 induction and CD147/EMMPRIN cell surface clustering. **J Biol Chem**, 279: (12): 11112-11118.

Tantipaiboonwong, P., Sinchaikul, S., Sriyam, S., et al. (2005) Different techniques for urinary protein analysis of normal and lung cancer patients. **Proteomics**, 5: (4): 1140-1149.

Taraboletti, G., D'Ascenzo, S., Borsotti, P., et al. (2002) Shedding of the matrix metalloproteinases MMP-2, MMP-9, and MT1-MMP as membrane vesicle-associated components by endothelial cells. **Am J Pathol**, 160: (2): 673-680.

Taylor, P.M., Woodfield, R.J., Hodgkin, M.N., et al. (2002) Breast cancer cell-derived EMMPRIN stimulates fibroblast MMP2 release through a phospholipase A(2) and 5-lipoxygenase catalyzed pathway. **Oncogene**, 21: (37): 5765-5772.

Theodorescu, D., Fliser, D., Wittke, S., et al. (2005) Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine. **Electrophoresis**, 26: (14): 2797-2808.

Thery, C., Boussac, M., Veron, P., et al. (2001) Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles. **J Immunol**, 166: (12): 7309-7318.

Thompson, D.A. and Weigel, R.J. (1998) hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines. **Biochem Biophys Res Commun**, 251: (1): 111-116.

Thomsen, L.S., Jochumsen, K.M. and Mogensen, O. (2006) Carcinoma in situ cervicis uteri and inheritance--a Danish twin study. **Gynecol Oncol**, 103: (2): 688-691.

Thomson, D.M., Krupey, J., Freedman, S.O., et al. (1969) The radioimmunoassay of circulating carcinoembryonic antigen of the human digestive system. **Proc Natl Acad Sci U S A**, 64: (1): 161-167.

Thorns, C., Feller, A.C. and Merz, H. (2002) EMMPRIN (CD 174) is expressed in Hodgkin's lymphoma and anaplastic large cell lymphoma. An immunohistochemical study of 60 cases. **Anticancer Res**, 22: (4): 1983-1986.

Toole, B.P. (2003) Emmprin (CD147), a cell surface regulator of matrix metalloproteinase production and function. **Curr Top Dev Biol**, 54: 371-389.

Unlu, M., Morgan, M.E. and Minden, J.S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. **Electrophoresis**, 18: (11): 2071-2077.

Valenti, R., Huber, V., Filipazzi, P., et al. (2006) Human tumor-released microvesicles promote the differentiation of myeloid cells with transforming growth factor-beta-mediated suppressive activity on T lymphocytes. **Cancer Res**, 66: (18): 9290-9298.

Valenti, R., Huber, V., Iero, M., et al. (2007) Tumor-released microvesicles as vehicles of immunosuppression. **Cancer Res**, 67: (7): 2912-2915.

van der Flier, L.G., Haegebarth, A., Stange, D.E., et al. (2009) OLFM4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells. **Gastroenterology**, 137: (1): 15-17.

VanWijk, M.J., VanBavel, E., Sturk, A., et al. (2003) Microparticles in cardiovascular diseases. **Cardiovasc Res**, 59: (2): 277-287.

Varnum, S.M., Covington, C.C., Woodbury, R.L., et al. (2003) Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer. **Breast Cancer Res Treat**, 80: (1): 87-97.

Veenstra, T.D.a.Y.I., J.R. (ed.) (2006) **Proteomics for Biological Discovery.**: John Wiley and Sons.

Volmer, M.W., Stuhler, K., Zapatka, M., et al. (2005) Differential proteome analysis of conditioned media to detect Smad4 regulated secreted biomarkers in colon cancer. **Proteomics**, 5: (10): 2587-2601.

von Haller, P.D., Yi, E., Donohoe, S., et al. (2003) The application of new software tools to quantitative protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry: II. Evaluation of tandem mass spectrometry methodologies for large-scale protein analysis, and the application of statistical tools for data analysis and interpretation. **Mol Cell Proteomics**, 2: (7): 428-442.

Wallace-Brodeur, R.R. and Lowe, S.W. (1999) Clinical implications of p53 mutations. **Cell Mol Life Sci**, 55: (1): 64-75.

Wang, H., Clouthier, S.G., Galchev, V., et al. (2005) Intact-protein-based high-resolution three-dimensional quantitative analysis system for proteome profiling of biological fluids. **Mol Cell Proteomics**, 4: (5): 618-625.

Wang, M.C., Papsidero, L.D., Kuriyama, M., et al. (1981) Prostate antigen: a new potential marker for prostatic cancer. **Prostate**, 2: (1): 89-96.

Ward, D.G., Cheng, Y., N'Kontchou, G., et al. (2006a) Changes in the serum proteome associated with the development of hepatocellular carcinoma in hepatitis C-related cirrhosis. **Br J Cancer**, 94: (2): 287-292.

Ward, D.G., Cheng, Y., N'Kontchou, G., et al. (2006b) Preclinical and post-treatment changes in the HCC-associated serum proteome. **Br J Cancer**, 95: (10): 1379-1383.

Ward, D.G., Nyangoma, S., Joy, H., et al. (2008a) Proteomic profiling of urine for the detection of colon cancer. **Proteome Sci**, 6: 19.

Ward, D.G., Roberts, K., Stonelake, P., et al. (2008b) SELDI-TOF-MS determination of hepcidin in clinical samples using stable isotope labelled hepcidin as an internal standard. **Proteome Sci**, 6: 28.

Ward, D.G., Suggett, N., Cheng, Y., et al. (2006c) Identification of serum biomarkers for colon cancer by proteomic analysis. **Br J Cancer**, 94: (12): 1898-1905.

Warner, J.R. and McIntosh, K.B. (2009) How common are extraribosomal functions of ribosomal proteins? **Mol Cell**, 34: (1): 3-11.

Whitesell, L. and Lindquist, S.L. (2005) HSP90 and the chaperoning of cancer. **Nat Rev Cancer**, 5: (10): 761-772.

Wieckowski, E. and Whiteside, T.L. (2006) Human tumor-derived vs dendritic cell-derived exosomes have distinct biologic roles and molecular profiles. **Immunol Res**, 36: (1-3): 247-254.

Williams, A.C., Hague, A., Elder, D.J., et al. (1996) In vitro models for studying colorectal carcinogenesis: cellular and molecular events including APC and Rb cleavage in the control of proliferation, differentiation and apoptosis. **Biochim Biophys Acta**, 1288: (1): F9-19.

Wilson, R., Whitelock, J.M. and Bateman, J.F. (2009) Proteomics makes progress in cartilage and arthritis research. **Matrix Biol**, 28: (3): 121-128.

Winkler, C., Denker, K., Wortelkamp, S., et al. (2007) Silver- and Coomassie-staining protocols: detection limits and compatibility with ESI MS. **Electrophoresis**, 28: (12): 2095-2099.

Wolf, P. (1967) The nature and significance of platelet products in human plasma. **Br J Haematol**, 13: (3): 269-288.

Wu, C.C., Chen, H.C., Chen, S.J., et al. (2008) Identification of collapsin response mediator protein-2 as a potential marker of colorectal carcinoma by comparative analysis of cancer cell secretomes. **Proteomics**, 8: (2): 316-332.

Wu, C.C., Chien, K.Y., Tsang, N.M., et al. (2005) Cancer cell-secreted proteomes as a basis for searching potential tumor markers: nasopharyngeal carcinoma as a model. **Proteomics**, 5: (12): 3173-3182.

Wu, S.L., Amato, H., Biringer, R., et al. (2002) Targeted proteomics of low-level proteins in human plasma by LC/MSn: using human growth hormone as a model system. **J Proteome Res**, 1: (5): 459-465.

Xiao, T., Ying, W., Li, L., et al. (2005) An approach to studying lung cancer-related proteins in human blood. **Mol Cell Proteomics**, 4: (10): 1480-1486.

Xue, H., Lu, B. and Lai, M. (2008) The cancer secretome: a reservoir of biomarkers. **J Transl Med**, 6: 52.

Yamaguchi, H. and Condeelis, J. (2007) Regulation of the actin cytoskeleton in cancer cell migration and invasion. **Biochim Biophys Acta**, 1773: (5): 642-652.

Yan, H. and Zhou, W. (2004) Allelic variations in gene expression. **Curr Opin Oncol**, 16: (1): 39-43.

Yao, X., Freas, A., Ramirez, J., et al. (2001) Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. **Anal Chem**, 73: (13): 2836-2842.

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

Ye X, L.B., Andresson T, Blonder J. (2009) 18O stable isotope labeling in MS-based proteomics. **Brief Funct Genomic Proteomic**, 8: (2): 136-144.

Yokoyama, Y., Kuramitsu, Y., Takashima, M., et al. (2004) Proteomic profiling of proteins decreased in hepatocellular carcinoma from patients infected with hepatitis C virus. **Proteomics**, 4: (7): 2111-2116.

Yonish-Rouach, E., Resnitzky, D., Lotem, J., et al. (1991) Wild-type p53 induces apoptosis of myeloid leukaemic cells that is inhibited by interleukin-6. **Nature**, 352: (6333): 345-347.

Yurchenko, V., O'Connor, M., Dai, W.W., et al. (2001) CD147 is a signaling receptor for cyclophilin B. **Biochem Biophys Res Commun**, 288: (4): 786-788.

Yurchenko, V., Pushkarsky, T., Li, J.H., et al. (2005) Regulation of CD147 cell surface expression: involvement of the proline residue in the CD147 transmembrane domain. **J Biol Chem**, 280: (17): 17013-17019.

Yurchenko, V., Zybarth, G., O'Connor, M., et al. (2002) Active site residues of cyclophilin A are crucial for its signaling activity via CD147. **J Biol Chem**, 277: (25): 22959-22965.

Zerr, I., Bodemer, M., Otto, M., et al. (1996) Diagnosis of Creutzfeldt-Jakob disease by two-dimensional gel electrophoresis of cerebrospinal fluid. **Lancet**, 348: (9031): 846-849.

Zhang, H., Yi, E.C., Li, X.J., et al. (2005a) High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. **Mol Cell Proteomics**, 4: (2): 144-155.

Zhang, J., Liu, W.L., Tang, D.C., et al. (2002) Identification and characterization of a novel member of olfactomedin-related protein family, hGC-1, expressed during myeloid lineage development. **Gene**, 283: (1-2): 83-93.

Zhang, J.S., Gong, A., Cheville, J.C., et al. (2005b) AGR2, an androgen-inducible secretory protein overexpressed in prostate cancer. **Genes Chromosomes Cancer**, 43: (3): 249-259.

Zhang, X., Huang, Q., Yang, Z., et al. (2004) GW112, a novel antiapoptotic protein that promotes tumor growth. **Cancer Res**, 64: (7): 2474-2481.

Zhong, L., Roybal, J., Chaerkady, R., et al. (2008) Identification of secreted proteins that mediate cell-cell interactions in an in vitro model of the lung cancer microenvironment. **Cancer Res**, 68: (17): 7237-7245.

Zhou, G., Li, H., DeCamp, D., et al. (2002) 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. **Mol Cell Proteomics**, 1: (2): 117-124.

Zhu, H., Lam, D.C., Han, K.C., et al. (2007) High resolution analysis of genomic aberrations by metaphase and array comparative genomic hybridization identifies candidate tumour genes in lung cancer cell lines. **Cancer Lett**, 245: (1-2): 303-314.

Zwaal, R.F., Comfurius, P., and Bevers, E.M. (2004) Scott syndrome, a bleeding disorder caused by defective scrambling of membrane phospholipids. **Biochim Biophys Acta**, 1636: 119-128.

Zwaal, R.F. and Schroit, A.J. (1997) Pathophysiologic implications of membrane phospholipid asymmetry in blood cells. **Blood**, 89: (4): 1121-1132.

Zybailov, B., Coleman, M.K., Florens, L., et al. (2005) Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. **Anal Chem**, 77: (19): 6218-6224.