Lung Cancer Biomarker Discovery using Proteomic Techniques

A thesis submitted for the degree of Ph.D.

by

Damian Pollard, B.Sc. Hons

Research work described in this thesis was performed

under the supervision of

Prof. Martin Clynes and Dr. Paul Dowling

National Institute for Cellular Biotechnology

Dublin City University

July 2013

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Ph.D. is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ ID No.: 50834459

Date: _____

Abbreviations

12-HETE	-	12-hydroxyeicosatetraenoic acid
2D-DIGE	-	2-dimensional difference gel electrophoresis
AA	-	Arachidonic acid
Ab	-	Antibody
AC	-	Atypical carcinoid
ACN	-	Acetonitrile
AD	-	Adenocarcinoma
ADP	-	Adenosine Diphosphate
Ag	-	Antigen
AIS	-	Amplified in Squamous cell carcinoma
APC	-	Adenomotous Polyposis Coli
APP	-	Acute Phase Proteins
ATCC	-	American Tissue Culture Collection
ATP	-	Adenosine Triphosphate
AUC	-	Area Under the Curve
BAC	-	Bronchioalveolar carcinoma
BAF	-	Bronchioalveolar lavage fluid
BCL2	-	B cell CLC/Lymphoma 2
BCR	-	B-cell Receptor
BH4	-	5,6,7,8-tetrahydrobiopterin
BLD	-	Benign Lung Disease
BRS-3	-	Bombesin Receptor Subtype-3
BSA	-	Bovine Serum Albumin
BVA	-	Biological Variation Analysis
CA-125	-	Cancer antigen 125
cDNA	-	Complementary DNA
CEA	-	Carcinoembryonic antigen
CHAPS	-	3- ((3-Cholamidopropyl)dimethylammonio)-1-Propanesulfonic Acid
CID	-	Collision-induced Dissociation
c-MYC	-	Cellular Myelocytomatosis Oncogene
c-SCLC	-	Combined Small Cell Lung Cancer

CT	-	Computed Tomography
CYFRA 21-1	-	Cytokeratin 19 Fragment
DAB	-	Diaminobenzidine
DDH	-	Dihydrodiol Dehydrogenase
DIA	-	Differential In-gel Analysis
DIGE	-	Difference Gel Electrophoresis
DMEM	-	Dulbeccos Modified Eagles Media
DMF	-	Dimethylformamide
DMSO	-	Dimethyl Sulfoximide
DNA	-	Deoxyribonucleic Acid
DTT	-	Dithiothreitol
ECM	-	Extra Cellular Matrix
EDTA	-	Ethylenediaminetetraacetic acid
EGF	-	Epidermal Growth Factor
EGFR	-	Epidermal Growth Factor Receptor
ELISA	-	Enzyme-linked Immunosorbent Assay
ESI	-	Electrospray Ionisation
EtOH	-	Ethanol
FCS	-	Foetal Calf Serum
FDA	-	Food and Drug Administration
FFPE	-	Formalin-fixed Paraffin Embedded Tissue
GAPDH	-	Glyceraldehyde-3-Phosphate Dehydrogenase
GC	-	Gas Chromatography
GRP	-	Gastrin-releasing Peptide
GRPR	-	Gastrin-releasing Peptide Receptor
GTP	-	Guanosine Triphosphate
H_2O_2	-	Hydrogen Peroxide
Hb	-	Haemoglobin
HCL	-	Hydrochloric Acid
HGF	-	Hepatocyte Growth factor
Hip	-	HSC70-interacting Protein
HIV	-	Human Immunodeficiency Virus
HnRNP	-	Heterogeneous nuclear ribonucleoproteins
Hp	-	Haptoglobin

HPLC	-	High-performance Liquid Chromatography
HRP	-	Horseradish Peroxidase
HSC70	-	Heat-shock Cognate 70
HSF	-	Heat-shock Factor
HSP	-	Heat-shock Protein
IASLC	-	International Association for the Study of Lung Cancer
iCAT	-	Isotope-coded Affinity Tags
ID	-	Identification
IEF	-	Isoelectric Focusing
Ig	-	Immunoglobulin
IHC	-	Immunohistochemistry
IMS	-	Industrial methylated Spirits
IP	-	Immunoprecipitation
IPG	-	Immobilised pH Gradient
iTRAQ	-	Isobaric Tags for Relative and Absolute Quantitation
kDA	-	Kilo Dalton
Kin	-	Kinesin
LCC	-	Large Cell Carcinoma
LC-MS	-	Liquid Chromatography – Mass Spectrometry
LCNEC	-	Large Cell Neuroendocrine Carcinoma
LDCT	-	Low Radiation Dose CT
LDH	-	Lactate Dehydrogenase
LIFE	-	Light-induced Fluorescence Endoscopy
LMW	-	Low Molecular Weight
LOH	-	Loss of Heterozygosity
LOX	-	Lipoxygenase
mA	-	Milliamps
MAb	-	Monoclonal Antibody
MALDI-TOF	-	Matrix-assisted laser desorption/ionization Time of Flight
MBAA	-	Multiplex Bead Array Assay
MeOH	-	Methanol
miRNA	-	Micro RNA
MMP	-	Matrix Metalloproteinase
MOPS	-	3-(N-morpholino)propanesulfonic acid

MRM	-	Multiple Reaction Monitoring
mRNA	-	Messenger RNA
MS	-	Mass Spectrometry
MS/MS	-	Tandem Mass Spectrometry
MudPIT	-	Multidimensional Protein Identification Technology
MW	-	Molecular Weight
m/z	-	Mass-to-charge Ratio
NaOH	-	Sodium Hydroxide
Na_2CO_3	-	Sodium Carbonate
NCI	-	National Cancer Institute
NCTCC	-	National Cell and Tissue Culture Collection
NET	-	Neuroendocrine Tumour
NH_4HCO_3	-	Ammonium Bicarbonate
NICB	-	National Institute for Cellular Biotechnology
NK	-	Natural Killer
NMB	-	Neuromedin B
NMBR	-	Neuromedin B Receptor
NMR	-	Nuclear Magnetic Resonance spectroscopy
NO	-	Nitric Oxide
NSCLC	-	Non-small Cell Lung Cancer
NSE	-	Neuro-specific Endolase
pAb	-	Polyclonal Antibody
PAGE	-	Polyacrylamide Gel Electrophoresis
PBS	-	Phosphate Buffered Saline
PDGF	-	Platelet-derived Growth Factor
PEP	-	Phosphophenolpyruvate
PET	-	Positron Emission Tomography
p*I*	-	Isoelectric Point
PKM2	-	Pyruvate Kinase M2
ProGRP	-	Pro-gastrin-releasing peptide
Prx	-	Peroxiredoxin
PVDF	-	Polyvinylidene Fluoride
QC	-	Quality Control
Ras	-	Rat Sarcoma

RBP	-	RNA binding protein
RNA	-	Ribonucleic Acid
RNAi	-	RNA interference
ROC	-	Receiver Operating Characteristic
ROS	-	Reactive Oxygen Specie
RP	-	Reverse Phase
RT	-	Room Temperature
SCC	-	Squamous cell carcinoma
SCCA	-	Squamous cell carcinoma Antigen
SCLC	-	Small Cell Lung Cancer
SCR	-	Scrambled
SCX	-	Strong Cation Exchange
SDS	-	Sodium Dodecyl Sulfate
SELDI	-	Surface-enhanced laser desorption/ionization
SF	-	Serum Free
SILAC	-	Stable Isotope Labelling by Amino acids in Cell culture
siRNA	-	Small Interfering RNA
SOP	-	Standard Operating Procedure
ST13	-	Suppression of Tumourigenicity 13
TB	-	Tuberculosis
TBS	-	Tris Buffered Saline
TC	-	Typical Carcinoid
TFA	-	Trifluoroacetic Acid
TGF	-	Transforming Growth Factors
TIMP	-	Tissue Inhibitor of MMP
TK	-	Tyrosine Kinase
TKI	-	Tyrosine Kinase Inhibitor
TMB	-	3, 3',5 ,5'-Tetramethylbenzidine
TN-C	-	Tenascin C
TNF	-	Tumour Necrosis Factor
TPA	-	Tissue Polypeptide Antigen
Tris	-	Tris (hydroxymethyl) aminomethane
TSG	-	Tumour Suppressor Gene
UHP	-	Ultra High Pure Water

UHPLC	-	Ultra High Performance Liquid Chromatography
VEGF	-	Vascular Endothelial Growth Factor
vol/vol	-	Volume to Volume ratio
WHO	-	World Health Organisation

ABSTRACT

Lung cancer has the highest mortality rate of any cancer, often due to the fact that it is detected at a late stage in its progression when it has already metastasised. The levels of certain biomarkers, such as proteins, metabolites and chemokines, in biological fluid or tissue could potentially detect cancer at an early stage, determine cancer subtype, or monitor the sensitivity to cancer treatment.

Currently available lung cancer markers lack the sensitivity and specificity to be of great benefit and there is room for improvement. The research in this thesis aims to discover new biomarkers, using proteomic techniques, with the potential to improve or supersede those used at present for diagnosis and prognosis of lung cancer.

Discovery phase was performed on conditioned media of lung cancer cell lines using 2D-DIGE in the hope it might mimic the serum/plasma environment of lung cancer patients. Further discovery phase was performed on serum using immunodepletion, proteominer, 2D-DIGE, label-free mass spectrometry followed by pathway analysis, metabolomic analysis, and multiplex assay analysis of cancer panels and a matrix-metalloproteinase panel.

Validation in serum and plasma was performed using ELISAs, biochemical assays, and multiplex platforms, and in tissue using immunohistochemistry. Lung cancer subtypes examined in validation phase were squamous cell carcinoma, adenocarcinoma, and small cell lung cancer. There were insufficient serum/plasma samples to include a large cell carcinoma group.

Potential biomarkers discovered include hnRNPA2B1, pyruvate kinase M2 (PKM2), HSC70-interacting protein (Hip), tenascin C, vascular endothelial growth factor alpha (VEGF-α), MMP-1, MMP-8, MMP-9, 12-HETE, and phenylalanine.

In serum hnRNPA2B1, PKM2, Hip, tenascin C, VEGF-α, MMP-1, -8, and -9, and 12-HETE were increased in cancer compared to normal. Phenylalanine had similar levels in normal and small cell lung cancer but was decreased in non-small cell lung cancer. Of those examined in plasma hnRNPA2B1, VEGF-α, MMP-1 and -9, and

12-HETE were decreased in cancer compared to normal whereas PKM2 and tenascin C were increased. Where benign lung disease controls were used the markers were present at levels similar to lung cancer except for MMP-9 in plasma, VEGF-α in serum, and PKM2 in both serum and plasma.

Immunohistochemistry of tissue showed an overall increase of expression of hnRNPA2B1 and HSC70-interacting protein in cancer tissue compared to normal.

Functional assays were performed on hnRNPA2B1 showing its potential role in invasion and migration of lung cancer; its knockdown in the DLKP-M lung cancer cell line with two siRNA's showed a 57% and 44% decrease in invasion and a 32% and 39% decrease in migration respectively.

The research provided in this thesis demonstrates the importance for intensive validation before conclusions can be drawn on the overall usefulness of candidate lung cancer markers; a good serum marker does not necessarily make a good plasma marker and a marker that differentiates normal conditions from cancer conditions does not necessarily differentiate lung diseases/benign tumour from cancer.

Table of Contents

CHAPTER ONE

INTRODUCTION

1.1 Cancer Biomarkers

After the genomic revolution science expected great progress to be made in the area of biomedical discovery, especially in relation to early diagnosis and identification of cancer risk. This has not been the case and the anticipated new range of diagnostic tools has been slow to emerge. Biomarkers are often proteins which can indicate the presence of a disease relatively early (Etzioni, Urban et al. 2003), and can also monitor the progression or recession of a cancer during treatment (Gail, Muenz et al. 1988). A biomarker, or more probably a panel of biomarkers (Han, Liew et al. 2008, Millar, Graham et al. 2011), ideally can be screened quite easily and the presence of these markers in the blood or other biological body fluids would allow this (Toyama, Nakagawa et al. 2011). Biological fluids are considered a better source of biomarkers than tissue due to several factors, including ease of accessibility, avoiding risks of invasive tissue sampling through biopsies, relative low cost obtainment, availability of monitoring based on multiple sampling, and the potential for large-scale, valuable prognostic/diagnostic tests (Good, Thongboonkerd et al. 2007). However, the detection of tissue-specific biomarkers in body fluids requires the biomarker to be identified as disease tissue-specific from thousands of other proteins in circulation and also requires it to be secreted in the first place, whereas tissue can be used to find tumour-specific protein biomarkers more directly from the source (Shiwa, Nishimura et al. 2003).

The hunt for blood-based biomarkers (and indeed from other biological fluids) is based on the hypothesis that tumours secrete specific proteins (including those harbouring characteristic post-translational modifications) that collectively

constitute a molecular fingerprint reflecting physiological conditions (Makawita, Smith et al. 2011). To be able to read and understand this fingerprint through blood samples would allow a non-invasive, unique look into the disease state of a patient.

While blood is a very appropriate non-invasive fluid for monitoring biomarkers (Hanash, Baik et al.), the presence of high abundant proteins such as albumin, transferrin, fibrinogen and IgGs can mask the presence of important, disease-specific low abundant proteins (Fountoulakis, Juranville et al. 2004). The most abundant blood protein, albumin, is present at 30mg/ml, in contrast to most biomarkers which are present in the ng and pg/ml range, meaning a dynamic range of detection of up to 12 orders of magnitude is required, whereas most mass spectrometry instruments can only measure 3-4 orders of magnitude (Hung and Yu). In fact, the 20 most abundant proteins account for 97% of the plasma proteome and some sort of fractionation step is needed to remove these proteins in order to have a chance of detecting the vast numbers of low abundant proteins (Kim and Kim 2007). However, examination of this low-molecular-weight (LMW) range of the circulatory proteome, often called the peptidome, due to the profusion of protein peptides and fragments, has an additional complexity caused by the binding of these disease markers to the high-abundant proteins such as albumin (Rogatsky, Tomuta et al. 2007).

To remove and discard the high abundant proteins could mean losing valuable diagnostic information (Tirumalai, Chan et al. 2003). Interestingly, by binding to albumin these LMW proteins acquire longer half-lives, since albumin protects the bound species from kidney clearance (Lowenthal, Mehta et al. 2005), thus

increasing the length of time these proteins will circulate in the blood, in turn improving the chances they will be detected by a clinical assay or reach their target (Dennis, Zhang et al. 2002). In addition to these problems protein concentrations are dynamic (Anderson and Anderson 2002), sometimes changing noticeably with stress, disease, or treatment, and the proteins themselves can be modified by cleavage or by addition of new functional groups (such as phosphorylation or glycosylation) which may effect their detection.

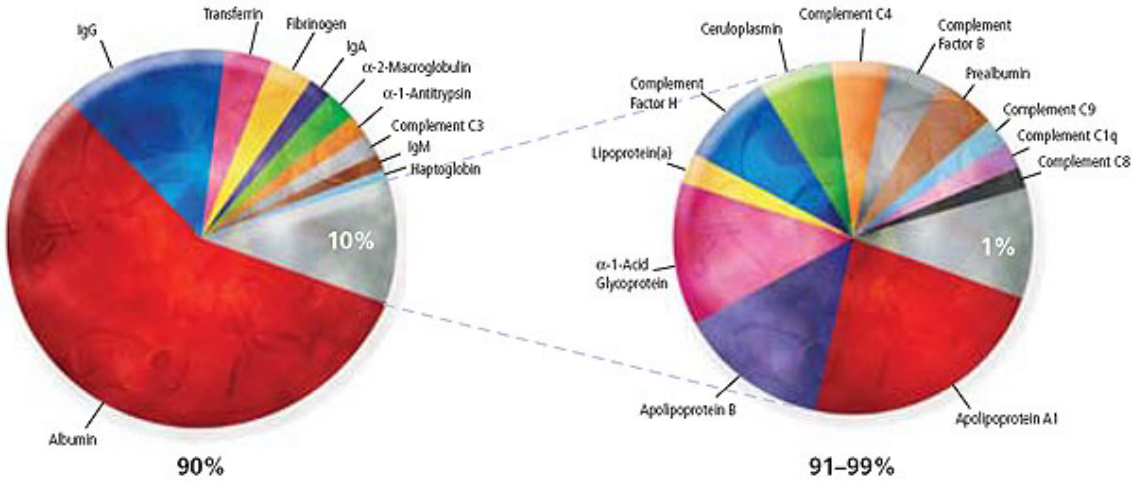


Figure 1.1: Low abundant proteins that may indicate the disease state of a patient are found in only one percent of the total protein content of serum (http://www.sigmaaldrich.com/life-science/proteomics/mass-spectrometry/protein-aqua/absolute-quantification-of-serum-proteome.html).

Examination of the low abundant proteins is helped by removal of high abundant proteins with the use of multiple affinity removal columns which contain affinity-purified polyclonal antibodies as seen with immunodepletion, which can be combined with other fractionation steps (Dayarathna, Hancock et al. 2008), and also

with newer technology such as ProteoMiner, which uses a large library of hexapeptide ligands with a limited number of binding sites that act as unique binders for proteins (Boschetti and Righetti 2008).

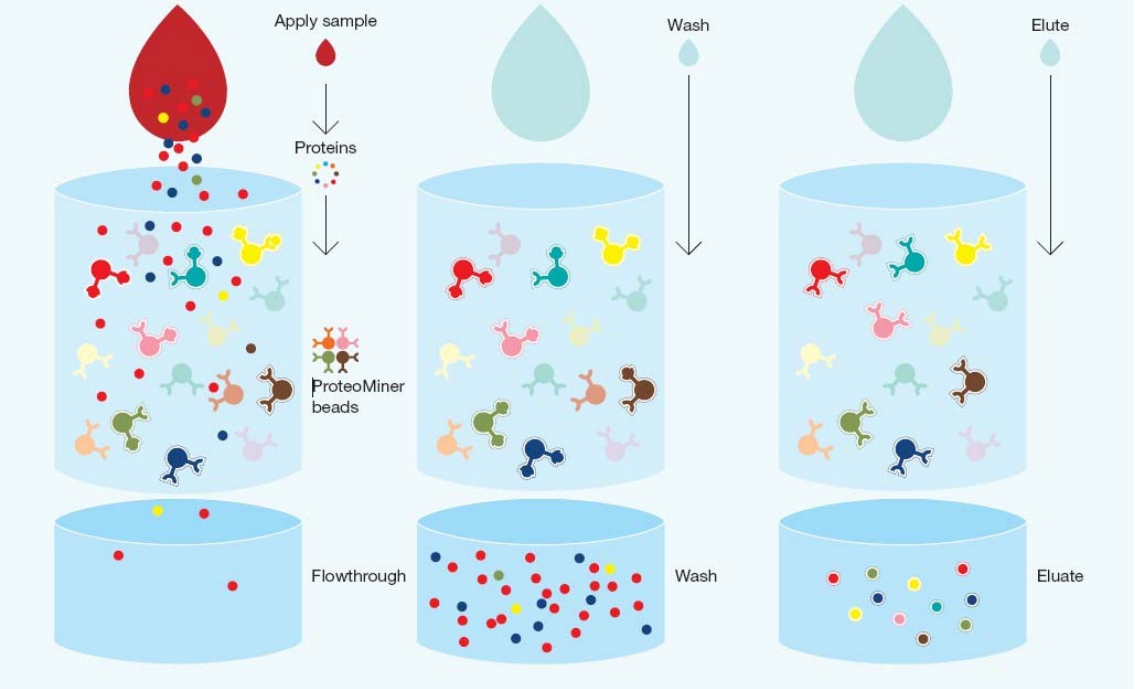


Figure 1.2: Proteominer Technology (BioRad)

ProteoMiner technology compresses the dynamic range of protein concentrations in complex biological samples such as serum by decreasing high-abundance proteins and capturing low-abundance proteins and can be used for quantitative analysis of the latter in complex biological samples (Hartwig, Czibere et al. 2009). ProteoMiner decreases the amount of high-abundance proteins without immunodepletion, preventing the loss of proteins bound to these proteins, which are inadvertently lost with immunodepletion and it enriches and concentrates low-abundance proteins that cannot be detected through traditional methods. It can be used to decrease the dynamic range of the protein concentration in a variety of

samples and is not dependent on a predefined set of antibodies as are immunodepletion columns, and importantly it is compatible with current downstream protein analysis techniques.

Studying the peptidome gives important diagnostic information such as; (i) the quantity of the peptide itself, (ii) the identity of the parent protein and the peptide fragment, and (iii) the nature of the carrier protein to which it is bound. Measuring panels of these peptidome markers instead of single markers can overcome sensitivity and specificity issues seen with single biomarkers and evaluating this combination as a fingerprint of disease is considered the way forward (Petricoin, Ardekani et al. 2002, Petricoin, Ornstein et al. 2004). The ability of biomarkers, whether single or a panel, to achieve the clinically required level of specificity and sensitivity depends on the intended use of the marker. For example, markers used for high-risk screening or relapse monitoring can have much lower specificity than markers for general population screening for rare diseases, which would require almost 100% specificity to be accepted.

It is clear that body fluids, and blood in particular, potentially harbour many markers of different diseases for diagnosis, prognosis and treatment monitoring (Hashiguchi, Tanaka et al. 2009, Shen, Tolic et al. 2010) and there are many proteomic techniques to make the most of this potential (Conrotto and Souchelnytskyi 2008).

1.2 Lung Cancer

Lung cancer is the most common cause of cancer-related death in Ireland and indeed the world. Between 1994 and 2000 there was a 33% increase in the incidence rates of lung cancer in Ireland and it accounts for 20% of all cancer deaths. In the United States in 2007, 203,536 people were diagnosed with lung cancer and of these 158,683 people died from the disease. The five-year survival rate in the U.S. is 49 percent when confined to the lungs, 16 percent when it has spread to the chest and as low as 2 percent when it has spread to other organs. Women with lung cancer are more likely than men to have adenocarcinoma, to be younger at diagnosis, to have smoked less, and to have a first-degree relative with lung cancer (Zang and Wynder 1996).

Small cell lung cancer (SCLC) makes up about 15% of lung cancer cases and non-small lung cancer (NSCLC) makes up the rest. NSCLC is further divided into adenocarcinoma (AD), squamous cell carcinoma (SCC), large cell carcinoma (LCLC), and bronchioalveolar carcinoma (BAC). A majority of SCLC patients present with extensive disease and the survival rate is less than 5% over 5 years. The main cause of lung cancer is smoking so it is easily preventable in most cases but unfortunately diagnosis generally occurs at a late stage when the cancer has metastasised and is therefore difficult to treat effectively. The association of SCLC with smoking is so strong that its diagnosis in a non-smoker is considered exceptional and would be re-evaluated (Rekhtman). Diagnosis at an early stage is essential for improvements in long-term survival.

Strategies used to detect lung cancer early include CT scans, bronchoscopy, and sputum analysis, though none of them have proven to be effective.

1.2.1 Pathogenesis of Lung Cancer

Significant strides have been made in understanding the molecular and cellular pathogenesis of lung cancer with focus being brought on genetic and epigenetic changes of tumour suppressor genes, abnormalitites of protooncogenes, and the role of angiogenesis in the many stages of lung cancer development, as well as detection of molecular abnormalities in preinvasive respiratory lesions. These findings can lead to clinical strategies for risk assessment, chemoprevention, treatment selection, early diagnosis and prognosis, and to provide new targets and methods of treatment for lung cancer patients.

The autocrine and paracrine systems are cell signalling systems that involve, in the former, a cell secreting a hormone or chemical messenger that binds to autocrine receptors on the same cell, and in the latter, secreting a hormone or chemical messenger which targets a nearby cell. A well-characterised autocrine system involves bombesin-like peptides, such as gastrin-releasing peptide (GRP) and neuromedin B (NMB). Bombesin-like peptides are important regulators of human lung development and in normal tissues stimulate growth of bronchial epithelial tissues (Spindel 1996). The family of G protein-coupled bombesin receptors mediate the effects of GRP, which includes GRP receptor (GRPR), NMB receptor (NMBR), and the bombesin receptor subtype 3 (BRS-3). In NSCLC the expression of NMBR is common, whereas that of GRPR and BRS-3 is less frequent (DeMichele, Davis et al. 1994). The NMB gene is expressed in all NSCLC and also

SCLC, and GRP is expressed in 20-60% of SCLC and less frequently in NSCLC. A proliferative response of bronchial cells to GRP as well as long-term tobacco use have been linked to activation of GRPR and it has been reported that GRPR mRNA is more frequently expressed in women than in men in the absence of smoking, and in response to tobacco exposure, expression is activated earlier in women (Cuttitta, Carney et al. 1985).

Another growth factor/receptor complex that could be involved in lung cancer development is the hepatocyte growth factor (HGF) and its receptor. Epithelial cells are stimulated by HGF to proliferate, move, and undergo differentiation. HGF is expressed in NSCLC (Harvey, Warn et al. 1996) and is linked with impaired survival (Siegfried, Weissfeld et al. 1997). The oncogene *MET* encodes the HGF receptor and is expressed in normal lung epithelium, and in both NSCLC and SCLC (Olivero, Rizzo et al. 1996).

Epidermal growth factor (EGFR) regulates epithelial proliferation and differentiation and the *EGFR* gene was found to be overexpressed in 13% of NSCLCs (Reissmann, Koga et al. 1999).

Proto-oncogenes are involved in normal cellular growth and development and their expression is tightly regulated. The *RAS* genes (*HRAS, KRAS, NRAS)* code for guanosine triphosphate (GTP)-binding proteins, and have an important role in signal transduction (which is when an extracellular signalling molecule activates a surface receptor of a cell). *RAS* oncogenes have been detected in about 20-30% of lung adenocarcinomas and in 15-20% of all NSCLC, and in almost all cases the mutation is present in codon 12 of the *KRAS* gene (Rodenhuis and Slebos 1990). Interestingly, *RAS* point mutations (a single base substitution causing the

replacement of a single base nucleotide with another nucleotide of the genetic material and can also include insertions or deletions of a single base pair) are never found in SCLCs (Zochbauer-Muller, Gazdar et al. 2002). Most *KRAS* mutations are G-T transversions of the type that are associated with carcinogens from cigarette smoke (Rodenhuis and Slebos 1990).

The *MYC* proto-oncogene family and their unregulated expression have involvement in cell proliferation and the formation of cancer. Abnormal *MYC* expression is often observed in SCLCs but less frequently so in NSCLCs (Richardson and Johnson 1993).

The *BCL-2* proto-oncogene is involved in the normal pathway for programmed cell death. It is negatively regulated by p53 and protects cells from apoptosis. Because overexpression of *MYC* or *RAS* in fibroblasts can lead to apoptosis it is possible tumour cells overexpress *BCL-2* to overcome these apoptotic signals, and the BCL-2 protein is often expressed in SCLCs and NSCLCs as immunohistochemical studies have shown (Pezzella, Turley et al. 1993, Kaiser, Schilli et al. 1996).

Overexpression of *Notch-3* is often found in NSCLCs and in metastatic lung carcinoma it was found to be translocated from its region on chromosome 19p to chromosome 15q (Hibi, Trink et al. 2000).

As its name suggests, *AIS* (amplified in squamous cell carcinoma) is often amplified in primary lung squamous cell carcinomas. It is a *p53* homologue with multiple protein products and overexpression of these proteins was observed in SCCs known to harbour a high frequency of *p53* mutations (Hibi, Trink et al. 2000).

Loss of function of **tumour suppressor genes (TSGs)** according to Knudson's two-hit hypothesis requires both alleles to be inactivated (Knudson 1971). There

must be deactivation of TSGs and activation of oncogenes for cancer to occur. One allele could be inactivated by mutation or methylation changes for example, whereas the other is inactivated as part of loss of genetic markers by nonreciprocal translocation, deletion, or mitotic recombination in the chromosomal region referred to as allele loss or loss of heterozygosity (LOH). The presence of one or more TSGs in a region of a gene could therefore be assumed if there is consistent LOH for genetic markers at a given locus. A search of LOH on SCLCs and NSCLCs was performed in order to detect new loci that may harbour TSGs with 13 regions showing a preference for SCLC, 7 for NSCLC, and 2 affecting both. Frequent LOH occurred in different regions with NSCLC having more regions than SCLC, which suggests they undergo different genetic alterations (Girard, Zochbauer-Muller et al. 2000).

The most frequent molecular alteration seen in lung cancer is that of the 3p allele loss. However this does not only occur in tumours but in epithelium of smokers without lung cancer, and also dysplasia, hyperplasias, and carcinoma *in situ* in the respiratory epithelium accompanying lung cancers, linking it to early pathogenesis of lung cancer (Wistuba, Behrens et al. 2000). In more than 90% of SCLCs and 80% of NSCLCs a LOH is found involving markers of 3p (Sekido, Fong et al. 1998).

The ***p53* gene** encodes a 53-kDa nuclear protein that acts as a transcription factor to turn on the expression of a DNA damage response program, stops cells going through the late G1 cell cycle phase and triggers apoptosis (Sidransky and Hollstein 1996). Phosphorylation of *p53* is triggered by DNA damage and acts as a specific DNA-binding transcription factor for several genes, the activation of which leads to

apoptosis, cell cycle arrest, and DNA repair (Sekido, Fong et al. 1998). Some of the most common genetic changes seen in cancer are mutations of the *p53* gene which can cause loss of tumour suppressor function and loss of ability to induce apoptosis. Interestingly, *p53* mutations were significantly more likely in smokers compared to non-smokers (Husgafvel-Pursiainen, Boffetta et al. 2000), although it was also more likely in alcohol drinkers who smoke compared with non-drinkers who smoke (Ahrendt, Chow et al. 2000). Immunostaining of abnormal p53 expression was seen in 40-60% of NSCLCs and 40-70% of SCLCs (Nishio, Koshikawa et al. 1996).

The *APC* gene (Adenomotous Polyposis Coli) encodes a protein with roles in signal transduction, intercellular adhesion mediation, cytoskeleton stabilisation, and even perhaps cell cycle and apoptosis regulation (Fearnhead, Britton et al. 2001). In primary lung cancer loss of expression of APC is often observed with aberrant methylation of APC an important mechanism for inactivation of this gene (Virmani, Rathi et al. 2001).

Angiogenesis, the formation of new blood vessels from pre-existing ones, is vital in normal growth and development as well as in wound healing. However this formation of new blood vessels also allows tumours to metastasise and grow. Inducers and inhibitors regulating endothelial cell proliferation and migration known as angiogenic factors can affect vasculature growth patterns, formation, and permeability as well as influence tumour invasion, metastasis and prognosis. Alterations in the oncogenes *p53* and *RAS* are involved in the secretion of angiogenic substances by tumour cells and their precursor cells (Rak, Filmus et al. 1995). The **vascular endothelial growth factor (VEGF)** is frequently expressed by

lung cancers (O'Byrne, Koukourakis et al. 2000). In NSCLC the expression of VEGF was linked with the formation of new blood vessels and was an adverse prognostic factor in patients with NSCLC (Fontanini, Vignati et al. 1997). Activated microvessel density in early operable NSCLC was seen to be lower in the normal lung distal to the tumour or the inner tumour areas compared to the invading front of the tumours and in normal lung adjacent to the tumours (Koukourakis, Giatromanolaki et al. 2000).

Tobacco smoke is the major cause of lung cancer, accounting for about 90% of all cases. In tobacco smoke there are three major classes of carcinogens; polycyclic hydrocarbons, nitrosamines, and aromatic amines (Gazdar and Minna 1997). Subjects sensitive to benzo(*a*)pyrene (a polycyclic hydrocarbon) -induced damage are susceptible to tobacco carcinogen exposure and may be at increased risk of developing lung cancer (Li, Firozi et al. 2001). Progressive pathological changes in the respiratory epithelium are found in preneoplastic/preinvasive lesions or even the in the cytologically normal bronchial epithelium of smokers which can lead to lung cancer.

Mutations of *p53* and *p16* methylation were found in the bronchial epithelium from chronic smokers before any clinical evidence of neoplasia (Kersting, Friedl et al. 2000).

The induction of high levels of reactive oxygen species (ROS) after exposure to tobacco smoke leads to impairment of epithelial and endothelial cell function as well as inflammation (Walser, Cui et al. 2008).

The capacity of tobacco carcinogens to mutagenise the respiratory epithelium extensively has been seen in multifocal premalignant lesions that occur throughout

the bronchial tree of tobacco smokers' respiratory epithelium (Auerbach, Stout et al. 1961).

1.2.2 Lung Cancer Classification

Lung cancer, also known as bronchogenic carcinoma, is broadly categorised into two types; small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). The classification into small cell and non-small cell is based upon the microscopic appearance of the tumour cells and because they grow, spread, and are treated in different ways a distinction between them is important. Carcinomas can be described as well-differentiated, moderately differentiated, or poorly differentiated. These terms describe how aggressive a carcinoma is likely to be with well-differentiated carcinomas more likely to grow slowly and have a better prognosis, poorly differentiated the most aggressive with a worse prognosis, and moderately differentiated falls between these two, with an intermediate prognosis/aggression. In this way stage 1 cancer will be well-differentiated, stage 2 will be moderately-differentiated, and stage 3 will be poorly differentiated.

Differentiation refers to how mature (developed) the cancer cells are in a tumour, with differentiated tumour cells resembling normal cells and growing and spreading at a relatively slow rate, whereas undifferentiated, or poorly differentiated, cells lack the structure and function of normal cells and grow out of control at a much faster rate. When a tumour cell becomes poorly differentiated it is therefore more difficult to ascertain the original cell type from which it came. Poorly differentiated adenocarcinoma for example does not resemble the normal gland cells from which

they began and are usually identified by the fact they stain positive for mucin, a substance produced in the glands.

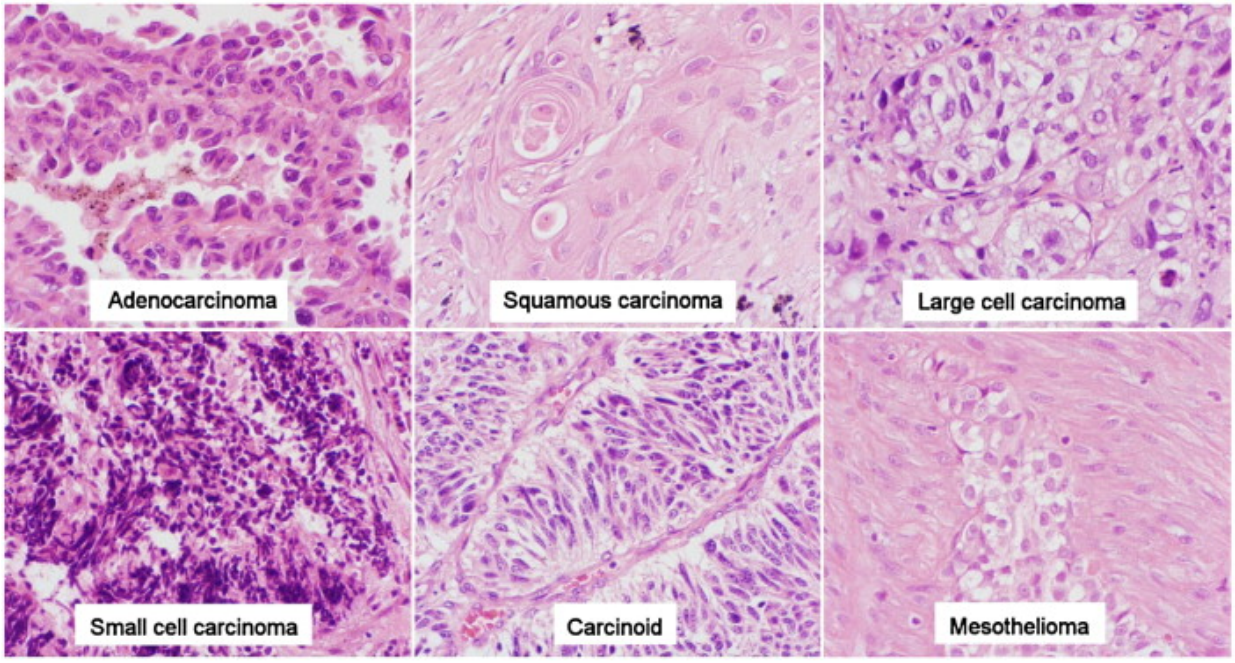


Figure 1.3: Main histological types of lung cancer; adenocarcinoma, squamous carcinoma, large cell carcinoma and small cell carcinoma, as well as carcinoid, a type of neuroendocrine tumour, and mesothelioma, a malignant carcinoma arising from mesothelial cells of the pleura (Motadi, Misso et al. 2007).

Small cell lung cancer is the most aggressive and rapidly growing lung cancer type and comprises about 20% of all lung cancers. It is strongly related to cigarette smoking with only 1% of these tumours occurring in patients who do not smoke. Because of its ability to rapidly metastasise it has a poorer prognosis than NSCLC and is often discovered after it has already spread extensively. It usually starts in the bronchi in the centre of the chest.

The lung is overwhelmingly the most common site of origin of small cell carcinoma in the body with more than 95% arising in that location, whereas extra-pulmonary

small cell carcinomas such as of bladder, prostate, esophagus, and cervix, are extremely rare (Walenkamp, Sonke et al. 2009). SCLC was once referred to as oat cell carcinoma, due to the resemblance of the cells to oats when viewed under the microscope; the flat cell shape and scanty cytoplasm giving it the oat-like appearance. A failure in the mechanism that controls the size of the cells has been attributed to the fact small cell carcinomas are smaller than normal cells and have little room for cytoplasm (Leslie). SCLC presents as a proliferation of small cells with the following morphological features: scant cytoplasm, ill-defined borders, finely granular 'salt and pepper' chromatin, absent or inconspicuous nucleoli, and a high mitotic count. Today SCLC is broken into two main types;

(1) **small cell lung carcinoma**

(2) **combined small cell lung carcinoma (c-SCLC)**

Small cell lung carcinoma is a neuroendocrine tumour (NET). As the name suggests the neuroendocrine system is made up of the nervous system and the endocrine system. The endocrine system is involved in the maintenance of homeostasis and is the systems of glands which secrete hormones directly into the bloodstream. There are four major types of lung neuroendocrine tumours: typical carcinoid (TC), atypical carcinoid (AC), small cell lung cancer, and large cell neuroendocrine carcinoma (LCNEC) (Rekhtman). As a group these tumours have a broad spectrum of clinical behaviour from lethargic in TC to rapidly fatal in SCLC. Patients with TC have a survival chance of >87% (Thomas, Tazelaar et al. 2001). TC is however defined as a low-grade malignancy because these tumours are capable of regional lymph node metastasis in 10-15% of cases and in 3-5% of cases distant metastases. AC is more aggressive with higher frequency of nodal (~50%)

and distant (~20%) metastases, and a 5-year survival rate of 60% (Rekhtman). SCLC is the most aggressive with survival measured in months and a 5-year survival rate of <5%. Two-thirds of SCLC patients present with distant metastases, typically in the brain, liver, adrenal, bone, and bone marrow (Elliott, Osterlind et al. 1987), and the amount of SCLCs presenting a solitary mass without evidence of metastases is very rare, occurring in <5%.

c-SCLC is a small cell carcinoma combined with an additional component consisting of any non-small cell histologic type, including adenocarcinoma, squamous cell carcinoma and large cell neuroendocrine carcinoma. Approximately 30% of SCLC will have this non-small cell component. Surgical resection is the mainstay for NSCLC and radiation and chemotherapy are used for SCLC so an accurate understanding of c-SCLCs is important for treatment strategies. Though at present c-SCLC is considered a subset of SCLC by the World health Organisation the biologic evidence to support this classification scheme is lacking and the validity of the current practice is yet unconfirmed (Wagner, Kitabayashi et al. 2009).

The three main types of **NSCLC** are adenocarcinoma (AD), squamous cell carcinoma (SCC), and large cell carcinoma (LCC).

Adenocarcinoma is the most common form of lung cancer, accounting for approximately 40%; a total of around 500,000 cases annually. The cells of adenocarcinoma resemble gland cells, such as the glands that secrete mucus in the lungs. It is the most common type of lung cancer in lifelong non-smokers and it has surpassed squamous cell carcinoma as the most common type in smokers. This shift

may be related to changes in exposure to tobacco products, dietary factors, environmental or occupational carcinogens, and host characteristics (Travis, Travis et al. 1995). Most adenocarcinomas arise in the outer, or peripheral, areas of the lungs, with a tendency to spread to the lymph nodes and beyond. This peripheral location has been attributed to the use of filters in cigarettes which allow smoke to be inhaled more deeply into the lungs. The four stages of lung adenocarcinoma are: (1) cancer is localised within the lung and has not spread to the lymph nodes (2) cancer has spread to lymph nodes or lining of the lungs, or is in the main bronchus (3) has spread to tissue close to the lungs, and (4) has metastasised to another part of the body. These stages can be applied to all lung cancers.

Squamous cell carcinoma (SCC) were formerly the most common type of lung cancer and today still account for 30% of NSCLC cases. They are also referred to as epidermoid carcinomas and frequently arise in the central chest area in the bronchi, the lung's major airways. SCC most often stays within the lung, spreads to lymph nodes, and grows quite large, forming a cavity. SCC is formed from reserve cells; round cells that replaced injured or damaged cells in the lining of the bronchi. SCC can spread to bones, adrenal glands, the liver, small intestine, or the brain. It is almost always caused by smoking and is a slow-growing cancer that can take years to develop from a confined tumour to an invasive one. Symptoms include wheezing, shortness of breath, chest pain, bloody sputum and coughing. SCC develop slowly, however due to their location are often found earlier than other forms of lung cancer.

I	**Adenocarcinoma**
	i. Adenocarcinoma with mixed subtypes
	1. Well-differentiated fetal adenocarcinoma
	2. Mucinous adenocarcinoma
	3. Mucinous cystadenocarcinoma
	4. Clear cell adenocarcinoma
	5. Signet ring adenocarcinoma
	ii. Acinar
	iii. Papillary
	iv. Bronchioloalveolar carcinoma
	1. Mucinous
	2. Nonmucinous
	3. Mixed mucinous and nonmucinous
	v. Solid adenocarcinoma with mucin
II	**Squamous**
	i. Papillary
	ii. Small-cell
	iii. Clear cell
	iv. Basaloid
III	**Large-cell**
	i. Large-cell neuroendocrine carcinoma
	ii. Basaloid carcinoma
	iii. Lymphoepithelioma-like carcinoma
	iv. Mixed large-cell neuroendocrine carcinoma
	v. Clear cell carcinoma with rhabdoid phenotype
IV	**Adenosquamous carcinoma**
V	**Carcinomas with pleomorphic, sarcomatous characteristics**
	i. Carcinosarcoma
	ii. Pulmonary blastoma
	iii. Carcinomas with spindle and/or giant cells
	1. Giant cell carcinoma
	2. Spindle cell carcinoma
	3. Pleomorphic carcinoma
	iv. Other
VI	**Carcinoid**
	i. Typical carcinoid
	ii. Atypical carcinoid
VII	**Carcinomas of salivary gland origin**
	i. Adenoid cystic carcinoma
	ii. Mucoepidermoid carcinoma
	iii. Others
VIII	**Unclassified**

Adapted from the World Health Organization (WHO): Histologic typing of lung tumors. In: International Classification of Tumors. Geneva, Switzerland: WHO, 1991; Travis WD, Colby TV, Corrin B, et al: World Health Organization Histological Typing of Lung and Pleural Tumours, 3rd ed. Berlin: Springer-Verlag, 1999.
IASLC = International Association for the Study of Lung Cancer

Figure 1.4: WHO and International Association for the Study of Lung cancer (IASLC) guidelines for the histologic classification of non-small cell lung cancer (www.cancernetwork.com).

Epithelia are formed of cells that line the cavities in the body, the epithelial cells bound together in sheets of tissue. The whole surface of the body is covered in epithelial tissue, which is specialised to form the covering or lining of all internal and external body surfaces. Epithelial cells are packed tightly together with almost no intercellular spaces. A thin sheet of connective tissue called basement membrane separates epithelial tissue from the underlying tissue and provides structural support and binds it to neighbouring structures. Squamous cell epithelium has the appearance of 'pavement'. The squamous cells themselves have the appearance of thin, flat plates and tend to have horizontal flattened, elliptical nucleus because of the thin flattened form of the cell. This nucleus shape usually corresponds to the cell form and helps to identify the type of epithelium. Squamous cells form the lining of cavities such as the mouth, blood vessels, heart and lungs and make up the outer layers of skin.

Large cell lung carcinomas (LCC) are the least common type of NSCLC, accounting for 10-15% of all lung cancers, and are often referred to as undifferentiated carcinoma. They get their name from the appearance of large round cells when examined under the microscope, although the tumours themselves tend to be large as well when diagnosed. They generally start in the outer regions of the lung and grow quite rapidly. Due to their presence in the outer regions they can cause fluid to develop in the space between the tissues that line the lung, a build-up that is known as pleural effusion, and invade the chest wall. A patient is generally diagnosed with LCC by 'diagnosis of exclusion', which is a process of elimination whereby the other lung cancer types have been completely ruled out. In this method the tumour cells would lack the light microscopic characteristics that would classify

a neoplasm as another type of lung cancer. For example it does not have the 'salt-and-pepper' chromatin and has a higher cytoplasmic-to-nuclear size ratio when compared to SCLC. One important subtype is large-cell neuroendocrine carcinoma (LCNEC) of the lung which displays morphologic and immunohistochemical characteristics common to neuroendocrine tumours and morphologic features of LCC. Again 'diagnosis by exclusion' is used here where LCNEC is diagnosed when other neuroendocrine tumours (typical carcinoid, atypical carcinoid, SCLC) have been ruled out. LCNEC has a very poor prognosis so there is a need for it to be accurately differentiated from other forms of NSCLC (Fernandez and Battafarano 2006).

1.2.3 Lung Diseases

The most common causes of lung diseases are smoking, infections, and genetic aberrations. Substances in the workplace can lead to lung disease such as dust from asbestos, fumes from metals, smoke from burning organic material, gases such as nitrogen oxides, vapors such as those given off by solvents, and mists or sprays from paint and pesticides. Lung disease incorporates lung cancer but it is a broad term that encompasses many afflictions; diseases affecting the airways such as asthma, chronic obstructive pulmonary disease (COPD), chronic bronchitis, emphysema, acute bronchitis, and cystic fibrosis; diseases affecting the alveoli such as pneumonia, tuberculosis, emphysema, pulmonary edema, acute respiratory distress syndrome (ARDS), and pneumoconiosis; diseases affecting the interstitium, the microscopically thin, delicate lining between the alveoli such as sarcoidosis, idiopathic pulmonary fibrosis, and autoimmune disease; diseases affecting the blood vessels such as pulmonary embolism and pulmonary hypertension; diseases

affecting the pleura, a thin layer that surrounds the lung and lines the inside of the chest wall, such as pleural effusion, pneumothorax, and mesothelioma; and diseases affecting the chest wall such as obesity hypoventilation syndrome and neuromuscular disorders.

With so many lung diseases a focus will be brought only to the ones that were used in some of the control groups during the biomarker validation phase.

Autoimmune disease is one where an inappropriate immune response occurs in the body against substances and tissues normally present in the body. The immune system mistakenly attacks and destroys healthy body tissue. Type 1 diabetes samples (although not a lung disease) were used as controls in some of our control groups. Type-1 diabetes involves the loss of the insulin-producing beta cells of the islets of Langerhans in the pancreas, leading to insulin deficiency and is a T-cell-mediated autoimmune attack (Weigmann, Franke et al.).

Pneumonia is an inflammation of the lung usually caused by an infection. The three most common causes are bacteria, viruses, and fungi. Symptoms include difficulty breathing, coughing, and a fever, and it can be very dangerous in older adults and young children because their immune systems are not as strong. It is usually caused by the inhalation of infectious air into the lungs. It can happen if the bacteria or virus is not cleared by the lung's immune system and so can happen to healthy people after a viral infection such as influenza which lowers the lung's defence mechanism. The most common cause is from bacteria called *streptococcus pneumoniae.*

Bronchitis is an inflammation of the bronchial tubes. Sufferers often have a cough which brings up mucus, the slimy substance made by the lining of the bronchial tubes. The two main types are acute bronchitis and chronic bronchitis. Acute is a short term ailment and chronic is an ongoing, serious condition. Acute bronchitis is caused by infections and lung irritants, the same viruses that cause colds and flu, such as rhinovirus, adenovirus, and influenza A and B, are the most common cause. Chronic bronchitis is a type of COPD, with inflamed bronchi producing a lot of mucus, leading to cough and difficulty getting air in and out of the lungs. Smoking is the most common cause.

Tuberculosis, or TB, is an infectious disease caused by a bacterium called *Mycobacterium tuberculosis.* It can affect the organs in the central nervous system, lymphatic system, and circulatory system, indeed almost any tissue or organ of the body, but mainly affects the lungs. Because of the way it 'consumes' from within it was once called consumption. On infection the bacteria in the lungs multiply and cause pneumonia along with chest pain, coughing up blood, and a prolonged cough. As it tries to spread to other regions of the body it is attacked by the immune system, which forms scar tissue or fibrosis around the TB bacteria. The bacteria will often break through the scar tissue and return to active state with pneumonia and damage to bones, kidneys, and the meninges that line the spinal cord and brain.

Sarcoidosis is an inflammatory disease which occurs in the lymph nodes, liver, eyes, skin, lung, and other tissues. What triggers this inflammation is not known and it cannot be caught from, or spread, to other people. It can appear suddenly and disappear or develop gradually with symptoms that can last a lifetime, though these

may come and go. Granulomas appear in the affected tissue, which are lumps/clusters of immune system cells, and if too many form they can affect how the organ works, with the resulting symptoms depending on the organ affected, though sarcoidosis often has no symptoms. If these granulomas form in the lungs it can led to wheezing, coughing, shortness of breath, or chest pain. Sarcoidosis has no known cure.

1.2.3.1 Benign lung tumours

Benign lung tumours are not cancerous as they lack the ability to metastasise. A benign tumour is a mass of tissue that serves no useful purpose and though many are harmless to human health some can compress on blood vessels, nerves, and the brain, and also for example a benign tumour of the endocrine tissue can overproduce certain hormones. They tend to grow more slowly than malignant tumours and because they stay in one part of the body and do not spread are referred to as localised. Benign tumours can of course lead to malignant tumours; colon polyps are benign tumours and most colon cancer develops from polyps. When benign tumours are surrounded by a fibrous sheath it inhibits their ability to become malignant, however many benign tumours can become malignant, teratomas in particular are known for this. In many cases benign tumours do not need any treatment, but in centralised benign lung tumours preference must be given to organ-preserving operations (Ismailov, Shishkin et al. 1993).

Some benign lung tumours can interfere with the airway and can cause lung infections. In turn inflammation from infections can lead to benign lung tumours. Hamartomas are a type of benign lung tumour and are firm marble-like tumours made up of tissue from the lung's lining as well as tissue such as fat and cartilage,

usually located in the lung periphery. Bronchial adenoma is a tumour in the trachea or bronchi that can usually block the airways although with the exception of mucous gland adenomas can spread to other areas of the body. There are also benign tumours made up of connective tissue or fatty tissue; rare neoplasms such as chondromas, fibroma, or lipomas. Despite improvements in diagnosing pulmonary nodules, a mass in the lung smaller than 3cm, there are still a significant number of cases where surgical resection is needed to differentiate cancer from benign lesions (Ohtsuka, Nomori et al. 2003).

1.2.4 Screening

The most significant factor for survival in lung cancer is the stage of disease at diagnosis. Lung cancer tends not to cause symptoms in its early stages, whilst in patients with more advanced disease the symptoms are non-specific; therefore it is classically picked up late in its development. It is estimated that a lung tumour of 1mm in size is halfway through its growth cycle, and more than three-quarters of the way through its growth cycle when it reaches 3-4cms in size (the average size at time of diagnosis), and plenty of time will have elapsed for metastatic spread before the cancer is diagnosed by conventional methods (Read, Janes et al. 2006). Histologic diagnoses of early-stage lung cancers are complicated by a lack of agreement among histopathologists on what constitutes a preneoplastic lesion in the lung (Price 2004). The first signs of invasive cancer could take 10 to 20 years to develop.

The most common symptom alerting patients and doctors to possible lung cancer is bloody sputum and a cough, however a cough is often ignored for a long time

before referral to a specialist (Buccheri and Ferrigno 2004). The link between abnormal sputum cytology and lung cancer development has been confirmed in various studies (Risse, Vooijs et al. 1988, Prindiville, Byers et al. 2003). The diagnostic yield of sputum cytology varies in relation to tumour location and identifies central tumours confidently but is of little value for identifying peripheral cancers, though recent developments in DNA analysis and nuclear image analysis show potential to improve or refine diagnosis beyond that achieved with conventional sputum cytology examination (Thunnissen 2003). Improvements have not been seen in disease-specific survival from screening tests for high-risk individuals using chest X-rays or sputum cytology (Marcus, Bergstralh et al. 2006).

Bronchoscopy is a very useful tool in the diagnosis of lung cancer and the endoscopic treatment of inoperable lung cancer, and it can improve the diagnosing and staging in patients (Herth, Eberhardt et al. 2006). Some epithelial changes, such as high-grade dysplasia and carcinoma in situ in the central airways, are early-stage squamous cell carcinomas (multidetector CT technology is unable to detect pre-invasive lesions and less suited to detect early-stage lung cancer in the central airways) and the most widely used and investigated technique for the detection of this kind of premalignant endobronchial lesion is the light-induced fluorescence endoscopy (LIFE) (Kennedy, McWilliams et al. 2007). Bronchoscopy with the help of auto-fluorescence bronchoscopy and narrow-band imaging can detect preinvasive lesions and early-stage lung cancer in the central airways that cannot be detected by CT technology (McWilliams, Lam et al. 2009).

X-ray computed tomography (CT) is used to detect early stage lung cancers (Swensen, Jett et al. 2005). Screening with low-dose spiral CT in patients at high risk for lung cancer may allow detection of many early-stage lung cancers (Swensen, Jett et al. 2003). Low radiation dose CT (LDCT) has been shown to have a much higher sensitivity for small pulmonary nodules, which are believed to be the most common presentation of early lung cancer though it should be noted small pulmonary nodules are common and most are not malignant (Diederich, Wormanns et al. 2003). LDCT allows a low-resolution image of the thorax to be obtained in a single breath-holding, with low radiation exposure (Jain and Arroliga 2001) and when compared with chest X-ray is able to detect approximately three to five times as many lung cancers and detect tumours with a smaller average size (Sone, Takashima et al. 1998). Although LDCT can increase the diagnosis of early stage lung cancer it does not show strong evidence of a mortality benefit (Kakinuma, Ohmatsu et al. 1999).

Low-dose computed tomography is a promising tool, as seen from preliminary screening trials results which employ this technique, but for the future a technique called positron emission tomography (PET) with the glucose analogue 18-fluorodeoxyglucose (FDG) which identifies malignant tumours on the basis of their increased metabolic rate is of particular interest, as is the analysis of biomolecule markers which can be detected in serum, sputum and exhaled air (Rossi, Maione et al. 2005).

Another approach is the detection and characterisation of circulating tumour cells (CTCs). The potential of CTC detection as a diagnostic parameter to distinguish malignant from benign lung disease has been evaluated and it was found that the

CTC count was higher in patients with lung cancer compared to non-malignant patients, however the sensitivity for detecting CTCs was low (Tanaka, Yoneda et al. 2009).

Routine histopathology is the current standard of lung cancer classification, but this method is unreliable. In one study addressing the reproducibility of histopathologic classification of resected lung cancers, the central expert pathologist disagreed with the regional pathologists in one third of the cases (Stang, Pohlabeln et al. 2006).

1.3 Mechanisms of Invasion and Migration

Migration and invasion mechanisms allow neoplastic cells to enter lymphatic and blood vessels for dissemination into the circulation, and then undergo metastatic growth in distant organs (Chambers, Groom et al. 2002). This spread of cells from the primary neoplasm to distant organs, and the subsequent rapid growth, is the main cause of death for cancer patients. The biological processes involved in moving tumour cells from their primary site to a distant location can potentially be targeted by therapeutic agents, however the molecular events that allow this to occur are highly complex and not well understood, and there are no effective treatments that target invading tumour cells. The mechanisms used in normal, neoplastic cells during physiological processes such as embryonic morphogenesis, wound healing and immune-cell trafficking, are quite similar to those used by tumour cells to migrate (Friedl and Brocker 2000).

Cancer cells disseminate from the primary tumour either as individual cells, using amoeboid, or mesenchymal-type movement, or as cell sheets, strands and clusters

using collective migration, this migration typically regulated by integrins, matrix-degrading enzymes, cell-cell adhesion molecules and cell-cell communication. Cancer therapeutics targeting the migration mechanism have not yet been effective in clinical trials, maybe due to the fact that the cancer cell's migration mechanisms can be reprogrammed, allowing it to maintain its invasive properties via morphological and functional de-differentiation (Friedl and Wolf 2003). Normal cells are anchorage dependent and will undergo apoptosis if detached from cells of their own kind; however in tumour cells this regulatory process is disrupted through genetic changes, allowing the cancer cell to survive.

The metastatic cascade involves the following steps; cancer cell from primary tumour locally invades the surrounding tissue (tumour cells and surrounding stromal cells interact and influence one another (Hanahan and Weinberg 2000)), enters the microvasculature of the lymph and blood systems (intravasation), survives and translocates largely through the bloodstream to microvessels of distant tissues, exits the bloodstream (extravasation), survives in the microenvironment of distant tissues, and finally adapts to the foreign microenvironment of these tissues in ways that facilitate cell proliferation and the formation of a macroscopic secondary tumour (Fidler 2003).

Nutrients for the expanding tumour mass must initially be supplied by simple diffusion, however extensive vascularisation must occur if the tumour mass is to exceed 1-2mm in diameter, with the synthesis and secretion of angiogenic factors establishing a capillary network from the surrounding host tissue (Folkman 1986).

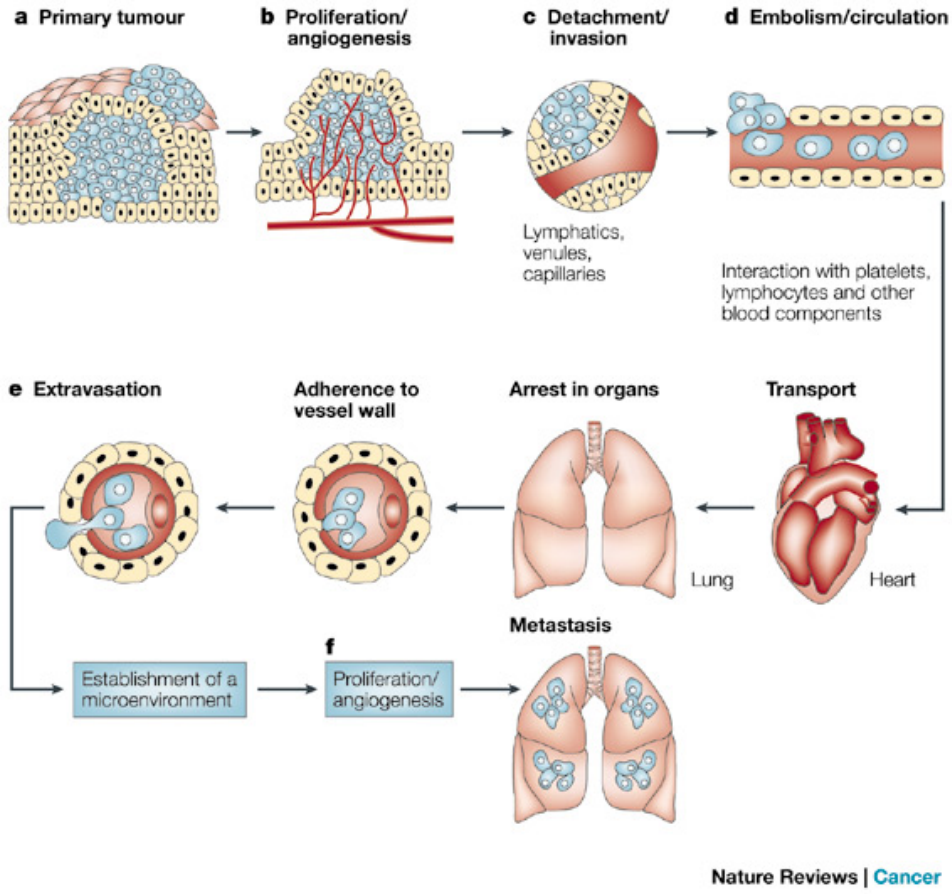


Figure 1.5: The main steps in the formation of a metastasis (Fidler 2003).

Local invasion of the host stroma by some tumour cells occurs by several parallel mechanisms such as this three-step process; (1) Tumour cell attachment via cell surface receptors which specifically bind to components of the matrix such as laminin (for basement membrane) and fibronectin (for the stroma), (2) The anchored tumour cell next secretes hydrolytic enzymes (or induces host cells to secrete enzymes) which can locally degrade the matrix, including degradation of the attachment componenets, (3) Tumour cell locomotion into the region of the matrix modified by proteolysis, with continued invasion of the matrix taking place by cyclic repetition of these three steps (Liotta 1986).

Tumour cells are generally heterogenous and contain numerous subpopulations of cells that have different biological characteristics including metastatic potential. Most primary tumour cells have a low metastatic potential but rare cells (estimated at less than one in ten million) within large primary tumour acquire metastatic capacity through somatic mutation (Poste and Fidler 1980). Somatic mutations are genetic alterations acquired by a cell, frequently caused by environmental factors, which can be passed to the progeny of the mutated cells in the course of cell division and differ from germ line mutations which are inherited genetic alterations that occur in the germ cells. The metastatic phenotype includes the ability to migrate from the primary tumour, survive in blood or lymphatic circulation, invade distant tissues and establish distant metastatic nodules. However it has been suggested that the metastatic potential of human tumours is encoded in the bulk of the primary tumour, challenging the notion that metastases arise from rare cells within a primary tumour that have the ability to metastasise (Ramaswamy, Ross et al. 2003).

1.4 Molecular Targeted Therapies for Lung Cancer

Drugs that specifically target certain molecular pathways leading to lung cancer phenotype have been clinically tested and there are others which are in development. Gene expression arrays are a genomic technique employed to discover changes in the DNA expression that occur in neoplastic transformation and have been used for lung cancer, with the goal not only of finding new markers for therapy but also to customise therapy based on individual tumour genetic composition. Genomic technology can also help with a better understanding of the

pathogenesis of lung cancer (molecular changes, pathways), with the ultimate goal of discovering genetic transformations that have occurred in a cancer cell that can be manipulated to kill the neoplastic cell (Singhal, Miller et al. 2008).

Small molecules that specifically inhibit the tyrosine kinase (TK) activity of the epidermal growth factor receptor (EGFR), such as gefitinib and erlotinib, were the first drugs to become clinically available in the treatment of NSCLC (Mitsudomi). A somatic mutation of the EGFR gene in Asian women with adenocarcinoma showed a tendency of the patients to be sensitive to EGFR-tyrosine kinase inhibitors (TKIs), and those with the mutation showed longer survival than those with wild-type EGFR (Han, Kim et al. 2005). Evidence that EGFR mutations are sensitive and specific predictors of response to single-agent epidermal growth factor receptor TKIs in advanced NSCLC has been shown elsewhere (Dahabreh, Linardou et al.).

Bevacizumab is a humanized antibody that targets vascular endothelial cell growth factor (VEGF), which is thought to play a pivotal role in tumour angiogenesis, though in a Phase II trial there was an increased risk of hemorrhage in patients with squamous cell carcinomas, especially those with a central location, meaning this set of patients was not eligible for subsequent clinical trials (Johnson, Fehrenbacher et al. 2004).

DNA synthesized from a mature messenger RNA (mRNA) template in a reaction catalyzed by the enzymes reverse transcriptase and DNA polymerase is called complementary DNA (cDNA). mRNA is a molecule which serves as a template for protein synthesis. On a cDNA array, messenger RNA clones are spotted on arrays

in a matrix where one analyzes the relative expression level of a gene by determining the amount of mRNA that is present. It allows the analysis of up to tens of thousands of genes in a single experiment though the analysis of this data can be quite complex and at each step there is the possibility of inheriting variability into the detection of the tissues genetic profile (Dobbin, Beer et al. 2005). Even though gene expression profiling is a high-throughput method for finding targets it does not provide information about whether the changes in gene expression are responsible for biological behaviour.

The principal molecular changes in lung cancer are seen in tumour suppressor genes, proto-oncogenes, growth factors, telomerase activity, and methylation status of promoters. Well-known agents include angiogenesis-stimulating factors (such as vascular endothelial growth factor), as well as factors related to tumour cell proliferation and apoptosis (epidermal growth factor receptor, p53, K-ras, retinoblastoma and BCL-2), and though several of these genetic factors have already been investigated, no single parameter has yet presented sufficient selectivity regarding prognostic value or therapeutic efficacy (Duarte and Paschoal 2006). The inactivation of p53 tumour suppressor gene is a common genetic event in lung cancer (Andjelkovic, Bankovic et al. 2011).

1.4.1 MicroRNAs in Lung Cancer

MicroRNAs (miRNA) are tiny non-coding RNA molecules which play important roles in the epigenetic control of cellular processes by preventing the translation of proteins from mRNAs. A single miRNA can target different mRNAs, and an mRNA can be targeted by multiple miRNAs, and the target specificity is determined by sequence complementarities between the miRNA and mRNA

molecules. These complex interplays underlie many molecular pathways in cells and changes in miRNA expression have been associated with a wide variety of disease states including the diagnosis and prognosis of lung cancer.

The effect of alteration in the expression of a single miRNA in a cell can depend on the concentration of the mRNAs that are targeted by it. miRNAs have been shown to control a wide range of biological functions such as cellular proliferation, differentiation and apoptosis and it has been suggested that miRNA might function as tumour suppressors and oncogenes (Esquela-Kerscher and Slack 2006). These alterations can be caused by various mechanisms, including deletions, amplifications or mutations involving miRNA loci, epigenetic silencing, or the dysregulation of transcription factors that target specific miRNAs.

As sequences of small RNAs are being explored in more types of cells using improved techniques, novel miRNAs are continually being discovered (Creighton, Benham et al. 2010). miRNAs have been shown to be remarkably well preserved in formalin-fixed, paraffin-embedded (FFPE) tissue, eliminating the need for fresh frozen tissue for miRNA expression studies (Szafranska, Davison et al. 2008).

MicroRNAs can be identified in serum or plasma, which has the potential to enhance the utility of miRNAs as non-invasive biomarkers of lung cancer. It is believed that tumour cells secrete cell membrane-enveloped microvesicles, known as exosomes, into the extracellular space allowing them to eventually be detected in the blood (Rabinowits, Gercel-Taylor et al. 2009). These exosomes are 30–100nm in size and contain proteins, cytosol, mRNA, and miRNAs. Their lipid membrane protects them from the circulating ribonucleases present in blood, and microRNAs have been shown to be stable in serum stored at 4 °C for up to 4 days (Taylor and

Gercel-Taylor 2008). Exosomes can bear cell-surface proteins characteristic of the cell of origin, so it may be possible to isolate exosomes in the blood that have originated from specific tissue (Mallick, Patnaik et al.).

One study has shown that expression of hsa-miR-205 was restricted to the squamous phenotype, and that the measurement of hsa-miR-205 expression was a very dependable method of distinguishing lung SCC from adenocarcinomas (Lebanony, Benjamin et al. 2009). This provides a rationale for the classification of NSCLC based on miRNA expression profiling.

Two members of the double-stranded RNA-specific endonuclease family, *Dicer* and *Drosha*, convert precursor forms of microRNA into their mature forms. *Dicer*, an RNAse III enzyme, expression levels are reduced in some lung cancers with a significant prognostic impact on the survival of surgically treated cases. In an examination of 67 non-small-cell lung cancer samples it was found that reduced expression of *Dicer* correlates with shortened post-operative survival, which indicates that Dicer might be able to prevent the transformation of lung tissue (Karube, Tanaka et al. 2005).

1.5 Proteomics in Lung Cancer

Early detection of lung cancer is challenging, due in part to a lack of adequate tumour markers. Discovery of biomarkers, and in particular serum biomarkers, that could indicate the presence of lung cancer at an early stage would be very beneficial and could save lives or lengthen the patient's survival time from months to years.

Even more beneficial would be a panel of biomarkers that diagnosed the particular type of lung cancer, allowing for more specific treatment.

Early detection is not the only area for which biomarkers would be advantageous; they could improve diagnosis, predict response after treatment, and monitor recurrence after treatment (Gail, Muenz et al. 1988). The clinical usefulness of serum biomarkers at present is very limited (McCarthy and Swain 2001). Individual studies are often retrospective and lacking statistically, hence, while discovery of clinically useful biomarkers is urgently required, larger prospective studies with clinically relevant modelling and hypothesis are vital to address the usefulness of novel potential tumour biomarkers in the management of lung cancer patients and also the treatment strategies must improve to accommodate any improvement in biomarker discovery (Bharti, Ma et al. 2007).

Proteomic-based approaches have greatly improved the data produced relating to lung cancer biomarkers but as yet this vast knowledge has not effectively been applied to the production of useful clinical assays that can be used on a routine basis in the hospital. There is a need to examine the data already available closely for trends that may lead to a better understanding of the disease and the pathways involved and perhaps then to useful biomarkers yet to be identified by other methods.

The general strategy is to analyse serum (indeed any body fluid) or tissue for proteins which show a difference in abundance levels between cancer samples and samples from healthy individuals or between patients who respond to therapy and those who don't. There have been numerous proteins identified with the potential to be useful biomarkers, many indicating high sensitivity and specificity. However,

validation by clinical trials in large cohorts of patients is necessary before cancer-related phenotypes can be translated into the clinic as reliable biomarkers (Cho 2007) but as this is a very highly regulated, exorbitantly expensive process only a few biomarkers are ever brought forward to the validation phase and for this there needs to be a high level of statistical confidence in the biomarker/biomarkers of choice.

1.5.1 Present state of lung cancer biomarkers

Some possible biomarkers for lung cancer often mentioned in the literature include cytokeratin 19 fragment (CYFRA 21-1), tissue polypeptide antigen (TPA) and squamous cell carcinoma antigen (SCCA) for NSCLC, particularly for SCC, carcinoembryonic antigen (CEA) and cancer antigen 125 (CA-125) in AD or LCLC, and progastrin-releasing peptide (ProGRP) and neuron-specific enolase (NSE) in SCLC.

In one study CYFRA 21-1 was investigated in 100 patients with lung cancer and detected with an immunoassay in the sera of 60% of patients, showing high sensitivity (86.4%) for squamous cell carcinoma and relatively low sensitivity for both adenocarcinoma (52.6%) and small cell carcinoma (50%) (Kinoshita, Watanabe et al. 1998). In another study CYFR21-1, at a cut-off level of 3.6ng/ml, was found to have sensitivity and specificity, respectively, of 56% and 89% for NSCLC, 63% and 91% for SCC, and 46% and 89% for SCLC (Pujol, Grenier et al. 1993). Also it was shown to have, at a cut-off of 3.3ng/ml, sensitivity and specificity, respectively, of 59% and 94.4% for NSCLC, 68% and 94.4% for SCC, and 19% and 94.4% for SCLC (Wieskopf, Demangeat et al. 1995). Combining

telomerase activity with levels of CYFRA 21-1 can increase the sensitivity and overall accuracy of differential diagnosis of benign and malignant pleural effusion caused by lung cancer (Li, Fu et al.). A decline in CYFRA 21-1, and CEA, during chemotherapy can predict response and survival in patients with advanced NSCLC (Ardizzoni, Cafferata et al. 2006). CYFRA 21-1 has also been seen to predict survival in patients receiving gefitinib as third-line therapy (Barlesi, Tchouhadjian et al. 2005).

CYFRA 21-1 has been shown to be a better marker than SCCA for early prediction of squamous cell carcinoma recurrence in the lung (Sun, Hsieh et al. 2000). The use of SCCA as a biomarker for NSCLC has been brought into question due to its low sensitivity and also by a study in which SCCA was shown to have no additional value when used in combination with CYFRA 21-1 (Kagohashi, Satoh et al. 2008).

CA-125 is more consistently elevated in epithelial ovarian cancer and not lung cancer (Bast, Xu et al. 1998) however one study showed that the positive rates of CA125 is higher than that of CEA in NSCLC patients, especially in large-cell lung cancers and advanced adenocarcinoma, and concluded that CA125 is more useful than CEA in diagnosis of NSCLC (Wang, Zhu et al. 2008).

A study on CEA and CA-125 in bronchoalveolar lavage fluid (BAF) in patients with NSCLC used cut-off values of 3ng/ml and 95 IU/ml respectively, and showed sensitivity and specificity of 100%, 84% and 92%, 80%, respectively (Dabrowska, Grubek-Jaworska et al. 2004).

The levels of CEA, NSE, CYFRA-21-1, CA-125, and CA 19-9 were noted as being significantly higher in patients with pathologically confirmed lung cancer than those patients with benign pulmonary disease or control subjects (Li, Asmitananda

et al. 2012). CA 15-3, an unspecific marker, showed it may be useful as a marker for treatment monitoring (Nutini, Cappelli et al. 1990).

ProGRP and NSE were shown to have complementary roles in diagnosis and prognosis of SCLC, indicating that ProGRP is more sensitive than NSE for diagnosis, while NSE is superior to ProGRP as a prognostic factor, concluding that both ProGRP and NSE are useful tumour markers (Shibayama, Ueoka et al. 2001). Detection of ProGRP, CYFRA21-1, NSE and CEA in pleural effusion could be of great clinical value in differential diagnosis and histological typing of malignant pleural effusion, with ProGRP being the optimal marker for malignant pleural effusion caused by SCLC (Liu, Yu et al. 2006).

One study showed TPS, CYFRA21-1 and soluble tumour necrosis factor receptor (STNFR) can be used as very useful and sensitive tumour markers in the diagnosis of lung cancer, with CYFRA 21-1 considered the most useful tumour marker for clinical application (Liao, Li et al. 2005). TPS, along with CYFRA 21-1, could be significant prognostic factors and effective monitors of therapy for NSCLC (Liao, Li et al. 2005).

Other potential lung cancer markers mentioned in the literature include plasma kallikrein, serum amyloid A (SAA), haptoglobin ß chain (Hp ß), and complement component 9. High levels of SAA in the sera of lung cancer patients has been detected by mass spectrometry, verified by western blot and then quantified by ELISA, while the same group demonstrated that the secretion of SAA1 and SAA2 (isoforms of SAA) stimulates infiltrating macrophages to induce MMP-9, which

may drive metastasis of cancer cells (Sung, Ahn et al. 2011). High levels of haptoglobin in serum have been reported in various cancers such as breast, ovarian, pancreatic, and bladder, and the Hp ß chain was suggested as a diagnostic marker for lung cancer for both NSCLC and SCLC types (Kang, Sung et al. 2011).

Dihydrodiol dehydrogenase (DDH) over-expression in adenocarcinoma cells has been shown to exhibit a much higher resistance to the anticancer drugs doxorubicin and cisplatin and to irradiation than cells with lower DDH expression, with the isoforms DDH1 and DDH2 mainly responsible for these effects (Hung, Chow et al. 2006). Over-expression of DDH has been previously shown to be a possible prognostic marker of NSCLC (Hsu, Ho et al. 2001) and in serum its level is higher in NSCLC patients than non-malignant lung tumours and healthy controls (Huang, Chen et al. 2006).

Pentraxin-3 has recently been highlighted as a new biomarker for lung cancer with diagnostic sensitivity and specificity similar to other clinically-used lung cancer biomarkers, though it's own clinical utility has yet to be established (Diamandis, Goodglick et al.).

Cigarette smoke, the main cause of lung cancer, induces an accumulation of reactive oxygen species which have multiple effects on cell defence, cell proliferation, and cell death, therefore compounds involved in the regulation of redox balance may play a fundamental role in both carcinogenesis and tumour progression. This is why Peroxiredoxins (Prxs), which represent a protein family with the capability of breaking down hydrogen peroxide, can participate in cellular antioxidant defence and regulate cell proliferation, and also act as indicators of lung

cancer (Lehtonen, Svensk et al. 2004). Peroxiredoxin 1 for example may serve as a new prognostic biomarker and therapeutic target in NSCLC (Kim, Bogner et al. 2008).

There are many published papers in the literature with lists showing abundance levels of proteins in lung cancer compared to normal, healthy controls, but with no clinical validation. These lists, especially when generated from non-serum based work such as conditioned media, are relatively obsolete without further validation because the expression of these proteins, even ones that have been secreted in cell culture for example, does not necessarily relate to their expression in serum. With the added difficulties faced when searching in serum for low abundant markers, the majority of proteins in these lists would not be viable markers. There is a need now for any study involving lung cancer biomarkers to include validation that can be considered clinically applicable. For example, a recent study on the conditioned media of four lung cancer cell lines included validation in serum, of patients with and without lung cancer, using enzyme linked immunosorbent assays (ELISAs) for five of their candidate biomarkers, which included pentraxin-3 and soluble tumour necrosis factor receptor (Planque, Kulasingam et al. 2009).

1.5.1.1 Preanalytical Phase

Standardisation and optimisation of the preanalytical conditions for clinical specimens is needed for the development of clinically applicable tests. Lack of these procedures, as well as specimen acquisition, handling, and storage account for more than 90% of the errors within the entire diagnostic process (Lippi, Salvagno et

al. 2006). To improve chances of generating validated biomarkers, there needs to be greater focus on collection, transport, preparation, and processing of the biological sample (Conrads, Hood et al. 2006). Improved quality management in areas such as establishment of large biorepositories (biobanks) is needed. Also, depending on the phase of biomarker development and validation, the sample needed, and in turn sample processing, could be quite different. For example, high-throughput proteomic profiling for clinical proteomics has considerably different requirements than the subsequent protein identification and in-depth characterisation of single protein samples (low-throughput) in the preclinical and discovery phase, and differences can arise due to the site of sample collection, the process of blood/biofluid collection, the material and liquid content of the sample container, the time until sample processing and the temperature (Apweiler, Aslanidis et al. 2009).

Hemolysis, the release of haemoglobin and other intracellular components from erythrocytes to the surrounding plasma following damage or disruption to the cell membrane, can influence the accuracy and reliability of laboratory testing. At the preanalytical stage there are many parameters related to patient preparation and blood collection which can effect this, such as difficulty in locating easy venous access, small or fragile veins, application of a negative pressure to the blood in the syringe, excessive shaking or mixing of the blood after collection, exposure to excessively hot or cold temperatures, and centrifugation at too high speed for a prolonged period of time (Lippi, Salvagno et al. 2006).

The time of sample collection can have an effect too, as can be seen in particular with urine collection where consideration must be given to the time that urine is stored in the bladder because due to bacterial contamination there is greater protein degradation in first void morning urine compared to urine collected at other times (Schaub, Wilkins et al. 2004).

The protein content can also differ depending on whether it is serum or plasma being examined. When serum and plasma are separated from the blood, plasma still retains the fibrinogen that helps in clotting while serum is that part of the blood after this fibrinogen (clotting factor) is removed. The clotting cascade causes a reduction in proteins so generally there are more peptides in plasma than serum, however the protein composition in plasma can be affected by the anticoagulant used (Tammen, Schulte et al. 2005).

Storing at -20°C or at -80°C can affect the protein content, although little difference was seen between samples frozen for 1-3 months (Hsieh, Chen et al. 2006), however this was probably due to the short storage time and greater differences would be seen in samples stored for longer periods. Storage conditions can also depend on the protein to be examined as one study on the HbA 1C protein showed that temperatures of 4°C and -80°C were best for short and long-term storage respectively, whereas temperatures of 20°C (room temperature) and -20°C were not recommended, even indicating that samples showed earlier degradation at -20°C than at 4°C (Szymezak, Lavalard et al. 2009).

The importance of standardisation of the preanalytical process can not be overemphasised. The high amount of variables to be taken into account for a clinically applicable assay means there is plenty of room for error and to lessen the

overall effect of these errors, standard protocols are needed which can be followed in all laboratories performing the same tests. Even in a well-used laboratory test such as the BCA-assay for measuring protein content, differences can be seen in the results depending on the reagents in the lysis buffer used to prepare your protein sample, which can interfere with the chemicals in the kit (Krieg, Dong et al. 2005).

1.5.1.2 Serum and Plasma

Blood is the most common body fluid for biomarker investigation, with both plasma and serum used. Plasma is prepared by centrifuging whole blood to remove red and white blood cells, with anti-coagulants to prevent clotting used; such as EDTA, heparin, and citrate, with EDTA the anticoagulant of choice if delayed blood processing is anticipated (Lam, Rainer et al. 2004). Serum is obtained when the blood is allowed to clot, the clot is removed by centrifugation, and the light yellow supernatant is removed. Serum is basically plasma with the clotting factors removed e.g. fibrin. Plasma and serum are a light yellow colour though plasma will tend to be cloudier due to the additional proteins.

It should be noted that a good serum biomarker does not necessarily make a good plasma biomarker (and vice versa) and differences are seen in their protein and metabolite profiles. One study found reproducibility of metabolite results significantly better for plasma than for serum though metabolite concentrations were generally higher in serum than in plasma (Yu, Kastenmuller et al. 2011). Another study comparing protein profiles of serum and plasma using an antibody suspension bead array approach found that while technical variability was equal, plasma offered a greater biological variability which gave rise to more potential discoveries than serum (Schwenk, Igel et al. 2010). It has been seen that an increase

in concentrations of some proteins in serum compared to plasma could be due to the stimulation of platelets by the coagulation cascade and the release of platelet factors, such as the chemokine NAP-2 which may stimulate leukocytes leading to the production of cytokines, growth factors, and other chemokines (Ayache, Panelli et al. 2006).

1.5.2 Assays for Detection of Biomarkers in Clinical Samples

1.5.2.1 Enzyme-linked Immunsorbent Assay (ELISA)

An enzyme-linked immunosorbent assay (ELISA) is used to detect the presence of an antigen or antibody in biological fluids such as serum, plasma, saliva, or urine, and also in cell culture supernatant and tissue samples, often performed on a 96-well microtiter plate. The enzyme is conjugated to an antibody which will react with a colourless substrate to generate a coloured reaction product, which is measured spectrophotometrically. The most commonly used enzymes are horseradish peroxidase, alkaline phosphatase, and p-nitrophenyl phosphatase ß galactosidase. When mixed with a suitable substrate (for example 3,3',5,5'-tetramethylbenzidine (TMB) is a colorimetric substrate for horseradish peroxidase enzyme) each of these enzymes generates a coloured reaction product, their absorbance read at a particular wavelength on a plate-reader, allowing qualitative detection or quantitative measurement of either antigen or antibody.

An ELISA begins with the generation of antibodies to a particular protein. Antibodies (also known as immunoglobulins) are proteins synthesised by an animal in response to the presence of a foreign substance (antigen) and normally function to protect the animal from infection. They are glycoproteins secreted by specialised

B lymphocytes known as plasma cells and are composed of four polypeptides; two identical copies of both a heavy (~55 kD) and light (~25 kD) chain are held together by disulfide and noncovalent bonds, and the resulting molecule is often represented by a schematic Y-shaped molecule of ~150 kD.

Antibodies occur in two forms: a soluble form secreted into the blood and tissue fluids, and a membrane-bound form attached to the surface of a B cell that is called the B cell receptor (BCR). The BCR allows a B cell to detect when a specific antigen is present in the body and triggers B cell activation. When a severe infection occurs pathogens spread through the bloodstream and B cells in the blood capture the pathogens, via their specific antigen receptors (surface immunoglobulins), then present the specific antigen to T cells in the spleen, thus increasing the degree of T-cell immune responses to systemic infection (Nagafuchi). T cells are a type of white blood cells called lymphocytes (as are B cells), the "T" standing for thymus, the organ in which their final stage of development occurs. T helper cells help with the maturation of B cells into antibody-producing plasma cells and memory B cells (which help with the quick response if the antigen invades again and are essential for the success of vaccinations) (Esser, Marchese et al. 2003).

Antibodies have specific and high affinity for the antigens that elicited their synthesis. The antibody recognises a specific group or cluster of amino acids on a large molecule called an antigen determinant, or epitope. Small foreign molecules, such as synthetic peptides, can also elicit antibodies, provided that the small molecule contains a recognised epitope and is attached to a macromolecular carrier; the small foreign molecule is called a hapten. Animals have a large repertoire of

antibody-producing cells, each producing an antibody of a single specificity. An antigen works by stimulating the proliferation of a small number of cells that were already forming an antibody capable of recognising the antigen. Monoclonal antibodies will detect only one epitope on an antigen, whereas polyclonal antibodies recognise multiple epitopes on any one antigen and are a heterogeneous mixture of antibodies of different affinities. Polyclonal antibodies can be produced in a short timescale and are relatively inexpensive but are multi-specific and there can often be batch-to-batch variability, whereas monoclonal antibodies are expensive with a long time scale for production but are mono-specific and once a hybridoma is made it is a constant and renewable source of identical batches (Lipman, Jackson et al. 2005). A hybridoma is an immortal cell line derived from a type of cancer, multiple myeloma, a malignant disorder of antibody-producing cells, resulting in a single transformed plasma cell dividing uncontrollably, generating a large number of cells of a single kind (Ling, Deng et al. 2003).

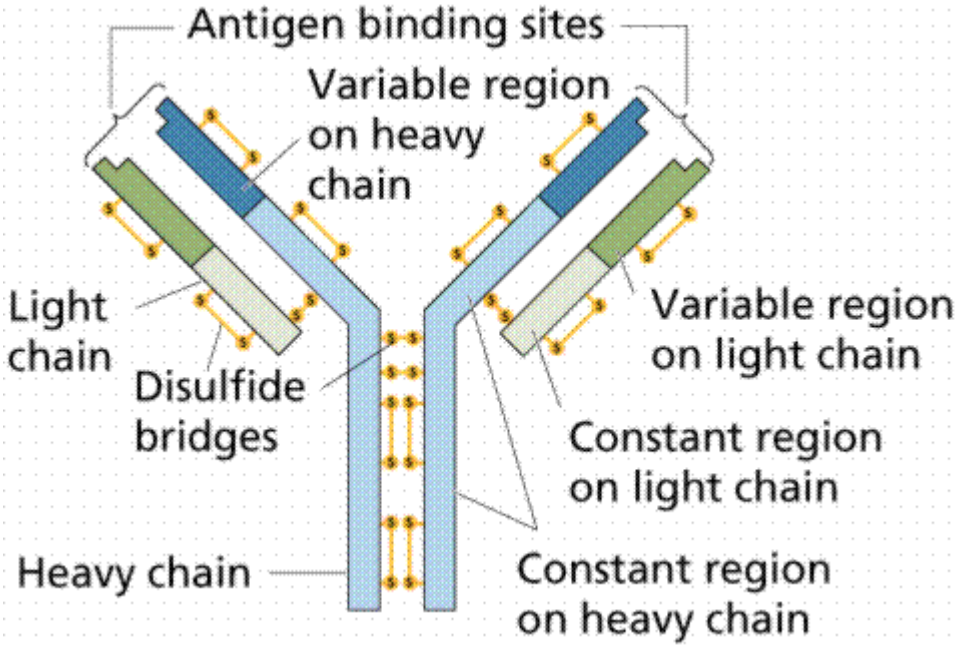


Figure 1.6: Structure of an antibody (www.cartage.org.lb)

1.5.2.2 Indirect, sandwich, and competitive ELISA

The direct method uses a labelled primary antibody that reacts directly with the antigen with detection being performed with an antigen that is immobilised on the assay plate. This method is not widely used in ELISA but is quite common for immunohistochemical staining of tissues and cells. The indirect and the "sandwich" methods are commonly used in ELISAs. The indirect method uses a known primary antibody to attach to the antigen of interest, which is attached to the well of your plate. A conjugated secondary antibody is then added to bind to your primary antibody and a substrate added to elicit a readable colour change. A sandwich ELISA is similar except that a capture antibody is used to pull out the antigen of interest before the primary and conjugated secondary are added.

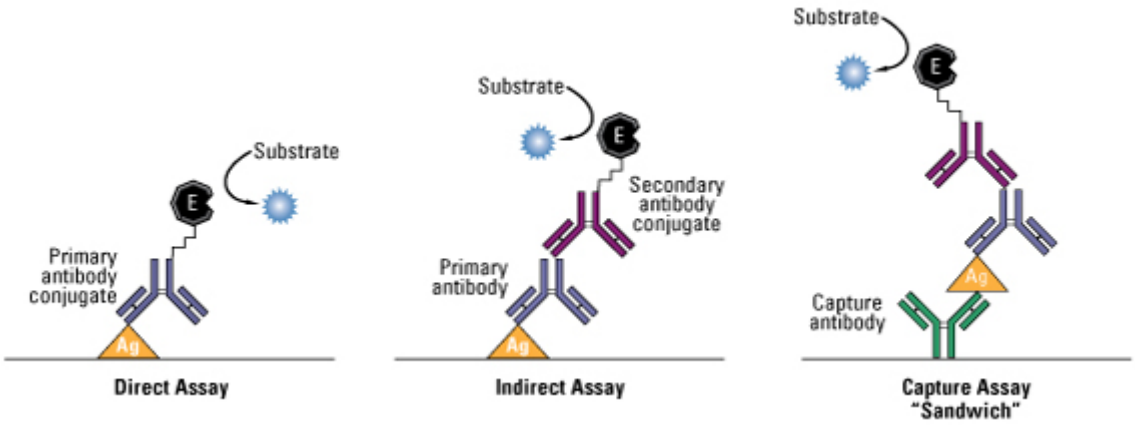


Figure 1.7: Direct, Indirect, and sandwich methods for detection of antigen of interest (www.piercenet.com)

A **sandwich ELISA** is more sensitive than the indirect method. The antigen immobilisation in the indirect method is not specific and small concentrations of analyte in serum for example must compete with other serum proteins when binding to the well surface. The sandwich ELISA provides a solution to this problem with

the use of a capture antibody specific for the antigen of interest so that when the primary antibody is added in the next step there is only the antigen of interest present to bind to. A sandwich ELISA does not need a conjugated secondary if the primary antibody is conjugated to an enzyme; however the use of a secondary-antibody conjugate avoids the expensive process of creating enzyme-linked antibodies for every antigen one might want to detect. Without the 'capture' antibody any proteins in the sample may competitively adsorb to the plate surface, lowering the quantity of antigen immobilised. Use of the purified specific antibody to attach the antigen to the well eliminates a need to purify the antigen from complicated mixtures before the measurement, simplifying the assay, and increasing the specificity and sensitivity of the assay. There is also a double-antibody sandwich ELISA technique which involves, for example, adding the samples to be tested (e.g. serum), together with a biotin antibody and streptavidin-HRP, into the ELISA well coated with capture antibody.

A **competitive ELISA** involves the initial incubation of unlabelled primary antibody with the sample of choice (e.g. serum). The samples, with the bound antibody/antigen complexes, are then added to an antigen coated well. The plate is washed so that unbound antibody is removed (the antibody/antigen complexes from initial incubation will be removed). The more antigen in the serum sample the less antibody will be available to bind this antigen, hence the 'competition'. Next a conjugated secondary is added followed by the substrate. The signal produced is inversely proportional to the amount of antigen in the sample, i.e. the higher the sample antigen concentration the weaker the eventual signal. A competitive ELISA allows the use of crude or impure samples and still selectively binds any antigen

that may be present. Some competitive ELISA kits include the enzyme-linked antigen rather than the enzyme-linked antibody.

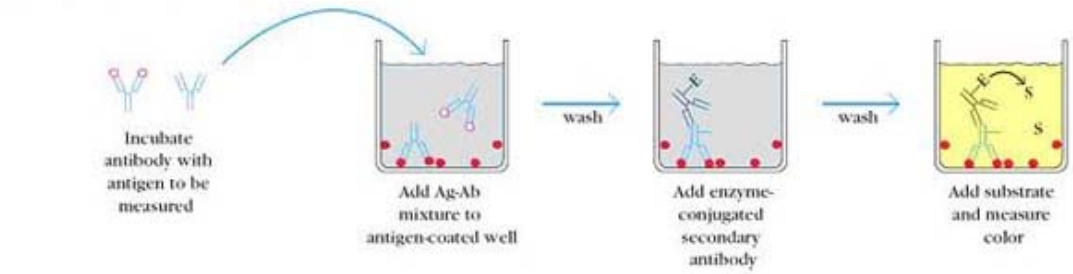


Figure 1.8: Competitive ELISA: the more antigen in sample the less antibody available to bind to antigen coated well, hence signal intensity is inversely proportional to antigen concentration in sample.

1.5.2.3 Multiplex Assays

The simultaneous quantitative assessment of multiple potential biomarkers across large cohorts presents a major challenge to the field. Multiplex assays are used for measuring the levels of multiple analytes in a single sample. These assays use substantially less sample and reagents than the traditional ELISA (which is further limited by its ability to measure only a single antigen) and have been shown to be reproducible, reliable, robust, and accurate, however their performance has also been shown to be analyte, sample type, and concentration dependent, and each assay's dynamic range, linearity, CV (the "coefficient of variation", describing the standard deviation as a percentage of the mean), and percentage recovery are important for obtaining accurate and reproducible assay results (Fu, Zhu et al.). Virtually all of the commercially available bead array kits are supplied with reference standards and thus provide quantitative information.

There are several differences between bead array assays and ELISAs. The multiplex bead array assays (MBAA) use fluorescence as a reporter system unlike ELISAs which use enzyme amplification of a colorimetric substrate. Luminex captures ligands onto spherical beads in suspension while ELISAs generally rely upon flat surfaces in 96-well plates. The multiplex nature of these assays can invariably lead to cross-reactivity, something generally avoided in ELISA methodologies as they study one analyte at a time (Elshal and McCoy 2006). A competing technology is protein microarray kits which capture antibodies and reporter antibodies in a multiplex fashion similar to MBAA, however these assays are relatively new, are not widely accepted as a 'gold standard' for clinical use, and may also be of limited sensitivity (Copeland, Siddiqui et al. 2004). Reliable MBAA results have been seen to be obstructed by anti-cytokine antibodies which may cross-react with other cytokines (Kellar and Douglass 2003), by cross-species antibodies, as well as by other interfering substances (Pang, Smith et al. 2005). It cannot therefore be assumed that a reliable single-plex assay can automatically be incorporated into a reliable multiplex array and instead each analyte must be tested for non-reactivity against all the other antibodies to be used in the multiplex format. Specialised platforms such as the Luminex system incorporate software capable of automatic gating as well as computation of absolute protein/cytokine levels, and can significantly reduce the complexity in performing these assays and also require less user interaction.

The Luminex multi-analyte xMAP technology combines advanced fluidics, optics, and digital signal processing with a microsphere technology. The microspheres are tiny colour-coded polystyrene beads dyed with distinct proportions of red and near-

infrared fluorophores. Each bead set can be coated with a reagent specific to a particular bioassay, allowing the capture and detection of specific analytes from a sample. A light source inside the luminex system excites the internal dyes that identify each microsphere particle and any reporter dye captured during the assay. Up to one hundred different detection reactions can be carried out simultaneously on the various bead populations in very small volumes (Earley, Vogt et al. 2002). The Luminex-100 for example has a three-colour fluorescence signal-detection system with two colours dedicated to microsphere classification and the third colour for measurement of the reporter fluorescence intensity. Many readings are made on each bead set, which further validates the results. Using the conjugation-ready bead sets provided by Luminex it is possible to develop MBAAs within any laboratory. Studies that looked at both ELISA and MBAAs demonstrated good correlations but often poor concurrence of quantitative values and a comparison of randomly selected bead array assays and ELISAs will likely demonstrate substantial differences, however if comparisons are made between MBAA and ELISA which use identical capture and reporter antibodies, as well as similar diluents and serum blockers, variability will be minimised, correlations will be good, and similar quantitative values will be achieved (Elshal and McCoy 2006).

1.6 Proteomic Methods

Proteomics has been defined as the study of the proteome, which is composed of the total collection of all the cellular proteins present at a given time point and the physiological state in a defined biological entity (Bharti, Ma et al. 2007). It has also been defined as the global analysis of gene expression at the protein level and

comes with the promise of unravelling the cellular mechanisms of diseases with the hope of developing reliable markers for disease diagnosis, prognosis, and therapy (Mirza and Olivier 2008).

Proteomics can involve separating proteins from a sample on SDS-PAGE gels followed by subsequent identification of the proteins via mass spectrometry, however this gives us no information about the differences in protein behaviour between normal and disease states and when you take into account the vast number of proteins, considered to be a figure of over one million (Jensen 2004) compared to gene numbers of between 20,000 and 25,000 (Pertea and Salzberg 2012), there is a great need to narrow down this quantity to a workable figure. There have been great advances in proteomics to help with this, with many labelling techniques available to help with quantitative comparisons between sample sets and there have been vast strides made in mass spectrometry methods to complement these techniques. There is now a move towards label-free spectrometry, which is less labour intensive and less expensive than label-based proteomic techniques and can also remove technical issues caused by inefficient labelling, a problem seen with iTRAQ and ICAT (Wu, Wang et al. 2006). These techniques, and various other proteomic techniques for biomarker discovery, will be examined in detail.

1.6.1 Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE)

The first method used in proteomics, and still used, involves a high-resolution separation of proteins based on two properties; two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) separates proteins by their isoelectric point (p*I*) in the first dimension and then by their molecular weight (MW) in the second dimension. The combination of two techniques to better resolve proteins has been around for a

long time (Poulik and Smithies 1958, O'Farrell 1975). An important step forward later on was the development of microanalytical techniques able to identify proteins at the amounts seen in 2D gels. Prior to loading samples on a 2D-gel accurate protein concentration assays must be performed to determine the amount of protein in a sample, which is important for a number of reasons such as ensuring that the amount of protein to be separated is appropriate for the gel size and visualization method being used, to facilitate comparison among similar samples as image-based analysis is simplified when equivalent quantities of proteins have been loaded on the gels to be compared, and it is necessary in cases where the protein sample is labelled with fluorescent dye before separation as the amount of dye needed is relative to the amount of protein in the sample (Berkelman 2008).

Isoelectric focusing (IEF) is used to separate proteins according to their p*I*, which is the pH at which the electric charge on the protein is neutral, therefore the protein carries no net electrical charge and can no longer move in an electrical field. Commercial gel strips are available with immobilised pH gradients (IPG) for this purpose (Rabilloud 2002). These strips are placed on top of SDS polyacrylamide gels and the proteins, via electrophoresis, are separated according to their molecular weight, resulting in a two dimensional representation of the proteins contained within a gel slab.

A number of staining methods are available to enable visualisation of the proteins in the gel such as Coomassie Blue G-250 and more sensitive methods such as silver staining when there are only very small quantities of protein available. Coomassie Blue staining remains the most used stain for a number of reasons; (i) it is relatively cheap, (ii) visible with the eye, (iii) basic desk top scanners can be employed for

image acquisition, (iv) better for quantitative analysis than silver staining, (v) possible modifications can be performed for fast or highly sensitive staining, and (vi) it is compatible with mass spectrometry (Westermeier 2006). However, for biomarker protein abundance comparison studies it is relatively ineffective.

1.6.1.1 2D-Difference In-Gel Electrophoresis (2D-DIGE)

2D-DIGE is a fluorescent dye based protein staining method which can give quantitative information on differences between sample sets. It involves covalent labelling prior to 2D-PAGE, employing three fluorescent cyanine dyes (Cy2, Cy3, and Cy5) which are size and charge matched. Traditional 2D-PAGE is time-consuming, labour-intensive and, due to the lack of reproducibility between gels, leads to significant system variability making it difficult to distinguish between system variation and induced biological change, which means that real differences between protein abundance attributed to a cancer state can seldom be predicted with any real confidence (Marouga, David et al. 2005). 2D-DIGE involves the use of an internal standard, included on all gels in the experiment, from which all other gels can be normalised against, resulting in accurate reproducibility and protein abundance comparisons between gels (Alban, David et al. 2003).

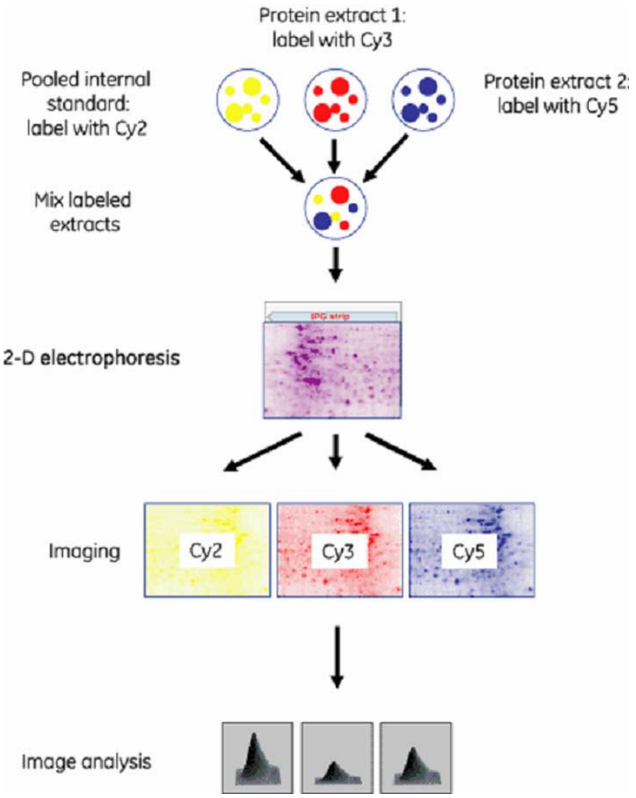


Figure 1.9: The workflow for a 2D-DIGE experiment (Andres Ritter).

Computer software such as DeCyder and Progenesis SameSpots are used to match gels to a reference image (generally the best resolved gel with the most number of protein spots) in order to perform comparison studies and there are statistical modules in place to indicate which protein abundance disparities can be examined with the most confidence. Coupled with immunodepletion and with subsequent mass spectrometry analysis, 2D-DIGE can be used to indicate possible disease markers in serum (Dowling, O'Driscoll et al. 2007). 2D-DIGE has also proved sensitive enough for proteomic profiling of the mitochondrial proteome of young adult rat versus aged rat skeletal muscle to find reliable mitochondrial marker candidates that could be used in the establishment of a biomarker signature of skeletal muscle aging (O'Connell and Ohlendieck 2009). However, spots on a given 2D gel often contain more than one protein, making quantification of the protein

ambiguous since it is not always clear which protein in the spot has actually changed. Also, gel-based methods have a number of restrictions on the proteins they can resolve; having difficulty handling or detecting proteins which are hydrophobic, have extreme molecular weights and p*I* values, or in samples with a high dynamic range.

1.6.2 Stable Isotope Labelling Proteomics

1.6.2.1 Stable isotope labelling with amino acids in cell culture (SILAC)

In contrast to other isotope labelling techniques, SILAC is an *in vivo* labelling strategy in which the proteome is labelled as cells grow in culture. SILAC relies on the incorporation of amino acids with substituted stable isotopic nuclei (e.g. deuterium, 13C, 15N) so that two cell populations can be grown in culture media that are identical except that one of them contains a 'light' and the other a 'heavy' form of a particular amino acid (e.g. 12C and 13C labelled L-lysine, respectively) which can be readily distinguished by mass spectrometry. If the labelled analogue of an amino acid is supplied instead of the natural abundant amino acid, it will be incorporated into each newly synthesized protein chain (Ong, Blagoev et al. 2002).

In fact, SILAC has the ability to compare up to five states in a single experiment, the number of states restricted by the limited availability of heavy forms of amino acids, though it can only be used in metabolically active cells so cannot be used for tissue samples (Harsha, Molina et al. 2008). It has been used to study posttranslational modifications such as protein phosphorylation and methylation, to characterise signalling pathways and to determine specific protein interactions (Gruhler and Kratchmarova 2008). The labelling process itself is uncomplicated and

highly efficient (100% of sample is available for labelling) and due to the combination of unlabelled and labelled samples prior to cell lysis it allows any method of protein or peptide purification to be used without introducing error (Ong, Foster et al. 2003).

SILAC has been used for phosphoproteomic analysis (Pimienta, Chaerkady et al. 2009), to discover differentially expressed plasma membrane proteins in renal cell carcinoma (Aggelis, Craven et al. 2009), to identify secreted proteins for pancreatic serum biomarker discovery (Yu, Barry et al. 2009), and even to help quantitatively identify proteins linked to human embryonic stem cells (Collier, Sarkar et al.), cells which could be used in the treatment of numerous medical conditions, showing its usefulness as a complementary technique for generating a list of possible protein disease markers.

1.6.2.2 Isotope-Coded Affinity Tags

The isotope coded affinity tag is a chemical modification strategy that allows quick and accurate quantification coupled with sequence identification of proteins in a complex mixture. Isotope-coded affinity tags (ICAT) and an improved version known as isobaric tags for relative and absolute quantification (iTRAQ) are the two main methods used. ICAT utilises reagents consisting of three functional groups; a sulfhydryl-reactive iodoacetate group, a biotin affinity group, and a linker carrying light or heavy isotopes (Sethuraman, McComb et al. 2004). ICAT-labelled peptides elute as pairs from a reverse-phase column so by calculating the ratio of the areas under the curve for identical peptide peaks labelled with the light and heavy ICAT reagent, the relative abundance of that peptide in each sample can be determined, which is directly related to the abundance of the corresponding protein. In addition,

because the ICAT reagents are specific for cysteinyl residues, the complexity of the original peptide mixture is greatly reduced (Li, Steen et al. 2003). Also, because it is based on post-isolation stable isotope labelling of proteins it is not limited to cells and tissues compatible with metabolic labelling.

The **ICAT** approach allows quantitative cataloguing and comparison of protein expression in a number of states; normal, developmental, and disease, with obvious uses in cancer biomarker discovery. The combination of ICAT and multidimensional chromatography followed by mass spectrometry of digested proteins is able to detect and quantify low abundant proteins in complex mixtures (Gygi, Rist et al. 2002).

There are a number of limitations to ICAT including missed identification of proteins with few or no cysteine residues, lost information for post-translational modifications (PTM's), differential reversed-phase elution of identical peptides labelled with the hydrogen/deuterium isotope pairs, and complicated interpretation of tandem mass spectrometry (MS/MS) spectra due to the addition of the biotin group (Goshe and Smith 2003, Leitner and Lindner 2004), however most of these problems have been solved by the use of cleavable isotope-coded affinity tags (cICAT).

cICAT involves labelling samples with the isotopically light or heavy cICAT reagent, combining them, digesting with a protease, retrieval using an avidin affinity column and importantly having the added bonus of being able to remove the retrieval ligand prior to MS, and subsequently can be used to reduce the complexity of digested proteins and to improve quantification of low-abundant proteins (Qu, Jusko et al. 2006).

Another method using isotope affinity tags, known as **iTRAQ**, allows up to eight samples to be analysed in a single experiment enabling simultaneous identification and quantification, both relative and absolute, using synthetic isobaric peptide standards that are indistinguishable by their MS spectra or MS/MS ion series, but exhibit intense, low-mass MS/MS signature ions that permit quantitation of members of the multiplex set (Ross, Huang et al. 2004). The distribution of isotopes in the different tags is such that when the tags fragment a reporter ion is released which is tag-specific and the ratio of signal intensities from these tags act as an indication of the relative proportions of that peptide between the different labelled samples (Unwin). Performed in 4plex or 8plex experiments, the amine specificity of the iTRAQ reagents makes most peptides in a sample amenable to this labelling strategy.

iTRAQ, generally in conjunction with either LC-MS/MS or MALDI-TOF/TOF MS or a combination of both, has been used to find new and better biomarkers for gastric cancer (Chong, Lee et al.), to uncover clinically relevant candidate markers for prostate cancer progression (Glen, Evans et al. 2010), to study p53-modulated proteins secreted in lung cancer cells (Chenau, Michelland et al. 2009), to identify serum biomarkers for oral squamous cell carcinoma (Bijian, Mlynarek et al. 2009), to investigate low-grade breast primary tumour tissues with and without metastases and metastasis for relevant biomarkers (Bouchal, Roumeliotis et al. 2009), and to identify serum biomarkers in brain-injured patients that may predict elevated intracranial pressure (Hergenroeder, Redell et al. 2008), to name a few areas of biomarker discovery iTRAQ has been used, and is presently used, for.

There are however limitations with labelling-based quantification approaches including increased time and complexity of sample preparation, requirement for higher sample concentration, the high cost of reagents, incomplete labelling, and the requirement for specialised computer software for quantification. Also, labelling strategies limit the number of samples that can be analysed in a single experiment and some labelling strategies can not be applied to all types of samples .

1.6.3 Mass Spectrometry

Mass spectrometry (MS) is used for peptide fragmentation and protein identification. The more common bottom-up mass spectrometry technique involves fragmenting peptides in the gas phase after protein digestion with a protease such as trypsin. The less common but becoming gradually more popular top-down technique involves fragmenting intact proteins directly.

Identifying proteins using MS requires an interconnected relationship between mass spectrometry instrumentation. For example there are differences in how molecules are ionised, activated, and detected. Gas-phase peptide chemistry determines which bonds are broken, at what rate, the cleavage depending on factors such as peptide/protein charge state, size, composition, and sequence. There is a huge variety of MS/MS instrument configurations with different aptitudes for speed, ionisation method, resolution, sensitivity, and mass/charge range (Wysocki, Resing et al. 2005). To identify proteins in complex biological samples, more sensitive mass spectrometers are used with low flow, high resolution separation technologies but the more sensitive the mass spectrometer generally the more expensive it is. There is a vast, and continuously expanding, amount of genome and proteome information stored in databases which can be used for MS and MS/MS applications

and the lists of peak intensities and mass to charge ratios produced by MS can be compared to lists generated from 'theoretical' fragmentation of a peptide or digestion of a protein to yield protein identifications.

For biomarker discovery, a benefit of using mass spectrometry is its ability to discern structural modifications to proteins (e.g. posttranslational modifications such as phosphorylation and glycosylation), and as proteomic techniques improve so too will discovery of these changes and whereas it is difficult to build conventional assays to such candidates, targeted mass spectrometry (such as multiple reaction monitoring (MRM)) can be applied to overcome this problem (Wang, Whiteaker et al. 2009).

1.6.3.1 Ionisation

To analyse a sample by MS it must be first ionised and vaporised. The two main ionisation techniques are electrospray ionisation (ESI) and matrix-assisted laser desorption/ionisation (MALDI). In ESI, ions are formed at atmospheric pressure, while ions in MALDI may be generated either at atmospheric pressure or under vacuum conditions (Canas, Lopez-Ferrer et al. 2006). Conventional **electrospray** involves initially dissolving your protein sample in a solvent, meaning it will exist in an ionised form, and this solution is pumped through a thin capillary which is raised to a high potential forming small charged droplets which are sprayed from the ES capillary into a bath gas at atmospheric pressure before travelling down a pressure and potential gradient towards an orifice in the MS high-vacuum system (Griffiths, Jonsson et al. 2001). As droplets make this journey they become desolvated and reduced in size to such an extent that surface-coulombic forces overcome surface-tension forces and subsequently the droplets are broken up into

even smaller droplets. There are two theories for ion formation in ESI, one being that ionised sample molecules are expelled from the droplets (Bakhtiar and Nelson 2001), the other that individual ionised molecules remain after continuous solvent evaporation and droplet fragmentation (Mora, Van Berkel et al. 2000).

In **matrix-assisted laser desorption/ionisation** the sample is usually embedded in an excess of a solid matrix, which, upon laser irradiation, assists in the volatilization and ionization of the analytes and additionally, the high matrix/sample ratio reduces associations between analyte molecules, and provides protonated and free-radical products that ionize the molecules of interest (Marvin, Roberts et al. 2003). The matrix is the primary absorber of the UV laser radiation and because it breaks down rapidly expanding into the gas phase it is used in high excess compared to the sample itself and in this way it minimises sample degradation from the laser radiation. The matrix is a crystal-like structure of weak organic acids; examples of such include a-Cyano-4-hydroxycinnamic acid (CHCA), Sinapic acid (SA), and 2-(4-Hydroxyphenylazo) benzoic acid (HABA), and the matrix of choice will have to strongly absorb the energy of the laser beam at a wavelength where analytes exhibit only weak absorption (Bonk and Humeny 2001). The sample and the matrix co-crystallise together on the MALDI plate.

1.6.3.2 Mass Analysis

ESI and MALDI mass spectrometers are combined in various ways with different mass analysers. The mass analysers generally used are quadrupoles, Time of Flight (TOF) instruments, and ion traps, either alone or in hybrid systems. One of the most important characteristics of mass analysers is their resolution, i.e. their ability to

differentiate two close signals. The key parameters for mass analysers in proteomics are sensitivity, resolution, mass accuracy, and the ability to generate information-rich ion mass spectra from peptide fragments (tandem MS or MS/MS spectra) (Aebersold and Mann 2003).

In ion trap MS (IT-MS), ions are stored until the trap is full and by varying the radio frequency (RF) voltage applied to the ring electrode, ions are ejected out of the endcap and are detected by the electron multiplier, however in quadrupole MS (Q-MS), ions are continuously formed, accelerated into the quadrupole and mass analyzed, consequently, both types of mass spectrometers have specific advantages; with IT-MS higher sensitivity is obtained in full scan, whereas ion ratio stability is better for Q-MS (Fitzgerald, O'Neal et al. 1997).

MALDI-TOF and the related technique, surface enhanced laser desorption/ionisation (SELDI)-TOF both use time-of-flight (TOF), in which a time measurement is used to measure ions mass-to-charge ratio. In MALDI-TOF MS, a small amount, typically ~1 μL, of specimen containing peptides and protein is dried on a target plate together with a light-absorbing matrix molecule. The protein/matrix sample is vaporised from surface deposits by nanosecond-duration laser pulses which release ionized protein molecules, often with a single charge (z =1), which are then accelerated in an electric field within a vacuum. Ions with low mass/charge ratios (m/z) are accelerated to higher velocities and reach the detector before ions with a high m/z. Ions reach the detector over a time span of ~0.01–1 ms; the millisecond timescale to analyze ions produced by a single laser pulse allows acquisition of data from multiple laser pulses for each specimen and provides short

analysis time per specimen, and therefore the technique lends itself to high-throughput analysis (Hortin 2006). MALDI-TOF sample analysis is independent of factors other than mass because separations occur in a vacuum. Its ability to desorb high-molecular-weight thermolabile molecules, its high accuracy and sensitivity, combined with its wide mass range (1-300kDa) makes MALDI-TOF MS a usual platform for identification of biomolecules in complex samples (Marvin, Roberts et al. 2003).

In SELDI-TOF the sample is spotted onto a protein-chip array, with each spot on the chip consisting of a chromatographic surface that allows attachment and enrichment of particular proteins under defined conditions from a complex protein mixture, such as serum. The surface may consist of various materials of different physico-chemical characteristics such as hydrophobic or hydrophilic properties, containing metal ions (such as immobilised metal affinity capture: IMAC) or anion and cation exchangers (Kiehntopf, Siegmund et al. 2007), or it can be pre-activated for the coupling of capture molecules such as protein, DNA, or RNA prior to sample loading. After the binding phase of the sample to these surfaces, the unbound proteins are washed away while retained molecules are overlaid with a laser-absorbing matrix, and then the sample is ionised with a laser and the mass spectra is recorded after ions are resolved with the TOF spectrometer.

Protein profiling using SELDI-TOF for biomarker discovery has become more popular in recent years. One of the key features of SELDI-TOF MS is its ability to provide rapid protein expression profiles from a variety of biological samples with minimum requirements for purification and separation of proteins prior to mass

spectrometry (De Bock, de Seny et al.). Biofluids used on SELDI chips include plasma, serum, saliva, urine, amniotic fluid, and nipple aspirate fluid, however, when working with biofluids great care must be taken as non-biological differences can be caused by variances in sample collection, processing, storage, and analytical techniques. The range detectable by SELDI-TOF and indeed MALDI-TOF is about 2 orders of magnitude so a complex sample such as serum over at least 9 orders of magnitude (Adkins, Varnum et al. 2002) must undergo some fractionation procedures before it can be analysed.

An advantage of SELDI-TOF-MS is its relatively high tolerance for salts and other impurities, and also sample requirement is low, and can range from 5ul to 400ul (Wiesner 2004). The major disadvantage is a lack of protein identification; identification by sequencing or peptide fingerprinting involves enrichment and purification of the biomarker, which is arduous and time-consuming. To partly solve this problem new ProteinChip interface coupled to a tandem mass spectrometer was recently developed allowing direct sequencing of peptides <6,000 Da .

A study to find biomarkers for early detection of ovarian cancer identified three biomarkers in serum using SELDI-TOF-MS; apolipoprotein A1, a truncated form of transthyretin, and a cleavage fragment of inter-α-trypsin inhibitor heavy chain H4, and when combined with a known ovarian cancer biomarker CA125, improve significantly on the specificity of CA125 alone while maintaining a relatively high sensitivity (Zhang, Bast et al. 2004).

1.6.3.3 Label-Free Mass Spectrometry

Isotope and fluorescent labelling have been widely used for quantitative proteomic research however there is a move towards label-free spectrometry as this can yield faster, cleaner, and simpler results. As an alternative to 2D-PAGE, two-dimensional chromatography coupled to mass spectrometry systems has been used for protein identification. This initially focused on protein separation and failed to identify large numbers of proteins in a sample. Improvements to this set-up resulted in the development of a method called multidimensional protein identification technology (MudPIT), which combines multidimensional liquid chromatography with electrospray ionization tandem mass spectrometry, and involves loading a complex peptide mixture generated from a biological sample on to a biphasic microcapillary column packed with reversed phase (RP) and strong cation exchange (SCX) high performance liquid chromatography (HPLC) grade materials (Wolters, Washburn et al. 2001). In HPLC, the sample is in a liquid phase known as the mobile phase and is forced through a column that is packed with a stationary phase composed of irregularly or spherically shaped particles, a porous monolithic layer, or a porous membrane, at high pressure. MudPIT is an online approach but SCX can also be run offline where SCX fractions are generated and then individually subjected to RP chromatography into a mass spectrometer, as has been done with a study on the proteome of *S. cerevisiae* (Peng, Elias et al. 2003).

Label-free spectrometry involves measurements of changes in chromatographic ion intensity such as peptide peak areas or peak heights, and also spectral counting of identified proteins. Label-free can be broken into two groups; (1) area under the curve (AUC) or signal intensity measurement based on precursor ion spectra, and (2) spectral counting, which is based on counting the number of peptides assigned

to a protein in an MS/MS experiment. Label-free shotgun proteomics produces large amounts of data that require meticulous statistical assessment.

The general process of quantitating proteins based on AUC involves the measurement of ion abundances at specific retention times for the given ionised peptides without the use of a stable isotope standard, which is often referred to as ion counts, and as ionised peptides elute from a reversed-phase column into the mass spectrometer, their ion intensities can be measured within the given detection limits of the experimental setup and differentially expressed peptides can be validated by LC-MS/MS identification either subsequently or simultaneously . This method requires that all data is collected in data-dependent 'Triple Play' mode (allowing MS scan, Zoom scan, and MS/MS scan) and these together with chromatographic retention time determine the analytical accuracy of protein identification and quantification by ion intensity (Higgs, Knierman et al. 2005).

There are concerns with LC signal resolution when peptide signals are spread over a large retention time causing overlap and co-eluting peptides, and concerns with biological variations resulting in multiple signals for the same peptide as well as technical variations in retention time MS intensity, and sample background noise from chemical interference.

For spectral counting, the more of a particular protein present in a sample the more MS/MS spectra are collected for peptides of that protein, therefore relative abundance of that protein can be obtained by comparing the number of MS/MS spectra among a set of experiments, the advantage of this being that both protein

quantification and identification can be achieved by extensive MS/MS data collection across the chromatographic time scale simultaneously (Wang, You et al. 2008). This method is dependent on the quality of MS/MS peptide identification and errors in this can lead to inaccurate protein quantification (Bantscheff, Schirle et al. 2007), however the correlation between the amount of protein and number of MS/MS spectra can generally be trusted (Gilchrist, Au et al. 2006). The spectral counting method has diversified from summing spectra to modifying counts with a normalisation factor and, more recently, combining strategies for increased accuracy, and many spectral counting strategies and statistical tools for analysing spectral count data have emerged in recent years (Lundgren, Hwang et al.).

Label-free quantitative proteomics has been used to find potential biomarkers for various conditions and diseases, for example kidney chronic allograft dysfunction (Quintana, Campistol et al. 2009), a common chronic liver disease called non-alcoholic fatty liver disease (Bell, Theodorakis et al.), melanoma tumour progression (Huang, Darfler et al. 2009), periodontal disease (Bostanci, Heywood et al. 2010), invasive ductal breast carcinoma (Rower, Vissers et al. 2009), and lung cancer (Zeng, Hood et al. 2011).

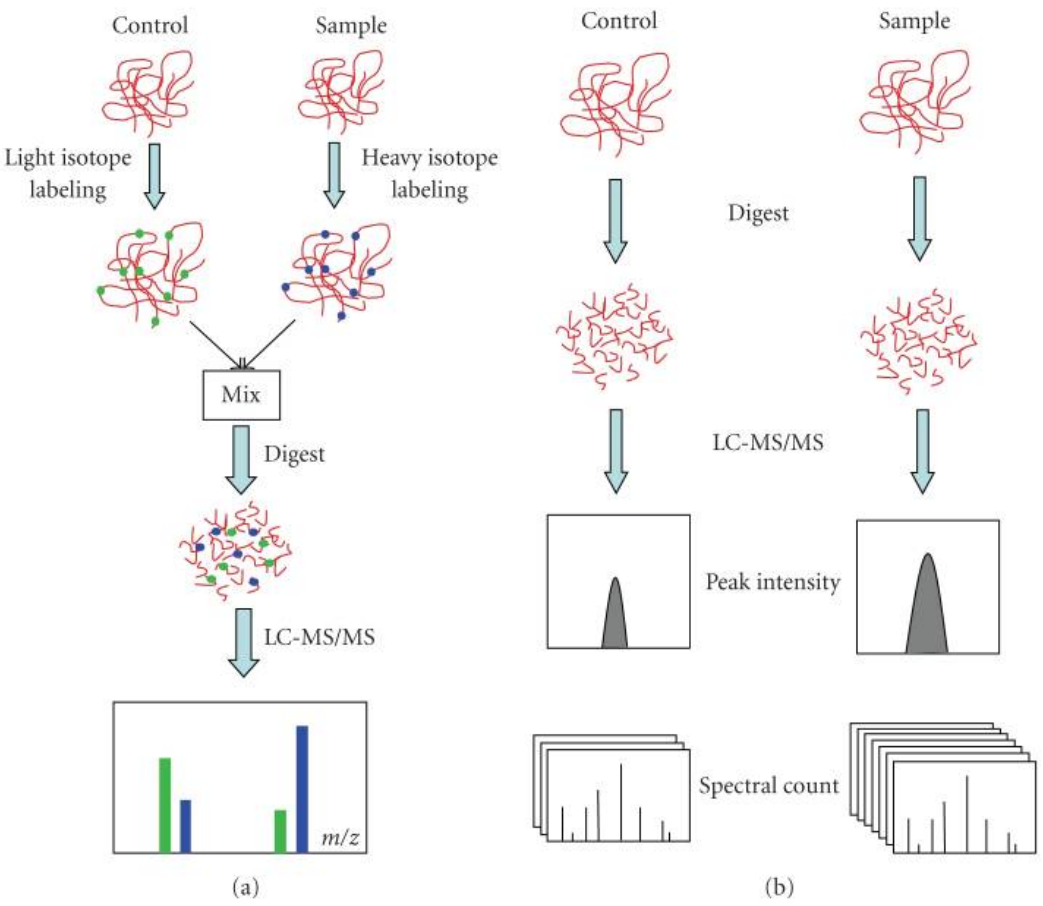


Figure 1.10: General approaches of quantitative proteomics. (a) Shotgun isotope labelling method. After labelling by light and heavy stable isotope, the control and sample are combined and analyzed by LC-MS/MS. The quantification is calculated based on the intensity ratio of isotope-labelled peptide pairs. (b) Label-free quantitative proteomics. Control and sample are subjected to individual LC-MS/MS analysis. Quantification is based on the comparison of peak intensity of the same peptide or the spectral count of the same protein (Zhu, Smith et al.).

1.6.3.4 Multiple Reaction Monitoring (MRM)

Multiple reaction monitoring (also known as selected reaction monitoring) uses mass spectrometry for highly sensitive and selective targeted quantitation of protein/peptide abundances in complex clinical samples. Untargeted MS-based

quantitation work flows concentrate on lengthy, exhaustive prefractionation sample methods. These methods involve both the protein and peptide level sometimes with the goal of detecting and quantifying an entire proteome. This untargeted approach can involve lengthy MS data acquisition and analysis times so can be limited to quantifying small sample sets. Therefore there is a need to reduce MS time and simplify sample preparation in order to quantitate larger clinical sample sets (Kuzyk, Smith et al. 2009). MRM is highly selective and allows the fine-tuning of MS instrumentation to specifically look for peptides, or protein fragments, of interest and is commonly used for the analysis of small molecules.

Multiple reaction monitoring (MRM) is capable of rapid, sensitive, and specific quantitation of analytes in highly complex matrices using tandem MS/MS scan mode unique to triple quadrupole MS instrumentation (Hoofnagle, Becker et al.). MRM is a targeted approach that requires knowledge of the molecular weight of an analyte and its fragmentation behaviour under CID. Electrospray ionisation is followed by two stages of mass selection, the first selecting the mass of the intact analyte, known as the parent ion, and after fragmentation of this ion by collision with gas atoms, the second stage then selects a specific fragment of the parent ion, collectively generating a selected reaction monitoring assay (plural MRM). The detection and integration of a peak for the selected analyte is possible due to the two mass filters producing a very specific and sensitive response in a simple 1D chromatographic separation of the sample. This structural specificity for the analyte, in combination with appropriate stable isotope-labelled internal standards (SISs) can provide absolute quantitation of analyte concentration (Anderson and Hunter 2006). These methods are used in the clinical laboratory for the routine

measurement of endogenous metabolites, for example immunosuppressants (Streit, Armstrong et al. 2002) and screening of newborns for a panel of inborn errors of metabolism (Roschinger, Olgemoller et al. 2003). MRM has also been used as a powerful analytical method for the prescreening of candidate protein biomarkers in serum prior to antibody and immunoassay development (Kuhn, Wu et al. 2004).

1.6.4 Immunohistochemistry

Immunohistochemistry (IHC) can be used in the routine diagnosis of lung cancer in order to identify biological markers. The objective of performing immunohistochemistry is to recognize cell antigens on tissue which are then viewed under microscopy (the visualization of the antigen-antibody complex is made possible through the addition of either a fluorochrome conjugate or an enzyme to the antibody) and, consequently, to identify and classify specific cells within a cell population whose morphology is heterogenous or apparently homogenous (Capelozzi 2009). Standardization of IHC for quantitation is becoming more important due to the requirement in translational research for quantitation of the differential expression of various prognostic markers for cancer, based on retrospective study of clinical cases with known clinical outcomes.

Quantitative evaluation of protein levels in tumour samples is challenging, with technical issues arising related to immunohistochemistry procedure and subjectivity of assessment. Due to the significant heterogeneity that can be seen in lung cancer small tumour specimens, such as those frequently obtained in advanced lung cancer, they may not be representative of the tumour itself (Taillade, Penault-Llorca et al. 2007). There may be as much as a 30% variance seen in protein expression in

primary versus metastatic tumour sites, though the clinical significance of this difference is unknown. Small samples are more susceptible to the "edge effect" with non-specific staining observed in the very edge of the specimen and it should be noted that the choice of primary antibody and antigen retrieval (AR) conditions are of utmost importance (Szutowicz and Dziadziuszko).

The **antigen retrieval technique** is a simple method of boiling archival paraffin-embedded tissue sections in water to enhance the signal of immunohistochemistry, though knowledge of the exact localization of a certain protein in tissue is critical to interpret not only the accuracy of IHC staining results but also the reliability of AR treatment (Shi, Cote et al. 2001). The heating conditions and the pH value of the AR solution are the two major factors that influence the effectiveness of IHC staining. The success of AR has shown that the modification of protein structure by formalin is reversible under certain conditions, such as high-temperature heating or strong alkaline treatment (Morgan, Navabi et al. 1997).

A pathologist performs a quantitative measurement of the observed immunostaining, an indication of protein level, with numerous scales and there is no consensus which scale is better or more practical, and often percentage of cells with positive staining of any intensity is recorded (Szutowicz and Dziadziuszko). To account for the protein abundance, the percentage of cells with positive staining is multiplied by the staining intensity (graded from 0 to 3 or from 0 to 4) to give the staining index (ranging from 0 to 300 or from 0 to 400, respectively) (Cappuzzo, Hirsch et al. 2005). In other systems, the staining index is composed of percentage of positive cells binned into categories and multiplied by staining intensity (0 to 3 or 0 to 4 scale) (Olaussen, Dunant et al. 2006). Regardless of which classification

you use, it is very important to note which part of the cell is stained, whether it's the cell membrane, cytoplasm, or nucleus, or in more than one of these cell compartments. Precise quantification of the protein level by the pathologist provides accurate data to be analyzed with patient outcome and to select the best prognostic or predictive cut-off points of the test. Reproducibility of immunohistochemical analysis is highly vulnerable to (i) staining variation as a function of tissue preservation and processing, (ii) to disparities in assay protocol and reagents, and (iii) to the interpretation bias inherent with observer-based assessment devoid of quantitative scoring systems (Taylor and Levenson 2006).

IHC analysis of lung adenocarcinoma was used to indicate paxillin over-expression occurs during the earliest stages of lung cancer development (Mackinnon, Tretiakova et al. 2011). Protein tyrosine phosphatase (SHP2) expression, with the help of IHC staining in the tissue of NSCLC patients who smoke, could play a role in the development of lung cancer (Zhan, Dong et al. 2010). IHC was used to investigate the expression of tumour necrosis factor-alpha (TNFalpha) in the tumour islets and stroma of patients with surgically resected NSCLC and its expression is associated with improved survival suggesting a role in the host anti-tumour immunological response (Ohri, Shikotra et al. 2010).

1.7 Metabolomics

Metabolomics is the global quantitative assessment of metabolites within a biological system. It is used to detect metabolites and their changes in biofluids or tissue. Like genomics, transcriptomics, and proteomics, metabolomics is a science

of systems biology, which is a biology-based inter-disciplinary study field that focuses on complex interactions in biological systems. Metabolomics could address the limitation associated with transcriptomics and proteomics where changes do not always result in altered biochemical phenotypes. Systems biology could be indispensible to the future discovery of biomarkers for early detection and diagnosis of cancer and for predictive and pharmacodynamic evaluation of drug effect. The level of metabolites at any given time represents a composite of both catabolic and anabolic processes and presents a snapshot of the metabolome at that particular time only, though this can be an advantage as time dependent snapshot sampling will reveal directed correlations of metabolic processes (Weckwerth 2003).

Metabolomics involves the detection of metabolites, either individually or grouped as a metabolomic profile, and is carried out in cells, tissues, or biofluids by either nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry, and when used as a translational research tool can provide a link between the laboratory and clinic, enabling the discrimination of metabolic markers noninvasively *in vivo* (Spratlin, Serkova et al. 2009).

Metabolomics involves a number of different analyses, including (i) metabolic fingerprinting, which measures a subset of the whole profile with little differentiation or quantitation of metabolites (Ryan and Robards 2006); (ii) metabolic profiling, the quantitative study of a group of metabolites, known or unknown, within or associated with a particular metabolic pathway (Dunn, Bailey et al. 2005) and (iii) target isotope-based analysis, which focuses on a particular segment of the metabolome by analyzing only a few selected metabolites that comprise a specific biochemical pathway (Boros, Lerner et al. 2005).

Metabolomics certainly has potential for finding useful biomarkers for cancer. It has been used to show that tumours display elevated phospholipid levels and increased glycolytic capacity including increased utilization of glucose carbons to drive synthetic processes, high glutaminolytic function, and overexpression of pyruvate kinase type M2 (Spratlin, Serkova et al. 2009). Using NMR-based metabolomics on blood samples it was shown that lipid metabolic profiles were 83% accurate at discriminating between cancer patients and controls (Bathen, Engan et al. 2000).

The urinary profile of lung cancer patients was studied and metabolic alterations enabled their discrimination from healthy subjects with high sensitivity and specificity (Carrola, Rocha et al. 2011).

One study using NMR and GC-MS metabolomic analysis compared non-cancerous lung and tumour tissues, showing accelerated glycolysis in tumour tissues (Fan, Lane et al. 2009).

1.8 Aims of Thesis:

The aims of this thesis are as follows:

1. To create lists of potential lung cancer biomarkers from the 2D-DIGE and subsequent DeCyder software analysis of conditioned media from the main lung cancer types; squamous cell carcinoma, adenocarcinoma, large cell carcinoma, and small cell carcinoma.

2. To validate some of the interesting proteins from the 2D-DIGE experiment of conditioned media study in serum from lung cancer patients and normal healthy patients.

3. To examine the expression of some of the interesting proteins from the 2D-DIGE experiment of conditioned media study in lung cancer and normal lung tissue using immunohistochemistry.

4. To examine the expression of the most interesting protein from the immuno-histochemistry study in different types of normal and cancer tissue.

5. To determine if the most interesting protein from the validation of the 2D-DIGE experiment of conditioned media study has any effect on the invasion and migration of lung cancer cells.

6. To perform 2D-DIGE on immunodepleted serum of lung cancer and healthy patients to discover potential lung cancer biomarkers.

7. To perform label-free mass spectrometry analyses of serum run through ProteoMiner columns to discover potential lung cancer biomarkers and subsequently validate the most interesting proteins in a larger cohort of serum samples.

8. To use the Luminex Multiplex assay platform to discover potential lung cancer biomarkers in serum and subsequently validate the most interesting proteins in a larger cohort of serum samples.

9. To use Metabolon diagnostics discovery platform to evaluate metabolites that may function as lung cancer biomarkers and subsequently validate the most interesting metabolites in a larger cohort of serum samples using ELISAs and biochemical assays.

10. To examine in plasma some of the interesting proteins from the serum validation phases.

11. To examine some of the interesting proteins from the validation phases in ELISAs/assays with the inclusion of a benign lung disease control group.

12. To create a lung cancer biomarker panel combination with high sensitivity and high specificity from the list of validated proteins/metabolites.

13. To examine, in serum and plasma, well-known cancer biomarkers on the Luminex system and compare panel of highest specificity and sensitivity to equivalent sized panel using at least one of the new potential lung cancer markers.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Cell Culture

2.1.1 Preparation of Cell Culture Media

Ultrapure water (UHP) was purified to a standard of 12-18 MΩ/cm resistance by a reverse osmosis system (Millipore MillI-RO 10 Plus, Elgastat UHP). Glassware required for cell culture related applications were soaked in a 2% RBS-25 (AGB Scientific) for 1 hour, washed in an industrial dishwasher, using Neodisher detergent and rinsed twice with UHP. All thermostable solutions, water and glassware were sterilised by autoclaving at 121°C for 20 minutes at 15 bar (Thermolabile solutions were filtered through 0.22 μm sterile filters (Millipore, Millex-GV SLGV025BS)).

All 1X basal media were prepared in-house as follows: 10X media was added to sterile UHP water, buffered with HEPES (N-(2-Hydroxyethyl) piperazine-N-(2-ethanesulfonic acid) and $NaHCO_3$ as required and adjusted to pH 7.5 using sterile 1.5 N NaOH or 1.5 N HCl. The media was then filtered through sterile 0.22μm bell filters (Gelman, 12158) and stored in sterile 500ml bottles at 4°C. Sterility checks on all media bottles for bacterial, yeast and fungal contamination were made using Colombia blood agar (Oxoid, CM217), Sabauraud dextrose (Oxoid, CM217) and Thioglycolate broths (Oxoid, CM 173) respectively. All sterility checks were then incubated at both 25°C and 37°C.

Basal media were stored at 4°C for up to three months in the dark. Complete media was then prepared as follows: supplements of 2mM L-glutamine (Gibco, 11140-0350) for all basal media and 1ml 100X non-essential amino acids (Gibco, 11140-

035) and 100mM sodium pyruvate (Gibco, 11360-035) were added to MEM. Other components were added as per cell line requirement. Complete media was stored at 4°C for a maximum of one month in the dark.

2.1.2 Cell lines and cell culture

All cell culture work was carried out in a class II laminar air-flow cabinet (Holten). Before and after use the LF cabinet was cleaned with 70% industrial methylated spirits (IMS). Any items brought into the cabinet were also swabbed down with IMS. At any time only one cell line was used in the LF cabinet and upon completion of work with any given cell line, 15 minutes clearance was given to eliminate any possibilities of cross-contamination between the various cell lines. The cabinet was cleaned weekly with Virkon (Antech International, P0550) and IMS. Details pertaining to the cell lines used for the experiments are provided in Table 2.1.1. All cells were incubated at 37°C and where required, in an atmosphere of 5% CO_2. Cells were fed with fresh media or subcultured (see section 2.5.1) every 2-3 days or as required in order to maintain active cell growth.

Cell line	Cell type	Source
DLKP	Poorly differentiated squamous cell lung carcinoma	NCTCC*
DLRP	Poorly differentiated squamous cell lung carcinoma	NCTCC
SK-MES-1	Squamous cell lung carcinoma	NCTCC
NCI-H226	Squamous cell lung carcinoma	NCI**
A549	Lung adenocarcinoma	NCTCC
CALU-3	Lung adenocarcinoma	NCTCC*
SK-LU-1	Lung adenocarcinoma	NCTCC
NCI-H23	Lung adenocarcinoma	NCI
COR-L23	Lung large cell carcinoma	NCTCC
NCI-H460	Lung large cell carcinoma	NCTCC
NCI-1299	Lung large cell cracinoma	NCTCC
HOP-18	Lung large cell carcinoma	NCI
COR-L24	Small cell lung carcinoma	NCTCC
DMS-53	Small cell lung carcinoma	NCTCC
NCI-H69	Small cell lung carcinoma	NCTCC
DMS-114	Small cell lung carcinoma	NCI
NHBE	Normal Human Bronchial/Tracheal Epithelial Cell Line	LONZA

Table 2.1.1: Description of cell lines used in conditioned media study.

*NCTCC – National Cell and Tissue Culture Centre

**NCI – National Cancer Institute

2.1.3 Subculturing of adherent cell lines

Waste cell culture medium was removed from the tissue culture flask and discarded into a sterile bottle. The flask was then rinsed out with 1ml of trypsin/EDTA solution (0.25% trypsin (Gibco, 043-05090), 0.01% w/v EDTA (Sigma, E9884) solution in PBS (Oxoid, BRI4a)) to ensure the removal of any residual media. Depending on the size of the flask, 2-5ml of trypsin was then added to the flask, which was then incubated at 37°C, for approximately 5 minutes, until all of the cells detached from the inside surface of the flask monitored by microscopic observation. Adding an equal volume of complete media to the flask deactivated the trypsin. The cell suspension was removed from the flask and placed in a sterile universal container (Sterilin, 128a) and centrifuged at 170*g* for 5 minutes. The supernatant was then discarded from the universal and the pellet was suspended gently in complete medium. A cell count was performed and an aliquot of cells was used to seed a flask at the required density. All cell waste and media exposed to cells were autoclaved before disposal.

2.1.4 Subculturing of floating aggregate cell lines

The cell suspension was removed from the flask and placed in a sterile universal container and centrifuged at 170*g* for 5 minutes. The supernatant was then discarded from the universal and the pellet was suspended in complete medium. The aggregates were broken up by gently pipetting the suspension up and down using a 10 ml pipette. A cell aliquot was used to seed a flask.

2.1.5 Assessment of cell number and viability

Prior to cell counts, cells were prepared for subculturing as detailed in 2.1.3 or 2.1.4 depending on adherent or floating aggregates. An aliquot of the cell suspension was

then added to trypan blue (Gibco, 525) at a ratio of 5:1. The mixture was incubated for 3 minutes at room temperature. An aliquot (10µl) was then applied to the chamber of a glass coverslip-enclosed haemocytometer. For each of the four grids, cells in the 16 squares were counted. The average of the four grids were multiplied by a factor of 10^4 volume of the grid) and the relevant dilution factor to determine the average cell number per ml in the original cell suspension. Non-viable cells stained blue, while viable cells excluded the trypan blue dye as their membrane remained intact and remained unstained. On this basis, percentage viability could be calculated.

2.1.6 Cryopreservation of cells

Cells for cryopreservation were harvested in the log phase of growth and counted as described in Section 2.1.5. Cell pellets were resuspended in a suitable volume of serum. An equal volume of a 10 - 20 % DMSO/serum solution was added dropwise with mixing, to the cell suspension. The suspension was then aliqoted in 1 ml volumes to cryovials (Greiner, 122278) and immediately placed in the vapour phase of a liquid nitrogen container. After four hours, vials were transferred to the liquid phase for long term storage (- 196°C).

2.1.7 Thawing of cryopreserved cells

A volume of 5ml of fresh growth medium was added to a sterile universal. The cryopreserved cells were removed from the liquid nitrogen and thawed at 37°C quickly. The cells were removed from the vials and transferred to the aliquoted media. The resulting cell suspension was centrifuged at 170*g* for 5 minutes. The supernatant was removed and the pellet resuspended in fresh culture medium. An

assessment of cell viability on thawing was then carried out (Section 2.1.5). Thawed cells were then added to an appropriately sized tissue culture flask with a suitable volume of growth medium and allowed to attach overnight. The following day, flasks were re-fed with fresh media to remove any non-viable cells.

2.1.8 Monitoring of sterility of cell culture solutions

Sterility testing was performed in the case of all cell culture media and cell culture related solutions. Samples of prepared basal media were incubated at 37°C for a period of seven days. This ensured that no bacterial or fungal contamination was present in the media or the solutions.

2.1.9 Serum batch testing

To prevent batch to batch variation (a common problem in foetal calf serum (FCS) in cell culture), a range of FCS batches were screened and the most suitable was chosen for a block of work (Sigma, F7524). Screening involved growing cells in 96-well plates and growth was recorded as a percentage of growth of a serum with known acceptable growth rate.

2.1.10 *Mycoplasma* analysis of cell lines

Mycoplasma testing was carried out in house. Both direct and indirect methods were used. In the indirect method, *Mycoplasma* negative NRK (Normal rat kidney fibroblast) cells were exposed to conditioned media of the test sample. Hoechst staining specific for DNA shows positive for *Mycoplasma* when viewed in the extracellular spaces. In the direct method, conditioned media were incubated on cells specifically to encourage *Mycoplasma* growth.

2.2 Conditioned Media Sample Preparation.

Cells ($3x10^6$) were seeded in four biological replicates in T-175cm^2 flasks. Cells were allowed to grow until 50-60% confluent. Cells were washed three times with 10mL of serum-free (SF) basal media and then incubated in 15mL of SF basal media for 60 min. After this time, cells were washed again twice with SF basal media. 15mL of SF basal media was added to the cells and incubated for 72hrs. After such time, conditioned media (CM) was collected, centrifuged for 15min at 250*g*, and stored at –80°C.

To concentrate conditioned media samples, 10ml was added to a Sartorius VivaSpin 20 with a 5,000 molecular weight cut-off and centrifuged at 4°C at 3220*g*. Samples were concentrated down to 200µl before undergoing a clean-up using Bio-Rad ReadyPrep 2-D Cleanup kit to remove substances like ionic detergents, salts, nucleic acids and lipids which are known to interfere with iso-electric focusing (IEF). Proteins were precipitated using ice-cold acetone and the subsequent protein pellets were resuspended in lysis buffer (20 mM Tris, 7 M Urea, 2 M Thiourea, 4% CHAPS, pH 8.5) prior to labelling.

2.3 Protein Quantification

Protein levels were determined using the Bio-Rad protein assay kit (Bio-Rad, 500-0006) as follows. A 2mg/ml bovine serum albumin (BSA) solution (Sigma, A9543) was prepared freshly in lysis buffer. A protein standard curve (0, 0.2, 0.4, 0.6, 0.8 and 1.0mg/ml) was prepared from the BSA stock with dilutions made in lysis buffer. The Bio-Rad reagent was diluted 1:5 in UHP water. A 20µl volume of protein standard dilution or sample (diluted 1:10) was added to 980µl of diluted dye reagent and the mixture vortexed. All samples were assayed in triplicate. After 5

minutes incubation, absorbance was assessed at 570nm. The concentration of the protein samples was determined from the plot of the absorbance at 570nm versus the concentration of the protein standard.

2.4 2D-DIGE Sample Preparation

2.4.1 Preparation of CyDye DIGE fluor minimal dye stock solution (1 nmol/µl)

The three CyDye DIGE Fluor Minimal dyes (Cy3, Cy5 and Cy2 (Amersham, 25-8010-65) were thawed from –20 ^{o}C to room temperature for 5 minutes. To each microfuge tube dimethylformamide (DMF) (Aldrich, 22,705-6) was added to a concentration of 1 nmol/ µl. Each microfuge tube was vortexed vigorously for 30 seconds to dissolve the dye. The tubes were then centrifuged for 30 seconds at 18,000*g* in a microcentrifuge. The reconstituted dyes can be stored at –20 ^{o}C for up to two months.

2.4.1.1 Preparation of 10 µl working dye solution (200 pmol/µl)

On thawing, the dye stock solutions were centrifuged in a microcentrifuge for 30 seconds. To make 10 µl of the three working dye solutions, 8 µl of DMF was added to three fresh eppendorfs labelled Cy2, Cy3 and Cy5. 2µl of dye stock was added to the 8µl DMF for a 0.2 nmol/µl concentration of each of the reconstituted dye stock solutions was added to their respective tubes. The dyes can be stored at -20 ^{o}C in tinfoil in the dark for 3 months.

2.4.2 Protein sample labelling

A volume of the protein samples equivalent to 50 µg was placed into eppendorf tubes. An eppendorf tube for the Cy2 pool made up from aliquots from each of the protein samples contained enough protein for 50µg for each gel. Each tube was

mixed by vortexing, centrifuged and then left on ice for 30 minutes in the dark. To stop the reaction, 1 µl of 10 mM lysine was added, the tubes were vortexed, centrifuged briefly and left on ice for 10 minutes in the dark. The labelled samples were stored at –80 ^{o}C. To this tube 1 µl of working dye solution was added.

2.4.3 Preparing the labelled samples for the first dimension

The protein samples labelled with Cy2 (pooled internal standard), Cy3 and Cy5 were thawed on ice (in the dark), combined by placing into a single eppendorf tube and mixed. An equal volume of 2X sample buffer (2.5 ml rehydration buffer stock solution (8M urea, 4 % CHAPS), pharmalyte broad range pH 4-7 (2%) (Amersham, 17-6000-86), DTT (2%) (Sigma, D9163)) was added to the labelled protein samples. The mixture was left on ice for at least 10 minutes then applied to Immobiline DryStrips for isoelectric focussing.

2.5 First dimension separation - isoelectric focussing methodologies

Isoelectric focussing of all samples was carried out using immobiline pH gradient (IPG) strips.

2.5.1 Strip rehydration using Immobiline DryStrip reswelling tray

The protective lid was removed from the Immobiline Dry Strip Reswelling tray. The tray was levelled using the spirit level. A 350µl volume of rehydration buffer solution (with 2 % pharmalyte broad range pH4-7 and 2 % DTT) was slowly pipetted into the centre of each slot; all air bubbles generated were removed. The cover film from the IPG strip (Amersham, 17-1233-01) was removed and positioned with the gel side down and lowered. To ensure the entire strip was

evenly coated the strip was gently lifted and lowered onto the entire surface of the solution avoiding trapping air bubbles.

Each strip was overlaid with about 3 ml IPG Cover Fluid (Amersham, 17-1335-01) starting on both ends of the strip, moving to the centre. The protective lid was then replaced and the strips were left at room temperature to rehydrate overnight (or at least 12 hours).

2.5.2 Isoelectric focussing using the IPGphor manifold

Following the rehydration procedure, the Manifold (Amersham) was placed onto the IPGphor unit by inserting the "T" shape into the hollow provided. A 9 ml volume of Cover Fluid was placed into each of the twelve lanes in the tray in order to cover the surface. Two wicks (Amersham, 80-6499-14) per strip were placed on tinfoil and 150 µl of UHP was pipetted onto each one to rehydrate them. The rehydrated strips were placed in the correct orientation (+ to anode) and aligned just below the indented mark, to allow for the wicks to overlap the strip. The rehydrated wicks were then placed over both the cathodic and anodic ends of all the strips. The wicks were checked to ensure they were positioned over the gel portion of the strip and avoiding the indent in the lane so as to guarantee a good contact with the electrodes. The sample cups (Amersham, 80-6498-95) were then positioned approximately 1 cm from the cathodic end of the strip and an insertion tool was used to securely "click" the cups into place. The electrodes were then fitted with their "Cams" open and in direct contact with the wicks.

The amount of protein loaded per strip was 150ug for DIGE or 400µg for spot picking. The protein samples were prepared by centrifuging to remove any

insoluble material and the appropriate volume was loaded with a pipette tip placed just beneath the surface of the cover fluid. The cover of the IPGphor unit was closed and the desired programme selected. The temperature was set for 20^{o}C with 50μA/strip. The IEF parameters are as follows: step 1: 300 volts for 3 hours (step-and-hold), step 2: 600 volts for 3 hours (gradient), step 3 1000 volts for 3 hours (gradient), step 4: 8000 volts for 3 hours (gradient). The IEF was left at 8000 volts (step-and –hold) until ready for SDS-PAGE step. On completion of the IEF run, the strips were drained of the cover fluid and stored in glass tubes at –80^{o}C or used directly in the second dimension.

2.6 Second Dimension – SDS polyacrylamide gel electrophoresis

2.6.1 Casting gels in the ETTAN Dalt-12 gel caster

The 12.5 % acrylamide gel solution was prepared in a glass beaker (acrylamide/bis 40 %, 1.5 M Tris pH 8.8, 10 % SDS). Prior to pouring, 10 % (w/v) ammonium persulfate and 50μl neat TEMED were added.

Two types of plates were used, low fluorescent for DIGE experiments and hinged for preparative and screening silver stained gels. All plates (both normal hinged and low fluorescent) and casting equipment were inspected to ensure they were clean. The gel caster frame was placed on a level bench leaning on its “legs” so that the back of the caster was open and facing the operator. The plates were assembled so that the front and back plates were evenly aligned and all seals and hinges in place. A thin spacer was placed in the gel caster unit followed by an assembled plate followed by a thin spacer then another plate. The plates were positioned in the caster unit so that the lower, front plate was the furthest away from the operator and the spacers packed with their curved edges to the top. This layering was repeated

until all 14 plates and spacers were in place. All plates and spacers were checked to ensure they were packed tightly together so as to minimise any gaps and air pockets. If all 14 gels were not required, up to 4 dummy plates could be substituted instead of the glass plates. When the desired amount of plates had been added, the thicker spacers were placed next to bring the level marginally over the edge of the back of the caster. The backing plate was then added to the caster frame and screwed into place with the 6 screws provided. The silicone tubing was added to the outlet of the glass beaker and the glass tube was inserted to the other end of the silicone tubing. The glass tube was inserted into the inlet of the reservoir and the glass beaker containing the gel solution was then clamped to a retort stand. The gel solution was held in place using arterial clamps on the top tube and the tube running down from the reservoir to the caster chamber. The top tube was unclamped and the gel solution was allowed to fill the tubing and the reservoir drain. Air bubbles that had been generated were dislodged by flicking the tube. When all air bubbles had been removed the bottom clamp was released allowing the gel solution into the gel caster. When the gel solution reached the indicator line across the top of the caster, the bottom and top tubes were reclamped. The displacement solution (0.375M Tris-Cl 1.5M pH8.8, 30% glycerol, UHP and bromophenol blue) was added to the reservoir and the glass tubing was slowly removed from the reservoir inlet. The clamp was removed from the bottom tube allowing the displacement solution into the tube and forcing the remaining gel solution into the gel caster. The gels were overlaid with 1 ml saturated butanol or sprayed with 0.1 % SDS solution. The gels were left to set for at least three hours at room temperature. Following this, the caster was gently unlocked and the gels removed and rinsed with distilled water. If

the gels were not used immediately they were stored for up to four days in 1X running buffer at 4 ^{o}C.

If gels were to be used for "spot picking" the plates were silanised to stick the acrylamide mixture to the plates. A volume of 2ml of (8ml ethanol, 200µl glacial acetic acid, 10µl bind-silane and 1.8ml UHP) was pipetted over the glass plate and wiped over with a lint free cloth. This was left to air dry for 15 minutes, after which 2ml ethanol and 2 ml UHP were each pipetted over the plate and wiped off respectively. The plate was left to air dry for approximately 1 hour 30 minutes.

2.6.2 Preparing the ETTAN DALT 12 electrophoresis unit

The electrophoresis chamber was prepared by adding 6.48 litres of UHP and 720 ml of 10X SDS running buffer. The pump was then turned on to cool the system to 10 ^{o}C.

2.6.3 Equilibration and loading of focussed Immobiline DryStrips

The SDS equilibration buffer (30% glycerol, 6M urea, 50nm 1.5M Tris-Cl pH 8.8, 2% SDS, bromophenol blue and UHP) which had been prepared, aliquotted into 30 ml volumes and frozen at –20^{o}C was allowed to thaw to room temperature. Two SDS equilibration buffer solutions with DTT (65 mM) or iodoacetamide (240 mM) (Sigma, I1149) were then prepared. Using a forceps, the IPG strips* were removed from the IPGphor unit, the cover fluid was drained off by holding the strips at an angle and they were placed into individual glass tubes with the support film toward the wall. Equilibration buffer (10 mls containing DTT) was added to each tube and incubated for 15 minutes with gentle agitation using an orbital shaker. During this equilibration step, the gel cassettes were rinsed with UHP and then the tops rinsed

with 1X running buffer. After the first equilibration, DTT containing equilibration solution was removed and 5 mls of the iodoacetamide containing equilibration buffer added. The strips were incubated for 10 minutes with gentle agitation. During this equilibration step, the agarose overlay solution (0.5% agarose in running buffer) was prepared and 50 ml of 1X running buffer was placed in a glass tube.

*If the strips had been frozen at this stage they were left at room temperature to thaw before the DTT-containing equilibration solution was added.

Using a forceps and holding the anode end, the IPG strips were rinsed in 1X SDS electrophoresis running buffer and placed between the two glass plates of the gel. The strip was pushed down gently using a thin plastic spacer until it came in contact with the surface of the gel. Any air bubbles trapped between the gel surface and the strip were gently removed. Approximately 1 ml of the 0.5 % agarose overlay solution was applied over the IPG strip to seal it in place.

2.6.4 Inserting the gels into the Ettan DALT 12 electrophoresis buffer tank

When the running buffer reached the desired temperature (10^{o}C) the loaded gel cassettes were wetted with UHP and inserted into the tank through the slots provided in the same orientation. When all 12 slots were filled the upper chamber was filled, 2X running buffer was added to the upper chamber until the mark on the side of the chamber was reached. The cover of the unit was replaced and the required running conditions selected. The unit was run for 18–24 hours at 1.5 Watts per gel at 10°C or until the bromophenol blue dye front reached the bottom of the gel. When the run was completed, the gel cassettes were removed from the tank one

at a time using the DALT cassette removal tool and rinsed with UHP to remove the running buffer.

2.7 Method for scanning DIGE labelled samples

The Typhoon Variable Mode Imager (GE Healthcare) was turned on and left to warm up for 30 minutes prior to scanning. The scanning control software was opened and the fluorescence mode was selected. The appropriate emission filters and lasers were then selected for the separate dyes (Cy2 520 BP40 Blue (488), Cy3 580 BP30 Green (532) and Cy5 670 BP 30 Red (633)). The first gel was placed in the scanner and pre-scanned at a 1000 pixel resolution in order to obtain the correct photo multiplier tube (PMT) value (to prevent saturation of the signal from high abundant spots). Once the correct PMT value was found, the gel was scanned at 100 pixel resolution, resulting in the generation of three images, one each for Cy2, Cy3 and Cy5. Once the scanning was completed, the gel images were imported into the ImageQuant software. All gels were cropped identically to facilitate spot matching in the Decyder BVA module.

2.8 Analysis of gel images

2.8.1 Differential in-gel analysis (DIA)

The DIA module processes a triplet of images from a single gel. The internal standard is loaded as the primary image followed by the secondary and tertiary image, derived from, for example, a control and treated sample. Spot detection and calculation of spots properties were performed for each image from the same gel. The software determined the margins of the spots, quantified the spot intensities and calculated the relative spot intensity as the ratio between the total intensity of the

gel and the intensity of each individual spot. The protein spots were then normalised using the in-gel linked internal standard. The data from the first gel was XML formatted and exported into the Biological Variation Analysis (BVA) software for further analysis. This procedure was repeated for each gel in the experiment.

2.8.2 Biological variation analysis (BVA)

Once all gels from the experiment were loaded into the BVA module, the experiment design was set up and the images were assigned into three groups (standard, control and treated). The spots on the gels were then matched across all gels in the experiment.

This module detects the consistency of the differences between samples across all the gels. The software standardises the relative spot intensity of the Cy5 image to that of the Cy3 image in the same gel. The standardised spot intensity was then averaged across the triplicate gels. The BVA module detected the consistency of the differences between samples across all the gels and applied statistics to associate a level of confidence for each of the differences. The protein spots with statistically significant protein expression changes were designated "proteins of interest" and placed in a pick list.

Preparative gels for spot picking with 400µg of protein/gel were focussed and run out on SDS-PAGE gels. The gels were then stained with colloidal coomassie (section 2.9). Spots that showed differential protein expression were picked with the ETTAN Spot Picker (section 2.10).

2.9 Brilliant blue G Colloidal Coomassie staining of preparative gels for spot picking

After electrophoresis, the smaller lower plates with the gels attached were placed in the gel boxes containing fixing solution (7% (v/v) glacial acetic acid in 40% (v/v) methanol (Aldrich, 200-659-6)) for at least one hour. During this step a 1X working solution of Brilliant Blue G colloidal coomassie (Sigma, B2025) was prepared by adding 800ml UHP to the stock bottle. When the fixing step had nearly elapsed, a solution containing 4 parts of 1X working colloidal coomassie solution and 1 part methanol was made, mixed by vortexing for 30 seconds and then placed on top of the gels. The gels were left to stain for 2 hours. To destain, a solution containing 10% acetic acid in 25% methanol was poured over the shaking gels for 60 seconds. The gels were then rinsed with 25% methanol for 30 seconds and then destained with 25% methanol for 24 hours. The glass surface was dried and two reference markers (Amersham) attached to the underside of the glass plate before scanning. The resulting image was imported into the ImageMaster software (Amersham) and the spots were detected, normalised and the reference markers selected. While keeping the shift key depressed, all spots of interest were manually selected. The resulting image was saved and exported into the Ettan Spot Picker software.

2.10 Spot picking

The stained gel was placed in the tray of the Ettan Spot Picker (Amersham, 18-1145-28) with reference markers (Amersham, 18-1143-34) aligned appropriately and covered with UHP. The imported pick list was opened, the syringe primed and the system was set up for picking the spots from the pick list. The spots were

robotically picked and placed in 96-well plates, which were stored at 4^{o}C until spot digestion.

2.11 Spot digestion and identification with MALDI-TOF

The 96-well plate was placed in the Ettan Digester (Amersham, 18-1142-68) to digest the protein as follows: Step 1 – the gel plugs were washed three times for 20 minutes each with 50µl 50mM ammonium bicarbonate (Sigma, A6141) in 50% methanol. Step 2 – the gel plugs were washed twice for 15 minutes with 50µl 70% acetonitrile (Sigma, 34967). The gel plugs were left to dry for at least 60 minutes. After drying, the individual gel pieces were rehydrated in 10µl digestion buffer (12.5ng trypsin (Promega, V5111) per µl of 10% acetonitrile, 40mM ammonium bicarbonate). Exhaustive digestion was carried out overnight at 37^{o}C. After digestion, the samples were transferred as follows: Step 1 – A volume of 40µl of 0.1% trifluoroacetic acid (Sigma, 302031) in 50% acetonitrile was added to the wells, mixed and left for 20 minutes. A volume of 60µl of this solution was transferred to a fresh 96-well plate. Step 2 - A volume of 30µl of 0.1% trifluoroacetic acid in 50% acetonitrile was added to the wells, mixed and left for 20 minutes. A volume of 50µl of this solution was transferred to the fresh 96-well plate. The liquid in the plate was vacuum-dried in a maxi dry plus. After drying, the 96-well plate was placed in the Ettan Spotter (Amersham, 18-1142-67) for spotting onto the target plates. A volume of 3µl of 0.5% trifluoroacetic acid in 50% acetonitrile was added to the desiccated peptides and mixed 5 times. A volume of 0.3µl of this mixture was spotted onto the target plate after which a volume of 0.3µl matrix solution [7.5mg/ml α-cyano-4-hydroxycinnamic acid (LaserBio labs, 28166-41-8) in 0.1% trifluoroacetic acid in 50% acetonitrile].

The target plate was placed in the MADLI-ToF (Amersham, 11-0010-87) instrument. The system was set up as follows: the target plate was disengaged from the machine. A new empty run list was opened. In the acquisition mode of the "favorites icon, the "spectrum processes", specifically PepMix 4 (LaserBio labs, C104) was picked, and "protein digest optimised" were successively selected and dragged to positions 1 and 2-24 respectively on the target slide. Within the identification section of favorites, "protein digest Homo sapiens IAA" was selected and dragged to positions 2-24 on the target slide. The run list for slide 1 was saved and associated to position 1 on the previously disengaged tray. Selecting process and play then started the MALDI.

Mass spectra were recorded operating in the positive reflector mode at the following parameters: accelerating voltage 20 kV; and pulsed extraction: on (focus mass 2500). Internal and external calibration was performed using trypsin autolysis peaks at 842.509 m/z, 2211.104 m/z and PepMix 4 respectively. Calibration using Pep4 was performed as follows: Once two spectra were generated for the PepMix 4 mix (position 1 on the slide), the acquisition of spectra was stopped. The first spectrum of sample one was selected and the calibrant peaks readjusted for accuracy. The five individual peaks cover the 500-3500 Da mass range and include bradykinin fragment 1-5 (573.315), angiotensin II human (1046.5424), neurotensin (1672.9176) and insulin B chain oxidised (3494.6514). Once calibration was completed it was saved as the new "system calibration". The MALDI was then restarted. The mass spectra generated for each of the proteins were analyzed using MALDI evaluation software (Amersham Biosciences). Protein identification was achieved with the PMF Pro-Found search engine for peptide mass fingerprints.

2.12 Identification of proteins with LC–MS/MS

Tryptic digest of proteins (for those which were more than 3 fold differentially expressed) were also analysed by 1-D LC-MS using the Ettan™ MDLC system (GE Healthcare) in high-throughput configuration directly connected to a Finnigan™ LTQ™ (Thermo Electron). Samples were concentrated and desalted on RPC trap columns (Zorbax™ 300SB C18, 0.3 mm65 mm, Agilent Technologies) and the peptides were separated on a nano-RPC column (Zorbax 300SB C18, 0.075 mm 6100 mm, Agilent Technologies) using a linear ACN gradient from 0 to 65% acetonitrile (Sigma, 34967) over 60 min. All buffers used for nano-LC separation contained 0.1% formic acid (Fluka, 94318) as the ion pairing reagent. Full scan mass spectra were recorded in profile mode and tandem mass spectra in centroid mode. A scan time of ~0.15 s (one microscan with a maximum ion injection time of 10ms) over an m/z range of 300-2000 was used followed by MS/MS analysis of the 3 most abundant peaks from each scan which were then excluded for the next 60 seconds followed by MS/MS of the next three abundant peaks which in turn were excluded for 60 seconds and so on. A "collision energy" setting of 35% was applied for ion fragmentation and dynamic exclusion was used to discriminate against previously analysed ions (data dependent analysis).

Protein identification search was performed using the Turbo-SEQUEST algorithm in the BioWorks 3.1 software package (Thermo Electron) and the Swiss-Prot human database (Swiss Institute of Bioinformatics, Geneva, Switzerland). The identified peptides were further evaluated using charge state versus cross-correlation number (XCorr). The criteria for positive identification of peptides was XCorr > 1.5 for singly charged ions, XCorr > 2.0 for doubly charged ions, and

XCorr > 2.5 for triply charged ions. The distinct peptides with p-value ≤0.05 were considered for protein identification and also had at least 2 peptides.

2.13 Western blot Analysis

2.13.1 Gel electrophoresis

Proteins for analysis by Western blotting were resolved using 12% NuPAGE Bis-Tris Gels (Invitrogen, NP0341BOX) in XCell SureLock™ Mini-Cell (Invitrogen, EI0001) running instrument. Western blotting samples (2μg/μL) were prepared by diluting samples with water and then an equal volume of 2X loading buffer (Sigma, S3401). 10-40μg of protein and 5μL of molecular weight marker (ISIS, P7708S) were loaded onto gels. The samples were electrophoretically separated at 200V and 45mA using a MOPS/SDS buffer (MOPS 1M (Fluka, 69949), 1M Tris base (Sigma, T8404), 2% SDS (Sigma, 23771), 20mM EDTA, pH 7.7 (Fluka, 03609)), until the bromophenol blue dye reached the bottom of the gel.

2.13.2 Western blotting

Once electrophoresis had been completed, the gel was equilibrated in transfer buffer (25mM Tris (Sigma, T8404), 192mM glycine (Sigma, G-7126), pH 8.3-8.5) for approximately 30 minutes. Five sheets of Whatman 3mm filter paper (Whatman, 1001824) were soaked in freshly prepared transfer buffer. These were then placed on the cathode plate of a semi-dry blotting apparatus (Bio-rad). Air pockets were then removed from between the filter paper. Nitrocellulose membrane (GE Healthcare, RPN3032D), which had been equilibrated in the same transfer buffer, was placed over the filter paper on the cathode plate. Air pockets were once again removed. The gels were then aligned onto the membrane. Five additional sheets of

transfer buffer-soaked filter paper were placed on top of the gel, all air pockets removed and excess transfer buffer removed from the cathode plate. The proteins were transferred from the gel to the membrane at a current of 34mA at 15V for 30-40 minutes, until all colour markers had transferred. Following protein transfer, membranes were stained using Ponceau (Sigma, P7170) to ensure efficient protein transfer. The membranes were then blocked overnight using 1-5% Marvel (Cadburys; Marvel skimmed milk) in PBS at 4°C. The membranes were washed with PBS prior to the addition of the primary antibody. Membranes were incubated with primary antibody overnight at 4°C. Antibodies were prepared in 1-5% Marvel in PBS at recommended dilutions. The membranes were then rinsed 3 times with PBS containing 0.5% Tween 20 (Sigma P1379) for a total of 15-30 minutes. Relevant secondary antibody (1/1000 dilution of anti-mouse (Dako Cytomation, P0260) or anti-rabbit (Dako Cytomation, P0448) or anti-goat (Santa Cruz Biotechnology, Sc2098) IgG peroxidase conjugate in 2-5% Marvel-TBS) was added for 1 hour at room temperature. The membranes were again washed three times thoroughly in TBS containing 0.5% Tween for 15 minutes.

2.14 Enzyme-Linked Immunosorbent Assay (ELISA)

The ELISAs performed for validation of lung cancer biomarker targets were sandwich ELISAs. These 96-well plates come with capture antibody precoated to the surface of the wells. Protein standards that come with the kit were prepared and along with serum samples at a kit specific dilution were added to the wells, with the standards in duplicate and a proportion of the serum samples in duplicate and triplicate. After incubation the plate was washed to remove any unbound antigen. An enzyme-linked detection antibody was added which binds specifically to the

antigen. After another incubation period the plate was washed to remove any unbound detection antibody. A substrate was added that is converted by the enzyme into a colour or fluorescent signal. This colour or fluorescence was measured at a wavelength recommended in the ELISA manual to determine the presence and quantity of the antigen. From the standard curve produced from the standards of known concentrations, the concentration of antigen present in the serum sample can be determined and comparisons can be made between the various types of samples. Standard protocols were followed for each ELISA.

2.15 Biochemical Fluorometric Assays

Biochemical assays involve a single reaction step. A standard curve was created to calculate concentrations in unknown samples. A reaction mix was created that contains assay buffer, a developer, and an enzyme mix, and was added to all standards and all unknown serum/plasma samples. These were then incubated in the dark (generally for a relatively short period e.g. 20-30min) and OD or fluorescence was measured in a microplate reader. Background is corrected by subtracting the zero standard/control from all sample readings.

2.16 Pyruvate Kinase Activity Assay

A pyruvate kinase assay kit from BioVision was used to measure the pyruvate kinase activity in serum samples. In the assay, phosphoenolpyruvate (PEP) and adenosine diphosphate (ADP) are catalysed by pyruvate kinase to generate pyruvate and adenosine triphosphate (ATP). The generated pyruvate is oxidised by pyruvate oxidase to produce a colour at 570nm and fluorescence at 535/587nm. Since the increase in colour or fluorescent intensity is proportional to the increase in pyruvate

amount, the pyruvate kinase activity can be accurately measured. One unit of pyruvate kinase activity is the amount of enzyme transfer of phosphate group from PEP to ADP, yielding 1.0 µmol of pyruvate per minute at 25ºC.

2.17 Luminex Multiplex Bead-based Assays

Cancer Panel 1, Cancer Panel 2, and the MMP Panel of biomarkers were performed according to Millipore Milliplex Map Kit protocol (Millipore). Serum/plasma samples were diluted as recommended with calibrator diluent. Initially, the 96-well filter bottom plate was prewet adding 200 µl assay buffer. The microparticles from a particular panel were pooled e.g. for discovery phase MMP-1, MMP-3, MMP-8, MMP-9, MMP-10, and MMP-13 were pooled together. 25µl of diluted microparticle solution and 25µl of sample and standards and controls were added to the necessary wells. 25µl assay buffer was added to all wells. Next the plate was incubated overnight at 4 ºC and next day washed three times with wash buffer. Afterwards, 50 µl of diluted biotin antibody was added to each well and incubated for 1 h. The plate was then washed as described above and 50µl of diluted Streptavidin-PE was added to each well and incubated for 30min. All incubations were performed at room temperature on an orbital shaker set at 15 *g*. Finally, the plate was washed again with 100 µl of Sheath Fluid. The median relative fluorescence units were measured using the Luminex 100 analyzer (Luminex, Austin, TX, USA).

2.18 Immunodepletion and ProteoMiner

Immunodepletion of serum was performed using ProteoPrep 20 Plasma Immunodepletion kit from Sigma. 8µl of serum was diluted to 100µl with 1x

Equilibration buffer and filtered through a Corning Spin-X centrifuge Tube Filter before being added to the immunodepletion column and following a standard protocol.

ProteoMiner was performed using ProteoMiner Protein Enrichment Kits from Bio-Rad. The small capacity kit was used in which 200µl of serum was added to the column. After following a standard protocol the elutions were pooled together.

2.19 Label-Free Mass Spectrometry Sample Preparation

20µg of protein (See Section 2.3) from the sample of interest was placed into an eppendorf. The 2D-Clean-up kit (Bio-rad) was used to precipitate the protein, which was then left overnight in acetone. The protein pellet was resuspended in 6M Urea, 2M Thiourea, 10mM Tris, pH 8. The sample was sonicated and vortexed to make sure of complete suspension. A protein assay was performed on this sample and 10ug of protein was transferred into a new eppendorf. 1ul of reduction buffer was added to the sample (5mM DTT in water), vortexed and incubated at 37ºC for 30 minutes. 1µl of alkylation buffer was added to the sample (25mM Iodoacetamide in 50mM Ammonium Bicarbonate), vortexed and incubated for 20 minutes at room temperature in the dark. Lys-C was added at a ratio of 1:50 (enzyme:protein, in this case 0.2ug of Lys-C) and incubated for 4 hours at 37ºC. Samples were diluted with 4 volumes of 50mM Ammonium Bicarbonate.

Trypsin was added at a ration of 1:25 (enzyme:protein, in this case 0.4ug of Trypsin) and incubated overnight at 37ºC. The following day, PepClean™ C-18 Spin Columns from Thermo were used to purify the peptides. 3 parts sample was mixed with 1 part Sample Buffer (2% TFA in 20% ACN). The final sample contained 0.5% TFA in 5% ACN. The column was tapped to settle resin. Top and

bottom cap were removed. The column was placed into a receiver tube. 200 µl of Activation Solution (50% Acn) was added to rinse walls of the spin column and to wet resin and centrifuged at 1,500 x g for 1 minute. Flow-through was discarded and this step was repeated. 200ul Equilibration Solution (0.5% TFA in 5% ACN) was added. Centrifuge at 1,500 x g for 1 minute. Flow-through was discarded and this step was repeated. The sample was loaded on top of the resin bed. The column was placed into a receiver tube and centrifuged at 1,500 x g for 1 minute. To ensure complete binding, flow-through was recovered and step was repeated 3 times. The column was placed into a receiver tube. 200 ul Wash Solution (0.5% TFA in 5% ACN) was added to column and centrifuged at 1,500 x g for 1 minute. Flow-through was discarded and this step was repeated four times. The column was placed in a new receiver tube. 20 ul of Elution Buffer (80% ACN) was added to the top of the resin bed and centrifuged at 1,500 x g for 1 minute. This was repeated 3 times with same receiver tube. The sample was gently dried in a vacuum evaporator and the peptides were re-suspended in 50ul of 0.1% TFA in 2% Acn with vortexing and sonication. 5µl of this sample was taken up per LC run.

Progenesis LC-MS software was used for statistical analysis of LC produced data. See section 7.3 for a more detailed description of the software.

2.20 RNA interference (RNAi)

RNAi using small interfering RNAs (siRNAs) was carried out to silence the hnRNPA2B1 gene. The siRNAs used were chemically synthesised (Ambion Inc). These siRNAs were 21-23 bps in length and were introduced to the cells via reverse transfection with the transfection agent siPORT^{TM} *Neo*FX^{TM} (Ambion Inc., 4511).

2.20.1 Transfection optimisation

In order to determine the optimal conditions for siRNA transfection, optimisation with kinesin siRNA (Ambion Inc., 16704) was carried out for the DLKP-M cell line.

Cell suspensions were prepared at 2.5 X 10^5, 1.5 X 10^5, 1 X 10^5, 0.75 X 10^5, and 0.5 X 10^5 cells per well. Solutions of negative control and kinesin siRNAs at a final concentration of 30nM were prepared in optiMEM (Gibco^{TM}, 31985047). NeoFX solutions at a range of concentrations were prepared in optiMEM in duplicate and incubated at room temperature for 10 minutes. After incubation, either negative control or kinesin siRNA solution was added to each neoFX concentration. These solutions were mixed well and incubated for a further 10 minutes at room temperature. 100µl of the siRNA/neoFX solutions were added to each well of a 6-well plate. 1 ml of the relevant cell concentrations were added to each well. The plates were mixed gently and incubated at 37°C for 24 hours. After 24 hours, the transfection mixture was removed from the cells and the plates were fed with fresh medium. The plates were assayed for changes in proliferation at 72 hours using the acid phosphatase assay (Section 2.20.3). Optimal conditions for siRNA transfection were determined as the combination of conditions, which gave the greatest reduction in cell number after kinesin siRNA transfection and also the least cell kill in the presence of transfection reagent. Western blot analysis was used to establish the optimum conditions for a siRNA transfection. The optimised conditions for the cell lines are shown in Table 2.18.1.

Cell Line	Seeding Density per 6-Well Plate	NeoFx Volume per 6-Well Plate (µl)	siRNA Volume per 6-Well Plate (µl)
DLKP-M	$0.5x10^5$	1	0.2

Table 2.18.1: - Optimised conditions for siRNA transfection

2.20.2 siRNA functional analysis of hnRNPA2B1 in DLKP-M

Two pre-designed siRNAs were chosen for the protein/gene targets and transfected into cells. Two siRNAs were sufficient if a validated siRNA was available. Validated siRNAs have been verified by real-time RT-PCR to reduce gene expression of >70% 48 hours post-transfection. For each set of siRNA transfections carried out, control, non-transfected cells and a scrambled (SCR) siRNA transfected control were used. Scrambled siRNA are sequences that do not have homology to any genomic sequence. The scrambled non-targeting siRNA used in this study is commercially produced, and guarantees siRNA with a sequence that does not target any gene product. It has also been functionally proven to have no significant effects on cell proliferation, morphology and viability. For each set of experiments investigating the effect of siRNA, the cells transfected with target-specific siRNAs were compared to cells transfected with scrambled siRNA. This took account of any effects due to the siRNA transfection procedure, reagents, and also any random effects of the scrambled siRNA. Kinesin was used as a control to assess the efficiency of the siRNA transfection. Kinesin plays an important role in cell division; facilitating cellular mitosis. Therefore, transfection of siRNA kinesin resulted in cell cycle arrest and confirmed efficient transfection. Western blots

(section 2.13) were used to determine if siRNA had an efficient knock-down effect at a Protein-Level. From Applied Biosystems, siRNAs s6713 and s6715 were used.

2.20.3 Acid Phosphatase Assay

Following the incubation period of 6-7 days, media was removed from the plates. Each well on the plate was washed twice with 100µl PBS. This was then removed and 100µl of freshly prepared phosphatase substrate (10mM *p*-nitrophenol phosphate (Sigma 104-0) in 0.1M sodium acetate (Sigma, S8625), 0.1% (v/v) Triton X-100 (Sigma, X100), pH 5.5) was added to each well. The plates were then incubated in the dark @ 37°C for 2h. Colour development was monitored during this time. The enzymatic reaction was stopped by the addition of 50µl of 1N NaOH. The plate was read in a dual beam plate reader at 405nm with a reference wavelength of 620nm.

2.20.4 Proliferation assays on siRNA transfected cells

As described in table 2.18.1, cells were seeded using 1µl NeoFX to transfect 30nM siRNA in a cell density of 0.5 X 10^5 per well of a 6-well plate. After 24 hrs, transfection medium was replaced with fresh media and cells were allowed to grow until they reached 80-90% confluency, a total of 5 days. Cell number was assessed using the acid phosphatase assay (section 2.20.3). All experiments were carried out independently at least three times.

2.20.5 Invasion Assays

Matrigel (Sigma, E-1270) was diluted to a working stock of 1 mg/ml in serum free DMEM. Aliquoted stocks were stored at -20°C.

Invasion assays were set up using a Costar 24 well plate (Costar, 3524), containing 8.0μm pore sized inserts (BD Biosciences, 353097). For invasion (but not motility) assays, a layer of Matrigel (BD Biosciences, 354234) was dispensed into the insert. Matrigel was diluted to 1mg/ml in serum-free DMEM medium. A volume of 100μl of this Matrigel was dispensed into each insert, and the plate was incubated overnight at 4°C. The following day, excess Matrigel was removed from the inserts. Cells were harvested at a concentration of 0.5 X 10^6 cells/ml in media, and 100μl was added to each insert, followed by another 100μl of media. A volume of 500μl media was then added to each plate well. For DLKP-M, cells were incubated at 37°C for 24h.

Following this, the inner area of each insert was wiped with a cotton bud soaked in PBS, to remove any cells, while the outside of the insert was stained with 0.25% crystal violet. Staining of the inserts was sustained for a period of 10 minutes, followed by a gentle rinse, done in triplicate, in UHP and then allowed to dry. Quantification of the cells was achieved by doing a cell count per area per view at 20X magnification using a light microscope. 10 fields of view were counted per insert. A minimum of 2 inserts were used per sample tested.

2.20.6 Motility Assays

Motility assays were carried out in an identical manner to invasion assays, as described in section 2.20.5, with the exception that the inserts were not coated in Matrigel.

2.20.7 Invasion assays on siRNA transfected cells

Using the optimised conditions in Table 2.18.1, each of the siRNAs was tested to see changes in invasion of the cells after transfection. Two-three separate siRNAs were used for each target gene (Appendix I, Table 5). All siRNAs were purchased from Ambion Inc.

To assay for changes in invasive capacity, siRNA experiments in 6-well plates were set up using 2µl NeoFX to transfect 30nM siRNA at a cell density of 3 X 10^5 per well of a 6-well plate. Transfection medium was removed after 24 hours and replaced with fresh growth medium. The transfected cells were assayed for changes in invasion capacity at 48 hours using the *in vitro* invasion assay described in Section 2.5.2. All experiments were carried out independently at least three times.

2.20.8 Motility assays on siRNA transfected cells

Motility assays were carried out in an identical manner to invasion assays, as described in section 2.20.7, with the exception that the inserts were not coated in Matrigel.

2.21 Immunohistochemistry

Antibodies were optimised (starting with recommended dilution on product data sheet) for immunohistochemistry staining purposes prior to using commercial lung cancer tissue arrays. Antigen retrieval was performed using the Dako PT Link, with pH 6 buffer being the optimum pH buffer. Staining of the tissue slide was performed using the Dako AutoStainerPlus according to a standard protocol. Stained tissue slides were prepared for mounting by placing in 70% ethanol for two 3 minute periods, followed by 90% ethanol for two 3 minute periods and then 100%

ethanol for two 3 minute periods. Next the slides were placed in xylene for two 5 minute periods before adding Sigma DPX Mountant and covering with a glass slide. The slides could then be viewed under a microscope and scored by trained personnel.

2.22 Metabolomics

Serum samples (140µL) were shipped on dry-ice to Metabolon Inc. where the metabolomic profiling was performed.

Metabolon incorporates three independent complementary analysis platforms to maximise the number of small molecules and metabolites that the combined systems can identify and measure. Two separate ultra-high performance liquid chromatography / tandem mass spectrometry (UHPLC/MS/MS2) injections (one optimised for basic compounds, and the other for acidic compounds) and one GC/MS injection per sample are performed.

Firstly, small molecules were extracted from serum specimens using methanol to allow precipitation of proteins. The extract supernatant was then split into four equal aliquots; two for UHPLC/MS, one for GC/MS and one reserve aliquot. Aliquots were then dried overnight to remove solvent.

For the UHPLC methods one aliquot was reconstituted in 50µL 0.1% formic acid, and the other in 50µL 6.5mM ammonium bicarbonate pH 8.0. For GC/MS analysis aliquots were derivatized using equal parts N,O-bistrimethylsilyltrifluoroacetamide and a solvent mixture of acetonitrile:dichloromethane:cyclohexane (5:4:1) with 5% triethylamine at 60ºC for 1 hour. All reconstitution solvents contained instrument internal standards used to monitor instrument performance.

UHPLC/MS was carried out using a Waters Acquity UHPLC coupled to an LTQ mass spectrometer equipped with an electrospray ionization source. Two separate UHPLC/MS injections were performed on each sample. The acidic injections were monitored for positive ions and the basic injections were monitored for negative ions. The derivatized samples for GC/MS were analyzed on a Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole MS.

The resulting MS/MS^2 data was then searched against Metabolon's reference standard library. This library was generated from 1500 standards and contains the retention time/index, mass to charge (m/z), and MS/MS spectral data for all molecules in the library, including their associated adducts, in-source fragments, and multimers. The library allows identification of experimentally detected metabolites based on a multi-parameter match basis. All identifications and quantifications were subjected to QC to verify the quality of the identification and peak integration.

CHAPTER THREE

BIOMARKER DISCOVERY FROM THE CONDITIONED MEDIA OF LUNG CANCER CELL LINES

3.1 Background

The secretion and leaking of proteins is commonly associated with cancer. The process of cancer initiation and progression, and the signalling pathways involved, are not confined to the cancer cell itself but extend to the tumour-host interface, a dynamic environment in which information passes between the tumour cells and the normal host tissue (Kulasingam and Diamandis 2007).

It is conceivable these leaked and/or secreted proteins from the tumour and its microenvironment will ultimately make their way to the serum proteome, where their detection could create a proteomic fingerprint of the tumour of origin. Of all cellular proteins about 20-25% are secreted, so one can predict with confidence that proteins or their fragments originating from cancer cells or their microenvironment may eventually enter the circulation (Liotta, Ferrari et al. 2003).

3.2 Lung Cancer Cell Line Model

In this study lung cancer cell lines were used as a model. Lung cancer cell lines secrete and shed proteins into the surrounding media and this environment could have similarities with the tumour microenvironment or the serum proteome of a lung cancer patient. This media, without serum which can mask the secreted or shed proteins, is known as conditioned media. Conditioned media from sixteen lung cancer cell lines and a normal human bronchial epithelial cell line were prepared and used for 2D-DIGE experiments. The four main types of lung cancer can be broadly categorised as squamous cell carcinoma, adenocarcinoma, large cell carcinoma, and small cell carcinoma; four cell lines from each of these categories was used in this study.

Conditioned media samples were concentrated from 10ml to 200µl before undergoing a 2-D cleanup and resuspended in lysis buffer. Accurate protein quantification is essential before labelling with the CyDye DIGE fluor minimal dyes. In total thirty gels were ran, each with a Cy3 and a Cy5 labelled individual cell line as well as a Cy2 labelled pool of all cell lines for internal standardisation. All sixteen lung cancer cell lines and normal human bronchial epithelial cell line were examined in triplicate, with some of the cancer lines examined four or five times.

3.3 Decyder analysis of 2D-DIGE conditioned media

All gels were scanned on a Typhoon Variable Mode Imager. The use of appropriate emission filters and lasers allowed the scanning of the three dyes per gel. The images are then cropped and imported into DeCyder for analysis. As per Materials and Methods section the DeCyder software allows quantitation of the protein spots, using the Cy2 internal pool to standardise spot intensity across all gels.

Protein spots are picked from Coomaisse stained preparative gels and identified via mass spectrometry, the subsequent identifications then entered into DeCyder where spots can be matched with confidence. DeCyder allows statistical comparison of all protein spots between all conceivable comparisons of each of the designated sample groups. Lists were generated of all interesting proteins with t-test and 1-WAY ANOVA scores of ≤ 0.1. The cut-off of 0.1 instead of the standard 0.05 was used due to potential inaccuracies of this discovery phase technique. Conditioned media itself is only a model for the tumour cell environment so can not be examined with full confidence. A total of 83 identifications were recorded in the DeCyder experiment, though a number of these were duplicate identifications.

Protein ID	Fold Change	ANOVA
Pyruvate kinase 3	4.88	0.0021
Stathmin	4.81	0.0049
KNP-I beta protein	4.1	0.0028
L-lactate dehydrogenase A chain	2.67	0.018
Fumarylacetoacetate hydrolase	2.59	0.0041
Ubiquitin carboxyl-terminal esterase L1	2.48	0.0013
Chain A; Cyclophilin B	2.37	0.028
Malate dehydrogenase; mitochondrial	2.27	0.049
Proteasome activator complex subunit 1	2.19	0.0082
Creatine kinase B-type	2.14	0.0091
Rho GDP dissociation inhibitor (GDI) alpha	1.99	0.0088
Fumarate hydratase	1.86	0.00092
Triosephosphate isomerase (TIM)	1.83	0.0002
Inorganic pyrophosphatase	1.78	0.054
WDR1 protein	1.65	0.039
Protein disulfide-isomerase A6	1.56	0.006
Proteasome subunit beta type-7	1.47	3.70E-05
Peroxiredoxin-2	1.43	0.0097
Acetyl-CoA acetyltransferase	1.35	0.0019
Phosphoglycerate kinase 1	1.29	0.046
Phosphatidylethanolamine-binding protein 1	1.28	0.014
TKT protein	1.22	0.022
Peroxiredoxin 1	1.09	0.0069
L-lactate dehydrogenase A chain	1.07	0.052
PARP1 protein	1.05	0.016
Pigment epithelium-derived factor	1.05	0.015
Malate dehydrogenase; cytoplasmic	-1.01	0.018
Synaptic vesicle membrane protein VAT-1 homolog	-1.2	0.015
Chain B; Human Platelet Profilin	-1.43	0.015
Adenosylhomocysteinase	-1.66	0.06
Aspartate aminotransferase; cytoplasmic	-1.75	0.037
Heat shock protein beta-1	-3.4	0.0014

Table 3.3.1: Lung cancer versus Normal. Statistically significant proteins identified from the 2D-DIGE experiments on conditioned media. A positive fold change indicates an increase in lung cancer and a negative fold change indicates a decrease in lung cancer for this protein compared to normal.

Protein ID	Fold Change	T-test
Pyruvate kinase 3	6.19	0.043
WDR1 protein	2.89	0.056
KNP-I beta protein	2.7	0.029
Hsc70-interacting protein	2.69	0.045
Inorganic pyrophosphatase	1.79	0.042
Heat shock protein beta-1	-2.48	0.017

Table 3.3.2: Large cell lung cancer versus Normal. Statistically significant proteins identified from the 2D-DIGE experiments on conditioned media. A positive fold change indicates an increase in large cell lung cancer and a negative fold change indicates a decrease in large cell carcinoma for this protein compared to normal.

Protein ID	Fold Change	T-test
Stathmin	6.55	0.04
Pyruvate kinase 3	4.82	0.015
KNP-I beta protein	4.19	0.0057
Hsc70-interacting protein	3.23	0.057
L-lactate dehydrogenase A chain	2.89	5.50E-05
Elongation factor 1-alpha 2	2.51	0.01
Triosephosphate isomerase (TIM)	2.18	2.90E-05
Protein DJ-1	1.95	0.049
Chain A; Crystal Structure Of Human Dj-1	1.81	0.0041
Chain L; Crystal Structure Of Human B Type Phosphoglycerate Mutase	1.77	0.047
Proteasome activator complex subunit 1	1.62	0.042
Phosphatidylethanolamine-binding protein 1	1.46	0.0014
Proteasome subunit beta type-7	1.28	0.045
Heat shock protein beta-1	-3.04	0.0039

Table 3.3.3: Adenocarcinoma versus Normal. Statistically significant proteins identified from the 2D-DIGE experiments on conditioned media. A positive fold change indicates an increase in adenocarcinoma and a negative fold change indicates a decrease in adenocarcinoma for this protein compared to normal.

Protein ID	Fold Change	T-test
KNP-I beta protein	5.81	0.0068
Stathmin	5.54	2.10E-06
Chain A; Cyclophilin B; Chain A; Cyclophilin B	3.64	0.017
Hsc70-interacting protein	3.6	0.045
Pyruvate kinase 3	3.59	0.015
Rho GDP dissociation inhibitor (GDI) alpha	3.43	0.00046
hnRNP protein A2	2.78	0.051
Ubiquitin carboxy-terminal hydrolase L1	2.29	0.02
Triosephosphate isomerase (TIM)	2.24	0.027
Peroxiredoxin-2	1.96	0.098
Peroxiredoxin 6	1.79	0.038
Ubiquitin carboxyl-terminal esterase L1	1.39	0.067
Chain B; Human Platelet Profilin	-2.47	0.015
Heat shock protein beta-1	-5.47	0.0057

Table 3.3.4: Small cell carcinoma versus Normal. Statistically significant proteins identified from the 2D-DIGE experiments on conditioned media. A positive fold change indicates an increase in small cell carcinoma and a negative fold change indicates a decrease in small cell carcinoma for this protein compared to normal.

Protein ID	Fold Change	T-test
Vimentin	7.56	6.70E-05
Pyruvate kinase 3	6.27	0.00015
Ubiquitin carboxyl-terminal esterase L1	5.04	0.011
KNP-I beta protein	3.8	0.015
hnRNP protein A2	3.58	0.0032
Elongation factor 1-alpha 2	3.54	0.011
Fumarylacetoacetate hydrolase	3.43	0.00043
Fumarate hydratase	3.39	0.04
Stathmin	3.27	0.014
L-lactate dehydrogenase A chain	3.16	0.052
Chaperonin 10	3.12	0.0065
Creatine kinase B-type	3.02	0.00078
Proteasome activator complex subunit 1	2.82	0.0099
Hsc70-interacting protein	2.79	0.015
Peroxiredoxin 3 isoform b	2.24	0.003
Chain L; Crystal Structure Of Human B Type Phosphoglycerate Mutase	2.12	0.012
Peroxiredoxin-1	2.03	0.0022
Inorganic pyrophosphatase	2.01	0.055
Proteasome subunit beta type-7	2	0.0029
Peroxiredoxin 6	1.83	0.022
Proteasome subunit alpha type-1	1.63	0.036
TKT protein	1.54	0.059
Triosephosphate isomerase (TIM)	1.39	0.013
Translin-associated protein X	-1.69	0.055
Heat shock protein beta-1	-3.28	0.04

Table 3.3.5: Squamous cell carcinoma versus Normal. Statistically significant proteins identified from the 2D-DIGE experiments on conditioned media. A positive fold change indicates an increase in squamous cell carcinoma and a negative fold change indicates a decrease in squamous cell carcinoma for this protein compared to normal.

Protein ID	Fold Change	T-test
Ubiquitin carboxyl-terminal esterase L1	2.89	0.0042
Peroxiredoxin-2	2.46	0.012
Pigment epithelium-derived factor	2.31	0.02
vimentin	2.17	0.009
fumarate hydratase	2.17	0.0099
PARP1 protein; PARP1 protein	2.05	0.01
Peroxiredoxin-1	1.77	2.90E-06
Proteasome activator complex subunit 1	1.74	0.0099
Chaperonin 10	1.74	0.04
Inorganic pyrophosphatase	1.72	0.01
Chloride intracellular channel protein 1	1.65	0.0049
Protein disulfide-isomerase A6	1.65	0.011
Malate dehydrogenase; mitochondrial	1.56	0.016
Proteasome subunit beta type-7	1.55	0.00058
Creatine kinase B-type	1.55	0.011
Peptidylprolyl isomerase A	1.48	0.046
Fumarylacetoacetate hydrolase	1.43	0.02
Chain A; Human Cyclophilin A	1.42	0.034
Tropomyosin 3	1.4	0.028
Adenosylhomocysteinase	1.39	0.03
Pyruvate kinase 3	1.3	0.088
Tyrosine 3/tryptophan 5 -monooxygenase activation protein	1.26	0.057
Chain A; Human Cyclophilin A	1.25	0.079
Chain A; Crystal Structure Of Human Dj-1	-1.33	0.022
L-lactate dehydrogenase A chain	-1.41	0.057
Phosphatidylethanolamine-binding protein 1	-1.46	0.00021
Peroxiredoxin 1	-1.5	0.059
Triosephosphate isomerase (TIM)	-1.56	4.50E-06
Purine nucleoside phosphorylase	-1.57	0.038
Creatine kinase B-type	-1.69	0.023
NM23-H1	-2.19	0.067
Fructose-biphosphate aldolase A	-2.81	0.0044

Table 3.3.6: Squamous cell carcinoma versus Adenocarcinoma. Statistically significant proteins identified from the 2D-DIGE experiments on conditioned media. A positive fold change indicates an increase in squamous cell carcinoma and a negative fold change indicates a decrease in squamous cell carcinoma for this protein compared to adenocarcinoma.

Protein ID	Fold Change	T-test
Pigment epithelium-derived factor	4.02	8.30E-05
Ubiquitin carboxyl-terminal esterase L1	3.61	0.00023
Fumarate hydratase	3.12	0.00033
Protein disulfide-isomerase A6	2.81	8.30E-05
TKT protein	2.52	9.20E-05
Proteasome activator complex subunit 1	2.38	0.00012
Acetyl-CoA acetyltransferase	2.13	0.0012
Aspartate aminotransferase; cytoplasmic	1.99	0.01
Fumarylacetoacetate hydrolase	1.97	0.0015
L-lactate dehydrogenase A chain	1.97	0.017
Uridine diphosphoglucose pyrophosphorylase	1.88	0.025
Creatine kinase B-type	1.85	0.0018
Chain B; Human Platelet Profilin	1.82	0.0014
Synaptic vesicle membrane protein VAT-1 homolog	1.78	0.008
Pyruvate kinase 3	1.75	0.0042
Proteasome subunit beta type-7	1.73	0.00029
Aspartate aminotransferase; mitochondrial	1.7	0.03
Phosphoglycerate kinase 1	1.65	0.0042
Tyrosine 3/tryptophan 5 -monooxygenase activation protein	1.64	0.0076
Malate dehydrogenase; mitochondrial	1.64	0.012
Chain F; Nucleoside Triphosphate	1.64	0.094
Actin-related protein 2	1.63	0.051
Adenosylhomocysteinase	1.61	0.0073
Chloride intracellular channel protein 1	1.57	0.02
Tropomyosin 3	1.54	0.013
Rhabdomyosarcoma antigen	1.54	0.038
UPF0568 protein C14orf166	1.51	0.023
Peptidylprolyl isomerase A	1.47	0.067
Malate dehydrogenase; cytoplasmic	1.45	0.011
ACTB protein; Isocitrate dehydrongenase [NADP] cytoplasmic	1.43	0.028
Chain A; Human Cyclophilin A	1.37	0.072
TAGLN2 protein	1.28	0.041
Chain L; Crystal Structure Of Human B Type Phosphoglycerate Mutase	1.23	0.099
Peroxiredoxin 1	-1.5	0.074
Triosephosphate isomerase (TIM)	-1.61	0.025
Phosphatidylinositol transfer protein alpha isoform	-1.64	0.015
Stathmin	-1.69	0.0033
Chain A; Cyclophilin B	-1.82	0.068
Rho GDP dissociation inhibitor (GDI) alpha	-2.43	3.30E-05
HNRPA2B1 protein	-9.37	0.00074

Table 3.3.7: Squamous cell carcinoma versus Small cell carcinoma. Statistically significant proteins identified from the 2D-DIGE experiments on conditioned media. A positive fold change indicates an increase in squamous cell carcinoma and a negative fold change indicates a decrease in squamous cell carcinoma for this protein compared to small cell carcinoma.

Protein ID	Fold Change	T-test
Chaperonin 10	3	0.00092
Ubiquitin carboxyl-terminal esterase L1	2.96	0.0055
Fumarate hydratase	2.41	0.011
Vimentin	2.39	0.0041
PARP1 protein	1.99	0.011
Malate dehydrogenase; mitochondrial	1.97	0.022
Creatine kinase B-type	1.62	0.0081
Pigment epithelium-derived factor	1.6	0.072
Proteasome subunit beta type-7	1.56	0.002
Glutathione S-transferase P	1.52	0.098
Malate dehydrogenase; cytoplasmic	1.49	0.02
Peroxiredoxin 3 isoform b	1.48	0.015
Tropomyosin 3	1.43	0.021
UPF0568 protein C14orf166	1.41	0.087
Peroxiredoxin-1	1.4	0.018
Chain A; Human Cyclophilin A	1.38	0.051
Proteasome subunit alpha type-1	1.35	0.049
Tyrosine 3/tryptophan 5 -monooxygenase activation protein	1.28	0.081
Triosephosphate isomerase (TIM)	1.28	0.096
Adenosylhomocysteinase	1.24	0.049
Heat shock protein beta-1	-1.32	0.071
Phosphoglycerate kinase 1	-1.63	0.037
Creatine kinase B-type	-2.12	0.095

Table 3.3.8: Squamous cell carcinoma versus Large cell carcinoma. Statistically significant proteins identified from the 2D-DIGE experiments on conditioned media. A positive fold change indicates an increase in squamous cell carcinoma and a negative fold change indicates a decrease in squamous cell carcinoma for this protein compared to large cell carcinoma.

Protein ID	Fold Change	T-test
NM23-H1	2.55	0.027
Triosephosphate isomerase (TIM)	1.99	0.0014
Malate dehydrogenase; cytoplasmic	1.98	0.06
Peroxiredoxin 1	1.92	0.0034
Glutathione S-transferase P	1.88	0.0066
Phosphatidylethanolamine-binding protein 1	1.78	0.0033
KNP-I beta protein	1.55	0.058
L-lactate dehydrogenase A chain	1.38	0.093
Chain A; Crystal Structure Of Human Dj-1	1.31	0.052
Inorganic pyrophosphatase	-1.53	0.027
Peroxiredoxin-2	-2	0.098
WDR1 protein	-2.61	0.043

Table 3.3.9: Adenocarcinoma versus Large cell carcinoma. Statistically significant proteins identified from the 2D-DIGE experiments on conditioned media. A positive fold change indicates an increase in adenocarcinoma and a negative fold change indicates a decrease adenocarcinoma for this protein compared to large cell carcinoma.

Protein ID	Fold Change	T-test
Fructose-biphosphate aldolase A	3.46	0.00056
NM23-H1	2.25	0.066
Chain B; Human Platelet Profilin	2.11	0.032
TKT protein	2.03	0.076
Acetyl-CoA acetyltransferase	1.97	0.004
Malate dehydrogenase; cytoplasmic	1.93	0.016
Synaptic vesicle membrane protein VAT-1 homolog	1.88	0.0098
L-lactate dehydrogenase A chain	1.8	0.013
Heat shock protein beta-1	1.8	0.0062
L-lactate dehydrogenase A chain	1.79	0.003
Protein disulfide-isomerase A6	1.71	0.078
Aspartate aminotransferase; mitochondrial	1.7	0.099
Phosphoglycerate kinase 1	1.61	0.021
Creatine kinase B-type	1.51	0.049
Fumarate hydratase	1.44	0.094
Proteasome activator complex subunit 1	1.37	0.026
Rho GDP dissociation inhibitor; Rho GDP dissociation inhibitor	-1.29	0.097
Ubiquitin carboxy-terminal hydrolase L1	-1.43	0.051
Inorganic pyrophosphatase	-1.86	0.093
PARP1 protein	-1.93	0.021
Chain A; Cyclophilin B	-1.95	0.065
Manganese-containing superoxide dismutase	-2.1	0.095
Rho GDP dissociation inhibitor (GDI) alpha	-2.39	0.0085
Peroxiredoxin-2	-2.87	7.90E-05
HNRPA2B1 protein	-29.99	0.0031

Table 3.3.10: Adenocarcinoma versus Small cell carcinoma. Statistically significant proteins identified from the 2D-DIGE experiments on conditioned media. A positive fold change indicates an increase in adenocarcinoma and a negative fold change indicates a decrease in adenocarcinoma for this protein compared to small cell carcinoma.

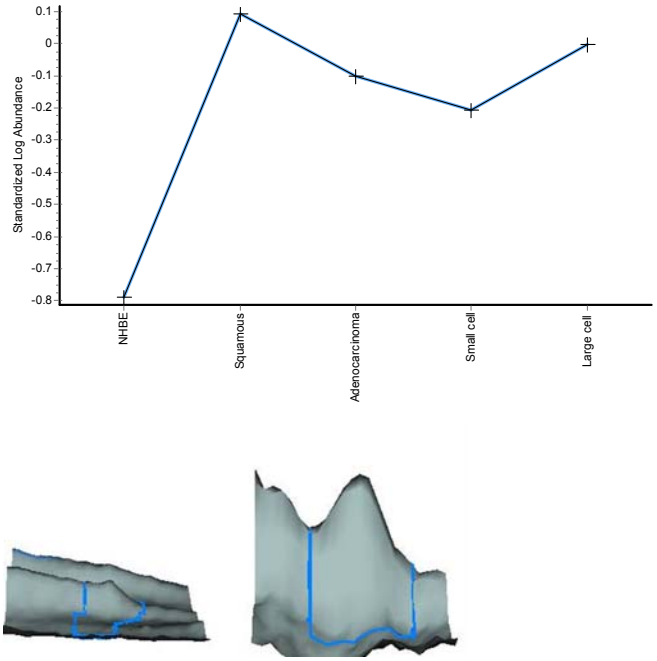
Protein ID	Fold Change	T-test
HNRPA2B1 protein	48.53	0.00093
Chaperonin 10	2.87	0.035
Manganese-containing superoxide dismutase	2.29	0.092
Peroxiredoxin 3	2.23	0.095
KNP-I beta protein	2.15	0.015
Chain A; Cyclophilin B	2.12	0.03
Triosephosphate isomerase (TIM)	2.05	0.012
Phosphatidylethanolamine-binding protein 1	1.99	0.042
Peroxiredoxin 1	1.92	0.0051
PARP1 protein	1.88	0.045
Stathmin	1.68	0.0049
Ubiquitin carboxy-terminal hydrolase L1	1.58	0.011
Fumarylacetoacetate hydrolase	-1.58	0.072
Synaptic vesicle membrane protein VAT-1 homolog	-1.61	0.06
Aspartate aminotransferase; cytoplasmic	-1.66	0.031
L-lactate dehydrogenase A chain	-1.86	0.067
Fructose-biphosphate aldolase A	-1.95	0.067
Chain B; Human Platelet Profilin	-2.09	0.012
Heat shock protein beta-1	-2.21	0.0033
WDR1 protein	-2.44	0.092
TKT protein	-2.52	0.021
Acetyl-CoA acetyltransferase	-2.6	0.004
Phosphoglycerate kinase 1	-2.76	0.017
Uridine diphosphoglucose pyrophosphorylase	-2.88	0.025

Table 3.3.11: Small cell carcinoma versus Large cell carcinoma. Statistically significant proteins identified from the 2D-DIGE experiments on conditioned media. A positive fold change indicates an increase in small cell carcinoma and a negative fold change indicates a decrease in small cell carcinoma for this protein compared to large cell carcinoma.

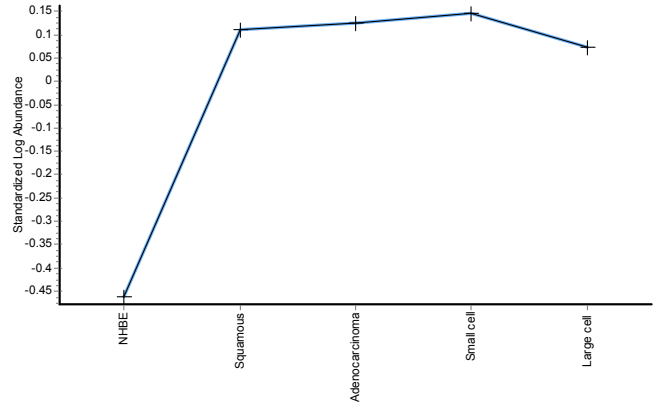
Protein ID	Fold Change	1-ANOVA
HNRPA2B1 protein	13.78	0.00025
Rho GDP dissociation inhibitor (GDI) alpha	2.08	0.0088
Chain A; Cyclophilin B	1.95	0.028
Peroxiredoxin 3 isoform b	1.72	0.088
KNP-I beta protein	1.64	0.0028
Peroxiredoxin-2	1.57	0.0097
Peroxiredoxin-1	1.50	0.096
Vimentin	1.48	0.059
chaperonin 10	1.47	0.033
PARP1 protein	1.45	0.016
Phosphatidylethanolamine-binding protein 1	1.44	0.014
Triosephosphate isomerase (TIM)	1.36	0.0002
Peroxiredoxin 1	1.35	0.0069
Inorganic pyrophosphatase	1.33	0.054
Ubiquitin carboxy-terminal hydrolase L1	1.25	0.076
Stathmin	1.22	0.0049
Malate dehydrogenase; mitochondrial	-1.18	0.049
Chloride intracellular channel protein 1	-1.23	0.061
Elongation factor 1-alpha 2	-1.31	0.049
Proteasome subunit beta type-7	-1.36	3.70E-05
Adenosylhomocysteinase	-1.37	0.06
Tyrosine 3/tryptophan 5 -monooxygenase activation protein	-1.41	0.063
Creatine kinase B-type	-1.42	0.0091
NM23-H1	-1.42	0.084
WDR1 protein	-1.49	0.039
L-lactate dehydrogenase A chain	-1.49	0.052
Malate dehydrogenase; cytoplasmic	-1.54	0.018
Pyruvate kinase 3	-1.56	0.0021
Fumarylacetoacetate hydrolase	-1.66	0.0041
Synaptic vesicle membrane protein VAT-1 homolog	-1.77	0.015
Aspartate aminotransferase; cytoplasmic	-1.85	0.037
Heat shock protein beta-1	-1.86	0.0014
L-lactate dehydrogenase A chain	-1.89	0.018
Phosphoglycerate kinase 1	-1.94	0.046
Chain B; Human Platelet Profilin	-2.01	0.015
Fumarate hydratase	-2.06	0.00092
Ubiquitin carboxyl-terminal esterase L1	-2.07	0.0013
Fructose-biphosphate aldolase A	-2.07	0.0035
Proteasome activator complex subunit 1	-2.18	0.0082
Acetyl-CoA acetyltransferase	-2.20	0.0019
Protein disulfide-isomerase A6	-2.20	0.006

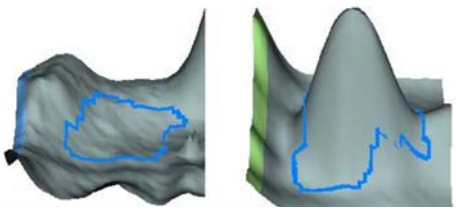
Table 3.3.12: Small cell carcinoma versus Non-Small Cell Lung Cancer. Statistically significant proteins identified from the 2D-DIGE experiments on conditioned media. A positive fold change indicates an increase in small cell carcinoma and a negative fold change indicates a decrease in small cell carcinoma for this protein compared to Non-Small Cell Lung Cancer.

(A)


Standardized Log Abundance
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
NHBE
Squamous
Adenocarcinoma
Small cell
Large cell

(B)


Standardized Log Abundance
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
-0.3
-0.35
-0.4
-0.45
NHBE
Squamous
Adenocarcinoma
Small cell
Large cell



(C)

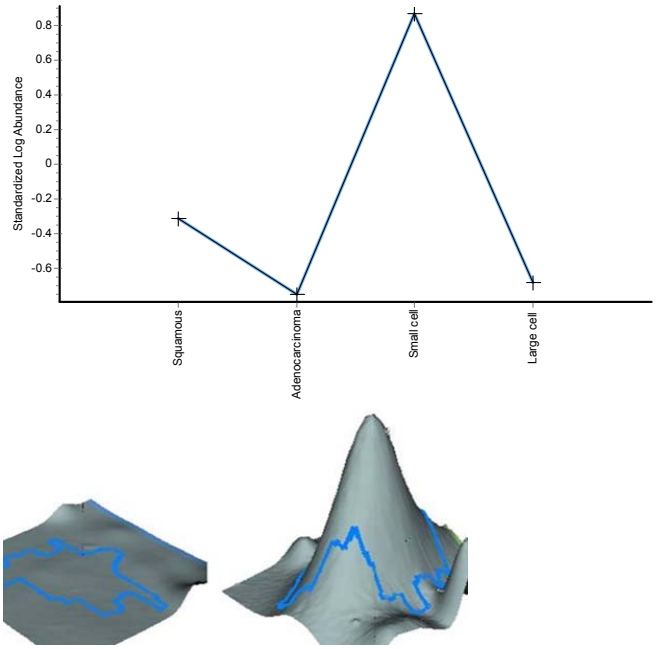
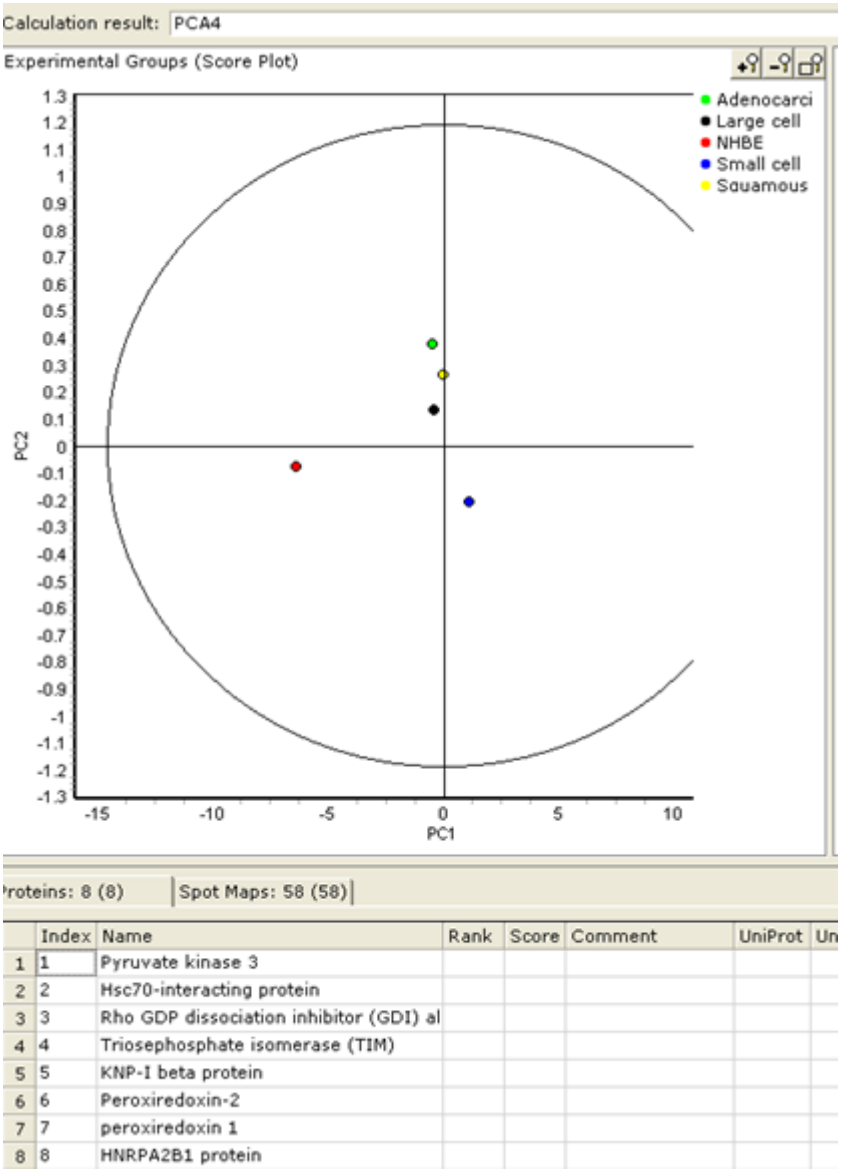


Figure 3.3.13: Example of DeCyder data from three interesting proteins. (A) Pyruvate kinase 3 showed a 4.88 fold increase and ANOVA value of 0.0021 for cancer compared to normal (B) Hsc-70 interacting protein showed a 2.69 and a 3.23 fold increase and ANOVA value of 0.045 and T-Test score of 0.057 for large cell carcinoma and adenocarcinoma respectively compared to normal (C) hnRNPA2B1 showed a 13.78 fold increase and ANOVA value of 0.00025 for small cell carcinoma compared to non-small cell carcinoma.



	Index	Name	Rank	Score	Comment	UniProt	Un
1	1	Pyruvate kinase 3					
2	2	Hsc70-interacting protein					
3	3	Rho GDP dissociation inhibitor (GDI) al					
4	4	Triosephosphate isomerase (TIM)					
5	5	KNP-I beta protein					
6	6	Peroxiredoxin-2					
7	7	peroxiredoxin 1					
8	8	HNRPA2B1 protein					

Figure 3.3.14: Principal component analysis showing the clustering of multivariate data of eight interesting proteins from fifty-eight spot maps in the DeCyder software for adenocarcinoma (Green), squamous cell carcinoma (Yellow), large cell carcinoma (Black), small cell carcinoma (Blue), and the normal human bronchial epithelial cell line (Red). A cumulative variance of 90% was reached after component 2.

3.4 Validation of Targets from Conditioned Media Discovery Phase

For biomarker discovery experiments to become clinically significant the targets of interest must be validated in biological fluids. Serum and plasma are commonly used due to their non-invasive nature. To examine the levels of a biomarker of interest in serum and/or plasma, kits such as enzyme-linked immunosorbent assays (ELISAs) or biochemical assays are commercially available, however due to the high cost of manufacturing these kits they are only available for a selection of proteins. The availability of these kits often determined which proteins of interest to validate, regardless of fold changes and statistical scores.

The clinical validation of these proteins was restricted to available serum/plasma samples, for example at the initial stages only serum samples were available and there were no patients used that were diagnosed with large cell carcinoma. In fact, through all the validation stages large cell carcinoma was not examined due to unavailability of clinical samples. Also, initial serum validation was performed in unevenly distributed sample sets (e.g. 20 normal samples, 22 SCC samples, 20 AD, and 5 SCLC), whereas validation further down the line could be performed in evenly distributed sample sets (e.g. 20 normal, 20 SCC, 20 AD, and 20 SCLC).

Benign lung disease (BLD) control samples were not available for validation performed on late stage lung cancer serum samples. Later, when certain biomarkers were examined in earlier stage lung cancer serum, a benign lung disease control group was used. When available, an autoimmune control group was used along with

the normal control and benign lung disease control. When examined, plasma included a benign lung disease control group.

Validation results (ELISA/biochemical assay) are presented in bar chart and box-and-whisker format. As the sample sets used were relatively small no standard cut-off point was used (e.g. mean +/- two-times the standard deviation), though any results that were obvious standout anomalies were excluded, though this was rare. For this reason box-and-whisker plots were included along with the standard bar chart as they show a more accurate distribution of results. From the box-and-whisker graphs the minimum, maximum, and median can be determined as well as the 25^{th} and 75^{th} percentile, giving a more accurate representation of where the overall results lie. Comparisons with T-test ≤ 0.05 are marked by a single asterisk (*) and ≤ 0.01 are marked by two asterisks (**).

ROC (receiver operating characteristic) curve analysis was performed on the ELISA results of some of the proteins that showed statistically significant comparisons. Using MedCalc software the sensitivity and specificity of the ELISA was calculated as well as the area under the curve. The sensitivity percentage refers to the chances of the test correctly identifying those patients with the disease, and the specificity percentage refers to the chances of the test correctly identifying those patients without the disease. In some cases where the same serum/plasma samples were used the results were combined for logistic regression, a percentage score created to indicate the chances of a correct diagnosis (whether diagnosing healthy or cancer).

3.4.1 hnRNPA2B1, PKM2, and HSC70-interacting Protein (Hip) show potential as lung cancer biomarkers

Figures 3.4.1 and 3.4.4 show an increase in hnRNPA2B1 in lung cancer serum when compared to hnRNPA2B1 levels in normal healthy serum. Both ELISAs (Fig. 3.4.1 and Fig. 3.4.4) were performed on an identical set of serum samples using the USCN kit E90323Hu. Although levels of hnPNPA2B1 are increased in the serum of all lung cancer types compared to normal, the quantities and the comparative differences vary between the two assays. In Fig. 3.4.1 hnRNPA2B1 was barely detectable in normal control serum (in the discovery phase 2D-DIGE conditioned media experiment hnRNPA2B1 was detected in all lung cancer types but not found at all in the NHBE cell line) but showed relatively high levels of almost 100pg/ml in normal control serum in the subsequent repeat ELISA (Fig. 3.4.4). This increase was not seen in the lung cancer types, indeed in Fig. 3.4.4 there is a large decrease in squamous and adneocarcinoma quantity (in pg/ml), however small cell carcinoma stays relatively consistent (in pg/ml). SCLC also showed higher levels in relation to the NSCLC groups in comparison to the first ELISA, indeed showing a higher average result (179pg/ml) compared to AD (163pg/ml), however the median value for AD (149pg/ml) is higher than SCLC (130pg/ml). The median values for the first assay for normal, SCC, AD, and SCLC were 1.6pg/ml, 53pg/ml, 9.8pg/ml, and 54pg/ml, whereas for the 2nd assay were 81pg/ml, 178pg/ml, 149pg/ml, and 129pg/ml respectively. As these assays were performed on an identical set of serum samples the differences seen could be attributed to the assay kit itself; perhaps to batch-to-batch variation. The trend however does not change and it was concluded

that hnRNPA2B1 is increased in lung cancer serum compared to normal healthy serum.

Figure 3.4.5 shows an ELISA on hnRNPA2B1 performed in plasma samples. These plasma samples are from earlier stage lung cancer compared to the late stages of the cancer serum used in Figures 3.4.1 and 3.4.4. The hnRNPA2B1 protein shows higher levels in normal plasma compared to plasma of benign lung disease and lung cancer patients. The quantity of hnRNPA2B1 in normal plasma is more than double than that in normal serum (when compared to Fig. 3.4.4), whereas for the same two ELISAs the quantity of hnRNPA2B1 in cancer serum far exceeds the quantity in normal serum. However, because the cancer stages used in serum and plasma are different, a direct comparison can not be made.

To address this, an ELISA (Fig. 3.4.6) was performed examining matching serum and plasma from the same patients, the cancer group being from early-stage cancer. Matching serum and plasma from normal patients was not included. This ELISA showed relatively similar levels of hnRNPA2B1 in plasma of both benign lung disease patients and lung cancer patients, a trend also seen in Fig. 3.4.5. The matching serum samples however showed lower quantities compared to the plasma samples for both the benign lung disease group and the lung cancer group. It was concluded that levels of hnRNPA2B1 may vary in serum and plasma of benign lung disease patients and lung cancer patients, a benign lung disease control group is essential in assays for potential lung cancer biomarkers, and hnRNPA2B1 has potential in serum as an indicator of advanced lung cancer and also for treatment monitoring.

Two serum ELISAs and two plasma ELISAs were performed on PKM2. There was also a biochemical assay measuring pyruvate kinase activity (of all pyruvate kinase not a particular isoform). One serum ELISA measured PKM2 in normal serum and late-stage SCC, AD, and SCLC (Fig. 3.4.7) and the other serum ELISA measured PKM2 in serum of normal, autoimmune, benign lung disease, and earlier-stage lung cancer (Fig. 3.4.8). Two plasma ELISAs were combined (Fig. 3.4.9) and compared normal, benign lung disease, NSCLC, and SCLC. The biochemical activity assay of PKM2 was performed on serum of normal and late-stage SCC, AD, and SCLC patients, though only 5 SCLC samples were available at the time.

Pyruvate kinase has potential as a biomarker of lung cancer. Although the results of Fig. 3.4.7 were not statistically significant there is a clear increase of PKM2 in lung cancer types compared to normal and even in Fig. 3.4.8 using earlier stage lung cancer PKM2 is elevated in lung cancer serum. Interestingly in Fig. 3.4.8 it is higher in lung cancer compared to benign lung disease too with an excellent T-test score and even in plasma in Fig. 3.4.9 it is elevated in both NSCLC and SCLC compared to the normal and benign lung disease control groups, though only the comparisons between normal and NSCLC, and normal and SCLC, are statistically significant. The conclusion can be made that PKM2 is increased in lung cancer serum and plasma compared to controls, and in the case of serum increased in both early-stage and late-stage lung cancers.

The biochemical assay measures pyruvate kinase activity by calculating the amount of pyruvate produced in a particular time-frame. Pyruvate kinase *activity* could be down in serum due to a lack of available substrate in the cancer serum.

An ELISA was performed for HSC70-interacting protein in serum (Fig. 3.4.11) comparing normal and late-stage SCC, AD, and SCLC. Although raised in all cancer serum compared to normal the biggest difference (and with a T-test less than 0.01) was seen between normal and AD. The levels in AD were almost double that in normal; in normal the average was 554pg/ml and the median was 502pg/ml while in AD they were 1089pg/ml and 759pg/ml respectively. In normal the minimum quantity was 278pg/ml and the maximum 1007pg/ml, while in AD they were 353pg/ml and 2496pg/ml respectively. Comparing the AD results to the normal results the ROC curve analysis showed sensitivity of 55.6% and specificity of 90.0%. The t-test score for AD versus SCC was better than that for Normal versus SCC (0.097 compared to 0.208 respectively), and ROC curve analysis comparing AD and SCC showed sensitivity of 93.3% and specificity of 38.5%. Hip shows some potential as a lung cancer biomarker, though perhaps only useful for adenocarcinoma.

Figure 3.4.12 shows how useful the three biomarkers (hnRNPA2B1, PKM2, Hip) would be when combined as a panel to determine normal serum from lung cancer serum. Although the combination of three can predict 81.58% of cases with an AUC value of 0.875 and a combination of just hnRNPA2B1 and PKM2 can predict 78.48% with an AUC value of 0.831 it should be noted that the combination of the latter was performed in a much larger sample set i.e. 20 normal and 59 cancer, compared to 8 normal and 30 cancer.

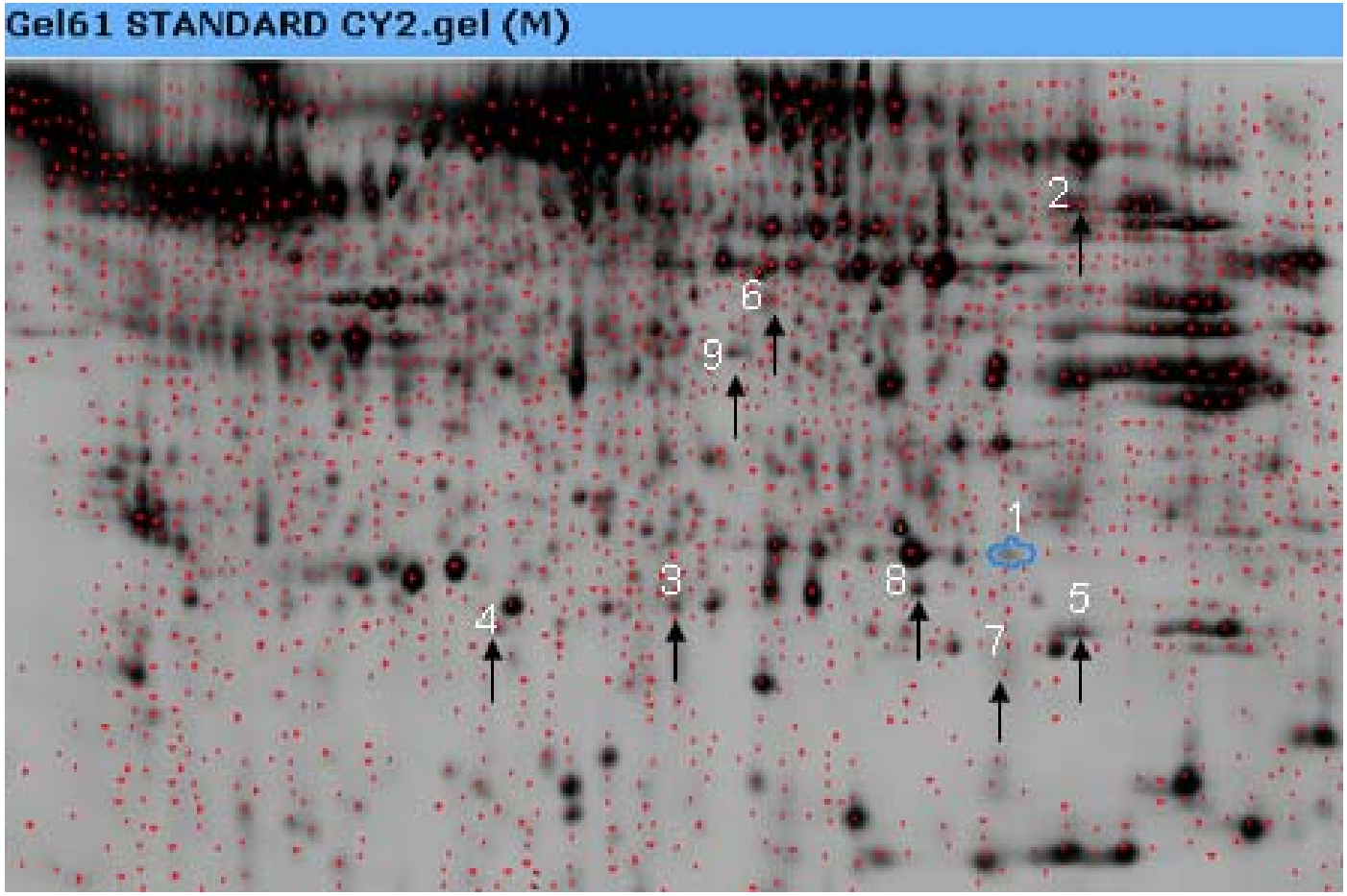
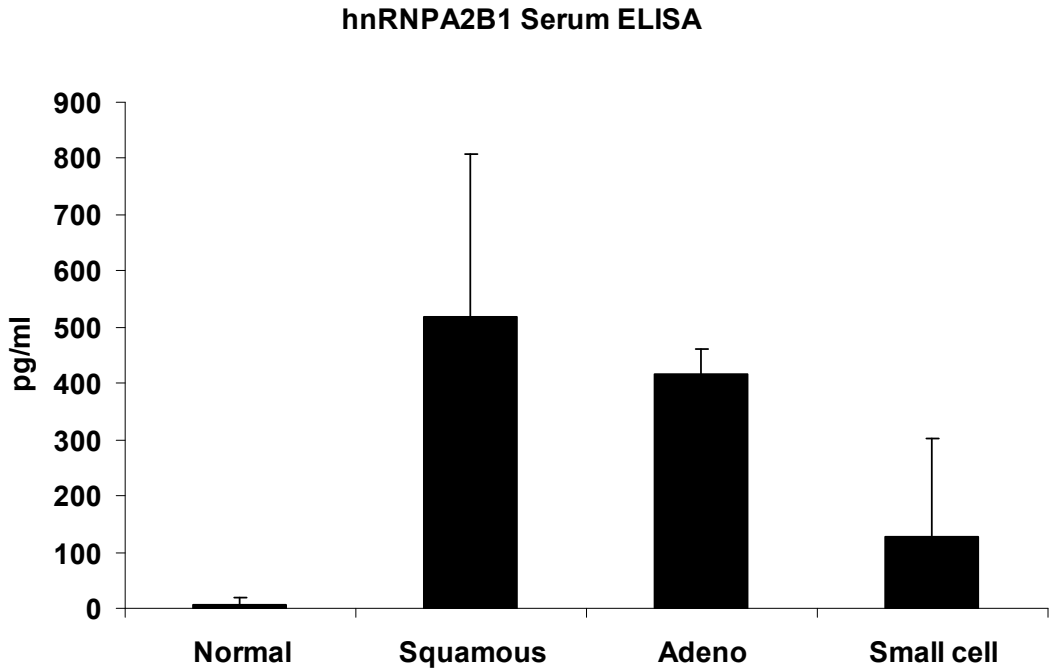


Figure 3.4.1: 2D-DIGE experiment Master Gel showing the nine proteins brought forward to the validation phase; (1) Triosephosphate isomerise, (2) Pyruvate kinase M2, (3) Peroxiredoxin 3, (4) Peroxiredoin 2, (5) Peroxiredoxin 3, (6) Hsc70-interacting protein, (7) hnRNPA2B1, (8) Lactate dehydrogenase, and (9) Glyceraldehyde-3-phosphate.



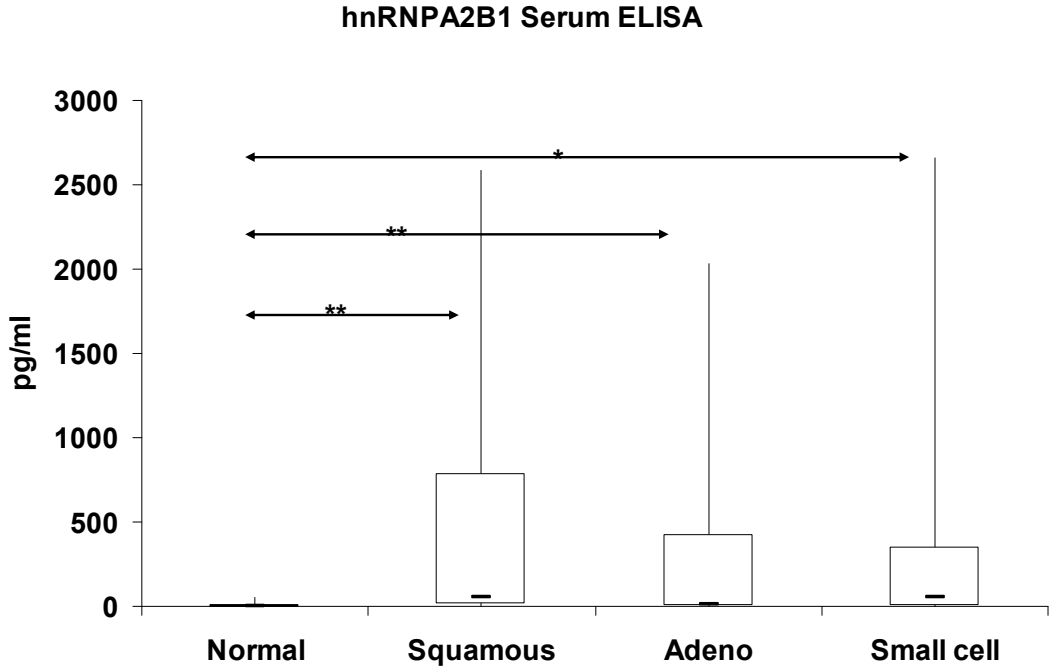
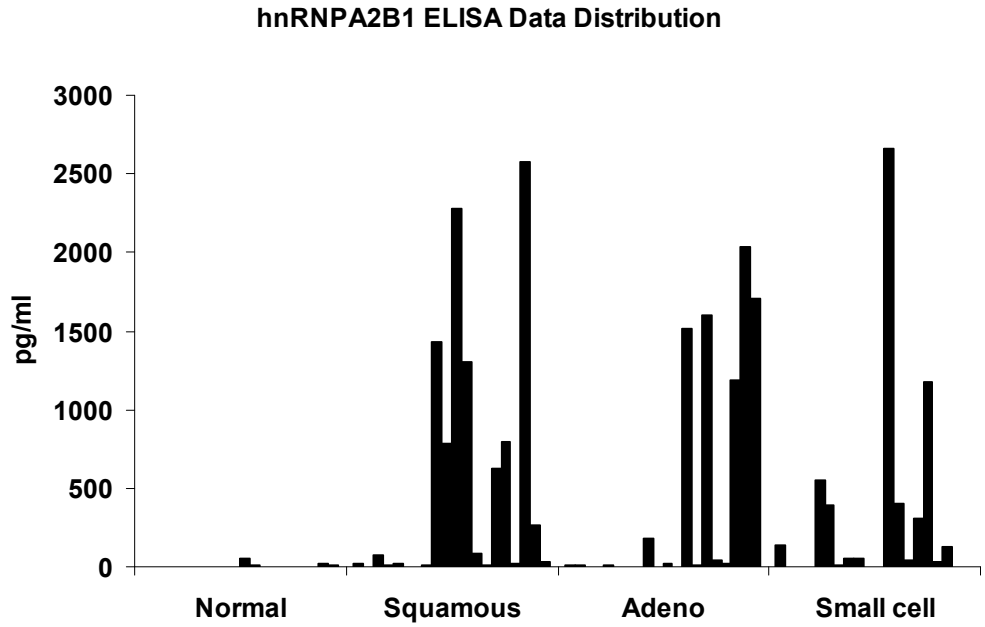


Figure 3.4.1: hnRNPA2B1 ELISA: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 20 squamous, 20 adenocarcinoma, and 19 small cell carcinoma. T-test score; Normal v SCC (*p* value=*0*.006), Normal v AD (*p* value=0.01), and Normal v SCLC (*p* value=0.04).



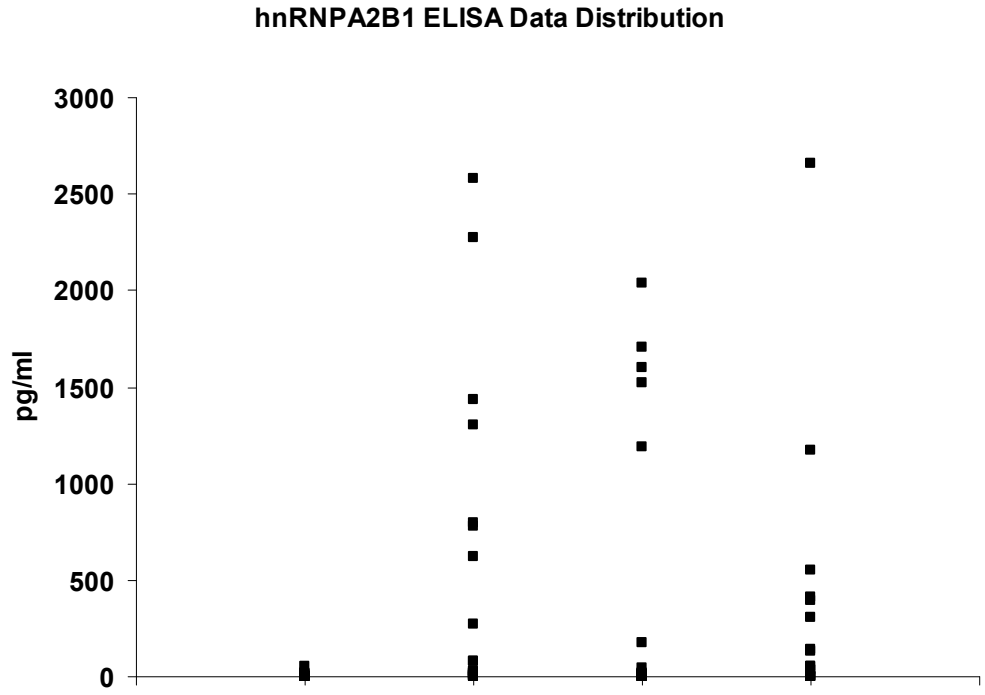


Figure 3.4.2: ELISA performed on hnRNPA2B1 showing the distribution of every data point. Patient serum samples used for this experiment included 20 normal, 20 squamous, 20 adenocarcinoma, and 19 small cell carcinoma. This is the same ELISA as seen in Figure 3.4.1.

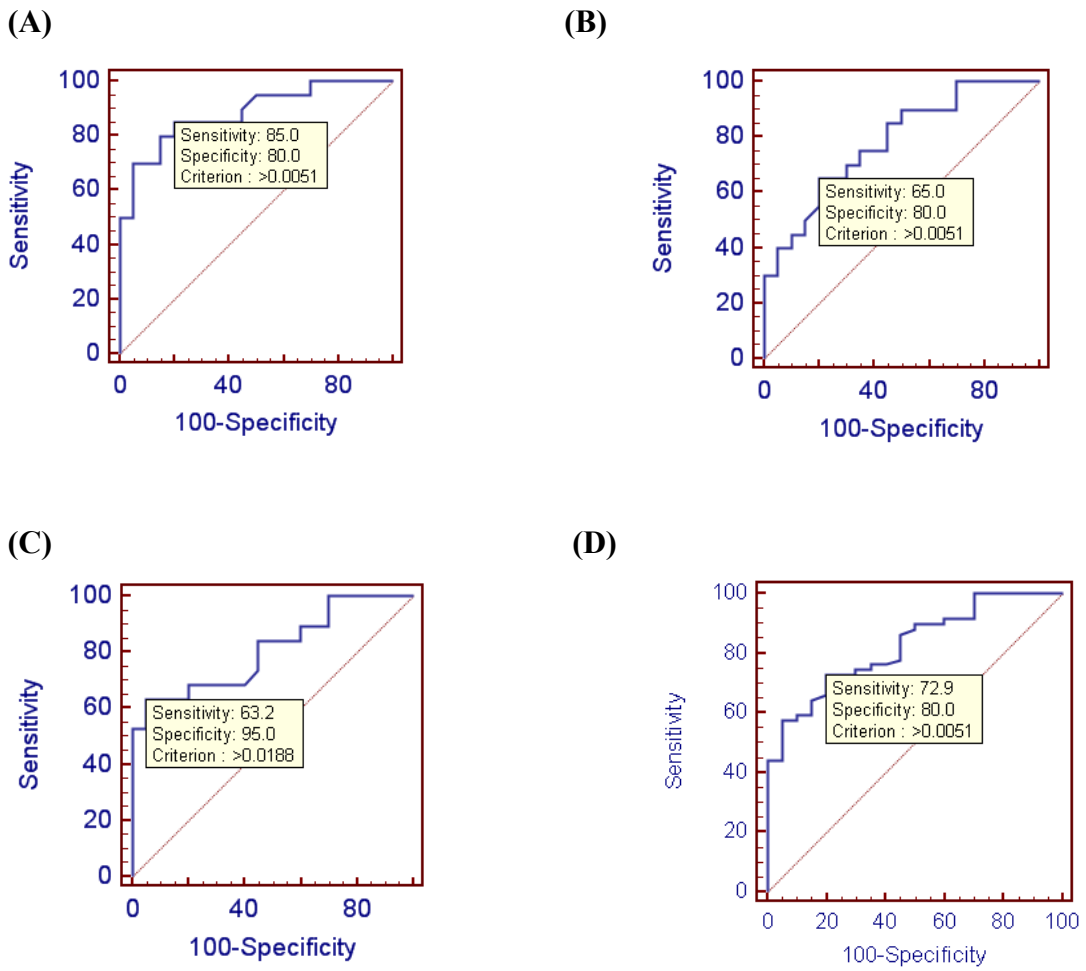


Figure 3.4.3: hnRNPA2B1 ELISA ROC Curve with sensitivity, specificity, and cut-off value in brackets respectively for statistically significant comparisons (A) Squamous versus Normal (85.0%, 80.0%, 0.0051ng/ml) (B) Adenocarcinoma versus Normal (65.0%, 80.0%, 0.0051ng/ml) (C) Small cell carcinoma versus normal (63.2%, 95.0%, 0.0188ng/ml) (D) Lung cancer versus Normal (72.9%, 80.0%, 0.0051ng/ml). X-axis refers to the true positive rate, the y-axis to the false positive rate.

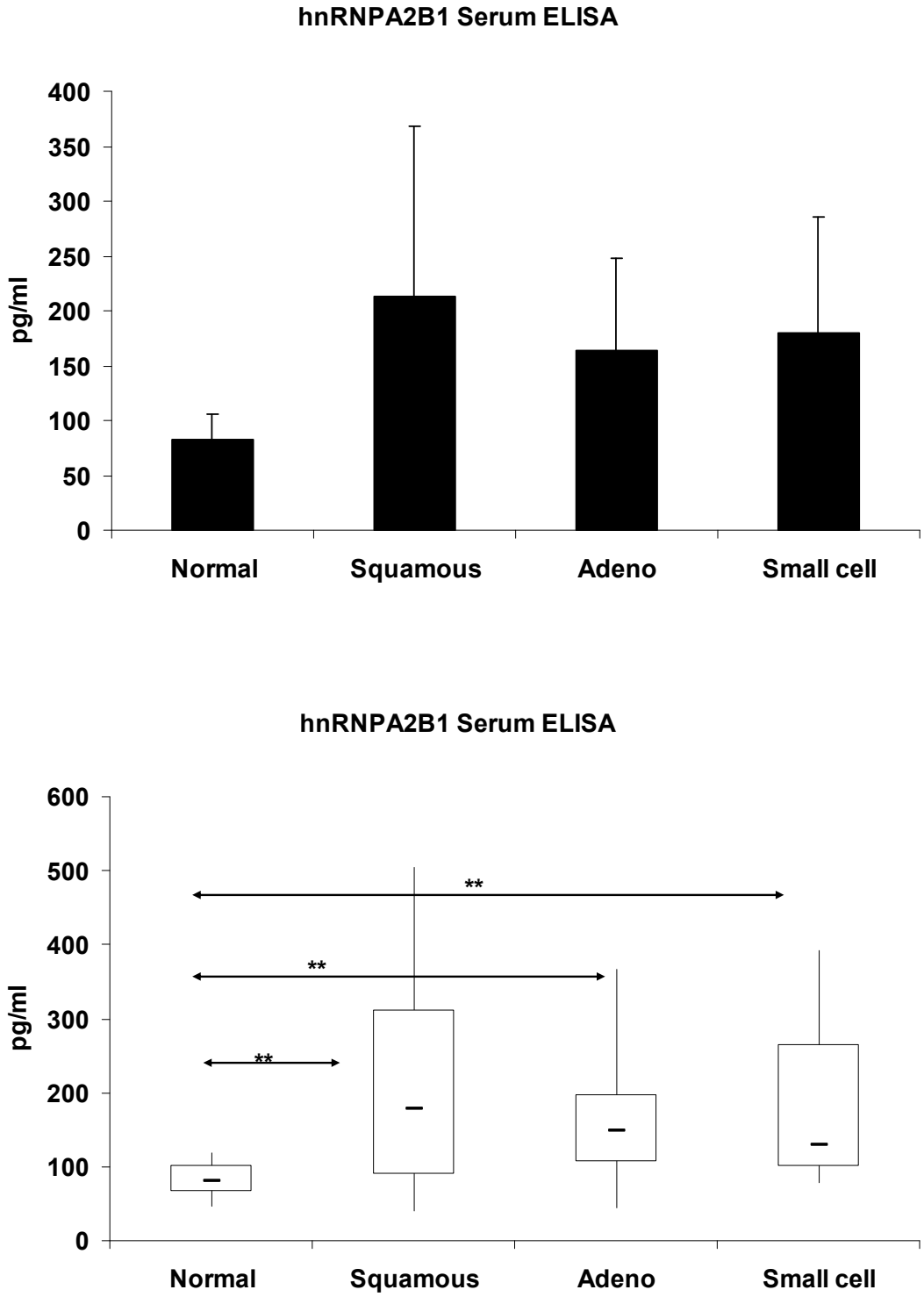
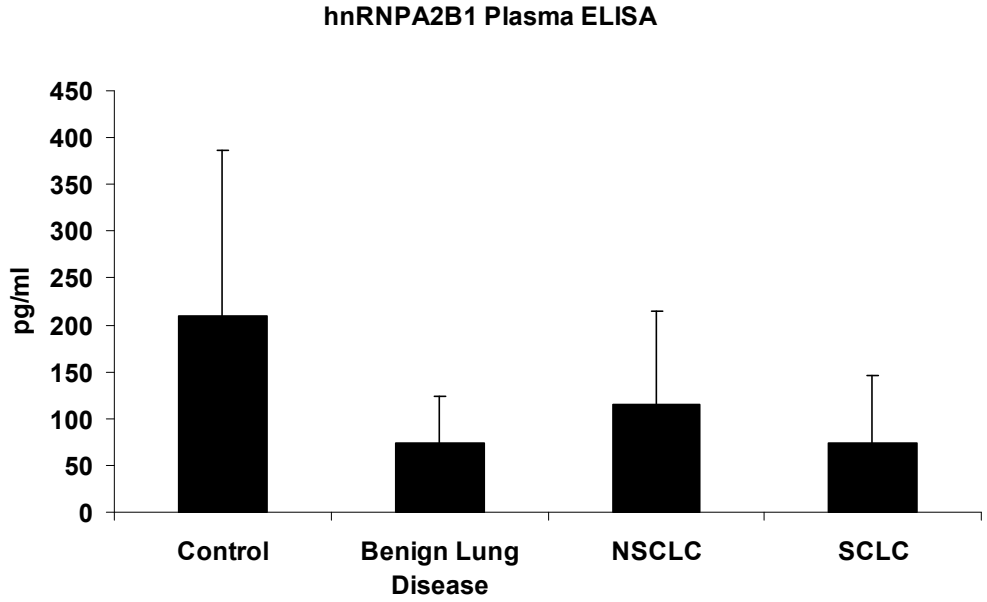


Figure 3.4.4: hnRNPA2B1 ELISA: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 16 normal, 15 squamous, 26 adenocarcinoma, and 19 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.002), Normal v AD (*p* value=0.0005), and Normal v SCLC (*p* value=0.001).



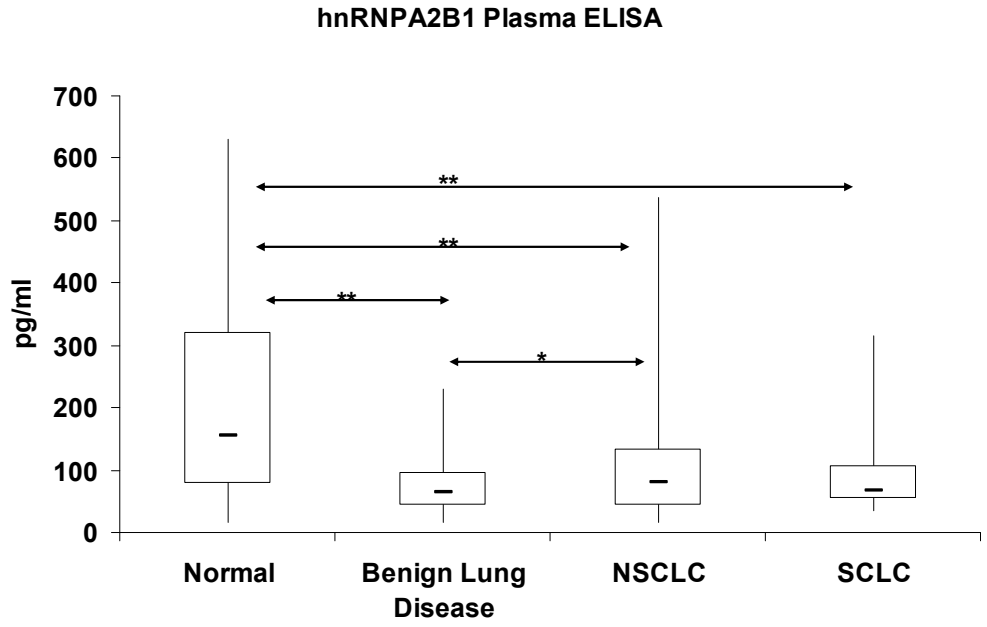
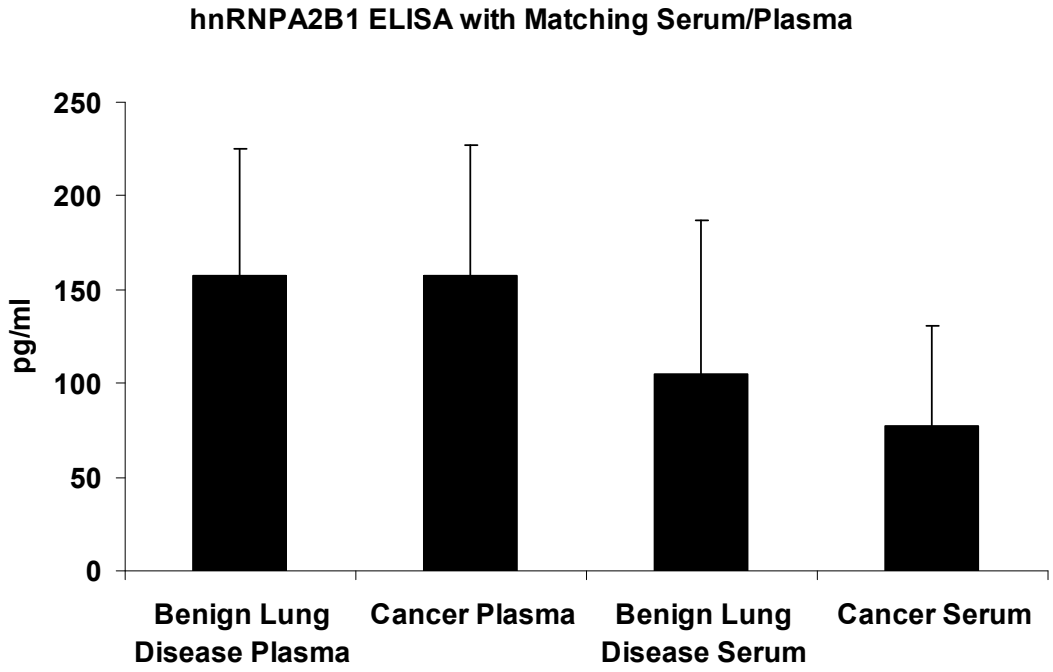


Figure 3.4.5: hnRNPA2B1 ELISA: Bar chart and Box and Whisker Plot for two ELISA's combined. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient plasma samples used for this experiment included 30 normal, 39 benign lung disease, 60 early-stage NSCLC, and 14 early-stage SCLC. T-test score; Normal v BLD (*p* value=3.8E-05), Normal v NSCLC (*p* value=0.001), Normal v SCLC (*p* value=0.01), BLD v NSCLC (*p* value=0.05).



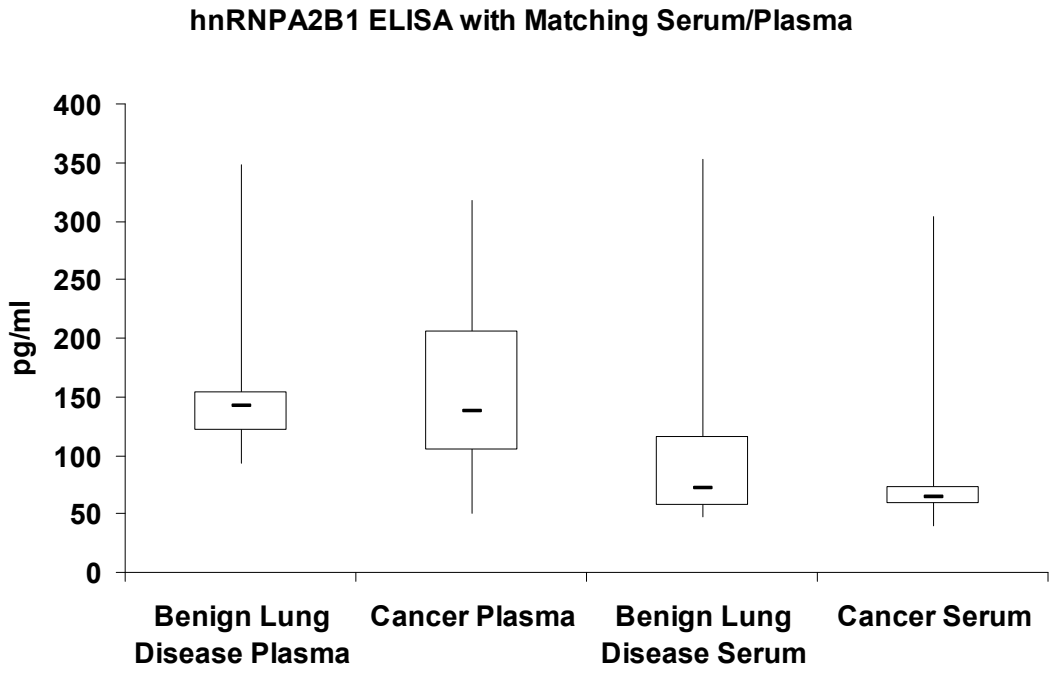
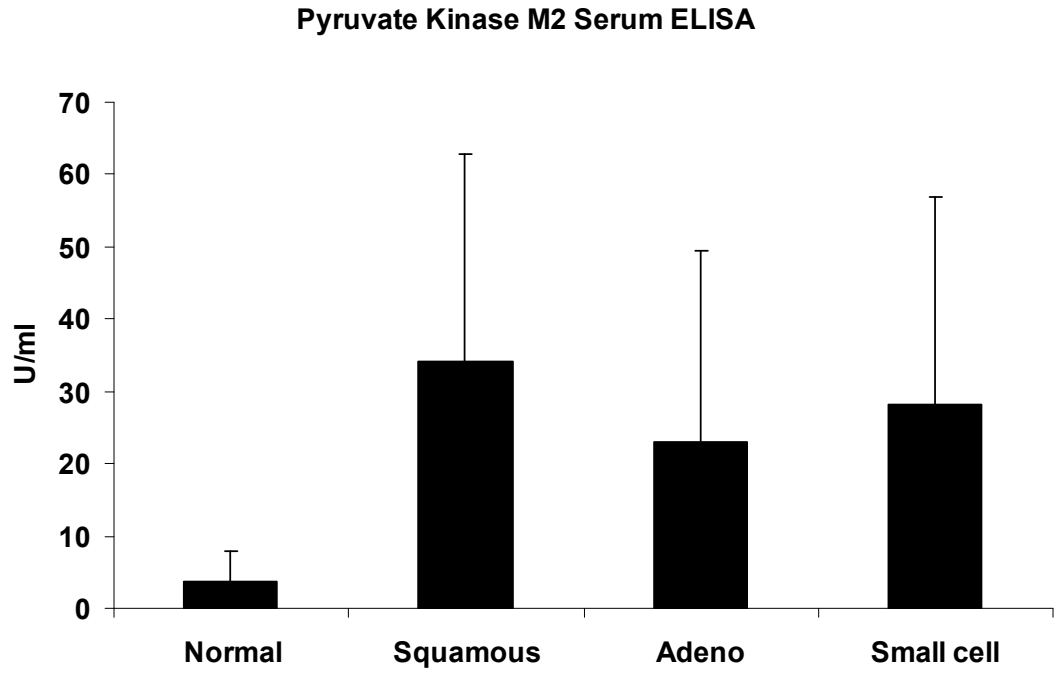


Figure 3.4.6: hnRNPA2B1 ELISA: Bar chart and Box and Whisker Plot for matching early-stage cancer serum and early-stage cancer plasma from the same patients. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient samples used for this experiment included 18 benign lung disease plasma, 18 benign lung disease serum, 22 cancer plasma, and 22 cancer serum.



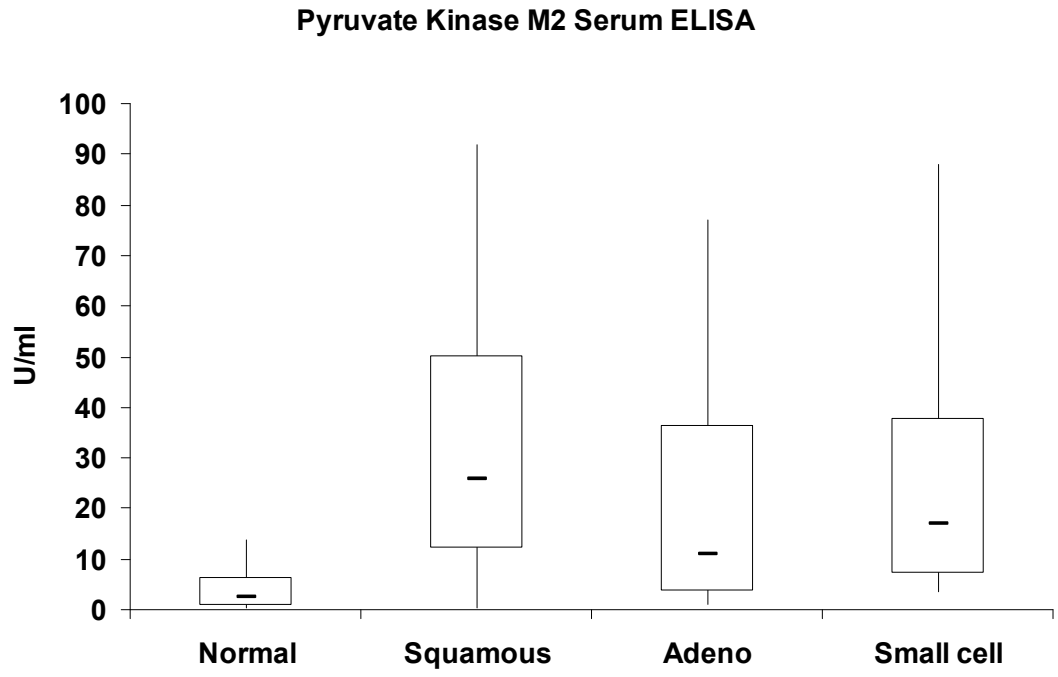
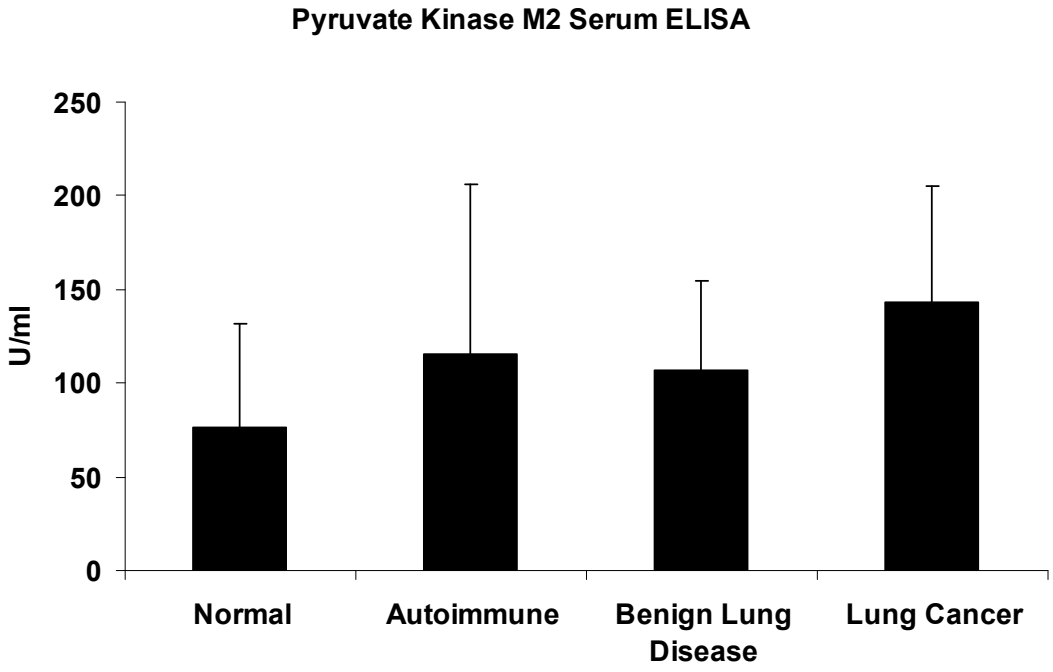


Figure 3.4.7: PKM2 ELISA: Bar chart and Box and Whisker Plot showing normal compared to late-stage lung cancer. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 20 squamous, 20 adenocarcinoma, and 20 small cell carcinoma.



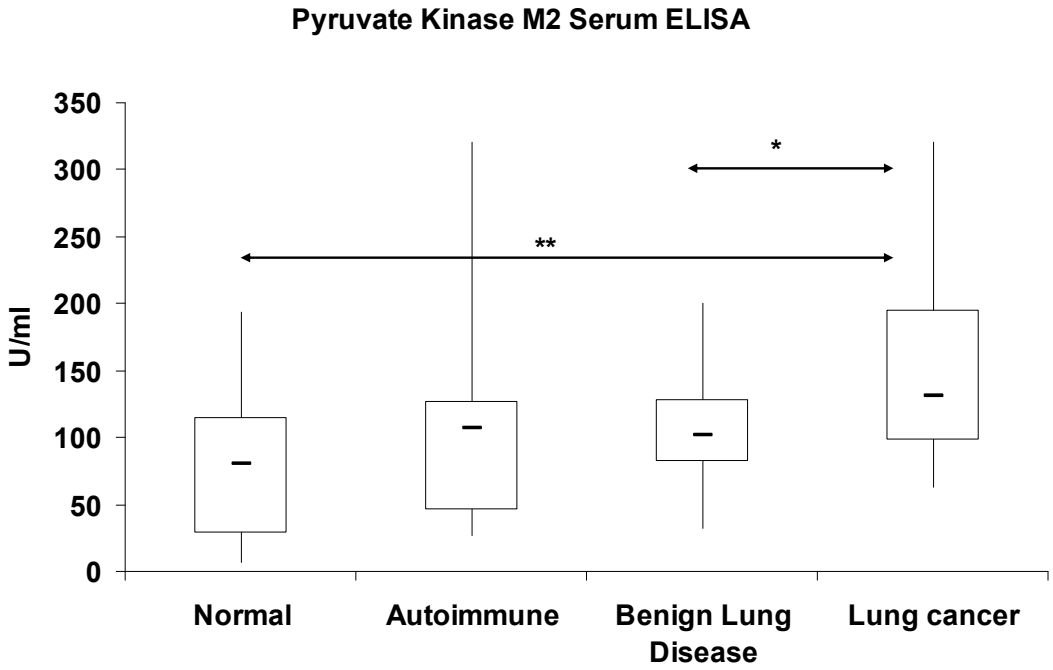
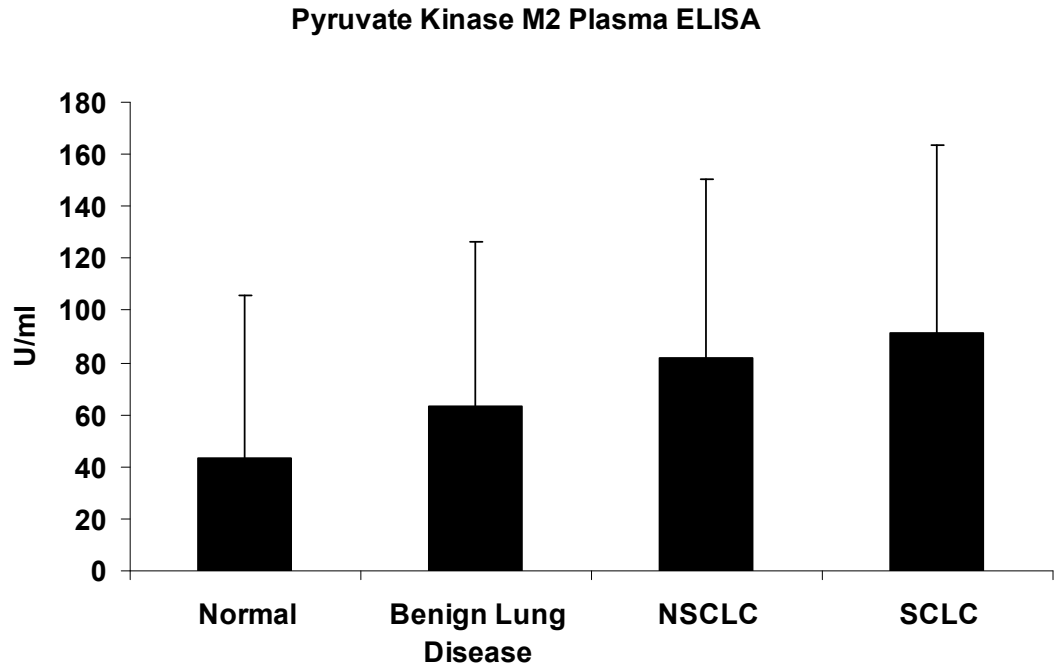


Figure 3.4.8: PKM2 ELISA: Bar chart and Box and Whisker Plot for PKM2 ELISA. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 16 normal, 16 autoimmune, 19 benign lung disease, and 25 early-stage lung cancer. T-test score; Normal v Lung Cancer (*p* value=0.001), BLD v Lung Cancer (*p* value=0.04).



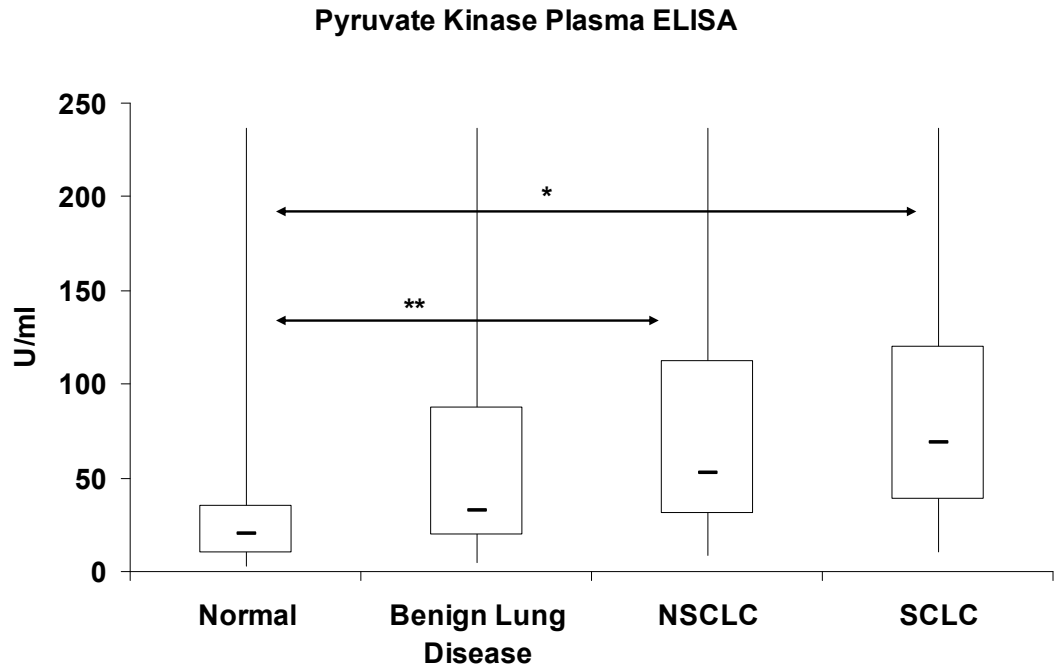
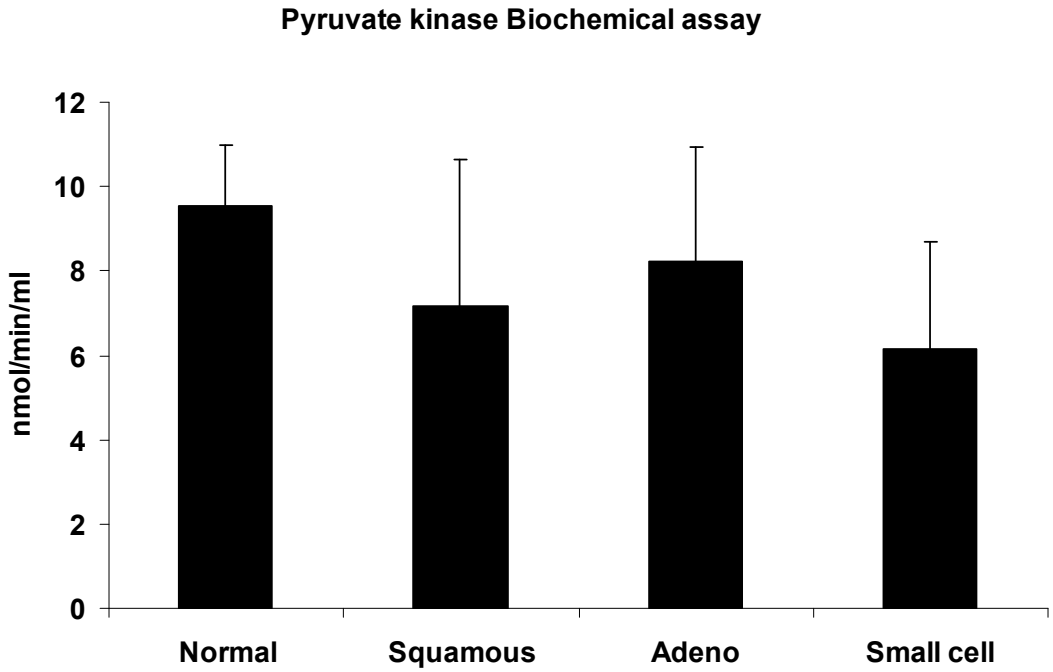


Figure 3.4.9: PKM2 ELISA: Bar chart and Box and Whisker Plot for two ELISAs combined. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient plasma samples used for this experiment included 30 normal, 29 benign lung disease, 60 early-stage NSCLC and 14 early-stage SCLC. T-test value; Normal v NSCLC (*p* value=0.01), Normal v SCLC (*p* value=0.03).



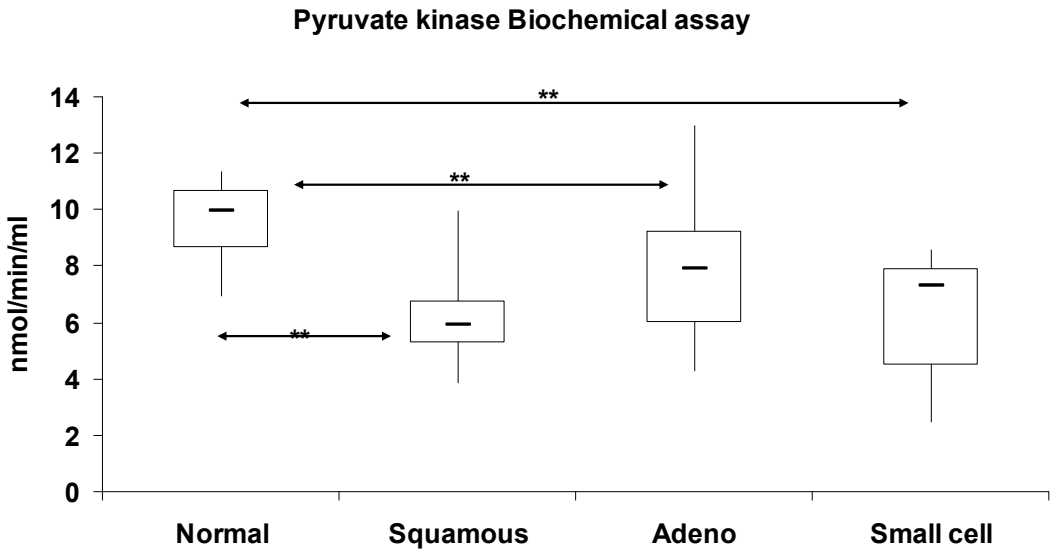
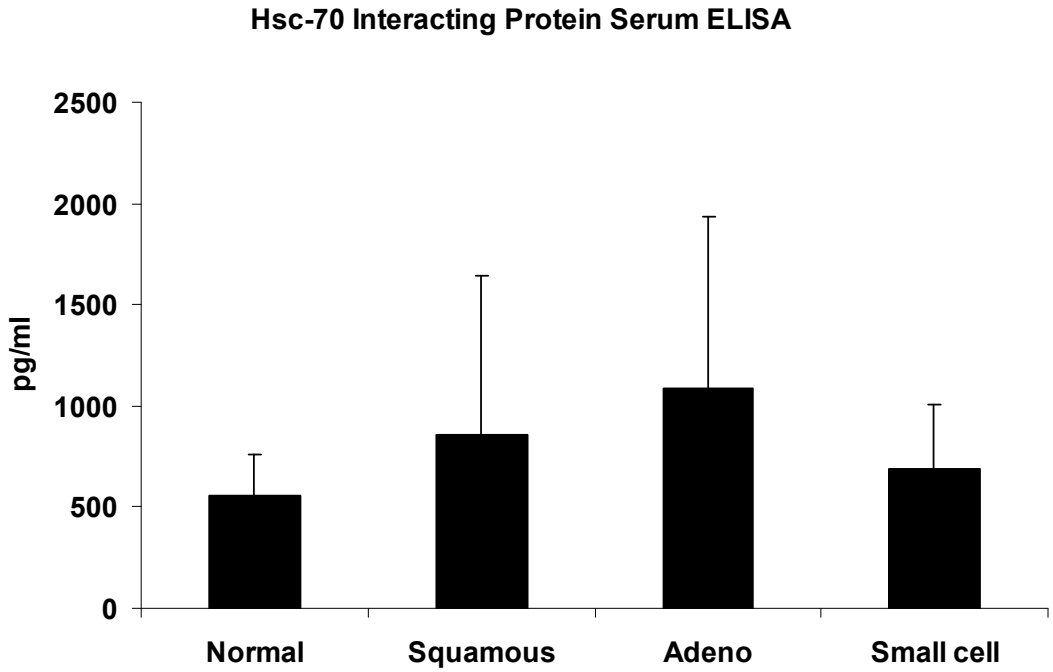


Figure 3.4.10: Pyruvate Kinase Biochemical Activity Assay: Bar chart and box and whisker showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 16 squamous, 17 adenocarcinoma, and 5 small cell carcinoma. T-test score; Normal v SCC (*p* value=6.6E-08), Normal v AD (*p* value=0.01), Normal v SCLC (*p* value=0.0005).



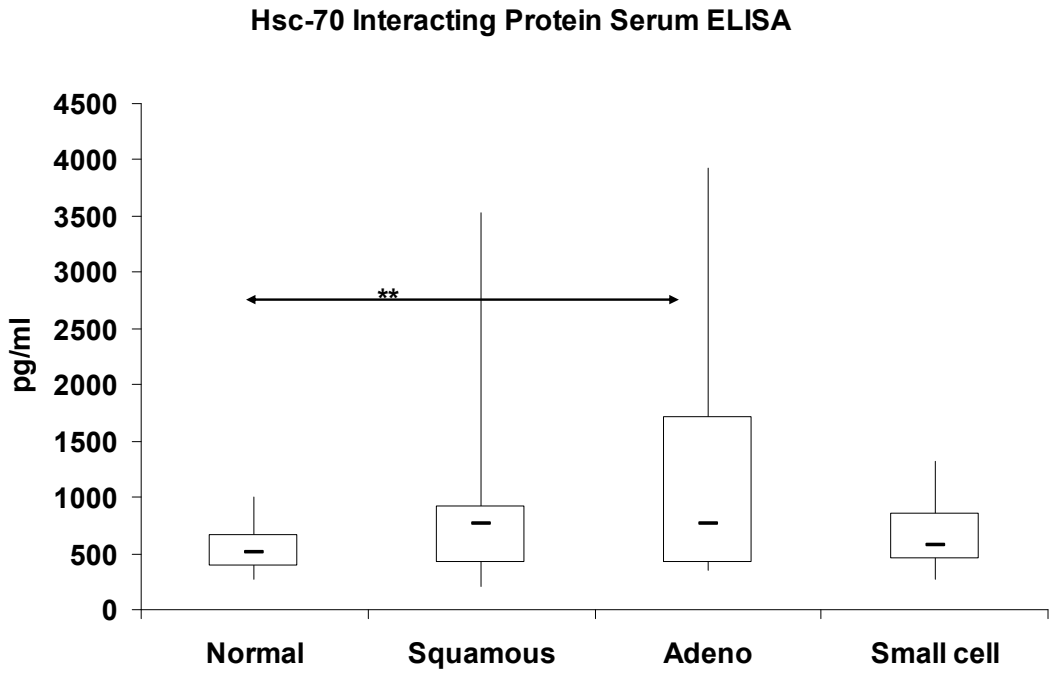


Figure 3.4.11: HSC70-interacting protein ELISA: Bar chart and box and whisker showing normal compared to late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles.Patient serum samples used for this experiment included 20 normal, 16 squamous, 28 adenocarcinoma, and 16 small cell carcinoma. T-test score; Normal v AD (*p* value=0.006).

(A)

Sample size	38
Cases with Y=0	8 (21.05%)
Cases with Y=1	30 (78.95%)

Overall Model Fit

Null model -2 Log Likelihood	39.114
Full model -2 Log Likelihood	25.776
Chi-square	13.338
DF	3
Significance level	P = 0.0040

Coefficients and Standard Errors

Variable	Coefficient	Std. Error	P
hnRNPA2B1	24.31602	24.31106	0.3172
HSC70_interacting_protein	0.0031212	0.0019173	0.1035
PKM2	0.027634	0.027011	0.3063
Constant	-1.6938		

Odds Ratios and 95% Confidence Intervals

Variable	Odds ratio	95% CI
hnRNPA2B1	36.3E+009	0.0000 to 18.0E+030
HSC70_interacting_protein	1.0031	0.9994 to 1.0069
PKM2	1.0280	0.9750 to 1.0839

Hosmer & Lemeshow test

Chi-square	4.7935
DF	8
Significance level	P = 0.7794

Classification table (cut-off value p=0.5)

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	4	4	50.00 %
Y = 1	3	27	90.00 %
Percent of cases correctly classified			81.58 %

ROC curve analysis

Area under the ROC curve (AUC)	0.875
Standard Error	0.0580
95% Confidence interval	0.727 to 0.960

(B)

Sample size	79
Cases with Y=0	20 (25.32%)
Cases with Y=1	59 (74.68%)

Overall Model Fit

Null model -2 Log Likelihood	89.394
Full model -2 Log Likelihood	66.650
Chi-square	22.744
DF	2
Significance level	P < 0.0001

Coefficients and Standard Errors

Variable	Coefficient	Std. Error	P
hnRNPA2B1	47.39639	26.04615	0.0688
PKM2	0.010815	0.0082647	0.1907
Constant	-0.1398		

Odds Ratios and 95% Confidence Intervals

Variable	Odds ratio	95% CI
hnRNPA2B1	384E+018	0.0259 to 5.69E+042
PKM2	1.0109	0.9946 to 1.0274

Classification table (cut-off value p=0.5)

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	9	11	45.00 %
Y = 1	6	53	89.83 %
Percent of cases correctly classified			78.48 %

ROC curve analysis

Area under the ROC curve (AUC)	0.831
Standard Error	0.0480
95% Confidence interval	0.729 to 0.906

Figure 3.4.12: Logistic regression performed on matching serum patient samples (A) combining ELISA results from hnRNPA2B1, HSC70-interacting protein, and PKM2 with 8 matching normal and 30 matching cancer samples, indicating 81.58% of cases would be correctly classified with an AUC value of 0.875, and (B) combining ELISA results from hnRNPA2B1 and PKM2 with 20 matching normal and 59 matching cancer samples, indicating 78.48% of cases would be correctly classified with an AUC value of 0.831.

3.4.2 Peroxiredoxins, Lactate Dehydrogenase, GAPDH, and Triosephosphate Isomerase show little potential as possible Lung Cancer Biomarkers

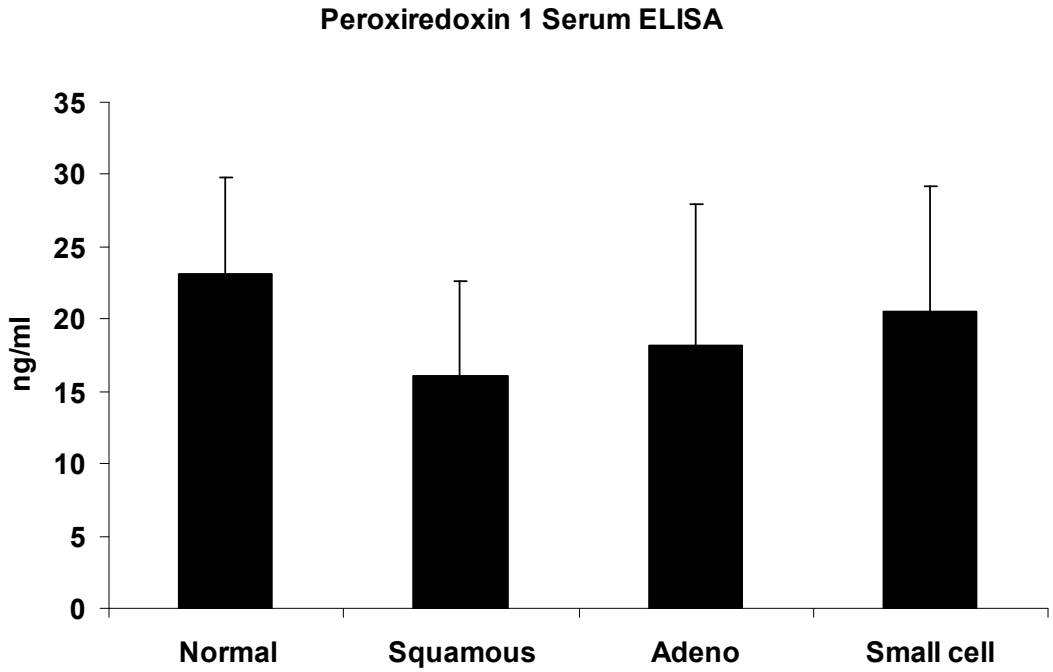
Serum ELISAs were performed for Peroxiredoxins 1, 2, and 3 comparing normal to late-stage lung cancer. For Prx 2 (Fig. 3.4.14) there were 20 samples for each of the cancers, whereas there were only five SCLC for Prx 1 ELISA (Fig. 3.4.13) and eight SCLC for Prx 3 ELISA (Fig. 3.4.15). All three peroxiredoxins were decreased in cancer serum compared to normal, though for Prx 3 the differences were negligible. In the Prx 1 ELISA the difference between the normal group and both the SCC and AD group was statistically significant; however there were only five SCLC serum samples available at the time compared to 21 for SCC and 20 for AD so the small sample set would affect the statistical comparison of SCLC versus normal group of 19 samples. In the Prx 2 ELISA the difference between the normal group and both SCC and SCLC were statistically significant.

Although peroxiredoxins often have elevated expression in lung cancer tissue compared to normal lung tissue perhaps the decrease of Prx 1 and Prx 2 in lung cancer serum has significance. If cancer serum has more reactive oxygen species than normal serum then any peroxiredoxins in the serum could be depleted as they attack these oxygen derived molecules. It could be interesting to examine the three peroxiredoxins in early-stage lung cancer serum because if their levels are elevated in this serum they could act as markers of disease progression.

The LDH assay (Fig. 3.4.16) shows higher levels in cancer serum compared to normal serum. This result was expected as LDH is known to be raised in many

cancers. This lack of specificity throws doubt on its usefulness as a lung cancer biomarker.

Both the GAPDH and triosephosphate isomerise serum assays (Fig. 3.4.17 and 3.4.18 respectively) showed little to no change in lung cancer compared to normal. Triosephosphate isomerise shows no potential as a lung cancer biomarker while the GAPDH assay could be treated as a loading control, as it indicated similar protein content throughout all serum samples used in this assay.



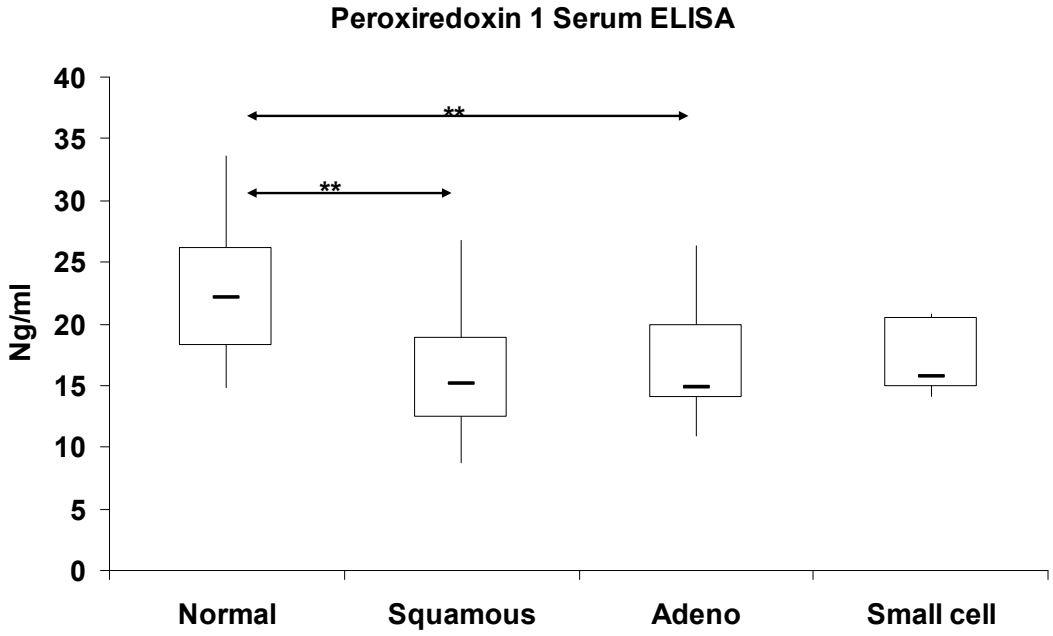
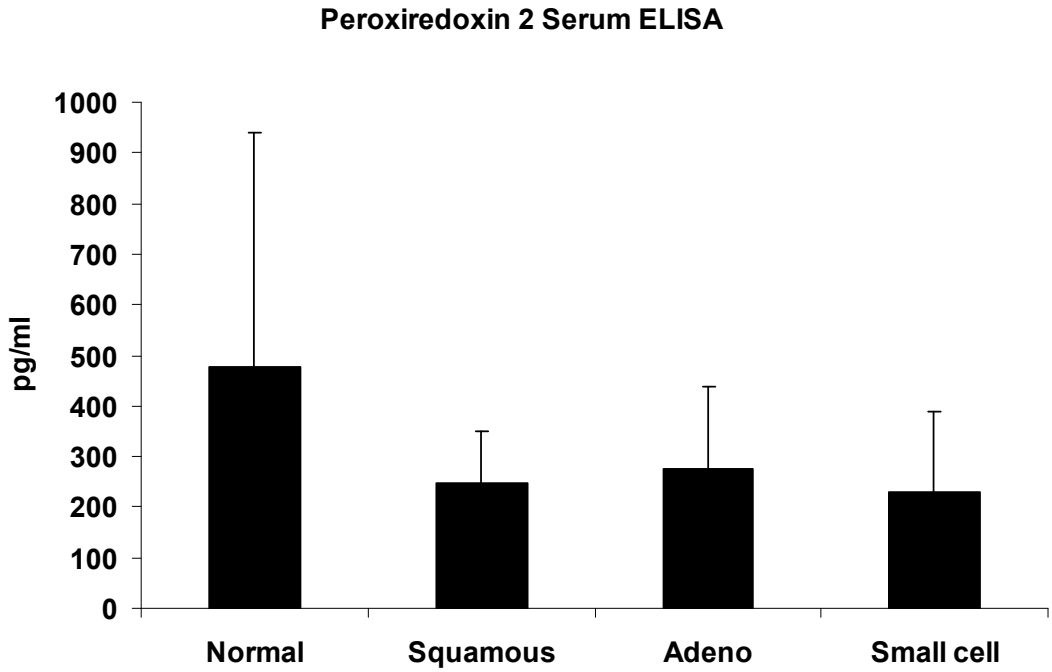


Figure 3.4.13: Peroxiredoxin 1 ELISA: Bar chart and box and whisker showing normal compared to late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 19 normal, 21 squamous, 20 adenocarcinoma, and 5 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.0002), Normal v AD (*p* value=0.002).



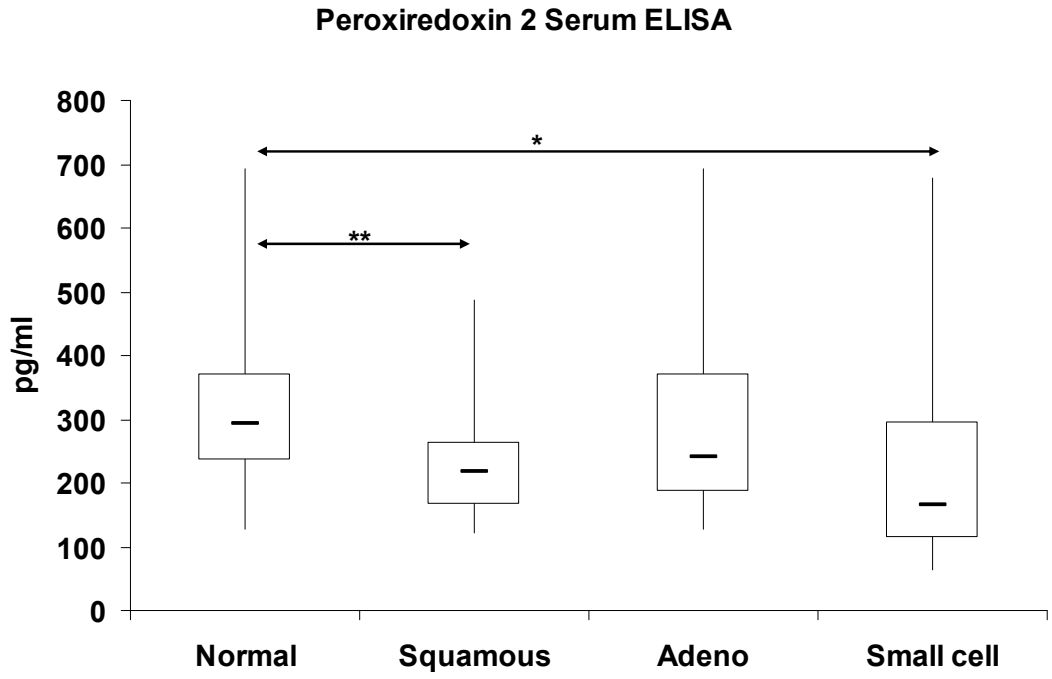
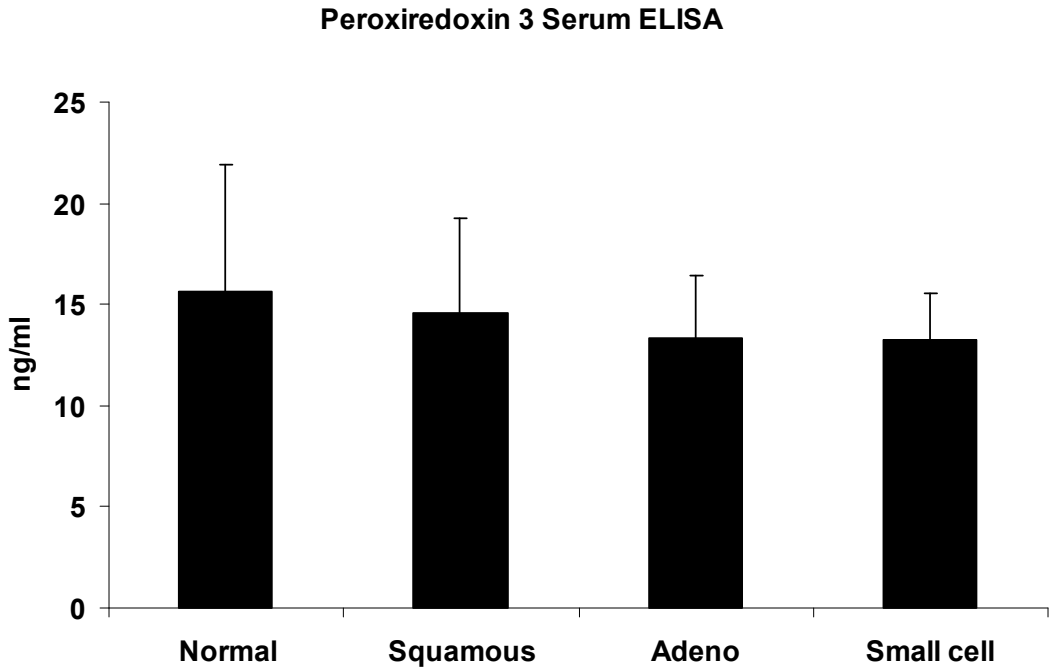


Figure 3.4.14: Peroxiredoxin 2 ELISA: Bar chart and box and whisker plot showing normal compared to late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 19 normal, 20 squamous, 20 adenocarcinoma and 20 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.01), Normal v SCLC (*p* value=0.02).



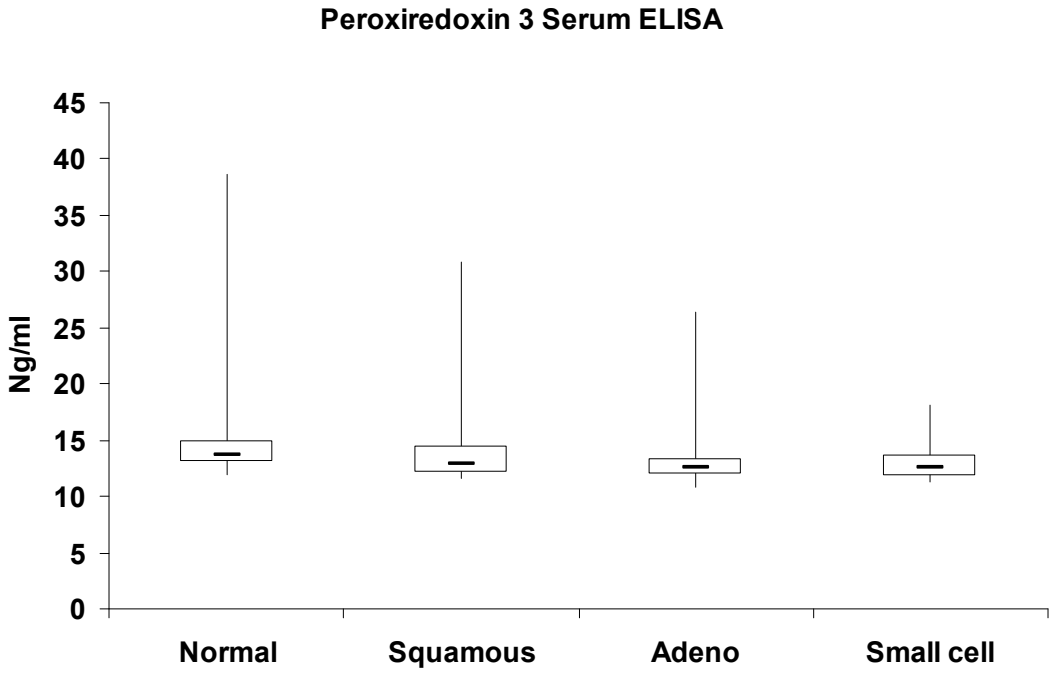
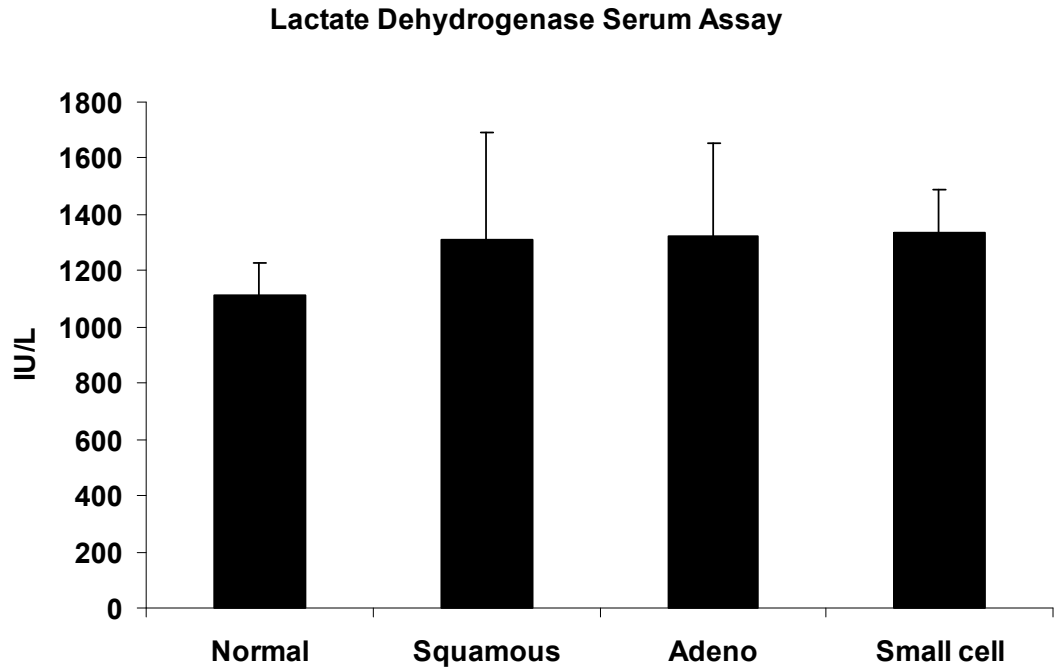


Figure 3.4.15: Peroxiredoxin 3 ELISA: Bar chart and box and whisker plot showing normal compared to late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 21 squamous, 24 adenocarcinoma and 8 small cell carcinoma.



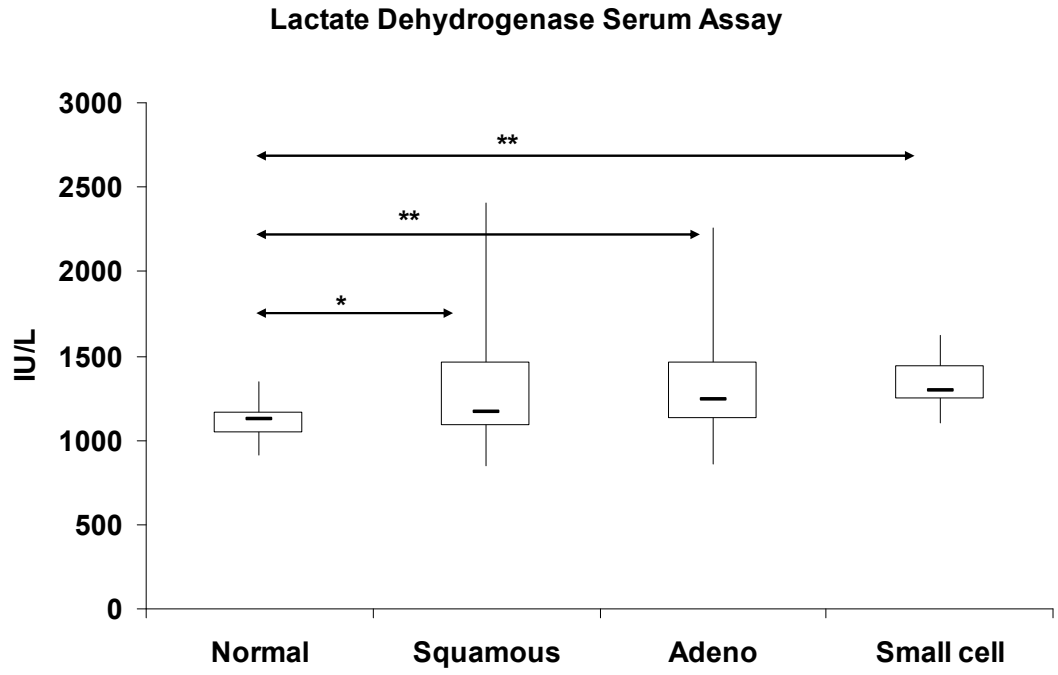


Figure 3.4.16: Lactate dehydrogenase ELISA: Bar chart and box and whisker plot showing normal compared to late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 19 squamous, 28 adenocarcinoma and 12 small cell carcinoma. T-test value; Normal v SCC (*p* value=0.03), Normal v AD (*p* value=0.01), Normal v SCLC (*p* value=5E-05).

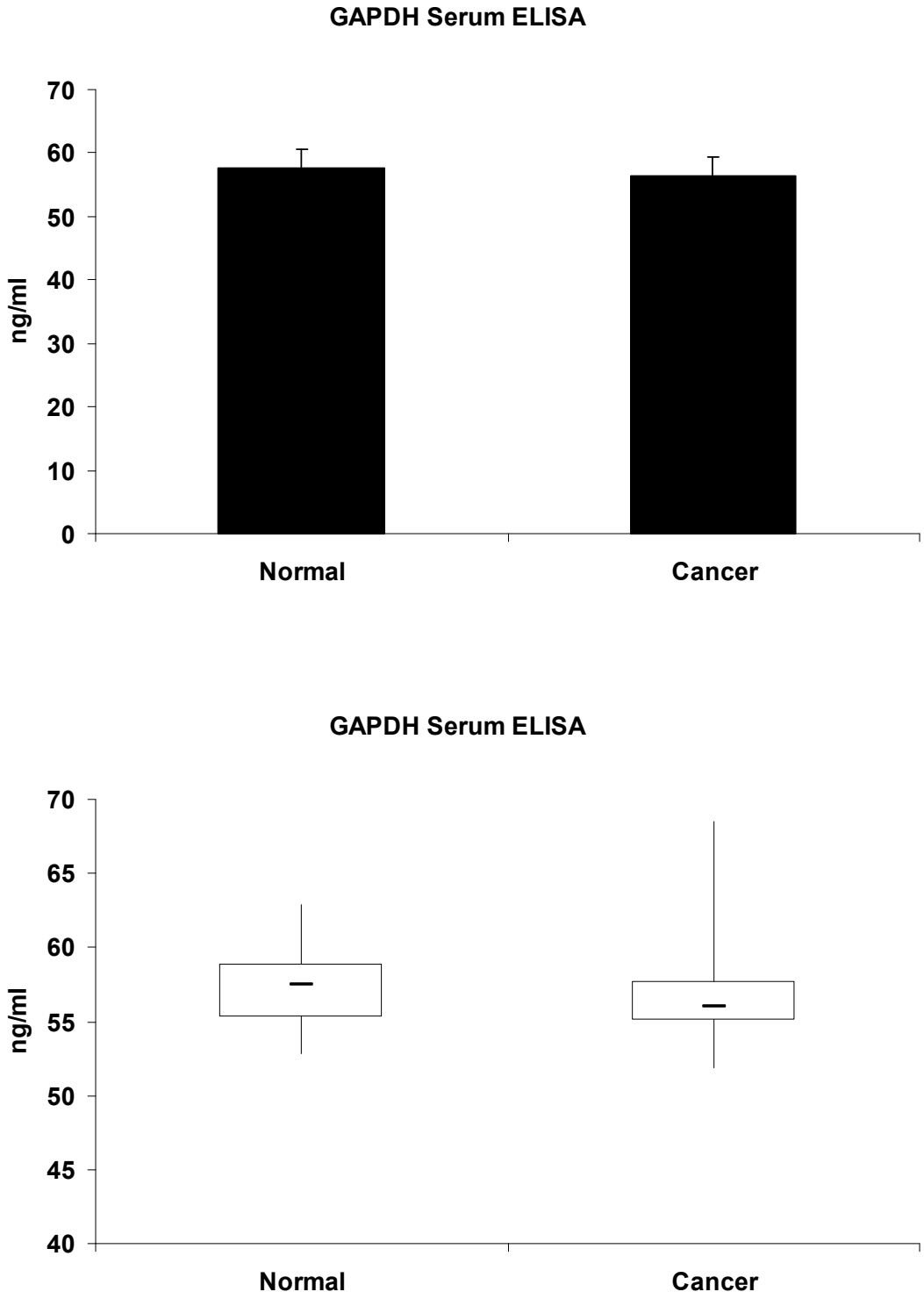
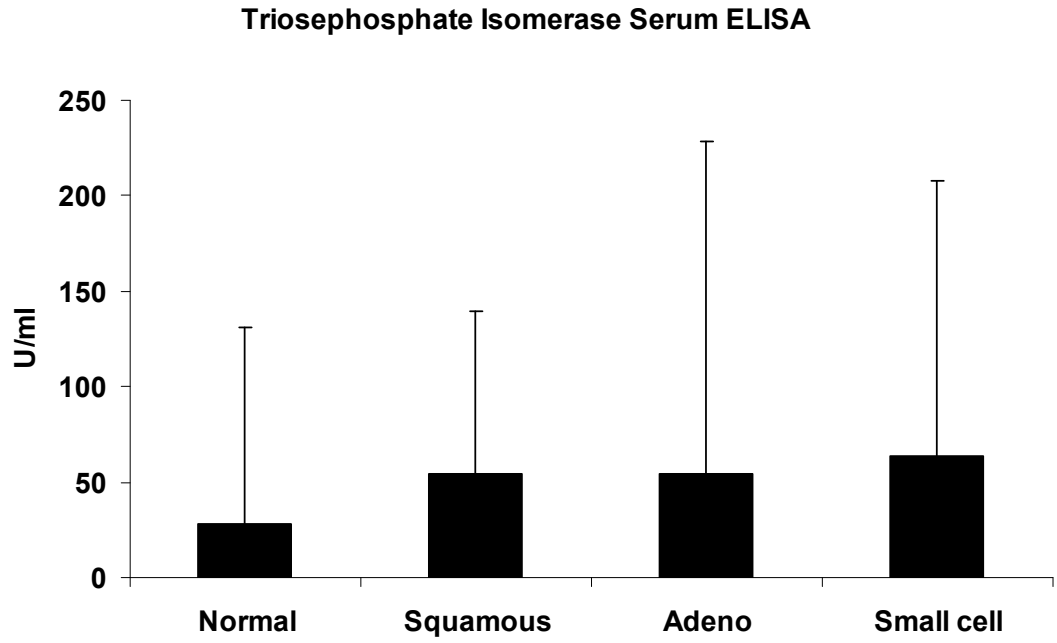


Figure 3.4.17: Glyceraldehyde-3-phosphate dehydrogenase ELISA: Bar chart and box and whisker plot showing normal compared to late-stage lung cancer. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 12 normal and 74 lung cancer samples.



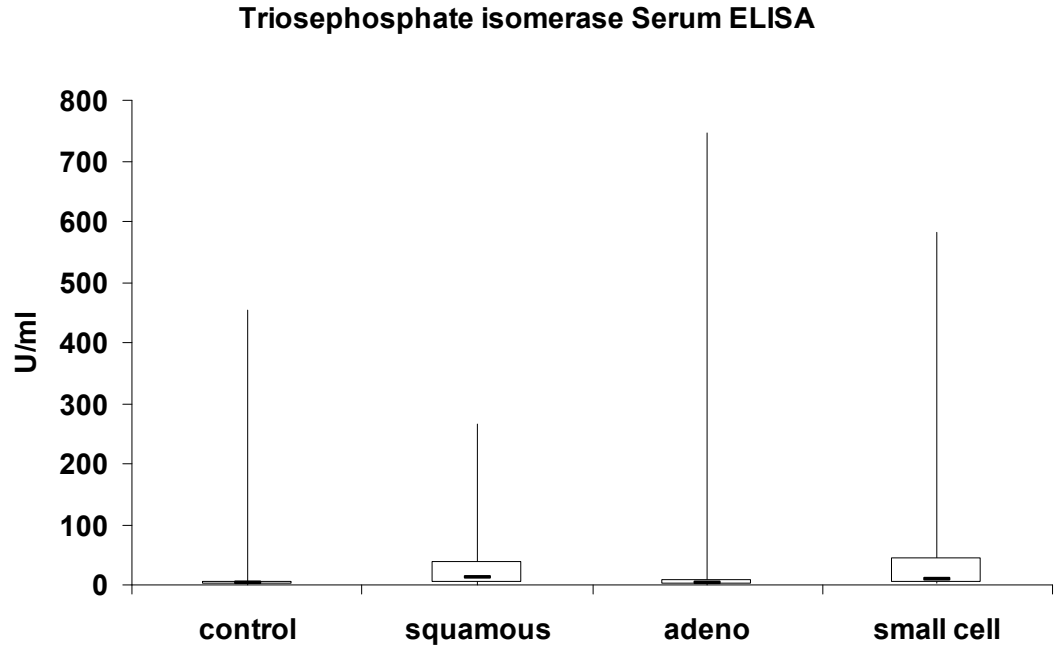


Figure 3.4.18: Triosephosphate isomerase ELISA: Bar chart and box and whisker plot showing normal compared to late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles.Patient serum samples used for this experiment included 19 normal, 18 squamous, 18 adenocarcinoma and 16 small cell carcinoma.

3.5 Immunohistochemistry Analysis of hnRNPA2B1, PKM2, and Hip

Immunohistochemistry (IHC) was performed on tissue slides to examine the staining intensity level for some of the interesting proteins from the conditioned media study. IHC can be used in the routine diagnosis of lung cancer in order to identify biological markers. In cancer diagnostics IHC can provide information on protein expression on top of tissue morphology and can be used as criteria for verification or rejection of a tentative diagnosis. Quantitative evaluation of protein levels in tumour samples is challenging, with technical issues arising related to immunohistochemistry procedure and subjectivity of assessment. Due to the significant heterogeneity that can be seen in small amounts of lung cancer tumour specimens, such as those frequently obtained in advanced lung cancer, they may not be representative of the tumour itself (Taillade, Penault-Llorca et al. 2007).

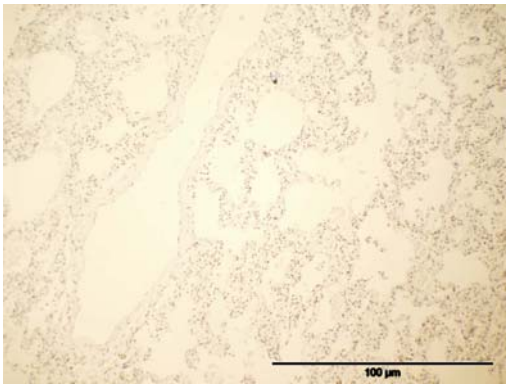
Immunohistochemistry was performed on Human Lung cancer Tissue slides using antibodies for hnRNPA2B1, HSC70-interacting protein, and pyruvate kinase M2. Optimisation determined antibody for hnRNPA2B1, Hip, and PKM2 should be diluted to 1:1600, 1:7500, and 1:50 respectively. Antibodies used were a mouse monoclonal to hnRNPA2B1 from Abcam (ab6102), a mouse monoclonal to Hip from Abcam (ab98960), and a rabbit polyclonal to PKM2 from Abcam (ab38237). The tissue slides contained tissue cores from adenocarcinoma, squamous cell carcinoma, small cell carcinoma, and normal lung tissue.

A mixed tissue slide was also used for hnRNPA2B1 immunohistochemistry analysis. This mixed tissue slide contained tissue spots for lung, colon, breast,

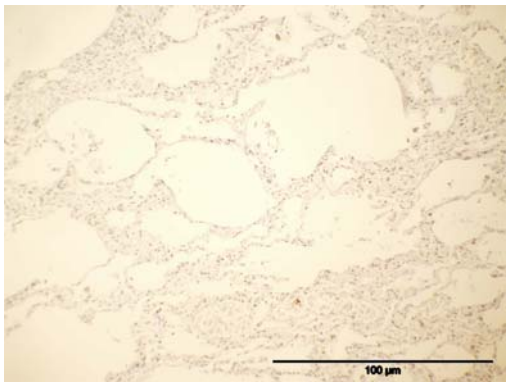
prostate, and pancreas cancers, as well as respective normal tissue (and for some cancer adjacent to normal tissue). All five cancers show similar levels with the majority of staining either intensity 3 or 2. For example 88% of colon AD, 66.66% of breast invasive duct carcinoma, 79.16% of prostate AD, 84% of pancreas AD, 100% of lung SCC and 93.33% of lung AD are scored as a 3 or a 2.

An intensity score of 1, 2, or 3, relating to staining that is weak, intermediate, or strong respectively, was assigned to each spot by a qualified person.

A.



B.



C.

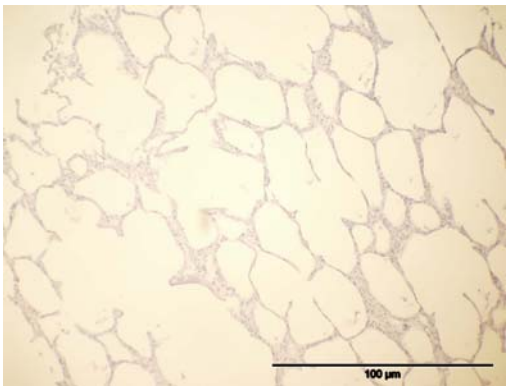
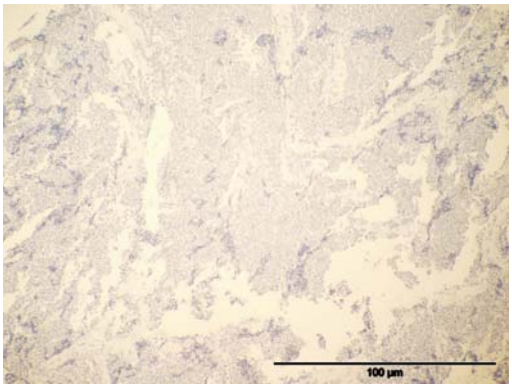


Figure 3.5.1: hnRNPA2B1 immunohistochemistry performed on lung cancer tissue array slide. Representative images show staining of normal lung tissue samples (A-C). Magnification, 40X. Scale Bar, 100µm.

A.



B.

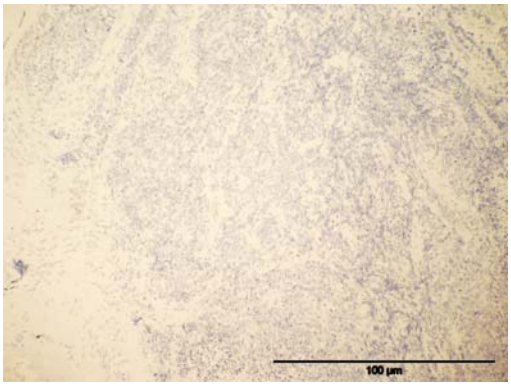
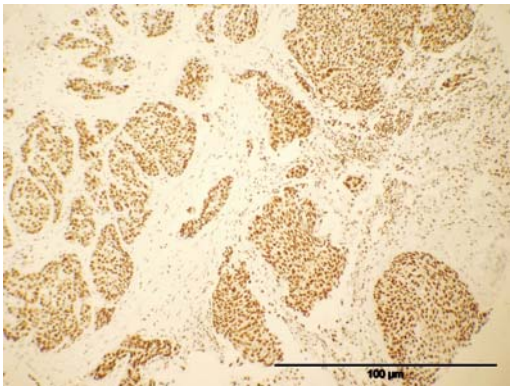
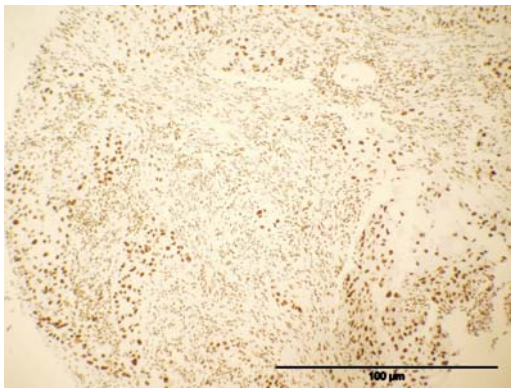


Figure 3.5.2: Staining for negative control (primary antibody omitted): Immunohistochemistry performed on lung cancer tissue array slide. Representative images show Grade 4 small cell lung cancer tissue samples (A+B). Magnification, 40X. Scale Bar, 100µm.

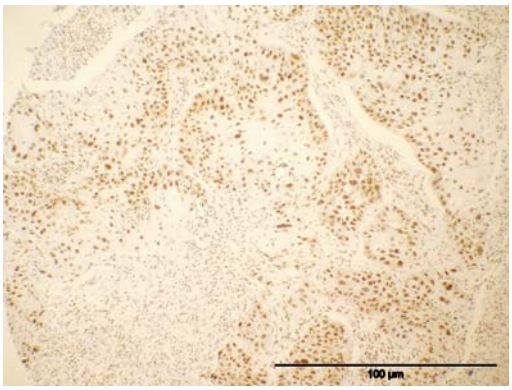
A.



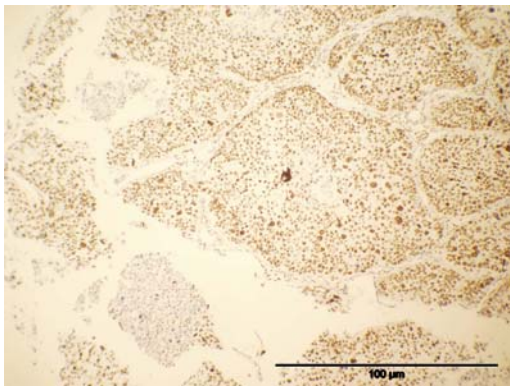
B.



C.



D.



E.

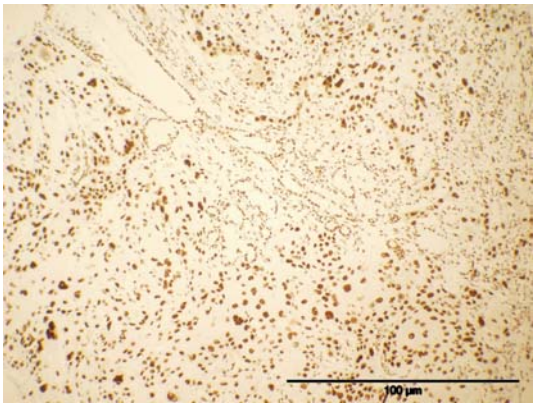
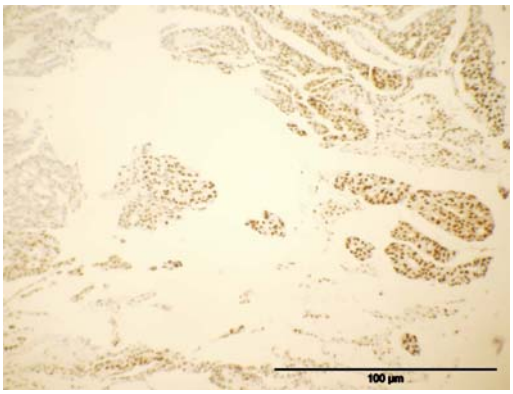
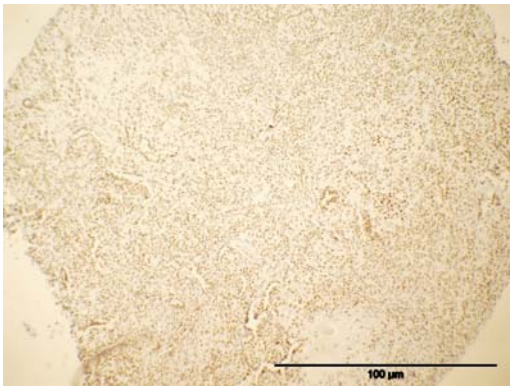


Figure 3.5.3: hnRNPA2B1 immunohistochemistry performed on lung cancer tissue array slide. Representative images show staining of squamous cell carcinoma lung cancer samples (A-E). Magnification, 40X. Scale Bar, 100µm.

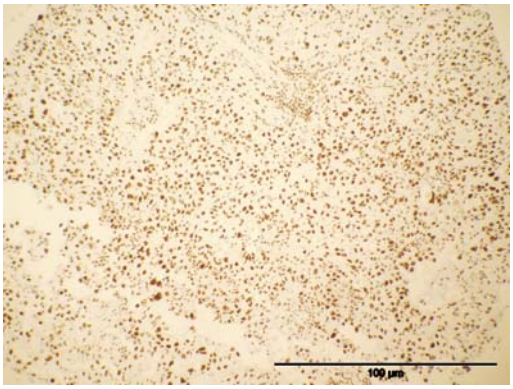
A.



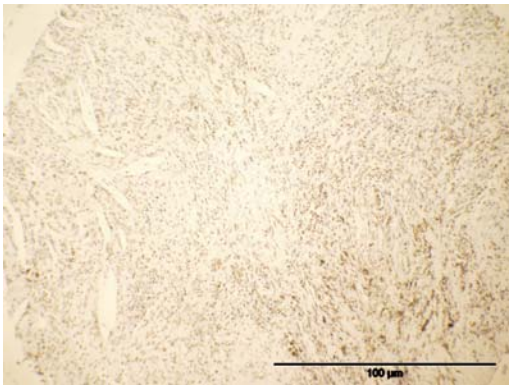
B.



C.



D.



E.

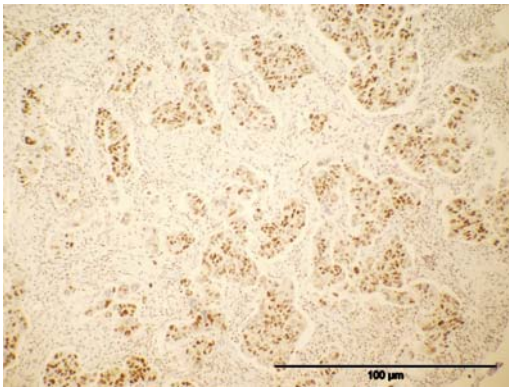
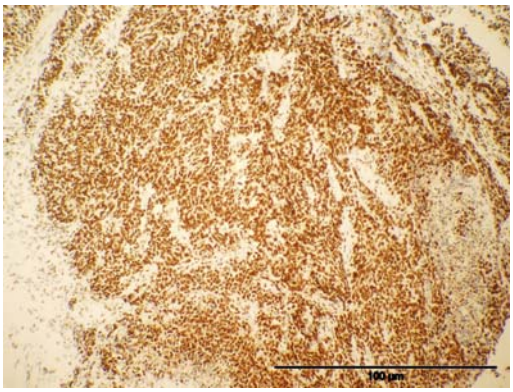
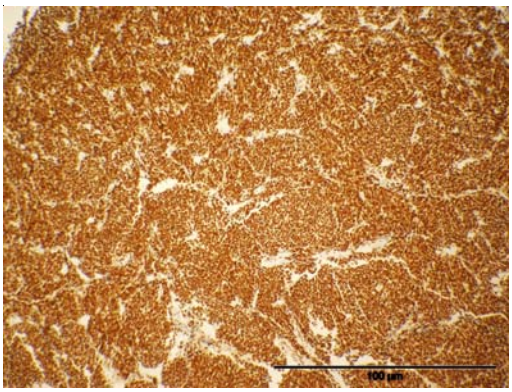


Figure 3.5.4: hnRNPA2B1 immunohistochemistry performed on lung cancer tissue array slide. Representative images (A-E) show staining of adenocarcinoma lung cancer samples. Magnification, 40X. Scale Bar, 100µm.

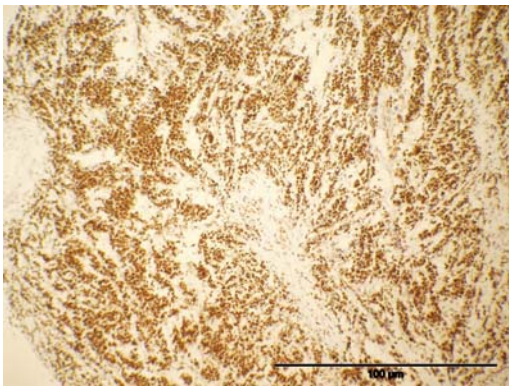
A.



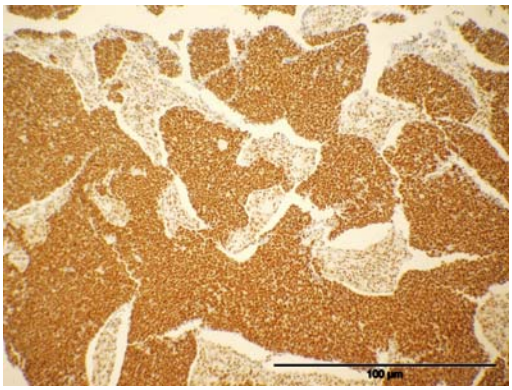
B.



C.



D.



E.

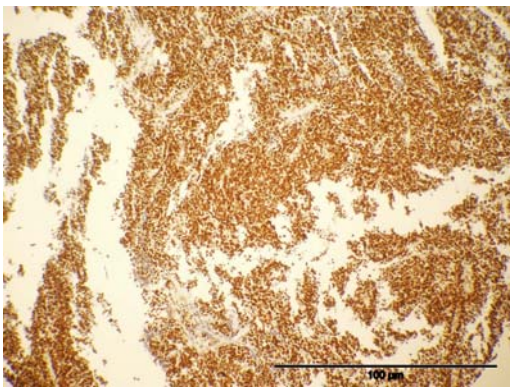


Figure 3.5.5: hnRNPA2B1 immunohistochemistry performed on lung cancer tissue array slide. Representative images show staining of Grade 4 (A-E) small cell lung cancer samples. Magnification, 40X. Scale Bar, 100µm.

Intensity Score	3	3V	2	2V	1	1V	neg	Total Cores
Squamous	3 (21.43)	2 (14.29)	3 (21.43)	2 (14.29)	2 (14.29)	1 (7.14)	1 (7.14)	14
Adeno	4 (12.5)	0	6 (18.75)	3 (9.37)	5 (15.65)	6 (18.75)	8 (25)	32
Small cell	12 (42.86)	2 (7.14)	5 (17.86)	1 (3.57)	2 (7.14)	3 (10.71)	3 (10.71)	28
Alveolar cell carcinoma	5 (38.46)	0	2 (15.38)	2 (15.38)	0	0	4 (30.77)	13
Normal lung tissue	0	0	3 (23.08)	0	2 (15.38)	0	8 (61.54)	13

Table 3.5.1: Intensity Scoring for immunohistochemistry of hnRNPA2B1 on Lung Cancer Tissue Slide. Scoring of 1, 2, and 3 relates to weak staining, intermediate staining, and strong staining, respectively. An intensity score with a "V" means there was variation in the staining. The percentage of total tissue cores with that particular intensity score is shown in brackets.

A. **B.**

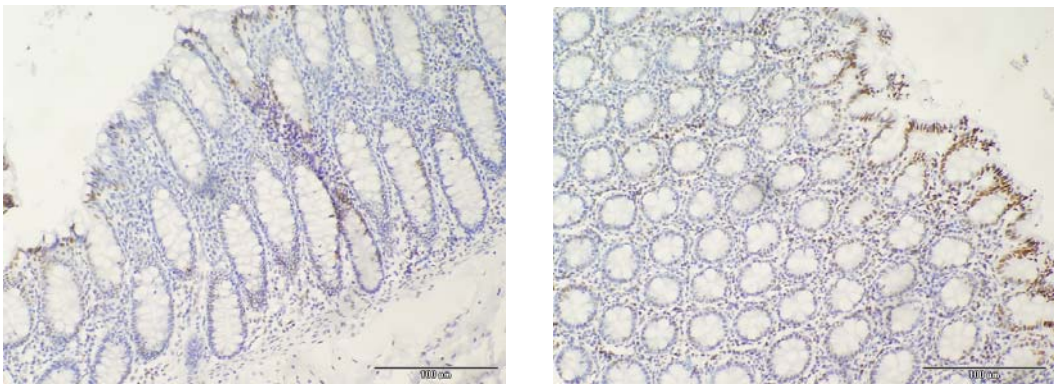


Figure 3.5.6: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of normal colon tissue (A+B). Magnification, 40X. Scale Bar, 100µm.

A. **B.**

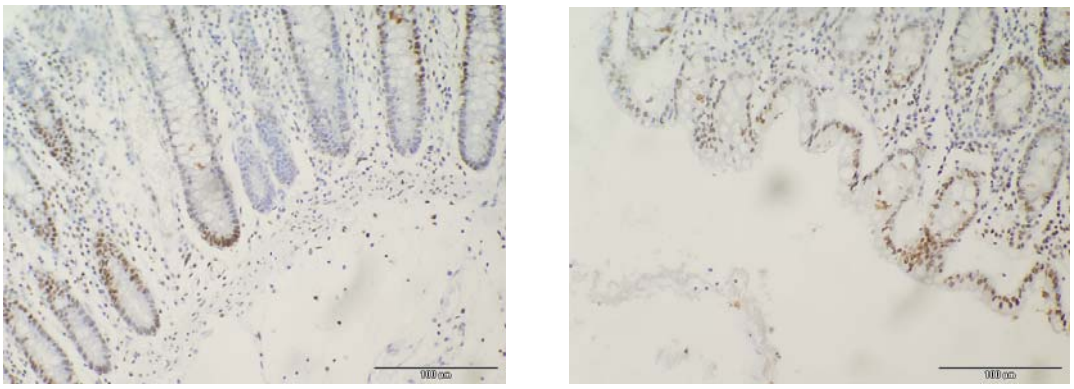
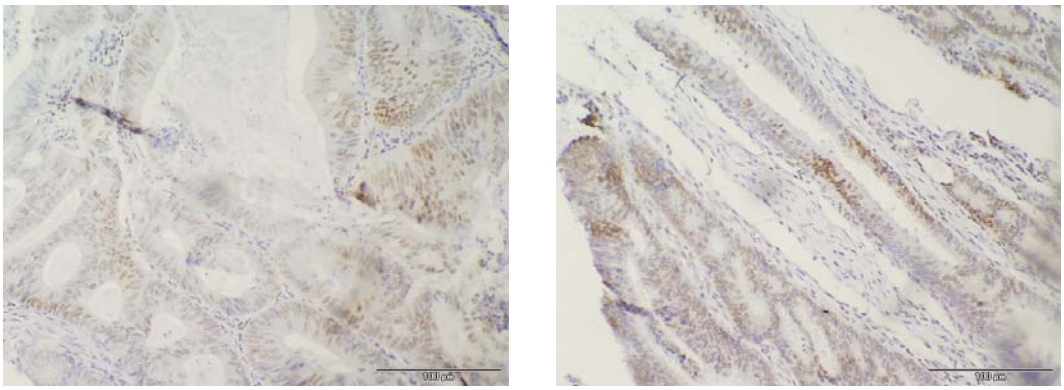


Figure 3.5.7: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of cancer tissue adjacent to normal colon tissue (A+B). Magnification, 40X. Scale Bar, 100µm.

A. **B.**



C. **D.**

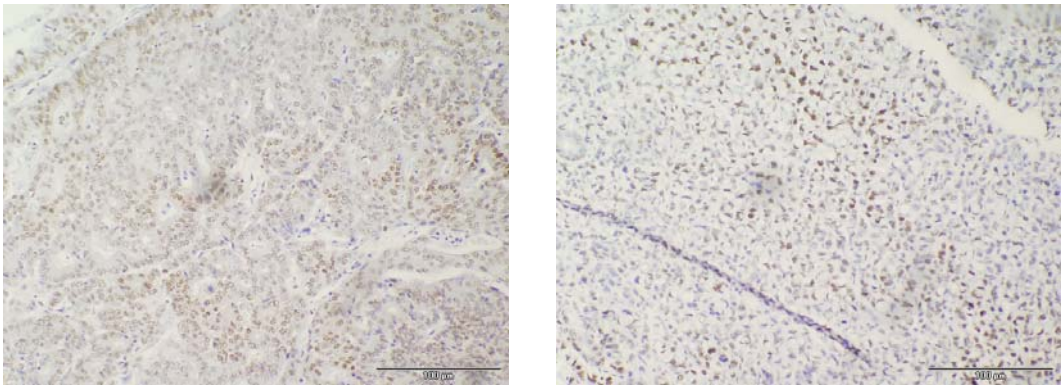


Figure 3.5.8: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of colon adenocarcinoma cancer tissue. (A) Grade 1 Stage II (B) Grade 1 Stage II (C) Grade 2 Stage II (D) Grade 3 Stage III Magnification, 40X. Scale Bar, 100µm.

A. **B.**

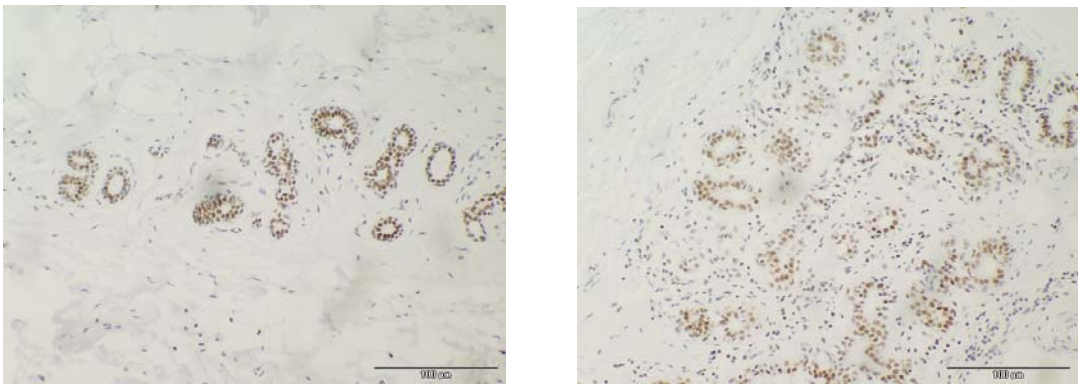


Figure 3.5.9: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of normal breast tissue (A+B). Magnification, 40X. Scale Bar, 100µm.

A. **B.**

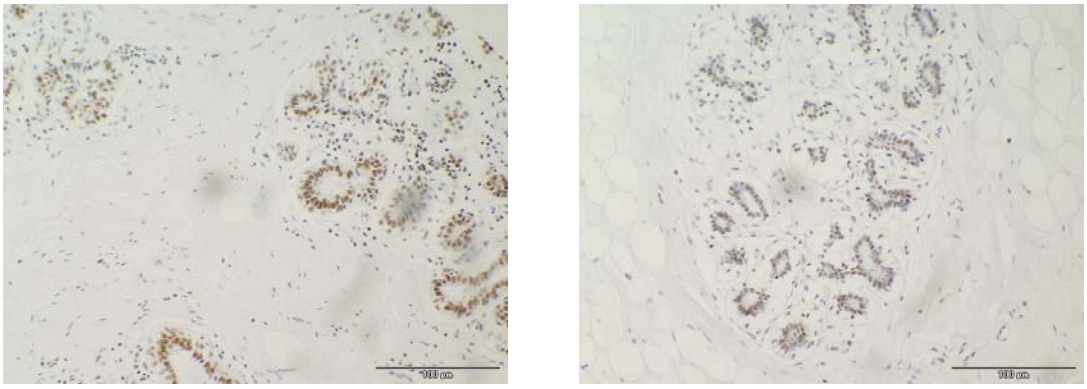
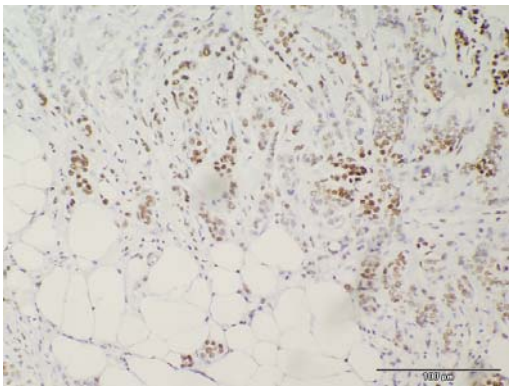
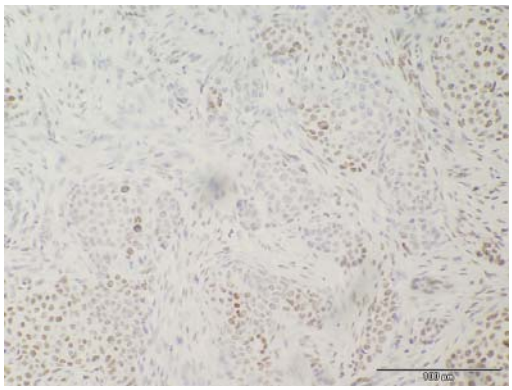


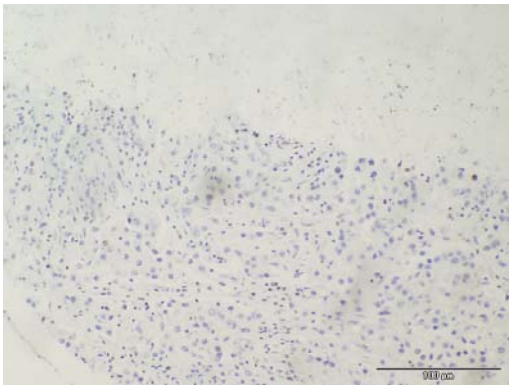
Figure 3.5.10: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of cancer tissue adjacent to normal breast tissue (A+B). Magnification, 40X. Scale Bar, 100µm.

A. **B.**





C. **D.**



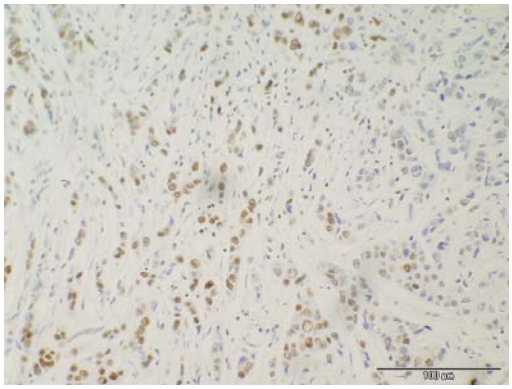


Figure 3.5.11: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of breast invasive ductal carcinoma cancer tissue. (A) Grade 1 Stage IIb (B) Grade 1 Stage IIa (C) Grade 2 Stage IIa (D) Grade 3 Stage IIb Magnification, 40X. Scale Bar, 100µm.

A. **B.**

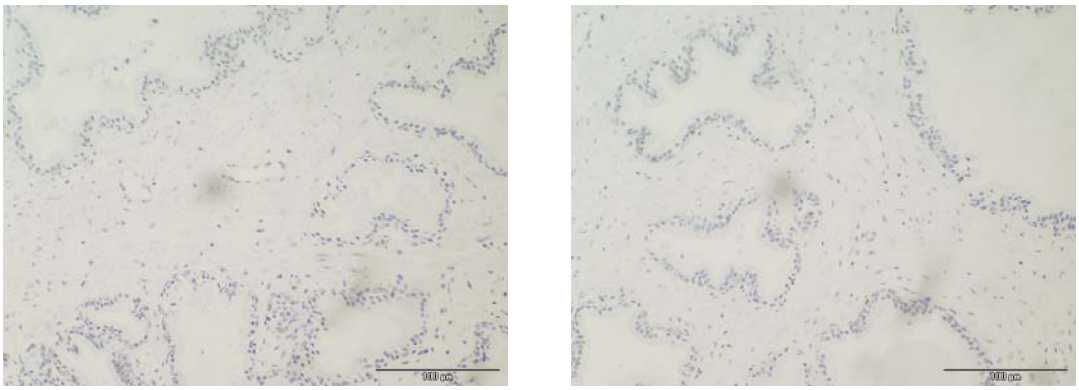


Figure 3.5.12: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of normal prostate tissue (A+B). Magnification, 40X. Scale Bar, 100µm.

A. **B.**

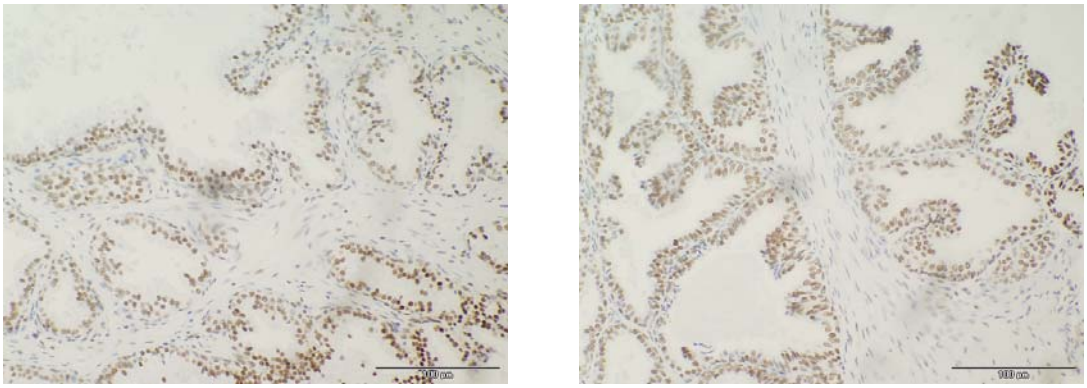
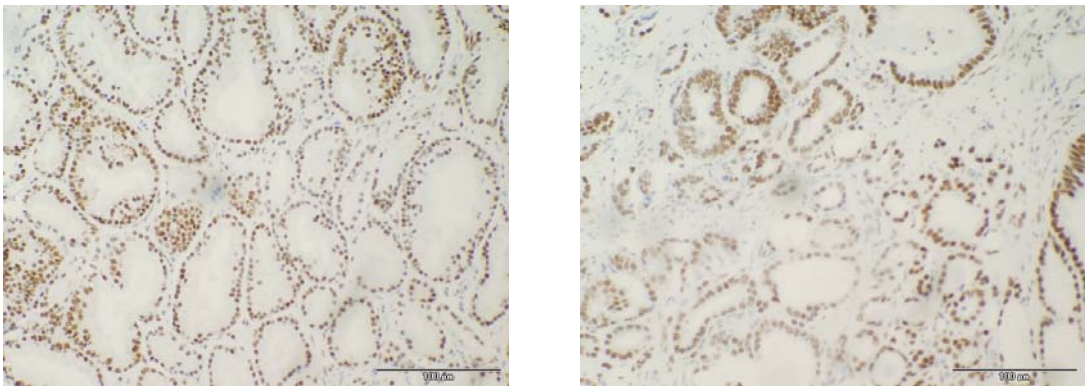


Figure 3.5.13: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of cancer adjacent to normal prostate tissue (A+B). Magnification, 40X. Scale Bar, 100µm.

A. **B.**



C. **D.**

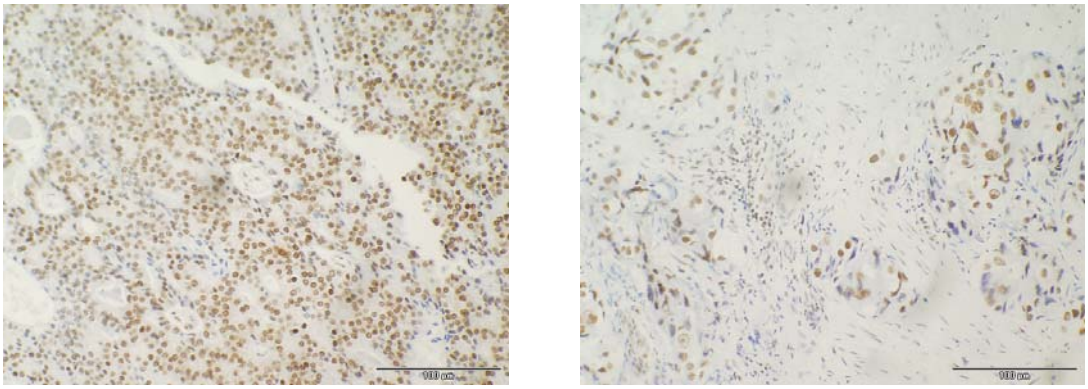
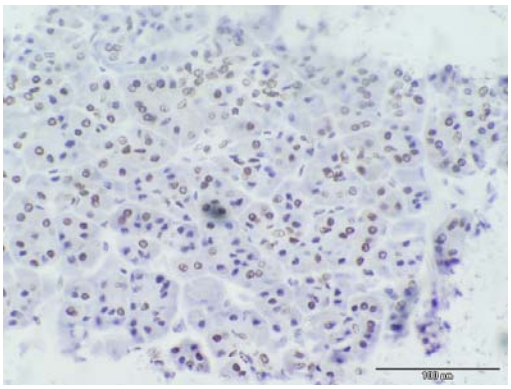
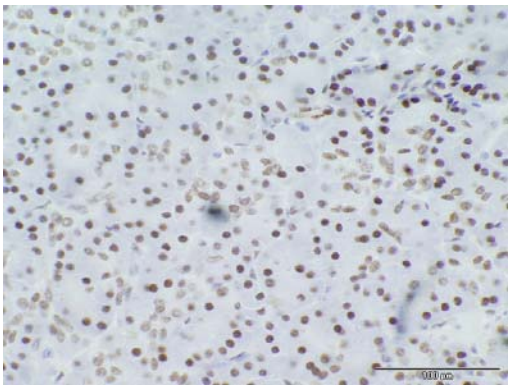


Figure 3.5.14: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of prostate adenocarcinoma cancer tissue. (A) Grade 1 Stage II (B) Grade 2 Stage III (C) Grade 2 Stage II (D) Grade 3 Stage IV Magnification, 40X. Scale Bar, 100µm.

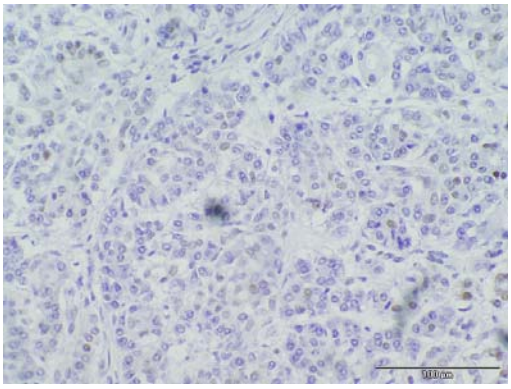
A.



B.



C.



D.

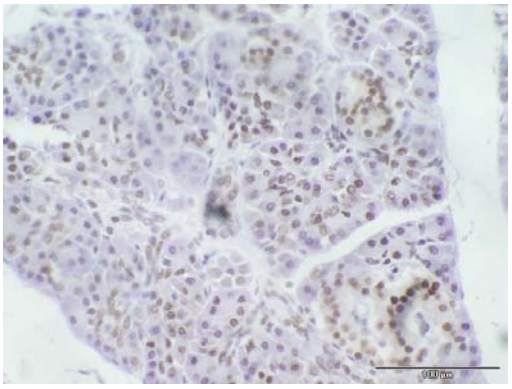
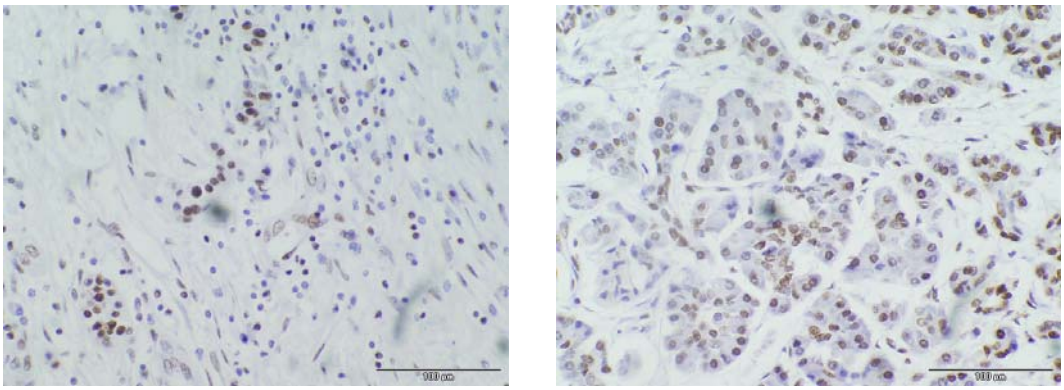


Figure 3.5.15: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of normal pancreas tissue (A-D). Magnification, 40X. Scale Bar, 100µm.

A. **B.**



C. **D.**

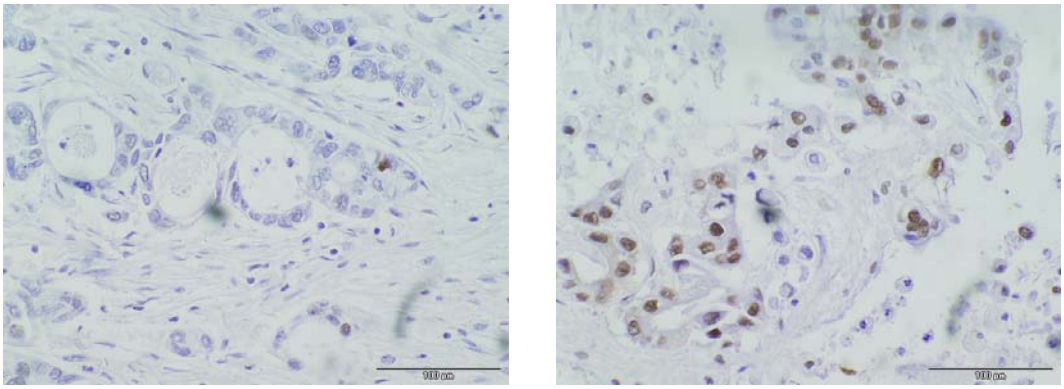
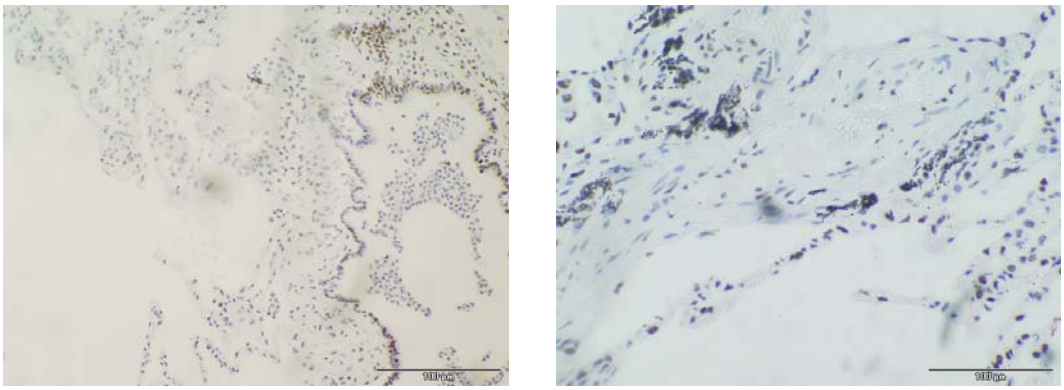


Figure 3.5.16: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of pancreas duct adenocarcinoma cancer tissue. (A) Grade 1-2 Stage II (B) Grade 1-2 Stage II (C) Grade 2 Stage II (D) Grade 2-3 Stage I Magnification, 40X. Scale Bar, 100µm.

A. **B.**



C. **D.**

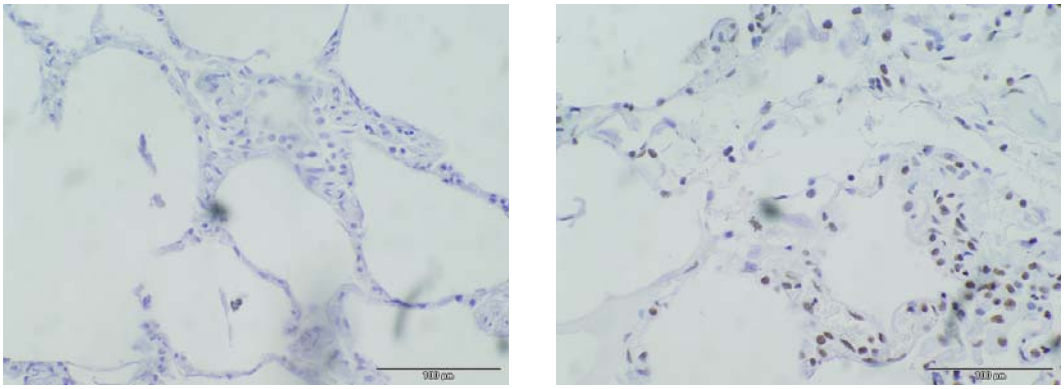
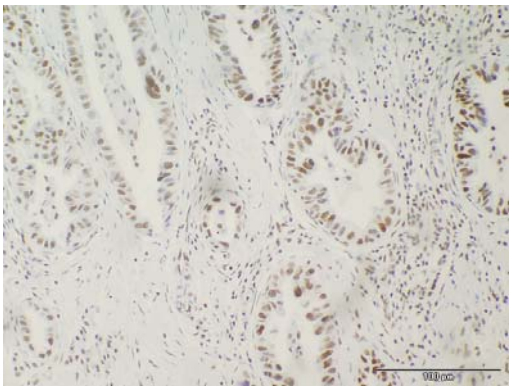
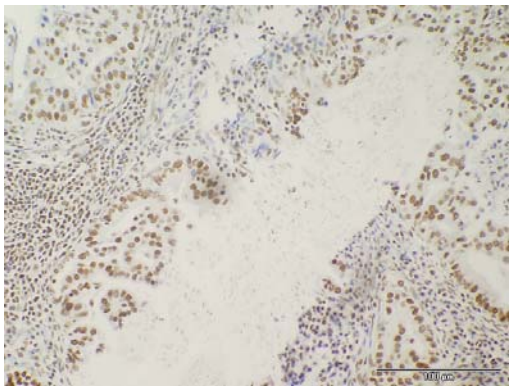


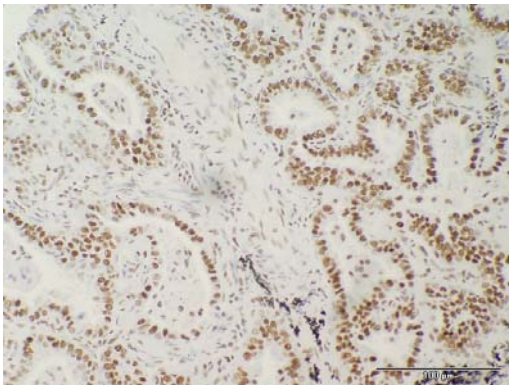
Figure 3.5.17: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of normal lung tissue. (A) staining on tissue spot in benign cancer tissue (B) staining on tissue spot in benign cancer tissue (C) normal lung tissue with congestion (D) normal lung tissue with congestion. Magnification, 40X. Scale Bar, 100µm.

A. **B.**





C. **D.**



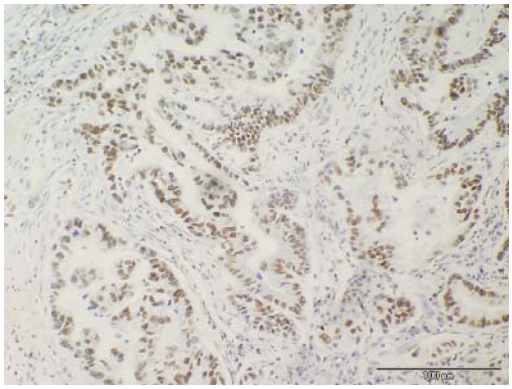
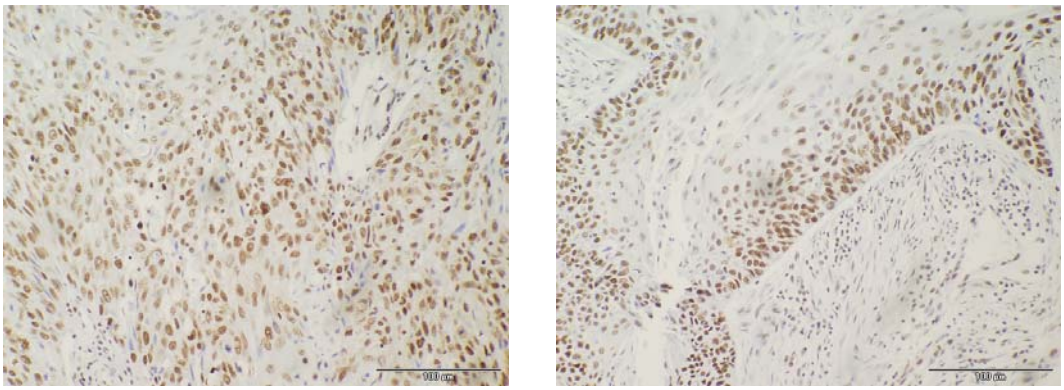


Figure 3.5.18: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of lung adenocarcinoma cancer tissue. (A) Grade 1 Stage II (B) Grade 2 Stage I (C) Grade 1 Stage IIIa (D) Grade 2 Stage IIIa. Magnification, 40X. Scale Bar, 100µm.

A. **B.**



C. **D.**

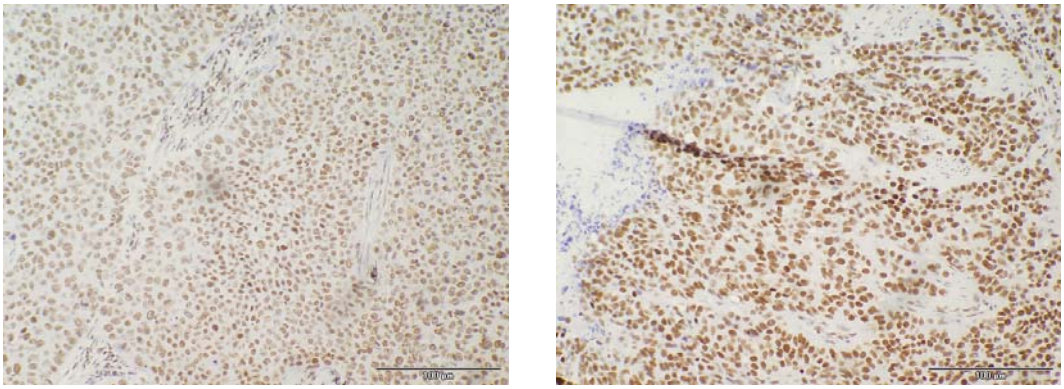
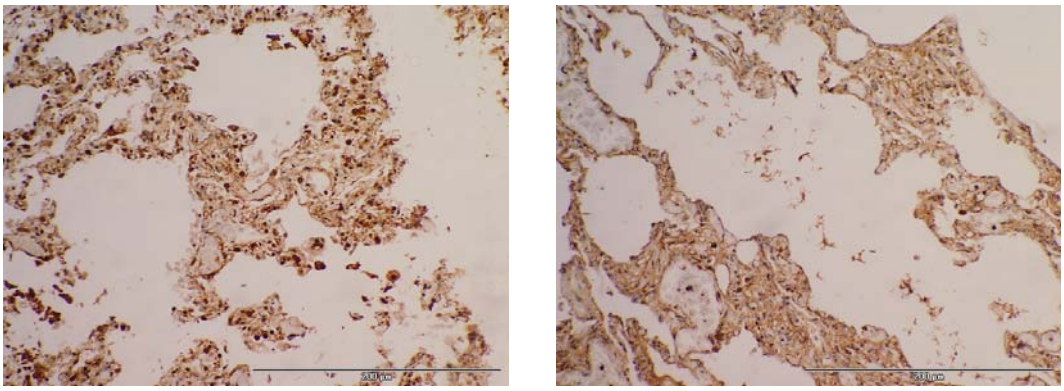


Figure 3.5.19: hnRNPA2B1 immunohistochemistry performed on mixed tissue array slide. Representative images show staining of lung squamous cell carcinoma tissue. (A) Grade 1 Stage I (B) Grade 1 Stage IIIb (C) Grade 2 Stage I (D) Grade 3 Stage II. Magnification, 40X. Scale Bar, 100µm.

Intensity Score	3	2	1	neg	Total Cores
Colon:					
Adenocarcinoma	5 (20)	17 (68)	3 (12)	0	25
Normal Colon Tissue	0	1 (25)	3 (75)	0	4
Cancer Adjacent Normal	2 (33.33)	2 (33.33)	0	2 (33.33)	6
Breast:					
Invasive ductal carcinoma	3 (12.5)	13 (54.16)	5 (20.83)	3 (12.5)	24
Normal Breast Tissue	0	4 (100)	0	0	4
Prostate:					
Adenocarcinoma	11 (45.83)	8 (33.33)	1 (4.16)	4 (16.66)	24
Normal Prostate Tissue	0	0	2 (50)	2 (50)	4
Cancer Adjacent Normal	1 (16.66)	5 (83.33)	0	0	6
Pancreas:					
Duct Adenocarcinoma	8 (32)	13 (52)	3 (12)	1 (4)	25
Normal Pancreas Tissue	0	6 (60)	2 (20)	2 (20)	10
Lung:					
Squamous	5 (50)	5 (50)	0	0	10
Adenocarcinoma	6 (40)	8 (53.33)	1 (6.66)	0	15
Normal Lung Tissue	0	1 (10)	1 (10)	8 (80)	10

Table 3.5.2: Intensity Scoring for immunohistochemistry of hnRNPA2B1 on Mixed Tissue Array Slide. Scoring of 1, 2, and 3 relates to weak staining, intermediate staining, and strong staining, respectively. The percentage of total tissue cores with that particular intensity score is shown in brackets.

A. **B.**



C. **D.**

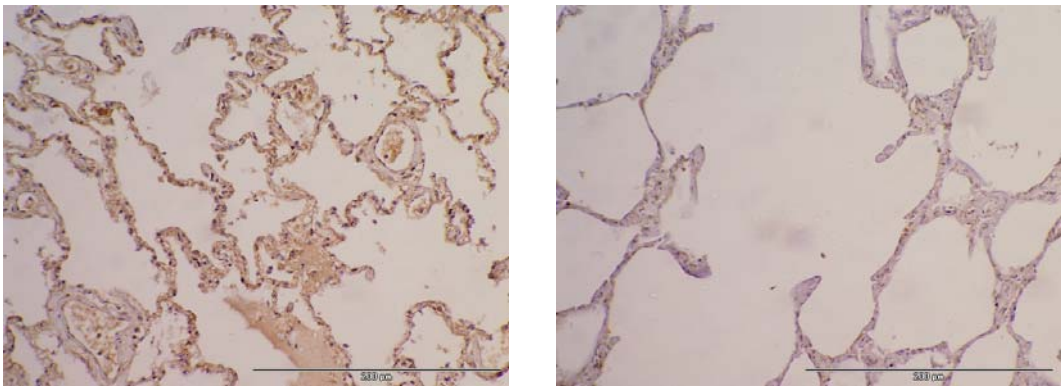
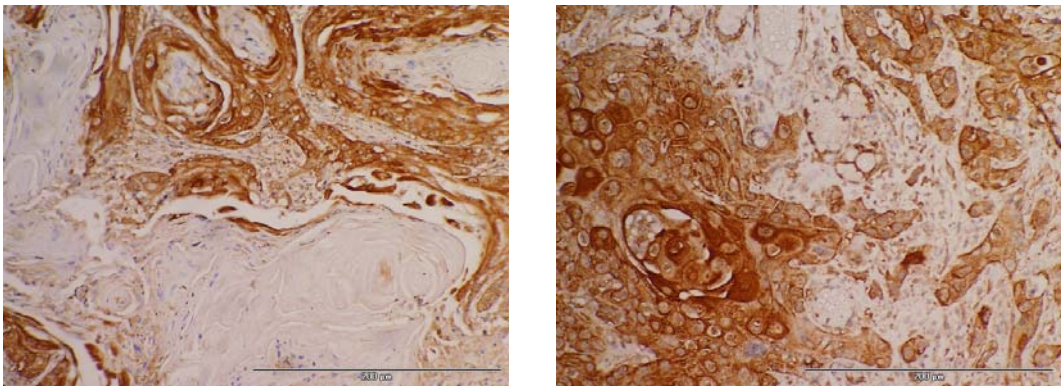


Figure 3.5.20: HSC70-interacting protein immunohistochemistry performed on lung tissue array slide. Representative images show staining of normal lung tissue (A-D). Magnification, 40X. Scale Bar, 200µm.

A. **B.**



C. **D.**

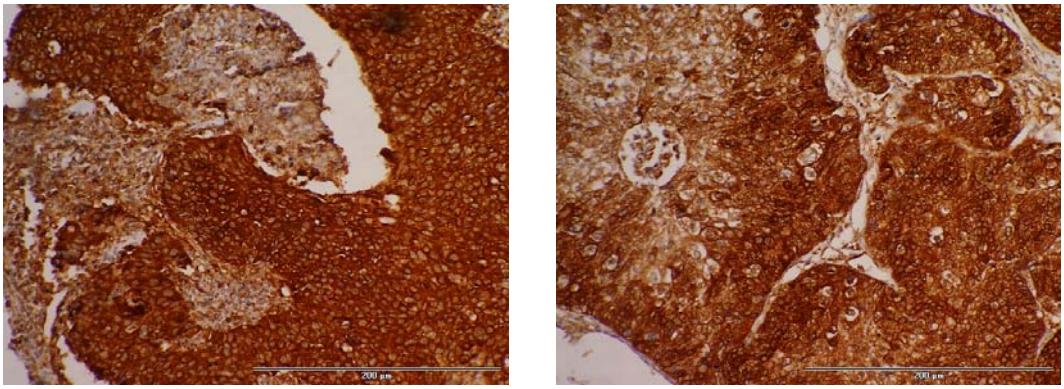
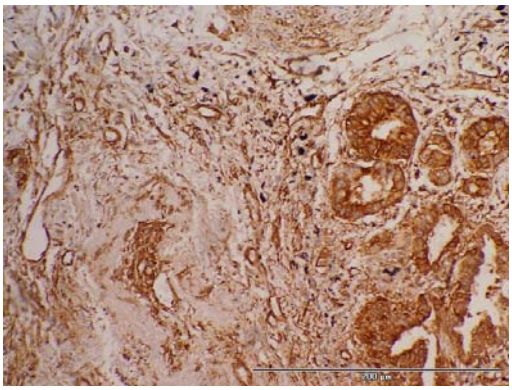
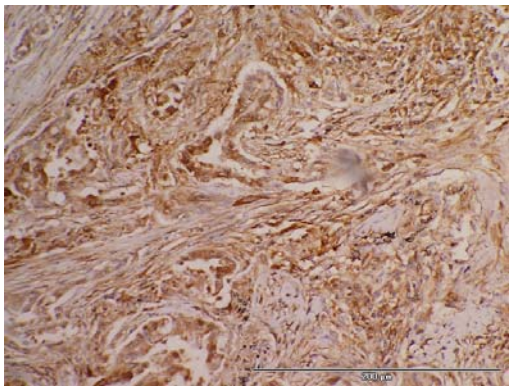


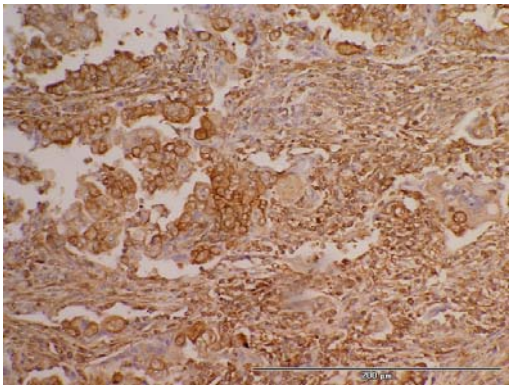
Figure 3.5.21: HSC70-interacting protein immunohistochemistry performed on lung tissue array slide. Representative images show staining of lung squamous cell carcinoma tissue. (A) Grade 1 (B) Grade 2 (C) Grade 3 (D) Grade 3. Magnification, 40X. Scale Bar, 200µm.

A. **B.**





C. **D.**



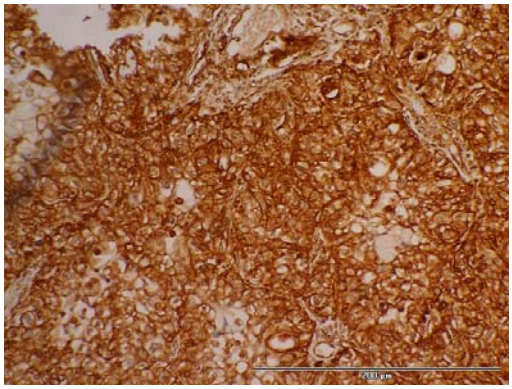
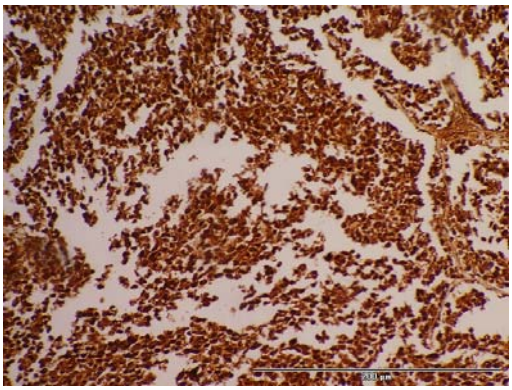
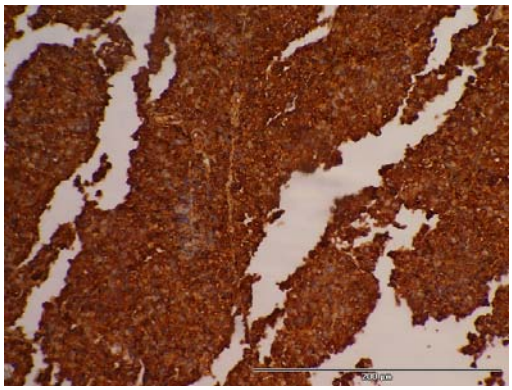


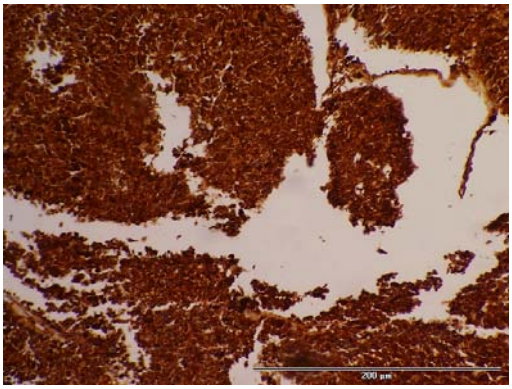
Figure 3.5.22: HSC70-interacting protein immunohistochemistry performed on lung tissue array slide. Representative images show staining of lung adenocarcinoma tissue. (A) Grade 1 (B) Grade 2 (C) Grade 3 (D) Grade 3. Magnification, 40X. Scale Bar, 200µm.

A. **B.**





C. **D.**



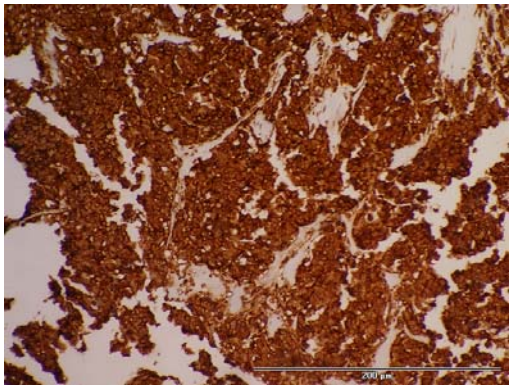


Figure 3.5.23: HSC70-interacting protein immunohistochemistry performed on lung tissue array slide. Representative images show staining of Grade 4 small cell lung carcinoma (A-D). Magnification, 40X. Scale Bar, 200µm.

Intensity Score	3	2	1	neg	Total Cores
Squamous	8 (61.54)	5 (38.46)	0	0	13
Adeno	5 (23.81)	9 (42.86)	7 (33.33)	0	21
Small cell carcinoma	15 (65.22)	4 (17.4)	4 (17.4)	0	23
Alveolar cell carcinoma	7 (53.85)	5 (38.46)	1 (7.7)	0	13
Normal Lung Tissue	0	3 (30)	7 (70)	0	10

Table 3.5.3: Intensity Scoring for immunohistochemistry of HSC70-interacting Protein on Lung Cancer Tissue Slide. Scoring of 1, 2, and 3 relates to weak staining, intermediate staining, and strong staining, respectively. The percentage of total tissue cores with that particular intensity score is shown in brackets.

A. **B.**

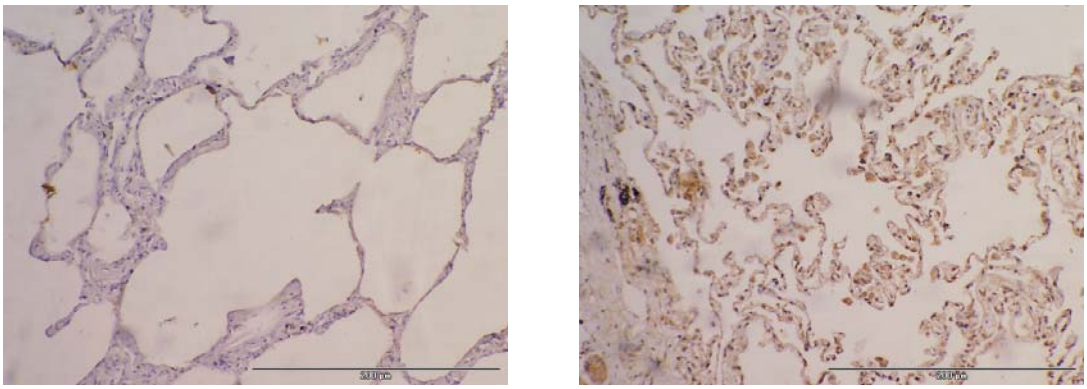


Figure 3.5.24: PKM2 immunohistochemistry performed on lung tissue array slide. Representative images show staining of normal lung tissue (A+B). Magnification, 40X. Scale Bar, 200µm.

A. **B.**

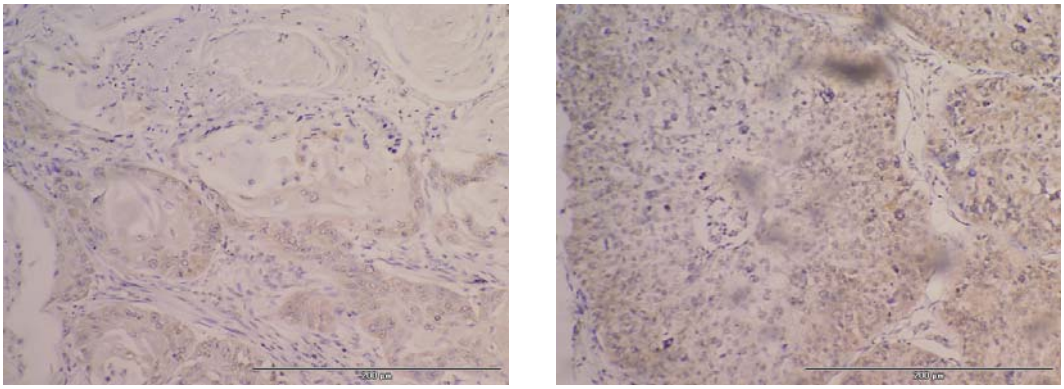


Figure 3.5.25: PKM2 immunohistochemistry performed on lung tissue array slide. Representative images show staining of squamous cell carcinoma. (A) Grade 1 (B) Grade 3. Magnification, 40X. Scale Bar, 200µm.

A. **B.**

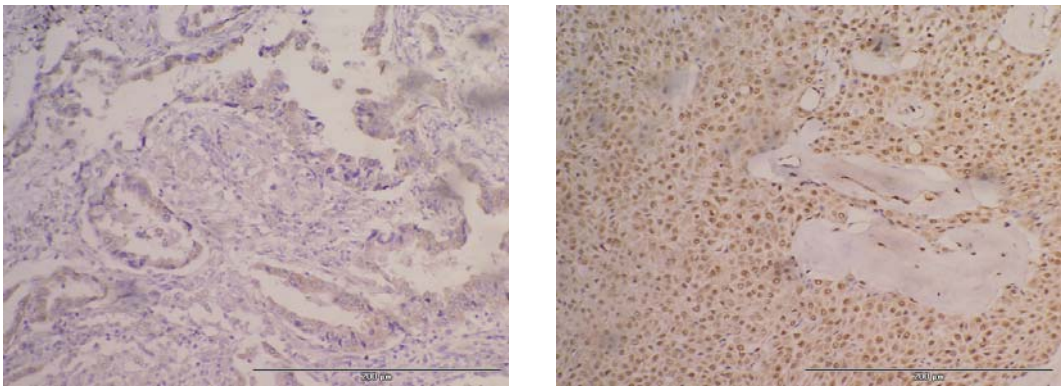


Figure 3.5.26: PKM2 immunohistochemistry performed on lung tissue array slide. Representative images show staining of Grade 3 adenocarcinoma (A+B). Magnification, 40X. Scale Bar, 200µm.

A. **B.**

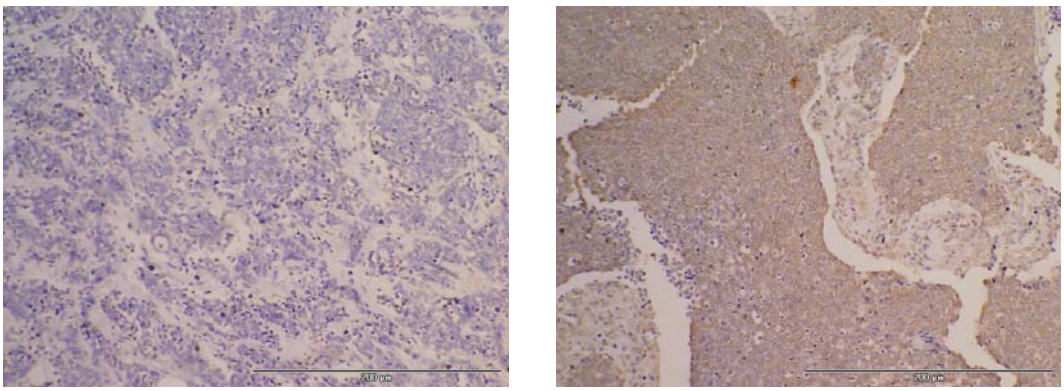


Figure 3.5.27: PKM2 immunohistochemistry performed on lung tissue array slide. Representative images show staining of Grade 4 small cell carcinoma (A+B). Magnification, 40X. Scale Bar, 200µm.

Intensity Score	3	2	1	neg	Total Cores
Squamous	2 (15.4)	5 (38.46)	5 (38.46)	1 (7.7)	13
Adeno	5 (23.81)	4 (19.05)	7 (33.33)	5 (23.81)	21
Small cell carcinoma	5 (21.74)	5 (21.74)	9 (39.13)	4 (17.4)	23
Alveolar cell carcinoma	2 (15.4)	5 (38.46)	5 (38.46)	1 (7.7)	13
Normal Lung Tissue	4 (40)	3 (30)	3 (30)	0	10

Table 3.5.4: Intensity Scoring for immunohistochemistry of PKM2 on Lung Cancer Tissue Slide. Scoring of 1, 2, and 3 relates to weak staining, intermediate staining, and strong staining, respectively. The percentage of total tissue cores with that particular intensity score is shown in brackets.

3.6 Functional Assays

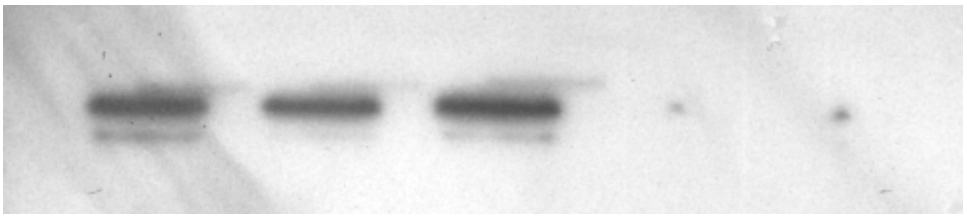
The results from the 2D-DIGE Decyder study on conditioned media, and the subsequent validation in biological fluids and tissue cores indicated the hnRNPA2B1 protein had potential as a marker for lung cancer; however these results give little indication of the role this protein may play in the progression of the cancer itself. The invasion and migration of neoplastic cells (normal cells will undergo apoptosis after detachment) allows these cells to enter lymphatic and blood vessels for dissemination into the circulation and subsequently undergo metastatic growth in distant organs. This formation of a metastasis is the main cause of death for cancer patients.

Invasion and migration assays are used to show how a change to the cancer cells will affect their ability to behave in a metastatic manner. Invasion assays give an indication of how well the cancer cells, under particular conditions, would move through tissue boundaries, a trait that separates benign from malignant lesions. Matrigel is used in the Boyden chamber to resemble the complex extracellular environment found in many tissues. Migration assays are set up the same as invasion, the only difference being (apart from the optimised cell count) that matrigel is NOT added to the chamber, and measure the motility of cancer cells under particular conditions.

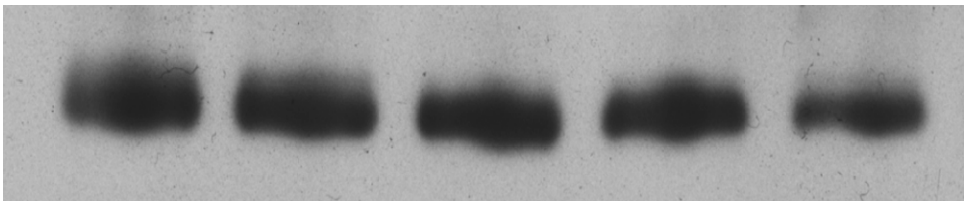
In this case the effect of knocking down the protein of interest on invasion and migration of cancer cells was examined. The lung squamous cell carcinoma cell line DLKP-M was chosen. Small interfering RNAs (siRNA) are used to suppress gene

expression of hnRNPA2B1; these 20-25 nucleotide double-stranded RNAs reduce expression of its corresponding gene. This potent inhibition of gene expression is achieved by transfection of the siRNA, a method by which the siRNA are delivered into the cancer cells.

As per Materials and Methods the transfection conditions were optimised before setting up invasion and migration assays. Control wells are used to show the effect, if any, of; (i) the experiment on untreated cells, (ii) transfection reagent on the cells, (iii) kinesin knockdown on the cells, and (iv) knocking down a random gene through the use of relative scrambled siRNA. Two siRNAs for hnRNPA2B1 were used to confirm the consistency of results. To assess whether any changes witnessed were caused by the proliferation of the cells and not as a result of gene knockdown, a proliferation assay was performed. Western blot analysis was performed to check if the knockdown was successful.



hnRNPA2B1



α-tubulin

Figure 3.6.1: Western blot analysis of whole cell lysates of DLKP-M transfected cells. A reduction in the levels of hnRNPA2B1 can be seen in the two knockdown siRNAs s6713 and s6715 when compared to the untreated control, NeoFx only and scrambled siRNA controls. (15µg protein loaded per lane).

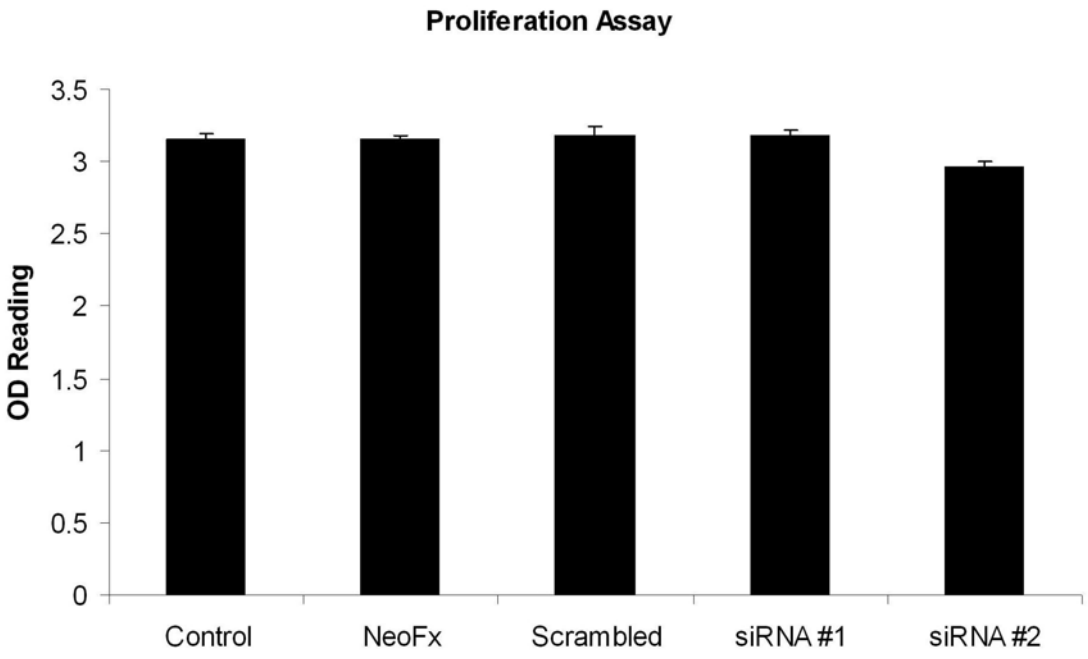


Figure 3.6.2: Proliferation assay performed on five conditions from invasion and migration assays. Proliferation of DLKP-M cells does not change under any of the conditions.

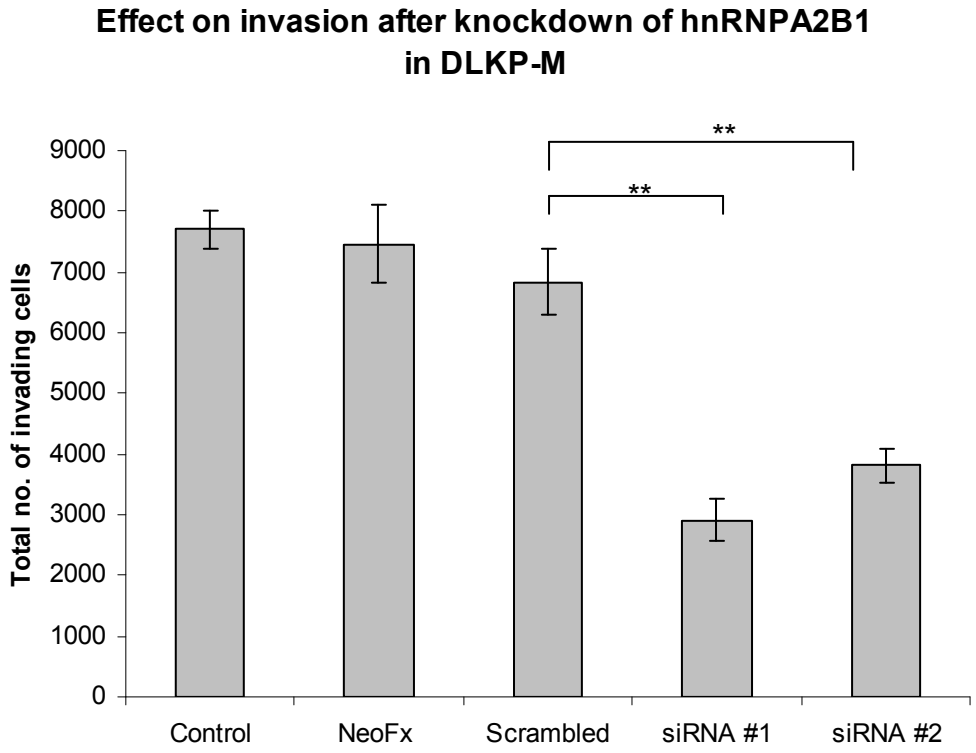
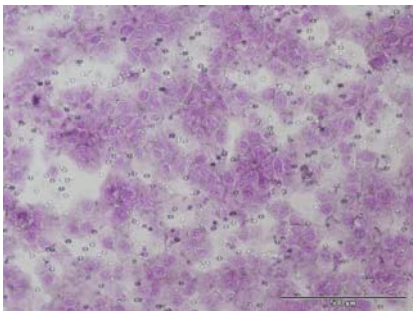
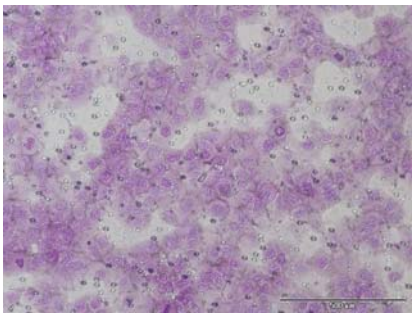


Figure 3.6.3: Histogram showing effect of hnRNPA2B1 knockdown on invasive behaviour of DLKP-M cells. Twenty-four hours post-transfection with hnRNPA2B1 siRNAs, invasion assays on DLKP-M were performed. Both siRNA #1 (s6713) and siRNA #2 (s6715) transfected cells show a significant decrease in invasion levels. T-test values; Scrambled versus siRNA#1 (*p* value = 4E-13), Scrambled versus siRNA#2 (*p* value = 6E-08).

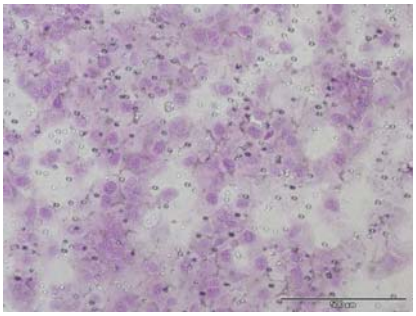
A. Untreated Control



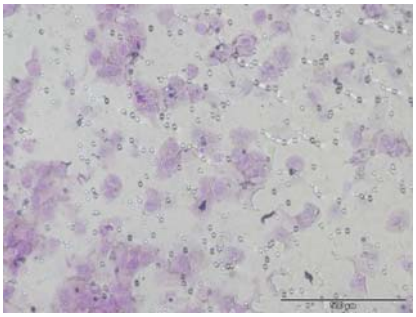
B. NeoFx Only Control



C. Scrambled siRNA Control



D. hnRNPA2B1 siRNA #1



E. hnRNPA2B1 siRNA #2

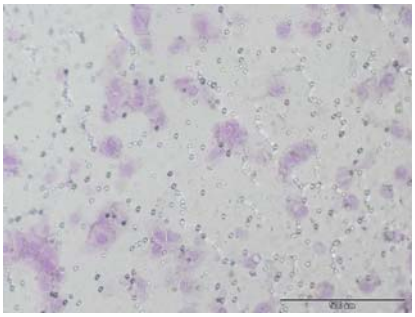


Figure 3.6.4: Representative photomicrographs showing invasion status of DLKP-M cells following transfection of hnRNPA2B1 siRNAs. (A) Untreated control (B) NeoFx only control (C) Scrambled siRNA control (D) hnRNPA2B1 siRNA #1 (E) hnRNPA2B1 siRNA #2. Inhibition of invasion can be seen in D and E. Magnification 40X. Scale Bar, 500µm.

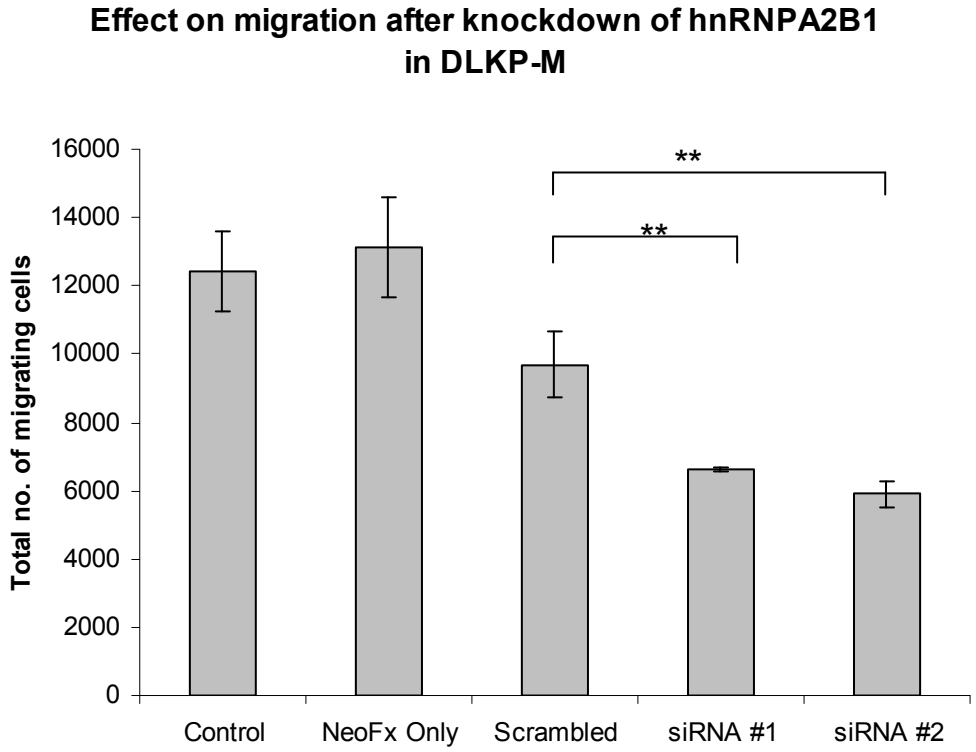
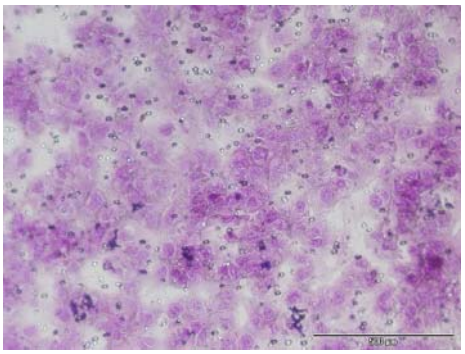
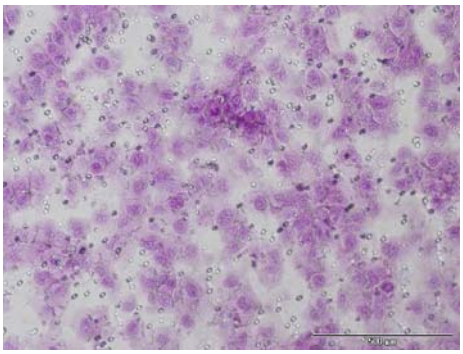


Figure 3.6.5: Histogram showing effect of hnRNPA2B1 knockdown on migration behaviour of DLKP-M cells. Twenty-four hours post-transfection with hnRNPA2B1 siRNAs, migration assays on DLKP-M were performed. Both siRNA #1 (s6713) and siRNA #2 (s6715) transfected cells show a decrease in migration levels. T-test score; Scrambled versus siRNA#1 (*p* value = 2E-06), Scrambled versus siRNA#2 (*p* value = 4E-08).

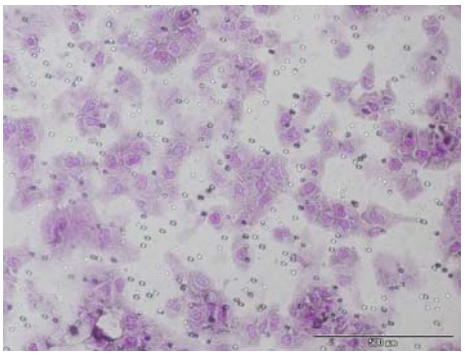
A. Untreated Control



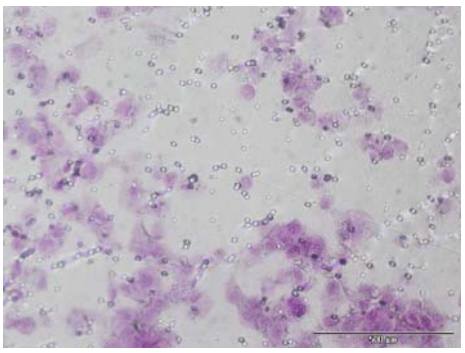
B. NeoFx Only Control



C. Scrambled siRNA Control



D. hnRNPA2B1 siRNA #1



E. hnRNPA2B1 siRNA #2

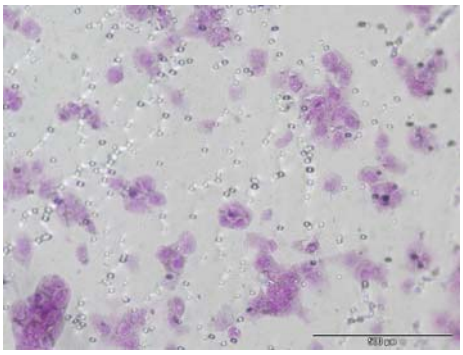


Figure 3.6.6: Representative photomicrographs showing migration status of DLKP-M cells following transfection of hnRNPA2B1 siRNAs. (A) Untreated control (B) NeoFx only control (C) Scrambled siRNA control (D) hnRNPA2B1 siRNA #1 (E) hnRNPA2B1 siRNA #2. Inhibition of migration can be seen in D and E, but also in the scrambled control C. Magnification 40X. Scale Bar, 500µm.

3.6.1 Effect of knockdown of hnRNPA2B1 on Invasion and Migration of DLKP-M lung Cancer Cells.

Knockdown of the hnRNPA2B1 gene in lung cancer cell line DLKP-M using two separate siRNAs was successful (Fig. 3.6.1), and had little effect on proliferation of cancer cells (Fig. 3.6.2).

The knockdown of hnRNPA2B1 had an effect on both invasion and migration. For the two siRNAs used, s6713 and s6715, it resulted in a 57% and a 44% decrease in invasion respectively when compared to the scrambled control (Fig. 3.6.3) and a 32% and 39% decrease in migration respectively when compared to scrambled control (Fig. 3.6.5).

hnRNPA2B1 certainly has a role to play in the migration and invasion of lung cancer cells further indicating its potential as a biomarker and also perhaps as a drug target to inhibit tumour growth.

CHAPTER FOUR

LUNG CANCER BIOMARKER DISCOVERY FROM 2D-DIGE OF IMMUNODEPLETED SERUM

4.1 Background

The 2D-DIGE analysis of conditioned media of lung cancer cell lines proved to be a relatively useful model for the discovery of lung cancer biomarkers. However, as this was just a model the ratio difference between abundance levels of proteins shown in the DyCyder software analysis between comparisons did not translate accurately to ELISA/biochemical assay analysis of clinical samples. This result was not unusual and was somewhat expected. Cell lines grown in culture will never secrete and/or leak proteins into the conditioned media in the same manner a tumour would into the microenvironment *in vivo*. Add to this the added complexity of serum compared to the conditioned media and it is understandable that potential biomarkers from *in vitro* models have no real significance until validated in biological samples or tumour tissue.

Instead of depending on a model, examining the biological sample directly (in this case serum) may prove more useful for finding biomarkers that will validate later in larger cohorts of clinical samples. Instead of performing 2D-DIGE on conditioned media why not perform 2D-DIGE on serum directly? Unfortunately, however, serum has far more complex protein content than conditioned media. High abundant proteins such as albumin and transferrin in the serum will mask any low abundant markers and have to be removed. For this immunodepletion kits are used, such as ProteoPrep 20, which specifically removes 20 of the most abundant proteins from human serum or plasma using small recombinant immunoaffinity ligands and conventional antibodies. The immunodepletion kit is supposed to remove 97-98% of total protein mass in human serum

4.2 2D-DIGE of Immunodepleted Serum

An initial 2D-DIGE experiment was performed on ten immunodepleted serum samples from adenocarcinoma patients and ten immunodepleted serum samples from normal healthy patients. Five normal and five adenocarcinoma samples were labelled with the Cy3 fluorophore and five normal and five adenocarcinoma samples were labelled with the Cy5 fluorophore to remove bias. Unfortunately the only proteins identified were high abundant proteins (See Table 4.2.1). A number of proteins were identified more than once, showing varying fold changes, for example alpha-1-antitrypsin, haptoglobin, leucine-rich alpha-2-glycoprotein, transferrin, and zinc-alpha-2-glycoprotein. This was an indication that a number of isoforms from the same protein had been identified.

The most interesting result from Table 4.2.1 (following page) was that for the haptoglobin isoforms. The six isoforms of haptoglobin in adenocarcinoma serum showed higher fold changes when compared to their counterparts in normal serum. Interestingly the highest fold change was seen in the isoform of highest molecular weight, the second highest fold change in the isoform of second highest molecular weight, and so on for the six isoforms. Alpha-1-trypsin showed a similar trend but for only four isoforms. Transferrin also showed a trend for four isoforms though these were up-regulated in normal compared to adenocarcinoma. The haptoglobin isoforms were also the most statistically significant.

Protein	T-test	Fold Change
Alpha-2-HS-glycoprotein	0.0011	-1.96
Alpha-1-Antitrypsin	0.00011	4.41
Alpha-1-Antitrypsin	0.00049	3.33
Alpha-1-Antitrypsin	0.00066	3.21
Alpha-1-Antitrypsin	0.0041	2.3
Haptoglobin (Isoform one)	5.90E-05	7.74
Haptoglobin (Isoform two)	5.90E-05	6.49
Haptoglobin (Isoform three)	5.90E-05	6
Haptoglobin (Isoform four)	5.90E-05	5.4
Haptoglobin (Isoform five)	6.80E-05	4.59
Haptoglobin (Isoform six)	0.00013	3.73
Human zinc-alpha-2-glycoprotein	0.0072	1.76
Human zinc-alpha-2-glycoprotein	0.036	1.59
Leucine-rich alpha-2-glycoprotein	0.00049	2.38
Leucine-rich alpha-2-glycoprotein	0.0011	2.39
Leucine-rich alpha-2-glycoprotein	0.0011	2.23
Leucine-rich alpha-2-glycoprotein	0.0016	2.39
Transferrin	0.034	-1.6
Transferrin	0.0043	-1.9
Transferrin	0.011	-1.89
Transferrin	0.026	-1.57

Table 4.2.1: 2D-DIGE Analysis of Immunodepleted Adenocarcinoma and Normal Serum. Multiple identifications for the same protein indicate multiple isoforms of that protein. A positive fold change indicates an increase in lung adenocarcinoma and a negative fold change indicates a decrease in adenocarcinoma for this protein, or its isoform, compared to normal control serum.

4.3 Analysis of Haptoglobin Isoforms in Raw Serum using Coomaisse Stained Gels

The haptoglobin isoform pattern seen in Table 4.2.1 could possibly be used itself as a biomarker. 2D-DIGE and immunodepletion however are time-consuming and expensive so could not be applied routinely in a clinical setting. Ideally this pattern could be picked up in gels stained with Coomaisse blue, a far quicker and cheaper alternative to 2D-DIGE for protein staining/labelling. If diluted raw serum was used (10µl of serum diluted in 440µl of rehydration buffer) again it would be a quicker and cheaper alternative to using immunodepletion columns, the high dilution hopefully lessening the masking effects of the high abundant proteins.

In total forty-eight 2D gels were stained with coomaisse before being scanned with the Typhoon and imported into the DeCyder software. The gels included eight normal, eight lung adenocarcinoma, eight lung squamous, eight breast cancer, and eight colorectal cancer serum samples. A table showing the fold change for the six isoforms in each cancer type compared to normal was created (Table 4.3.1).

An ELISA was performed in serum on haptoglobin (Fig. 4.3.3). The average quantity of Hp in normal serum was 566 ug/ml whereas in lung cancer it was more than four times this quantity for each of the conditions; 2311ug/ml in SCC, 2313ug/ml in AD, and 2662ug/ml in SCLC, while the median, min, and max results were significantly different, showing results of, respectively, 539ug/ml, 189ug/ml, and 1148ug/ml for normal, 1675ug/ml, 484ug/ml, and 8449ug/ml for SCC, 2126ug/ml, 307ug/ml, and 5690ug/ml for AD, and 1711ug/ml, 806ug/ml, and 9151ug/ml for SCLC.

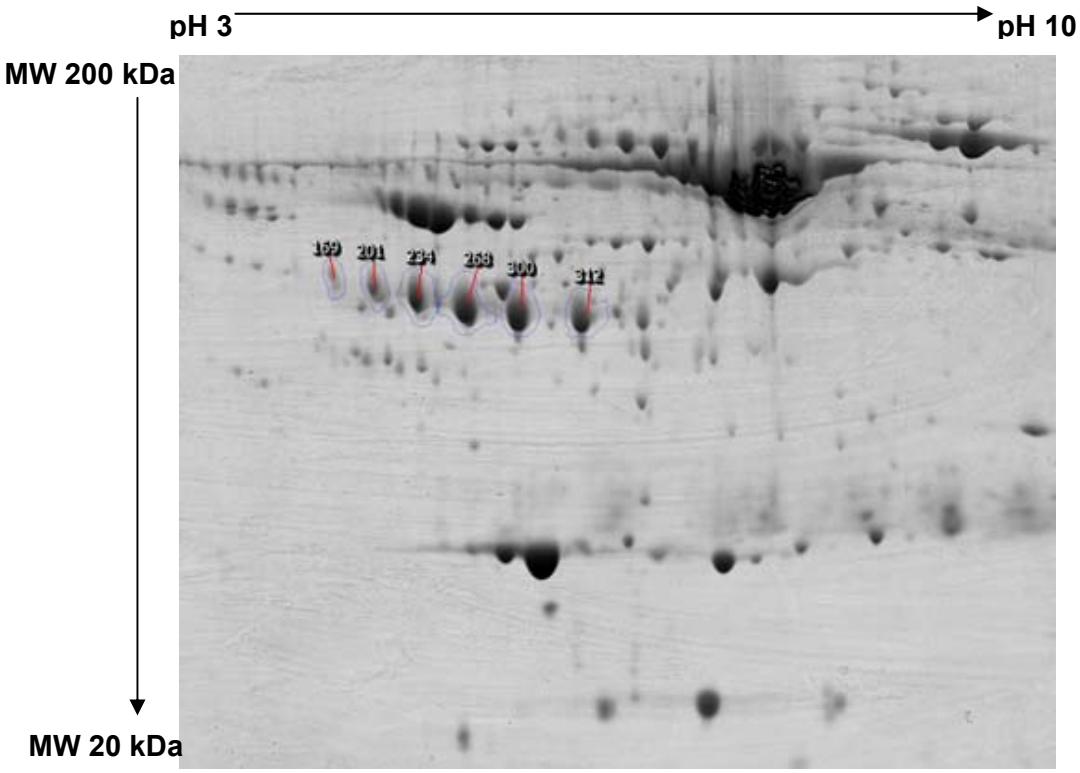


Figure 4.3.1: Progenesis analysis of coomaisse-blue stained 2D gel of diluted raw serum highlighted the six isoforms of haptoglobin as identified by mass spectrometry. In order of isoform of highest molecular weight to lowest the six isoforms from 1-6 in order are spots #169, #201, #234, #268, #300, and #312.

To confirm the six isoforms of haptoglobin were glycosylated a glycostain was performed. Also an ELISA for haptoglobin was performed on normal serum and lung cancer serum for squamous, adenocarcinoma, and small cell carcinoma (Figure 4.3.3).

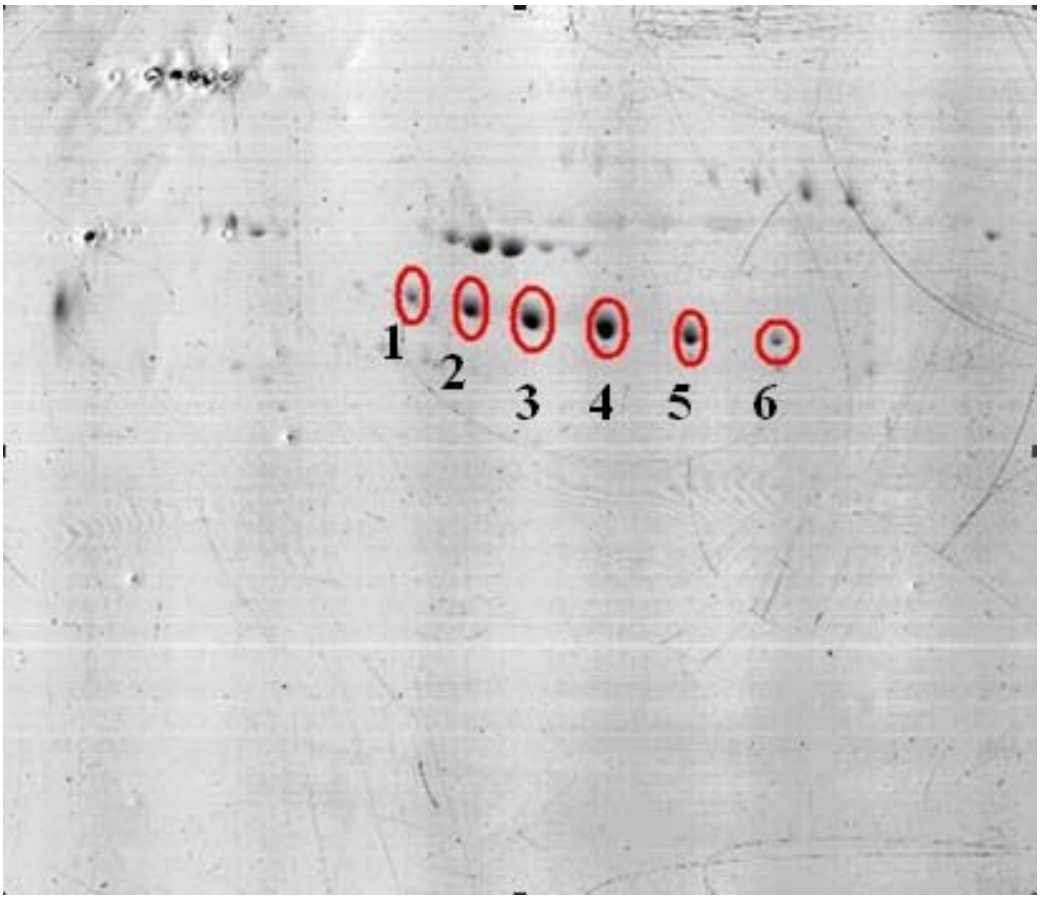


Figure 4.3.2: Glycostain of diluted raw serum sample ran on 2D-gel stained with Pierce Glycoprotein Staining Kit. Six haptoglobin isoforms circled in red indicating that they are glycosylated.

	Squamous	Adeno	Small Cell	Breast	Colon
Isoform 1	2.1 (0.079)	1.4 (0.077)	1.1 (0.768)	1.6 (0.308)	1.4 (0.392)
Isoform 2	2.1 (0.044)	1.5 (0.119)	1.2 (0.481)	1.5 (0.423)	1.4 (0.447)
Isoform 3	2 (0.041)	1.6 (0.033)	1.2 (0.263)	1.3 (0.442)	1.5 (0.161)
Isoform 4	1.7 (0.077)	1.6 (0.042)	1.4 (0.08)	1.4 (0.492)	1.6 (0.061)
Isoform 5	1.5 (0.097)	1.5 (0.042)	1.3 (0.155)	1 (0.812)	1.5 (0.086)
Isoform 6	1.1 (0.472)	1.4 (0.206)	1.1 (0.331)	1 (0.835)	1.2 (0.600)

Table 4.3.1: Haptoglobin isoforms: Fold change for each haptoglobin isoform for each cancer type compared to normal healthy control for Coomaisse stained gels using diluted raw serum that had not been immunodepleted. P value is recorded in brackets.

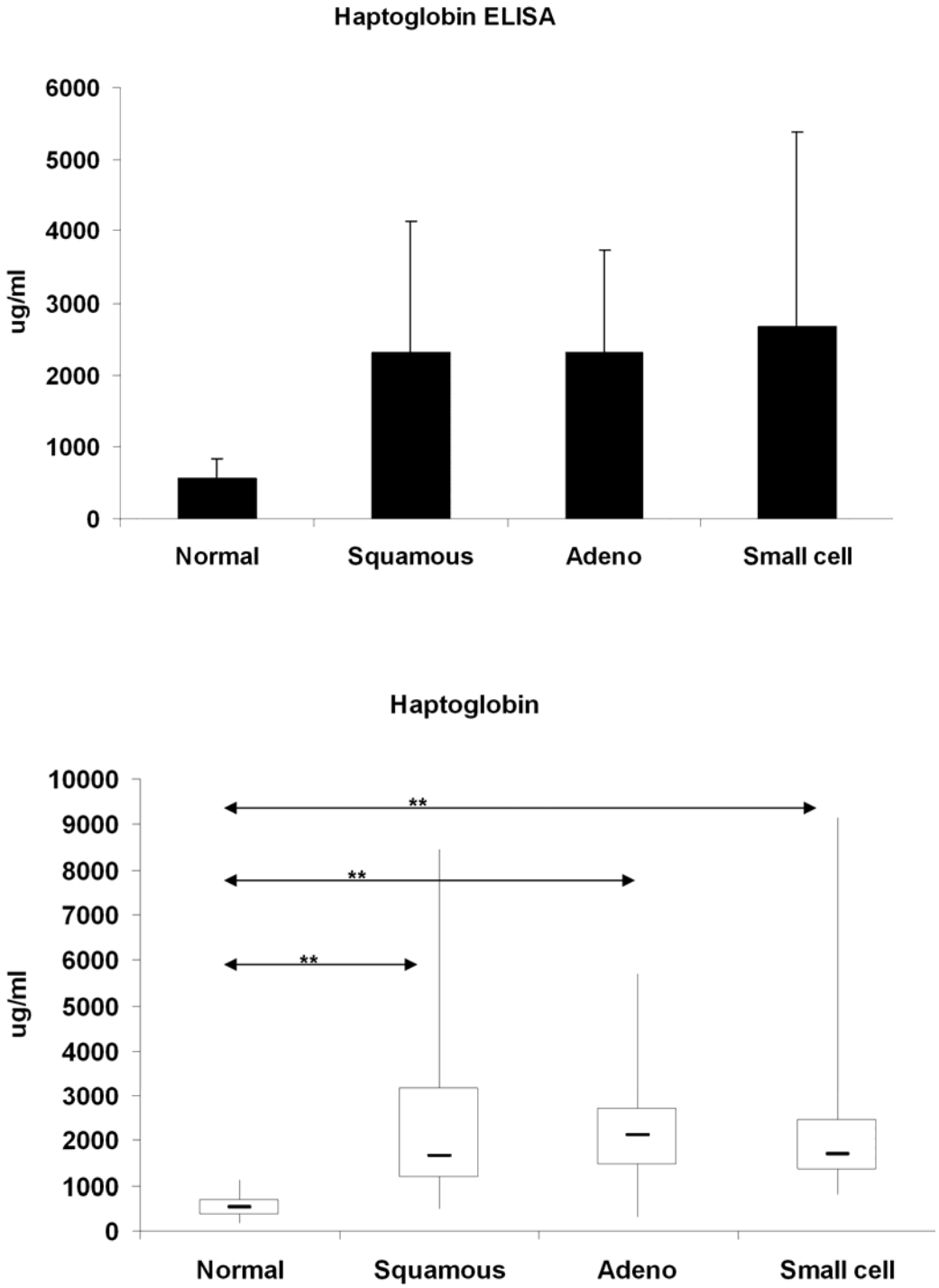


Figure 4.3.3: Haptoglobin ELISA: Bar chart and Box and Whisker Plot for normal compared to each lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum (not diluted or immunodepleted) samples used for this experiment included 20 normal, 21 squamous, 25 adenocarcinoma, and 8 small cell carcinoma. T-test scores; Normal v SCC (*p* value=0.0001), Normal v AD (*p* value=2E-06), Normal v SCLC (*p* value=0.002).

CHAPTER FIVE

LUNG CANCER BIOMARKER DISCOVERY FROM LABEL-FREE MASS SPECTROMETRY ANALYSIS OF PROTEOMINER SERUM

5.1 Background

Unlike immunodepletion, the ProteoMiner technique decreases the amount of high-abundant proteins while preventing the loss of proteins bound to them. It enriches and concentrates low-abundance proteins by decreasing the dynamic range of the overall protein concentration. It uses a large library of hexapeptide ligands with a limited number of binding sites that act as unique binders for proteins. Whereas immunodepletion will indiscriminately remove the high abundant proteins and any smaller proteins bound to it, ProteoMiner attempts to equalise all proteins throughout the sample.

In protein-labelling approaches different protein samples are combined together once labelling is finished and after sample preparation steps the pooled mixtures are analysed by a single LC-MS/MS experiment. In contrast, the label-free quantitative method involves preparing each sample separately and then running them individually by LC-MS/MS.

Progenesis LC-MS software allows statistical analysis of the label-free data. Each peptide ion in the experiment is automatically detected and quantified, even in complex samples such as serum which will have overlapping peptide ions. Peptide ion outlines are automatically applied to all runs providing a complete data set for analysis. Peptide ions are automatically ranked by ANOVA value and can be grouped in multiple ways depending on the data set you want to examine. Using a multivariate statistical analysis tool allows confident data exploration and visualisation of differentially expressed peptide ions. The user can validate and select MS/MS spectra for searches in protein databases such as Mascot or

SEQUEST. The imported search results are displayed alongside the differential analysis results. Identifications can then be filtered in a number of ways.

The software will provide a confidence score for the identified protein, as well as the number of peptides used for the identification. Proteins of interest can be tagged and exported as a group and a simple dropdown menu allows you to view each of your created comparison groups. The software shows the fold change between two groups and also indicates which group the fold change is increased or decreased in.

5.2 Progenesis Analysis of Label-Free Spectrometry on ProteoMiner Serum

The ProteoMiner technique was performed on ten serum samples for each of the following groups: normal, squamous cell carcinoma, adenocarcinoma, and small cell carcinoma. ProteoMiner samples underwent a 2D-clean-up to remove impurities followed by reduction, alkylation, digestion and a C-18 clean-up before being dried in speed vac and re-suspended in 40µl of 0.1% TFA in 2% Acn.

The Progenesis software that analyses results from label-free mass spectrometry generates lists of any conceivable comparisons, providing details on protein description, fold change, ANOVA value, confidence scores and number of peptides identified.

Accession	Peptides	Confidence	Anova (p)*	Fold change	Highest mean	Lowest mean	Description
P08670	1	64.28	0.050096357	**Infinity**	Lung Cancer	Normal	Vimentin
P07900	2	124.41	0.004490815	**20.31174351**	Lung Cancer	Normal	Heat shock protein HSP 90-alpha (HSP 86)
P04406	3	189.33	0.000986774	**12.46814311**	Lung Cancer	Normal	Glyceraldehyde-3-phosphate dehydrogenase
P02741	5	327.88	0.000199056	**12.38099907**	Lung Cancer	Normal	C-reactive protein precursor
P02787	12	759.43	5.99E-06	**11.51786162**	Normal	Lung Cancer	Serotransferrin precursor (Transferrin) (Siderophilin)
P02675	4	266.77	3.18E-05	**10.82589561**	Lung Cancer	Normal	Fibrinogen beta chain precursor
P02679	5	299.02	7.52E-05	**9.380321635**	Lung Cancer	Normal	Fibrinogen gamma chain precursor
P02790	2	107.69	0.001906784	**7.63513751**	Normal	Lung Cancer	Hemopexin precursor (Beta-1B-glycoprotein)
P02042	3	154.86	0.006614393	**7.218389773**	Normal	Lung Cancer	Hemoglobin subunit delta (Hemoglobin delta chain)
P06702	2	103.23	0.008782703	**4.778579307**	Lung Cancer	Normal	Protein S100-A9 (S100 calcium-binding protein A9)
P05156	1	81.7	0.007680159	**4.672195548**	Normal	Lung Cancer	Complement factor I precursor
P02652	2	209.48	0.003912751	**4.648122747**	Normal	Lung Cancer	Apolipoprotein A-II precursor (Apo-AII) (ApoA-II)
P02763	3	193.49	0.013053506	**4.310582007**	Normal	Lung Cancer	Alpha-1-acid glycoprotein 1 precursor (AGP 1) (Orosomucoid-1)
P10643	3	236.76	9.69E-06	**4.141475609**	Lung Cancer	Normal	Complement component C7 precursor
P01031	17	1043.78	6.83E-07	**4.069212425**	Lung Cancer	Normal	Complement C5 precursor
P01042	3	276.06	0.001041968	**3.468002367**	Normal	Lung Cancer	Kininogen-1 precursor (Alpha-2-thiol proteinase inhibitor)
P01023	10	674.68	0.020248803	**3.358464882**	Normal	Lung Cancer	Alpha-2-macroglobulin precursor (Alpha-2-M)
P02749	1	58.52	0.050323042	**3.32987504**	Normal	Lung Cancer	Beta-2-glycoprotein 1 precursor (Beta-2-glycoprotein I)
P19652	4	227.83	0.004587166	**3.294184059**	Normal	Lung Cancer	Alpha-1-acid glycoprotein 2 precursor (AGP 2)
P04196	5	342.3	5.39E-05	**3.25914217**	Normal	Lung Cancer	Histidine-rich glycoprotein precursor
P13671	4	215.68	0.000230245	**3.201675042**	Lung Cancer	Normal	Complement component C6 precursor
P02671	4	313.41	6.06E-05	**3.14507898**	Lung Cancer	Normal	Fibrinogen alpha chain precursor [Contains: Fibrinopeptide A]
P02748	10	677.56	1.69E-06	**3.128730081**	Lung Cancer	Normal	Complement component C9 precursor
P00738	6	353.91	0.000908213	**3.124602487**	Normal	Lung Cancer	Haptoglobin precursor
P02788	2	97.32	0.033839476	**3.009739383**	Lung Cancer	Normal	Lactotransferrin precursor
Q9Y238	1	43.39	0.013392696	**3.003353723**	Normal	Lung Cancer	Deleted in lung and esophageal cancer protein 1 (Deleted in lung cancer protein 1)
P02776	2	173.13	0.002776719	**2.94911617**	Normal	Lung Cancer	Platelet factor 4 precursor (PF-4) (CXCL4) (Oncostatin A) (Iroplact)
P00739	7	459.78	0.000166653	**2.885688187**	Normal	Lung Cancer	Haptoglobin-related protein precursor
P07358	3	160.19	2.12E-05	**2.774572774**	Lung Cancer	Normal	Complement component C8 beta chain precursor
P43652	2	102.75	0.000391901	**2.76600959**	Normal	Lung Cancer	Afamin precursor (Alpha-albumin) (Alpha-Alb)
P07360	2	106.66	4.25E-05	**2.734154226**	Lung Cancer	Normal	Complement component C8 gamma chain precursor
P00488	1	40.57	0.050661387	**2.667265174**	Normal	Lung Cancer	Coagulation factor XIII A chain precursor
P08697	2	109.72	0.003134112	**2.615252497**	Normal	Lung Cancer	Alpha-2-antiplasmin precursor (Alpha-2-plasmin inhibitor)
P06396	5	319.07	0.000239505	**2.608247046**	Normal	Lung Cancer	Gelsolin precursor (Actin-depolymerizing factor) (ADF) (Brevin) (AGEL)
P0C0L5	74	7226.76	1.12E-06	**2.599508943**	Lung Cancer	Normal	Complement C4-B precursor (Basic complement C4)
P07357	3	220.11	0.000206097	**2.408675298**	Lung Cancer	Normal	Complement component C8 alpha chain precursor
P07996	19	1382.02	8.62E-05	**2.278723301**	Normal	Lung Cancer	Thrombospondin-1 precursor
P02656	2	264.47	0.056950684	**2.217475399**	Normal	Lung Cancer	Apolipoprotein C-III precursor (Apo-CIII) (ApoC-III)
O14791	1	41.15	0.037597504	**2.177409003**	Normal	Lung Cancer	Apolipoprotein-L1 precursor (Apolipoprotein L-I) (Apolipoprotein L)

P04114	33	2270.23	0.002511154	**2.148358718**	Normal	Lung Cancer	Apolipoprotein B-100 precursor (Apo B-100)
Q15127	1	42.66	0.015963347	**2.086450413**	Normal	Lung Cancer	Phosphatidylinositol-glycan-specific phospholipase D 2 precursor
P01008	20	1364	0.000901754	**2.065675174**	Normal	Lung Cancer	Antithrombin-III precursor (ATIII)
P02765	2	147.69	0.004598383	**2.032232867**	Normal	Lung Cancer	Alpha-2-HS-glycoprotein precursor (Fetuin-A)
Q9UBC5	1	41.51	0.028556287	**1.936088123**	Lung Cancer	Normal	Myosin-Ia (Brush border myosin I) (BBM-I) (BBMI) (Myosin I heavy chain) (MIHC)
P02766	6	550.42	0.009639595	**1.886622488**	Normal	Lung Cancer	Transthyretin precursor (Prealbumin) (TBPA) (TTR) (ATTR)
P09871	1	43.66	0.009493832	**1.873163084**	Lung Cancer	Normal	Complement C1s subcomponent precursor (EC 3.4.21.42) (C1 esterase)
P18428	2	100.85	0.032022353	**1.868830114**	Lung Cancer	Normal	Lipopolysaccharide-binding protein precursor (LBP)
O95445	1	65.4	0.003809261	**1.864938637**	Normal	Lung Cancer	Apolipoprotein M (Apo-M) (ApoM) (Protein G3a)
P00747	3	155.37	0.002352292	**1.852312525**	Normal	Lung Cancer	Plasminogen precursor (EC 3.4.21.7)
P01024	65	6063.37	0.000287116	**1.827471851**	Lung Cancer	Normal	Complement C3 precursor
O75636	3	181.06	0.013903106	**1.823894203**	Normal	Lung Cancer	Ficolin-3 precursor
P04004	8	724.44	0.024664236	**1.810102205**	Normal	Lung Cancer	Vitronectin precursor (Serum-spreading factor)
Q9UK55	4	208.8	0.029061607	**1.807484568**	Normal	Lung Cancer	Protein Z-dependent protease inhibitor precursor
P00734	21	1584.26	0.001121278	**1.789251621**	Normal	Lung Cancer	Prothrombin precursor
P04180	1	55.09	0.002354515	**1.773866849**	Normal	Lung Cancer	Phosphatidylcholine-sterol acyltransferase precursor
P02647	12	857.22	0.002961719	**1.746514994**	Normal	Lung Cancer	Apolipoprotein A-I precursor (Apo-AI) (ApoA-I)
P02745	2	148.09	0.026868767	**1.745390695**	Lung Cancer	Normal	Complement C1q subcomponent subunit A precursor
Q08380	5	392.03	0.041189653	**1.743195217**	Normal	Lung Cancer	Galectin-3-binding protein precursor (Lectin galactoside-binding soluble 3-binding protein)
P10909	7	492.6	0.007354787	**1.737790518**	Normal	Lung Cancer	Clusterin precursor (Complement-associated protein SP-40,40)
P06727	9	578.77	0.005615092	**1.712094433**	Normal	Lung Cancer	Apolipoprotein A-IV precursor (Apo-AIV) (ApoA-IV)
Q06033	3	155.65	0.006877143	**1.688320939**	Lung Cancer	Normal	Inter-alpha-trypsin inhibitor heavy chain H3 precursor
P02649	10	678.49	0.010398747	**1.649726005**	Normal	Lung Cancer	Apolipoprotein E precursor (Apo-E)
Q8TDZ2	1	42	0.01025368	**1.62684871**	Lung Cancer	Normal	NEDD9-interacting protein with calponin homology and LIM domains
P08912	1	42	0.01025368	**1.62684871**	Lung Cancer	Normal	Muscarinic acetylcholine receptor M5
P02746	3	372.71	0.005347234	**1.556299129**	Lung Cancer	Normal	Complement C1q subcomponent subunit B precursor
P02747	3	232.65	0.013616266	**1.553880075**	Lung Cancer	Normal	Complement C1q subcomponent subunit C precursor
P19823	7	459.49	0.056391371	**1.473065446**	Normal	Lung Cancer	Inter-alpha-trypsin inhibitor heavy chain H2 precursor (ITI heavy chain H2)
P67936	1	72.45	0.037683702	**1.170005171**	Normal	Lung Cancer	Tropomyosin alpha-4 chain (Tropomyosin-4) (TM30p1)

Table 5.2.1: Progenesis Analysis of Label-Free Spectrometry on ProteoMiner serum samples showing Lung cancer versus Normal. For a particular protein the fold change indicates an increase to the sample set in the 'Highest Mean' column.

Accession	Peptides	Confidence	Anova (p)*	**Fold change**	Highest mean	Lowest mean	Description
P08670	1	64.28	0.050592958	**Infinity**	Squamous	Normal	Vimentin
P05109	1	46.23	0.019062082	**45.21316514**	Squamous	Normal	Protein S100-A8 (S100 calcium-binding protein A8) (Calgranulin-A)
P02741	5	327.88	0.008028932	**20.51512692**	Squamous	Normal	C-reactive protein precursor
P02787	12	759.43	0.000912215	**15.24374826**	Normal	Squamous	Serotransferrin precursor (Transferrin) (Siderophilin)
P05156	1	81.7	7.10E-06	**9.718043378**	Normal	Squamous	Complement factor I precursor
P04406	3	189.33	0.040803962	**8.377498357**	Squamous	Normal	Glyceraldehyde-3-phosphate dehydrogenase
P02675	4	266.77	1.39E-05	**7.483008912**	Squamous	Normal	Fibrinogen beta chain precursor
P02679	5	299.02	8.99E-05	**6.282996534**	Squamous	Normal	Fibrinogen gamma chain precursor
P01031	17	1043.78	1.97E-09	**5.903212888**	Squamous	Normal	Complement C5 precursor
P10643	3	236.76	5.28E-07	**5.844745043**	Squamous	Normal	Complement component C7 precursor -
Q6Q788	1	45.45	0.034220674	**5.77945457**	Normal	Squamous	Apolipoprotein A-V precursor (Apolipoprotein A5)
P02763	3	193.49	0.051304316	**4.483742642**	Normal	Squamous	Alpha-1-acid glycoprotein 1 precursor (AGP 1) (Orosomucoid-1)
P02654	3	163.07	0.006524973	**4.059903603**	Normal	Squamous	Apolipoprotein C-I precursor (Apo-CI) (ApoC-I)
P02748	10	677.56	9.81E-07	**4.034642408**	Squamous	Normal	Complement component C9 precursor
P02671	4	313.41	3.87E-05	**3.968450872**	Squamous	Normal	Fibrinogen alpha chain precursor
P13671	4	215.68	1.21E-06	**3.849337358**	Squamous	Normal	Complement component C6 precursor
P07360	2	106.66	0.000108813	**3.711334215**	Squamous	Normal	Complement component C8 gamma chain precursor
Q9Y238	1	43.39	0.039721658	**3.666607457**	Normal	Squamous	Deleted in lung and esophageal cancer protein 1 (Deleted in lung cancer protein 1) (DLC-1)
P07358	3	160.19	1.77E-05	**3.474502759**	Squamous	Normal	Complement component C8 beta chain precursor (Complement component 8 subunit beta)
P19652	4	227.83	0.03533365	**3.393322154**	Normal	Squamous	Alpha-1-acid glycoprotein 2 precursor (AGP 2) (Orosomucoid-2)
Q5T200	1	49.76	0.036728608	**3.368916286**	Normal	Squamous	Zinc finger CCCH domain-containing protein 13
P0C0L5	74	7226.76	0.000171263	**3.331114368**	Squamous	Normal	Complement C4-B precursor (Basic complement C4)
P07357	3	220.11	7.99E-06	**3.071053788**	Squamous	Normal	Complement component C8 alpha chain precursor
P07900	2	124.41	0.001392333	**3.003883654**	Squamous	Normal	Heat shock protein HSP 90-alpha (HSP 86)
P00738	6	353.91	0.033905456	**2.977230618**	Normal	Squamous	Haptoglobin precursor
P27169	2	98.61	0.022314085	**2.960562432**	Normal	Squamous	Serum paraoxonase/arylesterase 1
P43652	2	102.75	0.003411843	**2.83640862**	Normal	Squamous	Afamin precursor (Alpha-albumin) (Alpha-Alb)
P00739	7	459.78	0.011989534	**2.803714001**	Normal	Squamous	Haptoglobin-related protein precursor
Q14409	1	41.56	0.014491547	**2.715570513**	Normal	Squamous	Glycerol kinase, testis specific 1
P04196	5	342.3	0.02276141	**2.501459251**	Normal	Squamous	Histidine-rich glycoprotein precursor (Histidine-proline-rich glycoprotein) (HPRG)
P01024	65	6063.37	0.002664201	**2.410581428**	Squamous	Normal	Complement C3 precursor
P09871	1	43.66	0.006516534	**2.377969617**	Squamous	Normal	Complement C1s subcomponent precursor
Q8TDZ2	1	42	0.02657069	**2.279757394**	Squamous	Normal	NEDD9-interacting protein with calponin homology and LIM domains
P08912	1	42	0.02657069	**2.279757394**	Squamous	Normal	Muscarinic acetylcholine receptor M5
P06396	5	319.07	0.029326702	**2.272250709**	Normal	Squamous	Gelsolin precursor (Actin-depolymerizing factor) (ADF) (Brevin) (AGEL)
O75636	3	181.06	0.018509244	**2.181104601**	Normal	Squamous	Ficolin-3 precursor (Collagen/fibrinogen domain-containing protein 3)
P07996	19	1382.02	0.010241377	**2.118546811**	Normal	Squamous	Thrombospondin-1 precursor
P01008	20	1364	0.030836948	**2.099754695**	Normal	Squamous	Antithrombin-III precursor (ATIII)

O14791	1	41.15	0.008084209	**2.085217407**	Normal	Squamous	Apolipoprotein-L1 precursor (Apolipoprotein L-I) (Apolipoprotein L) (ApoL-I) (Apo-L) (ApoL)
P04114	33	2270.23	0.042776763	**2.041933053**	Normal	Squamous	Apolipoprotein B-100 precursor (Apo B-100) [Contains: Apolipoprotein B-48 (Apo B-48)]
P04180	1	55.09	0.010022081	**1.841829486**	Normal	Squamous	Phosphatidylcholine-sterol acyltransferase precursor
P02746	3	372.71	0.042069324	**1.729077412**	Squamous	Normal	Complement C1q subcomponent subunit B precursor
O95445	1	65.4	0.029721779	**1.701024688**	Normal	Squamous	Apolipoprotein M (Apo-M) (ApoM) (Protein G3a)
P51884	3	154.13	0.049620961	**1.654052876**	Squamous	Normal	Lumican precursor (Keratan sulfate proteoglycan lumican) (KSPG lumican)
Q14624	5	325.36	0.045281435	**1.412224808**	Squamous	Normal	Inter-alpha-trypsin inhibitor heavy chain H4 precursor (ITI heavy chain H4)

Table 5.2.2: Progenesis Analysis of Label-Free Spectrometry on ProteoMiner serum samples showing Squamous cell carcinoma versus Normal. For a particular protein the fold change indicates an increase to the sample set in the 'Highest Mean' column.

Accession	Peptides	Confidence	Anova (p)*	**Fold change**	Highest mean condition	Lowest mean condition	Description
P02675	4	266.77	0.000129	**22.14363622**	Adenocarcinoma	Normal	Fibrinogen beta chain precursor
P02679	5	299.02	0.000276	**18.90236356**	Adenocarcinoma	Normal	Fibrinogen gamma chain precursor
P02787	12	759.43	0.000515	**14.54467026**	Normal	Adenocarcinoma	Serotransferrin precursor (Transferrin) (Siderophilin)
P02741	5	327.88	3.38E-05	**11.95988051**	Adenocarcinoma	Normal	C-reactive protein precursor
P05156	1	81.7	0.003869	**11.94411207**	Normal	Adenocarcinoma	Complement factor I precursor
P02790	2	107.69	0.013694	**8.604268188**	Normal	Adenocarcinoma	Hemopexin precursor
P62805	1	49.14	0.043835	**8.434643856**	Adenocarcinoma	Normal	Histone H4
P04406	3	189.33	0.004051	**8.194487703**	Adenocarcinoma	Normal	Glyceraldehyde-3-phosphate dehydrogenase
P02652	2	209.48	0.002318	**6.691929995**	Normal	Adenocarcinoma	Apolipoprotein A-II precursor (Apo-AII) (ApoA-II)
P43652	2	102.75	4.09E-05	**4.995564697**	Normal	Adenocarcinoma	Afamin precursor (Alpha-albumin)
Q5T200	1	49.76	0.050142	**4.669152805**	Normal	Adenocarcinoma	Zinc finger CCCH domain-containing protein 13
P01042	3	276.06	0.004087	**4.449157874**	Normal	Adenocarcinoma	Kininogen-1 precursor (Alpha-2-thiol proteinase inhibitor)
P01031	17	1043.78	4.49E-07	**4.379638287**	Adenocarcinoma	Normal	Complement C5 precursor
P10643	3	236.76	1.17E-06	**4.322104028**	Adenocarcinoma	Normal	Complement component C7 precursor
P02749	1	58.52	0.054381	**4.176601336**	Normal	Adenocarcinoma	Beta-2-glycoprotein 1 precursor
P04196	5	342.3	0.0004	**4.074983889**	Normal	Adenocarcinoma	Histidine-rich glycoprotein precursor
P01023	10	674.68	0.0554	**3.899727079**	Normal	Adenocarcinoma	Alpha-2-macroglobulin precursor (Alpha-2-M)
P49908	1	75.1	0.053523	**3.798928819**	Normal	Adenocarcinoma	Selenoprotein P precursor (SeP)

P02671	4	313.41	0.000888	**3.769342648**	Adenocarcinoma	Normal	Fibrinogen alpha chain precursor
P13671	4	215.68	2.89E-05	**3.740257121**	Adenocarcinoma	Normal	Complement component C6 precursor
P06396	5	319.07	0.000426	**3.601427869**	Normal	Adenocarcinoma	Gelsolin precursor (Actin-depolymerizing factor)
P19652	4	227.83	0.032293	**3.526513963**	Normal	Adenocarcinoma	Alpha-1-acid glycoprotein 2 precursor (AGP 2)
P02748	10	677.56	8.84E-06	**3.500700364**	Adenocarcinoma	Normal	Complement component C9 precursor
O14791	1	41.15	0.000546	**3.459251289**	Normal	Adenocarcinoma	Apolipoprotein-L1 precursor (Apolipoprotein L-I)
P00739	7	459.78	0.002113	**3.292822349**	Normal	Adenocarcinoma	Haptoglobin-related protein precursor
P00738	6	353.91	0.011814	**3.275419041**	Normal	Adenocarcinoma	Haptoglobin precursor
P07900	2	124.41	0.000279	**3.010179356**	Adenocarcinoma	Normal	Heat shock protein HSP 90-alpha (HSP 86)
P07360	2	106.66	0.000217	**2.939400807**	Adenocarcinoma	Normal	Complement component C8 gamma chain precursor
P02654	3	163.07	0.036212	**2.934377062**	Normal	Adenocarcinoma	Apolipoprotein C-I precursor
P02788	2	97.32	0.037825	**2.911738306**	Adenocarcinoma	Normal	Lactotransferrin precursor
P08697	2	109.72	0.007922	**2.887385167**	Normal	Adenocarcinoma	Alpha-2-antiplasmin precursor (Alpha-2-plasmin inhibitor)
P04114	33	2270.23	0.002785	**2.820126763**	Normal	Adenocarcinoma	Apolipoprotein B-100 precursor (Apo B-100)
P08294	1	71.04	0.054037	**2.776883581**	Normal	Adenocarcinoma	Extracellular superoxide dismutase [Cu-Zn] precursor
P0C0L5	74	7226.76	3.67E-06	**2.709487429**	Adenocarcinoma	Normal	Complement C4-B precursor (Basic complement C4)
P07996	19	1382.02	0.000466	**2.691072606**	Normal	Adenocarcinoma	Thrombospondin-1 precursor
P07358	3	160.19	0.000102	**2.679410097**	Adenocarcinoma	Normal	Complement component C8 beta chain precursor
P02765	2	147.69	0.003645	**2.677701707**	Normal	Adenocarcinoma	Alpha-2-HS-glycoprotein precursor (Fetuin-A)
Q14409	1	41.56	0.016002	**2.559779772**	Normal	Adenocarcinoma	Glycerol kinase, testis specific 1
P02647	12	857.22	0.000227	**2.539082407**	Normal	Adenocarcinoma	Apolipoprotein A-I precursor (Apo-AI)
P02656	2	264.47	0.051945	**2.440115438**	Normal	Adenocarcinoma	Apolipoprotein C-III precursor (Apo-CIII) (ApoC-III)
P04180	1	55.09	0.000154	**2.42216972**	Normal	Adenocarcinoma	Phosphatidylcholine-sterol acyltransferase precursor
P07357	3	220.11	0.002099	**2.33555625**	Adenocarcinoma	Normal	Complement component C8 alpha chain precursor
Q08380	5	392.03	0.054541	**2.105425027**	Normal	Adenocarcinoma	Galectin-3-binding protein precursor
P01008	20	1364	0.029459	**2.057603265**	Normal	Adenocarcinoma	Antithrombin-III precursor (ATIII)
P00747	3	155.37	0.014531	**2.036439286**	Normal	Adenocarcinoma	Plasminogen precursor
P02766	6	550.42	0.044688	**2.028832419**	Normal	Adenocarcinoma	Transthyretin precursor
P09871	1	43.66	0.056167	**1.900908602**	Adenocarcinoma	Normal	Complement C1s subcomponent precursor
O75636	3	181.06	0.042446	**1.894058595**	Normal	Adenocarcinoma	Ficolin-3 precursor
P10909	7	492.6	0.041863	**1.807976188**	Normal	Adenocarcinoma	Clusterin precursor
P02649	10	678.49	0.021656	**1.791533775**	Normal	Adenocarcinoma	Apolipoprotein E precursor
P00734	21	1584.26	0.025771	**1.775204059**	Normal	Adenocarcinoma	Prothrombin precursor
Q06033	3	155.65	0.015111	**1.76723894**	Adenocarcinoma	Normal	Inter-alpha-trypsin inhibitor heavy chain H3 precursor
P06727	9	578.77	0.0464	**1.721722406**	Normal	Adenocarcinoma	Apolipoprotein A-IV precursor
P01024	65	6063.37	0.002245	**1.69323219**	Adenocarcinoma	Normal	Complement C3 precursor
P02747	3	232.65	0.048771	**1.611475424**	Adenocarcinoma	Normal	Complement C1q subcomponent subunit C precursor
P02746	3	372.71	0.051055	**1.604065348**	Adenocarcinoma	Normal	Complement C1q subcomponent subunit B precursor
P67936	1	72.45	0.044236	**1.534725533**	Normal	Adenocarcinoma	Tropomyosin alpha-4 chain (Tropomyosin-4) (TM30p1)

Table 5.2.3: Progenesis Analysis of Label-Free Spectrometry on ProteoMiner serum samples showing Adenocarcinoma versus Normal. For a particular protein the fold change indicates an increase to the sample set in the 'Highest Mean' column.

Accession	Peptides	Confidence	Anova (p)*	**Fold change**	Highest mean condition	Lowest mean condition	Description
P04406	3	189.33	3.57E-05	**20.4233788**	Small cell carcinoma	Normal	Glyceraldehyde-3-phosphate dehydrogenase
P02790	2	107.69	0.031437779	**8.083760734**	Normal	Small cell carcinoma	Hemopexin precursor (Beta-1B-glycoprotein)
P02787	12	759.43	0.012044232	**8.065261284**	Normal	Small cell carcinoma	Serotransferrin precursor (Transferrin) (Siderophilin)
P63104	2	97.49	0.052140846	**7.686471967**	Small cell carcinoma	Normal	14-3-3 protein zeta/delta (Protein kinase C inhibitor protein 1)
P02741	5	327.88	0.007440001	**5.481402557**	Small cell carcinoma	Normal	C-reactive protein precursor
P06702	2	103.23	0.001179068	**4.801386318**	Small cell carcinoma	Normal	Protein S100-A9 (S100 calcium-binding protein A9) (Calgranulin-B)
Q15485	2	148.97	0.039658333	**4.43578857**	Small cell carcinoma	Normal	Ficolin-2 precursor (Ficolin-B) (Ficolin-beta) (L-ficolin)
P02652	2	209.48	0.038541891	**4.358051842**	Normal	Small cell carcinoma	Apolipoprotein A-II precursor (Apo-AII) (ApoA-II)
P02749	1	58.52	0.049609017	**4.258663498**	Normal	Small cell carcinoma	Beta-2-glycoprotein 1 precursor (Beta-2-glycoprotein I) (Apolipoprotein H)
P04003	2	126.27	0.018493063	**4.035600165**	Normal	Small cell carcinoma	C4b-binding protein alpha chain precursor (C4bp) (Proline-rich protein)
P01042	3	276.06	0.009815122	**3.982164411**	Normal	Small cell carcinoma	Kininogen-1 precursor (Alpha-2-thiol proteinase inhibitor)
P04196	5	342.3	0.001367654	**3.513518756**	Normal	Small cell carcinoma	Histidine-rich glycoprotein precursor
P67936	1	72.45	0.011706915	**3.431840511**	Normal	Small cell carcinoma	Tropomyosin alpha-4 chain
P00738	6	353.91	0.020018769	**3.119937304**	Normal	Small cell carcinoma	Haptoglobin precursor
P19652	4	227.83	0.058117612	**3.01616877**	Normal	Small cell carcinoma	Alpha-1-acid glycoprotein 2 precursor (AGP 2)
P08697	2	109.72	0.0148864	**3.007641554**	Normal	Small cell carcinoma	Alpha-2-antiplasmin precursor
P22352	1	43.73	0.004788895	**2.909714649**	Normal	Small cell carcinoma	Glutathione peroxidase 3 precursor
P05546	3	167.57	0.015002638	**2.806373909**	Normal	Small cell carcinoma	Heparin cofactor 2 precursor (Heparin cofactor II)
P02679	5	299.02	0.038525098	**2.645872298**	Small cell carcinoma	Normal	Fibrinogen gamma chain precursor
P00739	7	459.78	0.011370529	**2.629738439**	Normal	Small cell carcinoma	Haptoglobin-related protein precursor
Q15127	1	42.66	0.024167586	**2.60510739**	Normal	Small cell carcinoma	Phosphatidylinositol-glycan-specific phospholipase D 2 precursor
P02675	4	266.77	0.016505274	**2.516753022**	Small cell carcinoma	Normal	Fibrinogen beta chain precursor
P10643	3	236.76	0.029102431	**2.4279047**	Small cell carcinoma	Normal	Complement component C7 precursor
P02765	2	147.69	0.009812318	**2.42314485**	Normal	Small cell carcinoma	Alpha-2-HS-glycoprotein precursor (Fetuin-A)
P04004	8	724.44	0.0167726	**2.383328129**	Normal	Small cell carcinoma	Vitronectin precursor (Serum-spreading factor) (S-protein)
P06396	5	319.07	0.025254494	**2.282546539**	Normal	Small cell carcinoma	Gelsolin precursor (Actin-depolymerizing factor) (ADF)
P02766	6	550.42	0.018562786	**2.250750028**	Normal	Small cell carcinoma	Transthyretin precursor
P07358	3	160.19	0.021449887	**2.239798466**	Small cell carcinoma	Normal	Complement component C8 beta chain precursor
P01031	17	1043.78	0.002362433	**2.108186146**	Small cell carcinoma	Normal	Complement C5 precursor
P07996	19	1382.02	0.007961609	**2.099852588**	Normal	Small cell carcinoma	Thrombospondin-1 precursor
P00734	21	1584.26	0.003265205	**2.058511331**	Normal	Small cell carcinoma	Prothrombin precursor

P01008	20	1364	0.030335506	**2.043838298**	Normal	Small cell carcinoma	Antithrombin-III precursor
P02745	2	148.09	0.007029892	**2.028489008**	Small cell carcinoma	Normal	Complement C1q subcomponent subunit A precursor
P10909	7	492.6	0.01727707	**1.976446671**	Normal	Small cell carcinoma	Clusterin precursor (Complement-associated protein SP-40,40)
Q9UK55	4	208.8	0.049035774	**1.975595701**	Normal	Small cell carcinoma	Protein Z-dependent protease inhibitor precursor
P02748	10	677.56	0.011059623	**1.941438703**	Small cell carcinoma	Normal	Complement component C9 precursor
P07357	3	220.11	0.045718995	**1.885653704**	Small cell carcinoma	Normal	Complement component C8 alpha chain precursor
P19827	10	743.77	0.007431948	**1.872435053**	Normal	Small cell carcinoma	Inter-alpha-trypsin inhibitor heavy chain H1 precursor
P19823	7	459.49	0.023815624	**1.856034944**	Normal	Small cell carcinoma	Inter-alpha-trypsin inhibitor heavy chain H2 precursor
P0C0L5	74	7226.76	0.00066799	**1.831085573**	Small cell carcinoma	Normal	Complement C4-B precursor (Basic complement C4)
P06727	9	578.77	0.028755148	**1.787358924**	Normal	Small cell carcinoma	Apolipoprotein A-IV precursor (Apo-AIV)
P02671	4	313.41	0.034159192	**1.779780611**	Small cell carcinoma	Normal	Fibrinogen alpha chain precursor
P07360	2	106.66	0.042471829	**1.649445656**	Small cell carcinoma	Normal	Complement component C8 gamma chain precursor
P12259	8	476.28	0.016789884	**1.544519466**	Small cell carcinoma	Normal	Coagulation factor V precursor (Activated protein C cofactor)
Q06033	3	155.65	0.053666255	**1.438094814**	Small cell carcinoma	Normal	Inter-alpha-trypsin inhibitor heavy chain H3 precursor
P01024	65	6063.37	0.013883731	**1.436912893**	Small cell carcinoma	Normal	Complement C3 precursor
P02746	3	372.71	0.043942821	**1.353032455**	Small cell carcinoma	Normal	Complement C1q subcomponent subunit B precursor

Table 5.2.4: Progenesis Analysis of Label-Free Spectrometry on ProteoMiner serum samples showing Small cell carcinoma versus Normal. For a particular protein the fold change indicates an increase to the sample set in the 'Highest Mean' column.

Accession	Peptides	Confidence	Anova (p)*	**Fold change**	Highest mean condition	Lowest mean condition	Description
P05109	1	46.23	0.052947732	**10.5605019**	Squamous cell carcinoma	Adenocarcinoma	Protein S100-A8 (S100 calcium-binding protein A8) (Calgranulin-A)
P08294	1	71.04	0.020418568	**6.918351356**	Squamous cell carcinoma	Adenocarcinoma	Extracellular superoxide dismutase [Cu-Zn] precursor
P00488	1	40.57	0.056420459	**3.134540466**	Squamous cell carcinoma	Adenocarcinoma	Coagulation factor XIII A chain precursor
P02765	2	147.69	0.027979633	**1.906703271**	Squamous cell carcinoma	Adenocarcinoma	Alpha-2-HS-glycoprotein precursor (Fetuin-A) (Alpha-2-Z-globulin)
P05090	3	254.63	0.012247958	**1.782471796**	Squamous cell carcinoma	Adenocarcinoma	Apolipoprotein D precursor (Apo-D) (ApoD)
P51884	3	154.13	0.052001058	**1.740268999**	Squamous cell carcinoma	Adenocarcinoma	Lumican precursor (Keratan sulfate proteoglycan lumican) (KSPG lumican)
Q02985	2	101.56	0.039861955	**1.641688398**	Squamous cell carcinoma	Adenocarcinoma	Complement factor H-related protein 3 precursor (FHR-3)
P04196	5	342.3	0.037322936	**1.629042683**	Squamous cell carcinoma	Adenocarcinoma	Histidine-rich glycoprotein precursor (Histidine-proline-rich glycoprotein) (HPRG)

Table 5.2.5: Progenesis Analysis of Label-Free Spectrometry on ProteoMiner serum samples showing Squamous cell carcinoma versus Adenocarcinoma. For a particular protein the fold change indicates an increase to the sample set in the 'Highest Mean' column.

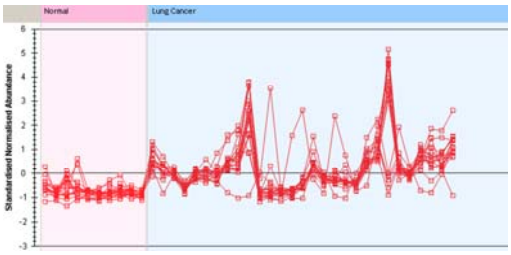
Accession	Peptides	Confidence	Anova (p)*	**Fold change**	Highest mean condition	Lowest mean condition	Description
P01031	17	1043.78	6.30E-05	**2.800138356**	Squamous cell carcinoma	Small cell carcinoma	Complement C5 precursor
P51884	3	154.13	6.98E-05	**3.411815613**	Squamous cell carcinoma	Small cell carcinoma	Lumican precursor (Keratan sulfate proteoglycan lumican)
P02654	3	163.07	0.000332992	**9.452261653**	Small cell carcinoma	Squamous cell carcinoma	Apolipoprotein C-I precursor (Apo-CI) (ApoC-I)
P02748	10	677.56	0.001349938	**2.07817141**	Squamous cell carcinoma	Small cell carcinoma	Complement component C9 precursor
P04003	2	126.27	0.001384598	**3.953193259**	Squamous cell carcinoma	Small cell carcinoma	C4b-binding protein alpha chain precursor (C4bp) (Proline-rich protein) (PRP)
P02675	4	266.77	0.002906855	**2.973279001**	Squamous cell carcinoma	Small cell carcinoma	Fibrinogen beta chain precursor
P35542	3	270.01	0.003315821	**1.976074848**	Small cell carcinoma	Squamous cell carcinoma	Serum amyloid A-4 protein precursor
P02671	4	313.41	0.004440963	**2.229741603**	Squamous cell carcinoma	Small cell carcinoma	Fibrinogen alpha chain precursor
P27169	2	98.61	0.004865035	**3.358153983**	Small cell carcinoma	Squamous cell carcinoma	Serum paraoxonase/arylesterase 1
P07360	2	106.66	0.005311063	**2.250049404**	Squamous cell carcinoma	Small cell carcinoma	Complement component C8 gamma chain precursor
P10643	3	236.76	0.007851931	**2.407320618**	Squamous cell carcinoma	Small cell carcinoma	Complement component C7 precursor
P04070	2	114.59	0.009029815	**2.183965527**	Squamous cell carcinoma	Small cell carcinoma	Vitamin K-dependent protein C precursor
P05546	3	167.57	0.009279239	**2.102772213**	Squamous cell carcinoma	Small cell carcinoma	Heparin cofactor 2 precursor (Heparin cofactor II)
P02679	5	299.02	0.010245799	**2.374640885**	Squamous cell carcinoma	Small cell carcinoma	Fibrinogen gamma chain precursor
P22352	1	43.73	0.01293374	**2.419969692**	Squamous cell carcinoma	Small cell carcinoma	Glutathione peroxidase 3 precursor
P13671	4	215.68	0.015364931	**1.850467807**	Squamous cell carcinoma	Small cell carcinoma	Complement component C6 precursor
P02787	12	759.43	0.01750537	**1.890050145**	Small cell carcinoma	Squamous cell carcinoma	Serotransferrin precursor (Transferrin) (Siderophilin)
P55056	1	63.04	0.017531791	**2.373025202**	Small cell carcinoma	Squamous cell carcinoma	Apolipoprotein C-IV precursor (Apo-CIV) (
P43652	2	102.75	0.017943194	**1.5060239**	Small cell carcinoma	Squamous cell carcinoma	Afamin precursor (Alpha-albumin) (Alpha-Alb)
P07357	3	220.11	0.022827503	**1.628641452**	Squamous cell carcinoma	Small cell carcinoma	Complement component C8 alpha chain precursor
P02753	2	103.89	0.027351744	**2.200059695**	Squamous cell carcinoma	Small cell carcinoma	Plasma retinol-binding protein precursor (PRBP) (RBP)
P0C0L5	74	7226.76	0.038734089	**1.819201907**	Squamous cell carcinoma	Small cell carcinoma	Complement C4-B precursor (Basic complement C4)
Q15485	2	148.97	0.038933494	**4.889998653**	Small cell carcinoma	Squamous cell carcinoma	Ficolin-2 precursor (Ficolin-B) (Ficolin-beta) (L-ficolin)
P08294	1	71.04	0.039990383	**4.808524596**	Squamous cell carcinoma	Small cell carcinoma	Extracellular superoxide dismutase [Cu-Zn] precursor
P02745	2	148.09	0.04892128	**1.63187255**	Small cell carcinoma	Squamous cell carcinoma	Complement C1q subcomponent subunit A precursor

Table 5.2.6: Progenesis Analysis of Label-Free Spectrometry on ProteoMiner serum samples showing Squamous cell carcinoma versus Small cell carcinoma. For a particular protein the fold change indicates an increase to the sample set in the 'Highest Mean' column.

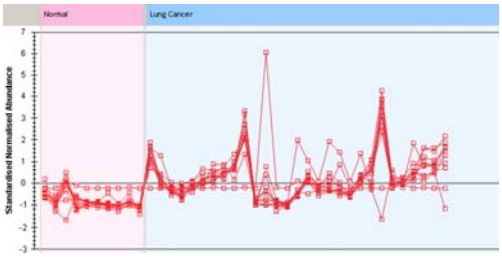
Accession	Peptides	Confidence	Anova (p)*	Fold change	Highest mean condition	Lowest mean condition	Description
P02675	4	266.77	0.008930687	**8.798493944**	Adenocarcinoma	Small cell carcinoma	Fibrinogen beta chain precursor
P02679	5	299.02	0.014798937	**7.144095192**	Adenocarcinoma	Small cell carcinoma	Fibrinogen gamma chain precursor
P02654	3	163.07	0.00126768	**6.831812401**	Small cell carcinoma	Adenocarcinoma	Apolipoprotein C-I precursor (Apo-CI) (ApoC-I)
P31151	2	112.28	0.027170703	**4.841443795**	Small cell carcinoma	Adenocarcinoma	Protein S100-A7 (S100 calcium-binding protein A7) (Psoriasin)
Q15485	2	148.97	0.055249851	**3.998104919**	Small cell carcinoma	Adenocarcinoma	Ficolin-2 precursor (Ficolin-B) (Ficolin-beta) (L-ficolin)
P06702	2	103.23	0.002283116	**3.56996585**	Small cell carcinoma	Adenocarcinoma	Protein S100-A9 (S100 calcium-binding protein A9) (Calgranulin-B)
P22352	1	43.73	0.004782377	**3.363908013**	Adenocarcinoma	Small cell carcinoma	Glutathione peroxidase 3 precursor
P27169	2	98.61	9.56E-05	**3.067633166**	Small cell carcinoma	Adenocarcinoma	Serum paraoxonase/arylesterase 1
P43652	2	102.75	6.26E-05	**2.652452744**	Small cell carcinoma	Adenocarcinoma	Afamin precursor (Alpha-albumin) (Alpha-Alb)
P04003	2	126.27	0.016083018	**2.215496287**	Adenocarcinoma	Small cell carcinoma	C4b-binding protein alpha chain precursor (C4bp) (Proline-rich protein)
P02741	5	327.88	0.055310399	**2.181901509**	Adenocarcinoma	Small cell carcinoma	C-reactive protein precursor [Contains: C-reactive protein(1-205)]
P01031	17	1043.78	0.004238585	**2.077443823**	Adenocarcinoma	Small cell carcinoma	Complement C5 precursor
P55056	1	63.04	0.026858468	**1.933404718**	Small cell carcinoma	Adenocarcinoma	Apolipoprotein C-IV precursor (Apo-CIV) (ApoC-IV)
P02787	12	759.43	0.015099852	**1.803372482**	Small cell carcinoma	Adenocarcinoma	Serotransferrin precursor (Transferrin) (Siderophilin) (
P02748	10	677.56	0.011396623	**1.80314751**	Adenocarcinoma	Small cell carcinoma	Complement component C9 precursor
P13671	4	215.68	0.038961473	**1.798030349**	Adenocarcinoma	Small cell carcinoma	Complement component C6 precursor
P07360	2	106.66	0.019349526	**1.782053744**	Adenocarcinoma	Small cell carcinoma	Complement component C8 gamma chain precursor
P10643	3	236.76	0.030812664	**1.780178616**	Adenocarcinoma	Small cell carcinoma	Complement component C7 precursor
P04180	1	55.09	0.008026835	**1.776295565**	Small cell carcinoma	Adenocarcinoma	Phosphatidylcholine-sterol acyltransferase precursor
P04070	2	114.59	0.021839728	**1.736331618**	Adenocarcinoma	Small cell carcinoma	Vitamin K-dependent protein C precursor
O95445	1	65.4	0.04881551	**1.706971369**	Small cell carcinoma	Adenocarcinoma	Apolipoprotein M (Apo-M) (ApoM) (Protein G3a)
P02647	12	857.22	0.00109676	**1.68465689**	Small cell carcinoma	Adenocarcinoma	Apolipoprotein A-I precursor (Apo-AI) (ApoA-I)
P05546	3	167.57	0.013607655	**1.601263102**	Adenocarcinoma	Small cell carcinoma	Heparin cofactor 2 precursor (Heparin cofactor II)
P06396	5	319.07	0.014600804	**1.577811365**	Small cell carcinoma	Adenocarcinoma	Gelsolin precursor (Actin-depolymerizing factor) (ADF) (Brevin) (AGEL)
P04114	33	2270.23	0.025006916	**1.563802232**	Small cell carcinoma	Adenocarcinoma	Apolipoprotein B-100 precursor (Apo B-100)
P0C0L5	74	7226.76	0.005729155	**1.479716442**	Adenocarcinoma	Small cell carcinoma	Complement C4-B precursor (Basic complement C4)
P01023	10	674.68	0.05571654	**1.447767297**	Small cell carcinoma	Adenocarcinoma	Alpha-2-macroglobulin precursor (Alpha-2-M)

Table 5.2.7: Progenesis Analysis of Label-Free Spectrometry on ProteoMiner serum samples showing Adenocarcinoma versus Small cell carcinoma. For a particular protein the fold change indicates an increase to the sample set in the 'Highest Mean' column.

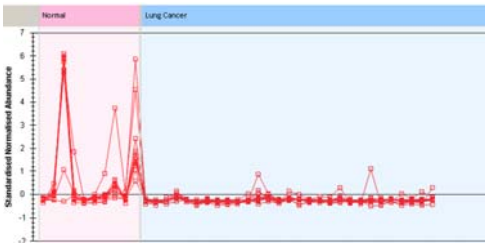
(A)



(B)



(C)

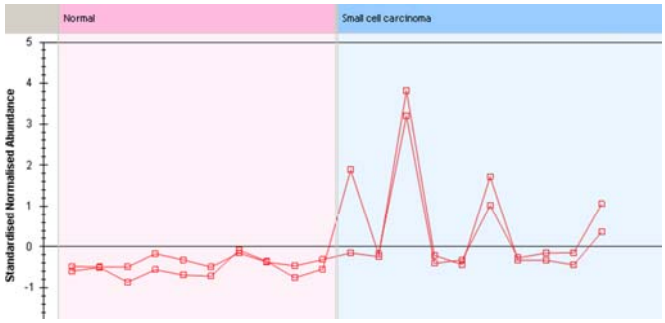


(D)

(E)

<u>Figure 5.2.1</u>: Label-Free Targets from Normal (pink) versus Lung Cancer (blue) comparison showing abundance levels of each peptide identified for a particular protein. (A) Complement C5 precursor, (B) complement c9 precursor, (C) serotransferrin, (D) histidine rich glycoprotein, and (E) Thrombospondin.

(A)



(B)

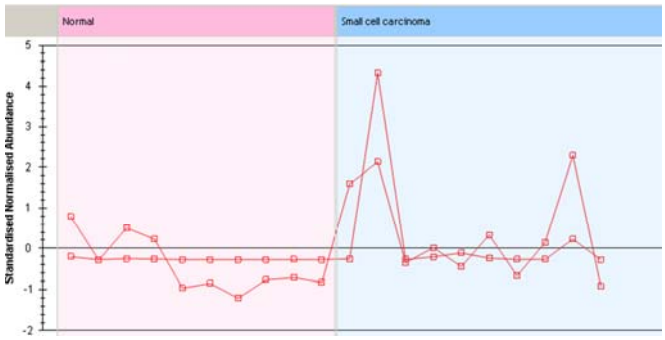
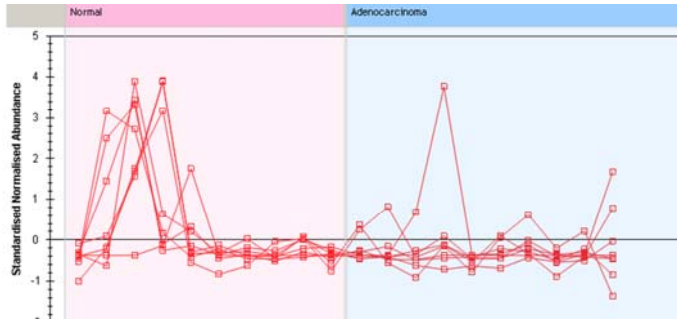


Figure 5.2.2: Label-Free Targets from Normal (pink) versus Small cell carcinoma (blue) comparison showing abundance levels of each peptide identified for a particular protein. (A) S100-A9, and (B) L-ficolin.

(A)



(B)

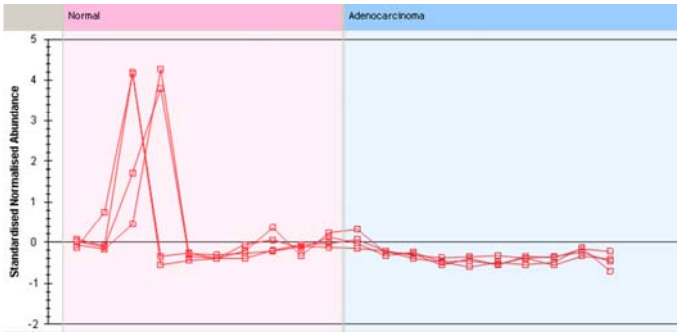


Figure 5.2.3: Label-Free Targets from Normal (pink) versus Adenocarcinoma (blue) comparison showing abundance levels of each peptide identified for a particular protein. (A) Clusterin precursor, and (B) Kininogen 1.

(A)

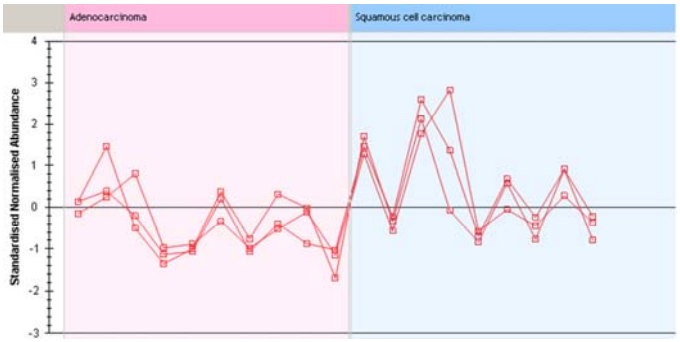


Figure 5.2.4: Label-Free Targets from Adenocarcinoma (pink) versus Squamous cell carcinoma (blue) comparison showing abundance levels of each peptide identified for a particular protein. (A) Lumican.

5.3 Pathway Studio Analysis of Targets from Label-Free Experiment

Pathway Studio pathway analysis software is used to find relationships among genes, proteins, cell processes and diseases, to interpret gene expression and other high throughput data, to build, expand and analyse pathways, and to create diagrams showing the pathway interaction between your proteins of interest. Proteins identified from the label-free experiment were put into the Pathway Analysis software and of the 64 proteins identified 23 proteins were found to have no direct interactions with any other proteins. The remaining 41 proteins are listed in Table 5.3.1.

Thrombospondin 1 (THBS1) and kininogen 1 (KNG1) were identified from Pathway Studio Analysis as 'hubs' of connectivity. In Figure 5.3.1 the diamond shape enclosing THBS1 and KNG1 denotes them as ligands, which can be described as signal triggering molecules. KNG1 is a constituent of the blood coagulation system and THBS1 plays a role in angiogenesis and tumourigenesis.
A serum ELISA was performed on both THBS1 (Fig. 5.3.3) and KNG1 (Fig. 5.3.4).

Name	Description
AMBP	alpha-1-microglobulin/bikunin precursor
C4BPA	complement component 4 binding protein, alpha
C5	complement component 5
CFH	complement factor H
CFHR5	complement factor H-related 5
CFI	complement factor I
CLU	clusterin
CP	ceruloplasmin (ferroxidase)
CRP	C-reactive protein, pentraxin-related
F13A1	coagulation factor XIII, A1 polypeptide
FBLN1	fibulin 1
FCN2	ficolin
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GPX3	glutathione peroxidase 3 (plasma)
GSN	gelsolin
HP	haptoglobin
HPR	haptoglobin-related protein
HPX	hemopexin
HRG	histidine-rich glycoprotein
HSP90AA1	heat shock protein 90kDa alpha (cytosolic)
KNG1	kininogen 1
LCAT	lecithin-cholesterol acyltransferase
LPA	lipoprotein, Lp(a)
LTF	lactotransferrin
MICAL1	microtubule associated monoxygenase, calponin and LIM domain
ORM1	orosomucoid 1
ORM2	orosomucoid 2
PON1	paraoxonase 1
PROC	protein C (inactivator of coagulation factors Va and VIIIa)
RBP4	retinol binding protein 4, plasma
S100A8	S100 calcium binding protein A8
S100A9	S100 calcium binding protein A9
SEPP1	selenoprotein P, plasma, 1
SERPINA3	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase)
SERPIND1	serpin peptidase inhibitor, clade D (heparin cofactor), member 1
SERPINF2	Plasmin inhibitor deficiency
THBS1	thrombospondin 1
TTR	transthyretin
VIM	vimentin
VTN	vitronectin
YWHAZ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein

Table 5.3.1: Protein Key for Pathway Studio Analysis of biomarker targets from Label-Free Spectrometry experiment on ProteoMiner serum.

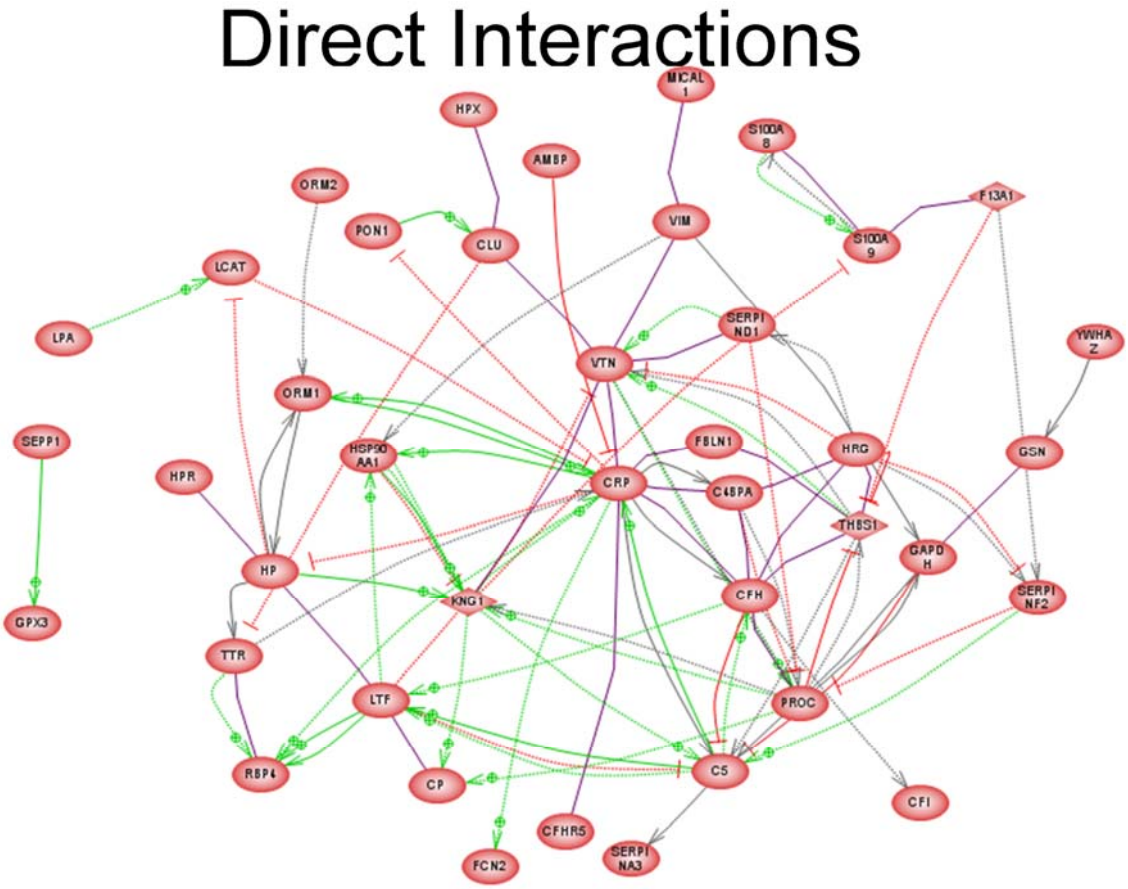


Figure 5.3.1: Proteins identified from label-free experiment and their interactions. Two major 'hubs' of connectivity were identified at KNG1 and THBS1.

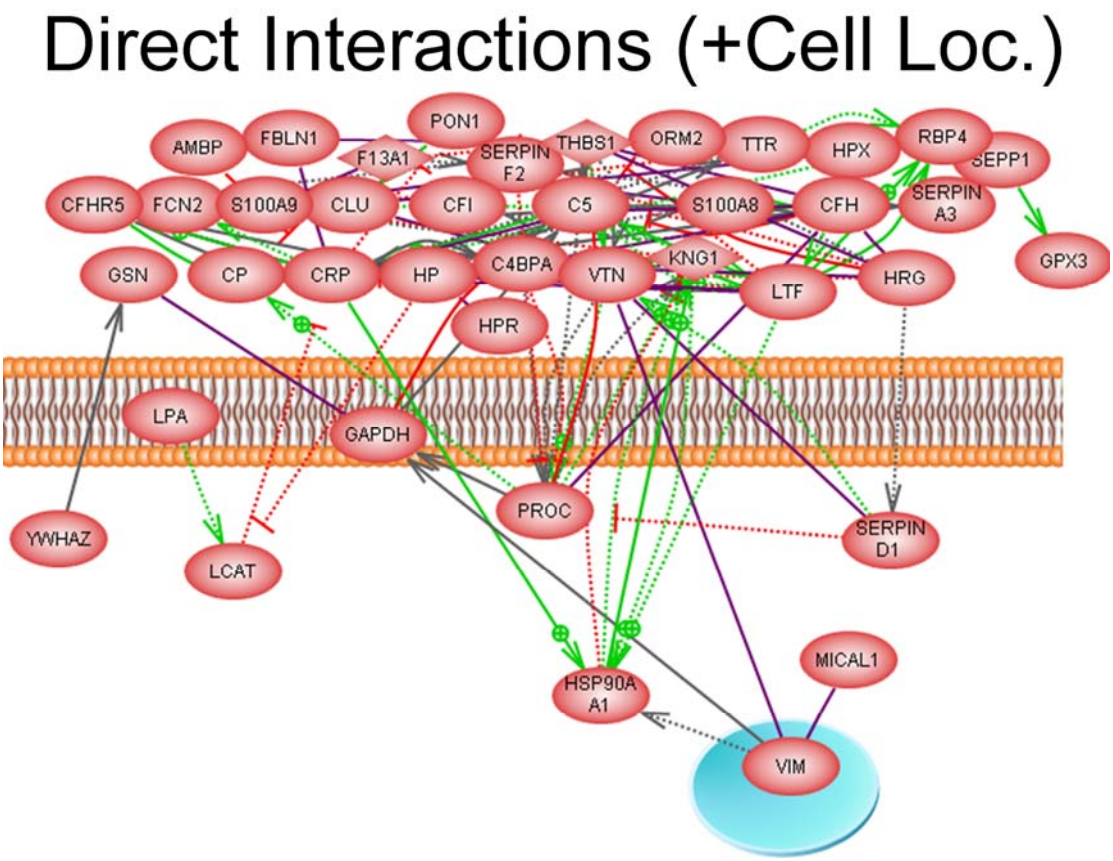


Figure 5.3.2: Cell Location of proteins identified from label-free experiment and their interactions. THBS1 and KNG1 are secreted proteins.

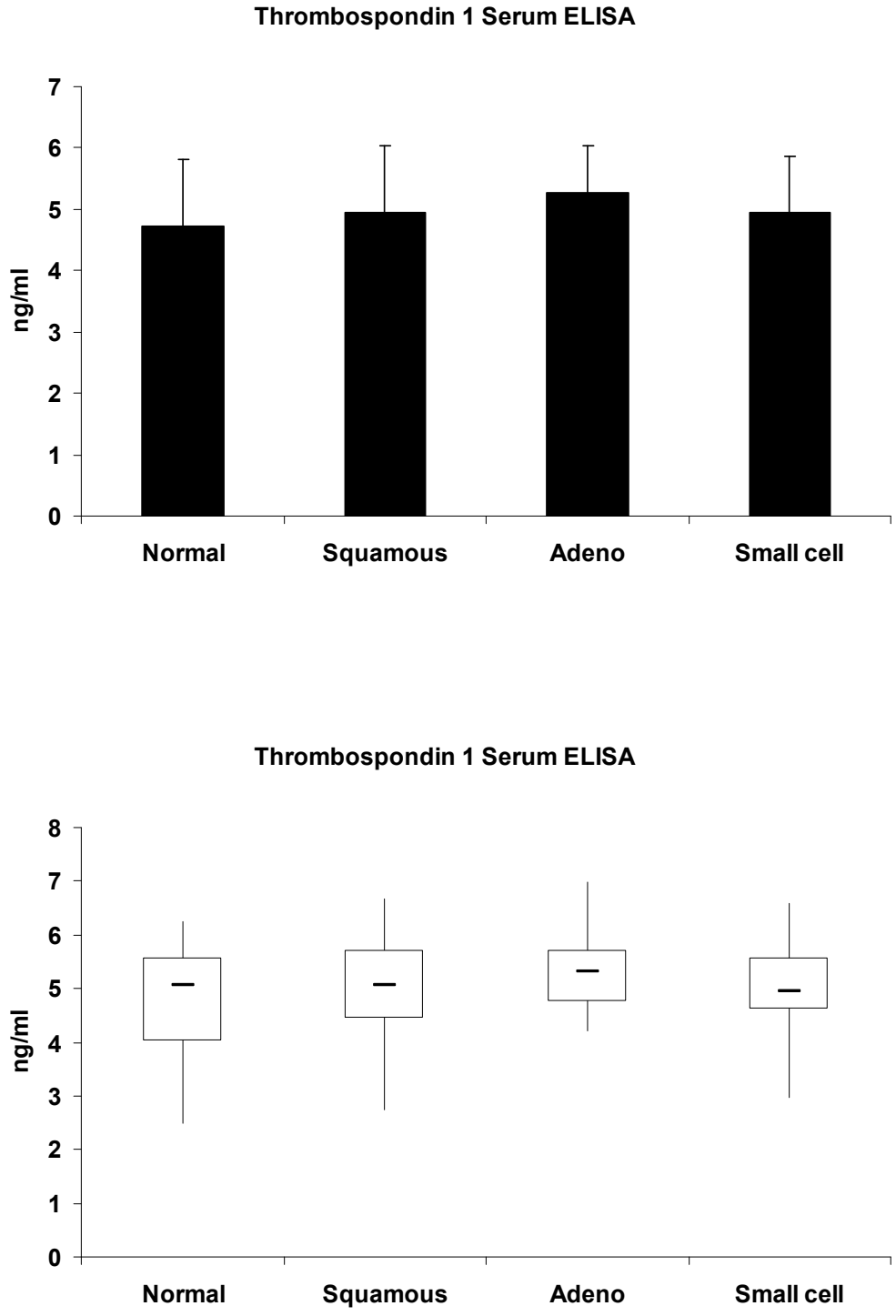
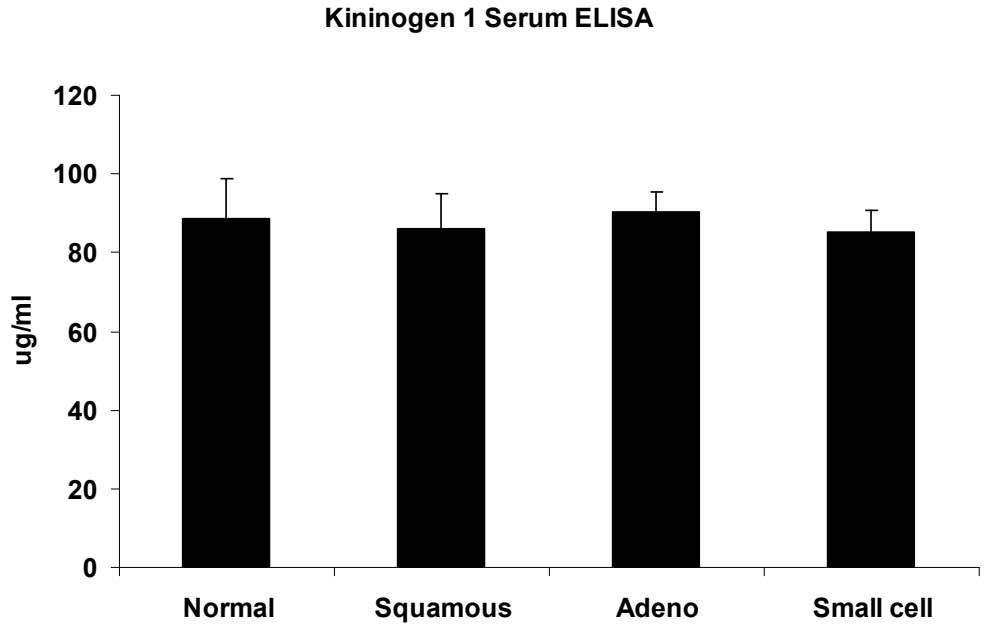


Figure 5.3.3: Thrombospondin 1 ELISA: Bar chart and Box and Whisker Plot showing normal compared to late-stage lung cancer types. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 13 squamous, 18 adenocarcinoma, and 17 small cell lung cancer.



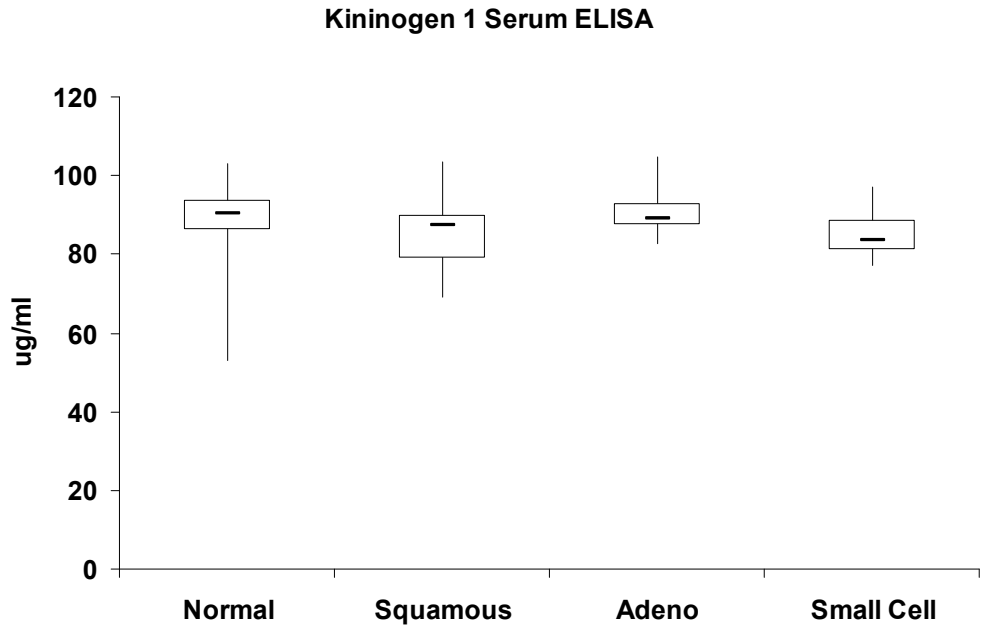


Figure 5.3.4: Kininogen 1 ELISA: Bar chart and Box and Whisker Plot showing normal compared to late-stage lung cancer types. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 13 squamous, 18 adenocarcinoma, and 17 small cell lung cancer.

CHAPTER SIX

LUMINEX MULTIPLEX PLATFORM FOR LUNG CANCER BIOMARKER DISCOVERY

6.1 Background

Multiple potential biomarkers can be quantitatively measured at the same time in the same serum or plasma sample using multiplex assay platforms. The sensitivity of the instruments means very small quantities of serum/plasma, for example as little as 5µl, can be used in a single experiment, a single experiment that may examine up to one hundred analytes. Due to cross-reactivity however, only certain combinations of analytes can be used in a single experiment, the combinations themselves restricted to those panels produced by the manufacturer.

Luminex multiplex uses coated microspheres, tiny colour-coded polystyrene beads with distinct proportions of coloured dyes, to capture and detect the specific analyte from a sample. To further validate the results, multiple readings are made on each bead set.

6.2 Experimental Design

Previous biomarker discovery techniques used, for example 2D-DIGE of conditioned media and immunodepleted serum, and label-free mass spectrometry of Proteominer serum, start off relatively blind; there is no way to know what proteins may show potential as lung cancer biomarkers, and due to their very nature as basic exploratory techniques the results could very well be different if the experiments were repeated. For multiplex assay platforms, because the markers are known, serum and plasma are used for discovery phase, meaning that those brought forward to be validated in larger sample sets can be done so with greater confidence that the results will be reproducible.

Initial discovery phase included cancer panel 1 with six proteins, cancer panel 2 with twelve proteins, and the matrix-metalloproteinase (MMP) panel with six proteins. From cancer panel 2, epidermal growth factor did not provide results so is not included (i.e. no data generated). As the discovery phase was performed over multiple plates with multiple cancers, the normal healthy sample size was larger than the sample size for any of the lung cancer types (i.e. normal $n = 20$, SCC $n = 9$, AD $n = 9$, and SCLC $n = 10$). All normal samples were performed in duplicate. For the cancer samples the majority were run as single samples, only five cancer samples ran in duplicate.

Interesting proteins were brought forward to be examined on the Luminex system in a larger cohort of samples. As well as investigating levels in serum, plasma was also looked at. Plasma assays include a benign lung disease group and in some serum assays performed on earlier stage lung cancer there is an autoimmune control as well as a benign lung disease control.

The Luminex multiplex platform reduces cost when panels of a certain size are purchased, i.e. the larger the panel the greater the savings. In some cases it was cheaper to buy in a separate ELISA for proteins interesting from the discovery phase and run them separately, or if further validation was required this was the cheaper alternative (e.g. in the case of MMP-8, prolactin, and tenascin C). Also, where the possibility presented itself to perform a multiplex assay on a protein not present in the discovery phase, a full set of samples was run and results included in the validation phase (e.g. MMP-2 and MMP-7). Logistic regression was performed on panels of some of the interesting proteins for both serum and plasma, as well as on the known cancer biomarkers (e.g. CA-125, CA19-9, and CEA) for comparison.

6.3 Luminex Discovery Phase

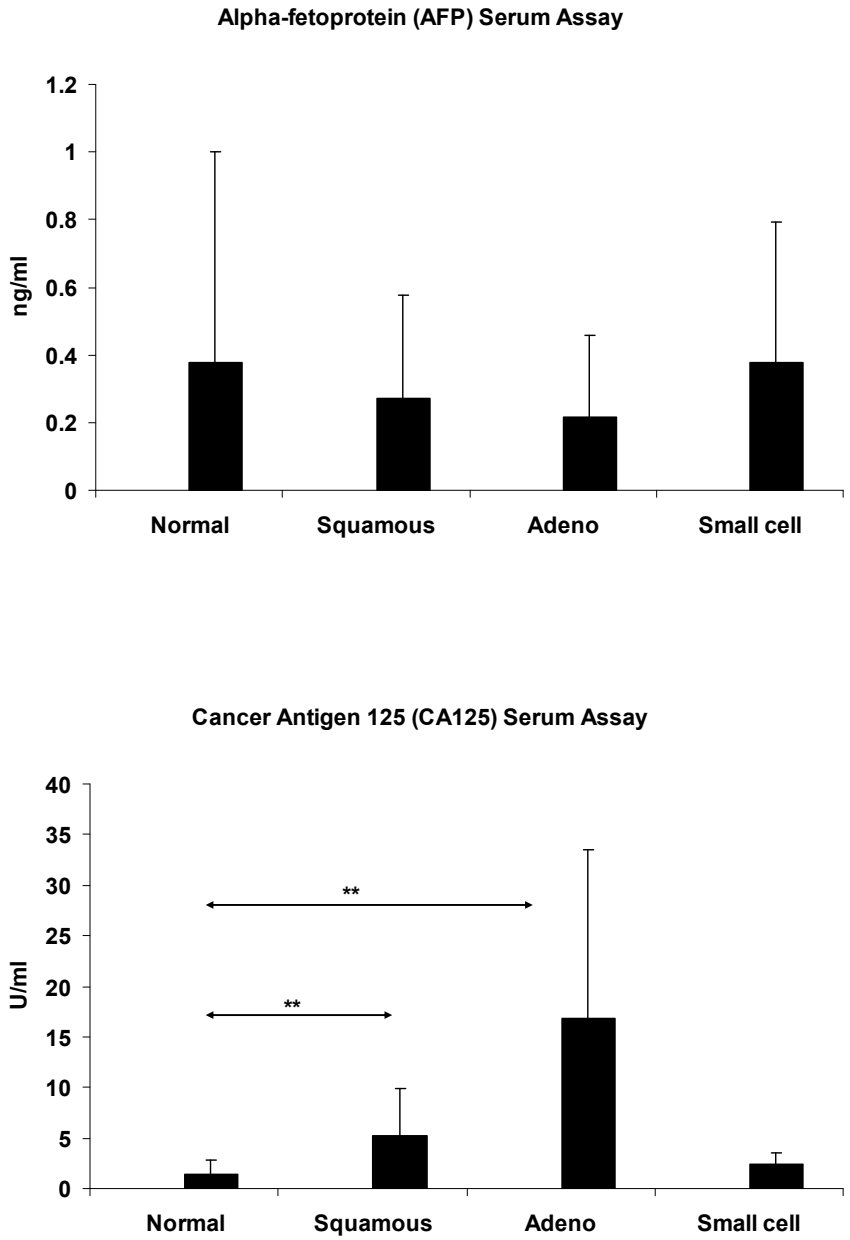


Figure 6.3.1: Alpha-fetoprotein (AFP) and cancer antigen-125 from Cancer Panel 1. Bar chart showing normal compared to late-stage lung cancer type. Patient serum samples included 20 normal, 9 squamous, 9 adenocarcinoma and 10 small cell carcinoma. Student's t-test with p value less than 0.01 is indicated by two asterisks (**).

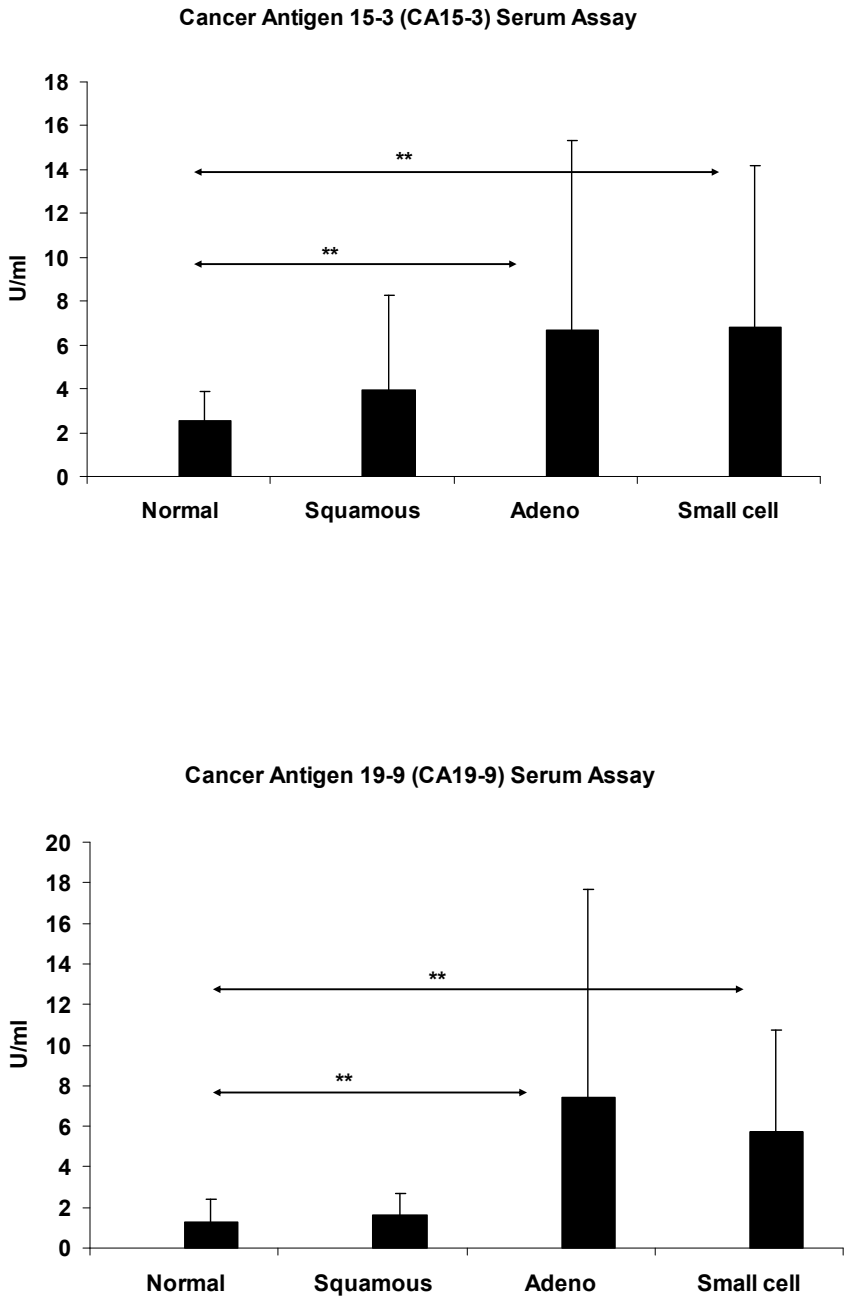


Figure 6.3.2: Cancer antigen 15-3 and cancer antigen 19-9 from Cancer Panel 1. Bar chart showing normal compared to late-stage lung cancer type. Patient serum samples included 20 normal, 9 squamous, 9 adenocarcinoma and 10 small cell carcinoma. Student's t-test with p value less than 0.01 is indicated by two asterisks (**).

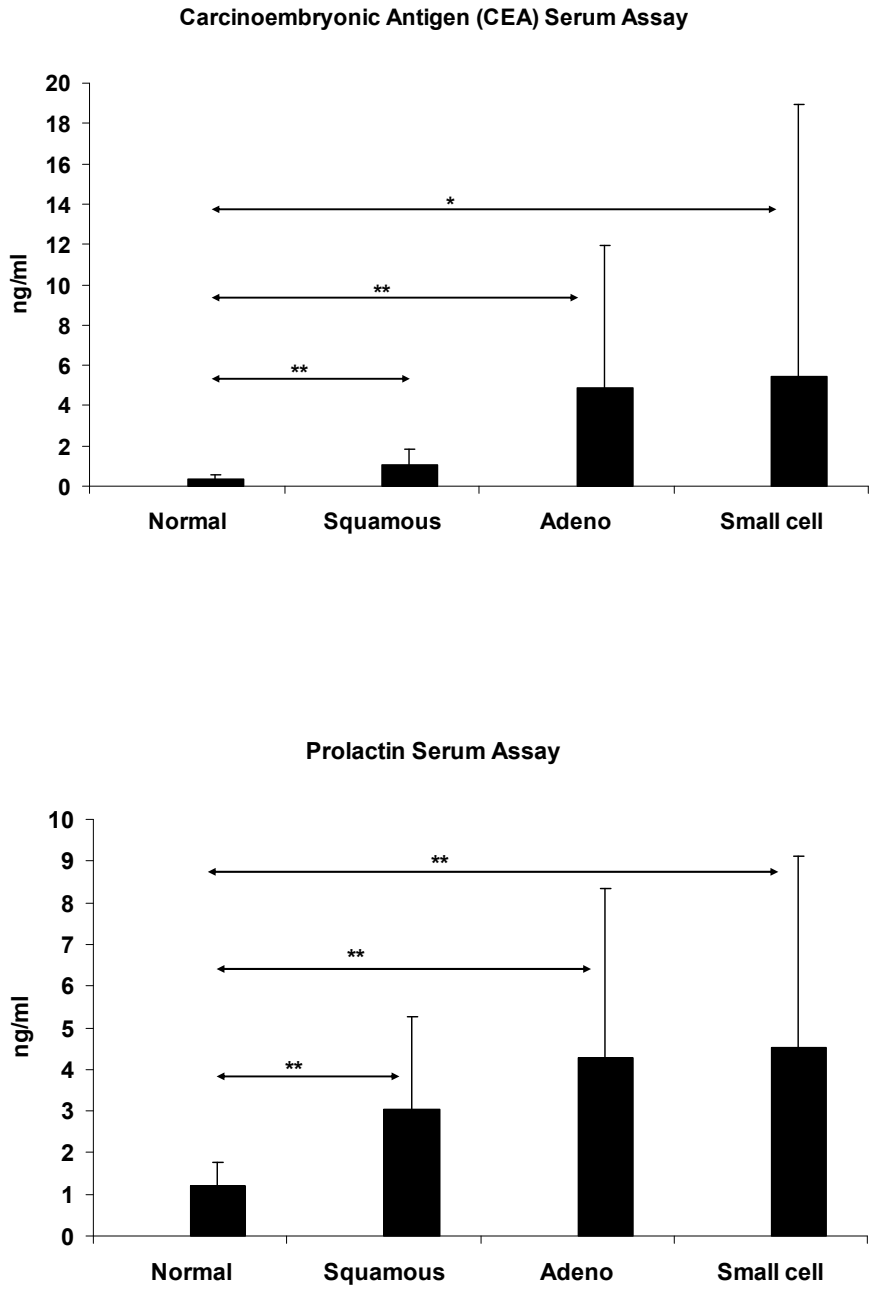


Figure 6.3.3: Carcinoembryonic antigen and prolactin from Cancer Panel 1. Bar chart showing normal compared to late-stage lung cancer type. Patient serum samples included 20 normal, 9 squamous, 9 adenocarcinoma and 10 small cell carcinoma. Student's t-test with p value less than 0.05 is indicated by a single asterisk (*) and p value less than 0.01 is indicated by two asterisks (**).

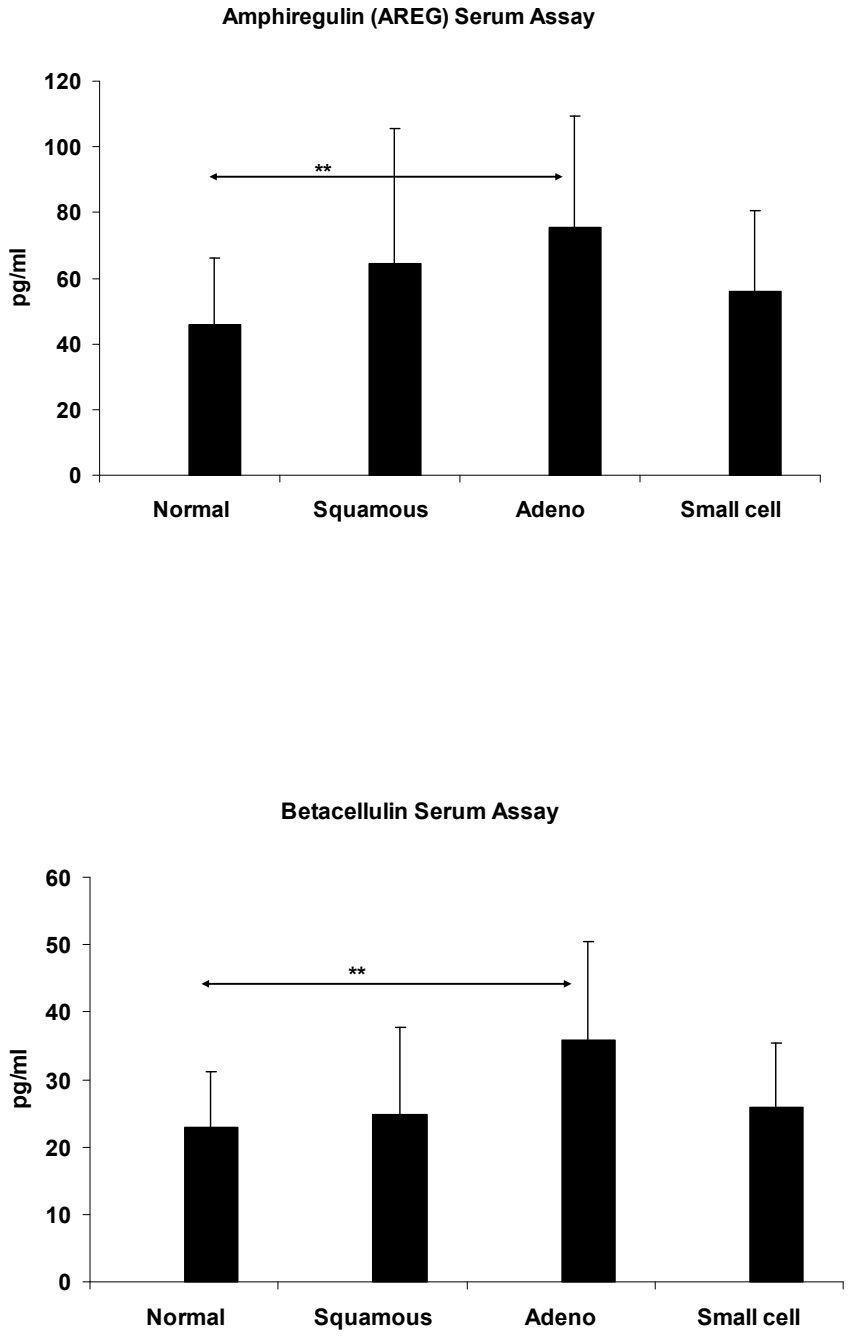


Figure 6.3.4: Amphiregulin and betacellulin from Cancer Panel 2. Bar chart showing normal compared to late-stage lung cancer type. Patient serum samples included 20 normal, 9 squamous, 9 adenocarcinoma and 10 small cell carcinoma. Student's t-test with p value less than 0.01 is indicated by two asterisks (**).

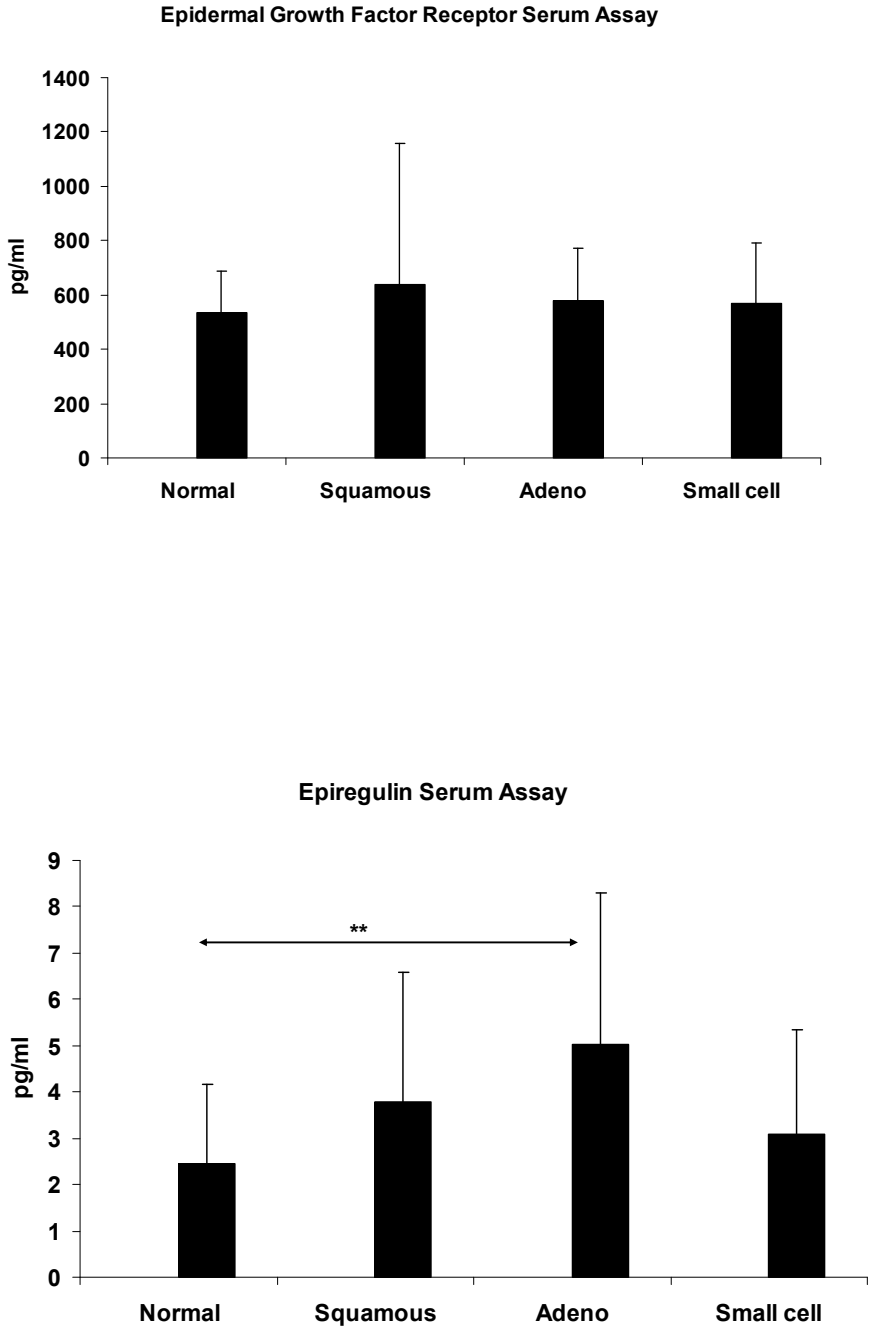
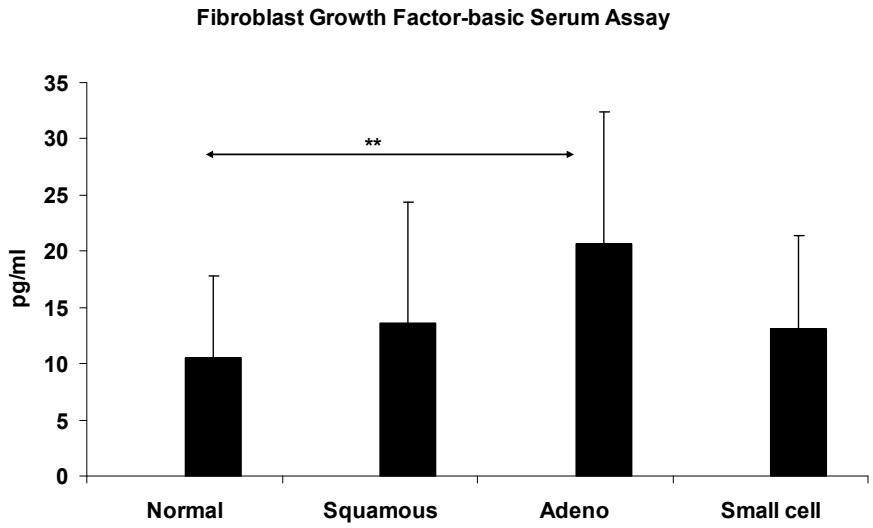


Figure 6.3.5: Epidermal growth factor receptor and epiregulin from Cancer Panel 2. Bar chart showing normal compared to late-stage lung cancer type. Patient serum samples included 20 normal, 9 squamous, 9 adenocarcinoma and 10 small cell carcinoma. Student's t-test with p value less than 0.01 is indicated by two asterisks (**).



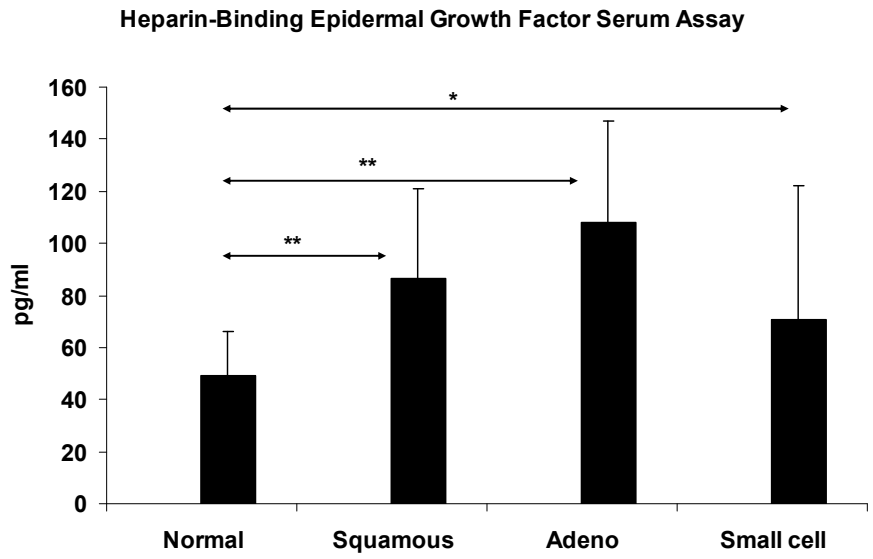


Figure 6.3.6: Fibroblast growth factor-basic and heparin-binding epidermal growth factor from Cancer Panel 2. Bar chart showing normal compared to late-stage lung cancer type. Patient serum samples included 20 normal, 9 squamous, 9 adenocarcinoma and 10 small cell carcinoma. Student's t-test with p value less than 0.05 is indicated by a single asterisk (*) and p value less than 0.01 is indicated by two asterisks (**).

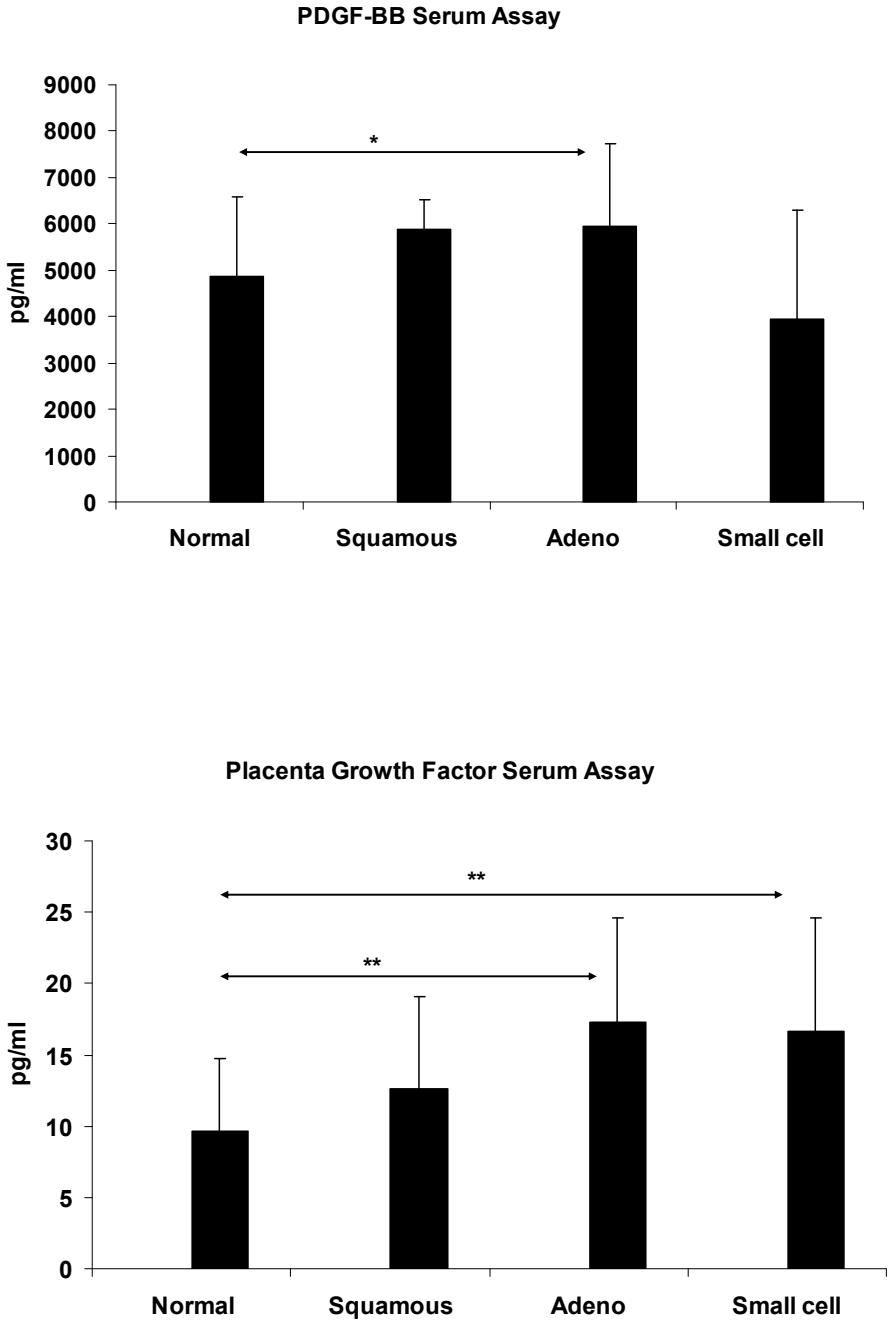
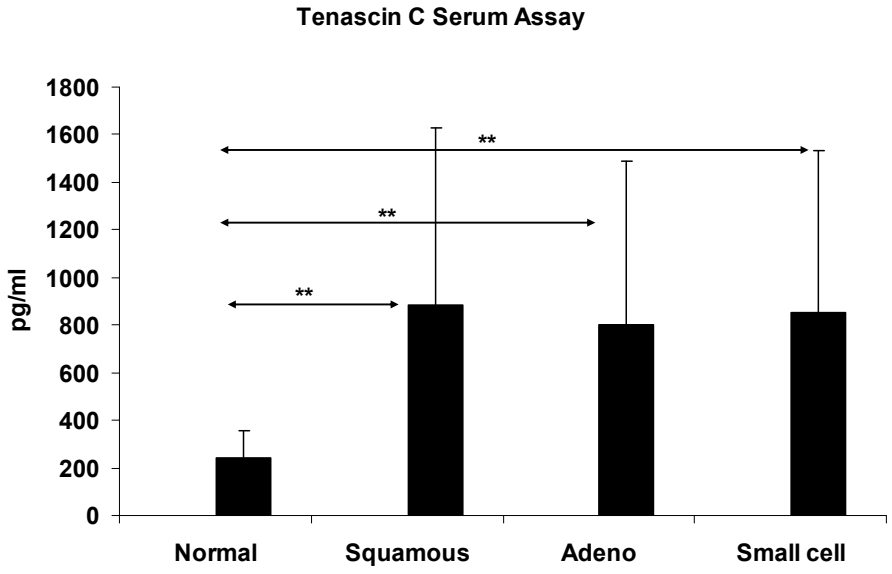


Figure 6.3.7: Platelet-derived growth factor and placenta growth factor from Cancer Panel 2. Bar chart showing normal compared to late-stage lung cancer type. Patient serum samples included 20 normal, 9 squamous, 9 adenocarcinoma and 10 small cell carcinoma. Student's t-test with p value less than 0.05 is indicated by a single asterisk (*) and p value less than 0.01 is indicated by two asterisks (**).



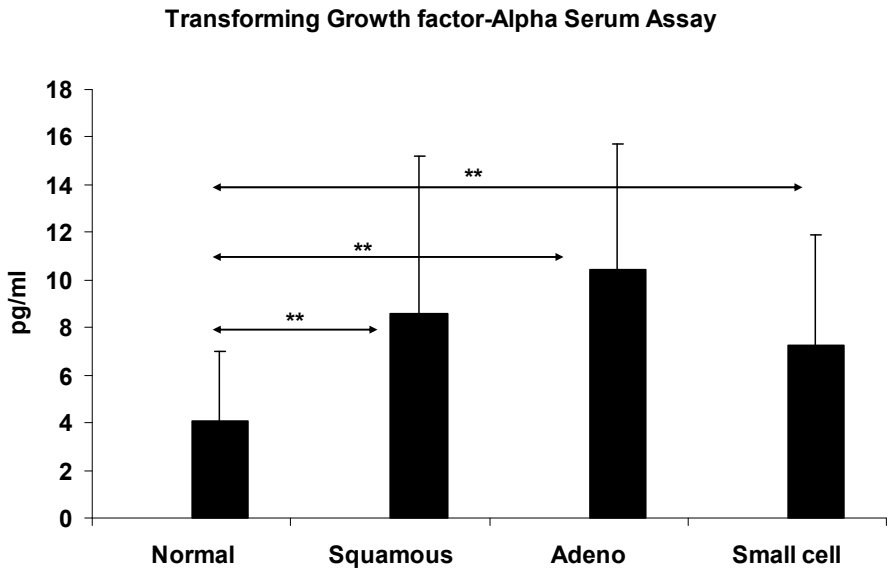


Figure 6.3.8: Tenascin C and transforming growth factor alpha from Cancer Panel 2. Bar chart showing normal compared to late-stage lung cancer type. Patient serum samples included 20 normal, 9 squamous, 9 adenocarcinoma and 10 small cell carcinoma. Student's t-test with p value less than 0.01 is indicated by two asterisks (**).

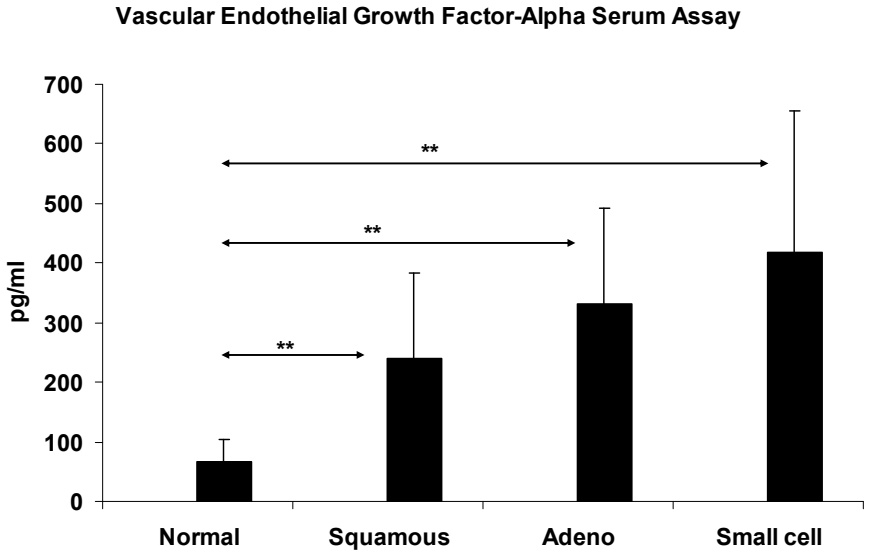


Figure 6.3.9: Vascular endothelial growth factor-alpha from Cancer Panel 2. Bar chart showing normal compared to late-stage lung cancer type. Patient serum samples included 20 normal, 9 squamous, 9 adenocarcinoma and 10 small cell carcinoma. Student's t-test with p value less than 0.01 is indicated by two asterisks (**).

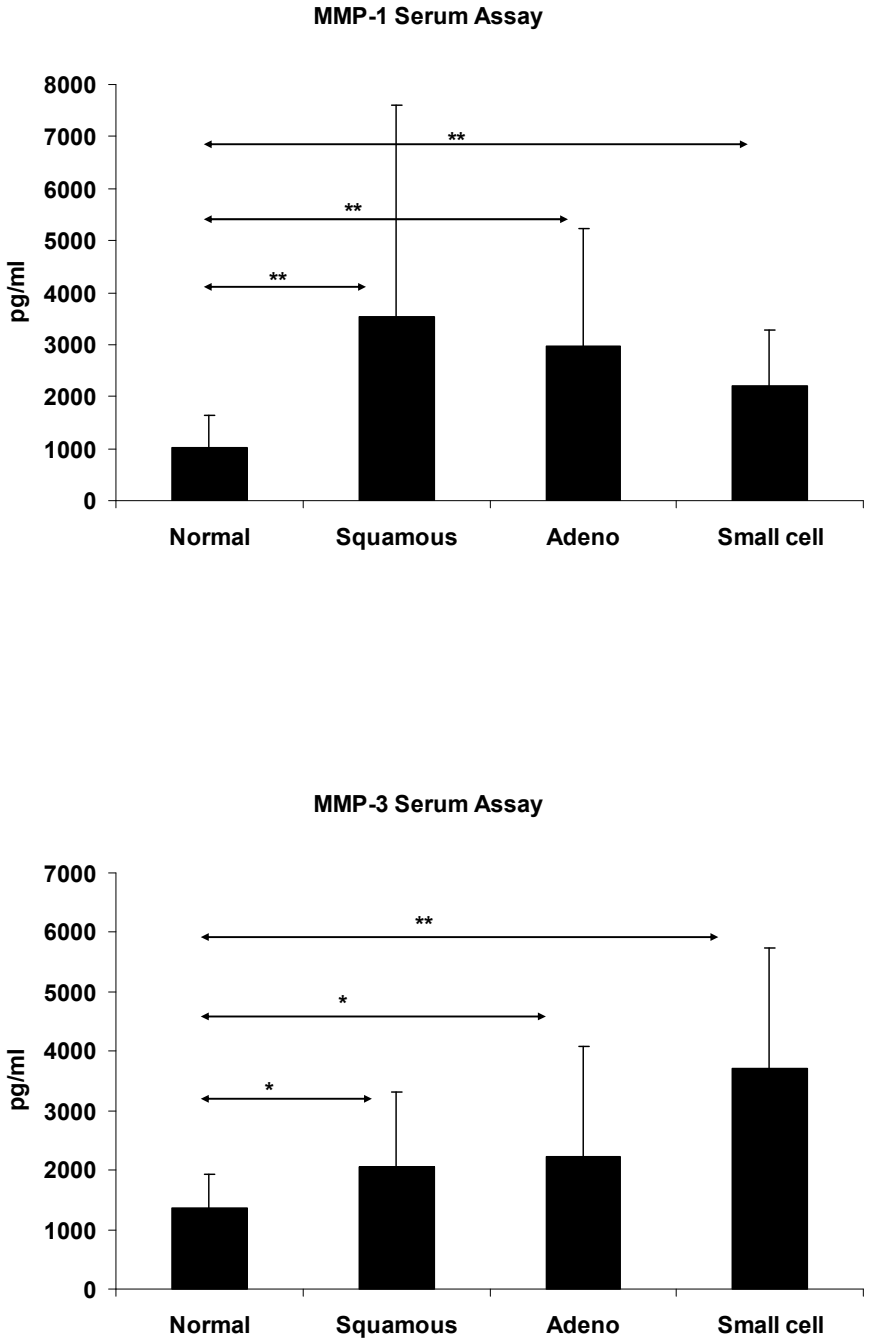
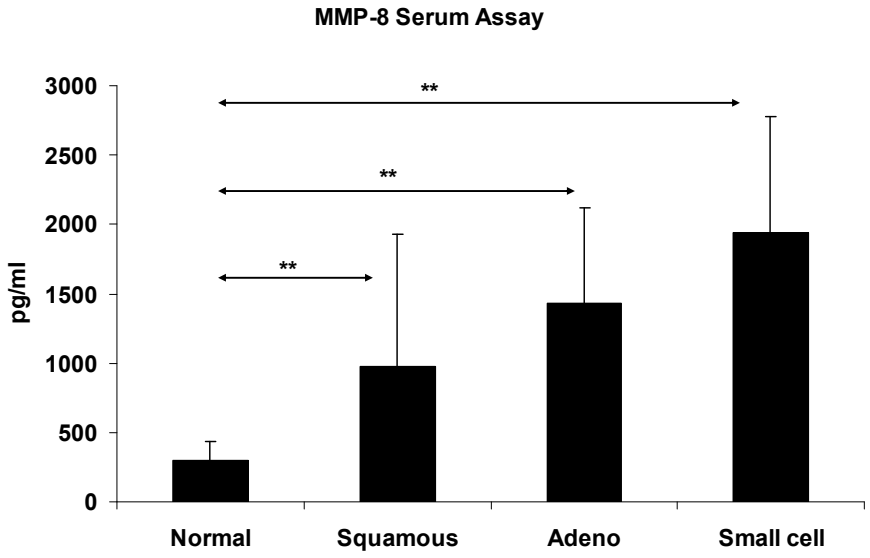


Figure 6.3.10: MMP-1 and MMP-3 from MMP Panel. Bar chart showing normal compared to late-stage lung cancer type. Patient serum samples included 20 normal, 9 squamous, 9 adenocarcinoma and 10 small cell carcinoma. Student's t-test with p value less than 0.05 is indicated by a single asterisk (*) and p value less than 0.01 is indicated by two asterisks (**).



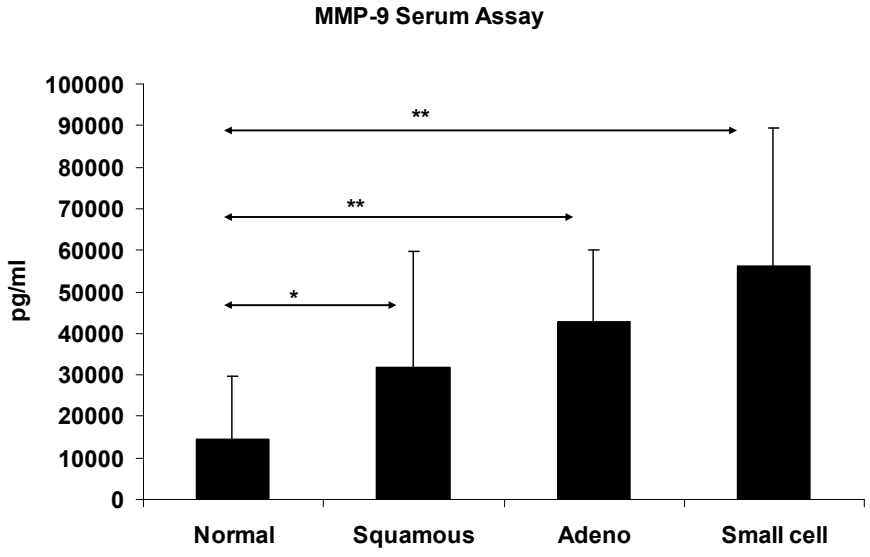
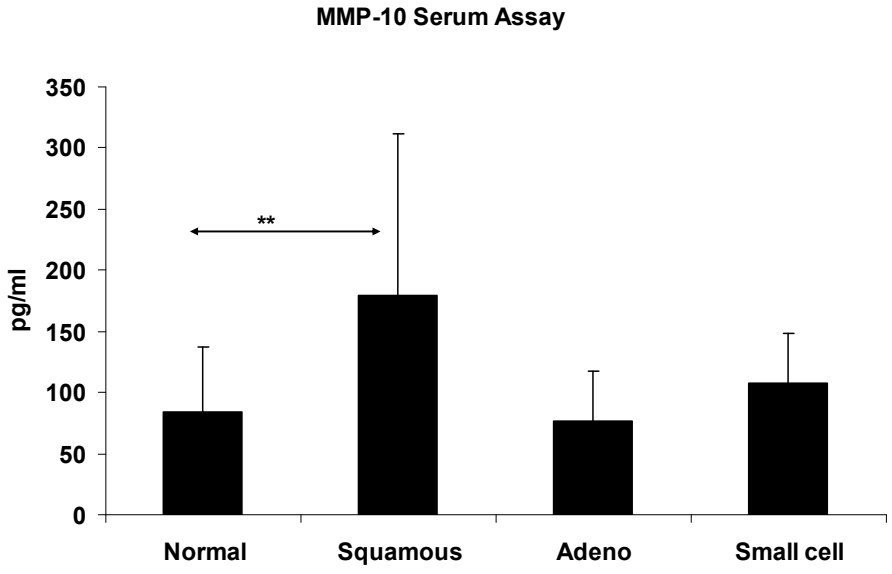


Figure 6.3.11: MMP-8 and MMP-9 from MMP Panel. Bar chart showing normal compared to late-stage lung cancer type. Patient serum samples included 20 normal, 9 squamous, 9 adenocarcinoma and 10 small cell carcinoma. Student's t-test with p value less than 0.05 is indicated by a single asterisk (*) and p value less than 0.01 is indicated by two asterisks (**).



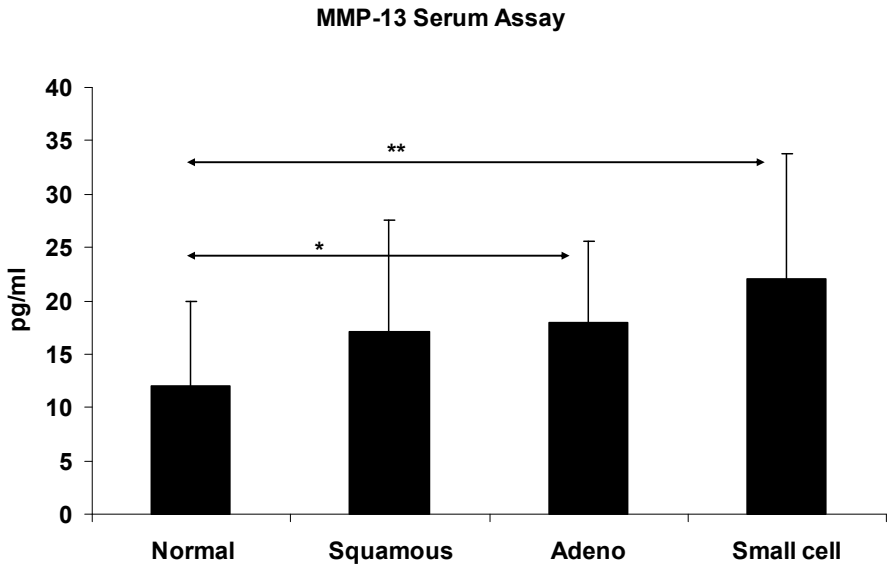


Figure 6.3.12: MMP-10 and MMP-13 from MMP Panel. Bar chart showing normal compared to late-stage lung cancer type. Patient serum samples included 20 normal, 9 squamous, 9 adenocarcinoma and 10 small cell carcinoma. Student's t-test with p value less than 0.05 is indicated by a single asterisk (*) and p value less than 0.01 is indicated by two asterisks (**).

6.4 Validation of targets from Luminex discovery phase: Cancer Panel 1

Cancer Panel 1 from Millipore was ran on the Luminex Multiplex platform includes the six proteins alpha-fetoprotein, CA 19-9, CEA, CA-125, CA 15-3, and prolactin. These proteins are known tumour markers, in some cases used for lung cancer though mainly used for other cancers. In the case where they are used for lung cancer there are issues with specificity and sensitivity, and often their main use is as a biological marker in other cancers.

Alpha-fetoprotein is a marker for hepatocellular carcinoma, a common type of liver cancer, and germ cell carcinoma. It is known to be elevated in normal pregnancy, benign liver diseases, as well as in cancer. It is less frequently elevated in bronchogenic cancers and rarely elevated in healthy patients.

CA-125 is an antigen present on eighty percent of nonmucinous ovarian carcinomas and is often elevated in patients with ovarian cancer, where it is used as a marker for patient response to surgical resection or chemotherapy. It is elevated in other cancers such as endometrial, pancreatic, breast, colon, and also lung cancer.

Elevated levels of CEA are found in many cancers such as colon, pancreatic, gastric, breast, and lung. However it is also detected in benign conditions such as pancreatitis, chronic lung disease, cirrhosis, and inflammatory bowel disease, as well as being found to be elevated in 19% of smokers and 3% of a healthy control population.

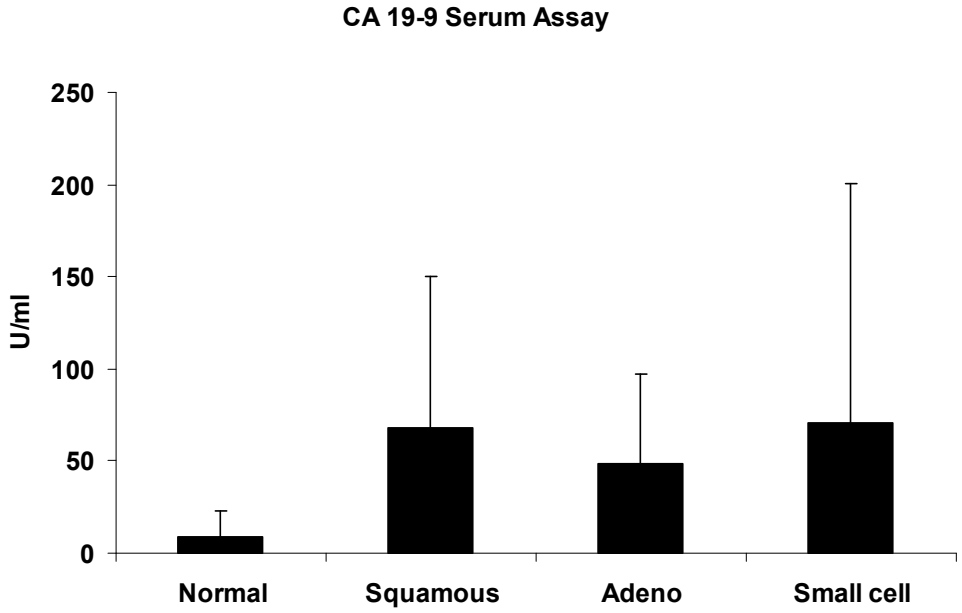
CA 19-9 is a primary marker in 80% of cases of pancreatic carcinoma, with good correlation between serum levels and tumour size. CA 19-9 can be used as a secondary marker after CEA in biliary carcinomas, colorectal cancer, liver

metastases and gastric carcinoma. Increases in CA 19-9 are also reported in breast, ovarian, and lung cancer.

CA 15-3 is the most widely used serum marker in breast cancer. It is used in pre-clinically detecting recurrent breast cancer and monitoring the treatment of patients with advanced breast cancer. Its levels are rarely elevated in women with early stage breast cancer. Elevated levels of CA 15-3 are seen in colon and lung cancer and also in noncancerous conditions such as benign ovarian or breast disease, pelvic inflammatory disease and hepatitis.

Prolactin has been mentioned as a possible marker for cancer of the cervix and also colorectal cancer. It has also been postulated that prolactin levels in plasma have a correlation to breast tumour size, though this has been refuted (Arslan, Serdar et al. 2000).

Although not specific for lung cancer, the inclusion of known cancer biomarkers is a useful benchmark for judging the overall effectiveness of the Luminex multiplex system. An expected result in these markers leads to confidence in the results of the less studied/new markers and vice versa.



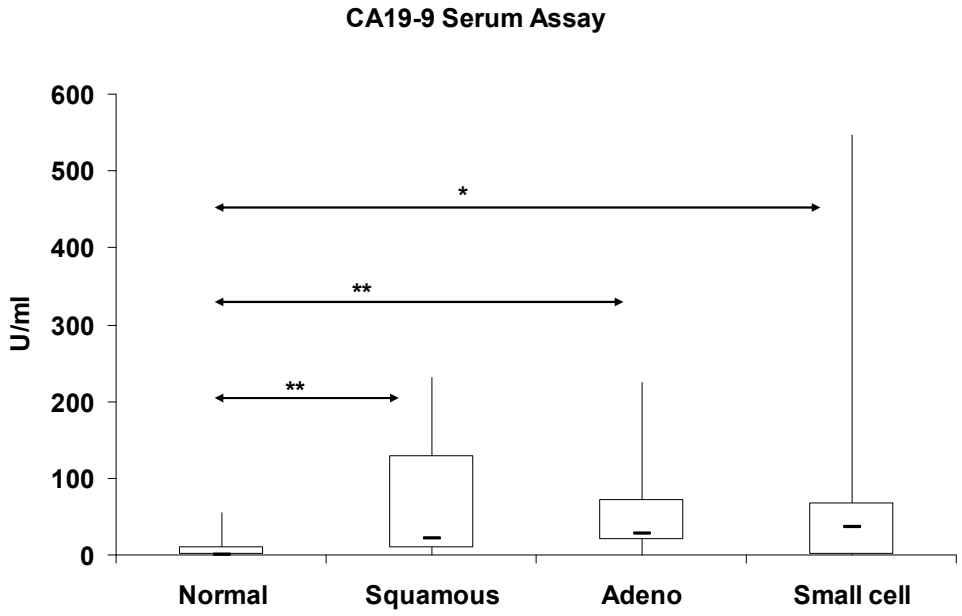
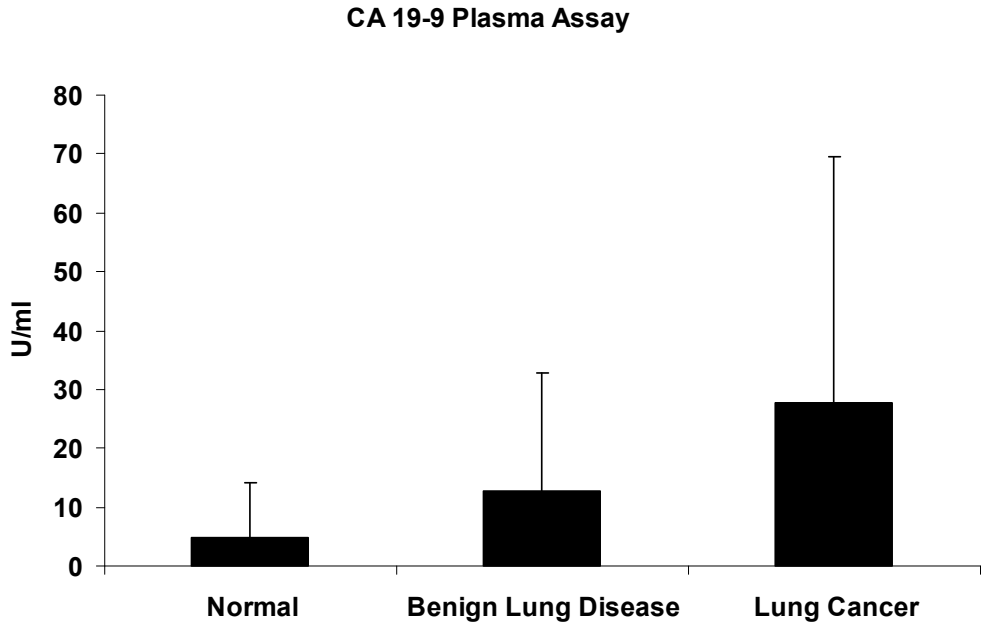


Figure 6.4.1: CA19-9 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.003), Normal v AD (*p* value=0.001), Normal v SCLC (*p* value=0.04).



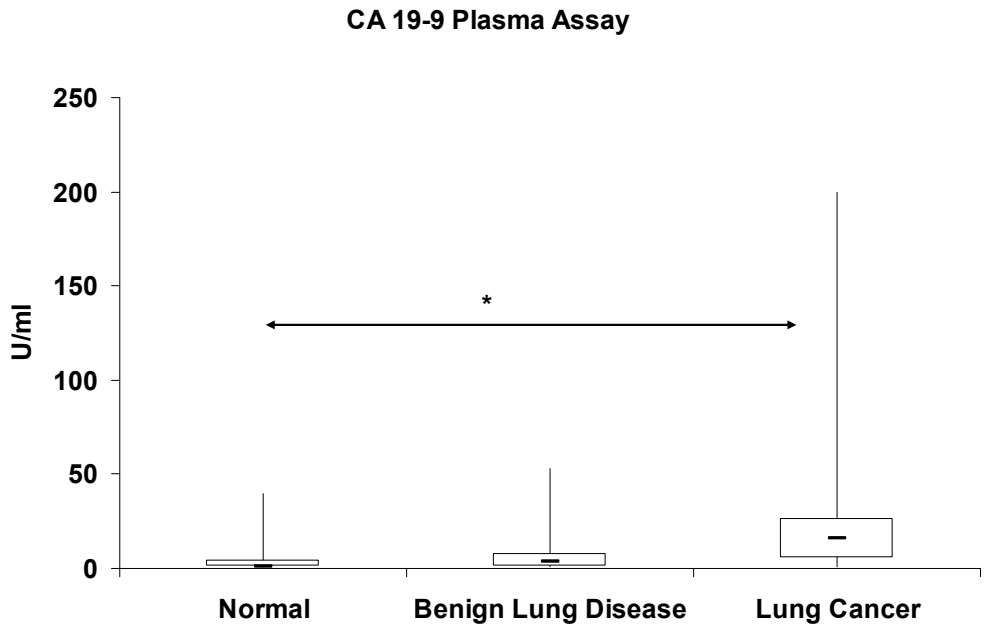


Figure 6.4.2: CA19-9 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to benign lung disease and lung cancer. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles.Patient plasma samples used for this experiment included 18 normal, 11 benign lung disease, and 35 lung cancer. T-test score; Normal v Lung Cancer (*p* value=0.03).

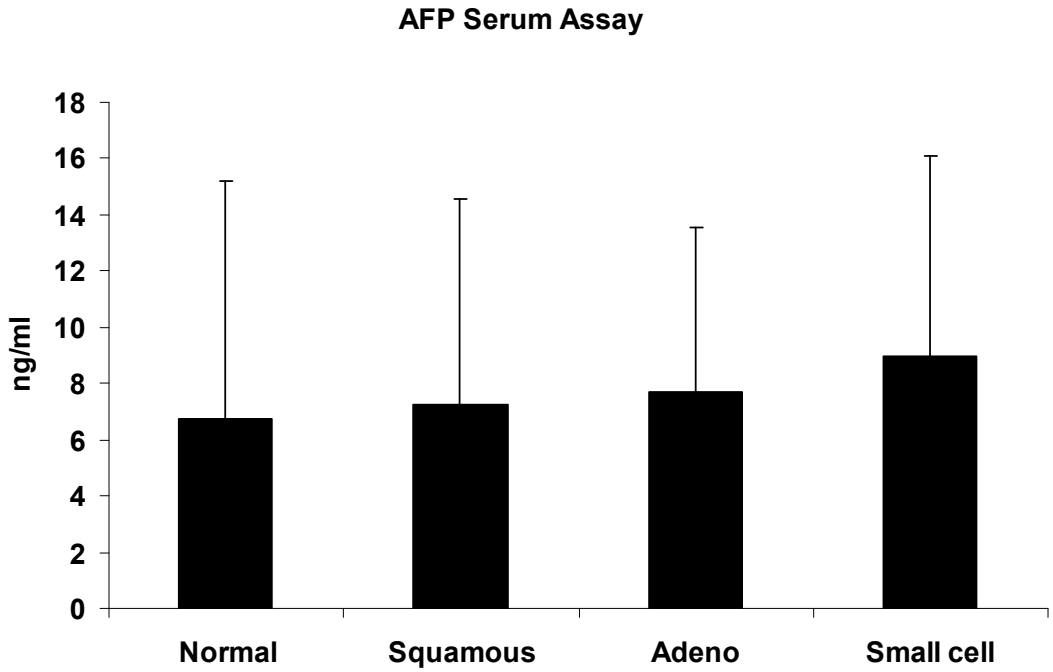
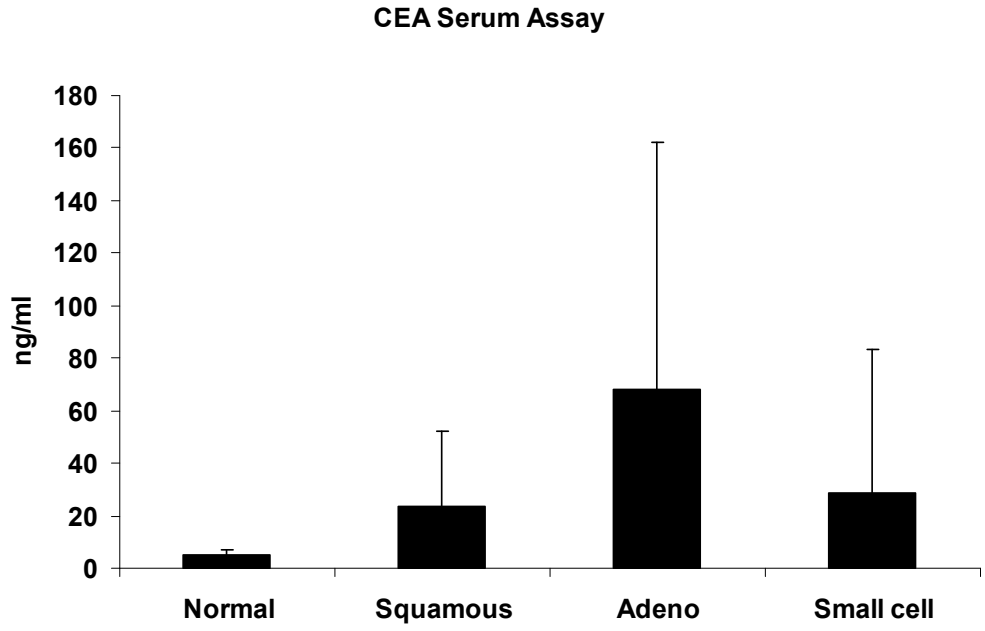


Figure 6.4.3: Alpha-fetoprotein Luminex Assay: Bar chart showing normal compared to each lung cancer type. Patient serum samples used for this experiment included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma.



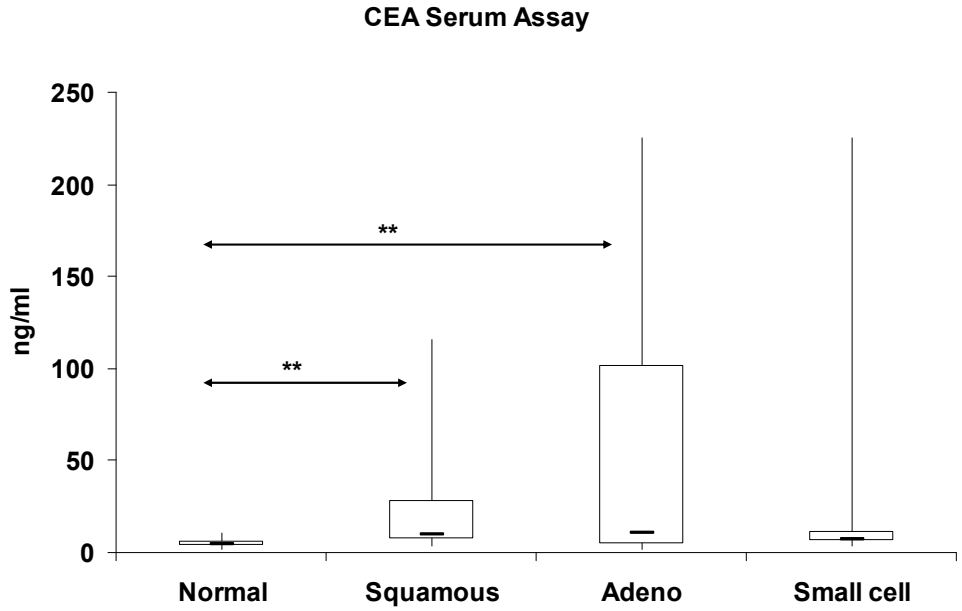
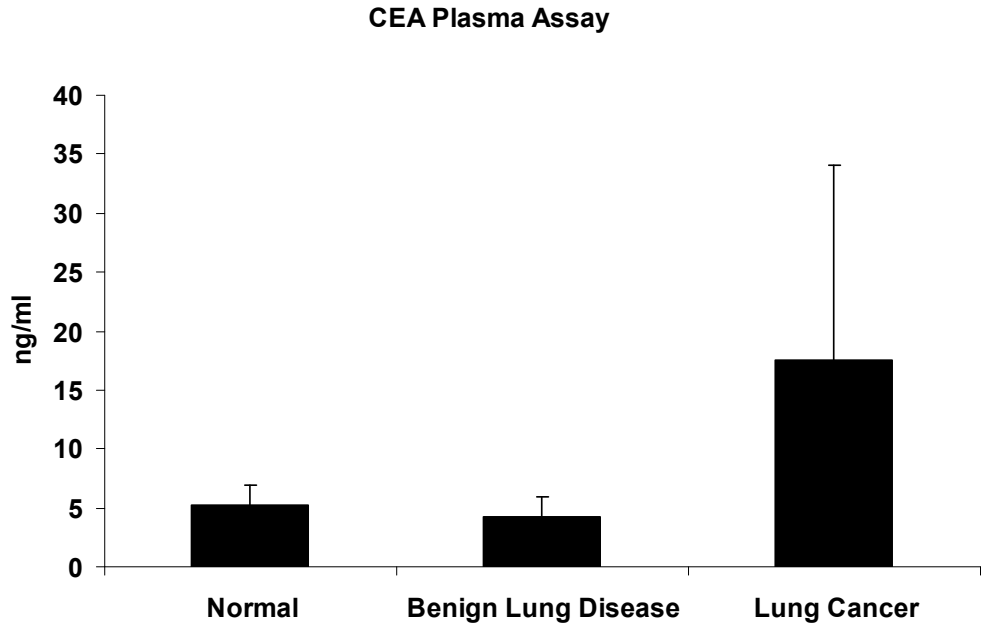


Figure 6.4.4: CEA Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles.Patient serum samples used for this experiment included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.008), Normal v AD (*p* value=0.005).



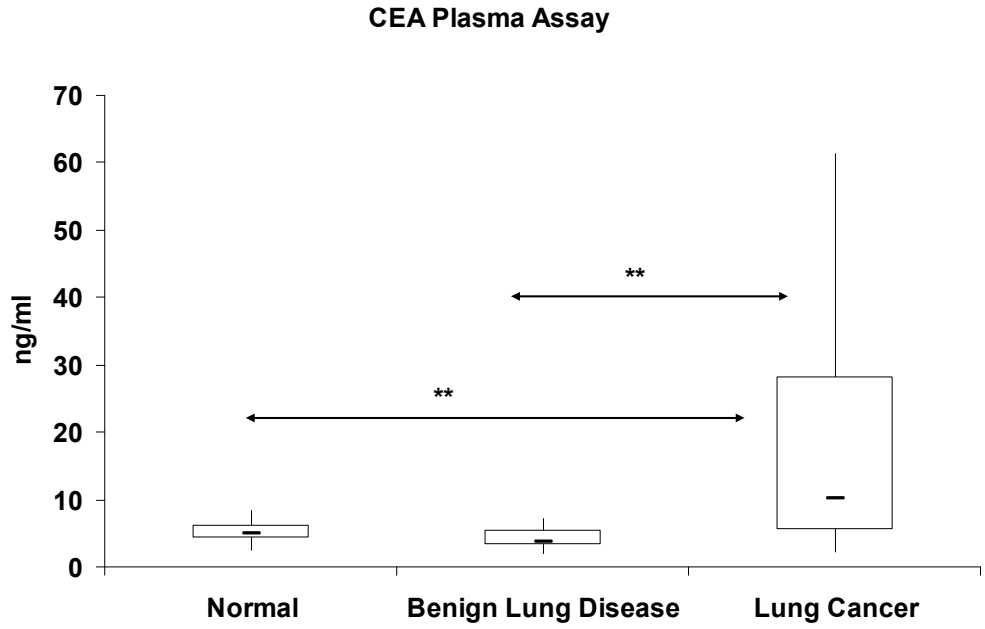


Figure 6.4.5: CEA Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to benign lung disease and early-stage lung cancer. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles.Patient plasma samples used for this experiment included 18 normal, 11 benign lung disease, and 35 lung cancer. T-test score; Normal v Lung Cancer (*p* value=0.003), BLD v Lung Cancer (*p* value=0.01).

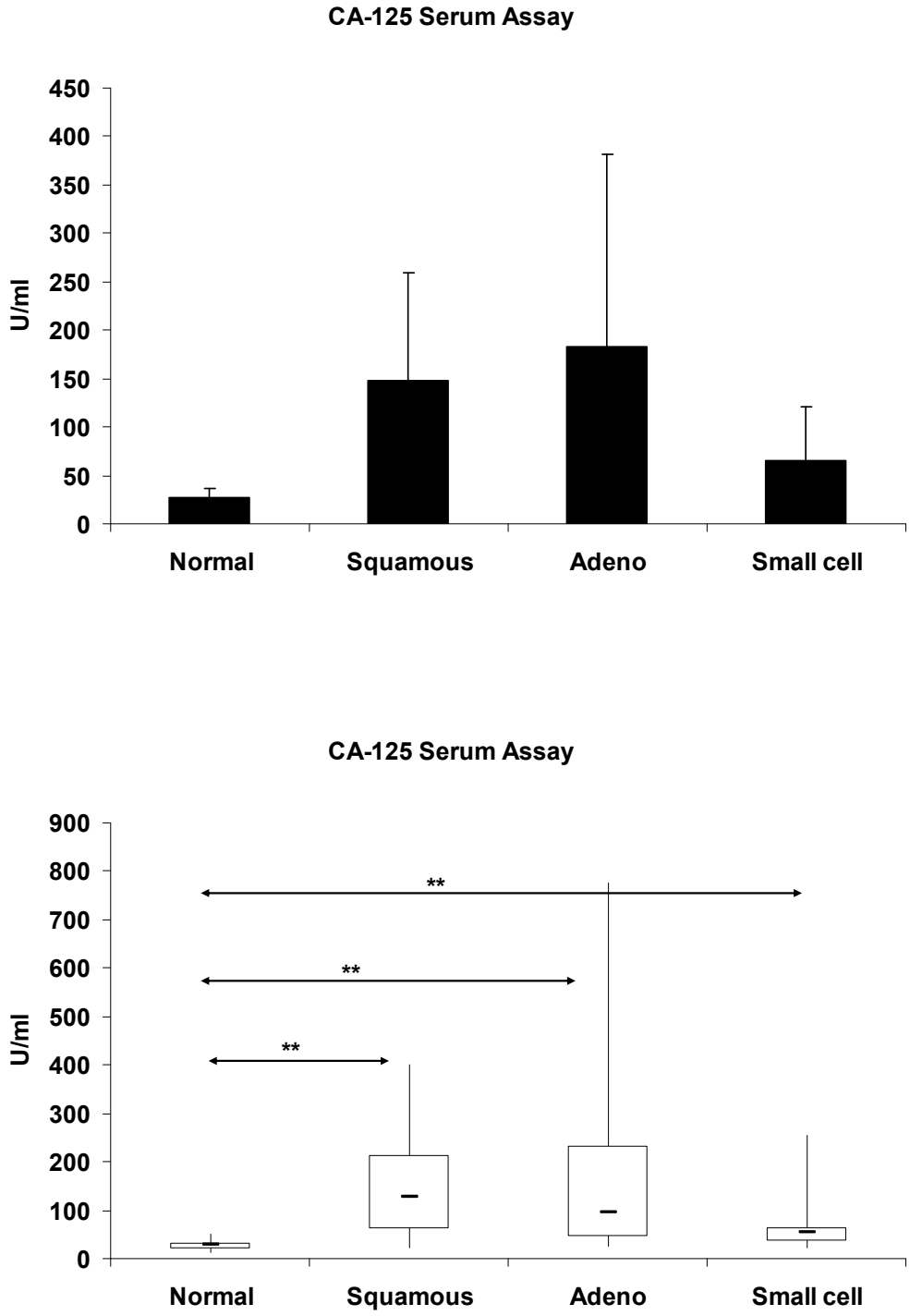


Figure 6.4.6: CA-125 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma. T-test scores; Normal v SCC (*p* value=3E-05), Normal v AD (*p* value=0.001), Normal v SCLC (*p* value=0.004).

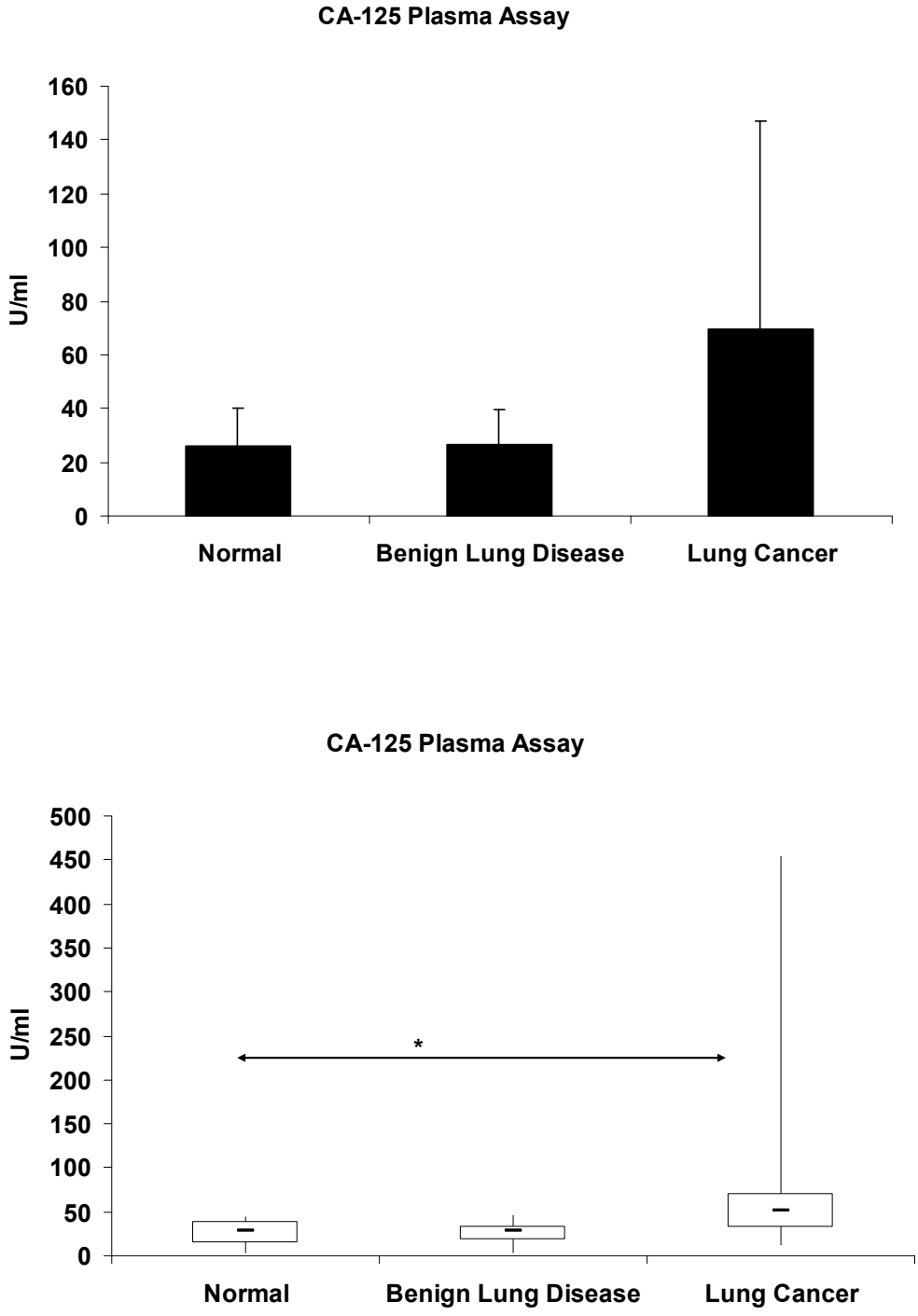
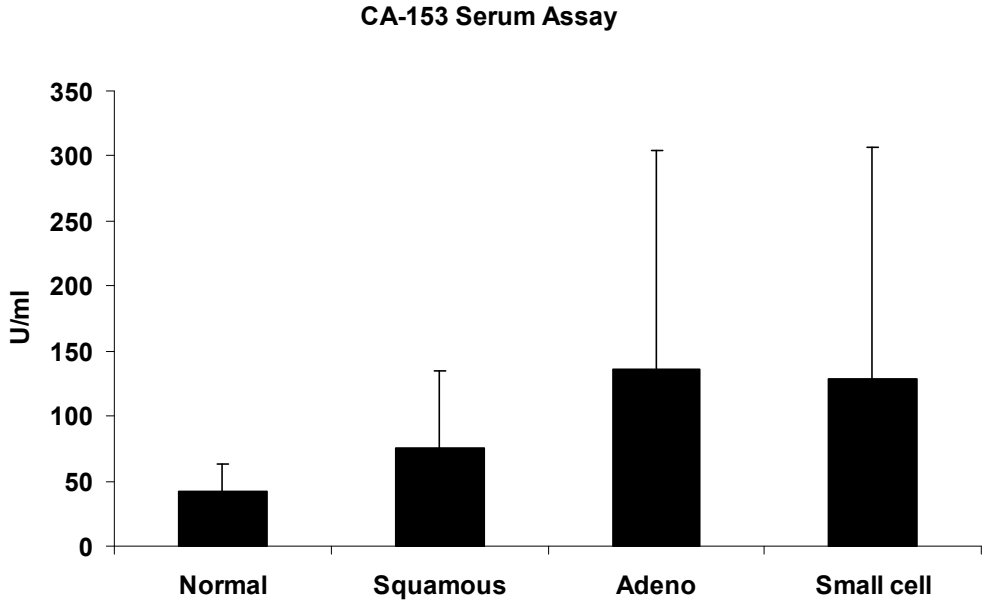


Figure 6.4.7: CA-125 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to benign lung disease and early-stage lung cancer. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles.Patient plasma samples used for this experiment included 18 normal, 11 benign lung disease, and 35 lung cancer. T-test score; Normal v Lung Cancer (*p* value=0.02).



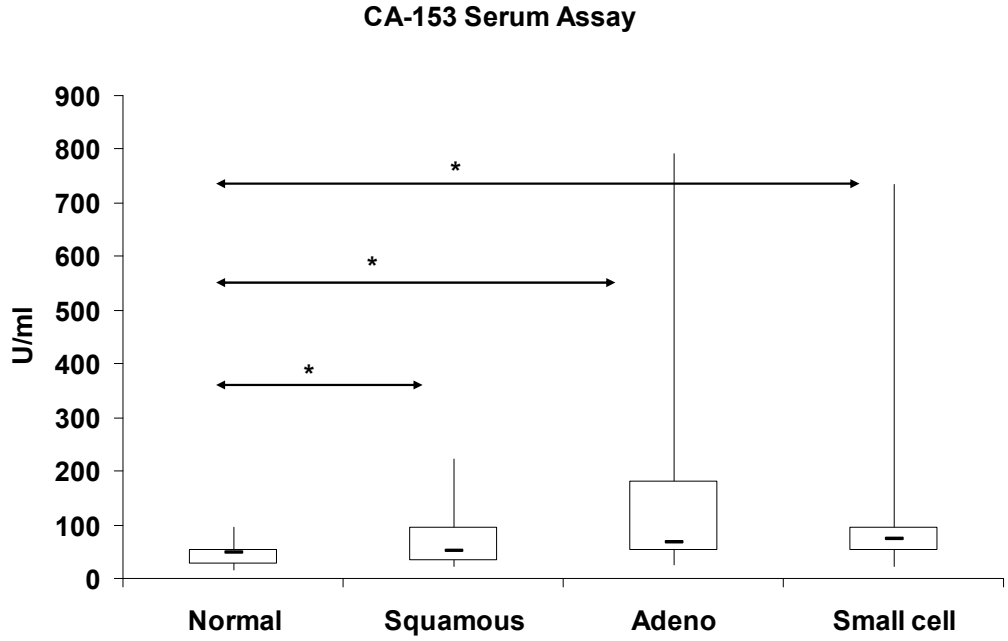


Figure 6.4.8: CA-153 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.02), Normal v AD (*p* value=0.02), Normal v SCLC (*p* value=0.04).

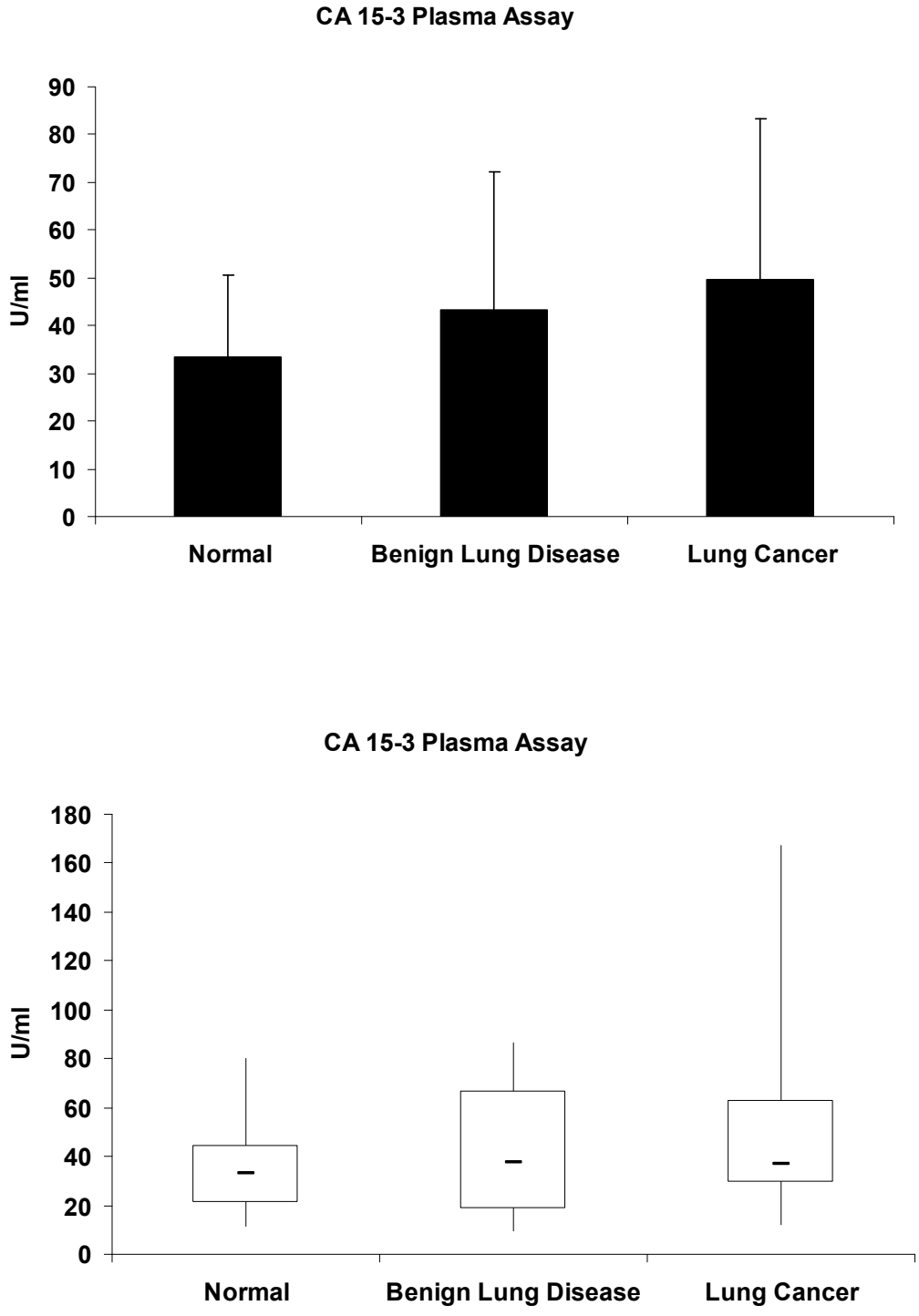
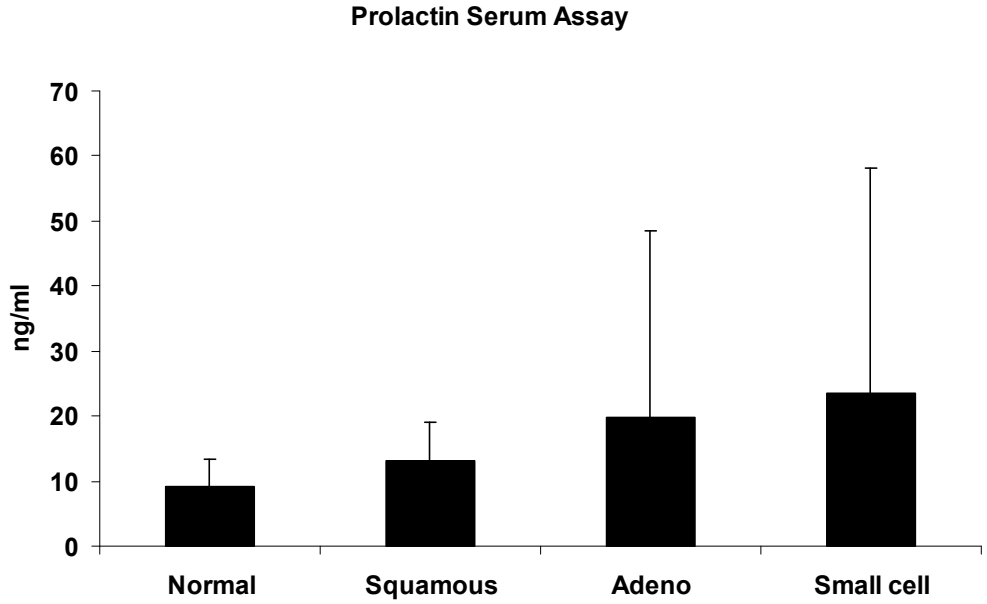


Figure 6.4.9: CA 15-3 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to benign lung disease and early-stage lung cancer. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles.Patient plasma samples used for this experiment included 18 normal, 11 benign lung disease, and 35 lung cancer.



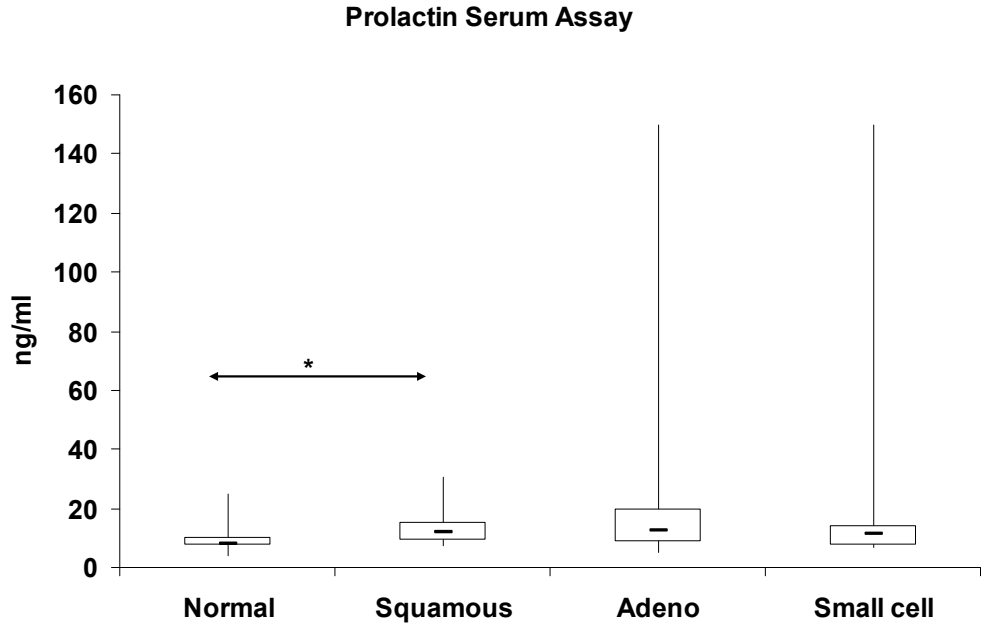
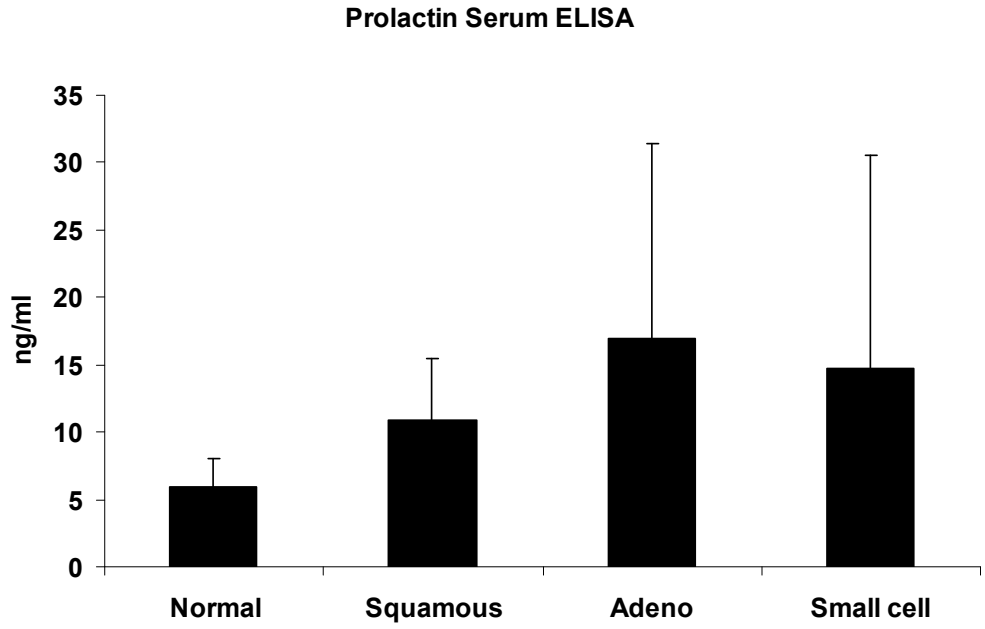


Figure 6.4.10: Prolactin Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each latre-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.02).



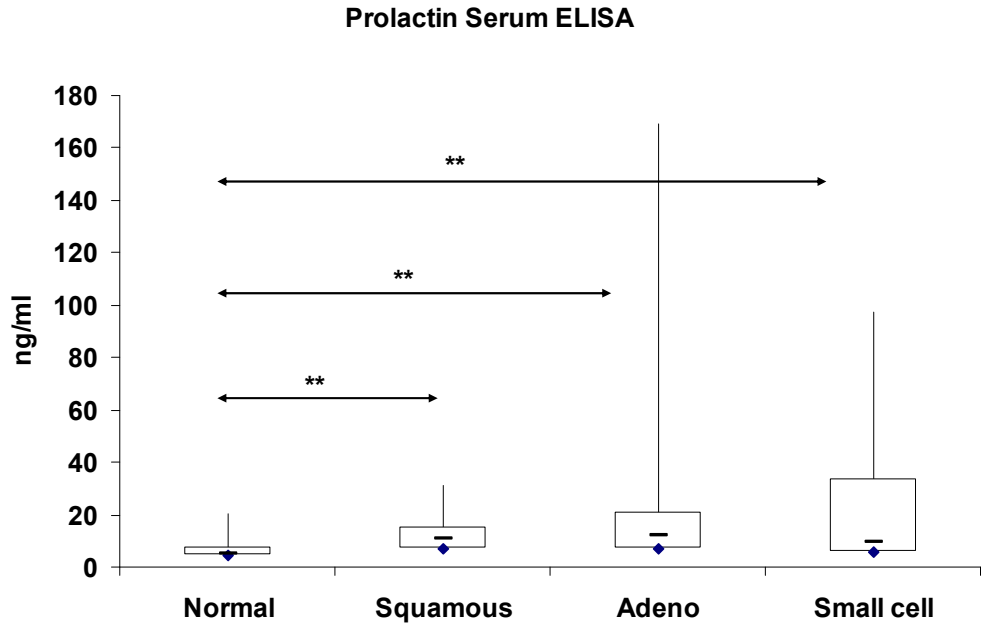


Figure 6.4.11: Prolactin ELISA: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 21 normal, 15 squamous, 23 adenocarcinoma, and 16 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.0001), Normal v AD (*p* value=0.001), Normal v SCLC (*p* value=0.01).

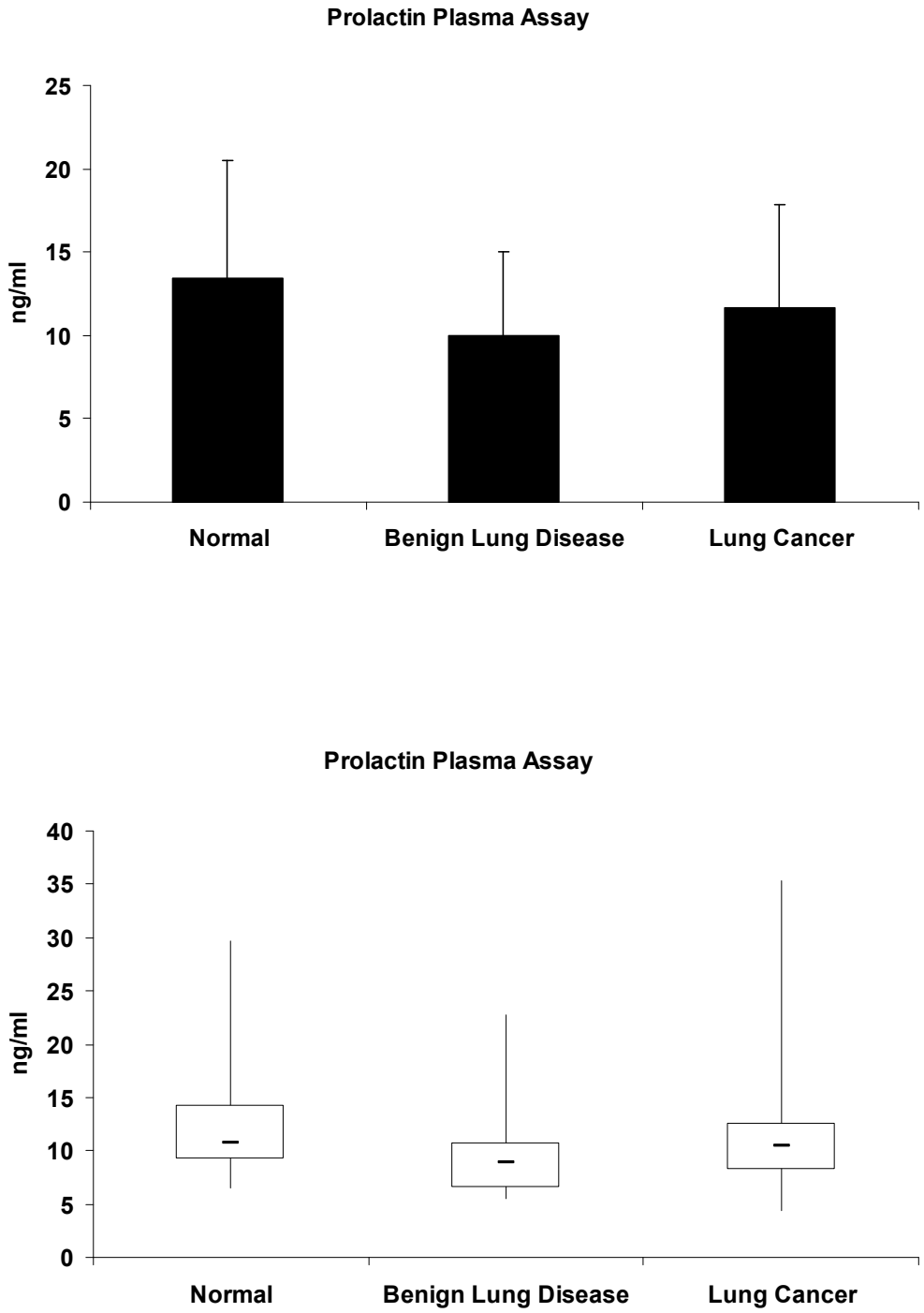


Figure 6.4.12: Prolactin Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to benign lung disease and early-stage lung cancer. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient plasma samples used for this experiment included 18 normal, 11 benign lung disease, and 35 lung cancer.

6.5 Validation of targets from Luminex discovery phase: Cancer Panel 2

From Cancer Panel 2 six proteins were brought forward for validation in larger serum sample sets. This 6plex included epiregulin, heparin-binding epidermal growth factor, placenta growth factor, transforming growth factor alpha, vascular epithelial growth factor (VEGF-α), and tenascin C. From this 6plex Luminex experiment, tenascin C and VEGF-α showed the most promise and were validated further in serum and also in plasma. One VEGF-α experiment in serum (without a tenascin C equivalent) included an autoimmune control, a benign lung disease control, and earlier stage lung cancer.

VEGF-α has a history of being linked to lung development and also lung cancer itself. It is known to play a role in lung development and maintenance of adult lung structure, and its effects on pulmonary endothelial cells are of importance in the pathology of lung diseases. Its potential as a lung cancer biomarker is well known, with high levels of VEGF in lung cancer serum associated with poor prognosis, and unfavourable survival and poor response to treatment with SCLC patients.

Tenascin-C is an extracellular matrix glycoprotein that is expressed upon damage to tissue. It is also known to modulate cell migration and proliferation making it an ideal candidate as a cancer biomarker. It has been suggested its levels in serum could be a predictive marker of angiogenesis in NSCLC and its expression in breast cancer cells has been linked to their ability to metastasise to the lungs.

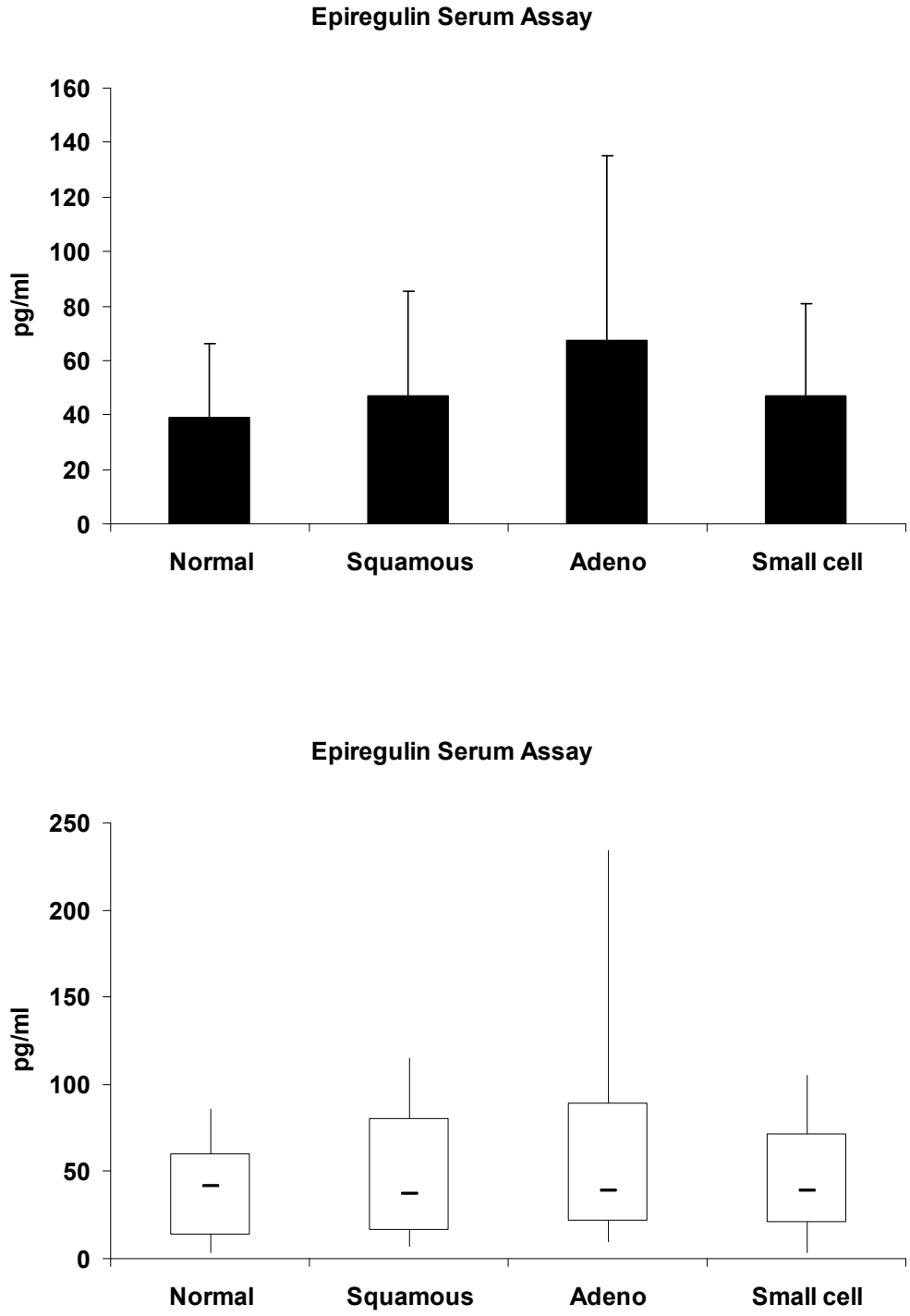
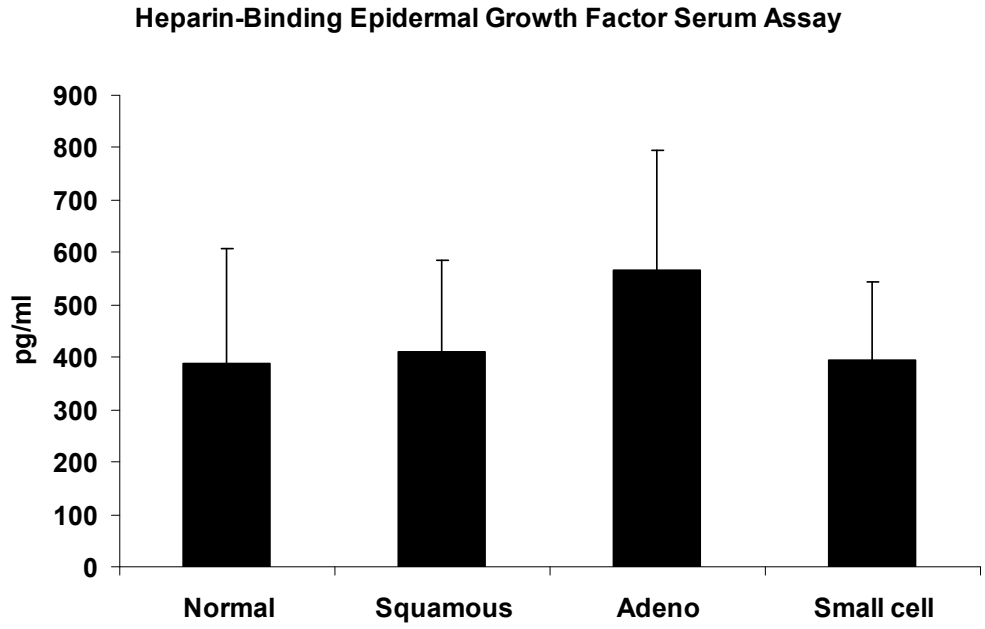


Figure 6.5.1: Epiregulin Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles.Patient serum samples used for this experiment included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma.



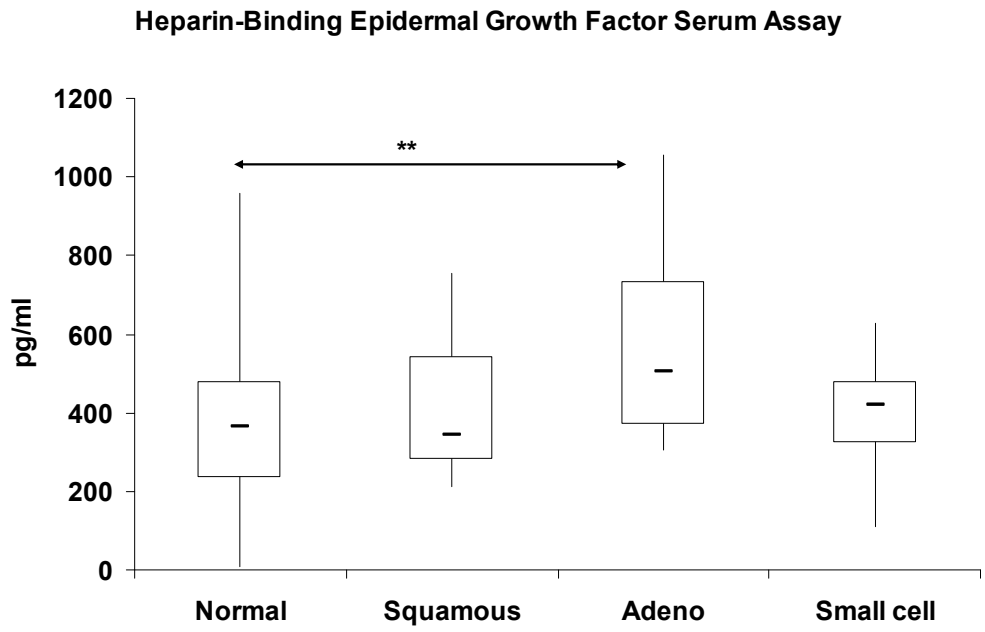
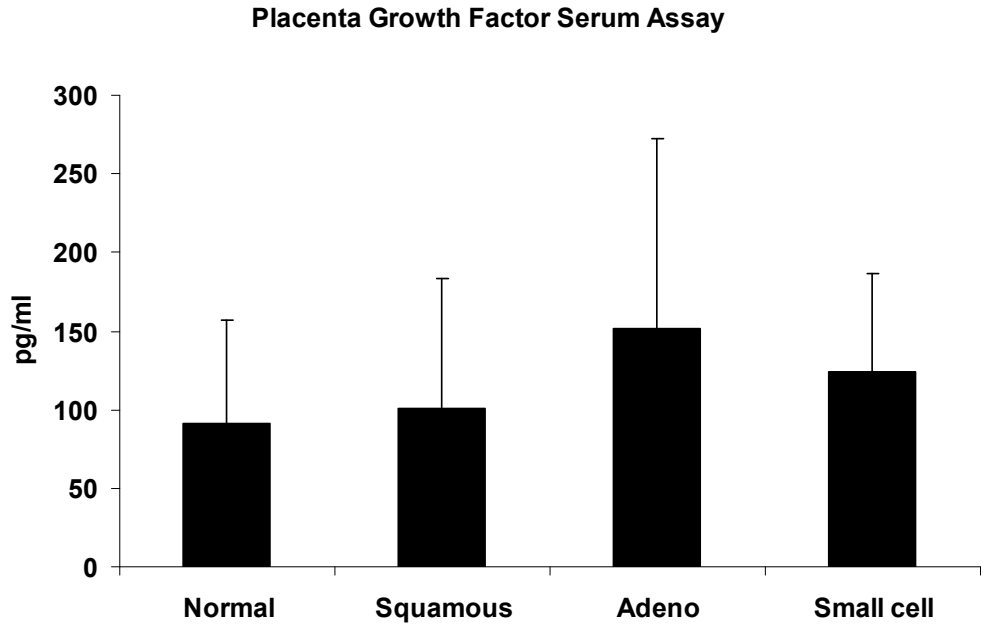


Figure 6.5.2: Heparin-binding Epidermal Growth Factor Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma. T-test score; Normal v AD (*p* value=0.01).



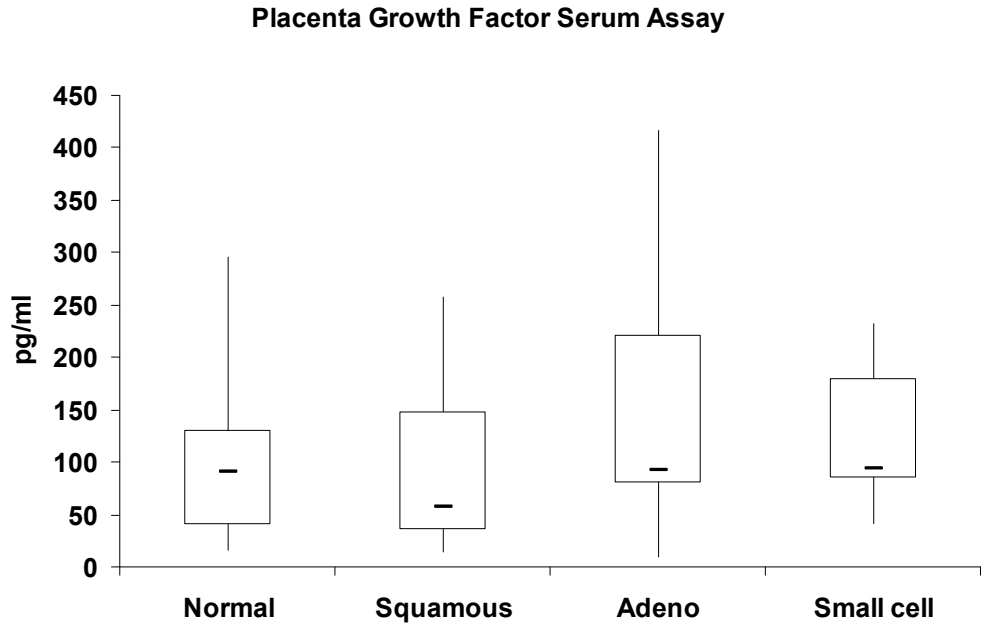
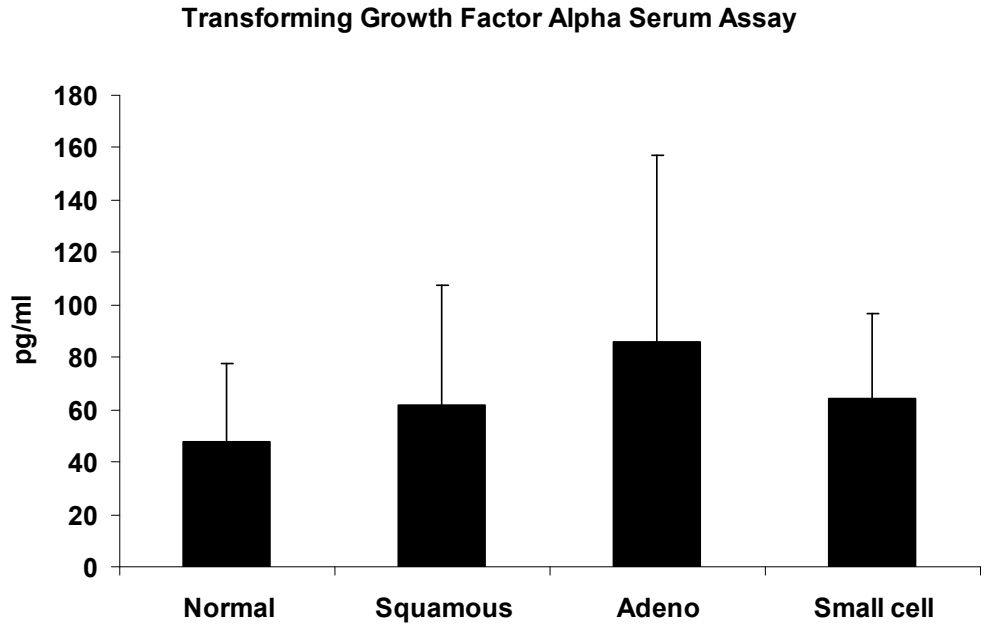


Figure 6.5.3: Placenta Growth Factor Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma.



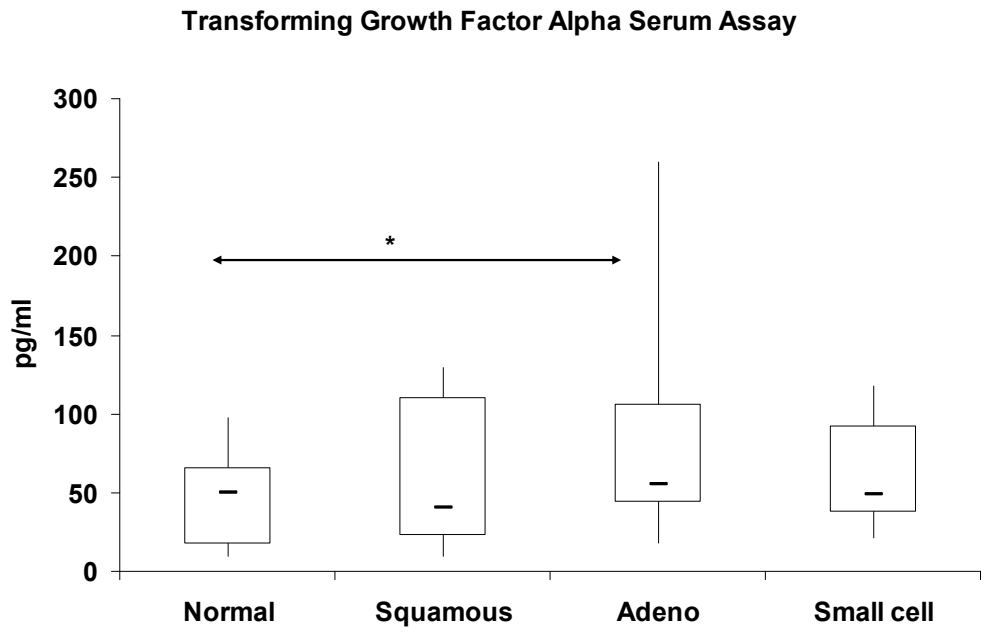
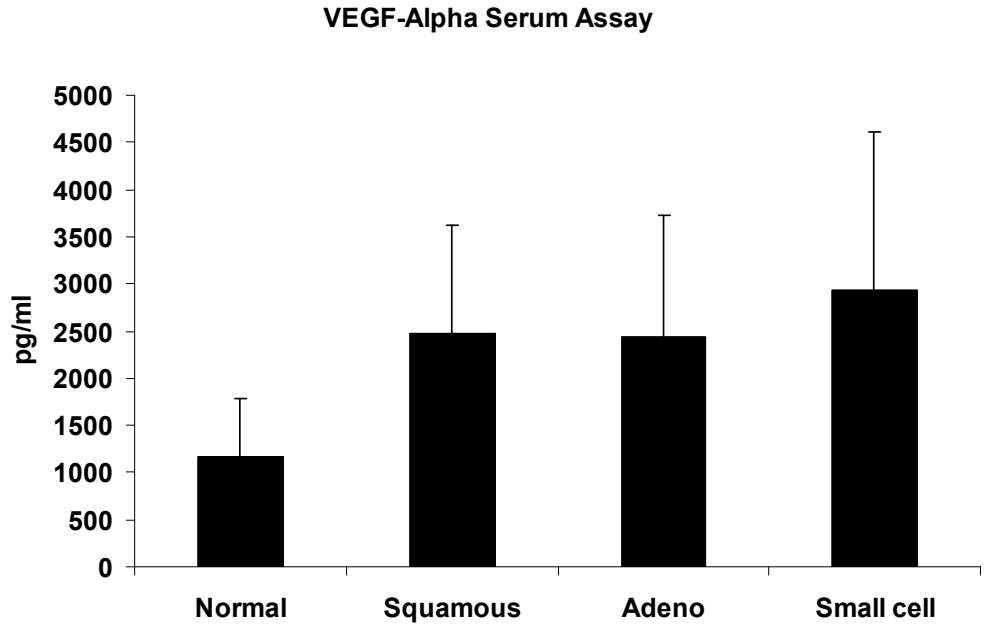


Figure 6.5.4: Transforming Growth Factor Alpha Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma. T-test score; Normal v AD (*p* value=0.04).



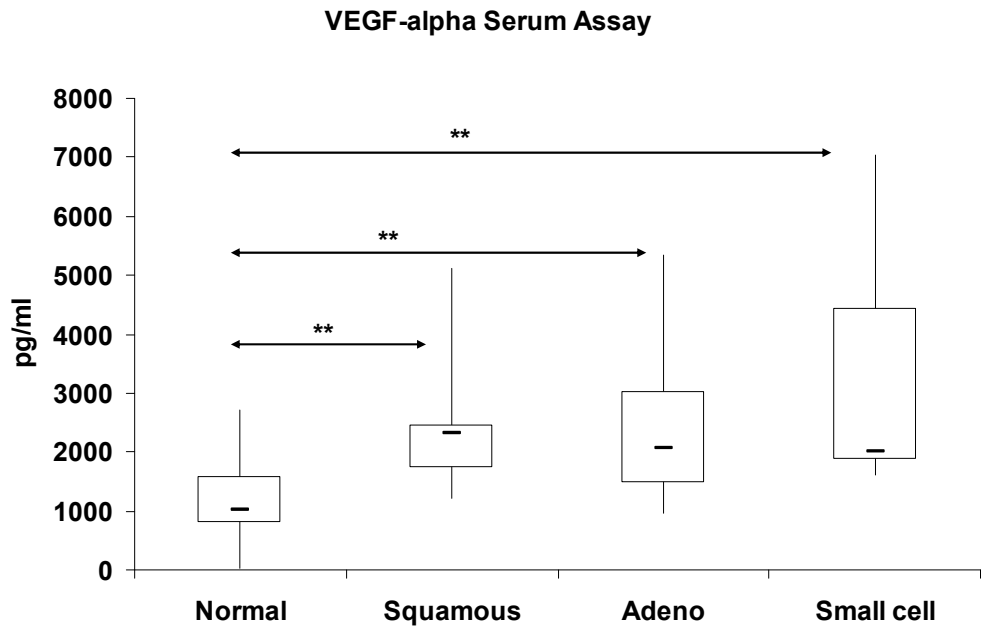
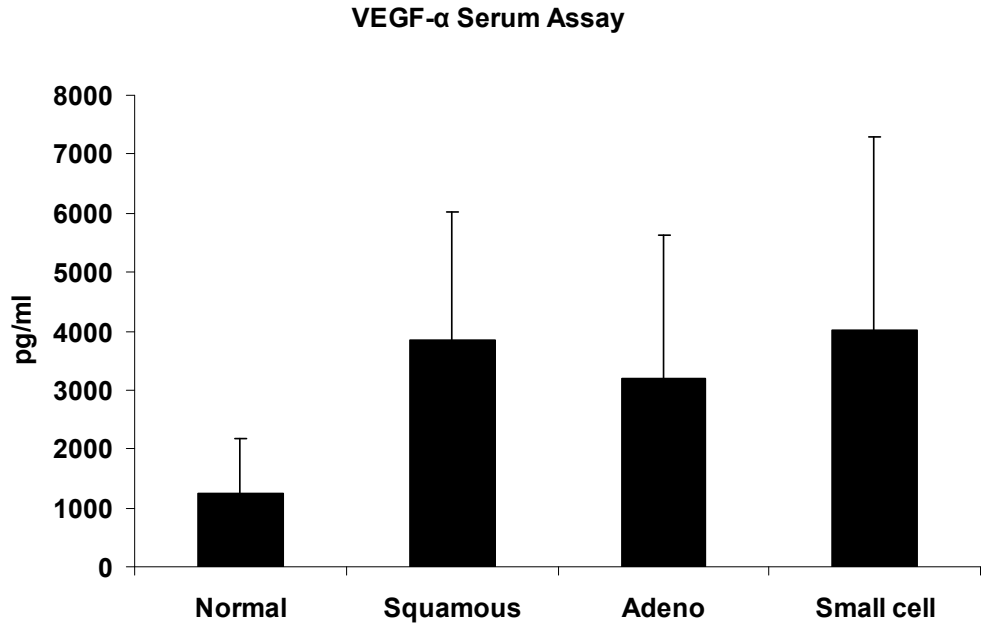


Figure 6.5.5: Vascular endothelial growth factor-alpha Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.0001), Normal v AD (*p* value=0.0003), Normal v SCLC (*p* value=0.0001).



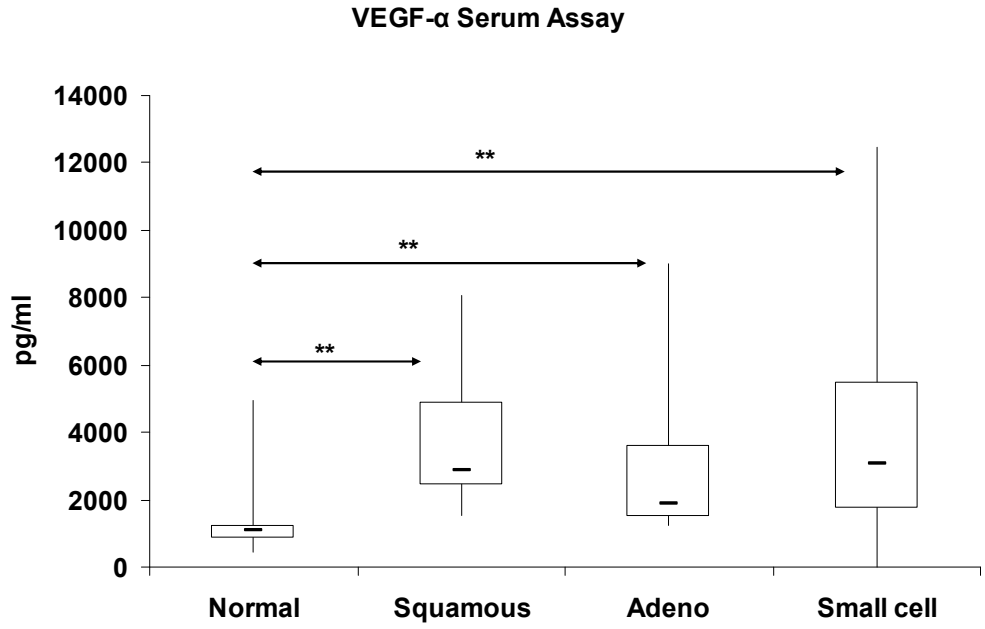


Figure 6.5.6: Vascular endothelial growth factor-alpha Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 14 squamous, 21 adenocarcinoma, and 16 small cell carcinoma. T-test score; Normal v SCC (*p* value=4E-05), Normal v AD (*p* value=0.002), Normal v SCLC (*p* value=0.001).

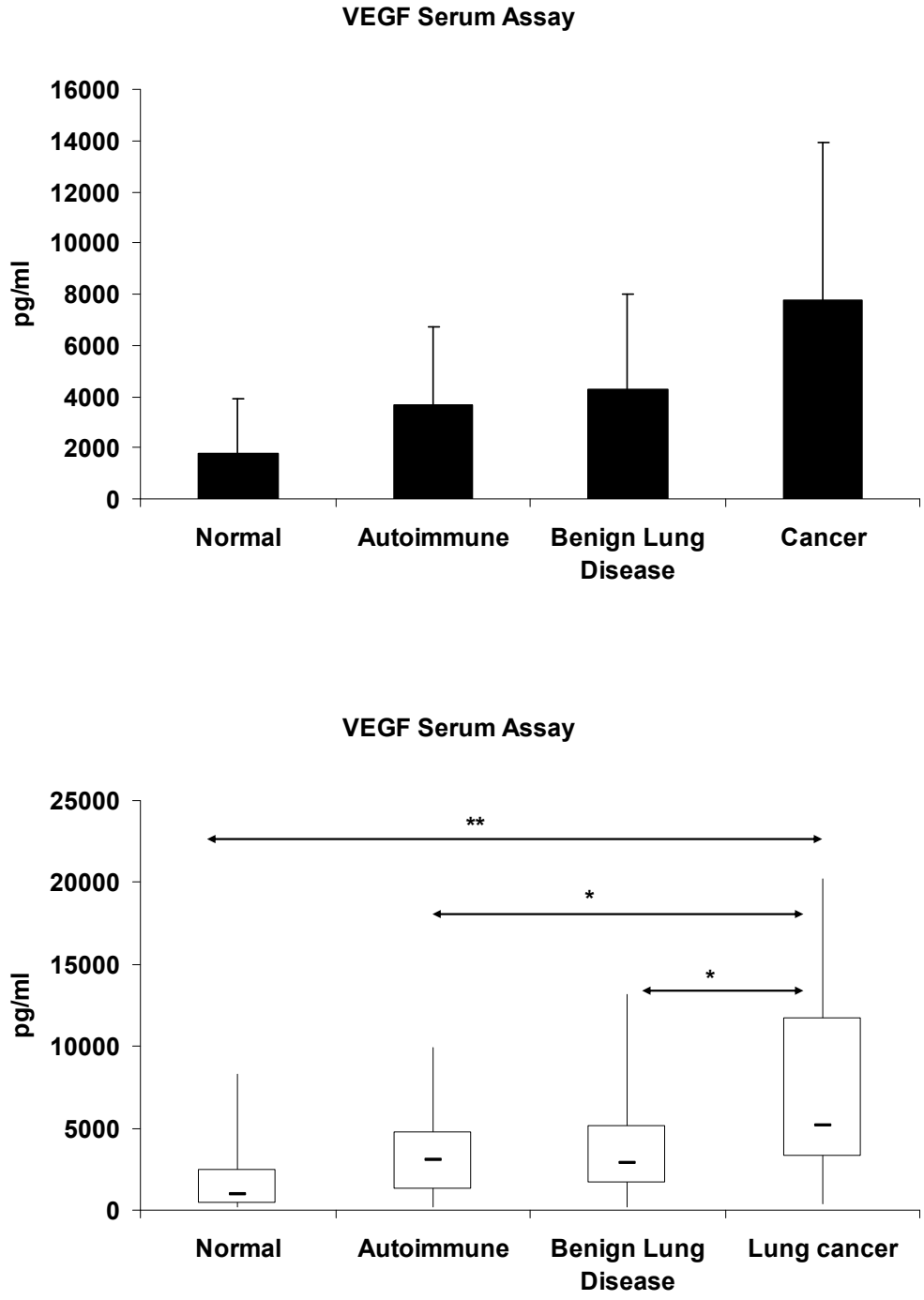
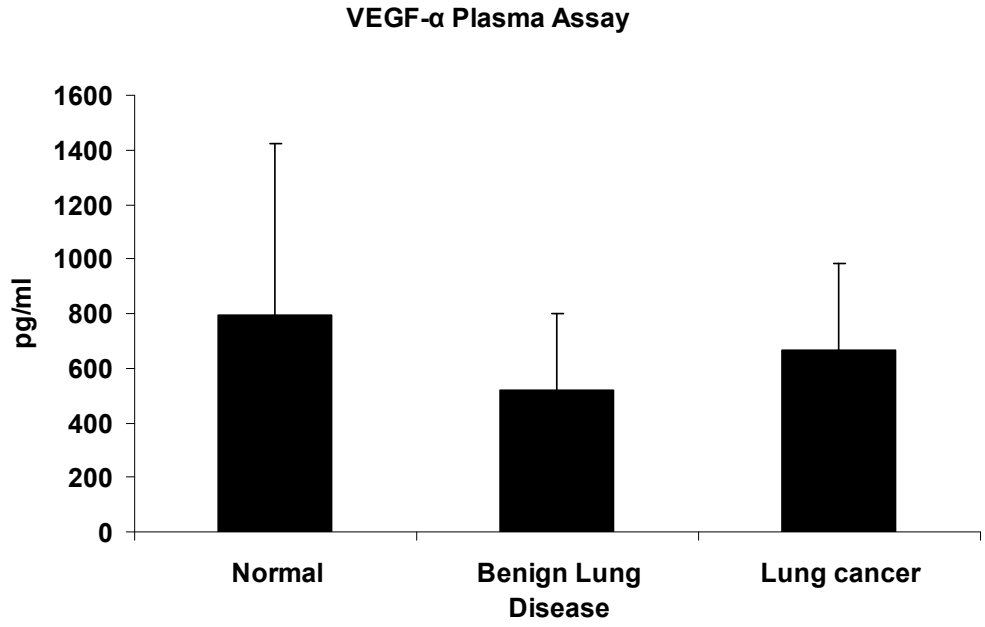


Figure 6.5.7: Vascular endothelial growth factor-alpha Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each early-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 16 normal, 16 autoimmune, 22 benign lung disease, and 26 lung cancer. T-test value; Normal v Lung Cancer (*p* value=0.0008), Autoimmune v Lung Cancer (*p* value=0.02), BLD v Lung Cancer (*p* value=0.03).



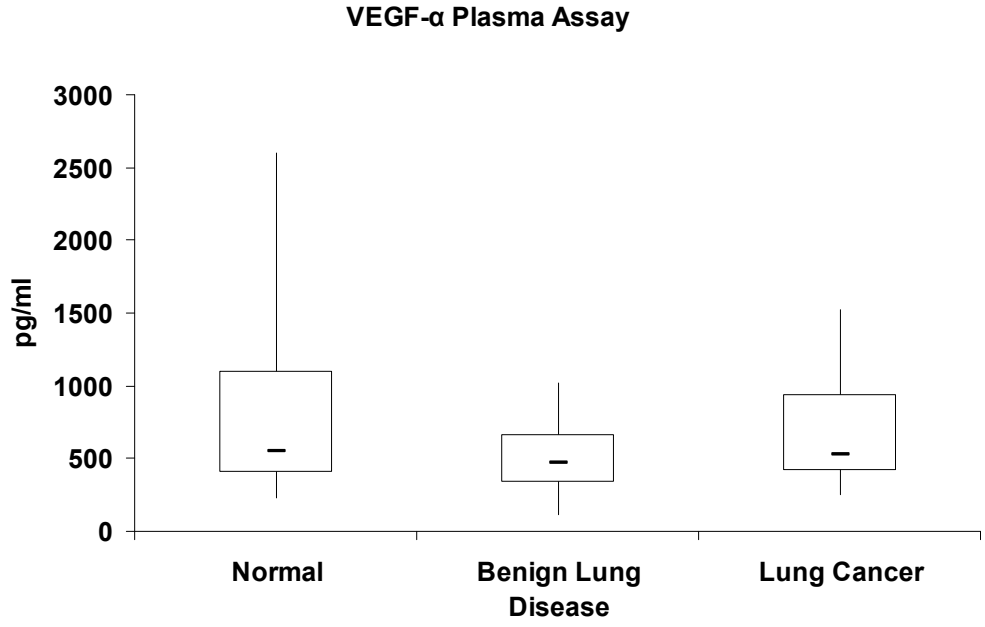
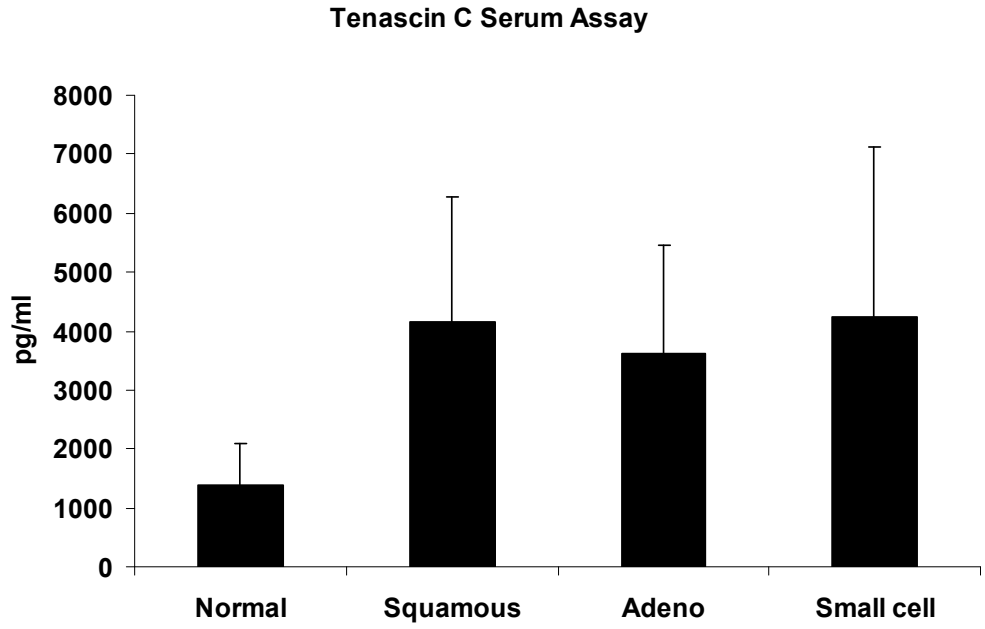


Figure 6.5.8: Vascular endothelial growth factor-alpha Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to benign lung disease and early-stage lung cancer. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient plasma samples used for this experiment included 18 normal, 11 benign lung disease, and 35 lung cancer.



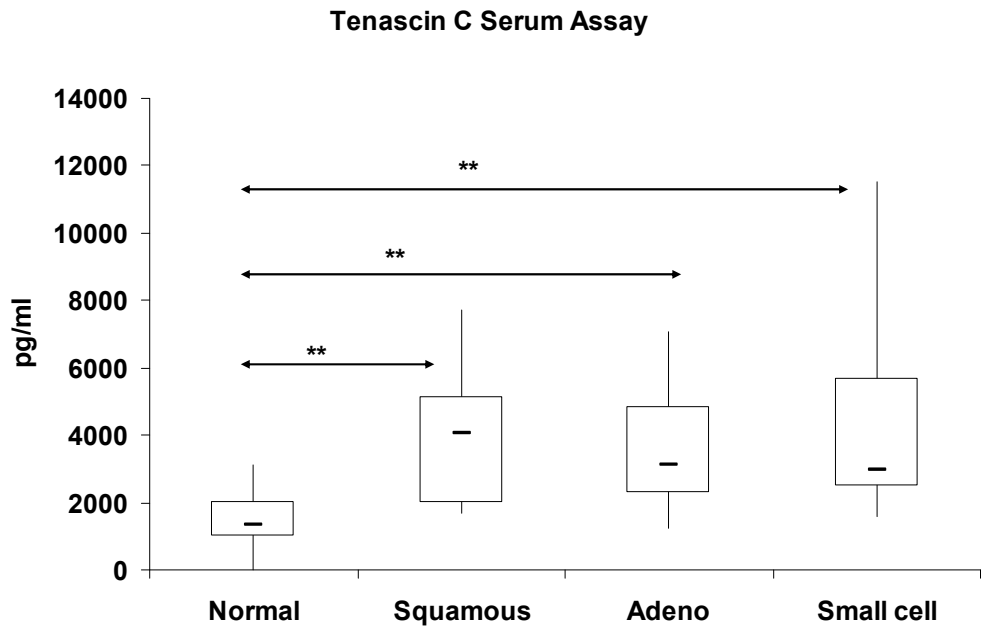
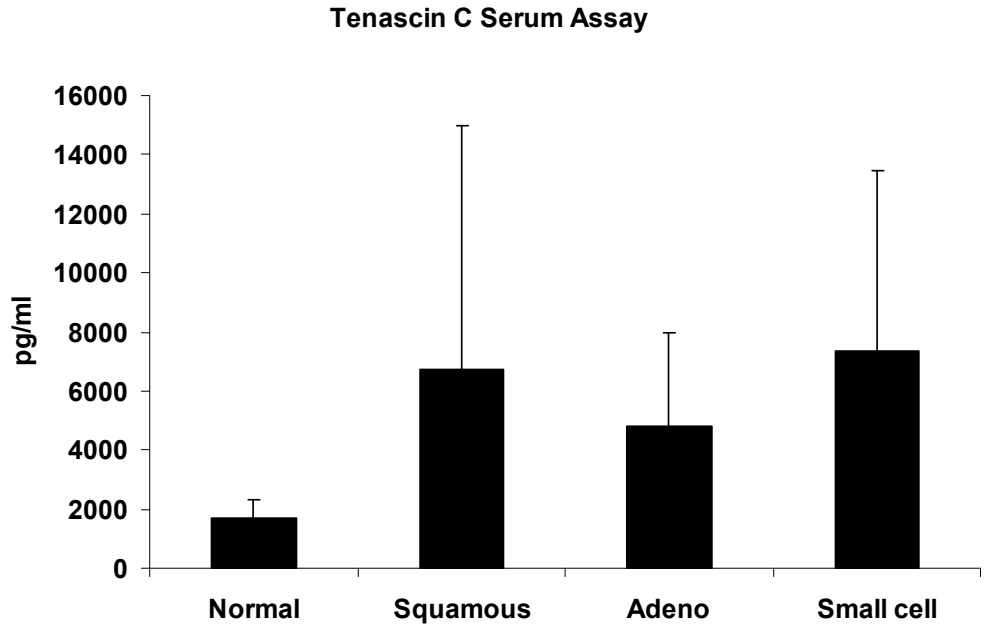


Figure 6.5.9: Tenascin C Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma. T-test score; Normal v SCC (*p* value=9E-06), Normal v AD (*p* value=2E-05), Normal v SCLC (*p* value=0.0002).



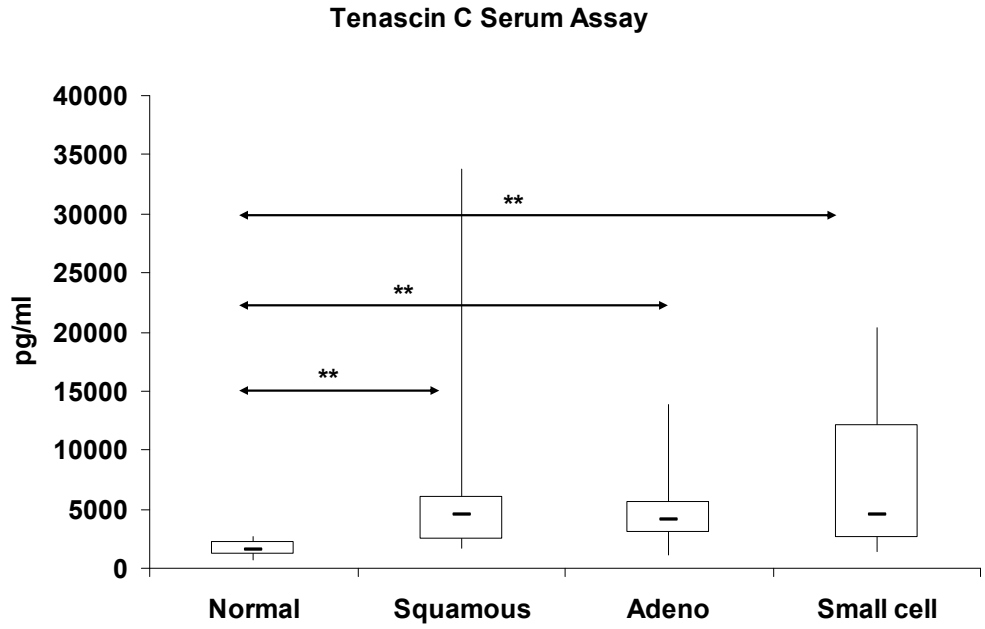
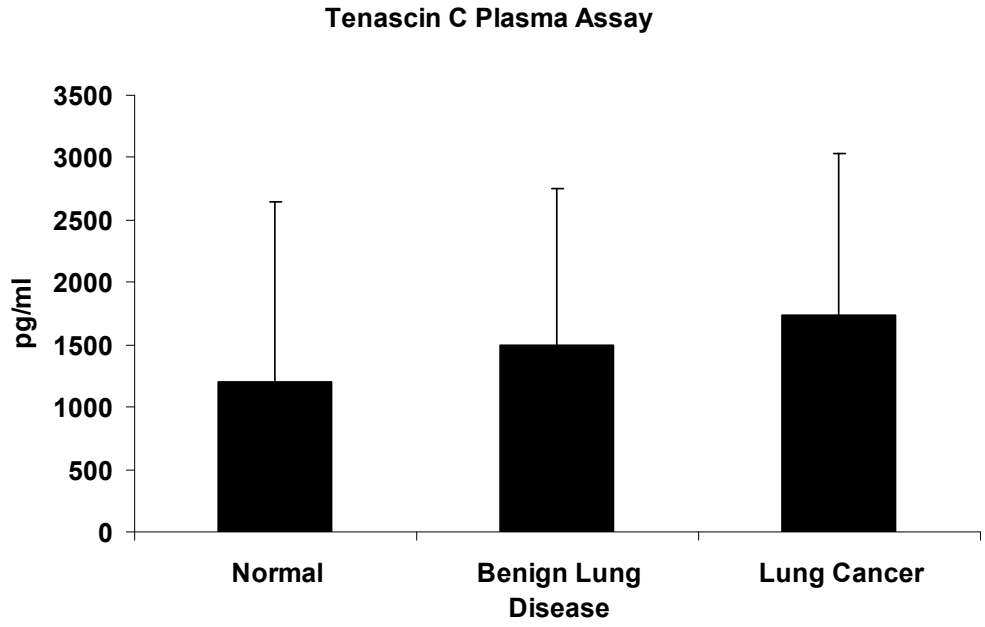


Figure 6.5.10: Tenascin C Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 14 squamous, 21 adenocarcinoma, and 16 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.01), Normal v AD (*p* value=0.0001), Normal v SCLC (*p* value=0.0002).



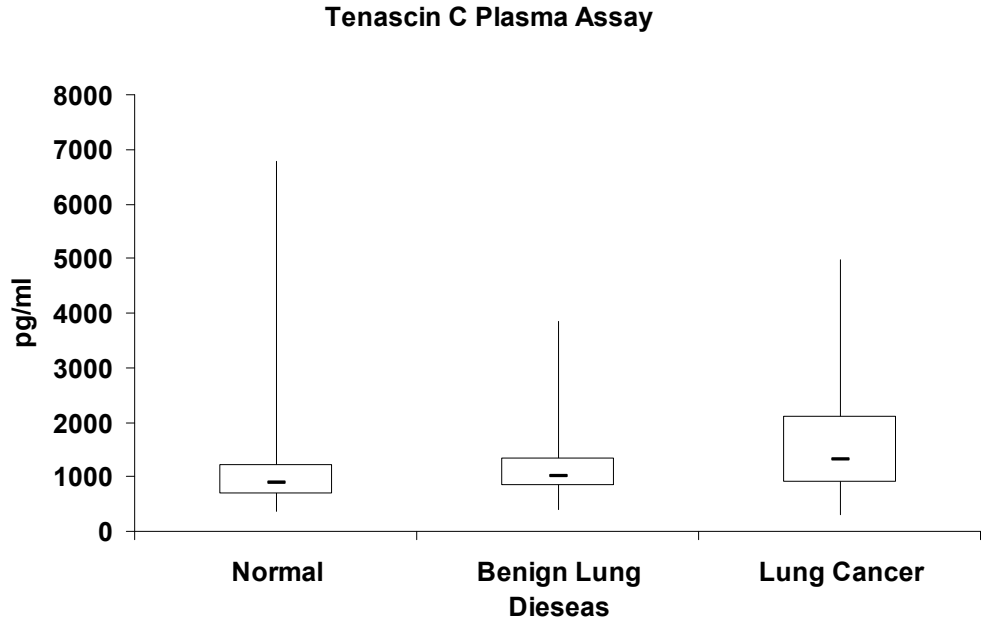


Figure 6.5.11: Tenascin C Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to benign lung disease and early-stage lung cancer. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient plasma samples used for this experiment included 18 normal, 11 benign lung disease, and 35 lung cancer.

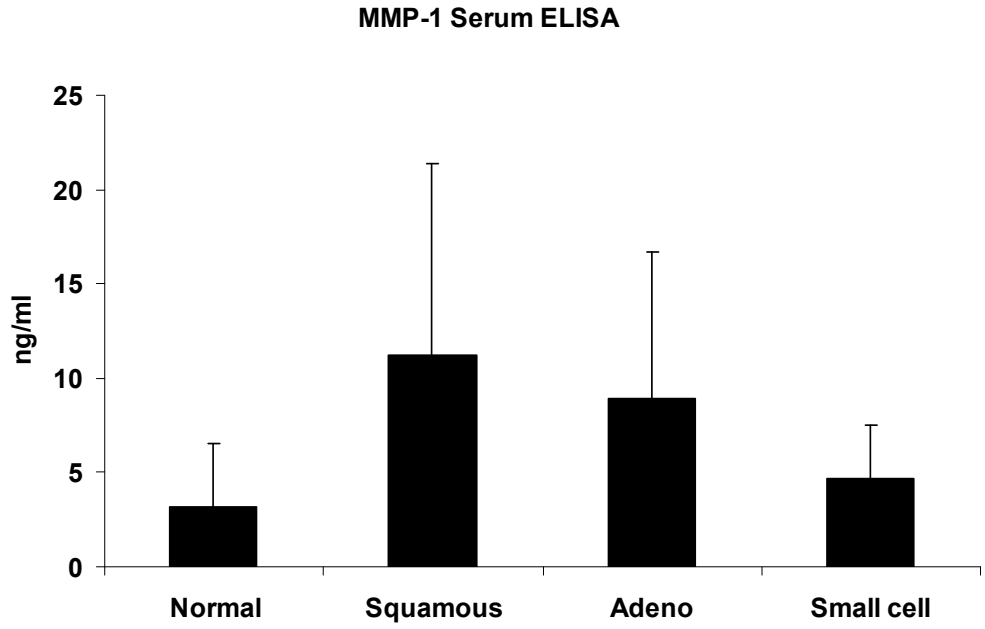
6.6 Validation of targets from Luminex discovery phase: MMP Panel

From the initial discovery phase of the MMP panel, which included MMPs -1, -3, -8, -9, -10, and -13, three were brought forward for validation in larger sample sets. The levels of MMP-1, MMP-8, and MMP-9 showed the greatest increase in cancer serum compared to normal control serum. These three assays also showed significant statistical scores for all of the comparisons of normal versus lung cancer type. The two MMP's -1 and -9 were most extensively studied, which included experiments in serum and plasma, and also in autoimmune serum, benign lung disease serum, and earlier stage lung cancer serum. MMP-8 validation was not performed on the Luminex system but on a commercial single-plex ELISA.

MMPs -2, -7, -10, were examined in larger sample sets due to the availability of Luminex bead sets at the time.

MMPs degrade all protein components of tissue extracellular matrices and basement membranes and are known to have roles in the late stage of tumour progression, invasion, and metastasis. They may also be involved in earlier tumourigenesis such as in malignant transformation, angiogenesis, and tumour growth both at the primary and metastatic sites.

MMP-1 and MMP-8 are collagenases, enzymes that break down collagen, the main component of connective tissue. MMP-9 is a gelatinase enzyme that breaks down gelatine, a mixture of proteins and peptides produced by the hydrolysis of collagen. MMP-1 has been suggested as possible marker for hematogenous metastasis of colorectal cancer, MMP-8 as a marker for improved prognosis in colorectal cancer, and MMP-9 as a marker for Dukes' B colorectal cancer.



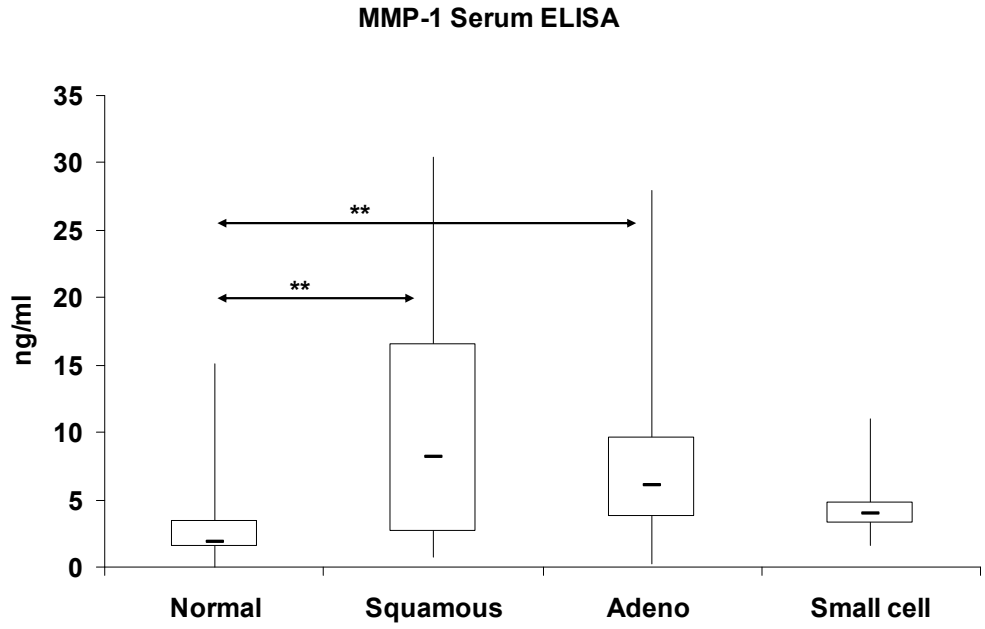


Figure 6.6.1: MMP-1 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.002), Normal v AD (*p* value=0.005).

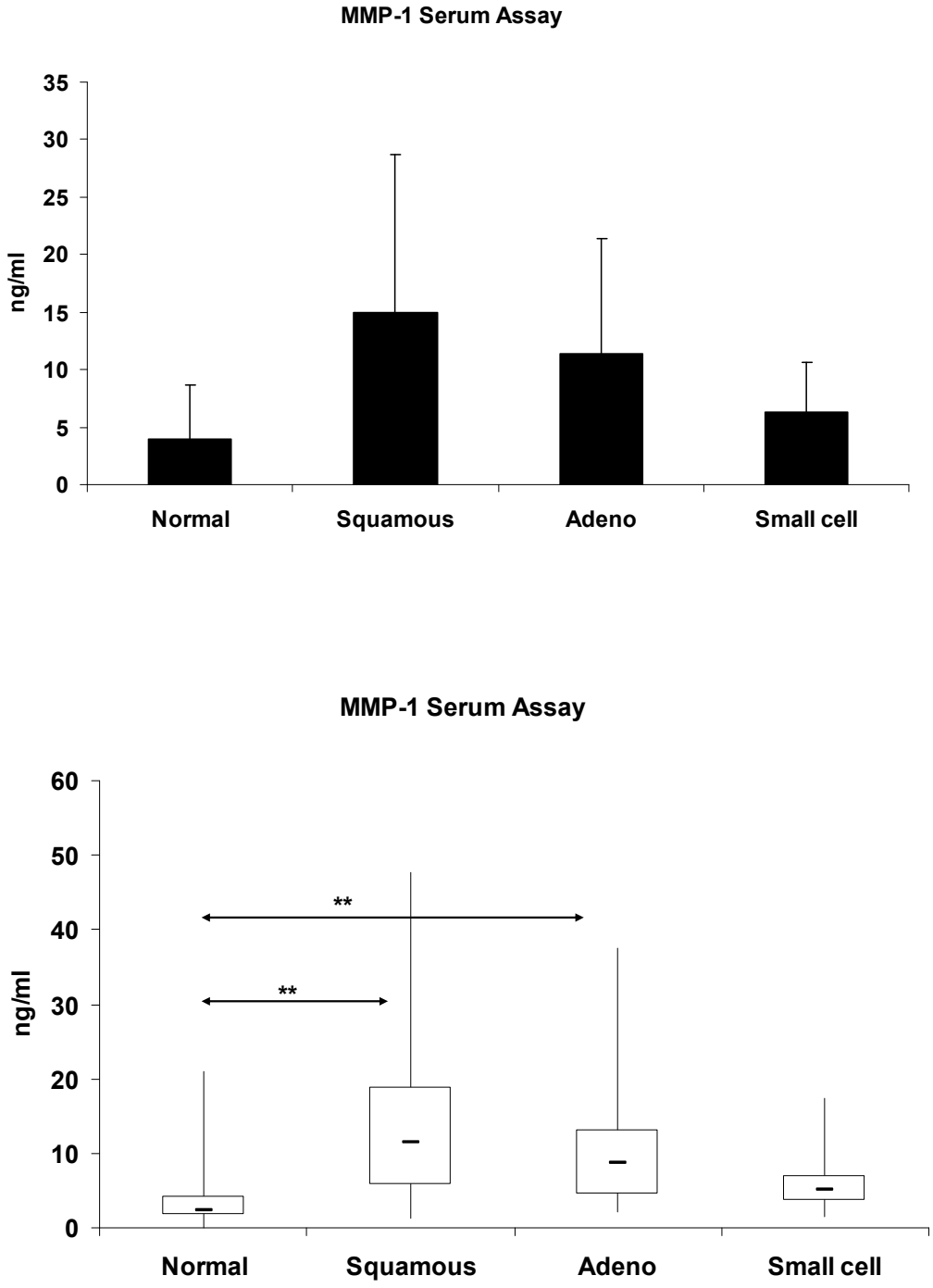
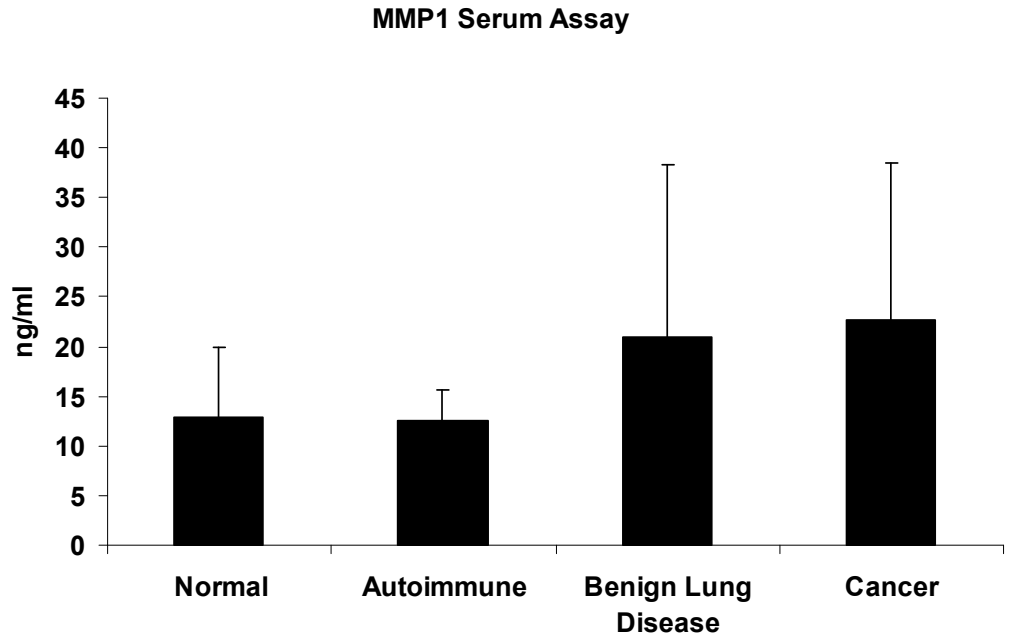


Figure 6.6.2: MMP-1 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 14 squamous, 21 adenocarcinoma, and 16 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.001), Normal v AD (*p* value=0.003).



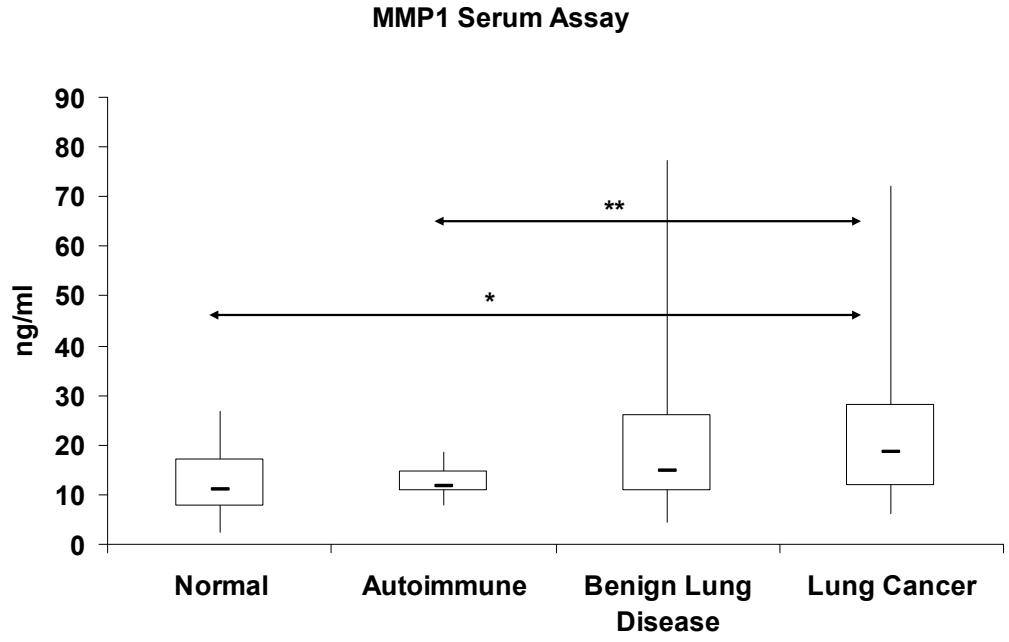
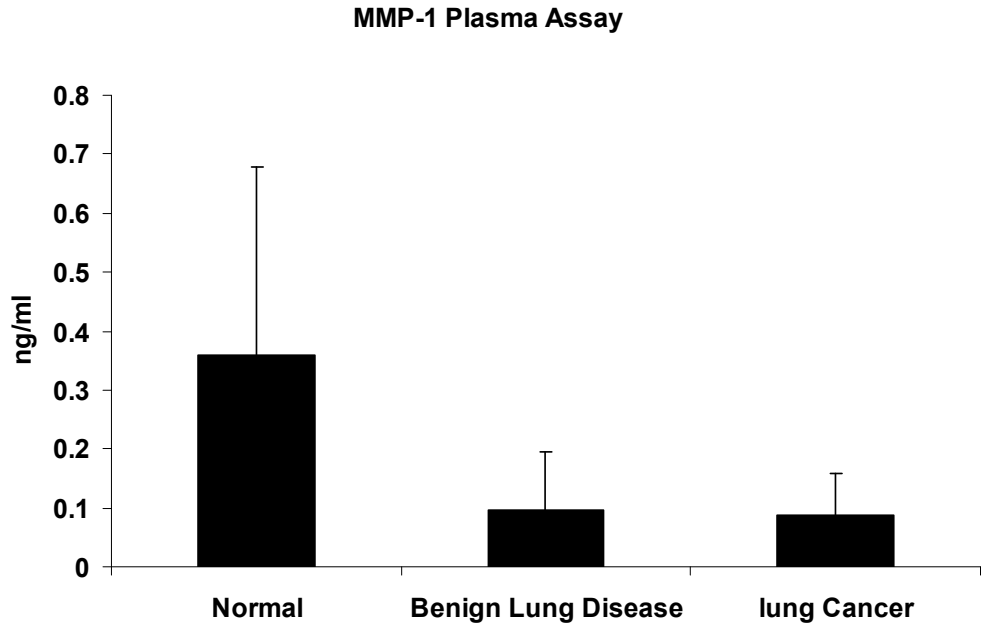


Figure 6.6.3: MMP-1 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each early-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 16 normal, 16 autoimmune, 22 benign lung disease, and 26 lung cancer. T-test score; Normal v Lung Cancer (*p* value=0.03), Autoimmune v Lung Cancer (*p* value=0.01).



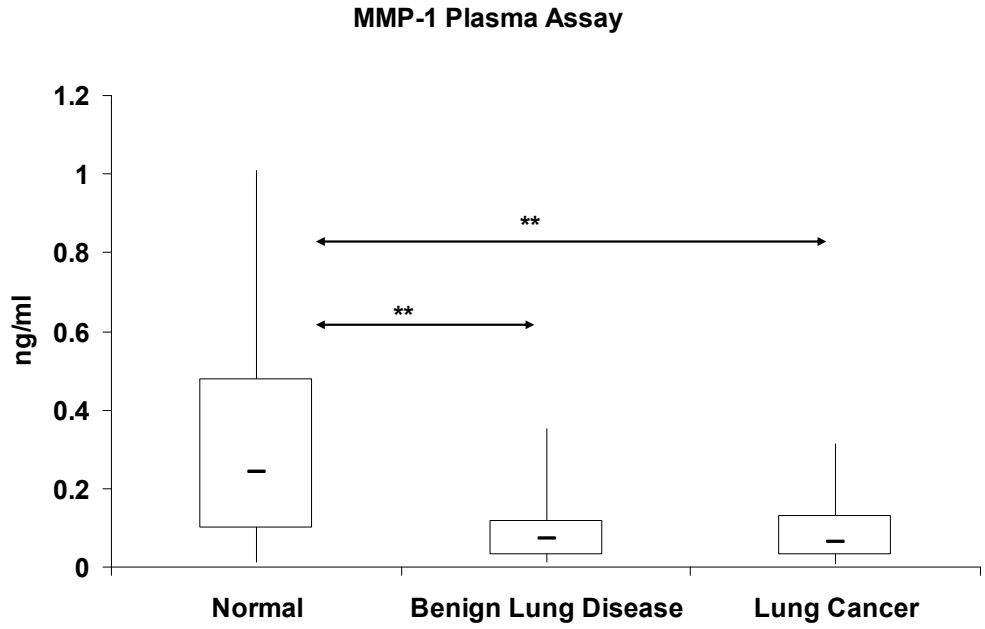
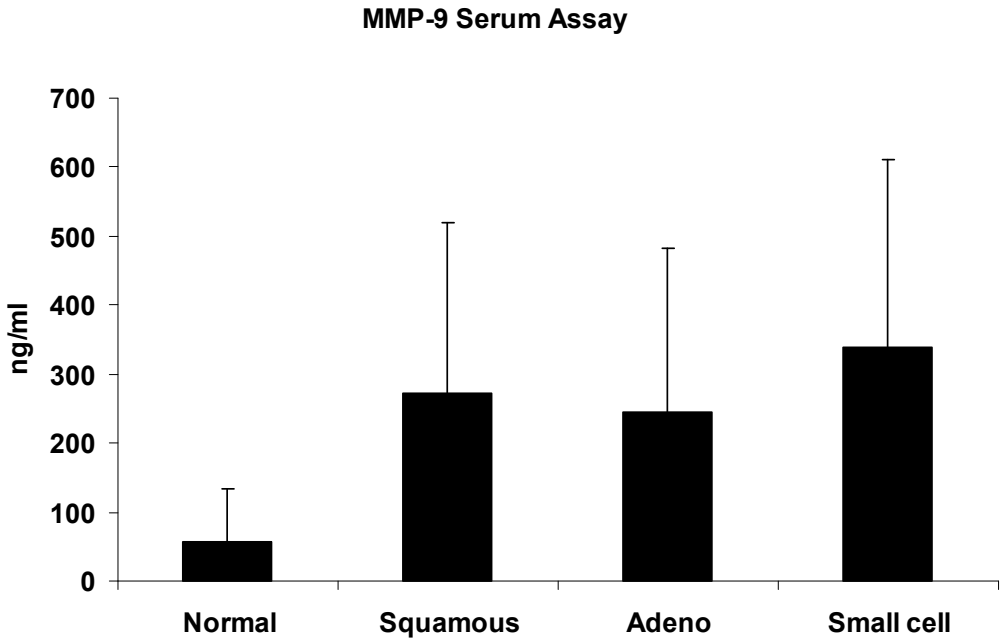


Figure 6.6.4: MMP-1 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to benign lung disease and early-stage lung cancer. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient plasma samples used for this experiment included 18 normal, 11 benign lung disease, and 35 lung cancer. T-test score; Normal v BLD (*p* value=0.01), Normal v Lung Cancer (*p* value=9.7E-06).



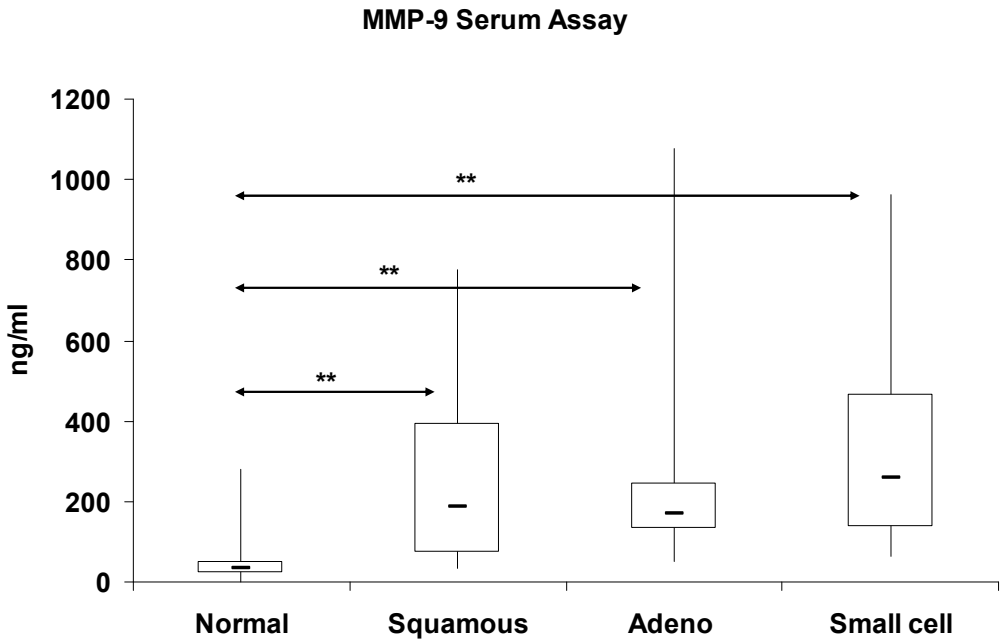
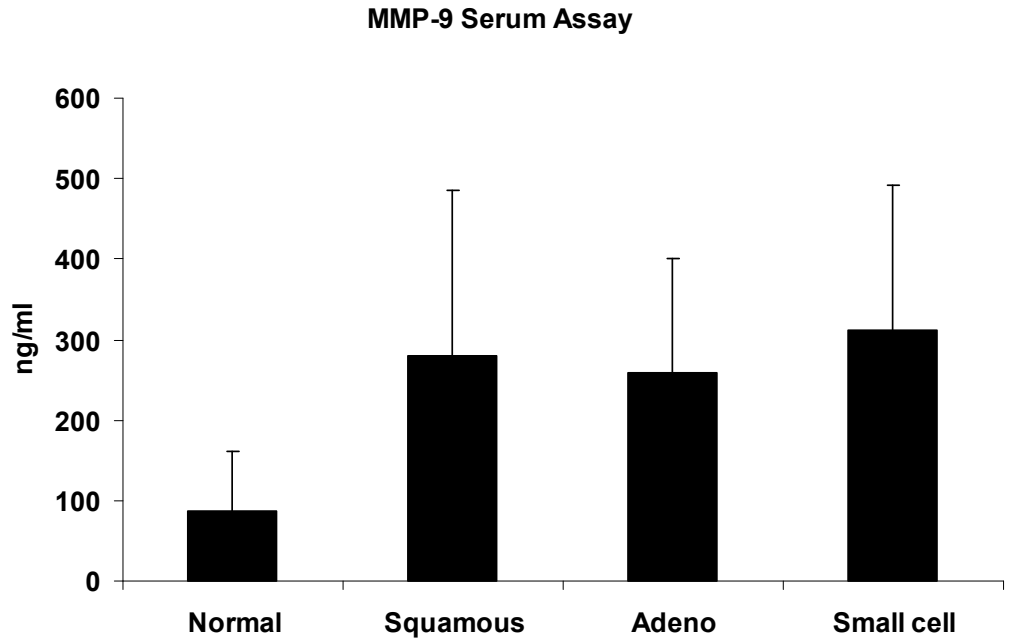


Figure 6.6.5: MMP-9 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used included 20 normal, 16 squamous, 21 adenocarcinoma, and 18 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.0007), Normal v AD (*p* value=0.001), Normal v SCLC (*p* value=5.5E-05).



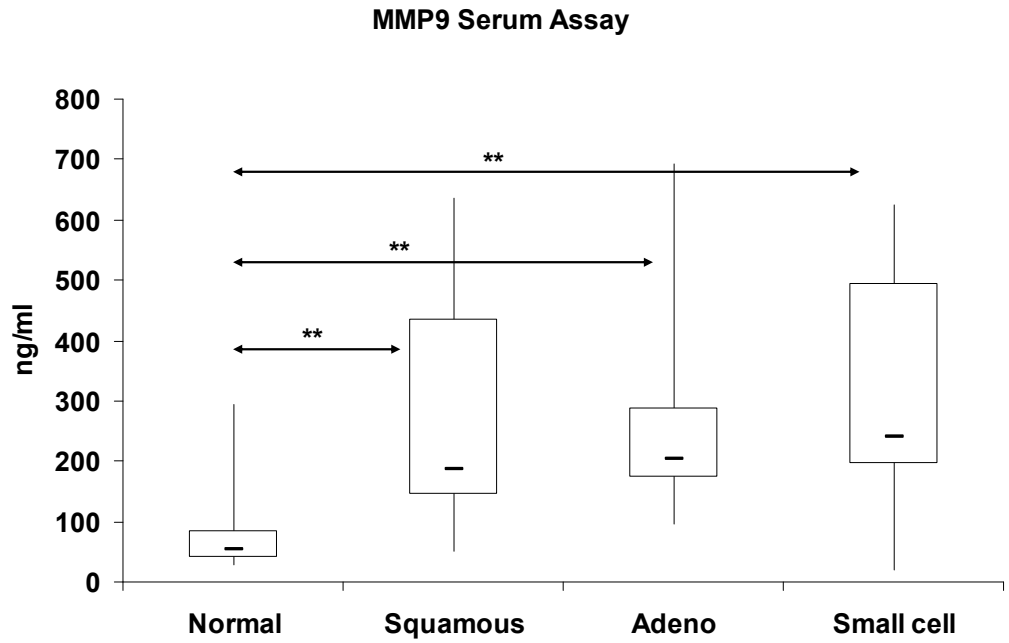


Figure 6.6.6: MMP-9 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 20 normal, 14 squamous, 21 adenocarcinoma, and 16 small cell carcinoma. T-test score; Normal v SCC (*p* value=0.0008), Normal v AD (*p* value=4E-05), Normal v SCLC (*p* value=2.9E-05).

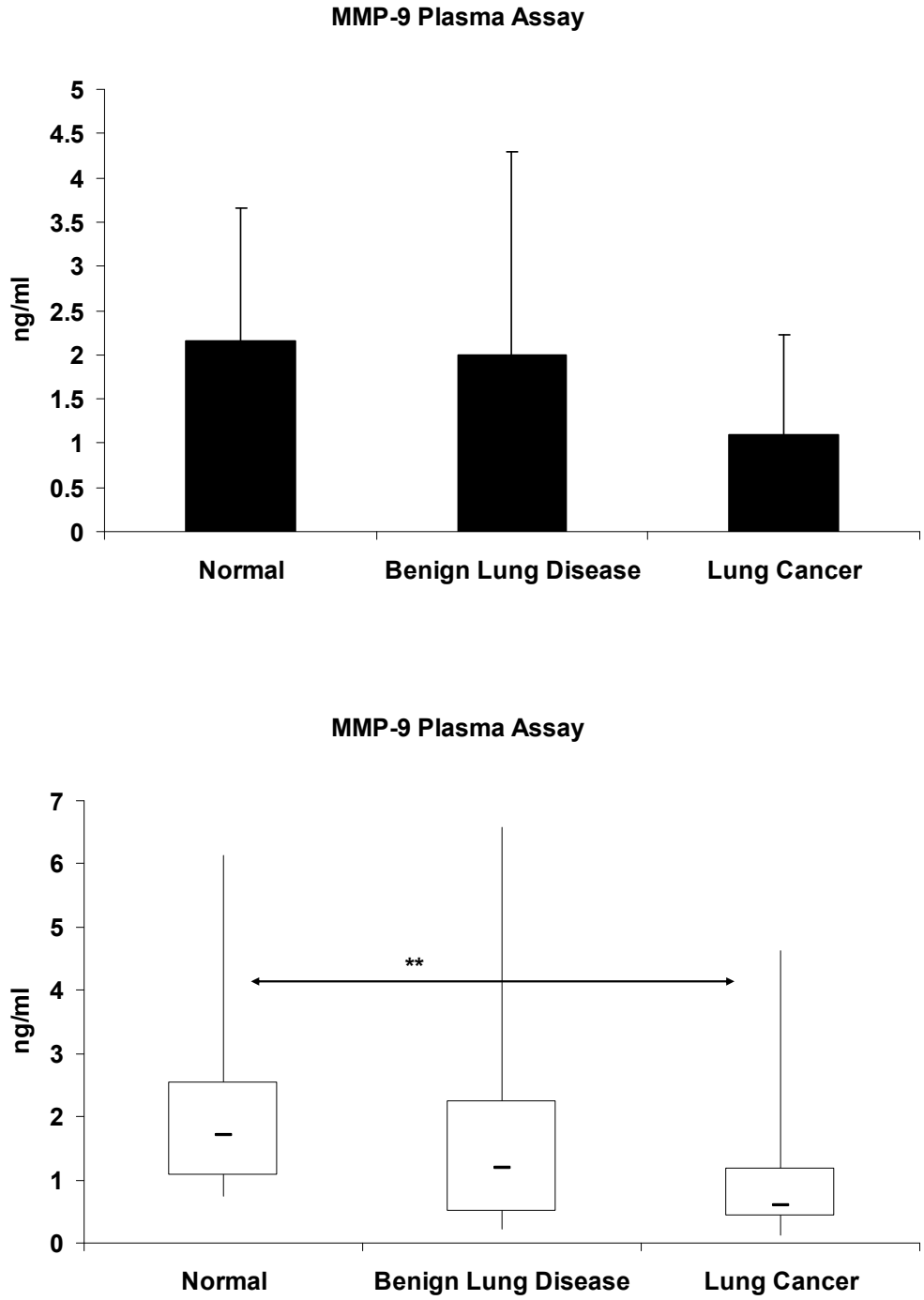
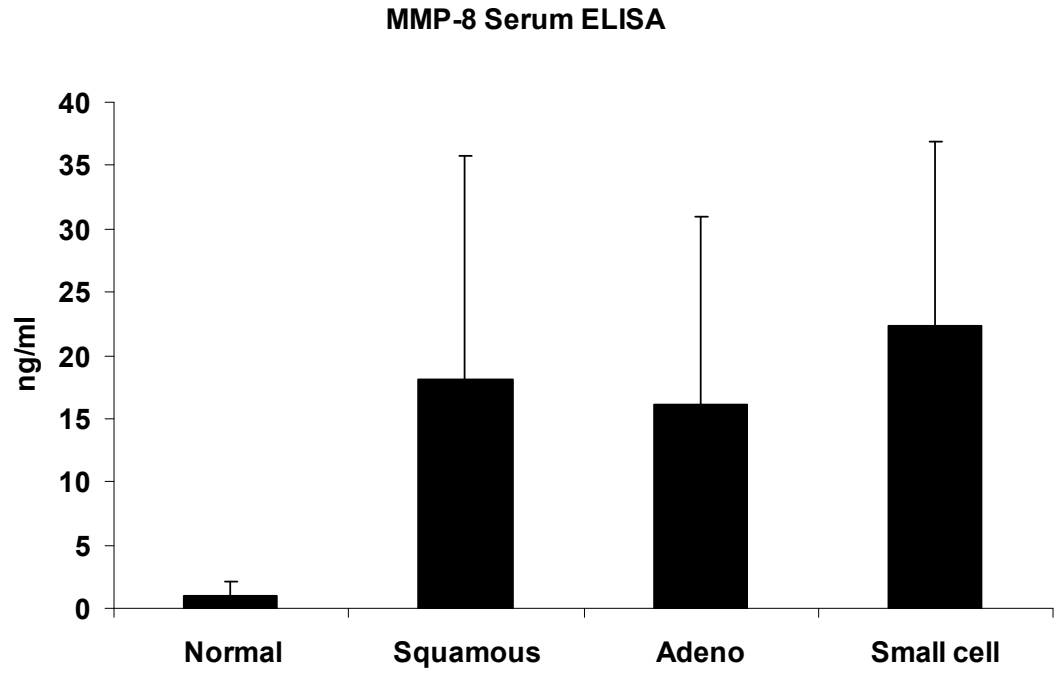


Figure 6.6.7: MMP-9 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to benign lung disease and early-stage lung cancer. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient plasma samples used for this experiment included 18 normal, 11 benign lung disease, and 35 lung cancer. T-test score; Normal v Lung Cancer (*p* value=0.004).



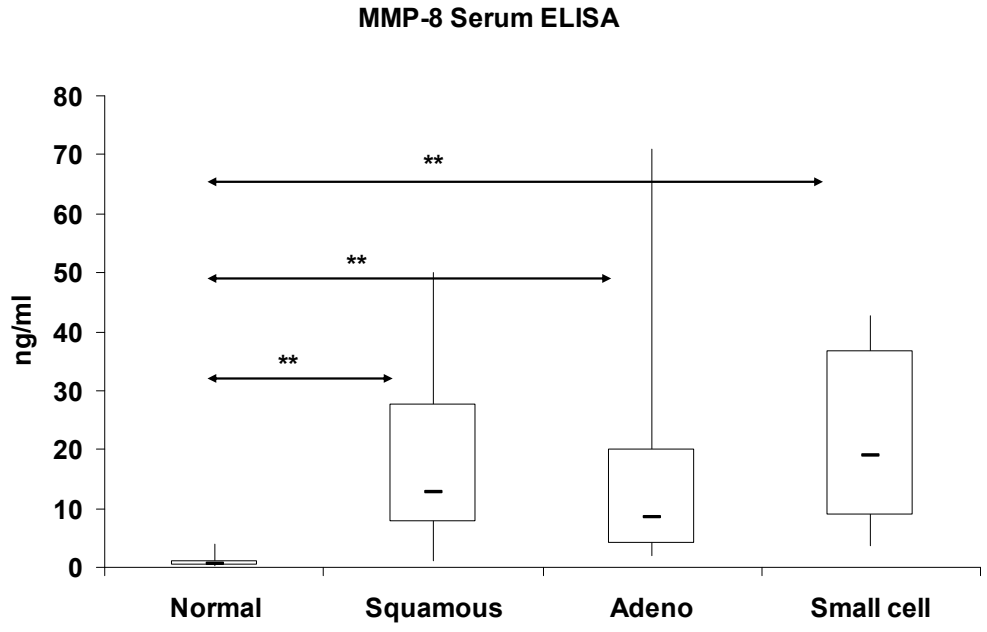


Figure 6.6.8: MMP-8 ELISA: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 22 normal, 16 squamous, 24 adenocarcinoma, and 18 small cell carcinoma. T-test score; Normal v SCC (*p* value=4E-06), normal v AD (*p* value=0.0002), Normal v SCLC (*p* value=3.5E-08).

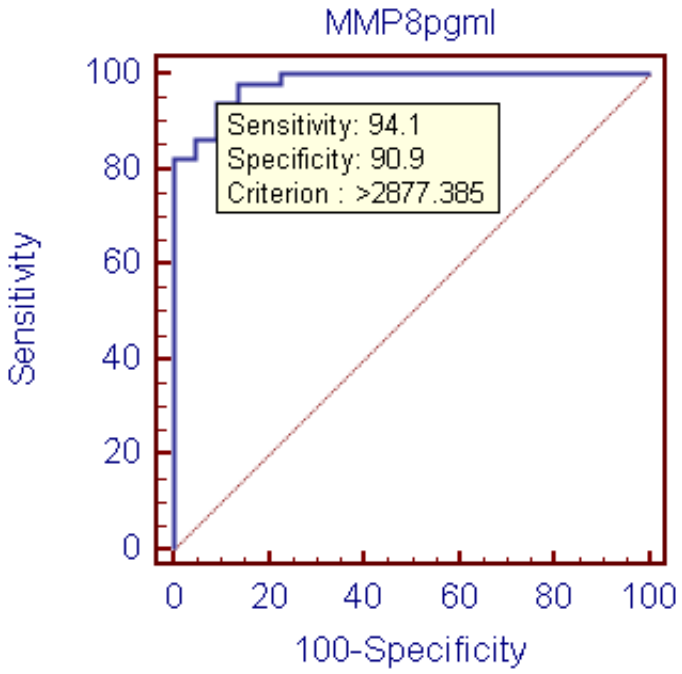
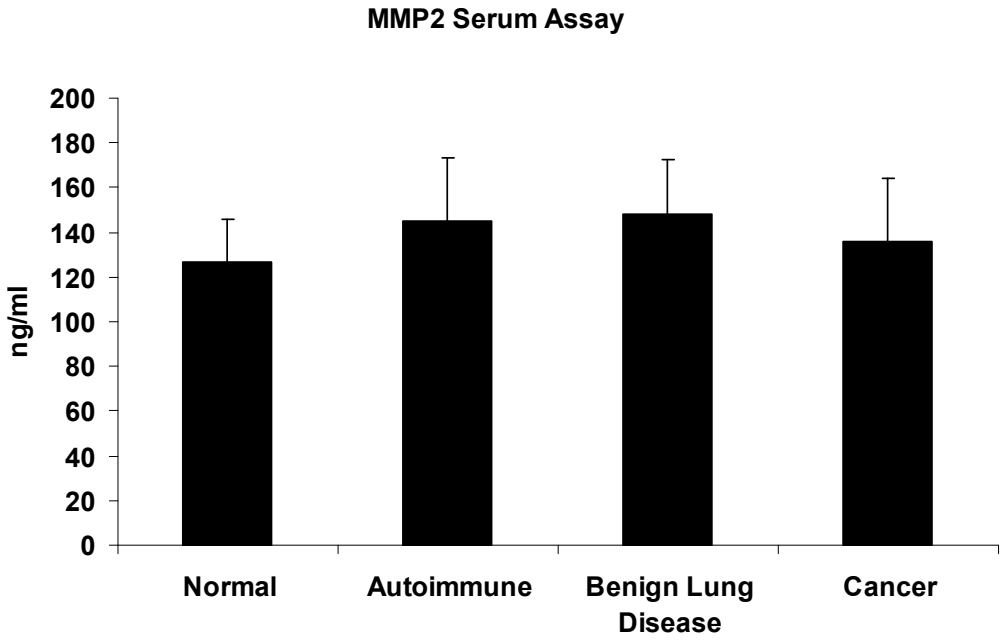


Figure 6.6.9: MMP-8 serum ELISA ROC Curve with sensitivity, specificity, and cut-off value in brackets respectively for statistically significant comparison of normal and lung cancer (94.1%, 90.9%, 2877 pg/ml). Area under the curve: 0.981.



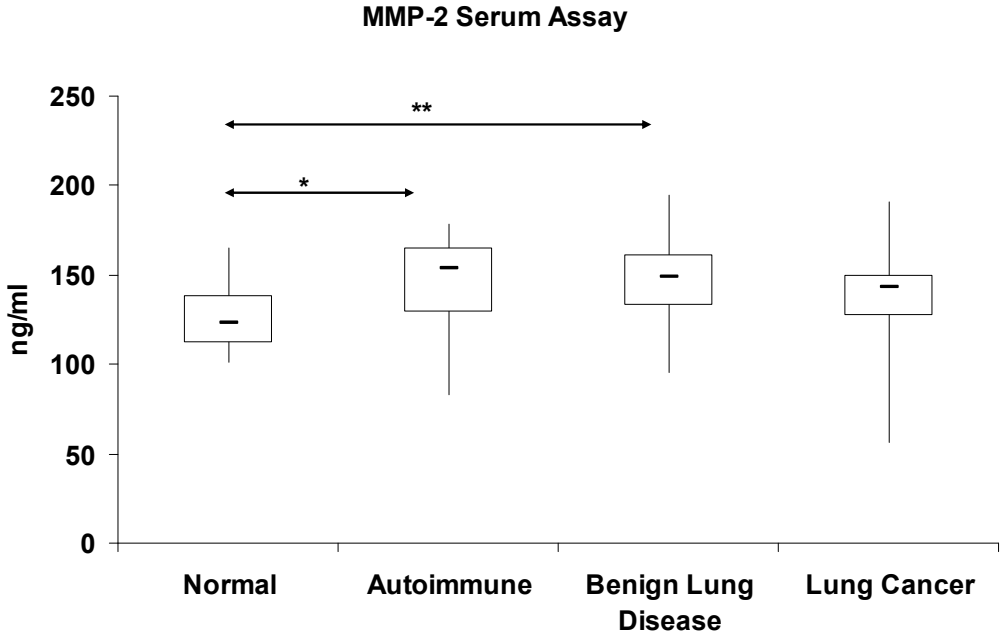
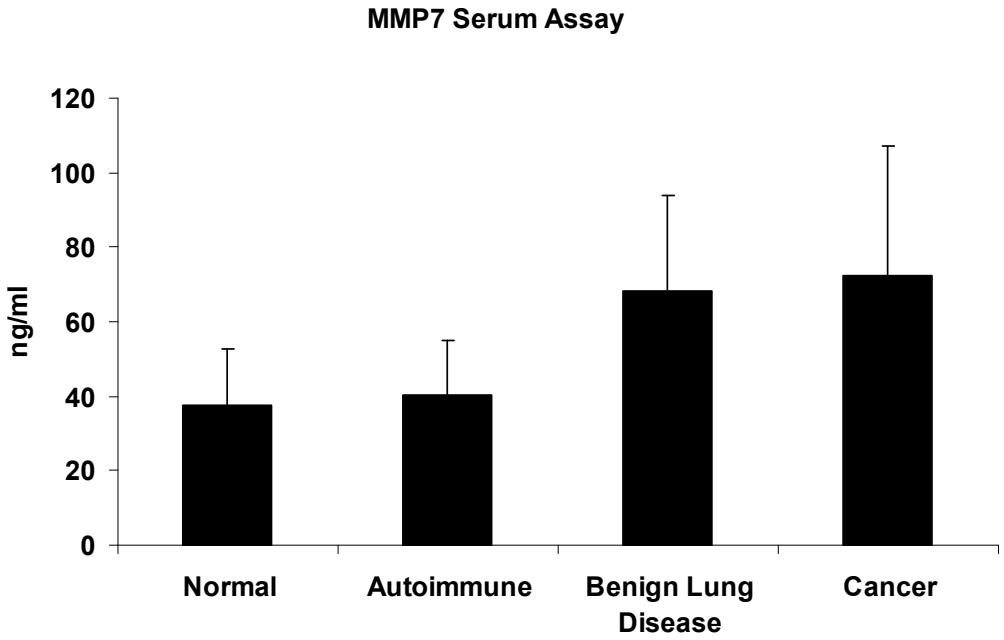


Figure 6.6.10: MMP-2 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each early-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 16 normal, 16 autoimmune, 22 benign lung disease, and 26 lung cancer. T-test score; Normal v Autoimmune (*p* value=0.04), Normal v BLD (*p* value=0.007).



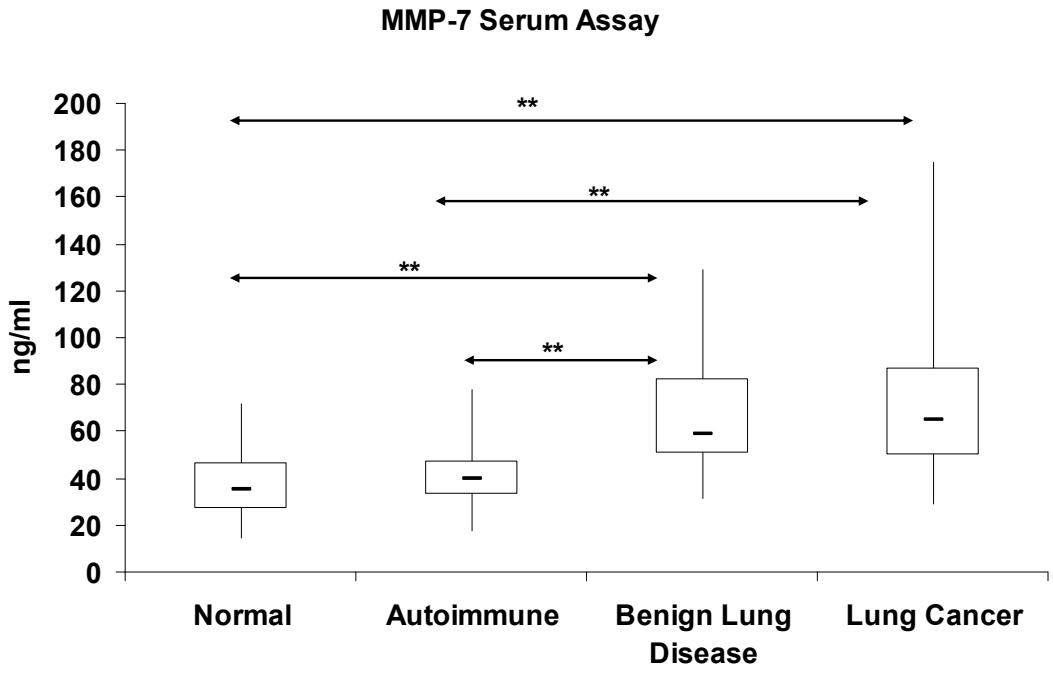
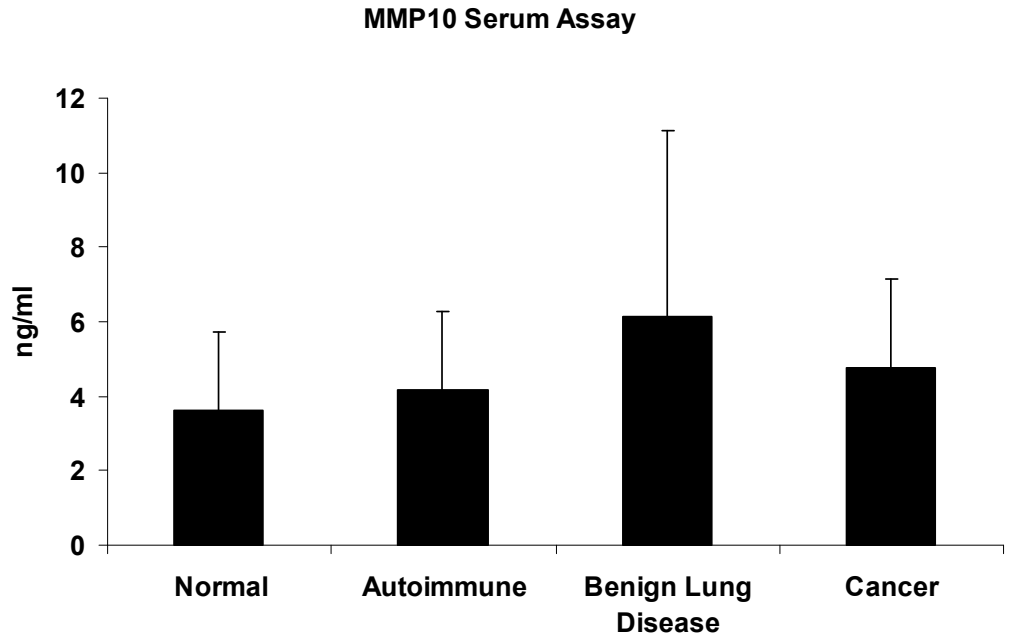


Figure 6.6.11: MMP-7 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each early-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 16 normal, 16 autoimmune, 22 benign lung disease, and 26 lung cancer. T-test score; Normal v BLD (*p* value=0.0001), Normal v Lung Cancer (*p* value=0.0005), Autoimmune v Lung Cancer (*p* value=0.001), Autoimmune v BLD (*p* value=0.0005).



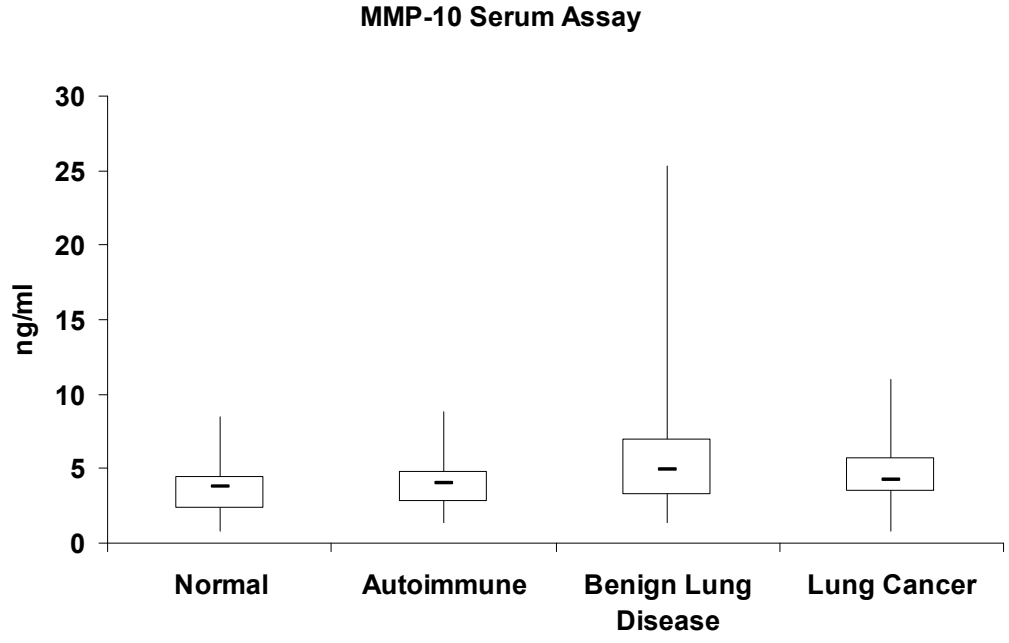


Figure 6.6.12: MMP-10 Luminex Assay: Bar chart and Box and Whisker Plot showing normal compared to each early-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples used for this experiment included 16 normal, 16 autoimmune, 22 benign lung disease, and 26 lung cancer.

6.7 Logistic Regression on Matching Clinical Samples

A.

Sample size	76
Cases with Y=0	20 (26.32%)
Cases with Y=1	56 (73.68%)

Overall Model Fit

Null model -2 Log Likelihood	87.603
Full model -2 Log Likelihood	33.318
Chi-square	54.285
DF	3
Significance level	P < 0.0001

Coefficients and Standard Errors

Variable	Coefficient	Std. Error	P
CA125serum	0.10644	0.043515	0.0144
CA199Serum	0.047411	0.022621	0.0361
CEAserum	0.34196	0.23358	0.1432
Constant	-6.2741		

Odds Ratios and 95% Confidence Intervals

Variable	Odds ratio	95% CI
CA125serum	1.1123	1.0214 to 1.2113
CA199Serum	1.0486	1.0031 to 1.0961
CEAserum	1.4077	0.8906 to 2.2251

Classification table (cut-off value p=0.5)

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	16	4	80.00 %
Y = 1	6	50	89.29 %
Percent of cases correctly classified			86.84 %

ROC curve analysis

Area under the ROC curve (AUC)	0.962
Standard Error	0.0187
95% Confidence interval	0.890 to 0.992

B.

Sample size	64
Cases with Y=0	20 (31.25%)
Cases with Y=1	44 (68.75%)

Overall Model Fit

Null model -2 Log Likelihood	79.499
Full model -2 Log Likelihood	13.716
Chi-square	65.783
DF	3
Significance level	P < 0.0001

Coefficients and Standard Errors

Variable	Coefficient	Std. Error	P
MMP1	0.36573	0.19918	0.0663
MMP9	0.011077	0.0056479	0.0499
TenascinC	0.0030043	0.0014447	0.0376
Constant	-10.1209		

Odds Ratios and 95% Confidence Intervals

Variable	Odds ratio	95% CI
MMP1	1.4416	0.9756 to 2.1300
MMP9	1.0111	1.0000 to 1.0224
TenascinC	1.0030	1.0002 to 1.0059

Classification table (cut-off value p=0.5)

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	19	1	95.00 %
Y = 1	1	43	97.73 %
Percent of cases correctly classified			96.87 %

ROC curve analysis

Area under the ROC curve (AUC)	0.992
Standard Error	0.00866
95% Confidence interval	0.929 to 1.000

Figure 6.7.1: Logistic regression performed on matching serum patient samples (A) combining Luminex assay results from CA-125, CA 19-9, and CEA with 20 matching normal and 56 matching cancer samples, indicating 86.84% of cases would be correctly classified, and (B) combining Luminex assay results from MMP-1, MMP-9, and tenascin-C with 20 matching normal and 44 matching cancer samples, indicating 96.87% of cases would be correctly classified.

A.

Variable	Coefficient	Std. Error	P
CA125	0.047771	0.023178	0.0393
CA199	0.074657	0.046251	0.1065
CEA	0.23662	0.16822	0.1595
Constant	-3.4325		

Odds Ratios and 95% Confidence Intervals

Variable	Odds ratio	95% CI
CA125	1.0489	1.0023 to 1.0977
CA199	1.0775	0.9841 to 1.1798
CEA	1.2670	0.9111 to 1.7618

Hosmer & Lemeshow test

Chi-square	7.8661
DF	9
Significance level	P = 0.5477

Contingency table for Hosmer & Lemeshow test [Hide]

Group	Y=0		Y=1		Total
	Observed	Expected	Observed	Expected	
1	5	4.275	0	0.725	5
2	2	3.716	3	1.284	5
3	4	3.234	1	1.766	5
4	3	2.615	2	2.385	5
5	3	2.077	2	2.923	5
6	0	1.400	5	3.600	5
7	1	0.522	4	4.478	5
8	0	0.150	5	4.850	5
9	0	0.0118	5	4.988	5
10	0	0.000101	5	5.000	5
11	0	1.9983E-015	2	2.000	2

Classification table (cut-off value p=0.5)

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	13	5	72.22 %
Y = 1	5	29	85.29 %
Percent of cases correctly classified			80.77 %

ROC curve analysis

Area under the ROC curve (AUC)	0.887
Standard Error	0.0450
95% Confidence interval	0.769 to 0.958

B.

Variable	Coefficient	Std. Error	P
MMP1	-0.014139	0.0056585	0.0125
CA125	0.071229	0.031696	0.0246
TenascinC	0.0011973	0.00063625	0.0599
Constant	-1.3213		

Odds Ratios and 95% Confidence Intervals

Variable	Odds ratio	95% CI
MMP1	0.9860	0.9751 to 0.9970
CA125	1.0738	1.0091 to 1.1427
TenascinC	1.0012	1.0000 to 1.0024

Hosmer & Lemeshow test

Chi-square	2.8901
DF	9
Significance level	P = 0.9685

Contingency table for Hosmer & Lemeshow test [Hide]

Group	Y=0		Y=1		Total
	Observed	Expected	Observed	Expected	
1	5	4.985	0	0.0147	5
2	5	4.737	0	0.263	5
3	3	3.404	2	1.596	5
4	1	2.029	4	2.971	5
5	2	1.435	3	3.565	5
6	1	0.805	4	4.195	5
7	1	0.398	4	4.602	5
8	0	0.150	5	4.850	5
9	0	0.0534	5	4.947	5
10	0	0.00354	5	4.996	5
11	0	0.0000224	3	3.000	3

Classification table (cut-off value p=0.5)

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	13	5	72.22 %
Y = 1	1	34	97.14 %
Percent of cases correctly classified			88.68 %

ROC curve analysis

Area under the ROC curve (AUC)	0.932
Standard Error	0.0349
95% Confidence interval	0.828 to 0.983

Figure 6.7.2: Logistic regression performed on matching plasma patient samples (A) combining Luminex assay results from CA-125, CA 19-9, and CEA with 18 matching normal and 34 matching cancer samples, indicating 80.77% of cases would be correctly classified, and (B) combining Luminex assay results from MMP-1, CA-125, and tenascin-C with 18 matching normal and 34 matching cancer samples, indicating 88.68% of cases would be correctly classified.

CHAPTER SEVEN

METABOLOMIC SERUM PROFILING FOR LUNG CANCER BIOMARKER DISCOVERY

7.1 Background

Metabolomics is the name given to the global quantitative assessment of endogenous metabolites within a biological system. Metabolites are the intermediates and products of metabolism and usually refer to small molecules, with primary metabolites directly involved in normal growth, development and reproduction. Much like proteins, metabolites present in the serum or plasma can be used as biomarkers for a particular cancer. Metabolomics could address the limitation associated with transcriptomics and proteomics where changes do not always result in altered biochemical phenotypes. The detection of these metabolites could lead to early cancer detection and diagnosis, and also for chemotherapy drug treatment monitoring.

Metabolon are a global biochemical and diagnostics company and provide a service of profiling serum samples. Lung cancer samples for squamous cell carcinoma, adenocarcinoma, and small cell carcinoma, as well as normal controls were sent to Metabolon, who in turn provided an extensive profile of up- and down-regulated metabolites. Eight samples, four male and four female, for each condition were analysed.

In total, information on over three hundred metabolites was provided. From this extensive list the most interesting ones were selected and from this a number were chosen for validation. The initial selection process was based on abundance levels of the metabolite in the lung cancer samples compared to the normal control. Those chosen to be brought forward for validation in larger sample sets were very much restricted by availability of commercial assays and in total four were chosen; 12-HETE, glutamate, aspartate, and phenylalanine.

12 (S)-HETE was examined in serum (including autoimmune and benign lung disease) and plasma. Glutamate and phenylalanine were examined in serum and plasma, and aspartate in serum only.

7.2 Metabolomics Discovery Phase

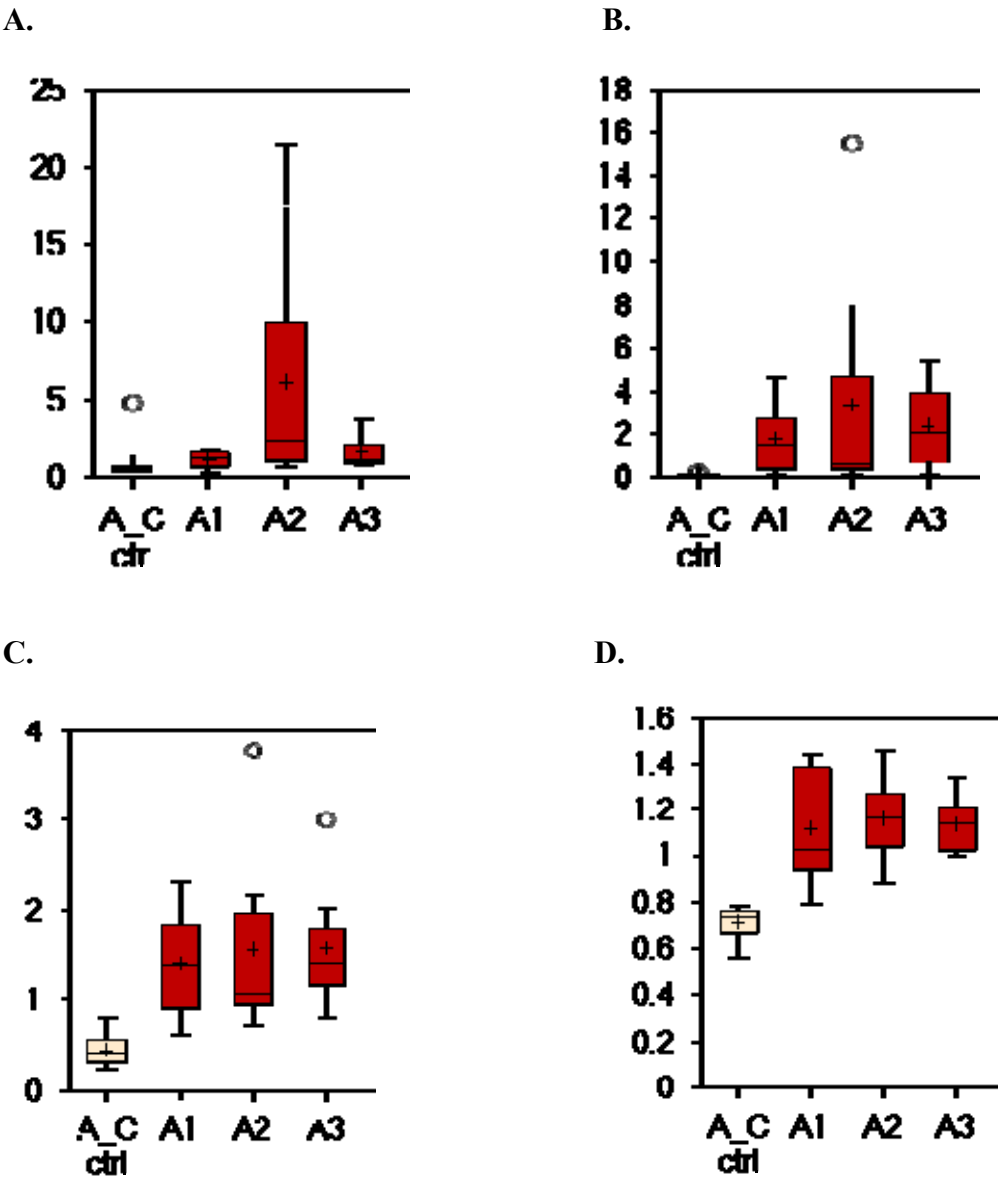


Figure 7.2.1: Box and whisker plots from metabolomic discovery phase showing (A) 3-hydroxybutyrate, (B) 12-HETE, (C) glutamate, and (D) phenylalanine. Serum samples on x-axis show normal (A_C ctrl), adenocarcinoma (A1), squamous cell carcinoma (A2), and small cell carcinoma (A3).

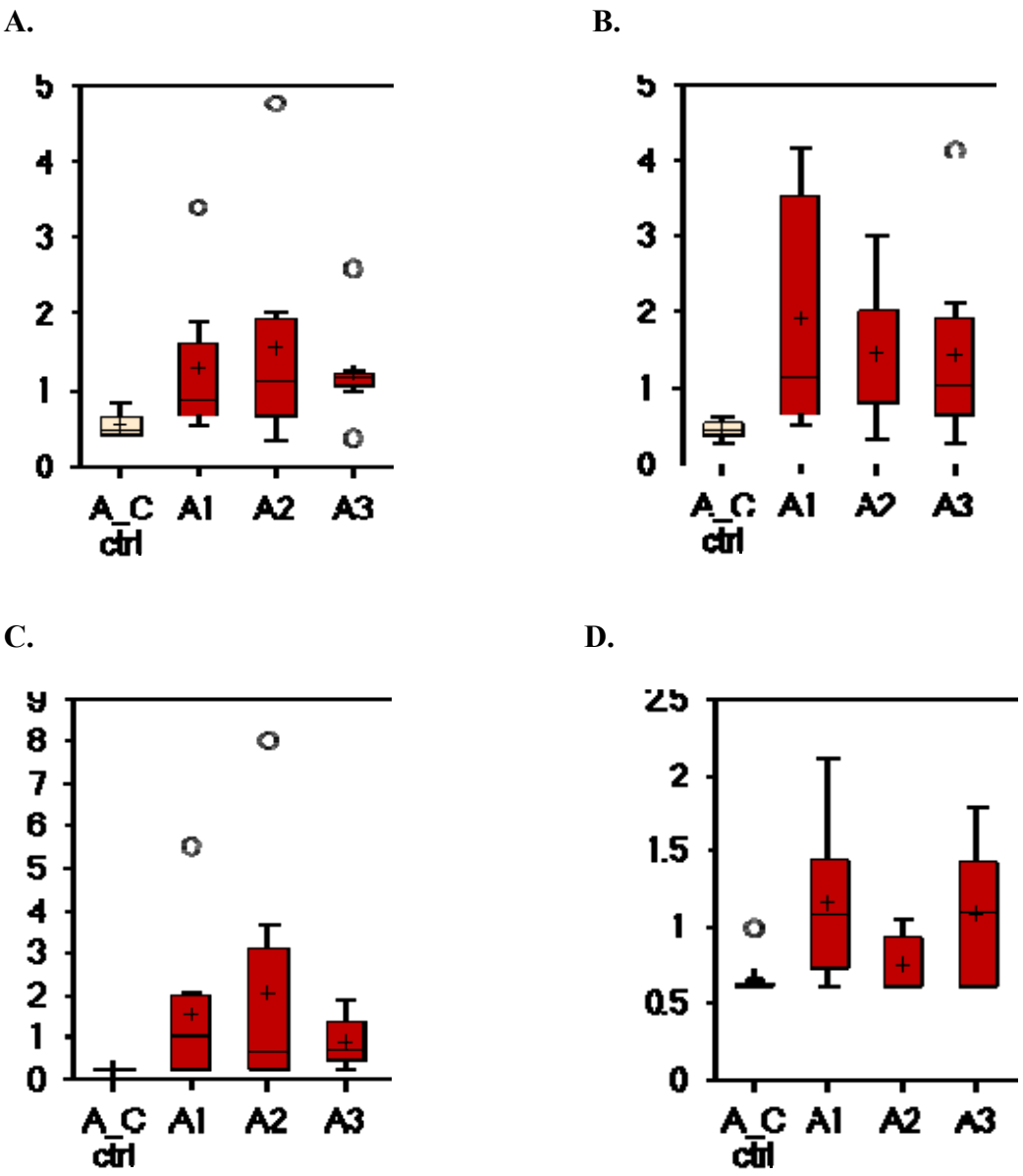


Figure 7.2.2: Box and whisker plots from metabolomic discovery phase showing (A) aspartate, (B) ornithine, (C) glycylvaline, and (D) 1,7-dimethylurate. Serum samples on x-axis show normal (A_C ctrl), adenocarcinoma (A1), squamous cell carcinoma (A2), and small cell carcinoma (A3).

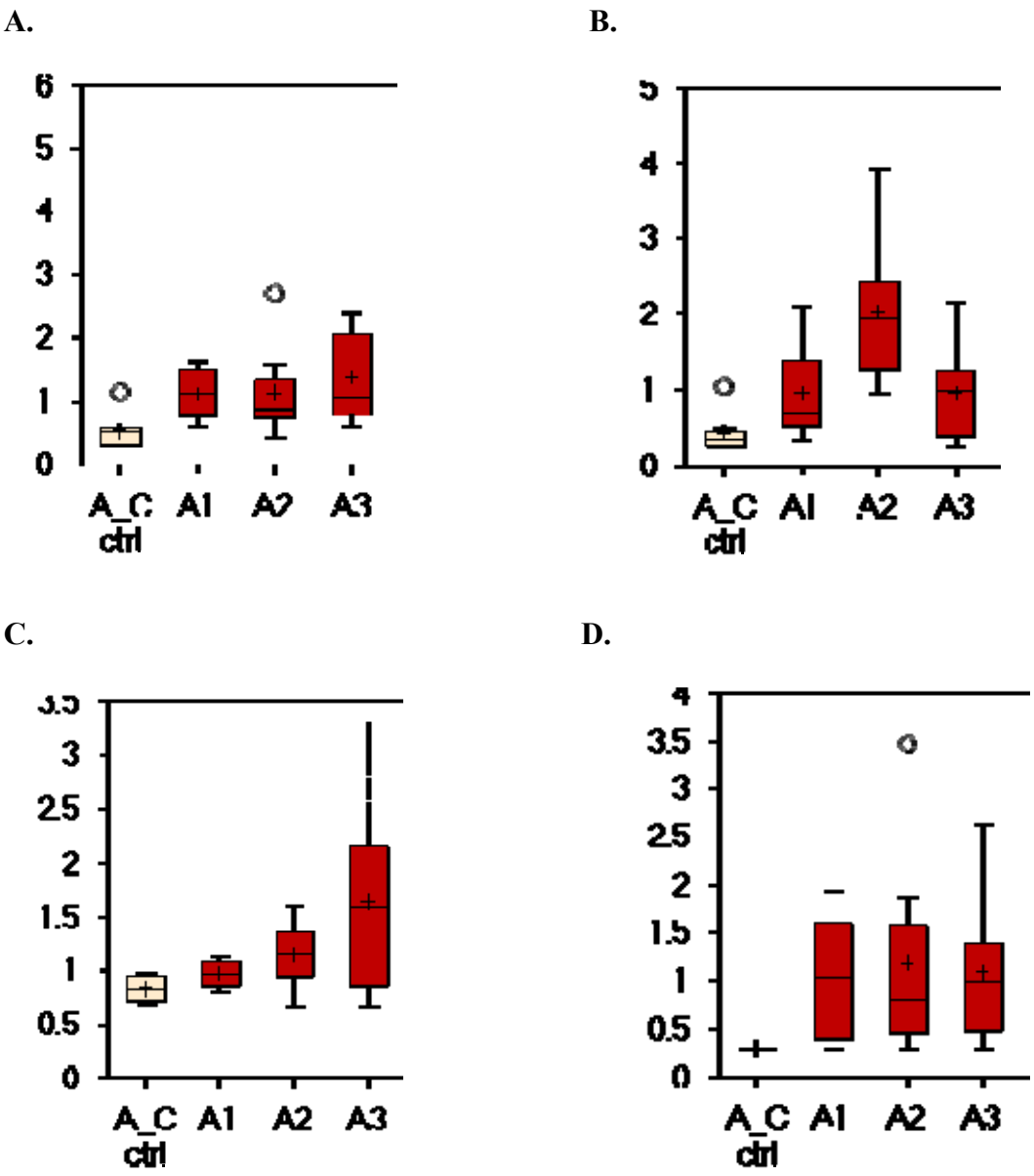


Figure 7.2.3: Box and whisker plots from metabolomic discovery phase showing (A) allantoin, (B) 10-nonadecenoate, (C) erythronate, and (D) phenylalanylleucine. Serum samples on x-axis show normal (A_C ctrl), adenocarcinoma (A1), squamous cell carcinoma (A2), and small cell carcinoma (A3).

7.3 Validation of Targets from Metabolomic Profiling of Serum

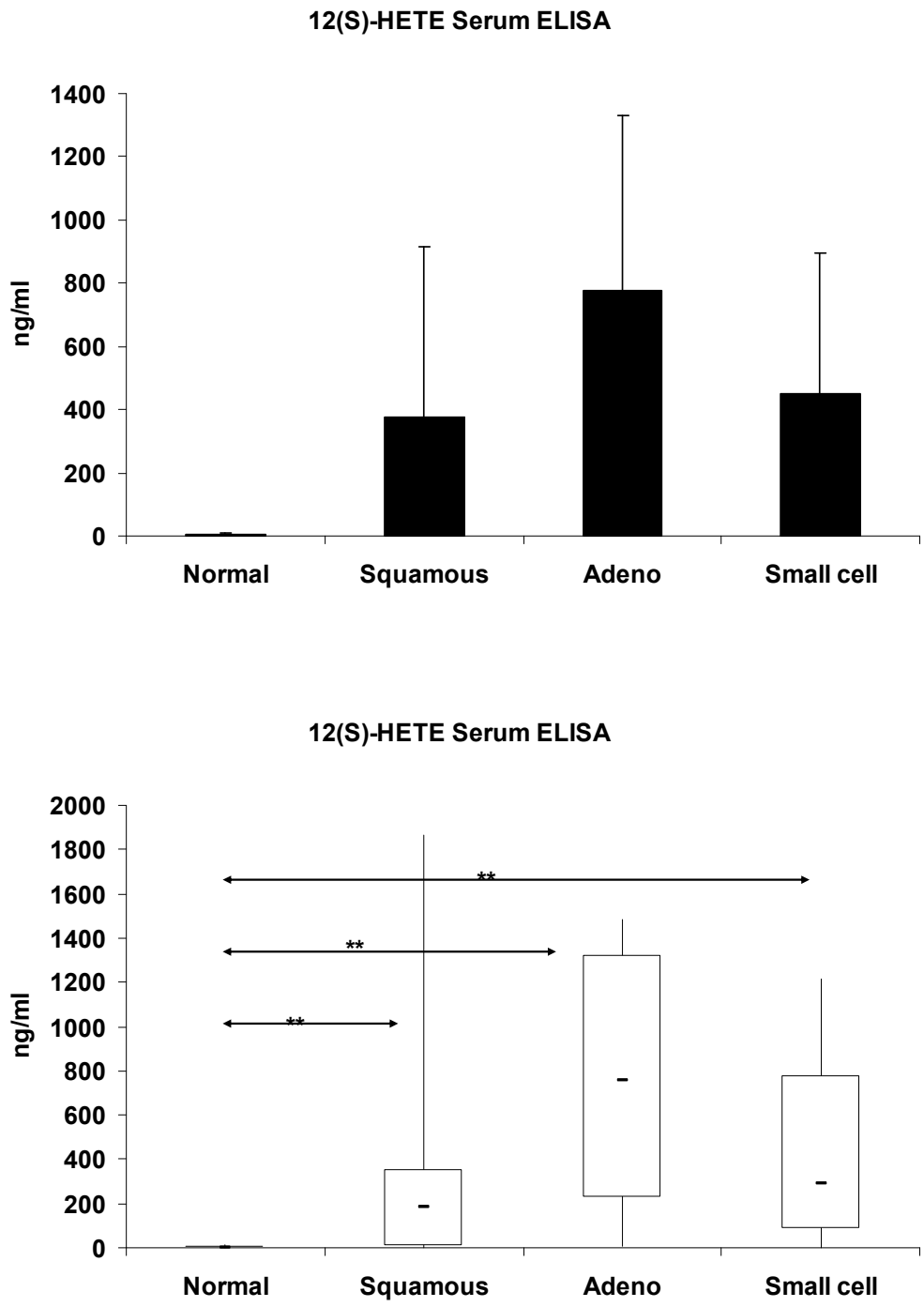


Figure 7.3.1: 12(S)-HETE ELISA: Bar chart and Box and Whisker Plot showing normal compared to each late stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples included 17 normal, 16 squamous, 17 adenocarcinoma, and 17 small cell carcinoma. T-test scores: Normal v SCC (*p* value=0.008), Normal v AD (*p* value=2E-06), Normal v SCLC (*p* value=0.0002).

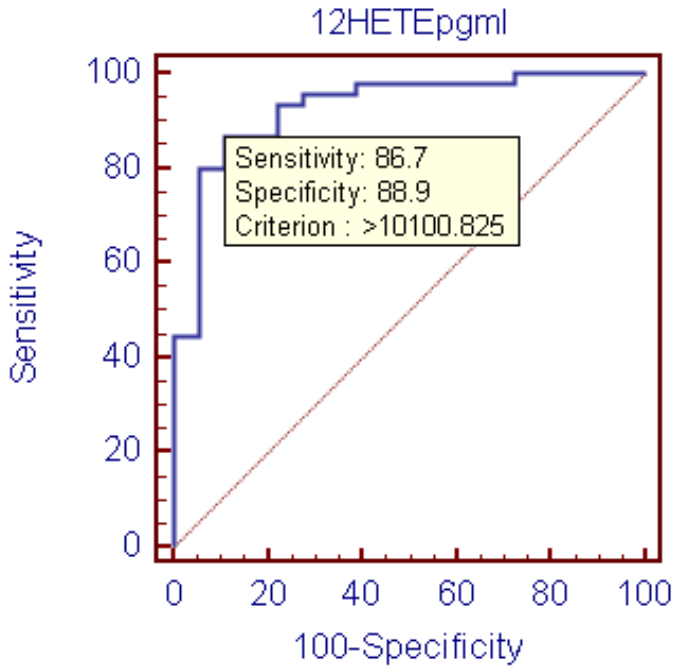


Figure 7.3.2: 12 (S)-HETE serum assay ROC Curve with sensitivity, specificity, and cut-off value in brackets respectively for statistically significant comparison of normal and lung cancer (86.7%, 88.9%, 10.10 ng/ml). Area under the curve: 0.927.

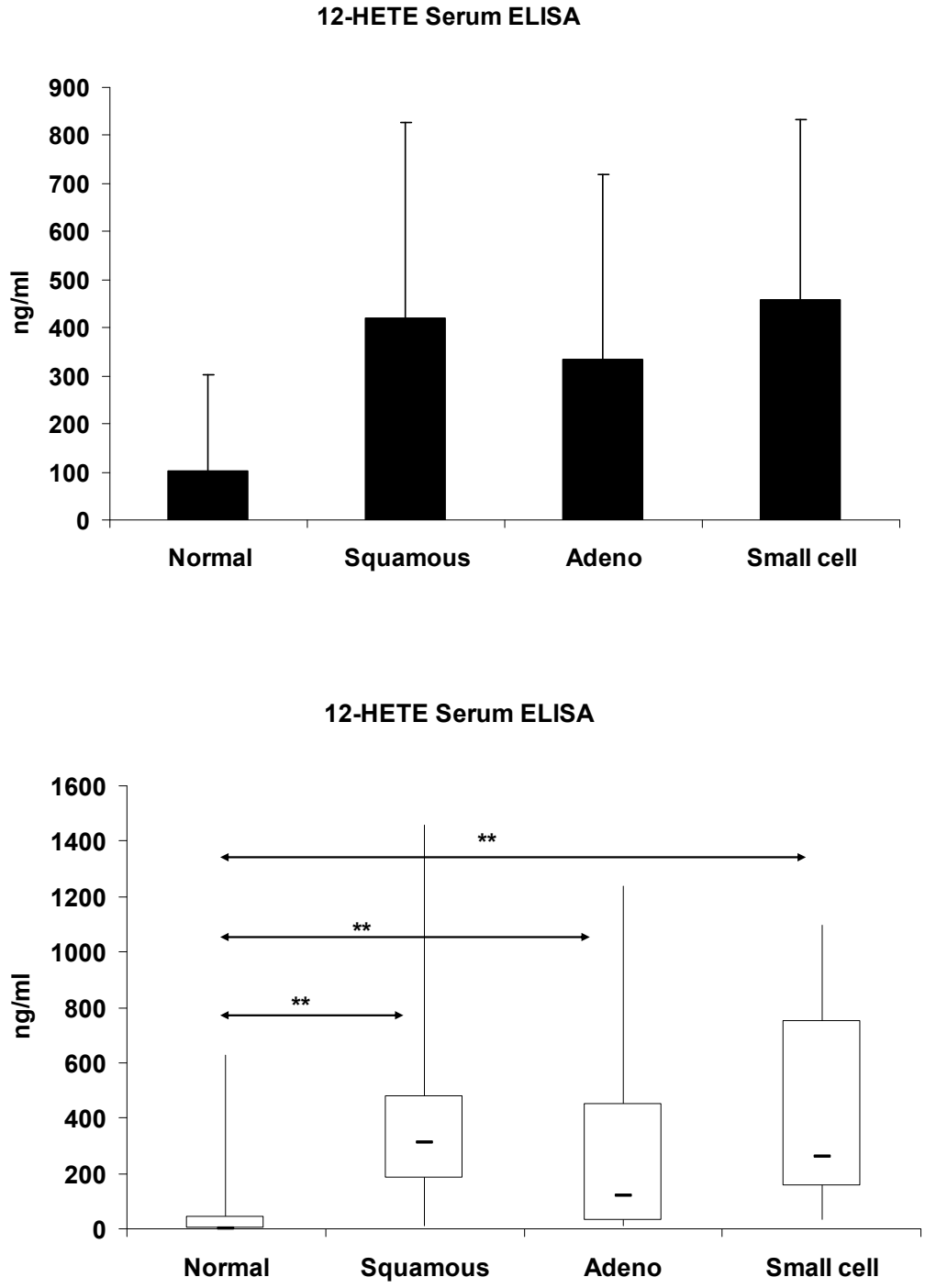
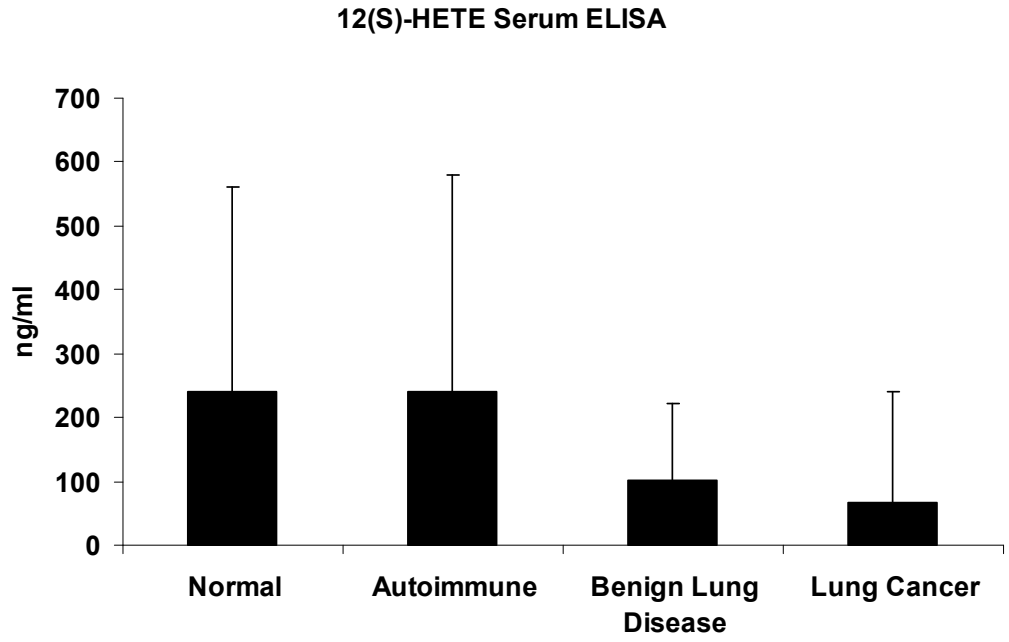


Figure 7.3.3: 12(S)-HETE ELISA: Bar chart and Box and Whisker Plot showing normal compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples included 20 normal, 16 squamous, 24 adenocarcinoma, and 18 small cell carcinoma. T-test scores: Normal v SCC (*p* value=0.004), Normal v AD (*p* value=0.004), Normal v SCLC (*p* value=0.0007).



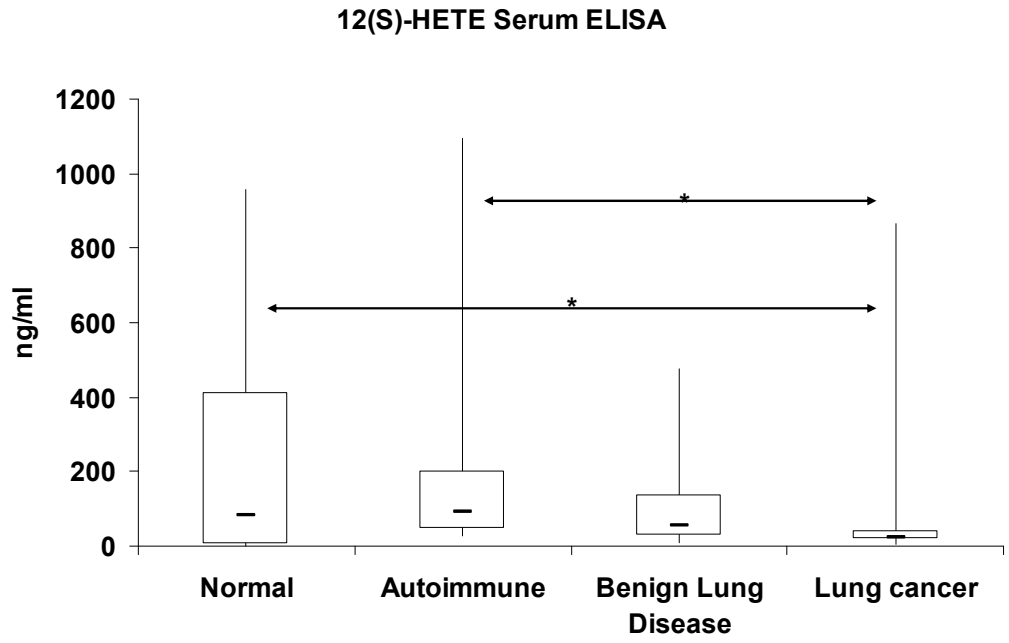
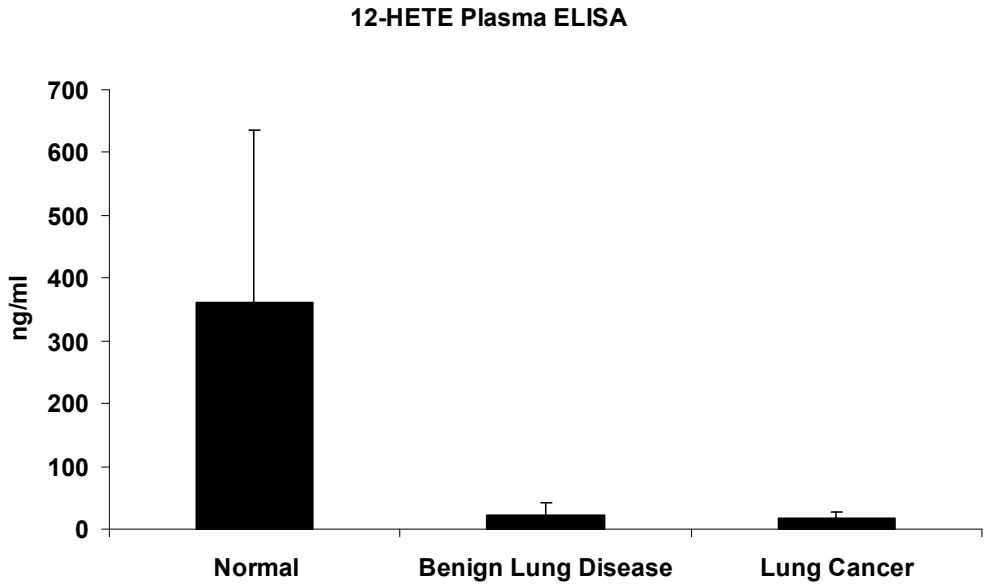


Figure 7.3.4: 12(S)-HETE ELISA: Bar chart and Box and Whisker Plot showing normal compared to each early-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples included 16 normal, 14 autoimmune, 19 benign lung disease, and 24 lung cancer. T-test score; Normal v Lung Cancer (*p* value=0.03), Autoimmune v Lung Cancer (*p* value=0.04).



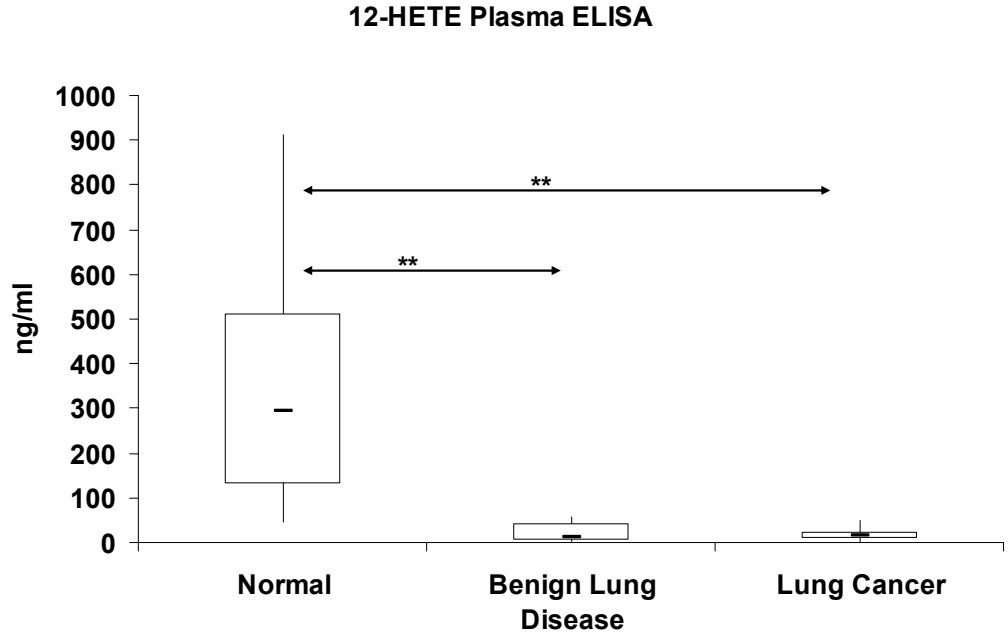


Figure 7.3.5: 12(S)-HETE ELISA: Bar chart and Box and Whisker Plot showing normal compared to each early-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient plasma samples included 18 normal, 14 benign lung disease, and 39 lung cancer. T-test score; Normal v BLD (*p* value=7E-05), Normal v Lung Cancer (*p* value=1E-10).

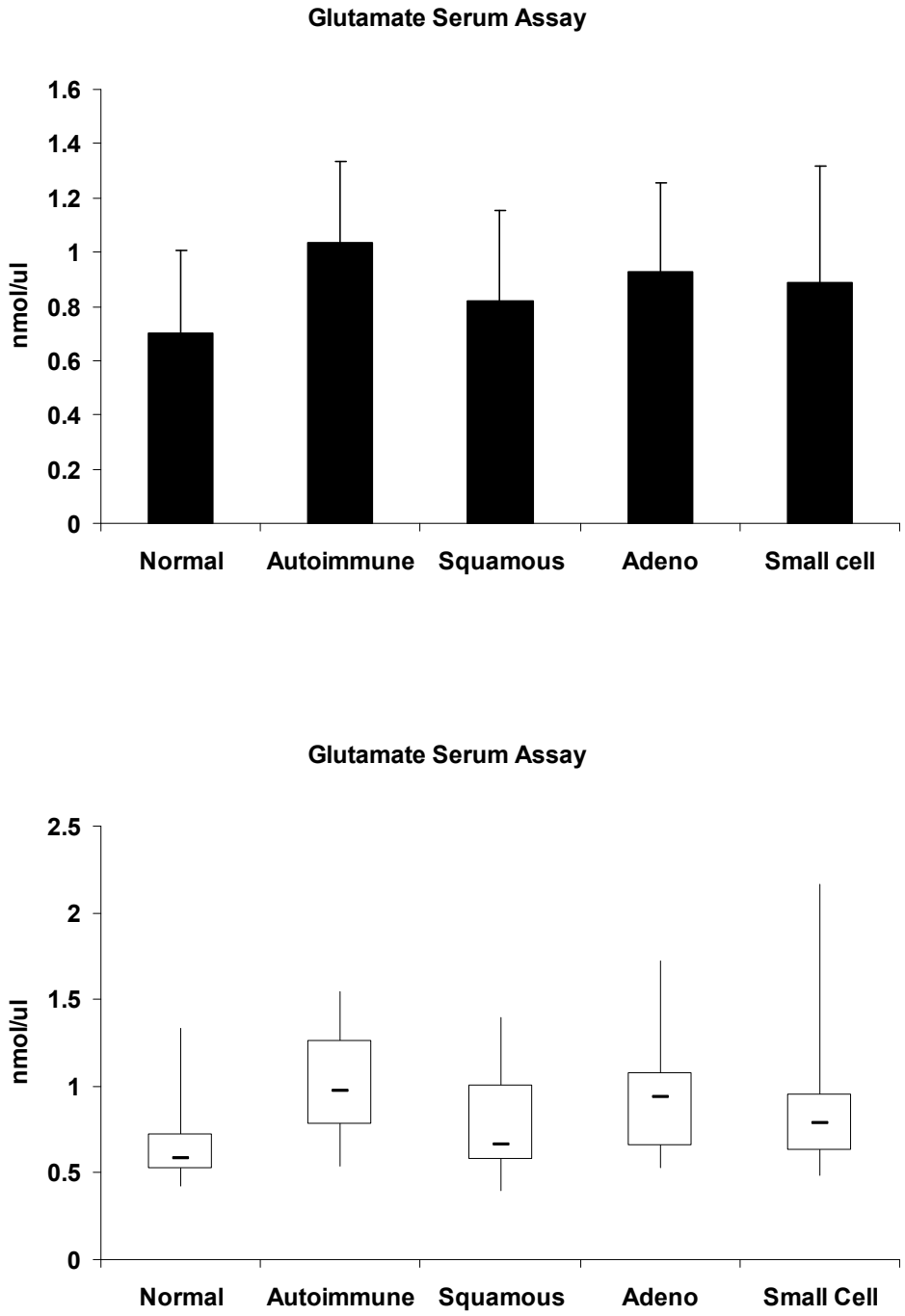
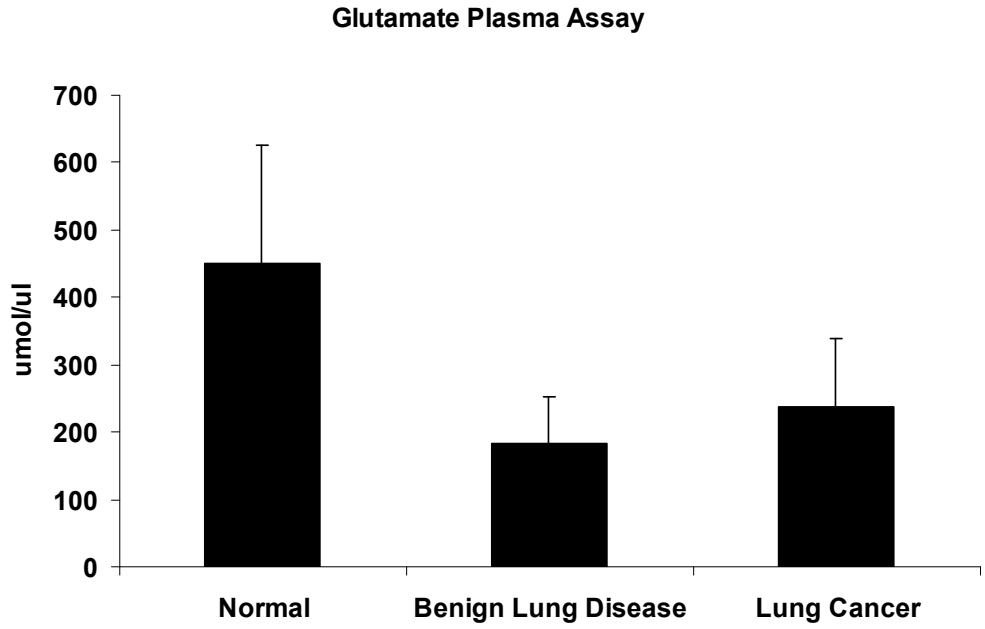


Figure 7.3.6: Glutamate serum ELISAs: Bar charts showing normal and autoimmune controls compared to each late-stage lung cancer type. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples included 10 normal, 20 autoimmune, 11 squamous, 23 adenocarcinoma, and 16 small cell carcinoma.



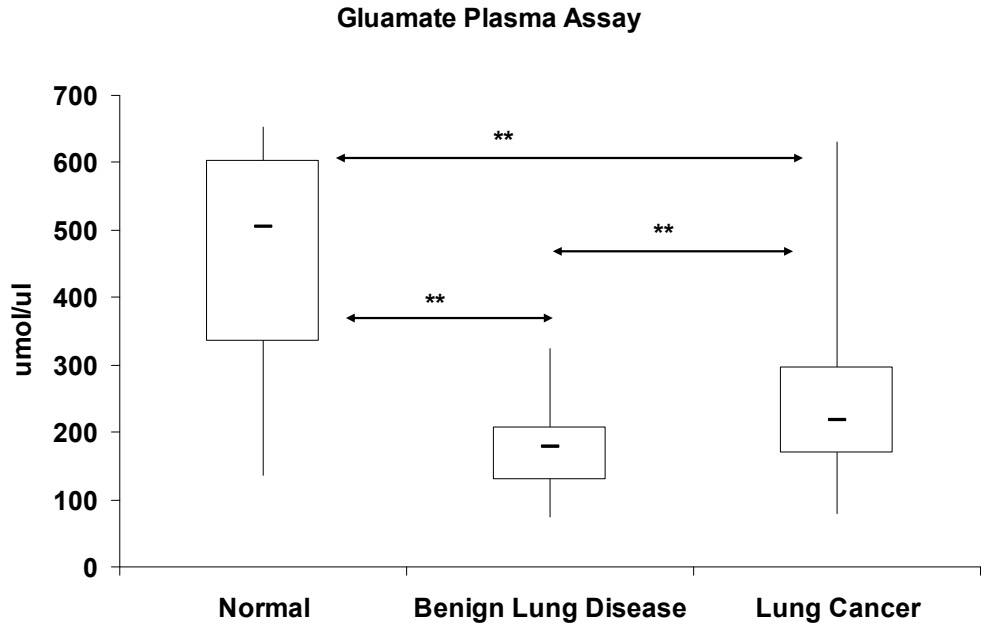


Figure 7.3.7: Glutamate plasma assay: Bar chart and Box and Whisker Plot showing two assays combined using early-stage cancer plasma. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient plasma samples included 30 normal, 40 benign lung disease, and 90 early-stage lung cancer. T-test score; Normal v BLD (*p* value=1E-12), Normal v Lung Cancer (*p* value=7E-13), BLD v Lung Cancer (*p* value=0.002).

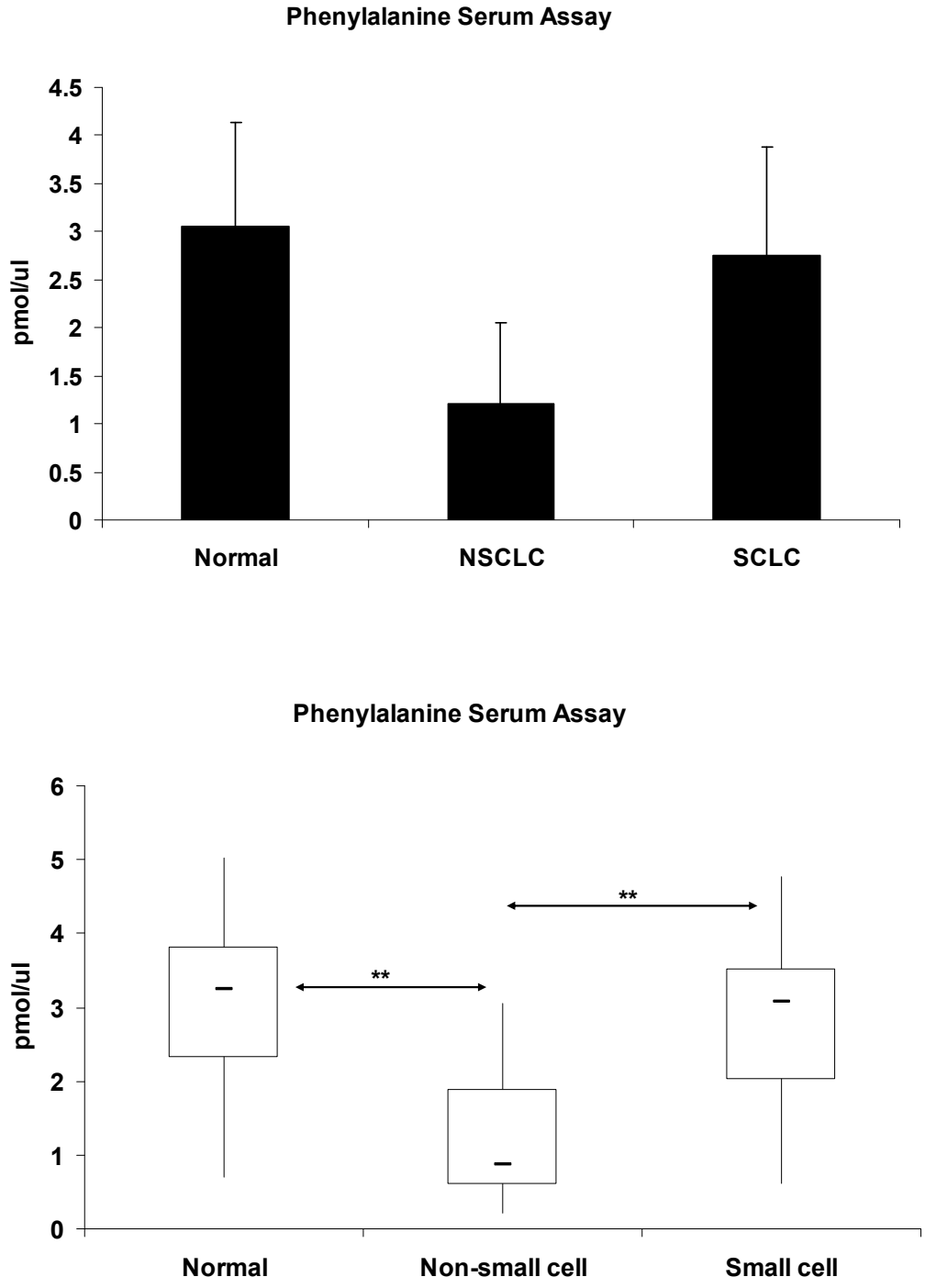


Figure 7.3.8: Phenylalanine serum assay: Bar chart and Box and Whisker Plot using late-stage cancer serum. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples included 18 normal, 40 non-small cell lung cancer, and 16 small cell lung cancer. T-test score; Normal v NSCLC (*p* value=3E-09), NSCLC v SCLC (*p* value=8E-07).

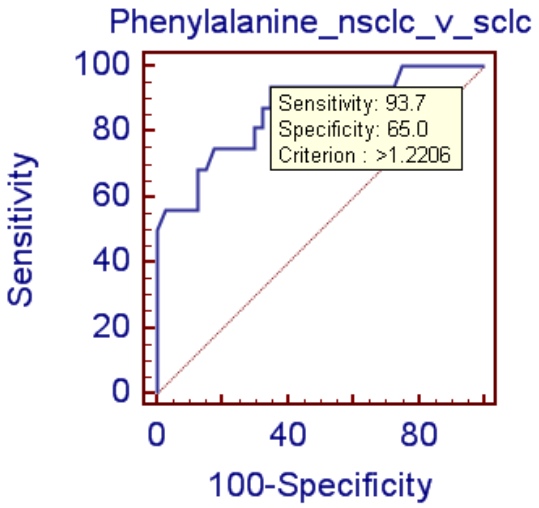
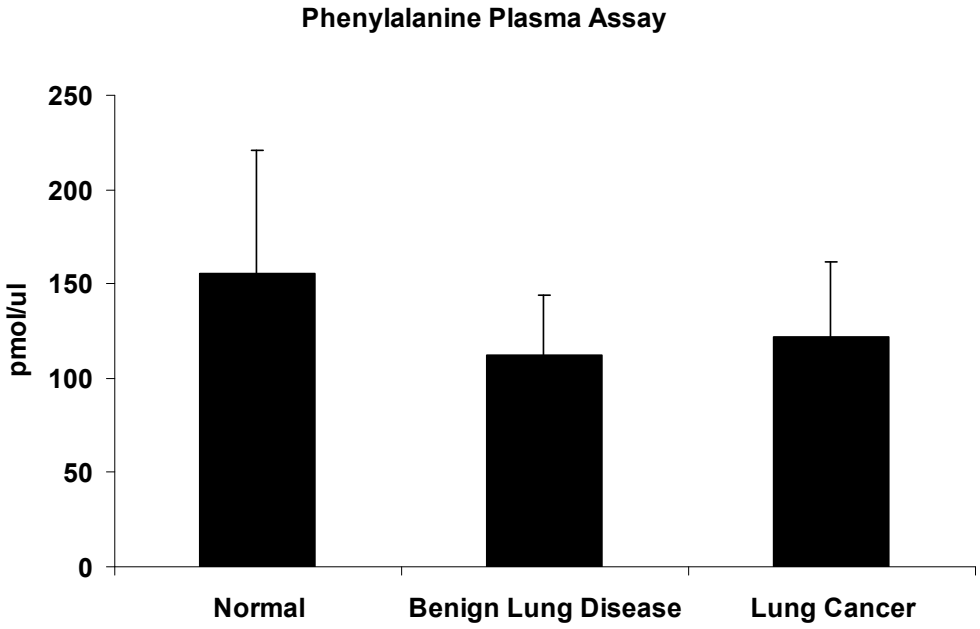


Figure 7.3.9: Phenylalanine serum assay ROC Curve with sensitivity, specificity, and cut-off value in brackets respectively for statistically significant comparison of non-small cell lung cancer and small cell lung cancer samples (93.7%, 65.0%, 1.2206 pmol/ul). AUC value was 0.866.



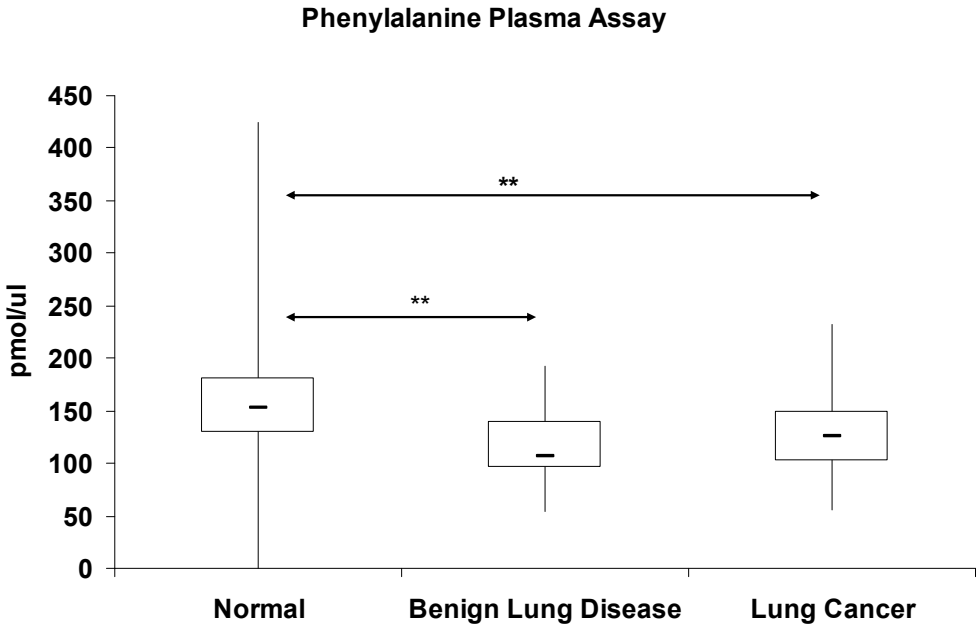
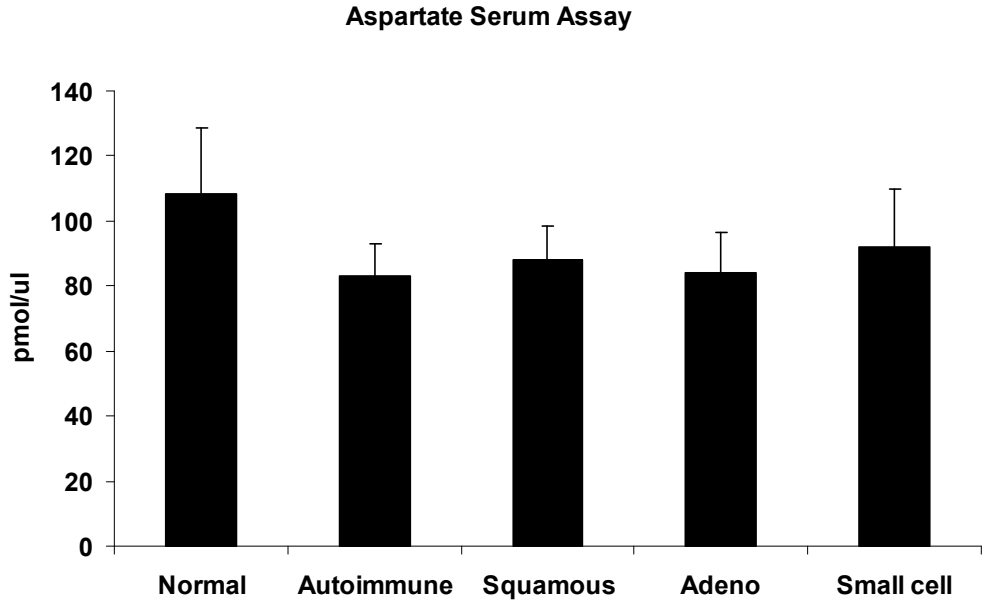


Figure 7.3.10: Phenylalanine plasma assay: Bar chart and Box and Whisker Plot showing two assays combined. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples included 30 normal, 40 benign lung disease, and 89 early-stage lung cancer. T-test score; Normal v BLD (*p* value=0.002), Normal v Lung Cancer (*p* value=0.005).



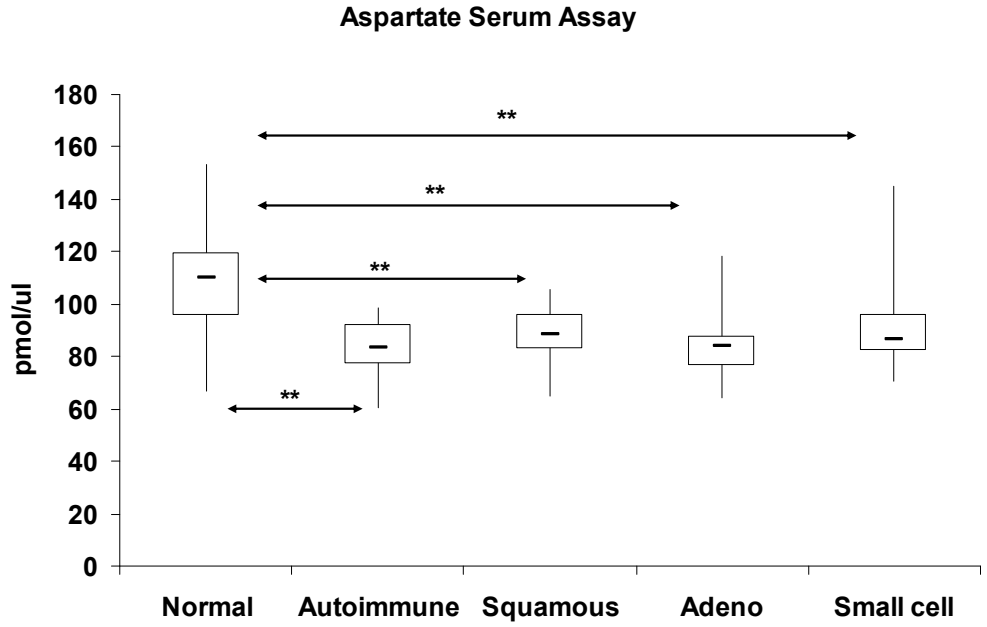


Figure 7.3.11: Aspartate serum assay: Bar chart and Box and Whisker Plot showing two assays combined. For box and whisker the box represents the interquartile range, the horizontal line the median, and the whiskers the highest and lowest quartiles. Patient serum samples included 20 normal, 40 autoimmune, 22 squamous, 46 adenocarcinoma, and 32 small cell carcinoma. All cancer samples were late-stage. T-test score; Normal v Autoimmune (*p* value=3E-08), Normal v SCC (*p* value=0.0002), Normal v AD (*p* value=2E-07), Normal v SCLC (*p* value=0.004).

CHAPTER EIGHT

AUSHON MULTIPLEX IMMUNOASSAY PLATFORM FOR LUNG CANCER BIOMARKER DISCOVERY

8.1 Background

Aushon Cira immunoassay platform can be used for protein biomarker discovery by combining the sensitivity and reproducibility of a singleplex ELISA with the advantages of multiplex. Using the standard 96-well plate format the Cirascan from Aushon can analyse up to 12 analytes per well, the arrays printed precisely in a circular pattern, using quantitative chemiluminescence. Unlike the bead-based method of the Luminex multiplex where beads are mixed by the user before addition to the plate, the Aushon arrays are delivered with chosen analytes pre-spotted. Imaging and analysis software allow the user to view the spots and adjust the standard curve (e.g. removal of outliers) if needed.

Cytokines are regulatory proteins released by cells of the immune system where they act as intercellular mediators for the generation of an immune response. Chemokines are a low-molecular weight type of cytokine classified by their ability to induce chemotaxis or chemokinesis, whereby cells can direct their movements according to certain chemicals in their environment, through controlled or random mechanisms respectively. Chemokines can play a role in cancer through the stimulation of angiogenesis and tumour growth through direct or indirect recruitment of tumour-associated macrophages (Vicari and Caux 2002) and also in cancer treatment by the combination of chemokine receptor antagonists with traditional chemotoxic agents (Wu, Lee et al. 2009).

Aushon biosystems provide an intensive catalogue of chemokines and cytokines to assay. Our experiment used a panel of nine chemokines; macrophage inflammatory

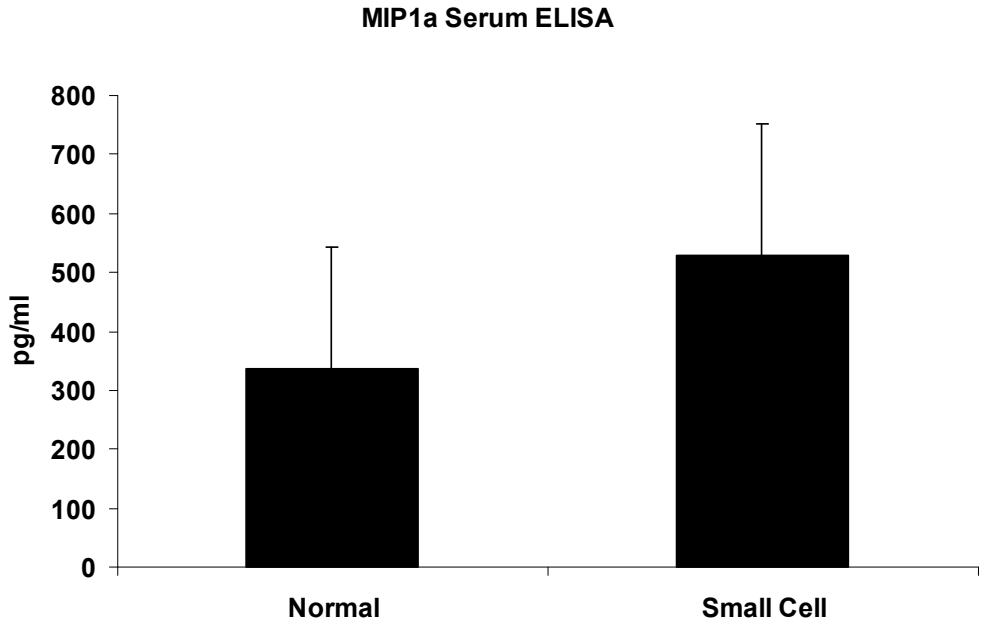
proteins alpha and beta (MIP1a and MIP1b, also known as chemokine (C-C motif ligand 3 and 4 respectively (CCL3 and CCL4))), monocyte chemotactic protein 1 (MCP-1, also known as CCL2), Regulated on Activation, Normal T-cell Expressed and Secreted (RANTES, also known as CCL5), I309 (also known as CCL1), Thymus and activation-regulated chemokine (TARC, also known as CCL17), Eotaxin (also known as CCL11), macrophage-derived chemokine (MDC, also known as CCL22), and interleukin-8 (IL-8).

CCL3 is involved in the acute inflammatory state in the recruitment and activation of polymorphonuclear leaukocytes, CCL4 acts as a chemoattractant for natural killer (NK) cells, monocytes and other immune cells, CCL2 recruits monocytes, memory T cells and dendritic cells to the sites of inflammation, CCL5 attracts T-cells, eosinophils and basophils and recruits leukocytes to inflammatory sites, CCL1 attracts monocytes, NK cells, immature B cells and dendritic cells, CCL17 binds and induces chemotaxis in T-cells, CCL11 selectively recruits eosinophils, CCL22 displays chemotactic activity for monocytes, dendritic cells, NK cells, and chronically activated T lymphocytes, and IL-8 is known to be a potent promoter of angiogenesis.

Due to an unavailability of assay wells only small cell lung cancer was used. In total nine normal serum samples were compared to nine small cell lung cancer serum samples. Although some chemokines showed potential as small cell lung cancer biomarkers in the discovery phase, no validation in larger cohorts of patient samples was performed.

8.2 Discovery Phase

A.



B.

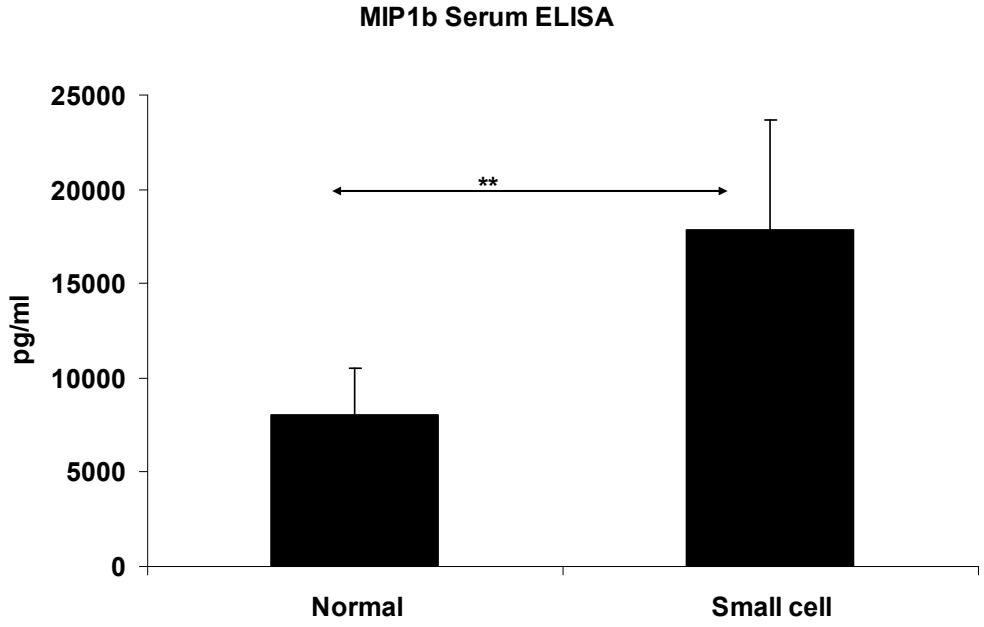
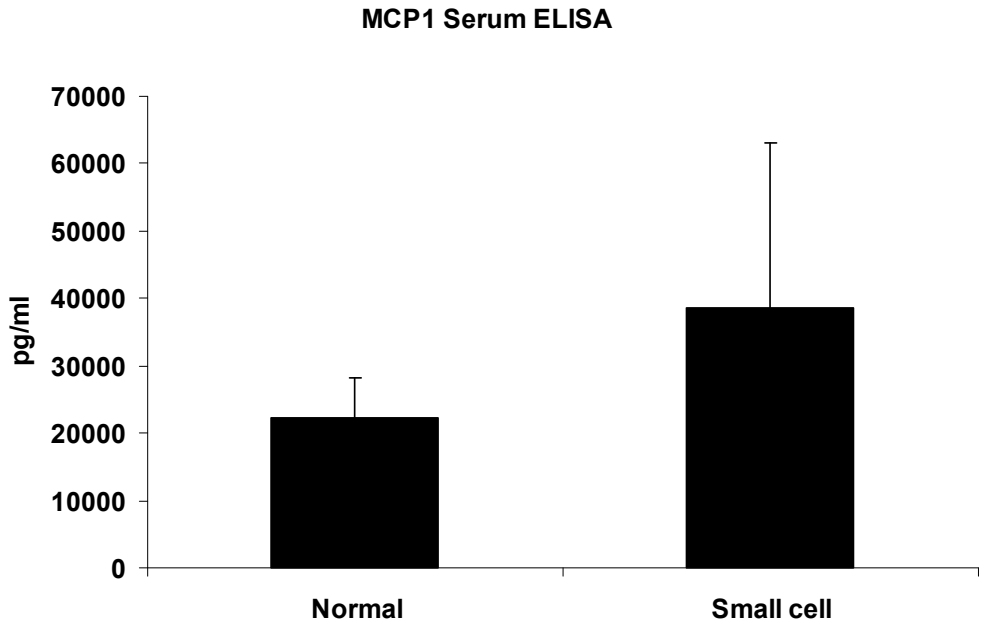


Figure 8.2.1: Macrophage inflammatory protein alpha (A) and beta (B) serum assay. Bar chart showing normal compared to late-stage small cell carcinoma. Patient serum samples included 9 normal and 9 small cell lung carcinoma. Student's t-test with p value less than 0.01 is indicated by two asterisks (**).

A.



B.

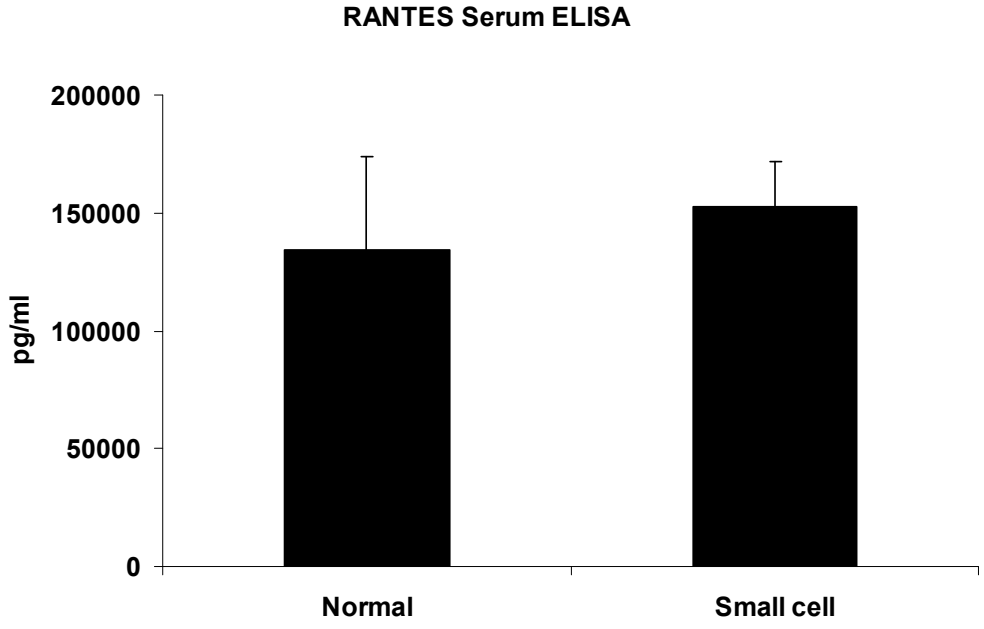
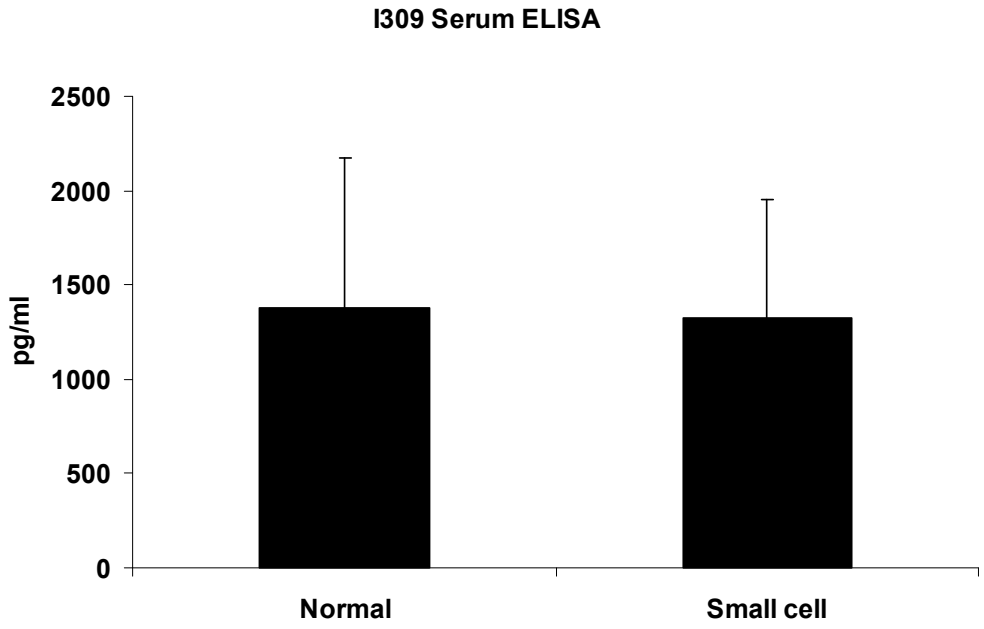


Figure 8.2.2: MCP1 (A) and RANTES (B) serum assay. Bar chart showing normal compared to late-stage small cell lung carcinoma. Patient serum samples included 9 normal and 9 small cell carcinoma.

A.



B.

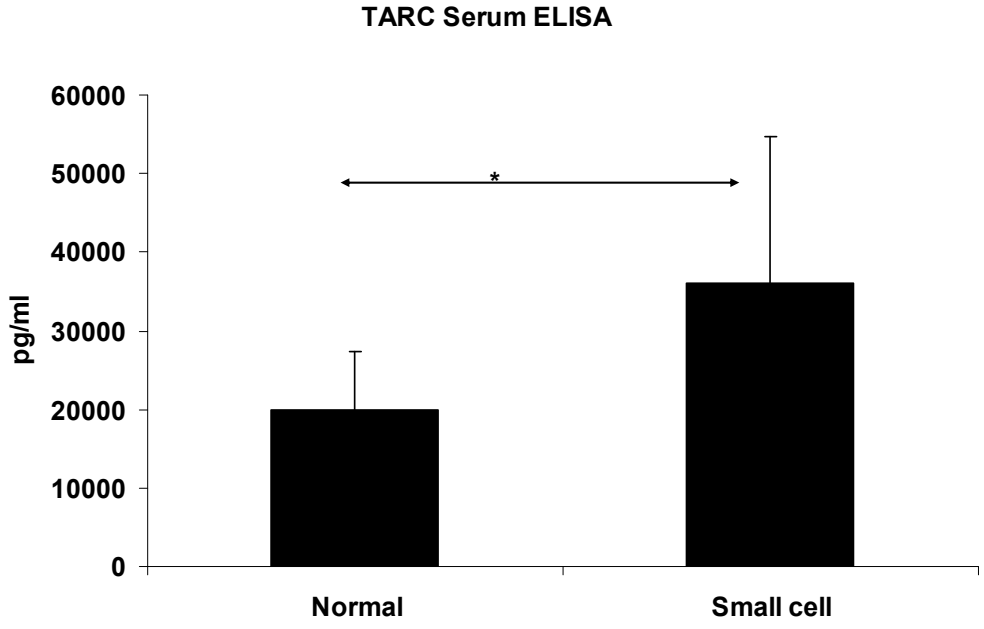
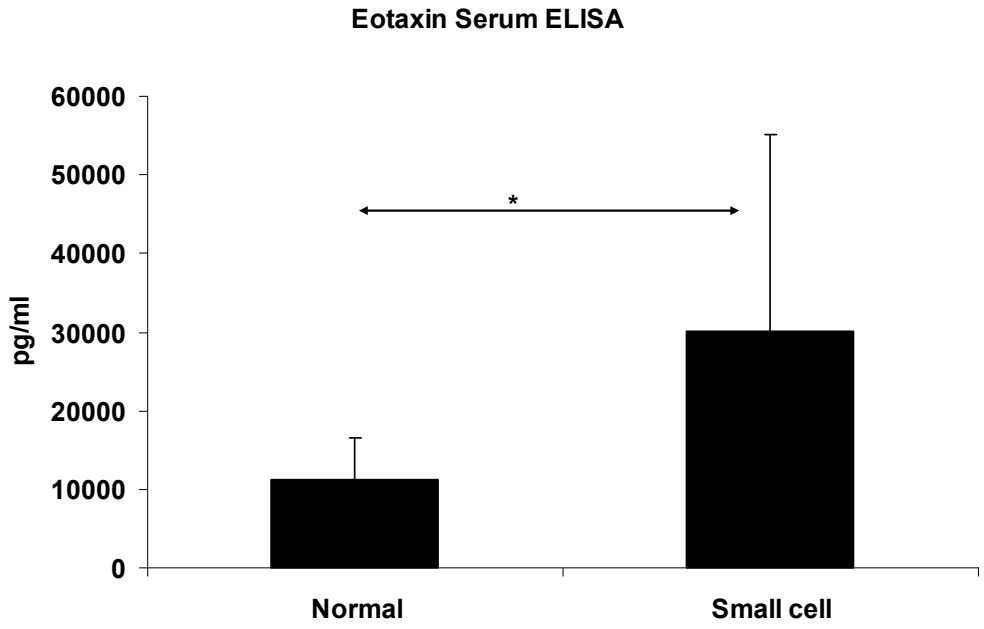


Figure 8.2.3: I309 (A) and TARC (B) serum assay. Bar chart showing normal compared to late-stage small cell carcinoma. Patient serum samples included 9 normal and 9 small cell lung carcinoma. Student's t-test with p value less than 0.05 is indicated by one asterisk (*).

A.



B.

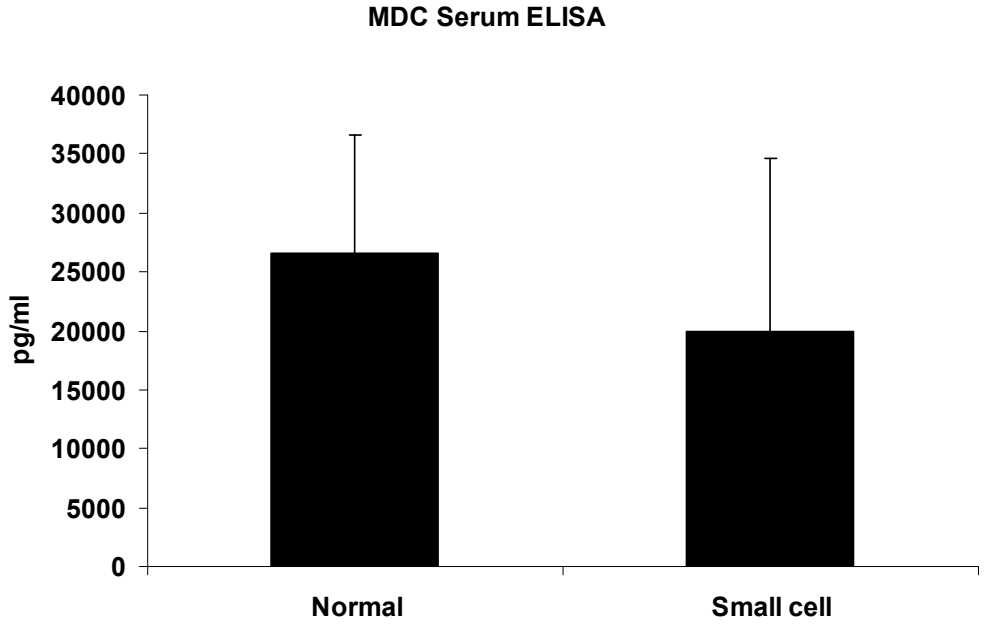


Figure 8.2.4: Eotaxin (A) and macrophage-derived chemokine (B) serum assay. Bar chart showing normal compared to late-stage small cell carcinoma. Patient serum samples included 9 normal and 9 small cell lung carcinoma. Student's t-test with p value less than 0.05 is indicated by one asterisk (*).

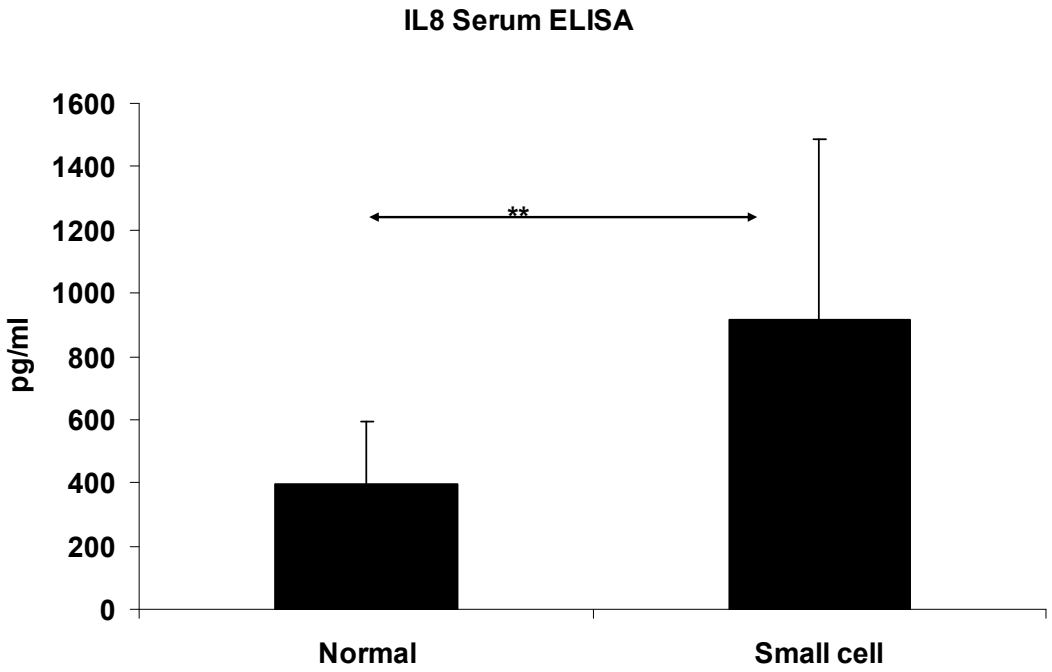


Figure 8.2.5: Interleukin-8 serum assay. Bar chart showing normal compared to late-stage small cell lung carcinoma. Patient serum samples included 9 normal and 9 small cell carcinoma. Student's t-test with p value less than 0.01 is indicated by two asterisks (**).

CHAPTER NINE

DISCUSSION

9.1 Introduction

Lung cancer is the most common cause of cancer-related death in the world. In Ireland it is the biggest cause of cancer death in both men and women, overtaking breast cancer in women. In 2010 in Ireland 1708 people died from lung cancer, 1006 of those men and 702 women. In the same period breast cancer deaths amounted to 634 (www.cancer.ie).

The American Cancer Society has estimated that in 2013 in the United States there will be 228,190 new cases of lung cancer, of which 159,480 patients will die, accounting for 27% of all cancer deaths, more than the cancer deaths from colon, breast, and prostate cancers combined (www.cancer.org).

The five-year survival rate of lung cancer in Ireland is 11% and though this is higher than England (8.4%), Scotland (8.2%), and Wales (10.4%) it is lower than many other European countries such as Germany (14.75%), Poland (14%), Sweden (14%) and the Netherlands (13%). In the US the five-year survival rate is 16.3%, however when it has metastasised to other organs it is only 3.5% (Wender, Fontham et al. 2013).

The biggest cause of lung cancer is smoking, accounting for 80-90% of cases. Often it is diagnosed at quite a late stage, usually when the cancer has already spread to other organs and curative treatment is not an option. Approximately 75% of patients present with symptoms due to metastatic or advanced local disease that is not amenable to cure. Symptoms can take many years to develop, often doing so when the disease is in an advanced stage. These symptoms vary and can often be overlooked as pertaining to lung cancer, both by patients and by doctors; these include persistent or intense coughing, pain in the chest, shoulder, or back from

coughing, hoarseness of voice, difficulty breathing and swallowing, and chronic bronchitis or pneumonia.

If lung cancer spreads, additional symptoms can present themselves in the newly affected area. For example if it spreads to the brain it can cause vertigo, headaches, or seizures. As it spreads and uses more of the body's energy patients can present additional symptoms that are often associated with other ailments; for example fever, fatigue, unexplained weight loss, joint pain, swelling in the neck or face, and general weakness.

Cancer screening is used to identify asymptomatic patients with unrecognised symptoms and to identify those at increased risk. It goes without saying that the earlier a cancer is diagnosed the less likely it has spread to other organs and the more treatable it is. However for screening to be a viable process it must benefit patients by increasing life expectancy, improving quality of life, or increase the chances of ridding the body of cancer altogether. For a screening technique to be ideal it must show low rates of false-positive results (i.e. results that show cancer is present when in fact it is not), must leave a large fraction of those tested unharmed (i.e. low risk), and should be not overly expensive.

Screening methods are covered in Section 1.2.4 though it has been shown that regular sputum examination and chest radiography programs did not effectively reduce lung cancer mortality, even though early detection of lung cancer was possible with these programs. Updated guidelines released by the American Cancer Society on lung cancer screening have indicated sufficient evidence for screening with low-dose computed tomography (LDCT) in certain high-risk individuals. An extensive study in the U.S. by an expert Lung Cancer Screening Panel recommended LDCT screening for select patients at high risk for lung cancer as

results from a large randomised national lung screening trial (NLST) show that lung cancer screening with LDCT can decrease lung cancer-specific mortality by 20% and even decrease all-cause mortality by 7%, though their findings do also indicate that 320 high-risk individuals must be screened with LDCT to prevent one death (Wood, Eapen et al.). Before this could be implemented large-scale the cost-effectiveness of LDCT screening would need to be analysed; reduction in mortality must be weighed against the harms from positive screening results and over-diagnosis, as well as the costs, which includes not only the screening examination but also the diagnostic follow-up and treatment, all of which will depend on how this screening method is implemented in particular with regard to the eligibility criteria, screening frequency, interpretation threshold, diagnostic follow-up, and treatment (Aberle, Adams et al.).

No single screening method is efficient for effective lung cancer detection and the use of companion diagnostics could help with diagnosis. This can involve the use of molecular assays that measure the levels of proteins, genes, or specific mutations that could help confirm the results from standard screening methods. These assays can not only help with early detection but also help provide a specific therapy for a patient's condition (known as personalised medicine) by stratifying disease status, selecting the correct medication, and tailoring the dosages of that medication to a particular patient's needs.

These molecular assays are designed by using biomarkers of a particular disease. A biomarker is a biological molecule that can be used to ascertain the disease state of a patient. By its very nature cancer changes the normal processes of the human body. By changing these processes different genes are affected and in turn the type

and level of proteins produced. The levels of these proteins therefore could indicate the presence of cancer and ideally the type of cancer and state of the cancer (e.g. whether it has contracted after treatment). Proteomic techniques can be used to find potential protein biomarkers that may help with the diagnosis and prognosis of lung cancer. As it stands no biomarker for lung cancer is in standard use, though some biomarkers (often more specific for other cancers) are used (See Section 1.5.1). Single biomarkers would never be sensitive or specific enough to accurately identify a disease state, so ideally a panel of biomarkers would be used to account for the heterogeneity of lung cancer.

The less invasive the method of detection for these biomarkers the better, so their presence in biological fluids such as serum/plasma, saliva, and urine are often examined. Tissue markers can also be effective indicators of disease state but involve the use of biopsies so are therefore more invasive than the use of biological fluids. Finding biomarkers that may detect lung cancer at an early stage, determine the type of lung cancer involved, and/or play an effective role in determining the usefulness of particular lung cancer treatment regimes, could greatly benefit the overall survival of lung cancer sufferers.

The aim of this thesis was to find potential lung cancer biomarkers using proteomic techniques and validate the most interesting candidates in biological fluids and/or tissue. The main focus was to validate biomarkers in serum and/or plasma, though initially serum was the biological fluid of choice. Serum however is very complex with many high abundant proteins that can mask the more interesting disease specific low abundant proteins. The removal of these high abundant proteins with techniques such as immunodepletion are insufficient to uncover the low abundant

proteins as often the column will become saturated and fail to remove a high enough percentage of high abundant proteins. An additional problem arises when the low abundant proteins bind to high abundant ones and are removed with them. Albumin alone constitutes over 50% of the protein content in serum and due to its 'sticky' nature many small low abundant proteins can bind to it and be removed with it. In fact the main 22 proteins in serum account for 99% of all protein content and together disguise the many thousands of other more disease-specific proteins.

9.2 2D-DIGE Analysis of Conditioned Media

To potentially unearth these disease-specific proteins before validating in biological fluids in very expensive commercial assays a cell culture model was initially proposed. In order to detect these biomarkers in serum they would need to find their way into the bloodstream. Tumours will secrete or leak proteins that can find their way into the serum/plasma of patients giving a potential molecular fingerprint reflecting physiological conditions. This secretion and leaking of proteins is commonly associated with cancer. The process of cancer initiation and progression, and the signalling pathways involved, are not confined to the cancer cell itself but extend to the tumour-host interface, a dynamic environment in which information passes between the tumour cells and the normal host tissue. It is conceivable these secreted and/or leaked proteins from the tumour and its microenvironment will ultimately make their way to the serum proteome, where their detection could create a proteomic fingerprint of the tumour of origin. Secreted proteins account for 10-15% of the proteins encoded by the human genome (Karagiannis, Pavlou et al. 2010), so one can predict with confidence that proteins or their fragments

originating from cancer cells or their microenvironment may eventually enter the circulation.

In this study lung cancer cell lines were used as a model. Lung cancer cell lines secrete and/or leaked proteins into the surrounding media and this environment could have similarities with the tumour microenvironment or the serum proteome of a lung cancer patient. This media, without serum which can mask the secreted or shed proteins, is known as conditioned media. The cell lines were initially grown in serum-containing media until healthy and ready to grow in serum-free media for 24/48h. During this period the cancer cell lines will secrete/shed proteins into the media that could potentially be used as biomarkers. In order to discover these biomarkers the conditioned media was collected, concentrated, cleaned up (e.g. removal of impurities), and in this case fluorescently labelled to determine differences in protein content and/or abundance for each of the cell lines. The technique used here was 2D-DIGE and its accompanying analysis software DeCyder. This allowed statistical comparisons to be made between all cell lines.

In total sixteen lung cancer cell lines were examined as well as a normal human bronchial epithelial cell line (NHBE). The four major lung cancer types were looked at; adenocarcinoma (AD), squamous cell carcinoma (SCC), large cell carcinoma (LCC), and small cell lung cancer (SCLC). Together AD, SCC, and LCC are classed as non-small cell lung cancer (NSCLC). For each of the four lung cell types four separate cell lines were examined. In the case of SCC this included two poorly differentiated squamous cell lines (DLKP and DLRP).

2D-DIGE was performed on the conditioned media and the subsequent gels scanned

by the Typhoon Image Analyser and imported into DeCyder. Using the DIA and BVA tools DeCyder can compare and match spots among multiple gels by using the Cy2 pool as reference. Abundance differences (fold changes) between the same spots in different gels are extrapolated and assigned confidence scores in the form of t-test or 1-WAY ANOVA depending on the comparison.

After the DeCyder analysis of the 2D-DIGE results the most interesting protein spots were chosen for mass spectrometry analysis and subsequent identification. After plugging these identifications into DeCyder tables were generated that showed statistically significant (p value cut-off of 0.1 was chosen) comparisons (See Tables 3.3.1 to 3.3.12). From these tables the proteins showing the biggest fold changes were selected for potential validation.

In order to validate these proteins in serum, commercial assay kits are needed. This involves either an ELISA or a biochemical assay (See Section 1.8). These assays can quantify the levels of particular proteins in the serum/plasma of healthy and disease patients. Unfortunately commercial assay kits were not available for many of the interesting proteins and therefore the selection process was restricted to those proteins with commercially available kits. Although this meant that some of the proteins selected did not show big fold changes in any of the comparisons, the possibility was there that the fold changes in serum for these proteins might be more pronounced than the conditioned media experiment indicated. Although this throws up a potential invalidity to this method for lung cancer biomarker discovery phase it should be noted that a cell culture based method such as this will never accurately reflect the proteome fingerprint of biological fluids and should be seen as a relatively crude initial discovery phase. This is why regardless of the fold change and p value, any protein from these experiments could not be considered potential

lung cancer biomarkers until some sort of validation step in cancer serum/plasma or in cancer tissue was performed.

From the DeCyder comparisons the following proteins were chosen for validation; hnRNPA2B1, pyruvate kinase 3 (also known as pyruvate kinase M2 or PKM2), HSC70-interacting protein (Hip), triosephosphate isomerase, GAPDH, lactate dehydrogenase, and peroxiredoxins 1, 2, and 3. Assays on serum of normal and lung cancer patients were performed on all these proteins. Lung cancer samples only included patients with squamous cell carcinoma, adenocarcinoma, and small cell carcinoma. There were insufficient large cell lung carcinoma serum/plasma samples to be included in any of the validation phases.

Further serum and plasma assays were performed on the proteins that showed the greatest abundance differences in serum between normal and cancer i.e. hnRNPA2B1 and PKM2. For these two proteins, as well as Hip, IHC was performed on lung tissue slides. In the case of hnRNPA2B1, IHC was performed on a mixed tissue slide also. As hnRNPA2B1 showed the most promising results functional assays were performed for this protein, involving the transient knockdown of this protein in the DLKP-M lung cancer cell line to examine its effect, if any, on the invasion and migration of lung cancer cells in the body.

9.3 Validation of Targets from 2D-DIGE Conditioned Media Experiment

9.3.1 Peroxiredoxin (Prx)

Reactive oxygen species (ROS) are derived from the metabolism of molecular oxygen (Halliwell 1999). They are formed and degraded by all aerobic organisms,

leading to either physiological concentrations required for normal cell function, or excessive quantities and the state called oxidative stress (Nordberg and Arner 2001). Oxygen derived species such as superoxide radical, hydrogen peroxide, singlet oxygen and hydroxyl radical are well known to be cytotoxic and have been implicated in the etiology of a wide array of human diseases, and as oxidative damage to cellular DNA can lead to mutations it may play an important role in the initiation and progression of multistage carcinogenesis (Waris and Ahsan 2006). Anti-oxidative enzymes, such as superoxide dismutase (SOD), catalase, and glutathione peroxidise (PGX), scavenge reactive oxygen species and in doing so protect cells against oxidative stress (Fujii and Ikeda 2002). ROS induces other signal transduction proteins and play a pivotal role in the genesis and invasiveness of tumours (Lavrovsky, Chatterjee et al. 2000).

Superoxide anions are formed from the incomplete reduction of oxygen to water during respiration and are spontaneously or enzymatically converted to hyrdrogen peroxide (H_2O_2), which, though not very reactive, can be further reduced to the very damaging hydroxyl radicals (Rhee, Kang et al. 2001). Hydroxyl radicals can reduce disulfide bonds in proteins, specifically fibrinogen, resulting in their unfolding and scrambled refolding into abnormal spatial configurations, a consequence of which has been observed in cancer (Lipinski). To combat this all aerobic cells have H_2O_2-removing enzymes which include catalase, glutathione peroxidise, haem-containing peroxidise, and peroxiredoxin.

Peroxiredoxin (Prx) is a family of peroxidases present in all animals, with molecular weight ranging from 20-30kDa. Prx exists in multiple isoforms, with at least six isoforms in humans (Tripathi, Bhatt et al. 2009). Each isoform distribute

differently in organelles, undergo different reaction intermediates during catalysis, and exhibit different expression patterns during development, with some providing defence against oxidative damage and others appearing to participate in signalling by controlling H_2O_2 concentration (Rhee, Kang et al. 2001). Prx 1 and Prx 2 are cytoplasmic, though small amounts of both have been reported to exist within the nucleus, while Prx 3 is found only in the mitochondria. The six isoforms in humans show strong amino acid sequence similarities and commonly exhibit peroxidase activity, which is dependent on reduced forms of thioredoxin (Trx) and/or glutathione (GSH). All members have shown antioxidant activity in the protection of glutamate synthase from oxidation and their reduction of peroxides, with a high affinity for H_2O_2 common to all (Fujii and Ikeda 2002).

Prxs have been implicated in regulating many cellular processes including cell proliferation, differentiation, and apoptosis. The continuous growth and proliferation of lung cancer and persistent exposure to carcinogens may produce large amounts of ROS, particularly H_2O_2 and because this is catalysed by Prx proteins it makes sense that they are overexpressed in many types of cancer, with one study suggesting the upregulation of Prx 1 and Prx 3 in lung cancer may represent an attempt by tumour cells to adjust to the tumour microenvironment in a way that is beneficial to proliferation and survival while maintaining their malignancy potential (Park, Kim et al. 2006).

Peroxiredoxins have been found to be elevated in several human cancer cells and tissues and to influence diverse cellular processes including cell survival, proliferation, and apoptosis (Kim, Bogner et al. 2008). Peroxiredoxin 1 (Prx1) has been shown to be elevated in oral, esophageal, pancreatic, follicular thyroid, and

lung cancer (Yanagawa, Ishikawa et al. 1999, Yanagawa, Iwasa et al. 2000, Chang, Jeon et al. 2001, Qi, Chiu et al. 2005) and to be overexpressed in NSCLC tissue (Chang, Lee et al. 2005), and Prx 1 and peroxiredoxin 2 (Prx2) have both been found to be elevated in breast cancer and head and neck cancers (Noh, Ahn et al. 2001, Karihtala, Mantyniemi et al. 2003). Prx 1 has also been found to be elevated in the serum of epithelial ovarian cancer versus normal/benign patients (Hoskins, Hood et al. 2011).

Although our 2D-DIGE experiments on conditioned media of lung cancer cell lines found peroxiredoxins to be elevated in lung cancer our validation in serum showed the levels of three peroxiredoxins to be decreased in the serum of lung cancer patients compared to normal patients. Prx 1 (Fig. 3.4.13) and Prx 2 (Fig. 3.4.14) were found to be decreased in lung cancer samples compared to normal samples, and while Prx 3 (Fig. 3.4.15) was also decreased, the differences were negligible.
Prx 3 did not show any potential as a lung cancer biomarker however the potential of Prx 1 and Prx 2 can not be dismissed. For example, lower levels of Prx 2 in colorectal tissue has been associated with poorer disease-free survival or poorer overall survival (Ji, Li et al.), and so in turn lower levels of Prx 1 and Prx 2 in lung cancer serum could act as an indicator of disease state. The lung cancer serum samples used for the Prx 1 and Prx 2 ELISAs were for late-stage disease. Though not performed for this Thesis, comparing the levels of Prx 1 and Prx 2 in early stage lung cancer serum to late stage lung cancer serum could prove to be very useful for treatment monitoring and disease staging.

9.3.2 Triosephosphate Isomerase (TPI)

Triosephosphate isomerise is an enzyme involved in glycolysis, catalysing the reversible interconversion of the triose phosphate isomers dihydroxyacetone phosphate (DHAP; a ketose) and D-glyceraldehyde-3-phosphate (GAP; an aldose). TPI is also involved in other metabolic processes such as gluconeogenesis, the pentose phosphate shunt, and fatty acid biosynthesis. Glycolysis and its role in cancer are covered in greater detail in Section 9.3.6. TPI has been mentioned as a possible lung cancer biomarker and one study showed in serum that it could distinguish squamous cell carcinoma patients from healthy patients, their ROC curve analysis taken at a particular cut-off point showed their ELISA could correctly classify 65.6% of SCC patients and 84.7% of healthy patients (Zhang, Xiao et al. 2009).

An ELISA was performed on TPI in serum (See Fig. 3.4.18). Although there was a slight increase in cancer serum compared to normal serum the results were not statistically significant and the Box-and-Whisker plot shows there were large differences between samples for the same condition. TPI as a possible lung cancer biomarker did not look promising and no further assays were performed. Our interest in TPI as a lung cancer biomarker ended with the results of this ELISA.

9.3.3 Lactate Dehydrogenase (LDH)

Lactate dehydrogenase catalyses the interconversion of pyruvate and lactate accompanied by interconversion of nicotinamide adenine dinucelotide (NADH) and NAD^+. When oxygen is absent or in short supply it converts the final product of glycolysis (i.e pyruvate) to lactate. LDH is not a specific cancer marker and tends to be elevated in almost all types of cancer. It is widely used as a prognostic factor for many cancers. Elevated serum levels of LDH was identified as a prognostic factor

for poor survival in lung cancer (Bremnes, Sundstrom et al. 2003), renal carcinoma (Atzpodien, Royston et al. 2003), head and neck cancer (Brizel, Schroeder et al. 2001), and prostate and colorectal cancer (Suh and Ahn 2007). It has also been described as an important indicator of in-hospital mortality for advanced cancer in general (Bozcuk, Bilge et al. 2004). High levels of LDH are an indicator of tissue damage and can be caused by illness or injury. Though high levels of LDH indicate cellular damage they are not specific enough to pin point the exact location of the damage or injury, therefore making it a useful prognosis marker and not diagnosis marker. Because LDH is involved in anaerobic glycolysis it can be assumed its levels will increase in cancer as the uncontrolled growth of cells will require more energy as postulated by the Warburg Effect (See Section 9.3.6). It has been suggested as a marker of airway inflammation (Faruqi, Wilmot et al.), so its usefulness as a lung cancer biomarker can be rightly questioned.

An assay for LDH was performed (Fig. 3.4.16) measuring its activity in serum. As expected its levels were increased in lung cancer compared to normal serum. LDH was elevated in SCC, AD, and SCLC compared to normal and this elevation was statistically significant for all three comparisons. Small cell carcinoma versus normal had a sensitivity of 91.7% and a specificity of 85%, highlighting its trustworthiness as a marker for cancer. High serum levels of LDH are so well characterised as a marker for cancer that a result in this case indicating lower levels of activity in cancer compared to control would raise doubts about the quality of the assay itself. Due to the differences between cancer and normal being relatively small and more importantly due to the fact LDH is elevated in so many conditions and already used as a marker for cancer prognosis in many cancers no further assays

were performed on this protein.

9.3.4 Glyceraldehyde-3-dehydrogenase (GAPDH)

GAPDH is an important enzyme involved in cell energy metabolism. Proliferating cells like cancer cells tend to utilise aerobic glycolysis for energy metabolism. GAPDH catalyses the sixth step in the glycolysis process (See Fig. 9.1). It is often used as a loading control in Western Blots to confirm similar quantities were loaded on the gel for all the different samples because its levels usually remain stable from condition to condition. However it has recently been shown to have an involvement in cell death, transfer RNA transport, DNA repair, membrane trafficking, histone biosynthesis and other activities (Tristan, Shahani et al.). GAPDH has been suggested as a promising target for anticancer therapy as inactivation of its enzymatic activity via targeted agents can likely affect its pro-apoptotic activity too, both in cancer and normal cells, and the bonus of a glycolysis-based therapy is that it can be applied to many types of cancers (Krasnov, Dmitriev et al.).

Although no comparisons of GAPDH from the conditioned media work were statistically significant and therefore not included in the DeCyder result Tables 3.3.1 to 3.3.12, it did show ~4-fold difference between lung cancer and NHBE and there was an ELISA kit available.

An ELISA was performed for GAPDH in serum samples (Fig. 3.4.17). The expected result was there would be no or little difference between cancer and normal samples but the result from the conditioned media work (data not shown) raised doubt in the certainty of this. As can be seen in Figure 3.4.17 there is practically no difference in the levels of GAPDH in normal serum compared to cancer serum, and the bigger spread in results in cancer can be attributed to the fact there were seventy-four cancer samples to 12 normal samples. It is unlikely

GAPDH could act as a biomarker for lung cancer or be used in lung cancer treatment monitoring. Taking GAPDH's function as a loading control into account perhaps what can be taken from this ELISA is that there are similar protein levels in all the normal serum and cancer serum samples. This is more significant than might first appear. The serum samples are a combination from different hospitals and there is always a worry that serum will be collected differently between hospitals that may affect the levels of proteins in the serum samples. Differences can also arise not only from the sampling technique used but the time before proper storage in the hospital and also during delivery from the hospital until proper storage in the freezers in the NICB (see Section 4.1 for greater detail on the possible affects of the preanalytical phases). Although it can not be taken with certainty it can be assumed that the GAPDH ELISA result may indicate that the overall protein content of these serum samples has not been changed despite the likely differences undergone during the preanalytical phase. No further assays were performed on GAPDH.

9.3.5 HSC70-interacting protein (Hip)

Heat-shock proteins (HSPs), present in all organisms, are regarded as intracellular molecules with important chaperone and cytoprotective functions (Hightower 1991). The expression of HSPs is induced by protein-damaging stresses and they have essential roles in protecting cells from the potentially lethal effects of stress and proteotoxicity (impairment of cell function due to misfolding of a protein). Conditions of environmental stress such as heat shock, oxidative stress, heavy metals, or pathologic conditions such as inflammation, tissue damage, infection and mutant proteins associated with genetic diseases, results in the expression of heat-shock proteins that function as molecular chaperones or proteases (Jolly and

Morimoto 2000). During cell stress, folding intermediates can lead to misfolded or otherwise damaged molecules and to prevent this happening a class of proteins called molecular chaperones interact with diverse protein substrates to assist in their folding. Heat-shock proteins, as molecular chaperones repair damaged proteins, or as proteases degrade damaged proteins, to restore protein homeostatis and promote cell survival, helping cells recover from stress and restore balance (Jolly and Morimoto 2000). HSPs are also known to afford protection against protein aggregation and induce solubilisation of loose protein aggregates (Hartl 1996).

HSPs make up 5-10% of the total protein content in a cell under healthy conditions. When protein-damaging stresses induce protein misfolding or aggregation, the affected proteins bind to chaperones and release heat shock factor (HSF) which acts as a transcription factor and binds to heat shock elements within the promoter of *HSP* genes, resulting in an increase, often two- to three-fold, of cellular Hsp concentration (Pockley 2003, Voellmy 2006, Soo, Yip et al. 2008).

Hsp70, designated so because of a molecular weight of 70kDa, has been studied extensively and can be considered the Hsp most inducible by stress (Suzuki, Ito et al. 2006). It is involved in inflammatory diseases, and serum levels of Hsp70 have been directly linked to a patient's inflammatory status (Njemini, Demanet et al. 2004). Due to the function of HSPs it has been proposed that their overexpression could be a potential therapy for neurodegenerative diseases characterised by the accumulation or aggregation of abnormal proteins (Mansilla, Montalban et al.). Serum levels of Hsp70 were shown to be elevated in patients with NSCLC diagnosed at an early or at an advanced stage when compared with healthy control groups (Zimmermann, Nickl et al. 2012). Hsp 70 has been shown to bind to the p53

tumour suppressor protein, the mutation of which is one of the most common events in cancer development (Lane, Midgley et al. 1993), though there is no conclusive evidence that the increased HSP levels are necessary to inactivate tumour suppressor molecules for malignant transformation to take place (Hollstein, Sidransky et al. 1991). Hsp70 expression has been correlated with poor prognosis in breast, endometrial, uterine cervical and bladder carcinomas (Ciocca and Calderwood 2005).

HSP70 and HSC70 are often used interchangeably to describe the heat shock protein 70 but in fact Hsc70 (also known as HSP73) is a constitutively expressed chaperone protein while HSP70 is a stress- and heat shock-induced protein. They are cytosolic isoforms of the heat shock protein family HSP70. HSC70-interacting protein (also described as HSP70-interacting protein or Hip) is a co-chaperone of the 70-kDa heat shock proteins HSC70/HSP70 encoded by the gene ST13 (suppression of tumourigenicity 13), though its role in cancer development is not certain. The expression of ST13 has however been shown to decrease in colorectal cancer tissue compared to adjacent normal tissue (Wang, Zheng et al. 2005). Hip may facilitate the chaperone function of HSC70/HSP70 in protein folding and repair, and in controlling the activity of regulatory proteins such as steroid receptors and regulators of proliferation or apoptosis (Shi, Zhang et al. 2007).

An ELISA was performed on HSC70-interacting protein in late-stage lung cancer serum (Fig. 3.4.11). There was an increase in quantity of Hip in serum for each of the three cancer types (SCC, AD, and SCLC) compared to normal serum. The only statistically significant comparison was normal compared to AD where 20 normal were compared to 28 AD samples. Though there is crossover in the results Hip

shows potential as a biomarker in serum for adenocarcinoma. Being able to differentiate adenocarcinoma from SCC is very useful for selecting appropriate chemotherapy for patients with advanced disease (Terry, Leung et al. 2010). The results from this ELISA are not overly impressive but the potential of Hip as a lung cancer biomarker remains. There are a number of other ELISAs that could be performed, perhaps one comparing a large cohort of SCC and AD samples only. Also there is a need to examine it in plasma samples and also in benign lung disease samples in both serum and plasma.

Immunohistochemistry was also performed on HSC70-interacting protein (Section 3.5). In Table 3.5.2 the intensity scores for the tissue cores is recorded. Tissue cores available were from SCC, AD, SCLC, alveolar cell carcinoma, and normal lung tissue. Alveolar cell carcinoma (ACC) is a primary neoplasm of relatively infrequent occurrence. When it is operable the tumour is typically peripheral and may involve a small area of pulmonary parenchyma or it may be extensive, involving all of a lung or an entire lobe, and tends to arise from the pulmonary parenchyma rather than from a major bronchus (Belgrad, Good et al. 1962). Alveolar cell carcinoma is now referred to as bronchioalveolar carcinoma (BAC), a variant of lung cancer that arises in the distal bronchioles or alveoli that initially exhibit a specific non-invasive pattern. BAC is a sub-group of adenocarcinoma, however the term bronchioalveolar carcinoma is no longer considered valid and has been re-categorised (Lee, Goo et al.).

Table 3.5.3 shows that SCC, SCLC, and ACC showed the strongest staining, with 61.54%, 65.22%, and 53.85% of cores respectively with an intensity score of 3. The scoring system covers; strong (designated 3), intermediate (2), weak (1), and

negative (neg) staining. For AD cores only 23.81% showed strong staining, with alveolar carcinoma showing staining intensity more similar to SCC and SCLC than AD. Staining overall was most intense in SCC, as the remaining 38.46% of cores were intensity 2. One third (33.33% exactly) of AD cores were intensity 1 while 0% of SCC cores and only 7.7% of ACC cores were scored the same. Staining was certainly less in normal lung tissue, with 30% at intensity 2 and 70% at intensity 1. No cores showed negative staining and with very strong staining in many cores it can be assumed that the antibody concentration was too high, despite the optimisation steps. This experiment would need to be repeated using a more dilute antibody concentration to give a more realistic representation of the staining patterns in lung cancer and normal lung tissue cores.

The conclusion can be made that HSC70-interacting protein is more strongly expressed in lung cancer tissue than normal lung tissue. By acting as a chaperone protein for heat shock proteins it is not surprising that Hip levels are increased in cancer. Heat shock proteins are over-expressed in many cancers and implicated in tumour cell proliferation, differentiation, invasion, metastasis, death, and recognition by the immune system (Ciocca and Calderwood 2005) so it is not surprising that a protein needed for its proper folding would also be increased.

9.3.6 Pyruvate Kinase M2 (PKM2)

Pyruvate kinase is an enzyme that catalyses the last, rate-limiting step of glycolysis. Glycolysis, literally meaning the lysis of glucose, first requires the conversion of glucose to pyruvate and then to the waste product lactic acid. In most mammalian cells glycolysis is inhibited by the presence of oxygen, which allows mitochondria to oxidise pyruvate to CO_2 and H_2O. For maintenance of energy production

throughout a wide range of oxygen concentrations it is essential that mammalian cells have versatile metabolisms and conversion of glucose to lactic acid in the presence of oxygen is known as aerobic glycolysis, and when seen in cancer cells as the Warburg Effect (Gatenby and Gillies 2004). Warburg proposed in the 1920's that cancer results from impaired mitochondria metabolism and although this hypothesis has been proven incorrect, increased glycolysis in tumours even in the presence of oxygen has been verified (Semenza, Artemov et al. 2001).

Pyruvate kinase catalyses the transfer of a phosphate group from phosphphenolpyruvate (PEP) to ADP, yielding one molecule of pyruvate and one molecule of ATP. Four isoforms are present in mammals; L, R, M1, and M2. The L and R isotypes are encoded by the *PKLR* gene and their expression is tissue specific and regulated by different promoters (Tamada, Suematsu et al. 2012). The L isotype is expressed in the liver, kidney, and intestine, and the R isotype is expressed in red blood cells (Mazurek, Boschek et al. 2005, Clower, Chatterjee et al. 2010, Mazurek 2011). On the other hand PKM1 and PKM2 are encoded by the *PKM* gene and are the products of 2 mutually exclusive alternatively spliced exons (exon 9 and exon 10 respectively) (Noguchi, Inoue et al. 1986, David, Chen et al. 2010).

In most adult differentiated tissue, such as brain and muscles, the M1 isoform is expressed, whereas M2 is expressed in embryonic cells, adult stem cells, and cancer cells (Mazurek, Boschek et al. 2005). However, the preference of cancer cells for the PKM2 isoform over PKM1 in tumourigenesis has been called into question, studies suggesting that an exchange in PKM1 to PKM2 isoform expression during cancer formation is not occurring, that the PKM2 isoform dominance in cancer is

not a result of a change in isoform expression since PKM2 was also the predominant PKM isoform in matched control tissues, in unaffected lung tissue for example PKM2 accounted for a minimum of 93% of total PKM (Bluemlein, Gruning et al. 2011). Linking PKM2 to tumour growth can not, however, be easily dismissed as it has been suggested the expression of this isoform is necessary, and indeed essential, for tumour growth (Christofk, Vander Heiden et al. 2008, Gupta and Bamezai 2010). Indeed it has been noted that the replacement of PKM2 with PKM1 *reduces* the capacity of human tumour cell lines to develop into a tumour, allowing for the notion that by a change in PK splicing, tumourigenesis can be induced through shifting cellular metabolism to aerobic glycolysis (Christofk, Vander Heiden et al. 2008).

PKM2 is present in very few types of proliferating normal cells but is present at high levels in cancer cells, slowing the passage of metabolites through glycolysis and promoting the shuttling of these substrates through the pentose phosphate pathway and other alternative pathways so that large quantities of reduced NADPH and other macromolecules are produced, which in turn are required for macromolecule biosynthesis and redox balance maintenance that are needed to support the rapid cell division that is undergone within a tumour (Cairns, Harris et al.). It has been suggested that silencing PKM2 alone can retard tumour cell growth by inducing apoptosis and inhibiting proliferation both in vitro and in vivo, making it a desirable target for cancer therapy (Peng, Gong et al. 2011). Several small molecule PKM2 inhibitors and activators have been developed, however as nearly complete knockdown of PKM2 does not completely inhibit cancer cell proliferation, caution should be maintained on the utility of PKM2 inhibition in targeting cancer (Tamada, Suematsu et al. 2012).

One study has shown in their study that tumour M2-PK (pyruvate kinase M2 tumour marker) had higher sensitivity (58%) in the detection of lung cancer than the CEA-Test (sensitivity 39%) or CYFRA 21-1 (sensitivity 48%) and generally showed a higher sensitivity for NSCLC (68%) (Schneider, Velcovsky et al. 2000). It has been shown that for colorectal cancer a combination of tumour M2-Pk with an appropriate first choice-marker (e.g. CEA + Tumour M2-PK) in plasma resulted in a remarkable increase in the sensitivities (Schulze 2000). One study found measurement of tumour M2-PK may be useful in patients with lung cancer as 78% of SCLC patients, 73% of AD patients, and 81% of NSCLC patients showed an increase of tumour M2-PK serum concentration, while its concentration in patients with non-malignant diseases (e.g. bronchitis or tuberculosis) was within normal limits (Oremek, Kukshaite et al. 2007).

Pyruvate kinase has potential as a biomarker of lung cancer. Interestingly PKM2 is elevated in lung cancer serum, both late-stage and early stage, as well as lung cancer plasma compared to controls. Although the results of Fig. 3.4.7 were not statistically significant there is a clear increase of PKM2 in lung cancer types compared to normal and even in Fig. 3.4.8 using earlier stage lung cancer PKM2 is elevated in lung cancer serum. Interestingly in Fig. 3.4.8 it is higher in lung cancer compared to benign lung disease too with significant T-test scores and even in plasma in Fig. 3.4.9 it is elevated in both NSCLC and SCLC compared to the control groups (both normal and benign lung disease). This increase in PKM2 in cancer serum and plasma could be attributed to the cancer cells undergoing aerobic glycolysis, the preferred energy pathway for cancer cells which utilises the PKM2 isoform.

It has been suggested that the PKM2 isoform plays an important role in the growth and metabolic reprogramming of cancer cells in conditions of stress and how upon glucose starvation or lack of glucose due to their proliferative behaviour PKM2 activity is stimulated to alter cellular energy levels, glucose uptake, and lactate production in cancer cells (Keller, Tan et al.). PKM2 can regulate the proportions of glucose carbons that are channelled to synthetic processes (an inactive dimeric form) or used for glycolytic energy production (active tetrameric form) and the use of the dimeric form in tumour cells is always predominant, allowing tumour cells to survive in environments with varying oxygen and nutrient supply, which is invariably the sort of environment cancer cells will find themselves due to their unnatural behaviour and location (Mazurek, Boschek et al. 2005).

The pyruvate kinase biochemical assay showed pyruvate kinase activity in serum to be decreased in cancer serum compared to normal serum. The pyruvate kinase biochemical assay calculates the rate at which pyruvate kinase is being used based on the amount of pyruvate being generated over a controlled time-point. A growing tumour will require more energy from the body and if pyruvate kinase is being produced in greater amounts then perhaps there is more than is required for the amount of PEP available for conversion to ADP. This could result in an increase of pyruvate kinase but a decrease of PEP finding its way into the serum of lung cancer patients. This could explain why pyruvate kinase levels in lung cancer serum are higher than normal serum, while at the same time PEP levels could be higher in normal serum compared to lung cancer serum meaning less pyruvate is being produced in lung cancer serum resulting in less pyruvate kinase activity.

Immunohistochemistry was performed on a lung tissue array slide for PKM2 (Figures 3.5.24 to 3.5.27). The intensity scoring is presented in Table 3.5.4. The results suggest PKM2 would not make a useful tissue marker. Indeed, surprisingly, the expression of PKM2 is highest in normal lung tissue, a total of 40% of normal lung tissue showing strong staining, more than any of the cancers SCC, AD, SCLC, and ACC. In fact all normal lung tissue showed some degree of staining, whereas some of the cancer cores exhibited negative staining e.g. ~24% of AD cores. There are no significant differences between the cancer tissue types to separate them from one another. Considering the elevated levels of PKM2 in lung cancer compared to controls for late-stage lung cancer serum, early-stage lung cancer serum, and early-stage lung cancer plasma, it was unexpected to see higher PKM2 expression in normal lung tissue than in the cancer tissue, however it is known that good serum markers do not necessarily make good tissue markers and vice versa.

9.3.7 HnRNPA2B1

hnRNPA2B1 belongs to the A/B subfamily of ubiquitously expressed heterogeneous nuclear ribonucleoproteins (hnRNPs) (Tauler, Zudaire et al.; Kamma, Horiguchi et al. 1999). A2 and B1 are produced by alternative splicing from a single copy gene (Kozu, Henrich et al. 1995), and differ from each other only by a 12-amino-acid insertion in the N-terminal RNA binding motif (RBD) of B1 (Burd, Swanson et al. 1989). hnRNPs are RNA binding proteins (RBPs) that form complexes with heterogeneous nuclear RNA (Wen, Shen et al.) and play important roles in regulating gene expression at transcriptional and also at post-transcriptional levels (He and Smith 2009), as well DNA repair, telomere biogenesis and cell signalling (Krecic and Swanson 1999; Dreyfuss, Kim et al. 2002). They are among the most

abundant nuclear proteins in mammalian cells (Patry, Bouchard et al. 2003) and appear to influence pre-mRNA processing and other aspects of mRNA metabolism and transport (Weighardt, Biamonti et al. 1996), and can prevent folding of pre-mRNA into secondary structures that may inhibit its interactions with other proteins (Golan-Gerstl, Cohen et al.). It has been noted that they are essential for RNA metabolism, coordinating post-transcriptional events (splicing, transport, cellular localisation, decay and translation of mRNA) by participating in an extensive network of RNA-RBP interactions (Boukakis, Patrinou-Georgoula et al.).

More than 20 distinct hnRNP proteins have been identified in human cells, designated hnRNPs A to U in increasing molecular size from 32 to 110kDa. This family of abundant nuclear proteins include many that share common structural motifs, exhibiting multiple isoforms (products of alternative splicing, as well as of post-translational) and having the ability to shuttle between the nucleus and the cytoplasm (Dreyfuss, Matunis et al. 1993; Burd and Dreyfuss 1994).

The hnRNP A/B subgroup includes members of 32-40kDa size range. They have in common two tandem N-terminal RNA-binding domains of the RRM/RBD type (RNA recognition motif/RNA binding domain) and a C-terminal auxiliary domain rich in glycine (Boukakis, Patrinou-Georgoula et al.). Along with hnRNPA2B1, hnRNPA1 and hnRNPA3 are the best characterised, all of which share a high degree of sequence homology and the presence of several isoforms originating mainly from alternative splicing (He, Brown et al. 2005). hnRNPA2B1 in particular refers to two isoforms, the major hnRNP A2 and the minor B1 form that results from the inclusion of an extra exon (Burd, Swanson et al. 1989). In both proteins and mRNA levels, their proportions vary in different cells and tissues, with B1 constituting

roughly 2-5% of A2 (Kozu, Henrich et al. 1995; Kamma, Horiguchi et al. 1999). The major nuclear function of hnRNPA2B1 is thought to be in splicing and in particular alternative splicing of mRNA which allows many gene products with different functions to be produced from a single coding sequence (Brett, Pospisil et al. 2002). hnRNPA2B1 has been shown to antagonise, in a concentration dependent manner, protein members of the splicing regulator (SR) group, notably ASF/SF2, and to influence the mode of splicing of mRNA target molecules (Hanamura, Caceres et al. 1998).

Overall expression of the hnRNP A/B proteins is tightly regulated during development and known to be tissue and cell type specific (Kamma, Horiguchi et al. 1999). It has been shown in human and rodent tissues that high hnRNP A/B levels occur during embryonic development and drop dramatically in adult lung tissue (Montuenga, Zhou et al. 1998). The expression of hnRNPs in cancer cells has been associated with alterations in their protein levels (over-expression and down-regulation), mRNA abundance, cellular localisation, isoform types and post-translational modifications (such as phosphorylation) (Tockman, Mulshine et al. 1997; Ostrowski and Bomsztyk 2003).

hnRNPA2B1 has previously been found to be a potential early marker of lung epithelial transformation and carcinogenesis (Zhou, Mulshine et al. 1996), up-regulated in bronchial lavage specimens (Fielding, Turnbull et al. 1999), a potential early marker for breast cancer (Zhou, Allred et al. 2001), overexpressed in human pancreatic cancer (Gu, Liu et al.), and a novel oncogene in glioblastoma and a potential new target for glioblastoma therapy (Golan-Gerstl, Cohen et al.). hnRNPA2B1 has also been shown to modulate epithelial-to-mesenchymal transition in lung cancer cell lines (Tauler, Zudaire et al.).

Two serum ELISAs were performed comparing normal serum to late-stage SCC, AD, and SCLC. In the first ELISA (fig. 3.4.1) comparing almost identical quantity of samples; normal (n=20), SCC (n=20), AD (n=20), and SCLC (n=19). HnRNPA2B1 was significantly increased in all lung cancer types compared to normal, where it was barely detectable. The results between NSCLC and SCLC overall were relatively similar with the median values for SCLC and SCC at 54.2pg/ml and 53.4pg/ml respectively. Median value for normal controls was very low in comparison (1.6pg/ml) and interestingly hnRNPA2B1 was not detected at all in the NHBE cell line conditioned media.

A few high results can obviously increase the average significantly, but by including a box and whisker plot alongside the bar chart that indicates q1, min, median, max, and q3 values then a more true understanding of the results can be deciphered. With such small sample sets being used it was not ideal to use strict cut-off points and by including both graphs to display data there is little chance of ambiguity when interpreting the trend. Figure 3.4.2 shows data distribution of every sample, indicating hnRNPA2B1 levels vary considerably in lung cancer serum.

ROC curve analysis was performed for statistically significant comparisons (Fig. 3.4.3), with normal versus SCC showing the best trade-off between sensitivity and specificity, 85% and 80% respectively. Even on its own hnRNPA2B1 showed potential as a lung cancer biomarker.

In the second ELISA (Fig. 3.4.4) hnRNPA2B1 is again significantly elevated in SCC, AD, and SCLC compared to normal, however the normal serum samples showed higher levels than those from the first ELISA. In both ELISAs the SCC group showed the highest levels of hnRNPA2B1. Overall the second ELISA (Fig.

3.4.4) shows more consistency in results and perhaps can be taken with more merit than the first ELISA (Fig. 3.4.1). It can be concluded that hnRNPA2B1 levels are increased in late-stage lung cancer serum compared to normal serum. The samples used in the first assay are actually identical to the samples used in the second assay so it can be assumed the ELISA kits themselves showed varying degrees of specificity and sensitivity perhaps due to batch-to-batch variation, or insufficient QC.

A plasma ELISA was performed on hnRNPA2B1 (Fig. 3.4.5) comparing normal, benign lung disease, NSCLC, and SCLC serum, though the number of samples for each group varied considerably (n=30, n=39, n=60, and n=14 respectively). Benign lung disease and SCLC show almost identical averages (74.1pg/ml and 74.6pg/ml respectively) and the higher average in NSCLC (114.6pg/ml) could be biased by the extra samples, especially in its comparison with SCLC. The average concentration in normal plasma is significantly higher (210ng/ml), its results statistically significant when compared to the other three groups. Interestingly, the benign versus NSCLC t-test was also significant but the benign versus SCLC was not (though of course this could be attributed to different sample sizes).

These samples were earlier stage lung cancer compared to the serum samples in Fig. 3.4.1 and Fig. 3.4.4 but levels of normal plasma and that of normal serum in Fig. 3.4.4 should be comparable and there are higher concentrations of hnRNPA2B1 in normal plasma, a ~2.5 fold increase. In cancer serum the concentrations of NSCLC and SCLC are higher compared to plasma but the concentration in plasma could rise in late-stage lung cancer to similar levels in serum of late-stage.

An ELISA was performed to investigate if the hnRNPA2B1 levels in early-stage

cancer serum are higher than those in early-stage cancer plasma (Fig. 3.4.6). The samples used were matching plasma and serum samples from the same patients. Interestingly, the levels of hnRNPA2B1 were higher in plasma than serum of early-stage lung cancer patients. The average levels of hnRNPA2B1 in both benign lung disease plasma and lung cancer plasma are double that in lung cancer serum, indeed benign lung disease serum shows higher levels too. Because many of these samples were used in the plasma ELISA (Fig. 3.4.5) then, as in that assay, it can be assumed that hnRNPA2B1 levels in normal plasma are higher than two plasma controls and hence higher than all four conditions in Fig. 3.4.6. This data further confirms that serum and plasma can not be treated as being the same biological fluid and assays must be run on both before any conclusions can be drawn. If, as expected, hnRNPA2B1 levels are lower than normal in early-stage lung cancer serum and far higher in late-stage lung cancer then this biomarker could be very beneficial for drug treatment monitoring. Ideally an assay would be run with many lung cancer samples diagnosed as different stages to elucidate any trend that relates tumour stage to hnRNPA2B1 serum levels. Also an assay should be run comparing hnRNPA2B1 levels in late-stage lung cancer to those in earlier-stage lung cancer, performed in the same run on the same plate. Normal and benign lung disease control groups would be essential also.

Immunohistochemistry was performed on hnRNPA2B1 on two lung cancer tissue slides and also a mixed cancer tissue slide. hnRNPA2B1 expression was examined in tissue slides on a total of 100 cores; 14 SCC, 32 AD, 28 SCLC, 13 ACC, and 13 normal lung tissue cores. The highest expression was seen in the SCLC tissue cores with a total of 50% of the cores either strong staining (42.86%) or strong staining

with variance (7.14%). A further 20% was intermediate staining or intermediate staining with variance and only 10% showed negative staining. SCC showed strong staining too (~36% strong or strong with variance). AD showed less staining than SCC and SCLC. There was significantly less staining in the normal lung tissue with over 61% of cores showing negative staining. Only 7.14% of SCC showed negative staining though this was higher in SCLC (10.71%) and AD (25%). hnRNPA2B1 certainly has potential as a tissue marker.

Immunohistochemistry was performed on a mixed tissue slide and scored (Table 3.5.2). This slide contained colon adenocarcinoma, breast invasive ductal carcinoma, prostate adenocarcinoma, pancreas duct adenocarcinoma, and also lung SCC and AD cores. There was also normal colon, breast, prostate, pancreas and lung tissue, as well as cancer tissue adjacent to normal for colon and prostate. This was necessary to examine if hnRNPA2B1 tissue staining was specific to lung cancer.

Table 3.5.2 shows that hnRNPA2B1 staining is present in all cancer types. Though intensity 3 is more likely in lung cancer it is still quite common to the other four cancers. However, though all the cancer tissues showed relatively similar levels of staining, the normal tissue did not. Interestingly there is far more negative staining in normal lung tissue compared to the other normal tissue, indeed some normal tissue showed quite intense staining; e.g. 75% of normal colon was intensity 1. On the other hand 80% of all normal lung tissue showed negative staining. This shows the potential of hnRNPA2B1 as a tissue marker of lung tumour mass and subsequently treatment monitoring.

Functional assays were performed for hnRNPA2B1. Invasion and migration assays were the assays of choice. By using small interfering RNA that 'interfere' with the expression of specific genes, the implications of effectively knocking down or reducing expression of a protein of interest can be examined in an assay of choice. As mentioned in Section 1.3 invasion and migration of cancer cells allow them to enter the blood or lymphatic system to grow at distant locations to the primary tumour in a process called metastasis. By 'knocking-down' expression of hnRNPA2B1 its role, if any, in migration and/or invasion of cancer cells can be elucidated.

The effectiveness of the knockdown, performed using two separate siRNAs, was examined using western blot (Fig. 3.6.1). The expression of hnRNPA2B1 was effectively knocked down using the optimised conditions.

Cells were seeded at $0.5x10^5$ in all wells for all conditions (i.e. Untreated, NeoFx only, kinesin, scrambled siRNA, siRNA#1, siRNA#2). The effect of knocking down hnRNPA2B1 will be calculated by cell counts e.g.for invasion by counting the number of cells that pass through the matrix, however if the knockdown effects the rate at which the cancer cells grow (i.e. proliferation) then the differences in cell counts compared to the controls can not solely be attributed to an effect on the invasion and/or migration potential. A proliferation assay performed on the same cells under all conditions is used to examine if the knockdown had any effect on proliferation. Figure 3.6.2 shows there was negligible effect on proliferation caused by hnRNPA2B1 knockdown.

The knockdown of hnRNPA2B1 had an effect on invasion and migration, having a bigger effect on the former. The knockdown was performed in the SCC cell line

DLKP-M. The effect can be measured by calculating percentage difference in cell counts between the scrambled control and both siRNA#1 and siRNA#2. For both invasion and migration both siRNAs had an effect. This was a transient knockdown, meaning the effect on expression is only temporary and not absolute. With this decrease in hnRNPA2B1 expression there was a 57% and a 44% knockdown in invasion for siRNA#1 and siRNA#2 respectively. For siRNA#1 and siRNA#2 there was a 32% and 39% decrease in migration. It can be concluded that hnRNPA2B1 expression is necessary for effective invasion and migration of SCC cancer cells. Interestingly it appears to play a more important role in the ability of SCC cancer cells to pass through surrounding tissue than the actual movement of the cells.

A number of functions of hnRNPA2B1 may explain why it has such a strong effect on invasion of SCC cells. The invasion of cancer cells is known to be initiated and maintained by signalling pathways that control cytoskeletal dynamics in tumour cells and the turnover of cell-matrix and cell-cell junctions, followed by cell migration into the adjacent tissue (Friedl and Alexander), and is associated with disorganisation and destruction of basement membranes (Lissitzky, Pourreau-Schneider et al. 1985), so the role of hnRNPs in regulating gene expression, in cell signalling, and in preventing folding of pre-mRNA into secondary structures that inhibit its interactions with other proteins could all be important for efficient invasion of cancer cells. Also their role in coordinating post-translational events such as cellular localisation could be very important for the ability of cancer cells to invade adjacent tissue and subsequently metastasise to other locations.

Logistic regression performed on hnRNPA2B1, PKM2, and Hip ELISAs (Fig.

3.4.12) show their potential as biomarkers of lung cancer. The results here are for normal versus lung cancer. However, hnRNPA2B1 and PKM2 combined appear to work just as well as the three markers together. It should be noted that in the classification table it shows the correct identification of patients with cancer at 90% for the combination of three and 89.83% for the combination of two, while the correct identification of patients that are healthy is 50% and 45% respectively. It is more important to correctly identify patients with cancer as having cancer than it is to lessen the chances of identifying healthy patients as having cancer when they don't.

9.4 2D-DIGE Analysis of Immunodepleted Serum

As already mentioned the analysis of conditioned media from lung cancer cell lines grown in culture can only be used for discovery phase and it is the subsequent validation of any interesting proteins that will decide their potential as lung cancer biomarkers. Validation of conditioned media targets was performed in serum and it was found that for the most part the differences seen in serum between normal and lung cancer did not correspond to the fold changes extrapolated by DeCyder for the 2D-DIGE experiments. Although conditioned media as a starting point showed potential, it was proposed that using serum in the discovery phase would increase the chances of targets subsequently validating in larger serum sets. As the 2D-DIGE experiments on conditioned media showed many potential biomarkers it was decided to use this technique again.

Due to the complexity of serum (covered in greater detail in the Introduction and Section 9.1) the many proteins present in high abundance can mask the more

interesting disease-specific low abundant proteins. Therefore before serum could be used in a 2D-DIGE experiment some sort of fractionation was needed. Immunodepletion columns are designed to remove ~99% of all protein content from serum, in this case using antibodies for the 22 most abundant serum proteins. The initial experiment compared six normal and six adenocarcinoma serum samples. Instead of DeCyder for analysis new software known as Progenesis SameSpots was used. Progenesis SameSpots allows greater manipulation of results than the DeCyder software.

This experiment failed to go as planned. After mass spectrometry identification of interesting spots, disease-specific low abundant proteins were not identified. Instead only high adundant proteins and their isoforms (See Table 4.2.1) were identified. This suggests the immunodepletion column could not remove all of the high abundant proteins and its antibodies may have become saturated. Although immunodepletion decreased overall protein content of the serum samples it did little to decrease the complexity of the sample and the small abundant proteins could not be detected. However, the results from Table 4.2.1 led to an unforeseen discovery.

Table 4.2.1 showed multiple identifications for mutiple proteins. There were two IDs for zinc-alpha-2-glycoprotein, four IDs for alpha-1-antitrypsin, leucine-rich alpha-2-glycoprotein, and transferrin, and six IDs for haptoglobin. The assumption was made that these represented isoforms of the same protein. Haptoglobin proved to be the most interesting because each of the six isoforms in adenocarcinoma showed different fold changes compared to their compatriots in normal serum, as well as showing the most significant T-Test scores. Interestingly the highest fold change was seen in the isoform of highest molecular weight, the next highest fold

change in the isoform of next highest molecular weight and so on for the six isoforms (See Fig. 4.3.1). It was concluded that the six isoforms of haptoglobin could serve as a potential panel of biomarkers but further experiments were needed.

9.4.1 Haptoglobin

Haptoglobin is an acute phase protein (APP). Acute phase response is a systemic reaction of an organism to local or systemic disturbances in its homeostatis caused by infection, tissue injury, trauma or surgery, neoplastic growth or immunological disorders. During this response protein synthesis by the liver is drastically altered resulting in an increase of some blood proteins, known as the acute phase proteins (Gruys, Toussaint et al. 2005). APPs are plasma proteins involved in the restoration of homeostasis. The plasma level of some proteins, albumins for example, is decreased during the acute phase response. These are known as negative acute phase proteins. Positive acute phase proteins are those, like haptoglobin for example, where their plasma levels increase during the acute phase response.

Haptoglobin (Hp) is a plasma glycoprotein composed of α^1, α^2 and ß polypeptide chains (around 9, 19 and 40kDa respectively), the three main phenotypes in humans designated Hp1-1, Hp2-2, and Hp2-1. Its main function is to bind haemoglobin (Hb) and prevent oxidative stress (Abdullah, Schultz et al. 2009). Hp binds free haemoglobin to protect against oxidative stress and to facilitate the uptake of Hb by Hb scavenger receptors (Kristiansen, Graversen et al. 2001), though some animals only synthesise haptoglobin in response to stress (Dobryszycka 1997).

Glycosylation forms the sugar-amino acid linkage in the biosynthesis of the

carbohydrate units of glycoproteins and leads to a complex series of posttranslational enzymatic steps that lead to the formation of many protein-bound oligosaccharides with diverse biological functions (Spiro 2002). It is an enzymatic process that leads to the addition of glycans, either polysaccharides or oligosaccharides, to proteins, hence modifying the protein in some way. Glycosylation of proteins is an important post-translational modification that enhances the functional diversity of the proteins and influences their biological activity and each glycoprotein, such as haptoglobin, will have a mixture of glycosylated variants known as glycoforms (Marino, Bones et al.).

A higher circulatory level of α haptoglobin has been reported in SCLC patients. Higher serum levels of α and ß haptoglobin isoforms have been observed in SCLC patient samples and also a ß chain variant (5kDa smaller than the full length) has been shown to be differentially expressed in SCLC serum (Shah, Singh et al.).

Immunodepletion of serum followed by 2D-DIGE analysis is an expensive and time-consuming technique. For routine diagnostic tests of patient serum samples it is not a viable option. An alternative cheaper and less time-consuming method of detecting haptoglobin isoforms in serum was proposed. This involved running diluted raw serum directly on 2D gels and staining with coomaisse blue. Before this however a glycostain was run on a 2D gel to confirm the haptoglobin isoforms were in fact glycosylated. In Fig. 4.3.1 the six haptoglobin isoforms are highlighted in red, the stain confirming they are glycosylated. In the literature one study analysed which chain of Hp is closely related to lung cancers and found the Hp ß chain showed the most potential as a serum biomarker for lung cancer, it being increased 4-fold in 190 lung adenocarcinoma patients versus 190 healthy controls while

showing levels in plasma of patients with other respiratory disease such as TB, idiopathic pulmonary fibrosis, and bronchial asthma were closer to those of healthy controls (Kang, Sung et al. 2011).

Our experiments hoped to use the fold changes of the six isoforms or indeed the ratio differences between some of the isoforms as lung cancer biomarkers. Table 4.2.1 showed for adenocarcinoma compared to normal serum fold changes, six isoforms from isoform with highest molecular weight to isoform with lowest molecular weight, of +7.74, +6.49, +6.00, +5.40, +4.59, and +3.73 respectively. This shows a decrease in fold change in tandem with a decrease in molecular weight for haptoglobin. Comparing isoform 1 to isoform 6 there is a ratio difference of ~2fold.

When running diluted raw serum, as well as examining lung cancer, colon cancer and breast cancer were also examined to investigate if the haptoglobin isoform pattern was cancer specific. In total forty-eight gels were stained (eight gels each for normal, SCC, AD, SCLC, colon cancer and breast cancer), scanned with the Typhoon Imaging System, and the isoforms quantified using Progenesis SameSpots Software. Table 4.3.1 shows the fold change for each haptoglobin isoform for each of the cancer types. The fold change seen from the immunodepleted serum was not repeated, with a lot less variance seen between each isoform. For AD for example there is a 1.4fold increase for isoform 1 and also a 1.4 fold increase for isoform 6, and very little difference (either 1.5 or 1.6 fold) for the other isoforms. SCC does show ~2-fold increase from isoform 1 to isoform 6 but shows such identical fold changes for isoforms 1-3 that no pattern from isoform of highest MW to lowest MW is discernible. The ~2fold increase for SCC isoforms 1-3 do however show

good p values so perhaps these three isoforms could be used for treatment monitoring of SCC patients.

The fold changes for each isoform were relatively similar between AD, SCLC, and breast and colon cancer. AD and colon cancer for example show almost identical fold changes for each of the isoforms. The isoform pattern for these four cancers shows no specificity.

There is no correlation between the haptoglobin isoform results from the 2D-DIGE immunodepleted serum experiment comparing AD and Normal (Table 4.2.1) and the same comparison in the subsequent Coomaisse experiments (Table 4.3.1). One explanation for this could be that the immunodepletion column exaggerated the results seen in the initial experiments. The adenocarcinoma serum samples may have saturated the haptoglobin binding sites in the column resulting in less percentage of total haptoglobin being removed in these samples compared to the normal samples, which could change a 2-fold increase to a 7-fold increase as seen for isoform 1.

Another reason for such a difference in results could be the use of 2D-DIGE on the immunodepleted serum compared to using Coomassie blue to stain the proteins in the raw serum. 2D-DIGE uses an internal standard to normalise all gels in an experiment. By pooling various samples together, labelling with its own fluorescent tag (generally Cy2), and including it in every gel, the DeCyder Software can normalise all spots using a Master gel (generally the gel with the most spots). This greatly lessens the effect of gel-to-gel variation and means the abundance levels of each protein spot in a gel can be taken with confidence. Without an internal standard when comparing gels to one another as is the case when using Coomassie

stain, it is impossible to know how much of the variance seen can be attributed to gel-to-gel variation.

Despite the disappointing results the use of haptoglobin isoforms as lung cancer biomarkers certainly has potential. Using abundance levels of specific haptoglobin isoforms could allow for improved classification of lung cancer subtype during diagnosis or perhaps as an indicator that tumour bulk is decreasing after treatment. Perhaps a different serum/plasma fractionation technique would allow for improved Hp isoform detection and result in a more accurate and consistent reading of the changes in the individual isoforms in cancer. Overall Hp levels would not be specific enough for early detection due to the fact it is a positive acute phase protein and its levels are raised in many disease states but evaluating a number of specific haptoglobin isoforms could improve specificity.

Recent studies have indicated a possible correlation between specific haptoglobin glycosylation and particular disease conditions and found differentially expressed haptoglobin isoforms in sera of SCLC versus normal and sera of prostate cancer versus benign prostate or normal subjects (Fujimura, Shinohara et al. 2008, Shah, Singh et al. 2010). Haptoglobin ß chain has been put forward as being potentially useful as a supportive biomarker for lung cancers (Kang, Sung et al. 2011).

An ELISA was performed in serum for haptoglobin comparing normal to SCC, AD, and SCLC (Fig. 4.3.2). Increased levels of Hp in serum has been seen for many cancers (Tabassum, Reddy et al. 2012) and it has been suggested that circulating Hp levels could be used as a prognostic indicator in patients with epithelial ovarian cancer as lower Hp levels in preoperative sera were associated with better survival outcome (Seal, Doe et al. 1978, Ye, Cramer et al. 2003, Zhao, Annamalai et al.

2007). Our ELISA results were as expected, with Hp increased in the sera of lung cancer types SCC, AD, and SCLC compared to normal control serum. All t-test scores for normal compared to a particular lung cancer type were all less than 0.01. Haptoglobin levels are vastly increased in lung cancer patients compared to normal, healthy patients. Measuring overall levels of haptoglobin could be used as a general biomarker for treatment/recovery response but lacks the specificity to be an early detection marker for a particular cancer due to its raised levels in sera of many different cancers and diseases.

9.5 Label-free Mass Spectrometry of ProteoMiner Serum

Unlike immunodepletion, the ProteoMiner method does not simply attempt to remove the high abundant proteins but instead tries to equalise all proteins in serum/plasma to relatively equal quantities so that no one protein or group of proteins in particular can mask the other proteins, while at the same time preventing the loss of proteins bound to high abundant proteins as is the case with immunodepletion (See Introduction and Materials and Methods). It effectively dilutes high abundant proteins while simultaneously concentrating low abundant proteins with the goal of making these low abundant proteins easier to detect.

To analyse the ProteoMiner serum, label-free mass spectrometry was used. Label-free spectrometry involves measurements of changes in chromatographic ion intensity such as peptide peak areas or peak heights, and also spectral counting of identified proteins (See Section 1.9.3.3). Shotgun isotope labelling methods label by using light and heavy stable isotopes, before combining control and cancer samples and analysing by LC-MS/MS, the subsequent quantification calculation based on

the intensity ratio of isotope-labelled peptide pairs. However in label-free mass spectrometry control and sample are subjected to individual LC-MS/MS analysis and quantification is based on the comparison of peak intensity of the same peptide or the spectral count of the same protein (Zhu, Smith et al.).

Progenesis LC-MS software allows statistical analysis of the label-free data and generates lists of any conceivable comparisons, providing details on protein description, fold change, ANOVA value, confidence scores and number of peptides identified (See Section 5.1).

ProteoMiner followed by label-free MS was performed on ten serum samples each for the following groups; normal, SCC, AD, and SCLC. Tables were generated for seven different comparisons (Table 5.2.1 to Table 5.2.7) and unfortunately there were very few novel low abundant proteins identified, the main proteins being precursors for high abundant proteins such as plasminogen, haptoglobin, apolipoprotein, prothrombin, vitronectin, and fibrinogen, as well as members of the albumin family (e.g. afamin) and proteins involved in blood clotting (e.g. antithrombin III). In total 111 identifications were made through Mascot, though this included many duplicate or similar identifications (e.g. apolipoprotein A, apolipoprotein A-IV, apolipoprotein A-V, apolipoprotein B and apolipoprotein L were all identified).

After the conditioned media study in which some of the proteins chosen to be validated in serum did not validate at all, showed opposite trends, or vastly different fold changes, compared to the 2D-DIGE Decyder analysis results, a different approach was decided upon here. Instead of automatically choosing the proteins with the highest fold changes, it was decided to use a Pathway Studio tool known as

PathwayAssist, which comes with a database of molecular networks automatically assembled from scientific abstracts and contains more than 100,000 events of regulation, interaction, and modification between proteins, cell processes, and small molecules (Nikitin, Egorov et al. 2003). By inputting our list of proteins it was hoped some proteins would stand out as being central to a number of pathways, perhaps indicating their importance to the development of lung cancer. From the list of 111 proteins, only 65 were linked with lung cancer and brought forward to the next step. The 65 proteins were examined for any interactions between one another and it was found 24 of these proteins had no interactions with any of the other proteins, resulting in 41 proteins of interest (Table 5.3.1). PathwayAssist is based on the fact that genes (and proteins) do not work alone but behave in an intricate network of interactions. A pathway-base approach has been used to identify a smoking-associated 6-gene signature of lung cancer risk and survival (Guo and Wan 2012).

From the resulting interaction analysis (Fig. 5.3.1) which displayed the relationships between the biological molecules, two proteins were identified as 'hubs' of connectivity. This interaction diagram does not show how these proteins react together in a biological system but rather the inter-connection of these proteins based on the literature. The two proteins chosen to bring forward for validation were kininogen 1 and thrombospondin 1, not only 'hubs' of connectivity but they also showed reasonable fold change differences in the label-free analysis. For example kininogen-1 precursor showed a 3.5 fold increase in normal serum compared to lung cancer serum and a 4.4 fold increase in normal compared to AD serum, and thrombospondin-1 showed a 2.3 fold increase in normal compared to

lung cancer serum. Peptide abundance levels recorded from label-free MS identification can be seen for thrombospondin normal versus lung cancer comparison (Fig. 5.2.1) and for kininogen-1 normal versus AD comparison (Fig. 5.2.3).

An ELISA was performed in serum for both thrombospondin-1 (Fig. 5.3.3) and kininogen-1 (Fig. 5.3.4). Both assays were performed on 20 normal, 13 SCC, 18 AD, and 17 SCLC patient samples. For both ELISAs there was little to no difference in the levels of the protein in serum of healthy patients compared to the serum of SCC, AD, and SCLC. As regards being potential lung cancer biomarkers there is nothing significant or noteworthy about the ELISAs for thrombospondin-1 and kininogen-1 and here our interest in these two proteins ended. Pathway Studio analysis clearly has its benefits but is perhaps not ideal when looking for cancer biomarkers that will need subsequent validation in a biological fluid. An importance in the progression and/or cause of lung cancer does not automatically translate into a good serum marker.

As mentioned already the label-free analysis of ProteoMiner serum threw up over one hundred identifications. From this list there are many other proteins that could be brought forward to be validated in serum without the need for any further analysis (e.g. Pathway Studio). For example heat-shock protein 86 showed a 20-fold increase in lung cancer versus normal, s100-A9 showed ~5-fold increase in lung cancer versus normal, vimentin showed a large increase (recorded as ‘infinity’) in lung cancer versus normal, and histone H4 showed a 8.4-fold increase in AD versus normal.

It is difficult from these experiments to determine how useful ProteoMiner and label-free MS are for the discovery of lung cancer biomarkers as one technique is very much dependent on the other. If the ProteoMiner technique does not equalise serum protein content and many high abundant proteins still mask the low abundant proteins then there is nothing the LC/MS system can do to identify these proteins but if the ProteoMiner technique worked then perhaps the label-free MS technique was not sensitive enough to identify these proteins. Most likely the high abundant proteins in the serum prevented the LC/MS system from identifying the disease-specific low abundant proteins and although MS instruments are now sensitive enough to identify the lowest abundant proteins the proper fractionation of serum remains the problem.

9.6 Luminex Multiplex Platform for Lung Cancer Biomarker Discovery

Multiplex platforms allow analyse of multiple analytes using the same serum sample. Single-plex ELISAs that examine serum/plasma levels of particular proteins are expensive and not only examine one protein at a time but also generally use 50-100ul of serum. With a dependency on the hospitals to provide enough serum samples in large enough volumes it is beneficial to be able to use small amounts of serum for large sets of proteins. For example our initial discovery phase on the Luminex system that examined 12 proteins in Cancer Panel 1 (Section 6.1), only 5ul of serum was required per sample. For 12 single-plex ELISAs the volume required could be as high as 1.2ml per sample, a 240-fold increase. Although the Luminex system is expensive, purchasing panels of biomarkers that can be examined together in the one experiment will be cost-efficient in the long run.

There is however a restriction on the analytes that be examined, based upon catalogues from the manufacturer, and only certain biomarkers can be examined together in the same panel. In previous experiments the discovery phase was performed using a particular proteomic method (e.g. 2D-DIGE, label-free MS) and interesting proteins validated in ELISAs, however both the discovery phase and validation phase are performed in an identical fashion using the same method, in this case utilising the Luminex platform. For discovery phase three panels were chosen based on their likelihood to be differentially expressed in cancer (namely Cancer Panel 1, Cancer Panel 2, and an MMP panel) and a small selection of lung cancer serum samples were ran alongside breast cancer and colon cancer samples for a colleague's discovery phase.

From these panels, a selection of proteins was chosen to be examined in a larger set of lung cancer serum samples. Cancer Panel 1 was treated differently. Cancer Panel 1 consists of six proteins; alpha-fetoprotein, CA 19-9, CEA, CA-125, CA 15-3, and prolactin, which are known cancer biomarkers. In particular CA 19-9, CEA, CA-125, and CA 15-3 are mentioned as possible lung cancer biomarkers (Section 1.5.1) but all six markers from Cancer Panel 1 are not specific to lung cancer (Section 6.4). Due to their use already as cancer biomarkers it was useful to examine the levels of each in lung cancer (both serum and plasma) for comparative purposes.

CA 19-9, CEA, CA-125, and CA 15-3 were examined in late-stage lung cancer serum and early-stage lung cancer plasma. CA 19-9 levels were increased in late-stage SCC, AD, and SCLC compared to normal with all three comparisons statistically significant, and increased early-stage lung cancer plasma compared to

normal plasma. Although levels of CA 19-9 were higher in lung cancer plasma than BLD plasma, the comparison was not statistically significant (T-test; 0.29).

CEA levels in late-stage lung cancer serum were higher in both NSCLC groups (SCC and AD) compared to the SCLC group. CEA levels in all three were raised compared to normal and although in SCLC this difference was less pronounced the T-test score was relatively significant (Normal versus SCLC T-test; 0.06). In early-stage lung cancer plasma CEA levels were significantly increased in lung cancer compared to normal and BLD plasma. Indeed normal and benign showed similar levels.

CA-125 levels in late-stage serum were significantly higher in SCC, AD, and SCLC compared to normal, though also higher in both NSCLC types compared to SCLC.

In early-stage cancer plasma levels of CA-125 are higher in lung cancer compared to both normal and BLD. Normal and BLD show very similar levels.

CA-153 levels are increased in all three lung cancer types compared to normal, more so in AD and SCLC than SCC. In plasma, the highest CA-153 levels are in early-stage cancer compared to normal and BLD.

From Cancer Panel 2 six proteins were brought forward; epiregulin, heparin-binding epidermal growth factor, placenta growth factor, transforming growth factor alpha (TGF-α), vascular epithelial growth factor (VEGF-α), and tenascin C. In total 20 normal, 16 SCC, 21 AD, and 18 SCLC were examined. Epiregulin is a member of the epidermal growth factor family and can function as a ligand for EGFR. It showed relatively similar levels in normal and each of the lung cancer types SCC, AD, and SCLC, and no statistically significant comparisons (Fig. 6.5.1). It was not brought forward for further validation.

Heparin-binding epidermal growth factor is also a member of the EGF family that bind to and activate the EGF receptor and has been reported to be a potential serological biomarker for ovarian cancer (Kasai, Kobayashi et al.). The only statistically significant comparison was normal versus AD (Fig. 6.5.2). It was not brought forward for further validation.

Placenta growth factor is a member of the VEGF subfamily and is an important molecule in angiogenesis and vasculogenesis, particularly during embryogenesis. Interestingly it has been suggested that higher expression of plcenta growth factor in tissue can contribute to differences between the progression of SCLC and NSCLC, especially in regard to the nature of SCLC metastasis (Woo, Park et al. 2004). Though our results showed an increase in lung cancer serum overall there was a relatively broad spread of results for all conditions and none of the comparisons were statistically significant (Fig. 6.5.3) and this protein was not brought forward for further validation.

TGF-α is closely related to epidermal growth factor (which is known to stimulate cell growth, proliferation, and differentiation) and induces epithelial development and can bind to the EGF-receptor with similar effects to EGF. One study claims that TGF-α expression can not serve as an independent marker for lung cancer (Pecur, Kapitanovic et al. 1994) and our results (Fig. 6.5.4) showed it to have similar levels in SCC and SCLC compared to normal. TGF-α was not brought forward for further validation.

Tenascin C and VEGF- α were the most interesting biomarkers from Cancer Panel 2 and were further validated in serum and also plasma. They will be discussed in detail.

The extra-cellular matrix (ECM) is a network of secreted proteins and carbohydrates that fills the intercellular spaces, the spaces between cells. It is formed by macromolecules secreted locally by cells resident in the matrix, the two main classes being polysaccharide glycosaminoglycans and fibrous proteins. It helps to bind together and regulate a number of cellular functions such as adhesion, migration, proliferation, and differentiation (Teti 1992). The ECM is not an inert structure surrounding the cells but varies depending on the cell type; in epithelium and muscle tissue for example cells are bound tightly together so the spaces between cells containing the ECM are very narrow, whereas in connective tissue the ECM is very abundant, with cells distributed sparsely within it. The ability of cancer cells to migrate in the vascular stream and to recognise the ECM of a target organ plays a vital role in their ability to invade an organ and form metastases (Teti 1992). Because the ECM provides a microenvironment for cells and its destruction is associated with cancer cell invasion, any proteins involved in its make-up become possible markers for cancer.

9.6.1 Tenascin C

Tenascin-C is an extracellular matrix glycoprotein that is expressed upon damage to tissue, playing a host of different roles that mediate both inflammatory and fibrotic processes that help in effective repair of the damaged tissue (Midwood, Hussenet et al. 2011). Low levels of tenascin-C is detected in healthy adult tissues though it is known to be re-expressed upon tissue injury, and down-regulated after completion of tissue repair (Midwood and Orend 2009). It is expressed in the developing brain, cartilage, and mesenchyme and re-expressed in tumours, wound healing, and inflammation (Jones and Jones 2000). It also modulates cell migration, proliferation

and cellular signalling through induction of pro-inflammatory cytokines and oncogenic signalling molecules, and given the causal role of inflammation in the progression of cancer perhaps tenascin-C regulates some mechanism common to both (Midwood and Orend 2009). One study postulated that a peptide derived from tenascin-C, termed TNIIIA2, stimulates cell adhesion to ECM through activation of membrane receptor ß1-integrin (regulated activation of integrins is critical for cell adhesion, motility and tissue homeostasis (Rantala, Pouwels et al. 2011)) and could not only enhance cell proliferation but also induce apoptotic cell death (Iyoda and Fukai).

The first assay performed on a full set of serum samples (20 normal, 16 SCC, 21 AD, and 18 SCLC) showed a significant increase in each of the cancer types compared to normal (Fig. 6.5.9). All comparisons of normal versus lung cancer type were statistically significant with T-test scores less than 0.01 and average results for SCC, AD, and SCLC were either more than double or more than triple those for normal serum patients.

Another assay performed on a full set of serum samples (Fig. 6.5.10) was performed on 20 normal, 14 SCC, 21 AD, and 16 SCLC serum samples. The trend of being up in all lung cancer types compared to normal was repeated and again all comparisons of normal versus lung cancer type were statistically significant with T-test scores less than 0.01. The average results in pg/ml were higher than in the first assay (Fig. 6.5.9) for all groups, which could be attributed to the new sample set (though many of the samples overlap with the first assay) but more likely can be attributed to a different standard curve and a different run being performed on the Luminex instrument. The excellent T-test scores in both assays, as well as results showing

similar trends, suggests the multiplex Luminex platform is reasonably consistent in its readings. As in the first assay the average results are lower in AD than for the other two lung cancer types, however the median values are very similar. The minimum and maximum concentrations in AD are also lower than their counterparts in SCC and SCLC.

It can be concluded with certainty than levels of tenascin C are increased in the serum of lung cancer patients compared to normal healthy patients. There is however a need to perform the assay again using more control groups, in particular a benign lung disease control group is essential. Due to the fact it is expressed upon damage to tissue it is likely Tenascin C is elevated in serum of other cancers and inflammatory diseases but its potential as a lung cancer biomarker is evident from these results.

Another assay was performed on a full set of plasma samples (Fig. 6.5.11). The plasma samples are from patients with an earlier stage lung cancer to those investigated in the serum assays. The groups investigated were categorised as normal (n=18), benign lung disease (n=11), and lung cancer (n=35). Though none of the comparisons were statistically significant tenascin C is elevated in lung cancer compared to normal, however it is also elevated in benign lung disease compared to normal though not to the same extent.

Tenascin-C levels in serum has been suggested as a potential predictive marker of angiogenesis in NSCLC (Ishiwata, Takahashi et al. 2005), in endobronchial epitheilial-lining fluid (ELF) of NSCLC it was upregulated in cancer patients with adeoncarcinoma or squamous cell carcinoma (Kahn, Meister et al. 2012), its

expression in breast cancer cells has been linked to their ability to metastasise to the lungs (Oskarsson, Acharyya et al. 2011), and its degradation could be a reliable marker for recurrence potential of stage-1 NSCLC (Cai, Onoda et al. 2002).

9.6.2 Vascular Endothelium Growth Factor alpha (VEGF-α)

VEGF-α is a 34-42 kDa dimeric, disulfide-bound glycoprotein. The highest levels of VEGF-α mRNA are found in adult lung, heart, adrenal gland, and kidney. Lower, yet still detectable, levels occur in the liver and spleen. The actions of VEGF are known to affect lung development (Gerber, Hillan et al. 1999) and maintenance of adult lung structure (Tuder, Petrache et al. 2003).

VEGF has vascular permeability and endothelial cell growth functions. Its actions as a prototypic growth factor for endothelial cells are well known and well researched, and its effects on pulmonary endothelial cells are of particular importance in the pathobiology of lung diseases (Voelkel, Vandivier et al. 2006). It has a strong affect on several functions of endothelial cells, functions relevant to lung function and pulmonary vascular properties, such as prostacyclin and nitric oxide (NO) synthesis. Endothelial cell NO synthase is activated by VEGF, which in turn mediates the proangiogenic and permeability effects of VEGF (Dimmeler and Zeiher 1999). VEGF synthesis is stimulated by a wide range of growth factors such as transforming growth factor – α and – ß, fibroblast growth factor, insulin growth factor-I, keratinocyte growth factor and platelet-derived growth factor (PDGF). The strong link between inflammation and angiogenesis is illustrated by the fact platelet-activating factor production is stimulated by VEGF (Jackson, Seed et al. 1997).

Three assays with a full set of serum samples were performed on the Luminex system for **VEGF-α**, two for late-stage lung cancer and one for earlier stage lung cancer. The two late stage assays showed increases of VEGF-α level in SCC, AD, and SCLC compared to normal serum. In all cases the comparisons of cancer type versus normal were statistically significant with p values of less than 0.01. In both assays the average SCLC result is higher than both NSCLC types but not significantly so. The results for the three cancer types vary between the two assays but do remain 2-3 times more elevated compared to normal. It can be concluded that VEGF-α is consistently higher in lung cancer serum compared to normal.

The other serum assay (Fig. 6.5.7) was performed on earlier stage lung cancer and examined a normal group (n=16), an autoimmune group (n=16), a benign lung disease group (n=22), and a lung cancer group (n=26). Using a control group with only normal, healthy people is not a sufficient comparison for the lung cancer group. Lung cancer patients will often go to their doctors due to the presentation of symptoms that they would not necessarily associate with lung cancer, and often inflammatory diseases and indeed benign lung diseases can present with many of the same symptoms. Being able to separate lung cancer patients from those with other diseases is essential for initial diagnosis.

Interestingly VEGF-α is increased in lung cancer compared to all three controls, with each comparison being statistically significant. The average results of both the autoimmune and benign lung disease group are more than double the normal group, however the level in lung cancer is much higher again compared to normal. Due to insufficient diagnosis information the lung cancer category can not be divided into lung cancer type. It can be concluded that VEGF-α is elevated in patients with

early-stage lung cancer compared to healthy patients and also patients with a number of disease conditions such as diabetes and benign lung diseases.

High levels of VEGF in lung cancer serum has been associated with poor prognosis (Kaya, Ciledag et al. 2004). A novel VEGF receptor antagonist has been shown to inhibit LC-6 growth, a human NSCLC cell line (Ueda, Yamagishi et al. 2004). Bevacizumab is an angiogenesis inhibitor and slows the growth of new blood vessels; it is an antibody to VEGF and its addition to paclitaxel plus carboplatin improves survival when compared to chemotherapy alone in patients with untreated metastatic non-squamous NSCLC (Subramanian, Morgensztern et al. 2010). High serum concentrations of VEGF has been associated with unfavourable survival and poor response to treatment with SCLC patients (Salven, Ruotsalainen et al. 1998).

Another study showed that although VEGF serum levels was strongly associated with platelet and leukocyte counts in NSCLC patients it did not correlate with prognosis and tumour burden of the patients (Choi, Kim et al. 2001).

One study used ELISAs to assess the clinical usefulness of circulating angiogenic factors as a predictor for tumour angiogenesis in primary lung cancer and found a significant increase in VEGF concentrations in both serum and plasma of cancer patients compared to the healthy control. This study also showed a strong correlation between microvessel density and VEGF levels in plasma, indicating plasma VEGF could be a useful indicator of tumour angiogenesis in patients with lung cancer (Tamura, Ohta et al. 2001).

9.6.3 Matrix Metalloproteinases (MMPs)

From the MMP panel discovery phase that examined MMP's -1, -3, -8, -9, -10, and -13, the MMP's -1, -8, and -9 were brought forward for validation in larger samples sets. Due to availability of an MMP panel, assays with large sample sets were also performed on MMP's -2 and -7 despite the fact two of these MMP's (2 and 7) were not in the discovery phase and on MMP-10 despite the fact it was not initially brought forward from the discovery phase.

As a group MMPs degrade all protein components of tissue extracellular matrices and basement membranes. They form a family of 23 endopeptidases which can be subdivided into five groups depending on their domain structure and substrate specificity; collagenases, gelatinases, stromelysins, matrilysins and membrane-type MMPs, and are known to participate in various physiological and pathological processes such as embryonic development, wound healing, arthritis, atherosclerosis and tumour progression (Decock, Thirkettle et al.). In general, the structure of the MMPs includes a signal peptide, a propeptide domain, a catalytic domain, and a zinc-binding motif.

MMPs are involved in cell signalling and are capable of activating specific cell receptors and growth factors or liberating them from the extracellular matrix, thereby regulating various cell behaviours such as cell growth, differentiation, apoptosis, and migration (Overall and Lopez-Otin 2002, Hojilla, Wood et al. 2008). They are secreted as inactive zymogens (inactive enzyme precursors) and are converted into active enzymes by specific proteolytic cleavages on the cell surface or in the pericellular environment (area around the cell), providing spatial control of their function. They are inhibited by their endogenous inhibitors, the tissue

inhibitors of MMPs (TIMPs), which have four family members in humans with characteristic expression patterns and properties (Baker, Edwards et al. 2002).

MMPs are known to have roles in the late stage of tumour progression, invasion, and metastasis, and there is increasing evidence that they are also involved in earlier tumourigenesis such as in malignant transformation, angiogenesis, and tumour growth both at the primary and metastatic sites (Hua, Li et al.). Interactions between malignant and surrounding stromal cells (fibroblasts, endothelial cells, and inflammatory cell) may establish an environment in which tumour cells grow and metastasise and also the extracellular matrix (ECM) has a strong impact on tumour progression, the major components of which include structural proteins such as fibrillin, fibronectin, and laminin, as well as proteoglycans (Stetler-Stevenson, Aznavoorian et al. 1993).

Due to their critical roles in inflammation and tumourigenesis, MMPs appear to be ideal drug targets, the development of MMP inhibitors seen as being very attractive (Hu, Van den Steen et al. 2007), but although many have shown promising effects in preclinical studies, the results from clinical trials with these agents appear to be disappointing in cancer treatment (Hua, Li et al.).

MMP-1 and MMP-8 are collagenases, enzymes that break down collagen, the main component of connective tissue. MMP-9 is a gelatinase, enzymes that break down gelatine, a mixture of proteins and peptides produced by the hydrolysis of collagen.

Increased expression of certain MMPs has been observed in many cancers, this over-expression often associated with poor survival (Egeblad and Werb 2002). Increased MMP-1 expression for example has been associated with the incidence or invasiveness of various types of cancer, including colorectal, esophageal,

pancreatic, gastric, breast, and malignant melanoma (Liu, Kato et al.). Also MMP-1 expression in atypical ductal hyperplastic tissues may serve as a marker for predicting which patients will develop invasive breast cancer (Poola, DeWitty et al. 2005). Elevated serum levels of MMP-8 in colorectal cancer (CRC) patients has suggested a role for MMP-8 in the course and progression of this cancer (Vayrynen, Vornanen et al.). In one study MMP-9 expression increased in lung cancer with tumour size and its expression was significantly higher in NSCLC cases with metastasis compared to those without metastasis, and also in lymph node metastasis compared to primary lesions, showing MMP-9 may play an important role in tumour growth and metastasis of NSCLC (Zheng, Chang et al. 2010).

The three most interesting MMP's were -1, -8, and -9 and further assays were performed on these three proteins.

9.6.3.1 MMP-1

In the first serum assay (Fig. 6.6.1) performed on **MMP-1** normal (n=20) was compared to SCC (n=16), AD (n=21), and SCLC (n=18). The cancer samples were late-stage. There was a notable increase in the NSCLC types compared to normal, whereas in SCLC the levels were similar to the control. Both SCC and AD comparisons with normal were statistically significant. Concentrations of MMP-1 in normal and SCLC were 3.2ng/ml and 4.6ng/ml respectively, however in SCC and AD the levels were far higher, 11.2ng/ml and 8.9ng/ml respectively. Results for NSCLC were very different from SCLC, with this comparison showing a t-test score of 0.0357.

In the second serum assay (Fig. 6.6.2) again there was a large increase in the levels of MMP-1 in the NSCLC types SCC and AD compared to both normal and SCLC groups. In serum MMP-1 appears to be highly elevated in NSCLC but only slightly elevated in SCLC. A lung cancer marker takes on added significance if it can make a diagnosis between NSCLC and SCLC as their treatment method varies. For this assay the t-test score for NSCLC versus SCLC was 0.0377, indicating a result for this assay could be assigned as NSCLC or SCLC with a strong degree of confidence. This t-test score was very similar to that in the first MMP-1 serum assay for NSCLC versus SCLC. Again the comparisons of normal versus both SCC and AD are statistically significant with T-test scores less than 0.01.

A serum assay was performed in earlier stage lung cancer and compared to three controls; normal, autoimmune, and benign lung disease (Fig. 6.6.3). The earlier-stage lung cancer samples are not separated into NSCLC and SCLC due to lack of clinical data. MMP-1 is elevated in lung cancer compared to normal yet again but the difference is not as pronounced as seen in the late-stage cancer comparison. The autoimmune group showed similar levels to the normal group; however the benign lung disease group had MMP-levels similar to the lung cancer group. Often serum assays only compare normal to cancer and if this was the case here the result would appear to be more significant than it is. With the addition of a benign lung disease control group it gives a more accurate overview of the usefulness of MMP-1 levels as a biomarker for lung cancer. The average for benign lung disease and lung cancer were notably higher than the normal and autoimmune groups, however the box and whisker plot shows this difference is somewhat exaggerated by a few high results as the median values show, e.g. median for autoimmune is 11.6ng/ml and for benign

lung disease it is 14.9ng/ml. Many symptoms of benign lung disease are similar to symptoms of early onset of lung cancer; so because MMP-1 serum levels could not make a distinction between lung cancer and lung diseases it is more likely to be a marker of inflammation. A marker of general lung inflammation is no good for early detection of lung cancer.

Interestingly, when MMP-1 levels were investigated in plasma for earlier-stage lung cancer a different trend was seen (Fig. 6.6.4). The plasma assay compared normal, benign lung disease, and lung cancer. The levels of MMP-1 in plasma were increased in normal plasma compared to both benign lung disease and lung cancer. Indeed these comparisons were also statistically significant. Levels of MMP-1 benign lung disease and lung cancer were relatively similar and dramatically lower than those seen in normal plasma. Indeed the concentrations in plasma (though subject to variability when using a new kit and a performing a different run on the Luminex system) are far lower than those in serum, e.g. MMP-1 levels in normal serum in Fig. 6.6.3 is 12.9ng/ml but levels in normal plasma are only 0.36ng/ml. Serum and plasma clearly having varying protein/biomarker content and must be treated as two distinct biological fluids for biomarker discovery until both are examined for the levels of a biomarker of choice. The differences between serum and plasma will be discussed in greater detail later.

9.6.3.2 MMP-9

MMP-9 was validated in two serum assays and one plasma assay using a full set of samples. In the first serum assay (Fig. 6.6.5) MMP-9 is elevated in all cancer types compared to normal, the comparisons all statistically significant. Unlike the levels

of MMP-1 there is little difference between small cell and the two NSCLC types. There is such a big difference between the results for the cancers compared to normal that the minimum results for the cancer are the same as or higher than the median value for the normal e.g. minimum values for SCC, AD, and SCLC are 32.6ng/ml, 51.9ng/ml, and 61.6ng/ml respectively, and the median value for normal is only 33ng/ml.

In the second serum assay (Fig. 6.6.6) there is a dramatic increase in MMP-9 levels in the cancer types compared to normal again, and though not as pronounced as in the first serum assay (Fig. 6.6.5) the comparisons are still statistically significant. It is increased in NSCLC and SCLC, though again it is slightly up in the SCLC group. It can be concluded that in late-stage lung cancer serum MMP-9 is significantly elevated in lung cancer serum compared to normal serum. MMP-9 was not examined using an autoimmune or a benign lung disease control group so before any further conclusions can be drawn it would need to be repeated in an assay including more control groups, the benign lung disease group in particular.

MMP-9 was also examined in plasma samples (Fig. 6.6.7) in earlier stage lung cancer. It was decreased in lung cancer compared to normal, as was MMP-1; however it was also decreased in lung compared to benign lung disease too. Levels in normal and benign were relatively similar and double that of lung cancer. The normal versus lung cancer comparison was statistically significant and even the benign lung disease versus lung cancer comparison showed a significant T-Test of 0.08. An MMP-9 assay was not performed in serum for early-stage lung cancer but comparing the levels in the normal it can be seen that, as was the case with the MMP-1 assays, the level of MMP-9 in plasma is lower than the levels in serum as

can be seen with the normal control. MMP-9 in plasma can be considered a better biomarker than MMP-1 in plasma due to the fact levels in lung cancer differ to those in the benign lung disease group.

9.6.3.3 MMP-8

MMP-8 validation was not performed in an assay on the Luminex system but in a single-plex commercial ELISA, performed on normal, SCC, AD, and SCLC serum (Fig. 6.6.8). There is a significant increase in levels of MMP-8 in cancer serum compared to normal. Interestingly there was very little overlap in any of results from normal with those in any of the cancers; indeed the maximum value in normal was 4ng/ml which wasn't much higher than the minimum value in SCC (1.3ng/ml), AD (2.1ng/ml), and SCLC (3.7ng/ml). It can be concluded that MMP-8 levels in lung cancer serum far exceed those in normal serum.

With the results showing such large variance in cancer serum compared to normal a ROC curve analysis was performed for the MMP-8 ELISA results for normal versus lung cancer serum (Fig. 6.6.9). With sensitivity of 94.1% and specificity of 90.9% and an AUC value of 0.981 it further emphasises the disparity between MMP-8 levels in normal serum compared to lung cancer serum. Though MMP-8 shows potential as a lung cancer biomarker, again further assays are required that will include more control groups and also examine it in plasma.

The final three MMP serum assays will be examined together; those of **MMP-2** (Fig. 6.6.10), **MMP-7** (Fig. 6.6.11), and **MMP-10** (Fig. 6.6.12). MMP-2 levels in lung cancer are more similar to those in normal serum than are the autoimmune and benign lung disease group and for the MMP-10 assay the average levels are very

similar for normal, autoimmune, and lung cancer. The most statistically significant comparison is normal versus benign lung disease so it would appear from this assay that MMP-10 has more potential as an inflammatory marker than a lung cancer marker. MMP-7 showed similar levels in lung cancer serum and serum of benign lung disease both of which showed statistically higher levels compared to both the normal and autoimmune groups. MMP-7 showed greater potential as a lung cancer marker than either MMP-2 or MMP-10.

9.6.4 Biomarker Panel from Luminex Multiplex Platform

The Luminex system proved to be an effective method for biomarker discovery. It is simple to run and uses very little serum/plasma compared to commercial ELISAs. In assays that were repeated using many of the same samples the results were relatively consistent and the highly significant T-Test scores for the comparisons that showed variance indicates a consistency in the systems reading of the beads. By expanding the number of proteins available in their library for detection there is no reason this system will not remain relevant for years to come. It also has the added benefit of allowing discovery phase and validation phase to be performed in serum/plasma on the same system using the same method. By using serum/plasma in the initial discovery phase it can be determined straight away if your protein of interest is even detectable in these biological fluids and saves time and money on validation.

As already mentioned Cancer Panel 1 includes many biomarkers already in clinical use. Though they are not specific for lung cancer they are often used in lung cancer and are a useful comparison for the new biomarkers. Excluding AFP, Cancer Panel

1 validated in serum reasonably effectively. CEA for example (Fig. 6.4.4) showed higher levels in NSCLC (AD in particular) compared to normal and SCLC and also showed statistically higher levels in lung cancer plasma compared to both normal and benign lung disease plasma (Fig. 6.4.5). CA-125 also showed higher levels in cancer serum compared to normal, with higher levels in NSCLC than SCLC, and also validated well in plasma showing an increase in lung cancer compared to both control groups (benign lung disease group versus lung cancer had a T-Test of 0.07).

By performing logistic regression analysis on panels of three biomarkers from the Luminex results it was investigated in both serum and plasma if a combination of any three proteins would show improved sensitivity and/or specificity over a panel of three biomarkers from Cancer Panel 1. The logistic analysis combines the sensitivity and specificity results to give a percentage based on the classification of correct cases identified overall. For both serum and plasma the combination was examined for the known markers from Cancer Panel 1 and then any number of combinations (including those from Cancer Panel 1 and at least one from Cancer Panel 2 or the MMP panel) and the best comparisons are shown in Fig. 6.7.1 for serum and Fig. 6.7.2 for plasma. Only matching samples for the three biomarkers in each combination are used for calculating results.

Interestingly it was found that a better combination of biomarkers in both serum and plasma included at least two not found in Cancer Panel 1. In fact in serum the best combination was for three biomarkers outside this panel. In serum a combination from Cancer Panel 1 of CA-125, CA 19-9, and CEA from a sample set of 76 could correctly predict **86.84%** of cases with an AUC of 0.962 (Fig. 6.7.1 A), whereas a

combination of MMP-1, MMP-9, and tenascin C from a sample set of 64 could correctly predict **96.87%** of cases with an AUC of 0.992 (Fig. 6.7.1 B)

The best combination in plasma included one from Cancer Panel 1 and two from outside it; CA-125, CA 19-9, and CEA from a sample set of 52 could predict **80.77%** of cases with an AUC of 0.887 (Fig. 6.7.2 A), whereas a combination of MMP-1, CA-125, and Tenascin C from a sample set of 53 could predict **88.68%** of cases with an AUC of 0.932 (Fig. 6.7.2 B).

9.7 Metabolomic Serum Profiling for Lung Cancer Biomarker Discovery

As the metabolism of cancer cells is different from that of normal cells it is not surprising that metabolites themselves can be biomarkers of lung cancer. Metabolites are intermediates and products of metabolism and are often involved cell growth, development, and reproduction (See Sections 1.10 and 7.1). There are many assays available to quantify metabolites in biological fluids. The study of metabolites in these biological fluids involves a quite technical discovery phase. A diagnostic company called Metabolon provides a service that globally analyses complex biological samples for the discovery of metabolic markers.

From the resulting discovery phase performed on eight samples of normal, SCC, AD, and SCLC serum that produced over 300 metabolites of interest a number of the most interesting were brought forward for validation in larger serum and/or plasma sets. These metabolites were 12-HETE, phenylalanine, glutamate, and aspartate (See Figures 7.2.1 and 7.2.2 for Metabolon results).

9.7.1 12-HETE

12-HETE was examined in serum in three assays, two for late-stage and one for earlier stage lung cancer, and in plasma in one assay.

The lipoxygenase (LOX) pathway leads to the formation of leukotrienes and also catalyses the conversion of arachidonic acid (AA) to hydroperoxyeicosatetraenoic acids that are then reduced to hydroxyeicosatetraenoic acids (HETE) by glutathione peroxidase. Arachidonic acid is a polyunsaturated fatty acid regarded as structural components of cell membranes whose main function is to regulate membrane permeability. There are four mammalian LOXs that produce 5-, 8-, 12- and 15-HETE, respectively (5-LOX produces 5-HETE, 12-LOX produces 12-HETE etc), 12-HETE being generated when cytochrome P-450 isozymes metabolise AA by bis-allylic oxidation (lipoxygenase-like reaction). The production of LOX metabolites has been observed in several tissues and cells, including epithelial cells, vascular smooth muscle cells, endothelial cells, monocytes and leukocytes. HETEs have important physiological and pathological functions that modulate ion transport, renal and pulmonary functions, vascular tone and reactivity, and inflammatory and growth responses. HETEs have also been observed to have a role in apoptosis, angiogenesis, and the proliferation and metastasis of cancer cells (Moreno 2009). 12(S)-HETE is the predominant enantiomer and is mainly produced by platelets and leukocytes.

Platelets metabolise AA through the lipoxygenase pathway producing 12(S)-HETE (Moncada and Higgs 1986). After intravasation, tumour cells interact with platelets and induce platelet aggregation and also synthesis of platelet 12(S)-HETE, indicating platelets can be a contributor of 12(S)-HETE to facilitate tumour cell

interaction with endothelial cells and subendothelial matrix. It has been previously shown that exogenous 12(S)-HETE increases tumour cell adhesion to endothelium, induces large vessel and mircovessel endothelial retraction, and stimulates tumour cell spreading on matrix (Chen, Duniec et al. 1994). Pretreatment of tumour cells with exogenous 12(S)-HETE enhances their adhesion to and spreading on fibronectin, which is mediated exclusively by integrin αIIbß3 receptors and in a cytoskeleton-dependent manner (Timar, Chen et al. 1992), and this pretreatment also stimulates tumour cell motility and release of cathepsin B, a proteinase linked to tumour cell invasiveness (Chen, Duniec et al. 1994).

The first 12-HETE serum ELISA (Fig. 7.3.1) showed 12-HETE to be significantly increased in SCC, AD, and SCLC compared to normal. All these comparisons were statistically significant; indeed the T-test for normal versus AD was 0.00000218. There was a large spread of results in each of the lung cancer types; for example minimum and maximum values for SCC, AD, and SCLC were respectively 1.2ngml and 1867ng/ml, 8.6ng/ml and 1485ng/ml, and 3.1ng/ml and 1214ng/ml, however the maximum concentration in normal serum was only 15ng/ml. ROC Curve analysis of these results showed high sensitivity and specificity for the comparison of normal versus lung cancer, 86.7% and 88.9% respectively, and a high AUC value of 0.927. As a marker for lung cancer the potential of 12-HETE was clear.

Another ELISA was performed in late-stage cancer serum (Fig. 7.3.3) and again a clear trend is present. 12-HETE is elevated in SCC, AD, and SCLC significantly compared to normal serum. In this assay 12-HETE was highly elevated in a few of the normal samples but the median value is a good indicator of how the results lie overall and median values for normal, SCC, AD, and SCLC were 2.3ng/ml,

311ng/ml, 119ng/ml, and 261ng/ml respectively, indicating big differences in the spread of results. Indeed the minimum values for each of the cancer types were significantly greater than the median of normal. All cancers compared to normal are statistically significant.

The third serum ELISA (Fig. 7.3.4) was performed in early-stage lung cancer and included an autoimmune and a benign lung disease control as well as the normal control group. Interestingly the levels of 12-HETE in earlier stage lung cancer is far lower than those in the normal control. The average in normal serum is 241ng/ml (similar level is seen in autoimmune controls) and in lung cancer it is 67ng/ml. The benign lung cancer group shows lower levels than the other two controls but higher than lung cancer. Although not directly comparable it is interesting to note that average 12-HETE serum levels in the two assays in serum of late-stage lung cancer are very high compared to early-stage. 12-HETE therefore shows potential as a marker for lung cancer staging and also for lung cancer treatment monitoring.

In plasma (Fig. 7.3.5) 12-HETE levels are significantly lower in lung cancer plasma and also benign lung disease plasma compared to normal plasma. As this was performed in early-stage lung cancer it makes sense to compare these results to Fig. 7.3.4 as this serum assay was performed in early-stage lung cancer too. The average benign lung disease result of 21.5ng/ml is similar to the lung cancer average of 16.2ng/ml, both being significantly lower than average of 361ng/ml in normal plasma. The difference between early-stage cancer serum and normal serum is not as big. Ideally an ELISA would be performed in plasma of late-stage lung cancer

patients to see if the difference shown in late-stage and earlier stage cancer serum transfers to plasma too.

9.7.2 Phenylalanine

A serum assay and a plasma assay were performed on phenylalanine. Phenylalanine is the precursor of tyrosine, an important amino acid for the biosynthesis of neurotransmitters like L-DOPA (L-3,4-dihydroxyphenylalanine) and catecholamines dopamine, epinephrine, and norepinephrine. Catecholamines are hormones produced by the adrenal glands, found on top of the kidneys, and are released into blood during times of physical or emotional stress. For enzymatic hydroxylation of phenylalanine to tyrosine by phenylalanine-4-hydroxylase (PAH), the cofactor 5,6,7,8-tetrahydrobiopterin (BH4), the reduced form of biopterin, is needed as a hydrogen donator (Shintaku 2002). Levels of phenylalanine in the blood are increased in consuming diseases such as HIV-1 infection, trauma, sepsis, burn, and malignancy, and though this increase is unexplained the clinical conditions mentioned are known to be linked with inflammation and immune activation and with increased concentrations of immune activation markers such as serum soluble 74kDa tumour necrosis factor-α (sTNF-R75) and neopterin (Neurauter, Grahmann et al. 2008). Subnormal enzymatic conversion of phenylalanine by PAH might explain the accumulation of phenylalanine in patients (Shintaku 2002).

In parallel to neopterin and sTNF-R75 formation, macrophages and other cells stimulated with IFN-γ produce reactive oxygen species (ROS) at a high rate, which may wipe out antioxidant defence systems and therefore help in oxidative stress

development. BH4 is very sensitive to oxidation so if BH4 is destroyed the conversion rate of phenylalanine to tyrosine by PAH is reduced, which is reflected by an increased phenylalanine to tyrosine ratio (phe/tyr) (Widner, Fuchs et al. 2001). Oxidative stress due to chronic immune activation and inflammation therefore could be involved in the increase of serum phenylalanine concentrations in patients, and a correlation between phe/tyr and neopterin concentrations in patients with multiple traumas has been described (Ploder, Neurauter et al. 2008).

Depressive mood disorders may develop in cancer patients with increased phenylalanine concentrations, maybe due to an impaired hydroxylation of phenylalanine (Hoekstra, van den Broek et al. 2001). One study showed correlations between increased sTNF-R75 and neopterin concentrations, and phenylalanine, phe/tyr, and 8-isoprostane in patients with ovarian cancer (Neurauter, Grahmann et al. 2008).

The serum assay for phenylalanine was performed in late-stage lung cancer (Fig. 7.3.8). Levels of phenylalanine are decreased in NSCLC compared to both normal serum and SCLC serum. Interestingly the comparison of normal versus NSCLC and NSCLC versus SCLC are statistically significant. This shows potential of phenylalanine as a marker to distinguish NSCLC from SCLC and a ROC curve performed for NSCLC versus SCLC (Fig. 7.3.9) showing a sensitivity of 73.7% and specificity of 65.0% with an AUC of 0.886. This shows phenylalanine may be useful for distinguishing patients with NSCLC and those with SCLC.

The plasma assay (Fig. 7.3.10) was performed in earlier stage lung cancer and compared to benign lung disease and normal plasma. Normal versus benign lung disease and normal versus lung cancer results were statistically significant. Average

concentration for normal, benign lung disease, and lung cancer were 155pmol/ul, 112pmol/ul, and 122pmol/ul respectively. When lung cancer plasma is divided into NSCLC and SCLC the average results were 128pmol/ul for eighty-nine NSCLC samples and 118pmol/ul for only fourteen SCLC (data not shown), and though different to that seen in late-stage lung serum there are too many variables for a trustworthy comparison. In the serum assay there were forty NSCLC versus sixteen SCLC. It is difficult to compare serum and plasma in this instant as the number of samples is different and the stage of lung cancer is different. The most interesting result is that in serum of late-stage lung cancer levels of phenylalanine has potential to differentiate between NSCLC and SCLC.

9.8 Aushon Multiplex Immunoassay Platform For Lung Cancer Biomarker Discovery

The Aushon Cira immunoassay platform can be using for multiplex analysis of up to 12 analytes of interest. Unlike the bead based system where the user mixes and adds the beads before addition to plate, the analytes of interest are pre-spotted on the array. Our discovery phase examined nine chemokines; low-molecular weight type of cytokine classified by their ability to induce chemotaxis or chemokinesis. Cytokines are regulatory proteins released by cells of the immune system where they act as intercellular mediators for the generation of an immune response. Chemokines stimulate angiogenesis and tumour growth by recruiting tumour-associated macrophages.

A single discovery phase was performed using normal and SCLC serum samples only. Nine samples of each were used for comparison. No further validation was

performed though some of the chemokines showed potential as SCLC biomarkers. The results are in Section 8.2.

MIP1a and MCP1 showed increased levels in SCLC compared to normal yet the results were not statistically significant (p value of 0.08 and 0.07 respectively). Considering the small sample set used however, these results are of potential interest. Both of these chemokines are involved in the inflammatory response.
MIP1b, TARC, Eotaxin, and IL-8 showed increased levels in SCLC compared to normal, the comparisons all statistically significant. The roles of four chemokines are explained in Section 8.1. Of the nine chemokines examined only three showed no potential as potential SCLC markers.

As this was only discovery phase using small sample sets for one control and one lung cancer type there is little to be said about the six interesting chemokines other than they show some potential as SCLC serum markers. A large sample set comparing normal, benign lung disease, SCC, AD, and SCLC is required for each of the six potential markers before any conclusions can be drawn. The Aushon system worked extremely well; used very little sample, was easy to use, and was fast and reliable. As an alternative multiplex platform to Luminex it shows potential and with a large catalogue of human analytes e.g. angiogenesis factors, chemokines, cytokines, growth factors, and MMPs, it is a very useful platform for biomarker discovery in serum.

CHAPTER TEN

CONCLUSIONS & FUTURE WORK

CONCLUSIONS

1. hnRNPA2B1 is a possible lung cancer biomarker in tissue and serum/plasma and has a role in invasion and migration of lung cancer cells.

2. Pyruvate kinase M2, Hsc70-interacting protein, tenascin C, VEGF-α, MMP-1, MMP-8, MMP-9, 12-HETE, and phenylalanine show potential as lung cancer biomarkers.

3. Haptoglobin isoforms have potential to differentiate lung cancer from healthy patients.

4. A good biomarker in serum does not necessarily make a good biomarker in plasma.

5. For lung cancer biomarker validation a benign lung disease control group is essential.

6. Intensive validation of candidate biomarkers is essential before conclusions on its validity as a biomarker can be made.

FUTURE WORK

1. To examine interesting biomarkers in late-stage and early-stage lung cancer samples (both serum and plasma) on the same array.

2. To validate more proteins from the discovery phases in serum and plasma in a larger cohort of patients.

3. To compare the proteome of patients who responded favourably, and those who did not, to drug treatment.

4. To perform an ELISA on Hsc70-interacting protein in plasma samples.

5. To perform siRNA knockdown of hnRNPA2B1 in more lung cancer cell lines.

6. To perform siRNA knockdown of other candidate biomarkers in lung cancer cell lines to examine their effect on migration and invasion of lung cancer.

7. To perform assays on interesting chemokines from Auschon multiplex experiments on non-small cell lung cancer samples as well as benign lung disease samples.

CHAPTER ELEVEN

BIBLIOGRAPHY

Abdullah, M., H. Schultz, et al. (2009). "Expression of the acute phase protein haptoglobin in human lung cancer and tumor-free lung tissues." Pathol Res Pract **205**(9): 639-47.

Aberle, D. R., A. M. Adams, et al. "Reduced lung-cancer mortality with low-dose computed tomographic screening." N Engl J Med **365**(5): 395-409.

Adkins, J. N., S. M. Varnum, et al. (2002). "Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry." Mol Cell Proteomics **1**(12): 947-55.

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

Aggelis, V., R. A. Craven, et al. (2009). "Proteomic identification of differentially expressed plasma membrane proteins in renal cell carcinoma by stable isotope labelling of a von Hippel-Lindau transfectant cell line model." Proteomics **9**(8): 2118-30.

Ahrendt, S. A., J. T. Chow, et al. (2000). "Alcohol consumption and cigarette smoking increase the frequency of p53 mutations in non-small cell lung cancer." Cancer Res **60**(12): 3155-9.

Alban, A., S. O. David, et al. (2003). "A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard." Proteomics **3**(1): 36-44.

Anderson, L. and C. L. Hunter (2006). "Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins." Mol Cell Proteomics **5**(4): 573-88.

Anderson, N. L. and N. G. Anderson (2002). "The human plasma proteome: history, character, and diagnostic prospects." Mol Cell Proteomics **1**(11): 845-67.

Andjelkovic, T., J. Bankovic, et al. "Coalterations of p53 and PTEN tumor suppressor genes in non-small cell lung carcinoma patients." Transl Res **157**(1): 19-28.

Apweiler, R., C. Aslanidis, et al. (2009). "Approaching clinical proteomics: current state and future fields of application in cellular proteomics." Cytometry A **75**(10): 816-32.

Ardizzoni, A., M. A. Cafferata, et al. (2006). "Decline in serum carcinoembryonic antigen and cytokeratin 19 fragment during chemotherapy predicts objective response and survival in patients with advanced nonsmall cell lung cancer." Cancer **107**(12): 2842-9.

Arslan, N., M. Serdar, et al. (2000). "Use of CA15-3, CEA and prolactin for the primary diagnosis of breast cancer and correlation with the prognostic factors at the time of initial diagnosis." Ann Nucl Med **14**(5): 395-9.

Atzpodien, J., P. Royston, et al. (2003). "Metastatic renal carcinoma comprehensive prognostic system." Br J Cancer **88**(3): 348-53.

Auerbach, O., A. P. Stout, et al. (1961). "Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer." N Engl J Med **265**: 253-67.

Ayache, S., M. C. Panelli, et al. (2006). "Comparison of proteomic profiles of serum, plasma, and modified media supplements used for cell culture and expansion." J Transl Med **4**: 40.

Baker, A. H., D. R. Edwards, et al. (2002). "Metalloproteinase inhibitors: biological actions and therapeutic opportunities." J Cell Sci **115**(Pt 19): 3719-27.

Bakhtiar, R. and R. W. Nelson (2001). "Mass spectrometry of the proteome." Mol Pharmacol **60**(3): 405-15.

Bantscheff, M., M. Schirle, et al. (2007). "Quantitative mass spectrometry in proteomics: a critical review." Anal Bioanal Chem **389**(4): 1017-31.

Barlesi, F., C. Tchouhadjian, et al. (2005). "CYFRA 21-1 level predicts survival in non-small-cell lung cancer patients receiving gefitinib as third-line therapy." Br J Cancer **92**(1): 13-4.

Bast, R. C., Jr., F. J. Xu, et al. (1998). "CA 125: the past and the future." Int J Biol Markers **13**(4): 179-87.

Bathen, T. F., T. Engan, et al. (2000). "Analysis and classification of proton NMR spectra of lipoprotein fractions from healthy volunteers and patients with cancer or CHD." Anticancer Res **20**(4): 2393-408.

Belgrad, R., C. A. Good, et al. (1962). "Alveolar-cell carcinoma (terminal bronchiolar carcinoma). A study of surgically excised tumors with special emphasis on localized lesions." Radiology **79**: 789-98.

Bell, L. N., J. L. Theodorakis, et al. "Serum proteomics and biomarker discovery across the spectrum of nonalcoholic fatty liver disease." Hepatology **51**(1): 111-20.

Berkelman, T. (2008). "Quantitation of protein in samples prepared for 2-D electrophoresis." Methods Mol Biol **424**: 43-9.

Bharti, A., P. C. Ma, et al. (2007). "Biomarker discovery in lung cancer--promises and challenges of clinical proteomics." Mass Spectrom Rev **26**(3): 451-66.

Bijian, K., A. M. Mlynarek, et al. (2009). "Serum proteomic approach for the identification of serum biomarkers contributed by oral squamous cell carcinoma and host tissue microenvironment." J Proteome Res **8**(5): 2173-85.

Bluemlein, K., N. M. Gruning, et al. "No evidence for a shift in pyruvate kinase PKM1 to PKM2 expression during tumorigenesis." Oncotarget **2**(5): 393-400.

Bonk, T. and A. Humeny (2001). "MALDI-TOF-MS analysis of protein and DNA." Neuroscientist **7**(1): 6-12.

Boros, L. G., M. R. Lerner, et al. (2005). "[1,2-13C2]-D-glucose profiles of the serum, liver, pancreas, and DMBA-induced pancreatic tumors of rats." Pancreas **31**(4): 337-43.

Boschetti, E. and P. G. Righetti (2008). "The ProteoMiner in the proteomic arena: a non-depleting tool for discovering low-abundance species." J Proteomics **71**(3): 255-64.

Bostanci, N., W. Heywood, et al. "Application of label-free absolute quantitative proteomics in human gingival crevicular fluid by LC/MS E (gingival exudatome)." J Proteome Res **9**(5): 2191-9.

Bouchal, P., T. Roumeliotis, et al. (2009). "Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis." J Proteome Res **8**(1): 362-73.

Boukakis, G., M. Patrinou-Georgoula, et al. "Deregulated expression of hnRNP A/B proteins in human non-small cell lung cancer: parallel assessment of protein and mRNA levels in paired tumour/non-tumour tissues." BMC Cancer **10**: 434.

Bozcuk, H., U. Bilge, et al. (2004). "An application of a genetic algorithm in conjunction with other data mining methods for estimating outcome after hospitalization in cancer patients." Med Sci Monit **10**(6): CR246-51.

Bremnes, R. M., S. Sundstrom, et al. (2003). "The value of prognostic factors in small cell lung cancer: results from a randomised multicenter study with minimum 5 year follow-up." Lung Cancer **39**(3): 303-13.

Brett, D., H. Pospisil, et al. (2002). "Alternative splicing and genome complexity." Nat Genet **30**(1): 29-30.

Brizel, D. M., T. Schroeder, et al. (2001). "Elevated tumor lactate concentrations predict for an increased risk of metastases in head-and-neck cancer." Int J Radiat Oncol Biol Phys **51**(2): 349-53.

Buccheri, G. and D. Ferrigno (2004). "Lung cancer: clinical presentation and specialist referral time." Eur Respir J **24**(6): 898-904.

Burd, C. G. and G. Dreyfuss (1994). "Conserved structures and diversity of functions of RNA-binding proteins." Science **265**(5172): 615-21.

Burd, C. G., M. S. Swanson, et al. (1989). "Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: a diversity of RNA binding proteins is generated by small peptide inserts." Proc Natl Acad Sci U S A **86**(24): 9788-92.

Cai, M., K. Onoda, et al. (2002). "Degradation of tenascin-C and activity of matrix metalloproteinase-2 are associated with tumor recurrence in early stage non-small cell lung cancer." Clin Cancer Res **8**(4): 1152-6.

Cairns, R. A., I. S. Harris, et al. "Regulation of cancer cell metabolism." Nat Rev Cancer **11**(2): 85-95.

Canas, B., D. Lopez-Ferrer, et al. (2006). "Mass spectrometry technologies for proteomics." Brief Funct Genomic Proteomic **4**(4): 295-320.

Capelozzi, V. L. (2009). "Role of immunohistochemistry in the diagnosis of lung cancer." J Bras Pneumol **35**(4): 375-82.

Cappuzzo, F., F. R. Hirsch, et al. (2005). "Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer." J Natl Cancer Inst **97**(9): 643-55.

Carrola, J., C. M. Rocha, et al. (2011). "Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine." J Proteome Res **10**(1): 221-30.

Chambers, A. F., A. C. Groom, et al. (2002). "Dissemination and growth of cancer cells in metastatic sites." Nat Rev Cancer **2**(8): 563-72.

Chang, J. W., H. B. Jeon, et al. (2001). "Augmented expression of peroxiredoxin I in lung cancer." Biochem Biophys Res Commun **289**(2): 507-12.

Chang, J. W., S. H. Lee, et al. (2005). "Peroxiredoxin-I is an autoimmunogenic tumor antigen in non-small cell lung cancer." FEBS Lett **579**(13): 2873-7.

Chen, Y. Q., Z. M. Duniec, et al. (1994). "Endogenous 12(S)-HETE production by tumor cells and its role in metastasis." Cancer Res **54**(6): 1574-9.

Chenau, J., S. Michelland, 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): 4579-91.

Cho, W. C. (2007). "Potentially useful biomarkers for the diagnosis, treatment and prognosis of lung cancer." Biomed Pharmacother **61**(9): 515-9.

Choi, J. H., H. C. Kim, et al. (2001). "Vascular endothelial growth factor in the serum of patients with non-small cell lung cancer: correlation with platelet and leukocyte counts." Lung Cancer **33**(2-3): 171-9.

Chong, P. K., H. Lee, et al. "Upregulation of plasma C9 protein in gastric cancer patients." Proteomics **10**(18): 3210-21.

Christofk, H. R., M. G. Vander Heiden, et al. (2008). "The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth." Nature **452**(7184): 230-3.

Ciocca, D. R. and S. K. Calderwood (2005). "Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications." Cell Stress Chaperones **10**(2): 86-103.

Clower, C. V., D. Chatterjee, et al. "The alternative splicing repressors hnRNP A1/A2 and PTB influence pyruvate kinase isoform expression and cell metabolism." Proc Natl Acad Sci U S A **107**(5): 1894-9.

Collier, T. S., P. Sarkar, et al. "Quantitative top-down proteomics of SILAC labeled human embryonic stem cells." J Am Soc Mass Spectrom **21**(6): 879-89.

Conrads, T. P., B. L. Hood, et al. (2006). "Sampling and analytical strategies for biomarker discovery using mass spectrometry." Biotechniques **40**(6): 799-805.

Conrotto, P. and S. Souchelnytskyi (2008). "Proteomic approaches in biological and medical sciences: principles and applications." Exp Oncol **30**(3): 171-80.

Copeland, S., J. Siddiqui, et al. (2004). "Direct comparison of traditional ELISAs and membrane protein arrays for detection and quantification of human cytokines." J Immunol Methods **284**(1-2): 99-106.

Creighton, C. J., A. L. Benham, et al. "Discovery of novel microRNAs in female reproductive tract using next generation sequencing." PLoS One **5**(3): e9637.

Cuttitta, F., D. N. Carney, et al. (1985). "Bombesin-like peptides can function as autocrine growth factors in human small-cell lung cancer." Nature **316**(6031): 823-6.

Dabrowska, M., H. Grubek-Jaworska, et al. (2004). "[Diagnostic usefulness of selected tumor markers (CA125, CEA, CYFRA 21-1) in bronchoalveolar lavage fluid in patients with non-small cell lung cancer]." Pol Arch Med Wewn **111**(6): 659-65.

Dahabreh, I. J., H. Linardou, et al. "Somatic EGFR mutation and gene copy gain as predictive biomarkers for response to tyrosine kinase inhibitors in non-small cell lung cancer." Clin Cancer Res **16**(1): 291-303.

David, C. J., M. Chen, et al. "HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer." Nature **463**(7279): 364-8.

Dayarathna, M. K., W. S. Hancock, et al. (2008). "A two step fractionation approach for plasma proteomics using immunodepletion of abundant proteins and multi-lectin affinity chromatography: Application to the analysis of obesity, diabetes, and hypertension diseases." J Sep Sci **31**(6-7): 1156-66.

De Bock, M., D. de Seny, et al. "Challenges for biomarker discovery in body fluids using SELDI-TOF-MS." J Biomed Biotechnol **2010**: 906082.

Decock, J., S. Thirkettle, et al. "Matrix metalloproteinases: protective roles in cancer." J Cell Mol Med **15**(6): 1254-65.

DeMichele, M. A., A. L. Davis, et al. (1994). "Expression of mRNA for three bombesin receptor subtypes in human bronchial epithelial cells." Am J Respir Cell Mol Biol **11**(1): 66-74.

Dennis, M. S., M. Zhang, et al. (2002). "Albumin binding as a general strategy for improving the pharmacokinetics of proteins." J Biol Chem **277**(38): 35035-43.

Diamandis, E. P., L. Goodglick, et al. "Pentraxin-3 is a Novel Biomarker of Lung Carcinoma." Clin Cancer Res.

Diederich, S., D. Wormanns, et al. (2003). "Lung cancer screening with low-dose CT." Eur J Radiol **45**(1): 2-7.

Dimmeler, S. and A. M. Zeiher (1999). "Nitric oxide-an endothelial cell survival factor." Cell Death Differ **6**(10): 964-8.

Dobbin, K. K., D. G. Beer, et al. (2005). "Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays." Clin Cancer Res **11**(2 Pt 1): 565-72.

Dobryszycka, W. (1997). "Biological functions of haptoglobin--new pieces to an old puzzle." Eur J Clin Chem Clin Biochem **35**(9): 647-54.

Dowling, P., L. O'Driscoll, et al. (2007). "2-D difference gel electrophoresis of the lung squamous cell carcinoma versus normal sera demonstrates consistent alterations in the levels of ten specific proteins." Electrophoresis **28**(23): 4302-10.

Dreyfuss, G., V. N. Kim, et al. (2002). "Messenger-RNA-binding proteins and the messages they carry." Nat Rev Mol Cell Biol **3**(3): 195-205.

Dreyfuss, G., M. J. Matunis, et al. (1993). "hnRNP proteins and the biogenesis of mRNA." Annu Rev Biochem **62**: 289-321.

Duarte, R. L. and M. E. Paschoal (2006). "Molecular markers in lung cancer: prognostic role and relationship to smoking." J Bras Pneumol **32**(1): 56-65.

Dunn, W. B., N. J. Bailey, et al. (2005). "Measuring the metabolome: current analytical technologies." Analyst **130**(5): 606-25.

Earley, M. C., R. F. Vogt, Jr., et al. (2002). "Report from a workshop on multianalyte microsphere assays." Cytometry **50**(5): 239-42.

Egeblad, M. and Z. Werb (2002). "New functions for the matrix metalloproteinases in cancer progression." Nat Rev Cancer **2**(3): 161-74.

Elliott, J. A., K. Osterlind, et al. (1987). "Metastatic patterns in small-cell lung cancer: correlation of autopsy findings with clinical parameters in 537 patients." J Clin Oncol **5**(2): 246-54.

Elshal, M. F. and J. P. McCoy (2006). "Multiplex bead array assays: performance evaluation and comparison of sensitivity to ELISA." Methods **38**(4): 317-23.

Esquela-Kerscher, A. and F. J. Slack (2006). "Oncomirs - microRNAs with a role in cancer." Nat Rev Cancer **6**(4): 259-69.

Esser, M. T., R. D. Marchese, et al. (2003). "Memory T cells and vaccines." Vaccine **21**(5-6): 419-30.

Etzioni, R., N. Urban, et al. (2003). "The case for early detection." Nat Rev Cancer **3**(4): 243-52.

Fan, T. W., A. N. Lane, et al. (2009). "Altered regulation of metabolic pathways in human lung cancer discerned by (13)C stable isotope-resolved metabolomics (SIRM)." Mol Cancer **8**: 41.

Faruqi, S., R. Wilmot, et al. "Serum LDH in chronic cough: a potential marker of airway inflammation." Clin Respir J **6**(2): 81-7.

Fearnhead, N. S., M. P. Britton, et al. (2001). "The ABC of APC." Hum Mol Genet **10**(7): 721-33.

Fernandez, F. G. and R. J. Battafarano (2006). "Large-cell neuroendocrine carcinoma of the lung." Cancer Control **13**(4): 270-5.

Fidler, I. J. (2003). "The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited." Nat Rev Cancer **3**(6): 453-8.

Fielding, P., L. Turnbull, et al. (1999). "Heterogeneous nuclear ribonucleoprotein A2/B1 up-regulation in bronchial lavage specimens: a clinical marker of early lung cancer detection." Clin Cancer Res **5**(12): 4048-52.

Fitzgerald, R. L., C. L. O'Neal, et al. (1997). "Comparison of an ion-trap and a quadrupole mass spectrometer using diazepam as a model compound." J Anal Toxicol **21**(6): 445-50.

Folkman, J. (1986). "How is blood vessel growth regulated in normal and neoplastic tissue? G.H.A. Clowes memorial Award lecture." Cancer Res **46**(2): 467-73.

Fontanini, G., S. Vignati, et al. (1997). "Vascular endothelial growth factor is associated with neovascularization and influences progression of non-small cell lung carcinoma." Clin Cancer Res **3**(6): 861-5.

Fountoulakis, M., J. F. Juranville, et al. (2004). "Depletion of the high-abundance plasma proteins." Amino Acids **27**(3-4): 249-59.

Friedl, P. and S. Alexander "Cancer invasion and the microenvironment: plasticity and reciprocity." Cell **147**(5): 992-1009.

Friedl, P. and E. B. Brocker (2000). "The biology of cell locomotion within three-dimensional extracellular matrix." Cell Mol Life Sci **57**(1): 41-64.

Friedl, P. and K. Wolf (2003). "Tumour-cell invasion and migration: diversity and escape mechanisms." Nat Rev Cancer **3**(5): 362-74.

Fu, Q., J. Zhu, et al. "Comparison of multiplex immunoassay platforms." Clin Chem **56**(2): 314-8.

Fujii, J. and Y. Ikeda (2002). "Advances in our understanding of peroxiredoxin, a multifunctional, mammalian redox protein." Redox Rep **7**(3): 123-30.

Fujimura, T., Y. Shinohara, et al. (2008). "Glycosylation status of haptoglobin in sera of patients with prostate cancer vs. benign prostate disease or normal subjects." Int J Cancer **122**(1): 39-49.

Gail, M. H., L. Muenz, et al. (1988). "Multiple markers for lung cancer diagnosis: validation of models for localized lung cancer." J Natl Cancer Inst **80**(2): 97-101.

Gatenby, R. A. and R. J. Gillies (2004). "Why do cancers have high aerobic glycolysis?" Nat Rev Cancer **4**(11): 891-9.

Gazdar, A. F. and J. D. Minna (1997). "Cigarettes, sex, and lung adenocarcinoma." J Natl Cancer Inst **89**(21): 1563-5.

Gerber, H. P., K. J. Hillan, et al. (1999). "VEGF is required for growth and survival in neonatal mice." Development **126**(6): 1149-59.

Gilchrist, A., C. E. Au, et al. (2006). "Quantitative proteomics analysis of the secretory pathway." Cell **127**(6): 1265-81.

Girard, L., S. Zochbauer-Muller, et al. (2000). "Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering." Cancer Res **60**(17): 4894-906.

Glen, A., C. A. Evans, et al. "Eight-plex iTRAQ analysis of variant metastatic human prostate cancer cells identifies candidate biomarkers of progression: An exploratory study." Prostate **70**(12): 1313-32.

Golan-Gerstl, R., M. Cohen, et al. "Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma." Cancer Res **71**(13): 4464-72.

Good, D. M., V. Thongboonkerd, et al. (2007). "Body fluid proteomics for biomarker discovery: lessons from the past hold the key to success in the future." J Proteome Res **6**(12): 4549-55.

Goshe, M. B. and R. D. Smith (2003). "Stable isotope-coded proteomic mass spectrometry." Curr Opin Biotechnol **14**(1): 101-9.

Griffiths, W. J., A. P. Jonsson, et al. (2001). "Electrospray and tandem mass spectrometry in biochemistry." Biochem J **355**(Pt 3): 545-61.

Gruhler, S. and I. Kratchmarova (2008). "Stable isotope labeling by amino acids in cell culture (SILAC)." Methods Mol Biol **424**: 101-11.

Gruys, E., M. J. Toussaint, et al. (2005). "Acute phase reaction and acute phase proteins." J Zhejiang Univ Sci B **6**(11): 1045-56.

Gu, W., W. Liu, et al. "Emergence of heterogeneous nuclear ribonucleoprotein A2/B1 vs loss of E-cadherin: their reciprocal immunoexpression profiles in human pancreatic cancer." Ann Diagn Pathol.

Guo, N. L. and Y. W. Wan "Pathway-based identification of a smoking associated 6-gene signature predictive of lung cancer risk and survival." Artif Intell Med **55**(2): 97-105.

Gupta, V. and R. N. Bamezai "Human pyruvate kinase M2: a multifunctional protein." Protein Sci **19**(11): 2031-44.

Gygi, S. P., B. Rist, et al. (2002). "Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags." J Proteome Res **1**(1): 47-54.

Halliwell, B. (1999). "Oxygen and nitrogen are pro-carcinogens. Damage to DNA by reactive oxygen, chlorine and nitrogen species: measurement, mechanism and the effects of nutrition." Mutat Res **443**(1-2): 37-52.

Han, M., C. T. Liew, et al. (2008). "Novel blood-based, five-gene biomarker set for the detection of colorectal cancer." Clin Cancer Res **14**(2): 455-60.

Han, S. W., T. Y. Kim, et al. (2005). "Predictive and prognostic impact of epidermal growth factor receptor mutation in non-small-cell lung cancer patients treated with gefitinib." J Clin Oncol **23**(11): 2493-501.

Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.

Hanamura, A., J. F. Caceres, et al. (1998). "Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors." Rna **4**(4): 430-44.

Hanash, S. M., C. S. Baik, et al. "Emerging molecular biomarkers--blood-based strategies to detect and monitor cancer." Nat Rev Clin Oncol **8**(3): 142-50.

Harsha, H. C., H. Molina, et al. (2008). "Quantitative proteomics using stable isotope labeling with amino acids in cell culture." Nat Protoc **3**(3): 505-16.

Hartl, F. U. (1996). "Molecular chaperones in cellular protein folding." Nature **381**(6583): 571-9.

Hartwig, S., A. Czibere, et al. (2009). "Combinatorial hexapeptide ligand libraries (ProteoMiner): an innovative fractionation tool for differential quantitative clinical proteomics." Arch Physiol Biochem **115**(3): 155-60.

Harvey, P., A. Warn, et al. (1996). "Immunoreactivity for hepatocyte growth factor/scatter factor and its receptor, met, in human lung carcinomas and malignant mesotheliomas." J Pathol **180**(4): 389-94.

Hashiguchi, T., K. Tanaka, et al. (2009). "Diagnostic value of serum peptidome analyses for protease activated pathological conditions beyond cancer diagnosis." Med Hypotheses **73**(5): 760-3.

He, Y., M. A. Brown, et al. (2005). "Roles of heterogeneous nuclear ribonucleoproteins A and B in cell proliferation." J Cell Sci **118**(Pt 14): 3173-83.

He, Y. and R. Smith (2009). "Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B." Cell Mol Life Sci **66**(7): 1239-56.

Hergenroeder, G., J. B. Redell, et al. (2008). "Identification of serum biomarkers in brain-injured adults: potential for predicting elevated intracranial pressure." J Neurotrauma **25**(2): 79-93.

Herth, F. J., R. Eberhardt, et al. (2006). "The future of bronchoscopy in diagnosing, staging and treatment of lung cancer." Respiration **73**(4): 399-409.

Hibi, K., B. Trink, et al. (2000). "AIS is an oncogene amplified in squamous cell carcinoma." Proc Natl Acad Sci U S A **97**(10): 5462-7.

Higgs, R. E., M. D. Knierman, et al. (2005). "Comprehensive label-free method for the relative quantification of proteins from biological samples." J Proteome Res **4**(4): 1442-50.

Hightower, L. E. (1991). "Heat shock, stress proteins, chaperones, and proteotoxicity." Cell **66**(2): 191-7.

Hoekstra, R., W. W. van den Broek, et al. (2001). "Effect of electroconvulsive therapy on biopterin and large neutral amino acids in severe, medication-resistant depression." Psychiatry Res **103**(2-3): 115-23.

Hojilla, C. V., G. A. Wood, et al. (2008). "Inflammation and breast cancer: metalloproteinases as common effectors of inflammation and extracellular matrix breakdown in breast cancer." Breast Cancer Res **10**(2): 205.

Hollstein, M., D. Sidransky, et al. (1991). "p53 mutations in human cancers." Science **253**(5015): 49-53.

Hoofnagle, A. N., J. O. Becker, et al. "Multiple-reaction monitoring-mass spectrometric assays can accurately measure the relative protein abundance in complex mixtures." Clin Chem **58**(4): 777-81.

Hortin, G. L. (2006). "The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome." Clin Chem **52**(7): 1223-37.

Hoskins, E. R., B. L. Hood, et al. "Proteomic analysis of ovarian cancer proximal fluids: validation of elevated peroxiredoxin 1 in patient peripheral circulation." PLoS One **6**(9): e25056.

Hsieh, S. Y., R. K. Chen, et al. (2006). "Systematical evaluation of the effects of sample collection procedures on low-molecular-weight serum/plasma proteome profiling." Proteomics **6**(10): 3189-98.

Hsu, N. Y., H. C. Ho, et al. (2001). "Overexpression of dihydrodiol dehydrogenase as a prognostic marker of non-small cell lung cancer." Cancer Res **61**(6): 2727-31.

Hu, J., P. E. Van den Steen, et al. (2007). "Matrix metalloproteinase inhibitors as therapy for inflammatory and vascular diseases." Nat Rev Drug Discov **6**(6): 480-98.

Hua, H., M. Li, et al. "Matrix metalloproteinases in tumorigenesis: an evolving paradigm." Cell Mol Life Sci **68**(23): 3853-68.

Huang, L. J., S. X. Chen, et al. (2006). "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.

Huang, S. K., M. M. Darfler, et al. (2009). "LC/MS-based quantitative proteomic analysis of paraffin-embedded archival melanomas reveals potential proteomic biomarkers associated with metastasis." PLoS One **4**(2): e4430.

Hung, J. J., K. C. Chow, et al. (2006). "Expression of dihydrodiol dehydrogenase and resistance to chemotherapy and radiotherapy in adenocarcinoma cells of lung." Anticancer Res **26**(4B): 2949-55.

Hung, K. E. and K. H. Yu "Proteomic approaches to cancer biomarkers." Gastroenterology **138**(1): 46-51 e1.

Husgafvel-Pursiainen, K., P. Boffetta, et al. (2000). "p53 mutations and exposure to environmental tobacco smoke in a multicenter study on lung cancer." Cancer Res **60**(11): 2906-11.

Ishiwata, T., K. Takahashi, et al. (2005). "Serum tenascin-C as a potential predictive marker of angiogenesis in non-small cell lung cancer." Anticancer Res **25**(1B): 489-95.

Ismailov, D. A., M. A. Shishkin, et al. (1993). "[Diagnosis and surgical treatment of benign tumors of the lungs]." Khirurgiia (Mosk)(12): 29-31.

Iyoda, T. and F. Fukai "Modulation of Tumor Cell Survival, Proliferation, and Differentiation by the Peptide Derived from Tenascin-C: Implication of beta1-Integrin Activation." Int J Cell Biol **2012**: 647594.

Jackson, J. R., M. P. Seed, et al. (1997). "The codependence of angiogenesis and chronic inflammation." Faseb J **11**(6): 457-65.

Jain, P. and A. C. Arroliga (2001). "Spiral CT for lung cancer screening: is it ready for prime time?" Cleve Clin J Med **68**(1): 74-81.

Jensen, O. N. (2004). "Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry." Curr Opin Chem Biol **8**(1): 33-41.

Ji, D., M. Li, et al. "Prognostic role of serum AZGP1, PEDF and PRDX2 in colorectal cancer patients." Carcinogenesis **34**(6): 1265-72.

Johnson, D. H., L. Fehrenbacher, et al. (2004). "Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer." J Clin Oncol **22**(11): 2184-91.

Jolly, C. and R. I. Morimoto (2000). "Role of the heat shock response and molecular chaperones in oncogenesis and cell death." J Natl Cancer Inst **92**(19): 1564-72.

Jones, F. S. and P. L. Jones (2000). "The tenascin family of ECM glycoproteins: structure, function, and regulation during embryonic development and tissue remodeling." Dev Dyn **218**(2): 235-59.

Kagohashi, K., H. Satoh, et al. (2008). "A re-evaluation of squamous cell carcinoma antigen (SCC) as a serum marker for non-small cell lung cancer." Med Oncol **25**(2): 187-9.

Kahn, N., M. Meister, et al. "Early detection of lung cancer by molecular markers in endobronchial epithelial-lining fluid." J Thorac Oncol **7**(6): 1001-8.

Kaiser, U., M. Schilli, et al. (1996). "Expression of bcl-2--protein in small cell lung cancer." Lung Cancer **15**(1): 31-40.

Kakinuma, R., H. Ohmatsu, et al. (1999). "Detection failures in spiral CT screening for lung cancer: analysis of CT findings." Radiology **212**(1): 61-6.

Kamma, H., H. Horiguchi, et al. (1999). "Molecular characterization of the hnRNP A2/B1 proteins: tissue-specific expression and novel isoforms." Exp Cell Res **246**(2): 399-411.

Kang, S. M., H. J. Sung, et al. "The Haptoglobin beta chain as a supportive biomarker for human lung cancers." Mol Biosyst **7**(4): 1167-75.

Karagiannis, G. S., M. P. Pavlou, et al. (2010). "Cancer secretomics reveal pathophysiological pathways in cancer molecular oncology." Mol Oncol **4**(6): 496-510.

Karihtala, P., A. Mantyniemi, et al. (2003). "Peroxiredoxins in breast carcinoma." Clin Cancer Res **9**(9): 3418-24.

Karube, Y., H. Tanaka, et al. (2005). "Reduced expression of Dicer associated with poor prognosis in lung cancer patients." Cancer Sci **96**(2): 111-5.

Kasai, N., K. Kobayashi, et al. "Soluble heparin-binding EGF-like growth factor (HB-EGF) detected by newly developed immuno-PCR method is a clear-cut serological biomarker for ovarian cancer." Am J Transl Res **4**(4): 415-21.

Kaya, A., A. Ciledag, et al. (2004). "The prognostic significance of vascular endothelial growth factor levels in sera of non-small cell lung cancer patients." Respir Med **98**(7): 632-6.

Kellar, K. L. and J. P. Douglass (2003). "Multiplexed microsphere-based flow cytometric immunoassays for human cytokines." J Immunol Methods **279**(1-2): 277-85.

Keller, K. E., I. S. Tan, et al. "SAICAR stimulates pyruvate kinase isoform M2 and promotes cancer cell survival in glucose-limited conditions." Science **338**(6110): 1069-72.

Kennedy, T. C., A. McWilliams, et al. (2007). "Bronchial intraepithelial neoplasia/early central airways lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition)." Chest **132**(3 Suppl): 221S-233S.

Kersting, M., C. Friedl, et al. (2000). "Differential frequencies of p16(INK4a) promoter hypermethylation, p53 mutation, and K-ras mutation in exfoliative material mark the development of lung cancer in symptomatic chronic smokers." J Clin Oncol **18**(18): 3221-9.

Kiehntopf, M., R. Siegmund, et al. (2007). "Use of SELDI-TOF mass spectrometry for identification of new biomarkers: potential and limitations." Clin Chem Lab Med **45**(11): 1435-49.

Kim, J. H., P. N. Bogner, et al. (2008). "Up-regulation of peroxiredoxin 1 in lung cancer and its implication as a prognostic and therapeutic target." Clin Cancer Res **14**(8): 2326-33.

Kim, M. R. and C. W. Kim (2007). "Human blood plasma preparation for two-dimensional gel electrophoresis." J Chromatogr B Analyt Technol Biomed Life Sci **849**(1-2): 203-10.

Kinoshita, M., H. Watanabe, et al. (1998). "Cyfra 21-1 as a marker of lung cancer." Kurume Med J **45**(1): 7-9.

Knudson, A. G., Jr. (1971). "Mutation and cancer: statistical study of retinoblastoma." Proc Natl Acad Sci U S A **68**(4): 820-3.

Koukourakis, M. I., A. Giatromanolaki, et al. (2000). "Vascular endothelial growth factor/KDR activated microvessel density versus CD31 standard microvessel density in non-small cell lung cancer." Cancer Res **60**(11): 3088-95.

Kozu, T., B. Henrich, et al. (1995). "Structure and expression of the gene (HNRPA2B1) encoding the human hnRNP protein A2/B1." Genomics **25**(2): 365-71.

Krasnov, G. S., A. A. Dmitriev, et al. "Deregulation of glycolysis in cancer: glyceraldehyde-3-phosphate dehydrogenase as a therapeutic target." Expert Opin Ther Targets **17**(6): 681-93.

Krecic, A. M. and M. S. Swanson (1999). "hnRNP complexes: composition, structure, and function." Curr Opin Cell Biol **11**(3): 363-71.

Krieg, R. C., Y. Dong, et al. (2005). "Protein quantification and its tolerance for different interfering reagents using the BCA-method with regard to 2D SDS PAGE." J Biochem Biophys Methods **65**(1): 13-9.

Kristiansen, M., J. H. Graversen, et al. (2001). "Identification of the haemoglobin scavenger receptor." Nature **409**(6817): 198-201.

Kuhn, E., J. Wu, et al. (2004). "Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and 13C-labeled peptide standards." Proteomics **4**(4): 1175-86.

Kulasingam, V. and E. P. Diamandis (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.

Kuzyk, M. A., D. Smith, et al. (2009). "Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma." Mol Cell Proteomics **8**(8): 1860-77.

Lam, N. Y., T. H. Rainer, et al. (2004). "EDTA is a better anticoagulant than heparin or citrate for delayed blood processing for plasma DNA analysis." Clin Chem **50**(1): 256-7.

Lane, D. P., C. Midgley, et al. (1993). "Tumour suppressor genes and molecular chaperones." Philos Trans R Soc Lond B Biol Sci **339**(1289): 369-72; discussion 372-3.

Lavrovsky, Y., B. Chatterjee, et al. (2000). "Role of redox-regulated transcription factors in inflammation, aging and age-related diseases." Exp Gerontol **35**(5): 521-32.

Lebanony, D., H. Benjamin, et al. (2009). "Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from nonsquamous non-small-cell lung carcinoma." J Clin Oncol **27**(12): 2030-7.

Lee, S. M., J. M. Goo, et al. "A new classification of adenocarcinoma: what the radiologists need to know." Diagn Interv Radiol **18**(6): 519-26.

Lehtonen, S. T., A. M. Svensk, et al. (2004). "Peroxiredoxins, a novel protein family in lung cancer." Int J Cancer **111**(4): 514-21.

Leitner, A. and W. Lindner (2004). "Current chemical tagging strategies for proteome analysis by mass spectrometry." J Chromatogr B Analyt Technol Biomed Life Sci **813**(1-2): 1-26.

Leslie, M. "Mysteries of the cell. How does a cell know its size?" Science **334**(6059): 1047-8.

Li, D., P. F. Firozi, et al. (2001). "Sensitivity to DNA damage induced by benzo(a)pyrene diol epoxide and risk of lung cancer: a case-control analysis." Cancer Res **61**(4): 1445-50.

Li, H., J. Fu, et al. "[Diagnostic significance of combining telomerase activity with CYFRA21-1 level in differentiating malignant pleural effusion caused by lung cancer from benign pleural effusion]." Zhongguo Fei Ai Za Zhi **13**(6): 652-4.

Li, J., H. Steen, et al. (2003). "Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents: the yeast salinity stress response." Mol Cell Proteomics **2**(11): 1198-204.

Li, X., T. Asmitananda, et al. "Biomarkers in the lung cancer diagnosis: a clinical perspective." Neoplasma **59**(5): 500-7.

Liao, Q., L. Li, et al. (2005). "[Comparison of TPS, CEA, CYFRA21-1 and STNFR in diagnosis of lung cancer.]." Zhongguo Fei Ai Za Zhi **8**(4): 309-312.

Ling, W. L., L. Deng, et al. (2003). "Improvement of monoclonal antibody production in hybridoma cells by dimethyl sulfoxide." Biotechnol Prog **19**(1): 158-62.

Liotta, L. A. (1986). "Tumor invasion and metastases--role of the extracellular matrix: Rhoads Memorial Award lecture." Cancer Res **46**(1): 1-7.

Liotta, L. A., M. Ferrari, et al. (2003). "Clinical proteomics: written in blood." Nature **425**(6961): 905.

Lipinski, B. "Hydroxyl radical and its scavengers in health and disease." Oxid Med Cell Longev **2011**: 809696.

Lipman, N. S., L. R. Jackson, et al. (2005). "Monoclonal versus polyclonal antibodies: distinguishing characteristics, applications, and information resources." Ilar J **46**(3): 258-68.

Lippi, G., G. L. Salvagno, et al. (2006). "Influence of hemolysis on routine clinical chemistry testing." Clin Chem Lab Med **44**(3): 311-6.

Lippi, G., G. L. Salvagno, et al. (2006). "Reliability of the thrombin-generation assay in frozen-thawed platelet-rich plasma." Clin Chem **52**(9): 1827-8.

Lissitzky, J. C., N. Pourreau-Schneider, et al. (1985). "[Invasion, metastases of solid tumors. Interaction of tumor cells with tissue and vascular basement membranes]." Bull Cancer **72**(5): 367-76.

Liu, H., Y. Kato, et al. (2012). "The role of MMP-1 in breast cancer growth and metastasis to the brain in a xenograft model." BMC Cancer **12**: 583.

Liu, Y., L. Yu, et al. (2006). "[Study on the value of tumor markers ProGRP, CYFRA21-1, NSE and CEA in the differential diagnosis of pleural effusion.]." Zhongguo Fei Ai Za Zhi **9**(3): 273-276.

Lowenthal, M. S., A. I. Mehta, et al. (2005). "Analysis of albumin-associated peptides and proteins from ovarian cancer patients." Clin Chem **51**(10): 1933-45.

Lundgren, D. H., S. I. Hwang, et al. "Role of spectral counting in quantitative proteomics." Expert Rev Proteomics **7**(1): 39-53.

Mackinnon, A. C., M. Tretiakova, et al. "Paxillin expression and amplification in early lung lesions of high-risk patients, lung adenocarcinoma and metastatic disease." J Clin Pathol **64**(1): 16-24.

Makawita, S., C. Smith, et al. (2011). "Integrated proteomic profiling of cell line conditioned media and pancreatic juice for the identification of pancreatic cancer biomarkers." Mol Cell Proteomics **10**(10): M111 008599.

Mallick, R., S. K. Patnaik, et al. "MicroRNAs and lung cancer: Biology and applications in diagnosis and prognosis." J Carcinog **9**.

Mansilla, M. J., X. Montalban, et al. "Heat shock protein 70: roles in multiple sclerosis." Mol Med **18**(9): 1018-28.

Marcus, P. M., E. J. Bergstralh, et al. (2006). "Extended lung cancer incidence follow-up in the Mayo Lung Project and overdiagnosis." J Natl Cancer Inst **98**(11): 748-56.

Marino, K., J. Bones, et al. "A systematic approach to protein glycosylation analysis: a path through the maze." Nat Chem Biol **6**(10): 713-23.

Marouga, R., S. David, et al. (2005). "The development of the DIGE system: 2D fluorescence difference gel analysis technology." Anal Bioanal Chem **382**(3): 669-78.

Marvin, L. F., M. A. Roberts, et al. (2003). "Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry." Clin Chim Acta **337**(1-2): 11-21.

Mazurek, S. "Pyruvate kinase type M2: a key regulator of the metabolic budget system in tumor cells." Int J Biochem Cell Biol **43**(7): 969-80.

Mazurek, S., C. B. Boschek, et al. (2005). "Pyruvate kinase type M2 and its role in tumor growth and spreading." Semin Cancer Biol **15**(4): 300-8.

McCarthy, N. J. and S. M. Swain (2001). "Tumor markers: should we or shouldn't we?" Cancer J **7**(3): 175-7.

McWilliams, A., B. Lam, et al. (2009). "Early proximal lung cancer diagnosis and treatment." Eur Respir J **33**(3): 656-65.

Midwood, K. S., T. Hussenet, et al. "Advances in tenascin-C biology." Cell Mol Life Sci **68**(19): 3175-99.

Midwood, K. S. and G. Orend (2009). "The role of tenascin-C in tissue injury and tumorigenesis." J Cell Commun Signal **3**(3-4): 287-310.

Millar, E. K., P. H. Graham, et al. (2011). "Prediction of outcome of early ER+ breast cancer is improved using a biomarker panel, which includes Ki-67 and p53." Br J Cancer **105**(2): 272-80.

Mirza, S. P. and M. Olivier (2008). "Methods and approaches for the comprehensive characterization and quantification of cellular proteomes using mass spectrometry." Physiol Genomics **33**(1): 3-11.

Mitsudomi, T. "Advances in target therapy for lung cancer." Jpn J Clin Oncol **40**(2): 101-6.

Moncada, S. and E. A. Higgs (1986). "Arachidonate metabolism in blood cells and the vessel wall." Clin Haematol **15**(2): 273-92.

Montuenga, L. M., J. Zhou, et al. (1998). "Expression of heterogeneous nuclear ribonucleoprotein A2/B1 changes with critical stages of mammalian lung development." Am J Respir Cell Mol Biol **19**(4): 554-62.

Mora, J. F., G. J. Van Berkel, et al. (2000). "Electrochemical processes in electrospray ionization mass spectrometry." J Mass Spectrom **35**(8): 939-52.

Moreno, J. J. (2009). "New aspects of the role of hydroxyeicosatetraenoic acids in cell growth and cancer development." Biochem Pharmacol **77**(1): 1-10.

Morgan, J. M., H. Navabi, et al. (1997). "Role of calcium chelation in high-temperature antigen retrieval at different pH values." J Pathol **182**(2): 233-7.

Motadi, L. R., N. L. Misso, et al. (2007). "Molecular genetics and mechanisms of apoptosis in carcinomas of the lung and pleura: therapeutic targets." Int Immunopharmacol **7**(14): 1934-47.

Nagafuchi, S. "The role of B cells in regulating the magnitude of immune response." Microbiol Immunol **54**(8): 487-90.

Neilson, K. A., N. A. Ali, et al. "Less label, more free: approaches in label-free quantitative mass spectrometry." Proteomics **11**(4): 535-53.

Neurauter, G., A. V. Grahmann, et al. (2008). "Serum phenylalanine concentrations in patients with ovarian carcinoma correlate with concentrations of immune activation markers and of isoprostane-8." Cancer Lett **272**(1): 141-7.

Nikitin, A., S. Egorov, et al. (2003). "Pathway studio--the analysis and navigation of molecular networks." Bioinformatics **19**(16): 2155-7.

Nishio, M., T. Koshikawa, et al. (1996). "Prognostic significance of abnormal p53 accumulation in primary, resected non-small-cell lung cancers." J Clin Oncol **14**(2): 497-502.

Njemini, R., C. Demanet, et al. (2004). "Inflammatory status as an important determinant of heat shock protein 70 serum concentrations during aging." Biogerontology **5**(1): 31-8.

Noguchi, T., H. Inoue, et al. (1986). "The M1- and M2-type isozymes of rat pyruvate kinase are produced from the same gene by alternative RNA splicing." J Biol Chem **261**(29): 13807-12.

Noh, D. Y., S. J. Ahn, et al. (2001). "Overexpression of peroxiredoxin in human breast cancer." Anticancer Res **21**(3B): 2085-90.

Nordberg, J. and E. S. Arner (2001). "Reactive oxygen species, antioxidants, and the mammalian thioredoxin system." Free Radic Biol Med **31**(11): 1287-312.

Nutini, S., G. Cappelli, et al. (1990). "Serum NSE, CEA, CT, CA 15-3 levels in human lung cancer." Int J Biol Markers **5**(4): 198-202.

O'Byrne, K. J., M. I. Koukourakis, et al. (2000). "Vascular endothelial growth factor, platelet-derived endothelial cell growth factor and angiogenesis in non-small-cell lung cancer." Br J Cancer **82**(8): 1427-32.

O'Connell, K. and K. Ohlendieck (2009). "Proteomic DIGE analysis of the mitochondria-enriched fraction from aged rat skeletal muscle." Proteomics **9**(24): 5509-24.

O'Farrell, P. H. (1975). "High resolution two-dimensional electrophoresis of proteins." J Biol Chem **250**(10): 4007-21.

Ohri, C. M., A. Shikotra, et al. "Tumour necrosis factor-alpha expression in tumour islets confers a survival advantage in non-small cell lung cancer." BMC Cancer **10**: 323.

Ohtsuka, T., H. Nomori, et al. (2003). "Radiological examination for peripheral lung cancers and benign nodules less than 10 mm." Lung Cancer **42**(3): 291-6.

Olaussen, K. A., A. Dunant, et al. (2006). "DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy." N Engl J Med **355**(10): 983-91.

Olivero, M., M. Rizzo, et al. (1996). "Overexpression and activation of hepatocyte growth factor/scatter factor in human non-small-cell lung carcinomas." Br J Cancer **74**(12): 1862-8.

Ong, S. E., B. Blagoev, 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): 376-86.

Ong, S. E., L. J. Foster, et al. (2003). "Mass spectrometric-based approaches in quantitative proteomics." Methods **29**(2): 124-30.

Oremek, G., R. Kukshaite, et al. (2007). "[The significance of TU M2-PK tumor marker for lung cancer diagnostics]." Klin Med (Mosk) **85**(7): 56-8.

Oskarsson, T., S. Acharyya, et al. "Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs." Nat Med **17**(7): 867-74.

Ostrowski, J. and K. Bomsztyk (2003). "Nuclear shift of hnRNP K protein in neoplasms and other states of enhanced cell proliferation." Br J Cancer **89**(8): 1493-501.

Overall, C. M. and C. Lopez-Otin (2002). "Strategies for MMP inhibition in cancer: innovations for the post-trial era." Nat Rev Cancer **2**(9): 657-72.

Pang, S., J. Smith, et al. (2005). "A comparability study of the emerging protein array platforms with established ELISA procedures." J Immunol Methods **302**(1-2): 1-12.

Park, J. H., Y. S. Kim, et al. (2006). "Expression of peroxiredoxin and thioredoxin in human lung cancer and paired normal lung." Respirology **11**(3): 269-75.

Patry, C., L. Bouchard, et al. (2003). "Small interfering RNA-mediated reduction in heterogeneous nuclear ribonucleoparticule A1/A2 proteins induces apoptosis in human cancer cells but not in normal mortal cell lines." Cancer Res **63**(22): 7679-88.

Pecur, L., S. Kapitanovic, et al. (1994). "Prognostic significance of transforming growth factor alpha TGF-alpha) in human lung carcinoma: an immunohistochemical study." Anticancer Res **14**(6B): 2839-43.

Peng, J., J. E. Elias, et al. (2003). "Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome." J Proteome Res **2**(1): 43-50.

Peng, X. C., F. M. Gong, et al. "Comparative proteomic approach identifies PKM2 and cofilin-1 as potential diagnostic, prognostic and therapeutic targets for pulmonary adenocarcinoma." PLoS One **6**(11): e27309.

Pertea, M. and S. L. Salzberg (2012). "Between a chicken and a grape: estimating the number of human genes." Genome Biol **11**(5): 206.

Petricoin, E. F., A. M. Ardekani, et al. (2002). "Use of proteomic patterns in serum to identify ovarian cancer." Lancet **359**(9306): 572-7.

Petricoin, E. F., D. K. Ornstein, et al. (2004). "Clinical proteomics: Applications for prostate cancer biomarker discovery and detection." Urol Oncol **22**(4): 322-8.

Pezzella, F., H. Turley, et al. (1993). "bcl-2 protein in non-small-cell lung carcinoma." N Engl J Med **329**(10): 690-4.

Pimienta, G., R. Chaerkady, et al. (2009). "SILAC for global phosphoproteomic analysis." Methods Mol Biol **527**: 107-16, x.

Planque, C., V. Kulasingam, et al. (2009). "Identification of five candidate lung cancer biomarkers by proteomics analysis of conditioned media of four lung cancer cell lines." Mol Cell Proteomics **8**(12): 2746-58.

Ploder, M., G. Neurauter, et al. (2008). "Serum phenylalanine in patients post trauma and with sepsis correlate to neopterin concentrations." Amino Acids **35**(2): 303-7.

Pockley, A. G. (2003). "Heat shock proteins as regulators of the immune response." Lancet **362**(9382): 469-76.

Poola, I., R. L. DeWitty, et al. (2005). "Identification of MMP-1 as a putative breast cancer predictive marker by global gene expression analysis." Nat Med **11**(5): 481-3.

Poste, G. and I. J. Fidler (1980). "The pathogenesis of cancer metastasis." Nature **283**(5743): 139-46.

Price, N. (2004). "Emerging molecular biomarkers for early detection of lung cancer in patients at high risk." Clin Lung Cancer **6**(3): 145-8.

Prindiville, S. A., T. Byers, et al. (2003). "Sputum cytological atypia as a predictor of incident lung cancer in a cohort of heavy smokers with airflow obstruction." Cancer Epidemiol Biomarkers Prev **12**(10): 987-93.

Pujol, J. L., J. Grenier, et al. (1993). "Serum fragment of cytokeratin subunit 19 measured by CYFRA 21-1 immunoradiometric assay as a marker of lung cancer." Cancer Res **53**(1): 61-6.

Qi, Y., J. F. Chiu, et al. (2005). "Comparative proteomic analysis of esophageal squamous cell carcinoma." Proteomics **5**(11): 2960-71.

Qu, J., W. J. Jusko, et al. (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): 4543-52.

Quintana, L. F., J. M. Campistol, et al. (2009). "Application of label-free quantitative peptidomics for the identification of urinary biomarkers of kidney chronic allograft dysfunction." Mol Cell Proteomics **8**(7): 1658-73.

Rabilloud, T. (2002). "Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains." Proteomics **2**(1): 3-10.

Rabinowits, G., C. Gercel-Taylor, et al. (2009). "Exosomal microRNA: a diagnostic marker for lung cancer." Clin Lung Cancer **10**(1): 42-6.

Rak, J., J. Filmus, et al. (1995). "Oncogenes as inducers of tumor angiogenesis." Cancer Metastasis Rev **14**(4): 263-77.

Ramaswamy, S., K. N. Ross, et al. (2003). "A molecular signature of metastasis in primary solid tumors." Nat Genet **33**(1): 49-54.

Rantala, J. K., J. Pouwels, et al. "SHARPIN is an endogenous inhibitor of beta1-integrin activation." Nat Cell Biol **13**(11): 1315-24.

Read, C., S. Janes, et al. (2006). "Early lung cancer: screening and detection." Prim Care Respir J **15**(6): 332-6.

Reissmann, P. T., H. Koga, et al. (1999). "Amplification and overexpression of the cyclin D1 and epidermal growth factor receptor genes in non-small-cell lung cancer. Lung Cancer Study Group." J Cancer Res Clin Oncol **125**(2): 61-70.

Rekhtman, N. "Neuroendocrine tumors of the lung: an update." Arch Pathol Lab Med **134**(11): 1628-38.

Rhee, S. G., S. W. Kang, et al. (2001). "Peroxiredoxin, a novel family of peroxidases." IUBMB Life **52**(1-2): 35-41.

Richardson, G. E. and B. E. Johnson (1993). "The biology of lung cancer." Semin Oncol **20**(2): 105-27.

Risse, E. K., G. P. Vooijs, et al. (1988). "Diagnostic significance of "severe dysplasia" in sputum cytology." Acta Cytol **32**(5): 629-34.

Rodenhuis, S. and R. J. Slebos (1990). "The ras oncogenes in human lung cancer." Am Rev Respir Dis **142**(6 Pt 2): S27-30.

Rogatsky, E., V. Tomuta, et al. (2007). "Nondigest liberation of biomarkers from plasma: a novel two-stage ultrafiltration approach." J Sep Sci **30**(2): 245-8.

Roschinger, W., B. Olgemoller, et al. (2003). "Advances in analytical mass spectrometry to improve screening for inherited metabolic diseases." Eur J Pediatr **162 Suppl 1**: S67-76.

Ross, P. L., Y. N. Huang, et al. (2004). "Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents." Mol Cell Proteomics **3**(12): 1154-69.

Rossi, A., P. Maione, et al. (2005). "Screening for lung cancer: New horizons?" Crit Rev Oncol Hematol **56**(3): 311-20.

Rower, C., J. P. Vissers, et al. (2009). "Towards a proteome signature for invasive ductal breast carcinoma derived from label-free nanoscale LC-MS protein expression profiling of tumorous and glandular tissue." Anal Bioanal Chem **395**(8): 2443-56.

Ryan, D. and K. Robards (2006). "Metabolomics: The greatest omics of them all?" Anal Chem **78**(23): 7954-8.

Salven, P., T. Ruotsalainen, et al. (1998). "High pre-treatment serum level of vascular endothelial growth factor (VEGF) is associated with poor outcome in small-cell lung cancer." Int J Cancer **79**(2): 144-6.

Schaub, S., J. Wilkins, et al. (2004). "Urine protein profiling with surface-enhanced laser-desorption/ionization time-of-flight mass spectrometry." Kidney Int **65**(1): 323-32.

Schneider, J., H. G. Velcovsky, et al. (2000). "Comparison of the tumor markers tumor M2-PK, CEA, CYFRA 21-1, NSE and SCC in the diagnosis of lung cancer." Anticancer Res **20**(6D): 5053-8.

Schulze, G. (2000). "The tumor marker tumor M2-PK: an application in the diagnosis of gastrointestinal cancer." Anticancer Res **20**(6D): 4961-4.

Schwenk, J. M., U. Igel, et al. "Comparative protein profiling of serum and plasma using an antibody suspension bead array approach." Proteomics **10**(3): 532-40.

Seal, U. S., R. P. Doe, et al. (1978). "Response of serum haptoglobin to hormone treatment and the relation of pretreatment values to mortality in patients with prostatic cancer." Cancer **42**(4): 1720-9.

Sekido, Y., K. M. Fong, et al. (1998). "Progress in understanding the molecular pathogenesis of human lung cancer." Biochim Biophys Acta **1378**(1): F21-59.

Semenza, G. L., D. Artemov, et al. (2001). "'The metabolism of tumours': 70 years later." Novartis Found Symp **240**: 251-60; discussion 260-4.

Seo, M. S., S. W. Kang, et al. (2000). "Identification of a new type of mammalian peroxiredoxin that forms an intramolecular disulfide as a reaction intermediate." J Biol Chem **275**(27): 20346-54.

Sethuraman, M., M. E. McComb, et al. (2004). "Isotope-coded affinity tag approach to identify and quantify oxidant-sensitive protein thiols." Mol Cell Proteomics **3**(3): 273-8.

Shah, A., H. Singh, et al. "Differential Serum Level of Specific Haptoglobin Isoforms in Small Cell Lung Cancer." Curr Proteomics **7**(1): 49-65.

Shen, Y., N. Tolic, et al. "Blood peptidome-degradome profile of breast cancer." PLoS One **5**(10): e13133.

Shi, S. R., R. J. Cote, et al. (2001). "Antigen retrieval techniques: current perspectives." J Histochem Cytochem **49**(8): 931-7.

Shi, Z. Z., J. W. Zhang, et al. (2007). "What we know about ST13, a co-factor of heat shock protein, or a tumor suppressor?" J Zhejiang Univ Sci B **8**(3): 170-6.

Shibayama, T., H. Ueoka, et al. (2001). "Complementary roles of pro-gastrin-releasing peptide (ProGRP) and neuron specific enolase (NSE) in diagnosis and prognosis of small-cell lung cancer (SCLC)." Lung Cancer **32**(1): 61-9.

Shintaku, H. (2002). "Disorders of tetrahydrobiopterin metabolism and their treatment." Curr Drug Metab **3**(2): 123-31.

Shiwa, M., Y. Nishimura, et al. (2003). "Rapid discovery and identification of a tissue-specific tumor biomarker from 39 human cancer cell lines using the SELDI ProteinChip platform." Biochem Biophys Res Commun **309**(1): 18-25.

Sidransky, D. and M. Hollstein (1996). "Clinical implications of the p53 gene." Annu Rev Med **47**: 285-301.

Siegfried, J. M., L. A. Weissfeld, et al. (1997). "Association of immunoreactive hepatocyte growth factor with poor survival in resectable non-small cell lung cancer." Cancer Res **57**(3): 433-9.

Singhal, S., D. Miller, et al. (2008). "Gene expression profiling of non-small cell lung cancer." Lung Cancer **60**(3): 313-24.

Sone, S., S. Takashima, et al. (1998). "Mass screening for lung cancer with mobile spiral computed tomography scanner." Lancet **351**(9111): 1242-5.

Soo, E. T., G. W. Yip, et al. (2008). "Heat shock proteins as novel therapeutic targets in cancer." In Vivo **22**(3): 311-5.

Spindel, E. R. (1996). "Roles of bombesin-like peptides in lung development and lung injury." Am J Respir Cell Mol Biol **14**(5): 407-8.

Spiro, R. G. (2002). "Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds." Glycobiology **12**(4): 43R-56R.

Spratlin, J. L., N. J. Serkova, et al. (2009). "Clinical applications of metabolomics in oncology: a review." Clin Cancer Res **15**(2): 431-40.

Stang, A., H. Pohlabeln, et al. (2006). "Diagnostic agreement in the histopathological evaluation of lung cancer tissue in a population-based case-control study." Lung Cancer **52**(1): 29-36.

Stetler-Stevenson, W. G., S. Aznavoorian, et al. (1993). "Tumor cell interactions with the extracellular matrix during invasion and metastasis." Annu Rev Cell Biol **9**: 541-73.

Streit, F., V. W. Armstrong, et al. (2002). "Rapid liquid chromatography-tandem mass spectrometry routine method for simultaneous determination of sirolimus, everolimus, tacrolimus, and cyclosporin A in whole blood." Clin Chem **48**(6 Pt 1): 955-8.

Subramanian, J., D. Morgensztern, et al. "Vascular endothelial growth factor receptor tyrosine kinase inhibitors in non-small-cell lung cancer." Clin Lung Cancer **11**(5): 311-9.

Suh, S. Y. and H. Y. Ahn (2007). "Lactate dehydrogenase as a prognostic factor for survival time of terminally ill cancer patients: a preliminary study." Eur J Cancer **43**(6): 1051-9.

Sun, S. S., J. F. Hsieh, et al. (2000). "Cytokeratin fragment 19 and squamous cell carcinoma antigen for early prediction of recurrence of squamous cell lung carcinoma." Am J Clin Oncol **23**(3): 241-3.

Sung, H. J., J. M. Ahn, et al. "Identification and validation of SAA as a potential lung cancer biomarker and its involvement in metastatic pathogenesis of lung cancer." J Proteome Res **10**(3): 1383-95.

Suzuki, K., Y. Ito, et al. (2006). "Serum heat shock protein 70 levels and lung cancer risk: a case-control study nested in a large cohort study." Cancer Epidemiol Biomarkers Prev **15**(9): 1733-7.

Swensen, S. J., J. R. Jett, et al. (2005). "CT screening for lung cancer: five-year prospective experience." Radiology **235**(1): 259-65.

Swensen, S. J., J. R. Jett, et al. (2003). "Lung cancer screening with CT: Mayo Clinic experience." Radiology **226**(3): 756-61.

Szafranska, A. E., T. S. Davison, et al. (2008). "Accurate molecular characterization of formalin-fixed, paraffin-embedded tissues by microRNA expression profiling." J Mol Diagn **10**(5): 415-23.

Szutowicz, E. and R. Dziadziuszko "Quantitative immunohistochemistry in lung cancer: clinical perspective." Folia Histochem Cytobiol **48**(1): 7-11.

Szymezak, J., E. Lavalard, et al. (2009). "[Incidence of sample storage temperature on HbA 1c determination by high performance liquid chromatography method]." Ann Biol Clin (Paris) **67**(6): 705-10.

Tabassum, U., O. Reddy, et al. "Elevated serum haptoglobin is associated with clinical outcome in triple-negative breast cancer patients." Asian Pac J Cancer Prev **13**(9): 4541-4.

Taillade, L., F. Penault-Llorca, et al. (2007). "Immunohistochemichal expression of biomarkers: a comparative study between diagnostic bronchial biopsies and surgical specimens of non-small-cell lung cancer." Ann Oncol **18**(6): 1043-50.

Tamada, M., M. Suematsu, et al. "Pyruvate kinase m2: multiple faces for conferring benefits on cancer cells." Clin Cancer Res **18**(20): 5554-61.

Tammen, H., I. Schulte, et al. (2005). "Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display." Proteomics **5**(13): 3414-22.

Tamura, M., Y. Ohta, et al. (2001). "Plasma VEGF concentration can predict the tumor angiogenic capacity in non-small cell lung cancer." Oncol Rep **8**(5): 1097-102.

Tanaka, F., K. Yoneda, et al. (2009). "Circulating tumor cell as a diagnostic marker in primary lung cancer." Clin Cancer Res **15**(22): 6980-6.

Tauler, J., E. Zudaire, et al. "hnRNP A2/B1 modulates epithelial-mesenchymal transition in lung cancer cell lines." Cancer Res **70**(18): 7137-47.

Taylor, C. R. and R. M. Levenson (2006). "Quantification of immunohistochemistry--issues concerning methods, utility and semiquantitative assessment II." Histopathology **49**(4): 411-24.

Taylor, D. D. and C. Gercel-Taylor (2008). "MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer." Gynecol Oncol **110**(1): 13-21.

Terry, J., S. Leung, et al. "Optimal immunohistochemical markers for distinguishing lung adenocarcinomas from squamous cell carcinomas in small tumor samples." Am J Surg Pathol **34**(12): 1805-11.

Teti, A. (1992). "Regulation of cellular functions by extracellular matrix." J Am Soc Nephrol **2**(10 Suppl): S83-7.

Thomas, C. F., Jr., H. D. Tazelaar, et al. (2001). "Typical and atypical pulmonary carcinoids : outcome in patients presenting with regional lymph node involvement." Chest **119**(4): 1143-50.

Thunnissen, F. B. (2003). "Sputum examination for early detection of lung cancer." J Clin Pathol **56**(11): 805-10.

Timar, J., Y. Q. Chen, et al. (1992). "The lipoxygenase metabolite 12(S)-HETE promotes alpha IIb beta 3 integrin-mediated tumor-cell spreading on fibronectin." Int J Cancer **52**(4): 594-603.

Tirumalai, R. S., K. C. Chan, et al. (2003). "Characterization of the low molecular weight human serum proteome." Mol Cell Proteomics **2**(10): 1096-103.

Tockman, M. S., J. L. Mulshine, et al. (1997). "Prospective detection of preclinical lung cancer: results from two studies of heterogeneous nuclear ribonucleoprotein A2/B1 overexpression." Clin Cancer Res **3**(12 Pt 1): 2237-46.

Toyama, A., H. Nakagawa, et al. (2011). "Deglycosylation and label-free quantitative LC-MALDI MS applied to efficient serum biomarker discovery of lung cancer." Proteome Sci **9**: 18.

Travis, W. D., L. B. Travis, et al. (1995). "Lung cancer." Cancer **75**(1 Suppl): 191-202.

Tripathi, B. N., I. Bhatt, et al. (2009). "Peroxiredoxins: a less studied component of hydrogen peroxide detoxification in photosynthetic organisms." Protoplasma **235**(1-4): 3-15.

Tristan, C., N. Shahani, et al. "The diverse functions of GAPDH: views from different subcellular compartments." Cell Signal **23**(2): 317-23.

Tuder, R. M., I. Petrache, et al. (2003). "Apoptosis and emphysema: the missing link." Am J Respir Cell Mol Biol **28**(5): 551-4.

Ueda, Y., T. Yamagishi, et al. (2004). "A novel low molecular weight VEGF receptor-binding antagonist, VGA1102, inhibits the function of VEGF and in vivo tumor growth." Cancer Chemother Pharmacol **54**(1): 16-24.

Unwin, R. D. "Quantification of proteins by iTRAQ." Methods Mol Biol **658**: 205-15.

Vayrynen, J. P., J. Vornanen, et al. "Serum MMP-8 levels increase in colorectal cancer and correlate with disease course and inflammatory properties of primary tumors." Int J Cancer **131**(4): E463-74.

Vicari, A. P. and C. Caux (2002). "Chemokines in cancer." Cytokine Growth Factor Rev **13**(2): 143-54.

Virmani, A. K., A. Rathi, et al. (2001). "Aberrant methylation of the adenomatous polyposis coli (APC) gene promoter 1A in breast and lung carcinomas." Clin Cancer Res **7**(7): 1998-2004.

Voelkel, N. F., R. W. Vandivier, et al. (2006). "Vascular endothelial growth factor in the lung." Am J Physiol Lung Cell Mol Physiol **290**(2): L209-21.

Voellmy, R. (2006). "Feedback regulation of the heat shock response." Handb Exp Pharmacol(172): 43-68.

Wagner, P. L., N. Kitabayashi, et al. (2009). "Combined small cell lung carcinomas: genotypic and immunophenotypic analysis of the separate morphologic components." Am J Clin Pathol **131**(3): 376-82.

Walenkamp, A. M., G. S. Sonke, et al. (2009). "Clinical and therapeutic aspects of extrapulmonary small cell carcinoma." Cancer Treat Rev **35**(3): 228-36.

Walser, T., X. Cui, et al. (2008). "Smoking and lung cancer: the role of inflammation." Proc Am Thorac Soc **5**(8): 811-5.

Wang, H., Z. Zhu, et al. (2008). "[The value of serum tumor marker CA125 and CEA in the diagonosis of non-small cell lung cancer.]." Zhongguo Fei Ai Za Zhi **11**(1): 97-100.

Wang, L. B., S. Zheng, et al. (2005). "Expression of ST13 in colorectal cancer and adjacent normal tissues." World J Gastroenterol **11**(3): 336-9.

Wang, M., J. You, et al. (2008). "Label-free mass spectrometry-based protein quantification technologies in proteomic analysis." Brief Funct Genomic Proteomic **7**(5): 329-39.

Wang, P., J. R. Whiteaker, et al. (2009). "The evolving role of mass spectrometry in cancer biomarker discovery." Cancer Biol Ther **8**(12): 1083-94.

Waris, G. and H. Ahsan (2006). "Reactive oxygen species: role in the development of cancer and various chronic conditions." J Carcinog **5**: 14.

Weckwerth, W. (2003). "Metabolomics in systems biology." Annu Rev Plant Biol **54**: 669-89.

Weighardt, F., G. Biamonti, et al. (1996). "The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism." Bioessays **18**(9): 747-56.

Weigmann, B., R. K. Franke, et al. "Immunotherapy in autoimmune type 1 diabetes." Rev Diabet Stud **9**(2-3): 68-81.

Wen, F., A. Shen, et al. "Higher expression of the heterogeneous nuclear ribonucleoprotein k in melanoma." Ann Surg Oncol **17**(10): 2619-27.

Wender, R., E. T. Fontham, et al. (2013). "American Cancer Society lung cancer screening guidelines." CA Cancer J Clin **63**(2): 107-17.

Westermeier, R. (2006). "Sensitive, quantitative, and fast modifications for Coomassie Blue staining of polyacrylamide gels." Proteomics **6 Suppl 2**: 61-4.

Widner, B., D. Fuchs, et al. (2001). "Does disturbed homocysteine and folate metabolism in depression result from enhanced oxidative stress?" J Neurol Neurosurg Psychiatry **70**(3): 419.

Wieskopf, B., C. Demangeat, et al. (1995). "Cyfra 21-1 as a biologic marker of non-small cell lung cancer. Evaluation of sensitivity, specificity, and prognostic role." Chest **108**(1): 163-9.

Wiesner, A. (2004). "Detection of tumor markers with ProteinChip technology." Curr Pharm Biotechnol **5**(1): 45-67.

Wistuba, II, C. Behrens, et al. (2000). "High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints." Cancer Res **60**(7): 1949-60.

Wolters, D. A., M. P. Washburn, et al. (2001). "An automated multidimensional protein identification technology for shotgun proteomics." Anal Chem **73**(23): 5683-90.

Woo, I. S., M. J. Park, et al. (2004). "Expression of placental growth factor gene in lung cancer." Tumour Biol **25**(1-2): 1-6.

Wood, D. E., G. A. Eapen, et al. "Lung cancer screening." J Natl Compr Canc Netw **10**(2): 240-65.

Wu, W. W., G. Wang, et al. (2006). "Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF." J Proteome Res **5**(3): 651-8.

Wu, X., V. C. Lee, et al. (2009). "Chemokine receptors as targets for cancer therapy." Curr Pharm Des **15**(7): 742-57.

Wysocki, V. H., K. A. Resing, et al. (2005). "Mass spectrometry of peptides and proteins." Methods **35**(3): 211-22.

Yanagawa, T., T. Ishikawa, et al. (1999). "Peroxiredoxin I expression in human thyroid tumors." Cancer Lett **145**(1-2): 127-32.

Yanagawa, T., S. Iwasa, et al. (2000). "Peroxiredoxin I expression in oral cancer: a potential new tumor marker." Cancer Lett **156**(1): 27-35.

Ye, B., D. W. Cramer, et al. (2003). "Haptoglobin-alpha subunit as potential serum biomarker in ovarian cancer: identification and characterization using proteomic profiling and mass spectrometry." Clin Cancer Res **9**(8): 2904-11.

Yu, K. H., C. G. Barry, et al. (2009). "Stable isotope dilution multidimensional liquid chromatography-tandem mass spectrometry for pancreatic cancer serum biomarker discovery." J Proteome Res **8**(3): 1565-76.

Yu, Z., G. Kastenmuller, et al. "Differences between human plasma and serum metabolite profiles." PLoS One **6**(7): e21230.

Zang, E. A. and E. L. Wynder (1996). "Differences in lung cancer risk between men and women: examination of the evidence." J Natl Cancer Inst **88**(3-4): 183-92.

Zeng, X., B. L. Hood, et al. "Lung Cancer Serum Biomarker Discovery Using Label-Free Liquid Chromatography-Tandem Mass Spectrometry." J Thorac Oncol.

Zhan, X., H. Dong, et al. "[Expression and clinical significance of SHP2 in the tumor tissues of smokers with lung cancer]." Zhongguo Fei Ai Za Zhi **13**(9): 877-81.

Zhang, X. Z., Z. F. Xiao, et al. (2009). "Triosephosphate isomerase and peroxiredoxin 6, two novel serum markers for human lung squamous cell carcinoma." Cancer Sci **100**(12): 2396-401.

Zhang, Z., R. C. Bast, Jr., et al. (2004). "Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer." Cancer Res **64**(16): 5882-90.

Zhao, C., L. Annamalai, et al. (2007). "Circulating haptoglobin is an independent prognostic factor in the sera of patients with epithelial ovarian cancer." Neoplasia **9**(1): 1-7.

Zheng, S., Y. Chang, et al. "Expression of KISS1 and MMP-9 in non-small cell lung cancer and their relations to metastasis and survival." Anticancer Res **30**(3): 713-8.

Zhou, J., D. C. Allred, et al. (2001). "Differential expression of the early lung cancer detection marker, heterogeneous nuclear ribonucleoprotein-A2/B1 (hnRNP-A2/B1) in normal breast and neoplastic breast cancer." Breast Cancer Res Treat **66**(3): 217-24.

Zhou, J., J. L. Mulshine, et al. (1996). "Purification and characterization of a protein that permits early detection of lung cancer. Identification of heterogeneous nuclear ribonucleoprotein-A2/B1 as the antigen for monoclonal antibody 703D4." J Biol Chem **271**(18): 10760-6.

Zhu, W., J. W. Smith, et al. "Mass spectrometry-based label-free quantitative proteomics." J Biomed Biotechnol **2010**: 840518.

Zimmermann, M., S. Nickl, et al. "Discrimination of clinical stages in non-small cell lung cancer patients by serum HSP27 and HSP70: a multi-institutional case-control study." Clin Chim Acta **413**(13-14): 1115-20.

Zochbauer-Muller, S., A. F. Gazdar, et al. (2002). "Molecular pathogenesis of lung cancer." Annu Rev Physiol **64**: 681-708.

APPENDIX

Protein Name	Protein Score	Peptide Count	Protein Score	Confidence Score	% Sequence	ID Method
hnRNPA2B1	N/A	N/A	N/A	0.004	20.10%	MALDI TOF
Pyruvate kinase 3	N/A	N/A	N/A	0	16.60%	MALDI TOF
Hsc70-interacting Protein	101	11	100.00%	N/A	N/A	MALDI TOF/TOF
Perox1	187	6	100.00%	N/A	N/A	MALDI TOF/TOF
Perox2	N/A	N/A	N/A	0	25.10%	MALDI TOF
Perox3	81	3	99.89%	N/A	N/A	MADLI TOF/TOF
Triosephosphate isomerase	147	2	100.00%	N/A	N/A	MADLI TOF/TOF
Lactate Dehydrogenase	170	12	100.00%	N/A	N/A	MADLI TOF/TOF
GAPDH	186	10	100.00%	N/A	N/A	MADLI TOF/TOF

Figure A1: Mass Spectrometry Identifications for Conditioned Media 2D-DIGE Experiments: MALDI TOF/TOF data shows protein score (> 60 is a score of high confidence), peptide count, and protein score (% chance of a correct identification). MALDI TOF data shows percentage of sequence covered and confidence score (with 0 showing the highest degree of confidence).

Protein Name	Protein Score	Peptide Count	Protein Score	ID Method
Haptoglobin (Isoform 1)	192	9	100.00%	MALDI TOF/TOF
Haptoglobin (Isoform 2)	86	7	99.99%	MALDI TOF/TOF
Haptoglobin (Isoform 3)	168	7	100.00%	MALDI TOF/TOF
Haptoglobin (Isoform 4)	224	8	100.00%	MALDI TOF/TOF
Haptoglobin (Isoform 5)	77	7	99.97%	MALDI TOF/TOF
Haptoglobin (Isoform 6)	191	9	100.00%	MALDI TOF/TOF

Figure A2: Mass Spectrometry Identifications for 2D-DIGE Analysis of Immunodepleted Serum Experiments: MALDI TOF/TOF data shows protein score (> 60 is a score of high confidence), peptide count, and protein score (% chance of a correct identification).