Proteomic Characterisation of Clonal Populations from a Human Lung Carcinoma Cell Line

A thesis submitted for the degree of Ph.D.

by

Shane Kelly, B.Sc. (Hons)

September 2016

Under the supervision of

Dr. Paula Meleady

&

Prof. Martin Clynes

National Institute for Cellular Biotechnology

School of Biotechnology

Dublin City University

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of PhD 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.: 53318681

Date: _____

This thesis is dedicated to my loving parents

Christy and Mary

Acknowledgements:

Firstly, I would like to thank Dr. Paula Meleady and Prof. Martin Clynes for giving me the opportunity to carry out this Ph.D. You both took quite a chance on a hobbit working in the basement storeroom and offering him a chance to pursue a career in research.

Paula, you were always patient and kind, ready to guide me in the right direction when projects didn't quite go as planned. Your patience during the write-up period was remarkable, and I won't forget how you gave your time so freely when I eventually had something for you to correct. Without you I would truly never have finished, thank you for everything.

Martin, it was you who gave me that initial start here in the NICB as a technical assistant. During that time, I felt your gentle guidance towards a more research based career, which I had initially thought was not for me. I have since learned I very much enjoy this type of work, and can only thank you for the opportunity to realise that. You have built a very special place here at the NICB, where the atmosphere is friendly, cooperative and encouraging. To have such a great environment to work in as an initial experience in research is rare and I am truly grateful to you for giving me this chance.

Michael Henry, what can I say. I may have driven you slightly mad at times (and vice versa), but you showed great generosity with the time you gave to me from when I first started working with you as a technical assistant. Your scientific knowledge and mass spec genius were invaluable, and helped form the bedrock upon which my Ph.D. was based.

A huge thank you to the people behind the scenes, pulling all the strings. Mairead "The Master of Coin" Callan for helping with every question I ever had, and keeping the centre running in general. Carol "The Hand of the King" McNamara for all her hard work in implementing Martin's decrees, and for your willingness to help with anything and everything. A big thank you to Gillian Smith for running the prep room and storeroom so well. We are very lucky to have you here and the NICB would not be the same without you.

A big thanks goes to Finbarr for all the help you gave so freely, particularly with the confocal imaging. Every thesis need some fancy pictures and they wouldn't be in here without you! Thanks to Nga for all your guidance with the overexpression experiments, it was great to learn a new discipline. Special thanks also to Annette Linge for all her help in the early days, teaching me all the various techniques which were critical to the Ph.D.

A very special thanks goes to all the office gang. It's great to work with people you genuinely get on with and who encouraging you when the will to work goes downhill. You guys are a large part of what makes this such a great place to work. Paul and Alan G, you guys were here since the beginning and we have chatted about everything under the sun at this stage during our legendary lunchtime debates! Also to Clair "Cersei" Gallagher, our inappropriate ragamuffin conversations shall never be repeated, but fondly remembered.

Gemma, Edel, Andrew and Alan C, thanks for all the laughs and great company. Take some comfort in this: if I can complete a Ph.D, you guys will have no problems!
Mark, I hope your prophecy about my death is not fulfilled but if it does come to pass, you must go into the west and remain…. Galadriel. Kevin, I don't know what the endgame of your spy mission to Ireland is but I'm very glad you ended up here, you helped so much I really appreciate it a great deal. You guys are friends for life. A very special thanks to Orla for keeping me sane and going above and beyond when I was losing my mind, I owe you a very great debt.

Finally, to my parents for accepting their son "still hasn't finished school". You have let me do my own thing, but I couldn't have done it without your support and understanding along the way. I hope you are proud that you have a (sort of) doctor in the family. Thanks for everything, and also for providing all the endless entertaining stories of your latest exploits. Thanks also to my sister Grace (and Andy!) for your support over the years. You've have always set the bar raaaather high and I hope you are proud of your little bro.

The Road goes ever on and on
Down from the door where it began.
Now far ahead the Road has gone,
And I must follow, if I can,
Pursuing it with eager feet,
Until it joins some larger way,
Where many paths and errands meet.
And whither then? I cannot say.

~B. B.

Table of Contents

Abbreviations:

2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
ACN	Acetonitrile
AD	Adenocarcinoma
ARF	Actin Retrograde Flow
BSA	Bovine serum albumin
C18	Octadecylsilyl
CAMs	Cell adhesion molecules
CID	Collision-induced dissociation
Co-IP	Co-Immunoprecipitation
CSC	Cancer stem cell
c-SCLC	Combined small cell lung carcinoma
CTTN	Cortactin
CV	Coefficient of variation
dH2O	Distilled water
DLAT	Dihydrolipoyllysine-residue acetyltransferase
DMSO	Dimethyl sulfoxide
DSG3	Desmoglein 3
ECL	Enhanced chemiluminescence
ECM	Extracellular matrix
ED	Effector domain
EGFR	Epidermal growth factor receptor
EMT	Epithelial to mesenchymal transition
ESCC	Esophageal squamous cell carcinoma
ESI	ElectroSpray Ionisation
GTP	Guanosine triphosphate
HGF	Hepatocyte Growth Factor
HPLC	High pressure liquid chromatography
ICAT	Isotope-Coded Affinity Tags
ICC	Immunocytochemical
IPG	Immobilised pH gradient
IT	Ion trap
iTRAQ	Isobaric Tag for Relative and Absolute Quantification

LCC	Large cell Carcinoma
LC-MS	Liquid Chromatography-Mass Spectrometry
LCNEC	Large cell neuroendocrine carcinoma
LTQ	Linear trap quadropole
MALDI	Matrix Assisted Laser Desorption Ionisation
MARCKS	Myristoylated alanine-rich C-kinase substrate
MMP	Matrix metalloproteinases
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NEBs	Neuroepithelial bodies
NSCLC	Non-small cell lung cancer
ORF	Open reading frame
PBS	Phosphate buffered saline
PCA	Principal component analysis
PG	Junction Plakoglobin
pI	Isoelectric point
PKC	Protein Kinase-C
PNECs	Pulmonary Neuroendocrine cells
PSGL-1	P-selectin glycoprotein ligand-1
Q	Quadropole mass analysers
qPCR	Quantitative PCR
RP	Reverse phase
RQ	Relative quantity
SCC	Squamous cell carcinoma
SCLC	Small cell lung cancer
SEMG1	Semenogelin-1
SILAC	Stable isotope labelling with amino acids in cell culture
siRNA	Small interfering RNA
STRN3	Striatin-3
TBS	Tris-buffered saline
TFA	Trifluoroacetic acid
TGF- β	Transforming growth factor- β
TKIs	Tyrosine kinase inhibitors
TOF	Time-of-flight

TSGs	Tumour suppressor genes
UHP	Ultrapure water

List of Tables:

List of Figures:

Abstract:

Title: Proteomic Characterisation of Clonal Populations from a Human Lung Carcinoma Cell Line

Author: Shane Kelly

Lung cancer is the leading cause of cancer related deaths worldwide. Although treatment strategies have improved, overall survival has not significantly increased. Studies have revealed a heterogeneous subclonal architecture to lung cancer, which contributes to the persistence of the disease by allowing the tumour to metastasise to distant sites, utilising the different phenotypic characteristics of the heterogeneous population.

The DLKP cell line was established from a lymph node metastasis of a primary lung tumour and is described as a poorly differentiated squamous lung carcinoma cell line. This cell line is comprised of three distinct subpopulations: DLKP-SQ, DLKP-M and DLKP-I, each with well-defined phenotypes. *In vitro* studies have shown DLKP-M to be extremely invasive relative to DLKP-SQ, whereas the latter cell line shows high anoikis resistance relative to DLKP-M. As the cells originate from the same tumour, these varied phenotypic characteristics make DLKP a particularly useful model for the study of cellular heterogeneity in lung cancer.

Quantitative label-free LC-MS/MS was used to analyse the DLKP cell line and its subpopulations at a proteomic level. This resulted in a cohort of significantly differentially expressed proteins between the cell lines. From this, three differentially expressed proteins; Shootin-1 (highly expressed in DLKP-SQ), MARCKS (highly expressed in DLKP-M), and Desmoglein-3 (highly expressed in DLKP-I) were chosen for a series of follow up studies. The expression patterns for these proteins were validated, and functional cell-based assays were carried out to establish the role of each protein in their respective clone. MARCKS and Desmoglein-3 were found to reduce cell motility *in vitro* upon transient protein knockdown using siRNA. Co-immunoprecipitation of Desmoglein-3 revealed a potential interaction with mitochondrial proteins, a novel finding in lung cancer. Shootin-1, a relatively uncharacterised protein, was also found to reduce cell motility upon protein knockdown. Shootin-1 was also found to have novel potential binding partners, such as Semenogelin-1. Shootin-1 and Desmoglein-3 expression were affected by co-culture of the clonal subpopulations, highlighting the heterogeneous nature of the original DLKP cell line and the influences this has on protein expression.

Chapter 1. Introduction

1.1. Lung Development

The primary function of the lungs is to exchange carbon dioxide in the cardiovascular system with oxygen in the external environment. This organ also carries out important secondary processes such as the regulation of blood pressure, and the filtering of blood for gas bubbles and blood clots. The entire inner surface of the lung is exposed to the external environment, and therefore faces many challenges when dealing with associated problems which arise from this. These include temperature changes, irritation by allergens, and constant bombardment of particulate matter. As a result, the lungs are a complex system comprised of multiple cell lineages, capable of carrying out all the functions of the lungs while dealing with the challenges mentioned above.

The lungs are an unnecessary organ *in utero*; however they must be developed to such an extent that they function immediately following birth. The development of the lung is a complex multistage process which involves growth factors, transcription factors and the extracellular matrix. Lung development can be divided into five main stages.

The *embryonic stage* of lung development begins with the appearance of a groove in the ventral lower pharynx. A bud develops in the ventral wall of the foregut and is known as the true lung primordium. The location of this bud is determined by signals from the surrounding mesenchyme, in particular from fibroblast growth factors (Al Alam et al. 2015). This bud undergoes further subdivision into the two main bronchi and the asymmetry of the bronchi present in adults is developed at this stage.

The *pseudoglandular stage* takes place from the 6^{th} to the 16^{th} week of prenatal lung development. Epithelial tubes lined with cuboidal epithelial cells undergo repeated branching to form primordial airways, and the lungs resemble exocrine glands at this stage. Separation occurs between the epithelium and the mesoderm by the development of a basement membrane. The bronchial tree is coated by cuboidal epithelial cells, which are the precursors of the ciliated epithelium and secretory cells. Neuroendocrine cells are the first type of cells to differentiate from the lung epithelium early on during this stage. They grow in clusters known as neuroepithelial bodies (NEBs) and are usually found at junctions of broncioalveolar ducts and airway bifurcations (Linnoila 2006, Song et al. 2012). This is followed by the development of ciliated cells, then

mucus secreting cells, and submucosal glands. At this stage, basal cells begin to differentiate; however mature basal cells are the last mature cell type to appear in the epithelium.

The *canalicular stage* (16^{th} week-26^{th} week) is characterised by the development of basic structures of the gas-exchanging parts of the lungs, vascularisation, and the formation of primordial alveoli called the terminal saccules. Invasion of capillaries into the mesenchyme occurs at this stage, and forms the foundation for the later exchange of gases. Differentiation of type I and type II alveolar epithelial cells (pneumocytes) occurs during this stage and arise from the undifferentiated cuboidal epithelial cells. From the cubic type II pneumocytes develop the flattened type I pneumocytes, and this is the beginning of alveolar epithelium development (Desai, Brownfield and Krasnow 2014).

The *saccular stage* (26^{th} week – Birth) is the time during which alveolar ducts and air sacs are developed. These air sacs develop on the terminal bronchioli which are the last subdivision of the passages which supply air to the lungs. Capillaries bulge into the thin epithelial lining of these sacs which are lined by type I and type II pneumocytes. Type I cells are numerous, covering ~95% of the alveolar surface. Type I cells are squamous and unable to replicate, and are involved in the process of gas exchange between alveoli and blood. Type II cells are scattered across the alveolar surface and represent ~ 5% of the total surface cell population. They function to secrete pulmonary surfactant which is critical to reduce the surface tension within alveoli allowing the lungs to expand, and to reduce fluid accumulation and keep epithelial surfaces dry (Wright 2004). In the event of damage to the epithelium, type II cells can differentiate into type I to compensate.

The *alveolar stage* ranges from the later stages of fetal development up to approximately 8 years of age. At birth, roughly one third of the alveoli in the lungs are fully developed. The remaining underdeveloped alveoli are present in their beginning forms; however these will continue to enlarge and expand until completion at approximately 3 years of age. New alveoli continue to be generated and increase up to 10 years of age.

1.1.1. Epithelial Cells of the Lung:

There are a number of different cell types in the developed lung which arise from what is believed to be a common stem cell. The upper respiratory airway is classified as the *ciliated pseudostratified columnar epithelium*, and is so called due to the multiple cell types which comprise this epithelial layer. Though several cell types are present in this layer, they all make contact with the basal layer, and it is therefore classified as pseudostratified. There are three main epithelial cells which make up this layer: Goblet cells; ciliated cells; and basal cells.

Goblet cells and ciliated cells are terminally differentiated epithelial cell types. They are located in the upper respiratory airways and work together to protect and clean the lung from environmental particulate matter as well as microorganisms such as fungal spores, bacteria and viruses. Goblet cells are columnar epithelial cells and are so called due to their shape which resembles a wine goblet. They contain membrane-bound mucous granules and their primary function is to secrete mucous as a method of maintaining the lung. The mucous helps to trap particles and pathogens which enter the airway, and also functions to maintain moisture in the epithelial layer. Goblet cells work together with ciliated cells, which are also columnar epithelial cells containing cilia protrusions. They sweep the mucous secretions of the goblet cells which have trapped undesirable matter and move it up and out of the airway towards the mouth, where it is swallowed or expelled from the body (Stannard and O'Callaghan 2006).

Basal cells are small cuboidal cells which are attached to the basal lamina. It was originally thought these cells have a primary role in adhering columnar epithelial cells to the basal lamina. Basal cells attach to the goblet and ciliated cells through desmosomes, and they also attach directly to the basal lamina through hemi-desmosomes. Both goblet and ciliated cells do not form desmosomal structures with the basal lamina, therefore it was thought that basal cells serve to anchor the *ciliated pseudostratified columnar epithelium* to the basal lamina through these interactions (Evans and Plopper 1988).

Basal cells are also known to be a stem cell population of the lung epithelium. These cells differentiate and mature last during the pseudoglandular stage of lung development as previously mentioned; therefore it is clear that they are not the

precursors of the other epithelial cell types. However, studies have shown that they have the ability to differentiate into other epithelial cell types under certain circumstances. A number of these studies examined the role of basal cells in denuded sections of the upper airway in rat tracheas. They found that basal cells are capable of proliferating and restoring a fully differentiated mucociliary and functional epithelium (Randell et al. 1991, Hajj et al. 2007, Rock, Randell and Hogan 2010).

Club cells are found in the small airways (bronchioles) of the lung and are known as exocrine cells. They function to protect the ciliated epithelium through the secretion of various substances such as glycosaminoglycans, and others similar to the surfactant secreted by type II alveolar cells. Club cells have been found to be associated with neuroendocrine cells. After the induced death of club cells by naphthalene injury, the only remaining club cells were found to be located bear neuroepithelial bodies (NEBs) at the broncioalveolar duct junction. These cells were found to have the ability to restore damaged lung epithelium, and the study led to the hypothesis that NEBs aid in club cell regeneration through paracrine signalling, and maintain a stem cell niche in this way (Reynolds et al. 2000a, Reynolds et al. 2000b).

Neuroendocrine cells are known as Pulmonary Neuroendocrine cells (PNECs) when they are present in the lung. As previously mentioned, these are the first type cells to differentiate in the developing lung. Here, they appear as solitary cells or in clusters as neuroepithelial bodies (NEBs). They are a distinct cell population of airway epithelial cells which display both endocrine and paracrine secretory mechanisms associated with nerve fibres. They synthesise and secrete a number of bioactive substances such as serotonin, calcitonin gene-related peptide and the mitogen bombesin (Van Lommel 2001). PNECs acting via their secretions are thought to function as local regulators of lung growth and differentiation, and bombesin has been shown to be involved in lung maturation (Aguayo et al. 1994).

PNECs have been found to share morphological features with cells of the nervous system and also cells from neuroendocrine carcinomas such as small cell lung carcinoma (Gu et al. 2014). A common protein expressed in neuronal cells and PNECs is ASH1, which is known to be essential for neuronal cell commitment and differentiation. Selectively expressed by PNECs, this protein has been has been found to be expressed by a range of lung cancers with neuroendocrine features (Westerman et

al. 2002). Inhibition of ASH1 expression has been shown to suppress growth of lung cancer cells which express it, potentially highlighting it as a therapeutic target (Osada et al. 2005).

1.2. Lung Cancer

Lung cancer is the fourth most common cancer in Ireland with approximately 2000 new cases diagnosed each year, and is the leading cause of cancer related death in this country (www.ncri.ie). In the United States, there were approximately 221,200 new cases of lung cancer diagnosed, and accounts for 13% all new cancers. Each year, more people die of lung cancer than of colon, breast, and prostate cancers combined.

Lung cancer is the abnormal growth of cells in one or both of the lungs. The abnormal cells are termed 'cancerous', and are caused by DNA damage which trigger the cells to divide continuously. Lung cancer cells usually originate in the epithelial lining of the lung and grow abnormally fast relative to non-cancerous cells and do not undergo apoptosis like healthy cells. Highly proliferative cancer cells eventually become so abundant they form tumours within the lung tissue. The increasing mass of these tumours restrict the ability of the lungs to provide the bloodstream with oxygenated blood, eventually resulting in death. In addition, cancerous cells can metastasise to other parts of the body, forming new tumours and significantly contributing to disease progression.

Lung cancer is classified based on histological grading by microscopic analysis. The classification of lung cancer type is crucial to predict the outcome of the disease and to determine the proper management and therapeutic strategies. Two broad classes of lung cancer have been distinguished: **small cell lung cancer** (SCLC) and **non-small cell lung cancer** (NSCLC). These two classes refer to primary lung cancer types and both behave quite differently. SCLC accounts for approximately 12% of all lung cancers, and is so called because the cancer cells are small. SCLC makes up approximately 13-15% of lung cancer cases worldwide and is characterised as being highly malignant with rapid growth and widespread metastasis. Treatment for this type of cancer is usually chemotherapy combined with radiotherapy. Over 70% of patients who present with SCLC have already developed secondary metastasis and despite extensive research over the last 30 years, no significant improvement in disease outcome has been achieved with the 5-year survival rate at < 7% (Jemal et al. 2010).

NSCLC makes up 87% of all lung cancers and this type is further subdivided into three subtypes of primary lung cancer which all behave in a similar way: Adenocarcinoma (AD); squamous cell carcinoma (SCC); and large cell carcinoma (LCLC). NSCLC is usually treated by surgical resection or a combination of surgery and chemotherapy depending on the stage of the disease, of which there are four. The five year survival rate ranges from 65% for Stage 1 to 7% for Stage 4. This form of lung cancer is the leading cause of cancer death, and many patients are beyond the curable stage at diagnosis.

1.2.1. Classification of Lung Cancer

Carcinomas may be described as well-differentiated, moderately-differentiated and poorly differentiated, and these classifications are based on microscopic evaluation of the cells by a histopathologist. Differentiation refers to how developed the cancer cells are within a tumour with regard to their morphology and growth rate, and how well they resemble their tissue of origin. Well-differentiated cells resemble normal surrounding cells, spreading and growing at a relatively slow rate. Tumours of this type tend to grow slowly and have a better prognosis. Poorly-differentiated cells bear minimal resemblance to the cells from which they arose and grow at a much faster rate. These tumour types are the most aggressive and have a poor prognosis. Moderately-differentiated tumours have features which are intermediate to the differentiation categories mentioned above. Differentiation classification refers to the staging system mentioned previously with Stage 1 being well-differentiated, Stage 2 being moderately-differentiated and Stage 3-4 being poorly differentiated.

As previously mentioned, there are two broad classes of lung cancer: SCLC and NSCLC. Within NSCLC there are three further subtypes: Adenocarcinoma; Squamous cell carcinoma; and Large cell carcinoma.

1.2.1.1. Non-Small Cell Carcinoma

Adenocarcinoma (AD) accounts for approximately 40% of lung cancer cases and is the most common form with ~500,000 cases presented annually. The tumour type is described as highly heterogeneous and is highly associated with smoking, now the leading form of cancer caused by this habit. AD is usually located around the periphery of the lungs relative to other cancers of the lung, and it is thought this development is related to the deep inhalation of tobacco smoke into the lungs. However, it is also the subtype which is presented most commonly in people who do not smoke. This cancer starts in early versions of glandular cells which would normally secrete substances such as mucin. This mucin is used as a marker in poorly-differentiated AD to verify the presence of these cancerous cells (Travis et al. 2013). AD does not respond well to chemotherapy or radiotherapy, and is therefore usually treated by resection or lobectomy.

Squamous cell carcinoma (SCC) is the second-most common type of cancer and accounts for about 30% of all cases of non-small cell lung cancer. Most cases of SCC begin in the central area of the lungs and are referred to epidermoid carcinomas. Typically arising from the bronchus, these tumour types invade the surrounding lung parenchyma and can extend into the chest wall (Miyazaki et al. 2011). Tumours can be very large and tend to undergo central necrosis. Cavitation is a frequent finding in primary lung SCC but can also be encountered in metastatic SCC and this process is secondary to tumour necrosis. In other instances, SCC can have a central scar with peripheral growth of tumour. SCC is characterised by intercellular bridging and the keratinisation of individual cells, depending on the degree of differentiation. Four subtypes are recognised: papillary, small-cell, clear-cell, and basaloid (Perez-Moreno et al. 2012). This slow growing tumour type can take many years become invasive, and the most common sites of metastasis are regional lymph nodes, adrenal glands, brain, bone and liver. SCC has a better prognosis than AD if they are at the same stage of development (Chansky et al. 2009).

Large cell Carcinoma (LCC) is the least common form of lung tumour and account for approximately 10-15% of all lung cancers. They are malignant epithelial tumours with large nuclei, abundant cytoplasm and well defined cell borders. Most LCCs grow without evidence of glandular or papillary differentiation, and have no distinguishing features which can be discerned by light microscopic examination. Risk of LCC increases in those with a history of tobacco smoking, with a relative risk of 2.3 compared to never-smokers which increases to 3.6 if smoking duration is > 40 years (Kenfield et al. 2008). Tumours of LCC generally start in the outer regions of the lung and grow rapidly. This causes fluid to develop in the lining of the lung which can result in what is known as a pleural effusion. Diagnosis of LCC is performed using a ‘diagnosis of exclusion’ method, where the tumour demonstrates none of the morphological features of AD, SCC or SCLC (Travis et al. 2013). Another clinically relevant subtype of LCC is large cell neuroendocrine carcinoma (LCNEC) which displays immunohistochemical and morphological features common to both neuroendocrine tumours and LCCs. Diagnosis of this subtype is critical as it has a much poorer prognosis compared to other forms of NSCLC (Fernandez and Battafarano 2006).

1.2.1.2. Small Cell Lung Cancer

Small cell carcinoma is a type of aggressive and malignant cancer which most commonly arises in the lung. It is also known as oat cell carcinoma due to its appearance under the microscope of cells with a flat shape and very little cytoplasm. It is a cancer which accounts for approximately 13-15% of all lung cancers and in this form is described as a highly aggressive neuroendocrine malignancy (Nishino et al. 2011). The majority of patients present with the metastatic form of the disease upon diagnosis, therefore surgical resection of rarely an option. The lung is the most common site of origin of small cell carcinoma in the body and more than 95% of cases occur in that location. Small cell carcinomas which develop outside of the lung in areas such as prostate, cervix and oesophagus are extremely rare (Walenkamp, Sonke and Sleijfer 2009).

There are two categories of SCLC: small cell lung carcinoma (SCLC) and combined small cell lung carcinoma (c-SCLC)

Small Cell Lung Carcinomas are a highly aggressive form of lung cancer and are considered a subtype of neuroendocrine tumours. SCLC is separate from NSCLC as it has unique presentation, imaging appearance, treatment and prognosis. Tumours with this classification are characterised as an epithelial malignancy made up of small tumour cells with ill-defined borders, scant cytoplasm and finely granular nuclear chromatin without obvious nucleoli. SCLC cells are associated with the secretion of a variety of substances such as serotonin, calcitonin and bombesin peptides (Alexander et al. 1988). For this reason SCLC is regarded as a neuroendocrine tumour because of its ability to synthesize various hormone peptides that can act as paracrine and autocrine growth factors, similar to PNECs of the lung.

The distinctive appearance of SCLC makes diagnosis of this subtype reliable, and a panel of immunohistochemical markers such as pancytokeratin as well as chromogranin and synaptophysin (neuroendocrine markers) can be used to confirm suspected SCLC (Miskovic et al. 2015). SCLC is rarely operable at the time of diagnosis and treatment is usually a combination of chemotherapy and radiotherapy. It is an extremely aggressive subtype with survival measured in months with a five year

survival rate of < 5%. The highly metastatic nature of SCLC results in approximately two-thirds of patients presenting with distant metastases usually in the liver, bone, brain, lung and adrenal gland. Studies have shown that metastasis to the liver, brain and pleural and/or pericardial fluid are associated with particularly poor prognosis (Nakazawa et al. 2012). The amount of SCLCs presenting without evidence of metastasis is extremely rare occurring in < 5% of cases.

c-SCLC or combined small cell lung carcinoma is a classification reserved for small cell carcinoma tumours which are found to additionally possess an additional component consisting of any non-small cell histological type. A NSCLC component can be one or more differentiated forms of lung cancer such as AD, SCC and LCC identified by histopathological examination. It is a relatively rare form of cancer and is thought to make up approximately 30% of SCLC cases (Wagner et al. 2009). As SCLC is the most aggressive form of lung cancer, it is recommended that all malignant lung tumours which are found to contain any proportion of SCLC cells should be classified as c-SCLC. Treatment for c-SCLC has traditionally been similar to SCLC and is usually a combination of chemotherapy and radiotherapy. However, studies have shown that surgery in combination with the above treatments can improve the outcomes of patients with stage 1 c-SCLC (Ihde 1984, Hage et al. 1998).

1.2.2. Causes of Lung Cancer

The cellular and molecular changes which give rise to lung cancer have become better understood over the past 25 years. Cancer cells contain many genetic and epigenetic abnormalities, and there is now a greater understanding of the role played by these factors in relation to the development of this disease. Numerous studies have identified changes in the expression of tumour suppressor genes, abnormal alterations in growth-stimulatory signalling pathways and overexpression of oncogenes. In addition, the external factors which contribute to acquired gene alterations and DNA damage which lead to the development of lung cancer have been extensively studied.

1.2.2.1. Proto-Oncogenes

Continued improvements in genetic technology have led to a greater understanding of the molecular changes involved in the development of lung cancer. It is thought that numerous mutations in diverse gene types lead to the development of lung cancer and drive the malignant phenotype. Though these numerous mutations may lead to the development of the disease, it has been found that proto-oncogenes play a crucial role in the metastatic phenotype. Proto-oncogenes are mutated genes which are overexpressed at a high level. Investigations into these genes have led to the concept of 'oncogene addiction', where tumour cells become dependent on overexpressed proteins as a result of genetic mutations (Weinstein 2002). Targeting such a proto-oncogene can profoundly inhibit the proliferation of cancer cells and increase overall patient survival (Weinstein and Joe 2008).

Epidermal growth factor receptor (EGFR) is a proto-oncogene associated with the malignant phenotype of lung cancer. It is overexpressed in 65% of Adenocarcinoma, 84% in Large cell carcinoma and has minimal expression in Small cell lung carcinoma (Bunn and Franklin 2002). The mutated version of EGFR and other proto-oncogenes are thought to drive the malignant phenotype of lung cancer. EGFR inhibitors such as gefitinib and erolitinib have shown a 67% response rate, increasing overall survival compared to wild-type EGFR, and outperforming conventional chemotherapeutic treatment (Jackman et al. 2009).

Ras genes (HRAS, KRAS and NRAS) code for proteins involved in guanosine triphosphate (GTP) binding, and approximately 20-25% of cases of NSCLC patients

have been found to present with KRAS mutations (Campos-Parra et al. 2015). Mutations of KRAS lead to constitutively activated proteins locked in the GTP-bound "on" state which in turn results in the continued activation of the MAPK and PI3K/AKT7/mTOR pathways, ultimately leading to increased proliferation and resistance to apoptosis (Shaw et al. 2011). The majority of mutations in the KRAS proto-oncogene in NSCLC occur at codon-12 with single point amino acid mutation being the most frequent. KRAS mutant tumours are independent of the EGFR signalling pathway and are therefore unaffected by EGFR inhibitors such as gefitinib, and they are also resistant to chemotherapy. Mutations in EGFR and KRAS are usually mutually exclusive, and patients who are positive for KRAS mutations have a low response rate to EGFR inhibitors at < 5% (Suda, Tomizawa and Mitsudomi 2010).

The Myc family of proteins have been found to be potent proto-oncogenes, mutated and persistently overexpressed in various cancers. c-Myc plays an essential role as a transcription regulator and plays a critical role in the regulation of processes such as cell-cycle control, apoptosis, protein synthesis and cell adhesion (Dang 2012). Abnormal overexpression of c-Myc is often found in SCLCs, and less frequently in NSCLCs (Bergh 1990). Mitogenic signalling pathways such as Wnt and EGF (via the MAPK/ERK pathway) activate c-Myc with a resulting upregulation of many genes such as those involved in cell proliferation. Constitutive overexpression of c-Myc through mutation drives the proliferation of cancer cells and contributes to the progression of the disease (Zhang et al. 2012).

1.2.2.2. Tumour Suppressor Genes

In contrast to proto-oncogenes, loss of function of tumour suppressor genes (TSGs) plays an equally important role in the development of lung cancer. TSGs are genes which protect cell from starting down the road towards a cancerous phenotype. These genes usually code for proteins which have a dampening effect on the regulation of the cell cycle, or may even promote apoptosis. It has been proposed that TSGs fall into two broad categories: gatekeeper genes and caretaker genes (Kinzler and Vogelstein 1997). Gatekeeper genes regulate the cell cycle and control how cells progress through division. Caretaker genes maintain the integrity of the genome and are involved in the repair of DNA damage. Loss of function in these TSGs requires both alleles of a gene to be damaged or mutated and is referred to as the 'two-hit model' (Knudson 1971). This model proposes that for the effect of a mutation in a TSG to be manifested and lead to a cancerous phenotype, both copies of the gene must undergo mutation as the remaining copy of the gene can still produce a functional version of the protein, though there are exceptions to his rule. In contrast, proto-oncogenes need only undergo a mutation in a single allele as they are gain of function mutations.

One of the most common TSGs which have undergone loss of function mutations in a wide variety of cancers is p53. It is mutated in 30-50% of human cancers (Hollstein et al. 1991, Freed-Pastor and Prives 2012) with high prevalence in lung cancer. It was originally believed to be a proto-oncogene; however these early data were based on findings from experimental overexpression using cDNA cloned from a mutated version of p53. Subsequently, wild-type p53 was found to suppress growth, and it was reclassified as a TSG (Michalovitz, Halevy and Oren 1990). Mutations in the TP53 gene are reported in 70% of SCLCs and 50% of NSCLCs, resulting in the synthesis of full-length p53 protein which is unable to bind directly to DNA and activate anti-proliferative and pro-apoptotic target genes (Toyooka, Tsuda and Gazdar 2003). P53 mutations are significantly higher in smokers than in non-smokers and the most prominent carcinogens in tobacco smoke such as polycyclic aromatic hydrocarbons has been found to form DNA adducts in the coding region of the TP53 gene (Abedi-Ardekani et al. 2011).

P16 and RB (cyclin-dependent kinase inhibitor 2A and retinoblastoma protein respectively) are well accepted TSGs which are mutated in lung cancers, with both alleles frequently inactivated in the diseased state. Alterations in both these genes have been reported within subtypes of lung cancer, with mutated versions of RB expressed in 85% of SCLC and 25% of NSCLC, and mutated P16 expressed in < 10% of SCLC and 60% of NSCLCs (Mori et al. 1990, Okamoto et al. 1995). Both proteins are involved in cell cycle regulation, decelerating the progression of cells transitioning from G1 phase of growth to synthesis (S) phase. Studies indicate that p16 and RB are inactivated in lung cancer cells, and the signalling pathways are estimated to be disturbed in 50-80% of lung cancers (Kelley et al. 1995). Loss of function of these genes has been shown to play a significant role in lung carcinogenesis through the unrestrained progression of the cell cycle.

1.2.2.3. Tobacco Smoke

Smoking of tobacco is by far the leading cause of lung cancer, responsible for ~90% of all lung cancer cases. It is estimated that approximately 1.25 billion people smoke cigarettes on a daily basis and 40% of patients with newly diagnosed lung cancer are still smokers. Tobacco smoke is an aerosol containing 4000 substances, of which 50 are known carcinogens including benzene, formaldehyde, lead, and polycyclic aromatic hydrocarbons (Abedi-Ardekani et al. 2011). It is thought that smoking will cause roughly 10 million deaths per annum by the year 2030, and one quarter of these deaths will be as a result of lung cancer (Lemjabbar-Alaoui et al. 2006). Smoking has been linked with aberrant methylation of both p53 and p16 causing silencing of the tumour suppressor genes. It is thought that analysis of promoter methylation on TSGs may provide a useful biomarker for the identification of groups with elevated risk of lung cancer such as smokers (Jarmalaite et al. 2003). Only 20% of smoker develop lung cancer, therefore it is logical to conclude that some people are more susceptible to the effects of tobacco related carcinogens than others.

1.2.2.4. Environmental Carcinogens

Radon is a naturally occurring radioactive gas which results from the breakdown of uranium in rocks and soil. Alpha-particles emitted by radon are highly effective at damaging tissues. Charged alpha-particles are attracted to dust due to their charged state, and can become lodged in the epithelial tissues of the lung if inhaled. This can lead to gene mutations, chromosome aberrations and eventually the development of lung cancer. Radon is recognised as the second most common cause of cancer in the United States causing about 21,000 deaths each year (Choi and Mazzone 2014).

Asbestos is a naturally occurring mineral which has been widely used as a building and fireproofing material due to its intrinsic properties, however it is now know to be a potent carcinogen. Asbestos fibres can become lodged deep in the lung and cause irritation and scarring. This eventually leads to the development of lung cancer. Asbestos fibres can penetrate into the outer lining of the lung and chest wall (known as the pleura), and lead to development of cancer there known as mesothelioma.

1.3. Cancer Metastasis

Cancer metastasis is a complex multistep process whereby malignant cells spread from the primary tumour site to distant organs of the body setting up secondary tumour sites. Approximately 90% of all cancer deaths are the result of complications arising from metastasis, and it is rare for death to occur from the effects of a primary tumour alone (Wirtz, Konstantopoulos and Searson 2011). Metastasising cancer cells travel via the bloodstream, the lymphatic system or invade adjoining tissue by direct extension. The processes of invasion and metastasis are poorly understood, and there are currently no therapies available which effectively target invading tumour cells.

The ability of cells to undergo metastasis is preceded by the acquisition of an invasive phenotype by tumour cells. Malignant cells must be able to breach the extracellular matrix (ECM) and escape the tumour environment. This is a prerequisite phenotypic characteristic required for the initiation of the metastatic cascade which then allows tumour cells to move from the primary site to a distant location. In lung cancer metastasis, this cascade involves a number of fundamental mechanisms such as cell motility, angiogenesis, and degradation of the ECM, and cell-cell adhesion (See Figure 1.1.)

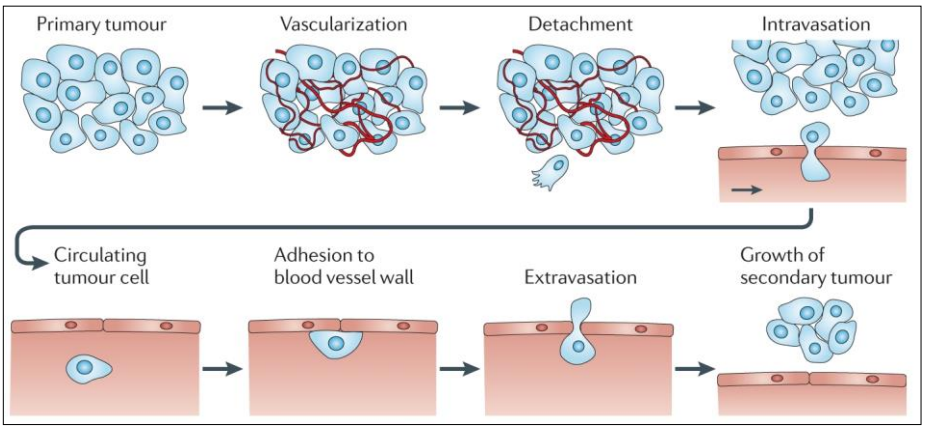


Figure 1.1: The steps involved in metastatic cascade resulting in the development of secondary tumours. (Wirtz, Konstantopoulos and Searson 2011)

The metastatic cascade involves the following critical steps which allow cells from a primary tumour site to set up a secondary tumour at a distant site in the body. Firstly, the cells must invade the surrounding tissue to escape the primary tumour site. The cell then must undergo trans-endothelial migration into lymph or blood vessels which is referred to as intravasation. Survival of the cells in these vessels is difficult due to the shear stresses involved, and if successful they will be translocated to distant sites. Next, the cells must leave the vessel through extravasation and survive the microenvironment of the new distant tissue type. Finally, the cells must undergo proliferation in order to set up a secondary tumour or metastasis (van Zijl, Krupitza and Mikulits 2011).

Fortunately, metastasis is thought to be an inefficient process, and studies suggest that less than 0.02% of tumour cells in circulation have the ability to form metastases (Chambers, Groom and MacDonald 2002). Tumour cells encounter significant obstacles once they have left the primary site, and subsequent colonisation is very limited due to incompatibility of distant sites with tumour cell proliferation. However, the ability of cancer cells to acquire a metastatic phenotype is one of the most critical and dangerous determining factors of tumour progression.

Healthy cells are programmed to be anchorage dependant, and to undergo apoptosis if detached or damaged. In cancer cells, the accumulation of genetic changes causes this regulatory process to be disrupted, leading to tumour development. A tumour which has not yet developed an invasive phenotype is termed ‘carcinoma *in situ*’. This form of carcinoma has a good prognosis and is usually treatable. However, if tumour cells start down the path of the metastatic cascade, the resulting metastases dramatically decrease the chances of survival.

A major problem which hinders the understanding of metastasis is how individual tumour cells acquire all the individual properties which are required for the metastatic phenotype. One possible solution to this problem is the existence of a multi-faceted biological program which allows carcinoma cells to attain a number of the characteristics to set them down the metastatic cascade path. This biological program is a process known as epithelial to mesenchymal transition (EMT).

1.3.1. Epithelial to Mesenchymal Transition (EMT)

Epithelial cells are closely packed cells connected to each other via adherens junctions, gap junctions and tight junctions. They are polarised with apico-basal polarity and usually interact with the basement membrane via their basal surface. Mesenchymal cells lack polarisation, have spindle-like morphology and connect to each other only through focal points. EMT describes a transition process by which epithelial cells lose their polarity and undergo biochemical changes which allows them to develop a mesenchymal phenotype. These phenotypic characteristics include enhanced migratory ability, invasiveness, resistance to apoptosis and increased production of ECM components (Kalluri and Weinberg 2009). This is a natural process which is known to be involved in embryonic development and wound healing processes (Kong et al. 2011).

The transition of epithelial cells in lung cancer to mesenchymal cells is thought to be a critical step in the development of metastasis. The acquisition of invasiveness is firstly demonstrated by the breaching of the basement membrane that confines the epithelial compartment. This step is thought to be the critical step heralding the initiation of a multi-step process leading to metastatic spread of malignant cells with life threatening consequences (Hanahan and Weinberg 2011).

Carcinoma cells in primary tumours undergoing EMT lose cell-cell adhesion, and a key step in this process is the reduction or loss of expression of E-cadherin, crucial to the progression of epithelial tumours to develop into metastatic cancers. The reduction or loss of expression of E-cadherin is usually seen to coincide with the gain of expression of N-cadherin which is a marker for mesenchymal cells (Nakajima et al. 2004). This is known as the ‘cadherin switch’, and is a hallmark of EMT. Several other interconnected pathways and signalling molecules have been identified which are involved in the process of EMT such as receptor tyrosine kinases, GTPases, β-catenin and Integrins. In addition, transcription factors such as Slug, Snail and Twist have all been implicated in inducing EMT in epithelial cells and have each been shown to be highly expressed in metastatic cells (Talmadge and Fidler 2010).

The central regulator of EMT has been shown to be transforming growth factor- β (TGF- β) which is an important suppressor of endothelial cell proliferation. It can serve as a positive regulator of tumour progression and metastasis, and in vitro

studies have shown it can induce EMT in certain types of cancer cells (Oft, Heider and Beug 1998, Song 2007). A number of pathways have been identified as mediators of TGF- β induced EMT. One pathway involves SMAD proteins, in which TGF-β induces SMAD-mediated signalling facilitating cell motility. Autocrine production of TGF-β can further amplify the EMT program via SMAD signalling. Signalling pathways that mediate the action of β-catenin have been found to also cooperate with SMAD inducing an EMT. Studies have collectively demonstrated that the TGF-β/SMAD/β-catenin axis is an important inducer of EMT in cancer cells (Yang, Lin and Liu 2006). The identification of pathways which lead to EMT provides a new perspective on the plasticity of cells and possible therapeutic routes.

1.3.2. Invasion/Metastasis

Once a cell undergoes EMT, it gains the ability to invade surrounding tissue. This involves the translocation of the cell through well-defined tissue boundaries and further through different types of ECM barriers. In the case of lung cancer, the cell must pass through the basement membrane to escape the primary tumour site within the lung. A number of proteins and processes are involved in this including: Matrix metalloproteinases, cellular adhesion molecules and Actin cytoskeletal regulation.

Matrix metalloproteinases (MMPs) are calcium dependant proteins which belong to the larger family of proteases. These enzymes are secreted in an inactive form as zymogens, are activated by the presence of other MMPs or serine proteases such as Kallikrein, Trypsin and Plasmin which remove the pro-peptide domain. MMPs are capable of degrading ECM proteins such as proteoglycans, fibronectin, collagen and laminin, and this ability significantly contributes to the invasive ability of malignant cancer cells. Studies have shown that MMP-10 is highly expressed in human NSCLC compared to normal lung tissues, rendering it a potential target for the development of novel therapeutics for lung cancer treatment (Gill et al. 2004).

In addition to providing cancer cells with the means to degrade the ECM and escape the primary tumour site, subgroups of MMPs have also been shown to cause EMT in the first instance. Overexpression of ADAM10 has been shown to lead to the shedding of E-cadherin in epithelial cells, leading to the disruption of cell-cell adhesions and increasing migratory ability (Maretzky et al. 2005). In addition, the shedding of E-cadherin induced by MMPs appear to contribute to EMT, as recent studies have implicated MMP-28 in the activation of TGF-β in lung carcinoma cells (Illman et al. 2006, Song 2007)

Cellular adhesion molecules (CAMs) are proteins expressed on the surface of cells which are involved in cell-cell binding or adhesion. They are transmembrane proteins which are composed of three domains: and intracellular domain which interacts with the cytoskeleton, and transmembrane domain, and the extracellular domain which interacts with other CAMs. Circulating tumour cells need to interact with the endothelium of blood and lymph vessels in order to undergo extravasation and ultimately metastasise. Studies of tumour cell-endothelial adhesion have been thus far based on parallels with leukocyte-endothelial interactions during inflammation, and it is thought that the mechanisms involved are potentially very similar. Sub-families of calcium dependant CAMs have been identified as participants in cancer metastasis such as Selectins, Integrins, and Cadherins (Paschos, Canovas and Bird 2009).

Selectins are vascular CAMs involved in adhesive interactions of leukocytes/platelets and the endothelium within the blood circulation. P-selectin glycoprotein ligand-1 (PSGL-1) is a Selectin protein expressed on the surface of white blood cells. The expression and activation of PSGL-1 on the surface of white blood cells being recruited to inflamed tissue allows the cell to interact with the endothelial cell surface of blood CAMs. This allows the cell to roll slowly along the wall until it reaches the site of inflammation, where it transmigrates into the tissue (Kang et al. 2012). The majority of Selectin ligands consist of distinct glycoprotein structures carrying the tetrasaccharide structure Sialyl Lewis^x (SLe^x) expressed on the cell surface (Kansas 1996) which is a modification required for its activation. During malignant transformation of epithelial cells, one of the major changes in Selectin proteins is the upregulation in expression of structures such as SLe^x. The enhanced expression of SLe^x is associated with progression and poor prognosis in various cancers including lung cancer (Kannagi et al. 2004).

Integrins are one of the major families of CAMs. It is well established that they are ubiquitously expressed ECM receptors, and serve as the main link between cells and the ECM through adhesion processes. Integrins mediate cell-ECM interactions through collagen, Vitronectin, Fibronectin and Fibrinogen (Humphries, Byron and Humphries 2006). These transmembrane proteins bind extracellularly to the ECM and intracellulary to the Actin cytoskeleton, integrating both environments. In this role, Integrins transduce signals from the outside of the cell to the inside and vice versa, and regulate cell adhesion, cell migration, ECM remodelling, and differentiation (Hynes 2002).

Integrins can also modulate the signalling cascade elicited by growth factors such as TGF-β (Margadant and Sonnenberg 2010).

Cadherins are calcium dependant cell surface glycoproteins which are capable of mediating cell-cell adhesion. The cadherin superfamily includes N/E-Cadherins, Proto-Cadherins, Desmogleins, and Desmocollins (Garrod and Chidgey 2008). These transmembrane proteins are concentrated at adherens junctions, particularly desmosomes, and are involved in cell-cell adhesion. Desmocollins are similar to E/N-Cadherins in overall structure, whereas Desmogleins have extra cytoplasmic sequenced added on to the basic Cadherin structure. This allows them bind to mediator proteins such as Plakoglobin, which in turn links to the Keratin intermediate filament network (Andl and Stanley 2001). Behaving as CAMs, these proteins have additional roles as both receptors and ligands for other molecules. As previously mentioned, loss of E-Cadherin expression is concomitant with a loss of cell-cell adhesion and is one of the primary steps involved in EMT (Nakajima et al. 2004).

Regulation of the Actin cytoskeleton is a key process involved in the normal function of numerous different types of cells and is highly associated with cell migration. Many naturally occurring cell types are migratory and are involved in various processes such as immune surveillance, tissue repair, and embryonic morphogenesis. Atypical regulation of the actin cytoskeleton has been found to contribute to the progression of many diseases including cancer with regard to migration and metastasis (Yamaguchi, Wyckoff and Condeelis 2005).

The initial step in cell migration is the protrusion of the cell membrane, and these protrusions are referred to as lamellipodia, filopodia, or invadopodia depending on their shape and function. Lamellipodia are flat structures which form at the leading edge of a motile cell and their extension is generated by the polymerisation of Actin filaments pushing against the cell membrane. The filaments of Actin can exert force as they polymerise at the leading edge of a cell, and flow back towards the cell body. By Actin filaments coupling with focal adhesions, force is generated against the cell membrane causing traction and movement (Zimmermann et al. 2012). This mechanism which gives rise to cell motility is referred to as Actin Retrograde Flow (ARF). The cell velocity is determined by polymerization forces at the lamellipodium leading edge, contraction of the Actin gel by myosin motors, cell adhesion to the ECM substrate via CAMs such as the Integrins, and the process of ARF (Ji, Lim and Danuser 2008).

In order for cells to migrate though the ECM barrier, they require the ability to attach to the ECM and move through it, as well as degrade it. Invasive cancer cells form Actin-rich protrusions which extend vertically from the ventral cell membrane, and are termed invadopodia. These structures are enriched with Actin regulatory proteins, CAMs, membrane remodelling proteins, and matrix degrading proteases (Baldassarre et al. 2003). Carcinoma cells migrating on ECM fibres extend invadopodial protrusions which attach to the ECM at the leading edge of the cells, and utilise these formations to migrate and invade through the stroma towards blood vessels in the process of metastasis. The formation of invadopodial structures has only been observed in highly invasive cancer cells, and is therefore implicated in tumour metastasis (Yamaguchi et al. 2005, Weaver 2006).

Regulation of the formation of protrusive structures is carried out by a number of Actin related proteins and complexes. One of these is known as the ARP 2/3 complex

and this plays a major role in the regulation of the Actin cytoskeleton. Many Actin related molecules sever pre-existing filaments and use the exposed ends as nucleation cores for Actin polymerisation, extending existing filaments. In contrast, ARP 2/3 stimulates the polymerisation of new Actin filaments by creating new nucleation cores, encouraging Actin branching (Aguda, Burtnick and Robinson 2005). The nucleation activity of ARP 2/3 is activated by upstream regulators such as members of the Wiskott-Aldrich syndrome family of proteins (WASP, N-WASP, WAVE, and WASH). These WASP proteins integrate multiple upstream signals to induce Actin polymerization through the Arp2/3 complex (Veltman and Insall 2010).

The expression of WASP and ARP 2/3 proteins has been associated with malignant phenotypes of cancer cells, indicating their importance in metastasis. Immunohistochemical investigation of ARP 2/3 expression in colorectal tumours found that while no positive staining was found in normal colorectal tissue or surrounding stroma, significantly high staining for ARP 2/3 was found in invasive carcinomas. Positive staining for this protein was typically seen at the invasive front of the colorectal carcinoma, and higher intensity staining correlated with increased invasive depth (Otsubo et al. 2004). Immunohistochemical staining of both WASP and ARP 2/3 complexes in lung adenocarcinoma found that co-staining of both proteins is an independent risk factor for tumour recurrence. They also concluded that WASP binds to ARP 2/3 in lung adenocarcinoma and drives the migratory phenotype through regulation of the actin cytoskeleton and the formation of protrusive and invasive structures (Semba et al. 2006).

Cortactin (CTTN) is a monomeric scaffolding protein involved in the regulation of the Actin cytoskeleton. The protein is located diffusely in the cytoplasm in its inactive state, and here Src-kinase can bind to its Proline-rich region and phosphorylate the protein. Cortactin assists in the creation of new nucleation cores in ARP 2/3 mediated F-Actin branching. It also binds and cross-links F-Actin filaments and stabilises existing interactions between ARP 2/3 complexes and F-Actin (Uruno et al. 2001). It is prominently expressed in cell protrusions such as lamellipodia and invadopodia which are essential structures required by cancer cells to achieve metastasis through migratory processes, and they have also been shown to be essential for the

degradation of the ECM through the localisation of proteolytic enzymes (Artym et al. 2006).

The development of invadopodia and podosomes is dependent on CTTN activity and the protein has been shown to be a key regulator in the formation of these structures in cancer cells. Examination of breast cancer cells found that inhibition of invadopodia formation was caused by the knockdown of CTTN by RNAi. Consequently, impeded development of invadopodial structures by CTTT depletion resulted in a block of ECM degradation due to failure of invadopodia formation (Bowden et al. 1999). In NSCLC, CTTN was found to be associated with poor prognosis and lymph node metastasis by immunohistochemistry. It was concluded that CTTN is involved in the progression of NSCLC (Noh et al. 2013).

Key proteins involved in the regulation of the Actin cytoskeleton are associated with the malignant and invasive phenotypes in a variety of cancers. The mode of invasion that is involved with invadopodial structures requires a dynamic Actin cytoskeleton, and there is growing evidence that Actin-binding proteins have multiple roles in tumourigenic and metastatic processes. Further investigation into how these proteins work together to promote the invasive phenotype may lead to a greater understanding of the migratory and metastatic phenotypes in cancers, and may present novel therapeutic routes of treatment.

1.4. Heterogeneity in Lung Tumours

The poor outcome of lung cancer is due to a number of factors specific to the disease. These include the early metastatic spread, difficult detection of the onset of disease, and the development of drug resistance following treatment with chemotherapeutic agents. One major factor which makes lung cancer a particularly difficult disease to treat is the characteristic heterogeneity of lung tumours. Functional and phenotypic heterogeneity of lung cancer cells has been long recognized, with experiments as early as the 1930s demonstrating that some but not other cells from mouse tumours could give rise to new tumours when transplanted (Marte 2013).

Tumour heterogeneity refers to the observation that multiple cell types can be found within individual tumours, and that each subpopulation can show distinct morphological and phenotypic characteristics. These can include cellular morphology, gene expression, protein expression, metabolism, and metastatic potential (Marusyk and Polyak 2010). It is thought that this complex population of cells within a single tumour type can be explained by the cancer stem cell (CSC) theory. This theory proposes that within tumours, there are subsets of stem cell-like cancer cells with the ability to self-renew and generate diverse tumour cells (Reya et al. 2001). These cells may be the source of intra-tumour heterogeneity which is a characteristic of lung cancer. The existence of CSC is supported by the fact that only a small population of lung adenocarcinoma cells (< 1.5%) possess clonal forming and tumourigenic ability (Kim et al. 2005, Carney et al. 1982).

In addition to the CSC theory, genomic and chromosomal instability is thought to contribute to the heterogeneity found in tumours. In normally functioning lung epithelial cells, the genome is replicated with high fidelity, with mutations being routinely eliminated by thorough molecular checking mechanisms such as double-strand break repair, and mismatch repair. In cancerous cells, these checking mechanisms malfunction, and lead to a proliferation of mutations in the tumour cells. This eventually results in further carcinogenesis and tumour heterogeneity (Pailler et al. 2015).

Diversity in subpopulations with the same tumour can be evident at a genetic level. Using next generation whole genome sequencing, it has been shown that heterogeneous somatic mutations and divergent allelic profiles are present in different

regions of micro-dissected samples from the same renal tumour. Genes associated with both good and bad prognoses were detected in different regions of the same tumour, reflecting the drastic subclonal diversity and heterogeneous nature of the tumour (Gerlinger et al. 2012). In lung cancer, analyses of samples from different regions of the same tumour are still lacking. One study conducted deep digital sequencing of tumour and adjacent normal tissue samples from 17 patients with NSCLC. Deep digital sequencing refers to a method in which multiple sequencing reads are carried out (100-10,000 times) to reliably identify potential mutations. Deep digital sequencing of somatic mutations in lung cancers from both smokers and never-smokers revealed multi-clonal signatures, indicated by distinct variant allele frequency features. This suggested that at a genetic level NSCLC had a heterogeneous nature, consisting of subclonal populations (Govindan et al. 2012).

Micro-environmental adaption of tumour cells is another factor which is thought to contribute to the heterogeneity of lung tumours. The immediate external tumour micro-environment area is made up of numerous cell types including stromal cells, fibroblasts, and immune cells. This results in the varying degrees of selective pressures such as oxygen availability, pH variations, and growth factors to name but a few (Albini et al. 2015). The surrounding microenvironment interacts with cancer cells to influence the development and progression of the neoplastic disease. For instance, secretion of Hepatocyte Growth Factor (HGF) by stromal fibroblasts under the stimulation of tumour-derived factors such as Interleukins strongly encourages tumour expansion and invasiveness by activating the MET oncogene pathway in NSCLC (Nakamura et al. 1997). Depending on the proximity of tumour cells to the stroma, these effects vary and induce the generation of subpopulations within a tumour with differing phenotypic abilities.

A major route in the treatment of NSCLC is the targeting of tumour-driving pathways such as epidermal growth factor receptor (EGFR) by tyrosine kinase inhibitors (TKIs). This treatment is based on the oncogene addiction theory (Weinstein and Joe 2008), and has been shown to be successful in advanced NSCLC patients, conferring prolonged survival and limited adverse effects. However, it has been found that tumours which initially respond to TKIs eventually become refractory to therapy and regrow (Mok et al. 2009, Gainor and Shaw 2013). The role of tumour heterogeneity

has been shown to be a critical factor in the development of drug resistance in lung cancer. It is thought that tumour development is a process of clonal evolution as described in the CSC theory, with every tumour possessing subclones with diverse phenotypic abilities. This allows some cells to escape therapeutic intervention and then induce regrowth of the tumour.

Tumour heterogeneity in lung cancer presents major challenges for the determination of appropriate treatments and indeed the diagnosis of the disease. Subpopulations within a tumour can survive through therapeutic intervention and cause re-growth of the tumour, severely compromising attempts at treatment. In addition, tumour heterogeneity may complicate the initial diagnosis and subsequent treatment regimen, as single biopsies of primary tumours, as is usual clinical practice, leads to a misleading characterisation of a heterogeneous lung tumour (Gerlinger et al. 2012). It is now recommended that multiple biopsies be taken of a tumour to mitigate the effects of intratumour heterogeneity. This may result in improved response to chemotherapy with more targeted therapies based on more detailed and accurate diagnoses (Hiley et al. 2014).

Despite the evidence for intratumour heterogeneity, the subject remains relatively poorly explored. It is thought that a more systemic approach is required for to characterise the extent of heterogeneity within tumours. An important factor to consider is to distinguish between cellular heterogeneity and clonal heterogeneity. Cellular heterogeneity refers to the differences between individual tumour cells, whereas clonal heterogeneity refers to differences which have been amplified by clonal expansion (Marusyk and Polyak 2010). Focusing on clonal heterogeneity eliminates some of the ‘noise’ from variants of clones which did not clonally expand due to poor adaptation to the tumour microenvironment.

A simplistic view of tumours as homogenous entities is not a true representation of the reality in the majority of lung cancers and cancers in general. Treating tumours as complex systems with subpopulations of interacting clones may provide a useful route to understanding the interactions in a tumour which leads expansion, EMT and subsequent metastasis. Unravelling the full extent of clonal interactions is unlikely; however this approach may uncover critical links which may aid in the elimination of tumours.

1.4.1. The DLKP Cell Line and its Subpopulations

The DLKP cell line presents a unique opportunity to study tumour heterogeneity *in vitro*. This cell line was derived from the lymph node metastasis of a primary lung tumour. Previous work determined that the DLKP cell line contains three morphologically distinct subpopulations (McBride et al. 1998a). These clonal subpopulations were termed: DLKP-SQ, DLKP-M, and DLKP-I. These cell lines can be used as a model system for the study of tumour heterogeneity in lung cancer, with respect to clonal heterogeneity. It also provides an *in vitro* model of invasion, due to the different phenotypic properties of the clones. Investigations into the phenotypes and proteomic profiles of the clonal subpopulations, as well as possible interactions between them, may reveal targets associated with those phenotypes, and add knowledge to the understanding of tumour heterogeneity in lung cancer.

1.4.1.1. Origin

The DLKP cell line was established by Dr Geraldine Grant (Law et al. 1992) during the course of routine primary culture of human lung tumours. The patient from which DLKP originated was a 52-year-old male who had smoked approximately 40 cigarettes a day for most of his adult life. The DLKP cell line was derived from a tumour which was histologically diagnosed as a poorly differentiated squamous cell carcinoma. A subsequent review further described the tumour as ‘poorly differentiated and necrotic carcinoma without obvious keratinisation, but of a larger size and with more cytoplasm than typical oat small cell carcinoma’. Therefore the most appropriate diagnosis became ‘poorly differentiated non-small cell carcinoma’. Routine cell culture of DLKP revealed the presence of at least three morphologically distinct subpopulations which were ever-present. Isolation of these different cell types was achieved using a cloning procedure, and the three DLKP clonal subpopulations were established (McBride et al. 1998a).

1.4.1.2. Subclonal populations of the DLKP Cell Line

The presence of morphologically distinct subpopulations with the DLKP cell line was confirmed by the establishment of three clonal subpopulations by McBride et al. 1998. It was found that the morphologies of the three individual clones are well defined, and remain so after approximately 20 passages in cell culture conditions. These clonal subpopulations were named based on their morphologies, and images of each cell line in monolayer are shown in Figure 1.2. The parental DLKP cell line is also included here.

DLKP-SQ is named for its squamous-like morphology. These cells grow to a relatively large size when grown in monolayer and form distinct cell boundaries. They are found to predominate in the parental DLKP cell population, making up approximately 70% of the total cell population from microscopic observations (McBride et al. 1998).

DLKP-M cells are named for their mesenchymal-like appearance in monolayer conditions. They are of intermediate size and have irregular elongated shapes with protruding outgrowths which extend from their ventral cell membrane. They grow in a scattered manner with cells extending across each other. In the parental DLKP cell line, they make up approximately 5% of the total cell population from microscopic observations (McBride et al. 1998).

DLKP-I cells account for approximately 25% of the parental DLKP population. They grow in tightly packed groups with indistinct cell boundaries. They are named after their ‘intermediate’ properties in relation to DLKP-SQ and DLKP-M.

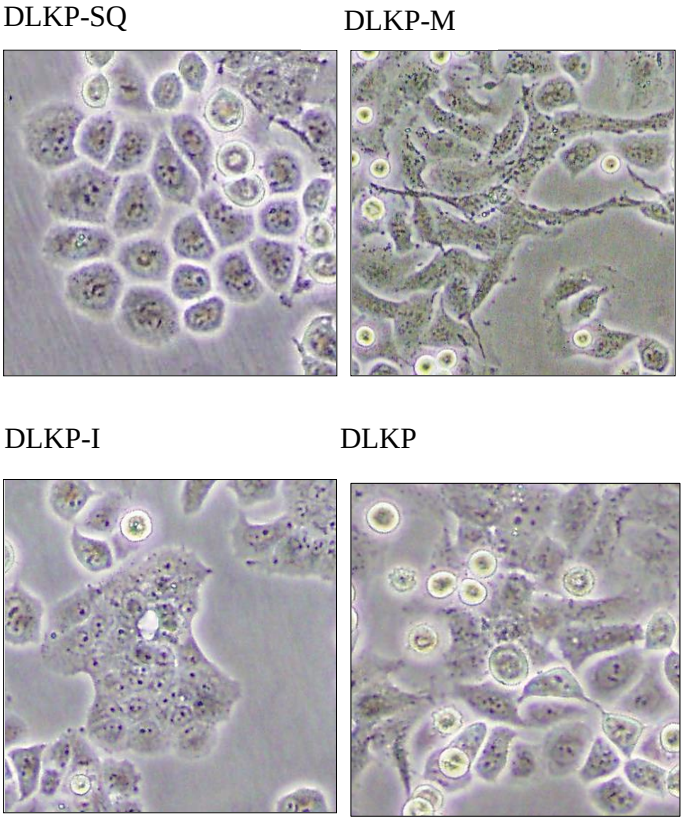


Figure 1.2: The three DLKP clonal subpopulations as they appear in monolayer: DLKP-SQ, DLKP-M, DLKP-I and DLKP.

1.4.1.3. Characteristics of the DLKP Clones

Follow-up studies carried out on the DLKP clones found that they displayed differences in the levels of in-vitro invasion, migration and anoikis resistance. The DLKP-SQ clone displays a poor invasive and migratory capacity, is anoikis resistant and displays low level expression of Integrin-αV. The DLKP-M clone displays a high invasive and migratory capacity but is anoikis sensitive (Keenan et al. 2012). The DLKP-I clone displays and intermediate level of invasive and migratory capacity, with an intermediate ability to resist anoikis. Both DLKP-I and DLKP-M display a high capacity to bind Fibronectin and Fitronectin, potentially explained by their higher expression of Integrin-αV relative to DLKP-SQ. The clones also show differences in expression of matrix-metalloproteinases (MMPs) which appear to correlate with the invasive profile of the clones, MMP2 is expressed in DLKP-I, while DLKP-M expressed MMP10, while DLKP-SQ appeared to express both as a lower level (Joyce H, PhD 2015).

1.4.1.4. The Inter-Conversion Model of DLKP

The ratio of clones present within the DLKP parental cell line seems to be tightly controlled through a model of inter-conversion proposed in the previous study performed by McBride et al. 1998. The model proposed that DLKP-SQ may inter-convert with DLKP-I, and DLKP-M may inter-convert with DLKP-I (See Figure 1.1). However, inter-conversion was not observed between DLKP-SQ and DLKP-M directly. It was therefore suggested that DLKP-I resembled a potential stem cell-like population in DLKP with its ability to interconvert and give rise to both DLKP-SQ and DLKP-M-like cells.

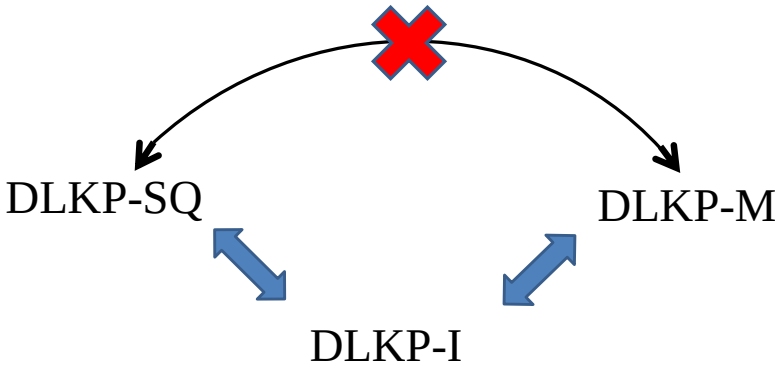


Figure 1.3: The inter conversion model of the DLKP clones as proposed by McBride et al. 1998.

Previous attempts were made to identify specific markers for the DLKP clones using immunocytochemical methods however this was unsuccessful (McBride et al. 1998a). To date, no markers have emerged that can allow for the differentiation between the individual DLKP clones. To identify proteins showing high expression in one of the clones relative to the others would be useful, as it may be involved in the phenotypic properties of that clone. Quantitative proteomic analysis could allow for the identification of proteins that are highly expressed in each of the clonal variants, and these could potentially act as markers for the clones. Also, identification of such proteins could provide the means to study the model of inter-conversion between the

clones, as well as their potential involvement in cell invasion and migration. The phenotypic differences between the DLKP clones present a prime opportunity to investigate tumour heterogeneity in lung cancer *in vitro*.

1.5. Proteomics

Studies in genomics were hoped to provide the data necessary for the comprehensive analysis of an organism, however it was found that a genome is predominantly static, and changes very little in response to intra and extracellular influences. Transcriptomics was the next avenue of investigation, as both the genome and proteome are dynamically linked to it. These studies provided mRNA quantification at certain points in time; however a severe lack in correlation was found between mRNA expression and protein expression (Chen et al. 2002). This has led to the direct study of biological systems through proteomic techniques, as proteins have been described as the ultimate effectors in a cell (Souchelnytskyi 2005).

Proteomics focuses on quantifying and identifying proteins, characterising expression levels in comparative analyses, discovering sub-cellular locations of certain proteins, and determining the roles of proteins of interest in relation to function and interactions. Expression-proteomics deals with quantitative comparisons of proteins that are changed between groups based on a sample set comparison. From this, proteins can be profiled based on the expression level changes which are present between groups in the comparison. This approach has been implemented extensively in the field of biomarker discovery, particularly in relation to cancer research. The use of quantitative proteomics can be used to compare normal vs. diseased state, and the identification of differentially expressed proteins can hypothetically be mapped back to the cause of the condition, or at least highlight proteins associated with the disease state. It is hoped that such an approach can lead to the development of diagnostic techniques which facilitate the early detection of said biomarkers, which may improve disease treatment and prognosis. This approach can also use the differentially expressed protein data to correlate with aspects of the disease state, leading to a greater understanding of the mechanisms involved.

Most proteomics research is focused on identifying differentially expressed proteins in a sample type by comparing its normal vs. diseased state. These samples can be tissues, bio-fluids, and cell lines to name but a few. Several protein profiling technologies have been developed which allow the quantitative analysis of sample types mentioned above, resulting in the identification of thousands of proteins. There have been major improvements in the speed, sensitivity and the resolution of **mass spectrometry** instrumentation, reproducible **nano-HPLC** and the associated advances in technologies for sample preparation and data analysis.

1.5.1. Sample Preparation

There are two main strategies employed when analyzing proteins using mass spectrometry. They are referred to as 'top-down and 'bottom-up'. Top-down methods analyse whole proteins, whereas bottom-up methods analyse the peptides generated from digested proteins (Armirotti and Damonte 2010). Bottom-up proteomics, or shotgun proteomics, is the most common method of sample preparation prior to mass spectrometric analysis. This method of identification characterises the amino acid sequence and post-translational modifications of a protein that has been enzymatically digested into its associated peptides. A crude protein sample is directly digested by adding an enzyme to the protein extract resulting in peptide formation. Numerous proteolytic enzymes are commercially available and differ in their cleavage sites of a protein. Trypsin has become the gold standard for protein digestion to peptides for bottom-up proteomics. Trypsin is a serine protease which cleaves proteins at the carboxyl side of arginine and lysine. These specific cleavage sites must be stated when sequence identification is searched for using protein databases.

Generally, peptide samples are separated by liquid chromatography prior to mass spectrometry to allow the MS adequate time to retrieve high resolution data on each peptide entering the instrument. The use of peptides instead of proteins in mass spectrometry gives higher sensitivity allowing for accurate identifications. By comparing the mass spectra of the proteolytic peptides, as determined by MS, with theoretical mass spectra generated from *in silico* digestion of a protein database, peptides can be accurately identified with several peptide identifications assembled into

a protein identification. By relying on the analysis of peptides, which are more easily fractionated, ionized, and fragmented, shotgun proteomics can be more universally adopted for protein analysis.

1.5.2. Reverse-phase HPLC

Separation of peptides using HPLC is carried out using two phases: the mobile phase (liquid), and the stationary phase. The mobile phase contains the analytes, and permeates through the stationary phase at high pressure, separating analytes which can be subsequently analysed by mass spectrometry. The stationary phase is typically a tightly packed column filled with irregular or spherically shaped particles which functions to separate the analytes in a complex sample.

In **Reverse phase** (RP) chromatography the stationary phase column in the analysis of peptides is composed of octadecylsilyl (C18) particles. Use of these particles takes advantage of the natural physiochemical properties of peptides which vary in their hydrophilic and hydrophobic properties. This method uses a polar mobile phase, and hydrophobic molecules (peptides) in this phase bind to the hydrophobic stationary phase (C18). By decreasing the polarity of the mobile phase over time, peptides can be eluted based on their hydrophobic interactions, which are reduced in strength by the decreasing polarity. Reverse phase LC columns are most suited for analysis of complex peptide mixtures (Tuli and Ressom 2009) with the most polar or hydrophilic peptides eluting from the stationary phase first and most non-polar hydrophobic peptides eluting towards the end of the HPLC separation. This allows good separation of peptides in a sample which can then be introduced into a mass spectrometer for analysis.

Prior to mass spectrometric analysis, the separation mechanism employed is a version of high pressure liquid chromatography (HPLC). This updated version is referred to as nano-HPLC. This method offers several advantages over regular HPLC which include: low sample volumes, improved due to small droplet formation, high efficiency packed columns that can be operated with MS friendly solvent systems, reproducible delivery of solvents to nano-columns (Qian et al. 2006).

1.5.3. Protein Mass Spectrometry (MS)

MS is an analytical technique which can be used for the fragmentation and detection of peptides, allowing their subsequent identification and quantification. A mass spectrometer is composed of three main parts, each with essential functions. There are: i) an ionization source which can convert eluting peptides into gas phase ions; ii) a mass analyser which separates ions based on their mass to charge ratios (m/z); iii) a detector which registers relative abundances of ions at discrete m/z values.

1.5.3.1. Ionisation

To analyse a peptide sample by mass spectrometry it must first be ionised and vaporised. The two commonly used techniques used for peptide ionisation are Matrix Assisted Laser Desorption Ionisation (MALDI), and ElectroSpray Ionisation (ESI). In MALDI, matrix absorbs laser energy and transfers it to the acidified peptide sample, where rapid laser heating causes desorption of matrix and peptide ions into the gas phase. ESI creates peptide ions by passing the peptide solution through a fine needle with a high electric potential across it. This electric potential produces charged droplets which shrink by evaporation resulting in increasing charge density. ESI produces multiply charged peptide ions at atmospheric pressure (Canas et al. 2006). This is carried out at the interface between the Nano-HPLC and the mass spectrometer. Peptides being separated by liquid chromatography can be continuously ionised with high efficiency and increased throughput. However this method of ionisation is not very tolerant of interfering compounds in the sample matrix such as salts, buffers, and detergents.

1.5.3.2. Mass Analysis

Mass analysers separate peptide ions based on their mass to charge ratio (m/z). The most common forms of these mass analysers are: time-of-flight (TOF), quadropole mass analysers (Q), and ion trap (IT). A mass spectrometer can have various arrangements of ion source and mass analysers, some including just one mass analyser or more than one. These are referred to as hybrid instruments. The most important characteristics of mass analysers are their resolution, which refers to their ability to

differentiate between two close signals. Other key parameters of these components in proteomic analyses are sensitivity, mass accuracy and the ability to generate information-rich ion mass spectra from peptide fragments (Aebersold and Mann 2003). Instruments with ion trap analysers have high sensitivity, fast scan rates, high duty cycle, and multiple MS scans with high resolution and mass accuracy. In quantitative label-free LC-MS/MS a hybrid instrument can be used such as a LTQ XL Orbitrap (Thermo Scientific).

Orbitrap is a high accuracy mass analyser which traps moving peptide ions in an electrostatic field and forces them to move in complex spiral patterns. Using mathematical algorithms, the m/z value of these ions can be determined. The Orbitrap can be coupled to other mass analyser such as a linear trap quadropole (LTQ) resulting in a hybrid system with high resolution and high mass accuracy along with faster scans and high sensitivity (Yates, Ruse and Nakorchevsky 2009).

In mass spectrometric analysis, the mass-to-charge ratios (*m*/*z*) of ionised peptides are determined. Each particle measured has an *m*/*z* signature detectable by the mass spectrometer. This *m/z* measurement is computed by bioinformatic tools and converted into a mass for the associated peptide. Peptides in the sample can subsequently be identified by comparing its actual mass as determined by MS to theoretical peptide masses which are predicted from databases of compounds and molecules.

Tandem mass spectrometry involves an instrument with MS/MS or MS^n capabilities which can provide additional information about an ion. In short, precursor ions are scanned in the first mass analyser and an m/z identified. The precursor ion is then fragmented and the product ions are scanned in a second mass analyser. In terms of tandem mass spectrometry of peptides, the mass spectrometer scans peptide ions which enter it. The analyser then isolates ions of a particular peptide and subjects them to collision-induced dissociation (CID) and scans the produced fragments in a tandem mass spectrum. Peptide fragment ions are generally scanned in the first analyser, fragmented and subsequent amino acid ions are scanned in the second analyser.

1.5.4. Protein Identification

Large sets of raw data generated from mass spectrometry analysis must be interrogated and converted to result files consisting of peptide and protein identifications. Data must be validated so that only information which will result in an identification are chosen and. Bioinformatics programs have the abilities to: set a threshold for only high resolution data to be accepted, assign charge states to ions, remove peaks below a certain intensity and merge MS/MS spectra that were derived from the same precursor ion (Cottrell 2011). The management of such large data sets has seen the introduction of various bioinformatics programs and algorithms to filter the raw data and convert data to protein identifications.

1.5.4.1. Protein Identification algorithms and databases

Sequest (MacCoss, Wu and Yates 2002) and MASCOT (Savitski, Nielsen and Zubarev 2005) are search engines used to process the data generated by MS. The software evaluates protein sequences from a database to calculate the list of peptides that could result from each, *in silico*. The intact mass of the peptide from a sample of interest is known from the mass spectrum, and Sequest uses this information to highlight candidate peptides sequences that could meaningfully be compared. For each candidate peptide, Sequest compares these theoretical spectra to the observed tandem mass spectrum by the use of cross correlation scoring. The candidate sequence with the best matching theoretical tandem mass spectrum is reported as the best identification for this spectrum.

By using an appropriate scoring algorithm, the closest match or matches can be identified. If the "unknown" protein is present in the sequence database, then the aim is to pull out that precise entry. If the sequence database does not contain the unknown protein, then the aim is to pull out those entries which exhibit the closest homology, often equivalent proteins from related species. The sequence databases that can be searched on the public Mascot server are:

SwissProt: a high quality, curated protein database. Sequences are non-redundant, rather than non-identical, so fewer matches are achieved for an MS/MS search than what would be obtained from a comprehensive database, such as NCBInr.

NCBInr: a comprehensive, non-identical protein database maintained by NCBI for use with their search tools BLAST and Entrez. The entries have been compiled from GenBank CDS translations, PIR, SWISS-PROT, PRF, and PDB.

1.5.5. Protein Quantification

Quantitative proteomics comes in two forms: absolute and relative. Relative quantitation compares the levels of a specific protein in different samples with results being expressed as a relative fold change of protein abundance. Absolute quantitation is the determination of the exact amount or mass concentration of a protein, for example, in units of ng/mL of a plasma biomarker.

Quantitative proteomics, as the name suggests, provides quantitative information (relative or absolute) on existing proteins within a sample set. Biological processes are mainly affected by the activity of proteins therefore it is useful to study these proteins directly. Obtaining accurate information on the behaviour of proteins is of the upmost importance as any change in response to external stimulation, or changes between sample sets in a proteomic comparison, may lead to the identification of proteins which control underlying mechanisms of potential interest. Though current quantitative proteomic techniques are far from characterising the entire proteome, there are a number of methods that are successful in determining quantitative information on proteins expression.

Spectral counting as a means of quantifying peptides is based on the idea that the more of a protein is present in a sample, the more MS/MS spectra will be collected for peptides belonging to that protein. The relative abundance of that protein can be obtained by comparing the number of MS/MS spectra between sample sets. Although this method has the advantage of simultaneous quantification and identification, it depends on the quality of MS/MS generated

Peak intensity measurement (also called 'signal intensity' or 'area under the curve' (AUC)), is a method used to calculate the relative quantity of peptides in a label-free quantitative analysis. In LC-MS, as peptides are eluted from the reverse phase column into the mass analyser, a particular mass to charge ratio (m/z) is recorded with a particular intensity, at a particular time. This results in a peptide of a particular charge and mass generating one mono-isotopic mass peak. It has been observed that the signal intensity of this peak correlates with ion concentration. The peak areas of identified peptides can therefore be extracted and calculated, and these peak areas increase linearly with increasing peptide concentration (Voyksner and Lee 1999). These peptides can be brought together to provide quantitative data on relative protein abundances between samples in a comparison. This method is applicable to data derived from high mass precision instruments, and uncouples quantification from the identification process. The accuracy of quantitative data can be further improved by the normalisation of peak areas. This method requires that all data is collected in data-dependent 'Triple Play' mode (allowing MS scan, and MS/MS scan). This AUC method of peptide quantification is used in this thesis to achieve quantitative proteomic profiles of the DLKP cell line and its subpopulations.

1.6. Proteomic Profiling

Proteomic profiling is a type of quantitative proteomics that identifies differences in protein expression across multiple samples. The quantitative analysis of protein levels is important for understanding the function of each protein and to provide insights into mechanisms of various biological processes and diseases. As previously mentioned, quantitative proteomics employs mass spectrometry to measure protein expression either in relative or absolute values. By gathering information about protein expression levels in a comparison between biological samples, proteomic profiles can be generated. Quantitative proteomic profiles can subsequently be exploited to target key proteins potentially involved in the known characteristics of each sample compared.

Since the proteome of an organism is far too complex to be analysed in a single step, simplification of the sample is necessary prior to analysis. Combined with the continued advances in mass spectrometry, gel-based quantitative proteomic profiling is

now a routine method for biomarker discovery using methods such as 2D-PAGE and 2D-DIGE. There have been great advances made in mass spectrometry based methods of protein quantification available using labelled techniques such as SILAC, iTRAQ, and TMT. The recent development in proteomic profiling technology is what is referred to as the quantitative label-free LC-MS/MS method, which is the basis of the work presented in this thesis.

1.6.1. Labelling techniques for Quantitative proteomic profiling

1.6.1.1. Two-Dimensional Polyacrylamide Gel Electrophoresis (2D-PAGE)

One of the earlier methods used for the separation of proteins is two-dimensional polyacrylamide gel electrophoresis (2-D PAGE). The first step involves the separation of equal quantities of proteins across a sample set, based on their isoelectric point (pI). The pI of a protein is described as the pH at which its net charge is zero. Using this property, protein samples are separated on immobilised pH gradient (IPG) strips in the first dimension and rest at certain locations along the IPG strip depending on their pI. Following this, the IPG strips are incorporated into the top of a large polyacrylamide gel sheet and a charge is applied. Proteins migrate from the strip into the gel and are separated based on their size, using sodium dodecyl sulphate electrophoresis, in the second dimension. The result of this process is a 2-D gel which upon staining produces a series of spots, with each spot corresponding to a protein (O'Farrell 1975). Commonly used protein stains include Coomassie Brilliant Blue and silver nitrate colloidal stain, both of which are global protein stains. Subsequently, different polyacrylamide gel sheets representing separate samples can be compared using imaging software to look for abundance changes in matching protein spots. These can be picked and identified by mass spectrometry.

This method is time consuming, labour intensive and results in poor reproducibility between gels. The inherent variability of this technique makes it difficult to distinguish between system variation and biological changes between samples being compared. This makes it hard to attribute protein abundance changes in a diseased state vs. normal conditions, resulting in a technique which is not very reliable (Marouga, David and Hawkins 2005).

1.6.1.2. Two-Dimensional Difference In-Gel Electrophoresis 2D-DIGE

2D-DIGE is based on the same principle of 2D-PAGE however; it aims to overcome the problems associated with the latter technique. It incorporates the use of fluorescent dyes which can be used to provide quantitative information on protein abundance differences between samples. This highly sensitive method of minimal protein labelling uses three fluorescent dyes (Cy2, Cy3, and Cy5) which are covalently bound to protein samples just prior to co-electrophoresis. This method has an advantage over 2D-PAGE in that the Cy labelled sample set can be run together using co-electrophoresis in both the first and second dimension. In addition, this method incorporates the use of a pooled sample internal standard which is included on all gels in the experiment. This allows normalisation of the resulting abundance data resulting in accurate reproducibility and protein comparisons between gels (Alban et al. 2003).

Gel based proteomic techniques, in particular 2D-DIGE, have been successfully used as a tool for examining pathological processes such as in Alzheimer's disease, as well identifying biomarkers in the saliva of lung cancer patients, and identifying proteins involved in radio-resistance in breast cancer (Souchelnytskyi 2005, Xiao et al. 2012, Duru et al. 2012). However, these techniques have drawbacks which limit the full potential of proteomic profiling. These include difficulties in accommodating hydrophobic proteins into the method, and an inability to achieve an entire representation of the proteome. Multiple proteins may be overlapping in one protein spot, which can lead to incorrect identification of a differentially expressed protein downstream. In addition, sample bottlenecks are a major limitation of 2D methodology due to the limited number of samples which can be analysed at any one time, and intensive image analysis required for data processing. In view of these limitations, many gel-free techniques have been developed which attempt to overcome these obstacles, and these techniques are centred on the use of mass spectrometry.

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

SILAC is an *in vivo* metabolic labelling technique where the isotopic labelling is performed when cells are in culture. Amino acids in growth medium are modified so they contain stable isotopic nuclei (e.g. deuterium, ^{13}C and ^{2}H), and two cell populations

can be grown so that they remain identical, except one of them contains 'light' amino acid while the other contains 'heavy' amino acids. These modifications are incorporated into each newly synthesised peptide chain, and only amino acids which are essential to the organisms in the comparative experiment can be used (Ong et al. 2002). These modifications result in a mass shift which can be readily distinguished by mass spectrometry analysis and relative quantitation is achieved by comparing heavy vs. light labelled cells. SILAC has the ability to compare up to five conditions in one experiment and is only limited by the availability of heavy forms of amino acids.

SILAC has been successfully used as a method to study posttranslational modifications such as protein phosphorylation and methylation in mammalian cells, and to analyse the dynamics of signal-dependent phosphorylation events in plants (Gruhler and Kratchmarova 2008, Engelsberger et al. 2006), highlighting its diverse applications. Other advantages include its compatibility with cell culture, no requirement for affinity purification steps, and the ability to include a pooled internal control to find real protein abundance differences independent of variability from processing steps. The main drawback of SILAC are the limited availability of heavy amino acid types, limiting sample set size (up to 5-plex), and introduction of variability during the labelling process (Beynon and Pratt 2005).

1.6.1.4. Isotope-Coded Affinity Tags (ICAT)

ICAT is an *in vitro* chemical labelling technique which allows protein quantification by LC-MS analysis through the use of heavy or light isotopes (Monteoliva and Albar 2004). Chemical incorporation of isotope tags is performed after protein extraction, with control and test conditions being modified by with heavy and light ICAT reagent followed by Tryptic digestion. The ICAT reagent has three components: a biotin tag, an oxyethylene linker region, and a sulfhydryl-reactive iodoacetate group which modifies Cysteine residues in proteins (Gulcicek et al. 2005). ICAT labelled peptides are separated and analysed using LC-MS, with peptide pairs co-eluting from different experimental conditions. Calculating the ratio of the areas under the curve for both peptides allows the relative abundance of that peptide to be determined, which corresponds to the protein level.

There are a number of limitations associated with the ICAT method. Similar to SILAC, the limited availability of isobaric tags limits the number of samples which can be included in an experiment. Labelling is also dependant on the presence of Cysteine residues, and therefore this method is unsuitable for proteins which lack these residues. (Monteoliva and Albar 2004). However, this can be turned into an advantage for experiments in which the goal is to enrich for proteins containing Cysteine residues. This method has been successfully implemented in the discovery of biomarkers and is able to detect and quantify low abundant proteins in complex mixtures (Gygi et al. 2002).

1.6.1.5. Isobaric Tag for Relative and Absolute Quantification (iTRAQ)

iTRAQ is another method which uses isobaric tags for the simultaneous identification, both relative and absolute, of proteins in a comparative sample set. It utilizes labelled amine modifying chemistry with tandem mass spectrometry (MS/MS). This multiplexing technique allows up to eight samples (8-plex) to be analysed in a single experiment (Choe et al. 2007). iTRAQ tags react with all primary amines of peptides, which means that all peptides are labelled and information about their post-translational modifications are retained. It requires the use of tandem-mass spectrometry with isobaric reporter ions being detected in the MS/MS scans.

ITRAQ is based on the incorporation of different isobaric tags, during fragmentation in the second stage of MS/MS, tags are fragmented along with peptides. 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 differentially labelled samples (Unwin 2010). The nature of the isobaric tags means that the same peptide from each of the samples in an experiment appears as a single peak in the mass spectrum. This reduces the data complexity associated with ICAT where heavy and light versions of each peptide need to be discerned (Fuller et al. 2010).

1.6.1.6. Tandem Mass Tags (TMT)

TMTs are a recently developed method for the isobaric labelling of peptides which permits simultaneous determination of both the identity and relative abundances of peptide pairs from multiple samples. These tags are designed so that on analysis by collision-induced dissociation (CID), the TMT fragment is released to give rise to an ion with a specific mass-to-charge ratio. The TMT approach is similar in principle to other peptide isotope labelling techniques however offers some additional advantages. TMT tagged peptides have the same overall mass and co-migrate in chromatographic separations and, thus, will act as more precise reciprocal internal standards, which leads to more accurate quantification (Thompson, 2003). The TMT 10-plex technique allows for the simultaneous identification of proteins from 10 different samples and this method is compatible with samples derived from cells, tissues or biological fluids.

1.6.2. Quantitative Label-Free Proteomics

Quantitative label-free LC-MS/MS is a mass spectrometry based method which aims to measure the quantity of proteins in two or more biological samples without the use of labelling. It is an alternative to both gel based and labelling techniques, and seeks to overcome the drawbacks associated with those methods. Isotopic labelling of proteins or peptides is not always practical and has several disadvantages. Labelling with stable isotopes is an expensive technique and can be time-consuming and a very labour-intensive method. Moreover, there may not be enough different isotopes to allow for simultaneous quantitation of proteins from multiple samples. The most current method of protein labelling is TMT however this relatively new technique only allows for 10-plex labelling. As an alternative, many groups have presented methods of label-free peptide and protein quantification by comparing peptide signal intensities measured in sequential MS analyses.

Quantitative label-free LC-MS/MS proteomic analysis is a popular current method of protein preparation whereby sample handling and potential contaminants are reduced allowing for high-throughput of samples. Quantitative label-free proteomics is based on the direct comparison of MS signal intensities between different peptides in a mass spectrometer (Zhu, Smith and Huang 2010). Quantitative proteomics is quickly becoming the method of choice for the analysis of complex samples due to its high

sensitivity and exceptional dynamic range. This technique involves the use of nano-high performance liquid chromatography (HPLC) with reverse phase chromatography prior to mass spectrometry analysis. Quantitative label-free LC-MS/MS was the main technique used in the discovery phase of the work presented in this thesis, and all subsequent work emerged from the proteomic data which was generated by it. Label-free proteomics identifies proteins through a combination of bioinformatics and mass spectrometry. The label-free approach is based on the separate LC-MS/MS analysis of all samples, followed by retention time control and normalisation between the generated MS/MS spectra through the use of specific label-free analysis software programmes such as Progenesis LC-MS (Non-Linear Dynamics). Quantitative LC-MS/MS can be broken into two groups: peak intensity measurement based on precursor ion spectra, and spectral counting both of which are explained in Section 1.5.4.

In order to provide a comprehensive overview of the basis of this quantitative proteomic profiling technique, what follows here is a short synopsis on each of the main elements involved. This will incorporate a brief outline of nano-HPLC coupled to mass spectrometry, the components which make this up and how proteins can be identified and quantified using this technology. The separation of complex samples is achieved through liquid chromatography prior to mass spectrometry. Liquid chromatography coupled to mass spectrometry (LC-MS) allows profiling of large numbers of peptides from complex mixtures and is an ideal method for comparing samples with different biological conditions. The introduction of nano-LC enables low flow separations and facilitate easy coupling to mass spectrometry to provide high resolution, sensitivity and selectivity for LC-MS proteomic applications (Gaspari and Cuda 2011). Quantitative label-free proteomics takes advantage of the robust and reproducible LC-MS technique in order to gather information from samples in sequential MS analyses.

As previously mentioned, mass spectrometric instruments have undergone extensive development in the last decade. Modern mass spectrometers allow high-resolution sequencing, identification of post-translational modifications and have also significantly increased the number of protein identifications per run. Tandem mass spectrometry has also greatly advanced the field of mass spectrometry. It is this advancement which allows for peptide and subsequent amino acid identification, facilitating label-free quantitative accuracy. The development of ion trap-orbitrap MS

machinery allows high resolution, increased sensitivity compared to MS alone and provides a wider linear range of quantitation. The development of such state-of-the-art MS instruments has allowed for precision quantitation of biological samples thus enabling reliable quantitative label-free LC-MS/MS to be achieved.

Increased numbers of protein databases, programs for quantitative raw data analysis, peptide search engines and protein scoring algorithms has provided great confidence in protein identifications and measurements. Data analysis programs are becoming more automated, user-friendly and accessible to researchers removing the need for bioinfomaticians to carry out in-depth interpretations of MS raw spectra and signal intensities. Current programs allow raw data files to be imported, basic filters to be applied such as choosing the genome of the sample being studied, and in a matter of minutes raw data will be converted to protein identification lists for each sample. Quantitative data analysis programs such as Progenesis, are slightly more complex, but advances in user friendly interfaces in these programs has enabled researchers to carry out their own differential analysis. Scoring algorithms such as Sequest and MASCOT give an indication of the confidence the program has in the peptide matches. Peptides can be filtered out of the data if their scoring or confidence is not high. Applying strict criteria and scoring thresholds to quantitative data analysis means only robust and strong protein candidates are presented. Increased confidence in protein identifications has thus given increased confidence to peptide and protein identifications obtained through quantitative label-free LC-MS/MS.

Many groups have employed quantitative label-free proteomics to analyse samples and thus create proteomic profiles for biological states. The quantitative label-free LC-MS/MS technique that is used for this thesis study was also employed by Linge et al. (2014) to investigate proteins involved in human breast cancer progression. Comparative label-free LC-MS/MS profiling was carried out using nano-LC coupled to a hybrid linear ion trap/ Orbitrap mass spectrometer. Label-free LC-MS quantitative analysis was carried out using Progenesis software, and MASCOT scoring criteria was applied to the data. This analysis revealed 45 proteins to be upregulated and 34 proteins to be downregulated when comparing the highly invasive to the poorly invasive breast cancer cell lines. Subsequent functional analyses of these deregulated proteins found RAD23B to have a key role in invasion and adhesion of breast cancer cells in vitro. This

study demonstrates that the utilisation of label-free quantitative LC-MS/MS to identify proteins is a valuable approach for characterising and profiling the phenotypes of various cell lines.

A similar breast cancer quantitative label-free proteomics was performed by Dowling et al. (2014) to profile the proteome of serum from patients with advanced breast cancer. Nano LC-MS/MS analysis of patient serum samples identified a number of proteins involved in the complement and coagulation cascades to be upregulated in cancer serum samples. The elevated levels of such proteins were found to reflect the advanced stage of cancer where most tumour processes are in over-drive. Such proteins could be used as potential biomarkers for disease staging and patient prognosis in breast cancer. A quantitative proteomic approach is therefore a useful tool in identifying proteins associated with disease progression which contribute to the functional processes allowing carcinogenesis.

Okayama et al. (2014) carried out proteomic analysis of proteins related to the prognosis of lung adenocarcinoma by employing shotgun proteomics. LC–MS/MS analysis of 21 cancerous tissue samples detected and identified a total of 875 proteins. Relative quantitative analysis revealed that 17 proteins were preferentially expressed in the poor prognosis group relative to the good prognosis group. Label-free protein relative quantification was performed using the Progenesis LC−MS software and scoring was achieved through the use of MASCOT. Among the 17 overexpressed proteins in the poor prognosis group, many of the proteins regulate cell adhesion and morphology and thus their overexpression might be involved in cancer progression. The results of this study strongly support the feasibility of a quantitative proteomic approach for identifying significant proteins associated with poor prognosis in cases of lung adenocarcinoma.

Quantitative label-free LC-MS/MS was the main technique used in the discovery phase of the work presented in this thesis and all subsequent work flowed from the proteomic data which was generated by it. The above mentioned studies justify the use of quantitative mass spectrometry as a tool to profiling the DLKP lung carcinoma cell line and its subpopulations. Quantitative proteomics is quickly becoming the method of choice for the analysis of complex samples due to its high sensitivity, high reproducibility and accurate identifications.

1.7. Aims of Thesis

- To identify differentially expressed proteins between the clonal subpopulations (DLKP-SQ, DLKP-I and DLKP-M) of a heterogeneous lung carcinoma cell line (DLKP) using quantitative label-free LC-MS/MS analysis, and interrogate the data to generate lists of the most significant differentially expressed protein targets.

- To perform multiple analyses using various comparative proteomic experimental designs to highlight the strongest protein candidates associated with high expression in each clonal subpopulation compared to the other clones.

- To determine potential individual protein markers for each clonal subpopulation based on their unique or high abundance in each cell line based on the quantitative LC-MS/MS analysis. Following this, validation studies will be carried out to support the use of the protein target as a marker for a clonal subpopulation within the DLKP cell line model.

- To study the isolated clonal subpopulations with respect to their associated high abundance protein using functional cell-based assays, and to elucidate their role in relation to invasion and migration in cancer cells *in vitro*. Using these techniques may improve our understanding of the metastatic phenotype of the disease.

- To investigate the potential binding partners of highly expressed proteins representing each clonal subpopulation. This may uncover unique protein interactions in this cell line model, and shed light on the role played by the protein targets in their respective cell lines.

- To investigate the effects of co-culture of the clonal subpopulations in order to determine the effects on protein expression. The tumour microenvironment is heterogeneous in nature, and cell-cell interactions may play an important role

in how proteins are expressed and consequently on the phenotypic behaviour of the cells.

Chapter 2. Materials and Methods

2.1. Preparation of Materials for Cell Culture

2.1.1. Ultrapure Water:

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).

2.1.2. Glassware:

All glassware and lids pertaining to cell culture in any capacity were prepared as follows: Items were soaked in a 2% RBS-25 (Chemical Products R. Borghgraef S.A.) for 1 hour, after which they were scrubbed and rinsed using tap water. Following this they were washed in an industrial dishwasher using Neodisher detergent, which is an organic, phosphate-based acid detergent. Prior to sterilisation all items were rinsed twice with UHP.

2.1.3. Sterilisation:

Water, glassware and all thermo-stable solutions were sterilised by autoclaving at 121°C for 20 minutes at 15psi. Thermo-labile solutions were filtered through 0.22 μm sterile filters (Millipore, Millex-GV SLGV025BS). Low protein-binding filters were used for all protein-containing solutions.

2.2. Cell Culture

All cell culture work was carried out in a class II laminar air-flow cabinet (Nuaire). The laminar air-flow cabinet was cleaned with 70% industrial methylated spirits (IMS) before and after use. Any items brought into the cabinet were also swabbed with IMS. At any time only one cell line was manipulated in the laminar air-flow cabinet, and upon completion of work, 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. All cells were incubated at 37°C and where required, in an atmosphere of 5% CO_2. Cells were fed with fresh media or sub-cultured every 2-3 days or as required in order to maintain active cell growth.

Details of cell culture media and cell lines used in this study are shown in Table 2.1 and Table 2.2. Cell lines were maintained in T25; T75; and T175 flasks (Corning).

2.2.1. Cell Culture Media and Cell Lines:

Media was stored in the dark at 4°C until one month before expiry. Complete medium was prepared by combining the components described in Table 2.1 . Complete media was stored at 4°C in the dark for up to one month. Aliquots of newly made cell culture media were taken and allowed to incubate at 37°C for up to 7 days to ensure sterility. Signs of contamination were assessed regularly by checking for turbidity in the growth media.

Cell Line	Passage Number	Media
DLKP	16	DMEM/F12 (D8437), 5% Fetal calf serum, Hyclone (A1011-1852).
DLKP-SQ	32	DMEM/F12 (D8437), 5% Fetal calf serum, Hyclone (A1011-1852).
DLKP-M	30	DMEM/F12 (D8437), 5% Fetal calf serum, Hyclone (A1011-1852).
DLKP-I	30	DMEM/F12 (D8437), 5% Fetal calf serum, Hyclone (A1011-1852).

Table 2.1: Details of media requirements for cell lines used in this study.

2.2.2. Subculture of Adherent Cell Lines

Exhausted cell culture medium was removed from the tissue culture flask and discarded into a sterile bottle. The flask was then rinsed out with 2 mL of PBS solution to ensure the removal of any residual media. Depending on the size of the flask, 2-5 mL of trypsin solution (0.25% (v/v) of Trypsin (Gibco, 043-05090) and 0.01% (v/v) of EDTA (Sigma, E9884) solution in PBS (Oxoid, BRI4a)) was then added. Cells were incubated at 37°C for approximately 5 minutes until all of the cells detached from the inside surface of the flask. This was monitored by microscopic observation. An equal volume of complete media was added to the flask to deactivate 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 fresh complete

medium. A cell count was performed as described in Section 2.2.3 and an aliquot of cells was used to seed a flask at the required density. All cell waste or media exposed to cells was autoclaved before disposal.

Cell Line	Details and Histology	Invasion status
DLKP **(McBride et al. 1998b)**	Poorly differentiated squamous cell lung carcinoma.	Low invasive
DLKP-SQ **(McBride et al. 1998b)**	Squamous subpopulation cloned from DLKP	Low invasive
DLKP-M **(McBride et al. 1998b)**	Mesenchymal-like subpopulation cloned from DLKP	High invasive
DLKP-I **(McBride et al. 1998b)**	Intermediate subpopulation cloned from DLKP	High invasive

Table 2.2: Cell lines used in this work with their details and invasive status.

2.2.3. Cell Counting

An aliquot of the cell suspension was then added to trypan blue (Gibco, 525) at a ratio of 1:1 (v/v). The mixture was incubated for 2-3 minutes at room temperature. An aliquot (10 µL) 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. Non-viable cells stained blue, while viable cells excluded the trypan blue dye. The average number of viable and dead cells per 16 squares was multiplied by a factor of 1 X 10^4 (volume of the grid) and the relevant dilution factor to determine the average cell number per mL in the original cell suspension. Using the data for viable and non-viable cells, percentage viability was calculated.

2.2.4. Cryopreservation of Cells

Cells for cryopreservation were harvested in the mid-log phase of growth and counted as described in Section 2.2.3. Cell pellets were resuspended in a suitable volume which of FCS which had been cooled to 4°C. An equal volume of filter sterilized solution of 10% (v/v) DMSO in serum was added drop-wise with mixing into the cell suspension. 1mL of cell suspension was then aliquoted into cryovials (Greiner, 122278) and immediately placed on ice for 15 minutes. Following this, cryovials were transferred into the -20 °C freezer for 1 hour and then placed in at -80°C for four hours or overnight. The cryovials were then transferred to liquid nitrogen tank for long term storage (-196°C).

2.2.5. Thawing of Cryopreserved Cells

A volume of fresh complete culture medium pre-warmed to 37°C and 8 mL was added to a sterile universal. The cryopreserved cells were removed from the liquid nitrogen and thawed at 37°C as quickly as possible. The cells were removed from the vials and transferred into the pre-warmed aliquoted media in the universal. The resulting cell suspension was centrifuged at 170 g for 5 minutes. The supernatant was removed and the pellet was resuspended in the pre-warmed culture medium. This cell suspension was then transferred to a T-75 cm^2 flask and allowed to attach and grow overnight in an incubator. The following day, the culture media was replaced with fresh complete culture medium to remove any non-viable cells and floating cells.

2.2.6. Mycoplasma Analysis of Cell Lines

Mycoplasma testing was carried out for all cell lines for possible Mycoplasma contamination in house by Mr. Michael O' Donoghue at the NICB.

2.2.6.1. Indirect Staining Procedure for Mycoplasma Analysis

Normal rat kidney fibroblast (NRK) cells were seeded onto sterile coverslips in sterile petri dishes (Greiner, 633 185) at a cell density of $2x10^3$ cells/mL and were allowed to attach overnight at 37°C in a 5% CO_2 humidified incubator. 1 mL of cell-free supernatant from each test cell line was inoculated onto an NRK petri-dish and incubated as before until the cells reached 20-50% confluency. After this time, the waste medium was removed from the dishes, the coverslips (Chance Propper, 22 x 22 mm) were washed twice with sterile PBS, once with a cold PBS/Carnoys (50/50) solution and fixed with 2 mL of Carnoys solution (acetic acid:methanol, 1:3) for 10 minutes. The fixative was removed and dried coverslips were washed twice in deionised water and stained with 2 mL of Hoechst stain (BDH) (50ng/mL) for 10 minutes. From this point on, work proceeded without direct light to limit quenching of the fluorescent stain. The coverslips were rinsed three times in PBS. They were then mounted in 50% (v/v) glycerol in 0.05 M citric acid and 0.1 M disodium phosphate and examined using a fluorescent microscope with a UV filter. A Mycoplasma infection would be seen as small fluorescent bodies in and sometimes outside the cells.

All cell lines used in this thesis are confirmed to be Mycoplasma free.

2.3. Western Blot Analysis

2.3.1. Preparation of Whole Cell Protein Lysates

Cells were grown to the desired confluency in monolayer culture flasks (T75/T175). Media was removed and cells were washed twice with sterile PBS. Excess PBS was removed and 500 μL of '2D lysis buffer' was added directly to the monolayer of cells in a T75 flask, 1 mL for a T175 flask. 2-D lysis buffer contains 7 M urea (Sigma Aldrich, 208884), 2 M thiourea, 4% (w/v) (Sigma Aldrich, T8656), CHAPS (Sigma Aldrich, C3023), 40 mM DTT (Sigma Aldrich, D9163). The flask was incubated on ice for 5 minutes, and gently swirled regularly to ensure complete coverage of the monolayer by the lysis buffer. The cell lysate was scraped down to a corner of the flask using a P1000 tip and gently aspirated five times. The cell lysate was then transferred to a 1.7 mL Eppendorf tube and incubated at room temperature for 30 minutes after the addition of protease inhibitor (Thermo Fisher Scientific, 78430). This was centrifuged at 18,620 g for 15 minutes at 4°C, and the supernatant was removed and stored at -80°C for future use.

2.3.2. Protein Quantification

Protein concentration was determined using the thiourea-compatible Quick Start Bradford Protein Assay Kit (Bio-Rad, 500-0201), containing 2 mg/mL of bovine serum albumin (BSA) solution as a known standard. Dilutions of BSA stock for 0.125, 0.25, 0.5 and 1.0 mg/mL were prepared and used for generating a protein standard curve. Samples for quantification were diluted to reduce their concentration so as to fit within the range of the assay. Dilution factor was determined based on a visual test by mixing neat protein lysate with the assay reagent prior to sample preparation. 5 μL of protein standard dilutions, and prepared samples were added to a 96-well plate. 250 μL of thiourea-compatible Bradford protein assay reagent (Bio-Rad, 500-0205) was then added each protein containing well in the plate. After 5 minutes incubation, the absorbance was assessed at 595nm. All samples were assayed in triplicate. The concentration of the protein samples was determined from the plot of the absorbance at 595nm versus the concentration of the protein standard.

Protein concentration was assayed using the Pierce BCA Protein Assay Kit (Cat#23227), when lysis buffers were incompatible with the Bradford assay.

2.3.3. 1-D Gel Electrophoresis

Proteins for analysis by Western blotting were resolved using 4-12% NuPAGE Bis-Tris Gels (Life Technologies, NP0322BOX) in XCell SureLock™ Mini-Cell (Life technologies, EI0001) running instrument. Protein samples were prepared for 1-D gel electrophoresis by taking the required volume to equalise the protein concentration across a sample set from the protein lysate sample preparation as described in Section 2.3.2. Volumes of all samples were then equalised by the addition of PBS. Finally an equal volume of 2X loading buffer (Sigma, S3401) containing *mercaptoethanol* and bromophenol blue was added to each sample. The samples were heated at 95 °C for 5 minutes on a heating block and allowed to cool to room temperature before being centrifuged. 10-40 µg of protein and 5 µL of molecular weight marker (Thermo, 26634) were loaded into the wells of a gel. The samples were electrophoretically separated at 200 V and 45 mA using a 1X MOPS/SDS running buffer (Life Technologies, NP001) until the bromophenol blue dye reached the bottom of the gel.

2.3.4. Western Blotting

Once electrophoresis had been completed, the gel was equilibrated in a 1X tris-glycine transfer buffer (Bio-Rad, 161-0734) containing 20% methanol for approximately 10 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 by rolling with a clean pipette. PVDF membrane (GE Healthcare, 10600021) was activated by soaking in 100% methanol and equilibrated using the transfer buffer prepared earlier. This was placed over the filter paper on the cathode plate. Air pockets were once again removed. The gels were then placed onto the membrane, and 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. Proteins were transferred from the gel to the membrane at a current of 340 mA at 15 V for 23-30 minutes depending on the Mw of the protein under investigation.

Following protein transfer, the PVDF membrane was blocked for two hours using 5% Marvel (Cadburys; Marvel skimmed milk) in TBS-T (a mixture of 1X tris-

buffered saline (TBS) (Sigma, t5912) containing 0.05% Tween 20 (Sigma, P5927)). The membrane was washed with TBS-T prior to the addition of the primary antibody, prepared in 5% Marvel in TBS-T at recommended dilutions, and then incubated overnight at 4 °C. The membrane was then rinsed 3 times with TBS-T for a total of 30 minutes. Relevant secondary antibody (1/2000 dilution of anti-mouse (Dako Cytomation, P0260) or anti-rabbit (Dako Cytomation, P0448) was added for 2 hours at room temperature. The membranes were again washed three times thoroughly in TBS-T for 30 minutes.

All primary and secondary antibodies used for Western Blot analysis are shown below in Table 2.3.

Antibody	**Dilution and Application**	**Details**
Shootin-1	1:1000	Pierce: PA5-17167
MARCKS	1:5000	Abcam: ab52616
Desmoglein-3	1:2000	Abcam: ab78448
Desmoglein-3	1:3000	Bio-Sciences: #326300
Striatin-3	1:500	Pierce: MA1-46461
Src-Substrate Cortactin	1:1000	Pierce: PA5-17273
Semenogelin-1	1:500	Abcam: ab139405
Junction Plakoglobin	1:1000	Pierce: PA5-17320
Pyruvate Dehydrogenase E1	1:500	Abcam: ab168329
Pyruvate Dehydrogenase E2	1:1000	Abcam: ab131200
Beta-Actin	1:10,000	Abcam: ab20272
Alpha Tubulin	1: 3000	Abcam: ab4074
Secondary Antibodies	**Dilution and Application**	**Details**
Anti-Mouse	1:2000	Dako: P0447
Anti-Rabbit	1:2000	Dako: P0448

Table 2.3: List of all antibodies used in this study for Western Blot along with their dilutions.

2.3.5. Enhanced Chemiluminescent Detection Using Autoradiographic Film

Western Blots were developed using an enhanced chemiluminescence (ECL) kit (GE Life Sciences, RPN2106), which facilitated the detection of bound peroxidase-conjugated secondary antibody. Following the final washing, membranes were subjected to ECL. The membrane was placed on a sheet of transparent plastic and 5 mL of a freshly prepared 1:1 (v/v) mixture of ECL reagent A and B was used to cover the membrane. The ECL reagent mixture was completely removed after a period of five minutes and the membrane was covered in a layer of transparent plastic. All excess air bubbles were removed. The membrane was then exposed to autoradiographic film (GE Life Sciences, 95017-681) for various times (from 10 seconds to 15 minutes depending on the intensity of the signal). The exposed autoradiographic film was developed for the appropriate time in developer solution (Kodak, LX24, diluted 1:5 in water) until clear bands developed. The film was then briefly washed in water and transferred to a fixative solution (Kodak, FX-40, diluted 1:5 in water) for 5 minutes. The film was washed with water for 5-10 minutes and left to dry at room temperature.

2.4. RNA Interference (RNAi)

RNAi using small interfering RNA (siRNA) was carried out to transiently knockdown the expression of Shootin-1, MARCKS, and Desmoglein-3 at the protein level. Two independent siRNA molecules were used for each protein (Life Technologies). These siRNAs were 21-23 bps in length and were introduced to the cells via reverse transfection with Lipofectamine 2000 Transfection Reagent (Life Technologies, 13778075).

2.4.1. Transfection Optimisation

In order to determine the optimal conditions for siRNA transfection, cell concentrations required for RNAi transient transfections were first established for each cell line. Cell suspensions were prepared at $5X10^4$, $7.5 X 10^4$, $1X 10^5$, $2.5 X 10^5$, and $5 X 10^5$ cells per well, and were allowed to grow for 72 hours. This was carried out in order to determine which cell concentration allows the cells to be in late exponential phase at the end of the transfection (after 72 hours incubation), ensuring healthy proliferating cells will be used in the follow on functional assays. Transfections were optimised further to determine optimal transfection reagent volume required to efficiently transfect each cell type, and optimal the siRNA concentrations to achieve the greatest knockdown of specific targets.

The general method used for siRNA transfection is as follows. Solutions of siRNA at their required concentrations were prepared in optiMEM (GibcoTM, 31985047). Separately, a Lipofectamine 2000 solution was made up as a master mix using optiMEM with enough volume required for all transfection samples. Both of these solutions were incubated at room temperature for 12 minutes before being combined together. Each of the newly combined siRNA/Lipofectamine solutions were incubated at room temperature for a further 12 minutes. 100 µL of the siRNA/Lipofectamine solutions were added in a drop-wise fashion to each well of a 6-well plate, with each well containing 1 mL of a cell suspension at the required concentration. The plates were mixed gently and incubated at 37 °C for 6 hours. After this time, the transfection mixture was removed from the cells and the plates were fed with fresh complete medium.

2.4.2. siRNA Transfection for Functional Analyses

Two pre-designed siRNAs were chosen for the protein targets and transfected into cells. For each set of siRNA transfections carried out a group of controls were always used which were: Non transfected cells-only control; Lipofectamine-only control cells; and a Negative siRNA transfected cells control. Negative siRNA are sequences that do not have homology to any genomic sequence. The Negative 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 Negative siRNA. This takes into account of any effects due to the siRNA transfection procedure, reagents, and also any random effects of the Negative siRNA. Western blots were used to determine if siRNA had an efficient knock-down effect at a protein-level.

All siRNA used in these experiments are shown in Table 2.4.

Target	siRNA Details	siRNA Concentration
Shootin-1	s33622; s33623, Ambion	20 nM
MARCKS	s8385; s8386, Ambion	20 nM
Desmoglein-3	s4327; s4328, Ambion	10 nM
Negative siRNA No. 1	4390843, Ambion	As required
Negative siRNA AllStars	1027280, Qiagen	As required
Kinesin (Kif11)	s7903, Ambion	5 nM

Table 2.4: siRNAs used in this study with their details and their required concentrations.

2.5. Functional Analyses

2.5.1. Acid Phosphatase Assay

Following an incubation period of 72 hours, media was removed from the plates. Each well on a 6-well plate was washed twice with 2 mL PBS. This was then removed and 2 mL of freshly prepared phosphatase substrate (10 mM p-nitrophenol phosphate (Sigma 104-0) in 0.1 M sodium acetate (Sigma, S8625), 0.1% triton X-100 (Sigma, X100), pH 5.5) was added to each well. The plates were then incubated in the dark at 37 °C for 2 hours. Colour development was monitored during this time. The enzymatic reaction was stopped by the addition of 1 mL of 1 M NaOH. A 150 μL aliquot of each sample (5 replicates) was transferred to a 96-well plate. The plate was read in a dual beam plate reader at 405 nm with a reference wavelength of 620 nm.

2.5.2. Proliferation Assay

Proliferation assays on transfection optimisation steps were carried out as described in Section 2.5.1. For functional analysis of proteins of interest, proliferation assays were performed using the cell counting method as described in Section 2.2.3 in order to assess cell viability in conjunction with proliferation.

2.5.3. Invasion Assay

2.5.3.1. Preparation of Inserts

Matrigel (BD Biosciences, 354234) was diluted to a working stock of 1 mg/mL in serum- free DMEM. Aliquoted stocks were stored at -20°C. A volume of 100 μL of Matrigel was placed into each insert (BD Biosciences, 353097) (8.0 μm pore size, 24 well format) and kept at 4°C for 24 hours. The insert and the plate were then incubated for one hour at 37°C to allow the proteins to polymerise. Cells were harvested and resuspended in culture media containing 5% FCS at 1×10^6 cells/mL. Excess media was removed from the inserts, and they were rinsed with 200 μL of culture media. A 100 μL volume of complete cell culture media was added to each insert. A 100 μL volume of cell suspension was added to each insert and 500 μL of culture media containing 5%

FCS was added to the well underneath the insert. Cells were incubated for 24 hours. Cell numbers seeded in each assay are shown in Table 2.5.

Cell Line	Cell Seed Number per Insert
DLKP-SQ	$1x10^5$ cells
DLKP-M	$7.5x10^4$ cells
DLKP-I	$1x10^5$ cells

Table 2.5: Seeding number for each cell line used for migration and invasion assays.

2.5.3.2. Staining of Invaded Cells

After the 24 hour incubation, the inside of the insert was wiped with a cotton swab dampened with PBS, while the outer side of the insert was stained with 0.25% crystal violet for 10 minutes and then rinsed in distilled water (dH2O) and allowed to dry. The inserts were viewed and photographed under the microscope.

2.5.3.3. Counting of Invaded Cells

To determine the total number of invading or migrating cells, the number of cells/field in 15 fields was counted at 200x magnification. The average number of cells per field was then multiplied by a conversion factor of 140 (growth area of membrane divided by field of viewed area at 200x magnification). All assays were subjected to statistical analysis using Student's t-tests (2-tailed, 2-sample unequal variance). A minimum of 2 inserts were used per sample tested.

2.5.4. Migration Assay

Migration assays were carried out as described in Section 2.5.3 except inserts were not coated with Matrigel. Cell counts were also carried out as for invasion assays.

2.6. Immunostaining of Fixed Cells.

2.6.1. Fixation of Cells Prior to Staining

30 μL volumes of cells a concentration of 1×10^6 cells/mL were plated directly onto sterilised 10 well, 7 mm microscope slides (Erie Scientific Company, 465-68X) from actively growing cultures. Cells were allowed to attach overnight. After this time, slides were washed 3 times in PBS before fixation. Slides were submerged in a solution of 4% Paraformaldehyde for a maximum of 10 minutes before being permeabilzed with a 0.2% TritonX-100 (Sigma) in PBS solution for 10 minutes. After this process, cell slides were kept in cold PBS, never being allowed to dry out, and used for follow up staining procedures within 1 hour.

2.6.2. Immunocytochemistry

All immunocytochemical (ICC) staining was performed using the DAKO Autostainer (DAKO, S3800). The slides were immersed in wash buffer (DAKO, S3006). On the Autostainer slides were blocked for 10 minutes with 200 μL HRP Block (DAKO, S2023). Cells were washed with 1× wash buffer and 200 μL of primary antibody of interest was added to the slides for 20 minutes. Slides were washed again with 1× wash buffer and then incubated with 200 μL Real EndVision DAKO, K4065) for 30 minutes. A positive control slide was included in each staining run. Each slide was also run with Negative Control Reagent (1× TBS/0.05% Tween-20), to allow evaluation of non-specific staining and allow better interpretation of specific staining at the antigen site. All slides were counterstained with haematoxylin (DAKO) for 5 minutes, and rinsed with deionised water, followed by wash buffer. All slides were then dehydrated in graded alcohols (2 immersions x 3 minutes each in 70% IMS, 90% IMS and 100% IMS), and cleared in Xylene (2 immersions x 5 minutes), and mounted with coverslips using DPX mountant (Sigma, 44581). Slides were sealed with clear nail varnish.

2.6.3. Immunofluorescence

Non-specific binding sites were blocked by incubating with 5% (v/v) normal serum (NS) in PBS for 30 minutes at room temp. The blocking serum was removed and the cells were incubated with the primary antibody of interest diluted in 1% v/v NS in PBS. This was incubated overnight at 4°C. Cells were washed 3× in PBS and secondary antibody (AlexoFluor, Thermo) which was diluted 1:1000 was added for 2 hours at RT. Secondary antibody was removed and cells were washed 3× in PBS. Finally, slides were counterstained using Dapi nuclear stain (Sigma, D9542). After washing 3× in PBS, slides were mounted with ProLong Gold mounting medium (P36930) and covered using a glass cover slip. Slides were sealed using clear nail varnish. Cells were viewed and photographed using a Nikon phase contrast microscope fitted with a mercury-vapour lamp.

2.6.4. Co-Labelling of Cells Transfected Using RNAi

Cells were set up for RNAi transfection as described in Section 2.4.2 with some modifications. In place of using 6-well plates to culture the cells during the transfection, the procedure was carried out using sterilized glass coverslips in their place. The glass coverslips were places in 30 mm tissue culture dishes and the transfection procedure was carried out. Coverslips were used as their thin profile allows better image quality when using confocal microscopy. 48 hours post-transfection, cell slides were scraped once using a P1000 tip to induce a migratory response prior to staining. Cells were stained 72 hours post transfection using an adapted version of the method described in Section 2.6.3. Slides were washed 3× in PBS and blocked using 5% (v/v) normal serum in PBS to reduce non-specific interactions. After blocking, cells were stained with the primary antibody of interest overnight at 4°C, followed by washing 3× in PBS. Staining using fluorophore-conjugated secondary antibody was then carried out using an incubation time of 2 hours at RT. Following another washing step, cells were stained using the F-Actin stain Phalloidin (Thermo, A12379) for 20 minutes at RT. Cells were washed 5× in PBS before being mounted on glass slides with ProLong Gold mounting medium. Imaging was performed using confocal microscopy with the invaluable help of Dr Finbarr O'Sullivan.

Details of antibodies used in staining procedures are shown in Table 2.6.

Antibody	Dilution and Application	Details
Shootin-1	IF-1/5000; ICC-1/5000	Atlas: HPA037943
MARCKS	IF- 1/200	Abcam: ab52616
Desmoglein-3	IF- 25µg/ml; ICC-25µg/ml	Bio-Sciences: #326300
Secondary Antibodies		
Anti-Mouse	1:2000 IF/ICC	Thermo: A11001
Anti-Rabbit	1:2000 IF/ICC	Thermo: A11008

Table 2.6: Details of antibodies used in cell staining methods.

2.7. Quantitative Label-Free LC-MS/MS Analysis

2.7.1. Preparation of Protein from Cell Lines

Cells for analysis by quantitative label-free LC-MS/MS were lysed to extract protein using the method described in Section 2.3.2. Quantification of protein concentrations of cell lysates was carried out with the Quick Start Bradford Protein Assay Kit (Bio-Rad, 500-0201) using the method described in Section 2.3.2. Following this, a buffer exchange step was performed in order to remove CHAPS detergent from the samples. 100 μg of protein from each sample (determined by Bradford assay) was placed into a clean Eppendorf and the protein was precipitated using a 2-D-clean-up kit (Bio-Rad, 163-2130), overnight in acetone at -20°C. Precipitated proteins were centrifuged for 15 minutes at 13,680 g at 4°C. The supernatant was discarded and the protein pellet was resuspended in 6M Urea, 2M Thiourea, 10 mM Tris, pH 8. 0.4% ProteaseMAX (Promega, V2071), a surfactant and Trypsin enhancer was added to improve protein solubility. The sample was sonicated and vortexed to ensure complete resuspension of precipitated proteins.

2.7.2. Quantification of Protein for Digestion

Quantification of buffer exchanged protein concentrations for each sample were determined with the Quick Start Bradford Protein Assay Kit (Bio-Rad, 500-0201) using a modified version of the method described in Section 2.3.2, and was performed prior to digestion and subsequent quantitative LC-MS/MS analysis. Accurate protein quantification at this stage is the foundation of a good label-free quantitative analysis to ensure protein concentrations can be precisely equalised across the entire sample set. 5 μL of standards and samples were added to the 96-well plate, with five replicates for each condition to be assayed. 250 μL of thiourea-compatible Bradford protein assay reagent (Bio-Rad, 500-0205) was then added each protein containing well in the plate. After 5 minutes incubation, the absorbance was read at 595nm. Sample replicates were required to have a coefficient of variation (CV) of < 1.5%.

2.7.3. In-Solution Digest of Proteins for LC-MS/MS Analysis

Preparation of individual protein samples for label-free LC-MS/MS analysis was carried out in accordance with the ProteaseMax Surfactant Trypsin enhancer in-solution digestion protocol (Promega). For label-free LC-MS/MS analysis, 10 µg of each sample was prepared and volumes were normalised to 50 µL using 50 mM ammonium bicarbonate (Sigma-Aldrich, 09830). All samples were reduced for 20 minutes with a final concentration of 5 mM Dithiothreitol in 50 mM ammonium bicarbonate at 56^{o}C followed by alkylation for 15 minutes in the dark at room temperature using a final concentration of 15 mM Iodoacetamide in 50 mM ammonium bicarbonate. A volume of 0.1% (v/v) ProteaseMAX (Promega, V2071) was then added to each fraction. Finally, protein digestion was carried out using sequencing-grade Trypsin Gold (Promega, V5280) at a ratio of 1:20 (protease:protein) overnight at 37^{o}C. Components and volumes required for this procedure are shown in Table 2.7. The enzymatic digestion was stopped by adding Trifluoroacetic acid (TFA) (Sigma-Aldrich, T6508) to a final volume of 0.1% (v/v).

Component	**Volume (µL)**
10 µg of protein for digestion	As required
50 mM ammonium bicarbonate	Up to 93.5
0.5 M DTT	1
0.55 M Iodoacetamide	2.7
Trypsin (1µg/µL)	1.8
1% ProteaseMAX Surfactant	1
Final Volume	**100**

Table 2.7: Components for in-solution digest prior to label-free LC-MS/MS analysis

2.7.4. C-18 spin column peptide concentration

In order to concentrate the peptides in each sample after the digestion process (Section 2.7.3) samples were purified using Pierce C-18 spin columns (Thermo Scientific, #89870). Buffers were prepared as shown in Table 2.8.

Buffer	Components
Activation Solution	50% Methanol
Equilibration Solution	0.5% TFA in 5% ACN
Sample Buffer	2% TFA in 20% ACN
Wash Solution	0.5% TFA in 5% ACN
Elution Buffer	70% ACN

Table 2.8: Buffers required for C-18 peptide cleanup.

All centrifugation steps were carried out at 1500 g for 1 minute. The spin column was prepared by gentle tapping until the resin was collected at the bottom after which it was placed into a collection tube. 200 µL of activation solution was used to rinse the column walls and to wet the resin, followed by centrifugation. Flow through was discarded and this step was repeated. Next, 200 µL of equilibration buffer was added and the column was centrifuged, flow-through discarded and the step repeated. Peptide samples were loaded into the resin bed of individual columns. The column was placed into a receiver tube and centrifuged. The flow-through was added to the resin bed once more to achieve maximum binding before re-centrifugation. The column was placed into a receiver tube and 200 µL of wash solution was added and centrifuged. The flow-through was discarded and the wash step was repeated. The column was placed into a new receiver tube. 20 µL of elution buffer was placed on top of the resin and centrifuged at 1500 g for 1 minute. This step was repeated with an additional 20 µL of elution buffer to yield a final volume of 40 µL of purified peptides. Samples were dried using a vacuum centrifuge (Maxi Dry Plus vacuum, MSC) until required. The resulting lyophilised peptides were re-suspended in 0.1% TFA in 2% acetonitrile with agitation and sonication prior to LC-MS/MS analysis.

2.7.5. LC-MS/MS Analysis

An Ultimate 3000 NanoLC system (Thermo Fisher Scientific) coupled to an LTQ Orbitrap XL mass spectrometer from Thermo Fisher Scientific (Dublin, Ireland) was used for the nano LC-MS/MS analysis. Digested peptide samples (1 μg) were loaded onto a C18 trap column (C18 PepMap, 300 μm id × 5 mm, 5 μm particle size, 100 Å pore size; Thermo Fisher Scientific). Peptide trapping was carried out at a flow rate of 25 μL/min in 0.1% TFA and 2% Acetonitrile (ACN) (Sigma-Aldrich, 34967) for 5 minutes. The trap column was switched on-line with an analytical PepMap C18 column (75 μm id × 500 mm, 3 μm particle, and 100 Å pore size; Thermo Fisher Scientific). Peptides were eluted using binary gradients of two solvent solutions A and B: Solvent A (2% ACN and 0.1% formic acid (Sigma-Aldrich, 94318) in LC-MS grade water (Sigma-Aldrich, product no. 39253); Solvent B (80% ACN and 0.08% formic acid in LC-MS grade water). 0%–25% of Solvent A was used for the first 240 min and 25%–50% of Solvent B for the remaining 60 min, resulting in a 5 hour total elution gradient. The column flow rate was set to 350 nL/min. Data was acquired with Xcalibur software, version 2.0.7 (Thermo Fisher Scientific).

Quantitative Label-Free MS: Separates the peptides using a 300 min linear gradient elution using Solvent A (0.1% formic acid) and Solvent B (0.08% formic acid, 80% acetonitrile) starting with 2% Solvent B to 25% B over 240 min and 25%-50% B for 60 minutes at a flow rate of 350 nL/min.

Qualitative MS Separates the peptides using a 60 min linear gradient elution using Solvent A (0.1% formic acid) and Solvent B (0.08% formic acid, 80% acetonitrile) starting with 2% Solvent B to 90% B over 60 min at a flow rate of 350 nL/min.

The MS was operated in data-dependent mode and externally calibrated. Survey MS scans were acquired in the Orbitrap in the 400–1800 m/z range with the resolution set to a value of 30,000 at m/z 400 and lock mass set to 445.120025 u. Collision induced dissociation (CID) fragmentation was carried out in the linear ion trap with up to three of the most intense ions per scan. Within 40s, a dynamic exclusion

window was applied. A normalised collision energy of 32%, an isolation window of 2 m/z, and one microscan were used to collect suitable tandem mass spectra.

2.7.6. Quantitative Profiling using Label-Free LC-MS/MS analysis

The following two sections describe the analysis of LC-MS/MS data to generate quantitative protein identifications based on the experimental design chosen in the software.

2.7.6.1. Detection of Differentially Expressed Peptides

Progenesis QI for Proteomics label-free LC-MS software version 3.1 (Nonlinear Dynamics, a Waters company) was used to process the MS data generated from the LC-MS/MS analysis. Data alignment was based on the LC retention time of each LC-MS sample, allowing for any drift in retention time given and adjusted retention time for all runs in the analysis. The sample run that yielded most features (i.e., peptide ions) was chosen to be the reference run. The retention times of all other runs were aligned to this reference run and peak intensities were normalised against it. Once alignments were complete, the following parameters were applied to filter out features which did not meet with the following criteria: (i) peptide features with ANOVA $p < 0.05$ between experimental groups, (ii) mass peaks (features) with charge states of +1, +2 and +3, and (iii) greater than one isotope per peptide. Comparative experimental designs were set up using the software, and differentially expressed peptides were subjected to the protein identification workflow.

2.7.6.2. Protein Identification

A MASCOT generic file was generated from all exported MS/MS spectra from Progenesis QI for Proteomics software after the comparative analysis was performed. The MASCOT generic file was used for peptide identification by importing the MS/MS data into Proteome Discoverer 1.4 (Thermo Fisher Scientific). The data was searched against a human FASTA database, downloaded from UniProKB/SwissProt in January 2015. Using both MASCOT and Sequest (HT) search algorithms, the following search

parameters were set for protein identification: (i) MS/MS mass tolerance set at 0.6 Da; (ii) peptide mass tolerance set to 20 ppm; (iii) up to two missed cleavages were allowed; (iv) cysteine carbamidomethylation set as a fixed modification; (v) methionine oxidation set as a variable modification; (vi) MASCOT peptide ion score set to a minimum of 40; (vii) Sequest HT XCorr was set at 1.9 for +1 ions, 2.2 for +2 ions, and 3.75 for +3 ions. Peptide identifications generated from Proteome Discoverer 1.4 were exported as a PepXML file and imported back into Progenesis QI software.

Once the peptide identification information was re-imported into Progenesis QI software, newly identified proteins were reviewed and only those which passed the following criteria were considered confidently identified and statistically significant: (i) an ANOVA p-value of <0.05 between experimental groups; (ii) proteins with ≥ 2 peptides contributing to the identification; (iii) a fold change between experimental groups of ≥ 2. Details on the experimental designs applied in Progenesis QI software, as well as the processing of the differentially expressed protein lists are described in Chapter 3.

2.8. Immunoprecipitation

2.8.1. Direct Immunoprecipitation

This method is also known as traditional immunoprecipitation. Proteins were isolated from cell lines using a gentle lysis buffer (Thermo Fisher Scientific, #87787) to preserve protein complexes using the method described in Section 2.3.1. Protein extracts were quantified using the Pierce BCA Protein assay Kit (Cat #23227) which is compatible with the gentle lysis buffer. Samples were diluted to a concentration of 1 mg/mL in a final volume of 1 mL using the gentle lysis buffer. Protein-G agarose beads (Sigma, P3296) were washed x3 in PBS before being added to the 1 mL test aliquot of protein extract for pre-clearing. 60 µL of Protein-G beads were added to the sample using a cut P-1000 pipette tip to minimise damage to the beads. All samples were incubated for 4 hours at 4°C on a rocking platform for pre-clearing. Beads were removed by spinning at 1000 g for 1 minute at 4°C. Supernatants were removed to clean Eppendorf tubes and divided evenly to represent the test sample and the negative

control. The primary antibody of interest was added to the test sample, and IgG matching the host species of the primary antibody was added to the negative control. Antibody/lysate mixtures were incubated at 4°C overnight on a rocking platform.

The following day, in order to precipitate the antibody-antigen complex, newly prepared Protein- G agarose beads were added to the samples as before and incubated at 4°C for 4 hours on a rocking platform. Beads were removed by spinning at 1000 g for 1 minute at 4°C, and the supernatant was saved as the 'unbound fraction'. The beads were then washed for 5 x 2-minute periods with gentle lysis buffer and pelleted at 1000 g between each wash for 1 minute. Following the final wash, the sample/bead solution was transferred to a spin column (Sigma, SC1000) before another centrifugation step as before. 70 μL of 2X Laemelli sample buffer was added to both sets of Protein-G beads before being heated to 95°C on a heating block. Samples were spun at 12000 g for 2 minutes and eluted into fresh Eppendorf tubes. Elutions were made to a 1x concentration with PBS and were stored at -20°C until required.

2.8.2. Cross-Linked Immunoprecipitation

Cross Linked IP were carried out using the Pierce Co-Immunoprecipitation Kit (Thermo Fisher Scientific, #26149). 200 μL of AminoLink resin was added to a spin column, and spun down at 1000 g for 1 minute (all subsequent centrifugation steps were at this speed and duration). The resin was washed twice with 0.4 mL of Coupling Buffer, centrifuged and excess solution discarded. The primary antibody of interest was added to the resin, along with 200 mL of Coupling Buffer (same for negative control IgG). 3 μL of Reducing Agent (Sodium Cyanoborohydride) was added to the mixture antibody/resin, which was then incubated at RT for 90 minutes. This step is to facilitate cross-linking of the antibody to the resin. The resin was then centrifuged and washed 2 times with 0.2 mL Coupling Buffer. Following this, 0.2 mL of Quenching Buffer was added to the suspension, centrifuged, and excess solution discarded. 0.4 mL of Quenching Buffer was then added to the gel/antibody complex, along with 3 μL of Reducing Agent. This suspension was incubated with end-over-end mixing, for 30 minutes at room temperature, followed by centrifugation and discarding of any excess

solution. The gel was then washed 2 times with 0.2 mL Quenching Buffer, followed by 6 washes in 0.4 mL Wash Solution, with centrifugation following each wash.

The protein sample was prepared as described in Section 2.8.1, and added to the resin/antibody complex. This solution was incubated overnight at 4°C on a rocking platform. The following day, the sample was centrifuged, and excess solution discarded. The resin/antibody-antigen complex was then washed 3 times in 0.4 mL Immunoprecipitation Buffer, with centrifugation after each wash. The antigen/antibody complex was then eluted with the addition of 200 μL of a low pH Elution Buffer to the solution, followed by 5-minute incubation at room temperature. The low pH of the buffer was neutralised with the addition of 5 μL of a Tris-HCl, pH 9.5, solution. Elutions were made to a 1x concentration with PBS and were stored at -20°C until required.

Both types of immunoprecipitation using direct and cross-linked methods are graphically displayed in Figure 2.1.

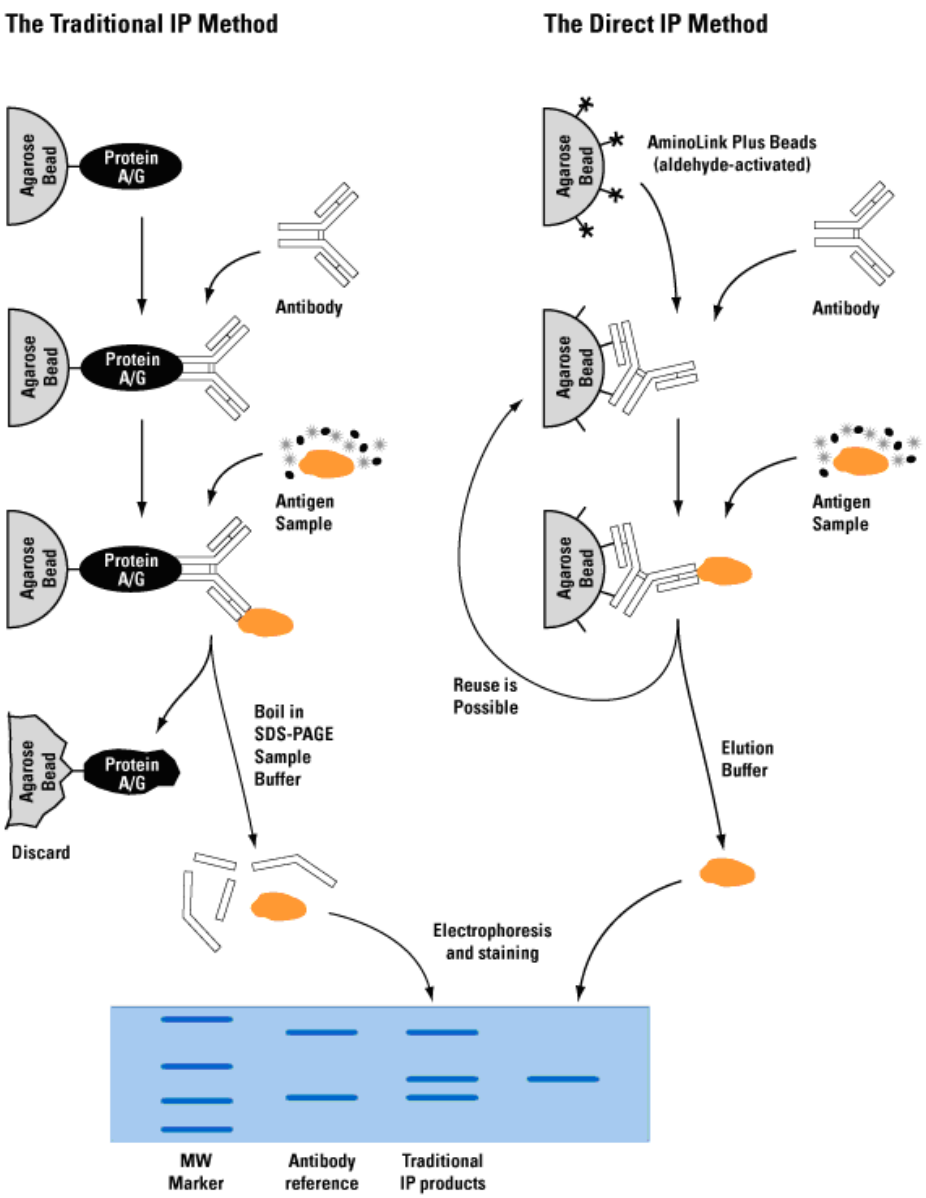


Figure 2.1: Traditional IP vs. Cross linked IP diagram. The advantage of cross linked IP is the antibody is not eluted with the target protein. Elimination of interfering heavy and light antibody chains and results in a clearer SDS gel for Western Blot analysis, or identification of proteins (www.piercenet.com)

2.8.3. 1-D Gel Electrophoresis of Co-Immunoprecipitated Proteins

1-D gel electrophoresis was carried out as described in Section 2.3.3 using 4-12% NuPAGE Bis-Tris Gels (Life Technologies, NP0321BOX) in XCell SureLock™ Mini-Cell (Life technologies, EI0001) running instrument). Following gel electrophoresis, the gel was either stained using Brilliant-blue G Colloidal Coomassie, or Western blot analysis was carried out as outlined in Section 2.3.4.

2.8.4. Staining of Gels Using Brilliant Blue G- Colloidal Concentrate

After electrophoresis, the gels were placed in a square petri dish (Sigma, Z617679) containing fixing solution (7% glacial acetic acid in 40% (v/v) methanol (ROMIL, HL109)) for 30 minutes. During this step a 1x working solution of Brilliant Blue G Colloidal Coomassie (Sigma, B2025) was prepared by combining 4 parts UHP to 1-part stain stock. Just prior to staining, a solution containing 4 parts of 1x working stock of Coomassie stain solution and 1-part methanol was made and thoroughly mixed. The gel was rinsed once with UHP water, and the Coomassie stain was poured over the gel. The gels were left to stain for 2 hours or overnight. To destain the background of the gel, UHP water was used with gentle rocking for 4 hours, or overnight.

2.8.5. Excision of Protein Bands from 1-D Gels

The stained gel was scanned and an image printed out for use as a map. The gel was rinsed 3x in LC-MS grade water and placed on a clean plastic plate which had been swabbed with 70% methanol. The work surface had also been cleaned with 70% methanol and great care was taken to reduce the probability of Keratin contamination from skin and hair. Each lane of the gel was sliced into thin sections, across the width of the lane using a scalpel, and cut into three smaller pieces. The approximate location of each gel slice was recorded on the image of the gel and an identifier was assigned to them. The pieces from each band were placed into individual wells of a polypropylene 96-well plate (Greiner Bio One, 651201) with 50 μL of LC-MS grade water in each, and these locations were recorded on the gel map also. The plate was then stored at 4 °C covered in parafilm, prior to protein destaining and subsequent extraction and digestion.

2.8.6. In-Gel Digestion of Proteins

2.8.6.1. Destaining of Gel Pieces

Gel pieces were washed three times in a 96-well polypropylene plate using 100 mM ammonium bicarbonate/acetonitrile (1:1, vol/vol) for 10 minutes each until the pieces were fully destained.

2.8.6.2. In-Gel Digestion

After destaining gel pieces, 50 μL of neat acetonitrile was then added to each well at room temperature until the pieces became dehydrated. The acetonitrile was removed and the gel pieces were dried. 50 μL of 10 mM DTT in 100 mM ammonium bicarbonate was added to cover the gel pieces and allowed to incubate for 30 minutes at 56 °C. The DTT solution was removed and the gel pieces were cooled to room temperature. 50 μL of 55 mM Iodoacetamide in 100 mM ammonium bicarbonate was then added. The gel pieces were incubated for 20 minutes in the dark at room temperature. After reduction and alkylation, 50 μL of neat acetonitrile was added to each gel piece at room temperature to dehydrate them. After drying, the individual gel pieces were rehydrated in 10 μL of digestion buffer containing Trypsin (Promega, V5111). The digestion buffer was made fresh and consisted of 12.5 ng sequence-grade Trypsin/μL in 10% acetonitrile, 40 mM ammonium bicarbonate. Sufficient buffer was added to cover the dried gel pieces and the plate was incubated at 37 °C for 30 minutes. After this time, the pieces were checked to ensure that the buffer was absorbed. An additional 10-20 μL of ammonium bicarbonate buffer was added to cover the gel pieces and maintain moisture for the digestion. Digestion was carried out overnight at 37 °C. Peptides were extracted from the gel pieces twice using 50 μL of 1:2 (vol/vol) 5% TFA in acetonitrile (made fresh) for 15 minutes each at 37 °C with gentle agitation. The extracted peptides were transferred to a new 96-well plate and lyophilised using a Maxi Dry Plus vacuum (MSC). The plate was then stored at -20 °C, wrapped in parafilm, or the proteins were resuspended in 20 μL of 0.1% TFA containing 0.1% acetonitrile and analysed using LC-MS/MS.

LC-MS/MS analysis and subsequent identification of proteins was performed as described in Section 2.7.5.

2.9. Gene Expression Assay

2.9.1. RNA Extraction

TRIzol Reagent (Thermo Fisher Scientific, 15596-026) was used for RNA extraction from cells in monolayer. 1 mL of TRIzol Reagent was added to each T75 flask for RNA isolation. These samples were allowed to stand for 5 minutes on ice to allow complete lysis of the cells. The TRIzol/lysed cells solution was transferred to a clean Eppendorf tube, and 200 µL of chloroform (Sigma, C2432) was added. Samples were vortexed at high speed for 15 seconds and allowed to stand for 15 minutes at room temperature. The resulting mixtures were centrifuged at 13,684 g in a microcentrifuge for 15 minutes at 4ºC. The aqueous layer containing RNA was carefully removed to a clean fresh Eppendorf tube. To this 500 µL of ice-cold Isopropanol (Sigma, I9516) was added. The samples were mixed and incubated at room temperature for 10 minutes, and then centrifuged at 13,684 g for 30 minutes at 4ºC to pellet the precipitated RNA. Taking care not to disturb RNA pellet, the supernatant was removed and the pellet was subsequently washed by the addition of 750 µL of 75% of ethanol and vortexed. The samples were centrifuged at 7,500 g for 5 minutes at 4ºC. The supernatant was removed and the wash step was repeated. The RNA pellet was allowed to air-dry for 10 minutes and then re-suspended in 50 µL of RNase free water.

2.9.2. Determination of DNA/RNA Quantity and Quality

Purified RNA samples were quantified using the Nanodrop® ND-1000 spectrophotometer (NanoDrop Technologies). Before applying the RNA/DNA sample, the pedestal was wiped down using a lint-free tissue dampened with UHP. The programme 'RNA-40' was selected on the NanoDrop software main screen to read the samples at 260nm, and the instrument was blanked with 2 µL of UHP. The concentration of RNA/DNA was calculated by software using the following formula:

OD260 nm x Dilution factor x 40 = ng/µL RNA
Samples were checked for quality (i.e. Phenol contamination) by assessing the A260/A280 and A260/A230 ratio values.

An A260/A280 ratio ~2, and an A260/A230 of ~1.8-2.2 is indicative of a pure RNA.

2.9.3. High Capacity cDNA Reverse Transcription

Reverse transcription of RNA to cDNA was carried out using the High Capacity cDNA reverse transcription kit (Applied Biosystems, 4368814) in accordance with the manufacturer's specifications. A maximum of 2 µg of RNA was made up to 10 µL in a 0.5 mL PCR tube using nuclease-free water. In a separate 0.5 mL PCR tube, the reverse transcription master mix was prepared. 10 µL of the RT master mix was added to the RNA sample and mixed using a pipette. The temperature profile for the reverse transcription reaction carried out using a bench top thermal cycler G-Storm GS1 PCR machine using the conditions shown in Table 2.9.

	Step1	Step2	Step3	Step 4
Temperature (°C)	25	37	85	4
Time (min)	10	120	5	∞

Table 2.9: Cycling conditions for reverse transcription.

2.9.4. Real-Time PCR Gene Expression Assay

RNA was extracted from cells as described in Section 2.9.1, and cDNA was synthesised from the RNA as described in Section 2.9.3. The PCR Master Mix used was TaqMan Fast Advanced Master Mix 2x (Thermo, 4444557). TaqMan assays used were Applied Biosystems TaqMan Gene Expression Assays (Thermo Fisher Scientific). The Real-Time PCR analysis was performed using a 7500 Fast Real-Time PCR system (Applied Biosystems). PCR reaction plates used were MicroAmp fast optical 96-well reaction plates (Applied Biosystems).

Master Mix for each TaqMan assay was set up in PCR tubes, using the volumes shown in Table 2.10. Volumes shown are for 1 PCR reaction and were scaled up for multiple samples, with an additional 10% to allow for pipetting loss. A map was drawn up of the reaction plate and wells were designated for certain samples. cDNA was then pipetted in these wells as per Table 2.10 and control samples were added also. Master Mix was then added to the appropriate wells and the plate was covered and sealed with optical adhesive film. Each PCR was performed in triplicate and an

endogenous control was included. Comparison of the Ct value of a target gene with that of the endogenous control gene allows the gene expression level of the target gene to be normalized to the amount of input cDNA. As an additional control, water on its own was amplified as a negative control to rule out presence of any contaminating RNA or DNA. Plates were run on the 7500 Fast Real-Time PCR system using the conditions shown in Table 2.11. Details of TaqMan assays used in this study are shown in Table 2.12.

Data were analysed using a Relative Quantification ddCt study. A calibrator sample was selected and set to a value of one, allowing for the comparison of all other samples in relation to the calibrator. For the analysis of target gene and endogenous control amplification, the baseline was set to average, normalised fluorescent signal before detectable increase (usually 3-15 cycles) and the cycle threshold was set in the exponential part of the curve. The Ct standard error was ideally less than +/-0.161 between replicate wells, and Ct errors with values greater than this were removed as outliers. The endogenous control was used automatically to normalise the data. When this was achieved for both the target and endogenous control, the relative quantity values for the run were generated and plotted relative to the calibrator sample.

Component	Volume µL/well
Master Mix	10
TaqMan Assay	1
cDNA Template	2
Nuclease free water	7
Total volume	20

Table 2.10: TaqMan assay components

Stage		Temperature (°C)	Time (Seconds)
Polymerase Activation		95	20
Denature	PCR (40 cycles)	95	3
Anneal/Extend		60	30

Table 2.11: Thermal cycling conditions for TaqMan assays

Gene	Catalogue Number
Shootin-1 (Kiaa1598)	hs01008867_m1
MARCKS	hs00158993_m1
Desmoglein-3	hs00951897_m1
Beta-Actin	hs1060665_g1

Table 2.12: TaqMan assays used in this study.

2.10. Overexpression of Targets

2.10.1. Polymerase Chain Reaction

The DreamTaqTM PCR Master Mix kit (Fermentas Life Sciences, #K1071) was the kit used to perform all PCR amplification reactions in accordance with the manufacturer's specifications. DreamtaqTM PCR Master Mix was vortexed after thawing and placed on ice until further required. Each PCR reaction was made up to a volume of 50 µL using the components Table 2.13. A 25 µL reaction volume was made up in a thin walled PCR tube using a maximum concentration range of 1-100 ng of cDNA and 10 pg of forward and reverse PCR primer. Once all components have been added, 25 µL of the 2X DreamTaqTM PCR Master Mix was added to each sample creating a 50 µL reaction volume.

Component	Volume/Concentration
DreamTaq 2x Master Mix	25 µL
Forward Primer	0.1 – 1.0 µM
Reverse Primer	0.1 – 1.0 µM
Template cDNA	1 – 100 ng
Nuclease Free Water	To 50 µL
Total Volume	**50 µL**

Table 2.13: DreamTaq Components

2.10.2. Gel Electrophoresis

Visual analysis of PCR products can be carried out using gel electrophoresis. cDNA samples were diluted in nuclease free water and mixed with a 10x loading buffer. A 0.8% agarose gel in 50 mL 1 x TAE was made with 5 µL ethidium bromide (EtBr) (Sigma, E1510) added for staining the cDNA. A DNA ladder was added to one well of the gel to estimate the location of the PCR products. The gel was run for a sufficient time to ensure clear separation of the cDNA bands.

2.10.3. PCR Purification

All PCR product cDNAs were purified using the QIAGEN QIAquick PCR Purification kit (QIAGEN, 28104). Into the PCR mix, 5 volumes of Buffer PB was added and mixed by pipetting. A QIAquick spin column was placed into a 2 mL

collection tube and the sample was added and centrifuged at 16,060 g for 1 minute. The flow through was discarded and a wash step was performed by adding 750 µL of Buffer PE to the QIAquick column. After centrifugation at 16,060 g for 1 minute, flow through was discarded and the QIAquick column was placed back into the collection tube and spun for a further 1 minute at 16,060 g to dry the membrane. Before elution of the purified DNA, the QIAquick spin column was placed into a clean microcentrifuge tube. DNA was eluted by adding 30-50 µL of Buffer EB or water directly to the QIAquick membrane and incubating for 1 minute at room temperature followed by spinning at 16,060 g for 1 minute. Quantification of DNA was determined using a NanoDrop spectrophotometer and measuring the OD260nm. Purified cDNA was stored at -20˚C until required.

2.10.4. Restriction Enzyme Digestion

Digestion reactions were set up in 0.5 mL PCR tubes at a reaction volume of 20 µL. The reaction volume can vary depending on the amount of DNA required. Components required are shown in Table 2.14 for the digestion of 2 µg of DNA with two restriction enzymes in a volume of 20 µL. Enzymatic digestions are carried out at a temperature of 37 ˚C on a hot plate until complete digestion is achieved.

Component	Required for 20 µL
cDNA	2 µg
NEB Buffer (1-4) 10x	2 µL
Restriction Enzymes	1U (0.5-1 µL)
Nuclease Free Water	To 20 µL

Table 2.14: Restriction enzyme digestion components.

2.10.5. Ligation

The T4 DNA ligase kit from ROCHE (11635379001) was used for all ligation reactions. No more than a total of 1 µg of DNA is to be used in any given ligation reaction and a range of 1-5 Units of T4 DNA ligase enzyme. A range of ~80-150 ng of PcDNA3.1+ (V79020, Thermo Fisher scientific) vector backbone was used in all

ligations. The concentration of insert used was determined based on the fragment size and the amount/size of vector being used (Ligation Calculator), and all reactions were carried out in 20 µL volumes. 2 µL of 10x T4 DNA ligase reaction buffer (660 mM Tris-HCl, 50 mM $MgCl_2$, 10 mM ATP, 50 mM EDTA, pH 7.5 at 25˚C) was added to the reaction mixture. Nuclease-free water was used to make up the reaction volume to 19 µL prior to the addition of the T4 DNA ligase enzyme. Finally 1 µL (1 U) of T4 DNA ligase enzyme was added to the reaction mixture creating a reaction volume of 20 µL. This reaction was incubated overnight on ice water. This allows the ligation reaction to commence over a temperature gradient. Ligated samples were stored at -20˚C. An identical reaction containing all components except the fragment to be inserted (no insert control) was set up in tandem to check for self-ligation of the vector. In addition, a no-enzyme control was also set up.

2.10.6. Transformation of Competent Cells

Competent cells used in this study were Subcloning Efficiency DH5αTM Competent cells (18265017, Thermo Fisher Scientific). A water bath was pre-heated to 42°C while 50 µL of competent cells were transferred into clean Eppendorf tubes and kept on ice. ~5 µL of ligation product was added to the competent cells and stirred gently with a pipette tip. This mixture was left on ice for 30 min. The competent cells/DNA were heat-shocked by placing in the water bath at 42°C for 45 seconds. Cells were then immediately placed on ice for a further 3 min. 500 µL of pre-heated (37 °C) Super Optimal broth with catabolite repression media (supplied in kit) was added to the transformed competent cells. Cells were incubated for 2 hours at 37°C on a shaker platform. Contents were transferred into an Eppendorf tube and centrifuged at 3000 g for 4 min. All but 50 µL of SOC media was decanted from the bacterial pellet. The pellet was resuspended using a pipette and transferred onto a LB agar plate containing Ampicillin (A9518, Sigma) at a concentration of 50 µg/ml for selection of successfully transformed cells. Using a sterile microbiological spreader, the transformed competent cells were homogenously spread over the LB agar plate (Sigma, L2897). The plates were then sealed with parafilm and incubated at 37 ˚C for 16-24 hours. Controls in this process were: untransformed DH5α cells; S.O.C media alone; ligation mix with no ligation enzyme (Section 2.10.5). Healthy growing colonies after 24 hours were deemed

potentially successful transformants. These were inoculated into 5 mL of fresh LB media containing 50 μg/ml Ampicillin for selection.

2.10.7. DNA mini-prep of plasmid DNA

The QIAGEN QIAprep Spin Miniprep kit (QIAGEN, 27104) was used according to the manufacturer's instructions for the isolation of plasmid DNA from a 5mL culture. Cells were pelleted by centrifugation at 4000 g for 10 min. Cell pellets were resuspended in 250 μL of buffer P1 and transferred to a microcentrifuge tube. 250 μL of buffer P2 was added to this solution and mixed gently by inversion. A 350 μL volume of buffer N3 was subsequently added to this solution and again mixed gently by inversion. This mixture was centrifuged for 10 min at 13,000 g. The decanted supernatant was applied to a QIAprep spin column and centrifuged for 60 seconds at 13,000 g, and flow though was discarded. The column was washed by adding 500 μL of Buffer PB and centrifuged for 60 seconds at 13,000 g. A second wash was performed using 750 μL of Buffer PE and centrifuged for 60 seconds. The column was dried by an additional centrifugation step for 60 seconds to remove any remaining wash buffer. The spin column was then placed into a clean microcentrifuge tube. The DNA was eluted by adding 30-50 μL (depending on the desired DNA yield) of Buffer EB or water to the spin column. This was allowed to stand for 1 minute at room temperature (RT) before centrifugation at 16,060 g for 1 minute. The eluted DNA was reapplied to the spin column and centrifuged for an additional 60 seconds to collect any remaining plasmid DNA. Quantification of the eluted plasmid DNA was determined using a NanoDrop spectrophotometer by measuring at OD260nm. Plasmid was stored at -20˚C until required.

Chapter 3. Quantitative Proteomic Profiling of DLKP by Label-Free LC-MS/MS

3.1. Differential Protein Expression Analysis of DLKP and its Subpopulations Using Quantitative Mass Spectrometry Based Proteomics

The DLKP cell line was established from a lymph node metastasis of a primary lung tumour and is described as a poorly differentiated squamous lung carcinoma cell line. This cell line was previously found to be comprised of three distinct subpopulations: DLKP-SQ, DLKP-M and DLKP-I, each with well-defined phenotypes (McBride et al. 1998a). Stark phenotypic differences between the subpopulations include the ability of the cells to migrate and invade *in vitro* (Joyce H, PhD, 2015), and the ability of cells to survive unattached to a surface or other cells (Keenan et al. 2012). These characteristics make DLKP a particularly useful model for studying cellular heterogeneity in poorly differentiated non-small cell carcinoma.

Liquid chromatography coupled to mass spectrometry (LC-MS) is one of the most popular and comprehensive profiling techniques used to analyse proteins from complex biological material (Christin, Bischoff and Horvatovich 2011). Labelling technologies such as SILAC, iTRAQ, ICAT and 2D-DIGE have been the main technologies used in MS-based protein quantification for profiling experiments in proteomics. While these techniques have been successful for certain experimental setups, they also have inherent limitations. These can include: increased time and complexity of sample preparation; high cost of reagents; incomplete sample labelling which leads to confused data interpretation; sample number limitation: up to 10 samples maximum for TMT isobaric labelling (Murphy et al. 2014). More recently, label-free LC-MS/MS has become an attractive method to proteomic investigators for the quantitative analysis of proteins. This technique is becoming more widely employed as it overcomes many of the issues associated with labelled techniques. It is generally accepted that label-free quantitation estimates protein abundance more accurately than gel based methods, and has a higher sensitivity and larger dynamic range than labelling techniques (Neilson et al. 2011). Conversely, this technique requires highly reproducible HPLC to perform well, and powerful bioinformatics tools for data processing. With a high end nano-LC system coupled to a hybrid linear ion trap/Orbitrap mass spectrometer in-house, label-free quantification was the method of choice for the proteomic analysis of the DLKP cell line and its subpopulations to understand cellular heterogeneity in lung cancer.

3.1.1. Preparation of DLKP Cell Lines

To identify proteins which may be used to characterise the DLKP cell lines, a comparative analysis was carried out using quantitative label-free mass spectrometry. The experimental design called for DLKP and clonal variants (DLKP-SQ, DLKP-M, DLKP-I) to be grown in biological triplicate with samples being taken at both exponential and stationary phases of growth. Both phases of growth were analysed separately in order to find highly differentially expressed proteins which were not growth phase dependent. Running two separate experiments would lead to strong protein identifications which could be used to characterise the DLKP clonal subpopulations. Previous optimisation had shown that the DLKP cell lines reach exponential phase at 72 hours and stationary phase 120 hours after seeding the cells at $2.5x10^4$cells/ml in a T75 flask, so these timepoints were used for both phases of growth. Extensive optimisation was carried out to determine the phases of growth for each cell line used in this study with both 72 and 120 hours representing exponential and stationary phases of growth respectively. Samples of each cell line were taken down within 6 passages of being thawed to reduce the chances of genetic drift due to prolonged time in culture.

Each cell line was grown in T75 cell culture flasks and seeded at a density of $2.5x10^4$cells/ml. Two flasks were designated for exponential phase and the other two for stationary phase. For exponential phase samples, one T75 flask was lysed using the method described in Section 2.3.1 after 72 hours and stored at -80°C. The remaining T75 flask was reseeded into two new flasks at a cell concentration of $2.5x10^4$cells/ml once again. This process was repeated twice more until all three biological replicates were harvested. The same procedure was carried out for stationary phase samples, the only difference being that each sample harvest and reseeding step took place at 120 hours.

In parallel, combination cell lines were also set up and grown until exponential and stationary phase. These combination cell lines were set up in order to monitor how protein expression was affected when the DLKP clones were grown in co-culture. The same cell culture method as described above was carried out for these samples, however they initially consisted of 50/50 mixtures of each clone by cell number ($1.25x10^4$cell/ml of A + $1.25x10^4$cell/ml of B). Alongside this, the DLKP cell line was reconstituted in

the proportions found previously (McBride et al. 1998a) with 70% DLKP-SQ, 25% DLKP-I and 5% DLKP-M and is referred to as “Remade”. A summary cell proportions used is shown in Table 3.1.

DLKP-SQ	DLKP-M	DLKP-I	Combination Cell Line
$1.25x10^4$cell/ml	$1.25x10^4$cell/ml	-	**SQ+M**
$1.25x10^4$cell/ml	-	$1.25x10^4$cell/ml	**SQ+I**
-	$1.25x10^4$cell/ml	$1.25x10^4$cell/ml	**I+M**
$1.75x10^4$cell/ml	$1.25x10^3$cell/ml	$6.25x10^3$cell/ml	**DLKP Remade**

Table 3.1: Proportions of the DLKP clones which were used to make up the combination cell lines.

3.1.2. Protein Preparation and Analysis by Quantitative Label-Free LC-MS/MS

Equal protein concentrations were prepared as outlined in Section 2.7.2. This step is essential and is the foundation of an accurate quantitative label-free study. For the protein assay step, high precision was required, with samples requiring a coefficient of variance no greater than 1.5%, and the standard curve is required to have an R^2 value of no less than 0.98. Following digestion of the protein samples as described in Section 2.7.3, each sample was analysed using quantitative label-free LC-MS/MS as described in Section 2.7.6.

All samples were analysed as one large group, directly after each other in order to reduce variability and retention time drift. To reduce potential bias incurred by running replicates from sample groups together as one block, samples were interleaved with those from other groups when being analysed using the label-free quantification method.

Data analysis was performed using Progenesis QI for Proteomics software. Raw MS files were imported into the software and aligned to each other based on a reference file chosen which was representative of the entire dataset. This file is chosen automatically by Progenesis software and is usually the file with the greatest number of peptide ions, which increases the chances of all other files to being strongly aligned to it. Once aligned, this matrix of sample files were analysed to detect features in each run.

Detected features were then filtered to only allow through those with an ANOVA p-value of <0.05. This p-value was chosen to only allow through statistically significantly differentially expressed peptide features. DLKP and its clonal subpopulations come from the same origin but are essentially different cell types, so there will be inherent protein abundance differences. A tight filtering criterion ensures that only the most significant differences come through the analysis. From the list of features generated after filtering, a PCA (principal component analysis) plot can then be generated in the software. Plotting the abundance data in this way allows the separation of samples according to abundance variation, and is a useful tool for identifying run outliers (See Figure 3.1 and Figure 3.2). <

The MS/MS data from this list of significant peptide features can then be exported from the software for protein identification. This is carried out by Proteome Discoverer 1.4 which searches the MS/MS information against the UniProt and SwissProt protein databases using MASCOT and Sequest algorithms (See Section 2.7.6.2). Once completed, the search results are then re-imported back into Progenesis QI software and identified proteins are assigned to peptide features.

A set of tight filtering criteria was applied to the data at this stage to select only the most statistically significant differentially expressed proteins. These filters were: ≥ 2 peptides contributing to protein identifications, a change of ≥ 2 fold between experimental groups for any identified protein, an ANOVA value of <0.05 for a protein between experimental groups. The list of resulting proteins depends on the initial comparative experimental design setup in the Progenesis software. The software allows for any permutation of the samples comparison and those chosen in this analysis are described below.

3.1.3. Experimental Designs

The LC-MS/MS data generated for the DLKP cell line and its clonal subpopulations resulted in a large dataset which could be arranged in any way for a comparative study using Progenesis software. Table 3.2 shows all the available samples for comparison. In the software, each biological triplicate for a given sample is grouped together as one to represent that cell line, and comparative studies are carried out on these.

Name	Exponential phase	Stationary phase
DLKP-SQ	3	3
DLKP-M	3	3
DLKP-I	3	3
DLKP	3	3
SQ+M	3	3
SQ+I	3	3
I+M	3	3
Remade	3	3

Table 3.2: Brief summary of the DLKP experimental design. 3 biological replicates for each cell line and combination cell lines were required, at both exponential and stationary phases of growth. This resulted in a total of 48 samples analysed by label-free LC-MS/MS.

It was decided to analyse the data using two different methods in order to extract the maximum fold change information, and complete the most thorough comparative analyses using Progenesis software. These two strategies were termed: The Global Analysis; and the Two-Sample Analysis.

The Global Analysis compares samples against each other as one large group. Using this method results in the availability of fold change data for a statistically significantly differentially expressed protein between the highest and lowest expressing cell lines. It also allows the expression profiles of a protein to be determined across multiple cell lines. However, it does not provide intermediary fold changes for a protein in cell lines which are not the highest and lowest expressers, nor does it provide statistical significance for these samples. For this reason, an additional experimental design was also used.

The Two-Sample Analysis compares just two cell lines against each other in a one vs. one comparison. All samples are imported into Progenesis software, allowing

the software the full wealth of data for the software to draw upon. However in the experimental design itself, only two cell lines are compared against each other directly. All replicates for both cell lines in the experimental design are included. This method provides individual fold change data between two cell lines, along with their statistical significance. This information is lost in the global analysis, so by using both methods, the data can result in protein expression patterns and all associated fold changes.

3.1.3.1. Global Comparison Experimental Design

The experimental designs for both exponential and stationary phases of growth were set up initially to compare the DLKP cell line and the clonal subpopulations, excluding the combination lines. These were compared initially using the global analysis method. Principal component analysis was carried out for both experiments in the Progenesis software and results are shown in Figure 3.1 and Figure 3.2. This tool performs statistical analysis on filtered data, and creates a graphical representation of the answers asked of the data by the experimental design. A number of details can be determined from this graphical representation: how tightly replicate samples cluster together, if there are outliers in groups, and how similar those groups are to each other. The PCA analysis was performed on statistically significant peptide features (ANOVA $p < 0.05$).

Principal component analysis of the exponential phase global analysis (Figure 3.1) shows that replicate samples cluster well into reasonably distinct groups based on the cell line. This indicates that sample preparation was uniform for each cell line, and growth in culture over three passages did not result in significant protein expression drift. The DLKP-SQ group clusters furthest away from the other cell lines, followed by DLKP-M. The principal component analysis for the stationary phase global analysis shows tighter clustering than during exponential phase. DLKP-M clusters furthest away compared to the other cell lines. In both phases of growth, good separation of the clones and DLKP cell line was found.

Interestingly, DLKP-SQ and DLKP cluster closely together during stationary phase of growth but not as much during exponential phase. This suggests that as the

cells become more tightly packed and reach stationary phase, DLKP-SQ and DLKP become more similar to each other. This highlights the plastic nature of these cell types and may support the interconversion theory proposed by (McBride et al. 1998a)

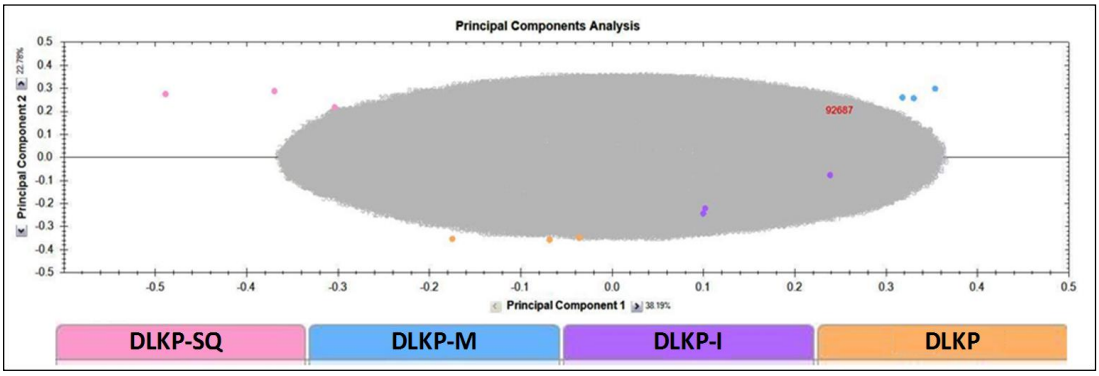


Figure 3.1: Principal component analysis for Exponential Phase global comparison of DLKP-SQ, DLKP-M, DLKP-I and DLKP cell lines only.

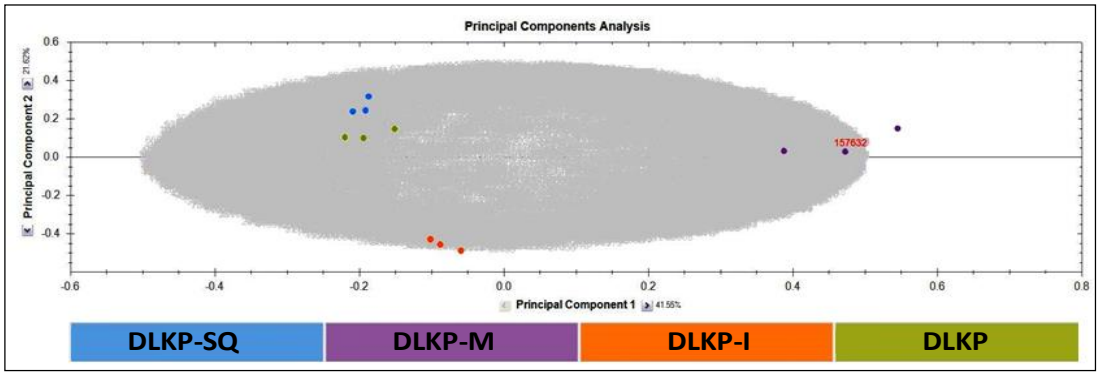


Figure 3.2: Principal component analysis for Stationary Phase global comparison of DLKP-SQ, DLKP-M, DLKP-I and DLKP cell lines only.

3.1.3.2. Differentially Expressed Proteins:

In this global analysis, the experimental design compared; DLKP-SQ; DLKP-M; DLKP-I; and DLKP all against each other at once. Analysis was performed on protein results which were filtered using the tight criteria described previously (≥ 2 peptides contributing to protein identifications, a change of ≥ 2 fold between comparisons for any protein, an ANOVA p-value of <0.05 for a protein between experimental groups). The exponential phase experiment highlighted 127 significantly differentially expressed proteins and the stationary phase highlighted 203 significantly differentially expressed proteins (Table 3.3). Overlapping these two lists revealed 63 common proteins which were found in both independent experiments. To get an overview of the results from these analyses, a number of differentially expressed proteins from exponential and stationary phases are displayed in Table 3.4 and Table 3.5. The top 10 proteins with the highest abundance in each cell line are shown. Full lists are in Section 8.1 (Appendices on Disc).

Total Number of Differentially Expressed Proteins

Comparison	DLKP-SQ vs. DLKP-M vs. DLKP-I vs. DLKP
Exponential	127 proteins
Stationary	203 proteins

Table 3.3: Displayed is the number of proteins which were shown to be statistically differentially expressed from a 'Global Comparison'. Proteins were required to have a fold change ≥2 between two samples, and ANOVA $p < 0.05$, and ≥2 peptides contributing to the identification.

Exponential vs. Stationary Phase

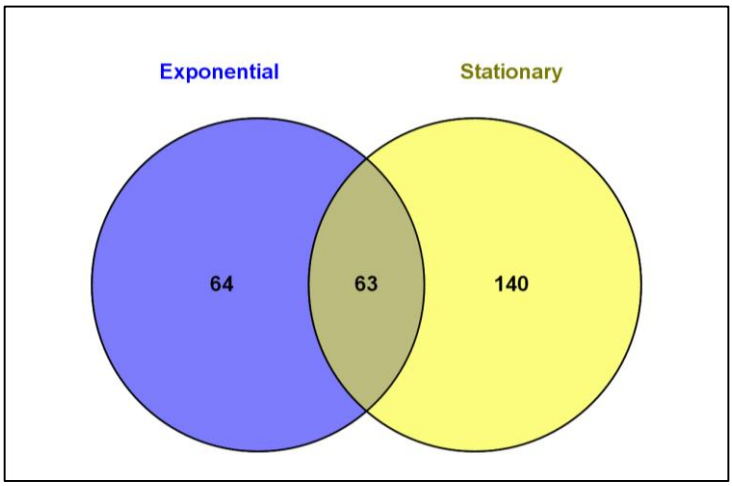


Figure 3.3: Venn diagram showing the overlapping proteins from the exponential and stationary phase experiments. 63 proteins were common to both global comparisons.

Exponential Phase: Top Fold Changes

Description	Max fold change	Highest
Serum albumin	88.0	DLKP SQ
Vitamin D-binding protein	40.7	DLKP SQ
UBX domain-containing protein 1	20.4	DLKP SQ
Lactotransferrin	17.8	DLKP SQ
Hemoglobin subunit alpha	13.5	DLKP SQ
Hyaluronan and proteoglycan link protein 1	9.6	DLKP SQ
D-3-phosphoglycerate dehydrogenase	5.8	DLKP SQ
Collagen alpha-1(XIV) chain	4.6	DLKP SQ
Leucine-rich repeat flightless-interacting protein 1	4.6	DLKP SQ
Shootin-1	**4.4**	**DLKP SQ**
Myristoylated alanine-rich C-kinase substrate	**51.7**	**DLKP M**
Collagen alpha-1(III) chain	12.5	DLKP M
Bifunctional ATP-dependent dihydroxyacetone kinase	9.3	DLKP M
PDZ and LIM domain protein 5	8.5	DLKP M
Dihydropyrimidinase-related protein 2	5.3	DLKP M
1,4-alpha-glucan-branching enzyme	4.8	DLKP M
3-hydroxyisobutyrate dehydrogenase, mitochondrial	3.7	DLKP M
Transferrin receptor protein 1	3.7	DLKP M
Cysteine and glycine-rich protein 2	3.5	DLKP M
Guanine nucleotide-binding protein G(i) subunit alpha-1	3.4	DLKP M
Desmoglein-3	**175.3**	**DLKP I**
Tissue factor pathway inhibitor 2	23.5	DLKP I
CD166 antigen	9.2	DLKP I
Protein S100-A10	8.0	DLKP I
Neurotensin/neuromedin N	7.5	DLKP I
Uridine phosphorylase 1	4.1	DLKP I
Glutaminase kidney isoform, mitochondrial	3.3	DLKP I
Neuroblast differentiation-associated protein AHNAK	3.0	DLKP I
Annexin A1	3.0	DLKP I
Signal transducer and activator of transcription 1-alpha/beta	2.8	DLKP I
Connective tissue growth factor	122.7	DLKP
Protein CYR61	22.6	DLKP
Transgelin	18.5	DLKP
Solute carrier family 2, facilitated glucose transporter member 3	15.3	DLKP
Hepatocyte growth factor	15.0	DLKP
Cathepsin Z	12.2	DLKP
F-box only protein 2	8.2	DLKP
Gamma-enolase	5.7	DLKP
Fascin	4.2	DLKP
Gap junction alpha-1 protein	3.4	DLKP

Table 3.4: List of proteins showing the greatest fold changes during exponential phase for each cell line. All proteins have ≥2 unique peptides contributing to their identification, and an ANOVA p-value <0.05, and a fold change ≥2 between any two cell lines. Full lists are shown in Appendix 8.1.

Stationary Phase: Top Fold Changes

Description	Max fold change	Highest Abundance
D-3-phosphoglycerate dehydrogenase	5.4	DLKP-SQ
Putative helicase MOV-10	4.2	DLKP-SQ
Fumarylacetoacetase	3.9	DLKP-SQ
Shootin-1	**3.5**	**DLKP-SQ**
Ataxin-10	3.4	DLKP-SQ
Serum albumin	3.2	DLKP-SQ
Heat shock protein beta-1	2.8	DLKP-SQ
Isoleucine--tRNA ligase, mitochondrial	2.7	DLKP-SQ
Aldose reductase	2.5	DLKP-SQ
E3 ubiquitin/ISG15 ligase TRIM25	2.5	DLKP-SQ
Myristoylated alanine-rich C-kinase substrate	**28.4**	**DLKP-M**
Collagen alpha-1(III) chain	20.2	DLKP-M
Histone H4	13.0	DLKP-M
Chromatin target of PRMT1 protein	7.5	DLKP-M
Cysteine and glycine-rich protein 2	6.8	DLKP-M
Insulin-like growth factor 2 mRNA-binding protein 3	6.4	DLKP-M
Protein NipSnap homolog 1	5.9	DLKP-M
Peroxiredoxin-2	5.7	DLKP-M
SH3 domain-binding glutamic acid-rich-like protein	5.5	DLKP-M
1,4-alpha-glucan-branching enzyme	5.3	DLKP-M
Tissue factor pathway inhibitor 2	14.9	DLKP-I
CD166 antigen	7.3	DLKP-I
Ras GTPase-activating-like protein IQGAP1	3.6	DLKP-I
Annexin A1	3.4	DLKP-I
Neuroblast differentiation-associated protein AHNAK	3.2	DLKP-I
Protein PRRC2C	3.2	DLKP-I
Neurotensin/neuromedin N	3.0	DLKP-I
U1 small nuclear ribonucleoprotein 70 kDa	2.9	DLKP-I
Glutaminase kidney isoform, mitochondrial	2.7	DLKP-I
Vasodilator-stimulated phosphoprotein	2.7	DLKP-I
Connective tissue growth factor	28.7	DLKP
Protein CYR61	20.0	DLKP
Transgelin	11.0	DLKP
Hepatocyte growth factor	6.7	DLKP
Heat shock 70 kDa protein 1A/1B	3.5	DLKP
Fascin	3.2	DLKP
B-cell receptor-associated protein 31	3.1	DLKP
Cathepsin Z	3.0	DLKP
Exportin-2	2.7	DLKP
Ubiquitin-like protein ISG15	2.5	DLKP

Table 3.5: List of proteins showing the greatest fold changes in the stationary phase for each cell line. All proteins have ≥2 unique peptides contributing to their identification, an ANOVA p-value <0.05, and a fold change ≥2 between any two cell lines. Full lists are shown in Appendix 8.1.

In Table 3.4 and Table 3.5, the top 10 proteins which had the greatest abundance in each of the cell lines is shown, resulting in a list of 40 proteins. These protein results are from the global comparisons which compared DLKP-SQ, DLKP-M, DLKP-I and DLKP against each other at exponential and stationary phases of growth separately. From these lists, 17 proteins were common to both experiments. Average fold change values were quite different between phases of growth. For the top 36 proteins in each experiment there was an average maximum fold change of **19.6** for exponential phase, and **7.64** for stationary phase.

To assess which cell line showed the greatest difference from the others at a proteomic level, cell lines were ranked based on those containing the largest proportion of highest abundance proteins. Results of this are shown in Table 3.6 and Table 3.7 below. During exponential phase of growth, the largest number of highest abundance proteins could be attributed to DLKP-SQ (45 out of 127 proteins). During stationary phase, the largest number of highest abundance proteins was attributed to DLKP-M (127 out of 203 proteins). In both phases of growth, this observation correlated with the PCA analysis on the peptide features in Figure 3.1 and Figure 3.2.

Highest Expressed in Exponential	Proteins by Number
DLKP-SQ	45
DLKP-M	35
DLKP	26
DLKP-I	21
Total	**127**

Table 3.6: Displayed is the total number of proteins from the global comparison during exponential phase, and how many proteins were the most abundantly expressed in each cell line.

Highest Expressed in Stationary	Proteins by Number
DLKP-M	137
DLKP-SQ	24
DLKP	26
DLKP-I	16
Total	**203**

Table 3.7: Displayed is the total number of proteins from the global comparison for stationary phase, and how many proteins were the most abundantly expressed in each cell line.

3.1.3.3. Two-Sample Comparisons:

The global comparison experimental designs described above are useful for an overall look at how proteins are expressed between all samples as one large set. Using that method, it is possible to track the expression profiles of interesting proteins across all sample groups. The two-sample analysis method allows the determination on intermediary cell line fold changes, and provides detailed differential protein information between just two cell lines in each analysis. Cell lines were compared using this method e.g. DLKP-M vs. DLKP-I. This way, the analysis is focused on one sample group versus another with the aim of pulling out the maximum number of statistically significant proteins for that one comparison.

Individual comparisons were processed by PCA analysis to ensure cell line replicates clustered well and groups separated well. A representative PCA graph is shown in Figure 3.4.

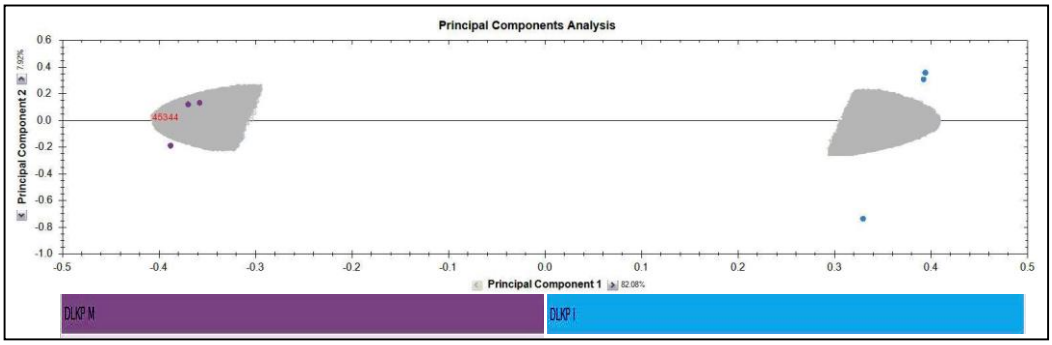


Figure 3.4: Principal component analysis for DLKP-M vs. DLKP-I during exponential phase of growth. PCA showed good separation of sample groups with reasonably tight clustering of samples. The PCA analysis was performed on statistically significant peptide features (ANOVA $p < 0.05$).

The following pages display the top ten highest abundance proteins found for each two-sample cell line comparison of DLKP and the clonal subpopulations. Proteins are sorted based on fold change, with the proteins showing the largest abundance change between any two samples ranked first. All proteins pass the criteria and statistical significance requirements, and full lists are shown in Appendix 8.1. From these lists, a number of proteins were highlighted as repeatedly highly expressed in a single clone for both phases of growth. Shootin-1 can be seen as highly expressed in DLKP-SQ during exponential and stationary phases of growth in Table 3.14 and Table 3.15 respectively for example, as well as others. Myristoylated alanine-rich C-kinase is associated with high expression in DLKP-M as seen in Table 3.8 and Table 3.9 during both exponential and stationary phases of growth. Desmoglein-3 showed extremely high expression in DLKP-I during exponential phase as seen in Table 3.10. These proteins were taken forward for further investigation.

DLKP-SQ vs. DLKP-M Exponential Phase

Name	Fold Change	Highest Abundance
Serum albumin	88.0	DLKP-SQ
Vitamin D-binding protein	40.7	DLKP-SQ
UBX domain-containing protein 1	23.4	DLKP-SQ
Lactotransferrin	17.8	DLKP-SQ
MICOS complex subunit MIC19	14.5	DLKP-SQ
Hemoglobin subunit alpha	13.5	DLKP-SQ
Keratin, type I cytoskeletal 16	11.5	DLKP-SQ
Protein CYR61	11.0	DLKP-SQ
Myristoylated alanine-rich C-kinase substrate	20.9	DLKP-M
Collagen alpha-1(III) chain	8.3	DLKP-M
PDZ and LIM domain protein 5	7.3	DLKP-M
Histone H4	4.1	DLKP-M
Gamma-enolase	4.0	DLKP-M
Histone H3.1t	3.4	DLKP-M
Guanine nucleotide-binding protein G(i) subunit alpha-1	3.4	DLKP-M
Tryptophan--tRNA ligase, cytoplasmic	3.1	DLKP-M

Table 3.8: Highest abundance proteins for DLKP-SQ vs. DLKP-M comparison. Top 10 highest fold changes for each cell line are shown.

DLKP-SQ vs. DLKP-M Stationary Phase

Name	Fold Change	Highest Abundance
Protein CYR61	9.5	DLKP-SQ
Actin, cytoplasmic 1	3.7	DLKP-SQ
D-3-phosphoglycerate dehydrogenase	3.6	DLKP-SQ
Shootin-1	3.5	DLKP-SQ
Serum albumin	3.2	DLKP-SQ
Heat shock 70 kDa protein 1A/1B	3.2	DLKP-SQ
Isoleucine--tRNA ligase, mitochondrial	2.7	DLKP-SQ
Fumarylacetoacetase	2.7	DLKP-SQ
Myristoylated alanine-rich C-kinase substrate	21.0	DLKP-M
Collagen alpha-1(III) chain	15.7	DLKP-M
Histone H4	12.1	DLKP-M
Insulin-like growth factor 2 mRNA-binding protein 3	6.4	DLKP-M
Chromatin target of PRMT1 protein	6.4	DLKP-M
Cysteine and glycine-rich protein 2	5.2	DLKP-M
Histone H2A type 1-B/E	4.9	DLKP-M
Sulfatase-modifying factor 2	4.6	DLKP-M

Table 3.9: Highest abundance proteins for DLKP-SQ vs. DLKP-M comparison. Top 10 highest fold changes in each cell line are shown.

DLKP-SQ vs. DLKP-I: Exponential Phase

Name	Fold Change	Highest Abundance
Vitamin D-binding protein	22.6	DLKP-SQ
Serum albumin	20.8	DLKP-SQ
Lactotransferrin	13.8	DLKP-SQ
D-3-phosphoglycerate dehydrogenase	5.8	DLKP-SQ
Bifunctional ATP-dependent dihydroxyacetone kinase	5.4	DLKP-SQ
Synaptosomal-associated protein 29	4.7	DLKP-SQ
Keratin, type II cytoskeletal 2 epidermal	4.4	DLKP-SQ
PDZ and LIM domain protein 1	4.4	DLKP-SQ
Desmoglein-3	175.3	DLKP-I
CD166 antigen	9.2	DLKP-I
Myristoylated alanine-rich C-kinase substrate	8.2	DLKP-I
Protein S100-A10	8.0	DLKP-I
40S ribosomal protein S29	5.3	DLKP-I
PDZ and LIM domain protein 5	3.6	DLKP-I
Glutaminase kidney isoform, mitochondrial	3.2	DLKP-I
Neuroblast differentiation-associated protein AHNAK	3.1	DLKP-I

Table 3.10: Highest abundance proteins for DLKP-SQ vs. DLKP-I comparison. Top 10 highest fold changes in each cell line are shown.

DLKP-SQ vs. DLKP-I: Stationary Phase

Name	Fold Change	Highest Abundance
D-3-phosphoglycerate dehydrogenase	5.4	DLKP-SQ
Peroxiredoxin-2	4.4	DLKP-SQ
Putative helicase MOV-10	4.2	DLKP-SQ
1,4-alpha-glucan-branching enzyme	3.6	DLKP-SQ
Fumarylacetoacetase	3.5	DLKP-SQ
Ataxin-10	3.4	DLKP-SQ
Heat shock protein beta-1	2.8	DLKP-SQ
Isoleucine--tRNA ligase, mitochondrial	2.7	DLKP-SQ
Tissue factor pathway inhibitor 2	14.9	DLKP-I
CD166 antigen	7.3	DLKP-I
Myristoylated alanine-rich C-kinase substrate	4.5	DLKP-I
Insulin-like growth factor 2 mRNA-binding protein 3	4.3	DLKP-I
Neuroblast differentiation-associated protein AHNAK	3.2	DLKP-I
Annexin A1	2.8	DLKP-I
Vasodilator-stimulated phosphoprotein	2.7	DLKP-I
Glutaminase kidney isoform, mitochondrial	2.5	DLKP-I

Table 3.11: Highest abundance proteins for DLKP-SQ vs. DLKP-I comparison. Top 10 highest fold changes in each cell line are shown.

DLKP-I vs. DLKP-M: Exponential Phase

Name	Fold Change	Highest Abundance
Desmoglein-3	97.3	DLKP-I
Tissue factor pathway inhibitor 2	19.0	DLKP-I
Neurotensin/neuromedin N	5.0	DLKP-I
CD166 antigen	4.6	DLKP-I
Uridine phosphorylase 1	4.1	DLKP-I
Hemoglobin subunit alpha	3.2	DLKP-I
Heat shock 70 kDa protein 1A/1B	2.9	DLKP-I
Protein S100-A10	2.8	DLKP-I
Bifunctional ATP-dependent dihydroxyacetone kinase	8.2	DLKP-M
1,4-alpha-glucan-branching enzyme	4.7	DLKP-M
Transferrin receptor protein 1	3.7	DLKP-M
KDEL motif-containing protein 2	3.4	DLKP-M
Caldesmon	3.3	DLKP-M
3-hydroxyisobutyrate dehydrogenase, mitochondrial	3.0	DLKP-M
Collagen alpha-1(III) chain	3.0	DLKP-M
Dihydropyrimidinase-related protein 2	3.0	DLKP-M

Table 3.12: Highest abundance proteins for DLKP-I vs. DLKP-M individual comparison. Top 10 highest fold changes in each cell line are shown.

DLKP-I vs. DLKP-M: Stationary Phase

Name	Fold Change	Highest Abundance
Protein CYR61	9.7	DLKP-I
Tissue factor pathway inhibitor 2	7.8	DLKP-I
CD166 antigen	4.7	DLKP-I
Hepatocyte growth factor	4.4	DLKP-I
Kinectin	3.8	DLKP-I
Hematopoietic progenitor cell antigen CD34	3.6	DLKP-I
Ras GTPase-activating-like protein IQGAP1	3.6	DLKP-I
Protein PRRC2C	3.2	DLKP-I
Histone H4	13.0	DLKP-M
Collagen alpha-1(III) chain	10.7	DLKP-M
Protein NipSnap homolog 1	5.9	DLKP-M
Peroxiredoxin-2	5.7	DLKP-M
1,4-alpha-glucan-branching enzyme	5.2	DLKP-M
Myristoylated alanine-rich C-kinase substrate	5.1	DLKP-M
Sulfatase-modifying factor 2	4.7	DLKP-M
Histone H2A type 1-B/E	4.4	DLKP-M

Table 3.13: Highest abundance proteins for DLKP-I vs. DLKP-M comparison. Top 10 highest fold changes in each cell line are shown.

DLKP-SQ vs. DLKP-: Exponential Phase

Name	Fold Change	Highest Abundance
Hyaluronan and proteoglycan link protein 1	9.6	DLKP-SQ
Collagen alpha-1(XIV) chain	4.7	DLKP-SQ
Shootin-1	4.3	DLKP-SQ
Neurotensin/neuromedin N	3.8	DLKP-SQ
D-3-phosphoglycerate dehydrogenase	3.0	DLKP-SQ
Four and a half LIM domains protein 1	2.9	DLKP-SQ
Chloride intracellular channel protein 4	2.7	DLKP-SQ
SH3 domain-binding glutamic acid-rich-like protein	2.7	DLKP-SQ
Transgelin	18.5	DLKP
Connective tissue growth factor	16.8	DLKP
Solute carrier family 2, facilitated glucose transporter member 3	15.3	DLKP
Cathepsin Z	12.2	DLKP
Gamma-enolase	5.7	DLKP
Fascin	4.2	DLKP
Hepatocyte growth factor	3.1	DLKP
Integrin beta-1	2.5	DLKP

Table 3.14: Highest abundance proteins for DLKP-SQ vs. DLKP comparison. Top 10 highest fold changes in each cell line are shown

DLKP-SQ vs. DLKP-: Stationary Phase

Name	Fold Change	Highest Abundance
Fumarylacetoacetase	3.9	DLKP-SQ
Collagen alpha-1(XIV) chain	3.8	DLKP-SQ
Myristoylated alanine-rich C-kinase substrate	3.6	DLKP-SQ
Shootin-1	3.5	DLKP-SQ
D-3-phosphoglycerate dehydrogenase	3.2	DLKP-SQ
Keratin, type II cytoskeletal 2 epidermal	3.1	DLKP-SQ
Keratin, type I cytoskeletal 10	2.8	DLKP-SQ
SH3 domain-binding glutamic acid-rich-like protein	2.6	DLKP-SQ
Connective tissue growth factor	15.6	DLKP
Transgelin	7.0	DLKP
Hematopoietic progenitor cell antigen CD34	4.2	DLKP
Hepatocyte growth factor	3.5	DLKP
Fascin	3.2	DLKP
Cathepsin Z	3.0	DLKP
Serpin H1	2.2	DLKP
Neuroblast differentiation-associated protein AHNAK	2.2	DLKP

Table 3.15: Highest abundance proteins for DLKP-SQ vs. DLKP comparison. Top 10 highest fold changes in each cell line are shown.

DLKP-M vs. DLKP: Exponential Phase

Name	Fold Change	Highest Abundance
Myristoylated alanine-rich C-kinase substrate	49.8	DLKP-M
Collagen alpha-1(III) chain	12.5	DLKP-M
Dihydropyrimidinase-related protein 2	5.3	DLKP-M
PDZ and LIM domain protein 5	4.4	DLKP-M
3-hydroxyisobutyrate dehydrogenase, mitochondrial	4.0	DLKP-M
Histone H4	3.6	DLKP-M
Collagen alpha-1(XIV) chain	3.6	DLKP-M
Cysteine and glycine-rich protein 2	3.5	DLKP-M
Connective tissue growth factor	122.7	DLKP
Protein CYR61	22.6	DLKP
Transgelin	18.3	DLKP
Hepatocyte growth factor	15.0	DLKP
Hemoglobin subunit alpha	11.4	DLKP
Solute carrier family 2, facilitated glucose transporter member 3	10.8	DLKP
F-box only protein 2	8.2	DLKP
Cathepsin Z	5.3	DLKP

Table 3.16: Highest abundance proteins for DLKP-M vs. DLKP individual comparison. Top 10 highest fold changes in each cell line are shown.

DLKP-M vs. DLKP: Stationary Phase

Name	Fold Change	Highest Abundance
Myristoylated alanine-rich C-kinase substrate OS	24.5	DLKP-M
Collagen alpha-1(III) chain OS	20.2	DLKP-M
Histone H4 OS	12.3	DLKP-M
Chromatin target of PRMT1 protein OS	7.5	DLKP-M
Cysteine and glycine-rich protein 2 OS	6.8	DLKP-M
SH3 domain-binding glutamic acid-rich-like protein OS	5.5	DLKP-M
Small nuclear ribonucleoprotein Sm D1 OS	4.8	DLKP-M
TATA-binding protein-associated factor 2N OS	4.7	DLKP-M
Connective tissue growth factor OS	28.7	DLKP
Protein CYR61 OS	20.0	DLKP
Hematopoietic progenitor cell antigen CD34 OS	9.7	DLKP
Hepatocyte growth factor OS	6.7	DLKP
Heat shock 70 kDa protein 1A/1B OS	3.5	DLKP
B-cell receptor-associated protein 31 OS	3.1	DLKP
Kinectin OS	3.1	DLKP
Gap junction alpha-1 protein OS	2.8	DLKP
Fascin OS	2.7	DLKP
Protein PRRC2C OS	2.5	DLKP

Table 3.17: Highest abundance proteins for DLKP-M vs. DLKP comparison. Top 10 highest fold changes in each cell line are shown.

DLKP-I vs. DLKP: Exponential Phase

Name	Fold Change	Highest Abundance
Desmoglein-3	81.6	DLKP-I
Myristoylated alanine-rich C-kinase substrate	19.7	DLKP-I
Neurotensin/neuromedin N	7.5	DLKP-I
Protein S100-A10	7.1	DLKP-I
CD166 antigen	5.1	DLKP-I
Collagen alpha-1(III) chain	4.2	DLKP-I
PDZ and LIM domain protein 5	4.2	DLKP-I
Glutaminase kidney isoform, mitochondrial	3.3	DLKP-I
Solute carrier family 2, facilitated glucose transporter member 3	14.6	DLKP
Transgelin	14.5	DLKP
Connective tissue growth factor	14.1	DLKP
Cathepsin Z	4.5	DLKP
1,4-alpha-glucan-branching enzyme	4.1	DLKP
Gamma-enolase	4.1	DLKP
Hemoglobin subunit alpha	2.9	DLKP
Peroxiredoxin-2	2.9	DLKP

Table 3.18: Highest abundance proteins for DLKP-I vs. DLKP comparison. Top 10 highest fold changes in each cell line are shown

DLKP-I vs. DLKP: Stationary Phase

Name	Fold Change	Highest Abundance
Tissue factor pathway inhibitor 2	10.8	DLKP-I
CD166 antigen	5.5	DLKP-I
Myristoylated alanine-rich C-kinase substrate	5.3	DLKP-I
PDZ and LIM domain protein 5	3.9	DLKP-I
Annexin A1	3.4	DLKP-I
Neurotensin/neuromedin N	3.0	DLKP-I
Insulin-like growth factor 2 mRNA-binding protein 3	2.8	DLKP-I
Glutaminase kidney isoform, mitochondrial	2.7	DLKP-I
Keratin, type I cytoskeletal 10	2.7	DLKP-I
Transgelin	11.0	DLKP
Connective tissue growth factor	11.0	DLKP
1,4-alpha-glucan-branching enzyme	4.0	DLKP
Peroxiredoxin-2	3.9	DLKP
Hematopoietic progenitor cell antigen CD34	2.7	DLKP
Exportin-2	2.7	DLKP
KDEL motif-containing protein 2	2.7	DLKP
Putative helicase MOV-10	2.6	DLKP

Table 3.19: Highest abundance proteins for DLKP-I vs. DLKP individual comparison. Top 10 highest fold changes in each cell line are shown.

3.1.3.4. Global Analysis Including the Combination Cell Lines

In a separate experimental design, the global analysis method was applied again, this time incorporating the combination cell lines which were created by combining certain proportions of the DLKP clonal subpopulations, as described in Section 3.1.1. This was done to allow the assessment of protein expression profiles in the combination lines of proteins which showed strong abundance in the individual clones. Combining these two analyses together may shed light on how the clones interact to enhance or inhibit protein expression in the DLKP cell line.

3.1.4. Expression Patterns of Proteins of Interest

Proteins which were highlighted as significant by the global and two-sample comparison methods show very stark differences in abundance between the DLKP cell line and its clonal subpopulations. To assess how each protein was expressed across all cell lines, proteins of interest were displayed graphically based on protein abundance values. The expression pattern of a protein can be displayed as a graph of these abundance values. A selection of proteins which were deemed as significantly differentially expressed are displayed in the following pages. Proteins were chosen based on the following criteria: i) passing all the stringent statistical criteria described previously ii) being very highly abundant in one cell line only iii) appearing in more than one of the two-sample comparisons.

The subset of protein targets showing highest abundance in individual cell lines are shown in Figure 3.6-Figure 3.23. These proteins were highlighted by the label-free analyses described previously in Section 3.1.3. In the case of DLKP-SQ, Hemoglobin subunit alpha (Figure 3.5) showed a maximum fold change of 13.5 between DLKP-SQ and DLKP-M (ANOVA, p= 2.73E-04) during exponential phase of growth. D-3-phosphoglycerate dehydrogenase (Figure 3.8) also showed highest abundance in DLKP-SQ with a maximum fold change of 5.77 between DLKP-SQ and DLKP-I (ANOVA, p= 5.58E-10) during exponential phase of growth, and a similar protein abundance profile

for stationary phase (Figure 3.9). Shootin-1 (Figure 3.6) was also highlighted and this protein was taken forward for further investigation as described in Section 3.1.5.

Proteins associated with high abundance in DLKP-M were also highlighted by the various analyses carried out in this study. Collagen alpha-1 (III) chain showed highest abundance in DLKP-M with a maximum fold change of 12.47 between DLKP-M and DLKP (ANOVA, p= 1.65E-11) during exponential phase of growth (Figure 3.12). A similar trend was also shown during stationary phase of growth (Figure 3.13) with a maximum fold change of 20.16 between DLKP-M and DLKP (ANOVA, p= 1.94E-10). Histone H4 showed a maximum fold change of 29.23 between DLKP-M and SQ+M (ANOVA, p= 1.40E-08) but only was highlighted in the stationary phase analyses (Figure 3.14). Myristoylated alanine-rich C-Kinase substrate was also highlighted by these analyses and taken forward for further investigation as described in Section 3.1.5.

In the case of DLKP-I, a number of proteins showed high expression in this cell line when protein expression profiles were generated. CD166 showed highest expression in DLKP-I with a maximum fold change of 9.18 between this cell line and DLKP-SQ (ANOVA, p= 1.13E-06) during exponential phase of growth (Figure 3.16). Similarly a maximum fold change of 7.3 between DLKP-I and DLKP-SQ (ANOVA, p= 3.26E-06) was found during stationary phase of growth (Figure 3.17). Neurotensin/Neuromedin N also showed highest expression in DLKP-I with a maximum fold change of 7.46 and 3.02 between DLKP-I and DLKP for exponential (Figure 3.18) and stationary phases (Figure 3.19) of growth respectively (ANOVA, p= 3.64E-08 and p= 5.31E-04). Desmoglein-3 was also highlighted as specifically highly expressed in DLKP-I and taken forward for further investigation as described in Section 3.1.5.

A number of proteins showed specifically high expression in the DLKP cell line itself, however as the aim of this thesis was to investigate protein expression in the clonal subpopulations of DLKP and look for potential marker, these were not pursued. Of particular interest however were Connective tissue growth factor, and Transgelin. Connective tissue growth factor showed maximum fold change of 122.69 and 28.66 between DLKP and DLKP-M during both exponential (Figure 3.20) and stationary phases (Figure 3.21) of growth respectively (ANOVA, p= 1.85E-07 and p= 3.56E-06).

Transgelin showed a maximum fold change between DLKP and DLKP-SQ of 18.46, highest in DLKP (Figure 3.22) during exponential phase of growth (ANOVA, p= 2.22E-06). Transgelin also showed a maximum fold change of 11.38 between DLKP and SQ+M, highest in DLKP (Figure 3.23), during stationary phase of growth (ANOVA, p= 3.53E-06).

Highest in DLKP-SQ

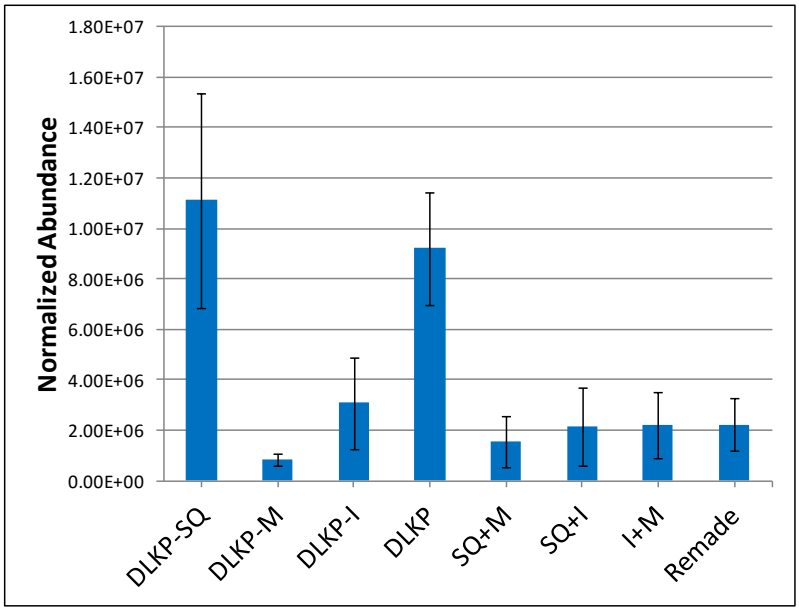


Figure 3.5: Protein expression profile for Hemoglobin subunit alpha during exponential phase of growth.

Highest in DLKP-SQ

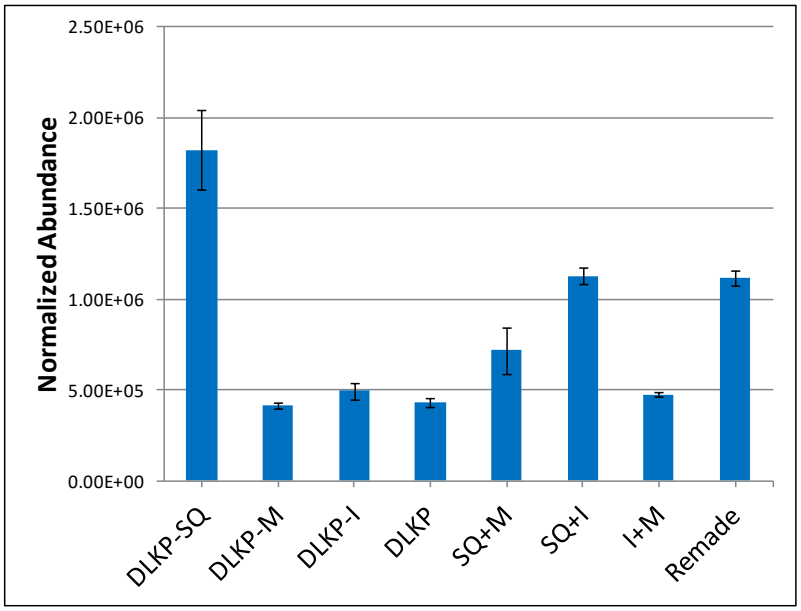


Figure 3.6: Protein expression profile for Shootin-1. It is expressed with highest abundance in DLKP-SQ during exponential phase of growth.

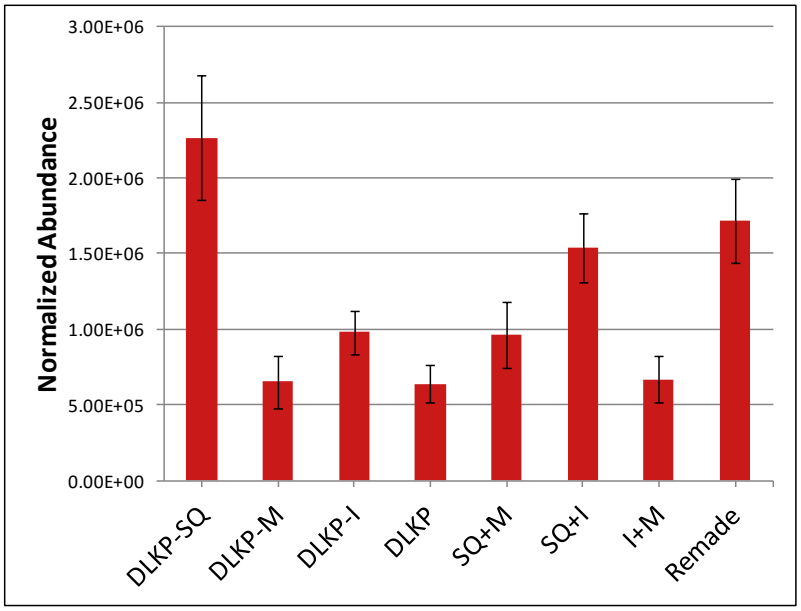


Figure 3.7: Protein expression profile for Shootin-1. Highest expression is seen in DLKP-SQ during stationary phase also.

Highest in DLKP-SQ

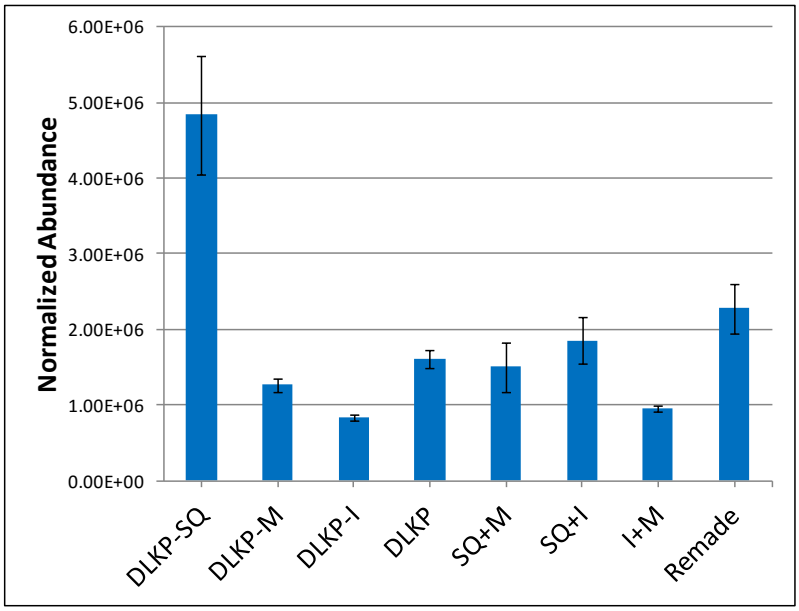


Figure 3.8: Protein expression profile for D-3-phosphoglycerate dehydrogenase during exponential phase of growth.

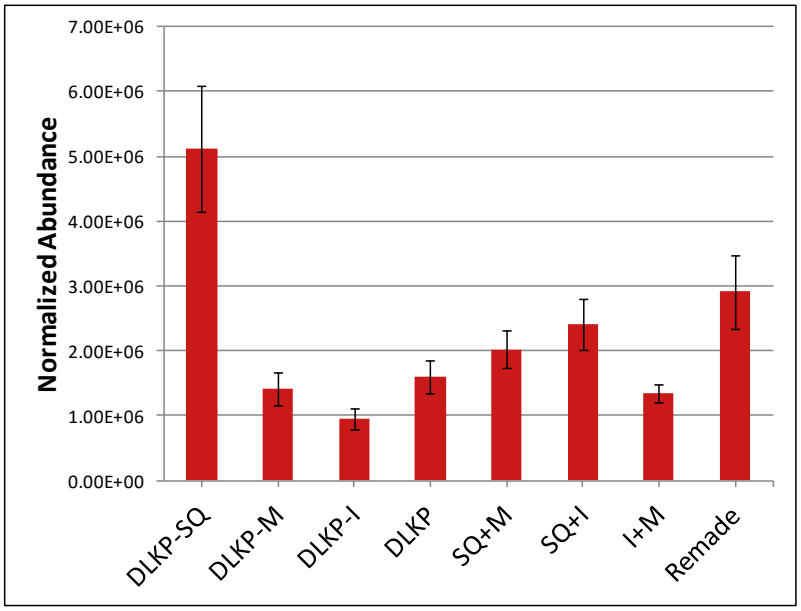


Figure 3.9: Protein expression profile for D-3-phosphoglycerate dehydrogenase during stationary phase of growth.

Highest in DLKP-M

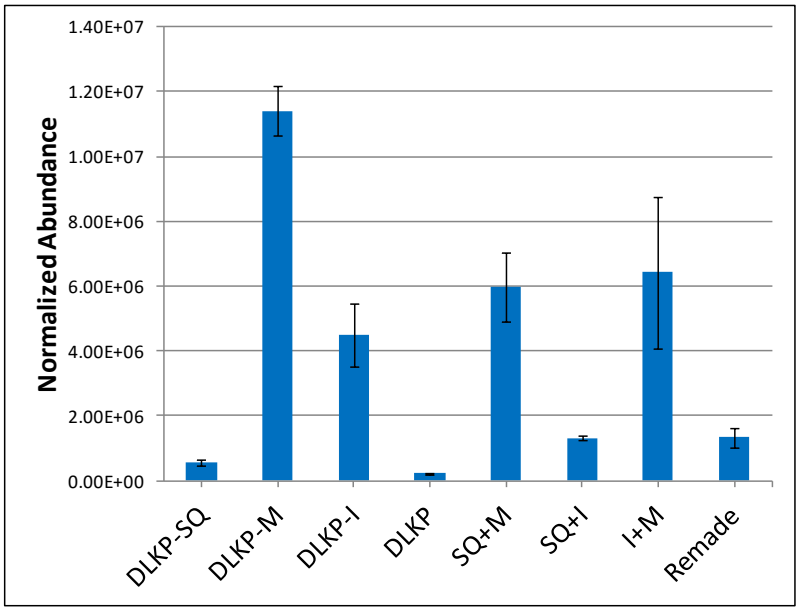


Figure 3.10: Protein expression profile for Myristoylated alanine-rich C-kinase substrate during exponential phase of growth.

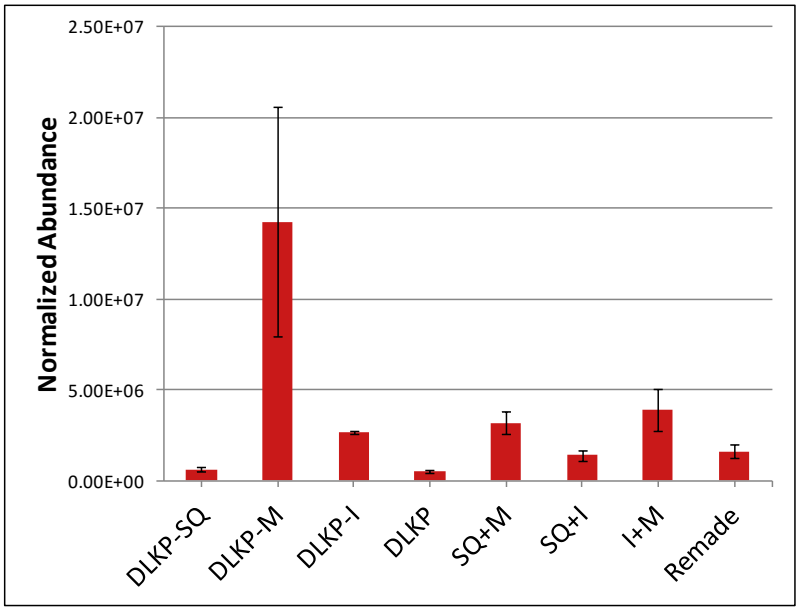


Figure 3.11: Protein expression profile for Myristoylated alanine-rich C-kinase substrate during stationary phase of growth.

Highest in DLKP-M

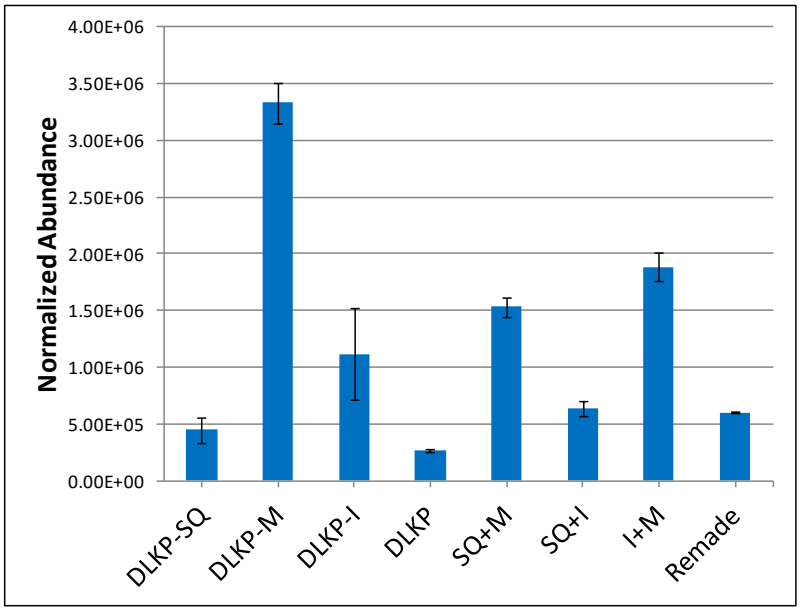


Figure 3.12: Protein expression profile for Collagen alpha-1(III) chain during exponential growth.

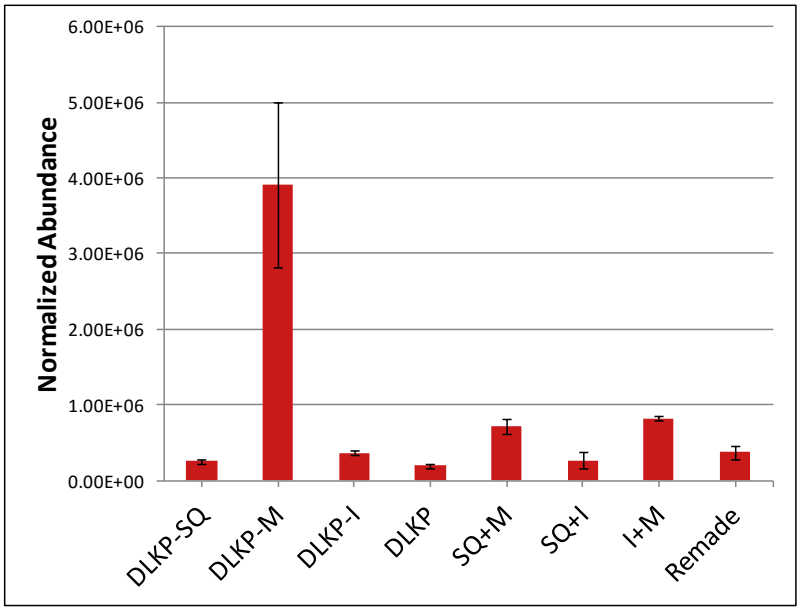


Figure 3.13: Protein expression profile for Collagen alpha-1(III) chain during stationary phase of growth.

Highest in DLKP-M

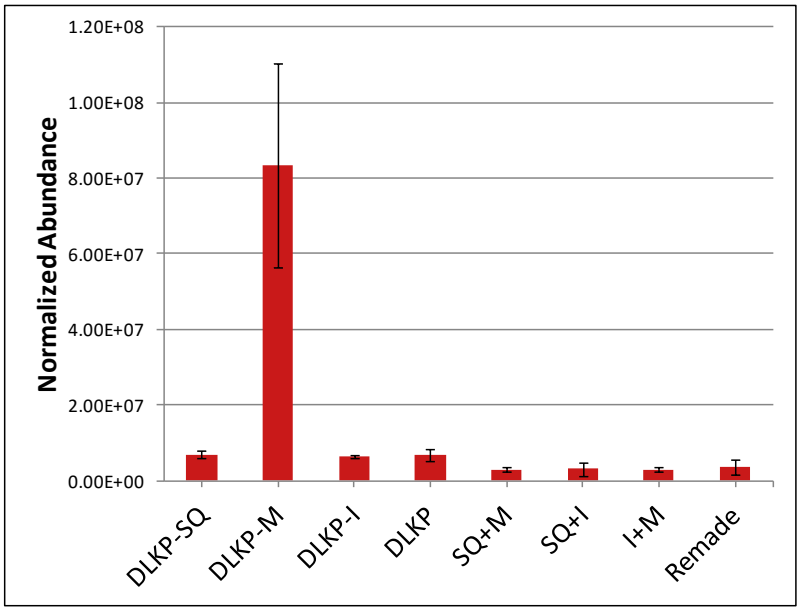


Figure 3.14: Protein expression profile for Histone H4 during stationary phase of growth.

Highest in DLKP-I

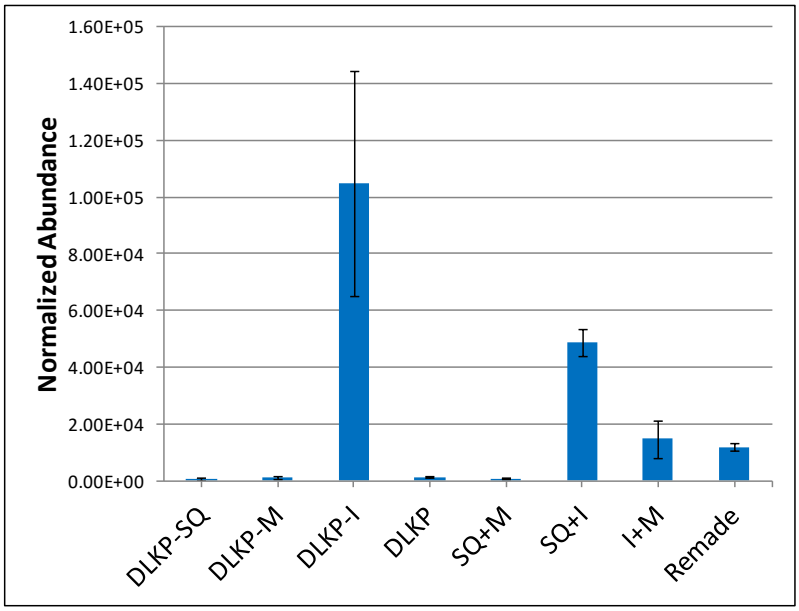


Figure 3.15: Protein expression profile for Desmoglein-3 during exponential phase of growth.

Highest in DLKP-I

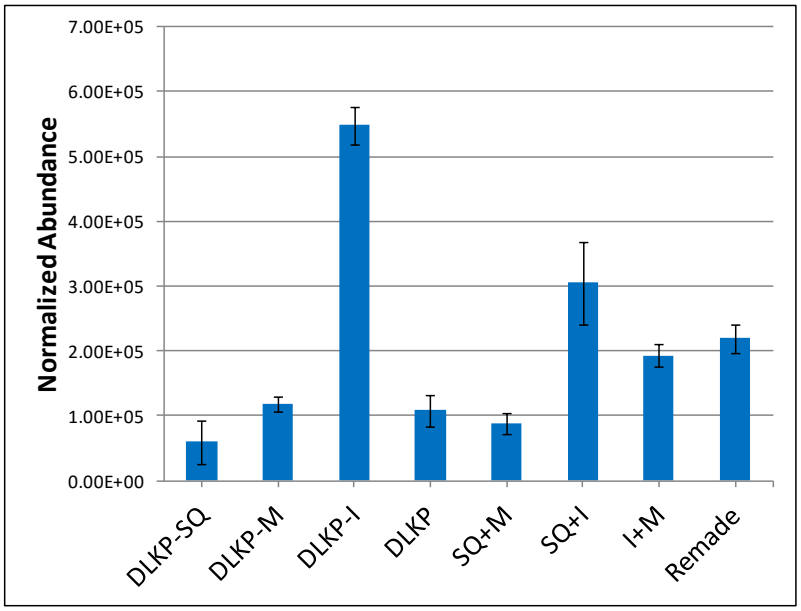


Figure 3.16: Protein expression profile for CD166 during exponential phase of growth.

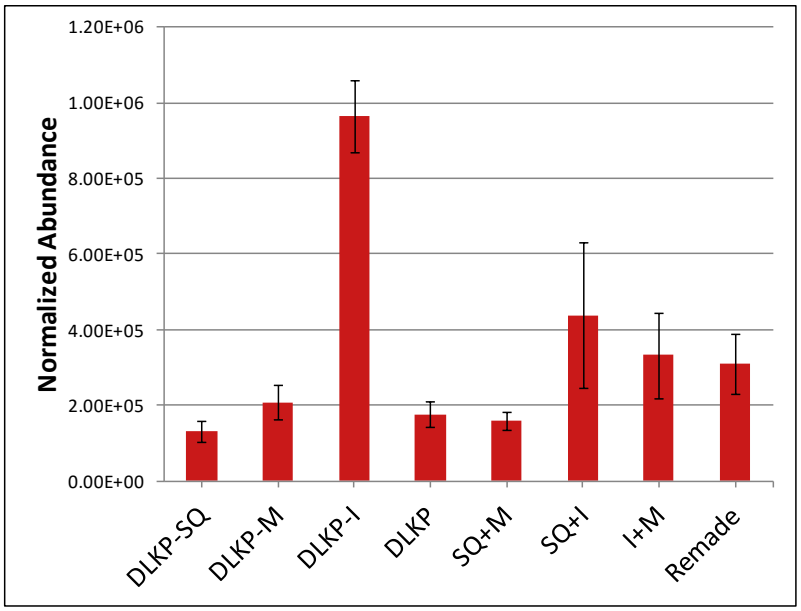


Figure 3.17: Protein expression profile for CD166 during stationary phase.

Highest in DLKP-I

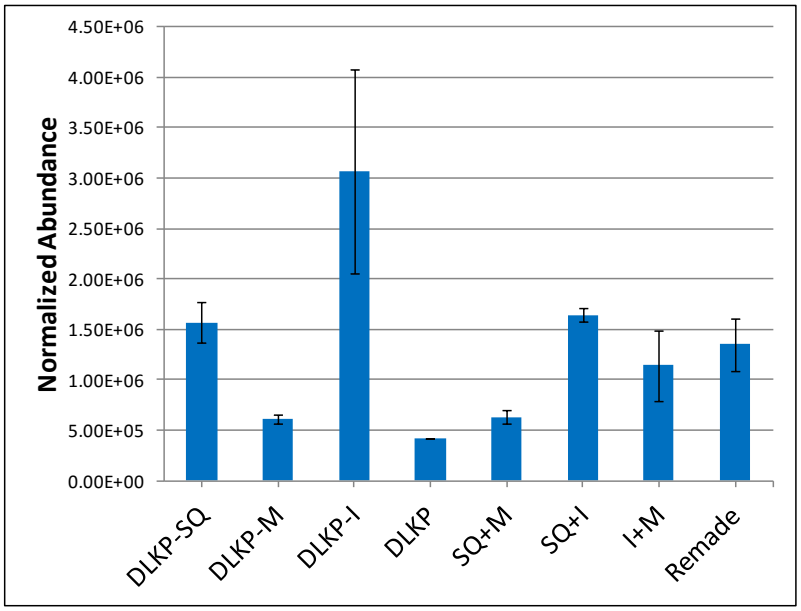


Figure 3.18: Protein expression profile for Neurotensin/Neuromedin N during exponential phase.

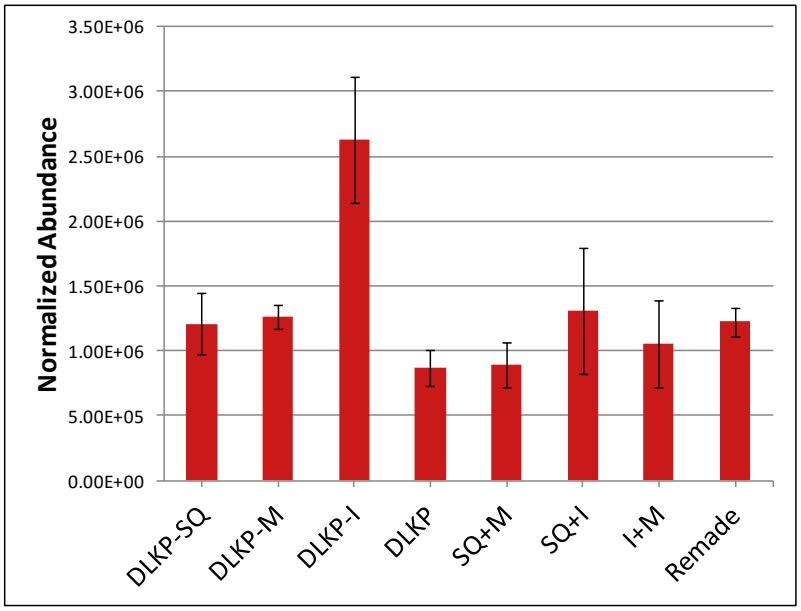


Figure 3.19: Protein expression profile for Neurotensin/Neuromedin N during stationary phase.

Highest in DLKP

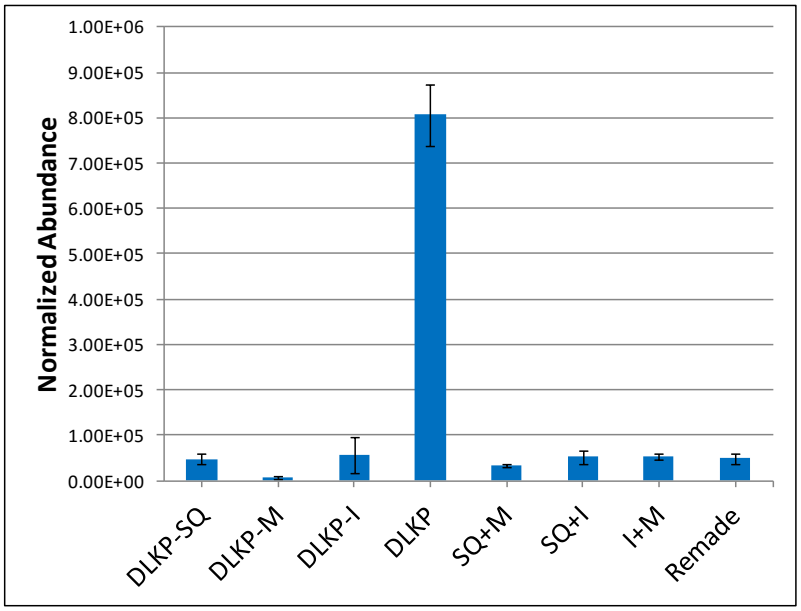


Figure 3.20: Protein expression profile for Connective tissue growth factor shows highest expression in DLKP during exponential phase of growth.

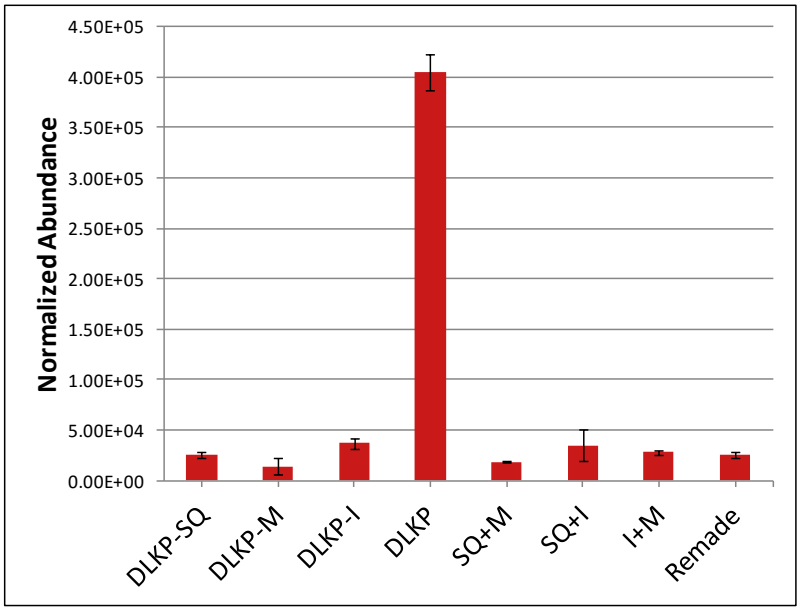


Figure 3.21: Protein expression profile for Connective tissue growth factor during stationary phase.

Highest in DLKP

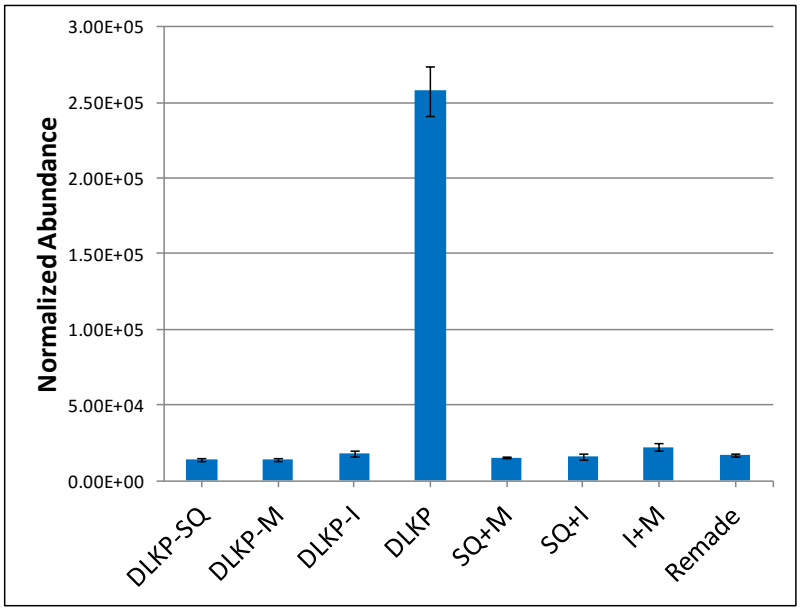


Figure 3.22: Protein expression profile for Transgelin shows highest expression in DLKP during exponential phase.

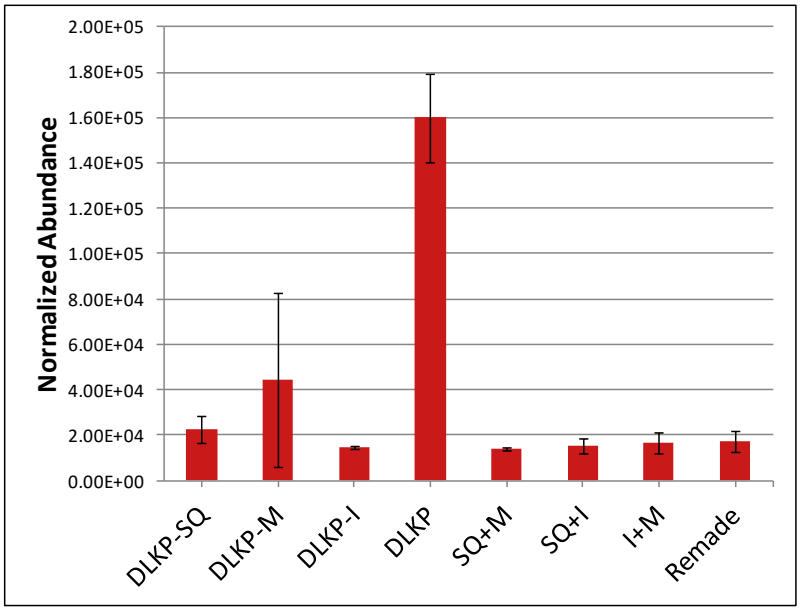


Figure 3.23: Protein expression profile for Transgelin shows the same expression pattern during stationary phase.

3.1.5. Selection of Proteins for Further Investigation

The quantitative label-free proteomic analysis resulted in the discovery of a large dataset of differentially expressed proteins between the cell lines. The global analysis method allowed the visualisation of the protein expression patterns of these proteins across the entire sample set. This enabled the highlighting of proteins which showed a high abundance in one clone relative to the other cell lines in the comparison. The addition of the combination cell lines into the global method of analysis facilitated the expression patterns of proteins to be determined during co-culture conditions. The two-sample analysis method supplemented the global method, by providing the intermediary fold change values and associated statistical significances for proteins of interest.

These analyses resulted in a number of proteins of interest for further investigation (See Section 3.1.4), which showed high fold changes and enriched protein expression in individual cell lines. These proteins are shown in Table 3.20.

Protein	DLKP Clone
Hemoglobin subunit alpha	DLKP-SQ
D-3-phosphoglycerate dehydrogenase	DLKP-SQ
Shootin-1	DLKP-SQ
Myristoylated alanine-rich C-Kinase substrate	DLKP-M
Collagen alpha-1 (III)-chain	DLKP-M
Histone H4	DLKP-M
Desmoglein-3	DLKP-I
CD166	DLKP-I

Table 3.20: A list of proteins which underwent preliminary investigation to validate their expression pattern in the DLKP cell lines.

Details of proteins shown in Table 3.2 can be found in Section 3.1.4 along with their protein expression patterns. The majority of these potential targets did not validate by western blot analysis, showing protein expression patterns in no way resembling

those found by the label-free analysis. For a number of the potential targets, the western blot analyses resulted in high background images, or a large number of non-specific bands being highlighted which obscured the result. Three proteins however validated well with the expression pattern found by western blot matching closely to the label-free protein abundance data. The three proteins were chosen for follow-up investigation: **Shootin-1; MARCKS; and Desmoglein-3.**

Shootin-1 is a relatively novel protein, exclusively studied in neuronal cells to date. It is a recently discovered protein thought to play a significant role in Actin-related cell motility (Sapir et al. 2013). Very few studies have been carried out on it and it represents an opportunity to characterise a novel protein in lung cancer. The protein expression profile during both exponential and stationary phases of growth (Figure 3.6 and Figure 3.7) show highest expression in DLKP-SQ relative to all other cell lines.

MARCKS is a ubiquitously expressed protein found to be especially abundant in the nervous system. It is the major substrate for Protein Kinase-C, and has been shown to play roles in cell migration and dendrite branching in neuronal cells (Larsson 2006). It has also been shown to bind actin, and is little studied in relation to lung cancer. This protein was found to be highly expressed in DLKP-M by the quantitative label-free proteomic analyses during both exponential and stationary phases of growth (Figure 3.10 and Figure 3.11 respectively).

Desmoglein-3 is a transmembrane protein exclusively expressed in stratified epithelium, and is a member of the Cadherin family. It is predominantly associated with its role in the desmosome cell-cell adhesion structure, and also plays a significant role in the autoimmune disease Pemphigus Vulgaris (Jennings et al. 2011). Little is known about its function in cancer, though it has been shown to be involved in cell migration in oral squamous cell carcinoma. This protein was found to be specifically highly expressed in DLKP-I during stationary phase (Figure 3.15), with very little to potentially no expression in the other clones.

These three proteins were taken forward for validation and further investigation into their roles in their respective clones. The relative novelty of Shootin-1 and MARCKS expression in lung cancer cells promoted their selection. The highly specific expression of Desmoglein-3 as well as the small number of studies in lung cancer

facilitated its selection. The following chapters describe the quantitative label-free data found for these proteins, and the follow-up investigations into their roles in their respective clones.

Chapter 4. Shootin-1

4.1. Shootin-1:

Shootin-1 was found to show highest abundance in DLKP-SQ in the quantitative label-free proteomic analysis. Shootin-1 was first described as a novel protein involved in neuron cell polarization (Toriyama et al. 2006, Sapir et al. 2013). This group sought to investigate proteins involved in the formation of axons in hippocampal neurons with a view to understanding the mechanism by which neuronal cell polarization occours. Shootin-1 was discovered as a key molecule involved in the process of neuronal polarization and in axon outgrowth. It was found that Shootin-1 accumulates at the leading edge of axonal growth cones and mediates the linkage between F-Actin Retrograde Flow (ARF) and the cell adhesion molecule L1-CAM (Shimada et al. 2008). Linkage between ARF and the CAM family of proteins is thought to transmit the force of actin filament polymerisation to extracellular substrates, resulting in cell motility. The ability of cancerous cells to migrate allows cells to change position within tissues, disseminate into circulation and undergo metastatic growth in distant organs. ARF is therefore an important mechanism in relation to the study of metastasis in cancer, and Shootin-1 may play an important role in this process. To date there are no studies on Shootin-1 in cancer as this protein is little studied, and the focus has been on neuronal cells thus far. The DLKP cell line was found to express neuronal markers (McBride et al. 1998a), therefore this was a reason to choose this protein for follow up.

4.2. Label-free Proteomic Analysis: Shootin-1

The purpose of the label-free analysis experiment was to highlight differentially expressed proteins between DLKP and its clonal subpopulations. Interesting targets are screened for using the protein expression pattern view in the Progenesis software, with the particular goal of highlighting proteins very highly expressed in one of the clones compared to the others. A specifically highly expressed protein in a single cell line compared to the others could potentially play a role in its known phenotypic characteristics. It could also become a marker for those characteristics and warrant further functional investigation. To focus in on potential candidate proteins, the data was screened initially using the protein expression profile

view shown in Figure 4.1. From an initial look at the data, Shootin-1 stood out as a prime candidate due to its high expression in DLKP-SQ compared to the other cell lines.

4.2.1. Global Analysis: Shootin-1

- During exponential phase of growth, Shootin-1 had a maximum fold change of **4.41** with highest abundance in **DLKP-SQ** compared to the lowest abundance found in DLKP-M.

- During stationary phase of growth the protein had a maximum fold change of **3.55** with highest abundance in **DLKP-SQ** and lowest in DLKP.

In both experiments, Shootin-1 passed the stringent criteria of: ≥ 2 peptides contributing to the protein identification, an abundance change of ≥ 2 fold between two experimental groups, an ANOVA p- value of <0.05 for a protein between experimental groups. A representative output protein abundance graph from the Progenesis software for exponential phase can be seen in Figure 4.1. Peptide sequences which contributed to the identification of Shootin-1 are shown in Figure 4.2. This image shows the normalised abundance values for Shootin-1 across the DLKP cell line and its clonal subpopulations. During both phases of growth, the abundance of Shootin-1 was very similar for DLKP-M and DLKP, slightly higher in DLKP-I, and approximately 4 fold higher in DLKP-SQ. Next, the expression levels of Shootin-1 were determined using experimental designs which compare the cell lines against each other using a two-sample method: e.g. DLKP-SQ vs. DLKP-M.

Figure 4.1: Label-Free normalised abundances for Shootin1 during exponential phase displayed by Progenesis software. Highest expression is seen in the DLKP-SQ clone followed by DLKP-I. Lowest expression levels are seen in DLKP-M and DLKP. 7 peptides contributed to this protein identification, with a max fold change of 4.41 between DLKP-SQ and DLKP-M, (ANOVA, p= 6.03E-08).

Accession: A0MZ66

Description: Shootin-1 OS=Homo sapiens GN=KIAA1598 PE=1 SV=4 - [SHOT1_HUMAN]

Σ	#	Score	Anova (p)	Max Fold Change	Highest Mean	Lowest Mean	Tag	Abundance	m/z	Charge	Retention Time (mins)	Mass error (ppm)	Peptide Sequence
•	8023	77.020	2.3E-10	2.71	DLKP SQ	DLKP Parents		2.511E+05	586.3318	2	136.352	-0.46	LVEVIEEVNK
•	8247	67.220	1.36E-09	3.93	DLKP SQ	DLKP Parents		2.715E+05	644.3488	2	126.574	-0.03	TVLNSEVLEQR
•	10128	64.060	7.41E-05	3.85	DLKP SQ	DLKP M		1.565E+05	573.2934	2	70.762	-0.22	LTQQLEEER
•	11778	65.970	3.49E-09	4.83	DLKP SQ	DLKP M		3.524E+05	781.3828	2	107.742	-0.51	ATQPETTEEVTDLK
•	12790	102.220	6.72E-12	6.04	DLKP SQ	DLKP M		4.241E+05	975.9803	2	133.816	1.59	VTAEADSSSPTGILATSESK
•	17318	59.520	5.11E-12	6.02	DLKP SQ	DLKP M		3.35E+05	779.7174	3	143.573	0.83	KTLEAEFNSPSPPTPEPGEGPR

Figure 4.2: Peptides which contributed to the identification of Shootin-1 are displayed by Progenesis software. Each peptide had an ANOVA $p < 0.05$, and MASCOT scores ≥40.

4.2.2. Two-Sample Experimental Designs: Shootin-1

Shootin-1 was highlighted as being highly expressed in a single clone when all samples were analysed using the global analysis experimental design (See Section 4.2.1). The expression profile clearly shows the protein as being highly expressed in DLKP-SQ compared to DLKP and the other clonal populations. In this analysis, Shootin-1 expression is assessed using experimental designs which compare two cell lines against each other in a one vs. one approach. This is carried out as described in Section 3.1.3.3. The purpose of this analysis is to uncover more detailed information on how a protein is expressed between each cell line, resulting in the availability of intermediary fold change data between any two samples, information which is not available from a group analysis. Results for Shootin-1 are summarised in Table 4.1.

The analysis was comprised six comparisons for each phase of growth, with every permutation accounted for. Each analysis has stringent filtering criteria applied to the resulting proteins which are: ≥ 2 peptides contributing to the protein identification, an abundance change of ≥ 2 fold between two experimental both groups, an ANOVA p-value of <0.05 for a protein between both experimental groups.

In each analysis containing DLKP-SQ in the comparison, Shootin-1 is identified as being significantly differentially expressed. Shootin-1 is not highlighted as a differentially expressed protein in comparisons which do not include DLKP-SQ in the experimental design. This indicates Shootin-1 is a specifically highly expressed protein in DLKP-SQ in both exponential and stationary phases of growth.

- During exponential phase of growth, Shootin-1 was only highlighted in comparisons which included DLKP-SQ, always being the most abundant in that cell line by **~4-fold** compared to other cell lines. The remaining comparisons did not find Shootin-1 to be statistically significantly differentially expressed and therefore no fold change data is available.
- During stationary phase of growth Shootin-1 was found to be most abundant in DLKP-SQ each time the cell line was compared to one other. Shootin-1 was also found to be statistically significantly differentially expressed between DLKP-I vs. DLKP-M (**2.07-fold**), with highest abundance in DLKP-I. This marked a

slight change from exponential phase data, where the clonal subpopulations expressed Shootin-1 at a similarly low level.

Exponential Phase:

(A)

Individual Comparison	Description	Max fold change	Highest Abundance	Lowest Abundance
DLKP-SQ vs. DLKP-M	**Shootin-1**	**4.41**	**DLKP-SQ**	**DLKP-M**
DLKP-SQ vs. DLKP-I	**Shootin-1**	**3.70**	**DLKP-SQ**	**DLKP-I**
DLKP-I vs. DLKP-M	-	-	-	-
DLKP-SQ vs. DLKP	**Shootin-1**	**4.25**	**DLKP-SQ**	**DLKP**
DLKP-M vs. DLKP	-	-	-	-
DLKP-I vs. DLKP	-	-	-	-

Stationary Phase:

(B)

Individual Comparison	Description	Max fold change	Highest Abundance	Lowest Abundance
DLKP-SQ vs. DLKP-M	**Shootin-1**	**3.48**	**DLKP-SQ**	**DLKP-M**
DLKP-SQ vs. DLKP-I	**Shootin-1**	**2.31**	**DLKP-SQ**	**DLKP-I**
DLKP-I vs. DLKP-M	Shootin-1	2.07	DLKP-I	DLKP-M
DLKP-SQ vs. DLKP	**Shootin-1**	**3.55**	**DLKP-SQ**	**DLKP**
DLKP-M vs. DLKP	-	-	-	-
DLKP-I vs. DLKP	-	-	-	-

Table 4.1: Individual comparisons of the DLKP cell line and clonal subpopulations for exponential (A) and stationary (B) phases of growth. In each comparison containing DLKP-SQ, Shootin-1 was highlighted as having the highest expression in that cell line. For each comparison, ≥ 2 peptides contributed to the identification, and ANOVA p-values were <0.05 in the protein abundance comparisons.

4.3. Validation of Shootin-1 Expression in DLKP and Clonal Subpopulations

The expression pattern of Shootin-1 was followed up by Western Blot validation using the method as described in Section 2.3. This protein has a predicted MW of ~70kDa according to UniProt. However, Shootin-1 was found to have a MW of ~100kDa in the DLKP cell lines. This unexpected MW for Shootin-1 was later validated as accurate by immunoprecipitation which identified Shootin-1 in the 100kDa region of a coomassie stained SDS-PAGE gel by mass spectrometry (See Figure 4.20).

The expression pattern of Shootin-1 found by western blot analysis correlates well with the quantitative label-free data. A representative image of the Western Blot for Shootin-1 in the DLKP and clonal subpopulations is shown in Figure 4.3 A. Highest expression of Shootin-1 is seen in DLKP-SQ. Low level bands are present in all other cell lines, with DLKP-I and DLKP-M showing similar abundance levels, and almost undetectable bands in DLKP. Normalised abundances from the label-free analysis were graphed for Shootin-1 and shown in Figure 4.3 B for comparison.

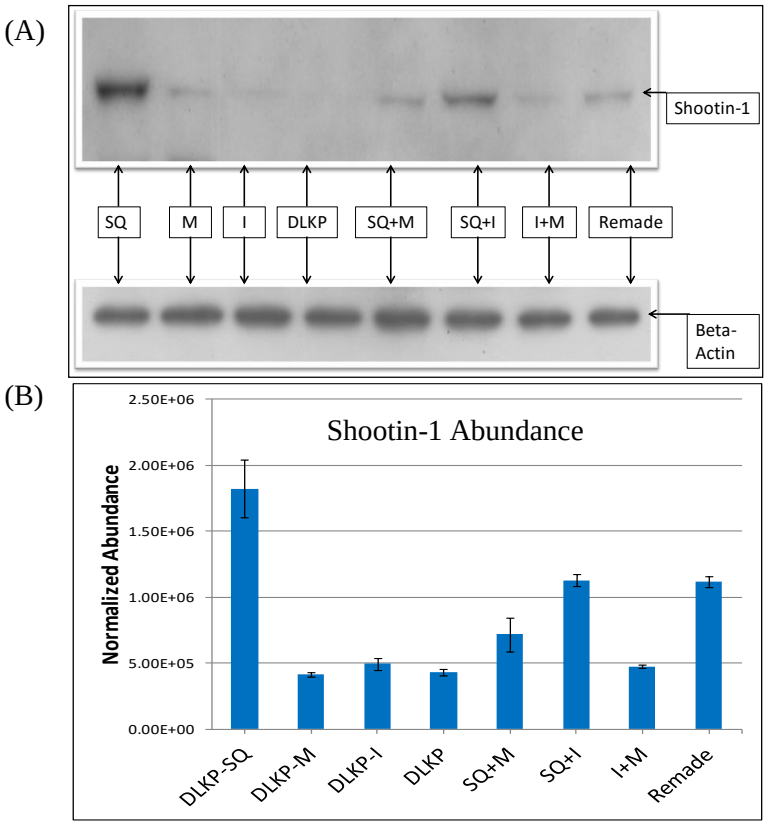


Figure 4.3: (A) Western Blot showing expression levels of Shootin-1 in the DLKP cell line and clonal subpopulations, as well as combination lines. Beta-Actin loading control included. **(B)** Label-free abundance values for Shootin-1 in the DLKP cell line and clonal subpopulations, as well as combination lines. Standard deviations for biological triplicate samples are shown.

4.4. Proteomic Comparison of DLKP Compared to the Three Clones and Combination Cell Lines

Shootin-1 was established as being highly expressed in DLKP-SQ when the experimental design compared DLKP and the clonal subpopulations against each other as a group. One vs. one comparisons showed the fold change of the protein when compared to the remaining cell lines. The expression pattern of Shootin-1 is next checked against the experimental design which incorporates all samples in the analysis. This experimental design includes the 'combination lines' which are newly created cell lines made up of 50:50 ratios of any two clonal subtypes, as well as a reconstituted DLKP based on proportions discovered in previous work (McBride et al. 1998b), as detailed in Section 3.1.1. A large global group analysis was created in Progenesis software which incorporates all samples into one large experimental design containing 24 samples (8 conditions with three biological replicates for each). The combination lines were included in the analysis to determine if co-culture of the DLKP clones had any effect on the expression of proteins of interest. These effects could be inhibition or increased expression of protein targets. Shootin-1 is highlighted as a statistically significantly differentially expressed protein by this analysis. In addition, the abundance levels of Shootin-1 in the combination lines are also readily available from this analysis.

Peptides which identify as belonging Shootin-1 and have contributed to its quantification are grouped together by the Progenesis software and abundances are displayed graphically across all samples. 6 peptides passed the filtering criteria previously mentioned in Section 3.1.2, and each of these contributed to the identification of Shootin-1. These peptides are shown in Figure 4.2.

The expression profile from the analysis which shows Shootin-1 peptide abundances are shown in Figure 4.4. Abundance levels for Shootin-1 are significantly higher in DLKP-SQ than all other sample groups in the experimental design. Additionally, the combination line samples which contain DLKP-SQ as a component show higher abundances of Shootin-1 than samples not containing that cell line. Interestingly, two of the combination lines containing DLKP-SQ as a component did not express Shootin-1 at a similar abundance level as expected. There is a fold change

of **1.57** between SQ+M and SQ+I for Shootin-1, with the highest expression found in SQ+I. These two combination lines would be expected to express Shootin-1 at a similar level based on the data from the clonal subpopulations only. However SQ+M shows a significantly lower expression level of the protein. This suggests inhibition of protein expression is occurring in co-culture of DLKP-SQ and DLKP-M.

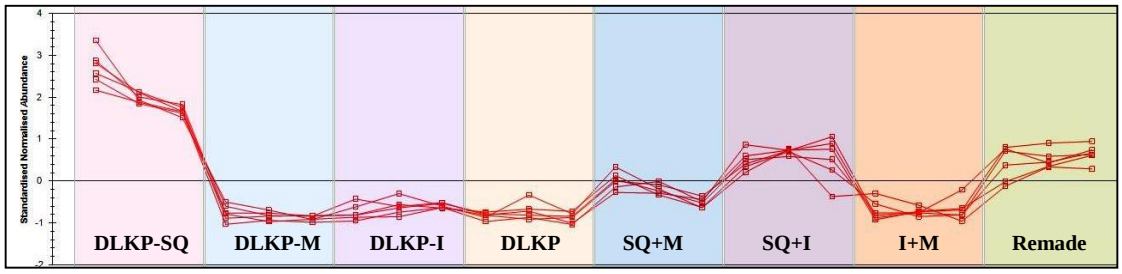


Figure 4.4: Peptide expression pattern as displayed by Progenesis software. Peptide abundances for Shootin-1 show changes between clonal cell lines as well as combination lines (Exponential phase). Highest abundance in DLKP-SQ, lowest in DLKP-M (ANOVA, p= **2.07E-12**).

4.5. Validation of Shootin-1 Expression in DLKP plus the Clones and Combination Cell Lines

4.5.1. Western Blot

The DLKP cell lines along with the combination lines were analysed by western blot to determine if the expression pattern found by the label-free proteomic analysis could be validated. The label-free data shows unexpected abundances of Shootin-1 in the combination lines, with a higher expression of the protein in SQ+I compared to SQ+M. Western Blot analysis validated the expression pattern in the exponential phase samples and can be seen in Figure 4.3 A. The label-free abundance values for Shootin-1 are displayed graphically in Figure 4.3 B for comparison.

4.5.2. mRNA Expression

Following on from western blot validation, DLKP samples were analysed using quantitative PCR (qPCR) to determine the relative levels of Shootin-1 mRNA levels in DLKP, clones and combination cell lines. Relative gene quantification was completed on all cell lines and combinations using beta-actin as the endogenous control. Biological duplicate samples were analysed and results are combined in Figure 4.5. The calculation of relative quantity (RQ) was achieved by normalizing the data to DLKP-SQ.

Gene expression of Shootin-1 correlated well with the label-free proteomics data in the clones. DLKP-SQ has the highest mRNA abundance out of the entire sample set. On average, there is a ~5-fold difference between both DLKP-M and DLKP-I when each is compared to DLKP-SQ. This correlates well with the label-free proteomic data which showed similar fold changes for the same comparisons. The abundance of Shootin-1 in the combinations is much lower than expected given the 1:1 ratio present in those samples. Though the mRNA level for Shootin-1 is low in these combination cell lines, they do show the same expression profile as was found in the proteomic analyses.

DLKP and Remade showed very similar levels of Shootin-1 gene expression. DLKP also shows higher levels of Shootin-1 gene expression than what is shown by label-free profiling data and western blot analysis in relation to the other samples. It is possible that the presence of DLKP-M cells in the DLKP cell line inhibit Shootin-1 expression at a translational level, though further work is required to establish this.

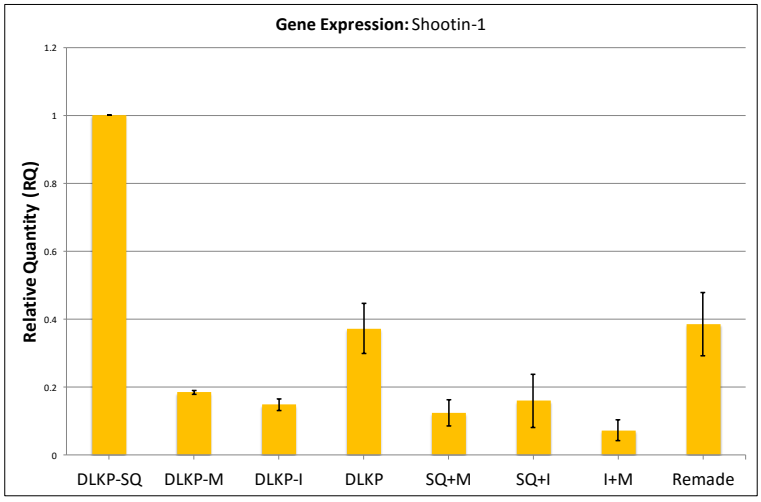


Figure 4.5: qPCR results showing gene expression of Shootin1 in the clones and combination lines. Samples analysed are in biological duplicate (n=2). Error bars shown are the standard deviations between the two replicate experiments. Data are normalised to DLKP-SQ.

4.6. Imaging of Shootin-1 in the DLKP Cell Lines:

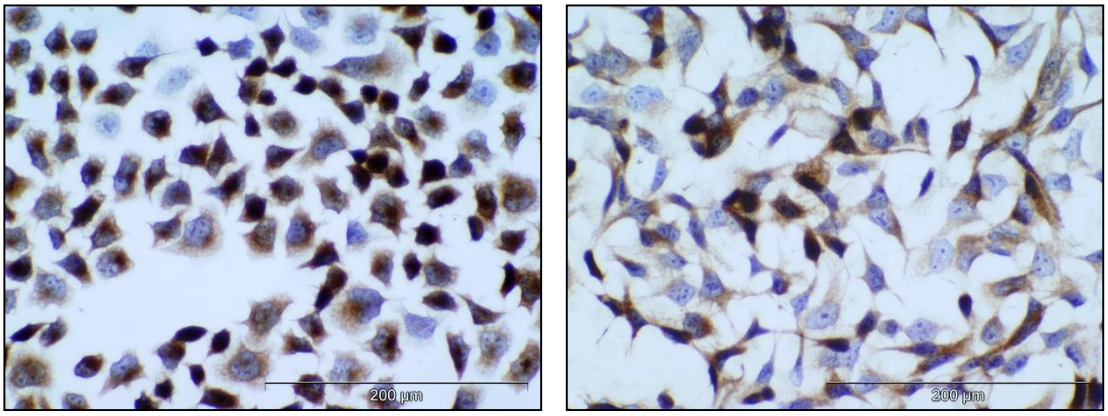
Validation studies confirm Shootin-1 as being highly expressed in DLKP-SQ compared to other cell types at the protein and mRNA level. In order to investigate the localization of Shootin-1 in the DLKP cell lines as well as further validate the expression pattern, Shootin-1 was stained using immunocytochemistry and immunofluorescence. Staining methods were performed as described in Section 2.6 on DLKP and clonal subpopulations fixed to glass slides.

4.6.1. Immunocytochemical Staining of Shootin-1 in DLKP Cell Lines

The staining intensity for Shootin-1 correlated well with the Label-free results, showing high staining levels in DLKP-SQ (See Figure 4.6). The staining in these cells is localized to the cytoplasm, with low staining in the nucleus/nuclear envelope. The next most intense staining is in DLKP-M. This clone has a morphology which presents as being projection rich with many cells showing dendritic-like outgrowths. Many of these projections stain positively for Shootin-1 which may lend support to the idea that the protein is involved in the formation of these outgrowths (Toriyama et al. 2006).

DLKP-I has the lowest level of Shootin-1 staining overall, with the majority of the cells showing low level staining in the cytoplasm of the cells. In the DLKP line, Shootin-1 stains strongly in subpopulations of cells revealing the heterogeneous nature of this cell line. Some of the more intensely stained cells may be DLKP-SQ, however there are some abnormally large cells present with positive staining for Shootin-1 also. These cells may represent a so far uncharacterised clone type. In each clonal subpopulation, there are subpopulations which show different intensities of Shootin-1 staining. This may indicate that each clone is not 100% one cell type only, and may support the theory of the inter-conversion properties of the DLKP clones as previously described (McBride et al. 1998b).

DLKP-SQ **DLKP-M**



DLKP-I **DLKP**

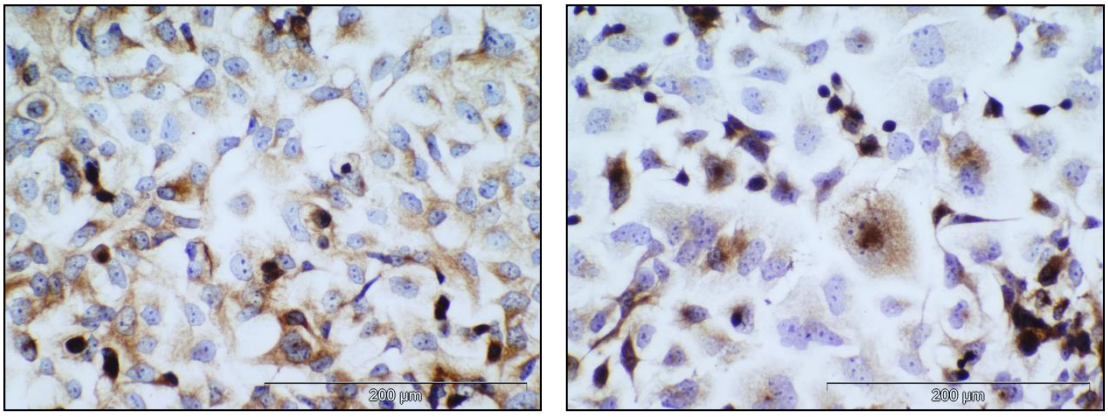


Figure 4.6: Immunocytochemical images of **DLKP-SQ**, **DLKP-M**, **DLKP-I** and **DLKP** stained with Shootin-1. Nuclei are stained with Dapi. Negative controls with secondary antibody only did not result in staining (images not shown). Original magnification used on all micro-pictographs shown at 200x, 200µm scale bar. (n=2)

4.6.2. Immunofluorescent Staining of Shootin-1 in DLKP Cell Lines

Immunofluorescence staining of Shootin-1 was carried out in DLKP and each of its clonal subpopulations to allow visualization of the distribution of the protein in each cell type, as well as a semi-quantitative assessment of protein abundances. Staining was carried out using the method described in Section 2.6.3 with resulting images shown in Figure 4.7. DLKP-SQ shows the most intense staining for Shootin-1 out of all cell lines. The staining is high in the cytoplasm and high areas of Shootin-1 can also be seen in what may be the leading edges of lamellipodia in some cells as described in a previous study (Shimada et al. 2008). In addition, there appears to be some perinuclear staining for Shootin-1 in the DLKP-SQ cells. DLKP-M has the next highest level of Shootin-1 staining. The morphology of DLKP-M results in many projecting structures similar to dendritic or axonal outgrowths, and staining of Shootin-1 is high in these structures. DLKP-I has the least intense staining, showing only diffuse low level Shootin-1 expression in the cytoplasm of the cells. DLKP shows intense staining of Shootin-1 in some subpopulations and none in others. Some of the stained cells resemble DLKP-SQ and DLKP-M as expected, with block like staining and outgrowth staining respectively. In addition, there are very large cells present staining very strongly for Shootin-1 which do not resemble any of the known clonal subpopulations and may represent a so far uncharacterised subpopulation within DLKP.

DLKP-SQ **DLKP-M**

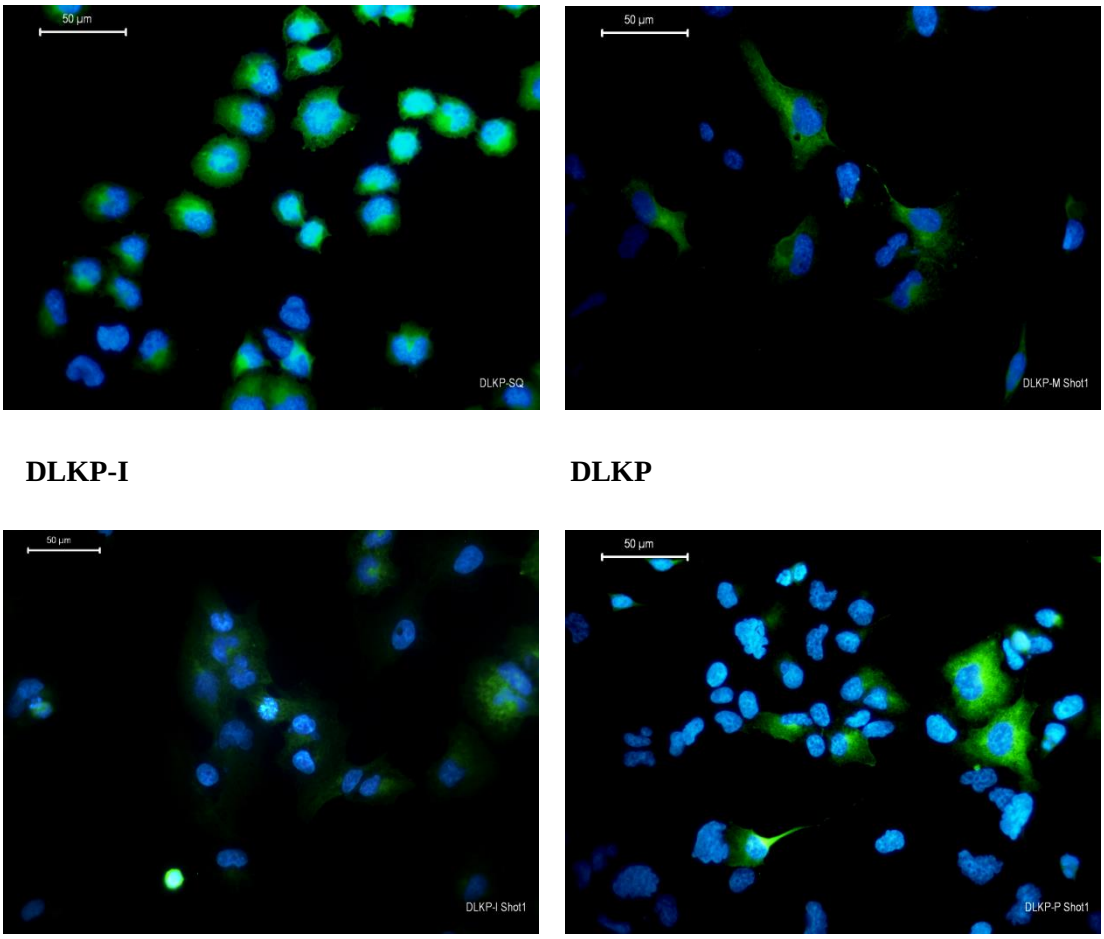


Figure 4.7: Immunofluorescence images staining for Shootin-1 in DLKP-SQ, DLKP-M, DLKP-I and DLKP. Nuclei are stained with Dapi. Negative controls with secondary antibody only did not result in staining (images not shown). Original magnification used on all micro-pictographs shown at 400x, 50µm scale bar. (n=2)

4.7. Functional Analysis of Shootin-1 Knockdown by RNAi

4.7.1. Transfection Optimisation

To investigate the role Shootin-1 plays in DLKP, RNA interference (RNAi) by transfection was chosen as a method to knockdown the expression of Shootin-1 in DLKP-SQ. This method is a relatively straightforward way of reducing the expression of a target protein and assessing the effects, using a panel of functional assays. Functional effects of protein knockdown can be determined by comparing the knockdown samples against the negative siRNA control. Here, a standard commercially available negative siRNA control is used that does not code for any known target sequence. This is used as a control to determine the effects of introducing siRNA into the cells. DLKP-SQ was chosen for Shootin-1 RNAi treatment as it has the highest expression levels of Shootin-1 at the protein and mRNA level. Before Shootin-1 knockdown was attempted, DLKP-SQ cells were first tested to determine the optimal transfection conditions required:

- The volume of transfection reagent necessary to effectively transfect DLKP-SQ was determined as described below.
- Next, the optimal concentration of siRNA was determined which significantly reduced expression of the target, yet did not over-saturate the cells with unnecessary siRNA.

4.7.1.1. Optimisation of Transfection Reagent Volume:

Transfection reagent is necessary to introduce siRNA into living mammalian cells using the method described in Section 2.4.1. In order to determine the most effective volume of transfection reagent to use for a particular cell type, a balance must be found between maximising transfection efficiency and limiting the toxic effects inherently associated with transfection reagent (Lipofectamine 2000). Rigorous optimisation is required in order to determine appropriate reagent volume. Transfection of DLKP-SQ cells was performed to test different volumes of transfection reagent in order to determine the optimal transfection reagent volume required. A proliferation assay subsequently performed using the acid phosphatase assay to determine cell viability (Section 2.5.1) is shown in Figure 4.8. Results are displayed as a percentage

survival value relative to the negative siRNA control (2 µl/well of transfection reagent). In all transfection conditions where siRNA is included, a 5 nM concentration is used. Four different transfection conditions were assayed in order to assess the optimal volume of Lipofectamine reagent to use for RNAi in DLKP-SQ: Cells-only; Lipo-only; Negative siRNA; Positive siRNA.

4.7.1.2. Transfection Reagent Optimisation

In order to focus in on the optimal transfection reagent volume to use for DLKP-SQ cells, 1 µl and 2 µl per well volumes of Lipofectamine per well in a 6-well plate were tested (after previous optimisation steps). Results are shown in Figure 4.8. A positive siRNA control is included which targets Kinesin, a motor protein required for the establishing of the bipolar spindle (Rapley et al. 2008)_. Inhibition of this protein arrests cells in mitosis, eventually leading to cell death. When compared to Cells-only, both 1 µl/well and 2 µl/well of transfection reagent show no effect on the proliferation of the Lipo-only control or the negative siRNA control. In contrast, the positive siRNA has a huge effect on the cell viability. A volume of 2 µl/well of transfection reagent effectively delivers positive control siRNA into the cells, resulting in a 90% reduction of cell viability compared to the negative siRNA control. This reagent volume shows no negative effects on cell viability and is twice as effective as a transfection reagent volume of 1 µl/well. Therefore, a volume was of 2 µl/well chosen as optimal for further transfections with DLKP-SQ.

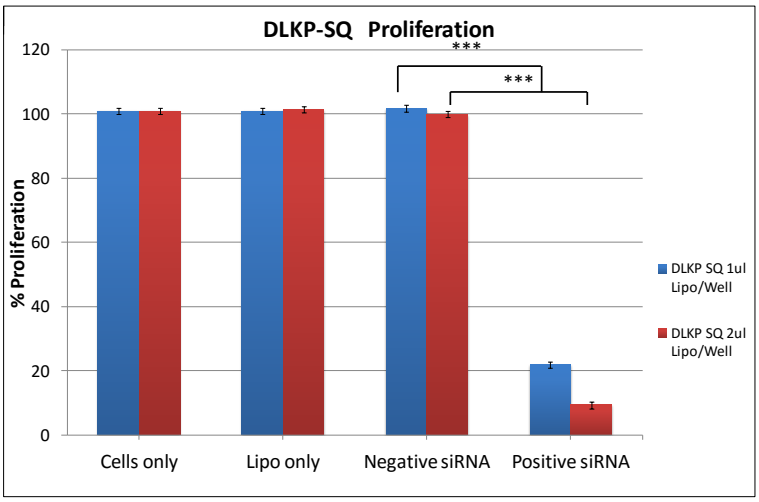


Figure 4.8: Proliferation assay performed on DLKP-SQ cells using 1µl/6-well (Blue) and 2µl/6-well (Red) of Lipofectamine transfection reagent. Percentage survival values are relative to the Negative siRNA control. (1µl /well: P-Value 1.8E-04) (2µl/well P-Value: 3.1E-05).

4.7.1.3. Optimisation of siRNA Concentration: Shootin-1

The optimization of Shootin-1 RNAi transfection conditions was finalized by establishing the optimal siRNA concentration to achieve maximum Shootin-1 protein knockdown. A transfection experiment was set up with a range of siRNA concentrations for Shootin-1 in DLKP-SQ. Transfection conditions were comprised of:

- Cell-only, Lipofectamine-only and Negative siRNA controls.
- 5 nM, 10 nM, 20 nM and 30 nM of siRNA were tested for Shootin-1.
- Two different siRNA molecules were tested in tandem which target Shootin-1.

Cell lysates were prepared from each transfection condition 72 hours post transfection using the method described in Section 2.3.1. Western blot analysis probed for Shootin-1 across all samples and the resulting blot image is shown in Figure 4.9. Two independent siRNA molecules were used to knockdown Shootin-1 expression at the four different concentrations described above. Bands for Shootin-1 are clearly visible at 100 kDa as expected in DLKP. A concentration of 30 nM siRNA was used in the negative siRNA sample to represent the maximum siRNA concentration in the Shootin-1 test panel. Based on the western blot results, 20 nM was determined as the optimal concentration of siRNA to use for Shootin-1 RNAi DLKP-SQ. This concentration shows total knockdown of the target, yet does not over-saturate the cells with unnecessary siRNA molecules.

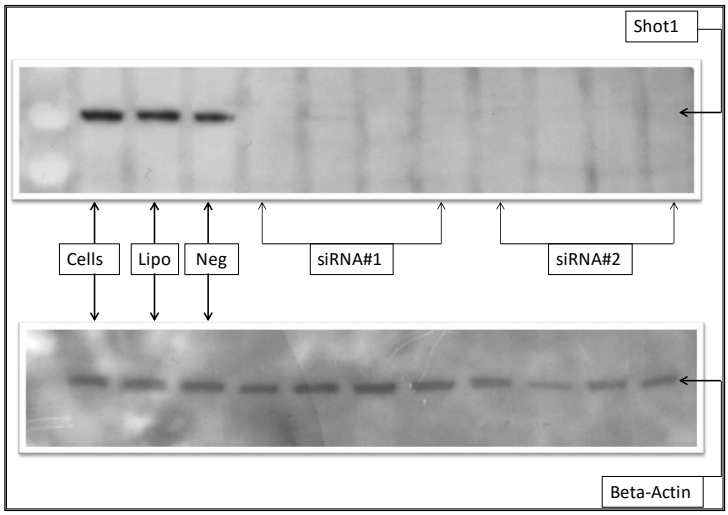


Figure 4.9: Western Blot probing for Shootin-1 (100kDa). Lanes 1-3 contain: Cells-only, Lipofectamine-only and Negative siRNA (30nM). Each Shootin-1 siRNA group (#1 and #2) is comprised of four sample lanes which were treated with 5 nM, 10 nM, 20 nM and 30 nM of siRNA from left to right. Beta-Actin loading control included.

4.7.2. Effects of Shootin-1 Knockdown on Proliferation, Migration and Invasion in DLKP-SQ

To investigate the potential role of Shootin-1 in DLKP-SQ, a panel of functional assays were carried out following siRNA transfection to knockdown the protein. Two individual siRNA molecules were used to knockdown Shootin-1 in DLKP-SQ using conditions optimised and described in Section 4.7.1. **Proliferation, migration** and **invasion** assays were carried out to investigate if a reduction in Shootin-1 expression resulted in a significant effect on the phenotype of DLKP-SQ.

4.7.2.1. Proliferation of DLKP-SQ after Shootin-1 Knockdown:

Proliferation of DLKP-SQ cells was measured 72 hours post transfection as described in Section 2.2.3. Knockdown of Shootin-1 by RNAi was found to have no significant effect on proliferation compared to the negative siRNA Control. Results are shown in Figure 4.10.

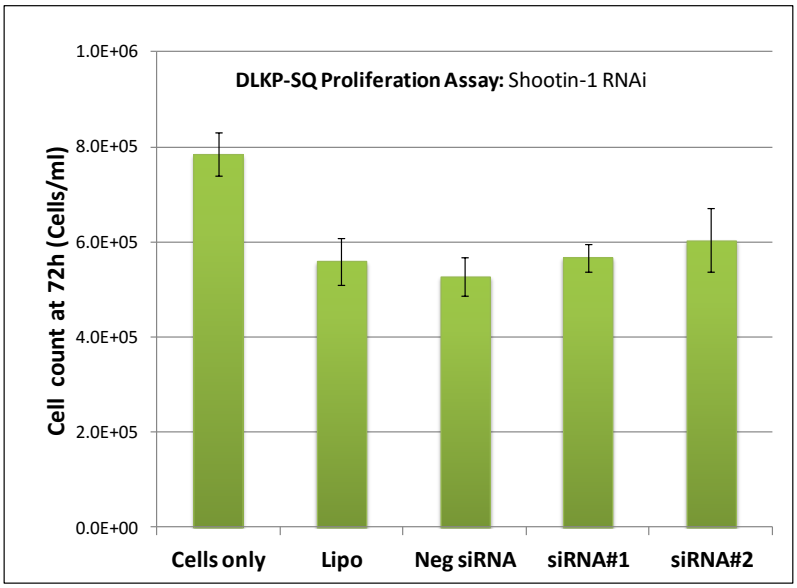


Figure 4.10: Proliferation assay on DLKP-SQ 72 hours post-transfection for Shootin-1 RNAi. There was no statistically significant difference in proliferation between siRNA knockdown samples (#1 and #2) and the Negative siRNA control (n=3). (**siRNA#1 P-value:** 0.23, **siRNA#2 P-value:** 0.16**).**

4.7.2.2. Migration of DLKP-SQ after Shootin-1 Knockdown:

Migration assays were performed 72 hours post transfection as described in Section 2.5.4. DLKP-SQ cells were allowed to pass through the uncoated membrane for 24 hours before being fixed and stained with crystal violet. A reduction in the migratory ability of DLKP-SQ cells was found when Shootin-1 knockdown samples were compared to the negative control. Migration levels are shown in Figure 4.11. Only one of the siRNA molecules show a significant reduction in migration; however the second molecule follows the same trend. This suggests that knockdown of Shootin-1 reduces the migratory ability of DLKP-SQ.

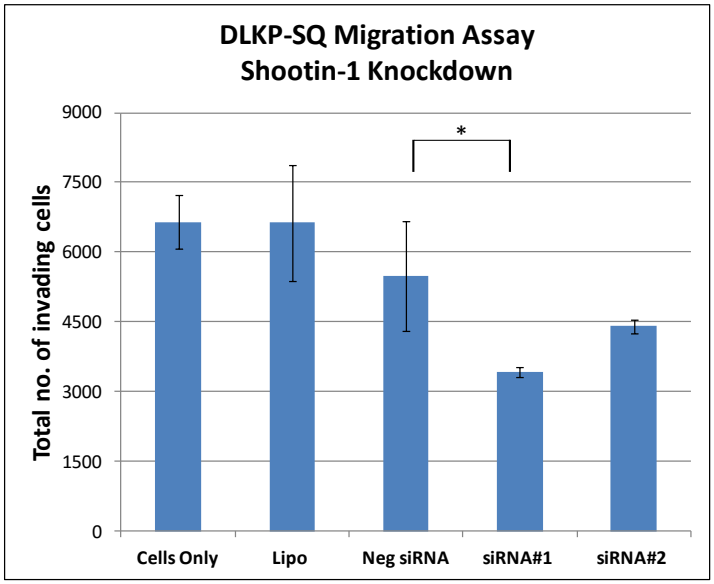


Figure 4.11: Migration assay results for DLKP-SQ cells following Shootin-1 knockdown with RNAi. Total number of cells which migrated through the uncoated membrane is shown (n=3). A statistically significant difference was found between knockdown siRNA sample #1 and the negative control. SiRNA #2 did not show a significant effect, but did show the same trend in reducing migration compared to the negative siRNA control. (siRNA#1 P-value: 0.04, siRNA#2 P-value: 0.19)

4.7.2.3. Invasion of DLKP-SQ after Shootin-1 Knockdown:

Invasion assays were performed 72 hours post transfection as described in Section 2.5.3. Cells which passed through the Matrigel coated membrane were counted and total numbers of invading cells per insert were calculated. No significant effect was found in the ability of DLKP-SQ to invade following Shootin-1 knockdown by RNAi when compared to the negative control. Invasion levels are shown in Figure 4.12.

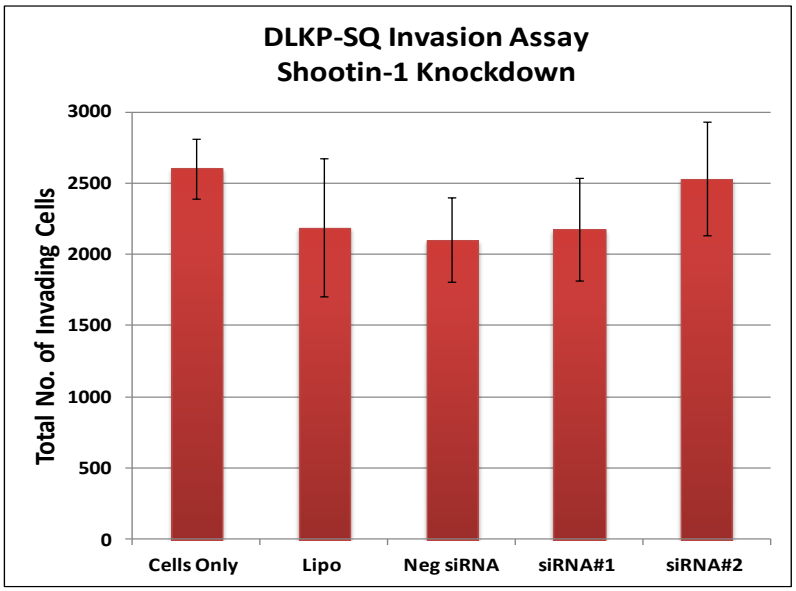


Figure 4.12: Invasion assay results of DLKP-SQ cells following Shootin-1 knockdown with RNAi. Total number of cells which invaded through the coated membrane is shown (n=3). There was no statistically significant difference between knockdown siRNA samples (#1 and #2) and the negative siRNA control (**siRNA#1 P-value:** 0.80, **siRNA#2 P-value:** 0.22).

4.7.3. Co-Staining of Shootin-1/F-Actin after Shootin-1 Knockdown by RNAi in DLKP-SQ

Literature studies have reported an association between Shootin-1 and Actin, particularly in the growth cones of neuronal cells. One research group sought to investigate the molecular mechanism linking cell adhesion molecules (CAMs) and Actin Retrograde Flow (ARF). They found Shootin-1 mediates the linkage between L1CAM and ARF in cultured rat hippocampal neurons (Shimada et al. 2008). Shootin-1 RNAi impaired this linkage, reducing L1CAM dependant axon outgrowth. At present, there are no reports of Shootin-1 expression in lung or lymph derived cell lines and therefore no links with Actin in these systems either. Due to these reported links between Shootin-1 and Actin mentioned above, it was decided to investigate this potential link in DLKP-SQ.

DLKP-SQ cells were stained simultaneously for Shootin-1 and Actin and imaged using confocal microscopy. Staining was performed after Shootin-1 expression was knocked down by RNAi. This was done to assess the effect of Shootin-1 reduction on Actin localization or staining intensity. In order to image the cells clearly, DLKP-SQ were grown on glass coverslips previous to Shootin-1 knockdown. Staining of Shootin-1 and Actin by confocal microscopy was performed after RNAi transfection. The transfection was performed on DLKP-SQ cells using the method as described in Section 2.4.2. To increase the likelihood of capturing motile cells mid-movement by confocal imaging, cells were scratched once with a P1000 tip 48 hours post transfection and washed with complete medium. Cells were fixed at 72 hours with 4% Paraformaldehyde before the staining process began. DLKP-SQ control and transfection samples were co-stained with Shootin-1, Actin and Dapi as described in Section 2.6.4.

Knockdown samples were compared to the controls: Cells-only, Lipo-only and negative siRNA. To rule out the possibility of non-specific staining, negative controls using secondary-only antibody treatment were imaged by confocal microscopy also. This resulted in no staining in DLKP-SQ cells, confirming that staining in the test samples is not due to sticky non-specific secondary binding (Images not shown). This experiment was performed in biological duplicate, and resulting images are shown in Figure 4.13 to Figure 4.17. Shootin-1 is referred to as ‘Shot1’ in the images.

DLKP-SQ: Cells Only

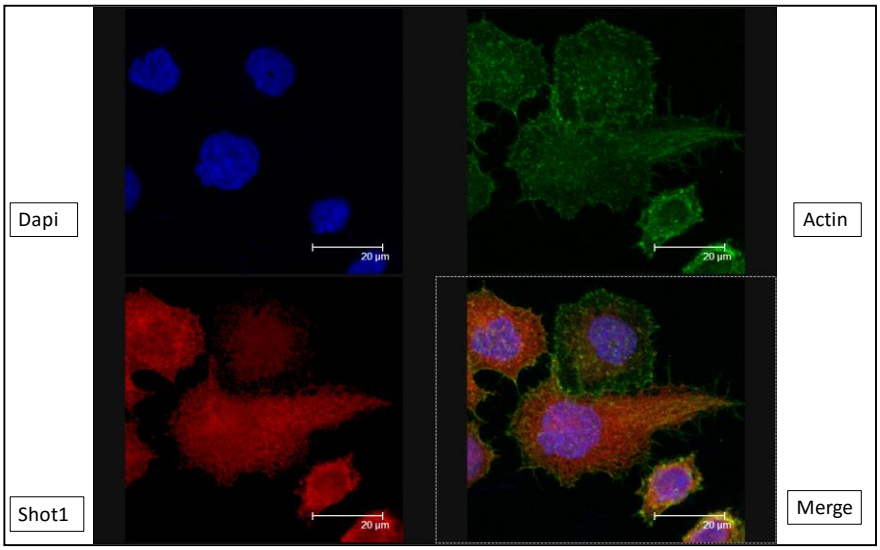


Figure 4.13: DLKP-SQ cells stained for Shootin-1, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are untreated by any RNAi method. Original magnification of all photomicrographs, ×400, scale bar = 20 µm.

DLKP-SQ: Lipofectamine-only Control

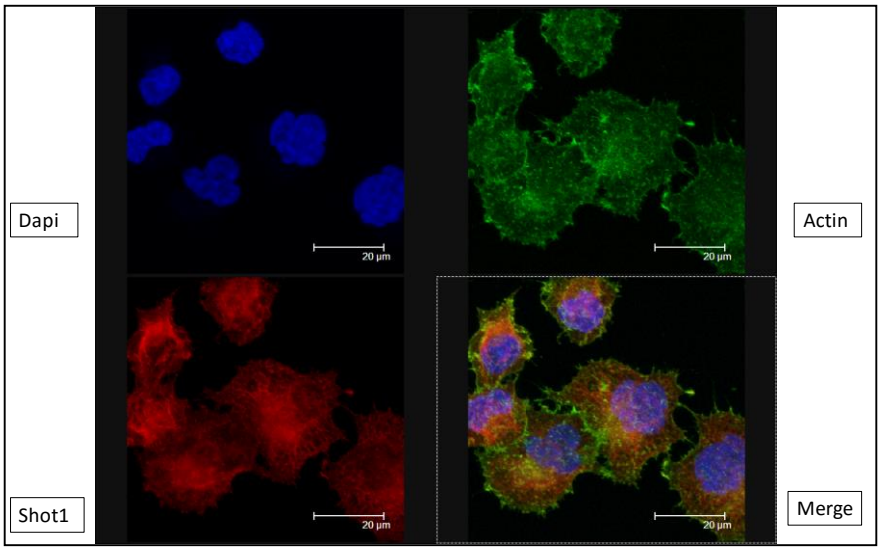


Figure 4.14: DLKP-SQ cells stained for Shootin-1, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are treated with transfection reagent only (Lipofectamine). Original magnification of all photomicrographs, ×400, scale bar = 20 µm.

DLKP-SQ: Negative siRNA Control

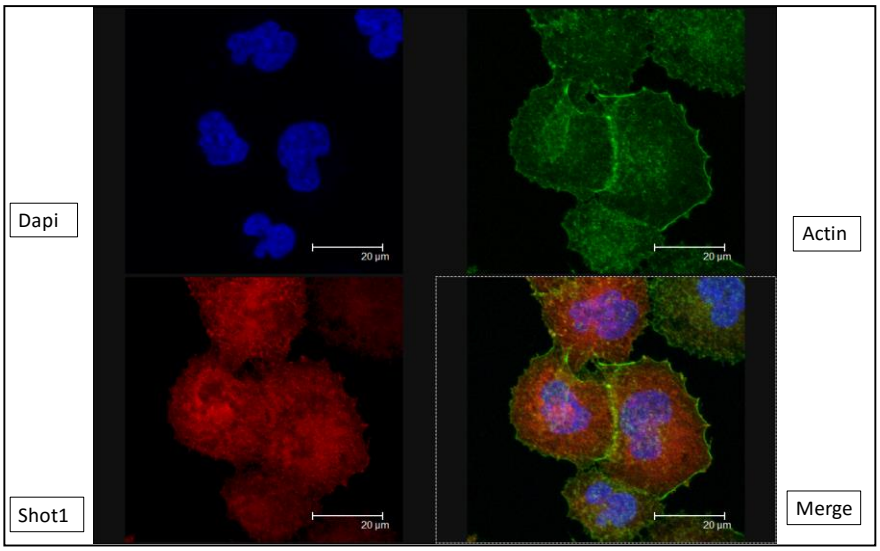


Figure 4.15: DLKP-SQ cells stained for Shootin-1, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are treated with transfection reagent and negative siRNA control. Original magnification of all photomicrographs, ×400, scale bar = 20 µm.

DLKP-SQ: Shootin-1 Knockdown- siRNA #1

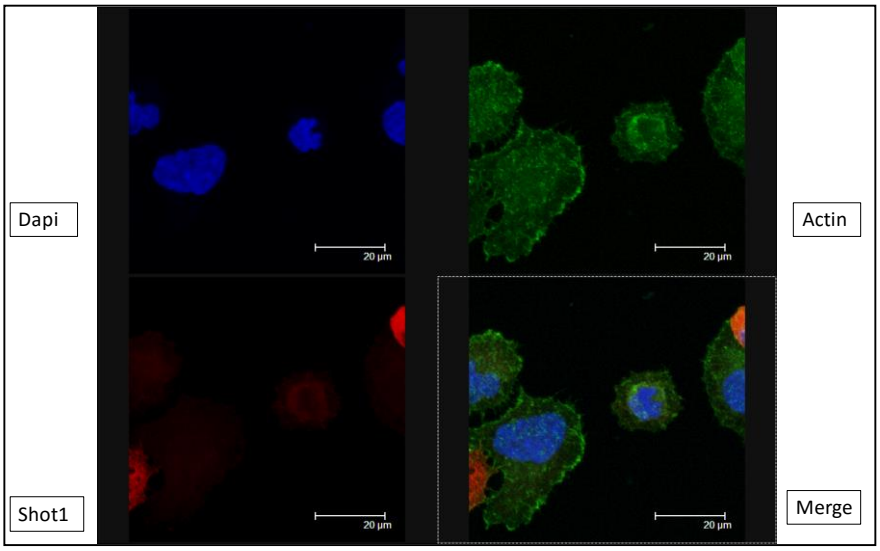


Figure 4.16: DLKP-SQ cells stained for Shootin-1, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are treated with siRNA#1 for Shootin-1. Original magnification of all photomicrographs, ×400, scale bar = 20 µm.

DLKP-SQ: Shootin-1 Knockdown- siRNA #2

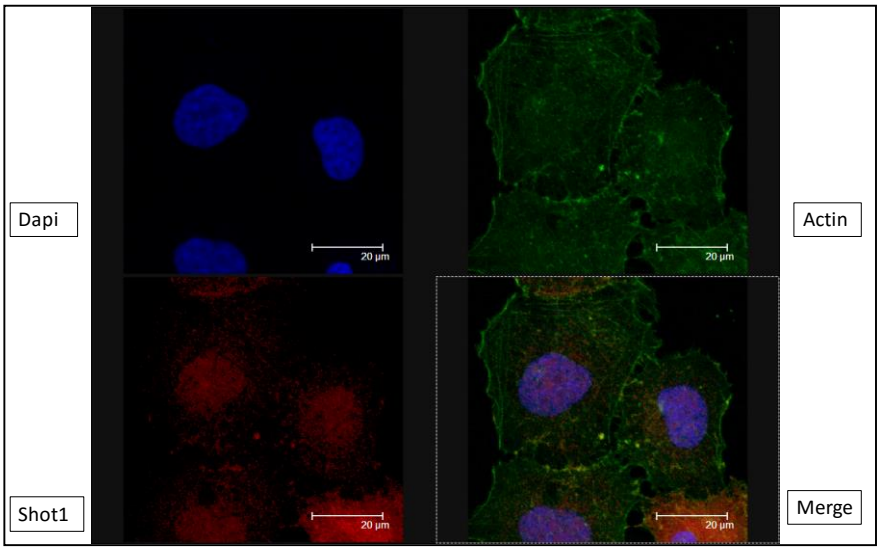


Figure 4.17: DLKP-SQ cells stained for Shootin-1, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are treated with siRNA#2 for Shootin-1. Original magnification of all photomicrographs, ×400, scale bar = 20 µm.

Negative Control: High Magnification

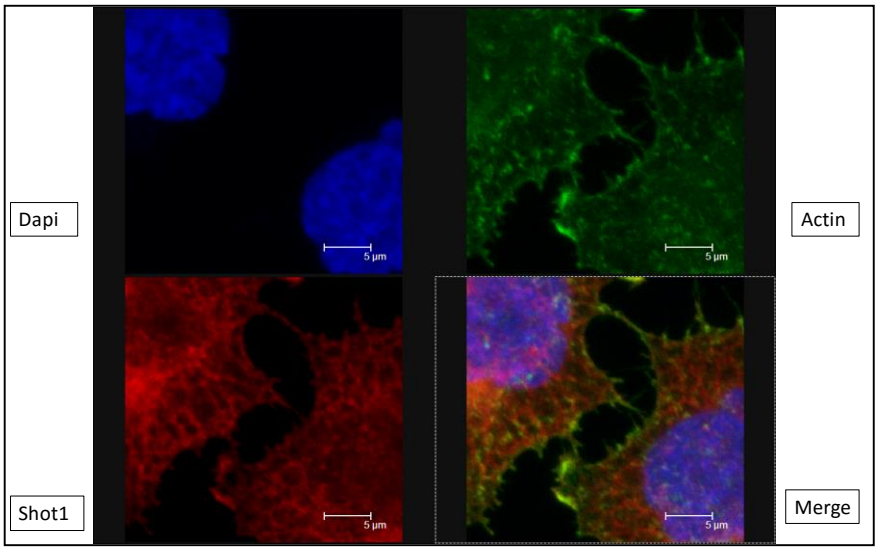


Figure 4.18: DLKP-SQ cells stained for Shootin-1, Actin and Dapi. A merged image is shown in the bottom right corner. DLKP-SQ cells at high magnification showing co-localization of Shootin-1 and Actin in cell projections. Original magnification of all photomicrographs, ×800, scale bar = 5µm.

Staining of Shootin-1 in DLKP-SQ (coloured red) shows a similar localization of the protein as found in the immunofluorescence work described in Section 4.6 for the control group. In the control group cells, there is high level staining of Shootin-1 visible throughout the cytoplasm with strong points of intense staining visible at the edges of the cells. The Dapi nuclear stain (coloured blue) is still visible in the merged photograph, which suggests low level staining for Shootin-1 in the nucleus or nuclear envelope of DLKP-SQ cells. Actin staining is high in DLKP-SQ (coloured green) showing high cytoplasmic staining throughout the cells as well as at the edges. The merged photographs reveal co-localization of Shootin-1 and Actin, particularly in Figure 4.14. This image shows most clearly an overlapping staining pattern for both proteins, notably in the projections emanating from the cells showing intense staining at the tips.

The confocal imaging of Shootin-1 also shows successful RNAi knockdown of the protein expression in Figure 4.16 and Figure 4.17. The two samples were treated with siRNA #1 and #2 for Shootin-1 and there is a large reduction in the levels of staining for Shootin-1 in both cases. Actin staining seems to be unaffected by this treatment, which indicates that Shootin-1 is neither involved in the expression levels of actin, nor its localization in DLKP-SQ cells. This finding somewhat corroborated by the use of beta-actin as a loading control in the western blot shown in Figure 4.9. Although the expression levels of actin were not expected to change upon knockdown of Shootin-1, it was thought that it could have had an effect on actin localization. The confocal imaging suggests that this is not the case in DLKP-SQ.

Shown in Figure 4.18 is a high magnification image of the negative control sample which clearly shows co-localization of Shootin-1 and Actin. This co-localization is visible throughout the cell, but especially on the protruding outgrowths which link neighbouring cells together. Shootin-1 and Actin can clearly be seen dotted along these filamentous outgrowths, suggesting a possible connection between the two proteins in DLKP-SQ.

4.8. Identification of Potential Binding Partners of Shootin-1

To shed some light on the role of Shootin-1 in DLKP-SQ, a co-immunoprecipitation (Co-IP) of Shootin-1 was performed on DLKP-SQ cells with a view to extracting potential binding partners with the target. Co-IP uses antibodies specific to the target to form an immune complex with the protein of interest which can be captured on a beaded support column. Along with the target itself, other macromolecules which may be bound to the target by native interactions will also be captured. This form of affinity chromatography can be an effective method of discovering binding partners of the target protein. Two strategies were attempted to Co-Immunoprecipitate Shootin-1 and potential binding partners: **Cross-linked Co-IP** and **Traditional Co-IP.**

4.8.1. Cross Linked Co-Immunoprecipitation:

Initially, Co-IP employing cross-linked antibody technology was used to immunoprecipitate Shootin-1 and any potential binding partners from DLKP-SQ cell lysate. This method leaves the antibody specific to the target covalently bound to the bead column after elution of the bound protein complexes. Protein elution is carried out using a low pH buffer to preserve the activity of the antibody for the protein of interest. The advantage of this method is once the eluted proteins are separated using SDS-PAGE, no antibody fragments (heavy and light chain fragments) co-elute in the sample. This results in easier identification of captured proteins in the sample by MS or western blot, as they are not masked by the presence of relatively large abundances of antibody fragments. Also, the column can be reused for further Co-Immunoprecipitation experiments. Rabbit IgG antibody was used as a negative control in parallel to account for non-specific binding of protein to the target antibody isotype. Unfortunately this method was unsuccessful at purifying Shootin-1 by immunoprecipitation as shown in Figure 4.19.

Western blot analysis of Shootin-1 for the Co-IP sample set is shown in Figure 4.19. Shootin-1 is visible at 100 kDa in the positive control lysate, showing the protein can be detected when DLKP-SQ cells have been lysed using the IP compatible lysis buffer. For both Shootin-1 and rabbit IgG control samples, double elutions were

performed in order to retrieve the maximum amount of bound protein from each column. Elutions of Shootin-1 samples do not contain the protein (Lanes 2 and 3) and only heavy chain antibody fragments are visible. Shootin-1 can be clearly seen in the unbound fraction (Lane 4), suggesting Shootin-1 was not captured by the cross-linked antibody technology. Rabbit IgG control sample elutions (Lanes 5 and 6) show extremely strong heavy and light chain antibody fragments as well as high background. This is unusual as the same concentration of antibody was used for both test and control. Shootin-1 is visible once again in the unbound fraction of this control group (Lane 7). These results show the failure of this technique on two fronts. The target protein was not captured, and heavy and light chain antibody fragments were eluted from the beads even after cross-linking.

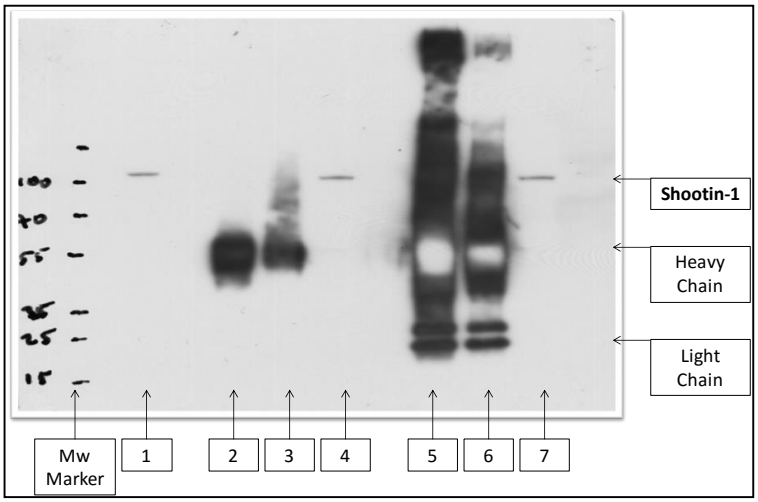


Figure 4.19: Western Blot probing for Shootin-1 in DLKP-SQ samples after cross linked Co-IP. **Lane 1:** DLKP-SQ lysate control. **Lanes 2 and 3:** First and second elution from Shootin-1 Co-IP. **Lane 4:** Shootin-1 Co-IP unbound fraction. **Lanes 5 and 6:** First and second elution from Rabbit IgG control. **Lane 7:** Rabbit IgG unbound fraction. Shootin-1 is visible at 100kDa. Heavy and light antibody chains are prominent.

4.8.2. Traditional Co-Immunoprecipitation: Shootin-1

This method differs from Cross-Linked IP in that the target antibody is not covalently bound to the Protein-G bead column. During elution of the target protein and any binding partners, the antibody is also released from the bead support. Because of this, the heavy and light chains of the denatured antibody are present when the sample is separated by SDS-PAGE, and mask the detection of proteins by western blot or MS. However, it is an effective and simple method of performing a Co-IP, particularly if the protein of interest does not have a similar Mw to the antibody fragments.

Shootin-1 was immunoprecipitated from DLKP-SQ and analysed by western blot. This analysis was unable to detect the protein due to huge interference by the antibody fragments present in the elution samples. Fragments of antibody were detected in the entire length of the lane, thereby masking detection of any specific protein (image not shown). However, a coomassie stained gel on which the test and control elutions were separated by SDS-PAGE revealed a weak but obvious band at 100 kDa in the Shootin-1 elution lane which was not present in the rabbit IgG control elution shown in Figure 4.20. Each lane was excised from the gel and sliced into 2 mm fragments along its entire length. In-gel digestion was performed on the gel pieces and proteins were identified using LC-MS/MS using methods described in Section 2.8.6. MS analysis was successful in detecting Shootin-1 and potential binding partners in two independent traditional Co-IP experiments with results summarised in Table 4.2.

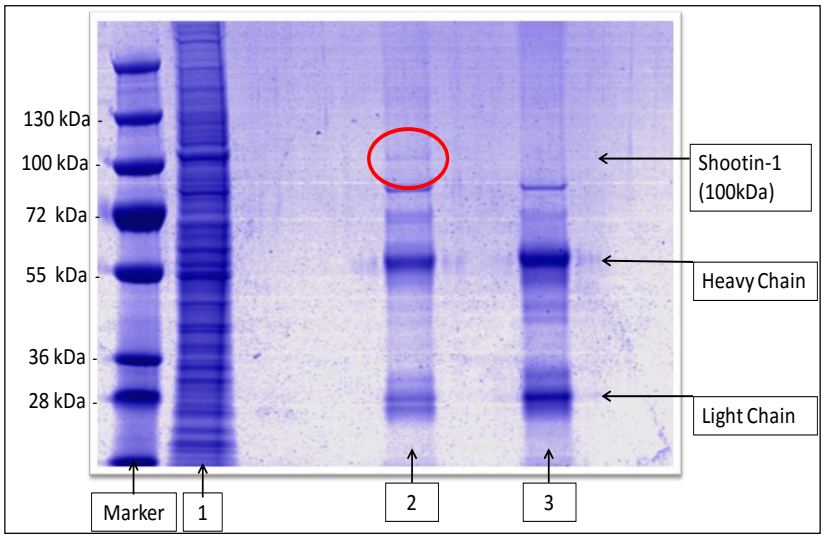


Figure 4.20: Coomassie stained gel of Shootin-1 Co-IP samples. **Lane 1:** DLKP-SQ positive control lysate. **Lane 2:** Shootin-1 Co-IP elution. **Lane 3:** Negative control Rb IgG to account for non-specific binding events. Shootin-1 band is circled in red at 100kDa in lane 2. Heavy and light antibody chains are clearly visible.

Proteins identified as present in the negative control (Rabbit IgG) lane were eliminated from the list of proteins identified in the Shootin-1 immunoprecipitate. Results from both replicate experiments were overlapped and common proteins are shown in Table 4.2. The sequence coverage for each protein identified by MS is shown for both experiments.

Shootin-1 was the most abundant unique protein in each experiment, with sequence coverage of 54% and 36% respectively. The majority of peptides which contributed to this identification were obtained from the gel slice circled in red in Figure 4.20. This finding confirms Shootin-1 as being expressed at 100kDa in DLKP. Even though the western blot detection of Shootin-1 was made impossible by the presence of antibody fragments, MS detection was used to show its presence at the expected weight in the immunoprecipitation. This indicates that successful immunoprecipitation of Shootin-1 and potential binding partners was achieved.

A number of proteins were found to co-immunoprecipitate with Shootin-1 in both replicate experiments. This panel of proteins (Table 4.2) were not present in the negative control, ruling out non-specific binding as the reason for their capture. They

also passed all the stringent criteria associated with MS analysis in this experiment as described in Section 2.7.6.2, and were therefore followed up by western blot analysis.

Protein	Sequence Coverage Exp 1	Sequence Coverage Exp 2
Shootin-1	54%	36%
Small nuclear ribonucleoprotein Sm D3	25%	8%
60S ribosomal protein L12	24%	9%
Striatin-3	6%	2%
Src substrate cortactin	6%	2%
Semenogelin-1	5%	10%

Table 4.2: Mass spec identifications showing immunoprecipitated Shootin-1 protein, with co-eluted proteins from two replicate experiments. Percentage of each protein sequence identified by MS is shown for each experiment.

4.8.3. Western Blot Analysis of Shootin-1 Binding Partners in DLKP

Western Blot analysis was used to validate the presence of three of the potential Shootin-1 binding partners. Unfortunately, this could not be done on the Shootin-1 immunoprecipitate elution, as the antibody fragment problem prevents detection of specific proteins due to masking effects. The analysis was therefore performed on DLKP-SQ cell lysate to ensure the proteins are being expressed by the cell line. Alongside this, the protein panel was also analysed by western blot in the remaining DLKP cell lines as well as the combination lines in order to establish their expression pattern in the DLKP cell line model. This information may correlate with the known phenotypes of the clones and shed light on role in this system. The three proteins chosen for validations are: **Semenogelin-1**, **Striatin-3** and **Src Substrate Cortactin**.

4.8.3.1. Semenogelin-1:

Semenogelin-1 (SEMG1) is a secreted protein usually expressed in the seminal vesicles and urinary bladder with little to no expression in other tissues of the body (Bjartell et al. 1996). However, it has been shown to be expressed very specifically in the cerebellum and only in Purkinje cells located there. Semenogelin1 has been found to be ectopically expressed in small cell lung cancer (SCLC) and expression is not limited to males in these cases. The protein has been found to be secreted into growth medium by SCLC cell lines and was proposed as a useful biomarker for the disease (Rodrigues et al. 2001).

SEMG1 was successfully detected by western blot in all DLKP and subclonal cell lines. Looking at DLKP and the clones only (Figure 4.21), highest expression of SEMG1 is in DLKP-M which shows a much higher abundance of this protein when compared to all other samples. The next highest expression level is in DLKP-I followed by DLKP-SQ and DLKP. The combination lines show unexpected expression patterns. The abundance of SEMG1 is higher in SQ+M than in I+M, which is not what would be the case if the band intensities from the clone constituents are added together. It is therefore possible that DLKP-I has an inhibitory effect on the expression of SEMG1. The DLKP has a lower abundance of SEMG1 compared to the ‘Remade’ cell line. The ‘Remade’ cell line was created to reform DLKP in the proportions found previously

(McBride et al. 1998b). In the case of SEMG1 the abundance level of the protein is not equal between the two, suggesting the presence of other cell types in the DLKP cell line. SEMG1 was not found to be a significantly differentially expressed protein in the label-free LC-MS proteomic analysis.

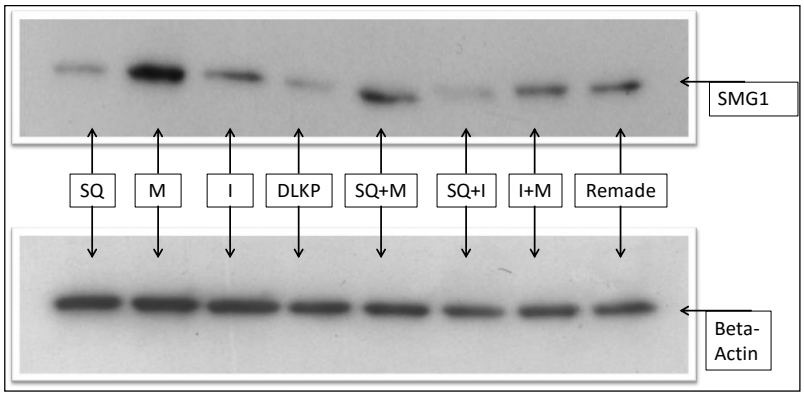


Figure 4.21: Western Blot probing for Semenogelin-1 in DLKP and clonal subpopulation lines. Combination lines are included also. Bands are visible at the expected weight of ~52kDa. Beta-Actin is used as a loading control.

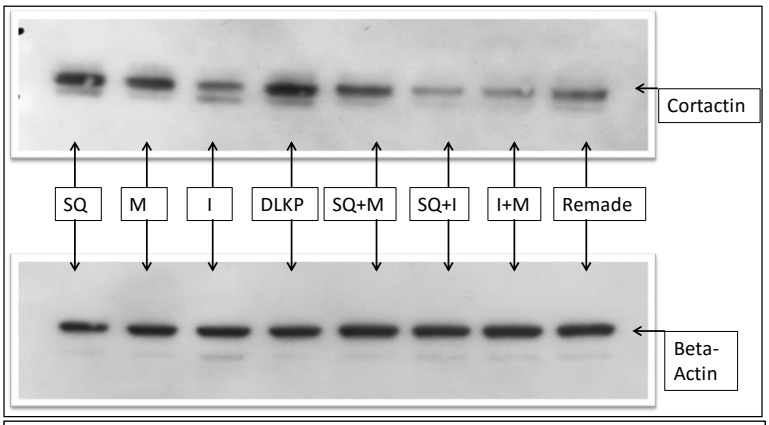
4.8.3.2. Src Substrate Cortactin

Src Substrate Cortactin (CCTN) is a cytoplasmic protein involved in actin polymerization and rearrangement of the actin cytoskeleton (Wei et al. 2014). Cortactin-assisted actin branching is most prominent around the outside of the cell and occours when cortactin has been phosphorylated. The protein then stabilises the ARP2/3 complex which provides more nucleation sites for actin branching, resulting in enhanced formation of lamellipodia and cell spreading (Uruno et al. 2001, Weed and Parsons 2001). CCTN has been found to be overexpressed in malignant tumours in oesophageal squamous cell carcinoma and has been found to be a prognostic marker for poor survival in colorectal adenocarcinoma (Kim et al. 2012, Luo et al. 2006).

CCTN was validated as being expressed in the DLKP cell line model by western blot analysis (Figure 4.22A). DLKP and DLKP-SQ have the highest abundance of the protein, followed next by DLKP-M and then DLKP-I. Combination lines showed unexpected expression patterns. SQ+M has a lower abundance of CCTN than the two individual clones which combine to make it. This suggests an inhibitory effect on the expression of CCTN when the two cell types are grown in co-culture. A similar inhibitory effect can be seen on the other combinations SQ+I and I+M, which show much lower abundances of CCTN that would be expected. DLKP and Remade also show different levels of CCTN expression, suggesting an expression difference between them.

CCTN was also discovered by the label-free analysis as a differentially expressed protein (Figure 4.22B). It has a maximum fold change of **1.79** between DLKP-SQ and DLKP-I (ANOVA, p= 1.73E-05) and 4 peptides contributed to its identification. With a fold change <2, it did not meet with the criteria necessary to be included in the results described in Section 3.1.2. However, the fold change is still significant and the data can be used to assess protein expression patterns in the DLKP cell line model. In the case of CTTN, there is a good correlation between the label-free data, and that discovered by western blot analysis. In both cases, DLKP-SQ and DLKP showed the highest levels of CTTN abundance. CCTN is involved in actin polymerisation at the leading edges of cells. Immunofluorescence staining showed localization of Shootin-1 at these areas also, which suggests an interaction between the two proteins, however further work is necessary to show this.

(A)



(B)

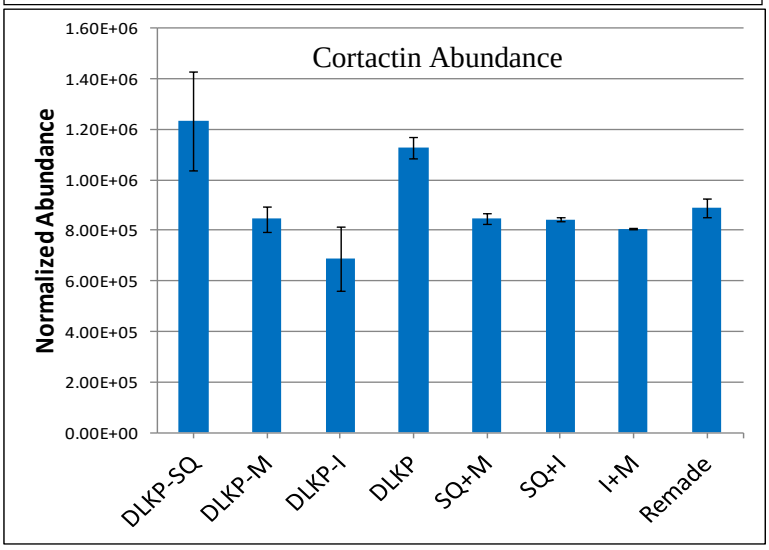


Figure 4.22: (A) Western blot probing for Src Substrate Cortactin. Bands are visible at the expected weight of ~61kDa. Beta-actin is used as a loading control. **(B)** Label-Free analysis result for Src Substrate Cortactin.

4.8.3.3. Striatin-3

Striatin-3 (STRN3) is a member of the striatin family of proteins which are found to be mostly expressed in the dendritic spines of neurons in the central nervous system. They contain multiple protein-binding domains such as caveolin and coiled coil-binding domains. They are also known to bind calmodulin and Ca^{2+} (Bartoli et al. 1999, Benoist, Gaillard and Castets 2006). Expressed in a variety of cancer cell lines and tissues, striatin family members are reported to be involved in cell migration and invasion. Knockdown of a striatin family member has been shown to reduce the metastatic phenotype of oesophageal and ovarian cancer cell lines (Wong et al. 2014).

STRN3 was identified as a potential binding partner of Shootin-1 by Co-IP. Western blot validation of STRN3 revealed its expression profile in the DLKP cell line model (Figure 4.23). Lowest expression of STRN3 is seen in DLKP-SQ (the least invasive cell line). Highest expression is in DLKP-M (the most invasive cell line). Similar levels are in both DLKP-I and DLKP with the latter showing slightly less abundance. In the case of the combination lines, the level of STRN3 expression in both of the cell lines containing DLKP-M is higher than expected. Abundance levels for the protein remained similar to the undiluted DLKP-M clone. This suggests that STRN3 expression in DLKP-M is unaffected by co-culture with DLKP-I or DLKP-SQ, with abnormally high abundance levels of the protein visible in SQ+M and I+M. Expression levels of this protein correlate with the known phenotypes of the DLKP clones with regards to invasion and migration. Highest expression is visible in DLKP-M while lowest expression is in DLKP-SQ. Further work is needed to determine if STRN-3 is involved with the invasive phenotype of DLKP-M. STRN-3 was not found as a differentially expressed protein in any of the label-free quantitative analyses.

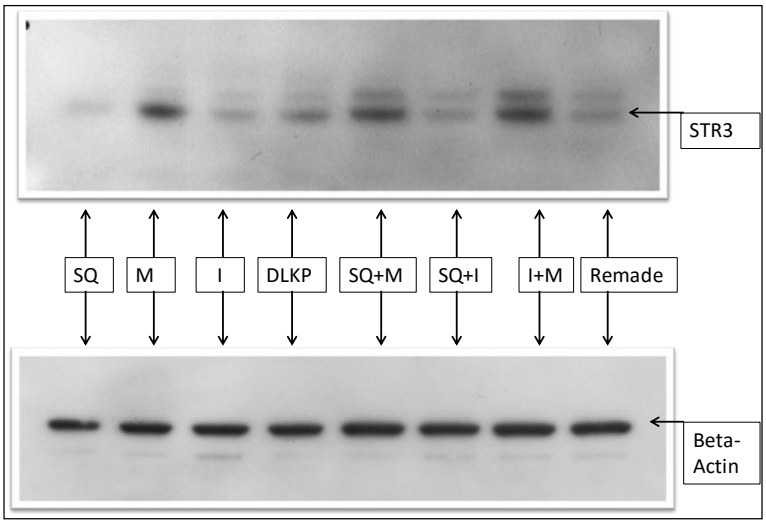


Figure 4.23: Western blot probing for Striatin-3. Bands are visible at the expected weight of ~87kDa. Beta-Actin is used as a loading control.

4.9. Over-Expression of Shootin-1 in DLKP and Clonal Subpopulations

To further investigate the role of Shootin-1 in DLKP, it was decided that over-expression of the protein in the DLKP clones may shed light on its function in the cell line model. If successful over-expression was achieved, functional assays could be performed to assess the effects of Shootin-1 over-expression. In order to create an over-expression vector, the approach taken uses cDNA generated from DLKP-SQ RNA as a template to amplify the Shootin-1 open reading frame (ORF) sequence. This sequence is then ligated into a vector and can be transfected into each clonal subpopulation to promote gene over-expression. This method is cost effective and uses the endogenous sequence for Shootin-1 being expressed by DLKP-SQ. The first step is to design primers which successfully amplify the target sequence for Shootin-1. Next, that sequence is cloned into an over-expression vector and transformed into competent bacteria for amplification. Following this, the vector is extracted and digested with restriction enzymes to ensure the correct size fragment has been incorporated into the vector. Finally, the potential Shootin-1 fragment is sequenced to ensure complete homology to the known open reading frame sequence.

4.9.1. PCR Amplification of the Shootin-1 Coding Sequence from DLKP-SQ

Firstly, the DLKP-SQ cell line was used as a basis for acquiring the Shootin-1 coding sequence as it is expressed at the highest abundance in that cell line at the protein and mRNA levels. Primers were designed which flank either end of the Shootin-1 coding sequence and also incorporate different restriction enzyme sites at their 5 prime ends. Restriction enzyme sites flanking Shootin-1 were chosen based on their presence and position in the vector (pcDNA 3.1 Hygro+) and were designed into the primers to ensure the resulting amplicon would be oriented correctly in the vector after ligation. Restriction enzyme sites used were KPN1 and XHO1.

Following extensive optimisation, a PCR was carried out to amplify the Shootin-1 coding sequence. The starting material for this process was cDNA generated from DLKP-SQ RNA. Synthesis of cDNA was carried out using a standard reverse

transcription kit with random primers as described in Section 2.9.3. To amplify the Shootin-1 sequence by PCR, multiple different annealing temperatures were tested using the Shootin-1 primers and resulting DNA fragments were separated by agarose gel electrophoresis and visualised using SafeView (Figure 4.24A). In the final panel tested, 2 samples resulted in faint bands at the expected weight of ~1950bp PCR products were cleaned up using a PCR purification kit to remove small DNA fragments (primers) and incompatible buffers from the PCR reaction. Following this, the PCR product was digested using KPN1 and XHO1 to create the sticky ends necessary for ligation into the pcDNA3.1 vector. In tandem, the vector was also digested with the same restriction enzymes to linearize it and prepare it for ligation (Figure 4.24B).

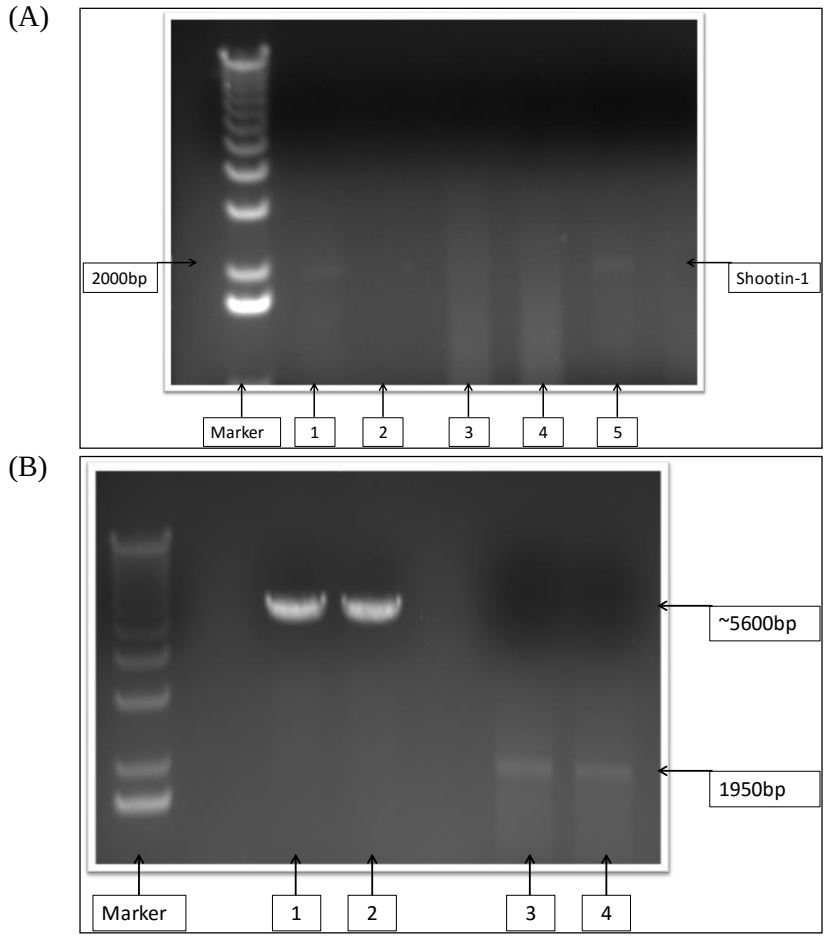


Figure 4.24: (A) PCR amplification of the Shootin-1 coding sequence. Five samples at different annealing temperatures were attempted (1-5). Samples 1 and 5 show faint bands at the expected weight of 1950bp. **(B)** Samples are separated on an agarose gel after digestion by KPN1 and Xho1. The pcDNA3.1 vector has been linearized by the digestion process (Lanes 1 and 2), and Shootin-1 sequence can be assumed to be digested also (Lanes 3 and 4).

4.9.2. Ligation of Shootin-1 Fragment and Over-Expression Vector

The vector and Shootin-1 fragments were ligated together overnight at 16°C using a Roche Rapid Ligation kit. The following day, 7 µl of ligation mix was transformed into DH5α *E. coli* and plated on ampicillin LB agar plates. Four colonies were picked and miniprepped to extract the vector. Extracted vectors from all picked colonies were quantified and digested using KPN1 and Xho1 to determine if any contained the insert of expected size (1950 bp). The resulting product was visualized on an agarose gel. Four colonies were picked and mini-prepped, and these correspond to the last four lanes in Figure 4.25. The image shows that sample 3 has a secondary band at the expected size of ~1950bp, which suggests the Shootin-1 fragment has been successfully incorporated into the vector backbone.

Sample 3 was sent for sequencing to ensure the fragment which has been inserted is indeed the coding sequence for Shootin-1. Primers were designed to allow sequencing of the entire length of the Shootin-1 ORF. Sequencing was completed and showed 100% homology to the Shootin-1. Once this step was completed, the vector was the midi-prepped to scale it up and remove endo-toxins so it can be used for transfection and ultimately, overexpression of Shootin-1.

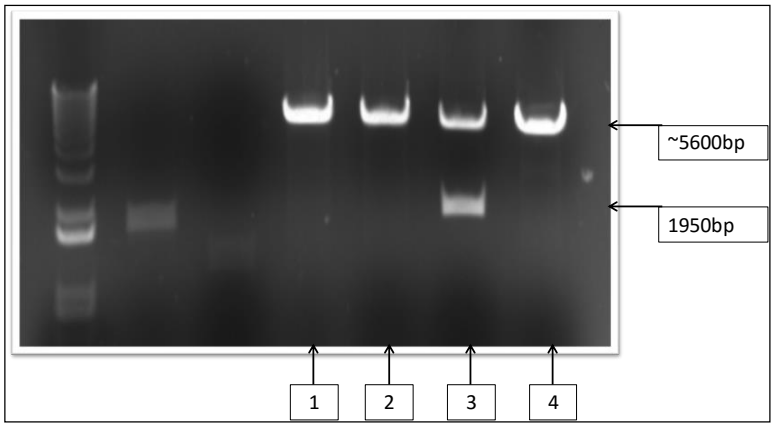


Figure 4.25: Digested vectors from positive bacterial colonies. Four were picked and digested with KPN1 and XHO1. Sample 3 resulted in the expected weights of 1950 bp for Shootin-1 and 5600bp for the vector backbone.

4.9.3. Shootin-1 Overexpression

To assess the role Shootin-1 may play in the DLKP cell line model, DLKP-SQ, DLKP-M and DLKP-I were transfected with the newly created vector with the goal of overexpressing the protein and following up with functional assays. Transfection was carried out and cells were placed under antibiotic selection (Hygromycin). Antibiotic concentration was determined by setting up a kill curve. Each clone was transfected with both the Shootin-1 overexpression vector (O.E), and an empty vector control (E.V) to generate cell lines for comparison. Untreated cells which were not transfected at all are included as a control. Results of western blot analysis analysing Shootin-1 expression are shown in Figure 4.26.

DLKP-SQ showed no change in expression for Shootin-1 in any of the three samples which were analysed by western blot. This suggests the maximum level of Shootin-1 is currently being expressed by DLKP-SQ and post translational regulation may be prohibiting over-expression. DLKP-M samples showed low level Shootin-1 expression in each of the three samples. The over-expression sample shows a higher level of Shootin-1 expression compared to the empty vector and control samples, suggesting over-expression has been achieved. DLKP-I samples showed an increase in Shootin-1 expression in the overexpression sample compared to both the empty vector and untreated cells control also. This western blot has only been carried out once and will need to be repeated to verify the findings. Functional assays will then be carried out to look for phenotypic changes related to the overexpression of Shootin-1. Experiments would include investigating the migratory and invasive abilities of each clone when Shootin-1 is being overexpressed. A quantitative label-free proteomic analysis could also be performed, comparing Shootin-1 overexpressing clones to their normal and empty vector counterparts, This could highlight proteins which are affected by changes in Shootin-1 expression, potentially revealing new interacting partners.

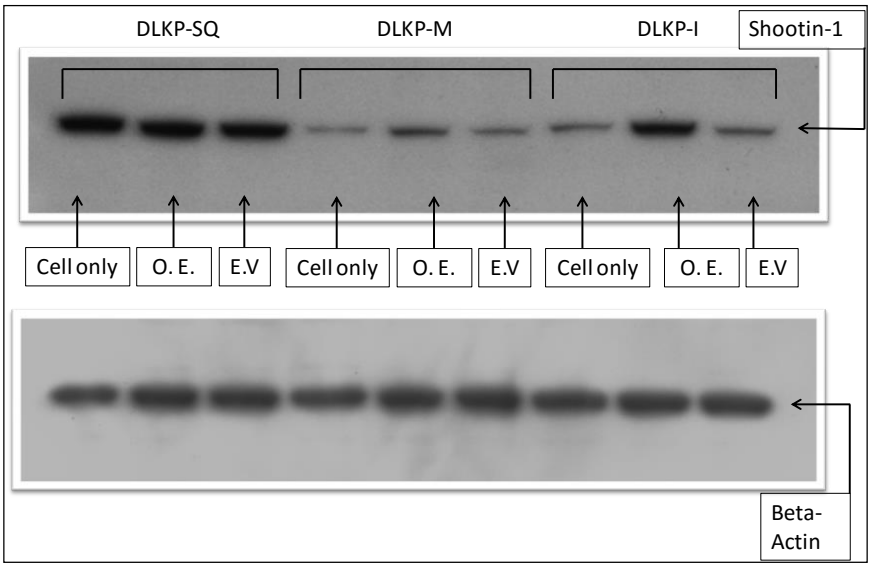


Figure 4.26: Western blot probed for Shootin-1. Each DLKP clone was transfected with Shootin-1 over expressing vector (O.E) and an empty vector control (E.V). Untreated cells and Beta-actin loading control are included also.

4.10. Shootin-1 Expression in a Panel of Cell Lines

Western blot analysis was used to determine the presence and expression levels of Shootin-1 in a panel of cell lines. Results are shown in Figure 4.27

- Pancreatic cell lines: PANC-1, AsPC-1, BxPC-3 and MIA PaCa 2.
- Lung cancer cell lines: A549, SK-LU-1, and NCI-H82.
- DLKP clones: DLKP-SQ, DLKP-M and DLKP-I.

Western blot analysis of Shootin-1 shows the expression levels of the protein in the panel of pancreatic and lung cancer cell lines mentioned above (Figure 4.27). Shootin-1 has its characteristic expression pattern in the DLKP clones, with very high abundance in DLKP-SQ compared to the others. In the pancreatic panel, the two with the most invasive phenotypes (PANC-1 and BxPC-3) show the highest expression of Shootin-1. AsPC-1 has low expression of Shootin-1 while Mia PaCa 2 had very little to none at all. Both these cell lines are known to be lowly invasive from preliminary studies carried out in the NICB by others. Regarding the lung cancer cell lines, expression of Shootin-1 was only found in A549, with no detectable levels of the protein in either SK-LU-1, or NCI-H82.

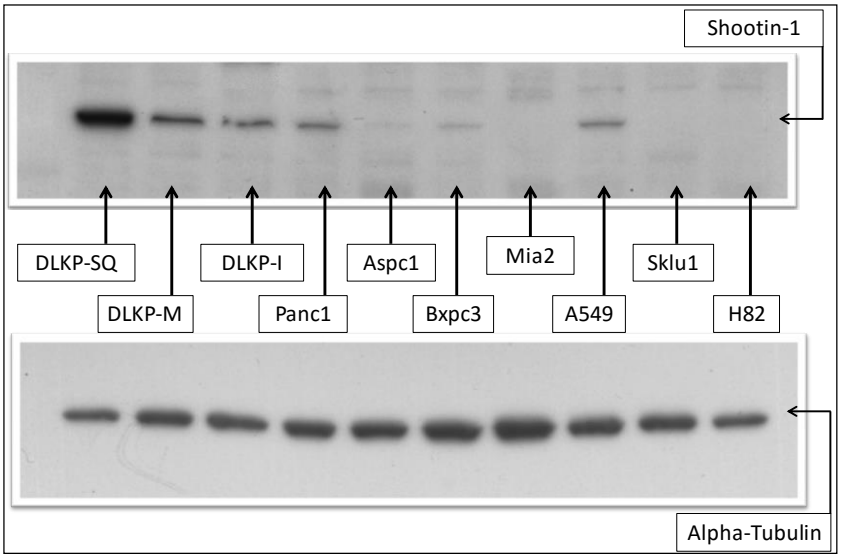


Figure 4.27: Western blot probed for Shootin-1 in a panel of cell lines. Shootin-1 bands came up at ~100kDa in cell lines where it was present. Alpha Tubulin loading control included.

4.11. Summary

Label-free quantitative LC-MS/MS analysis of the DLKP cell line and its clonal subpopulations identified Shootin-1 as a significantly differentially expressed protein, with highest expression in DLKP-SQ relative to all other cell lines. This was found to be the case for both exponential and stationary phases of growth, indicating a strong expression pattern for the protein, conserved across multiple growth phases. Follow up experiments validated this expression profile at a protein and mRNA level, confirming the high expression in DLKP-SQ. In addition, DLKP-M was found to express the next highest abundance of Shootin-1 using the techniques mentioned above. In order to further investigate Shootin-1 in the DLKP cell line and clonal subpopulations, immunocytochemical and immunofluorescent staining of fixed cells was carried out. These analyses highlighted the localisation of Shootin-1 and also further validated the expression pattern described in the proteomic analyses. Localisation of Shootin-1 in DLKP-SQ was found to be cytoplasmic and also concentrated in lamellipodial-like structures at the edges of the cells. In addition, ICC staining also revealed Shootin-1 localisation in fine dendritic-like outgrowths extending from DLKP-SQ. Staining for Shootin-1 was also found in DLKP-M, and was seen to be high in the protrusions which extend from that cell type.

Transient protein knockdown of Shootin-1 in DLKP-SQ followed by functional analyses showed a significant reduction in the migratory ability of the cells. No effects on proliferation or the invasive ability of DLKP-SQ were observed upon knockdown of Shootin-1. Functional analyses and staining of Shootin-1 performed in this study suggest a potential link between Shootin-1 and actin in this lung cancer cell line, as actin is known to be intimately involved in the outgrowth of cell protrusions. As this link has only been shown in neuronal cells to date, co-staining of Shootin-1 and F-actin was therefore carried out using confocal microscopy. This analysis provided strong evidence of co-localisation of Shootin-1 and F-actin in DLKP-SQ cells, strengthening the possible interaction between the two proteins. Transient RNAi knockdown of Shootin-1 did not affect the abundance or localisation of F-Actin as assessed by confocal microscopy, suggesting that Shootin-1 does not regulate Actin expression or localisation, but may still act as an interacting partner in DLKP-SQ.

To investigate this link further, co-immunoprecipitation of Shootin-1 was performed on DLKP-SQ cells in order to discover the binding partners of Shootin-1, and to determine if actin is among them. Repeated co-IP analyses did not find actin as a binding partner of Shootin-1 however a panel of proteins were highlighted which are strong potential binding partners. Investigations into a subset of these (Semenogelin-1; Src-Substrate Cortactin; and Striatin-3, were found to be expressed in all DLKP cell lines by western blot analyses. These proteins have neuronal roles, as well as actin binding roles, which suggest Shootin-1 may interact with actin through other proteins in DLKP-SQ. In addition, the neuronal role of Semenogelin-1 and Striatin may strengthen the neuroendocrine qualities of the DLKP cell line as previously described (McBride et al. 1998a).

To determine the functional role on Shootin-1 in the DLKP clonal subpopulations, an overexpression vector was constructed using the endogenous sequence for the protein extracted from the DLKP-SQ cell line. The vector was successfully constructed and DNA sequencing confirmed the coding sequence for Shootin-1 was present on the vector. Following transfection of the vector, overexpression of Shootin-1 was achieved in DLKP-M and DLKP-I compared to empty vector controls. Unfortunately, time did not allow for subsequent functional analysis of the effects of Shootin-1 overexpression.

Chapter 5. MARCKS

5.1. MARCKS

Myristoylated alanine-rich C-kinase substrate (MARCKS) was found to be a significantly differentially expressed protein by LC-MS/MS analysis, with highest expression in DLKP-M. MARCKS is a ubiquitously expressed multi-functional protein which has been found to be especially abundant in the nervous system (Albert, Nairn and Greengard 1987). Originally identified as a major substrate for Protein Kinase-C (PKC), MARCKS is tethered to the plasma membrane via a myristoyl group at its N-terminus and also by an electrostatic interaction between membrane phospholipids and the effector domain (ED) of the protein (Arbuzova, Schmitz and Vergeres 2002). Upon phosphorylation by PKC at the ED, MARCKS dissociates from the plasma membrane and moves to the cytosol where it has been shown to play roles in cell migration through direct binding to actin resulting in cell remodelling (Li et al. 2008a). The ED of MARCKS has also been shown to be a target for calcium (Ca^{2+}) activated calmodulin. Increased Ca^{2+} concentration results in calmodulin binding to MARCKS at the ED and initiates its reversible dissociation from the plasma membrane (Matsubara et al. 2003). The association of MARCKS with actin has shown the protein to be a key regulator of dendritic outgrowth in neuronal cells. Knockdown of MARCKS significantly reduced dendritic complexity, while over-expression enhanced dendritic development in neuronal cells (Calabrese and Halpain 2005, Li et al. 2008a). Thus far, there have only been two studies on MARCKS in lung cancer. It has been identified as a potential biomarker associated with poor prognosis in human SCC using immunohistochemical methods (Hanada et al. 2013). The second study detected high expression of MARCKS in highly invasive lung cancer cell lines and found that MARCKS knockdown by siRNA significantly reduced cell migration (Chen et al. 2014b).

5.2. Label-free Analysis of DLKP Cell Lines: MARCKS

The label-free profiling experiment sought to highlight differentially expressed proteins between the DLKP cell line as well as the clonal subpopulations. A protein with a high abundance in one cell line compared to the others may play a role in the known phenotypic characteristics of the cell. DLKP-M is characterised by it high migratory and invasive abilities relative to the other clonal subpopulations. Therefore, protein expression patterns were screened to look for a highly abundant protein in DLKP-M.

MARCKS protein stood out as an interesting candidate due to its high abundance in DLKP-M compared to DLKP and clonal subpopulations (Figure 5.1) during both exponential and stationary phase of growth. The protein was found as strongly expressed in DLKP-M in both the global analysis and the two-sample experimental designs.

5.2.1. Global Analysis: MARCKS

- During exponential phase of growth MARCKS abundance shows a maximum fold change of **51.70** with highest abundance in DLKP-M compared to the lowest in DLKP.

- During stationary phase of growth, the protein abundance has a maximum fold change of **28.37** with highest abundance in DLKP-M and lowest in DLKP.

During both exponential and stationary phases, MARCKS passed the stringent criteria of: ≥ 2 peptides contributing to the protein identification, an abundance fold change of ≥ 2 between two experiment groups, an ANOVA p-value of < 0.05 for a protein between experimental groups. A representative protein abundance graph from Progenesis software is shown for exponential phase. This graph displays the normalised abundance values for MARCKS across the DLKP cell line and its clonal subpopulations (Figure 5.1).

MARCKS shows the highest abundance in DLKP-M, followed by DLKP-I. The expression levels of MARCKS is at its lowest in DLKP and DLKP-SQ, with both cell lines showing the greatest fold change in abundance when compared to DLKP-M. Next, the expression level of MARCKS was determined using experimental designs which compared each cell line against one other using a one vs. one design: e.g. DLKP-SQ vs. DLKP-M. Peptides which contributed to the identification of MARCKS are shown in Figure 5.2.

Figure 5.1: Label-free abundances for MARCKS during exponential phase of growth viewed in Progenesis software. Highest expression is in DLKP-M, followed by DLKP-I. 4 peptides contributed to this identification with a maximum fold change of 51.70 between DLKP-M and DLKP (ANOVA, p= 3.14E-09).

Accession: P29966

Description: Myristoylated alanine-rich C-kinase substrate OS=Homo sapiens GN=MARCKS PE=1 SV=4 - [MARCS_HUMAN]

Σ	#	Score	Anova (p)	Max Fold Change	Highest Mean	Lowest Mean	Tag	Abundance	m/z	Charge	Retention Time (mins)	Mass error (ppm)	Peptide Sequence
•	1139	85.840	3.45E-13	91.3	DLKP M	DLKP Parents		1.67E+06	595.6303	3	65.779	-0.49	GEAAAERPGEAAVASSPSK
•	1719	99.000	1.11E-12	19.5	DLKP M	DLKP Parents		3.335E+06	811.9063	2	93.842	2.00	GEPAAAAAPEAGASPVEK
•	20353	61.790	4.46E-10	1.74E+03	DLKP M	DLKP Parents		7.799E+05	1458.1577	2	140.823	1.95	EAPAEGEAAEPGSPTAAEGEAASAASSTSSPK
•	7849	91.850	5.81E-10	134	DLKP M	DLKP Parents		1.007E+06	994.8035	3	104.076	1.28	EELQANGSAPAADKEEPAAAGSGAASPSAAEK

Figure 5.2: Peptides which contributed to the identification of MARCKS are displayed by Progenesis software. Each peptide had an ANOVA $p < 0.05$, and MASCOT scores ≥40.

5.2.2. Two-Sample Experimental Designs

MARCKS was highlighted in the global analysis with DLKP-M expressing the protein at a much higher abundance than all other cell lines in the comparison. In the two-sample analysis, the expression of MARCKS was determined for cell lines showing intermediary fold changes for the protein. Results are summarised in Table 5.1.

The analysis is made up of six comparisons for each phase of growth, with every permutation accounted for. Each analysis has stringent filtering criteria applied to the resulting proteins which are: ≥ 2 peptides contributing to the protein identification, an abundance change of ≥ 2 fold between the two experimental groups, an ANOVA p-value of <0.05 for a protein between the experimental groups. In each analysis containing DLKP-M in the comparison, MARCKS is identified as being the most abundant protein. This indicates MARCKS is specifically highly expressed in DLKP-M.

- During exponential phase of growth, MARCKS is highlighted in every comparison as a statistically significantly differentially expressed protein (Table 5.1A). The data shows that this protein has the highest abundance in DLKP-M out of all the cell lines. The largest fold change in this phase of growth is between DLKP-M and DLKP at **49.82**, with highest MARCKS abundance in DLKP-M. In each other comparison, MARCKS consistently is shown as being highest in DLKP-M.

- Stationary phase analysis (Table 5.1B) revealed the same trend as exponential with expression of MARCKS being highest in DLKP-M in each comparison containing that cell line. The highest fold change is once again between DLKP-M and DLKP at **24.50**, highest in DLKP-M. The remaining comparisons matched the findings for exponential phase.

Exponential Phase:

(A)

Individual Comparison		Description	Max fold change	Highest Abundance	Lowest Abundance
DLKP-SQ	**vs. DLKP-M**	**Myristoylated alanine-rich C-kinase substrate**	**20.94**	**DLKP-M**	**DLKP-SQ**
DLKP-SQ	vs. DLKP-I	Myristoylated alanine-rich C-kinase substrate	8.16	DLKP-I	DLKP-SQ
DLKP-I	**vs. DLKP-M**	**Myristoylated alanine-rich C-kinase substrate**	**2.53**	**DLKP-M**	**DLKP-I**
DLKP-SQ	vs. DLKP	Myristoylated alanine-rich C-kinase substrate	2.38	DLKP-SQ	DLKP
DLKP-M	**vs. DLKP**	**Myristoylated alanine-rich C-kinase substrate**	**49.82**	**DLKP-M**	**DLKP**
DLKP-I	vs. DLKP	Myristoylated alanine-rich C-kinase substrate	19.68	DLKP-I	DLKP

Stationary Phase:

(B)

Individual Comparison		Description	Max fold change	Highest Abundance	Lowest Abundance
DLKP-SQ	**vs. DLKP-M**	**Myristoylated alanine-rich C-kinase substrate**	**20.98**	**DLKP-M**	**DLKP-SQ**
DLKP-SQ	vs. DLKP-I	Myristoylated alanine-rich C-kinase substrate	4.47	DLKP-I	DLKP-SQ
DLKP-I	**vs. DLKP-M**	**Myristoylated alanine-rich C-kinase substrate**	**5.11**	**DLKP-M**	**DLKP-I**
DLKP-SQ	vs. DLKP	Myristoylated alanine-rich C-kinase substrate	3.59	DLKP-SQ	DLKP
DLKP-M	**vs. DLKP**	**Myristoylated alanine-rich C-kinase substrate**	**24.50**	**DLKP-M**	**DLKP**
DLKP-I	vs. DLKP	Myristoylated alanine-rich C-kinase substrate	5.34	DLKP-I	DLKP

Table 5.1: Individual comparisons of the DLKP cell line and its subpopulations for exponential **(A)** and stationary **(B)** phases of growth. In each comparison containing DLKP-M, MARCKS was highlighted as being most abundant in that cell line. For each comparison, ≥ 2 peptides contributed to the protein identification, and ANOVA p-values were < 0.05 in the protein abundance comparisons.

5.3. Validation of MARCKS Expression in DLKP and Clonal Subpopulations

The expression pattern of MARCKS was validated by western blot using the method described in Section 2.3. The protein came up with bands visible at 80 kDa in the DLKP cell line model which is the expected molecular weight of MARCKS. The expression pattern of MARCKS found by western blot analysis is shown in Figure 5.3A. MARCKS expression profile was graphed using the label-free normalized abundance values. The expression pattern correlated well with label-free data for MARCKS, shown in Figure 5.3B for comparison.

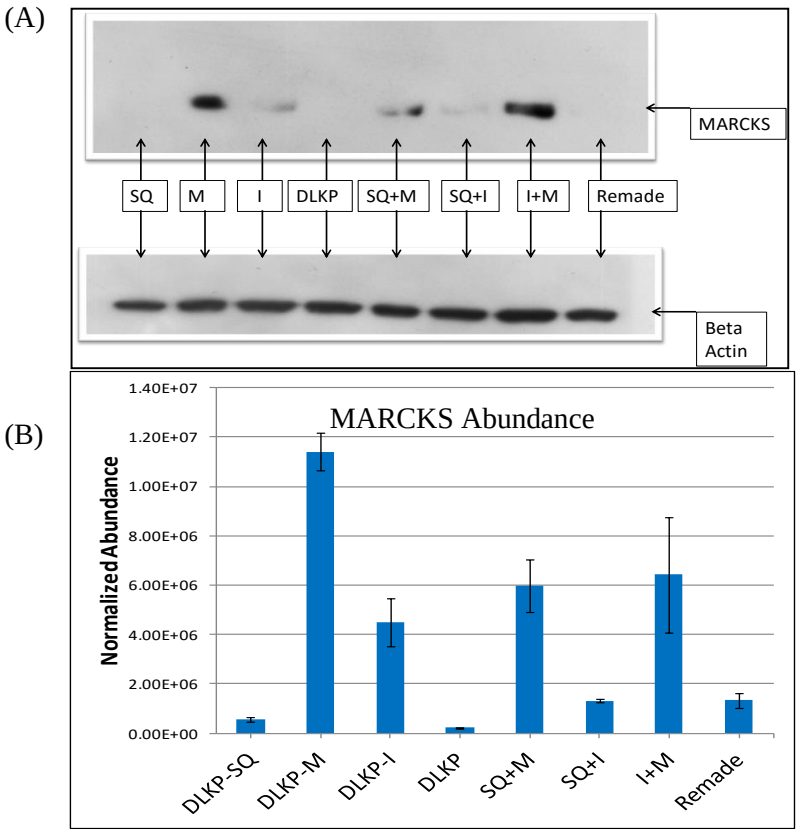


Figure 5.3:(A) Western blot showing the expression levels of MARCKS in all cell lines. Beta-actin was used as a loading control. MARCKS bands are visible at ~80 kDa as expected (n=3). **(B)** Label-free abundance values for MARCKS in all DLKP cell lines during exponential phase. Standard deviations are shown. (n=3).

5.4. Proteomic Comparison of DLKP Compared to the Three Clones and Combination Cell Lines

The expression pattern of MARCKS was next analysed using an experimental design which incorporated all samples in a global analysis. The design includes the combination cell lines, which are 1:1 combinations of the clonal subpopulations and a reconstituted version of DLKP made from the clones in proportions described previously (McBride et al. 1998b). Details of the creation of these cell lines can be found in Section 3.1.1.

A large global analysis was created in Progenesis software which incorporated all samples into one large experimental design, including the combination cell lines. were included in the analysis to determine if co-culture of the DLKP clones had any effect on the expression of MARCKS. Progenesis software groups together peptides which contribute to the identification of MARCKS, and displays their abundance levels for each sample group. The standardised normalised abundance values for peptides belonging to MARCKS are shown in Figure 5.4, which demonstrates the expression profile for each peptide. Peptides which contributed to the identification of MARCKS are shown in Figure 5.2.

Peptides belonging to MARCKS are more abundant in DLKP-M compared to all other samples in the global experimental design. In the combination lines which contain DLKP-M as a component, the abundances of peptides belonging to MARCKS are significantly higher than in all other combination cell lines. In addition, I+M seems to show a slightly higher abundance of MARCKS than is shown for SQ+M which is most likely due to the higher abundance of the protein in the DLKP-I clone. DLKP-Remade shows a slightly higher abundance of MARCKS than the DLKP cell line.

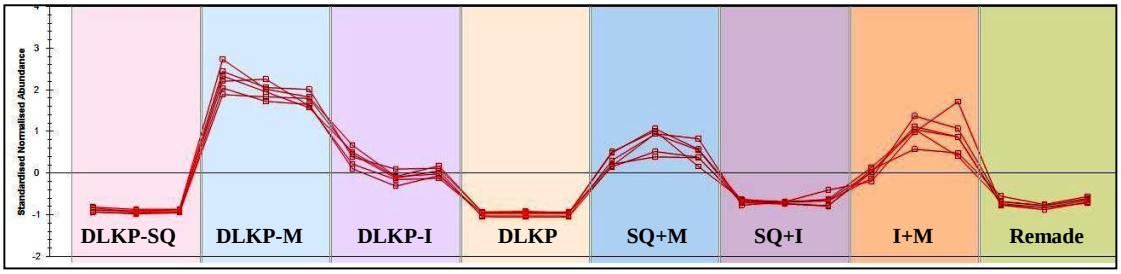


Figure 5.4: Peptide expression pattern for MARCKS as displayed by Progenesis software. Peptides identifying as MARCKS cluster well together showing highest expression of the protein in DLKP-M and lowest in DLKP (ANOVA, p= **9.70E-13**).

5.5. Validation of MARCKS Expression in DLKP plus the Clones and Combination Cell Lines

5.5.1. Western Blot

The DLKP cell lines along with the combination lines were analysed by western blot to determine if the expression pattern found by the label-free analysis could be validated. Results of the western blot analysis are shown in Figure 5.3A, showing DLKP-M as a cell line containing a relatively high abundance of MARCKS in the clones and combination lines. The label-free analysis output for MARCKS (Figure 5.3B) shows the abundance of MARCKS is also relatively high in the combination lines containing the DLKP-M clone as a constituent.

The combination lines show high expression of MARCKS in the cell lines which contain DLKP-M as a constituent. However, SQ+M shows lower abundance of MARCKS compared to I+M. This is an expected finding as DLKP-I expresses the second highest abundance of MARCKS out of all the DLKP clonal subpopulations. Neither DLKP nor DLKP-Remade show detectable levels of MARCKS by Western Blot analysis. Label-free proteomic analysis showed DLKP-Remade to express MARCKS at a higher abundance than DLKP, however in both cases the abundance was very low. Western Blot analysis therefore correlates well with the label-free analysis.

5.5.2. mRNA Expression

Following on from the label-free analysis and western blot validation, DLKP samples were analysed using quantitative PCR (qPCR) to determine the relative levels of MARCKS mRNA expression between each cell line. Relative quantification was performed on all cell lines and combinations using beta-actin as the endogenous control gene. Biological duplicate samples were analysed and results are combined in Figure 5.5. The calculation of relative quantity (RQ) was achieved by normalizing the data to DLKP-M.

mRNA level of MARCKS correlated well with the label-free analysis of the DLKP cell lines. DLKP-M shows the highest expression of MARCKS, followed by DLKP-I and then DLKP-SQ. Fold changes in the mRNA transcript are comparable to those found in the proteomics analysis between highest and lowest expressing cell lines. This expression pattern also matches what was found by western blot analysis. In the combination cell lines, SQ+M and I+M show the highest expression of MARCKS, with lowest expression in SQ+I and Remade. A slightly higher gene expression of MARCKS is seen in SQ+M compared to I+M, which contradicts the western blot findings, however this is not a statistically significant difference. In addition, DLKP-Remade expresses a higher abundance of MARCKS than the DLKP cell line. This expression pattern correlates with the findings of the label-free proteomic data. Different expression levels of MARCKS between DLKP and DLKP Remade may suggest the presence of uncharacterised cell types in DLKP which reduce the expression of MARCKS in that cell line.

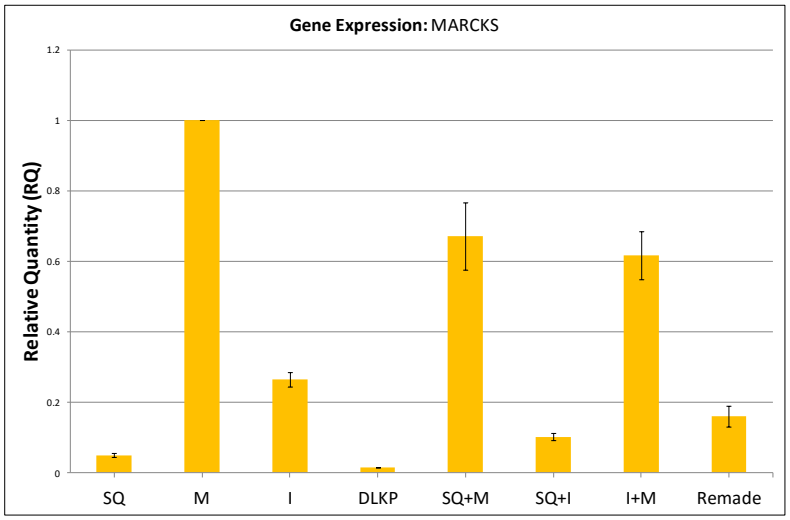


Figure 5.5: qPCR results showing gene expression of MARCKS in the clones and combination cell lines. Samples are analysed in biological duplicate. Error bars shown are the standard deviations between the two replicate experiments. Data are normalized to DLKP-M.

5.6. Imaging of MARCKS in the DLKP Cell Lines:

Validation of MARCKS expression by western blot and mRNA analysis in the DLKP cell lines confirmed the pattern shown by the label-free data. In order to investigate the localization of MARCKS protein in the DLKP cell lines and further validate the expression pattern found by the techniques mentioned above, MARCKS was stained using immunocytochemistry (ICC) and immunofluorescence (IF). Staining was performed using the method described in Section 2.6 on DLKP and clonal subpopulations fixed to glass slides by Paraformaldehyde. Unfortunately, the immunocytochemistry staining resulted in high non-specific staining with the IHC compatible antibody. Immunofluorescence on the other hand resulted in staining specific for MARCKS in the DLKP cell lines.

5.6.1. Immunofluorescent Staining of MARCKS in DLKP Cell Lines

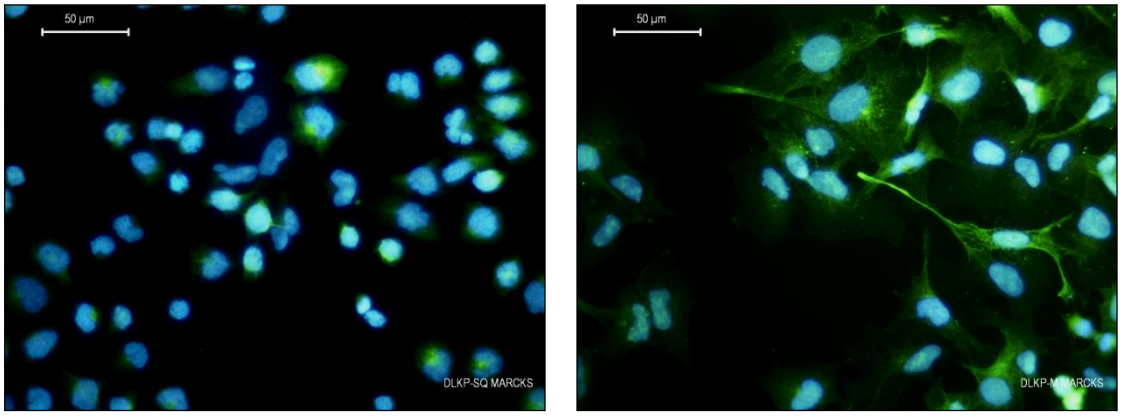
Immunofluorescence staining of MARCKS was carried out in DLKP and each of its clonal subpopulations to allow visualization of the distribution of the protein in each cell type, as well as a semi-quantitative assessment of MARCKS abundance. Staining was carried out using the method described in Section 2.6.3 with resulting images shown in Figure 5.6. MARCKS staining showed very high levels in DLKP-M compared to the other cell lines which correlated well with the label-free proteomic data. High levels of staining can be seen throughout the cytoplasm of DLKP-M, and in many cases the localization of the protein is concentrated in the cell projections which are characteristic of this cell type. There is also punctate staining clearly visible in DLKP-M cells with intense dots of MARCKS staining throughout the cells.

DLKP-I had the next highest level of MARCKS staining. In the case of this cell type, staining was intense in a subpopulation within DLKP-I, with low level staining in the remaining population. Punctate staining for MARCKS is also visible in the cytoplasm of DLKP-I as well as accumulation of the protein in the tips of projection structures. DLKP itself shows a varied staining pattern also, with some cells in the population more intensely stained compared to others. This staining pattern is expected as DLKP is made up of the clonal subpopulations, with DLKP-M comprising approximately 5% of the total population (McBride et al. 1998b). Similar localization of

MARCKS is found in this cell type with projections showing high MARCKS staining intensity.

DLKP-SQ has the lowest level of staining for MARCKS with very low level cytoplasmic staining visible in the majority of the population. Similar to the other cell types analysed here, a subpopulation within DLKP-SQ shows higher staining for MARCKS than the majority of cells. The staining found in subpopulations of the clones themselves may support the theory of clone inter-conversion discussed in Section 1.4.1.4. The staining localization of MARCKS suggests it may play a role in actin polymerization, similar to Shootin-1. MARCKS punctate staining is visible in all cell types and it is particularly high in DLKP-M. This suggests an additional role for the protein in DLKP-M.

DLKP-SQ **DLKP-M**



DLKP-I **DLKP**

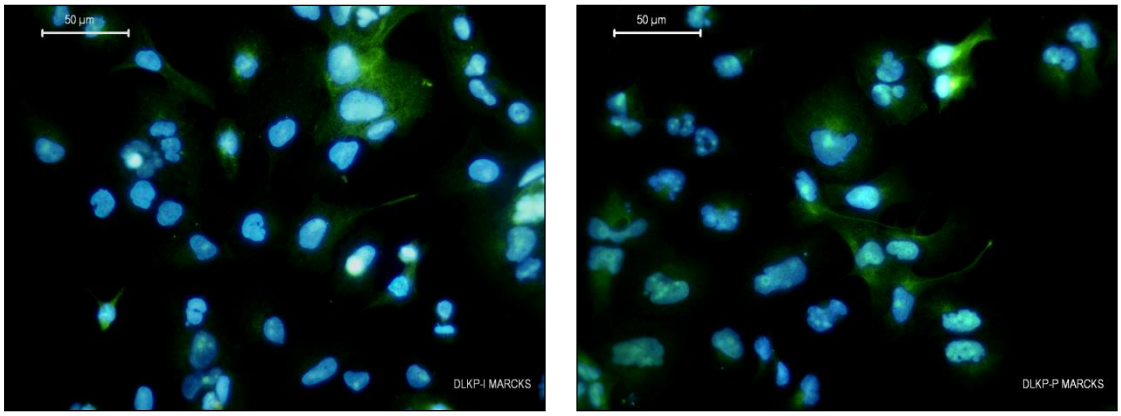


Figure 5.6: Immunofluorescence images staining for MARCKS in DLKP-SQ, DLKP-M, DLKP-I and DLKP. Nuclei are stained with Dapi. Negative controls with secondary antibody only did not result in staining (images not shown). Original magnification used on all micro-pictographs shown at 400x, 50µm scale bar. (n=2)

5.7. Functional Analysis of MARCKS Knockdown by RNAi

5.7.1. Transfection Optimisation

To investigate the potential role played by MARCKS in the DLKP model, protein knockdown was carried out and followed up by functional analysis in DLKP-M. Protein knockdown was performed using RNA interference (RNAi). DLKP-M was chosen as the cell line to investigate the functional effects of protein knockdown due to the high expression of MARCKS in that cell line at the protein and mRNA level. Following knockdown of MARCKS by RNAi, the effects of this were assessed using a panel of functional assays. A standard commercially available negative siRNA control was used to compare against for the knockdown samples. This negative siRNA is homologous to no known sequence and therefore acts as a control to assess the effects of introducing siRNA into cells. Before MARCKS knockdown was attempted, DLKP-M cells underwent rigorous optimisation steps as previously described in Section 4.7.1.

5.7.1.1. Optimisation of Transfection Reagent Volume:

To optimise the volume of transfection that would effectively deliver siRNA into DLKP-M cells, 1 μl and 2 μl per well volumes of Lipofectamine were tested (after previous optimisation). Cell viability was analysed using an acid phosphatase proliferation assay as described in Section 2.5.1, and results are shown in Figure 5.7. Results are displayed as a percentage survival value relative to the negative siRNA control (2 μl/well of transfection reagent). In all conditions where siRNA was included, a 5 nM concentration was used. Four different transfection conditions were tested in order to determine the optimal volume of Lipofectamine reagent to use for RNAi in DLKP-M: Cells-only; Lipo-only; Negative siRNA; Positive siRNA (Kinesin siRNA).

Both volumes (1 μl/well and 2 μl/well) of transfection reagent had no effect on the Lipo-only control samples compared to the non-transfected Cells-only condition (Figure 5.7). Similarly, the negative siRNA control had a negligible effect on viability using both volumes of transfection reagent. The positive siRNA control had a large effect on the viability of DLKP-M cells at both volumes. This Kinesin siRNA induces

cell death by arresting growth during mitosis, preventing formation of the bipolar spindle. Using a volume of 1 µl of transfection reagent reduced the cell viability by ~35%, whereas using 2 µl of transfection reagent reduced the cell viability by ~62%. It was determined in previous optimisation experiments that higher volumes of transfection reagent had a detrimental effect on the viability of DLKP-M in the Lipo-only and negative control samples. Therefore, a volume of 2 µl transfection reagent was used for all RNAi experiments involving DLKP-M moving forward.

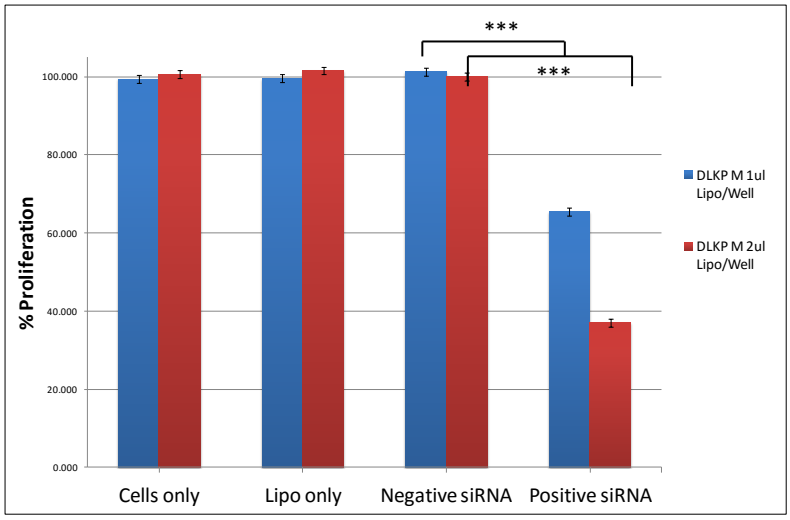


Figure 5.7: Proliferation assay performed on DLKP-M cells using 1µl/6-well (Blue) and 2µl/6-well (Red) of Lipofectamine transfection reagent. Percentage survival values are relative to the negative siRNA control. (1µl /well: P-Value 1.3E-03) (2µl/well: P-Value: 2.5E-04).

5.7.1.2. Optimisation of siRNA Concentration: MARCKS

The optimization of MARCKS RNAi transfection conditions was finalized by determining the optimal siRNA concentration to achieve maximum MARCKS protein knockdown. A transfection experiment was set up with a range of siRNA concentrations for MARCKS in DLKP-M:

- Cell-only, Lipofectamine-only and Negative siRNA controls.
- 5 nM, 10 nM, 20 nM and 30 nM of siRNA were tested for MARCKS.
- Two different siRNA molecules were tested in tandem which target MARCKS.

Cell lysates were prepared from the various transfection conditions as described in Section 2.3.1, 72 hours post transfection. Western blot analysis was performed on these samples and probed using an antibody specific for MARCKS. The results are shown in Figure 5.8. Untreated cells, Lipofectamine-only and Negative siRNA samples take up lanes 1-3 respectively, and bands for MARCKS are clearly visible at ~80kDa. A concentration of 30 nM siRNA was used in the negative siRNA sample to represent the maximum siRNA concentration in the MARCKS test panel. Each concentration tested effectively reduced the expression of MARCKS in DLKP-M to near undetectable levels. Based on the Western Blot results, 20 nM of siRNA was determined as the optimal concentration to use for MARCKS RNAi before functional assessment of the role of the protein in DLKP-M.

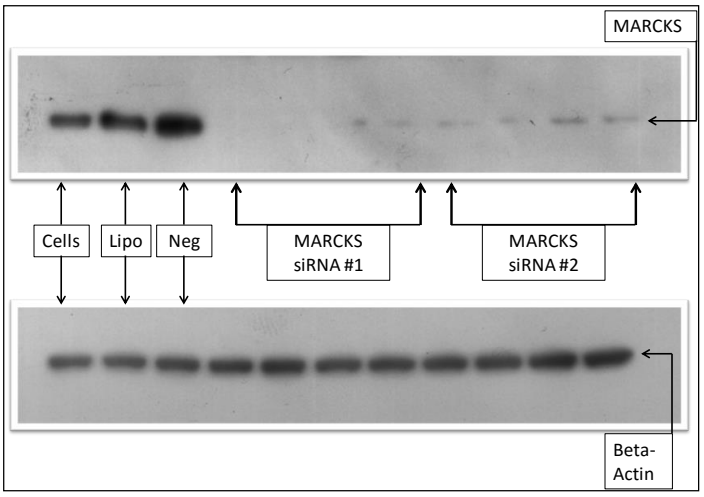


Figure 5.8: Western blot probing for MARCKS-1 (80kDa). Lanes 1-3 contain: Cells-only, Lipofectamine-only and Negative siRNA (30 nM). Each MARCKS siRNA group (#1 and #2) is comprised of four sample lanes which were treated with 30 nM, 20 nM, 10 nM and 5 nM of siRNA from left to right. Beta-actin loading control included.

5.7.2. Effects of MARCKS Knockdown on Proliferation, Migration and Invasion in DLKP-M

To study the potential role of MARCKS in DLKP-M, a panel of functional assays were carried out following siRNA transfection to knock down the target. Two individual siRNA molecules were used to knockdown MARCKS in DLKP-M using conditions optimised and described in Section 5.7.1. **Proliferation, migration** and **invasion** assays were carried out to investigate if a reduction in MARCKS expression had a significant effect.

5.7.2.1. Proliferation of DLKP-M Post-MARCKS Knockdown

Proliferation of DLKP-M cells was measured 72 hours post transfection as described in Section 2.2.3. Knockdown of MARCKS by RNAi was found to have no significant effect on proliferation compared to the negative control (n=3). Results are shown in Figure 5.9.

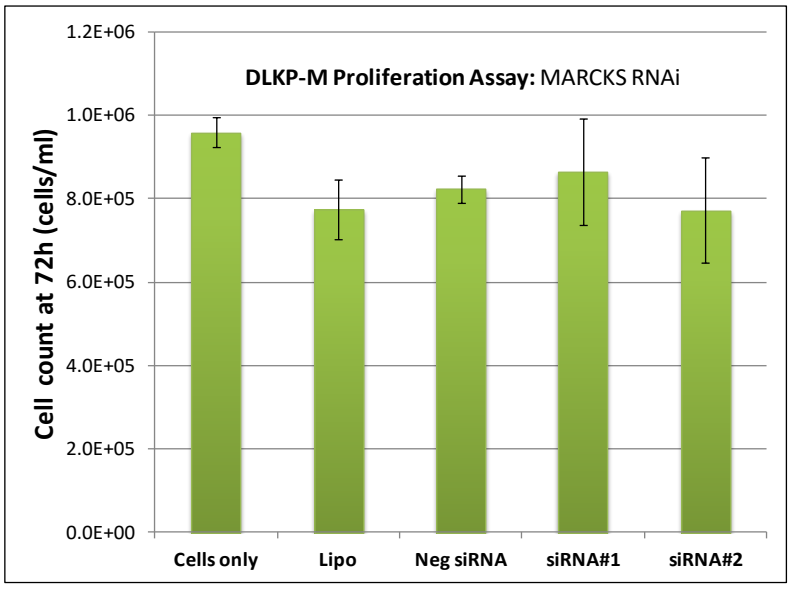


Figure 5.9: Proliferation assay on DLKP-M post transfection for MARCKS RNAi. There was no statistically significant difference in proliferation between siRNA knockdown samples (#1 and #2) and the negative control (n=3). (siRNA#1 P-value: 0.63, siRNA#2 P-value: 0.56)**.**

5.7.2.2. Migration of DLKP-M Post-MARCKS Knockdown

Migration assays were performed 72 hours post transfection using the method as described in Section 2.5.4. Cells were allowed to migrate through the uncoated membrane for 24 hours before being fixed and stained by crystal violet. Upon comparing the MARCKS RNAi samples to the negative control, a significant inhibitory effect on the migratory ability of the cells was found (n=3). Results are shown in Figure 5.10

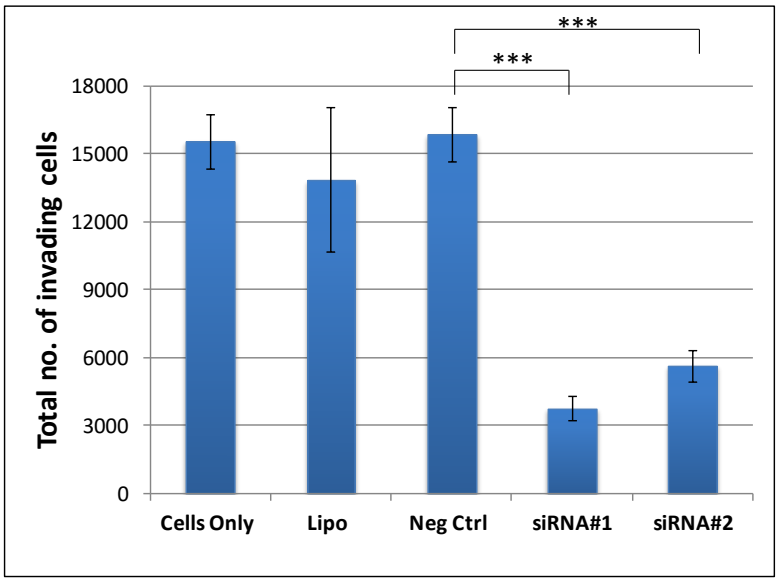


Figure 5.10: Migration assay results for DLKP-M cells following MARCKS knockdown using RNAi. Total number of cells which migrated through the uncoated membrane is shown (n=3). A statistically significant difference between siRNA knockdown samples (#1 and #2) and the Negative siRNA control was found. (**siRNA#1 P-value**: 8.7E-04, **siRNA#2 P-value**: 7.7E-04)

5.7.2.3. Invasion of DLKP-M Post-MARCKS Knockdown:

Invasion assays performed 72 hours post transfection using the method as described in Section 2.5.3. Knockdown of MARCKS had a significant effect on the invasive ability of DLKP-M cells when compared to the negative control sample (n=3). Results are shown in Figure 5.11.

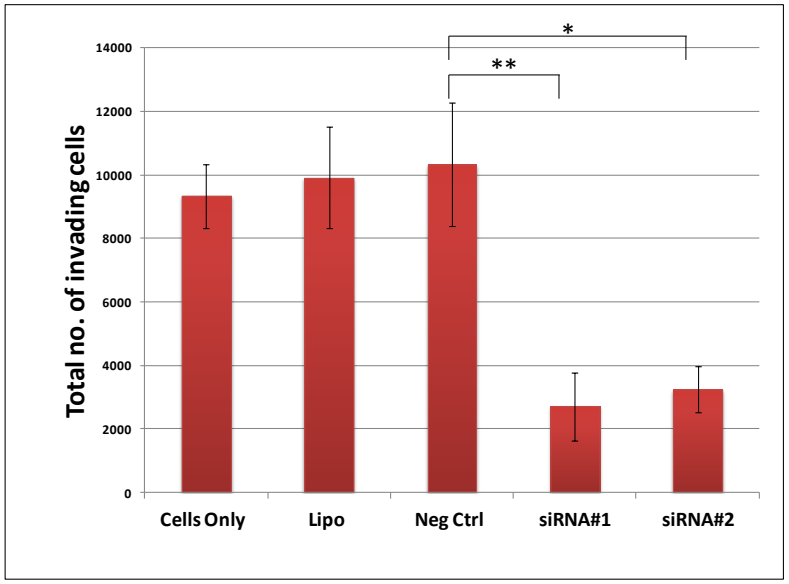


Figure 5.11: Invasion assay results for DLKP-M cells following MARCKS knockdown using RNAi. Total number of cells which migrated through the Matrigel coated membrane is shown (n=3). A statistically significant difference between siRNA knockdown samples (#1 and #2) and the negative siRNA control was found. (**siRNA#1 P-value**: 8.5E-03, **siRNA#2 P-value**: 1.5E-02)

5.7.3. Co-Staining of MARCKS/F-Actin after MARCKS Knockdown by RNAi in DLKP-M

MARCKS is a membrane bound protein involved in the cross-linking of actin filaments. The N-terminus of the protein is attached to the membrane via a myristoyl group and anchors the actin network to the inner side of the plasma membrane. Upon phosphorylation by protein kinase C (PKC) or binding by calmodulin/actin, MARCKS becomes detached from the plasma membrane and its association with actin filaments is inhibited, leading to its presence in the cytoplasm (Matsubara et al. 2003). Due to these reported links between MARCKS and actin, it was decided to investigate this potential link in DLKP-M using confocal microscopy. Cells were treated with RNAi to knock down expression of MARCKS to assess the effect on actin, and also localization of MARCKS itself.

Once RNAi was performed, DLKP-M cells were allowed to grow for 48 hours before being scratched with a P1000 tip, washed and had growth media replaced. Cells were allowed to grow for a further 24 hours before being fixed with 4% paraformaldehyde. Scratching was done to encourage the cells to grow in a particular direction to close the created wound. This enables the visualisation of motile cells by confocal microscopy once the cells are stained. Resulting images are shown in Figure 5.12 to Figure 5.16.

DLKP-M: Cells Only

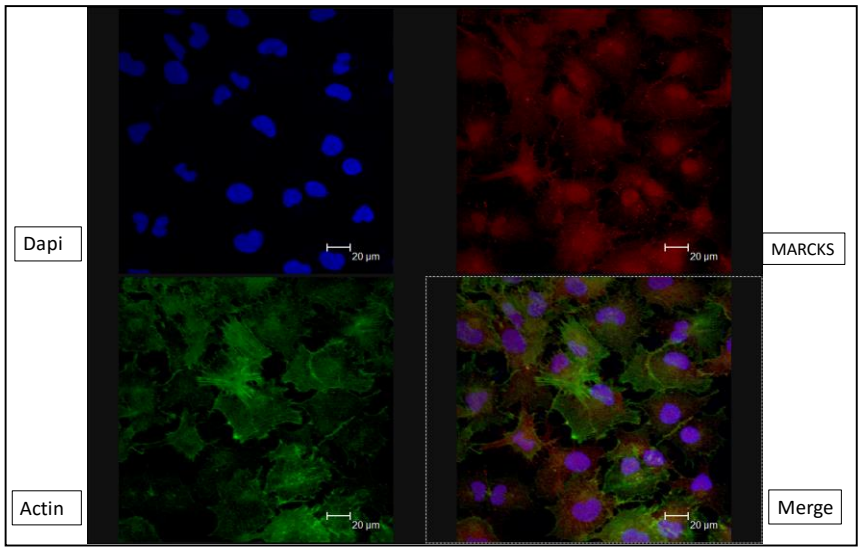


Figure 5.12: DLKP-M cells stained for MARCKS, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are untreated by any RNAi method. Original magnification of all photomicrographs, ×400, scale bar = 20 µm.

DLKP-M: Lipofectamine-only Control

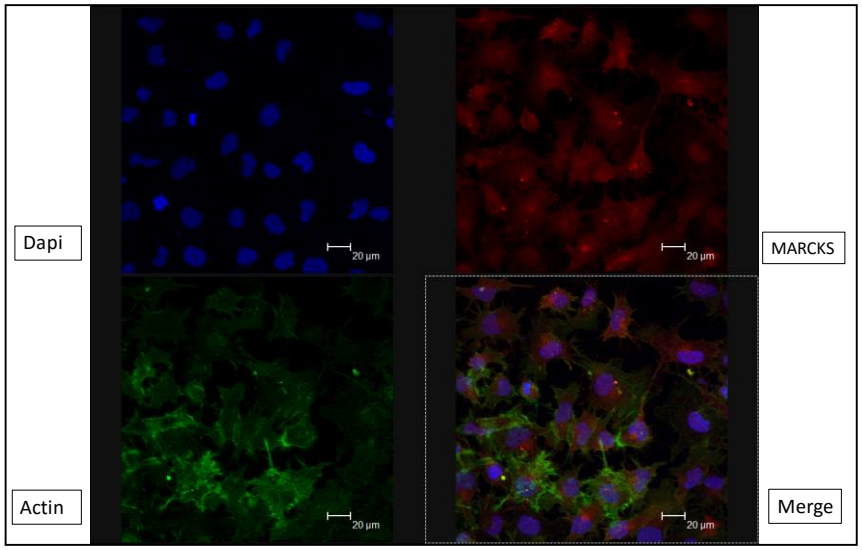


Figure 5.13: DLKP-M cells stained for MARCKS, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are treated with transfection reagent only. Original magnification of all photomicrographs, ×400, scale bar = 20 µm.

DLKP-M: Negative siRNA Control

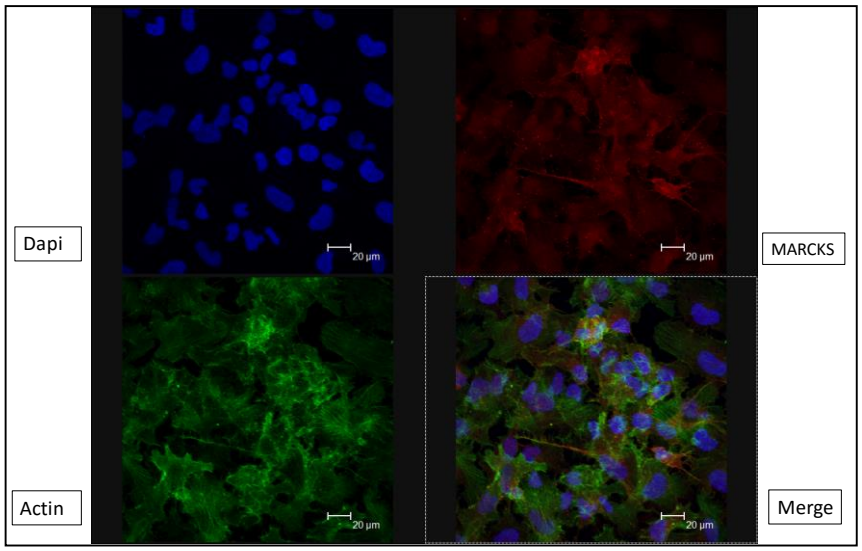


Figure 5.14: DLKP-M cells stained for MARCKS, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are treated with transfection reagent and negative siRNA control. Original magnification of all photomicrographs, ×400, scale bar = 20 µm.

DLKP-M: MARCKS Knockdown- siRNA #1

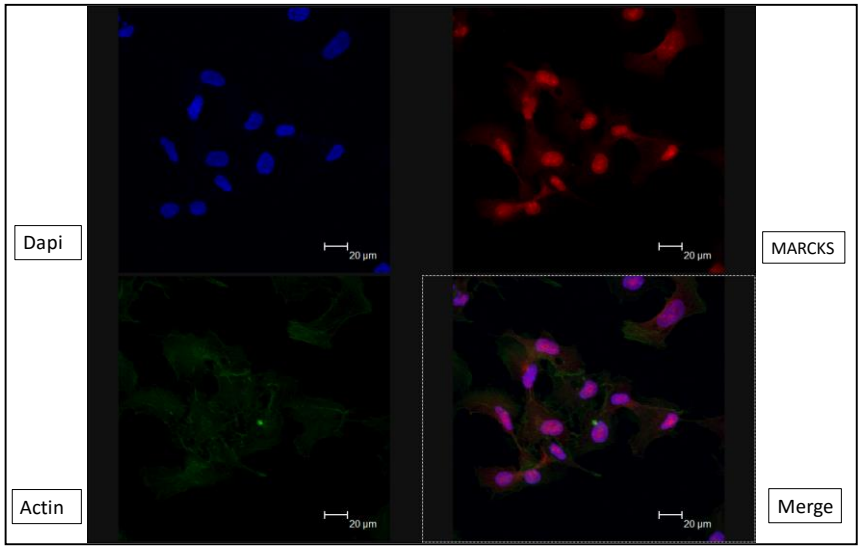


Figure 5.15: DLKP-M cells stained for MARCKS, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are treated with siRNA#1 for MARCKS. Original magnification of all photomicrographs, ×400, scale bar = 20 µm.

DLKP-M: MARCKS Knockdown- siRNA #2

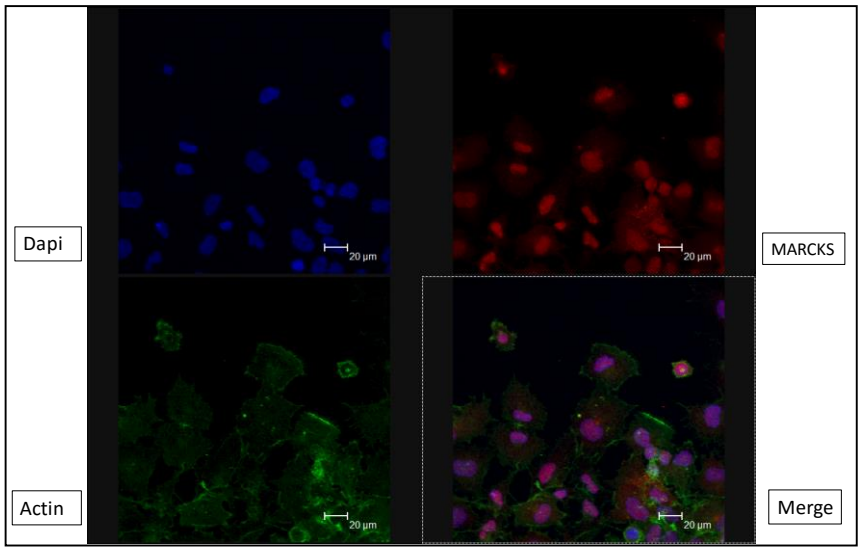


Figure 5.16: DLKP-M cells stained for MARCKS, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are treated with siRNA#2 for MARCKS. Original magnification of all photomicrographs, ×400, scale bar = 20 µm.

Negative siRNA Control: DLKP-M High Magnification

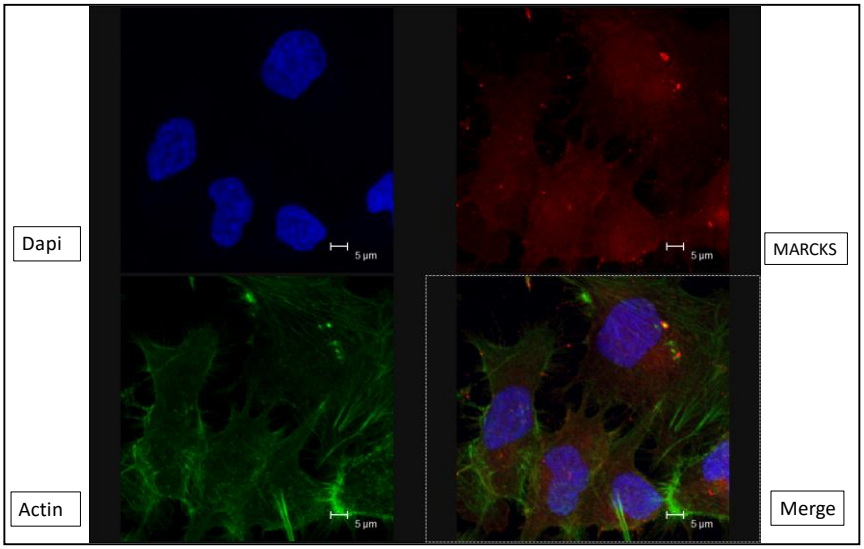


Figure 5.17: DLKP-M cells stained for MARCKS, Actin and Dapi. A merged image is shown in the bottom right corner. DLKP-M cells at high magnification showing co-localization of MARCKS and Actin in a punctate staining pattern. Cells are treated with negative siRNA control. Original magnification of all photomicrographs, ×800, scale bar = 5µm.

The staining of MARCKS (coloured red) in DLKP-M was followed by imaging using confocal microscopy and shows a similar staining pattern to that found by the immunofluorescence work described in Section 5.6.1. Strong staining for MARCKS is visible throughout the cytoplasm of the cells in the control group: Cells-only; Lipo-only; Negative siRNA (Figure 5.12 and Figure 5.13), with MARCKS expression and localization unaffected by the transfection control conditions. Punctate staining is present with intense points of MARCKS protein scattered throughout the cytoplasm. This staining pattern potentially represents accumulations of the MARCKS protein at distinct cellular locations, and may be indicative of a role as part of a protein complex.

MARCKS can also be seen at the tips of projections emanating from the DLKP-M cells. Dapi nuclear staining is clearly visible, and in the merged photograph nuclei have a light purple colour upon combining with red MARCKS staining. This indicates a low level of MARCKS localization in the nucleus or nuclear envelope. F-actin staining (coloured green) is very high in the control DLKP-M cells and can be seen throughout the cytoplasm as well as at the edges of cells. These may be lamellipodia where actin is reported to polymerise aiding in cell movement. Overlap of MARCKS and F-actin staining shows strong co-localization of both proteins. Punctate staining as well as cell projection staining overlap is particularly apparent for both proteins as shown in Figure 5.14.

RNAi knockdown of MARCKS in DLKP-M revealed an interesting development which can be seen in Figure 5.15 and Figure 5.16. Cells here treated with siRNA #1 and siRNA #2 for MARCKS, and in both cases there is large reduction in MARCKS expression for both siRNA. However, the nuclei in both knockdown samples appear to have become enriched with remaining MARCKS once they are treated with RNAi targeting the protein. This is clear from the confocal images which show the overlap between MARCKS and the nuclear stain Dapi. The nuclei are a rich purple colour compared to the control samples, which suggests that knockdown of MARCKS triggered the protein to be translocated the nuclei. Localization of MARCKS is visible in the nuclei of the control samples at a low level; however RNAi appears to have significantly increased staining for MARCKS in this cellular component. F-actin staining appears to be reduced in the knockdown samples compared to the control

group. Western blot analysis probing for beta-actin in these samples did not show a reduction in the expression of actin (Figure 5.8). F-actin was not assayed.

A high magnification image is shown in Figure 5.17 of the negative siRNA control. Punctate cytoplasmic staining of MARCKS is visible, which in many cases shows co-localization with actin. Nuclear staining for MARCKS is also visible, but at a low level compared to the knockdown samples in. Another control was set up using secondary antibody only. This was done to test for the specificity of the target antibody and to check for non-specific binding of the secondary. Images remained clear of staining (not shown).

5.7.4. Over-Expression of MARCKS in DLKP and Clonal Subpopulations

Functional assays performed on DLKP-M after MARCKS knockdown show the protein is highly involved in the phenotypic characteristics of the cell line. Knockdown of MARCKS by RNAi results in a significant reduction of both the migratory and invasive capabilities of DLKP-M. In order to further illuminate the role MARCKS plays in DLKP, over-expression of the protein was next investigated.

To create an over-expression vector for MARCKS, a similar approach was taken as was done for Shootin-1, described in Section 4.9. The open reading frame (ORF) the MARCKS nucleotide sequence was isolated using custom designed primers, and cDNA generated from DLKP-M RNA was used as a template. Primer sequences have restriction enzymes sites incorporated into their design to facilitate the creation of sticky ends for ligation further downstream. PCR amplification of the sequence targeted by the primers resulted in faint bands at the expected size of ~1000bp when they were separated on an agarose gel using electrophoresis. An image of this gel is shown in Figure 5.18.

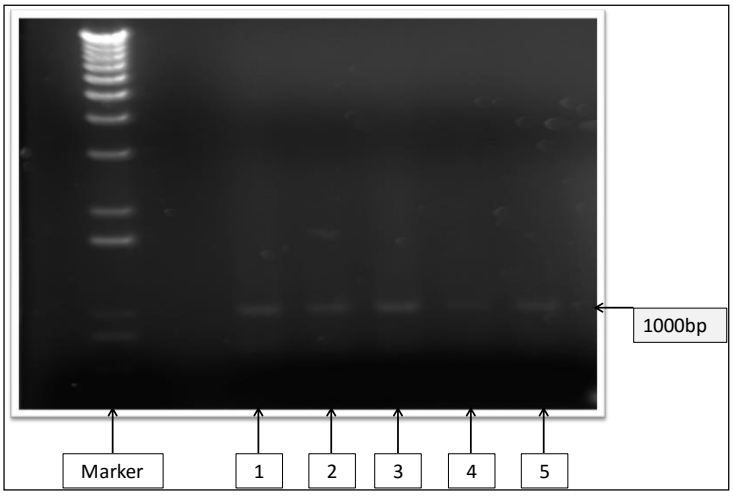


Figure 5.18: PCR for MARCKS using cDNA reverse transcribed from DLKP-M RNA. Bands are visible in the sample set (1-5 lanes) at ~1000bp. MARCKS has a predicted size of 1010bp based on primer design.

These bands were excised from the gel and digested to create sticky ends in preparation for ligation into a previously linearized overexpression vector. The newly digested fragment for MARCKS coding sequence was ligated into the pcDNA3.1 overexpression vector before being amplified by DH5α *E. coli.* The following day the bacteria cultures were miniprepped to extract the amplified vectors. Following this, vectors were digested using the restriction enzymes designed into the primers to assess if they had incorporated the MARCKS coding sequence of expected size (~1000 bp) along with the vector backbone (5600 bp). The resulting products were separated on an agarose gel and visualised (Figure 5.19).

The vector sample from lane 2 was sequenced which confirmed that it has 100% homology to the known ORF sequence for MARCKS. Next this vector was midi-prepped to scale it up and remove endo-toxins in preparation for transfection. In tandem, an empty vector was also midi-prepped in the same way to act as a control for all future transfections.

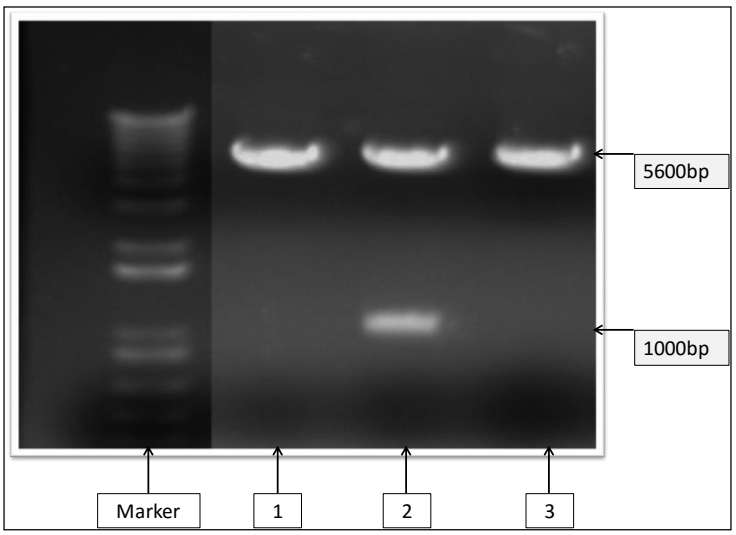


Figure 5.19: Three bacteria colonies grew successfully on Ampicillin LB agar plates, and each was picked and miniprepped. Each was digested with KPN1 and XHO1 and products were run on a 0.8% agarose gel. Lane 2 resulted in the expected fragment sizes of 5600bp for vector backbone and ~1000bp for MARCKS.

To assess the role MARCKS may play in the DLKP cell line model, clonal subpopulations: DLKP-SQ, DLKP-M, DLKP-I and DLKP were transfected with the newly created vector with the goal of overexpressing the protein. Cells were put under antibiotic selection to generate stable over-expressing clones. Each clone was transfected with both the MARCKS overexpression vector (O.E), and an empty vector control (E.V). Untreated cells which were not transfected at all are included as a control. Results of western blot analysis are shown in Figure 5.20.

DLKP-SQ showed no expression of MARCKS in any of the three samples which were analysed by western blot. This cell line does not express MARCKS as shown in Figure 5.3A. DLKP-M samples showed low level MARCKS expression in each of the three samples, all at the same level of band intensity. The low level of MARCKS expression is unusual and may be as a result of low exposure during the imaging of the western blot. DLKP-I samples showed a slight increase in MARCKS expression in the overexpression sample compared to both the Empty Vector and untreated cells control. This western blot has only been carried out once and will need to be repeated to verify the findings, at least in DLKP-I. Functional assays will then be carried out to look for phenotypic changes related to the overexpression of MARCKS.

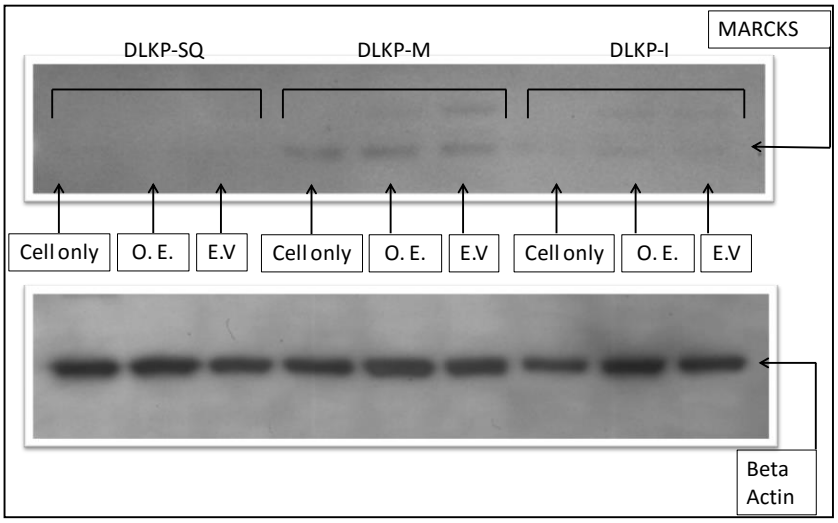


Figure 5.20: Western Blot probed for MARCKS. Each DLKP clone was transfected with MARCKS over expressing vector (O.E) and an empty vector control (E.V). Untreated cells are included also.

5.8. Cell Line Panel: MARCKS

Western blot analysis was used to determine the presence and expression levels of MARCKS in a panel of cell lines. Results are shown in Figure 5.21.

- **Pancreatic cell lines**: PANC-1, AsPC-1, BxPC-3 and MIA PaCa 2.
- **Lung cell lines**:A549, SK-LU-1, and NCI-H82.
- **DLKP clones**: DLKP-SQ, DLKP-M, and DLKP-I.

MARCKS showed its characteristic expression pattern in the DLKP clones, being highest in DLKP-M, followed by DLKP-I. Little to no expression was visible in DLKP-SQ. The pancreatic cell lines showed expression of MARCKS in BxPC-3 and PANC-1 cell lines. Little or no expression was seen in AsPC-1 or MIA PaCa 2. Both cell lines expressing MARCKS are known to have an invasive phenotype, whereas the cell lines with no expression of MARCKS are considered lowly invasive. The three lung cancer cell lines all showed expression of MARCKS. SK-LU-1 had the highest expression followed by A549 which showed low levels of expression. NCI-H82 had extremely faint bands for MARCKS indicating an extremely low level of expression in that cell line.

Western Blot: MARCKS

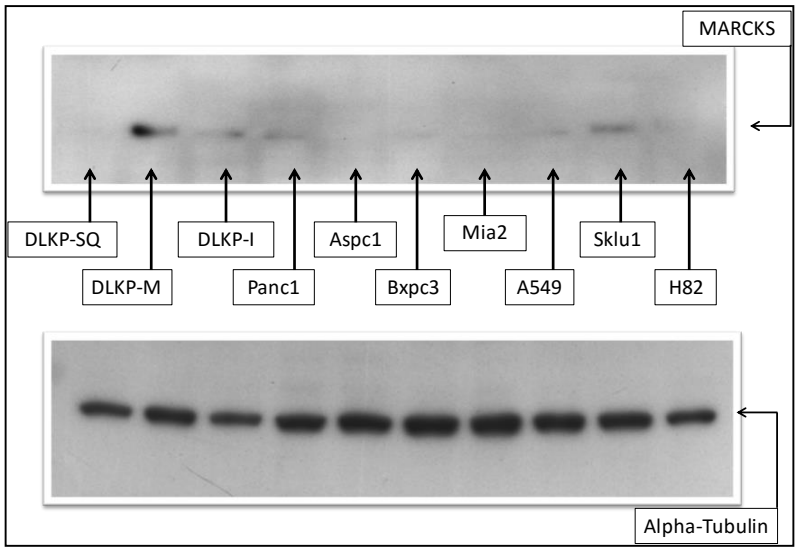


Figure 5.21: Western Blot probed for MARCKS in a panel of cell lines. Alpha-Tubulin loading control included.

5.9. Summary

MARCKS was identified as a significantly differentially expressed protein when the DLKP cell line and its clonal subpopulations were compared against each other using quantitative label-free LC-MS/MS. The DLKP-M cell line was found to have the highest abundance of MARCKS for both exponential and stationary phases of growth. The difference in fold change was shown to be very high between the highest expresser in DLKP-M compared to the lowest in DLKP, at 49.82 during exponential phase, and this expression pattern was validated at a protein and mRNA level. Immunofluorescent staining of MARCKS in each cell line further validated this expression pattern, and found MARCKS to localise to invadopodial-like structures in DLKP-M and subpopulations within DLKP, similar to what was found for Shootin-1. Punctate staining is also visible for MARCKS in all cell lines, suggesting it may co-localise with discrete focal contacts in these cell lines as this protein is tethered to the plasma membrane in its unphosphorylated form.

Transient knockdown of MARCKS protein in DLKP-M was found to have a significant effect on the migratory and invasive abilities of this cell line. Both of these abilities were significantly reduced when MARCKS expression was reduced in DLKP-M compared to the negative control samples. No effect was found on proliferation upon MARCKS knockdown in DLKP-M. Further knockdown studies on MARCKS were carried out to look for an effect on the behaviour of F-actin in DLKP-M, as there are reported associations between MARCKS and actin in neuronal cells and hepatic stellate cells (Li et al. 2008a, Rombouts et al. 2013). Both proteins were co-stained and imaged using confocal microscopy. In control cells, co-localisation of MARCKS and F-actin were found, with high staining for both proteins throughout the cytoplasm and at the edges of cells. Co-localisation of punctate staining is also visible, which may indicate both proteins form a complex in DLKP-M cells. Co-staining of MARCKS and F-actin in transient knockdown samples revealed a potential translocation of remaining MARCKS to the nuclei of DLKP-M cells. The cytoplasmic staining of MARCKS was reduced by the knockdown process, however the nuclei of the cells show visibly enriched MARCKS expression, suggesting a translocation effect. This observation requires further investigations to determine if nuclei are enriched for MARCKS and will be completed as future work. In addition, co-staining of MARCKS and F-Actin showed

a reduction in F-actin staining intensity in knockdown samples. This reduction is slight however, and was not validated by western blot probing for Beta-actin. This will also need to be investigated further.

Overexpression of MARCKS was attempted through the construction of an overexpression vector. The coding sequence for MARCKS was successfully extracted from DLKP-M cells and amplified. It was ligated into an overexpression vector and DNA sequencing confirmed successful construction of the vector. This was transfected into the DLKP clones, however western blot analysis did not find successful overexpression of MARCKS in DLKP-SQ; DLKP-M; or DLKP-I. Time did not allow for further exploration of this side of the MARCKS investigation, however it will be performed again in the future, and followed up with functional assays.

Chapter 6. Desmoglein-3

6.1. Desmoglein-3

Desmoglein 3 (DSG3) was identified in the quantitative label-free study as a very specifically highly expressed protein in DLKP-I. Little or no protein is present in the other cell lines, potentially meaning the protein could be a marker for DLKP-I. DSG3 is a transmembrane glycoprotein exclusively expressed in stratified epithelium. It is member of the desmoglein family of proteins, which in turn are a part of the cadherin superfamily of cell adhesion molecules. Expression of DSG3 is limited to the basal and supra-basal layers of the skin, however it is expressed homogeneously throughout the stratified epithelium in oral mucosa (Amagai et al. 1996, Garrod and Chidgey 2008). DSG3 is a major component of the desmosome, which is an adhesive intercellular junction critical to the integrity of tissues that experience shear stress. Homophilic binding of the DSG3 extracellular domain to adjacent cells leads to cell-cell adhesion, while the cytoplasmic tail of DSG3 links to internal keratin intermediate filaments via linker proteins such as the armadillo and plakin families (Dehner et al. 2014, Garrod and Chidgey 2008, Delva, Tucker and Kowalczyk 2009). DSG3 has been extensively studied in relation to the auto-immune disease *Pemphigus Vulgaris* (PV), in which autoantibodies bind to the extracellular domain of DSG3 and cause internalization of the protein (Jennings et al. 2011, Sato, Aoyama and Kitajima 2000). This results in a signalling cascade which leads to loss of cell-cell adhesion in the basal and immediate supra-basal layers of stratified squamous epithelia as well as the oral mucosa, causing painful blistering and lesions.

DSG3 has been found to be overexpressed in head and neck cancer (HNC) by PCR analysis in tissue samples when compared to normal keratinocytes and other cancer types (bladder, neuron, colon and liver) (Chen et al. 2007). Using immunohistochemical staining, DSG3 was found to be redistributed to the cytoplasm from the membrane in the majority of ESCC cases (Wang et al. 2007). This paper hypothesised that DSG3 may play a role other than functioning as a cell adhesion molecule, and may be involved in cell migration and invasion. DSG3 has also been studied in lung cancer. Immunohistochemical studies aiming to identify markers for the differential diagnosis of non-small-cell carcinomas of the lung identified DSG3 as a potentially useful marker for the identification of squamous cell carcinoma (Gomez-

Morales et al. 2013). No in vitro functional studies have been carried out in lung cancer investigating DSG3 to date.

6.2. Label-Free Proteomic Analysis: Desmoglein-3

Label-free proteomic profiling was carried out on DLKP and clonal subpopulations in order to discover proteins exhibiting significant differentially expressed proteins between the cell lines. The resulting proteomics data was screened to find proteins showing large abundances in a single clonal subpopulation compared to the others using the protein expression profile view shown in Figure 6.1. The purpose of highlighting such proteins is discussed previously in Section 4.2. One such protein which shows large protein abundance difference in DLKP and clonal subpopulation lines is DSG3.

6.2.1. Global Analysis: Desmoglein-3

DSG3 was highlighted by the label-free profiling work which quantitatively analysed the DLKP cell line and its clonal subpopulations and compared the data in one large experimental design. This design is termed a 'Global Analysis' and compares DLKP-SQ, DLKP-M, DLKP-I and DLKP against each other as one large group at both exponential and stationary phase of growth. DSG3 stood out as one of the top fold change proteins discovered by this analysis during exponential phase of growth only.

- During exponential phase of growth, DSG3 has a maximum fold change of **175.33** during with highest abundance in DLKP-I and lowest abundance in DLKP-SQ.

- Three peptides contributed to the identification of DSG3.

- During stationary phase of growth, DSG3 was not highlighted as a statistically significantly differentially expressed protein.

DSG3 was positively identified by mass spectrometry by passing the stringent criteria of: ≥ 2 peptides contributing to the protein identification, an abundance change of ≥ 2 fold between two experimental groups, an ANOVA p-value of <0.05 attributed to a protein between experimental groups. A representative protein abundance graph from the Progenesis software for DSG3 can be seen below in Figure 6.1. This image shows the normalised abundance values for DSG3 across the DLKP cell line and its clonal subpopulations. During exponential phase of growth, highest abundance of DSG3 is in DLKP-I compared to all other cell lines in the global analysis. Next highest abundance is seen in DLKP-M and DLKP, both showing similar levels judging by the protein expression profile. Very low level abundance is in DLKP-SQ. Following this initial assessment, the levels of DSG3 expression were analysed using experimental designs which compare the cell lines against each other using a one vs. one method instead of as one group: e.g. DLKP-SQ vs. DLKP-I. Peptides contributing to the identification of DSG3 are shown in Figure 6.2.

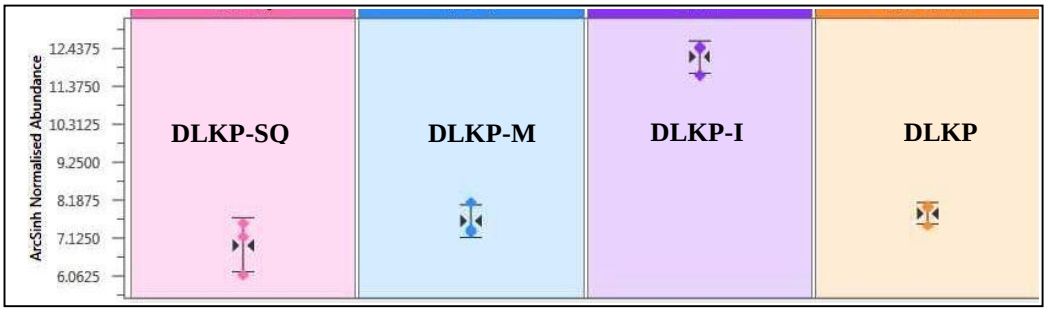


Figure 6.1: Peptide expression profile for DSG3 during exponential phase of growth. Highest expression of DSG3 can be seen in DLKP-I, with a 175.33-fold difference between DLKP-I and the lowest abundance found in DLKP-SQ (ANOVA, p-value is 5.94E-06).

Accession: P32926

Description: Desmoglein-3 OS=Homo sapiens GN=DSG3 PE=1 SV=2 - [DSG3_HUMAN]

Σ	#	Score	Anova (p)	Max Fold Change	Highest Mean	Lowest Mean	Tag	Abundance	m/z	Charge	Retention Time (mins)	Mass error (ppm)	Peptide Sequence	Modifications
•	58757	40.930	6.97E-10	284	DLKP I	DLKP SQ		4.563E+04	777.4543	2	169.66	0.19	VQSTPVTIQVINVR	
•	62910	47.580	6.45E-05	245	DLKP I	DLKP M		3.322E+04	750.3397	2	94.669	0.34	GTAVEGTSGMEMTTK	
•	84123	40.770	5.33E-06	1.84E+05	DLKP I	DLKP Parents		2.609E+04	805.8433	2	97.826	0.52	DGEGLSTQCECNIK	[9] C+57.02[11] C+57.02

Figure 6.2: Peptides which contributed to the identification of DSG3 as shown by Progenesis software. Each peptide had an ANOVA $p < 0.05$, and MASCOT scores ≥40.

6.2.2. Two-Sample Experimental Designs: Desmoglein-3

DSG3 was found to be expressed with very high abundance in a single clone population when samples were grouped together in a global analysis. The expression profile in Figure 6.1 shows DSG3 as being highly expressed in DLKP-I when compared to DLKP and the remaining clonal subpopulations. In this analysis, DSG3 expression is determined using an approach which compares individual cell lines against each other in a one vs. one experimental design. Each cell line is compared to another individually, using Progenesis software as described previously in Section 4.2.2. This analysis results in the availability of individual fold change data between cell lines with the associated statistical information. Intermediary fold change data can be obtained using this analysis. The usual strong criteria are applied to the resulting proteins to ensure strong identifications with high statistical significance as previously described in Section 3.1.2. Results of this analysis for DSG3 are summarised in Table 6.1.

Each experimental design which contains DLKP-I in the comparison highlights DSG3 as the top fold change result with greatest abundance in DLKP-I. The protein is not highlighted as a significantly differentially expressed protein in comparisons which do not include DLKP-I in the experimental design. This indicates that DSG3 is a specifically highly expressed protein in DLKP-I during exponential phase. DSG3 was identified as highly expressed in DLKP-I during stationary phase, however only one peptide contributed to this identification and was therefore screened out.

- During exponential phase of growth DSG3 is highlighted only in the experimental designs which contain DLKP-I, indicating that the protein is very specifically expressed in DLKP-I compared to the other cell lines.

- DSG3 shows the largest abundance difference between DLKP-I and DLKP-SQ with a fold change of **175.33**, highest in DLKP-I. The next highest abundance is in DLKP followed finally by DLKP-M.

Exponential Phase:

Individual	Comparison	Description	Max fold change	Highest Abundance	Lowest Abundance
DLKP-SQ	vs. DLKP-M	-	-	-	-
DLKP-SQ	**vs. DLKP-I**	**Desmoglein-3**	**175.33**	**DLKP-I**	**DLKP-SQ**
DLKP-I	**vs. DLKP-M**	**Desmoglein-3**	**97.31**	**DLKP-I**	**DLKP-M**
DLKP-SQ	vs. DLKP	-	-	-	-
DLKP-M	vs. DLKP	-	-	-	-
DLKP-I	**vs. DLKP**	**Desmoglein-3**	**81.55**	**DLKP-I**	**DLKP**

Table 6.1: Individual comparisons of the DLKP cell line and clonal subpopulations for exponential phase of growth. In each comparison containing DLKP-I, DSG3 was highlighted as having the highest expression in that cell line. For each comparison, ≥ 2 peptides contributed to the identification, and ANOVA p-values were <0.05 in the protein abundance comparisons.

6.2.3. Validation of DSG3 Expression in DLKP and Clonal Subpopulations

To validate the expression pattern for DSG3 found by the label-free analysis comparing DLKP and the clonal subpopulations, each cell line was analysed by Western Blot. Samples were prepared as described in Section 2.3.1 and probed for DSG3 using the appropriate primary antibody. Western Blots analysis revealed bands at a MW of 130 kDa which is the expected weight of this protein (Koga et al. 2013).

Highest expression of DSG3 is in DLKP-I, with no detectable DSG3 in any of the other clones. A representative image of the western blot analysis is shown in Figure 6.3A of DSG3 expression in the DLKP cell lines. For comparative purposes, a graphical representation of the label-free abundances for DSG3 is shown in Figure 6.3B.

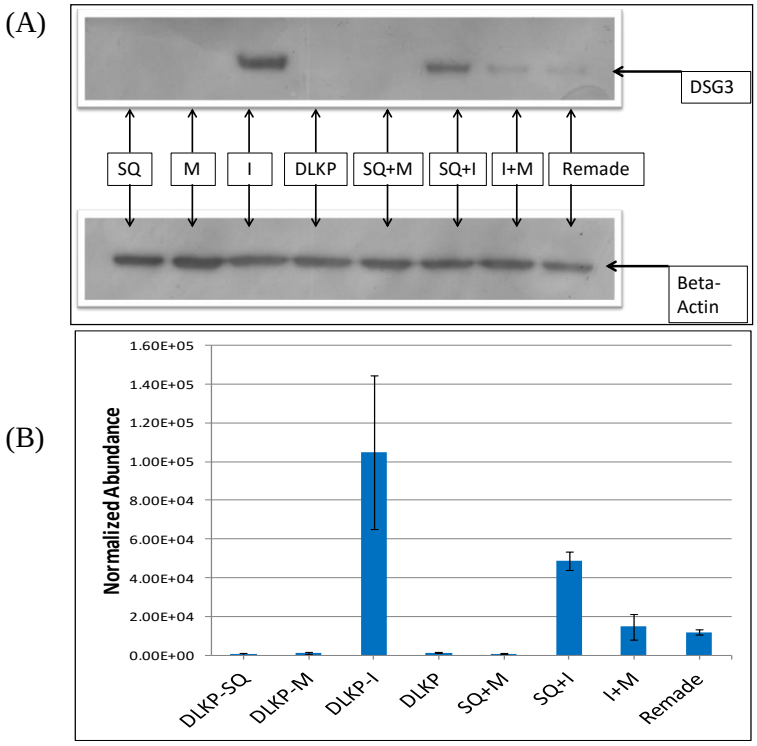


Figure 6.3: (A) Western Blot showing expression levels of DSG3 in all DLKP cell lines. Beta-Actin was used as a loading control. DSG3 bands are visible at ~130kDa in the DLKP model. (n=3). **(B)** Label-free abundance values for DSG3 in the DLKP cell line and clonal subpopulations. Standard deviations are shown (n=3).

6.3. Proteomic Comparison of DLKP Compared to the Three Clones and Combination Cell Lines

DSG3 was established as highly expressed in the DLKP-I clonal subpopulation in an analysis comparing the cell lines together as a group. Further analysis showed the individual fold changes between the high expressing DLKP-I clone against the remaining cell lines. To further explore the expression pattern of DSG3 using the label-free profiling data, an experimental design was set up which incorporates all samples into the analysis. This design includes the 'combination lines' which are newly created cell lines made up of 50:50 ratios of any two clonal subtypes, as well as a reconstituted version of DLKP based on previous work (McBride et al. 1998b). Details of the combination cell lines are described in Section 3.1.1.

A large global group analysis was created in Progenesis software which incorporates all samples into one large experimental design, consisting of 24 individual samples representing eight different conditions in biological triplicate. The combination lines inclusion into the analysis was done to determine if co-culture of the DLKP clonal subpopulations would have any effect on the expression of proteins of interest. Such effects include inhibited or increased expression of protein targets. In this analysis, DSG3 is highlighted as a statistically significantly differentially expressed protein during exponential phase, and its expression in the combination lines is available for analysis also.

Peptides which belong to DSG3 and have contributed to its identification and quantification are displayed by Progenesis software (Figure 6.2). Each peptide had an ANOVA p-value <0.05, and Mascot scores ≥ 40. Each peptide showed highest abundance in DLKP-I. These peptides are grouped together by the software and quantification data is displayed as a peptide expression profile. Label-free data can be displayed as a peptide or protein expression pattern. The peptide expression view (Figure 6.4) allows for the detection of peptide outliers in the data which may skew the results. In the case of DSG3, the peptides cluster together well, indicating a strong quantitative analysis for the protein identified. The abundance levels of DSG3 are significantly higher in DLKP-I than all other samples in the analysis. Moreover, the combination lines which contain DLKP-I as a component show a higher abundance of DSG3 when compared to samples which do not contain that cell line. The two

combination lines containing DLKP-I as a component are: SQ+I and I+M. The peptide expression profile for DSG3 shows that these two samples do not express the same level of the protein as would be expected. This is due to the fact both DLKP-SQ and DLKP-M show little to no expression of DSG3 based on the data described in Section 6.2.1. This unusual expression pattern in the combination lines suggests that co-culture of DLKP-I with DLKP-M results in decreased expression of DSG3.

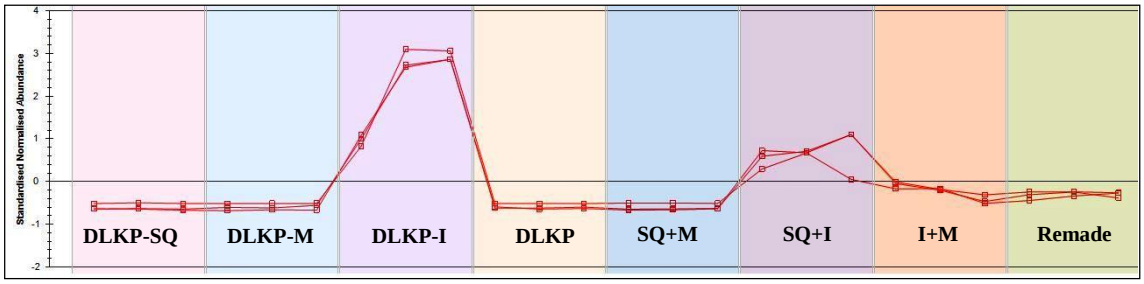


Figure 6.4: Peptide expression profile for DSG3 across all DLKP cell lines and combination lines. Peptide abundance for DSG3 show clear differences between clonal cell lines, as well as the combination lines. Highest abundance is in DLKP-I and lowest in DLKP-SQ (ANOVA p-value= **1.44E-10**).

6.4. Validation of Desmoglein-3 Expression in DLKP plus the Clones and Combination Cell Lines

6.4.1. Western Blot

Western Blot analysis was performed on the DLKP cell lines along with the combination lines to determine if the expression pattern found by the label-free analysis could be validated. The label-free analysis had revealed an unexpected abundance pattern of DSG3 in the combination line samples. SQ+I and I+M would be expected to result in similar abundance levels. However, the label-free data reveals a lower abundance of DSG3 in I+M than is found in SQ+I. This suggests an inhibition of DSG3 expression by the DLKP-M clone when it is in co-culture with DLKP-I. Western blot analysis validated this expression pattern and can be seen in Figure 6.3 A, with SQ+I showing higher expression of DSG3 than I+M. A graphical representation of the label-free abundances for DSG3 in DLKP clones and combinations is shown in Figure 6.3B for comparison.

6.4.2. mRNA Expression

Western blot analysis validated the expression of DSG3 found by the label-free proteomic profiling. To further investigate this expression pattern, DLKP samples were analysed using quantitative PCR (qPCR) to determine the relative levels of DSG3 mRNA expression in DLKP, clones and combination cell lines. Relative gene quantification was performed on all cell lines as one experiment, and beta-actin was used as the endogenous control. Biological duplicate samples were analysed and the experiment was carried out twice. The calculation of relative quantity (RQ) was achieved by normalizing the data to DLKP-I. Results are combined in Figure 6.5.

mRNA levels of DSG3 correlated well with label-free proteomics data as well as the western blot analysis. Highest mRNA level can be seen in DLKP-I, further showing the target as a potential marker for this cell type. mRNA for DSG3 in DLKP-SQ and DLKP-M is virtually non-existent. There is a low expression level of the target in the DLKP cell line, which is expected as DLKP is made up of approximately 25% DLKP-I

(McBride et al. 1998b). The mRNA abundance in the combinations correlated well with the findings from the proteomic analysis. SQ+M shows no expression of DSG3, while both SQ+I and I+M have detectable levels of the target. I+M has approximately half the abundance of mRNA levels for DSG3 compared to SQ+I. This further supports the hypothesis that co-culture of DLKP-I with DLKP-M has inhibited the expression level of DSG3 at the protein, and now mRNA level.

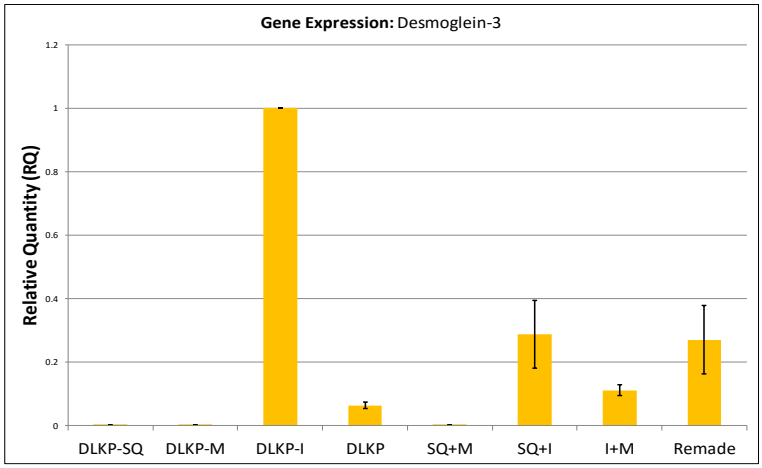


Figure 6.5: qPCR results showing mRNA levels of DSG3 in the clones and combination lines. Samples are analysed in biological duplicate (n=2). Error bars shown are the standard deviations between two replicate experiments. Data are normalized to DLKP-I.

6.5. Imaging of Desmoglein-3 in the DLKP Cell Lines

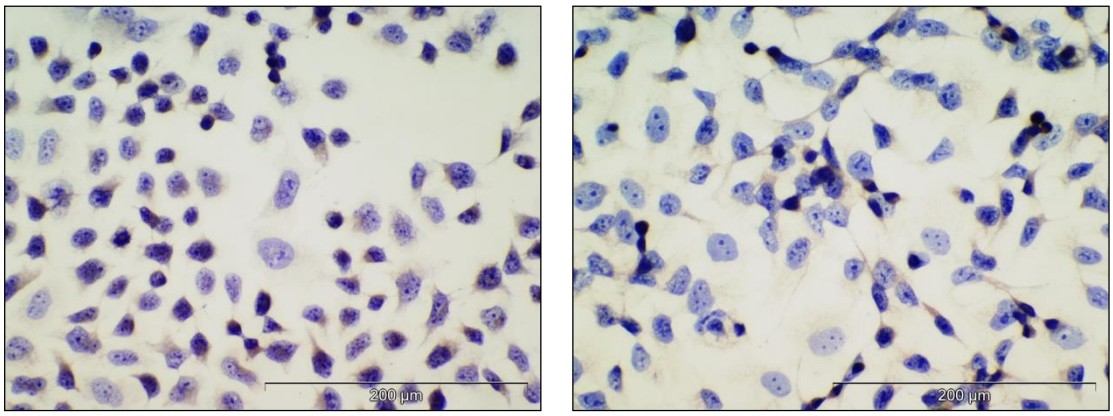
DSG3 was validated as being a significantly differentially expressed target at the protein and mRNA level, showing very high abundance in the DLKP-I clonal subpopulation. In order to investigate the localization of this protein in the DLKP cell line model, as well as further validate the expression pattern, DSG3 was stained and imaged using both immunocytochemistry and immunofluorescence. Staining methods were performed as described in Section 2.6 on DLKP and clonal subpopulations fixed to glass slides by paraformaldehyde.

6.5.1. Immunocytochemical Staining of Desmoglein-3 in DLKP Cell Lines

Expression of DSG3 in the DLKP cell line and clonal subpopulations shows drastic differences in protein abundance according to the label-free proteomics data. DLKP-I has the highest and only expression of DSG3 with the other clonal subpopulations showing no detectable levels of the protein by western blot analysis or qPCR analysis. The staining intensity of DSG3 in the DLKP cell lines by immunocytochemistry correlated very well with these findings and resulting images are shown in Figure 6.6.

DLKP-I shows extremely high intensity of DSG3 staining compared to the other cell lines, further validating its specific expression in DLKP-I. Localization of the protein can be clearly seen in the cell-cell junctions of actively dividing DLKP-I cells, which is expected due to its known role as a component of the desmosome (Garrod and Chidgey 2008). In normal stratified epithelium, DSG3 expression is predominantly in the cell membrane. However, strong cytoplasmic and perinuclear staining is clearly visible in DLKP-I cells. This cytoplasmic localization has previously been found in oral squamous cell carcinoma (OSCC) and in esophageal squamous cell carcinoma (ESCC). In both cases, this abnormal cytoplasmic localization correlated with poor prognosis (Wang et al. 2007, Fang et al.). DLKP-SQ and DLKP-M shows no staining for DSG3, while the Dapi nuclear stain remains clearly visible. The DLKP cell line itself shows a subpopulation of cells which stain positively and strongly for DSG3. This is most likely the DLKP-I subpopulation which makes up approximately 25% of the DLKP cell line as described in Section 1.4.1.2.

DLKP-SQ **DLKP-M**



DLKP-I **DLKP**

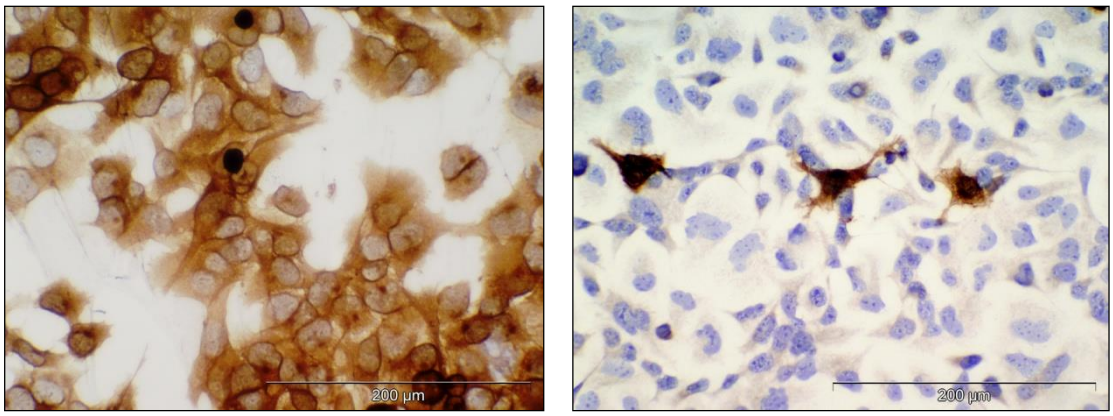


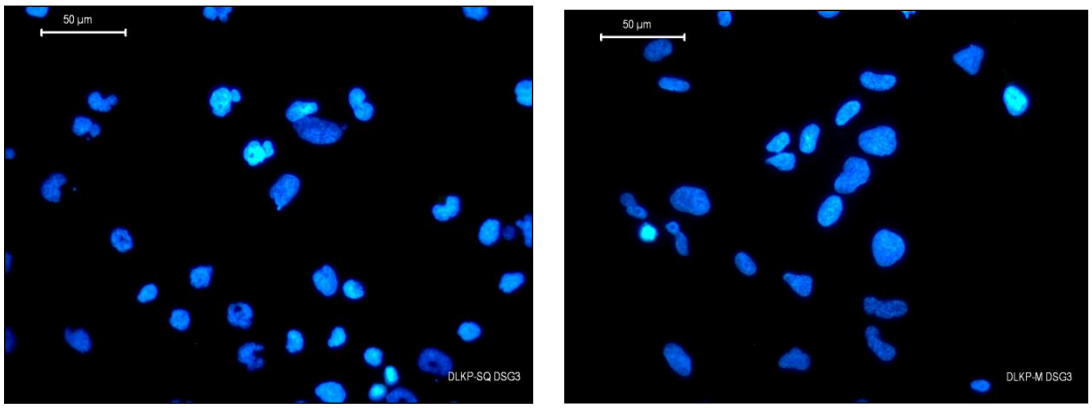
Figure 6.6: Immunocytochemistry staining images of DLKP-SQ, DLKP-M, DLKP-I and DLKP stained with DSG3. Nuclei are counterstained with Dapi. Negative controls with secondary antibody only did not result in staining (images not shown). Original magnification used on all micro-pictographs shown at 200x, 200µm scale bar (n=2).

6.5.2. Immunofluorescence Staining of Desmoglein-3 in DLKP Cell Lines

Immunofluorescence staining of DSG3 was carried out in DLKP and each of its clonal subpopulations to allow visualization of the distribution of the protein in each cell type, as well as a semi-quantitative assessment of protein abundances. Staining was carried out using the method described in Section 2.6.3, with resulting images shown in Figure 6.7.

Immunofluorescence staining of DSG3 followed the same trend as was found for immunocytochemistry. The DLKP-I cell line has the most intense staining of all the samples, and is the only cell line from the clonal subpopulations to stain positively for the protein. The localization of DSG3 in this cell line is both membranous and cytoplasmic in nature. There is clear enhanced staining for the protein at cell-cell junctions which is expected. In addition, there is punctate staining present on the cell surface which may be the DSG3 component of the desmosomes dotted across the cell membrane. Perinuclear and cytoplasmic staining is also visible, which is an abnormal localization for this protein (Chidgey and Dawson 2007). Both the DLKP-M and DLKP-SQ cell lines show no detectable levels of DSG3 staining, which correlates well with the findings for the protein so far. This stark difference in protein expression is quite unusual as the clonal cell lines are derived from the same heterogeneous population (DLKP). The DLKP cell line shows positive staining for DSG3 in a small subpopulation, which is most likely DLKP-I. However, the morphology of this cell line is slightly different from the clonal DLKP-I cell line. This indicates that in the heterogeneous population, co-culture conditions play a significant role on the behaviour of the cells. Investigating cellular heterogeneity and how subpopulations interact in lung cancer may lead to improved treatments and the discovery of useful prognostic markers.

DLKP-SQ **DLKP-M**



DLKP-I **DLKP**

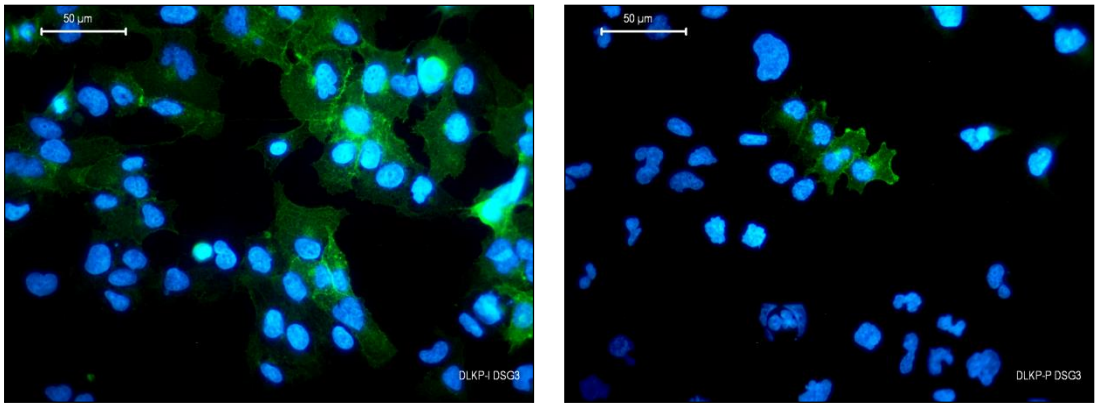


Figure 6.7: Immunofluorescence images stained for DSG3 in DLKP-SQ, DLKP-M, DLKP-I and DLKP. Nuclei are counterstained with Dapi. Negative controls with secondary antibody did not result in staining (images not shown). Original magnification used on all micro-pictographs shown at 400x, 50µm scale bar (n=2).

6.6. Functional Analysis of Desmoglein-3 Knockdown by RNAi

DSG3 plays an unknown role in the DLKP cell line model. It shows extremely high expression in one clonal subpopulation only: DLKP-I. Expression in the remaining clonal cell lines is undetectable by both western blot analysis and gene expression assays. In order to investigate the role of DSG3 at a functional level, knockdown of the protein was carried out using RNA interference (RNAi). DLKP-I was used as the cell line to carry out the investigation of this process as it expresses the highest abundance of DSG3 in the DLKP cell line model. The effects of DSG3 knockdown were determined by a panel of functional assays and comparing test samples against a negative siRNA control. In this case a standard commercially available negative siRNA was used for this purpose. Previous to the knockdown of DSG3, cells were first tested to determine the optimal transfection conditions for DLKP-I. Conditions were optimised as described in Section 2.4.1.

6.6.1. Transfection Optimisation

6.6.1.1. Optimisation of Transfection Reagent Volume:

The transfection reagent volume necessary to effectively introduce siRNA in cells is an important optimisation step previously described in Section 4.7.1. Prior optimisation narrowed down the required transfection reagent volume to 1 µl or 2 µl per 6-well, and both these volumes are tested here. Post-transfection cell viability was measured using an acid phosphatase proliferation assay on a series of controls and tests as described previously in Section.4.7.1. Results are shown in Figure 6.8.

In the case of DLKP-I, both of the volumes of transfection reagent tested had a negligible effect on the Lipo-only control compared to the Cells-only condition. This indicates that neither volume is detrimental to the survival of DLKP-I cells. The negative siRNA control similarly shows no reduction in cell proliferation or viability which indicates that introducing non-coding siRNA into these cells has no effect on these conditions. The positive control siRNA is designed to induce cell death by targeting kinesin, a protein crucial to proper cell functioning. Both 1 µl or 2 µl volumes of transfection reagent effectively delivered the siRNA into DLKP-I cells, resulting in

~37% and ~62% cell death respectively. This result led to choosing the 2 µl volume of transfection reagent for future transfections with DLKP-I.

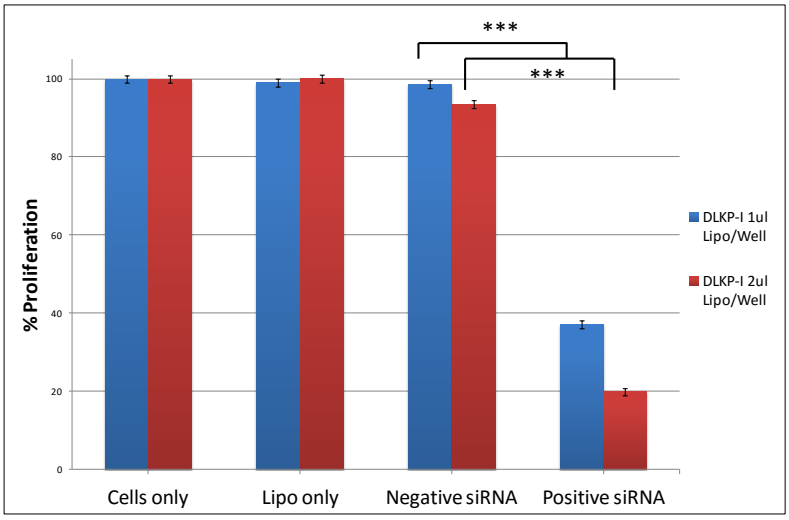


Figure 6.8: Proliferation assay performed on DLKP-I cells using 1µl/6-well (Blue) and 2µl/6-well (Red) of Lipofectamine transfection reagent. Percentage survival values are relative to the negative siRNA control. (1µl /well: P-Value 7.8E-03) (2µl/well: P-Value: 9.4E-04)

6.6.1.2. Optimisation of siRNA Concentration: Desmoglein-3

Optimisation of siRNA concentration for DSG3 is essential to ensure the target is being knocked down effectively. In addition, it is important to use the minimum concentration of siRNA which achieves this, as over saturating the cells with surplus siRNA can have additional effects aside from knockdown of the protein. Optimisation was carried out using the following conditions:

- Cells-only, Lipofectamine-only and Negative siRNA controls.
- 30 nM, 20 nM, 10 nM and 5 nM of siRNA for DSG3.
- Two different siRNA molecules were tested in tandem which target DSG3.

After transfection using the conditions described above, cell lysates were prepared as described in Section 2.3.1. Western blot analysis for DSG3 was carried out on the entire sample set and the resulting image is shown in Figure 6.9. In the first three lanes containing the control samples, DSG3 can be seen at 130kDa, with expression unaffected by the transfection process. Each siRNA group shows no expression of DSG3 at any of the concentrations tested. However, the blot is bad quality and has high background, therefore the experiment was repeated using the two lowest siRNA concentrations: 10 nM and 5 nM (Figure 6.10).

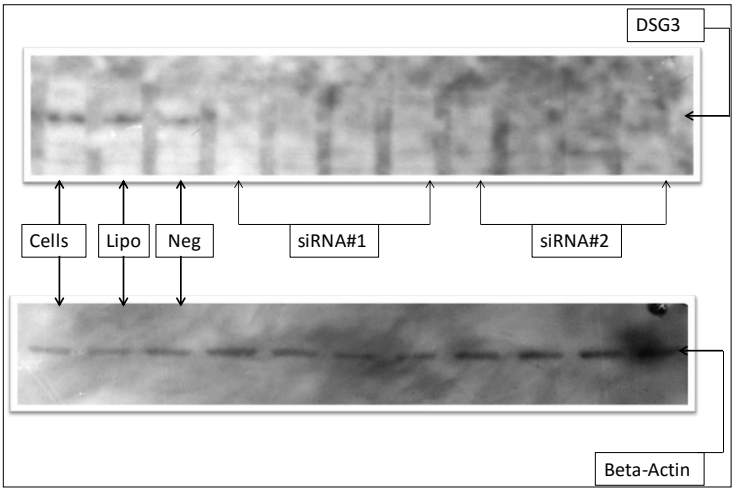


Figure 6.9: Western Blot probing for DSG3 (130 kDa). Lanes 1-3 contain: Cells-only, Lipofectamine-only and Negative siRNA (30 nM). Each DSG3 (DSG3) siRNA group (#1 and #2) is comprised of four sample lanes which were treated with 30 nM, 20 nM, 10 nM and 5 nM of siRNA from left to right. Beta-actin loading control included.

The transfection experiment was repeated once again with the middle concentrations of siRNA (20 nM and 10 nM), as these concentrations in the previous optimisation experiment seemed to result in complete knockdown of DSG3. A concentration of 5 nM was ruled out on the basis of practicability. This concentration requires a siRNA volume which is simply too small to be accurately and repeatedly pipetted in transfection experiments. The resulting western blot analysis of DSG3 in these samples is shown in Figure 6.10. The results show that both concentrations of siRNA effectively reduce the expression of DSG3 to undetectable levels. A concentration of 10 nM was chosen for future knockdown of DSG3 by RNAi.

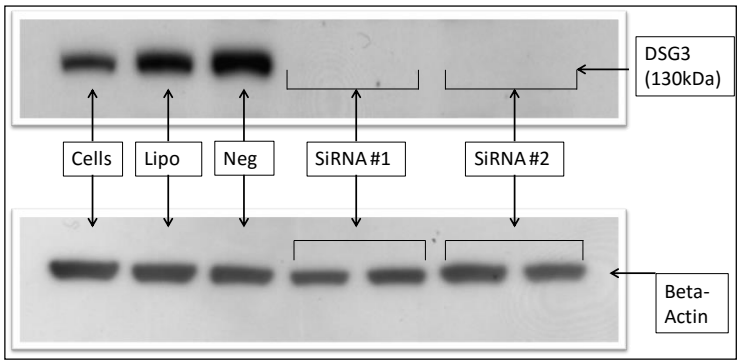


Figure 6.10: Western Blot probing for DSG3 (DSG3). Lanes 1-3 contain: Cells-only, Lipofectamine-only and negative siRNA (30nM). Each DSG3 siRNA group (#1 and #2) is comprised of two sample lanes treated with 20 nM and 10 nM of siRNA in the left and right lane respectively. Beta-Actin loading control included.

6.6.2. Functional Effects of Desmoglein-3 Knockdown on Proliferation, Migration and Invasion in DLKP-I

In order to shed light on the role played by DSG3 in DLKP-I, siRNA knockdown of the protein was carried out. This was followed up by a panel of functional assays. These assays measure proliferation, migration and invasion and significance is calculated by comparing the knockdown samples to the negative siRNA control. Each assay was carried out as described in Section 2.5.

6.6.2.1. Proliferation of DLKP-I Post-DSG3 Knockdown:

Proliferation of DLKP-I was measured 72 hours post transfection as described in Section 2.2.3. Knockdown of DSG3 in DLKP-I had no significant effect on the proliferation of the cells compared to the negative siRNA control (n=3). Results are shown in Figure 6.11.

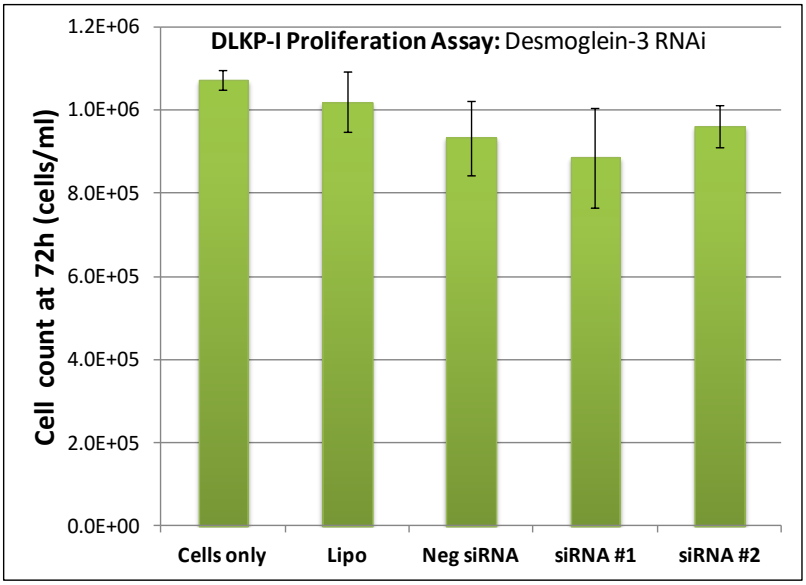


Figure 6.11: Proliferation assay on DLKP-I post transfection for DSG3 RNAi. There was no statistically significant difference in proliferation between siRNA knockdown samples (#1 and #2) and the negative control (n=3). (**siRNA#1 P-value**: 0.61, **siRNA#2 P-value:** 0.68)**.**

6.6.2.2. Migration of DLKP-I Post Desmoglein-3 Knockdown:

Migration assays on DLKP-I were carried out 72 hours post DSG3 knockdown by RNAi. The assay was performed using the method described in Section 2.5.4. DLKP-I cells were allowed to migrate through the uncoated membrane for 24 hours before being fixed and stained with crystal violet. Knockdown of DSG3 in DLKP-I had a significant inhibitory effect on the ability of the cells to migrate when compared to the negative siRNA control sample (n=3). Results are shown in Figure 6.12.

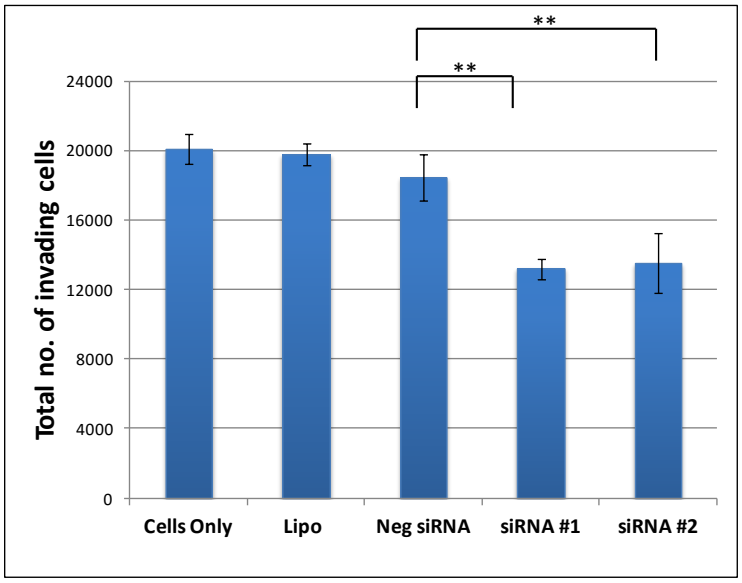


Figure 6.12: Migration assay results for DLKP-I cells following DSG3 knockdown using RNAi. Total number of cells which migrated through the uncoated membrane is shown (n=3). A statistically significant difference between siRNA knockdown samples (#1 and #2) and the negative siRNA control was found. (**siRNA#1 P-value**: 9.6E-03, **siRNA#2 P-value**: 1.0E-02).

6.6.2.3. Invasion of DLKP-I Post-Desmoglein-3 Knockdown:

Knockdown of DSG3 was performed 72 hours post transfection as described in Section 2.5.3. Cells were allowed to pass through the Matrigel coated membrane for 24 hours before fixation and staining by crystal violet. DSG3 knockdown had a significant inhibitory effect on the ability of DLKP-I cells to invade when they were compared to the negative siRNA control sample (n=3). Results are shown in Figure 6.13.

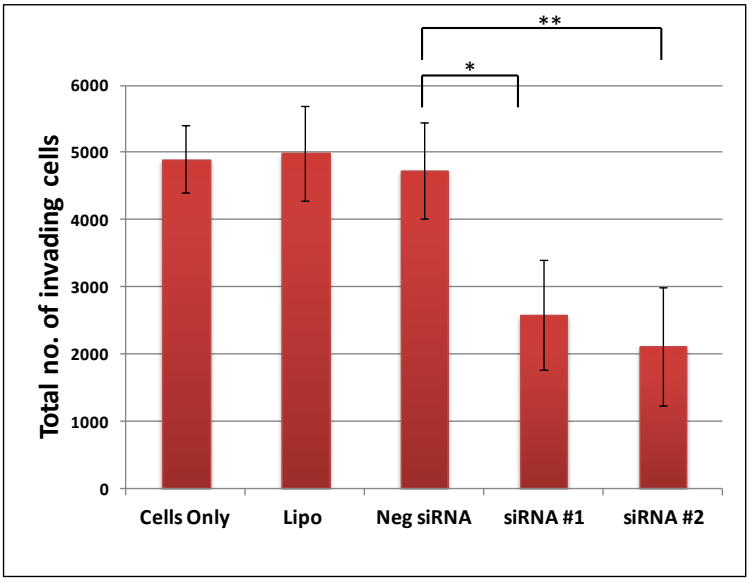


Figure 6.13: Invasion assay results for DLKP-I cells following DSG3 knockdown using RNAi. Total number of cells which migrated through the Matrigel coated membrane is shown (n=3). A statistically significant difference between siRNA knockdown samples (#1 and #2) and the negative siRNA control was found. (**siRNA#1 P-value**: 2.7E-02, **siRNA#2 P-value**: 1.5E-02).

6.6.3. Co-Staining of Desmoglein-3/F-Actin after Dsemoglein-3 Knockdown by RNAi in DLKP-I

DSG3 is a membrane bound protein which is predominantly found as a component of the desmosome. Its importance as a cell adhesion molecule is highlighted by its role in the autoimmune disease *Pemphigus vulgaris*, in which DSG3 has been shown to be the target of autoimmune antibodies (Koga et al. 2013). Upon binding of these antibodies to DSG3, a signalling cascade leading to the degradation of the protein from desmosomes is initiated. This in turn leads to the disorganisation of the actin cytoskeleton resulting in a loss of cell-cell adhesion which leads to blistering and lesions. In addition, knockdown of DSG3 in epidermoid carcinoma cells caused in a significant loss of cortical F-actin bundles, resulting in decreased cell height indicating a loss of contractility in cells with DSG3 depletion (Gliem et al. 2010, Tsang et al. 2012). These reported links between DSG3 and actin may be relevant to the phenotypic characteristics of DLKP-I, as knockdown of the protein has resulted in decreased migration and invasion. It was decided to investigate this link by staining both DSG3 and F-actin in DLKP-I and imaging the proteins by confocal microscopy. DLKP-I cells were also treated by RNAi to knockdown the expression of DSG3 and determine if there was an effect on F-actin organization, location or abundance. DSG3 expression and localization was also assessed.

DLKP-I cells were prepared as previously described in Section 4.7.3 and were co-stained with Dapi nuclear stain. Resulting images are shown in Figure 6.14 to Figure 6.17.

DLKP-I: Cells-Only

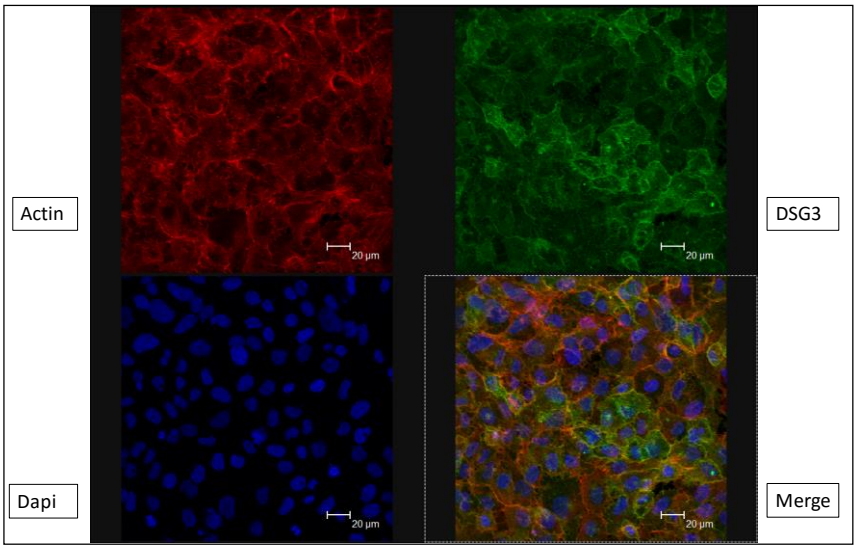


Figure 6.14: DLKP-I cells stained for DSG3, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are untreated by any RNAi method. Original magnification of all photomicrographs, ×400, scale bar = 20µm.

DLKP-I: Lipofectamine-Only Control

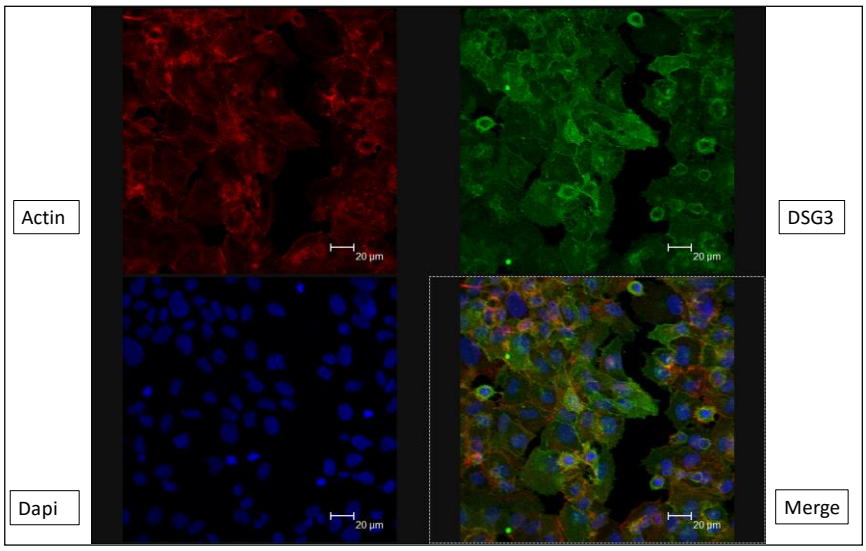


Figure 6.15: DLKP-I cells stained for DSG3, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are treated with transfection reagent only (Lipofectamine). Original magnification of all photomicrographs, ×400, scale bar = 20µm.

DLKP-I: Negative siRNA Control

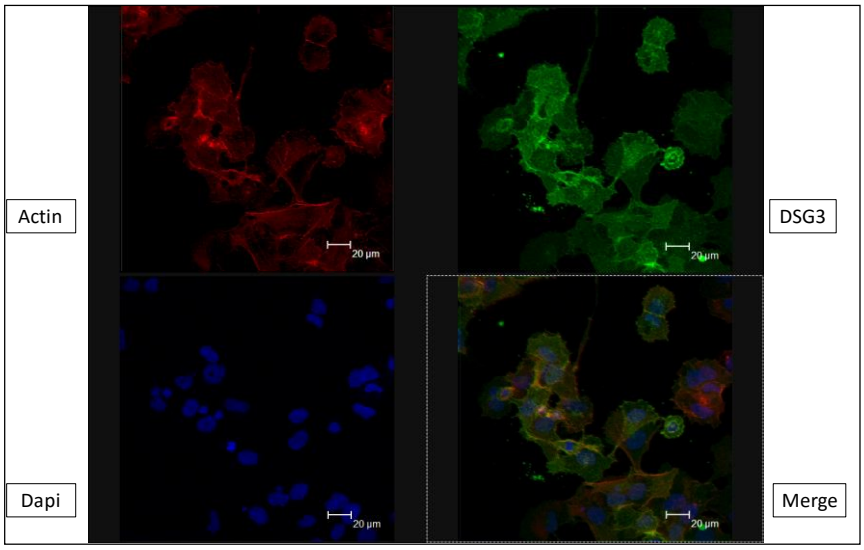


Figure 6.16: DLKP-I cells stained for DSG3, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are treated with transfection reagent and negative siRNA control. Original magnification of all photomicrographs, ×400, scale bar = 20µm.

DLKP-I: DSG3 Knockdown-siRNA #1

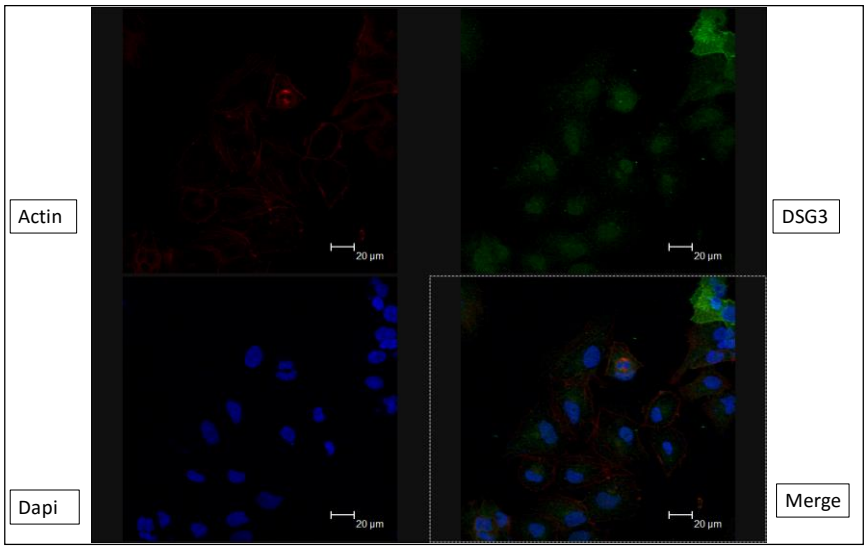


Figure 6.17: DLKP-I cells stained for DSG3, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are treated with siRNA#1 for DSG3. Original magnification of all photomicrographs, ×400, scale bar = 20µm.

DLKP-I: DSG3 Knockdown-siRNA #2

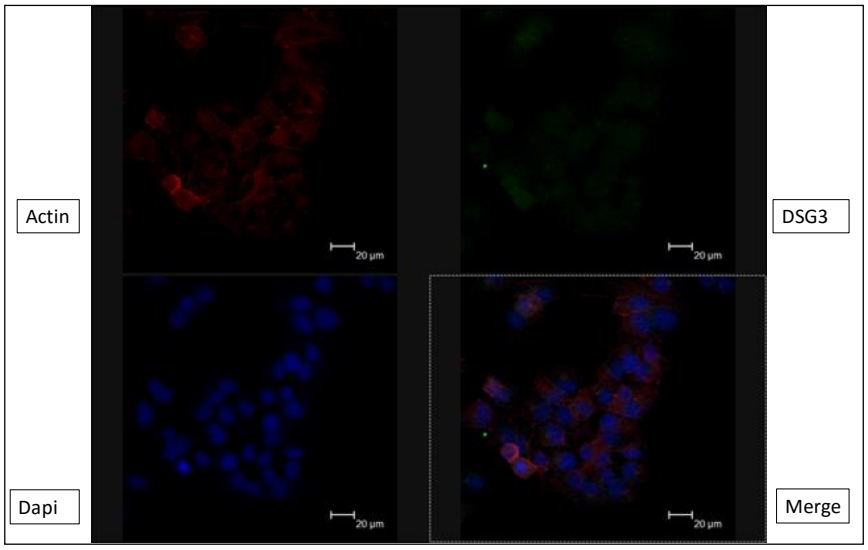


Figure 6.18: DLKP-I cells stained for DSG3, Actin and Dapi. A merged image is shown in the bottom right corner. Cells are treated with siRNA#2 for DSG3. Original magnification of all photomicrographs, ×400, scale bar = 20µm.

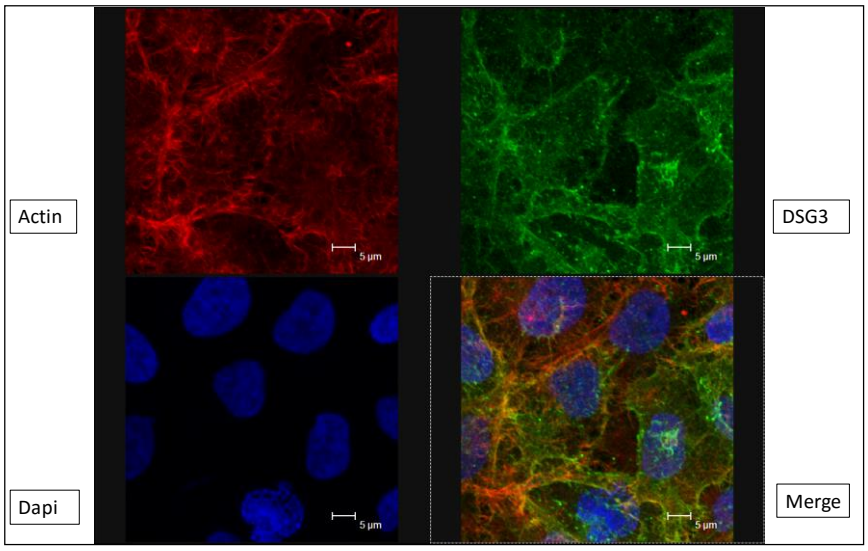


Figure 6.19: DLKP-I treated with negative siRNA control. Cells are stained for DSG3, Actin and Dapi. A merged image is shown in the bottom right corner. Intense staining of DSG3 can be seen at cell-cell junctions as well as punctate staining throughout the cytoplasm. Original magnification of all photomicrographs, ×800, scale bar = 5µm.

Confocal imaging of DLKP-I cells resulted in highly specific staining for both DSG3 and F-actin. This cell line clearly expresses a high abundance of DSG3 which

can be seen in the control group of images: Cells-only, Lipofectamine-only and Negative siRNA (Figure 6.14-Figure 6.16). In these groups, staining for DSG3 is visible throughout the cell. However, increased staining intensity can be seen at the periphery of many cells and at cell-cell junctions. This junction localization of the protein is presumably desmosomal bound DSG3 acting in its role as a cell adhesion molecule. There is also a significant amount of punctate staining of DSG3 present in the control group images which vary in size. Some are small and may be membrane bound desmosomal structures. Others are relatively large and are unlikely to be desmosomes, and therefore are possibly larger cell structures containing DSG3 as a component. F-Actin staining is high in the control group also, in many cases showing its characteristic filamentous expression pattern. Co-localization of F-actin and DSG3 can be seen where intense staining for DSG3 is present. The negative siRNA control image Figure 6.16 shows overlapping staining for both F-actin and DSG3 at cell-cell junctions and at fillopodial outgrowths.

DSG3 knockdown by RNAi (Figure 6.17 and Figure 6.18) resulted in a significant loss of staining intensity for the protein when compared to the control group. Faint perinuclear staining and potentially desmosomal punctate staining remains in the knockdown samples, but overall DSG3 expression is drastically reduced. F-actin expression seems to have been affected and shows reduced staining intensity in response to DSG3 knockdown for both siRNA molecules. This is evident especially in Figure 6.18, where relatively high DSG3 expression remains in a subpopulation of the cells in the field of view, and F-actin shows relatively high staining in the same subpopulation. This supports the findings of a previous study where overexpression of DSG3 was found to increase the rate of actin turnover while knockdown had the opposite effect (Tsang et al. 2012).

A high magnification image of the Cells-only control is shown in Figure 6.19 displaying DSG3/F-actin staining. Co-localization of DSG3 and F-actin is clearly visible in this example, with intense staining of both proteins at the cell periphery and at cell-cell junctions. Co-staining is also visible in fillopodial protrusions and membrane ruffles, suggesting that both proteins may interact at these locations in DLKP-I. This interaction is investigated in Section 6.7 where DSG3 binding partners were determined using co-immunoprecipitation.

6.7. Identification of Potential Binding Partners of Desmoglien-3 in DLKP-I

In order to further investigate the role of DSG3 in DLKP-I, co-immunoprecipitation analysis was performed to pull the protein out the cell lysate along with any binding partners which may be complexed to it. The basis behind Co-immunoprecipitation is discussed in Section 2.8. The two methods described previously in Section 4.8 were used in an attempt the Immunoprecipitation of DSG3 and its binding partners: **Cross Linked Co-IP** and **Traditional Co-IP.**

6.7.1. Cross-Linked Co-Immunoprecipitation

Cross-linked Immunoprecipitation technology was used initially as a method to pull DSG3 and its binding partners out of DLKP-I cell lysate. Within this, two different systems were used. The first used a magnetic Co-IP kit which cross-links the antibody for the target protein to Protein A/G magnetic beads permanently. Beads are separated from the cell lysate and during wash steps by magnetism in place of centrifugation. The second method uses a standard Cross-Linked Co-IP kit in which the target antibody is covalently bound to Protein A/G bead column using amide bonds. In both cases, the target antibody and binding partners are eluted from the beads by a low pH buffer which dissociates the bound complex from the beads without disrupting the activity of the target antibody. Unfortunately, neither of these cross-linked techniques was in any way successful in achieving immunoprecipitation of the target. Western Blot analysis from both techniques shows that the immunoprecipitation of DSG3 and recovery was not achieved. Results of the standard Cross-Linked Co-IP technique are shown in Figure 6.20.

Lane-1 shows DSG3 being successfully detected at 130kDa when DLKP-I cells have been lysed using the IP compatible gentle lysis buffer. Lane-2 containing the elution of immunoprecipitated DSG3 has not detected the protein, indicating that the protein was not pulled out of the lysate and bound to the Protein A/G beads via the

primary antibody. Lane-3 contains the unbound fraction in which a faint band for DSG3 can be seen, as well as faint antibody heavy and light chains. Lane-5 and 6 contain the negative control mouse IgG samples. DSG3 was not detected in the elution, but was detected in the unbound fraction as expected. Each sample run on this western Blot has an equal concentration of protein. Therefore, these results suggest that DSG3 was depleted from the cell lysate as the intensity of the band in Lane-3 is much lower than in Lane-1 or Lane-5. However, as the protein is not detected in the elution, it was either lost during wash steps, or remains bound to the beads even after the elution step. Antibody heavy and light chains in the elutions for both test and control groups further solidify the failure of this technique. Both the Magnetic and Cross-Linked kits were tried multiple times with various optimisation steps with no success. Therefore the Traditional IP method was employed.

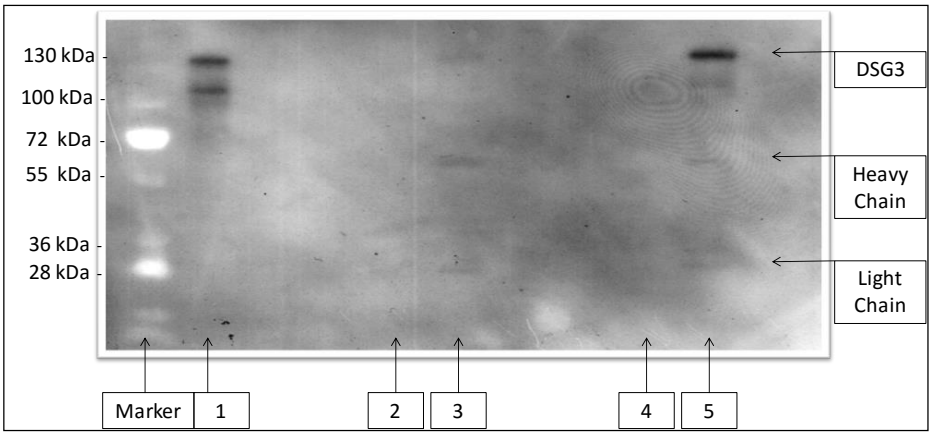


Figure 6.20: Western Blot analysis of DSG3 on cross-linked Co-IP samples. Lane 1: DLKP-I control lysate. Lane 2: DSG3 immunoprecipitate elution. Lane 3: DSG3 unbound fraction. Lane 4: Mouse IgG immunoprecipitate elution. Lane 5: Mouse IgG unbound fraction. DSG3 is visible at 130kDa in the control lysate and the mouse IgG unbound fraction.

6.7.2. Traditional Co-Immunoprecipitation: DSG3

DLKP-I cells were lysed using a gentle IP compatible lysis buffer and DSG3 was immunoprecipitated along with potential binding partners using the method outlined in Section 4.8.2. Elutions from the DSG3 test, and IgG negative control were analysed by western blot.

Western blot analysis of DSG3 was successful at detecting the protein in the test elution sample. Lane-1 contains the DLKP-I lysate control and DSG3 is visible at 130kDa as expected. Lane-2 contains the elution of immunoprecipitated DSG3 along with the heavy and light antibody chains. Lane-3 contains the elution of the control mouse IgG. An enriched band for DSG3 can be clearly seen in Lane-2, indicating successful immunoprecipitation of the target. This experiment was repeated twice and a representative blot is shown in Figure 6.21.

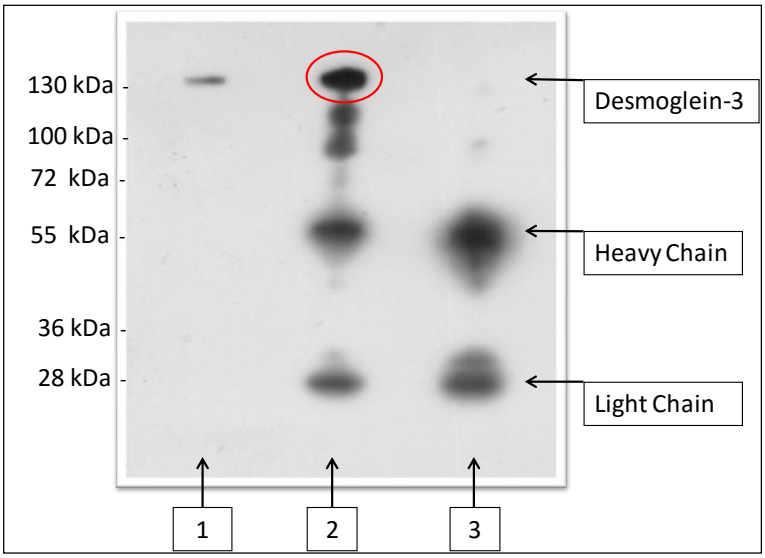


Figure 6.21: Western Blot analysis of DSG3 on Co- IP elution samples. **Lane 1:** DLKP-I lysate positive control. **Lane 2:** DSG3 Co-IP elution**. Lane 3:** Mouse IgG Negative Control elution. DSG3 is visible at 130kDa in positive control and DSG3 Co-IP elution (circled in Red). Heavy and light antibody chains are visible in both elutions.

6.7.3. Identification of Immunoprecipitated Proteins by Mass Spectrometry

Two independent Co-IP experiments were performed using the Traditional IP method with the goal of achieving co-immunoprecipitation of DSG3 and its binding partners in DLKP-I. Positive lysate control, DSG3 elution and negative control mouse IgG elution samples were separated on a polyacrylamide gel by SDS-PAGE (Section 2.3.3) and stained using Coomassie Blue (Figure 6.22). Staining revealed a faint band in the DSG3 elution (Lane-2) at 130 kDa which was not present in the mouse IgG control lane (Lane-3). Each lane was excised from the gel and sliced into 2 mm fragments from top to bottom. In-gel digestion of each gel piece was performed and proteins were identified using LC-MS/MS analysis using methods described in Section 2.8.6. MS analysis was successful in detecting DSG3 in two independent Co-IP experiments, and results are summarised in Table 6.2.

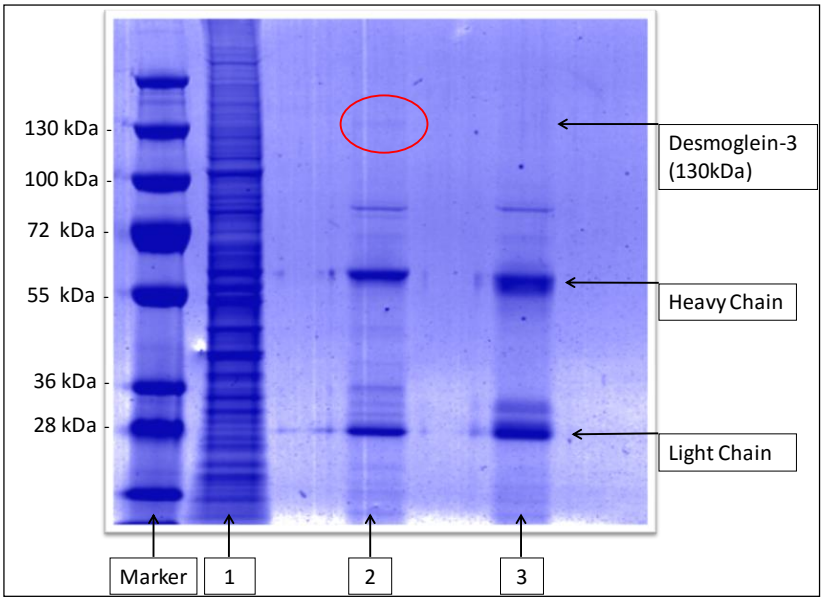


Figure 6.22: Coomassie stained gel of DSG3 Co-IP samples. **Lane 1:** DLKP-I positive control lysate. **Lane 2:** DSG3 Co-IP elution. **Lane 3:** Negative control Ms IgG to account for non-specific binding events. DSG3 band is circled in red at 130 kDa in lane-2. Heavy and light antibody chains are clearly visible.

6.7.3.1. Proteins Identified by Mass Spectrometry: Desmoglein-3 Co-IP

Immunoprecipitation elutions were analysed by LC-MS/MS after in-gel digestion as described in Section 2.7.5. Proteins identified in the negative control (mouse IgG) were eliminated from the list of proteins identified in the DSG3 immunoprecipitate. Results from both independent experiments were overlapped and proteins identified as common to both are shown in Table 6.2. The peptides numbers contributing to protein identifications are shown for each experiment (Expt 1 and Expt 2). DSG3 was found to have the most peptides contributing to its identification, 28 and 29 peptides respectively. The majority of the peptides contributing to the identification of DSG3 came from the gel slice circled in red in Figure 6.22.

Potential binding partners of DSG3 are shown in Table 6.2. A subset of these proteins were followed up by western blot analysis to see if they could be detected in the DSG3 immunoprecipitation elution, as well as to validate their expression in the DLKP cell line model and ascertain their expression patterns.

Protein	Peptide Number Expt 1	Peptide Number Expt 2
DSG3	28	29
Junction plakoglobin	4	6
Pyruvate dehydrogenase E1 component subunit alpha, mitochondrial	12	13
Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	11	10
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	15	11
Tubulin alpha-1C chain	7	5
Pyruvate dehydrogenase protein X component, mitochondrial	5	7
ADP/ATP translocase 2	3	4
ADP-ribosylation factor 4	4	5

Table 6.2: Mass spectrometry results showing Immunoprecipitated DSG3 protein, with co-eluted proteins from two replicate experiments. Number of peptides identified by MS is shown for each experiment.

6.7.4. Western Blot Analysis of Desmoglein-3 Binding Partners in DLKP

The co-IP of DSG3 in DLKP-I resulted in a panel of proteins which were repeatedly found to co-elute with the target protein. A large proportion of these proteins are associated with the mitochondria, which is unexpected for a cell adhesion protein. A subset of proteins was chosen for further validation by western blot analysis:

- **Junction Plakoglobin**
- **Pyruvate dehydrogenase E1 component subunit alpha, mitochondrial**
- **Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial**

6.7.4.1. Junction Plakoglobin:

Junction plakoglobin (PG) is a member of the armadillo family of proteins and has been identified as a component of the desmosome (Andl and Stanley 2001). PG has been found to interact directly with the cytoplasmic domain of desmogleins and desmocollins, tethering these proteins to desmoplakin (DP). This large protein links the cytoskeleton to the plasma membrane, and integrates actin, microtubules and keratin intermediate filaments (KIF) (Delva, Tucker and Kowalczyk 2009). It has been shown that PG is weakly expressed or absent in several NSCLC cell lines and that restoration of protein expression in these cell lines had anti-proliferative effects (Winn et al. 2002). In addition, over-expression of PG in SCC-9 squamous carcinoma cells induced a mesenchymal to epidermoid phenotype (Parker et al. 1998).

PG was analysed by Western Blot in the DSG3 elution and negative control sample from the co-IP experiment to see if MS analysis had correctly detected its presence as a binding partner of DSG3, and results are shown in Figure 6.23. PG can be detected in the DSG3 co-IP elution (Lane-2), indicating that it is a binding partner of DSG3 in DLKP-I. No PG is present in the negative control, ruling out non-specific binding of the protein. This DSG3/PG interaction is not a novel discovery has been long reported in reported extensively in the literature (Chitaev et al. 1996). However, the

finding is a good validation of the method employed here to discover binding partners of DSG3 by co-IP using the traditional IP protocol.

Western Blot of Junction Plakoglobin on Desmoglein-3 Co-IP Samples:

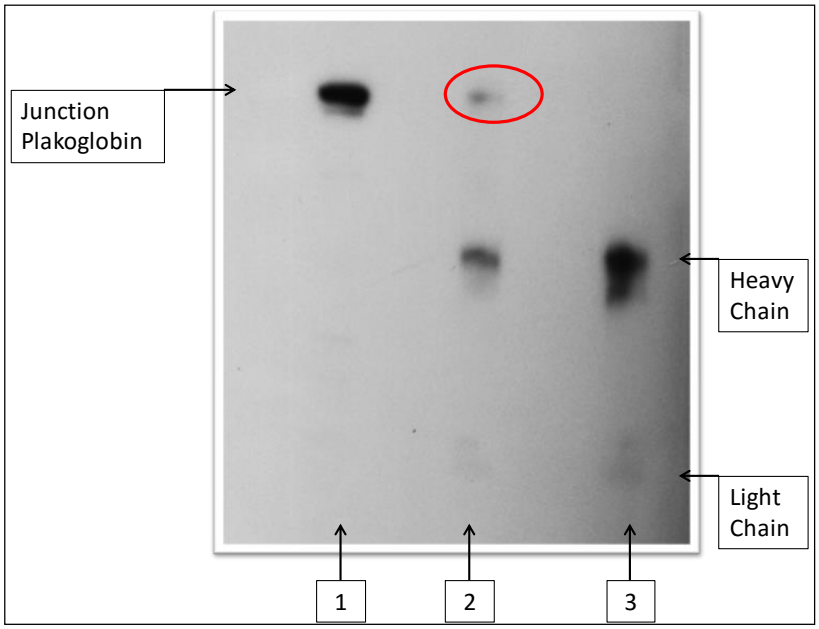


Figure 6.23: Western Blot probing for Junction Plakoglobin (PG) on Co-IP samples which targeted DSG3 and its binding partners. **Lane-1:** DLKP-I control lysate. **Lane 2:** DSG3 Elution. **Lane 3:** Mouse IgG Negative Control elution. PG is visible at ~80kda in **Lane-1** and **Lane-2** (circled in red).

6.7.4.2. Expression of Junction Plakoglobin in DLKP plus the Clones and Combination Cell Lines

Further investigation into the expression of PG was carried out by determining its expression in the DLKP cell lines as well as the combination lines by western blot, and results are shown in Figure 6.24A. The expression pattern of PG followed the same trend as was previously found for DSG3. Western blot for DSG3 is reproduced here in Figure 6.24B for comparison.

DLKP-I shows the highest expression of PG out of all the samples. DLKP-M has the next highest expression in the clones with a faint band visible at the expected MW. Neither DLKP-SQ nor DLKP show any expression of PG. Each of the combination cell lines shows expression of PG, with similar and highest abundance in DLKP-I containing cell lines: SQ+I and I+M. The Remade cell line has the next highest abundance (25% DLKP-I) and SQ+M has extremely faint expression levels. The expression pattern in co-culture therefore correlates well with the finding that PG is predominantly expressed in DLKP-I. It does not seem to be affected by any of the inhibitory effects experienced by DSG3 in co-culture, as described in Section 6.4.

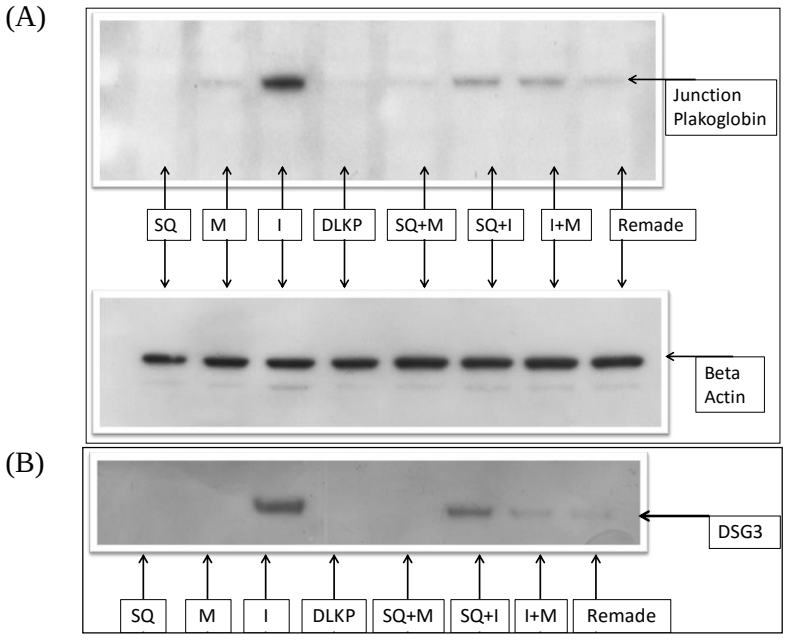


Figure 6.24: (A) Western Blot analysis of Junction Plakoglobin in the DLKP cell lines and combination lines. Bands are visible at ~80kDa and beta-actin loading control is included. **(B)** Western Blot analysis of DSG3 reproduced from Fig 5A for comparative purposes.

6.7.4.3. The Pyruvate Dehydrogenase Complex:

Co-IP of DSG3 resulted in a panel of proteins which repeated were pulled out of DLKP-I with the target protein in two independent experiments and results are shown in Table 6.2. Eight proteins were found as potential binding partners of DSG3, and five of these are mitochondrial associated proteins:

- Pyruvate dehydrogenase E1 component subunit alpha, mitochondrial **(PDHE1-A).**
- Pyruvate dehydrogenase E1 component subunit beta, mitochondrial **(PDHE1-B).**
- Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial **(DLAT).**
- Pyruvate dehydrogenase protein X component, mitochondrial **(PDHX).**
- ADP/ATP translocase 2 **(ANT2)**.

The first four proteins shown above (**PDHEA1-A; PDHE1-B; DLAT; PDHX**) are each a component of the Pyruvate Dehydrogenase Complex (PDC), which is a multi-enzyme complex located in the mitochondrial matrix.

6.7.4.4. Western Blot of PDHE1-A on Desmoglein-3 Co-IP Samples:

PDHE1-A was analysed by western blot to determine if the protein is present in the DSG3 co-IP elution. Positive and negative control samples were analysed in tandem and results are shown in Figure 6.25. PDHE1-A was successfully detected in the positive control DLKP-I lysate as well as the DSG3 elution sample. The protein was not detected in the negative mouse IgG control ruling out non-specific binding of the PDHE1-A to Protein-G. Bands for PG are visible on the blot as it had been used previously to detect that protein.

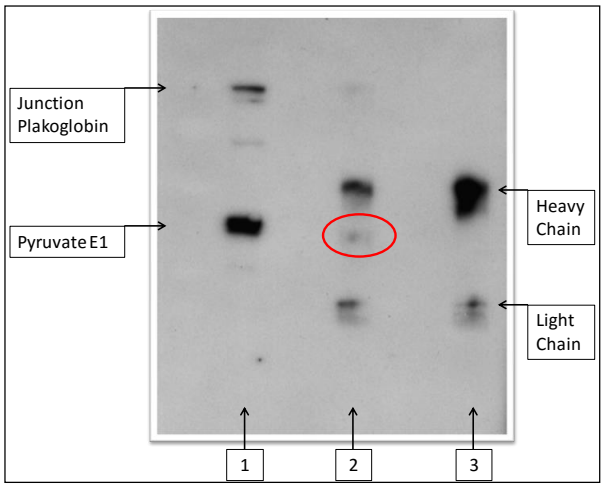


Figure 6.25: Western blot probing for pyruvate dehydrogenase E1 component subunit alpha, mitochondrial (Pyruvate E1/PDHE1-A) on Co-IP samples which targeted DSG3 and its binding partners. **Lane-1:** DLKP-I control lysate. **Lane 2:** DSG3 Elution. **Lane 3:** Mouse IgG Negative Control elution. PDHE1-A is visible at ~43kda in **Lane-1** and **Lane-2.** PG is also visible from the previous western blot analysis.

6.7.4.5. Expression of PDHE1-A in DLKP plus the Clones and Combination Cell Lines

Further investigation into the expression of PDHE1-A was performed by western blot analysis of the protein in the DLKP cell lines as well as the combination lines. This was carried out to validate the expression of the protein in the DLKP cell lines, and to determine its expression pattern. Results are shown in Figure 6.26.

PDHE1-A shows highest expression levels in both DLKP-SQ and DLKP-M at a similar abundance. DLKP-I has the lowest expression of the protein and DLKP shows intermediate expression compared to the clonal subpopulations. The combination lines showed an unexpected expression pattern with SQ+I having a higher abundance of PDHE1-A than SQ+M. Considering DLKP-SQ and DLKP-M were found to express the protein the most, it is unusual that their co-culture has resulted in decreased protein abundance. In the same vein, SQ+I shows higher than expected expression of PDHE1-A despite the DLKP-I clone showing the lowest abundance. This further implicates the co-culture and heterogeneous nature of the DLKP cell line as a critical factor to consider when analysing the behaviour of these cells.

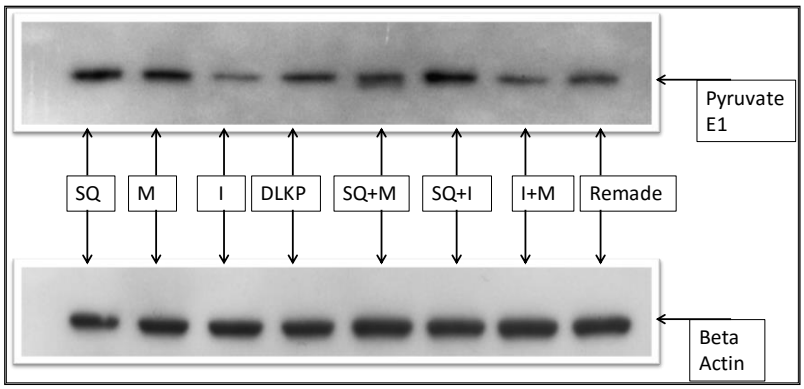


Figure 6.26: Western Blot probing for Pyruvate dehydrogenase E1 component subunit alpha, mitochondrial (Pyruvate E1/PDHE1-A) in the DLKP cell lines and combination lines. Bands are visible at ~43kDa and beta-actin loading control is included.

6.7.4.6. Expression of DLAT in DLKP plus the Clones and Combination Cell Lines

Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial (**DLAT**) is another major component of the PDC and is therefore also localized to the mitochondrial matrix. Its function is to finalise the conversion of Pyruvate to Acetyl-CoA in conjunction with the other proteins of the PDC (PDHE1-A, DLD).

Western blot analysis of DLAT in the DLKP cell lines and combinations validates the expression of the protein as present in the DLKP samples Figure 6.27. In addition, it shows the expression pattern across all the cell lines. Similar to PDHE1-A, highest expression was found in DLKP-SQ and DLKP-M. Both DLKP-I and DLKP show low expression of DLAT and at a similar level. Based on these findings, it is possible that both DLKP-SQ and DLKP-M have a higher proportion of mitochondria than DLKP-I or DLKP. Further work would be required to show this however. In the combination lines, SQ+M showed an unexpectedly low expression of DLAT considering both of its clonal constituents have the highest expression level of the protein. This suggests a possible inhibition of protein expression when the cells are in co-culture, and a similar effect was found for PDHE1-A (Figure 6.26).

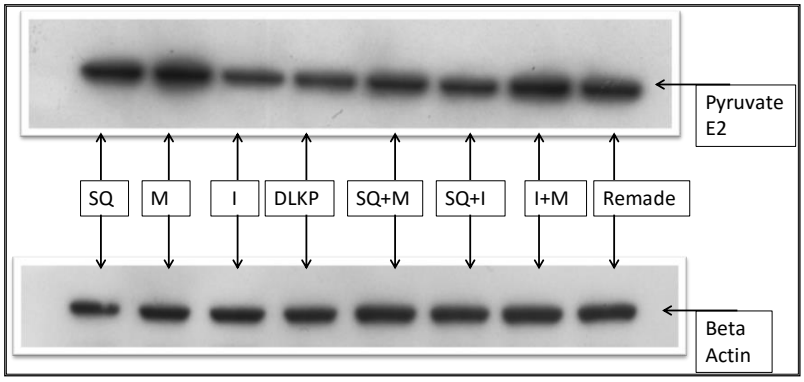


Figure 6.27: Western blot probing for DLAT (Pyruvate E2) in the DLKP cell lines and combination lines. Bands are visible at ~68kDa and beta-actin loading control is included.

6.8. Desmoglein-3 Expression in a Panel of Cell Lines

Western blot analysis was used to determine the presence and expression levels of DSG3 in a panel of cell lines. Results are shown in Figure 6.28.

- Pancreatic cell lines: PANC-1, AsPC-1, BxPC-3 and MIA PaCa 2.
- Lung cancer cell lines: A549, SK-LU-1, and NCI-H82.
- DLKP clones: DLKP-SQ, DLKP-M and DLKP-I.

The expression of DSG3 in the DLKP clones follow the expression pattern determined previously, with high abundance in DLKP-I and no detectable levels in DLKP-SQ or DLKP-M. From the panel outlined above, only two other cell lines show detectable levels of DSG3. These are AsPC-1 and BxPC-3, both pancreatic adenocarcinoma cell lines. AsPC-1 is derived from ascites and shows low level DSG3 expression. BxPC-3 is an epithelial derived cell line, and has a higher abundance of DSG3 than AsPC-1. Neither PANC1 nor MIA PaCa 2 expressed DSG3 and the protein was not detected in any of the lung cancer cell lines either.

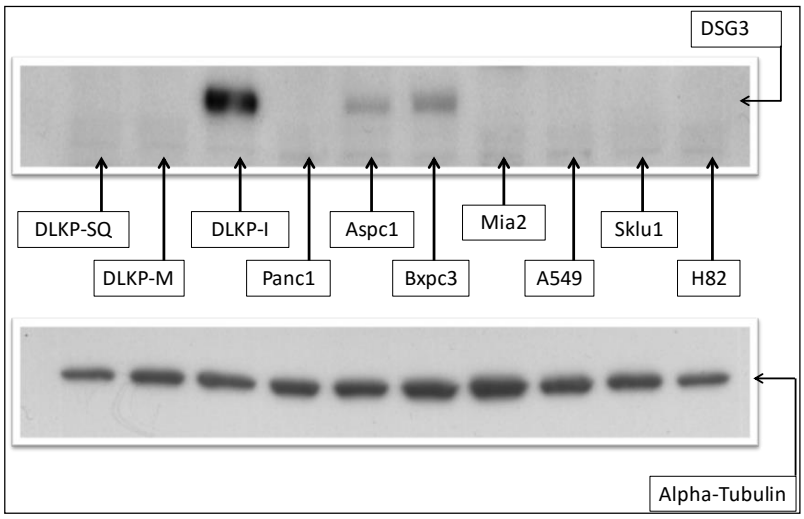


Figure 6.28: Western Blot probed for DSG3 in a panel of cell lines. DSG3 bands came up at ~130kDa in cell lines where it was present. alpha tubulin loading control included.

6.9. Summary

Desmoglein-3 was highlighted as a highly differentially expressed protein when the DLKP cell line and its subpopulations were compared against each other using quantitative label-free LC-MS/MS during exponential phase of growth. Highest expression was found in DLKP-I, with a 175.33-fold change between this cell line and the lowest expresser in DLKP-SQ. The other cell lines (DLKP-M and DLKP) also showed very low to no expression of Desmoglein-3 in this analysis, suggesting the protein may be a specific marker for DLKP-I. This expression pattern was validated at a protein and mRNA level, with DLKP-I showing the highest expression of Desmoglein-3, with no expression in the other clones. The DLKP cell line itself was found to express a low abundance of Desmoglein-3 protein and mRNA, as expected.

Immunocytochemical and immunofluorescent staining of Desmoglein-3 further validated the expression pattern already found, with only DLKP-I positively staining for the protein out of all the clones. Staining was found at cell-cell junctions as expected, however abnormal cytoplasmic staining is also visible. For both staining methods, clear staining for Desmoglein-3 is visible in a subpopulation with the DLKP, most likely representing DLKP-I. This further strengthened the potential of Desmoglein-3 as a potential specific marker for DLKP-I.

Transient knockdown of Desmoglein-3 protein expression was investigated by functional analyses, and this revealed a role for the protein in migration and invasion. Both of these phenotypic traits were significantly reduced when Desmoglein-3 protein expression was knocked down, suggesting additional roles for the protein in lung cancer other than its known role in cell-cell adhesion as previously reported (Delva, Tucker and Kowalczyk 2009). The literature also reports associations between Desmoglein-3 and the actin cytoskeleton, which may be how invasion and migration is affected by Desmoglein-3 knockdown, so this link was also investigated.

Confocal imaging of co-stained Desmoglein-3 and F-actin showed high staining for Desmoglein-3 in cell-cell junctions, and this overlapped with F-actin staining in many cases. Punctate staining is also visible for Desmoglein-3 which may be an indicator of its role in desmosomes, complexing with these structures at focal adhesions. Knockdown samples showed a visible decrease in Desmoglein-3 expression,

further validating the successful knockdown of the protein in DLKP-I. In addition, there seems to be a decrease in the staining intensity for F-actin in the Desmoglein-3 knockdown samples, though this finding requires further investigation.

Co-immunoprecipitation studies on Desmoglein-3 repeatedly found the majority of potential binding partners for the protein to be associated with the mitochondria. One of these, Junction Plakoglobin, is a known binding partner of DSG3 from the literature (Chitaev et al. 1996), and western blot analysis in our study found it shows the same expression pattern as Desmoglein-3 in these cell lines. The mitochondrial proteins found as potential binding partners are components of the Pyruvate Dehydrogenase Complex, and two of the members validated in our study as expressed in the DLKP cell line and its subpopulations by western blot analysis also (PDHE1-A and DLAT).

Chapter 7. Discussion

Tumour heterogeneity refers to the observation that multiple cell types can be found within individual tumours, and that each subpopulation can show distinct morphological and phenotypic characteristics. These can include cellular morphology, gene expression, protein expression, metabolism, and metastatic potential (Marusyk and Polyak 2010). Tumour heterogeneity in lung cancer presents major challenges for the determination of appropriate treatments and indeed the diagnosis of the disease. Subpopulations within a tumour can survive through therapeutic intervention and cause re-growth of the tumour, severely compromising attempts at treatment. Despite the evidence for intratumour heterogeneity, the subject remains relatively poorly explored. It is thought that a more systematic approach is required to characterise the extent of heterogeneity within tumours. An important factor to consider is to distinguish between cellular heterogeneity and clonal heterogeneity (Marusyk and Polyak 2010).

The DLKP cell line represents a heterogeneous lung cancer derived cell line which was found to be comprised of at least three clonal subpopulations, each with specific morphologies and phenotypic characteristics (McBride et al. 1998a, Keenan et al. 2012). This distinctive cell line model represents a unique opportunity to study tumour heterogeneity *in vitro*. In this study, we undertook to investigate the clonal subpopulations of DLKP using quantitative label-free proteomic analysis. It was hoped that proteins could be identified which are highly expressed in each of the clonal subpopulations relative to the others. These could be related to their phenotypes or potentially even act as markers for particular clones. In addition, these clones were co-cultured and included in the proteomic analysis to look for effects on highly abundant proteins representing the clones. This may provide insight into how the clones behave in their heterogeneous environment, inhibiting or enhancing protein expression stimulated by neighbouring cells.

A simplistic view of tumours as homogenous entities is not a true representation of the reality in the majority of lung cancers or cancers in general. Intratumour heterogeneity is a significant hurdle in the development of effective treatments, as undetected subclonal cancer cells can develop drug resistance rendering the treatment ineffective (Gerlinger and Swanton 2010). It can also lead to the false diagnosis of cancer subtypes based on single biopsy methods which underestimate the genetic and proteomic heterogeneity present in individual tumours. A number of approaches can be taken to investigate the tumour heterogeneity. Our study aimed to

increase the knowledge of the roles of specific proteins in a heterogeneous lung cancer cell line, and investigate the roles of these proteins in its clonal subpopulations. An alternative or even supplementary strategy could use next generation sequencing techniques to investigate genetic aberrations which give rise to lung cancer. Developments in next generation sequencing techniques have been employed to identify extensive genetic heterogeneity in individual tumours (Cancer Genome Atlas Research Network et al. 2013). These studies have shown the wide variety of aberrant gene fusions, dysfunctional transcriptomes with single-nucleotide variants, and alternative splice variants which contribute to tumour heterogeneity.

One such study investigated intratumour heterogeneity in clear-cell renal carcinomas (ccRCC) where multiple biopsies were taken from each of 10 ccRCC tumours and analysed using multi-region exome sequencing (Gerlinger et al. 2014). The analysis revealed a heterogeneous genomic profile within each of the 10 ccRCCs with spatially separated subclones present in each. The spatial separation can give the illusion of a dominant clone when the tumour is assessed by a single biopsy resulting in ineffective targeted therapies. They also screened for the known genetic drivers of ccRCC and found that 73-75% of genetic driver events were subclonal. The presence of these subclonal driver events may explain the inevitable acquisition of resistance to targeted therapeutic treatment.

Studies such as the one carried out by Gerlinger et al highlight the importance of understanding cancer at the genetic and transcriptomic levels, however it is equally important to include the proteomic level in order to achieve a greater understanding of the mechanisms involved in cancer. Proteins are central to the normal function of cells and aberrant protein expression can initiate tumour development, drive the development of heterogeneous subclones, and ultimately dictate treatment options. High throughput options for analysing the proteome of cancers are critical to allow the identification of changes in signalling pathways and protein expression, which themselves may become therapeutic targets. A mass spectrometry based study using a method such as the one presented in this thesis represents such an option, which would be greatly enhanced by using fresh frozen tumour tissue as the starting material. Such a proteomic study could be complemented by a genetic study which would allow the correlation between mRNA and protein expression to be investigated on a large scale. There is only a partial

correlation between mRNA levels and protein expression, therefore integrating both disciplines into a proteogenomic approach may shed light on how the information flow from genome to proteome is altered in tumours.

One of the main aims of this thesis was to use quantitative proteomic profiling to identify proteins highly expressed each of the DLKP clonal subpopulations. The identification of such proteins may lead a greater understanding of the heterogeneous nature of lung cancer. These proteins can be exploited to investigate the phenotypic differences of the clones as found in previous work (Joyce H, PhD 2015), and may serve to characterise their known phenotypic properties.

In order to highlight proteins which may be potential markers for each clonal subpopulation, samples were first analysed in a 'global analysis' experimental design method using Progenesis software. The global analyses resulted in the identification of 127 proteins during exponential phase, and 203 proteins during stationary phase of growth (See Section 3.1.3.2). These proteins showed high fold changes and strong, statistically significant changes in expression levels between at least two samples. Within the two growth-phase datasets, 63 proteins overlapped, indicating conservation of expression for a subgroup of proteins across multiple growth phases.

Cell lines were then compared against each other in two-sample experimental designs to elucidate the fold changes for a protein between individual cell lines in a comparison, along with their associated statistical significance. This method was referred to as the two-sample analysis. The data generated from this analysis was used to re-enforce the findings of the global analysis, and to provide the intermediary fold changes and associated statistical significances. A protein found to have an interesting expression pattern in the global analysis, i.e. showing highest abundance in one of the clones relative to all the other cell lines, could be re-enforced using data available from the two-sample analysis. This strengthens the protein discovery and provides extra fold change information.

7.1. Proteins of Interest

Protein expression patterns across the whole sample set were screened to look for proteins showing interesting expression patterns. An interesting expression pattern was assigned to proteins where high expression was visible in one clone only or at least significantly highly expressed in one clone. This may indicate a potential clone marker, or a protein associated with the known phenotype of that cell line. A number of proteins showed these patterns are shown in Section 3.1.4. We followed up Shootin-1, MARCKS, and Desmoglein-3 due to their relative novelty in lung cancer studies, and the role of Desmoglein-3 in particular as a potential marker for DLKP-I. **Table 7.1** shows a summary of the findings of this study for each protein investigated.

Protein Target	Highly Expressed In	Cellular Localisation	Functional Effect after RNAi	Potential Binding Partners
Shootin-1	DLKP-SQ	Cytoplasmic staining with enrichment in projecting structures.	Slight reduction in migration	Semenogelin-1
				Src-Substrate Cortactin
				Striatin-3
MARCKS	DLKP-M	Punctate staining throughout the cytoplasm with enrichment in the tips of projecting structures.	Reduction in migration	No data
			Reduction in invasion	
Desmoglein-3	DLKP-I	Cytoplasmic and nuclear staining. Intense staining at cell-cell junctions.	Reduction in migration	Junction plakoglobin
			Reduction in invasion	Pyruvate dehydrogenase E1
				DLAT

Table 7.1: Summary table of the three main proteins investigated in this study. Main findings are displayed.

7.2. Shootin-1

Shootin-1 was highlighted as a highly expressed protein in DLKP-SQ as discussed in Section 4.2. Shootin-1 is a recently discovered protein which has been exclusively studied in neuronal cells to date. It was originally discovered in a study which identified it as a protein involved in the polarization of rat hippocampal neurons (Toriyama et al. 2006), and is described as a brain-specific protein. This group was investigating the formation of axons *in vitro* using rat hippocampal neuronal cells, and found Shootin-1 to be involved in the generation of internal asymmetric signals required for neuronal polarization and neurite outgrowth, strongly localising to the tips of outgrowing axons. They identified the protein as a potential mediator of axon outgrowth in these cells, and found that overexpression of Shootin-1 leads to the generation of surplus axons, while siRNA knockdown reduces axon formation.

Follow up work found that Shootin-1 interacts with actin retrograde flow (ARF), a process integral to providing the force required for axon outgrowth in neuronal cells (Shimada et al. 2008, Van Goor et al. 2012). They concluded that Shootin-1 acts as a mediator or 'clutch molecule', linking filamentous actin and the cell adhesion molecule L1-CAM. Impairing this link impairs axon outgrowth, while enhancing the link promotes outgrowth. Another recent study showed that p21-activated kinase 1 (Pak1) mediates the phosphorylation of Shootin-1, promoting the clutch engagement and transmission of traction forces at axonal growth cones through ARF (Toriyama et al. 2013). They also found the chemo-attractant Netrin1 positively regulated the Pak1 mediated phosphorylation of Shootin-1, resulting in increased axon outgrowth. In addition, Shootin1 was also shown to co-localise with active pools of phosphoinositide 3-kinase (PI3K) and inhibition of PI3K activity reduces the ability of shootin1 to induce surplus axons. A separate study found Shootin-1 to be a novel cargo of the motor protein Kif20B, and the interaction with this member of the kinesin superfamily localises Shootin-1 to the tip of elongating axons by transport along microtubules (Sapir et al. 2013). Taken together, these studies suggest Shootin-1 is a protein involved in cell motility, however this has not yet been investigated in cancer cell lines.

The Shootin-1 protein does not show significant homology to previously known proteins, which suggests that it is a member of a novel class of proteins. Invertebrate homologues of Shootin-1 were not found in the databases, therefore is

thought that Shootin-1 is probably a late addition to the genome during the evolution of animals (Toriyama et al. 2006). The work carried out by Toriyama et al. investigated the expression of Shootin-1 in a range of rat tissues: brain; heart; lung; liver; spleen; kidney; muscle; and placenta, and they determined that Shootin-1 is a brain specific protein not expressed in the other tissues tested. Shootin-1 has not been studied at length in any type of cancer as of yet and has only been studied in neuronal cells. One group investigating single nucleotide polymorphisms (SNPs) associated Shootin-1 with breast cancer risk as part of a large dataset (Higginbotham et al. 2012), therefore Shootin-1 remains un-researched in cancer or explored in any other cells than neuronal types.

In our study we wanted to determine if Shootin-1 is highly expressed in DLKP-SQ compared to the other clones. Therefore validation studies were performed on all DLKP cell lines using western blot analysis. Label-free proteomic analysis had found Shootin-1 to be expressed at ~4-fold higher abundance in DLKP-SQ compared to DLKP-M, DLKP-I and DLKP, each of which expressed the protein at a similarly low level. The findings from the western blot analysis of Shootin-1 across the sample set correlates well with the label-free proteomics data. Expression of Shootin-1 is significantly higher in DLKP-SQ relative to DLKP-M, DLKP-I and DLKP. A minor discrepancy was found in that DLKP-M shows higher expression of Shootin-1 than DLKP-I, the opposite of the quantitative proteomic analysis findings. However, both cell lines express Shootin-1 at a similarly low level and the expression pattern as a whole remains unchanged.

The expression pattern for Shootin-1 in the combination lines highlighted unexpected protein abundances from label-free proteomic analysis. The combination of SQ+M shows a much lower abundance of Shootin-1 than SQ+I. Given that both DLKP-M and DLKP-I express Shootin-1 at similar levels, this expression pattern is surprising. Western blot analysis of the combination lines agrees with the finding that Shootin-1 expression is higher in SQ+I compared to SQ+M. This result suggests there may be an inhibitory effect on the expression of Shootin-1 in DLKP-SQ by co culture with DLKP-M. The interaction between mesenchymal and squamous cells has been previously found to directly affect protein expression as well as the phenotypes of cells in heterogeneous cell populations. Cancer associated fibroblasts have been shown to exert significant paracrine effects on squamous tumour cells by excreting IL-6 which in turn

affected protein expression and promoted tumour growth in NSCLS (Vicent et al. 2012). The heterogeneous nature of DLKP may function in a similar way, with cells interacting with each other through paracrine effects, regulating protein expression and altering phenotypes to adapt to the tumour microenvironment.

The comparison of DLKP and DLKP-Remade shows a significant difference in Shootin-1 abundance by both proteomic analysis methods described above. Abundance levels for Shootin-1 are higher in the DLKP-Remade with low levels of the protein detectable in DLKP. This suggests that DLKP-Remade is not an exact reconstitution of the original DLKP cell line as described previously (McBride et al. 1998a). It is possible DLKP-Remade requires a longer time in culture than 5 passages to fully mimic the protein expression patterns of DLKP, or there are as of yet undescribed cell types present in DLKP affecting Shootin-1 expression. Further work is needed to establish this.

Following on from the western blot analysis, quantitative PCR was performed to determine the relative levels of Shootin-1 gene expression across the sample set. The abundance of Shootin-1 mRNA in the sample set may highlight post transcriptional regulation factors in certain cell lines, and shed light on how co-culture influences the expression of the protein. The results from the gene expression assays show a direct relationship between protein expression abundance as determined by the label-free and western blot proteomic analyses. On average, DLKP-SQ shows a ~4 fold increase in Shootin-1 protein abundance when compared to DLKP, DLKP-M and DLKP-I. The mRNA assay results show a similar abundance change with DLKP-SQ showing ~5-fold higher Shootin-1 mRNA abundance compared to DLKP-M and DLKP-I. This further solidifies Shootin-1 as a marker for DLKP-SQ at both the protein and mRNA level.

The quantitative PCR analysis on combination cell lines shows Shootin-1 mRNA is significantly decreased when the clonal subpopulations are co-cultured. This is especially evident in the combination lines which contain DLKP-SQ as a constituent. Using the mRNA abundances in the clones as a guide, it was expected that co-culture of DLKP-SQ with either DLKP-M or DLKP-I would reduce Shootin-1 mRNA expression by ~60%, due to the diluting effect of co-culture. However qPCR analysis reveals the mRNA abundance levels for Shootin-1 to reduce sharply by ~80% relative to DLKP-SQ levels. This suggests that inhibition of Shootin-1 expression is occurring at the

transcriptional level when the clones are in co-culture conditions. In addition, the mRNA abundances in DLKP and DLKP-Remade show similar expression at the mRNA level, whereas they differ at the protein level. Taken together, it is possible that an undiscovered cell line in the DLKP cell population inhibits the expression of Shootin-1 at the post-transcriptional level which is not present in DLKP-Remade. This may support the theory that DLKP parental cell line contains cell populations which have yet to been discovered, again outlining the heterogeneity of this cell line.

Shootin-1 is a highly expressed protein in DLKP-SQ relative to the other clones as shown by this work so far, at the protein and mRNA level. The evidence shows Shootin-1 is expressed at a significantly higher abundance in DLKP-SQ relative to DLKP-M, DLKP-I, or DLKP. In order to further validate the expression pattern of Shootin-1, and elucidate the role it may play in the DLKP cell line model, staining of Shootin-1 was performed by immunocytochemistry and immunofluorescence in the four cell lines (DLKP and its clonal subpopulations). Each of the DLKP cell lines were expected to stain positively for Shootin-1 as the protein is detectable by the proteomic analyses already discussed. Visualisation of Shootin-1 in the four cell lines shows the protein staining pattern correlates well with the expected staining pattern. DLKP-SQ has the most intense staining for Shootin-1 out of all the cell lines in the analysis. The next most intense staining is in DLKP-M, followed by DLKP and finally DLKP-I.

In the case of DLKP-SQ, Shootin-1 localization is evident in the cytoplasm with little or no staining in the nucleus or nuclear envelope. The immunocytochemical analysis shows Shootin-1 staining positively in small filopodial outgrowths in DLKP-SQ. These outgrowths are not usually visible in day to day cell culture of DLKP-SQ due to its tightly packed block-like morphology in monolayer (See Figure 1.2), however immunocytochemical staining of Shootin-1 allows them to be visualized. In addition, immunofluorescence staining shows high levels of Shootin-1 in the cytoplasm, and high intensity areas of staining can be seen at the edges of the cytoplasm in DLKP-SQ cells.

DLKP-M stains positively for Shootin-1, but at a lower intensity than DLKP-SQ. The DLKP-M clone has a morphology which can be described as mesenchymal-like, with an elongated cytoplasm and long neurite-like outgrowths.

Immunocytochemical staining of Shootin-1 is high in these outgrowths and in some cases staining can only be seen in the protrusions emanating from the cell. Immunofluorescent staining demonstrates this staining pattern very clearly, cell protrusions staining strongly for Shootin-1. The immunocytochemical and immunofluorescent staining in DLKP-I shows the lowest intensity of Shootin-1 relative to the other cell types. Diffuse cytoplasmic localization of Shootin-1 can be seen in both techniques with no remarkable staining pattern visible in DLKP-I.

DLKP shows a very diverse staining pattern for Shootin-1. This finding is expected due to the heterogeneous nature of the cell line. The most intense staining for Shootin-1 is visible in block-like cells with thin filopodial protrusions. These cells are assumed to be DLKP-SQ as they have the highest Shootin-1 staining intensity as well as the distinctive DLKP-SQ morphology. What are assumed to be the remaining two clones are visible in the DLKP cell line; however it is difficult to distinguish these from each other by using Shootin-1 staining intensity as a guide. The DLKP cell line also contains abnormally large cells which stain positively for Shootin-1. These cells do not match the characteristics of any of the known clonal subpopulations of DLKP.

This imaging of Shootin-1 in the DLKP cell lines revealed a common characteristic in staining patterns which correlates with previous finding from the literature. The high staining of Shootin-1 in the cell outgrowths was previously described in studies originally performed on the protein in neuronal cells. Shootin-1 was first described in a study which investigated the formation of axons in rat hippocampal cells (Toriyama et al. 2006). In this study, rat hippocampal neurons were cultured and then dissected to separate the axon networks from the cell bodies. A proteomic comparison of axons vs. the cell bodies found Shootin-1 to be significantly highly expressed in the axons, and this study resulted the first charaterisation of the protein. This description of Shootin-1 localization in neuronal cells found by Toriyama et al. closely matches what was found in our study for DLKP cell lines. This discovery of Shootin-1 in the DLKP cell line model with its previously described role in neuronal cells suggests it may act as a neuroendcrine marker.

Shootin-1 is highly expressed in DLKP-SQ, at both the protein and mRNA level relative to all the other DLKP cell lines. Co-culture has a significant effect on the expression of Shootin-1, and it is hypothesised that DLKP-M plays an inhibitory role in

this regard. Immunocytochemical and Immunofluorescent imaging techniques confirm DLKP-SQ as having the highest Shootin-1 abundance out of all the DLKP clonal subpopulations, further validating the expression pattern established by other techniques. In addition, evidence of other cell types in DLKP has been found. This may support the hypothesis that the DLKP cell line contains as of yet uncharacterised subpopulations. Literature studies show that Shootin-1 plays a role in the outgrowth of axons in neuronal cells. Staining of the protein in DLKP revealed a similar staining pattern, with highest localization of the protein found in cell outgrowths.

To investigate the role Shootin-1 plays in DLKP-SQ *in vitro*, the protein was knocked down using RNAi and follow up functional assays were performed. Shootin-1 knockdown had been verified by western blot analysis before functional assays were pursued. These assays measure proliferation, migration and invasion of the cells to assess if knockdown of Shootin-1 has any effect on these phenotypic characteristics. DLKP-SQ was chosen as the subject for the follow up functional analyses as the abundance of the Shootin-1 is significantly higher in this cell line relative to the other clonal subpopulations at the protein and mRNA level.

Transient siRNA knockdown of Shootin-1 by RNAi has no effect on proliferation of DLKP-SQ. This finding is unsurprising as there is no evidence in the literature that Shootin-1 plays any role in the proliferation or viability of the neuronal cells in which it has been studied. The migratory ability of DLKP-SQ is significantly decreased upon knockdown of Shootin-1 in DLKP-SQ. This significant effect is present in only one of the siRNA molecules used in the experiments when compared to the negative control. The second siRNA molecule which targets Shootin-1 does not have a significant effect on migration, but it follows the same trend tending towards a reduction in migration also. Knockdown of Shootin-1 in DLKP-SQ has no effect on invasion for either siRNA molecule tested, indicating that knockdown of Shootin-1 does not reduce or increase the invasive ability of this cell line.

The original studies on Shootin-1 found the protein to interact with F-actin and as a result actively promote axon outgrowth (Toriyama et al. 2006, Shimada et al. 2008). Overexpression of the protein was found to enhance axon outgrowth, while

interference by RNAi was found to have the opposite effect and inhibit the formation of outgrowths. For our study, it was hypothesised based on the literature that a reduction in Shootin-1 expression may have a similar effect on actin related phenotypes such as cell spreading, motility or cell shape in DLKP-SQ. The findings in our study suggest that this potentially is the case for cell migration.

The reduction in migration of DLKP-SQ by Shootin-1 knockdown may be explained by its known associations with actin retrograde flow (ARF) investigated in neuronal cells by Shimada et al. 2008. They found that Shootin-1 is involved in ARF which is an instrumental process in wound healing, cell migration and tumour metastasis (Zimmermann et al. 2012). ARF describes the process by which actin filaments polymerize in the lamellipodium, which is a flat projection structure at the leading edge of a migrating cell. The retrograde polymerization of actin (back towards the cell body) in the leading edge of this structure pushes against the plasma membrane, conferring movement to the cell. However, in order for the cell to gain traction by using ARF, the meshwork of actin must be mechanically coupled to cell adhesion molecules (CAMs) attached to extracellular substrates thereby providing the mechanical tension required for movement (Suter and Forscher 2000, Yamashiro and Watanabe 2014). The study by Shimada et al. found evidence that Shootin-1 mediates the link between actin and the cell adhesion molecule L1-CAM, acting as a 'clutch molecule', regulating the effects of ARF on neuronal axon outgrowth. They concluded that Shootin-1 plays a significant role in cell motility through ARF. It is therefore possible that the role found by these studies Shootin-1 as a mediator between the Actin cytoskeleton and CAMs may also be the case in DLKP-SQ in our study, though further work would be required to establish this link.

Our study represents the first investigation of Shootin-1 at a functional level in cells other than neuronal cells. In addition, this is the first study to investigate the role of Shootin-1 in a lung cancer cell line, and to have found a functional role for the protein in relation to cell migration. No studies link Shootin-1 with a functional cancer phenotype, however its binding partner L1-CAM, as described in Shimada et al. 2008, has been found to be expressed in a variety of cancer types. L1-CAM is a transmembrane protein predominantly expressed in developing neurons where it plays important roles in the developing nervous system mediating neuron adhesion, migration

and axon guidance (Kleene et al. 2001, Hortsch 2000). In addition to its physiological role in neural development, recent studies indicate that L1-CAM also plays a role in the pathology of a diverse variety of cancers types. Its expression has been detected in colon, breast, ovarian and pancreatic cancers (Gavert et al. 2005, Fogel et al. 2003, Ben et al. 2010) and high expression has been found to correlate with poor prognosis. Its localization in human tumours is exclusively found at the invasive front, suggesting that L1-CAM may have an active role in the migratory and invasive processes.

The functional assays carried out in this study found while Shootin-1 has no effect on proliferation or invasion, a significant effect was found on cell migration. Knockdown of Shootin-1 in DLKP-SQ reduces the migratory ability of the cells. Our study shows for the first time that Shootin-1 knockdown significantly reduces cell migration in a lung cancer cell line.

With both siRNA molecules inducing a reduction in the migratory ability of DLKP-SQ upon knockdown, it was concluded that Shootin-1 potentially plays a role in the motility of DLKP-SQ. Literature studies implicate a strong link between F-Actin and Shootin-1 in neuronal cells. It has been found to mediate the linkage between actin retrograde flow and CAMs, working as a 'clutch molecule' to increase mechanical tension depending on its phosphorylation state by Pak1 (Shimada et al. 2008, Toriyama et al. 2013). To investigate the mechanism by which Shootin-1 is involved in the migratory ability of DLKP-SQ, the studies mentioned above were taken into account, and the expression and localisation of F-actin were investigated relative to Shootin-1. Using RNAi, Shootin-1 was knocked down and co-stained with F-actin as described in Section 4.7.3.

Co-staining of Shootin-1 and F-actin revealed an overlapping staining pattern in the control group samples (Cell-only; Lipo-only; Negative siRNA control). This overlap is most clearly visible in the thin filopodial protrusions which can be seen extending from the cytoplasmic membrane of DLKP-SQ. In addition, the leading edge of lamellipodial-like structures also shows enrichment of Shootin-1, and similar findings were reported by Toriyama et al. in their investigations into Shootin-1 in neuronal cells. These structures were visible in our immunofluorescence work (Section 4.6.2) and it was thought these may be associated with actin also due to the strong links reported in the literature. Combined with the knowledge that Shootin-1 knockdown

reduces the migratory ability of DLKP-SQ, it can be hypothesised that a similar mechanism involving Shootin-1/F-actin coupling may be involved in the migratory ability of this lung cancer cell line.

The confocal imaging of DLKP-SQ was also used to further validate the transient knockdown of Shootin-1by RNAi. There was a visible decrease in the expression of Shootin-1 in the majority of the cells in each field of view, with the largest reduction of the protein visible in the cytoplasm of the cells. Some minor perinuclear as well as focused cytoplasmic staining remains after RNAi knockdown. Our results also show that F-actin staining is unaffected by the knockdown of Shootin-1, with the protein showing a similar staining intensity and localisation when compared to control group samples. This implies that Shootin-1 is not critical to the expression and localisation of F-actin in DLKP-SQ, but it may still act as a mediator molecule similar to its role in neuronal cells. The work carried out by Shimada et al. shows that Shootin-1 acts as a mediator between F-actin and CAMs, but does not enhance or inhibit F-actin levels. Therefore, we hypothesise that Shootin-1 does not affect the expression, localisation, or rate of actin polymerisation in DLKP-SQ. However it may increase or decrease the coupling of F-actin filaments to CAMs and in so doing, regulate cell migratory ability of DLKP-SQ.

Shootin-1 was successfully co-immunoprecipitated from DLKP-SQ whole cell lysate along with a panel of 5 other proteins. This protein panel was reproducibly co-immunoprecipitated with Shootin-1 from the lysate of DLKP-SQ cells and identified using LC-MS/MS. These are Semenogelin-1, Src-Substrate Cortactin, Striatin-3, Small Nuclear Ribonucleoprotein Sm D3, and 60s Ribosomal Protein L12. For the purposes of this discussion, the focus will be on the first three proteins in the list, as no follow up work was carried out on the last two proteins.

Semenogelin-1 was identified in our study as a potential binding partner of Shootin-1 in DLKP-SQ. The semenogelin proteins (SEMG1 and SEMG2) are mainly expressed and secreted by the seminal vesicles, and are the main structural component proteins which make up human semen. They both originate from the glandular epithelium of the seminal vesicles and are secreted in high concentrations. SEMG1

plays a specific role in the gel-like formation of semen, and is readily cleaved by prostate-specific antigen (PSA) into small peptides which are thought to have functions of their own such as self-assembling to form a hydrogel (Frohm et al. 2015), and binding to HIV-virions resulting in enhanced HIV infection (Roan et al. 2014).

We found SEMG1 to be expressed in a human lung cancer cell line using co-IP as a potential binding partner of Shootin-1. Neither of these two proteins have been reported to be expressed in healthy lung tissue. SEMG1 is normally a very highly specifically expressed protein, and is only found in the seminal vesicles and urinary bladder, with little or no expression in other tissues of the body. Interestingly, low levels of SEMG1 have been found to be expressed in neuronal cells of the brain, specifically the Purkinje cells of the cerebellum (Human protein atlas). These neuronal cells are characterised by their elaborate and high number of dendritic outgrowths, and suggests that SEMG1 may have a neuronal role, in addition to its established role from which it acquired its name. This may strengthen the role of SEMG1 as a potential binding partner of Shootin-1 in our study, as the latter protein is highly associated with dendritic and axonal outgrowths in neuronal cells (Sapir et al. 2013, Toriyama et al. 2013).

We found SEMG1 to be expressed by the DLKP cell line and its subpopulations using western blot analysis (see Figure 4.21). DLKP-M showed the highest levels of SEMG1 expression, while DLKP-SQ had the lowest. DLKP-M cells are characterised by their prominent outgrowths and cell protrusions, most closely resembling neuronal cell morphology, and they also express the second highest levels of Shootin-1 out of all the DLKP clones as determined by the quantitative label-free and western blot analyses. It is therefore possible that SEMG1 and Shootin-1 have neuronal roles in the DLKP cell lines. Further work is required to validate SEMG1 as a binding partner of Shootin-1 using reverse co-IP experiments. The other DLKP clones, particularly DLKP-M, will be investigated to assess the same binding partners of Shootin-1 are found across all cell lines and to discover new binding partners.

SEMG1 has been described in one other study to be expressed in SCLC and was found to be associated with adhesion complexes in lung cancer cell lines (Rodrigues et al. 2001). In the context of the potential interaction of SEMG1 and Shootin-1 found by our study, the association of SEMG1 with adhesion complexes ties in with previous literature reports which linked Shootin-1 to CAMs in neuronal cells

(Shimada et al. 2008). This strengthens our findings that SEMG1 may be a binding partner of Shootin-1 in the DLKP-SQ cell line. The study by Rodrigues et al. also found SEMG1 to be a secreted protein, and concluded that it may be potentially used as a biomarker for the presence of SCLC. Further work will need to be carried out to determine if SEMG1 is secreted by DLKP-SQ, or any other the other DLKP cell lines.

Src-Substrate Cortactin (CCTN) was another protein identified in our study as a potential binding partner of Shootin-1 in DLKP-SQ. It is a multi-domain, ubiquitously expressed protein that was first identified as a major substrate of the Src oncogene (Wu et al. 1991). It is localized to the cytoplasm and is highly associated with F-actin. It has been shown to play a role in many actin based cellular processes such as migration and invasion (Yamaguchi and Condeelis 2007) as well as axon guidance (Knoll and Drescher 2004). CCTN is activated by phosphorylation events, and once this occours it binds directly to the Arp2/3 complex and F-actin through its Proline-rich domain. This activates the Arp2/3 complex to promote nucleation of actin filaments and induces rearrangements of the actin cytoskeleton such as the formation of filopodia and lamellipodia (Uruno et al. 2001). It is highly associated with the formation and maintenance of invadopodial structures by stabilising the ARP 2/3 complex on filamentous actin, and has also been shown to be essential for the trafficking of the key invadopodia proteases, MT1-MMP, MMP2, and MMP9 in head and neck squamous carcinoma cells (Clark and Weaver 2008, Clark et al. 2007).

Our study found CTTN to be a potential binding partner of Shootin-1 in DLKP-SQ cells by co-IP in repeatable experiments. In addition, CTTN was also identified as a differentially expressed protein in our quantitative label-free study on the DLKP cell line and its clonal subpopulations, showing highest expression in DLKP-SQ. Follow up by western blot analysis for CTTN validated the expression pattern found in the label-free study as shown in Figure 4.22. CTTN has been shown to be overexpressed in a variety of cancers such as colorectal, laryngeal and esophageal squamous cell carcinoma (Cai et al. 2010, Rodrigo et al. 2011, Luo et al. 2006). Immunohistochemical analysis of CTTN expression in colorectal carcinoma found that staining scores were significantly higher in well, moderately, and poorly differentiated colorectal adenocarcinomas compared to normal colorectal epithelia (Lee et al. 2009). In a subset of *in vitro* studies, CTTN over-expression induced tumour invasion and metastasis in esophageal and head/neck squamous cell carcinoma (Luo et al. 2006).

Reports on investigations of the role of CTTN in lung cancer are rare, however in NSCLC it was found to be associated with poor prognosis and lymph node metastasis by immunohistochemical analysis. It was concluded that CTTN is involved in the progression of NSCLC (Noh et al. 2013).

CTTN clearly plays a role in migration and invasion in different systems. We hypothesise that this protein may serve a similar function in the DLKP cell lines. In our study, transient knockdown of Shootin-1 expression resulted in a significant reduction in the migratory ability DLKP-SQ. As a potential binding partner of Shootin-1, it is possible CTTN was affected, and was unable to fulfil its role as an F-actin/Arp 2/3 binding protein, resulting in reduced migration. Shootin-1 has been shown to be a mediator of F-actin-CAM interactions, so our discovery of CTTN as a potential binding partner of Shootin-1 represents a novel finding in lung cancer. The potential of Shootin-1 and CTTN acting as binding partners is strengthened by a very recent study which found a direct interaction between them in rat hippocampal neurons (Kubo et al. 2015).

Striatin-3 (STRN3) was investigated as a prospective binding partner of Shootin-1 after co-IP experiments in our study repeatedly found it to be potentially complexed with Shootin-1 in DLKP-SQ. STRN3 is described as a quantitatively minor, Ca^{2+}/Calmodulin binding protein, and is mostly expressed in neurons of the motor and olfactory structures of the central nervous system (Bartoli, Monneron and Ladant 1998). It is exclusively expressed in the somatodendritic compartment of neurons, at the exclusion of axons, and is localised mostly to dendritic spines. Striatin family members contain several protein–protein association domains: from the N- to the C-terminus, a putative Caveolin-binding motif, a coiled-coil structure, a Ca^{2+}/Calmodulin-binding site and a large WD-repeat domain (Castets et al. 1996, Bartoli et al. 1999). Due to its potential for bind many different protein types, Striatin is thought to be a scaffolding protein which functions in a Ca^{2+}dependent manner. Down-regulation of Striatin in cultured rat spinal motor-neurons was found to impair the growth of dendrites, indicating the protein has a potential role in the formation of cell protrusions (Bartoli et al. 1999).

We analysed the expression of STRN3 in the DLKP cell line and its clonal subpopulations using western blot analysis. All cell lines showed detectable levels of STRN3 expression, with highest expression in DLKP-M. Our study is not the first to detect the expression of Striatin family members in cancer cell lines, as Striatin-4 has been found to be expressed in pancreatic; esophageal; colorectal; lung; and ovarian carcinomas (Wong et al. 2014). However our study is the first to specifically detect STRN3 in a lung cancer cell line, and demonstrate heterogeneous expression between subpopulations of that cell line. Similar to SEMG1, STRN3 is thought to have a specific neuronal role, and both proteins show highest expression in DLKP-M. As mentioned before, DLKP-M has the most neuron-like morphology, with protruding dendritic-like outgrowths. It is therefore possible STRN3 in combination with Shootin-1 may play a role in the formation of these structures, which may have a concomitant effect on the migratory ability of the DLKP cell lines, particularly DLKP-SQ and DLKP-M.

Shootin-1 was previously found to be highly expressed in DLKP-SQ in a preliminary label-free experiment carried out by Michael Henry and Dr Joanne Keenan in the NICB on whole cell lysates of DLKP and its clonal subpopulations. This work was not followed up; however the preliminary findings strengthen the discovery of Shootin-1 as associated with high expression in the DLKP-SQ clonal subpopulation. In addition, microarray analysis carried out by Dr Helena Joyce on the DLKP clones found the Shootin-1 gene to be expressed highly in DLKP-SQ relative to both DLKP-M and DLKP-I (Joyce H, PhD 2015, data not published). Taken together, these findings increased confidence in Shootin-1 being a target associated with high expression in DLKP-SQ.

In order to shed light on the clinical relevance of Shootin-1 with respect to lung cancer, a Kaplan Meier plot was created using KM plotter to determine if the target is a potential biomarker or target for lung cancer treatment. The Kaplan Meier plotter is a bioinformatics tool capable of assessing the effect of 54,675 genes on survival using 10,188 cancer samples (www.kmplot.com). These include 4,142 breast, 1,648 ovarian, 2,437 lung and 1,065 gastric cancer patients. The resulting plot can be used to estimate survival function from lifetime data, with respect to proteins of interest in the case of our study presented here.

The resulting plot is shown in Figure 7.1 and shows the overall survival data based on 1926 patients. The plot suggests that Shootin-1 may be a good prognosis indicator, as high levels of protein expression are associated with a better outcome and prolonged survival. This suggests that Shootin-1 may not be very useful as a therapeutic target as low levels are associated with poorer outcome; however it may still be useful as a prognostic biomarker. Highest levels of Shootin-1 were found to be expressed in DLKP-SQ which is the least migratory and invasive clone as determined by previous work and also this study. These findings fit in with the results of the Kaplan-Meier plot shown in Figure 7.1 as better prognosis is associated with high levels of Shootin-1 expression, suggesting Shootin-1 is associated with a less aggressive metastatic phenotype.

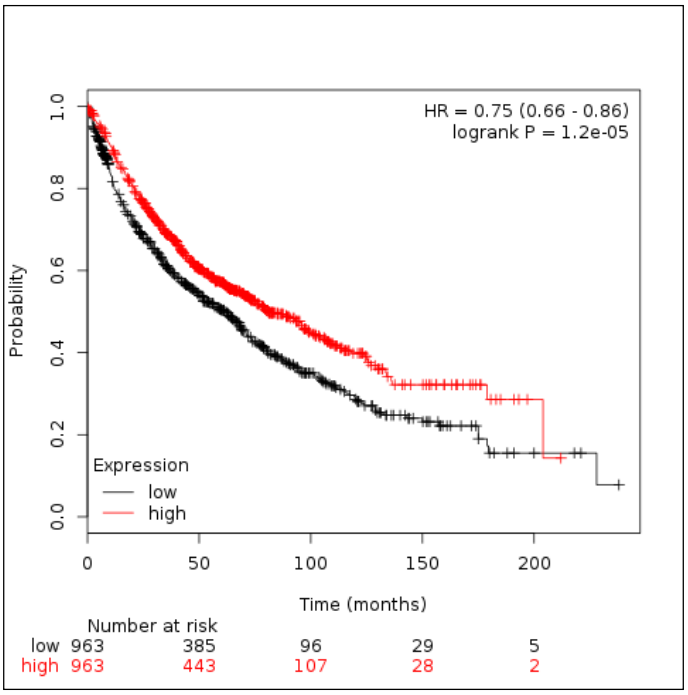


Figure 7.1: Kaplan Meier plot for Shootin-1 showing overall survival data regarding lung cancer patients.

7.3. MARCKS

Label-free proteomic analyses highlighted MARCKS as highly expressed in DLKP-M. The protein expression pattern suggests MARCKS expression is significantly higher in DLKP-M compared to the remaining clonal subpopulations and the DLKP cell line. The combination lines also show relatively high expression of MARCKS in the samples which contain DLKP-M as a component.

Myristoylated alanine-rich C-kinase substrate is a ubiquitously expressed heat-stable protein localized to the plasma membrane. It is reversibly tethered to the plasma membrane via a myristoyl group located at its N-terminus, and also by ionic interactions between membrane phospholipids and the MARCKS effector domain (ED) (Albert, Nairn and Greengard 1987, Arbuzova, Schmitz and Vergeres 2002). MARCKS is phosphorylated by Protein Kinase C (PKC) and depending on its phosphorylation status, the protein shuttles between the plasma membrane and the cytosol. This translocation to the cytoplasm is also induced by the binding of Calmodulin (CAM) in a Calcium (Ca^{2+}) dependant manner to the ED (Hartwig et al. 1992, Yarmola et al. 2001). At the plasma membrane, MARCKS cross-links Actin filaments into bundles, significantly altering the Actin cytoskeleton.

MARCKS has been reported to be highly expressed in a wide range of cancers such as breast, melanoma, and cholangiocarcinoma (Jonsdottir et al. 2012, Micallef et al. 2009, Techasen et al. 2010). MARCKS has been directly linked to motility, where it was found that phosphorylation of MARCKS directly promotes motility of melanoma cells (Chen and Rotenberg 2010). MARCKS has been extensively studied in the lungs in relation to its role in regulating the secretions of granules in goblet cells (Haddock et al. 2014); however, it has only been the subject of two studies to date in lung cancer. It has been identified as a potential biomarker associated with poor prognosis in human SCC using immunohistochemical methods (Hanada et al. 2013). The second study detected high expression of MARCKS in highly invasive lung cancer cell lines and found that MARCKS knockdown by siRNA significantly reduced cell migration (Chen et al. 2014b).

In order to validate the expression pattern found for MARCKS by quantitative label-free LC-MS/MS, all cell line samples were analysed by western blot to determine the protein expression pattern for MARCKS using a second proteomic technique. The western blot analysis results correlate well with the label-free proteomic data, validating the protein expression pattern for MARCKS. DLKP-M shows the highest abundance of MARCKS with strong bands visible at the expected MW of 80kDa. DLKP-I shows the next highest abundance as expected, with weak bands visible. Both DLKP-SQ and DLKP show no detectable expression of MARCKS in either cell line.

Analysis of the combination lines by western blot show MARCKS abundance range from highest to lowest as follows: I+M; SQ+M; SQ+I; DLKP-Remade. These are the expected abundance levels for the combinations based on the data generated from the clones. The combination lines are made up in a 1:1 ratio, therefore these findings show that MARCKS expression is not significantly affected by co-culture in DLKP-SQ and DLKP-M. The western blot findings also clarify the unexpectedly low MARCKS abundance value found by label-free proteomic analysis in the I+M cell line described in Section 5.5.1. This cell line now shows an expression level for MARCKS by western blot at the expected higher abundance. It is likely that instead of DLKP-I inhibiting MARCKS, the low abundance value is most likely down to the large standard deviation found for that sample, skewing the levels downward. This is somewhat corroborated by the peptide expression profile for MARCKS in Figure 5.4 which shows a peptide outlier bringing the value down. Finally, both DLKP and DLKP-Remade show similarly low MARCKS abundances by label-free proteomic and western blot analyses methods, both being undetectable in the latter cell line. This is unsurprising as DLKP-M constitutes approximately 5% of the cell population in both cell lines.

Overall, western blot analysis correlates well with the label-free proteomic analyses, validating the expression pattern found for MARCKS by the global analysis, and also by the two-sample experimental designs. MARCKS is shown to be significantly highly expressed in DLKP-M relative to all other samples denoting the protein as a strongly abundant in DLKP-M relative to the other cell lines. Following on from this, MARCKS was analysed at an mRNA level using qPCR.

Quantitative PCR was carried out across the sample set to assess the mRNA level of MARCKS in the DLKP cell lines as well as the combinations. Analysing the abundance of MARCKS mRNA in relation to its known protein expression abundances may highlight post-transcriptional processes which are known to play key roles in the final synthesis of a protein. The findings from the MARCKS gene expression assay strongly correlate with the label-free proteomics data in that they show the same expression profile in both datasets. DLKP-M shows significantly high expression of MARCKS in DLKP-M at both mRNA and protein levels. Protein analysis showed a fold change of 49.42 between DLKP-M and DLKP, and at the mRNA level this figure is increased to a fold change of 66.66. This shows that DLKP-M is the strongest expresser of MARCKS at both the protein and mRNA level by a significant margin. DLKP-I expression of MARCKS mRNA is lower than expected, with a fold change of 3.78 between DLKP-M and DLKP-I. This further separates DLKP-M from DLKP-I in relation to MARCKS expression, reinforcing MARCKS expression as being associated strongly with DLKP-M. Both DLKP and DLKP-SQ show very low levels of MARCKS mRNA expression, similar to the finding of the proteomic analyses, with DLKP-SQ being the slightly higher expresser.

The combination cell lines show the same expression profile for MARCKS at the mRNA level as they show at a protein level. The mRNA abundances for MARCKS are slightly higher than expected in the cell lines SQ+M and I+M relative to DLKP-M when the data is compared with the protein abundance findings. High expression of mRNA relative to the known protein expression may suggest that MARCKS mRNA has a long half-life, and does not degrade quickly. It has been shown previously that MARCKS mRNA has a long half-life in murine fibroblast cells (Wein et al. 2003). This may explain the high mRNA levels relative to protein levels in our combination cell lines, indicating that post transcriptional factors keep the protein level lower than the mRNA level.

Validation studies in this work so far have shown MARCKS to be a highly expressed in the DLKP-M cell line compared to the other cell lines. With this in mind, MARCKS was stained using immunofluorescence in each of the four cell lines: DLKP-SQ, DLKP-M, DLKP-I and DLKP. This was carried out to further validate the

expression pattern as found by the methods previously discussed, and to establish the localization of MARCKS in the DLKP cell lines, with a focus on DLKP-M. It was expected that DLKP-M would show the strongest staining for MARCKS based on the previous findings, with the next most intense staining in DLKP-I, and lowest in both DLKP-SQ and DLKP. Visualisation of MARCKS by immunofluorescence correlates with the proteomic analyses in general.

Immunofluorescence analysis shows that DLKP-M has the strongest staining for MARCKS out of all the cell lines analysed. This cell line is characterized as mesenchymal-like, and has morphology rich in outgrowths and extended cytoplasmic formations. Interestingly, extremely strong staining for MARCKS shows localization in these protruding outgrowths. The protein has previously been implicated in Actin dynamics and cellular outgrowths in neuronal cells (Calabrese and Halpain 2005, Larsson 2006). This suggests that MARCKS may play a role in cell motility in the DLKP cell lines also. In DLKP-M, MARCKS also shows localization to the plasma membrane/cytoplasm, with high punctate staining patterns visible here, and it is absent from the nuclear area. Punctate staining patterns indicate that MARCKS accumulates in highly concentrated forms in DLKP-M. Previous work has found MARCKS to show a similar punctate staining pattern in macrophages (Rosen et al. 1990), and it was determined that MARCKS was localized to the substrate-adherent surface in the filopodia of these cells. Subsequent phosphorylation of MARCKS by PKC liberates MARCKS from these punctate structures and induces its translocation to the cytosol (Green et al. 2012).

DLKP-I has the second strongest staining intensity for MARCKS as determined by Immunofluorescence imaging. This correlates with the proteomic analyses findings which show DLKP-I as the second highest MARCKS expresser. However, the positive staining in DLKP-I for MARCKS is found in subpopulations within the DLKP-I cell line. The presence of this subpopulation may be explained by the plastic nature of DLKP-I as previously described (McBride et al. 1998b). This work found that DLKP-I may be an intermediate cell line capable of inter-converting to either DLKP-M or DLKP-SQ. It was hypothesised that DLKP-I acts as a stemcell-like population which can give rise to either of the other two clones depending on growth conditions. Therefore, it is possible that subpopulations of the cells within DLKP-I are tending towards converting to DLKP-M morphology, indicated by the strong expression

of MARCKS in these cells. This hypothesis is strengthened by the presence of outgrowths in the DLKP-I cells expressing MARCKS, showing them tending towards established DLKP-M morphology.

The DLKP-SQ cell line shows no staining for MARCKS in ~90% of the cell population. However, the protein does stain positive in a small subpopulation of cells within DLKP-SQ. The presence of these cells expressing MARCKS helps shed light on the unexpected finding that DLKP-SQ expresses a higher abundance of MARCKS than the DLKP cell line as found by our proteomic analyses. It is surprising DLKP-SQ expresses MARCKS at a higher level than DLKP once proportions of each clone present in DLKP is considered: 5% DLKP-M and 25% DLKP-I and 70% DLKP-SQ as was reported by McBride et al. 1998. It would be expected that the presence of DLKP-M and DLKP-I in the DLKP cell line would boost the abundance of the protein in the DLKP cell line as they both express MARCKS. The abundance is so unexpectedly low in DLKP that it is likely an inhibitory effect is occurring. This strengthens the hypothesis that there may be other uncharacterised cell types present in the DLKP cell line which reduce MARCKS expression by inhibitory processes.

The staining of MARCKS in DLKP results in varied intensities which can be explained by the heterogeneous nature of the cell line. Clear staining for MARCKS can be seen in cellular outgrowths and, similar to Shootin-1, localization is strong in the tips of these outgrowths. This increases the likelihood of MARCKS as being involved in cell motility processes in the DLKP cell line model as has previously been found in mouse melanoma cells, vascular smooth muscle cells, and neuronal cells (Chen and Rotenberg 2010, Yu et al. 2015, Li et al. 2008b).

DLKP-M is known to show the highest cell motility out of all the clonal subpopulations (Joyce H, PhD 2015). These phenotypic differences are stark, and DLKP-M is associated with very high migration and invasion relative to the other DLKP clones. Though MARCKS has been shown to be over-expressed in a variety of aggressive cancers such as pancreatic and breast cancer (Gardner-Thorpe et al. 2002, Jonsdottir et al. 2012), there have been limited studies on MARCKS in cancer metastasis. MARCKS has however been shown to be involved in the migration and invasion of melanoma cells (Chen and Rotenberg 2010). They found that MARCKS is a major determinant in the motility of mouse melanoma cells *in vitro*. Weakly metastatic

melanoma cells show undetectable levels of phosphorylated MARCKS, whereas aggressively metastatic cells show abundant expression of the same protein. Knockdown and overexpression studies confirmed MARCKS as being a key determinant in the metastatic process, and also elucidated a role for phospho-MARCKS which was previously unclear. The high staining of MARCKS in the outgrowths of cells in DLKP-M and DLKP suggest the protein may play a role in cell motility in the DLKP cell line model. Therefore it is hypothesised that MARCKS may play a role in both migration and possibly invasion, promoting these phenotypes. This was investigated by functional analysis assays after siRNA knockdown of MARCKS in DLKP-M.

To investigate the functional role played by MARCKS in DLKP-M, the protein was knocked down using RNAi and the effects of this process were subsequently determined using a panel of functional assays. The effects of MARCKS RNAi were verified by western blot analysis to ensure the protein was effectively knocked down prior to the functional analyses being carried out. DLKP-M was chosen as the cell line to investigate due to its significantly high expression of MARCKS in this cell type. Functional assays measured proliferation, migration and invasion of the DLKP-M cell line

The functional analysis of MARCKS after RNAi knockdown shows that MARCKS plays a significant role in the motility of DLKP-M for both migration and invasion. Firstly, RNAi of MARCKS in DLKP-M significantly reduces the migratory ability of the cell line. The effect is observable in both siRNA molecules which are used to knockdown the MARCKS protein and in both cases the effect is highly significant. Migration of DLKP-M in knockdown samples is reduced by an average of 70% compared to the negative control. This finding is consistent with our hypothesis that MARCKS is involved in the motility of the DLKP-M cell line. In addition, knockdown of MARCKS also reduces the invasive ability of DLKP-M with high significance. Invasive ability is reduced by an average of 71% when knockdown samples are compared to the negative control. These findings demonstrate that MARCKS is intimately involved in the motility of DLKP-M as suggested in Section 5.6.1.

Previous studies shed light on the mechanisms by which MARCKS may be involved in cell migration. As discussed earlier, MARCKS has been shown to shuttle

between the plasma membrane and the cytoplasm depending on its phosphorylation state (Hartwig et al. 1992, Larsson 2006). Unphosphorylated MARCKS is located to the plasma membrane and cross-links actin filaments to the cytoplasmic side which results in alterations to the actin cytoskeleton. The translocation of MARCKS to the cytoplasm upon phosphorylation promotes cell spreading by releasing actin-membrane cross links, resulting in cytoskeletal plasticity which in turn promotes the ability of cells to spread. De-phosphorylation of MARCKS causes its re-localisation to the plasma membrane, induces actin cross-linking and causes cytoskeletal conformational changes once more. These transient interactions with the actin cytoskeleton are thought to regulate the dynamics of F-actin which in turn impacts cell motility (Chen and Rotenberg 2010).

This reversible localization of MARCKS is controlled by phosphorylation of the protein, mainly by PKC. This mechanism may be a key system which regulates cell motility, and this seems to be the case for DLKP-M. This cell line has been shown to be highly motile and a reduction in the abundance of MARCKS by RNAi has significantly reduced the ability of the cells to migrate.

Knockdown of MARCKS in DLKP-M was found to have no effect on cell proliferation when compared to the negative control sample. Previous studies have found conflicting results on the role MARCKS plays in proliferation. One group studying glioblastoma multiforme (GBM) found that RNAi knockdown of MARCKS promoted proliferation of the cells, and over-expression of the protein greatly reduced the growth rate and induced senescence (Jarboe et al. 2012). This suggested that high MARCKS expression reduces cell proliferation. A different study found that treatment of fibroblasts by a cell permeable peptide which inhibits the function of MARCKS had no effect on the proliferation of the cells (Chen et al. 2014b). Therefore, the finding that MARCKS knockdown has no effect on the proliferation rate of DLKP-M is unsurprising.

In our study so far, MARCKS has been shown to significantly reduce migration and invasion in the DLKP-M cell line. Immunofluorescence studies have shown the localization of MARCKS to be at both the cytoplasm and at plasma membrane in concentrated punctate structures. MARCKS also has particularly strong localization in cell projections. Literature studies reveal that MARCKS is highly associated with F-actin, and is involved with cross linking this protein to the plasma

membrane through the two binding sites located in the effector domain of MARCKS (Yarmola et al. 2001). The association between MARCKS and F-Actin has not been studied as of yet in lung cancer. It was therefore decided to knockdown MARCKS by RNAi, and co-stain both F-actin and MARCKS using confocal laser microscopy. It was hoped this investigation would clarify the relationship between F-actin and MARCKS in DLKP-M as we hypothesise both proteins interact. This hypothesis is based on the staining pattern found by our immunofluorescence in combination with what is known from the literature (Larsson 2006, Rombouts et al. 2013). These studies found MARCKS to be an F-actin binding protein in neuronal and hepatic stellate cells involved in the dynamic reorganisation of the actin cytoskeleton.

Our results from the confocal imaging of MARCKS and F-actin in the DLKP-M cell line show strong staining for both proteins in the control conditions where MARCKS was not knocked down. Similar to the immunofluorescence findings, MARCKS shows localization to the cytoplasm and plasma membrane and intense staining in cellular protrusions characteristic of DLKP-M. MARCKS also shows strong punctate staining patterns which suggest high concentrations of the protein localize to particular regions of the cell. There is faint nuclear staining in some samples within the control groups, but it is weak. These findings corroborate the results from the immunofluorescence work which show the same staining patterns for MARCKS. The imaging of both proteins shows that co-localisation for MARCKS and F-actin is present in DLKP-M. This is particularly apparent in the cellular outgrowths and punctate staining regions, and both proteins show strong overlap in these cases. This strengthens the hypothesis that MARCKS and F-actin interact in DLKP-M, and this may be one of the key processes which is interrupted upon RNAi, resulting in decreased migration and invasion of the cell line.

Confocal imaging of DLKP-M treated with MARCKS RNAi results in a decrease of MARCKS staining for both siRNA molecules used. This is accompanied by a visible decrease in the staining strength of F-actin. In the control group (Cell-only, Lipofectamine-only and Negative control), F-actin staining is moderately high, and localisation is present in cell protrusions and punctate structures similar to MARCKS. F-actin staining is also strong in cell-cell junctions and at the edges of cells, which seem to be the leading edges, or lamellipodia, of the DLKP-M cells. MARCKS knockdown samples show a marked decrease in staining of F-actin, particularly in the areas

overlapping with MARCKS staining. However low abundance staining remains in some of the lamellipodial structures, indicating knockdown of MARCKS does not strongly affect the expression of F-actin at these regions. There is a visible decrease in punctate and cell protrusion staining for F-actin in the MARCKS knockdown samples, which suggests that a reduction in MARCKS expression has resulted in a concomitant reduction in F-actin localisation to these areas of the cell. These results show that MARCKS is very likely an actin interacting partner in DLKP-M. To further investigate this link, and establish if the interaction is direct or indirect, immunoprecipitation studies were attempted which aimed to pull out MARCKS from DLKP-M and any binding partners which came with it. Unfortunately, repeated attempts using two independent MARCKS antibodies failed as the protein was not immunoprecipitated using the antibodies available. This will be pursued in the future if suitable antibodies for immunoprecipitation can be sourced.

Confocal imaging of MARCKS in the knockdown samples by RNAi resulted in an unusual localisation of the protein. Firstly, the knockdown samples show decreased staining for MARCKS overall, with markedly reduced cytoplasmic and membrane staining, as expected. Additionally, there is also a decrease in cellular protrusions visible in the knockdown samples compared to the control group, which suggests MARCKS is involved in the morphology of the DLKP-M cell line. This finding is in-line with the effect of knockdown of MARCKS on cell migration and invasion, decreasing them both significantly. However, imaging of MARCKS in the knockdown samples shows that the MARCKS protein which remains in the cells potentially shows significant re-localisation to the nuclear area. This striking translocation of MARCKS to the nucleus is clearly visible in the confocal images for both siRNA molecules which were used to reduce the expression of the protein.

Western blot analysis of DLKP-M cells treated by MARCKS RNAi show that the highest concentration of siRNA results in almost complete knockdown of the protein as seen in Figure 5.8. Remaining protein can be seen for both siRNAs used, and is more abundant in siRNA#2. It is this remaining protein which may be translocated to the nucleus of the cells after RNAi treatment. MARCKS which remains after RNAi is much more visible in the confocal images compared to the western blot images. This may be due to the much higher concentration of MARCKS primary antibody (25 times more) used in the cell staining method compared to the western blot analysis, as per the

manufacturers antibody usage guidelines. MARCKS translocation is not a certainty, and further work is required to confirm this finding by the use of nuclear isolation and MARCKS quantification. If it is the case, translocation of MARCKS to the nucleus of DLKP-M is an unexpected effect of the protein being knocked down by RNAi. Translocation of MARCKS to the nucleus is not completely unheard of, and literature studies show three papers which previously described MARCKS behaving in this manner, though not after RNAi.

One study found MARCKS to translocate to the nucleus from the cytoplasm upon over-expression in HEK 293 cells (Spizz and Blackshear 2001). Following this, a separate group (Rohrbach et al. 2015) investigated the mechanism by which MARCKS translocates to the nucleus upon overexpression in glioblastoma cell lines. They induced mutations into various domains of MARCKS to determine if it resulted in the inhibition of nuclear localisation. Mutations in the myristoyl tail and other regions of MARCKS did not inhibit translocation to the nucleus, therefore these protein domains were deemed as having no relation to the translocation of MARCKS. However, mutations in the Effector Domain (ED) of MARCKS effectively prevented translocation of MARCKS to the nucleus, an indication that this domain is involved in the process. In addition, they determined the ED of MARCKS contains homologous sequences to known nuclear localisation sequences (NLS). Therefore, Rohrbach et al. 2015 determined that the ED of MARCKS is the critical domain that regulates nuclear localization of MARCKS upon over-expression.

A third study focused on the expression of MARCKS in cholangiocarcinoma (CCA), where it had been found to be over-expressed by immunohistochemical analysis of paraffin embedded liver tissue samples (Techasen et al. 2010). They followed up this discovery by investigating the effects of MARCKS phosphorylation on CCA cell lines and assayed these effects by confocal microscopy, migration and invasion assays. Induced phosphorylation of MARCKS in this case was stimulated by the use of a PKC activator called TPA. Confocal microscopy revealed that upon stimulation of MARCKS using TPA induced phosphorylation by PKC, MARCKS visibly translocated to the nucleus. This process also resulted in a significant increase in the migratory and invasive abilities of the CCA cell lines tested (Techasen et al. 2010).

Our study is the first to report nuclear translocation of MARCKS in a lung cancer cell line. In addition, our study found this translocation to the nucleus to be stimulated by knockdown of the protein, rather than over expression or induced phosphorylation as previously reported (Rohrbach et al. 2015, Techasen et al. 2010). Further work is required to validate this observation, and phosphorylated MARCKS will also need to be investigated to determine whether it is MARCKS or phospho-MARCKS which undergoes translocation. The association of MARCKS with high expression is strengthened by unpublished work carried out by Dr Helena Joyce (Joyce H, PhD 2015). DNA microarray analysis showed the MARCKS gene to be highly expressed in the DLKP-M clone relative to both DLKP-SQ and DLKP-I. Lowest expression was found in DLKP-SQ and MARCKS expression in DLKP-I was approximately half that found in DLKP-M. This correlates with the expression pattern in our study, further strengthening this expression pattern.

Examination of the clinical relevance of MARCKS in to lung cancer progression was carried out using KM plotter to generate a Kaplan-Meier plot showing overall survival of lung cancer patients with respect to MARCKS expression (Figure 7.2). From the plot, MARCKS appears to be a marker of poor prognosis. High expression of the protein is associated with a lower probability of prolonged survival in lung cancer. Our study found MARCKS to be highly expressed in the DLKP-M clonal subpopulation relative to the other cell lines, and it is DLKP-M which has been shown to have the highest metastatic phenotypes. It has the highest levels of both migration and invasion relative to the other clones, and these finding correlate with the Kaplan-Meier plot showing high expression of MARCKS associated with poor outcome. The two papers which investigated MARCKS expression in lung cancer also corroborate this finding. A study by (Hanada et al. 2013) compared a cohort of primary SCC tissue samples against non-cancerous adjacent tissues using mass spectrometry and immunohistochemical analyses found that MARCKS was upregulated by ~2 fold in the SCC samples. Subsequent immunohistochemical evaluation in another set of SCC patients found high expression of MARCKS to be associated with poor outcome, suggesting the protein could be a useful biomarker for outcome in SCC. A separate study by (Chen et al. 2014a) found that higher levels of phospho-MARCKS were associated with shorter overall survival of lung cancer patients. Blocking the phosphorylation site of MARCKS with a peptide synthesised in the lab resulted in a

reduction in MARCKS phosphorylation and a concomitant reduction in tumour growth and metastasis *in vivo*. Taken together, these studies suggest MARCKS may play an important role in the clinical setting as both a biomarker for poor outcome in SCC and also as a therapeutic target.

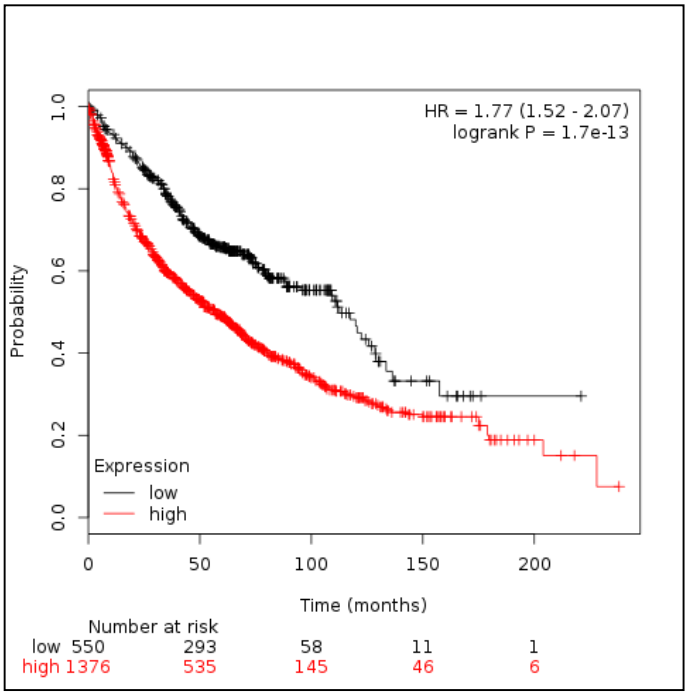


Figure 7.2: Kaplan Meier plot for MARCKS showing overall survival data for lung cancer patients.

7.4. Desmoglein-3

DSG3 is shown to be a strong candidate marker for DLKP-I by label-free proteomic analyses. Desmoglein 3 (DSG3) is a transmembrane glycoprotein exclusively expressed in stratified epithelium. It is member of the desmoglein family of proteins, which in turn are a part of the cadherin superfamily of cell adhesion molecules. Expression of DSG3 is limited to the basal and supra-basal layers of the skin, however it is expressed homogeneously throughout the stratified epithelium in oral mucosa (Amagai et al. 1996, Garrod and Chidgey 2008). DSG3 is a major component of the desmosome, which is an adhesive intercellular junction critical to the integrity of tissues that experience shear stress. Homophilic binding of the DSG3 extracellular domain to adjacent cells leads to cell-cell adhesion, while the cytoplasmic tail of DSG3 links to internal keratin intermediate filaments via linker proteins such as the Armadillo and Plakin families (Dehner et al. 2014, Garrod and Chidgey 2008, Delva, Tucker and Kowalczyk 2009). DSG3 has been extensively studied in relation to the auto-immune disease *Pemphigus Vulgaris* (PV), in which autoantibodies bind to the extracellular domain of DSG3 and cause internalization of the protein (Jennings et al. 2011, Sato, Aoyama and Kitajima 2000). This results in a signalling cascade which leads to loss of cell-cell adhesion in the basal and immediate supra-basal layers of stratified squamous epithelia as well as the oral mucosa, causing painful blistering and lesions.

DSG3 has been found to be overexpressed in head and neck cancer (HNC) by PCR analysis in tissue samples when compared to normal keratinocytes and other cancer types (bladder, neuron, colon and liver) (Chen et al. 2007). DSG3 was also found to exhibit upregulation and abnormal localization in esophageal squamous cell carcinoma (ESCC). Using immunohistochemical staining, DSG3 was found to be redistributed to the cytoplasm from the membrane in the majority of ESCC cases (Wang et al. 2007). This paper hypothesised that DSG3 may play a role other than functioning as a cell adhesion molecule, and may be involved in cell migration and invasion.

The quantitative proteomic analyses performed on DLKP and its clonal subpopulations highlight DSG3 as specifically and highly expressed in DLKP-I, with little or no expression of the protein in the other three cell lines (DLKP-SQ; DLKP-M; DLKP). To investigate these findings and validate the expression pattern found by the label-free proteomic analyses, validation studies were carried out using western blot

analysis. Furthermore, qPCR analysis was carried out on the sample set to determine the relative abundance of DSG3 mRNA in all the cell lines.

The expression pattern for DSG3 by WB analysis correlates very well with the label-free quantitative proteomics findings. Analysis of DSG3 by WB shows up bands at the expected weight of 130kDa in all samples analysed. DLKP-I has the highest abundance of DSG3 out of the four DLKP cell lines. DSG3 expression in the remaining cell lines (DLKP-SQ; DLKP-M; DLKP) is not detectable by WB and suggests the protein is extremely low or absent in these cell lines. The WB analysis clarifies the expression levels found by label-free proteomic analyses of DSG3 in these cell lines. Results suggest that these cell lines most likely do not express DSG3, at least to levels which are detectable by Western Blot. This strengthens the position of DSG3 as a marker for DLKP-I

WB analysis for DSG3 also supports the differential expression found between SQ+I and I+M with higher abundance validating in the former cell line. These findings suggest that the co-culture cell line I+M has resulted in an inhibition of DSG3 expression in DLKP-I by the presence of DLKP-M. It is unlikely that the DSG3 abundance difference between these two combination lines is due to one population outgrowing another. This can be deduced from the label-free proteomic analysis where standard deviations for these combination lines show little variation across the three biological replicates which make them up (Figure 6.3B). This effect may be induced by paracrine signalling from DLKP-M resulting in inhibited DSG3 expression. The finding highlights the effects of cellular heterogeneity in the tumour microenvironment with regard to protein expression and potentially cell phenotypic characteristics.

Another surprising finding from the proteomic analysis on the combination cell lines was that the abundance of DSG3 in DLKP-Remade is higher than what is found for the DLKP cell line. This protein expression pattern was also subsequently validated by WB. The difference in DSG3 abundance levels between DLKP and DLKP-Remade strengthens the ‘undiscovered subpopulation’ hypothesis which was formulated upon finding similar results for MARCKS and Shootin-1. This hypothesis suggests that the DLKP cell line may contain as of yet undiscovered and uncharacterised cell populations which interfere with the expression of some proteins. If DLKP is made up of only DLKP-SQ, DLKP-M and DLKP-I in proportions which mimic DLKP, one would

expect the protein expression patterns of both cell lines to match each other. As this is not the case, the possibility of undiscovered clonal subpopulation remains. The difference in DSG3 abundance between DLKP-Remade and the parental cell line could also be due to cross-talk between the known DLKP subpopulations, resulting in altered protein expression. Further work is needed to establish these hypotheses.

Quantitative PCR analysis was performed on the sample set to measure the relative gene expression level of DSG3 in the DLKP cell lines as well as the combinations. The findings of the gene expression assay correlate closely with both the label-free and WB proteomic analyses. DLKP-I shows significantly high expression of DSG3 mRNA levels relative to all other samples in the analysis. Proteomic data had shown a maximum fold change of 175.33 for DSG3 between DLKP-I and DLKP-SQ. Quantitative PCR analysis shows a fold change of 6288 between the same two cell lines indicating that DSG3 is highly expressed in DLKP-I. The analysis also suggests that DSG3 is expressed at extremely low or absent levels in DLKP-SQ and DLKP-M at the mRNA level as the relative quantity abundance levels between these cell lines and DLKP-I are very high.

The qPCR analysis on the combination cell lines shows the same expression pattern as is found in the proteomic analyses. The cell line SQ+I shows a 2.6 higher DSG3 mRNA level than is found for I+M. This suggests that inhibition of DSG3 expression in DLKP-I by DLKP-M is occurring at a post-transcriptional level, preventing or interfering with the synthesis of DSG3 mRNA in co-culture conditions. This further highlights the range of interacting effects heterogeneity in the tumour microenvironment can have on cell populations at both the protein and mRNA levels.

The difference in DSG3 protein abundance between DLKP and DLKP-Remade has also been corroborated by the gene expression analysis. This finding supports the idea that there are uncharacterised cell populations remaining in DLKP which have yet to be discovered. The existence of such populations may be the reason why DLKP and DLKP-Remade do not behave the same at the protein and mRNA level. This data also shows that DSG3 is expressed at an mRNA level, whereas the proteomic analyses suggest extremely low or absent expression of DSG3. The mRNA expression pattern correlates well with the cDNA microarray analysis carried out by Dr Helena Joyce where each clone was compared to each other as a group (Joyce H, PhD 2015) in

unpublished work. DSG3 was found to be essentially absent from both DLKP-M and DLKP-SQ, which is what has been found in the current study also. This strengthens DSG3 as a potential marker for DLKP-I. Combination cell lines also showed high expression of DSG3 in the DLKP-I containing cell lines. These findings were validated at the protein and mRNA levels.

Thus far, this work has shown DSG3 to be specifically highly expressed in DLKP-I compared to the other clones. The various assays performed on the sample set suggest that the high abundance of DSG3 is specific to DLKP-I and is either very low or absent from the other clonal subpopulations. To investigate this expression pattern further, Immunocytochemical and Immunofluorescent staining of DSG3 were performed on the DLKP cell line and its clonal subpopulations. This was carried out to validate the expression pattern found by this work so far, and to establish the localisation of DSG3 in the DLKP cell lines. It was expected that DLKP-I would show the most intense staining for DSG3, with little or no staining in the other clones. In addition, it was uncertain whether the DLKP cell line would show positive staining for DSG3, as this target shows detectable levels at an mRNA level, but extremely low to undetectable protein levels in this cell line.

Immunocytochemical staining of DSG3 shows DLKP-I as having very high expression relative to DLKP-SQ; DLKP-M; and DLKP. This correlates well with the previous proteomic analyses which established this same expression pattern. Staining is highest in DLKP-I and essentially absent from the other two clonal subpopulations. The DLKP cell line shows positive staining for DSG3, where it seems to be expressed very specifically by a subpopulation within the cell line. This could be reasonably hypothesised to represent the DLKP-I component within DLKP. The stark staining contrast between the DLKP-I subpopulation and the surrounding cells in DLKP suggest that expression of DSG3 is highly specifically expressed, and supports its role as a potential marker of this clonal subpopulation.

Localisation of DSG3 in DLKP-I shows high intensity at cell-cell junctions as would be expected due to the known role of the protein as a desmosomal component (Dusek and Attardi 2011). However, our study found that DSG3 shows abnormal cytoplasmic and perinuclear staining in a large proportion of the DLKP-I cells.

Uncharacteristic localisation of DSG3 has been previously reported in relation to esophageal squamous-cell carcinoma (Fang et al. 2014). They reported that abnormal cytoplasmic localisation of DSG3 correlated with regional lymph node metastasis in esophageal squamous-cell carcinoma, and in patients without lymph node metastasis the cytoplasmic localisation of DSG3 correlated with poor prognosis. They also refer to a previous study which investigated a panel of desmosomal proteins and found DSG2, a Cadherin family member homologous to DSG3, showed cytoplasmic localisation correlated with poor tumour differentiation in esophageal squamous-cell carcinoma also (Fang et al. 2010). A study by a different group reported on the localisation of DSG3 in oral squamous-cell carcinoma, and found that perinuclear staining for the protein as well as other Cadherin proteins significantly correlated with disease progression (Wang et al. 2007).

Immunofluorescent staining of DSG3 in the DLKP cell lines shows the same trend as was found using the previous staining method; however this higher resolution technique detected additional DSG3 localisation not visible by immunocytochemistry. Where DSG3 staining is present, it is strong in the cell-cell interactions as was previously found. In addition, a punctate staining pattern is visible which is not visible by ICC. Highly specific dots of DSG3 are visible in distinct points in the DLKP-I cells. These may be desmosomal-bound DSG3 dotted across the cell membrane. High nuclear staining can also be seen in some of the cells, and these add to what was shown in the previous staining method.

In our study, cytoplasmic and perinuclear staining of DSG3 is visible in DLKP-I. This suggests that DSG3 functions as more than merely a cell-cell adhesion molecule and may play a significant role in the metastatic characteristics of DLKP-I. Previous studies have suggested that DSG3 may serve as a signalling centre which extends the role of the protein beyond just cell-cell adhesion, and a study has implicated the protein to be an upstream regulator of Src activity in epithelial cell lines (Chidgey and Dawson 2007, Tsang et al. 2010). A recent study found that overexpression of DSG3 increased the motility of an oral squamous carcinoma cell line by causing enhanced phosphorylation of Ezrin by PKC(Brown et al. 2014). They found DSG3 forms a complex with Ezrin at the plasma membrane in OCC which is required for the proper formation of the Ezrin-F-actin interaction, and knockdown of DSG3 impaired this interaction. Taken together, these reports implicate DSG3 as having a wide range of

potential roles in addition to its function in the desmosome as a cell adhesion molecule. The abnormal localisation of DSG3 in DLKP-I may involve some of the pathways described above, and the protein may play a role in the motility of the DLKP-I cell line.

The subpopulation of DLKP cells staining positive for DSG3 in our ICC and IF analyses also shows localisation of the protein at the edges of these cells. Intense staining can be seen at lamellipodial-like structures where the cells appear to be spreading outwards. This suggests DSG3 may co-localise with F-Actin in DLKP-I as was found by Brown et al. 2014, and may be involved in the regulation of cell motility in this role. A previous study found DSG3 to interact with Actin in human epithelial cell lines and showed overexpression of DSG3 promotes Actin dynamics leading to enhanced formation of lamellipodial and filopodial protrusions (Tsang et al. 2012). This supports the idea that DSG3 may be involved in cell motility in the DLKP-I clonal cell line and the DLKP lung cancer cell line model.

To investigate the role of DSG3, a panel of functional analyses were carried out on the DLKP-I cell line after knockdown of DSG3 by RNAi. DLKP-I was chosen as the subject for analysis as DSG3 is potentially a marker for this cell type, and it is also the only DLKP clone to express DSG3 protein. The effects of DSG3 RNAi were verified by Western Blot analysis to ensure the protein was effectively knocked down prior to functional analyses. The assays which were performed on DLKP-I measured proliferation, migration and invasion.

The functional analyses of DSG3 shows the protein to play a major role in the motility of DLKP-I. Both the migratory and invasive abilities of DLKP-I show a significant reduction upon knockdown of DSG3, with invasion being more strongly affected than migration. The migratory ability of DLKP-I shows an average of 28% reduction in the siRNA treated samples compared to the Negative control, and both siRNA molecules have the same effect. Invasiveness of DLKP-I is reduced by 50% in knockdown samples compared to the negative control. These findings suggest that DSG3 plays a key role in the motility of DLKP-I. No effect on proliferation was found upon DSG3 knockdown. This finding marks the first time knockdown of DSG3 has reduced cell motility in a lung cancer cell line.

This finding that DSG3 knockdown reduces cell motility is supported by previous studies which have shown the protein to be involved in the metastatic phenotype of other cancer types. Previous studies have found overexpression of DSG3 to be associated with the metastatic phenotype of Inverted Papilloma SCC and also lymph node metastasis in HNSCC (Huang et al. 2010, Patel et al. 2013). Another study compared normal epithelial tissue against tissue samples from head and neck cancer (HNC) patients with a view to finding targets which could serve as therapeutic markers of the disease (Chen et al. 2007). They found DSG3 to be overexpressed in the HNC samples at the protein and mRNA levels, and knockdown of the protein significantly reduced both migration and invasion of HNC cell lines.

The mechanisms by which DSG3 may play a role in the motility of DLKP-I are unclear, however previous work may provide an insight the role played by the protein in these phenotypes. As discussed earlier, DSG3 was investigated in oral squamous- cell carcinoma cell lines (Brown et al. 2014). It was found that DSG3 forms a complex with Ezrin, another protein known to be involved in metastatic phenotypes in cancers such as osteosarcoma and hepatocellular carcinoma (Ren et al. 2009, Chen et al. 2011). This complex is necessary for the interaction between Ezrin and F-actin to induce cytoskeletal remodelling. The study by Brown et al. 2014 identified DSG3 is an upstream cell surface activator of the PKC/Ezrin pathway and showed it to be a regulating factor of cell motility and invasion in OCC. They demonstrated that DSG3/Ezrin complex forms at the plasma membrane and regulates Ezrin phosphorylation by PKC. The phosphorylation of Ezrin is a prerequisite for proper interaction with F-Actin and cytoskeletal remodelling leading to cell migration and invasion. Overexpression of DSG3 resulted in enhanced phosphorylation of Ezrin which resulted in increased filopodial protrusions, cell spreading and invasive phenotype. DSG3 was also found by this study to regulate the activity of AP-1, a known proto-oncogene which is show to be a regulator of cancer cell metastasis (Ozanne et al. 2007).

The reduction in invasion of DLKP-I cells upon knockdown of DSG3 may be supported by the findings of a previous work in which the effects of DSG3 knockdown in invasive HNCC cell lines were investigated (Chen et al. 2013). They found that silencing of DSG3 in these cell lines reduced the interaction between DSG3 and its known binding partner Plakoglobin (PG) (Hu et al. 2003), resulting in the translocation of PG from the plasma membrane to the nucleus of the cells. In the nucleus, they found

PG to bind to TCF/LEF transcription factors which suppress the transcriptional activity normally associated with this Wnt-signalling pathway (Hoppler and Kavanagh 2007). The effect of this disruption among others things was an inhibition of MMP7 and a concomitant reduction in the invasive ability of the HNCC cells.

Literature studies show that DSG3 is implicated in the metastatic phenotypes of a number of cancer cell lines, however our study is the first to show that knockdown of DSG3 results in decreased migration and invasion in a lung cancer cell line. It is unclear how DSG3 plays a role in these phenotypes, however the literature suggests a number of possibilities that require further investigation in DLKP-I. The interaction of DSG3 with Ezrin and the effects on F-Actin found in literature studies present a potential avenue of investigation, and it is possible that F-Actin plays an important role in cell motility in DLKP-I.

Co-staining of DSG3 and F-Actin was carried out on DLKP-I cells and imaged using confocal microscopy. A number of conditions were included in this experiment as described in Section 6.6.3 In short, both proteins were co-stained to assess if co-localisation was occurring in this cell line. In addition, the effect of DSG3 knockdown was investigated to determine if loss of the protein had an effect on F-Actin expression or cell morphology. F-actin has been found to be an interacting partner of DSG3 in other cancer cell lines such as ESCC (Brown et al. 2014), so this potential link was investigated in our lung cancer cell line using a co-staining method.

It was found that both DSG3 and F-Actin show strong staining in the DLKP-I cell line. DSG3 shows localisation to cell-cell junctions as well as punctate staining patterns throughout the cytoplasm, similar to what was found in the IF work discussed in Section 6.5.2. In addition, strong staining for DSG3 can be seen around the periphery of the DLKP-I cells. F-actin staining shows a localisation pattern which is similar to that found for DSG3. Strong accumulations of F-actin can be seen in cell-cell junctions also and around the cell peripheries. This may be cortical Actin as this protein has been shown to be enriched around the cell periphery where it interacts with the inner side of the plasma membrane (Cosen-Binker and Kapus 2006). In the control group, co-localisation of DSG3 and F-actin is visible. This is particularly noticeable in the cell-cell junctions and cell peripheries. In addition, where the cells show filopodial outgrowths, strong co-staining of both proteins can be seen. This concurs with the

findings of Brown et al. 2014 where DSG3, Ezrin and F-Actin were found to co-localise and promote cytoskeletal remodelling.

RNAi of DSG3 resulted in a visible reduction of staining for the protein in the knockdown samples compared to the negative siRNA control. This did not result in a change in the morphology of the DLKP-I cells; however it did seem to affect the intensity of F-Actin staining by noticeably reducing it. F-Actin staining is still present and it has not been reduced to the same degree as DSG3, but the reduction is strong. It is possible that the knockdown of DSG3 has interfered with actin polymerization, resulting in the decreased staining intensity of F-actin. Western blot analysis probing for Beta-actin (Figure 6.10) in DSG3 knockdown samples did not show a reduction in expression of the protein. However, F-actin staining highlights the protein in its filamentous form, and it is possible that the knockdown of DSG3 has affected actin polymerisation rather than the expression of the protein itself. This hypothesis contributes to explaining the reduction in cell motility which was also found in DLKP-I upon transient DSG3 knockdown by this study. Interfering with actin polymerisation may inhibit cytoskeletal dynamics and reduce the ability of the affected cells to migrate through ARF.

Thus far, this work has shown DSG3 to play a role in migration and invasion in the DLKP-I cell line. The involvement of DSG3 in these phenotypic characteristics is unclear; however co-staining studies have shown that there is a potential interaction between DSG3 and F-Actin, a protein known to be involved in cell motility. In order to investigate this link further, Co-Immunoprecipitation (Co-IP) of DSG3 was carried out using the method described in Section 2.8.1. It was hoped that this study would validate the potential interaction between DSG3 and F-actin, and also discover other binding partners of DSG3 in DLKP-I which could shed light on the role of the protein in this cell line.

This co-IP study was successful in extracting DSG3 protein from DLKP-I cell lysate. In addition, the technique reproducibly co-immunoprecipitated a small panel of proteins which were deemed as potential interacting partners of DSG3 in the DLKP-I cell line. F-Actin was not found to be a binding partner of DSG3 in this study. It was

pulled out along with DSG3 repeatedly; however it also was present in the negative control sample suggesting it non-specifically bound to the Protein-G bead-antibody complex. Of the panel of proteins which are potentially binding partners of DSG3, a subset were deemed the most interesting based on their known function in combination with what is known from the literature. These are:

- Junction Plakoglobin (PG).
- **Pyruvate dehydrogenase E1 component subunit beta, mitochondrial (PDHE1-A).**
- **Pyruvate dehydrogenase E1 component subunit beta, mitochondrial** (PDHE1-B)**.**
- Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex (DLAT).
- **Pyruvate dehydrogenase protein X component, mitochondrial** (PDHX).

Junction Plakoglobin is a member of the Armadillo family of proteins and has been previously identified as a component of the desmosome. This protein links the cytoskeleton to the plasma membrane and integrates actin, microtubules and keratin intermediate filaments (Delva, Tucker and Kowalczyk 2009). Junction plakoglobin is a known direct binding partner of DSG3 in cultured keratinocytes and has been extensively studied along with DSG3 in the *Pemphigus Vulgaris* disease (Aoyama et al. 2009). Knockdown of DSG3 has been shown to result in the translocation of this protein to the nucleus in HNCC cell lines resulting in decreased expression of MMP7 (Hoppler and Kavanagh 2007).

The presence of this protein in our co-IP of DSG3 has not been described before in a lung cancer cell line. However, the interaction of DSG3 and PG in general is not a novel discovery. This finding also represents a validation of the method for discovering binding partners of DSG3 in DLKP-I. PG was validated as expressed in the DLKP cell lines with the same expression pattern as was found for DSG3. This indicates that both these proteins are strongly linked in this cell line model and confirms their interacting dynamic.

Co-IP of DSG3 in DLKP-I resulted in a surprising high proportion of mitochondrial related proteins being found as potential binding partners of DSG3. Eight proteins were found as potential binding partners of DSG3, and four of these are mitochondrial associated proteins: PDHE1-A; PDHE1-B; DLAT; PDHX

These proteins are each a component of the Pyruvate Dehydrogenase Complex (PDC), which is a multi-enzyme complex located in the mitochondrial matrix. This complex is a key component in aerobic energy metabolism which functions to convert Pyruvate to Acetyl-CoA by Pyruvate decarboxylation. The resulting Acetyl-CoA may then be used for cellular respiration as part of the citric acid cycle (Sun et al. 2015). The primary enzymes which form this complex are PDHEA1-A; PDHE1-B; DLAT and Dihydrolipoyl dehydrogenase, mitochondria (DLD) which was not identified here.

Direct links between DSG3 and mitochondria are sparse in the literature. A number of studies in the 1980's found mitochondria attached to the cytoplasmic sides of desmosomes in a panel of ciliary epithelia and diseased liver biopsies (Freddo 1988, Rassat, Robenek and Themann 1981). More recently a study has found an association between the autoimmune disease *Pemphigus Vulgaris* (PV) and damage to mitochondria. Exposure of keratinocytes to the sera of patients suffering PV resulted in damage to mitochondria and release of anti-mitochondrial antibodies (AMAs). A protein microarray study of auto-antibodies in patients with PV identified a number of mitochondrial associated proteins targeted by these antibodies including PDHE1-A (Chen et al. 2015, Kalantari-Dehaghi et al. 2013). Mitochondrial damage is a symptom of PV and AMA screening is currently used as a diagnostic test for the disease. PDHE1-A/B and DLAT are the primary proteins (of three) which comprise the PDC and are expressed exclusively in the mitochondria.

Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial (**DLAT**) is another major component of the PDC and is therefore also localized to the mitochondrial matrix. Its function is to finalise the conversion of Pyruvate to Acetyl-CoA in conjunction with the other proteins of the PDC (PDHE1-A, DLD). In addition, DLAT is also the primary antigen targeted in the autoimmune disease known as Primary Biliary Cirrhosis (PBC) in which the small bile ducts of the liver are degraded to the point of rendering the liver seriously damaged. The primary cause of the disease is thought to be as a result of AMAs

targeting DLAT and causing an immune response, destroying the cells (Lindor et al. 2009). Disease diagnosis is performed by screening for the AMAs, and it has been found that >95% of patients with the disease test positive for antibodies against DLAT.

DSG3 appears to be a relatively poor prognosis biomarker as shown in the Kaplan-Meier plot in Figure 7.3. High expression of DSG3 can be seen to be associated with a poorer outcome for lung cancer patients. This finding is reflected in previous research carried out on DSG3 in other cancer types. A study comparing tissues from head and neck cancer (HNC) against normal epithelial tissue found that high expression of DSG3 correlated highly with T-stage and extracapsular spread of the cancer (Chen et al. 2007). From this they concluded that DSG3 takes part in carcinogenesis, and *in vitro* follow up work found that knockdown of DSG3 in HNC cell lines significantly inhibited their migratory and invasive abilities. A separate study investigating DSG3 expression in oesophageal squamous cell carcinoma (ESCC) found the protein to be significantly upregulated in tumour samples relative to normal epithelial tissue and concluded that it contributes to the aggressive features of ESCC (Fang et al. 2014). Studies of DSG3 in lung cancer are few, with only a handful investigating the role of DSG3 in the disease.

One group carried out an immunohistochemical analysis comparing 300 surgical non-small cell lung cancer and neuroendocrine tumours (Fukuoka et al. 2007). They found positive immunohistochemical staining of DSG3 to be correlated with favourable prognosis in non-small cell lung cancer and carcinoid tumours, and could be useful as a prognostic marker. This contrasts with the findings of the role of DSG3 in other cancer types where high expression of the protein was associated with poorer outcome. A second study of DSG3 in lung cancer did not provide survival data, but investigated DSG3 as a potential diagnostic marker (Savci-Heijink et al. 2009). They found that staining for DSG3 had a sensitivity of 99% and specificity of 87% for SQCC, and concluded that DSG3 may be useful as a diagnostic marker to distinguish SQCC from other subtypes of lung cancer when included in a panel of markers. Further studies are required in lung cancer to determine the role of DSG3 in the disease. If DSG3 proves to be a therapeutic target, it has a natural advantage as a transmembrane protein expressed on the cell surface which would make it potentially susceptible to antibody-

targeting therapies as is being currently explored in relation to its known role in Pemphigus Vulgaris (Koga et al. 2013).

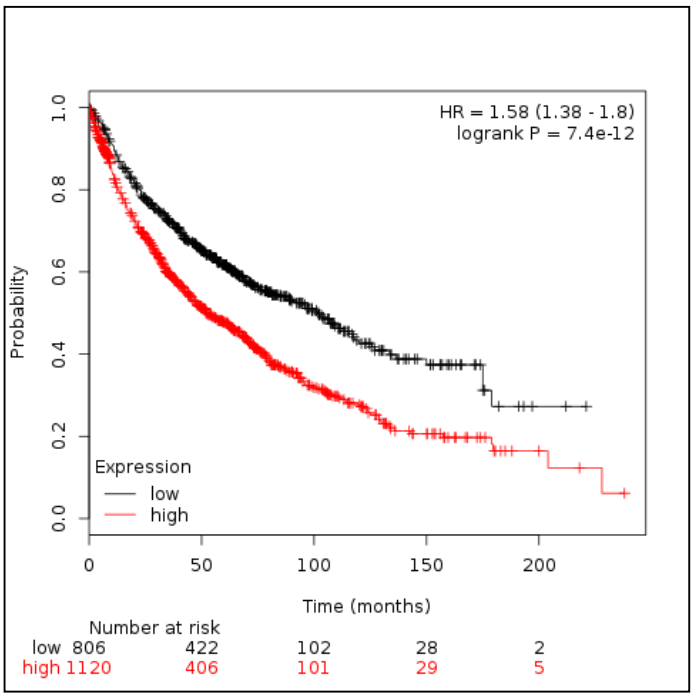


Figure 7.3: Kaplan Meier plot for Desmoglein-3 showing overall survival for lung cancer patients.

7.5. Conclusions:

Proteomic profiling using quantitative LC-MS/MS was carried out on a lung carcinoma cell line comprised of three distinct subpopulations with known phenotypic differences with regard to their migratory and invasive ability. Comparative proteomic analyses of the cell lines facilitated the discovery of highly abundant proteins enriched in each subpopulation. These markers confirm the phenotypic and proteomic variations between the clonal subpopulations which underlie the heterogeneous nature of this cell line. Conclusions drawn from this work are as follows:

1. LC-MS/MS comparative proteomic analyses were used to identify differentially expressed proteins between clonal subpopulations in a heterogeneous lung cancer cell line model. High stringency filtering criteria identified 127 and 203 significantly differentially expressed proteins during exponential and stationary phases of growth respectively. The analysis allowed for the identification of highly abundant proteins associated with each of the clones, providing candidates to potentially represent each one. The technology is an effective method for characterising the different proteomic profiles in a sample set.

2. DLKP and its clonal subpopulations are proteomically distinct cell lines, as found by the LC-MS/MS comparative analyses. Individual protein profiles were determined for each cell line using the quantitative proteomic analysis software. Myristoylated Alanine-Rich C-Kinase Substrate, Desmoglein-3 and Shootin-1 were each found to be differentially expressed between the three clonal subpopulations with highest abundance in DLKP-M, DLKP-I and DLKP-SQ respectively. Each protein was significantly more highly expressed in its associated clonal cell line. Desmoglein-3 in particular serves as a potential marker for DLKP-I as it is highly expressed in DLKP-I relative to DLKP-SQ and DLKP-M, with little or no expression in the latter two cell lines. The proteins representing each of the clonal subpopulations were taken forward for further investigation.

3. Shootin-1 was found to be highly expressed in DLKP-SQ. This protein is relatively uncharacterised and has little research carried out on it to date, with no functional studies carried out in cancer to date.
    - Shootin-1 shows a significantly higher protein expression in DLKP-SQ compared to DLKP and the other clonal subpopulations. Western Blot and qPCR analyses confirmed this finding at the protein and gene level.
    - Immunocytochemistry and Immunofluorescence staining of Shootin-1 supported the expression pattern found by LC-MS/MS. In addition, Shootin-1 localization was found to be high in cell projections, supporting the research done to date which characterizes the protein as involved with actin polymerization in dendritic outgrowths.
    - Co-immunoprecipitation studies found potential binding partners of Shootin-1 which to date have not yet been described. Of particular interest was the discovery of Semenogelin-1 (SEMG1) as a novel binding partner of Shootin-1. SEMG1 expression is very limited in the body, and its ectopic expression in lung cancer was an unusual finding. SMG proteins are secreted, and therefore may be useful as a biomarker for lung cancer, though further work would need to be carried out to support this hypothesis.

4. Myristoylated Alanine-Rich C-Kinase Substrate (MARCKS) was determined to be highly expressed in DLKP-M relative to the other clones:
    - MARCKS is abundantly expressed in the DLKP-M clone relative to all the other cell lines. Western Blot and qPCR analyses confirmed the expression pattern found by the LC-MS/MS analysis at the protein and mRNA level.

    - Immunofluorescence staining of MARCKS in each of the cell lines further validated the expression pattern found in the initial comparative proteomic study. In addition, this analysis showed the protein as localised to the cytoplasm with higher concentrations in structures projecting from the cells, such as filopodia and dendritic-like outgrowths.
    - Transient siRNA knockdown of MARCKS expression resulted in a significant reduction in the ability of DLKP-M cells to migrate and invade *in vitro*. DLKP-

M is the most invasive and migratory clonal subpopulation of DLKP, and these findings show that MARCKS plays an important role in this phenotype.

- Confocal imaging of MARCKS and F-actin was carried on DLKP-M cells to assess the effects of MARCKS knockdown on both proteins. The remaining MARCKS showed translocation from the cytoplasm to the nucleus of the cells. This suggests that MARCKS potentially becomes localized to the nucleus when the protein expression is reduced and may play a role in gene expression as a result; however further work is needed to confirm this. In addition, the intensity of F-Actin staining was decreased after MARCKS RNAi.
- Both Shootin-1 and MARCKS are known to interact with actin. Reduction in the expression of each target protein resulted in a significant reduction in cell migration in their respective clones. These findings may be as a result of their interaction with actin, with reduced expression of Shootin-1 and MARCKS causing reduced actin-mediated cell motility.

5. Desmoglein-3 (DSG3) was found to be abundantly expressed in DLKP-I only and can therefore be used as a marker for that cell line:
    - DSG3 shows significantly high expression in DLKP-I with no detectable levels of the protein in the other clonal subpopulations. Western Blot and qPCR analysis confirmed the high expression of DSG3 at the protein and mRNA level.
    - Immunocytochemistry and immunofluorescent staining of DSG3 in DLKP and the clonal subpopulations confirmed the high specificity of expression in DLKP-I only. These techniques confirmed the specific expression DSG3, especially in the heterogeneous DLKP cell line where stark differences in protein expression between the cells were observed.
    - Transient siRNA knockdown of DSG3 expression in DLKP-I results in a significant reduction in the ability of the cells to migrate and invade. This finding represents the first report of DSG3 knockdown reducing migration and invasion in a lung cancer cell line.
    - Co-immunoprecipitation of Desmoglein-3 in DLKP-I resulted in the discovery of a panel of potential binding partners which were identified using LC-MS/MS. One protein in the panel, Junction Plakoglobin, is a known binding partner of Desmoglein-3 and therefore helped to validate the method. A subpanel of

mitochondrial proteins were detected as potential binding partners of Desmoglein-3. This suggests DSG3 has additional roles relating to mitochondria, which so far remain little studied. Reciprocal Co-IP experiments are required to validate these potential binding partner interactions.

- Co-culture of DLKP-I with DLKP-M reduced expression levels of DSG3. This shows that in the heterogeneous population of DLKP, cell-cell interactions significantly alter the protein expression profiles of the clonal subpopulations.

7.6. Future Work:

1. MARCKS was found as a highly expressed protein in DLKP-M and has been shown to play a functional role in this cell line. There are a number of follow on studies which would need to be performed to clarify the role of this protein in the DLKP cell line model:

    - Functional analyses of MARCKS knockdown will need be carried out in both DLKP-SQ and DLKP-I. Though expression of the MARCKS is lower in these cell lines, it may play a similar role in cell migration and invasion. This could strengthen the findings of this study, and also potentially highlight new roles for MARCKS.
    - Confocal analysis revealed a potential translocation of MARCKS to the nucleus of DLKP-M upon transient knockdown of the protein using RNAi. This will require further investigation to validate the finding. A nuclear extraction will be performed on knockdown and control cells, and these will be compared by Western Blot analysis to determine if there is enrichment for the protein in knockdown samples. A protein segment within MARCKS (the effector domain) has been shown to have a homologous sequence to known nuclear localisation sequences. This could be investigated in DLKP-M to assess if it is responsible for nuclear localisation of MARCKS, and what specifically stimulated this response.
    - Co-Immunoprecipitation studies were unsuccessful for MARCKS due to a lack of good quality antibodies. In the future, this avenue of investigation will be reopened to look for MARCKS binding partners; for example alternative suitable antibodies for Co-IP will be sourced. This may shed light on its role in the DLKP cell line model, and elucidate the mechanism behind its functional roles in migration and invasion. In addition, Co-IP could be carried out on MARCKS after it has translocated to the nucleus upon knockdown, to determine what binding partners differ depending on its localisation within the cell.
    - Though an overexpression vector was successfully constructed, preliminary attempts at inducing a high expression of MARCKS in the DLKP clones were unsuccessful. This will be explored again and if overexpression is achieved, functional assays will be carried out to determine the effects of this process. It

would be hoped that overexpression of MARCKS in DLKP-SQ could induce a highly invasive phenotype for example.

- MARCKS has not been extensively studied in lung cancer. It would be interesting to carry out immunohistochemical staining on lung cancer tissue with varying invasive and migratory properties to assess if there is a correlation with MARCKS expression.

2. Desmoglein-3 was identified as a potential marker for DLKP-I with high specific expression in this cell line. This protein target requires additional investigation to clarify its role in the DLKP cell line model:

    - Further work will need to be carried out to determine how useful Desmoglein-3 is as a marker for DLKP-I. This protein could potentially be used to re-clone DLKP-I from the parent DLKP cell line using Fluorescently-Activated Cell Sorting (FACS), and compared to the original DLKP-I. Following on from this, Desmogelin-3 could then be used as a potential marker for inter-conversion events which these cells are hypothesised to do through DLKP-I (McBride et al. 1998a).
    - Co-Immunoprecipitation studies found Desmoglein-3 to be potentially associated with mitochondrial proteins. Reverse Co-IPs will need to be performed in order to validate these interactions. Mitochondrial extractions could also be performed from DLKP-I cells and analysed by Western Blot to look for the presence of Desmoglein-3, or desmosomal proteins in general.
    - Co-localisation of F-Actin and Desmoglein-3 was found in DLKP-I, and knockdown of Desmoglein-3 resulted in a potential reduction of F-Actin abundance. This link will need to be investigated further, as well as the potential effects of Desmogelin-3 Actin cytoskeletal dynamics.
    - Construction of an overexpression vector for Desmoglein-3 was unsuccessful. A synthesised version has been purchased, and this will be transfected into the DLKP clonal subpopulations to look for effects on the phenotypic characteristics of the cells. This is particularly interesting as expression of Desmoglein-3 is specific to DLKP-I, and overexpression may lead to inter-conversion events in DLKP-SQ or DLKP-M. In addition, if overexpression is successful, it may be

worth investigating if Co-Immunoprecipitation extracts mitochondrial proteins in the other DLKP clones also.

3. Shootin-1 was found to be highly expressed protein in DLKP-SQ relative to the other cell lines. It requires further work to explore the findings of this study.
    - Transient siRNA knockdown of Shootin-1 was found to reduce the migratory ability of DLKP-SQ. It would be worth carrying out the same process in the other DLKP clones to assess the functional role of Shootin-1 in these cell lines. DLKP-SQ is an anoikis resistant cell line, and the mechanism behind this phenotypic trait has not yet been established. Shootin-1, being relatively highly abundant in DLKP-SQ, may represent a possible role in this process. Preliminary attempts at Anoikis assays were unsuccessful, however once a reproducible method is established, this phenotypic trait should be explored with regard to Shootin-1.
    - Co-Immunoprecipitation studies found a number of potential Shootin-1 binding partners require validation by reverse Co-IP methods. Functional analyses of these targets, in particular Semenogelin-1, are required to elucidate the role of these proteins in a lung cancer cell line such as DLKP-SQ. They should also be explored in the other DLKP clones.
    - Our study found a potential interaction between Shootin-1 and F-Actin, which is also reported in the literature in neuronal cells. This link was not validated by Co-IP, however it would still be worth investigating if Shootin-1 is involved with cell motility in its role as a mediator in Actin retrograde flow in the DLKP cell line model.
    - Overexpression of Shootin-1 was achieved by the construction of a vector, and this was successfully transfected into the DLKP clones. Both DLKP-M and DLKP-I show overexpression of Shootin-1 relative to empty vector controls. In the future, functional analyses of the effects of this overexpression will be explored.

4. Knockdown of each protein target by RNAi induced phenotypic changes in the behaviour of their respective cell lines. It would be interesting to assess if similar effects are observed when each target is knocked out using a different method such as the CRISPR/Cas9 system. This would effectively knockout a target at the molecular level and the phenotypic changes associated with this could be assessed. Evaluation of the effects of target knockdown at the molecular level could be carried by functional analysis of the migratory and invasive abilities of the cells.

5. Co-culture of the DLKP clones resulted in changes in expression of target proteins. It would be interesting to determine if the targets explored in this study are secreted, and assess the effects of this on the clones which do not express the proteins at high levels. Studies could also be carried out to assess the effects of conditioned medium from one clone on the growth and phenotypic characteristics of another, and effects on targets of interest could be determined also.

6. The quantitative label-free data contains a wealth of protein expression information which has yet to be explored. This study looked at high abundance proteins from each DLKP clone, however there are many more targets which are yet be explored.

Chapter 8.

Bibliography:

Abedi-Ardekani, B., Kamangar, F., Sotoudeh, M., Villar, S., Islami, F., Aghcheli, K., Nasrollahzadeh, D., Taghavi, N., Dawsey, S.M., Abnet, C.C., Hewitt, S.M., Fahimi, S., Saidi, F., Brennan, P., Boffetta, P., Malekzadeh, R. and Hainaut, P. 2011. Extremely high Tp53 mutation load in esophageal squamous cell carcinoma in golestan province, iran. *PloS One,* 6(12), pp.e29488.

Aebersold, R. and Mann, M. 2003. Mass spectrometry-based proteomics. *Nature,* 422(6928), pp.198-207.

Aguayo, S.M., Schuyler, W.E., Murtagh, J.J.,Jr and Roman, J. 1994. Regulation of lung branching morphogenesis by bombesin-like peptides and neutral endopeptidase. *American Journal of Respiratory Cell and Molecular Biology,* 10(6), pp.635-642.

Aguda, A.H., Burtnick, L.D. and Robinson, R.C. 2005. The state of the filament. *EMBO Reports,* 6(3), pp.220-226.

Al Alam, D., El Agha, E., Sakurai, R., Kheirollahi, V., Moiseenko, A., Danopoulos, S., Shrestha, A., Schmoldt, C., Quantius, J., Herold, S., Chao, C.M., Tiozzo, C., De Langhe, S., Plikus, M.V., Thornton, M., Grubbs, B., Minoo, P., Rehan, V.K. and Bellusci, S. 2015. Evidence for the involvement of fibroblast growth factor 10 in lipofibroblast formation during embryonic lung development. *Development (Cambridge, England),* 142(23), pp.4139-4150.

Alban, A., David, S.O., Bjorkesten, L., Andersson, C., Sloge, E., Lewis, S. and Currie, I. 2003. A novel experimental design for comparative two-dimensional gel analysis: Two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics,* 3(1), pp.36-44.

Albert, K.A., Nairn, A.C. and Greengard, P. 1987. The 87-kDa protein, a major specific substrate for protein kinase C: Purification from bovine brain and characterization. *Proceedings of the National Academy of Sciences of the United States of America,* 84(20), pp.7046-7050.

Albini, A., Bruno, A., Gallo, C., Pajardi, G., Noonan, D.M. and Dallaglio, K. 2015. Cancer stem cells and the tumor microenvironment: Interplay in tumor heterogeneity. *Connective Tissue Research,* 56(5), pp.414-425.

Alexander, R.W., Upp, J.R.,Jr, Poston, G.J., Gupta, V., Townsend, C.M.,Jr and Thompson, J.C. 1988. Effects of bombesin on growth of human small cell lung carcinoma in vivo. *Cancer Research,* 48(6), pp.1439-1441.

Amagai, M., Koch, P.J., Nishikawa, T. and Stanley, J.R. 1996. Pemphigus vulgaris antigen (desmoglein 3) is localized in the lower epidermis, the site of blister formation in patients. *The Journal of Investigative Dermatology,* 106(2), pp.351-355.

Andl, C.D. and Stanley, J.R. 2001. Central role of the plakoglobin-binding domain for desmoglein 3 incorporation into desmosomes. *The Journal of Investigative Dermatology,* 117(5), pp.1068-1074.

Aoyama, Y., Yamamoto, Y., Yamaguchi, F. and Kitajima, Y. 2009. Low to high Ca2+ -switch causes phosphorylation and association of desmocollin 3 with plakoglobin and desmoglein 3 in cultured keratinocytes. *Experimental Dermatology,* 18(4), pp.404-408.

Arbuzova, A., Schmitz, A.A. and Vergeres, G. 2002. Cross-talk unfolded: MARCKS proteins. *The Biochemical Journal,* 362(Pt 1), pp.1-12.

Armirotti, A. and Damonte, G. 2010. Achievements and perspectives of top-down proteomics. *Proteomics,* 10(20), pp.3566-3576.

Artym, V.V., Zhang, Y., Seillier-Moiseiwitsch, F., Yamada, K.M. and Mueller, S.C. 2006. Dynamic interactions of cortactin and membrane type 1 matrix metalloproteinase at invadopodia: Defining the stages of invadopodia formation and function. *Cancer Research,* 66(6), pp.3034-3043.

Baldassarre, M., Pompeo, A., Beznoussenko, G., Castaldi, C., Cortellino, S., McNiven, M.A., Luini, A. and Buccione, R. 2003. Dynamin participates in focal extracellular matrix degradation by invasive cells. *Molecular Biology of the Cell,* 14(3), pp.1074-1084.

Bartoli, M., Monneron, A. and Ladant, D. 1998. Interaction of calmodulin with striatin, a WD-repeat protein present in neuronal dendritic spines. *The Journal of Biological Chemistry,* 273(35), pp.22248-22253.

Bartoli, M., Ternaux, J.P., Forni, C., Portalier, P., Salin, P., Amalric, M. and Monneron, A. 1999. Down-regulation of striatin, a neuronal calmodulin-binding protein, impairs rat locomotor activity. *Journal of Neurobiology,* 40(2), pp.234-243.

Ben, Q.W., Wang, J.C., Liu, J., Zhu, Y., Yuan, F., Yao, W.Y. and Yuan, Y.Z. 2010. Positive expression of L1-CAM is associated with perineural invasion and poor outcome in pancreatic ductal adenocarcinoma. *Annals of Surgical Oncology,* 17(8), pp.2213-2221.

Benoist, M., Gaillard, S. and Castets, F. 2006. The striatin family: A new signaling platform in dendritic spines. *Journal of Physiology, Paris,* 99(2-3), pp.146-153.

Bergh, J.C. 1990. Gene amplification in human lung cancer. the myc family genes and other proto-oncogenes and growth factor genes. *The American Review of Respiratory Disease,* 142(6 Pt 2), pp.S20-6.

Beynon, R.J. and Pratt, J.M. 2005. Metabolic labeling of proteins for proteomics. *Molecular & Cellular Proteomics : MCP,* 4(7), pp.857-872.

Bjartell, A., Malm, J., Moller, C., Gunnarsson, M., Lundwall, A. and H, L. 1996. Distribution and tissue expression of semenogelin I and II in man as demonstrated by in situ hybridization and immunocytochemistry. *Journal of Andrology,* 17(1), pp.17-26.

Bowden, E.T., Barth, M., Thomas, D., Glazer, R.I. and Mueller, S.C. 1999. An invasion-related complex of cortactin, paxillin and PKCmu associates with invadopodia at sites of extracellular matrix degradation. *Oncogene,* 18(31), pp.4440-4449.

Brown, L., Waseem, A., Cruz, I.N., Szary, J., Gunic, E., Mannan, T., Unadkat, M., Yang, M., Valderrama, F., O Toole, E.A. and Wan, H. 2014. Desmoglein 3 promotes cancer cell migration and invasion by regulating activator protein 1 and protein kinase C-dependent-ezrin activation. *Oncogene,* 33(18), pp.2363-2374.

Bunn, P.A.,Jr and Franklin, W. 2002. Epidermal growth factor receptor expression, signal pathway, and inhibitors in non-small cell lung cancer. *Seminars in Oncology,* 29(5 Suppl 14), pp.38-44.

Cai, J.H., Zhao, R., Zhu, J.W., Jin, X.L., Wan, F.J., Liu, K., Ji, X.P., Zhu, Y.B. and Zhu, Z.G. 2010. Expression of cortactin correlates with a poor prognosis in patients with stages II-III colorectal adenocarcinoma. *Journal of Gastrointestinal Surgery : Official Journal of the Society for Surgery of the Alimentary Tract,* 14(8), pp.1248-1257.

Calabrese, B. and Halpain, S. 2005. Essential role for the PKC target MARCKS in maintaining dendritic spine morphology. *Neuron,* 48(1), pp.77-90.

Campos-Parra, A.D., Zuloaga, C., Manriquez, M.E., Aviles, A., Borbolla-Escoboza, J., Cardona, A., Meneses, A. and Arrieta, O. 2015. KRAS mutation as the biomarker of response to chemotherapy and EGFR-TKIs in patients with advanced non-small cell lung cancer: Clues for its potential use in second-line therapy decision making. *American Journal of Clinical Oncology,* 38(1), pp.33-40.

Canas, B., Lopez-Ferrer, D., Ramos-Fernandez, A., Camafeita, E. and Calvo, E. 2006. Mass spectrometry technologies for proteomics. *Briefings in Functional Genomics & Proteomics,* 4(4), pp.295-320.

Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C. and Stuart, J.M. 2013. The cancer genome atlas pan-cancer analysis project. *Nature Genetics,* 45(10), pp.1113-1120.

Carney, D.N., Gazdar, A.F., Bunn, P.A.,Jr and Guccion, J.G. 1982. Demonstration of the stem cell nature of clonogenic tumor cells from lung cancer patients. *Stem Cells,* 1(3), pp.149-164.

Castets, F., Bartoli, M., Barnier, J.V., Baillat, G., Salin, P., Moqrich, A., Bourgeois, J.P., Denizot, F., Rougon, G., Calothy, G. and Monneron, A. 1996. A novel calmodulin-binding protein, belonging to the WD-repeat family, is localized in dendrites of a subset of CNS neurons. *The Journal of Cell Biology,* 134(4), pp.1051-1062.

Chambers, A.F., Groom, A.C. and MacDonald, I.C. 2002. Dissemination and growth of cancer cells in metastatic sites. *Nature Reviews.Cancer,* 2(8), pp.563-572.

Chansky, K., Sculier, J.P., Crowley, J.J., Giroux, D., Van Meerbeeck, J., Goldstraw, P. and International Staging Committee and Participating Institutions. 2009. The international association for the study of lung cancer staging project: Prognostic factors and pathologic TNM stage in surgically managed non-small cell lung cancer. *Journal of Thoracic Oncology : Official Publication of the International Association for the Study of Lung Cancer,* 4(7), pp.792-801.

Chen, C.H., Statt, S., Chiu, C.L., Thai, P., Arif, M., Adler, K.B. and Wu, R. 2014a. Targeting myristoylated alanine-rich C kinase substrate phosphorylation site domain in lung cancer. mechanisms and therapeutic implications. *American Journal of Respiratory and Critical Care Medicine,* 190(10), pp.1127-1138.

Chen, C.H., Thai, P., Yoneda, K., Adler, K.B., Yang, P.C. and Wu, R. 2014b. A peptide that inhibits function of myristoylated alanine-rich C kinase substrate (MARCKS) reduces lung cancer metastasis. *Oncogene,* 33(28), pp.3696-3706.

Chen, G., Gharib, T.G., Huang, C.C., Taylor, J.M., Misek, D.E., Kardia, S.L., Giordano, T.J., Iannettoni, M.D., Orringer, M.B., Hanash, S.M. and Beer, D.G. 2002. Discordant protein and mRNA expression in lung adenocarcinomas. *Molecular & Cellular Proteomics : MCP,* 1(4), pp.304-313.

Chen, X. and Rotenberg, S.A. 2010. PhosphoMARCKS drives motility of mouse melanoma cells. *Cellular Signalling,* 22(7), pp.1097-1103.

Chen, Y., Chernyavsky, A., Webber, R.J., Grando, S.A. and Wang, P.H. 2015. Critical role of the neonatal fc receptor (FcRn) in the pathogenic action of antimitochondrial autoantibodies synergizing with anti-desmoglein autoantibodies in pemphigus vulgaris. *The Journal of Biological Chemistry,* 290(39), pp.23826-23837.

Chen, Y., Wang, D., Guo, Z., Zhao, J., Wu, B., Deng, H., Zhou, T., Xiang, H., Gao, F., Yu, X., Liao, J., Ward, T., Xia, P., Emenari, C., Ding, X., Thompson, W., Ma, K., Zhu, J., Aikhionbare, F., Dou, K., Cheng, S.Y. and Yao, X. 2011. Rho kinase phosphorylation promotes ezrin-mediated metastasis in hepatocellular carcinoma. *Cancer Research,* 71(5), pp.1721-1729.

Chen, Y.J., Chang, J.T., Lee, L., Wang, H.M., Liao, C.T., Chiu, C.C., Chen, P.J. and Cheng, A.J. 2007. DSG3 is overexpressed in head neck cancer and is a potential molecular target for inhibition of oncogenesis. *Oncogene,* 26(3), pp.467-476.

Chen, Y.J., Lee, L.Y., Chao, Y.K., Chang, J.T., Lu, Y.C., Li, H.F., Chiu, C.C., Li, Y.C., Li, Y.L., Chiou, J.F. and Cheng, A.J. 2013. DSG3 facilitates cancer cell growth and invasion through the DSG3-plakoglobin-TCF/LEF-myc/cyclin D1/MMP signaling pathway. *PloS One,* 8(5), pp.e64088.

Chidgey, M. and Dawson, C. 2007. Desmosomes: A role in cancer? *British Journal of Cancer,* 96(12), pp.1783-1787.

Chitaev, N.A., Leube, R.E., Troyanovsky, R.B., Eshkind, L.G., Franke, W.W. and Troyanovsky, S.M. 1996. The binding of plakoglobin to desmosomal cadherins: Patterns of binding sites and topogenic potential. *The Journal of Cell Biology,* 133(2), pp.359-369.

Choe, L., D'Ascenzo, M., Relkin, N.R., Pappin, D., Ross, P., Williamson, B., Guertin, S., Pribil, P. and Lee, K.H. 2007. 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for alzheimer's disease. *Proteomics,* 7(20), pp.3651-3660.

Choi, H. and Mazzone, P. 2014. Radon and lung cancer: Assessing and mitigating the risk. *Cleveland Clinic Journal of Medicine,* 81(9), pp.567-575.

Christin, C., Bischoff, R. and Horvatovich, P. 2011. Data processing pipelines for comprehensive profiling of proteomics samples by label-free LC-MS for biomarker discovery. *Talanta,* 83(4), pp.1209-1224.

Clark, E.S. and Weaver, A.M. 2008. A new role for cortactin in invadopodia: Regulation of protease secretion. *European Journal of Cell Biology,* 87(8-9), pp.581-590.

Clark, E.S., Whigham, A.S., Yarbrough, W.G. and Weaver, A.M. 2007. Cortactin is an essential regulator of matrix metalloproteinase secretion and extracellular matrix degradation in invadopodia. *Cancer Research,* 67(9), pp.4227-4235.

Cosen-Binker, L.I. and Kapus, A. 2006. Cortactin: The gray eminence of the cytoskeleton. *Physiology (Bethesda, Md.),* 21pp.352-361.

Cottrell, J.S. 2011. Protein identification using MS/MS data. *Journal of Proteomics,* 74(10), pp.1842-1851.

Dang, C.V. 2012. MYC on the path to cancer. *Cell,* 149(1), pp.22-35.

Dehner, C., Rotzer, V., Waschke, J. and Spindler, V. 2014. A desmoplakin point mutation with enhanced keratin association ameliorates pemphigus vulgaris autoantibody-mediated loss of cell cohesion. *The American Journal of Pathology,* 184(9), pp.2528-2536.

Delva, E., Tucker, D.K. and Kowalczyk, A.P. 2009. The desmosome. *Cold Spring Harbor Perspectives in Biology,* 1(2), pp.a002543.

Desai, T.J., Brownfield, D.G. and Krasnow, M.A. 2014. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature,* 507(7491), pp.190-194.

Duru, N., Fan, M., Candas, D., Menaa, C., Liu, H.C., Nantajit, D., Wen, Y., Xiao, K., Eldridge, A., Chromy, B.A., Li, S., Spitz, D.R., Lam, K.S., Wicha, M.S. and Li, J.J. 2012. HER2-associated radioresistance of breast cancer stem cells isolated from HER2-negative breast cancer cells. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research,* 18(24), pp.6634-6647.

Dusek, R.L. and Attardi, L.D. 2011. Desmosomes: New perpetrators in tumour suppression. *Nature Reviews.Cancer,* 11(5), pp.317-323.

Engelsberger, W.R., Erban, A., Kopka, J. and Schulze, W.X. 2006. Metabolic labeling of plant cell cultures with K(15)NO3 as a tool for quantitative analysis of proteins and metabolites. *Plant Methods,* 2pp.14.

Evans, M.J. and Plopper, C.G. 1988. The role of basal cells in adhesion of columnar epithelium to airway basement membrane. *The American Review of Respiratory Disease,* 138(2), pp.481-483.

Fang, W.K., Chen, B., Xu, X.E., Liao, L.D., Wu, Z.Y., Wu, J.Y., Shen, J., Xu, L.Y. and Li, E.M. 2014. Altered expression and localization of desmoglein 3 in esophageal squamous cell carcinoma. *Acta Histochemica,* 116(5), pp.803-809.

Fang, W.K., Gu, W., Li, E.M., Wu, Z.Y., Shen, Z.Y., Shen, J.H., Wu, J.Y., Pan, F., Lv, Z., Xu, X.E., Huang, Q. and Xu, L.Y. 2010. Reduced membranous and ectopic cytoplasmic expression of DSC2 in esophageal squamous cell carcinoma: An independent prognostic factor. *Human Pathology,* 41(10), pp.1456-1465.

Fang, W., Chen, B., Xu, X., Liao, L., Wu, Z., Wu, J., Shen, J., Xu, L. and Li, E. Altered expression and localization of desmoglein 3 in esophageal squamous cell carcinoma. *Acta Histochemica,* (0),

Fernandez, F.G. and Battafarano, R.J. 2006. Large-cell neuroendocrine carcinoma of the lung. *Cancer Control : Journal of the Moffitt Cancer Center,* 13(4), pp.270-275.

Fogel, M., Gutwein, P., Mechtersheimer, S., Riedle, S., Stoeck, A., Smirnov, A., Edler, L., Ben-Arie, A., Huszar, M. and Altevogt, P. 2003. L1 expression as a predictor of progression and

survival in patients with uterine and ovarian carcinomas. *Lancet (London, England),* 362(9387), pp.869-875.

Freddo, T.F. 1988. Mitochondria attached to desmosomes in the ciliary epithelia of human, monkey, and rabbit eyes. *Cell and Tissue Research,* 251(3), pp.671-675.

Freed-Pastor, W.A. and Prives, C. 2012. Mutant p53: One name, many proteins. *Genes & Development,* 26(12), pp.1268-1286.

Frohm, B., DeNizio, J.E., Lee, D.S., Gentile, L., Olsson, U., Malm, J., Akerfeldt, K.S. and Linse, S. 2015. A peptide from human semenogelin I self-assembles into a pH-responsive hydrogel. *Soft Matter,* 11(2), pp.414-421.

Fukuoka, J., Dracheva, T., Shih, J.H., Hewitt, S.M., Fujii, T., Kishor, A., Mann, F., Shilo, K., Franks, T.J., Travis, W.D. and Jen, J. 2007. Desmoglein 3 as a prognostic factor in lung cancer. *Human Pathology,* 38(2), pp.276-283.

Fuller, H.R., Man, N.T., Lam le, T., Shamanin, V.A., Androphy, E.J. and Morris, G.E. 2010. Valproate and bone loss: iTRAQ proteomics show that valproate reduces collagens and osteonectin in SMA cells. *Journal of Proteome Research,* 9(8), pp.4228-4233.

Gainor, J.F. and Shaw, A.T. 2013. Emerging paradigms in the development of resistance to tyrosine kinase inhibitors in lung cancer. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology,* 31(31), pp.3987-3996.

Gardner-Thorpe, J., Ito, H., Ashley, S.W. and Whang, E.E. 2002. Differential display of expressed genes in pancreatic cancer cells. *Biochemical and Biophysical Research Communications,* 293(1), pp.391-395.

Garrod, D. and Chidgey, M. 2008. Desmosome structure, composition and function. *Biochimica Et Biophysica Acta,* 1778(3), pp.572-587.

Gaspari, M. and Cuda, G. 2011. Nano LC-MS/MS: A robust setup for proteomic analysis. *Methods in Molecular Biology (Clifton, N.J.),* 790pp.115-126.

Gavert, N., Conacci-Sorrell, M., Gast, D., Schneider, A., Altevogt, P., Brabletz, T. and Ben-Ze'ev, A. 2005. L1, a novel target of beta-catenin signaling, transforms cells and is expressed at the invasive front of colon cancers. *The Journal of Cell Biology,* 168(4), pp.633-642.

Gerlinger, M., Horswell, S., Larkin, J., Rowan, A.J., Salm, M.P., Varela, I., Fisher, R., McGranahan, N., Matthews, N., Santos, C.R., Martinez, P., Phillimore, B., Begum, S., Rabinowitz, A., Spencer-Dene, B., Gulati, S., Bates, P.A., Stamp, G., Pickering, L., Gore, M., Nicol, D.L., Hazell, S., Futreal, P.A., Stewart, A. and Swanton, C. 2014. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature Genetics,* 46(3), pp.225-233.

Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S., McDonald, N.Q., Butler, A., Jones, D., Raine, K., Latimer, C., Santos, C.R., Nohadani, M., Eklund, A.C., Spencer-Dene, B., Clark, G., Pickering, L., Stamp, G., Gore, M., Szallasi, Z., Downward, J., Futreal, P.A. and Swanton, C. 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England Journal of Medicine,* 366(10), pp.883-892.

Gerlinger, M. and Swanton, C. 2010. How darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *British Journal of Cancer,* 103(8), pp.1139-1143.

Gill, J.H., Kirwan, I.G., Seargent, J.M., Martin, S.W., Tijani, S., Anikin, V.A., Mearns, A.J., Bibby, M.C., Anthoney, A. and Loadman, P.M. 2004. MMP-10 is overexpressed, proteolytically active, and a potential target for therapeutic intervention in human lung carcinomas. *Neoplasia (New York, N.Y.),* 6(6), pp.777-785.

Gliem, M., Heupel, W.M., Spindler, V., Harms, G.S. and Waschke, J. 2010. Actin reorganization contributes to loss of cell adhesion in pemphigus vulgaris. *American Journal of Physiology.Cell Physiology,* 299(3), pp.C606-13.

Gomez-Morales, M., Camara-Pulido, M., Miranda-Leon, M.T., Sanchez-Palencia, A., Boyero, L., Gomez-Capilla, J.A. and Farez-Vidal, M.E. 2013. Differential immunohistochemical localization of desmosomal plaque-related proteins in non-small-cell lung cancer. *Histopathology,* 63(1), pp.103-113.

Govindan, R., Ding, L., Griffith, M., Subramanian, J., Dees, N.D., Kanchi, K.L., Maher, C.A., Fulton, R., Fulton, L., Wallis, J., Chen, K., Walker, J., McDonald, S., Bose, R., Ornitz, D., Xiong, D., You, M., Dooling, D.J., Watson, M., Mardis, E.R. and Wilson, R.K. 2012. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell,* 150(6), pp.1121-1134.

Green, T.D., Park, J., Yin, Q., Fang, S., Crews, A.L., Jones, S.L. and Adler, K.B. 2012. Directed migration of mouse macrophages in vitro involves myristoylated alanine-rich C-kinase substrate (MARCKS) protein. *Journal of Leukocyte Biology,* 92(3), pp.633-639.

Gruhler, S. and Kratchmarova, I. 2008. Stable isotope labeling by amino acids in cell culture (SILAC). *Methods in Molecular Biology (Clifton, N.J.),* 424pp.101-111.

Gu, X., Karp, P.H., Brody, S.L., Pierce, R.A., Welsh, M.J., Holtzman, M.J. and Ben-Shahar, Y. 2014. Chemosensory functions for pulmonary neuroendocrine cells. *American Journal of Respiratory Cell and Molecular Biology,* 50(3), pp.637-646.

Gulcicek, E.E., Colangelo, C.M., McMurray, W., Stone, K., Williams, K., Wu, T., Zhao, H., Spratt, H., Kurosky, A. and Wu, B. 2005. Proteomics and the analysis of proteomic data: An overview of current protein-profiling technologies. *Current Protocols in Bioinformatics / Editoral Board, Andreas D.Baxevanis ...[Et Al.],* Chapter 13pp.Unit 13.1.

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

Haddock, B.J., Zhu, Y., Doyle, S.P., Abdullah, L.H. and Davis, C.W. 2014. Role of MARCKS in regulated secretion from mast cells and airway goblet cells. *American Journal of Physiology.Lung Cellular and Molecular Physiology,* 306(10), pp.L925-36.

Hage, R., Elbers, J.R., Brutel de la Riviere, A. and van den Bosch, J.M. 1998. Surgery for combined type small cell lung carcinoma. *Thorax,* 53(6), pp.450-453.

Hajj, R., Baranek, T., Le Naour, R., Lesimple, P., Puchelle, E. and Coraux, C. 2007. Basal cells of the human adult airway surface epithelium retain transit-amplifying cell properties. *Stem Cells (Dayton, Ohio),* 25(1), pp.139-148.

Hanada, S., Kakehashi, A., Nishiyama, N., Wei, M., Yamano, S., Chung, K., Komatsu, H., Inoue, H., Suehiro, S. and Wanibuchi, H. 2013. Myristoylated alanine-rich C-kinase substrate as a prognostic biomarker in human primary lung squamous cell carcinoma. *Cancer Biomarkers : Section A of Disease Markers,* 13(4), pp.289-298.

Hanahan, D. and Weinberg, R.A. 2011. Hallmarks of cancer: The next generation. *Cell,* 144(5), pp.646-674.

Hartwig, J.H., Thelen, M., Rosen, A., Janmey, P.A., Nairn, A.C. and Aderem, A. 1992. MARCKS is an actin filament crosslinking protein regulated by protein kinase C and calcium-calmodulin. *Nature,* 356(6370), pp.618-622.

Higginbotham, K.S., Breyer, J.P., McReynolds, K.M., Bradley, K.M., Schuyler, P.A., Plummer, W.D., Freudenthal, M.E., Trentham-Dietz, A., Newcomb, P.A., Parl, F.F., Sanders, M.E., Page, D.L., Egan, K.M., Dupont, W.D. and Smith, J.R. 2012. A multistage genetic association study identifies breast cancer risk loci at 10q25 and 16q24. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology,* 21(9), pp.1565-1573.

Hiley, C., de Bruin, E.C., McGranahan, N. and Swanton, C. 2014. Deciphering intratumor heterogeneity and temporal acquisition of driver events to refine precision medicine. *Genome Biology,* 15(8), pp.453-014-0453-8.

Hollstein, M., Sidransky, D., Vogelstein, B. and Harris, C.C. 1991. P53 mutations in human cancers. *Science (New York, N.Y.),* 253(5015), pp.49-53.

Hoppler, S. and Kavanagh, C.L. 2007. Wnt signalling: Variety at the core. *Journal of Cell Science,* 120(Pt 3), pp.385-393.

Hortsch, M. 2000. Structural and functional evolution of the L1 family: Are four adhesion molecules better than one? *Molecular and Cellular Neurosciences,* 15(1), pp.1-10.

Hu, P., Berkowitz, P., O'Keefe, E.J. and Rubenstein, D.S. 2003. Keratinocyte adherens junctions initiate nuclear signaling by translocation of plakoglobin from the membrane to the nucleus. *The Journal of Investigative Dermatology,* 121(2), pp.242-251.

Huang, C.C., Lee, T.J., Chang, P.H., Lee, Y.S., Chuang, C.C., Jhang, Y.J., Chen, Y.W., Chen, C.W. and Tsai, C.N. 2010. Desmoglein 3 is overexpressed in inverted papilloma and squamous cell carcinoma of sinonasal cavity. *The Laryngoscope,* 120(1), pp.26-29.

Humphries, J.D., Byron, A. and Humphries, M.J. 2006. Integrin ligands at a glance. *Journal of Cell Science,* 119(Pt 19), pp.3901-3903.

Hynes, R.O. 2002. Integrins: Bidirectional, allosteric signaling machines. *Cell,* 110(6), pp.673-687.

Ihde, D.C. 1984. Current status of therapy for small cell carcinoma of the lung. *Cancer,* 54(11 Suppl), pp.2722-2728.

Illman, S.A., Lehti, K., Keski-Oja, J. and Lohi, J. 2006. Epilysin (MMP-28) induces TGF-beta mediated epithelial to mesenchymal transition in lung carcinoma cells. *Journal of Cell Science,* 119(Pt 18), pp.3856-3865.

Jackman, D.M., Miller, V.A., Cioffredi, L.A., Yeap, B.Y., Janne, P.A., Riely, G.J., Ruiz, M.G., Giaccone, G., Sequist, L.V. and Johnson, B.E. 2009. Impact of epidermal growth factor receptor and KRAS mutations on clinical outcomes in previously untreated non-small cell lung cancer patients: Results of an online tumor registry of clinical trials. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research,* 15(16), pp.5267-5273.

Jarboe, J.S., Anderson, J.C., Duarte, C.W., Mehta, T., Nowsheen, S., Hicks, P.H., Whitley, A.C., Rohrbach, T.D., McCubrey, R.O., Chiu, S., Burleson, T.M., Bonner, J.A., Gillespie, G.Y., Yang, E.S. and Willey, C.D. 2012. MARCKS regulates growth and radiation sensitivity and is a novel prognostic factor for glioma. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research,* 18(11), pp.3030-3041.

Jarmalaite, S., Kannio, A., Anttila, S., Lazutka, J.R. and Husgafvel-Pursiainen, K. 2003. Aberrant p16 promoter methylation in smokers and former smokers with nonsmall cell lung cancer. *International Journal of Cancer.Journal International Du Cancer,* 106(6), pp.913-918.

Jemal, A., Siegel, R., Xu, J. and Ward, E. 2010. Cancer statistics, 2010. *CA: A Cancer Journal for Clinicians,* 60(5), pp.277-300.

Jennings, J.M., Tucker, D.K., Kottke, M.D., Saito, M., Delva, E., Hanakawa, Y., Amagai, M. and Kowalczyk, A.P. 2011. Desmosome disassembly in response to pemphigus vulgaris IgG occurs in distinct phases and can be reversed by expression of exogenous Dsg3. *The Journal of Investigative Dermatology,* 131(3), pp.706-718.

Ji, L., Lim, J. and Danuser, G. 2008. Fluctuations of intracellular forces during cell protrusion. *Nature Cell Biology,* 10(12), pp.1393-1400.

Jonsdottir, K., Zhang, H., Jhagroe, D., Skaland, I., Slewa, A., Bjorkblom, B., Coffey, E.T., Gudlaugsson, E., Smaaland, R., Janssen, E.A. and Baak, J.P. 2012. The prognostic value of MARCKS-like 1 in lymph node-negative breast cancer. *Breast Cancer Research and Treatment,* 135(2), pp.381-390.

Kalantari-Dehaghi, M., Chen, Y., Deng, W., Chernyavsky, A., Marchenko, S., Wang, P.H. and Grando, S.A. 2013. Mechanisms of mitochondrial damage in keratinocytes by pemphigus vulgaris antibodies. *The Journal of Biological Chemistry,* 288(23), pp.16916-16925.

Kalluri, R. and Weinberg, R.A. 2009. The basics of epithelial-mesenchymal transition. *The Journal of Clinical Investigation,* 119(6), pp.1420-1428.

Kang, Y., Lu, S., Ren, P., Huo, B. and Long, M. 2012. Molecular dynamics simulation of shear- and stretch-induced dissociation of P-selectin/PSGL-1 complex. *Biophysical Journal,* 102(1), pp.112-120.

Kannagi, R., Izawa, M., Koike, T., Miyazaki, K. and Kimura, N. 2004. Carbohydrate-mediated cell adhesion in cancer metastasis and angiogenesis. *Cancer Science,* 95(5), pp.377-384.

Kansas, G.S. 1996. Selectins and their ligands: Current concepts and controversies. *Blood,* 88(9), pp.3259-3287.

Keenan, J., Joyce, H., Aherne, S., O'Dea, S., Doolan, P., Lynch, V. and Clynes, M. 2012. Olfactomedin III expression contributes to anoikis-resistance in clonal variants of a human lung squamous carcinoma cell line. *Experimental Cell Research,* 318(5), pp.593-602.

Kelley, M.J., Nakagawa, K., Steinberg, S.M., Mulshine, J.L., Kamb, A. and Johnson, B.E. 1995. Differential inactivation of CDKN2 and rb protein in non-small-cell and small-cell lung cancer cell lines. *Journal of the National Cancer Institute,* 87(10), pp.756-761.

Kenfield, S.A., Wei, E.K., Stampfer, M.J., Rosner, B.A. and Colditz, G.A. 2008. Comparison of aspects of smoking among the four histological types of lung cancer. *Tobacco Control,* 17(3), pp.198-204.

Kim, C.F., Jackson, E.L., Woolfenden, A.E., Lawrence, S., Babar, I., Vogel, S., Crowley, D., Bronson, R.T. and Jacks, T. 2005. Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell,* 121(6), pp.823-835.

Kim, Y.N., Choi, J.E., Bae, J.S., Jang, K.Y., Chung, M.J., Moon, W.S., Kang, M.J., Lee, D.G. and Park, H.S. 2012. Expression of cortactin and focal adhesion kinase in colorectal adenocarcinoma: Correlation with clinicopathologic parameters and their prognostic implication. *Korean Journal of Pathology,* 46(5), pp.454-462.

Kinzler, K.W. and Vogelstein, B. 1997. Cancer-susceptibility genes. gatekeepers and caretakers. *Nature,* 386(6627), pp.761, 763.

Kleene, R., Yang, H., Kutsche, M. and Schachner, M. 2001. The neural recognition molecule L1 is a sialic acid-binding lectin for CD24, which induces promotion and inhibition of neurite outgrowth. *The Journal of Biological Chemistry,* 276(24), pp.21656-21663.

Knoll, B. and Drescher, U. 2004. Src family kinases are involved in EphA receptor-mediated retinal axon guidance. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience,* 24(28), pp.6248-6257.

Knudson, A.G.,Jr. 1971. Mutation and cancer: Statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America,* 68(4), pp.820-823.

Koga, H., Tsuruta, D., Ohyama, B., Ishii, N., Hamada, T., Ohata, C., Furumura, M. and Hashimoto, T. 2013. Desmoglein 3, its pathogenecity and a possibility for therapeutic target in pemphigus vulgaris. *Expert Opinion on Therapeutic Targets,* 17(3), pp.293-306.

Kong, D., Li, Y., Wang, Z. and Sarkar, F.H. 2011. Cancer stem cells and epithelial-to-mesenchymal transition (EMT)-phenotypic cells: Are they cousins or twins? *Cancers,* 3(1), pp.716-729.

Kubo, Y., Baba, K., Toriyama, M., Minegishi, T., Sugiura, T., Kozawa, S., Ikeda, K. and Inagaki, N. 2015. Shootin1-cortactin interaction mediates signal-force transduction for axon outgrowth. *The Journal of Cell Biology,* 210(4), pp.663-676.

Larsson, C. 2006. Protein kinase C and the regulation of the actin cytoskeleton. *Cellular Signalling,* 18(3), pp.276-284.

Law, E., Gilvarry, U., Lynch, V., Gregory, B., Grant, G. and Clynes, M. 1992. Cytogenetic comparison of two poorly differentiated human lung squamous cell carcinoma lines. *Cancer Genetics and Cytogenetics,* 59(2), pp.111-118.

Lee, Y.Y., Yu, C.P., Lin, C.K., Nieh, S., Hsu, K.F., Chiang, H. and Jin, J.S. 2009. Expression of survivin and cortactin in colorectal adenocarcinoma: Association with clinicopathological parameters. *Disease Markers,* 26(1), pp.9-18.

Lemjabbar-Alaoui, H., Dasari, V., Sidhu, S.S., Mengistab, A., Finkbeiner, W., Gallup, M. and Basbaum, C. 2006. Wnt and hedgehog are critical mediators of cigarette smoke-induced lung cancer. *PloS One,* 1pp.e93.

Li, H., Chen, G., Zhou, B. and Duan, S. 2008a. Actin filament assembly by myristoylated alanine-rich C kinase substrate-phosphatidylinositol-4,5-diphosphate signaling is critical for dendrite branching. *Molecular Biology of the Cell,* 19(11), pp.4804-4813.

Li, H., Chen, G., Zhou, B. and Duan, S. 2008b. Actin filament assembly by myristoylated alanine-rich C kinase substrate-phosphatidylinositol-4,5-diphosphate signaling is critical for dendrite branching. *Molecular Biology of the Cell,* 19(11), pp.4804-4813.

Lindor, K.D., Gershwin, M.E., Poupon, R., Kaplan, M., Bergasa, N.V., Heathcote, E.J. and American Association for Study of Liver Diseases. 2009. Primary biliary cirrhosis. *Hepatology (Baltimore, Md.),* 50(1), pp.291-308.

Linnoila, R.I. 2006. Functional facets of the pulmonary neuroendocrine system. *Laboratory Investigation; a Journal of Technical Methods and Pathology,* 86(5), pp.425-444.

Luo, M.L., Shen, X.M., Zhang, Y., Wei, F., Xu, X., Cai, Y., Zhang, X., Sun, Y.T., Zhan, Q.M., Wu, M. and Wang, M.R. 2006. Amplification and overexpression of CTTN (EMS1) contribute to the metastasis of esophageal squamous cell carcinoma by promoting cell migration and anoikis resistance. *Cancer Research,* 66(24), pp.11690-11699.

MacCoss, M.J., Wu, C.C. and Yates, J.R.,3rd. 2002. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Analytical Chemistry,* 74(21), pp.5593-5599.

Maretzky, T., Reiss, K., Ludwig, A., Buchholz, J., Scholz, F., Proksch, E., de Strooper, B., Hartmann, D. and Saftig, P. 2005. ADAM10 mediates E-cadherin shedding and regulates epithelial cell-cell adhesion, migration, and beta-catenin translocation. *Proceedings of the National Academy of Sciences of the United States of America,* 102(26), pp.9182-9187.

Margadant, C. and Sonnenberg, A. 2010. Integrin-TGF-beta crosstalk in fibrosis, cancer and wound healing. *EMBO Reports,* 11(2), pp.97-105.

Marouga, R., David, S. and Hawkins, E. 2005. The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Analytical and Bioanalytical Chemistry,* 382(3), pp.669-678.

Marte, B. 2013. Tumour heterogeneity. *Nature,* 501(7467), pp.327.

Marusyk, A. and Polyak, K. 2010. Tumor heterogeneity: Causes and consequences. *Biochimica Et Biophysica Acta,* 1805(1), pp.105-117.

Matsubara, M., Titani, K., Taniguchi, H. and Hayashi, N. 2003. Direct involvement of protein myristoylation in myristoylated alanine-rich C kinase substrate (MARCKS)-calmodulin interaction. *The Journal of Biological Chemistry,* 278(49), pp.48898-48902.

McBride, S., Meleady, P., Baird, A., Dinsdale, D. and Clynes, M. 1998a. Human lung carcinoma cell line DLKP contains 3 distinct subpopulations with different growth and attachment properties. *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine,* 19(2), pp.88-103.

McBride, S., Meleady, P., Baird, A., Dinsdale, D. and Clynes, M. 1998b. Human lung carcinoma cell line DLKP contains 3 distinct subpopulations with different growth and attachment properties. *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine,* 19(2), pp.88-103.

Micallef, J., Taccone, M., Mukherjee, J., Croul, S., Busby, J., Moran, M.F. and Guha, A. 2009. Epidermal growth factor receptor variant III-induced glioma invasion is mediated through myristoylated alanine-rich protein kinase C substrate overexpression. *Cancer Research,* 69(19), pp.7548-7556.

Michalovitz, D., Halevy, O. and Oren, M. 1990. Conditional inhibition of transformation and of cell proliferation by a temperature-sensitive mutant of p53. *Cell,* 62(4), pp.671-680.

Miskovic, J., Brekalo, Z., Vukojevic, K., Miskovic, H.R., Kraljevic, D., Todorovic, J. and Soljic, V. 2015. Co-expression of TTF-1 and neuroendocrine markers in the human fetal lung and pulmonary neuroendocrine tumors. *Acta Histochemica,* 117(4-5), pp.451-459.

Miyazaki, H., Goto, A., Hino, R., Ota, S., Okudaira, R., Murakawa, T., Nakajima, J. and Fukayama, M. 2011. Pleural cavity angiosarcoma arising in chronic expanding hematoma after pneumonectomy. *Human Pathology,* 42(10), pp.1576-1579.

Mok, T.S., Wu, Y.L., Thongprasert, S., Yang, C.H., Chu, D.T., Saijo, N., Sunpaweravong, P., Han, B., Margono, B., Ichinose, Y., Nishiwaki, Y., Ohe, Y., Yang, J.J., Chewaskulyong, B., Jiang, H., Duffield, E.L., Watkins, C.L., Armour, A.A. and Fukuoka, M. 2009. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *The New England Journal of Medicine,* 361(10), pp.947-957.

Monteoliva, L. and Albar, J.P. 2004. Differential proteomics: An overview of gel and non-gel based approaches. *Briefings in Functional Genomics & Proteomics,* 3(3), pp.220-239.

Mori, N., Yokota, J., Akiyama, T., Sameshima, Y., Okamoto, A., Mizoguchi, H., Toyoshima, K., Sugimura, T. and Terada, M. 1990. Variable mutations of the RB gene in small-cell lung carcinoma. *Oncogene,* 5(11), pp.1713-1717.

Murphy, J.P., Everley, R.A., Coloff, J.L. and Gygi, S.P. 2014. Combining amine metabolomics and quantitative proteomics of cancer cells using derivatization with isobaric tags. *Analytical Chemistry,* 86(7), pp.3585-3593.

Nakajima, S., Doi, R., Toyoda, E., Tsuji, S., Wada, M., Koizumi, M., Tulachan, S.S., Ito, D., Kami, K., Mori, T., Kawaguchi, Y., Fujimoto, K., Hosotani, R. and Imamura, M. 2004. N-cadherin expression and epithelial-mesenchymal transition in pancreatic carcinoma. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research,* 10(12 Pt 1), pp.4125-4133.

Nakamura, T., Matsumoto, K., Kiritoshi, A., Tano, Y. and Nakamura, T. 1997. Induction of hepatocyte growth factor in fibroblasts by tumor-derived factors affects invasive growth of tumor cells: In vitro analysis of tumor-stromal interactions. *Cancer Research,* 57(15), pp.3305-3313.

Nakazawa, K., Kurishima, K., Tamura, T., Kagohashi, K., Ishikawa, H., Satoh, H. and Hizawa, N. 2012. Specific organ metastases and survival in small cell lung cancer. *Oncology Letters,* 4(4), pp.617-620.

Neilson, K.A., Ali, N.A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S.C. and Haynes, P.A. 2011. Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics,* 11(4), pp.535-553.

Nishino, K., Imamura, F., Kumagai, T., Uchida, J., Akazawa, Y., Okuyama, T. and Tomita, Y. 2011. Small-cell lung carcinoma with long-term survival: A case report. *Oncology Letters,* 2(5), pp.827-830.

Noh, S.J., Baek, H.A., Park, H.S., Jang, K.Y., Moon, W.S., Kang, M.J., Lee, D.G., Kim, M.H., Lee, J.H. and Chung, M.J. 2013. Expression of SIRT1 and cortactin is associated with progression of non-small cell lung cancer. *Pathology, Research and Practice,* 209(6), pp.365-370.

O'Farrell, P.H. 1975. High resolution two-dimensional electrophoresis of proteins. *The Journal of Biological Chemistry,* 250(10), pp.4007-4021.

Oft, M., Heider, K.H. and Beug, H. 1998. TGFbeta signaling is necessary for carcinoma cell invasiveness and metastasis. *Current Biology : CB,* 8(23), pp.1243-1252.

Okamoto, A., Hussain, S.P., Hagiwara, K., Spillare, E.A., Rusin, M.R., Demetrick, D.J., Serrano, M., Hannon, G.J., Shiseki, M. and Zariwala, M. 1995. Mutations in the p16INK4/MTS1/CDKN2, p15INK4B/MTS2, and p18 genes in primary and metastatic lung cancer. *Cancer Research,* 55(7), pp.1448-1451.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics : MCP,* 1(5), pp.376-386.

Osada, H., Tatematsu, Y., Yatabe, Y., Horio, Y. and Takahashi, T. 2005. ASH1 gene is a specific therapeutic target for lung cancers with neuroendocrine features. *Cancer Research,* 65(23), pp.10680-10685.

Otsubo, T., Iwaya, K., Mukai, Y., Mizokami, Y., Serizawa, H., Matsuoka, T. and Mukai, K. 2004. Involvement of Arp2/3 complex in the process of colorectal carcinogenesis. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc,* 17(4), pp.461-467.

Ozanne, B.W., Spence, H.J., McGarry, L.C. and Hennigan, R.F. 2007. Transcription factors control invasion: AP-1 the first among equals. *Oncogene,* 26(1), pp.1-10.

Pailler, E., Auger, N., Lindsay, C.R., Vielh, P., Islas-Morris-Hernandez, A., Borget, I., Ngo-Camus, M., Planchard, D., Soria, J.C., Besse, B. and Farace, F. 2015. High level of chromosomal instability in circulating tumor cells of ROS1-rearranged non-small-cell lung cancer. *Annals of Oncology : Official Journal of the European Society for Medical Oncology / ESMO,* 26(7), pp.1408-1415.

Parker, H.R., Li, Z., Sheinin, H., Lauzon, G. and Pasdar, M. 1998. Plakoglobin induces desmosome formation and epidermoid phenotype in N-cadherin-expressing squamous carcinoma cells deficient in plakoglobin and E-cadherin. *Cell Motility and the Cytoskeleton,* 40(1), pp.87-100.

Paschos, K.A., Canovas, D. and Bird, N.C. 2009. The role of cell adhesion molecules in the progression of colorectal cancer and the development of liver metastasis. *Cellular Signalling,* 21(5), pp.665-674.

Patel, V., Martin, D., Malhotra, R., Marsh, C.A., Doci, C.L., Veenstra, T.D., Nathan, C.A., Sinha, U.K., Singh, B., Molinolo, A.A., Rusling, J.F. and Gutkind, J.S. 2013. DSG3 as a biomarker for the ultrasensitive detection of occult lymph node metastasis in oral cancer using nanostructured immunoarrays. *Oral Oncology,* 49(2), pp.93-101.

Perez-Moreno, P., Brambilla, E., Thomas, R. and Soria, J.C. 2012. Squamous cell carcinoma of the lung: Molecular subtypes and therapeutic opportunities. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research,* 18(9), pp.2443-2451.

Qian, W.J., Jacobs, J.M., Liu, T., Camp, D.G.,2nd and Smith, R.D. 2006. Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications. *Molecular & Cellular Proteomics : MCP,* 5(10), pp.1727-1744.

Randell, S.H., Comment, C.E., Ramaekers, F.C. and Nettesheim, P. 1991. Properties of rat tracheal epithelial cells separated based on expression of cell surface alpha-galactosyl end groups. *American Journal of Respiratory Cell and Molecular Biology,* 4(6), pp.544-554.

Rapley, J., Nicolas, M., Groen, A., Regue, L., Bertran, M.T., Caelles, C., Avruch, J. and Roig, J. 2008. The NIMA-family kinase Nek6 phosphorylates the kinesin Eg5 at a novel site necessary for mitotic spindle formation. *Journal of Cell Science,* 121(Pt 23), pp.3912-3921.

Rassat, J., Robenek, H. and Themann, H. 1981. Structural relationship between desmosomes and mitochondria in human livers exhibiting a wide range of diseases. *The American Journal of Pathology,* 105(3), pp.207-211.

Ren, L., Hong, S.H., Cassavaugh, J., Osborne, T., Chou, A.J., Kim, S.Y., Gorlick, R., Hewitt, S.M. and Khanna, C. 2009. The actin-cytoskeleton linker protein ezrin is regulated during osteosarcoma metastasis by PKC. *Oncogene,* 28(6), pp.792-802.

Reya, T., Morrison, S.J., Clarke, M.F. and Weissman, I.L. 2001. Stem cells, cancer, and cancer stem cells. *Nature,* 414(6859), pp.105-111.

Reynolds, S.D., Giangreco, A., Power, J.H. and Stripp, B.R. 2000a. Neuroepithelial bodies of pulmonary airways serve as a reservoir of progenitor cells capable of epithelial regeneration. *The American Journal of Pathology,* 156(1), pp.269-278.

Reynolds, S.D., Hong, K.U., Giangreco, A., Mango, G.W., Guron, C., Morimoto, Y. and Stripp, B.R. 2000b. Conditional clara cell ablation reveals a self-renewing progenitor function of pulmonary neuroendocrine cells. *American Journal of Physiology.Lung Cellular and Molecular Physiology,* 278(6), pp.L1256-63.

Roan, N.R., Liu, H., Usmani, S.M., Neidleman, J., Muller, J.A., Avila-Herrera, A., Gawanbacht, A., Zirafi, O., Chu, S., Dong, M., Kumar, S.T., Smith, J.F., Pollard, K.S., Fandrich, M., Kirchhoff, F., Munch, J., Witkowska, H.E. and Greene, W.C. 2014. Liquefaction of semen generates and later degrades a conserved semenogelin peptide that enhances HIV infection. *Journal of Virology,* 88(13), pp.7221-7234.

Rock, J.R., Randell, S.H. and Hogan, B.L. 2010. Airway basal stem cells: A perspective on their roles in epithelial homeostasis and remodeling. *Disease Models & Mechanisms,* 3(9-10), pp.545-556.

Rodrigo, J.P., Alvarez-Alija, G., Menendez, S.T., Mancebo, G., Allonca, E., Garcia-Carracedo, D., Fresno, M.F., Suarez, C. and Garcia-Pedrero, J.M. 2011. Cortactin and focal adhesion kinase as predictors of cancer risk in patients with laryngeal premalignancy. *Cancer Prevention Research (Philadelphia, Pa.),* 4(8), pp.1333-1341.

Rodrigues, R.G., Panizo-Santos, A., Cashel, J.A., Krutzsch, H.C., Merino, M.J. and Roberts, D.D. 2001. Semenogelins are ectopically expressed in small cell lung carcinoma. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research,* 7(4), pp.854-860.

Rohrbach, T.D., Shah, N., Jackson, W.P., Feeney, E.V., Scanlon, S., Gish, R., Khodadadi, R., Hyde, S.O., Hicks, P.H., Anderson, J.C., Jarboe, J.S. and Willey, C.D. 2015. The effector domain of MARCKS is a nuclear localization signal that regulates cellular PIP2 levels and nuclear PIP2 localization. *PloS One,* 10(10), pp.e0140870.

Rombouts, K., Carloni, V., Mello, T., Omenetti, S., Galastri, S., Madiai, S., Galli, A. and Pinzani, M. 2013. Myristoylated alanine-rich protein kinase C substrate (MARCKS) expression modulates the metastatic phenotype in human and murine colon carcinoma in vitro and in vivo. *Cancer Letters,* 333(2), pp.244-252.

Rosen, A., Keenan, K.F., Thelen, M., Nairn, A.C. and Aderem, A. 1990. Activation of protein kinase C results in the displacement of its myristoylated, alanine-rich substrate from punctate structures in macrophage filopodia. *The Journal of Experimental Medicine,* 172(4), pp.1211-1215.

Sapir, T., Levy, T., Sakakibara, A., Rabinkov, A., Miyata, T. and Reiner, O. 2013. Shootin1 acts in concert with KIF20B to promote polarization of migrating neurons. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience,* 33(29), pp.11932-11948.

Sato, M., Aoyama, Y. and Kitajima, Y. 2000. Assembly pathway of desmoglein 3 to desmosomes and its perturbation by pemphigus vulgaris-IgG in cultured keratinocytes, as revealed by time-lapsed labeling immunoelectron microscopy. *Laboratory Investigation; a Journal of Technical Methods and Pathology,* 80(10), pp.1583-1592.

Savci-Heijink, C.D., Kosari, F., Aubry, M.C., Caron, B.L., Sun, Z., Yang, P. and Vasmatzis, G. 2009. The role of desmoglein-3 in the diagnosis of squamous cell carcinoma of the lung. *The American Journal of Pathology,* 174(5), pp.1629-1637.

Savitski, M.M., Nielsen, M.L. and Zubarev, R.A. 2005. New data base-independent, sequence tag-based scoring of peptide MS/MS data validates mowse scores, recovers below threshold data, singles out modified peptides, and assesses the quality of MS/MS techniques. *Molecular & Cellular Proteomics : MCP,* 4(8), pp.1180-1188.

Semba, S., Iwaya, K., Matsubayashi, J., Serizawa, H., Kataba, H., Hirano, T., Kato, H., Matsuoka, T. and Mukai, K. 2006. Coexpression of actin-related protein 2 and wiskott-aldrich syndrome family verproline-homologous protein 2 in adenocarcinoma of the lung. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research,* 12(8), pp.2449-2454.

Shaw, A.T., Winslow, M.M., Magendantz, M., Ouyang, C., Dowdle, J., Subramanian, A., Lewis, T.A., Maglathin, R.L., Tolliday, N. and Jacks, T. 2011. Selective killing of K-ras mutant cancer cells by small molecule inducers of oxidative stress. *Proceedings of the National Academy of Sciences of the United States of America,* 108(21), pp.8773-8778.

Shimada, T., Toriyama, M., Uemura, K., Kamiguchi, H., Sugiura, T., Watanabe, N. and Inagaki, N. 2008. Shootin1 interacts with actin retrograde flow and L1-CAM to promote axon outgrowth. *The Journal of Cell Biology,* 181(5), pp.817-829.

Song, H., Yao, E., Lin, C., Gacayan, R., Chen, M.H. and Chuang, P.T. 2012. Functional characterization of pulmonary neuroendocrine cells in lung development, injury, and tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America,* 109(43), pp.17531-17536.

Song, J. 2007. EMT or apoptosis: A decision for TGF-beta. *Cell Research,* 17(4), pp.289-290.

Souchelnytskyi, S. 2005. Bridging proteomics and systems biology: What are the roads to be traveled? *Proteomics,* 5(16), pp.4123-4137.

Spizz, G. and Blackshear, P.J. 2001. Overexpression of the myristoylated alanine-rich C-kinase substrate inhibits cell adhesion to extracellular matrix components. *The Journal of Biological Chemistry,* 276(34), pp.32264-32273.

Stannard, W. and O'Callaghan, C. 2006. Ciliary function and the role of cilia in clearance. *Journal of Aerosol Medicine : The Official Journal of the International Society for Aerosols in Medicine,* 19(1), pp.110-115.

Suda, K., Tomizawa, K. and Mitsudomi, T. 2010. Biological and clinical significance of KRAS mutations in lung cancer: An oncogenic driver that contrasts with EGFR mutation. *Cancer Metastasis Reviews,* 29(1), pp.49-60.

Sun, W., Liu, Q., Leng, J., Zheng, Y. and Li, J. 2015. The role of pyruvate dehydrogenase complex in cardiovascular diseases. *Life Sciences,* 121pp.97-103.

Suter, D.M. and Forscher, P. 2000. Substrate-cytoskeletal coupling as a mechanism for the regulation of growth cone motility and guidance. *Journal of Neurobiology,* 44(2), pp.97-113.

Talmadge, J.E. and Fidler, I.J. 2010. AACR centennial series: The biology of cancer metastasis: Historical perspective. *Cancer Research,* 70(14), pp.5649-5669.

Techasen, A., Loilome, W., Namwat, N., Takahashi, E., Sugihara, E., Puapairoj, A., Miwa, M., Saya, H. and Yongvanit, P. 2010. Myristoylated alanine-rich C kinase substrate phosphorylation promotes cholangiocarcinoma cell migration and metastasis via the protein kinase C-dependent pathway. *Cancer Science,* 101(3), pp.658-665.

Toriyama, M., Kozawa, S., Sakumura, Y. and Inagaki, N. 2013. Conversion of a signal into forces for axon outgrowth through Pak1-mediated shootin1 phosphorylation. *Current Biology : CB,* 23(6), pp.529-534.

Toriyama, M., Shimada, T., Kim, K.B., Mitsuba, M., Nomura, E., Katsuta, K., Sakumura, Y., Roepstorff, P. and Inagaki, N. 2006. Shootin1: A protein involved in the organization of an asymmetric signal for neuronal polarization. *The Journal of Cell Biology,* 175(1), pp.147-157.

Toyooka, S., Tsuda, T. and Gazdar, A.F. 2003. The TP53 gene, tobacco exposure, and lung cancer. *Human Mutation,* 21(3), pp.229-239.

Travis, W.D., Brambilla, E., Noguchi, M., Nicholson, A.G., Geisinger, K., Yatabe, Y., Ishikawa, Y., Wistuba, I., Flieder, D.B., Franklin, W., Gazdar, A., Hasleton, P.S., Henderson, D.W., Kerr, K.M., Petersen, I., Roggli, V., Thunnissen, E. and Tsao, M. 2013. Diagnosis of lung cancer in small biopsies and cytology: Implications of the 2011 international association for the study of lung cancer/american thoracic society/european respiratory society classification. *Archives of Pathology & Laboratory Medicine,* 137(5), pp.668-684.

Tsang, S.M., Brown, L., Gadmor, H., Gammon, L., Fortune, F., Wheeler, A. and Wan, H. 2012. Desmoglein 3 acting as an upstream regulator of rho GTPases, rac-1/Cdc42 in the regulation of actin organisation and dynamics. *Experimental Cell Research,* 318(18), pp.2269-2283.

Tsang, S.M., Liu, L., Teh, M.T., Wheeler, A., Grose, R., Hart, I.R., Garrod, D.R., Fortune, F. and Wan, H. 2010. Desmoglein 3, via an interaction with E-cadherin, is associated with activation of src. *PloS One,* 5(12), pp.e14211.

Tuli, L. and Ressom, H.W. 2009. LC-MS based detection of differential protein expression. *Journal of Proteomics & Bioinformatics,* 2pp.416-438.

Unwin, R.D. 2010. Quantification of proteins by iTRAQ. *Methods in Molecular Biology (Clifton, N.J.),* 658pp.205-215.

Uruno, T., Liu, J., Zhang, P., Fan, Y., Egile, C., Li, R., Mueller, S.C. and Zhan, X. 2001. Activation of Arp2/3 complex-mediated actin polymerization by cortactin. *Nature Cell Biology,* 3(3), pp.259-266.

Van Goor, D., Hyland, C., Schaefer, A.W. and Forscher, P. 2012. The role of actin turnover in retrograde actin network flow in neuronal growth cones. *PloS One,* 7(2), pp.e30959.

Van Lommel, A. 2001. Pulmonary neuroendocrine cells (PNEC) and neuroepithelial bodies (NEB): Chemoreceptors and regulators of lung development. *Paediatric Respiratory Reviews,* 2(2), pp.171-176.

van Zijl, F., Krupitza, G. and Mikulits, W. 2011. Initial steps of metastasis: Cell invasion and endothelial transmigration. *Mutation Research,* 728(1-2), pp.23-34.

Veltman, D.M. and Insall, R.H. 2010. WASP family proteins: Their evolution and its physiological implications. *Molecular Biology of the Cell,* 21(16), pp.2880-2893.

Voyksner, R.D. and Lee, H. 1999. Investigating the use of an octupole ion guide for ion storage and high-pass mass filtering to improve the quantitative performance of electrospray ion trap mass spectrometry. *Rapid Communications in Mass Spectrometry : RCM,* 13(14), pp.1427-1437.

Wagner, P.L., Kitabayashi, N., Chen, Y.T. and Saqi, A. 2009. Combined small cell lung carcinomas: Genotypic and immunophenotypic analysis of the separate morphologic components. *American Journal of Clinical Pathology,* 131(3), pp.376-382.

Walenkamp, A.M., Sonke, G.S. and Sleijfer, D.T. 2009. Clinical and therapeutic aspects of extrapulmonary small cell carcinoma. *Cancer Treatment Reviews,* 35(3), pp.228-236.

Wang, L., Liu, T., Wang, Y., Cao, L., Nishioka, M., Aguirre, R.L., Ishikawa, A., Geng, L. and Okada, N. 2007. Altered expression of desmocollin 3, desmoglein 3, and beta-catenin in oral squamous cell carcinoma: Correlation with lymph node metastasis and cell proliferation. *Virchows Archiv : An International Journal of Pathology,* 451(5), pp.959-966.

Weaver, A.M. 2006. Invadopodia: Specialized cell structures for cancer invasion. *Clinical & Experimental Metastasis,* 23(2), pp.97-105.

Weed, S.A. and Parsons, J.T. 2001. Cortactin: Coupling membrane dynamics to cortical actin assembly. *Oncogene,* 20(44), pp.6418-6434.

Wei, J., Zhao, Z.X., Li, Y., Zhou, Z.Q. and You, T.G. 2014. Cortactin expression confers a more malignant phenotype to gastric cancer SGC-7901 cells. *World Journal of Gastroenterology : WJG,* 20(12), pp.3287-3300.

Wein, G., Rossler, M., Klug, R. and Herget, T. 2003. The 3'-UTR of the mRNA coding for the major protein kinase C substrate MARCKS contains a novel CU-rich element interacting with the mRNA stabilizing factors HuD and HuR. *European Journal of Biochemistry / FEBS,* 270(2), pp.350-365.

Weinstein, I.B. 2002. Cancer. addiction to oncogenes--the achilles heal of cancer. *Science (New York, N.Y.),* 297(5578), pp.63-64.

Weinstein, I.B. and Joe, A. 2008. Oncogene addiction. *Cancer Research,* 68(9), pp.3077-80; discussion 3080.

Westerman, B.A., Neijenhuis, S., Poutsma, A., Steenbergen, R.D., Breuer, R.H., Egging, M., van Wijk, I.J. and Oudejans, C.B. 2002. Quantitative reverse transcription-polymerase chain reaction measurement of HASH1 (ASCL1), a marker for small cell lung carcinomas with neuroendocrine features. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research,* 8(4), pp.1082-1086.

Winn, R.A., Bremnes, R.M., Bemis, L., Franklin, W.A., Miller, Y.E., Cool, C. and Heasley, L.E. 2002. Gamma-catenin expression is reduced or absent in a subset of human lung cancers and re-expression inhibits transformed cell growth. *Oncogene,* 21(49), pp.7497-7506.

Wirtz, D., Konstantopoulos, K. and Searson, P.C. 2011. The physics of cancer: The role of physical interactions and mechanical forces in metastasis. *Nature Reviews.Cancer,* 11(7), pp.512-522.

Wong, M., Hyodo, T., Asano, E., Funasaka, K., Miyahara, R., Hirooka, Y., Goto, H., Hamaguchi, M. and Senga, T. 2014. Silencing of STRN4 suppresses the malignant characteristics of cancer cells. *Cancer Science,* 105(12), pp.1526-1532.

Wright, J.R. 2004. Host defense functions of pulmonary surfactant. *Biology of the Neonate,* 85(4), pp.326-332.

Wu, H., Reynolds, A.B., Kanner, S.B., Vines, R.R. and Parsons, J.T. 1991. Identification and characterization of a novel cytoskeleton-associated pp60src substrate. *Molecular and Cellular Biology,* 11(10), pp.5113-5124.

Xiao, H., Zhang, L., Zhou, H., Lee, J.M., Garon, E.B. and Wong, D.T. 2012. Proteomic analysis of human saliva from lung cancer patients using two-dimensional difference gel electrophoresis and mass spectrometry. *Molecular & Cellular Proteomics : MCP,* 11(2), pp.M111.012112.

Yamaguchi, H. and Condeelis, J. 2007. Regulation of the actin cytoskeleton in cancer cell migration and invasion. *Biochimica Et Biophysica Acta,* 1773(5), pp.642-652.

Yamaguchi, H., Lorenz, M., Kempiak, S., Sarmiento, C., Coniglio, S., Symons, M., Segall, J., Eddy, R., Miki, H., Takenawa, T. and Condeelis, J. 2005. Molecular mechanisms of invadopodium formation: The role of the N-WASP-Arp2/3 complex pathway and cofilin. *The Journal of Cell Biology,* 168(3), pp.441-452.

Yamaguchi, H., Wyckoff, J. and Condeelis, J. 2005. Cell migration in tumors. *Current Opinion in Cell Biology,* 17(5), pp.559-564.

Yamashiro, S. and Watanabe, N. 2014. A new link between the retrograde actin flow and focal adhesions. *Journal of Biochemistry,* 156(5), pp.239-248.

Yang, L., Lin, C. and Liu, Z.R. 2006. P68 RNA helicase mediates PDGF-induced epithelial mesenchymal transition by displacing axin from beta-catenin. *Cell,* 127(1), pp.139-155.

Yarmola, E.G., Edison, A.S., Lenox, R.H. and Bubb, M.R. 2001. Actin filament cross-linking by MARCKS: Characterization of two actin-binding sites within the phosphorylation site domain. *The Journal of Biological Chemistry,* 276(25), pp.22351-22358.

Yates, J.R., Ruse, C.I. and Nakorchevsky, A. 2009. Proteomics by mass spectrometry: Approaches, advances, and applications. *Annual Review of Biomedical Engineering,* 11pp.49-79.

Yu, D., Makkar, G., Strickland, D.K., Blanpied, T.A., Stumpo, D.J., Blackshear, P.J., Sarkar, R. and Monahan, T.S. 2015. Myristoylated alanine-rich protein kinase substrate (MARCKS) regulates small GTPase Rac1 and Cdc42 activity and is a critical mediator of vascular smooth muscle cell migration in intimal hyperplasia formation. *Journal of the American Heart Association,* 4(10), pp.e002255.

Zhang, S., Li, Y., Wu, Y., Shi, K., Bing, L. and Hao, J. 2012. Wnt/beta-catenin signaling pathway upregulates c-myc expression to promote cell proliferation of P19 teratocarcinoma cells. *Anatomical Record (Hoboken, N.J.: 2007),* 295(12), pp.2104-2113.

Zhu, W., Smith, J.W. and Huang, C.M. 2010. Mass spectrometry-based label-free quantitative proteomics. *Journal of Biomedicine & Biotechnology,* 2010pp.840518.

Zimmermann, J., Brunner, C., Enculescu, M., Goegler, M., Ehrlicher, A., Kas, J. and Falcke, M. 2012. Actin filament elasticity and retrograde flow shape the force-velocity relation of motile cells. *Biophysical Journal,* 102(2), pp.287-295.

8.1. Appendices on Disc

A. Exponential Phase
B. Stationary Phase