Discovery and development of liquid biomarkers for ovarian and lung cancer

A thesis submitted for the degree of
Doctor of Philosophy

by

Dimple CHUDASAMA

College of Health and Life Sciences

BRUNEL UNIVERSITY LONDON

Abstract

Survival rates in cancers have improved vastly over the years. However, some survival rates remain extremely low, as is the case for ovarian and lung cancer. The lack of robust and reliable biomarkers is strongly reflected in the absence of pre-screening programs, and as such, most patients in these cancer types are diagnosed only in advanced stages, leaving few treatment options. Moreover, relapse and resistance to therapies adds to the complexities of treating these diseases, even in the era of targeted drug development.

Research has shown the presence of cancer material, in the form of circulating cancer cells (CTCs) and genomic material in the blood of patients, opening the possibility of 'liquid biopsies'. Liquid biopsies allow sampling of the disease to provide phenotypic and genomic data on the cancer in real-time and on a routine basis. Moreover, they overcome obstacles currently faced by conventional tissue biopsies.

In this work we evaluate the use of a novel CTC imaging flow-cytometry platform, and report the ability to characterise and quantify these cells in blood samples. Moreover, we report significantly higher levels of CTCs in cancer patients compared to controls, and found them to be associated with a poorer prognosis. In particular, in lung cancer we observe these findings even in the early stages, suggesting a potential diagnostic use for this assay. We detect a similar trend in when analysing the ctDNA and suggest the possibility of using this technique with a prognostic value in the advanced setting.

We also report on the analysis of existing microarray data by use of unique gene regulatory networks to identify biomarkers of interest. RAD51AP1 was identified by this process. Clinical validation revealed an over-expression of this gene in both tissue and blood of ovarian and lung cancers. Moreover, the gene over-expression was associated with a poor overall survival. Functional analysis *in vitro* revealed silencing RAD51AP1 suppressed tumour growth, in addition, various tumorigenic proteins were down-regulated, whilst apoptotic and immune genes were up-regulated. These results suggest a role for RAD51AP1 as a potential therapeutic target.

In this study, we also demonstrate the ability to further exploit tumour genomic material in the blood by means of RNAseq, cancer panels, and CNI scoring to identify novel markers, that play an important role in disease genesis and evolution. RNAseq analysis identified XIST as a gene up-regulated in the blood and tissue of lung cancers. The ovarian cancer panels revealed 2 unique gene signatures in the ovarian cancer patients. With the CNI analyses also highlighting chromosomal aberrations from plasma analysis of cancer patients.

Collectively, the use of all these techniques and exploitation of available blood based biomarkers could see significant improvements to survival rates in these, currently devastating diseases.

Contents

List of Figures

List of Tables

To my mum, for all the sacrifices & to my beautiful daughter Binti-Ria, you are the world to me…

Acknowledgements

Firstly, I owe a huge thank you to two very important individuals, Dr Emmanouil (Manos) Karteris and Mr Vladimir Anikin. I wouldn't be here if it wasn't for you both. Manos, thank you for believing in me and giving me this opportunity, and the countless times you had to convince me I could do this. I am incredibly grateful for your guidance, support and understanding at tough times.

Mr Anikin, you have given me so much, from opportunity to knowledge, experience, and confidence in myself. You always pushed me to go one step further, and here I am today. I have learnt and seen so much working with you. I could never have done this without you. Your guidance as a clinical supervisor has been invaluable.I can't thank you enough for all you have done over the past 5 years.

I would also like to thank Dr Marica Hall for her support as a clinical supervisor. Your ideas, opinions and guidance were invaluable to this project.

I was very fortunate to have several academic and clinical supervisors, one of whom I owe so much to is Dr Cristina Sisu, I truly would not have finished this thesis to the high standard without you. You gave up so much of your time for me, and introduced me to ShareLatex, and taught me how to use 'R'. You were there day and night to support me when I was stuck. Despite joining my project in the final year, you have done so much and made a huge impact, and I am so very grateful for all you have done for me. I was very fortunate to have you as a second supervisor, and I value all your support, emotional, professional, and personal, thank you so much!

I want to say thank you to all my friends and colleagues, Haroon, Stephen, Karly, Sheila and Rooben for all the support in the labs, and more importantly, all the jokes and laughs. Also I would like to say a special thank you to Dr Terry Roberts, for helping me figure out why my qPCRs weren't working, and for all the support and guidance you gave me throughout my PhD.

Finally I want to say a special thank you to my wonderful family. To my loving husband Hash, for being patient and supportive always, to my lovely daughter Binti-Ria for always encouraging me to keep going. To all my siblings, Divendra and Daniel for silently caring and my amazing baby sister Bobby, for always listening to my qualms and slapping me back to reality. To Trish and Dillu and my wonderful mother whose always been there for me. I love you all soooo much!!!! Thank you Bhagwan, this would never have happened without you.

Declaration of Authorship

I, Dimple CHUDASAMA, declare that this thesis titled, "Discovery and development of liquid biomarkers for ovarian and lung cancer" and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Publications, Presentations & Awards

Peer reviewed publications resulting from this work

- (*Chapter 6*) Chudasama D, *et al*. A Transcriptomics Liquid Biopsy Approach in Lung Cancer Diagnosis and Prognosis. *In preparation.*
- (*Chapter 3 & 5*) Chudasama D, *et al*. Liquid Biomarker Approaches in Ovarian Cancer Diagnosis and Prognosis. *In preparation.*
- (*Chapter 5*) Chudasama DY, Aladag Z, Felicien MI, Hall M, Beeson J, Asadi N, Gidron Y, Karteris E & Anikin VB. Prognostic value of the DNA Integrity Index in patients with malignant lung tumours. *Oncotarget* (2018) 9:21281-21288.
- (*Chapter 4*) Chudasama D, Bo V, Hall M, Anikin V, Jeyaneethi J, Gregory J, Pados G, Tucker A, Harvey A, Pink R, Karteris E. Identification of novel cancer biomarkers of prognostic value using specific gene regulatory networks (GRN): a novel role of RAD51AP1 for ovarian and lung cancers. *Carcinogenesis* (2017) 10.1093/carcin/bgx122.
- (*Chapter 3*) Chudasama DY, Freydina DV, Freidin MB, Leung M, Montero Fernandez A, Rice A, Nicholson AG, Karteris E, Anikin V, Lim E. Inertia based microfluidic capture and characterisation of circulating tumour cells for the diagnosis of lung cancer. *Annals of Translation Medicine* (2016) 4:480.

Conference poster presentations

- European Association of Cancer Research - American Association of Cancer Research - SIC Special Conference - Cancer Biology, Florence (Italy), 2017
- European Association of Cancer Research - Cancer Genomics, Manchester (UK), 2016
- European Association of Cancer Research - Cancer Genomics, Cambridge (UK), 2015

Awards

- **1st Prize, Poster Presentation** College of Health & Sciences, Brunel University London, 2017
- **Vice Chancellors Travel Prize**, 2017
- **Vice Chancellors Travel Prize**, 2016
- **5th Prize, Poster Presentation** College of Health & Sciences Vice Chancellor Award, Brunel University London, 2014

List of Abbreviations

ALK	anaplastic lymphoma kinase
AQ	absolute quantification
AUC	area under the curve
BN	Baysian network
CA125	cancer antigen 125
cDNA	complementary DNA
cfDNA	circulating free DNA
CNI	copy number instability
COL12A1	collagen type XII, alpha 1
CRUK	Cancer Research UK
CTC	circulating tumour cells
ctDNA	circulating DNA
ctNA	circulating nucleic acids
ctRNA	circulating RNA
DAB	3,3-diaminobenzidine
DMEM	Dulbecco's modified eagles medium
DMSO	dimethyl sulfoxide
DNA	dezoxyribonucleic acid
DRS	diffuse reflectance spectroscopy
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
EGFR	epidermal growth factor receptor
EMT	epithelial-to- mesenchymal transition
EpCAM	epithelial cell adhesion molecule
ESS	elastic scattering spectroscopy
FBS	etal bovine serum
FDA	Food and Drugs Administration
FFPE	fresh frozen paraffin embedded
FIGO	Federation of Gynaecology and Obstetrics
FLIM	fluorescence lifetime imaging microscopy
FPKM	fragments per kilobase of transcript per million mapped reads
FSTL1	follistatin like protein 1

FTE	fallopian tube epithelium
GO	gene ontology
GP	general practitioners
GRN	gene regulatory network
H&E	haematoxylin and eosin
HD	high definition
HDHD1	haloacid dehalogenase-like hydrolase domain-containing 1
HGSOC	high-grade serous ovarian cancer
HRP	horseradish peroxidase
HTA	Human Tissue Act
IHC	immunohistochemistry
IMS	industrial methylated spirits
IR	infrared
KEGG	Kyoto Encyclopedia of Genes and Genomes
KM	Kaplan Meier
LC	lung cancer
LDA	linear discriminant analysis
MET	mesenchymal–epithelial transition
MIQE	minimum information for publication of quantitative real-time PCR experiments
MPAK	mitogen-activated protein kinase
mRNA	messenger RNA
MTA	material transfer agreement
MW	molecular weight
NCBI	National Center for Biotechnology In- formation
NHS	national health service
NICE	National Institute for Health and Care Excellence
NRES	National Research Ethical Authority
NSCLC	non-small cell lung cancer
NTC	non template control
OCT	optical coherence tomography
OD	optical density
OS	overall survival
OSE	ovarian surface epithelium
OV	ovarian cancer
PARP	poly ADP ribose polymerase
PBS	phosphate buffered saline
PCA	principal component analysis
PCR	polymerase chain reaction

PFA	paraformaldehyde
PI	propidium iodide
PMT	photomultiplier tube
qPCR	quantitative polymerase chain reaction
RAD51AP1	RAD51 associated protein 1
RBC	red blood cell
RNA	ribonucleic acid
RNAseq	RNA sequencing
ROC	receiver operating characteristic curve
ROS	reactive oxygen species
RPM	rotations per minute
RQ	relative quantification
rRNA	ribosomal RNA
SCLC	small cell lung cancer
SDS	sodium-dodecyl-sulphate
SDS-PAGE	sodium-dodecyl-sulphate polyacrylamide gel electrophoresis
SINE	short interspersed elements
siRNA	small interfering RNA
SPRR1A	small proline rich protein 1A
TBE	tris/borate/EDTA buffer
TBS	tris-buffered saline
TCGA	The Cancer Genome Atlas
TEMED	tetramethylethylenediamine
TKI	tyrosine kinase inhibitor
TNF	tumour necrosis factor
TVUS	transvaginal ultrasound
UNIP	unique network identification pipeline
VEGF	vascular endothelial growth factor
WBC	white blood cell

Chapter 1

Introduction

1.1 Prolegomenon

Cancer diagnosis and treatment has come a long way over the past four decades; despite this, survival rates in some cancers remain poor. This is particularly true in lung and ovarian cancers, where ten-year survival rates are only 5 and 35%, respectively (CRUK, 2017). An important contributing factor to these poor survival rates is the late diagnosis. Even for the advanced disease, the symptoms are often vague which results in the majority of patients being diagnosed only at late stages (e.g. III and IV). Moreover, it is well known that prognosis in cancer is related to staging, with earlier diagnosis offering more treatment options and better chances of survival. To complicate things further, the limited knowledge of interpatient variability, which defines the disease and modulates response to therapeutics, is also an important issue clinicians' face when treating patients. In light of this, much interest is now focused on 'precision medicine' or personalised therapy, whereby individual patients are given a treatment tailored specifically to their disease, taking into account their individual genetic profile and potential responses to therapeutics. This allows the identification of the best treatment regime for the patient based on their exact needs.

In support of precision medicine, much research is currently being channelled in to the development of robust 'liquid biopsy' biomarkers. 'Liquid biopsy', is an innovative technology that comes to improve and potentially replace the classic invasive tumour biopsy. It consists of exploiting blood samples for various biomarkers, that can reflect the genetic diversity of the tumour profile, and act as a surrogate biomarker to aid in cancer diagnosis and treatment. Several such biomarkers have been identified, including circulating tumour cells (CTCs), circulating nucleic acids (ctNAs) (i.e. DNA and RNA), and genomic material in whole blood. Here, we look holistically at these biomarkers, evaluate their efficacy, and judge their clinical utility and relevance to identify a robust biomarker that can be used in 'liquid biopsy', in the move towards personalised treatments.

1.2 Ovarian cancer

1.2.1 Incidence of ovarian cancer

Over 7,000 new ovarian cancer cases are diagnosed in the UK every year, making it the fifth most common cancer in the UK (CRUK, 2017). Despite a slight decrease in the number of new cases per year in the UK over the past 20 years, the gap between incidence and deaths remains unchanged, suggesting little or no improvement in overall survival rates, Fig 1.1. As a result, ovarian cancer is still a deadly disease, with a cure rate of only 30% (Lengyel, 2010).

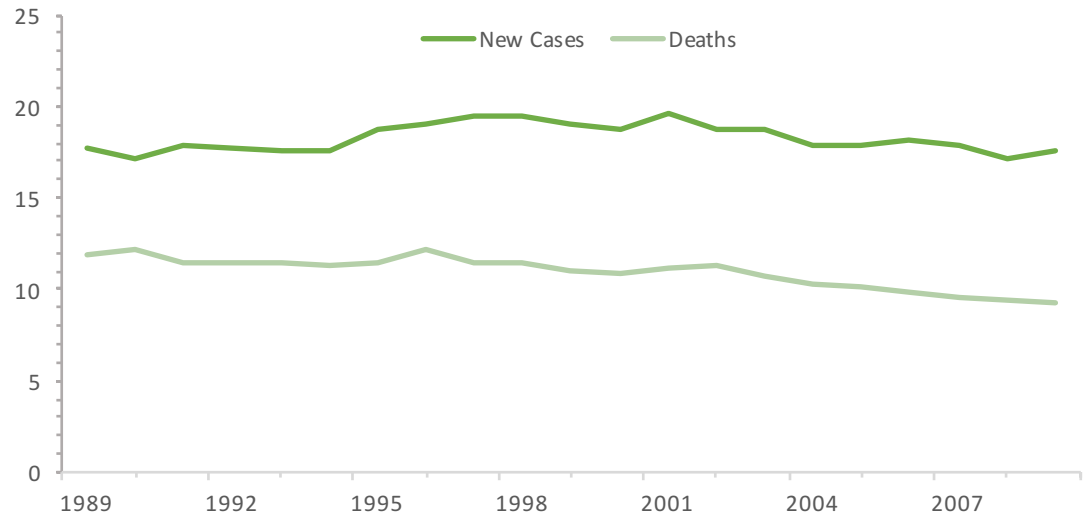


FIGURE 1.1: **Incidence and deaths of Ovarian cancer by year.** Rate of ovarian cancer incidence and deaths per 100,000 by year from 1989 to 2009 in UK. Data sourced from Poole and Nordin, 2012.

Poor survival is partly contributed to the late diagnosis, where treatment options become limited, and metastatic spread adds a new level of complexity. On average, 70% of patients are being diagnosed at stages III and IV (Buys et al., 2011), with 47% being classed as distant due to cancer spread, Fig 1.2. This has far implications and can be associated with eventual relapses despite the use of conventional treatment approaches (surgery, chemotherapy) (Li et al., 2004).

Typically, if the disease is confined to the ovaries, the cancer is classed as localised. Once the cancer begins to spread into the nodes and around the body, the cancer is classed as regional or distant (CRUK, 2017). The majority of ovarian cancers are of epithelial origin (Vargas, 2014).

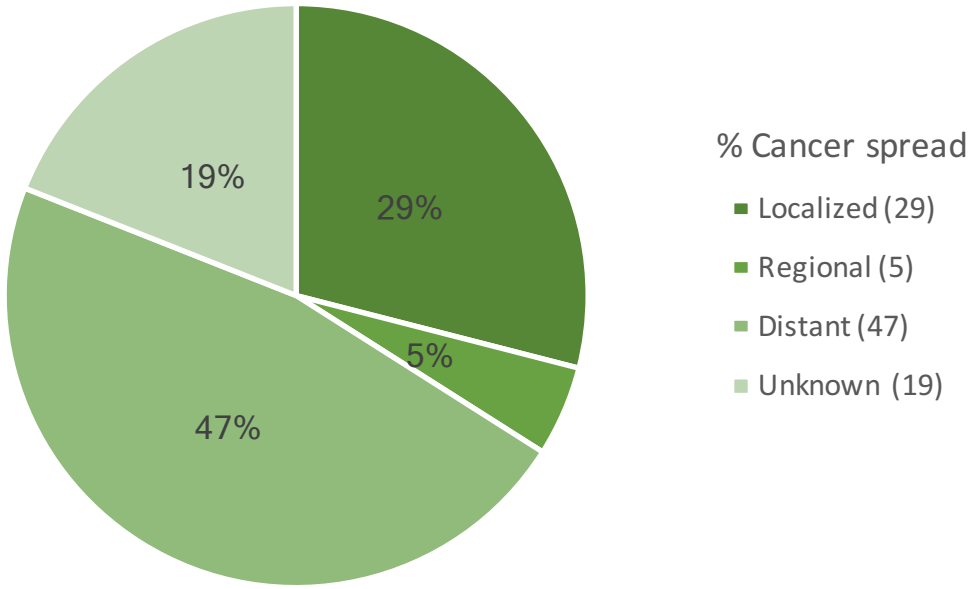


FIGURE 1.2: **Extent of cancer spread at time of diagnosis.** The extent of cancer spread at time of diagnosis, based on data from patients diagnosed in 2012, 2013 and 2014 and followed up in 2015 in UK. Data sourced from Bannister and Broggio, 2016.

Ovarian cancer is primarily a post-menopausal disease, with the majority of cases diagnosed in patients at the age of 55-74 years (45.5%), Fig 1.3.

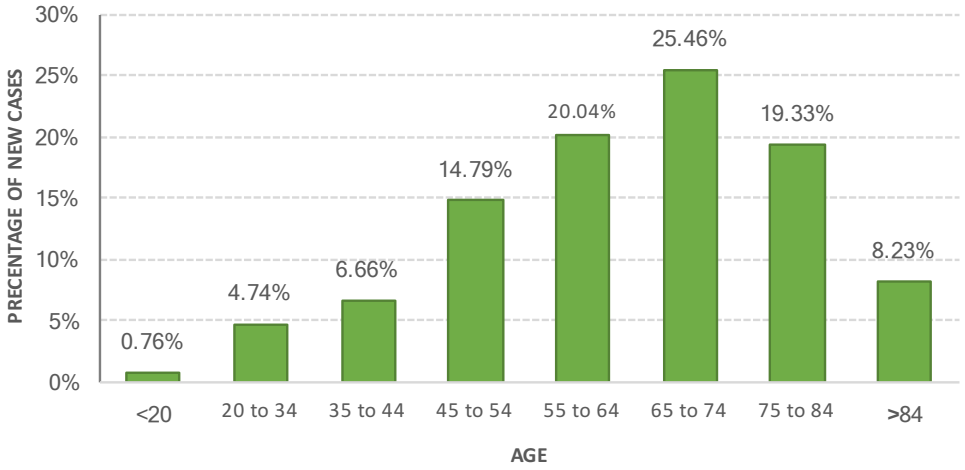


FIGURE 1.3: **Cancer diagnosis by age.** Chart showing age of cancer diagnosis, based on 2012 to 2014 figures. Data sourced from CRUK, 2017.

1.2.2 Staging and disease progression in ovarian cancer

All patients diagnosed with ovarian cancer, are staged on their disease according to the Federation of Gynaecology and Obstetrics (FIGO) staging system, detailed in Table 1.1, and diagrammatic representation of disease and stage can be seen in Fig 1.4 below.

Stage	Sub-stage	Description
I	Ia	The tumour is confined to one ovary with no signs of tumour on the surface.
	Ib	As Ia but involving both ovaries.
	Ic	The tumour is confined to one or both ovaries with either or all of the following: signs of the tumour on the surface of the ovary, rupture of tumour capsule before or during surgery, malignant cells found in ascites.
II	IIa	Metastasis outside the ovaries in the uterus or fallopian tubes.
	IIb	Metastasis to pelvic cavity organs for example the bladder.
III	IIIa	Metastasis to retroperitoneal lymph nodes or microscopic malignancy found outside the pelvis.
	IIIb	Tumour smaller than or equal to 2cm found outside the pelvic cavity including surface of liver and/or spleen.
	IIIc	Tumour bigger than 2cm found outside the pelvic cavity including surface of liver and/or spleen.
IV	IVa	Pleural effusion (fluid around the lungs) positive for malignant cells.
	IVb	Metastasis to distant sites including extra-abdominal and parenchymal liver or spleen involvement.

TABLE 1.1: **Ovarian cancer staging.** Ovarian cancer staging parameters, as defined by FIGO (Society for Gynecologic Oncology, 2014).

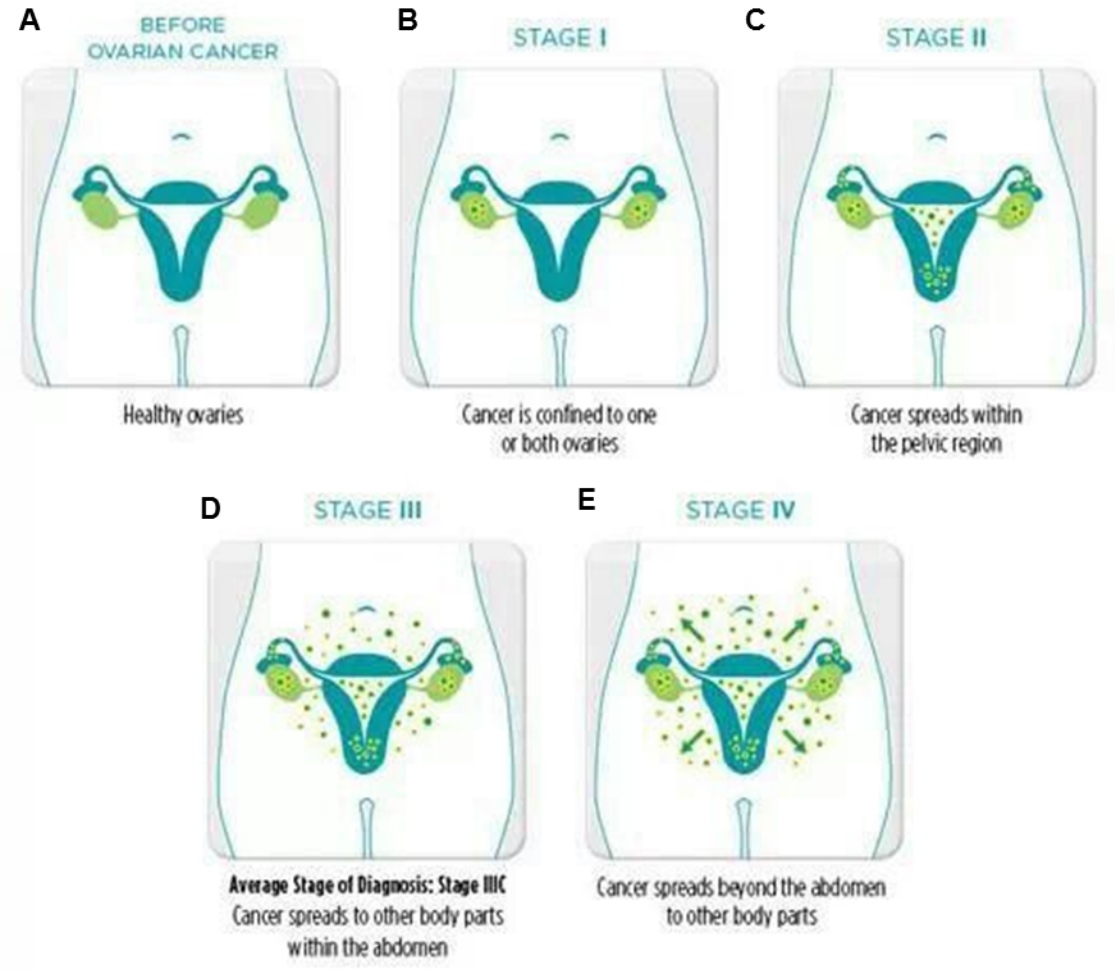


FIGURE 1.4: **Ovarian cancer location by stage.** Ovarian cancer location by stage. (A) Healthy ovaries with no cancer. (B) Stage I ovarian cancer, where the cancer is confined to one or both ovaries. (C) Stage II ovarian cancer, where the cancer spreads within the pelvic region. (D) Stage III ovarian cancer, the cancer has metastasised within the abdomen, this is also the average stage of diagnosis. (E) Stage IV ovarian cancer, metastasis has now spread beyond the abdomen to distant organs. Taken from http://smrfteal.org November 2017.

In addition to staging, patients are also graded on their ovarian cancer as detailed in Table 1.2.

Grade	Description
I	Well differentiated
II	Moderately differentiated
III	Poorly differentiated

TABLE 1.2: **Ovarian cancer grading.** Description of grading parameters in ovarian cancers.

The biology of ovarian carcinoma differs from that of hematogenous-initiatied metastasising tumours, as ovarian cancer cells primarily disseminate within the peritoneal cavity, and thus are only superficially invasive. The rapidly proliferating tumours however compress the visceral organs (Lengyel, 2010).

Ovarian carcinomas are most likely to originate from the surfaces of the ovary, the fallopian tube, or the mesothelium-lined peritoneal cavity. Growing evidence supports the fallopian tube epithelia as an etiological site for the development of high grade serous carcinoma of the ovaries and consequently, salpingectomy alone is emerging as a prophylactic option. The fallopian tube epithelium (FTE) and the ovarian surface epithelium (OSE), have both been shown to share common mesodermal embryological origin and a close anatomic proximity (George, Garcia, and Slomovitz, 2016). Ovarian cancer tumorigenesis progresses either from a slow growing borderline tumour to a well-differentiated carcinoma (type I), or involves a genetically unstable high-grade serous carcinoma that can metastasise rapidly (type II). During the initial stages of tumorigenesis, ovarian cancer cells undergo an epithelial-to-mesenchymal transition (EMT); this involves a change in cadherin and integrin expression and up-regulates the proteolytic pathways (Huber, Kraut, and Beug, 2005; Cavallaro and Christofori, 2004). These cancer cells are carried on the peritoneal fluid, where cancer cell spheroids overcome anoikis, and attach preferentially on the abdominal peritoneum or the omentum, at which point the cancer cells revert to their epithelial phenotype (Sawada et al., 2007).

Metastasis is initially regulated by a controlled interaction of adhesion receptors and proteases, whereas late metastasis is characterised by the oncogene-driven fast growth of tumour nodules on mesothelium covered surfaces, in turn causing ascites, bowel obstruction, and tumour cachexia (Lengyel, 2010).

1.2.3 Ovarian cancer survival

Long term survival rates in ovarian cancer remain comparatively poor, with a 10 year survival of only 35%, Fig 1.5. Ovarian cancer is the most common cause of deaths in genealogical cancers in the UK (CRUK, 2017).

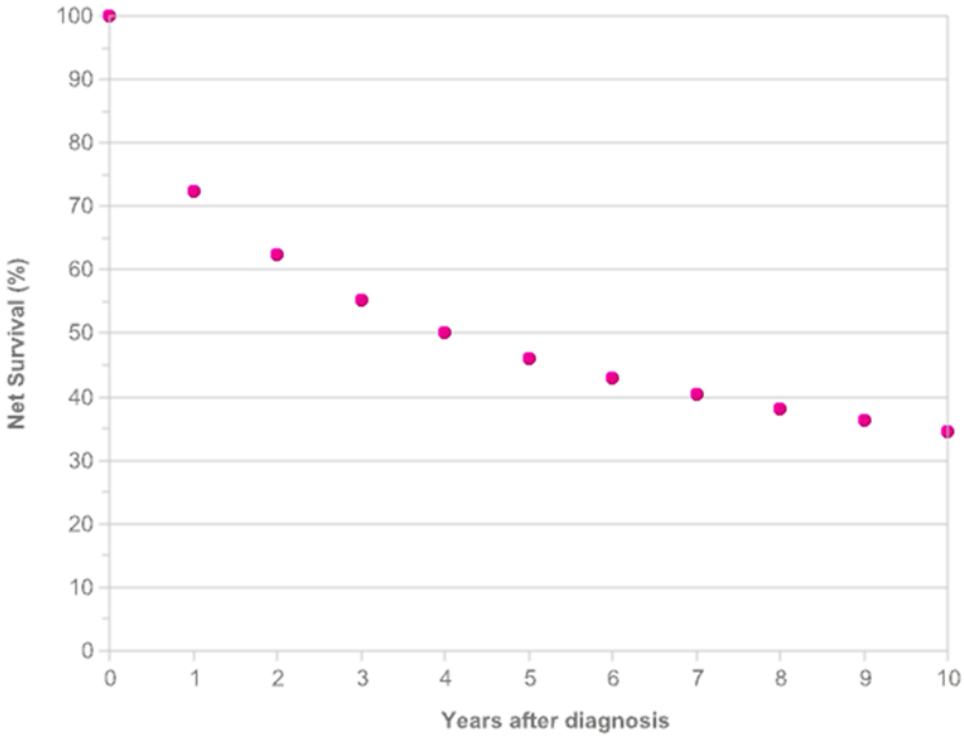


FIGURE 1.5: **Ovarian cancer survival rates over 10 years.** Chart showing a decline from approximately 72% at 1 year to approximately 35% at 10 years. Taken from CRUK, 2017.

Survival is also dependent on stage of the cancer with advanced cases showing poorer survival as presented in Fig 1.6.

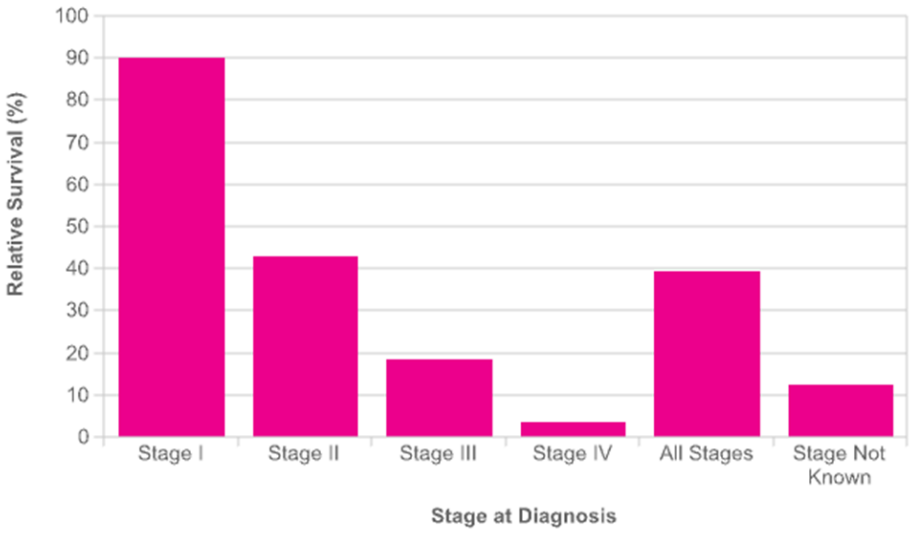


FIGURE 1.6: **Ovarian cancer survival by stage.** Ovarian 5 year survival data by stage of cancer. Data shows poorer survival with advancement in stage. Taken from CRUK, 2017.

The absence of a robust screening tool has hindered potential early diagnosis efforts, in part due to the absence of a reliable biomarker for this disease. Typical ovarian cancer screening exists as a transvaginal vaginal scan (Transvaginal ultrasound - TVUS), and detection of the cancer antigen 125 (CA125), neither of which have been successful. CA125 is commonly utilised as a means of monitoring treatment management of patients diagnosed with ovarian cancer (Gupta and Lis, 2009). CA125 is known to be elevated in ovarian cancers, however CA125 levels are also known to be elevated in various other non-ovarian cancer related diseases, deeming this inadequate as a screening tool (Gupta and Lis, 2009). Although due to lack of other markers, the CA125 serum levels are commonly used as a means of monitoring responses to chemotherapy, relapse, and disease progression.

CA125 is a membrane associated glycoprotein, known to be involved in cell adhesion, migration and metastasis. CA125 is commonly used clinically to monitor disease progression in patients undergoing ovarian cancer treatment, and had been suggested as a possible screening/diagnostic tool (Menon et al., 2015; Giannakouros et al., 2015; Reinartz et al., 2012; Rump et al., 2004). The efficacy of CA125 is questionable due to a high propensity for false positives. Conditions such as heart failure, endometriosis, and breast cancer also raise CA125 levels, as do demographic and lifestyle factors such as ethnicity, obesity, and tobacco (Folga et al., 2012; Mol et al., 1998; Johnson et al., 2008). A CA125 value of 35 units per mL is considered abnormally high, however Johnson *et al.* 2008 found this to be the case in 1.6% of non-cancer individuals (Johnson et al., 2008). Buys *et al.* 2011 carried out a randomised trial in the US to explore the utility of CA125 as a screening tool in conjunction with TVUS

(Buys et al., 2011). General findings suggested simultaneous screening of TVUS (for 4 years) and CA125 (offered annually for 6 months), compared with the usual care offered by their centre did not reduce ovarian cancer mortality. The false positive rate was reported as high as 8.4%, of which a diagnostic evaluation was based on, resulting in further complications for 15% of patients. All these results highlight the limitations and even the inadequacy of this approach (Buys et al., 2011). However, more recently a study described the benefits of early screening using the Risk of Ovarian Cancer Algorithm, along with the frequent of CA125 testing in women at an increased familial risk of ovarian cancer (Skates et al., 2017).

1.2.4 Treatment of ovarian cancer

Treatment in the UK for ovarian cancers is dependent on the type and stage of cancer, and consists of a combination of surgery, chemotherapy, and radiotherapy, with the majority of patients undergoing primary debulking surgery to remove malignant tissue, followed by further treatment (NHS choices).

Many agents are being developed against a range of potential new molecular therapeutic targets for the treatment of ovarian cancer. For this, research has been focused on tumour growth and metastasis. Several clinical trials are being run in the UK, Table1.3 lists a few targeted drug clinical trials.

Drug name	**Treatment/Target**
Cediranib	Inhibits the vascular endothelial growth factor (VEGF)
Thalidomide	antiangiogenic and immunomodulatory properties
Pazopanib	Receptor tyrosine kinase inhibitor (TKI)
Olaparib	inhibits poly ADP ribose polymerase (PARP)
Niraparib	PARP inhibitor

TABLE 1.3: **Details of current targeted therapies in ovarian cancer.** List of some current targeted drug clinical trials within the UK from Clinicaltrials.gov, 2018.

Angiogenesis is one of the processes central to tumour growth and metastasis. It is generally triggered by localised tissue conditions, such as hypoxia, which cause the release of angiogenic promoters from cells in the affected tissue. Vascular endothelial growth factor (VEGF) is one of the most important promoters produced by normal and neoplastic cells. It is a diffusible glycoprotein, and is an important regulator of physiological and pathological angiogenesis. Increased levels of VEGF expression have been found in most human malignancies including tumours of the lung, breast, thyroid, gastrointestinal tract, kidney, bladder, ovary, cervix, and pancreas, as well as angiosarcomas and glioblastomas. Increased VEGF serum levels have been correlated with poor survival (Li, Zeng, and Shen, 2014).

In ovarian cancer, angiogenesis has been shown to have a central role in both disease progression and prognosis. A direct relationship has been demonstrated between the expression of biomarkers for angiogenesis such as VEGF, the degree of neovascularization and the behaviour of epithelial ovarian cancers (Yoneda et al., 1998; Burger et al., 2007). These data suggest that pharmacological inhibitors of angiogenesis may have the capacity to arrest tumour progression. Several phase II trials of different antiangiogenic drugs have been reported to demonstrate activity against relapsed ovarian cancer (Burger et al., 2007; Friedlander et al., 2010).

1.2.5 Ovarian cancer relapse

Once ovarian cancer recurs, a curative option is almost non-existent, beyond this point the disease is considered chronic with sequential relapses and remissions. Over 80% of patients with advanced ovarian cancer will relapse and despite further good remissions from additional chemotherapy, they will usually die from their disease (Luvero, Milani, and Ledermann, 2014). Nationally there are likely to be about 4,200 patients (70% of the forecasted $\sim$7,000 new cases per annum), who will have recurrence of their cancer.

Sequential treatment strategies are employed to maximise quality and length of life. Many of these patients will receive three or more lines of chemotherapy but will eventually become resistant to standard therapies. Further treatment is required in order to prevent the onset of symptoms in this group of resistant patients. Preference is generally given for simple non-toxic therapies that will maintain or improve quality of life. In addition, after several lines of chemotherapy, many patients do not want further intravenous treatment with its' associated often weekly hospital visits. The adverse effects and ease of administration of any future treatments are of paramount importance to maximise patients' quality of life.

1.3 Lung cancer

1.3.1 Incidence of lung cancer

Lung cancer incidences in the UK, were reported at 46,403 in 2014, accounting for 13% of all cancer cases in the UK (CRUK, 2017). Despite an overall decrease in incidence over the years, the rate of deaths has remained constant, Fig 1.7.

Moreover, similarly to ovarian cancer, late diagnosis is also responsible for the poor overall survival rates, where there are no effective screening programs. Over 60% of patients are diagnosed in advanced stages, where treatment options become limited and prognosis is poor (Bannister and Broggio, 2016), Fig 1.8.

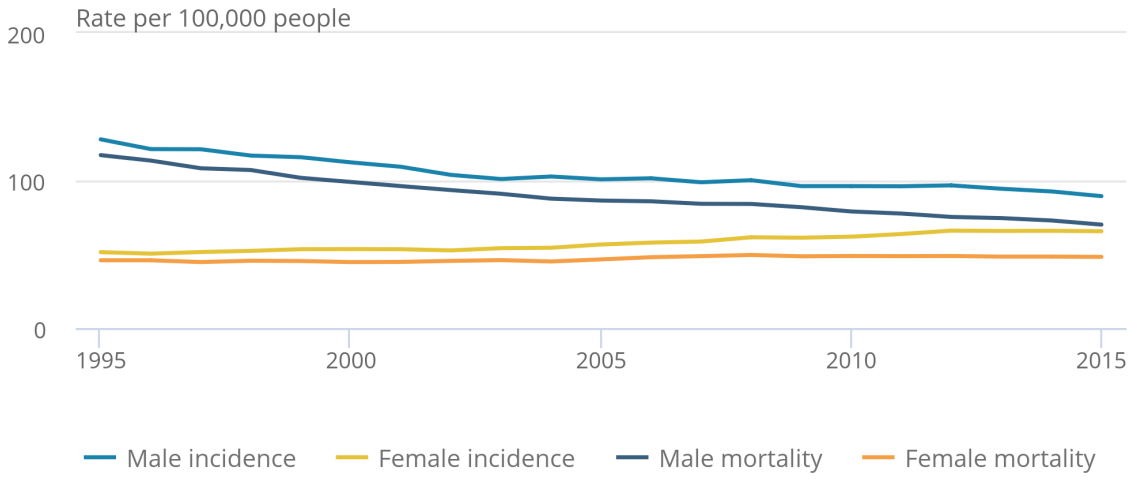


FIGURE 1.7: **Lung cancer incident rates.** Lung cancer incidents per 100,000 persons from 1995 to 2015 in the UK for male and female patients. Figure extracted from Kaur and Poole, 2017.

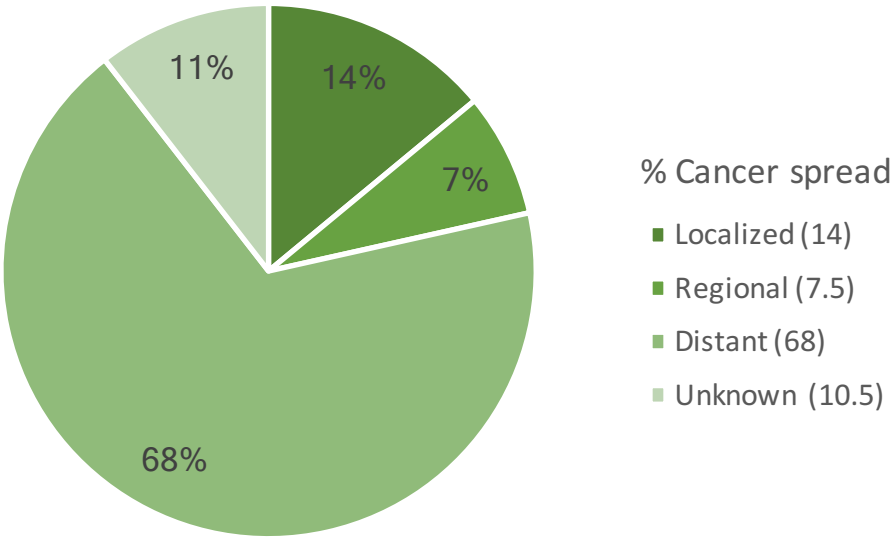


FIGURE 1.8: **Lung cancer diagnosed by stage.** Percentage of cases diagnosed by stage, 2012-2014 and followed up in 2015 in UK. Data sourced from Bannister and Broggio, 2016.

1.3.2 Staging and disease progression in lung cancer

Primary lung cancers are commonly staged and graded according to the TNM staging system. The **T** refers to the size and extent of the tumour, **N** refers to the number of nearby lymph nodes affected by the tumour, and **M** signifies metastasis. Details of the TNM staging system are shown in Tables 1.4, 1.5, 1.6 and Fig 1.9.

T Stage	Sub-stage	Description
Tx	-	Primary tumour cannot be assessed, or tumour proved by the presence of malignant cells in sputum or bronchial washings, but not visualised by imaging or bronchoscopy.
TI	Ia	Tumour is $\leq$2cm in greatest dimension, surrounded by lungs or visceral pleura, without bronchoscopic evidence invasion, more proximal then the lobar bronchus.
	Ib	As Ia but tumour is 2-3cm.
TII	IIa	Tumour $\leq$5cm in greatest dimension. Involves main bronchus, $\geq$2cm distal to the carina. Involves the visceral pleura. Associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung.
	IIb	As IIa but tumour is 5-7cm.
TIII	-	Tumour of any size that directly invades: chest wall, diaphragm, mediastinal pleura, parietal pericardium; or Tumour in the main bronchus <2cm distal to carina, but without involvement of the carina; or associated atelectasis or obstructive pneumonitis of the entire lung.
TIV	-	Tumour of any size that invades any of the following: mediastinum, heart great vessels, trachea, oesophagus, vertebral body, carina; or tumour with a malignant pleural or pericardial effusion, or with satellite tumour nodule(s) within the ipsilateral primary-tumour lobe of the lung.

TABLE 1.4: **Lung cancer staging classification.** Details of the T classification of the TNM staging system and sub sections within, referring specifically to the size and extent of the primary tumour. Adapted from Cancer.org, 2018.

N Stage	Description
Nx	Reginal lymph nodes cannot be assessed
N0	Reginal lymph nodes cannot be assessed
N1	Metastasis to ipsilateral peribronchial and/or ipsilateral hilar lymph nodes, interpulmonary nodes involved by direct extension of the primary tumour
N2	Metastasis to ipsilateral peribronchial and/or ipsilateral hilar lymph nodes, interpulmonary nodes involved by direct extension of the primary tumour
N3	Metastasis to contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s). Details of regional lymph node involvement determined by the N classification of the TNM staging system.

TABLE 1.5: **Lung cancer N stage classification.** Details of regional lymph node involvement determined by the N classification of the TNM staging system. Adapted from Cancer.org, 2018.

M Stage	Description
Mx	Presence of distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis present

TABLE 1.6: **Lung cancer M stage classification.** M classification details of the TNM staging system, determining distant metastasis. Adapted from Cancer.org, 2018.

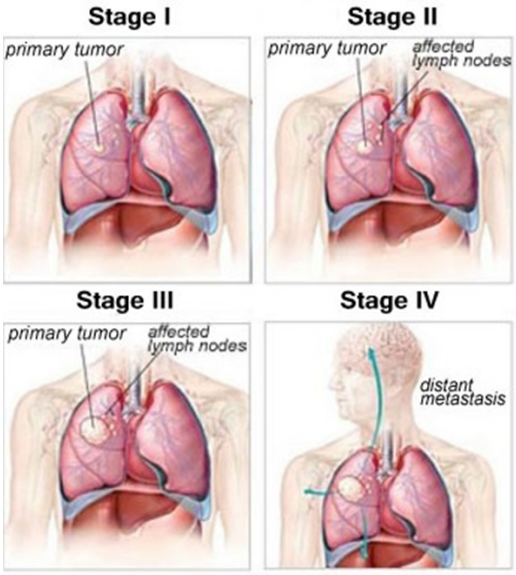


FIGURE 1.9: **Diagrammatic representation of lung cancer stages I – IV.** Where at stage I the primary tumour is small and localised to one lung, stage II depicts growth and spread of the tumour to nearby lymph nodes. Stage III represents further growth of the tumour and spread to additional lymph nodes, by stage IV the tumour is shown to spread to distant organs. Taken from Asbestos.ucoz.net, 2018.

Lung cancers arise from the cells of the respiratory epithelium, and can be separated into two broad categories. Small cell lung cancer (SCLC), a highly malignant neuroendocrine tumour subtype derived from cells exhibiting neuroendocrine characteristics. SCLC account for 15% of all lung cancer cases. The other category being non–small cell lung cancer (NSCLC) (epithelial cancer), which accounts for the remaining 85% of cases, can be further divided into 3 major pathologic subtypes: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Adenocarcinoma itself accounts for 38.5% of all lung cancer cases, squamous cell carcinoma for 20%, while large cell carcinomas covers for 2.9% of all lung cancer cases, with the rest made up from other less common subtypes including, adenosquamous carcinoma and sarcomatoid carcinoma (Herbst, Heymach, and Lippman, 2008; Dela Cruz, Tanoue, and Matthay, 2011). Over the past several decades, incidence of adenocarcinomas has increased greatly, replacing squamous cell carcinoma as the most prevalent type of NSCLC (Dela Cruz, Tanoue, and Matthay, 2011).

There is also a genetic component in the pathogenesis of lung cancer, these have been shown to differ between smokers and non-smokers. Various genetic mutations have been associated with lung cancers (frequency in NSCLCs), including epidermal growth factor receptor (EGFR) (10-35%), KRAS (15-25%), PTEN (4-8%), and anaplastic lymphoma kinase

(ALK) (3-7%), as the most frequently mutated (MyCancerGenome, 2017; Brambilla and Gazdar, 2009).

A lung cancer risk prediction analysis developed by Spitz *et al.* 2007, reported the influence of a family history of cancer increases the risk of developing lung cancer in non-smokers, ex-smokers, and current smokers (Spitz et al., 2007; Spitz et al., 2008).

Multiple genetic, cellular, and local tissue alterations have been known to be involved in a chronic process that can lead to lung carcinogenesis. The transformation of normal cells to pre-neoplastic cells to malignant cells, involves changes including, DNA damages, genetic and epigenetic alterations, and the progression and proliferation of cells, leading to invasion outside of the boundaries of the local lung tissues that is characterised as metastases. Exposure to various carcinogens can alter normal cells before clinically detectable, invasive malignant tumours occur (Dela Cruz, Tanoue, and Matthay, 2011).

1.3.3 Lung cancer survival

Lung cancer is one of the leading causes of death around the world, in America it accounts for more deaths than breast, prostate, and colon cancers combined (Dela Cruz, Tanoue, and Matthay, 2011). Reports show five year survival as low as 18.1% (based on 2007-2013 data), accounting for 25.9% in the USA (2008-2013 data) of all cancer deaths (SEER, 2017). In the UK, survival at 1 year is around the 35% in both genders, and falls to around only 5% survival by ten years (CRUK, 2017), Fig 1.10.

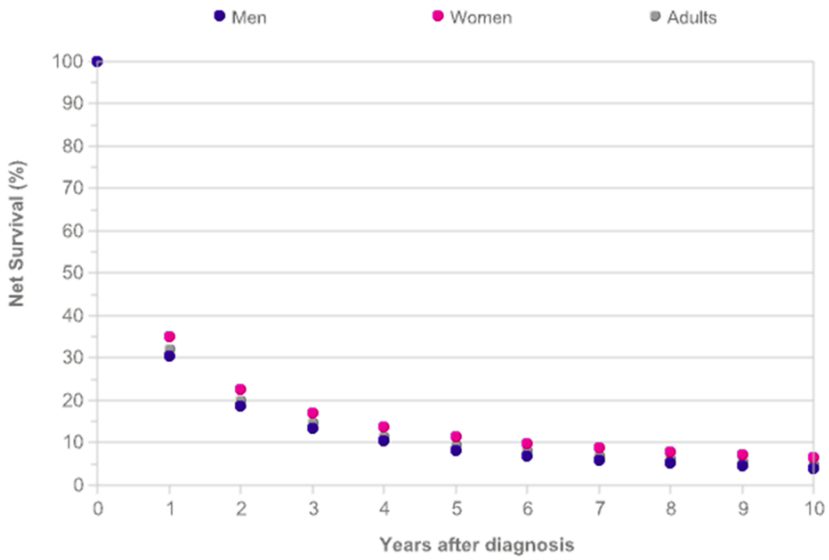


FIGURE 1.10: **Ten year lung cancer survival by gender in the UK.** Taken from (CRUK, 2017).

Once again, survival is dependent to the stage of the cancer, with advanced cases in stage IV having the worst survival rates, of less than 20% in both males and females, Fig 1.11. This is mostly due to the lack or absence of conspicuous symptoms, and consequently a late diagnosis.

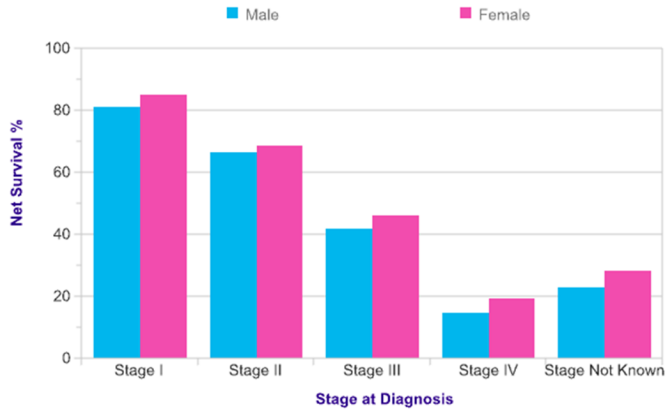


FIGURE 1.11: **Net survival (%) in lung cancer patients by staging, broken down by gender, in the UK.** Taken from (CRUK, 2017).

The lack of robust biomarkers as part of a screening tool, also contribute to the late diagnosis and poor associated outcomes (CRUK, 2017). Currently most lung cancer findings are incidental, due to the lack of conspicuous symptoms. Those patients who present symptoms (respiratory related) are usually in more advanced stages of the disease (Midthun, 2016). Currently most findings of lung cancer are incidental, usually from X-rays and CT scans. The incidence of lung cancer and related mortality can be reduced by: early detection, treatment of disease, chemoprevention, and smoking avoidance and cessation (Kelley and McCrory, 2003).

1.3.4 Lung cancer treatment

Surgery is the gold standard treatment in cancer, followed by radio- and chemotherapy, either independently, if the patient is unfit for surgery or the cancer is too advanced, or combined. Limitations of radiotherapy, and chemotherapy more so, are the toxicity to patients, but also the risk of no response to the treatment. Monitoring response to chemotherapy can be complicated as scans may not necessarily show much change in the tumour in the short term, and tissue biopsies are impractical on a routine basis. Currently, response to chemotherapy is monitored by scans after cycles are completed (with some cycles lasting several months), meaning substantial time lapse before efficiency of the drug can be measured, and the patient is reassessed for other treatment should the current therapy be

ineffective (Housman et al., 2014). Various clinical trials are currently underway, targeting therapies. Table 1.7 lists some of these.

Drug name	**Treatment/Target**
Gefitinib	Receptor tyrosine kinase inhibitor (TKI)
erlotinib	Receptor tyrosine kinase inhibitor (TKI)
Pazopanib	Receptor tyrosine kinase inhibitor (TKI)
Olaparib	Inhibits poly ADP ribose polymerase (PARP)

TABLE 1.7: **Lung cancer targetted therapies.** Lists of some current targeted drug clinical trials for lung cancers within the UK (Clinicaltrials.gov).

Surgery is the only potentially curative treatment for early-stage NSCLC patients (Mitsudomi, Suda, and Yatabe, 2013). Many patients with NSCLC have been cured by surgery; however, 30 to 55% of patients develop recurrence and die of their disease despite curative resection (Kattan et al., 1997; Hoffman, Mauer, and Vokes, 2000; Carnio et al., 2013).

Numerous studies have been conducted to determine causes of recurrence after complete resection. These studies have shown that complete cancer removal needs to be ensured both macroscopically and microscopically (Matsutani et al., 2017; Chudasama et al., 2017a). It has been suggested that there are occult micro-metastatic cancer cells, undetected by standard staging methods, already present systemically in the patients at the time of surgery (Uramoto and Tanaka, 2014). Moreover, handling of the tumour during surgery itself may also lead to the further dissemination of cancer cells (Hayashi et al., 1999; Hashimoto et al., 2014). Reports on these cancer cells, known as circulating tumour cells (CTCs) have been previously described (Pantel, Brakenhoff, and Brandt, 2008; Sienel et al., 2003; Tanaka et al., 2009). Numerous studies have reported an association between the presence of CTCs during tumour resection and patient outcomes (Pantel, Brakenhoff, and Brandt, 2008).

1.4 Metastatic spread

The development of metastatic disease is responsible for 90% of all cancer related deaths (Moore et al., 2008). The general underlying metastatic process is thought to be similar in all cancer types. Once metastasis has been established, prognosis is often very poor in individuals due to the aggressive nature of the disease. The process of metastasis is propagated by cancerous cells, which intravasate to allow lymphohematogenous dissemination to other areas of the body (Fidler, 2003).

The three major routes of tumour cell invasion and formation of tumour metastasis are: i) the lymphatic vessels, ii) the bloodstream, iii) and the transcoelomic (across cavities) spread into the pleural, pericardial, and abdominal cavities (Fidler, 1978). The haematogenous

system is the most common, and regarded as the primary route for distant metastasis. The exact mechanisms are not fully understood. Figure 1.12 shows the cancer cell proliferation leading to metastasis.

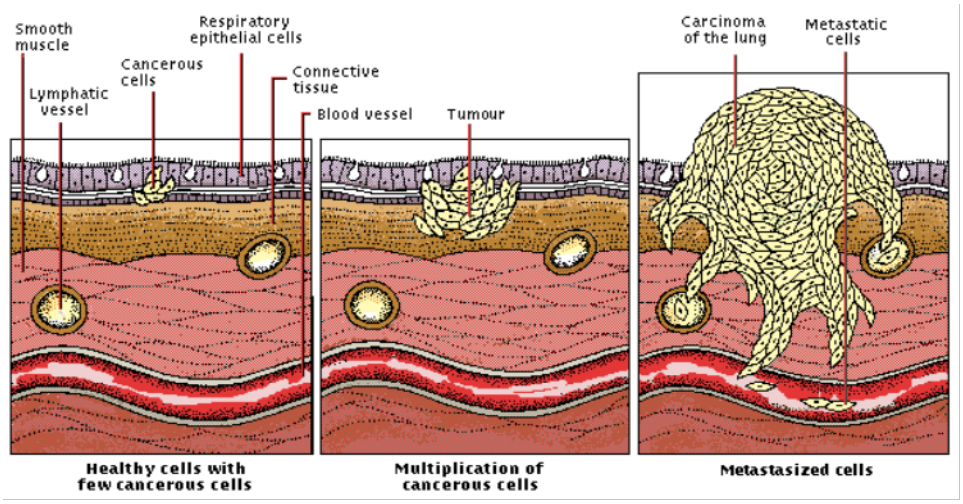


FIGURE 1.12: **Diagrammatic representation of the metastatic process.** Schematic showing general tumour growth leading to metastasis. The initiation of tumour growth can be seen from the proliferation of a few cancerous cells, with growth and development of the tumour cells seen to invade surrounding tissue and enter the lymphatic vessels and blood stream where they are carried to distant sites, leading to metastasis (taken from Knoji.com, 2017).

Tumour cells can also lay dormant in the bone marrow (known as disseminating tumour cells), for a number of years, before they re-enter the bloodstream en-route to a secondary site (Slade et al., 2009). This theory could explain variations in metastasis amongst individuals (e.g. time to metastasis and where secondary tumour growth occurs); as at present it is not well understood what other factors trigger the metastatic pathway.

1.5 Limitations in ovarian and lung cancer contributing to poor survival rates

As highlighted previously, there are several limitations that if by-passed could potentially improve survival rates. In order to achieve this, there is a need for a robust biomarker or genetic signature that could be used for diagnostic and prognostic purposes. Below lists the limitations that could be addressed by this approach:

- Absence of a robust screening tool for early diagnosis resulting in mostly late diagnosis in both cancers.

- Absence of a robust biomarker for routine real-time treatment monitoring. Current measures allow infrequent monitoring to assess efficiency of the treatment, particularly in the case of chemotherapy.
- Absence of a robust biomarker for monitoring of relapse post curative surgery was shown to be an area of concern, with alarmingly high rates of recurrence, owing to the presence of macro and microscopic tumour remnants, known as CTCs.

Most of these limitations exist due to the restrictive use of tissue biopsies; in the next section we discuss developments in liquid biopsies and their role in addressing these shortcomings.

1.6 Liquid biopsy & biomarkers

Recent research around cancer cells in the blood, such as CTCs and ctDNA has generated much interest, opening the potential to a 'Liquid Biopsy' option, as an alternative to a conventional tissue biopsy. Thus, allowing for a better routine monitoring system, but also robust genetic characterisation of the cancer. The concept of tumour and cancer profiling from blood components is an exciting field, opening the potential to solve many issues currently surrounding cancer diagnosis and treatment.

Several limitations have been identified in tissue biopsies compared to blood liquid biomarkers, Table 1.8 details these.

Tissue Biopsy vs Liquid Biopsy	
Tissue Biomarker	**Blood Biomarker**
Can probe many features	Can probe many features
Single location, limited sampling	Tissue volume , full tumour burden sampling possible
Invasive (tissue biopsy)	Non-invasive
Serial assay challenging	Serial assay possible
Widely available central assay	Widely available local and central assay

TABLE 1.8: **Comparison of the use of tissue and liquid biomarkers.**

Several advantages exist with the use of liquid over tissue biomarkers. There are various applications and clinical uses for liquid biopsies, as highlighted above. This section discusses in more detail the potential uses and benefits of a liquid biopsy in these particular areas, to improve prognosis and survival.

The 'liquid biopsy' is appealing due to its non-invasive nature, and by comparison with conventional biopsies is also much more cost effective. Furthermore, blood analysis allows

for routine sampling on even a daily basis, where this would be impractical with a tissue biopsy, which may be obtainable 2-3 times a year, particularly in the case of ovarian and lung cancers.

This study looks to evaluate various blood based indicators, in an attempt to identify a robust 'liquid biopsy' marker for use in cancer clinical practice.

1.6.1 Circulating tumour cells

The general definition of CTCs remains universal, as cancer cells of epithelial origin that have detached from the primary tumour, or a metastatic site, and entered the blood circulation (Allard et al., 2004; Hou et al., 2012; Park et al., 2012). Some CTCs are thought to arise as a consequence of passive tumour cell shedding. However, a more complex epithelial-to-mesenchymal transition has also been described, whereby cells lose their differentiated epithelial characteristics and acquire mesenchymal features including motility, invasiveness, and a resistance to apoptosis (Polyak and Weinberg, 2009; Munzone et al., 2012). Nevertheless the cells are shed from the tumour and express tumour-specific characteristics (Hiltermann, Wekken, and Groen, 2012). Most CTCs that detach from a primary tumour die; it has been suggested by pre-clinical models that only as few as 0.01% will give rise to metastasis.

The passage of CTCs in peripheral blood is key to the genesis of distant metastases in various cancers and hence crucial to patient outcomes. Fig 1.13 details the concept of CTCs.

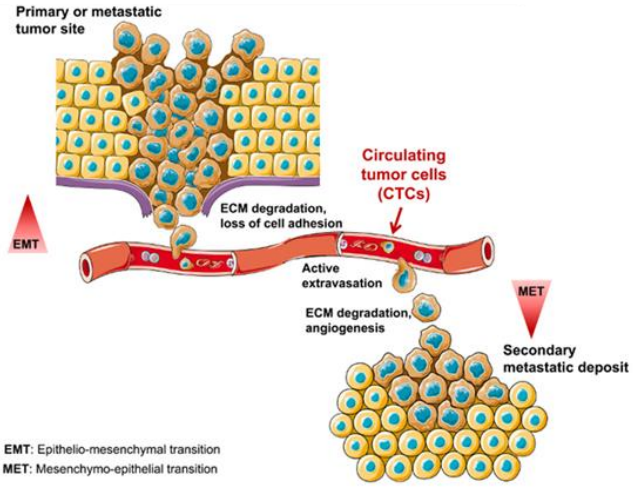


FIGURE 1.13: **CTCs entering the blood circulation.** Diagrammatic representation of the process of CTCs entering the general blood circulation, with highlights on the role of EMT and MET in the spread of cancer (Krzeszinski and Wan, 2015).

CTCs have long been hypothesised as being heavily implicated with metastatic disease (Hiltermann, Wekken, and Groen, 2012), with the haematogenous spread of carcinoma resulting in incurable metastasis, yet the basic characteristics of cancer cells in the bloodstream are relatively unknown, including their ability to evade apoptosis, and mechanisms that influence and facilitate migration, particularly to specific secondary sites (i.e. ovarian metastasis occurs mostly in the bowel, bladder, and abdomen, whereas lung cancers commonly metastasizes to the liver, brain, and bones) (Hong, Fang, and Zhang, 2016; Dalum, Holland, and Terstappen, 2012; Valastyan and Weinberg, 2011). Previous studies have detected CTCs in both ovarian and lung cancers, reporting positive findings and associations with prognosis (Normanno et al., 2016; Hamilton and Rath, 2016; Nurwidya et al., 2016; Zhou et al., 2015; Kolostova et al., 2016; Romero-Laorden et al., 2014).

Detection of potentially small populations of CTCs within the large number of normal blood cells represent a significant technical challenge (Hou et al., 2013; Dent et al., 2016). Isolation of CTCs has been attempted by using physical features such as their larger size and weight. Microfluidic techniques, where spaces and flows are commensurate with a scale of single cells, allowing CTCs to be captured (Autebert et al., 2012). Table 1.9 lists some CTC isolation platforms.

Lable-Free		**Affinity based**	
Function	**Platform**	**Function**	**Platform**
Gradient Based	Ficoll Paque	Immunomagnetic	CellSearch
	RareCyte		Adna Test
			MACs
Size Based	Circulogix	Microfludic	CTC-chip
	ISET		GEDI Chip
	ScreenCell		OncoCEE
	Parsotix		Clearbridge
Dielectrophoresis	ApoStream	Surface Based	Herringbone Chip
	DEPArray		Graphene oxide
			ImageStream

TABLE 1.9: **CTC isolation platforms.** Details of some existing CTC isolation platforms grouped in lable-free and affinity based classes.

However, the physical features of some CTCs overlap with some leucocytes, which may compromise accurate detection. Efforts to purify or enrich the CTC populations rely on surface proteins, such as EpCAM or MUC1, and ultimately antibody specificity. However, the EMT process has been shown to down-regulate and alter many surface proteins, particularly EpCAM expression has been shown to be lost in several cancers including lung (Hyun

et al., 2016; Polyak and Weinberg, 2009). Moreover, the frequency of CTCs with mesenchymal and/or EMT markers have been shown to rise in breast cancer patients who are resistant to treatment (Aktas et al., 2009) and in those with more advanced disease compared to localised cancers (Kallergi et al., 2011). More recently, a new High Definition CTC (HD-CTC) assay was introduced as a fluid biopsy approach that identifies CTCs without using surface protein-based enrichment. This method presents CTCs in sufficiently high definition (HD) to satisfy diagnostic pathology image quality requirements (Marrinucci et al., 2012; Phillips et al., 2012).

The current literature supports the clinical utility of CTCs, and there is increasing evidence that CTCs can be used as predictive markers for diagnosis, prognosis, and response to treatment (Fidler, 2003; Hou et al., 2011; Phillips et al., 2012; Lecharpentier et al., 2011).

1.6.2 Circulating nucleic acids

Another form of 'liquid biopsy' are circulating nucleic acids (ctNAs), these include DNA (ctDNA) and RNA. The process of ctNAs shedding is not well understood. A schematic of this process is depicted in Fig 1.14.

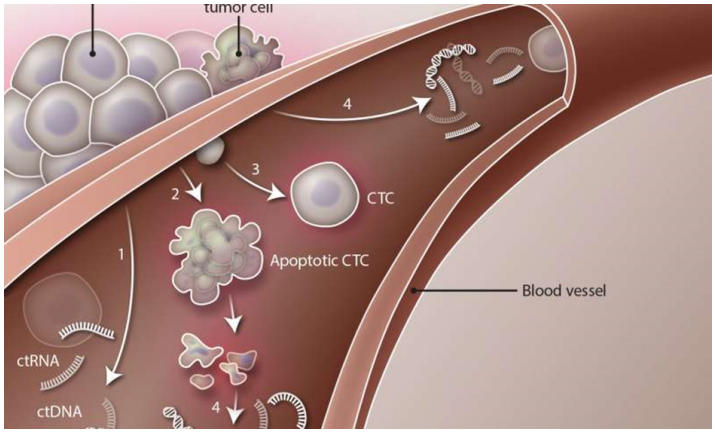


FIGURE 1.14: **Source and process of some circulating liquid biomarkers.** Source and process of ctNAs, tumour DNA & RNA and CTCs entering the blood circulation. Taken from Medicalexpress, 2017.

ctDNA is tumour derived and originating from the cancer. The exact nature of ctDNA shedding like CTCs is unknown, but it is postulated to occur due to apoptosis and necrosis of cancer cells, in addition to active release from viable tumour cells. Apoptosis of normal cells results in larger uniformly truncated cfDNA fragments, in contrast to ctDNA, whereby cancerous cells undergo necrosis and mitotic catastrophe resulting in smaller fragmented DNA (Huang et al., 2016).

Cell-free nucleic acids were first discovered in plasma samples by Mandel and Metais, 1948, but only much later in 1987, Stroun *et al.* determined that ctDNA in cancer patients' plasma had originated from tumour cells (Stroun et al., 1987). Currently, ctDNA have been detected in several bodily fluids, such as blood, urine, stools, milk, bronchial lavages and ascites.

ctDNA half-life has been estimated at about 16 min (Fournié et al., 1995; Diehl et al., 2008; Lo et al., 1999), suggesting that ctDNA is not 'naked' but rather complexed with cellular or non-cellular components. ctDNA physico-chemical characteristics are poorly documented but it might be associated with cell membrane parts, specific or non-specific DNA-binding proteins (Kawamura, Paschoal, and Carvalho, 2000), apoptotic bodies (Nagata, 2000) or multi-nucleosome complexes (Holdenrieder et al., 2008a; Deligezer et al., 2006). Discrepancies about ctDNA size in serum/plasma exist in the literature certainly due to technical limitations. ctDNA size was found to range from 500 bp to >30 kb (Dennin, 1979; Giacona et al., 1998; Chan et al., 2004); however, recent studies described ctDNA fragments as smaller than 250 bp, whereas cfDNA released from normal cells are believed to be larger (Jahr et al., 2001; Diehl et al., 2008; Diehl et al., 2008). The size distribution of ctDNA fragments within the same plasma/serum sample has been poorly studied (Chan et al., 2004; Fan et al., 2008).

The genomic profiling of ctDNA has become increasingly popular, based on the hypothesis that ctDNA harbours the same profile of somatic mutations and genomic rearrangements as the tumour itself (Freidin et al., 2015). As cancer cells frequently rely on the activation of oncogenes for proliferation and survival, the presence of these specific gene alterations and mutations can be of diagnostic value, reflect patient's responsiveness to the treatment and therapies, and predict survival (Gold et al., 2015). At present, advanced technologies have turned precise ctDNA mutation detection to reality (Patel and Tsui, 2015; Thierry et al., 2010), including the FDA approved Cobas kits for EGFR mutation analysis.

1.7 Applications of liquid biopsy

1.7.1 Diagnosis

As discussed earlier, late diagnosis in lung and ovarian cancers contributes to the poor survival rates. We know this is due to the lack of obvious symptoms and the absence of any robust screening tool.

Cancer screening programmes are fundamental to support early diagnosis, and have been seen to dramatically improve breast, cervical, and prostate cancer survival rates. A recent study evaluated the breast cancer screening programme, concluding that the screening

programme significantly reduced mortality rates in breast cancers, and the risks of population screening were significantly outweighed by the benefits (Johns et al., 2017).

In breast cancers, mammograms (breast's X-rays) can be suitably used for cancer diagnosis, as is the detection of abnormal cells in cervical screening tests, and identification of the prostate serum antigen (PSA) in prostate screening tests. However, there is an absence of a robust biomarker in both ovarian and lung cancers.

Characterisation and quantification of CTCs and ctNAs (including ctDNA) could be exploited as potential diagnostic markers. Several studies have evaluated and reported the diagnostic capabilities of CTCs (Ankeny et al., 2016; Ilie et al., 2014; Wu et al., 2016; Cen et al., 2012; Liu et al., 2013), and ctDNA (Vidal et al., 2017; Fernandez-Cuesta et al., 2016; Cheng, Su, and Qian, 2016) in various cancers, however none of these have been implemented into clinical practice due to limitations in technologies. These limitations include, low sensitivity and specificity, and poor correlation with clinical characteristics in some cases (Liu et al., 2013).

As liquid biopsy testing is simple and non-invasive for patients, it could be potentially offered at a primary care site, e.g. general practitioners (GP) surgeries, walk in clinics etc., avoiding thus long NHS waiting times for tertiary or specialist hospital appointments. The test could be patient stratified, and offered to those who are at high risk, pertaining to age, life-style, family history, etc.

1.7.2 Predicting and treatment planning

Clinicians are always faced with the challenge of treatment planning for cancer patients. It is well known that all patients' cancers are individual and will respond differently to therapies and drugs, this phenomenon is defined as inter-patient variability. Even the tumour itself is known to be heterogeneous and constantly evolving and changing its biological and genetic make up, increasing the disease complexity (Gay, Baker, and Graham, 2016).

With the emergence and continual development of target specific drugs, also known as targeted therapies, there is a requirement to identify potential candidate patients who would benefit from these, over broad conventional chemotherapy or therapeutics.

Target therapies are drugs designed to target particular genes and mutations known to drive or be involved in the development and growth of the tumour. Specific examples include: Gefitinib, used in non-small cell lung cancers (NSCLC) acting as a tyrosine kinase inhibitor (TKI), Trastuzumab in human epidermal growth factor receptor 2 (HER2) breast cancer positive patients, Bevacizumab (Avastin) targeting the VEGF ligand in various cancers, including ovarian, NSCLC, colorectal, and cervical cancers. In many cases these drugs have been shown to be more effective than conventional chemotherapy.

Recent advances in the analysis of CTCs and circulating tumour nucleic acids (ctNAs) offer the prospect of better routine and real-time monitoring of the patients cancer, with the potential of identifying changes in genetics more rapidly, informing thus of appropriate treatment options. Furthermore, robust sensitive and specific assays could allow for the characterisation and mutation profiling of patients offering the possibility of a truly personalised treatment plan.

1.7.3 Monitoring treatment response

Currently, when surgery is not an option, patients are offered chemotherapy and radiotherapy as alternatives. These therapeutic options come with risks, such as the attack of healthy cells, side effects, and even failure to treat the cancer itself. Monitoring this is currently done by means of CT scans and X-rays at the end of a cycle of treatment, which could mean several months may go by before the efficacy of the approach is assessed. Time is incredibly critical for cancer patients, particularly those in advanced cases (Graham et al., 2014; Thoeny and Ross, 2010).

Routine and real-time monitoring during therapies, of CTCs and ctNAs could inform of patients' response and allow more rapid decisions to be made. Routine mutational profiling could also measure changes in the cancers' genetics, providing valuable information that could lead to clinically relevant treatment decisions being taken.

1.7.4 Monitoring relapse

Cancer relapse, as mentioned earlier, contributes to the high mortality rates in the ovarian and lung cancers. Despite curative surgery, there is a risk of cancer recurrence. This has been reported to occur due to ineffective removal of all the tumour, and also manipulation of the cancer leading to further cancer cells being shed, entering the general circulation, and going on to seed and develop in new sites (Hayashi et al., 1999; Hashimoto et al., 2014; Pantel, Brakenhoff, and Brandt, 2008).

For this reason, monitoring CTCs and ctNAs, directly post-surgery and thereafter on a regular basis could inform of presence of cancer cells in the blood, and the potential of relapse, thus enabling rapid clinical decisions on further intervention or treatment if required. Currently, patients have a CT scan every 6 months post-surgery, which may not necessarily pick up small nodules. Moreover, studying the mechanisms and biology of CTCs and ctNAs in this context may offer more insight in the process of metastasis and relapse.

1.7.5 Identifying novel and important gene signatures

Mutational and marker analysis of CTCs and ctNAs could have huge potential benefits in identifying novel liquid biomarkers and gene signatures in cancer. The biomarker discovery is also important in unlocking vital information on cancer biology, including critical associated pathways and mechanisms currently unknown.

Ultimately, the application of CTCs and ctNAs could sustainably support and reform current clinical practice in the management and treatment of cancers. All the mentioned liquid biomarkers have been shown to be clinically relevant. Application in the clinical setting of these biomarkers requires the assays to be robust, sensitive and specific, and cost-effective. Overcoming these obstacles could see a potentially revolutionary biomarker enter the clinical setting.

Potential benefits of liquid biopsies:

- Liquid biopsies could overcome the sampling bias that results from tissue heterogeneity.
- Liquid biopsies have the potential of early diagnosis.
- Liquid biopsies could provide early insights to response to treatment.
- Liquid biopsies have the potential of informing of early resistance to drugs.
- Liquid biopsies provide the opportunity of serial real-time monitoring of patients.
- Liquid biopsies have the potential to monitor and inform of possible relapses following curative surgery/treatment.

Precision medicine is the goal of the future, where optimal individual patient treatment is determined by means of in-depth analysis of multiple materials and markers from the patient. Fig 1.15 describes a proposed integrated profiling of cancer to define patient treatment.

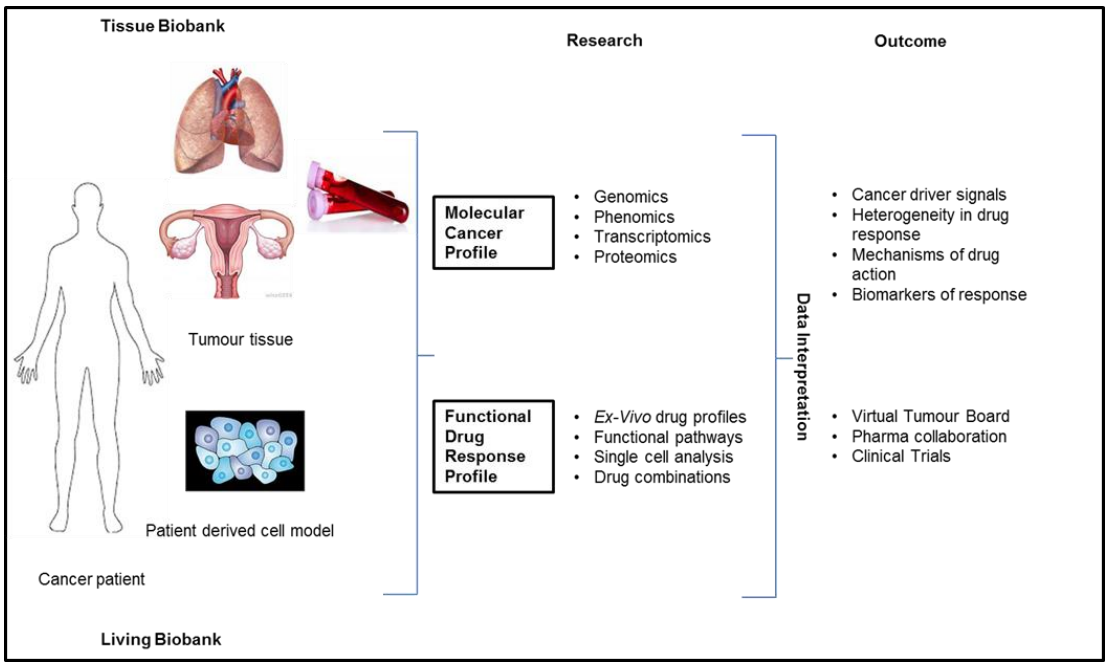


FIGURE 1.15: **Integrated profiling of cancer.** Pictorial representation of an integrated profiling of cancer to define patient treatment, as a form of precision medicine.

1.7.6 Aims and objectives

This study aims to evaluate the use of blood based liquid biomarkers to address current limitations in the diagnosis, treatment, and management of ovarian and lung cancers as identified earlier, ultimately in a bid to increase survival in these two diseases. This will include identifying one or several robust liquid biomarkers and proposing a pipeline of tests to be implemented into the clinical setting, addressing in detail the following objectives:

- The ability to detect CTCs in the blood of lung and ovarian cancer patients using the Imagestream^{TM} and Clearbridge^{TM} devices, and it's clinical utility.
- The clinical relevance of detecting ctDNA in the blood of lung and ovarian cancer patients for diagnostic and prognostic purposes.
- Evaluate the use of available microarray data to identify clinically relevant liquid biomarkers in the blood of lung and ovarian cancers.
- Evaluate the ability to identify clinical relevant biomarkers and gene signatures by means of RNAseq data interrogation and chromosomal analysis.

Chapter 2

Methodology

2.1 Tissue culture

2.1.1 Cell lines

SKOV-3 (clear cell adenocarcinoma), MDAH-2774 (endometrioid adenocarcinoma), A549 (non-small cell adenocarcinoma) and NCI-H358 (non-small cell lung bronchialveolar carcinoma) human adherent epithelial cell were used as *in vitro* models of human ovarian cancer, and human lung cancer (latter 2) respectively. Cell line details are summarised in Table 2.1.

Cell line	Origin	Grade	Key gene mutations
SKOV-3	Ascites	1/2	TP53
			PIK3CA
MDAH-2774	Ascites	2	TP53
			PIK3CA
			KRAS
			BRCA1 (silent)
			BRCA2 (silent)
A549	Lung	1	RAS
			CDKN2A
NCI-H358	Lung/bronchiole	1	KRAS

TABLE 2.1: **Cell lines.** Characteristics of SKOV-3, MDAH-2774, A549, and NCI-H358 cell lines.

A549, SKOV-3, MDAH-2774, and NCI-H358 cells were grown in DMEM (Dulbecco's Modified Eagles Medium, Gibco) supplemented with 10% FBS (fetal bovine serum, Gibco), 1% penicillinstreptomycin (Gibco) and 1% L-glutamine (Gibco). All cell lines were cultured at 37°C 5% CO_2 and subcultured when approaching 80% confluency, approximately three times a week.

2.1.2 Tissue culture practice

An aseptic environment was maintained with the use of a HERAsafe laminar flow cabinet (Heraeus), and repeated applications of 2% TriGene Advance (Medimark Scientific) and 70% industrial methylated spirits (IMS) in dH_2O to all surfaces and equipment used. Commercial pre-packed and pre-sterilised items were used, where possible. All plasticware was sterilised by autoclave prior to use.

2.1.3 Thawing cryopreserved cells

When growing cells from stored frozen liquid nitrogen stocks, 20mL of media was first stored in a T-75 cell culture flask in the incubator for at least one hour to allow the media to equilibrate to the temperature and CO_2 conditions. After which cells were moved with speed from liquid nitrogen, defrosted in a water bath at 37°C, and immediately transferred to the pre-warmed media which was then left undisturbed in the incubator for 12 hours. At 12 hours the media was replaced with fresh media that had been warmed to 37°C, removing the high concentration of DMSO remaining from the freezing medium and replenishing the nutrients.

2.1.4 Subculturing cells

Cells were subcultured at approximately 80% confluency by aspirating the media, incubating with 2.5mL TrypLE^{TM} Express (Invitrogen) per about 75cm^2 growth surface area and manually disturbing the flask to detach adherent cells. The detached cells were then suspended in 17.5mL of their appropriate media and 5mL transferred to each of four new T-75 cell culture flasks already containing 15mL of media. All media used in the subculturing process was first warmed to 37°C. All cell lines required a 1 in 4 split approximately three times per week.

2.1.5 Cryopreserving cells

Stocks of each cell line were stored in liquid nitrogen. This was done by aspirating the media and detaching the cells by incubating them for two minutes with 2.5mL TrypLE^{TM} Express (Invitrogen), per $\sim$75cm^2 surface area, and manually disturbing the flask to detach adherent cells. The cells were resuspended in 4mL of media, the suspension moved to a 12mL centrifuge tube and centrifuged for five minutes at 1,500 RPM to form a cell pellet. The supernatant was removed and 0.5mL of FBS and 0.4mL of media was used to resuspended the cells. The cell suspension with 100μl DMSO was added to a cryovial and cooled slowly, first to -80°C in a Mr. Frosty freezing container (Nalgene) and then in liquid nitrogen 24 hours later.

2.1.6 Seeding cells

Cells were seeded in 24 and 6-well plates at a specific seeding density. Cell counts to determine seeding requirements were performed with a Countess® Automated Cell Counter (Invitrogen) using an equal volume of Trypan Blue (0.4%) stain to cell suspension for cell viability. The seeding volume was calculated as follows:

Total number of cells = Viable cells per mL $\times$ Volume (in mL) of suspension
Number of cells per μL = Total number of cells/Volume (in μL) of suspension
Volume of suspension to add to well = Seeding density required/Number of cells per μL

Cells were allowed to proliferate for 24 hours before processing.

2.2 Clinical samples

2.2.1 Ethical approval process

In order to obtain clinical material from patients, firstly I sought ethical approval via the Human Tissue Act (HTA) and National Research Ethical Authority (NRES). Ethical approval was already in place for the collection of Ovarian blood and tissue samples, under the METRO-BIBF clinical trials ethical consent. Ethical approval was sought for collection of lung blood and tissue samples from The Royal Brompton & Harefield NHS Hospital Trust. This involved completion of an online NRES ethical approval form and production of patient specific documents, followed by attendance at a panel review meeting to answer questions from health professional and members of the public. I was notified a week later that our application had been successful, and would run under the following reference: 14/LO/1284. Material transfer agreements (MTA) were also put in place for transfer of ovarian and lung tissue from their respective sites to Brunel University.

2.2.2 Fresh ovarian tissue

Ovarian clinical tissue samples from patients with ovarian and endometrial cancer, as well as unaffected female controls were obtained from the University of Thessanoliki, Greece (ethical approval obtained by the local hospital authority and by Brunel University ethics committee, Table 2.2). Samples were shipped on dry ice in RNAlater® (Life Technologies) to preserve RNA integrity.

	Stage/Grade		
Pathology	**2/IIIC**	**3/IIIC**	**3/IV**
Serous adenocarcinoma	2	7	1
Endometriod	1	-	-

TABLE 2.2: **Ovarian cancer samples.** Table detailing clinical samples, including pathology, staging and grade of cancer.

2.2.3 Fresh lung tissue

Lung tissue samples were collected from The Royal Brompton & Harefield NHS Trust, from patients undergoing surgical resection of their known lung cancer, tumour tissue and normal tissue was collected for these patients. Ethical approval was sought prior and all patients consented to the study. All tissue was resected by the surgeon, and immediately snap frozen in liquid nitrogen and stored at -80°C, until further use. Sample information is shown in Table 2.3.

		Stage				
Pathology		**I**	**II**	**III**	**IV**	**Un-staged**
Non-small cell lung cancer	Adenocarcinoma	5	6	1	1	-
	Squamous Cell Carcinoma	-	5	-	2	-
	Metastatic lung cancer	-	-	-	-	5
Controls	Bullectomy					5

TABLE 2.3: **Lung cancer samples.** Table detailing clinical samples, including pathology and staging of cancer.

2.2.4 Ovarian, breast, and lung cancer, and normal tissue cDNA array plate (Origen^{TM})

An array plate containing lyophilised ovarian, breast and lung tissue cDNA for cancer and control patients was also purchased from Origene^{TM}. The kit also contained kidney cancer samples, that were subsequently discarded. Details of samples are listed in Table 2.4.

Position	Gender	Age	Tissue	Type	Diagnosis	Stage
D01	Female	47	Brest	Normal	Adenocarcinoma of breast, ductal	Normal
D02	Female	52	Brest	Tumour	Adenocarcinoma of breast, ductal	IIA

D03	NA	78	Brest	Tumour	Adenocarcinoma of breast, ductal	IIA
D04	Female	50	Brest	Tumour	Adenocarcinoma of breast, ductal	IIB
D05	Female	55	Brest	Tumour	Adenocarcinoma of breast, ductal	IIIC
D06	Female	68	Brest	Tumour	Adenocarcinoma of breast, ductal	IIIC
D07	Male	77	Kidney	Normal	Carcinoma of kidney, renal cell, clear cell	Normal
D08	Male	43	Kidney	Tumour	Carcinoma of kidney, renal cell, papillary	I
D09	Male	64	Kidney	Tumour	Carcinoma of kidney, renal cell, papillary	II
D10	Male	4	Kidney	Tumour	Nephroblastoma	II
D11	Male	59	Kidney	Tumour	Carcinoma of kidney, renal cell, papillary	II
D12	Male	59	Kidney	Tumour	Carcinoma of kidney, renal cell, clear cell	III
E01	Female	72	Lung	Normal	Adenocarcinoma of lung	Normal
E02	Female	82	Lung	Tumour	Adenocarcinoma of lung	IA
E03	Male	65	Lung	Tumour	Carcinoma of lung, squamous cell	IIB
E04	Male	76	Lung	Tumour	Carcinoma of lung, squamous cell	IIIA
E05	Female	46	Lung	Tumour	Adenocarcinoma of lung	IV
E06	Female	64	Lung	Tumour	Tumor of lung, carcinoid, atypical	Unknown
E07	Female	54	Ovary	Normal	Adenocarcinoma of ovary, clear cell	Normal
E08	Female	58	Ovary:	Tumour	Adenocarcinoma of ovary, endometrioid	IIB
E09	Female	91	Ovary	Tumour	Adenocarcinoma of ovary, papillary serous	III
E10	Female	49	Ovary	Tumour	Tumor of ovary, mucinous, borderline	IIIA
E11	Female	47	Ovary	Tumour	Adenocarcinoma of ovary, clear cell	IIIC

E12	Female	68	Ovary	Tumour	Adenocarcinoma of ovary, serous	IV

TABLE 2.4: **Information on cDNA from tissue samples, purchased from Origene^{TM}.** Clinical details of samples provided in the commercial kit.

2.2.5 Fresh frozen paraffin embedded blocks of ovarian and lung tissue

Matching fresh frozen paraffin embedded (FFPE) blocks for ovarian and lung patients from Mount Vernon hospital and The Royal Brompton & Harefield NHS Trust respectively, were provided. Blocks were cut using a microtome at 5μm and 15μm for microscope slides and scrolls for RNA extraction respectively. Slides were stored at room temperature and scrolls at -20°C until use. Clinical tissue arrays containing paraffin embedded ovarian and lung cancer and normal samples were purchased from US Biomax (Tables 2.5, 2.6, and 2.7 respectively).

Locus	**Sex**	**Age**	**Organ**	**Pathology**	**Grade**	**Stage**
A1	F	40	Ovary	Clear cell carcinoma	-	I
A2	F	57	Ovary	Serous papillary carcinoma	II	Ic
A3	F	48	Ovary	Clear cell carcinoma	-	II
A4	F	57	Ovary	Serous papillary carcinoma	II	IIIc
A5	F	43	Ovary	Serous papillary carcinoma	III	IIIc
A6	F	54	Ovary	Serous papillary carcinoma	I	Ic
A7	F	63	Ovary	Serous papillary carcinoma	III	IV
A8	F	46	Ovary	Serous papillary carcinoma	III	IIIc
A9	F	54	Ovary	Serous papillary carcinoma	III	IIIc
A10	F	56	Ovary	Hyperplastic fibrous tissue	-	-
B1	F	44	Ovary	Granular cell tumor	-	-
B2	F	49	Ovary	Serous papillary carcinoma	III	II
B3	F	18	Ovary	Immature teratoma	-	-
B4	F	15	Ovary	Endodermal sinus carcinoma	-	IIa
B5	F	38	Ovary	Metastatic adenocarcinoma	III	-
B6	F	39	Ovary	Serous papillary carcinoma	III	IV
B7	F	24	Ovary	Endodermal sinus carcinoma	-	II
B8	F	42	Ovary	Serous papillary carcinoma	III	II
B9	F	50	Ovary	Serous papillary carcinoma	III	I
B10	F	49	Ovary	Serous papillary carcinoma	III	IIIc
C1	F	62	Ovary	Serous papillary carcinoma	III	II
C2	F	53	Ovary	Mucinous papillary carcinoma	I	IV
C3	F	38	Ovary	Metastatic adenocarcinoma	III	-

C4	F	43	Ovary	Clear cell carcinoma	-	Ia
C5	F	26	Ovary	Serous papillary carcinoma	I	Ic
C6	F	47	Ovary	Serous papillary carcinoma	I	I
C7	F	62	Ovary	Squamous cell carcinoma	III	I
C8	F	35	Ovary	Dysgerminoma	-	Ia
C9	F	41	Ovary	Dysgerminoma	-	I
C10	F	47	Ovary	Serous papillary carcinoma	III	I
D1	F	42	Ovary	Clear cell carcinoma	-	Ic
D2	F	39	Ovary	Metastatic adenocarcinoma	I	-
D3	F	66	Ovary	Metastatic adenocarcinoma	III	-
D4	F	48	Ovary	Malignant follicular theca cytoma	-	III
D5	F	51	Ovary	Serous papillary carcinoma	I	IIIc
D6	F	33	Ovary	Metastatic signet-ring cell carcinoma	-	
D7	F	18	Ovary	Mixed germ cell tumor	-	Ib
D8	F	40	Ovary	Metastatic signet-ring cell carcinoma	-	-
D9	F	43	Ovary	Granular cell tumor	-	-
D10	F	55	Ovary	Serous papillary carcinoma	III	II
E1	F	46	Ovary	Serous papillary carcinoma	III	IIIc
E2	F	57	Ovary	Serous papillary carcinoma	III	IIIc
E3	F	75	Ovary	Serous papillary carcinoma	III	IIIc
E4	F	69	Ovary	Serous papillary carcinoma (sparse)	III	Ia
E5	F	30	Ovary	Serous papillary carcinoma	III	I
E6	F	42	Ovary	Serous papillary carcinoma	II	IIIc
E7	F	48	Ovary	Clear cell carcinoma	-	I
E8	F	22	Ovary	Serous papillary carcinoma	I	IIb
E9	F	50	Ovary	Clear cell carcinoma	-	I
E10	F	32	Ovary	Serous papillary carcinoma	II	I
F1	F	48	Ovary	Serous papillary carcinoma	II	I
F2	F	50	Ovary	Serous papillary carcinoma	II	II
F3	F	65	Ovary	Serous papillary carcinoma	III	IIIc
F4	F	38	Ovary	Serous papillary carcinoma	III	IIIc
F5	F	31	Ovary	Metastatic adenocarcinoma	III	-
F6	F	55	Ovary	Metastatic adenocarcinoma	III	-
F7	F	51	Ovary	Serous papillary carcinoma	III	II
F8	F	65	Ovary	Serous papillary carcinoma	II	I
F9	F	26	Ovary	Serous papillary carcinoma	II	IIIc
F10	F	55	Ovary	Serous papillary carcinoma	II	I
G1	F	49	Ovary	Serous papillary carcinoma	III	II
G2	F	48	Ovary	Metastatic signet-ring cell carcinoma	-	-

G3	F	46	Ovary	Serous papillary carcinoma	III	IIIc
G4	F	63	Ovary	Serous papillary carcinoma	III	II
G5	F	37	Ovary	Serous papillary carcinoma	II	IV
G6	F	35	Ovary	Malignant tumour (sparse)	-	IIIc
G7	F	12	Ovary	Dysgerminoma	-	Ib
G8	F	55	Ovary	Serous papillary carcinoma	II	I
G9	F	20	Ovary	Malignant tumour cell (sparse)	-	I
G10	F	55	Ovary	Serous papillary carcinoma	II	I
H1	F	40	Ovary	Cancer adjacent normal ovarian tissue of No. 01	-	-
H2	F	22	Ovary	Cancer adjacent normal ovarian tissue	-	-
H3	F	30	Ovary	Cancer adjacent normal ovarian tissue	-	-
H4	F	39	Ovary	Cancer adjacent normal ovarian tissue	-	-
H5	F	32	Ovary	Cancer adjacent normal ovarian tissue of No. 50	-	-
H6	F	63	Ovary	Normal ovarian tissue	-	-
H7	F	17	Ovary	Normal ovarian tissue	-	-
H8	F	29	Ovary	Normal ovarian tissue	-	-
H9	F	14	Ovary	Normal ovarian tissue	-	-
H10	F	20	Ovary	Normal ovarian tissue	-	-

TABLE 2.5: **Information on paraffin embedded tissue of 80 ovarian cancer and normal tissues, from the OV802 array chip.** Clinical details of FFPE samples of ovarian cancer patients.

Locus	**Sex**	**Age**	**Organ**	**Pathology**	**Grade**
A1	F	40	Ovary	Endometrioid adenocarcinoma	II
A2	F	40	Ovary	Necrosis tissue	-
A3	F	57	Ovary	Serous papillary cystadenocarcinoma	II
A4	F	57	Ovary	Serous papillary cystadenocarcinoma	II
A5	F	48	Ovary	Clear cell carcinoma	-
A6	F	48	Ovary	Clear cell carcinoma	-
B1	F	57	Ovary	Serous papillary cystadenocarcinoma	II
B2	F	57	Ovary	Serous papillary cystadenocarcinoma	II
B3	F	46	Ovary	Fibrofatty tissue	-
B4	F	46	Ovary	Serous papillary cystadenocarcinoma	III
B5	F	42	Ovary	Endometrioid adenocarcinoma	I
B6	F	42	Ovary	Endometrioid adenocarcinoma	I
C1	F	66	Ovary	Serous papillary cystadenocarcinoma	III
C2	F	66	Ovary	Serous papillary cystadenocarcinoma	III

C3	F	53	Ovary	Serous papillary cystadenocarcinoma	III
C4	F	53	Ovary	Serous papillary cystadenocarcinoma	III
C5	F	35	Ovary	Serous papillary cystadenocarcinoma	II
C6	F	35	Ovary	Serous papillary cystadenocarcinoma	II
D1	F	52	Ovary	Serous papillary cystadenocarcinoma	III
D2	F	52	Ovary	Serous papillary cystadenocarcinoma	III
D3	F	41	Ovary	Cancer adjacent ovary tissue	-
D4	F	41	Ovary	Cancer adjacent ovary tissue	-
D5	F	49	Ovary	Cancer adjacent ovary tissue	-
D6	F	49	Ovary	Cancer adjacent ovary tissue	-

TABLE 2.6: **Information on paraffin embedded tissue of 23 ovarian cancer and normal tissues, from the OV242 array chip.** Clinical details of samples obtained from the origene kits for ovarian cancer.

Locus	**Sex**	**Age**	**Organ**	**Pathology**	**Grade**	**Stage**
A1	M	49	Lung	Squamous cell carcinoma	1-2	IIIb
A2	M	43	Lung	Squamous cell carcinoma	1	II
A3	M	51	Lung	Squamous cell carcinoma	1	I
A4	F	40	Lung	Squamous cell carcinoma	1	II
A5	M	62	Lung	Squamous cell carcinoma (sparse)	-	II
A6	M	40	Lung	Squamous cell carcinoma	1	II
A7	M	64	Lung	Squamous cell carcinoma	1	I
A8	M	60	Lung	Squamous cell carcinoma	1	II
B1	M	65	Lung	Squamous cell carcinoma	2	II
B2	F	54	Lung	Squamous cell carcinoma	2	IIIa
B3	M	70	Lung	Squamous cell carcinoma	2	I
B4	M	39	Lung	Squamous cell carcinoma	2	I
B5	M	63	Lung	Squamous cell carcinoma	2	I
B6	M	63	Lung	Squamous cell carcinoma (sparse)	-	II
B7	M	50	Lung	Squamous cell carcinoma	2	II
B8	M	68	Lung	Squamous cell carcinoma	-	IIIb
C1	M	49	Lung	Squamous cell carcinoma	2	II
C2	M	35	Lung	Squamous cell carcinoma	3	I
C3	M	61	Lung	Squamous cell carcinoma	3	I
C4	M	72	Lung	Adenocarcinoma	1	I
C5	M	68	Lung	Adenocarcinoma	1	I
C6	M	62	Lung	Adenocarcinoma	2	I
C7	F	59	Lung	Adenocarcinoma	2	II
C8	M	55	Lung	Adenocarcinoma	2	I

D1	F	42	Lung	Adenocarcinoma	2-3	I
D2	M	59	Lung	Adenocarcinoma	2	I
D3	F	60	Lung	Adenocarcinoma (sparse) with necrosis	2	I
D4	F	46	Lung	Adenocarcinoma	2	I
D5	F	53	Lung	Adenocarcinoma	2	I
D6	F	41	Lung	Adenocarcinoma	2	I
D7	M	58	Lung	Adenocarcinoma	2	I
D8	F	56	Lung	Adenocarcinoma	2	II
E1	M	60	Lung	Adenocarcinoma	3	I
E2	M	65	Lung	Adenocarcinoma with necrosis	3	I
E3	M	71	Lung	Adenosquamous carcinoma	-	II
E4	M	50	Lung	Adenosquamous carcinoma	-	II
E5	M	70	Lung	Adenosquamous carcinoma	-	IIIb
E6	F	67	Lung	Adenosquamous carcinoma	-	I
E7	M	60	Lung	Atypical carcinoid	-	I
E8	M	67	Lung	Atypical carcinoid	-	I
F1	F	21	Lung	Normal lung tissue	-	-
F2	F	21	Lung	Normal lung tissue	-	-
F3	F	40	Lung	Normal lung tissue	-	-
F4	F	40	Lung	Normal lung tissue	-	-
F5	F	18	Lung	Normal lung tissue	-	-
F6	M	49	Lung	Normal lung tissue	-	-
F7	M	50	Lung	Normal lung tissue	-	-
F8	M	47	Lung	Normal lung tissue	-	-

TABLE 2.7: **Information on paraffin embedded tissue of 48 lung cancer and normal tissues, from the LC485 array chip.** Clinical details of lung cancer samples obtained from this kit.

2.2.6 Whole blood

Ovarian cancer blood samples were collected from ovarian cancer patients enrolled on a clinical trial at Mount Vernon Hospital, London, known as the METRO-BIBF trial (appropriate ethical approval was in place). Healthy control blood samples were collected from female volunteers. All patients were stage III & IV, Ovarian Serous Adenocarcinoma, enrolled on the METRO-BIBF trial to begin third line dual chemotherapy treatment.

Blood samples from lung cancer patients were collected from consenting patients (with appropriate NRES ethical approval in place) at The Royal Brompton & Harefield NHS Trust. Samples were also collected from patients with non-cancer diseases and normal healthy volunteers to act as controls. Sample details are summarised in Table 2.8.

		Stage				
Pathology		**I**	**II**	**III**	**IV**	**Un-staged**
Non-small cell lung cancer	Adenocarcinoma	8	9	3	1	-
	Squamous Cell Carcinoma	-	6	1	2	-
	Metastatic lung cancer	-	-	-	-	5
Controls	Bullectomy					5
	Healthy volunteers					20

TABLE 2.8: **Clinical samples information.** Details including pathology and staging of cancer.

All blood samples were collected in ethylenediaminetetraacetic acid (EDTA) tubes to prevent coagulation of the blood. The samples were inverted 10 times, before decanting 5mL into a 15mL falcon tube and spinning for 10 min at 1,500 RPM, 2mL of the plasma layer is then aspirated, carefully not to disturb the red blood sediment, in to a 2mL eppendorf tubes and stored at -80°C, until use. If RNA extraction was not done immediately, 0.5mL of whole blood was added to 1.5mL of RNAlater® (Life Technologies) in an eppendorf tube and stored at room temperature for up to 72 hours, or -20 indefinitely.

2.3 Isolation and imaging of cells

2.3.1 ImageStream X^{TM}

Protein expression and localisation was investigated using ImageStream X^{TM} (Amnis) high resolution flow cytometry. Expression and cellular location of AE1/AE3, RAD51 associated protein 1 (RAD51AP1) and Follistatin like protein 1 (FSTL1) was assessed using ImageStream X^{TM} (Amnis) imaging flow cytometry.

2.3.2 Preparing cultured cells

Cells were cultured in T-75 tissue culture flasks until approximately 90% confluent. Media was aspirated and the cells were washed briefly in PBS (Gibco). The cells were incubated with 2.5mL of TrypLE^{TM} Express (Invitrogen) per $\sim$75cm^2 growth surface area and the flask was manually disturbed to detach adherent cells, which were resuspended in 2.5mL of prewarmed PBS (Gibco). The cell suspension was transferred into a 15mL centrifuge tube and centrifuged for 5 minutes at 1,500 RPM to form a pellet. Supernatant was removed and cells were resuspended in 5mL of prewarmed PBS (Gibco) a further time to remove debris. Samples were fixed immediately as described in Section 2.3.4.

2.3.3 Preparing patient samples for ImageStream X^{TM}

One mL of whole blood from patient samples was decanted from the EDTA blood tubes into a 15mL falcon tube and mixed with 9mL of red blood cell (RBC) lysis buffer (G Biosciences), inverted several times and incubated for 10 mins with gentle agitation. The solution was then spun at 2,500 RPM for 10 mins, the supernatant removed and a further 2mL of RBC lysis buffer added to resuspend the pellet, before incubating for 10 mins at room temperature with gentle agitation. The solution was spun for a further 10 mins at 2,500 RPM, and supernatant aspirated. The pellet was then washed in 1.5mL of PBS, and solution moved to a 2mL microcentrifuge tube and spun at 3,600 RPM for 3 mins. Samples were then fixed immediately as described in Section 2.3.4. For spiking in experiments, 1mL of healthy blood was used, and an approximate amount of cells calculated (as described in Section 2.1.6) and spiked in accordingly. Samples were then processed as per patient samples described above in this section.

2.3.4 Fixing cells

All cell pellets (cultured cells and patient samples) were then transferred to a 2mL microcentrifuge tube and resusended in 1mL of ice cold 4% paraformaldehyde for 7 minutes to crosslink proteins in the cells. The cell suspension was centrifuged for 2 minutes at 3,600 RPM and the PFA removed. The cells were washed in prewarmed PBS a further two times, and centrifuging for 2 minutes at 3,600 RPM between each wash. All cells were immediately stained.

2.3.5 Staining cells

The cells were incubated in blocking buffer (10% fetal bovine serum, Gibco, in PBS) for 30 minutes with gentle agitation. Cells were centrifuged for 3 minutes at 3,600 RPM and the blocking buffer was removed. The cells were then incubated in the appropriate primary antibody, details listed in (Table 2.9), diluted in blocking buffer overnight at 4°C with gentle agitation. Following primary antibody incubation, cells were centrifuged for 3 minutes at 3,600 RPM and the antibody aspirated. The cells were washed in PBS (Gibco) tween (0.2%) to remove any remaining antibody and centrifuged again for 3 minutes at 3,600 RPM. The PBS tween was removed and the cells incubated in secondary antibody (Table 2.9) diluted in blocking buffer for 1 hour with gentle agitation. From this step onwards the cell were protected from light as the fluorphore conjugated to the secondary antibody is light sensitive. After secondary antibody incubation the cells were centrifuged for 3 minutes at 3,600 RPM and the secondary antibody removed. The cells were washed once in PBS tween for 3 minutes at 3,600 RPM to remove any remaining antibody. PBS tween was removed and the cells

were resuspended in 50μL Accumax (Innovative Cell Technologies) to dissociate any cellular aggregates. 1μL of Draq5 (Biostatus Ltd) nuclear stain was added before visualisation on the ImageStream X. All the data files were then analysed on the IDEAS software.

Primary Antibody	**Dilution**	**Species**	**Suplier**	**Secondary Antibody**		**Dilution**	**Suplier**
Anti-AE1/AE3	1:200	Rabbit	Dako	Anti-rabbit, conjugated	alexifluor	1:1000	Molecular Probes
Anti-RAD51AP1	1:200	Rabbit	Sigma	Anti-rabbit, conjugated	alexifluor	1:1000	Molecular Probes
Anti-FSTL1	1:200	Rabbit	Sigma	Anti-rabbit, conjugated	alexifluor	1:1000	Molecular Probes

TABLE 2.9: **Details of antibodies used in Imagestream X experiment.** Details of all primary and respective secondary antibodies used in the Imagestream X experiments.

2.3.6 Clearbridge processing of cell culture pellets and lung samples

Samples were processed on the ClearBridge ClearCell FX system (ClearBridge Biomedics^{TM}). Cultured H358 cells were prepared as described in Section 2.1.3 and 2.1.4. Approximately 100,000 cells (counted as described 2.1.6) were spiked in to 7.5mL of blood from a healthy volunteer and processed as patient samples. Peripheral blood samples from lung cancer patients, and patients suffering from other lung malignancies (controls) were collected in 9mL EDTA vacutainers or Streck tubes, either prior to surgery for surgical patients, and in clinics for all other samples, all samples were processed within 24 hours for EDTA samples, and within 72h for Streck tubes. The blood was first inverted 10 times before adding 4.3mL of a suspension reagent supplied by the manufacturer. The samples were then processed through the ClearCell^{TM} system using a CTC Chip^{TM} FR1 and microfluidics technology to separate tumour cells from white blood cells, to form an enriched suspension of CTCs. Processed samples were air dried on to glass slides and subjected to standard haematoxylin and eosin (H&E) staining, and then reported independently by two histopathologists. Results were categorized into negative or positive cells suggestive of tumour, based on the presence of nucleated cells with high nuclearcytoplasmic ratios that were larger than resting lymphocytes and/or cells having irregular nuclear outlines.

Inter-pathologist agreement was measured by Cohen's Kappa statistic (Medicalc) and diagnostic utility was calculated as sensitivity, specificity, positive and negative predictive values and expressed with a 95% confidence interval. Statistical analyses were performed on Stata 13 (College Station, Texas, USA).

2.4 Immunohistochemistry

Immunohistochemistry and DAB (3,3-Diaminobenzidine) staining was used to visualise mTOR, DEPTOR, rictor and raptor protein expression in paraffin embedded ovarian and lung cancer tissue samples. A PAP pen was used to encircle the tissue sections on the slide, to allow liquids to stay within the desired region. Tissue was deparaffinised and rehydrated by incubation in Histo-Clear (National Diagnostics) and decreasing concentrations of ethanol as follows (Table 2.10):

Solution	**Time**
Histoclear	3×5 minutes
Histoclear:ethanol (1:1)	3 minutes
100% ethanol	3 minutes
95% ethanol	3 minutes
70% ethanol	3 minutes
50% ethanol	3 minutes
Running tap water	3 minutes

TABLE 2.10: **IHC deparaffinisation process.** Incubation times to deparaffinise and rehydrate formalin fixed paraffin embedded tissue samples.

Antigen retrieval was achieved by boiling the slides in sodium citrate (pH6, Table 2.11) for 20 minutes. Additional sodium citrate was added every 5 minutes to avoid the slides boiling dry.

Reagent	**Quantity**
Sodium citrate	2.94g
dH_2O	1L
Tween 20	500μL

TABLE 2.11: **Formulation of antigen retrieval solution.**

The slides were then washed in running tap water for 10 mins followed by washing with TBS 0.025% Triton X 2 times for 5 minutes to remove sodium citrate. Endogenous hydrogen peroxidase activity was blocked by incubation with 0.3% hydrogen peroxide for 30 mins. The slides were then washed in TBS 0.025% Triton X 3 times for 5 mins each on a shaker. Slides were placed in a humidity chamber (glass chamber with dH_2O in surrounding wells) for subsequent incubations to avoid drying out. The tissue was blocked in 10% rabbit serum in TBS for 1 hour at room temperature by dispensing 200μL onto the tissue area of the slide and using a small square of parafilm to ensure the blocking buffer spreads over the tissue evenly. Blocking buffer was carefully removed from the slides and the primary antibody (diluted in blocking buffer) was applied to the tissue in the same way as the blocking buffer and incubated over night at 4°C. After primary antibody incubation, the slides were washed

3 times for 5 mins each time in TBS 0.025% Triton X on a shaker to remove any unbound primary antibody. The tissue was then incubated in horseradish peroxidase (HRP) conjugated secondary antibody (diluted in blocking buffer) at room temperature for 1 hour in the same way as blocking buffer. The slides were then washed 3 times for 5 mins in TBS 0.025% Triton X on a shaker to remove any unbound secondary antibody. DAB solution (DAKO) was prepared as follows: one drop of chrom added to 1mL of DAB. 200μL of the DAB solution was applied to the tissue and incubated at room temperature for 2-5 minutes until a brown colour develops. DAB solution was washed off the slide by incubation in dH_2O for 5 mins on a shaker. Haematoxylin (Fisher Scientific) was used to stain nuclei blue to provide contrast to the brown DAB staining. The tissue was incubated in haematoxylin for 1 min and rinsed off with running tap water followed by incubation in 0.1% sodium bicarbonate for 1 min and a second wash in running tap water. Tissue was dehydrated prior to coverslip mounting as follows (Table 2.12):

Solution	**Time**
50% ethanol	3 minutes
70% ethanol	3 minutes
95% ethanol	3 minutes
100% ethanol	3 minutes
Histoclear:ethanol (1:1)	3 minutes
Histoclear	3 minutes

TABLE 2.12: **Incubation times to dehydrate stained tissue samples.**

Coverslips were mounted with aqueous mounting media (DAKO) to protect the tissue, the slide was allowed to dry for 30 mins before observation.

2.5 RNA extraction

mRNA was extracted from cells for examination of gene expression transcripts. All surfaces and equipment were cleaned thoroughly with 70% IMS and 2% Trigene before and throughout any RNA work. RNA was stored at -80°C when not in use and kept on ice at all other times to prevent degradation.

2.5.1 RNA extraction from cultured cells using GenElute^{TM} (Sigma Aldrich)

The GenElute^{TM} Mammalian Total RNA MiniPrep Kit is a commercially available silica membrane, spin column based RNA extraction kit. Media was removed from cells in culture. Cells were briefly washed in phosphate buffered saline (PBS, Gibco). Cells were then trypsinised with 2mL TrypLE^{TM} Express (Invitrogen) and incubated for 5 mins at 37°C to allow cells to detach. The cell solution was collected in to a 15mL falcon tube, and spun at

1,500 RPM for 5 mins. Supernatant was then aspirated and 500μL of lysis buffer containing 1% β-mercaptoethanol was used to resuspend the cell pellet and lyse cell membranes and inactivate RNases. The lysate was transferred into a blue filtration column contained within a collection tube and centrifuged for 2 minutes at 12,000$\times$g to remove cellular debris and shear DNA. The filtered lysate was then mixed with 500μL of 70% ethanol to enhance the binding of nucleic acids to the silica membrane and loaded into a red silica column contained within a collection tube. The column was centrifuged for 15 seconds at 12,000$\times$g; the flow-through was discarded as the RNA had now bound to the membrane. The RNA was washed three times, once with Wash Solution I and twice with Wash Solution II (supplied in kit) by centrifuging for 15 seconds at 12,000$\times$g to remove any impurities including proteins, cellular debris and salts. A final centrifugation of 12,000$\times$g for 2 mins was carried out to ensure that the ethanol was completely removed from the membrane. RNA was eluted by placing the membrane column in a fresh collection tube, applying 50μL of pure water (elution solution) to the membrane and centrifuging at 12,000$\times$g for 60 seconds. The RNA was stored at -80°C until further use.

2.5.2 RNA extraction of tissue samples (GenElute^{TM} kit by Sigma Aldrich)

Tissue samples were removed from -80°C, and 40mg of tissue was placed in a sterile mortar, liquid nitrogen was then poured over the tissue and the pestle used to smash the tissue to a powder residue. The tissue was then added to a 1.5mL eppendorf tube and 500μL of lysis buffer containing 1% β-Mercaptoethanol was added and retropippeted to resuspend the tissue. The lysate was transferred into a blue filtration column contained within a collection tube and centrifuged for 2 mins at 12,000$\times$g to remove cellular debris and shear DNA. The filtered lysate was then mixed with 500μL of 70% ethanol to enhance binding of nucleic acids to the silica membrane and loaded into a red silica column contained within a collection tube. GenElute^{TM} protocol was then continued as above from Section 2.5.1.

2.5.3 RNA extraction from Formalin Fixed Paraffin Embedded tissue blocks with Promega ReliaPrep^{TM} FFPE RNA extraction mini kit

FFPE blocks were cut to 15μm thick scrolls using a microtome, and approximately 3 scrolls were added to a microcentrifuge tube. FFPEs were deparaffinised by adding 300μL of mineral oil (supplied) to the microcentrifuge tubes containing the scrolls and incubated at 80°C for 1 min, followed by vortexing to mix. Samples were lysed by adding 100μl of Lysis Buffer (supplied). Samples were centrifuged at 10,000$\times$g for 15 seconds at room temperature, forming 2 phases, lower (aqueous) phase, and upper (oil) phase. 10μL of Proteinase K was added to the lower aqueous phase and pipetted gently to mix. The sample was then incubated at 56°C for 15 mins, followed by 80°C for 1 min. Tubes were then immediately

placed on ice for 1 min, followed by room temperature for 2 mins. 30μL of DNAase treatment was added to the lower phase, and mixed by gentle pipetting, and incubated for 15 mins at room temperature. Followed by addition of 325μL of BL buffer and 200μL of 100% isopropanol, sample was vortexed to mix. Samples were then centrifuged at 10,000$\times$g for 15 sec. Two phases were formed as earlier. A binding column was fitted into a collecting tube, and the entire lower aqueous phase added to the binding column. Samples were then centrifuged at 10,000$\times$g for 30 sec. The flow through was discarded, and 500μL of 1$\times$ wash buffer was added to the column, followed by centrifugation at 10,000$\times$g for 15 sec, flow through was discarded and wash step repeated. Samples were then centrifuged at 16,000$\times$g for 30 sec to dry column. The column was then transferred to a fresh 2 ml eppendorf tube, and 50μL of elution solution added. Samples were then centrifuged at 16,000$\times$g for 1 min. Collected RNA was stored at -80°C until use.

2.5.4 RNA extraction from whole blood with Ribopure RNA extraction kit (Life Technologies)

All blood samples stored in RNALater were processed using the Ribopure RNA extraction Kit (Life Technologies), including for fresh samples immediately processed. Samples stored in RNALater, were first defrosted if stored at -20°C at room temperature. Defrosted samples or refrigerated/fresh samples were then centrifuged at 13,000 RPM for 1 min, followed by aspirating supernatant. From this point on the process is the same for fresh and RNALater stored samples. 800μL of lysis solution and 50μL of sodium acetate were added to the samples (500μL of fresh anticoagulated blood), samples were then vortexed vigorously. 500μL of Acid-Phenol:Chloroform was added to the cell lysate, and vortexed vigorously for 30 secs, and stored at room temperature for 5 min. Samples were then centrifuged at 13,000 RPM for 1 min, to separate the aqueous and organic phase. The upper aqueous phase was then transferred to a 2mL centrifuge tube and 600μL of 100% ethanol was added and vortexed briefly. 700μL of solution was then added to a filter cartridge assembly and spun for 30 secs at 13,000 RPM to allow solution to filter through. The supernatant in the collection tube is discarded and the next 700μL of sample solution added and spinning repeated, until all the sample solution had been filtered through. 700μL of wash solution 1 was then added to the filter cartridge and spun again for 30 secs 13,000 RPM, collection tube was replaced and 700μL of wash solution 2/3 added to the filter tube, and re-spun for 30 secs 13,000 RPM. Wash solution 2/3 step was repeated, followed by an additional spin at 1 min at 13,000 RPM, to remove residual fluid. The filter cartridge was then added to a new centrifuge tube, and 50μL of elution solution added, and spun at 13,000 RPM for 1 min. Eluted RNA was stored at -80°C until use.

2.5.5 RNA quantification

1μL of extracted RNA per sample was analysed spectrophotometrically for concentration and purity using the NanoDrop 2000C (Thermo Fisher Scientific). Concentration was calculated from absorbance at 280nm and purity was calculated by A260/A280 ratio with an acceptable range of 1.7-2.1.

2.5.6 cDNA synthesis

Complementary DNA (cDNA) was synthesised from extracted RNA for use as template in qPCR experimentations. All surfaces and equipment were cleaned thoroughly with 70% IMS before and throughout cDNA synthesis work. Random primers were chosen over oligo-dT in order to avoid 3' poly-A tail weighting. RNA volume input was calculated by the following:

RNA input = desired cDNA concentration/RNA concentration (ng/μL).

2.5.7 High throughput cDNA synthesis kit by Applied Biosystems (Thermo Fisher)

RNA input was calculated and the appropriate amount, made up to 10μL with pure H_2O, was mixed in one 0.6mL, thin walled PCR tube for each sample. The reverse transcriptase mastermix was made up using reagents and volumes listed in Table 2.13.

Reagent	**Volume Added**
10× RT Buffer	2.0μL
25× dNTP Mix (100 mM)	0.8μL
10× RT Random Primers	2.0μL
MultiScribe^{TM} Reverse Transcriptase	1.0μL
RNase Inhibitor	1.0μL
Nuclease-free H_2O	3.2μL
Total per reaction 10.0μL	10.0μL

TABLE 2.13: **cDNA synthesis process.** Reagent volumes for reverse transcriptase for one sample for use with the High-throughput cDNA reverse transcriptase kit (Thermo Fisher).

10μL of the mastermix was then added to the RNA/H_2O samples, to make a 20μL final volume, samples were briefly vortexed. Samples were then incubated at 25°C for 25 minutes to anneal the primers, 37°C for 120 minutes to extend the synthesis, and 85°C for 5 mins to inactivate the enzyme. cDNA was stored at -20°C until further processing.

2.5.8 Quantitative PCR (qPCR)

qPCR (quantitative polymerase chain reaction) was used to assess relative gene expression (RQ) in a number of different experiments. qPCR experiments were prepared in an exclusive area or PCR dedicated hoods when possible and all equipment used was sterilised using 2% TriGene (Medimark Scientific) and 70% IMS initially and repeatedly throughout the preparation. All reagents were defrosted and kept on ice when in use to preserve their integrity and SYBR® Green containing reagents were kept in a light proof storage whenever possible.

2.5.9 Reference gene assessment

The MIQE (minimum information for the publication of qPCR experiments) guidelines outline the importance of an assessment of the most appropriate reference genes specific to the samples used must be carried out prior to any qPCR experiment. In light of this, a selection of samples representing the whole cohort in each experiment were assessed using the either the geNorm™ human 12 gene kit or geNorm™ human 6 gene kit (Primerdesign). The genes assessed are detailed in Table 2.14.

Gene name	Description
B2M*	Beta-2-microglobulin (mRNA)
YWHAZ*	Phospholipase A2 (mRNA)
RPL13A*	Ribosomal protein L13A (mRNA)
18S	18S RNA (rRNA)
UBC*	Ubiquitin C (mRNA)
ATP5B	ATP synthase subunit (mRNA)

TABLE 2.14: **The geNorm™ human 6 gene kit from Primerdesign.**

2.5.10 geNorm™ with PrecisionPLUS™ Mastermix (Primerdesign)

GeNorm™ experiments were prepared in a qPCR exclusive area and all equipment used was sterilised using 2% TriGene (Medimark Scientific) and 70% IMS initially and repeatedly throughout the preparation. All reagents were defrosted and stored on ice when in use to preserve their integrity and SYBR® Green containing reagents were kept in light proof containers whenever possible. GeNorm™ experiments were carried out on the Quantistudio 7 cycler (Life Technologies). PrecisionPLUS™ Mastermix by Primerdesign was used for the geNorm™ experiments as per manufacturer's instructions. PrecisionPlus™ Mastermix is pre-prepared master mix containing a hot start Taq polymerase enzyme, SYBR® Green I dye and deoxynucleoside triphosphates (dNTPs). Primers in the geNorm™ kit were supplied lyophilised, thus prior to use tubes were briefly centrifuged to collect dry components at the bottom of the tubes which were then rehydrated with 220μL of pure H_2O. Forward

and reverse primers were contained within the same tube. Samples were run in duplicate and so small mixes were made up for each primer for (n $\times$ 2)+1 containing the following components as shown in Table 2.15.

Reagent	**Volume n=1 (μL)**
Primer Mixture 1	1
PrecisionPLUS^{TM} Mastermix	10
Pure H_2O	4
cDNA	5
Total	20

TABLE 2.15: **geNorm primer mix.** Volume of components for primer mixes in geNorm^{TM} experiment.

cDNAs for all representative samples were diluted to a concentration of 5ng/μL. 15μL of the primer mix was pipetted into each well (two reactions for each sample) of a MicroAmp® Fast Optical 96-Well Reaction Plate (Life Technologies) followed by 5μL of appropriate cDNA. The wells were sealed hermetically with a transparent, contact adhesive sealing film and the plate briefly centrifuged to ensure all the reagents had collected and there were no visible air bubbles. The reagents were then subject to the following thermal protocol detailed in Table 2.16.

Step	**Temperature**	**Time**	**Cycles**
Start	95°C	2 minutes	1
Amplification	95°C	15 seconds	40
	60°C	60 seconds	
Fluorescent measurement	-	-	
Dissociation curve	60-95°C	-	1

TABLE 2.16: **Thermal protocol of a geNorm^{TM} experiment on the Quantstudio 7 (Life Technologies).**

Following the amplification protocol, the dissociation temperature of the PCR products generated were assessed by a dissociation protocol. In the protocol, the temperature was increased by 1°C from 60-95°C, whereby a fluorescence measurement is taken at each increase. The differential fluorescence was plotted against the temperature (°C) to determine the temperature at which the product strands dissociated.

2.5.11 geNorm^{TM} analysis

geNorm^{TM} was analysed using the qBase+ software (Biogazelle). qPCR raw data was uploaded to qBase+, where the software calculated an M (measure) and V (variation) value for expressional stability and optimum number of reference genes to use respectively.

2.5.12 Primers

Primer sequences were sourced from the literature, and ordered from Sigma Aldrich. Primers were supplied lyophilised and so were briefly centrifuged to collect dry components at the bottom of the tubes. Primers were then rehydrated according to manufacturer's direction. All reference genes are shown in Table 2.17 and all primer sequences are listed in Table 2.18.

Species	**Gene Symbol**	**Accession Number**	**Anchor Nucleotide**	**Context Length Sequence (bp)**
HUMAN	18S	NM_10098	235	99
HUMAN	B2M	NM_004048	362	141
HUMAN	GAPDH	NM_002046	1087	142
HUMAN	ACTB	NM_001101	1194	106

TABLE 2.17: **Primer sequence of reference genes.** Primer sequence for all reference genes tested as purchased from Primerdesign. Primerdesign does not release primer sequence details, but provided the amplicon context sequence for all reference genes, accepted by MIQE guidelines for publication purposes.

Primer	**Amplicon (bps)**	**Strand**	**Size (base)**	**Sequence**
RAD51AP1	121	Forward	20	AGTGAAGTAAAATCCCAGTAGA
		Reverse	20	TGGCAAGGACTGAGATTCTGAT
FSTL1	150	Forward	22	TAAAGGGAGCAACTACAGYGAAAT
		Reverse	20	TTGTTCTCCTGGTCTGGATAGC
COL12A1	98	Forward	19	GTGCCTGGACTGATTGGTTT
		Reverse	21	AGACACAAGAGCAGCAATGAAG
SPRR1A	137	Forward	19	GCCACTGGATACTGAACACC
		Reverse	21	AGAATGAGGGTAAGGGACATCTT
SOX2	210	Forward	19	CATCGACGAGGCTAAGCGG
		Reverse	18	TAACTGTCCATGCGCTGGTT
SNAI1	208	Forward	21	GAGCTGAACCTCCCTGTCAGA
		Reverse	20	GTTGAAGGCCTTTCGAGCCT
MDM2	285	Forward	22	ATGTCTGTACCTACTGATGGTGC
		Reverse	20	TCACAGAGAAGCTTGGCACG
AIP	199	Forward	19	TCAAGGCCTACTTCAAGCGG
		Reverse	19	CAATGGGAGAAGATCCCCCG
FAS	192	Forward	18	ACACTCACCAGCAACACCAA
		Reverse	19	TGCCACTGTTTCAGGATTTAA
SOD1	78	Forward	23	TAAAGTAGTCGCGGAGACGGG

		Reverse	22	CGGCCTCGCAACACAAGCCT
BAX	204	Forward	20	TGCTTCAGGGTTTCATCCA
		Reverse	21	GGAAAAAGACCTCTCGGGG
SHISA2	244	Forward	19	ACGATTCGACCATCTGCTG
		Reverse	21	CAGTTGGTTTGGGATCGAGT
THBS1	550	Forward	22	CCAGCTGTACATCGACTGTGA
		Reverse	22	GCAGATGGTAACTGAGTTCTGA
XIST	110	Forward	20	AGGTCAGGCAGAGGAAGTCA
		Reverse	20	CTCCCGATACAACAATCACG
		Reverse	19	CAGTGGGTGTTGCAGGATG
GSTT1	989	Forward	20	GGCGAGAGAGCAAGACTCAG
		Reverse	20	GGCAGCATAAGCAGGACTTC
NBPF14	201	Forward	19	TGGCTCATCCTCTATGTTG
		Reverse	21	AGATTCCTGGGAGACACTGGT
IL1A	119	Forward	21	CGCCAATGACTCAGAGGAAGA
		Reverse	19	AGGCGTCATTCAGGATGAA
IL1B	118	Forward	18	AATCTGTACCTGCGTGTT
		Reverse	25	TGGTAATTTTTGGGATCTACACTCT
EGR1	160	Forward	21	CTTCAACCCTCAGGC GGACA
		Reverse	21	GGAAAAGCGGCCAGTATAGGT
F2RL2	109	Forward	23	AGCTCCACGAAAGGTCTTAATGG
		Reverse	18	AGCTGCTGGCCTGCTGCT
MEGF6	452	Forward	24	CCTCACAGCTGTTCACGATTTCCA
		Reverse	22	GATTGGCCTTGAATGCCTGTCA
CTFG	117	Forward	19	AATGCTGCGAGGAGTGGGT
		Reverse	23	CGGCTCTAATCATAGTTGGGTCT
mTOR	135	Forward	20	TGCCAACTACCTTCGGAACC
		Reverse	19	GCTCGCTTCACCTCAATTC
DEPTOR	202	Forward	19	CACCATGTGTGTATGAGCA
		Reverse	19	TGAAGGTGCGCGCTCATTG
rictor	117	Forward	20	GGAAGCCTGTTGATGGTGAT
		Reverse	19	GGCAGCCTTTTTATGGTGT
raptor	170	Forward	19	ACTGATGGAGTCCGAATGC
		Reverse	19	TCATCCGATCCTTCATCCTC

TABLE 2.18: **Primer sequences of all genes.** Primer sequence for all genes tested, all primers were synthesised by Sigma-Aldrich.

2.5.13 Gel electrophoresis

Primers were validated by running the PCR products on a 2% agarose gel. 2g of agarose (Fisher Scientific) was added to 100mL of 1 × tris-borate EDTA (TBE, 89mM Tris-borate, 2mM EDTA, pH 8.3) buffer and microwaved for approximately 2 mins or until the agarose had dissolved and the mixture was clear. The mixture was cooled to approximately 50°C, and 5μL ethidium bromide was added and mixed well. The gel was poured into a prepared casting tray on a level surface and a 20 well comb inserted at one end. The gel was left to set for approximately 25 mins and transferred into a tank containing 1×TBE buffer. 2μL of DNA loading buffer was mixed with 10μL qPCR product to aid loading and visualisation of migration through the gel. The mixture was added to each appropriate well. In a separate well 5μL of 1kb Plus DNA Ladder™ (Life Technologies) was added as a guide of product size. The gel was run at 100V and 400mA for approximately 45 minutes until the product had migrated a satisfactory distance. The gel was visualised using a Gel Doc™ XR+ Imaging System (Bio-Rad) and analysed visually for consistency of product size, to predicted product size and the presence of extra products which may represent contamination or primer dimers.

2.5.14 Power SYBR® Master Mix (Life Technologies)

Power SYBR® Master Mix (Life Technologies) was used with the Quantstudio 7 (Life Technologies). Power SYBR® Master Mix is a pre-prepared master mix containing AmpliTaq Gold DNA Polymerase, SYBR® Green I dye and dNTPs (with a blend of dUTP and dTTP). Reactions were formulated as follows (Table 2.19).

Reagent	**Volume n = 1 (μl)**
Power SYBR® Master Mix (2×)	10
Pure H_2O	4
Forward Primer (10μM)	0.5
Reverse Primer (10μM)	0.5
cDNA (100ng/μL)	5
Total	20

TABLE 2.19: **qPCR mastermix preparation.** Volume of components for primer mixes in qPCR experiment with Power SYBR® Master Mix (Life Technologies).

The above reagent mix was made up separately for each primer set. 15μL of the appropriate mix and 5μL of cDNA was dispensed into a MicroAmp® Fast Optical 96-Well Reaction Plate (Life Technologies) according to a pre-designed well plan. NTCs (non-template controls), where water was substituted for cDNA, were included for each of the primer

mixes as a negative control. According to MIQE guidelines and geNorm^{TM} analysis (completed previously), 2 reference genes, B2M and 18S, were shown to be expressionally stable and thus were included as an endogenous controls in this set of experiments. The wells were sealed hermetically with a transparent, contact adhesive sealing film and the plate centrifuged to ensure all the reagents had collected. The sample plate was subjected to a temperature protocol as described in Table 2.20.

Step	**Temperature**	**Time**	**Cycles**
Hot Start	95°C	10 minutes	1
Amplification	95°C	15 seconds	40
	60°C	60 seconds	
Fluorescent measurement	60°C	-	
Dissociation curve	60-95°C	-	1

TABLE 2.20: **qPCR thermal cycler details.** The thermal protocol of a qPCR experiment using Power SYBR® Master Mix on a Quantstudio 7 thermal cycler (both Life Technologies).

Following the amplification protocol, the dissociation temperature of the PCR products generated were assessed by a dissociation protocol as described above.

2.5.15 Multiplex TAQMan qPCR

Multiplex TAQman qPCR was used to measure RAD51AP1 in clinical tissue and blood samples. The RAD51AP1 gene was tagged with FAM-MGB, the 2 reference genes Beta Actin and GAPDH were used, and both were tagged with VIC-MGB, and FAM-MGB, respectively. Multiplex was carried out on the RAD51AP1 and Beta Actin, whereas GAPDH was run separately. The TAQman Multiplex real time PCR mastermix (ABI) was used in a 20μL reaction on the Quantstudio 7. Reactions were formulated as show in Table 2.21.

Reagent	**Volume n=1 (μL)**
TAQman Multiplex real time PCR mastermix	10
Pure H_2O	3*
Primer 1	1
Primer 2	1
cDNA (100ng/μL)	5
Total	20

TABLE 2.21: **Multiplex reagent mix.** Volume of components for primer mixes in qPCR experiment with TAQman Multiplex real time PCR mastermix (ABI). *In GAPDH reactions only, H_2O was increased to 4μL to make up a final 20μL working reaction.

The above reagent mix was made up for RAD51AP1 and Beta Actin simultaneously and GAPDH separately. 15μL of the appropriate mix and 5μL of cDNA was dispensed into a MicroAmp® Fast Optical 96-Well Reaction Plate (Life Technologies) according to a pre-designed well plan. NTCs, where water was substituted for cDNA, were included for each of the primer mixes as a negative control. Beta Actin and GAPDH were used as reference genes. The wells were sealed hermetically with a transparent, contact adhesive sealing film and the plate centrifuged to ensure all the reagents had collected. The sample plate was subjected to a temperature protocol as described in Table 2.22.

Step	**Temperature**	**Time**	**Cycles**
Hot Start	50°C	2 minutes	1
	95°C	10 seconds	1
Amplification	95°C	15 seconds	40
	60°C	60 seconds	
Fluorescence Measurement	60°C	-	
Dissociation Curve	60-95°C	1	1

TABLE 2.22: **Multiplex qPCR thermal cycler process.** The thermal protocol of a qPCR experiment using Power SYBR® Master Mix on a Quantstudio 7 thermal cycler (both Life Technologies).

Following the amplification protocol, CT values were generated and relative quantification calculated as described in Section 2.5.19.

2.5.16 DNA extraction

DNA was extracted from frozen plasma samples from clinical samples, using a commercially available kit, Qiagen DNA blood mini kit (cat #51104). Briefly, plasma was defrosted at room temperature. 20μL of protease K was added to the bottom of a clean 1.5mL eppendorf tube, followed by addition of 200μL of plasma sample, and 200μL of buffer AL, the solution was then pulse mixed for 15s. The sample was then incubated for 10 mins at 56°C in a water bath. 200μL of 100% ethanol was then added to the samples, followed by pulse vortexing for 15 s. The solution was then spun through a QIAamp Mini spin column with a collection tube attached, at 8,000 RPM for 1 min, discarding the filtrate. 500μL of buffer AW1 was added to each spin column and spun at 8,000 RPM for 1 min. The flow through was discarded and the spin column placed in a new collection tube. 500μL of buffer AW2 was added, and samples spun for 3 mins at 14,000 RPM. The flow through was discarded and the spin column spun for an additional 1 min at 14,000 RPM to eliminate carryover of buffer AW2. The spin column was then placed in a new 1.5mL eppendorf tube, and 200μL of buffer AE added, incubated for 1 min at room temperature before being spun at 8,000 RPM for 1 min. The DNA elute was stored at -20°C until further analysis.

2.5.17 ALU repeat primers

Sequences for the ALU 114 and 247 were taken from Umetani *et al.* (2006), and synthesised by Sigma-Aldrich (Table 2.23).

Primer	Forward Sequence	bp	Reverse Sequence	bp
ALU 114	CCTGAGGTCAGGAGTTCGAG	20	CCCGAGTAGCTGGGATTACA	20
ALU 247	GTGGCTCACGCCTGTAATC	19	CAGGCTGGAGTGCAGTGG	19

TABLE 2.23: **Alu repeat primer sequence.** Primer sequences for Alu repeat sequences 114 and 247. Sequences obtained from Umetani *et al.*, 2006, and synthesised by Sigma-Aldrich.

2.5.18 DNA qPCR

Power SYBR® Master Mix (Life Technologies) was used with the Quantstudio 7 (Life Technologies). Reactions were formulated as follows (Table 2.24).

Reagent	Volume n = 1 (μL)
Power SYBR® Master Mix (2×)	10
Pure H_2O	4
Forward Primer (10μM)	0.5
Reverse Primer (10μM)	0.5
DNA (100ng/μL)	5
Total	20

TABLE 2.24: **Alu repeat PCR reagent mix.** Volume of components for primer mixes in qPCR experiment with Power SYBR® Master Mix (Life Technologies).

The above reagent mix was made up separately for each primer set. 15μL of the appropriate mix and 5μL of cDNA was dispensed into a MicroAmp® Fast Optical 96-Well Reaction Plate (Life Technologies) according to a pre-designed well plan. NTCs where water was substituted for cDNA, were included for each of the primer mixes as a negative control. The wells were sealed hermetically with a transparent, contact adhesive sealing film and the plate centrifuged to ensure all the reagents had collected. The sample plate was subjected to a temperature protocol consisting of the following (Table 2.25).

Step	Temperature	Time	Cycles
Hot Start	95°C	10 minutes	1
Amplification	95°C	30 seconds	35
	64°C	30 seconds	
	95°C	15 seconds	
	60°C	60 seconds	
	95°C	15 seconds	
Fluorescence Measurement	60°C	-	
Dissociation Curve	60-95°C	1	1

TABLE 2.25: **qPCR thermal cycler process.** The thermal protocol of a qPCR experiment using Power SYBR® Master Mix on a Quantstudio 7 thermal cycler (both Life Technologies).

Following the amplification protocol, CT values were generated and relative quantification calculated as described below. A DNA Integrity Index calculated as follows: RQ Alu 247/RQ Alu 114.

2.5.19 qPCR analysis

Amplicon load can be measured in by either absolute or relative quantification. Absolute quantification (AQ) measures amplification in relation to a known standard to provide an absolute copy number. Relative quantification (RQ) measures amplification in relation to a reference gene or genes which are stably expressed under the conditions of the experiment (for example after treatment of cultured cells with a drug compound). The most common way to analyse relative quantification qPCR data is the ΔCq or ΔΔCq method detailed below:

Δ**Cq**=Cq (gene of interest - GOI) - Cq (reference gene)
(Normalises GOI expression to the reference gene)

ΔΔ**Cq**=ΔCq (sample) - ΔCq (calibrator)
(Normalises GOI expression in a sample to that of a calibrator, for example untreated cells)

For ΔCq method: $RQ = 2^{-\Delta Cq}$
For ΔΔCq method: $RQ = 2^{-\Delta\Delta Cq}$
(Inverts value to show change whilst assuming 100% efficiency)

The ΔΔCq method demonstrates fold change in expression in comparison to a calibrator and is only appropriate when a valid calibrator is available; for example where cultured

cells treated with a drug compound, the cells under basal conditions would be the calibrator. In the case of gene expression in clinical samples, unless each sample is matched to non-affected adjacent tissue and the assumption is made that the tissue is 'normal', it is not acceptable to match the ΔCq value from one patient to that of a different patient as natural individual differences will mean that gene expression will differ from person to person. If this is the case, the ΔCq method can be used to provide a comparison between samples and calibrators.

2.6 siRNA transfection in cell lines

In order to further understand the role and function of RAD51AP1 we sought to carryout functional studies. We used siRNA targeted to RAD51AP1 (SMARTpool: ON-TARGETplus, Dharmacon, CO, USA) to suppress RAD51AP1 expression in SKOV-3, MDAH-2774 and A549 cells. Transfection efficiency was first determined with siGLO-labelled siRNA, at different seeding densities, siRNA concentration and volume of transfection reagent. All cell lines were seeded at 2×10^4 and 1×10^5 cells, per well in 6-well (RNA) and 24-well plates (protein lysates and cell count) respectively, based on optimisation results. The siRNA transfection was carried out once at a concentration of 50nmol/L per well/dish using Dharmafect 1 transfection reagent, a scrambled control was also included (ON-TARGETplus Non-targeting Pool; Thermo Scientific). Transfection was established by, seeding cells in serum-free DMEM media (Gibco) overnight. SiRNA was diluted to a 5μM working concentration from a 20μM stock solution, by a 1 in 4 dilution in siRNA buffer (Dharmacon). Transfection was made up for n + 1 working concentration. Tables 2.26 and 2.27 detail the reagents and volumes for the reaction.

Tube A		**Tube B**	
Reagent	**Volume μL (n=1)**	**Reagent**	**Volume μL (n=1)**
siRNA	5	Transfection reagent 1	1.5
Serum free media	45	Serum free media	48.5
Serum free media - 400μL			

TABLE 2.26: **Reagents and quantities for siRNA transfection cocktail for a 24 well plate.**

Tube A and B are made up independently and allowed to incubate at room temperature for 5 mins, before being mixed together, and incubated for 20 mins at room temperature to allow the complexes to form. Following incubation the tubes were spun briefly for 1 min at 3,000 RPM. Serum free media was then added to the solution and mixed by pipetting, followed by added 500μl and 2mL to cells in the 24 well and 6 well plates respectively (following removal of existing media). The transfection media was replaced with fresh serum

Tube A		Tube B	
Reagent	**Volume μL (n=1)**	**Reagent**	**Volume μL (n=1)**
siRNA	20	Transfection reagent 1	10
Serum free media	180	Serum free media	190
Serum free media - 1600μL			

TABLE 2.27: **Reagents and quantities for siRNA transfection cocktail for a 6 well plate.**

free media at 24 hours to avoid toxicity, and samples collected at 48 and 72h post transfection. Scrambled controls were conducted in the same way, and carried for all experiments in all 3 cell lines.

2.7 Western Blotting

Protein RAD51AP1, FSTL1, mTOR, DEPTOR, rictor, and raptor lysates was analysed by Western blot. Proteins were separated by mass by SDS-PAGE (Sodium-Dodecyl-Sulphate Polyacrylamide gel electrophoresis).

2.7.1 Protein extraction

Protein was extracted from cultured cells from 24-well plates in order to examine expression patterns of RAD51AP1, FSTL1, mTOR, DEPTOR, rictor, and raptor. Media was aspirated and cells were washed briefly in PBS (Gibco). 100μL of LaemmLi buffer (Table 2.28) was added and the plate gently rocked to cover all cells. Cells were scraped with a pipette tip and lysate was transferred into a sterile 1.5mL sample tube. Lysates were denatured by boiling for 10 minutes at 95°C in a heat block before storage at -20°C until further use.

Reagent	Volume
Tris (pH 6.8)	1mL
10% SDS	4mL
Glycerol	2mL
β-mercaptoethanol	0.5mL
dH_2O	2.5mL
Bromophenol blue	Tip of pipette amount to achieve desired colour to aid visualisation

TABLE 2.28: **Volume of components of 10mL LaemmLi buffer.**

2.7.2 Western gel

A 12% resolving and 5% stacking gel were used and were prepared as follows. The gels were poured at a thickness of 1mm between glass plates according to the recipes in Table 2.29.

Gel Reagent	**Resolving (mL)**	**Stacking (mL)**
dH_2O	1.6	3.4
30% Acrylamide/Bisacrylamide (Sigma-Aldrich)	2	0.83
1.5M Tris (pH 8.8)	1.3	N/A
1M Tris (pH 6.8)	N/A	0.63
10% SDS	0.05	0.05
10% Ammonium Persulphate in dH_2O	0.05	0.05
TEMED (tetramethylethylenediamine)	0.002	0.005
Total	5	5

TABLE 2.29: **Formulations of resolving and stacking gels, in mL.**

Resolving gel was topped with 100% methanol to level and allowed to set for 20-30 minutes at room temperature. Once the resolving gel had set the methanol was poured away, the gel was washed with distilled H_2O and excess water removed with filter paper. The stacking gel was poured over the resolving gel at an approximate height of 2cm, and a comb placed between the glass plates to form wells. The stacking gel was left to set for a further 20-30 minutes at room temperature. Cell lysates were incubated at 100°C for at least 10 minutes to linearise the protein, break down disulphide bonds and remove tertiary structures. The SDS in the protein buffer and gel serves to coat the protein in a consistent negative charge to avoid the varying charges on amino acids to affect migration of the protein through the gel. 10μl of each sample was loaded into each appropriate well. 5μl of PageRuler^{TM} Prestained Protein Ladder (Life Technologies) was used as a reference of mass. The gels were run in 1$\times$ Running Buffer (Table 2.30) at 100V and 40mA per gel until the visible blue band of LaemmLi buffer was almost at the bottom of the gel and the ladder was fully separated.

Reagent	**Quantity**
Tris Base	30g
Glycine	44g
10% SDS	100mL
dH_2O	Up to 1L

TABLE 2.30: **Formulation of 10$\times$ Running Buffer.**

When the samples move through the stacking gel they are concentrated between a chloride and glycine front and so enter the resolving gel simultaneously. The proteins are

then able to separate based on their individual mass by their differential rate of movement through the resolving gel.

The separated proteins were then electrophoretically transferred onto a nitrocellulose membrane (Thermo Scientific) in Wet-Transfer Buffer (Table 2.31). The transfer apparatus were assembled whist immersed in Wet-Transfer Buffer to avoid the gel and membrane drying out. The membrane and gel were sandwiched between sponges and filter paper and placed in a cassette before being placed in a tank of transfer buffer and run at 100V and 300mA for 1.5 hours.

Reagent	**Quantity**
Tris Base	2.41g
Glycine	11.25g
dH_2O	Up to 800mL
Methanol	200mL

TABLE 2.31: **Formulation of 1× Wet-Transfer Buffer.**

After successful transfer the protein ladder was visible on the nitrocellulose membrane, and visualisation of bands in the presence of a Ponceau stain. The membrane was then rinsed briefly in 1× TBS Tween (Tables 2.32 and 2.33) and incubated in 50mL of 5% cow's milk (Marvel milk powder in TBS Tween) for one hour at room temperature on a shaker.

Reagent	**Quantity**
Tris Base	24.2g
NaCl	80g
dH_2O	Up to 1L

TABLE 2.32: **10× TBS recipe.**

Reagent	**Quantity**
10× TBS	100mL
dH_2O	899mL
Tween 20	1mL

TABLE 2.33: **Formulation of 1× TBS Tween.**

After blocking, the membrane was washed three times for ten minutes each in TBS Tween to remove any unbound milk proteins. A primary antibody (Table 2.34) diluted in 5% bovine serum albumin/TBS Tween was applied to the membrane in sealed pouch and incubated overnight at 4°C on a shaker. The membrane was then washed a further three times for 10 mins each in TBS Tween to remove any unbound primary antibody. The secondary

antibody (Table 2.34) was diluted in 5% bovine serum albumin/TBS Tween and incubated with the membrane for one hour at room temperature. The membrane was washed three times for 10 mins in TBS Tween to remove any unbound secondary antibody; the membrane was stored in TBS Tween while reagents are prepared for visualisation of the proteins by enhanced chemiluminescence (ECL). Development of the membranes was carried out in a dark room so as not to expose the light-sensitive X-ray films. Solution A and B were prepared (Table 2.35) beforehand, mixed and applied to the drained membranes in the dark room immediately prior to developing. The membranes were incubated for one minute, blotted on filter paper and exposed to X-ray film within a light impermeable cassette for a range of times (usually 30 seconds to 5 minutes). The films were developed on a Curix60 (AGFA) automatic developing machine.

Primary Antibody	**Dilution**	**Species**	**MW (kDA)**	**Supplier**	**Secondary Antibody**	**Dilution**	**Supplier**
RAD51AP1	1:1000	Rabbit	37	Sigma	Anti-rabbit, HRP conjugated	1:2000	Sigma
FSTL1	1:1000	Rabbit	50	Sigma	Anti-rabbit, HRP conjugated	1:2000	Sigma
mTOR	1:1000	Rabbit	289	Abcam	Anti-rabbit, HRP conjugated	1:2000	Sigma
DEPTOR	1:1000	Mouse	46	Abcam	Anti-mouse, HRP conjugated	1:2000	Sigma
rictor	1:1000	Mouse	192	Abcam	Anti-mouse, HRP conjugated	1:2000	Sigma
raptor	1:1000	Rabbit	150	Abcam	Anti-rabbit, HRP conjugated	1:2000	Sigma
B Actin	1:1000	Mouse	42	Source Bioscience	Anti-mouse, HRP conjugated	1:2000	Sigma

TABLE 2.34: **Western Blot antibody details.** Details of antibodies used in Western blotting experiments. MW is the molecular weight.

Solution A		**Solution B**	
Reagent	**Quantity**	**Reagent**	**Quantity**
100 mM Tris (pH 8)	5mL	100 mM Tris (pH 8)	5mL
Luminol	50μL	Hydrogen Peroxide (30%)	3μL
Coumaric Acid	22μL		

TABLE 2.35: **Formulation of solutions A and B for enhanced chemiluminescence.**

2.8 Cell cycle assay

Cell cycle analysis was carried out on the RAD51AP1 transfected cell lines, using a commercial kit, Propidium Iodide Flow Cytometry Kit for Cell Cycle Analysis (Abcam ab139418). Samples for cell cycle analysis were collected from transfected cells at baseline, 48 and 72h post siRNA transfection with RAD51AP1, and scrambled siRNA to act as a control, from 24 well plates. Media was aspirated and collected in a 15mL centrifuge tube (to retain apoptotic and loosely adherent cells), followed by addition of 200μL of TrypLE^{TM} Express (Invitrogen), cells were incubated for 5 mins in a 37°C incubator, before collecting the cell suspension in to the same 15mL tube. Cells were then spun at 500xg in a centrifuge and supernatant discarded. The cells were then washed in 1mL of ice cold PBS, re-spun at 500 g for 5 mins and supernatant discarded. The cells were then fixed in 66% ethanol on ice, by resuspending the cell pellet in 400μL of ice cold PBS, then slowly adding 800μL of ice cold 100% ethanol; this was mixed well by retro pipetting. The samples were then stored at 4°C and processed within 4 weeks.

Samples were run on the ACEA Novocyte^{TM} Flow cytometer; samples were first equilibrated to room temperature, followed by gentle inversion of cells to resuspend the pellet. Cells were pelleted by centrifuging at 500g for 5 mins, aspirating the supernatant. The cells were then washed in 1mL of PBS, and respun at 500g for 5 mins, before removing the supernatant. The pellet was then resuspended in 200μL of 1$\times$ Propidium iodide and RNase staining solution (Abcam). The solution was then incubated in the dark for 30 mins at 37°C. Samples were then placed on ice, and mixed well before running on the flow cytometer. Analysis was conducted using the cell cycle assay function.

2.9 Oncology protein analysis

Further analysis was carried out on RAD51AP1 transfected cell lines to ascertain if cancer specific proteins were affected by the RAD51AP1 silencing. This was done by a commercial kit, Human XL Oncology Array Kit (R&D systems, cat #ARY026). Samples were collected at 60 hours post transfection, as only one sample was possible due to cost constraints; it was felt this would be the optimal transfection time between 48 and 72h. Treated and scrambled control sample were run for 2 of the cell lines only, SKOV-3 and A549 (due to cost constraints), all samples were run in duplicate.

Cell lysates were collected by aspirating media from the cells, and washing with PBS. The PBS was aspirated off. Cells were solubilised at 1×10^7 cells/mL in Lysis buffer 17. The solution was resuspended by pipetting, followed by rocking the lysate gently on a rocker at 2-8°C for 30 mins. The solution was then microcentrifuged at 14,000$\times$g for 50 mins, and supernatant transferred in to a clean Eppendorf and stored at -80°C until processing.

Samples are processed on an array plate, provided. Firstly, 2mL of Array Buffer 6 was added to each well of the of the 4-Well Multi-dish. Array Buffer 6 serves as a block buffer. The membranes containing the antibody is placed in the well, and incubated for 1 hour on a rocking shaker. 500μL of sample were mixed with 500μL of Array Buffer 4 in separate tunes. The final volume was adjusted to 1.5mL with Array Buffer 6. Array Buffer 6 was aspirated from the wells of the 4-Well Multi-dish and the prepared samples added to the membranes, and covered by a lid. The membranes were incubated overnight 2-8°C on a rocking shaker. The membranes were carefully removed and placed in individual plastic containers, with 20mL 1× Wash Buffer, and washed for 10 mins on a rocking shaker, this was repeated for a total of 3 washes. The array was then carefully removed from the container, allowing excess Wash Buffer to drain off, returning to the 4-Well Multi-dish containing the Detection Antibody Cocktail, covering the dish with a lid. The plates were then incubated for an hour at room temperature on a shaking rocker. The membranes were re-washed in 20mL of 1× Wash Buffer, three times on a rocking shaker for 10 mins each. Following the final wash, all the solution was discarded and 2mL of 1× Streptavidin HRP was added to each well, of the 4-well multi dish. The membrane was then removed from the wash container, allowing excess Wash Buffer to drain off, and returned to the 4-Well Multi-dish containing the 1× Streptavidin HRP, closing the lid. The membranes were incubated for 30 mins at room temperature on a rocking shaker. The membranes were re-washed in 20mL of 1× Wash Buffer, three times on a rocking shaker for 10 mins each. The membrane was removed from the wash container, allowing excess Wash Buffer to drain from the membrane by blotting the lower edge onto paper towels. Each membrane was placed on the bottom sheet of the plastic sheet protector provided. 1mL of Chemi Reagent mix was added to the membrane, followed by covering the membrane with the top sheet plastic protector, smoothing out any bubbles. This was incubated at room temperature for 1 min, followed by squeezing out the excess Chemi Reagent. The membranes were wrapped in plastic wrap whilst in the protective cover, and placed in an autoradiography cassette. The membrane was then exposed to X-ray film from 1-10 mins, followed by developing on a Curix60 (AGFA) automatic developing machine.

2.9.1 Oncology array analysis

The membranes were analysed densitometrically using ImageJ software (National Institutes of Health). Optical density (OD) of the spots/bands representing protein expression was measured and a ratio calculated comparing transfected OD by scrambled control OD.

2.10 Microarray

Microarray analysis was carried out on SKOV-3 72h transfected cells and scrambled controls, due to cost constraints. The microarray experimentation was carried out by collaborators at Oxford Brookes University. Experiments were performed in pooled biological replicates. The extracted RNA was quality verified on a 2100 expert Agilent bioanalyzer; all samples had an RNA yield greater than 30μg and RIN values of 10. Samples were amplified and labelled with Agilent Low Input Quick Amp labelling two colour kit and hybridised with a dye swap to an Agilent G3 8plex x 60k gene Human transcriptome microarray and washed to supplier's protocol.

The array slide was scanned at 3μM and 20 bit Tiff file dynamic range on an Agilent Surescan G2565CA microarray scanner at 100% PMT (photomultiplier tube) gain for both red and green lasers, and exported TIFF images. The TIFF images were then aligned to their design files, quality controlled and converted into probe intensity values using the Agilent Feature Extraction Software Ver 12.0.1.1. Using this software this data was probe- and Loess-normalised and then quantile-normalised for subarray variation using Genespring 14.8 GX-Build 7274 (Agilent), and baseline transformed by the median of all samples. Genes showing 2-fold change between groups and Moderated T-Test significance $p\text{-}value < 0.05$ to remove technical variation (no FDR was carried out due to small sample numbers) were enriched and loaded into Pathways Analysis module of Genespring and compared to Wikipathways Jan 2017 release, BioCyc Ver 20.5 and KEGG Release 81.0 annotated pathways to report on significantly enriched pathways.

2.11 RNAseq

RNA sequencing was carried out on lung cancer patient samples only (due to cost constraints). Sampling was carried out on tissue and blood samples of lung cancer and healthy controls. The RNA sequencing was carried out by collaborators at the Wellcome Trust Centre for Human Genetics (Oxford). RNA samples were normalised to 630ng total RNA and the libraries prepared with the Illumina TruSeq Stranded mRNA Library Prep Kit which involves isolation of the polyA containing mRNA molecules using poly-T oligo attached magnetic beads. Manufacturer's instructions were followed with only minor modifications. All libraries were pooled equimolar and sequenced on one lane of HiSeq4000 at 75bp paired end according to Illumina specifications.

2.11.1 RNAseq data analysis

The RNA-seq data was analysed using open source software programs from the Tuxedo suite: namely TopHat2 (Kim et al., 2013) and Cufflinks (Trapnell et al., 2010). The paired

end raw reads were mapped to the human reference genome hg37 (Ensembl 74) using the annotations from GENCODE 19 (Harrow et al., 2012), withTopHat2 under standard conditions. The resulting alignments were filtered for high quality hits using samtools (Li et al., 2009) with a threshold score of 30. Next, we used Cufflinks to assemble the mapped reads into transcripts and quantify their expression levels in the patient and control samples. Finally, we used Cuffdiff, as part of the Cufflinks package, to identify differentially transcribed genes and transcripts between any two states (cancer tissue vs normal, cancer blood vs normal, cancer tissue vs cancer blood, and normal tissue vs normal blood). The RNA-seq reads manipulation was done in collaboration with Dr Cristina Sisu (Brunel University).

2.11.2 *In silico* analysis using online tools

In silico analysis was carried out on various subsets of data. For this we used an online tool, Oncomine (Oncomine, 2018), to determine gene expression levels of existing microarray data for specific genes in cancer and normal tissues, across cancer types. We also used another online tool known as Kaplan Meier Plotter, (Kaplan-Meier Plotter, 2018), this tool also utilises existing microarray data, but for the purposes of gene expression associations on prognosis and overall survival.

2.11.3 Statistical analysis

A Student's t-test was used to asses statistical significance of any changes observed in experiments. An assessment for homoscedasticity of data from each category was made using the F-test. If homoscedasticity was proven, an unpaired, two-tailed Student's t-test was performed to assess significance in all cases as no matched pairs of samples were used. If data were not homoscedastic, an unpaired, two-tailed Student's t-test with Welch's correction was performed to account for variance. All statistical tests were performed using GraphPad Prism® Software (GraphPad Software). p-values were denoted on graphs and interpreted as follows (Table 2.36):

p-value	Denotation
0.01-0.05	*
0.001-0.009	**
<0.0009	***

TABLE 2.36: **The asterisk denotations of p-value on graphs and ranges.**

Chapter 3

Imagestream^{TM} and ClearCell^{TM} analysis of circulating tumour cells in cell lines and the blood of ovarian and lung cancer patients

3.1 Introduction

Circulating tumour cells (CTCs) are cancer cells shown to be present in the peripheral blood of cancer patients. These cells are shed from the tumour and express tumour-specific characteristics (Hiltermann et al., 2012). The passage and adherence of CTCs in peripheral blood is key to the genesis of distant metastases in various cancers and hence crucial to patient outcomes. Emerging evidence supports the clinical utility of CTCs, and there is increasing evidence that CTCs can be used as predictive markers for diagnosis, prognosis, and response to treatment (Fidler, 2003). However, detection of small populations of CTCs within the large number of normal blood cells represents a significant technical challenge (Hou et al., 2013). Conventional cytogenetics, morphology, and flow cytometry have been used to detect circulating tumour cells at a level of 1 in 100; more sensitive methods including immunocytochemistry may detect one tumour cell in 10^5 normal cells (Molino et al., 1991; Hou et al., 2011). However, examination of CTCs has been limited by several factors including biological heterogeneity and the lack of surface signals on CTCs as opposed to static cells (Gorges et al., 2012). Most assays established for CTC enumeration rely on the expression of the cell surface marker epithelial cell adhesion molecule (EpCAM). However, these approaches may not detect CTCs that express no/low levels of EpCAM, e.g. by undergoing epithelial-to-mesenchymal transition (EMT; (Schneck et al., 2015)). Other methods include size based isolation, as CTCs were shown to be larger than typical blood leukocytes, in studies by Vona et al., 2000 and Dent et al., 2016.

Developments in imaging flow cytometry has enabled the technology to overcome the

earlier hurdles of acquiring sufficient fluorescence sensitivity, leading to the production of high spatial resolution imaging, the ability to combine fluorescence imaging with other imaging modes, and to image all of the cells in flow (Basiji et al., 2007). The Imagestream Mark II^{TM} is an advanced imaging flow cytometer that combines microscopy with flow cytometry. The Imagestream Mark II^{TM} is capable of producing 60,000 images of 10,000 cells in approximately 30 seconds. The IDEAS software analyses captures images of cells as it compares cellular morphology, fluorescent signal strength, signal locations and other aspects. The device is equipped with a software package that provides 40 quantitative features per image, allowing for approximately 250 features to be analysed per cell (Basiji et al., 2007).

Other recent developments in technology include the Clearbridge ClearCell^{TM} FX system device (Clearbridge Biomedics, Singapore). The platform involves the use of spiral microfluidics to separate and enrich CTCs from whole blood based on cell size. Smaller haematological cells (red blood cells and leucocytes), 8-15μm are affected by the Dean drag force and migrate to the outer walls, whereas larger CTCs (15-20μm) are subjected to stronger inertial lift forces and are focussed along the microchannel inner wall (Hou et al., 2013; Chudasama et al., 2016), resulting in a concentrated cell suspension. Previous studies using microfluidic based platforms have reported positive results, with high sensitivity and specificity. Moreover they report minimal cell loss due to no or little pre-processing required. The resulting sample is also available as an enriched CTC suspension that can be further utilised for genomic analysis (Nagrath et al., 2007; Maheswaran et al., 2008; Xu et al., 2017).

Overall isolation, detection, and characterisation of CTCs has proved challenging over the years. Despite the development of many CTC isolation platforms, all have shown some shortfalls; hence none have been approved for clinical NHS practice.

3.2 Aims and objectives

- To evaluate the ability of the Imagestream^{TM} to detect and quantify CTCs using SKOV-3, MDAH-2774, and A549 cell lines as initial *in vitro* experimental models.
- To assess the efficacy of quantifying CTCs in clinical ovarian and lung cancer blood samples using the Imagestream^{TM} technology.
- To assess the efficacy of quantifying CTCs in clinical lung cancer patients using the ClearCell^{TM} device.
- To extrapolate CTCs results and correlate with patient CA125 levels (ovarian only), and patient outcomes and survival.

3.3 Results

3.3.1 Imagestream^{TM} analysis of SKOV-3, MDAH-2774 & A549 cell lines

We conducted *in vitro* proof of principle experiments to ascertain whether we could stain SKOV-3, MDAH, and A549 cancer cell lines with the AE1/AE3 antibody, and assess the efficiency of this staining. Cells were grown in culture, and counted before conducting the experiment. Both SKOV-3 and MDAH-2774 were used as ovarian models, as they portrayed different subsets within this cancer (clear cell adenocarcinoma and endometriod carcinoma respectively), and would allow for any differences to be identified.

The cell pellets were stained with AE1/AE3 antibody, and tagged with a secondary Alexafluor® 488 (AF488) antibody, emitting a green fluorescence. AE1/AE3 is a pan-cytokeratin marker used widely in NHS histopathology laboratories as part of routine testing in cancer (Travis, 2012). As a cytokeratin marker, a positive stain on cells would resemble a green halo like staining around the entire cells surface. Once samples were stained they were run through the Imagestream^{TM} flow cytometer. The machine is comprised of 12 channels; for the purposes of this experiment the following channels were used as described in Table 3.1.

Chanel	**Image**	**Colour**
Channel 1	Brightfield image	Brightfield
Channel 2	AE1/AE3 image	Green (AF488)
Channel 5	Draq5 nuclear stain	Red (642)

TABLE 3.1: **Stains used and relevant Imagestream^{TM} channels.** Details of the different stains used in the Imagestream analysis, including AE1/AE3 a broad cytokeratin marker and DRAQ5, a nuclear marker, and the respective channels the florescence is captured.

Upon cell capture with the Imagestream^{TM}, a scatterplot is produced (see Fig 3.1), based on aspect ratio and area of cell captured, to allow identification of populations of cells. Here, the further left are beads from the machinery, cancer cells (SKOV-3 cells) are to the right, while the far right cells are most likely clumps of multiple cells.

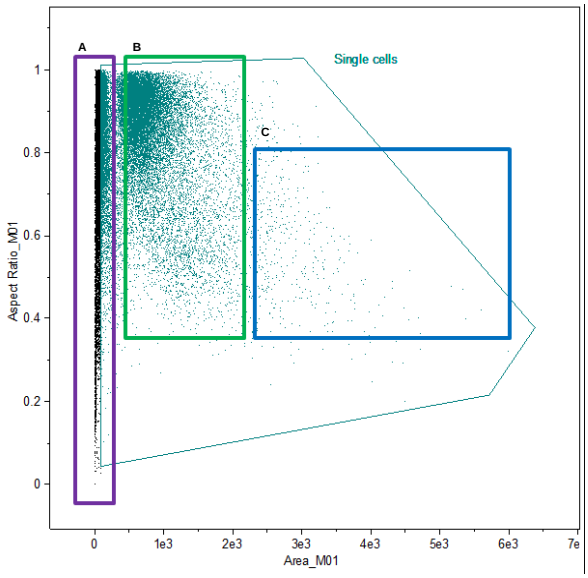
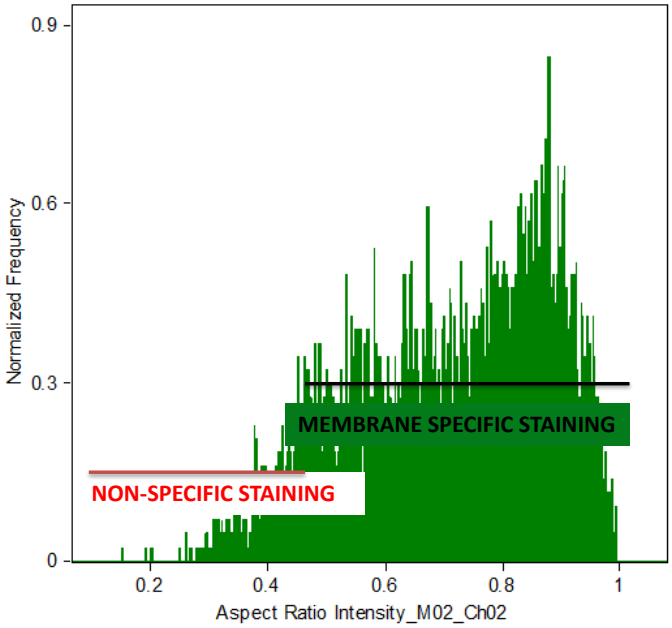


FIGURE 3.1: **Gating of single cells captured during Imagestream^{TM} analysis.** Scattter plot portraying the gating of single cells captured by the Imagestream^{TM} analysis. SKOV-3 cancer cells were run analysed with the system as a proof of principle experiment to ascertain if single cancer cells could be detected. A - represents beads and debris of much smaller density and size; B - represents single SKOV-3 cells; C - shows larger clusters of SKOV-3 cells.

Once the cells of interest were gated, unfocused cells (images of cells that appeared distorted) were excluded from further analysis. The software allows various parameters to be set to group and identify cell populations, and allow enumeration. For the purposes of this study cells were grouped by membrane specific staining of AE1/AE3 using AF488 with positive DRAQ5 nuclear staining (positive CTCs), and non-specific staining (negative CTCs), as is seen in Fig 3.2, and images of CTCs in Fig 3.3. Automatic quantification's was done for the gated regions using the available software, while gating was done manually by viewing cells and staining.



Population	Count	%Gated	Mean	Median	Std. Dev.	Geo. Mean
Focused cells & Single cells	4370	100	0.7127	0.7402	0.1694	0.6897

FIGURE 3.2: **Quantification of AE1/AE3 positive staining in SKOV-3 cells, measured by the Imagestream^{TM}.** Quantification was done following gating of relevant single cells and interrogation staining using the IDEAS software. Charts were constructed for quantification purposes and data tabulated below. Similar charts were constructed for A549 and MDAH-2774 cell lines using AE1/AE3 antibody, with similar patterns of staining. Positive staining was defined as cells with a green halo encircling the entire cell, and a nucleus, depicted by a positive DRAQ5 stain, red.

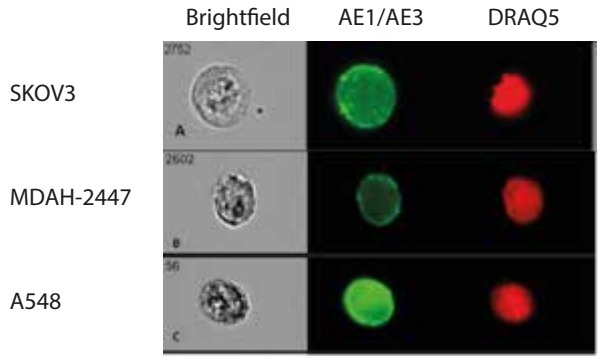


FIGURE 3.3: **Positive AE1/AE3 staining of SKOV-3, MDAH-2774 & A549 cells.** Positive AE1/AE3 staining, is signified by a deep green halo encircling the entire cells cytoplasm. Positive CTCs are depicted as having a positive AE1/AE3 staining in addition to positive DRAQ5 (red) staining of the nucleus This figure portrays positive staining of these cancer cells, for all 3 cell lines. 40× magnification.

Negative control experiments were also run alongside, whereby all 3 cell lines were stained as described, omitting the primary antibody (AE1/AE3), Fig 3.4.

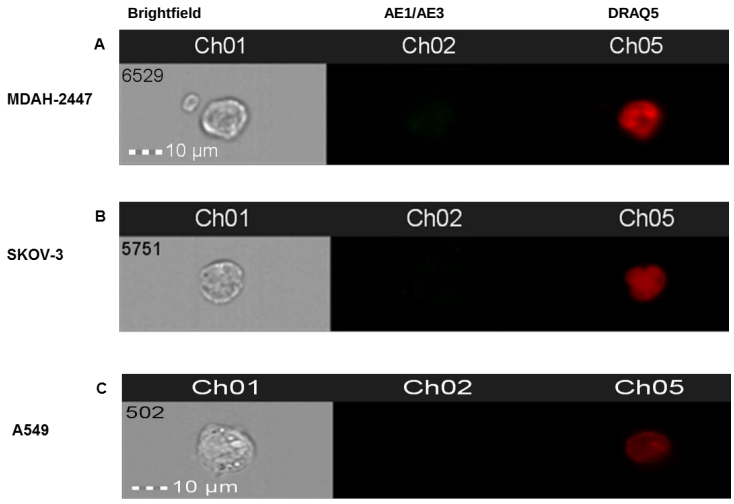


FIGURE 3.4: **Negative controls for AE1/AE3 in SKOV-3, MDAH-2774 & A549 cells.** A negative control experiment was also run to prove results seen in previous experiment are valid. The experiment was run in the absence of a primary antibody, no staining is seen for the AE1/AE3 antibody, signifying a successful negative control experiment. The cell nucleus can be seen in red from DRAQ5 staining. 40× magnification.

Initial results of the three cell lines independent staining with AE1/AE3 proved promising. Next, we focused on developing a realistic model, using whole blood spiked individually with the 3 cancer cell lines, to evaluate the staining in the medium of blood mimicking a liquid biopsy, and the ability to differentiate cancer cells from white blood cells. Approximately 200,000 SKOV-3, MDAH-2774, and A549 cells were spiked in to 1mL of whole blood. Initial experimentations appeared to be inconclusive, with very faint to no staining at all. To address this issue, we increased the primary antibody incubation time from 1h to overnight in order to eliminate the potential effects of other blood components (i.e. leukocytes, erythrocytes) on the staining. The overnight incubations were successful, and all subsequent experiments were conducted with a primary overnight incubation. The ability to separate cell populations upon cell size (with CTCs typically being much larger than WBCs) (Fig 3.5) and staining of AE1/AE3 (Fig 3.6) were assessed.

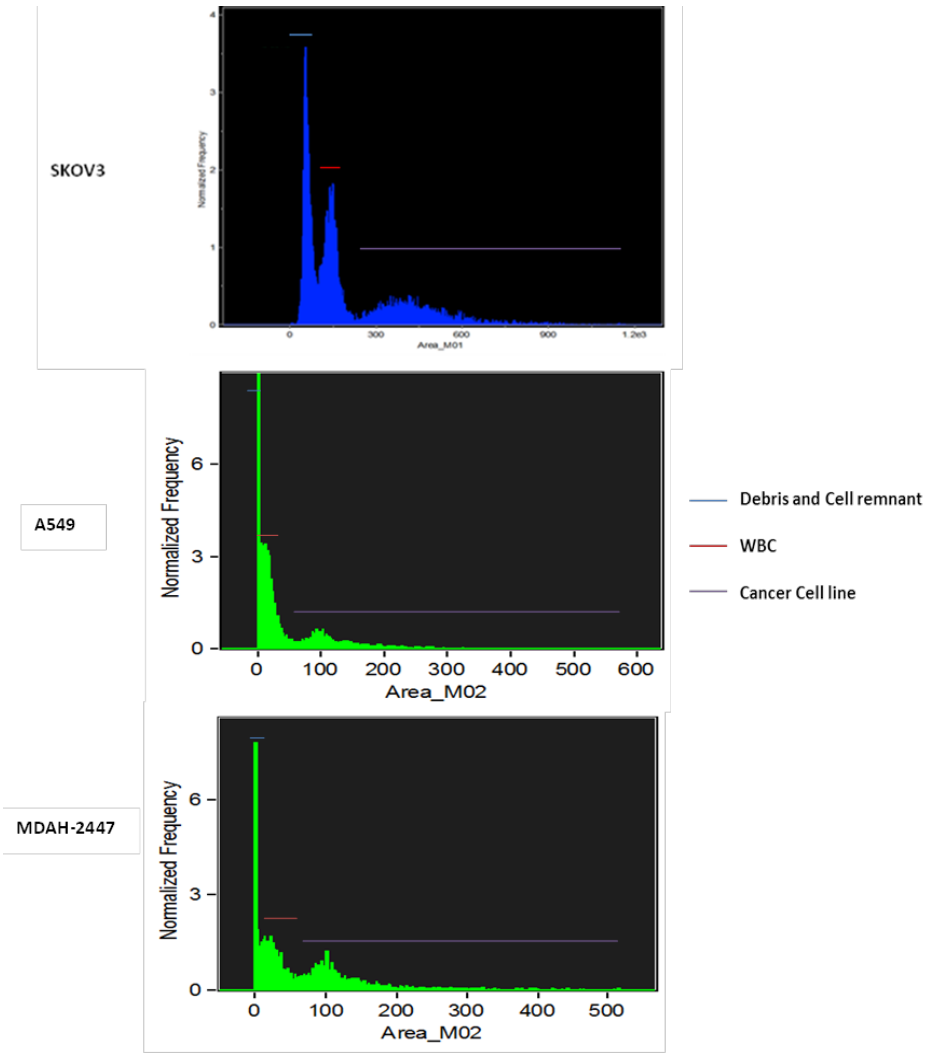


FIGURE 3.5: **Separation of the cell population based on cell size.** Separation shows 3 distinct populations, as can be seen labelled on the chart as debris, white blood cells (WBC) & SKOV-3, MDAH-2774, and A549 cells respectively. Debris is present in the form of beads from the machine and cell debris and remnants. Beads and cell debris would be expected to be smallest in aize, followed by WBCs, with cancer cells being the largest.

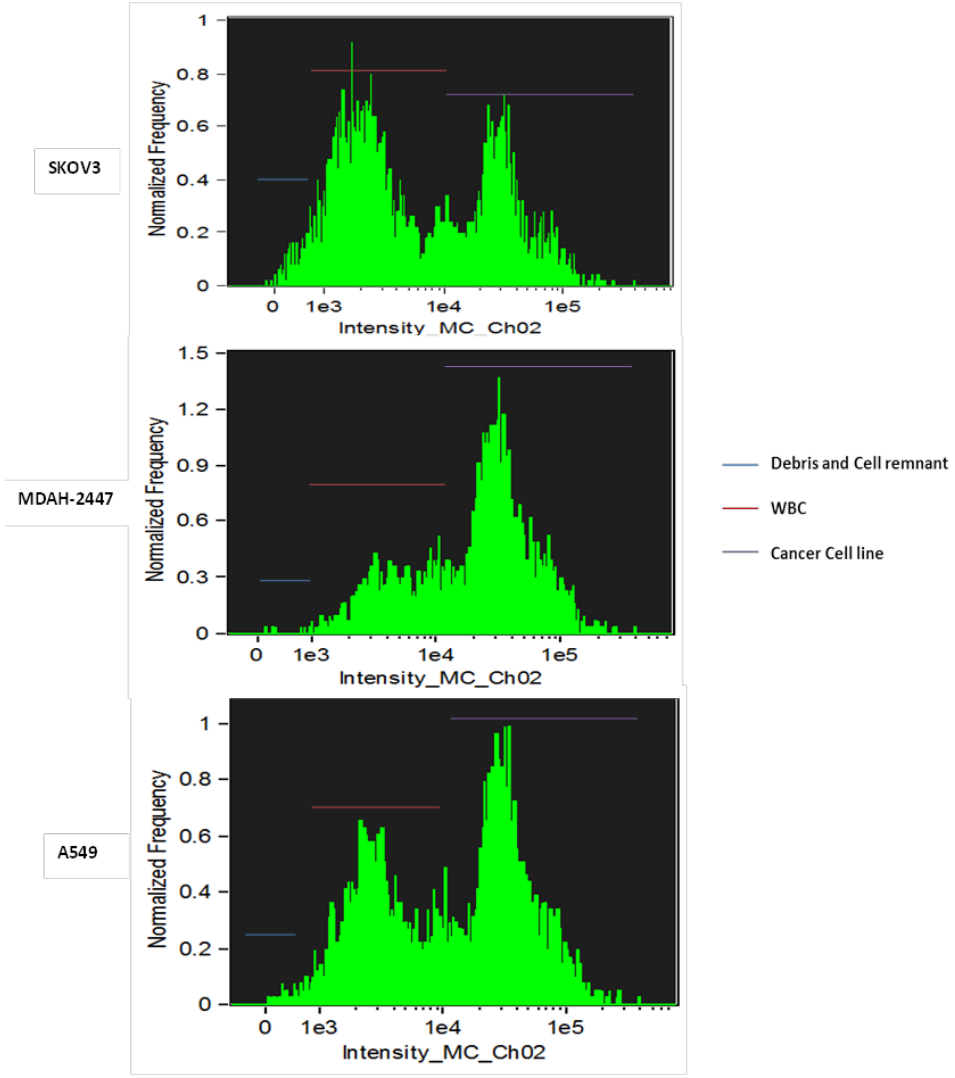


FIGURE 3.6: **Separation of the captured cell population based on cell size.** Cell populations, were measured by intensity of AF488 staining of the AE1/AE3 antibody in channel 2. The 3 populations highlighted in the charts were identified based on staining pattern, as labelled above.

The spiking experiments allowed us to differentiate cell populations based on both cell size and staining. Observations were expanded to look at cell retrieval and cell loss during the process and to evaluate efficiency of cell capture. Varying levels of cancer cells were spiked in whole blood and cell retrieval was quantified as detailed in Table 3.2.

	AE1/AE3 stained cells approximate retrieval		
Total cells spiked in to 1mL	**SKOV-3**	**MDAH-2774**	**A549**
200,000	3.58% (7,168)	3.76% (7,520)	3.48% (6,958)
20,000	7.12% (1,421)	5.85% (1,169)	9.43% (1,885)
2000	10.70% (213)	28.10% (561)	38.60% (771)
200	32.50% (65)	48.50% (97)	41.00% (82)
20	20.00% (8)	60.00% (12)	55.00% (11)

TABLE 3.2: **Quantification of positive CTCs.** Quantification of positive CTC cells was made using the enumeration features of the Imagestream^{TM}. A cell retrieval experiment was carried out on average cell capture from spoking in experiments to ascertain cell recovery, from a total of 200,000, 20,000, 2000, 200, and 20 cells across all 3 cell lines. Findings showed a recovery of 3.61%, 7.48%, 25.8%, 40.7%, and 45% respectively.

Retrieval of cancer cells amongst the 3 different cell lines is relatively consistent, throughout the different number of spiked-in cells. Interestingly, the general consensus shows poor cancer cell retrieval when a high number of cells are spiked in. We observed a cancer cell retrival as low as below 10% across all 3 cell lines when 200,000 and 20,000 cells were spiked in. By contrast, the highest cell retrieval is seen at 60% in MDAH-2774 cells when only 20 cells are spiked in. While these results do not follow previously observed trends, they suggest that a smaller number of spiked cells can be used successfully for further experiments. Thus we proceeded with subsequent analysis in a clinical setting using patient samples.

3.3.2 Imagestream^{TM} analysis of ovarian cancer blood samples

Patient blood samples were collected from Mount Vernon Hospital, from ovarian cancer patients on the METRO-BIBF trial. Details of collection and processing can be seen in flow diagram below (Fig 3.7).

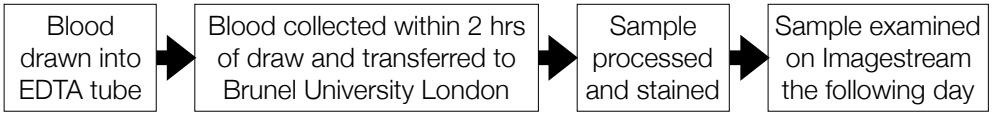


FIGURE 3.7: **Ovarian cancer blood sample processing.** Flowchart detailing the steps in the collection and processing of the ovarian cancer blood sample.

All patients were chemotherapy free for one month before commencing the trial, as per required guidelines. Blood samples were taken from ovarian cancer patients (n=22, and control samples taken from female volunteers (n=16). All blood samples were processed within 4 hours of blood draw. Results can be seen in Fig 3.8.

Staining discrepancies were observed in some cells, particularly for the AF488 staining for AE1/AE3 as can be seen in Fig 3.9. These cells were discarded from all analyses as non-specific staining.

All ovarian cancer and normal blood samples were analysed through the Imagestream^{TM} software in the same fashion while applying the same selection criteria, positive cytoplasmic staining throughout of AE1/AE3 antibody, positive staining of DRAQ5 nuclear stain, and CTCs quantification (Fig 3.10A). Quantification was calculated as positive cells per 10,000 cells rather than per sample volume (i.e. 1mL), in order to by-pass the variability nature of the sample volume of the flow cytometry acquisition (see Chapter 2). Diagnostic tool evaluation was assessed by Receiver Operator Characteristic (ROC) curve, and Area Under the Curve (AUC) as a means of sensitivity and specificity testing, as can be seen in Fig 3.10B. As the ovarian patients were on the METRO-BIBF and CICATRIX trial, multiple samples were taken in most cases at different points of treatment. Overall survival data was also calculated, and can be seen in Fig 3.11, while the median survival data is summarised in Table 3.3.

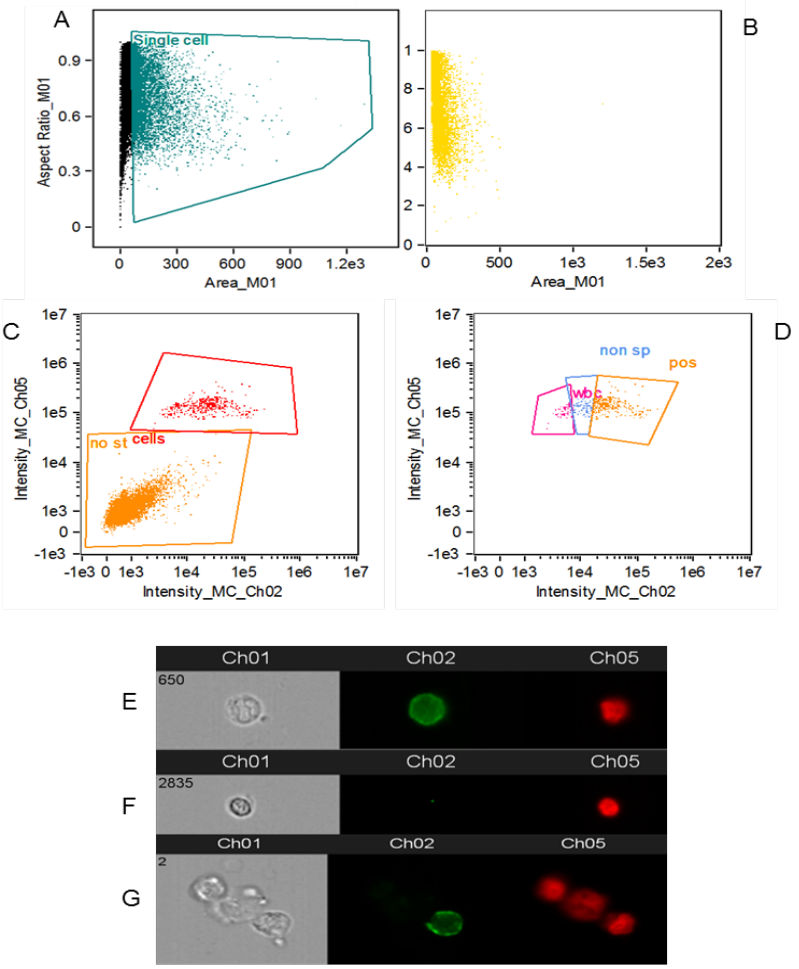


FIGURE 3.8: **Scatter plots of processed patient samples, based on cell size.** (A) population of an ovarian cancer sample shows much wider spread and presence of larger cells, whereas (B) cells are more confined to left hand region, suggestive of generally smaller non cancer cells in the entire population. CTCs were identified by intensity of AF488 AE1/AE3 staining, non-specific cell/debris were easily distinguished at the bottom left corner with very minimal AF488 staining, these cells were excluded from the analysis. The population of cells in top middle consisted of CTCs, WBCs and non-specifically stained cell, upon interrogation of the cells visually these areas could be gated and quantified. (C) captured cells from patient sample based on AF488 staining in a scatter image generated by the Imagestream^{TM} , (D) further interrogation of patient sample A to differentiate CTCs, to WBCs and other non-specifically stained cells. Captured images of cells in the ovarian cancer sample of (E) positive CTC (positive cytoplasmic staining of AE1/AE3 & DRAQ5 nuclear staining), (F) WBC (negative for AE1/AE3 but positive for DRAQ5), (G) positive CTC attached to WBCs. 40× magnification.

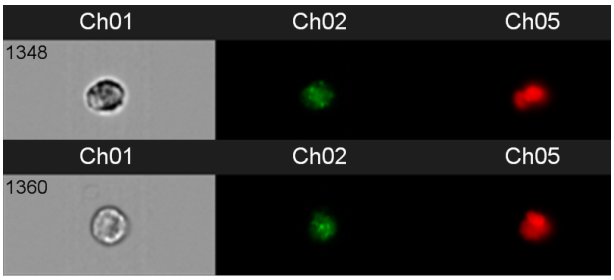


FIGURE 3.9: **Non-specific staining pattern of the AF488 for AE1/AE3 in cells captured from patient samples.** This is depicted by a dark green staining pattern covering the entire cell, rather than a green halo as seen in positive images, 40× magnification.

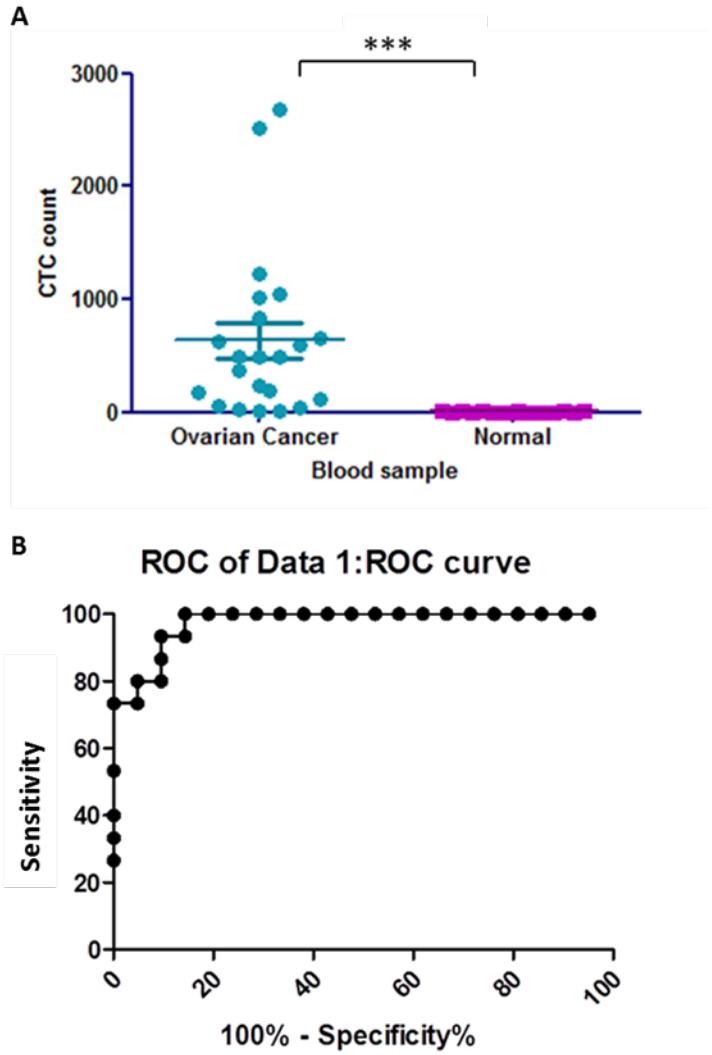


FIGURE 3.10: **Ovarian cancer patients blood CTC quantification.** CTCs were quantified in the blood of ovarian cancer patients and normal disease free females using the Imagestream^{TM} , with AE1/AE3 antibody staining. CTC quantification was done per 10,000 cells. An F-test was performed to assess the variance between the two groups, and two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance. (A) The Ovarian cancer group showed substantially increased positive CTCs in comparison to the normal group, and was shown to be statistically significant with a *p-value>0.0006*. (B) ROC curve analysis was used to measure sensitivity and specificity, an AUC of 0.97 (*p-value<0.0001*) was calculated indicating strong sensitivity and specificity.

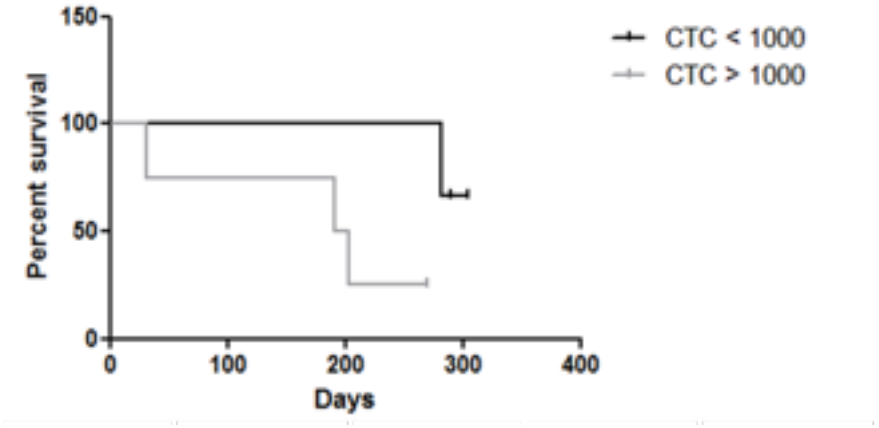


FIGURE 3.11: **Overall survival.** Overall survival was measured by Kaplan Meier plot, patients with an excess of 1,000 CTCs in 10,000 cells showed a poorer overall survival compared to those with less than 1,000 CTCs, Chi-squared value 3.18 (*p-value=0.07*), Hazards ratio 0.12 (95% confidence interval 0.013-1.23), however this was not shown to be statistically significant.

CTC value	Median Survival (Days)
<1,000	290
>1,000	197

TABLE 3.3: **Median survival.** Median survival in days was measured as days of survival post beginning of therapy. Median survival is shown to be longer in the <1,000 CTCs/10,000 cell group compared to the >1,000 CTC/10,000 cells by as much as 93 days.

Enumeration of CTCs in ovarian cancer blood samples revealed higher values when compared to controls (*p-value<0.006*). However, we also observed a large spread of the CTC values in the ovarian cancer group, some of which appear to overlap with the control cohort.

CTC size measurements were also taken throughout the METRO-BIBF (third line ovarian stage III & IV) trial in patients where possible. Here samples were taken at 6 week intervals, for every cycle of the treatment. The CTC load as calculated by the Imagestream^{TM} was compared to the CA125 values at that time (Fig 3.12), obtained from the patients clinical records.

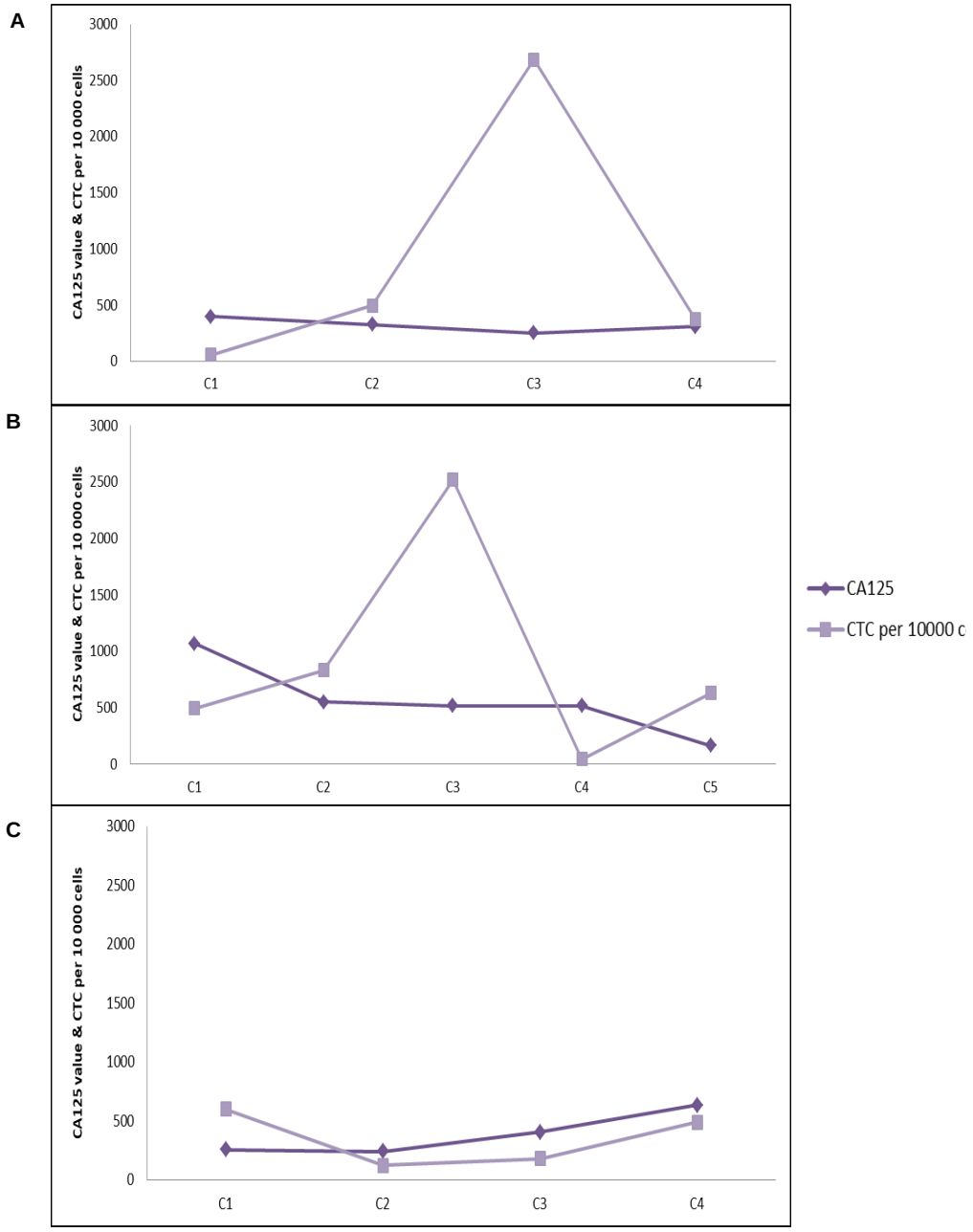


FIGURE 3.12: **Patient blood CTC count comparison to CA125 blood count.** CTC counts from ovarian patient blood samples were compared to CA125 blood counts during the METR-BIBF treatment for individual patients, during each treatment cycle for individual patients. Patients A & B show very little to no correlation, however patient C appears to show a correlation with the CTC and CA125 levels.

We observed little or no correlation between the CTC and CA125 counts. This could be a result of inter-patient variability, and also suggests a lack of specificity for the CA125 biomarker.

3.3.3 Imagestream^{TM} analysis of lung cancer blood samples

CTCs were quantified from blood samples of lung cancer patients (n=21), prior to undergoing surgical resection of their known cancers at Harefield Hosptial (Fig 3.13 & 3.14). Control blood samples were obtained from healthy volunteers (n=21). The same staining criterion was applied as in the ovarian cancer analysis. Samples were obtained from patients with early stage (I-II), late stage (III-IV), and metastatic lung cancer. Due to small 'n' numbers, for statistical analysis purpose, the late stage and metastatic patients were grouped together, as they are both categorised as advanced cases of lung cancer.

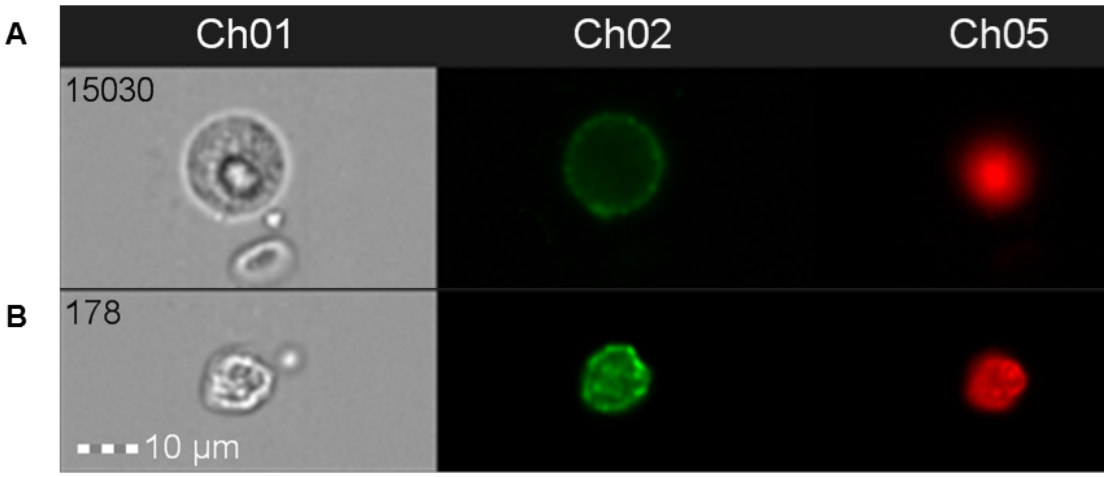


FIGURE 3.13: **Positive CTC image from the blood of lung cancer patient.** A positive cell is depicted by a strong green halo staining around the cell, and a positive nuclear stain see in red, as seen on the images above (A) large positive CTC, (B) smaller positive CTC as visualised by the Imagestream^{TM}, magnification $\times$20.

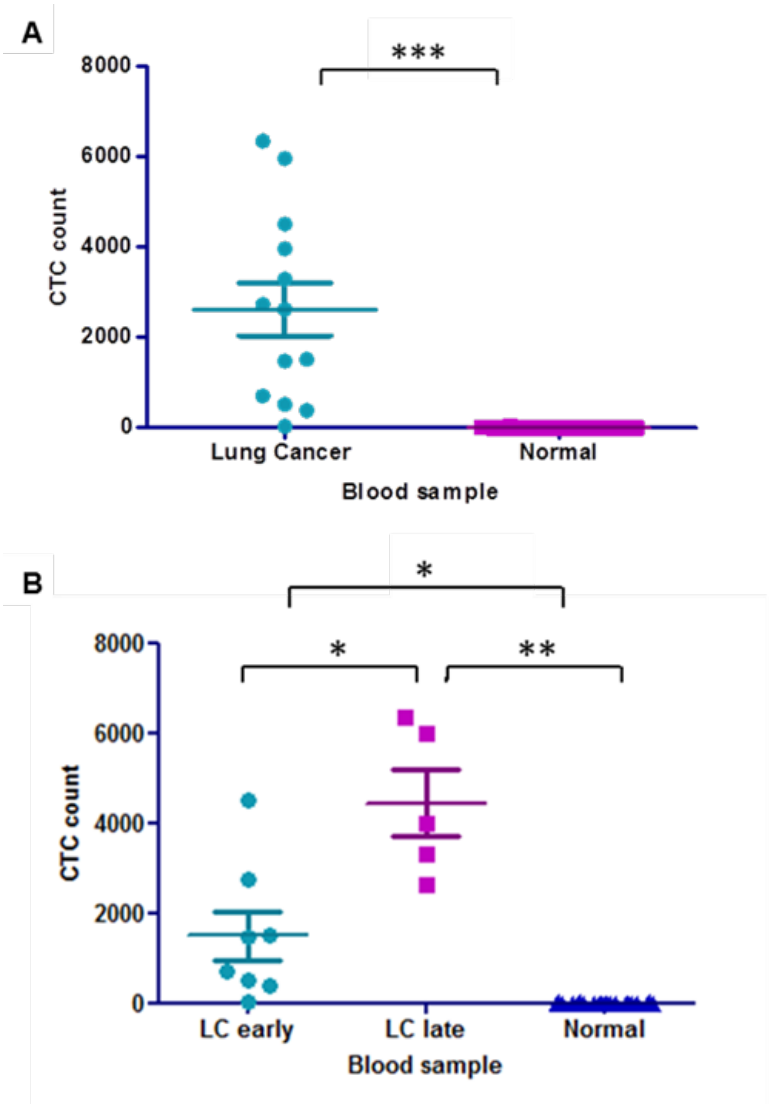


FIGURE 3.14: **Imagestream^{TM} CTCs quantification in blood of lung cancer patients vs healthy volunteers with AE1/AE3 antibody staining.** CTC quantification was done per 10,000 cells. An F-test was performed to assess the variance between the two groups, and two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance. (A) the lung cancer (LC) group showed substantially increased positive CTCs in comparison to the normal group, and was shown to be statistically significant with a *p-value>0.0007*. (B) breakdown by staging and cancer advancement showed higher CTC counts in the late stage and metastatic patients compared to early stage and controls, with statistical significant differences in early stage vs normal, late stage vs normal and early stage vs late stage and metastatic, *p-value=0.0260*, *p-value=0.0038*, *p-value= 0.0139* respectively.

We observed a statistical significant enrichment in the cancer cohort CTC counts compared to normal (*p-value<0.0007*). Further analysis in the lung cancer cohort revealed variations at the staging and cancer advancement level. In particular, the late stage and metastatic patients showed higher CTCs compared to both normal controls and early stage patients (*p-value=0.0038* and *p-value=0.0139* respectively). Next, we assessed the sensitivity and specificity of the method by ROC curve analysis (Fig 3.15), and measured the overall survival by Kaplan Meier analysis (Fig 3.16).

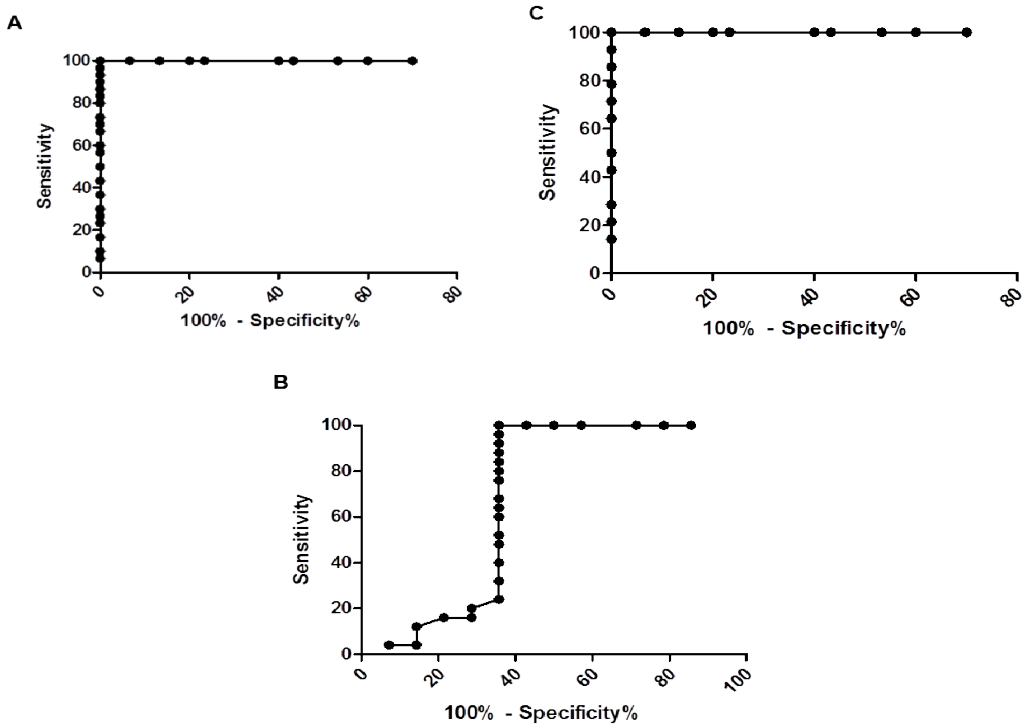


FIGURE 3.15: **ROC curve analysis of the CTC data from lung cancer samples.** Diagnostic tool evaluation using ROC curve analysis of the CTC data from lung cancer samples was done to assess sensitivity and specificity. Strong positive sensitivity and specificity was shown in (A) early stage vs normal, AUC 1.000 (*p-value<0.0001*); poor sensitivity and specificity is seen in (B) early stage vs late stage lung cancer, AUC 0.68 (*p-value=0.06*), in contrast to, (C) late stage cancer vs normal for which we calculated a perfect AUC 1.000 (*p-value<0.0001*).

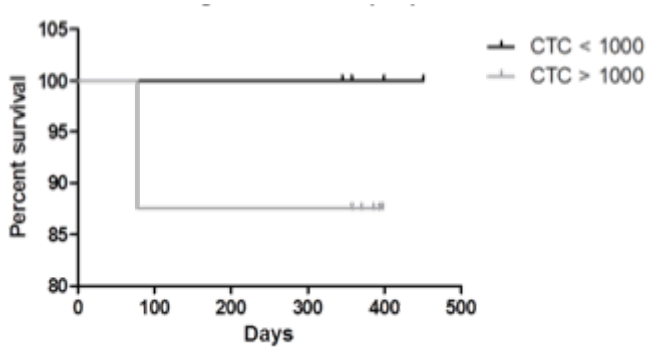


FIGURE 3.16: **Overall survival plot based on blood CTC levels.** Overall survival was calculated using a Kaplan Meier plot, Chi-squared value of 0.50, *p-value=0.46*, Hazard ratio 0.22 (95% confidence interval 0.0036-14.2). Data appears to show poorer overall survival in patients displaying >1,000 CTCs in their blood samples, data is not shown to be significant but there is a trend emerging based on CTC enumeration.

Overall the results suggest there is a general trend where higher CTC counts are associated with more advanced cases, and also a poorer overall survival (although survival data analysis was not statistically significant). Median survival analysis (measured in days), however suggests that there is no difference in survival based on CTC counts (Table 3.4).

CTC value	Median Survival (Days)
<1,000	378
>1,000	378

TABLE 3.4: **Median Survival in days in lung cancer patients.** Median survival is shown to be the same, regardless CTC counts being < or > 1,000 CTCs/10,000 cells.

3.3.4 Analysis of Imagestream^{TM} data based on size of CTCs

Following the clinical validation experiment, we extended our analysis to look at specific characteristics of the captured cells, in particular at their size. Images of captured cells positive for AE1/AE3, showed a variation in size for both lung and ovarian cancer groups (Fig 3.17). Thus we compared lymphocyte size properties to those of CTCs to evaluate the level of overlap in size.

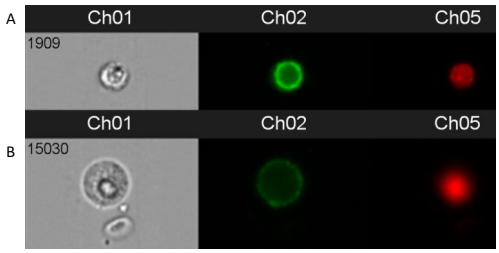


FIGURE 3.17: **Positive CTCs for lung and ovarian cancer samples.** White blood cells and positive CTCs, were viewed on the Imagestream^{TM} and cell size measured by taking the diameter by using the scale tool available. Results show a small positive CTC (A) with a diameter of 10μm (area = 78.5μm^2) compared to a much larger positive CTC (B) with a diameter of 16.6μm (area = 216.4μm^2). 40$\times$ magnification.

Cell size was measured in all cell lines of AE1/AE3 positive cells and WBC, to observe variations and overlap in cell size amongst cancer cells and WBCs. This analysis was performed for all patient and normal blood samples investigated with the Imagestream^{TM}, by taking the diameter of a fraction of CTCs and WBCs. In patient blood, all cells positive for the DRAQ5 nuclear stain were counted, and separated into AE1/AE3 positive (deemed CTC positive), and AE1/AE3 negative (presumed WBC). Cells with non-specific staining were excluded from the analysis. Average cell size and ranges of all patient samples are recorded in Table 3.5.

	Largest Tumour Cell (μm)	**Smallest Tumour Cell (μm)**	**Largest WBC (μm)**	**Smallest WBC (μm)**
			Ovarian	
Average	13.44	10.28	10.03	7.05
Range	10-18.6	8-11.3	6.7-13.3	4.7-8.6
			Lung	
Average	13.68	10.53	10.06	6.35
Range	11.3-16.7	6.6-13.3	6.6-14.6	4.7-8

TABLE 3.5: **Average and range of cell diameters of CTCs in ovarian and lung cancer samples.** Cell sizes were calculated using the Imagestream^{TM} scale feature. The average cell sizes do not appear to overlap at any point. There is an overlap seen in the smallest tumour cells and the largest WBC in both cancer types, where smallest CTC range from 8-11.3μm and largest WBC range from 6.7-13.3μm, in ovarian cancer samples, and 6.6-13.3μm and 6.6-14.6μm in in smallest CTCs and largest WBCs, respectively, in lung cancer samples.

Cell size (diameter) of the three cell lines are larger than those of WBCs as can be seen in

Fig 3.18A, *p-value<0.0001*. SKOV-3 cells has the largest cells compared to A549 and MDAH-2774 cell lines, although there is considerable overlap of cell sizes amongst the 3 cell lines. MDAH-2774 and A549 cells share a similar distribution of cell sizes, however A549 show the smallest cancer cell sizes between all cell lines.

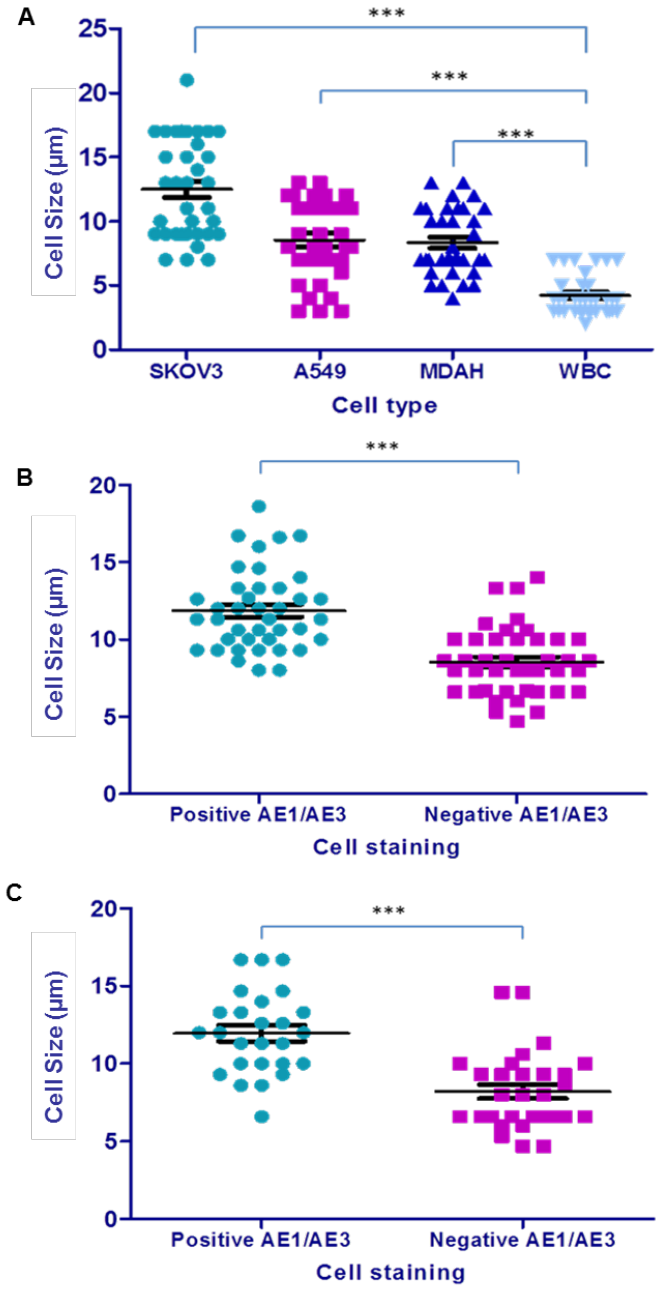


FIGURE 3.18: **Cell sizes.** Diameters of all captured cells in μm, plotted for (A) cell lines, (B) ovarian cancer patient samples, (C) lung cancer patient samples, compared to WBCs. All showed significantly larger cell diameters for CTCs compared to WBCs/negative AE1/AE3 (*p-value<0.0001*), for all samples.

Average cell size of the smallest CTC does not overlap with the average size of the largest WBC captured, cell area 141.9μm^2 and 78.6μm^2, and 147.0μm^2 and 79.49μm^2 in ovarian and lung, respectively, as can be seen with Table 3.5. When comparing ranges, there is an overlap in the measured cell diameter of the smallest CTC and largest WBC in both ovarian and lung samples, cell area ranges 50.3μm^2 to 100.3μm^2 and 35.3μm^2 to 138.9μm^2, and 34.2μm^2 to 138.9μm^2 and 34.2μm^2 to 167.4μm^2, respectively.

3.3.5 ClearCell^{TM} proof of principal experiment

A proof of principal experiment was carried out to test the efficacy of the ClearCell^{TM} device. This was done by using healthy blood donated from 3 volunteers, with each of the 3 samples being split into two, where one sample was left untreated, the other was spiked with 10,000 H358 lung cancer cells. The samples were run through the ClearCell^{TM} device and the resulting cell suspension dried onto a glass slide and stained with haematoxylin and eosin (H&E). The slides (Fig 3.19) were viewed by two independent pathologists and results tabulated in Table 3.6. The results were given as either positive or negative for CTCs.

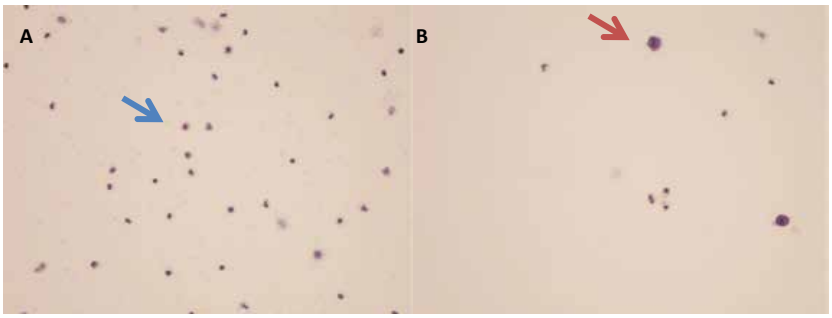


FIGURE 3.19: **H&E stained slides from a positive and negatively identified CTCs.** Preliminary experimentation on voluntary samples, with and without the lung cancer cell line H358 spiked in. Stained H&E slides were viewed and scored by an independent pathologist, who were also blind to the details of the sample. Results show (A) a slide scored as negative fot CTCs, the blue arrow highlighting a WBC on the slide, and no CTCs are visible. This sample was not spiked with the cancer cell line. (B) Shows a slide positive for CTCs, with the red arrow highlighting a CTC. This sample was spiked with the cancer cell line. CTCs are characterised as a large irregular shaped cell, imaged at 20× magnification.

The positive results from the proof of principle experiment prompted clinical validation. In total 51 patients were recruited from the Royal Brompton hospital, consisting of 32 lung cancer patients and 19 non-cancer patients with demographics and pathology summarised in Table 3.7. A total of 7.5mL of blood were taken from each patient and processed through the ClearCell^{TM} device as described in methods chapter 2, and the results are tabulated in

Sample	Phatology Result	Sample	Phatology Result
B1	Negative	B1+H358	Positive
B2	Negative	B2+H358	Positive
B3	Negative	B3+H358	Positive

TABLE 3.6: **CTC positive and negative reporting by 2 independent pathologist of cells isolated using the ClearCell™ device.** All normal samples spiked with H358 cells were reported as positive, whereas all normal blood samples with no spiked in cancer cells were reported negative for CTCs.

Table 3.8. Each reporting group consisted of two independent pathologists, reporting the slides (Fig 3.20) independently. All samples were anonymised.

Variable	Value	Percent
Total	51	100
Mean age (±SD)	57±15.1	-
Males/Females	27/24	52.9/47.1
Pathology		
Primary lung cancer	24	47.1
Adenocarcinoma	17	70.8
Squamous cell carcinoma	6	25.0
Small cell carcinoma	1	4.2
Metastatic Cancer	8	15.7
Non-cancer control	19	37.3
Staging		
I-II	18	90.0
III-IV	2	10.0

TABLE 3.7: **Patient demographics.** Details on demographics and histology of patients recruited for ClearCell™ study.

Sample Type	First reporting team (%)	Second reporting team (%)
All samples, n=51	23 (45.1)	28 (54.9)
Cancer Samples, n=32	22 (68.8)	28 (88.0)
Primary Cancer, n =24	9 (37.5)	15 (62.5)
Metastatic cancer, n =8	4 (50.0)	4 (50.0)
Benign, n =19	9 (47.3)	9 (47.3)
Early stage cancer, n =18	9 (50.0)	12 (66.7)
Late stage cancer, n =2	0 (0.0)	2 (100.0)

TABLE 3.8: **Results as reported by two independent pathologists of blood samples, positive and negative for CTCs.** The first reporting team reported on 23 samples, whereas the second reported on 28. Average reporting of samples positive for CTCs for across both team were 49%, 78.4%, 50%, 50%, 47.3%, 58.4%, 50%, in all samples, cancer samples, primary cancers, metastatic cancers, benign disease, early stage cancer and late stage cancer respectively.

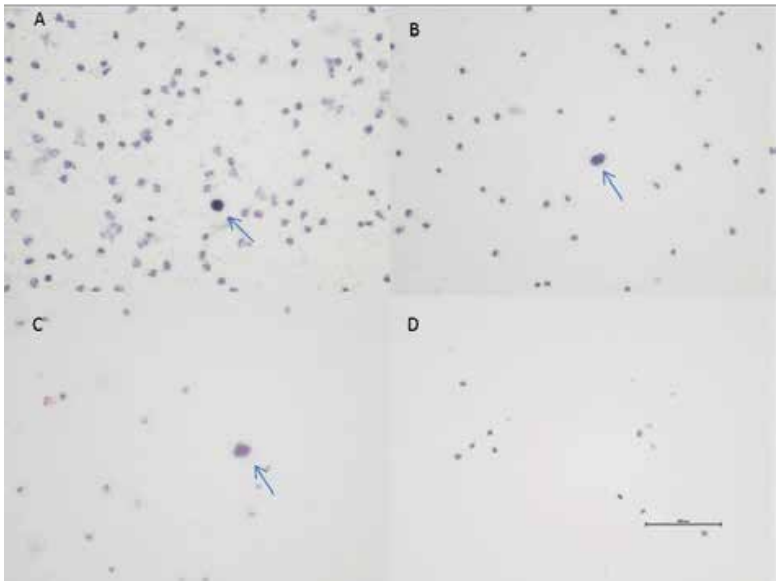


FIGURE 3.20: **H&E staining of patient samples.** Panel of H&E stained slides from 4 patient samples, samples were processed through the ClearCell^{TM} system, and viewed and scored independently by a pathologst as being either negative or positive for CTCs, in a blind fashion to clinical details. Slides with larger cells, with a nucleus were deemed positive, (A-C) positive slides, arrow pointing to a positive CTC, due to large cell, with a nucleus. (D) shows a negative slide with only WBCs, depicted by smaller rounded cells with a nucleus.

CTCs were detected on average in 49% of all cases and 78.4% of all cancer cases, by both reporting teams. In primary and metastatic groups on average 50% of cases were reported positive for CTCs in both cohorts and 47.3% in all benign cases. Positive CTCs were reported

in an average of 58.4% and 50% of early stage (I-II) and late stage (I-IV) cases, respectively. Sensitivity and specificity values were calculated and tabulated in Table 3.9.

	First reporting team	**Second reporting team**
Sensitivity	59% (41 to 76%)	41% (24 to 59%)
Specificity	53% (29 to 76%)	53% (29 to 76%)
Positive predictive value	68% (48 to 84%)	59% (36 to 79%)
Negative predictive value	44% (23 to 66%)	34% (18 to 54%)

TABLE 3.9: **Sensitivity and specificity reporting of the ClearCell™ device.** Sensitivity and specificity values for patient samples reported on were recorded as, 59% and 53%, and 41% and 53% were calculated for the first and second reporting team respectively, using Medcalc online tool.

Sensitivity and specificity values were calculated by reporting team, with values of 59% and 53% for the first reporting team, and a lower sensitivity of 41% and specificity of 53% for the second reporting team. Positive and negative predictive values were also calculated. For assessment of the agreement between the observers, an inter observer agreement was calculated at 80.4%, with a Kappa statistic of 0.6±1.1 (*p-value<0.001*), showing substantial agreement.

3.4 Discussion

This study investigated the efficacy of both the Imagestream™ and ClearCell™ device in the detection of circulating tumour cells from the blood of lung and ovarian cancer patients.

Staining of SKOV-3, MDAH-2774, and A549 cell lines, proved successful with the broad spectrum cytokeratin marker AE1/AE3, using the Imagestream™ platform. This was also the case when these cancer cells were spiked in to whole blood, and could be differentiated in the presence of WBCs with the criterion applied to identify CTCs being AE1/AE3 positive and DRAQ5 positive. This selection criterion is similar to the typical EpCAM positive, CD45 negative, and nucleus positive used by CellSearch (Farace et al., 2011; Miller, Doyle, and Terstappen, 2010; Riethdorf et al., 2007; Cristofanilli et al., 2004). EpCAM was not used, as several studies have shown loss and downregulation of EpCAM in CTCs (Hou et al., 2011; Gorges et al., 2012). Cells negative for AE1/AE3 and positive for DRAQ5 were presumed WBCs.

Cell retrieval experiments, showed poor CTC capture/retrieval of higher spiked-in volumes of cancer cells, compared to smaller volumes. Retrieval was as low as 10% in all cell lines when 20,000 cells were spiked in to 1mL of whole blood. This increased close to 50% when 20 cells where spiked in. However, this behaviour was not observed in SKOV-3. These

results do not follow previously reported trends, and could be attributed to certain system limitations such as poor uptake of antibody in a dense population and cell loss from processing. A study by Dent (2016) published similar cell retrieval rates using the Imagestream, and in a subsequent study by the same group it was reported a 44% cell loss due to processing and a further 6.7% during cell collection, i.e. sample uptake, cells lost/unfocused (Ogle et al., 2016; Dent et al., 2016). This experiment also highlights the difficulty of CTC capture and detection in the context of whole blood, which is echoed in the literature (Hou et al., 2013; Joosse, Gorges, and Pantel, 2015). General CTC detection is challenging as few as only one CTC may be found in the background of 10^5-10^6 peripheral blood mononuclear cells (Hou et al., 2011; Hou et al., 2013; Dent et al., 2016), hence common referencing of CTC capture being likened to 'finding a needle in a haystack'.

Tumour cell plasticity and heterogeneity have also been associated with difficulties in detection. In particular, in the context of epithelial antibody based detection, there is evidence of epithelial marker loss at the cell surface in some cancers, with one explanation for this loss attributed to the controversial epithelial-mesenchymal transition (EMT) (Krebs et al., 2011; Hou et al., 2011). There is growing evidence supporting EMT, as being essential in the metastatic process (Thiery, 2002; Kalluri and Weinberg, 2009). Cancer cells are believed to lose their epithelial properties such as cell-cell adhesion, leading to the transition of firmly attached and immobile cells to become more flexible and mobile as mesenchymal cells. Mesenchymal cells have the ability to invade through the basement membrane and enter the vasculature and travel to distant organs to instigate secondary metastasis (Munzone et al., 2012; Krebs et al., 2011).

Extending the experiments to clinical samples, we were able to detect, characterise and quantify CTCs in both ovarian and lung cancer samples. Data acquired from clinical samples showed varying and in some cases large numbers of captured CTCs. A higher number of CTCs were detected in ovarian cancer samples compared to normal controls, and this was mirrored in the lung patient cohort. When interrogated on staging differences in lung cancer, an increase in CTCs was seen between early (I-II) and advanced and metastatic cases (III-IV). Diagnostic tool evaluation by means of ROC cure analysis, provided strong positive results amongst the ovarian vs normal, and lung cancer vs normal comparison. However, this was not the case when comparing early stage lung cancer with late and metastatic cases. Analysis of overall survival (OS) showed no statistical significance, in both ovarian and lung cancer, despite a poorer OS appearing to be associated with the presence of more than 1,000 CTCs per 10,000 cells in both cancers. These observations are in line with other findings of higher CTC counts being associated with poorer OS in both lung (Shen et al., 2017; Krebs et al., 2011; Hirose et al., 2012) and ovarian cancers (Poveda et al., 2011; Pearl et al., 2014). Median survival in days within the ovarian group shows a similar trend where survival is

shown to be poorer in those with more than 1,000 CTCs/10,000 cells, which is in contrast to the lung group which shows no difference in median survival.

When comparing CTC and CA125 in ovarian cancer patients, no correlation is visible, except for 1 patient (SL in Fig 3.12C). This is unsurprising, as CA125 is known to be non-specific (Poveda et al., 2011), and inter-patient variability is an important confounding factor. Interestingly both patients (Fig 3.12A & B) are known to have responded well to the treatment, which may be signified by the initial spike in CTC counts, eluding to a response (tumour necrosis and thus further CTC shedding), followed by a reduction in CTC counts as the patient continues to do well on the treatment. Patient C, however is known to have not responded to the treatment at all (based on scans and clinical data), and subsequently died quite soon after; one could speculate that the absence of a change in CTC counts (initial spike) as seen with the other 2 cases may signify the lack of response.

Interrogation of cell staining and size showed a variation amongst the cancer types, particularly in the cell size distribution. Non-specific staining was seen in all samples including cell line experiments, and subsequently these cells were excluded from the analysis. Cell size was collected for a proportion of captured cells from both extremes (largest and smallest) and in-between, and plotted. Differences in sizes were seen amongst all the cell lines, this was also shown in a recent study listing mean cancer cell diameters of hepatocellular ($21\pm0.6\mu$m), oesophageal ($17.2\pm0.4\mu$m), thyroid ($16\pm0.36\mu$m) and ovarian cancer ($13.6\pm0.6\mu$m) cells, eluding once more to the heterogeneous nature of cancers, both within themselves and in relation to other cancers, results further echoed by subsequent studies by Dent et al., 2016 and Ogle et al., 2016. A degree of overlap between the cancer cells and WBCs was visible, which was enhanced in the clinical samples. This suggests a cell size differentiation would be futile. Remarking on this, Dent et al., 2016 suggests that the phenotypic criteria for CTC identification varies amongst groups dependant on the technique used, which could potentially lead to varying results and danger of inclusion of non-cellular objects and cellular debris. This data also hints at possible staining issues, where larger cells negative for AE1/AE3, positive for DRAQ5, presumed WBCs, may in fact be CTCs, however, in the absence of any other markers it is impossible to speculate. Similar findings were also reported by Ogle et al., 2016, with similar mean areas of recovered large cells both CTC positive and negative. The authors though concluded this was due to tumour cell heterogeneity, and as the CTC negative cells were positive for the DNA marker and exhibited evidence of hyperploidy, they were of tumour origin (Ogle et al., 2016). The same could not be said for all the larger AE1/AE3 negative cells captured in this experiment. These results also highlight the danger of false positive reporting (larger WBCs i.e. monocytes/macrophages as antibody negative CTCs) or false negative (true CTC negative for antibody as a larger WBC). The use of WBC specific markers, such as CD45 may help to alleviate this. Dent et al., 2016, demonstrated the use of multiple markers including epithelial and tumour specific

markers as well as the WBC marker CD45, to successfully characterise CTCs in a variety of cancer types, concluding the adaptability of the Imagestream^{TM} technique to identify CTCs across many cancers by use of more specific tumour markers.

Data from the ClearCell^{TM} platform show the potential to detect CTCs in clinical blood samples, but in the absence of any quantification tools, the clinical utility remains limited, with results providing only a binary (negative or positive) answer, in the absence of tumour specific markers. The sensitivity and specificity levels in this experiment fall well below the expected level for clinical use, and differences in results amongst the pathologists shed light to a level of subjectivity, despite the substantial agreement calculated. This has been highlighted previously with the use of varying phenotypic criteria to identify CTCs (Dent et al., 2016), particularly in this case where there are no specific cancer/epithelial markers. In contrast, the 'CTC chip', another microfluidic based device utilising EpCAM beads as a marker without the requirement of a pre-processing step, effectively isolating CTCs from WBCs, reported the achievement of high sensitivity and impressive enrichment (proven by purity ranging from 1-80% CTCs to WBCs), in lung cancer and metastatic lung, breast, prostate, pancreatic, and colon cancers (Sequist et al., 2009; Nagrath et al., 2007; Maheswaran et al., 2008). Similarly Xu et al., 2017, used a microfluidic technique, the 'Rare cell sorter' to isolate breast CTCs and compare them to healthy controls, and reported a significant increase in CTCs in the breast cancer group compared to normal. The authors went on to further evaluate the RNA integrity of the enriched CTC population and concluded that the processing was mild and had minimal effect on the RNA integrity for further genomic work (Xu et al., 2017). The enrichment of the CTC population in the ClearCell^{TM} device does offer the prospect of further characterisation work using immunohistochemistry, and importantly genomic analysis, and thus the technique could be further developed and exploited. However clinical application would only be successful if pathologists were able to score in line with one another universally.

The findings from this study show that the use of the Imagestream^{TM} and ClearCell^{TM} for CTC detection and characterisation is possible. Various shortfalls must be overcome before clinical application can be considered, along with further larger sample size testing. Enumeration of CTCs with the Imagestream^{TM}, has the ability to differentiate cancer patients from normal controls. Results were particularly strong in lung cancer; between early stage and late stage and metastatic. CTC quantification for prognosis and treatment monitoring appears to also hold some correlation as was seen in the ovarian cancer treatment monitoring results.

One of the limitations in this study would be the use of one broad spectrum marker for CTC detection. The use of multiple and highly tumour specific markers would allow detection not only of those cells with ambiguous cytokeratin staining but some tumour specific

markers could also identify the source and type of cancer i.e. TTF-1 for lung adenocarcinoma. Due to the nature of this thesis exploring and assessing various techniques, this was not developed further. With the limited data available from the H&E staining of clinical samples, the clinical utility for the Clearbridge^{TM} platform would currently be limited. Genomic analysis of the enriched CTC suspension could be an exciting area of further development.

Collectively, these results show the development of a robust CTC quantification assay could act as a potential non-invasive early diagnostic tool, and could be expanded for prognostic and treatment monitoring purposes.

Chapter 4

DNA and plasma based 'liquid biopsy' approaches for diagnosis and prognosis in ovarian and lung cancer patients: use of Alu repeat sequences and Raman spectroscopy

4.1 Introduction

In addition to circulating tumour cells, there is also evidence of circulating tumour DNA being present in the blood of cancer patients. Various applications have been explored including quantification and genomic interrogation to analyse these markers. There is also emerging research on the metabolomics analysis of plasma from cancer patients, as a potential 'liquid biopsy', these methods are described and evaluated in this chapter.

Typically in healthy individuals apoptosis of cells occurs naturally, where DNA (circulating free DNA or cfDNA) is released into the blood circulation and uniformly truncated at 185-200 bp (Umetani et al., 2006a). Circulating tumour DNA (ctDNA), is tumour derived fragmented DNA, known to originate directly from the cancer itself (Umetani et al., 2006b). ctDNA is found circulating in the blood of cancer patients similarly to circulating tumour cells. DNA fragments released from malignant cells tend to vary in size and are typically >200bp, due to the pathological process of cell death consisting not only of apoptosis, but also necrosis, autophagy, and mitotic catastrophe (Wang et al., 2003). Ineffective nuclease activity is also reported to contribute to longer DNA fragments. The exact nature of how the ctDNA is shed is unknown, but it is postulated to occur as a consequence of apoptosis and necrosis of cancer cells, but also as active release from viable circulating tumour cells (Umetani et al., 2006a; Umetani et al., 2006b).

Genomic profiling of ctDNA has become increasingly popular, based on the hypothesis that ctDNA harbours the same profile of somatic mutations and genomic rearrangements as the tumour itself (Han, Wang, and Sun, 2017; Freidin et al., 2015). This is an additional advantage over CTCs, which may not adequately represent the entire tumour due to its heterogeneous nature (Freidin et al., 2015). Despite ctDNA being fragmented (as DNA is shed into the blood stream) in comparison with the more intact DNA extracted from viable CTCs, ctDNA provides a more comprehensive picture of the overall mutation profile as isolated CTCs are not able to adequately represent the entire tumour (Freidin et al., 2015). Furthermore, capturing of CTCs remains a challenging subject (Yu et al., 2011). To date, a number of successful studies have been performed to validate this concept, in a variety of cancers, including lung, breast, pancreatic, bladder, gastroesophageal, colorectal, melanoma, hepatocellular, and head and neck cancers (Punnoose et al., 2012; Dawson et al., 2013; Bettegowda et al., 2014), by comparing both CTCs and ctDNA. All the studies report higher detection and concordance with tissue results using ctDNA over CTCs.

The percentage of ctDNA originating from tumour cells however, has been estimated to range from 10 to 90% of the total cfDNA. Thus, the applicability of measuring the cfDNA length (i.e. DNA integrity) in plasma may therefore depend on the type of disease (Jahr et al., 2001). Detection of these longer ctDNA fragments and quantification of their relative abundance in plasma compared to short cfDNA fragments and the calculation of a DNA Integrity Index has been explored as a potential cancer monitoring technique (Agostini et al., 2012; Silva Filho et al., 2013). To this end, several studies have shown that the size distribution of the cfDNA can be measured using qPCR on amplified Alu-repeats (Cancer Genome Atlas Research Network, 2011; Umetani et al., 2006a; Umetani et al., 2006b). Classified as Short Interspersed Elements (SINE), Alu is the most abundant mobile element in the human genome (Zhang et al., 2011). A full length element can span approximately 300 base pairs (bp), and include two tandem monomer units, separated by a poly 'A' stretch (Umetani et al., 2006b). Alu-repeat measurements were shown to adequately predict the disease in cases of colorectal, ovarian, and breast cancer, however the presence of pancreatic cancer could not be diagnosed (Umetani et al., 2006a; Umetani et al., 2006b; Wang et al., 2003; Sikora et al., 2015) using the same approach.

To date, various applications have been explored in the analysis of ctDNA including quantification and genomic interrogation.

A more recent and alternative approach to the 'liquid biopsy' field, is the metabolic analysis on plasma and serum samples, by use of Raman Spectroscopy. Typically spectroscopy techniques have been utilised for various applications, e.g. interpretation of tissue samples. A number of optical techniques, such as elastic scattering spectroscopy (ESS) (Rodriguez-Diaz et al., 2014; Bigio et al., 2000), optical coherence tomography (OCT) (Escobar et al.,

2004; Mogensen et al., 2009), fluorescence spectroscopy (Ramanujam et al., 1994; Ramanujam, 2000), fluorescence lifetime imaging microscopy (FLIM) (Verveer, Squire, and Bastiaens, 2000) and diffuse reflectance spectroscopy (DRS) (Zonios et al., 1999; Jayanthi et al., 2011) have been studied for their potential applications in cancer diagnosis. However, all these techniques have limitations, including low specificity (ESS, fluorescence spectroscopy), low penetration depth and low specificity (OCT), low signal-to-noise ratio (FLIM), low accuracy (DRS), and an invasive nature.

Vibrational spectroscopies such as Raman and mid-infrared (IR), are strongly sensitive to the chemical composition of the sample, and have been tested for application in cancer diagnosis using tissue samples as input. Between these two techniques, IR is significantly affected by the presence of water in the tissue, and therefore cannot be used in non-invasive measurements *in vivo*. By contrast, Raman (near-infrared wavelengths) remains largely unaffected by water content, and for that reason it can be used directly in real-time for non-invasive measurements. Due to its high specificity and sensitivity, Raman spectroscopy is currently used in a number of applications including material characterisation, security (Eliasson, Macleod, and Matousek, 2007), pharmaceutical industry (Griffen, Owen, and Matousek, 2015), art history (Volpati et al., 2015), and geology (Edwards, Middleton, and Hargreaves, 2009), and emerging contribution to cancer.

Raman scattering (Raman effect) is a specific case of photon scattering, whereby the photon is scattered in a medium with a change in its energy (inelastic scattering). When incident light falls onto a material surface, it will be typically scattered in random directions. The vast majority of the light will be elastically scattered, maintaining the same energy (i.e. frequency). A tiny amount of light (less than 0.001%) will be scattered with a slightly different frequency from the incident light (inelastic scattering). This difference in the frequency is characteristic of the bond vibrations of the molecules, and provides important chemical and structural information of the sampled material (Feng et al., 2015; Edwards, Middleton, and Hargreaves, 2009; Griffen, Owen, and Matousek, 2015).

To date, this method has been used to study tissue material, however more recently has been explored in plasma, to test its value as a non-invasive 'liquid biopsy' technique. Studies are not only confined to cancer, but include Alzheimer's disease, whereby levels of amyloid β-peptide were successfully detected in plasma samples of patients suffering from Alzheimer's (Huang and Isidoro, 2017). Few studies have explored the use of Raman spectroscopy in plasma sample of cancer patients, to test its efficacy as 'liquid biopsy'. Successful results were obtained by Feng et al., 2015 in colorectal patients, with the authors reporting sensitivity and specificity values of 86.4 and 80% respectively. More recently, Medipally et al., 2017 reported various drawbacks of the classical Raman spectroscopy including sample preparation, equipment and expensive substrates, and long sample acquisition times. As

an alternative, the authors proposed a 'High-Throughput Raman Spectroscopy' platform developed by adapting in use methods and techniques, and successfully reported sensitivity and specificity values of 96.5 and 95% respectively, in prostate cancers. Moreover, the described improvements allowed for high throughput and increased the method's applicability as a routine practice (Medipally et al., 2017), providing evidence of the potential clinical utility of this technique as a 'liquid biopsy'.

4.2 Aims and objectives

- Evaluate the use of Alu repeats and DNA Integrity Index in ovarian and lung cancer patients for diagnostic and prognostic purposes.
- Evaluate use of Raman spectroscopy on plasma samples in ovarian and lung cancers, to explore its utility as a potential 'liquid biopsy' method.

4.3 Declaration of contribution to the work presented in this chapter

The University of Exeter - Raman Spectroscopy

- I collected patient samples and processed for plasma samples. I then organised delivery of these samples to The University of Exeter on dry ice.
- The University of Exeter ran the Raman Spectroscopy experiments. The processed data was then sent to me.
- I then cross checked the data supplied with the patient details to make correlations and further analyse this data.

Table 4.1 gives a detailed description of contributions from each participating unit.

Contributions	**The University of Exeter**	**DC**
Patient and sample collection	-	x
Sample preparation (RNA/DNA/Plasma)	-	x
Sample processing	x	-
Processing of raw data	x	-
Further analysis and correlation with clinical samples	-	x

TABLE 4.1: **Contribution to the work presented in this chapter.** Details on specific tasks performed by The University of Exeter and Dimple Chudasama (DC).

4.4 Results

4.4.1 Alu repeat ratio and DNA Integrity Index in ovarian cancer patients

DNA extracted from 31 ovarian cancer plasma samples on the METRO-BIBF trial throughout their treatment, were processed by qPCR against the Alu repeat primers Alu 115 and Alu 247 (as commonly used in literature Umetani et al., 2006a; Umetani et al., 2006b. Healthy female samples were also collected from 11 individuals, details of patients are listed in Table 4.2. Data was analysed and DNA Integrity Index was calculated as the Alu 247/115 ratio, and plotted as can be seen in Fig 4.1.

Variable	**Value**	**Percent (%)**
Total	42	100
Mean age (±SD)	64±11	-
Males/Females	0/45	0/100
Pathology		
All cancer	31	73.8
Serous Ovarian Carcinoma	31	100
Staging		
III	31	100
Non-cancer control	11	26.2
III-IV	5	22.7

TABLE 4.2: **Information on ovarian cancer patients.** Details on pathology and ovrian cancer staging for patients recruited to this study.

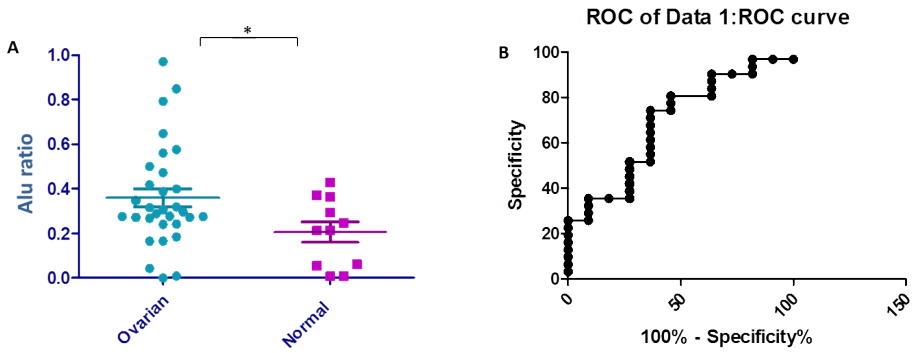


FIGURE 4.1: **DNA Integrity Index in ovarian cancer.** DNA Integrity Index as calculated by Alu247/115 ratio in ovarian cancer samples, shows higher ratios compared to normal (*p-value=0.0194*), considerable overlap exists when compared to normal samples. Reporter operating characteristic (ROC) curve analysis, gives an area under the curve value of 0.69 (**p-value=0.07*), suggesting moderate sensitivity and specificity, however this is not significant.

Results of DNA Integrity Index show higher values are reached in the ovarian cancer group compared with controls (**p-value=0.0194*). While the difference is statistically significant, the average DNA Integrity Index values are close, 0.36 (0.08-0.97) and 0.21 (0.007-0.43) respectively. Diagnostic tool evaluation by ROC analysis shows poor diagnostic utility with AUC of 0.69 and **p-value=0.07*.

Based on these results, we selected DNA Integrity Index value of >0.5 as an indicator of a higher proportion of Alu 115, and consequently a higher tumour DNA burden. The DNA Integrity Index was used to assess impact on overall and median survival times. Overall survival calculation was made using a Kaplan Meier chart as seen in Fig 4.2, by taking the mid-way point and separating patients below and above a DNA Integrity Index of 0.5, median survival in days from trial commencement is tabulated in Table 4.3.

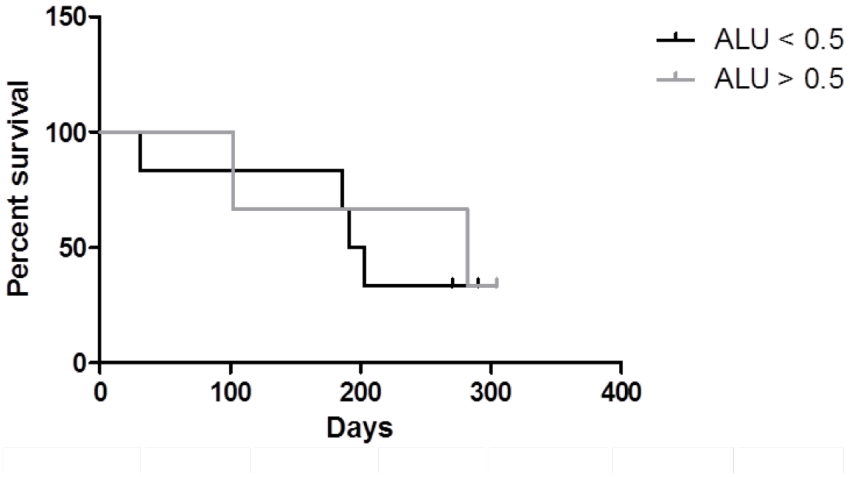


FIGURE 4.2: **Kaplan Meier plot of overall survival in ovarian cancer patients.** A Kaplan Meier was plotted to measure overall survival, Chi-squared value of 0.12, *p-value=0.70*, Hazards ratio 1.34 (95% confidence interval 0.25-7.26). Data shows no difference between the two groups, indicating Alu ratio has no impact on OS.

DNA Integrity Index	Median Survival (Days)
<0.5	197
>0.5	282

TABLE 4.3: **Median survival in days from the start of METRO-BIBF treatment.** Calculation of median survival in days as of 1st March 2017. A higher Alu ratio (>0.5) shown to be associated with longer median survival.

The DNA Integrity Index was also compared to CTC values (see Chapter 3) from the same patient over their treatment cycles and plotted in Fig 4.3. This analysis investigates whether a similar pattern of change is apparent in the two biomarkers over treatment time.

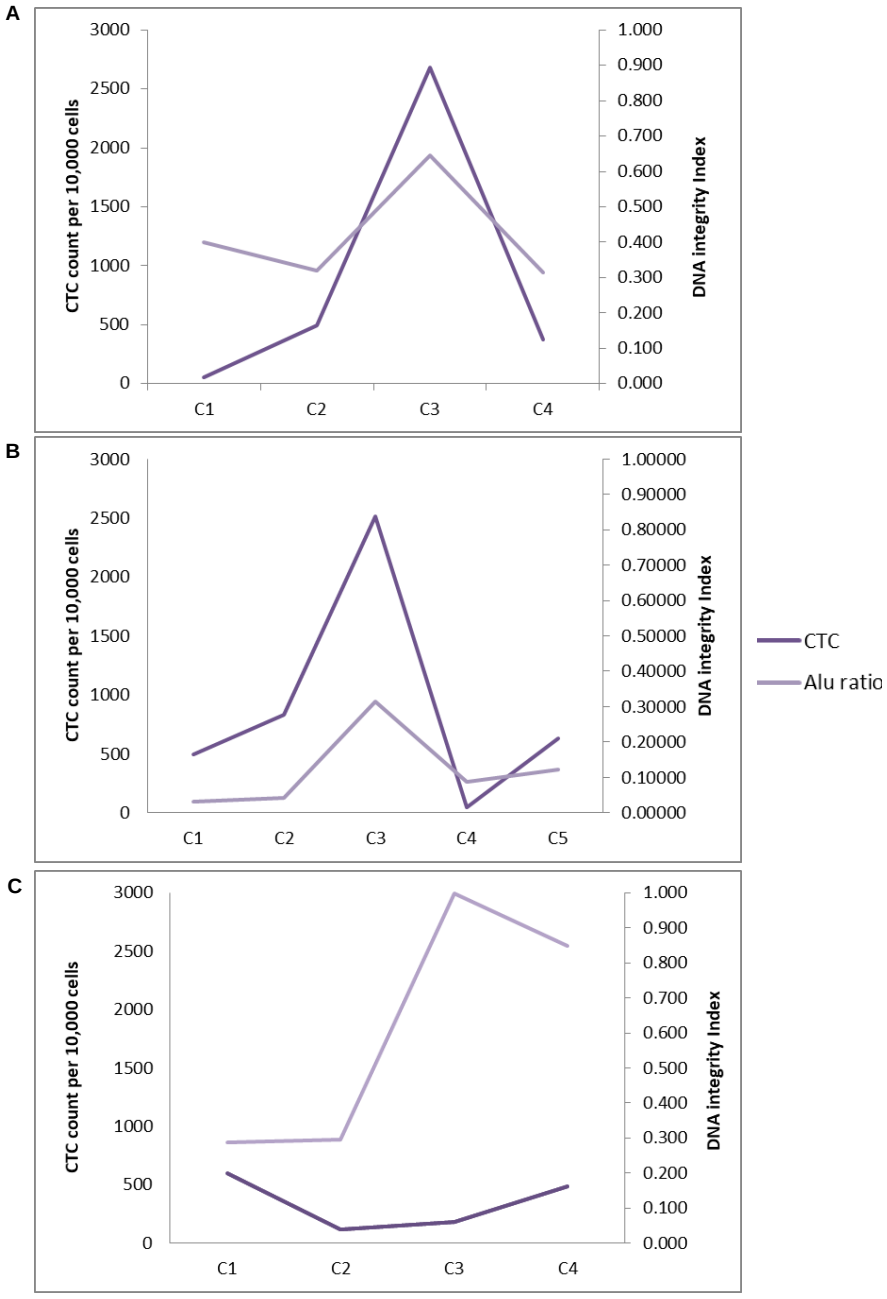


FIGURE 4.3: **CTC vs Alu ratio in ovarian cancer patients.** Observations of the CTC value and Alu ratio in individual patients over the course of treatment shows some correlation in patients A & B, however this correlation does not appear in patient C.

A general trend of little to no change in CTC and ctDNA followed by a spike in both at around cycle 3 is evident in patients A & B followed by a significant drop. This was described earlier in Chapter 3, and postulated as a signature of positive response to therapy. The spike indicates a positive response and hence release of CTC and ctDNA as the tumour may be shrinking/responding, followed by reduction in both biomarkers as tumour burden is lessened. Patient C shows little to no correlation between the 2 biomarkers, the patient is known to have not responded well to the therapy and subsequently died. The variation could be attributed to inter-patient variability, and poor response to therapy.

4.4.2 Alu repeat ratio and DNA Integrity Index in lung cancer patients

The same analysis was carried out on samples from 29 lung cancer patients and 19 non-cancer controls. Details of patients and pathology are listed in Table 4.4.

Variable	**Value**	**Percent (%)**
Total	48	100
Mean age (±SD)	60±15	-
Males/Females	25/23	52.1/47.9
Pathology		
All cancer	29	60.4
Primary lung cancer:	22	75.9
- Adenocarcinoma	15	68.2
- Squamous cell carcinoma	7	31.8
Metastatic Cancer	7	24.1
Non-cancer control	19*	40
Staging		
I-II	17	77.2
III-IV	5	22.7

TABLE 4.4: **Lung cancer study patient information.** Details on patients selected in lung cancer study including pathological staging.* Includes 2 non-cancer patients undergoing bullectomy surgery.

DNA samples were processed as earlier described using qPCR with the same Alu repeat sequences as above. The DNA Integrity Index was calculated and plotted, as can be seen in Fig 4.4. Due to small sample size (n) late stage and metastatic cases were pooled together, as they would be categorised as advanced lung cancers, compared to localised early stage lung cancer. ROC curve analysis and AUC were also calculated to measure sensitivity and specificity of this technique in lung cancer patients, and can be seen in Fig 4.5.

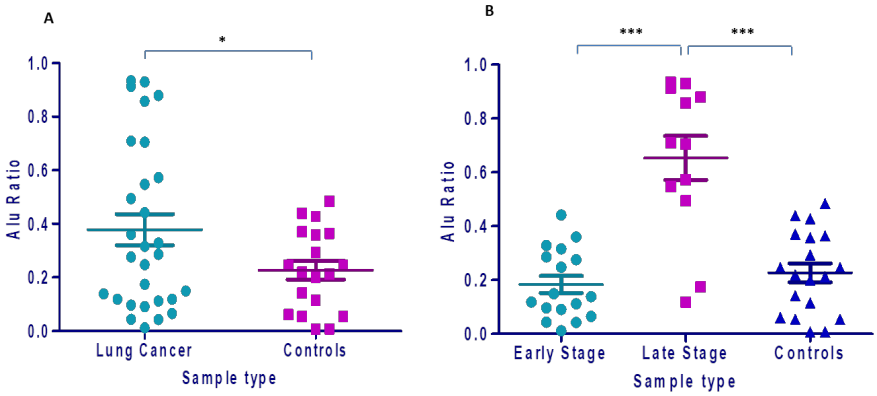


FIGURE 4.4: **Scatter plot showing DNA Integrity Index.** (A) shows an increased DNA Integrity Index lung cancer vs controls, *p-value=0.035*. (B) shows an increased DNA Integrity Index in the advanced cancer group (III-IV including metastatic) compared to both early stage (I-II) and controls, ****p-value<0.0001* and ****p-value=0.0006* respectively.

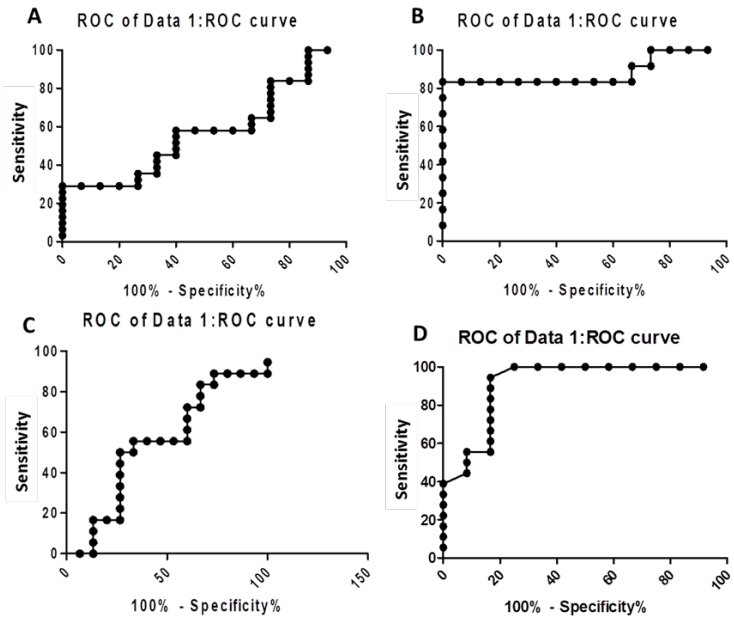


FIGURE 4.5: **ROC curve analysis of diagnostic sensitivity and specificity of using ALU repeat ratios.** (A) All lung cancer vs controls, showed poor sensitivity and specificity AUC 0.61, p-value= 0.22 compared to; (B) Late stage vs controls, AUC 0.88, ****p-value=0.0007* showing much better performance, in contrast to; (C) Early stage vs controls, sensitivity and specificity were poor with, AUC 0.67, *p-value=0.23*; (D) Late stage vs early stage, AUC 0.92, ****p-value=0.0002*, suggests very good sensitivity and specificity as a diagnostic tool.

Patient survival data was also collected and patient status confirmed on 1st May 2017. Patient median survival (Table 4.5), and overall survival was plotted as seen in Fig 4.6, in terms of DNA Integrity Index, with a value of >0.5 used as an indication of higher proportion of tumour DNA presence.

Pathology	**Patients (N)**	**Mean DNA Integrity Index (Range)**	**Median survival**
Primary lung cancer:			
- Early stage (I-II)	17	0.18 (0.01-0.44)	434
- Late stage (III-IV)	5	0.86 (0.71-0.93)	460
Metastatic	7	0.51 (0.12-0.93)	454
Non-cancer controls	19	0.22 (0.007-0.44)	-

TABLE 4.5: **DNA Integrity Index by staging, including median survival times.** Interrogation of DNA Integrity Index by staging show advanced cases and metastatic patients to have a higher Alu repeat ratios, compared to early stage and non-cancer controls, 0.86, 0.51, 0.18, and 0.22 respectively. Median survival times however do not seem to be significantly influenced by a higher DNA Integrity Index.

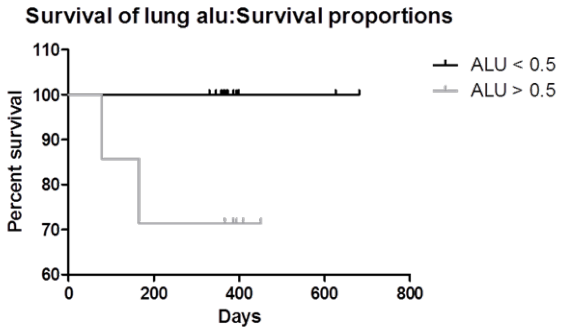


FIGURE 4.6: **Kaplan Meier plot of overall survival in lung cancer patients based on DNA Integrity Index.** Patients with ALU reading >0.5 show poorer survival, Chi-squared value of 5.55, **p-value=0.03*, Hazards ratio 0.023 (95% confidence interval 0.00099-0.53).

Median survival post operation in days was also calculated and tabulated in Table 4.6.

Overall survival is shown to be poorer in patients with a DNA Integrity Index of >0.5, *p-value=0.03*, in contrast median survival in days post-operatively shows little difference between the two groups, with survival increased by 23 days for the <0.5 DNA Integrity Index group.

We further interrogated the data, using known factors for lung cancer that can influence survival. No significant correlations were found between Alu and CT tumour size (r = 0.21, *p-value>0.05*), Fig 4.7.

DNA Integrity Index	Median Survival (Days)
<0.5	434
>0.5	457

TABLE 4.6: **Median survival by DNA Integrity Index.** DNA Integrity Index was measured against median survival, by stratifying the Alu ration as <0.5 or >0.5. Data show a better median survival in the group with a higher DNA Integrity Index, compared to the lower group

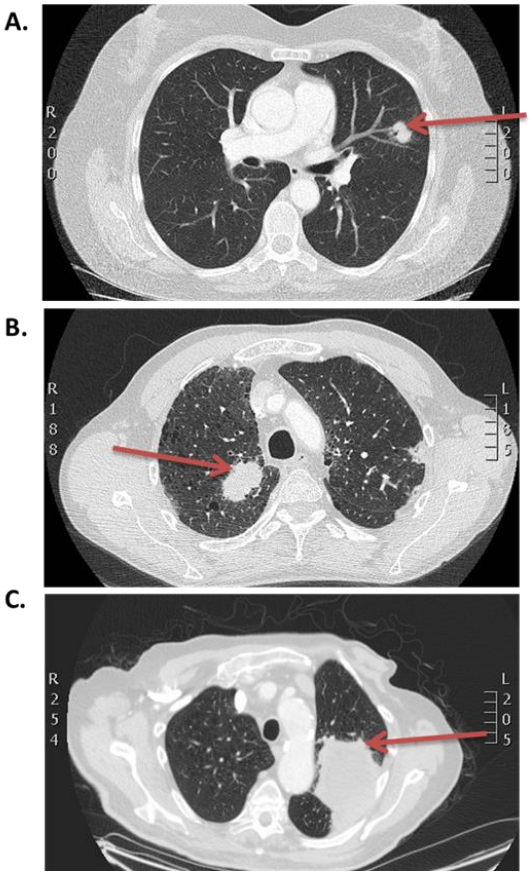


FIGURE 4.7: **CT Images of patients tumour.** CT mass were interrogated against DNA Integrity Index, by means of CT scans to assess tumour size. (A) Stage IA Non-mucinous Adenocarcinoma of the Lung; (B) Stage IIB Adenocarcinoma of the lung; (C) Stage IV NSCLC. Red arrow denoted tumour.

Alu was also unrelated to age, smoking, or diabetes mellitus. The only background variable associated with Alu was gender: women had significantly higher Alu (0.40) than men (*p-value<0.001*). Subsequently, we found that the main group effect on Alu remained significant also after statistically controlling for gender (*p-value<0.001)*.

4.4.3 Plasma interrogation of ovarian patient cancer samples using Raman Spectroscopy

Plasma samples from 10 ovarian cancer patients, and 5 healthy female individuals were processed by Raman Spectroscopy by our collaborators at Exeter University. Samples were analysed by drop Raman mapping. Lung samples were not included due to time constraints. Plasma samples were defrosted from -80°C freezer and a 1.5μL drop was pipetted onto a clean 'Raman grade' Calcium Fluoride slide and allowed to dry for 30 mins. Twenty Raman spectra were measured from the middle of the 'coffee ring' where most proteins were deposited onto the slide. The 20 spectra were collected at equally spaced points around the circumference of the ring using 830nm laser and acquired using a Renishaw in via microspectrometer with a Leica x50 long working distance objective (Fig 4.8).

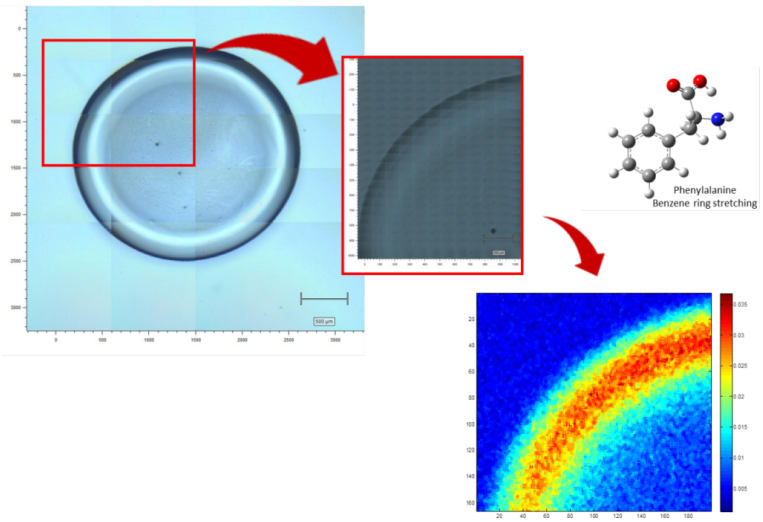


FIGURE 4.8: **Plasma drop on a 'Raman grade' calcium fluoride slide.** The image depicts the process of Raman spectroscopy, with an image of a plasma drop on a 'Raman grade' calcium fluoride slide, and image using the 830nm laser, x50 magnification.

The baseline was subtracted and the samples spectra were analysed for various parameters. Using this technique, various chemical properties of a sample can be measured, based on the Raman shift resulted from the laser interaction with the samples' molecules, eliciting excitation. These readings can identify specific chemicals and molecules, Fig 4.9.

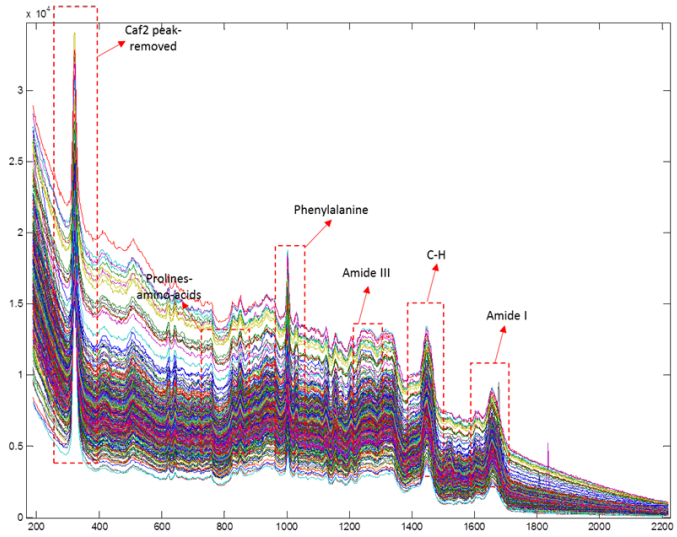


FIGURE 4.9: **Raman spectra of all ovarian cancer and control samples.** Image shows differences in chemical balance of the following molecules: CaF_2, Prolines-amino acids, phenylalanine, Amide III, C-H and Amide I.

The initial Raman spectra analysis shows multiple chemical property differences amongst all the samples. First we compared the Raman shift of ovarian cancer patients, to female controls, Fig 4.10.

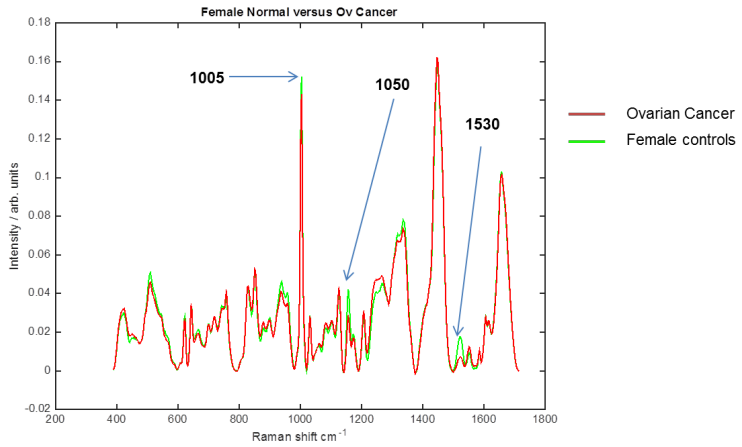


FIGURE 4.10: **Raman shift comparison of ovarian cancer and control samples.** A clear loss in carotentoid signals (1005/1150/1530 cm^{-1} peaks) in cancer patients. Other subtle peaks in the cancer group can also be seen.

There are clear distinctions in the loss of carotenoids in cancer patients compared to control samples, which are denoted by the identified peaks.

Principal component analysis (PCA) was also carried out to assess commonalities across the ovarian cancer samples. Loadings for the dataset contained 300 spectra, from 20 samples (cancer and control). PCA scores were calculated by ANOVA analysis for the principal components with highest statistical significance of separation between the groups.

Using the PCA scores for each of the 300 spectra a linear discriminant (LDA) function (Fig 4.11) was used to maximise the separation, and minimises the variance within groups.

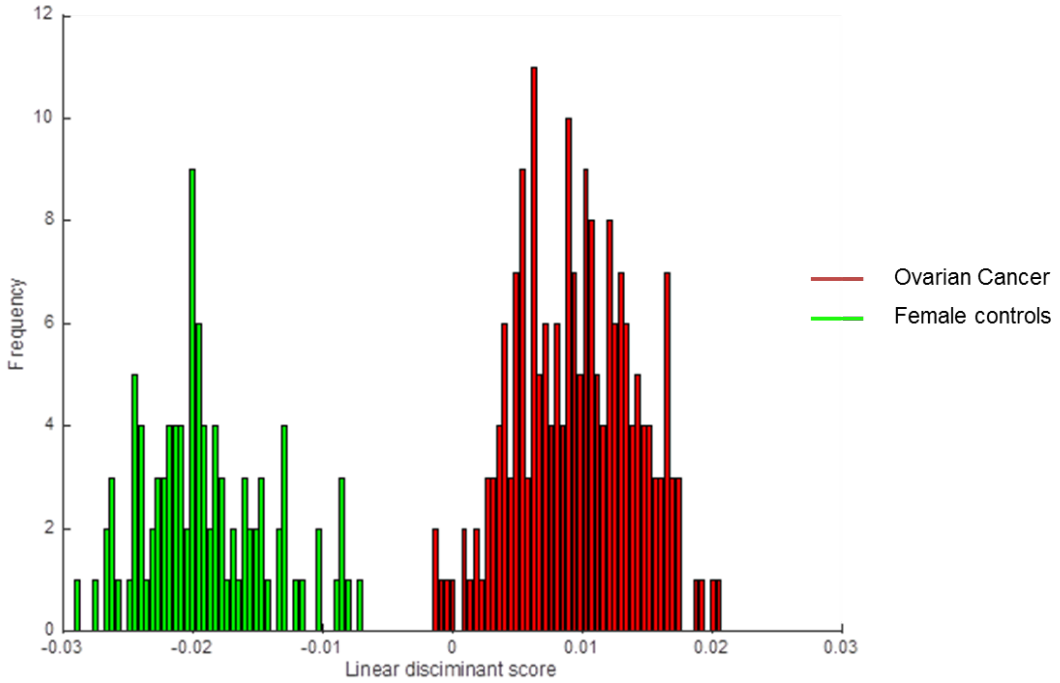


FIGURE 4.11: **Differentiation between cancer and control samples using linear discriminat analysis.** Histogram of a linear discrimination function of ovarian cancer and control samples from Raman spectroscopy analysis. A clear distinction in chemical properties is visible between the two groups.

The LDA plot reveals a 100% correct prediction; however, these results required further validation. This was done by extracting 20 spectra from each individual and keeping them back while a new PCA-LDA model was constructed from the remaining spectra to maximise separation. The held out spectra were then projected on the new model. The results are shown below Table 4.7.

	Prediction	
	Control	**Ovarian Cancer**
Control	100	0
Ovarian Cancer	40	160

TABLE 4.7: **Cross validation of 20 spectra from each individual sample.** Results generated show 80% sensitivity and 100% specificity.

Validation results support the findings of the PCA-LDA correlation model, and enforce strong sensitivity and specificity values within this technique.

4.5 Discussion

With ctDNA readily available in the plasma of cancer patients, various efforts have been made to exploit their clinical utility, one being the measurement of the DNA Integrity Index (Wang et al., 2003; Umetani et al., 2006a). We evaluated the efficacy of the DNA Integrity Index, by means of Alu repeat analysis of ctDNA found in the plasma of ovarian and lung cancer patients.

Based on the premise that tumour cells undergo a more chaotic cell death, compared to mostly apoptosis in normal cells (Umetani et al., 2006a; Lehner et al., 2013), we expect to find a higher ratio of smaller fragmented DNA to larger in the cancer cohort. As the disease progresses one would expect an increased DNA Integrity Index, reflecting the increased tumour burden and thus shedding of fragmented DNA. This can potentially provide means to not only identify and diagnose cancer patients, but also differentiate them based on staging and advancement of disease. Umetami *et al.* (2006) demonstrated a significantly increased DNA Integrity Index, even in patients with localised disease for colorectal cancer, demonstrating its utility as a diagnostic tool (Umetani et al., 2006a). More recently, a study in 95 breast cancer patients, revealed a significantly higher DNA Integrity Index, compared to benign and control samples (*p-value<0.001*). Moreover, sensitivity and specificity values of 85 and 100% were reported, concluding the clinical utility of ctDNA, and correlation with TNM staging (Kamel et al., 2016).

The results generated from this study show that a difference can be observed in the DNA Integrity Index of cancer patients compared to controls, in both ovarian and lung cancers. DNA Integrity Index in both the ovarian (*p-value=0.0194*) and lung cancer (*p-value=0.035*) cohorts were significantly higher than in their respective controls with average ratios of 0.37 (range 0.009-0.97) vs 0.20 (0.007-0.42); 0.38 (range 0.01-0.93) and 0.22 (0.007-0.44), in cancer and control respectively. ROC curve analysis reported an AUC of 0.69 (*p-value=0.07*) and respectively of 0.61 (*p-value=0.22*) in ovarian and lung cancer, indicating poor sensitivity and

specificity. This is in contrast to results from a study by Wang *et al.* (2003), in gynaecological and breast cancers where an AUC of 0.911 was observed (Wang et al., 2003). However, we do have to account for differences in the analysis method and primers used, including larger patient cohort, consisting of early and late stages.

In addition we noticed a considerable overlap between the cancer and non-cancer cohorts in both cancer types, which has been previously reported (Umetani et al., 2006a; Umetani et al., 2006b). No normal established baseline exists, healthy non-cancer patients will be expected to show both shorter and larger DNA fragments due to biological cell death processes. Inflammation and auto immune diseases are contributing factors to cell death rates, explaining higher DNA Integrity Index values in non-cancerous controls, with the lower ratio seen in the cancer group being potentially attributed to effective DNA clearance, as well as minimal cell death (Holdenrieder et al., 2008b). Other factors such as trauma or stroke can also affect the Alu ratio, demonstrating poor specificity (Chiu et al., 2006; Zeerleder et al., 2003; Lam et al., 2003; Rainer et al., 2006; Pares and Whitecross, 1982); the overlap in the DNA Integrity Index between the cancer and non-cancer group further corroborates this.

Further analysis was carried out to compare ctDNA values and corresponding CTCs from ovarian cancer patient samples on the METRO-BIBF trial. Data shows a similar trend for both biomarkers in patients A & B (Fig 4.3), while this is not apparent in patient C. These differences could be attributed to inter-patient variability. It is known that patients A & B responded well to the therapy, which may be represented by the spike in both biomarkers around cycle 3, possibly indicating tumour necrosis. By contrast, patient C did not respond well to the therapy and subsequently died, hence this may be reflected with by the little change in Alu ratio compared to patients A & B. Patient C also shows less of a decline in CTC value following a spike, as is seen with the other 2 patients, perhaps indicating little or no change to the tumour. As seen previously in chapter 3, CA125 levels did not correlate with CTC levels, the same is also true with ctDNA, these findings corroborating again the lack of value of the CA125 biomarker.

For lung cancer patients, further break down by staging revealed a significantly higher DNA Integrity Index in the advanced patients (III-IV and metastatic), compared to both the early stage (II-II) (*p-value<0.0001*) and controls (*p-value=0.0006*). Furthermore, ROC curve analysis between advanced cases and early stage, and advanced cases and normal indicated high sensitivity and specificity with AUC of 0.92, *p-value=0.0002* and 0.88, *p-value=0.0077* respectively, whereas early stage vs normal has a poorer AUC of 0.67, *p-value=0.23*. The results of this study are generally in line with recent reports (Umetani et al., 2006a; Umetani et al., 2006b), and hint at a prognostic role for this biomarker in advanced cancers. Literature also suggests that DNA released from malignant tumours into the bloodstream is

enhanced by lymphovascular invasion, as direct lymphatic or blood flow through the tumours enables dissemination of viable tumour cells, and thus can enhance the diffusion of DNA released from dead tumour cells into the bloodstream. As a result, circulating DNA may be directly related to tumour progression and rate of tumour cell turnover, representing biologic tumour aggressiveness (Umetani et al., 2006b), which would agree with the results of this study. Analysis of DNA Integrity Index in lung patients against staging showed Alu repeat ratios of >0.5 in the advanced stages, 0.61 and 0.58 in the advanced cases (III-IV) and metastatic patients, respectively, when compared to controls and early stage (I-II), 0.23 and 0.27 respectively, supporting the notion that with disease progression more ctDNA is shed into the circulation.

Short term follow up data was also acquired, and overall survival was plotted against DNA Integrity Index, whereby patients were divided by a ratio of either <0.5 or >0.5. Based on the results of average DNA Integrity Index in both ovarian and lung cancer, and given that the DNA Integrity Index falls between 0 and 1, the middle point seemed sensible to distinguish higher ratios from lower. A similar approach and cut-off was utilised in a recent study in breast cancers (Kamel et al., 2016). Overall survival was seen to be unaffected in the ovarian cohort (*p-value=0.7*), but was significantly better in the <0.5 DNA Integrity Index cohort (*p-value=0.03*) for lung cancer patients. This result suggests a potential role for lung cancer prognosis, however the median survival in days shows (Tables 4.3 & 4.6), that survival is slightly worse or not significantly different in the lower DNA Integrity Index group compared with the higher DNA Integrity Index group for lung cancer. Interestingly, no differences are observed in overall survival for ovarian cancers, while median survival in days is shown to be increased by approximately 3 months (85 days), suggesting again a potential prognostic use for this assay. These disparities allude to the complications of inter-patient variability, different cancer types, and one must consider the small number of samples analysed for both cancer types, particularly in the case of advanced and metastatic lung cancer patients.

Studies by Wang et al., 2003 corroborate our findings, as they report a significant increase in the DNA Integrity Index in ovarian and gynaecological cancers, concluding its clinical utility. Our initial findings showed similar results in ovarian cancer patients with an increased DNA Integrity Index compared to controls, however statistical significance was not reached and these did not translate into clinical outcomes which appears not to have been measured by Wang et al., 2003. Similarly studies in colorectal, periampullary, breast, ovarian, head and neck and prostate cancers (Yu et al., 2011; Lehner et al., 2013; Umetani et al., 2006a; Umetani et al., 2006b; Jiang et al., 2006; Hanley et al., 2006) also report positive results for the use of Alu repeats both diagnostically and prognostically. In contrast studies in pancreatic, gastrointestinal, colorectal cancer (Utomo et al., 2016; Boddy et al., 2006) have concluded against their utility as a clinical tool due to poor sensitivity and specificity. The

results from this finding suggest the assay lacks value in a diagnostic setting, however can be used in the advanced setting for prognostic purposes at least for lung cancer.

Raman spectroscopy analysis of the ovarian cancer samples, revealed interesting variations amongst the cancer and control samples. Metabolomic differences were also observed in a study by Winnard et al., 2017, when comparing primary breast cancer cell lines to normal, using Raman spectroscopy. Another recent study in melanoma patients, revealed differences in the biochemical and structural make-up of tumour tissue compared to controls, allowing the group to develop a model and identify specific components to be affected in this cancer cohort (Mazurenka et al., 2017). These studies are in line with our findings.

In this analysis we observed a significant loss in carotenoids in cancer patients compared to controls. Carotenoids are a structurally and functionally diverse group of natural pigments of the polyene type, known to be very efficient physical and chemical quenchers of both singlet oxygen ($^1O_2$), and potent scavengers of other reactive oxygen species (ROS) (Fiedor et al., 2005). The uncontrolled generation, and increase of ROS levels in the body, results in 'oxidative stress', a contributor to the pathogenic processes of many diseases, including cancers (Fiedor and Burda, 2014). It has been reported that various natural carotenoids have been proven to possess anti-carcinogenic properties (Nishino et al., 2000). Furthermore, carotenoids such as β-cryptoxanthin are suggested to stimulate the expression of an anti-oncogene, and p73 gene that is known as one of the p53-related genes (Nishino et al., 2000). Thus, this observed loss in carotenoids could explain tumorigenesis in these patients, in addition to acting as a blood based biomarker.

This technique has the potential to detect molecular changes prior to any morphological changes occurring in the tissue, offering thus many possibilities to aid the early detection of cancers. Raman spectroscopy is non-invasive with high specificity, and could be used for differentiating between benign and malignant tissue, in place of conventional biopsies (Mazurenka et al., 2017).

ctDNA and Raman spectroscopy show potential as 'liquid biopsy' techniques. Using a larger sample size when analysing ctDNA, may alleviate some limitations observed in this study. In this work we show the diagnostic and prognostic value of ctDNA in the advanced setting, by means of Alu repeats and DNA integrity analysis. Raman spectroscopy has also shown promising results, with particular emphasis on levels of carotenoids to differentiate cancer patients from controls. The results generated are shown to be in line with other findings, and further evaluation of specific identified components could be developed further.

Chapter 5

Identification of novel cancer biomarkers of prognostic value using specific gene regulatory networks: a novel role of RAD51AP1 for ovarian and lung cancers

5.1 Introduction

In addition to the use of CTCs as liquid biomarkers, microarray data has been utilised for biomarker and gene discovery. This chapter explores the use of microarrays in the discovery of robust liquid biomarkers and unique gene signatures using gene regulatory networks.

Microarray is a powerful technology that allows large scale interrogation of biological samples. The wealth of clinical cancer microarray data has been utilised over the years in the search for new biomarkers and disease specific 'molecular signatures' of genes. However, despite the successful use of microarray data in cancer biomarker discovery, there are still limitations in its main stream adoption for diagnostic or prognostic purposes. In addition, most studies concentrate on tissues which pose a greater level of difficulty in a clinical setting as they are are usually not readily accessible by comparison to liquid biopsies.

To make matters more complicated, a recent study revealed extensive genetic similarities across 14 major cancer types (breast, lung, and ovarian amongst them) based on whole exome somatic mutation profiles (Heim et al., 2014). This has been further corroborated by research demonstrating similarities between carcinomas from different organs, such as high-grade serous ovarian carcinomas, basal like breast cancer, and uterine serous carcinomas (Cancer Genome Atlas Network, 2012; Cancer Genome Atlas Research Network, 2011; Klein, 2013).

Thus, generic analyses may reveal distinct and common genetic features across cancer types, and these features could have the potential to be biomarkers of diagnostic or prognostic value. For this reason, we worked in collaboration with the computer science division at Brunel University London, to use a Unique Network Identification Pipeline (UNIP) for biomarker discovery focusing on four different cancer datasets: ovarian, lung, triple negative breast, and medullary breast cancer. UNIP aims to semi-automatically identify an enriched set of genes and the relationships between them, specific to one or a number of independent studies (Bo et al., 2014).

In this work, a list of four potential genes, known to be involved uniquely in each type of cancer has been identified, through mining of the National Center for Biotechnology Information (NCBI) data (Rebhan et al., 1997): RAD51 associated protein 1 (RAD51AP1), Follistatin like 1 (FSTL1), Collagen type XII, alpha 1 (COL12A1), and Small proline rich protein 1A (SPRR1A). This study evaluates their validity and clinical relevance.

5.2 Aims and objectives

- Clinically validate all four identified biomarkers in tissue samples, with emphasis in ovarian and lung cancer.
- Evaluate the function and relevance of successfully validated genes by means of functional studies.

5.3 Declaration of contribution to the work presented in this chapter

Oxford Brookes University- Microarrray

- **I have obtained RNA from SKOV3 cells transfected with siRNA for RAD51AP1 and controls.**
- **Oxford Brookes University ran the Microarray experiments and raw data was Then provided to me in a form of an excel file.**
- **I carried out bioinformatic analysis using a wide repertoire of *in silico* tools.**
- **I then cross checked the data with the patient details for any correlations.**

Table 5.1 gives a detailed description of contributions from each participating unit.

Contributions	**Oxford Brookes University**	**DC**
Patient and sample collection	-	x
Sample preparation (RNA/DNA/Plasma)	-	x
Sample processing	x	-
Analysis of raw/partly processed data	-	x
Further analysis and correlation with clinical samples	-	x

TABLE 5.1: **Contribution to the work presented in this chapter.** Details on specific tasks performed by Oxford Brookes University and Dimple Chudasama (DC).

5.4 Results

5.4.1 Unique-networks for prediction of unique genes

Development of the UNIP was done independently by a computing mathematics team, as part of a collaborative project. The UNIP was applied to analyse four different cancer datasets:ovarian, lung, triple-negative and medullary breast cancers. Initially, enrichment of the identified genes was undertaken by using principal component analysis (PCA) with a threshold on the standard deviation. Glasso was then applied to identify a Gene Regulatory Network (GRN) for each cancer sub-type, and derive unique connections common to the networks. Baysian Networks (BN) was used to obtain the unique-network structure (Fig 5.1) for each cancer subtype and identify, through inference, the most predictive (how well it predicts other expression level values) and predictable (how well its expression level values are predicted) genes within (intra) and outside (inter) the cancer subtypes using the leave one out cross validation technique.

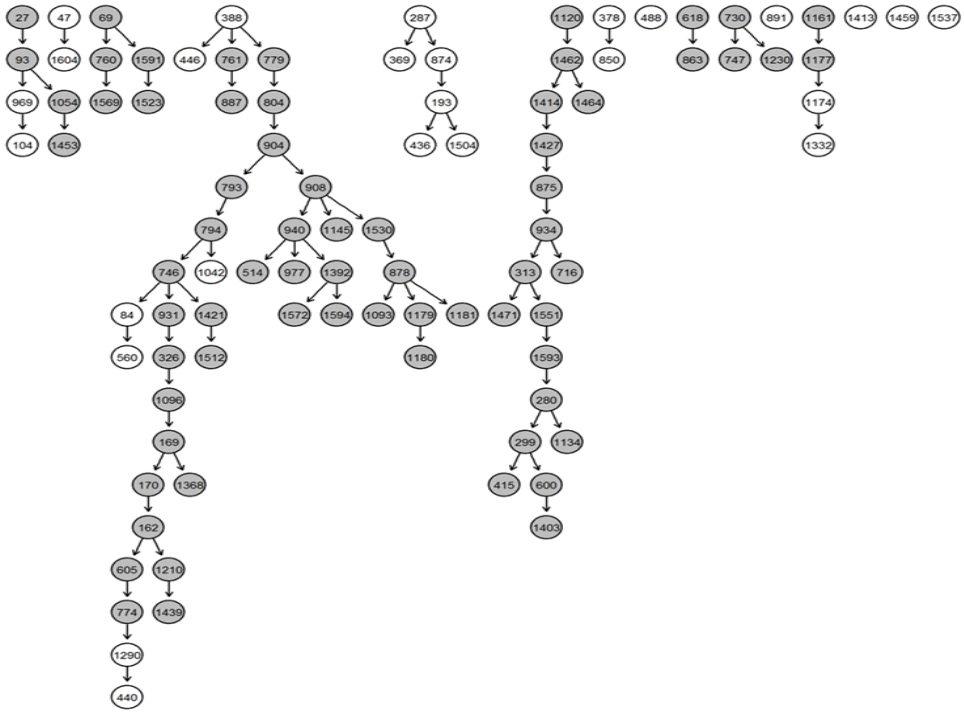


FIGURE 5.1: **Unique gene regulatory network.** A GRN was constructed using the methods described earlier. The nodes with a grey background indicate genes with intra-study predicted accuracy higher than 0.6. Numerical values within the nodes refer to the gene specific ID in the dataset. This pipeline identified 4 novel markers: SPRR1A, FSTL1, COL12A1, and RAD51AP1.

To capture study-specific genes, GeneCards encyclopaedia was used to obtain the list

of genes that are known to be involved in each cancer. These lists were compared amongst themselves and with the genes included in the unique-network list and, finally, we selected the genes that appeared in the ovarian cancer study.

Identification of these four genes prompted *in silico* analysis of existing datasets to further investigate the relevance of all genes identified in relation to prognosis and survival (see Fig 5.2, 5.3, 5.4, and 5.5). This was done using Oncomine and Kaplan Meier plotter, online tools that mine existing clinical data for these purposes. Oncomine normal data is denoted as a 1 and diseased (cancer) as 2.

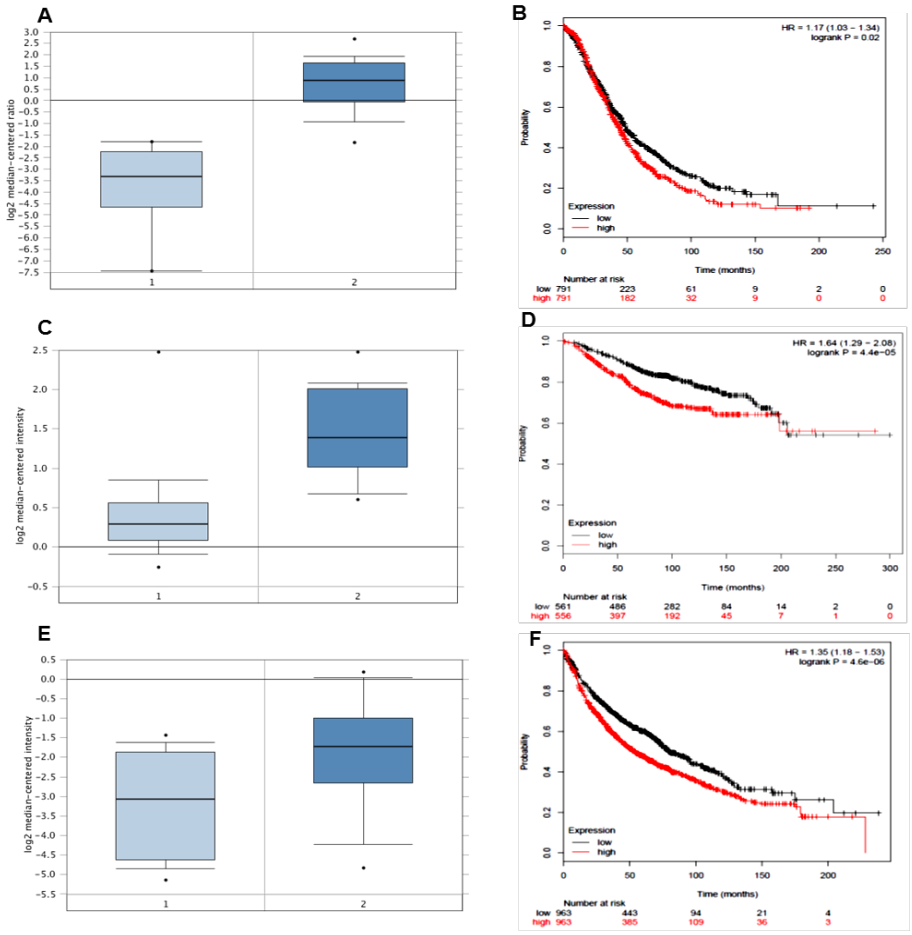


FIGURE 5.2: **Oncomine analysis for the RAD51AP1 gene.** Oncomine results for (A) Yoshimoto ovarian dataset (1, n=9; 2, ovarian serous adenocarcinoma, n=43), (C) Curtis breast dataset (1, n=144; 2, breast carcinoma, n=14); (E) Shearman lung dataset (1, n=19; 2, lung adenocarcinoma, n=20). RAD51AP1 is shown to be over-expressed across all the cancer types and associated with a poorer overall survival by Kaplan Meier analysis in all 3, ovarian (B), breast (D) and lung cancer (F), all shown to be statistically significant.

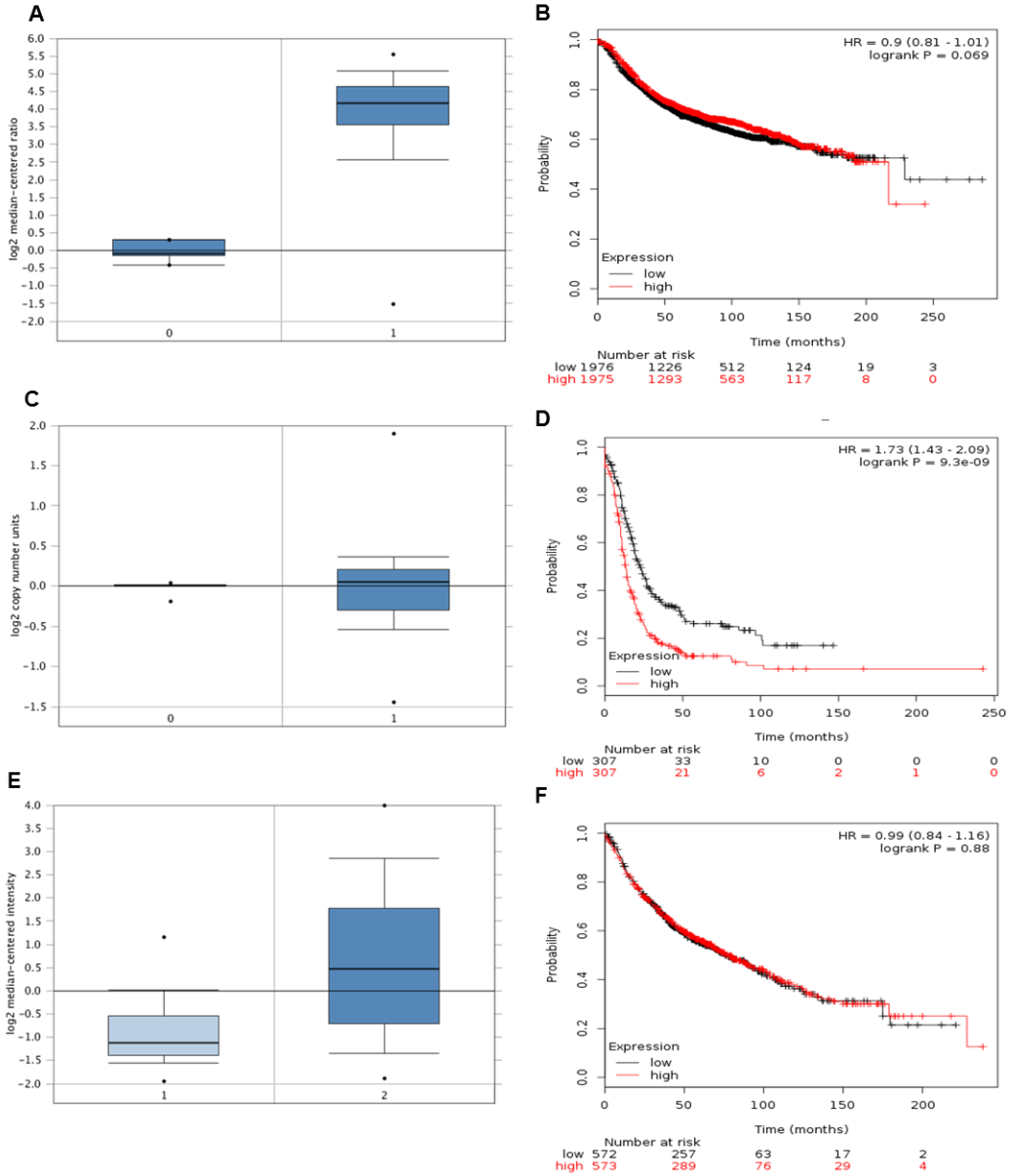


FIGURE 5.3: **Oncomine analysis for the FSTL1 gene.** Oncomine analysis denoted as 0 for controls and 1 for cancer, across (A) Richardson breast dataset (0, n=7; 1, ductal adenocarcinoma n=79), (C) Adib ovarian dataset (0, =4; 1, serous ovarian adenocarcinoma n=12), (E) Landi lung dataset (0=49; 1, adenocarcinoma n=79). Kaplan Meier analysis of FSTL1 is shown to be overexpressed across all 3 cancer types, although it appears to be weakly associated with a poorer overall survival for (B) breast and (D) ovarian cancers, this is not shown to be statistically significant. (F) Lung cancer however shows a poor overall survival with increased FSTL1 expression, which is seen to be statistically significant.

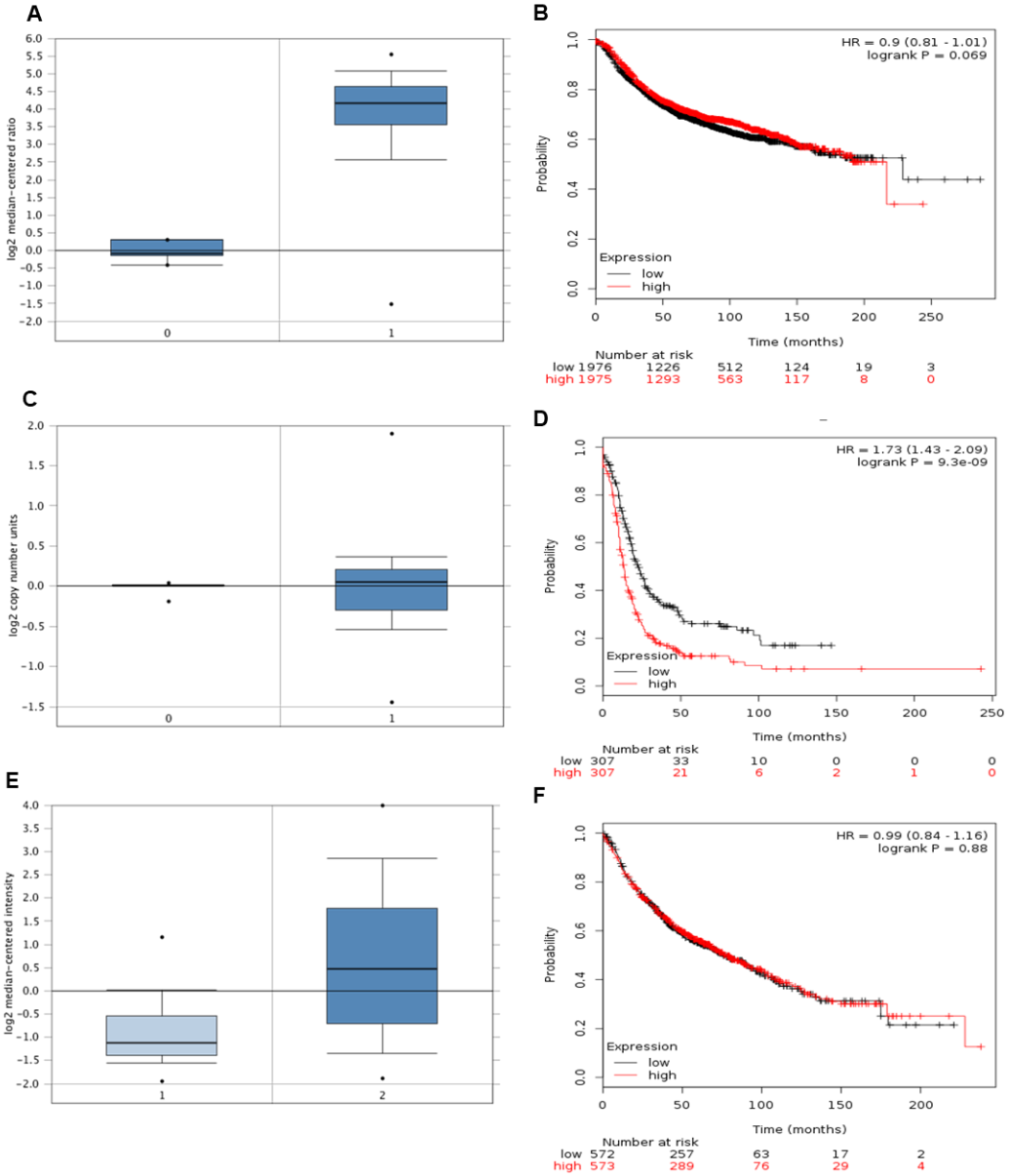


FIGURE 5.4: **Oncomine analysis for the COL12A1 gene.** Oncomine results for (A) Finask breast dataset (0, n-6; 1, invasive breast carcinoma n=53), (C) ovarian TCGA dataset (0, n=561; 1, ovarian serous cystoadenocarcinoma n=607), (E) Hou lung dataset (1, control n=65; 2, lung adenocarcinoma, n=45). COL12A1 is shown to be over-expressed across all the cancer types, ovarian being the exception where there is considerable overlap with controls. Kaplan Meier analysis of overall survival is shown to be poorer in (B) breast cancer patients who overexpress COL12A1, shown to be statistically significant. Overall survival is not shown to be affected in (D) ovarian and (F) lung cancers.

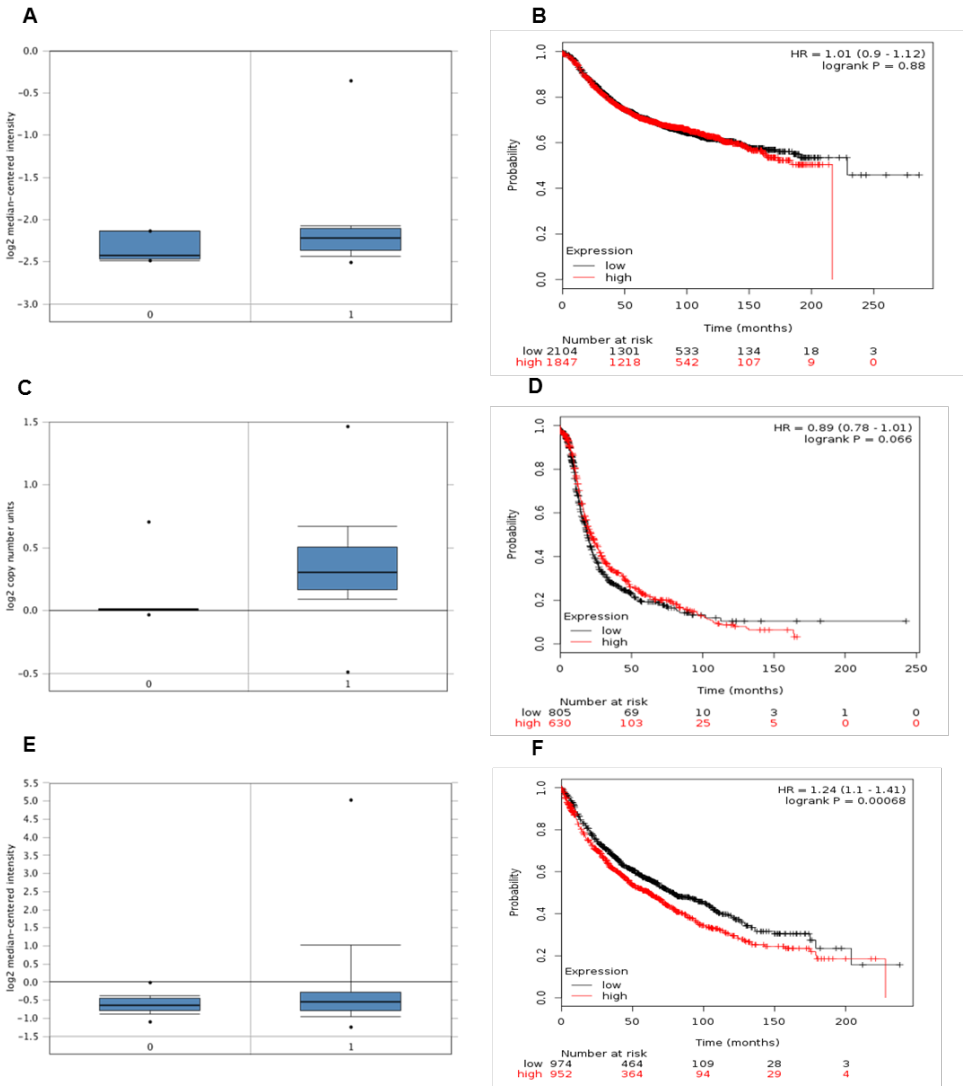


FIGURE 5.5: **Oncomine analysis for the SPR1A gene.** Oncomine results for (A) Richardson breast dataset (n=7; 1, ductal breast carcinoma, n=40), (C) TGCA ovarian dataset (0, n=561; 1, ovarian serous cystoadenocarcinoma, n=607), (E) Landi lung dataset (0, n=49; 1, adenoxarcinoma, n=58). SPRR1A is shown to have differential expression across all the cancer types. Similar levels appear to be shown in breast and lung cancer, with higher levels of expression seen in ovarian cancer compared to controls. Overall survival by Kaplan Meier analysis does not appear to be affected in breast, ovarian or lung cancer, (B), (D) & (E) respectively.

***In silico* analysis of the four identified genes portraits a very similar pattern of over-expression in breast, ovarian, and lung cancers, for the RAD51AP1, FSTL1, and COL12A1 (with the exception of ovarian) genes, according to the available microarray datasets, thus corroborating the findings of the GRN modelling. The SPRR1A gene however, exhibited a lower expression profile compared to the 3 other identified genes according to the GRN model. Oncomine analysis also showed a similar down-regulation compared to controls in breast and lung cancers, however higher expression in ovarian cancers. Kaplan Meier analysis of overall survival is mixed amongst the four genes and cancer subtypes. A strong association with over-expression and poorer overall survivals is seen in RAD51AP1, and to a lesser extent in FSTL1, both showing statistically significant results.**

5.4.2 Validation of identified genes in cell lines and clinical samples

We expanded on the *in silico* analysis results, by assessing the relative expression of the four genes common to the four cancer subtypes in relevant immortalised cell lines (see Table 5.2) by means of RT-qPCR, (as described in Section 2.5.8) using primers for RAD51AP1, FSTL1, COL12A1, SPRR1A, and reference genes Beta Actin and GAPDH. Biological and technical triplicates were performed for each experiment, as seen in Fig 5.6 to 5.9. Only normal (i.e. non-cancerous) cell lines were available for breast cancer cell lines, hence additional analysis on pooled cancer vs pooled control was carried out on these cell lines only, with breast cancer cell lines n=6, and control breast cell lines n=2.

Cell line	Site	Subtype
MDA-MB 231	Breast cancer	Invasive ductal carcinoma
MDA-MB 436	Breast cancer	Triple-negative breast cancer
MDA-MB 361	Breast cancer	Adenocarcinoma
Gi101	Breast cancer	Metastatic breast carcinoma
T-47D	Breast cancer	Ductal Carcinoma
SK-BR	Breast cancer	Adenocarcinoma
MCF-10A	Breast normal	Fibrocystic disease
HMEC	Breast normal	Primary human mammalian epithelial cells
PEO1	Ovarian cancer	Ovarian cell carcinoma
SKOV-3	Ovarian cancer	Adenocarcinoma
MDAH-2774	Ovarian cancer	Endometriod ovarian adenocarcinoma
A-431	Lung Cancer	Squamous cell carcinoma,
A549	Lung cancer	Adenocarcinoma

TABLE 5.2: **Cell lines tested against the panel of 4 GRN identified genes.** Details of site and disease subtype for each analysed cell line.

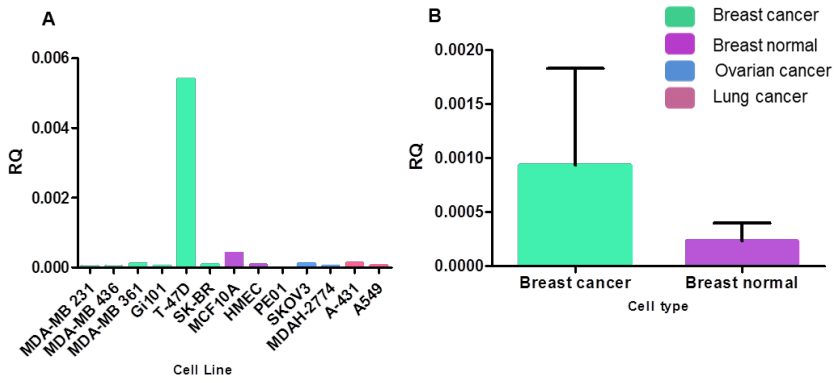


FIGURE 5.6: **Expression of SPRR1A across breast, ovarian and cancer cell lines.** An F-test was performed to assess the variance between the two groups, and a two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance. (A) Low expression was detected across all cell lines, with the exception of T47D breast cell line where a markedly increased expression in comparison to all other cell lines was noted. (B) Direct comparison of breast cancer and normal cell lines shows increased expression in the cancer cell lines, however this did not reach significance.

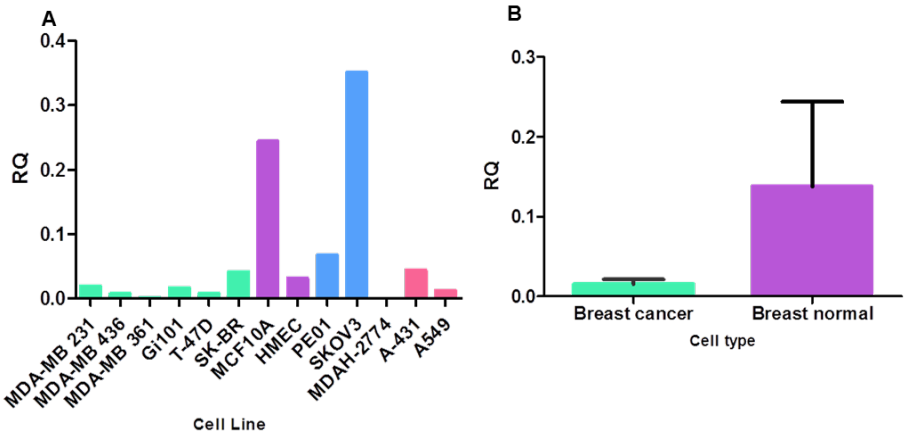


FIGURE 5.7: **Expression of FSTL1 across breast, ovarian and cancer cell lines.** An F-test was performed to assess the variance between the two groups, and a two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance. (A) A variable gene expression is seen across all cell lines. (B) Normal breast cell lines are shown to have higher FSTL1 expression compared to cancer; however this did not reach statistical significance.

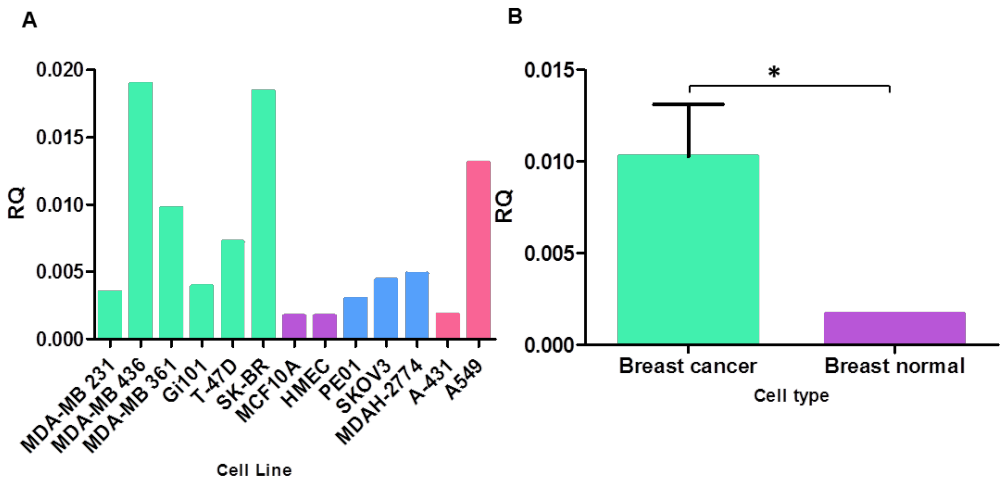


FIGURE 5.8: **Expression of RAD51AP1 across breast, ovarian and cancer cell lines.** An F-test was performed to assess the variance between the two groups, and a two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance. (A) a variable but generally increased gene expression is seen across all cell lines. (B) Breast cancer cell lines are shown to have higher RAD51AP1 expression compared to normal breast cell lines, **p-value=0.0285.*

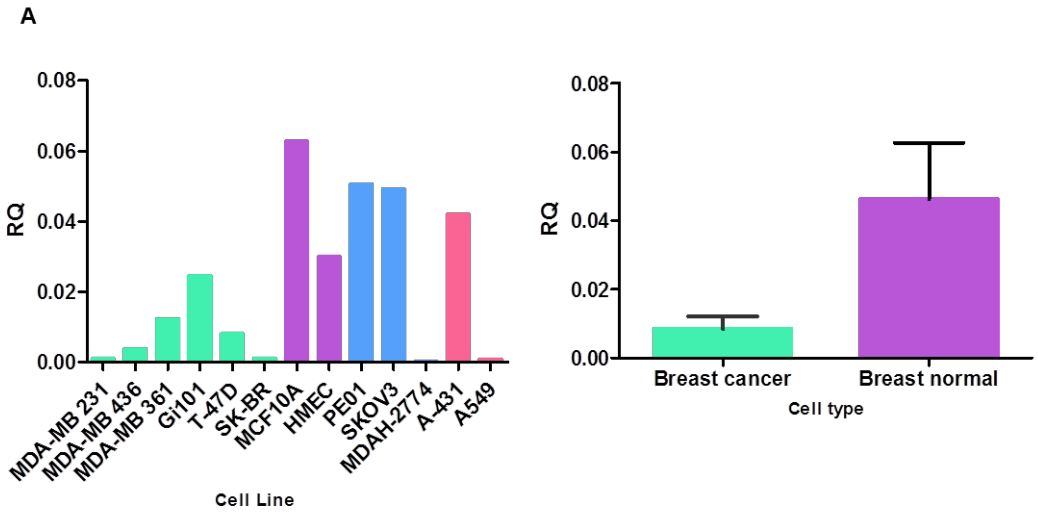


FIGURE 5.9: **Expression of COL12A1 across breast, ovarian and cancer cell lines.** An F-test was performed to assess the variance between the two groups, and two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance. (A) Variable but generally increased expression of COL12A1 are seen in ovarian and lung cancer cell lines, compared to low gene expression seen in breast cancer patients in comparison to controls, which is seen better in (B), although not shown to be statistically significant.

The cell line expression data generated for the four genes, showed a cell-type specific response. To further validate the expression of the 4 genes clinically, patient tissue samples were compared with normal controls (Fig 5.10). Tissue samples for ovarian cancer patients were obtained from the University of Thessaloniki (Table 2.2, detailed in Section 2.2.2). Lung tissue samples were obtained from Harefield Hospital, London, in the form of tissue or formalin fixed paraffin embedded (FFPE) blocks, breast cancer cDNA was purchased. A tissue array kit was purchased, which consisted of a 96 well plate preloaded with ovarian, lung, and breast cancer, and control cDNA (Origene). Information of all samples in this study are tabulated in Table 5.3. qPCR on cDNA was performed and reference genes Beta Actin (as recommended by Origene array plate manufacturers) and GAPDH were used.

Lung Tissue		**Ovarian Tissue**		**Breast Tissue**	
Pathology	**N**	**Pathology**	**N**	**Pathology**	**N**
Normal	16	Normal	12	Normal	1
Adenocarcinoma	19	Ovarian serous adenocarcinoma	25	Breast ductal adenocarcinoma	5
Squamous carcinoma	6	Clear cell adenocarinoimoma	2		
Carcinoid	1	Endometriod adenocarcinoma	1		

TABLE 5.3: **Clinical samples information.** Summary of the pathology and numbers pertaining to the clinical tissue samples used in this study.

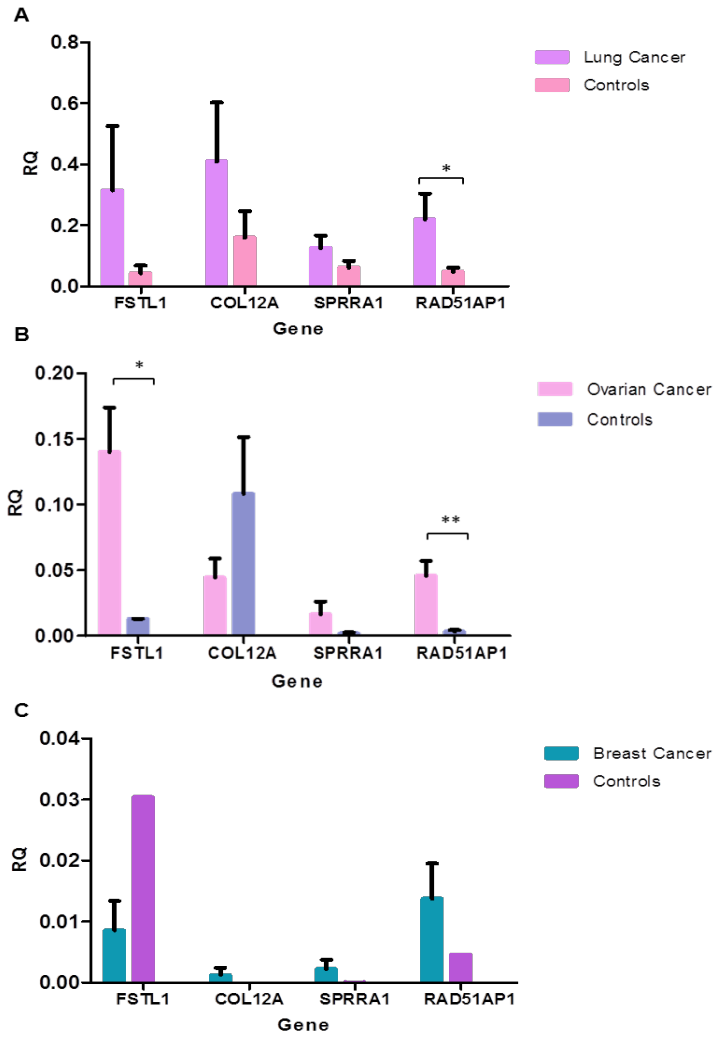


FIGURE 5.10: **Results of RT-qPCR of all four genes across breast, lung and ovarian cancer cell line.** RAD51AP1 gene expression in breast tissue samples was measured by RT-qPCR. An F-test was performed to assess the variance between the two groups, and a two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance. Results show RAD51AP1, COL12A1 and SPRR1A as significantly over-expressed in lung, when compared with ovarian and breast cancer, and FSTL1 was significantly over-expressed in ovarian cancer compared with lung and breast cancer. Within lung cancer RAD51AP1 was shown to be significantly over-expressed in cancer compared to control tissue, (**p-value=0.04*). Whereas in ovarian cancer FSTL1 and RAD51AP1 were shown to be significantly over-expressed in diseased tissue compared to healthy controls, **p-value=0.04* and ***p-value=0.0012*, respectively.

Short term outcomes and survival data was available for the lung cancer cohort only, as no outcomes data was available for the ovarian samples for ethical reasons. Overall survival for the RAD51AP1 gene was analysed by a Kaplan Meier plot (Fig 5.11). The mean RAD51AP1 expression in controls being used as a cut-off of high and low expression.

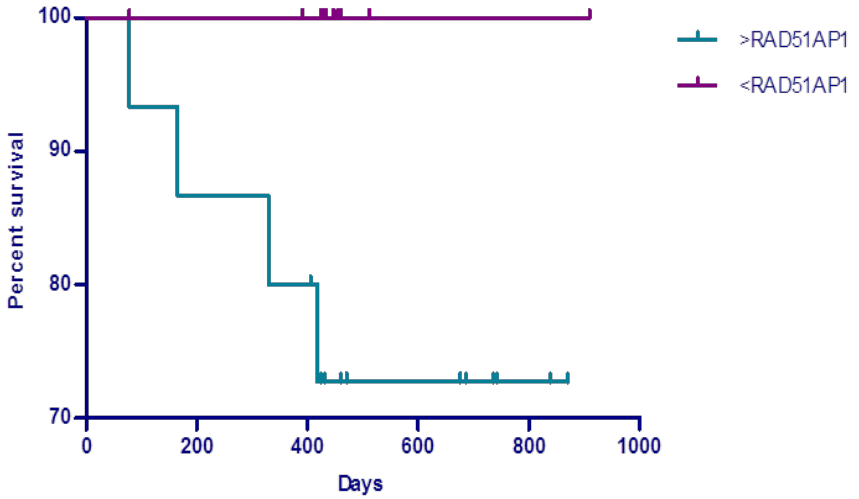


FIGURE 5.11: **Over-expression of RAD51AP1 in lung cancer patients.** Over-expression of RAD51AP1 is associated with poorer overall survival, *p-value=0.0230*, Chi-squared 5.166, Hazards Ratio 9.9 (95%, 1.37-71.6).

Data generated from clinical validation of the four genes showed RAD51AP1 to be the most promising of the four, with significant over-expression in lung and ovarian cancer tissue samples compared to their relevant controls. A similar pattern was observed in breast cancer samples as well but the results were not statistically significant. Furthermore, overall survival data of lung cancer patients has shown that a poorer prognosis association with higher RAD51AP1 expression levels was statistically significant. For these reasons, we have decided to concentrate on further investigation of RAD51AP1.

To validate RAD51AP1 in a clinical setting, we used immunohistochemistry techniques to measure RAD51AP1 expression at the protein level in tissue samples from both lung and ovarian cancer patients using a commercially-available tissue microarray (Biomax Ltd). Formalin fixed paraffin embedded (FFPE) tissue slides were acquired from Harefield Hospital (lung) and Mount Vernon Cancer Centre (ovarian). Sections were sliced and mounted on to slides, and stained using an antibody against RAD51AP1 (Sigma), and visualised by DAB-chromogen (Dako), as described in Section 2.3.5. Slides were then scored on positive staining and plotted, Fig 5.12.

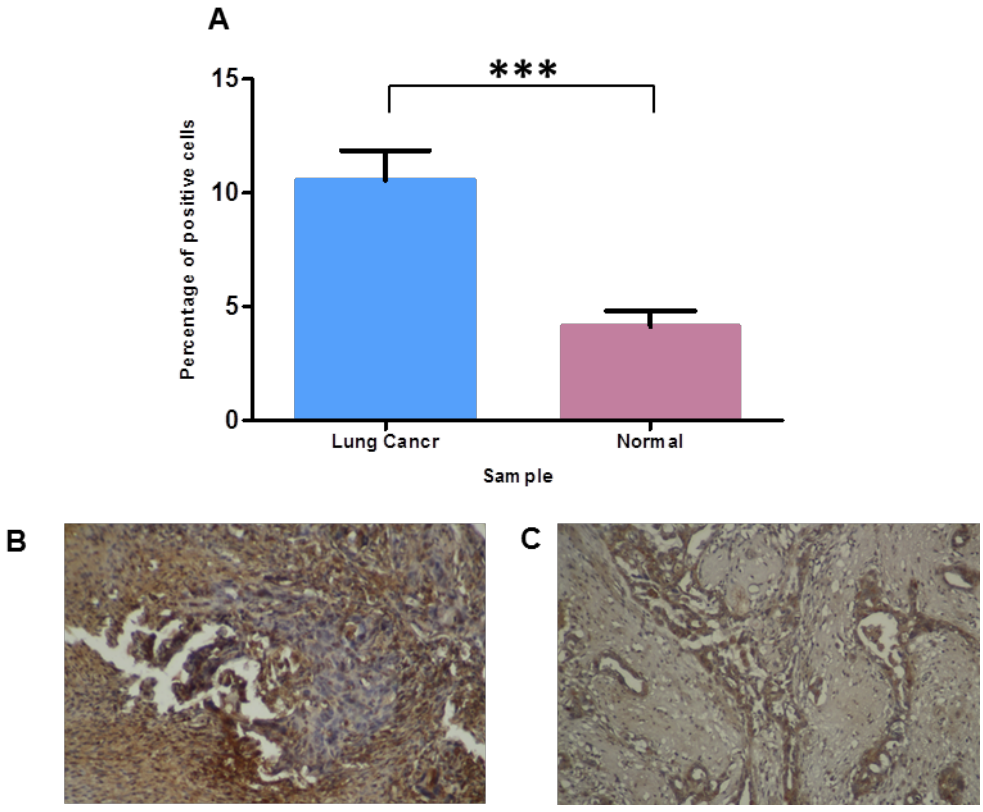


FIGURE 5.12: **RAD51AP1 protein expression was measured in lung cancer tissue samples embedded in FFPE blocks.** The scoring of lung FFPE slides are as follows, (A) shows a significantly higher proportion of RAD51AP1 expression at the protein level compared to normal controls,****p-value<0.0001*. (B) Image represents an FFPE slide from a lung cancer patient, stained with the RAD51AP1 antibody, strong positive staining can be visualised by the brown staining. (C) Control non-lung cancer lung tissue, with little to no RAD51AP1 staining on FFPE tissue. Images at $\times$20 magnification.

RAD51AP1 is shown to be over-expressed significantly in lung cancer patients (n=43) compared to controls (n=7). Analysis of the staining pattern of RAD51AP1 revealed both cytoplasmic and nuclear staining. Further analysis was carried out on staging and subtype, Fig 5.13.

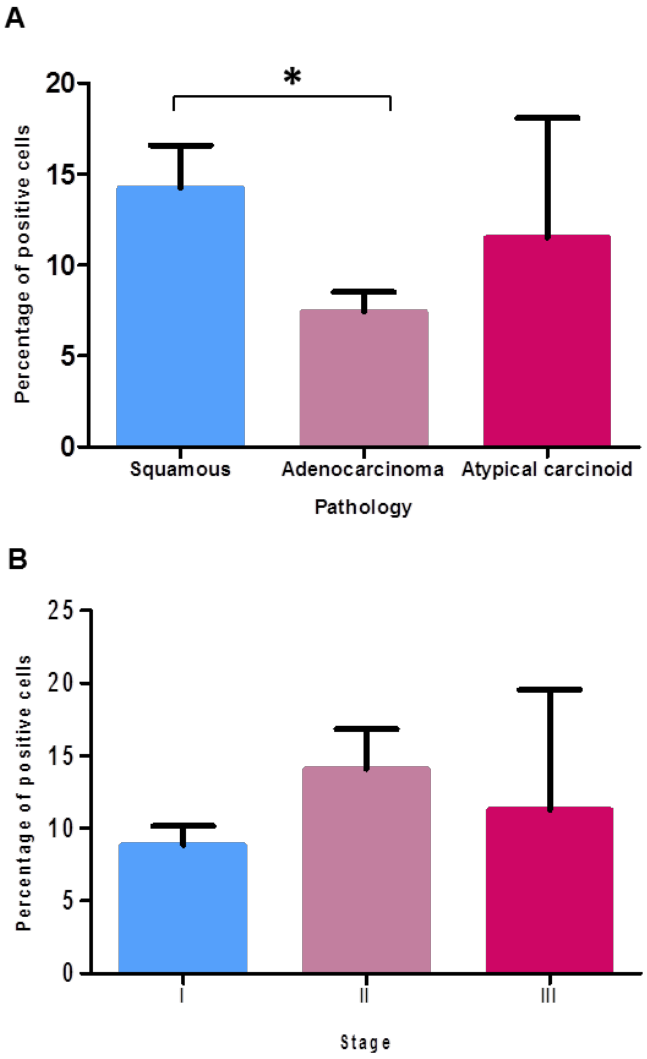


FIGURE 5.13: **RAD51AP1 protein expression from scoring of lung FFPE slides.** (A) shows a significantly higher proportion of RAD51AP1 expression in squamous cell carcinoma compared to adenocarcinoma, *p-value<0.0001*. RAD51AP1 expression appears increased compared to adenocarcinomas but not squamous carcinoma, however this is not statistically significant. (B) RAD51AP1 expression in the tissue appears highly expressed in stages II and III, however not in a stage specific manner. This data is not seen to be statistically significant.

Differences were observed in RAD51AP1 protein level expression in lung cancer patients amongst cancer subtype and staging. Immunohistochemistry and RAD51AP1 protein expression was also tested on FFPE tissues obtained from ovarian cancer patients. Details of scoring were plotted, Fig 5.14.

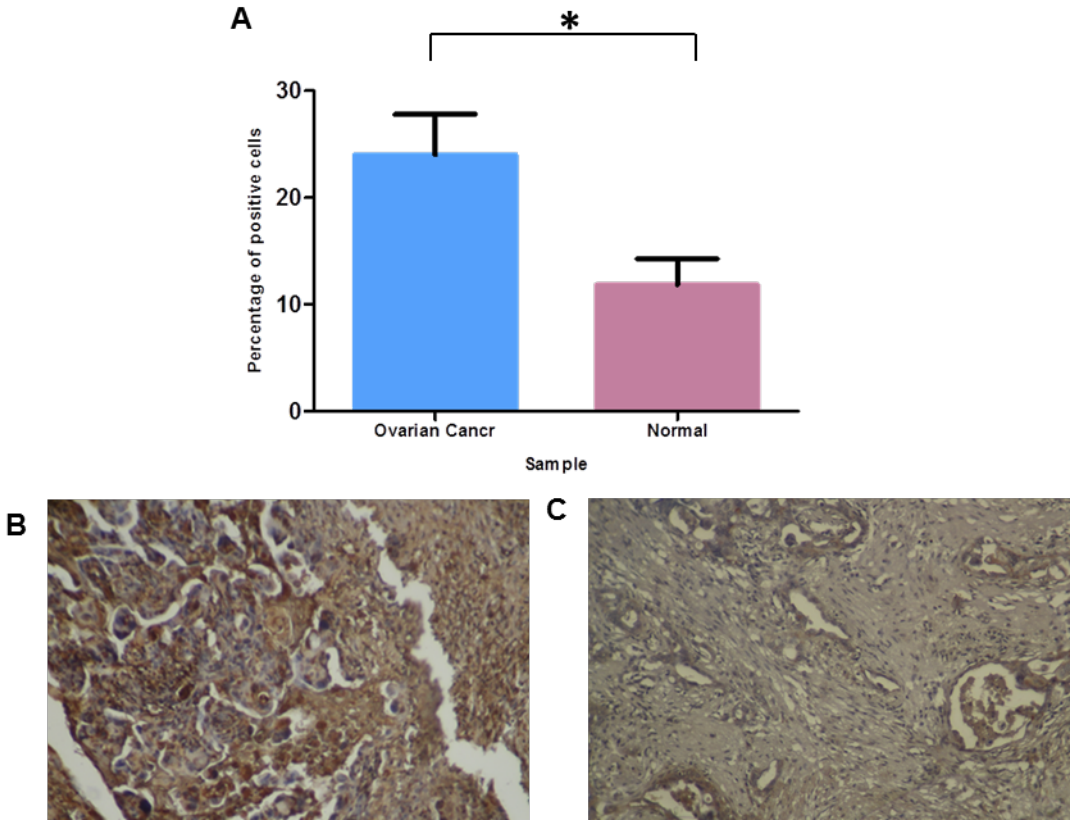


FIGURE 5.14: **RAD51AP1 protein expression from scoring of ovarian FFPE slides.** (A) shows a significantly higher proportion of RAD51AP1 expression at the protein level compared to normal controls, *p-value=0.0157*. (B) image represents an FFPE slide from an ovarian cancer patient, stained with the RAD51AP1 antibody, slide shows positive staining denoted by the brown. (C) Control ovarian FFPE tissue from non-cancer patient, shows weak RAD51AP1 staining.

RAD51AP1 is shown to be significantly over-expressed in ovarian cancer patients (n=40) compared to controls (n=5). Again a similar staining pattern of RAD51AP1 was observed, in cytoplasm and nucleus. Further analysis was carried out on staging and subtype, Fig 5.15.

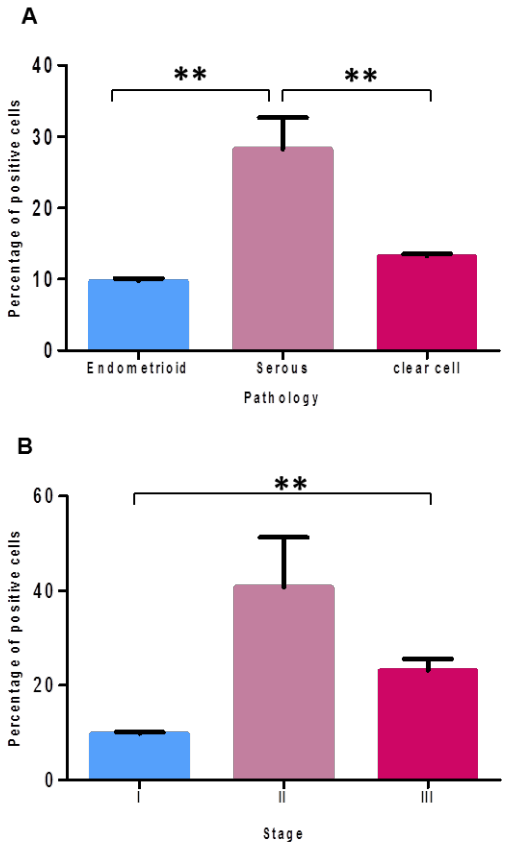


FIGURE 5.15: **RAD51AP1 protein expression from scoring of ovarian FFPE slides.** (A) shows a significantly higher proportion of RAD51AP1 expression in serous cell carcinoma compared to endometriod and clear cell carcinoma, ***p-value=0.0016* and *0.0063*, respectively. (B) RAD51AP1 expression in the tissue appears highly expressed in stages II and III, however not in a stage specific manner. Statistical significance is only seen when comparing stage I to III, where stage III shows increased expression, ***p-value=0.0016*.

Ovarian cancer subtype and staging also displays difference in RAD51AP1 expression at the protein level suggesting differential expression is dependent on cancer pathology.

Next we analysed blood samples for elevated expression levels in RAD51AP1. Blood samples were only available for ovarian (diseased n=30, controls n=12), and lung cancers (diseased n=18, controls n=16), due to ethical restrictions breast cancer blood samples were unattainable. All blood samples were collected and processed within 4 hours.

RNA was extracted using the Ribopure™ RNA blood extraction kit (Fisher) or Qiagen mini RNA™ extraction blood kit (Qiagen). cDNA was synthesised using the high capacity cDNA Reverse Transcriptase kit (Thermo Fisher, as described in Section 2.5.7). A 6 geNorm™ analysis was carried out to assess the reference gene stability, as it is known that common reference genes can lack expressional stability.

Multiplex qPCR on cDNA was performed as previously described on the Quantistudio 7® (ABI) thermal cycler using Multiplex Taqman™ PCR Master Mix (Thermo Fisher) (Section 2.5.15) using probes for RAD51AP1 (FAM-MGB), Beta Actin (VIC-MGB), and GAPDH (FAM-MGB). Relative quantification is shown in Fig 5.16.

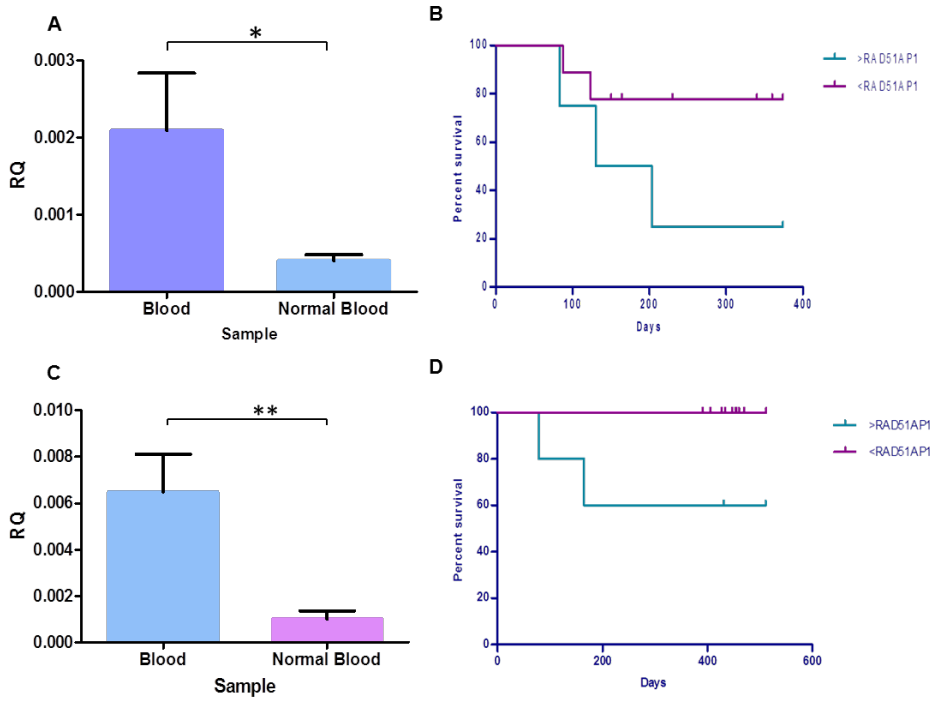


FIGURE 5.16: **RT-qPCR results of RAD51AP1 expression bloods of ovarian and lung cancer patients compared to healthy controls.** An F-test was performed to assess the variance between the two groups, and two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance. RAD51AP1 was shown to be over-expressed in the bloods of (A) ovarian and (C) lung cancer patients, in comparison to control blood samples, **p-value=0.0320* and ***p-value=0.0040*, respectively. Moreover, Kaplan Meier analysis demonstrates a poorer overall survival being associated with RAD51AP1 expression in both (B) ovarian and (D) lung cancer patients, **p-value=0.0478*, Chi-squared - 3.92, Hazards ratio - 8.72 (95% CI, 1.02 to 72.6); **p-value=0.0230*, Chi-squared - 6.26, Hazards ratio - 62.2 (95% CI, 2.44 to 1582), respectively.

Analysis of RNA extracted from whole blood showed significant up-regulation of the expression level of RAD51AP1 in diseased patients compared to healthy controls, furthermore this was associated with a poorer overall survival in both ovarian and lung cancer patients.

Data generated from *in silico* and qPCR of clinical samples, demonstrate that RAD51AP1 is an important gene which is implicated in both ovarian and lung cancers and has a great potential as a prognostic predictor. For this reason, RAD51AP1 was therefore chosen for further *in vitro* functional studies, using the SKOV-3, MDAH-2774, and A549 cells lines as experimental models for ovarian cancers, and the latter for lung cancer respectively.

5.4.3 Knockdown of the RAD51AP1 gene in cell lines

Suitable cell lines for the knockdown experiments were established by testing protein expression of RAD51AP1 by means of immunofluorescence and Imagstream analysis (Fig 5.17 and 5.18) of the two ovarian cancer cell lines, MDAH-2774 and SKOV-3, and lung cell line A549. Stained cells were viewed under the Leica DM4000 microscope for the IF experimentation. The Imagestream^{TM} was used to also view RAD51AP1 protein expression, in particular due to its enhanced microscopic and individual cell analysis capabilities.

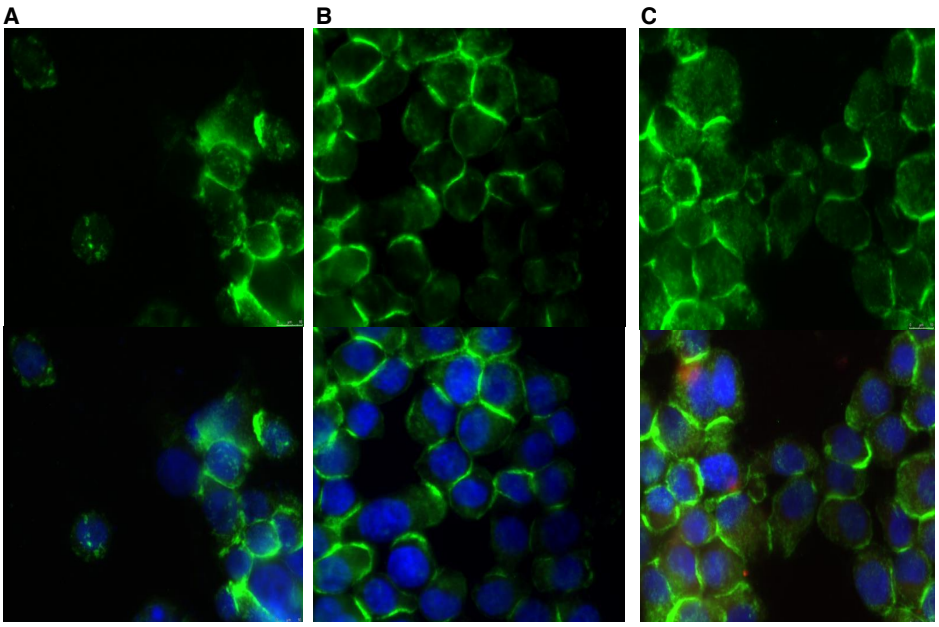


FIGURE 5.17: **Immunofluorescent staining of cell lines for RAD51AP1 expression.** Staining of (A) MDAH-2774, (B) SKOV-3, and (c) A549 cell lines of RAD51AP1 protein expression, at 40x magnification. Images are divided into two with the top - RAD51AP1 seen in green (Alexafluor488) and the bottom panel a combined image of the RAD51AP1 by the Alexafluor488 (green) and blue DAPI staining of the nucleus. A strong cytoplasmic green staining in addition to a moderate to strong green loci type staining is visible throughout most cells, suggesting aberrant expression of RAD51AP1 in all three cell lines.

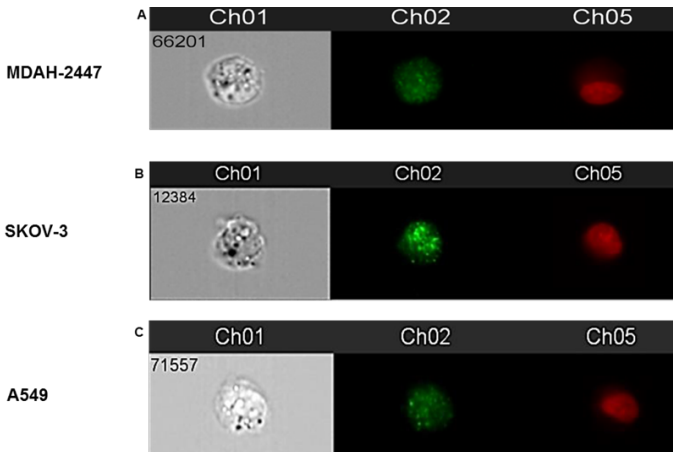


FIGURE 5.18: **RAD51AP1 protein level expression in MDAH-2774, SKOV-3, and A549 cell lines.** Imagestream analysis was used to measure protein expression of RAD51AP1 (green) and nuclear stainig with DARQ5 (red). Results demonstrate an aberrant protein expression in (A) MDAH-2774, (B) SKOV-3, and (C) A549 cell lines.

Negative controls for each cell line were also carried out, where primary antibody was omitted. Images can be seen in Fig 5.19.

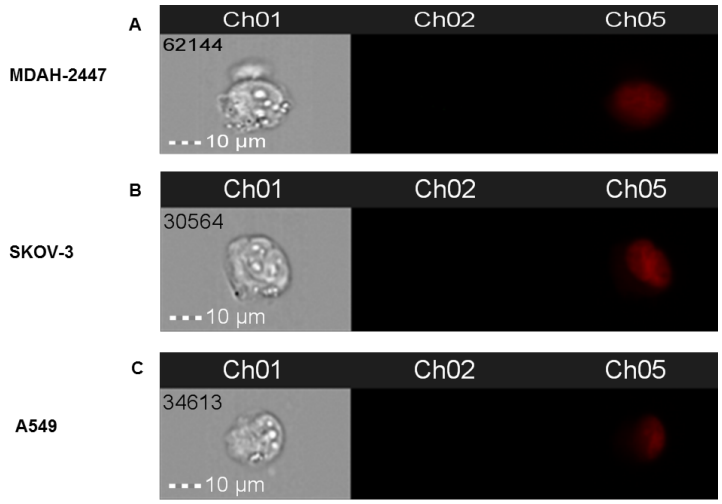


FIGURE 5.19: **Imagestream analysis of negative control experiment in MDAH-2774, SKOV-3, and A549 cell lines.** Results of all 3 cells lines in a negative control experiment, whereby RAD51AP1 primary antibody was omitted, are shown to have worked and no staining (green) is visible as would be expected, only the nucleus is stained by the DRAQ5 (red) . (A) MDAH-2774, (B) SKOV-3 & (C) A549 cell line.

Using ImageStream and immunofluorescence analyses we show that RAD51AP1 is aberrantly expressed as a protein in all 3 cells lines, while the lack of staining in the negative controls of all 3 cell lines renders them suitable models for functional studies.

Next we used siRNA targeted to RAD51AP1 (SMARTpool: ON-TARGETplus, Dharmacon, CO, USA) to silence RAD51AP1 expression in MDAH-2774, SKOV-3, and A549 cells. Transfection efficiency was first determined with a series of optimisation assays using siGLO-labelled siRNA, carried out as per manufacturer's instructions (Fig 5.20).

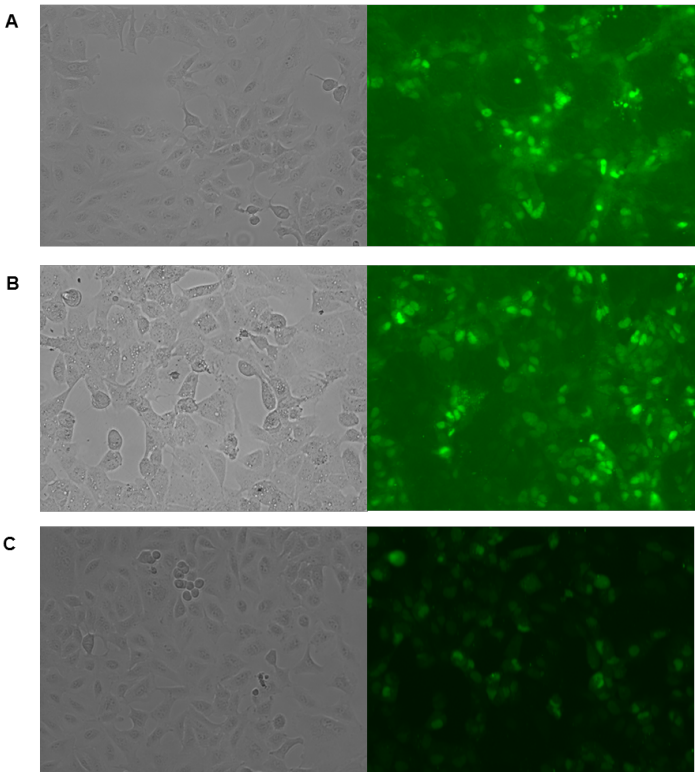


FIGURE 5.20: **Phase and fluorescent imaging of SIGLO-labelled cell lines at a siRNA concentration of 50nmol/L per well.** Phase imaging, showing the level of siRNA uptake in (A) MDAH-2774, (B) SKOV-3, and (C) A549. Imaging shows approximately >90% uptake efficiency in all three cell lines, seen by the green fluorescence uptake.

Transfection efficiency was determined to be greater than 90% with the siGLO-labelled siRNA, using a concentration of 50nmol/L per well/dish and Dharmafect 1 transfection reagent for all cell lines.

According to MIQE (Minimum Information for the Publication of Quantitative Real-Time qPCR) guidelines, reference gene expression should be stable and strongly correlated with the total mRNA expression. As cancer is a heterogeneous disease, it is advisable to assess reference gene stability in these samples. geNorm^{TM} (Primerdesign) is a commercially available assessment kit for suitable reference genes within a sampleset. The geNorm^{TM} human 6 gene kit, assesses six different human reference genes for average expression stability and the optimum number of reference genes required in the qPCR experiments for each set of samples.

Six samples were selected (2 of each cell line, one transfected and one scrambled control) to represent the whole cohort and were tested in triplicate using the geNorm^{TM} 6 gene kit (Primerdesign). An equal number of samples were used from each group to

eliminate bias towards any variable. The kit contains primers for ATB5P, B2M, UBC, 18S, YWHAZ and RPL13A (detailed in Section 2.5.9). 2X PrecisionPlus^{TM} qPCRMaster-Mix (Primerdesign) was used on the ABI Quantistudio 7 (Life Technologies, described in Section 2.5.15). cDNA was used at a concentration of 5ng/μL as per manufacturer's instruction. qPCR results were analysed using the qBase+ software (Biogazelle) (see Figs 5.21 & 5.22).

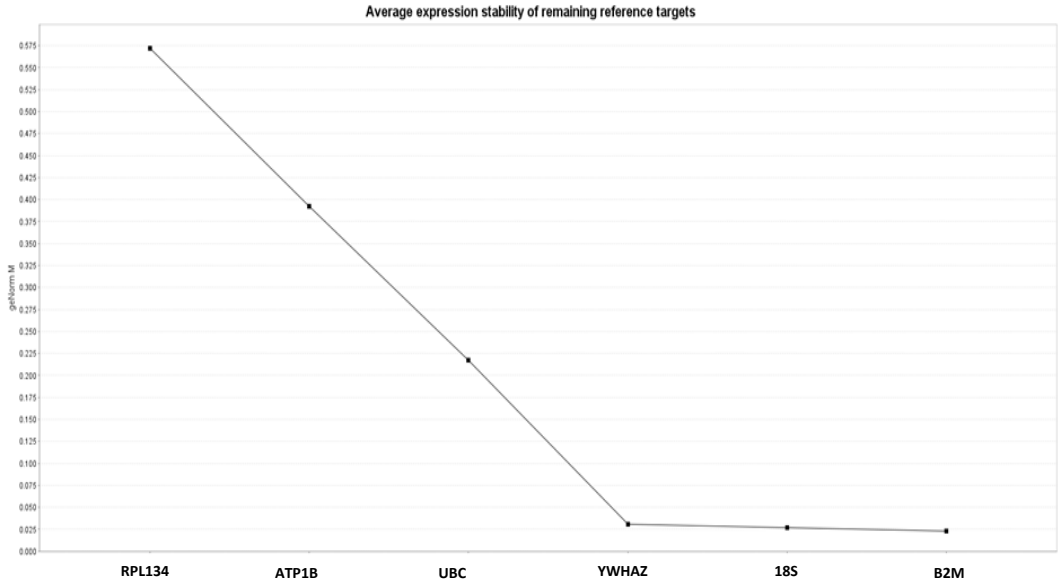


FIGURE 5.21: **Six gene analysis of the geNorm^{TM} 6 gene kit (Primerdesign).** House-keeping genes from the 6 gene geNorm kit were assessed for expressional stability in a selection of cDNAs from clinical samples, representative of the entire cohort. Data were analysed using qBase software (Biogazelle). qBase+ provides a geNorm^{TM} M value, representing the expressional stability; a lower M value is indicative of a greater expressional stability. The graph orders genes of least stability to most (left to right). The two most stable genes were identified as 18S and B2M.

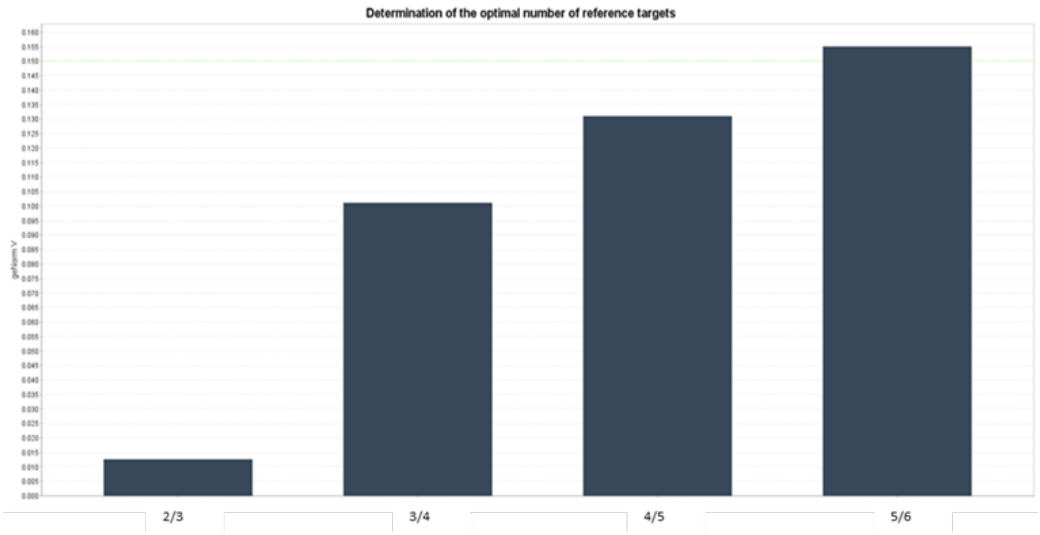


FIGURE 5.22: **Six gene analysis of the geNorm^{TM} 6 gene kit (Primerdesign)**. The 6 genes in the geNorm kit were assessed for expressional stability in a selection of cDNAs from clinical samples, representative of the entire cohort. Data were analysed using qBase software (Biogazelle). qBase+ provides a geNorm^{TM} V graph showing cumulative variability of successive genes, thus indicating how many reference genes should be used in each qPCR experiment for optimum normalisation, with a value below 0.15 represents the optimal number. In this case 2-3 genes have been shown to be sufficient.

geNorm analysis, indicates the requirement of at least 2 to 3 reference genes for these experimentations. As it was earlier indicated in Fig 5.21 that 18S and B2M were the most stable genes, it was decided to proceed with these 2 genes in these clinical samples to ensure robust normalisation.

A scrambled control was also included (ON-TARGETplus Non-targeting Pool; Thermo Scientific). Twenty-four hours after transfection, the medium in the wells was replaced with fresh serum-free complete media, to avoid toxicity. RNA and lysate cell samples were collected at baseline, 48, and 72 hours post transfection. Level and efficiency of RAD51AP1 silencing was established by Western blotting and qPCR analysis (Fig 5.23).

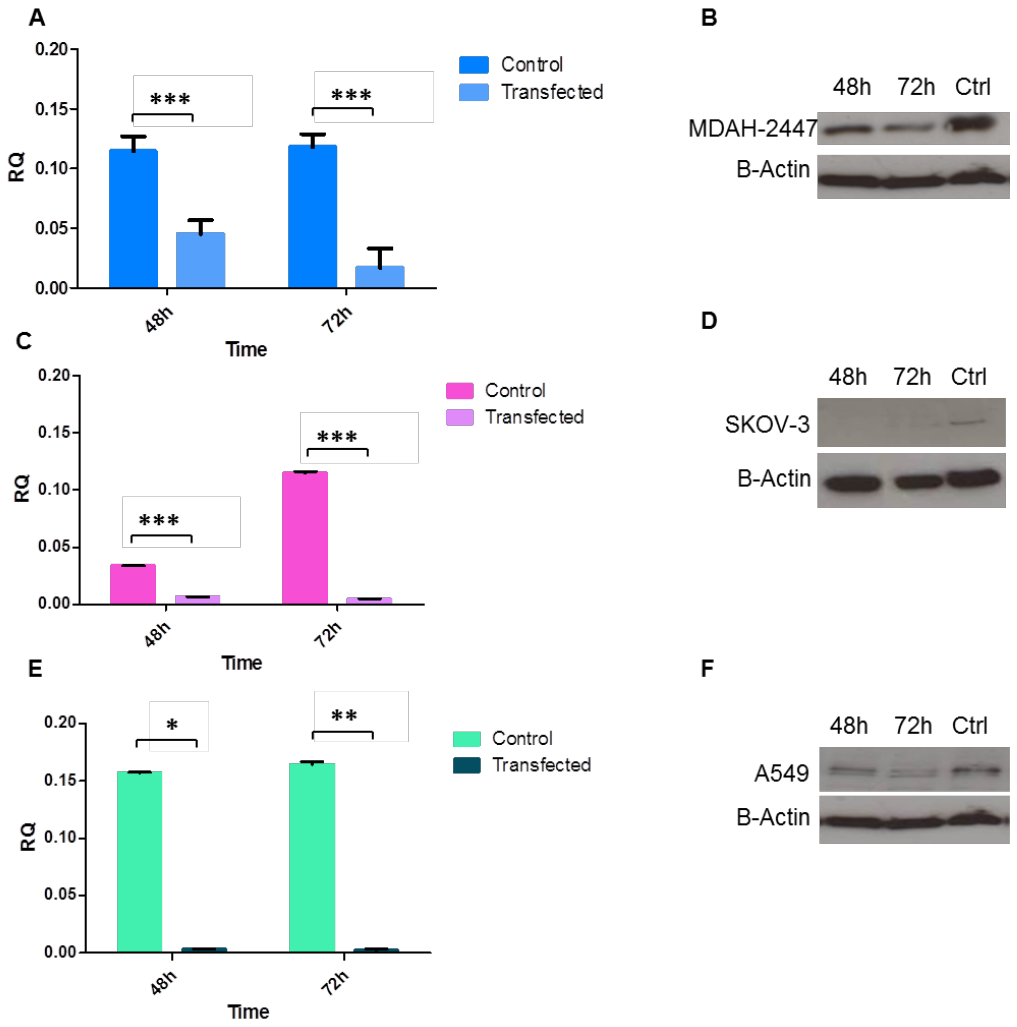


FIGURE 5.23: **RT-qPCR and Western blot analysis of RAD51AP1 silencing in cell lines.** Silencing of the RAD51AP1 gene was measured by RT-qPCR and western blotting. An F-test was performed to assess the variance between the two groups, and two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance. A substantial and significant down-regulation in RAD51AP1 expression post transfection was seen at 48 and 72h at qPCR (mRNA) level and mirrored in Western blot (protein) data, compared to scrambled controls, across (A-B) MDAH-2774, (C-D) SKOV-3 and (E-F) A549 cell lines, ****p-value<0.0001* and ****p-value=0.0002; ****p-value<0.0001* and ****p-value<0.0001*, **p-value=0.0147* and ***p-value=0.0067*, respectively.

A complete downregulation of RAD51AP1 at gene level and protein level was evident as early as 48h post-transfection for SKOV-3 cells. A similar reduction in mRNA level was evident for MDAH-2774 and A549 cells, however further protein inhibition was achieved at 72h post-transfection.

5.4.4 Proliferative potential of RAD51AP1 knockdown cell lines

Next, we investigated the effects of RAD51AP1 knockdown on proliferation within the three cancer cell lines. Images of cells in culture were taken to visualise any differences in cell growth at baseline, 48, and 72h, using the Fliod microscope (Applied Biosystems), as shown in Fig 5.24 to 5.26. A scrambled control for each cell line was used as a comparison, to rule out toxicity from the transfection reagents as a contributor. Cells were also counted using an automated cell counting system, (Countess cell counting chamber slides from Thermo Fisher), whereby cells were scraped off at baseline, 48, and 72h, and all cells were collected to establish levels of cell death.

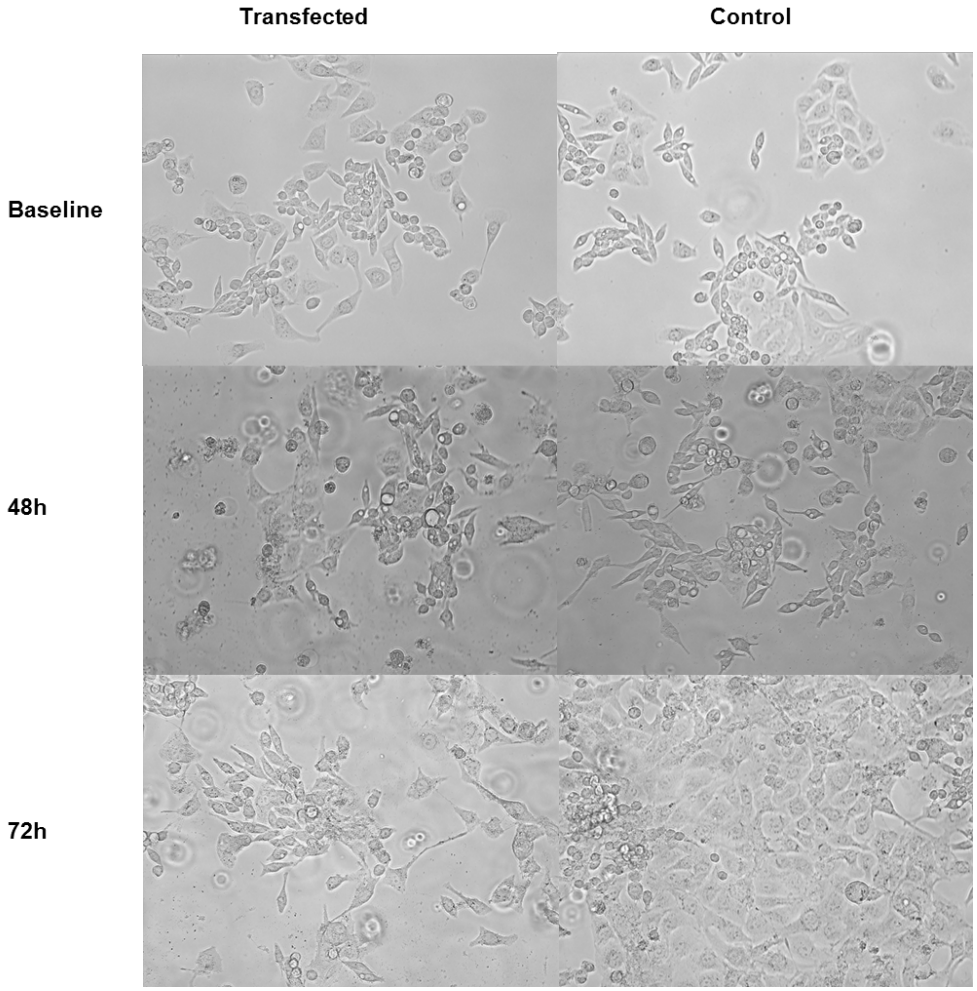


FIGURE 5.24: **MDAH-2774 transfected cells.** Images of MDAH-2774 transfected cells taken whilst in culture post transfection with siRAD51AP1, at baseline, 48, and 72h. Cell growth is shown to be slower in transfected cells compared to controls x40 magnification

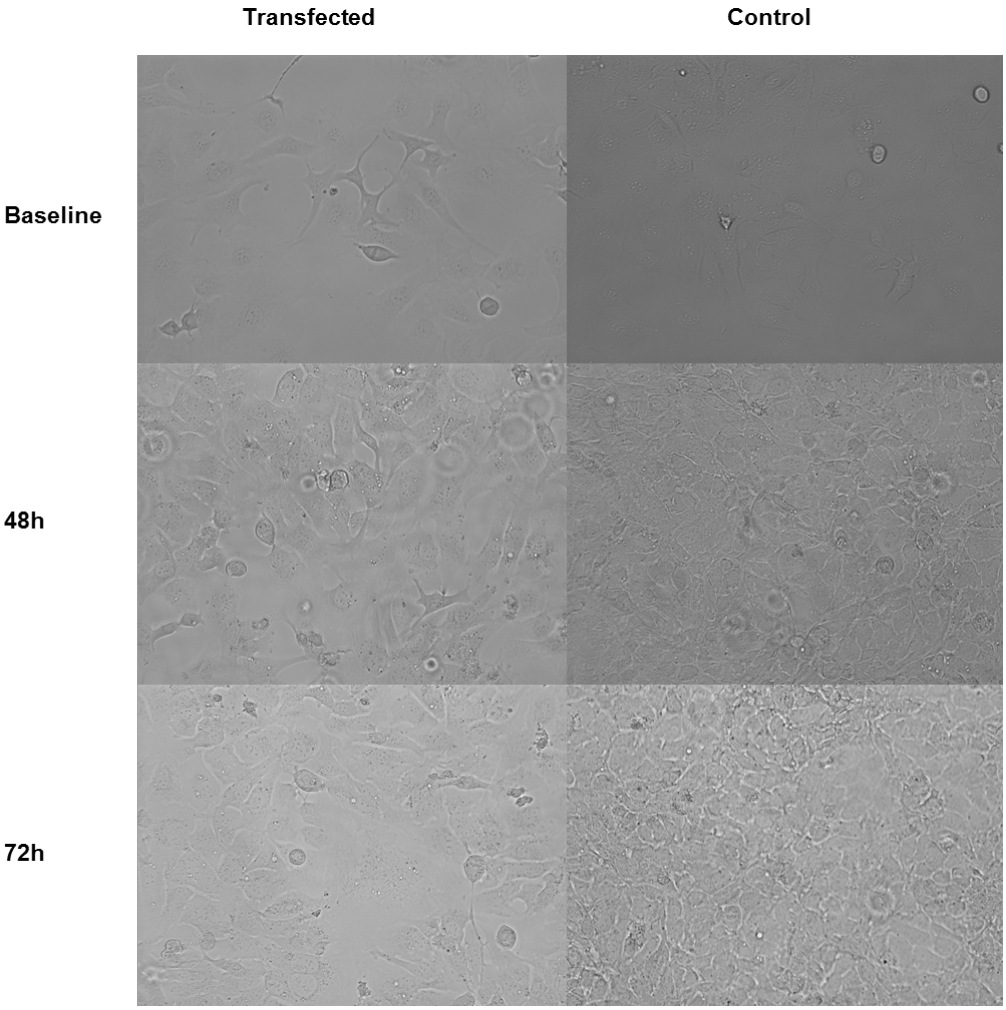


FIGURE 5.25: **Transfected SKOV-3 cells.** Images of transfected SKOV-3 cells taken whilst in culture post transfection with siRAD51AP1, at baseline, 48, and 72h. Cell growth is shown to be slower in transfected cells compared to controls.

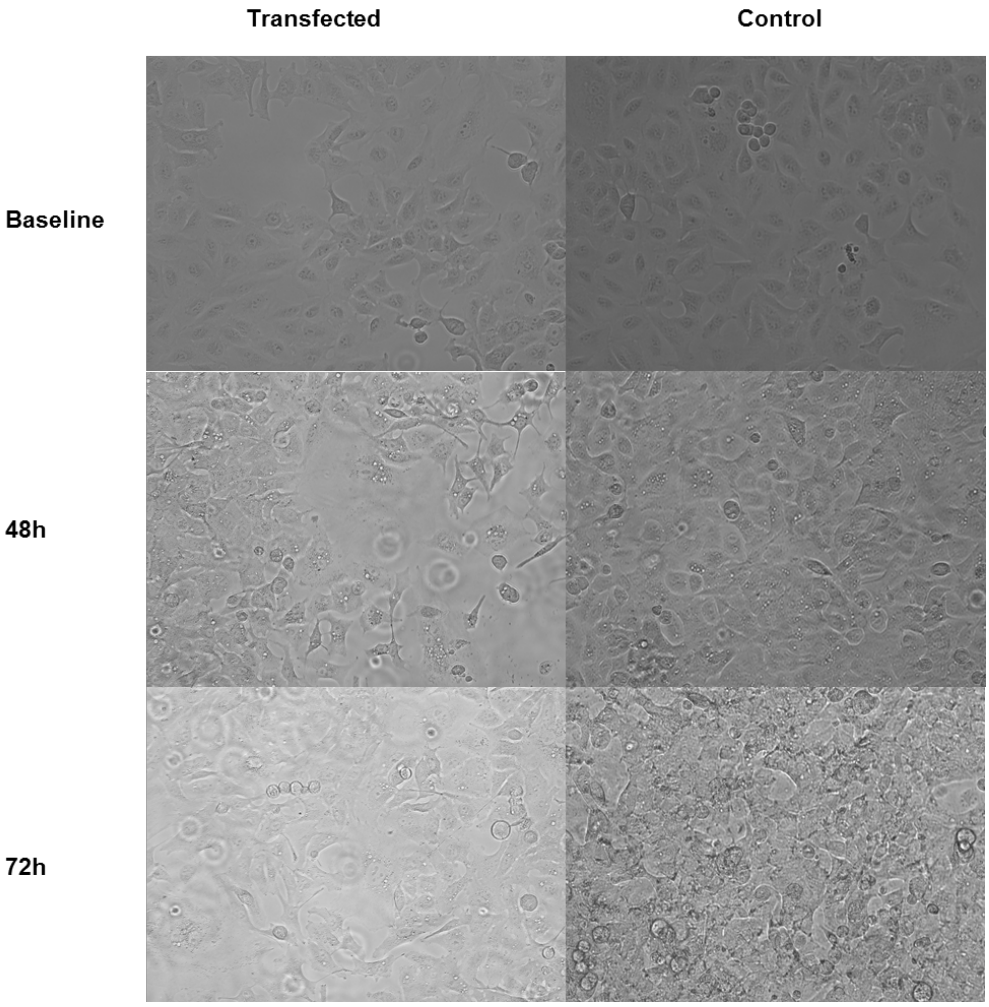
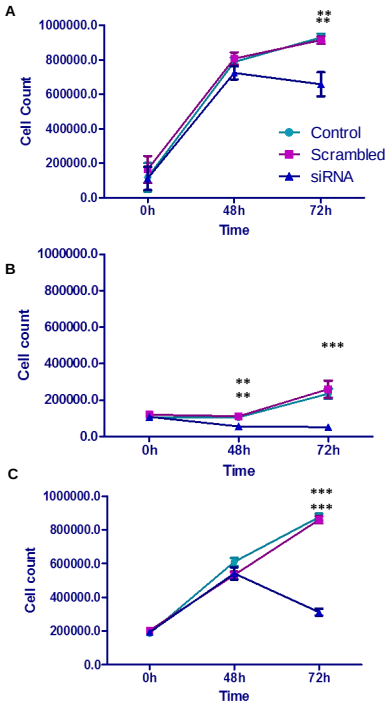


FIGURE 5.26: **Transfected A549 cells.** Images of transfected A549 cells taken whilst in culture post transfection with siRAD51AP1, at baseline, 48, and 72h. Cell growth is shown to be slower in transfected cells compared to controls.

Phase microscopy images show a slowing down of the growth of all 3 cell lines once transfected with the RAD51AP1 siRNA, compared to the scrambled control. Cell proliferation was quantified by use of the Countess cell counting chamber slides (Thermo Fisher), and plotted in Fig 5.27. An F-test was performed to assess the variance between the two groups, and two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance.



MDAH-2447	Average Cell death count	
	48h	72h
siRNA	270 000	240 000
Scrambled Control	270 000	220 000

SKOV-3	Average Cell death count	
	48h	72h
siRNA	325 000	390 000
Scrambled Control	350 000	400 000

A549	Average Cell death count	
	48h	72h
siRNA	300 000	365 000
Scrambled Control	310 000	370 000

FIGURE 5.27: **Cell proliferation charts, reporting viable cell counts at baseline, 48, and 72h post transfection.** Cell proliferation is shown to be reduced in transfected cell lines compared to scrambled controls and controls (non-treated cells). (A) MDAH-2774 at 72h, *p-value=0.0066* and *0.0049*, respectively. (B)SKOV-3 cells at 48 h, *p-value<0.0001* and *p-value=0.004*, respectively, and at 72h, significance only reached in scrambled controls, *p-value=0.0002*. (C) A549 cells at 72h, *p-value<0.0001* for both scrambled controls and untreated controls. In addition, average cell death figures were also tabulated, showing no difference in cell death for siRNA transfected cells and scrambled control cells.

Silencing of RAD51AP1 resulted in significant inhibition of cell growth at 72 hours for all 3 cell lines compared to untransfected control and scrambled siRNA cells; whereas a modest but significant decrease in cell proliferation was also evident for SKOV-3 at 48h. The similar average cell death values demonstrate that the reduction in cell numbers is not due to increased cell death in the siRNA cells (due to possible toxicity), but rather a true reflection of reduction in cell growth.

A cell cycle assay was also carried out to determine the frequency of cells at different cell cycle phases, particularly apoptotic cells. This was achieved by use of Propidium Iodide (PI). Transfected cells were collected at 48 and 72h and stained with PI, before being observed by flow cytometry. The results were plotted as percentages in Fig 5.28.

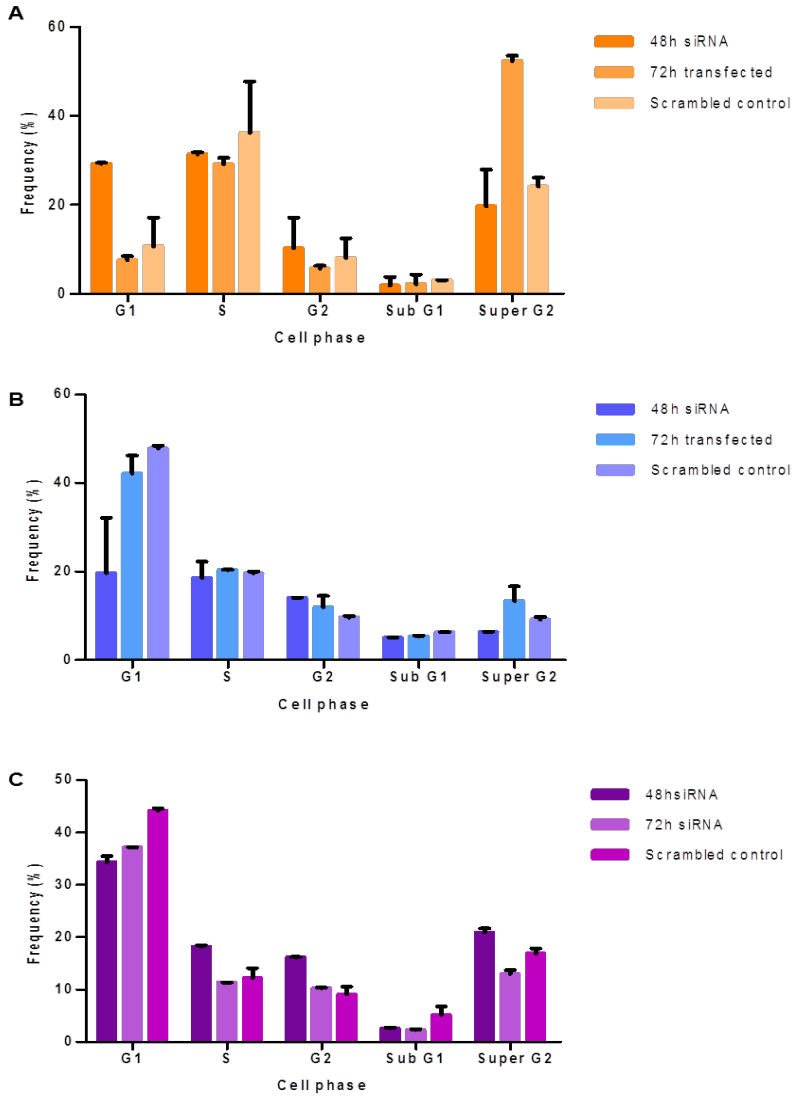


FIGURE 5.28: **Cell cycle assay of MDAH-2774, SKOV-3 and A549 cell lines.** Assay results for (A) MDAH-2774, (B) SKOV-3 and (C) A549 cell lines. All show very low frequencies of cells in sub G1 phase, suggesting a low proportion of apoptotic cells.

Cell cycle analysis also shows a small proportion of sub G1 phase cells in all cell lines, suggestive of little apoptotic activity. These results support the previous findings of cell growth suppression as opposed to cell death due to toxicity, suggesting the effects seen are cytostatic rather than cytotoxic.

5.4.5 Effects of RAD51AP1 silencing on apoptotic, metastatic markers, and the pro-survival mTOR signalling pathway

To gain a better insight into the inhibition of cell growth by siRNA, the expression of certain apoptotic genes and pro-metastatic markers known to be involved in lung and ovarian cancer were measured by RT-qPCR (see Fig 5.29 and 5.30). An F-test was performed to assess the variance between cancer and control groups, and two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance.

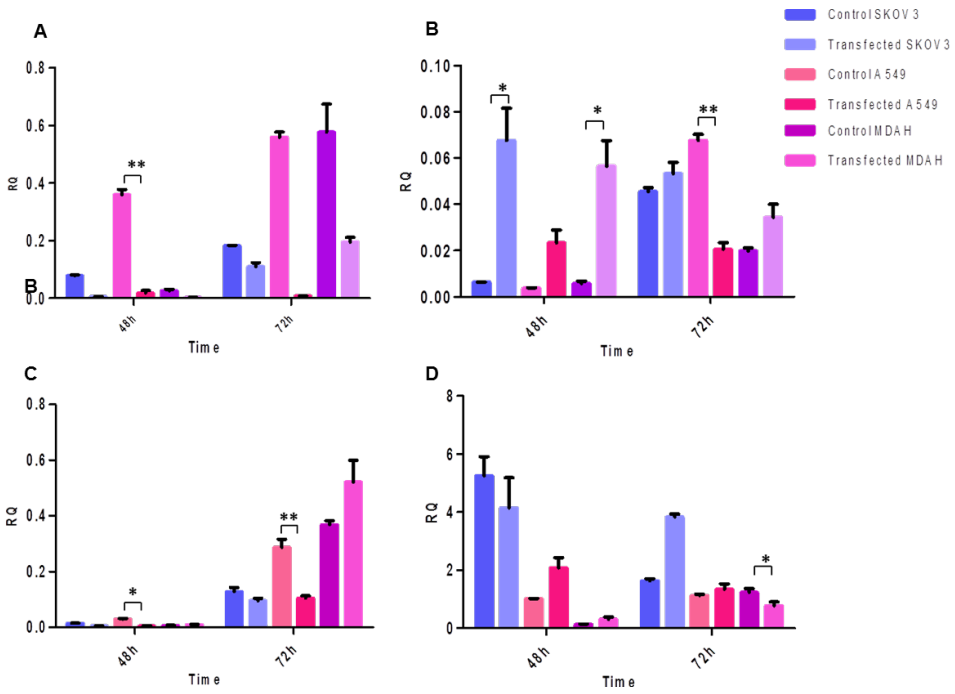


FIGURE 5.29: **RT-qPCR results of expression of pro-apoptotic markers in RAD51AP1 silenced lung and ovarian cell lines compared to scrambled controls.** (A) BAX expression decreases in all cell lines at 48 and 72h, with significant decreases in A549 at 48h, ***p-value=0.0036*. (B) FAS expression increases in all cell lines at 48h, however by 72h, this increase is reduced, and in A549 significantly reduced, ***p-value=0.0014*. (C) AIP decreases in expression across all cell lines, with significance being reached in A549 at 48 and 72h (**p-value=0.0159* and **p-value=0.0271*, respectively), except for MDAH-2774 at 72h. (D) SOD, shows variable results with increase in expression in some cell lines and decreases in others compared to controls, although significance is only reached in MDAH-2774 at 72h, showing a decrease in expression, **p-value=0.0201*.

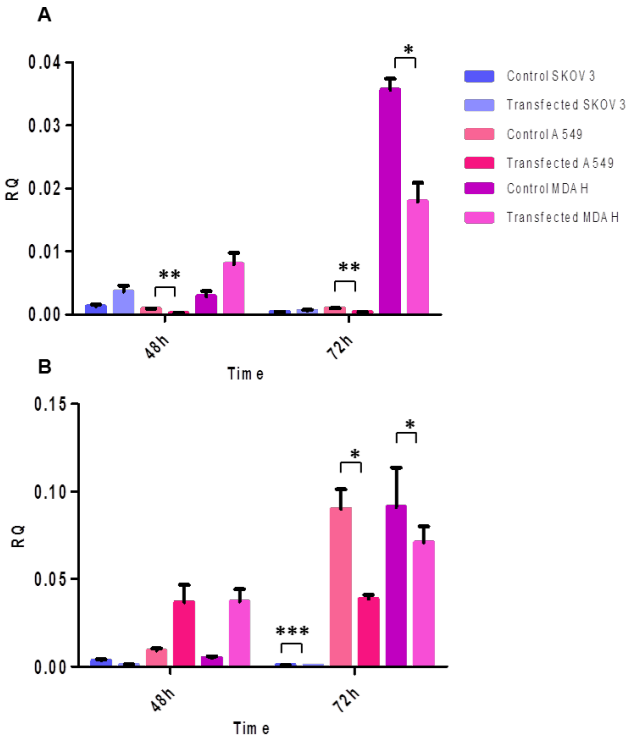


FIGURE 5.30: **RT-qPCR results of expression of pro-metastatic markers in RAD51AP1 silenced lung and ovarian cell lines compared to scrambled controls.** In general there is a decrease in pro-metastatic markers in the silenced cell lines, particularly seen at 72h. (A) SNAI1, significant decrease in expression is seen in transfected A549 at 48 and 72h and MDAH-2774 at 72h, p-value=0.0046, 0.0010 and 0.0144, respectively. (B) SOX2 expression is significantly reduced at 72h in SKOV-3, A549 and MDAH-2774 cell lines, ****p-value=0.0006*, ***0.00457* and **0.0459*, respectively

Expression of these pro-apoptotic markers was variable across the cell lines, suggesting a cell and disease specific response. BAX showed a general reduction across all cell lines in apoptosis compared to controls. FAS, AIP, and SOD showed more variable results, where increases in apoptosis were seen at 48h, this was reduced or reversed in most cases by 72h. No significant increases in pro-apoptotic markers were seen at 72h in any of the transfected cell lines. Collectively these data suggest once again a cytostatic effect for RAD51AP1 rather a cytotoxic effect.

These findings were then expanded to look at the effects of silencing RAD51AP1 on pro-metastatic markers, SNAI1 and SOX2. Silencing of RAD51P1 in all 3 cell lines led to a reduction in expression of both metastatic markers SNAI1 and SOX2, the exception being SNAI1 in SKOV-3 cells at 72h, however significance was not reached.

We also sought to investigate the effects of RAD51AP1 silencing on the mTOR signalling pathway both at the genetic and protein level (Western blot analysis), *in vitro* across all 3 cell lines, see Fig 5.31 to 5.34. F-tests were performed to assess the variance between cancer and control groups, and two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance.

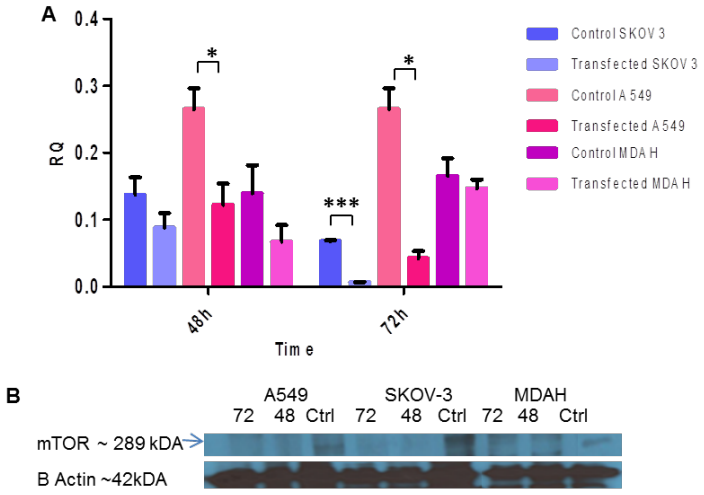


FIGURE 5.31: **Gene and protein level expression of mTOR in RAD51AP1 transfected cell lines.** (A) mTOR gene expression. A decrease in mTOR expression is evident in all cell lines, significance however is only reached in A549 at 48h (**p-value=0.0497*), and SKOV-3 and A549 at 72h, ****p-value=0.0007* and ***0.00123*, respectively. (B) Western blot analysis showed similar results of downregulation in mTOR expression across the cell lines, this is not as apparent in the MDAH-2774 cell line.

The expression of mTOR was significantly reduced in all 3 cell lines, in the siRNA transfected samples, whereas DEPTOR depicted cancer-specific expression; being upregulated in SKOV-3 and MDAH-2774, and downregulated in A549 compared to controls. The expression of key components for mTORC1 and mTORC2, namely raptor and rictor remained unaltered in SKOV-3 but was markedly reduced in A549 cells, and MDAH-2774 for raptor, but increased for rictor, again depicting a cancer specific expression.

Next, we sought to investigate changes in the expression of cancer specific proteins, by using the Human Oncology Array (R&D systems) commercial kit. Protein lysates from SKOV-3 and A549 cells were collected and processed as described in Section 2.7. Only 2 cell lines were analysed due to cost constraints. The kit functions like a Western blot assay, with the ability to process 84 cancer markers in duplicate on one blot, Fig 5.35.

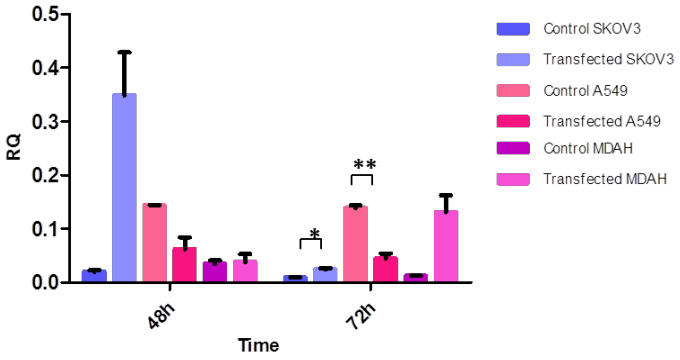


FIGURE 5.32: **Gene and protein level expression of in RAD51AP1 transfected cell lines.** A variable DEPTOR expression is visible in all cell lines, a significant increase DEPTOR is seen in SKOV-3 at 72h (**p-value=0.0345*), and significantly decreased in A549 at 72h compared to controls (***p-value=0.0032*).

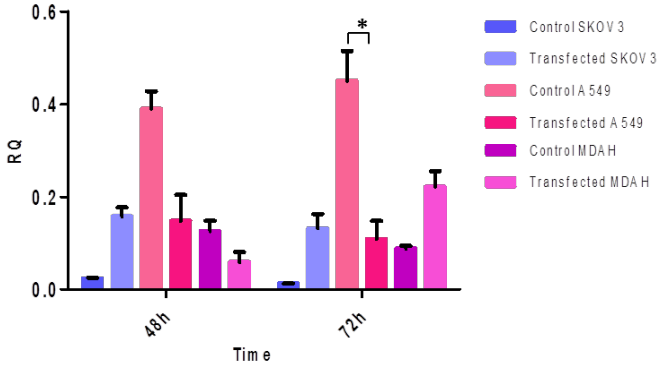


FIGURE 5.33: **Gene and protein level expression of rictor in RAD51AP1 transfected cell lines.** A variable rictor expression is visible in all cell lines, a significant decrease in rictor expression is seen in A549 at 72h compared to the control (**p-value=0.0201*).

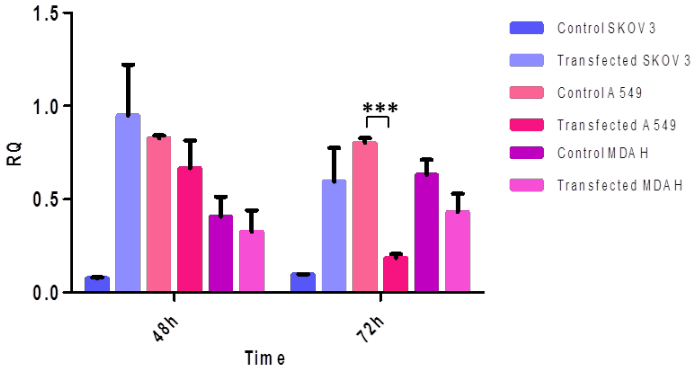


FIGURE 5.34: **Gene and protein level expression of raptor in RAD51AP1 transfected cell lines.** A variable raptor expression is also visible in all cell lines, a significant increase in raptor is seen in A549 at 72h compared to controls (****p-value=0.0004*).

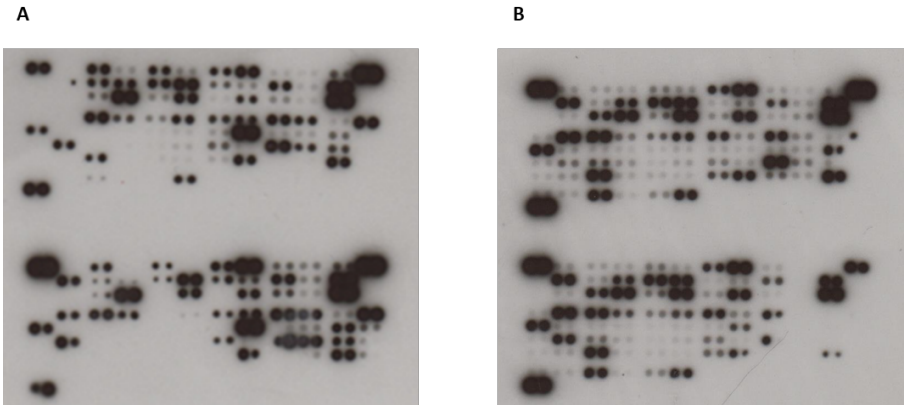


FIGURE 5.35: **Western blots of Oncology Array (R&D systems) used for the detection of predetermined targets.** The Oncology Array (R&D systems) was used to test various oncology markers in all 2 cancer cell lines, (A) A549 (B) and SKOV-3 cell lines. 10 min exposure in both cell lines.

Several exposures were taken, and optical density (OD) measured using the ImageJ software. Results can be seen plotted in Tables 5.4 and 5.5.

Up-regulated	**Function**	**Down-regulated**	**Function**
FoxO1*	Transcription factor regulates cell death	Angiopoietin-1*	Vascular development and angiogenesis
CA-125	Tumour biomarker	Progranulin*	Promotes tumorigenesis
Mesothelin	Over-expressed in cancers	SPARC*	Promotes tumour cell growth
E-Cadherin	Cell-to-cell adhesion	Vimentin*	Facilitates EMT & cancer cell migration
CCL7	Anti-tumour activity	Leptin	Involved in angiogenesis

TABLE 5.4: **Details on protein targets found to be up- or down-regulated in RAD51AP1 silenced A549 cell lines using OD analysis.** Cancer promoting proteins are shown to be down-regulated, and immune and apoptotic are up-regulated.
**expression shown to be switched off*

Up-regulated	Function	Down-regulated	Function
CCL20	Regulates immune response	DLL-1	Cell-to-cell communication
CCL2	Anti-tumour activity	ErbB3	Cell proliferation
GM-CSF	Regulates immune response	HGFR	Cell migration and invasion
Hif-1α	Role in apoptosis and angiogenesis	MMP-3	Tumour initiation
BCL-x	Regulates cell death	Survivin	Cell proliferation and prevention of apoptosis

TABLE 5.5: **Details of protein targets found to be up-regulated and down-regulated in RAD51AP1 silenced SKOV-3 cell lines using OD analysis.** The results are similar to those seen for the A549 cell lines, with mostly cancer promoting proteins to be down-regulated and immune and apoptotic to be up-regulated.

These results suggest that the silencing of the RAD51AP1 gene has regulatory effects on multiple cancer proteins and pathways in cancer. In particular many apoptotic and cell death regulated genes appear to be up-regulated, whilst cell proliferation, angiogenic, and tumorigenic proteins appear to be down-regulated, across both cell types. Interestingly, in A549 cells, vimentin is shown to be switched off (no expression) in RAD51AP1 transfected cells, and E-Cadherin expression up-regulated by over 3 fold, suggesting a mesenchymal to epithelial transition (MET) effect. A similar trend is seen in SKOV-3 cells to a lesser extent, with down-regulation of vimentin of just over 1 fold, and increase in CEACAM-5 expression, another cell-cell adhesion gene.

5.4.6 Microarray analysis of transfected SKOV-3 cells

To further analyse the effects of RAD51AP1 silencing, we carried out whole genome microarray analysis, as a non-biased transcriptomics screen. Due to financial constraints we were only able to carry out this analysis on one cell line. We have decided to concentrate on the *in vitro* model of ovarian cancer, given that screening is ineffective for this disease, allowing us to gain a better insight into the signalling of this potential biomarker. Moreover, complete protein knock down was achieved at 72h post-transfection in SKOV-3 but not in A549 or MDAH-2774 cells.

Gene expression values with a *p-value* derived from the unpaired Student t-test of less than 0.05 compared to the control, and a 2-fold change compared to the control were included, all others were discounted as displaying no significance. Overall, silencing of the RAD51AP1 gene significantly up-regulated 124 genes and down-regulated 21. Tables 5.6 and 5.7 list the top 4 up-regulated and down-regulated genes.

Gene	Log_2Fold Change	p-value	Role
Interleukin-1A	2.123492	0.003458	Immune responses, inflammatory processes, and hematopoiesis
Interleukin-1B	1.982557	0.003777	Mediates inflammatory response, cell proliferation and apoptosis
EGR1	2.833716	0.004206	Transcription factor: Tumour suppressor gene
F2RL2	1.582566	0.007408	Coagulation factor

TABLE 5.6: **Top 4 up-regulated genes generated from Microarray analysis.** Details of gene expression and functionality.

Gene	Log_2Fold Change	p-value	Role
RAD51AP1	-3.80083	6.22E-04	DNA double strand break repair
MEGF6	-2.0819	0.004201	Transcription factor
CTGF	-1.34039	0.010976	Growth factor: cell adhesion, migration, proliferation and angiogenesis
SHISA2	-1.27854	0.017224	Plays an essential role in the maturation of presomitic mesoderm cells by individual attenuation of both FGF and WNT signalling

TABLE 5.7: **Top 4 down-regulated genes generated from Microarray analysis.** Details of gene expression and functionality.

The tables above list the 4 most up-regulated and down-regulated genes. RAD51AP1 being the most down-regulated, demonstrates the level of successful silencing by the transfection. In order to validate the results of the identified genes, we performed RT-qPCR on the SKOV-3 cell line, Fig 5.36 and 5.37. Bars denote standard error of the mean. An F-test was performed to assess the variance between the two groups, and two-tailed unpaired Student's t-tests with Welch's correction for unequal variance were performed to assess significance.

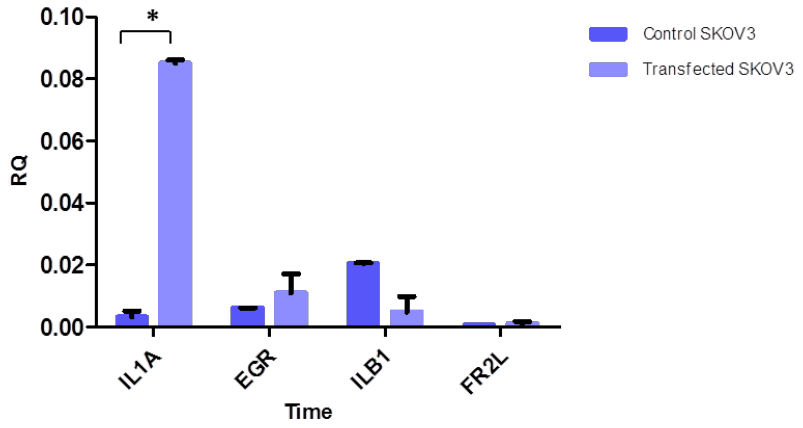


FIGURE 5.36: **RT-qPCR validation of 4 most up-regulated genes, as reported by microarray analysis.** Microarray results show an up-regulation of expression in all 4 identified genes statistical significance was only reached in ILA1 **p-value=0.0175*, with the exception of FR2L, where little or no difference is seen in expression

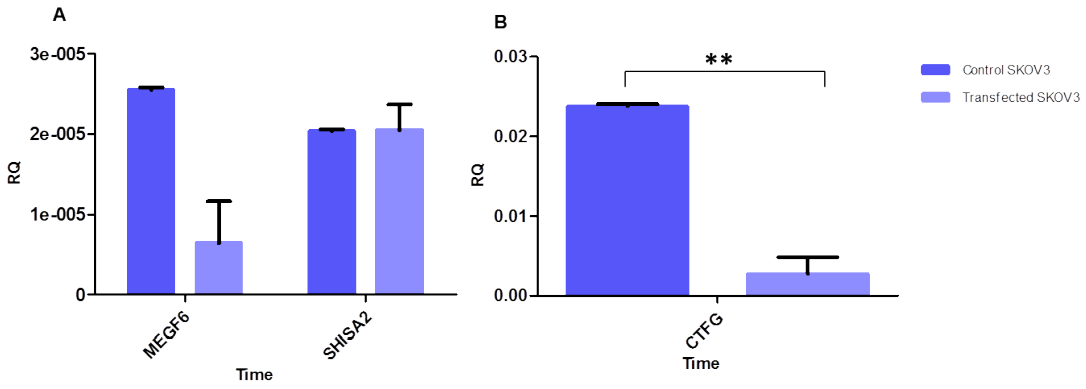


FIGURE 5.37: **RT-qPCR validation of 3 most down-regulated genes, as reported by microarray analysis.** Microarray results show (A) SHISA2 with little or no expression change, whereas MEGF6 and (B) CTCG (*p-value=0.0098*) show downregulation of expression.

Validation of the genes identified from the microarray data generally show consensus, although 2 genes showed little or no difference in expression, namely FR2L and SHISA2. This data corroborates the robustness of the generated microarray data.

Enrichment analysis using GeneSpring Software using KEGG, Reactome and Wikipathways produced 32 pathways that have 4 more genes which show changes in the expression patterns including: apoptotic signalling, MAPK signalling, Toll-like Receptor and

PI3K-Akt-mTOR signalling, Fig 5.38 to 5.41. Various genes within these pathways have been shown to be up-regulated, these a highlighted by a red dot. These results demonstrate the various different pathways affected as a result of RAD51AP1 silencing in SKOV-3 cell lines.

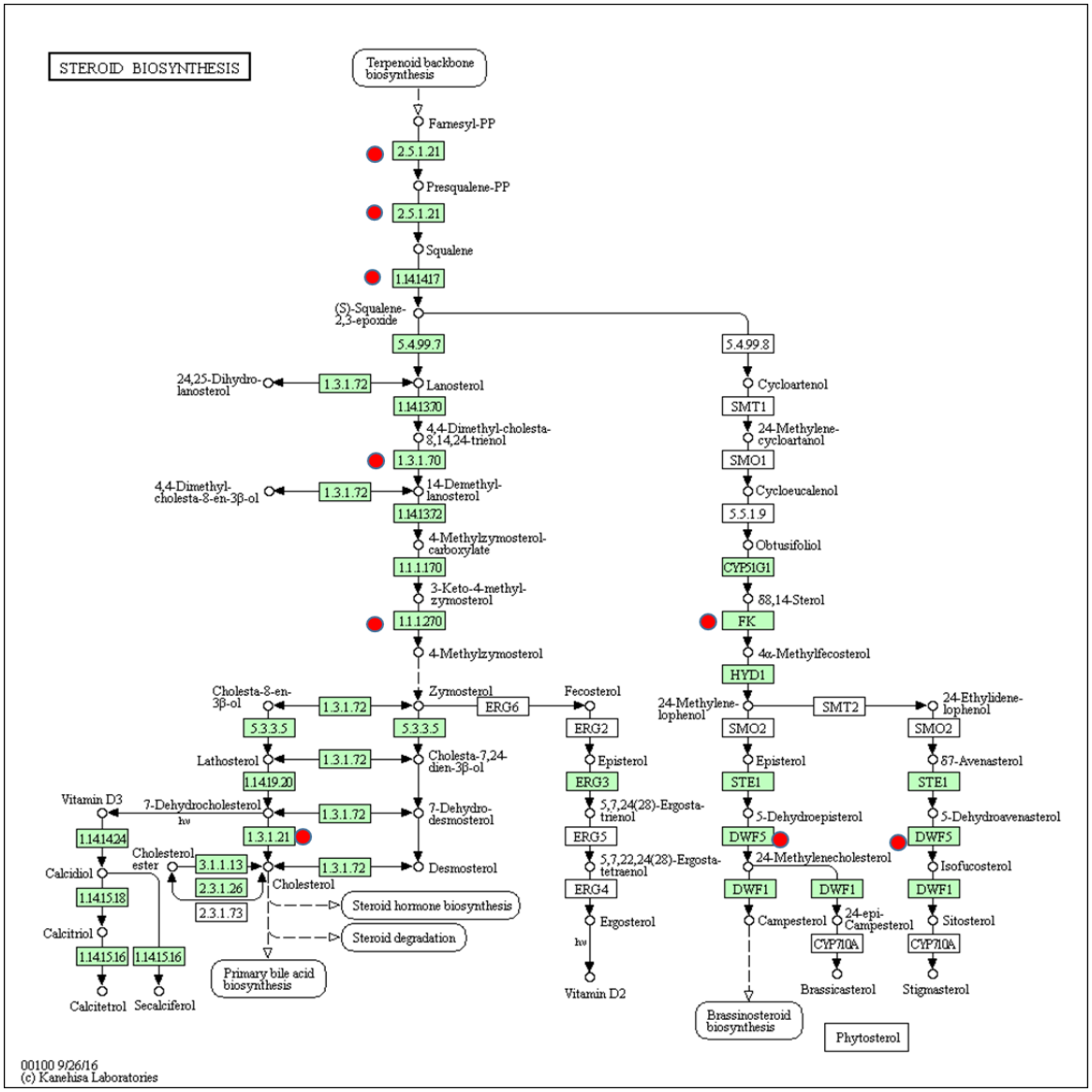


FIGURE 5.38: **Effects of RAD51AP1 silencing on the steroid biosynthesis pathway.** The steroid synthesis pathway is one of the pathways significantly affected by the silencing of RAD51AP1 in SKOV-3 cell lines. The affected genes within the pathway are highlighted with a red dot.

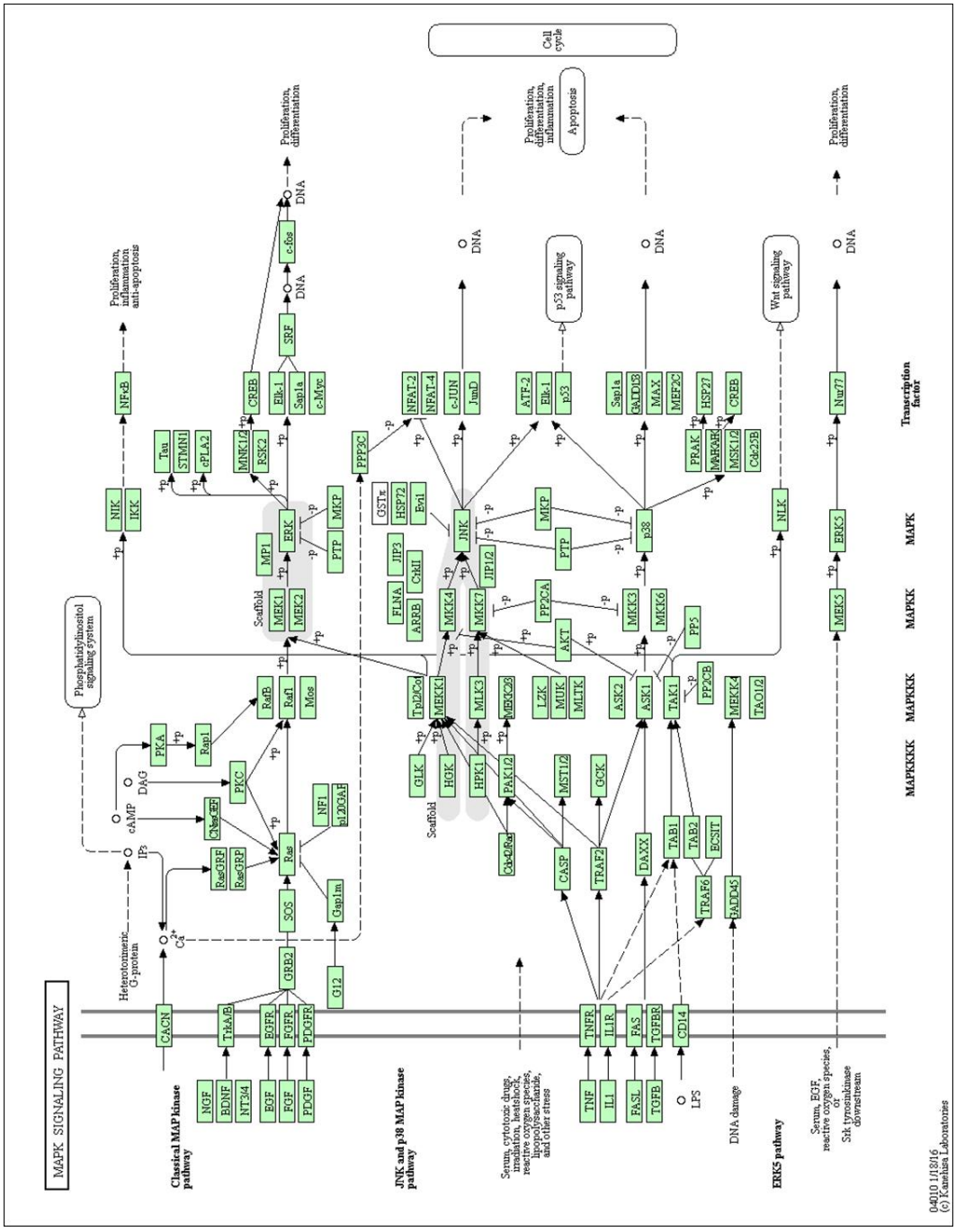


FIGURE 5.39: **Effects of RAD51AP1 silencing on the mitogen-activated protein kinase (MAPK) pathway.** The MAPK pathway is one of the pathways significantly affected by the silencing of RAD51AP1 in SKOV-3 cell lines. The affected genes within the pathway are highlighted with a red dot.

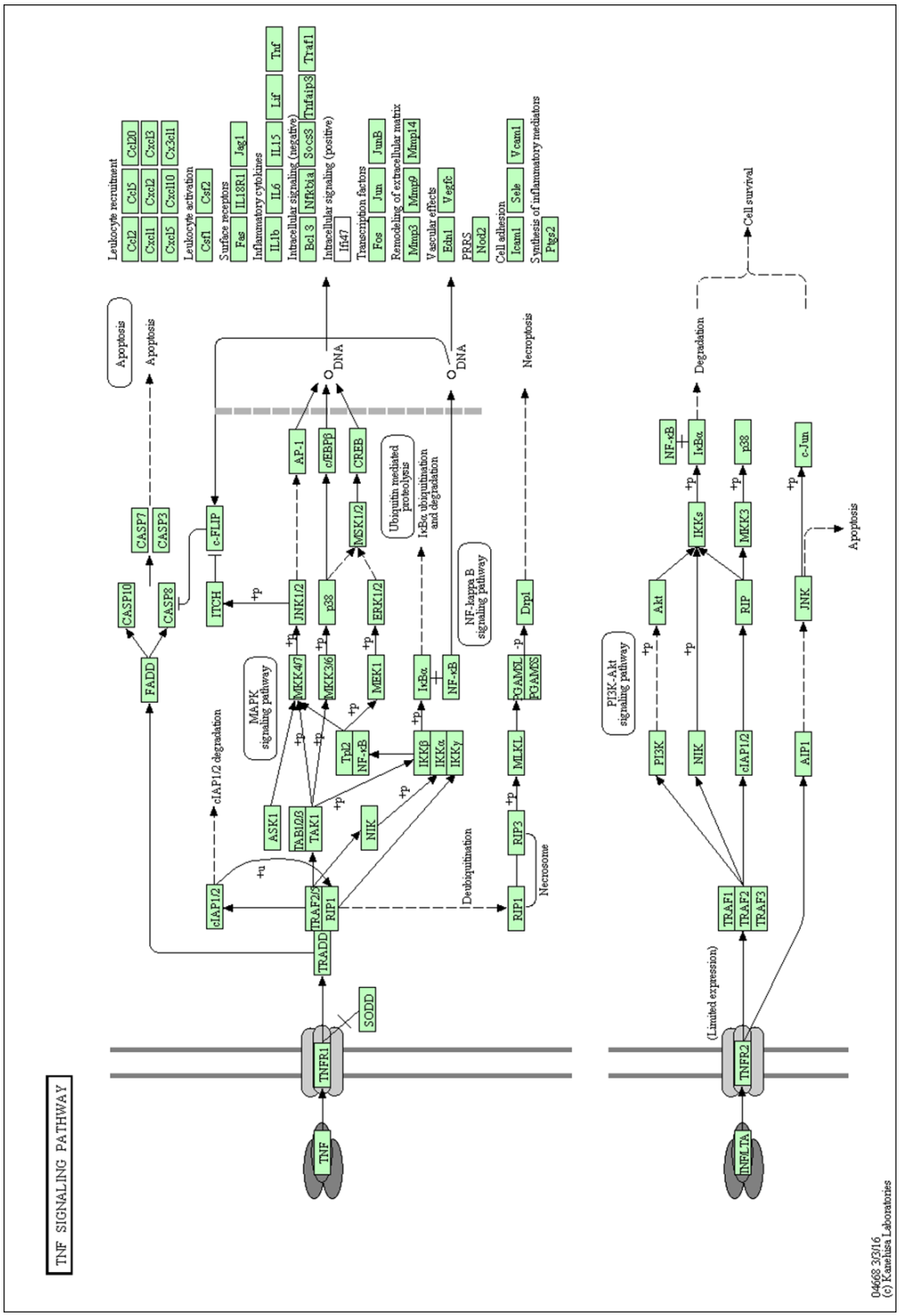


FIGURE 5.40: **Effects of RAD51AP1 silencing on the tumour necrosis factor (TNF) signalling pathway.** The TNF pathway is one of the pathways significantly affected by the silencing of RAD51AP1 in SKOV-3 cell lines. The affected genes within the pathway are highlighted with a red dot.

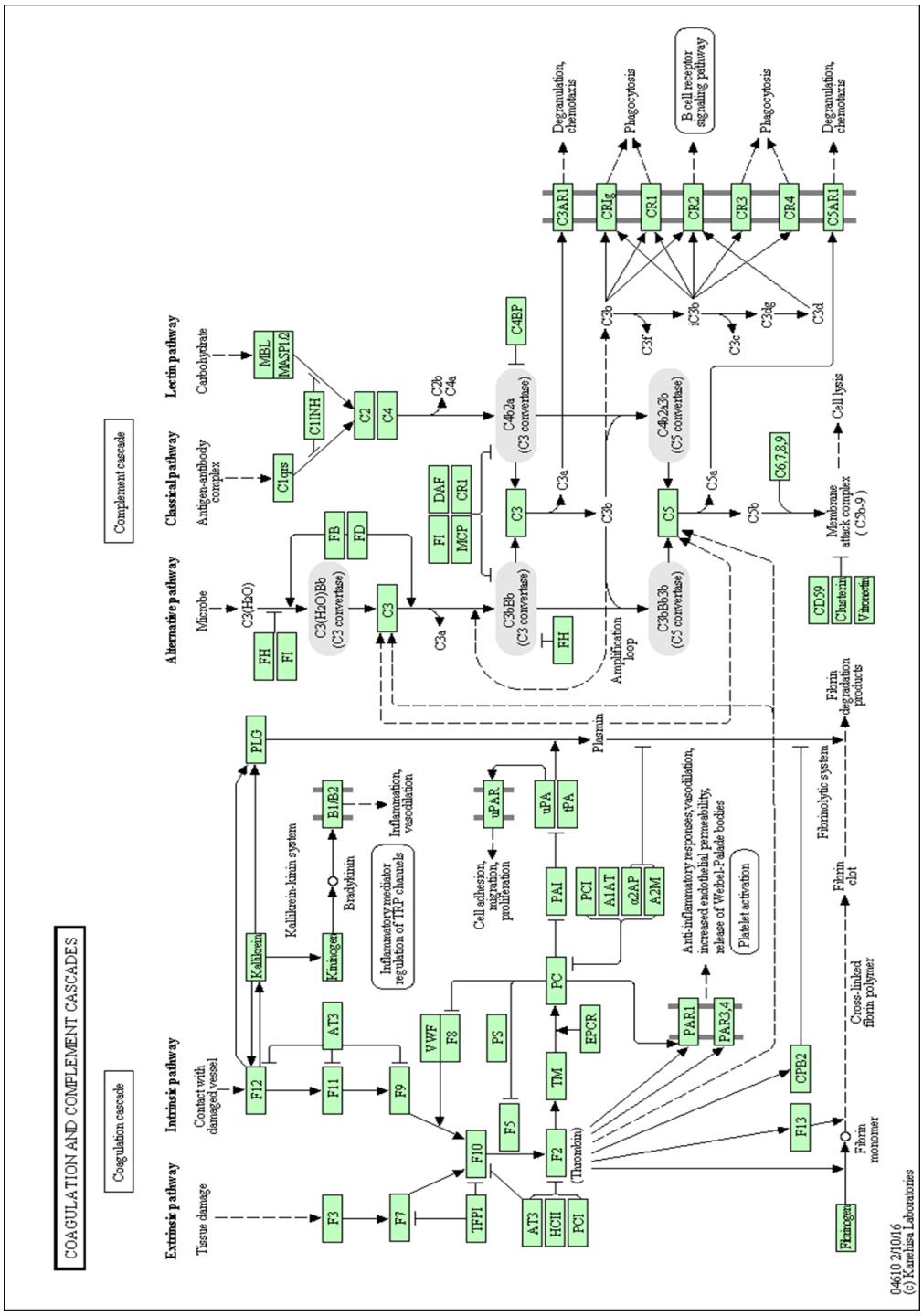


FIGURE 5.41: **Effects of RAD51AP1 silencing on the coagulation and complement cascade** The coagulation and complement cascade is one of the pathways significantly affected by the silencing of RAD51AP1 in SKOV-3 cell lines. The affected genes within the pathway are highlighted with a red dot.

5.5 Discussion

In this chapter, we provide evidence of how genes derived from unique networks can be of importance as potential biomarkers or therapeutic targets. The combination of glasso network learning, graph theory and Bayesian network prediction, and the Unique Network approach enabled the exploration and extraction of useful biological information from the existing data. The approach focuses on those genes that appear specifically in a subset of conditions - here four different cancer types. This is in stark contrast to many data integration techniques that risk "averaging" over multiple datasets. Only genes that are predictive in the subset of cancer types are preserved in the unique networks and this approach has enabled the identification of one key gene, RAD51AP1.

RAD51AP1 has been shown to participate in the homologous recombination DNA damage response pathway, as an accessory protein to RAD51, known as the main mediator (Modesti et al., 2007). Interaction of the RAD51AP1 with RAD51 greatly enhances its recombinase activity, stimulating the RAD51-mediated D loop reaction (Wiese et al., 2007). Over the past years, novel roles for RAD51AP1 have emerged, including acting as a growth promoting signalling molecule (Pathania et al., 2016), as well as being critical during the early stages of neoplasia, where replication stress occurs at higher levels than normal. Moreover, elevated levels of RAD51AP1 can then shift the balance from a precancerous lesion to cancer (Parplys et al., 2014). In the same study, RAD51AP1-deficient cells appeared to be sensitised to cisplatin's cytotoxic effects. Here, RAD51AP1 up-regulation is described in both tissue and matching blood from patients with ovarian and lung cancer compared to healthy controls. These results expand and compliment findings from previous cancer studies, where over-expression of RAD51AP1 has been described in cholangiocarcinoma tissues (Obama et al., 2008), hepatocellular carcinomas (Song et al., 2015) and acute myeloid leukaemia with complex karyotypic abnormalities (Schoch et al., 2005). Moreover, publicly available microarray data corroborate our findings, demonstrating significant up-regulation of the gene in ovarian and lung malignancies.

The over-expression of RAD51AP1 described above is associated with reduced overall survival in ovarian and lung cancer patients, and similar observations are also reported for breast cancer (Pathania et al., 2016). Furthermore, Miles *et al.* (2012), identified novel microRNA/mRNA mechanisms in ovarian cancers, demonstrating that the expression of RAD51AP1 is strongly inversely correlated with the expression of hsa-miR-140-3p, which was significantly down-regulated in the ovarian cancer samples. This down-regulation was not seen in normal ovarian tissue samples, suggesting a direct causal dysregulation of RAD51AP1 by hsa-miR-140-3p microRNA in the ovary, thus potentially playing a part in tumorigenesis (Miles et al., 2012).

With the emergence of circulating tumour cells (CTCs) (Chudasama et al., 2016), and circulating nucleic acids (ctNAs), readily available in the blood of lung and ovary cancer patients there is renewed hope of developing this tool further to identify circulating surrogate biomarkers, i.e. 'liquid biopsy' biomarkers, which would be considerably more accessible than repeated tissue biopsies. Given the over-expression of RAD51AP1 in ovary and lung cancer patients tissue and peripheral blood, compared with normal healthy controls described here, RAD51AP1 could be explored as potential biomarker for both ovarian and lung cancer.

Taking advantage of the availability of parallel tumour tissue samples for the lung cancer patients, we compared them against each other to assess the relationship of RAD51AP1 expression levels in tissue and blood samples from the same patients. In 50% of the patient samples we observed a correlation between the expression levels of RAD51AP1 in tumour tissue and blood. A weaker association was observed in a further 27% of the patients, while the remaining samples showed very low or no correlation at all between the expression levels of RAD51AP1 in tumour and blood. These results suggests that blood 'liquid biopsy', while a powerful approach might be contingent in some degree to inter-patient variability. Further analysis using a much larger cohort in the future will evaluate this potential inter-patient variation.

To investigate the role of RAD51AP1 in tumorigenesis, we generated RAD51AP1 silenced ovarian and lung cell lines using siRNA. We demonstrate for the first time, that suppressing RAD51AP1 in SKOV-3, MDAH-2774, and A549 cells leads to a significant decrease in cell proliferation at 72h post-transfection. Similar results were produced in a study with RAD51AP1 suppression using intrahepatic cholangiocarcinoma (Obama et al., 2008). This marked reduction in cell proliferation could be attributed to reduced expression of pro-metastatic genes. Following RAD51AP1 siRNA transfection into SKOV-3, MDAH-2774 and A549 cells, SOX2 (Sex-determining region Y (SRY)-Box2) was significantly down-regulated. SOX2, is a transcription factor belonging to the SOX family. SOX2 is involved in embryonic development and regulating stem cell fate, as well as conferring and maintaining stem cell identities (self-renewal, an important hallmark of cancer-initiating cells). SOX2 has also been linked to epithelial mesenchymal transition (EMT), with studies showing that knocking down SOX2 resulted in decreased expression of EMT drivers, including SNAI1 (Herreros-Villanueva et al., 2013). Interestingly, SOX2 over-expression occurs in the majority of patients with high-grade serous ovarian cancer (HGSOC) irrespective of tumour stage (Hellner et al., 2016). However, a more cell-specific effect was noted for SNAI1, since transfection of RAD51AP1 siRNA did not alter its expression in SKOV-3 cells but was significantly down-regulated in MDAH-2774 and A549

cells. SNAI1 is a transcription factor, shown to down regulate E-cadherin expression, resulting in loss of cell to cell adhesion, and facilitation of EMT, promoting metastasis (Bell and Watson, 2009).

With regards to Bax, this is not the first time that a variable expression is documented. For example, when the FSTL1 gene was silenced in ovarian cancer cells, a decrease in proliferation was noted while the Bax gene expression remained unchanged (Chan et al., 2009). It is possible therefore, that the mitochondrial-initiated intrinsic pathway of apoptotic events might not be involved, and the decrease in cell proliferation is due to compromised mTOR signalling.

mTOR signalling regulates growth, proliferation, controls cellular behaviour and acts as a nutrient and amino acid sensor (Foster et al., 2014; Mparmpakas et al., 2012; Foster et al., 2010; Mparmpakas et al., 2010). Previous studies from our laboratory have shown involvement of this pathway in drug resistance in ovarian cancer, mediation of responses of the complement protein C1q (Kaur et al., 2016), and as a therapeutic target, given that rapalogues can exert an inhibitory effect on ovarian cancer cells (Rogers-Broadway et al., 2016). Here we demonstrate that mTOR (the key component for both mTORC1 and mTORC2 complexes) is significantly down-regulated in all 3 cell lines both at gene and protein level. For the remaining of the mTOR pathway genes, there is some cell-specificity. For example, DEPTOR appeared to be upregulated in SKOV-3 and MDAH-2774, whereas the key components for mTORC1 and mTORC2 namely raptor and rictor respectively, were down-regulated only in A549 cells, and raptor only in MDAH-2774, but not in SKOV-3 cells. Collectively, these results point towards a compromised mTOR signalling in all 3 cell lines *in vitro*. In SKOV-3 and MDAH-2774 cells, the decreased mTOR signalling along with the increase in DEPTOR would inhibit the activity of complexes. In A549 cells, similarly compromised activity occurs through the reduction in expression of the key components (raptor and rictor) of mTORC1 and mTORC2 complexes. Future work on protein expression of these components or changes in the phosphorylation status of downstream components, like S6 kinase (S6K) would provide a more detailed insight into the effects of RAD51AP1 on mTOR signalling.

Silencing of the RAD51AP1 gene has been shown to influence multiple cancer proteins, as demonstrated in the results of the Human Oncology array assay. Multiple apoptotic markers where shown to be up-regulated in RAD51AP1 silenced cells, and a down-regulation of tumour promoting markers. Moreover, evidence suggested the process of MET in RAD51AP1 knockdown cells, with a decrease in vimentin and up-regulation in epithelial cell-to-cell adhesion markers, hinting at the reversion of cancer migration and progression (Heerboth et al., 2015). It is well known that EMT is involved in tumour migration and metastasis. Whereby cells take on mesenchymal features and loose their rigid

cell structure by loss of epithelial properties and gain of mesenchymal, with vimentin being one of these drivers of EMT (Bell and Watson, 2009).

To summarise, using unique regulatory network analysis we were able to filter and select genes of interest with clinically relevant expression signatures, from a wealth of existing microarray data. The main risk of unique networks is small sample sizes that may affect the significance of the connections identified.

Finally, we found that RAD51AP1 is not only over-expressed in lung and ovarian cancer tissue but also readily identified in the peripheral blood of these patients as opposed to normal healthy controls, thus, showing potential as a biomarker, and even as a 'liquid biopsy' biomarker. Further studies using 'liquid biopsies' will also provide evidence as to whether this gene has diagnostic value. In addition, the knock down of RAD51AP1 reduced cell proliferation *in vitro* possibly offering itself as a therapeutic target for ovarian and lung cancers. Preliminary work here suggests that RAD51AP1 has the potential to influence many pathways from pro-metastatic genes such as SOX2, transcription factors such as SNAI1 and the mTOR pathway.

Chapter 6

Evaluation of molecular signatures in the blood of cancer patients, using cancer panels, copy number instability analysis, and RNA sequencing

6.1 Introduction

'Liquid biopsies' open the potential to carry out biological characterisation of tumours by sampling blood or plasma, offering prospects of biomarker and unique gene signature discovery. Researchers have explored the use of liquid biopsies in many forms, including quantification and characterisation of circulating tumour cells and circulating tumour DNA, and their potential clinical utility to inform on prognosis and response to treatment. The question however remains on how similar the tumour profile identified in blood is to that of the tumour, and how this may influence clinical decisions.

In order to address this issue, a number of technologies are available to create a comprehensive picture of the tumour genome and advance the field of 'liquid biopsy', by exploitation of circulating tumour DNA & RNA, tumour genes expression profiles, and copy number changes. Copy Number Instability (CNI) is a measure of copy number changes (Beck et al., 2013). Specific human cancers are reported to share recurrent copy number changes (Beroukhim et al., 2010). CNI assays can be obtained by various sources, including circulating tumour free DNA (ctDNA) derived from plasma/serum samples and tissue. Various methods have been used to assess chromosomal instability, including, next generation sequencing (NGS) (Weiss et al., 2017), and digital droplet PCR (Beck et al., 2013). Elevated CNI scores are indicative of increased copy number changes and instability, a hallmark of cancers, often characterised by large physical or functional somatic gains or losses in the ctDNA (Weiss et al., 2017; Beck et al., 2013; Beck et al., 2010).

Several recent studies report positive results, including Weiss *et al.* (2017), who showed the ability of CNI scores to accurately predict response to immunotherapy in advanced and metastatic cancers (including, colorectal, lung, breast, pancreatic, renal, colon, melanoma and others) (Weiss et al., 2017). Other studies report impressive specificity and sensitivity results, 95% and 90%, respectively, in the early detection of breast cancers (Beck et al., 2010), showing this to be a promising tool for 'liquid biopsies'.

Another method of 'liquid biopsy' evaluation is the use of whole blood transcriptomics analysis. RNAseq is an attractive approach to transcriptomics profiling. RNAseq provides precise measurement of levels of transcripts and their isoforms compared to other methods (Wang, Gerstein, and Snyder, 2009). This technology allows for a high-throughput, and overcomes several limitations faced with classical methods, e.g. limited outputs. Moreover, RNAseq is able to identify transcriptomic complexities, such as allele-specific expression and novel promoters and isoforms (Oshlack, Robinson, and Young, 2010). A study by Zhang *et al.* (2015), compared microarray and RNAseq based transcriptomic models as an end point predictor in lung cancers, concluding that RNAseq outperforms microarrays in determining the transcriptomic characteristics of cancer, but performed similarly in end point prediction (Zhang et al., 2015). However, some levels of inaccuracy in using RNAseq results have also been reported (Su et al., 2014).

Recent studies have shown value in transcriptomics analysis of RNA extracted from whole blood. Karkokawa *et al.* (2017), analysed whole blood RNA extracted from colorectal cancer patients by microarray analysis, assessing sense and anti-sense RNA. They observed significant differences between the cancer and control cohort, including an anti-sense RNA, haloacid dehalogenase-like hydrolase domain-containing 1 (HDHD1). Further analyses revealed that samples analysed 3 months post-surgical excision of the tumour showed significantly reduced levels of HDHD1 compared to samples from the same patient one week post-surgery (Kurokawa et al., 2017). Another study in castration resistant prostate cancers, reported the prognostic value of whole blood RNA analysis. The study hybridised whole blood RNA on to the Affymetrix U133plus2 microarrays. A nine-gene signature was identified by this method, shown to be associated with overall survival. The authors concluded the benefits of a whole blood RNA transcriptomics approach in identifying novel gene signatures, in this case that could stratify patients with castration resistant prostate cancer in to 2 distinct prognostic groups (Olmos et al., 2012). Friedlander *et al.* 2017, found that whole blood RNA transcriptomics in melanoma patients, provided a validated pre-treatment mRNA classifier model that predicted clinical response. Moreover, the data suggested that the model captured a biological signature representative of genes needed for a robust anti-cancer immune response. It also identified non-responders to Tremelimumab at baseline prior to treatment (Friedlander et al., 2017).

All these findings support the utility of a transcriptomics approach to cancer, in the form of a 'liquid biopsy'. This chapter evaluates several approaches to transcriptomics evaluation in clinical samples.

6.2 Aims and objectives

- **Evaluate the use of a 79-gene cancer panel to identify novel biomarkers using peripheral blood from ovarian cancer patients and comparing it to controls." Changes in red.**
- **Evaluate the use of copy number instability analysis to identify copy number changes in lung and ovarian cancer patients.**
- **Identify novel gene signatures in matched tissue and blood samples from lung cancer patients compared to controls, using RNA sequencing.**

6.3 Declaration of contribution to the work presented in this chapter

TATAA Biocenter – 79 cancer panel

- **I extracted total RNA from whole blood of patients and controls. I then organised delivery of these samples to TATAA Biocenter in Sweden on dry ice.**
- **TATAA Biocenter, ran cancer panels on the samples. The resulting qPCR data was then sent to me by TATAA.**
- **I then cross checked the data supplied with Oncomine and Kaplan Meyer plots to dissect the data further in terms of expression in tissue and overall survival.**

Table 6.1 gives a detailed description of contributions from each participating unit.

Contributions	TATAA Biocenters	DC
Patient and sample collection	-	x
Sample preparation (RNA/DNA/Plasma)	-	x
Sample processing	x	-
Processing of raw data	x	-
Data analysis and correlation with clinical samples	-	x

TABLE 6.1: **Contribution to the work presented in this chapter.** Details on specific tasks performed by TATAA Biocenters and Dimple Chudasama (DC).

Chronix Biomedics - Chromosome Number Instability

- **I collected patient FFPPE samples and processed their matched plasma. I then organised delivery of these samples to Chronix Biomedics in Germany on dry ice**
- **Chronix Biomedics, ran the copy number instability (CNI) experiments. The CNI score 171 and Circos plots were then sent to me. then cross checked the data supplied with the patient details to make correlations with stage of disease and further analyse this data in terms of chromosomal instability.**

Table 6.2 gives a detailed description of contributions from each participating unit.

Contributions	**Chronix Biomedics**	**DC**
Patient and sample collection	-	x
Sample preparation (RNA/DNA/Plasma)	-	x
Sample processing	x	-
Processing of raw data	x	-
Data analysis and correlation with clinical samples	-	x

TABLE 6.2: **Contribution to the work presented in this chapter.** Details on specific tasks performed by Chronix Biomedics and Dimple Chudasama (DC).

The Wellcome Trust Centre for Human Genetics - RNASeq

- **I collected patient samples and extracted RNA from matched blood and tissue of control and cancer patients. I then organised delivery of these samples to The Wellcome Trust Centre for Human Genetics in Oxford, on dry ice.**
- **The Wellcome Trust Centre for Human Genetics, ran the RNASeq experiments. The raw data was sent to us.**
- **I carried out the analysis with the use of a bioinformatician using 'R' and 'Python'.**
- **I then cross checked the data supplied with the patient details to make correlations and further analyse this data.**

Table 6.3 gives a detailed description of contributions from each participating unit.

Contributions	**The Wellcome Trust Centre for Human Genetics**	**DC**
Patient and sample collection	-	x
Sample preparation (RNA/DNA/Plasma)	-	x
Sample processing	x	-
Processing of raw data	-	x
Data analysis and correlation with clinical samples	-	x

TABLE 6.3: **Contribution to the work presented in this chapter.** Details on specific tasks performed by The Wellcome Trust Centre for Human Genetics and Dimple Chudasama (DC).

6.4 Results

6.4.1 79 cancer gene panel evaluation of ovarian cancer blood samples

The 79 cancer gene panel evaluation was carried out on 25 ovarian cancer blood samples and 14 female controls. Table 6.4 lists all 79 genes contained in the panel.

Genes				
UPA	CD45/PTPRC	Hjurp	MUC1	SRD5A1
ACTB	CDH1_1	HPRT1	Myc	STAT3
ADAM17	CDH2	IGFR	NCOA1	TBP
AHR	CTSD	KI67	PARP	TNFSF11
AKR1C3	CXCR4	KIT_1	PI3KCA	TOP2A
AKT2	CYP11A1	KRAS	PPIC	TP53
ALDH	CYP19A1	KRT19	PROM1	TUBB
ANXA2R	DDR1	MAL2	PTCH1	VEGFA
AR	EPCAM	MCL1	PTEN	VEGFR1
AURKA	ERBB2	MCM4	RAD51	VEGFR2
B2M	ESR1	MET_1	RPLP	VIM
BCL2	FOXO	MRP1	RUNX2	WHSC1L1_L1
CCND1	GAPDH	MRP2	SATB1	WHSC1S1_S1
CCNE2	GUSB	MRP4	SLC6A8	XIAP
CD24L4	H2AFZ	MRP5	SNAI1	YWHAZ
CD44_all	HDAC2	MTOR	SPINK1	

TABLE 6.4: **List of all 79 markers contained on the cancer gene panel.**

Expression levels of the diseased samples were compared against controls and plotted accordingly. The analyses revealed 28 genes to be significantly up or down-regulated in cancer blood samples compared to controls, Fig 6.1 and Table 6.5.

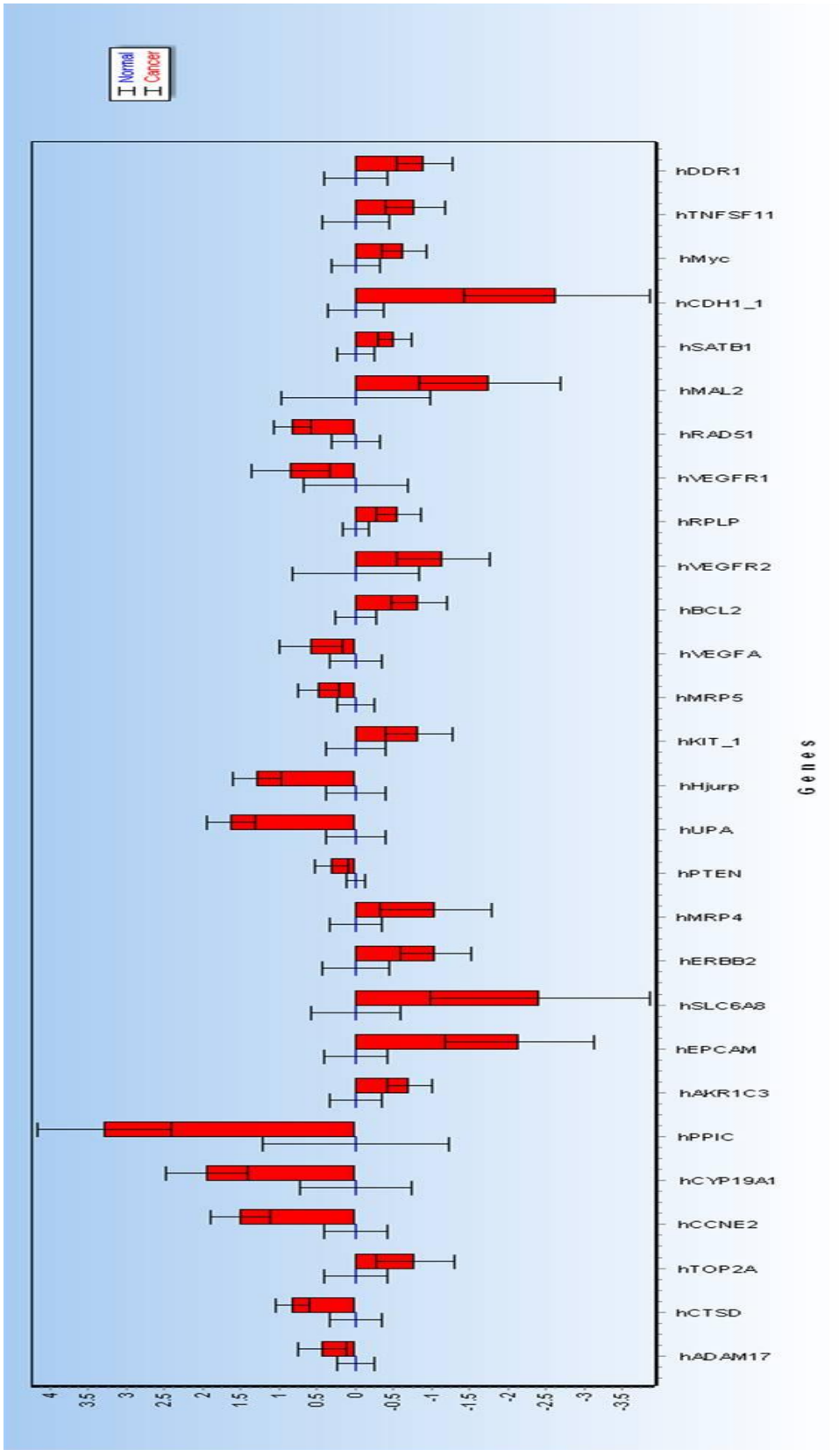


FIGURE 6.1: **Gene expression of ovarian cancer blood samples compared to controls.** Data generated from ovarian cancer panels is plotted as a box plot chart for both ovarian cancer and controls. Expression of 28 genes found to be most significantly (p-value<0.05) up or down-regulated in cancer cohort compared to controls.

GeneID	Regulation change (Cancer vs Normal)	P-Value
UPA	Up-regulated	0.000000125
Hjurp	Up-regulated	0.000005849
CCNE2	Up-regulated	0.000006464
PPIC	Up-regulated	0.000040492
CYP19A1	Up-regulated	0.00006855
CTSD	Up-regulated	0.000081921
RAD51	Up-regulated	0.000178834
DDR1	Down-regulated	0.001770788
EpCAM	Down-regulated	0.002151251
BCL2	Up-regulated	0.00218176
CDH1_1	Up-regulated	0.002229509
AKR1C3	Up-regulated	0.002304223
ERBB2	Up-regulated	0.004033908
SATB1	Up-regulated	0.005105688
Myc	Up-regulated	0.006669233
RPLP	Up-regulated	0.007967613
KIT_1	Up-regulated	0.012608669
TNFSF11	Up-regulated	0.013825421
MAL2	Up-regulated	0.014203491
SLC6A8	Up-regulated	0.016099728
MRP5	Down-regulated	0.020914345
VEGFR2	Up-regulated	0.026447178
TOP2A	Up-regulated	0.034451156
VEGFR1	Down-regulated	0.040150036
MRP4	Up-regulated	0.040338224
PTEN	Down-regulated	0.043608118
ADAM17	Down-regulated	0.055675039
VEGFA	Up-regulated	0.05752768

TABLE 6.5: **Data of change in expression and *p-values* (generated by a two-tailed t-test) comparing cancer samples to controls.** Genes are listed from most to least significant, with colour coding separates most significant (*p-value<0.0001*) to significant (*p-value<0.0001-0.049*), and borderline significant (*p-value<0.05-0.06*).

Several genes were shown to be differentially expressed in cancer samples compared to controls, highlighting that genetic material from the cancer is carried in blood and could be potentially exploited as a 'liquid biopsy'. Particularly of interest is the down-regulation of EpCAM, which has also been shown in the literature (Barriere et al., 2014; Raimondi, Nicolazzo, and Gradilone, 2015). RAD51 is also shown to be over-expressed in cancer blood samples, and is of interest due to its close relationship with RAD51AP1 discussed in Chapter 5.

To further evaluate these results we sought to exploit existing *in silico* data of expression level and overall survival, using Oncomine and Kaplan Meier plotter. This analysis revealed 5 genes to be up-regulated in both blood from the cancer panel analysis, and tissue according to the Oncomine data, Table 6.6. Overall survival data was also evaluated for these 5 genes, using the KM plotter online tool.

Genes	Expression in blood	Expression in tissue (Oncomine)	Overall Survival (KM Plotter)
VEGFA	Up-regulated	Up-regulated	Very Significant
Hjurp	Up-regulated	Up-regulated	Very Significant
CCNE2	Up-regulated	Up-regulated	Significant
CTSD	Up-regulated	Up-regulated	Significant
RAD51	Up-regulated	Up-regulated	Very Significant

TABLE 6.6: **Five genes that are significantly up-regulated compared to controls.** The ovarian cancer panel revealed 5 of the 79 genes to be significantly up-regulated. The results mirror the Oncomine data. Kaplan Meier plotter shows significant (*p-value 0.05-0.0001*) to very significant (*p-value<0.0001*) association with overall survival.

This data not only validates the expression pattern of the 5 identified genes from the 79 cancer panel, but shows that expression levels in blood mirror the ones observed in tissue. Furthermore, the expression pattern of these 5 genes has been shown to be associated with a poorer prognosis and overall survival. These results are detailed in Figs 6.2 to 6.6.

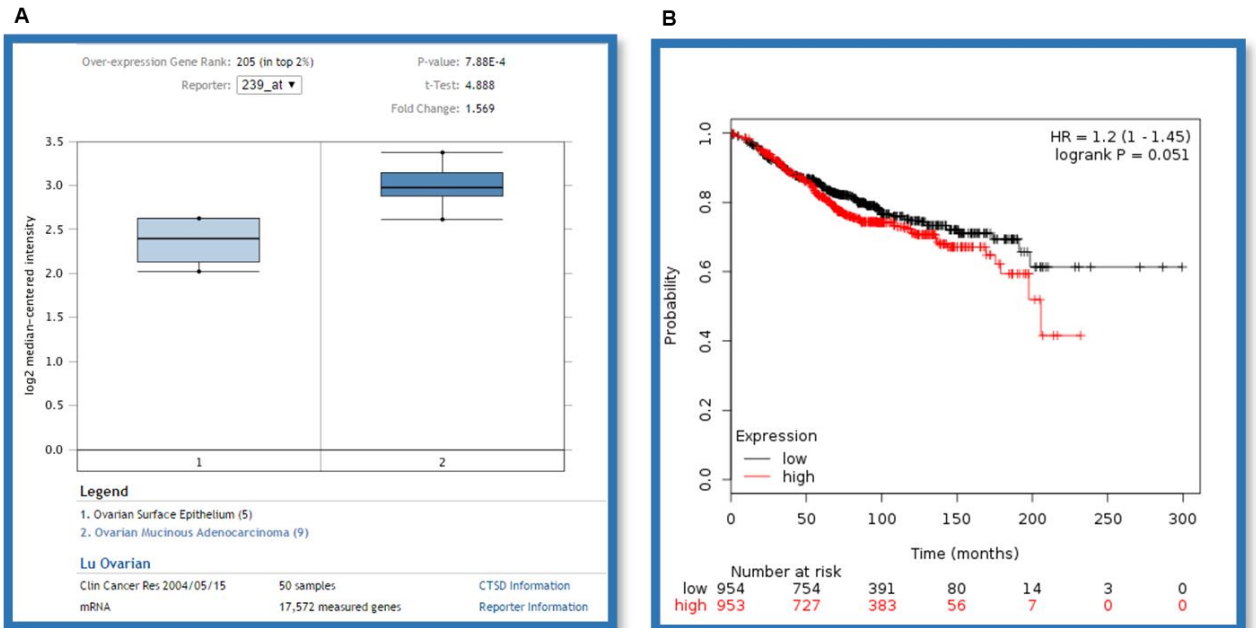


FIGURE 6.2: **Oncomine analysis for the CTSD gene.** Results in (A) Lu ovarian dataset (1, n=5; 2, ovarian mucinous adenocarcinoma, n=9), show up-regulation of CTSD in the cancer cohort compared to controls, which was shown to be statistically significant. (B) Overall survival plotted as a Kaplan Meier, shows poorer overall survival in the CTSD high expression group, also shown to be significant.

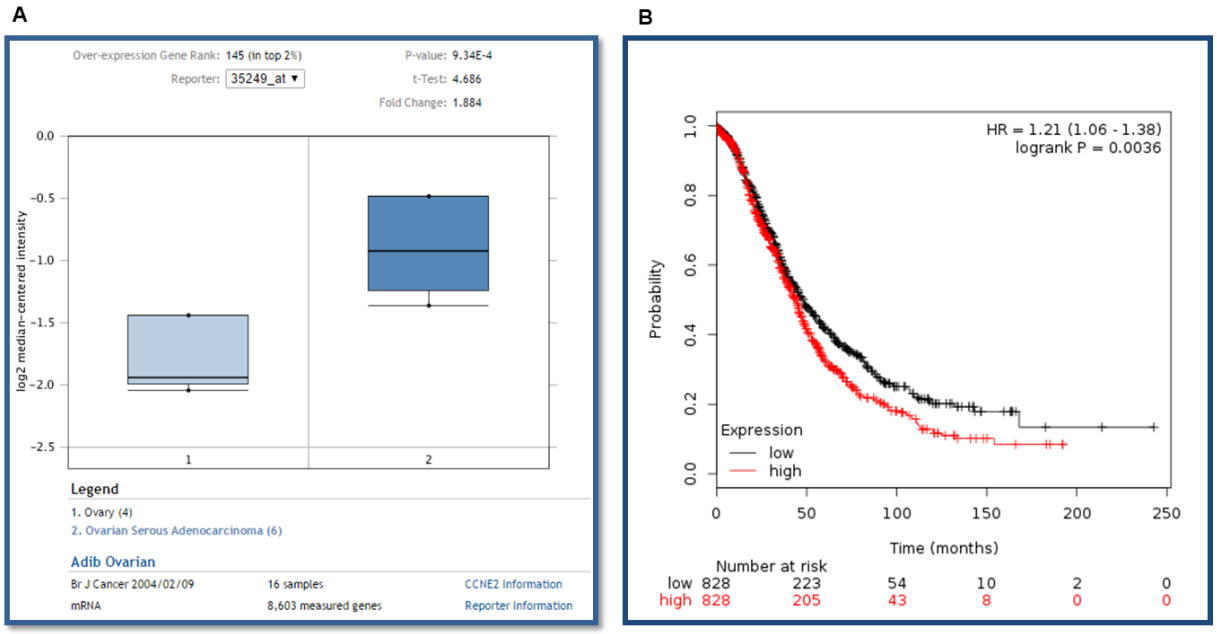


FIGURE 6.3: **Oncomine analysis for the CCNE2 gene.** Results in (A) Adib ovarian dataset (1, n=4; ovarian serous adenocarcinoma, n=6), show an up-regulation in gene expression of CCNE2 in the cancer cohort compared to controls, that is statistically significant. (B) Overall survival plotted as a Kaplan Meier, shows poorer overall survival in the CCNE2 high expression group, this was shown to be significant.

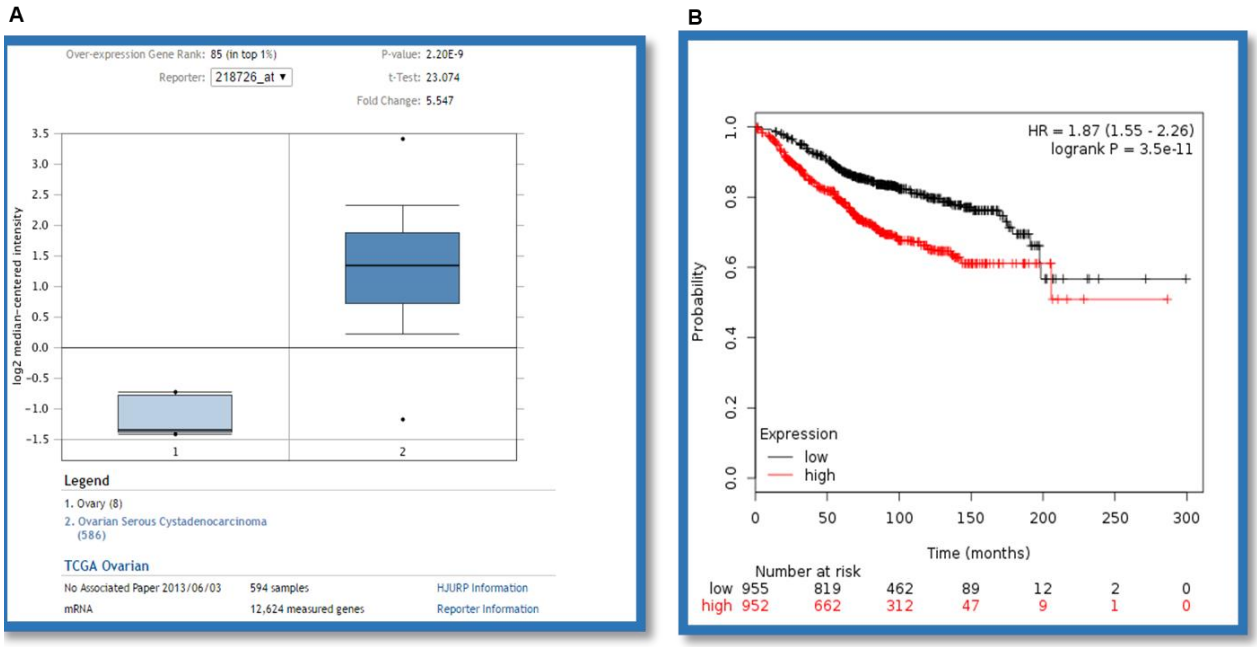


FIGURE 6.4: **Oncomine analysisfor the HJURP gene.** Results in (A) TGCA ovarian dataset (0, n=8; 1, ovarian cystoadenocarcinoma, n=586), show that HJURP expression is up-regulated in the cancer cohort compared to controls, statistically significance was reached. (B) Overall survival plotted as a Kaplan Meier, shows poorer overall survival in the HJURP high expression group, this was also shown to be significant.

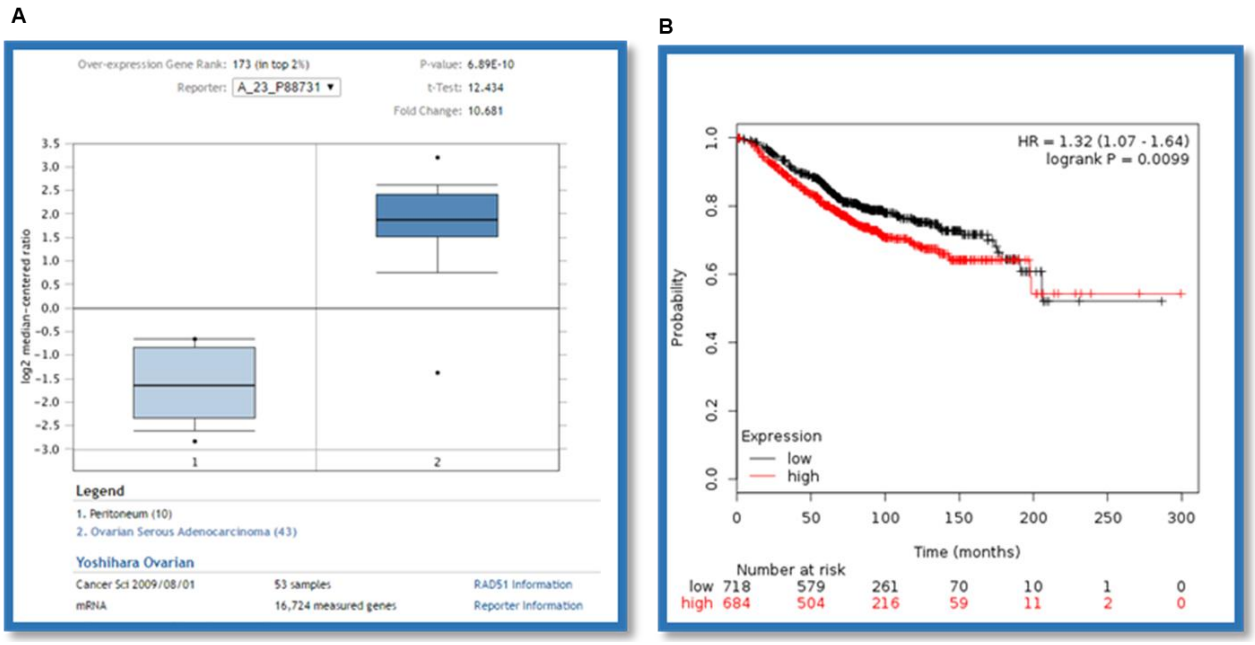


FIGURE 6.5: **Oncomine analysis for the RAD51 gene.** Results in (A) Yoshihara ovarian dataset (0, n=10; 1, ovarian adenocarcinoma, n=43), show that RAD51 expression is up-regulated in ovarian cancers compared to controls, statistical significance was reached. (B) Overall survival plotted as a Kaplan Meier, shows poorer overall survival in RAD51 over-expressed patients, again this was shown to be significant.

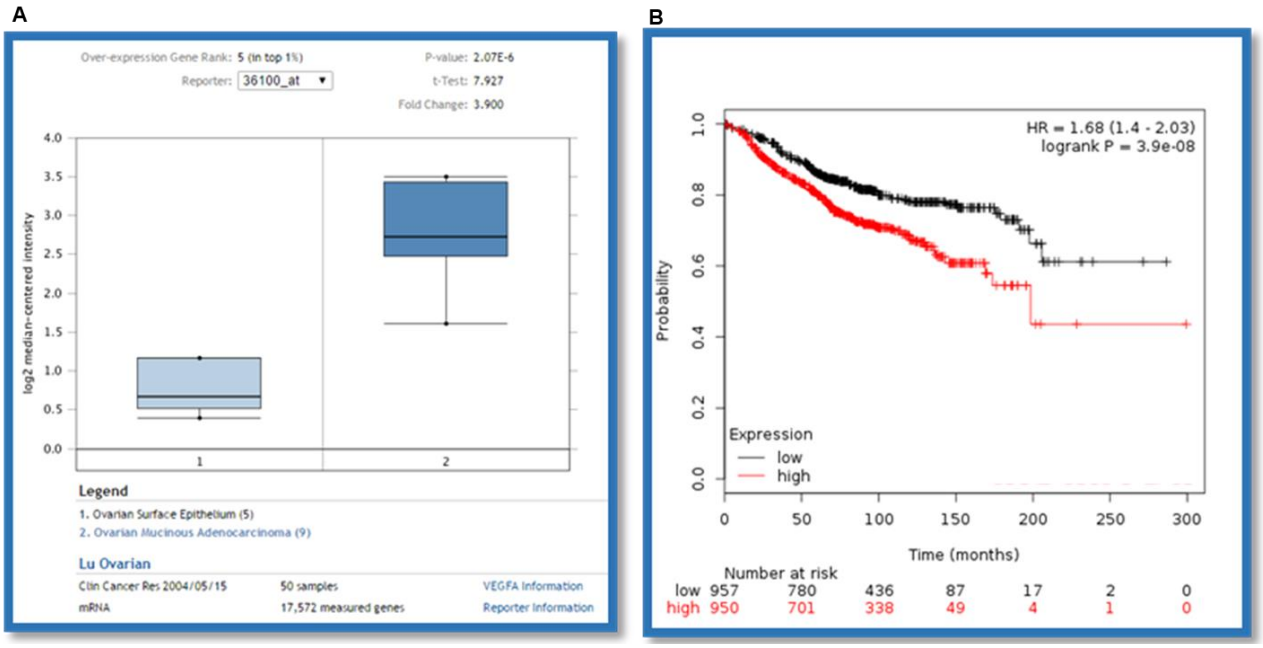


FIGURE 6.6: **Oncomine analysis for the VEGFA gene.** Results in (A) Lu ovarian dataset (0, n=5; 1, ovarian cystoadenocarcinoma, n=586), show an up-regulation in gene expression of VEGFA in the cancer cohort compared to controls, statistical significance was not reached. (B) Overall survival plotted as a Kaplan Meier, shows poorer overall survival in the VEGFA over-expressed group, shown to be statistically significant.

***In silico* analysis of the 5 genes identified from the ovarian 79 cancer panel, against available microarray data on Oncomine also reveal an increased expression of these 5 genes in ovarian cancer tissue samples, interestingly a poorer overall survival can also be seen from the Kaplan Meier plots. These results highlight the importance of such a technique to identify unique and clinically relevant gene signatures.**

To further assess associations and gene expression patterns in the samples, a principle component analysis (see Fig 6.7) was conducted.

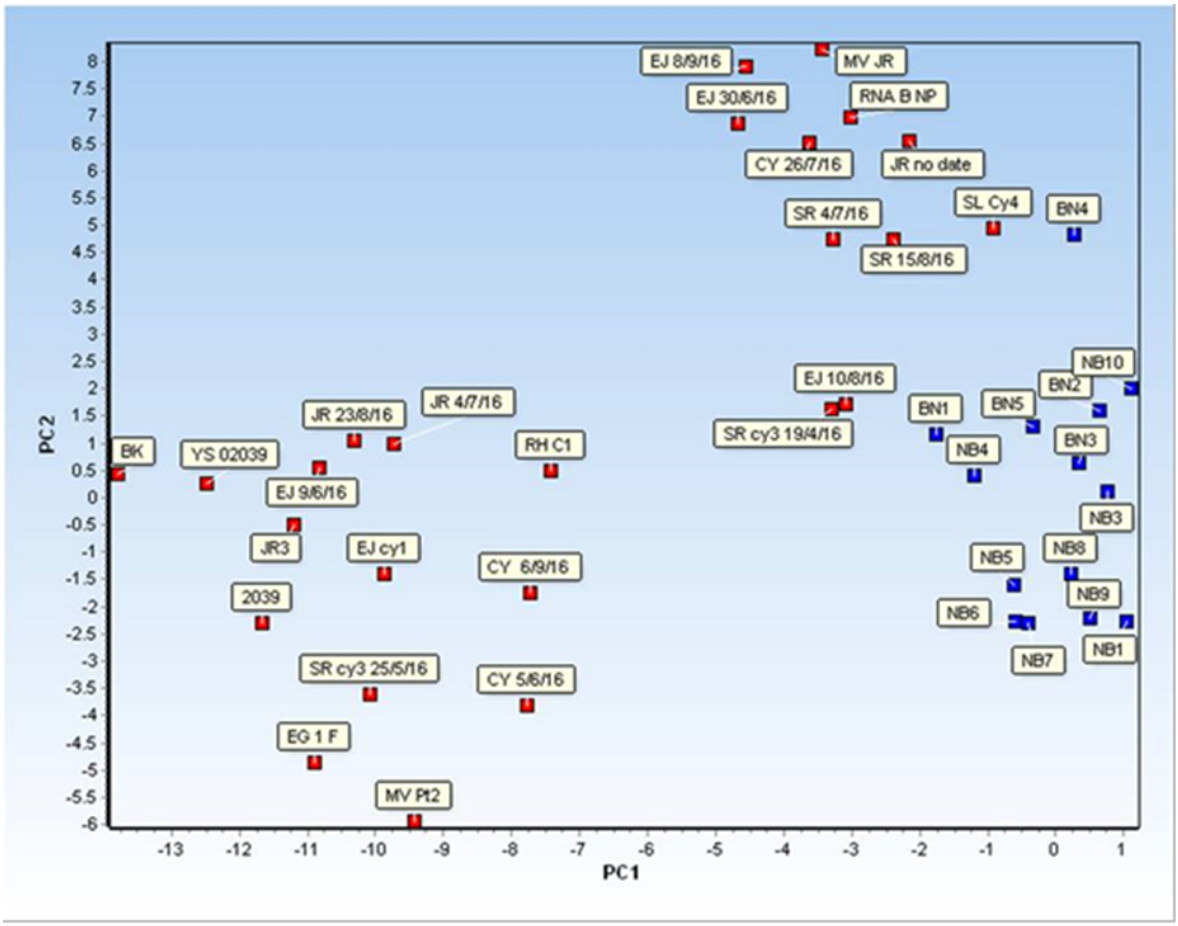


FIGURE 6.7: **Principle component analysis of all samples tested on the ovarian cancer gene panel.** PC1 vs PC2 scatter plot for the 79 analysed genes shows a good separation into three distinct gene clusters, 2 containing the cancer cohort (in red), and a final one containing the controls (in blue).

The principle component plot revealed 3 distinct populations of patients, sharing simi-lar gene expression profiles. As would be expected, the control group has formed one cluster. Interestingly the cancer cohort formed 2 clusters, suggesting 2 different expression profiles being shared amongst the cancer samples, 1 control sample was also

found to be present in a cancer cluster. Further analysis on a patient level reveals no trends, patients on the chemotherapy trial appear to move from one cancer cluster to another and back. This could be explained by changes in the disease state and stage influencing the genetic profiles, however there is no correlation to outcome. These results could be explained by the very nature of the chemotherapy treatment, which could alter the cancer, along with influences from inter-patient variability. This data provides strong evidence of common expression profiles shared in this cancer cohort, highlighting the value of cancer gene expression panel analysis to identify unique gene signatures.

6.4.2 Cell-free Copy Number Instability in lung cancer patients

Next we looked at the cell free Copy Number Instability (CNI). We collaborated with Chronix Biomedical in Germany to analyse the CNI profile on 3 lung cancer (blood, tumour tissue, and normal adjacent tissue) and 3 control samples from patients undergoing surgery for other respiratory diseases (blood and lung tissue). Two ovarian samples were also run, however further samples could not be run due to cost constraints. CNI analyses was conducted on DNA extracted from the tissue (tissue analysis) or plasma (blood/liquid biomarker analysis). Sample collection and DNA extraction was done prior to sending for CNI processing and sequencing to the company, as described in these studies (Beck et al., 2013; Beck et al., 2010).

Samples are assigned a CNI 171 score. A copy number instability (CNI) score is calculated from the number of copies of a particular genetic region, which varies from one individual's sample compared to a previous sample or to a normal database. The CNI score can be interpreted as a general measure of genomic instability and is directly related (within the technology limits) to the regional chromosomal DNA ploidy. Data is then plotted in a circos plot. Circos plots are images for visualising data and information. It visualizes data in a circular layout enabling exploration of relationships of genomic positions. In these reports, a red dot symbolizes where a genomic region has a gain of copy number of that region and a purple dot symbolized where a genomic region has a loss of copy number of the region. Z-values per evaluated region are used for display. Each sample has a Z value calculated, A Z-value is a statistical measurement of a value's relationship to the average in a group unaffected individuals in multiples of the standard deviation. A Z-value of 0 means the value is the same as the average, 1 means the value is one standard deviation above the average, a Z-value can also be positive or negative, indicating whether it is above or below the average.

Patient CNI scoring can be seen in Table 6.7, with CNI circus plots in Fig 6.8.

Patient Sample	CNI 171 score	Sample type
MAS4	524	Cancer
MAS12	24	Cancer
MAS23	74	Cancer
N1	0	Control
N2	3	Control
N3	54	Control

TABLE 6.7: **CNI 171 score for all lung cancer patient and control samples.** All cancer and control samples were processed and a CNI171 score generated. A normal CNI 171 score would be seen as 0 to 31, values above this would indicate disease, with a higher CNI 171 scores indicating increased chromosomal instability and disease. All cancer patients are seen to have a higher than normal CNI 171 score, except MAS 12. All normal patients show a CNI 171 score within the normal range, with N3 being the exception.

Read counts per bin were normalised to the median read counts over all bins. Ratios are displayed as $\log_2$ values. $\log_2$ values > 0.15 (gain) or < -0.15 (loss) are displayed as red or purple dots, respectively.

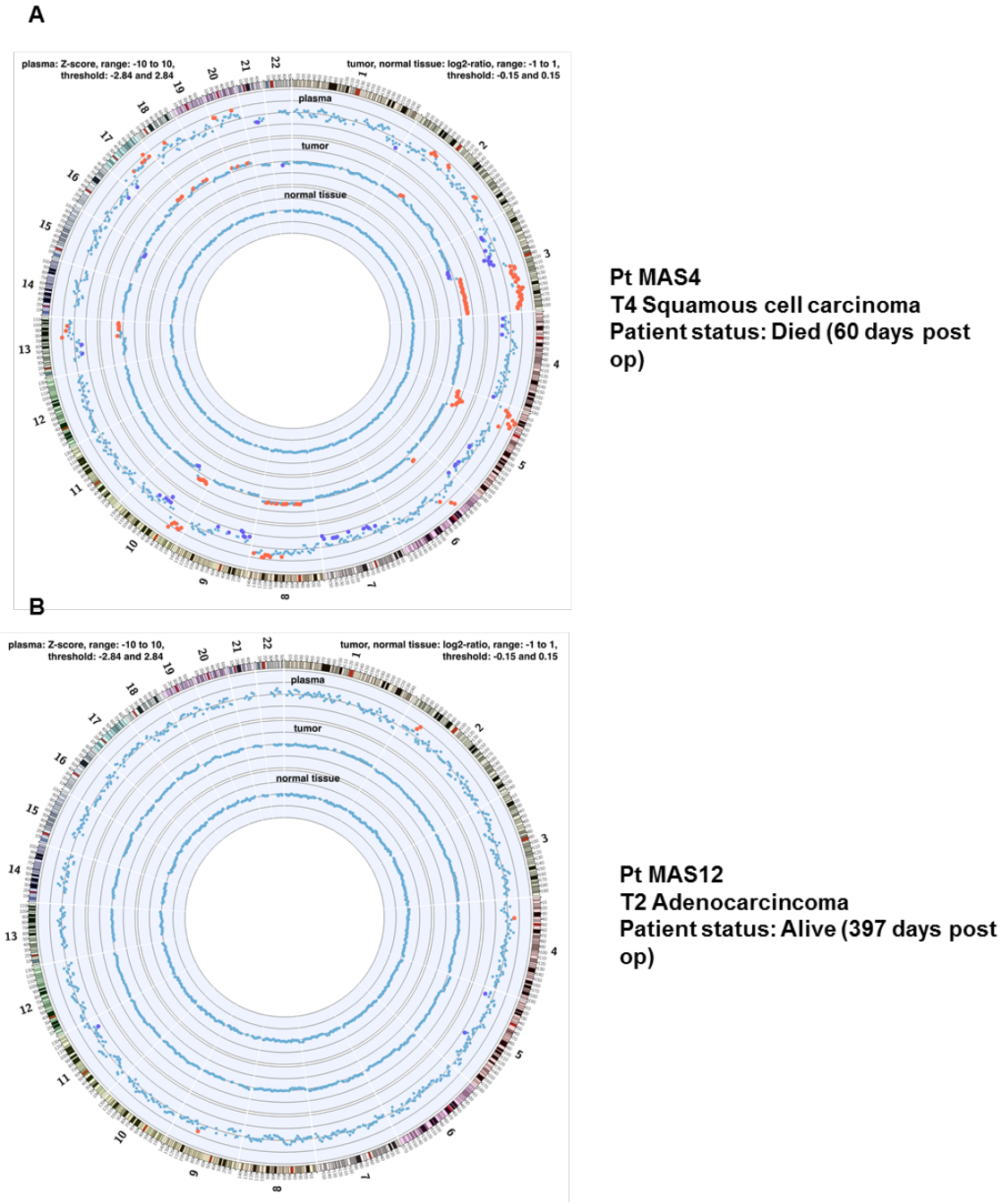


FIGURE 6.8: **Circos plots CNI for samples MAS4 and MAS12.** Circos plots were generated for 2 of the lung cancer cases, highlighting the chromosomal copy number instability and corresponding chromosome and location (A) MAS4 shows extensive copy number changes, visualised by the red and purple dots (loss and gains based on values that are significantly different to normal). Interestingly the copy number changes are identical and occur in the same regions in the tumour and plasma sample. (B) MAS12 shows very little chromosomal copy number changes, and these can only be seen in the plasma.

The results generally show that cancer patients display a CNI 171 score above the normal range, with one exception of a cancer patient (MAS12), with a CNI 171 score within the normal range. This could be due to patient variability, or the stage and nature of the disease. This particular patient's cancer is known to be diagnosed as a stage II adenocarcinoma, which would be classed as early stage. The patient may have a stable slow growing tumour or be in remission. Correspondingly all normal patients displayed a CNI 171 score within the normal range, with one exception of a normal patient (N3) with a CNI 171 score above the normal range; again this could be due to other underlying or undiagnosed conditions the patient may have.

Similar analysis was carried out on 2 ovarian cancer patients. Table 6.8 lists the CNI 171 scores for all the ovarian patient samples, while circos plots are depicted in Fig 6.9.

Patient	CNI 171 score	Sample type
Ov1	984	Cancer
Ov2	18	Cancer

TABLE 6.8: **CNI 171 score for all ovarian patient samples.** All ovarian cancer samples were processed and a CNI 171 score generated as with the lung samples. A normal CNI 171 score would be seen as 0 to 31, values above this would indicate disease. Patient Ov1 scored 984, much higher than the normal range, whereas patient Ov2 scored 18 which is within the normal range

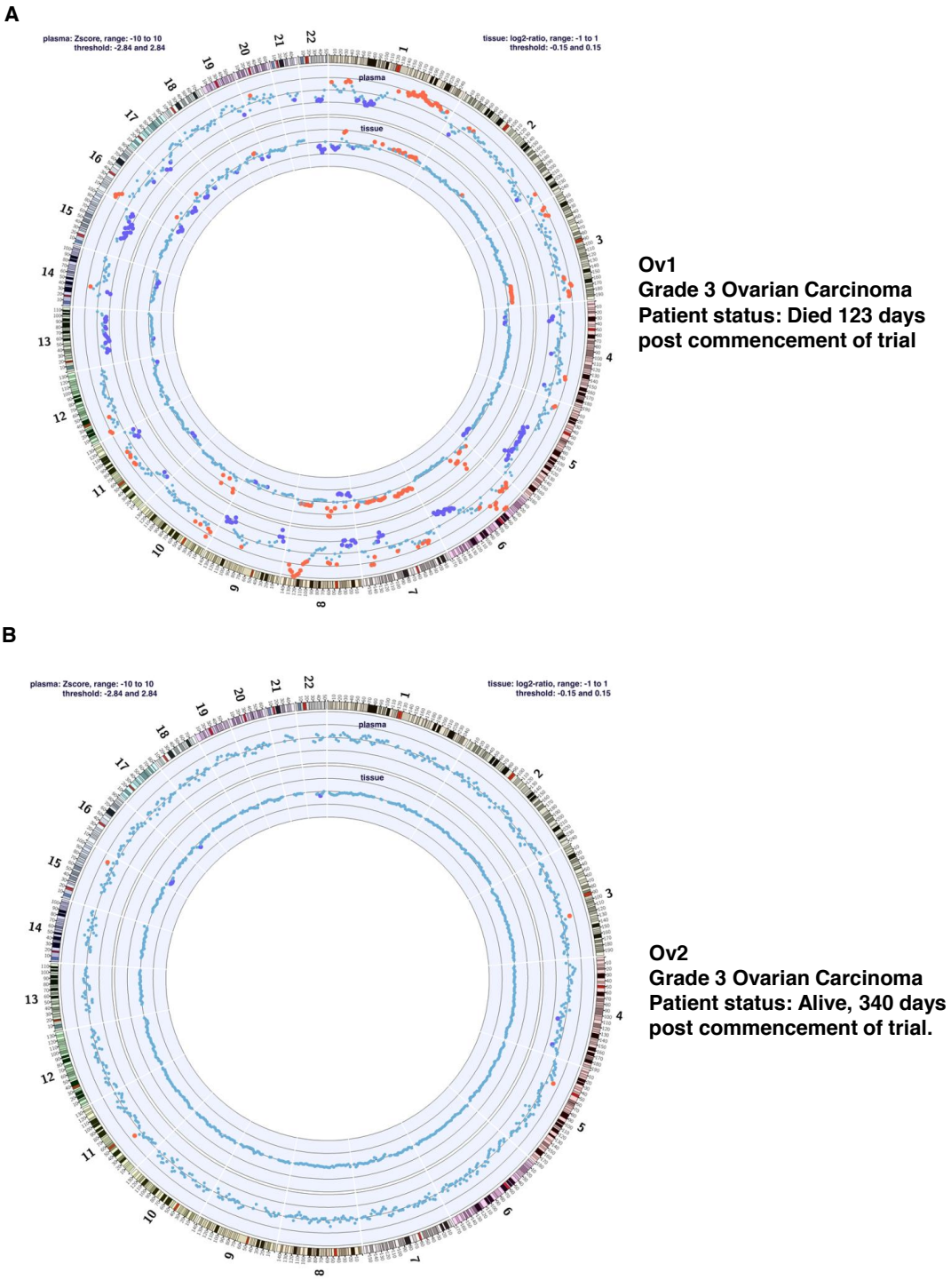


FIGURE 6.9: **Circos plots CNI for samples Ov1 & Ov2.** A circus plot was generated, highlighting the chromosomal copy number instabilities identified, in relation to corresponding chromosome and location. (A) Ov1 shows extensive copy number changes, visualised by the red and purple dots (loss and gains based on values that are significantly different to normal). The copy number changes are identical and occur in the same regions in both tumour and plasma. (B) Ov2 shows no real chromosomal copy number instability.

When these results above are compared to overall survival, it is apparent that a higher CNI 171 score correlates with a poorer prognosis, a similar trend is also seen for the lung cancer patients. Interestingly, although patient Ov2 scored within the normal ranges, it is important to note that this patient has gone into remission and has subsequently been discharged, thus suggesting a strong predictive and prognostic role for CNI scoring.

6.4.3 Transcriptomics analysis

RNA sequencing was carried out by our collaborators at the High-Throughput Genomics, Wellcome Trust Centre for Human Genetics, Oxford. RNAseq was run on 3 lung cancer matched tissue and blood samples, with 3 matched tissue and blood samples from controls. All data was analysed as described in Chapter 2.11.1.

A scatterplot was formulated to visualise the general distribution of gene expression levels in the studied samples for quality assurance purpose. The blood and tissue samples were compared amongst themselves, respectively (Fig 6.10 and 6.11).

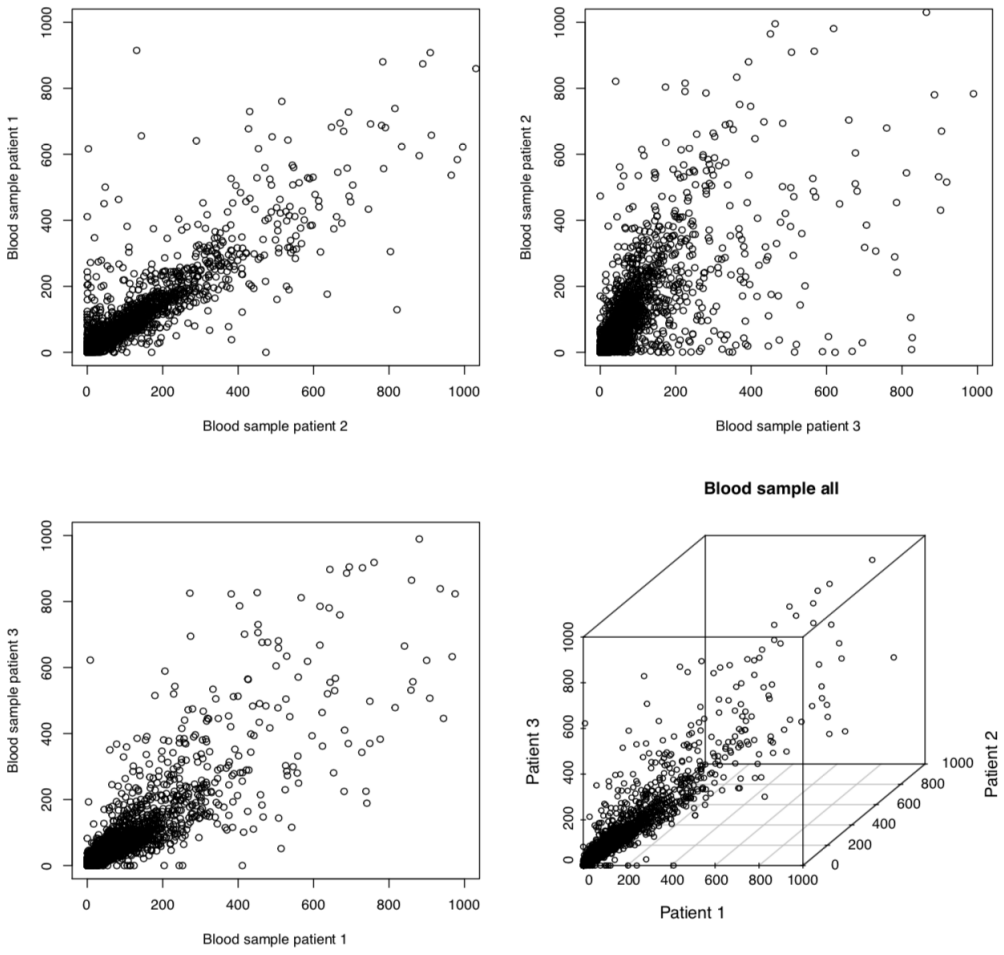


FIGURE 6.10: **Comparison of gene expression values between all patient blood samples.** Pairwise scatter plots show the expression values (as measured by the FPKM value) for each gene (represented by a circle). If the value is the same in both samples the gene would be found on the diagonal. Any difference between the gene expression in the two sample will appear of diagonal. The scatterplots show a homogenous distribution of gene expression for all blood samples tested.

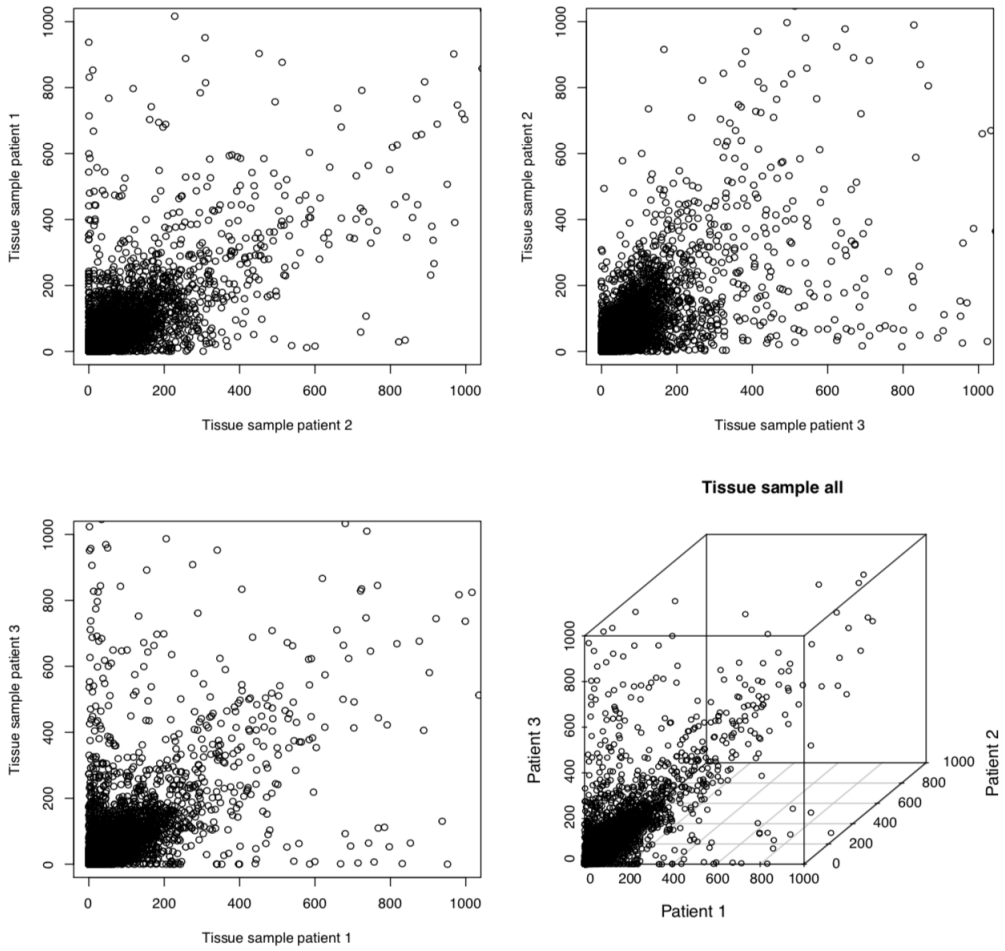


FIGURE 6.11: **Comparison of gene expression values between all patient tissue samples.** Pairwise scatter plots show the expression values (as measured by the FPKM value) for each gene (represented by a circle). If the value is the same in both samples the gene would be found on the diagonal. Any difference between the gene expression in the two sample will appear of diagonal. The scatterplot shows a heterogenous distribution of gene expression of all the tissue samples tested, particularly in comparison to the blod sample as seen in the previous scatterplot.

As expected, the gene expression levels in blood show a higher level of homogeneity than in tissue. This result comes in accord with previous observations on the effect of tissue cell heterogeneity on gene expression (Zhao and Simon, 2010).

Next, we compared the landscape of gene expression levels in the cancer tissue and blood samples vs normal (Fig 6.12 & 6.13) as well cancer tissue vs cancer blood for each of the three patients (Fig 6.14). The data was plotted as a volcano plot. We distinguished

three types of gene groups. First, there are genes (shown in red) that show a statistical significant differential transcription pattern characterised by an FPKM (Fragments Per Kilobase of transcript per Million mapped reads) value >5, the absolute value of $\log_2$FoldChange between the two analysed conditions >2, and *p-value*$<10^{-4}$. Second, there are genes (seen in green) with FPKM>5, absolute value of $\log_2$FoldChange>2 and with a *p-value*<0.05. Finally, there are genes that do not show any significant change in the transcription levels between the analysed states (shown here in grey).

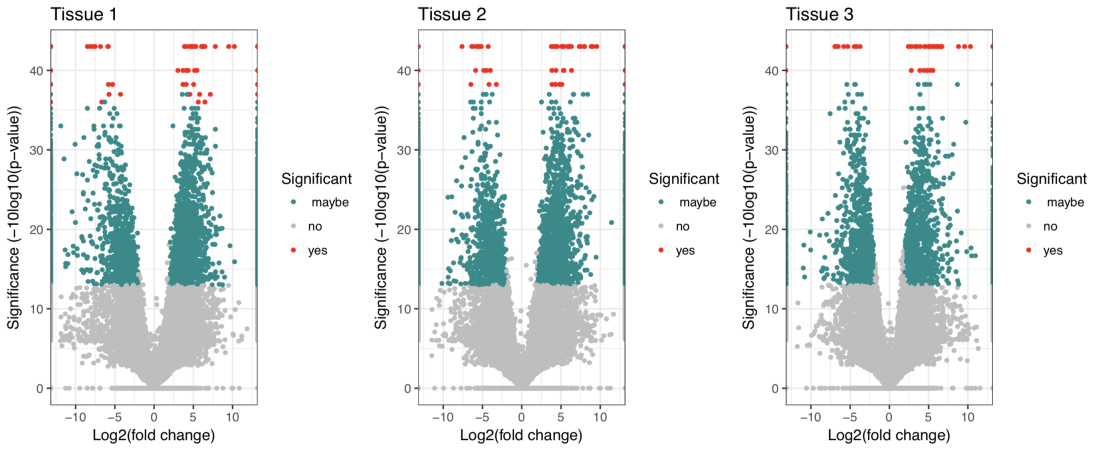


FIGURE 6.12: **Volcano plot of all 3 cancer tissue samples compared to control tissues.** Data shows mostly up-regulation or down-regulation of genes in the cancer samples in comparison to controls (left of centre = down-regulation, right of centre = up-regulation), with a handful of genes shown to be very significantly up or down-regulated (red, *p-value*<0.0001).

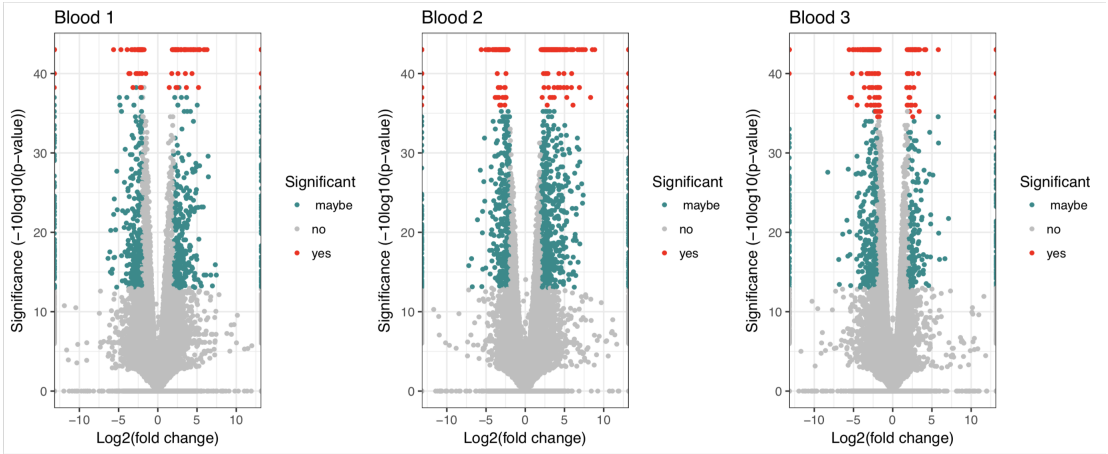


FIGURE 6.13: **Volcano plot of all 3 cancer blood samples compared to blood samples.** Data shows mostly up-regulation and down-regulation of genes in the cancer samples in comparison to controls, with a handful of genes shown to be very significantly up or down-regulated (red, *p-value<0.0001*).

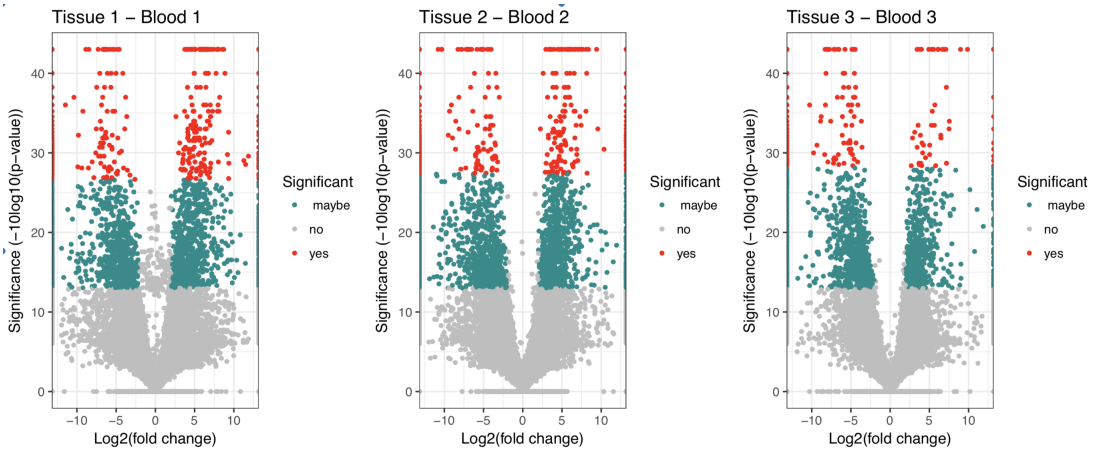


FIGURE 6.14: **Volcano plot comparing all 3 cancer patient tissue samples to respective blood samples.** Volcano plots were also constructed to look at the differences in gene expression between parallel tissue and blood samples taken from the same patient. The data shows a huge variation in gene expression between the 2 sample types in the same patient. (Tumour 1 vs Blood 1) shows a large proportion of samples with the same expression pattern seen in both the tumour tissue and blood, seen in the middle region, although these are not seen to be significant.

Overall, the gene expression data shows that a large number of genes are up or down-regulated in the tumour samples compared to control. In particular, there are 356 differentially expressed genes in patients' blood compared to controls, and 293 differentially expressed genes in the tumour tissue samples compared to controls. Moreover, we observed a large difference in the gene expression profiles when we compared the tumour tissue and blood samples in individual patients. This suggests that the tumour tissue cancer genetics are not necessarily mirrored in the blood, even though we see differential gene expression in both the blood and tissue of cancer patients compared to controls. This could again be attributed to inter-patient variability. Evaluation of these genes using GO term analysis showed these genes are mostly immune related.

Next, we evaluated in more detail the genes that show a statistical significant differential expression (red dots in volcano plots) in the three patients lung tissue compared to controls (see Table 6.9). Patient 1 is a stage IV squamous cell carcinoma, whilst patients 2 & 3 share the same pathology of T1B adenocarcinoma.

	Patient 1	**Patient 2**	**Patient 3**
Total genes	149	116	80
Up-regulated	94	54	34
Down-regulated	55	62	46
Common up-regulated (All)		3	
Common up-regulated (Pt 2&3)		16	
Common down-regulated (All)		3	
Common down-regulated (Pt 2&3)		29	

TABLE 6.9: **Differentially expressed genes within the three patient tissue samples.** Greyed areas indicate the number of common differentially expressed genes across multiple samples (e.g. across all patients or common only to patients sharing the same disease stage and type as is the case with patients 2 & 3.

Overall the results are heterogeneous, with patient 1 (most advanced disease, squamous cell carcinoma) having the largest number of differentially expressed genes in the tumour tissue. Moreover, patient 1 also shows a higher number of genes being over-expressed compared to down-regulated vs controls. By contrast, patients 2 and 3 show a similar ratio of up- to down-regulated genes compared to control. Interestingly, when looking at genes differentially expressed across all three patients, only 3 are commonly up-regulated across all 3 patients, and the same number down-regulated (within patient 2 and 3 analysis only this becomes 16 and 29, respectively). In particular, MUC4 a proto-oncogene, and MIR663A a tumor progression associated gene, have been shown to be over-expressed, while the RUNX1T1, tumour supressor gene, as expected was down-regulated in all three patients.

GO term analysis reveals that most of these genes are commonly associated with immune responses. A summary of these genes can be seen in Tables 6.10 and 6.11.

Up-regulated		**Down-regulated**	
Gene	**Biological Function**	**Gene**	**Biological Function**
MUC4	Proto oncogene	AGER	Immunoglobulin family
RIMS2	Presynaptic protein	RUNX1T1	Tumour suppressor
MIR663A	Tumor progression	SFTPA1	Immune response

TABLE 6.10: **List of genes commonly expressed in the tumour tissue across all three patients.** Table listing details on biological function for genes showing similar expression across all 3 patients.

Up-regulated		**Down-regulated**	
Gene	**Biological Function**	**Gene**	**Biological Function**
BRDT	Transcription regulation	AGER	Immunoglobulin family
CALML5	Calcium binding protein	RUNX1T1	Tumour supressor
DPF1	double PHD fingers 1	SFTPA1	Immune response
GSTT1	Glutathione S-transferase theta 1	ADRA1A	Cells growth and ploripheration
KCNJ10	Potassium channel protein	CD300LG	Immune escape of lung cancer cells
LINC00511	Long noncoding RNA	DDX3Y	Y-linked DEAD-box helicase
LINC00668	Long noncoding RNA	EIF1AY	Eukaryotic translation initiation factor 1A
MIR663A	Tumor progression	FAM110D	Protein FAM110D
MSMB	Microseminoprotein beta	FAM153A	Renal carcinoma antigen
MUC4	Proto oncogene	FAM153B	Protein FAM153B
RIMS2	Presynaptic protein	FGFR4	Growth factor
RP11-169F17.1	Ribosomal protein	FHL1	Ion channel binding
TMPRSS11E	Serine protease	GPM6A	Calcium ion transport
XAGE1B	Tumor associated	GRIA1	Glutamate receptor
XIST	Pseudogene related to lung & testicular cancer	MYO15B	Myosin
ZIC2	Zing finger protein	NOSTRIN	Endocytosis
		PRKY	Kinase pseudogene
		RPS4Y1	Ribosomal protein
		SEMA3B	Tumour supressor
		SFTPA2	Immune response
		SFTPB	Pulmonary-associated surfactant protein
		SFTPC	Pulmonary-associated surfactant protein
		SORBS2	Muscle cell development
		TPSD1	Serine-type peptidase
		TTTY15	Testis-specific transcrip
		TXLNG2P	Pseudogene
		USP9Y	Ubiquitin specific peptidase
		VIPR1	Signal transduction
		ZFY	Organism growth

TABLE 6.11: **List of genes similarly expressed in the tumour tissue across patient 2 and 3.** Details on biological function for genes with similar expression across 2 patients with the same cancer subtype and staging.

The increased number of genes shared between patients 2 & 3, in contrast to all 3 patients, suggests that similarities in gene profiles exist within similar cancer subtypes as would be expected. This can potentially allow for pathological subtyping and possibly

staging based on specific gene profiles.

We repeated the analysis workflow on the data from patient blood samples, and identified a number of genes that are commonly expressed across the patient blood samples (see Table 6.12).

	Patient 1	**Patient 2**	**Patient 3**
Total genes	121	188	190
Up regulated	61	77	110
Down-regulated	60	111	80
Common up-regulated (All)		7	
Common up-regulated (Pt 2&3)		38	
Common down-regulated (All)		8	
Common down-regulated (Pt 2&3)		48	

TABLE 6.12: **Number of differentially expressed genes within the three patient blood samples.** The number of differentially expressed genes within the three patient blood samples were also assessed. Greyed areas indicate the number of common differentially expressed genes across multiple samples (e.g. across all patients or common only to patients sharing the same disease stage and type as is the case with patients 2 & 3).

The blood samples show overall the same heterogeneity when comparing the three patient samples. Specifically, we observe a higher number of differentially expressed genes in patients 2 and 3 compared to patient 1. Further analysis reveals that these genes are not commonly shared between patients 2 & 3, however as previously seen with the tissue sample in Table 6.9, more genes are shared between these 2 patients with the same pathological cancer than across all three patients. This result provides further evidence that similar gene profiles are seen in the same disease subtypes. The variation in genes also amplifies the challenge of heterogeneity we know exists within cancers, also those of the same subtype. This also highlights issues seen with inter-patient variability.

Using GO term analysis, we found once more that these genes are related to the organism's immune response. A summary of these differentially expressed genes in blood is shown in Tables 6.13 and 6.14.

Up-regulated		Down-regulated	
Gene	**Biological Function**	**Gene**	**Biological Function**
KIAA1324	Survival in carcinomas	IGJ	Immune response
NBPF14	Neuroblastoma family; linked to cancer	RPS17	Ribosomal protein
IFITM3	Membrane protein; confers immunity to H1N1	CFD	Immune response
AC104135.3	Long noncoding RNA	IGLC2	Antigen binding
HLA-DRB6	Pseudogene	GCG	Negative regulation of apoptotic process
HLA-DQA2	Immune response	IGHG1	Immunoglobulin receptor binding
DEFA3	Host defence	REG1A	Cell regeneration
		TTR	Proteolysis, autophagy & glucose homeostasis

TABLE 6.13: **List of genes similarly expressed in blood across all patients.** Details on gene ID and biological function.

Up-regulated		Down-regulated	
Gene	**Biological Function**	**Gene**	**Biological Function**
AC104135.3	**Long noncoding RNA**	**CFD**	**Immune response**
DEFA3	**Host defense**	**GCG**	**Negative regulation of apoptotic process**
HLA-DQA2	**Immune response**	**IGHG1**	**Immune response**
HLA-DRB6	**Immune response**	**IGJ**	**Immune response**
IFITM3	**Immune response**	**IGLC2**	**Antigen binding**
KIAA1324	**Associated with survival in carcinoma**	**REG1A**	**Cell regeneration**
NBPF14	**Nueroblastoma family; alterations linked to cancer**	**RPS17**	**Ribosomal protein**
AHSP	**Hemoglobin assembly**	**TTR**	**Proteolysis, autophagy & glucose homeostasis**
ALAS2	**Heme synthesis pathway**	**BCORP1**	**Pseudogene**
BTNL3	**Butyrophilin-like**	**C4BPA**	**Immune response**
CD177	**Immune response**	**DDX3Y**	**DEAD-box helicase**
CLEC4D	**Immune response**	**EIF1AY**	**Translation initiation**
CYP27A1	**Bile synthesis pathway**	**FAM118A**	**Protein family 118A**
FAM21B	**Protein transport**	**HLA-DRB5**	**Immune response**

GSTT1	**Role in carcinogenesis**	**HLA-G**	**Immune response**
HBD	**Oxygen binding**	**IGHA1**	**Immune response**
IGHA1	**Immune response**	**IGHM**	**Immune response**
IGHA2	**Immune response**	**IGHV2-5**	**Immune response**
IGHG2	**Immune response**	**IGHV3-15**	**Immune response**
ITGA2B	**Platelet coagulation**	**IGHV3-23**	**Immune response**
LINC00958	**Long noncoding RNA**	**IGHV5-51**	**Immune response**
MEG3	**Long noncoding RNA**	**IGKC**	**Immune response**
MSRB2	**Protein repair**	**IGKV1-5**	**Immune response**
RAMP3	**Coreceptor activity**	**IGKV1D16**	**Immune response**
RP11203B9.4	**Ribosomal protein**	**IGKV3-20**	**Immune response**
RP11661A12.5	**Ribosomal protein**	**IGLC1**	**Immune response**
RPH3A	**Neurotransmitter release**	**IGLV3-1**	**Immune response**
S100A12	**Calcium binding**	**IGLV3-10**	**Immune response**
S100P	**Calcium binding**	**IGLV3-21**	**Immune response**
SELENBP1	**Selenium bindin**	**IGLV3-27**	**Immune response**
SERPING1	**Endopeptidase inhibitor**	**KDM5D**	**T-cell antigen processing and presentation**
SLC37A3	**Transmembrane transport**	**KRT77**	**Sturctural molecule**
THBS1	**Platelet aggregation, angiogenesis, tumorigenesis**	**MRGPRE**	**Signal transduction**
TMTC1	**RNA processing**	**MS4A1**	**B-cells development**
TUBBP5	**Pseudogene**	**PRKY**	**Kinase pseudogene**
VNN1	**Immune response**	**PRSS1**	**Serine protease**
XIST	**Pseudogene related to lung & testicular cancer**	**RP11-424G14.1**	**Ribosomal protein**
ZFP57	**Zing finger protein**	**RPS4Y1**	**Ribosomal protein**
		SHISA4	**Membrane component**
		SLC12A7	**Ion transport**
		SORBS2	**Muscle cell development**
		SOX2-OT	**Long noncoding RNA**
		SPON2	**Defense response**
		TMSB4Y	**Acting sequestrin protein**
		TTTY15	**Noncoding RNA**
		TXLNG2P	**Pseudogene**
		USP9Y	**Ubiquitin specific peptidase**

	ZFY	**Organism growth**

TABLE 6.14: **List of genes similarly expressed in blood across patients 2 and 3.** Details on gene ID and biological function.

Next, we analysed genes expression signatures in tumour tissue and blood compared with control (Fig 6.15) we identified 21 genes that are differentially expressed in both tissue and blood.

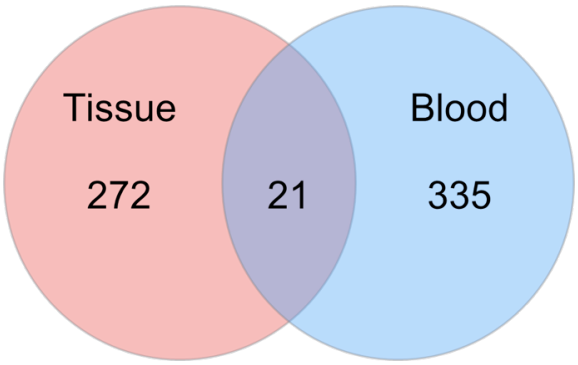


FIGURE 6.15: **Summary of differential expressed genes in blood and tissue.** A total of 272 genes were shown to be differentially expressed in tumour tissue compared to controls, and 335 in cancer blood samples compared to controls. 21 genes show statistical significant differential expression pattern in both tissue and blood samples compared to control (see Table 6.15).

Genes		
TTTY15	DDX3Y	DMKN
EIF1AY	KDM5D	C10orf9
TXLNGY	ELN	KRT77
RPS4Y1	SLC6A8	SPD1C
USP9Y	FRCLA	IGHV4-31
PRKY	ME132	XIST
ZFY	SORB5	GSTT1

TABLE 6.15: **Differentially expressed genes, across blood and tissue samples.** 21 genes were shown to have significantly different expression patterns in blood and tissue samples.

We observed that while the majority of genes are down-regulated in patient samples compared to controls, there was only 1 gene that was over-expressed in both patient tissue and blood, with no or very little expression in both blood and tissue of control patients, X Inactive Specific Transcript (XIST). All results were shown to be highly statistically significant (*p-value<0.0001*).

To expand our biomarker search horizon, we looked next at genes that are overexpressed in patient blood compared to control. Out of the total number of differentially expressed genes in blood (335+21) we selected 18 genes that have the following characteristics: have an expression level higher than 5 FPKM (minimal expression threshold for

calling a protein coding gene transcribed (Harrow et al., 2012)), show a statistical significant differential expression, and the expression level in the tumour blood is at least 2 fold higher than in control. The 18 genes are summarised in the Table 6.16.

Genes					
BTNL3	DEFA3	HLA-DQA2	KIAA1324	RPH3A	TMTC1
CD177	ECHDC3	HLA-DRB6	MSRB2	SERPING1	VNN1
CEBPE	GSTT1	IGHG2	NBPF14	THBS1	XIST

TABLE 6.16: **Over-expressed genes in lung cancer blood.** A total of 18 genes were shown to have significantly higher expression levels in the blood of lung cancer patients compared to controls.

We functionally characterised the differentially expressed genes, 335 (+21) & 272 (+21), in blood and tissue, respectively. We also characterised the 21 genes significantly differentially expressed in both blood and tissue, and the 18 genes known to be up-regulated in blood (Fig 6.16) using FunRich software.

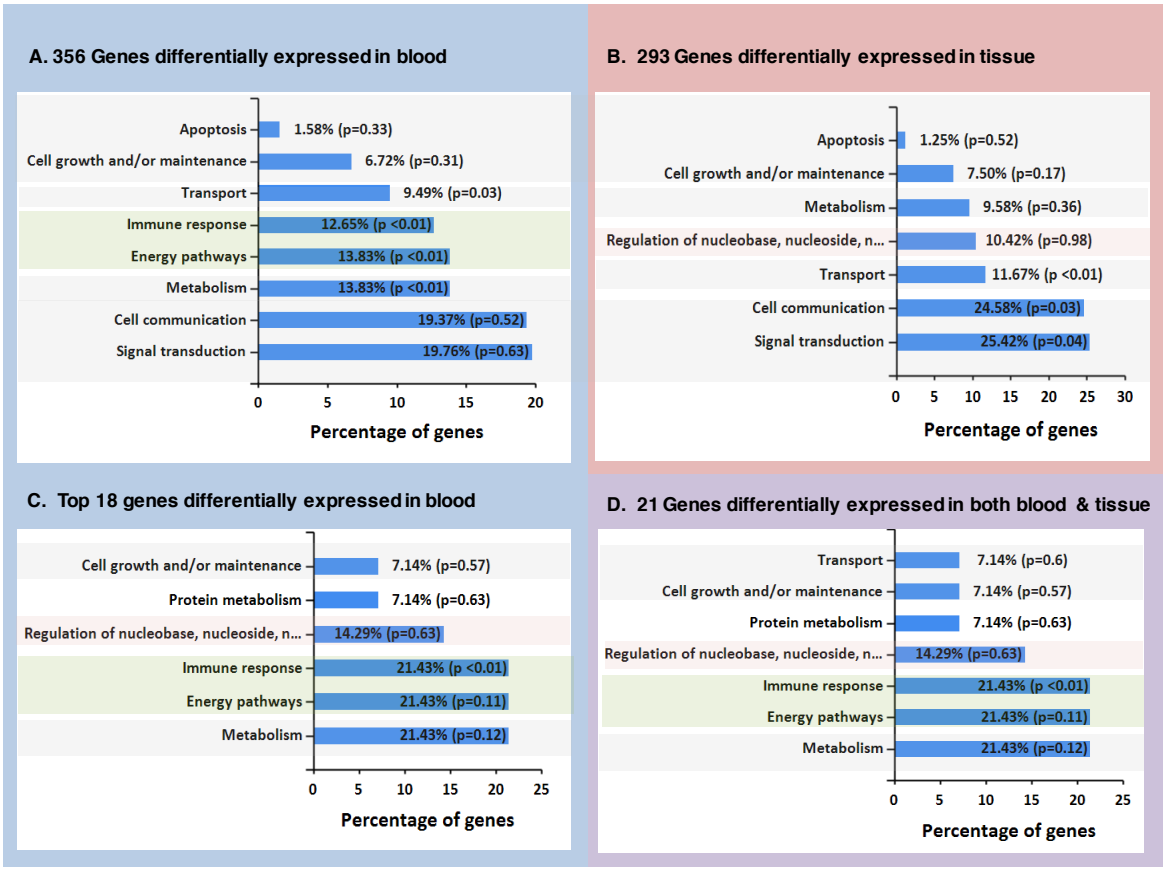


FIGURE 6.16: **Functional enrichment analysis.** Details on top GO terms resulting form the functional enrichment analysis using Fun Rich software, for differentially expressed genes in blood (356), tumour tissue (293), blood only (18), and in both blood and tumour (21) compared to controls.

We observed that overall, in both tumour tissue and blood, cell communication and signal transduction are the most enriched pathways. For tissue samples, immune response and energy pathways are top gene ontology (GO) terms for differentially expressed genes in blood, while regulation of the nucleobase and nucleosides is the top tumour tissue specific GO term. The set of enriched GO terms for the 21 differentially expressed genes in both tissue and blood is composed of a mixture of both blood and tumour tissue relevant terms, with metabolism, energy pathways, immune response, and regulation of nucleoside being the top terms.

The functional analysis using FunRich software revealed immune response, energy pathway and metabolism to be the most common functions of these 18 genes (see Fig 6.16). Moreover, it is interesting to point out that the top GO terms enriched in the 18 genes match the top enriched terms in the set of 21 genes that are differentially expressed in both tissue and blood, highlighting the fact that the functions of the top blood differentially expressed genes match at certain levels, the ones observed in tissue.

To verify the results from the RNAseq analysis, validation experiments were carried out on the 4 statistically significant over-expressed genes in blood from the 18 identified. These include, XIST, GSTT1, THBS1 and NBPF14. These were tested in 5 patient blood samples (cancer and controls). Validation was conducted by qPCR, (Figs 6.17 to 6.20). Samples were taken from cancer and control patients not included in the RNAseq experimentations. Samples consisted of 3 cancer patients, and 2 controls. For each of the figures, (A) represents the qPCR validation experiment, and (B) the RNAseq generated data.

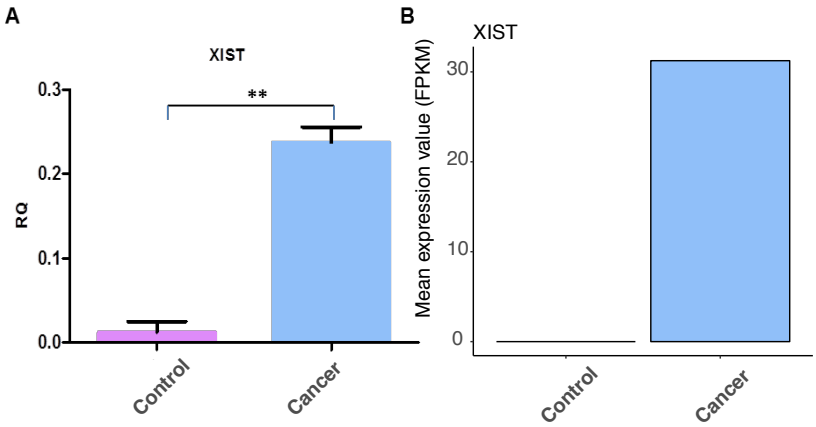


FIGURE 6.17: **qPCR and RNAseq expression of XIST.** Validation using qPCR, shows XIST expression in blood samples for control and cancer patients. Validation data shows the same up-regulation of XIST expression in patient blood samples compared to controls.

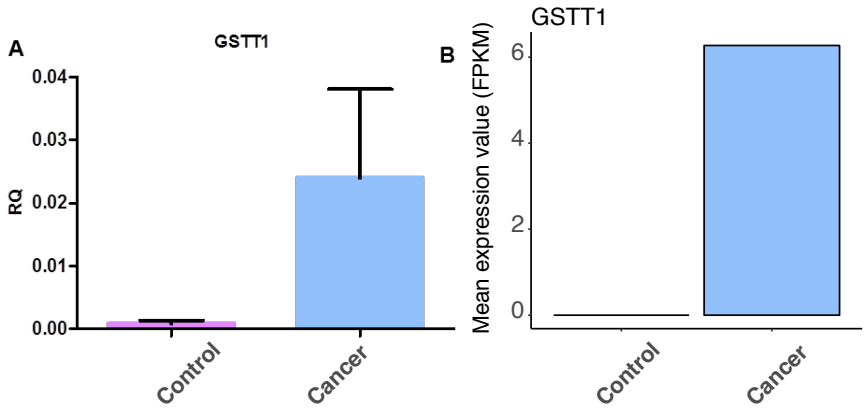


FIGURE 6.18: **qPCR and RNAseq expression of GSTT1.** Validation using qPCR shows GSTT1 expression in blood samples for control and cancer patients. The data shows similar up-regulation of expression in patient blood samples compared to controls, in both the validation experiments, and RNAseq data.

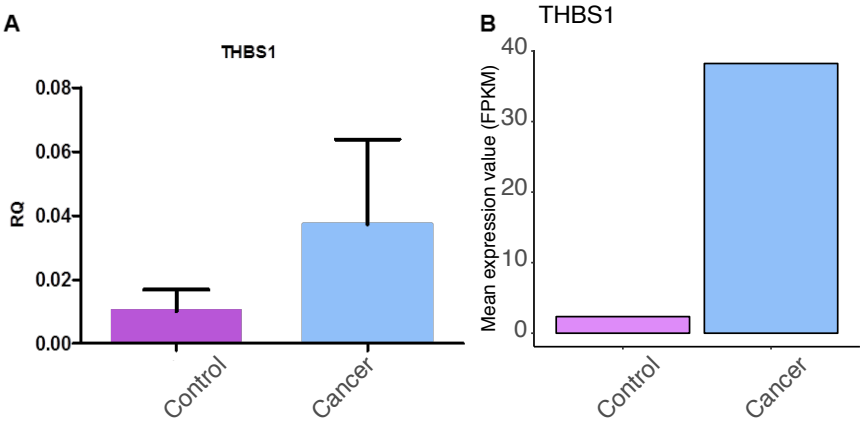


FIGURE 6.19: **qPCR and RNAseq expression of THBS1.** Validation using qPCR, shows THBS1 expression in blood and tissue samples for control and cancer patients. The data shows similar up-regulation of expression in patient blood samples compared to controls, in both the validation experiments, and RNAseq data.

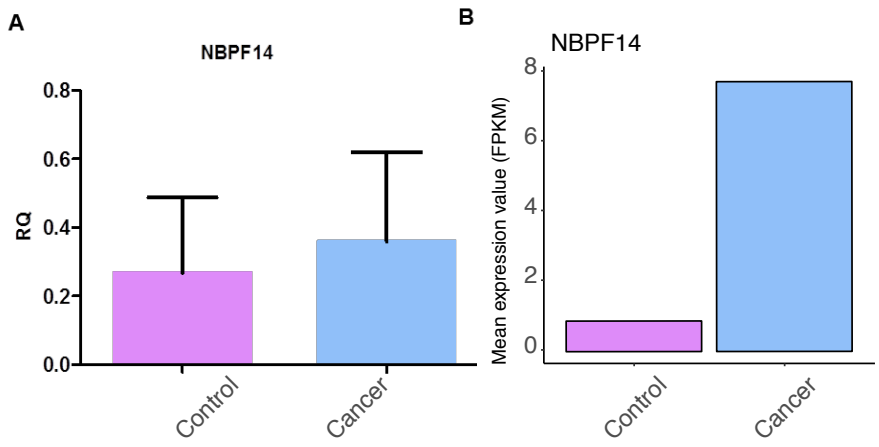


FIGURE 6.20: **qPCR and RNAseq expression of NBPF14A.** Validation using qPCR, shows NBPF14 expression in blood samples for control and cancer patients. The data shows similar upregulation of expression in patient blood and tissue samples compared to controls, in both the validation experiments, and RNAseq data.

The validation experiments of the 4 genes appear to mirror that of the RNAseq findings. Thus, a validation success rate of 100% adds confidence to our RNAseq results.

To further assess this data, average Log_2 fold change values for the qPCR and RNAseq are tabulated in Table 6.17.

Genes	qPCR Log_2 fold change	RNAseq Log_2 fold change
GSTT1	3.60	inf*
XIST	4.28	inf*
THBS1	0.15	4.02
NBPF14A	1.00	3.14

TABLE 6.17: **Levels of changes in expression for 4 validated genes, according to qPCR and RNAseq results.** The Log_2 fold change data generated from the validation and RNAseq analysis are listed above.
** No expression observed in control blood samples (FPKM=0)*

These results corroborate the findings above, and evidence the genes seen to be up-regulated in blood samples as per the RNAseq, are also seen to be up-regulated in the qPCR validation experiment, although not to the same extent. THBS1 shows a much smaller increase in expression levels in cancer blood samples compared to controls, in

contrast to the RNAseq results. Overall the results suggest that our transcriptomics analysis is robust.

6.5 Discussion

Genomic analysis is fast becoming a growing area of interest in cancer diagnostics and prognosis. In this chapter, we tested and evaluated three current genomic techniques for their suitability in liquid biomarker discovery.

First, we focused on cancer panels; in particular we studied a 79 ovarian cancer gene panel, processed by qPCR. This technique was used to evaluate differential gene signatures in ovarian patient blood samples compared to controls. The attraction of this method is the ease of use, directed analysis, high throughput and fast response time. This type of assay has tremendous potential in the clinical setting as the results are available to the patients within 2-3 hours. However, we also observed a number of downfalls such as the need to know beforehand the genes of interest, thus limiting the investigation to known onco and proto-onco genes.

Results of the cancer panel analysis revealed statically significant differential expression for 28 of the 79 genes when comparing the cancer and control samples. Several genes of interest were identified. In particular, we observed an upregulation of RAD51. RAD51 is involved in DNA repair by homologous recombination. Studies have shown an association of RAD51 mutations found in breast cancer, with an increased risk of developing cancer (Jara et al., 2017; Le Calvez-Kelm et al., 2012).

Interestingly the epithelial adhesion molecule (EpCAM) was shown to be down-regulated in cancer patients compared to controls. EpCAM confers cell-to-cell adhesion, and is known to be lost in many cancers due to the epithelial mesenchymal transition (EMT). The EMT process involves cells losing their rigid epithelial structure to take on more mobile mesenchymal features, allowing cancerous cells to invade through the basal membrane and enter the general vasculature, leading to metastasis. These results are supported by previous reports (Hyun et al., 2016; Imrich, Hachmeister, and Gires, 2012; Hyun et al., 2016) and match the expected cancer cell behaviour. Similarly, we found as expected, that the genes VEGFR1 and VEGFA were up-regulated in the cancer samples. These two genes are known to play a role in promoting angiogenesis and vasculogensis (Shibuya, 2011; Yao et al., 2011; Cao, 2009).

Furthermore, the expression pattern of the 79 genes was used as a finger print for each of the 30 samples (cancer n=15, control n=15). The principal component analysis of the 30 sample gene expression finger prints, revealed 3 distinct populations sharing

similar genetic profiles. One population in particular, consisted primarily of controls only (with the exception of 1 case), while the other two populations consisted mainly of cancer patients, again with the exception of one control case in one of these cancer clusters. Further analyses of this data did not reveal any trends within the cancer clusters amongst the patients, in terms of prognosis or treatment stage. In fact, patients were seen to move between the 2 predominant cancer clusters throughout their chemotherapy treatment. These results suggest that there are 2 unique gene profiles shared by these advanced ovarian cancer patients. These shifts in gene profiles could be driven by the treatment or due to inter-patient variability and changes in their disease profile.

***In silico* analysis of the 79 genes, revealed 5 genes with significantly up-regulated expression in the tumour tissue, in addition to being significantly up-resulted in blood samples compared to controls, as observed from the 79 ovarian cancer panel. Furthermore, evaluation of overall survival with the KM plotter tool, revealed a poorer overall survival to be associated with this over-expression in these 5 genes. Interestingly, 3 of the 5 identified genes are known to be involved in the DNA damage pathways, CCNE2, RAD51, and HJURP. The other 2 genes identified include, VEGFA, a growth factor, involved in angiogenesis and vascularisation and CTSD, part of the family of peptidases, known to be mutated in some cancers. These 5 genes could be explored further as potential 'liquid biomarkers', as gene expression for these 5 targets appear to be mirrored in both the tumour and blood. Moreover, these genes could be exploited further to understand mechanisms and pathways and their involvement in cancer. As mentioned a significant proportion of these 5 genes are related to the DNA damage pathway, which would be expected due to the nature and mechanisms of cancer.**

Overall the results suggest unique gene signatures exist amongst the ovarian cancer populations, which could be exploited further in disease stage grading and survival rate predictions, potentially allowing a better stratification of ovarian cancers based on their genetic profile.

Next, we evaluated the use of CNI in the analysing of patient samples (3 cancer and 3 control). Overall, we observed a high CNI 171 score for the cancer samples and a low (normal) CNI 171 score for the controls. However, there were some exceptions, including one cancer patient scoring within the normal range, and a control patient scoring above the normal range. Although it must be noted, the cancer case scored in the higher range of the normal score, and the control patient scored in the lower range of the cancer scoring. These outliers could be attributed to a number of underlying factors influencing the score. In respect to the cancer patient, the tumour may be slow growing and localised, in this case the patient was a stage IIB which would be classed as early stage. The control patient could have another undiagnosed underlying condition (other than the bulla,

the reason for this surgery) that could have attributed to the high CNI 171 score. Various diseases influenced by genetic disorders exist, any of which may be responsible for the high CNI 171 score. Other more general factors that could influence the CNI 171 score in both these patients include inter-patient variation, previous treatment or therapeutics, or exposure to damaging agents or materials, such as X-rays, asbestos, and harmful radiation that can cause DNA damage (Kamp, 2009; Borrego-Soto, Ortiz-López, and Rojas-Martínez, 2015). As expected we observed the highest CNI 171 score for the most advanced cancer case with several copy number changes apparent on the circos plot. Interestingly the same copy number changes were mirrored across both their tumour tissue and their matched blood plasma sample. Similar results were observed in the ovarian cancer patients, with both patients being diagnosed as Grade 3, advanced ovarian cancer, however only 1 patient demonstrated an enhanced CNI 171 score, and subsequently (as with the high scoring lung patient) died. The other ovarian cancer patient (Ov2), despite having an advanced cancer, exhibited a lower (normal range) CNI 171 score, but the patient also responded well to the treatment and has been reported to be in remission. Thus, these results offer not only the possibility of a blood based test providing chromosomal level data and ability to profile the cancer, but the potential of predictive and prognostic evaluation. These findings are in line with previously published reports (Beck et al., 2010; Weiss et al., 2017). The results are promising and allow further exploitation of the chromosomes and associated copy number changes. However, they also highlight the necessity for high accuracy and specificity when translating the research into a clinical setting.

Finally, we focused our study on designing a framework for identifying genes as potential liquid biomarkers using transcriptomics data. For this, we used the RNAseq data available from 3 patients and 3 controls provided by Oxford Wellcome Trust Centre. The data was mapped to the human genome annotation GENCODE 19 and we were able to identify 8,910 and 13,230 genes expressed with an expression level higher than 5 Fragments Per Kilobase of transcript per Million (FPKM) in the 3 patients, in blood and tissue respectively. By contrast we found 9,400 and 11,755 genes expressed with and FPKM larger than 5, in the control blood and tissue respectively. Moreover, by comparing the cancer and control samples we observed a number of differentially expressed genes across the three patients. It is interesting to note that we noticed a higher number of significantly differentially expressed gene in blood compared to tissue. Similar results were seen in a study by Huang et al., 2017.

In order to define a gene as a potential biomarker we required to see a similar change in the expression pattern in both blood and tissue. We found 21 genes that match this criterion. However, of them only X inactive-specific transcript (XIST) has shown an increase in expression in the cancer samples for both blood and tissue compared to control.

XIST is a non-coding RNA gene, involved in X chromosome silencing in female cells, and allows X chromosome equilibrium in males (Weakley et al., 2011). Long non-coding RNAs are known to often contribute to unrestricted growth and invasion of cancer cells, with XIST shown to be up-regulated in several cancers, including non-small cell lung cancer (NSCLC), colorectal, gastric, and giloma cancers (Ma et al., 2017; Fang, Sun, and Gong, 2016; Yu et al., 2017). A study in NSCLC, also showed up-regulation of XIST. Silencing of the XIST gene resulted in suppressive functions, including, inhibition of cell proliferation and invasion, and induced apoptosis (Wang et al., 2017). Fang *et al.* (2016), also observed the role of XIST as an oncogene in NSCLC. A study in colorectal cancers, also reported XIST as a prognostic factor, and its down-regulation was shown to overcome resistance to 5-fluorouracil, acting as potential therapeutic target (Xiao, Yurievich, and Yosypovych, 2017). Due to its up-regulation in tumours, as well the effects of it's silencing on cell proliferation, invasion and apoptosis, XIST is important not only as a potential biomarker but also a potential druggable target.

Next, to expand our liquid biopsy marker horizon we focused our analysis on genes that are differentially expressed in the blood of the cancer patients compared to control. In particular, we selected for future investigation only genes that are over-expressed in the cancer blood samples. Here we identified 18 genes that match our imposed selection threshold as described in Section 6.4.3. Relaxing this strict selection method increases considerably the number of potential search avenues. For practical purpose and given the limited sample availability we selected only 4 genes to pursue further for experimental validation using qPCR. The experimental results reflected the RNAseq observations for most of the genes in the context of blood, with the 5 validated genes seen to be up-regulated in both the validation experiments and RNAseq results, for the cancer blood samples compared to controls. Subtle differences between the results of the two analyses can be attributed to the small sample size and again inter-patient variability where specific patient disease profiles will not be identical. It must also be considered that the RNAseq and qPCR were conducted on samples from different individuals (both for cancer and control).

Finally, we integrated our CNI and RNAseq results in order to get an overall picture of the tumour. Thus, looking at the CNI circos plots for patient 1 (MAS4), it is interesting to note that while chromosome 3, 5, 8, 10, and 13 seem to exhibit a larger number of chromosomal abnormalities, this is not reflected in changes in the gene expression levels for those locations. However, a direct effect of the genome abnormality on the expression levels has been previously observed in cancers (Lavarino et al., 2009). The discrepancies could be attributed to the tissue/blood differences, as well as a decrease in expression level in those genes in the blood sample, which did not constitute the focus of our RNAseq analysis.

Findings from this chapter show the huge potential of genomic analysis, with promising initial results from the ovarian 79 cancer gene panel providing valuable insights on molecu-lar signatures. CNI scoring method gives a snapshot of genomic disruptions that are the result of disease evolution and can even potentially drive it. Therefore, a larger cohort study is necessary in order to evaluate the method's usefulness in a clinical setting. Finally, trans-criptomics analysis provided exciting insights into new potential markers of interest parti-cularly in a 'liquid biopsy' context. However, the contrast between the gene expression in the tissue and blood of the same patients raises some questions on how well the clinical blood biomarkers can describe the tumour behaviour and heterogeneity.

Despite these challenges, the three methods studied here opened up significant new possibilities for biomarker discovery.

Chapter 7

General discussion and concluding remarks

7.1 Impact and importance

Over the past decades a vast improvement in cancer survival has been achieved, with some cancer types having a 10 year survival as high as 80%. e.g. breast and prostate. However, in other cancers i.e. ovarian and lung, survival rates remain poor, being as low as 35 and even 5% in some cases (CRUK, 2017) (see Fig 7.1).

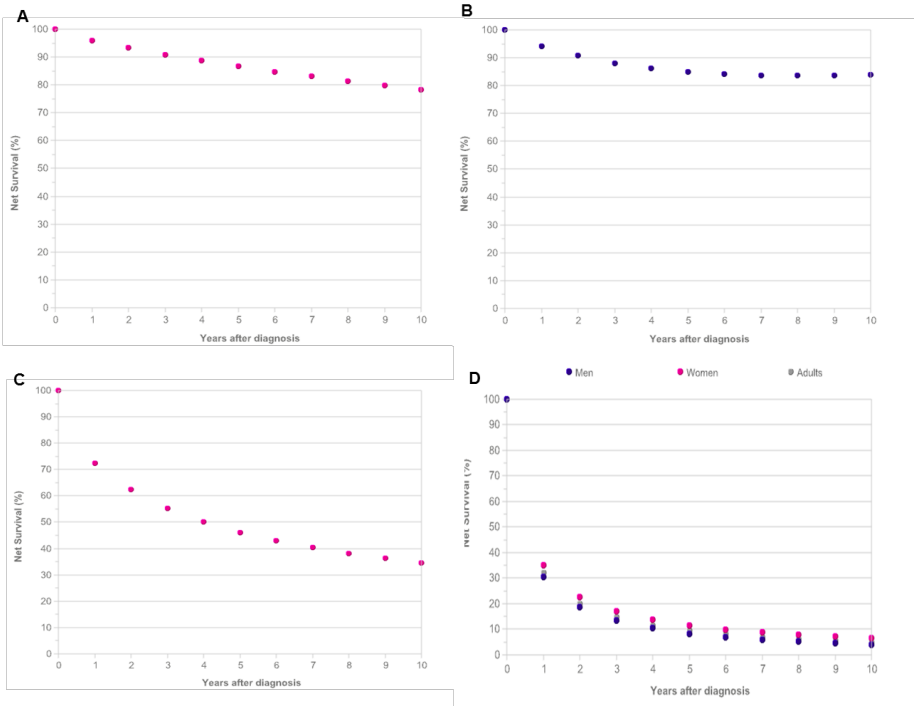


FIGURE 7.1: **Ten-year survival rates in different cancer types.** (A) Breast, (B) Prostate, (C) Ovarian, (D) Lung cancer (CRUK, 2017).

This disparity can be related to a number of reasons, including lack of early diagnosis, real-time disease monitoring, and limited therapeutic options. The introduction of robust screening tools in both, breast and prostate cancers has hugely benefitted cancer survival, as treatment in early stages whilst the disease is still localised offers more options, and is associated with a promising outcome (CRUK, 2017; Goldstraw et al., 2007). Cancers detected at stage I would benefit from a 98% in ovarian, and just over 80% survival rate in prostate (CRUK, 2017).

Response to therapies is highly patient and cancer type specific, therefore not all patients can benefit from the same therapy. Due to lack of real-time screening and monitoring options, valuable time could pass before non-response is observed and treatment can be changed. Furthermore, the risk of recurrence in both ovarian and lung cancers is high, particularly in advanced stages, which make up the majority of diagnosed cases. Once the cancer recurs the prognosis is poor, as a curative options are almost non-existent at this point (Luvero, Milani, and Ledermann, 2014; Consonni et al., 2015).

Over the years, several therapeutic targets have been developed, some with very successful results, i.e. anti-HER2 treatment in breast cancer (Vrbic et al., 2013; Rimawi, De Angelis, and Schiff, 2015). However, cancers evolve and adapt over time, becoming resistant to drugs, a behaviour described as an evolutionary pressure imposed by the treatment (Friedman, 2016). Thus, discovery of new biomarkers to characterise and provide information on the disease stage and evolution is crucial in the development of new therapies. Identification and evaluation of novel liquid biomarkers could not only allow development of new thera-peutic targets, but could also inform on the biology of cancer spread. This is particularly poignant as circulating cancer cells (CTCs) in the blood are known to propagate metastasis (Polyak and Weinberg, 2009; Munzone et al., 2012). Moreover, real-time measurements of liquid biomarkers, could inform of variations in the cancer biology, particularly during drug therapies where these changes may render the patients treatment ineffective, prompting rapid changes in therapeutic regimes to be implemented.

All the aforementioned areas within ovarian and lung cancers would benefit from development of personalised treatment plans, which is being widely accepted as the future of effective cancer treatment and management (Verma, 2012).

Here, we developed a pipeline for biomarker discovery to identify robust liquid biomarkers, as shown in Fig 7.2.

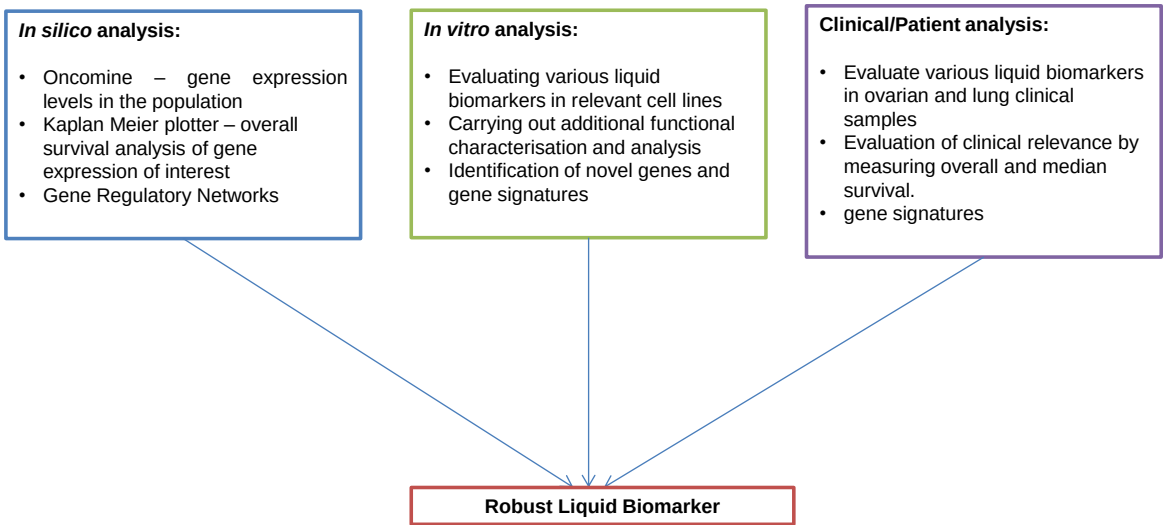


FIGURE 7.2: **Flow diagram of liquid biopsy identification study design.** Liquid biomarker discovery is informed by *in silico*, *in vitro* and clinical analysis.

7.2 Circulating tumour cells, circulating nucleic acids and liquid biomarker discovery in ovarian and lung cancer

Cancerous tumours are shown to shed from their primary sites circulating tumour cells (CTCs) (Allard et al., 2004; Hou et al., 2012; Park et al., 2012), and circulating nucleic acids, (ctNAs, i.e. DNA and RNA) into the blood (Volinia et al., 2008; Freidin et al., 2015). These CTCs and ctNAs originate directly from the cancer site, and have been shown to share biological tumour-specific characteristics and genomic profiles (Hiltermann, Wekken, and Groen, 2012; Freidin et al., 2015). Thus, the presence of genetic tumour material in the blood opens the potential for liquid biomarker discovery. It can potentially facilitate the development of novel therapeutics and provide new insights in to the mechanisms and biology of the disease (Beck et al., 2013; Weiss et al., 2017; Zeng et al., 2011).

From our studies, we show that we can identify, characterise, and quantify CTCs in the blood of ovarian and lung cancer patients, using the ImagestreamX technology. We also showed that CTCs were significantly over-expressed in both ovarian and lung cancers compared to control samples, even in the early setting (for lung cancer patients only). A strong prognostic value was observed in advanced stages of lung cancer patient compared to early stage using ctDNA analysis; with our findings mirroring previously published reports in other cancers (Umetani et al., 2006a; Umetani et al., 2006b; Wang et al., 2003).

We also demonstrated the presence of unique biomarkers in the blood of ovarian and

lung cancer patients, which are not seen in control samples. Various gene signatures and individual genes were identified using several approaches such as GRNs, gene cancer panel evaluation and RNAseq.

More specifically, cancer panel evaluation of 79 genes in ovarian cancer patients, also revealed significantly differentially expressed genes in cancer patients compared to controls. Further analysis revealed specific gene signatures in cancer patients.

Analysing RNAseq data from lung cancer patients, we observed differential gene expressions in cancer patients compared to controls. More in-depth analysis revealed the presence of lung cancer subtypes specific expression profiles. The analysis also identified a novel gene, XIST as being significantly over-expressed in the tissue and blood of lung cancer patients compared to controls.

Finally, we report the identification a novel marker, RAD51AP1, exploiting existing microarray data through the use of GRNs and UNIPs. *In silico* clinical validation revealed that this gene was over-expressed in both ovarian and lung cancer tissue and blood samples compared to controls. *In silico* analysis using Oncomine also revealed a similar over-expression in the tissue of ovarian and lung cancers. KM plotter showed that this over-expression is associated with a poorer overall survival.

7.3 The clinical utility of CTCs, ctNAs and liquid biomarker discovery in lung and ovarian cancer

In this study we showed that the over-expression of CTCs in the blood of ovarian and lung cancer patients was associated with a poorer prognosis. Moreover, in-depth analysis of CTCs in lung cancer patients by stage revealed CTC quantification could be used to differentiate early stage from advanced cancers. Most importantly, diagnostic utility evaluations revealed high sensitivity and specificity in both early and late stage lung cancers compared to controls. As all recruited ovarian cancer patients were diagnosed at stage III and IV, a si-milar analysis was not possible. However, high sensitivity and specificity values for ovarian cancers CTC detection were also calculated, suggesting a strong diagnostic utility of CTC evaluation and quantification using the ImagestreamX. These results show the potential use of CTCs as a non-invasive screening tool in support of early diagnosis in lung and possibly ovarian cancer. Moreover, our observations inform the utility of this biomarker for prognostic purposes, as increased CTCs burden correlates with an increased tumour burden, potentially broadening its use to real-time monitoring of treatment and relapses. By conducting a short-term treatment monitoring study we observed a specific trend in CTC levels for patients who responded well to treatment. Expanding this to longer term monitoring in the future, could provide better insights.

Evaluation of ctDNA was conducted by calculating the DNA Integrity Index, from measurements of Alu repeats in the plasma of patients. We observed a significantly increased DNA Integrity Index (ctDNA burden) compared to controls, in both cancer types. Diagnostic utility evaluation however revealed poor sensitivity and specificity values for both cancer types. Further analysis by lung cancer staging revealed a significantly higher DNA Integrity Index in advanced lung cancer cases, compared to both early stage and controls, with diagnostic utility evaluation showing high sensitivity and specificity. This increased DNA Integrity Index was also shown to be associated with a poorer overall survival, demonstrating the prognostic value of this technique in the advanced setting.

In our work we also demonstrate the ability to identify novel gene signatures as potential liquid biomarkers in ovarian and lung cancers. These findings support the notion that cancer genomic material is present in patients' blood, and can be exploited for various uses, i.e. development of therapeutic target, monitoring of disease and genomic changes to cancer profile, as well as informing on treatment options.

Specifically, here we describe the clinical significance of RAD51AP1 gene in ovarian and lung cancers. Our findings reveal that over-expression RAD51AP1 is also associated with a poorer prognosis. Further functional analysis of this gene demonstrated its key role in cancer progression. By conducting gene knockouts *in vitro*, we have shown that the cancer growth was reduced *in vitro*. Protein analysis revealed various apoptotic and immune response genes to be up-regulated, whilst tumorigenic ones were down-regulated (Chudasama et al., 2017b). These findings suggest RAD51AP1 plays a critical role in both ovarian and lung cancers, and could be further exploited as a therapeutic target, by means of down-regulation.

Our findings underpin the clinical utility of all the techniques and assays described above. Moreover our analyses address the current limitations faced by cancer treatment and management, and provide potential solutions (see Table 7.1).

Problem area	**Solution: liquid biomarker**	**Cancer Evaluation Results**
Early diagnosis	CTCs	Ovarian & Lung
Real-time treatment monitoring and relapse	CTCs	Ovarian & Lung
Biomarker discovery	Circulating nucleic acids and tumour material	Ovarian & Lung
Monitoring of genomic changes to cancer	ctDNA (CNI), Circulating tumour material	Lung only for CNI; Ovarian & Lung
Prognostic value	CTCs, ctDNA	Ovarian & Lung (ctDNA advanced lung only)

TABLE 7.1: **Limitations currently faced in ovarian and lung cancers.** Current identified limitations, with liquid biomarkers solutions from this study.

Collectively these data are indicative of potential use of a combination of these liquid biomarkers to support current clinical practice and improve diagnosis and prognosis in these two cancer types.

7.4 Implementation of liquid biopsies in to the clinical setting

Various factors must be taken in to consideration in order for a technique or assay to be implemented into routine clinical practice in the NHS, including addressing an unmet clinical need, methodology robustness, repeatability and reliability, clinical value and validity, and cost-effectiveness. In this respect, liquid biomarkers have proven their huge potential in a clinical setting. Moreover, early diagnosis will enable treatment in localised stages, where costs are lower. It has been reported by the National Institute for Health and Care Excellence (NICE) that treatment of ovarian cancer in Stage I is approximately £5,328 per patient, but it increases to £17,810 at Stage IV. NICE predicts savings of over £16 million with early diagnosis in ovarian cancers (NICE, 2017). Predicting and treatment monitoring would also inform on the best drugs for the patient based on their genomic profile, as well as an effective treatment update upon a poor response, due to drug resistance for example. In the long term we feel the benefits and costs will outweigh the current burdens and poor survival rates.

In addition, an assay that can be used universally across multiple cancer types, would also prove advantageous. Currently, we know this is problematic due to each diseases complexity and heterogeneity. For all these reason, Cancer Research UK (CRUK) have developed a biomarker roadmap for development of a robust biomarker assay, Fig 7.3.

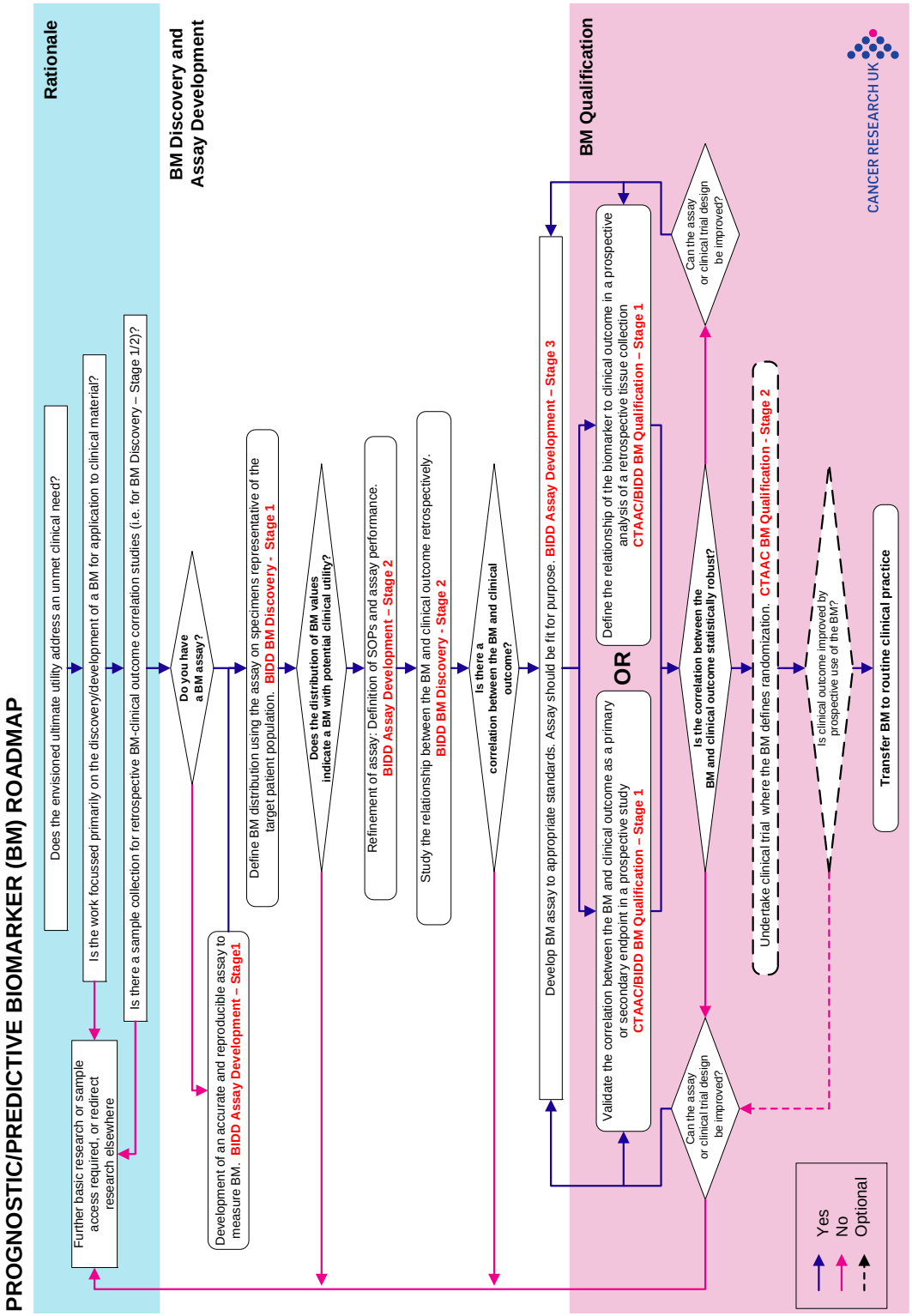


FIGURE 7.3: **CRUK Biomarker development flowchart.** Biomarker roadmap structure outlined by CRUK.

In the present work, we showed that CTCs yielded similar results across both cancer types. By contrast, transcriptomic evaluation, as would be expected, yielded very different gene signatures, with the exception of RAD51AP1, which was shown to be overexpressed in tissue and blood samples across both cancer types.

These strict criteria are essential, and must be met amongst other factors in order for the assay to be implemented into routine clinical practice.

7.5 Limitations

One of the biggest limiting factors of any study is inter-patient variability. Inter-patient variability is described as differences in genetics and/or pharmacokinetics from individual to individual (Holliday et al., 2014). Inter-patient variation is therefore likely to mask trends in gene and protein changes, particularly if these changes are subtle.

In particular, we observed the effects of inter-patient variability when we measured CTC levels in the blood of ovarian and lung cancer patients. As highlighted earlier, the exact mechanism of CTC shedding is not well understood (Polyak and Weinberg, 2009; Munzone et al., 2012), and the frequency of CTCs will differ from patient to patient. In turn this impacts the level of genomic material present in blood, as it is dependent on the body's mechanisms of controlling and discarding these aberrant cells.

Furthermore, differences in clinical samples were also reflected in the assessment of expression signatures of various genes. Transcriptomics evaluation has shown varying genetic profiles from patient to patient. Moreover, both ovarian and lung cancers are recognised as strikingly heterogeneous diseases. The high standard deviation bars from qPCR analysis can thus be interpreted as a result of inter-patient variation, and may translate into a lack of statistical significant results in some cases. In one example, RAD51AP1 expression in breast cancer compared to controls did not reach significance, despite a marked increase in cancer patients. Thus, inter-patient variation, provides yet another hurdle in the ever complicated saga of improving cancer management and treatment (Jordan et al., 2003).

Tumour heterogeneity is an added complexity in the world of cancer research. The plane/collection area could thus affect the genomic and phenotypical characteristics of the analysed samples (Fidler, 1978; Fisher, Pusztai, and Swanton, 2013). Findings from single-cell CTC analysis have also shown that cells from the same tumour may be genetically and phenotypically different from one another, dependent on the exact origin on the tumour (Saadatpour et al., 2015). Furthermore, single cell studies have reported

modulations in oncogenic, proliferation and immune signalling, differential growth factor receptors and KRAS expression, and ER receptor expression in breast cancers (Miwa et al., 2015; Kim et al., 2015; Janiszewska et al., 2015; Shalek et al., 2013; Patel et al., 2014; Babayan et al., 2013).

Another factor contributing to increased variability, particularly in immunohistochemistry (IHC), is the nature of the analysis method. As mentioned, ovarian and lung tumours can be highly heterogeneous. Slides for IHC analysis are cut from the FFPE block housing the tumour sample. Dependent on the plane and area of the tumour tissue sliced for experimentation, results could vary and highlight differential protein expression levels, as we have seen in the case of RAD51AP1. As each cancer sample is taken from an unspecified area of the tumour, it may contain multiple tissue types that were phenotypically different. Furthermore, IHC by DAB staining is not only a semi-quantitative method of analysis; it also gives an indication on the number of cells showing expression. However, the expression levels cannot be determined accurately. While this information is useful, the counting is manual, and thus the method is also subjective to the presence of residual background noise. An automated counting system using stringent marking criteria could overcome this. Alternatives for DAB staining are also now commercially available.

On the matter of subjectivity, various aspects of this study require a level of individual/personal judgment and assessment. This is the case with flow-cytometry analysis using ImagestreamX, whereby gating of cells is required in order to apply various parameters for analysis and quantification. For these reasons, initiatives have been developed including ACCEPT, the automated imaging analysis software, shown to increase concordance from 30% in manual scoring to 51% (Zeune et al., 2017). Although extensive care is taken to ensure these are applied accurately, the measurements are done "manually", by human eye, and results could vary subtly between operators. The Clearbridge study in Chapter 3 required a similar approach, where slides were stained with Haematoxylin and Eosin for cancer cells, however this is not a confirmatory stain and two pathologists were asked to score these independently for the presence or absence of CTCs. Inter observer agreement was calculated at 80.4%, with a Kappa statistic of 0.6 $\pm$ 1.1 (p-*value<0.001*), highlighting subjectivity of the assessment procedure. Even in the presence of strict criteria and rules, inter observer subjectivity will be a continuing issue.

7.6 Future directions

7.6.1 Large patient cohorts

In this project, different numbers of samples were used for various assays, dependent on the time of the experimentations and type of sample required. In all instances, sample

levels did not exceed 20 for any given type (i.e. cancer or controls) due to limitations in patient recruitment and ethical approval. For the RNAseq and CNI work, sample sizes were limited due to cost constraints to 6, including controls. The same constraint limited our analyses to lung only samples as well. Future perspectives include expansion of sample types to ovarian cancer, as well as increasing the cohort numbers in both disease and controls, in order to confirm the observed results. This increase in cohort size should take in to consideration subtypes and grading of cancer to account for any specific changes as a result of cancer pathology. Moreover, the research will benefit by further expanding the analysis to account for variations according for gender and ethnicity, as these has been shown to affect gene expressions (Knappskog et al., 2014). Genetic variations ascribed to ethnicity have been associated to differences in treatment response (Rotger, Csajka, and Telenti, 2006), and drug susceptibility (O'Donnell and Dolan, 2009).

7.6.2 Developing an assay for long term monitoring of patients for relapse and treatment

In this work, we reported a significant increase in CTCs in the blood of ovarian and lung cancer patients compared to controls. We followed ovarian cancer patients on the METRO-BIBF trial, with correlations observed in CTC trends overtime with responders compared to non-responders. However, this study was limited by the number of cycles (maximum of 5 cycles for some patients) followed and patients recruited, with only short term follow-up data available, up to 2 years for some patients. Long term monitoring data could provide more insights in to the trends of CTC changes and their association with response to treatment. Furthermore, coupling this analysis with molecular profiling at various time points would inform of crucial biological and genomic changes undergone as a result of the treatment.

Long term monitoring of both ovarian and lung cancer patients can also measure any relapse post treatment or surgery. Prompt treatment of the recurrent cancer may result in a better prognosis than at present.

7.6.3 Elucidating the cytostatic effect of RAD51AP1

In the present study, we reported on the cytostatic effect exerted by silencing the RAD51AP1 gene in SKOV-3, MDAH-2774, and A549 cell lines. Future work could provide valuable insight into this effect by focusing on the analysis of cell cycle progression markers, including the expression of cyclin B1 and the phosphorylation of cyclin dependent kinase 1 (CDK1), which are increased at the G_2/M transition (Stewart and Dell'Orco, 1992) in addition to the proliferating cell nuclear antigen (PCNA), which increases at the G_1/S transition (Stewart and Dell'Orco, 1992). Furthermore, a better understanding of the protein

interactions of RAD51AP1, see Fig 7.4, with other predicted proteins other than RAD51 will provide a novel insight into its signalling mechanisms.

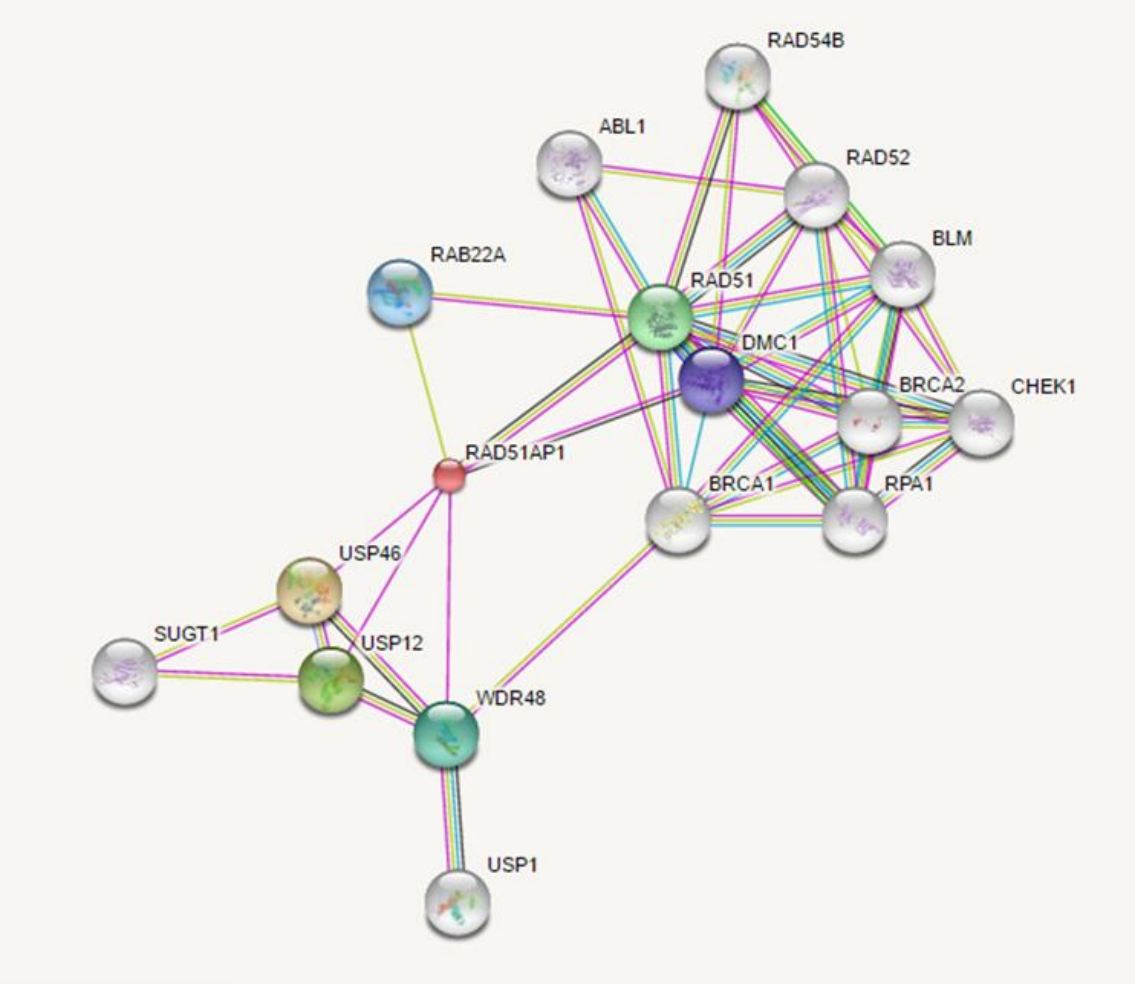


FIGURE 7.4: **String motif of associated proteins to RAD51AP1 using STRING (String-db.org, 2018).** Schematic of RAD51AP1 protein-protein association network. Each node represents a protein and each edge indicates an association. Known interaction are coloured in cyan and pink, while predicted associations (e.g. gene neighborhood, gene fusions, gene co-occurrence) are shown in green, red, and blue.

7.6.4 Use of additional *in vitro* models to study the effects of RAD51AP1

In this study we used two epithelial ovarian adenocarcinoma cell lines and one lung cancer adenocarcinoma cell line, as *in vitro* models of ovarian and lung cancer. By expanding the analysis to include various cancer cell lines reflecting the different cancer types and stages, would provide a better understanding of the observed variations within cancer

types of differing pathologies. Furthermore, the addition of both control healthy ovarian and lung cell lines would be beneficial.

7.6.5 Further elucidation of the effects of RAD51AP1 on the mTOR pathway

Findings from this work revealed compromised mTOR signalling in all three cell lines *in vitro*. In SKOV-3 cells, decreased mTOR signalling along with DEPTOR increase would inhibit the activity of complexes. Future studies concentrating on the protein changes of the same key components would be useful. Future work on protein expression of these components or changes in the phosphorylation status of downstream components, such as S6 kinase (S6K), as well as upstream, components including Akt would provide a more detailed insight into the effects of RAD51AP1 on mTOR signalling.

7.6.6 Development of a rapid liquid biomarker test: Concluding Remarks

The ultimate aim of this study was to provide a framework for developing a liquid biopsy test, allowing for real-time testing and monitoring of cancer patients to support various aspects of the treatment. Thus, the natural route of building on the results provided by this project is to develop a pipeline of tests using the assays described here in order to address early diagnosis, relapse and treatment monitoring. Fig 7.5 and 7.6 outline these proposals.

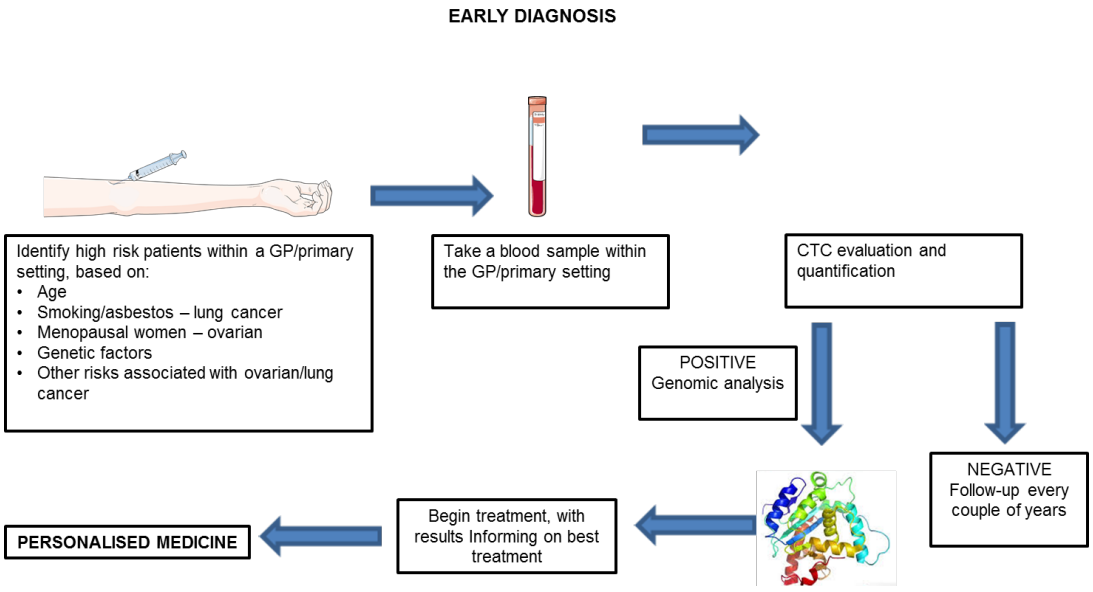


FIGURE 7.5: **Liquid biomarkers screening pipeline.** Proposed screening method and management for ovarian and lung cancers using liquid biomarkers.

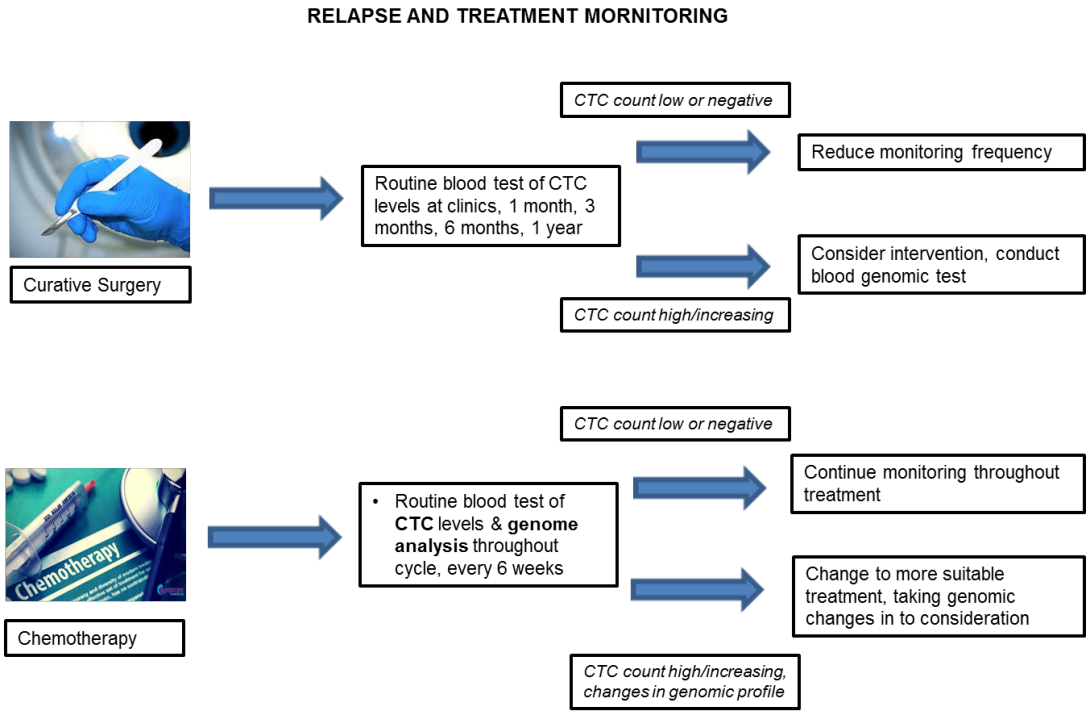


FIGURE 7.6: **Liquid biomarkers relapse and treatment monitoring pipeline.** Proposed relapse and treatment monitoring method and management for ovarian and lung cancers using liquid biomarkers.

These proposed workflows utilise a wide repertoire of liquid biomarker assays, enabling a detailed picture of the cancer to be obtained on a regular basis. The test results will form the basis of any subsequent treatment or management of that patient. This detailed and more systematic approach will enable a personalised treatment plan to be developed for patients.

Finally, personalised medicine is the ultimate goal in cancer diagnostics, treatment and management. Thus it is crucial to determine detectable biomarkers in cancer, in a view to developing better diagnostic and treatment plans. A great deal can be achieved through the development of liquid biomarkers, and their implementation in to clinical practice.

Appendix A

Supplementary Information

A.1 Extended Tables

Date	Patient code	Age	Gender	Pathology	Cancer Type	Stage
04/11/14	EC001	56	F	Adenocarcinoma	Primary	T3NXM1A
16/11/14	RR002	65	M	Adenocarcinoma	Primary	T2BN2M0
09/11/14	TK003	71	M	Adenocarcinoma	Primary	T23N0M2
14/12/14	PD004	72	M	Squamous Cell Carcinoma	Primary	T2BN2MX
13/01/15	WS005	58	M	Papillary Adenocarcinoma	Primary	T2AN0MX
09/02/15	PG006	67	F	Adenocarcinoma	Primary	T1AN0M0
20/04/15	JH007	57	F	Squamous Cell Carcinoma	Primary	T2AN1M0
27/04/15	PW008	50	F	Adenocarcinoma	Primary	T1AN2MX
14/06/15	JC009	64	M	Adenocarcinoma	Primary	T3N1M0
15/06/15	TB010	84	F	Adenocarcinoma	Primary	T1BN0M0
25/06/15	TB011	57	F	Non-mucinous adenocarcinoma	Primary	T1AN0MX
25/06/16	JB012	76	F	Metastatic lung, from lung primary	Metastatic	No staging
06/12/15	MAS01	56	M	Primary Adenocarinoma (squamous differentiation)	Primary	T4N0MXR1
07/12/15	MAS02	77	F	Primary multifocal mucinous Adenocarcimona (bronchoalveolar)	Primary	T3N0MXR0

17/01/16	**MAS03 66**	**M**	**2 metastatic renal and bladder cell carcinoma, 1 primary lung adeno**	**Metastatic**	**No staging**
18/01/16	**MAS04 71**	**M**	**Squamous CC 2 nodules in 2 separate lobes**	**Primary**	**T4N1MXPL0R0**
27/01/16	**MAS05 76**	**M**	**Keratinising squamous cell carcinoma**	**Primary**	**pT2aPL1N0R0**
27/01/16	**MAS06 75**	**F**	**Undifferentiated carcinoma**	**Primary**	**pT1bPL0N1R0**
28/01/16	**MAS07 56**	**F**	**Mucinous adenocarcinoma acinar predominant**	**Primary**	**pT2aPL0N2 (station 6) ?R1**
01/02/16	**MAS08 76**	**M**	**Adenocarcinoma (local recurrence?)**	**Primary**	**pT1aN0MxPL0R0 (or pT2aN0MxPL1?)**
02/02/16	**MAS09 60**	**M**	**Metastatic adenocarcinoma (acinar/cribriform architecture, nuclear stratification and necrosis morphologically favouring metastasis from a colorectal primary)**	**Metastatic**	**No staging**
08/02/16	**MAS10 71**	**M**	**Keratinising squamous cell carcinoma**	**Primary**	**T2BN0MX**
09/02/16	**MAS11 70**	**F**	**Non-mucinous adenocarcinoma**	**Primary**	**T1AN0MX**
09/02/16	**MAS12 50**	**F**	**Non-mucinous adenocarcinoma**	**Primary**	**T2AN0MX**
22/02/16	**MAS13 67**	**F**	**Non-mucinous adenocarcinoma**	**Primary**	**T2BN2MX**
22/02/16	**MAS14 56**	**F**	**Non-mucinous adenocarcinoma**	**Primary**	**T1AN0MX**

23/02/16	**MAS15 70**	**M**	**Non keratinising squamous cell carcinoma**	**Primary**	**T2AN0MX**
25/02/16	**MAS16 34**	**M**	**Metastatic colorectal adenocarinoma**	**Metastatic**	**No staging**
25/02/16	**MAS17 72**	**M**	**Oesphageal cancer - adenocarcinoma**	**Primary**	**No staging**
29/02/16	**MAS18 73**	**M**	**Keratinising squamous cell carcinoma**	**Primary**	**T4N2MX**
29/02/16	**MAS19 72**	**M**	**Non-mucinous adenocarcinoma**	**Primary**	**T1BN0MX**
03/03/16	**MAS20 56**	**M**	**Squamous Cell Carcinoma secondary from lung**	**Metastatic**	**No staging**
03/03/16	**MAS21 77**	**F**	**None-keratinising squamous carcinoma**	**Primary**	**T2AN0MX**
08/03/16	**MAS22 72**	**F**	**Adenocarcinoma**	**Primary**	**T2AN2MX**
21/03/16	**MAS23 62**	**F**	**Adenocarcinoma**	**Primary**	**T2AN0MX**
05/04/16	**MAS24 79**	**F**	**Adenocarcinoma**	**Primary**	**T2AN0MX**

TABLE A.1: **Patient information.** Phenotypic details of patients used in this study.

Bibliography

Agostini, M et al. (2012). "Circulating cell-free DNA: a promising marker of regional lymphonode metastasis in breast cancer patients". In: *Cancer Biomark* 11.2-3, pp. 89–98. DOI: 10.3233/CBM-2012-0263.

Aktas, B. et al. (2009). "Stem cell and epithelial-mesenchymal transition markers are frequently overexpressed in circulating tumor cells of metastatic breast cancer patients". In: *Breast Cancer Res* 11.4, R46. DOI: 10.1186/bcr2333.

Allard, W. J. et al. (2004). "Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases". In: *Clin Cancer Res* 10.20, pp. 6897–904. DOI: 10.1158/1078-0432.CCR-04-0378.

Ankeny, J. S. et al. (2016). "Circulating tumour cells as a biomarker for diagnosis and staging in pancreatic cancer". In: *Br J Cancer* 114.12, pp. 1367–75. DOI: 10.1038/bjc.2016.121.

Asbestos.ucoz.net (2018). "http://asbestos.ucoz.net/photo/lung_cancer/2-3". In: *31 Jan 2018*.

Autebert, J. et al. (2012). "Microfluidic: an innovative tool for efficient cell sorting". In: *Methods* 57.3, pp. 297–307. DOI: 10.1016/j.ymeth.2012.07.002.

Babayan, A. et al. (2013). "Heterogeneity of estrogen receptor expression in circulating tumor cells from metastatic breast cancer patients". In: *PLoS One* 8.9, e75038. DOI: 10.1371/journal.pone.0075038.

Bannister, N. and J. Broggio (2016). "Survival for people diagnosed by cancer type, split by stage at diagnosis. Data based on people diagnosed in England in 2012, 2013 and 2014. Produced in collaboration with Public Health England (PHE)". In: *Statistical bulletin: Cancer survival by stage at diagnosis for England (experimental statistics)*, pp. 1–30.

Barriere, G. et al. (2014). "Circulating tumor cells and epithelial, mesenchymal and stemness markers: characterization of cell subpopulations". In: *Ann Transl Med* 2.11, p. 109. DOI: 10.3978/j.issn.2305-5839.2014.10.04.

Basiji, D. A. et al. (2007). "Cellular image analysis and imaging by flow cytometry". In: *Clin Lab Med* 27.3, pp. 653–70, viii. DOI: 10.1016/j.cll.2007.05.008.

Beck, J. et al. (2010). "Next generation sequencing of serum circulating nucleic acids from patients with invasive ductal breast cancer reveals differences to healthy and nonmalignant controls". In: *Mol Cancer Res* 8.3, pp. 335–42. DOI: 10.1158/1541-7786.MCR-09-0314.

Beck, J. et al. (2013). "Genome aberrations in canine mammary carcinomas and their detection in cell-free plasma DNA". In: *PLoS One* 8.9, e75485. DOI: 10.1371/journal.pone.0075485.

Bell, C. E. and A. J. Watson (2009). "SNAI1 and SNAI2 are asymmetrically expressed at the 2-cell stage and become segregated to the TE in the mouse blastocyst". In: *PLoS One* 4.12, e8530. DOI: 10.1371/journal.pone.0008530.

Beroukhim, R. et al. (2010). "The landscape of somatic copy-number alteration across human cancers". In: *Nature* 463.7283, pp. 899–905. DOI: 10.1038/nature08822.

Bettegowda, C. et al. (2014). "Detection of circulating tumor DNA in early- and late-stage human malignancies". In: *Sci Transl Med* 6.224, 224ra24. DOI: 10.1126/scitranslmed.3007094.

Bigio, I. J. et al. (2000). "Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results". In: *J Biomed Opt* 5.2, pp. 221–8. DOI: 10.1117/1.429990.

Bo, V. et al. (2014). "Discovering study-specific gene regulatory networks". In: *PLoS One* 9.9, e106524. DOI: 10.1371/journal.pone.0106524.

Boddy, J. L. et al. (2006). "The role of cell-free DNA size distribution in the management of prostate cancer". In: *Oncol Res* 16.1, pp. 35–41.

Borrego-Soto, G., R. Ortiz-López, and A. Rojas-Martínez (2015). "Ionizing radiation-induced DNA injury and damage detection in patients with breast cancer". In: *Genet Mol Biol* 38.4, pp. 420–32. DOI: 10.1590/S1415-475738420150019.

Brambilla, E and A Gazdar (2009). "Pathogenesis of lung cancer signalling pathways: roadmap for therapies". In: *Eur Respir J* 33.6, pp. 1485–97. DOI: 10.1183/09031936.00014009.

Burger, R. A. et al. (2007). "Phase II trial of bevacizumab in persistent or recurrent epithelial ovarian cancer or primary peritoneal cancer: a Gynecologic Oncology Group Study". In: *J Clin Oncol* 25.33, pp. 5165–71. DOI: 10.1200/JCO.2007.11.5345.

Buys, S. S. et al. (2011). "Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial". In: *JAMA* 305.22, pp. 2295–303. DOI: 10.1001/jama.2011.766.

Cancer Genome Atlas Network (2012). "Comprehensive molecular portraits of human breast tumours". In: *Nature* 490.7418, pp. 61–70. DOI: 10.1038/nature11412.

Cancer Genome Atlas Research Network (2011). "Integrated genomic analyses of ovarian carcinoma". In: *Nature* 474.7353, pp. 609–15. DOI: 10.1038/nature10166.

Cancer.org (2018). "https://www.cancer.org/cancer/non-small-cell-lung-cancer/detection-diagnosis-staging/staging.html". In: *6 Jan 2018*.

Cao, Y. (2009). "Positive and negative modulation of angiogenesis by VEGFR1 ligands". In: *Sci Signal* 2.59, re1. DOI: 10.1126/scisignal.259re1.

Carnio, S. et al. (2013). "Prognostic and predictive biomarkers in early stage non-small cell lung cancer: tumor based approaches including gene signatures". In: *Transl Lung Cancer Res* 2.5, pp. 372–81. DOI: 10.3978/j.issn.2218-6751.2013.10.05.

Cavallaro, U. and G. Christofori (2004). "Cell adhesion and signalling by cadherins and Ig-CAMs in cancer". In: *Nat Rev Cancer* 4.2, pp. 118–32. DOI: 10.1038/nrc1276.

Cen, P. et al. (2012). "Circulating tumor cells in the diagnosis and management of pancreatic cancer". In: *Biochim Biophys Acta* 1826.2, pp. 350–6. DOI: 10.1016/j.bbcan.2012.05.007.

Chan, K. C. A. et al. (2004). "Size distributions of maternal and fetal DNA in maternal plasma". In: *Clin Chem* 50.1, pp. 88–92. DOI: 10.1373/clinchem.2003.024893.

Chan, Q. K. Y. et al. (2009). "Tumor suppressor effect of follistatin-like 1 in ovarian and endometrial carcinogenesis: a differential expression and functional analysis". In: *Carcinogenesis* 30.1, pp. 114–21. DOI: 10.1093/carcin/bgn215.

Cheng, F., L. Su, and C. Qian (2016). "Circulating tumor DNA: a promising biomarker in the liquid biopsy of cancer". In: *Oncotarget* 7.30, pp. 48832–48841. DOI: 10.18632/oncotarget.9453.

Chiu, T. W. et al. (2006). "Plasma cell-free DNA as an indicator of severity of injury in burn patients". In: *Clin Chem Lab Med* 44.1, pp. 13–7. DOI: 10.1515/CCLM.2006.003.

Chudasama, D. et al. (2017a). "Detection of Circulating Tumour Cells and Survival of Patients with Non-small Cell Lung Cancer". In: *Anticancer Res* 37.1, pp. 169–173. DOI: 10.21873/anticanres.11302.

Chudasama, D. et al. (2017b). "Identification of novel cancer biomarkers of prognostic value using specific gene regulatory networks (GRN): a novel role of RAD51AP1 for ovarian and lung cancers". In: *Carcinogenesis*. DOI: 10.1093/carcin/bgx122.

Chudasama, D. Y. et al. (2016). "Inertia based microfluidic capture and characterisation of circulating tumour cells for the diagnosis of lung cancer". In: *Ann Transl Med* 4.24, p. 480. DOI: 10.21037/atm.2016.12.28.

Clinicaltrials.gov (2018). "https://clinicaltrials.gov". In: *31 Jan 2018*.

Consonni, D. et al. (2015). "Lung cancer prognosis before and after recurrence in a population-based setting". In: *J Natl Cancer Inst* 107.6, djv059. DOI: 10.1093/jnci/djv059.

Cristofanilli, M. et al. (2004). "Circulating tumor cells, disease progression, and survival in metastatic breast cancer". In: *N Engl J Med* 351.8, pp. 781–91. DOI: 10.1056/NEJMoa040766.

CRUK (2017). "http://www.cancerresearchuk.org/health-professional/cancer-statistics". In: *6 Dec 2017*.

Dalum, G. van, L. Holland, and L. W. Terstappen (2012). "Metastasis and Circulating Tumor Cells". In: *EJIFCC* 23.3, pp. 87–97.

Dawson, S.-J. et al. (2013). "Analysis of circulating tumor DNA to monitor metastatic breast cancer". In: *N Engl J Med* 368.13, pp. 1199–209. DOI: 10.1056/NEJMoa1213261.

Dela Cruz, C. S., L. T. Tanoue, and R. A. Matthay (2011). "Lung cancer: epidemiology, etiology, and prevention". In: *Clin Chest Med* 32.4, pp. 605–44. DOI: 10.1016/j.ccm.2011.09.001.

Deligezer, U. et al. (2006). "Circulating fragmented nucleosomal DNA and caspase-3 mRNA in patients with lymphoma and myeloma". In: *Exp Mol Pathol* 80.1, pp. 72–6. DOI: 10.1016/j.yexmp.2005.05.001.

Dennin, R. H. (1979). "DNA of free and complexed origin in human plasma: concentration and length distribution". In: *Klin Wochenschr* 57.9, pp. 451–6.

Dent, B. M. et al. (2016). "High-resolution imaging for the detection and characterisation of circulating tumour cells from patients with oesophageal, hepatocellular, thyroid and ovarian cancers". In: *Int J Cancer* 138.1, pp. 206–16. DOI: 10.1002/ijc.29680.

Diehl, F. et al. (2008). "Circulating mutant DNA to assess tumor dynamics". In: *Nat Med* 14.9, pp. 985–90. DOI: 10.1038/nm.1789.

Edwards, H. G. M., P. S. Middleton, and M. D. Hargreaves (2009). "Romano-British wall paintings: Raman spectroscopic analysis of fragments from two urban sites of early military colonisation". In: *Spectrochim Acta A Mol Biomol Spectrosc* 73.3, pp. 553–60. DOI: 10.1016/j.saa.2008.10.027.

Eliasson, C, N. A. Macleod, and P Matousek (2007). "Noninvasive detection of concealed liquid explosives using Raman spectroscopy". In: *Anal Chem* 79.21, pp. 8185–9. DOI: 10.1021/ac071383n.

Escobar, P. F. et al. (2004). "Diagnostic efficacy of optical coherence tomography in the management of preinvasive and invasive cancer of uterine cervix and vulva". In: *Int J Gynecol Cancer* 14.3, pp. 470–4. DOI: 10.1111/j.1048-891x.2004.14307.x.

Fan, H. C. et al. (2008). "Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood". In: *Proc Natl Acad Sci U S A* 105.42, pp. 16266–71. DOI: 10.1073/pnas.0808319105.

Fang, J., C.-C. Sun, and C. Gong (2016). "Long noncoding RNA XIST acts as an oncogene in non-small cell lung cancer by epigenetically repressing KLF2 expression". In: *Biochem Biophys Res Commun* 478.2, pp. 811–7. DOI: 10.1016/j.bbrc.2016.08.030.

Farace, F et al. (2011). "A direct comparison of CellSearch and ISET for circulating tumour-cell detection in patients with metastatic carcinomas". In: *Br J Cancer* 105.6, pp. 847–53. DOI: 10.1038/bjc.2011.294.

Feng, S. et al. (2015). "Label-free surface-enhanced Raman spectroscopy for detection of colorectal cancer and precursor lesions using blood plasma". In: *Biomed Opt Express* 6.9, pp. 3494–502. DOI: 10.1364/BOE.6.003494.

Fernandez-Cuesta, L. et al. (2016). "Identification of Circulating Tumor DNA for the Early Detection of Small-cell Lung Cancer". In: *EBioMedicine* 10, pp. 117–23. DOI: 10.1016/j.ebiom.2016.06.032.

Fidler, I. J. (1978). "Tumor heterogeneity and the biology of cancer invasion and metastasis". In: *Cancer Res* 38.9, pp. 2651–60.

— (2003). "The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited". In: *Nat Rev Cancer* 3.6, pp. 453–8. DOI: 10.1038/nrc1098.

Fiedor, J. and K. Burda (2014). "Potential role of carotenoids as antioxidants in human health and disease". In: *Nutrients* 6.2, pp. 466–88. DOI: 10.3390/nu6020466.

Fiedor, J. et al. (2005). "Cyclic endoperoxides of beta-carotene, potential pro-oxidants, as products of chemical quenching of singlet oxygen". In: *Biochim Biophys Acta* 1709.1, pp. 1–4. DOI: 10.1016/j.bbabio.2005.05.008.

Fisher, R, L Pusztai, and C Swanton (2013). "Cancer heterogeneity: implications for targeted therapeutics". In: *Br J Cancer* 108.3, pp. 479–85. DOI: 10.1038/bjc.2012.581.

Folga, A. et al. (2012). "Simultaneous predictive value of NT-proBNP and CA-125 in patients newly diagnosed with advanced heart failure: preliminary results". In: *Arch Med Sci* 8.4, pp. 637–43. DOI: 10.5114/aoms.2012.30287.

Foster, H. et al. (2010). "Differential expression of mTOR signalling components in drug resistance in ovarian cancer". In: *Anticancer Res* 30.9, pp. 3529–34.

Foster, H. A. et al. (2014). "The human myometrium differentially expresses mTOR signalling components before and during pregnancy: evidence for regulation by progesterone". In: *J Steroid Biochem Mol Biol* 139, pp. 166–72. DOI: 10.1016/j.jsbmb.2013.02.017.

Fournié, G. J. et al. (1995). "Plasma DNA as a marker of cancerous cell death. Investigations in patients suffering from lung cancer and in nude mice bearing human tumours". In: *Cancer Lett* 91.2, pp. 221–7.

Freidin, M. B. et al. (2015). "Circulating tumor DNA outperforms circulating tumor cells for KRAS mutation detection in thoracic malignancies". In: *Clin Chem* 61.10, pp. 1299–304. DOI: 10.1373/clinchem.2015.242453.

Friedlander, M. et al. (2010). "A Phase II, open-label study evaluating pazopanib in patients with recurrent ovarian cancer". In: *Gynecol Oncol* 119.1, pp. 32–7. DOI: 10.1016/j.ygyno.2010.05.033.

Friedlander, P. et al. (2017). "Whole-blood RNA transcript-based models can predict clinical response in two large independent clinical studies of patients with advanced melanoma treated with the checkpoint inhibitor, tremelimumab". In: *J Immunother Cancer* 5.1, p. 67. DOI: 10.1186/s40425-017-0272-z.

Friedman, R. (2016). "Drug resistance in cancer: molecular evolution and compensatory proliferation". In: *Oncotarget* 7.11, pp. 11746–55. DOI: 10.18632/oncotarget.7459.

Gay, L., A.-M. Baker, and T. A. Graham (2016). "Tumour Cell Heterogeneity". In: *F1000Res* 5. DOI: 10.12688/f1000research.7210.1.

George, S. H. L., R. Garcia, and B. M. Slomovitz (2016). "Ovarian Cancer: The Fallopian Tube as the Site of Origin and Opportunities for Prevention". In: *Front Oncol* 6, p. 108. DOI: 10.3389/fonc.2016.00108.

Giacona, M. B. et al. (1998). "Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls". In: *Pancreas* 17.1, pp. 89–97.

Giannakouros, P. et al. (2015). "MUC16 mucin (CA125) regulates the formation of multicellular aggregates by altering -catenin signaling". In: *Am J Cancer Res* 5.1, pp. 219–30.

Gold, B. et al. (2015). "Do circulating tumor cells, exosomes, and circulating tumor nucleic acids have clinical utility? A report of the association for molecular pathology". In: *J Mol Diagn* 17.3, pp. 209–24. DOI: 10.1016/j.jmoldx.2015.02.001.

Goldstraw, P. et al. (2007). "The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours". In: *J Thorac Oncol* 2.8, pp. 706–14. DOI: 10.1097/JTO.0b013e31812f3c1a.

Gorges, T. M. et al. (2012). "Circulating tumour cells escape from EpCAM-based detection due to epithelial-to-mesenchymal transition". In: *BMC Cancer* 12, p. 178. DOI: 10.1186/1471-2407-12-178.

Graham, L. J. et al. (2014). "Current approaches and challenges in monitoring treatment responses in breast cancer". In: *J Cancer* 5.1, pp. 58–68. DOI: 10.7150/jca.7047.

Griffen, J., A. Owen, and P. Matousek (2015). "Comprehensive quantification of tablets with multiple active pharmaceutical ingredients using transmission Raman spectroscopy–a proof of concept study". In: *J Pharm Biomed Anal* 115, pp. 277–82. DOI: 10.1016/j.jpba.2015.07.019.

Gupta, D. and C. G. Lis (2009). "Role of CA125 in predicting ovarian cancer survival - a review of the epidemiological literature". In: *J Ovarian Res* 2, p. 13. DOI: 10.1186/1757-2215-2-13.

Hamilton, G. and B. Rath (2016). "Detection of circulating tumor cells in non-small cell lung cancer". In: *J Thorac Dis* 8.6, pp. 1024–8. DOI: 10.21037/jtd.2016.03.86.

Han, X., J. Wang, and Y. Sun (2017). "Circulating Tumor DNA as Biomarkers for Cancer Detection". In: *Genomics Proteomics Bioinformatics* 15.2, pp. 59–72. DOI: 10.1016/j.gpb.2016.12.004.

Hanley, R. et al. (2006). "DNA integrity assay: a plasma-based screening tool for the detection of prostate cancer". In: *Clin Cancer Res* 12.15, pp. 4569–74. DOI: 10.1158/1078-0432.CCR-06-0130.

Harrow, J. et al. (2012). "GENCODE: the reference human genome annotation for The ENCODE Project". In: *Genome Res* 22.9, pp. 1760–74. DOI: 10.1101/gr.135350.111.

Hashimoto, M. et al. (2014). "Significant increase in circulating tumour cells in pulmonary venous blood during surgical manipulation in patients with primary lung cancer". In: *Interact Cardiovasc Thorac Surg* 18.6, pp. 775–83. DOI: 10.1093/icvts/ivu048.

Hayashi, N. et al. (1999). "No-touch isolation technique reduces intraoperative shedding of tumor cells into the portal vein during resection of colorectal cancer". In: *Surgery* 125.4, pp. 369–74.

Heerboth, S. et al. (2015). "EMT and tumor metastasis". In: *Clin Transl Med* 4, p. 6. DOI: 10.1186/s40169-015-0048-3.

Heim, D. et al. (2014). "Cancer beyond organ and tissue specificity: next-generation-sequencing gene mutation data reveal complex genetic similarities across major cancers". In: *Int J Cancer* 135.10, pp. 2362–9. DOI: 10.1002/ijc.28882.

Hellner, K. et al. (2016). "Premalignant SOX2 overexpression in the fallopian tubes of ovarian cancer patients: Discovery and validation studies". In: *EBioMedicine* 10, pp. 137–49. DOI: 10.1016/j.ebiom.2016.06.048.

Herbst, R. S., J. V. Heymach, and S. M. Lippman (2008). "Lung cancer". In: *N Engl J Med* 359.13, pp. 1367–80. DOI: 10.1056/NEJMra0802714.

Herreros-Villanueva, M et al. (2013). "SOX2 promotes dedifferentiation and imparts stem cell-like features to pancreatic cancer cells". In: *Oncogenesis* 2, e61. DOI: 10.1038/oncsis.2013.23.

Hiltermann, T. J. N., A. J. van der Wekken, and H. J. M. Groen (2012). "Moving forward with circulating tumor cells and lung cancer". In: *J Thorac Dis* 4.5, pp. 440–1. DOI: 10.3978/j.issn.2072-1439.2012.08.08.

Hiltermann, T. J. N. et al. (2012). "Circulating tumor cells in small-cell lung cancer: a predictive and prognostic factor". In: *Ann Oncol* 23.11, pp. 2937–42. DOI: 10.1093/annonc/mds138.

Hirose, T. et al. (2012). "Relationship of circulating tumor cells to the effectiveness of cytotoxic chemotherapy in patients with metastatic non-small-cell lung cancer". In: *Oncol Res* 20.2-3, pp. 131–7.

Hoffman, P. C., A. M. Mauer, and E. E. Vokes (2000). "Lung cancer". In: *Lancet* 355.9202, pp. 479–85. DOI: 10.1016/S0140-6736(00)82038-3.

Holdenrieder, S. et al. (2008a). "Clinical relevance of circulating nucleosomes in cancer". In: *Ann N Y Acad Sci* 1137, pp. 180–9. DOI: 10.1196/annals.1448.012.

Holdenrieder, S. et al. (2008b). "DNA integrity in plasma and serum of patients with malignant and benign diseases". In: *Ann N Y Acad Sci* 1137, pp. 162–70. DOI: 10.1196/annals.1448.013.

Holliday, S. F. et al. (2014). "Interpatient variability in dexmedetomidine response: a survey of the literature". In: *ScientificWorldJournal* 2014, p. 805013. DOI: 10.1155/2014/805013.

Hong, Y., F. Fang, and Q. Zhang (2016). "Circulating tumor cell clusters: What we know and what we expect (Review)". In: *Int J Oncol* 49.6, pp. 2206–2216. DOI: 10.3892/ijo.2016.3747.

Hou, H. W. et al. (2013). "Isolation and retrieval of circulating tumor cells using centrifugal forces". In: *Sci Rep* 3, p. 1259. DOI: 10.1038/srep01259.

Hou, J.-M. et al. (2011). "Circulating tumor cells as a window on metastasis biology in lung cancer". In: *Am J Pathol* 178.3, pp. 989–96. DOI: 10.1016/j.ajpath.2010.12.003.

Hou, J.-M. et al. (2012). "Clinical significance and molecular characteristics of circulating tumor cells and circulating tumor microemboli in patients with small-cell lung cancer". In: *J Clin Oncol* 30.5, pp. 525–32. DOI: 10.1200/JCO.2010.33.3716.

Housman, G. et al. (2014). "Drug resistance in cancer: an overview". In: *Cancers (Basel)* 6.3, pp. 1769–92. DOI: 10.3390/cancers6031769.

Huang, A. et al. (2016). "Plasma Circulating Cell-free DNA Integrity as a Promising Biomarker for Diagnosis and Surveillance in Patients with Hepatocellular Carcinoma". In: *J Cancer* 7.13, pp. 1798–1803. DOI: 10.7150/jca.15618.

Huang, A. et al. (2017). "Circumventing intratumoral heterogeneity to identify potential therapeutic targets in hepatocellular carcinoma". In: *J Hepatol* 67.2, pp. 293–301. DOI: 10.1016/j.jhep.2017.03.005.

Huang, C.-C. and C. Isidoro (2017). "Raman Spectrometric Detection Methods for Early and Non-Invasive Diagnosis of Alzheimer's Disease". In: *J Alzheimers Dis* 57.4, pp. 1145–1156. DOI: 10.3233/JAD-161238.

Huber, M. A., N. Kraut, and H. Beug (2005). "Molecular requirements for epithelial-mesenchymal transition during tumor progression". In: *Curr Opin Cell Biol* 17.5, pp. 548–58. DOI: 10.1016/j.ceb.2005.08.001.

Hyun, K.-A. et al. (2016). "Epithelial-to-mesenchymal transition leads to loss of EpCAM and different physical properties in circulating tumor cells from metastatic breast cancer". In: *Oncotarget* 7.17, pp. 24677–87. DOI: 10.18632/oncotarget.8250.

Ilie, M. et al. (2014). "'"Sentinel" circulating tumor cells allow early diagnosis of lung cancer in patients with chronic obstructive pulmonary disease". In: *PLoS One* 9.10, e111597. DOI: 10.1371/journal.pone.0111597.

Imrich, S., M. Hachmeister, and O. Gires (2012). "EpCAM and its potential role in tumor-initiating cells". In: *Cell Adh Migr* 6.1, pp. 30–8. DOI: 10.4161/cam.18953.

Jahr, S et al. (2001). "DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells". In: *Cancer Res* 61.4, pp. 1659–65.

Janiszewska, M. et al. (2015). "In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer". In: *Nat Genet* 47.10, pp. 1212–9. DOI: 10.1038/ng.3391.

Jara, L. et al. (2017). "Mutations in BRCA1, BRCA2 and other breast and ovarian cancer susceptibility genes in Central and South American populations". In: *Biol Res* 50.1, p. 35. DOI: 10.1186/s40659-017-0139-2.

Jayanthi, J. L. et al. (2011). "Diffuse reflectance spectroscopy: diagnostic accuracy of a non-invasive screening technique for early detection of malignant changes in the oral cavity". In: *BMJ Open* 1.1, e000071. DOI: 10.1136/bmjopen-2011-000071.

Jiang, W.-W. et al. (2006). "Increased plasma DNA integrity index in head and neck cancer patients". In: *Int J Cancer* 119.11, pp. 2673–6. DOI: 10.1002/ijc.22250.

Johns, L. E. et al. (2017). "Effect of population breast screening on breast cancer mortality up to 2005 in England and Wales: an individual-level cohort study". In: *Br J Cancer* 116.2, pp. 246–252. DOI: 10.1038/bjc.2016.415.

Johnson, C. C. et al. (2008). "The epidemiology of CA-125 in women without evidence of ovarian cancer in the Prostate, Lung, Colorectal and Ovarian Cancer (PLCO) Screening Trial". In: *Gynecol Oncol* 110.3, pp. 383–9. DOI: 10.1016/j.ygyno.2008.05.006.

Joosse, S. A., T. M. Gorges, and K. Pantel (2015). "Biology, detection, and clinical implications of circulating tumor cells". In: *EMBO Mol Med* 7.1, pp. 1–11. DOI: 10.15252/emmm.201303698.

Jordan, S. D. et al. (2003). "A retrospective evaluation of the feasibility of intrapatient dose escalation as appropriate methodology for dose-ranging studies for combination cytotoxic regimens". In: *Cancer Chemother Pharmacol* 52.2, pp. 113–8. DOI: 10.1007/s00280-003-0634-8.

Kallergi, G. et al. (2011). "Epithelial to mesenchymal transition markers expressed in circulating tumour cells of early and metastatic breast cancer patients". In: *Breast Cancer Res* 13.3, R59. DOI: 10.1186/bcr2896.

Kalluri, R. and R. A. Weinberg (2009). "The basics of epithelial-mesenchymal transition". In: *J Clin Invest* 119.6, pp. 1420–8. DOI: 10.1172/JCI39104.

Kamel, A. M. et al. (2016). "Plasma DNA integrity index as a potential molecular diagnostic marker for breast cancer". In: *Tumour Biol* 37.6, pp. 7565–72. DOI: 10.1007/s13277-015-4624-3.

Kamp, D. W. (2009). "Asbestos-induced lung diseases: an update". In: *Transl Res* 153.4, pp. 143–52. DOI: 10.1016/j.trsl.2009.01.004.

Kaplan-Meier Plotter (2018). "http://kmplot.com/analysis/index.php". In: *31 Jan 2018*.

Kattan, K al et al. (1997). "Disease recurrence after resection for stage I lung cancer". In: *Eur J Cardiothorac Surg* 12.3, pp. 380–4.

Kaur, A. et al. (2016). "Human C1q Induces Apoptosis in an Ovarian Cancer Cell Line via Tumor Necrosis Factor Pathway". In: *Front Immunol* 7, p. 599. DOI: 10.3389/fimmu.2016.00599.

Kaur, J. and J. Poole (2017). "Cancer diagnoses and age-standardised incidence rates for all cancer sites by age, sex and region." In: *Cancer Registration Statistics, England*, pp. 1–23.

Kawamura, M. T., M. E. Paschoal, and M. d. G. Carvalho (2000). "Profile of proteins complexed with circulating DNA of a lung cancer patient". In: *Ann N Y Acad Sci* 906, pp. 51–4.

Kelley, M. J. and D. C. McCrory (2003). "Prevention of lung cancer: summary of published evidence". In: *Chest* 123.1 Suppl, 50S–59S.

Kim, D. et al. (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions". In: *Genome Biol* 14.4, R36. DOI: 10.1186/gb-2013-14-4-r36.

Kim, K.-T. et al. (2015). "Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells". In: *Genome Biol* 16, p. 127. DOI: 10.1186/s13059-015-0692-3.

Klein, C. A. (2013). "Selection and adaptation during metastatic cancer progression". In: *Nature* 501.7467, pp. 365–72. DOI: 10.1038/nature12628.

Knappskog, S. et al. (2014). "Population distribution and ancestry of the cancer protective MDM2 SNP285 (rs117039649)". In: *Oncotarget* 5.18, pp. 8223–34. DOI: 10.18632/oncotarget.1910.

Knoji.com (2017). "https://lung-conditions.knoji.com/lung-cancer-its-risk-in-related-careers-and-environment/". In: *6 Dec 2017*.

Kolostova, K. et al. (2016). "Molecular characterization of circulating tumor cells in ovarian cancer". In: *Am J Cancer Res* 6.5, pp. 973–80.

Krebs, M. G. et al. (2011). "Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer". In: *J Clin Oncol* 29.12, pp. 1556–63. DOI: 10.1200/JCO.2010.28.7045.

Krzeszinski, J. Y. and Y. Wan (2015). "New therapeutic targets for cancer bone metastasis". In: *Trends Pharmacol Sci* 36.6, pp. 360–73. DOI: 10.1016/j.tips.2015.04.006.

Kurokawa, T. et al. (2017). "Antisense RNA transcripts in the blood may be novel diagnostic markers for colorectal cancer". In: *Oncol Lett* 14.3, pp. 3487–3493. DOI: 10.3892/ol.2017.6572.

Lam, N. Y. L. et al. (2003). "Time course of early and late changes in plasma DNA in trauma patients". In: *Clin Chem* 49.8, pp. 1286–91.

Lavarino, C. et al. (2009). "Specific gene expression profiles and chromosomal abnormalities are associated with infant disseminated neuroblastoma". In: *BMC Cancer* 9, p. 44. DOI: 10.1186/1471-2407-9-44.

Le Calvez-Kelm, F. et al. (2012). "RAD51 and breast cancer susceptibility: no evidence for rare variant association in the Breast Cancer Family Registry study". In: *PLoS One* 7.12, e52374. DOI: 10.1371/journal.pone.0052374.

Lecharpentier, A et al. (2011). "Detection of circulating tumour cells with a hybrid (epithelial/mesenchymal) phenotype in patients with metastatic non-small cell lung cancer". In: *Br J Cancer* 105.9, pp. 1338–41. DOI: 10.1038/bjc.2011.405.

Lehner, J. et al. (2013). "Plasma DNA integrity indicates response to neoadjuvant chemotherapy in patients with locally confined breast cancer". In: *Int J Clin Pharmacol Ther* 51.1, pp. 59–62.

Lengyel, E. (2010). "Ovarian cancer development and metastasis". In: *Am J Pathol* 177.3, pp. 1053–64. DOI: 10.2353/ajpath.2010.100105.

Li, H., J. Zeng, and K. Shen (2014). "PI3K/AKT/mTOR signaling pathway as a therapeutic target for ovarian cancer". In: *Arch Gynecol Obstet* 290.6, pp. 1067–78. DOI: 10.1007/s00404-014-3377-3.

Li, H. et al. (2009). "The Sequence Alignment/Map format and SAMtools". In: *Bioinformatics* 25.16, pp. 2078–9. DOI: 10.1093/bioinformatics/btp352.

Li, L. et al. (2004). "Correlation of serum VEGF levels with clinical stage, therapy efficacy, tumor metastasis and patient survival in ovarian cancer". In: *Anticancer Res* 24.3b, pp. 1973–9.

Liu, J. F. et al. (2013). "Predictive value of circulating tumor cells (CTCs) in newly-diagnosed and recurrent ovarian cancer patients". In: *Gynecol Oncol* 131.2, pp. 352–6. DOI: 10.1016/j.ygyno.2013.08.006.

Lo, Y. M. et al. (1999). "Rapid clearance of fetal DNA from maternal plasma". In: *Am J Hum Genet* 64.1, pp. 218–24. DOI: 10.1086/302205.

Luvero, D., A. Milani, and J. A. Ledermann (2014). "Treatment options in recurrent ovarian cancer: latest evidence and clinical potential". In: *Ther Adv Med Oncol* 6.5, pp. 229–39. DOI: 10.1177/1758834014544121.

Ma, L. et al. (2017). "Long non-coding RNA XIST promotes cell growth and invasion through regulating miR-497/MACC1 axis in gastric cancer". In: *Oncotarget* 8.3, pp. 4125–4135. DOI: 10.18632/oncotarget.13670.

Maheswaran, S. et al. (2008). "Detection of mutations in EGFR in circulating lung-cancer cells". In: *N Engl J Med* 359.4, pp. 366–77. DOI: 10.1056/NEJMoa0800668.

Mandel, P and P Metais (1948). "Les acides nucleiques du plasma sanguin chez l'homme". In: *C R Seances Soc Biol Fil* 142.3-4, pp. 241–3.

Marrinucci, D. et al. (2012). "Fluid biopsy in patients with metastatic prostate, pancreatic and breast cancers". In: *Phys Biol* 9.1, p. 016003. DOI: 10.1088/1478-3975/9/1/016003.

Matsutani, N. et al. (2017). "Does lung cancer surgery cause circulating tumor cells?-A multicenter, prospective study". In: *J Thorac Dis* 9.8, pp. 2419–2426. DOI: 10.21037/jtd.2017.07.33.

Mazurenka, M et al. (2017). "Development of a combined OCT-Raman probe for the prospective in vivo clinical melanoma skin cancer screening". In: *Rev Sci Instrum* 88.10, p. 105103. DOI: 10.1063/1.5004999.

Medicalexpress (2017). "https://medicalxpress.com/news/2016-06-liquid-biopsy-techniques-edge-cancer.html". In: *29 Dec 2017*.

Medipally, D. K. R. et al. (2017). "Development of a high throughput (HT) Raman spectroscopy method for rapid screening of liquid blood plasma from prostate cancer patients". In: *Analyst* 142.8, pp. 1216–1226. DOI: 10.1039/c6an02100j.

Menon, U. et al. (2015). "Risk Algorithm Using Serial Biomarker Measurements Doubles the Number of Screen-Detected Cancers Compared With a Single-Threshold Rule in the United Kingdom Collaborative Trial of Ovarian Cancer Screening". In: *J Clin Oncol* 33.18, pp. 2062–71. DOI: 10.1200/JCO.2014.59.4945.

Midthun, D. E. (2016). "Early detection of lung cancer". In: *F1000Res* 5. DOI: 10.12688/f1000research.7313.1.

Miles, G. D. et al. (2012). "Identifying microRNA/mRNA dysregulations in ovarian cancer". In: *BMC Res Notes* 5, p. 164. DOI: 10.1186/1756-0500-5-164.

Miller, M. C., G. V. Doyle, and L. W.M. M. Terstappen (2010). "Significance of Circulating Tumor Cells Detected by the CellSearch System in Patients with Metastatic Breast Colorectal and Prostate Cancer". In: *J Oncol* 2010, p. 617421. DOI: 10.1155/2010/617421.

Mitsudomi, T., K. Suda, and Y. Yatabe (2013). "Surgery for NSCLC in the era of personalized medicine". In: *Nat Rev Clin Oncol* 10.4, pp. 235–44. DOI: 10.1038/nrclinonc.2013.22.

Miwa, S. et al. (2015). "Heterogeneous cell-cycle behavior in response to UVB irradiation by a population of single cancer cells visualized by time-lapse FUCCI imaging". In: *Cell Cycle* 14.12, pp. 1932–7. DOI: 10.1080/15384101.2015.1033598.

Modesti, M. et al. (2007). "RAD51AP1 is a structure-specific DNA binding protein that stimulates joint molecule formation during RAD51-mediated homologous recombination". In: *Mol Cell* 28.3, pp. 468–81. DOI: 10.1016/j.molcel.2007.08.025.

Mogensen, M. et al. (2009). "Assessment of optical coherence tomography imaging in the diagnosis of non-melanoma skin cancer and benign lesions versus normal skin: observer-blinded evaluation by dermatologists and pathologists". In: *Dermatol Surg* 35.6, pp. 965–72. DOI: 10.1111/j.1524-4725.2009.01164.x.

Mol, B. W. et al. (1998). "The performance of CA-125 measurement in the detection of endometriosis: a meta-analysis". In: *Fertil Steril* 70.6, pp. 1101–8.

Molino, A et al. (1991). "A comparative analysis of three different techniques for the detection of breast cancer cells in bone marrow". In: *Cancer* 67.4, pp. 1033–6.

Moore, R. G. et al. (2008). "Utility of a novel serum tumor biomarker HE4 in patients with endometrioid adenocarcinoma of the uterus". In: *Gynecol Oncol* 110.2, pp. 196–201. DOI: 10.1016/j.ygyno.2008.04.002.

Mparmpakas, D. et al. (2010). "Expression of mTOR and downstream signalling components in the JEG-3 and BeWo human placental choriocarcinoma cell lines". In: *Int J Mol Med* 25.1, pp. 65–9.

Mparmpakas, D. et al. (2012). "Placental DEPTOR as a stress sensor during pregnancy". In: *Clin Sci (Lond)* 122.7, pp. 349–59. DOI: 10.1042/CS20110378.

Munzone, E. et al. (2012). "Prognostic value of circulating tumor cells according to immunohistochemically defined molecular subtypes in advanced breast cancer". In: *Clin Breast Cancer* 12.5, pp. 340–6. DOI: 10.1016/j.clbc.2012.07.001.

MyCancerGenome (2017). "https://www.mycancergenome.org/content/disease/lung-cancer/". In: *29 Dec 2017*.

Nagata, S (2000). "Apoptotic DNA fragmentation". In: *Exp Cell Res* 256.1, pp. 12–8. DOI: 10.1006/excr.2000.4834.

Nagrath, S. et al. (2007). "Isolation of rare circulating tumour cells in cancer patients by microchip technology". In: *Nature* 450.7173, pp. 1235–9. DOI: 10.1038/nature06385.

NICE (2017). "https://www.nice.org.uk". In: *29 Dec 2017*.

Nishino, H et al. (2000). "Cancer prevention by natural carotenoids". In: *Biofactors* 13.1-4, pp. 89–94.

Normanno, N. et al. (2016). "The prognostic role of circulating tumor cells in lung cancer". In: *Expert Rev Anticancer Ther* 16.8, pp. 859–67. DOI: 10.1080/14737140.2016.1202767.

Nurwidya, F. et al. (2016). "Circulating Tumor Cell and Cell-free Circulating Tumor DNA in Lung Cancer". In: *Chonnam Med J* 52.3, pp. 151–8. DOI: 10.4068/cmj.2016.52.3.151.

Obama, K. et al. (2008). "Enhanced expression of RAD51 associating protein-1 is involved in the growth of intrahepatic cholangiocarcinoma cells". In: *Clin Cancer Res* 14.5, pp. 1333–9. DOI: 10.1158/1078-0432.CCR-07-1381.

O'Donnell, P. H. and M. E. Dolan (2009). "Cancer pharmacoethnicity: ethnic differences in susceptibility to the effects of chemotherapy". In: *Clin Cancer Res* 15.15, pp. 4806–14. DOI: 10.1158/1078-0432.CCR-09-0344.

Ogle, L. F. et al. (2016). "Imagestream detection and characterisation of circulating tumour cells - A liquid biopsy for hepatocellular carcinoma?" In: *J Hepatol* 65.2, pp. 305–13. DOI: 10.1016/j.jhep.2016.04.014.

Olmos, D. et al. (2012). "Prognostic value of blood mRNA expression signatures in castration-resistant prostate cancer: a prospective, two-stage study". In: *Lancet Oncol* 13.11, pp. 1114–24. DOI: 10.1016/S1470-2045(12)70372-8.

Oncomine (2018). "https://www.oncomine.org". In: *31 Jan 2018*.

Oshlack, A., M. D. Robinson, and M. D. Young (2010). "From RNA-seq reads to differential expression results". In: *Genome Biol* 11.12, p. 220. DOI: 10.1186/gb-2010-11-12-220.

Pantel, K., R. H. Brakenhoff, and B. Brandt (2008). "Detection, clinical relevance and specific biological properties of disseminating tumour cells". In: *Nat Rev Cancer* 8.5, pp. 329–40. DOI: 10.1038/nrc2375.

Pares, R. D. and M. I. Whitecross (1982). "Gold-labelled antibody decoration (GLAD) in the diagnosis of plant viruses by immuno-electron microscopy". In: *J Immunol Methods* 51.1, pp. 23–8.

Park, J.-M. et al. (2012). "Highly efficient assay of circulating tumor cells by selective sedimentation with a density gradient medium and microfiltration from whole blood". In: *Anal Chem* 84.17, pp. 7400–7. DOI: 10.1021/ac3011704.

Parplys, A. C. et al. (2014). "RAD51AP1-deficiency in vertebrate cells impairs DNA replication". In: *DNA Repair (Amst)* 24, pp. 87–97. DOI: 10.1016/j.dnarep.2014.09.007.

Patel, A. P. et al. (2014). "Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma". In: *Science* 344.6190, pp. 1396–401. DOI: 10.1126/science.1254257.

Patel, K. M. and D. W. Y. Tsui (2015). "The translational potential of circulating tumour DNA in oncology". In: *Clin Biochem* 48.15, pp. 957–61. DOI: 10.1016/j.clinbiochem.2015.04.005.

Pathania, R. et al. (2016). "Combined Inhibition of DNMT and HDAC Blocks the Tumorigenicity of Cancer Stem-like Cells and Attenuates Mammary Tumor Growth". In: *Cancer Res* 76.11, pp. 3224–35. DOI: 10.1158/0008-5472.CAN-15-2249.

Pearl, M. L. et al. (2014). "Prognostic analysis of invasive circulating tumor cells (iCTCs) in epithelial ovarian cancer". In: *Gynecol Oncol* 134.3, pp. 581–90. DOI: 10.1016/j.ygyno.2014.06.013.

Phillips, K. G. et al. (2012). "Optical quantification of cellular mass, volume, and density of circulating tumor cells identified in an ovarian cancer patient". In: *Front Oncol* 2, p. 72. DOI: 10.3389/fonc.2012.00072.

Polyak, K. and R. A. Weinberg (2009). "Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits". In: *Nat Rev Cancer* 9.4, pp. 265–73. DOI: 10.1038/nrc2620.

Poole, J. and A. Nordin (2012). "Overview of Ovarian Cancer in England: Incidence, Mortality and Survival". In: *Trent Cancer Registry - National Cancer Intelligence Network*, pp. 1–59.

Poveda, A. et al. (2011). "Circulating tumor cells predict progression free survival and overall survival in patients with relapsed/recurrent advanced ovarian cancer". In: *Gynecol Oncol* 122.3, pp. 567–72. DOI: 10.1016/j.ygyno.2011.05.028.

Punnoose, E. A. et al. (2012). "Evaluation of circulating tumor cells and circulating tumor DNA in non-small cell lung cancer: association with clinical endpoints in a phase II clinical trial of pertuzumab and erlotinib". In: *Clin Cancer Res* 18.8, pp. 2391–401. DOI: 10.1158/1078-0432.CCR-11-3148.

Raimondi, C., C. Nicolazzo, and A. Gradilone (2015). "Circulating tumor cells isolation: the "post-EpCAM era"". In: *Chin J Cancer Res* 27.5, pp. 461–70. DOI: 10.3978/j.issn.1000-9604.2015.06.02.

Rainer, T. H. et al. (2006). "Plasma beta-globin DNA as a prognostic marker in chest pain patients". In: *Clin Chim Acta* 368.1-2, pp. 110–3. DOI: 10.1016/j.cca.2005.12.021.

Ramanujam, N (2000). "Fluorescence spectroscopy of neoplastic and non-neoplastic tissues". In: *Neoplasia* 2.1-2, pp. 89–117.

Ramanujam, N et al. (1994). "Fluorescence spectroscopy: a diagnostic tool for cervical intraepithelial neoplasia (CIN)". In: *Gynecol Oncol* 52.1, pp. 31–8. DOI: 10.1006/gyno.1994.1007.

Rebhan, M et al. (1997). "GeneCards: integrating information about genes, proteins and diseases". In: *Trends Genet* 13.4, p. 163.

Reinartz, S. et al. (2012). "CA125 (MUC16) gene silencing suppresses growth properties of ovarian and breast cancer cells". In: *Eur J Cancer* 48.10, pp. 1558–69. DOI: 10.1016/j.ejca.2011.07.004.

Riethdorf, S. et al. (2007). "Detection of circulating tumor cells in peripheral blood of patients with metastatic breast cancer: a validation study of the CellSearch system". In: *Clin Cancer Res* 13.3, pp. 920–8. DOI: 10.1158/1078-0432.CCR-06-1695.

Rimawi, M. F., C. De Angelis, and R. Schiff (2015). "Resistance to Anti-HER2 Therapies in Breast Cancer". In: *Am Soc Clin Oncol Educ Book*, e157–64. DOI: 10.14694/EdBook_AM.2015.35.e157.

Rodriguez-Diaz, E. et al. (2014). "Elastic scattering spectroscopy as an optical marker of inflammatory bowel disease activity and subtypes". In: *Inflamm Bowel Dis* 20.6, pp. 1029–36. DOI: 10.1097/MIB.0000000000000058.

Rogers-Broadway, K.-R. et al. (2016). "Differential effects of rapalogues, dual kinase inhibitors on human ovarian carcinoma cells in vitro". In: *Int J Oncol* 49.1, pp. 133–43. DOI: 10.3892/ijo.2016.3531.

Romero-Laorden, N. et al. (2014). "Circulating and disseminated tumor cells in ovarian cancer: a systematic review". In: *Gynecol Oncol* 133.3, pp. 632–9. DOI: 10.1016/j.ygyno.2014.03.016.

Rotger, M., C. Csajka, and A. Telenti (2006). "Genetic, ethnic, and gender differences in the pharmacokinetics of antiretroviral agents". In: *Curr HIV/AIDS Rep* 3.3, pp. 118–25.

Rump, A. et al. (2004). "Binding of ovarian cancer antigen CA125/MUC16 to mesothelin mediates cell adhesion". In: *J Biol Chem* 279.10, pp. 9190–8. DOI: 10.1074/jbc.M312372200.

Saadatpour, A. et al. (2015). "Single-Cell Analysis in Cancer Genomics". In: *Trends Genet* 31.10, pp. 576–86. DOI: 10.1016/j.tig.2015.07.003.

Sawada, K. et al. (2007). "c-Met overexpression is a prognostic factor in ovarian cancer and an effective target for inhibition of peritoneal dissemination and invasion". In: *Cancer Res* 67.4, pp. 1670–9. DOI: 10.1158/0008-5472.CAN-06-1147.

Schneck, H. et al. (2015). "EpCAM-Independent Enrichment of Circulating Tumor Cells in Metastatic Breast Cancer". In: *PLoS One* 10.12, e0144535. DOI: 10.1371/journal.pone.0144535.

Schoch, C. et al. (2005). "Acute myeloid leukemia with a complex aberrant karyotype is a distinct biological entity characterized by genomic imbalances and a specific gene expression profile". In: *Genes Chromosomes Cancer* 43.3, pp. 227–38. DOI: 10.1002/gcc.20193.

SEER (2017). "https://seer.cancer.gov/statfacts/html/ovary.html". In: *4 Dec 2017*.

Sequist, L. V. et al. (2009). "The CTC-chip: an exciting new tool to detect circulating tumor cells in lung cancer patients". In: *J Thorac Oncol* 4.3, pp. 281–3. DOI: 10.1097/JTO.0b013e3181989565.

Shalek, A. K. et al. (2013). "Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells". In: *Nature* 498.7453, pp. 236–40. DOI: 10.1038/nature12172.

Shen, J. et al. (2017). "Predictive and prognostic value of folate receptor-positive circulating tumor cells in small cell lung cancer patients treated with first-line chemotherapy". In: *Oncotarget* 8.30, pp. 49044–49052. DOI: 10.18632/oncotarget.17039.

Shibuya, M. (2011). "Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies". In: *Genes Cancer* 2.12, pp. 1097–105. DOI: 10.1177/1947601911423031.

Sienel, W. et al. (2003). "Tumour cells in the tumour draining vein of patients with non-small cell lung cancer: detection rate and clinical significance". In: *Eur J Cardiothorac Surg* 23.4, pp. 451–6.

Sikora, K. et al. (2015). "Evaluation of cell-free DNA as a biomarker for pancreatic malignancies". In: *Int J Biol Markers* 30.1, e136–41. DOI: 10.5301/jbm.5000088.

Silva Filho, B. F. da et al. (2013). "Circulating cell-free DNA in serum as a biomarker of colorectal cancer". In: *J Clin Pathol* 66.9, pp. 775–8. DOI: 10.1136/jclinpath-2013-201521.

Skates, S. J. et al. (2017). "Early Detection of Ovarian Cancer using the Risk of Ovarian Cancer Algorithm with Frequent CA125 Testing in Women at Increased Familial Risk - Combined Results from Two Screening Trials". In: *Clin Cancer Res* 23.14, pp. 3628–3637. DOI: 10.1158/1078-0432.CCR-15-2750.

Slade, M. J. et al. (2009). "Comparison of bone marrow, disseminated tumour cells and blood-circulating tumour cells in breast cancer patients after primary treatment". In: *Br J Cancer* 100.1, pp. 160–6. DOI: 10.1038/sj.bjc.6604773.

Song, L.-J. et al. (2015). "PU.1 Is Identified as a Novel Metastasis Suppressor in Hepatocellular Carcinoma Regulating the miR-615-5p/IGF2 Axis". In: *Asian Pac J Cancer Prev* 16.9, pp. 3667–71.

Spitz, M. R. et al. (2007). "A risk model for prediction of lung cancer". In: *J Natl Cancer Inst* 99.9, pp. 715–26. DOI: 10.1093/jnci/djk153.

Spitz, M. R. et al. (2008). "An expanded risk prediction model for lung cancer". In: *Cancer Prev Res (Phila)* 1.4, pp. 250–4. DOI: 10.1158/1940-6207.CAPR-08-0060.

Stewart, C. A. and R. T. Dell'Orco (1992). "Expression of proliferating cell nuclear antigen during the cell cycle of human diploid fibroblasts". In: *In Vitro Cell Dev Biol* 28A.3 Pt 1, pp. 211–4.

String-db.org (2018). "https://string-db.org/". In: *31 Jan 2018*.

Stroun, M et al. (1987). "Isolation and characterization of DNA from the plasma of cancer patients". In: *Eur J Cancer Clin Oncol* 23.6, pp. 707–12.

Su, Z. et al. (2014). "An investigation of biomarkers derived from legacy microarray data for their utility in the RNA-seq era". In: *Genome Biol* 15.12, p. 523. DOI: 10.1186/s13059-014-0523-y.

Tanaka, F. et al. (2009). "Circulating tumor cell as a diagnostic marker in primary lung cancer". In: *Clin Cancer Res* 15.22, pp. 6980–6. DOI: 10.1158/1078-0432.CCR-09-1095.

Thierry, A. R. et al. (2010). "Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts". In: *Nucleic Acids Res* 38.18, pp. 6159–75. DOI: 10.1093/nar/gkq421.

Thiery, J. P. (2002). "Epithelial-mesenchymal transitions in tumour progression". In: *Nat Rev Cancer* 2.6, pp. 442–54. DOI: 10.1038/nrc822.

Thoeny, H. C. and B. D. Ross (2010). "Predicting and monitoring cancer treatment response with diffusion-weighted MRI". In: *J Magn Reson Imaging* 32.1, pp. 2–16. DOI: 10.1002/jmri.22167.

Trapnell, C. et al. (2010). "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation". In: *Nat Biotechnol* 28.5, pp. 511–5. DOI: 10.1038/nbt.1621.

Travis, W. D. (2012). "Update on small cell carcinoma and its differentiation from squamous cell carcinoma and other non-small cell carcinomas". In: *Mod Pathol* 25 Suppl 1, S18–30. DOI: 10.1038/modpathol.2011.150.

Umetani, N. et al. (2006a). "Increased integrity of free circulating DNA in sera of patients with colorectal or periampullary cancer: direct quantitative PCR for ALU repeats". In: *Clin Chem* 52.6, pp. 1062–9. DOI: 10.1373/clinchem.2006.068577.

Umetani, N. et al. (2006b). "Prediction of breast tumor progression by integrity of free circulating DNA in serum". In: *J Clin Oncol* 24.26, pp. 4270–6. DOI: 10.1200/JCO.2006.05.9493.

Uramoto, H. and F. Tanaka (2014). "Recurrence after surgery in patients with NSCLC". In: *Transl Lung Cancer Res* 3.4, pp. 242–9. DOI: 10.3978/j.issn.2218-6751.2013.12.05.

Utomo, W. K. et al. (2016). "DNA integrity as biomarker in pancreatic cyst fluid". In: *Am J Cancer Res* 6.8, pp. 1837–41.

Valastyan, S. and R. A. Weinberg (2011). "Tumor metastasis: molecular insights and evolving paradigms". In: *Cell* 147.2, pp. 275–92. DOI: 10.1016/j.cell.2011.09.024.

Vargas, A. N. (2014). "Natural history of ovarian cancer". In: *Ecancermedicalscience* 8, p. 465. DOI: 10.3332/ecancer.2014.465.

Verma, M. (2012). "Personalized medicine and cancer". In: *J Pers Med* 2.1, pp. 1–14. DOI: 10.3390/jpm2010001.

Verveer, P. J., A Squire, and P. I. Bastiaens (2000). "Global analysis of fluorescence lifetime imaging microscopy data". In: *Biophys J* 78.4, pp. 2127–37. DOI: 10.1016/S0006-3495(00)76759-2.

Vidal, J et al. (2017). "Plasma ctDNA RAS mutation analysis for the diagnosis and treatment monitoring of metastatic colorectal cancer patients". In: *Ann Oncol* 28.6, pp. 1325–1332. DOI: 10.1093/annonc/mdx125.

Volinia, S. et al. (2008). "Genome wide identification of recessive cancer genes by combinatorial mutation analysis". In: *PLoS One* 3.10, e3380. DOI: 10.1371/journal.pone.0003380.

Volpati, D et al. (2015). "Exploring copper nanostructures as highly uniform and reproducible substrates for plasmon-enhanced fluorescence". In: *Analyst* 140.2, pp. 476–82. DOI: 10.1039/c4an00889h.

Vona, G et al. (2000). "Isolation by size of epithelial tumor cells : a new method for the immunomorphological and molecular characterization of circulatingtumor cells". In: *Am J Pathol* 156.1, pp. 57–63. DOI: 10.1016/S0002-9440(10)64706-2.

Vrbic, S et al. (2013). "Current and future anti-HER2 therapy in breast cancer". In: *J BUON* 18.1, pp. 4–16.

Wang, B. G. et al. (2003). "Increased plasma DNA integrity in cancer patients". In: *Cancer Res* 63.14, pp. 3966–8.

Wang, H. et al. (2017). "The Long Non-Coding RNA XIST Controls Non-Small Cell Lung Cancer Proliferation and Invasion by Modulating miR-186-5p". In: *Cell Physiol Biochem* 41.6, pp. 2221–2229. DOI: 10.1159/000475637.

Wang, Z., M. Gerstein, and M. Snyder (2009). "RNA-Seq: a revolutionary tool for transcriptomics". In: *Nat Rev Genet* 10.1, pp. 57–63. DOI: 10.1038/nrg2484.

Weakley, S. M. et al. (2011). "Expression and function of a large non-coding RNA gene XIST in human cancer". In: *World J Surg* 35.8, pp. 1751–6. DOI: 10.1007/s00268-010-0951-0.

Weiss, G. J. et al. (2017). "Tumor Cell-Free DNA Copy Number Instability Predicts Therapeutic Response to Immunotherapy". In: *Clin Cancer Res* 23.17, pp. 5074–5081. DOI: 10.1158/1078-0432.CCR-17-0231.

Wiese, C. et al. (2007). "Promotion of homologous recombination and genomic stability by RAD51AP1 via RAD51 recombinase enhancement". In: *Mol Cell* 28.3, pp. 482–90. DOI: 10.1016/j.molcel.2007.08.027.

Winnard Jr, P. T. et al. (2017). "Organ-specific isogenic metastatic breast cancer cell lines exhibit distinct Raman spectral signatures and metabolomes". In: *Oncotarget* 8.12, pp. 20266–20287. DOI: 10.18632/oncotarget.14865.

Wu, X.-L. et al. (2016). "Diagnostic and Prognostic Value of Circulating Tumor Cells in Head and Neck Squamous Cell Carcinoma: a systematic review and meta-analysis". In: *Sci Rep* 6, p. 20210. DOI: 10.1038/srep20210.

Xiao, Y., U. A. Yurievich, and S. V. Yosypovych (2017). "Long noncoding RNA XIST is a prognostic factor in colorectal cancer and inhibits 5-fluorouracil-induced cell cytotoxicity through promoting thymidylate synthase expression". In: *Oncotarget* 8.47, pp. 83171–83182. DOI: 10.18632/oncotarget.20487.

Xu, W. et al. (2017). "Comparison of three different methods for the detection of circulating tumor cells in mice with lung metastasis". In: *Oncol Rep* 37.6, pp. 3219–3226. DOI: 10.3892/or.2017.5613.

Yao, J. et al. (2011). "Expression of a functional VEGFR-1 in tumor cells is a major determinant of anti-PlGF antibodies efficacy". In: *Proc Natl Acad Sci U S A* 108.28, pp. 11590–5. DOI: 10.1073/pnas.1109029108.

Yoneda, J et al. (1998). "Expression of angiogenesis-related genes and progression of human ovarian carcinomas in nude mice". In: *J Natl Cancer Inst* 90.6, pp. 447–54.

Yu, H et al. (2017). "Knockdown of long non-coding RNA XIST increases blood-tumor barrier permeability and inhibits glioma angiogenesis by targeting miR-137". In: *Oncogenesis* 6.3, e303. DOI: 10.1038/oncsis.2017.7.

Yu, M. et al. (2011). "Circulating tumor cells: approaches to isolation and characterization". In: *J Cell Biol* 192.3, pp. 373–82. DOI: 10.1083/jcb.201010021.

Zeerleder, S. et al. (2003). "Elevated nucleosome levels in systemic inflammation and sepsis". In: *Crit Care Med* 31.7, pp. 1947–51. DOI: 10.1097/01.CCM.0000074719.40109.95.

Zeng, X. et al. (2011). "Lung cancer serum biomarker discovery using label-free liquid chromatography-tandem mass spectrometry". In: *J Thorac Oncol* 6.4, pp. 725–34. DOI: 10.1097/JTO.0b013e31820c312e.

Zeune, L. et al. (2017). "Quantifying HER-2 expression on circulating tumor cells by ACCEPT". In: *PLoS One* 12.10, e0186562. DOI: 10.1371/journal.pone.0186562.

Zhang, W. et al. (2015). "Comparison of RNA-seq and microarray-based models for clinical endpoint prediction". In: *Genome Biol* 16, p. 133. DOI: 10.1186/s13059-015-0694-1.

Zhang, W. et al. (2011). "Alu distribution and mutation types of cancer genes". In: *BMC Genomics* 12, p. 157. DOI: 10.1186/1471-2164-12-157.

Zhao, Y. and R. Simon (2010). "Gene expression deconvolution in clinical samples". In: *Genome Med* 2.12, p. 93. DOI: 10.1186/gm214.

Zhou, Y. et al. (2015). "Prognostic Value of Circulating Tumor Cells in Ovarian Cancer: A Meta-Analysis". In: *PLoS One* 10.6, e0130873. DOI: 10.1371/journal.pone.0130873.

Zonios, G et al. (1999). "Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo". In: *Appl Opt* 38.31, pp. 6628–37.