Use of liquid biopsies as prognostic and diagnostic marker for Non-Small Cell Lung Cancer

A thesis submitted for the degree of
Doctor of Philosophy

by

PERIKLIS KATOPODIS

College of Health and Life Sciences
BRUNEL UNIVERSITY LONDON

Abstract

Lung cancer survival remains poor in the western world due to late presentation in most cases, leading to difficulty of treatment in these advanced and metastatic patients. Therefore, the development of robust biomarkers for prognosis and monitoring treatment response and relapse would be of great benefit. Blood based biomarkers, referred to as 'liquid biopsies' are becoming increasingly popular in the cancer field, due to their non-invasive nature compared to conventional biopsies. Circulating tumour cells (CTCs) are cells that have detached from the primary tumour, enter the blood stream and can be of diagnostic or prognostic value in several cancers. Similarly, circulating free tumour DNA (ctDNA), are fragments of tumour derived DNA, that can be detected in plasma. In this study, we used liquid biopsies to identify novel biomarkers of prognostic or diagnostic value, by characterising and enumerating of CTCs, measuring of circulating cell-free DNA (ccfDNA) and inflammatory biomarkers and identification of differentially expressed genes in vitro, following silencing of a key long non-coding RNA. Following series of validation experiments using lung cancer cell lines (A549 and H1975), we have characterised CTCs from NSCLC patients using multi-spectral imaging flow cytometry (ImageStream) against multiple targets such as EpCAM, PanCK, TTF-1, CK7 and CD45 as a marker to exclude white blood cells. We have investigated the impact of surgically induced trauma using peripheral blood from pre- and post-operative non-small cell lung cancer (NSCLC) patients undergoing thoracotomy or video-assisted thoracoscopic surgery (VATS) resection. CTCs were increased in post-operative blood samples in 54 NSCLC patients. Patients that underwent thoracotomy instead of VATS had higher numbers of EpCAM (p=0.004) and PanCK-labelled (p=0.03) CTCs postoperatively (statistical test; paired t-test). ccfDNA and DNA integrity index were also significantly increased in post-operative samples (p=0.0009 and p=0.04), with concomitant increase of IL-6 and IL-10 levels in the same cohorts (p=0.0004 and p=0.034 respectively). In this part of the thesis we have shown the potential clinical utility of several biomarkers from liquid biopsies to guide peri-operative management. This is of particular importance since the long-term goals of lung cancer surgery include improved length and quality of life, and surgical intervention is not always an option. Finally, based on previous studies from laboratory, the lncRNA X-inactive specific transcript (XIST) appeared to be of diagnostic value. Here, we have made use of siRNA to silence this gene in two lung cancer cell lines (A549 and

H1975) and RNA-Seq to identify differentially expressed genes (DEGs). We have found that these genes are involved in signal transduction, energy and metabolic pathways, and cell communication. Moreover, we expanded our observations making use of in-silico tools, demonstrating an association of XIST expression associated with TSIX, hnRNPu, Bcl-2, and BRCA1. This cluster of genes collectively, has a strong discriminatory power between controls and lung cancer, implying a diagnostic potential. The holistic approach into the interrogation of multiple read-outs using liquid biopsies, indicated a potential clinical utility of several biomarkers from liquid biopsies to monitor disease burden peri-operatively includes more accurate prognostication for patients as well as guiding the role of adjuvant therapy (chemotherapy and radiotherapy). Moreover, it paves a research path for further studies into the role on lncRNAs in this particular malignancy.

Contents

List of Figures

List of Tables

To my family...

Acknowledgements

Declaration of Authorship

I, Periklis Katopodis, declare that this thesis titled, "Use of liquid biopsies as prognostic and diagnostic marker for Non-Small Cell 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

Peer reviewed publications resulting from this work

1. **Katopodis P.**, Dong Q., Halai H., Fratila C., Polychronis A., Sisu C., Anikin V., Karteris E.: In silico and in vitro analysis of lncRNA XIST reveals a panel of possible lung cancer regulators and a five-gene diagnostic signature. Cancers (2020)

2. Chudasama D., **Katopodis P.**, Stone N., Haskell J., Sheridan H., Gardner B., Urnowitz H., Schuetz E., Beck J., Hall M, Barr J., Sisu C., Rice A., Polychronis A., Anikin V., Karteris E.: Liquid Biopsies in Lung Cancer: Four Emerging Technologies and Potential Clinical Applications. Cancers (2019)

3. **Katopodis P.**, Chudasama D., Wander G., Sales L., Kumar J., Pandhal M., Anikin V., Chatterjee J., Hall M., Karteris E.: Kinase inhibitors and ovarian cancer. Cancers (2019)

4. **Katopodis P.**, Anikin V., Kishore U., Carter T., Hall M., Polychronis A., Karteris E.: Peri-operative circulating tumour cells and ccfDNA in open thoracotomy vs. VATS. (submitted)

Peer reviewed publications not resulting from this work

1. **Katopodis P.**, Kerslake R., Davies J. , Randeva H. S., Chatha K. , Hall M., Spandidos D. A., Anikin V., Polychronis A., Robertus J. L., Kyrou I., Karteris E.: COVID-19 and SARS-CoV-2 host cell entry mediators: Expression profiling of TMRSS4 in health and disease. International Journal of Molecular Medicine (2021)

2. Empeslidis T., Tsaousis K., **Katopodis P.**, Tyradellis S.: Patient's satisfactory outcomes after implementation of follow up "virtual clinic" in age related macular degeneration treatment. Clinical and Experimental Optometry (2021)

3. Klonos E, **Katopodis P.**, Karteris E., Papanikolaou E., Tarlatzis B., Pados G.: Endometrial changes of estrogen and progesterone receptors expression during implantation process in an oocyte donation program. Experimental and Therapeutic Medicine (2020)

4. Fiorentzis M., **Katopodis P.,** Kalirai H., Seitz B., Viestenz A., Coupland S.E.: Image Analysis of 3D Conjunctival Melanoma Cell Cultures Following Electrochemotherapy. Biomedicines (2020)

5. **Katopodis P.,** Anikin V., Harpal S. Randeva, D. A. Spandidos, Chatha K., Kyrou I., Karteris E.: Pan-cancer analysis of transmembrane protease serine 2 and cathepsin L that mediate cellular SARS-CoV2 infection leading to COVID19. International Journal of Oncology (2020)

6. **Katopodis P.,** Karteris E., Katopodis K.P.: Pathophysiology of Drugs-induced Hypomagnesemia. Drugs Safety (2020)

7. **Katopodis P.,** Khalifa M.S., Anikin V.: Molecular characteristics of uveal melanoma and intraocular tumors. Oncology Letters (2020)

8. Pappas M.E., Mpournaka S., Chardalias A., Tsakas S., **Katopodis P.,** Katopodis KP, Goumenos D.S.: The effect of Dialysis modality and Membrane performance on native immunity in Hemodialyzed patients.Pril (Makedon Akad Nauk Umet Odd Med Nauki) (2019)

9. Fiorentzis M., Kalirai H., **Katopodis P.,** Seitz B., Viestenz A., Coupland S.E.: Electrochemotherapy with bleomycin and cisplatin enhances cytotoxicity in primary and metastatic uveal melanoma cell lines in vitro. Neoplasma (2018)

10. Fiorentzis M., **Katopodis P.,** Kalirai H., Seitz B., Viestenz A., Coupland SE.: Conjunctival Melanoma and Electrochemotherapy: Preliminary results using 2D and 3D cell culture models in vitro. Acta Ophthalmologica (2018)

11. Fiorentzis M., Kalirai H., **Katopodis P.,** Coupland S.E.: Adipophilin expression in primary and metastatic uveal melanoma: a pilot study. Graefe's archive for clinical and experimental ophthalmology (2017)

List of Abbreviations

General

A	Active
AC	Adenocarcinoma
ACa	Atypical Carcinoid
BC	Blood cell
BM	Biphasic mesothelioma
cDNA	Complementary DNA
cfDNA	Circulating cell-free DNA
CNI	Copy Number Instability
CT	Computed Tomography
CTC	Circulating Tumour Cell
ctDNA	Circulating tumor DNA
ECL	Enhanced Chemiluminescence
ECM	Extracellular Matrix
EDTA	Ethylene-diamine-tetra-acetic acid
EM	Epithelial Mesothelioma
EMT	Epithelial to mesenchymal transition
EP	Extrapleural Pneumonectomy
FFPE	Formalin-fixed paraffin-embedded
IMIG	International Mesothelioma Interest Group
ISET	Isolation by Size of Epithelial Tumor cells
LCC	Large cell carcinoma
LCL	Lower Control Limit
LCNEC	Large-Cell Neuroendocrine Tumors
lncRNA	Long non-coding RNA
MET	Mesenchymal to epithlial transition
MM	Malignant Mesothelioma
MPM	Malignant pleural mesothelioma
NR	Not Recruiting
NSCLC	Non-Small cell lung cancer
P/D	Pleurectomy/Decortication
PB	Peripheral Blood
PBL	Peripheral Blood Leukocytes
PET	Positron Emission Tomography
qPCR	Quantitative Polymerase Chain Reaction

R	Recruiting
RBC	Red Blood Cell
RT	Room Temperature
SC	Single Cell
SCLC	Small cell lung cancer
siRNA	small interfering RNA
SM	Sarcomatoid Mesothelioma
SSC	Squamous cell carcinoma
TC	Typical carcinoma
TDI	Time Delay Integration
TNM	Tumour Nodes Metastases
UCL	Upper Control Limit
XIC	X inactivation centre

Molecules & Drugs

AAK	Aurora A Kinase
AKT2	RAC-beta serine/threonine-protein kinase
ALK	Anaplastic lymphoma kinase
BCR-ABL	BCR-ABL mutation
BRAF	Serine/threonine-protein kinase B-Raf
CART	Chimeric antigen receptor T
CDK4/6	Cyclin-dependent kinase 4/6
CTLA-4	Cytotoxic T-Lymphocyte-associated protein
DART B7-H3	Dual-affinity re-targeting protein B7-H3 Histone Complex
DD3	Delta Protein 3
DDR2	Discoidin domain receptor tyrosine kinase 2
EGFR	Epidermal growth factor receptor
EpCAM	Epithelial cell adhesion molecule
EZH2	Histone-lysine N-Methyltransferase enzyme (H3K27)
FAK	Focal Adhesion Kinase
FGFR1	Fibroblast growth factor receptor 1
HER2	Human epidermal growth factor receptor 2
Hsp90	Heat Shock Protein 90
KRAS	Kirsten RAt Sarcoma virus
LAG-3	Lymphocyte-activation gene 3
Mesothelin Mab	Mesothelin Monoclonal Antibody
MET	Tyrosine-protein kinase Met
MSLN	Mesothelin
NAE	NEDD8-activating enzyme
NRAS	Neuroblastoma RAS viral oncogene homolog

PARP	Poly)ADP-ribose) polymerase
PD-1	Programmed cell death protein 1
PD-L1	Programmed cell death protein ligand 1
PEG-Arginine	Pegylated Arginine Deiminase
PIK3CA	Phosphatidylinositol- 4,5-bisphosphate 3-kinase catalytic subunit alpha
PTEN	Phosphatase and tensin homolog
RET	Ret proto-oncogene
ROS1	Proto-oncogene tyrosine-protein kinase ROS
SMRP	Serum Mesothelin Related Protein
TGF	Tumour Growth Factor
TGF	Transforming growth factor
TKI	Tyrosine Kinase Inhibitor
TK	Tyrosine Kinase
TTF-1	Thyroid transcription factor 1
TWIST	Twist-related protein 1
VEGF	Vascular Endothelium Growth Factor
XIST	X-inactive specific transcript

Chemicals & Buffers

DAB	3,3'-Diaminobenzidine
FBS	Fetal Bovine Serum
PBS	Phosphate Buffered Saline
Poly-HEMA	Poly 2-hydroxyethyl methacrylate

Chapter 1

Introduction

1.1 Lung Cancer

Lung cancer is the most common cause of cancer related death in the world. It continues to have an enormous impact on health systems of all countries. The number of new cases is increasing in a rate of about 3% annually. Despite the advances in the detection and treatment of lung cancer, the overall 5-year survival remains grim. Cigarette smoking remains the single biggest risk factor on the incidence of cancer, with 90% of all lung cancers occurring in smokers. Adenocarcinoma is the most frequent histologic type (approximately 50%) while squamous cell carcinoma accounts for approximately one third of lung cancers, and small cell carcinoma for 15%. Prognosis is influenced by the stage of the disease at diagnosis and by the treatment. Screening trials that have begun in the early 1950s based on chest X-ray and sputum cytology did not produce improvement in overall mortality [1]. Refinements in the staging classification of lung cancer and advances in stage identification were introduced in the 1990s. Post-surgical mortality has declined since the 1950s but the 5-year survival rates have improved only minimally. A gradual improvement is seen in locally advanced inoperable non-small cell lung cancer, mainly due to the advanced chemotherapy and radical radiotherapy. Chemotherapy offers small improvement for patients with non-small cell lung cancer (NSCLC). The management of NSCLC, which appeared so promising in the 1970s has hit a plateau with very little advance in the last years. The biological active agents currently in phase III trials offer some hope in the advance of therapy of lung cancer. The most important and cost-effective management for lung cancer is smoking cessation, but for those with this disease novel methods of treatment are urgently needed [1, 2].

Patients present with a variety of symptoms usually relating to the primary tumour. The commonest symptoms include cough, dyspnoea, weight loss and chest pain [3]. Haemoptysis and bone pain are also relatively common symptoms. Less frequently (0-20% of patients) finger clubbing and fever may be present (CancerResearchUK; CRUK). Treatments for small-cell lung cancer (SCLC) and NSCLC are very different and therefore accurate histological diagnosis is essential. Molecular genetic studies of lung cancer biopsies have revealed that clinically evident lung cancers have multiple genetic and epigenetic abnormalities, including DNA sequence alterations, copy number changes, and aberrant promoter hypermethylation [4]. Together, these abnormalities result in the activation of oncogenes and inactivation of tumour-suppressor genes. Extensive molecular genetic studies of lung cancer show that clinically overt lung cancers have multiple genetic and epigenetic alterations (>20 per tumour) [5–7].

In addition, several studies have demonstrated that pre-neoplastic cells and histologically normal bronchial epithelium harbour many of these abnormalities, suggesting that human lung cancer develops from normal epithelial cells through a multistep process involving successive genetic and epigenetic abnormalities, usually coincident with cigarette smoking. These abnormalities contribute to the initiation, development, and maintenance of lung cancer [8–10].

1.1.1   Incidence

Lung cancer by year and age specific incidence rates in UK

Lung cancer is the most commonly occurring cancer in men and the third most commonly occurring cancer in women worldwide. Lung cancer incidence is strongly related to age, with the highest incidence rates being in older men and women. In the UK between 2014 and 2016, an average of 44% of the cases were diagnosed in men and women aged 75 and over (Fig. 1.1). Age-specific incidence rates rise steeply from around age 45-49 and peak in the 85-89 in males and 80-84 in females. Age-specific incidence rates are similar for men and women up to the age of 50-54, but afterwards, incidence rates are higher for males than for females, and this gap widens with increase in age. At age 55-59, the male: female incidence ratio of age-specific rates (to account for the different proportions of males to females in each age group) is around 11:10; by ages 85-89, it is around 19:10 [11, 12].
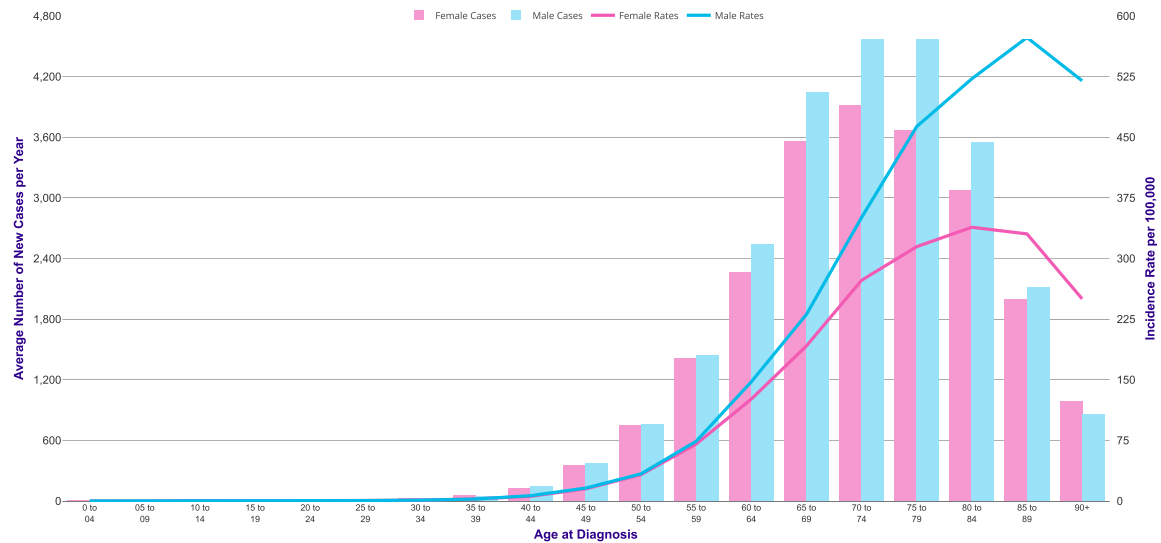
FIGURE 1.1: **Average number of new cases.** Average Number of New Cases per Year and Age-Specific Incidence Rates per 100,000 Population, UK, Source: cruk.org/cancerstats

Lung cancer incidence by sex and UK region

Lung cancer is the third most common cancer in the UK (CRUK, 2016), accounting for 13% of all new cases. It is the second most common cancer in both males (13% of all male cases) and females (13% of all female cases) after prostate cancer (26%) and breast cancer (30%). In 2016, there were 47,388 new cases of lung cancer in the UK: 25,046 (53%) in men and 22,342 (47%) in women. The crude incidence rate shows that there are 77 new lung cancer cases for every 100,000 males in the UK, and 67 for every 100,000 females per year (Table 1.1). The European Age-Standardised incidence rates (AS rates) are significantly higher in Scotland compared with the other parts of the UK for both males and females.

TABLE 1.1: Number of New Cases, Crude and European Age-Standardised (AS) Incidence Rates per 100,000 Population, UK (95% Lower Confidence Limit - LCL and 95% Upper Confidence Limits - UCL are around the Age Standardised Rate, Source: cruk.org/cancerstats)

		England	Scotland	Wales	Northern Ireland	UK
Female	Cases	17.821	2.586	1.237	698	22.342
	Crude Rate	63,7	93,1	78,3	73,7	67,2
	AS Rate	65,5	91	72,1	81,8	68,5
	AS Rate 95% LCI	64,6	87,5	68,1	75,7	67,6
	AS Rate 95% UCI	66,5	94,5	76,2	87,9	69,4
Male	Cases	20.560	2.526	1.307	653	25.046
	Crude Rate	75,3	96,1	85,2	71,3	77,4
	AS Rate	89,8	111	90,2	93,4	91,7
	AS Rate 95% LCI	88,6	106,5	105,3	106,2	90,5
	AS Rate 95% UCI	91,0	115,4	95,1	100,5	92,8
Persons	Cases	38.381	5.112	2.544	1.351	47.388
	Crude Rate	69,4	94,6	81,7	72,6	72,2
	AS Rate	76,1	99,3	79,8	86,4	78,5
	AS Rate 95% LCI	75,3	96,5	76,7	81,8	77,8
	AS Rate 95% UCI	76,8	102,0	82,9	91,0	79,2

1.1.2 Mortality

Lung cancer mortality by age

Lung cancer mortality is strongly related to age, with the highest mortality rates being in older men and women. In the UK between 2015 and 2017, an average of 49% of lung cancer deaths were in men and women aged 75 years and over (Fig. 1.2). Age-specific mortality rates rise sharply from around age 45-49, with the highest rates in the 90+ age group in men and at age 85-89 years in women. Mortality rates are generally similar between males and females until age 50-54, after which time rates are higher for males than for females and this gap is widest in the 85+ age group [2, 11, 12].
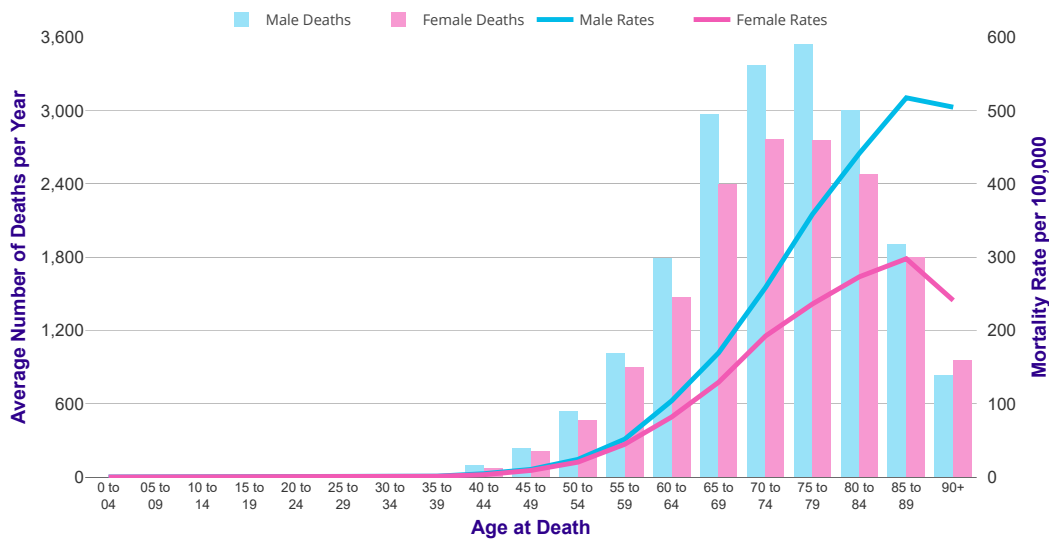
FIGURE 1.2: Average numbers of deaths per year.Rates per 100,000 Population, UK, Source: cruk.org/cancerstats

Lung cancer mortality statistics by sex and UK region

Lung cancer is the most common cause of cancer death in the UK (2017), accounting for 21% of all deaths from cancer. Lung cancer is the most common cause of cancer death among men in the UK (2017), accounting for 21% of both male and female deaths from cancer. (Table 1.2).

In 2017, there were 35.148 deaths from lung cancer in the UK: 18.810 (54%) in men and 16.338 (46%). The crude mortality rate shows that there are 62 lung cancer deaths for every 100,000 males in the UK, and 50 for every 100,000 females. The European age-standardised mortality rates are significantly higher in Scotland compared with the other constituent countries of the UK for both males and females. The mortality rate in England is also significantly lower than that in Northern Ireland for males only.

TABLE 1.2: Number of Deaths, Crude and European Age-Standardised (AS) Mortality Rates per 100,000 Population, UK (95% Lower Confidence Limit - LCL and 95% Upper Confidence Limits - UCL are around the Age Standardised Rate, Source: cruk.org/cancerstats)

		England	Scotland	Wales	Northern Ireland	UK
Female	Deaths	12.959	2.033	890	456	16.338
	Crude Rate	46,1	73,0	56,2	48,0	48,8
	AS Rate	46,1	70,4	50,7	52,5	48,6
	AS Rate 95% LCI	45,3	67,4	47,4	47,6	47,9
	AS Rate 95% UCI	46,9	73,5	54,1	57,3	49,4
Male	Deaths	15.229	2.036	1.001	544	18.810
	Crude Rate	55,4	77,1	65,0	59,1	57,7
	AS Rate	65,8	88,8	68,0	78,0	68,1
	AS Rate 95% LCI	64,7	84,9	63,8	71,4	67,1
	AS Rate 95% UCI	66,8	92,7	72,3	84,5	69,1
Persons	Deaths	28.188	4.069	1.891	1.000	35.148
	Crude Rate	50,7	75,0	60,5	53,5	53,2
	AS Rate	54,7	78,1	58,1	63,2	57,1
	AS Rate 95% LCI	54,1	75,7	55,5	59,3	56,5
	AS Rate 95% UCI	55,1	80,5	60,8	67,1	57,7

1.1.3 Lung cancer classification

Histological classification

Lung cancer is only rarely a tumour of the true lung parenchyma, arising far more frequently in large and medium sized bronchi (Fig. 1.3). There are many histological types of lung cancer. However, it is clinically and histologically convenient to consider the commonest varieties in four major groups, although there is substantial histological variation within each of these. There are two main types of lung cancers: around 80% are non-small cell lung cancers (NSCLC) and 20% are small cell lung cancers (SCLC) and the remainders. The main types of NSCLC are squamous cell carcinoma (SqCC), adenocarcinoma (AC) and large cell carcinoma (LCC), which account for approximately 35%, 27% and 10% of all lung cancer cases, respectively [13]. Given the significant impact on healthcare worldwide, a uniform system of tumour classification is needed to be established [2, 14]. From a clinical standpoint, lung carcinomas are broadly divided into SqCC and NSCLC for treatment purposes.

As mentioned, NSCLC traditionally include SqCC, AC and LCC, but in the broadest sense may include any epithelial tumour that lacks a small cell component [14].
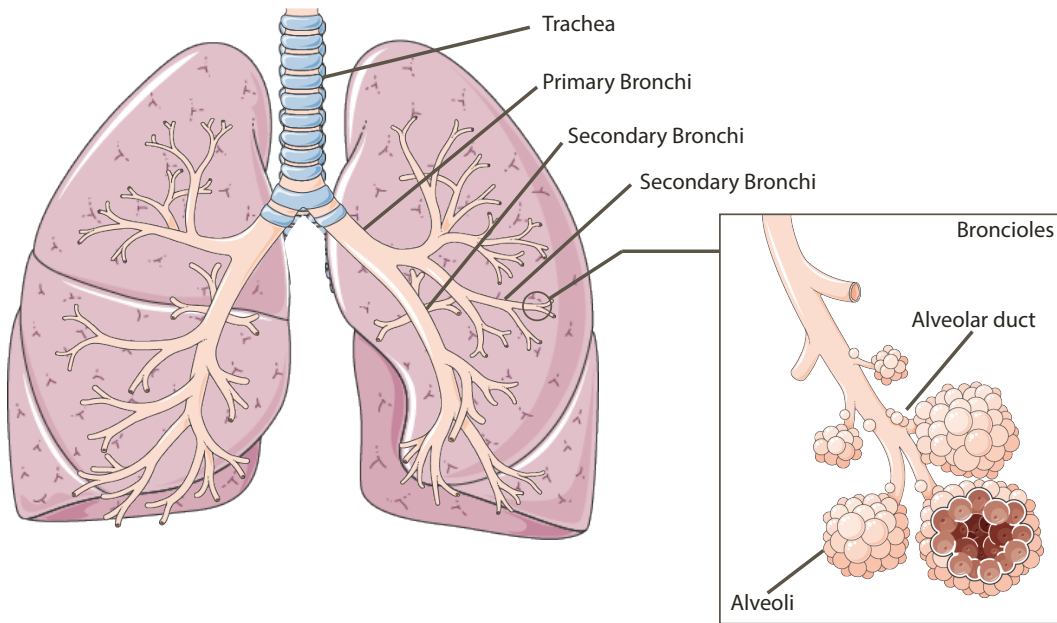
FIGURE 1.3: The main bronchi divide into lobar or secondary bronchi within each lung. Two lobar bronchi exist in the left lung, and three exist in the right lung. The lobar bronchi, in turn give rise to segmental or tertiary bronchi. The tertiary bronchi supply the bronchopulmonary segments.

Non-Small Cell Lung Cancer

Squamous cell carcinoma (SqCC)

SqCC accounts for 35% of the cases and it is the most common histological type and is characterised by the presence of keratinization and intracellular bridging (Fig. 1.4). Most of these tumours arise proximally in large bronchi and tend to be polypoid or infiltrating. Mutations and loss of heterozygosity of the p53 gene have been found in many of these tumours.
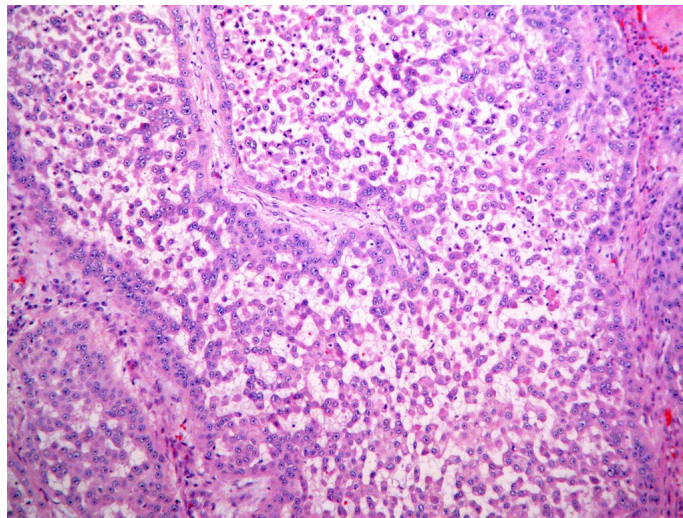
FIGURE 1.4: Squamous cell carcinoma. Biopsy of a highly differentiated SqCC of the mouth. Haematoxylin and eosin stain. Permission and photo by Yale Rosen.

Adenocarcinoma (AC)

The ACs are the most common type of lung cancer in many countries [13]. These show the typical features of neoplastic cells derived from glandular epithelium. Histologically, AC of the lung is sub classified into acinar (Fig. 1.5), papillary, bronchioloalveolar and solid types. Many ACs are peripheral in site of origin, frequently invading the pleura. They sometimes arise in scar tissue and may occur in fibrotic lung disease. This type is less related with cigarette smoking than the rest. Unlike other forms of lung cancer, it is slightly more common in females. ACs are typically heterogeneous and contain more than one subtype in over 80% of cases. Such cases are classified as AC with mixed patterns [15].
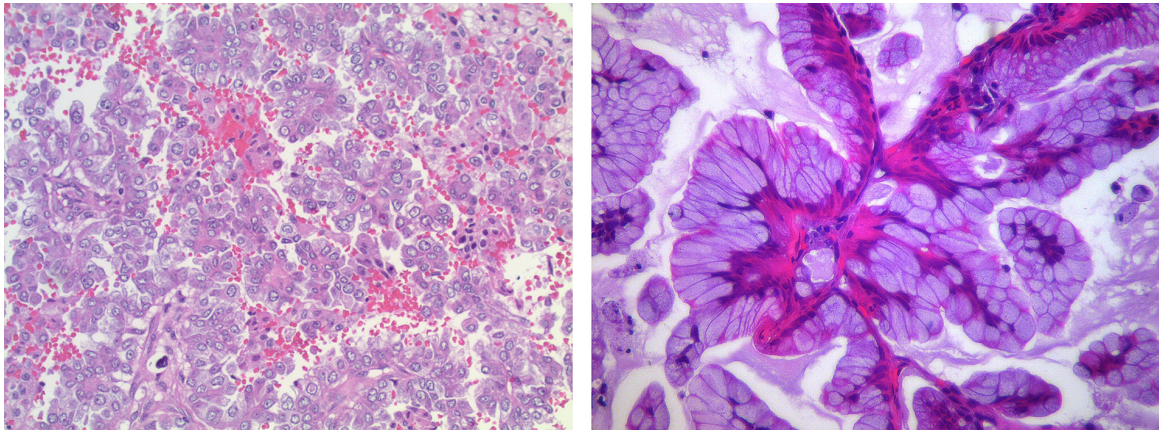
FIGURE 1.5: Adenocarcinomas. Left: Adenocarcinoma, invasive, acinar. These tumour cells are forming glands or acini indicating adenocarcinoma differentiation. Right: Bronchi alveolar carcinoma, mucinous type. The tumour cells lining the alveoli are tall columnar goblet cells with basally located nuclei. Permission and photo by Yale Rosen.

Large cell carcinoma (LCC)

This subtype refers to undifferentiated tumours with a variety of appearances. The cells are large, often with featureless cytoplasm and show little tendency towards keratinization or acinar formation (Fig. 1.6). No clear-cut pattern of clinical or radiological presentation distinguishes them from other malignant tumours.

Findings include abundant mitosis, prominent nucleoli, abundant cytoplasm and well-defined cell borders [16]. Neuroendocrine tumours of the lung are a distinctive subset of lung cancers which share certain morphologic, immuno-histochemical, and ultrastructural features. The main tumour types include the low-grade typical carcinoid (TC), the intermediate grade atypical carcinoid (ACa), and the two high-grade tumours: large-cell neuroendocrine tumours (LCNEC) and SCLC. TC and ACa are categorized together under the heading of carcinoid tumours; LCNEC is retained as a subtype of LCC, and SCLC is retained as an independent category. All four tumours are histologically characterised by varying degrees of neuroendocrine morphologic features. These features consist primarily of an organoid or trabecular growth pattern, peripheral palisading of tumour cells around the periphery of tumour nests, and the formation of rosette structures [2].
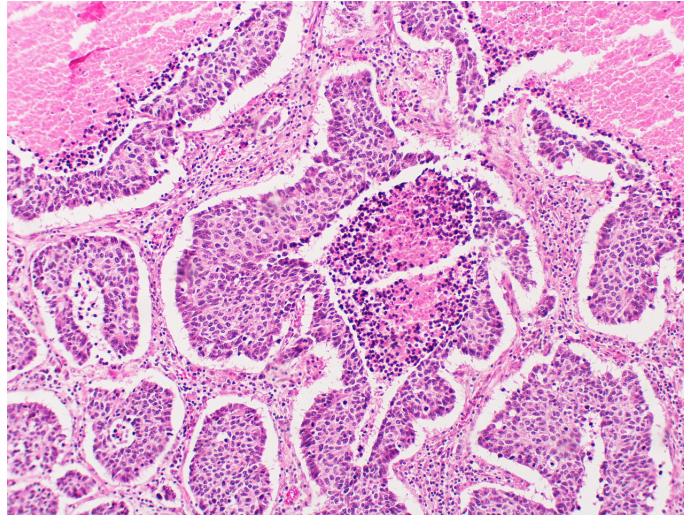
FIGURE 1.6: Large cell carcinoma. Large Cell Neuroendocrine Carcinoma (LCNE) This low magnification image shows an insular/organoid growth pattern and extensive areas of necrosis. Permission and photo by Yale Rosen.

Small cell lung cancer (SCLC)

SCLC is characterised by a diffuse growth of small cells with fine granular nuclei, inconspicuous nucleoli and scanty cytoplasm (Fig. 1.7). Neurosecretory granules are often present on electron microscopy. Numerous peptide hormones bind to receptors on the SCLC cell surface. Karyotypically, there is an almost invariable deletion of part of the short arm of chromosome 3 and loss of heterozygosity at the site of p53 gene [15].
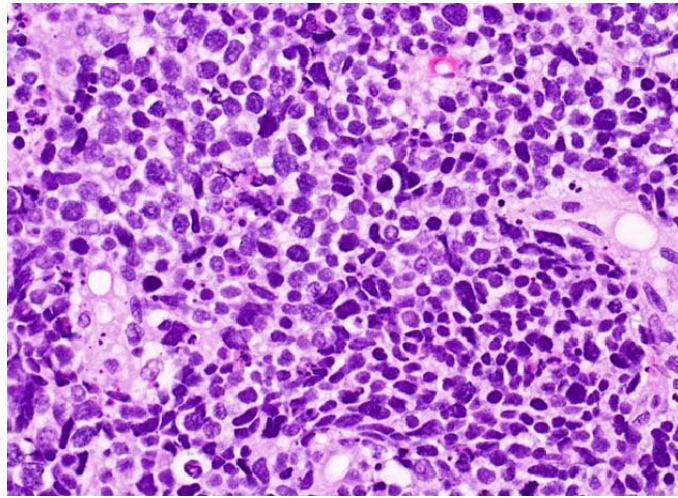
FIGURE 1.7: Small cell lung cancer. Tumour cells show pleomorphic hyperchromatic nuclei with finely granular chromatin, inconspicuous nucleoli, scanty cytoplasm and ill-defined cell borders. Nuclear molding, mitoses and individual cell necrosis are frequently seen. Permission and photo by Yale Rosen,

TNM Staging

Stage is one of the most important prognostic factors for diagnosis and is essential when making treatment decisions, it is important that it is as accurate as possible. At present, the extent of NSCLC disease is staged according to the 2009 tumour–node–metastasis (TNM) classification (Adapted from Cancer.org, 2018).

The TNM staging system for NSCLC is the following:

TX: Primary tumour cannot be assessed or tumour is proven by presence of malignant cells in sputum or bronchial washing but not visualised by imaging or bronchoscopy.

Tis: Carcinoma in situ
T1: < 3 cm in greatest dimension surrounded by lung or visceral pleura, without bronchoscopic evidence of invasion more proximal than lobar bronchus

T1a: $\leq$ 2 cm

T1b: > 2 cm but $\leq$ 3 cm

T2: > 3 cm but $\leq$ 7 cm or tumour with any of:

- Main bronchus involvement > 2 cm distal to the carina

- Invasion of visceral pleura

- Tumour with partial atelectasis [complete or partial collapse of the entire lung or area (lobe) of the lung]

T2a: > 3 cm but ≤ 5 cm

T2b: > 5 cm but ≤ 7 cm

T3: > 7 cm

T4: Tumour of any size that invades any of: Mediastinum, heart great vessels, trachea, recurrent laryngeal nerve, carina, vertebral body or esophagus or separate nodule(s) in a different ipsilateral lobe

NX: Regional lymph nodes cannot be assessed

N0: No regional node metastasis

N1: Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and itrapulmonary nodes, including involvement

N2: Metastasis in ipsilateral mediastinal and/or subcanical lymph node(s)

N3: Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene or supraclavicular lymph nodes.

MX: Distant metastasis cannot be assessed

M0: No distant metastasis

M1a: Chest cavity metastasis

M1b: Distant metastases outside the chest cavity

The TNM system helps describe cancer in great detail. But, for many cancers, the TNM combinations are grouped into five less-detailed stages (Fig. 1.8).

ANATOMIC STAGE/PROGNOSTIC GROUPS			
Occult Carcinoma	TX	N0	M0
Stage 0	Tis	N0	M0
Stage IA	T1a	N0	M0
	T1b	N0	M0
Stage IB	T2a	N0	M0
Stage IIA	T2b	N0	M0
	T1a	N1	M0
	T1b	N1	M0
	T2a	N1	M0
Stage IIB	T2b	N1	M0
	T3	N0	M0
Stage IIIA	T1a	N2	M0
	T1b	N2	M0
	T2a	N2	M0
	T2b	N2	M0
	T3	N1	M0
	T3	N2	M0
	T4	N0	M0
	T4	N1	M0
Stage IIIB	T1a	N3	M0
	T1b	N3	M0
	T2a	N3	M0
	T2b	N3	M0
	T3	N3	M0
	T4	N2	M0
	T4	N3	M0
Stage IV	Any T	Any N	M1a
	Any T	Any N	M1b

FIGURE 1.8: Cancer Stage grouping. Stage 0. This stage describes cancer in situ, which means "in place." Stage 0 cancers are still located in the place they started and have not spread to nearby tissues. This stage of cancer is often highly curable, usually by removing the entire tumor with surgery; Stage I. This stage is usually a small cancer or tumor that has not grown deeply into nearby tissues. It also has not spread to the lymph nodes or other parts of the body. It is often called early-stage cancer; Stage II and Stage III. In general, these 2 stages indicate larger cancers or tumours that have grown more deeply into nearby tissue. They may have also spread to lymph nodes but not to other parts of the body; Stage IV. This stage means that the cancer has spread to other organs or parts of the body. It may also be called advanced or metastatic cancer.

Surgical treatment for lung cancer

Surgical resection of early-stage NSCLC

Surgery is the gold standard treatment for early-stage (I and II) lung cancer. Surgical procedures continue to evolve along a 'less is more' trajectory, with an increasing trend toward minimally invasive techniques (video-assisted thoracoscopic surgery [VATS] or robotics) and the removal of as little lung parenchyma as possible (lobectomy > segmentectomy > wedge resection (Fig. 1.9). The type of resection selected for the treatment of T1aN0M0 tumours depends on the patient's age, cardiopulmonary reserve, tumour size and histology (Fig. 1.9).
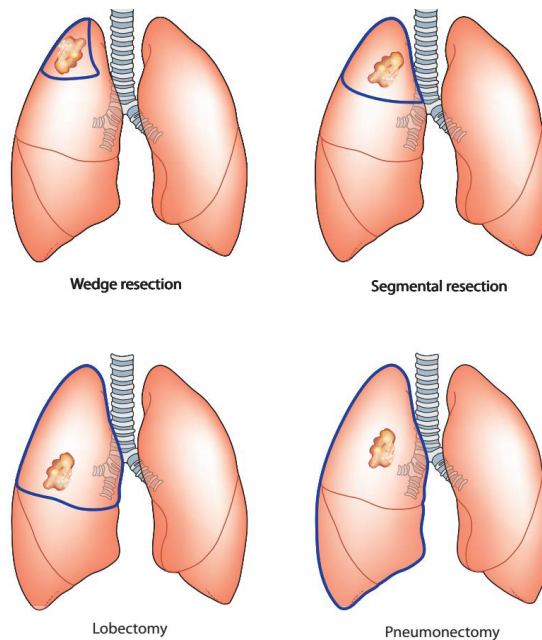
FIGURE 1.9: Surgical options for NSCLC

1.2   Mesothelioma

Mesothelioma is an uncommon neoplasm arising from the mesothelial cells lining the pleura. Rarely, pleural mesothelioma is localised, benign, and readily resectable for cure. A variant of localised pleural mesothelioma is a fibrous tumour of the pleura that probably arises from a different layer of cells in the pleura, and it is usually completely resectable [17]. Malignant mesothelioma (MM) is a highly aggressive tumour of the mesothelial origin associated with asbestos exposure, and most commonly develops in the pleura with a very poor prognosis. MM is rare, but is increasingly prevalent in many industrialised countries even after a ban of asbestos usage probably because of its long latency period between asbestos exposure and development of MM. Thus, the prevention, diagnosis and therapy of MM are important issues in occupational medicine. The diagnosis of MM is usually confirmed by histological examination of biopsied samples, which are usually obtained by thoracoscopic pleural biopsy. These invasive procedures may not be appropriate for mass-screening to identify MM patients among high-risk population with history of asbestos-exposure or cannot be performed for patients with impaired organ functions. Among less invasive procedures for the diagnosis, radiographic examinations such as chest roentgenogram and computed tomography (CT) are most commonly employed, but do not provide definitive diagnosis of Malignant Pleural Mesothelioma (MPM). Blood-based tests may be promising, but the serum mesothelin related protein, the only clinically approved blood-test, may not provide sufficient diagnostic sensitivity. Accordingly, a novel blood-based test for the diagnosis of MM should be established for early diagnosis as well as improvement of prognosis of MM patients. MPM early diagnosis is not easy and radiological surveillance is far from allowing early disease detection even when high-risk populations are screened. Therefore, one speculates that the use of blood-based biomarkers might solve this difficulty and allow detection of MPM at an early stage and/or to predict efficacy of treatments involving surgery, chemotherapy and radiotherapy.

1.2.1   Staging

As with all malignancies, proper staging is crucial in MPM for rational treatment planning. Over the years, many staging systems have been proposed. The most widely accepted is the TNM-type system of the International Mesothelioma Interest

Group (IMIG) [18]. The IMIG system is the most comprehensive classification, albeit somewhat more detailed. A brief explanation of the classification follows:

Stage I includes lymph node-negative patients with minimal tumour confined to the parietal pleura (stage Ia) or with minimal visceral pleural involvement (stage Ib).

Stage II includes lymph node-negative patients with confluent superficial tumour on all pleural surfaces or involvement of the diaphragmatic muscle or lung parenchyma. Stage I and II patients have potentially resectable tumour.

Stage III is the most common presenting stage and includes patients with metastasis to hilar (N1) or ipsilateral mediastinal (N2) lymph nodes, or those with extension of tumour into the soft tissues of the chest wall, the endothoracic fascia, mediastinal fat or pericardium (T3 tumour).

Stage IV includes patients who have locally advanced tumour invading the spine or ribs, the chest wall extensively, trans diaphragmatic spread, or contralateral pleural spread. Patients with stage IV disease also may have contralateral or supraclavicular lymph node involvement (N3) or distant metastases.

Treatment

Surgery

Three surgical procedures may be used with MPM for palliation and/or treatment: a) VATS talc pleurodesis, b) pleurectomy/decortication (P/D), or c) extrapleural pneumonectomy (EPP) [17].

Radiation

Unlike most tumours, MPM grows as a diffuse sheet of tumour throughout the pleural cavity, enveloping the lung. As a result, it is difficult to deliver to the entire neoplasm the radiotherapy needed to be tumouricidal (>60Gy) because of the limitations on dose to the underlying structures. Radiation pneumonitis, myelitis, and hepatitis have been well described in early series attempting primary treatment with whole chest radiotherapy [17].

1.3 Metastasis

Metastasis is a very complex sequence of events that involves the completion of several steps. These include shedding of cells from the primary tumour into the blood and lymph stream, survival of the tumour cells in circulation, anchor at a new site, extravasation, initiation and maintenance of growth and at last vascularization that allows the growth of the metastatic tumour at its new location (Fig. 1.10) [19–21]. Although metastasis is a widely studied process, the exact events and the genetic and biochemical features are still incompletely understood. During the tumour development, it is believed that subsets of tumour cells within the primary tumours that possess numerous genetic abnormalities and display unregulated growth, at some point acquire the ability to invade nearby tissues and get some motility features that allow them to gain entrance in the circulation (Fig. 1.10) [22]. Metastatic cascade can last from weeks to years and during metastatic dissemination, a cancer cell from a primary tumour must execute a sequence of steps. Firstly, after receiving soluble signals (growth factors, cytokines), cell-cell interactions, epigenetic factors (circadian rhythm, ageing), it is locally invading the surrounding tissue, breaking its basal membrane if it is epithelial and often it acquires a mesenchymal phenotype with an epithelial to mesenchymal transition (EMT) [23, 24]. Secondly, it breaks the extracellular matrix (ECM) since it secretes matrix metalloproteinases and other enzymes. Now, it is free to enter the microvasculature of the circulating and lymph system [23]. If the cell survives in the vessels, it uses them to travel through the bloodstream to distant tissues. It exits the bloodstream (extravasation) and if it survives again in the microenvironment, it adapts to that and starts the proliferation and forms a secondary tumour (colonisation) [23, 24].
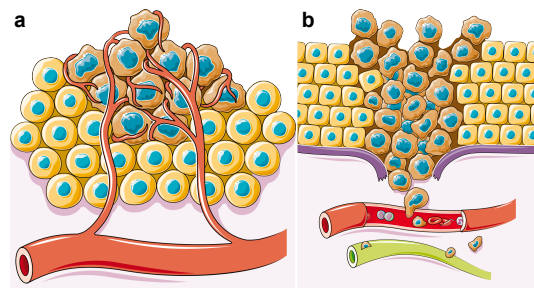
FIGURE 1.10: a) Angiogenesis and b) dissemination. Tumour has the ability to recruit blood vessels that provide nutrients and oxygen and an exit to enter the circulation

1.4   Liquid biopsies

Recent research around cancer cells and nucleic acids in the blood, such as circulating tumour cells (CTCs) and circulating tumour DNA (ctDNA) has generated much interest, opening the potential to a 'Liquid Biopsy' option, as an alternative to a conventional tissue biopsy. Liquid biopsies allow a better monitoring system, but also robust genetic characterisation of the cancer and its origin. Several advantages exist with the use of liquid over tissue biopsies. Liquid biopsies are non-invasive, and in comparison with conventional biopsies is also much more cost effective. Body liquids like blood, saliva and cerebrospinal fluid can be collected often, something that it is not feasible and impractical with a tissue biopsy (obtainable 2-3 times a year).

1.5   Circulating Tumour Cells (CTCs)

During the tumour development, cancer cells acquire some major genetic and epigenetic alterations that helps them to adapt easily in different micro-environments and form heterogenous tumour cell populations [25]. This heterogeneity in tumour cell population results in development of resistance to radiotherapy and chemotherapy. The resistance is acquired due to adaptation-related evolution of the tumour cells that eventually evolve to novel clones and metastasise [25]. CTCs are tumour cells of epithelial origin present in the bloodstream that shed either from the primary tumour or its metastasis. They possess antigenic and/or genetic properties of a specific tumour type and which makes them relatively identified. CTCs are found in the peripheral blood (PB) of patients with different types of carcinomas and they are rare cells since 1 CTC per 10^5 to 10^7 cells can be found in peripheral blood leukocytes (PBL) [26, 27]. CTCs follow the metastasis steps described above, which is intravasation $\rightarrow$ translocation $\rightarrow$ extravasation $\rightarrow$ colonisation. All cancer cells are communicating and are tightly connected with the neighbour cells with desmosomes and tight junctions. Even though they are close with surrounding cells, they are surrounded by a mix of ECM and neo-angiogenic blood vessels (Fig. 1.12). The organism creates this complex stroma to prevent distant metastasis and that is why carcinomas need to overcome this barrier. To do that, they enhance their motility in the stroma and they start invading into the vessels though EMT or non-EMT entry (Fig.1.12). Few invading CTCs will survive but if they do, they might be arrested and they extravasate and colonise the new distant site. The extravasation

and engraftment in a new site could be helped by a MET [28, 29]. The MET is supposed to play a fundamental role when CTCs are settling down in distant organs and start to form metastasis in the new site. A little is known about what is triggering the cells to balance between the EMT and MET during this metastasis cascade (Fig. 1.11). However, there are reports that suggest that the most aggressive CTCs sub-population cannot be captured with assays using epithelial antigens [30–32].
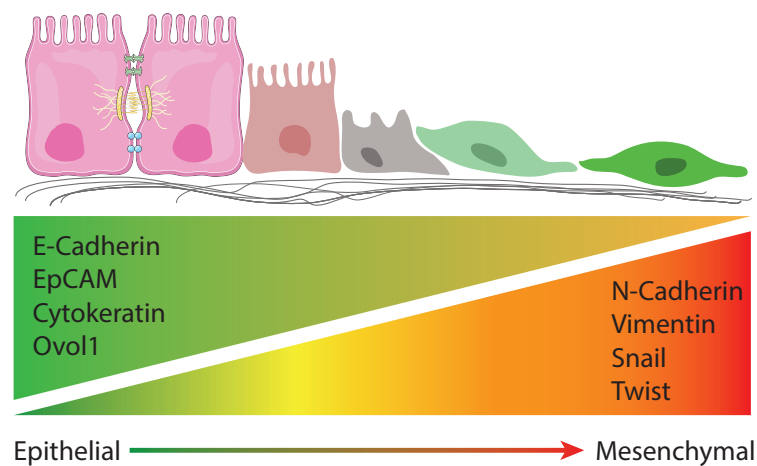
FIGURE 1.11: Schematic figure of EMT. This process results in a transformation of epithelial cells toward mobile mesenchymal cells. Expression of E-Cadherin, EpCAM and cytokeratins is a feature of epithelial cells and the increased expression of Vimentin, N-cadherin, Snail and Twist is common in mesenchymal cells. The E-cadherin to N-cadherin switch indicates epithelial cells are progressing through EMT.
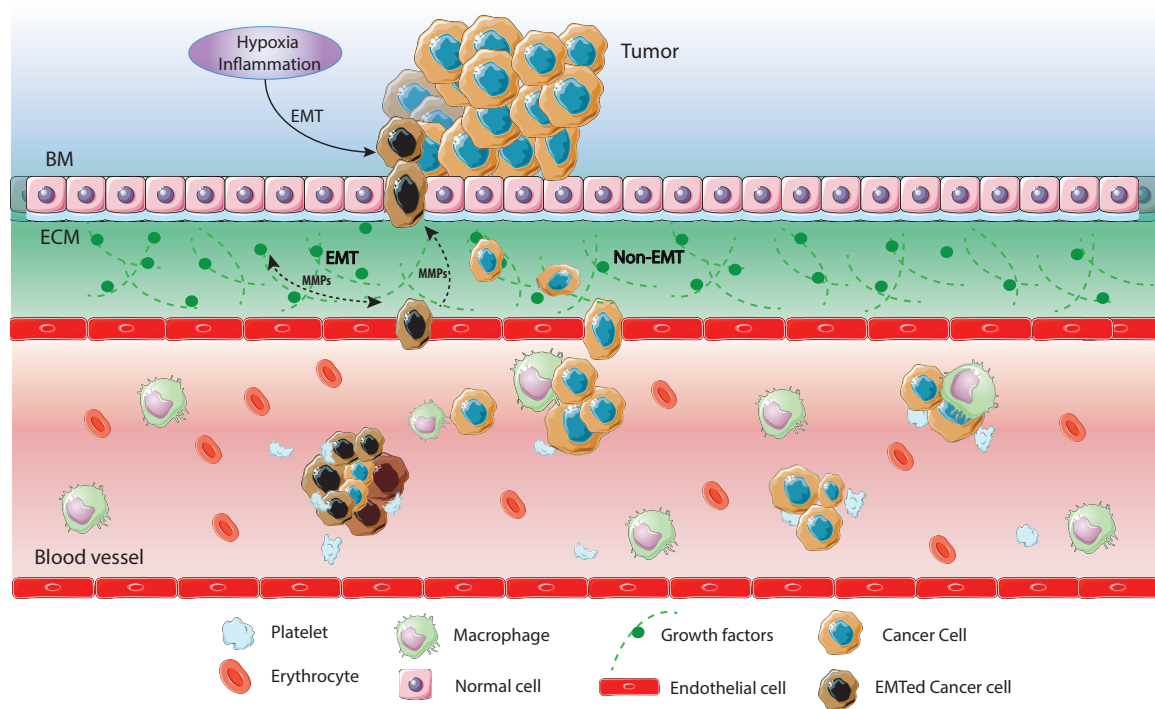
FIGURE 1.12: Mechanisms of CTCs generation and circulation. EMT and non-EMT are the main mechanisms involved in the movement of CTC to the blood. In EMT, cancer cells are breaking down the basement membrane (BM) with the help of matrix metalloproteinases (MMPs) and they migrate through the extracellular matrix (ECM) and enter the blood vessel. In non-EMT CTCs leave the tumours as single cells or in clusters. In both cases, all CTCs can be cloaked by platelets and macrophages

1.5.1 CTCs isolation techniques in Lung Cancer

The characterisation and enumeration of CTCs in the last years is considered as an attractive diagnostic tool, since it gives an idea of the aggressiveness of the tumour and could assist in therapeutic decisions in patients with solid malignancies. There are many detection methods that include cytometric and nucleic-acid based techniques [33, 34].

1.5.2 CTCs detection methods

There is a variety of approaches that researchers used to detect CTCs and they rely on the psychical properties of the cell, the expression of specific biomarkers and some functional characteristics of the CTCs [33].

Physical properties

There are many physical differences that have been found between CTCs and blood cells (BC) and in particular the white blood cells (WBC, CD45$^+$) and these are:

- Larger size than most epithelial cells

- Density

- Charge

- Migratory properties

Buoyant density is the main separation factor from the red blood cells and the increased size of the CTCs separated them from the leukocytes [35–37]. The distinction between the two can be done with filtration-based approaches, such as isolation by size and microelectrochemical systems [33]. Several interesting approaches have been developed like the ones that use pores and filters to capture CTCs but still none of the current approaches consist of an optimal platform for CTCs isolation [38–40]. One of these is the Isolation by Size of Epithelial tumour Cells (ISET) that isolates tumour cells by filtration based on their large size (>8μm) compared to leukocytes.

Surface antigens

Normal leukocytes lack epithelial surface markers and thus, CTCs are easily detectable with techniques that rely on antibody-based capture. The most common marker is epithelial adhesion molecule (EpCAM). CellSearch system (Veridex), the only one FDA-approved method uses ferrofluids full of EpCAM antibodies to capture CTCs, that are subsequently visualized by staining with a mix of antibodies against the cytoplasmic epithelial cytokeratins [41]. With ImageStream system, a blood sample is counted as CTC positive if at least one CTC is detected. AKT2, PI3CA and TTF-1/NKX2.1 are only a few of the markers that can be found to be heterogeneously expressed in CTCs of NSCLC and markers like the mesenchymal marker TWIST cannot be detected in CTCs [42–45]. In order to exclude the leukocytes contained in the blood samples, the leukocyte-specific marker CD45 is used [33]. There are times that both markers for cytokeratins and CD45 stain the same cell and this double-positive cell is excluded from enumeration. Positive for selection will eventually be positive for cytokeratin and negative for CD45, but this should be standardised. The reason why the CTCs may be found CD45 positive

remains unanswered and presumably they may acquire the CD45 while they are residing in the bone marrow stem cell niche.

1.5.3   CTCs as diagnostic and prognostic factors in LC

CTCs are a promising tool to approach early diagnosis, prediction of treatment efficacy and early detection of LC [46]. There are many research groups that have reported a positive association between the number of CTCs and the clinical stage of the patient or the existence of distant metastases [47–50]. CTCs nowadays are easily recognised with staining technologies in combination with fluorescent digital scanners that also allow us to perform *in situ* molecular assays to find genomic mutations on a single cell basis [51–54].

Early Diagnosis

The current difficulty is the early diagnosis, many cancers are not picked up until in advanced stages, where treatment becomes difficult and is often associated with a poor prognosis. Novel devices have been tested for their efficiency in the detection and collection of viable CTCs for further analysis. Recent studies have shown the efficacy of these devices in the detection of CTCs from the peripheral blood, even with minute volumes of blood. One of these studies evaluates the efficiency of microchip technology, the 'CTC-chip' [55]. A study conducted in 2007 tested the efficacy of this device and reported that the CTC chip successfully identified CTC cells in the peripheral blood of patients with metastasising lung, prostate, pancreatic, breast and colon cancer in 115 out of 116 (99%) samples, with a range of 5-1,281 CTCs per ml and approximately 50% purity [55]. The study also reported successful isolation of CTCs in 7 out of 7 patients with early-stage prostate cancer. Suggesting that the device has high sensitivity and specificity and in some cases can allow for detection in the early stages of cancer [55]. Another study published in 2012, focused on NSCLC patients [56]. Paired peripheral blood samples were collected from the patient and CTCs were enumerated by using the CellSearch device, and filtrated/isolated by the ISET device. The results were promising, and CTCs were detected in 32 of 40 (80%) patients using the ISET device, compared with 9 of 40 (23%) with CellSearch. It is clear that there are varying results amongst the devices and would require extensive work to identify a device capable of high specificity and sensitivity with the ability to allow a more advanced analysis. It is also clear from the results that CellSearch

shows variable results with different cancers. The device has been FDA approved and has shown promising results in breast, prostate, and colorectal cancer, however, it is not the case with lung cancer yet. A recently developed device 'tLN' tethered lipoplex nanoparticles, has the ability to simultaneously detect and capture CTCs and circulating nucleic acids [57, 58]. The tLN device allows both cell sorting and intracellular biomarker detection of both CTCs and circulating nucleic acids simultaneously [59].

Prognosis

Detection of CTCs in the peripheral blood of cancer patients would act as an indicator for the presence of metastatic disease. Routine sampling of CTCs would allow monitoring of tumour progression and stage through CTC levels. Regression of disease coincides with a significant decrease in CTCs, whereas a rise in CTC count indicates advancement in malignancy. Other similar studies shown more than 5 CTCs per 7.5 ml of peripheral blood indicates progressive disease. Detection and enumeration of CTCs is a predictive marker of clinical outcomes and overall survival rates. Multivariate analysis have shown that CTC count can act as an independent prognostic indicator irrespective of other variables. A study conducted in 2011 by (Krebs et al.), is a strong example of how CTC analysis can be used as a prognostic indicator and prediction of progression-free survival [59]. The ability to detect the presence of CTCs at the time of diagnosis can act as an indicator of whether adjuvant chemotherapy is required in early-stage cancer patients. Similarities between CTCs and cancer stem cells (i.e. longevity characteristics, tumour initiation ability, self-renewing, and proliferative capacity), may explain the eventual relapse in patients who have previously been in remission after primary therapy if CTCs are detected in peripheral circulation at the time of diagnosis and treatment [60, 61].

Treatment monitoring

Predicting the biological fate of cancer is difficult from biopsies obtained from the primary cancer. Following on from primary treatment a significant proportion of patients will experience metastatic disease and development of secondary cancers, due to the inability to target the more aggressive metastasising population of cells.

The detection and bio-characterisation of CTCs can aid in determining the biological fate of malignancies. Chemo-sensitivity testing on CTC populations can identify treatments that are likely to induce apoptosis of metastasising cells, allowing for individualisation of diagnosis and treatment plans, based on CTC analysis. The risk of re-occurrence following curative therapy is a potential issue and affects a large proportion of patients. Approximately 20% of patients undergo curative resection, of which only 25-30% of these are alive at five years due to local/chest relapse in a quarter of this number and further afield metastatic spread in three quarters. A patient with stage I tumour has a greater than 50% chance of being alive and disease-free at this time point, but due to delays in diagnosis, such patients account for only 5% of the population of patients. The ability to assess individuals on the possibility of relapse and the development of secondary cancer would be of great importance in determining prognosis and could aid in treatment planning. CTC counts in peripheral blood prior to surgery and chemotherapy or other treatments are being shown as a marker that can independently predict the early re-occurrence in patients with cancer [62]. With the use of novel enrichment and molecular analytic techniques, it has been made possible to detect metastasising disease that may be undetectable using conventional imaging techniques.

1.6   Circulating cell free DNA (ccfDNA)

Almost 30 years ago, Kopreski et al. reported that they found more free RNA and DNA in the plasma or serum of cancer patients than healthy subjects [63, 64]. These observations are the reason of development of circulating cell-free DNA/RNA assays as early diagnostic and prognostic tools. Even though surgery and targeted therapies have improved cancer outcomes over the past 20 years, patients still die due to metastases and/or drug resistance [65]. The heterogeneity of the tumour and its clonal evolution is the main challenge for an effective treatment strategy [66, 67]. Tumours from different tissues show distinct morphological and phenotypical characteristics in cellular morphology, gene expression, proliferation and the ability to metastasize [68–71]. Differences within the cells in the same tumour is called intratumoural heterogeneity and this includes the chromosomal imbalances and the somatic mutations in either the primary tumour or its metastasis. During the metastasis, as mentioned above, the clones are evolving and due to selection pressure of anti-cancer drugs, they tend to develop resistance [72, 73]. Obtaining tissue biopsies to monitor treatment and diagnose the disease is invasive, costly and cannot be performed sequentially. For this reason, CTCs and ctDNA (circulating tumor DNA, tumor-derived fragmented DNA) that can be analysed from blood samples, are preferred. Not only they are rapid and non-invasive but also their low cost are the main factors that made the liquid biopsies seem so attractive. In addition, ctDNA enrichment is not depending on special equipment as CTC capturing [74].

There are many ways in which the DNA and RNA fragments are being discharged by the tumour cells. ccfDNA and ctDNA can be found in blood plasma and serum due to apoptosis or necrosis of the normal or primary and metastatic tumour cells [75]. ctDNA should not be confused with ccfDNA, a broader term which describes DNA that is freely circulating in the bloodstream, but is not necessarily of tumor origin.

ccfDNA as biomarker for cancer detection and treatment

The ctDNA is fragmented DNA, that originates directly from the cancer cell itself [76]. Emerging studies relate ctDNA to tumour progression and rate of tumour cell turnover, a biological measure of tumour aggressiveness [77, 78] ccfDNA released from apoptotic cells is usually found to be in 160-200bp fragments whereas ccfDNA from necrotic cells varies in length [79–81]. In both scenarios, ccfDNA concentration in the plasma or serum is increased in patients than healthy subjects just like the CTC number [82]. In numbers, normal subjects have an average mean of 13ng/ml free DNA concentration, whereas, according to Leon et al. various cancer patients' mean is around 180ng/ml [83] and in metastatic disease around 209 ng/ml. Stroun et al. showed that free DNA of cancer patients is originated from the tumour cells and many mutations in oncogenes and/or tumour suppressor genes have been found in the free-DNA in plasma [77, 84, 85].

Personalised treatment and drug monitoring

To choose an optimal targeted treatment, a series of experiments and analyses must be conducted to identify the heterogeneity in CTCs sub-populations. EpCAM positive/negative CTCs and mRNA/DNA assays alongside staining and counting CTCs will provide information about the tumour driving pathway and the mutations in tumour genome [86]. Combining all these, the best drug treatment can be chosen and if none of these work due to low CTC number, a leukapheresis can follow and a personal CTC marker for the patient could be identified. After deciding the drug treatment, the therapy response should be monitored, and this can be done with less invasive techniques like liquid biopsies, which is blood samples, that can easily be taken from the patient. It is important to monitor the therapy response because this will enable the doctor to switch to another more effective treatment if the tumour is resisting. The most common monitoring procedures are the CT, MRI and PET scan but all these procedures lack sensitivity and they cannot give any information about early stage of disease or treatment [87, 88]. There are examples that the size of the tumour remains the same in this kind of imaging monitors, but this could be the inflammation caused by the tumour and not the tumour itself [89]. On the other hand, counting CTCs in the blood stream and analysing them, they provide valuable information that can be combined with the classic monitoring methods and lead to the ultimate diagnostic and therapy monitoring tool [88, 90].

1.7   Treatments for NSCLC

Tyrosine-kinases are classified as a group of enzymes that consist of a catalytic sub-unit, which transfers a phosphate from nucleotide triphosphate to the hydroxyl group of one or more tyrosine residues on signal transduction molecules, result-ing in a conformational change affecting protein function. Upon activation, they function to auto-phosphorylate as well as phosphorylate other signalling molecules carrying out an important role in signal transduction and acting to activate and promote a variety of biological processes including cell growth, migration, differ-entiation and apoptosis. Amongst the most important cytoplasmic signalling path-ways activated are the phosphoinositide 3- kinase/Akt pathway/mechanistic target of rapamycin (PI3K/AKT/mTOR), the Ras/Raf/ mitogen-activated protein kinase (MAPK) pathway, the Raf/MEK/ERK1/2 pathway and the protein kinase C (PKC) pathway [91]. Following the success of a pure VEGF1 receptor inhibitor, the mon-oclonal antibody bevacizumab, it was hoped that the TKIs could target alternate angiogenic pathways in cancer growth. Additionally, the TKIs could potentially be useful in overcoming resistance to VEGF blockade [91, 92]. Tyrosine kinase in-hibitors (TKIs) utilise different mechanisms such as competing for the substrate and bind in that ATP-binding pocket during an active conformation, they can occupy a site adjacent to the ATP-binding pocket, this allows both the inhibitor and ATP to bind to the same protein and/or bind irreversibly to the protein kinase target [91, 93]. It has also been shown that TKIs can block protein kinase recruitment to the Hsp90-Cdc37 system. This is of particular importance in cancer cells, where these inhibitors deprive oncogenic kinases of access to this complex, leading to their degradation [91, 94]. Most research on plasma ctDNA for lung cancer samples is limited to the most frequently mutated genes like EGFR, KRAS, BRAF, TP53 and PIK3CA (Fig. 1.13) [95–97]. Epidermal growth factor receptor (EGFR) mutations are the most common ones, since they can be found in 10-40% of the NSCLC cases. These mutations are usually the reason for the use of the first-generation TKIs like gefitinib and erlotinib (Tarceva, Genentech) which are used as effective first-line treatment of advanced NSCLC (Fig. 1.13) [60, 80]. Usually, after 12-24 months after treatment initiation, resistance to EGFR-TKI and ALK-TKI has been observed and second and third- generation TKIs are being used. Some of them are the second generation Crizotinib, Ceritinib (Iressa, AstraZeneca), and Alectinib and the third-generation Rociletinib, Osimertinib and the AZD9291 [98–101] .

New drug targets

Thoracic oncology is under an era of new perspectives due to targeted therapy, which includes tyrosine kinase agents, immunobiological, immune checkpoint inhibitors and other agents. The first thoracic neoplasms to be at the spot of improved survival were the NSCLC. Prognosis for mesothelioma is often dismissal, with some trials reporting about 13 months with combination platinum-based chemotherapy opposite to 10 months without treatment. Programmed death-ligand 1 (PD-L1) inhibitors, for instance, have been reported to shown medium survival up to 18.8 months in combination with traditional combination chemotherapy approach. Immune checkpoints inhibitors, as anti PD-1, PD-L1 and CTLA-4 (Cytotoxic T - Lymphocyte - associated protein), have been regarded as promising drugs. Around 20-40% of PD-L1 expression has been observed in mesothelioma and stromal cells in tissue immunohistochemically. Overall survival and disease-free data is still poorly available in the literature, although encouraging its prescription, yet there is no sufficient data to sustain FDA approval, as most clinical trials are phase I and II trials. Another on-going class of promising anti-neoplastic agents on development are Mesothelin (MSLN) Immunobiological agents. MSLN is a membrane glycoprotein specific to the mesothelial cells; this class of agents, under development and initial phase trials, relies in specific mesothelial cancer cells elimination through diverse immunological mechanisms, immunotoxins, vaccines, chimeric monoclonal Abs, chimeric antigen receptor T cells (CAR-T) and Ab-drug conjugates. Summarised latest target-therapy trials on course enlisted at Clinicaltrials.gov, up to 1st of May 2018. Of those, active and recruiting trials, 70 were selected, encompassing a total of 8732 expected patients to be enrolled in the United States and Europe, most of them on Immune Checkpoints therapy (PD-1, PD-L1, CTLA-4) and novel drugs, as lymphocyte - activation gene 3 (LAG-3), anti-Focal Adhesion Kinase (anti-FAK); MSLN directed therapy also is ongoing, including CAR-T.

Molecular testing for tumour gene (somatic) mutations

The discovery of specific gene mutations in NSCLC has led to the development of targeted therapies. The presence of EGFR gene mutations, found primarily in adenocarcinomas, is predictive of responsiveness to EGFR tyrosine kinase inhibitors [102]. Furthermore, ACs with ALK-MET rearrangements are responsive to Crizotinib and patients with AC or NSCLC–NOS (Not Otherwise Specified), are more responsive to Pemetrexed (Alimta) than those with squamous cell carcinoma [103]. In the initial randomized phase 2 study of bevacizumab and chemotherapy in advanced NSCLC, bevacizumab was associated with life-threatening hemorrhage in patients with SqCC therefore, it is contraindicated in patients with this NSCLC histology (Table 1.3).

TABLE 1.3: Frequency of gene mutations in NSCLC

Gene	Alteration	Frequency in NSCLC
EGFR	Mutation	10–35%
KRAS	Mutation	15–25%
FGFR1	Amplification	20%
PTEN	Mutation	4–8%
DDR2	Mutation	4%
ALK	Rearrangement	3–7%
MET	Amplification	2–4%
HER2	Mutation	2–4%
BRAF	Mutation	1–3%
PIK3CA	Mutation	1–3%
AKT1	Mutation	1%
MEK1	Mutation	1%
NRAS	Mutation	1%
RET	Rearrangement	1%
ROS1	Rearrangement	1%
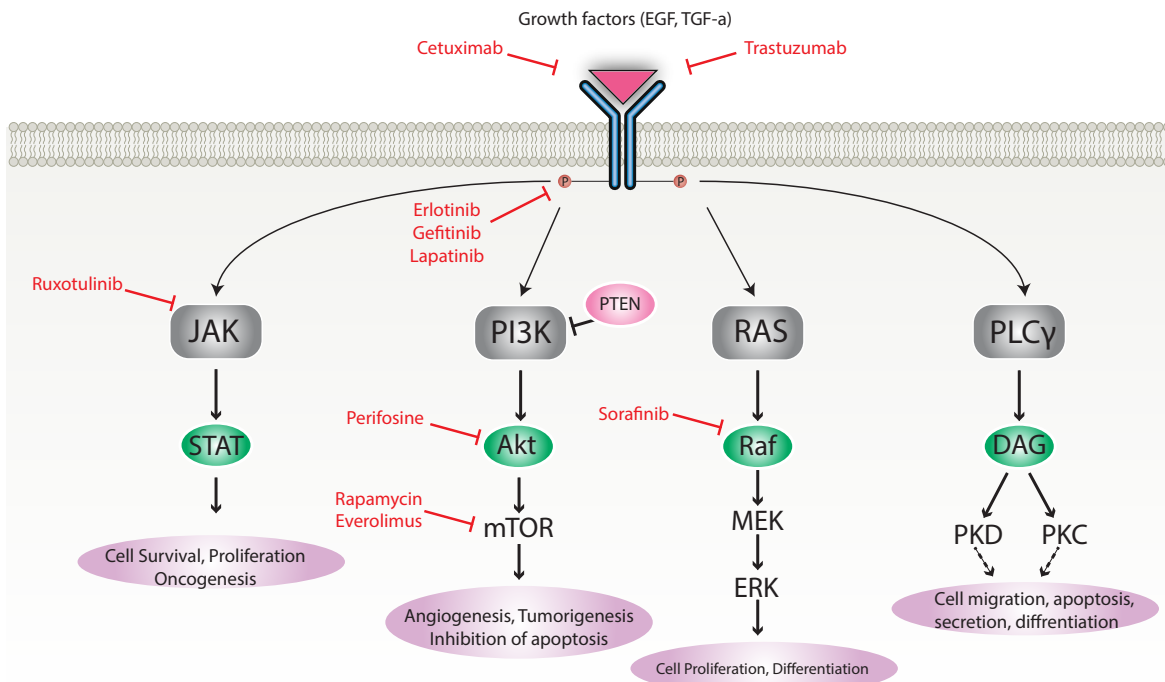
FIGURE 1.13: Clinically approved and in-development EGFR signalling pathway inhibitors for chronic respiratory disease. Three strategies are used to antagonise aberrant EGFR signalling. First, monoclonal antibodies bind the extracellular domain of EGF receptors and prevent endogenous ligand engagement, thereby attenuating downstream pathway activation (cetuximab and trastuzumab). Secondly, small molecule tyrosine kinase inhibitors target the intracellular portion of EGF receptors; in the absence of kinase activity EGFR autophosphorylation and downstream signalling is inhibited (erlotinib, gefitinib, lapatinib, HKI-272 and BIBW-2948). Finally, small molecule inhibitors which target specific downstream EGFR signalling pathway components have also been developed and may exhibit clinical efficacy in chronic respiratory disease. PI3K: phosphatidylinositol-3- kinase; PIP3: phosphatidylinositol (3,4,5)-triphosphate; Akt: protein kinase B; GSK3: glycogen synthase kinase 3; mTOR: mechanistic target of rapamycin; MEK: MAPK kinase; ERK: extracellular signal-regulated kinase; JAK: Janus kinase; STAT: signal transducer and activator of transcription.

1.8 Epidermal growth factor receptor

The EGFR is a member of the ErbB family of receptors, a subfamily of four closely related receptor tyrosine kinases: EGFR (ErbB-1), Her2/neu (ErbB-2), Her 3 (ErbB-3) and Her 4 (ErbB-4). In many cancer types, mutations affecting EGFR expression or activity could result in cancer (Fig. 1.14).

ErbB-1   ErbB-2   ErbB-3   ErbB-4
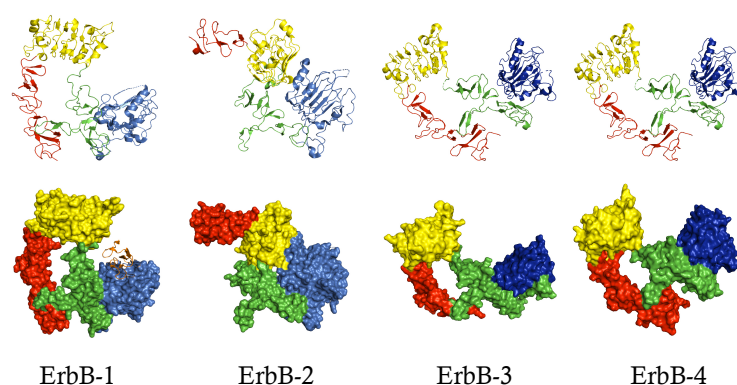
FIGURE 1.14: **Comparison of ErbB extracellular domain structures.** This figure shows the 3D structure of ErbB family proteins using the pdb files 1NQL (ErbB-1), 1S78 (ErbB-2), 1M6B (ErbB-3) and 2AHX (ErbB-4)

EGFR is a trans-membrane protein that is activated by binding of its specific ligands, including epidermal growth factor and transforming growth factor α (TGFα) (Fig. 1.15). ErbB2 has no known direct activating ligand and may be in an activated state constitutively or become active upon heterodimerization with other family members such as EGFR. Upon activation by its growth factor ligands, EGFR undergoes a transition from an inactive monomeric form to an active homodimer (Fig. 1.16) [104]. Although there is some evidence that preformed inactive dimers may also exist before ligand binding. In addition to forming homodimers after ligand binding, EGFR may pair with another member of the ErbB receptor family, such as ErbB2, to create an activated heterodimer (Fig. 1.16). There is also evidence to suggest that clusters of activated EGFRs are being formed, although it remains unclear whether this clustering is important for activation itself or occurs subsequent to activation of individual dimers. EGFR has been found to be over-expressed in many cancers such as gliomas and NSCLC. Drugs such as panitumumab, cetuximab, gefitinib, erlotinib, afatinib, and lapatinib are used to inhibit it (Figs. 1.17 and 1.18) [98, 99, 105–110].
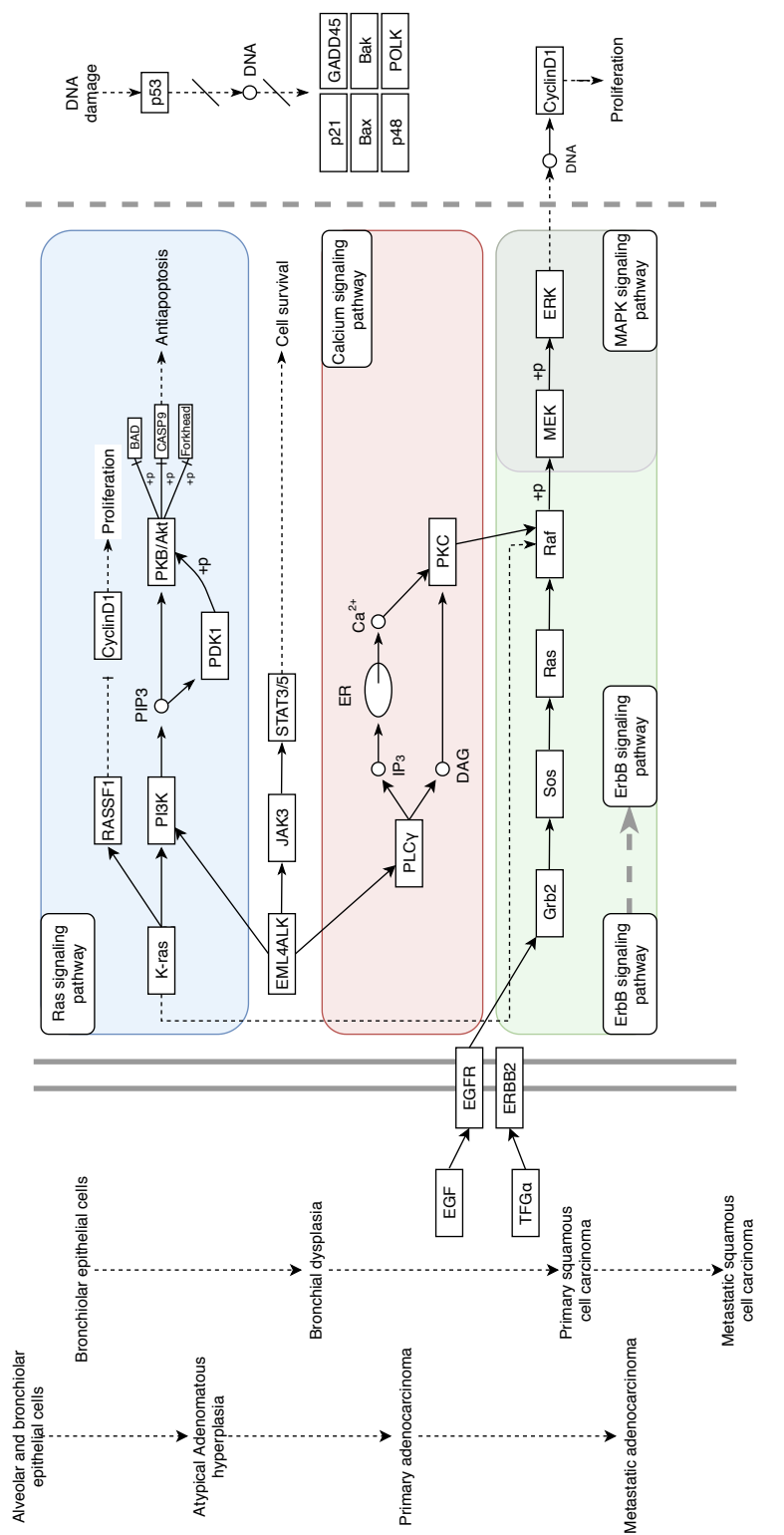
FIGURE 1.15: EGFR pathway in NSCLC. Activation of EGFR leads to homodimerisation and/or heterodimerisation, phosphorylation of specific tyrosine residues which leads to the recruitment of several proteins at the intracellular portion of the receptors. Phospholipase C and STAT transcription factors bind directly to the receptor, whereas Ras-Raf-MAPK pathway and PI3K pathway need several specific adaptor molecules. PI3K can also bind directly any of the ErbB partners of EGFR heterodimers.
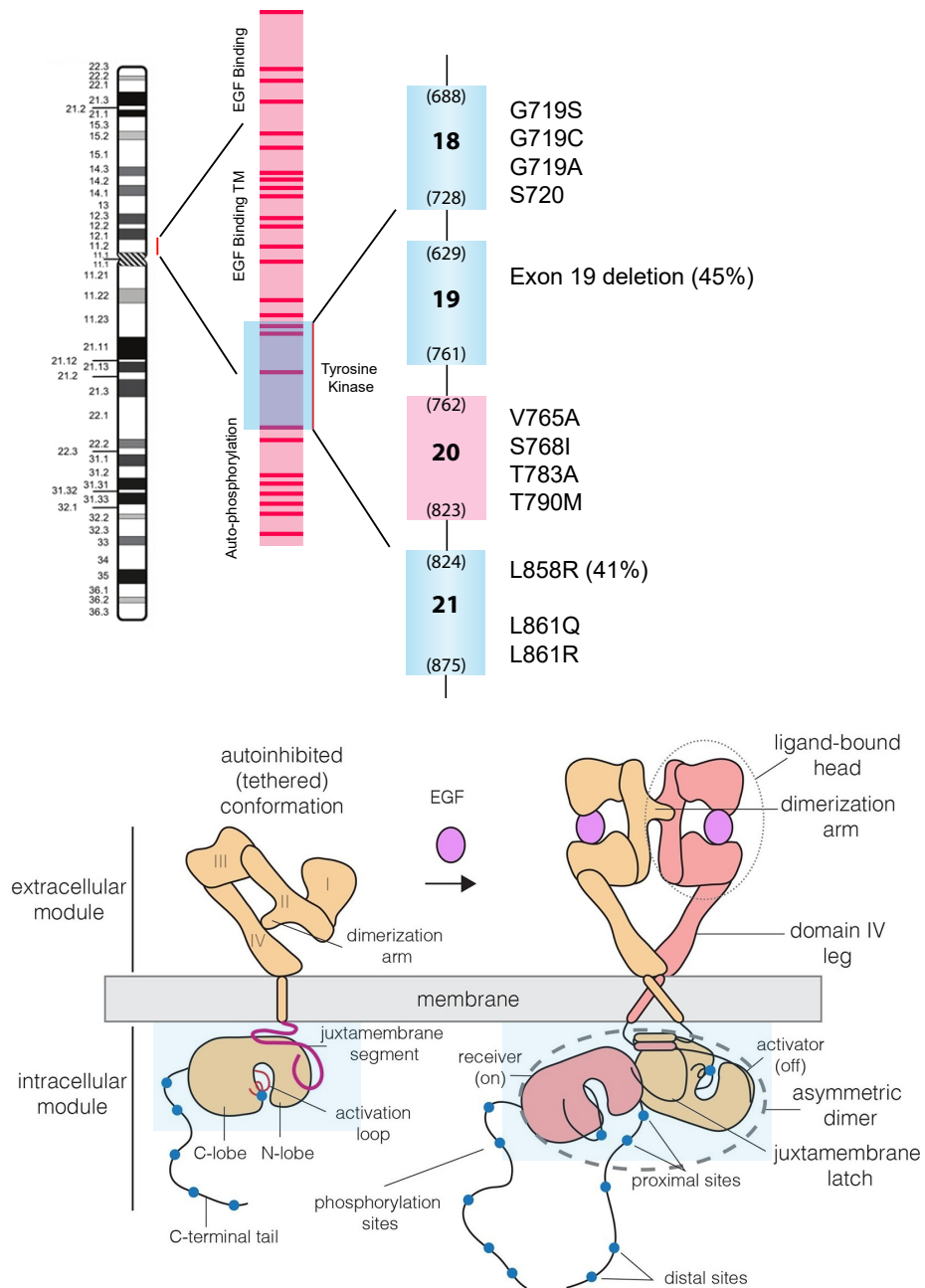
FIGURE 1.16: EGFR TK domain. This exon boundary map shows the location of regions within the EGFR-TK domain wherein mutations activate the kinase activity by a ligand-independent mechanism. Deletions in exon 19 and the point mutation of L858R are common activating mutations and these 'classical' mutations are associated with sensitivity to gefitinib and erlotinib in patients with NSCLC. T790M is a secondary point mutation found in tumours that were previously responsive to these agents but have developed acquired resistance. Below, is a schematic representation of EGFR-TK activation and EGFR kinase domain mutations. Upon binding of the extracellular ligand, the EGFR receptor dimerizes, leading to the activation of cytoplasmic TK activity [111]

1.9   Resistance to EGFR-TKIs

Many mechanisms have been postulated on how a EGFR-TKIs resistance occurs on tumours in response to first exposures of erlotinib and gefitinib [112]. Usually, tumours with insertions in exon 20 are insensitive to EGFR-TKIs and this occurs in 5% of the NSCLC patients [113, 114]. Primary resistance, which account more that 20% of the NSCLC cases is mediated by mutation in the KRAS protein [115, 116]. Rare mutations including these on PTEN, MEK and ALK, are responsible for the primary resistance to EGFR-TKI as well. T790M mutation which substitutes Threonine with Methionine at the 790 position of the exon 20 affects the ATP binding site of the TKIs and they consist the 50-60% of EGFR-TKIs resistance cases (Figs. 1.15, 1.16, 1.17) [110]. Another really interesting way of how tumour cells are gaining the resistance to the EGFR-TKIs is the phenomenon called 'kinase switch' in which EGFR mutated cells that are surviving from prolonged exposures to EGFR-TK inhibition develop resistance by becoming dependent on another kinase [117] (Fig. 1.18). Moreover, resistance to TKIs may occur when activating mutations occur downstream in BRAF and KRAS [96, 115].

Inactive EGFR with AMP   Inactive EGFR with erlotinib   Inactive EGFR with gefitinib
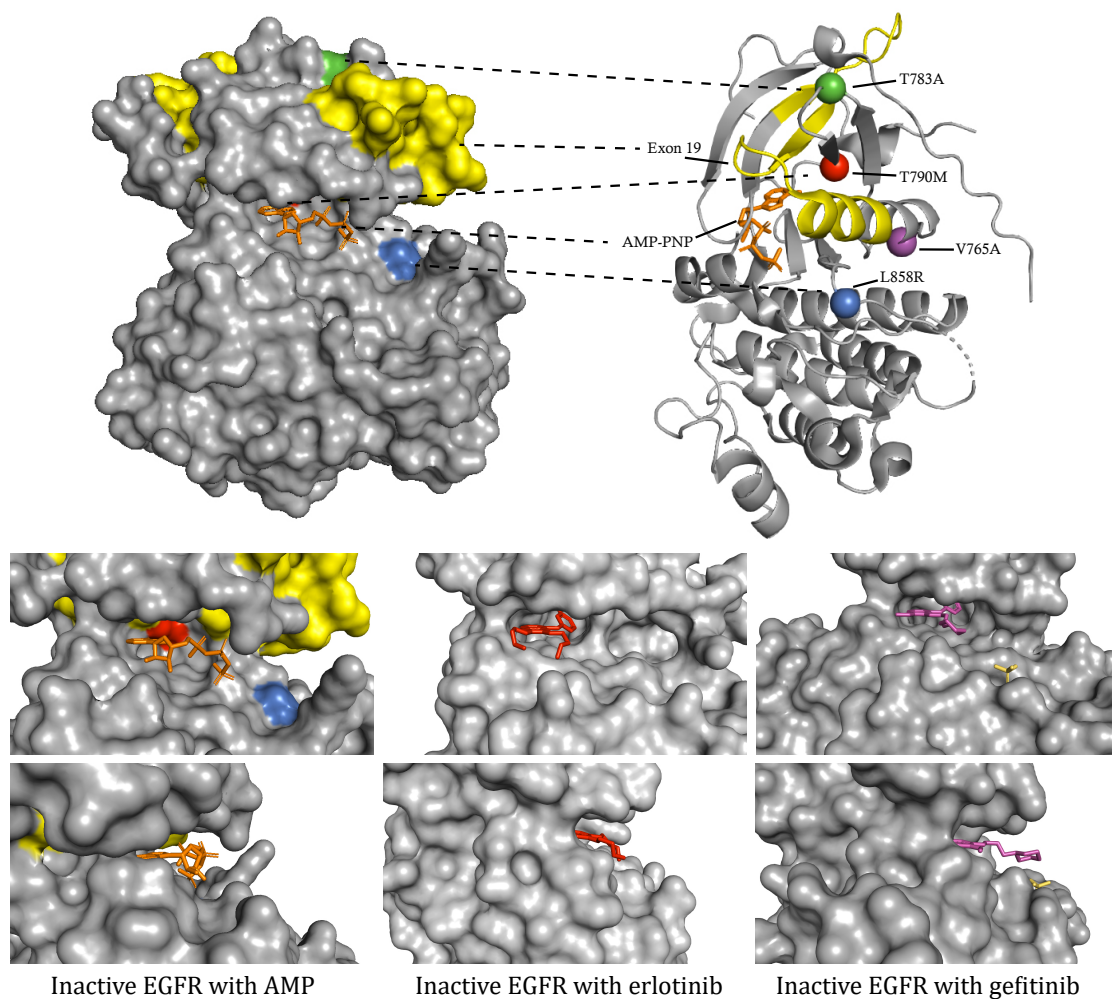
FIGURE 1.17: Crystal structures of inactive EGFR-TK domain with AMP, Erlotinib and Gefitinib. Activation of EGFR is achieved by a protein-protein interaction that forces a structural rearrangement of the kinase towards the active state. In this figure, kinase mutation points are colored and indicated (above) and the C-Helix movements are shown underneath.
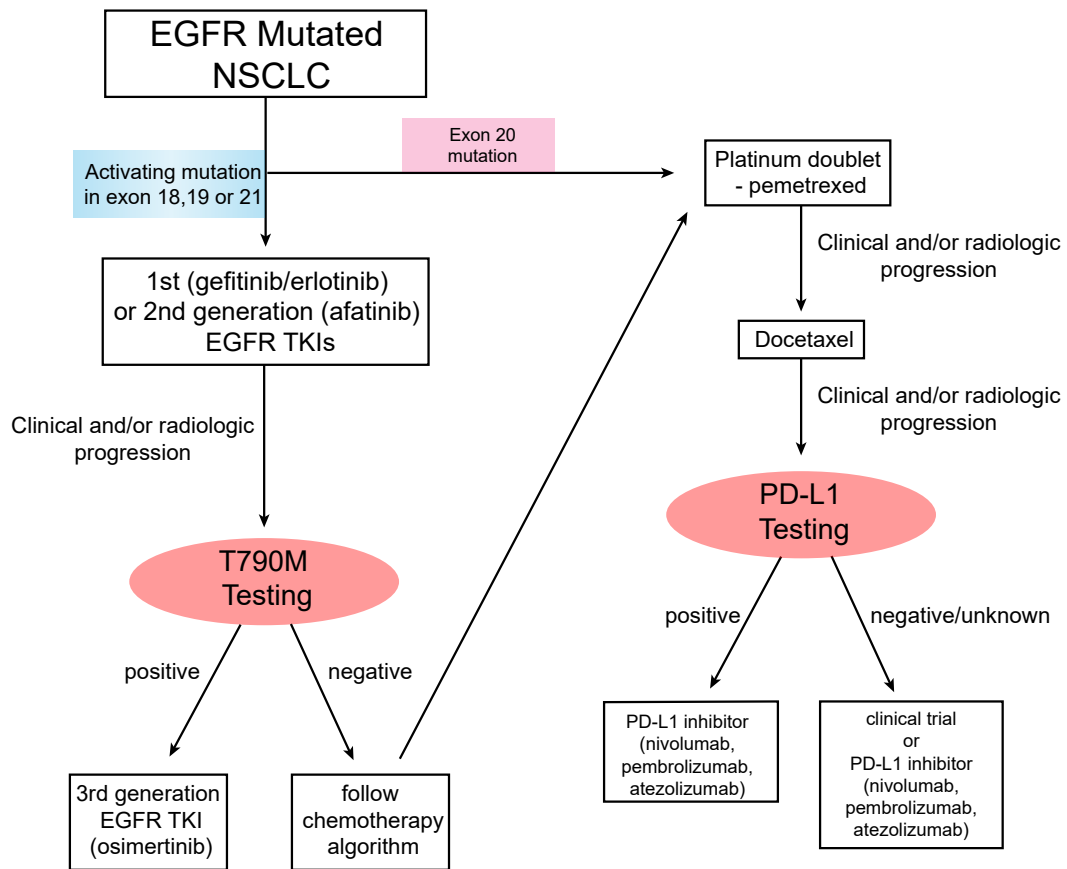
FIGURE 1.18: Proposed algorithm for the management of EGFR-mutated NSCLC. If oligo-progression occurs local therapy and continuation of tyrosine kinase inhibitor is considered. Then, a re-biopsy of the growing lesion is recommended if possible. Another alternative is testing plasma free DNA for key point mutations and then depending on the results, the appropriate therapy will be followed.

1.10 Long non-coding RNAs

Long non-coding RNAs (lncRNAs) are RNA molecules with more than 200nt length that do not encode proteins. About 98% of the transcriptome consists of lncRNAs whose function is not yet well understood. They are thought to have a wide range of functions in developmental and cellular processes, including gene expression, chromatin remodelling and modification, direct transcriptional regulation, splicing, editing, translation and degradation of the RNA and gene silencing with endogenous small interfering RNA (siRNA) [118–120]. LncRNAs found to be expressed in a variety of diseases including cancer [121, 122]. LncRNAs appear to play significant role in cancer genomics, creating the need for a validation of their cancer roles. It remains debated whether mutated lncRNAs can drive tumourigenesis, and whether such functions could be conserved during evolution. Apart from their role in tumourigenesis, they have been reported to mediate drug resistance in chemotherapies or even sensitisation of cancer cells [123]. For this reason, they are targets for identification and potentially be used as biomarkers and drug targets. Researchers have already identified key lncRNA pseudogenes that regulate oncogenes or tumour suppressor genes like PTEN and KRAS [124–127]. Other lncRNAs are extremely important for the genomic imprinting and they regulate epigenetic procedures [128, 129]. One of them is the lncRNA X-inactive specific transcript (XIST) that inactivates one of the two X chromosomes [130, 131].

1.10.1 X-inactive specific transcript (XIST)

LncRNA XIST is a pseudogene that acts as a major effector in the X chromosome inactivation process. It is expressed only on the inactive X chromosome providing a dosage equivalence between males and females [130, 132]. On the X chromosome there is a region called X inactivation centre (XIC) where many protein-coding and non-coding genes can be found with XIST being the first non-coding gene identified in this region [133]. Expression of this pseudogene is the essential step for the initiation of the X inactivation. Even if the XIST is going through splicing and polyadenylation like the mRNAs, yet, it remains untranslated (Fig. 1.19). XIST is coating and inactivates one of the X chromosomes, but lack of the XIST leads to failed inactivation and duplication of the gene on the other chromosome which leads to inactivation of that chromosome (Fig. 1.20 and Fig. 1.21) [133]. Another key lncRNA that mediates the X chromosome inactivation is TSIX. TSIX acts as a XIST repressor.

TSIX is the antisense RNA of XIST and their differential expression patterns define the activation or inactivation of the X chromosome (Fig. 1.21). Apart from the XIST-TSIX mechanism, XIST RNA and the nuclear matrix protein -heterogeneous nuclear ribonucleoprotein U (hnRNPu)- interact and upon depletion of hnRNPu, XIST is detached from the inactive X chromosome (Xi) and diffusely localised into the nucleoplasm [130, 134].
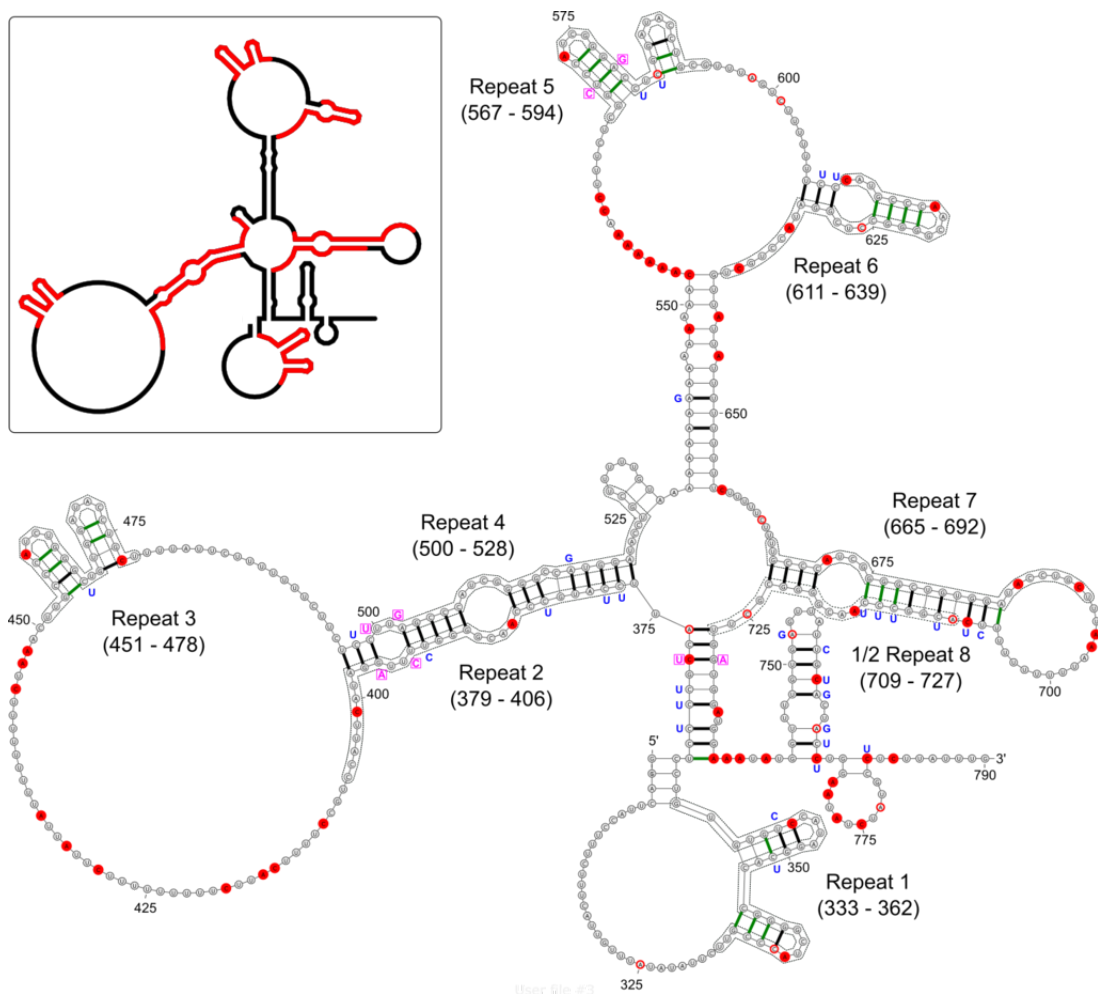
FIGURE 1.19: LncRNA XIST. Structure model of the repeat A (repA) region of XIST based on structural model and comparative sequence analysis. The data and model are taken from Fang et. al [135].
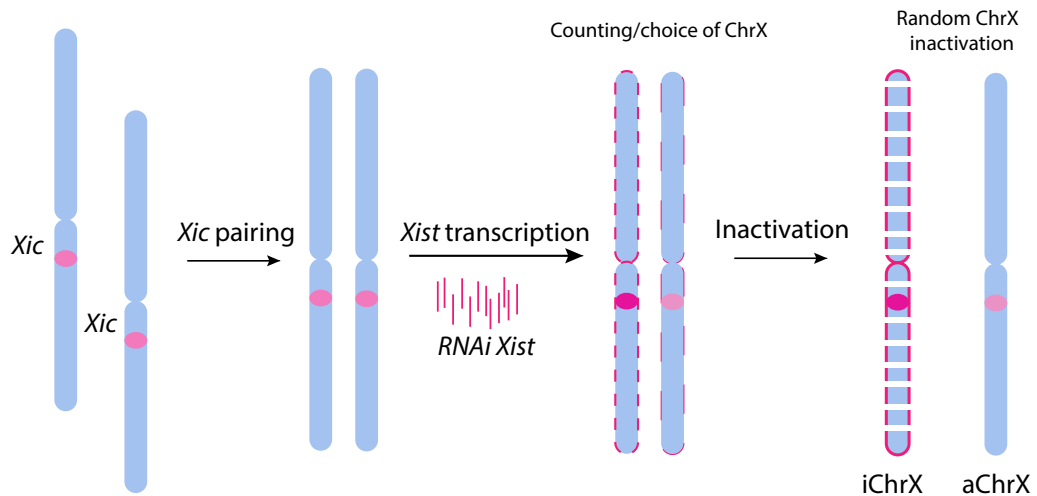
FIGURE 1.20: Chromosome X inactivation. LncRNA XIST it is expressed in high levels and it coats one of the two X chromosomes and it inactivates it to provide a dosage equivalence between males and females. It is being expressed in large numbers and randomly coats one of the two X-Chromosomes more than the other and it silences it. Xic: X-inactivation centre; iChrX: Inactive X-Chromosome; aChrX: Active X-Chromosome.
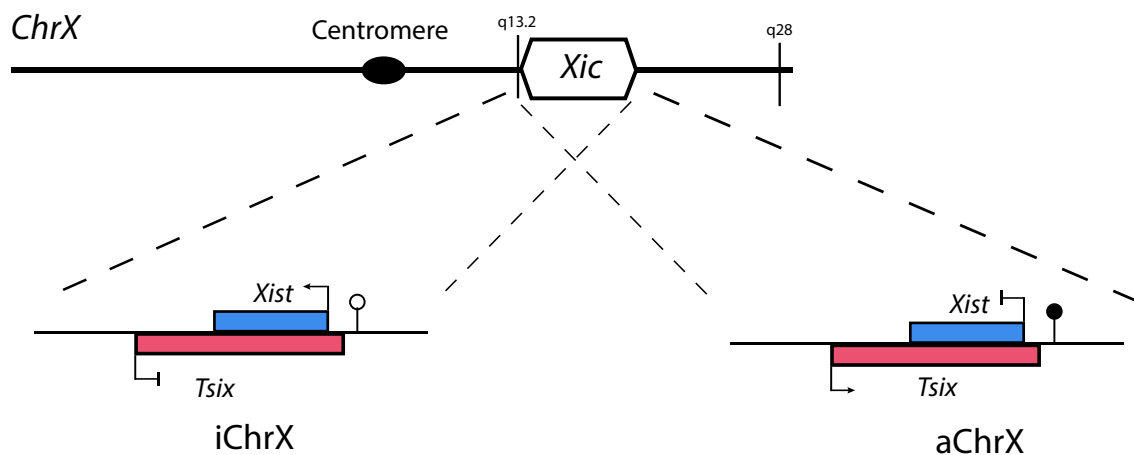
FIGURE 1.21: TSIX-mediated repression of XIST. TSIX is the antisense RNA of XIST and whenever one of them is more expressed than the other the X chromosome is either active or inactive.

1.10.2 XIST and Cancer

XIST seems to have tumour suppressing roles via inhibition of processes like migration, proliferation and invasion in cancers like glioblastoma and NSCLC [136, 137]. XIST is upregulated in NSCLC and after knocking it down, inhibition of cancer cell proliferation and invasion is possible [136]. Wang et al. showed that cell lines and patient samples of NSCLC overexpress XIST and they managed to confirm that XIST knockdown inhibits tumour growth in vivo. All these data indicate that XIST could be an important novel biomarker for the detection of NSCLC and inhibition of cell proliferation and invasion through gene editing and knocking down techniques [136, 138–140]. Recent studies have shown that XIST has an aberrant expression pattern in breast cancer [141], cervical squamous cell carcinoma [142], colorectal cancer [143], gastric cancer [144], glioma [145], hepatocellular carcinoma [146], nasopharyngeal carcinoma [147], NSCLC [138], pancreatic cancer [148], osteosarcoma [149], and ovarian cancer [150]. Furthermore, XIST was shown to regulate tumour cell migration, proliferation, and invasion in NSCLC [151, 152]. Wang *et al.* showed that cell lines and patient samples of NSCLC over-expressed XIST and shown that XIST knockdown inhibits tumour growth in vivo. Moreover, XIST exhibits oncogenic properties by regulating the miR-449a and B-cell lymphoma 2 (Bcl-2) gene in NSCLC [119]. The same study has shown that BRCA1 also influences the concentration of XIST on the Xi. Specifically, RNAi of BRCA1 decreases the concentration of XIST on Xi, and the reduction of BRCA1 by Cre-mediated excision also decreases XIST concentration on Xi [153]. Collectively these data indicate that XIST could be an important novel biomarker for the detection of NSCLC [154, 155].

1.11 Aims and Objectives

This study aims to evaluate the use of liquid biopsies in in the diagnosis of NSCLC. This will include identifying one or several robust liquid biomarkers and proposing a pipeline of tests to be implemented in to the clinical setting. Moreover this study will examine the role of XIST pseudogene in LC with functional and bioinformatic analysis to investigate further its function. The following objectives will be addressed in detail:

- Imaging flow cytometry analysis of CTCs in the blood of lung cancer patients undergoing surgery (i.e., pre-operatively vs post-operatively)

- Quantification of ccfDNA and ctDNA in the same cohort of patients

- Measurement of inflammatory cytokines in the same cohort of patients

- Measurement of EMT markers from whole blood

- Bioinformatic analyses on the role of lncRNA XIST in lung cancer

- Assessment of the effect of down-regulation of XIST in vitro

Chapter 2

Methodology

2.1 Tissue culture

2.1.1 Cell lines

A549, H1975 and MSTO-211H (Mesothelioma) human adherent epithelial cells were
used as in vitro models of human lung cancer. Cell line details summarised in Table
2.1.

TABLE 2.1: Characteristics of the tested cell lines

Cell line	Origin
A549 (CCL-185)	Lung-AC
H1975 (CRL-5908)	Lung-AC
MSTO-211H (CRL-2081)	Pleura, Lung

A549 cells were grown in complete DMEM (Dulbecco's Modified Eagle's Medium,
Gibco) with 10% fetal bovine serum (FBS, Gibco), 1% penicillin/streptomycin (Pen-
Strep, Gibco) and 1% L-Glutamine (Gibco). H1975 and MSTO-211H cells were grown
in RPMI + 1% L-Glutamine (Gibco), 10% FBS and 1% Pen/strep. Cell lines were
cultured at 37 ^{o}C - 5% CO_2 and subcultured when approached 80% confluency, ap-
proximately two 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 plastic ware was sterilised by autoclave prior to use.

2.1.3 Thawing Cryo-preserved Cells

When growing cells from stored frozen liquid nitrogen stocks, media was first stored in T-25 cm^2 cell culture flasks 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 from liquid nitrogen, defrosted in a water bath at 37 ^oC, and immediately transferred to a 15 mL falcon tube, spun down. The supernatant was removed, and the pellet was then suspended in 1 mL of their appropriate media and transferred in a T-25 cm^2 cell culture flask for 24h. At 24h, the media was replaced with fresh media that had been warmed to 37 ^oC, 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 4 mL TrypLE^{TM} Express (Invitrogen) per $\sim$75 cm^2 growth surface area and manually disturbing the flask to detach adherent cells. The cells were re-suspended in 4 mL of media, the suspension (8 mL) moved to a 15 mL centrifuge tube and centrifuged for three minutes at 1.000 RPM to form a cell pellet. The supernatant was removed and the pellet was then suspended in 4 mL of their appropriate media and 1 mL transferred to each of four new T-75 cell culture flasks already containing 10 mL of media. All media used in the subculturing process was first warmed to 37 ^oC. All four cell lines required a 1 in 4 splits approximately two times per week.

2.1.5 Cryopreserving Cells

When no longer needed, 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 4 mL TrypLE^{TM} Express (Invitrogen), per 75 cm^2 surface area, and manually disturbing the flask to detach adherent cells. The cells were resuspended in 4 mL of media, the suspension moved to a 15-mL centrifuge tube and centrifuged

for three minutes at 1.000 RPM to form a cell pellet. The supernatant was removed and 4 mL of Recovery^TM Cell Culture Freezing Medium (ThermoFisher) was used to resuspend the cells. 1 mL (x4) of the cell suspension was added to a cryovial and cooled slowly, first to - 80 ^{o}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 x 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.

2.1.7   3D Cell Culture

The cell lines A549 and H1975 were tested if they form spheroids with the hanging drops technique using the lid of a petri dish in a variety of cell densities and culture days. Both A549 and H1975 were plated in a Poly-HEMA coated 24-well plate to form spheroids and the growing medium was being changed every two days until the staining step. For the low adhesion dishes – coating with Poly-HEMA (Poly 2-hydroxyethyl methacrylate; Sigma, cat.# P3932), I dissolved 2 g Poly-HEMA powder in 100 ml 95% EtOH by rotating several hours at 65°C or leaving it overnight in a waterbath at 65°C to obtain a 20mg/mL solution. Then, i coated the plates with 4 ml/10 cm plate. For 6-well plates or 35 mm dishes, i used 700 μL/well. After drying

overnight at 37°C, the plates can be wrapped in parafilm and stored indefinitely at RT prior to use.

2.2 Clinical Samples

2.2.1 Whole Blood

Blood samples from lung cancer patients were collected from consented patients (with appropriate NRES ethical approval in place) at The Royal Brompton & Harefield NHS Trust and the Mt Vernon (The Hillington Hospitals NHS Foundation Trust). Samples were also collected from patients with non-cancer diseases and normal healthy volunteers to act as controls (Table 2.2).

Lung cancer samples from Harefield

TABLE 2.2: **Patient demographics for Harefield.** Demographics and histology of patients recruited for Harefield study

Variable	Value
Total	54
Mean age ($\pm$SD)	67.5$\pm$9.5
Males/Females(%)	23 (45) / 31 (55)
Pathology - no. (%)	
Adenocarcinoma	38 (70)
SqCC	16 (30)
Control	10
Males/Females(%)	5 (50) / 5 (50)
Mean age ($\pm$SD)	45$\pm$5.5
Staging - no. (%)	
I	35 (65)
II	10 (18.5)
III	7 (13)
IV	2 (3.5)
Operation - no. (%)	
Thoracotomy	12 (22)
VATS	42 (78)

All blood samples were collected in Ethylene-diamine-tetra-acetic acid (EDTA) tubes (for up to 3 days) or cfDNA collection Roche tube (Roche Diagnostics, Mannheim, Germany cat.# 07785666001) for up to 6 days to prevent coagulation of the blood. The samples were inverted 10 times, and 3 mL of blood were transferred in 3 different 1.5 mL tubes for further DNA work. One EDTA tube was spun down at 2.500 rpm for 15 minutes to separate the plasma and 2 mL of the plasma layer is then aspirated, carefully not to disturb the red blood sediment, in to a 2 mL Eppendorf and stored at - 80 ^{o}C, until use. If RNA extraction is not done immediately, 0.5 mL of whole blood was added to 1 mL of RNAlater$^®$ (Life Technologies) in an Eppendorf and stored at -80ºC (Fig. 2.1).

FIGURE 2.1: **Blood process flowchart**

2.3 Isolation and imaging of cells

2.3.1 ImageStream

Protein expression and localisation was investigated using ImageStream (Amnis) high resolution flow cytometry. Expression and cellular location of pan-cytokeratin AE1/AE3 (PanCK, Abcam), Thyroid transcription factor-1 TFF-1/NX2.1 (Novus Biologicals), Vimentin phospho-S38, Epithelial cell adhesion molecule, EpCAM (Abcam), Recombinant Anti-Cytokeratin 7 antibody (CK7, Abcam) and the leukocyte marker CD45, was assessed using Image Stream™ (Amnis) imaging flow cytometry. The cells that have been stained with EpCAM primary antibody, then tagged with a secondary Alexafluor® 488 (AF488) antibody, emitting a green fluorescence.

2.3.2 Preparation of cell lines

Cells were cultured in T-75 cm^2 tissue culture flasks until approximately 90% confluent. Media was aspirated, and the cells were washed briefly in PBS (Gibco). The cells were incubated with 4 mL of TrypLE™ Express (Invitrogen) per 75 cm^2 growth surface area and the flask was manually disturbed to detach adherent cells, which were resuspended in 4 mL of pre-warmed medium. The cell suspension was transferred into a 15 mL centrifuge tube and centrifuged for 3 minutes at 1.000 rpm to form a pellet. Supernatant was removed, and cells were resuspended in 5 mL of pre-warmed PBS (Gibco) a further time to remove debris. Samples were fixed immediately as described in section 2.3.5.

2.3.3 Proof of principle experiments

H1975 and A549 cell lines were cultured as above and then the cell suspension was resuspended in different cell concentrations which were 5.000, 2.500, 1.200, 600, 300 and 150 cells/ml. The cells then were transferred and mixed in 1 ml blood from donor. The protocol of fixation and permeabilization can be found in detail at 2.3.5 - 2.3.6.

2.3.4 Preparation of blood samples

One mL of whole blood from patient samples was decanted from the EDTA blood tubes into a 15 mL falcon tube and mixed with 9 mL of red blood cell (RBC) lysis

buffer (G Biosciences), inverted 8 times and incubated for 10 minutes with gentle agitation. The solution was then spun for 10 minutes at 2.500 rpm, the supernatant removed and 3 mL of RBC lysis buffer added to resuspend the pellet, before incubating for 10 minutes at RT with gentle agitation. One more centrifuge step followed for 10 minutes at 2.500 rpm, and again the supernatant aspirated. The pellet was then washed in 1.5 mL of PBS, and solution moved to a 2 mL microcentrifuge tube and spun at 3.600 rpm for 3 minutes. Samples were then fixed immediately as described in section 2.3.5. For spiking in experiments, 1 mL of healthy blood was used, and an approximate number of cells calculated and spiked in accordingly. Samples were then processed as per patient samples described above in this section.

Fixing

All cell pellets (cultured cells and patient samples) were then transferred to a 2 mL microcentrifuge tube and resuspended in 1 mL of ice cold 4% paraformaldehyde (PFA, Sigma) for 5 minutes to cross-link proteins in the cells. The cell suspension was centrifuged for 3.5 minutes at 3.500 rpm and the PFA removed. The cells were washed with pre-warmed PBS two times and centrifuged for 3.5 minutes at 3.500 rpm between each wash. All cells were immediately stained or a permeabilization and blocking step followed.

Permeabilization

One ml of permeabilising buffer of 0.5% TritonX 100 in PBS was added in each tube and incubated for 10 minutes on ice. Then, they have been centrifuged for 3.5 minutes at 3.500 rpm and the supernatant removed to wash them again with PBS and re-centrifuge and discard the supernatant again.

Staining

The cells were incubated in blocking buffer (10% FBS, Gibco, in PBS) for 60 minutes with gentle agitation in RT. Cells were centrifuged for 3.5 minutes at 3.500 rpm and the blocking buffer was removed. The cells were then incubated in the appropriate primary antibody, details listed in Table 2.3, diluted in blocking buffer overnight at 4^oC with gentle agitation. Following primary antibody incubation, cells were centrifuged for 3.5 minutes at 3.500 rpm and the antibody aspirated. The cells were

resuspended in washing buffer (0.1% Tween in PBS) to remove any remaining antibody and centrifuged again for 3.5 minutes at 3.500 rpm the PBS was removed and the cells incubated in secondary antibody (Table 2.3) diluted in blocking buffer for 1h with gentle agitation. From this step, onward, the cells were protected from light as the fluorophore conjugated to the secondary antibody is light sensitive. After secondary antibody incubation, the cells were centrifuged for 3.5 minutes at 3.500 rpm and the secondary antibody removed. The cells were washed once in PBS (Gibco) for 3.5 minutes at 3.500 rpm to remove any remaining antibody. PBS was removed, and the cells were resuspended in 100 μL Accumax (Innovative Cell Technologies) to dissociate any cellular aggregates. 0.5 μL of Draq5 DNA at 37 ^{o}C (Biostatus Ltd) nuclear stain was added before visualisation on the ImageStream. All the data files were then analysed on the IDEAS$^®$ software.

TABLE 2.3: Details of antibodies used for the Imagestream.

Primary Antibody	Dilution	Species	Supplier	2nd Ab	Dilution	Wavelength
EpCAM	1:100	Mouse	Abcam	Anti mouse, alexafluor	1:1000	488nm
AE1/AE3	Rabbit	Rabbit	Abcam			488nm
Vimentin	1:100	Mouse	Abcam			594nm
CK7	1:100	Mouse	Abcam			488nm
TFF-1/NKX2.1	1:100	Mouse	NovusBio			488nm

2.4 Staining and Imaging of the spheroids

After 3 days of incubation as a hanging drop, the spheroids were transferred carefully into a 15 mL falcon tube and the medium was removed without touching the spheroids. Then, 1 mL of cold 4% PFA was added and left for 30 minutes on ice. The spheroids were then washed 2 times for 5 minutes with 1 mL PBS and blocked with 1 mL of 10% FBS for 30 minutes at RT and washed 3 times with 0.1% Triton-X and left for 5 minutes at RT. Following the washing step, the spheroids were incubated with the appropriate primary antibody as 1 μL/100 μL of BSA for 90 minutes at RT and then washed 3 times for 5 minutes with Triton-X. If there was a need for secondary antibody, 1 μL/ 1000 μL FBS of the antibody was added for 60 minutes in RT and then washed as previously. After the staining, 100 μL of 1 μg/mL DAPI/PBS was added in the pellet and left for 5 minutes in RT and washed 2-3 times with PBS and finally with ddH$_2$O, then the spheroids were transferred on a slide for observation. Similar protocol was followed for the staining of the cover slips used for the siRNA experiments.

2.5 ImageStream Analysis

The IDEAS® software is used to analyse samples captured with the ImageStreamX. The software firstly shows all the captured objects (cells, debris, etc.) and allows the user to gate them (based on manual examination) and tag them as single viable cells, focused cells and stained cells and the analysis can then can be done only on the single cells. It is worth mentioning that the INSPIRE® capture software allows for single cells to be isolated and exclusively identified; however, careful examination of collected cells is important, this is done to exclude small clusters of cells that might have been accidentally included in the original capture. The next step is to narrow the analysis only to focused cells which have a clear cell structure and morphology. In the beginning I analysed results from A549, H1975 and MSTO-211H cells alone with the EpCAM, TTF-1/NKX2.1, AE1/AE3 PanCK and CK7 antibodies. Following the detection of each cell line alone, I conducted proof of principle experiments be spiking different densities of the cell line in 1 ml of blood and tried to detect them. For the analysis of the cell size, morphology, shape and signal strength I used the features that the IDEAS® software provide following carefully the guidelines for the analysis of each feature.

2.5.1 Size Features in IDEAS®

Area Feature

By using the 'Area Feature' on IDEAS software, we can measure the Area of a mask in pixels and according to the calculation described in the manual 1 pixel = 0.25 μm^2. That means, that if we have a cell with a mask that include 2000 pixels this equals to 500 μm^2 (Fig. 2.2). For the following analysis, I used the Width of the circular, single and focused cells and I divided it by 0.25 μm^2 = $\frac{1}{4}$.
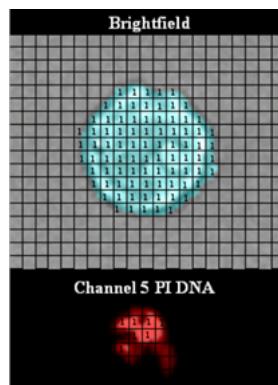
FIGURE 2.2: **Area feature by pixel measurement**

Diameter Feature

The Diameter feature provides the diameter of the circle that has the same area as the object. The accuracy of the diameter is highly dependent on a close-fitting mask and roundness of the cell (equation 2.1).

$$Area = \pi \times r^2 = 2 \times \sqrt{\frac{Area}{\pi}} \tag{2.1}$$

2.5.2   Advanced Imaging flow cytometry analysis

Data acquisition in imaging flow cytometry

The fluorescence image- based method for quantifying nuclear translocation described in this project relies on the spectral isolation of PanCK, CK7, TTF-1 and Ep-CAM images from ImageStream analysis. Cells probed with fluorescently labelled antibody to each target excited with a 488nm laser light and imaged on a time delay integration (TDI) CCD camera. Light is passed through a spectral decomposition element that directs different spectral bands to spatially distinct channels on the TDI CCD detector. With this technique, a single cell image is optically decomposed into a set of sub-images, each corresponding to a different colour component and each with an identical spatial registry of pixels from channel to channel. Images of fixed cells collected on the ImageStream were analysed using ImageStream Data Exploration and Analysis Software (IDEAS) [156, 157]. Because fluorescent probes have broad emission spectra, spectral compensation is digitally performed on a pixel-by-pixel basis prior to data analysis. After compensation, similarity analysis was done on in-focus single cell images. In-focus single cells were classified based on their small bright field area, high bright field aspect ratio (width to height ratio) and high nuclear contrast (as measured by the gradient max feature). Gradient max measures the magnitude of the largest pixel intensity slope over the whole image and in focus images with sharp edges at their borders tend to have high max gradient scores. The degree of nuclear translocation can be assessed in a qualitative manner by visually comparing a cell's nuclear fluorescence image to the pattern of fluorescence produced by the antibodies tested [158–161]. The INSPIRE acquisition software generates data in the form of a raw image file (.rif file) which can then be directly loaded into IDEAS for further analysis. The IDEAS software analyses captured 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 [162].

When the .rif file is loaded into IDEAS, a compensation matrix generated from the fluorescence control experiments can be used to produce a compensated image file (.cif file). In the IDEAS environment, the user can plot features derived from the bright-field, dark-field and fluorescence single cell images in the form of histograms or bivariate scatter plots. Gating can be performed using these plots to generate

subpopulations that can be then be studied in further detail. The plots, gating and sub-population information from a session can then be saved as a data analysis file (.daf file). It is also possible to generate individual .tiff images from each channel for each cell to analyse outside of the IDEAS framework. IDEAS is especially suited for visually inspecting the data irrespective of the further analysis pipeline the user wishes to perform. The important first steps of identifying out-of-focus cells and removing debris or multiple cells are best carried out using this software platform. IDEAS suggest using a measure of the gradient RMS of the bright-field image to determine the focus quality of each cell. By gating the high values in the gradient RMS histogram, a sub-population of in-focus cells is defined. The next step is to identify the single cells by plotting the cell mask aspect ratio versus the cell mask area. A 2D gating window is defined to select cells with an aspect ratio close to 1, which removes clumped cells, while also rejecting high and low areas, which removes debris. Once subpopulations are identified via gating they can be saved as a new .cif file in IDEAS [161, 163–166].

Fluorescently labelled or unlabelled cells in 100 μL solution were run through the ImageStream and the data were acquired using the INSPIRE software. The INSPIRE software generates data in raw image file (.rif) format that can be analysed into IDEAS and produce a compensated image file (.cif). During the analysis the user can generate plots for subpopulations and study them further and save the file as a data analysis file (.daf). It is also possible for the user to generate individual .tiff images from each channel of a selected population to analyse the cells outside the software's framework. Firstly, an out-of-focus exclusion was made to remove any debris or clusters of cells from my selection by gating accordingly the focused and single cells. Once the subpopulations were identified via getting they were being saved as a new .cif file in IDEAS.

From data acquisition to high-throughput data analysis

For the analysis, a simple pipeline introduced by Hennig et al. has been used [167]. Using CellProfiler, a cell image analysis software that operates image-wise I could detect hundreds or thousands of objects in an image. I accessed the .cif file with the Bio-Formats .cif file reader that Carpender and his team provided to Bio-Formats [168]. The protocol I used for imaging flow cytometry is the following:

1. Load .cif file in CellProfile (drag&drop)

2. Segment images and extract hundreds of features per cell channel using Cell-Profiler using the pipeline that can be found at cellprofiler website.

3. Analysis

2.6 RNA isolation

Total RNA 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 ^{o}C when not in use and kept on ice at all other times to prevent degradation.

2.6.1 Cell lines

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 4 mL TrypLE^{TM} Express (Invitrogen) and incubated for 3 minutes at 37 ^{o}C to allow cells to detach. The cell solution was collected in to a 15 mL falcon tube and spun at 1.500 rpm for 5 minutes. Supernatant was then aspirated and 500 μL 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 x g to remove cellular debris and shear DNA. The filtered lysate was then mixed with 500 μL 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 x 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 x g to remove any impurities including proteins, cellular debris and salts. A final centrifugation of 12.000 x g for 2 minutes 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 50 μL of pure water (elution solution) to the membrane and centrifuging at 12.000 x g for 60 seconds. The RNA was stored at -80 ^{o}C until further use.

2.6.2 Blood samples

For the RNA extraction from RNA later samples,the Ribopure$^{\text{TM}}$ RNA Purification Kit (cat.# AM1924) by ThermoFisher scientific was used. RNAlater samples were removed from -20 ^{o}C and thawed. Then, the samples were spun down at full speed (13.000 rpm) to remove the supernatant. 800 μL of lysis buffer and 50 μL of sodium acetate solution was added to each sample followed by vigorous vortex to obtain a homogenous solution and 500 μL of Acid-Phenol Chloroform were added to the samples and incubated at RT for 5 minutes. After a brief vortex step the samples were spun down at full speed for 1 minute and the supernanatant (aqueous phase) was transferred into a new 2 mL tube. Afterwards, 600 μL of 100% EtOH were added in each sample and briefly vortexed and 700 μL of the mixture was transferred into the filter cartridge and spun down for 10 seconds at full speed. This step was repeated until all the mixture was passed through the filter cartridge. In addition, 700 μL of Wash Solution 1 was added to the filter cartridge and centrifuged for 10 seconds to pass the solution through. The flow-through was discarded and 700 μL of Wash Solution 2/3 was added twice and discarded as Wash Solution 1. A dry spin was followed to ensure that the filter was dry and the filter cartridge was transferred into a new collection tube. Lastly, 100 μL of preheated Elution Buffer was added straight on the filter, allowed to incubate for 1 minute at RT and spun down in order to elute the RNA and store it in -80 ^{o}C until further use.

2.6.3 Quantification

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

2.6.4 cDNA synthesis

Complementary DNA (cDNA) was synthesised from extracted RNA for use as template in qPCR experimentation. All surfaces and equipment were cleaned thoroughly with 70% IMS before and throughout cDNA synthesis work. Random primers were chosen over oligo-dT to avoid 3′ poly-A tail weighting. Oligo(dT) primers amplify only mRNAs containing a poly(A) tail, since that is where the primer binds to promote reverse transcription. Random primers amplify most RNA species and they are great for synthesizing large pools of cDNA and they are also ideal for non-polyadenylated RNA. RNA volume input was calculated by the following: RNA input = desired cDNA concentration/RNA concentration (ng/ 500 μL).

2.6.5 Applied Biosystems - High capacity cDNA synthesis kit

RNA input was calculated and the appropriate amount, made up to 10 μL with pure H_2O, was mixed in one 0.6 mL, thin walled sample tube for each sample. The reverse transcriptase mastermix was made up using reagents and volumes listed in Table 2.4.

TABLE 2.4: Reagent volumes for reverse transcriptase for one sample for use with the High-throughput cDNA reverse transcriptase kit.

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

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

Quantitative PCR (qPCR)

qPCR (quantitative polymerase chain reaction) was used to assess relative gene expression (RQ) in several 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.

2.6.6 Power SYBR® Master Mix

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 2.5):

TABLE 2.5: Volume of components for primer mixes in qPCR experiment with Power SYBR® Master Mix (Life Technologies).

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

The above reagent mix was made up separately for each primer set. 15 μL of the appropriate mix and 5 μL 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™ analysis (completed previously), reference gene, 18S RNA, was shown to be expressional stable and thus was included as endogenous control 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 at

the bottom of the wells. The sample plate was subjected to a temperature protocol consisting of the following (Table 2.6).

TABLE 2.6: The thermal protocol of a qPCR experiment using Power SYBR® Master Mix on a Quantstudio 7 thermal cycler (Life Technologies).

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

2.7 Primers

All primers used for this project were sourced from the literature or previous studies of the lab and ordered from Sigma Aldrich. Primers were supplied lyophilised and were centrifuged and then re-hydrated according to manufacturers guidelines. All primers used, are shown in the following table (Table 2.7).

Primer	Strand	Sequence
18S RNA	Forward	ATGGCCGTTCTGAGTTGGTG
	Reverse	CGCTGAGCCAGTCAGTGTAG
RHOH	Forward	GAGAAGTAACATTCTGCAAATCGC
	Reverse	AGCACACGCCATTCAGCAAG
PRKCQ	Forward	CTTGTGGCAGCTTTGGATGT
	Reverse	CGTTTCTGACGCACATGTTT
CDKN1A	Forward	GCAGACCAGCATGACAGATTT
	Reverse	GGATTAGGGCTTCCTCTTGGA
SFT1AP	Forward	CAGCATTCCAGGTGGGCTTT
	Reverse	CCTTGTTTGGCTTACTCGTGC
NrCAM	Forward	TTGTGCAAAGAGGGAGCATG
	Reverse	GGGCAGTTCCCTGTTGTCCT
MYC	Forward	TGGACATCCGCAAAGACCTGTAC
	Reverse	TCAGGAGGAGCAATGATCTTGA
EpCAM	Forward	GAAGGCTGAGATAAAGGAGATGGG
	Reverse	TTAACGATGGAGTCCAAGTTCTGG
Vimentin	Forward	TGTCCAAATCGATGTGGATGTTTC
	Reverse	TTGTACCATTCTTCTGCCTCCTG
N-Cadherin	Forward	GCGTCTGTAGAGGCTTCTGG
	Reverse	GCCACTTGCCACTTTTCCTG
E-Cadherin	Forward	CATTGCCACATACACTCTCTTCT
	Reverse	CGGTTACCGTGATCAAAATCTC
XIST	Forward	AGGTCAGGCAGAGGAAGTCA
	Reverse	CTCCCGATACAACAATCACG
	Reverse	CAGTGGGTGTTGCAGGATG

TABLE 2.7: **Primer sequences of all genes.** Primer sequences for all genes tested for this study. All primers were synthesised by Sigma-Aldrich.

2.7.1   Gel electrophoresis

Primers were validated by running the PCR products on a 2% agarose gel. 2g of agarose (Fisher Scientific) were added to 100 mL of 1 $\times$ tris-borate EDTA (TBE, 89 mM Tris-borate, 2 mM 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 ^{o}C, and 5 μL 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$\times$TBE buffer. 2 μL of DNA loading buffer was mixed with 10 μL qPCR product to aid loading and visualisation of migration through the gel. The mixture was added to each appropriate well. In a separate well 5 μL of 1kb Plus DNA Ladder^{TM} (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^{TM} 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.

2.8   DNA Extraction

2.8.1   QIAamp$^®$ DNA Blood Mini Kit

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 RT. 20 μL of proteinase K was added to the bottom of a clean 1.5 mL Eppendorf tube, followed by addition of 200 μL of plasma sample, and 200 μL of buffer AL, the solution was then pulse mixed for 15 seconds. The sample was then incubated for 10 minutes at 56 ^{o}C in a water bath. 200 μL of 100% ethanol was then added to the samples, followed by pulse vortexing for 15 seconds. The solution was then spun through a QIAamp Mini spin column with a collection tube attached, at 8.000 rpm for 1 minute, discarding the filtrate. 500 μL of buffer AW1 was added to each spin column and spun at 8.000 rpm for 1 minute. The flow through was discarded and the spin column placed in a new collection tube. 500 μL of buffer AW2 was added, and samples spun for 3 minutes at 14.000 rpm. The flow through was discarded and the spin column spun for an additional 1 minute at 14.000 rpm

to eliminate carryover of buffer AW2. The spin column was then placed in a new 1.5 mL Eppendorf tube, and 50 μL of buffer AE added, incubated for 1 minute at RT before being spun at 8.000 rpm for 1 minute. The DNA elute was stored at -20 ^{o}C until further analysis.

2.8.2 Maxwell® RSC ccfDNA Plasma Kit by Promega

The Maxwell® RSC ccfDNA Plasma Kit is used with the Maxwell® RSC or Maxwell® RSC 48 Instruments to provide an easy method for efficient, automated purification of circulating cell-free DNA (ccfDNA) from human plasma samples. The Maxwell® RSC and Maxwell® RSC 48 Instruments are supplied with pre-programmed purification procedures and are designed for use with predisposed reagent cartridges, maximizing simplicity and convenience (Fig. 2.3). The Maxwell® RSC Instrument can process up to 16 samples ranging in volume from 0.2–1.0 ml in approximately 70 minutes. The purified DNA can be used directly in a variety of downstream applications such as PCR. ccfDNA was extracted using the Maxwell® RSC ccfDNA Plasma Kit (cat.# AS1480) following step by step the manufacturers' protocol using 400 μL of plasma.

FIGURE 2.3: Maxwell RSC instrument containing several cartridge samples for purification.

2.8.3   DNA concentration measurement

The measurement of the DNA concentration for the variety of the experiments was made with either Nanodrop of Qubit.

Qubit$^{\text{TM}}$ 1X dsDNA HS Assay Kit

For the measurement of ccfDNA the Qubit$^{\text{TM}}$ dsDNA HS Assay kit (cat.# Q33230) was used and performed at RT with the samples being protected from light and heat (avoiding holding the tubes on hand before measurement). Before each run, new calibration was made if there was any temperature fluctuation (seasons) and the calibration curved was used to generate the quantification results. For both the standard curve and sample preparation the Qubit$^{\text{TM}}$ 1X dsDNA HS Assay kit protocol has been followed step by step.

To calculate the concentration of the samples the following equation was used:

$$\text{Concentration of sample} = \text{QF value x } 200/\text{X}$$

where

QF value = the value given by the Fluorometer

x = the number of microliters of sample added to the assay tube

The equation generates a result of ng/ml units.

QuantiFluor$^{®}$ dsDNA System by Promega

To accurately determine the concentration of the eluted ccfDNA I used Fluorescence technique. The sensitivity of dsDNA-specific dyes makes them a better choice for quantitating ccfDNA. The measurement of all ccfDNA extracted with the Maxwell$^{®}$ RSC ccfDNA Plasma kit was made by the QuantiFluor$^{®}$ dsDNA System (cat.# E2670) and the Quantus$^{\text{TM}}$ Fluorometer (cat.# E6150) and the samples were prepared in thin-walled 0.5 ml PCR tubes (cat.# E4941). The Quantus$^{\text{TM}}$ Fluorometer operating manual TM396 and QuantiFluor$^{®}$ dsDNA System technical manual #TM346 (promega.com/protocols) was used to measure the concentration of 2 μL of each sample after being mixed with 198 μL of QuantiFluor$^{®}$ dsDNA Dye working solution, vortexed and kept in dark till measurement.

2.9 ALU repeats

The quantification of the DNA fragments was performed by qPCR at the QuantStudio 7 (Life Technologies) which quantified and amplified the shorter and longer fragments. The Alu repeats are the most abundant repeat sequence in the human genome, hence they were used a target of the PCR. Sequences for the ALU 115 and 247 described by Umetani et al., represent the total free plasma DNA and the amount of DNA released from apoptotic cells respectively (Table 2.8 and Table 2.9) [169]. The quantitative values from the 115-bp primers represent the total level of ccfDNA (ng/ml) while the ratio of longer to shorter fragments (ALU247/115) shows the integrity of ccfDNA in each sample.

TABLE 2.8: Primer sequences for Alu repeat sequences 115 and 247 (Sigma-Aldrich)

Primer	Forward Sequence	Reverse Sequence
ALU 115	CCTGAGGTCAGGAGTTCGAG	CCCGAGTAGCTGGGATTACA
ALU 247	GTGGCTCACGCCTGTAATC	CAGGCTGGAGTGCAGTGG

TABLE 2.9: Volume of components for primer mixes in qPCR experiment with Power SYBR® Master Mix (Life Technologies)

Reagent	Volume n = 1 (μL)
Power SYBR® Master Mix (2×)	10
Pure H_2O	4
Forward Primer (10 μM)	0.5
Reverse Primer (10 μM)	0.5
ccfDNA (0.2 ng/μL)	5
Total	20

2.9.1 DNA qPCR

Power SYBR® Master Mix (Life Technologies) was used with the Quantstudio 7 (Life Technologies). Reactions were formulated as follows (Table 2.10).

TABLE 2.10: Volume of components for primer mixes in qPCR experiment with Power SYBR® Master Mix (Life Technologies)

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

The above reagent mix was made up separately for each primer set. 15 μL of the appropriate mix and 5 μL of DNA 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 DNA, 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 2.11):

TABLE 2.11: The thermal protocol of a qPCR experiment using Power SYBR® Master Mix on a Quantstudio 7 thermal cycler

Step	Temperature	Time	Cycles
Hot Start	95°C	10 minutes	1
Amplification	95°C	30 seconds	5
	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

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 115 and ratio of concentration (ng/ml) [Alu247]/ [Alu115].

2.9.2   Standard curves

The absolute amount of ccfDNA in each sample was determined by a standard curve with serial dilutions (from 10.000 to 0.01 ng/ml) of human genomic DNA (Promega, Madison, WI, USA, cat.# G3041). Standard curves were created for both 115 and 247 primer sets by PCR amplifying 10-fold serially diluted human genomic DNA samples. A negative control (without template – NTC) and a positive control (human genomic DNA) were included in each plate and a melting curve was drawn after each PCR to ensure that only one peak was amplified for all samples. The 115-bp amplicon was producing a peak at 79 ^{o}C and the 247-bp had a higher melting temperature at approximately 83 ^{o}C. The genomic human DNA is highly concentrated and for the proper re-suspension and dilution I had to pipette, vortex, quick spin and leave for 10 minutes at RT before I take the volume desired for the next dilution. The same procedure followed for all the serial dilutions.

2.10   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:

ΔCq=Cq (gene of interest -GOI) - Cq (reference gene)

Normalises GOI expression to the reference gene

ΔΔCq=Cq (Sample) - ΔCq (Calibrator)

Normalises GOI expression in a sample to that of a calibrator, for example untreated cells

$$\text{\textit{For } }\Delta Cq\text{ \textit{method: }} RQ = 2^{-\Delta Ct}$$

$$\text{\textit{For } }\Delta\Delta Cq\text{ \textit{method: }} RQ = 2^{-\Delta\Delta Ct}$$

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.

2.11   EntroGen ctDNA EGFR mutation Detection kit

The most frequently occurring mutations in EGFR (L858R, exon 19 deletions) are associated with sensitivity to EGFR tyrosine kinase inhibitors (TKIs). The T790M gatekeeper mutation has been implicated in acquired resistance to first generation EGFR TKIs, however it confers sensitivity to more recently developed targeted therapies. EGFR mutations are found in up to 15% of Caucasian and 4% of Asian populations. The cell-free EGFR assay requires circulating tumor DNA extracted from plasma. Mutation detection is based on allele-specific amplification and detection with hydrolysis probes.

The ctDNA EGFR mutation detection kit (Entrogen, CE-IVD was used for preliminary experiments with the Quantstudio 7 (Life Technologies) and the CFX96 (Bio-Rad). Reactions were formulated as follows (Table 2.12).

TABLE 2.12: Volume of components for primer mixes in qPCR experiment with ctDNA EGFR mutation detection kit. N = number of samples to be run (including controls)

Reagent	Volume n = 1 (μL)
CT Mutation Detection Reaction Mix (2X)	15
Primer Mix	6
Cf-DNA sample (2.5-25 ng) or controls	1-9
Molecular grade water	Adjust to 30
Total	30
Reagent	**Volume n = 1 (μL)**
CT Mutation Detection Reaction Mix (2X)	N $\times$ 15 μL $\times$ 1.05
Primer Mix	N $\times$ 15 μL $\times$ 1.05
Molecular grade water	N $\times$ 15 μL $\times$ 1.05

The reagent mix was made up separately for each primer set. 15 μL of the appropriate mix and 5 μL 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 (Tables 2.13, 2.14):

TABLE 2.13: The thermal protocol of a qPCR experiment using EGFR mutation reaction mix on a Quantstudio 7 thermal cycler

Step	Temperature	Time	Cycles
Hot Start	95^0C	10 min	1
Amplification	65^0C	15 secs	40
Fluorescence Measurement	63^0C	30 secs	CY5,FAM, ROX, VIC

TABLE 2.14: The thermal protocol of a qPCR experiment using EGFR mutation reaction mix on a Quantstudio 7 thermal cycler

Step	Temperature	Time	Cycles
1	95^0C	10 min	1
2	65^0C	15 secs	40
3	63^0C	30 secs +Plate Read	CY5,FAM, ROX, VIC
GOTO step 2 - 39 more times			

2.11.1 Entrogen qPCR data analysis

Analysis on the ABI Quantstudio 7 and Bio-Rad 96

Applied Biosystems -Quantstudio 7

To analyse the data in the real-time PCR instrument software, I had to follow specific criteria that the Entrogen company is providing:

Instrument	CY5 THLD	FAM THLD	ROX THLD	VIC THLD
QuantStudio7	30.000	50.000	100.000	20.000

And manually adjust the baseline:

- First cycle: 3

- Last cycle: 22

Bio-Rad CFX96

To analyse the data in the real-time PCR instrument software I followed the following criteria:

- Ct Determination with a single Threshold

- Baseline setting: Subtracted curve fit

To analyse the data in the real-time PCR instrument software, I had to follow specific criteria that the Entrogen company is providing:

Target	Baseline	Threshold
HEX (VIC	3-22	500
FAM	3-22	400
TX RED (ROX)	3-22	450
CY5	3-22	300

Results interpretation

Applied Biosystems -Quantstudio 7

In the beginning of the analysis, I had to select the low positive and high positive control wells and check the detector for a signal. Both controls should have the same VIC (Yellow) Ct Values and the target Cts of the low positive should be around 4 Cts away from the high positive control. After this step, I had to select each one of the reaction wells and check the signal for each detector and according to the table below determine the results. The EGFR mutation targets are not having a specific Dye that detects them, but the combination and the Ct value range of each sample is related to a specific mutation.

CY5 Ct is	ROX or FAM Ct is	VIC/yellow Ct is	Result
Ct = any	Ct = any	0<Ct<26	Rerun with less DNA
36<Ct; Ct=0	37<Ct; Ct=0	26≤Ct≤32	Negative sample. No need to rerun
Ct≤36	Ct≤37	26≤Ct≤32	Positive for Ex19Del (Cy5) and/or L858R (ROX) and/or T790M (FAM)
36<Ct; Ct=0	37<Ct; Ct=0	Ct>32; Ct=0	Rerun with more DNA
Ct≤36	Ct≤37	Ct>32; Ct=0	Positive for Ex19Del (Cy5) and/or L858R (ROX) and/or T790M (FAM)

Bio-Rad CFX96

Target Ct is	VIC (HEX) Ct is	VIC/yellow Ct is	Result
Ct = any	Ct = any	0<Ct<26	Rerun with less DNA
36<Ct; Ct=0	28≤Ct≤33	26≤Ct≤32	Negative sample. No need to rerun
Ct≤38	28≤Ct≤33	26≤Ct≤322	Positive for Ex19Del (Cy5) and/or L858R (ROX) and/or T790M (FAM)
36<Ct; Ct=0	Ct>33; Ct=0	Ct>32; Ct=0	Rerun with more DNA

2.12 siRNA transfection in cell lines

To further understand the role and function of XIST I sought to carryout functional studies. I used siRNA targeted to XIST (SMARTpool: ON-TARGETplus, Dharmacon, CO, USA) to suppress XIST expression in A549 and H1975 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 1×10^5 per well in 6-well (for RNA extraction) and in a spare well, 4 cover slips were placed. The siRNA transfection was carried out once at a concentration of 25 nmol per well 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 (A549) and RPMI (H1975) media (Gibco) overnight. siRNA was diluted to a 5 μM working concentration from a 20 μM stock solution, by a 1X in siRNA buffer (Dharmacon). Transfection was made up for n + 1 working concentration. Table 2.15 and Table 2.16 details the reagents and volumes for the reaction.

TABLE 2.15: Reagents and quantities for siRNA transfection mix for a 24 well plate

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 - 400 μL			

TABLE 2.16: Reagents and quantities for siRNA transfection mix for a 6 well plate

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 - 1600 μL			

Tube A and B are made up independently and allowed to incubate at RT for 5 minutes, before being mixed with together, and incubated for 20 minutes at RT to allow the complexes to form. Following incubation, the tubes were spun briefly for 1

minute at 3.000 rpm. Serum free media was then added to the solution and mixed by pipetting, followed by added 2 mL to cells in the 6 well plates (following removal of existing media). The transfection media was replaced with fresh serum free media at 24 h to avoid toxicity, and samples collected at 48 h and 72 h post transfection. Scrambled controls were conducted in the same way and carried for all experiments in both cell lines.

2.13   Advanced imaging wound healing data analysis

For the wound healing data analysis, the CellProfiler 2.0 was used that has a user interface for editing pipelines including dragdrop operations. A newly designed test mode allows the user to repeatedly adjust settings to optimise image analysis. I loaded high quality images of the wounds and imported the Example-Wound-Healing pipeline that can be freely downloaded from cellprofiler.org. Rather than identifying individual cells, this pipeline quantifies the area occupied by the tissue sample and gives the area covered by the mono-layer of cells in percentage.

2.14   RNAseq

RNA sequencing was carried out on A549 and H1975 cell lines treated with siRNA XIST for 48 hours. The RNA sequencing was carried out by collaborators at the Welcome Trust Centre for Human Genetics (Oxford). RNA samples were normalised to 550 ng/μl 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 two lanes of HiSeq4000 and the reference genome used for the run was the GRCh37.EBVB95-8wt.ERCC.

2.14.1   RNAseq data analysis

All samples were sequenced at 75bp paired end according to Illumina specifications. The data was analysed using open source software from the Tuxedo suite. Reads were mapped to the human reference genome hg19 using the GENCODE 24 annotations, with TopHat2 (Bowtie 2) under standard conditions. The alignments were filtered for high quality hits with a minimum selection threshold score of 30. The mapped reads were assembled into transcripts and quantified using Cufflinks package. Functional enrichment analyses and Venn diagrams were performed in the open software FunRich. The statistical cut-off of functional enrichment analyses using this stand-alone software was kept at default setting with a p-value < 0.05 after Bonferroni correction.

2.15   Magnetic Luminex Screening Assay

For this assay a custom kit of a human cytokine/chemokine magnetic bead panel was used (EMD Millipore, cat.# HTH17MAG-14K) (Table 2.17). Prior to the experiment, all reagents were allowed to warm to RT and wash buffer was added to each well of the plate and was mixed for 10 minutes on a plate shaker at RT. The wash buffer was then discarded, and assay buffer and controls were added to the appropriate wells. Matrix solution (control culture media) was added to background, standard and control wells. 25 μL of plasma from the pre- post-op patients was added in the appropriate wells in duplicates. Premixed beads were then added to the wells and the plate was sealed and incubated overnight at 4 ^{o}C. The next day, the well contents were gently removed and the plate was washed twice. Detection antibodies were added to each well, and the plate was incubated for an hour at RT on a plate shaker. After the incubation, streptavidin-phycoerythrin was added to each well and the plate was brought back to the plate shaker for an incubation of 30 minutes at RT. The wells were washed twice and sheath fluid was added to all of them. The beads were resuspended on a plate shaker for 5 minutes. The plate was then read using a Luminex 100/200 analyser (Luminex, U.S.A).

TABLE 2.17: **Human Th17 Antibody-Immobilized Magnetic Beads**
The Analyte names that were used in the assay.

Bead/Analyte Name	
Anti-Human IL-17F	Anti-Human IL-12P70
Anti-Human GM-CSF	Anti-Human IL-13
Anti-Human CCL20/MIP3α	Anti-Human IL-15
Anti-Human IFNγ	Anti-Human IL-17A
Anti-Human IL-10	Anti-Human IL-22
Anti Human IL-9	Anti-Human IL-1β
Anti-Human IL-33	Anti-Human IL-2
Anti-Human IL-21	Anti-Human IL-4
Anti-Human IL-23	Anti-Human IL-5
Anti-Human IL-6	Anti-Human IL-17E/IL-25
Anti-Human IL-27	Anti-Human IL-31
Anti-Human TNFα	Anti-Human TNFβ
Anti-Human IL-28A	

2.16 Bio-informatics

2.16.1 The Cancer Genome Atlas

The cancer genome atlas program (TCGA) carried by national cancer institute in The United States from 2006. This database selected and collected a high numbers of malignant tumor data from various sources include Nature and Cell Reports journals. This database opens to public, provided significant helps to everyone who needs cancer data.

2.16.2 The Genotype-Tissue Expression project

The Genotype-tissue expression project (GTEx) has also been carried by national cancer institute of The United States, very similar to TCGA but contain many healthy tissues genotype information for public to study.

2.16.3 Analysis tool

Xenabrower

An online tool to search and download data from various database, include TCGA and GTEx (UCSC Xenabrowser). This website is created by University of California Santa Cruz (UCSC), allowed public to study and download any data within those cohorts they included. It also allowed online visualisation of data.

R studio and R language

For this study, we used R packages of ggplot2, ggpubr, reshape2, openxlsx, plyr, survminer, survival and gridextra.

2.16.4 Analysis method and statistical test

Whitney-Mann Wilcoxon test

A test often used in compare two samples of non-parametric results. This test assumed two groups of samples were independent and calculated ranks rather than the actual observation values. We used 5% significance level in all Wilcoxon tests carried by this study.

Log-rank test

A test to compare survival data in Kaplan-Meier's graph. This statistical test aimed to test if there is no difference between the populations in probability of death at all individual time points (Null hypothesis). All results of individual time points of where the events happening were collected to estimate the risk of death and performed a Chi-square test to see if it was significant. Overall, this test aimed to analyse if the risk of death between two groups was different.

Spearman's Rank Correlation

This method was used to determine if there is any two-way association or relationships between two groups. Results of a correlation test is given by a correlation coefficient, a positive coefficient means a positive correlation, refer to if X increased, Y increased in the same way; negative correlation refer to opposite meaning. This test used ranks to calculate, provided a more robust correlation coefficient when data were skewed or contained extreme values.

2.17   Statistical analysis

A student's t-test was used to assess statistical significance of any changes observed in experiments. An assessment for homoscedasticity of data from each category was made using the F-test (F-hypothesis under a null hypothesis). 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 Graph-Pad Prism Software (GraphPad Prism Software). p values were denoted on graphs and interpreted as follows (Table 2.18).

TABLE 2.18: The asterisk denotations of p-value on graphs and ranges.

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

The Wilcoxon test was used to compare the gene expression level between two cohorts and Spearman's rank test was used to determine the correlation coefficient between two gene expression patterns and one-way ANOVA was used for the TNM stage and gene expression correlation in chapter 5. All statistical tests performed on GraphPad Prism or R studio v3.5.0.T-SNE analysis was performed using the RtSNE v0.15 package in R v3.5.0.

ccfDNA statistics

The Mann-Whiney U-test was used to compare the [ALU115] and ALU247/115 of plasma DNA between the different patients and a Kruskal-Wallis H-test was used for the comparisons between the groups. p value of < 0.05 was considered statistically significant.

Chapter 3

ImageStream™ analysis of cell lines and circulating tumour cells in blood of lung cancer patients

3.1 Introduction

As mentioned, circulating tumour cells 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. 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. However, detection of small populations of CTCs within the large number of normal blood cells represents a significant technical challenge [33, 37, 170].

CTCs in cancer patients not only contribute to metastasis but also hold information about disease progression and drug resistance. As a result, early detection of CTCs and their precise monitoring at the molecular level can positively impact the prognosis, and the efficiency of surgical and chemotherapeutic interventions [171]. CTCs can be biochemically, genetically or bio-physically characterised. Biochemically, a cell can be defined as CTC if it co-expresses the epithelial cell adhesion molecule (EpCAM) and cytokeratin (CKs – e.g. CK7) but lacks the leukocyte common antigen CD45 [172, 173]. All cells (CTCs and WBCs) can be stained positive with the

nuclear stain Draq5 to ensure the integrity of the nucleus. However, CTCs may undergo numerous genetic and phenotypic changes and can be expressed in several different sub-populations, each with different biomarkers [47, 174–176].

Conventional cytogenetics, morphology and flow cytometry have been used to detect circulating tumour cells at a low level of detection. To date, 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. Most assays established for CTC enumeration rely on the expression of the cell surface marker EpCAM. However, these approaches may not detect CTCs that express no/low levels of EpCAM, e.g. by undergoing EMT [177, 178]. Developments in imaging flow cytometry has enabled the technology to overcome the earlier hurdles of acquiring enough fluorescence sensitivity, production of high spatial resolution imagery, combining fluorescence imagery with other imaging modes, and the ability to image all the cells in flow.

The fluorescence image- based method for quantifying nuclear translocation (an ImageStream pipeline) described in this project relies on the spectral isolation of PanCK, CK7, TTF-1 and EpCAM images from ImageStream analysis. Fluorescently labelled or unlabelled cells in solution were run through the ImageStream and the data were acquired using the INSPIRE software. The INSPIRE software generates data in raw image file (.rif) format that can be analysed into IDEAS and produce a compensated image file (.cif). During the analysis the user can generate plots for subpopulations and study them further and save the file as a data analysis file (.daf). It is also possible for the user to generate individual .tiff images from each channel of a selected population to analyse the cells outside the software's framework. Firstly, an out-of-focus exclusion was made to remove any debris or clusters of cells from my selection by gating accordingly the focused and single cells. Once the subpopulations were identified via gating they were being saved as a new .cif file in IDEAS.

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. So far, only one instrument (CellSearch) has gained FDA approval but clinical adoption is lacking.

3.2 Aims and Objectives

Quantification of CTCs in lung cancer blood samples using the Imagestream$^{\text{TM}}$ technology

- Assessment the efficacy of quantifying CTCs from lung cancer blood samples using the Imagestream$^{\text{TM}}$ technology

- Evaluation of the ability of the Imagestream$^{\text{TM}}$ to detect and quantify CTCs using A549, H1975 and MSTO-211H cell lines as initial *in vitro* experimental models

- Comparison CTCs numbers for pre- and post- operated NSCLC patients

- Extrapolation of CTC results and correlation with patient outcomes and survival

3.3 Results

3.3.1 ImageStream^{TM} analysis of cell lines

I first conducted proof of principle experiments to try and stain the A549, H1975 and MSTO-211H cell lines with the AE1/AE3 PanCK, EpCAM, TTF-1 and CK7 antibody and validate the efficiency of this staining. A549 and H1975 were used as lung cancer models and the MSTO-211H cell line as mesothelioma model. The cells that had been stained with the EpCAM diluted primary antibody were then stained with a secondary Alexafluor® 488 (AF488) antibody that emits a green fluorescence. Pan-cytokeratin (PanCK) a cytoplasmic and membrane marker, is a widely used antibody in histopathology laboratories and appears as a halo staining around the cell surface. The TTF-1 is an antibody mainly used for the detection of lung cancer and thyroid cancer. CK7 is a type II cytokeratin that is found in epithelial cells of the lung and is a useful immunohistochemical marker for diagnosis of lung tumours. Once the samples were processed and stained, they were run through the Imagestream^{TM} to obtain the raw file from the fluorescent flow cytometry process.

3.3.2 Identification of A549 single cells using IDEAS®

In the Fig. 3.1 and Fig. 3.2 Y-axis shows the distribution of the cells in terms of their aspect ratio while the X-axis shows their distribution in terms of size. Each one of the blue dots represents one captured object (e.g. cell, debris, cell clusters). The "M01" at the end of the axis' name refers to the channel it applies to the software automatically generates statistics about the captured data such as the number of cells, mean and more and these statistics are displayed in a table below the graph. Careful examination of these dots or cells was then followed by gating single cells (SC) to exclusively analyse them. EpCAM, PanCK and CK7 antibodies are conjugated and they stain the cytoplasm and the membrane, while TTF-1 needs a further permeabilization step to penetrate the nucleus and stain it. Figures 3.1 and 3.2 represent examples of the selection process of positive, single and focused cells.
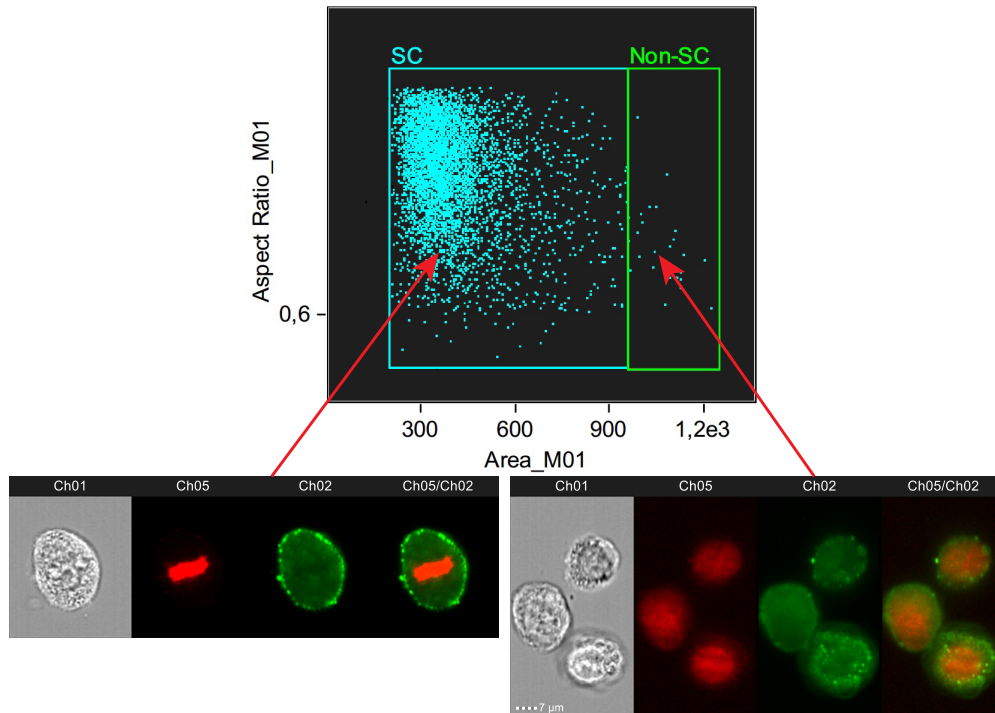
FIGURE 3.1: **Selection of the A549 EpCAM+ single cells and exclusion of the debris or clusters.** A selection by gating was taking place for every run to isolate the Single cells from the Non-Single cells. Here with blue colour are the single cells and with green colour are the non-single cells. SC: Single cells; Channel 1 (Ch1): Bright Field; Channel 2 (Ch2): EpCAM+; Channel 5 (Ch5): DRAQ5 (Nuclei staining)
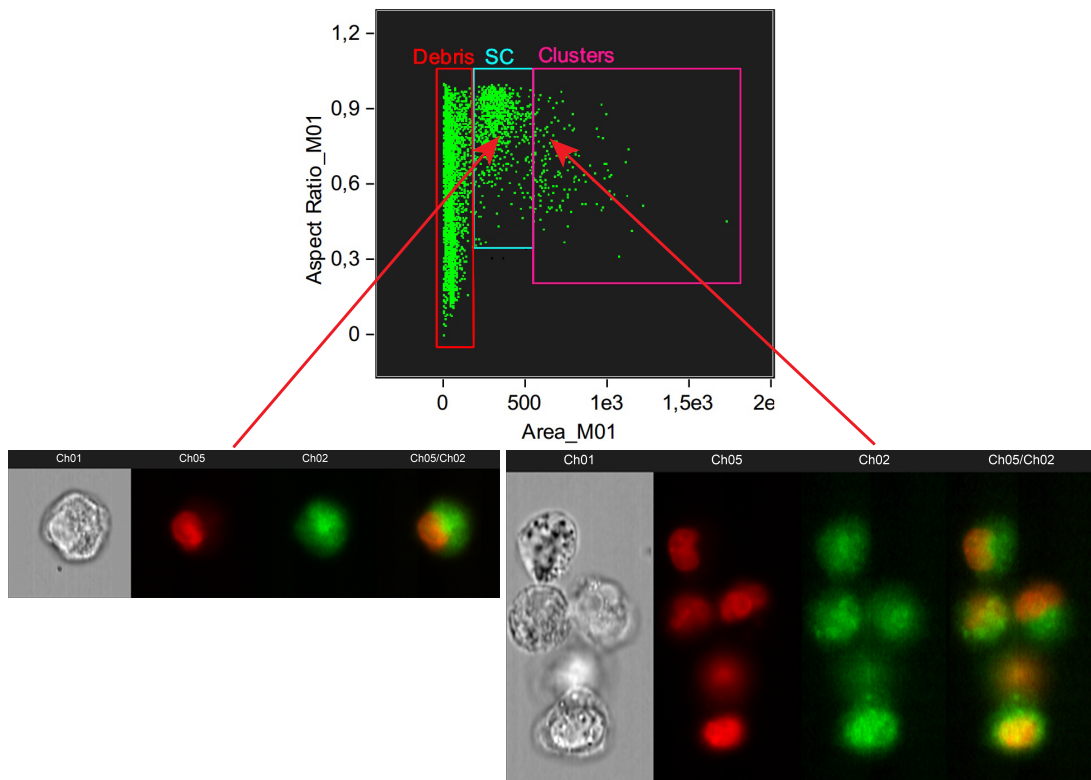
FIGURE 3.2: **Selection of the A549 TTF-1+ single cells and exclusion of the debris or clusters.** SC: Single cells; Channel 1 (Ch1): Bright Field; Channel 2 (Ch2): TTF-1+; Channel 5 (Ch5): DRAQ5 (Nuclei staining). Debris gated events are fragments of cells or unbinded antibodies

3.3.3 Identification of A549 focused cells

After the selection of the SC, I created a new graph from where I gated the Focused cells (F). This selection excludes any blurred cells from further analysis. On the X-axis, a gradient of cells based on image focus is used, while the Y-axis represented the normalised frequency. Following examination, a line gate was created (in yellow) separating single focused cells from blurred unclear cells (Fig. 3.3). Cells with less than 50% focus were excluded. The table next to the graph is automatically generated by the analysis which give us the statistics of the gate.
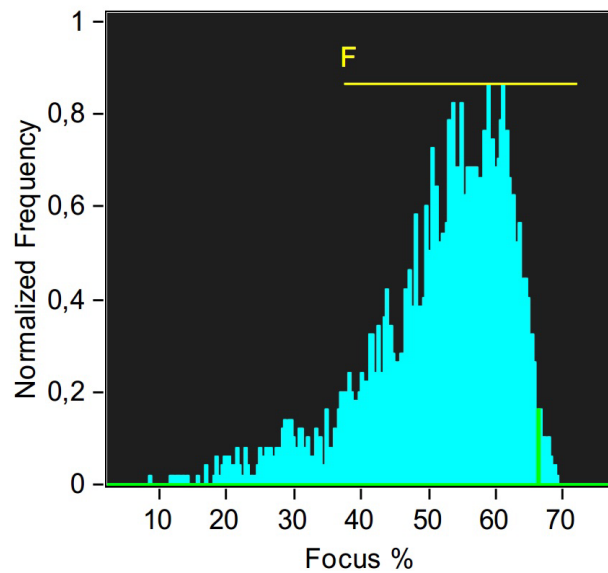
FIGURE 3.3: **Selection of the A549 Focused Single Cells** Selection was made for cells that were focused 40-100%. F: Focused

Figures 3.4 and 3.5 compare the intensity of the primary antibody staining on the Y-axis against the intensity of DRAQ5 (nuclei staining) on the X-axis. After manual examination of various cells, the population of cells that was EpCAM positive was gated as can be seen in the red box in Fig. 3.3. The same process was followed for the PanCK and CK7 antibodies.
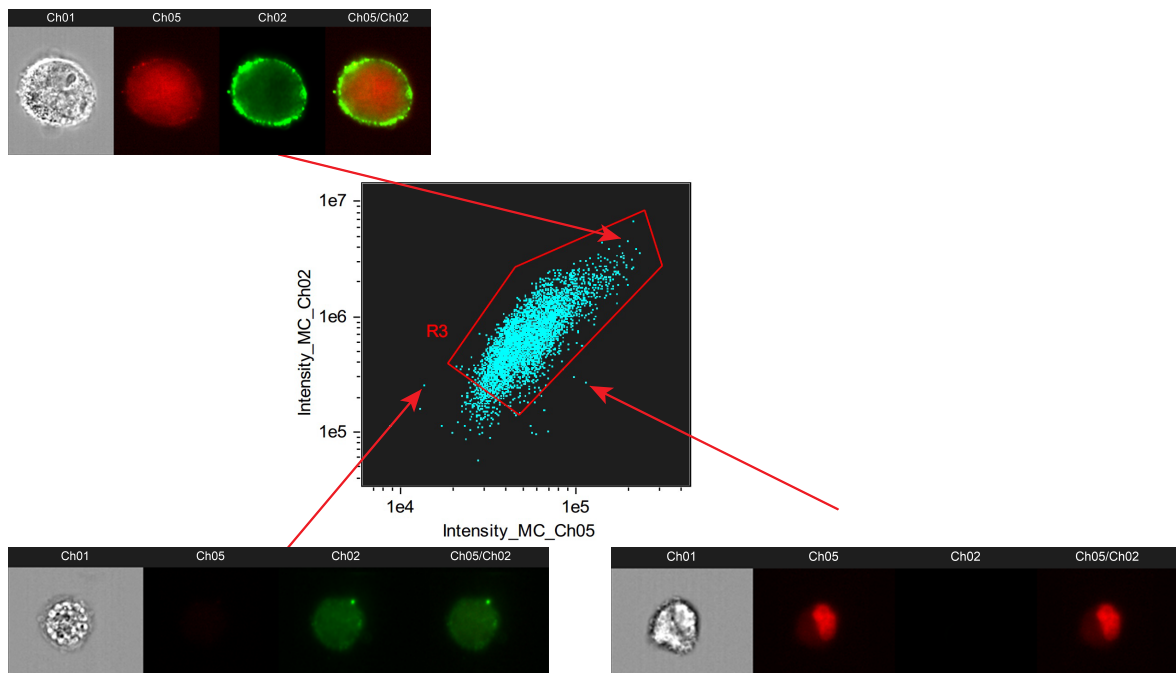
FIGURE 3.4: **Identification of EpCAM+/- A549 cells.** The first image shows an A549 EpCAM+ cell with a good staining pattern and the other two images show only one channel positive for the staining.
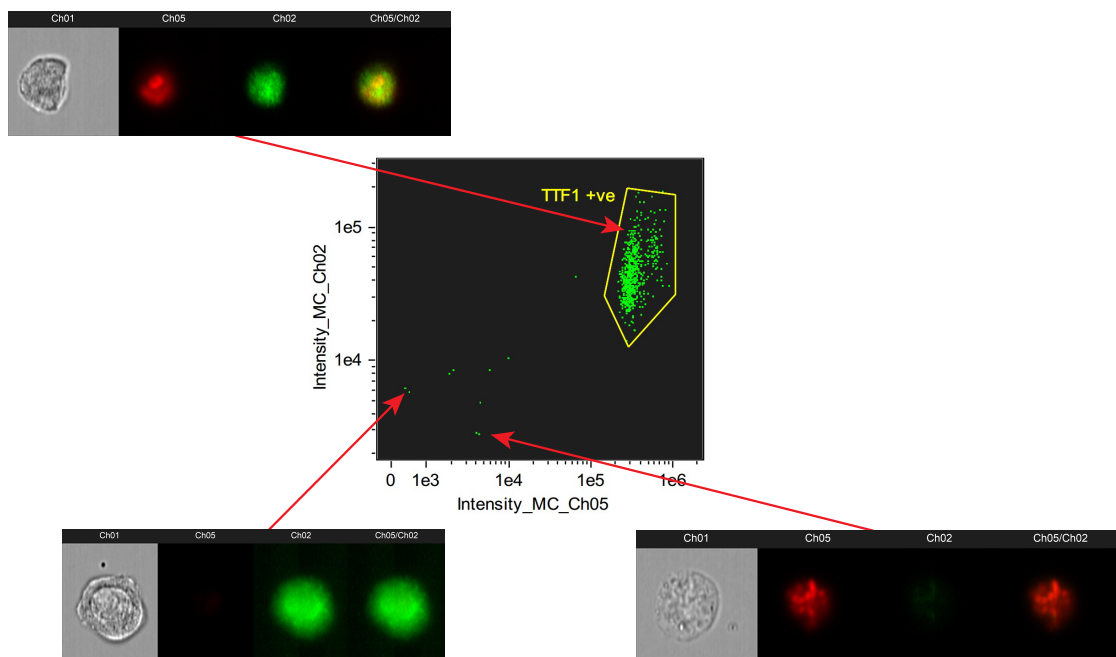
FIGURE 3.5: **Identification of TTF-1+/- A549 cells.**The first image shows an A549 TTF-1+ cell with a good staining pattern and the other two images show only one channel positive for the staining.

3.3.4 Size analysis

By using the Area Feature on IDEAS software, we can measure the Area of a mask in pixels and according to the calculation described in the manual 1 pixel = $0.25\mu m^2$. That means, that if we have a cell with a mask that include 2000 pixels this equals to 500 μm^2. For the following analysis, I used the Width of the circular, single and focused cells and I divided it by 0.25 μm^2 = 1/4.
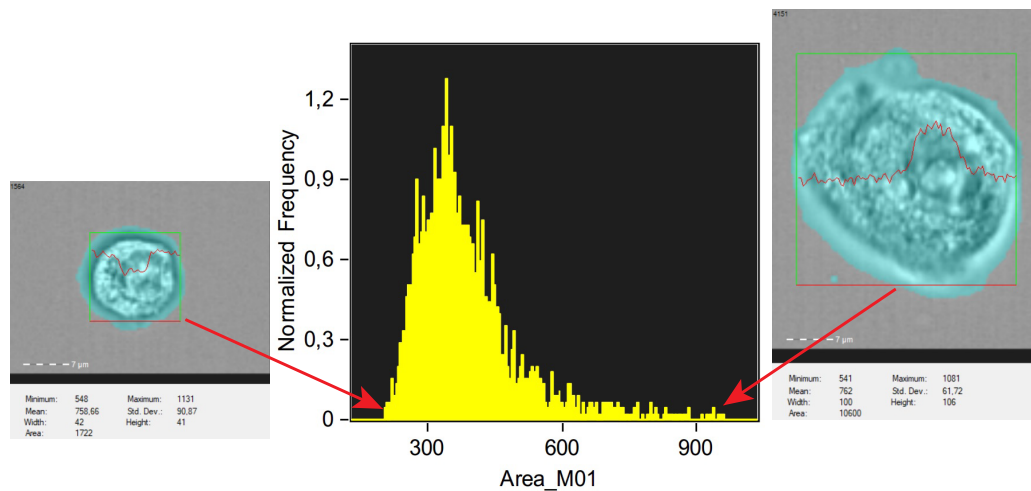
FIGURE 3.6: **Example of the smallest and biggest A549 EpCAM+ cells.**
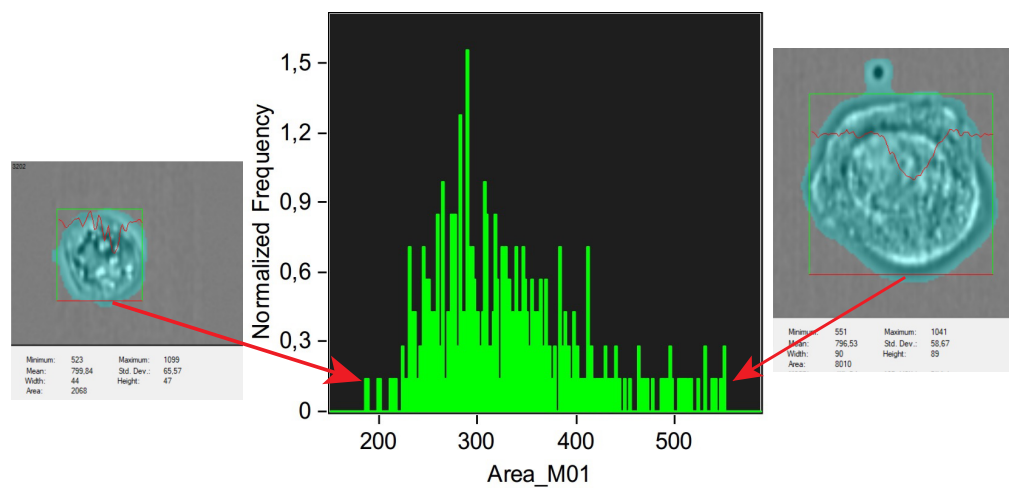
FIGURE 3.7: **Example of the smallest and biggest A549 TTF-1+ cells.**

The figures above show the cell size scale (Figs. 3.6 and 3.7). The A549 cell size is measured in pixels area and expands from 44 pixels width = 11 μm diameter to 100 pixels width = 25 μm diameter. Cells on the left is the smallest cell detected, whereas cell on the right is the largest A549 cell observed in these samples. Detailed analysis and structure figures are given below each cell. Both bright field images are highlighted with a blue mask as the software measured variable statistics for both cells and calculated their size. Images shown in this figure are captured at 60x magnification.

3.3.5 Identification of H1975 single cells using IDEAS®

Following the same analysis as for the A549, I detected the smallest and the biggest H1975 cells so as I can possibly detect them later in blood according to their size.
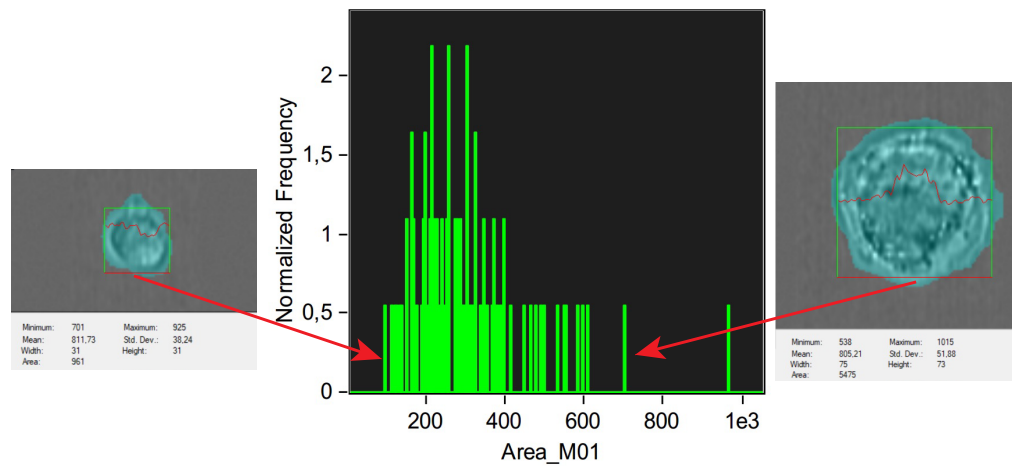
FIGURE 3.8: **The smallest and biggest H1975 TTF-1+ cells.** Details of selected region were automatically generated and presented under each image
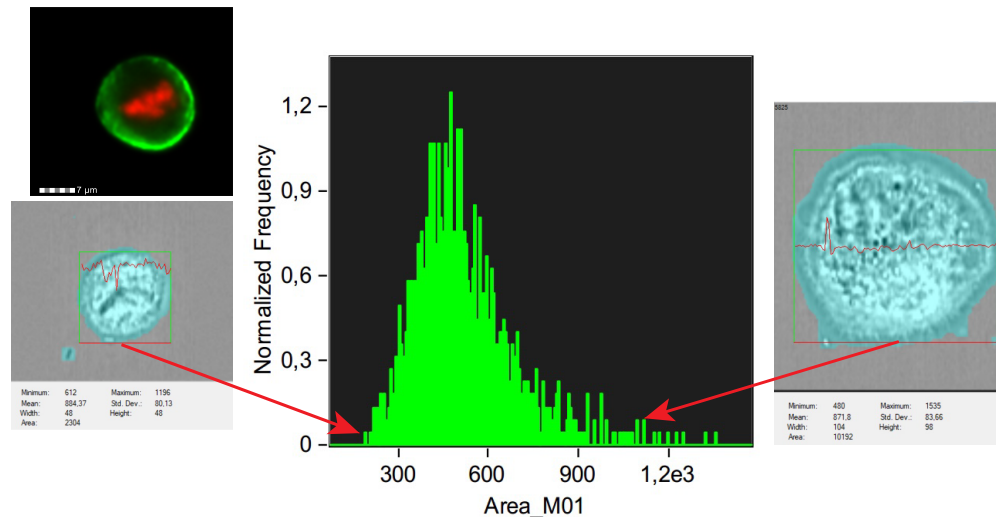
FIGURE 3.9: **The biggest and smallest H1975 PanCK+ cells.** An example of H1975 cell undergoing mitosis. Details of selected region were automatically generated and presented under each image
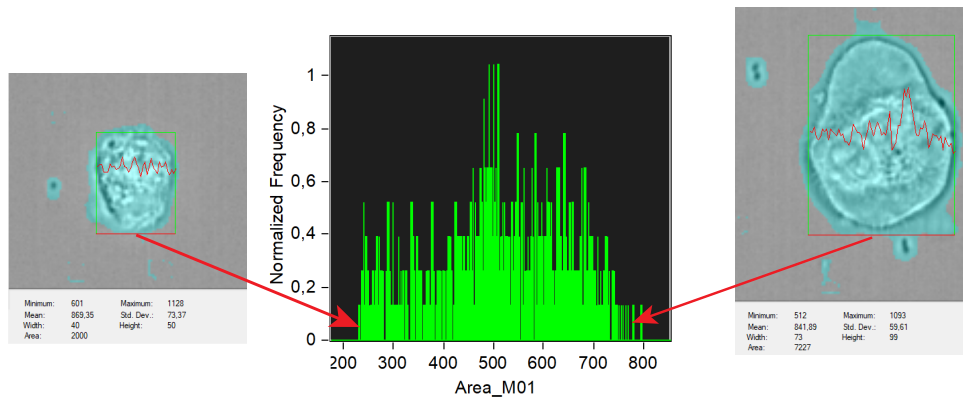
FIGURE 3.10: **The biggest and smallest H1975 EpCAM+ cells.** Details of selected region were automatically generated and presented under each image

The figures above show the cell size scale (Fig. 3.8, 3.9 and 3.10). The H1975 cell size is measured in pixels area and expands from 48 pixels = 12 μm diameter to 104 pixels = 26 μm diameter. Cell on the left are the smallest cells detected, whereas the cells on the right are the largest H1975 cells observed in these samples. Both bright field images are highlighted with a blue mask as the software measured variable statistics for both cells and calculated their size. Images shown in this figure are captured at 60× magnification.

3.3.6 Proof of Principle experiment, cells lines spiked in blood

Initial results of the three cell lines (A549, H1975, MSTO-211H) with TTF-1, PanCK, CK7 and EpCAM staining proved that the antibodies stain all cell lines. Subsequently, I spiked individually the cell lines and stained with 4 different antibodies, to evaluate the staining in 1 ml of healthy donors' blood mimicking a liquid biopsy, and the ability to differentiate the cancer cell lines from white blood cells. Approximately 5.000 and serial dilutions down to 150 cells of A549, H1975 and MSTO-211H were spiked in 1 ml of whole blood from a healthy donor (Tables 3.1, 3.2, 3.3, Figs. 3.11, 3.12 and 3.13).

TABLE 3.1: Retrieval number of positive for the antibodies tested in A549 cells for EpCAM, PanCK, TTF-1 and CK7.

Total A549 spiked in to 1mL	TTF-1+	PanCK+	CK7+	EpCAM+
5000	57% (2850)	42% (2100)	65% (3286)	27% (1375)
2500	75% (1875)	34% (850)	60% (1500)	32% (818)
1200	78% (213)	41% (500)	41% (500)	41% (494)
600	64% (386)	34% (208)	39% (235)	50% (303)
300	46% (141)	63% (160)	37% (112)	74% (222)
150	23% (35)	65% (98)	60% (91)	33% (50)
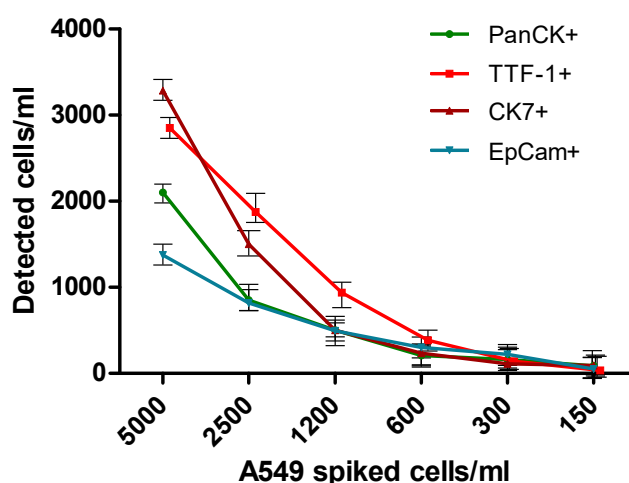
FIGURE 3.11: Average cell captures at 5.000, 2.500, 1200, 600, 300 and 150 positive A549 cells for EpCAM, PanCK, TTF-1 and CK7.

TABLE 3.2: Retrieval number of positive for the antibodies tested in H1975 cells for EpCAM, PanCK, TTF-1 and CK7.

Total H1975 spiked in to 1mL	TTF-1+	PanCK+	CK7+	EpCAM+
5000	57% (2850)	42% (2100)	62% (3125)	15% (750)
2500	75% (1875)	34% (850)	50% (1250)	25% (625)
1200	78% (213)	41% (500)	31% (375)	41% (494)
600	64% (386)	34% (208)	19% (118)	62% (357)
300	46% (141)	63% (160)	37% (114)	37% (111)
150	23% (35)	65% (98)	66% (100)	66% (100)
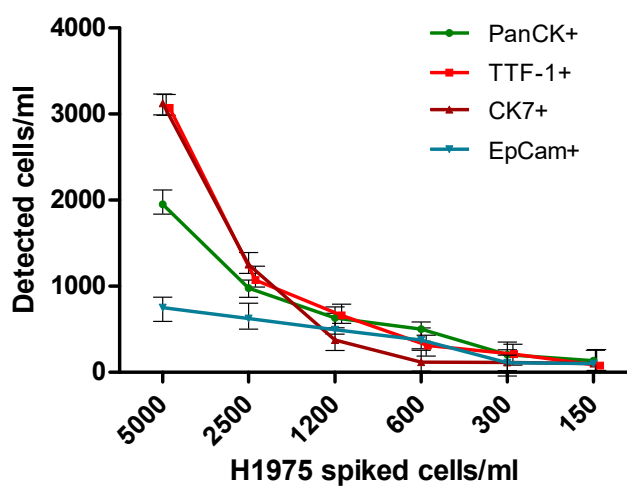
FIGURE 3.12: Average cell captures at 5.000, 2.500, 1200, 600, 300 and 150 positive H1975 cells for EpCAM, PanCK, TTF-1 and CK7.

TABLE 3.3: Retrieval number of positive for the antibodies tested in MSTO-211H cells for EpCAM, PanCK, TTF-1 and CK7.

Total MSTO-211H spiked in to 1mL	TTF-1+	PanCK+	CK7+	EpCAM+
5000	23% (1179)	20% (1000)	27% (1389)	80% (4023)
2500	19% (478)	28% (714)	65% (1636)	25% (625)
1200	30% (363)	52% (625)	56% (682)	55% (667)
600	62% (375)	41% (247)	52% (313)	37% (222)
300	46% (277)	42% (128)	37% (125)	66% (200)
150	62% (94)	44% (67)	74% (111)	44% (67)
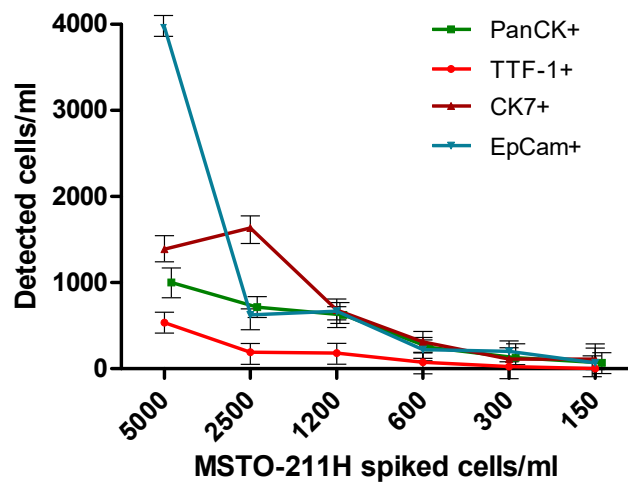
FIGURE 3.13: Average cell captures at 5.000, 2.500, 1200, 600, 300 and 150 positive MSTO-211H cells for EpCAM, PanCK, TTF-1 and CK7

3.3.7 ImageStream® analysis of lung cancer cells

Patient blood samples were collected from Mount Vernon Hospital and Harefield Hospital from lung cancer and mesothelioma patients. CTCs were quantified from blood samples of lung cancer patients before and after operation for samples from Harefield Hospital (n=54) (Table 3.4). Control blood samples were obtained from healthy volunteers (n=10). The same staining criterion was applied as in cell lines. Samples were obtained from patients with early stage (I-II), late stage (III-IV), and metastatic lung cancer.

Lung cancer samples from Harefield

TABLE 3.4: **Patient demographics for Harefield.** Demographics and histology of patients recruited for Harefield study

Variable	Value
Total	54
Mean age ($\pm$SD)	67.5$\pm$9.5
Males/Females(%)	23 (45) / 31 (55)
Pathology - no. (%)	
Adenocarcinoma	38 (70)
SqCC	16 (30)
Control	10
Males/Females(%)	5 (50) / 5 (50)
Mean age ($\pm$SD)	45$\pm$5.5
Staging - no. (%)	
I	35 (65)
II	10 (18.5)
III	7 (13)
IV	2 (3.5)
Operation - no. (%)	
Thoracotomy	12 (22)
VATS	42 (78)

All samples were processed and analysed on ImageStream™ by using the optimised template with the gates as explained above (Section 3.3.5). Quantification was calculated as positive cells per 1 ml. This calculation may have variability in

the exact sample volume but it is the most clinical correct way of translating the result. During the analysis process of all samples and all the tested antibodies, some discrepancies were observed in the specificity of the staining. These events were accounting about 10% of the total event captured per run. For these cells, a more careful examination had to be done in order to determine whether or not a cell was positive for the tested antibody. After every run, $20 \mu l$ of the cell suspension were dispersed on a microscope slide to verify the staining and characterize the positive cells further using the Leica DM4000 (Leica Microsystems) (Figs. 3.14 and 3.15).
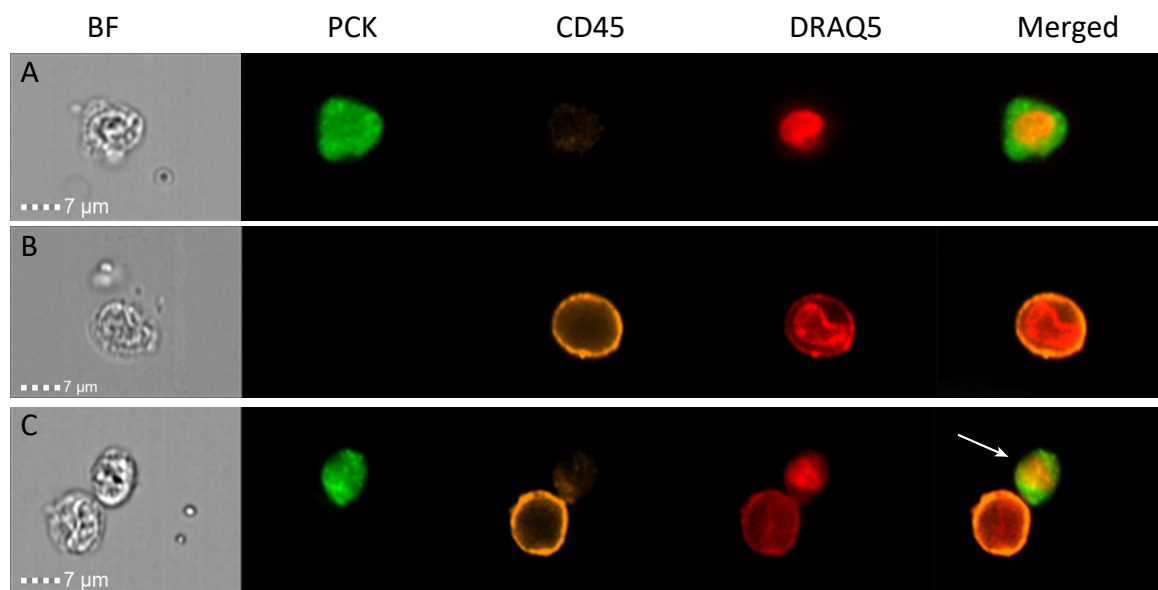
FIGURE 3.14: **Representative images of PanCK immunofluorescence staining of patients' blood sample.** A) PanCK+ (green) CTC; B) PanCK-/CD45+ (orange) WBC; C) PanCK+/CD45– and PanCK-/CD45+ cells captured in the same frame. All images have been captured on the Imagestream^{TM}, magnification $\times 60$. The white arrow shows a CTC attached on a WBC. BF: Bright Field
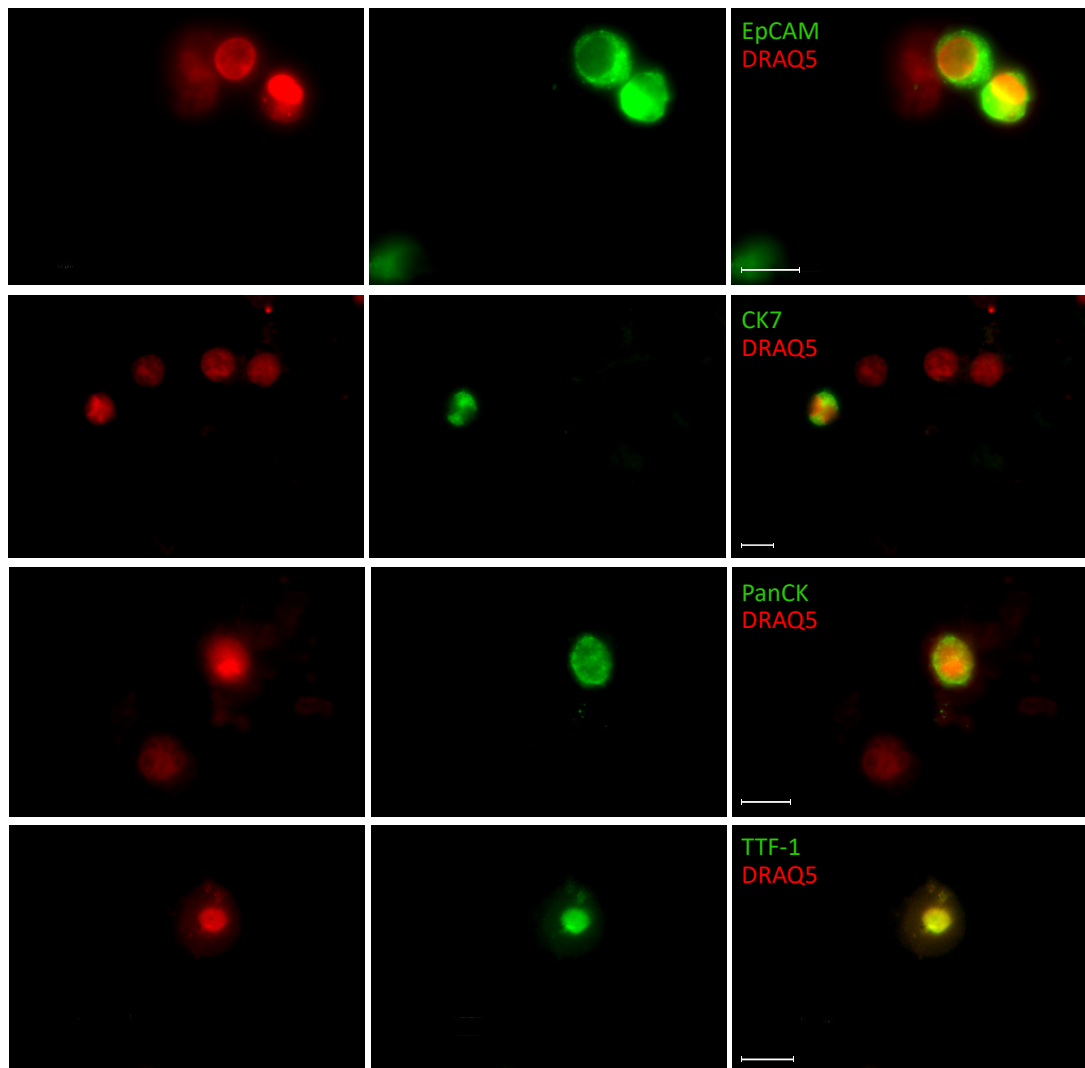
FIGURE 3.15: Representative images of EpCAM, CK7, PanCK and TTF-1 immunofluorescence staining of patients' blood sample on Leica DM4000. Scale bar: 20μm

Expression of CTCs in pre-op versus post-op NSCLC patients

We enumerated and characterised the EpCAM, PanCK, TTF-1 and CK7 positive cells in patients undergoing thoracotomy vs. VATS, as well as their histological subtype and stage. Each of the 4 chosen CTC markers was analysed separately with pre-operative samples compared to post-operative ones (n=54). We observed a statistically significant increase in the numbers of PanCK+ cells (mean value: pre-op=151, post-op=361 cells, p=0.0003) whilst a slight increase was observed for EpCAM (mean value pre-op=418, post-op=625, p=0.24), TTF-1 (mean value: pre-op=181, post-op=464 cells, p=0.09), CK7 (mean value: pre-op=95 , post-op=119 cells, p=0.15 (Fig. 3.16).

Subgroup analysis was then conducted on pre/post-operative paired samples for each of the 4 studied CTCs markers, with analysis performed based on (a) Surgical procedure (VATS vs. open thoracotomy), (b) Histology (AC vs. SqCC), and (c) post-operative tumour stage (I-IV), with results shown in Figures 3.17, 3.18 and 3.19.

CTCs levels for each of the studied markers were first compared between patients undergoing thoracotomy and VATS (Fig. 3.17). A trend towards increased EpCAM+ CTCs post-operatively was observed in thoracotomy patients (pre-op: 403, post-op: 1200, p=0.146), with minimal change in EpCAM+ CTCs observed following VATS (pre-op: 412, post-op: 386, p=0.716). When comparing surgical method directly, a significant difference was seen in the numbers of post-operative EpCAM+ CTCs between thoracotomy (1200 cells/mL, n=13), and VATS (386 cells/mL, n=41) (p=0.004). TTF-1+ CTCs were found to increase post-operatively in all patients regardless of surgery, and whilst the mean numbers of post-operative TTF-1+ CTCs were higher in thoracotomy (825 cells/mL, n=13) vs. VATS (353 cells/mL, n=41) patients, this was non-statistically significant (p=0.079). PanCK+/CD45- CTCs were found to significantly increase post-operatively in both thoracotomy (pre-op: 212, post-op: 618, p=0.021) and VATS patients (pre-op: 127, post-op: 259, p=0.022), whilst in direct comparison, post-operative PanCK+/CD45- CTCs numbers were significantly higher (p=0.03) in thoracotomy (618 cells/mL, n=13) patients compared to those who had VATS (259 cells/mL, n=41).

Overall, we observed a statistically significant increase in the numbers of positive PanCK (p=0.0003) cells and a trend of an increase in the EpCAM (p=0.248), TTF-1 (p=0.09) and CK7 (p=0.155) positive cells (Fig. 3.16).
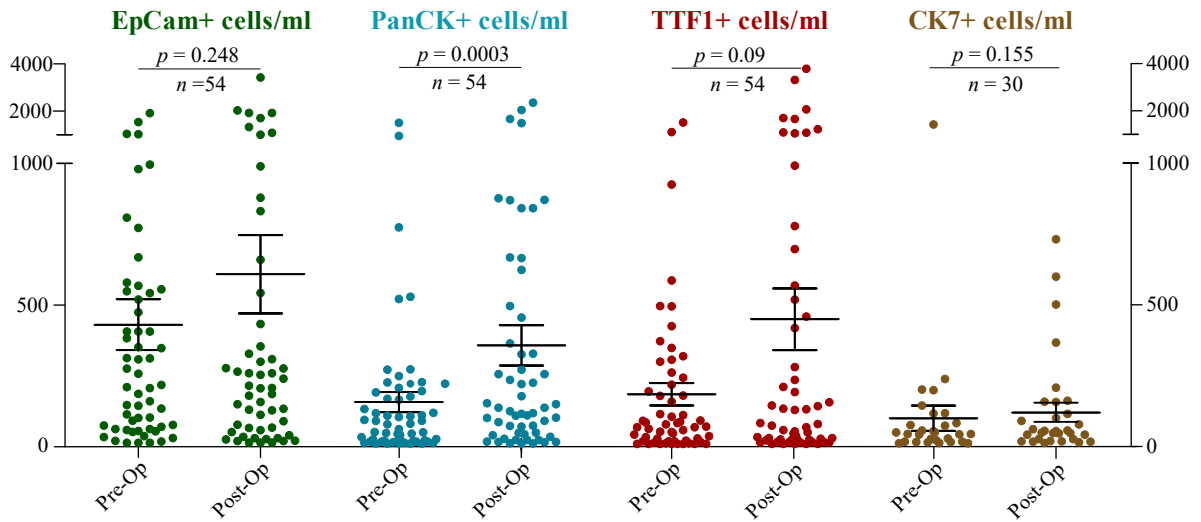
FIGURE 3.16: **EpCAM+, TTF-1+, CK7 and PanCK+/CD45- cells in pre-and post-operative patients.** Most of the patients showed elevated levels of detectable CTCs after the operation. All paired groups were analysed with Wilcoxon signed-rank test. Mean values and SEM bar is presented in all scatter plots.

Further analysis in the pre/post- op cohort revealed variations at the TNM stage, the type of surgery and the tumor type that these patients had. In particular, patients that underwent thoracotomy instead of VATS, had higher numbers of PanCK (*p=0.021* and *p*=0.022) while again positive cells for EpCAM (*p*=0.146 and *p*=0.716), TTF-1 (*p*=0.077 and *p*=0.357) and CK7 (*p*=0.12 and *p*=0.477) were a little bit higher than VATS but not statistically significant (Fig. 3.17). Post-operative patients presented higher numbers of positive cells for all the antibodies tested in thoracotomy compared to VATS (Fig. 3.17).
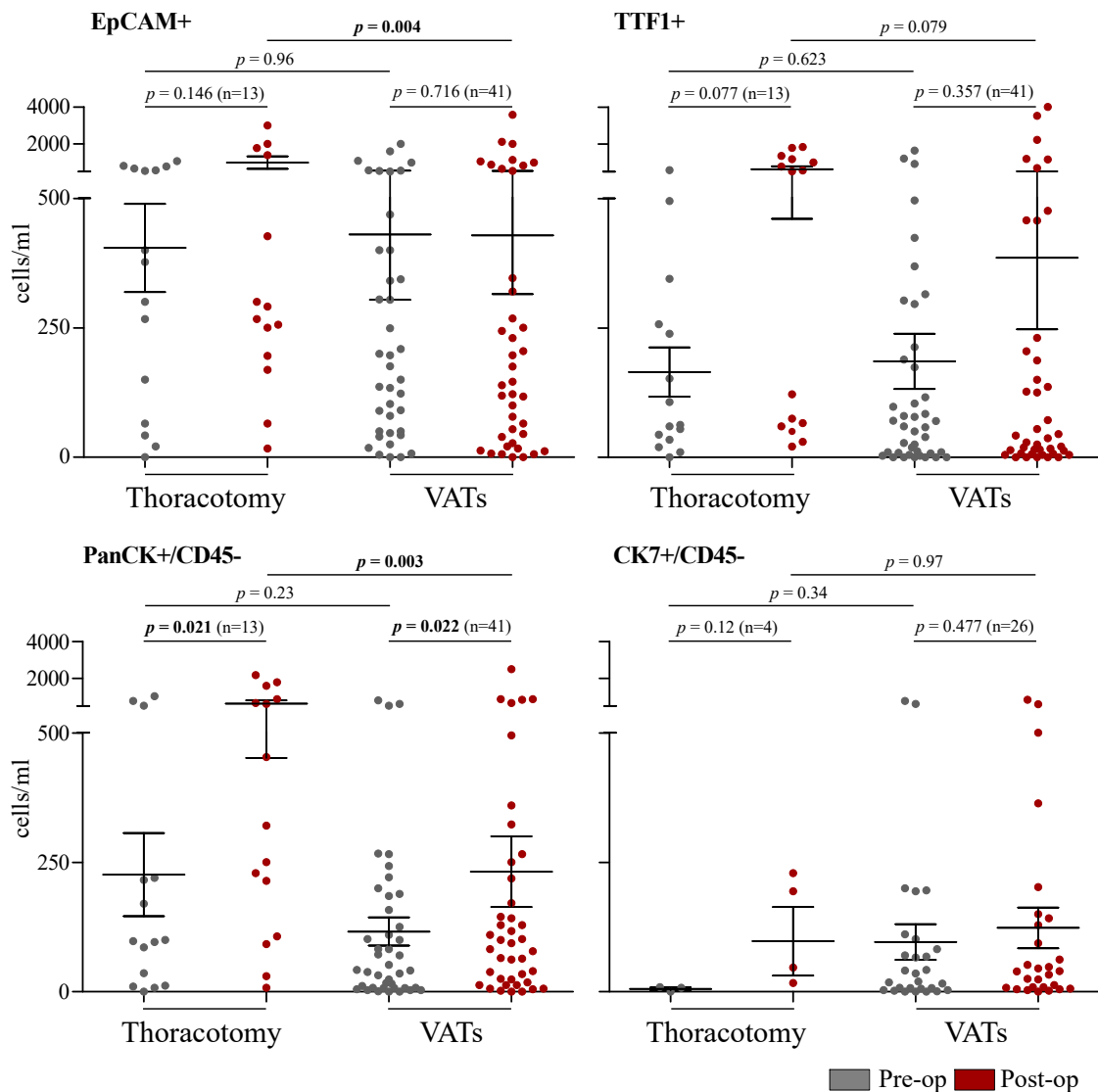
FIGURE 3.17: **Comparison of pre-and post-operative patients underwent thoracotomy or VATS.** Both operation techniques showed increased number of positive for all antibodies cells. All paired groups were analysed with Wilcoxon signed-rank test, and the unpaired groups with an unpaired t-test. Mean values and SEM bar is presented in all scatter plots.

Tumour associated CTC levels were then compared between AC and SqCC patients (Fig. 3.18). For patients with AC (n=38), CTC numbers increased significantly following surgery for all 4 markers (EpCAM, *p*=0.029; TTF-1+, *p*=0.015; PanCK+, *p*=0.0005, CK7+, *p*=0.004). For SqCC patients (n=10), no significant difference was observed between pre- and post-operative samples (Fig. 3.18).
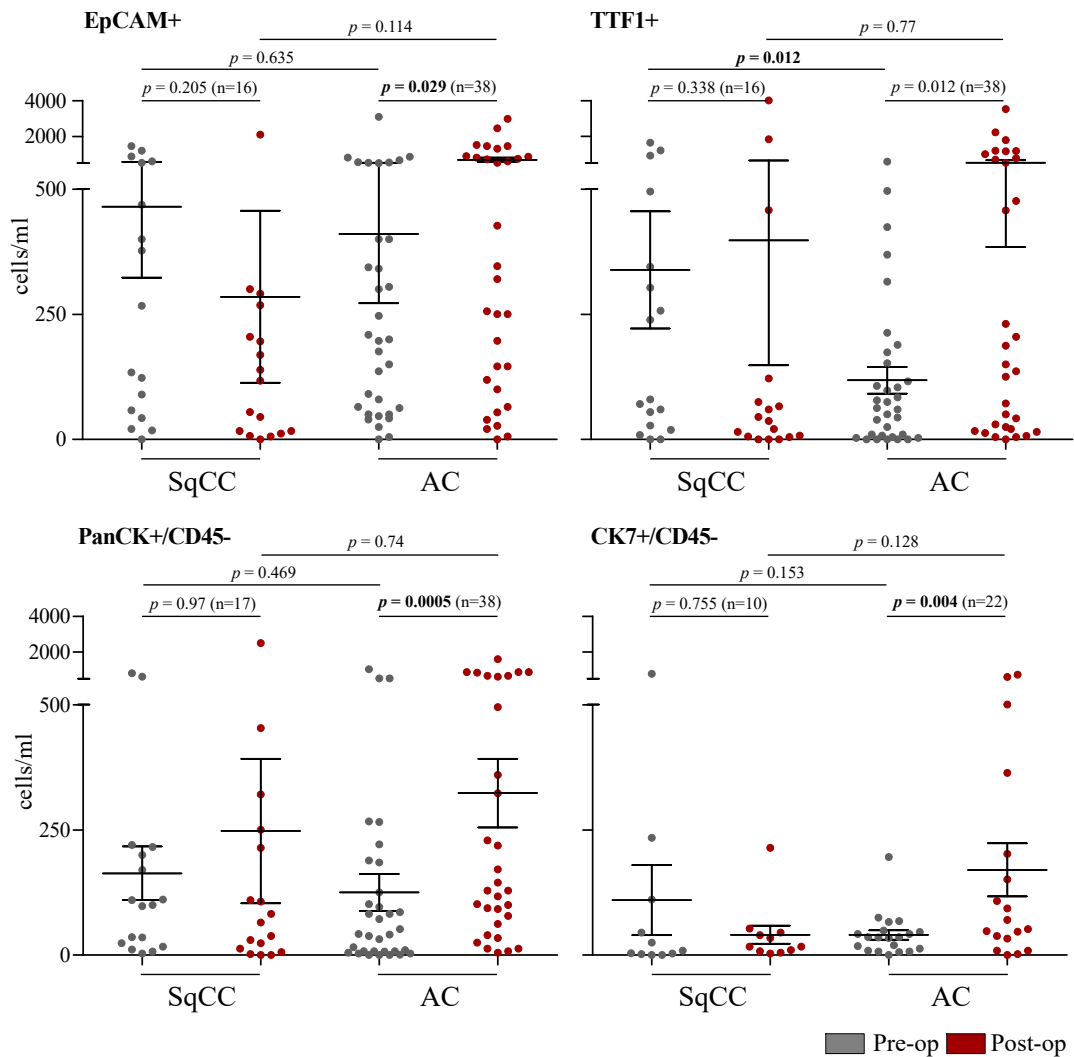
FIGURE 3.18: **Comparison of CTCs number in different lung cancers.** Ep-CAM+, PanCK+ and CK7+ cells showed higher numbers in AC patients after the operation (*p*=0.02, *p*=0.0005 and *p*=0.0004 respectively). All paired groups were analysed with Wilcoxon signed-rank test, and the unpaired groups with a Mann Whitney test. Mean values and SEM bar is presented in all scatter plots.

Finally, CTCs levels were compared between tumour stage (I-IV, Fig. 3.19). For patients with stage I disease (n=35), post-operative CTCs numbers were higher for all four CTCs markers studied, with significant differences seen in PanCK+ (p=0.04) CTCs. Stages I and II post-op samples had higher numbers of positive cells for all antibodies but same or lower number of cells in the stage III and IV population. In stage I the PanCK+/CD45- and CK7+/CD45- cells were more statistically significant after the surgery (p=0.04 and p=0.04 respectively). TTF-1+ and EpCAM+ cells were elevated but not significantly higher in numbers after the operation (p=0.5 and p=0.2 respectively). The same thing observed in the stages III and IV where no statistical significant difference was found in the numbers of cells for all antibodies before and after the surgery (Fig. 3.19).
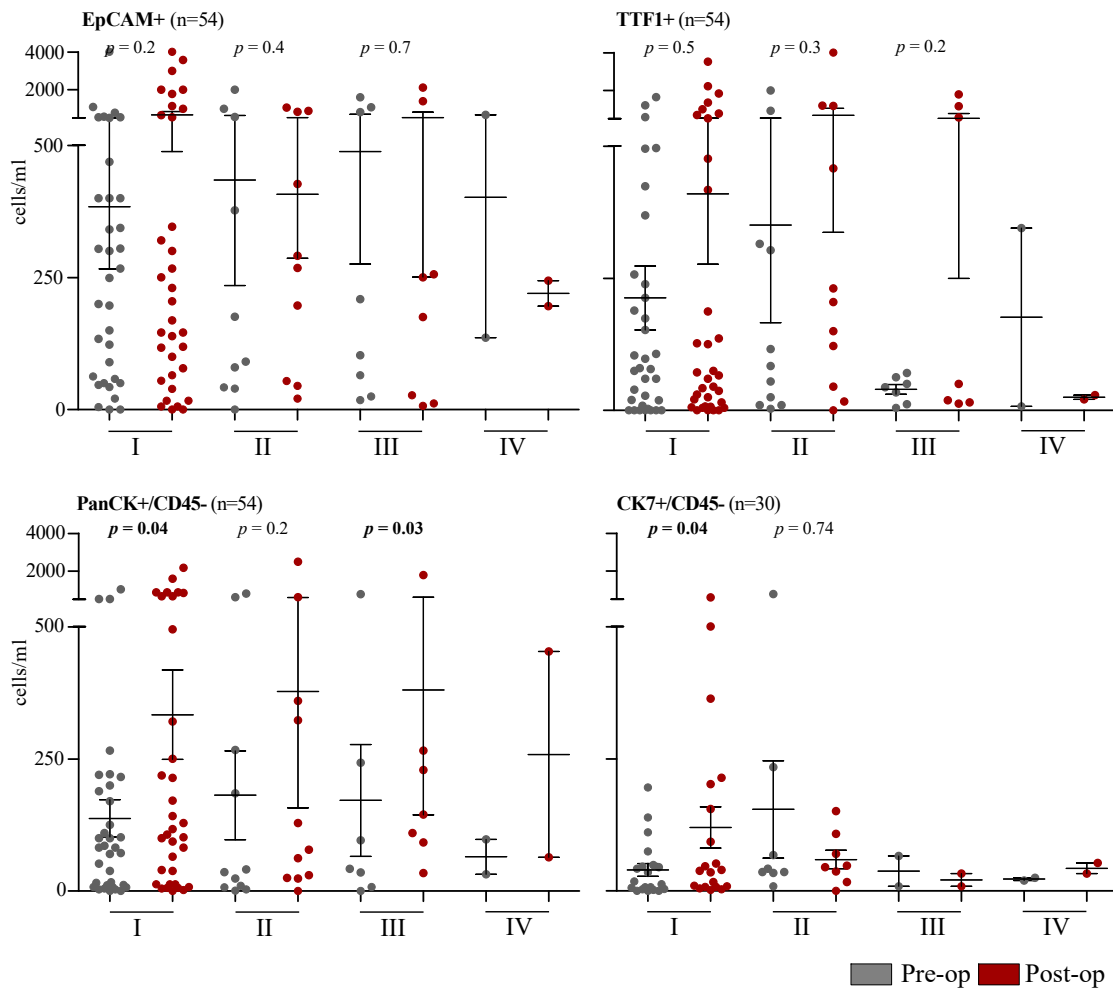
FIGURE 3.19: **EpCAM+, TTF-1+, CK7+ and PanCK+/CD45- cells in pre-and post-operative patients.** Number of positive CTCs for all the antibodies were higher after the surgery for all the different stages. PanCK+/CD45- and CK7+/CD45- cells were significantly elevated in stage I patients *p*=0.04 and *p*=0.04 respectively. All paired groups were analysed with Wilcoxon signed-rank test, and the unpaired groups with with a Mann Whitney test. Mean values and SEM bar is presented in all scatter plots.

3.4 Discussion

Following validation of multiple parameters, we aimed to describe how surgically induced trauma affects post-operative CTC levels. The first variable evaluated in this study was the effect of surgery on CTC levels pre- and post-operatively. Our data suggests that CTC levels were elevated in post-op patients in all groups stained with epithelial, cytokeratin or TTF-1 markers; reaching statistical significance with the use of PanCK marker. To the best of our knowledge, this is the first study documented in NSCLC patients and corroborates previous findings in patients undergoing surgery for hepatic metastases from colorectal cancer, where a substantial number of CTCs were detected intra-operatively when compared to their preoperative status [179]. Following identification of CTCs, we sought to establish whether there were differences between the type of operation, tumour stage and histological subtype. In this study, we demonstrate that patients that underwent thoracotomy had higher numbers of EpCAM (p=0.004) and PanCK (p=0.03) positive cells post-operatively when compared to post-op patients underwent VATS.

Due to the larger incision and manual manipulation required to perform thoracotomy, there is a higher likelihood of epithelial cells making their way into the circulation, which could explain why the mean number of PanCK+ (pre-op: 212 cells; post-op: 618 cells), EpCAM+ (pre-op: 403 cells; post-op: 1,200 cells) and TTF-1+ (pre-op: 142 cells; post-op: 825 cells) was higher after thoracotomy compared to VATS. We recognise that one limitation of this study was the inclusion of patients who were found to be stage IV post-operatively, in whom we would expect levels of CTCs to be higher regardless of surgical intervention. Future studies would aim to include more patients and analyses early stage (I and II) and advanced stage (III and IV) NSCLC patients as separate cohorts.

Despite a number of advances in systemic and local treatment options, more research in needed to improve outcomes for all patients [180]. Over the past decade, single-port VATS has been used extensively for major lung resections. Although originally limited to resection of simple small lung tumours, more recently, complex lung resections using single-port VATS have been performed [181]. Moreover, randomized trials comparing VATS with open thoracotomy in patients with NSCLC have been quite small in terms of size . It has been reported for example that VATS is an appropriate procedure for patients with early-stage NSCLC when compared with

open surgery, since there is a reduced systemic recurrence rate and an improved 5-year mortality rate of patients undergoing VATS [182]. Similar findings have been reported in previous studies [183].

CTC detection and enumeration is affected by the tumour cell plasticity and heterogeneity. It has been shown that in some cancers, the cells lose the epithelial marker due to EMT a process that is essential for their metastatic ability [184–187]. This ability is one of the main challenges for the detection of circulating cells with a variety of markers. Interrogation of positive staining and size of the cells revealed that cells vary in sizes and non-specific staining may be seen in the samples. Cells may also appear large and negative for a pan-cytokeratin marker or CD45, but positive for nuclear marker which could mean that tumour cell heterogeneity and hyperploidy may play a significant role in the CTC characterisation [188]. In order to define a cell as a true positive CTC Dent et al. showed that the use of multiple epithelial and CD45 markers can lead to a successful identification of CTCs across different cancers with specific tumour markers [158]. For this reason, both general epithelial markers (EpCAM and PanCK) and specific for lung cancer markers (TTF-1 and CK7) have been used to facilitate the selection of true CTCs.

Chapter 4

Implementation of plasma and nucleic acids analysis in biomarker development

4.1 Introduction

Circulating tumour DNA (ctDNA), a portion of ccfDNA specifically derived from cancer cells, has been estimated to range from 10 to 90% of the total ccfDNA in cancer patients. ctDNA, is tumour derived fragmented DNA, known to originate directly from the cancer itself and 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 >200 bp, due to the pathological process of cell death consisting not only of apoptosis, but also necrosis, autophagy, and mitotic catastrophe [189, 190]. 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. 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. This is an additional advantage over CTCs, which may not adequately represent the entire tumour due to its heterogeneous nature.

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 ccfDNA length (i.e. DNA integrity) in plasma may therefore depend on

the type of disease. Detection of these longer cfDNA 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.

ALU elements are found throughout the genome and they make up about 10% of the genome and are the largest family of mobile elements in the human genome [191]. They are a repeated sequence of DNA that are approximately 300bp and are derived from 7SL RNA gene [169, 192]. Retrotransposition is a process where the amplification of Alu elements occurs through reverse transcribing ALU-derived RNA polymerase III transcripts. ALU is an example of SINE (short interspersed elements). Each ALU structure is comprised of two parts. The 3' end consists of a 31-bp insertion in relation to the 5' end. Each ALU sequence base is approximately 300-bp long, this is dependent on the length of the 3' oligo (dA)-rich tail. ALU consists of a central A-rich region which is resultant from the area of insertion, they are lined by short intact direct repeats. The 5' end of the ALU sequence is comprised of an RNA-polymerase-III promoter. The 3' end is usually comprised of a multitude of as which rarely contains other bases in between it [191]. Studies measuring ALU have been carried out on breast, colorectal and lung cancer to predict the disease [193–195].

As the tumour progresses, the DNA integrity index is expected to be higher since the cell death of tumour cells becomes a predominantly necrotic process compared to the programmed apoptotic cell death of normal cells [196]. Detection of these longer ctDNA fragments and quantification of their relative abundance in plasma compared to short ccfDNA fragments and the calculation of a DNA integrity index has been explored as a potential cancer monitoring technique. To date, a number of studies have successfully validated these readouts in a variety of tumour types, including NSCLC [196, 197].

In our study, we focused on NSCLC samples and performed real-time PCR on ALU repeats (ALU115 and ALU247) on pre-op and post-op patient cohort. ALU115 value obtained by qPCR is a good reflection of the total amount of DNA in the serum as they amplify both short and long DNA fragments [194]. On the other hand ALU247 amplify long DNA fragments therefore detecting tumour DNA [198]. A DNA integrity index is calculated as a ratio (ALU247/ALU115). Typically, in healthy individuals' apoptosis of cells occurs naturally, where cfDNA is released in to the blood

circulation and uniformly truncated at 150-200 bp. ALU repeats can be detected of either >200 bp (indicative of necrotic DNA), or of <200 bp (detecting both necrotic and apoptotic DNA) [193, 199].

Cytokines are signaling molecules that mediate and regulate immunity and inflammation, are an important component of the biological milieu associated with cancers. In tumour, specific cytokines like IL-6, IL-10, and TNFα appear to be some components of a consistent cancer-associated cytokine network resulting in both a systemic immune stimulation and a microenvironment of cancer-induced immune suppression that ultimately protects the cancer cells. Because cytokines operate in integrated networks, a more complete understanding is gained as multiple cytokines are examined for patterns that may be associated with symptoms and prognosis as an effect of different types of surgical approaches. In this study we analysed the concentration of 25 analytes in 19 pre- and post-operative plasma of lung cancer patients to investigate how the cytokine levels are affected upon surgery.

4.2   Aims and Objectives

Implementation of plasma and nucleic acids analysis in biomarker development and management of lung cancer

- Quantification of circulating cell free DNA in case control study

- Comparison of the ccfDNA levels with the tumour type, size and operation method

- Evaluation of ALU repeats and DNA integrity in lung cancer patients

- Investigation of cytokines levels from pre-and post-operative patients

- Evaluation of the expression of EMT markers from mRNA extracted from whole blood of pre- and post-operative patients

4.3 Results

ccfDNA was extracted with the robotic Promega Maxwell system that is an automated protocol of extracting nucleic acids with minimum user interference. ccfDNA concentration measured by Promega Quantus in pre-operative patients (n=46, median: 133.5 ng/ml, range 20 – 5.000 ng/ml) and post-operative patients (n=46, median: 515 ng/ml, range 10 – 6.698) had a *p=0.0009* (Table 4.1, Fig. 6.2). While ALU115 marker was used for the quantification of cfDNA by using standard curves of known human genomic DNA. Its concentration in pre-operative patients (n=46, median: 441 ng/ml, range 124 – 21.210 ng/ml) was significantly lower than that in the post-operative patient samples (median: 2.127 ng/ml, range 192 – 9.959 ng/ ml with *p<0.0001*.

TABLE 4.1: **Patient demographics for Harefield study.** Demographics and histology of patients recruited for Harefield study and analysed for ccfDNA.

Variable	Value
Total	46
Mean age ($\pm$SD)	68$\pm$8.9
Males/Females(%)	19(41)/27(59)
Pathology - no. (%)	
Adenocarcinoma	33 (72)
SqCC	13 (28)
Control	10
Males/Females(%)	4 (50) / 4 (50)
Mean age ($\pm$SD)	44$\pm$6.5
Staging - no. (%)	
I	29 (63)
II	9 (19.5)
III	6 (13)
IV	2 (4.5)
Operation - no. (%)	
Thoracotomy	9 (20)
VATS	37 (80)
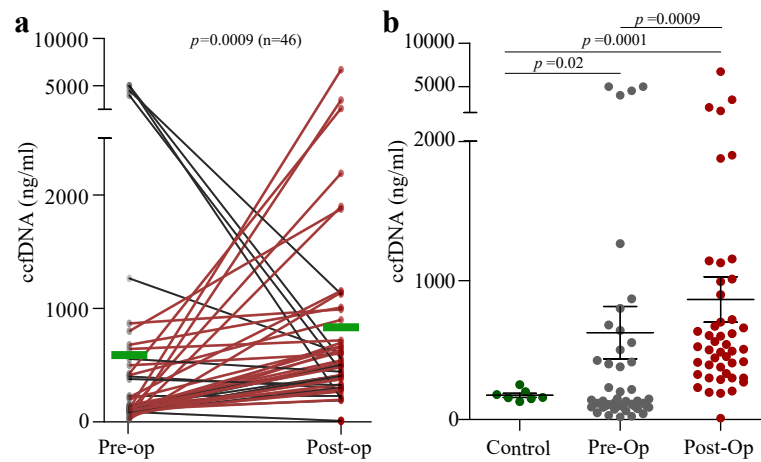
FIGURE 4.1: **Concentration of ccfDNA in pre- post-operative patient samples.** a) Line graph of pre- and post-operative ccfDNA concentrations showing that most of the patients after the surgery presented higher concentration of ccfDNA. b) Significant increase of ccfDNA concentration of the cancer and operated patients compared to control samples.

In combination with the results presented in Chapter 3, we compared the CTC numbers with the ccfDNA concentration. The results showed again that, the post-op samples had an increase in all the parameters tested with the EpCAM+ cells to be higher in both pre-op and post-op samples (Fig. 4.2).
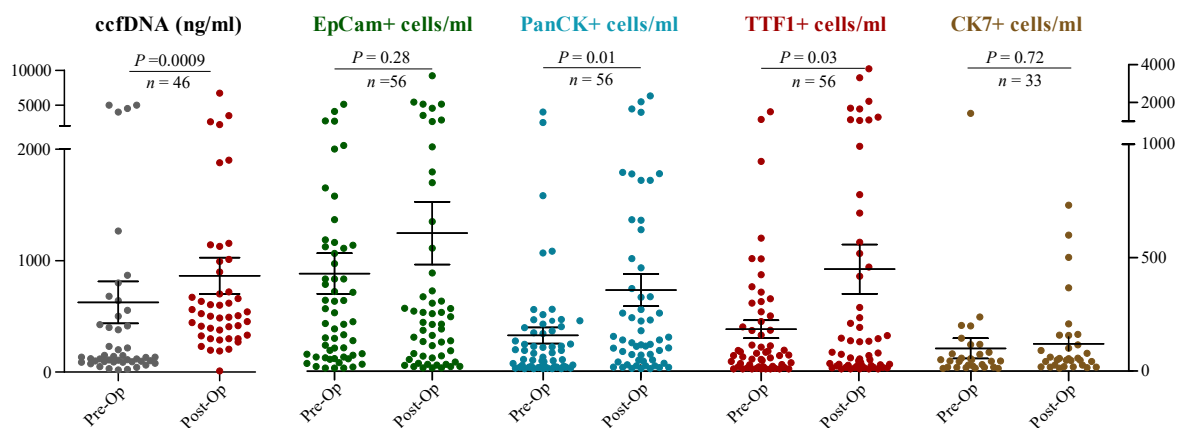
FIGURE 4.2: **EpCAM+, TTF-1+ and PanCK+/CD45- cells in pre-and post-operative patients in comparison with the ccfDNA concentration.** PanCK+ and TTF-1+ cell numbers were significantly elevated after the operation. All paired groups were analysed with Wilcoxon signed-rank test

4.3.1 DNA Integrity Index in lung cancer patients

Changes in ccfDNA following surgical trauma

DNA extracted from n=46 lung cancer plasma samples from pre/post-op patients, were processed by qPCR against the ALU repeat primers ALU115 and ALU247 (as commonly used in literature) [169]. Healthy samples were also collected from N=8 individuals, details of patients are listed in Table 4.1. Data was analysed and DNA Integrity Index was calculated as the ALU247/115 ratio, and plotted as can be seen in Fig. 4.5. In order to calculate the concentration of ALU elements, a standard curve of known concentrations (0.1 - 10.000 ng/ml) of human genomic DNA was made (Fig. 6.1). A single standard curve was drawn for each of the ALU element and the concentration was calculated from the Ct values taken from the qPCR results (Fig. 4.4).
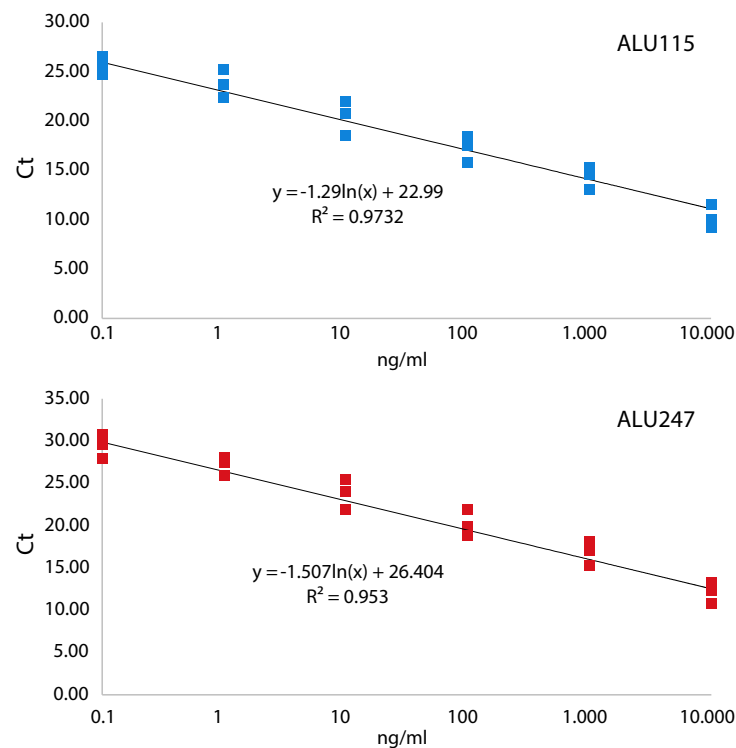
FIGURE 4.3: DNA from human genomic DNA was serially diluted (from 0.1 - 10.000 ng/ml) and used to construct the standard curves from which the quantity of DNA of the patients was calculated.

The ALU115 marker was used for the quantification of cfDNA. Its concentration in pre-operative patients (n=46, median: 441 ng/ml, range 124 – 21.210 ng/ml) was significantly lower than that in the post-operative patient samples (median: 2.624 ng/ml, range 192 – 9.959 ng/ ml with $p<0.0001$). The ALU247 marker allows quantification of circulating cell-free DNA originated from non-apoptotic dead cells. Its concentration in pre-operative patients (median: 125 ng/ml, range 24 – 2.014 ng/ml) was significantly lower than that in the than that in the post-operative patient samples (median: 378 ng/ml, range 15 – 7.480 ng/ml) with a $p<0.0001$. The DNA integrity ratio: [247]/[115] was significant different before and after the surgery ($p<0.046$) but not for the different types of tumours AC ($p<0.75$) and SqCC ($p<0.10$). The ROC curve of serum DNA integrity for discriminating pre-operative patients with LC (n=46) from healthy individuals had an area under the curve (AUC) value of 0.97 (95%CI, 0.91 – 1.02) and post-operative patients an AUC value of 0.94 (95%CI, 0.84- 1.04).
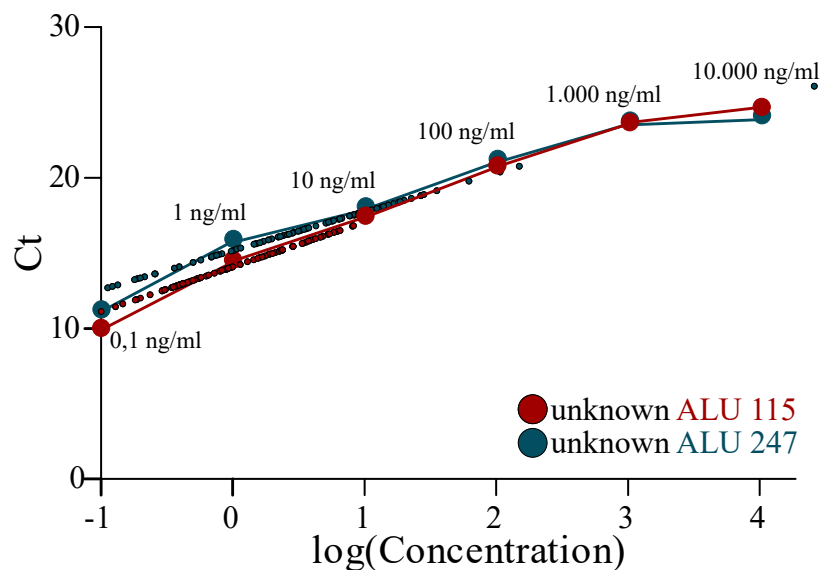
FIGURE 4.4: Patients' ALU repeats concentrations aligned on the standard curve of from human genomic DNA. The black dots indicate patient sample.
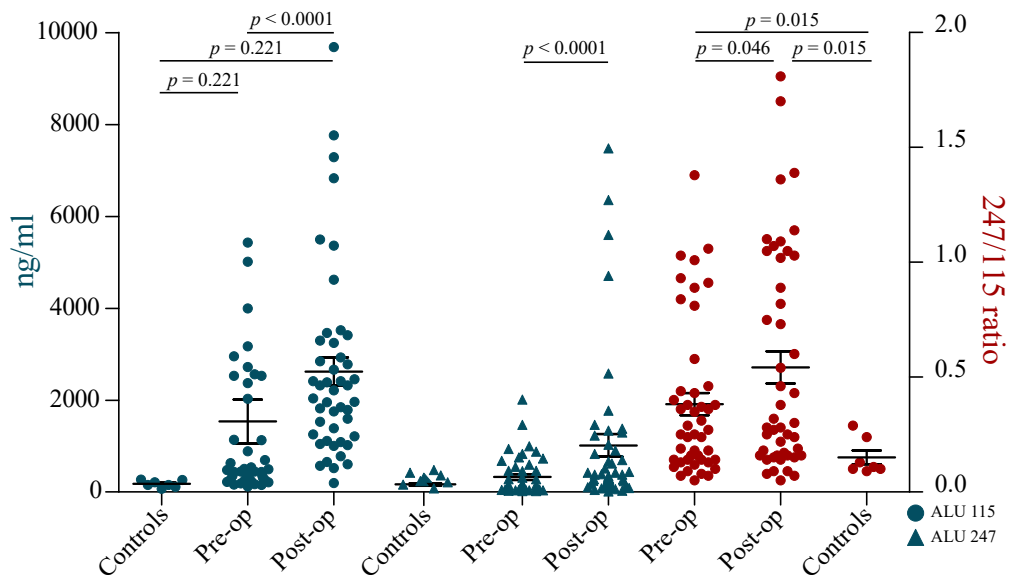
FIGURE 4.5: ALU115 (*p*<0.0001) and ALU247 (*p*<0.0001) concentration of 46 samples (blue) and ALU ratio of 247/115 (*p*=0.046) analysis pre- vs post-op.
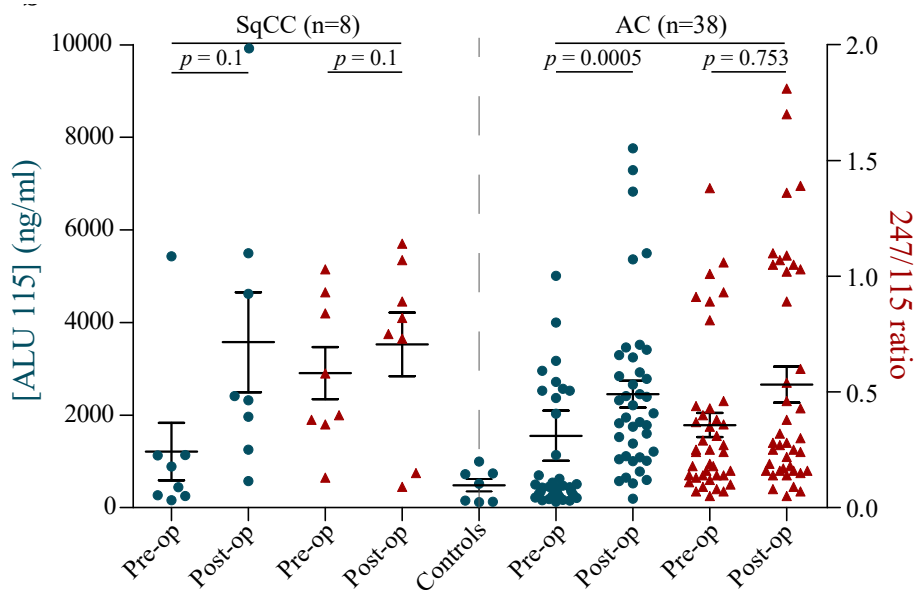
FIGURE 4.6: ALU115 (*p*<0.0001) and ALU247 (*p*<0.0001) concentration of 46 samples (blue) and ALU ratio of 247/115 (*p*=0.046) analysis before and after treatment.
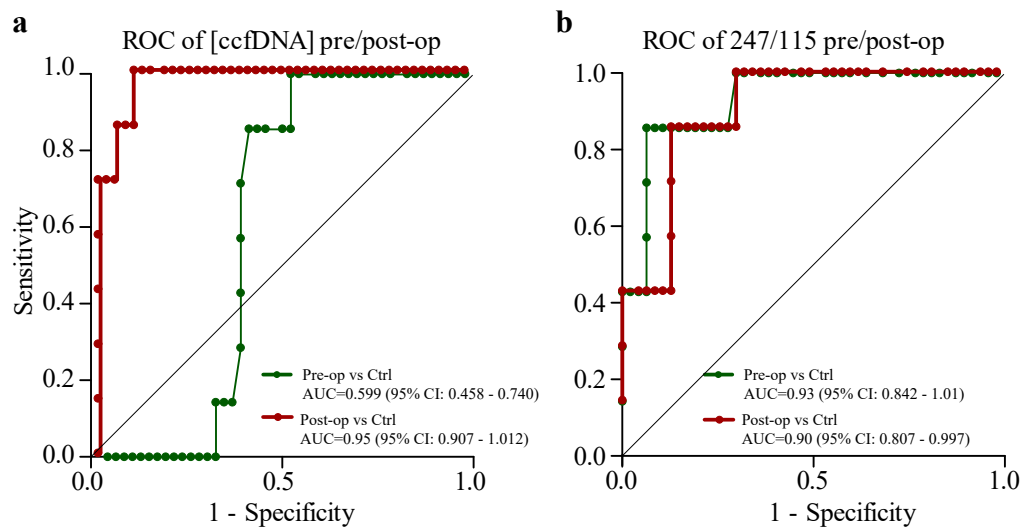
FIGURE 4.7: **ROC curve analysis of diagnostic sensitivity and specificity of using concentration of ccfDNA and ALU repeat ratios** a) ROC curves for ccfDNA indicate that pre-operative patients vs control had an area of AUC=0.59 (95% CI: 0.45-0.74) and post-operative patients vs control had an area of AUC=0.95 (95% CI: 0.90-1.01); b) the ratio of ALU247/115 concentration had high sensitivity and specificity in pre- and post-operative analysis compared to controls.

Results of DNA Integrity Index show higher values are reached in the pre-op cancer group compared with controls (*p*<0.0001) and post-op cancer group (*p*=0.0001). Diagnostic tool evaluation by ROC analysis shows strong diagnostic utility of the post-op ccfDNA concentration compared to controls with AUC of 0.95 and *p*=0.0001 (Table 4.2).

4.3.2   Cytokine microarray analysis in patients' plasma

Changes in circulating cytokine levels following surgical trauma

Blood from lung cancer patients was collected before and three to four days after surgery and plasma was isolated and stored in -80°C (Table 4.3). When everything was ready, plasma was thawed and 50μL were used for a cytokine microarray. Plasma was loaded on a custom kit of a human cytokine/chemokine magnetic bead panel that consisted of 25 analytes to study their expression (Fig. 4.8 and 4.9). The

	Mean ng/ml($\pm$SD))	Significance
ALU115 Pre-op	1534 (3250)	*p<0.0001*
ALU115 Post-op	2624 (2071)	
ALU247 Pre-op	325 (421)	*p<0.0001*
ALU247 Post-op	1015 (1680)	
247/115 Pre-op	0.38 (0.33))	*p<0.046*
247/115 Post-op	0.54 (0.47)	

TABLE 4.2: **ALU concentrations and DNA Integrity Index** Interrogation of DNA Integrity Index show that post-operative patients have a higher concentration in both ALU elements and ALU repeat ratios, compared to the pre-operative ones.

performed multiplex assay demonstrated elevated IL-6 levels post-operatively compared to the baseline pre-operative samples ($p=0.0003$), a result that remained significant even with the exclusion of a single outlier ($p=0.0004$). For the remainder of the cytokines studied, IL-10 was also elevated in post-operative samples ($p=0.034$).

The analysis of the Multiplex results indicates that IL-6 is elevated not only in cancer samples compared to control but also to the post-operation samples ($p=0.0003$) (Fig. 4.8, 4.9, 4.10). IL-6 which acts as an inflammatory cytokine, inhibits the TNFα ($p=0.538$), IL-1 ($p=0.538$) and activated IL-10 ($p=0.034$). As well known, IL-6 plays significant role in the tumour microenvironment regulation and metastasis through the down-regulation of E-Cadherin [200, 201]. Now, following the increase of the Il-6, the IL-10 has been found to be increased in the tested samples (Fig. 4.10). IL-10 which is an anti-inflammatory cytokine which is activated by the IL-6, effects in immunoregulation by blocking the NF-κB pathway and down-regulating the pro-inflammatory cytokines NFα ($p=0.532$), IL-1β ($p=0.532$), IL-12 ($p=0.365$) and IFNγ ($p=0.586$) [202–204]. IL-10 also down-regulates the CCL20/MIP3α ($p=0.1313$) which chemo-tactically attracts the lymphocytes and dendritic cells to the epithelial cells.

TABLE 4.3: **Patient demographics** Demographics and histology of patients recruited for Harefield study and analysed for Multiplex Assay

Variable	Value
Total	19
Mean age ($\pm$SD)	66$\pm$9.7
Males/Females(%)	6 (31) / 13 (68)
Pathology - no. (%)	
Adenocarcinoma	15 (79)
SqCC	4 (21)
Control	1
Mean age	65
Males/Females(%)	1 (100) / 0 (0)
Staging - no. (%)	
I	12 (63.5)
II	6 (31.5)
III	1 (5)
IV	0 (0)
Operation - no. (%)	
Thoracotomy	2 (10)
VATS	17 (90)
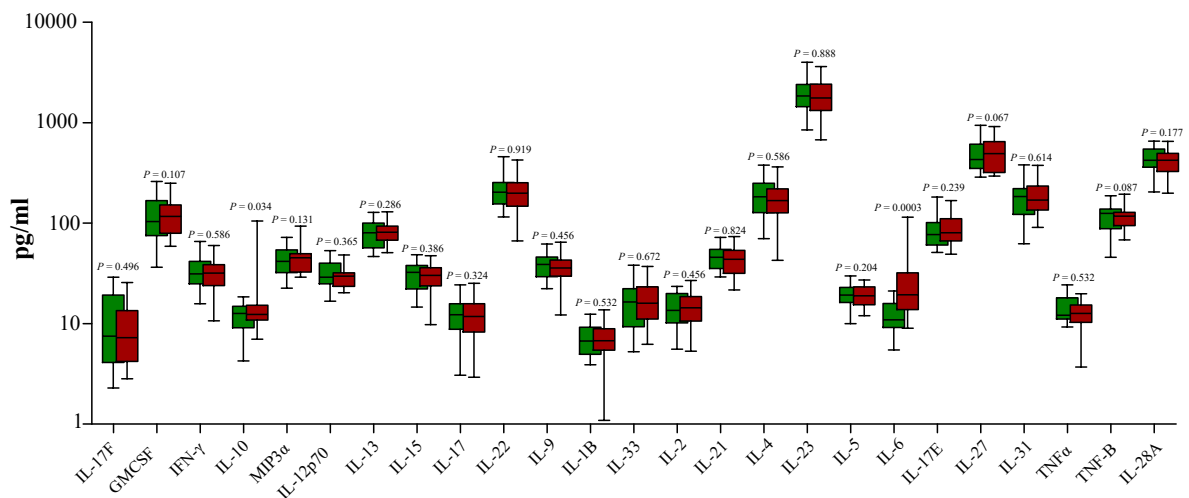
FIGURE 4.8: Multiplex cytokine array analysis in plasma from pre- and post-operative patients.
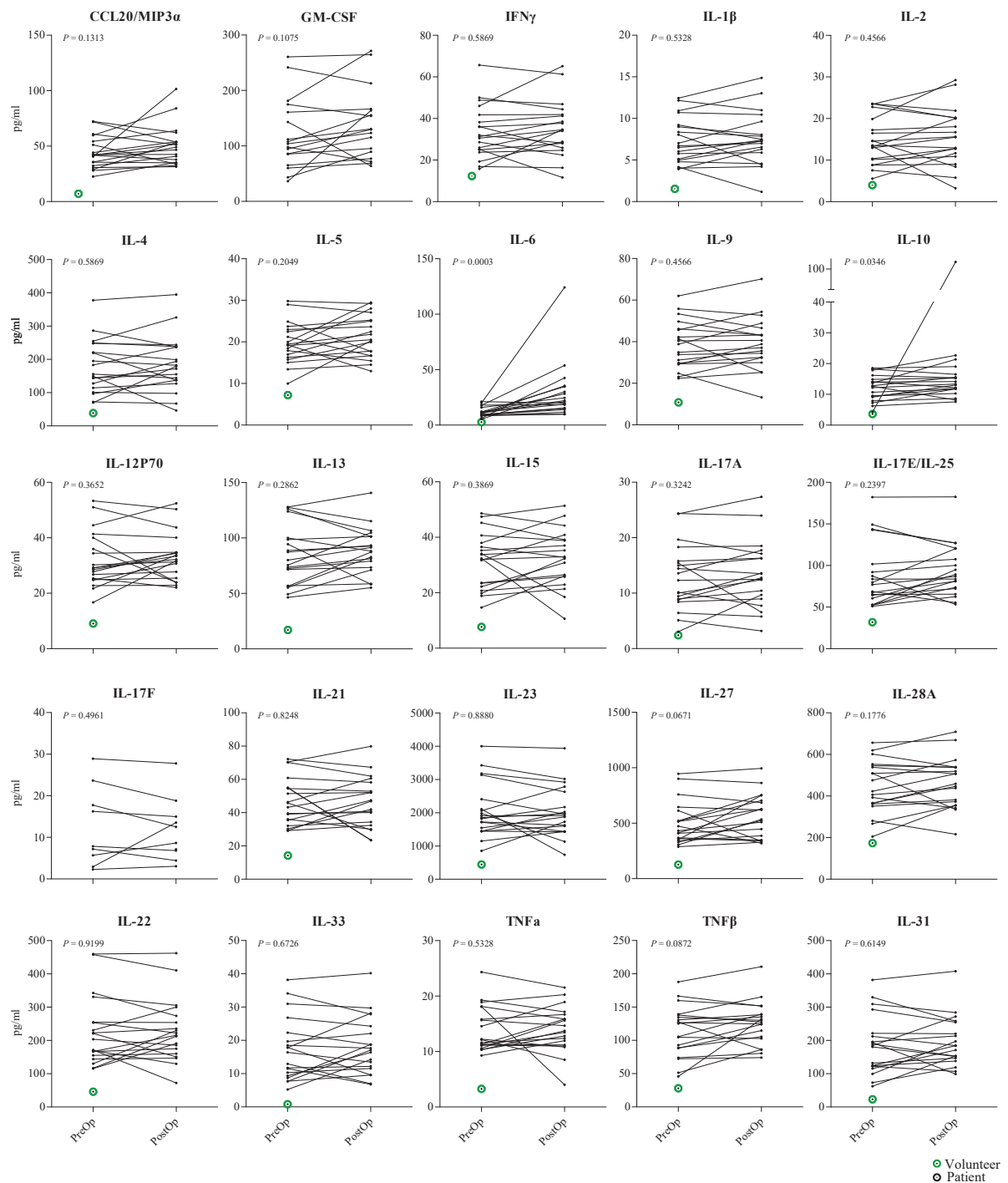
FIGURE 4.9: Multiplex cytokine array analysis in plasma from pre-and post-operative patients patients. All the chemokines detected in higher levels compared to the volunteer samples with the IL-6 and IL-10 having the most significant difference post-op compared to the pre-op samples.
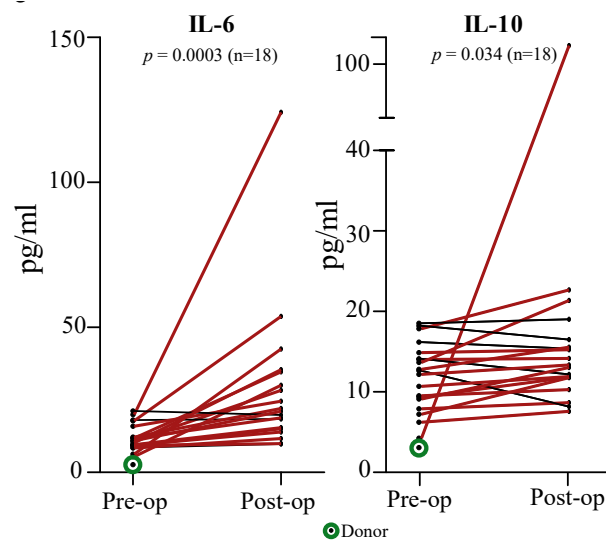
FIGURE 4.10: IL-6 and IL-10 levels before and after surgery.

4.3.3 RNA expression of EMT markers

As presented in Fig. 1.11, during the EMT the cells lose their cell membrane epithelial markers and acquire mesenchymal characteristics. Here we investigated the differences in gene expression of two epithelial (EpCAM and E-Cadherin) and two mesenchymal (Vimentin and N-Cadherin) markers in whole blood lung cancer samples. We used blood from volunteers as a control and as shown in Fig. 4.11, all the markers tested except the E-Cadherin had lower expression compared to the patient ones. E-cadherin, as well as N-cadherin, build the so-called adherence junctions and are needed for proper functioning of cell-to-cell adhesion mechanisms. It is known that higher expression of E-cadherin in normal blood compared to the N-cadherin indicates that more epithelial origin cells are circulating (Figs. 4.11 and 4.12).
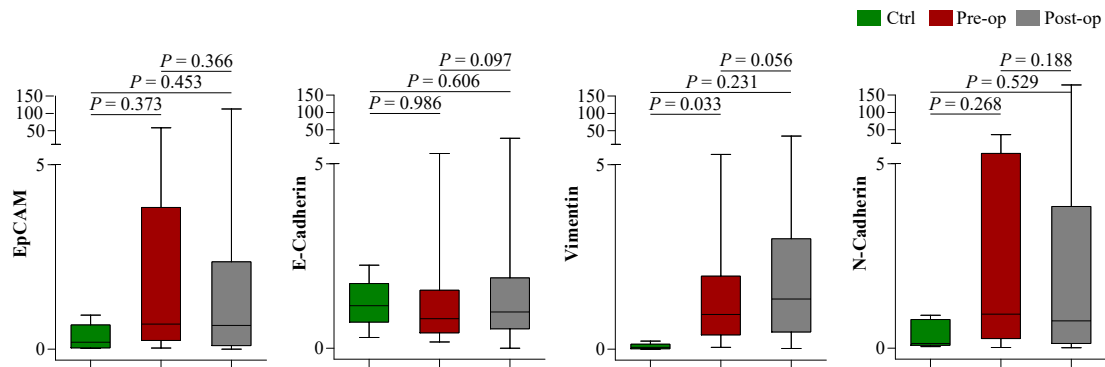
FIGURE 4.11: RNA expression in EMT markers in pre-op post-op patient samples compared to control. For this experiment all 54 patients and 5 control individuals were included but no significant difference was observed in the expression of all markers apart from the vimentin expression of pre-operative patients compared to control samples.
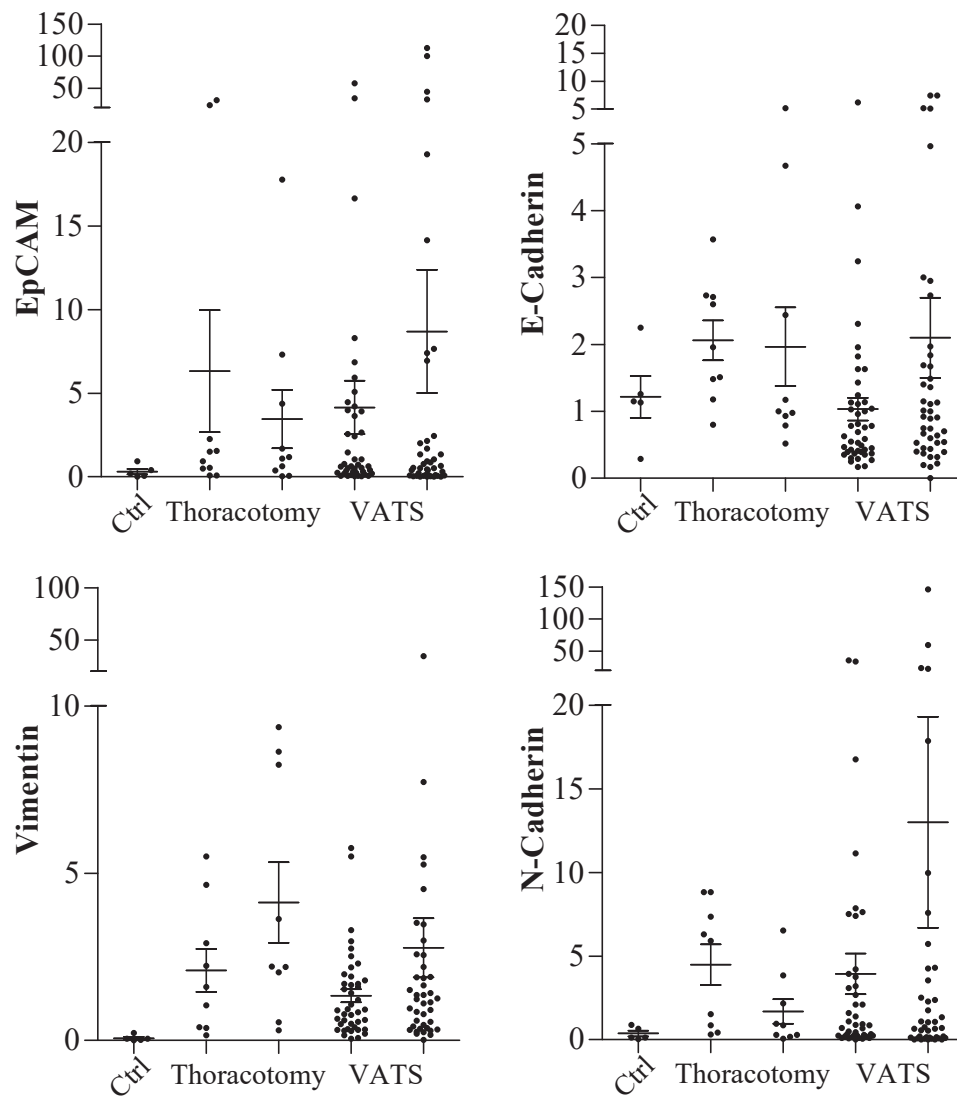
FIGURE 4.12: Comparison of RNA expression in EMT markers in Thoracotomy vs. VATS operation in patients. For this experiment all 54 patients were included but no significant difference was observed in the expression of all markers. Left scatter plots for operations are pre-operative measurements and the right scatter plot of each operation type refer to the post-operative measurements.

4.4 Discussion

Circulating tumour DNA are small fragments of DNA that are released when dead cells get broken-down due to tumour growth. They are less than 200bp long that are found in the bloodstream. However, the quantity of ctDNA varies between individuals as it depends on the location, type, and stage of tumour [205]. Recent studies have shown that ctDNA could be a good diagnostic tool in regards to detecting a tumour. It has been suggested that "detectable ctDNA after surgery for early-stage cancers are associated with a very high risk of recurrence" [206]. Another advantage is that it comes from the entire tumour and therefore better reflects tumour heterogeneity than tissue biopsies. Therefore, analysis can reveal the entire mutational landscape, making it a clear reflection of accumulated mutational patterns present in the main tumour and metastases [207].

To date, several 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, by comparing both CTCs and ctDNA. All the studies report higher detection and concordance with tissue results using ctDNA over CTCs [208]. A variety of efforts has been made to investigate the ctDNA signature in plasma from a clinical point of view. DNA integrity index measurement is one of the many ways to exploit the clinical utility of the plasma ccfDNA. We evaluated the efficacy of the DNA Integrity Index, by means of ALU repeat analysis of ccfDNA found in the plasma of lung cancer patients that underwent operation. As the tumour progresses, the DNA integrity index is expected to be higher since the cell death of tumor cells is more of a necrotic process compared to the programmed apoptosis of normal cells [169]. This information could benefit the clinicians to determine whether there is an advancement of the disease and diagnose patients earlier. Kamel et al. showed that DNA integrity index is significantly higher in breast cancer patients compared to controls and demonstrated a sensitivity and specificity value of 85 and 100% [209].

The amount of ccfDNA in serum can be found to be 2-4 times higher than that in plasma and this is why most researchers use serum for ctDNA preparation [210]. 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 [211]. Other factors such as trauma or stroke can also affect the ALU ratio, demonstrating poor specificity and the overlap in the DNA Integrity Index between the cancer and non-cancer group further corroborates this [212–215].

In this study, we tested the difference in the DNA integrity index in pre-operation and post-operation lung cancer patients to investigate if the type of surgery sheds more circulating nucleic acid material. ccfDNA refers to DNA fragments detected in cell-free components (i.e. plasma) that are released into the bloodstream by various circulating cells. We observed a significant increase (3.8-fold) in ccfDNA following tumour removal. Increased post-op levels of ccfDNA have also been documented in patients who underwent the resection of liver metastases [216]. Similarly, patients undergoing elective surgery for stage I–III colorectal cancer, had elevated levels of ccfDNA persisting up to 4 weeks after surgery [217]. Here, we show the presence of a higher DNA integrity index in the post-op cohort when compared to the pre-op in accordance with the ccfDNA findings. Of note, both measurements were significantly higher when compared to controls as it is reflected by the AUC measurements denoting sensitivity of 97% for the post-op and 94% for the pre-op cohort. This is in agreement with previous studies demonstrating excellent diagnostic potential of this measurement for other malignancies like endometrial and breast cancer [196, 218, 219]. This supports the hypothesis that surgical trauma due to tumour removal leads to higher levels of ccfDNA and ctDNA in this cohort. It has been previously argued that the excess of ccfDNA release post-operatively could hamper the detection of ctDNA, however this does not appear to be the case in this study. Future work will aim to further validate these findings in a larger cohort of NSCLC patients with blood samples drawn consecutively over a period of 5 weeks post-op to determine whether the levels of ccfDNA or ctDNA fall below the pre-op detection measurements. Moreover, the mutation status of these samples should be cross-referenced to the tissue biopsy in order to see whether these liquid biopsies are representative of all the genetic changes of the actual tumour.

There are around 47.800 new lung cancer cases in the UK every year with 38% of them are undergoing surgery, so 18.164 patients. According to sample size calculator for a study like this, and a margin of error at 10% we would need 96 patients.

Based on these results, higher number of patients need to be included in the study not only for CTCs enumeration but also for plasma analysis, cytokines levels and survival curves. Moreover, more control samples should be added in the future study to compare all the potential biomarkers.

Inflammatory changes following surgery are well documented [220]. For this reason, we have measured a wide repertoire of cytokines in a limited sample of our cohort. We have detected a significant increase in IL-6 and IL-10 levels post-operatively when compared with matched pre-op samples. There is some evidence to suggest that certain laparoscopic procedures have been associated with lower cytokine levels when compared with their open counterparts [220–227]. For example, VATS reduced cytokine levels compared with conventional surgery in 38 patients with clinical stage I NSCLC recruited for this study [220]. IL-6 is of particular interest, given that its' release has been linked with post-op systemic inflammatory response syndrome [228]. We acknowledge that a limitation of this study was that we did not randomize the cohorts, hence the majority (18 out of 20) were patients undergoing VATS [180]. Future work should concentrate on measuring IL-6 levels on a larger population and stratify the cohorts on those undergoing VATS vs. thoracotomy. Despite the inherent flaw, it is clear that IL-6 can be a very useful marker to surgically induced trauma.

EMT is a cellular process that epithelial cells develop invasive ability by becoming mesenchymal. This process is very important for the progression and the invasion of cancer. In order for the cells to obtain this ability, the expression of specific proteins and the down-regulation of epithelial ones, is needed. Studies have been reported reduced of the epithelial markers expression like E-Cadherin with a diffuse cytoplasmic staining pattern in tumour tissues [229]. Moreover, Vimentin, a critical pre-requisite mesenchymal marker for tumour metastasis, has been found to be involved in many cancers not only in early stage but also in advanced stages as well [230]. In this project, gene expression of epithelial (EpCAM and E-Cadherin) and mesenchymal (Vimentin and N-Cadherin) markers was investigated although did not reveal any significant increase in the expression of the markers after the surgery and no correlation with the type of operation (Thoracotomy vs. VATS). This study, included 54 patients with post-operative blood taken 2-4 days after the operation. In this study these markers were measured from RNA extracted from whole blood and as a result might mask the actual changes seen in CTCs. Future studies should

focus measuring EMT markers on isolated CTCs only.

Chapter 5

Identification of novel cancer biomarkers of prognostic value: Pseudogene XIST in lung cancer

5.1   Introduction

The large-scale genome sequencing studies of the last decade have documented the pervasive transcription of almost 90% of the human genome [231], with 98% of the transcriptome consisting of long non-coding RNAs (lncRNAs) [232]. However, despite these efforts, our current understanding of lncRNAs' complex biology is still limited. Long non-coding RNAs are a class of RNA molecules that do not encode proteins and range in length from 200 nucleotides to 100 kilobases [233]. Previous analyses have pointed to a wide range of functions for lncRNAs in developmental and cellular processes, including gene expression, chromatin remodelling and modification, splicing, editing, translation and degradation of the RNA, and gene silencing with endogenous small interfering RNA (siRNA) [152, 234, 235]. LncRNAs were found to be expressed in a variety of diseases including cancer suggesting potential roles as biomarkers or even therapeutic targets. Studies have identified key lncRNAs as regulators of oncogenes and tumour suppressors such as PTEN and KRAS [236–238]. Other lncRNAs were shown to be important for genomic imprinting, and for regulating epigenetic procedures. This is the case for lncRNA X-inactive specific transcript (XIST) that inactivates one of the two X chromosomes in females [130, 239, 240]. In literature, XIST is described interchangeably as either a lncRNA or a pseudogene and acts as a major effector in the X chromosome inactivation process.

It is expressed only on the inactive X chromosome providing a dosage equivalence between males and females [130, 132]. XIST was the first non-coding gene identified in the X inactivation center (XIC) region [133]. The expression of this lncRNA is the essential step for the initiation of the X inactivation. XIST is transcribed, spliced, and poly-adenylated resulting in an mRNA molecule, however, no protein products have ever been observed. XIST is coating and inactivates one of the X chromosomes. However, the lack of XIST activity leads to failed inactivation and duplication of the gene on the second X chromosome resulting in its activation [133, 134]. Another key lncRNA that mediates the X chromosome inactivation is TSIX and acts as a XIST repressor. TSIX is the antisense RNA of XIST and their differential expression patterns define the activation or inactivation of the X chromosome. Apart from the XIST-TSIX mechanism, lncRNA XIST and the nuclear matrix protein-heterogeneous nuclear ribonucleoprotein U (HNRNPU)- interact and upon depletion of HNRNPU, XIST is detached from the inactive X chromosome (Xi) and diffusely localized into the nucleoplasm [130, 134]. Recent studies have shown that XIST has an aberrant expression pattern in breast cancer [141], cervical squamous cell carcinoma [142], colorectal cancer [143], gastric cancer [144], glioma [145], hepatocellular carcinoma [146], nasopharyngeal carcinoma [147], NSCLC [138], pancreatic cancer [148], osteosarcoma [149], and ovarian cancer [149]. Furthermore, XIST was shown to regulate tumour cell migration, proliferation, and invasion, in NSCLC [151, 152]. Wang et al. showed that cell lines and patient samples of NSCLC over-expressed XIST and shown that XIST knockdown inhibits tumour growth in vivo. Moreover, XIST exhibits oncogenic properties by regulating the miR-449a and B-cell lymphoma 2 (Bcl-2) gene in NSCLC [235]. The same study has shown that BRCA1 also influences the concentration of XIST on the Xi. Specifically, RNAi of BRCA1 decreases the concentration of XIST on Xi, and the reduction of BRCA1 by Cre-mediated excision also decreases XIST concentration on Xi [153]. Collectively these data indicate that XIST could be an important novel biomarker for the detection of NSCLC [138, 154]. We hypothesise that there is a higher order of complexity in the regulation of XIST and its impact in multiple signalling pathways. In this study, we investigate changes in the transcriptional landscape of cell lines from male and female patients with NSCLC where XIST was down-regulated and studied the expression and correlation of associated genes in lung cancer.

5.2 Aims and Objectives

– Evaluate the role of XIST pseudogene in lung cancer cell lines

– Elucidate the role of XIST in lung cancer from *in silico* analysis and its relation with other genes

5.3 Results

RNA sequencing that was carried out in previous work by our collaborators at the High-Throughput Genomics (Wellcome Trust Centre for Human Genetics, Oxford) on 3 lung cancer matched tissue and blood samples revealed a number of differentially expressed genes in blood and tissue samples.

Our group has previously analysed gene expression signatures in paired tumour tissue and blood compared with control (Fig. 5.1). We identified 21 genes that are differentially expressed in both tissue and blood (Table 5.1).
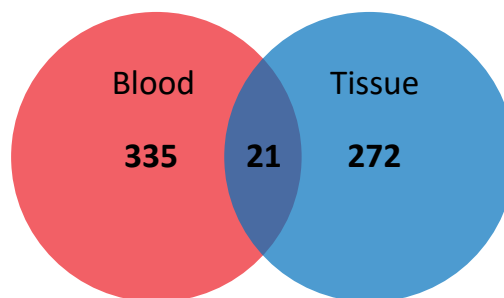
FIGURE 5.1: **Summary of differential expressed genes in blood and tissue.** A total of 272 genes were shown to be differentially expressed in tumour tissue compared to controls, and 335 in cancer blood samples compared to controls. 21 genes show statistical significant differential expression pattern in both tissue and blood samples compared to control (Table 5.1).

We observed that while the majority of genes are down-regulated in patient samples compared to controls, there was only 1 gene [(X Inactive Specific Transcript (XIST)] that was over-expressed in both patient tissue and blood, with no or very little expression in both blood and tissue of control patients. From these 21 genes, XIST was selected for further investigation of its functional role.

TABLE 5.1: **Differentially expressed genes, across blood and tissue samples.** 21 genes were shown to have significantly different expression patterns in blood and tissue samples.

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

5.3.1 Knockdown of the XIST in cell lines

We used siRNA to target the XIST (SMARTpool: ON-TARGETplus, Dharmacon, CO, USA) and silence it in H1975 and A549 cell lines. Transfection efficiency has been determined in previous experiments and optimisation assays using the siGLO-labelled siRNA following the manufacturer's instructions. Images of the cells were taken to visualise any differences in cell growth at baseline, 48h and 72h using the Leica microscope as shown in Fig. 5.2 and 5.3. From the previous scratch assay of NSCLC cell lines, it is shown that cell growth is slower in transfected cells in both the A549 and H1975 cell line compared to their controls. The control cells of both cell lines showed a significant and faster growth into the wound. This was obvious after 48h post transfection. While, on the other hand, transfected A549 and H1975 cells had slower gap closure, which shows that the knockdown of XIST slowed down the gap closure in the transfected cells of both cell lines (Fig. 5.2, 5.3). Those cells appeared under the microscope to be floating in the medium rather than attaching to the plate. Cell death at this point could be due to less efficient transfection; cell line sensitivity to transfection reagents and degree of confluency. Measurements obtained from microscopic images were used to generate the statistical analysis of the difference between the tested conditions. It was measured with microscope and photoshop tools that 1100 pixels = 1 mm, and the equation used was:

$$\text{wound width }\% = (wt = oh - wt = \Delta h)/wt = 0h$$

Wt = 0h: width of the wound measured immediately after scratching (t = 0h) Wt=Δh: width of the wound measured -h- hours after the scratch is performed.

XIST siRNA in A549 cell line
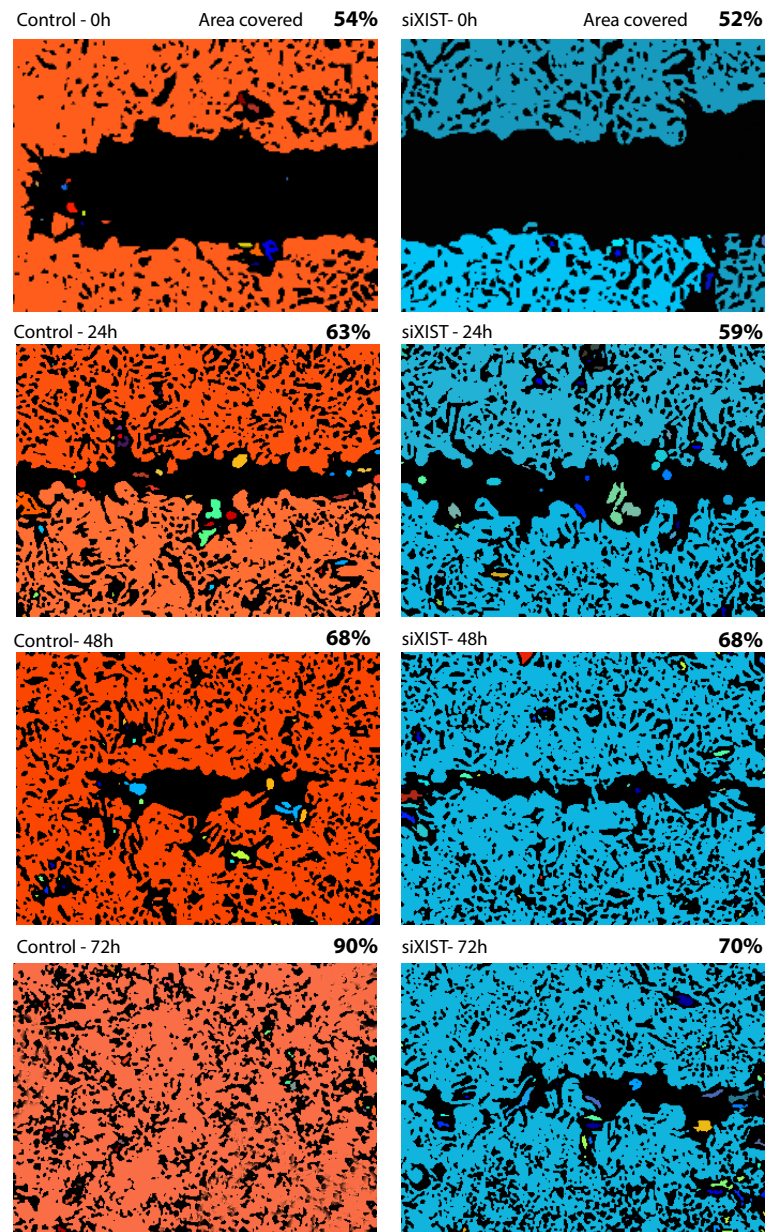
FIGURE 5.2: Wound healing of A549 treated with a non-targeting pool (Control) and siRNA XIST for 48 and 72h. As shown, control cells, covered 90% of the area compared to the 70% of the treated cells.

XIST siRNA in H1975 cell line
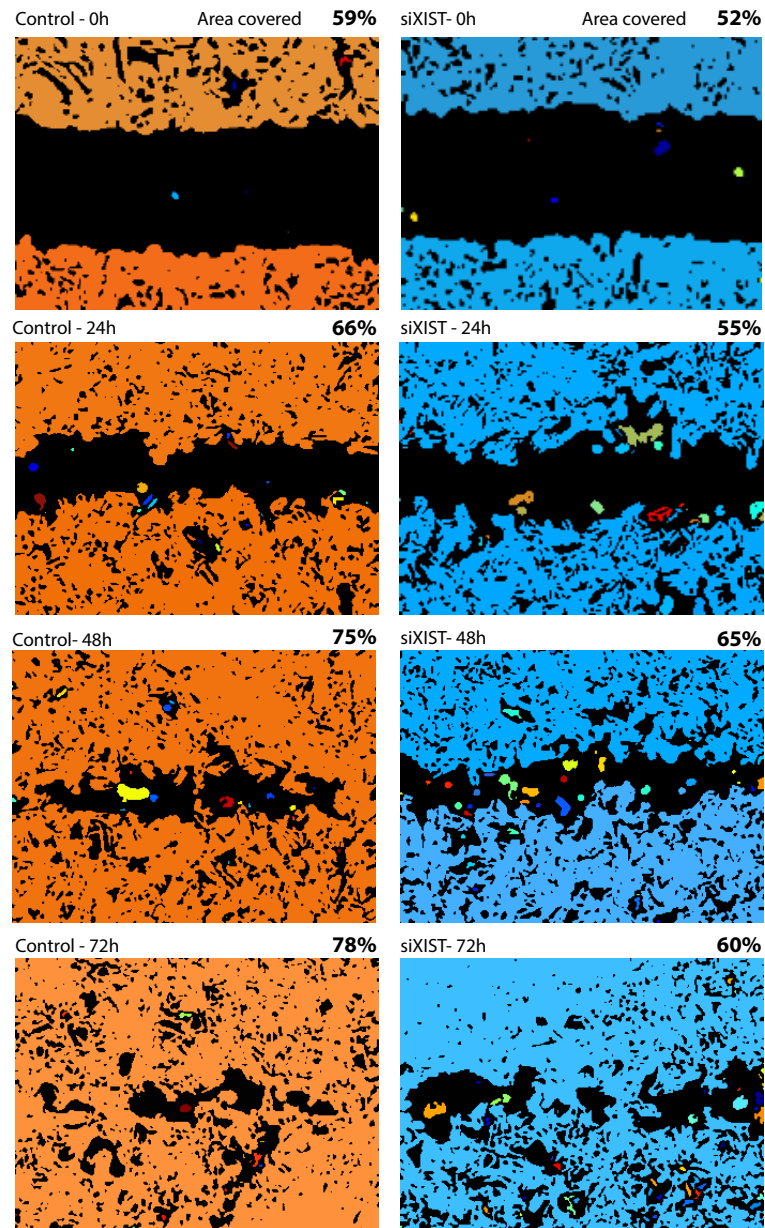
FIGURE 5.3: Wound healing of H1975 cells treated with a non-targeting pool (Control) and siRNA XIST for 48 and 72h. After the 72h, the non-treated sample covered the area at 78% while the treated with siRNA XIST covered it only at 60%.

5.3.2 Effects of XIST silencing

RNA was collected at 48 and 72h post transfection and the level of XIST silencing efficiency was determined be qPCR analysis. There was a down-regulation of XIST after the transfection with the siRNA in both time points.

5.3.3 Transcriptomic analysis

RNA sequencing was carried out by our collaborators at the High-Throughput Genomics, Wellcome Centre for Cancer Genetics, Oxford. RNAseq was run on 2 lung cancer cell lines (A549 and H1975) that had been treated with siRNA XIST for 48 hours. Next, we compared the landscape of gene expression levels in the Non-treated and siXIST samples. The data was plotted as Vulcan plots (Fig. 5.4). We distinguished three types of gene groups. First, there are genes (shown in red) that show a statistical significant differential transcription pattern, Fragments Per Kilobase of transcript per Million mapped reads (FPKM >5, the absolute value of $\log_2$FoldChange >2, p-value <10^{-4}) between the two analysed conditions. 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). What is most obvious is that there is a significant down-regulation for differential expressed genes in the XIST samples compared with NT ones. The fold change is ratio of siRNA XIST over NT. So, if the ratio is smaller than 1, log will be smaller than 0 (so on the left side) and that means that the expression in siRNA XIST is smaller than in NT. Overall, 944 genes were significantly dysregulated after the treatment with the siRNA XIST for the A549 cell line. From those genes, the 683 were down-regulated (*p<0.05*) and 261 significantly down-regulated (*p<5e-5*). In the H1975, 751 genes were significantly significantly dysregulated after the treatment with the siRNA XIST. 536 genes were down-regulated (*p<0.05*) and 115 significantly down-regulated (*p<5e-5*) (Fig. 5.5, full list of genes: Tables A.2,A.3). The two cell lines had 34 down-regulated genes in common with a p value of *p<0.05*, 24 down-regulated genes with a *p<5e-5* with one gene in common (RHOH) that was up-regulated in the H1975 and down-regulated for the A549 (Fig. 5.6). Subsequent analysis of the differentially expressed genes (DEGs) using FunRich revealed distinct biological processes for the two cell lines, including signal transduction, cell communication, metabolism, and energy pathways (Fig. 5.7).
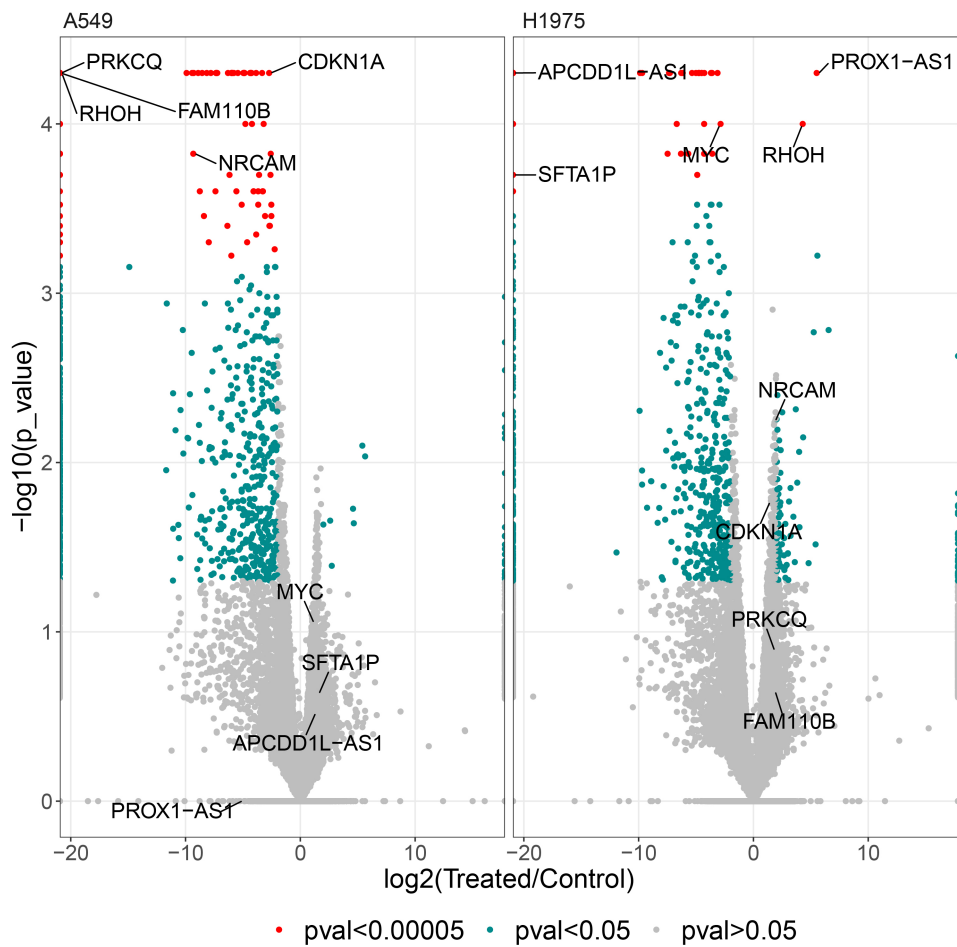
FIGURE 5.4: **Vulcan plots for the A549 and H1975 gene expression**. Data shows mostly up-regulation or down-regulation of genes in the siRNA XIST treated samples in comparison to non-treated ones (left of center: down-regulation, right of center: up-regulation), with a handful of genes shown to be very significantly up or down-regulated.

FIGURE 5.5: Venn diagram showing the total genes significantly dysregulated for both cell lines
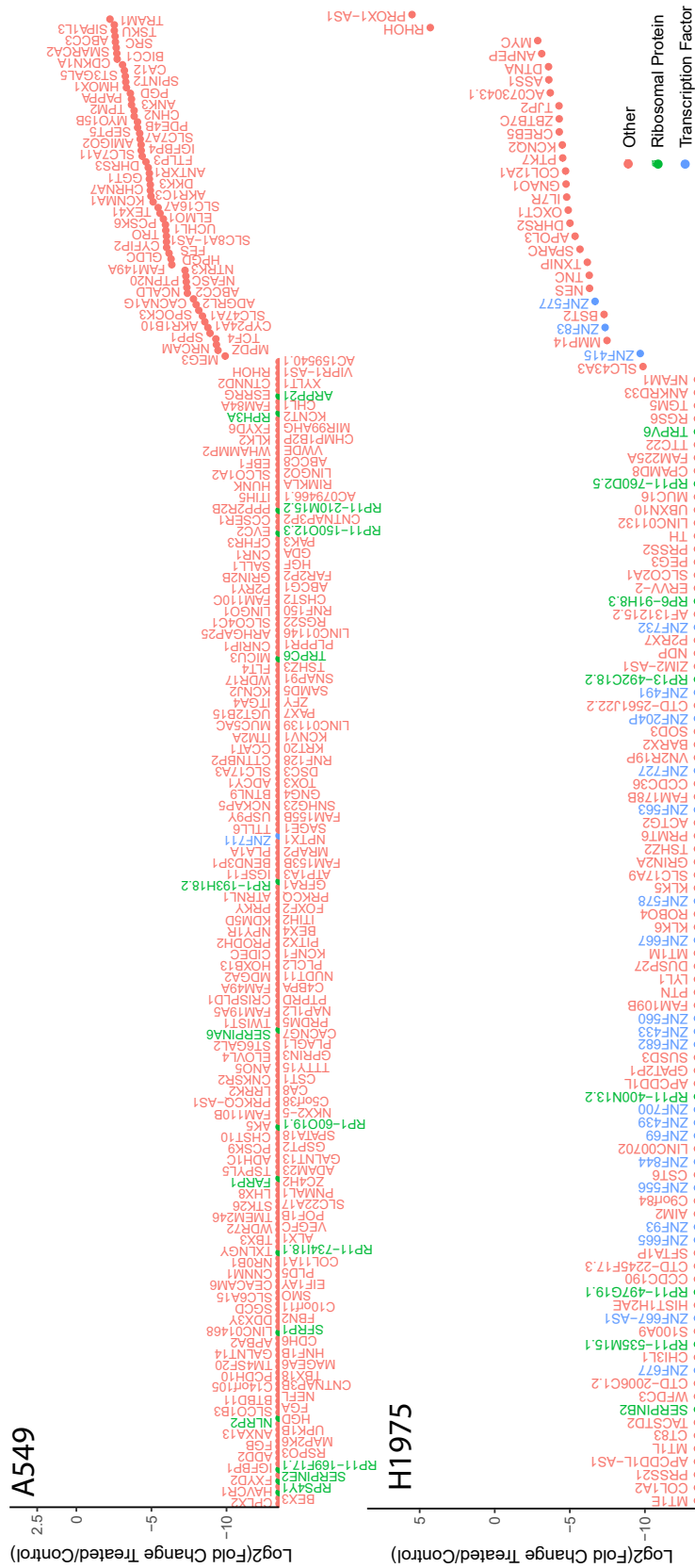
FIGURE 5.6: Ranking of all the significantly dysregulated genes in both cells lines ($p<5e\text{-}5$). Full list of genes: Tables A.2,A.3).
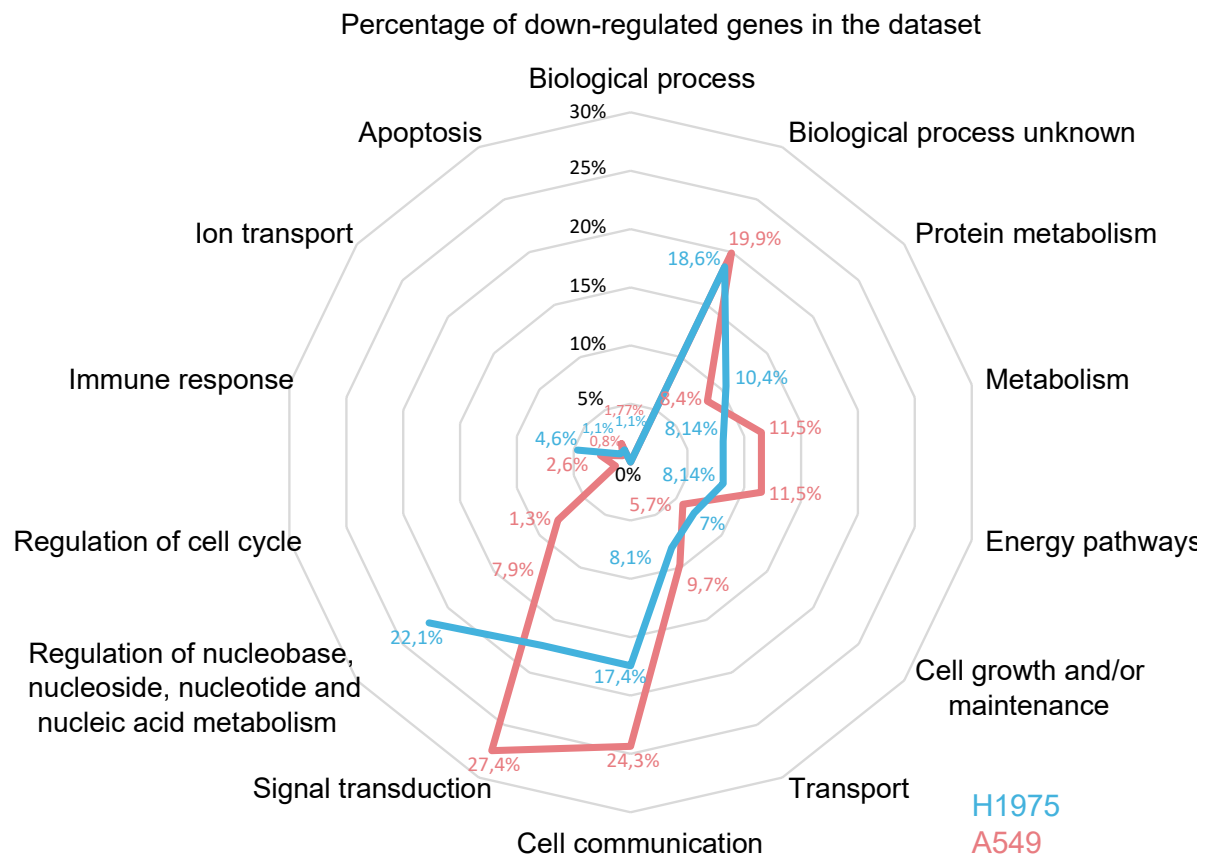
FIGURE 5.7: Summary of biological processes of both cell lines using Fun-Rich. Most of the dysregulated genes in both cells lines are playing a significant role in signal transduction (27.4% for the A549), cell-to-cell communication (24.3% for the A549 and 17.4% for the H1975) and regulation of nucleic acid metabolism (22.1% for the H1975 and 7.9% for the A549). The biological processes where the down-regulated genes are found. Most of the dysregulated genes in both cells lines are playing a significant role in signal transduction (27.4% for the A549), cell-to-cell communication (24.3% for the A549 and 17.4% for the H1975) and regulation of nucleic acid metabolism (22.1% for the H1975 and 7.9% for the A549).

Differentially expressed genes in A549 after siRNA XIST

From all those genes that there was a differential expression after the siRNA transfection, 4 targets for each cell line were selected to be verified with qPCR (Tables 5.2 and 5.3). In detail, in A549, the PRKCQ/(PCK-θ) after the siRNA XIST treatment, was significantly down-regulated. PRKCQ/(PCK-θ) enhances anchorage- independent survival, growth factor independent proliferation, and migration. Byerly et al. showed that down-regulation of PRKCQ/(PCK-θ) in cells enhances anoikis while enhancement of PRKCQ/(PCK-θ) expression can promote growth factor independent growth, anoikis resistance, and migration (Fig. 5.8) [241].

TABLE 5.2: Details on gene targets selected from RNAseq analysis in A549.

Gene Name	Function
RHOH	A negative regulator of cell growth and survival
PRKCQ (PCK-θ)	Required for the activation of the transcription factors NF-κB and AP-1
NrCAM	Required for normal responses to cell-cell contacts in the brain and peripheral nervous system
CDKN1A	Involved in p53 mediated inhibition of cellular proliferation to DNA damage
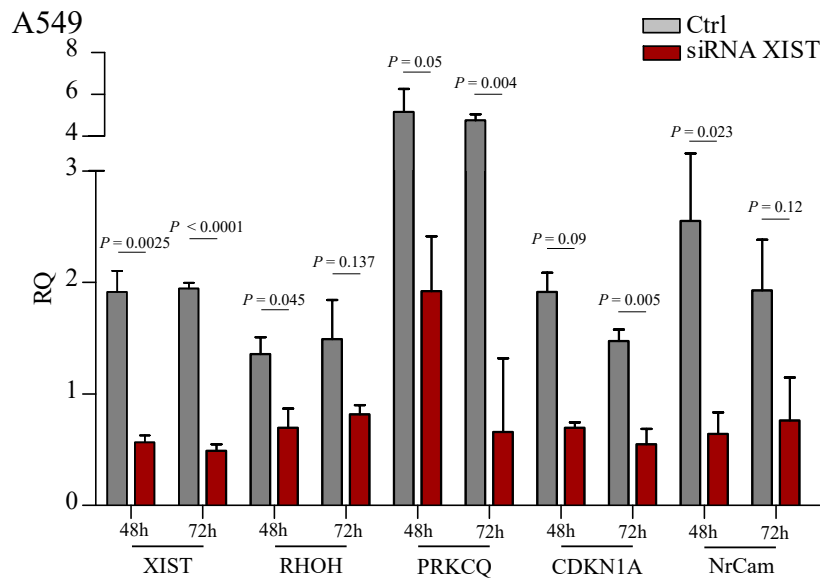
FIGURE 5.8: **Relative expression of selected targets validating RNAseq data.**
For A549 cell line, XIST, RHOH, PRKCQ, CDKN1A and NrCam were validated
using qRT-PCR. Significant down-regulation corroborated the RNA-seq data for
all genes tested

CDKN1A which was also down-regulated after the transfection with siRNA XIST is
a downstream gene target for SOX2. As CDKN1A, SOX2 was also down-regulated
after the transfection and this result comes in agreement with a study by Fukazawa
et al. that showed that SOX2 targets CDKN1A influencing the tumourigenicity in
lung cancer cells [242]. Neuro-glia-related cell-adhesion molecule (NrCAM) has also
been found to be down-regulated after the XIST silencing. NrCAM is up-regulated
in a series of carcinomas like papillary thyroid carcinomas in mRNA and protein
level. The fact that XIST silencing influences NrCAM with a down-regulation could
explain the loose cell adhesion of the cells post transfection. Ras Homolog Family
Member H, RHOH is a negative regulator of cell growth and survival. RHOH can
be related with cancer with two mechanisms. Firstly, it acquires an aberrant somatic
hypermutaion and secondly, it forms fusion transcripts with genes like transcrip-
tional repressors [243]. It activates ZAP70 a protein that is part of T cell receptor
and facilitated the cell signalling pathways. RHOH and ZAP70 that appeared to be
down-regulated after the transfection are a prognostic marker for the different form
of chronic lymphocytic leukaemia (CLL). Lower expression of RHOH and ZAP70
by silencing XIST has to be investigated further to elucidate their connection [244].

In this Rho GTPases pathway, there are more genes that have been negatively affected after the transfection. ARHGAP25, LINGO1, GSPT2 and RGS22 are some of the genes that presented a significant down-regulation due to siRNA XIST.

Differentially expressed genes in H1975 after siRNA XIST

As mentioned above, RHOH-ZAP70 connection is a regulator for CLL [244]. Counter to what happens in A549, RHOH was up-regulated in H1975, lung adenocarcinoma a cell line. Apart from this difference of RHOH over-expression in H1975, none of the other Rho GTPase genes have been affected except the RGS6. The reason of this strange result in the two different cell lines could be down to the fact that H1975, is as EGFR mutated cell line and these mutations may contribute in the aberrant expression of RHOH. The second gene that was up-regulated post transfection was the PROX1-AS1. This gene has been found to be expressed in papillary thyroid carcinoma and when silenced, inhibition of invasion and migration occurs [245]. Silencing XIST, affected 113 genes with a down-regulation. Two of these genes were the MYC and SFTAP1. SFTAP1 is a lncRNA that has been strongly correlated with with advanced tumour lymph node metastasis stage, larger tumour size, lymphatic metastasis, and poor prognosis of patients with gastric cancer. Over-expression of SFTAP1 inhibits cell proliferation, migration, and invasion, thus verifying the tumour inhibitory role of SFTAP1 and experiments have shown that down-regulation of SFTAP1 decreases TP53 which indicated that is acts as a tumour suppressor factor (Fig. 5.9, Table 5.3) [245].

TABLE 5.3: Details on gene targets selected from RNAseq analysis in H1975

Up-regulated	Function	Down-regulated	Function
RHOH	A negative regulator of cell growth and survival	MYC	Regulates cell proliferation, apoptosis, senescence, cell differentiation
PROX1-AS1	Antisense RNA	SFTAP1	Pseudogene associated with SqCC
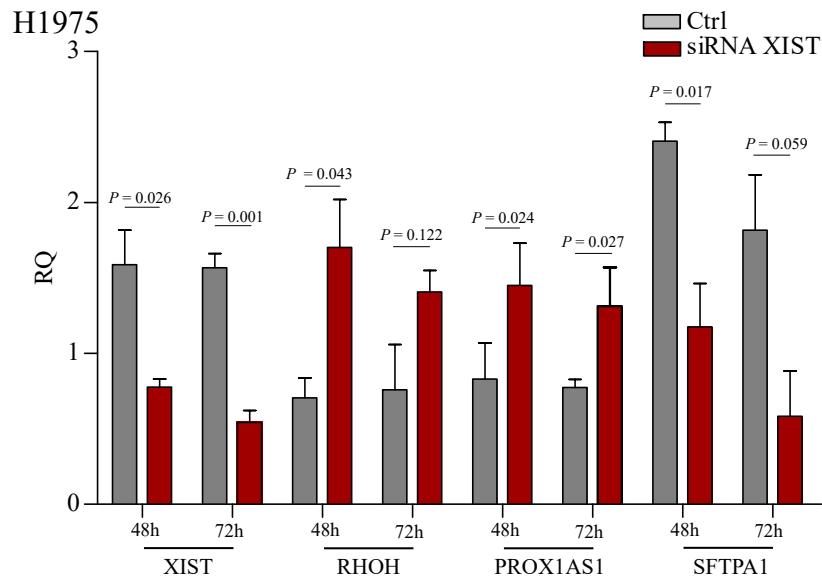
FIGURE 5.9: **Relative expression of selected targets validating RNAseq data**. For H1975 cell lines, XIST, RHOH, PROX1AS and SFT1AP were validated using qRT-PCR. Significant down-regulation corroborated the RNA-seq data for all genes tested.

5.3.4   Bioinformatic Analysis

This study used GDC and GTEx data from XENA Browser. After removing samples from different primary tissues or different diseases, we removed all samples with no gender information, as XIST expression level is gender specific. All details of sample removed and the flowchart of data removing by steps showed in Fig. 5.10.
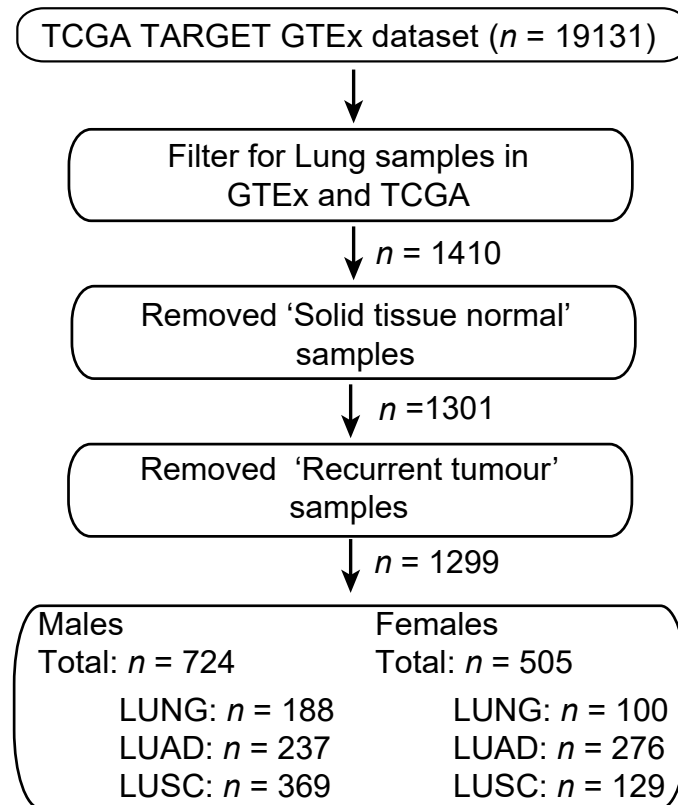
FIGURE 5.10: Flowchart of steps of initial data cleaning, XIST expression data were used originally.

With the remaining data, i first plotted the mean expression of XIST for all the tissues in two graphs to show the expression of XIST in all tissues. I performed Mann-Whitney test on each pair of primary tissue's healthy samples and tumour samples, and obtained the p value to see if they have a statistical significant difference in distribution, then plotted it with mean expression of XIST expression level in the primary tissue, shown in Fig. 5.11. I organised all primary sample into digestive tracts, inner organs, sex specific tissues and other and found that mean expression of all tissue apart from sex specific organs were lied within 5-10 log2(FPKM + 1).

Apart from prostate that gives the lowest expression level of XIST, all other sex specific tissues have a higher mean value of XIST expression level over other tissues. This log2(fold change) versus -log10(p-value), showed that apart from testis and breast, none of primary tissue pairs have an up-regulation of XIST in tumour tissue compare to healthy tissue. All primary tissues pairs were significantly different.
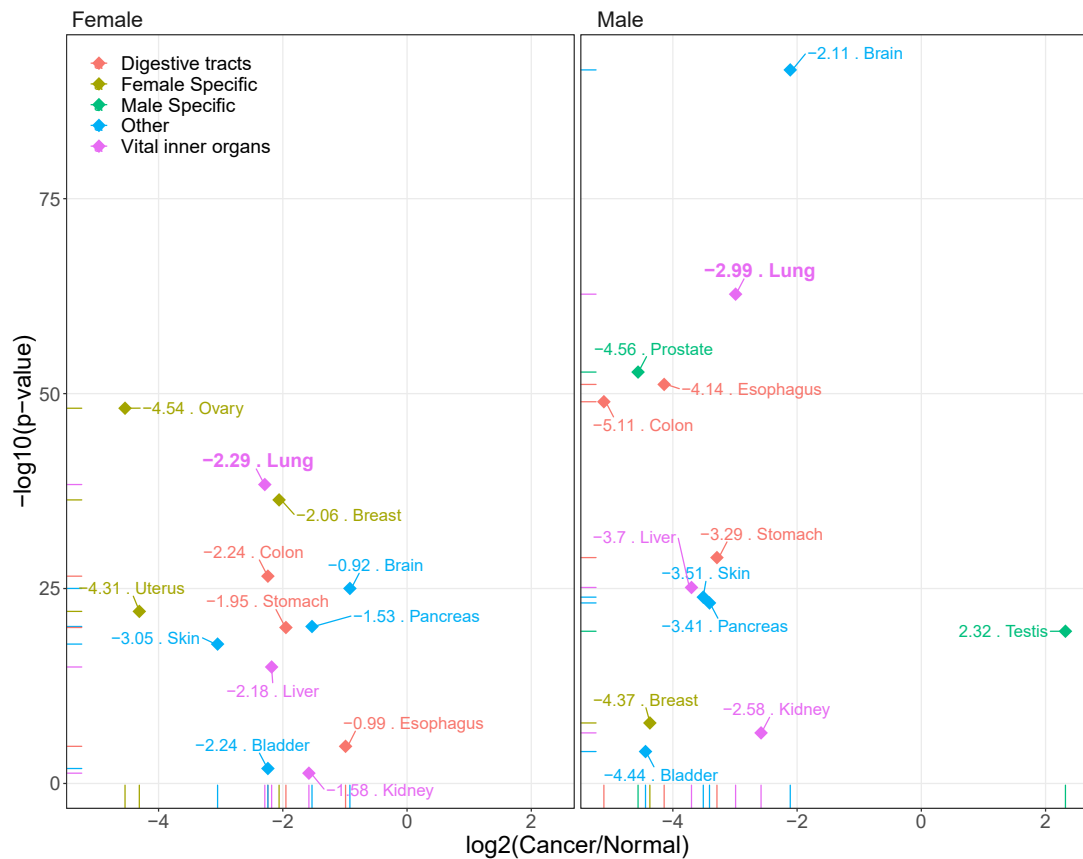
FIGURE 5.11: A log2(fold change) of XIST expression level in primary tissues from healthy tissue to tumour versus -log10(p-value) of Whitney-Mann Wilcoxon test result

After obtaining the gene expression of XIST, I started analysing the lung cancer data set to see whether there is any significant difference in gender, smoking status, pathological stage etc. In the analysis of NSCLC, some samples had to be removed due to missing information in some of the variables to be investigated. The flowchart shows the removed samples by each step in Fig. 5.10.

Expression level of XIST, TSIX and HNRNPU in NSCLC

First, I investigated the distribution of XIST expression between the two types of NSCLC samples, the lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) and also the GTEx normal lung samples. A violin plot with all three types of samples was created and median, interquartile range (IQR) and the sample size of each group (count) were labelled in Fig. 5.12. In the XIST expression level (log2(FPKM + 1)) analysis, outliers calculated by any sample over upper quartile + 1.5*IQR or below lower quartile - 1.5*IQR automatically. High values of p-value indicate a high level of significantly differences between each pair while no outliers have been removed when analysed with median instead of mean. By looking at median, I observed a significantly down-regulation on XIST from GTEx to GDC dataset regardless gender, along with the increased interquartile range.

In male samples, the data indicated deregulation as increased numbers of outliers and IQR in GDC dataset. More interestingly, male population has less expression of XIST in the GTEx dataset, which is different than what Brown et al. suggested, but many samples with zero expression level of XIST have been indicated in LUAD and LUSC instead of GTEx data (Table 5.4) [246]. In figure 5.12a red boxplot of overall XIST expression is showing the big difference between males and females.
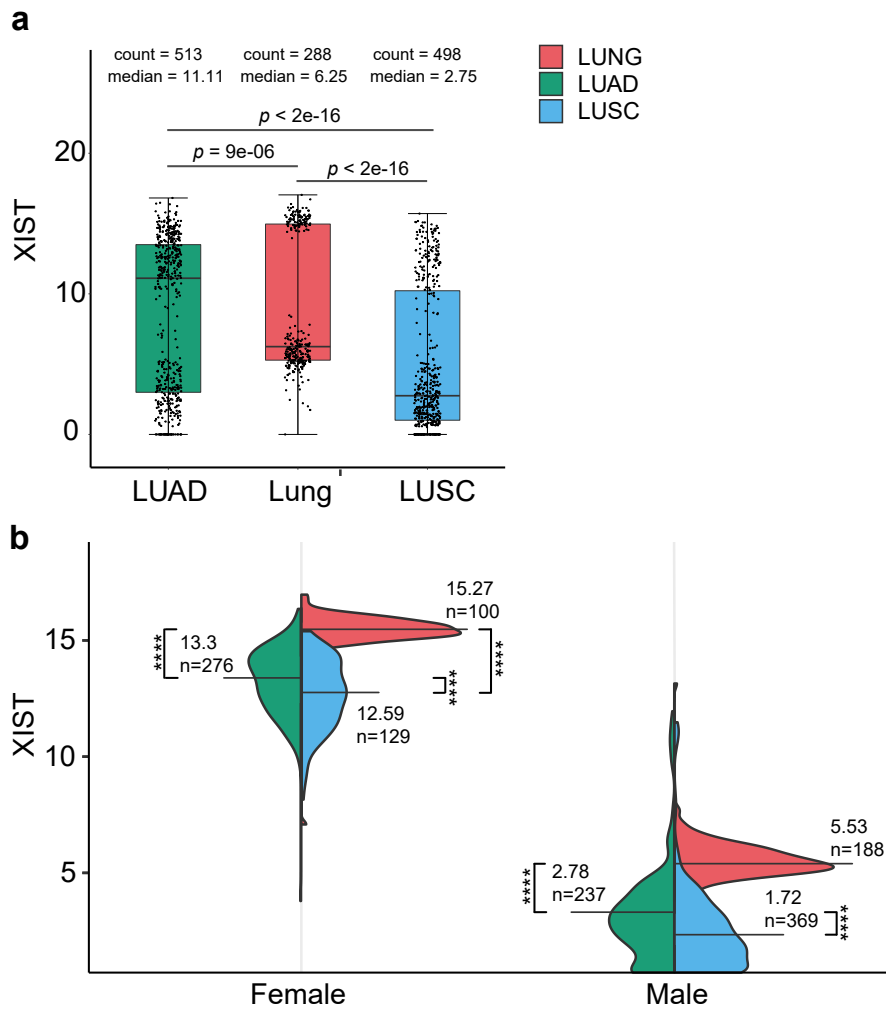
FIGURE 5.12: (a) Boxplots of the overall XIST expression in LUAD, LUSC and normal lung, (b) Violin plots of the gender-specific expression of XIST in LUAD, LUSC and normal lung. Given the two distinct sub-populations, the data has been further presented as gender specific. The violin plot and boxplot of XIST expression level of 3 types of lung sample from GDC and GTEX and their sample size, median of XIST expression and IQR, separate by gender. Between all groups, expression level was significantly different.

TABLE 5.4: The Mann-Whitney results and log2(Fold change) of XIST expression level in lung and NSCLC samples. Fold change calculated by log2(mean expression of group 2) - log2(mean expression of group 1), positive shows up-regulation and negative shows down-regulation.

Gender	Group 1	Group 2	p-value	Significance	Fold Change
Male	Lung (188)	LUAD (237)	<2e-16	****	-2,45
Male	Lung (188)	LUSC (369)	<2e-16	****	-3,34
Male	LUAD (237)	LUSC (369)	1e-06	****	-0,89
Female	Lung (100)	LUAD (276)	<2e-16	****	-2,08
Female	Lung (100)	LUSC (129)	<2e-16	****	-2,75
Female	LUAD (276)	LUSC (129)	0.0004	****	-0,67

We further investigated the expression of 2 other factors closely associated with XIST in NSCLC (HNRPNU and TSIX). HNRPNU is the protein that is required for the localization of the lncRNA XIST on the inactive chromosome X and the subsequent initiation and maintenance of X-linked transcriptional gene silencing during X-inactivation.

The expression level of HNRNPU is stable with no significant changes between tumour and healthy tissues and gender (Fig. 5.13). However, the sample that has no expression of HNRNPU within GTEx lung samples of male is the same as the sample with no expression in XIST which indicates that this zero-expression sample is due to experiment failure and not valid (Table 5.5).
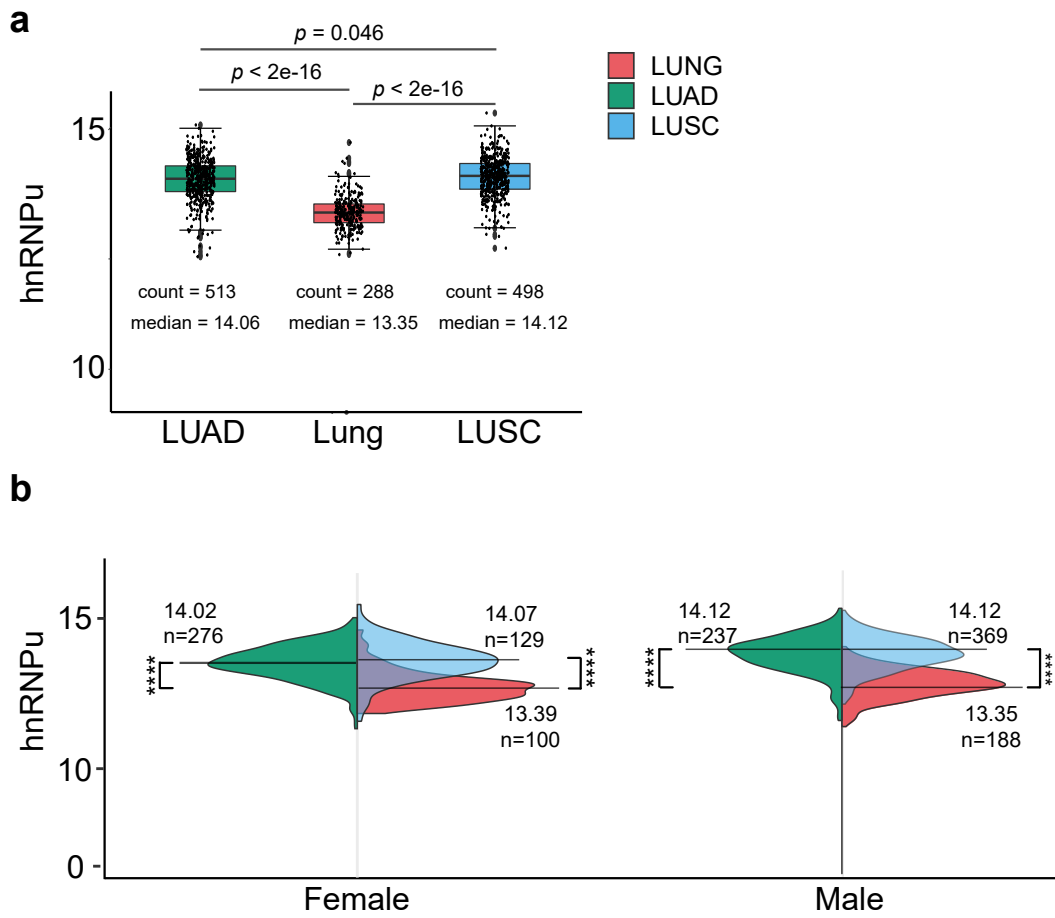
FIGURE 5.13: Expression level of HNRNPU in three different sample groups from GDC and GTEx in NSCLC. Sample size, median and IQR are presented.

TABLE 5.5: The Mann-Whitney results and log2(Fold change) of HN-RNPU expression level in lung and NSCLC samples. Fold change calculated by log2(mean expression of group 2) - log2(mean expression of group 1), positive shows up-regulation and negative shows down-regulation.

Gender	Group 1	Group 2	p-value	Significance	Fold Change
Male	Lung (188)	LUAD (283)	<2e-16	****	0,82
Male	Lung (188)	LUSC (411)	<2e-16	****	0,87
Male	LUAD (283)	LUSC (411)	0.40	ns	0,04
Female	Lung (100)	LUAD (329)	<2e-16	****	0,59
Female	Lung (100)	LUSC (146)	<2e-16	****	0,64
Female	LUAD (329)	LUSC (146)	0.29	ns	0,05

Next, we plotted graph use the antisense repressor TSIX within the same samples (Fig. 5.14, Table 5.6). There was a statistically significant difference between each comparable group pairs in this data for both genders and types of cancers compared to control.
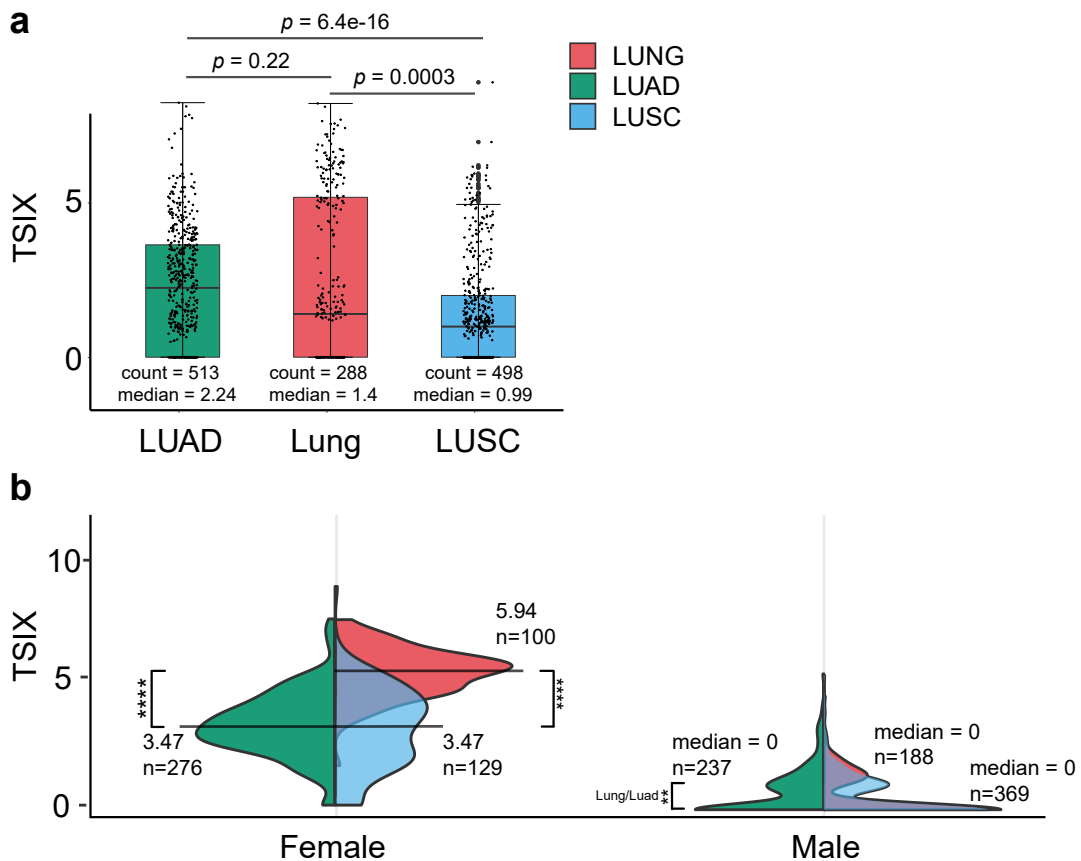
FIGURE 5.14: (a) Boxplots of the overall TSIX expression in LUAD, LUSC, and normal lung, (b) Violin plots of the gender-specific expression of TSIX in LUAD, LUSC, and normal lung

TABLE 5.6: The Mann-Whitney results and log2(Fold change) of TSIX expression level in lung and NSCLC samples. Fold change calculated by log2(mean expression of group 2) - log2(mean expression of group 1), positive shows up-regulation and negative shows down-regulation.

Gender	Group 1	Group 2	p-value	Significance	Fold Change
Male	Lung (188)	LUAD (237)	<0.007	**	0,25
Male	Lung (188)	LUSC (369)	0.068	ns	0,12
Male	LUAD (237)	LUSC (369)	0.287	ns	-0,13
Female	Lung (100)	LUAD (276)	<2e-16	****	-2,31
Female	Lung (100)	LUSC (129)	<2e-16	****	-2,47
Female	LUAD (276)	LUSC (129)	0.589	ns	-0,16

Next, we plotted a graph using the the Bcl-2 and BRCA1 within the same samples (Fig. 5.15). XIST exhibits oncogenic properties by regulating the miR-449a and B-cell lymphoma 2 (Bcl-2) gene in NSCLC [235]. The same study has shown that BRCA1 also influences the concentration of XIST on the inactive X chromosome. There was a statistically significant difference between each comparable group pairs in this data for both genders and types of cancers compared to control.

Overall, we found that the expression of HNRNPU and BRCA1 in LUAD and LUSC is significantly up-regulated compared to the normal lung samples in both males and females (Figs 5.13, 5.15). On the other hand, Bcl-2 appears to be down-regulated only in LUAD when compared to controls (p<2.7e-16). Moreover, the three genes did not display any gender-specific expression patterns as observed earlier for XIST and TSIX suggesting that the transcriptional landscape of HNRPNU, Bcl-2, and BRCA1 is gender agnostic.
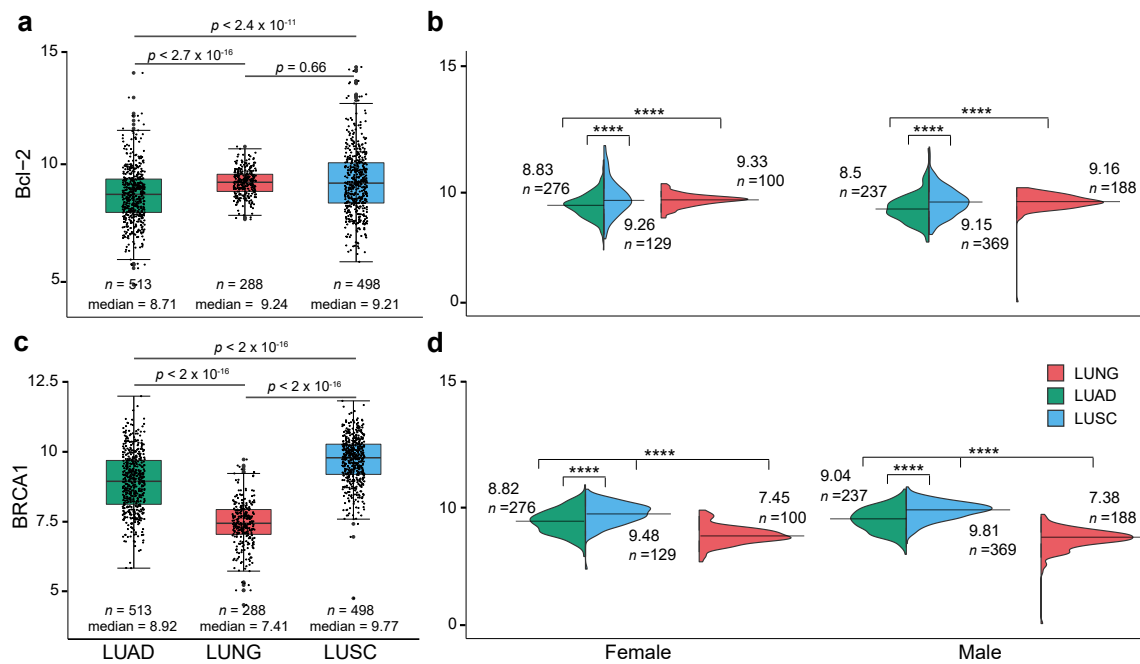
FIGURE 5.15: a) Boxplots of the overall Bcl-2 expression in LUAD, LUSC, and normal lung, b) Violin plots of the gender-specific expression of Bcl-2 in LUAD, LUSC, and normal lung, c) Boxplots of the overall BRCA1 expression in LUAD, LUSC, and normal lung, d) Violin plots of the gender-specific expression of BRCA1 in LUAD, LUSC, and normal lung.

Spearman's rank correlation test of XIST in lung tissues

In order to see if there are any factors associated with XIST expression in NSCLC, we carried Spearman's rank correlation tests. The data were merged with phenotype data of GDC LUAD and LUSC datasets, all downloaded from Xenabrowser and then the dataset was filtered appropriately. First, we investigated the relationships between XIST expression level and the age at initial diagnosis of patients, separated by gender and results are shown in Fig. 5.16. Eight (n=8) samples in male and twelve (n=12) samples in female population were removed due to lack of information. Only the LUSC samples in male showed a small positive correlation which is significantly different, all others showed a smaller negative correlation with a large p-value. This suggested that XIST expression likely have no correlation with age in NSCLC.
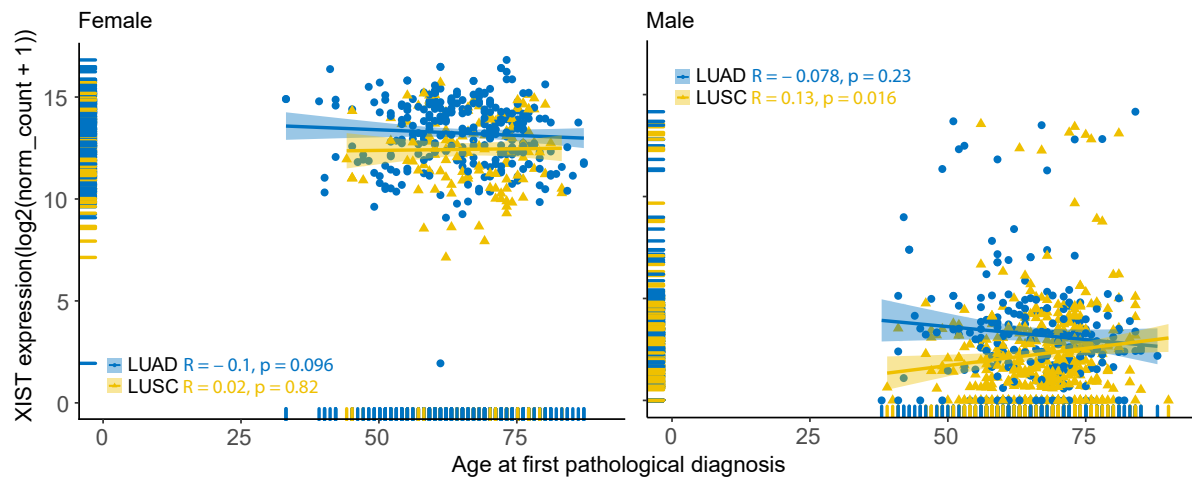
FIGURE 5.16: Spearman's rank correlation of level of XIST expression in LUAD and LUSC in female (n=434) and male (n=644) samples versus their age at initial pathological diagnosis. R represent the correlation coefficient and p: p-value.

Furthermore, I investigated the XIST expression level with cancer staging. I combined the stages IA and IB together as stage 1, the stages IIA and IIB as stage 2, the stages IIIA and IIIB as stage 3 and lastly the IV as stage 4. Nine (n=9) samples for male and three (n=3) samples for female were removed due to lack of data with labels in the dataset as "NA" or "Discrepancy". This result indicated that regardless gender, XIST expression level in LUAD has a relatively small negative correlation, and an even smaller one in LUSC but not statistically significant. oth populations from the two genders gave a very similar result with very similar level of correlation which makes the result more valid and reliable, but due to the small correlation coefficient, it would be reckless to make any conclusion without carefully evaluate other factors.The patient samples in the data set, did not seem to have any differences in XIST expression from the lymph nodes invasion and primary to the metastatic stage for both LUAD and LUSC and both genders (Fig. 5.17, 5.18, 5.19).
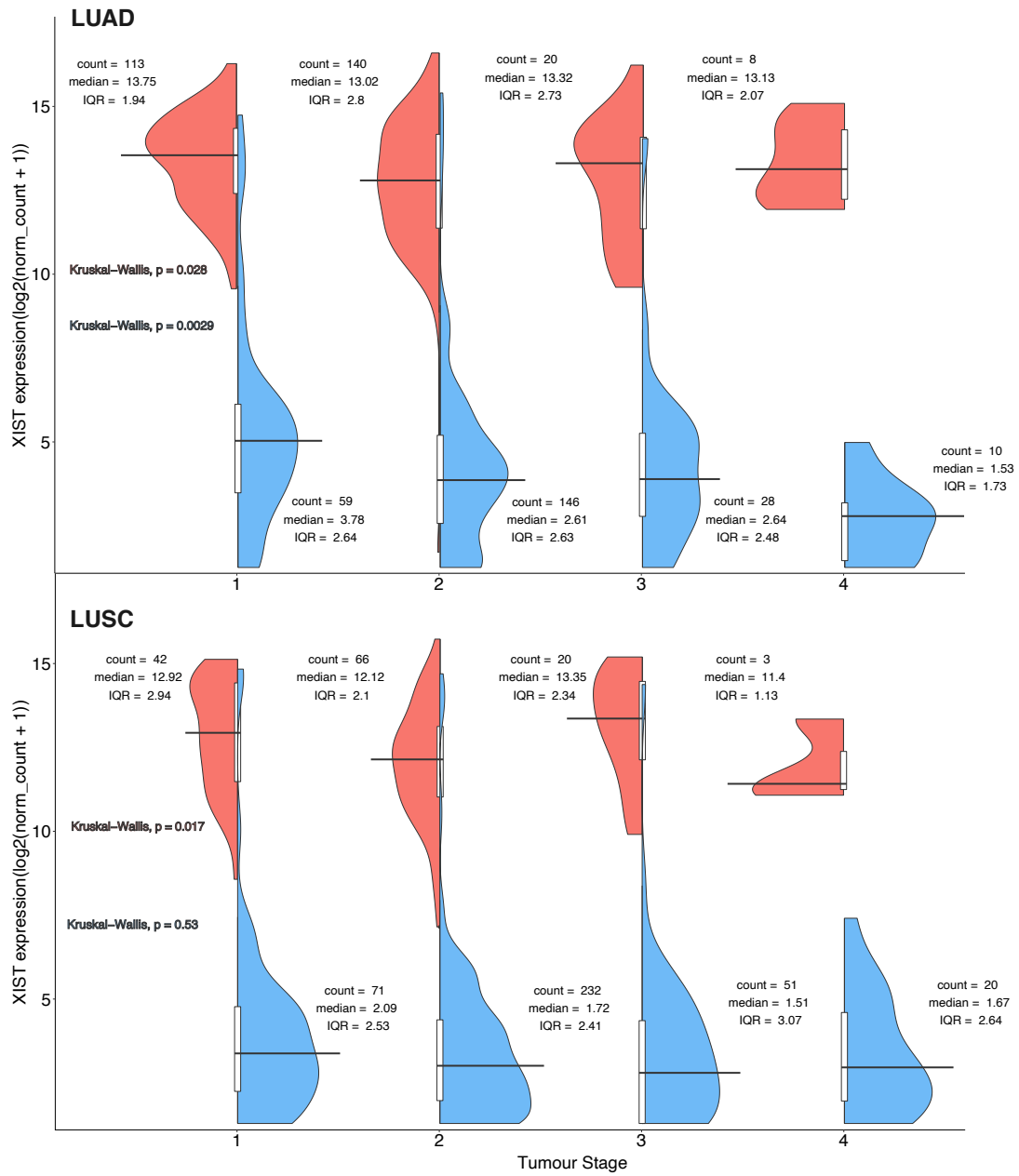
FIGURE 5.17: XIST expression in LUAD (above) and LUSC (under) in female and male samples versus their tumour staging.
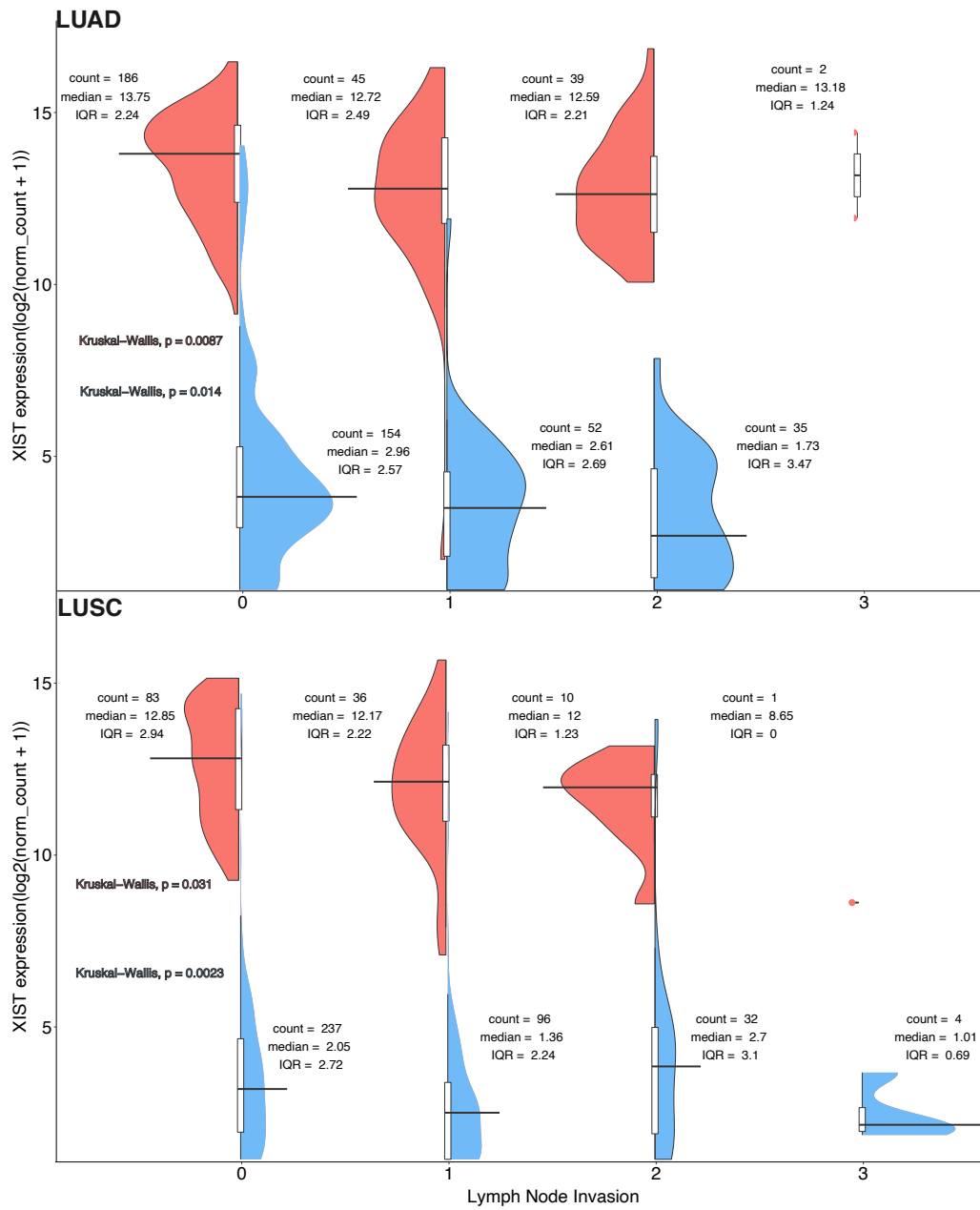
FIGURE 5.18: XIST expression in LUAD (above) and LUSC (under) in female and male samples versus their lymph node invasion status.
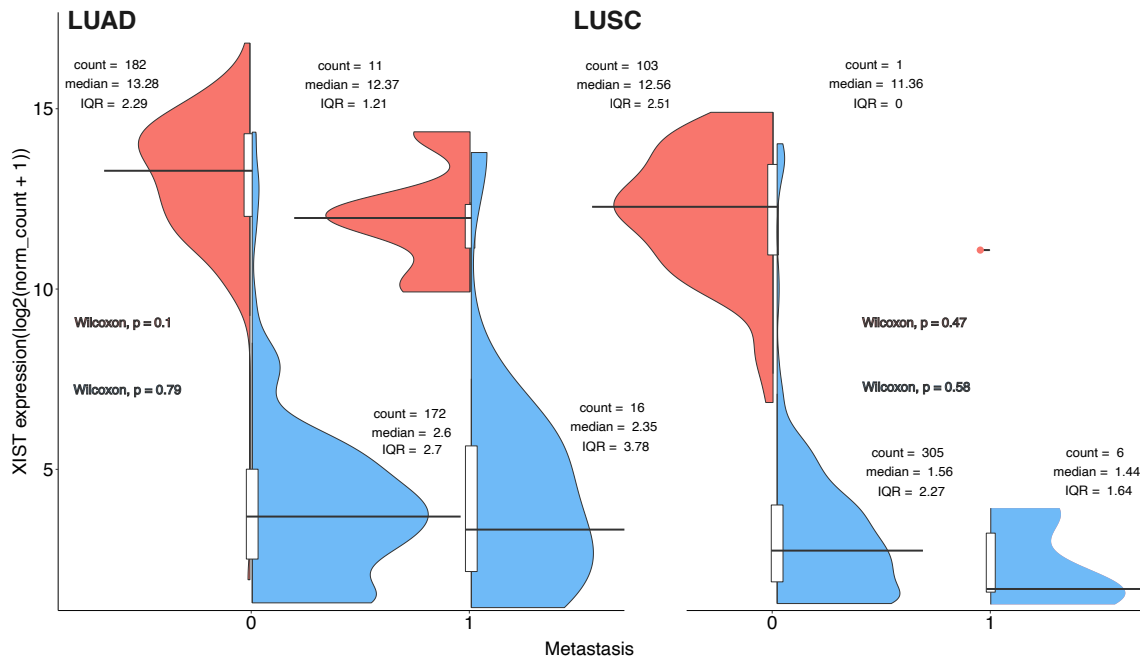
FIGURE 5.19: XIST expression in LUAD (above) and LUSC (under) in female and male samples versus their metastasis status.
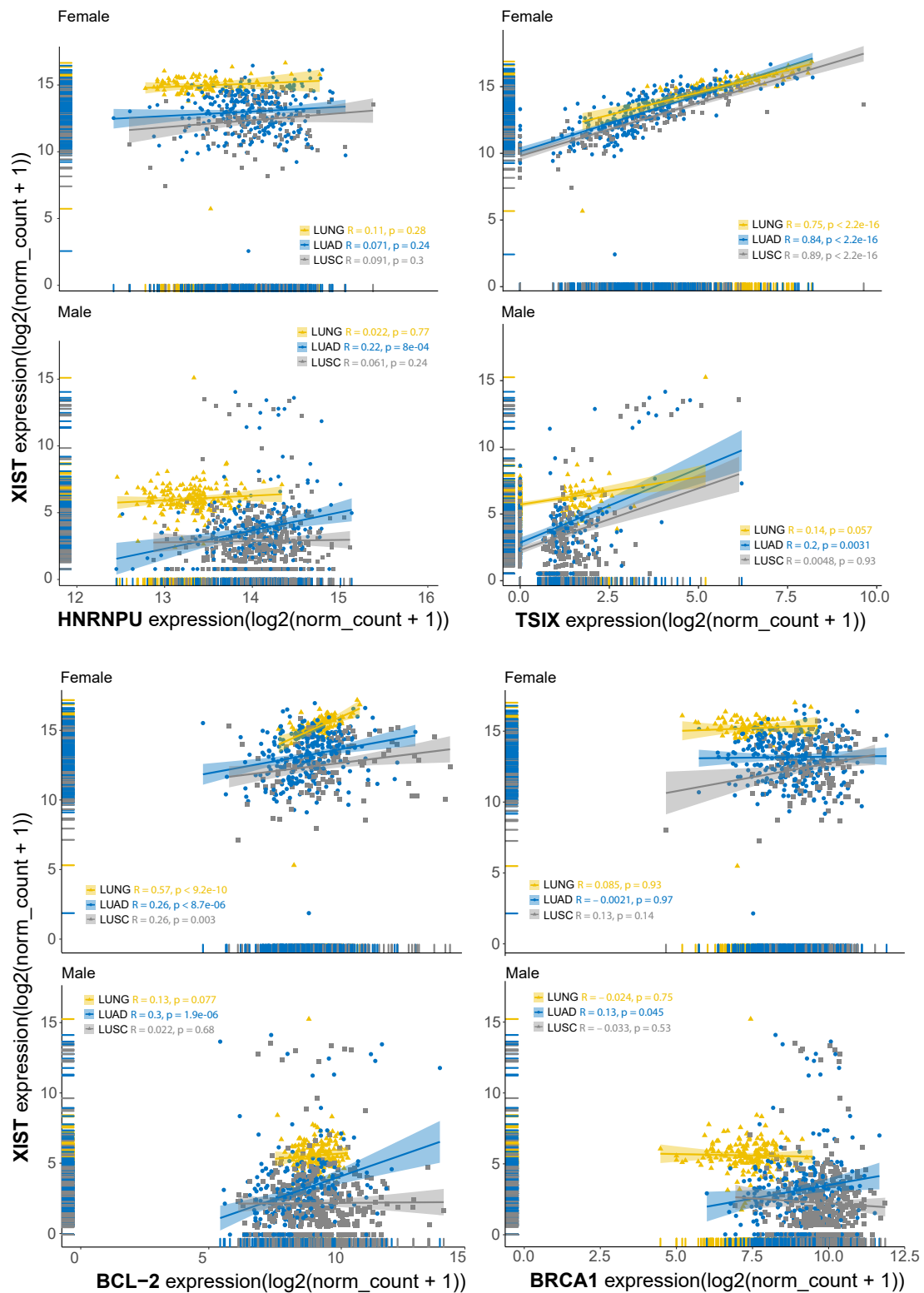
FIGURE 5.20: Correlation of gene expression of (a) HNRNPU, (b) TSIX, (c) Bcl-2 and (d) BRCA1 against XIST. Spearman's rank correlation test was used to generate the correlation coefficient. Lung, LUAD, and LUSC were compared separately in both sexes. Total sample number is 1299, 794 derived from male patients and 505 from female patients.

Next, we investigated whether the XIST and the four associated genes have corre-
lated expression patterns in NSCLC. For this, we used Spearman's rank on test cor-
relation evaluating male and female samples individually (Fig. 5.20). We found that
XIST has an independent expression pattern with respect to HNRNPU, Bcl-2, and
BRCA1. However, as expected we observed a strong correlation between XIST and
TSIX expressions in samples derived from female patients (R=0.75 for Lung, 0.84 for
LUAD and 0.89 for LUSC) (Fig. 5.20). Following the sample and gender-specific ex-
pression analysis, we used T-distributed Stochastic Neighbour Embedding (t-SNE)
to evaluate the discriminatory power of the five genes' expression patterns to dif-
ferentiate between normal and tumour samples. Visual inspection of the t-SNE plot
showed a clear and distinct clustering of both male and female healthy lung sam-
ples to cancer. This clustering of healthy samples to cancer showed a good collective
diagnosis biomarker potential for all five genes (Fig. 5.21). Notably, some cancer
samples are overlapping healthy samples, suggesting that some healthy individ-
ual's expression pattern of the 5 genes is similar to the ones observed commonly in
tumour samples, hinting at the potential existence of cancer precursors in healthy
individuals.
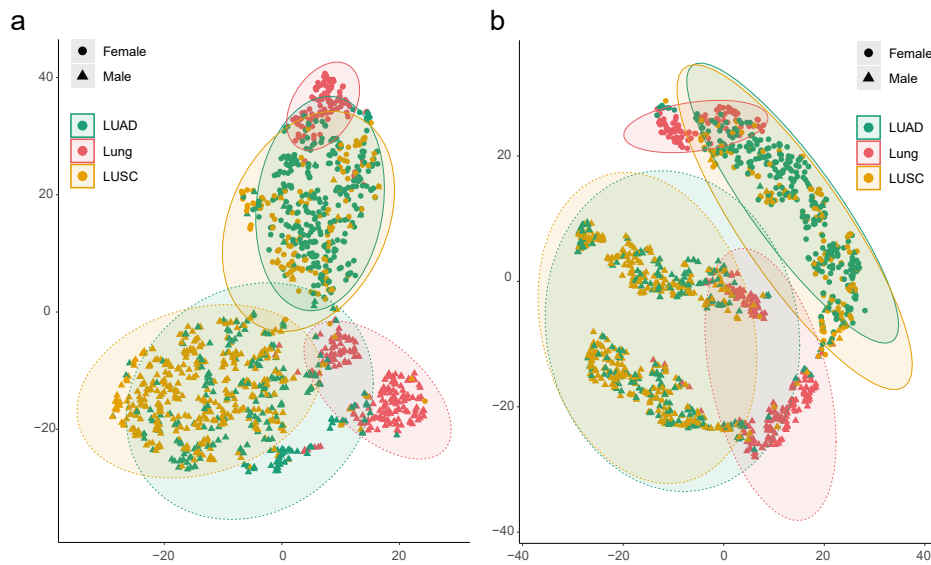
FIGURE 5.21: Clustering of tumour and healthy lung samples in a 2D map
using t-SNE algorithm based on the normalized expression levels of a)
XIST, TSIX, HNRNPU, Bcl-2, and BRCA1 and b) XIST, TSIX and HNRNPU

In order to understand each genes' contribution to the tumour-normal segregation, we evaluated all possible gene groupings using t-SNE (Fig. A.1). Overall, we found that the five gene collection provides the strongest discriminant between normal and cancer samples. After correlation analysis, a heatmap was produced to identify specific clusters or individuals that seem to have a strange expression in the genes examined (Fig. 5.22).
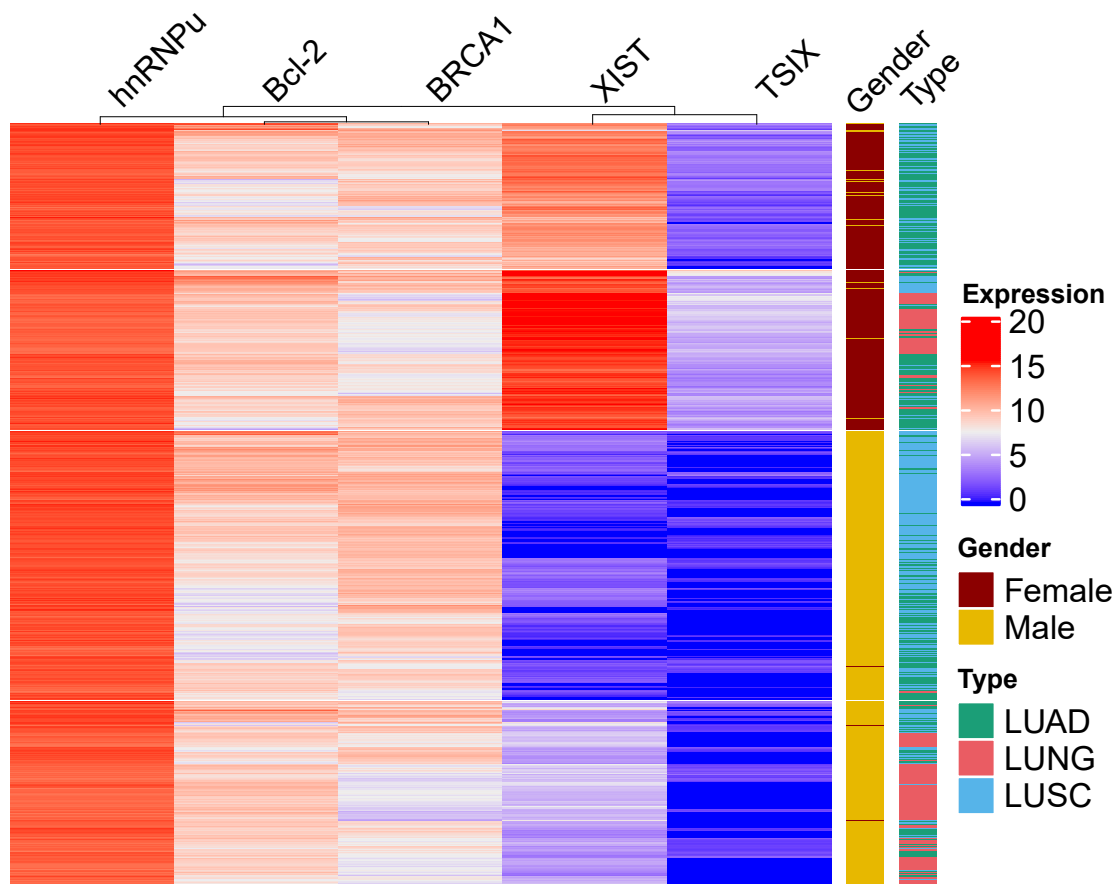
FIGURE 5.22: Heatmap of XIST level and TSIX, HNRNPU, Bcl-2, BRCA1 levels in lung cancer

Survival analysis for XIST expression in NSCLC

In XIST expression survival graph, we compared the XIST expression from cancer samples with the normal expression level in the healthy ones. A simple diagram explained these ranges is presented in Fig. 5.23. As only TCGA data have the survival data matching, we defined normal range ("!") as those TCGA samples having the same XIST expression levels within a range of expression defined by GTEx samples. This range was defined by from upper quartile + 1.5 * interquartile range to lower quartile – 1.5 * interquartile range, which is a common way to exclude outlier. Any TCGA samples having expression level within this range were defined as "normal expression range", where other TCGA samples defined as "non-normal expression range". Two Kaplan-Meier's graphs plotted to compare XIST expression level within "normal range" to any samples with XIST expression level lower than "normal range" for male samples (Fig. 5.24) and female samples (Fig. 5.25).
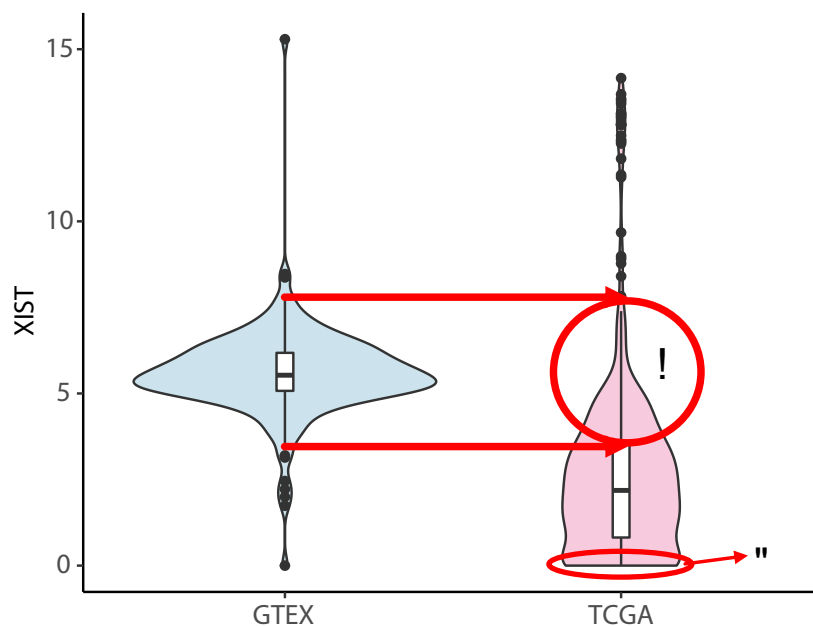
FIGURE 5.23: The '!' Area is defined by the samples from TCGA that have the same ranged expression level of XIST as from a range of GTEx sample, which is upper quartile + 1.5 * IQR and lower quartile - 1.5*IQR. The 'area' is where the XIST has zero expression level in TCGA samples.

None of the KM plots in Figs. 5.24 and 5.25 showed any significant difference, but in female, the line is closer and harder to distinguish between groups of patients

with different expression level of XIST. In males, compared to samples with normal range of XIST expression and those having zero expression gives similar results with no significant differences. Overall, XIST in NSCLC did not show any significant differences thus XIST expression level does not affect survival time of NSCLC patients.
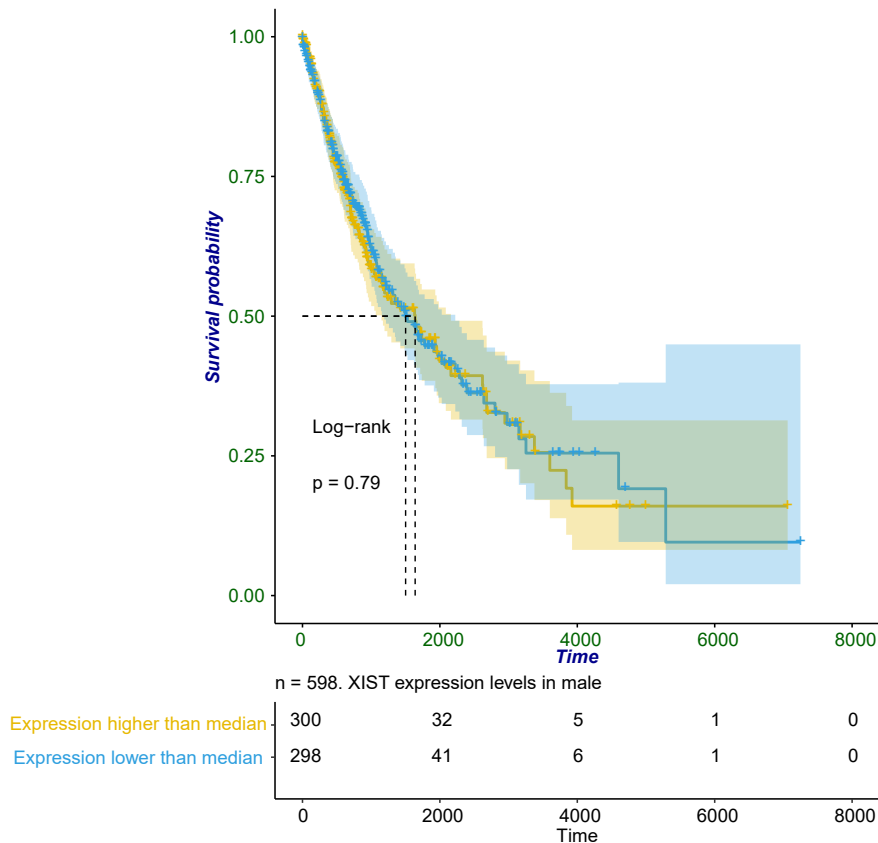
FIGURE 5.24: Kaplan-Meier's graph analysis for male samples regards to XIST expression in NSCLC include LUAD and LUSC.
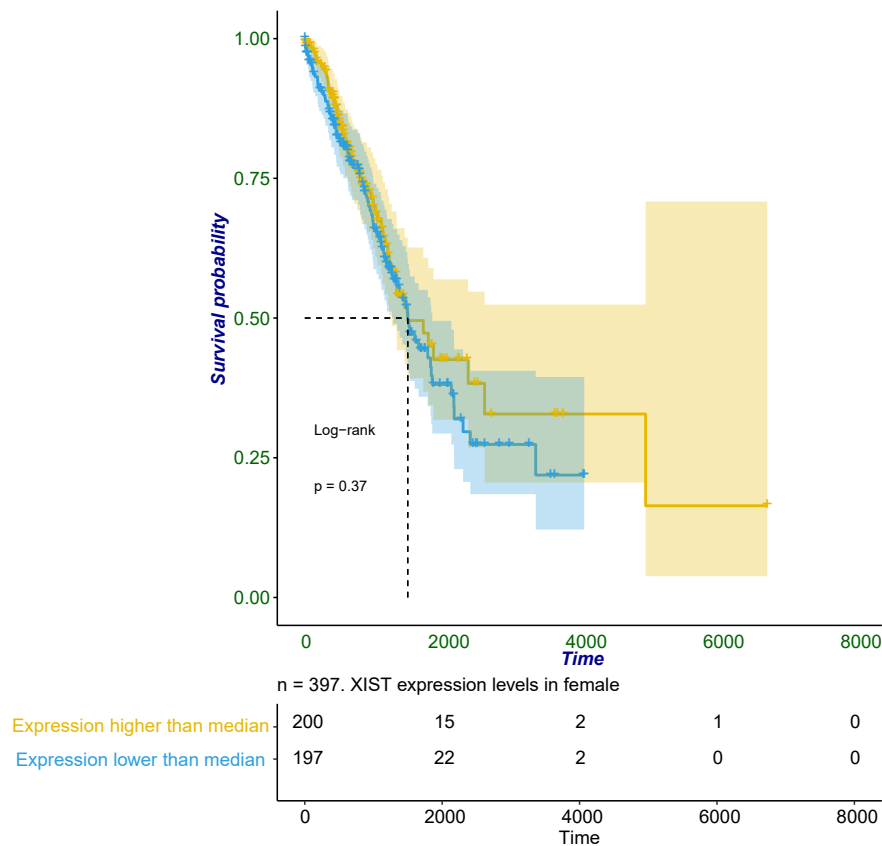
FIGURE 5.25: Kaplan-Meier's graph analysis for female samples regards to XIST expression in NSCLC include LUAD and LUSC.

5.4 Discussion

XIST is being studied in many cancers such as breast cancer, colorectal cancer, cervical cancer, bladder cancer, and prostate cancer and lung cancer. It acts as an oncogene or tumour suppressor for different cancers. XIST is expressed by female somatic cells therefore it is associated with breast, cervical and ovarian cancer. Our focus is particularly on the expression of XIST in lung cancer. LncRNAs play a crucial role in the development and progression of NSCLC [132]. A study carried out on NSCLC samples to investigate the role of XIST in pathogenesis and the mechanisms concluded that XIST was up-regulated in NSCLC compared to normal lung tissue. The paper determined that XIST promoted tumour growth in NSCLC by inducing G0/G1 and promoting proliferation and migration [247]. Another study on XIST, performed a knockdown to identify if XIST acts as an oncogene in NSCLC. XIST

was bound to an EZH2 enhancer that epigenetically silences KLF2 (tumour suppressor) resulting in proliferation and tumour growth and its knockdown resulted in decreased cell proliferation [138]. Data from our wound healing assays corroborates previous findings showing that indeed, silencing XIST results in a reduced tumour growth and proliferation.

MicroRNAs are regulatory non-coding RNA usually 19-25 nucleotides long that play a vital role tumourigenesis especially regulation of EMT [248]. MicroRNA miR-16 has been studied as bio-informatic analysis showed a potential target region in between XIST and miR-16. The results showed that miR-16 was down-regulated and it reversed the effect of XIST [247]. Another microRNA that is widely used for lung cancer is miR-137. miR-137 is suggested to be down-regulated in lung cancer and acts as a tumour suppressor. A study by Wang et al. concluded that the miR-137 expression in NSCLC cells and tissue decreased and XIST expression increased [248]. However, the knockdown of XIST resulted in inhibition of proliferation and TGF-β-induced EMT through XIST- miR-137- Notch-1 signalling pathway [248]. In contrast, a study to investigate the role of XIST in glioblastoma and the effect of the knockdown of XIST found that XIST and miR-137 were both down-regulated in glioma endothelial cells and it was concluded that miR-137 regulated glioma angiogenesis and blood tumour barrier [249]. Another microRNA that seems to be a target for XIST is miR-141. miR-141 could be a potential target for XIST as XIST contains binding sites for miR-141. Evidence has shown that XIST binds directly to miR-141 in papillary thyroid carcinoma cells and their expressions are directly correlated. Similar with NSCLC the knockdown of XIST encouraged metastasis and cell invasion via miR-367/miR-141-ZEB2 axis [250].

In this study, we provide a novel insight into the expression of XIST and key co-expressed genes. Leveraging the available large-scale sequence data from TCGA and GTEX, we demonstrate that XIST, TSIX, HNRNPU, Bcl-2, and BRCA1 are differentially expressed in two different types of lung cancer when compared to controls. Initial observations suggested that XIST was significantly up-regulated in LUAD compared to controls. This finding corroborates previous studies where XIST has been shown to be over-expressed in lung cancer and other cancers [251–257]. However, subsequent stratification by sex revealed that XIST is highly expressed in females and down-regulated in LUAD and LUSC when compared to normal controls. Although this might appear as a discrepancy, it is evident that there is a vast range

in XIST's gene expression when all samples are measured independently of gender. Thus, depending on data preparation and processing a potentially different picture can emerge. For example, in all earlier studies, XIST expression was not stratified in male and female lung cancer patients and numbers were considerably lower to the ones assessed in this study. Of note, our detailed analysis revealed the presence of two groups in control and LUAD cohorts. Our data provides a deeper insight into the differential expression of XIST and highlights the need of standardisation of RNA preparation protocols that would ultimately increase the transcriptomic comparability between different RNA-seq datasets, thus enabling a better mapping of lncRNAs [258, 259].

Further, it is not a surprise that a very strong correlation between XIST and TSIX was seen. TSIX is the antisense of XIST. It is the complementary sequence to XIST that is 40kb long as is transcribed in the opposite direction across the XIST gene. Much like XIST, TSIX only acts on the chromosome it is produced by. The relationship between TSIX and XIST is inverse meaning that when the expression of TSIX is increased, XIST expression is reduced therefore it blocks inactivation in the cis or same X chromosome. When TSIX expression is reduced, XIST expression is increased and causes the inactivation of the X chromosome [118]. One key observation emerging from our t-SNE analysis is the ability to use XIST in conjunction with the four genes of interest as a potential collective diagnostic marker.

Silencing XIST *in vitro* resulted in 944 DEGs for A549 and 751 DEGs for H1975 cell lines. A common trend for both cell lines was that the majority of DEGs were down-regulated compared to controls, providing further evidence for a critical role of XIST in cell proliferation. As mentioned, in A549 cells, PRKCQ/(PCK-θ) was significantly down-regulated. PRKCQ/(PCK-θ) enhances anchorage-independent survival, growth factor-independent proliferation, and migration and its down-regulation enhances anoikis [241]. Of note, cell cycle checkpoint pathways are regulated by PRKCQ/(PCK-θ) in lung cancer [260].

CDKN1A was also down-regulated after the transfection with siRNA. Cyclin - dependent kinase inhibitor 1A (CDKN1A; p21) plays an important role in the progression of the cell cycle particularly in cell differentiation, DNA repair, and apoptosis. p21 responds to DNA damage and is therefore transported to the site of damage to aid in the repair. In addition, p21 functions as an oncogene which promotes the growth of cancer cells by inhibiting apoptosis. This highlights the potential role of

p21 as being a therapeutic response indicator of the pemetrexed-cisplatin combination in NSCLC [261]. Furthermore, another role of p21 is that it can be used as a predictive biomarker of response to therapy in TP53 and KRAS mutated NSCLC [261–263]. Neuro-glia-related cell-adhesion molecule (NrCAM) was also down-regulated after XIST silencing. NrCAM is up-regulated in a series of carcinomas like papillary thyroid carcinomas [264, 265] and significant over-expression of the NrCAM in SCLC was also noted in comparison to normal lung tissue [266]. In addition, stable expression of the NrCAM ectodomain in NIH3T3 cells induced tumourigenesis in mice [267]. Ras Homolog Family Member H, RHOH is a negative regulator of cell growth and survival that was down-regulated in A549 cells upon silencing XIST. Rho GTPases regulate cell migration, proliferation, survival, and death. All these cellular processes are crucial for the maintenance of normal tissues, but also contribute to cancer progression [244]. For example, RHOH expression levels correlate with prostate cancer progression [268]. However, RHOH was up-regulated in H1975 cells. This warrants further research as it might be a case of a differential expression in a cell- or sex-specific manner.

Silencing XIST, lead to the down-regulation of 113 genes in H1975 cells. One of these genes was MYC, a family of regulator genes and proto-oncogenes that code for transcription factors. For example, c-myc oncogene is frequently amplified in lung tumours and has been linked to the malignancy of these cancers [269]. In a conditional model for metastasis of NSCLC, c-myc played a key role in this process [270]. c-myc is over-expressed in 40-75% of NSCLC cases [271]. Of note, inhibiting MYC on NSCLC in mice, lead to rapid regression of tumours with mild reversible side effects [272]. It is possible therefore that MYC can be targeted via silencing XIST.

Gene enrichment analyses revealed that for A549 the main biological processes (in terms of percentage of genes involved) are signal transduction, cell communication, energy pathways, and metabolism (genes list: Table A.2). The same subset of dysregulated genes appears to be involved in energy pathways and cell metabolism and the involvement of some of them in lung cancer is well documented. For example, a down-regulated genes from A549 lung cancer cells treated with siRNA XIST, expresses higher CYP24A1 mRNA and has oncogenic properties mediated by increasing RAS signalling in lung adenocarcinoma [273, 274]. Whereas AKR1B10 has been shown to induce the metastatic potential of lung cancer cells to the brain in

vitro, using an organ-on-a-chip model. MicroRNA-1304 inhibited the growth of an NSCLC cell line by targeting HMOX1 [275].

In terms of cell signalling, gene enrichment revealed several different candidate genes for these processes. CHRNA7 is up-regulated in NSCLC and associated with an unfavourable prognosis [276]. An interesting target of XIST is HGF (or MET proto-oncogene) that is aberrantly expressed in NSCLC and has become a drug target over recent years [277]. Indeed, dual MET and SMO negative modulators overcome resistance to EGFR inhibitors in NSCLC [278]. Another target of XIST was the proto-oncogene siSRC that promotes metastatic events in NSCLC and has become a target for cancer treatments [279]. Two other genes correlate with poor outcomes: high IGFBP1 expression is correlated with poor OS in NSCLC [280]. Whereas, VEGF-C expression is significantly correlated with lymph node metastasis in NSCLC [281].

Silencing XIST in H1975 cell line affected many different processes including cell communication and signal transduction involving genes like MUC16 (CA125) that its levels relate to different stages of NSCLC (genes list: Table A.3) [282], and PTK7 that is associated with lymph node metastasis, ALK and EGFR Mutations in lung cancer [283]. In addition, PTN another gene effected by XIST is a heparin-binding growth factor that is involved with tumour progression [284]. In the same cell line, XIST affected the expression of 13 different zinc finger proteins. This is of increasing importance since they constitute the largest transcription factor family in the human genome and have been implicated in cancer progression [285]. Collectively our study raises issues regarding detection of lncRNA using RNA-seq and provides a novel insight into a sex-specific role of XIST that warrants further investigation that can be of translational value in terms of the development of biomarkers and new therapeutic approaches targeting XIST and/or components of the signalling pathways it regulates in NSCLC.

Chapter 6

General Discussion

6.1 Impact and importance

Lung cancer is the third most common cancer in the UK and every year there are around 47,800 cases (130 every day). Lung cancer is the second most common cancer that affects both males and females. Every year around 35,300 people die of lung cancer (97 every day). This is due to patients being diagnosed at a late stage. In 2014 three-quarters of the cases for lung cancer were diagnosed at a late stage and only 19% of people survive when diagnosed this late. Over the past 40 years, no improvement has been shown in the survival of lung cancer in the UK because of limited therapeutic options available, lack of early diagnosis and no real-time monitoring of cancer. Therefore, an urge for new diagnosis methods and treatment is needed. Patients who are diagnosed early have an 88% chance of surviving a year or more (CRUK, 2019). The general options for treatment available for primary treatment are removal of the tumour, radiotherapy or chemotherapy. Although there are many difficulties clinicians have to overcome to give effective treatment. The problem with removing the whole tumour is that they must make sure everything is removed and not even one cancer cell is left otherwise there will be a relapse. Besides the whole tumour can only be removed if it has not metastasised. On the other hand, radiotherapy and chemotherapy are effective ways to control ad minimise cancer however some patients may not react well to it. The best possible way forward is to create a more personalised treatment. Over time, cancers adapt and evolve and become resistant to drugs due to evolutionary pressure by treatment therefore new treatments need to be discovered. An important aspect of developing new therapies is the discovery of new biomarkers. They might help provide

information about the stage of the disease and how it is evolving. Liquid biomarkers such as circulating tumour cells in the blood give real-time measurements of the changes in cancer biology especially during drug therapies.

In our study, we show that we can identify, characterise, and quantify CTCs in the blood of lung cancer patients, using the Imagestream technology. We also showed that CTCs were found higher in numbers in lung cancer patients compared to control samples, and in patients underwent thoracotomy vs. VATS. A potential prognostic value was observed in advanced stages of lung cancer patients compared to early stage using ctDNA analysis with our findings. Evaluation of ctDNA was conducted by calculating the DNA Integrity Index ([ALU247]/[ALU115]), from measurements of ALU repeats concentration in the plasma of patients. We observed a significantly increased DNA integrity index compared to controls. Diagnostic utility evaluation however revealed poor sensitivity and specificity values for pre-operative patients but high sensitivity and specificity for post-operative plasma samples. DNA integrity index and the cytokines levels before and after the surgery, provide more details for a more holistic approach of liquid biopsies.

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. Silencing XIST with siRNA knockdown techniques, in two lung cancer cell lines revealed a list of DEGs that have been affected and that they play a significant role in many biological processes according to the bio-informatic analysis we performed.

6.2 Implementation of liquid biopsies in to the clinical setting

Liquid biopsies allow us to examine the real-time disease status by the use of tumour-specific biomarkers in the blood. These biomarkers are essential when deciding the treatment and to evaluate response for targeted therapies [286]. The most studied type of liquid biopsy involves circulating tumour cells. Circulating tumour cells are

nucleated cells that lack hematopoietic differentiation. Another way is by detecting and quantifying cell-free DNA from apoptotic and dividing cancer cells [286]. However, numerous factors need to be considered if the technique or assay is implemented into clinical practice in the NHS. One of the major factors is cost. According to CRUK treatment for stage 4 lung cancer costs £13,078 and the cost for stage 1 lung cancer is much less at £7,952. In the long run, the benefits made in the poor survival rate will outweigh the costs. Predicting and monitoring treatment will help identify the best drugs to use depending on the genomic profile of the patient. For future aspirations, a universal assay that can be used for multiple cancer types would be advantageous.

6.3 Limitations of the research

One of the biggest limiting factors of any study is inter-patient variability. All the parameters tested in this research (CTCs, ccfDNA, mRNA expression) indicated that each individual even at the same tumour stage, has each own numbers. While measuring the CTCs in blood of lung cancer patients, we observed a big inter-patient variability. The reason why we experience such differences from patient to patient and the exact mechanism of CTC shedding in the bloodstream in not well understood. One thing that we can be sure about is that the larger number of CTCs travelling in the blood, increase the genomic material that is present. The body, initiates specific processes to discard free circulating epithelial cells, but when these cells undergo EMT, it is hard for the leukocytes to detect them and detain them. Apart from the epithelial to mesenchymal transition, the circulating tumour cells, may start expressing specific surface markers that attract platelets so as they create a layer that covers them and allows them to travel in 'stealth' mode.

In addition, differences may be appear in the molecular level of these samples. Transcriptomic, proteomic and metabolomic analyses have shown a variety in genetic profile from patient to patient and even on the same patient before and after chemotherapy, radiation or surgery.

Another significant issue with the CTCs research is the non-specific staining. Although many proof-of-principle experiments have been conducted, yet some cells appear to be positive for a variety of antibodies, an indication on non specific binding. This ability of some some to be positive in both epithelial and hematopoietic

markers has been the main challenge in the enrichment and isolation of CTCs [287]. In this research 'double positive' cells have been excluded from the counting since the origin and the nature of such cells remains unclear. Some CTCs that travel in the circulation may not express mesenchymal markers and maybe the CD45+ cells are attracted and potentially kill them. During this process, the CD45 and the epithelial markers may be present [288, 289]. In addition to the classic EMT markers, more markers should be used to identify cells not of epithelial or mesenchymal origin but endothelial cells (CECs) that could provide valuable information for the disease [290]. Increasing number of CECs could be an indicator of vascular dysfunction and damage [290].

A series of experiments and optimisation steps in the staining process took place for this study, and there was no possibility to get a 'clean' of double positive cells result. Using the ImageStream^{TM} technology, we managed to look into each individual cells' morphology and determine whether we can count it as a positive for a marker or not. Cell size and the aspect ratio and the focus of the cells were the key factors for the clearance of the results. After optimisations in the histograms and gating of the cells, we managed to have a universal analysis template to not be biased on the results. It has been shown that cell sizes in breast cancer changes after chemotherapy treatments which makes the characterisation of a positive CTC even more difficult. Due to variety of cell sizes in the various cancers or even in the same patient, staining with multiple antibodies is preferred from the size analysis alone.

If technologies like ImageStream^{TM} are going to be used in clinical use, then optimised templates to detect and gate the positive cells have to be made. So far, it is up to each individual scientist and his/her empirical opinion to determine the positiveness of a cell. In order for this technology to be part of the everyday clinic and monitoring of the patients, artificial intelligence and automated tools need to be used instead of each individual researcher. Not only they will offer an unbiased analysis but also they can go through hundreds of images that would take a lot of time to be processed by researchers.

6.4 Future directions

6.4.1 Large patient cohorts

There are around 47.800 new lung cancer cases in the UK every year with 38% of them are undergoing surgery, so 18.164 patients. According to sample size calculator for a study like this, and a margin of error at 10% we would need 96 patients. The number of samples in this project did not exceed 60 for any given type (cancer or control) because of the limitations in the recruitment of patients and healthy individuals. The different numbers of samples were used for the various assays because of the type of samples required, time of the experimentation and availability of samples. In regards to RNAseq, the sample size was limited due to the cost constrains. Furthermore, for future experiments, a larger cohort of samples for cancer and control should be considered for better confirmation of results especially considering the different sub-types and grading of cancer. Also when analysing the variations of the results in gender and ethnicity should be accountable as they affect gene expression [291]. It has been suggested that genetic variations due to ethnicity have been associated with drug susceptibility and treatment response [292].

6.4.2 Detection and characterisation of CTCs and CECs

In this study, 4 antibodies were used that are being used in histopathology labs in daily basis. These 4 antibodies stain specific types of cells of epithelial origin but not endothelial cells or cells that are undergoing EMT. In future, more antibodies can be used like Vimentin (mesenchymal marker) or CD105 (CECs) marker. Using more antibodies that stain not only different types of cells but also different parts of a cell or in different stage of cellular cycle could provide important information for the progress of a treatment of the effect of a surgical approach. Vimentin has been shown to be associated with metastasis and is a predictor of poor progression in patients with advanced lung cancer [230, 293]. By using isolation methods and staining with FISH techniques, we can not only characterise the Vimentin+ cells but also identify the differences in their phenotype [230, 293].

6.4.3 Artificial Inteligence (AI) and liquid biopsies

Collecting and processing multi-modal cancer data namely clinical, radiology, pathology and molecular, as well as liquid biopsy data (RNA-seq, DNA-seq, circulating tumour cell imaging) is quite challenging and yet time critical. Data collected under differing methodologies, conditions (research or clinical), or even from separate countries cannot be easily compared and is exponentially more complicated to normalize than the typical mismatched column header. Platforms that rely on the data analyst, researcher, or clinician to conform their data to the system can be time consuming, leading to a lack of adoption. Overall this has led to the current data environment that is composed of sets of redundant data collection campaigns, each unable to communicate with the other without hours of meticulous normalization activity. This means that a wealth of data is unavailable to clinicians, researchers, and decision makers although multiple studies have shown that by integrating patient multi-modal data, more personalized phenotypes can be reached and thus, more precise diagnosis and treatment selection can be provided. In the future, the focus should be on collecting and processing real lung cancer data that is un-normalized. In other words, data that has not been aligned and synchronized will be entered into an AI system. Following construction of algorithms that will read these data, the platform can provide probable outcomes. These outcomes can then determine the best course of treatment.
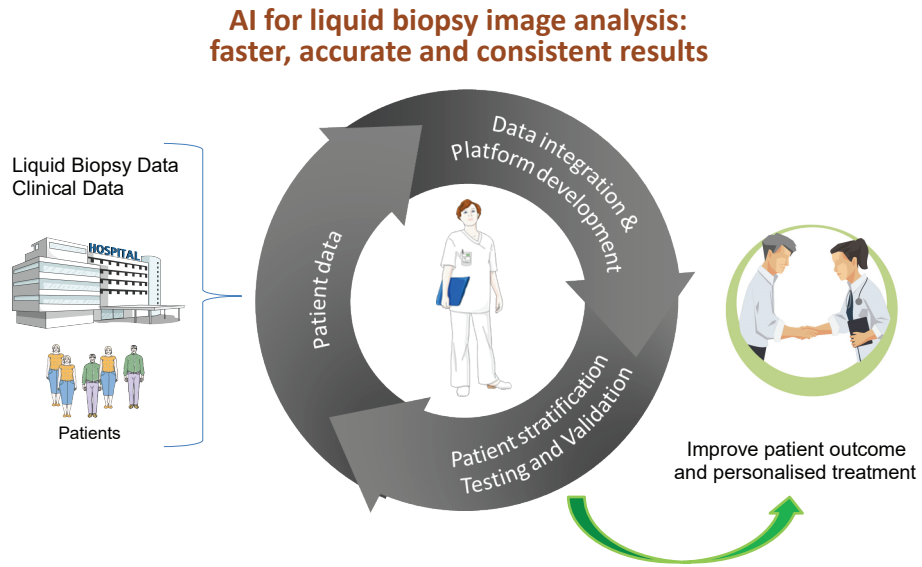
FIGURE 6.1: Machine and deep learning to harmonise multiple data types while developing predictive algorithms that will create a personalized treatment recommendation tool.

6.4.4   EGFR Mutations and cancer

Epidermal growth factor receptor (EGFR) is a membrane protein that plays a central role in transmitting signals that promote cell growth and proliferation. Its tyrosine kinase (TK) domain activates several downstream effectors that lead to activation of the Ras-Raf-MAPK pathway. Over-expression and oncogenic mutations that constitutively activate the TK domain of EGFR have been found in various solid tumours. Additionally, excessive activation of EGFR has been shown to be associated with advanced stages of cancer and a poor prognosis. Analysing EGFR mutations and monitoring the point of the kinase switch that a cell becomes dependent on another kinase, offers the ability for personalized therapy.

Entrogen EGFR Mutation Detection Kit

EntroGen's ctEGFR Mutation Detection Kit is a CE-IVD approved non-invasive, ultra-sensitive test to accurately detect the presence of EGFR activating mutations in ctDNA extracted from plasma. The test identifies the most common sensitising (exon 19 deletions and L858R) and resistance (T790M) mutations in EGFR. Furthermore, quantitative calibrators enable quantification of absolute copy numbers of each variant in unknown specimens, making the test ideal for monitoring response over the course of therapy. EntroGen's ctEGFR mutation analysis kit is a real-time polymerase chain reaction (PCR)-based assay that uses mutant-specific probes to identify the presence of EGFR mutations (Fig. 6.2).
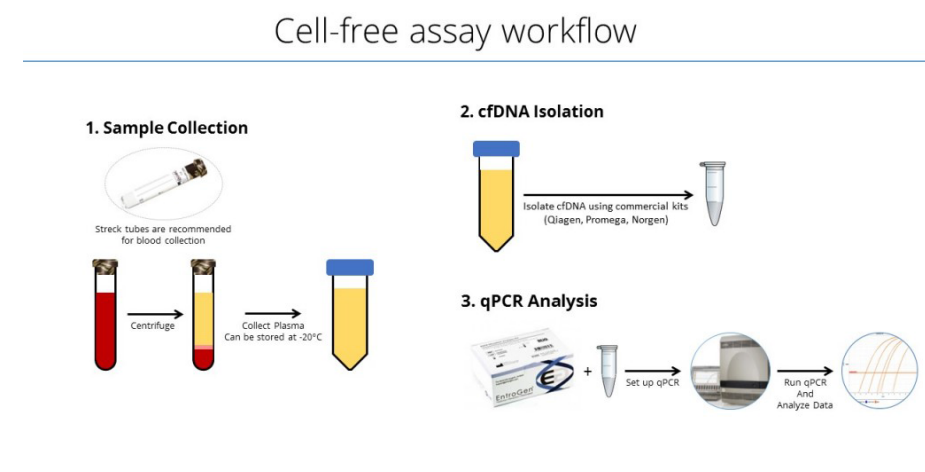
FIGURE 6.2: EntroGen cell-free assay workflow

6.4.5 Use of additional in vitro models to study the effects of XIST

To study the effects of XIST we used two adenocarcinoma cell lines for lung cancer to establish in vitro models. To expand the use of additional cell lines would be beneficial to reflect on the different stages and types of cancer. Moreover, 3D cultures of silenced XIST cells could provide better insight in the morphology of the cells after the knockdown. In figures 6.3, 6.4 and 6.5, some preliminary results of Immunofluorescence and scanning electron microscopy (SEM) experiments.
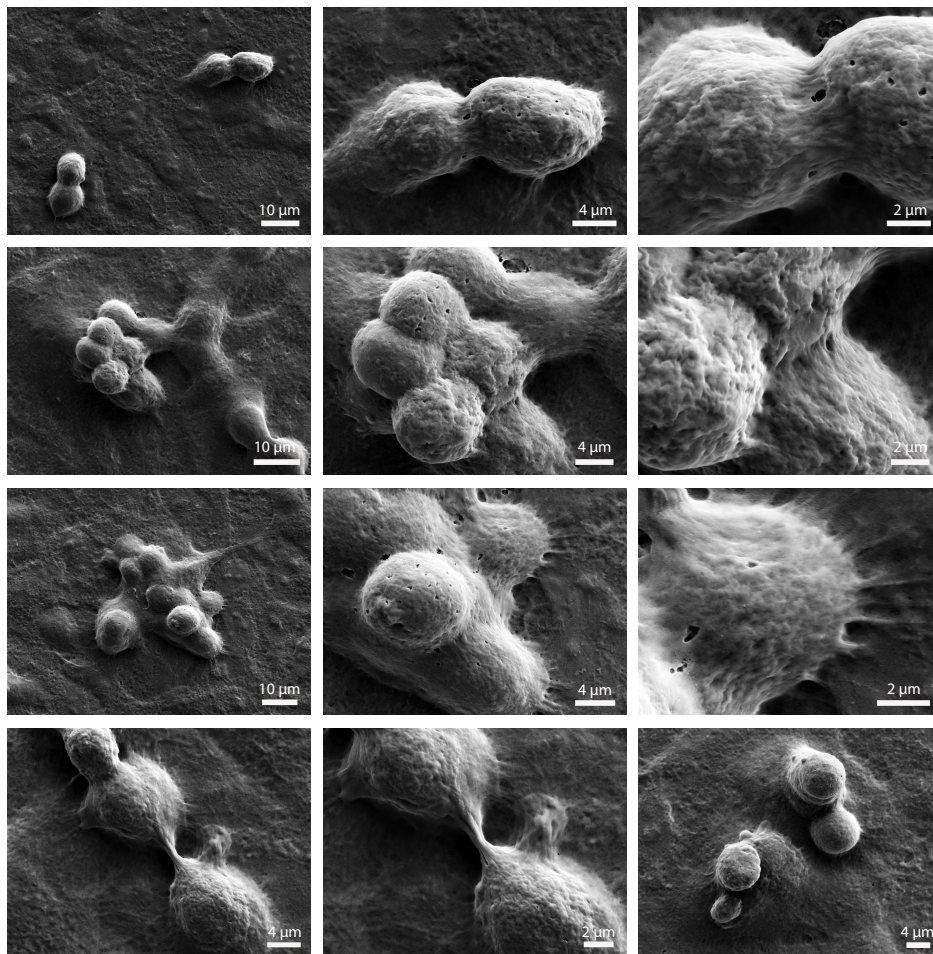
FIGURE 6.3: SEM in A549 cell line. Clusters of cells and cells undergoing mitosis can be seen.
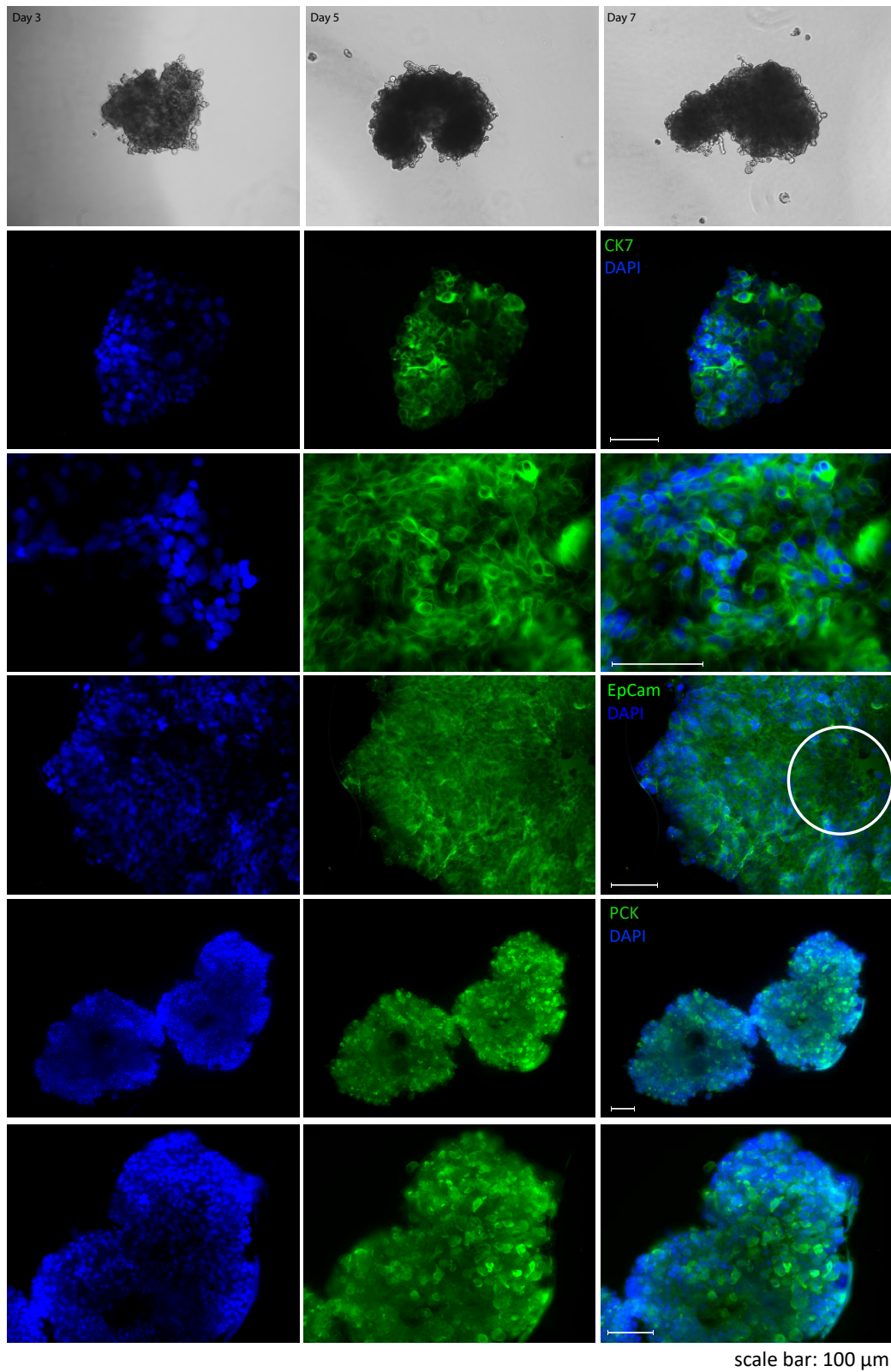
FIGURE 6.4: Immunofluorescence images of A549 spheroids stained with the three cytoplasm/membrane markers EpCAM, PCK and CK7

scale bar: 100 μm
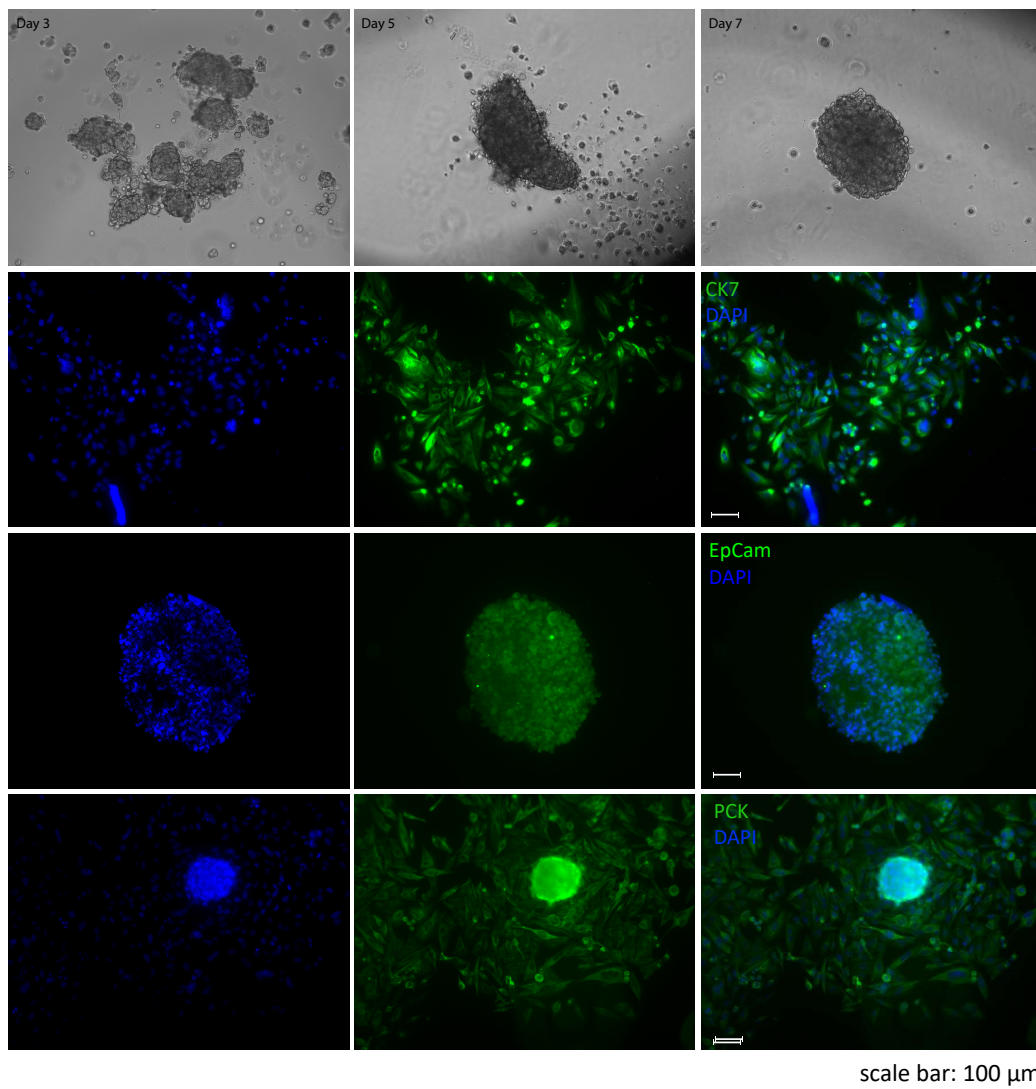
FIGURE 6.5: Immunofluorescence images of H1975 spheroids stained with the three cytoplasm/membrane markers EpCAM, PCK and CK7

6.5   Concluding remarks

Personalised medicine is the target for everyone in cancer diagnostics, treatment and management. Determination of detectable markers like CTCs and ccfDNA is has to be done and be combined with the daily clinical protocols. As presented in this projects, from a single blood collection tube we can extrapolate and measure a wide repertoire of markers without any need of a solid biopsy. Clinical practice can benefit from these non invasive techniques and evolve them even further with technologies like singe cell RNAseq and detection of key mutated genes. The combination of all those parameters will allow the clinicians to provide a more personalised targeted therapy designed to each individual patient.

Appendix A

Supplementary Information

A.1   Extended Figures

ATCC

Product Sheet

A549 (ATCC® CCL185 ™)

Please read this FIRST
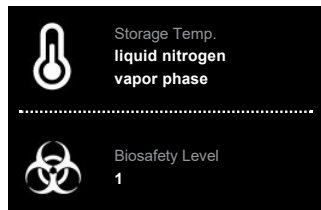
Storage Temp.
**liquid nitrogen
vapor phase**

Biosafety Level
1

Intended Use

This product is intended for research use only. It is not intended for any animal or human therapeutic or diagnostic use.

Complete Growth Medium

The base medium for this cell line is ATCCformulated F12K Medium, Catalog No. 302004.   To make the complete growth medium, add the following components to the base medium: fetal bovine serum to a final concentration of 10%.

Citation of Strain

If use of this culture results in a scientific publication, it should be cited in that manuscript in the following manner: A549 (ATCC® CCL185 ™)

American Type Culture Collection
PO Box 1549
Manassas, VA 20108 USA
www.atcc.org

800.638.6597 or 703.365.2700
Fax: 703.365.2750
Email: Tech@atcc.org

Or contact your local distributor

Page 1 of 3

Description

Organism: *Homo sapiens*, human
Tissue: lung
Disease: Carcinoma
Cell Type: epithelial
Age: 58 years
Gender: male
Morphology: epitheliallike
Growth Properties: adherent
Isoenzymes:
G6PD, B
DNA Profile:
Amelogenin: X,Y
CSF1PO: 10,12
D13S317: 11
D16S539: 11,12
D5S818: 11
D7S820: 8,11
THO1: 8,9.3
TPOX: 8,11
vWA: 14
Cytogenetic Analysis: This is a hypotriploid human cell line with the modal chromosome number of 66, occurring in 24% of cells. Cells with 64 (22%), 65, and 67 chromosome counts also occurred at relatively high frequencies; the rate with higher ploidies was low at 0.4%. There were 6 markers present in single copies in all cells. They include der(6)t(1;6) (q11;q27); ?del(6) (p23); del(11) (q21), del(2) (q11), M4 and M5. Most cells had two X and two Y chromosomes. However, one or both Y chromosomes were lost in 40% of 50 cells analyzed. Chromosomes N2 and N6 had single copies per cell; and N12 and N17 usually had 4 copies. Note: Cytogenetic information is based on initial seed stock at ATCC. Cytogenetic instability has been reported in the literature for some cell lines.

BatchSpecific Information

Refer to the Certificate of Analysis for batchspecific test results.

⚠ SAFETY PRECAUTION

ATCC highly recommends that protective gloves and clothing always be used and a full face mask always be worn when handling frozen vials. It is important to note that some vials leak when submersed in liquid nitrogen and will slowly fill with liquid nitrogen. Upon thawing, the conversion of the liquid nitrogen back to its gas phase may result in the vessel exploding or blowing off its cap with dangerous force creating flying debris.

Unpacking & Storage Instructions

1. Check all containers for leakage or breakage.
2. Remove the frozen cells from the dry ice packaging and immediately place the cells at a temperature below 130°C, preferably in liquid nitrogen vapor, until ready for use.

Handling Procedure for Frozen Cells

To insure the highest level of viability, thaw the vial and initiate the culture as soon as possible upon receipt. If upon arrival, continued storage of the frozen culture is necessary, it should be stored in liquid nitrogen vapor phase and not at 70°C. Storage at 70°C will result in loss of viability.

1. Thaw the vial by gentle agitation in a 37°C water bath. To reduce the possibility of contamination, keep the Oring and cap out of the water. Thawing should be rapid (approximately 2 minutes).
2. Remove the vial from the water bath as soon as the contents are thawed, and decontaminate by dipping in or spraying with 70% ethanol. All of the operations from this point on should be carried out under strict aseptic conditions.
3. Transfer the vial contents to a centrifuge tube containing 9.0 mL complete culture medium. and spin at approximately 125 x *g* for 5 to 7 minutes.
4. Resuspend cell pellet with the recommended complete medium (see the specific batch information for the culture recommended dilution ratio). It is important to avoid excessive alkalinity of the medium during recovery of the cells. It is suggested that, prior to the addition of the vial contents, the culture vessel containing the complete growth medium be placed into the incubator for at least 15 minutes to

ATCC

Product Sheet

NCIH1975 [H1975, H1975] (ATCC® CRL5908 ™)

Please read this FIRST

Storage Temp.
liquid nitrogen vapor phase

Biosafety Level
1

Intended Use

This product is intended for research use only. It is not intended for any animal or human therapeutic or diagnostic use.

Complete Growth Medium

The base medium for this cell line is ATCCformulated RPMI1640 Medium, ATCC 302001. To make the complete growth medium, add the following components to the base medium: fetal bovine serum (ATCC 302020) to a final concentration of 10%.

Citation of Strain

If use of this culture results in a scientific publication, it should be cited in that manuscript in the following manner: NCIH1975 [H1975, H1975] (ATCC ® CRL 5908™)

Page 1 of 2

Description

Organism: *Homo sapiens*, human
Tissue: lung
Disease: adenocarcinoma; nonsmall cell lung cancer
Gender: female
Morphology: epithelial
Growth Properties: adherent
DNA Profile:
Amelogenin: X
CSF1PO: 12
D13S317: 10,13
D16S539: 9,12
D5S818: 11,12
D7S820: 8,11
THO1: 7
TPOX: 8,11
vWA: 18

BatchSpecific Information

Refer to the Certificate of Analysis for batchspecific test results.

⚠ SAFETY PRECAUTION

ATCC highly recommends that protective gloves and clothing always be used and a full face mask always be worn when handling frozen vials. It is important to note that some vials leak when submersed in liquid nitrogen and will slowly fill with liquid nitrogen. Upon thawing, the conversion of the liquid nitrogen back to its gas phase may result in the vessel exploding or blowing off its cap with dangerous force creating flying debris.

Unpacking & Storage Instructions

1. Check all containers for leakage or breakage.
2. Remove the frozen cells from the dry ice packaging and immediately place the cells at a temperature below 130°C, preferably in liquid nitrogen vapor, until ready for use.

Handling Procedure for Frozen Cells

To insure the highest level of viability, thaw the vial and initiate the culture as soon as possible upon receipt. If upon arrival, continued storage of the frozen culture is necessary, it should be stored in liquid nitrogen vapor phase and not at 70°C. Storage at 70°C will result in loss of viability.

1. Thaw the vial by gentle agitation in a 37°C water bath. To reduce the possibility of contamination, keep the Oring and cap out of the water. Thawing should be rapid (approximately 2 minutes).
2. Remove the vial from the water bath as soon as the contents are thawed, and decontaminate by dipping in or spraying with 70% ethanol. All of the operations from this point on should be carried out under strict aseptic conditions.
3. Transfer the vial contents to a centrifuge tube containing 9.0 mL complete culture medium. and spin at approximately 125 x g for 5 to 7 minutes.
4. Resuspend cell pellet with the recommended complete medium (see the specific batch information for the culture recommended dilution ratio). and dispense into a 25 cm² or a 75 cm² culture flask. It is important to avoid excessive alkalinity of the medium during recovery of the cells. It is suggested that, prior to the addition of the vial contents, the culture vessel containing the complete growth medium be placed into the incubator for at least 15 minutes to allow the medium to reach its normal pH (7.0 to 7.6).
5. Incubate the culture at 37°C in a suitable incubator. A 5% CO_2 in air atmosphere is recommended if using the medium described on this product sheet.

Handling Procedure for Flask Cultures

The flask was seeded with cells (see specific batch information) grown and completely filled with medium at ATCC to prevent loss of cells during shipping.

1. Upon receipt visually examine the culture for macroscopic evidence of any microbial contamination. Using an inverted microscope (preferably equipped with phasecontrast optics), carefully check for any evidence of microbial contamination. Also check to determine if the majority of cells are still attached to the bottom of the flask; during shipping the cultures are sometimes handled roughly and

ATCC

Product Sheet

MSTO211H (ATCC® CRL 2081™)

Please read this FIRST
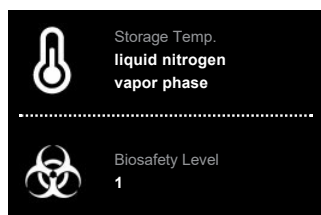
Storage Temp.
liquid nitrogen vapor phase

Biosafety Level
1

Intended Use

This product is intended for research use only. It is not intended for any animal or human therapeutic or diagnostic use.

Complete Growth Medium

The base medium for this cell line is ATCCformulated RPMI1640 Medium, ATCC 302001 . To make the complete growth medium, add the following components to the base medium: fetal bovine serum (ATCC 302020) to a final concentration of 10%.

Citation of Strain

If use of this culture results in a scientific publication, it should be cited in that manuscript in the following manner: MSTO211H (ATCC® CRL2081 ™)

American Type Culture Collection
PO Box 1549
Manassas, VA 20108 USA
www.atcc.org

800.638.6597 or 703.365.2700
Fax: 703.365.2750
Email: Tech@atcc.org

Or contact your local distributor

Page 1 of 3

Description

Organism: *Homo sapiens*, human
Tissue: lung; derived from metastatic site: pleural effusion
Disease: biphasic mesothelioma
Age: 62 years
Gender: male
Morphology: fibroblast
Growth Properties: adherent
DNA Profile:
Amelogenin: X,Y
CSF1PO: 11,12
D13S317: 11,14
D16S539: 13
D5S818: 12
D7S820: 8,12
TH01: 8,9.3
TPOX: 11
vWA: 16,18
Cytogenetic Analysis: modal number = 72; range = 70 to 78

BatchSpecific Information

Refer to the Certificate of Analysis for batchspecific test results.

SAFETY PRECAUTION

ATCC highly recommends that protective gloves and clothing always be used and a full face mask always be worn when handling frozen vials. It is important to note that some vials leak when submersed in liquid nitrogen and will slowly fill with liquid nitrogen. Upon thawing, the conversion of the liquid nitrogen back to its gas phase may result in the vessel exploding or blowing off its cap with dangerous force creating flying debris.

Unpacking & Storage Instructions

1. Check all containers for leakage or breakage.
2. Remove the frozen cells from the dry ice packaging and immediately place the cells at a temperature below 130°C, preferably in liquid nitrogen vapor, until ready for use.

Handling Procedure for Frozen Cells

To insure the highest level of viability, thaw the vial and initiate the culture as soon as possible upon receipt. If upon arrival, continued storage of the frozen culture is necessary, it should be stored in liquid nitrogen vapor phase and not at 70°C. Storage at 70°C will result in loss of viability.

1. Thaw the vial by gentle agitation in a **37°C** water bath. To reduce the possibility of contamination, keep the Oring and cap out of the water. Thawing should be rapid (approximately 2 minutes).
2. Remove the vial from the water bath as soon as the contents are thawed, and decontaminate by dipping in or spraying with 70% ethanol. All of the operations from this point on should be carried out under strict aseptic conditions.
3. Transfer the vial contents to a 75 cm² tissue culture flask and dilute with the recommended complete culture medium (see the specific batch information for the recommended dilution ratio). It is important to avoid excessive alkalinity of the medium during recovery of the cells. It is suggested that, prior to the addition of the vial contents, the culture vessel containing the growth medium be placed into the incubator for at least 15 minutes to allow the medium to reach its normal pH (7.0 to 7.6).
4. Incubate the culture at **37°C** in a suitable incubator. A 5% CO_2 in air atmosphere is recommended if using the medium described on this product sheet.

If it is desired that the cryoprotective agent be removed immediately, or that a more concentrated cell suspension be obtained, centrifuge the cell suspension at approximately 125 xg for 5 to 10 minutes. Discard the supernatant and resuspend the cells with fresh growth medium at the dilution ratio recommended in the specific batch information.

Handling Procedure for Flask Cultures

The flask was seeded with cells (see specific batch information) grown and completely filled with medium at ATCC to prevent loss of cells during shipping.

A.2 Extended Tables

No	Age	Sex	Tumour Type	Operation	TNM	PL	Stage
1	75	M	SqCC	Thoracotomy	T2aN0M0	PL0	IA
2	59	M	AC	Thoracotomy - Lobectomy	T4N0M0	PL3	IA
3	59	F	AC	VATS - Lobectomy	T1N0M0	PL0	IA
4	84	M	AC	Thoracotomy - Lobectomy	T1aN0M0	PL0	IA
5	71	F	AC	Thoracotomy - Lobectomy	T1bN0M0	PL0	IA
6	57	F	AC	VATS - Lobectomy	T1aN0M0	PL0	IA
7	71	M	AC	VATS - Lobectomy	T1cN0M0	PL0	IA
8	53	F	SqCC	Thoracotomy - Lobectomy	T1bN0M0	PL0	IA
9	59	M	AC	VATS - Lobectomy	T1bN2M0	PL0	IIIA
10	40	F	AC	VATS - Lobectomy	T1N0M0	PL0	IA
11	70	F	SqCC	VATS - Lobectomy	T1bN0M0	PL0	IA
12	74	M	AC	VATS - Lobectomy	T1bN2M0	PL0	IIIA
13	71	F	AC	VATS - Lobectomy	T2aN0M0	PL0	IB
14	77	F	AC	Thoracotomy - Lobectomy	T4N1M0	Pl1	IIIA
15	71	M	AC	VATS - Lobectomy	T2bN0M0	PL0	IIA
16	66	F	SqCC	VATS - Lobectomy	T1bN0M0	PL0	IA
17	80	F	AC	VATS - Lobectomy	T1bN0M0	PL0	IA
18	80	M	AC	VATS - Lobectomy	T3N1M0	PL0	IIIA
19	77	F	SqCC	VATS - Lobectomy	T3N0M0	PL1	IIB
20	36	F	AC	VATS - Lobectomy	T1bN0M0	PL0	IA
21	69	M	AC	Thoracotomy - Lobectomy	T3N0M0	PL0	IB
22	62	M	AC	Thoracotomy - Lobectomy	T3N0M0	PL1	IB
23	83	M	SqCC	horacotomy - Lobectomy	T2N0M1a	PL2	IV
24	71	F	SqCC	Thoracotomy - Lobectomy	T2aN0M0	PL0	IB
25	71	F	AC	VATS - Lobectomy	pT2N1bM0	PL1	IIB
26	73	F	AC	VATS - Lobectomy	pT1cN0M0	PL0	IA
27	57	F	VA	VATS - Lobectomy	n/a	n/a	
28	67	F	AC	VATS - Lobectomy	T2bN0M0	PL0	IIA
29	61	M	AC	VATS - Lobectomy	pT1cN0M0	PL0	IA
30	68	F	AC	Thoracotomy - Lobectomy	pT1cN0M0	PL0	IA
31	51	F	AC	VATS - Lobectomy	M1	PL0	IV

32	79	F	AC	VATS - Lobectomy	n/a	n/a	n/a
33	66	F	AC	VATS - Lobectomy	pT1cN0M0	PL0	IA
34	63	F	AC	VATS - Lobectomy	pT1cN0M0	PL0	IA
35	78	M	SqCC	VATS - Lobectomy	T3N0M0	PL0	IIB
36	68	F	SqCC	VATS - Lobectomy	pT1bN0M0	PL0	IA
37	57	F	AC	VATS - Lobectomy	pT1bN0M0	PL0	IA
38	57	F	AC	VATS - Lobectomy	T1cN0M0	PL0	IA
39	79	M	AC	VATS - Lobectomy	T3N0M0	PL0	IIB
40	61	F	AC	Thoracotomy - Lobectomy	T3N2bM0	PL0	IIIA
41	68	F	SqCC	VATS - Lobectomy	T3N0M0	PL1	IIB
42	74	M	SqCC	VATS - Lobectomy	T1N0M0	PL0	IA
43	81	M	AC	VATS - Lobectomy	T2aN0M0	PL0	IB
44	77	M	AC	VATS - Lobectomy	T1N0M0	PL0	IA
45	55	M	AC	VATS - Lobectomy	T3N0M0	PL0	IIB
46	62	F	CA	VATS - Lobectomy	T1N0M0	PL0	IA
47	61	M	AC	VATS - Lobectomy	T1N0M0	PL0	IA
49	74	M	SqCC	VATS - Lobectomy	T2bN0M0	PL1	IB
50	68	F	AC	VATS - Lobectomy	T2aN0M0	PL1	IB
51	68	F	AC	VATS - Lobectomy	T1aN0M0	PL0	IA
52	68	M	SqCC	VATS - Lobectomy	T2aN0M0	PL0	IB
53	68	M	AC	VATS - Lobectomy	pT1bNxM0	PL0	IA
54	69	F	AC	VATS - Lobectomy	T1bN0M0	PL0	IA
55	71	M	AC	VATS - Lobectomy	T4N0M0	PL0	IIIA
56	70	M	SqCC	Thoracotomy - Lobectomy	T1bN1M0	PL0	IIB
57	80	F	SqCC	VATS - Lobectomy	T1cN0M0	PL0	IA

TABLE A.1: Detailed information about the patients of the study. Patients 32, 43, 48 were not included in the study due to unavailability of post-operative blood collection or other issue.

Gene ID	Gene Symbol	Gene Name
ENSG00000023839.10	ABCC2	ATP binding cassette C2
ENSG00000108846.15	ABCC3	ATP binding cassette C3
ENSG00000006071.12	ABCC8	ATP binding cassette subfamily C8
ENSG00000160179.18	ABCG1	ATP binding cassette G1
ENSG00000114948.12	ADAM23	ADAM metallopeptidase domain 23
ENSG00000164742.14	ADCY1	adenylate cyclase 1
ENSG00000075340.22	ADD2	adducin 2
ENSG00000117114.19	ADGRL2	adhesion G protein-coupled receptor L2
ENSG00000248144.5	ADH1C	alcohol dehydrogenase 1C (class I), gamma polypeptide
ENSG00000154027.18	AK5	adenylate kinase 5
ENSG00000198074.9	AKR1B10	aldo-keto reductase 1-B10
ENSG00000196139.13	AKR1C3	aldo-keto reductase 1-C3
ENSG00000180318.3	ALX1	ALX homeobox 1
ENSG00000139211.6	AMIGO2	adhesion molecule with Ig like domain 2
ENSG00000151150.21	ANK3	ankyrin 3(ANK3)
ENSG00000171714.10	ANO5	anoctamin 5(ANO5)
ENSG00000169604.19	ANTXR1	anthrax toxin receptor 1
ENSG00000104537.16	ANXA13	annexin A13(ANXA13)
ENSG00000034053.14	APBA2	amyloid beta precursor protein A-2
ENSG00000163219.11	ARHGAP25	Rho GTPase activating protein 25
ENSG00000172995.16	ARPP21	cAMP regulated phosphoprotein 21
ENSG00000105409.16	ATP1A3	ATPase Na+/K+ transporting subunit a3
ENSG00000107518.16	ATRNL1	attractin like 1
ENSG00000231345.3	BEND3P1	BEN domain containing 3 pseudogene 1
ENSG00000166681.13	BEX3	brain expressed X-linked 3
ENSG00000102409.9	BEX4	brain expressed X-linked 4
ENSG00000122870.11	BICC1	BicC family RNA binding protein 1
ENSG00000151136.14	BTBD11	BTB domain containing 11
ENSG00000165810.16	BTNL9	butyrophilin like 9
ENSG00000148655.14	C10orf11	chromosome 10 open reading frame 11
ENSG00000100557.9	C14orf105	chromosome 14 open reading frame 105

ENSG00000123838.10	C4BPA	complement component 4 binding protein a
ENSG00000186493.11	C5orf38	chromosome 5 open reading frame 38
ENSG00000074410.13	CA12	carbonic anhydrase 12
ENSG00000178538.9	CA8	carbonic anhydrase 8
ENSG00000006283.17	CACNA1G	calcium voltage-gated channel subunit alpha1
ENSG00000105605.7	CACNG7	calcium voltage-gated channel auxiliary subunit gamma 7
ENSG00000247844.1	CCAT1	colon cancer associated transcript 1
ENSG00000184305.14	CCSER1	coiled-coil serine rich protein 1
ENSG00000113361.12	CDH6	cadherin 6
ENSG00000124762.13	CDKN1A	cyclin dependent kinase inhibitor 1A
ENSG00000086548.8	CEACAM6	carcinoembryonic antigen - cell adhesion molecule 6
ENSG00000116785.13	CFHR3	complement factor H related 3
ENSG00000134121.9	CHL1	cell adhesion molecule L1 like
ENSG00000278530.4	CHMP1B2P	charged multivesicular body protein 1B2
ENSG00000106069.21	CHN2	chimerin 2
ENSG00000175344.17	CHRNA7	cholinergic receptor nicotinic alpha 7 subunit
ENSG00000115526.10	CHST10	carbohydrate sulfotransferase 10
ENSG00000175040.5	CHST2	carbohydrate sulfotransferase 2
ENSG00000187288.10	CIDEC	cell death inducing DFFA like effector c
ENSG00000149970.14	CNKSR2	connector enhancer of kinase suppressor of Ras 2
ENSG00000119946.10	CNNM1	cyclin domain divalent metal cation transport mediator
ENSG00000118432.12	CNR1	cannabinoid receptor 1
ENSG00000119865.8	CNRIP1	cannabinoid interacting protein 1
ENSG00000154529.14	CNTNAP3B	contactin associated protein-like 3B
ENSG00000276386.1	CNTNAP3P2	contactin associated protein-like 3 pseudogene 2
ENSG00000060718.20	COL11A1	collagen type XI alpha 1 chain
ENSG00000145920.14	CPLX2	complexin 2

ENSG00000121005.8	CRISPLD1	cysteine rich secretory protein LCCL-1
ENSG00000170373.8	CST1	cystatin SN
ENSG00000169862.18	CTNND2	catenin delta 2
ENSG00000077063.10	CTTNBP2	cortactin binding protein 2
ENSG00000055163.19	CYFIP2	cytoplasmic FMR1 interacting protein 2
ENSG00000019186.9	CYP24A1	cytochrome P450 family 24 subfamily A member 1
ENSG00000067048.16	DDX3Y	DEAD-box helicase 3, Y-linked
ENSG00000162496.8	DHRS3	dehydrogenase/reductase 3
ENSG00000050165.17	DKK3	dickkopf WNT pathway inhibitor 3
ENSG00000134762.16	DSC3	desmocollin 3(DSC3)
ENSG00000164330.16	EBF1	early B-cell factor 1
ENSG00000198692.9	EIF1AY	eukaryotic translation factor 1A
ENSG00000155849.15	ELMO1	engulfment and cell motility
ENSG00000118402.5	ELOVL4	ELOVL fatty acid elongase 4
ENSG00000196482.16	ESRRG	estrogen related receptor gamma
ENSG00000173040.12	EVC2	EvC ciliary complex subunit 2
ENSG00000169122.11	FAM110B	family with sequence similarity 110-B
ENSG00000184731.5	FAM110C	family with sequence similarity 110-C
ENSG00000109794.13	FAM149A	family with sequence similarity 149-A
ENSG00000182230.11	FAM153B	family with sequence similarity 153-B
ENSG00000130054.4	FAM155B	family with sequence similarity 155-B
ENSG00000219438.8	FAM19A5	family with sequence similarity 19-A5
ENSG00000197872.11	FAM49A	family with sequence similarity 49-A
ENSG00000162981.13	FAM84A	family with sequence similarity 84-A
ENSG00000178162.8	FAR2P2	fatty acyl-CoA reductase 2 pseudogene
ENSG00000152767.16	FARP1	FERM and pleckstrin domain protein 1
ENSG00000138829.11	FBN2	fibrillin 2
ENSG00000182511.11	FES	FES proto-oncogene, tyrosine kinase
ENSG00000171560.14	FGA	fibrinogen alpha chain
ENSG00000171564.11	FGB	fibrinogen beta chain
ENSG00000037280.15	FLT4	fms related tyrosine kinase 4
ENSG00000137273.3	FOXF2	forkhead box F2
ENSG00000226608.3	FTLP3	ferritin light chain pseudogene 3
ENSG00000137731.13	FXYD2	FXYD domain-ion transport regulator 2

ENSG00000137726.16	FXYD6	FXYD domain-ion transport regulator 6
ENSG00000144278.14	GALNT13	N-acetylgalactosaminyl transferase 13
ENSG00000158089.14	GALNT14	N-acetylgalactosaminyl transferase 14
ENSG00000119125.16	GDA	guanine deaminase
ENSG00000151892.14	GFRA1	GDNF family receptor alpha 1
ENSG00000100031.18	GGT1	gamma-glutamyltransferase 1
ENSG00000178445.9	GLDC	glycine decarboxylase
ENSG00000168243.10	GNG4	G protein subunit gamma 4
ENSG00000185477.4	GPRIN3	GPRIN family member 3
ENSG00000273079.5	GRIN2B	glutamate ionotropic receptor NMDA type - 2B
ENSG00000189369.8	GSPT2	G1 to S phase transition 2
ENSG00000113249.12	HAVCR1	hepatitis A virus cellular receptor 1
ENSG00000113924.11	HGD	homogentisate 1,2-dioxygenase
ENSG00000019991.15	HGF	hepatocyte growth factor
ENSG00000100292.16	HMOX1	heme oxygenase 1
ENSG00000275410.4	HNF1B	HNF1 homeobox B
ENSG00000159184.7	HOXB13	homeobox B13
ENSG00000164120.13	HPGD	hydroxyprostaglandin dehydrogenase
ENSG00000142149.8	HUNK	Neu-associated kinase
ENSG00000146678.9	IGFBP1	insulin binding protein 1
ENSG00000141753.6	IGFBP4	insulin binding protein 4
ENSG00000144847.12	IGSF11	immunoglobulin superfamily -11
ENSG00000115232.13	ITGA4	integrin subunit alpha 4
ENSG00000151655.18	ITIH2	inter-alpha-trypsin inhibitor heavy chain 2
ENSG00000123243.14	ITIH5	inter-alpha-trypsin inhibitor heavy chain 5
ENSG00000078596.10	ITM2A	integral membrane protein 2A
ENSG00000162975.4	KCNF1	potassium voltage-gated channel modifier F1
ENSG00000123700.4	KCNJ2	potassium voltage-gated channel J2
ENSG00000156113.22	KCNMA1	potassium calcium channel M1
ENSG00000162687.16	KCNT2	potassium sodium-activated channel T2
ENSG00000164794.8	KCNV1	potassium voltage-gated channel V1

ENSG00000012817.15	KDM5D	lysine demethylase 5D
ENSG00000167751.12	KLK2	kallikrein related peptidase 2
ENSG00000171431.3	KRT20	keratin 20
ENSG00000162624.14	LHX8	LIM homeobox 8
ENSG00000215808.2	LINC01139	long intergenic ncRNA 1139
ENSG00000258867.5	LINC01146	long intergenic ncRNANA 1146
ENSG00000231131.6	LINC01468	long intergenic ncRNA 1468
ENSG00000169783.12	LINGO1	leucine rich repeat and Ig domain 1
ENSG00000174482.10	LINGO2	leucine rich repeat and Ig domain 2
ENSG00000188906.14	LRRK2	leucine rich repeat kinase 2
ENSG00000197172.10	MAGEA6	MAGE family member A6
ENSG00000108984.13	MAP2K6	mitogen-activated protein kinase 6
ENSG00000139915.18	MDGA2	MAM containing glycosylphosphatidylinositol anchor 2
ENSG00000214548.14	MEG3	maternally expressed 3 (non-protein coding)
ENSG00000155970.11	MICU3	mitochondrial calcium uptake family member 3
ENSG00000215386.11	MIR99AHG	mir-99a-let-7c cluster host gene
ENSG00000107186.16	MPDZ	PDZ domain crumbs cell polarity complex
ENSG00000135324.5	MRAP2	melanocortin 2 receptor accessory protein
ENSG00000215182.8	MUC5AC	mucin 5AC, oligomeric mucus-forming
ENSG00000266714.6	MYO15B	myosin XVB(MYO15B)
ENSG00000186462.8	NAP1L2	nucleosome assembly protein 1 like
ENSG00000104490.17	NCALD	neurocalcin delta
ENSG00000176771.16	NCKAP5	NCK associated protein 5
ENSG00000277586.2	NEFL	neurofilament, light polypeptide
ENSG00000163531.15	NFASC	neurofascin
ENSG00000183072.9	NKX2-5	NK2 homeobox 5
ENSG00000022556.15	NLRP2	NLR family pyrin domain containing 2
ENSG00000171246.5	NPTX1	neuronal pentraxin 1
ENSG00000164128.6	NPY1R	neuropeptide Y receptor Y1
ENSG00000169297.7	NR0B1	nuclear receptor subfamily 0 B-1

ENSG00000091129.19	NRCAM	neuronal cell adhesion molecule
ENSG00000140538.16	NTRK3	neurotrophic receptor tyrosine kinase 3
ENSG00000196368.4	NUDT11	nudix hydrolase 11
ENSG00000169860.6	P2RY1	purinergic receptor P2Y1
ENSG00000077264.14	PAK3	p21 (RAC1) activated kinase 3
ENSG00000182752.9	PAPPA	pappalysin 1
ENSG00000009709.11	PAX7	paired box 7
ENSG00000138650.8	PCDH10	protocadherin 10
ENSG00000140479.16	PCSK6	proprotein convertase subtilisin/kexin type 6
ENSG00000169174.10	PCSK9	proprotein convertase subtilisin/kexin type 9
ENSG00000184588.17	PDE4B	phosphodiesterase 4B
ENSG00000142657.20	PGD	phosphogluconate dehydrogenase
ENSG00000164093.15	PITX2	paired like homeodomain 2
ENSG00000144837.8	PLA1A	phospholipase A1 member A
ENSG00000118495.18	PLAGL1	PLAG1 like zinc finger 1
ENSG00000154822.17	PLCL2	phospholipase C like 2
ENSG00000180287.16	PLD5	phospholipase D family member 5
ENSG00000148123.14	PLPPR1	phospholipid phosphatase related 1
ENSG00000182013.17	PNMAL1	paraneoplastic Ma antigen family like 1
ENSG00000124429.17	POF1B	premature ovarian failure, 1B
ENSG00000156475.18	PPP2R2B	protein phosphatase 2 regulatory subunit Bbeta
ENSG00000138738.10	PRDM5	PR/SET domain 5(PRDM5)
ENSG00000065675.14	PRKCQ	protein kinase C theta
ENSG00000237943.6	PRKCQ-AS1	PRKCQ antisense RNA 1
ENSG00000099725.14	PRKY	protein kinase, Y-linked, pseudogene
ENSG00000250799.9	PRODH2	proline dehydrogenase 2
ENSG00000204179.10	PTPN20	protein tyrosine phosphatase, non-receptor 20
ENSG00000153707.16	PTPRD	protein tyrosine phosphatase, receptor D
ENSG00000132554.19	RGS22	regulator of G-protein signaling 22
ENSG00000168421.12	RHOH	ras homolog family member H

ENSG00000177181.14	RIMKLA	ribosomal modification protein rimK like
ENSG00000133135.13	RNF128	ring finger protein 128, E3 ub protein
ENSG00000170153.10	RNF150	ring finger protein 150
ENSG00000089169.14	RPH3A	rabphilin 3A
ENSG00000129824.15	RPS4Y1	ribosomal protein S4, Y-linked 1
ENSG00000146374.13	RSPO3	R-spondin 3(RSPO3)
ENSG00000181433.9	SAGE1	sarcoma antigen 1(SAGE1)
ENSG00000103449.11	SALL1	spalt like transcription factor 1
ENSG00000203727.3	SAMD5	sterile alpha motif domain containing 5
ENSG00000170099.5	SERPINA6	serpin family A member 6
ENSG00000135919.12	SERPINE2	serpin family E member 2
ENSG00000104332.11	SFRP1	secreted frizzled related protein 1
ENSG00000170624.13	SGCD	sarcoglycan delta
ENSG00000105738.10	SIPA1L3	signal induced proliferation associated 1 like 3
ENSG00000118596.11	SLC16A7	solute carrier family 16 member 7
ENSG00000124564.17	SLC17A3	solute carrier family 17 member 3
ENSG00000092096.16	SLC22A17	solute carrier family 22 member 17
ENSG00000142494.13	SLC47A1	solute carrier family 47 member 1
ENSG00000072041.16	SLC6A15	solute carrier family 6 member 15
ENSG00000151012.13	SLC7A11	solute carrier family 7 member 11
ENSG00000155465.18	SLC7A7	solute carrier family 7 member
ENSG00000227028.6	SLC8A1-AS1	SLC8A1 antisense RNA 1
ENSG00000084453.16	SLCO1A2	solute carrier organic anion transporter 1A2
ENSG00000111700.12	SLCO1B3	solute carrier organic anion transporter 1B3
ENSG00000173930.8	SLCO4C1	solute carrier organic anion transporter 4C1
ENSG00000080503.22	SMARCA2	SWI/SNF related,actin dependent regulator of chr
ENSG00000128602.9	SMO	smoothened, frizzled class receptor
ENSG00000065609.14	SNAP91	synaptosome associated protein 91
ENSG00000225746.9	SNHG23	small nucleolar RNA host gene 23

ENSG00000163071.10	SPATA18	spermatogenesis associated 18
ENSG00000167642.12	SPINT2	serine peptidase inhibitor, Kunitz type 2
ENSG00000196104.10	SPOCK3	SPARC/osteonectin, proteoglycan 3
ENSG00000118785.13	SPP1	secreted phosphoprotein 1
ENSG00000197122.11	SRC	SRC proto-oncogene, non-receptor tyrosine kinase
ENSG00000115525.17	ST3GAL5	ST3 beta-galactoside alpha-2,3-sialyltransferase 5
ENSG00000144057.15	ST6GAL2	ST6 beta-galactoside alpha-2,6-sialyltransferase 2
ENSG00000134602.15	STK26	serine/threonine protein kinase 26
ENSG00000112837.16	TBX18	T-box 18
ENSG00000135111.15	TBX3	T-box 3
ENSG00000196628.15	TCF4	transcription factor 4
ENSG00000226674.8	TEX41	testis expressed 41 (non-protein coding)
ENSG00000168955.3	TM4SF20	transmembrane 4 L six family member 20
ENSG00000165152.8	TMEM246	transmembrane protein 246
ENSG00000103460.16	TOX3	TOX high mobility group box family member 3
ENSG00000198467.13	TPM2	tropomyosin 2 (beta)
ENSG00000067167.7	TRAM1	translocation associated membrane protein 1
ENSG00000067445.20	TRO	trophinin
ENSG00000137672.12	TRPC6	transient receptor potential cation channel member 6
ENSG00000121297.6	TSHZ3	teashirt zinc finger homeobox 3
ENSG00000182704.7	TSKU	tsukushi, small leucine rich proteoglycan
ENSG00000180543.4	TSPYL5	TSPY like 5
ENSG00000170703.15	TTLL6	tubulin tyrosine ligase like 6
ENSG00000233864.7	TTTY15	testis-specific transcript, Y-linked 15
ENSG00000122691.12	TWIST1	twist family bHLH transcription factor 1
ENSG00000131002.11	TXLNGY	taxilin gamma pseudogene, Y-linked
ENSG00000154277.12	UCHL1	ubiquitin C-terminal hydrolase L1

ENSG00000196620.9	UGT2B15	UDP glucuronosyltransferase family 2 member B15
ENSG00000114638.7	UPK1B	uroplakin 1B
ENSG00000114374.12	USP9Y	ubiquitin specific peptidase 9, Y-linked
ENSG00000150630.3	VEGFC	vascular endothelial growth factor C
ENSG00000232354.8	VIPR1-AS1	VIPR1 antisense RNA 1
ENSG00000146530.11	VWDE	von Willebrand factor D and EGF domains
ENSG00000150627.15	WDR17	WD repeat domain 17
ENSG00000166415.14	WDR72	WD repeat domain 72
ENSG00000248334.6	WHAMMP2	WAS protein homolog associated with actin
ENSG00000103489.11	XYLT1	xylosyltransferase 1
ENSG00000126970.15	ZC4H2	zinc finger C4H2-type containing
ENSG00000067646.11	ZFY	zinc finger protein, Y-linked
ENSG00000147180.16	ZNF711	zinc finger protein 711

TABLE A.2: All significantly differentially regulated genes in siRNA XIST in A549 cell line

Gene ID	Gene Symbol	Gene Name
ENSG00000163017.13	ACTG2	actin, gamma 2, smooth muscle, enteric
ENSG00000163568.14	AIM2	absent in melanoma 2
ENSG00000167612.12	ANKRD33	ankyrin repeat domain 33
ENSG00000166825.13	ANPEP	alanyl aminopeptidase, membrane)
ENSG00000198768.10	APCDD1L	APC down-regulated 1 like
ENSG00000231290.5	APCDD1L-AS1	APCDD1L antisense RNA 1
ENSG00000128284.19	APOL3	apolipoprotein L3
ENSG00000130707.17	ASS1	argininosuccinate synthase 1
ENSG00000043039.6	BARX2	BARX homeobox 2
ENSG00000130303.12	BST2	bone marrow stromal cell antigen 2
ENSG00000165181.16	C9orf84	chromosome 9 open reading frame 84
ENSG00000185860.13	CCDC190	coiled-coil domain containing 190
ENSG00000173421.16	CCDC36	coiled-coil domain containing 36
ENSG00000133048.12	CHI3L1	chitinase 3 like 1
ENSG00000111799.20	COL12A1	collagen type XII alpha 1 chain
ENSG00000164692.17	COL1A2	collagen type I alpha 2 chain
ENSG00000160111.12	CPAMD8	alpha-2-macroglobulin domain containing 8
ENSG00000146592.16	CREB5	cAMP responsive element binding protein 5
ENSG00000175315.2	CST6	cystatin E/M
ENSG00000204019.4	CT83	cancer/testis antigen 83
ENSG00000100867.14	DHRS2	dehydrogenase/reductase 2
ENSG00000134769.21	DTNA	dystrobrevin alpha
ENSG00000198842.9	DUSP27	dual specificity phosphatase 27
ENSG00000268964.1	ERVV-2	endogenous retrovirus group V member 2
ENSG00000177096.8	FAM109B	family with sequence similarity 109 member B
ENSG00000168754.14	FAM178B	family with sequence similarity 178 member B

ENSG00000231528.2	FAM225A	family with sequence similarity 225 member A
ENSG00000087258.14	GNAO1	G protein subunit alpha o1
ENSG00000237510.7	GPAT2P1	glycerol-3-phosphate acyltransferase 2 pseudogene 1
ENSG00000183454.16	GRIN2A	glutamate ionotropic receptor NMDA type subunit 2A
ENSG00000277075.2	HIST1H2AE	histone cluster 1 H2A family member e
ENSG00000168685.14	IL7R	interleukin 7 receptor
ENSG00000075043.18	KCNQ2	potassium voltage-gated channel subfamily Q member 2
ENSG00000167754.12	KLK5	kallikrein related peptidase 5
ENSG00000167755.14	KLK6	kallikrein related peptidase 6
ENSG00000233117.2	LINC00702	long intergenic non-protein coding RNA 702
ENSG00000227630.2	LINC01132	long intergenic non-protein coding RNA 1132
ENSG00000104903.4	LYL1	LYL1, basic helix-loop-helix family member
ENSG00000157227.12	MMP14	matrix metallopeptidase 14
ENSG00000169715.14	MT1E	metallothionein 1E
ENSG00000260549.1	MT1L	metallothionein 1L (gene/pseudogene)
ENSG00000205364.3	MT1M	metallothionein 1M
ENSG00000181143.15	MUC16	mucin 16, cell surface associated
ENSG00000136997.16	MYC	v-myc avian myelocytomatosis viral oncogene homolog
ENSG00000124479.8	NDP	NDP, norrin cystine knot growth factor
ENSG00000132688.10	NES	nestin
ENSG00000235568.6	NFAM1	NFAT activating protein with ITAM motif 1
ENSG00000083720.12	OXCT1	3-oxoacid CoA-transferase 1
ENSG00000089041.16	P2RX7	purinergic receptor P2X 7
ENSG00000198300.12	PEG3	paternally expressed 3
ENSG00000198890.7	PRMT6	protein arginine methyltransferase 6
ENSG00000230461.8	PROX1-AS1	PROX1 antisense RNA 1

ENSG00000275896.5	PRSS2	protease, serine 2
ENSG00000007038.10	PRSS21	protease, serine 21
ENSG00000112655.15	PTK7	protein tyrosine kinase 7 (inactive)
ENSG00000105894.11	PTN	pleiotrophin
ENSG00000182732.16	RGS6	regulator of G-protein signaling 6
ENSG00000168421.12	RHOH	ras homolog family member H
ENSG00000154133.14	ROBO4	roundabout guidance receptor 4
ENSG00000163220.10	S100A9	S100 calcium binding protein A9
ENSG00000197632.8	SERPINB2	serpin family B member 2
ENSG00000225383.7	SFTA1P	surfactant associated 1, pseudogene
ENSG00000101194.17	SLC17A9	solute carrier family 17 member 9
ENSG00000134802.17	SLC43A3	solute carrier family 43 member 3
ENSG00000174640.12	SLCO2A1	organic anion transporter family member 2A1
ENSG00000109610.5	SOD3	superoxide dismutase 3, extracellular
ENSG00000113140.10	SPARC	secreted protein acidic and cysteine rich
ENSG00000157303.10	SUSD3	sushi domain containing 3
ENSG00000184292.6	TACSTD2	tumor-associated calcium signal transducer 2
ENSG00000104055.15	TGM5	transglutaminase 5
ENSG00000180176.14	TH	tyrosine hydroxylase
ENSG00000119139.17	TJP2	tight junction protein 2
ENSG00000041982.15	TNC	tenascin C
ENSG00000165125.19	TRPV6	transient receptor potential cation channel V
ENSG00000182463.15	TSHZ2	teashirt zinc finger homeobox 2
ENSG00000006555.10	TTC22	tetratricopeptide repeat domain 22
ENSG00000265972.5	TXNIP	thioredoxin interacting protein
ENSG00000162543.5	UBXN10	UBX domain protein 10
ENSG00000235974.1	VN2R19P	vomeronasal 2 receptor 19 pseudogene
ENSG00000124116.18	WFDC3	WAP four-disulfide core domain 3
ENSG00000184828.9	ZBTB7C	zinc finger and BTB domain containing 7C
ENSG00000269793.5	ZIM2-AS1	ZIM2 antisense RNA 1
ENSG00000204789.4	ZNF204P	zinc finger protein 204, pseudogene

ENSG00000170954.11	ZNF415	zinc finger protein 415
ENSG00000197647.11	ZNF433	zinc finger protein 433
ENSG00000171291.8	ZNF439	zinc finger protein 439
ENSG00000177599.12	ZNF491	zinc finger protein 491
ENSG00000172000.7	ZNF556	zinc finger protein 556
ENSG00000198028.3	ZNF560	zinc finger protein 560
ENSG00000188868.13	ZNF563	zinc finger protein 563
ENSG00000161551.14	ZNF577	zinc finger protein 577
ENSG00000258405.9	ZNF578	zinc finger protein 578
ENSG00000197497.10	ZNF665	zinc finger protein 665
ENSG00000198046.11	ZNF667	zinc finger protein 667
ENSG00000166770.10	ZNF667-AS1	ZNF667 antisense RNA 1
ENSG00000197928.10	ZNF677	zinc finger protein 677
ENSG00000197124.11	ZNF682	zinc finger protein 682
ENSG00000198429.9	ZNF69	zinc finger protein 69
ENSG00000196757.7	ZNF700	zinc finger protein 700
ENSG00000214652.5	ZNF727	zinc finger protein 727
ENSG00000186777.11	ZNF732	zinc finger protein 732
ENSG00000167766.18	ZNF83	zinc finger protein 83
ENSG00000223547.9	ZNF844	zinc finger protein 844
ENSG00000184635.14	ZNF93	zinc finger protein 93

TABLE A.3: All significantly differentially regulated genes in siRNA XIST in H1975 cell line

A.3 Extended Data
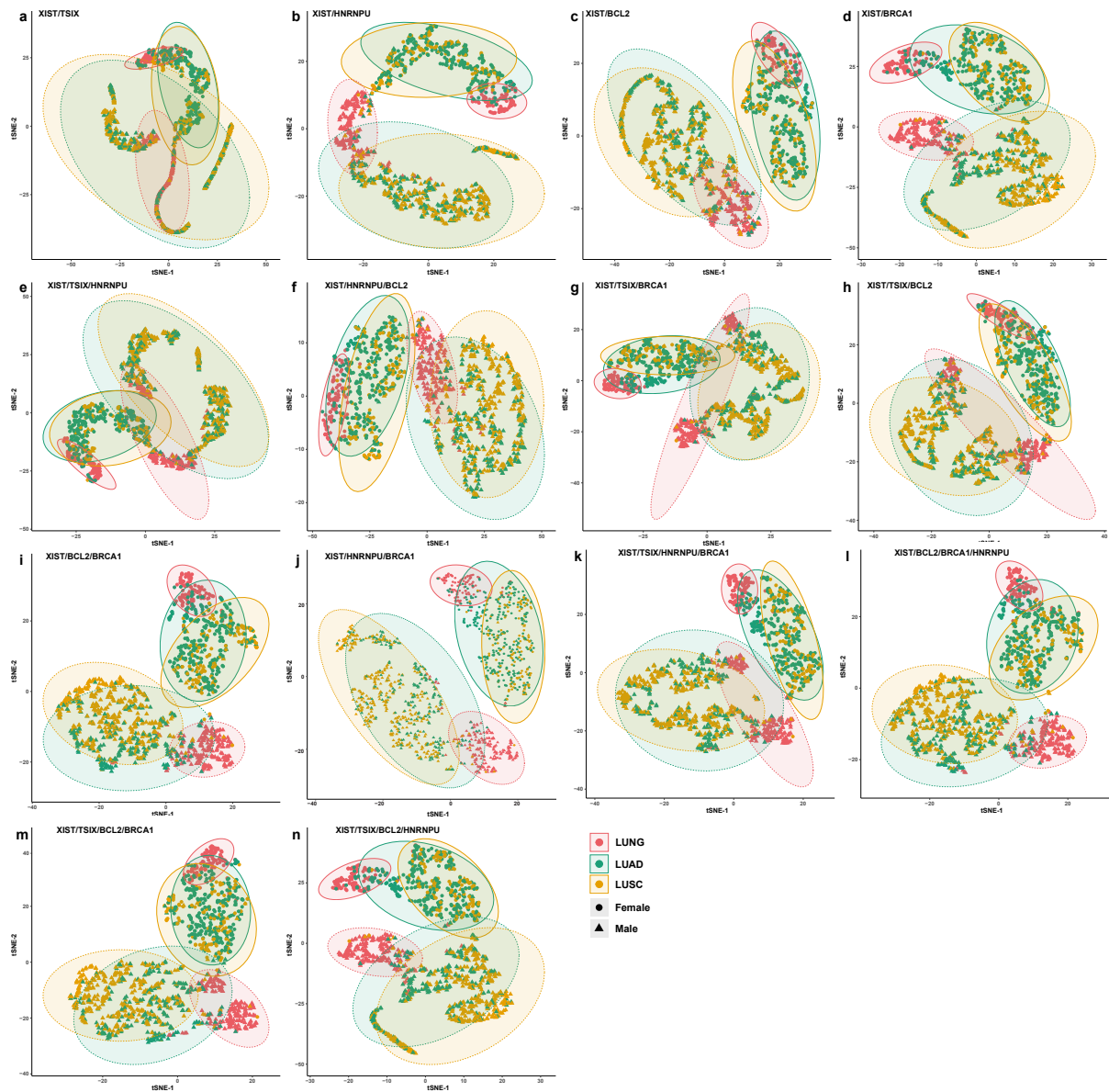
FIGURE A.1: tSNE clustering of various gene combinations differentiating for gender and sample type: (a) XIST/TSIX, (b) XIST/HNRNPU, (c)XIST/BCL2, (d) XIST/BRCA1, (e) XIST/TSIX/HNRNPU, (f) XIST/HNRNPU/BCL2, (g) XIST/TSIX/BRCA1, (h)XIST/TSIX/BCL2, (i) XIST/BCL2/BRCA1, (j) XIST/HNRNPU/BRCA1, (k) XIST/TSIX/HNRNPU/BRCA1, (l) XIST/BCL2/BRCA1/HNRNPU, (m) XIST/TSIX/BCL2/BRCA1, (n) XIST/TSIX/BCL2/HNRNPU.
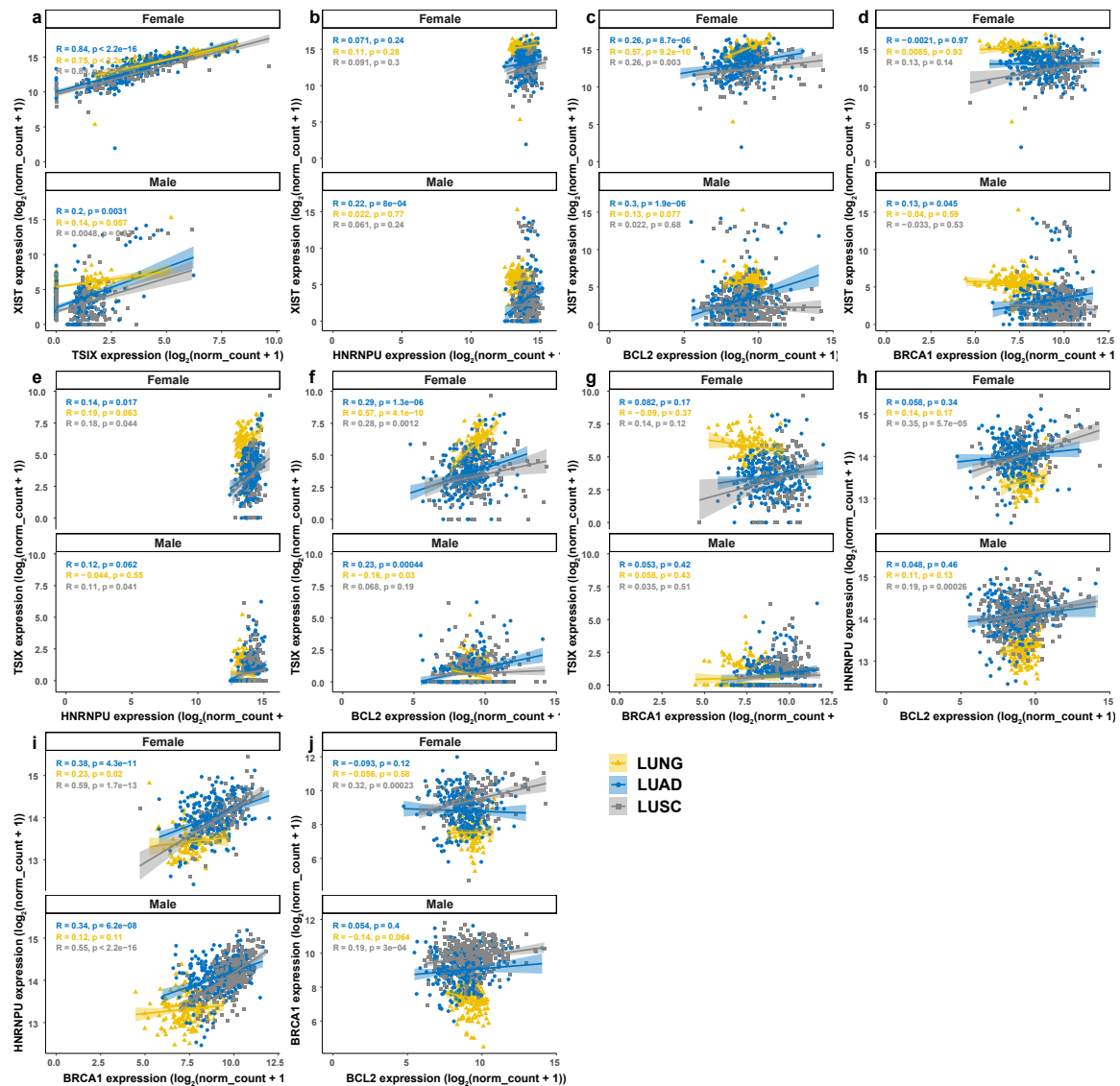
FIGURE A.2: Correlation with regression line for gene expression for all pair-wise combinations of genes from XIST, TSIX, hnRNPu, TSIX, Bcl-2 and BRCA1. Spearman's rank correlation test was used to generate the correlation coefficient represented as R in the graph, p represents p-value. Healthy lung, LUAD, and LUSC were compared separately in both sexes. Total sample number is 1299, 794 derived from male patients and 505 from female patients. There are 116 samples from male patients with 0 expression in XIST, 461 in TSIX, 1 in hnRNPu, Bcl-2, and BRCA1. In female patients derived samples, 17 show zero expression for TSIX gene. All samples with zero expression in both genes investigated were removed the analysis. Table S1: Spreadsheet with the differentially expressed genes in A549 and H1975 cell lines following siRNA of XIST.

Bibliography

1. Spiro, S. C. & Silvestri, G. A. One hundred years of lung cancer. *American Journal of Respiratory and Critical Care Medicine* **172,** 523–529. ISSN: 1073449X (2005).

2. Beadsmoore, C. J. & Screaton, N. J. Classification, staging and prognosis of lung cancer. *European Journal of Radiology* **45,** 8–17. ISSN: 0720048X (2003).

3. Athey, V. L., Walters, S. J. & Rogers, T. K. Symptoms at lung cancer diagnosis are associated with major differences in prognosis. *Thorax* **73,** 1177–1181. ISSN: 1468-3296 (Dec. 2018).

4. Sato, M., Shames, D. S., Gazdar, A. F. & Minna, J. D. A translational view of the molecular pathogenesis of lung cancer. *Journal of Thoracic Oncology* **2,** 327–343. ISSN: 15560864 (2007).

5. Herceg, Z. & Hainaut, P. Genetic and epigenetic alterations as biomarkers for cancer detection, diagnosis and prognosis. *Molecular oncology* **1,** 26–41. ISSN: 1878-0261 (June 2007).

6. Langevin, S. M., Kratzke, R. A. & Kelsey, K. T. Epigenetics of lung cancer. *Translational research : the journal of laboratory and clinical medicine* **165,** 74–90. ISSN: 1878-1810 (Jan. 2015).

7. Kim, D., Lee, Y.-S., Kim, D.-H. & Bae, S.-C. Lung Cancer Staging and Associated Genetic and Epigenetic Events. *Molecules and cells* **43,** 1–9. ISSN: 0219-1032 (Jan. 2020).

8. Gazdar, A. F. & Brambilla, E. Preneoplasia of lung cancer. *Cancer biomarkers : section A of Disease markers* **9,** 385–396. ISSN: 1875-8592 (2010).

9. Wistuba, I. I. *et al.* Molecular changes in the bronchial epithelium of patients with small cell lung cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **6,** 2604–2610. ISSN: 1078-0432 (July 2000).

10. Greenberg, A. K., Yee, H. & Rom, W. N. Preneoplastic lesions of the lung. *Respiratory Research* **3,** 13. ISSN: 1465-993X (2002).

11. De Groot, P. M., Wu, C. C., Carter, B. W. & Munden, R. F. The epidemiology of lung cancer. *Translational lung cancer research* **7,** 220–233. ISSN: 2218-6751 (June 2018).

12. Wong, M. C. S., Lao, X. Q., Ho, K.-F., Goggins, W. B. & Tse, S. L. A. Incidence and mortality of lung cancer: global trends and association with socioeconomic status. *Scientific Reports* **7,** 14300. ISSN: 2045-2322 (2017).

13. Dela Cruz, C. S., Tanoue, L. T. & Matthay, R. A. Lung cancer: epidemiology, etiology, and prevention. *Clinics in chest medicine* **32,** 605–644 (Dec. 2011).

14. Cheung, W. K. & Nguyen, D. X. Lineage factors and differentiation states in lung cancer progression. *Oncogene* **34,** 5771–5780. ISSN: 14765594 (2015).

15. Tobias, J. & Hochhauser, D. in *Cancer and its Management* 197–222 (Wiley-Blackwell, 2009). ISBN: 9781444306361.

16. Travis, W. D. *et al.* The 2015 World Health Organization Classification of Lung Tumors. *Journal of Thoracic Oncology* **10,** 1243–1260. ISSN: 15560864 (2015).

17. Ismail-Khan, R. *et al.* Malignant pleural mesothelioma: A comprehensive review. *Cancer Control* **13,** 255–263. ISSN: 10732748 (2006).

18. Kaufman, A. J. & Flores, R. M. Surgical treatment of malignant pleural mesothelioma. *Current treatment options in oncology* **12,** 201–216. ISSN: 1534-6277 (June 2011).

19. Jiang, Y., Goldberg, I. D. & Shi, Y. E. Complex roles of tissue inhibitors of metalloproteinases in cancer. *Oncogene* **21,** 2245–2252. ISSN: 0950-9232 (2002).

20. Pantel, K. & Brakenhoff, R. H. Dissecting the metastatic cascade. *Nature Reviews Cancer* **4,** 448–456. ISSN: 1474-175X (2004).

21. Chambers, A. F., Groom, A. C. & MacDonald, I. C. Metastasis: Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer* **2,** 563–572. ISSN: 1474-175X (Aug. 2002).

22. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70. ISSN: 0092-8674. eprint: 0208024 (gr-qc) (2000).

23. Fares, J., Fares, M. Y., Khachfe, H. H., Salhab, H. A. & Fares, Y. Molecular principles of metastasis: a hallmark of cancer revisited. *Signal transduction and targeted therapy* **5,** 28 (Mar. 2020).

24. Valastyan, S. & Weinberg, R. A. Tumor metastasis: molecular insights and evolving paradigms. *Cell* **147,** 275–292. ISSN: 1097-4172 (Oct. 2011).

25. Chen, L., Bode, A. M. & Dong, Z. Circulating tumor cells: Moving biological insights into detection. *Theranostics* **7,** 2606–2619. ISSN: 18387640 (2017).

26. Harris, L. *et al.* American society of clinical oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *Journal of Clinical Oncology* **25,** 5287–5312. ISSN: 0732183X (2007).

27. Mego, M. *et al.* Characterization of metastatic breast cancer patients with non-detectable circulating tumor cells. *International Journal of Cancer* **129,** 417–423. ISSN: 00207136 (2011).

28. Alix-Panabières, C., Schwarzenbach, H. & Pantel, K. Circulating Tumor Cells and Circulating Tumor DNA. *Annual Review of Medicine* **63,** 199–215. ISSN: 0066-4219 (Jan. 2012).

29. Alix-Panabieres, C. & Pantel, K. Circulating tumor cells: liquid biopsy of cancer. *Clinical chemistry* **59,** 110–118. ISSN: 1530-8561 (Jan. 2013).

30. Bednarz-Knoll, N., Alix-Panabières, C. & Pantel, K. Plasticity of disseminating cancer cells in patients with epithelial malignancies. *Cancer and Metastasis Reviews* **31,** 673–687. ISSN: 1573-7233 (Dec. 2012).

31. Li, J., Gregory, S. G., Garcia-Blanco, M. A. & Armstrong, A. J. Using circulating tumor cells to inform on prostate cancer biology and clinical utility. *Critical reviews in clinical laboratory sciences* **52,** 191–210. ISSN: 1549-781X (2015).

32. Schneck, H. *et al.* EpCAM-Independent Enrichment of Circulating Tumor Cells in Metastatic Breast Cancer. *PloS one* **10,** e0144535 (2015).

33. Yu, M., Stott, S., Toner, M., Maheswaran, S. & Haber, D. A. Circulating tumor cells: Approaches to isolation and characterization. *Journal of Cell Biology* **192,** 373–382. ISSN: 00219525 (2011).

34. Han, X., Wang, J. & Sun, Y. Circulating Tumor DNA as Biomarkers for Cancer Detection. *Genomics, proteomics & bioinformatics* **15,** 59–72. ISSN: 2210-3244 (Apr. 2017).

35. Müller, V. *et al.* Circulating Tumor Cells in Breast Cancer : Correlation to Bone Marrow Micrometastases , Heterogeneous Response to Systemic Therapy and Low Proliferative Activity. *Clin Cancer Res* **11,** 3678–3685. ISSN: 1078-0432 (2005).

36. Wülfing, P. *et al.* HER2-positive circulating tumor cells indicate poor clinical outcome in stage I to III breast cancer patients. *Clinical Cancer Research* **12,** 1715–1720. ISSN: 10780432 (2006).

37. Vona, G. *et al.* Isolation by Size of Epithelial Tumor Cells. *The American Journal of Pathology* **156,** 57–63. ISSN: 00029440 (2000).

38. Marrinucci, D. *et al.* Case study of the morphologic variation of circulating tumor cells. *Human Pathology* **38,** 514–519. ISSN: 00468177 (2007).

39. Tan, S. J., Yobas, L., Lee, G. Y. H., Ong, C. N. & Lim, C. T. Microdevice for the isolation and enumeration of cancer cells from blood. *Biomedical Microdevices* **11,** 883–892. ISSN: 13872176 (2009).

40. Mohamed, H., Murray, M., Turner, J. N. & Caggana, M. Isolation of tumor cells using size and deformation. *Journal of Chromatography A* **1216,** 8289–8295. ISSN: 00219673 (2009).

41. Hanssen, A. *et al.* Characterization of different CTC subpopulations in non-small cell lung cancer. *Scientific reports* **6,** 28010. ISSN: 2045-2322 (2016).

42. Smit, D. J. *et al.* High Sensitivity of Circulating Tumor Cells Derived from a Colorectal Cancer Patient for Dual Inhibition with AKT and mTOR Inhibitors. *Cells* **9,** 2129. ISSN: 2073-4409 (Sept. 2020).

43. Yin, L.-C. *et al.* Twist Expression in Circulating Hepatocellular Carcinoma Cells Predicts Metastasis and Prognoses. *BioMed research international* **2018,** 3789613. ISSN: 2314-6141 (June 2018).

44. Jhi, J. H. *et al. Circulating Tumor Cells and TWIST Expression in Patients with Metastatic Gastric Cancer* 2020.

45. Poruk, K. E. *et al.* Circulating Tumor Cells Expressing Markers of Tumor-Initiating Cells Predict Poor Survival and Cancer Recurrence in Patients with Pancreatic Ductal Adenocarcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* **23,** 2681–2690. ISSN: 1078-0432 (June 2017).

46. Calabuig-Fariñas, S., Jantus-Lewintre, E., Herreros-Pomares, A. & Camps, C. Circulating tumor cells versus circulating tumor DNA in lung cancer—which one will win? *Translational Lung Cancer Research* **5,** 466–482. ISSN: 22186751 (2016).

47. Tanaka, F. *et al.* Circulating tumor cell as a diagnostic marker in primary lung cancer. *Clinical Cancer Research* **15,** 6980–6986. ISSN: 10780432 (2009).

48. Isobe, K. *et al.* Clinical significance of circulating tumor cells and free DNA in non-small cell lung cancer. *Anticancer Research* **32,** 3339–3344. ISSN: 02507005 (2012).

49. Pailler, E. *et al.* High level of chromosomal instability in circulating tumor cells of ROS1-rearranged non-small-cell lung cancer. *Annals of Oncology* **26,** 1408–1415. ISSN: 15698041 (2015).

50. Hashimoto, M. *et al.* Significant increase in circulating tumour cells in pulmonary venous blood during surgical manipulation in patients with primary

lung cancer. *Interactive Cardiovascular and Thoracic Surgery* **18,** 775–783. ISSN: 15699285 (2014).

51. Hansen, C. H. & Oudenaarden, A. V. HHS Public Access. **10,** 869–871 (2014).

52. Grundberg, I. *et al.* In situ mutation detection and visualization of intratumor heterogeneity for cancer research and diagnostics ABSTRACT : **4** (2013).

53. Ke, R., Nong, R. Y., Fredriksson, S., Landegren, U. & Nilsson, M. Improving Precision of Proximity Ligation Assay by Amplified Single Molecule Detection. **8,** 1–5 (2013).

54. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. **10** (2013).

55. Nagrath, S. *et al.* Isolation of rare circulating tumour cells in cancer patients by microchip technology. *Nature* **450,** 1235–1239 (Dec. 2007).

56. Wendel, M. *et al.* Fluid biopsy for circulating tumor cell identification in patients with early-and late-stage non-small cell lung cancer: a glimpse into lung cancer biology. *Physical Biology* **9,** 16005. ISSN: 1478-3967 (2012).

57. Wang, X. *et al.* Extracellular mRNA detected by molecular beacons in tethered lipoplex nanoparticles for diagnosis of human hepatocellular carcinoma. *PloS one* **13,** e0198552 (2018).

58. Lee, L. J. *et al.* Extracellular mRNA Detected by Tethered Lipoplex Nanoparticle Biochip for Lung Adenocarcinoma Detection. *American journal of respiratory and critical care medicine* **193,** 1431–1433. ISSN: 1535-4970 (June 2016).

59. Siemer, S. *et al.* Nano meets micro-translational nanotechnology in medicine: Nano-based applications for early tumor detection and therapy. *Nanomaterials* **10,** 1–28. ISSN: 20794991 (2020).

60. Liao, B.-C., Lin, C.-C. & Yang, J. C.-H. Second and third-generation epidermal growth factor receptor tyrosine kinase inhibitors in advanced nonsmall cell lung cancer. *Current Opinion in Oncology* **27.** ISSN: 1040-8746 (2015).

61. Sayed, M., Zahran, A. M., Hassan, M. S. F. & Mohamed, D. O. Circulating Tumor Cells and Cancer Stem Cells: Clinical Implications in Nonmetastatic Breast Cancer. *Breast Cancer: Basic and Clinical Research* **10,** BCBCR.S40856. ISSN: 1178-2234 (Jan. 2016).

62. Chudasama, D. *et al.* Detection of Circulating Tumour Cells and Survival of Patients with Non-small Cell Lung Cancer. *Anticancer research* **37,** 169–173 (Jan. 2017).

63. Kopreski, M. S. *et al.* Detection of mutant K-ras DNA in plasma or serum of patients with colorectal cancer. **76,** 1293–1299 (1997).

64. Kopreski, M. S. *et al.* Somatic Mutation Screening : Identification of Individuals Harboring K-ras Mutations With the Use of Plasma DNA. **92** (2000).

65. Valastyan, S. & Weinberg, R. A. Review Tumor Metastasis : Molecular Insights and Evolving Paradigms. *Cell* **147,** 275–292. ISSN: 0092-8674 (2011).

66. Vogelstein, B. *et al.* Cancer genome landscapes. *Science (New York, N.Y.)* **339,** 1546–1558. ISSN: 1095-9203 (Mar. 2013).

67. Meric-bernstam, F. & Mills, G. B. Overcoming implementation challenges of personalized cancer therapy. *Nature Publishing Group* **9,** 542–548. ISSN: 1759-4774 (2012).

68. Marklein, R. A., Lam, J., Guvendiren, M., Sung, K. E. & Bauer, S. R. Functionally-Relevant Morphological Profiling: A Tool to Assess Cellular Heterogeneity. *Trends in Biotechnology* **36,** 105–118. ISSN: 0167-7799 (Jan. 2018).

69. Marusyk, A. & Polyak, K. Tumor heterogeneity: causes and consequences. *Biochimica et biophysica acta* **1805,** 105–117. ISSN: 0006-3002 (Jan. 2010).

70. Ramón y Cajal, S. *et al.* Clinical implications of intratumor heterogeneity: challenges and opportunities. *Journal of Molecular Medicine* **98,** 161–177. ISSN: 1432-1440 (2020).

71. Tellez-Gabriel, M., Ory, B., Lamoureux, F., Heymann, M.-F. & Heymann, D. Tumour Heterogeneity: The Key Advantages of Single-Cell Analysis. *International journal of molecular sciences* **17,** 2142. ISSN: 1422-0067 (Dec. 2016).

72. Endesfelder, D. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366,** 883–92 (2012).

73. Yates, L. R. & Campbell, P. J. Evolution of the cancer genome. *Nature Reviews Genetics* **13,** 795 (Oct. 2012).

74. Ignatiadis, M., Lee, M. & Jeffrey, S. S. Circulating Tumor Cells and Circulating Tumor DNA : Challenges and Opportunities on the Path to Clinical Utility. **21,** 4786–4801 (2015).

75. Friel, A. M., Corcoran, C., Crown, J. & Driscoll, L. O. Relevance of circulating tumor cells , extracellular nucleic acids , and exosomes in breast cancer, 613–625 (2010).

76. Huang, A. *et al.* Plasma Circulating Cell-free DNA Integrity as a Promising Biomarker for Diagnosis and Surveillance in Patients with Hepatocellular Carcinoma. *Journal of Cancer* **7,** 1798–1803. ISSN: 1837-9664 (2016).

77. Fournié, G. J. *et al.* Plasma DNA as a marker of cancerous cell death. Investigations in patients suffering from lung cancer and in nude mice bearing human tumours. *Cancer Letters* **91,** 221–227. ISSN: 03043835 (1995).

78. Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics. *Nature medicine* **14,** 985–990. ISSN: 1078-8956 (Sept. 2008).

79. Umetani, N. *et al.* Increased integrity of free circulating DNA in sera of patients with colorectal or periampullary cancer: direct quantitative PCR for ALU repeats. *Clinical chemistry* **52,** 1062–1069. ISSN: 0009-9147 (June 2006).

80. Gormally, E., Caboux, E., Vineis, P. & Hainaut, P. Circulating free DNA in plasma or serum as biomarker of carcinogenesis: Practical aspects and biological significance. *Mutation Research - Reviews in Mutation Research* **635,** 105–117. ISSN: 13835742 (2007).

81. Jiang, T., Ren, S. & Zhou, C. Role of circulating-tumor DNA analysis in non-small cell lung cancer. *Lung Cancer* **90,** 128–134. ISSN: 18728332 (2015).

82. Jahr, S. *et al.* DNA Fragments in the Blood Plasma of Cancer Patients : Quantitations and Evidence for Their Origin from Apoptotic and Necrotic Cells 1, 1659–1665 (2001).

83. Leon, S. A. *et al.* Free DNA in the Serum of Cancer Patients and the Effect of Therapy Free DNA in the Serum of Cancer Patients and the Effect of Therapy. *Cancer Res* **37,** 646–650. ISSN: 15387445 (1977).

84. Anker, P., Mulcahy, H., Chen, X. Q. & Stroun, M. Detection of circulating tumour DNA in the blood (plasma/serum) of cancer patients. *Cancer and Metastasis reviews* **18,** 65–73. ISSN: 0167-7659 (1999).

85. Stroun, M. *et al.* Neoplastic Characteristics of the DNA Found in the Plasma of Cancer Patients. *Oncology* **46,** 318–322. ISSN: 0030-2414 (1989).

86. Van de Stolpe, A. & den Toonder, J. M. J. Circulating tumor cells: What is in it for the patient? A vision towards the future. *Cancers* **6,** 1195–1207. ISSN: 20726694 (2014).

87. Neri, E. *et al.* Radiomics and liquid biopsy in oncology: the holons of systems medicine. *Insights into imaging* **9,** 915–924. ISSN: 1869-4101 (Dec. 2018).

88. Wong, S. Q. & Dawson, S.-J. Combining liquid biopsies and PET-CT for early cancer detection. *Nature Medicine* **26,** 1010–1011. ISSN: 1546-170X (2020).

89. Ganten, M. K., Ganten, T. M., Schlemmer, H. P., Cancer, G. & Cancer, G. Radiological Monitoring of the Treatment of Solid Tumors in Practice Radiologisches Monitoring der Therapie von soliden Tumoren in der Praxis, 466–473 (2014).

90. Skougaard, K. *et al.* Surveillance With PET/CT and Liquid Biopsies of Stage I-III Lung Cancer Patients After Completion of Definitive Therapy: A Randomized Controlled Trial (SUPER). *Clinical Lung Cancer* **21,** e61–e64. ISSN: 1525-7304 (2020).

91. Katopodis, P. *et al.* Kinase inhibitors and ovarian cancer. *Cancers* **11,** 1–17. ISSN: 20726694 (2019).

92. Marchetti, C. *et al.* Tyrosine-kinases inhibitors in recurrent platinum-resistant ovarian cancer patients. *CANCER TREATMENT REVIEWS.* ISSN: 0305-7372 (2015).

93. Bhullar, K. S. *et al.* Kinase-targeted cancer therapies: progress, challenges and future directions. *Molecular Cancer* **17,** 48. ISSN: 1476-4598 (2018).

94. Polier, S. *et al.* ATP-competitive inhibitors block protein kinase recruitment to the Hsp90-Cdc37 system. *Nature Chemical Biology* **9,** 307 (Mar. 2013).

95. Kukita, Y. *et al.* Quantitative identification of mutant alleles derived from lung cancer in plasma cell-free DNA via anomaly detection using deep sequencing data. *PLoS ONE* **8,** 1–10. ISSN: 19326203 (2013).

96. Kobayashi, S. *et al.* <i>EGFR</i> Mutation and Resistance of Non–Small-Cell Lung Cancer to Gefitinib. *New England Journal of Medicine* **352,** 786–792. ISSN: 0028-4793 (2005).

97. Guo, N. *et al.* Circulating tumor DNA detection in lung cancer patients before and after surgery. *Scientific Reports* **6,** 33519. ISSN: 2045-2322 (2016).

98. Wang, S., Cang, S. & Liu, D. Third-generation inhibitors targeting EGFR T790M mutation in advanced non-small cell lung cancer. *Journal of Hematology & Oncology* **9,** 34. ISSN: 1756-8722 (2016).

99. Cross, D. A. E. *et al.* NIH Public Access. **4,** 1046–1061 (2015).

100. Kasner, E., Hunter, C. A., Ph, D., Kariko, K. & Ph, D. NIH Public Access. **70,** 646–656. ISSN: 09652140 (2013).

101. Thompson, J. C. *et al.* Detection of therapeutically targetable driver and resistance mutations in lung cancer patients by next generation sequencing of cell-free circulating tumor DNA. *Clinical Cancer Research* (Jan. 2016).

102. Bethune, G., Bethune, D., Ridgway, N. & Xu, Z. Epidermal growth factor receptor (EGFR) in lung cancer: an overview and update. *Journal of thoracic disease* **2,** 48–51. ISSN: 2077-6624 (Mar. 2010).

103. Awad, M. M. & Shaw, A. T. ALK Inhibitors in Non–Small Cell Lung Cancer: Crizotinib and Beyond. *Clinical advances in hematology & oncology : H&O* **12,** 429–439. ISSN: 1543-0790 (July 2014).

104. Oda, K., Matsuoka, Y., Funahashi, A. & Kitano, H. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology* **1.** ISSN: 1744-4292 (2005).

105. Gazdar, A. Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene* **28,** 1–14. ISSN: 1476-5594 (2009).

106. Joshi, M., Rizvi, S. M. & Belani, C. P. Afatinib for the treatment of metastatic non-small cell lung cancer. *Cancer Management and Research* **7,** 75–82. ISSN: 11791322 (2015).

107. Mok, T. S. *et al.* Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma. *New England Journal,* 947–957 (2009).

108. Zhou, C. *et al.* Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *The Lancet Oncology* **12,** 735–742. ISSN: 14702045 (2011).

109. Wu, K., House, L., Liu, W. & Cho, W. C. Personalized targeted therapy for lung cancer. *International Journal of Molecular Sciences* **13,** 11471–11496. ISSN: 16616596 (2012).

110. Barnes, T. A., O'Kane, G. M., Vincent, M. D. & Leighl, N. B. Third-Generation Tyrosine Kinase Inhibitors Targeting Epidermal Growth Factor Receptor Mutations in Non-Small Cell Lung Cancer. *Frontiers in Oncology* **7,** 1–9. ISSN: 2234-943X (2017).

111. Kumar, A., Petri, E. T., Halmos, B. & Boggon, T. J. The Structure and Clinical Relevance of the EGF Receptor in Human Cancer. *Journal of Clinical Oncology* **26,** 1742–1751 (2008).

112. Greulich, H. *et al.* Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. *PLoS Medicine* **2,** 1167–1176. ISSN: 15491277 (2005).

113. Wu, J. Y. *et al.* Lung cancer with epidermal growth factor receptor exon 20 mutations is associated with poor gefitinib treatment response. *Clinical Cancer Research* **14,** 4877. ISSN: 10780432 (2008).

114. Wu, J. Y. *et al.* Lung cancer with epidermal growth factor receptor exon 20 mutations is associated with poor gefitinib treatment response. *Clinical Cancer Research* **14,** 4877. ISSN: 10780432 (2008).

115. Pao, W. *et al.* KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Medicine* **2,** 0057–0061. ISSN: 15491277 (2005).

116. Stella, G. M. *et al.* Unexpected responses to EGFR inhibition in NSCLC. *Respiratory Medicine Case Reports* **16,** 32–34. ISSN: 22130071 (2015).

117. Ogunleye, F., Ibrahim, M., Stender, M., Kalemkerian, G. & Jaiyesimi, I. Epidermal growth factor receptor tyrosine kinase inhibitors in advanced non-small cell lung cancer a paradigm shift in stage IV non-small cell lung cancer treatment. *American Journal of Hematology/Oncology®* **11** (2015).

118. Hannon, G. J. & Dela, M. J. lncRNAs in development and disease : from functions to mechanisms. *Open Biology,* 1–10 (2017).

119. Zhang, Y.-L. *et al.* The lncRNA XIST exhibits oncogenic properties via regulation of miR-449a and Bcl-2 in human non-small cell lung cancerThis article has been corrected since Advanced Online Publication, and an erratum is also printed in this issue. *Acta pharmacologica Sinica* **38,** 371–381 (Mar. 2017).

120. Wang, H. *et al.* The Long Non-Coding RNA XIST Controls Non-Small Cell Lung Cancer Proliferation and Invasion by Modulating miR-186-5p. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* **41,** 2221–2229. ISSN: 1421-9778 (2017).

121. Prensner, J. R. & Chinnaiyan, A. M. The emergence of lncRNAs in cancer biology. *Cancer discovery* **1,** 391–407. ISSN: 2159-8290 (Oct. 2011).

122. Niland, C. N., Merry, C. R. & Khalil, A. M. Emerging Roles for Long Non-Coding RNAs in Cancer and Neurological Disorders. *Frontiers in genetics* **3,** 25. ISSN: 1664-8021 (Feb. 2012).

123. Liu, K. *et al.* Long non-coding RNAs regulate drug resistance in cancer. *Molecular Cancer* **19,** 54. ISSN: 1476-4598 (2020).

124. Li, W., Zhang, T., Guo, L. & Huang, L. Regulation of PTEN expression by noncoding RNAs. *Journal of experimental & clinical cancer research : CR* **37,** 223. ISSN: 1756-9966 (Sept. 2018).

125. Mu, M., Niu, W., Zhang, X., Hu, S. & Niu, C. LncRNA BCYRN1 inhibits glioma tumorigenesis by competitively binding with miR-619-5p to regulate CUEDC2 expression and the PTEN/AKT/p21 pathway. *Oncogene* **39,** 6879–6892. ISSN: 1476-5594 (2020).

126. Zhang, D. *et al.* Oncogenic RAS Regulates Long Noncoding RNA Orilnc1 in Human Cancer. *Cancer research* **77,** 3745–3757. ISSN: 1538-7445 (July 2017).

127. Yu, S.-N. *et al.* KRAS-related noncoding RNAs in pancreatic ductal adenocarcinoma. *Chronic diseases and translational medicine* **2,** 215–222. ISSN: 2095-882X (Dec. 2016).

128. Autuoro, J. M., Pirnie, S. P. & Carmichael, G. G. Long noncoding RNAs in imprinting and X chromosome inactivation. *Biomolecules* **4,** 76–100. ISSN: 2218-273X (Jan. 2014).

129. Kanduri, C. Long noncoding RNAs: Lessons from genomic imprinting. *Biochimica et biophysica acta* **1859,** 102–111. ISSN: 0006-3002 (Jan. 2016).

130. Gendrel, A. V. & Heard, E. Fifty years of x-inactivation research. *Development* **138,** 5049–5055. ISSN: 09501991 (2011).

131. Panning, B. X-chromosome inactivation: the molecular basis of silencing. *Journal of Biology* **7,** 30. ISSN: 1475-4924 (2008).

132. Yang, Z., Jiang, X., Jiang, X. & Zhao, H. X-inactive-specific transcript: A long noncoding RNA with complex roles in human cancers. *Gene* **679,** 28–35. ISSN: 18790038 (2018).

133. Pintacuda, G., Young, A. N. & Cerase, A. Function by Structure: Spotlights on Xist Long Non-coding RNA. *Frontiers in Molecular Biosciences* **4,** 1–11. ISSN: 2296-889X (2017).

134. Cerase, A., Pintacuda, G., Tattermusch, A. & Avner, P. Xist localization and function: New insights from multiple levels. *Genome Biology* **16,** 1–12. ISSN: 1474760X (2015).

135. Fang, R., Moss, W. N., Rutenberg-Schoenberg, M. & Simon, M. D. Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. *PLoS genetics* **11,** e1005668–e1005668. ISSN: 1553-7404 (Dec. 2015).

136. Wang, H. *et al.* The long non-coding RNA XIST controls non-small cell lung cancer proliferation and invasion by modulating miR-186-5p. *Cellular Physiology and Biochemistry* **41,** 2221–2229. ISSN: 14219778 (2017).

137. Yao, Y. *et al.* Knockdown of long non-coding RNA XIST exerts tumor-suppressive functions in human glioblastoma stem cells by up-regulating miR-152. *Cancer Letters* **359,** 75–86. ISSN: 18727980 (2015).

138. Fang, J., Sun, C.-c. C. & Gong, C. Long noncoding RNA XIST acts as an oncogene in non-small cell lung cancer by epigenetically repressing KLF2 expression. *Biochemical and Biophysical Research Communications* **478,** 811–817. ISSN: 10902104 (2016).

139. Huang, Y.-S., Chang, C.-C., Lee, S.-S., Jou, Y.-S. & Shih, H.-M. Xist reduction in breast cancer upregulates AKT phosphorylation via HDAC3-mediated repression of PHLPP1 expression. *Oncotarget* **7,** 43256–43266. ISSN: 1949-2553 (July 2016).

140. Katopodis, P. *et al.* In Silico and In Vitro Analysis of lncRNA XIST Reveals a Panel of Possible Lung Cancer Regulators and a Five-Gene Diagnostic Signature. *Cancers* **12.** ISSN: 2072-6694 (Nov. 2020).

141. Salvador, M. A. *et al.* The histone deacetylase inhibitor abexinostat induces cancer stem cells differentiation in breast cancer with low Xist expression. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19,** 6520–6531. ISSN: 1078-0432 (Dec. 2013).

142. Kobayashi, R. *et al.* Increased expression of long non-coding RNA XIST predicts favorable prognosis of cervical squamous cell carcinoma subsequent to definitive chemoradiation therapy. *Oncology letters* **12,** 3066–3074. ISSN: 1792-1074 (Nov. 2016).

143. Chen, D.-L. *et al.* Long noncoding RNA XIST expedites metastasis and modulates epithelial-mesenchymal transition in colorectal cancer. *Cell death & disease* **8,** e3011 (Aug. 2017).

144. Chen, D., Ren, J., Wang, Y., Li, B. & Gu, Y. Intraoperative monitoring of blood perfusion in port wine stains by laser Doppler imaging during vascular targeted photodynamic therapy: A preliminary study. *Photodiagnosis and photodynamic therapy* **14,** 142–151 (June 2016).

145. Du, P. *et al.* LncRNA-XIST interacts with miR-29c to modulate the chemoresistance of glioma cell to TMZ through DNA mismatch repair pathway. *Bioscience reports* **37** (Oct. 2017).

146. Mo, Y. *et al.* Long non-coding RNA XIST promotes cell growth by regulating miR-139-5p/PDK1/AKT axis in hepatocellular carcinoma. *Tumour biology : the*

     journal of the International Society for Oncodevelopmental Biology and Medicine **39,** 1010428317690999 (Feb. 2017).

147. Song, P., Ye, L.-F., Zhang, C., Peng, T. & Zhou, X.-H. Long non-coding RNA XIST exerts oncogenic functions in human nasopharyngeal carcinoma by targeting miR-34a-5p. *Gene* **592,** 8–14 (Oct. 2016).

148. Wei, W., Liu, Y., Lu, Y., Yang, B. & Tang, L. LncRNA XIST Promotes Pancreatic Cancer Proliferation Through miR-133a/EGFR. *Journal of cellular biochemistry* **118,** 3349–3358 (Oct. 2017).

149. Li, G.-L., Wu, Y.-X., Li, Y.-M. & Li, J. High expression of long non-coding RNA XIST in osteosarcoma is associated with cell proliferation and poor prognosis. *European review for medical and pharmacological sciences* **21,** 2829–2834 (June 2017).

150. Huang, K.-C. *et al.* Relationship of XIST expression and responses of ovarian cancer to chemotherapy. *Molecular cancer therapeutics* **1,** 769–776. ISSN: 1535-7163 (Aug. 2002).

151. Yao, Y. *et al.* Knockdown of long non-coding RNA XIST exerts tumor-suppressive functions in human glioblastoma stem cells by up-regulating miR-152. *Cancer letters* **359,** 75–86 (Apr. 2015).

152. Wang, H. *et al.* The Long Non-Coding RNA XIST Controls Non-Small Cell Lung Cancer Proliferation and Invasion by Modulating miR-186-5p. *Cellular Physiology and Biochemistry* **41,** 2221–2229. ISSN: 1015-8987 (2017).

153. Silver, D. P. *et al. Further evidence for BRCA1 communication with the inactive X chromosome.* Mar. 2007.

154. Wang, P., Chen, D., Ma, H. & Li, Y. Long non-coding RNA MEG3 regulates proliferation and apoptosis in non-small cell lung cancer: Via the miR-205-5p/LRP1 pathway. *RSC Advances* **7,** 49710–49719. ISSN: 20462069 (2017).

155. Shen, L. *et al.* Prognostic and clinicopathological roles of long non-coding RNA XIST in human cancers: a meta-analysis. *Translational Cancer Research; Vol 7, No 6 (December 2018): Translational Cancer Research* (2018).

156. Hennig, H. *et al.* An open-source solution for advanced imaging flow cytometry data analysis using machine learning. *Methods* **112,** 1–10. ISSN: 10462023 (2016).

157. Schehr, J. L. *et al.* High specificity in circulating tumor cell identification is required for accurate evaluation of programmed death-ligand 1. *PLoS ONE* **11,** 1–15. ISSN: 19326203 (2016).

158. Dent, B. M. *et al. of Circulating Tumour Cells From Patients With Oesophageal , Hepatocellular , Thyroid and Ovarian Cancers* 2016.

159. Dominical, V., Samsel, L. & McCoy, J. P. J. Masks in imaging flow cytometry. *Methods (San Diego, Calif.)* **112,** 9–17 (Jan. 2017).

160. George, T. C. *et al.* Quantitative measurement of nuclear translocation events using similarity analysis of multispectral cellular images obtained in flow. *Journal of Immunological Methods* **311,** 117–129. ISSN: 00221759 (2006).

161. Ang, R. R. Spectral image cytometry for circulating tumor cell identification (2017).

162. Basiji, D. A., Ortyn, W. E., Liang, L., Venkatachalam, V. & Morrissey, P. Cellular image analysis and imaging by flow cytometry. *Clinics in laboratory medicine* **27,** 653–70, viii. ISSN: 0272-2712 (Sept. 2007).

163. Dowski, E. R. & Cathey, W. T. Dowski - 1995 Extended depth of field through wave-front coding.pdf. **34,** 1859–1866 (1995).

164. Depth, E. Extended Depth of Field. *Japanese Journal of Visual Science* **37,** 142–144. ISSN: 2188-0522 (2016).

165. Design, E., Imagestream, T. & Type, C. Sample Preparation Guide ISX Quick Start Guide : Sample Preparation Sample Preparation Guide ISX Quick Start Guide : 7–9 (2009).

166. Zuba-Surma, E. K., Kucia, M., Abdel-Latif, A., Lillard, J. W. & Ratajczak, M. Z. The ImageStream system: A key step to a new era in imaging. *Folia Histochemica et Cytobiologica* **45,** 279–290. ISSN: 02398508 (2007).

167. Hennig, H. *et al.* An open-source solution for advanced imaging flow cytometry data analysis using machine learning. *Methods (San Diego, Calif.)* **112,** 201–210. ISSN: 1095-9130 (Jan. 2017).

168. Jones, T. R. *et al.* CellProfiler Analyst: data exploration and analysis software for complex image-based screens. *BMC Bioinformatics* **9,** 482. ISSN: 1471-2105 (2008).

169. Umetani, N. *et al.* Prediction of breast tumor progression by integrity of free circulating DNA in serum. *Journal of Clinical Oncology* **24,** 4270–4276. ISSN: 0732183X (2006).

170. Joosse, S. A., Gorges, T. M. & Pantel, K. Biology, detection, and clinical implications of circulating tumor cells. *EMBO molecular medicine* **7,** 1–11 (2015).

171. Chudasama, D. *et al.* Liquid Biopsies in Lung Cancer: Four Emerging Technologies and Potential Clinical Applications. *Cancers* **11.** ISSN: 2072-6694 (Mar. 2019).

172. Bailey, M. E., Brown, R. W., Mody, D. R., Cagle, P. & Ramzy, I. Ber-EP4 for differentiating adenocarcinoma from reactive and neoplastic mesothelial cells in serous effusions: Comparison with carcinoembryonic antigen, B72.3 and Leu-M1. *Acta Cytologica* **40,** 1212–1216. ISSN: 19382650 (1996).

173. Adams, D. L. *et al.* Cytometric characterization of Circulating Tumor Cells Captured by microfiltration and their correlation to the cellsearch® CTC test. *Cytometry Part A* **87,** 137–144. ISSN: 1552-4922 (Feb. 2015).

174. Poruk, K. E. *et al.* Circulating tumor cell phenotype predicts recurrence and survival in pancreatic adenocarcinoma. *Annals of Surgery* **264,** 1073–1081. ISSN: 15281140. arXiv: 15334406 (2016).

175. Hager, G. *et al.* The use of a panel of monoclonal antibodies to enrich circulating breast cancer cells facilitates their detection. *Gynecologic Oncology* **98,** 211–216. ISSN: 00908258 (2005).

176. Wei, T. *et al.* Vimentin-positive circulating tumor cells as a biomarker for diagnosis and treatment monitoring in patients with pancreatic cancer. *Cancer Letters.* ISSN: 0304-3835 (2019).

177. Breuninger, S. *et al.* Membrane Hsp70—A Novel Target for the Isolation of Circulating Tumor Cells After Epithelial-to-Mesenchymal Transition. *Frontiers in Oncology* **8,** 1–13. ISSN: 2234-943X (2018).

178. Tsubakihara, Y. & Moustakas, A. Epithelial-Mesenchymal Transition and Metastasis under the Control of Transforming Growth Factor β. *International journal of molecular sciences* **19,** 1–30. ISSN: 14220067 (2018).

179. Papavasiliou, P., Fisher, T., Kuhn, J., Nemunaitis, J. & Lamont, J. Circulating tumor cells in patients undergoing surgery for hepatic metastases from colorectal cancer. *Proceedings (Baylor University. Medical Center)* **23,** 11–14 (Jan. 2010).

180. Vallières, E., Peters, S., Van Houtte, P., Dalal, P. & Lim, E. Therapeutic advances in non-small cell lung cancer. *Thorax* **67,** 1097–1101 (Dec. 2012).

181. Ng, C. S. H., Lau, K. K. W., Gonzalez-Rivas, D. & Rocco, G. *Evolution in surgical approach and techniques for lung cancer.* July 2013.

182. Yan, T. D., Black, D., Bannon, P. G. & McCaughan, B. C. Systematic review and meta-analysis of randomized and nonrandomized trials on safety and efficacy of video-assisted thoracic surgery lobectomy for early-stage non-small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **27,** 2553–2562 (May 2009).

183. Sugi, K., Kaneda, Y. & Esato, K. Video-assisted thoracoscopic lobectomy achieves a satisfactory long-term prognosis in patients with clinical stage IA lung cancer. *World journal of surgery* **24,** 21–27. ISSN: 0364-2313 (Jan. 2000).

184. Krebs, M. G. *et al.* Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **29,** 1556–1563 (Apr. 2011).

185. Hou, J.-M. *et al.* Circulating Tumor Cells as a Window on Metastasis Biology in Lung Cancer. *The American Journal of Pathology* **178,** 989–996. ISSN: 0002-9440 (Mar. 2011).

186. Thiery, J. P. Epithelial–mesenchymal transitions in tumour progression. *Nature Reviews Cancer* **2,** 442–454. ISSN: 1474-1768 (2002).

187. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *The Journal of Clinical Investigation* **119,** 1420–1428. ISSN: 0021-9738 (June 2009).

188. Ogle, L. F. *et al.* Imagestream detection and characterisation of circulating tumour cells – A liquid biopsy for hepatocellular carcinoma? *Journal of Hepatology* **65,** 305–313. ISSN: 16000641 (2016).

189. Leers, M. P. Circulating tumor DNA and their added value in molecular oncology. *Clinical Chemistry and Laboratory Medicine,* 1–10. ISSN: 14374331 (2019).

190. Underhill, H. R. *et al.* Fragment Length of Circulating Tumor DNA. *PLoS Genetics* **12,** 1–24. ISSN: 15537404. eprint: `101219451` (2016).

191. Batzer, M. A. & Deininger, P. L. Alu repeats and human genomic diversity. *Nature Reviews Genetics* **3,** 370–379. ISSN: 14710056 (2002).

192. Rowold, D. J. & Herrera, R. J. Alu elements and the human genome. *Genetica* **108,** 57–72. ISSN: 00166707 (2000).

193. Hao, T. B. *et al.* Circulating cell-free DNA in serum as a biomarker for diagnosis and prognostic prediction of colorectal cancer, 1482–1489 (2014).

194. El-Shazly, S. F., Eid, M. A., El-Sourogy, H. A., Attia, G. F. & Ezzat, S. A. Evaluation of serum DNA integrity as a screening and prognostic tool in patients

with hepatitis C virus-related hepatocellular carcinoma. *International Journal of Biological Markers* **25,** 79–86. ISSN: 03936155 (2010).

195. Zhang, R. *et al.* Clinical value of ALU concentration and integrity index for the early diagnosis of ovarian cancer : A retrospective cohort trial, 1–9 (2018).

196. Vizza, E. *et al.* Serum DNA integrity index as a potential molecular biomarker in endometrial cancer. *Journal of experimental & clinical cancer research : CR* **37,** 16 (Jan. 2018).

197. Umetani, N. *et al.* Prediction of breast tumor progression by integrity of free circulating DNA in serum. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **24,** 4270–4276 (Sept. 2006).

198. Utomo, W. K. *et al.* DNA integrity as biomarker in pancreatic cyst fluid. *American Journal of Cancer Research* **6,** 1837–1841. ISSN: 21566976 (2016).

199. Agostini, M. *et al.* Circulating Cell-Free DNA a promising marker of pathologic tumor response in rectal cancer patients receiving preoperative chemoradiotherapy, 2461–2468 (2011).

200. Liu, Q. *et al.* Targeting interlukin-6 to relieve immunosuppression in tumor microenvironment. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **39,** 1010428317712445. ISSN: 1423-0380 (June 2017).

201. Jin-Wei Miao Li-Jiang Liu, J. H. Interleukin-6-induced epithelial-mesenchymal transition through signal transducer and activator of transcription 3 in human cervical carcinoma. *International journal of oncology* **45,** 165–176. ISSN: 1791-2423 (July 2014).

202. Opp, M. R., Smith, E. M. & Hughes Jr, T. K. Interleukin-10 (cytokine synthesis inhibitory factor) acts in the central nervous system of rats to reduce sleep. *Journal of neuroimmunology* **60,** 165–168. ISSN: 0165-5728 (July 1995).

203. Aste-Amezaga, M., Ma, X., Sartori, A. & Trinchieri, G. Molecular mechanisms of the induction of IL-12 and its inhibition by IL-10. *Journal of immunology (Baltimore, Md. : 1950)* **160,** 5936–5944. ISSN: 0022-1767 (June 1998).

204. Varma, T. K. *et al.* Cellular mechanisms that cause suppressed gamma interferon secretion in endotoxin-tolerant mice. *Infection and immunity* **69,** 5249–5263. ISSN: 0019-9567 (Sept. 2001).

205. Bernabé, R., Hickson, N., Wallace, A. & Blackhall, F. H. What do we need to make circulating tumour DNA (ctDNA) a routine diagnostic test in lung cancer? *European Journal of Cancer* **81,** 66–73. ISSN: 18790852 (2017).

206. Tie, J. *et al.* Circulating Tumor DNA Analyses as Markers of Recurrence Risk and Benefit of Adjuvant Therapy for Stage III Colon Cancer. *JAMA oncology,* 1–8. ISSN: 2374-2445 (2019).

207. Koessler, T., Addeo, A. & Nouspikel, T. *Implementing circulating tumor DNA analysis in a clinical laboratory: A user manual* 1st ed., 1–58. ISBN: 0065-2423 (Elsevier Inc., 2019).

208. Coco, S. *et al.* Circulating cell-free DNA and circulating tumor cells as prognostic and predictive biomarkers in advanced non-small cell lung cancer patients treated with first-line chemotherapy. *International Journal of Molecular Sciences* **18.** ISSN: 14220067 (2017).

209. Kamel, A. M., Teama, S., Fawzy, A. & El Deftar, M. Plasma DNA integrity index as a potential molecular diagnostic marker for breast cancer. *Tumor Biology* **37,** 7565–7572. ISSN: 1423-0380 (2016).

210. Fleischhacker, M. Letters Changes in Concentration of DNA in Serum and Plasma during Storage of Blood Samples, 1028–1029.

211. Holdenrieder, S. *et al.* Clinical Relevance of Circulating Nucleosomes in Cancer. *Annals of the New York Academy of Sciences* **1137,** 180–189. ISSN: 0077-8923 (Aug. 2008).

212. Chiu, T. W., Young, R., Chan, L. Y. S., Burd, A. & Lo, D. Y. M. Plasma cell-free DNA as an indicator of severity of injury in burn patients. English. *Clinical Chemistry and Laboratory Medicine (CCLM)* **44,** 13–17 (2006).

213. Zeerleder, S. *et al.* Elevated nucleosome levels in systemic inflammation and sepsis*. *Critical Care Medicine* **31.** ISSN: 0090-3493 (2003).

214. Lam, N. Time Course of Early and Late Changes in Plasma DNA in Trauma Patients. *Clinical Chemistry* **49,** 1286–1291 (Aug. 2003).

215. Rainer, T. H. *et al.* Plasma β-globin DNA as a prognostic marker in chest pain patients. *Clinica Chimica Acta* **368,** 110–113. ISSN: 0009-8981 (2006).

216. He, Y. *et al.* Perioperative Circulating Tumor DNA in Colorectal Liver Metastases: Concordance with Metastatic Tissue and Predictive Value for Tumor Burden and Prognosis. *Cancer management and research* **12,** 1621–1630. ISSN: 1179-1322 (2020).

217. Henriksen, T. V. *et al.* The effect of surgical trauma on circulating free DNA levels in cancer patients—implications for studies of circulating tumor DNA. *Molecular Oncology* **14,** 1670–1679. ISSN: 18780261 (2020).

218. Sobhani, N., Generali, D., Zanconati, F., Bortul, M. & Scaggiante, B. Cell-free DNA integrity for the monitoring of breast cancer: Future perspectives? *World journal of clinical oncology* **9,** 26–32. ISSN: 2218-4333 (Apr. 2018).

219. Kamel, A. M., Teama, S., Fawzy, A. & El Deftar, M. Plasma DNA integrity index as a potential molecular diagnostic marker for breast cancer. *Tumor Biology* **37,** 7565–7572. ISSN: 14230380 (2016).

220. Yim, A. P., Wan, S., Lee, T. W. & Arifi, A. A. VATS lobectomy reduces cytokine responses compared with conventional surgery. *The Annals of thoracic surgery* **70,** 243–247. ISSN: 0003-4975 (July 2000).

221. Cho, J. M. *et al.* Response of serum cytokines in patients undergoing laparoscopic cholecystectomy. *Surgical endoscopy* **8,** 1380–1384. ISSN: 0930-2794 (Dec. 1994).

222. Ueo, H. *et al.* Minimal increase in serum interleukin-6 levels during laparoscopic cholecystectomy. *American journal of surgery* **168,** 358–360. ISSN: 0002-9610 (Oct. 1994).

223. Karayiannakis, A. J., Makri, G. G., Mantzioka, A., Karousos, D. & Karatzas, G. Systemic stress response after laparoscopic or open cholecystectomy: a randomized trial. *The British journal of surgery* **84,** 467–471. ISSN: 0007-1323 (Apr. 1997).

224. Harmon, G. D., Senagore, A. J., Kilbride, M. J. & Warzynski, M. J. Interleukin-6 response to laparoscopic and open colectomy. *Diseases of the colon and rectum* **37,** 754–759. ISSN: 0012-3706 (Aug. 1994).

225. Ellström, M. *et al.* Evaluation of tissue trauma after laparoscopic and abdominal hysterectomy: measurements of neutrophil activation and release of IL-6, cortisol, and C-reactive protein. *Journal of the American College of Surgeons* **182,** 423–430. ISSN: 1072-7515 (May 1996).

226. Yuen, P. M. *et al.* Metabolic and inflammatory responses after laparoscopic and abdominal hysterectomy. *American journal of obstetrics and gynecology* **179,** 1–5. ISSN: 0002-9378 (July 1998).

227. Sietses, C. *et al.* A prospective randomized study of the systemic immune response after laparoscopic and conventional Nissen fundoplication. *Surgery* **126,** 5–9. ISSN: 0039-6060 (July 1999).

228. Wan, S., LeClerc, J. L. & Vincent, J. L. Cytokine responses to cardiopulmonary bypass: lessons learned from cardiac transplantation. *The Annals of thoracic surgery* **63,** 269–276. ISSN: 0003-4975 (Jan. 1997).

229. Koumarianou, A. *et al.* Prognostic markers in early-stage colorectal cancer: significance of TYMS mRNA expression. *Anticancer research* **34,** 4949–4962 (Sept. 2014).

230. Wang, Y. *et al.* Vimentin expression in circulating tumor cells (CTCs) associated with liver metastases predicts poor progression-free survival in patients with advanced lung cancer. *Journal of cancer research and clinical oncology* **145,** 2911–2920. ISSN: 1432-1335 (Electronic) (Dec. 2019).

231. Tay, Y., Rinn, J. & Pandolfi, P. P. The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505,** 344–352. ISSN: 1476-4687 (Jan. 2014).

232. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489,** 101–108. ISSN: 1476-4687 (Sept. 2012).

233. Nagano, T. & Fraser, P. No-nonsense functions for long noncoding RNAs. *Cell* **145,** 178–181. ISSN: 1097-4172 (Apr. 2011).

234. Delás, M. J. & Hannon, G. J. lncRNAs in development and disease: from functions to mechanisms. *Open biology* **7,** 170121. ISSN: 2046-2441 (July 2017).

235. Zhang, Y.-L. *et al.* The lncRNA XIST exhibits oncogenic properties via regulation of miR-449a and Bcl-2 in human non-small cell lung cancer. *Acta pharmacologica Sinica* **38,** 371–381. ISSN: 1745-7254 (Mar. 2017).

236. Inamura, K. Major Tumor Suppressor and Oncogenic Non-Coding RNAs: Clinical Relevance in Lung Cancer. *Cells* **6,** 12. ISSN: 2073-4409 (May 2017).

237. Do, H. & Kim, W. Roles of Oncogenic Long Non-coding RNAs in Cancer Development. *Genomics & informatics* **16,** e18–e18. ISSN: 1598-866X (Dec. 2018).

238. Du, Z. *et al.* Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. *Nature Communications* **7,** 10982. ISSN: 2041-1723 (2016).

239. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379,** 131–137. ISSN: 0028-0836 (Jan. 1996).

240. Jonkers, I. *et al.* RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. *Cell* **139,** 999–1011. ISSN: 1097-4172 (Nov. 2009).

241. Byerly, J., Halstead-Nussloch, G., Ito, K., Katsyv, I. & Irie, H. Y. PRKCQ promotes oncogenic growth and anoikis resistance of a subset of triple-negative breast cancer cells. *Breast cancer research : BCR* **18,** 95. ISSN: 1465-542X (Sept. 2016).

242. Fukazawa, T. *et al.* SOX2 suppresses CDKN1A to sustain growth of lung squamous cell carcinoma. *Scientific Reports* **6,** 20113. ISSN: 2045-2322 (2016).

243. Fueller, F. & Kubatzky, K. F. The small GTPase RhoH is an atypical regulator of haematopoietic cells. *Cell communication and signaling : CCS* **6,** 6. ISSN: 1478-811X (Sept. 2008).

244. Haga, R. B. & Ridley, A. J. Rho GTPases: Regulation and roles in cancer cell biology. *Small GTPases* **7,** 207–221. ISSN: 2154-1256 (Oct. 2016).

245. Shen, Y., Xia, E., Bhandari, A., Wang, X. & Guo, G. LncRNA PROX1-AS1 promotes proliferation, invasion, and migration in papillary thyroid carcinoma. *Bioscience Reports* **38.** ISSN: 0144-8463 (Sept. 2018).

246. Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349,** 38–44. ISSN: 0028-0836 (Jan. 1991).

247. Zhou, X., Xu, X., Gao, C. & Cui, Y. XIST promote the proliferation and migration of non-small cell lung cancer cells via sponging miR-16 and regulating CDK8 expression. *American journal of translational research* **11,** 6196–6206. ISSN: 1943-8141 (2019).

248. Wang, X. *et al.* Knockdown of LncRNA-XIST Suppresses Proliferation and TGF-β1-Induced EMT in NSCLC Through the Notch-1 Pathway by Regulation of miR-137. *Genetic Testing and Molecular Biomarkers* **22,** 333–342. ISSN: 1945-0265 (2018).

249. Yu, H. *et al.* Knockdown of long non-coding RNA XIST increases blood–tumor barrier permeability and inhibits glioma angiogenesis by targeting miR-137. *Oncogenesis* **6,** e303. ISSN: 2157-9024 (Mar. 2017).

250. Li, C. *et al.* Long non-coding RNA XIST promotes TGF-β-induced epithelial-mesenchymal transition by regulating miR-367/141-ZEB2 axis in non-small-cell lung cancer. *Cancer Letters* **418,** 185–195. ISSN: 18727980 (2018).

251. Fang, J., Sun, C.-C. & Gong, C. Long noncoding RNA XIST acts as an oncogene in non-small cell lung cancer by epigenetically repressing KLF2 expression. *Biochemical and biophysical research communications* **478,** 811–817. ISSN: 1090-2104 (Sept. 2016).

252. Sun, W., Zu, Y., Fu, X. & Deng, Y. Knockdown of lncRNA-XIST enhances the chemosensitivity of NSCLC cells via suppression of autophagy. *Oncology Reports* **38,** 3347–3354. ISSN: 17912431 (2017).

253. Hu, B., Shi, G., Li, Q., Li, W. & Zhou, H. Long noncoding RNA XIST participates in bladder cancer by downregulating p53 via binding to TET1. *Journal of Cellular Biochemistry.* ISSN: 10974644 (2018).

254. Tantai, J., Hu, D., Yang, Y. & Geng, J. Combined identification of long noncoding RNA XIST and HIF1A-AS1 in serum as an effective screening for nonsmall cell lung cancer. *International Journal of Clinical and Experimental Pathology* **8,** 7887–7895. ISSN: 19362625 (2015).

255. Zhou, Q. *et al.* Clinica Chimica Acta Long non coding RNA XIST as a prognostic cancer marker – A meta-analysis. *Clinica Chimica Acta* **482,** 1–7. ISSN: 0009-8981 (2018).

256. Liu, J. L., Zhang, W. Q., Zhao, M. & Huang, M. Y. Upregulation of long noncoding RNA XIST is associated with poor prognosis in human cancers. *Journal of Cellular Physiology,* 1–7. ISSN: 10974652 (2018).

257. Xu, Y., Wang, J. & Wang, J. Long noncoding RNA XIST promotes proliferation and invasion by targeting miR-141 in papillary thyroid carcinoma, 5035–5043 (2018).

258. Huang, R. *et al.* An RNA-Seq Strategy to Detect the Complete Coding and Non-Coding Transcriptome Including Full-Length Imprinted Macro ncRNAs. *PLOS ONE* **6,** e27288 (Nov. 2011).

259. Ilott, N. E. & Ponting, C. P. Predicting long non-coding RNAs using RNA sequencing. *Methods* **63,** 50–59. ISSN: 1046-2023 (2013).

260. Deeds, L., Teodorescu, S., Chu, M., Yu, Q. & Chen, C.-Y. A p53-independent G1 cell cycle checkpoint induced by the suppression of protein kinase C alpha and theta isoforms. *The Journal of biological chemistry* **278,** 39782–39793. ISSN: 0021-9258 (Oct. 2003).

261. Zamagni, A. *et al.* CDKN1A upregulation and cisplatin-pemetrexed resistance in non-small cell lung cancer cells. *International journal of oncology* **56,** 1574–1584. ISSN: 1791-2423 (June 2020).

262. Shoji, T. *et al.* Clinical significance of p21 expression in non-small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **20,** 3865–3871. ISSN: 0732-183X (Sept. 2002).

263. Groeger, A. M. *et al.* Expression of p21 in non small cell lung cancer relationship with PCNA. *Anticancer research* **20,** 3301–3305. ISSN: 0250-7005 (2000).

264. Górka, B. *et al.* NrCAM, a neuronal system cell-adhesion molecule, is induced in papillary thyroid carcinomas. *British journal of cancer* **97,** 531–538. ISSN: 0007-0920 (Aug. 2007).

265. Zhang, Y. *et al.* Positive Feedback Loops Between NrCAM and Major Signaling Pathways Contribute to Thyroid Tumorigenesis. *The Journal of clinical endocrinology and metabolism* **102,** 613–624. ISSN: 1945-7197 (Feb. 2017).

266. Rohrbeck, A. *et al.* Gene expression profiling for molecular distinction and characterization of laser captured primary lung cancers. *Journal of translational medicine* **6,** 69. ISSN: 1479-5876 (Nov. 2008).

267. Conacci-Sorrell, M. *et al.* The shed ectodomain of Nr-CAM stimulates cell proliferation and motility, and confers cell transformation. *Cancer research* **65,** 11605–11612. ISSN: 0008-5472 (Dec. 2005).

268. Tajadura-Ortega, V. *et al.* An RNAi screen of Rho signalling networks identifies RhoH as a regulator of Rac1 in prostate cancer cell migration. *BMC Biology* **16,** 29. ISSN: 1741-7007 (2018).

269. Barr, L. F. *et al.* c-Myc suppresses the tumorigenicity of lung cancer cells and down-regulates vascular endothelial growth factor expression. *Cancer research* **60,** 143–149. ISSN: 0008-5472 (Jan. 2000).

270. Rapp, U. R. *et al.* MYC is a metastasis gene for non-small-cell lung cancer. *PloS one* **4,** e6029. ISSN: 1932-6203 (June 2009).

271. Chen, H., Liu, H. & Qing, G. Targeting oncogenic Myc as a strategy for cancer treatment. *Signal Transduction and Targeted Therapy* **3,** 5. ISSN: 2059-3635 (2018).

272. Soucek, L. *et al.* Inhibition of Myc family proteins eradicates KRas-driven lung cancer in mice. *Genes & development* **27,** 504–513. ISSN: 1549-5477 (Mar. 2013).

273. Shiratsuchi, H. *et al.* Oncogenic Potential of CYP24A1 in Lung Adenocarcinoma. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* **12,** 269–280. ISSN: 1556-1380 (Feb. 2017).

274. Liu, W. *et al.* AKR1B10 (Aldo-keto reductase family 1 B10) promotes brain metastasis of lung cancer cells in a multi-organ microfluidic chip model. *Acta Biomaterialia* **91,** 195–208. ISSN: 1742-7061 (2019).

275. Li, C.-G. *et al.* MicroRNA-1304 suppresses human non-small cell lung cancer cell growth in vitro by targeting heme oxygenase-1. *Acta pharmacologica Sinica* **38,** 110–119. ISSN: 1745-7254 (Jan. 2017).

276. Ma, G. *et al.* Mining and validating the expression pattern and prognostic value of acetylcholine receptors in non-small cell lung cancer. *Medicine* **98,** e15555. ISSN: 1536-5964 (May 2019).

277. Salgia, R. MET in Lung Cancer: Biomarker Selection Based on Scientific Rationale. *Molecular cancer therapeutics* **16,** 555–565. ISSN: 1538-8514 (Apr. 2017).

278. Morgillo, F. *et al.* Dual MET and SMO Negative Modulators Overcome Resistance to EGFR Inhibitors in Human Nonsmall Cell Lung Cancer. *Journal of medicinal chemistry* **60,** 7447–7458. ISSN: 1520-4804 (Sept. 2017).

279. Wang, W. *et al.* Src Promotes Metastasis of Human Non-Small Cell Lung Cancer Cells through Fn14-Mediated NF-κB Signaling. *Medical science monitor : international medical journal of experimental and clinical research* **24,** 1282–1294. ISSN: 1643-3750 (Mar. 2018).

280. Wang, J., Hu, Z.-G., Li, D., Xu, J.-X. & Zeng, Z.-G. Gene expression and prognosis of insulin-like growth factor-binding protein family members in non-small cell lung cancer. *Oncology reports* **42,** 1981–1995. ISSN: 1791-2431 (Nov. 2019).

281. Bi, M. M., Shang, B., Wang, Z. & Chen, G. Expression of CXCR4 and VEGF-C is correlated with lymph node metastasis in non-small cell lung cancer. *Thoracic cancer* **8,** 634–641. ISSN: 1759-7714 (Nov. 2017).

282. Wang, L. *et al.* Clinical evaluation and therapeutic monitoring value of serum tumor markers in lung cancer. *The International journal of biological markers* **31,** e80–7. ISSN: 1724-6008 (Feb. 2016).

283. Jiang, W. *et al.* PTK7 expression is associated with lymph node metastasis, ALK and EGFR mutations in lung adenocarcinomas. *Histology and histopathology* **35,** 489–495. ISSN: 1699-5848 (May 2020).

284. Wang, H. Q. & Wang, J. *Expression of pleiotrophin in small cell lung cancer.* 2015.

285. Jen, J. & Wang, Y.-C. Zinc finger proteins in cancer progression. *Journal of biomedical science* **23,** 53. ISSN: 1423-0127 (July 2016).

286. Lustberg, M. B., Stover, D. G. & Chalmers, J. J. Implementing Liquid Biopsies in Clinical Trials: State of Affairs, Opportunities, and Challenges. *Cancer journal (Sudbury, Mass.)* **24,** 61–64. ISSN: 1540-336X (2018).

287. Stott, S. L. *et al.* Isolation of circulating tumor cells using a microvortex-generating herringbone-chip. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 18392–18397. ISSN: 1091-6490 (Oct. 2010).

288. Mikolajczyk, S. D. *et al.* Detection of EpCAM-Negative and Cytokeratin-Negative Circulating Tumor Cells in Peripheral Blood. *Journal of oncology* **2011,** 252361. ISSN: 1687-8469 (2011).

289. Chinen, L. T. D. *et al.* Cytokeratin-based CTC counting unrelated to clinical follow up. *Journal of thoracic disease* **5,** 593–599. ISSN: 2072-1439 (Oct. 2013).

290. Damani, S. *et al.* Characterization of circulating endothelial cells in acute myocardial infarction. *Science translational medicine* **4,** 126ra33 (Mar. 2012).

291. Gansmo, L. B. *et al.* Influence of MDM2 SNP309 and SNP285 status on the risk of cancer in the breast, prostate, lung and colon. *International journal of cancer* **137,** 96–103. ISSN: 1097-0215 (July 2015).

292. Rotger, M., Csajka, C. & Telenti, A. Genetic, ethnic, and gender differences in the pharmacokinetics of antiretroviral agents. *Current HIV/AIDS reports* **3,** 118–125. ISSN: 1548-3568 (Sept. 2006).

293. Lindsay, C. R. *et al.* Vimentin and Ki67 expression in circulating tumour cells derived from castrate-resistant prostate cancer. *BMC Cancer* **16,** 168. ISSN: 1471-2407 (2016).