Department of Molecular and Clinical Cancer Medicine

DEREGULATION OF LONG NON-CODING RNAs IN NON-SMALL CELL LUNG CANCER

Thesis submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy

Bubaraye Robson Uko

University of Liverpool, Liverpool

August 2020

Supervisors: Dr. Triantafillos Liloglou, Dr. Russell Hyde & Prof. John K. Field

DEDICATION

I owe immense gratitude to God and dedicate this work to the memory of my late mum, Mrs. Diweri Kosuowei. Thanks to my husband, Udeme Robson who has never stopped believing in me, my adorable children – Emmanuel, Margaret, Rhoda and Ethan (who have travelled this road with me, bearing so many inconveniences just for me to achieve this feat) and finally, thanks to my friends, dad and siblings for all their encouragement.

ACKNOWLEDGEMENT

I am most grateful to my supervisors, Lakis Liloglou, Russell Hyde and John Field for their support, guidance, understanding and encouragement throughout the duration of my studies and onto the completion of my thesis. I deeply appreciate the patience and support of Dr. Lakis Liloglou through the course of my research.

I will not fail to appreciate all the staff, colleagues, collaborators and friends who have been of tremendous support to me. Thanks to Dr. Amelia Acha-Sagredo- for her immense and untiring contribution and support, Dr. Michael P.A. Davies- for his invaluable ideas and contributions during presentations, Dr. Michael W. Marcus for rendering guidance on questions relating to my statistical analyses, Dr. Ahmed Al-khafaji, Dr. Israa Al-Humairi, Ms Ghaliah Alnefaie, Mr Ben Brown, Ms. Paschalia Pantazi, Ms. EfpraxiaTzortzini, Ms. Emma Price, Ms. Chryssanthi Moschandrea, Ms. Yakinthi Chrysochoidou, Mrs Stephanie Tate, Mr Nial Hodge, Ms. Patricia Gerard, Dr. Matthew Agwae, Dr. Thompson Gana, International Advice & Guidance Team, ITM post graduate team and Professor Andrea Varro for her support.

Finally, I would like to acknowledge the Niger Delta Development Commission (NDDC), Nigeria and the Roy Castle Lung Cancer Foundation, UK for funding this project.

STATEMENT OF ORIGINALITY:

All the work presented in this thesis, unless otherwise stated, is the work of the author.

ABSTRACT

Lung cancer is the second most common cancer and the leading cause of cancer mortality in the United Kingdom. Non-small cell lung cancer (NSCLC) accounts for over 80% of lung neoplasms. The high mortality is largely due to detection of the disease at advanced stages and resistance to drugs used in the treatment of the disease. Suitable biomarkers (diagnostic, prognostic and therapeutics) are therefore needed for the management of the disease.

Long non-coding RNAs (lncRNAs) are involved in the regulation of numerous biological processes and have been implicated in the pathogenesis of many human diseases such as cancer. While the list of deregulated lncRNAs in NSCLC is increasing, most studies lack validation. In addition, the mechanisms involved in deregulation in most of these lncRNAs still needs to be elucidated. This thesis was therefore aimed at identifying and validating lncRNAs that are deregulated in NSCLC, investigating the epigenetic causes of their deregulation and analysing the functional contribution to development of NSCLC and effect on cell phenotype and drug resistance.

Following microarray screening of paired tumour and normal surgical tissues from 44 NSCLC patients, 20 induced and 30 repressed lncRNAs were discovered in tumours compared to normal tissues. Top 13 of the 50 lncRNAs (*FEZF1-AS1*, *LINC01214*, *LOC105376287, PCAT6, LOC101927229, LINC00673, NUTM2A-AS1, RNF139-AS1, LANCL1-AS1, FENDRR, LINC00968, SVIL-AS1*, and *PCAT19*) were selected for validation, following careful inspection of the annotation in databases and sequence verification in relation to the microarray probes used for each gene. Validation by qPCR was performed in sample pairs from 29/44 microarray patients (technical validation) and 38 independent patients. With the exception of

LOC105376287 and *RNF139*, the remaining 11 targets were successfully validated in both groups. Levels of all validated lncRNAs were independent of patient age and histology, however; higher expression of *FEZF1-AS1* was observed in males compared to females (Mann Whitney's Test, *p=0.02*). Variable expression levels were observed for all validated lncRNAs in NSCLC cell lines and human bronchial epithelial cell line and its isogenic derivatives. Furthermore, the epigenetic influence on lncRNA deregulation study showed significant hypermethylation of the promoter regions of *FEZF1-AS1* and *FENDRR,* and hypomethylation of *NUTM2A-AS1* and *LINC00673*.

LINC00968 was selected for exogenous overexpression in NSCLC cell line A549. Some of the derived clones demonstrated lower proliferation, however, not proportional to the level of *LINC00968* expression. Suppression of migration and invasion was demonstrated for the clone with maximum expression of *LINC00968*, and anchorage independent assay showed a reduction in cell colony count for maximum expressing clone. *LINC00968* overexpression did not correlate with resistance to 5FU, gemcitabine, cisplatin and vinorelbine, on the contrary, the presence of *LINC00968* increased cell viability in methotrexate and pemetrexed treated A549 cells and sensitized the cells to the effects of hydrogen peroxide in an expression dependent manner.

In conclusion, this study identified previously uncharacterized lncRNAs that are deregulated in NSCLC, showed that epigenetic modulation of lncRNAs is an important reason for their deregulation in NSCLC in a direct or indirect manner and that *LINC00968* is involved in tumour suppression and may be a potential prognostic marker or therapeutic target.

Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

5mC	5-methylcytosine
ALK	Anaplastic Lymphoma Kinase
APS	Adenosine 5'-phosphosulfate
ATP	Adenosine Triphosphate
CBX7	Chromobox 7
CCD	Charge-coupled Device
cDNA	Complementary DNA
CoREST	Co-repressor for element-1-silencing transcription factor
CpG	Cytosine Guanosine dinucleotide
CT	Computer tomography
DAC	Decitabine
DDP	Diamminedichloride platinum (Cisplatin)
DEPC	Diethyl pyrocarbonate
DMEM	Dulbecco's Modified Eagles Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNMTs	DNA Methyltransferases
DOCK 8	Dedicator of cytokinesis 8
dsRNA	Double stranded RNA
EGFR	Epidermal Growth Factor Receptor
EZH2	Enhancer of zeste homolog 2
FBS	Fetal Bovine Serine
FDA	Food and Drug Administration
FGFR2	Fibroblast growth receptor 2
GWAS	Genome-wide Association Studies
H and E	Hematoxylin and Eosin
HBEC	Human Bronchial Epithelial Cell
hnRNP-K	Heterogenous nuclear ribonucleoprotein K
HOXD	Homeobox D
KDM2A	Lysine demethylase 2A
LADC	Lung Adenocarcinoma
LCC	Large Cell Carcinoma
lncRNA	Long non-coding RNA
LSCC	Lung Squamous Cell Carcinoma
miRNA	Micro RNA
MRI	Magnetic Resonance Imaging
MTT	3-(4, 5-dimethylthiazol)-2-yl)-2, 5-diphenyltetrazolium bromide
NATs	Natural Antisense Transcripts
ncRNA	Non-coding RNA
NSCLC	Non-Small Cell Lung Cancer
ORF	Open Reading Frames
PBS	Phosphate Buffered Saline

Pen-Strep	Penicillin and streptomycin
PET	Positron Emission Tomography
piRNA	PIWI interacting RNA
PRC	Polycomb Repressive Complex
PTEN	Phosphatase and tensin homolog
PTGS	Post Transcriptional Gene Silencing
PVT1	Plasmacytoma variant translocation 1
qPCR	Quantitative Polymerase Chain Reaction
RFA	Radio Frequency Ablation
RISC	RNA-induced silencing complex
RNA	Ribonucleic Acid
RNAi	RNA interference
RQ	Relative Quantitation
rRNA	Ribosomal RNA
RTEs	Repetitive Transposable Elements
SAM	S-adenosyl-L-methionine
SBRT	Stereotactic Body Radiation Therapy
SCLC	Small Cell Lung Cancer
scRNA	Small cytoplasmic RNA
SFM	Serine Free Medium
siRNA	Small interfering RNA
SNIP1	Small nuclear-interacting protein 1
snoRNA	Small nucleolar RNA
SNP	Single Nucleotide Polymorphism
snRNA	Small nuclear RNA
SOC Media	Super optimal broth with catabolite repression
STR	Short Tandem Repeat
TBP	TATA box binding protein
TCGA	The Cancer Genome Atlas
TFs	Transcription Factors
TNBC	Triple Negative Breast Cancer
TNM	Tumour Node Metastasis
tRNA	Transfer RNA
UCR	Ultraconserved genomic region
VPA	Valproic Acid

1. INTRODUCTION

1.1 Lung Cancer: Incidence and Mortality

Lung cancer is a complex disease of the lung that arises from uncontrolled proliferation of abnormal cells that line the lungs and is associated with various genetic, epigenetic and cytological changes (Giaconne, 2012). It is the third most common cancer, following breast cancer in women and prostate cancer in men; and is the leading cause of all cancer mortalities in the United Kingdom (UK) among males and females [Figures 1.1 and 1.2] (Cancer Research, UK 2016a). It has continued to be associated with the highest mortality rates in the UK accounting for 22% and 19.4% of cancer deaths in the UK and Worldwide respectively (Cancer Research, UK 2016b). It is regarded largely as a disease of the elderly as over 60% of lung cancer cases are diagnosed in patients aged over 70 years (Ferlay, *et al*., 2013). The lack of specificity of symptoms of this disease has made early detection difficult, the 5-year survival rate has improved slightly to 15.9% after over four decades compared to 65%, 90% and 99% observed in colon, breast and prostate cancer respectively (Chheang & Brown, 2013; Chen *et al.,* 2014a). Advances in high throughput technology, diagnosis, therapy and management have made only minimal improvements in overall survival (Giaconne,. 2012).

1.2 Symptoms

Although symptoms of lung cancer are not specific, initial symptoms observed include: persistent cough, shortness of breath, haemoptysis (coughing blood in the sputum), pain in the chest and ribs, loss of appetite, weight loss and extreme tiredness. These symptoms are followed by hoarseness, difficulty in swallowing, swollen lymph nodes. in advanced stages of the disease (Bradley *et al.,* 2019).

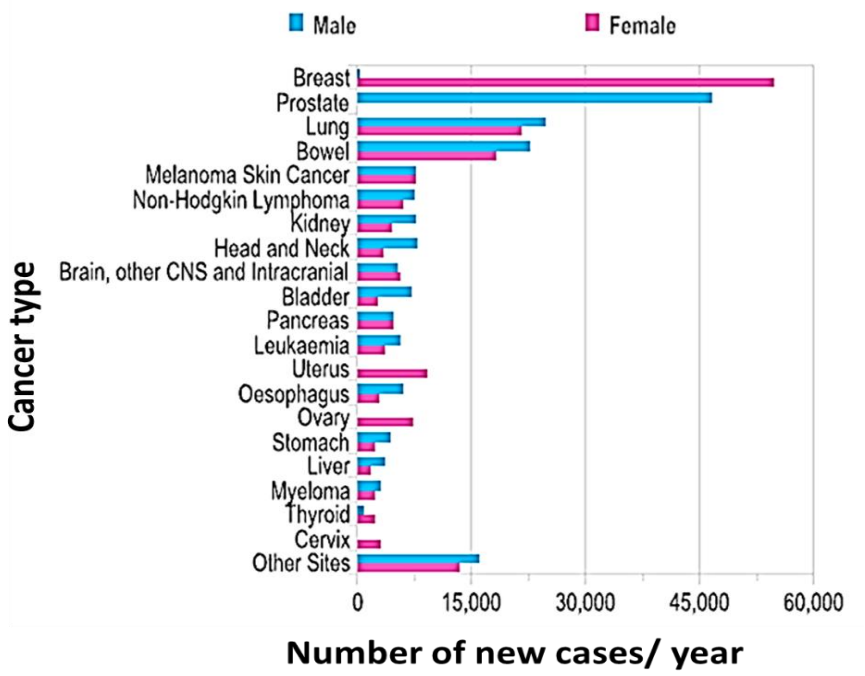


Figure 1.1: Incidence of Major Cancer Types, including Lung in the UK (www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/incidence).

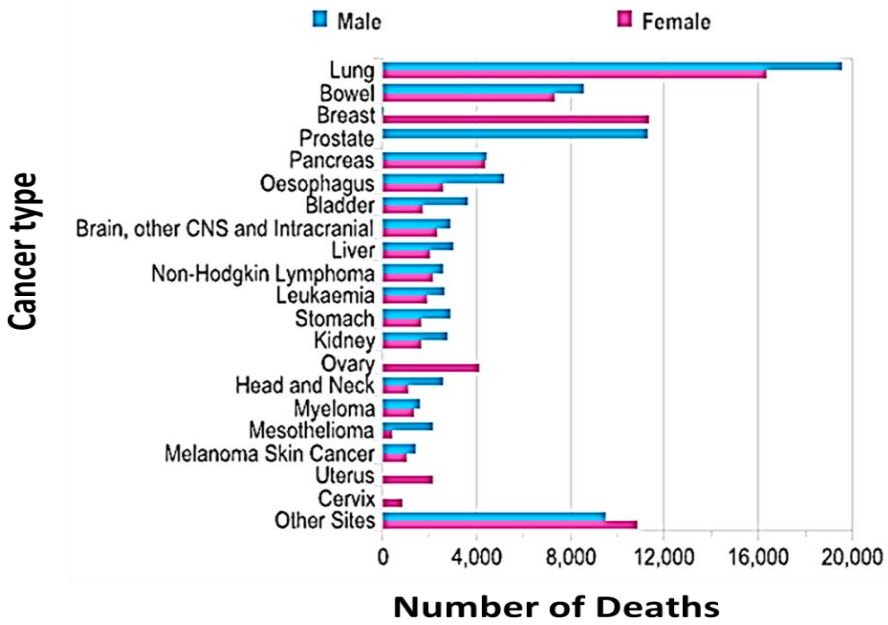


Figure 1.2: Lung cancer mortality rates in comparison to other cancers (www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/mortality)

1.3 Histological classification

Lung cancer is generally classified into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), based on the differences observed in their clinical characteristics (which includes, the high response rate of SCLC to first-line chemotherapy, its early progression and poor prognosis). SCLC and NSCLC account for approximately 15% and 85% respectively of all lung cancer cases worldwide (Gazdar *et al*., 2017; Herbst *et al.,* 2018).

1.3.1 Small Cell Lung Cancer (SCLC)

SCLC comprises lung cancer cells that are relatively small, contain dense neurosecretory granules and are very aggressive lung tumours (Figure 1.3). They are highly sensitive to chemotherapy at the initial stage of the disease (Abidin *et al.,* 2010; Zhang & He, 2013a; Gazdar *et al*., 2017) and are also known as oat cell carcinomas, small cell anaplastic carcinomas or undifferentiated small cell carcinomas. SCLC is of neuroendocrine origin (i.e. they originate from the nerve cells or hormone producing cells of the lung), and therefore typically begin in the bronchi, at the centre of the chest (Park *et al*., 2011). SCLC is predominantly caused by smoking, and reports have associated increased amounts of cigarettes smoked per day and the duration of smoking to the risk of developing SCLC (Morabito *et al*., 2014).

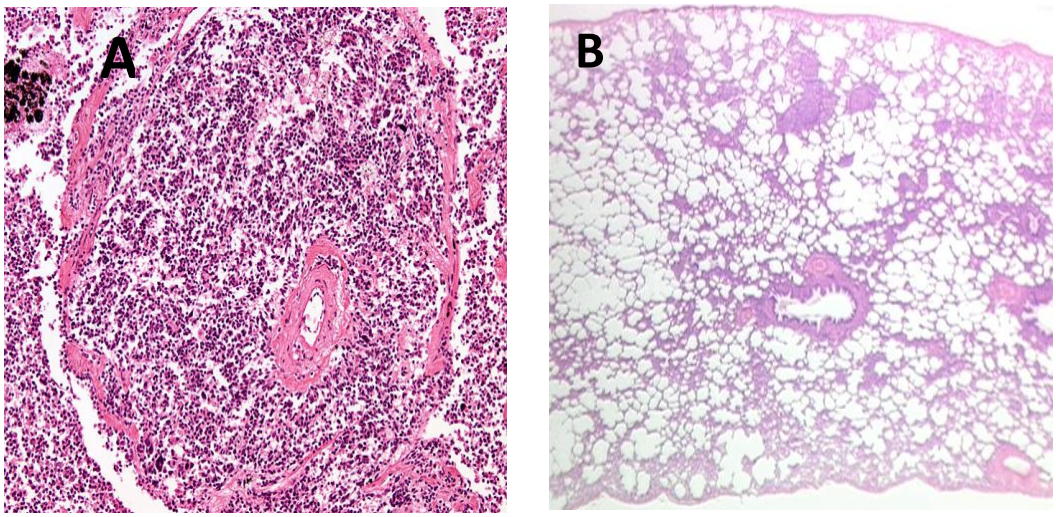


Figure 1.3: H & E staining of (A) small cell lung cancer at 10X magnification compared to (B) Normal healthy lung cell at 10X magnification. While the small cell lung cancer is characterised by oat-shaped malignant cells, the normal lung cells have visible alveoli (www.microscopyu.com and www.bio.rutgers.edu).

1.3.2 Non-Small Cell Lung Cancer (NSCLC)

Although there are several types of NSCLC based on histological classifications, the three most common are lung adenocarcinoma (LADC), lung squamous cell carcinoma (LSCC) and large cell carcinoma (LCC) (Girard *et al*., 2016).

Lung Adenocarcinoma

LADC makes up 38.5% of lung cancer cases and is a more common type of primary lung cancer (Sardenberg *et al*., 2013). It arises from the peripheral lung tissues, appears glandular and is characterized by production of mucin (Figure 1.4). LADC has been associated with early and current smokers and is the only type of NSCLC associated with never-smokers. Presently a lot of emphasis is been placed on LADC for biomarker testing because of its prevalence and its association with Epidermal Growth Factor Receptor (*EGFR*) mutation, Kirsten rat sarcoma (*KRAS*) mutation and Anaplastic lymphoma kinase (ALK) rearrangement (Brandao *et al*., 2012; Dong *et al*., 2016; Herbst *et al*., 2018).

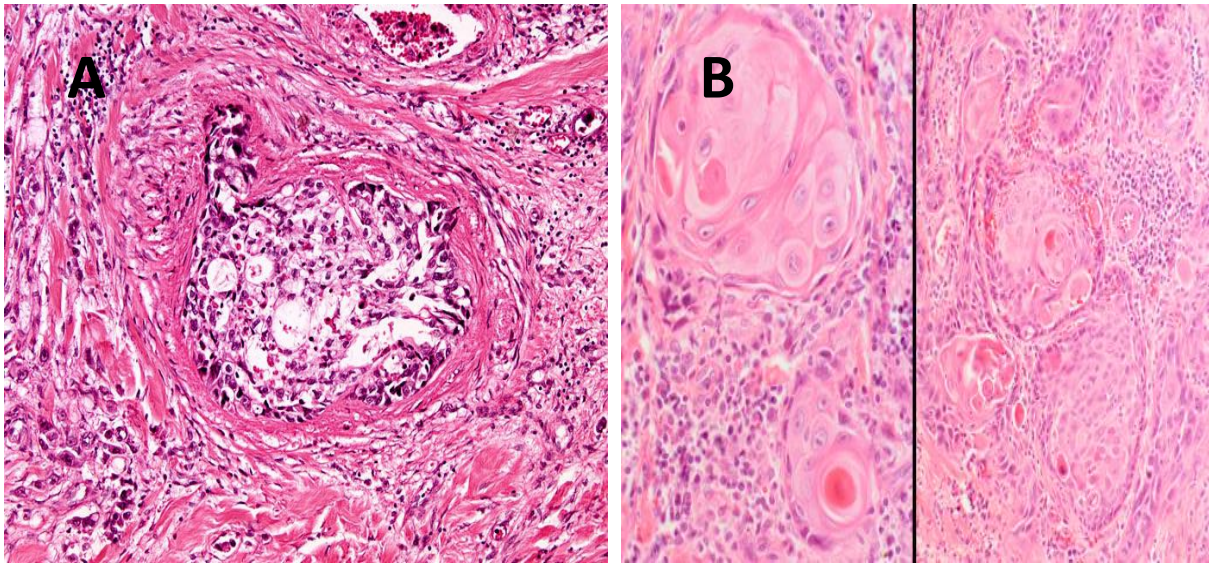


Figure 1.4: Representative H & E staining of (A.) 10x magnification of lung adenocarcinoma and (B.) 10X magnification of well-differentiated lung squamous cell carcinoma (www.microscopyu.com)

Lung Squamous cell carcinoma

LSCC also known as epidermoid carcinoma accounts for about 25% to 30% of all lung cancer cases and often originates from the squamous cells of the central part of the lung or either bronchi. It is majorly associated with smoking (Herbst *et al*., 2018).

Large cell carcinoma

LCC is most often diagnosed by exclusion, that is, if the tumour cells lack the phenotypic characteristics of LADC or LSCC, or the respective biomarkers associated with either lung LADC or LSCC (Chen *et al*., 2014) or from SCLC by its size. They make up approximately 5 - 10% of lung cancer cases and can originate from any part of the lung (Zappa and Mousa, 2016).

1.4 Risk factors

1.4.1 Smoking

There are several well-established factors responsible for the development of lung cancer; however, the single major cause of this disease is tobacco smoking. About 80% of lung cancer deaths are attributed to smoking (Furrukh, 2013; WHO, 2016). There are over 5000 chemicals and 60 known carcinogens in tobacco smoke, these chemicals like nicotine, tar, formaldehyde, nitrosamines, arsenic, hydrogen cyanide, polycyclic aromatic hydrocarbons, uranium etc. penetrate lung tissues, produce intermediates that react with deoxyribonucleic acid (DNA) bases or bind to specific receptors and cause DNA damage or stimulate signalling pathways that increase cell proliferation and survival (Sanner & Grimsrud, 2015). Although the rate of smoking in both males and females has significantly reduced in comparison to the 1970s, the gender gap has significantly narrowed due to increased rate of women smokers (Figure 1.5).

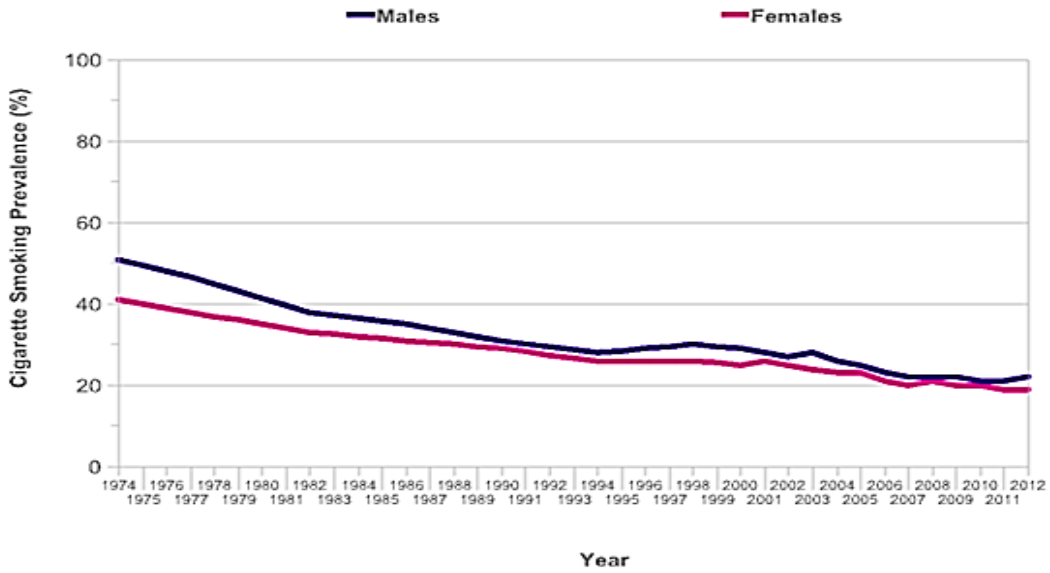


Figure 1.5: Prevalence of smoking between males and females in the UK between 1974 and 2012 (www.cancerresearchuk.org/health-professional/cancer-statistics/risk/tobacco).

1.4.2 Environmental Factors

Other factors are environmental, which includes exposure to second-hand smoke, residential radon, radiation therapy and occupational exposure to chemicals like asbestos, arsenic, coal products, sulphur mustard (mustard gas), silica, chromium etc. (Straif *et al*., 2009; Field & Withers, 2012; Spyratos *et al*., 2013).

1.4.3 Genetic Factors

Genetic predisposition to lung cancer is also another risk factor (Yang *et al*., 2013). Several studies have shown family history of lung cancer in lung cancer patients. Genome-wide association studies (GWAS) have been able to associate lung cancer development in smokers to their susceptibility to the disease, they have also reported an increased risk of development of the disease in first degree relatives (Thorgeirsson, 2008; Kanwal *et al*., 2017).

Through GWAS, chromosomal regions associated with lung cancer risk in smokers and never-smokers have been identified. Amos *et al*., in their study performed in Caucasian smoking population demonstrated significant association between single nucleotide polymorphisms (SNPs) in the 15q, 5p and 6p chromosomal regions and susceptibility to lung cancer (Amos *et al*., 2008). In an independent study of Asian smoking population, similar association was observed at the same loci (Hung *et al*., 2008). These regions (15q25.1, 5p15.33 and 6p21.33) contain *CHRNA5*/ *CHRNA3*/*CHRNB4*, *TERT*/*CLPTM1L* and *BAG6* (previously known as *BAT3*/*M5H5*) genes respectively, which have been shown to be strongly associated with smoking quantity, nicotine dependence and lung cancer. *CHRNA/ CHRNB* (cholinergic receptor nicotinic alpha and beta) genes are nicotinic acetylcholine receptor subunits that can be stimulated by nitrosamines resulting in tumour growth (Wu *et*

al., 2014a; Bossé and Amos, 2018). Interestingly, the 5p15.33 susceptibility locus has been associated with lung adenocarcinoma, while 6p21 locus was shown to be strongly associated with lung squamous carcinoma (Timofeeva *et al*., 2012). Li *et al*., identified a SNP associated with neversmokers at the 13q31.3 chromosomal region which harbours the *GPC5* (Glypican Proteoglycan 5) gene. *GPC5* is thought to contribute to lung cancer development by regulating certain signalling pathways (Wnt, Hedgehogs and Fibroblast growth factors) (Li *et al*., 2010).

1.5 Diagnosis, Staging and Management

Early diagnosis and accurate staging are necessary for the administration of appropriate therapy and the estimation of disease outcome, hence the continued development of strategies for early detection and the review of the tumour node metastasis (TNM) staging system.

1.5.1 Diagnosis

Initial screening for lung cancer is determined by carrying out a chest x-ray, low dose CT (computer tomography), an MRI (magnetic resonance imaging) or a PET (positron emission tomography) scan. Other screening methods include, bronchoscopy and liquid biopsies. (Silvestri *et al*., 2013; Blandin Knight *et al*., 2017).

1.5.2 Lung Cancer Staging: TNM Staging

The TNM staging system for NSCLC is a staging system internationally accepted for the determination of the disease stage, which is measured by the extent to which the disease has spread. The latest TNM staging (8th edition) developed by the International Association for the Study of Lung Cancer (IASLC) and approved by the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC) classifies NSCLC staging

into five major prognostic stages and these include 0, I, II, III and IV based on the absence/ presence of a tumour, nodal infiltration and the presence or absence of metastasis. And a further sub-classification to indicate the size of the tumour or the extent of metastasis (Detterbeck *et al*., 2017).

Determination of T staging is based on the size of the primary tumour in long axis or its extent into adjoining structures like the chest wall, the T0 descriptor indicates absence of tumour to T4 where the tumour is > 7cm. Nodal staging considers the degree of spread to the surrounding lymph nodes, while the M staging describes the presence of metastases beyond the surrounding lymph nodes (Table 1.1 and Figure 1.6 to 1.8).

Table 1.1: TNM Classification of lung cancer, showing the different stages of the disease adapted from (Detterbeck et al. 2017)

Anatomical Staging and Prognostic Groups			
Occult Carcinoma	**TX**	**N0**	**M0**
Stage 0	Tis	N0	M0
Stage IA	T1a (≤ 1cm)	N0	M0
	T1b (> 1cm ≤ 2cm)	N0	M0
	T1c (> 2cm ≤ 3cm)	N0	M0
Stage IB	T2a (> 3cm ≤ 4cm)	N0	M0
Stage IIA	T2b (> 4cm ≤ 5cm)	N0	M0
	T3	N0	M0
Stage IIB	T1a/ T1b/ T1c	N1	M0
	T2a/ T2b	N1	M0
Stage IIIA	T1a/ T1b/ T1c	N2	M0
	T2a/ T2b	N2	M0
	T3 (> 5cm ≤ 7cm)	N1	M0
	T4 (≥7cm)	N0	M0
	T4	N1	M0
Stage IIIB	T1a/ T1b/ T1c	N3	M0
	T2a/ T2b	N3	M0
	T3	N2	M0
	T4	N1	M0
Stage IIIC	T3/ T4	N3	M0
Stage IVA	Any T	Any N	M1a/M1b
Stage IVB	Any T	Any N	M1c (Multi)

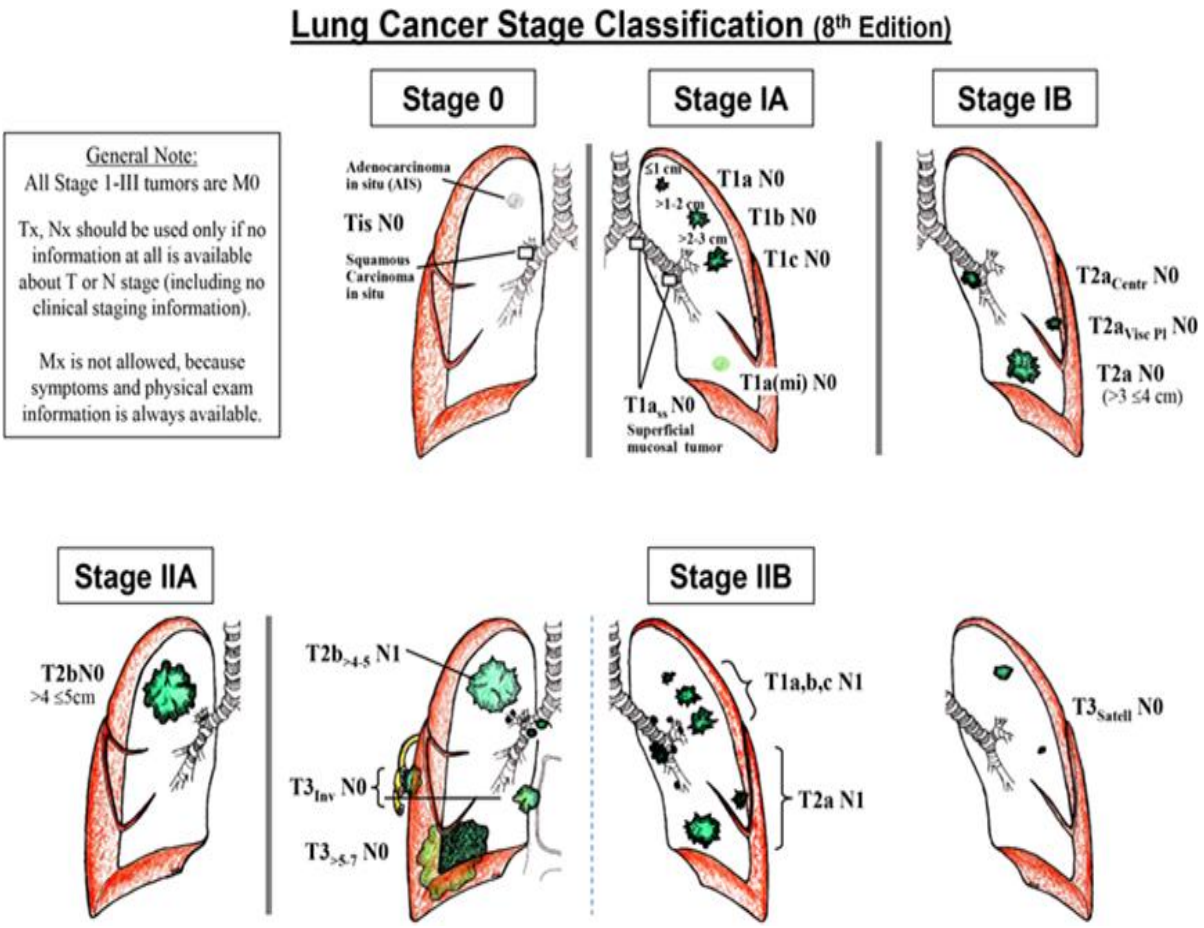


Figure 1-6: Graphical representation of the TNM staging showing stages 0 to IIB. The presence of non-invasive adenocarcinoma and squamous carcinoma without nodal infiltration can be observed in stage 0, while increase in tumour size up to 7 cm with/without nodal infiltration is classified under stages IA to IIB. At these stages tumours are M0 (Detterbeck *et al.*, 2017).

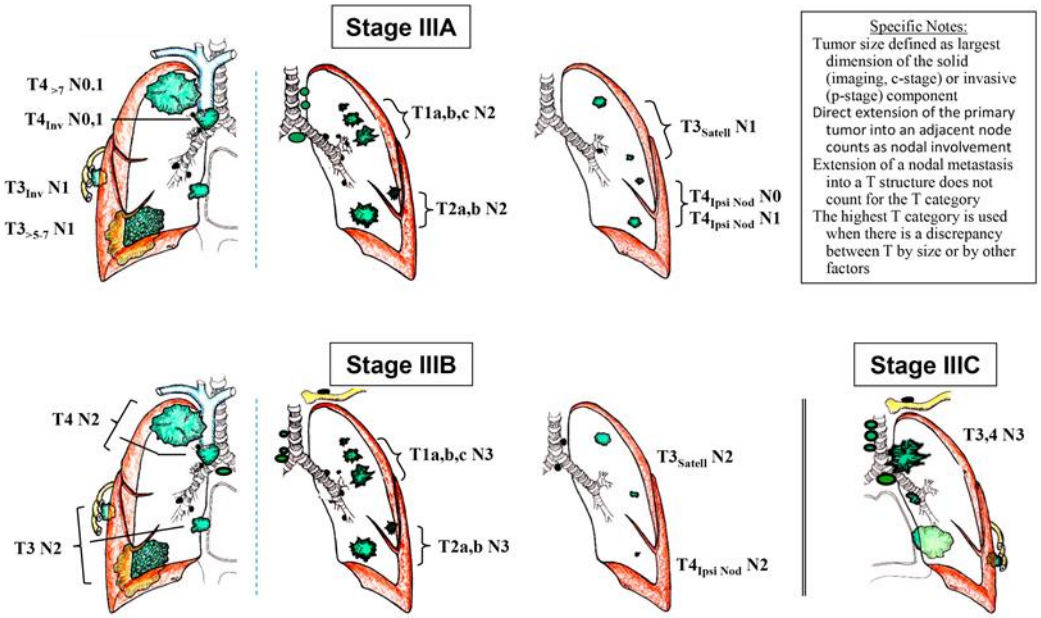


Figure 1-7: Illustration of stages IIIA to IIIC classification of NSCLC. Stage IIIA to IIIC is characterised by large tumours greater than 7 cm associated with nodal infiltration (including spread to the mediastinum, diaphragm, heart, great vessels of the heart, trachea and ipsilateral peribronchial, hilar, mediastinal nodes) Detterbeck *et al.,* 2017.

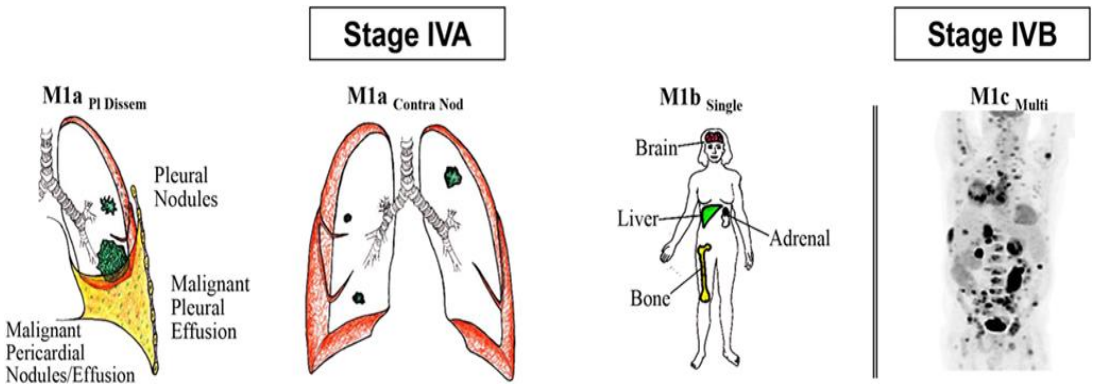


Figure 1-8: Stages IVA and IVB in the TNM staging as shown above is characterised by metasis which is classed into the regional metastatic disease (M1a) characterised by presence of tumour in both lungs or pericardial pleural effusion, the solitary extrathoracic disease (M1b) where a single organ outside the thoracic region (viz the brain, liver, bone, adrenal and liver) is involved and the multiple extrathoracic disease (M1c) where multiple organs have been affected .

1.5.3 Management

The management employed in the treatment of NSCLC patients depends on the stage of the disease. Treatment options include surgical resection, radiofrequency ablation (RFA), radiotherapy, chemotherapy, targeted therapy and immunotherapy.

Surgical resection is the standard treatment option for stages I and II NSCLC, when the tumour is still localised, however, most NSCLC patients are diagnosed at a later stage when surgical resection alone cannot cure the disease (Lackey *et al*., 2013; Van Schil *et al*., 2013).

RFA (an imaging guided technique) is a minimally invasive procedure also used for the treatment of NSCLC at the early stages of NSCLC in patients unable to undergo surgery (due to poor performance status); it requires the use of high frequency electrical currents producing heat through electrodes to destroy small cancer cells which are less than 2 cm in diameter (or <3cm as recommended by the American College of Chest Surgeons). The 1-, 3-, and 5-year survival rates after RFA therapy were reported to be 97.7%, 72.9% and 55.7% respectively (Dupuy *et al*., 2000; Howington *et al*., 2013; de Baere *et al*., 2013; Hiraki *et al*., 2014).

Radical radiotherapy is considered in early stage treatment of NSCLC in case of comorbidity in some patients or as an adjunct therapy in locally advanced disease; it can also be administered as part of a palliative measure in metastasized disease. Although there are different types of radiation therapy, advanced radiation techniques such as the stereotactic body radiation therapy (SBRT) (which involves the precise administration of high radiation doses to lung tumours in short durations as possible) have been shown to be a very effective treatment and attains similar results as surgery with minimal toxicity and damage to healthy lung tissues (Amini *et al*., 2010; Zheng *et al*., 2014).

Chemotherapy is used for patients with locally advanced NSCLC as well as in patients with metastatic NSCLC but having good performance status. Prior to the late 1990s, the singular line of treatment for lung cancer was the use of platinum-based combination chemotherapy (Cosaert *et al*., 2013) Cisplatin or carboplatin with a non-platinum based drug have been a

promising and well-established two-drug combination, they are still the first-line treatment for most patients with advanced stage (IIIB and IV) NSCLC (Rossi & Di Maio. 2016). However, current therapeutic interventions consider the different histologic sub types of NSCLC and the involvement of driver mutations if known, this kind of treatment is referred to as targeted therapy. Typical examples are the epidermal growth factor receptor (*EGFR*) tyrosine kinase inhibitors used as first-line therapy for patients with advanced lung adenocarcinomas with *EGFR* mutations and Crizotinib which has been approved by the US Food and Drug Administration (FDA) for advanced adenocarcinomas with anaplastic lymphoma kinase (*ALK*) rearrangements (Travis et al. 2012; Chan & Hughes, 2015).

1.6 Epigenetics and Epigenetic Biomarkers

In recent years, epigenetics has emerged as a major concept in the understanding of normal developmental processes as well as disease development and progression (Liloglou *et al*., 2014). Lung cancer, much like any other cancer, is now known to develop through a series of genetic and epigenetic alterations resulting in qualitative changes in the function of many genes (Langevin *et al*., 2015).

Conrad Waddington first used the term epigenetics and defined it “as change in phenotype without change in genotype” (Waddington, 2012; Goldberg *et al*., 2007). Today, epigenetics is defined as the study of stable heritable alterations in gene expression without corresponding change in the DNA sequence (Jaenisch & Bird, 2003). Epigenetic mechanisms like DNA methylation, histone modification, chromatin remodelling and non-coding RNAs have been shown to control gene expression patterns (Liloglou *et al*., 2014).

1.6.1 DNA Methylation

DNA methylation is an epigenetic mechanism that involves the covalent addition of a methyl group from S-adenosyl-L-methionine (SAM) to the fifth carbon atom of a cytosine ring of DNA resulting in the formation of 5-methylcytosine (5mC). This process is catalysed by DNA methyltransferases (DNMTs) encoded in many genomes of bacteria, plants and mammals; and the conservation of these enzymes is an indication of the importance of DNA methylation in the organism (Figure 1.9) (Jin *et al*., 2011; Zhu *et al*., 2016; Zhang and Xu, 2017).

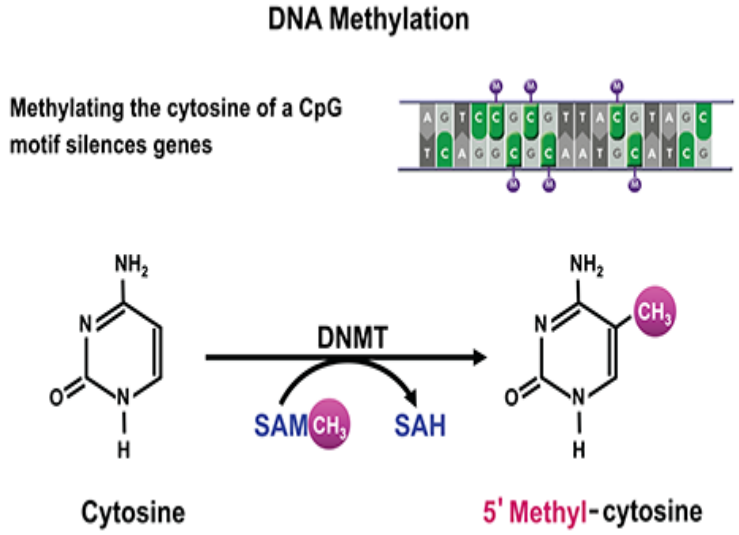


Figure 1-9: Chemical illustration of DNA methylation at the 5 position of cytosine catalysed by DNMT in the presence of SAM (Zakhari, 2013).

As a very important epigenetic factor and stable repressive mark, DNA methylation is essential in regulating gene expression for proper mammalian development (e.g., genomic imprinting and X-chromosome inactivation in females) and plays a role in maintaining genomic stability; it closely interacts with histone modification and chromatin-remodelling complexes to regulate the proper functioning of the genome (Esteller, 2005; Tost, 2009;

Kulis & Esteller, 2010). However, dysregulation of DNA methylation has been shown to be linked to several diseases including cancer (Jin *et al*., 2011; Hernando-Herraez *et al*., 2015).

1.6.2 DNA Methyltransferases (DNMTs)

DNA Methyltransferases (DNMTs) are a family of enzymes (DNMT1, DNMT3A, DNMT3B and DNMT3L) responsible for the transfer of a methyl group from SAM to the 5 position of cytosine (Jin & Robertson, 2013). Although they are similar structurally, they differ in their functions and expression patterns. *DNMT1* is responsible for the replication of the DNA methylation pattern from the parent DNA (hemimethylated DNA) to the newly synthesized DNA strand, it is involved in the maintenance and repair of DNA methylation (Goll & Bestor, 2005). DNMT3A and DNMT3B are known as *de novo DNMTs* and play a role in establishing new methylation patterns to unmodified DNA (Moore *et al*., 2013). DNMT3L is an enzyme that in human is encoded in the *DNMT3L* gene, expressed solely in germ cells and functions as a regulatory factor. Although it lacks detectable cytosine methyltransferase activity, it stimulates de novo methylation by DNMT3A (Gujar *et al*., 2019).

1.6.3 DNA Methylation and Cancer

In cancer, DNA hypermethylation in the upstream promoter region is associated with silencing of growth regulators like tumour suppressor genes (Figure 1.10) (Baylin, 2005; Kazanets *et al*., 2016).

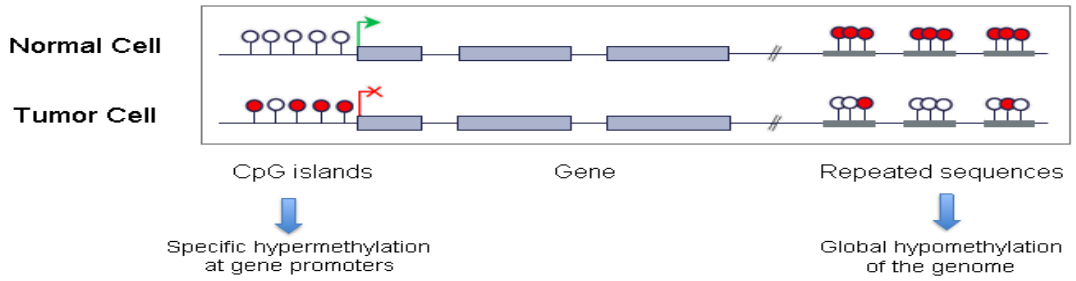


Figure 1.10: DNA hypermethylation can be seen in the CpG islands in tumour cell and gene silencing effect. White circles are unmethylated cytosine, while red circles are methylated cytosines (Moison *et al*., 2014).

It mainly occurs in CpG islands (CGI) of many gene promoters (which are most often, largely unmethylated during normal development and in adult cells) preventing the binding of transcription factors to DNA and subsequent transcription inactivation and gene silencing (Pfeifer, 2018). Gene silencing occurs as a result of the methylation of cytosine dinucleotides in the promoter region; however, there can be gene body methylation where methylation occurs at the transcribed regions (after the first exon) resulting in either silencing of two or more alternative promoters of a gene or retrotransposon elements; or increased gene expression in rapidly dividing cells at a global genome scale. The former resulting in a change in expression of specific transcript isoforms, while the latter may promote oncogenesis if the methylated genes have oncogenic properties (Moore *et al*., 2013; Pfeifer, 2018). Detailed knowledge of the relations between methylation and gene expression, and the triggers for the changes in methylation associated with development and progression for the multitudes of cancers remains largely unclear.

1.7 Non-Coding RNAs (ncRNAs)

The completion of the Encyclopaedia of DNA Elements (ENCODE) project revealed that only about 2% of the human genome has protein-coding potential, while the rest of the genome

is made up of repetitive transposable elements (RTEs) and ncRNAs (The Encode Project Consortium, 2012; Rinn, & Chang, 2012; Richard, 2017). The non-protein-coding regions of the genome are believed to serve as substrates for DNA-binding proteins (that control the expression and 3D architecture of the genome) as well as a template for the transcription of vast numbers of ncRNAs that are capable of carrying out a wide array of regulatory functions (Carninci *et al*., 2005; Amaral *et al*., 2008; Mercer *et al*., 2009).

ncRNAs are basically classified into two major categories based on their length; small ncRNAs which are less than 200 nucleotides in length (~<200 nt in length) and long ncRNAs (lncRNAs) that are equal to or greater than 200 nucleotides in length (~≥200 nt). The smallest ncRNAs include RNAs of 20–30 nucleotides which are involved in RNA silencing; e.g., microRNA (miRNA), small interfering RNA (siRNA) and PIWI (P-element-induced wimpy testis)-interacting RNAs (piRNA); other small ncRNAs of 100–200 nucleotides include the abundant housekeeping ncRNAs like transfer RNAs(tRNAs), ribosomal RNAs (rRNAs) and the ncRNAs classified based on their subcellular localization; i.e. small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), and small cytoplasmic RNA (scRNA) (Table 1.2) (Gutschner & Diederichs, 2012; Yoshimoto *et al*., 2015; Hombach & Kretz, 2016). Despite the discovery of several classes of small ncRNAs; miRNAs, piRNAs and siRNAs are generally recognized as the three main categories of small ncRNAs due to their biogenesis, structure, function and associated effector proteins (Carthew & Sontheimer, 2009).

1.7.1 RNA interference (RNAi) Pathway

RNAi also known previously as Post Transcriptional Gene Silencing (PTGS) is an evolutionary conserved biological process that triggers gene silencing by small RNA molecules (double-stranded RNA and pre-miRNA precursors). It was initially discovered in fungi and plants and

subsequently in *Caenorhabditis elegans, Drosophillia* and finally in the mammals. siRNAs and miRNAs are processed from long dsRNA and pre-miRNA respectively. While siRNA is cleaved into short double-stranded fragments of about 21 nucleotides in length by the RNAse III enzyme Dicer, miRNA processing is catalysed by Drosha in the nucleus and Dicer in the cytoplasm (Kim and Rossi, 2007; NCBI 2017).

The dsRNA molecules are separated into two single-stranded RNA having 2nt overhang on the 3' end of each strand; the passenger strand is degraded, while the guide strand is incorporated into a multi-protein silencing complex known as RNA-induced silencing complex (RISC). The RISC complex is guided to the target mRNA in a sequence-specific manner, and the mRNA is cleaved by Argonaute 2a member of the Argonaute protein family important in RNAi, regulation of stem cell development and tumorigenesis. On the other hand, pre-miRNAs in the cytoplasm are enzymatically cleaved by Dicer into mature miRNA which then binds to RISC. The miRNA-RISC complex binding to a target mRNA inhibits translation thereby silencing the gene (Dana *et al*., 2017; Kim and Kim, 2012).

1.7.2 MicroRNAs

miRNAs are small ncRNAs that were first discovered in *C. elegans*, they are 18–25 nucleotides long, single-stranded, endogenous, and evolutionarily conserved (Bartel, 2004). They target protein-coding mRNAs at the post-transcriptional level and bind to the 3'-untranslated regions (3' UTRs) resulting in target mRNA repression and consequently translational inhibition and gene silencing (Ricciuti *et al*., 2014; Ha & Kim, 2014). miRNAs are well characterised, and the best studied of all the small ncRNAs, they have been shown to either act as tumour suppressors or oncogenic initiators when dysregulated. To date over

2,469 new human miRNAs have been reported, with over 1,098 validated (Friedlander *et al*., 2014; Bracken *et al*., 2016; O'Brien *et al*., 2018).

1.7.3 Small interfering RNAs

siRNAs also called silencing RNAs are 19- 25 nucleotides in length and are similar to miRNAs. They operate within the RNA interference pathway, leading to a sequence-specific degradation of complementary target mRNAs, thereby inhibiting translation (McManus *et al*., 2002; Rand *et al*., 2005; Dana *et al*., 2017)

1.7.4 PIWI-interacting RNAs

piRNAs are the most recent and longest class of small non-coding RNAs and are often expressed in the germline. They are 24–33 nucleotides long and are generated by Dicer-independent mechanisms, unlike siRNAs and miRNAs. piRNAs interact only with PIWI clade proteins (a germ-line specific subfamily of Argonaute proteins) also involved in the gene-silencing pathways (Moazed, 2009). It has been reported that there are over 30,000 piRNAs in the human genome, 80% of which developed from intergenic sequences. They are subdivided into three types based on their source of origin viz:- the transposon-derived piRNA from the, mRNA-derived piRNA and lncRNA-derived piRNA (Robine *et al*., 2009). Recent evidence has shown their role in the preservation of genomic integrity during germline development by silencing transposable elements (TEs) via the PIWI-piRNA pathway (Siomi *et al*., 2011; Iwasaki *et al*., 2015).

The PIWI-piRNA pathway is a gene regulatory machinery that is important in inhibition of TEs (Sturm *et al*., 2017). The association between piRNA and the PIWI proteins induces the piRNA-induced silencing complex (pi-RISC) which enables the recognition of target mRNA for cleavage and eventual gene silencing. The PIWI-piRNA complex has also been reported to

regulate HP1, H3K9 and DNA methylation in the nucleus silencing transposable elements (Meseure and Alsiba, 2018).

These molecules are important not only in regulation of gene expression and alternative splicing but also in epigenetic control and guidance of chromatin remodeling complexes, and have become potential target for diagnostics, prognostics and therapeutics.

1.8 Long non-coding RNAs

lncRNAs are common both in prokaryotic and eukaryotic genomes. They were originally discovered through large-scale sequencing of full-length cDNA libraries in mice and microarray tiling arrays (Gutschner & Diederichs, 2012; Kashi *et al*., 2016). lncRNAs share many features of mRNAs (they are frequently transcribed by RNA polymerase II, polyadenylated and can show complex splicing patterns), but lack functional open reading frames (ORF) (Cheetham *et al*., 2013). In comparison with mRNAs, lncRNAs are generally less abundant, more likely to be expressed in a tissue-specific manner, and poorly conserved between species (Ulitsky & Bartel, 2013; Yoshimoto *et al*., 2016). They are defined as endogenous cellular RNAs of more than 200 nucleotides in length (Gutschner & Diederichs, 2012) based on RNA purification protocols which excludes small RNAs. However, recent studies have led to the redefining of lncRNAs to include "RNA molecules that may function as primary or spliced transcripts and do not fit into known classes of small RNAs or structural RNAs" examples like *BCI* and *snaR* having nucleotide lengths less than 200 (Amaral *et al*., 2008; Quinn and Chang, 2016).

They have been shown to be important in numerous biological processes (transcription, translation, splicing, imprinting), and have also been implicated in the development and progression of many diseases including cancer (Clark & Mattick, 2011; Chen *et al*., 2016).

lncRNAs often overlap with or are distributed between coding and non-coding transcripts. There are different types of lncRNAs based on their genomic location in relation to protein-coding genes as shown in Figure 1.11 (Kung *et al*., 2013). These include (i) sense lncRNAs- which are transcribed from the sense strand of protein-coding genes and may overlap these genes; (ii) antisense lncRNAs also called Natural Antisense Transcripts [NATs]- that overlap one or more exons of a protein-coding gene on the opposite strand partially or completely; (iii) bi-directional lncRNAs- that are located on the opposite strand of a protein-coding gene and their transcription initiated in close genomic proximity [i.e. less than 1000 base pairs]; (iv) intronic lncRNAs- which are derived from an intron of a protein-coding gene; (v) intergenic lncRNAs- that are transcribed from intergenic regions (between protein-coding genes) and more recently, (vi) enhancer lncRNAs and (vii) circular RNAs (Gutschner & Diederichs, 2012; Qiu *et al*., 2013; Ma *et al*., 2013; Schein *et al*., 2016; Lorenzen & Thum, 2016).

lncRNAs are also classified functionally into *cis-acting* and *trans-acting* based on the influence they exert on DNA sequences (Ma *et al*., 2013). *Cis-acting* lncRNAs restrict their action to their site of synthesis and regulate the expression of genes on the same chromosomal location. Examples include *Xist* (X inactive-specific transcript), *Tsix*, *ANRIL* (antisense non-coding RNA in the *INK4* locus), *HOTTIP* (HOXA transcript at the distal tip), *COLDAIR* (cold assisted intronic non-coding RNA) etc. However, *trans-acting* lncRNAs disperse from their site of synthesis and act directly on many genes at great distances, including other chromosomal locations. Examples of this class of lncRNAs are *HOTAIR* (HOX antisense intergenic RNA), *GAS5* (growth arrest-specific transcript 5), *Jpx* and *MALAT1*

(Metastasis associated lung adenocarcinoma transcript 1) (Guil & Esteller, 2012; Lee, 2012; Vance & Ponting, 2014).

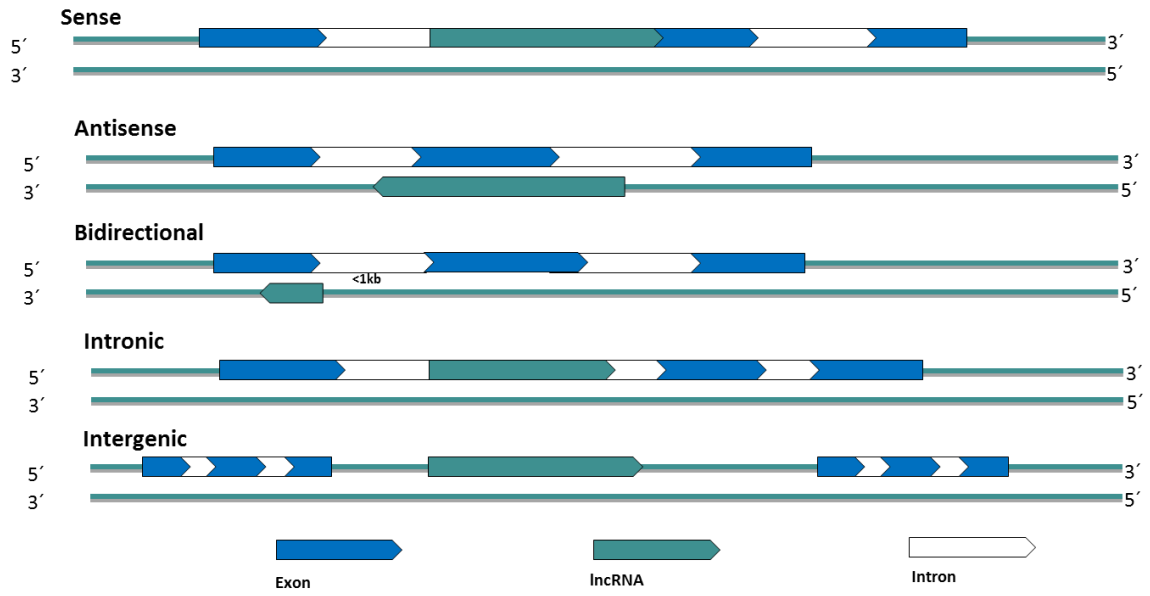


Figure 1.11: Categories of lncRNAs based on their genomic location in relation to the protein genes.

Table 1.2: Types of ncRNAs (Adapted from (Esteller, 2011)

Group	Type	Length (nt)	Functions	Examples	References
Short ncRNAs	miRNA	18 – 25bp	mRNA degradation and RNA silencing	miR-15, miR-16, miR-21, miR-155	(Calin *et al*., 2002; Lawrie *et al*., 2008; Volinia *et al*., 2006; Yanaihara *et al*., 2006)
	tiRNA	17 -18bp	Regulation of transcription	Associated with CAP1 genes	(Taft *et al*., 2010)
	piRNA	24 -31bp	Silencing of transposons and maintaining genome stability	They exist as clusters	(Sai Lakshmi & Agrawal, 2008; Luteijn & Ketting, 2013; Ross *et al*., 2014)
Mid-size ncRNAs	snoRNAs	60-300bp	rRNA modifications	*U12*, *U50*, *SNORD115*, *SNORD116*	(Kiss-Laįszla *et al*., 1996)
	PASRs	22-200bp	Transcriptional activation or repression	Interact with 50% of polycomb group of proteins	(Kapranov *et al*., 2007; Kaikkonen *et al*., 2011)
	TSSa-RNAs	20-90bp	Maintenance of transcription	Associated with *RNF12* and *CCDC52* genes	(Seila *et al*., 2008)
Long ncRNAs	lincRNAs	>200bp	Regulation of gene expression by chromatin complex modification	*HOTAIR*, *MALAT1*, *lincRNA-p21* and *ANRIL*	(Khalil *et al*., 2009; Huarte *et al*., 2010; Tsai *et a*l., 2010)
	T-UCRs	>200bp	Antisense inhibitors of protein coding genes	uc.202, uc.328, uc. 230	(Braconi *et al*., 2011; Calin *et al*., 2007; Fassan *et al*., 2014)
	Other lncRNAs	>200bp	X-chromosome inactivation, telomere regulation etc.	*XIST*, *TSIX*, *TERRAs*, *H19*	(Rinn *et al*., 2007; Gupta *et al*., 2010)

1.8.1 Cellular functions and mechanisms of lncRNAs

Despite the vast array of lncRNAs identified to date, much is yet to be understood about their function. Their conservation, however, has strengthened the debate over their functional relevance. Some of these transcripts are derived from ultra-conserved genomic regions (UCR) and their aberrant expression has been associated with numerous diseases including cancers (Braconi *et al*., 2011; Fang & Fullwood, 2016).

It has become apparent that lncRNAs play an important role in regulating gene expression at various levels. They have been shown to be important in chromatin modification, as well as transcriptional and post-transcriptional regulation (Ponting *et al*., 2009; Chen & Carmichael, 2010). There are several mechanisms that have been postulated that will aid the understanding of lncRNA functions (Cheetham *et al*., 2013).

Chromatin Modification

lncRNAs can act as scaffold molecules, controlling the delivery of regulatory proteins to specific loci where they are required to effect epigenetic changes either in *cis* or *trans*. lncRNAs like *Xist, COLDAIR, HOTTIP, HOTAIR, KCNQ1OT1* (KCNQ1 overlapping transcript 1) interact with chromatin remodelling complexes resulting in heterochromatin formation in specific genomic loci consequently leading to reduction in target gene expression and gene silencing (Wang & Chang, 2011; Gutschner & Diederichs, 2012). *ANRIL* antisense transcript also known as CDKN2B-AS is an example of a lncRNA that acts as a scaffold. It recruits and binds to the chromobox 7 (CBX7) subunit of the PRC1 (Polycomb repressive complex 1) and SUZ12 component of PRC2 (Polycomb repressive complex 2). These complexes then direct H3K27me (methylation of histone H3 at Lys-27) to the target loci bringing about the silencing of sense transcripts expressed from this region (Yap *et al*., 2010). In a similar way,

HOTAIR binds to and recruits PRC2 and the LSD1 (lysine-specific demethylase 1)- a component of the CoREST (co-repressor for element-1-silencing transcription factor) complex. This brings about specific alterations in the methylation status of H3K27me, change in the nature of the chromatin surrounding the *HOTAIR* targets in the HOXD (homeobox D) cluster and consequent gene silencing (Rinn *et al*., 2007; Gupta *et al*., 2010; Fang & Fullwood, 2016).

Regulation of Alternative Splicing

Some lncRNAs act at post-transcriptional levels and function in RNA processing by controlling alternative splicing of pre-mRNAs (precursor mRNAs). A typical example is *MALAT1* also known as *NEAT2* (noncoding nuclear enriched abundant transcript 2), which interacts with the serine/arginine-rich (SR) splicing regulatory proteins (a member of the trans-acting protein factors) that control alternative splicing patterns. The interaction of *MALAT1* with these factors results in their localization to nuclear speckles (mRNA processing site) and subsequent release when needed. Disruptions to the expression or activity of *MALAT1* have been implicated in some cancers (Tripathi *et al*., 2010). More recently, Gonzalez et al, have reported the regulation of alternative splicing of *FGFR2* mRNA (Fibroblast growth factor receptor 2) by a lncRNA *asFGFR2* (antisense *FGFR2*) transcript via the recruitment of PRC2 and KDM2A to the *FGFR2* locus (Gonzalez *et al*., 2015).

Transcriptional Co-activation and inhibition

lncRNAs can also acts as transcriptional co-activators or inhibitors, their expression has been reported to be induced by transcriptional factors in response to specific stimuli as a result of DNA damage. An Example is the *lincRNA-p21* (long intergenic non-coding RNA p21) that

possesses p53 binding sites in its promoter and is directly induced by p53. *lincRNA-p21* consequently represses genes that inhibit cell cycle arrest and apoptosis by binding to heterogeneous nuclear ribonucleoprotein K (hnRNP-K) (Huarte *et al*., 2010; Liu & Lu, 2012; Quan *et al*., 2015).

Decoy Molecules

lncRNAs have also been reported to affect the transcription process by acting as decoys and sequestering transcription factors from degradation by target genes or competing for the binding sites of these transcription factors (Kung *et al*., 2013). Examples include, *PTENP1* (phosphatase and tensin homolog pseudogene 1) which binds to miRNAs that downregulate *PTEN* (a tumour suppressor gene) thereby increasing expression of *PTEN* (Poliseno et al. 2010). Gas5 acts as a decoy by binding to the DNA binding domain of glucocorticoid receptor and inhibiting the binding of glucocorticoid response elements thereby modulating its activity (Kino *et al*., 2010).

1.8.2 lncRNAs and Cancer

Aberrant expression of lncRNAs in cancer had been reported before the availability of high-throughput sequencing technologies; *H19, PCA3* (prostate cancer associated 3), *MALAT1* were shown to be over expressed in tumours (Zhang *et al*., 1993; de Kok *et al*., 2002; Bartonicek *et al*., 2016). However, increasing numbers of dysregulated lncRNAs have now been identified in cancer tissues and cell lines through genome-wide transcriptome analyses (Qiu *et al*., 2013; Huarte, 2015; Bartonicek *et al*., 2016). In view of the large-scale dysregulation of lncRNAs in different cancers, it is plausible to infer that lncRNAs are not just secondary effects of cancer but are involved in tumorigenic or metastatic processes. However, lncRNA research is still at the cradle stage and only a relatively small number of

these group of transcripts have been characterized and their role in cancer fully understood, despite the large array of lncRNAs identified to date.

1.8.3 Examples of Well Characterized lncRNAs Deregulated in Cancer

MALAT1: MALAT1 (Metastasis-associated lung adenocarcinoma transcript 1) also known as *NEAT2*, is an 8.7 kb lncRNA gene that is located on chromosome 11q13 (GeneCard, 2017). It has been previously reported to contribute to tumour development, invasion and metastasis. It was shown to be over-expressed in non-small cell lung cancer and this was associated with metastasis and poor prognosis in a histology-specific manner (Ji *et al*., 2003), it was further reported to promote cell motility of lung cancer cells (Tano *et al*., 2010). A study by Schmidt *et al*., (2011) revealed the role of the metastasis marker as a potential key player in the metastatic process. *MALAT1* is also over-expressed in ovarian cancer (Zhou *et al*., 2016), breast cancer (Meseure *et al*., 2016) and colorectal cancer (Ji *et al*., 2014). *MALAT1* has been shown to be involved in the modulation of several molecular signalling pathways, resulting in the modification of different associated cancer phenotypes (Li *et al*., 2018). In lung cancer, *MALAT1* knock down in cisplatin-resistant A549/DDP cells is shown to increase cisplatin sensitivity through upregulation of miR-101-3p and *MCL1* downregulation (Wang *et al*., 2017). Although *MALAT1* may not be an independent diagnostic biomarker in NSCLC as a result of its low sensitivity based on small sample size, it has been shown to be a potential prognostic biomarker (Lin *et al*., 2018).

HOTAIR: The HOX antisense intergenic RNA (*HOTAIR*) is a 2.2 kb transcript which is transcribed from HOX C gene cluster on chromosome 12 and was first identified by Rinn and colleagues (Rinn *et al*., 2007; GeneCard, 2017). The *HOTAIR* gene was found to be over-expressed in breast cancer (Gupta *et al*., 2010), oesophageal cancer (Chen *et al*., 2013)

hepatocellular cancer (Yang *et al*., 2011) and its high expression correlated with metastasis, recurrence, and poor prognosis.

PTENP 1: PTENP1 is the pseudogene of the tumour suppressor gene *PTEN* (phosphatase and tensin homolog) and is located on chromosome 9p13.3 (GeneCard, 2017). Highly similar to *PTEN* with homology of ~98% at the 3' untranslated region (3'UTR), it has been shown to function as a "decoy" interacting with miRNA miR21 that represses *PTEN*, thereby increasing the expression of *PTEN* (Poliseno *et al*., 2010; Gao *et al*., 2017). Its reduced expression has been associated with poor survival outcomes in head and neck squamous cell carcinoma (Liu *et al*., 2017) and increased proliferation and metastasis in clear-cell renal cell carcinoma (Yu *et al*., 2014).

1.8.4 lncRNAs Deregulated in NSCLC

Apart from the well characterized *MALAT1* in lung cancer, there are other lncRNAs that have also been identified to be deregulated in NSCLC but are yet to be fully studied (Table 1.3).

PVT1: The *PVT1* (plasmacytoma variant translocation 1) lncRNA encoded in the *PVT1* gene located on chromosome 8q24.21, has been reported to be upregulated in a large variety of human tumours (Colombo *et al*., 2015; Yang *et al*., 2019).

Takahashi *et al*., showed that colorectal cancer patients with high *PVT1* expression had significantly poorer prognosis than those with low *PVT1* expression (Takahashi *et al*., 2014). Wang *et al*., revealed that *PVT1* could promote cell proliferation, cell cycling, and the acquisition of stem cell-like properties in hepatocellular carcinoma cells by stabilizing NOP2 protein (Wang *et al*., 2014; Colombo *et al*., 2015).

Studies carried out by Cui *et al.*, showed significant increase in *PVT1* levels in NSCLC tissues and cell lines (A549, H157, H226, H460, and HCC827). Their results further revealed that elevated *PVT1* level was associated with poor overall survival and advanced T-stage and tumour-node-metastasis. It has also been previously reported to possess oncogenic properties in numerous cancers, with its overexpression resulting in the promotion of cell proliferation and cell cycle progression in NSCLC (Cui *et al*., 2016). Wan *et al*., in their findings also showed that regulation of LATS2 (large tumour suppressor kinase 2) by *PVT1* through its interaction with EZH2 was responsible for the promotion of cell proliferation by *PVT1* (Wan *et al*., 2016).

HOXA11-AS1: HOXA11 antisense 1 is located on chromosome7p15.2 and transcribed from the antisense strand of the HOXA11 gene. It has been shown to inhibit oncogenic phenotype in ovarian cancer (Richards *et al*., 2015) and is a potential biomarker target for glioma progression (Wang *et al*., 2016). Zhang *et al*., in their study based on microarray and other bioinformatic data, as well as original data obtained from The Cancer Genome Atlas (TCGA) reported its upregulation in LAD and LSCC. Further studies showed they were upregulated in NSCLC tumour tissues compared to the paired normal tissues and that they could drive the development and progression of NSCLC by regulating *DOCK8* expression (Zhang *et al*., 2016; Chen *et al*., 2017a).

Table 1.3: Examples of lncRNAs deregulated in NSCLC

lncRNA	Deregulation Pattern	Lung cancer Type	Sample No	References
MALAT1	Upregulated	Adenocarcinoma	70	(Ji *et al*., 2003; Schmidt *et al*., 2011)
HOTAIR	Upregulated	Adenocarcinoma/Squamous cell carcinoma	77	(Nakagawa *et al*., 2013)
ANRIL	Upregulated	-	68	(Nie *et al*., 2015)
SOX2OT	Upregulated	Squamous cell carcinoma	83	(Hou *et al*., 2014)
HNF1A-AS1	Upregulated	Adenocarcinoma	40	(Wu *et al*., 2015)
PVT1	Upregulated	-	82	(Yang *et al*., 2014)
H19	Upregulated	Squamous cell carcinoma	70	(Zhang *et al*., 2016)
CARLO-5	Upregulated	-	-	(Luo *et al*., 2014)
BCAR4	Upregulated	Adenocarcinoma/Squamous cell carcinoma	76	(Li *et al*., 2017a)
MEG3	Downregulated	-	44	(Lu *et al*., 2013)
SPRY4-IT1	Downregulated	-	121	(Sun *et al*., 2014)
GAS5	Downregulated	Adenocarcinoma/Squamous cell carcinoma	72	(Shi *et al*., 2015)
TUG1	Downregulated	-	192	(Zhang *et al*., 2014)

TUBA4B

A novel lncRNA reported to be downregulated in epithelial ovarian cancer and associated with poor prognosis, has now been shown to also be downregulated in NSCLC. Overexpression of this gene in A549 and NCI-H1299 cell lines was associated with reduced proliferation (Chen *et al*., 2017).

BCAR4

Breast cancer anti-oestrogen resistance 4(*BCAR4*) located on chromosomal position 16p13.13 has been implicated in breast cancer metastasis via chemokine mediated binding of *BCAR4* to transcription factors (TFs) SNIP1 and PNUTS (Xing *et al*., 2014). More recently, this lncRNA has been shown to be upregulated in NSCLC and promotes proliferation, invasion and metastasis (Li *et al*., 2017a).

miR31HG

miR31HG also known as *LOC554202* is a 2,246 bases lncRNA located on chromosome 9p21.3 and is known to encapsulate miR31 in its first intron and contain a CpG island in its first exon. It has been reported to be present in the nucleus and cytoplasm of fibroblasts and to be downregulated in triple-negative breast cancer (TNBC) cell lines of luminal subtype (Montes *et al*., 2015). However, Shi *et al*., showed upregulation of *miR31HG* in breast cancer tissues compared to paired normals (Shi *et al*., 2014). Recent studies by Wang *et al*., revealed that *miR31HG* was differentially expressed in gefitinib-sensitive and gefitinib-resistant PC9 cell lines respectively and that increased expression of *miR31HG* resulted in gefitinib resistance in the NSCLC cell line through the EGFR/P13K/AKT signalling pathway (Wang *et al*., 2017a).

SPRY4-IT1

Sprouty4-Intron 1 is a long non-coding RNA that is transcribed from the intron of the *SPRY4* gene. It has been shown to be overexpressed in different kinds of tumour tissues (including melanoma, breast cancer, gastric cancer, colon cancer, hepatocellular carcinoma and NSCLC) (Xie *et al*., 2015; Cao *et al*., 2016; Li *et al*., 2017b).

XLOC_008466

It has also been shown to be upregulated and to have oncogenic properties in cervical cancer (Guo *et al*., 2018). It has been reported to be upregulated in NSCLC tumours and functions as an oncogene in NSCLC (Yang *et al*., 2017a).

1.8.5 lncRNAs: as biomarkers for early detection

A biomarker is a biological molecule found in blood, other bodily fluids or tissues, that can be used to measure and evaluate normal or abnormal biological processes objectively. They can be used for screening, prediction of disease outcome, diagnostic or therapeutic purposes (Strimbu & Tavel, 2010).

Current diagnostic tools for NSCLC include the use of imaging techniques or body fluid analysis, at which time the disease may have metastasized, and available treatment options cannot provide a cure. Therefore, the development of non-invasive molecular biomarkers for early detection of NSCLC is a major translational research challenge. The recent evidence on the extent of lncRNA deregulation in NSCLC provisionally makes them worth investigating for biomarker properties. They, however, must be specific, sensitive, and detectable in specimens obtained through minimally invasive procedures to be clinically useful as biomarkers (Mazzone *et al*., 2017).

Some lncRNAs have been identified as potential non-invasive biomarkers for early detection of cancer. A typical example, is the FDA approved prostate cancer antigen 3 (*PCA3*) also known as *DD3*, a prostate cancer specific lncRNA first identified in 1999 and is present at high levels in prostate cancer tissues (Deng *et al*., 2017; Bolha *et al*., 2017). Since prostate specific antigen (*PSA*), which is the frequently used biomarker for prostate cancer is prostate specific, the *PCA3* is used to determine if the PSA increase is due to prostate cancer. As a biomarker for early detection of prostate cancer, it can be detected by RT-qPCR in urine or urine sediments obtained after prostatic massage or digital rectal examination (Tinzl *et al*., 2004; Marks *et al*., 2007; van Gils *et al*., 2007).

lncRNA16 is a lncRNA that has been identified as a potential biomarker for the early detection of lung cancer. Studies carried out by Zhu *et al*, showed lncRNA16 to be highly expressed in lung tissues and elevated in plasma of lung cancer patients. They further demonstrated that *lncRNA16* was detected at the ground glass opacity, early and advanced stage of the disease making it a suitable biomarker for early detection. As a diagnostic biomarker, *lncRNA16* was found to have 73.97% and 100% sensitivity and specificity respectively when compared to currently used tumour markers like carcinoembryonic antigen (CEA) and cancer antigen 125 (CA125)(Zhu *et al*., 2017).

1.9 Hypotheses

Deregulation of long non-coding RNAs is extensive in NSCLC and many of these lncRNAs may be drivers of the development and progression of NSCLC.

1.10 Aim and Objectives

The aim of this study is to identify novel lncRNAs that are deregulated in NSCLC and evaluate their role in lung carcinogenesis.

Objectives of this study are: -

- To evaluate expression profile of identified lncRNAs in primary NSCLC tissues and adjacent normal tissues, NSCLC cell lines and explore their biomarker potential.
- To investigate the epigenetic causes of the deregulation in the identified lncRNAs.
- To analyse the functional contribution of chosen lncRNA target to the development of NSCLC and explore their effect in cell phenotype and drug resistance.

2. MATERIALS AND METHOD

2.1. Materials and Instruments

Absolute Ethanol

Agarose (Sigma-Aldrich)

Agilent RNA 6000 Nano kit (Agilent Technologies) Catalogue No. 5067-1511

Ampicillin

Annealing Buffer (Qiagen)

Beakers

Bioanalyzer 2100 (Agilent Technologies)

CO_2 Incubator

Corning Cryogenic Vials

Decitabine (5-aza-2-deoxycytidine) (Sigma Aldrich) Catalogue No. 11390

DEPC (diethyl pyrocarbonate) Water

Direct-zol RNA miniprep kit (Zymo Research) Catalogue Nos. R2050-R2053

Dri-Block DB-2A (Techne)

Dulbecco's Modified Eagle's Medium (DMEM) (Sigma-Aldrich)

Electrophoresis Chamber

e-Myco^{TM} plus Mycoplasma PCR Detection Kit (Intron Biotechnology)

Eppendorf tubes (0.6 ml, 1.5 ml, 2 ml)

EVOS Cell Imaging System (Thermo Fisher Scientific)

EZ DNA Methylation-Gold^{TM} Kits (Zymo Research. Catalogue No. D5005 and D5006)

Fetal Bovine Serum (FBS) (Sigma-Aldrich)

Flasks (25 cm^2 and 75 cm^2)

FuGENE HD Transfection Reagent (Promega)

G418 Sulfate (Geneticin) Thermo Fisher Scientific)

Geltrex Matrix (Thermo Fisher Scientific)

GenePrint 10 System (Promega) Catalogue No. B9510

GENios Microplate Reader (Tecan)

Haemocytometer (Neubauer)

High Capacity Reverse Transcription Kit (Applied Biosystems) Catalogue No. 4368813

Human recombinant Epidermal Growth Factor (HrEGF) (Life Technologies)

Human TATA-box Binding Protein (TBP) (Applied Biosystems)

HyperLadder 50 bp (Bioline)

Microscope

MTT 3-(4, 5-dimethylthiazol)-2-yl)-2, 5-diphenyltetrazolium bromide (Sigma-Aldrich)

NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific)

Nitroblue Tetrazolium Chloride (Thermo Fisher Scientific)

Oligo 7 Primer Analysis Software (Molecular Biology Insights MBI)

PCR Machine (Thermo Fisher Scientific)

pH Strip

Phosphate Buffered Saline ((PBS) (Sigma Aldrich)

Polystyrene Cloning Rings (Thermo Fisher Scientific)

Primer / Probe sets for lncRNAs

pTARGET Mammalian Vector Expression System (Promega)

PyroMark Assay Design Software 2.0 (Qiagen)

PyroMark Binding Buffer (Qiagen)

PyroMark Q96 Vacuum Workstation (Qiagen)

QIAamp DNA Blood Mini Kit (Qiagen). Catalogue No. 51104

QuantiTect Probe Master Mix (Qiagen)

Real Time PCR Instrument 7500 & 7500 Fast (Applied Biosystems)

RNase-free Water (Thermo Fisher Scientific)

RNaseZAP (Ambion) Catalogue No. 9780

Safe View Nucleic Acid Stain (NBS Biologicals)

SOC Media (Sigma-Aldrich)

Sodium Hydroxide (NaOH)

Streptavidin Sepharose High Performance Beads (GE Healthcare)

SurePrint G3 Comparative Genomic Hybridization (CGH) Microarray Platform (Agilent Technologies)

ThinCert Cell Culture Inserts (24 well plate) (Greinier Bio-One)

Trigene 2%

Trypsin-EDTA Solution (Sigma Aldrich)

Tryptone, Yeast Extract, Sodium Chloride (Thermo Fisher Scientific)

Universal Tubes (25 ml)

Valproic Acid (VPA) (Sigma-Aldrich) Catalogue No. P4543

Water Bath

Well Plates (24, 48 and 96)

Zyppy Plasmid Midiprep Kit (Zymoresearch)

2.2. Methods

2.2.1. NSCLC Primary Tissues

All the samples used in this study were obtained from Liverpool Lung Project (LLP) and adequate ethical guidelines stipulated by the University of Liverpool ethics committee have been adhered to in relation to sample handling, patient consent and confidentialities (https://www.ethicsguidebook.ac.uk).

2.2.2. NSCLC Cell lines and Cell Culture

Nine (9) NSCLC human cell lines (A549, CALU6, CORL23, SKLU1, CALU3, SKMES, LUDLU1, LUNG14 and H358) and four (4) Non-tumorigenic immortalized Human Bronchial Epithelial Cells with its isogenic derivatives or variants (KRAS mutation and TP53 knockdown) (HBEC3 KT, HBEC3 KTR. HBEC3 KTP53 and HBEC3 KTR53) were utilized in this study. An overview of the cell lines used, and their histological origin is shown in Table 2.1.

All cell culture procedures were carried out under sterile conditions in a Class II Biological Safety Cabinet (Esco). NSCLC cells were maintained in a humidified incubator with a constant supply of 5% CO_2 at 37^{o}C and were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 5% or 10% FBS and F-12 Hams nutrient mixture. All HBECs were maintained in Keratinocyte-SFM medium supplemented with 50 µg/mL Bovine Pituitary Extract (BPE) and 5 ng/ml human recombinant Epidermal Growth Factor (rEGF). Cells were grown in 25 cm^2 and 75 cm^2 flasks and routinely passaged when monolayer cell growth was about 85%-95% confluent.

For passaging, cells were washed twice with sterile Dulbecco's phosphate buffered saline (PBS), treated with trypsin-EDTA and incubated at 37 °C and 5% CO_2 incubator until cells had detached from the flasks, after incubation equal volume of medium was added to the flasks

to pipet the cells up and down to disperse them. Cells were pelleted by centrifugation at 300 g for 5 minutes and resuspended with medium and transferred to flasks containing fresh medium and incubated at 5% CO_2 and at 37^{o}C.

All cell lines were tested for mycoplasma using the e-Myco^{TM} plus Mycoplasma PCR Detection Kit (Catalogue no. 17341) according to manufacturer's protocol and authenticated using the short tandem repeat (STR) profiling kit and software GenePrint 10 Kit prior to the commencement of the study. STR profile obtained for the cell lines was compared with STR profile of the same cell lines from https://www.lgcstandards-atcc.org.

Table 2.1: Cell lines utilized in this study and their histological subtypes (source: ATCC, ECACC)

Cell lines	**Histological Subtype**
A549	Adenocarcinoma
CALU6	Adenocarcinoma
CORL23	Large-cell carcinoma
SKLU1	Adenocarcinoma
CALU3	Adenocarcinoma
SKMES	Squamous cell carcinoma
LUDLU1	Squamous cell carcinoma
LUNG14 (PC9)	Adenocarcinoma
H358	Bronchioalveolar adenocarcinoma
HBEC3 KT	Non-tumorigenic human bronchial epithelial
HBEC3 KTR	Non-tumorigenic human bronchial epithelial
HBEC3 KT53	Non-tumorigenic human bronchial epithelial
HBEC3 KTR53*	Non-tumorigenic human bronchial epithelial

*cell line with combination of TP53 knockdown and KRAS mutation (Davies et al. 2014).

2.2.3. RNA Expression Analysis

2.2.3.1. RNA Extraction

RNA extraction was carried out in a laminar flow hood and all materials sterilized with 2% trigene and 70% ethanol. Total RNA was extracted using Direct-zol RNA miniprep kit following manufacturer's protocol. Cells grown in a 75 cm^2 flask to ~70% confluency were aspirated of old medium, washed with 10 ml DMEM PBS twice, lysed with 1 ml TRI-Reagent and collected in a 1.5 ml Eppendorf tube. Equal volume (1:1) of absolute ethanol (95-100%) was added to the TRI-Reagent and mixed thoroughly by vortexing for 30 seconds. The lysate was loaded into a Zymo-Spin™ IIC Column and centrifuged at 16,000 g for 1 minute. The column was transferred into a new collection tube and the flow through discarded, for lysate volumes more than 700 µl, the column was reloaded with remaining lysate and step 2 repeated, the column was then transferred into a new collection tube and the flow through discarded. DNase treatment was carried out by adding 80 µl DNase I reaction mix (Table 2.2) directly into the column and incubated at room temperature for 15 minutes before centrifuging for 1 minute. After which, 400 µl of Direct-zol RNA PreWash Buffer was added to the column and centrifuged at 16,000 g for 1 minute, this step was repeated.

Table 2.2: Composition of DNase I Reaction Mix

DNase I' (lyophilized)	5 µl
DNase I Reaction Buffer (10X)	8 µl
DNase/RNase Free Water	3 µl
RNA Wash Buffer (with ethanol)	64 µl

700 µl of RNA Wash Buffer was then added to the column, centrifuged for 1 minute at 16,000 g and the column transferred into a collection tube and centrifuged for 4 minutes at

16,000 g in order to discard any residual volume. To elute RNA, the column was transferred in a clean 1.5 ml new RNase-free collection tube, 50 µl of DNase/RNase-Free Water was carefully added to the column and incubated at room temperature for 1minute. After which the column was centrifuged at 16,000 g for 1 minute and the eluted RNA was used immediately or stored at -80^{o}C until use.

2.2.3.2. RNA Quantitation

Quality control assessment of the RNA samples was carried out to determine their concentration, purity and integrity.

RNA concentration and purity

The concentration and purity of the RNA samples obtained from NSCLC cell lines were determined using the NanoDrop 2000 Spectrophotometer and absorbance measured at 260 nm and 280 nm. After initializing from the NanoDrop 2000 software and RNA selected as the nucleic acid to be measured, the upper and lower optical surfaces of the machine were cleaned with RNAse-free water and wiped off using sterile dry lint-free wipes. Blank measurement was taken with 1.5 µl of RNAse free-water and the optical surfaces wiped clean. 1.5 µl of RNA sample was dispensed unto the lower optical surface, the upper lever arm closed and measured. The same procedure was repeated for each individual sample and the optical surfaces wiped clean after each measurement.

RNA integrity

The concentration and quality of RNA samples from NSCLC tumour and their adjacent normal tissues was determined using the Agilent RNA 6000 Nano kit on an Agilent 2100 Bioanalyzer (a chip-based capillary electrophoresis). Summarily, the bioanalyzer electrodes were decontaminated with 400 µl RNaseZAP and DEPC (diethyl pyrocarbonate) water

respectively. 550 µl of RNA 6000 gel matrix was vortexed for 1 minute, transferred by pipetting into a spin filter, centrifuged at 1,500 g for 10 minutes and aliquots of 65 µl used per chip. RNA Gel-dye mix was prepared by adding 1 µl of vortexed RNA dye concentrate to 65 µl of gel matrix, the mix was vortexed for 30 seconds and centrifuged at 14,000 g for 10 minutes before use. RNA chip was placed on the chip priming station and 9 µl of gel-dye mix was loaded into well marked **G** on the chip. The gel-dye mix was spread across the channels of the chip by pressing down the plunger and holding for 30 seconds, after which the clip was released. 9 µl of gel-dye mix was added to the wells marked **G**, 5 µl of the marker was then pipetted into all the wells (including the well designated with the ladder symbol). 1 µl of each RNA sample and 1 µl of ladder previously denatured at 70^{o}C for 2 minutes were added to each sample well and the well that is marked with the ladder symbol respectively. The chip was vortexed for 1 minute at 2400 r.p.m. and placed on the Agilent 2100 Bioanalyzer to run. RNA samples with RIN values ≥ 8 were selected for microarray.

2.2.4. Expression Microarray

Expression microarray profiling was carried out by the Centre for Genomic research (CGR), University of Liverpool. The human 8x60k array manufactured by Agilent Technology with design ID: 047718 generated by Gencode Consortium (Version 15) and the SurePrint G3 Human lncRNA chip were used in this study. The array had 2 probes per lncRNA transcript and targeted 22,001 lncRNA transcripts and 17,535 randomly selected protein coding targets.

RNA samples from forty-four (44) paired NSCLC tumour tissues and adjacent normal tissues obtained from the LLP were reverse transcribed and cDNAs labelled with cy3 (tumour

tissues) and cy5 (normal tissues) cDNA probes respectively; and hybridized to a two-colour array platform.

The array data was processed and generated by the CGR, University of Liverpool and then analyzed by Dr. Russell Hyde.

2.2.4.1. Reverse Transcription

Reverse transcription of RNA samples obtained from tissues and cell lines was carried out using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems). 2X master mix was prepared by mixing 2 µl of 10X Reverse Transcription Buffer, 0.8 µl of 25X dNTP mix (100 mM), 0.6 µl of Anchored Oligo (dT) (100 µM), 1 µl of MultiScribe™ Reverse Transcriptase (50 U/µl) and 5.6 µl of nuclease-free water. RNA concentration was standardized to 1 µg per 20 µl concentration (50 ng/µl). 500 ng RNA to be reverse transcribed was denatured by heating at 70°C for 5 minutes and left to cool on ice for 2 minutes, after which 10 µl of 2X master mix was added to the RNA sample and mixed by pipetting and centrifugation. The reaction was performed using a thermal cycler with the following conditions: 25^{o}C for 10 minutes, 37^{o}C for 120 minutes, 85^{o}C for 5 minutes and 6^{o}C for ∞. cDNAs obtained after reverse transcription were diluted 5X and stored at -20^{o}C for further use.

2.2.4.2. Primer Design

qRT-PCR primers and probes were designed for lncRNA expression analysis in NSCLC tissues and cell lines using primer design software Oligo7 (MBI) (Table 2.3). Primer sets were optimized and run on a 2% gel to confirm the appropriate size.

2.2.4.3. Quantitative Real Time PCR (qRT-PCR)

RNA level analyses were carried out for selected lncRNA transcripts chosen from the microarray data. qRT-PCR assays were performed using the TaqMan^{TM} gene expression assays (Applied Biosystems). Expression assays were carried out in a final reaction volume of 15 µl in duplicates and contained 7.5 µl of 2X QuantiTect Expression Probe Mix (Qiagen), 0.75 µl of primer/probe mix of my targets (900 nM:250 nM), 0.75 µl TBP (0.70 µl for down-regulated lncRNAs), 3 µl of cDNA and the remaining volume made up with double distilled water (ddH_2O). Assay reactions were run on a 7500/7500Fast Thermal Cycler (Applied Biosystems) under the following conditions: - 95°C for 15 minutes (activation step), 50 cycles of 94°C for 15 seconds (denaturation step), varying annealing temperatures and 60°C for 45 seconds (extension). RNA levels were expressed as relative quantification (RQ) and calculated as: RQ=2^$^{-\Delta\Delta CT}$.

Table 2.3: Primers and probes designed for validation of lncRNAs discovered through microarray

lncRNA		**Primer Sequence**			
		Forward 5' - 3'	**Reverse 5' - 3'**	**Probe 5' - 3'**	**Ta (°C)[1]**
Induced	*FEZF1-AS1*	TGGCTATGGTTACTGCAATTC	CCATAAAGTCCAACCCTGAGT	AAAAGGCCTGTGAGGTGTGTCCC	53
	LINC01214	GCATACTTCTGGTAGCAATGG	TAGGGATTATGTGTCTTCATTCTG	ACCTCAAGTCCCCTTTGACCCG	54
	LOC105376287	CACCCTCCTCCACTGTCCT	CCTCCTGCTTTGTTTCCTGT	TCCCTGCACTAGGTCAGACAATCCC	54
	PCAT6	ACCCCACTTTCCAGCCTG	AGGGAGGCTCACGGACAC	CCAGATCTGCAGCCTTCGCCC	55
	LOC101927229	CTTGACACGACTTCAGAAGCCTC	GCAGAGCTCGACCAGGACAG	TCTGGCCCATCGTGGCATGGT	58
	LINC00673	GAAAGGACAAGAAAGAGGATGG	AGAGGTGGTCCAGCCTGA	TTCCCACCAGGAAGTTTAGCAGAACC	54
	NUTM2A-AS1	GGCTCATATGACATTAACAGACAA	TATCGCCTCCTGTACTATCAAAAT	AAGACAGGCAACGTGTTGGACCTTC	54
	RNF139-AS1	GCGACTGAAGGGCAAGAAC	CCAACTTGTGTTTTAGATGAGTCCT	CCCATAATGGCCTCTCTCCTTTTGCT	58
Repressed	*LANCL1-AS1*	GTGAAAGTATTCTCTGACTGCAA	TGCATTGGCCAGAACATA	TGACCACCTGTCTTTCTATATCAGAACCC	55
	FENDRR	GCTTCTGTCCAAGGCACT	CAAGCTTGCTAACTTCTTTGC	AGCCTACTCGTCAAAAGCCCGA	55
	LINC00968	CTACAGCAAGGCAACTTATCTCAC	TGGGAGGGAAGGATGACAA	TCACCAAGATATTCTGCACTTTCAGTGGC	57
	SVIL-AS1	ACCTTTGATCCAGAACTTGCAG	CTAAGGGGTGGCTGCATTC	TCTTCGGTTGTGAATCCGGCCC	59
	PCAT19	TGTTATTTGGCTGGAGTGAGG	AATTCATTCCACTGTAAGCCTTC	ATGAGTATCTCCAATGGTTCCCTGTTCTG	60

[1] Ta is the temperature at which the single stranded DNA anneal to the template in the RT-qPCR process.

2.2.5. DNA Methylation Analysis

2.2.5.1. Genomic DNA Extraction

DNA extraction was performed using QIAamp DNA Blood Mini Kit (Qiagen). Medium was discarded from the 75 cm^2 flasks and the cells were washed with 10 ml PBS twice. 500 µl of ATL buffer and proteinase K (stock solution 10 mg/ml) to a final concentration of 100 µg/ml were added and the cells were scraped off and transferred in a 2 ml tube. 500 ul of AL buffer was added and mixed thoroughly by vortexing and the samples were incubated at 56^oC for 30 minutes, after which 520 µl absolute ethanol was added in a 1:1 ratio and mixed thoroughly. The suspension was transferred to an IC spin column and centrifuged at 16,000 g for 1 minute, the flow-through was discarded and the collection tube was replaced with a new one. 700 µl of AW1 buffer (with ethanol) was added to the column, centrifuged for 1 minute at 16,000 g and flow-through was discarded. This was followed by addition of 680 µl of AW2 buffer (with ethanol), centrifugation at 16,000 g for 1 minute and flow-through discarded (this step was repeated). The IC spin column was transferred in a new collection tube and centrifuged for 4 minutes at 16,000 g to be completely dried out. The column was transferred in a 1.5 ml tube and 50 µl of AE elution buffer was added directly to the silica membrane, allowed to stand for 5 minutes at room temperature and then centrifuged at 16,000 g for 1 minute. The eluted DNA was collected and stored at -20^oC for further use.

2.2.5.2. DNA Quantitation

The NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific) was used to quantify and assess the purity of DNA. The procedure was similar to that of RNA quantitation; however, DNA was selected as the nucleic acid to be quantified. Measurement was taken for the

quantity of DNA in concentration (ng/µl) and the purity of the DNA samples represented by A_{260}/A_{280} and A_{260}/A_{230} ratios.

2.2.5.3. Bisulfite Conversion

Bisulfite treatment of DNA sample was carried out using the EZ DNA Methylation-Gold™ Kit (ZymoResearch) to assess the methylation status of the gene promoters. The CT Conversion Reagent was prepared following manufacturer's protocol. 900 µl of pre-warmed (60°C) water, 300 µl of M-Dilution Buffer and 50 µl M-Dissolving Buffer were added to a tube of CT conversion reagent. It was vortexed for about 2 minutes and left on the rotating mixer for further 8 minutes. 130 µl of the CT Conversion Reagent was added to 20 µl of DNA sample (1 µg) in a PCR tube, the samples were mixed by flicking the tube or pipetting up and down, and then spun down briefly. Afterwards, they were placed in a thermal cycler under the following conditions: - 98°C for 10 minutes, 64°C for 2.5 hours and 4°C storage up to 20 hours. At the expiration of PCR incubation time, the bisulfite-treated DNA was cleaned up. 600 µl of M-Binding Buffer was added to 150 µl of bisulphite conversion reaction and mixed thoroughly by pipetting in a 1.5 ml tube. The mix was transferred in a Zymo-Spin™ IC Column and placed in a collection tube. The sample was centrifuged at 16,000 g for 1 minute and the flow-through discarded. 200 µl of M-Wash Buffer was added to the column and centrifuged at 16,000 g for 1 minute. 200 µl of M-Desulphonation Buffer was added to the column and left to stand at room temperature for 20 minutes. Then, the column was centrifuged for 30 seconds at 16,000 g. 300 µl of M-Wash Buffer was added to the column and centrifuged at 16,000 g for 1 minute. Another 300 µl of M-Wash Buffer was added and centrifuged for an additional 4 minutes. The column was placed into a 1.5 ml tube and 50 µl of pre-warmed (65°C) M-Elution Buffer was added directly to the column matrix. The

column was incubated at room temperature for 5 minutes and centrifuged for 1 minute at 16,000 g to elute the DNA. The bisulphite converted DNA was stored at -20°C for about a week or -70°C until use.

2.2.5.4. Pyrosequencing Methylation Analysis

The pyrosequencing technique allows for the detection and quantification of DNA methylation at specific CpG sites within a specified region of interest in real time. The protocol involved the preparation of samples, the design of pyrosequencing primers and pyrosequencing.

2.2.5.5. Polymerase Chain Reaction (PCR)

The methylated DNA samples were amplified by PCR using two amplification primers forward biotinylated (Fb) and reverse non-biotinylated (R) in a 1:2 ratio (Table 2.4); and performed in a 25 µl volume containing 2.5 µl of coral load buffer, 1 µl of 5 mM dNTP, 1 µl of primer mix, 0.125 µl of HotStar Taq polymerase, 3 µl of DNA sample and ddH_2O. Thermal cycling conditions were set at 95^{o}C for 5 minutes and 40 cycles of 94^{o}C for 30 seconds, 58^{o}C for 45 seconds, 72^{o}C for 45 seconds and 72^{o}C for 10 minutes.

Table 2.4: LINE1 primer used for the evaluation of global methylation analysis

Primer Name	**Primer Sequence 5' – 3'**	**Modification**	**Tm(^{o}C[2])**
LINE1-Fb (Forward)	TAGGGAGTGTTAGATAGTGG	5'-	52.1
LINE1-R (Reverse)	AACTCCCTAACCCCTTAC	-	51.8
LINE1-S	CAAATAAAACAATACCTC	-	

[2] Tm is the primer melting temperature at which 50% of the DNA duplex becomes single stranded.

2.2.5.6. Agarose Gel Electrophoresis

DNA fragments were separated using 2% agarose gel electrophoresis. The gel was prepared by dissolving agarose in 0.5X TBE buffer. 1.5 µl of safe view nucleic acid was added to 65 ml (2%) agarose solution to visualize the DNA. To load the samples, the DNA was mixed in equal volume ratios with the agarose gel loading buffer. Electrophoresis was performed at 80 V for 60 minutes. DNA was detected using UV light and the size of the DNA was determined using 50 bp ladder. Successful bisulfite conversion was seen in the strong bands formed after agarose gel electrophoresis separation.

2.2.5.7. Primer Design

Prior to primer design, the DNA region of interest of specified length was verified and obtained from GeneBank (NCBI), and the primers (forward, reverse and sequencing) used in the pyrosequencing assay were designed using the PyroMark Assay Design Software.

Table 2.5: Pyrosequencing primers sequence of targeted lncRNAs of interest

lncRNA	Forward Primer (5' – 3')	Reverse Primer (5' – 3')	Sequencing Primer (5' – 3')	Ta (°C)[3]
FEZF1-AS1	TTTTGGGTTTGGTATTAGGA	BIO-CCCAAACTCAACTACAACATT	GGGTTTGGTATTAGGAG	54
PCAT6	TTAGTTAAGGGAGTTGATTGGT	BIO-CAAACTAAAATTACCCAAATCC	GTTGATTGGTAGGTAGT	53
LINC00673	TATATAAGGGTAGAATGGTTTAGT	BIO-CAAAACCACAAATACCAAAAAC	GTTTAGTAGATTGTAGGG	51
NUTM2A-AS1	GGAGGAAGAATGTAGGGAGA	BIO-AAACCAACACCAACCAATAA	GGGAGATTTTTAGTGGAT	57
RNF139-AS1	BIO-TATTTTAGGGAGTTTGAAAGT	ACCACACCCAACCTAATA	CCACACCCAACCTAATACC	52
FENDRR	BIO-TGTTTTTAGGAATTTGGTTTG	CCTTAACCTAATACCCCTATAAACTC	ACCTAATACCCCTATAAACT	53
SVIL-AS1	TGGTGTTAGTAAAGATATATAAGGAA	BIO-CCCTCTCAACTATTCAACAAC	GTTAGTAAAGATATATAAGGAA	55

4

[3] Ta is the annealing temperature.
[4] Letters in red represents methylated

2.2.5.8. Pyrosequencing

All reagents were allowed to equilibrate at room temperature before use, assay plate layout was set up using the PyroMark software and the pre-run information was calculated to determine the volume of enzyme, substrate and nucleotides required in the cartridge. 75 µl of Binding Mix (made up of 50 µl PyroMark Binding Buffer, 23 µl of ddH_20 and 2 µl Streptavidin- Sepharose (S-S) beads) was added to 20 µl of PCR product, mixed thoroughly and transferred to a 96 well plate. The plate was sealed with adhesive film and agitated on a monoshaker for 10 minutes at 350 r.p.m to ensure proper mixing of the S-S beads with the PCR products. 45 µl of annealing mix (made up of 43.5 µl PyroMark Annealing Buffer and 1.5 µl methylation sequencing primer (10 µM)) was prepared per sample and placed in appropriate wells in a "soft" 96-well plate. The PyroMark vacuum station trough was filled with 70% Ethanol, 0.2 M NaOH (denaturation solution), Tris-acetate (pH 7.5) and ddH_20. The Vacuum tool was turned on and immersed into ddH_20 for 30 seconds to allow the water to be flushed through the tool and switched off. It was then lowered into the 96 well plate containing the PCR products and the S-S beads and mixed gently to allow the samples to be attached to the probes. The vacuum tool was washed by immersing into 70% Ethanol, 0.2 M NaOH solution and Tris-acetate for 10 seconds respectively and held vertically to allow to drain. The tool with the S-S beads and PCR product was lowered directly into the 96-well plate containing the annealing mix, and gently agitated to release the beads. The 96 well plate was incubated on a heating block (Dri-Block DB-2A) at 80^oC for 2 minutes and left to cool for 3 minutes at room temperature. The 96 well plate was then transferred into the PyroMark Q96 ID and the assay run saved and started.

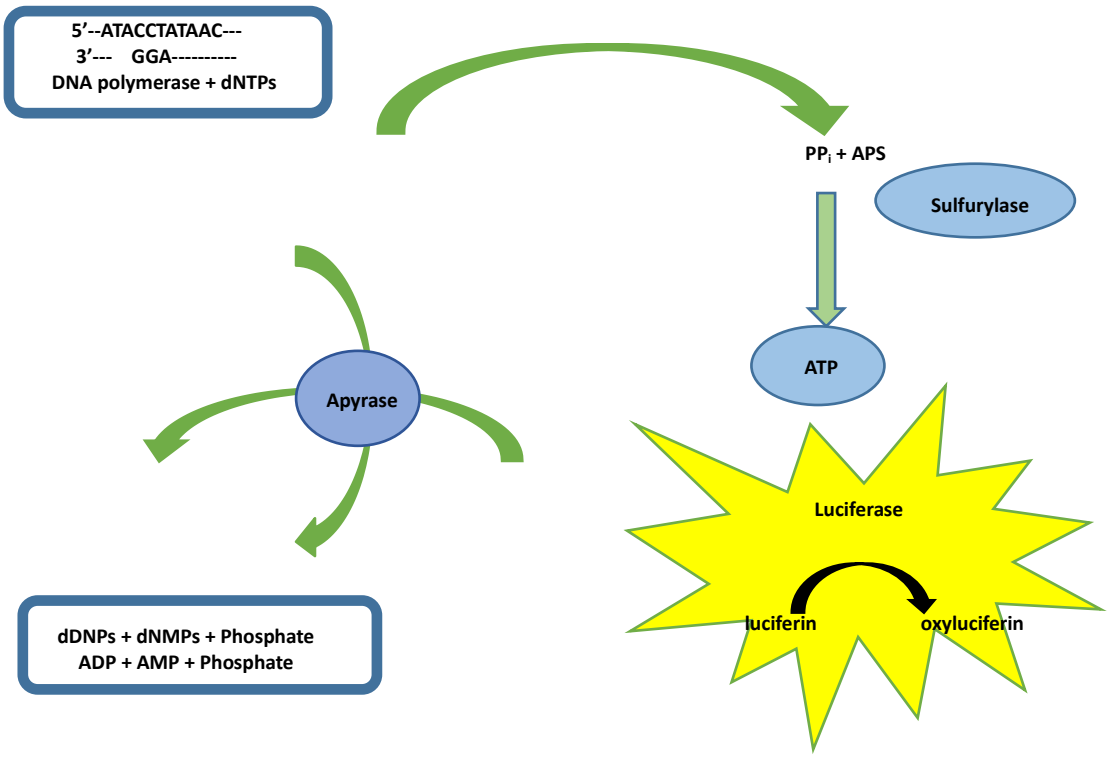


Figure 2.1: Illustration of pyrosequencing principle adapted from Coyler et al. 2012. The first nucleotide (dNTPs) is released and if complementary to the template strand, it is then incorporated by DNA polymerase, leading to the release of inorganic pyrophosphate (PPi) in a quantity that is equimolar to the amount of nucleotide incorporated. PPi is converted to adenosine triphosphate (ATP) by ATP sulfurylase in the presence of adenosine 5'-phosphosulfate (APS). Luciferase utilizes ATP and drives the conversion of luciferin to oxyluciferin, generating visible light in amounts proportional to the number of incorporated bases in the sequence, while unincorporated nucleotides are degraded by apyrase before the release of another base. The light produced is detected by a charge-coupled device (CCD) camera and seen as a peak in a pyrogram.

2.2.6. Decitabine Modulatory Effect on Global DNA Methylation

5-aza-2-deoxycytidine (Decitabine - DAC) and Valproic Acid (VPA) treatments were carried out on 4 NSCLC cell lines (A549, CALU6, H358 AND SKLU1) to assess global DNA methylation. 3 x 10^4 cells were seeded in four 12 well plates in triplicate and grown in 500 µl DMEM supplemented with 5% FBS. At 70% confluency cells were exposed to 100 nM DAC and 1

mM VPA and incubated at 37°C and 5% CO_2 for 48 hours (with medium replenished with the drugs in 24 hours). Old medium was vacuumed off, cells were washed with 1 ml PBS, RNA and DNA were extracted and stored at -80^{o}C for further use. Efficiency of DAC and VPA was determined by measuring global DNA methylation by pyrosequencing as shown in Section 2.2.6.4 and Figure 2.1.

2.2.7. Transfection

Preparation of Luria Broth (LB) Plates

LB medium (solid culture) was prepared in a class II cabinet by dissolving 5 g tryptone, 2.5 g yeast extract, 5 g sodium chloride (NaCl) and 7.5 g agar in 500ml of ddH_2O, autoclaved and allowed to cool to 50^{o}C prior to the addition of 100 µg/ml ampicillin. The medium was mixed thoroughly, and 20 ml transferred into each 10 cm^2 plate, swirled gently and left to set at room temperature before storing in sterile films at 4°C for further use.

Ligation using pTARGET Vector _ LINC00968

The pTARGET mammalian vector expression system purchased from Promega was used in this study. LINC00968 sequence was inserted into the pTarget vector (Promega, USA) (Figure 2.2).

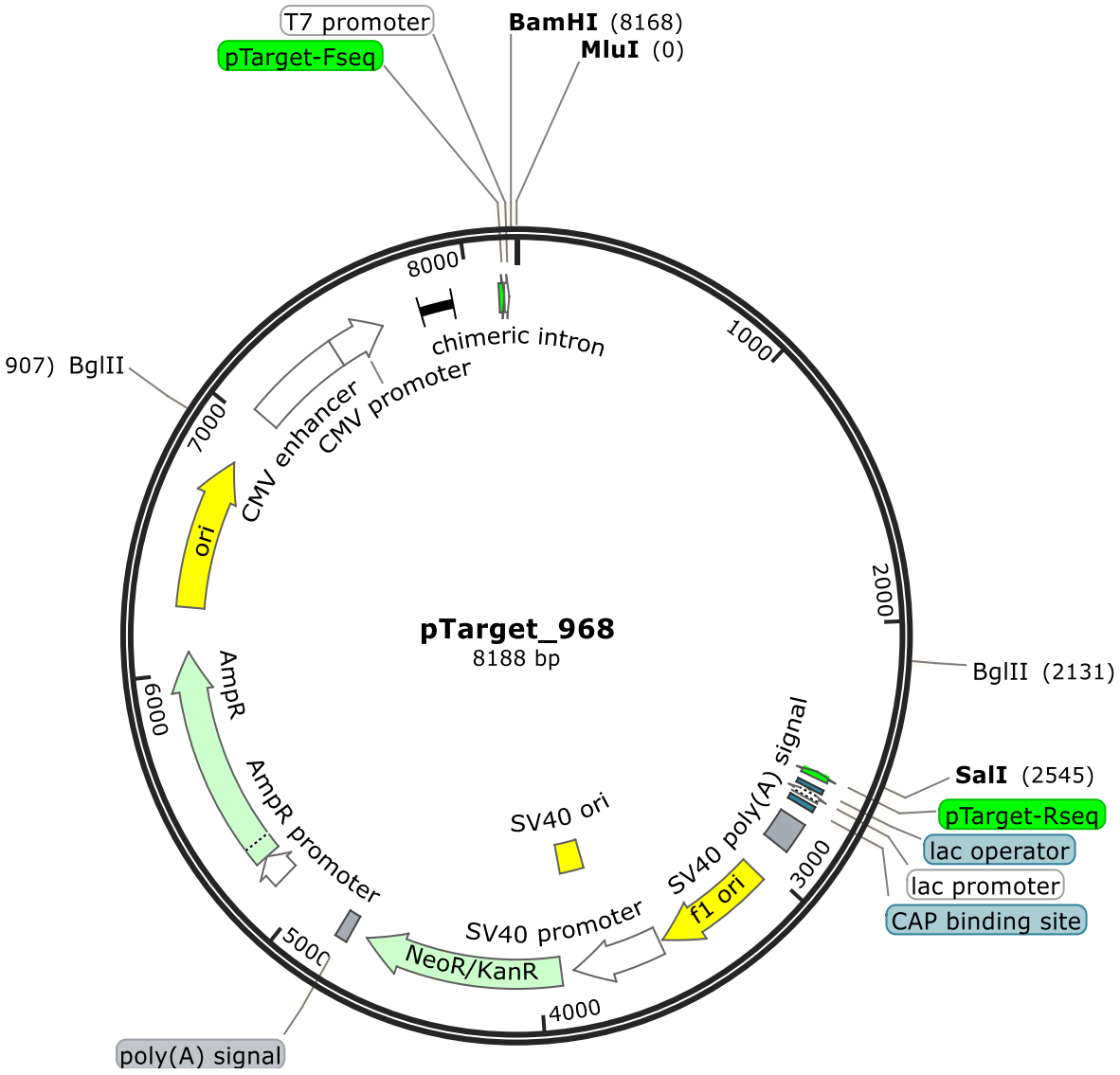


Figure 2.1: pTarget_968 vector. Figure extracted from SnapGene.

2.2.8. Transformation

Eschericia coli competent bacteria used for transformation was stored at -80°C prior to use. An aliquot of 50 µl *E. coli* competent bacteria was allowed to thaw on ice and 1 µl of DNA was added and incubated on ice for 30 minutes. The cells were subjected to heat shock by incubating in the water bath at 42°C for 30 seconds and returned to ice for 2 minutes. 900 µl of pre-warmed SOC medium was added to the cells and incubated at 37°C for 1 hour in a

shaking incubator. 200 µl of the reaction mix was spread onto each LB plate containing the ampicillin and incubated overnight at 37°C.

2.2.8.1. Plasmid DNA Extraction

Plasmid DNA was extracted using the Zyppy Plasmid Midiprep Kit (ZymoResearch) following manufacturer's protocol. Summarily, 6 ml of bacterial culture in LB medium was transferred into a 25 ml falcon tube, 1 ml of 7X lysis buffer (blue) was added to the culture medium and mixed by inverting the tube 2 - 4 times and left to sit at room temperature for 2 minutes. 5 ml of neutralization buffer (yellow) was then added gently to the mixture and inverted 4 - 6 times until a homogenous mix was achieved and the tube placed on ice for 5 minutes. The mixture was transferred into the Zymo-Midi Filter/ Zymo-Spin column assembly placed on a vacuum manifold and the vacuum switched on to allow the liquid to flow through. The Zymo-Spin column was transferred to a collection tube and centrifuged at >11,000 g for 30 seconds to remove all traces of the lysate. 400 µl of Endo-Wash Buffer was added to the zymo-spin column and centrifuged at >11,000 g for 30 seconds, this was followed by addition of 400 µl of Zippy Wash Buffer and centrifuged at >11,000 g for 30 seconds. The zymo-spin column was finally transferred to a 1.5 ml microcentrifuge tube and 150 µl of pre-warmed (65^{O}C) Zyppy Elution Buffer added directly to the centre of the tube and centrifuged at >11,000 g for 1 minute. Eluted plasmid DNA was quantified and dried out at 65^{O}C for 30 minutes and stored at -20^{O}C for further use.

2.2.8.2. G418 (Geneticin) Kill Curve

To use G418 as a selection marker, the optimal working concentration was determined by carrying out a kill curve. A549 cells were plated in triplicate in 800 µl DMEM supplemented with 5% FBS and 10X Penicillin-Streptomycin (pen-strep) per well in a 24 well plate and

incubated at 37^{o}C until 70% confluent. Cells were then exposed to varying concentrations (400 - 900 µg/ml) of G418 including control and observed for 7 days (with replacement of media/G418 every 48 hours).

2.2.8.3. Stable Transfection with pTarget_968

A549 cells were transfected using the FuGENE HD Transfection Reagent following manufacturer's protocol. 20 µg of Plasmid DNA was resuspended in 100 µl of sterile ddH_20, mixed by vortexing and incubated at 4^{o}C overnight. A549 cells were seeded at a very low density in a 10 cm^2 dish to achieve 50% confluency before transfection. At 50% confluency, the old medium was vacuumed off, the cells were washed with 10 ml PBS and replaced with 8 ml DMEM supplemented with 5% FBS and 10X Pen-Strep. Cells were transfected with the transfection reagent mix (Table 2.6) dropwise into the dish for even distribution and incubated at 37°C and 5% CO_2 for 24 hours.

Table 2.6: Transfection Reagent Mix

Reagent/ Sample	**Volume**	Reagents were mixed in a tube, vortexed, spun briefly and incubated for 7 minutes at room temperature before use.
Plasmid DNA (4 µg)	20 µl	
DMEM + FBS	368 µl	
*FuGENE HD	12 µl	
Total	**400 µl**	

*FuGENE to plasmid DNA ratio was 3:1

After 48 hours, the cells were washed with 10 ml PBS and 8 ml DMEM containing 2.5% FBS, 10X Pen-Strep and 700 µg/ml G418 was added to select the resistant clones (this process was repeated until single clones were formed). 12 single clones were isolated using sterile polystyrene clone rings, trypsinized with Trypsin-EDTA and 2.5% FBS, transferred to a properly labelled 24 well plate and incubated for 24 hours. Following which, the clones were maintained with DMEM containing 5% FBS, 10X pen-strep and 500 µg/ml G418. The clones

were further harvested and grown in 25 cm^2 and 75 cm^2 flasks respectively for expansion and further use. Successful transfection was confirmed by qRT-PCR as shown in Section 2.2.4.3.

2.2.9. Phenotypic Analysis

2.2.9.1. MTT Assay

The viability of the A549 parent cell line and overexpression derived clones were measured using the MTT (3-(4, 5-Dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide) assay. A549-LINC00968 cell clones and A549 parent cell line were counted and 30,000 cells plated per well in a 48-well plate and incubated for 24 hours at 37^{o}C and 5% CO_2. After incubation period, the medium was vacuumed off and the cells washed with PBS. Fresh medium containing 0.75 mg/ml MTT was added to the cells and incubated for up to 4 hours at 37°C and 5% CO_2, until purple formazan crystals are visible under a microscope. Medium containing MTT was then vacuumed off and cells were solubilized by addition of 200 µl/ well of equimolar ratios of DMSO and isopropyl alcohol, and incubated at 37°C for 2 hours, until cells have lysed and purple crystals have dissolved. Absorbance was measured at 595 nm with 630 nm as reference using a GENios Microplate Reader at 8, 24, 48, 72 and 96 hours respectively.

2.2.9.2. Soft Agar (Anchorage Independence) Assay

The soft agar assay made up of a top and a base layer of soft agar was used to monitor Anchorage-Independent Cell Growth. The base layer was prepared by mixing 1% agarose (Noble Agar) and 2X DMEM containing 10% FBS, while the top agar layer was prepared by dissolving 0.6% agarose (Noble Agar) in 1X DMEM supplemented with 10% FBS (both agar solutions were sterilized by autoclaving before use). A549_LINC00968 cells were counted

and 3 x 10^4 cells from each A549_LINC00968 clone were added to the top agar mixture and plated in a 10 cm^2 dish and incubated at 37°C for 21 days. Cells were washed with 10 ml PBS (Sigma-Aldrich), colonies stained with 200 µl of nitroblue tetrazolium chloride (Thermo Fisher Scientific) and counted.

2.2.9.3. Wound Healing Assay

The wound healing assay also called the scratch assay was employed to determine LINC00968 involvement in cell migration and interaction. Selected A549_LINC00968 clones and the A549 parent cell line were plated in 24 well plates and incubated at 37°C and 5% CO_2 until confluent as a monolayer. A cross-shaped scratch (like a wound) was made on the cell surface area at an angle of 30^oC using a p200 pipette tip and the wells washed gently twice with 1 ml DMEM to remove detached cells. Fresh medium was added, the location of the scratch was marked at the bottom of the 24 well plate as reference point, and the migration rate of the cells was monitored and quantified using the EVOS Cell Imaging System at 0 hours and subsequent intervals.

2.2.9.4. Boyden Chamber Assay

The invasive potentials of some selected A549_LINC00968 clones and the A549 parent cell line were studied using an adaptation of the Boyden Chamber Assay involving the use of ThinCert Cell Culture Inserts (24-well plate) and Geltrex^{TM} matrix. For the invasive study, Geltrex matrix was allowed to thaw and mixed by pipetting up and down gently. 100 µl of the Geltrex matrix was used for coating each ThinCert Cell Culture chamber and incubated at 37^oC and 5% CO_2 for 60 minutes for gelling. 200 µl of 2 x 10^4 cells maintained in serum-free DMEM medium were seeded into each ThinCert Cell Culture chamber, transferred into a 24-well plate containing 750 µl of DMEM supplemented with 5% FBS and incubated for 16

hours at 37^{o}C and 5% CO_2. Following incubation, medium was discarded and the chambers were washed with PBS, fixed with 70% ethanol and washed with PBS before staining with 0.2% crystal violet. Non-invasive cells were cleaned off with a cotton swab and invasive cells visualized and counted using the EVOS Cell Imaging System. Same protocol was applied for the cell migration study (without the Geltrex coating in the ThinCert Cell Culture chambers).

2.2.9.5. Oxidative Stress Analysis

A549_parent cell line and the selected A549_LINC00968 clones were subjected to varying concentrations of Hydrogen Peroxide and cell survival measurement taken after 24 hours using the MTT assay.

2.2.9.6 Drug Treatment Assay

Select anti-cancer drugs used currently in NSCLC therapy were tested on A549_LINC00968 clones and the parent cell line, and their response monitored. A549_LINC00968 clones and the A549_parent cell line were maintained in culture as stated above (2.2). 2 x 10^3 cells were seeded into 48 well plates in DMEM supplemented with 5% FBS and Pen-Strep and incubated at 37^{o}C and 5% CO_2 until cells were 70% confluent, carried out in 4 technical replicates. The old medium was discarded and replaced with new medium containing the compounds of interest. After 72 hours of exposure to the drugs, cell survival was measured via MTT assay as stated in Section 2.2.9.1.

2.2.10. Data Interpretation and Statistical Analyses

All statistical analyses were carried out using SPSS v24.0 (IBM, USA). The qRT PCR results were quantified as fold change between TBP and target lncRNAs using the RQ formula $2^{-\Delta\Delta CT}$. The Wilcoxon Signed-Rank non-parametric test for paired samples was used for analysis of the target lncRNA expression in NSCLC tumour compared to normal tissues, while the

association between the expression of target lncRNAs and the clinicopathological parameters of the patients was analyzed using the Mann-Whitney test. P values < 0.05 were considered significant.

3. DISCOVERY AND VALIDATION OF DEREGULATED lncRNA TRANSCRIPTS IN PRIMARY NSCLC TISSUES

3.1. Introduction

lncRNAs have been implicated in the development and progression of cancer, with a recent study reporting a total of 19,175 potentially functional lncRNA genes in the human genome (Hon *et al*., 2017). The urgent need for the development of biomarkers for the early detection of NSCLC and the need for novel drug therapies to combat the high mortality rate of lung cancer has led to this current investigation.

3.2. Bioinformatic Analysis

At the beginning of the study and prior to our own microarray analysis, we carried out a bioinformatics analysis involving gene ontology and pathway analysis in order to identify lncRNAs with high frequency of deregulation in lung cancer. Data was sourced from publicly available databases such as Oncomine and The Cancer Genome Atlas (TCGA) databases (overseen by my 3rd supervisor, Dr. Russell Hyde).

The analysis showed 13 transcripts (lncRNAs and pseudogenes) deregulated in NSCLC. We attempted to validate this result by RT-qPCR analysis of *COL6A4P1* (*DVWA/344875*) and *LINC00669* (*LOC647946*) transcripts in 48 paired lung samples obtained from the Liverpool Lung Project (LLP) Biobank. *COL6A4P1* transcripts were significantly overexpressed in lung tumour samples compared to their adjacent normals (Wilcoxon's test $p < 0.001$), expression studies for *LINC00669* were inconclusive. Results obtained for *COL6A4P1* when compared with histological and clinicopathological data showed no significant difference between histology subtypes, differentiation and tumour/nodal status(See Appendix I & II) Both the *LINC00669* and *COL6A4P1* were however, not identified in our microarray analysis.

The first objective of this study was to evaluate the transcript levels of novel lncRNAs in primary NSCLC tissues in comparison to adjacent normal tissues and explore their biomarker potential. To achieve this objective, we carried out a microarray analysis to identify novel lncRNAs in NSCLC, we then examined the levels of these transcripts in NSCLC tumour and normal tissues by qRT-PCR and further explored the correlation of the levels of these transcripts with the clinicopathological parameters of the patients.

3.3. Microarray Analysis

To identify novel lncRNAs deregulated in NSCLC tissues, RNA samples obtained from a set of 44 paired NSCLC tumour and adjacent normal samples provided by the LLP were screened for their quality. We carried out RNA quantitation and bioanalysis to select RNA samples with high RIN (RNA Integrity Number), before samples were sent for Microarray.

Bioinformatic analysis was undertaken by Dr. Russell Hyde and the raw data was analysed using the *limma* (linear models for microarray) package (Ritchie *et al*., 2015).

The two elements considered in the selection of the lncRNAs in this study were the consistency of the difference in each lncRNA in the tumours (expressed as the pValue) and the actual magnitude of the differences (log fold change). Following the array analysis, 20 induced and 30 repressed lncRNAs were identified. (Appendix III and IV).Using Ensembl and NCBI BLAST (Basic Local Alignment Search Tool), we further selected 13 lncRNAs with > 80% homology (sequence similarity), and designed primers and probes for further analysis. Selected lncRNA transcripts discovered to be deregulated from the microarray analysis are shown in Table 3.1.

Table 3.1: Selected top deregulated RNAs identified by microarray

lncRNA		Target ID (Ensembl)	*Probe sequence on microarray 5' – 3'
Induced	*FEZF1-AS1*	ENST00000428449.1	CACGGTTCGACTGTTTCCTTGACACTACCCACGAAGTTTAAAGCATTTTTTATGTTATTT
			GAAACAAGCAGACACACACACAAAACAACCATATTCAAGACACGGTTCGACTGTTTCCTT
	LINC01214	ENST00000471222.2	ACTTGAGGTTATCAGCTTTCAGAATGAAGACACATAATCCCTATTAGAGTGGAAGACTTA
			ATGGGGACACGGGTCAAAGGGGACTTGAGGTTATCAGCTTTCAGAATGAAGACACATAAT
	LOC105376287 (HMGA1P4)	ENST00000428643.1	TTTGTCCCAGCCTGGGGCTCCCCCTCTGGTTTCCTATTTGTAGTTACTAGAATGAAAAAA
			GTTTGTCCCAGCCTGGGGCTCCCCCTCTGGTTTCCTATTTGTAGTTACTAGAATGAAAAA
	PCAT6	ENST00000425295.1	TATTTTGTGTAGTCCTACAACGTCTTGTTACTACCCCCTATTACAACACTTATAACTCAG
			TTATTTTGTGTAGTCCTACAACGTCTTGTTACTACCCCCTATTACAACACTTATAACTCA
	LOC101927229 (592405)	ENST00000592405.1	CCAGATCACTTCTGCATAAAACCACATGGAAGAATAAGAAGGGAAAAACAATCAAGCAAT
			ACCAGATCACTTCTGCATAAAACCACATGGAAGAATAAGAAGGGAAAAACAATCAAGCAA
	LINC00673 (LINC00511)	ENST00000580948.1	CCTCTTGAGAGGCAGGAGCTCTGGATTTGATCAAGAATTCTTTGCTGAGCATGGTGCCTC
			CTCTTGAGAGGCAGGAGCTCTGGATTTGATCAAGAATTCTTTGCTGAGCATGGTGCCTCA
	NUTM2A-AS1	ENST00000413722.1	AGACTTGGACCGATGCTTGCATTTGTTTCTGAGAATTAAACGTTATGTTTTCTTCAATCC
			TGACAAAGGTTTGGGAAGGAAGACTTGGACCGATGCTTGCATTTGTTTCTGAGAATTAAA
	RNF139-AS1	ENST00000519861.1	GAGGAGATAGGTGAGTACTATTATATTATGTTCATGGAAAATGAGTTCATGGGCTTTCCT
		ENST00000530778.1	AGGAGATAGGTGAGTACTATTATATTATGTTCATGGAAAATGAGTTCATGGGCTTTCCTC
Repressed	*LANCL1-AS1*	ENST00000420418.1	TGAAAGTATTCTCTGACTGCAAGTATGACCACCTGTCTTTCTATATCAGAACCCAAGCTA
			GTGAAAGTATTCTCTGACTGCAAGTATGACCACCTGTCTTTCTATATCAGAACCCAAGCT
	FENDRR	ENST00000595886.1	TAAAAATGCTAGAAGCTTTAGTCATAGAATTACCATATGATACAGCCTACTGCAGAGTCC
			GGCTCTGTTTCATGCTGACTTTACCATCATGTAAGCAGTTTTAAAAATGCTAGAAGCTTT
	LINC00968	ENST00000499425.1	TAACTTCATAAGAAAATGATAAGAAAAATGTGGTCACCGGTATCATGAGGAACTTCCAGA
			CAACTTGCAGACAGGCAAGAATTTGGCTTTGTATGTTTAAAAATATGTACATCTTGGTTC
	SVIL-AS1	ENST00000446807.1	TGCTTTGGAATCTGGCATTTATGTTTTGAAGCGTTGTGAGCAAGCGATGTGGCAGATTGC
			TTGAAAGGGCTCTTGGTTTCTGCTTTGGAATCTGGCATTTATGTTTTGAAGCGTTGTGAG
	PCAT19	ENST00000594315.1	ACAAAAATAAAGTGTTATTTGGCTGGAGTGAGGTCTCATGTCTGCTTATGCGGTGGCTCG
			TACAAAAATAAAGTGTTATTTGGCTGGAGTGAGGTCTCATGTCTGCTTATGCGGTGGCTC

***Probe sequence is specific for the different ProbeID not shown in the table. Letter Cs in red indicate methylated cytosines.**

3.4. Validation of Deregulated Transcripts in NSCLC Tumour and Normal Tissues

The 13 lncRNAs identified from the microarray analysis in the discovery phase where validated by qRT-PCR in a total of 67 NSCLC tumour and adjacent normal tissues. The 67 samples were made up of 29 tumour and normal paired samples taken from the 44 samples hybridized in the microarray (termed "technical validation set"), and 38 samples were from an independent set of frozen tumour and normal paired samples from NSCLC patients (termed "the biological validation set") (Table 3.2). RNA was extracted from the samples and reverse transcribed as described in Section 2.2.3. Primers and probes were designed for the different induced and repressed lncRNAs respectively and RT qPCR performed as shown in Section 2.2.4.3.

Table 3.2 Clinicopathological data of patients used for expression analysis of discovered lncRNAs

	Technical set	**Validation set**	**Total**
N	29	38	67
Age			
mean (s.d.)	66.4 (8.5)	67.6 (8.4)	67.1 (8.4)
Gender			
Male:Female	21:8	29:9	50:17
Histology			
Adenocarcinoma	14	12	26
Squamous cell	15	24	39
Other	-	2	2
Tumour stage			
T1	1	4	5
T2	21	30	51
T3	5	3	8
T4	2	1	3
Nodal stage			
N0	15	19	34
N1	9	12	21
N2	5	7	12
Differentiation			
1	8	9	17
1.5	1	2	3
2	17	16	33
2.5	-	2	2
3	3	4	7
Missing	-	5	5

A total of 67 patient samples were used (of which 29 were from the discovery set and 38 for the biological validation set, as earlier stated). The mean age was 66.7 years. 74.6% of the patients were males and 25.3% were females. In the technical validation set, equal numbers of adenocarcinoma (14) and squamous cell carcinoma (15) subtypes were observed, while adenocarcinomas were less than the squamous cell carcinoma subtype in the biological validation set (12 and 24 respectively). Most of the samples were shown to be resected at stage IB and IIA based on the TNM staging (8^{th} edition Detterbeck *et al*., 2017), however, this was as a result of the amount of T1 samples made available by the pathologists. 17 out of the total 67 patient samples were well-differentiated tumours, while 33 and 7 were moderately and poorly differentiated respectively. Information about differentiation for 5 patients in the validation set was missing.

From our findings, 8 out of the 13 lncRNAs qualified in the microarray analysis as upregulated, were significantly elevated in lung tumours compared with paired normal tissues both in the technical and biological sets (Figure 3.1 – Figure 3.4). *FEZF1-AS1* was identified as upregulated in the microarray analysis; interestingly, validation by qRT PCR also showed significant elevation of *FEZF1-AS1* in NSCLC tumour tissues when compared to the corresponding normal tissues in both the technical and the biological validation sets ($p= 4.8 \times 10^{-6}$ and 6.6×10^{-7} respectively Figure 3.1)**.** Increased levels of *LINC01214* can be observed for tumours in both sample sets ($p= 5.9 \times 10^{-6}$ and 2.63×10^{-6} Figure 3.1). No significant difference was observed in *LOC105376287* levels in NSCLC tumours and adjacent normal tissues (p= 0.754 and 0.765, Wilcoxon Signed-Rank test Figure 3.2). *PCAT6* was significantly upregulated in tumours compared to normal tissues ($p=2.56 \times 10^{-6}$ and 5.26×10^{-6} Figure 3.2). *LOC101927229* levels was shown to be induced in tumour tissues in comparison to

adjacent normals (p=0.018 and 0.032 Figure 3.3). There was a significant difference in *LINC00673* levels in NSCLC tumours compared to the normal tissues (p=3.17 x 10^{-6} and 8.30 x 10^{-6} Figure 3.3). *NUMT2A-AS1* was upregulated in tumours compared to normal tissues (p=0.004 and 0.008 respectively for technical and biological validation sets Figure 3.4*).* There was no significant difference in the expression of *RNF139-AS1* in tumours compared to the normal tissues in the technical validation set (p=0.325), however, it seemed to be considerably elevated in the biological validation set (Wilcoxon Signed-Rank Test, p=0.03 Figure 3.4).

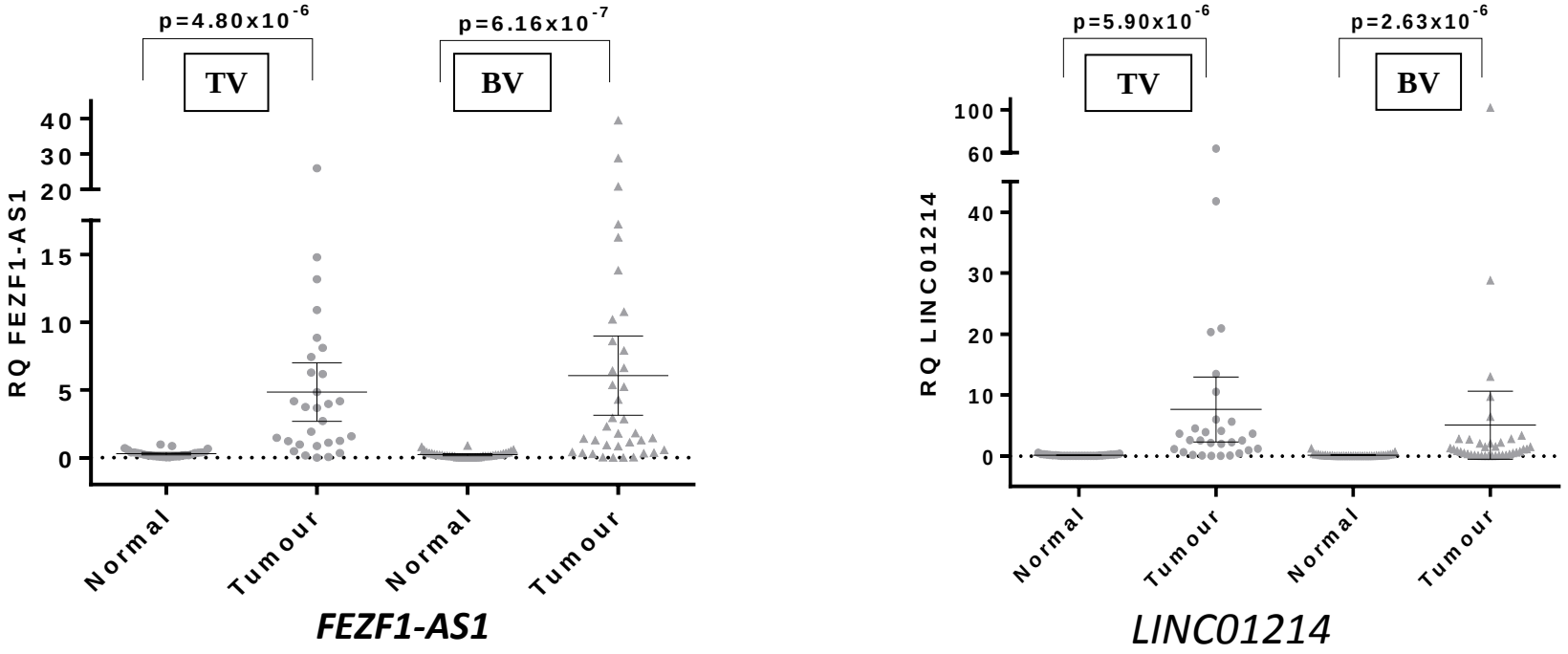
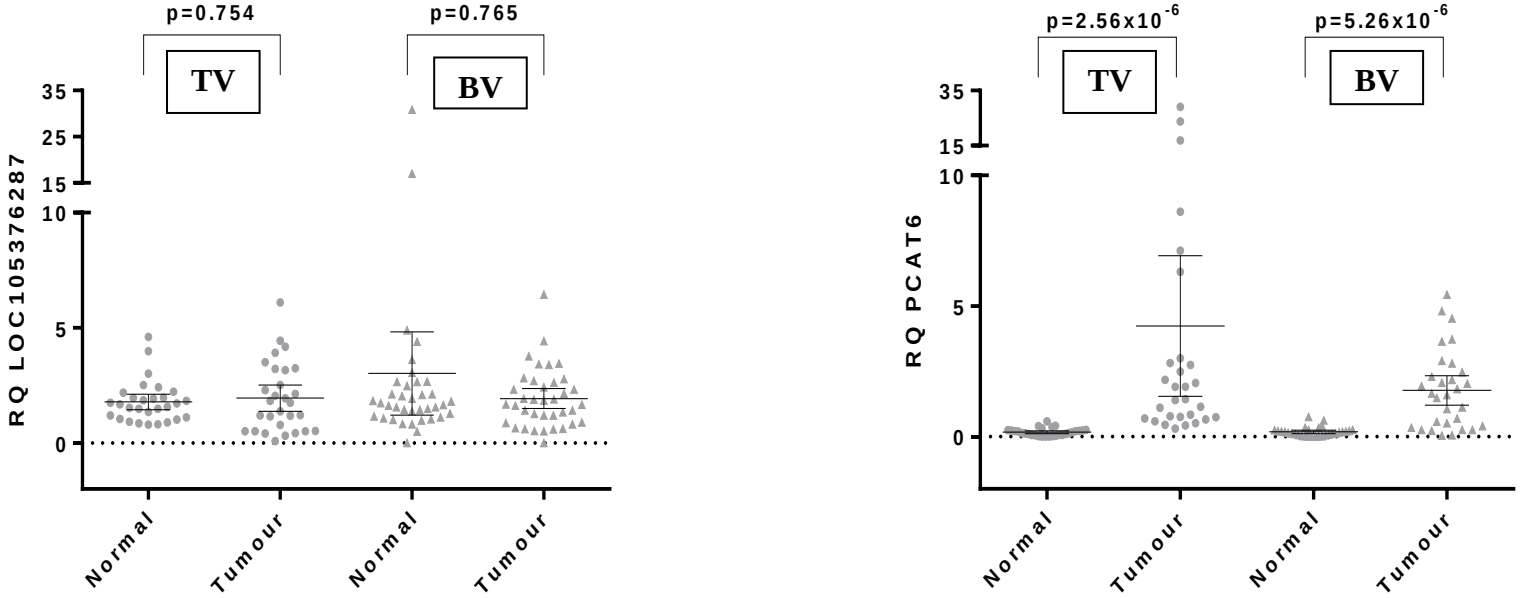


Figure 3.1: Levels of identified upregulated lncRNA *FEZF1-AS1* and *LINC01214* following RT qPCR analysis. *FEZF1-AS1* level in tumours compared to adjacent paired normals is shown to be significantly upregulated both in the technical validation (TV) sets and the biological validation (BV) sets (n=67) p-Values less than 0.05 were considered significant and were derived from Wilcoxon signed-rank test.



LOC105376287 *PCAT6*

Figure3.2: *LOC105376287* though identified as upregulated in the microarray analysis, did not show significant difference between the tumours and paired normals both in the TV and BV sets. *PCAT6* levels were shown to be significantly elevated in lung tumours compared to paired normal tissues both in TV and BV sets. (n=67), with p-values less than 0.05 considered as significant and derived from Wilcoxon signed-rank test.

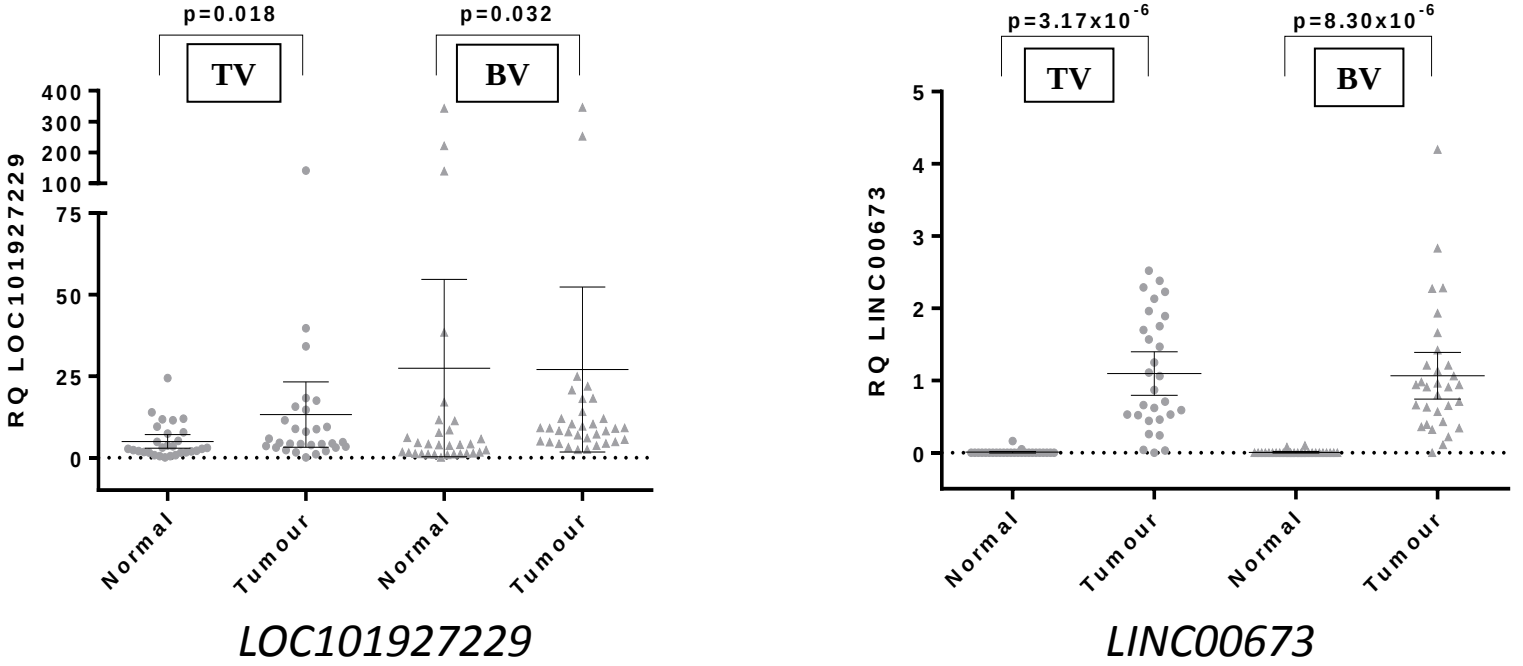


Figure3.3: *LOC101927229* and *LINC00673* levels are shown to be significantly upregulated in tumours compared to paired normal tissues both in the TV and BV (n=67)(Wilcoxon signed=rank test).

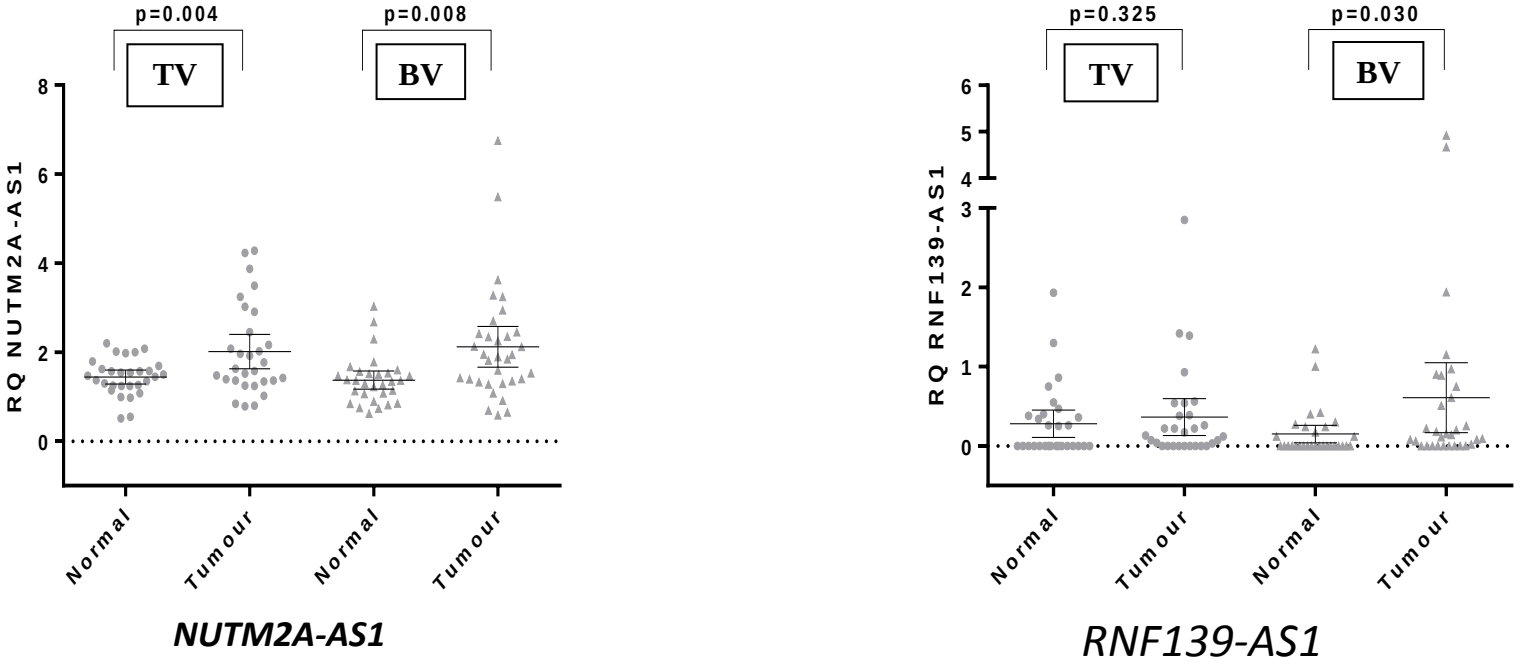


Figure 3.4: Scatterplot graph showing *NUTM2A* and *RNF139-AS1* levels in tumours compared to adjacent paired normals. A significant upregulation can be observed in *NUTM2A* levels in both TV and BV sets, and in TV sets of *RNF139-AS1*. (n=67, p value 0f 0.05 considered as significant).

The levels of the repressed lncRNA transcripts that were discovered were also validated in the same technical and biological validation sets. All the selected lncRNAs identified as repressed, were significantly reduced in tumours compared to paired normal tissues (Figure 3.5 – 3.7).

LANCL1-AS1 was significantly downregulated in tumours compared to paired normal tissues in both technical and biological validation sets, with p=2.56 x 10^{-}6 and p=1.14 x 10^{-}7 respectively Wilcoxon signed-rank test (Figure 3.5). Expression of *FENDRR* was repressed in tumours in both validation sets as shown in Figure 3.5 (Wilcoxon Signed-Rank Test p= 2.56 x 10^{-}6 and 4.54 x 10^{-}7). This result is consistent with the findings from the microarray analysis shown in Table 3.1. Similar results were observed for *SVIL-AS1* and *PCAT19*. Both genes were also significantly downregulated in tumour tissues compared to paired normal tissues (Figure 3.6 and 3.7). *LINC00968* was found to be downregulated in tumours compared to the paired normal and this was observed both in the technical and biological validation sets (Wicoxon signed rank test p= 3.79 x 10^{-6} and 5.6 x 10^{-5} respectively).

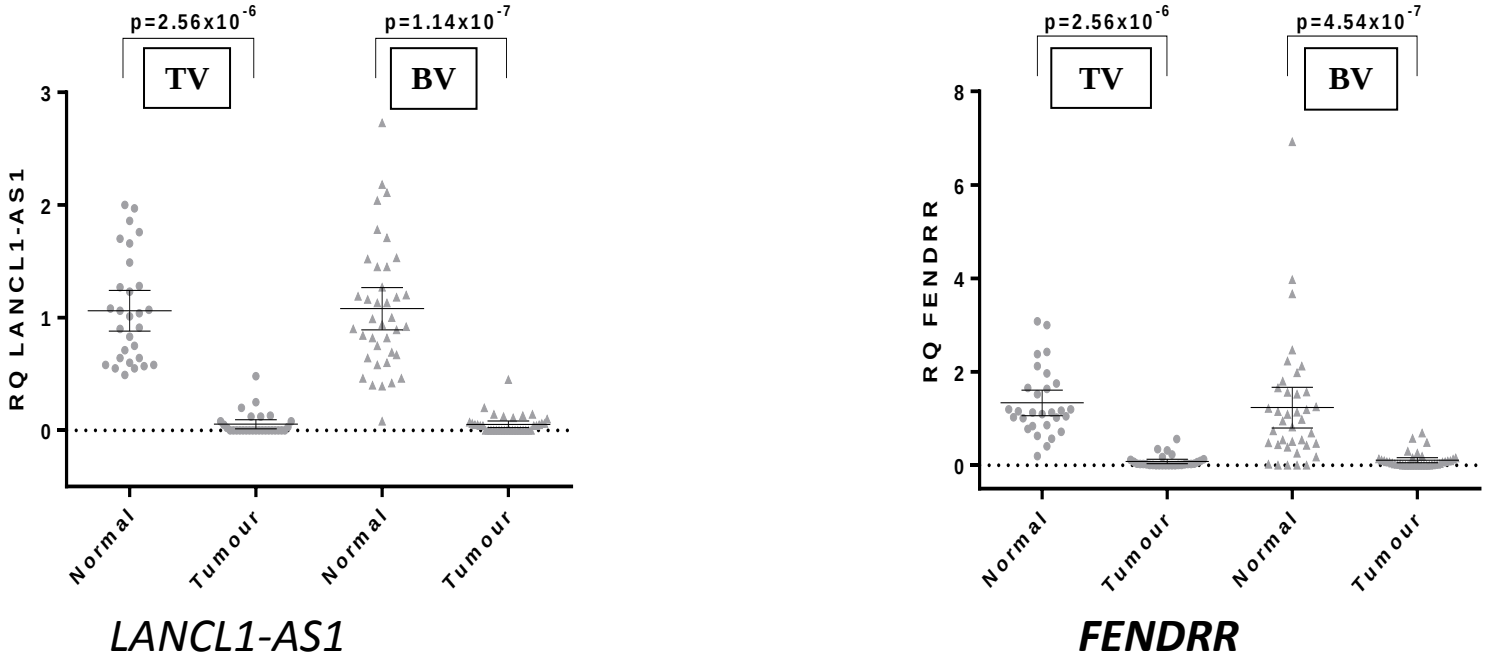


Figure 3.5: Scatter plot showing significant downregulation of repressed *LANCL1-AS1* AND *FENDRR* following RT-qPCR analysis both in the TV and BV sets. p values less than 0.05 were considered significant and were derived from Wilcoxon signed-rank test.

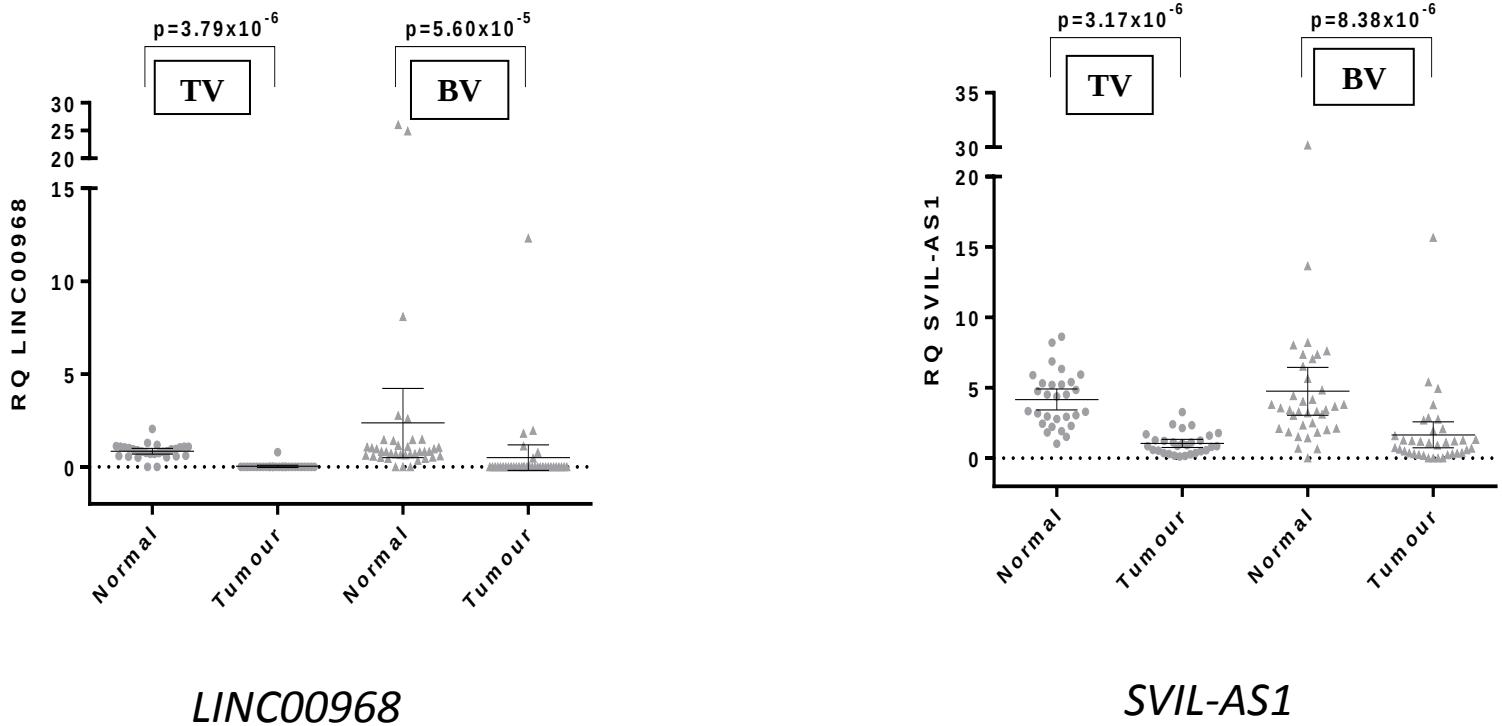
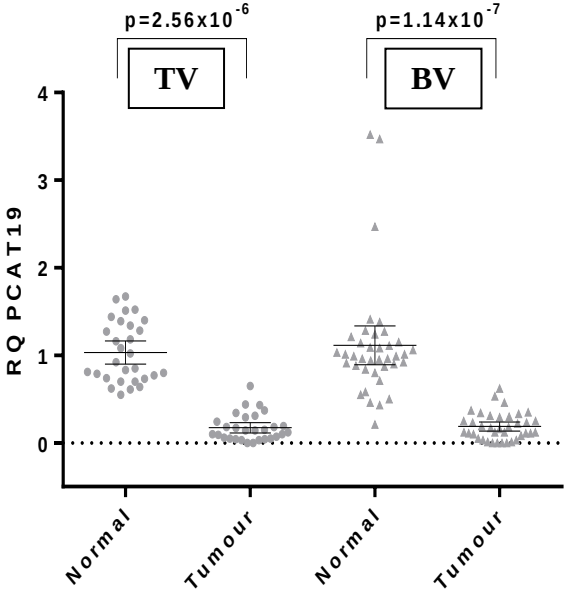


Figure 3.6: Repressed lncRNAs *LINC00968* and *SVIL-AS1* are shown to be significantly downregulated in tumours compared to adjacent paired normal tissues in the TV and BV sets following RT-qPCR analysis. P values were derived from Wilcoxon signed-rank test.



PCAT19

Figure 3.7: *PCAT19* is shown to be also significantly downregulated in tumour tissues compared to adjacent paired normal tissues in TV and BV sets. p values less than 0.05 considered significant (Wilcoxon signed-rank test).

3.5. Clinicopathological Correlation with Discovered lncRNAs Expression

The clinicopathological parameters (age, gender, histology subtype, TNM stage and degree of differentiation) of the patients were analysed using SPSS v24 and the levels of identified lncRNA transcript was correlated with patient clinicopathological characteristics using the Mann-Whitney's test. There were no significant differences in the clinicopathological parameters of patients in the discovery set and the validation set, hence both were combined to increase our study power. The levels of all validated lncRNAs were independent of age and histology, however, *FEZF1-AS1* was observed to be elevated in males compared to females (Mann Whitney's Test, p=0.03).

We also found that *LINC00968* levels were significantly higher in patients with lymph node involvement (N1 or N2 stages when compared to N0), while the opposite relationship was found for *LOC101927229* (Table 3.3). Additionally, both *PCAT6* and *LANCL1-AS1* showed significant difference between the tumour grades (*PCAT6* and *LANCL1-AS1* levels were significantly lower in well-differentiated tumours when compared to poor or moderately differentiated ones (Table 3.3).

Table 3.3: Clinicopathological correlation of lncRNA expression with patient characteristics

	lncRNA		Mean RQ	Std. Deviation	Percentiles			Mann-Whitney test	
					25th	50th	75th	U-score	p-value
Gender	*FEZF1-AS1*	Male	6.49	8.31	1.03	3.66	8.47	268	**0.03**
		Female	2.72	3.68	0.23	1.22	4.75		
Lymph node involvement	*LOC101927229*	N(-)	31.63	74.00	4.59	9.05	17.79	339	**0.05**
		N(+)	8.00	7.39	3.75	5.01	9.29		
	LINC00968	N(-)	0.20	0.50	0.00	0.00	0.00	435	**0.03**
		N(+)	0.38	2.18	0.00	0.00	0.00		
Differentiation	*PCAT6*	Well	2.03	3.85	0.50	0.73	1.96	212	**0.02**
		Poor-Mod	3.60	5.83	0.82	2.04	3.16		
	LANCL1-AS1	Well	0.02	0.05	0.00	0.00	0.00	262	**0.01**
		Poor-Mod	0.07	0.11	0.00	0.04	0.10		

3.6. Profiling NSCLC cell lines for expression of identified lncRNAs

Profiling of the levels of identified lncRNAs was carried out in 9 NSCLC cell lines, and HBEC3KT and its isogenic derivatives by qRT PCR, to select cell lines suitable for downstream genetic manipulation. A representation of the levels of some of the lncRNAs in the different cell lines is shown in Figure 3.8 and 3.9. The levels of the selected lncRNAs were variable in the lung cancer cell lines. *FEZF1-AS1* was high in SKLU1 and LUDLU1, moderate in CALU3 and H358 but was not detectable in A549, CALU6, LUNG 14 and the HBECs. Although *LINC01214* was upregulated in NSCLC tumours, interestingly its level was observed to be high in H358 only, and present at low levels in all other NSCLC cell lines and the HBECs. *LOC105376287* was highly expressed in H358 and LUDLU1, but moderately expressed in LUNG14 and HBEC3KT 53.

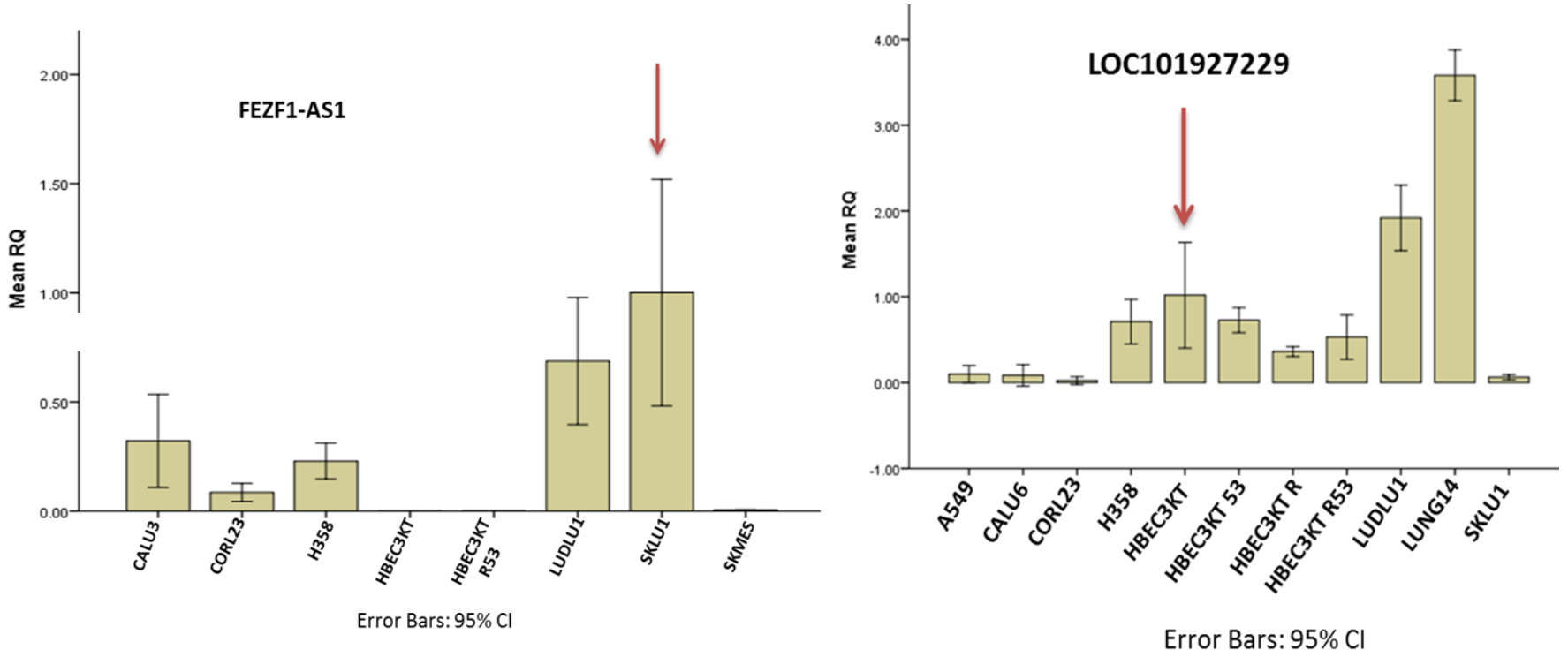


Figure 3.8: Bar chart showing the level of upregulated lncRNAs *FEZF1-AS1* and *LOC101927229* in NSCLC cell lines, HBEC3KT and its isogenic derivatives (HBEC3KT 53, HBEC3KT R and HBEC3KT R53). Red arrow indicates reference cell line for calculation of relative quantity (RQ=1), For *FEZF1-AS1*, SKLU1 was used as the reference since its level was undetected in HBEC3KT. Error bars represent 95% confidence interval (CI).

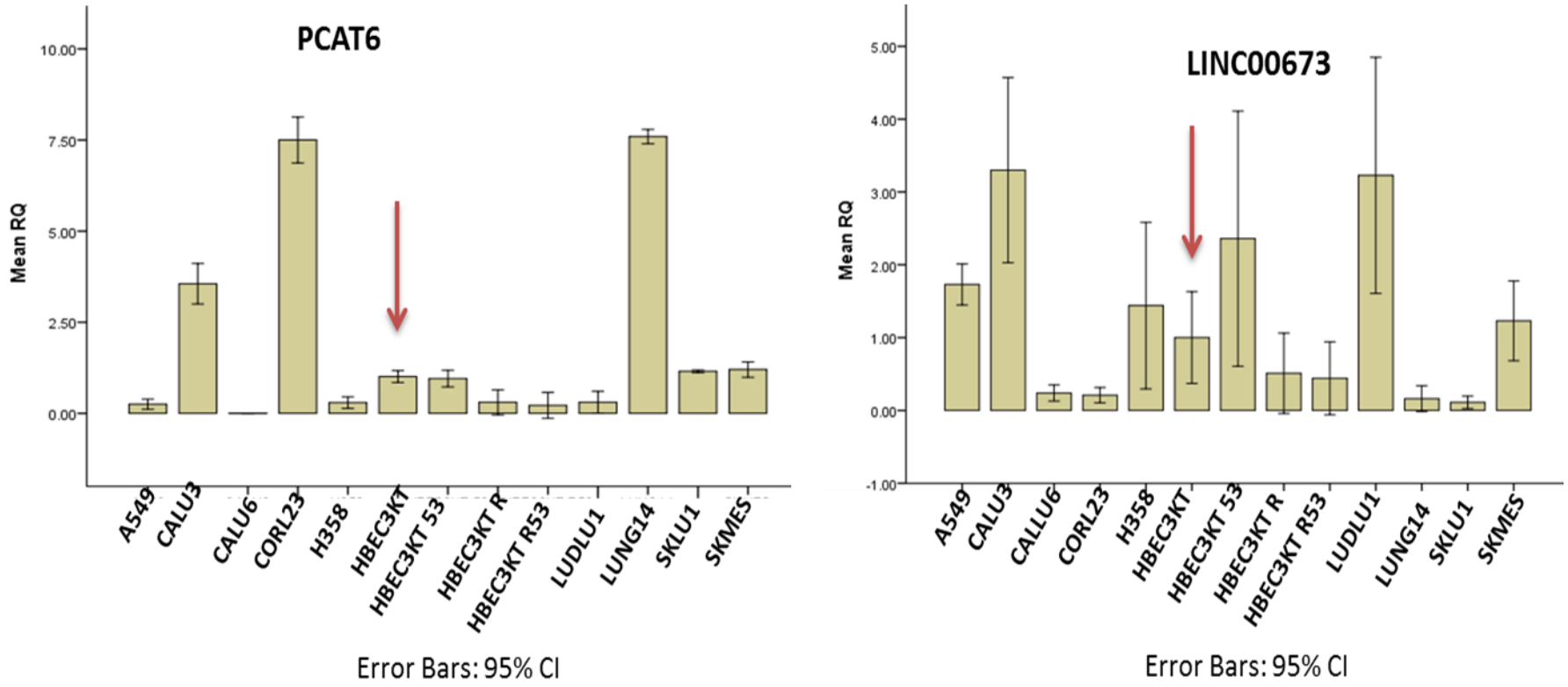


Figure 3.9: Bar chart showing the level of validated lncRNAs, *PCAT6* and *LINC00673* in NSCLC cell lines and HBEC3KT with its isogenic derivatives. HBEC3KT was used as reference cell line (RG=1). Error bars represent 95% CI.

3.7. Discussion

The high mortality rate of NSCLC and the persistently low 5-year survival rate for decades has been the basis for the search for biomarkers for the early detection of the disease and for the discovery of novel therapeutics (Chheang & Brown, 2013; Cancer Research, 2016). The relatively recent discovery of the lncRNAs has added a new level of understanding in disease pathogenesis as well as a new pool of candidate biomarkers.

This study aimed, through a discovery and validation process, to identify lncRNA transcripts that are deregulated in NSCLC and which could serve as potential biomarkers using the microarray technology and validating the data obtained by RT-qPCR. The microarray technology first described about two decades ago (Kononen *et al*., 1998), has been a reliable tool in the study of genetic alterations in cancer and has been instrumental in the identification of potential biomarkers for the early detection of cancer (Giltnane & Rimm, 2004; Holgersson *et al*., 2010).

Following microarray analysis in our study, 50 lncRNA transcripts that are deregulated in NSCLC (20 highly upregulated and 30 highly downregulated) were identified in tumour tissues in relation to their paired normals ($P<0.05$). 13 transcripts across different p-values and fold-changes were selected for validation by RT-qPCR. BLAST searching and optimization of primers and probes enabled the identification of all transcripts. Previously incorrectly annotated lncRNAs (as in the case of *LOC105376287*: originally annotated as *HMGA1P4* in the microarray, *LOC101927229*: Assay 592405 and LINC00673: originally annotated as *LINC00511*) were recovered (Table 3.1).

Several studies have previously demonstrated that lncRNAs are differentially expressed in tumours compared to normals. Xu *et al*, in their study, identified 2420 lncRNAs that were

differentially expressed between lung adenocarcinomas and normal tissue samples, with 1213 of the 2420 upregulated and 1207 lncRNAs downregulated. In another study 47 lncRNAs were shown to be differentially expressed in tumours (14 upregulated and 33 downregulated) from gene expression data of five NSCLC cohorts that were deposited in the Gene Expression Omnibus (GEO) database (Xu *et al*., 2014; Yang *et al*., 2014). A larger number of lncRNAs were differentially expressed in the study carried out by Xu *et al*, while Yang *et al* identified a similar number of lncRNAs that were differentially expressed as shown in our study. The targets identified by Xu *et al* and Yang *et al* differ from those identified in our study, and may be related to the difference in the microarray platforms and probe used in the different studies. Results obtained by Xu *et al*, are based on differential expression between lung adenocarcinoma and normal tissue samples, while our studies involved both lung adenocarcinoma and lung squamous cell carcinoma.

The transcript *FEZF1-AS1* (FEZF1 antisense 1) shown to be upregulated both in the array analysis and qRT PCR, is a non-protein coding gene located antisense to the FEZF1 (FEZ family zinc finger 1) gene on chromosome 7q31.32, with a transcript size of 2653 bases. *FEZF1-AS1* was first identified in studies carried out by Ota *et al*., in 2004, it has been reported to be upregulated in human primary colorectal carcinoma and to facilitate cell proliferation and migration in colorectal carcinoma. It has also been implicated in gastric cancer by its repressive activity on CDKN1A expression, and also identified among the top 50 genes highly expressed in human invasive mucinous adenocarcinoma (IMA) of the lungs (Liu *et al*., 2017). He *et al*., in their study, showed that *FEZF1-AS1* was upregulated in NSCLC tissues, and that downregulation of *FEZF1-AS*1 inhibited cell proliferation and cell invasion (He *et al*., 2017). *FEZF1-AS1* was upregulated in NSCLC tumours and these high levels

correlated with sex in our study. *FEZF1-AS1* was evaluated in a panel of NSCLC cell lines and found to be upregulated in CALU3, CORL23, H358, LUDLU1 and SKLU1 (Figure 3.2A).

Our findings revealed that *LINC01214* (long intergenic non-protein coding RNA 1214) is upregulated in tumours in the array analyses as well as in the validation sets ($p< 0.05$). Its expression was reported to be repressed in studies involving the dual inhibition of TNKS and MEK in the SW480 KRAS-mutant colorectal cancer cell line, however, expression level of *LINC01214* in NSCLC has never been reported. *LINC01214*, an uncharacterized lncRNA is genomically located on chromosome 3q25.1, spanning 150265407 to 150323747 bp. The amplification of the distal portion of this chromosomal region in lung cancer has been shown to be a major indicator of neoplastic transformation, is found in early stages of lung cancer development and maintained throughout the progression of cancer until metastatic stages (Qian and Massion, 2008). Amplification on chromosome 3q25–3q26 associated with a preinvasive lesion that progressed to a subsequent carcinoma has also been described (Foster *et al*., 2005). There may be the possibility of a link between this genetic mechanism and *LINC01214* expression in NSCLC (You & Jones, 2012).

LOC105376287 located on chromosome 9q34.11 showed no significant upregulation in the discovery set, and validation by qRT PCR confirmed this finding. Therefore, *LOC105376287* is not upregulated in NSCLC tumours in comparison to the normal tissues in this study. However, it is elevated in 2 lung cell lines (H358 and LUDLU1).

PCAT6 (prostate cancer associated transcript 6) also known as *KDM5B/KDM5B-AS1* an oncogenic lncRNA located on chromosome 1q32.1, was shown to be induced in both sample sets in this study and in 3 NSCLC cell lines (CALU3, CORL23 and LUNG14). This finding is consistent with studies carried out by Wan *et al*, which showed significant increase in *PCAT6*

expression in lung cancer tissues when compared to normals, and in lung cancer cell lines compared to normal human bronchial epithelial cells. Knockdown of *PCAT6* was associated with inhibition of cell proliferation and metastasis (Wan *et al*., 2016).

LOC101927229 was upregulated in NSCLC tumours compared to adjacent normals in this study. There has been no published information on this lncRNA target. It was shown to be high in LUNG14, moderate in LUDLU1 and was not detectable in CALU3 and SKMES.

LINC00673 was upregulated in this study, and its high level correlates with previous studies. Shi *et al* reported an upregulation of *LINC00673* expression in their study, following bioinformatic analyses and qRT PCR in NSCLC (Shi *et al*., 2016). They further demonstrated that knockdown of *LINC00673* was associated with inhibition of cell proliferation and ability to form colonies. *LINC00673* has also been reported to regulate NSCLC proliferation, migration, invasion and epithelial mesenchymal transition by sponging miR-150-5p (Lu *et al*., 2017). It has been shown to promote metastasis in NSCLC by binding to EZH2 resulting in epigenetic silencing of HOXA5 which is known to regulate gene expression (Ma *et al*., 2017).

NUTM2A-AS1 levels were shown to be elevated in this study. It is located on chromosome 10q23.2 of the cytogenic band, on the opposite strand of the NUTM2A gene. Studies have linked chromosome 10q21 – 10qter to genetic susceptibility to loss of heterozygosity (LOH), and showed a link between LOH and advanced tumour stage in LSSC (Petersen *et al*., 1998). *NUTM2A-AS1* is among a sub-set of lncRNAs that have been reported previously to exhibit significant expression change following genetic perturbation of PNN in the investigation of its role in alternative splicing of lncRNAs (Joo *et al*., 2013; Peixing *et al*., 2017). There have been no reported findings about this identified transcript in NSCLC.

RNF139-AS1 is genomically located on chromosome 8q24.13 of the cytogenic band and its levels were significantly upregulated in the validation set in this study.

LANCL1-AS1 is a lncRNA located on chromosome 2q34. There is no previously published information available about its expression in NSCLC to the best of my knowledge.

FENDRR (FOXF1 Adjacent Non-coding Developmental Regulatory RNA) is located on chromosome 16q24.1 and is believed to act by binding to PRC2 which are important in gene silencing. *FENDRR* has been reported to be important in the proper development of the heart and body of mice, down-regulated in gastric cancer tissues and cell lines, and involved in regulating metastasis by affecting fibronectin1 expression (Grote *et al*., 2013; Xu *et al*., 2014). Navarro *et al*, in their computational study revealed that *FENDRR* regulates FOXF1 expression in lung cancer by interaction with the promoter region of FOXF1 (Navarro *et al*., 2016). However, there has been little or no published information available for its expression in NSCLC. Herrera-Merchan *et al*, have however suggested that both *FENDRR* and FOXF1 are silenced by DNA methylation in p53 mutated human lung cancer cells lines, and that expression of FOXF1 is regulated by *FENDRR* (Herrera-Merchan *et al*., 2016).

LINC00968 (long intergenic non-coding RNA 968) is located on chromosome 8 (8q12.1) and was among the top 10 aberrantly expressed lncRNAs in LUSC according to studies by Wen-Jie *et al*., 2017. It has recently been shown to be significantly increased in LAD and LSCC tissues, as well as in NSCLC cell lines in comparison to controls (Wang *et al*., 2017). They further indicated that inhibition of *LINC00968* brought about repression of NSCLC growth, migration and invasion. However, *LINC00968* expression was repressed in NSCLC tissues in comparison to adjacent normal tissues, as well as in NSCLC cell lines in this investigation, and this finding correlates with studies carried out by Li *et al*, they observed *LINC00968* to be significantly down regulated in LADC (Li *et al*., 2017c).

SVIL-AS1 (SVIL antisense RNA) is genomically located on chromosome 10p11.23. Findings from this study reveal that expression of this transcript is repressed in NSCLC tissues compared to normals, and no previous report of its expression and function has been published to date.

PCAT19 (Prostate Cancer Associated Transcript 19) is located on chromosome 19 (19q13.8). The analysis carried out by Li et al. showed down-regulation of *PCAT19* in LADC (Li *et al*., 2017d).

In conclusion, the findings of this study showed that 13 lncRNAs deregulated in NSCLC tissues were identified through the microarray analysis and predicted to be either induced or repressed (as seen in Table 3.1). qRT PCR validation analysis confirmed the microarray results (Figures 3.1 A-G and Figure 3.2 A-E), apart from *LOC105376287* expression that was not upregulated in the validation sample sets. The expression change observed in tumours and normals was not significantly associated with the clinicopathological parameters.

Comparative analysis of lncRNA expression with patient clinicopathological characteristics produced varying results. No significant differences were observed in the expression levels of all validated lncRNAs with age and histology. *FEZF1-AS1* was higher in males compared to females (Mann Whitney's Test, p=0.02), in contrast, high level of *FEZF1-AS1* was not significantly correlated with gender in studies carried out by Jin *et al*. They however, reported an association between high expression level and tumour size, TNM stage and lymph node metastasis. Chen *et al*, also observed that patients with high expression of *FEZF1-AS1* in colorectal cancer had advanced T-stage, lymph node, or distant metastasis (Chen *et al*., 2016).

4. EVALUATION OF EPIGENETIC INFLUENCE ON lncRNA DEREGULATION

4.1. Introduction

Epigenetic regulation is a key mechanism that is important in controlling developmental processes and disease development and progression (Liloglou *et al*., 2014). Abnormal gene promoter DNA methylation has been implicated in tumorigenesis and has become a target for developing strategies for early detection of cancer (Belinsky, 2004; Esteller, 2005 & Wu *et al*., 2016). In this study, we aimed to investigate the epigenetic causes of deregulation of the identified genes in NSCLC tumours and also to evaluate the influence of epigenetic modulators in some NSCLC cell lines.

4.2. Promoter Methylation Levels of Selected lncRNAs in Tumour and Normal Tissues

Pyrosequencing mediated DNA methylation analysis was undertaken for the promoters of those lncRNA target genes identified in this study that have a CpG island[5] (*FEZF1-AS1, PCAT6, NUTM2A-AS1, LINC00673, RNF139-AS1, FENDRR* and *SVIL-AS1*). This was analysed by testing 1000 bp upstream exon 1 (including exon 1) in the online tool EBI (www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/), an example of which is shown in Figure 4.1.

[5] A 200bp region of DNA having a GC content above 50% and an observed CpG versus expected CpG ratio greater than or equal to 0.6.

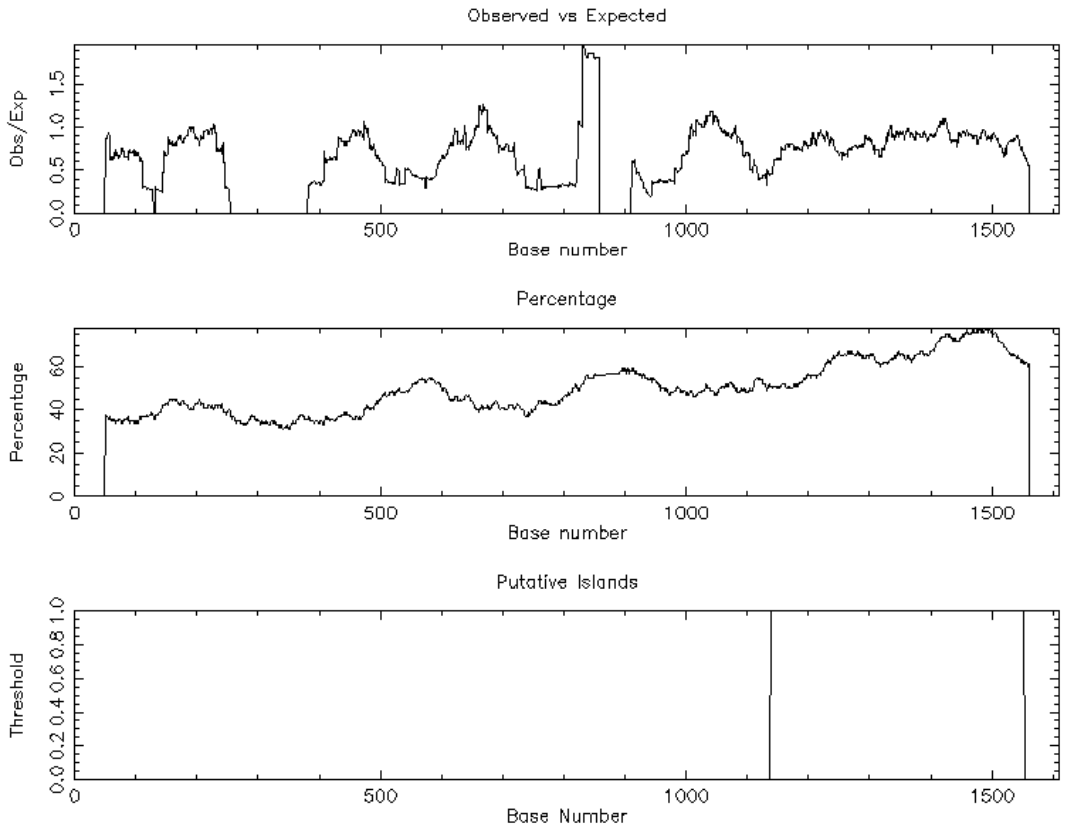


Figure 4.1: An example of a Cpgplot showing the location of a CpG island identified in the FEZF1-AS1 promoter (the density of the CpG Island; the percentage of the C + G nucleotide content and the core promoter, based on standard conditions (Observed/Expected ratio> 0.60, percent C + G> 50, Length > 200).

DNA methylation analysis carried out on 129 NSCLC tumours and 36 normal tissues for which DNA was available. The threshold for significant methylation levels was set to 5%, as below this is considered technical noise (Shaw *et al*., 2006). The clinicopathological characteristics of these set of samples are given in Table 4.1. The mean age of the patients in this study was 65.57 and the male to female ratio was 88:41 respectively. From a histological perspective, patient samples comprised of 76 (58.9%) LSCCs and 52 (40.3%) LADCs. Most of the tumours were stage II disease [(107 (82.95%)] and 62% of patients had moderately differentiated carcinomas.

Table 4.1: Clinicopathological data of patients used for methylation analysis of identified lncRNAs

	Methylation Set
N	129
Age	
Mean (s.d.)	65.57 (9.14)
Gender	
Male: Female	88: 41
Histology	
Adenocarcinoma	52
Squamous	76
Other	1
Tumour Stage	
T1	10
T2	107
T3	10
T4	1
Missing	1
Nodal Stage	
N0	70
N1	41
N2	16
Missing	2
Differentiation	
Well	32
Well - Moderate	4
Moderate	80
Moderate - Poor	2
Poor	10
Missing	1

Pyrosequencing analysis demonstrated that two out of the 7 genes (*FEZF1-AS1* and *FENDRR*) were hypermethylated (Figures 4.2A and 4.2F). *NUTM2A-AS1* and *LINC00673* demonstrated a consistent and thus significant (regarding the derived p value) hypomethylation in the tumour tissue (as shown in Figures 4.2C and 4.2D). Despite the statistical significance of the changes observed, the biological significance is questionable, considering the methylation

levels (which are < 5% and therefore within the technical noise) (Shaw *et al*., 2006). *PCAT6*, *RNF139-AS1* and *SVIL-AS1* demonstrated no significant difference in DNA methylation between the normal or tumour tissues.

The *FEZF1-AS1* promoter was shown to be hypermethylated in tumour samples relative to the normal tissues (Mann Whitney test, p= 0.048). A representative pyrogram of the *FEZF1-AS1* promoter (Figure 4.3) shows the methylation levels of the individual CpG islands in tumour samples compared to normal lung tissues. Lower methylation levels were observed for *NUTM2A-AS1* and *LINC00673* in tumours compared to the adjacent normal tissues (Mann Whitney test, $p=8.5 \times 10^{-6}$ and 1.37×10^{-6} respectively). Representative Pyrograms showing methylation levels of *FEZF1-AS1, FENDRR, NUTM2A-AS1* and *LINC00673* are shown in Figures 4.3 to 4.6.

Statistical analysis revealed no association between the DNA methylation deregulation shown in *FEZF1-AS1, FENDRR, NUTM2A-AS1* and *LINC00673* and the clinicopathological parameters (age, gender, histological sub-type, tumour/nodal staging and differentiation) in the examined sample set, indicating this deregulation is universal across histologies and stages of NSCLC.

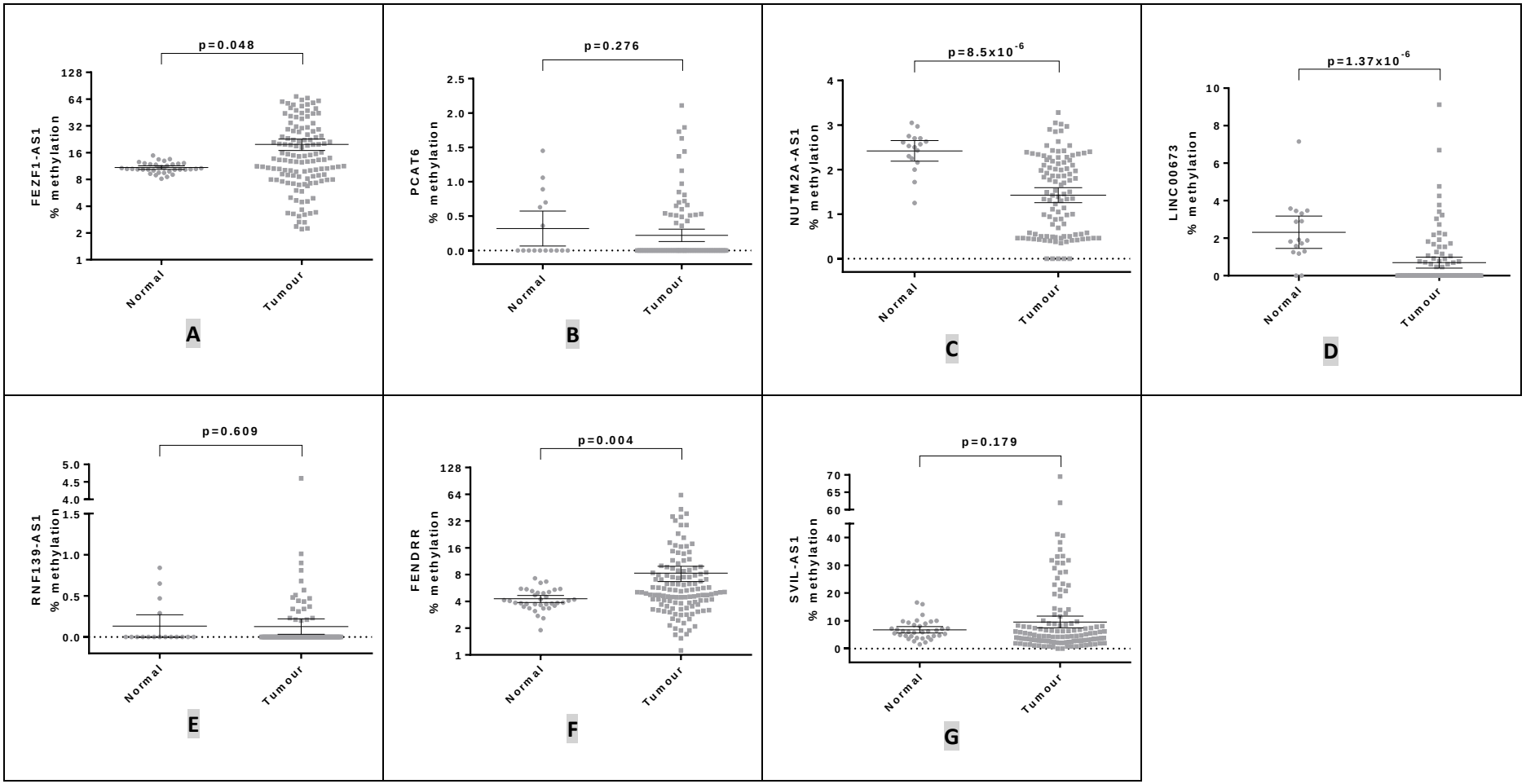


Figure 4.2: Scatter plots showing DNA methylation of 8 lncRNA promoters in NSCLC tumours and paired normal tissues. Mann-Whitney Test, $p < 0.05$ was considered significant.

FEZF1-AS1

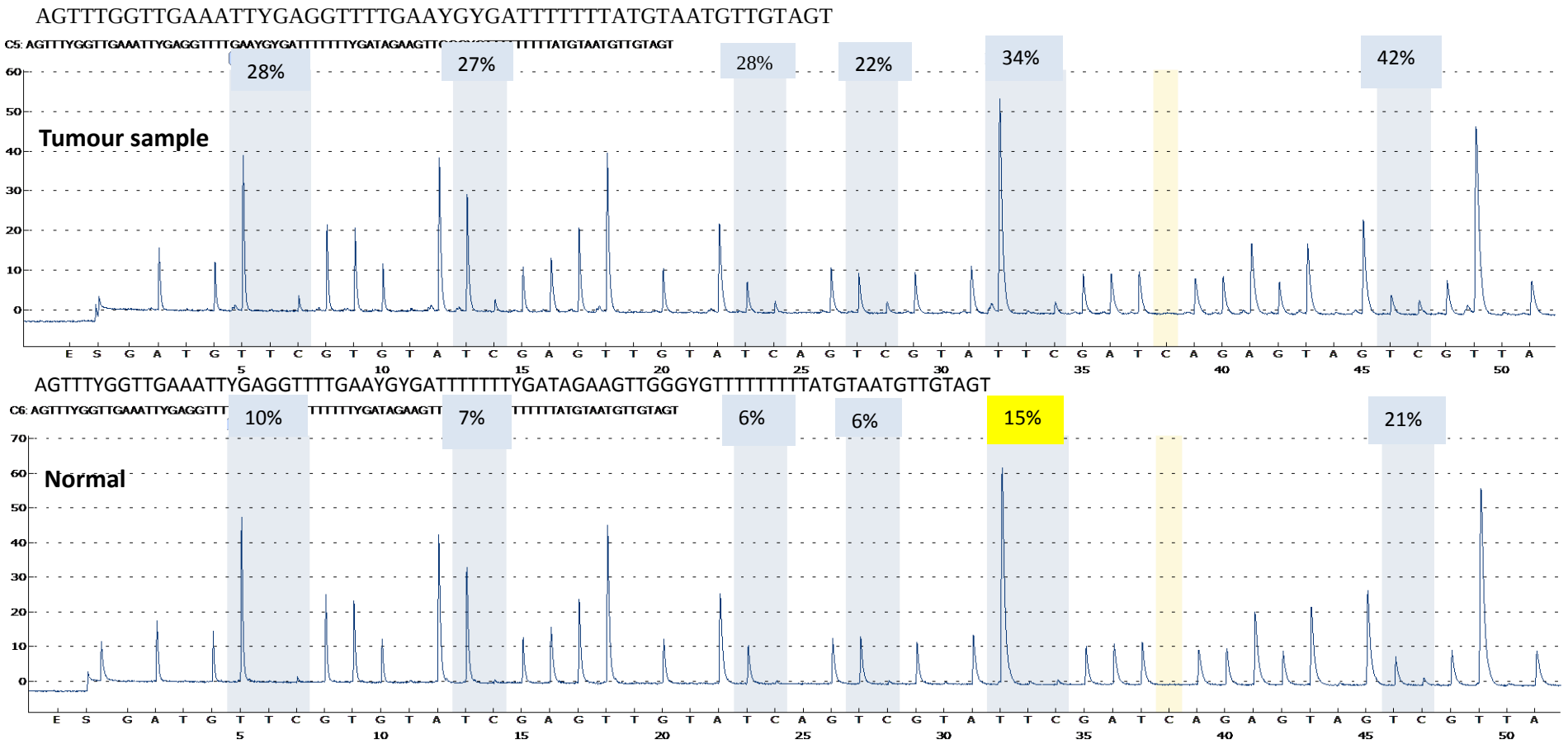


Figure 4.3: Pyrogram showing the methylation of 6 positions in tumour and adjacent normal lung tissues of the FEZF1-AS1 promoter. Sequence analysed is shown at the top left of both tumour and normal samples. X-axis demonstrates the dispensation order. Examined CpGs are indicated by blue vertical zones while the bisulphite conversion control is denoted by the yellow vertical zone. The methylation level of individual CpGs are shown at the top of each zone. The peak with the yellow bar is an indication of a probable false positive.

FENDRR

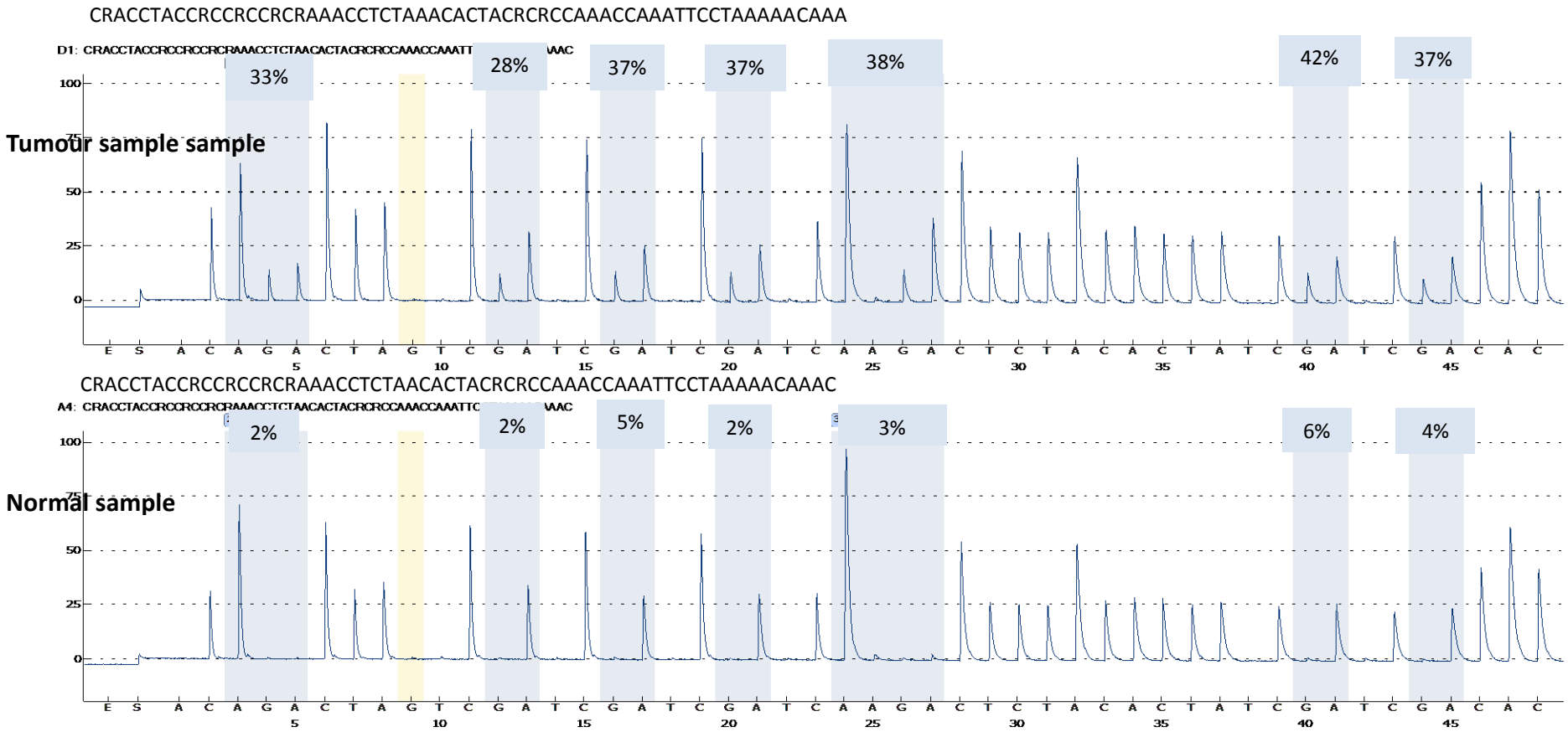


Figure 4.4: Hypermethylation in the promoter region of FENDRR promoter can be observed for all the 7 CpG islands in tumour samples compared to the normal samples. The sequence analysed is shown at the top left of both tumour and normal samples. X-axis demonstrates the dispensation order. Examined CpGs are indicated by blue vertical zones while the bisulphite conversion control is denoted by the yellow vertical zone, while the methylation level of individual CpGs are shown at the top of each zone

NUTM2A-AS1

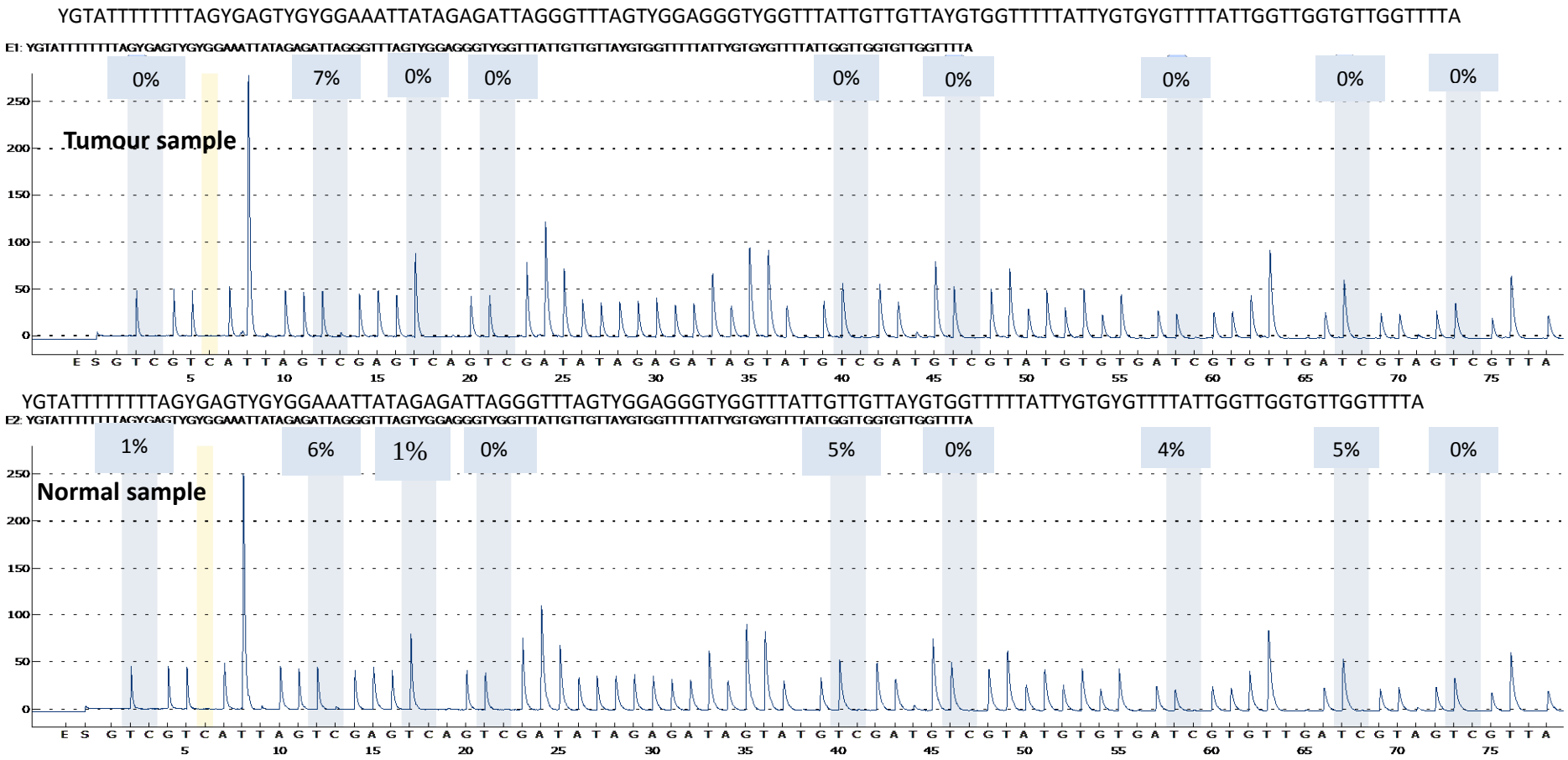


Figure 4.5: Pyrogram showing the methylation levels at the promoter of *NUTM2A-AS1* gene in tumour and normal lung tissues. All 9 CpG islands were hypomethylated in tumour and normal samples. The yellow bar is the internal quality control. CpG islands highlighted in red are indicative of error data picked up by pyrosequencing.

LINC00673

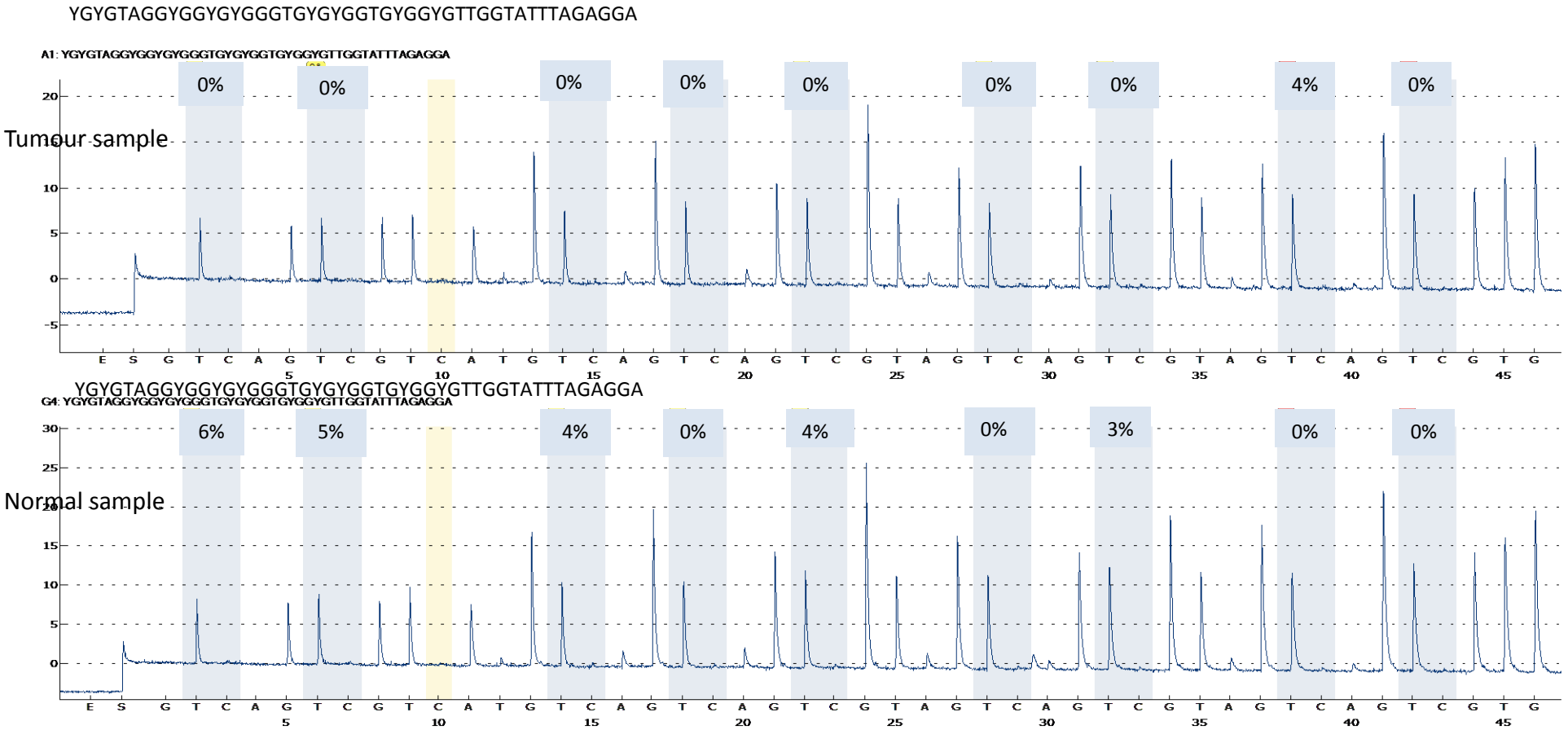


Figure 4.6: Pyrogram showing the methylation levels at the promoter of LINC00673 gene in tumour and normal lung tissues. CpG islands highlighted in blue are shown to be hypomethylated in both tumour and normal samples.

4.3. Influence of Epigenetic Modifiers on NSCLC Cell Lines

In order to gain additional evidence on the potential epigenetic control of the identified lncRNAs, we studied the impact of two epigenetic modifiers, namely 5-aza-deoxycytidine (DAC) and valproic acid (VPA), on their levels in 4 NSCLC cell lines (A549, CALU6, H358 and SKLU1). The efficiency of DAC on cell lines was assessed by measuring global DNA methylation using the LINE1 repetitive element sequence (Daskalos, 2009). Line1 being a gold standard for the measurement of global methylation is used in this study to show that methylation differences in lncRNA promoters were not primarily following global demethylation trends, therefore there might be certain selective advantages for the tumour biology.

From our findings, we observed that the use of LINE1 in the cell lines showed reduced levels of global methylation of LINE1 in SKLU1 cell line after treatment with DAC compared to the control. SKLU1 untreated cells had a methylation index of 62%, while DAC treated cells had a reduced methylation index of 41%. Methylation index of LINE1 in the untreated H358 cell line was only 51% (which did not signify heavy methylation), however, there was a marked reduction in the DAC treated cell line (37%). No significant difference was observed between control and DAC treated A549 and CALU6 cell lines. There was no change observed in all the cell lines treated with VPA, but it is interesting to observe a potentiating effect in VPA/DAC treated H358 and SKLU1 cell lines. A summary of the results obtained for all the cell lines and a representative pyrogram for SKLU1 are shown in Table 4.2 and Figure 4.7 respectively.

Table 4.2: Methylation status of LINE1 in cell lines before and after treatment with DAC and VPA. Results are presented as means of 3 technical replicates.

Cell Line	Methylation Status (%)			
	Control	DAC	VPA	VPA/DAC
A549	54	49	53	51
CALU6	44	41	44	39
H358	51	37	51	38
SKLU1	62	41	63	42

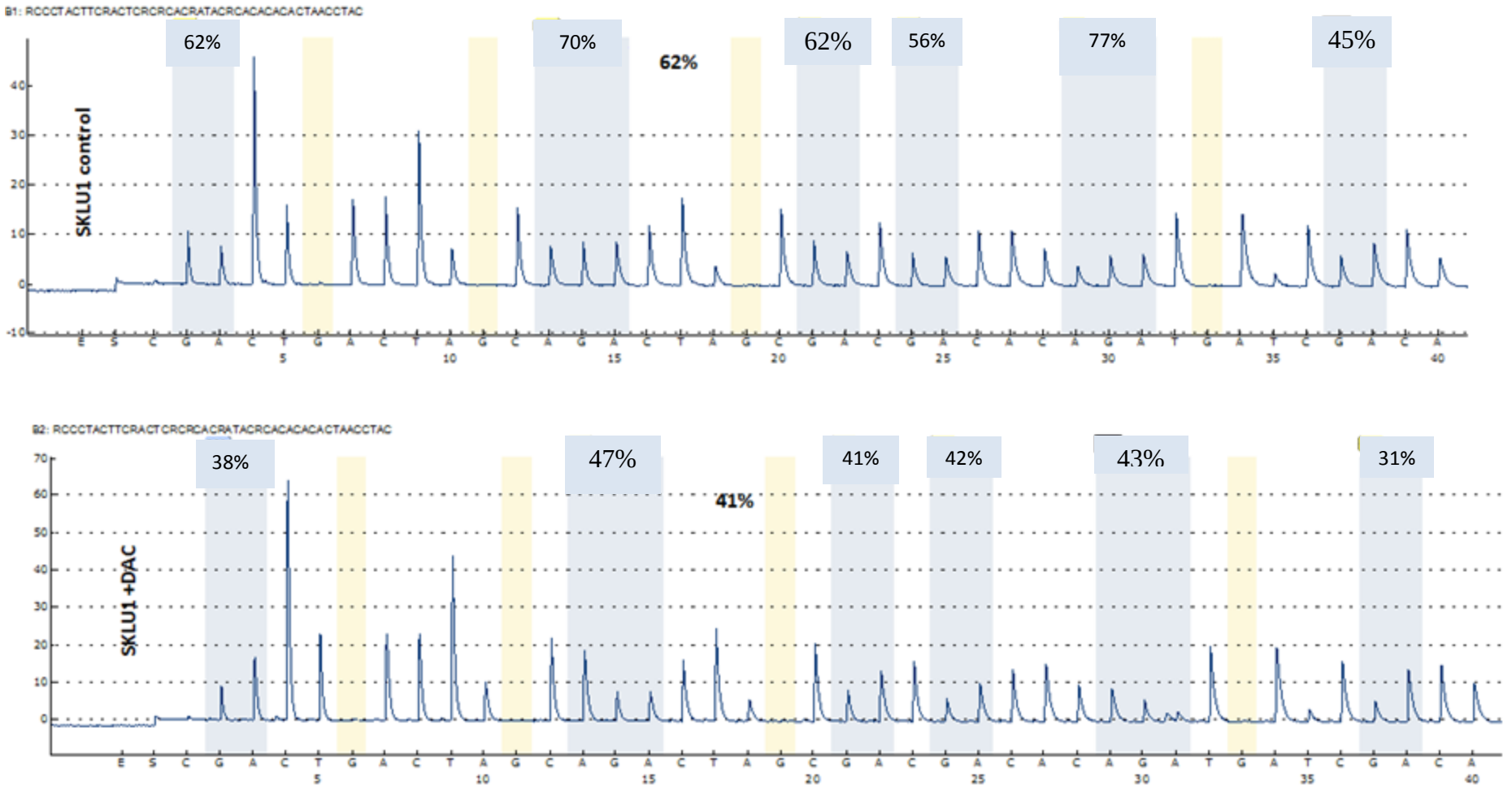


Figure 4.7: Representative pyrogram showing the methylation status of LINE1 in the untreated SKLU1 and the SKLU1 cell line exposed to DAC. Hypomethylation of LINE1 can be observed following treatment with DAC. The sequence to be analysed is shown at the top left of both pyrograms, and percent methylation of individual CpG sites indicated at the top of the grey-shaded positions.

Levels of the individual lncRNAs with or without the epigenetic modifiers are shown in Figure 4.8. *FEZF1-AS1* level analysis showed varying levels in the different cell lines and the results were compared with controls. A549 and CALU6 did not express *FEZF1-AS1* and treatment with DAC had no effect on its expression in both cell lines; however, its expression was found to be altered by VPA and VPA/DAC in A549, and by VPA in CALU6. A similar expression pattern was observed for H358 and SKLU1, expression of *FEZF1-AS1* was reduced following DAC treatment in H358 and SKLU1. A synergistic effect was observed for combination treatment with VPA/DAC in H358 significantly (Figure 4.8). *LINC01214* gene expression was increased after treatment with DAC in A549, CALU6 and SKLU1. In contrast, expression of *LINC01214* in H358 and SKLU1 was significantly increased with VPA/DAC treatment (Figure 4.8). There was a trend of reduced expression of *LOC105376287* in DAC treated A549, H358 and SKLU1 (Figure 4.9) and all the VPA treated A549, CALU6, H358 and SKLU1. Expression of *PCAT6* was increased in VPA treated A549, H358 and SKLU1, and a synergistic effect could be observed in VPA/DAC treated A549 and SKLU1. However, its expression was reduced in DAC treated A549, H358 and SKLU1 (Figure 4.9).

DAC treated CALU6 and H358 showed increased expression of *LOC101927229*, while a lower expression was observed in SKLU1 treated with DAC, VPA and VPA/DAC (Figure 4.10). *LINC0073* expression was reduced in DAC treated A549, H358 and SKLU1, as well as in VPA treated A549, CALU6 and H358 (Figure 4.10). There was no significant difference in expression of *NUTM2A-AS1* expression in all the treated cell lines compared to the controls (Figure 4.11). Expression of *RNF139-AS1* was reduced in A549 exposed to DAC and VPA; however, it was increased in H358 exposed to DAC (Figure 4.11).

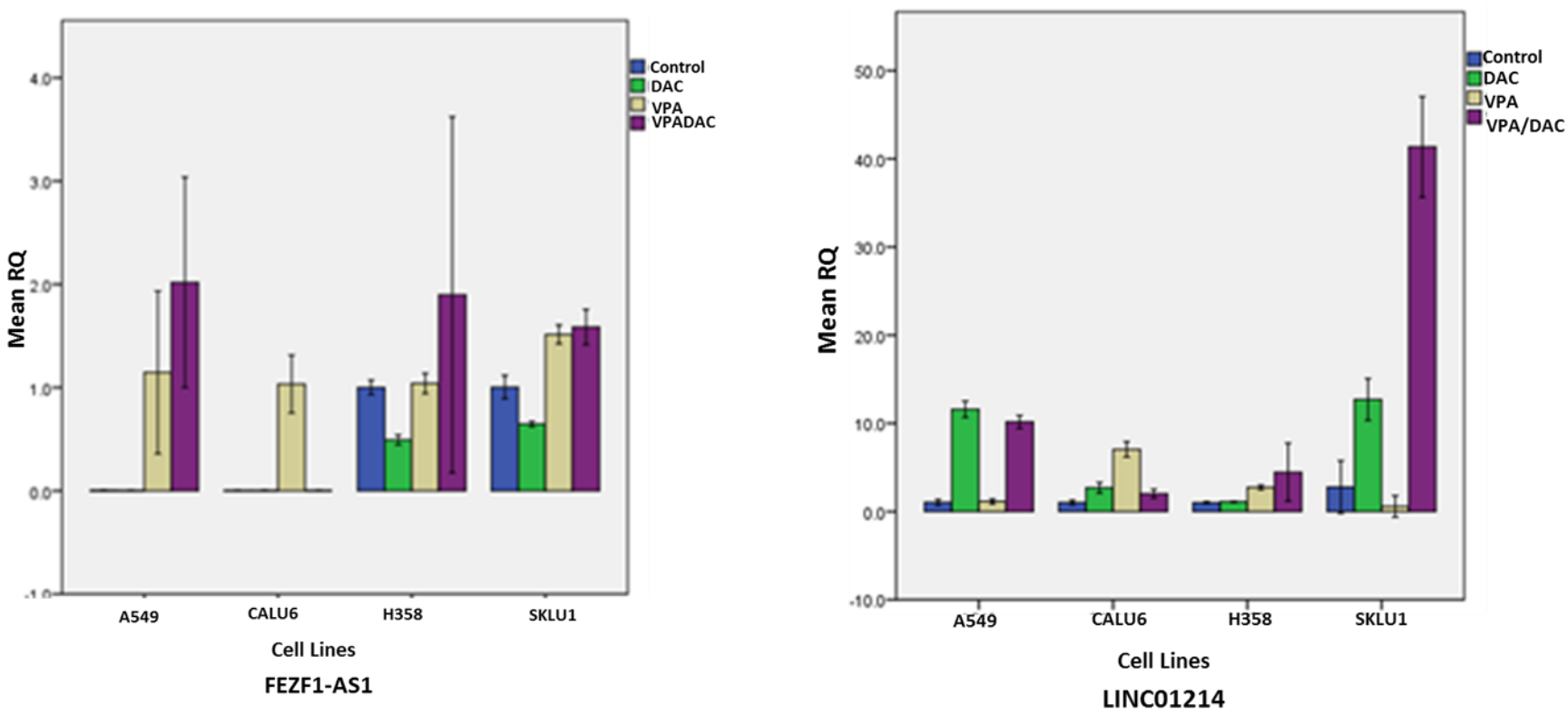


Figure 4.8: Bar chart representation showing *FEZF1-AS*1and *LINC01214* levels in A549, CALU6, H358 and SKLU1 NSCLC cell lines following treatment with of DAC and VPA. DAC treatment is shown to increase LINC01214 levels in A549, CALU6 and SKLU1 consistently. Error bars are set at 95% confidence interval.

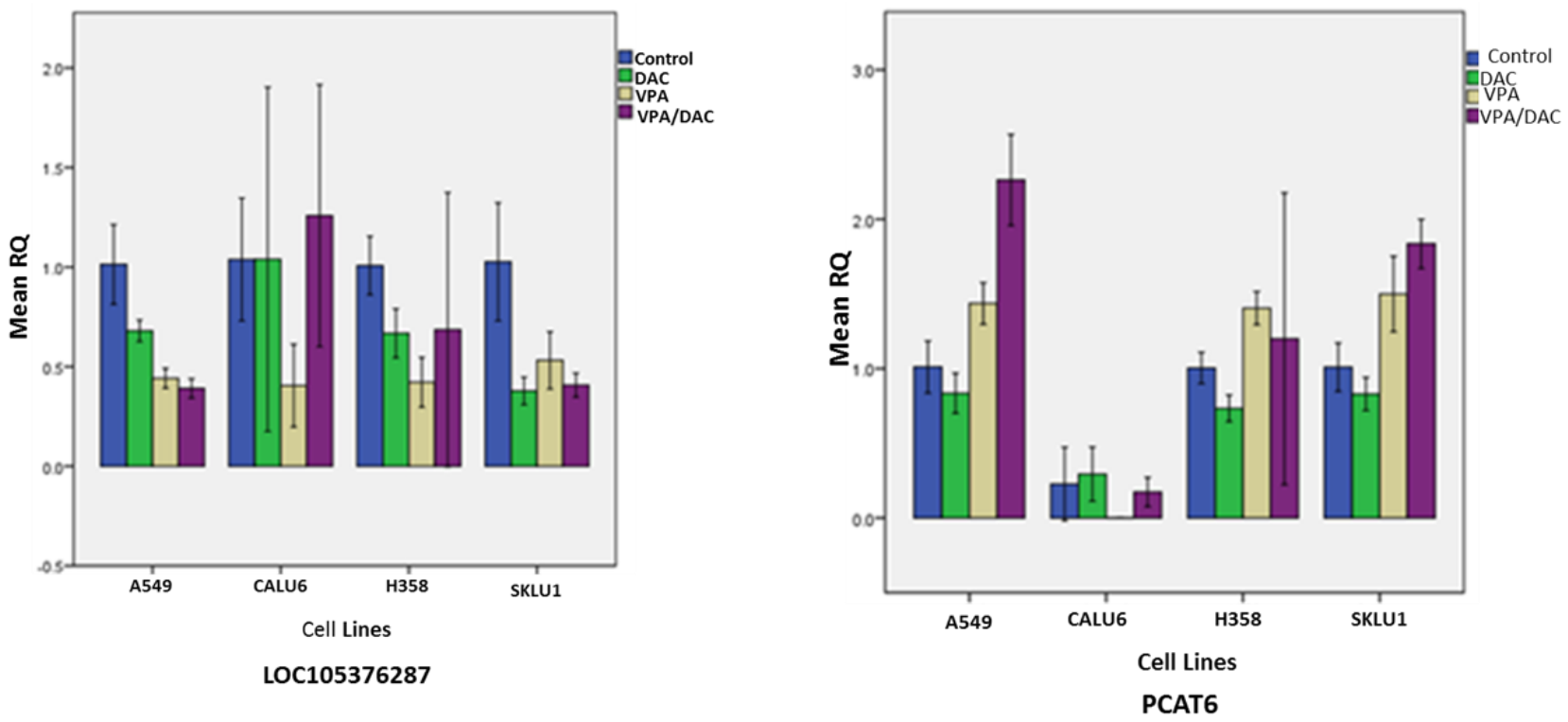


Figure 4.9: Bar chart representation showing *LOC105376287* and *PCAT6* levels following treatment with DAC, VPA and a combination of both. Treatment with DAC reduced *LOC105376287* and *PCAT6* levels in A549, H358 and SKLU1.

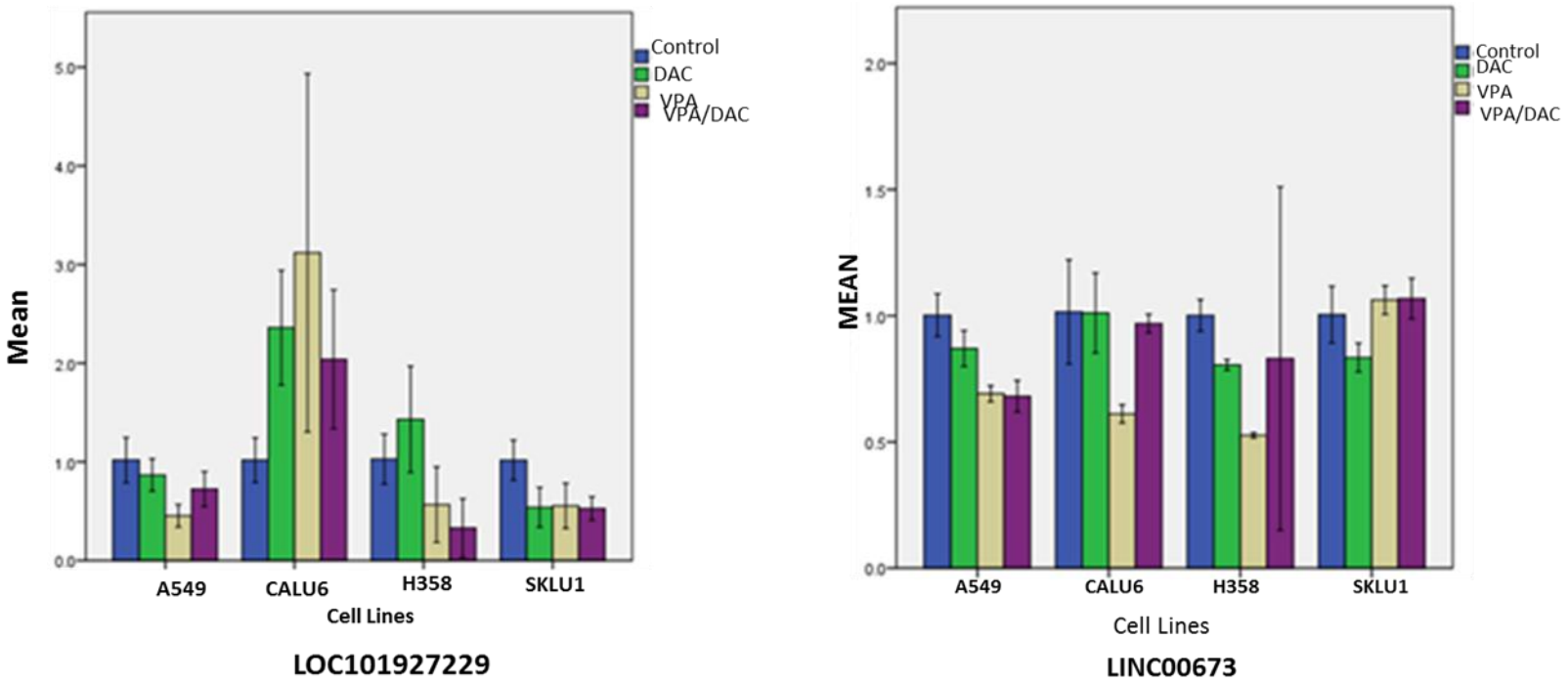


Figure 4.10: Bar chart showing effect of DAC and VPA on *LOC101927229* and *LINC00673* in the A549, CALU6, H358 and SKLU1 cell lines. DAC treated cells showed a reduction in *LINC00673* levels A549, H358 and SKLU1.Error bars are set at 95% CI.

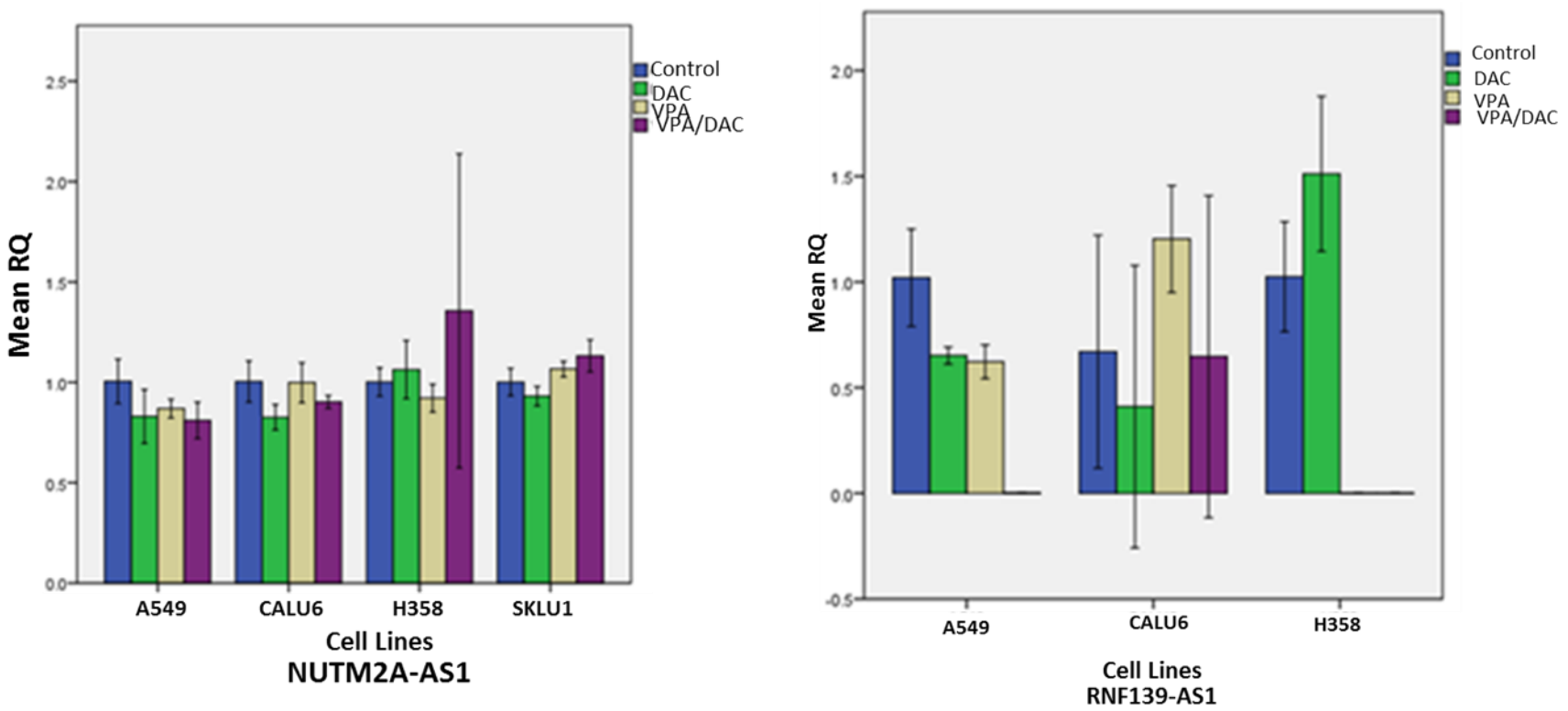


Figure 4.11: Bar chart representation of *NUTM2A-AS1* and *RNF139-AS1* levels in some cell lines after exposure to DAC and VPA.In the presence of DAC *NUTM2A-AS1* levelsreduced in A549, CALU6 and SKLU1.

4.4. Discussion

There are different mechanisms by which lncRNAs appear to regulate gene expression, however, epigenetic mechanisms are known to play a crucial role in regulating gene expression and are involved in the development of cancer. Several studies have shown that about 20-30% of lncRNAs have been able to physically interact with specific epigenetic enzymes, driving them to specific genomic loci (Karapetyan *et al*., 2013). This interaction with chromatin-modifying proteins, like PRC2 and CoREST, results in modulation of the chromatin states thereby controlling normal biological processes (Cheetham *et al*., 2013), while interaction with promoters and transcription factors control transcription of genes in *cis* or *trans* resulting in gene silencing etc.(Qureshi *et al*., 2010). Epigenetic mechanisms (like DNA methylation, histone modification and ncRNAs) control the access of transcriptional machinery to their target genes thereby modulating transition from the condensed heterochromatin to relaxed euchromatin and *vice versa*. These are synergistically operated to regulate the chromatin structure and to achieve the required degree of gene expression needed for normal biological processes (Seidel *et al*., 2012; Liloglou *et al*., 2014). The methylation of CpG islands in the promoter region of genes is known to result in gene silencing and there is an established link between aberrant global DNA methylation and genomic instability (Jones and Takai, 2001).

lncRNAs deregulation a probable cause for the onset and development of epigenetic diseases, including cancer (Zhao *et al*., 2016). Feng et al, in their study linked lncRNA deregulation and DNA methylation in NSCLC (Feng *et al*., 2016). In this study, methylation changes were shown in the promoters of *FEZF1-AS1, FENDRR, NUTM2A-AS1* and *LINC00673*.

From our findings, we observed that aberrant DNA methylation occurs in NSCLC tumour tissues when compared to their adjacent paired normals. We identified two hypermethylated and two hypomethylated lncRNA genes, which, to our knowledge have no published report of their expression or methylation status. *FEZF1-AS1* expression was induced in NSCLC tumours and was found to be hypermethylated in this study. Although promoter hypermethylation is associated with gene silencing (repression), *FEZF1-AS1* was shown to be overexpressed in NSCLC tumour tissues. Global DNA hypomethylation (mainly associated with repeated DNA) and hypermethylation in non-repeat DNA stretches of gene promoter sequences have been reported to contribute in cancer development and it has been revealed that DNA hypomethylation at the promoter regions of oncogenes can activate their expression, while DNA hypermethylation at the promoters of tumour suppressor genes can lead to their silencing (Ehrlich, 2002; Ehrlich, 2009). Although promoter hypermethylation is associated with gene silencing, *FEZF1-AS1* is hypermethylated but also upregulated in NSCLC tumours compared to adjacent normal tissues. It has been reported to act as an oncogene in gastric cancer and as an anti-oncogene in pancreatic cancer (Liu *et al*., 2017). The methylation status of *FEZF1-AS1* promoter in gastric and pancreatic tumours is unknown. *FEZF1-AS1* may be acting as an oncogene in NSCLC. *FENDRR* gene expression was shown to be repressed in tumours compared to the normal lung tissues and was also significantly hypermethylated in lung tumour tissues (Figures 4.6 and 4.9), this finding agrees with the gene silencing function of promoter methylation. Several lncRNAs that are down regulated in cancer are shown to act as tumour suppressor genes (*MEG3 , GAS6-AS1 , BANCR* have all been reported to be significantly down-regulated in NSCLC tissues and other malignancies) (Lu *et al*., 2013; Han *et al*., 2013; Sun *et al*., 2014). *NUTM2A-AS1* and *LINC00673* are two lncRNAs overexpressed in NSCLC but also

hypomethylated at their promoters. Udomsinprasert et al have reported an association between promoter hypomethylation and overexpression (Udomsinprasert *et al*., 2017). DNA hypomethylation has also been linked to overexpression of *PLS3*, GATA6, and *TWIST1* in Sezary Syndrome (Wong *et al*., 2015). For over 30 years, aberrant promoter hypermethylation that results in silencing of tumour suppressor genes in cancer has been the focus of several studies and have led to the development of DNA hypomethylating therapeutic agents to reverse DNA hypermethylation. However, the emergence of new high-resolution methylation profiling techniques, like whole genome bisulfite sequencing (WGBS), have revealed that the methylome of cancer cells frequently contain promoters that are hypomethylated in comparison to their paired normal cells. Such aberrant hypomethylation has been shown to be frequently accompanied by increased gene expression at differentially methylated loci (Upchurch et al. 2016). The hypomethylation of the promoter regions of *NUTM2A-AS1* and *LINC00673* may be responsible for increased expression of this lncRNAs and they may be acting as oncogenes.

DNA methylation is an early event in tumourigenesis, and the profiling of global methylation can lead to identification of disease-state biomarkers and early diagnosis. Several studies have used pyrosequencing of LINE1 assay as a standard method for studying global DNA methylation (Yang *et al*., 2004; Lisanti *et al*., 2013; Delaney *et al*., 2015). LINE1 (long interspersed nuclear elements 1) are repetitive elements that are abundant in the human genome (making up approximately 17% of the genome i.e. about 500,000 copies per genome). They are relatively inactive, majorly due to 5' truncations (Lander *et al*., 2001, Kazazian, 2004), however, the prevalence of CpG islands and the associated aberrant DNA methylation (hypomethylation and hypermethylation) in these regions of the genome is

reported to cause the retrotransposition of these elements resulting in diseases like cancer etc. (Portela and Esteller, 2010). Loss of methylation at repeat elements has also been reported to be associated with increased chromosomal instability, aggressive development of colorectal cancer and NSCLC (Igarashi *et al*., 2010; Saito *et al*., 2010).).

5. FUNCTIONAL *IN VITRO* ANALYSIS OF LINC00968

5.1. Introduction

lncRNAs have diverse functions, however, to date we only understand the function of a small proportion of the known lncRNAs. While the observation of lncRNA deregulation in lung cancer (which was the starting point of this study) is very important, it is imperative to understand the functional relevance of our identified lncRNAs in NSCLC and explore their functional contribution to cell phenotype and/or drug resistance if any.

We selected *LINC00968* one of the top repressed lncRNAs in our study (Figure 3.2C), for further analysis. In order to examine the biological importance of *LINC00968* in NSCLC, we overexpressed the gene in A549 cell line and assessed the phenotypic changes including potential modulation of resistance to commonly used chemotherapeutic drugs in NSCLC.

5.2. *LINC00968* Selection and Cloning

LINC00968 was chosen for further studies because of its evident low levels as observed in our microarray study, the fact that its expression was completely diminished in most of the tested lung cancer cell lines (See Appendix V) and the extremely limited knowledge available on its function in the literature [Li *et al*, in their study also showed that *LINC00968* was differentially expressed in tumours compared to normals (Li *et al*., 2017c)].

Because of the apparent absence of expression observed in tissues and cell lines, we hypothesized that *LINC00968* could be functioning as a tumour suppressor in NSCLC; therefore, the restoration of expression could potentially reverse malignant phenotype.

To achieve the above objective, the pTarget_968 plasmid was generated to induce *LINC00968* expression as described in Section 2.2.8.3. Following which, the A549 lung

adenocarcinoma cell line was transfected with the plasmid to generate stable clones expressing the gene.

Transcript levels of LINC00968 in A549 parental cell and the clones was assessed by qRT PCR. *LINC00968* levels was not detected in A549 parental cell line, Clones 3, 6, 7, 8, 10 and 11. However, variable levels of the gene were detected in Clones 1, 2, 4, 9 and 12, and Clone 2 was chosen as a reference (set to 100%) because it demonstrated the highest level of *LINC00968* (Figure 5.1). To assess its level for comparative purposes, all other clones were measured in reference to Clone 2 shown in Table 5.1.

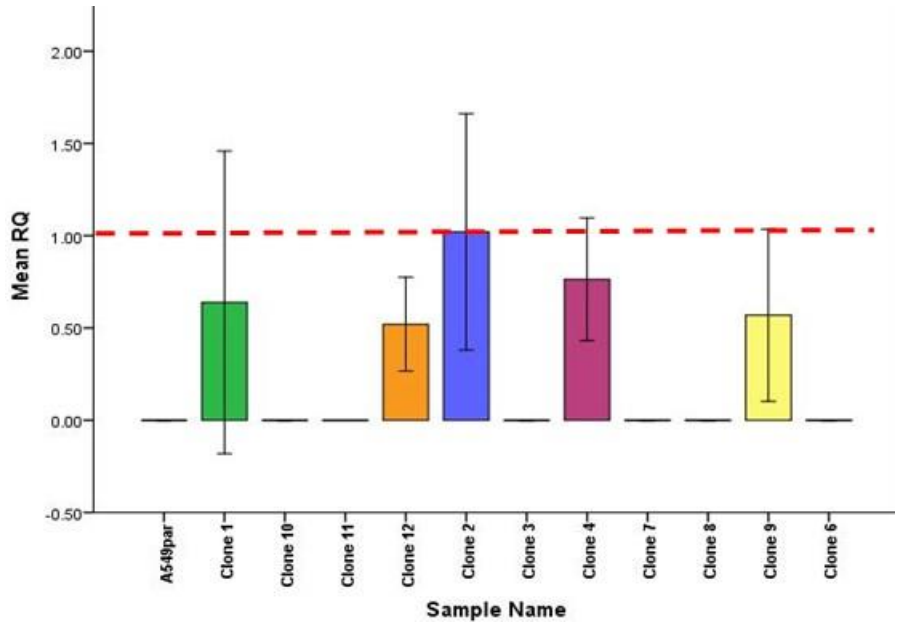


Figure 5.1: Bar chart showing *LINC00968* levels in A549 parental cell line and the different clones following transfection with pTarget_968. The order follows the pattern of the qRT PCR analysis. Error bars are representative of 95% confidence interval.

Table 5.1: Comparative levels of *LINC00968* in the A549 parental cell line and the Clones as a percentage in reference to Clone 2.

A549_LINC00968 Clone	LINC00968 Comparative Expression Level	LINC00968 Expression
A549Parental	0	0
Clone 1	59	Moderate
Clone 2	100	Maximum
Clone 3	0	
Clone 4	75	High
Clone 5	0	
Clone 6	0	
Clone 7	0	0
Clone 8	0	
Clone 9	53	Moderate
Clone 10	0	
Clone 11	0	
Clone 12	33	Low

5 of the clones (2 representing maximum transcript levels, 4 as high levels, 9 as moderate levels, 12 as low levels and 7 acting as vector only) were selected for the evaluation of functional relevance of the *LINC00968* lncRNA in the different assays and drug treatments.

5.3. MTT Assay

The ability to proliferate is one of the hallmarks of cancer cells, therefore, cell proliferation rate was assessed using the MTT assay as described in Section 2.2.9.1. Although this assay measures cell metabolic activity, it is commonly used in assessing proliferation rates. Interestingly, the greater repression in proliferation was not demonstrated in the clones expressing high levels of *LINC00968* but in the clone with lower levels (Figure 5.2), while further increased levels demonstrated a tendency toward the proliferation rates of parental or vector only cell. This is interesting but also difficult to interpret at this point, as the physiological levels of *LINC00968* are not well understood and comparison to the normal

human lung tissue is not technically valid, given the fact that the latter is a mix of different cell types.

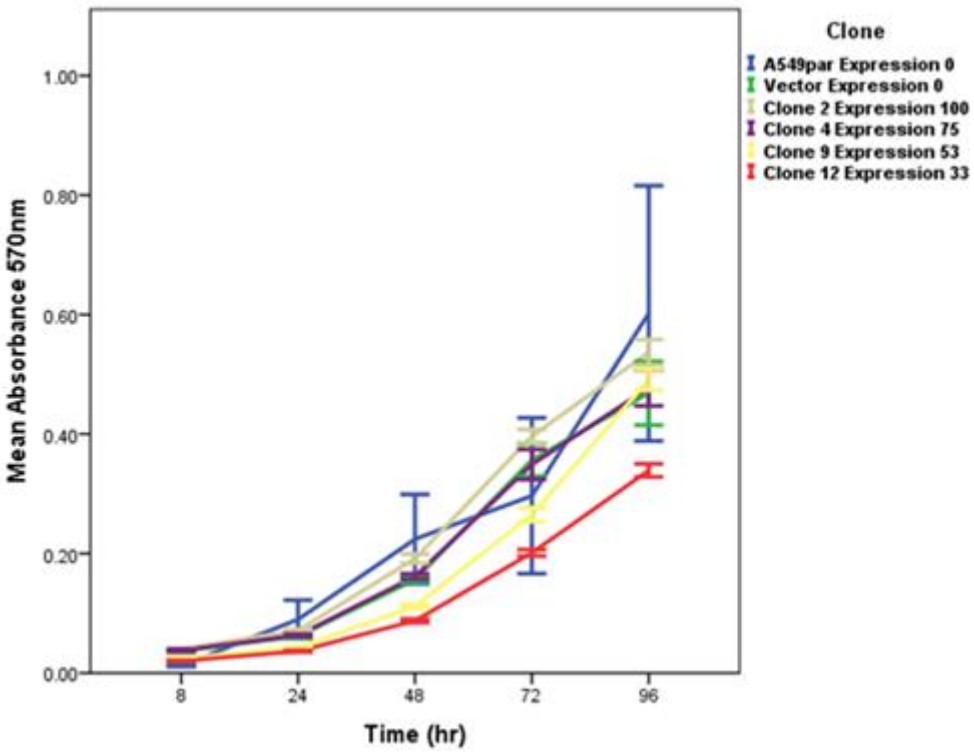


Figure 5.2: Line graph showing the viability of the parental A549 cell line and the respective selected Clones. Higher proliferation rate was observed for the A549 parental cell line in comparison to the Clones. Error bars are representative of 95% confidence interval.

5.4. Anchorage Independence Assay

The ability of transformed cells to grow independently of a solid surface is another hallmark of tumourigenesis, therefore in this study we explored the anchorage independent activity of the parental A549 and the selected A549_linc00968 clones as described in Section 2.2.9.2. We observed that the presence of *LINC00968* reduces the cell line's anchorage independent growth (Figure 5.3a), and reduction in cell colony counts reflects the inhibition of colony formation as a result of *LINC00968* elevated levels (Table 5.2 and Figure 5.3b).

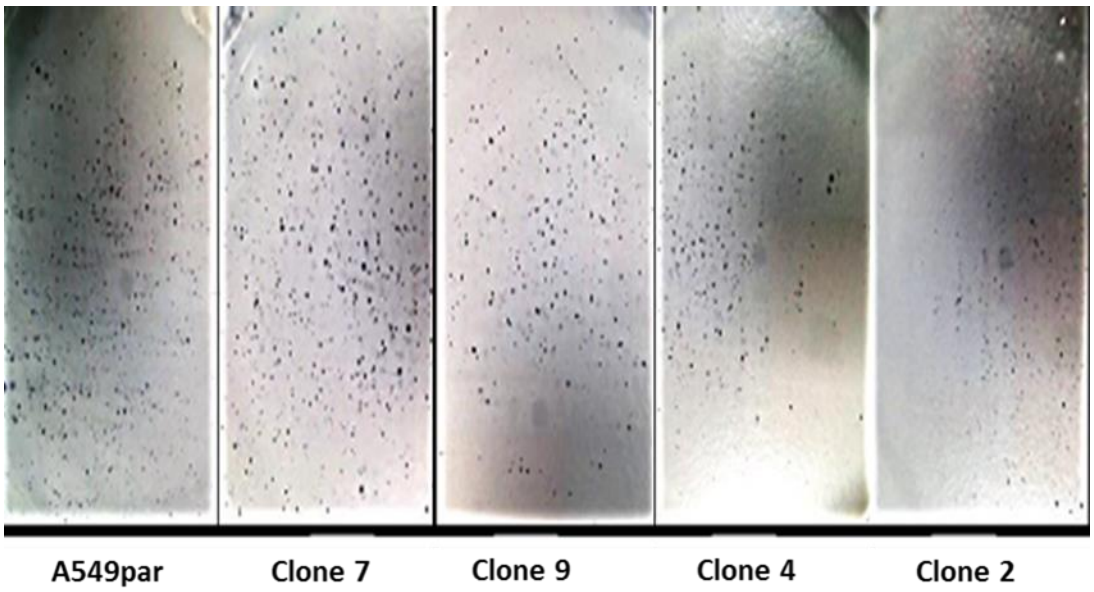


Figure 5.3a; Images of anchorage independent growth assay of A549 arental cell and A549_LINC00968 overexpressing clones. The number of colonies created was inversely correlated with the amount of LINC00968 expressed by the cells (there was a marked decrease in cell colony count between the A549 parental, A549_LINC00968 clone 7 and the clones that expressed *LINC00968*.

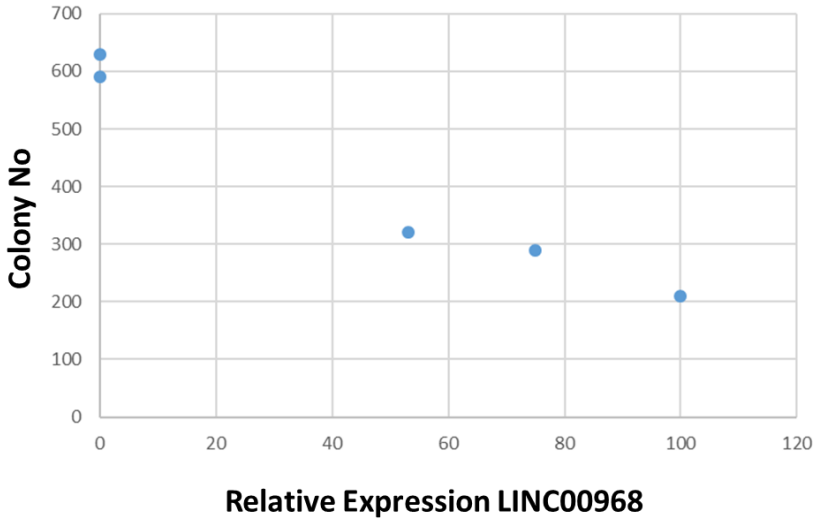


Figure 5.3b: Graph showing the inverse correlation between the relative expression of *LINC00968* and the cell colony count.

Table 5.2: Reduction in cell colony count observed following anchorage –independent growth analysis. Percentage expression is relative to *LINC00968* levels in Clone 2.

Cell Line/ Clones	Cell Count (Average)
A549 parental (0 expression)	630
Clone 7 (0 expression)	590
Clone 9 (53% expression)	320
Clone 4 (75% expression)	290
Clone 2 (100% expression)	210

5.5. Migration and Invasion Assay

Cancer metastasis is a complex and multistep process that is extremely related to cell motility, and also to the migratory and invasive potentials of cancer cells (Webb *et al*., 2011). Cancer cells spread by migrating and invading extracellular matrixes (ECM), entering into the blood circulation, attaching to distant sites, and finally erupting to form distant foci (Justus *et al*., 2014).

Wound Healing Assay

The wound healing assay as described in Section 2.2.9.3 was used to determine the role of LINC00968 in cell migration and interaction. Figures 5.4 (A – B) shows the gap area created by the wound analysed over time, the difference in gap area for the A549 parental cell line compared to the clones was taken over a 72-hour period. A complete closure of the wound area was observed at 72 hours for A549 parental cells and empty vector (Clone 7), both of which do not show detectable levels of the *LINC00968* . In contrast, the gaps in clones 2 and 4 (maximum and high *LINC00968* Clones) were not completely closed.

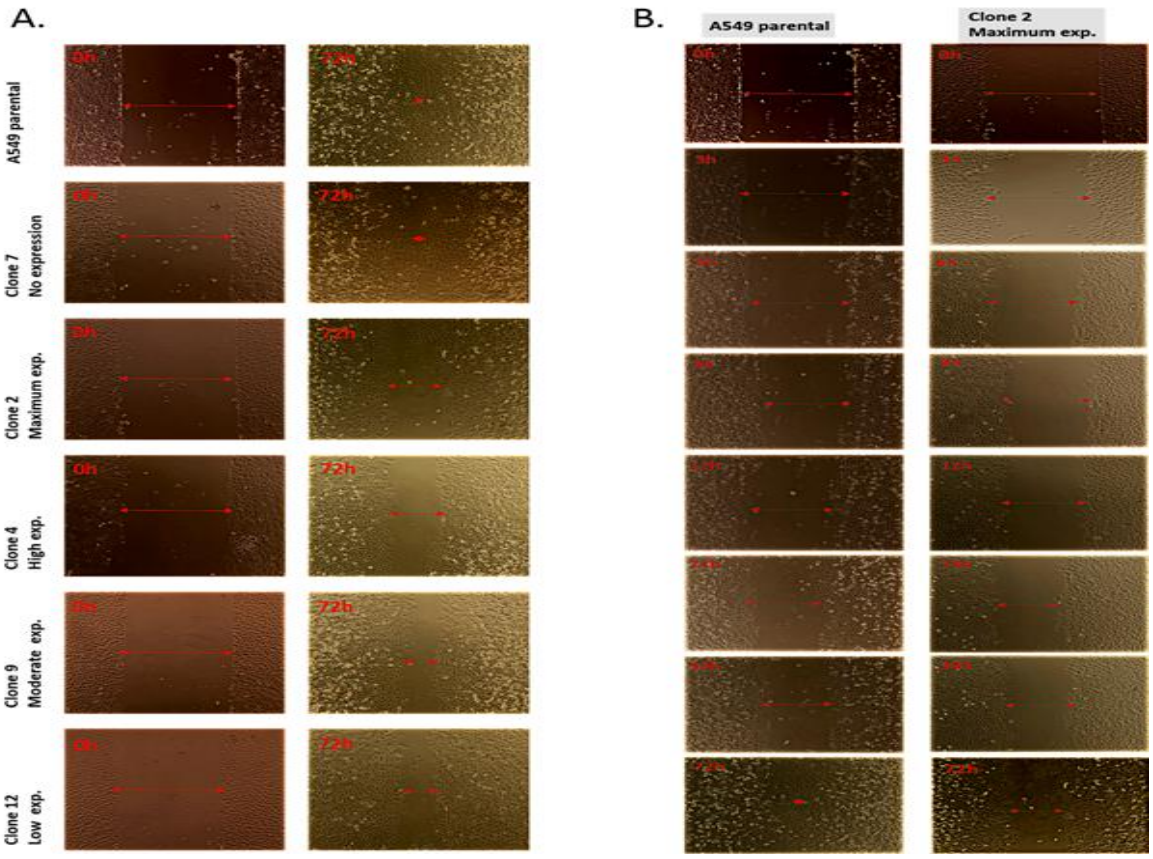


Figure 5.4: Scratch assay image showing extent of wound closure (A) between the A549 parental and the different Clones at 0h and 72h. (B) Wound closure in the A549 parental cell and Clone 2 throughout the different time points of the assay.

Boyden Chamber Assay

We went on further to explore the metastatic and invasive effects of *LINC00968* in our A549 parent cell line and selected A549 LINC00968 Clones using an adaptation of the Boyden Chamber Assay (Section 2.2.9.4).

There was markedly repressed migration observed in Clones 2 and 4 (maximum and high *LINC00968* levels respectively) compared to A549 parental cell and empty vector (Figure 5.5A). Presence of *LINC00968* also had an effect on invasion, reduction in number of cells can be observed in the Clones with maximum and high levels of *LINC00968* compared to the A549 parental and vector only which both had undetectable levels of the *LINC00968* lncRNA (Figure 5.5B).

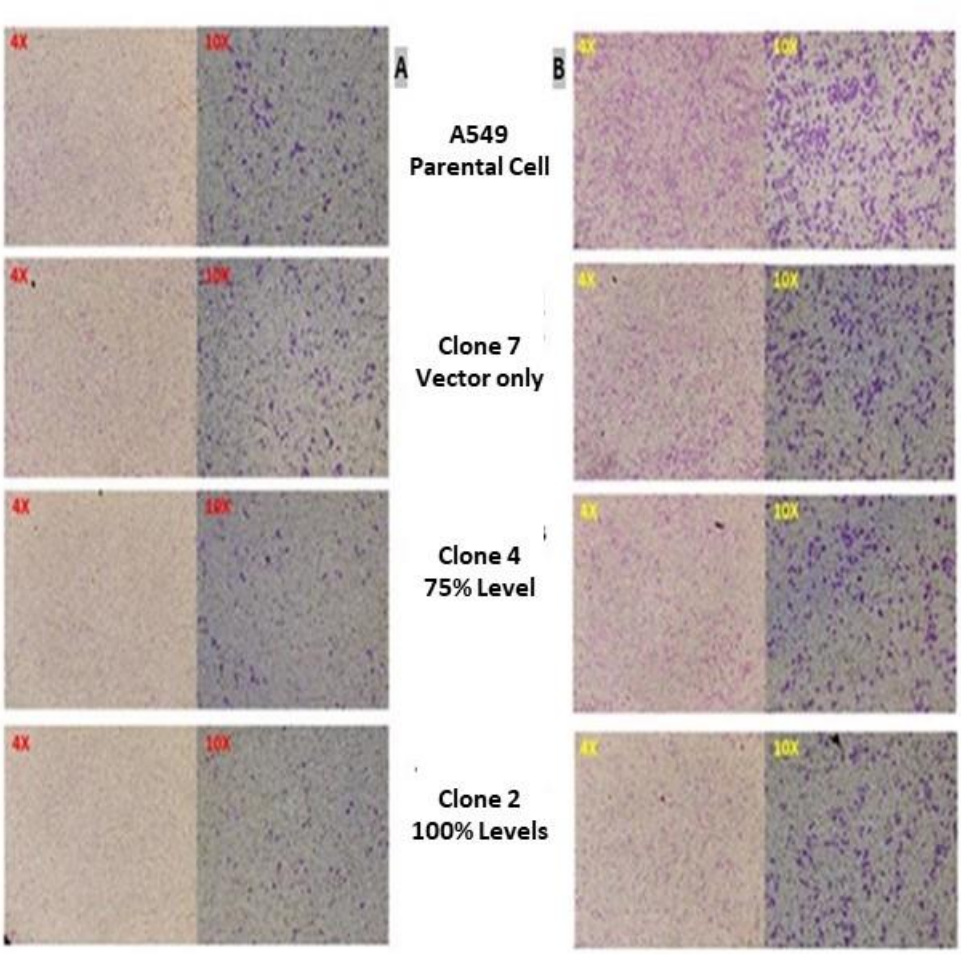


Figure 5.5: Migratory and invasive effect of *LINC00968* on the A549 parent cell line and the different A549 LINC00968 Clones. (A.) Migration assay in the absence of Geltrex matrix: overexpression reduced migration in the A549 clones with 100% and 75% elevated levels. (B.) Invasion assay using Geltrex matrix: marked reduction in invasive ability of LINC00968 expressing Clones was observed compared to A549 parental and the vector only clone.

5.6. *LINC00968* and Oxidative Stress

In order to assess the potential impact of *LINC00968* as an antioxidant in the cell's response to oxidative stress, the A549 parental cell line and the derivative clones were exposed to different concentrations of H_2O_2. Viability was measured using the MTT assay.

As it becomes evident from Figure 5.6, that low concentrations (25 μM) of H_2O_2 overexpression of *LINC00968* appears to have a negative impact on cell viability. The effect

is almost proportional to *LINC00968* levels. Dunnet's post-hoc test demonstrated differential significance (p=0.107, p= 0.026, and p=0.008 for the clones 9, 4 and 2 respectively) between parental and *LINC00968* overexpressing clones while there was no significance between parental and the vector only clone (p=0.661). At higher concentrations this effect is minimised. A significant difference was observed in all clones at 150 μM, however, the viability of cells is already limited and there is no difference between the vector only and *LINC00968* clones.

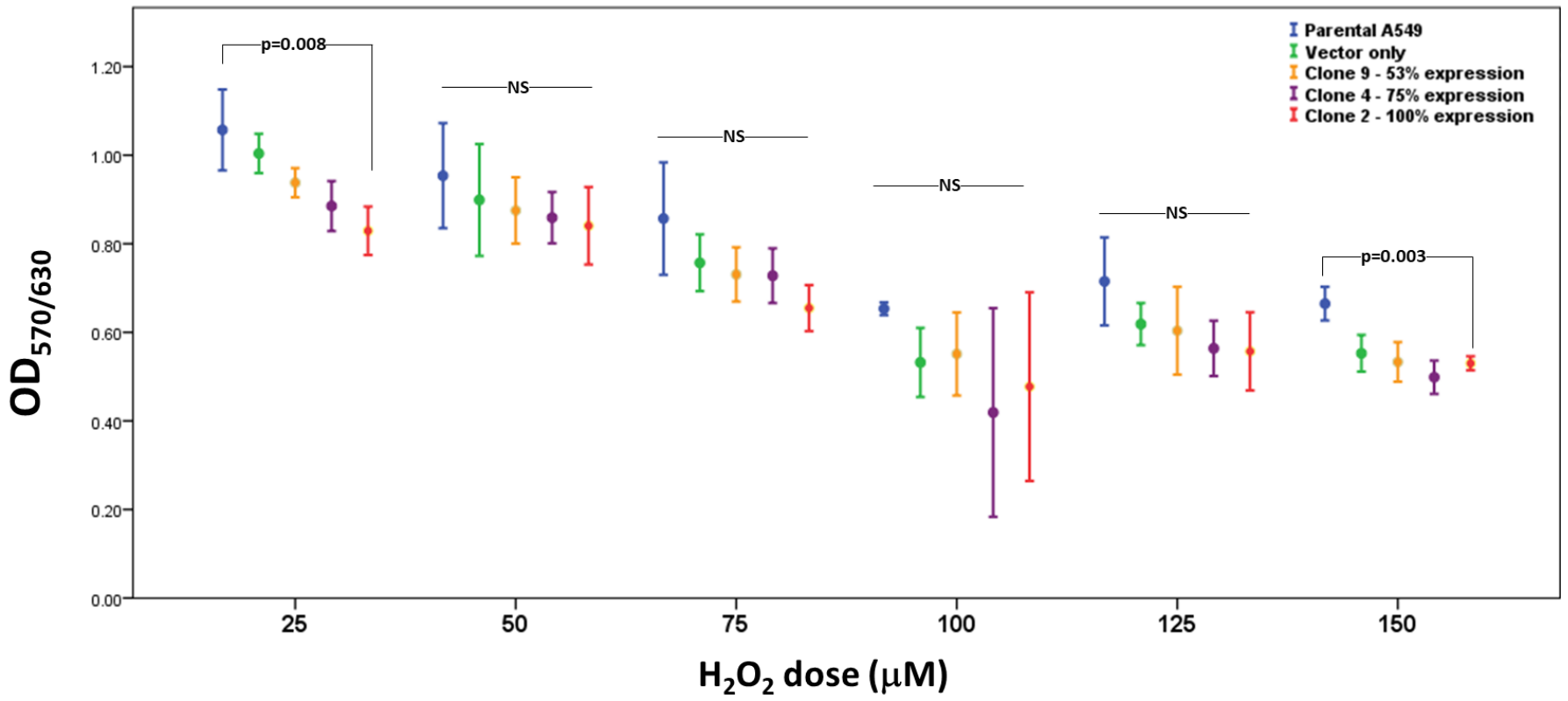


Figure 5.6: Line graph showing response of A549 parental cell line and the respective clones to different concentrations of H_2O_2.

5.7. Impact of *LINC00968* Expression on Drug Resistance

Resistance to anti-cancer drugs is one of the major reasons for the failure of lung cancer treatment. In this study, the role of *LINC00968* in modulating resistance ito some widely used drugs in lung cancer was investigated. A549 parental cell and selected A549 *LINC00968* clones were exposed to fluorouracil (5FU), gemcitabine, cisplatin, vinorelbine, methotrexate and pemetrexed to assess the potential modulation of resistance to commonly used chemotherapeutic drugs in NSCLC and their involvement in molecular pathways associated with the use of these drugs

Following drug treatment with 5FU, there was no significant difference in cell viability of the Clones (2, 4, 9, 12, 7) and the A549 parental cell line (Figure 5.7, Kruskal Wallis *test* p=0.37), irrespective of the expression level of the *LINC00968*. However, there was a statistical difference detected in the untreated cells and the clones expressing *LINC00968* (Figure 5.7C Mann Whitney test $p=5.6 \times 10^{-5}$).

The results obtained for gemcitabine and cisplatin were similar with no significant difference observed in relative cell survival between the A549 parental cell line, the empty vector and the selected clones (Figure 5.8 and 5.9).

Exposure of the A549 parental and Clones to methotrexate and pemetrexed showed a difference in cell viability. Clones 2, 4 and 9 demonstrated an increase in relative cell survival in an expression dependent manner compared to the parental cell and vector only. *LINC00968* may be involved in increasing the sensitivity of cancer cells to methotrexate and pemetrexed and may also be involved in their signalling pathways (Figures 5.11 and 5,12 respectively).

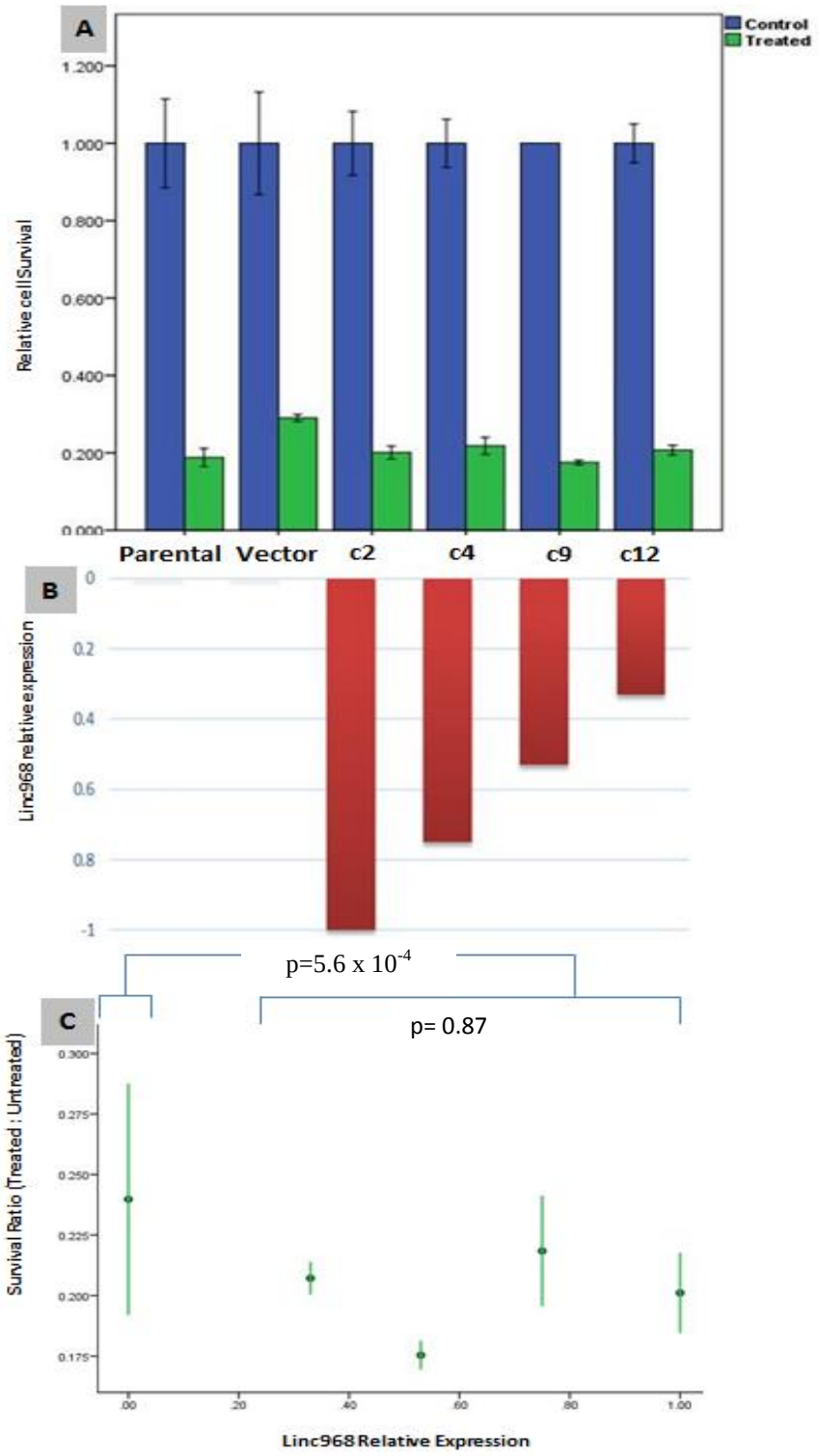


Figure 5.7: Bar chart and simple error bar showing cell viability response of the A549 parental cell and the respective Clones to 5-FU and their corresponding LINC00968 level. Significant effect on cell viability can be observed in the untreated cells. Error bars represent 95% confidence interval.

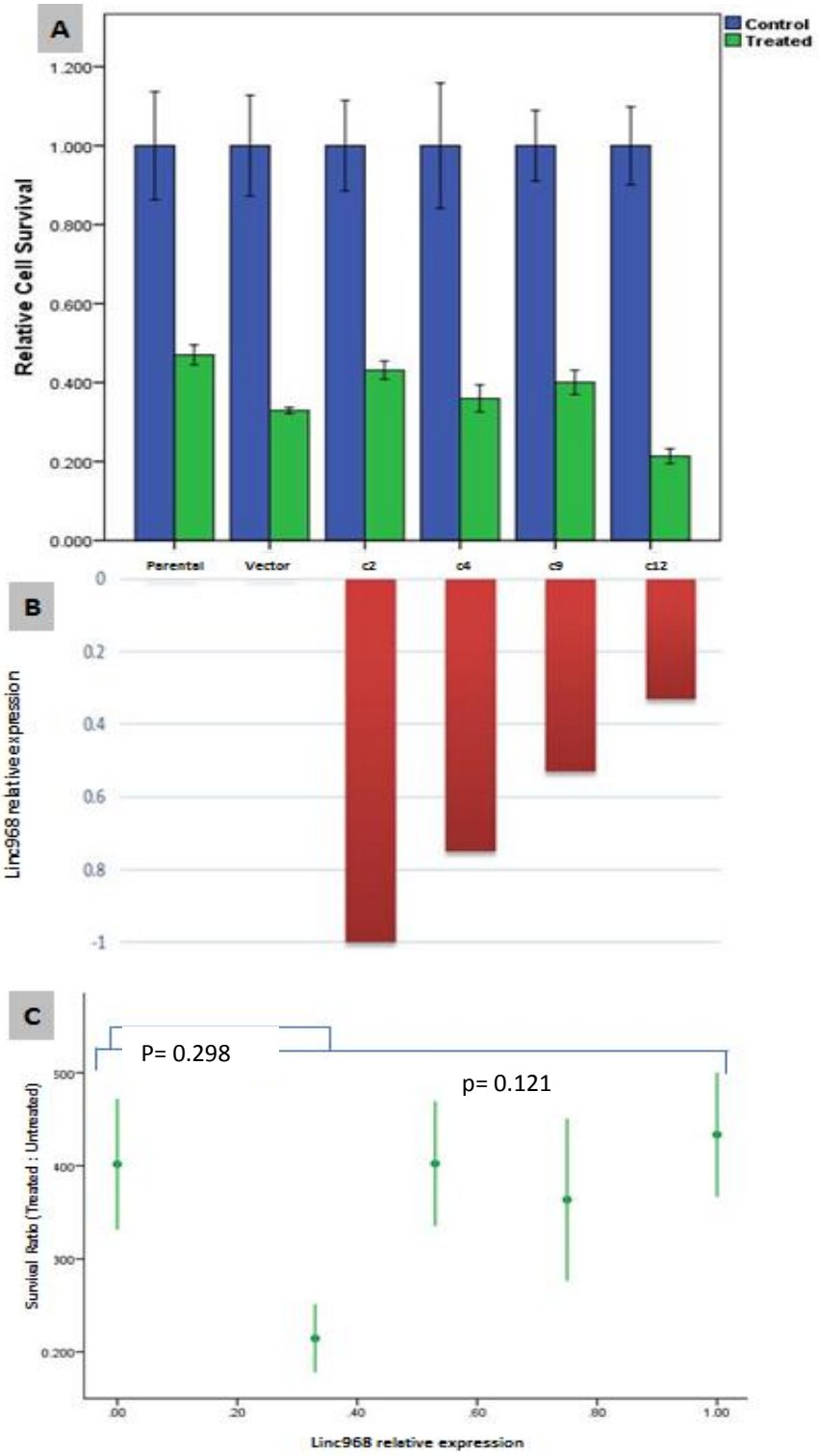


Figure 5.8: Bar chart and simple error bar showing cell viability response of the A549 parental cell and the respective Clones to Gemcitabine and their corresponding *LINC00968* levels. There is no statistical difference in cell viability of treated and untreated cells. Error bars represent 95% confidence interval.

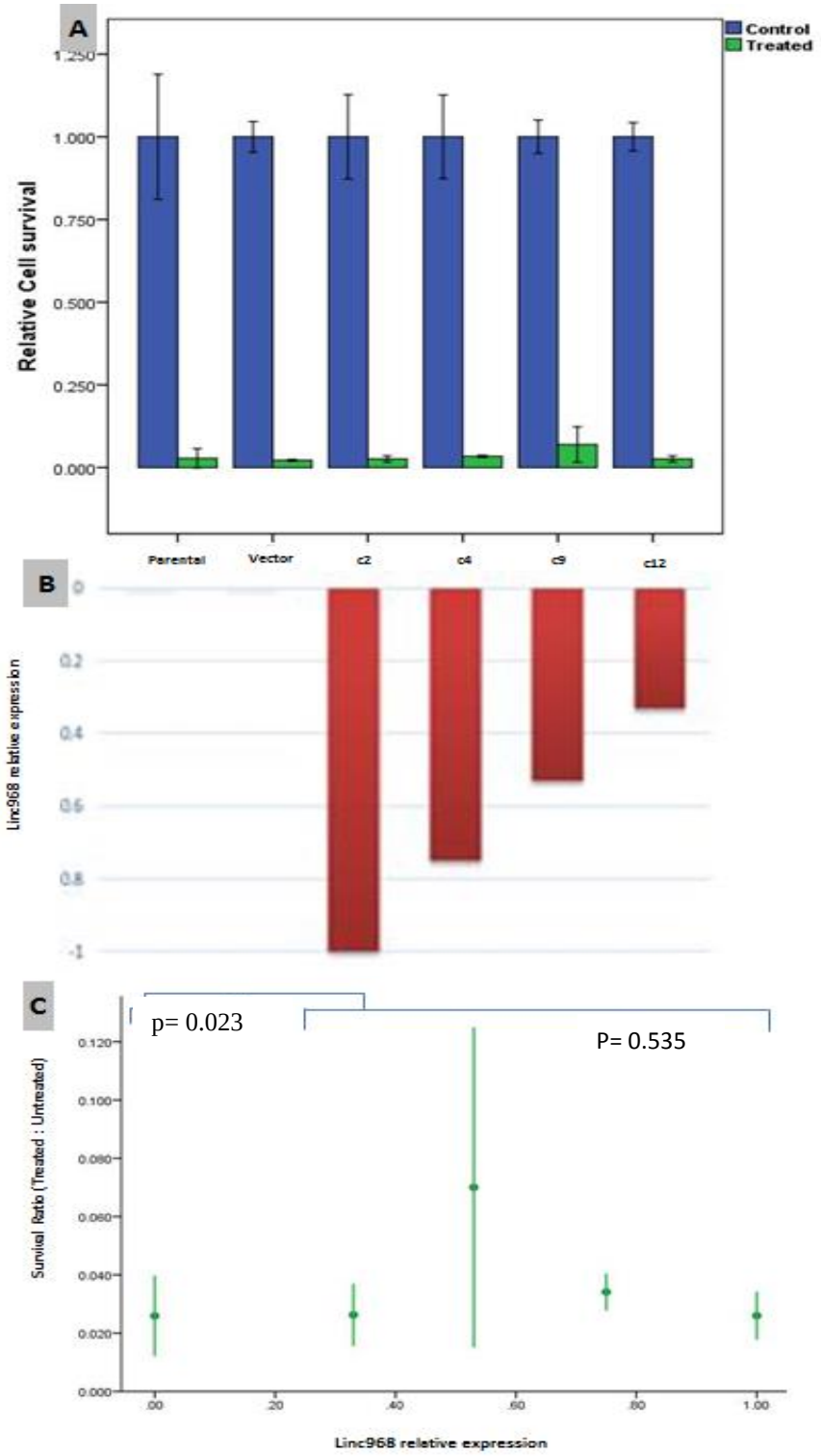


Figure 5.9: Bar chart and simple error bar showing cell viability response of the A549 parental cell and the respective clones to Cisplatin and their corresponding *LINC00968* levels. No statistical difference was observed in cell viability of treated and untreated cells. Error bars represent 95% confidence interval.

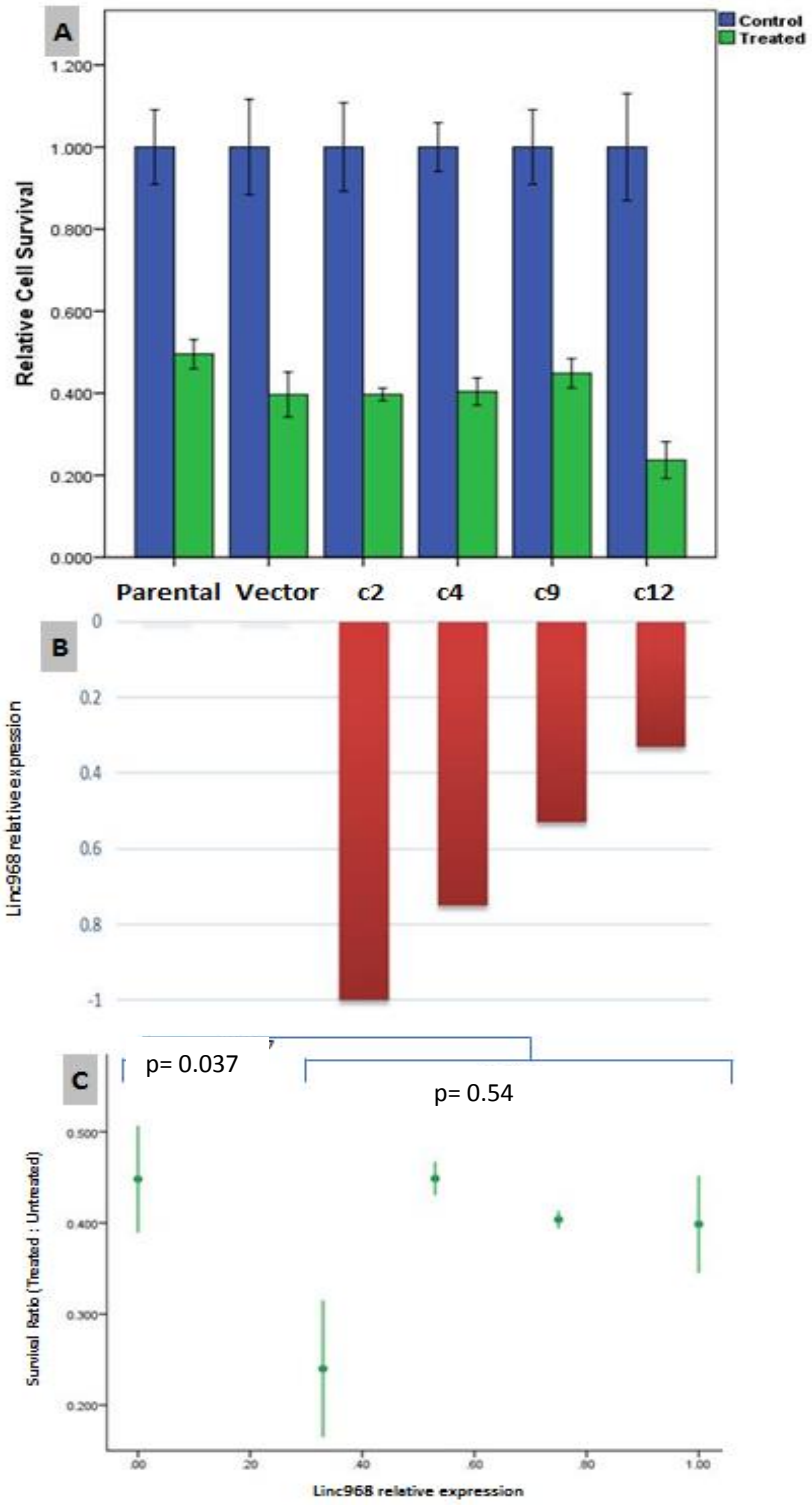


Figure 5.10: Bar chart and simple error bar showing cell viability response of the A549 parental cell and the respective Clones to Vinorelbine and their corresponding *LINC00968* levels. Error bars represent 95% confidence interval.

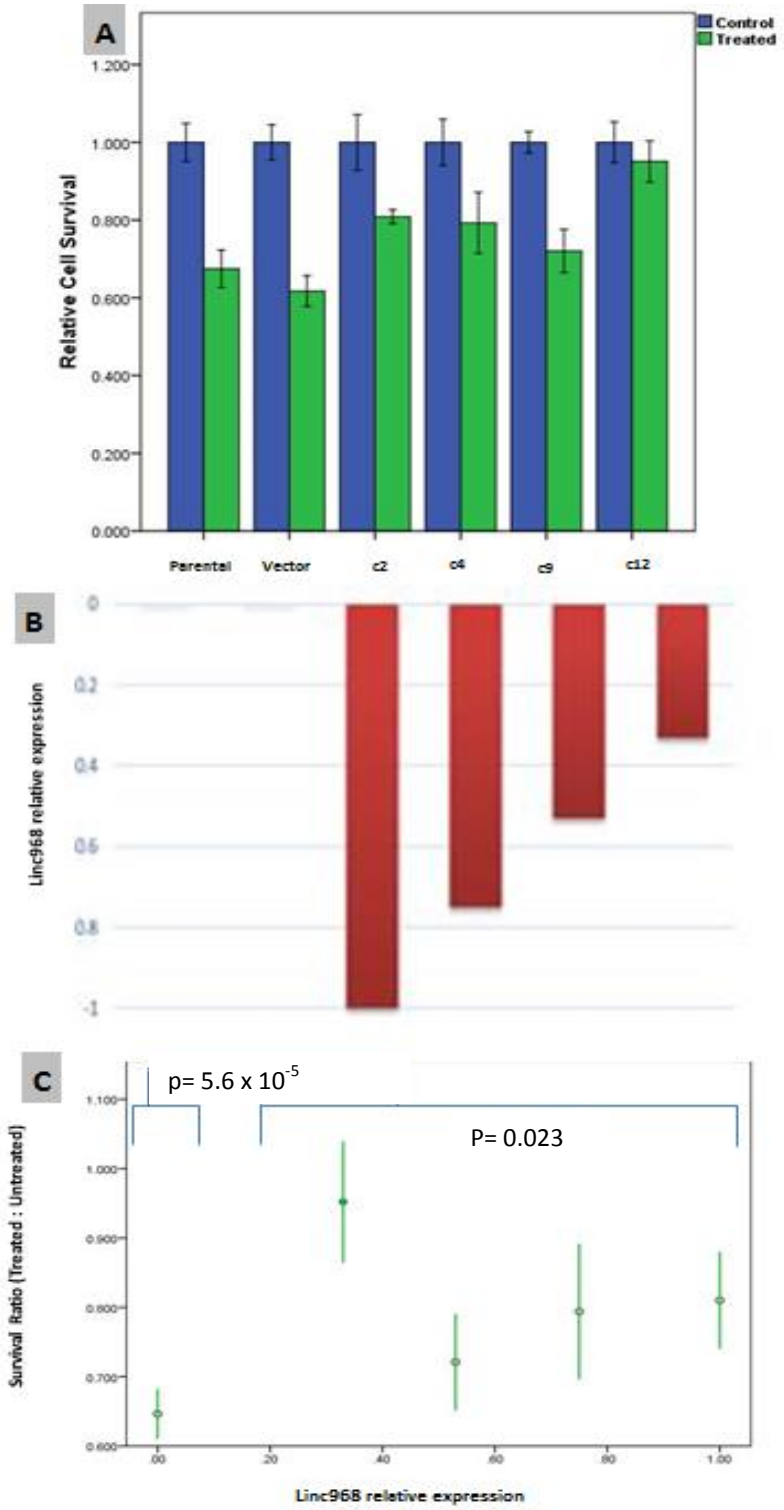


Figure 5.11: A.] Effect of methotrexate on cell viability, showing significant cell survival in the Clones expressing *LINC00968* compared to A549 parental cell and the empty vector. B.] Corresponding *LINC00968* levels of the A549 parental cell line and the clones C.] Survival ratio between the treated and untreated. Error bars are representative of 95% confidence interval.

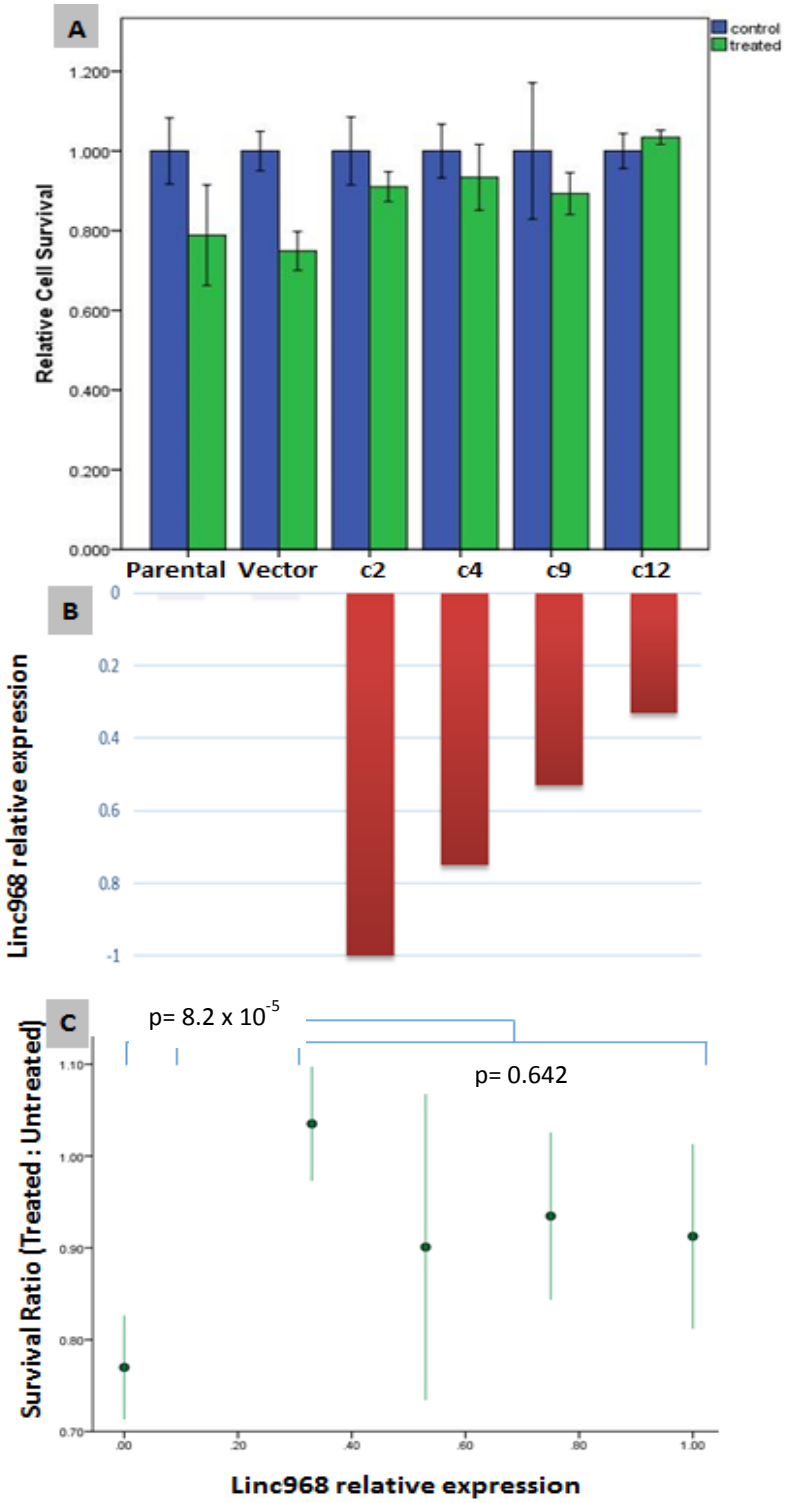


Figure 5.12: Effect of Pemetrexed on the A549 parental cell line and the different clones, showing significant cell survival in the clones expressing LINC00968 compared to A549 parental cell and the empty vector. B.] Corresponding *LINC00968* expression levels of the A549 parental cell line and the clones C.] Survival ratio between the treated and untreated. Error bars are representative of 95% confidence interval.

5.8. Discussion

LINC00968 was selected for functional study as one of the top repressed lncRNAs identified in our study and having minimal information on its function. A recent publication reported its upregulation in NSCLC, and that its inhibition repressed growth, migration and invasion by activating the Wnt signalling pathway (Wang *et al*., 2018). Although their findings do not correlate with the findings in our study, the possibility of the upregulation and downregulation of a gene promoting cancer in the same cell has been reported by Wang *et al*, who demonstrated that *SRPK1* (serine-arginine protein kinase 1) was shown to be either downregulated or overexpressed and that this aberrant expression of *SRPK1* in either direction brings about modulation of the Akt pathway (Wang *et al*., 2014). Similarly, there are numerous signal molecules that have demonstrated tumour suppressing and oncogenic functions in different cells (Feng, 2012).

Interestingly, the lowest proliferation rates were observed in clone 12, which was the clone with the least *LINC00968* levels. This is not currently understood, however, it is not necessarily a paradox; since the physiological levels of *LINC00968* gene expression are unknown and the 100% expression used in this study as a comparative term is an arbitrary selection referring to the highest achieved overexpression through our inducible system and not the physiological one.

In contrast to proliferation, the presence of *LINC00968* in A549 also resulted in suppression of anchorage independence in almost a proportional manner. Many cell types are known to require a form of anchorage dependence to survive, anchorage independence is therefore a phenotypic characteristic of transformed cells, and this has been linked to metastasis. A more moderate but clear effect was observed in migration and invasion properties of the

LINC00968 overexpressing clones, demonstrating suppression of both characteristics but in a manner not proportional to the level of the lncRNA. While cell migration by definition is the movement of individual cells, cell sheets and clusters from one location to another, cell invasion on the other hand is the 3-dimensional migration of cells as they penetrate an extracellular matrix (ECM) and is a process typically associated with cancer cell metastasis (Hulkower & Herber, 2011). Both cell migration and cell invasion occur as a result of epithelial-mesenchymal transition (EMT), which though integral to development have been shown to be reactivated in wound healing, fibrosis and cancer progression (Kalluri & Weinberg, 2009). Several molecular processes are involved in the initiation and completion of the EMT process and these include the activation of transcription factors, expression of specific cell-surface proteins, reorganization and expression of cytoskeletal proteins, production of ECM-degrading enzymes, changes in the expression of specific microRNAs (Kalluri & Weinberg, 2009) and more recently changes in the expression of lncRNAs in cancer (Dhamija, & Diederichs, 2016). We can state that *LINC00968* is an important regulator of migration and invasion in NSCLC *in vitro*, however, we are yet to understand the mechanism by which this is accomplished.

Oxidative stress is an important factor for cancer cell growth and survival, especially when tumours are treated with radiation therapy and certain types of chemotherapy (Conklin, 2004; Chen et al, 2007; Morry et al, 2017). We therefore tested the impact of *LINC00968* expression on oxidative stress. It is apparent that, under low H_2O_2 concentrations, restoration of *LINC00968* expression leads to reduced viability (Figure 5.6). As the function of *LINC00968* is not yet known, it is not possible to speculate the underlying mechanism. The next reasonable experiment to perform is to test apoptosis, hypothesizing that

LINC00968 expression triggers programmed cell death in the presence of oxidative stress, however there was not enough time to perform this experiment.

At higher H_2O_2 concentrations this effect is lost as the overall viability is gradually being reduced. One can hypothesize that as other molecules may be triggering cell death at those concentrations, the effect of *LINC00968* is hidden or indeed the function of the gene may become obsolete. The clarification of the mechanistics behind this observation opens an interesting research question and requires further investigation. At this point, the statistical difference seen at the highest H_2O_2 is not considered of any biological importance because the overall viability of all cells is low and the only difference is seen in the vector only clone, so it is considered circumstantial and irrelevant to *LINC00968* expression.

Chemotherapy is a major tool of treatment for the management of all cancers; however, the drug resistance development is by far the most important problem for successful treatment of cancer across a wide range of anti-cancer drugs. Drug resistance occurs not only due to individual pharmacogenetic variations in patients, but also as a result of genetic and epigenetic differences in tumours themselves. Recent studies have revealed that lncRNAs play a vital role in drug resistance in different cancers including lung cancer. Dysregulation of multiple targets and pathways by lncRNAs have been shown to result in the occurrence of drug resistance. *HOTAIR* has been shown to target P21 in NSCLC resulting in resistance to cisplatin, *GAS5* targets IG-1R to bring about gefitinib resistance in NSCLC and *MEG3* contributed to cisplatin resistance in NSCLC by targeting P53 and bcl-xl (Chen et al, 2017). We investigated *LINC00968* involvement in resistance of 6 anti-cancer drugs, commonly used in NSCLC therapy, to get a first impression of the involvement of this lncRNA in drug resistance. Overexpression of *LINC00968* did not have a significant effect on the cell survival

when compared to the A549 parental cell and the empty vector for fluorouracil (5FU), gemcitabine, cisplatin and vinorelbine. On the contrary, exposure to methotrexate and pemetrexed demonstrated that cells expressing *LINC00968* were more resistant to these drugs, compared with the A549 parental and the empty vector. While at the moment it is very difficult to draw conclusions of the mechanism involved in this resistance, it is safe to exclude the possibility of this being a secondary effect of the low proliferation rate that *LINC00968* induces. Both methotrexate and pemetrexed bind to DHFR (dihydrofolate reductase) enzyme inhibiting the conversion of folate to tetrahydrofolate, consequently affecting DNA synthesis and cell growth (Chen *et al*., 1984; Zhang *et al*., 2011). Although pemetrexed is also involved in the inhibition of thymidylate synthase (TS) and glycinamide ribonucleotide formyltransferase (GARFT) enzymes in the folate pathway, we suggest that *LINC00968* may be involved in methotrexate and pemetrexed resistance by interacting with the DHFR binding site.

The phenotypic changes induced by *LINC00968* overexpression in A549 cell line, while encompassing some variability, clearly demonstrate the negative impact of this lncRNA in proliferation, migration and invasion. *LINC00968* therefore demonstrates the properties of a putative tumour suppressor gene and may affect drug response,

6. CONCLUSION AND FUTURE WORK

6.1. Conclusion

This study is one of the very few comprehensive screenings of lncRNA expression in NSCLC. It successfully demonstrated that a number of lncRNAs are differentially expressed in NSCLC tumours when compared with paired normal tissues. We identified 20 upregulated and 30 down regulated lncRNAs through the expression microarray analysis and validated the top 13 by qRT PCR. We showed that the expression pattern of selected top induced (*FEZF1-AS1, LINC01214, , PCAT6, LOC101927229, LINC00673, NUTM2A-AS1* and *RNF139-AS1),* and top repressed *(LANCL1-AS1, FENDRR, LINC00968, SVIL-AS1,* and *PCAT19*) lncRNAs in our expression microarray was consistent with results obtained following qRT PCR validation except for *LOC105376287*. We discovered 6 uncharacterized lncRNAs (*LOC105376287, LOC101927229, NUTM2A-AS1, RNF139-AS1, SVIL-AS1* and *PCAT19*) in our study and have been able to determine their expression pattern in NSCLC tissues and a number of NSCLC cell lines.

There have been no major correlations with clinical descriptors, indicating that their deregulation is a universal phenomenon in NSCLC and most likely appear early in tumour development. It is possible that *FEZF1-AS1, LINC00968, LOC101927299* and *PCAT6* may be valuable as surrogate diagnostic markers in NSCLC but this will be subject to their detectability in clinical specimens such as blood, sputum or bronchial washings.

We also investigated DNA methylation as a potential molecular reason for the observed deregulation of those lncRNAs levels. We established that *FEZF1-AS1* and *FENDRR* were hypermethylated, while *NUTM2A-AS1* and *LINC0067*3 were hypomethylated, suggesting that epigenetic control of these lncRNAs is important in lung cancer development. These

results were strengthened by the *in* vitro analysis of *FEZF1-AS1* and *LINC00673* expression, following exposure to DNA methylation histone deacetylase inhibitors in four NSCLC cell lines. These experiments have clearly demonstrated the differential action of DAC and VPA on the different cell lines emphasizing on the dependence of the drug efficacy on the existing combination of genome and epigenome of tumour cells.

Finally, we selected one of the deregulated lncRNAs (*LINC00968*) for functional *in vitro* analysis. Restoration of expression of *LINC00968* in A549 cells demonstrated a clear shift in the phenotype. While the inhibition of proliferation was not proportional to *LINC00968* levels, a clearer inverse correlation was observed between *LINC00968* levels and anchorage-independent growth, *LINC00968* overexpression also repressed migration and invasion. One of the most interesting findings of this thesis is that *LINC00968* modulated resistance to methotrexate and pemetrexed in *LINC00968* expressing A549 cells. These are very popular second line chemotherapeutics used in NSCLC treatment and further research should be undertaken to establish the potential clinical value of this observation.

6.2. Future Work

The results from this study pose interesting research questions on two levels: cancer biology and clinical utility. On the first front, a major goal for future research will be to explore the relationship between the expression of identified lncRNAs and the expression of any associated mRNAs and proteins to enhance our understanding of the specific signalling/molecular pathways through which the deregulated lncRNAs interact.

6.2.1 Explore lncRNAs interactions to related mRNAs/proteins.

LINC00968 has been shown to act as an oncogene in NSCLC **by activating the Wnt signalling pathway**, although the findings suggested that *LINC00968* was significantly increased in LAD tissues, LSC tissues and NSCLC cell lines (Wang *et al*., 2018), which was contrary to our findings. It was also found to play an important role in the reduction of drug resistance in breast cancer cells by **inhibiting the Wnt2/B-catenin signalling pathway through silencing Wnt2** (Xiu *et al*., 2019). Therefore, it is of interest to consider examining the Wnt signalling pathway and any associated mRNA/protein in relation to *LINC00968* down regulation in NSCLC.

Experiment to enable the identification of mRNAs and/or proteins associated with down regulation of *LINC00968* will be carried out. To achieve this, RNA immunoprecipitation (RIP) and/or Chromatin immunoprecipitation (ChIP) technique will be employed using total RNA extracted from patient samples available from the Liverpool Lung Project Biobank. Cells will be cultured and harvested following the recommended protocol, the harvested cells will then be lysed and *LINC00968* co-precipitated with the protein of interest from the lysate. The RIP procedure involving the antibody prebinding step, cross-linking and wash step is performed, after which protein analysis (using western blot) and RNA analysis (using RT qPCR) may be carried out to analyse the protein or the RNA associated with *LINC00968*.

6.2.2 *In vitro* functional analysis on a selected lncRNA target

Of the 13 lncRNAs selected for further studies following the microarray analysis, only *LINC00968* was evaluated for functional *in vitro* analysis in this study, it is important therefore that the other lncRNAs targets be studied. Of interest is the *LOC101927229* that

was found to be upregulated in the microarray analysis and further validated by qPCR to be significantly upregulated in tumours compared to the adjacent paired normals. *LOC101927229* is still uncharacterised and to date no information is available to the best of my knowledge. To carry out in vitro functional analysis, shRNA knockdown protocol will be employed. *LOC101927229* was shown to be upregulated in the LUDLU1 NSCLC cell line, consequently LUDLU1 will be used for the knockdown experiment. shRNA construct for *LOC101927229* and negative control will be purchased, primer design will be done using the Primer3 Input (and purchased), while LUDLU1 cells will be cultured and plated under the recommended conditions and protocols before transfection using the FuGENE Transfection Reagent kit. *LOC101927229* shRNA will be transfected into LUDLU1 cell line. Following successful knockdown, clones with over 50% efficiency will be used for functional studies (cell proliferation, migration and invasion).

6.2.3 Identified lncRNAs as a biomarker in non-invasive samples

Finally, at the clinical utility front, the levels of the discovered lncRNAs will be tested in clinical samples such as sputum, plasma and bronchial washings, in order to determine their value as biomarkers for clinical management of the disease.

The high stability of lncRNAs while circulating bodily fluids makes them suitable for development as biomarkers that are none or minimally invasive. For a lncRNA to be considered as a biomarker (putative, prognostic, diagnostic or predictive), it must be differentially expressed in tumours compared to the adjacent normal tissues and validated by qPCR to be differentially expressed. It should also be measurable and specific. Although *LINC00968* was shown to be down regulated in the microarray results, which was confirmed by qPCR, it can be considered a putative biomarker, as further studies are still needed to

establish its specificity and availability in non-invasive samples. The biomarker screening will target a sub-population of the elderly (65 years and above in both men and women).

It will also be interesting to look at *LOC101927229, PCAT6* and *LANCL1-AS1* in relation to their potential as prognostic biomarkers due to their correlation with some of the clinicopathological parameters. We may also in the future consider biomarker development from elevated tumours, as it will enable the use of the siRNA or shRNA for knockdown experiments. Identification of biomarkers will aid in early detection of lung cancer, increase the 5 year survival time and consequently reduce mortality.

REFERENCES

Abidin, A.Z., Garassino, M.C., Calfano, R., Harle, A., & Blackhall, F. (2010). Targeted therapies in small cell lung cancer: A review. *Therapeutic Advances in Medical Oncology, 2*(1), 25-37. doi:10.1177/1758834009356014

Amaral, P. P., Dinger, M. E., Mercer, T. R. & Mattick, J. S. (2008). The eukaryotic genome as an RNA machine. *Science, 319*(5871), 1787-1789. doi:10.1126/science.1155472

Amini, A., Yeh, N., Gaspar, L.E., Kavanagh, B., & Karam, S.D. (2014). Stereotactic body radiation therapy (SBRT) for lung cancer patients previously treated with conventional radiotherapy: A review. *Radiation Oncology, 9*, 210. doi:10.1186/1748-717X-9-210

Amos, C. I., Wu, X., Broderick, P., Gorlov, I. P., Gu, J., Eisen, T., … Houlston, R. S. (2008). Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nature Genetics, 40*(5), 616-622. doi:10.1038/ng.109

Bartel, D. P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell, 116*(2), 281-297. doi:10.1016/S0092-8674(04)00045-5

Bartonicek, N., Maag, Jesper, L. V., & Dinger, M. E. (2016). Long noncoding RNAs in cancer: Mechanisms of action and technological advancements. *Molecular Cancer, 15*(1), 43. doi:10.1186/s12943-016-0530-6

Baylin, S. (2005). DNA methylation and gene silencing in cancer. *Nature Clinical Practice Oncology, 2*, S4-S11. doi:10.1038/ncponc0354

Belinsky, S. A. (2004). Gene-promoter hypermethylation as a biomarker in lung cancer. *Nature Reviews Cancer, 4*, 707-717. doi:10.1038/nrc1432

Blandin Knight, S., Crosbie, P. A., Balata, H., Chudziak, J., Hussell, T., & Dive, C. (2017). Progress and prospects of early detection in lung cancer. *Open Biology, 7*(9), 170070. doi:10.1098/rsob.170070

Bolha, L., Ravnik-Glavač, M., & Glavač, D. (2017). Long noncoding RNAs as biomarkers in cancer. *Disease Markers, 7243968* doi:10.1155/2017/7243968

Bossé, Y., and Amos, C. I. (2018). A decade of GWAS results in lung cancer. *Cancer Epidemiol Biomarkers Preview, 27*(4), 363-379. doi:10.1158/1055-9965.EPI-16-0794

Bracken, C. P., Scott, H. S., & Goodall, G. J. (2016). A network-biology perspective of microRNA function and dysfunction in cancer *Nature Reviews Genetics, 17*, 719-732. doi:10.1038/nrg.2016.134

Braconi, C., Valeri, N., Kogure, T., Gasparini, P., Huang, N., Nuovo, Gerard J., Terracciano, L., Croce, C. M. Patel, T. (2011). Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma *Proceedings of the National Academy of Sciences, 108*(2), 786-791. doi:10.1073/pnas.1011098108

Bradley, S. H., Kennedy, M., & Neal, R. D. (2019). Recognising lung cancer in primary care. *Advances in Therapy, 36*(1), 19-30. doi:10.1007/s12325-018-0843-5

Brandao, G.D.A., Brega, E.F., & Spatz, A. (2012). The role of molecular pathology in non-small cell lung carcinoma- now and in the future. *Current Oncology, 19*(Suppl 1), S24-S32. doi:10.3747/co.19.1058

Calin, G. A., Dumitru, C. D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M.... & Croce, C. M. (2002). Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences, 99*(24), 15524-15529. doi:10.1073/pnas.242606799

Calin, G. A., Liu, C-g., Ferracin, M., Hyslop, T., Spizzo, R., Sevignani, C., Fabbri, M.... & Croce, Carlo M. (2007). Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell, 12*(3), 215-229. doi:10.1016/j.ccr.2007.07.027

Cao, D., Ding, Q., Yu, W., Gao, M., & Wang, Y. (2016). Long noncoding RNA *SPRY4-IT1* promotes malignant development of colorectal cancer by targeting epithelial–mesenchymal transition. *Oncotargets and Therapy, 9*, 5417-5425. doi:10.2147/OTT.S111794

Cao, J. (2014). The functional role of long non-coding RNAs and epigenetics. *Biological Procedures Online, 16*(11) doi:10.1186/1480-9222-16-11

Carninci, T., Kusukawa, P., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R..., FANTOM Consortium., & RIKEN Genome Exploration Group. (2005). The transcriptional landscape of the mammallian genome. *Science, 309*(5740), 1559-1563. doi:10.1126/science.1112014

Carthew, R. W., & Sontheimer, E. J. (2009). Origins and mechanisms of miRNAs and siRNAs. *Cell, 136*(4), 642-655. doi:10.1016/j.cell.2009.01.035

Chan, B.A., & Hughes, B.G.M. (2015). Targeted therapy for non-small cell lung cancer: Current standards and the promise of the future. *Translational Lung Cancer Research, 4*(1), 36-54. doi:10.3978/j.issn.2218-6751.2014.05.01

Cheetham, S. W., Gruhl, F., Mattick, J. S., & Dinger, M. E. (2013). Long noncoding RNAs and the genetics of cancer *British Journal of Cancer, 108*, 2419. doi:10.1038/bjc.2013.233

Chen, F-J., Sun, M., Li, S-Q., Wu, Q-Q., Ji, Lv., Liu, Z-L., Zhou, G-Z., Cao, G. Jin, L., Xie, H-W., Wang, C-M., Lv,J., De, W., Wu, M., Cao, X-F. (2013). Upregulation of the long non-coding rna hotair promotes esophageal squamous cell carcinoma metastasis and poor prognosis. *Molecular Carcinogenesis, 52*(11), 908-915. doi:10.1002/mc.21944

Chen, J., Hu, L., Wang, J., Zhang, F., Chen, J., Xu, G., Wang, Y., Pan, Q. (2017). Low expression LncRNA TUBA4B is a poor predictor of prognosis and regulates cell proliferation in non-small cell lung cancer. *Pathology & Oncology Research., 23*, 265-270. doi:10.1007/s12253-016-0089-y

Chen, J.-H., Zhou, L.-Y., Xu, S., Zheng, Y.-L., Wan, Y.-F., & Hu, C.-P. (2017a). Overexpression of lncRNA HOXA11-AS promotes cell epithelial–mesenchymal transition by repressing miR-200b in non-small cell lung cancer. *Cancer Cell International, 17*(64) doi:10.1186/s12935-017-0433-7

Chen, L-L., Carmichael, G. G. (2010). Long noncoding RNAs in mammalian cells: What, where, and why? *Wiley Interdisciplinary Reviews - RNA, 1*(1), 2-21. doi:10.1002/wrna.5

Chen, N., Guo, D., Xu, Q., Yang, M., Wang, D., Peng, M., … Zhou, J. (2016). Long non-coding RNA *FEZF1-AS1* facilitates cell proliferation and migration in colorectal carcinoma. *Oncotarget, 7*(10), 11271-11283. doi:10.18632/oncotarget.7168

Chen, Q., Wei, C., Wang, Z., & Sun, M. (2017). Long non-coding RNAs in anti-cancer drug resistance. *Oncotarget, 8*(1), 1925-1936. doi:10.18632/oncotarget.12461

Chen, V.W., Ruiz, B.A., Hsieh, M.-C., Wu, X. -C., Ries, L., & Lewis, D.R. (2014a). Analysis of stage and Clinical/Prognostic factors for lung cancer from SEER registries: AJCC staging and collaborative stage data collection system. *Cancer, 120*, 3781-3792. doi:10.1002/cncr29045

Chen, X., Yan, C. C., Zhang, Xu., You, Z-H. (2016). Long non-coding RNAs and complex diseases: From experimental results to computational models. *Briefings in Bioinformatics, 18*(4), 558-576. doi:10.1093/bib/bbw060

Chen, Y., Jungsuuwadee, P., Vore, M., Butterfield, D. A., & St Clair, D. K. (2007). Collateral damage in cancer chemotherapy: Oxidative stress in nontargeted tissues. *Molecular Interventions, 7*(3), 147-156. doi:10.1124/mi.7.3.6

Chheang, S., & Brown, K. (2013). Lung cancer staging: Clinical and radiologic perspectives. *Seminars in Interventional Radiology, 30*(2), 99-113. doi:10.1055/s-0033-1342950

Clark, M. B., & Mattick, J. S. (2011). Long noncoding RNAs in cell biology. *Seminars in Cell & Developmental Biology, 22*(4), 366-376. doi:10.1016/j.semcdb.2011.01.001

Colombo, T., Farina, L., Macino, G., & Paci, P. (2015). PVT1: A rising star among oncogenic long noncoding RNAs. *BioMed Research International, 304208* doi:10.1155/2015/304208

Conklin, K. A. (2004). Chemotherapy-associated oxidative stress: Impact on chemotherapeutic effectivesness. *Integrative Cancer Therapies, 3*(4), 294-300. doi:10.1177/1534735404270335

Cosaert, j., & Quoix, E. (2002). Platinum drugs in the treatment of non-small cell lung cancer. *British Journal of Cancer, 87*(8), 825-833. doi:10.1038/sj.bjc.6600540

Cui, D., Yu, C. H., Liu, M., Xia, Q-Q., Zhang, Y-F., & Jiang, W-L. (2016). Long non-coding RNA PVT1 as a novel biomarker for diagnosis and prognosis of non-small cell lung cancer. *Tumor Biology, 37*(3), 4127-4134. doi:10.1007/s13277-015-4261-x

Dana, H., Chalbatani, G. M., Mahmoodzadeh, H., Karimloo, R., Rezaiean, O., Moradzadeh, A., … Gharagouzlo, E. (2017). Molecular mechanisms and biological functions of siRNA. *International Journal of Biomedical Science, 13*(2), 48-57.

Daskalos, A., Nikolaidis, G., Xinarianos, G., Savvari, P., Cassidy, A., Zakopoulou, R., . . . Liloglou, T. (2009). Hypomethylation of retrotransposable elements correlates with genomic

instability in non-small cell lung cancer. *International Journal of Cancer, 124*(1), 81-87. doi:10.1002/ijc.23849

de Baere, T., Farouil, G., & Deschamps, F. (2013). Lung cancer ablation: What is the evidence? *Seminars in Interventional Radiology, 30*(2), 151-156. doi:10.1055/s-0033-1342956

de Kok, J. B., Verhaegh, G. W., Roelofs, R. W., Hessels, D., Kiemeney, L. A., Aalders, T. W., Swinkels, D. W., & Schalken, J. A. (2002). A very sensitive and specific marker to detect prostate tumors. *Cancer Research, 62*(9), 2695-2698.

Delaney, C., Garg, S. K., & Yung, R. (2015). Analysis of DNA methylation by pyrosequencing. *Methods in Molecular Biology, 1343*, 249-264. doi:10.1007/978-1-4939-2963-4_19

Deng, J., Tang, J., Wang, G., & Zhu, Y.-S. (2017). Long non-coding RNA as potential biomarker for prostate cancer: Is it making a difference? *International Journal of Environmental Research and Public Health, 14*(3), 270. doi:10.3390/ijerph14030270

Detterbeck, F.C., Boffa, D.J., Kim, A.W., & Tanoue, L.T. (2017). The eight edition lung cancer stage classification. *Chest, 151*(1), 193-203. doi:10.1016/j.chest.2016.10.010

Dhamija, S., & Diederichs, S. (2016). From junk to master regulators of invasion: LncRNA functions in migration, EMT and metastasis. *International Journal of Cancer, 139*(2), 269-280. doi:10.1002/ijc.30039

Dong, Y., Ren, W., Qi, J., Jin, B., Li, Y., Tao, H., ...Han, B. (2016). EGFR, ALK, RET, KRAS and BRAF alterations in never-smokers with non-small cell lung cancer. *Oncology Letters, 11*(4), 2371-2378. doi:10.3892/ol.2016.4235

Dupuy, D.E., Zagoria, R.J., Akerley, W., Mayo-Smith, W.W., Kavanagh, P.V., & Safran, H. (2000). Percutaneous radiofrequency ablation of malignancies in the lung. *American Journal of Roentgenology, 174*, 57-59. doi:10.2214/ajr.174.1.1740057

Ehrlich, M. (2002). DNA methylation in cancer: Too much, but also too little. *Oncogene, 21*, 5400-5413. doi:10.1038/sj.onc. 1205651

Ehrlich, M. (2009). DNA hypomethylation in cancer cells. *Epigenomics, 1*(2), 239-259. doi:10.2217/epi.09.33

Esteller, M. (2005). Aberrant DNA methylation as cancer-inducing mechanism. *Annual Review of Pharmacology and Toxicology, 45*, 629-656. doi:10.1146/annurev.pharmtox.45.120403.095832

Esteller, M. (2011). Non-coding RNAs in human disease *Nature Reviews Genetics, 12*, 861. doi:10.1038/nrg3074

Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., … Uhlén, M. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics, 13*(2), 397-406. doi:10.1074/mcp.M113.035600

Fang, Y., & Fullwood, M. J. (2016). Roles, functions, and mechanisms of long non-coding RNAs in cancer. *Genomics, Proteomics & Bioinformatics, 14*(1), 42-54. doi:10.1016/j.gpb.2015.09.006

Fassan, M., Dall'Olmo, L., Galasso, M., Braconi, C., Pizzi, M., Realdon, S., … Rugge, M. (2014). Transcribed ultraconserved noncoding RNAs (T-UCR) are involved in Barrett's esophagus carcinogenesis. *Oncotarget, 5*(16), 7162-7171. doi:10.18632/oncotarget.2249

Feng, G. (2012). Conflicting roles of molecules in hepatocarcinogenesis: Paradigm or paradox. *Cancer Cell, 21*(2), 150-154. doi:10.1016/j.ccr.2012.01.001

Feng, N., Ching, T., Wang, Y., Liu, B., Lin, H., Shi, O., ... Qian, B. (2016). Analysis of microarray data on gene expression and methylation to identify long non-coding RNAs in non-small cell lung cancer. *Scientific Reports, 6*(37233) doi:10.1038/srep37233

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coeburgh, J.W.W., Comber, H., Forman, D., & Bray, F. (2013). Cancer incidence and mortality patterns in europe: Estimates for 40 countries in 2012. *European Journal of Cancer, 49*(6), 1374-1403. doi:10.1016/j.ejca.2012.12.027

Field, R.W., & Withers, B.L. (2012). Occupational and environmental causes of lung cancer. *Clinics in Chest Medicine, 33*(4) doi:10.1016/j.ccm.2012.07.001

Foster, N. A., Banerjee, A. K., Xian, J., Roberts I.,. Rabbitts, P. H. (2005). Somatic genetic changes accompanying lung tumor development. *Genes Chromosomes Cancer, 44*(1), 65-75.

FriedlÃ¤nder, M. R., Lizano, E., Houben, A. J. S., Bezdan, D., .. & Estivill, X. (2014). Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biology, 15*(4), R57. doi:10.1186/gb-2014-15-4-r57

Furrukh, M. (2013). Tobacco smoking and lung cancer: Perception-changing facts. *Sultan Qaboos University Medical Journal, 13*(3), 345-358.

Gao, L., Ren, W., Zhang, L., Li, S., Kong, X., Zhang, H., Dong, J., Cai, G., Jin, C., Zheng, D., & Zhi K. (2017). PTENp1, a natural sponge of miR-21, mediates PTEN expression to inhibit the proliferation of oral squamous cell carcinoma. *Molecular Carcinogenesis, 56*(4), 1322-1334. doi:10.1002/mc.22594.

Garcia-Perez, J. L., Morell, M., Scheys, J. O., Kulpa, D. A., Morell, S., Carter, C. C., … Moran, J. V. (2010). Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature, 466*(7307), 769-773. doi:10.1038/nature09209

Gazdar, A F., Bunn, P A. & Minna, J D. (2017). Small-cell lung cancer: What we know, what we need to know and the path forward. *Nature Reviews Cancer, 17*, 725-737. doi:10.1038/nrc.2017.87

Giaconne, G. (Ed.). (2012). *Systemic treatment of non-small cell lug cancer*. Oxford, United Kingdom: Oxford University Press.

Giltnane, J. M. & Rimm, D. L. (2004). Technology insight: Identification of biomarkers with tissue microarray technology . *Nature Clinical Practice Oncology, 1*, 104-111. doi:10.1038/ncponc0046

Girard, L., Rodriguez-Canales, J., Behrens, C., Thompson, D.M., Botros, I.W., Tang, H., Xie, Y., Rekhtman, N, Travis, W.D, Wistuba, I.I., Minna, J.D. & Gazdar, A.F. (2016). An expression signature as an aid to the histologic classification of non-small cell lung cancer. *Clinical Cancer Research, 22*(19), 4880-4889. doi:10.1158/1078-0432.CCR-15-2900

Globocan (IARC). (2018). Lung cancer mortality. Retrieved from www.globocan.iarc.fr/Pages/fact_sheets_cancer.aspx?cancer=lung

Goldberg, A.D., Allis,C.D., Bernstein, E. (2007). Epigenetics: A landscape takes shape. *Cell, 128*(4), 635-638. doi:10.1016/j.cell.2007.02.006

Goll, M. G., & Bestor,T. H. (2005). Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry, 74*(1), 481-514. doi:10.1146/annurev.biochem.74.010904.153721

Gonzalez, I., Munita, R., Agirre, E., Dittmer, T. A., Gysling, K., Misteli, T., & Luco, R. F. (2015). A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature *Nature Structural and Molecular Biology, 22*, 370. doi:10.1038/nsmb.3005

Grote, P., Wittler, L., Währisch, S., Hendrix, D., Beisaw, A., Macura, K., … Herrmann, B. G. (2013). The tissue-specific lncRNA *fendrr* is an essential regulator of heart and body wall development in the mouse. *Developmental Cell, 24*(2), 206-214. doi:10.1016/j.devcel.2012.12.012

Guil, S., & Esteller, M. (2012). Cis-acting noncoding RNAs: Friends and foes. *Nature Structural & Molecular Biology, 19*, 1068-1075. doi:10.1038/nsmb.2428

Gujar, H., Weisenberger, D. J., & Liang, G. (2019). The roles of human DNA methyltransferases and their isoforms in shaping the epigenome. *Genes, 10*(2), 172. doi:10.3390/genes10020172

Guo, J., Yu, J., Song, X., & Mi, H. (2017). Serum CA125, CA199 and CEA combined detection for epithelial ovarian cancer diagnosis: A meta-analysis. *Open Medicine, 12*, 131-137. doi:10.1515/med-2017-0020

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., … Chang, H. Y. (2010). Long noncoding RNA HOTAIR reprograms chromatin state to promote cancer metastasis.

Nature, 464(7291), 1071-1076. doi:10.1038/nature08975

Gutschner, T., & Diederichs, S. (2012). The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biology, 9*(6), 703-719. doi:10.4161/rna.20481

Ha, M., & Kim, V. M. (2014). Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology, 15*(8), 509-524. doi:10.1038/nrm3838

Han L, Kong R, Yin DD, Zhang EB, Xu TP, De W, Shu YQ. (2013). Low expression of long noncoding RNA GAS6-AS1 predicts a poor prognosis in patients with NSCLC. *Medical Oncology, 30*(4), 694.

He, R., Zhang, F h., & Shen, N. (2017). LncRNA FEZF1-AS1 enhances epithelial-mesenchymal transition (EMT) through suppressing E-cadherin and regulating WNT pathway in non-small cell lung cancer (NSCLC). *Biomedicine & Pharmacotherapy, 95*(Supplement C), 331-338. doi:https://doi.org/10.1016/j.biopha.2017.08.057

Herbst, R S., Morgensztern, D. & Boshoff, C. (2018). The biology and management of non-small cell lung cancer *Nature, 553*, 446-454. doi:10.1038/nature25183

Hernando-Herraez, I., Garcia-Perez, R., Sharp, A. J., & Marques-Bonet, T. (2015). DNA methylation: Insights into human evolution. *PLoS Genetics, 11*(12), e1005661. doi:10.1371/journal.pgen.1005661

Herrera-Merchan, H., Cuadros, M., Rodriguez, S., Torres, R., Farez, E., & Medina, P. (2016). Long non-coding RNA fendrr as biomarker for lung cancer. [Abstract]. *Proceedings of the 107th Annual Meeting of the American Association for Cancer Research, 76*(S14) doi:10.1158/1538-7445.AM2016-976

Hiraki, T., Gobara, H.,Iguchi, T., Fujiwara, H., Matsui, Y., & Kanazawa, S. (2014). Radiofrequency ablation for early-stage non-small cell lung cancer. *BioMed Research International, 152087* doi:10.1155/2014/152087

Hirsh, V. (2017). New developments in the treatment of advanced squamous cell lung cancer: Focus on afatinib. *OncoTargets and Therapy, 10*, 2513-2526.

Holgersson, G., Ekman, S., Reizenstein, J., Bergqvist, M., Pontén, F., ... & Bergström, S. (2010). Molecular profiling using tissue microarrays as a tool to identify predictive

biomarkers in laryngeal cancer treated with radiotherapy. *Cancer Genomics - Proteomics, 7*(1), 1-7.

Hombach, S., & Kretz, M. (Ed.). (2016). *Non-coding RNAs: Classification, biology and functioning. in: Slaby O., Calin G. (eds) non-coding RNAs in colorectal cancer. advances in experimental medicine and biology*. Switzerland: Springer, Cham. doi:10.1007/978-3-319-42059-2_1

Hon, C-C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J. L.,Gough, J., Denisenko, E.,… & Forrest, A. R. R. (2017). An atlas of human long non-coding RNAs with accurate 5' ends *Nature, 543*(7644), 199-204. doi:10.1038/nature21374

Hou, Z., Zhao, W., Zhou, J., Shen, L., Zhan, P., Xu, C., Chang, C., Bi, H., Zou, J., Yao, X., Huang, R., Yu, L., & Yan, J. (2014). A long noncoding RNA Sox2ot regulates lung cancer cell proliferation and is a prognostic indicator of poor survival. *The International Journal of Biochemistry & Cell Biology, 53*(Supplement C), 380-388. doi:10.1016/j.biocel.2014.06.004

Howington, J. A., Blum, M. G., Chang, A. C., Balekian, A. A., & Murthy, S. C. (2013). Treatment of stage I and II non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest, 143*(5, Supplement), e278S-e313S. doi:https://doi.org/10.1378/chest.12-2359

Hu, Z., Chen, J., Tian, T., Zhou, X., Gu, H., Xu, L., ... Shen, H. (2008). Genetic variants of miRNA sequences and non-small cell lung cancer survival. *Clinical Investigation, 118*(7), 2600-2608. doi:10.1172/JCI34934

Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nature Medicine, 21*, 1253-1261. doi:10.1038/nm.3981

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M. J., Kenzelmann-Broz, D., … Rinn, J. L. (2010). A large intergenic non-coding RNA induced by p53 mediates global gene repression in the p53 response. *Cell, 142*(3), 409-419. doi:10.1016/j.cell.2010.06.040

Hulkower, K. I., & Herber, R. L. (2011). Cell migration and invasion assays as tools for drug discovery. *Pharmaceutics, 3*(1), 107-124. doi:10.3390/pharmaceutics3010107

Hung Rj, Mckay Jd, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, . . . & Brennan P. (2008). A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature, 452*, 633-637. doi:10.1038/nature06885

Igarashi, S., Suzuki, H., Niinuma, T., Shimizu, H., Nojima, M., Iwaki, H., Nobuoka, T., ...Shinomura, Y. (2010). A novel correlation between *LINE-1* Hypomethylation and the malignancy of gastrointestinal stromal tumors *Clinical Cancer Research, 16*(21), 5114-5123. doi:10.1158/1078-0432.CCR-10-0581

Iwasaki, Y. W., Siomi, M. C., & Siomi, H. (2015). PIWI-interacting RNA: Its biogenesis and functions. *Annual Review of Biochemistry, 84*, 405-433. doi:10.1146/annurev-biochem-060614-034258

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals *Nature Genetics, 33*, 245. doi:10.1038/ng1089

Ji, P., Diederichs, S., Wang, W., Boing, S., Metzger, R., Schneider, P. M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., Thomas, M., Berdel, W. E., Serve, H., & Muller-Tidow, C. (2003). MALAT-1, a novel noncoding RNA, and thymosin *β*4 predict metastasis and survival in early-stage non-small cell lung cancer *Oncogene, 22*, 8031-8041. doi:10.1038/sj.onc.1206928

Ji, Q., Zhang, L., Liu, X., Zhou, L., Wang, W., Han, Z., … Li, Q. (2014). Long non-coding RNA MALAT1 promotes tumour growth and metastasis in colorectal cancer through binding to

SFPQ and releasing oncogene PTBP2 from SFPQ/PTBP2 complex. *British Journal of Cancer, 111*(4), 736-748. doi:10.1038/bjc.2014.383

Jin, B., & Robertson, K. D. (2013). DNA methyltransferases (DNMTs), DNA damage repair, and cancer. *Advances in Experimental Medicine and Biology, 754*, 3-29. doi:10.1007/978-1-4419-9967-2_1

Jin, B., Li, Y., & Robertson, K. D. (2011). DNA methylation: Superior or subordinate in the epigenetic hierarchy? *Genes & Cancer, 2*(6), 607-617. doi:10.1177/1947601910393957

Jingcheng, Y., Jinyuan, L., Tianxiao, L., Ting, C., Shangling, P., Weiqiang, H., & Shikang, L. (2014). Analysis of lncRNA expression profiles in non-small cell lung cancers (NSCLC) and their clinical subtypes. *Lung Cancer, 85*, 110-115. doi:10.1016/j.lungcan.2014.05.011

Jones, P.A., and Takai, D. (2001). The role of DNA methylation in mammalian epigenetics. *Science, 293*(5532), 1068-1070. doi:10.1126/science.1063852

Joo, J. H., Correira, G. P., Li, J. L., Lopez, M. C., Baker, H. V., & Sugrue, S. P. (2013). Transcriptomic analysis of PNN- and ESRP1-regulated alternative pre-mRNA splicing in human corneal epithelial cells. *Investigative Ophthalmology and Visual Science, 54*, 697-707. doi:10.1167/iovs.12-10695

Justus, C. R., Leffler, N., Ruiz-Echevarria, M., & Yang, L. V. (2014). *in vitro* Cell migration and invasion assays. *Journal of Visualized Experiments, 88* doi.org/10.3791/51046

Kalluri, R., & Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *The Journal of Clinical Investigation, 119*(6), 1420-1428. doi:10.1172/JCI39104

Kanwal, M., Ding, X. J., & Cao, Y. (2017). Familial risk of lung cancer. *Oncology Letters, 13*(2), 535-542. doi:10.3892/ol.2016.5518

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., Stadler, Peter F., Hertel, J., ...& Drenkow, J. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription *Science, 316*(5830), 1484-1488. doi:10.1126/science.1138341

Karapetyan, A. R., Buiting, C., Kuiper, R. A., & Coolen, M. W. (2013). Regulatory roles for long ncRNA and mRNA. *Cancers, 5*(2), 462-490. doi:10.3390/cancers5020462

Kashi, K., Henderson, L., Bonetti, A., & Carninci, P. (2016). Discovery and functional analysis of lncRNAs: Methodologies to investigate an uncharacterized transcriptome. *Biochimica Et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 1859*(1), 3-15. doi:https://doi.org/10.1016/j.bbagrm.2015.10.010

Kazanets, A., Shorstova, T., Hilmi, K, & Marques, M., & Witcher, M. (2016). Epigenetic silencing of tumor suppressor genes: Paradigms, puzzles, and potential. *Biochimica Et Biophysica Acta (BBA) - Reviews on Cancer, 1865*(2), 275-288. doi:10.1016/j.bbcan.2016.04.001

Kazazian JR, H. H. (2004). Mobile elements: Drivers of genome evolution. *Science, 303*(5664), 1626-1632. doi:10.1126/science.1089670

Kenfield, S.A., Wei, E.K., Stampfer, M.J., Rosner, B.A., & Colditz, G.A. (2008). Comparison of aspects of smoking among four histologic types of cancer. *Tobacco Control, 17*(3), 198-204. doi:10.1136/tc.2007.022582

Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., … Rinn, J. L. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying

complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America, 106*(28), 11667-11672. doi:10.1073/pnas.0904715106

Kim, D. H. and Rossi, J. J. (2007). Strategies for silencing human disease using RNA interference. *Nature Reviews Genetics, 8*, 173-184. doi:10.1038/nrg2006

Kim, Y., and Kim, V. N. (2012). MicroRNA factory: RISC assembly from precursor MicroRNAs. *Molecular Cell, 46*, 384-386. doi:10.1016/j.molcel.2012.05.012

Kino, T., Hurt, D. E., Ichijo, T., Nader, N., & Chrousos, G. P. (2010). Noncoding RNA Gas5 is a growth arrest and starvation-associated repressor of the glucocorticoid receptor. *Science Signaling, 3*(107) doi:10.1126/scisignal.2000568

Kiss-LÃ¡szlÃ³, Z., Henry, Y., Bachellerie, J-P., Caizergues-Ferrer, M., & Ki Kiss-LÃ¡szlÃ³, Z. T. (1996). Site-specific ribose methylation of preribosomal RNA: A novel function for small nucleolar RNAs *Cell, 85*(7), 1077-1088. doi:10.1016/S0092-8674(00)81308-2

Kononen, J., Bubendorf, L., Kallioniemi, A., Bärlund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M. J., Sauter, G., & Kallioniemi, O. P. (1998). Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature Medicine, 4*(7), 844-847.

Kulis, M., & Esteller, M. (2010). DNA methylation and cancer. *Advances in Genetics, 70*(Supplement C), 27-56. doi:doi.org/10.1016/B978-0-12-380866-0.60002-2

Kung, J. T. Y., Colognori, D., & Lee, J. T. (2013). Long noncoding RNAs: Past, present, and future. *Genetics, 193*(3), 651-669. doi:10.1534/genetics.112.146704

Lackey, A., & Donington, J. (2013). Surgical management of lung cancer. *Seminars in Interventional Radiology, 30*(2), 133-140. doi:10.1055/s-0033-1342954

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J.,.. Morgan, J. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), 860-921. doi:http://hdl.handle.net/2027.42/62798

Langevin, S. M., Kratzke, R. A., & Kelsey, K. T. (2015). Epigenetics of lung cancer. *Translational Research: The Journal of Laboratory and Clinical Medicine, 165*(1), 74-90. doi:10.1016/j.trsl.2014.03.001

Lawrie, C. H., Gal, S., Dunlop, H. M., Pushkaran, B., Liggins, A. P., Pulford, K., Banham, A. H., Pezzella, F., Boultwood, J., Wainscoat, J. S., Hatton, C. S. R. and Harris, A. L. (2008). Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *British Journal of Haematology, 141*, 672-675. doi:10.1111/j.1365-2141.2008.07077.x

Lee, J. T. (2012). Epigenetic regulation by long noncoding RNAs. *Science, 338*(6113), 1435-1439. doi:10.1126/science.1231776

Li, D. Y., Chen, W. J., Luo, L., Wang, Y. K., Shang, J., Zhang, Y., Chen, G., Li, S. K. (2017c). Prospective lncRNA-miRNA-mRNA regulatory network of long non-coding RNA LINC00968 in non-small cell lung cancer A549 cells: A miRNA microarray and bioinformatics investigation. *International Journal of Molecular Medicine, 40*(6), 1895-1906. doi:10.3892/ijmm.2017.3187

Li, J., Chen, Y., Chen, Z., He, A., Xie, H., Zhang, Q., Cai, Z., Liu, Y., Huang, W. (2017b). SPRY4-IT1: A novel oncogenic long non-coding RNA in human cancers. *Tumor Biology, 39*(6) doi:10.1177/1010428317711406

Li, N., Gao, W-J., & Liu, N-S. (2017a). LncRNA BCAR4 promotes proliferation, invasion and metastasis of non-small cell lung cancer cells by affecting epithelial-mesenchymal transition. *European Review of Medical & Pharmacological Sciences, 29*(9), 2075-2086.

Li, S., Li, B., Zheng, Y., Li,M., Shi, L., & Pu, X. (2017d). Exploring functions of long noncoding RNAs across multiple cancers through co-expression network. *Scientific Reports, 7*(1), 754. doi:10.1038/s41508-017-00856-8

Li, W., Liu, X., Sun, M., Zhang, M., Wu, W., Xie, W., & Hou, Y-y. (2013). Long non-coding RNA MEG3 inhibits NSCLC cells proliferation and induces apoptosis by affecting p53 expression. *BioMedical Central Cancer, 13*, 461. doi:10.1186/1471-2407-13-461

Li, Y., Sheu, C. C., Ye, Y., de Andrade, M., Wang, L., Chang, S. C., … Yang, P. (2010). Genetic variants and risk of lung cancer in never smokers: A genome-wide association study. *The Lancet Oncology, 11*(4), 321-330. doi:10.1016/S1470-2045(10)70042-5

Li, Z. X., Zhu, Q. N., Zhang, H. B., Hu, Y., Wang, G., & Zhu, Y. S. (2018). MALAT1: A potential biomarker in cancer. *Cancer Management and Research, 10*, 6757-6768. doi:10.2147/CMAR.S169406

Liloglou, T., Bediaga, N. G., Brown, B. R. B., Field, J. K., & Davies, M. P. A. (2014). Epigenetic biomarkers in lung cancer. *Cancer Letters, 342*(2), 200-212. doi:10.1016/j.canlet.2012.04.018

Lin, Q., Guan, W., Ren, W., Zhang, L., Zhang, J. and Xu, G. (2018). MALAT1 affects ovarian cancer cell behavior and patient survival. *Oncology Reports, 39*, 2644-2652. doi:10.3892/or.2018.6384

Lisanti, S., Omar, W. A. W., Tomaszewski, B., De Prins, S., Jacobs, G., Koppen, G., … Langie, S. A. S. (2013). Comparison of methods for quantification of global DNA methylation in human cells and tissues. *Plos One, 8*(11), e79044. doi:10.1371/journal.pone.0079044

Liu, J., Xing, Y., Xu, L., Chen, W., Cao, W., & Zhang, C. ((2017). Decreased expression of pseudogene *PTENP1* promotes malignant behaviours and is associated with the poor survival of patients with HNSCC. *Scientific Reports, 7*(41179) doi:10.1038/srep41179

Liu, Y., & Lu, X. (2012). Non-coding RNAs in DNA damage response. *American Journal of Cancer Research, 2*(6), 658-675.

Liu, Y., Xia, R., Lu, K., Xie, M., Yang, F., Sun, M., … Ji, G. (2017). LincRNA*FEZF1-AS1* represses *p21 expression to promote* gastric cancer proliferation *through LSD1-mediated H3K4me2 demethylation*. *Molecular Cancer, 16*, 39. doi:10.1186/s12943-017-0588-9

Lorenzen, J. M., & Thum, T. (2016). Long noncoding RNAs in kidney and cardiovascular diseases *Nature Reviews Nephrology, 12*, 360-373. doi:10.1038/nrneph.2016.51

Lu, K-h., Li, W., Liu, X-h., Sun, M., Zhang, M-l., Wu, W-q., Xie, W-p., Hou, Y-y. (2013). Long non-coding RNA MEG3 inhibits NSCLC cells proliferation and induces apoptosis by affecting p53 expression 461. *BioMedical Central Cancer, 13*(1), 461. doi:10.1186/1471-2407-13-461

Lu, W., Zhang, H., Niu, Y., Wu, Y., Sun, W., Li, H., ... Wu, Y. (2017). Long non-coding RNA linc00673 regulated non-small cell lung cancer proliferation, migration, invasion and epithelial mesenchymal transition by sponging miR-150-5p. *Molecular Cancer, 16*(118) doi:10.1186/s12943-017-0685-9

Luo, J., Tang, L., Zhang, J., Ni, J., Zhang, H-p., Zhang, L., Xu, J-f & Zheng, D. (2014). Long non-coding RNA *CARLo-5* is a negative prognostic factor and exhibits tumor pro-oncogenic

activity in non-small cell lung cancer. *Tumor Biology, 35*(11), 11541-11549. doi:10.1007/s13277-014-24427

Luteijn, M. J., & Ketting, R. (2013). PIWI-interacting RNAs: From generation to transgenerational epigenetics *Nature Reviews Genetics, 14*, 523. doi:10.1038/nrg3495

Ma, C., Wu, G., Zhu, Q., Liu, H., Yao, Y., Yuan, D., … Song, Y. (2017). Long intergenic noncoding RNA 00673 promotes non-small-cell lung cancer metastasis by binding with EZH2 and causing epigenetic silencing of HOXA5. *Oncotarget, 8*(20), 32696-32705. doi:10.18632/oncotarget.16158

Ma, L., Bajic, V. B., & Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA Biology, 10*(6), 924-933. doi:10.4161/rna.24604

Marks, L. S., Fradet, Y., Deras, I. L., Blase, A., Mathis, J., Aubin, S. M. J., Cancio, A. T., Desaulniers, M., Ellis, W. J., Rittenhouse, H., & Groskopf, J. (2007). PCA3 molecular urine assay for prostate cancer in men undergoing repeat biopsy. *Urology, 69*(3), 532-535. doi:10.1016/j.urology.2006.12.014

Martin, A., & Gaya, A. (2010). Stereotactic body radiotherapy: A review. *Clinical Oncology, 22*(3), 157-172. doi:https://doi.org/10.1016/j.clon.2009.12.003

Mavrou, A., Brakspear, K., Hamdollah-Zadeh, M., Damodaran, G., Babaei-Jadidi, R., Oxley, J., … Oltean, S. (2015). Serine arginine protein kinase-1 (srpk1) inhibition as a potential novel targeted therapeutic strategy in prostate cancer. *Oncogene, 34*(33), 4311-4319. doi:10.1038/onc.2014.360

Mazzone, P. J., Sears, C. R., Arenberg, D. A., Gaga, M., Gould, M. K., Massion, P. P., …. (2017). Evaluating molecular biomarkers for the early detection of lung cancer: When is a biomarker ready for clinical use? an official American thoracic society policy

statement. *American Journal of Respiratory and Critical Care Medicine, 196*(7), e15-e29. doi:10.1164/rccm.201708-1678ST

McManus, M. T., & Sharp, P. A. (2002). Gene silencing in mammals by small interfering RNAs. *Nature Reviews Genetics, 3*, 737-747. doi:10.1038/nrg908

Mercer, T. R., Dinger, M. E. & Mattick, J. S. (2009). Long non-coding RNAs: Insights into functions. *Nature Reviews Genetics, 10*, 155-159. doi:10.1038/nrg2521

Meseure, D. and Alsiba, K. D. (2018). Part 1: The PIWI-piRNA pathway is an immune-like surveillance process that controls genome integrity by silencing transposable elements. *https://www.intechopen.com/online-first/part-1-the-piwi-pirna-pathway-is-an-immune-like-surveillance-process-that-controls-genome-integrity-* doi:10.5772/intechopen.79974

Meseure, D., Vacher, S., Lallemand, F., Alsibai, K. D., Hatem, R., Chemlali, W., … Bieche, I. (2016). Prognostic value of a newly identified *MALAT1*alternatively spliced transcript in breast cancer. *British Journal of Cancer, 114*(12), 1395-1404. doi:10.1038/bjc.2016.123

Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. *Nature, 457*(7228), 413-420. doi:10.1038/nature07756

Moison, C., Guieysse-Peugeot, A. L., & Arimondo, P. B. (2014). DNA methylation in cancer. *Atlas Genetics & Cytogenetics in Oncology & Haematology, 18*(4), 285-292. doi:10.4267/2042/53543

Molina, J.R., Yang, P., Cassivi, S.D., Schild, S.E., & Adjei, A.A. (2008). Non-small cell lung cancer: Epidemiology, risk factors, treatment and survivorship. *Mayo Clinic Proceedings, Mayo Clinic, 83*(5), 584-594. doi:10.4065/83.5.584

Montes, M., Nielsen, M. M., Maglieri, G., Jacobsen, A., HÃ¸jfeldt, J., Agrawal-Singh, S., Hansen, K., Helin, K., van de Werken, Harmen J. G., Pedersen, J. S., & Lund, Anders H. (2015). The lncRNA MIR31HG regulates p16INK4A expression to modulate senescence. *Nature Communications, 6*, 6967. doi:10.1038/ncomms7967

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology, 38*(1), 23-28. doi:10.1038/npp.2012.112

Morabito, A., Carillio, G., Daniele, G., Piccirillo, M.C., Montanino, A., Costanzo, R., Sandomenico, C., Giordano, P., Normanno, N., Perrone, F., Rocco, G., Di Maio, M. (2014). Treatment of small cell lung cancer. *Critical Reviews in Oncology Heamatology, 91*(3), 257-270. doi:10.1016/j.critrevonc.2014.03.003

Moran, C. A. (2016). Large cell carcinoma. Retrieved from www.cancernetwork.com/articles/slide-show-non-small-cell-lung-cancer-pathology

Morry, J., Ngamcherdtrakul, W., Yantasee, W. (2017). Oxidative stress in cancer and fibrosis: Opportunity for therapeutic intervention with antioxidant compounds, enzymes, and nanoparticles. *Redox Biology, 11*, 240-253. doi:10.1016/j.redox.2016.12.011

Nakagawa, T., Endo, H., Yokoyama, M., Abe, J., Tamai, K., Tanaka, N., Sato, I., Takahashi, S., Kondo,T., Satoh, K. (2013). Large noncoding RNA HOTAIR enhances aggressive biological behavior and is associated with short disease-free survival in human non-small cell lung cancer. *Biochemical and Biophysical Research Communications, 436*(2), 319-324. doi:10.1016/j.bbrc.2013.05.101

Navarro, C. C., Cano, C., Cuadros, M., Herrera-Merchan, A., Molina, M., & Blanco, A. (2016). In OrtuÃ±o, F., & Rojas, I. (Ed.), *A mechanistic study of lncRNA fendrr regulation of FoxF1*

lung cancer tumor supressor. Switzerland: Springer International Publishing. doi:10.1007/978-3-319-31744-1_67

NCBI. (2017). RNA interference. Retrieved from https://www.ncbi.nlm.nih.gov/probe/docs/techrnai/

Nie, F. Q., Sun, M., Yang, J. S., Xie, M., Xu, T. P., Xia, R., Liu, Y. W., Liu, X. H., Zhang, E. B., Lu, K. H., & Shu, Y. Q. (2015). Long noncoding RNA ANRIL promotes Non–Small cell lung cancer cell proliferation and inhibits apoptosis by silencing KLF2 and P21 expression. *Molecular Cancer Therapeutics, 14*(1), 268-277. doi:10.1158/1535-7163

Nikon Microscopy U. (2016). Lung adenocarcinoma. Retrieved from Nikon.magnet.fsu.edu/staticgallery/pathology/adenocarcinomaoflung40x01.html

O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of MicroRNA biogenesis, mechanisms of actions, and circulation. *Frontiers in Endocrinology, 9*(402) doi:10.3389/fendo.2018.00402

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sigiyama, T., Irie, R., ... & Sugano, S. (2004). Complete sequencing and characterization of 21,243 full length cDNAs. *Nature Genetics, 36*, 40-45. doi:10.1038/ng1285

Park, K.-S., Martelotto, L. G., Peifer, M., Sos, M. L., Kamezis, A.N., Mahjoub, M.R, ... Sage, J. (2011). A crucial requirement for hedgehog signalling in small cell lung cancer. *Nature Medicine, 17*(11), 1504-1508. doi:10.1038/nm.2473

Peixing, W., Wenru, S., & Yehong, Z. (2017). Precise long non-coding RNA modulation in visual maintenance and impairment. *Medical Genetics, 54*(7), 450-459. doi:10.1136/jmedgenet-2016-104266

Petersen, S., Wolf, G., BockmA1/4hl, U., Gellert, K., Dietel, M., & Petersen, I. (1998). Allelic loss on chromosome 10q in human lung cancer: Association with tumour progression and metastatic phenotype. *British Journal of Cancer, 77*(2), 270-276.

Pfeifer G. P. (2018). Defining driver DNA methylation changes in human cancer. *International Journal of Molecular Sciences, 19*(4), 1166. doi:10.3390/ijms19041166

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., & Pandolfi, P. P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature, 465*(7301), 1033-1038. doi:10.1038/nature09144

Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell, 136*(4), 629-641. doi:10.1016/j.cell.2009.02.006

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. *Nature Biotechnology, 28*(10), 1057-1068. doi:10.1038/nbt.1685

Qian, J., and Massion, P. P. (2008). Role of chromosome 3q amplification in lung cancer. *Journal of Thoracic Oncology, 3*(3), 212-215. doi:10.1097/JTO.0b013e3181663544

Qiu, M. T., Hu, J. W., Yin, R., & Xu, L. (2013). Long noncoding RNA: An emerging paradigm of cancer research. *Tumor Biology, 34*(2), 613-620. doi:10.1007/s13277-013-0658-6

Quan, M., Chen, J., & Zhang, D. (2015). Exploring the secrets of long noncoding RNAs. *International Journal of Molecular Sciences, 16*(3), 5467-5496. doi:10.3390/ijms16035467

Quinn, J. J., & Chang, H. Y. (2016). Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics, 17*(1), 47-62. doi:10.1038/nrg.2015.10

Qureshi, I. A., Mattick, J. S., & Mehler, M. F. (2010). Long non-coding RNAs in nervous system function and disease. *Brain Research, 1338C*, 20-35. doi:10.1016/j.brainres.2010.03.110

Rand, T. A., Petersen, S., Du, F., & Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell, 123*(4), 621-629. doi:10.1016/j.cell.2005.10.020

Reck, M., Heigener, D. F., Mok, T., Soria, J., & Rabe, K. F. (2013). *Management of non-small-cell lung cancer: Recent developments* doi:10.1016/S0140-6736(13)61502-0

Ricciuti, B., Mecca, C., Crinò, L., Baglivo, S., Cenci, M., & Metro, G. (2014). Non-coding RNAs in lung cancer. *Oncoscience, 1*(11), 674-705. doi:10.18632/oncoscience.98

Richard, B. C. (2017). Non-coding RNA: It's not junk. *Digestive Diseases and Sciences, 62*(5), 1107-1109. doi:10.1007/s10620-017-4506-1

Richards, E. J., Permuth-Wey, J., Li, Y., Chen, Y. A., Coppola, D., Reid, B. M., … Cheng, J. Q. (2015). A functional variant in *HOXA11-AS*, a novel long non-coding RNA, inhibits the oncogenic phenotype of epithelial ovarian cancer. *Oncotarget, 6*(33), 34745-34757. doi:10.18632/oncotarget.5784

Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry, 81* doi:10.1146/annurev-biochem-051410-092902

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., … Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by non-coding RNAs. *Cell, 129*(7), 1311-1323. doi:10.1016/j.cell.2007.05.022

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research, 43*(7), e47. doi:10.1093/nar/gkv007

Robine, N., Lau, N. C., Balla, S., Jin, Z., Okamura, K., Kuramochi-Miyagawa, S., … Lai, E. C. (2009). A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Current Biology, 19*(24), 2066-2076. doi:10.1016/j.cub.2009.11.064

Rodriguez-Canales, J., Parra-Cuentas, E., & Wistuba, I.I. (2016). Diagnosis and molecular classification of lung cancer. In K. L. Reckamp (Ed.), *Lung cancer treatment and research* (pp. 25-46). Switzerland: Springer International Publishing. doi:10.1007/978-3-319-40389-2_2

Ross, Robert J., Weiner, M. M., & Lin, H. (2014). PIWI proteins and PIWI-interacting RNAs in the soma *Nature, 505*, 353. doi:10.1038/nature12987

Rossi, A., & Di Maio, M. (2016). Platinum-based chemotherapy in advanced non-small cell lung cancer: Optimal number of treatment cycles. *Expert Review of Anticancer Therapy, 16*(6), 653-660. doi:10.1586/14737140.2016.1170596

Rouillard, A. D., Gundersen, G. W., Fernandez, N. Z., Wang, Z., Monteiro, C. D., McDermmot, M. G., & Ma'ayan, A. (2016). The harmonizome: A collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database, 2016* doi:10.1093/database/baw100

Sai Lakshmi, S., & Agrawal, S. (2008). piRNABank: A web resource on classified and clustered piwi-interacting RNAs. *nucleic acids research*. *Nucleic Acids Research, 36*(Database Issue), D173-D177. doi:10.1093/nar/gkm696

Saito, K., Kawakami, K., Matsumoto, I., Oda, M., Watanabe, G. and Minamoto, T. (2010). Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage

IA non–small cell lung cancer. *Clinical Cancer Research, 16*(8), 2418-2426. doi:10.1158/1078-0432.CCR-09-2819

Sanner, T., & Grimsrud, T. K. (2015). Nicotine: Carcinogenicity and effects on response to cancer treatment - A review. *Frontiers in Oncology, 5*, 196-10.3389/fonc.2015.00196.

Sardenberg, R.A.S., Pinto, C., Bueno, C.A., & Younes, R.N. (2013). Non-small cell lung cancer stage IV long-term survival with isolated spleen metastasis. *The Annals of Thoracic Surgery, 95*(4), 1432-1434. doi:10.1016/j.athoracsur.2012.08.086

Schein, A., Zucchelli, S., Kauppinen, S., Gustincich, S., & Carninci, P. (2016). Identification of antisense long noncoding RNAs that function as SINEUPs in human cells. *Scientific Reports, 6*, 33605. doi:10.1038/srep33605

Schmidt, L. H., Spieker, T., Koschmieder, S., Humberg, J., Jungen, D., Bulk, E., Hascher, A., Wittmer, D., Marra, A., Hillejan, L., Wiebe, K., Berdel, W. E., Wiewrodt, R., & Muller-Tidow, C. (2011). The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. *Journal of Thoracic Oncology, 6*(12), 1984-1992. doi:10.1097/JTO.0b013e3182307eac

Schoumacher, M., Hurov, K. E., Lehar, J., Yan-Neale, Y.,... & Shao, W. (2014). InhibitingTankyrase sensitizes KRAS-mutant cancer cells to MEK inhibitors via FGFR2 feedback signalling. *Cancer Research, 74*(12), 3294-3305. doi:10.1158/0008-5472.CAN-14-0138-T

Seidel, C., Florean, C., Schnekenburger, M., Dicato, M., Diederich, M. (2012). Chromatin-modifying agents in anti-cancer therapy. *Biochimie, 94*(11), 2264-2279. doi:10.1016/j.biochi.2012.05.012

Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., …& Sharp, P. A. (2008). (). Divergent transcription from active promoters. 1851. *Science, 322*(5909), 1849. doi:10.1126/science.1162253

Serghiou, S., Kyriakopoulou, A. and Ioannidis, J. P. A. (2016). Long noncoding RNAs as novel predictors of survival in human cancer: A systematic review and meta-analysis. *Molecular Cancer, 15*(50) doi:10.1186/s12943-016-0535-1

Shaw, R. J., Liloglou, T., Rogers, S. N., Brown, J. S., Vaughan, E.D., Lowe, D., Field, J.K. and Risk, J. M. (2006). Promoter methylation of *P16*, *RARß*, E-cadherin, cyclin A1 and cytoglobin in oral cancer: Quantitative evaluation using pyrosequencing. *British Journal of Cancer, 94*, 561-568. doi:10.1038/sj.bjc.6602972

Shi, X., Ma, C., Zhu, Q., Yuan, D., Sun, M., Gu, X., ... Song, Y. (2016). Upregulation of long intergenic noncoding RNA 00673 promotes tumour proliferation via LSD1 interaction and repression of NCALD in non-small cell lung cancer. *Oncotarget, 7*(18), 25558-25575. doi:10.18632/oncotarget.8338

Shi, X., Sun, M., Liu, H., Yao, Y., Kong, R., Chen, F., & Song, Y. (2015). A Critical role for the long non-coding RNA GAS5 in proliferation and apoptosis in non-small-cell lung cancer. *Molecular Carcinogenesis, Supplement 1*, E1-E12. doi:10.1002/mc.22120

Shi, Y., Lu, J., Zhou, J., Tan, X., He, Y., Ding, J., Tian, Y., Wang, L., & Wang, K. (2014). Long non-coding RNA Loc554202 regulates proliferation and migration in breast cancer cells. *Biochemica Biophysica Research Communication, 446*(2), 448-453. doi:10.1016/j.bbrc.2014.02.144

Silvestri, G.A., Gonzalez, A.V., Jantz, M.A., Margolis, M.L., Gould, M. K., Tanoue, L.T.,, Harris, L.J and Detterbeck, F.C. (2013). Methods for staging non-small cell lung cancer: Diagnosis

and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest, 143*(5), e211S-e250S. doi:10.1378/chest.12-2355

Siomi, M. C., Sato, K., Pezic, D., & Aravin, A. A. (2011). PIWI-interacting small RNAs: The vanguard of genome defence. *Nature Reviews Molecular Cell Biology, 12*(4), 246-258. doi:10.1038/nrm3089.

Spyratos, D., Zarogoulidis, P., Porpodis, K., Tsakiridis, K., Machairiotis, n., Katsikogiannis, N., ... Zarogoulidis, K. (2013). Occupational exposure and lung cancer. *Journal of Thoracic Disease, 5*(Suppl 4), S440-S445. doi:10.3978/j.issn.2072-1439.2013.07.09

Stinchcombe, T. E. (2014). Unmet needs in squamous cell carcinoma of the lung: Potential role of immunotherapy. *Medical Oncology, 31*(5), 960. doi:10.1007/s12032-014-0960-1

Straif, k., Benbrahim-Talaa, L., Baan, R., Grosse, Y., Secretan, B., ... El Ghissassi, F., & WHO International Agency on Cancer Monograph Working Group. (2009). A review of human carcinogens-part C: Metals, arsenic, dusts and fibres. *Lancet Oncology, 10*(5), 453-454.

Strimbu, K., & Tavel, J. A. (2010). What are biomarkers? *Current Opinion in HIV and AIDS, 5*(6), 463-466. doi:10.1097/COH.0b013e32833ed177

Sturm, A., Perczel, A., Ivics, Z. and Vellai, T. (2017). The Piwi-piRNA pathway: Road to immortality. *Aging Cell, 16*(5), 906-911. doi:10.1111/acel.12630

Sun, M., Liu, X.-H., Wang, K.-M., Nie, F., Kong, R., Yang, J., … Wang, Z.-X. (2014). Downregulation of BRAF activated non-coding RNA is associated with poor prognosis for non-small cell lung cancer and promotes metastasis by affecting epithelial-mesenchymal transition. *Molecular Cancer, 13*, 68. doi:10.1186/1476-4598-13-68

Sun, M., Liu, X-H., Lu, K-H., Nie, F-Q., Xia, R., Kong, R., Yang, J-S., Xu, T-P., Liu, Y-W., Zou, Y-F., Lu, B-B., Yin, R., Zhang, E-B., Xu, L., De, W., & Wang, Z-X. (2014). EZH2-mediated epigenetic suppression of long noncoding RNA SPRY4-IT1 promotes NSCLC cell proliferation and metastasis by affecting the epithelial mesenchymal transition *Cell Death & Disease, 5*, e1298. doi:10.1038/cddis.2014.256

Taft, R. J., Pang, K. C., Mercer, T. R., Dinger, M. and Mattick, J. S. (2010). Non-coding RNAs: Regulators of disease. *Journal of Pathology, 220*, 126-139. doi:10.1002/path.2638

Takahashi, Y., Sawada, G., Kurashige, J., Uchi, R., Matsumura, T., Ueo, H., … Mimori, K. (2014). Amplification of *PVT-1* is involved in poor prognosis via apoptosis inhibition in colorectal cancers. *British Journal of Cancer, 110*(1), 164-171. doi:10.1038/bjc.2013.698

Tano, K., Mizuno, R., Okada, T., Rakwal, R., Shibato, J., Masuo, Y., Ijiri, K., & Akimitsu, N. (2010). MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. *FEBS Letters, 584*(22), 4575-4580. doi:10.1016/j.febslet.2010.10.008

The Cancer Genome Research Network. (2016). Squamous cell carcinoma. Retrieved from cancergenome.nih.gov/newsevents/newsannouncements/LUSCseptemberpressrelease

The Encode Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*, 57-74. doi:10.1038/nature11247

Thorgeirsson, T. E., Geller, F., Sulem, P., Rafner, T., Wiste, A., Magnusson, K. P., . . . Steffansson, K. (2008). A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature, 452*(7187), 638-642. doi:10.1038/nature06846

Timofeeva, M. N., Hung, R. J., Rafnar, T., Christiani, D. C., Field, J. K., Bickeböller, H.,.. (2012). Influence of common genetic variation on lung cancer risk: Meta-analysis of 14 900 cases and 29 485 controls. *Human Molecular Genetics, 21*(22), 4980-4995. doi:10.1093/hmg/dds334

Tinzl, M., Marberger, M., Horvath, S., & Chypre, C. (2004). DD3^{PCA3} RNA analysis in urine – A new perspective for detecting prostate cancer. *European Urology, 46*(2), 182-187. doi:10.1016/j.eururo.2004.06.004

Tost, J. (Ed.). (2009). *DNA methylation: Methods and protocols* (Second ed.). New York: Humana Press. doi:10.1007/978-1-59745-522-0

Travis, W., Brambilla, E., Noguchi, M., Nicholson, A., Geisinger, K.,.. Tsao, M. (2012). Diagnosis of lung adenocarcinoma in resected specimens: Implications of the 2011 international association for the study of lung Cancer/American thoracic Society/European respiratory. *Archives of Pathology & Laboratory Medicine, 137*(5), 685-705. doi:10.5858/arpa.2012-0264-RA

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., … Prasanth, K. V. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular Cell, 39*(6), 925-938. doi:10.1016/j.molcel.2010.08.011

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., … Chang, H. Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science, 329*(5992), 689-693. doi:10.1126/science.1192002

Udomsinprasert, W., Kitkumthorn, N., Mutirangura, A., Chongsrisawat ,V, Poovorawan Y, Honsawek S. (2017). Association between promoter hypomethylation and overexpression of

autotaxin with outcome parameters in biliary atresia. *Plos One, 12*(1), e0169306. doi:10.1371/journal.pone.0169306

Ulitsky, I., & Bartel, D. P. (2013). lincRNAs: Genomics, evolution, and mechanisms. *Cell, 154*(1), 26-46. doi:10.1016/j.cell.2013.06.020

Upchurch, G. M., Haney, S. L., & Opavsky, R. (2016). Aberrant promoter hypomethylation in CLL: Does it matter for disease development? *Frontiers in Oncology, 6*, 182. doi:10.3389/fonc.2016.00182

van Gils, M. P.M.Q., Hessels, D., van Hooij, O., Jannink, S. A., Peelen, W. P., Hanssen, S. L.J., Witjes, J. A., Cornel, E. B., Karthaus, H. F.M., Smits, G. A.H.J., Dijkman, G. A., Mulders, P. F.A., & Schalken, J. A. (2007). The time-resolved fluorescence-based PCA3 test on urinary sediments after digital rectal examination; a Dutch multicenter validation of the diagnostic performance. *Clinical Cancer Research, 13*(3), 939-943. doi:10.1158/1078-0432.CCR-06-2679

Van Schil, P.E., Balduyck, B., De Waele, m., Hendriks, J.M., Hertoghs, M., & Lauwers, P. (2013). Surgical treatment of early stage non-small cell lung cancer. *EJC Supplements, 11*(2), 110-122. doi:10.1016/j.ejcsup.2013.07.021

Vance, K. W., & Ponting, C. P. (2014). Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends in Genetics, 30*(8), 348-355. doi:10.1016/j.tig.2014.06.001

Volinia, S., Calin, G. A., Liu, C.-G., Ambs, S., Cimmino, A., Petrocca, F., … Croce, C. M. (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. *Proceedings of the National Academy of Sciences of the United States of America, 103*(7), 2257-2261. doi:10.1073/pnas.0510565103

Waddington, C. H. (2012). The epigenotype 1942. *International Journal of Epidemiology, 41*(1), 10-13. doi:10.1093/ije/dyr184

Wan, L., Sun, M., Liu, G-J., Wei, C-C., Zhang, E-B., Kong, R., Xu, T-P., Huang, M-D., Wang, Z-X. (2016). Long noncoding RNA PVT1 promotes Non–Small cell lung cancer cell proliferation through epigenetically regulating LATS2 expression. *Molecular Cancer Therapeutics, 15*(5), 1082-1094. doi:10.1158/1535-7163.MCT-15-0707

Wan, L., Zhang, L., Fan, K., Cheng, Z. X., Sun, Q. C., & Wang, J. J. (2016). Knockdown of long noncoding RNA PCAT6 inhibits proliferation and invasion in lung cancer cells. *Oncology Research, 24*(3), 161-170. doi:10.3727/096504016X14618564639178

Wang, B., Jiang, H., Wang, L., Chen, X., Wu, K., Zhang, S., … Xia, B. (2017a). Increased MIR31HG lncRNA expression increases gefitinib resistance in non-small cell lung cancer cell lines through the EGFR/PI3K/AKT signaling pathway. *Oncology Letters, 13*(5), 3494-3500. doi:10.3892/ol.2017.5878

Wang, F., Yuan, J.-H., Wang, S.-B., Yang, F., Yuan, S.-X., Ye, C., Yang, N., Zhou, W.-P., Li, W.-L., Li, W. and Sun, S.-H. (2014). Oncofetal long noncoding RNA PVT1 promotes proliferation and stem cell-like property of hepatocellular carcinoma cells by stabilizing NOP2. *Hepatology, 60*(4), 1278-1290. doi:10.1002/hep.27239

Wang, K. C., & Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Molecular Cell, 43*(6), 904-914. doi:10.1016/j.molcel.2011.08.018

Wang, P., Zhou, Z., Hu, A., de Albuquerque, C. P., Zhou, Y., Hong, L., … Fu, X.-D. (2014). Both decreased and increased SRPK1 levels promote cancer by interfering with PHLPP-mediated dephosphorylation of akt. *Molecular Cell, 54*(3), 378-391. doi:10.1016/j.molcel.2014.03.007

Wang, Q., Zhang, J., Liu, Y., Zhang, W., Zhou, J., Duan, R., Pu, P., Kang, C., & Han, L. (2016). A novel cell cycle-associated lncRNA, HOXA11-AS, is transcribed from the 5-prime end of the

HOXA transcript and is a biomarker of progression in glioma. *Cancer Letters, 373*, 251-259. doi:10.1016/j.canlet.2016.01.039

Wang, Y., Zhou, J., Xu, Y. J., & Hu, H. B. (2018). Long non-coding RNA LINC00968 acts as oncogene in NSCLC by activating the wnt signalling pathway. *Journal of Cell Physiology,* doi:10.1002/jcp.26186

Webb, B. A., Chimenti, M., Jacobson, M. P., and Barber, D. L. (2011). Dysregulated pH: A perfect storm for cancer progression. *Nature Reviews Cancer, 11*, 671-677. doi:10.1038/nrc3110

Wen-Jie, C., Rui-Xue, T., Rong-Quan, H., Dong-Yao, L., Liang, L., Jiang-Hui, Z., Xiao-Hua, H., Jie, M., Shi-Kang, L., & Gang, C. (2017). Clinical roles of the aberrantly expressed lncRNAs in lung squamous cell carcinoma: A study based on RNA-sequencing and microarray data mining. *Oncotarget, 8*, 61282-61304. doi:10.18632/oncotarget.18058

WHO. (2016). Tobacco. Retrieved from http://www.who.int/mediacentre/factsheets/fs339/en/

Wong, H. K.., Gibson, H., Hake, T., Geyer, S., Frederickson, J., Marcucci, G., Caligiuri, M. A.., Porcu, P., Mishra, A. (2015). Promoter-specific hypomethylation is associated with overexpression of PLS3, GATA6, and TWIST1 in the sezary syndrome. *Journal of Investigative Dermatology, 135*(8), 2084-2092. doi:10.1038/jid.2015.116

Wu, L., Shen, Y., Peng, X., Zhang, S., Wang, M., Xu, G., … & Lu, C. (2016). Aberrant promoter methylation of cancer-related genes in human breast cancer. *Oncology Letters, 12*(6), 5145-5155. doi:10.3892/ol.2016.5351

Wu, X., Pu, X., & Lin, J. (2014a). Lung cancer susceptibility and risk assessment models. In J. A. Roth, & Waun, K.H. and Ritsuko, U.K (Eds.), *Lung cancer* (4th ed., pp. 25-47). United Kingdom: Wiley-Blackwell.

Wu, Y., Liu, H., Shi, X., Yao, Y., Yang, W., & Song, Y. (2015). The long non-coding RNA HNF1A-AS1 regulates proliferation and metastasis in lung adenocarcinoma. *Oncotarget, 6*(11), 9160-9172. doi:10.18632/oncotarget.3247

Xie, M., Nie, F., Sun, M., Xia, R., Liu, Y., Zhou, P., … Liu, X. (2015). Decreased long noncoding RNA SPRY4-IT1 contributing to gastric cancer cell metastasis partly via affecting epithelial–mesenchymal transition. *Journal of Translational Medicine, 13*, 250. doi:10.1186/s12967-015-0595-9

Xing, Z., Lin, A., Li, C., Liang, K., Wang, S., Liu, Y., … & Yang, L. (2014). LncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell, 159*(5), 1110-1125. doi:10.1016/j.cell.2014.10.013

Xu, G., Chen, J., Pan, Q., Huang, K., Pan, J., Zhang, W., ... & Wang, Y. (2014). Long noncoding RNA expression profiles of lung adenocarcinoma ascertained by microarray analysis. *Plos One, 9*(8), e104044. doi:10.1371/journal.pone.0104044

Xu, T., Huang, M., Xia, R., Liu, X., Sun, M., Yin, L., … Shu, Y. (2014). (2014). Decreased expression of the long non-coding RNA *FENDRR* is associated with poor prognosis in gastric cancer and *FENDRR* regulates gastric cancer cell metastasis by affecting fibronectin1 expression. *Hematology & Oncology, 7*(63) doi:10.1186/s13045-014-0063-7

Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., Stephens, R. M., Okamoto, A., Yokota, J., Tanaka, T., Calin, G. A., Liu, C-G., Croce, C. M., Harris, C. C. (2006).

Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell, 9*(3), 189-198. doi:https://doi.org/10.1016/j.ccr.2006.01.025

Yang R., Li P., Zhang G., Lu C., Wang H., Zhao, G. (2017a). Long non-coding RNA XLOC_008466 functions as an oncogene in human non-small cell lung cancer by targeting miR-874. *Cell Physiology and Biochemistry, 42*, 126-136. doi:10.1159/000477121

Yang, A. S., Estécio, M. R. H., Doshi, K., Kondo, Y., Tajara, E. H., & Issa, J.-P. J. (2004). A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Research, 32*(3), e38. doi:10.1093/nar/gnh032

Yang, I. A., Holloway, J. W., & Fong, K. M. (2013). Genetic susceptibility to lung cancer and co-morbidities. *Journal of Thoracic Disease, 5*(5), S454-S462. doi:10.3978/j.

Yang, J., Lin, J., Liu, T., Chen, T., Pan, S., Huang, W., & Li, S. (2014). Analysis of lncRNA expression profiles in non-small cell lung cancers (NSCLC) and their clinical subtypes. *Lung Cancer, 85*, 110-115. doi:10.1016/j.lungcan.2014.05.011

Yang, L., Peng, X., Jin, H., Liu, J. (2019). Long non-coding RNA PVT1 promotes autophagy as ceRNA to target ATG3 by sponging microRNA-365 in hepatocellular carcinoma. *Gene, 697*, 94-102. doi:10.1016/j.gene.2019.02.036

Yang, Y.-R., Zang, S.-Z., Zhong, C.-L., Li, Y.-X., Zhao, S.-S., & Feng, X.-J. (2014). Increased expression of the lncRNA PVT1 promotes tumorigenesis in non-small cell lung cancer. *International Journal of Clinical and Experimental Pathology, 7*(10), 6929-6935.

Yang, Z., Zhou, L., Wu, L. M., Lai, M-C., Xie, H-Y., Zhang, F., & Zheng, S-S. (2011). Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. *Annals of Surgical Oncology, 18*, 1243. doi:10.1245/s10434-011-1581-y

Yap, K. L., Li, S., Muñoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., … Zhou, M.-M. (2010). Molecular interplay of the non-coding RNA *ANRIL* and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of *INK4a*. *Molecular Cell, 38*(5), 662-674. doi:10.1016/j.molcel.2010.03.021

Yoshimoto, R., Mayeda, A., Yoshida, M., Nakagawa, S. (2016). MALAT1 long non-coding RNA in cancer. *Biochimica Et Biophysica Acta (BBA), 1859*(1), 192-199. doi:10.1016/j.bbagrm.2015.09.012

You, J. S., & Jones, P. A. (2012). Cancer genetics and epigenetics: Two sides of the same coin? *Cancer Cell, 22*(1), 9-20. doi:10.1016/j.ccr.2012.06.008

Yu, G., Yao, W., Gumireddy, K., Li, A., Wang, J., Xiao, W., … Xu, H. (2014). Pseudogene PTENP1 functions as a competing endogenous RNA to suppress clear-cell renal cell carcinoma progression. *Molecular Cancer Therapeutics, 13*(12), 3086-3097. doi:10.1158/1535-7163.MCT-14-0245

Zakhari, S. (2013). Alcohol metabolism and epigenetics changes. *Alcohol Research: Current Reviews, 35*(1), 6-16.

Zappa, C., & Mousa, S. A. (2016). Non-small cell lung cancer: Current treatment and future advances. *Translational Lung Cancer Research, 5*(3), 288-300. doi:10.21037/tlcr.2016.06.07

Zhang, D., Ochi, N., Takigawa, N., Tanimoto, Y., Chen, Y., Ichihara, E., Hotta, K., Tabata, M., Tanimoto, M., Kiura, K. (2011). Establishment of pemetrexed-resistant non-small cell lung cancer cell lines. *Cancer Letters, 309*(2), 228-235. doi:10.1016/j.canlet.2011.06.006

Zhang, E., Li, W., Yin, D., De, W., Zhu, L., Sun, S., & Han, L. (2016). c-myc-regulated long non-coding RNA H19 indicates a poor prognosis and affects cell proliferation in non-small-cell lung cancer. *Tumor Biology, 37*, 4007-4015. doi:10.1007/s13277-015-4185-5

Zhang, E-b., Yin, D-d., Sun, M., Kong, R., Liu, X-h., You, L-h., Han, L., Xia, R., Wang, K-m., Yang, J-s., De, W., Shu, Y-q., & Wang, Z-x. (2014). P53-regulated long non-coding RNA TUG1 affects cell proliferation in human non-small cell lung cancer, partly through epigenetically regulating HOXB7 expression. *Cell Death & Disease, 5*, e1243. doi:10.1038/cddis.2014.201

Zhang, L., Li, S., Choi, Y-L., Lee, J., Gong, Z., Liu, X., Pei, Y., Jiang, A., Ye, M., Mao, M., Zhang, X., Kim, J., Chen, R. (2017). Systematic identification of cancer-related long noncoding RNAs and aberrant alternative splicing of quintuple-negative lung adenocarcinoma through RNA-seq. *Lung Cancer, 109*(Supplement C), 21-27. doi:10.1016/j.lungcan.2017.04.009

Zhang, Y., & He, J. (2013a). The development of targeted therapy in small cell lung cancer. *Journal of Thoracic Disease, 5*(4), 538-548. doi:10.3978/j.issn.2072-1439.2013.07.04

Zhang, Y., He, R., Dang, Y., Zhang, X., Wang, X., Huang, S., … Gan, T. (2016). Comprehensive analysis of the long noncoding RNA HOXA11-AS gene interaction regulatory network in NSCLC cells. *Cancer Cell International, 16*, 89. doi:10.1186/s12935-016-0366-6

Zhang, Y., Shields, T., Crenshaw, T., Hao, Y., Moulton, T., & Tycko, B. (1993). Imprinting of human H19: Allele-specific CpG methylation, loss of the active allele in wilms tumor, and potential for somatic allele switching. *American Journal of Human Genetics, 53*(1), 113-124.

Zhao, Y., Sun, H., & Wang, H. (2016). Long noncoding RNAs in DNA methylation: New players stepping into the old game. *Cell & Bioscience, 6*(45) doi:10.1186/s13578-016-0109-3

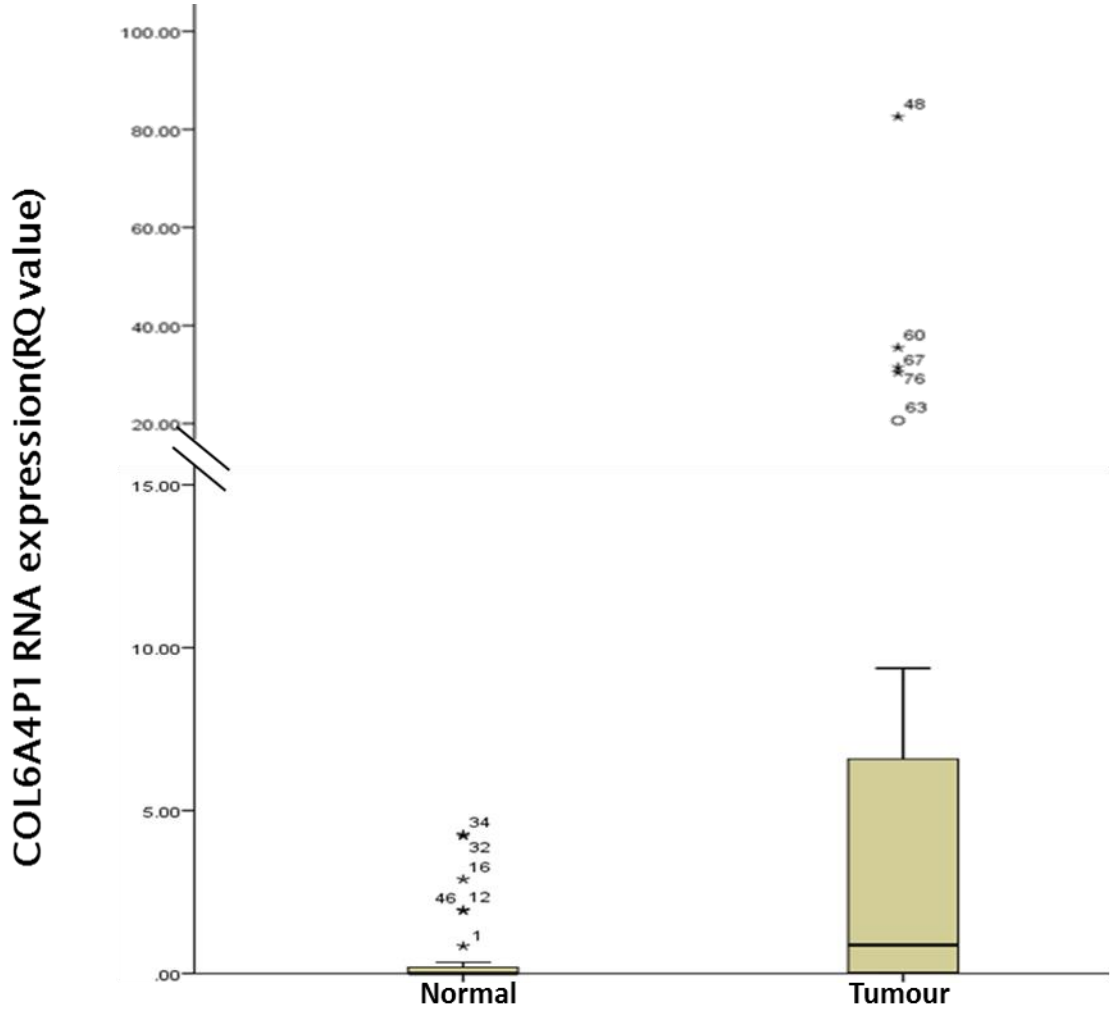
Zheng, X., Schipper, M., Kidwell, K., Lin, J., Reddy, R., Ren, Y., Chang, A., Lv, F., Orringer, M., & Spring Kong, F-M. (2014). Survival outcome after stereotactic body radiation therapy and surgery for stage I non-small cell lung cancer: A meta-analysis. *International Journal of Radiation Oncology*Biology*Physics, 90*(3), 603-611. doi:https://doi.org/10.1016/j.ijrobp.2014.05.055

Zhou, Y., Xu, X., Lv, H., Wen, Q., Li, J., Tan, L., Li, J., & Sheng, X. (2016). The long noncoding RNA MALAT-1 is highly expressed in ovarian cancer and induces cell growth and migration *Plos One, 11*(5), e0155250. doi:10.1371/journal.pone.0155250

Zhu, H., Wang, G., & Qian, J. (2016). Transcription factors as readers and effectors of DNA methylation *Nature Reviews Genetics, 17*, 551-565. doi:10.1038/nrg.2016.83

Zhu, H., Zhang, L., Yan, S., Li, W., Cui, J., Zhu, M., … Wu, N. (2017). LncRNA16 is a potential biomarker for diagnosis of early-stage lung cancer that promotes cell proliferation by regulating the cell cycle. *Oncotarget, 8*(5), 7867-7877. doi:10.18632/oncotarget.13980

APPENDIX I



Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
RQ tumour - RQ normal	Negative Ranks	7[a]	15.43	108.00
	Positive Ranks	33[b]	21.58	712.00
	Ties	6[c]		
	Total	46		

a. RQ tumour < RQ normal

b. RQ tumour > RQ normal

c. RQ tumour = RQ normal

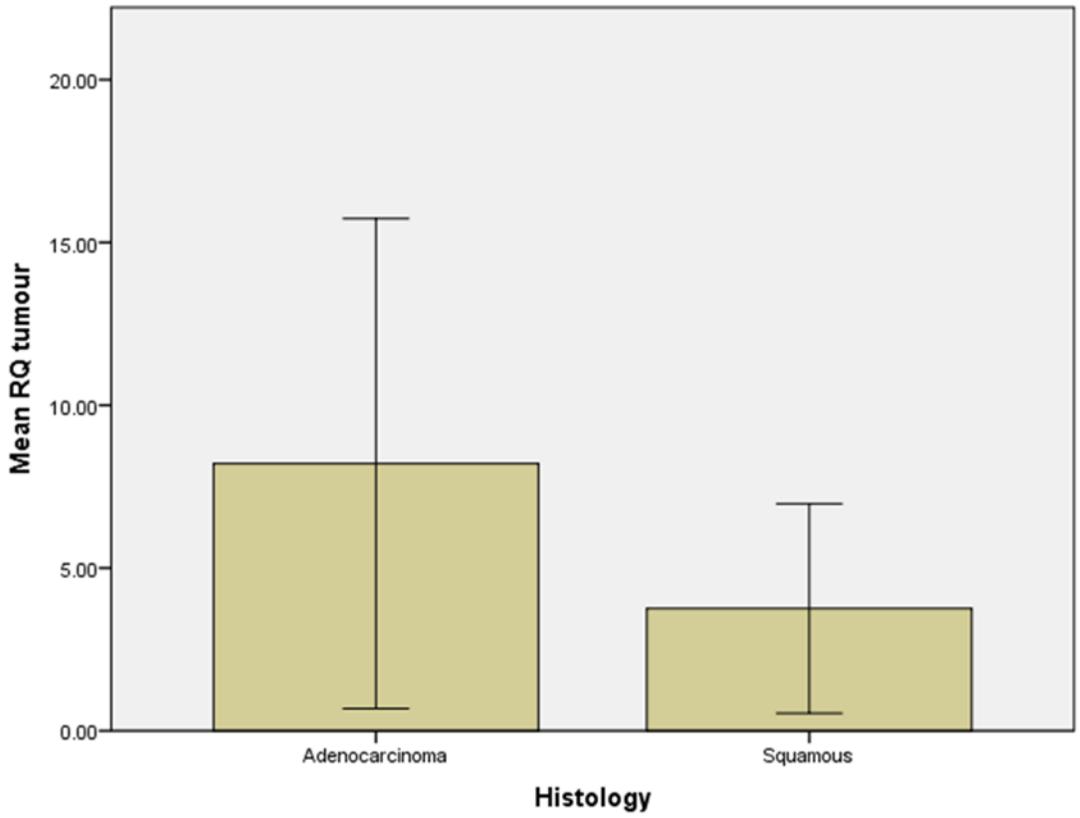
Test Statistics[a]

	RQ tumour - RQ normal
Z	-4.060[b]
Asymp. Sig. (2-tailed)	.000

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

APPENDIX II



Ranks

	Histology	N	Mean Rank	Sum of Ranks
RQ tumour	Adenocarcinoma	25	22.90	572.50
	Squamous	21	24.21	508.50
	Total	46		

Test Statistics[a]

	RQ tumour
Mann-Whitney U	247.500
Wilcoxon W	572.500
Z	-.332
Asymp. Sig. (2-tailed)	.740

a. Grouping Variable: Histology

APPENDIX III

Top Induced lncRNA	p-Value	log fold change
ENST00000366437.3	4.35E-13	-3.4314046
ENST00000452675.1	1.95E-10	-1.7173267
ENST00000555801.1	9.88E-10	-1.3814849
ENST00000423943.1	1.48E-09	-1.9907353
ENST00000521292.2	4.17E-09	0.9183526
NR_033877	2.80E-08	-0.5892284
ENST00000449737.1	1.50E-07	-1.3789755
ENST00000545372.1	2.43E-07	-1.3939459
ENST00000427109.1	4.08E-07	-0.5855934
ENST00000560267.1	5.21E-07	-0.8471578
ENST00000504230.2	1.65E-06	-2.1188106
ENST00000425358.2	2.19E-06	-1.1664112
ENST00000581856.1	4.00E-06	-0.8719493
ENST00000448587.1	5.87E-06	-1.4686119
ENST00000558575.1	7.66E-06	-0.5573358
ENST00000457079.1	1.59E-05	-0.5615499
ENST00000439192.1	1.63E-05	-0.971341
ENST00000297163.3	3.81E-05	-0.5357253
ENST00000419650.1	4.42E-05	-0.6017524
ENST00000437597.1	5.32E-05	0.6371909

APPENDIX IV

Top Repressed lncRNA	p-Value	log fold change
ENST00000439088.1	1.76E-06	-1.664139
ENST00000499425.1	4.94E-06	-3.633362
ENST00000515227.1	5.49E-06	-1.394228
ENST00000409590.1	5.83E-06	-1.776927
ENST00000566787.1	1.49E-05	-5.151783
ENST00000594315.1	1.68E-05	-3.483867
ENST00000563424.2	2.07E-05	-5.256712
ENST00000416105.1	3.03E-05	-1.185832
ENST00000595886.1	3.29E-05	-3,600,458
LOC723809	4.35E-05	-6.660856
ENST00000458364.1	6.78E-05	-2.378006
ENST00000448411.1	7.32E-05	-1.125427
ENST00000415590.1	7.32E-05	-4.032923
ENST00000574016.1	7.85E-05	-2.114123
ENST00000413945.1	9.58E-05	-2.105705
ENST00000490006.2	1.69E-04	-3.911208
ENST00000442069.1	2.15E-04	-1.995611
LINC00312	2.15E-04	-3.991329
ENST00000555688.1	2.37E-04	-1.104316
ENST00000433843.1	2.98E-04	-2.555832
ENST00000556899.1	3.02E-04	-2.00974
C14orf48	4.35E-04	-1.39501
ENST00000452840.1	4.39E-04	-1.689849
ENST00000413221.1	4.40E-04	-1.754503
ENST00000435574.1	4.70E-04	-1.213978
ENST00000432535.1	5.64E-04	-1.322973
ENST00000562038.1	5.83E-04	-1.237423
ENST00000564492.1	6.47E-04	-3.941881
ENST00000425887.1	8.40E-04	-3.433299
ENST00000583490.1	1.09E-03	-2.732779

APPENDIX V

Well	Sample Name	Target Name	Cт	ΔCт	ΔΔCт	RQ				
A1	A549	LINC968	Undetermined				A1	A549	TBP	27.76
B1	A549	LINC968	Undetermined				B1	A549	TBP	27.69
A5	CALU3	LINC968	Undetermined				A5	CALU3	TBP	29.62
B5	CALU3	LINC968	Undetermined				B5	CALU3	TBP	29.58
A2	CALU6	LINC968	Undetermined				A2	CALU6	TBP	27.71
B2	CALU6	LINC968	Undetermined				B2	CALU6	TBP	27.57
C3	cDNApoolNormal	LINC968	29.98	-0.81	=		C3	cDNApoolNormal	TBP	30.79
D3	cDNApoolNormal	LINC968	29.62	-1.05			D3	cDNApoolNormal	TBP	30.68
A3	CORL23	LINC968	Undetermined				A3	CORL23	TBP	27.77
B3	CORL23	LINC968	Undetermined				B3	CORL23	TBP	27.68
A9	H358	LINC968	25.61	-1.38			A9	H358	TBP	26.99
B9	H358	LINC968	25.76	-1.26			B9	H358	TBP	27.02
A10	HBEC3KT	LINC968	27.01	-0.20			A10	HBEC3KT	TBP	27.21
B10	HBEC3KT	LINC968	41.92	14.29			B10	HBEC3KT	TBP	27.63
A12	HBEC3KT 53	LINC968	27.84	-0.93			A12	HBEC3KT 53	TBP	28.78
B12	HBEC3KT 53	LINC968	28.07	-0.59			B12	HBEC3KT 53	TBP	28.66
A11	HBEC3KT R	LINC968	Undetermined				A11	HBEC3KT R	TBP	27.68
B11	HBEC3KT R	LINC968	Undetermined				B11	HBEC3KT R	TBP	27.65
C1	HBEC3KT R53	LINC968	42.80	15.10			C1	HBEC3KT R53	TBP	27.70
D1	HBEC3KT R53	LINC968	42.36	14.73			D1	HBEC3KT R53	TBP	27.62
A7	LUDLU1	LINC968	27.11	-1.78			A7	LUDLU1	TBP	28.89
B7	LUDLU1	LINC968	27.04	-1.76			B7	LUDLU1	TBP	28.80
A8	LUNG14	LINC968	26.21	-2.26			A8	LUNG14	TBP	28.47
B8	LUNG14	LINC968	26.30	-2.13			B8	LUNG14	TBP	28.43
A4	SKLU1	LINC968	43.89	16.08			A4	SKLU1	TBP	27.81
B4	SKLU1	LINC968	48.43	20.62			B4	SKLU1	TBP	27.81
A6	SKMES	LINC968	Undetermined				A6	SKMES	TBP	26.95
B6	SKMES	LINC968	Undetermined				B6	SKMES	TBP	26.87

PRESENTATIONS

Poster Presentation: Deregulation of long non-coding RNAs in lung cancer- National Cancer Research Institute (NCRI) Conference, Liverpool, UK, 2015.

Poster Presentation: Long non-coding RNA deregulation in non-small cell lung cancer- The 2nd International Symposium on Frontiers in Molecular Science, Basel, Switzerland, 2017.

PUBLICATION

Acha-Sagredo, A., Uko, B., Pantazi, P. et al. Long non-coding RNA dysregulation is a frequent event in non-small cell lung carcinoma pathogenesis. Br J Cancer 122, 1050–1058 (2020). https://doi.org/10.1038/s41416-020-0742-9