A Demethylation Screen Identifies Novel Targets of Epigenetic Regulation in Lung Cancer

Karen O'Leary

Ph.D **2014**

Table of Contents

Table of Figures

List of Abbreviations

5'UTR	5' untranslated region
aa	amino acid
AEC	alveolar epithelial cell
AIS	adenocarcinoma *in situ*
AKT	protein kinase B
ALK	anaplastic lymphoma kinase
ALK1	activin receptor-like kinase 1
ALK2	activin receptor-like kinase 2
ALK5	activin receptor-like kinase 5
ANOVA	analysis of variance
ATP	adenosine triphosphate
AZA	5-aza-2’-deoxycytidine
BASC	bronchioalveolar stem cell
BCL-2	B-cell lymphoma 2
BCP	1-bromo-3-chloro-propane
BMP	bone morphogenic protein
BSA	bovine serum albumin
CC10	Clara cell antigen 10
CD	cluster of differentiation
CDK	cyclin-dependent kinase
cDNA	complementary DNA
CELSR	cadherin, EGF LAG seven-pass G-type receptor
CGRP	calcitonin gene related peptide
CI	confidence interval
CKI	cyclin-dependent kinase inhibitor
CMV	cytomegalovirus
CpG	cytosine-phosphate-guanine
CT	computerised tomography
DAB2	Dab, mitogen-responsive phosphoprotein, homolog 2
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
DUSP	dual specificity phosphatase
ECL	electrochemiluminescence
EDTA	ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
EML4	microtubule-associated protein-like 4

EMP3	epithelial membrane protein 3
EMT	epithelial to mesenchymal transition
ENG	Endoglin
ERBB2	human epidermal growth factor receptor 2
ERK	extracellular signal-regulated kinase
ESMO	European Society for Medical Oncology
FBS	fetal bovine serum
FFPE	formalin-fixed, paraffin-embedded
G418	Geneticin
GDF	growth and differentiation factor protein
GEF	guanine nucleotide exchange factor
GIPC	Gα-interacting protein C-terminus-interacting protein
HCL	hydrochloride
HDAC	histone deacetylase
HER2	human epidermal growth factor 2
HHT	hereditary haemorrhagic telangiectasia
HR	hazard ratio
HRP	horseradish peroxidase
IGF-1R	insulin-like growth factor type 1 receptor
IGFBP7	insulin-like growth factor-binding protein 7
IgG	immunoglobulin G
kDa	kilodalton
KLF	kruppel-like zinc finger
KRAB	kruppel associated box
KRAS	Kirsten rat sarcoma viral oncogene homolog
KRT	Keratin
LAD1	ladinin 1
LCC	Large cell carcinoma
LOH	Loss of heterozygosity
LRIG1	leucine-rich repeats and immunoglobulin-like domains 1
mAb	monoclonal antibody
MAPK	mitogen-activated protein kinase
MET	Met proto-oncogene
MGMT	O-6-methylguanine-DNA methyltransferase
MIA	minimally invasive adenocarcinoma
MRI	magnetic resonance imaging
mRNA	messenger RNA
MS-HRM	methylation-sensitive high resolution melting
MSP	methylation-specific PCR
MTHFR	Methylenetetrahydrofolate reductase
mTOR	Mammalian target of rapamycin

MTT	3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide
MVD	microvessel density
Ncam1	neural cell adhesion molecule 1
NFAT	nuclear factor of activated T cells
NSCLC	Non-small cell lung cancer
p16	tumour protein p16
p53	tumour protein p53
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PET	positron emission tomography
PI3K	Phosphoinositide 3-kinase
PNEC	pulmonary neuroendocrine cell
PS	performance status
PVDF	polyvinylidene fluoride
qRT-PCR	reverse transcription and semi-quantitative real-time PCR
Rb	retinoblastoma gene
RIN	RNA integrity number
RNA	ribonucleic acid
RNA-seq	RNA-sequencing
RPLP0	ribosomal protein, large, P0
RPMI	Roswell Park Memorial Institute medium
RSEM	RNA-Seq by Expectation Maximization
RT	room temperature
SCLC	Small cell lung cancer
SD	standard deviation
SDS	sodium dodecyl sulfate
siRNA	small interfering RNA
SMAD	mothers against decapentaplegic (MAD) homolog protein
SOC	super optimal broth with catabolite repression
SPC	surfactant protein C
SSC	Squamous cell carcinoma
STAT3	signal transducer and activator of transcription 3
TBST	Tris-buffered saline with tween
TCGA	The Cancer Genome Atlas
TGFBI	transforming growth factor, beta-induced
TGF-β	transforming growth factor beta
TKI	tyrosine kinase inhibitor
TLDA	Taqman low density array
TNM	tumour, node, metastasis
TSA	trichostatin A

TSG	Tumour suppressor gene
ULA	ultra low attachment
UV	ultraviolet
Vav-1	guanine nucleotide exchange factor 1
VEGF	vascular endothelial growth factor
VIK	Vav 1-interacting Kruppel-like protein
VIM	vimentin
WHO	World Health Organisation
ZNF655	zinc finger protein 655

Declaration

I declare that the research contained in this thesis, unless otherwise formally indicated within the text, is the original work of the author. The thesis has not been previously submitted to these or any other university for a degree, and does not incorporate any material already submitted for a degree.

_____

Karen O'Leary

Date:

Abstract

Lung cancer is the leading cause of death by cancer worldwide. There is an urgent need to expand the range of molecular targets and therapies, as well as biomarkers to diagnose disease, stratify high-risk patients and predict response to therapy. The importance of epigenetic transcriptional silencing as a mechanism of tumour suppressor gene inactivation is well recognized. DNA methylation is the dominant process necessary for this epigenetic inactivation and represents a valuable source of new biomarkers and therapeutic targets. The objective of this project was to seek *de novo*, candidate lung cancer tumour suppressor genes that are subject to transcriptional silencing by means of promoter methylation. Selected candidates would have their potential as prognostic biomarkers in non-small cell lung cancer characterized and any functional consequences of their silencing investigated. A genome-wide, microarray-based screen involving methylation and gene expression analysis in a panel of cell lines, incorporating treatment with the demethylating agent 5-aza-2′-deoxycytidine, was completed. From this, the TGF-β accessory receptor Endoglin and the transcription factor ZNF655 were identified and confirmed as novel genes subject to methylation-dependent silencing in lung cancer. Methylation of both genes was detected in primary lung cancer tissues. There was a strong trend towards shorter survival for patients with methylated Endoglin in two separate cohorts. Functional experiments demonstrated that silencing of Endoglin leads to increased invasion and epithelial to mesenchymal transition (EMT) in 2 NSCLC cell lines *in vitro*. In summary, I have validated the use of a high-throughput microarray screening strategy to identify novel epigenetically silenced tumour suppressor genes. I have identified Endoglin and ZNF655 as common targets of epigenetic silencing in lung cancer and have shown a functional role for Endoglin silencing in lung cancer progression.

Acknowledgements

I would like to thank my supervisor Prof. Peter Schmid for giving me the opportunity to do this PhD and for his direction and guidance. Thank you also to everyone in the Schmid group for their help, advice and friendship over the last 4 years, in particular Alice Shia and Francesca Cavicchioli. Thank you to all my friends and family for their support and encouragement. And thanks to Sarah Dobson for her beautiful illustration on page 17.

1 Introduction

1.1 Introduction to lung cancer

1.1.1 Prevalence and classification

Recent statistics show that worldwide, lung cancer is the most commonly diagnosed cancer in men, and the leading cause of death by cancer, while in women it is the 4th most common cancer and the 2nd highest cause of death by cancer (Jemal *et al.* 2011). In 2012, lung cancer was the leading cause of death by cancer in the US (Siegel *et al.* 2012) and in Europe (Ferlay *et al.* 2013) for both men and women. Global figures suggest that smoking accounts for 80% of lung cancers in men and at least 50% of those in women, and regional variations in incidence largely reflect variations in the smoking epidemic (Jemal *et al.* 2011). Lung cancer incidence in the UK appears to have increased by about 4% per 3-year period during the years 2000-2009, although this may be due in part to the increase in systematic, centralised recording of cancer incidence (Iyen-Omofoman *et al.* 2011).

Lung cancer has been the leading cause of cancer-related death for many decades, but its epidemiology has changed over time. Notably, the incidence of lung cancer in women has increased 4-fold in the last 30 years (Thomas *et al.* 2005). Adenocarcinoma is the most common type of cancer among females and lifetime never-smokers. Worldwide incidence of adenocarcinoma has increased, replacing squamous cell carcinoma as the most common subtype (Nakamura and Saji 2013).

5-year survival rates for lung cancer in the UK were 10 and 12% for men and women, respectively, during the period 2000-2009 (Iyen-Omofoman *et al.* 2011). According to the American Cancer Society the 5-year survival rate for lung cancer patients is still

only 16%, having increased by only 3% in the last few decades (Jemal *et al.* 2009). Diagnosis often occurs at an advanced stage and 5-year survival rates decrease from approximately 50% for clinical stage Ia to as low as 2% for clinical stage IV (Hammerschmidt and Wirtz 2009). While 10-20% of patients with non-small cell lung cancer (NSCLC) in England present with surgically resectable, early stage disease, the majority of small cell lung cancer (SCLC) patients have advanced, metastatic disease at the time of diagnosis and therefore have a much poorer survival rate (Luchtenborg *et al.* 2013).

1.1.2 Development of lung cancer

Development of cancer has long been known to involve activation of oncogenic pathways and deactivation of tumour suppressor pathways (Bowden *et al.* 1994; Kopnin 2000). The genetic basis of non-inherited, sporadic cancer development was first proposed by Alfred Knudson and his 'two-hit hypothesis' in 1971. From his study of childhood retinoblastoma, he hypothesised that two mutation events or 'hits' to a tumour suppressor gene were needed in order for the cancer to develop. In other words, both copies of the gene must be inactivated, and inheriting a single inactivated gene would confer a higher risk of an individual developing cancer since a single additional acquired mutation would be enough (Knudson 2001).

This hypothesis has provided the foundation for our understanding of tumour suppressor genes (TSGs) and the development of cancer. The term 'loss of heterozygosity' (LOH) is frequently used to describe this dual knockout of both alleles of a TSG. The hypothesis was found to apply to a number of important TSGs in cancers other than the Rb gene in retinoblastoma. The most notable of these is probably p53 which is frequently inactivated in most human cancers, including lung (Campling and el-Deiry 2003; Petitjean *et al.* 2007). Mutational events and non-

mutational or ‘epigenetic’ events can cooperate to bring about biallelic loss of TSGs (Garinis *et al.* 2002). These epigenetic events will be discussed in detail in later sections.

Another proposed mechanism of cancer initiation is that of cancer stem cells. Stem cell niches exist within many organs of the body in order to maintain the specific cell types within that organ. These are widely believed to be the originators of solid cancers, for a number of reasons. Firstly, many mechanisms of self-renewal of normal stem cells are shared by cancer cells, for example expression of the anti-apoptotic Bcl-2, as well as the Notch, Sonic hedgehog and Wnt signalling pathways. Also, their long lifespan allows for accumulation of multiple cancer-associated mutations in stem cells (Reya *et al.* 2001). Therefore tumour formation starts from hyperplastic tissue of known cell types, which continues to form a mass. The origin of the tissue forms the initial subtype classification, which will be described below.

The World Health Organisation (WHO) classification of lung tumours lists four main subtypes of malignant epithelial lung tumour. Three of these, squamous cell carcinoma (SSC), adenocarcinoma and large cell carcinoma fall into the more general category of non-small cell lung cancer (NSCLC), which accounts for ~80% of all lung cancers. The fourth subtype is small cell lung carcinoma (SCLC) which accounts for the remaining 20%. A number of less common subtypes exist such as adenosquamous carcinoma (<5% of all lung cancers), sarcomatid carcinoma (<2% of all lung cancers) as well as very rare subtypes such as squamous cell papilloma (<0.5% of all lung cancers). Each of these types can be further classified into more specific subtypes (Travis *et al.* 2004).

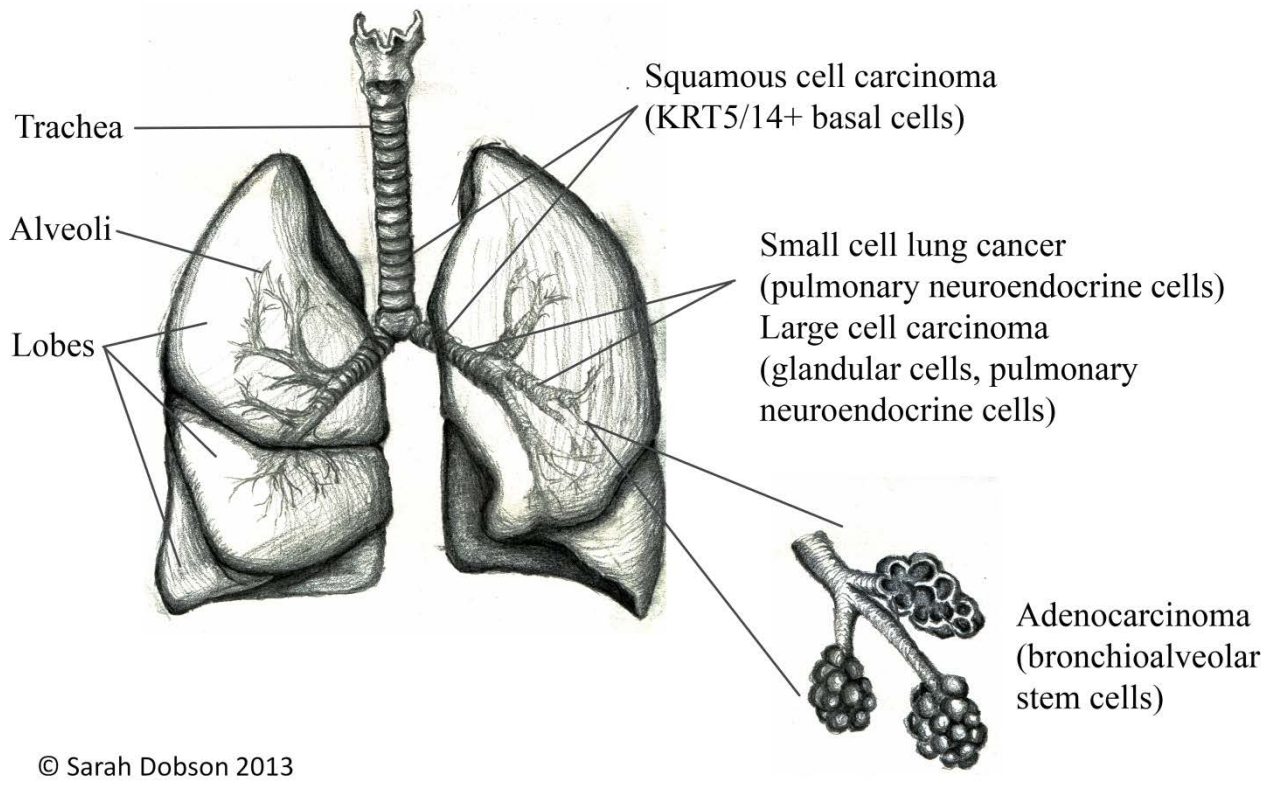


Figure 1-1 Origins of lung cancer. Moving distally from the trachea to the alveolar sacs, the distribution pattern of lung cancer subtypes is as follows: proximally near larger airways, squamous cell carcinoma (SCC), centrally within the lung, small cell lung cancer (SCLC), centrally and towards the periphery, large cell carcinoma and in the periphery, adenocarcinoma. Cell types from which these cancers have been proposed to originate from are shown in brackets.

1.1.3 Squamous cell carcinoma (SCC)

Squamous cell carcinomas (SSCs) generally occur centrally in the lung, within the upper airways, and can be further classified into papillary, clear cell, small cell and basaloid carcinomas. The main histological features of these tumours are cellular keratinisation and intercellular bridges (Travis *et al.* 2004). Keratin proteins can be used to identify specific types of epithelial cell, for example Keratin (KRT) 5 and 14 are expressed by basal cells of the trachea (Cole *et al.* 2010). These KRT5/14 positive basal cells are multipotent progenitor cells which are capable of self-renewal and of differentiating into, and replenishing, cells of the upper airway (Hong *et al.* 2004; Rock *et al.* 2009). This, together with the fact that basal cells resemble squamous cell carcinomas (SSC) in terms of their gene expression and histology, has led to the hypothesis that KRT5/14+ basal cells are the originators of these cancers (Sutherland and Berns 2010; Hanna and Onaitis 2013).

Unlike adenocarcinomas, there are not many oncogenic alterations known to occur at high frequency in SSC. However alterations involving the PI3K/AKT/mTOR pathway have been identified in 47% of SSCs and alteration of this pathway is more common to SSC than adenocarcinoma. The pathway is an important regulator of proliferation and is activated by a number of membrane tyrosine kinase receptors including EGFR and HER2. Gene amplification of the hepatocyte growth receptor (MET) has been reported at frequencies of up to 1-7% and mainly in SCC. Amplification results in over-activation of downstream signalling pathways such as the PI3K/AKT/mTOR pathway mentioned above (Cooper *et al.* 2013).

A B

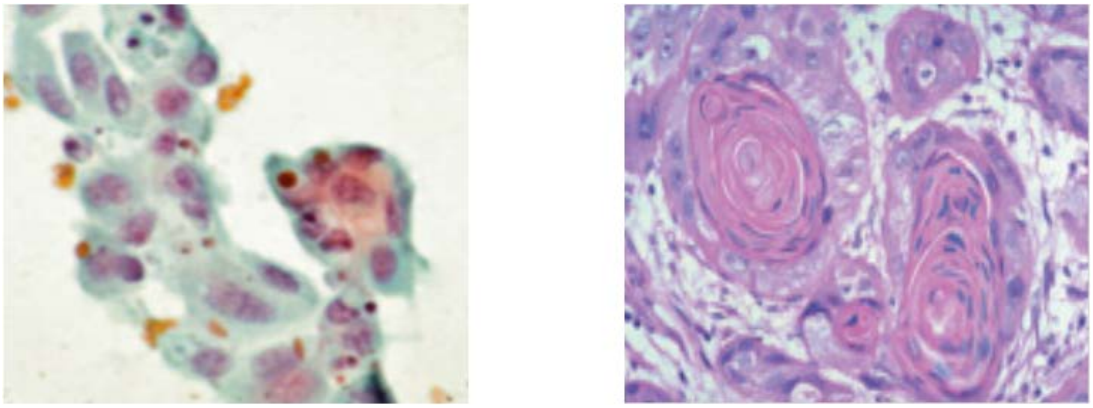


Figure 1-2 Squamous cell cancer cytology and histology. (A) Squamous cell cancer cytology. Image shows a squamous cell carcinoma without keratinisation, obtained by fine needle aspiration and stained with Papanicolaou stain. Gaps between cells and distinct cell borders identify the tumour as a squamous cell carcinoma. (B) Squamous cell cancer histology. Squamous cell differentiation is evident by the keratin pearls and prominent keratinisation. From Travis et al. 2004.

1.1.4 Adenocarcinoma

Adenocarcinomas occur within the periphery of the lung. The WHO classification of lung tumours from 2004 lists adenocarcinoma subtypes as mixed subtype, acinar, papillary, bronchioalveolar and solid adenocarcinoma with mucin production (Travis *et al.* 2004). Primary histological features of adenocarcinomas are glandular differentiation and/or mucin production. The majority however display mixed histological patterns, hence the need for extensive subtype classification (Davidson *et al.* 2013). Due to the heterogeneity and rising incidence of this disease, an international multidisciplinary reclassification has been recently proposed. Under this new system, bronchioalveolar and mixed subtype classifications are no longer used, and others have been introduced such as minimally invasive adenocarcinoma (MIA), adenocarcinoma in situ (AIS) and micropapillary adenocarcinoma (Travis *et al.* 2011)

In the peripheral lung, Clara cells replace the mucous cells of the upper airway. They produce a more watery secretion, important for airway clearance and reducing surface

tension. Most of the alveolar surface consists of alveolar epithelial cells (AECs), of which there are two types. Type I AECs comprise 40% of the alveolar cells but make up 95% of alveolar surface. Type II AECs make up the remainder of alveolar cells and their primary function is to manufacture surfactant to reduce surface tension (Shi *et al.* 2009). In mouse models, KRAS-mutated adenomas were found to be made up of cells co-expressing the Clara cell marker Clara Cell Antigen 10 (CC10) and the type II AEC marker Surfactant Protein C (SPC). Bronchioalveolar stem cells (BASCs) also express both these markers therefore it has been hypothesised that BASCs may be the cell of origin of adenocarcinoma (Jackson *et al.* 2001). In further support of this, BASCs of the normal lung respond to injury by proliferating and replenishing Clara cells and type II AECs, but will also expand in response to activation of the KRAS oncogene (Kim *et al.* 2005).

The most common molecular abnormality in adenocarcinoma is activating KRAS mutation, which occurs in 25-40% of cases (Travis *et al.* 2004; Cooper *et al.* 2013). KRAS belongs to the RAS family of proto-oncogenes which regulate cell proliferation, differentiation and survival. They are inactive in normal, quiescent cells. When activated they trigger downstream signalling within the MAPK and PI3-K pathways (Cooper *et al.* 2013). KRAS mutations in lung adenocarcinoma generally occur in codon 12 and occasionally in codon 13. Mutations commonly take the form of a G to A transition in non-smokers and G to T in smokers, and both are more frequent in Caucasian compared to Asian populations (Cooper *et al.* 2013; Takamochi *et al.* 2013).

Another well-characterised set of mutations are those of the epidermal growth factor receptor (EGFR), a member of the receptor tyrosine kinase family of proteins. These occur in 10-15% of Western patients and 30-40% of Asian patients. They occur

almost exclusively in adenocarcinomas, and primarily in female non-smokers (Cooper *et al.* 2013). Mutation of HER2, which also belongs to the EGFR family of tyrosine kinase receptors, occurs in only about 2% of NSCLC and again, adenocarcinomas account for the majority of these. However overexpression of HER2 can occur in up to 30% (MacKinnon *et al.* 2010).

Finally, genetic translocation event can result in fusion of the anaplastic lymphoma kinase (ALK) gene with the echinoderm microtubule-associated protein-like 4 (EML4) in approximately 4% of NSCLCs. The vast majority occur in adenocarcinomas of light/never smokers. This genetic event leads to constitutive activation of the ALK protein, which drives proliferation and inhibition of apoptosis (Cooper *et al.* 2013).

A B

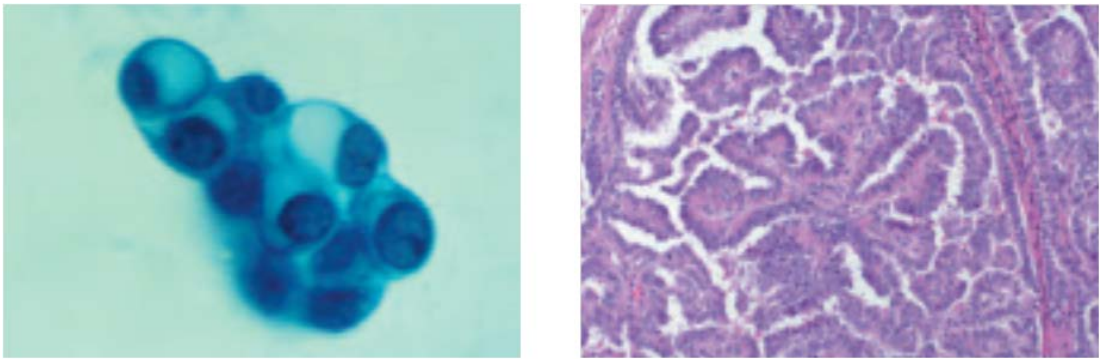


Figure 1-3 Adenocarcinoma cytology and histology. (A) Adenocarcinoma cytology. Image shows mucinous adenocarcinoma cells with Papanicolaou stain from a sputum sample. Most of the cellular volume is taken up with cytoplasmic mucin vacuoles therefore the nuclear to cytoplasmic ratio is low. (B) Adenocarcinoma histology. Image shows a papillary adenocarcinoma, which shows glandular proliferation along fibrovascular cores. From Travis *et al.* 2004.

1.1.5 Large cell carcinoma (LCC)

Large cell carcinomas generally occur in the lung periphery and can be further classified as large cell neuroendocrine, combined large cell neuroendocrine, basaloid, lymphoepithelioma-like, clear cell and large cell with rhabdoid phenotype (Travis *et*

al. 2004). They are predominantly diagnosed by a process of exclusion, whereby they cannot be classified as any other subtype, and appear as large, partially necrotic tumours composed of sheets and nests of large, polygonal cells (Davidson *et al.* 2013).

Large cell carcinomas may are differentiated from either glandular, squamous or neuroendocrine cell lineages according to cell-specific marker expression (Rossi *et al.* 2013). Most squamous epithelium of the lung is made up of either type I AECs in the periphery, or goblet, ciliated and basal cells in the upper airways, as mentioned above. Glandular cells of the upper airways are called goblet cells. They release mucous into the bronchial lumen, preventing drying-out and trapping particulate matter. Concentrated regions of mucous-producing cells called submucosal glands provide 95% of all upper airway mucous, with approximately one gland per mm^2 in the trachea (Wine and Joo 2004). On the airway surface next to these submucosal glands, pulmonary neuroendocrine cells collect into clusters called neuroendocrine bodies (NEBs). These have an endocrine function and they synthesise, store and release a number of bioactive substances such as serotonin and calcitonin gene related peptide (CGRP) (Van Lommel 2001). LCCs from neuroendocrine origins may actually have differentiated from SCLC cells. In cell line and mouse models, SCLC has been shown to transition completely into large cell tumours (Goodwin *et al.* 1983). LCC typically displays molecular markers characteristic of the cells from which it originated, and mutations characteristic of other cancers such as SCC or adenocarcinoma originating from these same sources (Rossi *et al.* 2013).

A B

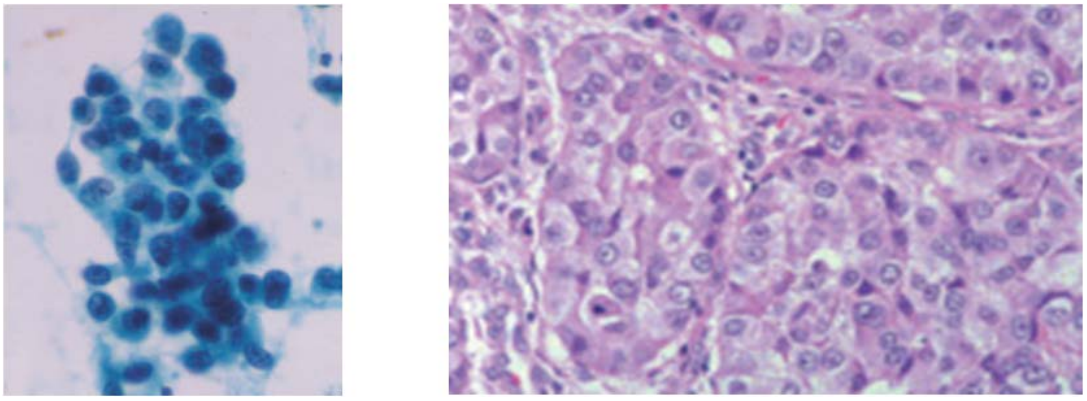


Figure 1-4 Large cell carcinoma cytology and histology. (A) Cytology of large cell carcinoma. Image shows cells obtained from a bronchial brushing specimen, with Papanicolaou stain. Cells have prominent and irregular nucleoli, and coarsely granular, dark chromatin. Nuclei have thick, irregular membranes. (B) Histology of large cell carcinoma. Cells are large with abundant cytoplasm, vesicular nuclear chromatin and prominent nucleoli. There is no evidence of glandular or squamous differentiation. From Travis *et al.* 2004

1.1.6 Small cell lung cancer (SCLC)

Small cell lung carcinoma (SCLC) generally occurs centrally in the lung and there are only two subtypes according to the WHO classification system; pure SCLC and combined (Travis *et al.* 2004). Its previous classification as 'oat cell carcinoma' (Travis *et al.* 2004) describes its appearance; it consists of small tumour cells with poorly defined borders and granular chromatin without obvious nucleoli (Davidson *et al.* 2013).

Pulmonary neuroendocrine cells (PNECs) have been suggested as the originators of small cell lung cancer (SCLC). Evidence comes from the fact that SSLC cells express neuroendocrine markers such as calcitonin gene related peptide (CGRP) and neural cell adhesion molecule 1 (Ncam1) (Sutherland and Berns 2010), and neuroendocrine peptides such as bombesin and gastrin-releasing peptide are abundant in SCLC (Wistuba *et al.* 2001). Mouse models of SCLC support PNECs as the originators of this disease, possibly due to loss of expression of p53 and Rb (Park *et al.* 2011). In

contrast to proposed stem cell/progenitor origins of SSC and adenocarcinoma, these PNECs are differentiated cells which do not proliferate during normal homeostasis. However, a recent study on PNECs in a mouse *in vivo* model has demonstrated plasticity of these cells, and the ability to contribute to Clara cell and ciliated cell populations following lung injury (Song *et al.* 2012).

SCLC has its own distinct molecular signature. In contrast to NSCLC, KRAS is never mutated in SCLC. However, retinoblastoma (Rb) is mutated far more frequently in over 90% of SCLC compared to only 15% of NSCLC. In addition, the frequency of p53 mutations in SCLC is almost double that of NSCLC (Wistuba *et al.* 2001) .

A B

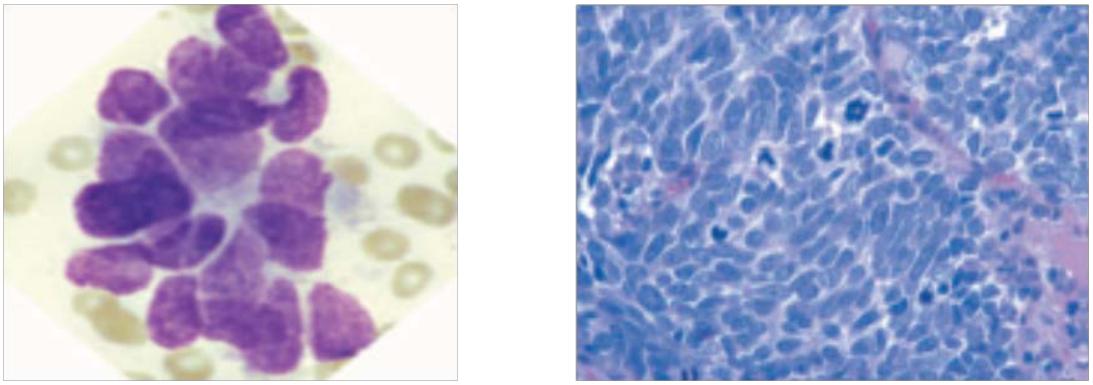


Figure 1-5 Small cell lung cancer cytology and histology. (A) Small cell lung cancer cytology. Image shows a cluster of cells with very little cytoplasm, nuclear molding and fine, granular chromatin. (B) Histology of small cell lung cancer. Image shows densely packed cells, again with finely granular chromatin and scant cytoplasm. From Travis *et al*. 2004

1.2 Current diagnostic and treatment issues

1.2.1 Diagnosis and staging of lung cancer

After classification of disease subtype, a decision must be made to determine the extent of disease and appropriate treatment. Optimal treatment of lung cancer depends on accurate staging, and early detection is crucial. Imaging techniques are widely used on a systematic basis for lung cancer diagnosis, most commonly chest X-ray,

computerised tomography (CT) and positron emission tomography (PET) scans. CT is more sensitive than X-ray for identification of early stage and asymptomatic cancers (Saintigny and Burger 2012). Integrated PET-CT scanning is superior again, providing metabolic information as well as anatomical detail (Vansteenkiste *et al.* 2010). MRI scanning can be used to evaluate brain or bone involvement if these are suspected (Peters *et al.* 2012). Biopsies are mandatory for histological confirmation of imaging results (Crino *et al.* 2010).

Screening strategies have been suggested in an attempt to increase early detection of lung cancer. Low dose CT scanning allows for sensitive detection of small lung nodules in early stage and asymptomatic patients, at an acceptable level radiation but without loss of sensitivity compared to regular dose CT (Diederich 2011). Therefore this may represent a suitable screening strategy for high risk patients. However there is conflicting evidence with regard to the benefit to patients in terms of outcome and survival. While some studies report up to 20% reduction in mortality with low dose CT screening at 1-year intervals, others could not find a significant reduction (Crino *et al.* 2010).

All cancers are staged using the international Tumour, Node, Metastasis (TNM) system published by the Union for International Cancer Control. These staging guidelines are continuously being revised and updated. The T descriptor refers to the size and spread of the primary tumour, starting at T1a for small tumours of ≤2mm and increasing thereafter. The N descriptor refers to lymph node involvement, starting at N0 for no involvement. The M descriptor refers to metastasis, with M1a indicating metastasis within the lung, M1b indicating metastasis elsewhere in the body. Various combinations of TNM values can then be grouped into a more singular stage value, from I to IV. Stage Ia and Ib are the smallest tumours (T1 and T2), with no lymph

node involvement or metastasis (N0, M0), while stage IV tumours can be any size but will have metastasized (Marshall *et al.* 2012). In addition to confirmed diagnosis and staging, evaluation of performance status (PS) of the patient and lung function status are the other main factors which informs treatment decisions (Hammerschmidt and Wirtz 2009).

1.2.2 Conventional treatments

For SCLC, if diagnosed with limited disease at stage I-III, chemoradiotherapy with etoposide/cisplatin is recommended. 5-year survival rate can be 20-25% for these patients, but the vast majority of SCLCs are diagnosed with extensive disease. Combined chemotherapy as above can be administered but the median survival time is just 10 months (Sorensen *et al.* 2010), making this cancer subtype particularly difficult to treat.

For early stage NSCLC (stage I and II), resection is the standard treatment for patients without contraindications for surgery. Surgical procedures include lobectomy (removal of a single affected lobe), bilobectomy (removal of two lobes), or pneumonectomy (removal of a whole lung) (Hammerschmidt and Wirtz 2009), but most commonly involve lobectomy with systematic lumph node dissection (Crino *et al.* 2010). Following surgery, platinum-based chemotherapy is recommended for stage II but not stage I tumors. For patients who cannot have surgery, radiotherapy with curative intent is recommended (Hammerschmidt and Wirtz 2009). Neo-adjuvant chemotherapy is not common practice and is still being evaluated, but a review of clinical trial data for the European Society for Medical Oncology (ESMO), found that while a significant benefit could not be shown, the evidence was in favour of neo-adjuvant platinum-based chemotherapy for early stage NSCLC (Crino *et al.* 2010).

Locally advanced (stage III) are a more heterogeneous group of cancers and so have been divided into 6 sub-groups to define more specifically the anatomy of nodal involvement. These tumours may be resectable, but survival after resection alone is low therefore adjuvant platinum-based therapy is recommended. For unresectable stage III tumours, chemoradiotherapy may be administered either sequentially or concomitantly (Hammerschmidt and Wirtz 2009; Crino *et al.* 2010).

As many as 70% of NSCLC patients present with advanced, stage IV lung cancer, for which only palliative treatment can be offered (Saintigny and Burger 2012). In patients with a performance status (PS) of 0-2, several regimens of combined chemotherapies may be considered. Histology is vitally important in treatment of advanced cancers, as the two main cancer sub-types, SCC and adenocarcinoma, respond differently to treatment. Chemotherapy is not suitable for patients with PS 3-4. The chemotherapy agent Pemetrexed is suitable for use in non-squamous NSCLC only, as is the anti-angiogenic monoclonal antibody Bevacizumab, which targets vascular endothelial growth factor (VEGF) (Peters *et al.* 2012). Targeted therapies such as this will be described below.

1.2.3 Targeted therapies

The emergence of targeted therapies against driver mutations means a new treatment option for patients whose tumours have specific genetic alterations. Currently available targeted therapies for lung cancer consist of the small-molecule tyrosine kinase inhibitors (TKIs) gefitinib and erlotinib, which were approved by the FDA in 2003 and 2004. These drugs are prescribed for advanced NSCLC patients harbouring EGFR activating mutations, who have not responded to conventional therapies (Sharma *et al.* 2007). They are also now approved for first-line treatment of advanced-stage NSCLCs with EGFR activating mutations, including those with PS 3-4, due to

an improvement in survival and quality of life (Peters *et al.* 2012). While the response to TKIs in patients with activating mutations can be dramatic, not all individuals with these mutations will show tumour shrinkage as a result of treatment. Stabilisation of disease rather than shrinkage will occur in some cases, and resistance to targeted therapy can develop within 6-12 months (Sharma *et al.* 2007).

It is imperative therefore that new therapies are introduced. Cetuximab, a monoclonal antibody against EGFR, is one emerging example. The clinical effects of Cetuximab in lung cancer have not been investigated as thoroughly as the TKIs, but evidence suggests it may increase responsiveness in combination with first line chemotherapy (Chi *et al.* 2013). It is currently approved for use in metastatic colorectal cancer and locally advanced head and neck cancer, and is being evaluated in clinical trials for NSCLC (Cavazzoni *et al.* 2012).

Another emerging target is the ALK tyrosine kinase. As previously mentioned, the ALK-EML4 gene rearrangement acts as an oncogenic driver in NSCLC. Although only discovered in 2007, development of a targeted therapy against ALK amplification proceeded rapidly. Crizotinib (an inhibitor of the ALK tyrosine kinase) was approved by the FDA in 2011 for treatment of ALK-EML4-positive NSCLC patients (Ma 2012). It is now recommended in clinical practice guidelines by ESMO for first-line treatment of these patients, and for second or third-line treatment if not used previously (Peters *et al.* 2012). Very recently a Phase 3 trial has demonstrated improved progression-free survival in patients treated with Crizotinib compared to patients receiving chemotherapy, confirming superiority of targeted treatment in specific subsets of patients (Shaw *et al.* 2013).

1.2.4 Future possibilities for targeted therapy

The therapies described above have broken new ground in the last decade and have introduced new alternatives to the standard chemotherapy drugs. However, not all patients will harbour a driver mutation for which a targeted therapy is available and for those that do, resistance is the norm rather than the exception.

Resistance to TKIs such as erlotinib typically develops within a median of 10-14 months (Oxnard *et al.* 2011). There are a number of reasons for this acquired resistance. One is the initiation of alternative routes to activate the signalling pathways formerly activated by EGFR. For example, drug-inhibited EGFR can be bypassed by overexpression of MET, as coupling of MET and HER2 leads to sustained activation of the PI3K/AKT pathway otherwise activated by EGFR (Oxnard *et al.* 2011; Sechler *et al.* 2013). EGFR inhibitors work by competitively binding to the ATP-binding site on the EGFR molecule. Absence of ATP-binding renders downstream signalling impossible. However, the tumour frequently evades this by means of a secondary mutation; a single amino acid substitution at position 790 results in an increased affinity for ATP by the EGFR, so that inhibitor drugs cannot sufficiently compete (Sechler *et al.* 2013).

A number of solutions present themselves for this; one involves building on what we know and maximising its utility. Research continues to improve the efficacy of existing targeted treatments, optimise drug combinations and developing more effective drugs for known driver mutations. Further research is ongoing towards development and approval of EGFR inhibitors which target multiple receptors and bind irreversibly to these, as well as novel receptor-targeting monoclonal antibodies (Hirsch *et al.* 2013). The novel monoclonal antibody Figitumumab is in advanced stages of clinical development, targeting insulin-like growth factor type 1 receptor

(IGF-1R) which has been shown to bypass EGFR (Dempke *et al.* 2010). As MET amplification also bypasses EGFR, dual targeting of these two receptors is being explored in an effort to combat resistance (Ma 2012).

A different take on the traditional approach would be to identify novel targets and biomarkers. Existing targeted treatments are only suitable for a subset of patients, and usually only those with adenocarcinoma since the majority of sensitising mutations occur in this group. Therefore there is an urgent need to expand the range of molecular targets and therapies, as well as biomarkers to predict outcome and response. While the current targeted therapies are directed against activating oncogenic processes, genetic events leading to loss of tumour suppressor functions are of equal importance. These may be considered as biomarkers, and potentially new therapeutic targets; loss of a tumour suppressor will result in loss of inhibition of the proto-oncogenic genes/proteins/pathways which it regulates, and these may be targetable.

One of the more common mechanisms of gene silencing in cancer is DNA methylation. Methylation signatures are being explored for biomarker use in early diagnosis, risk stratification, prediction of response to therapy and monitoring of resistance (Baylin and Herman 2000). DNA methylation will be discussed in detail in the following section.

1.3 Epigenetics and cancer

1.3.1 Introduction to epigenetics

Epigenetic processes are those that result in stable gene silencing which can be passed on through generations of cell division. The main features of this type of regulation are DNA methylation, histone modification and non-coding RNAs. The first two are

closely related in that the addition of a methyl group may lead to modifications of the histone which cause the chromatin to adopt a closed, condensed form, preventing transcription of the gene at that site (Holliday 2006). This mechanism of gene silencing is crucial to many natural processes in mammals such as embryogenesis, placental development, genomic imprinting and X-chromosome inactivation. However, epigenetic defects can accumulate over time and play a major role in cancer development (Schuebel *et al.* 2007).

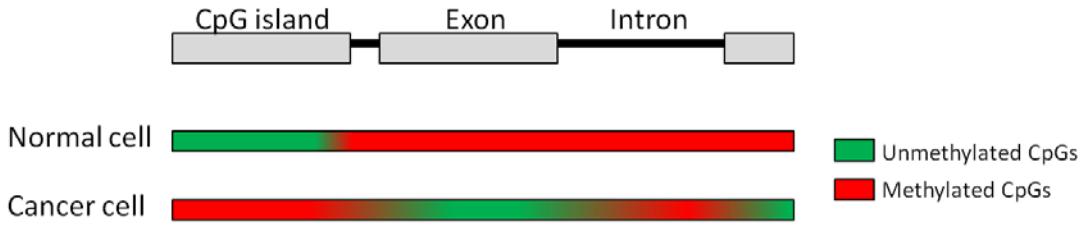


Figure 1-6 Global methylation changes in cancer cell DNA. In normal cells, CpG islands in gene promoter regions are unmethylated, while CpG dinucleotides outside of these are generally methylated. Promoter hypermethylation and genome-wide hypomethylation are hallmarks of cancer.

Addition of a methyl group to DNA only occurs at CpG (cytosine-phosphate-guanine) dinucleotides and clusters of CpG dinucleotides are known as CpG islands. These dinucleotides occur throughout the genome and 50-70% in total are methylated. In normal cells, CpG islands often found within gene promoters are unmethylated while CpGs outside of these islands are generally methylated. In cancer cells however, this pattern may be inverted. Common features of cancer cells are global hypomethylation which can lead to genomic instability, and promoter hypermethylation (Figure 1-6). This promoter hypermethylation can result in silencing of tumour suppressor genes and others crucial for metabolism of exogenous toxins, DNA damage repair and maintenance of normal cell cycle (Gronbaek *et al.* 2007). Methylation patterns are maintained in successive generations of replicating cells by the enzyme DNA

methyltransferase 1 (DNMT1), while other DNMTs such as DNMT3a and 3b are responsible for *de novo* methylation. This methylation is accompanied by histone modifications, the most common of which is deacetylation, catalysed by histone deacetylases (HDACs). These enzymes continually remove acetyl groups from histone residues and cause the chromatin to adopt a closed, transcriptionally silenced structure. Other histone modifications such as methylation also affect chromatin structure (Esteller 2007).

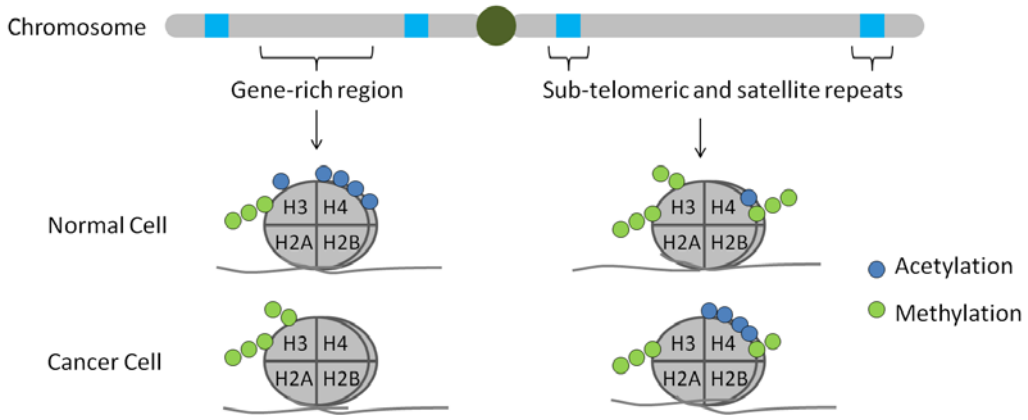


Figure 1-7 Histone modifications of cancer cells. Acetylation and methylation patterns of core histones H2-H4 are shown. Histones adjacent to promoter regions of tumour suppressor genes are normally characterised by transcriptionally active histone marks such as trimethylation of H3 and acetylation of H3 and H4. In cancer cells this pattern is disrupted. In particular, deacetylation of histones H3 and H4 is important in transcriptional silencing of tumour suppressors. Disruption of histone marks on sub-telomeric and satellite repeat regions contribute to genomic instability. Image adapted from (Esteller 2007).

In human cancers, aberrant methylation of genes is a far more frequent event than genetic mutation (Taby and Issa 2010); therefore it represents a valuable source of new biomarkers and therapeutic targets, which will be discussed in the following section. Global epigenetic changes are also a target for treatment with inhibitors of DNMTs and HDACs. By inhibiting the enzymes which maintain DNA methylation and histone deacetylation, these drugs allow re-expression of epigenetically silenced

tumour suppressor genes. A number of DNMT inhibitors (Azacytidine, Decitabine) and HDAC inhibitors (Vorinostat, Romidepsin) are FDA-approved for treatment of myeloblastic leukaemia and lymphoma (Nebbioso *et al.* 2012). A number of epigenetic therapies are currently being evaluated in lung cancer clinical trials both as monotherapies and in combination with chemotherapeutic drugs, but these have so far failed to live up to the promise of pre-clinical studies. However, combinations of epigenetic drugs for targeting of both DNA methylation and histone modifications are showing promise and extensive clinical evaluation is ongoing (Jakopovic *et al.* 2013; Liu *et al.* 2013).

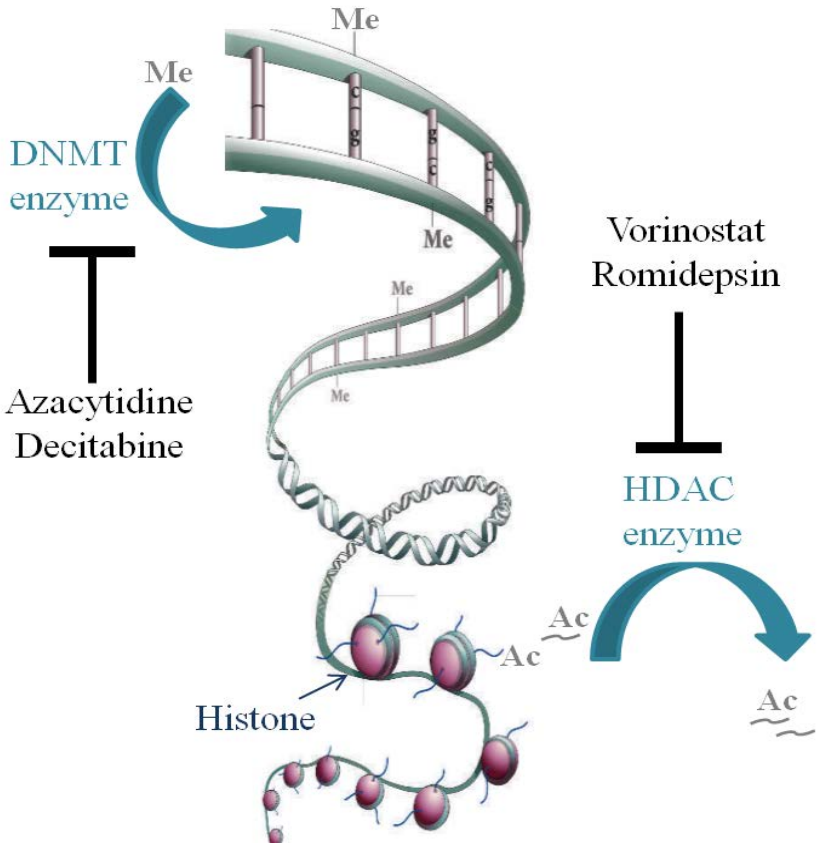


Figure 1-8 Epigenetic gene silencing. Two main processes maintain the silenced state of a gene. DNA methyltransferase (DNMT) enzymes maintain methylation patterns on DNA through sequential cell divisions. Histone deacetylases (HDAC) enzymes remove acetyl groups from lysine residues on histones, allowing them to bind tightly to DNA, packaging it into a closed format which transcription factors cannot access. Many inhibitors of DNMTs and HDACs are being explored as cancer therapies, examples of each are shown above. Image adapted from Adapted from Qiu 2006.

1.3.2 DNA methylation as a biomarker

Research is ongoing to develop diagnostic, prognostic and predictive biomarkers for lung cancer. These markers mainly consist of proteins, mRNA and more recently DNA methylation markers. DNA methylation patterns are relatively stable over time and can be amplified by PCR; therefore they avoid many of the pitfalls associated with RNA and protein biomarkers (Tost 2009). As methylation patterns are not reflected in the DNA sequence, they cannot be detected by PCR in their naturally-occurring state. However, treatment of DNA with sodium bisulfite results in deamination of unmethylated cytosine residues to uracil, while 5-methylcytosines will remain unchanged (Frommer *et al.* 1992). This results in differing sequences for methylated and unmethylated DNA (Figure 2-2), which can then be quantified by PCR and sequencing-based methods such as methylation-specific PCR (Herman *et al.* 1996), bisulfite sequencing (Frommer *et al.* 1992), methylation-sensitive high resolution melting (MS-HRM) (Wojdacz and Dobrovic 2007), and pyrosequencing (Ronaghi *et al.* 1998; Tost and Gut 2007). A biomarker should ideally be measured in a minimally invasive way. In cancer the amount of plasma DNA is increased and contains DNA fragments originating from the site of tissue damage. This will display the same genetic and epigenetic characteristics as the tumour DNA, therefore methylation biomarkers in plasma can be measured in a relatively non-invasive way (Ziegler *et al.* 2002).

DNA methylation patterns satisfy another crucial criterion for biomarker use in that many are tumour-specific, non-random and contribute to the distinct gene expression profiles of the tumours in which they are found (Costello *et al.* 2000). Methylation patterns have been shown to be capable of distinguishing between squamous cell carcinoma and adenocarcinoma, using single gene markers (Kusakabe *et al.* 2010) or

using panels of genes (Brena *et al.* 2007; Anglim *et al.* 2008). This specificity may be explained by the fact that the type of tumour which will develop in a tissue depends largely on the combination of environmental risk factors to which that tissue is exposed, and specific environmental factors have been shown to target specific genes. For example in lung cancer there is a strong association between MTHFR methylation and tobacco smoking (Vaissiere *et al.* 2009). Like disease-associated genetic events, epigenetic ones will also begin to accumulate within the genome long before tumour presentation. In genes such as P16 and MGMT, promoter methylation has been detected in sputum up to 3 years before diagnosis in patients with squamous cell carcinoma, suggesting a possible use for screening populations of high-risk individuals (Palmisano *et al.* 2000).

Therefore, it is of great interest to investigate novel epigenetic markers in lung cancer which can be used for early diagnosis, to stratify high risk patients who may benefit from adjuvant therapy or who may respond to targeted therapy, and to illuminate new therapeutic targets and mechanisms of cancer progression. In this thesis, I will present a novel screening mechanism to identify genes which may be of interest and expand on two genes, one of which presents itself as a potential biomarker.

2 Materials and Methods

2.1 DNA preparation and analysis

2.1.1 DNA extraction

DNA was extracted from cell lines using the DNeasy Blood &Tissue Kit (Qiagen) as per manufacturer's instructions and quantified using a spectrophotometer (Nanodrop, Thermo Scientific). DNA from FFPE tissues was extracted using an optimised protocol. Briefly, tissue slices were deparaffinised in 10ml xylene overnight while shaking, then washed twice in ethanol by spinning at 4,000 rpm and decanting liquid from pellet. Tissue was digested by incubating in 0.2mg/ml proteinase K in 300-500µl lysis buffer (Tris 100 mM, EDTA 5 mM, SDS 0.2%, NaCl 200 mM) at 55°C for 1-5 days until digestion was complete. DNA was then purified using phenol saturated with Tris (Sigma). 300-500µl of phenol were added to the sample, mixed then centrifuged at 14,000rpm for 10 minutes at 4°C. The aqueous layer containing the DNA was removed from the top of the sample without disturbing the interface or bottom layer. DNA was precipitated by adding 1.2 volumes isopropanol to the aqueous phase, spinning at 14,000rpm for 20 minutes at 4°C and then 500µl 100% ethanol at -80°C for 1 hour, followed by spinning as before. After removal of ethanol the DNA pellet was dried for 15 minutes, resuspended in ddH_2O and quantified using a spectrophotometer (Nanodrop, Thermo Scientific).

2.1.2 Bisulfite conversion of DNA

Bisulfite modification of DNA was carried out using the EZ DNA methylation kit (Zymo Research) exactly as per manufacturer's instructions, using either 500ng of cell line DNA or 1µg of FFPE tissue DNA. The kit is based on the three-step reaction that takes place between unmethylated cytosine and sodium bisulfite where cytosine is

converted into uracil (Figure 2-1). The converted DNA sequence therefore reflects the methylation status and is suitable for downstream PCR and sequencing analyses.

2.1.3 Microarrays

13 lung cancer cell lines were profiled using the Infinium HumanMethylation450 Beadchip Kit. Genomic DNA, extracted as previously described, was quantified against a lambda DNA standard curve using Picogreen assay (Invitrogen). 1µg of DNA was then sent to Barts and the London Genome Centre (Queen Mary University of London) for microarray hybridisation, scanning (Illumina iScan system) and quality control according to standard operating procedures within the facility.

The scanned data were normalised using GenomeStudio software (Illumina) to reduce differences between the two types of probes contained within the array. Infinium I probes are designed across the CpG dinucleotide so there are two for each CpG site, whereas Infinium II probes are directly adjacent to the CpG so that there is only one for each dinucleotide. After normalisation, methylation status was calculated as beta values, i.e. the ratio of methylated signal to total signal.

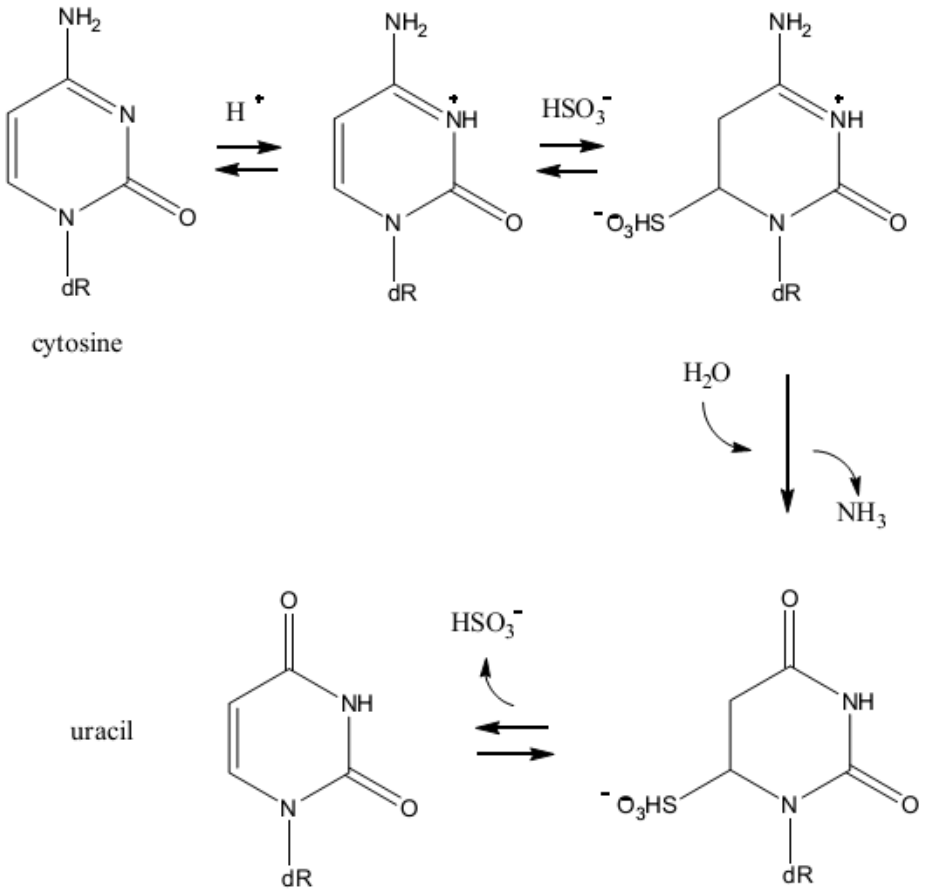


Figure 2-1 Conversion of unmethylated cytosine to uracil: Stepwise reaction of bisulfite with protonated C leads to deamination to form uracil. Cytosine will react with bisulfite but 5-Me-Cytosine does not react (Applied_Biosystems 2007).

2.1.4 Methylation-specific PCR (MSP)

MSP relies on sequence differences arising from bisulfite conversion of DNA to distinguish methylated from unmethylated sequence. For each region of DNA analysed, primers (Table 2-1) targeting specifically methylated or unmethylated sequence were each used to amplify DNA by PCR (Figure 2-2). Where possible, multiple primer pairs were used. PCR was carried out using ThermoStart *Taq* PCR Master Mix (Thermo Scientific) with addition of 0.5µM primers and 10-20ng bisulfite-converted DNA in a 20µl reaction. The following cycling conditions were used: 95°C for 13 minutes, followed by 40 cycles of 95°C for 30 seconds, primer

annealing temperature 30 seconds, 72°C for 30 seconds, followed by a final extension step at 72°C for 7 minutes in a Veriti 96-well thermal cycler (Applied Biosystems). Commercially available unmethylated and methylated DNA (Millipore) were included in all bisulfite conversion and PCR steps as controls. PCR products were run on 2% agarose gels with ethidium bromide at 0.1µg/ml (Sigma) and visualised under UV light (Syngene G:BOX Chemi).

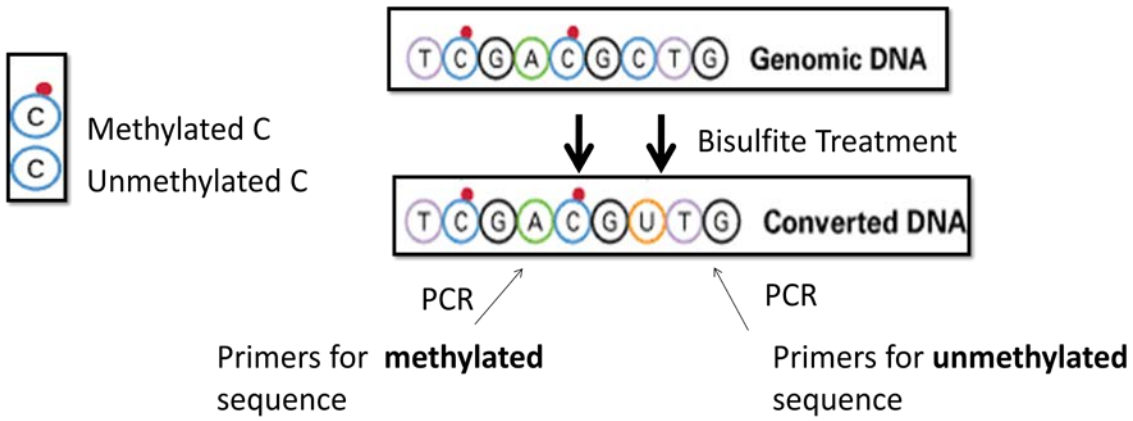


Figure 2-2 Methylation-specific PCR: unmethylated cytosine is converted to uracil, methylated cytosine remains unchanged. Uracil forms a base pair with adenine and is replaced by thymine during strand synthesis. Two primer sets are designed to target methylated and unmethylated sequences of bisulfite-modified DNA, i.e. the same region of DNA with CpG dinucleotides as either CpG or UpG, respectively.

Gene name	Forward primer	5'-3' sequence	Reverse primer	5'-3' sequence	Annealing temp (°C)	Source
Endoglin	M	TATTTTAGAAGGTTGGAGTAGGGAC	M	CTATCCACGTAAAAACCTATACGCT	58	(Wong *et al.* 2008)
	U	TTTTAGAAGGTTGGAGTAGGGATG	U	CTATCCACATAAAAACCTATACACT		
VIK-1	M(1)	TTCGGTGAGTATTCGGAAGTC	M(1)	CGAAAAAACGAACTACCGTC	52	(Nelofer Syed, Imperial College London)
	U(1)	TTTTTGGTGAGTATTTGGAAGTTG	U(1)	CACTCTCAAAAAAACAAACTACCATC		
VIK-1	M(2)	TCGACGGTAGTTCGTTTTTTC	M(2)	CCGATTTACTACACGAACTTAACGAA	52	
	U(2)	TTGATGGTAGTTTGTTTTTTTGA	U(2)	CAATTTACTACACAAACTTAACAAA		
VIK-1	M(3)	GTTTTTAGGATTTCGGGAAAAGGTC	M(3)	CCGATTTACTACACGAACTTAACGAA	52	
	U(3)	TTTTAGGATTTTGGGAAAAGGTTGG	U(3)	CAATTTACTACACAAACTTAACAAA		

Table 2-1 MSP primer pairs. M= methylated primer, U= unmethylated primer. Numbers in brackets differentiate sets of primers within different regions of the same CpG island.

Gene Name	Primer Forward 5’	Primer Reverse 3’	Annealing temp (°C)	Source
Endoglin	GAAGGTTGGAGTAGGGA	CCTTATCCTATATCCAATAAC-5’[BTN]	52	Cristiana lo Nigro, S. Croce General Hospital, Cuneo, Italy
VIK-1	GGGTAAGGTTTTTTGAGGA-5’[BTN]	CCACTTTTAAAATCAAACCCTT	54	Nelofer Syed, Imperial College London

Table 2-2 Pyrosequencing primer pairs. BIOT = biotinylated primer.

2.1.5 Pyrosequencing

The region to be sequenced was amplified by PCR prior to sequencing. Bisulfite-converted DNA (50ng cell line and control DNA, 200-300ng FFPE tissue DNA), was amplified by PCR using the primers listed in Table 2-2. PCR was carried out with the Pyromark PCR Kit (Qiagen) with the addition of 0.5µM primers in a 30µl reaction, using the following cycling conditions: 95°C for 15 minutes, followed by 40 cycles of 94°C for 30 seconds, primer annealing temperature 30 seconds, 72°C for 30 seconds, followed by a final extension step at 72°C for 10 minutes in a Veriti 96-well thermal cycler (Applied Biosystems). Primers were designed on regions lacking any CpG dinucleotides to ensure amplification regardless of methylation status. As DNA recovered from FFPE tissues is considerably fragmented, primers were designed to amplify the smallest possible fragment which would cover several CpG dinucleotides. One primer of each pair was biotinylated at the 5'end. 3µl of PCR product was resolved on 2% agarose gels to confirm the success of the reaction prior to sequencing.

For pyrosequencing, the biotinylated PCR products were captured by binding to streptavidin-coated sepharose beads and used as the template for the sequencing reaction. Using the PyroMark Q96 Vacuum Prep Workstation (Qiagen), streptavidin-bound PCR products were immobilised, washed, denatured and allowed to anneal to the sequencing primer for 2 minutes at 80°C. The non-biotinylated PCR primer was used as the sequencing primer and was extended by addition of nucleotides by the PyroMark Q96 ID pyrosequencer (Qiagen). Briefly, in pyrosequencing, sequential incorporation of nucleotides results in light emissions which are detected by the instrument and allow determination of the DNA sequence. C not followed by G should always be converted by bisulfite treatment and serves as an internal control

within each sequencing reaction. Methylation at each CpG dinucleotide was calculated from the ratio of C to T in the bisulfite-modified sequence (Figure 2-2).

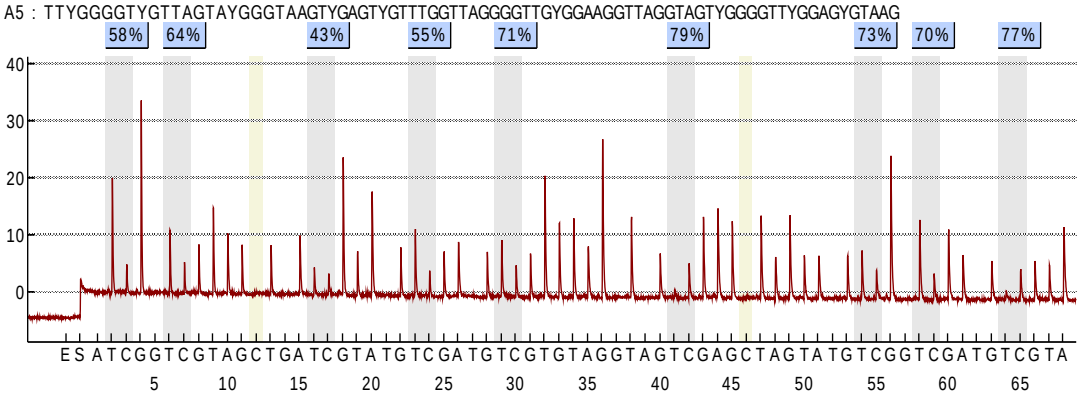


Figure 2-3 Representative program generated by the Pyro Q-CpG™ Software showing methylation levels of 9 CpG dinucleotides. The x-axis shows the nucleotide dispensation and peaks represent light emission following nucleotide incorporation. Ratio of C to T is automatically calculated by the software at each CpG site.

2.2 RNA and protein extraction and analysis

2.2.1 RNA extraction

RNA was extracted using a phenol-based protocol. TRI-Reagent (Sigma) was added to cells for 5 minutes, and homogenised using a pipette tip. The mixture was spun to remove debris and 1-bromo-3-chloro-propane (BCP) was added at a 1:10 dilution. Cells were vortexed then centrifuged at 12,000g for 10 minutes at 4°C. The top layer consisting of the aqueous phase was removed to a separate tube and 1x volume of 100% propan-2-ol was added to this. After 10 minutes incubation at room temperature (RT), the sample was spun at 12,000g for 8 minutes at RT. Supernatant was removed and 75% ethanol added to the pellet. The pellet was spun for 5 minutes at 7,500g, air-dried to remove excess ethanol and re-suspended in RNAse-free H_2O. RNA was quantified using a spectrophotometer (Nanodrop, Thermo Scientific).

2.2.2 Microarrays

13 lung cancer cell lines were profiled using the HumanHT-12 v4 Expression BeadChip Kit (Illumina). RNA was extracted using the mirVana™ RNA isolation kit (Ambion), for isolation of both small and large RNA molecules. RNA quality was verified on a Bioanalyser instrument (Agilent Technologies) using microfluidic chips from the Agilent RNA 6000 Nano Kit (Agilent Technologies). Only RNA samples with an RNA integrity number (RIN) of 7.0 or greater were used for microarray analysis. RNA of satisfactory quality was then amplified with the Illumina TotalPrep ™-96 RNA amplification kit, which generates large scale amounts of cRNA suitable for hybridisation with Illumina arrays. RNA was then sent to the UCL Genomics facility (University College London) for array hybridisation, scanning (Illumina iScan system) and quality control according to standard operating procedures within the facility. Expression values were quantile-normalised on GenomeStudio software (Illumina) and represented as the average of biological duplicates.

2.2.3 Reverse transcription and semi-quantitative real-time PCR (qRT-PCR)

100ng of RNA were reverse transcribed into cDNA in a 20µl reaction using the High Capacity cDNA Kit (Applied Biosystems), as per manufacturer's instructions. Cycling conditions were as follows: 25°C for 25 minutes, 37°C for 120 minutes and 85°C for 5 minutes. qRT-PCR reactions were conducted in triplicate using 1ng of cDNA for each reaction. Commercially available Taqman gene expression assays (Applied Biosystems) together with Brilliant II QPCR Master Mix with low ROX (Agilent Technologies) were used for each gene (see table 2-3 for assays used). Taqman assays operate by generating a fluorescent signal proportional to the amount of target cDNA in the PCR reaction. Briefly, the assay contains unlabelled forward and reverse primers and a probe labelled with a fluorophore at the 5' end and a quencher at the 3'

end. As DNA polymerase extends the template strand of DNA, it cleaves the fluorophore molecule of the DNA-bound probe. When released from the quencher, this molecule emits a fluorescent signal which is quantified by the instrument. The cycling program was as follows: 95°C for 10 minutes followed by 40 cycles of 95°C for 15 seconds, 60°C for 1 minute on the Stratagene Mx3000P QPCR System (Agilent Technologies). For all qRT-PCR analyses, RPLP0 was used as a housekeeping reference gene for each sample. For screening of the cell line panel for baseline expression of a gene, the method $2^{-\Delta C_T}$ method of quantification was used, whereby samples are normalised to a reference gene only. Therefore fold changes of each cell line are independent of one another, rather than normalised to one of the samples. For all other experiments, the $2^{-\Delta\Delta C_T}$ method was used, whereby data are normalised to the reference gene and then fold change is shown relative to the untreated control (Livak and Schmittgen 2001).

Gene	Assay
Endoglin	Hs00923996_m1
Vimentin	Hs00185584_m1
ZNF655	Hs01002760_m1
Vav-1	Hs00232108_m1
CDKN2A (p16)	Hs00923894_m1
RPLPO	Hs02992885_s1

Table 2-3 Taqman expression assays (Applied Biosystems) used for each gene.

2.2.4 Cell lysis and Western blot

Cells were washed in PBS and lysed in ice-cold RIPA buffer (150 mM sodium chloride, 1.0% NP-40 or Triton X-100, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulphate, 50 mM Tris, pH 8.0) with protease inhibitor (cOmplete Mini Protease Inhibitor, Roche) with constant agitation for 30 minutes. Lysates were centrifuged at 10,000rpm for 10 minutes to remove debris and the protein was quantified using the *DC* Protein Assay (Bio-Rad) as per manufacturer's protocol. Briefly, 5µl bovine serum albumin (BSA) standard solutions from 0.125 to 2mg/ml were added to a 96-well microplate, as well as 5µl of lysates. Reagents were added as per protocol and the plate was read at 750nm on a plate reader (Biotek Synergy HT). Lysates were quantified by comparing absorbance to that of standards using a standard curve. Lysates were diluted in 2x Laemmli buffer (4% SDS, 20% glycerol, 0.004% bromophenol blue, 0.125 M Tris HCl, pH6.8) with 10% β-mercaptoethanol, and denatured for 10 minutes at 90 degrees. Proteins were separated by electrophoresis on an 8% polyacrylamide gel using a standard migration buffer (25 mM Tris base, 190 mM glycine, 0.1% SDS, pH8.3) in a Mini Trans-Blot® Cell (Bio-Rad). The gel was equilibrated in transfer buffer (25 mM Tris base, 190 mM glycine, 20% methanol) and proteins were blotted onto a PVDF membrane under wet transfer conditions (1 hour at 300mM on ice). Membranes were blocked in 5% non-fat milk in TBST (50 mM Tris, 150 mM NaCl, 0.05% Tween20) for 1 hour, shaking, at room temperature. After washing 3 times for 5 minutes in TBST, membranes were incubated with primary antibody diluted in blocking buffer overnight, shaking, at 4°C. Primary antibodies used were mouse anti-CD105 (611314, BD Transduction Laboratories) at 1:600 dilution, rabbit anti-ZNF655 (ARP39587, Aviva Systems Biology) at 1:3000 dilution, rabbit anti-β-actin (4970L, Cell Signalling) at 1:5000 dilution. Secondary antibodies

used were HRP-linked anti-mouse IgG (7076S, Cell Signalling) and HRP-linked anti-rabbit (7074S, Cell Signalling). Membranes were washed as before and incubated with secondary antibody for 1 hour, shaking, at room temperature. Membranes were washed and bound antibody was visualised using ECL Plus Western Blotting Detection Reagent (GE Healthcare) and the SynGene G:Box Chemi gel documentation system using optimised capture times.

2.3 Cell Culture Assays

2.3.1 Cell culture and reagents

All cell lines are listed in Table 2-3. The cell lines A549, HCC193 and HCC95 were maintained in DMEM (PAA Laboratories), and all others in RPMI (PAA Laboratories) supplemented with 2mM L-glutamine, 50µg/ml streptomycin, 0.5 Units/ml penicillin and 10% heat inactivated foetal calf serum (Gibco) in a 37°C humidified chamber with 5% CO_2 and 95% humidity. Transfected cell lines were maintained in medium containing G418 (Sigma) as follows: 300µg/ml for HCC95 and H23, 500µg/ml for A549, HCC193, H322M and MOR.

2.3.1 Methylation reversal assays

Cells were grown in culture medium until 60% confluent. Fresh medium containing 5 µM of 5-aza-2'-deoxycytidine (AZA) (Sigma) was added and cells were incubated for 5 days. The medium containing AZA was refreshed every 2 days. 0.3 nM of Trichostatin A (TSA) (Sigma) was added to media for the final 16 hours prior to harvest. On day 5, cells were counted, washed in PBS and pelleted at $1x10^6$ cells per pellet and stored at -20°C until DNA, RNA and protein extraction.

Cell line	Tumour Type	Sub-type	Growth mode	Source Information
EKVX*	NSCLC	Adenocarcinoma	Adherent	DCTD tumour repository, U.S. National Institutes of Health
Hop62	NSCLC	Adenocarcinoma	Adherent	DCTD tumour repository, U.S. National Institutes of Health
H23*	NSCLC	Adenocarcinoma	Adherent	DCTD tumour repository, U.S. National Institutes of Health
A549*	NSCLC	Adenocarcinoma	Adherent	DCTD tumour repository, U.S. National Institutes of Health
HCC193*	NSCLC	Adenocarcinoma	Adherent	A gift from Michael Seckl, Imperial College, London
HCC78*	NSCLC	Adenocarcinoma	Adherent	Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures
H322M*	NSCLC	Adenocarcinoma	Adherent	DCTD tumour repository, U.S. National Institutes of Health
MOR*	NSCLC	Adenocarcinoma	Adherent	HPA culture collections, Public Health England
H460*	NSCLC	Large cell carcinoma	Adherent	DCTD tumour repository, U.S. National Institutes of Health
Hop92*	NSCLC	Large cell undifferentiated	Adherent	DCTD tumour repository, U.S. National Institutes of Health
HCC95*	NSCLC	Squamous cell	Adherent	A gift from Michael Seckl, Imperial College, London
H226	NSCLC	Squamous cell; mesothelioma	Adherent	DCTD tumour repository, U.S. National Institutes of Health
N417	SCLC	Transformed into large cell carcinoma *ex vivo*	Suspension	A gift from Peter Szlosarek, Barts Cancer Institute, Queen Mary University of London
H510	SCLC		Suspension/ adherent	HPA culture collections, Public Health England
H209	SCLC		Suspension	American Type Culture Collection (ATCC)
HCC33*	SSLC		Suspension	Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures
H69*	SCLC		Suspension	American Type Culture Collection (ATCC)
H2171*	SCLC		Suspension	American Type Culture Collection (ATCC)

Table 2-3 Properties of lung cancer cell lines used in this study. NSCLC = non-small cell lung cancer, SCLC = small cell lung cancer. * Cell lines included in microarray panel.

2.3.1 siRNA transfections

Cells were reverse transfected by adding 100µl of siRNA-lipofectamine complex to each well of a 24-well plate and seeding 400µl of cell suspension on top of this, or by adding 20µl of siRNA-lipofectamine complex to each well of a 96-well plate and seeding 100µl of cell suspension. siRNA-lipofectamine complexes were formed as follows: siRNA (Hs_ENG_5 FlexiTube siRNA, catalogue number SI02663024, Qiagen) was diluted in culture medium without FBS for a final concentration of 10nM per well. 2.52µl of Lipofectamine® RNAiMax (Invitrogen) per well of a 24-well plate or 0.12µl per well of a 96-well plate was added and the complex solution was incubated at room temperature for 20 minutes. Cells were added to each well on top of complexes for a total of $6x10^4$ cells per well of a 24-well plate, or $1.8x10^4$ per well of a 96-well plate. Fresh medium was added to cells the day after transfection.

2.3.2 Vector amplification

A pcDNA3.1 plasmid vector alone and a pcDNA3.1 plasmid containing the Endoglin long isoform cDNA (both gifts from Professor Clare Isacke, Institute of Cancer Research, London), and a pcDNA3.1 plasmid vector (Figure 2-4) containing the Vik-1 cDNA (a gift from Nadine Varin-Blank, Departement d'Hematologie, Institut Cochin, Paris, France) were transformed into DH5α™ competent bacterial cells (Invitrogen) as per the manufacturer's protocol. Cells were thawed on ice and 10ng of DNA was added to 50µl of cells for each transformation. Cells were incubated with DNA for 30 minutes, and then heat shocked for 20 seconds at 42°C in a water bath and placed immediately on ice for 2 minutes. 950µl of warm S.O.C (super optimal broth with catabolite repression) medium (Invitrogen) was added and cells incubated at 37°C for 1 hour, shaking, at 225rpm. Volumes from 10-100µl were plated onto Luria-Bertani

(LB) agar containing ampicillin at 100µg/ml and 20µl of x-galactosidase at 40mg/ml and incubated overnight at 37°C. Multiple single colonies were picked from plates and plasmids were purified from these using the QIAprep Spin Miniprep Kit (Qiagen) as per manufacturer's instructions. This kit contains endotoxin removal buffers and yields DNA suitable for immediate transfection. Presence of plasmids plus or minus the Endoglin cDNA or VIK-1 cDNA were verified by running on 1% agarose gel with ethidium bromide at 0.1µg/ml (Sigma) alongside a 1kb ladder. Selected colonies were then grown up in LB broth at 37°C overnight, shaking. Plasmids were purified using the Qiagen Plasmid Plus Maxi Kit (Qiagen) as per manufacturer's instructions. All sequences were verified by sequencing (GATC Biotech).

2.3.3 Plasmid transfections

For ectopic expression of Endoglin and VIK-1, cells were transfected with purified pcDNA3.1 alone or pcDNA3.1 containing Endoglin or VIK-1 cDNA. See figure 2-3 for plasmid diagram. Cells were seeded into 6-well culture plates (Corning) so as to be 80-90% confluent on the day of transfection. 5µg of DNA was diluted in 250µl of culture medium without FBS. 5µl of X-tremeGENE HP DNA Transfection Reagent (Roche) was added at room temperature for 30 minutes. The transfection complex was added directly to cells in a drop-wise manner while gently swirling the plate. Fresh culture medium was added to cells the day after transfection for 8 hours, then medium containing G418 was added to select for transfected cells. G418 concentration was determined for each cell line individually using a kill curve.

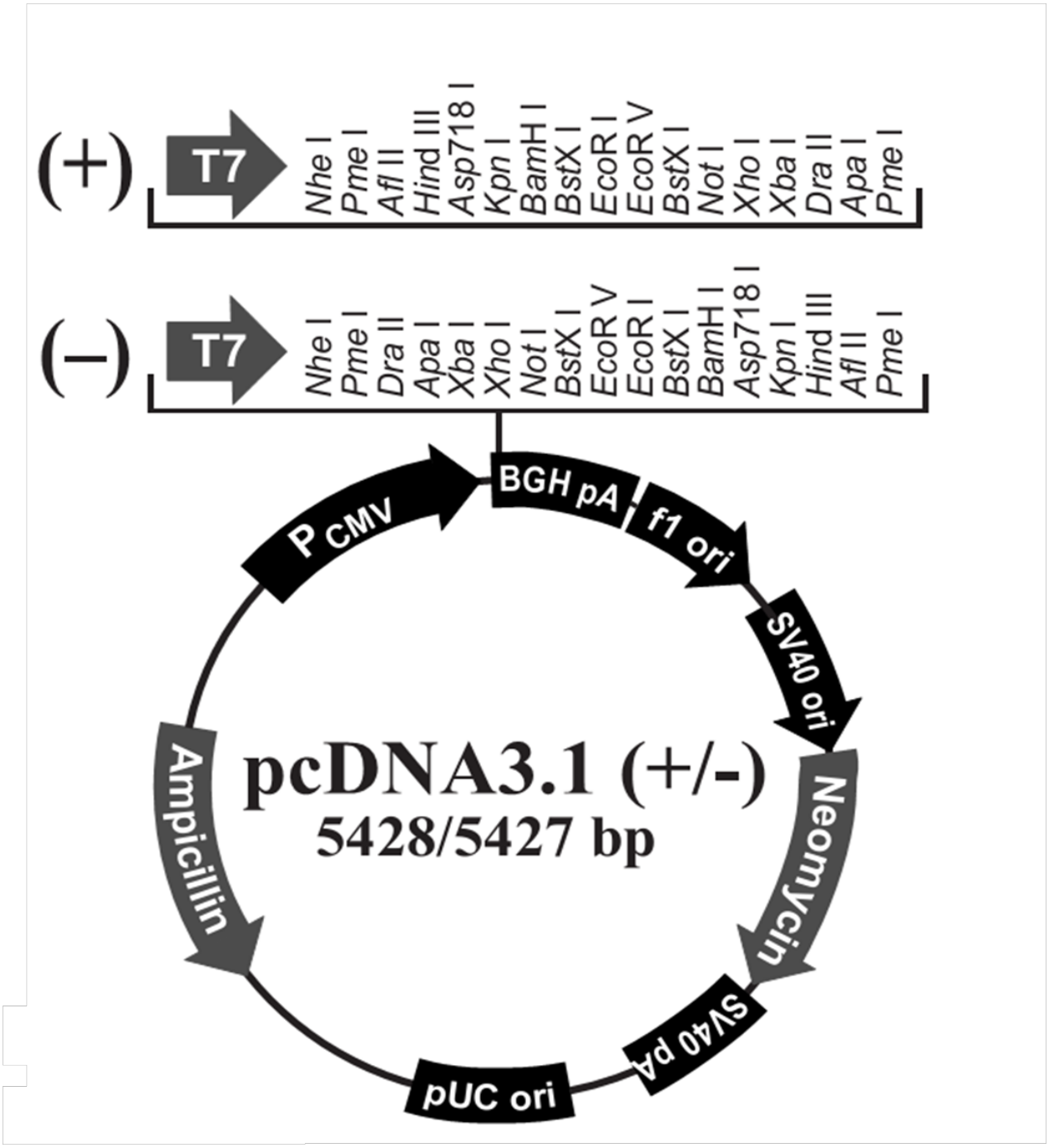


Figure 2-4 Diagramatic representation of the pcDNA3.1 plasmid (Life Technologies). The plasmid contains a cytomegalovirus (CMV) promoter for high level, stable expression, multiple cloning sites, a neomycin cassette allowing for selection using G418, and an ampicillin resistance gene for selection in DH5α bacterial cells.

2.3.4 Cell proliferation assays

Cell viability and proliferation were measured using the MTT (3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide) assay. MTT (NBS Biologicals) is reduced by metabolically active cells, resulting in formation of intracellular purple formazan stock solution, which can be solubilised and measured spectrophotometrically. Cells were allowed to attach overnight and the next day was treated as day zero. TGF-β was added where indicated at 10ng/ml. MTT assays were carried out on day 0, 1, 3 and 5. MTT solution (stock 2mg/ml in culture medium) was diluted 1:3 in culture medium and added to cells for 2-4 hours until a purple precipitate was visible. Culture medium was carefully removed from cells and 100µl of DMSO added to each well, to solubilise the precipitate. Plates were incubated in the dark, shaking gently, for 30 minutes and absorbance read on the spectrophotometer (Biotek Synergy HT) at 570nm and 650nm. The 650nm reading is subtracted from the 570nm reading as a background correction. Data are represented as background corrected raw values, then each sample is normalised to its respective day zero baseline value.

2.3.5 Scratch wound healing assay

Cells were seeded in a 24-well plate (Corning) to be 90-100% confluent on the day of the assay. Cells were serum-starved overnight and scratched across the diameter of the well with a sterile pipette tip in the morning. Cells were washed twice with PBS to remove cell debris and culture medium with or without TGF-β1 (10ng/ml) was added. Scratch wounds were photographed immediately and after 16-30 hours depending on speed of migration of each cell line. Distance migrated was quantified using ImageJ software, version 1.45 (Abramoff *et al.* 2004) as follows: the boundary perimeter of the monolayer on each side of the scratch was drawn manually per image and the area within was measured by the software. The area of the gap at indicated time points was

subtracted from the area at time zero for each individual well of the plate. Scratches were performed in triplicate per sample for each experiment. These were averaged and normalised to the untreated empty vector control, and experiments were reproduced as biological replicates.

2.3.6 Tumour spheroid-based invasion assay

For spheroid generation, 200µl of cell suspension were seeded into ultra low attachment 96-well round-bottomed plates (Corning) and incubated for 4 days at 37°C, 5% CO_2, 95% humidity. On day 4, 70µl of medium was removed without disrupting the sphere and carefully replaced with 70µl of Matrigel (BD Biosciences), with or without TGF-β (final concentration 10ng/ml per well) (Peprotech). Since Matrigel contains small variable amounts of TGF-β (average 2.7ng/ml), the TGF-β inhibitor SB 431542 (Sigma) was added to TGF-β-negative wells at a final concentration 20µM per well. After 4 days, spheroids were photographed at 40x magnification using an inverted microscope (Olympus CKX-41), equipped with a digital camera (Micropix).

For seeding of siRNA transfected cells into the tumour spheroid-based invasion assay, cells were first reverse-transfected in a 24-well plate as described above. The next day, cells were trypsinised, counted and seeded into a 96-well, U-bottom ultra low attachment (ULA) plate as described below for the invasion assay.

2.3.7 Transwell invasion assay

The BD BioCoat™ Tumor Invasion System (BD Biosciences) is a transwell assay in which inserts in each well are coated with Matrigel and a FluoroBlok™ membrane. This allows for repeated measurement of fluorescently-labelled cells which have crossed to the other side of the insert, in a non-destructive manner. Cells were stained

in situ with BD DilC_{12}(3) fluorescent dye (BD Biosciences) at 10µg/ml in normal serum-containing medium for 1 hour at 37°C. Cells were then trypsinised, resuspended in serum-free medium and counted. For TGF-β wells, TGF-β1 was added for a final concentration of 10ng/µl. For TGF-β-negative wells, the TGF-β inhibitor SB 431542 (Sigma) was added to cell suspensions at a final concentration of 20µM per well (inhibitor could not be added directly into Matrigel as inserts were already Matrigel-coated). Cells were seeded into transwell inserts at a density of $3x10^4$ per well and medium with 20% FBS as a chemo attractant was added to each basal chamber (below the insert). Fluorescence of invaded cells was read on a plate-reader (BioTek Synergy HT) at 549/565nm (Ex/Em) at time zero and after incubation at 37°C for 24 hours. Data are represented as fluorescence at 24 hours – fluorescence at time zero for each well. Technical replicates (n=4) are then averaged and normalised to the untreated empty vector control.

2.3.8 EMT gene expression analysis

To determine expression levels of EMT-related genes, cell line cDNA was analyzed using a Taqman Low Density Array (TLDA, Applied Biosystems) at Genentech (San Francisco). Arrays were designed to include 42 genes related to EMT. The 42 genes were comprised of 20 genes whose expression was previously reported to be associated with an epithelial-like or mesenchymal-like phenotype in non-small cell lung cancer cell lines (Walter et al., Clinical Cancer Research 2012) plus 22 additional candidate EMT-related genes showing evidence of differential methylation between epithelial and mesenchymal-like non-small cell lung cancer cell lines. 200ng of total cDNA was amplified using Taqman Universal PCR MasterMix (Applied Biosystems) on a TLDA card using the following thermocycling conditions: 1 cycle of 50˚C for 2 min, 1 cycle of 95˚C for 10 min, and 40 cycles of 95˚C for 15 sec followed by 60˚C

for 1 min. All samples were assayed in triplicate. Two custom-designed reference genes, AL-1377271 and VPS-33B, and a commercial reference gene assay, 18S (Hs99999901_s1, Applied Biosystems) were also included in the TLDA panel. A mean of the Ct values for the three reference genes was calculated for each sample, and expression levels of EMT target genes were determined using the $2^{-\Delta C_T}$ method as described earlier. dCt values were used to cluster cell lines according to EMT gene expression levels using Cluster v.3.0 and Treeview v.1.60 software (http://rana.lbl.gov/EisenSoftware.htm).

2.4 Primary tissues

2.4.1 Cohort 1 (Cuneo)

134 lung cancer tissues were obtained as formalin-fixed, paraffin-embedded (FFPE) specimens with informed patient consent and local ethics committee approval from the S. Croce General Hospital, Cuneo, Italy. Tissues were from diagnostic biopsies or surgical resection samples of NSCLC patients who underwent surgery between the years 2003-2006. In each case, the diagnosis and presence of adequate tumour representation in the specimen was confirmed by histopathological analysis. Clinical and pathological data included age at diagnosis, tumour grade, type, stage, smoking status and overall survival. All patients were allocated an ID number and were completely anonymised. 20 tumour-adjacent tissues and 10 biopsy tissues from cancer-free emphysema patients were also obtained as control tissues, and were processed exactly as above.

2.4.2 Cohort 2 (London)

234 lung cancer tissues from biopsies or surgical resection samples were obtained as FFPE sections as described above, from Hammersmith Hospital and Imperial College

Healthcare NHS Trust, London. Samples were from NSCLC patients operated on between the years 2004-2011. Clinical and pathological data were available as above and samples were all anonymised. 20 tumour-adjacent tissues processed in the same way were obtained as controls.

2.5 Statistical Analysis

SPSS software was used for all primary tissue analyses. Normality of methylation data for both primary tissue cohorts was tested using the Shapiro-Wilk test (Shapiro and Wilk 1965; Ghasemi and Zahediasl 2012). Based on this, non-parametric tests were used for all primary tissue analyses.

Kaplan-Meier analysis was used to compare survival of methylated and unmethylated populations. Differences between curves are reported as either log-rank or generalised Wilcoxon p-values. Cox regression analyses were used to generate hazard ratios, which are reported with 95% confidence intervals and log-rank p values. For *in silico* analysis of primary tissues, Spearman's rank correlation coefficient was used to assess correlation between methylation and expression. Correlation is reported as the correlation coefficient value (R) with 95% confidence interval and p value.

All statistical analyses on cell line work were carried out on the software program Prism 6 (GraphPad). Differences in functional traits of cell lines are reported as p-values generated by 1 or 2-way analysis of variance (ANOVA), as appropriate.

3 Methylation and expression microarrays

3.1 Introduction

DNA methylation and epigenetic gene silencing are stable events, however they can be reversed by inhibiting the enzymes that are responsible for maintaining the transcriptionally inactive state. 5-aza-2'-deoxycytidine (AZA) and Trichostatin A (TSA) can be used for this purpose. AZA inhibits the activity of DNA methyltransferase which maintains methylation patterns on DNA through sequential cell divisions. TSA inhibits histone deacetylase, which removes acetyl groups from lysine residues on histones, thereby allowing the histone to bind and package DNA so that it is inaccessible to transcription factors (Cameron *et al.* 1999). With the development of high throughput microarray analysis technologies, this ability to globally re-activate epigenetically silenced genes represents a powerful tool to identify novel epigenetically silenced, potential tumour suppressor genes in cancer.

Genes which become re-activated by AZA and/or TSA may do so directly due to reversal of epigenetic silencing of that gene, but re-activation can also occur indirectly. For example, re-activation of gene A which is an inducer of gene B would result in a possible misclassification of gene B as epigenetically regulated. For this reason, combination of expression analysis with genome-wide methylation analysis provides a very robust, multi-faceted approach for identifying true epigenetically silenced genes across the genome. The efficacy of this integrated array approach has been proven by a number of studies in different cancers, such as neuroblastoma (Caren *et al.* 2011), hepatocellular carcinoma (Matsumura *et al.* 2012) and colorectal cancer (Yagi *et al.* 2010).

Although DNA methylation and histone deacetylation can act synergistically to maintain gene silencing, DNA methylation is the dominant process necessary for this epigenetic inactivation (Cameron *et al.* 1999). Therefore we combined treatment of cell lines with AZA followed by expression array, with genome-wide methylation array analysis to identify novel epigenetically regulated genes in lung cancer.

The Illumina Human HT-12 v4 Expression BeadChip contains probes for more than 47,000 transcripts across the genome, covering NCBI RefSeq (National Center for Biotechnology Information Reference Sequence) Release 38 annotated genes. The Illumina Human 450K Methylation BeadChip has 99% coverage of genome-wide CpG sites. Coverage of the CpG island region extends to shore areas (defined as 2kb up and downstream of the island) and shelf areas (2kb up and downstream of shores). There are two types of probe on this array, type 1 Infinium probes which have 2 beads per target CpG site (for methylated and unmethylated DNA) sharing a single colour channel, while type II probe design requires just a single bead per CpG target, with two separate colour channels (Bibikova *et al.* 2011).

The aim of this study was to carry out a comprehensive genome-wide screen to identify epigenetically silenced tumour suppressor genes in lung cancer. To this end, an experimental strategy was designed which would be highly informative, which would maximise discovery and generate novel, functionally relevant targets for investigation. An outline of this experimental design is shown in Figure 3-1.

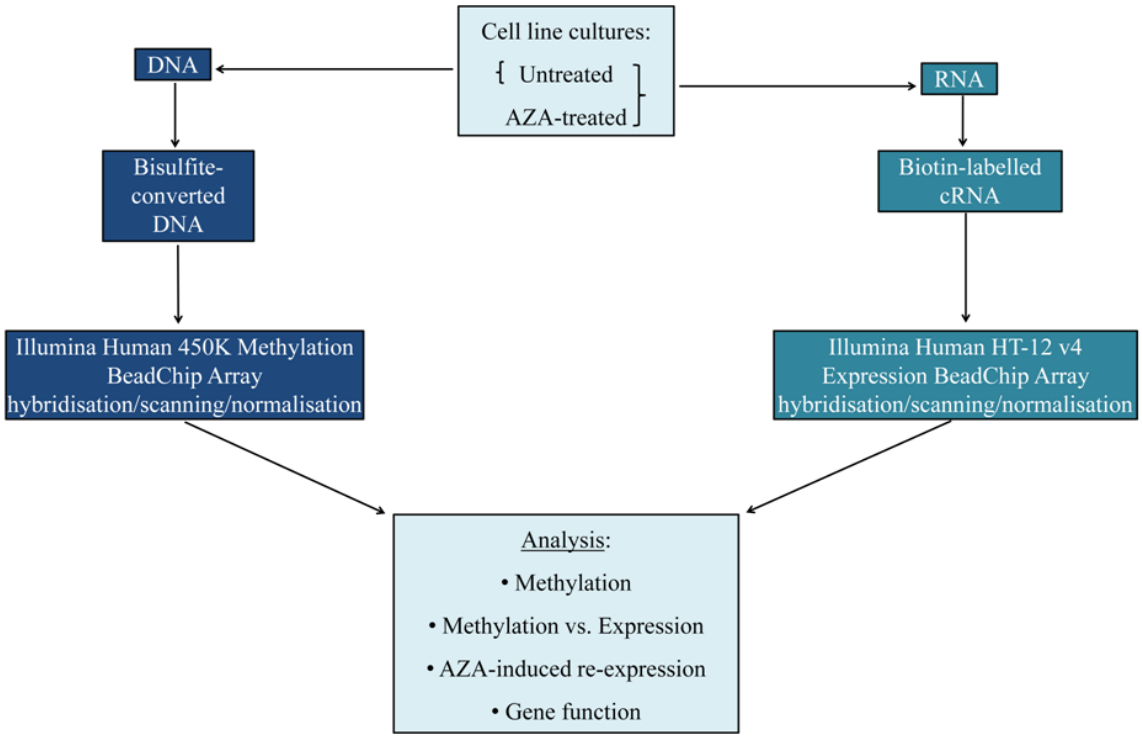


Figure 3-1 Workflow for microarray experiment showing integrated methylation and expression analysis.

3.2 Results

Experimental duplicates of the cell line A549 were first analysed by methylation array as a pilot study to check variability between replicates. Using linear regression analysis, there was a very high concordance in genome-wide methylation values between replicates, with an r^2 value of 0.9928 (Figure 3-2). Therefore all remaining cell lines were analysed as a single sample.

Prior to preparing AZA-treated samples for microarrays, dose-response assays were performed on all cell lines (data not shown) using MTT cell viability assays to ensure that the treatment resulted in demethylation without significant toxicity to the cells. Demethylation was confirmed by pyrosequencing assays of the commonly methylated genes OPCML and CMTM5.

Several criteria were used to generate a list of candidate genes from microarray data. Firstly, only genes which contained a CpG island were selected. From these, genes which were found to be highly methylated within the island region in a proportion of the cell line panel were shortlisted. Next, those for which high methylation corresponded with low baseline expression levels and/or re-expression after AZA treatment were shortlisted further. Finally, genes from this list which were of interest functionally were selected as candidate genes for validation and further investigation. Examples of candidate genes are shown in Table 3-1 and two of these were prioritised for validation and further investigation.

One of the genes selected from these criteria is Endoglin (CD105). It is located on chromosome 9q34.11 and contains a small CpG island of 238bp (24 CpG dinucleotides) located within the first exon, which overlaps the coding region start site. A single array probe mapped to this CpG island, and displayed high methylation

values for more than half of the cell line panel (Figure 3-3). These highly methylated cell lines showed minimal basal expression of the gene, while cell lines with lower methylation values showed higher expression levels, e.g. EKVX, H460 and Hop92 have beta values <0.5 and expression intensity >200 (Figure 3-4). Beta values represent methylation levels and range from 0-1, 0 representing the lowest methylation value, 1 the highest. Endoglin mRNA expression was up-regulated by 1.2 fold or greater following AZA treatment in 6 cell lines (Figure 3-4), 5 of which had beta values greater than 0.5. All results from the arrays will be further validated in an expanded panel of cell lines.

The other gene of interest is ZNF655, which is located on chromosome 7 and contains a larger CpG island of just over 1kb (90 CpG sites). This island overlaps the first exon containing the 5'untranslated region (5'UTR). There are 13 array probes located within the island region. From the array, methylation levels appear uniform across the entire island, with two cell lines showing very high methylation levels (Figure 3-5). These cell lines, HCC95 and MOR, had the lowest baseline expression levels of ZNF655 of the whole panel, suggesting that CpG island methylation may be silencing the gene. However, none of the cell line panel showed up-regulation by 1.2 fold or greater of ZNF655 RNA following treatment with AZA (Figure 3-6). Both genes will be described in detail in the following chapters.

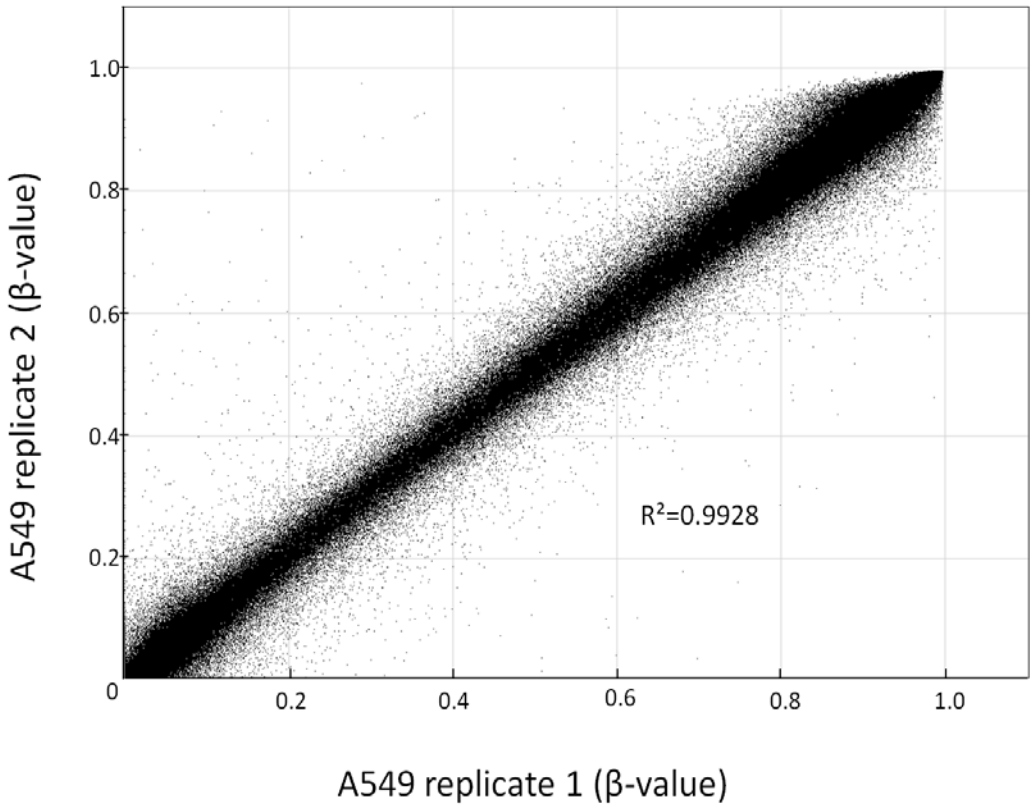


Figure 3-2 Evaluation of reproducibility of methylation array data. The cell line A549 was run in duplicate as a pilot experiment to assess variability of methylation array data (Illumina Human 450K Methylation BeadChip). Linear regression shows highly replicable methylation beta values for duplicate experiments (r^2=0.9928) when cells were treated with the demethylating agent 5-aza-2'-deoxycytidine (AZA) (5µM) for 5 days.

Gene	Cell line methylation (beta values) ≤0.33	0.34-0.66	≥0.67	Re-expression with Aza Y/N
CELSR1	11/13	2/13	-	N
CELSR2	11/13	2/13	-	Y (2/13)
CELSR3	12/13	-	1/13	Y (1/13)
DAB2	13/13	-	-	Y (6/13)
ENDOGLIN	4/13	3/13	6/13	Y (6/13)
DUSP1	-	13/13	-	Y (8/13)
DUSP4	11/13	1/13	1/13	Y (1/13)
DUSP5	11/13	2/13	-	Y (12/13)
IGFBP7	7/13	4/13	2/13	Y (6/13)
LRIG1	13/13	-	-	Y (3/13)
SMAD3	13/13	-	-	Y (5/13)
TGFBI	8/13	2/13	3/13	Y (7/13)
ZNF655	11/13	-	2/13	N

Table 3-1 Candidate gene targets of epigenetic silencing in lung cancer cell lines. Data refer to the number of cell lines out of a total of 13 with beta values within ranges as indicated, and the number showing re-expression upon treatment with the demethylating agent AZA. Y=yes, N=no. Beta values refer to methylation levels according to the Illumina Human 450K Methylation BeadChip, calculated as the ratio of methylated signal to total signal. For each gene, re-expression was defined as an increase of ≥1.2 fold in expression intensity (Illumina Human HT-12 v4 Expression BeadChip) following treatment with AZA.

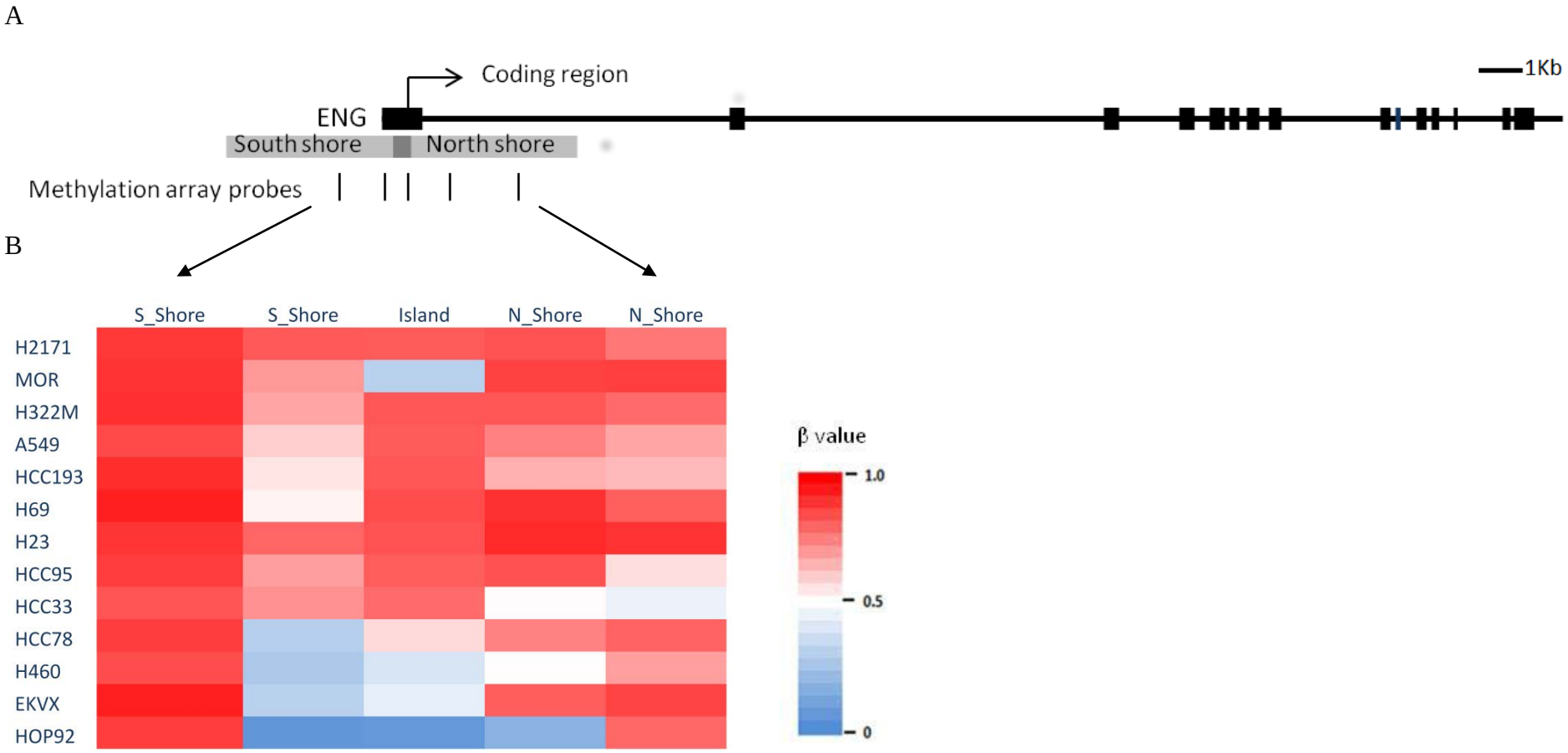
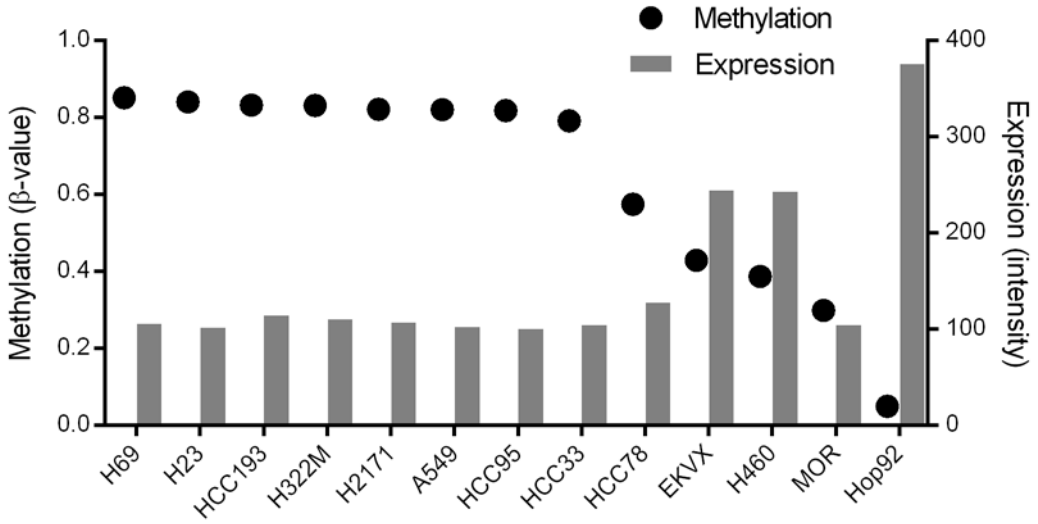


Figure 3-3 Microarray analysis of Endoglin methylation in NSCLC cell lines. (A) Diagrammatic representation of the Endoglin gene with location of methylation array probes. Thick black horizontal lines represent exons. The CpG island is shown in dark grey with the shore areas in light grey. Location of methylation array probes which target a single CpG dinucleotide are shown underneath as vertical black lines. (B) Heat map representation of methylation levels of cell lines within the CpG island and shore areas. Heat map shows beta values, calculated as the ratio of methylated signal to total signal.

A



B

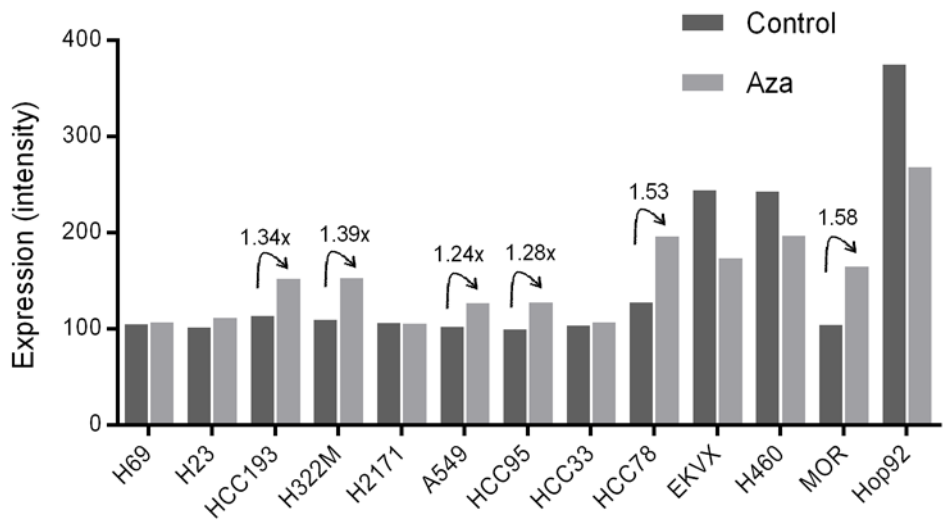


Figure 3-4 Microarray analysis of Endoglin expression in NSCLC cell lines. (A) Baseline Endoglin expression values from the Illumina Human HT-12 v4 Expression BeadChip. Cell lines are arranged in order of decreasing methylation (Illumina Human 450K Methylation BeadChip), represented by black dots. Methylation values refer to beta value of the single probe within the Endoglin CpG island. Expression values are intensity values post-quantile normalisation, shown as the average of biological duplicates for each sample. (B) Microarray expression values for Endoglin in lung cancer cell lines before and after treatment with the demethylating agent 5-aza-2'-deoxycytidine (AZA) (5μM). Arrows indicate cell lines up-regulated by 1.2-fold or greater by AZA treatment, numbers above arrows are the exact fold change.

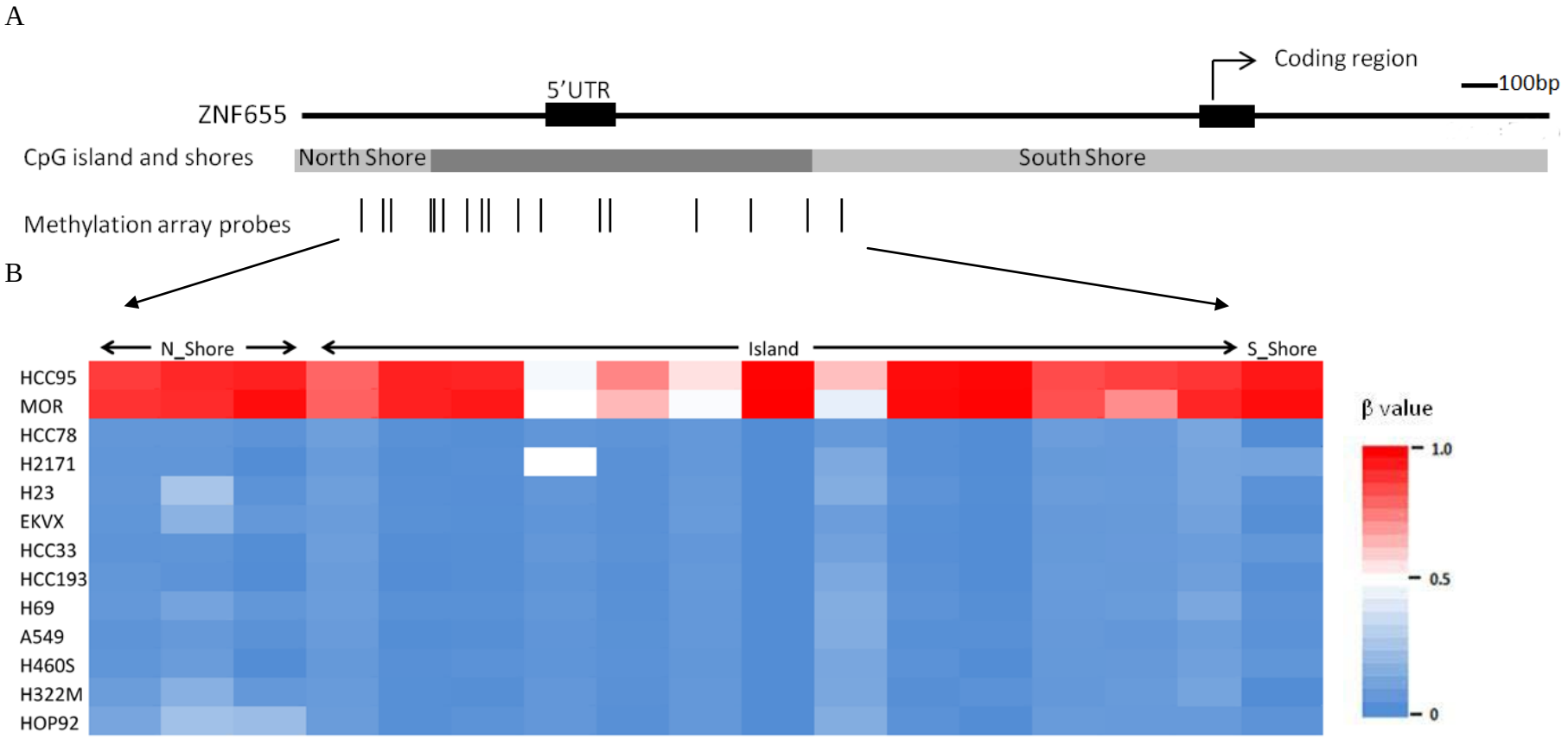
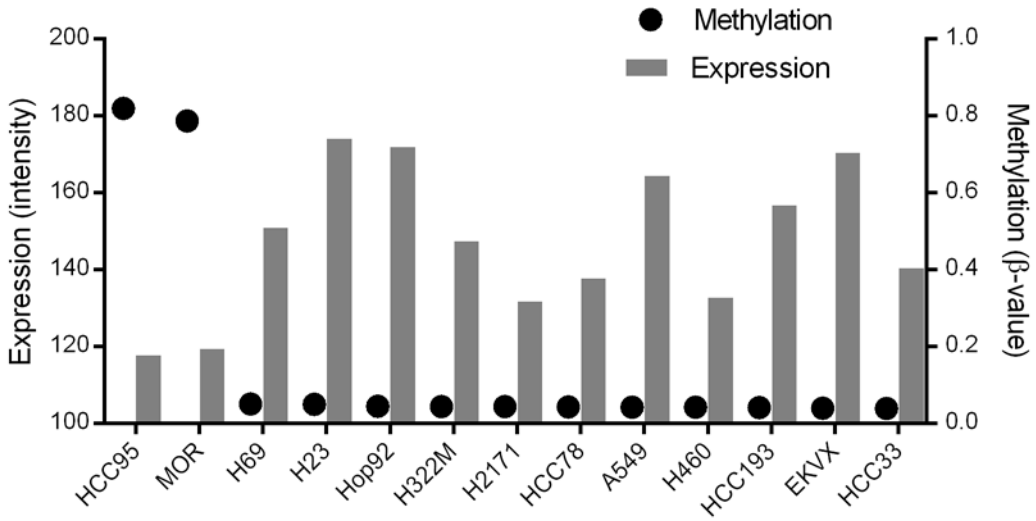


Figure 3-5 Microarray analysis of ZNF655 methylation in NSCLC cell lines. (A) Diagramatic representation of the VIK-1 gene with location of methylation array probes. Thick black horizontal lines represent exons, thin black horizontal lines are introns. The CpG island is shown in dark grey with the shore areas in light grey. Location of methylation array probes which target a single CpG dinucleotide are shown underneath as vertical black lines. (B) Heat map representation of methylation levels of cell lines within the ZNF655 CpG island and shore areas. Map shows beta values, calculated as the ratio of methylated signal to total signal.

A



B

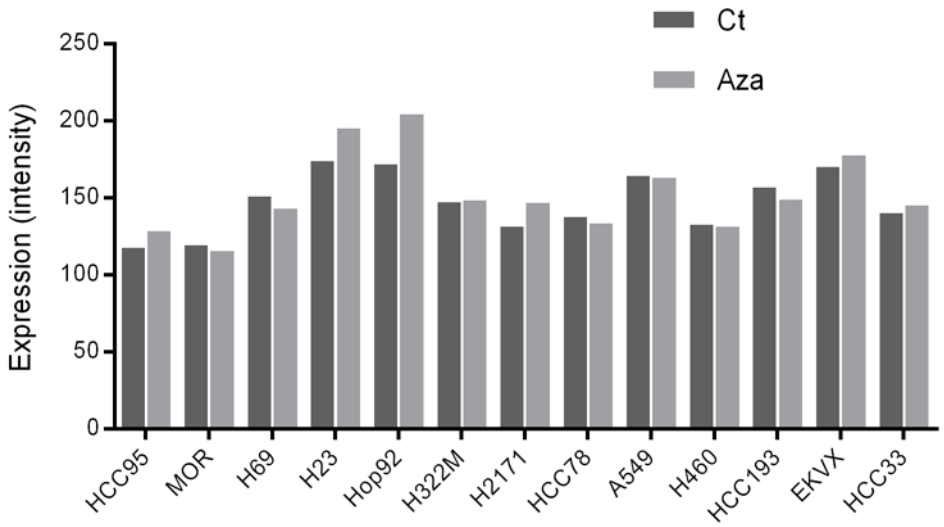


Figure 3-6 Microarray analysis of ZNF655 expression in NSCLC cell lines. (A) Baseline ZNF655 expression values from the Illumina Human HT-12 v4 Expression BeadChip. Cell lines are arranged in order of decreasing methylation (Illumina Human 450K Methylation BeadChip), represented by black dots. Methylation values for each sample refer to average beta values of array probes within the ZNF655 CpG island. Expression values are intensity values post-quantile normalisation, shown as the average of biological duplicates for each sample. (B) Microarray expression values for Endoglin in lung cancer cell lines before and after AZA treatment, panel arranged by decreasing levels of methylation.

3.3 Discussion

This screening approach has yielded a multitude of interesting candidate genes, and further investigation of these is ongoing. The experimental design has allowed for maximum generation of data from a panel of cell lines encompassing squamous cell carcinoma (HCC95), large cell carcinoma (Hop92, H460), adenocarcinoma (MOR, H23, H322M, HCC78, A549, HCC193, and EKVX) and small cell lung cancer (H69, H2171, and HCC33). Due to budget issues, a large sample size came at the cost of additional replicates. As this is an initial screening panel we completed biological duplicates, anticipating further validation and clarification of all candidate genes identified. Experimental protocols were designed to anticipate and minimise variability at all stages.

Methylation was analysed as a single array for each sample. Although the high reproducibility of the Illumina 450K methylation arrays has been reported in the literature (Bibikova *et al.* 2011; Wu *et al.* 2013), a pilot analysis was run on experimental replicates of the A549 cell line to assess variability within the specific treatment conditions of this experiment. As excellent concordance between replicates was observed (r^2 =0.9928) (Figure 3-2), we were confident that analysis of a single biological sample would be reproducible.

In contrast to the relative stability of methylation patterns, gene expression levels are inherently dynamic and this introduces a higher level of variability into expression microarrays. Important considerations in reducing batch effects and technical noise are appropriate normalisation of data and adequate sample size (Qiu *et al.* 2013). Quantile normalisation has been applied to all array data in the current experiment. Alternatives such as rank normalisation exist, but the former, which is designed to

give each array in an experiment the same distribution of probe intensities, is the most widely used and effective method (Bolstad *et al.* 2003; Qiu *et al.* 2013). For expression arrays, we opted for a low sample number (biological duplicates). This decision was taken due to cost issues, and due to the fact that full validation will be carried out for all candidate genes identified, thereby reducing the need for further replicates. The parallel analysis of samples on methylation microarrays serves as an additional control which we believe compensates for the lack of additional expression replicates, as well as being highly informative. All experimental samples were processed through array hybridisation and scanning in a single lot to minimise batch effects at this critical stage.

Highlighted here are two of these candidate epigenetically regulated tumour suppressor genes. Of the criteria described above for short-listing, the Endoglin gene satisfies all. It is very interesting from a functional point of view. It is a TGF-β accessory receptor which interacts with multiple members of the TGF-β signalling pathway (Barbara *et al.* 1999). It has been implicated in driving tumour development via its crucial role in angiogenesis, as well as suppressing tumour development via other mechanisms in various cancer types, which will be discussed in detail in chapter 4. While it has been widely studied, it is a protein with a dual capability for tumour progression and tumour suppression, whose role in the context of lung cancer is yet to be revealed. Endoglin was therefore prioritised for validation experiments and further investigation.

While high methylation levels correspond to minimum baseline expression of ZNF655, it is not re-expressed in methylated cell lines following demethylation in our microarray screen. However it is a very novel and interesting candidate. Very little is known of the function of ZNF655, but the protein was first described in 2005 and has

some properties consistent with a potential tumour suppressor function (Houlard *et al.* 2005), which will be discussed in detail in chapter 5. In particular, the inhibitory interaction between ZNF655 and CDK4 described by Houlard *et al.* suggested a possible therapeutic application for patients with ZNF655 methylation by means of CDK4 inhibitors.

Since ZNF655 does not meet the requirement for re-expression with AZA, its place in our list of candidate genes could be debated. Relaxing the selection criteria for a large-scale study such as this will lead to a greater occurrence of false-positives even if it leads to discovery of some epigenetically regulated genes which would have otherwise been missed. While we chose in this case that ZNF655 was worth validating due to its functional relevance, we recognise the risks associated with adapting a less-stringent selection approach. It was decided that ZNF655 should be analysed further to ascertain whether in fact it is epigenetically regulated and if so, to functionally characterise the effect of this regulation in NSCLC.

4 Role of Endoglin as an epigenetically regulated candidate tumour suppressor gene.

4.1 Introduction

4.1.1 Introduction to Endoglin

The Endoglin gene is located on chromosome 9q34.11 (Fernandez-Ruiz *et al.* 1993), and the protein functions as an accessory receptor within the TGF-β signalling pathway (Massague 1998). It was first identified as a predominantly endothelial-expressed trans-membrane protein with an extracellular domain of 561 amino acids (aa), a membrane-spanning region of 25aa and a cytoplasmic tail of 47aa (Gougos and Letarte 1990). A second isoform was then identified, whose structure differs only in that it has a shorter cytoplasmic tail of just 14aa (Bellon *et al.* 1993). These isoforms have become known as S-Endoglin for the shorter isoform and L-Endoglin for the long isoform. L-Endoglin has been far more widely characterised and studied, and is the predominantly expressed isoform. While S-Endoglin is expressed in several tissues including lung, L-Endoglin levels have been found to be higher and more consistent in all tissues analysed (Perez-Gomez *et al.* 2005). For the purposes of this study, we are only investigating L-Endoglin.

Endoglin is a member of the TGF-β signalling pathway. Along with Betaglycan, it belongs to a class of TGF-β receptor known as the accessory receptors. These do not directly participate in intracellular TGF-β signalling, but they influence it by regulating ligand access to the type I and type II receptors (Massague 1998). Tightly controlled levels of TGF-β signalling are essential for normal lung development and for repair in the mature lung. However, deregulation of TGF-β signalling can result in chronic inflammation and fibrosis (Bartram and Speer 2004; Shi *et al.* 2009), and

plays a vital role in cancer progression and metastasis. TGF-β signalling has been shown to have a protective effect in early stage cancers, but can then switch to a pro-invasion, pro-metastatic profile (Wendt *et al.* 2011).

4.1.2 The TGF-β signalling pathway

The transforming growth factor β (TGF-β) family of proteins are a group of structurally related growth factors which include TGF-β1, 2 and 3, as well as subfamilies of bone morphogenic proteins (BMPs), activins, growth and differentiation factor proteins (GDFs). These growth factors regulate a variety of intracellular processes, many of which determine cell fate, such as apoptosis, differentiation and repair. TGF-β signalling is mediated by two families of serine/threonine kinase receptors, named type I and type II receptors, which internalise the signal upon binding of ligand (Massague 1998). Activated TGF-β binds first to the constitutively active type II receptor which then phosphorylates and forms a complex with a type I receptor. A signal is then transduced from the activated type I receptor to within the cell by means of phosphorylation of SMAD proteins. This signal is transduced into the nucleus resulting in direct binding of SMAD complexes to DNA and transcription of distinct subsets of TGF-β-regulated genes (Massague 1998; Bartram and Speer 2004).

4.1.3 Endoglin within the TGF-β pathway

Endoglin binds directly to both ligand and receptor. It binds TGF-β1 and 3, but can only do so by association with the type II receptor within the ligand-receptor complex. It also binds other ligands of the TGF-β family within their receptor complexes, such as BMP-7 (Barbara *et al.* 1999). It has been proposed that Endoglin binds to type I and type II receptors, resulting in an altered phosphorylation state of Endoglin itself and of

the type I receptor. This altered phosphorylation state then allows ligand access to the receptors and transduction of the intracellular signal (Guerrero-Esteo *et al.* 2002).

The role of Endoglin at a molecular level has for the most part, only been examined in endothelial cells, wherein various lines of evidence suggest that Endoglin affects the TGF-β pathway by specifically driving signalling down the ALK1 route. Mutations within Endoglin and ALK1 cause two major types of the same disease, hereditary haemorrhagic telangiectasia, characterised by abnormal vascular development (Abdalla and Letarte 2006). It has been shown in endothelial cells that Endoglin not only directly binds ALK1 but potentiates signalling down this route and hinders signalling via ALK5 (Blanco *et al.* 2005), with the result that TGF-β induced growth arrest is reduced and proliferation is increased (Lebrin *et al.* 2004). In addition to TGF-β1 and 3, Endoglin also acts as an accessory receptor for Activin A, BMP-7 and BMP-2 in the presence of their respective receptors. Moreover, Endoglin has been shown to affect TGF-β signalling by interacting with the scaffolding protein β-arrestin, which results in internalisation of Endoglin within endocytic vesicles (Lee and Blobe 2007), and with another scaffolding protein GIPC (Lee *et al.* 2008). The latter interaction again supports a role for Endoglin in ALK1 signalling, while the former antagonises TGF-β-mediated ERK signalling.

Whether the specific function of Endoglin in potentiating TGF-β/ALK1 signalling is relevant to non-endothelial tissues is largely unknown. A single study has shown that Endoglin also drives signalling down this route in normal human chondrocytes (Finnson *et al.* 2010), but the dynamics of Endoglin signalling in the lung have yet to be revealed.

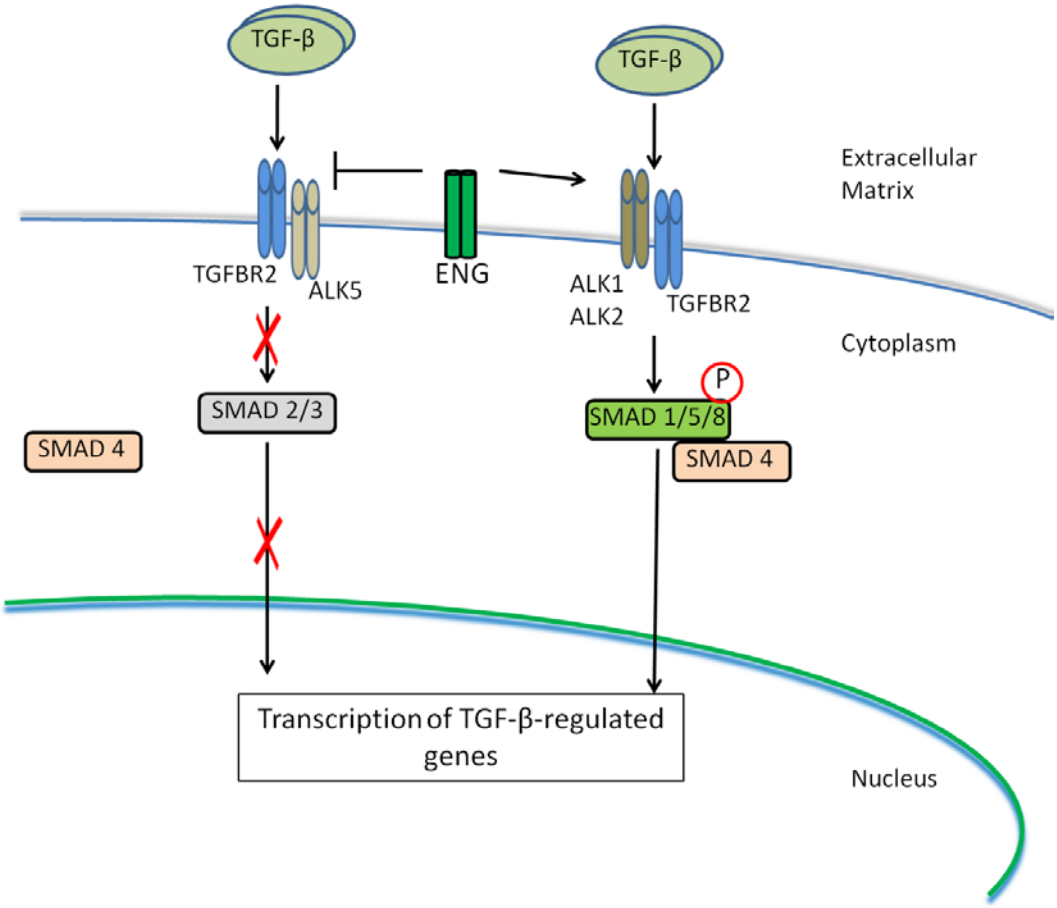


Figure 4-1 Hypothetical model for the role of Endoglin in TGF-β signalling. TGF-β binds to a type II receptor homodimer, which in turn forms a heteromeric complex with a type I receptor. In endothelial cells, Endoglin potentiates signalling via the type I receptor ALK1, while inhibiting signalling down the ALK5 route. ALK1 is an endothelial-specific protein and the dynamics of Endoglin function in non-endothelial cells is largely unknown. However in prostate cancer cells, it has been shown that Endoglin regulates signalling via ALK2, which has the highest sequence homology with ALK1, and leads to phosphorylation of SMAD1 (Craft *et al.*2007)

4.1.4 Endoglin in tumour angiogenesis

Endoglin has long been known to play an important role in angiogenesis. As mentioned above, mutations of Endoglin and/or ALK1 both cause the vascular disease HHT. Endoglin is highly expressed throughout the developing embryonic mouse

vasculature (Jonker and Arthur 2002) and mice lacking Endoglin die by gestational day 11.5 due to defective vascular development (Li *et al.* 1999). This angiogenic role of Endoglin is carried out via TGF-β-dependant mechanisms (Raab *et al.* 1999; Park *et al.* 2013).

Since vascularisation of solid tumours is an essential survival mechanism for cancer cells, Endoglin has been widely studied in terms of its role in tumour angiogenesis. Endoglin protein is a highly selective endothelial marker. It has a distinct advantage over pan-endothelial markers such as CD31 in that Endoglin is specific to newly formed vasculature; it is therefore highly selective for tumour angiogenesis and will not stain normal vasculature as previously published on brain, lung, breast, stomach and colon cancers (Miller *et al.* 1999; Minhajat *et al.* 2006; Takase *et al.* 2010). Quantification of intratumoural microvessel density (MVD) by immunohistochemical staining of Endoglin has been shown to be a good prognostic marker and indicator of metastatic disease. Support for this comes from publications on colon cancer (Saad *et al.* 2004; Romani *et al.* 2006; Bellone *et al.* 2010), breast cancer (Gluz *et al.* 2011), pancreatic cancer (Yoshitomi *et al.* 2008), oesophageal squamous cell carcinoma (Li *et al.* 2007), ovarian carcinoma (Taskiran *et al.* 2006), endometrial cancer (Erdem *et al.* 2006), squamous cell carcinoma of the hypopharynx (Chien *et al.* 2006), prostate cancer (Wikstrom *et al.* 2002) and cervical cancer (Brewer *et al.* 2000; Zijlmans *et al.* 2009).

Endoglin can be cleaved from the membrane and released as a soluble form known as sEndoglin (Li *et al.* 1998). The level of soluble Endoglin in plasma has also been suggested as a useful and easily measured prognostic marker, but support for this is not as consistent as in the case of endothelial Endoglin. Positive evidence for soluble Endoglin as a prognostic marker was found in breast cancer (Li *et al.* 2000) and in a

cohort of colorectal, breast and other solid tumours (Takahashi *et al.* 2001). However one study on a cohort of cervical cancer patients found that while levels of soluble Endoglin were independently associated with invasiveness and recurrence, it was not specific enough for clinical use (Landt *et al.* 2011).

The above evidence has led to much work on development of targeted anti-angiogenic therapy against Endoglin, mostly in the form of monoclonal antibodies (mAbs) against the protein. The development and approval of Bevacizumab, a mAb against the pro-angiogenic molecule vascular endothelial growth factor (VEGF) for use in cancer treatment, has paved the way for other treatments of this type (Welti *et al.* 2013). Endoglin mAbs have been shown to effectively suppress MVD and angiogenesis in mouse models (Uneda *et al.* 2009), primates and humans (Seon *et al.* 2011).

4.1.5 A role for Endoglin in tumour suppression

There is some evidence for expression of Endoglin in non-endothelial tissues, and more evidence of an angiogenesis-independent role for Endoglin in the pathogenesis of cancer in several tissues. Endoglin is expressed to varying degrees in human sarcoma cell lines and lung, kidney, prostate, breast, thyroid, liver and skin carcinoma cell lines, and has been shown to be present in the cytoplasm as well as the membrane (Postiglione *et al.* 2005). Endoglin expression has been found on primary prostate cancer epithelial and stromal cells as well as the tumour vasculature of these tissues. Although expression levels were not high (<25% of cancer cells were immunoreactive in 59% of specimens), it was significantly correlated with disease development (Kassouf *et al.* 2004).

In cancer tissues, expression of Endoglin has been shown to suppress migration and invasion, and conversely, loss of Endoglin leads to an increase in these processes. This

is likely to involve both TGF-β-dependent and TGF-β-independent mechanisms. Endoglin suppression of invasion and migration has been shown independently of TGF-β in prostate cancer (Liu *et al.* 2002), and esophageal squamous cell carcinoma (Wong *et al.* 2008). However, studies investigating Endoglin function in the context of TGF-β have found that presence of the ligand amplifies the above effect, as demonstrated by *in vitro* cell line experiments on extravillous trophoblasts (Mano *et al.* 2011), prostate cancer cells (Craft *et al.* 2007), HER2 positive breast cancer cells (Henry *et al.* 2011), and mouse keratinocyte cells (Perez-Gomez *et al.* 2007). This clearly demonstrates Endoglin has a protective effect in these tissues.

These *in vitro* findings may be clinically important due to the fact that *in vivo*, loss of Endoglin expression has been shown to occur in primary cancer tissues due to promoter methylation (Henry *et al.* 2011; Jin *et al.* 2013) and/or loss of heterozygosity (Wong *et al.* 2008). Loss of Endoglin expression in tumours could therefore be a critical factor in driving invasive behaviour of these cells. This is particularly relevant to lung cancer, in which TGF-β is frequently overexpressed and has been shown to have a pro-invasive effect (Xu *et al.* 2011).

There are a number of possibilities for molecular mechanisms by which this tumour-suppressive effect is brought about, which have been suggested by the studies mentioned above. In their study on prostate cancer mentioned above, Craft *et al.* (2007) looked to the well-documented Endoglin-dependent signalling molecule, ALK1. Upon finding a lack of ALK1 expression in prostate cancer cells, sequence homology searching found ALK2 to be expressed and necessary for Endoglin-mediated suppression of motility via SMAD1. This mechanism of action is supported by the authors of the aforementioned study in tumourigenic mouse keratinocytes, who have shown that Endoglin suppresses activation of SMADs 2 and 3 (Perez-Gomez *et*

al. 2007). This would imply that signalling would be driven down the alternative SMAD1/5/8 route as shown by Craft *et al.*(2007) and as shown repeatedly for endothelial cells, though unfortunately SMADs1/5/8 were not analysed in this study. They did however additionally demonstrate an induction of epithelial to mesenchymal transition (EMT) as a mechanism of acquired invasiveness. Henry *et al.* (2011) found that while Endoglin suppressed the pro-invasive and pro-migratory effects of TGF-β, there was no involvement of SMAD signalling in this breast cancer cell response.

4.1.6 Endoglin in NSCLC

In the context of NSCLC, research on Endoglin has largely been restricted to its potential as an angiogenic marker. With regard to the assessment of tumour angiogenesis, Endoglin expression has been found to be superior to other markers such as VEGF (Medetoglu *et al.* 2010) and CD34 (Tanaka *et al.* 2001). It has also been suggested for use in radiolabelling of angiogenic tumour tissue for radiotherapy treatment of NSCLC (Lee *et al.* 2009).

The research on lung and other cancers described here supports a role for Endoglin as potentially a tumour promoter or a tumour suppressor. Tissue-specificity is clearly a factor in this dual capability; in endothelial cells Endoglin promotes tumour angiogenesis, thereby facilitating growth. In cancer cells, when expressed, it has a protective role, but loss of expression can increase invasive potential of these cells. Our microarray screen has suggested a possible loss of Endoglin expression by means of epigenetic inactivation. Methylation of lung cancer cell lines and primary tissues has been reported on one previous occasion, although a total of only 16 primary tissues were analysed, and no patient outcome or functional data were presented (Dammann *et al.* 2005). Given what we know of the function of Endoglin, our hypothesis is that epigenetic silencing of this gene may significantly affect behaviour

of NSCLC tumours and may be a prognostic marker is these cancers. The aim of this study therefore was to confirm the existence of epigenetic silencing of Endoglin in cell lines and in two substantial patient cohorts, as well as demonstrating a functional or prognostic effect on NSCLC.

4.2 Results

4.2.1 Validation of Endoglin as an epigenetically regulated gene in lung cancer cell lines

Methylation of Endoglin was confirmed in an extended panel of 18 cell lines using methylation-specific PCR (MSP) and pyrosequencing (Figure 4-3). Methylation status as measured by MSP and pyrosequencing correlated for all cell lines. Primers for MSP were designed to cover at least a single CpG site within the CpG island and were designed to target methylated sequence and unmethylated sequence of bisulfite-modified DNA, respectively. Pyrosequencing primers were designed on a region of bisulfite-modified DNA within the CpG island in which there were no CpG dinucleotides. This sequence would therefore be amplified irrespective of methylation status. Pyrosequencing primers were also designed to generate as small an amplicon as possible so that the assays would be compatible with FFPE DNA which may be fragmented.

DNA from the 13 cell lines with methylation above 50% by pyrosequencing (H69, A549, N417, H226, H510, H23, H2171, HCC193, MOR, H322M, HCC95, HCC33, and H209) was only amplified by primers specific for methylated DNA. Those with methylation between 10 and 50% (HCC78, H460 and EKVX), were amplified by both primer sets and cell lines with methylation levels below 10% (Hop62, Hop92) were only amplified by primers specific for unmethylated DNA . Methylation status as measured by pyrosequencing and MSP correlated with methylation array results for 12/13 cell lines (Figure 4-2), as shown by linear regression analysis (r square=0.6714, p=0.006). The only outlier was the cell line MOR, which in our validation assays were confirmed to be more highly methylated than in the array (Figure 4-2, Figure 4-3).

Expression of Endoglin in our extended panel of cell lines was measured by qPCR quantification of Endoglin mRNA and by Western blot quantification of the 98kDa Endoglin protein (Figure 4-4). All highly methylated cell lines have little or no detectable Endoglin mRNA and no detectable Endoglin protein, while those with methylation <50% are all expressing the gene, to varying degrees. Protein levels reflect mRNA levels, and both are silenced in cell lines with methylation above 50% (Figure 4-4).

To validate epigenetics as a direct mechanism of Endoglin transcriptional silencing, we treated the highly methylated cell lines MOR, H322M and HCC95 with the demethylating agent 5-aza-2'-deoxycytidine (AZA) and the HDAC inhibitor Trichostatin A (TSA). All three cell lines show ≥20-fold re-expression of Endoglin mRNA by RT-qPCR. Cells showing up-regulation of Endoglin mRNA also show increased protein expression. Re-expression was seen upon treatment with AZA and AZA combined with TSA, although no synergistic effect of the combined treatment was observed. Treatment with TSA alone had no effect on Endoglin expression. To demonstrate that the effect of AZA on Endoglin expression was due directly to demethylation of the gene and not any other non-specific effects of the treatment, the unmethylated cell line Hop62 was used as a control. There was no increase in Endoglin expression following treatment with AZA and/or TSA in this cell line (Figure 4-5). Cumulatively, these results point towards methylation-dependent epigenetic regulation as a common mechanism of Endoglin silencing in this panel of lung cancer cell lines.

A

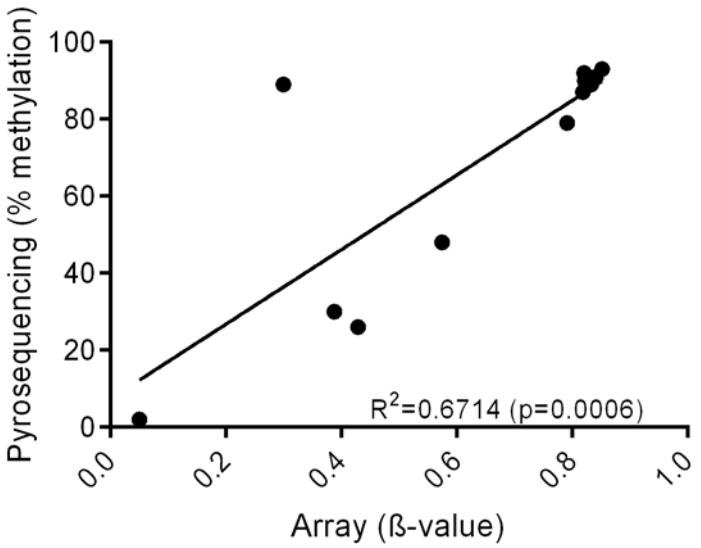


Figure 4-2 Methylation of Endoglin in cell lines as measured by pyrosequencing and microarray (Illumina Human 450K Methylation BeadChip). The x-axis shows beta values for each cell line of the single array probe within the Endoglin CpG island. Pyrosequencing methylation values on the y-axis are the mean of biological triplicates for each cell line. Each one is calculated as the average methylation across the 13 CpG sites covered by the pyrosequencing primers. Using linear regression analysis, the R^2 value=0.6714 (p=0.0006).

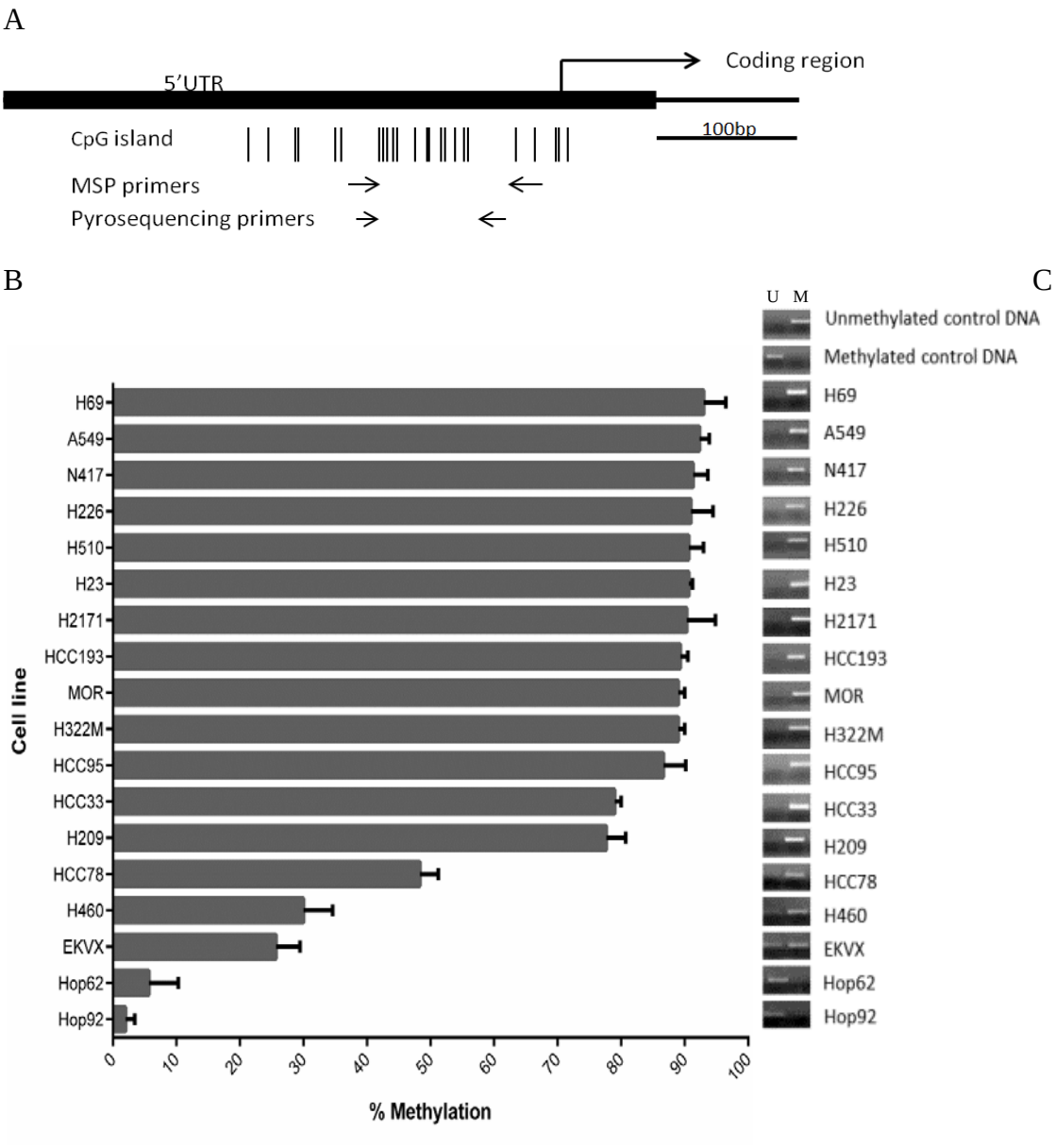
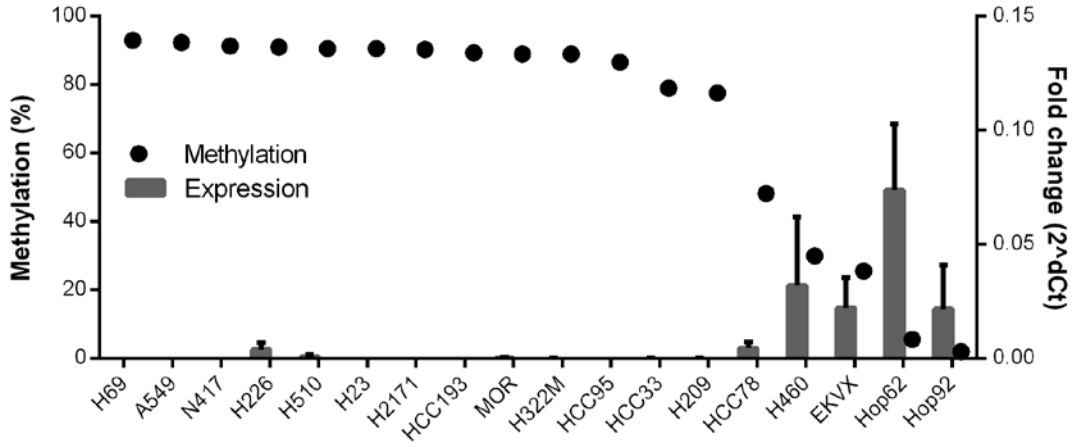


Figure 4-3 Methylation analysis of the Endoglin promoter region. (A) Diagrammatic representation of the Endoglin gene with CpG island and primer locations. Individual CpG sites within the island are represented by vertical lines. The island overlaps the 5'UTR of the gene and the beginning of the coding region. Forward and reverse primers are represented by horizontal arrows. (B) Endoglin methylation as measured by pyrosequencing. Methylation values are calculated as the average methylation across the 13 CpG sites covered by the pyrosequencing primers. Histogram bars show mean of biological triplicates with standard deviation. (C) Endoglin methylation measured by MSP. M and U refer to PCR reactions carried out with primers targeting methylated and unmethylated sequence, respectively. Results are shown as PCR products of both reactions on ethidium bromide-stained agarose gel for each cell line.

A



B

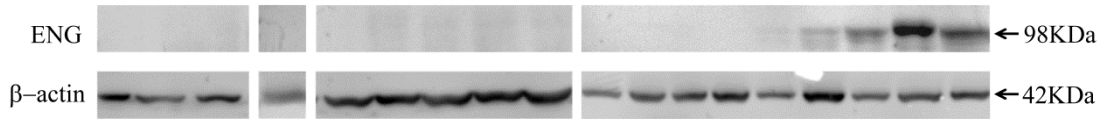
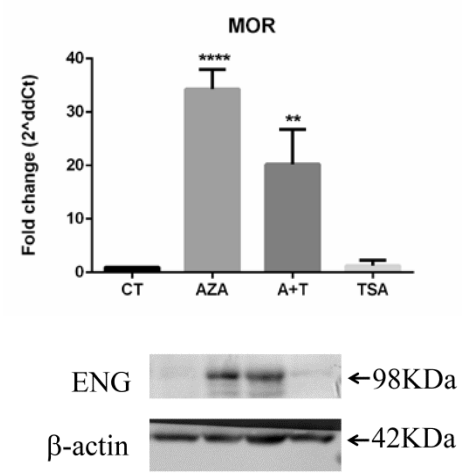
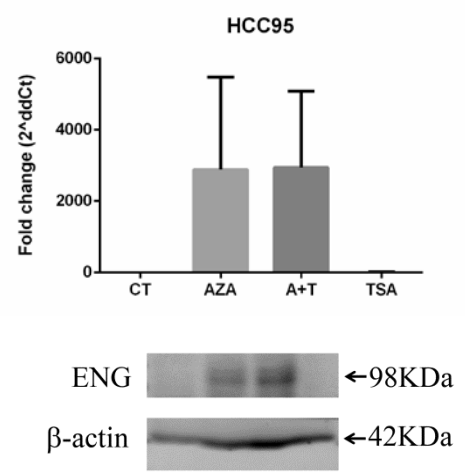


Figure 4-4 Endoglin methylation correlates with reduced mRNA and protein expression. (A) Endoglin mRNA was measured by qPCR. RNA expression is shown as the mean 2^dCt value of 3 biological replicates with standard deviation for each cell line (right Y axis). Cell lines are shown in order of decreasing methylation, as measured by pyrosequencing and represented by black dots (left Y axis). Methylation values shown are mean of 3 biological replicates. (B) Endoglin protein was measured by western blot and correlates with mRNA levels. Blot shows cell lines in same order as histogram above, with β-actin as a reference for each sample.

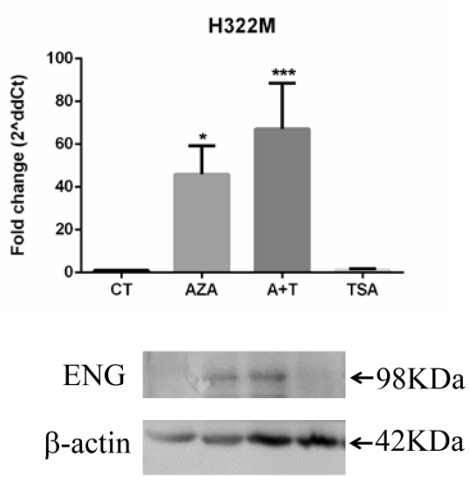
A



B



C



D

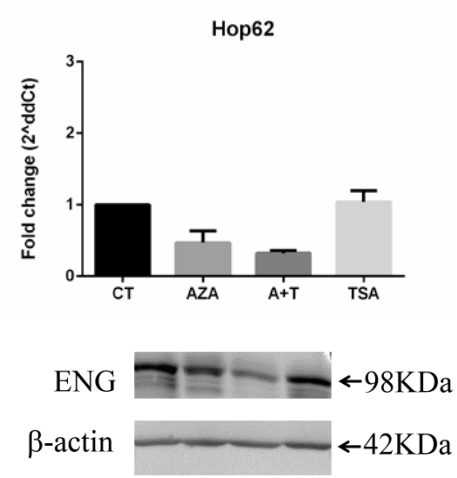


Figure 4-5 Demethylation treatment results in re-expression of Endoglin. (A-C) Cell lines which are methylated within the Endoglin promoter show up-regulation of expression following treatment with AZA. No effect of TSA on expression was observed. (D) No increase in Endoglin expression was observed following treatment with AZA and/or TSA in an unmethylated cell line, suggesting a direct effect of methylation on Endoglin expression.

4.2.1 Endoglin methylation in primary tissues

In our cell line panel, 13/18 (72%) have methylation levels above 50% in the Endoglin promoter region (Figure 4-3). The high frequency of methylation in cell lines prompted the question of whether this pattern is mirrored in primary tissues. To address this, two independent cohorts of formalin-fixed, paraffin-embedded (FFPE) tissues from biopsy or surgical resection were analysed for methylation using the pyrosequencing assay validated in our cell line panel. The characteristics of the patient cohorts are summarised in Table 4-1 and Table 4-2. Endoglin methylation data for cohort 1 and cohort 2 were tested to ascertain whether they had a normal distribution using the Shapiro-Wilk test (Shapiro and Wilk 1965), currently the gold standard for normality testing, having more statistical power than the traditionally used Kolmogorov-Smirnov test (Ghasemi and Zahediasl 2012). Data were found not to be normally distributed (cohort 1 $p=0.016$, cohort 2 $p<0.001$). Therefore non-parametric tests were used for all further analyses.

As with cell lines, methylation of primary tissues was calculated as the average across 13 CpG sites within the amplicon generated by the pyrosequencing primers. A panel of non-tumour lung tissue samples were used as controls with which to define a cut-off for methylation in tumour samples. Cohort 1 control tissues were histologically normal, tumour-adjacent lung tissues (n=19) and lung tissues from cancer-free emphysema patients (n=9). Cohort 2 control tissues were histologically normal, tumour-adjacent tissues only (n=20). Distribution of Endoglin methylation in control tissues compared to tumour tissues is shown in Figure 4-7(A) and Figure 4-8(A). For each cohort, the cut-off was defined as the average of control tissues plus 3 standard deviations. Applying this to our panel of tumour samples, cut-off values of 47% and 32% methylation were used, for cohort 1 and cohort 2, respectively. This led to 43% of cohort 1 and 24% of

cohort 2 being classified as methylated for Endoglin. For cohort 1, Endoglin methylation was not significantly affected by age or gender, or any of the clinical characteristics measured in this cohort (Table 4-1). For cohort 2, there was a significant relationship between histological subtype and Endoglin methylation according to the Kruskal-Wallis test for analysis of variance ($p<0.001$) Table 4-2).

Patients were stratified according to Endoglin methylation status and survival times were plotted in Kaplan-Meier curves (Figure 4-7, Figure 4-8). The cohorts are not matched; cohort 1 is a predominantly chemotherapy-naive early-stage cohort, while cohort 2 is made up of a more mixed population with patients ranging from stage I to stage IV. The Kaplan-Meier curves reflect this, with cohort 2 having a worse prognosis overall (Log Rank $p<0.001$) (Figure 4-6). However both cohorts show a trend towards shorter survival for patients with methylated Endoglin (Figure 4-7 B, Figure 4-8 B).

To try and compare both cohorts together, and to ascertain if Endoglin methylation might be more relevant in early stage lung cancers, survival analysis was done on stage 1 patients alone within cohort 2. Stage 1 patients comprised 56% (n=123) of cohort 2. When comparing survival functions, the effect was found to be stronger in this early stage group (generalised Wilcoxon $p=0.011$) (Figure 4-8 D). Therefore stage I and stage II-III patients from both cohorts were combined and entered into a multivariate Cox proportional hazards analysis to control for potential confounding factors such as grade, stage, age, gender and histology. Endoglin was confirmed a significant predictor of overall survival in stage I but not stage II-III cancers (Figure 4-9).

To confirm that the relationship between methylation and gene silencing observed in cell lines also exists in primary tissues, an *in silico* analysis was carried out using data from the TCGA (The Cancer Genome Atlas) database, a publicly accessible source of

data with clinical information, high-throughput sequencing and array data on various cancer types (Weinstein *et al.* 2013). Matched methylation (Illumina Human 450K Methylation BeadChip) and expression (Illumina HiSeq 2000 RNA sequencing) data were available for 205 lung adenocarcinoma primary tissues. Expression values are normalised RNA-Seq by Expectation Maximization (RSEM) (Li and Dewey 2011) count estimates for sequence reads mapping to the Endoglin gene. Methylation values for the array probe located within the Endoglin CpG island were filtered by preprocessing so the probe closest to this was used as a surrogate. This probe is located 297bp upstream of the island probe, within the south shore region. In our panel of cell lines, beta values for this probe differed by <0.3 for 11 out of 13 cell lines. There was a significant correlation between methylation of this probe and expression of Endoglin in the TCGA dataset; Spearman's rank correlation coefficient (r) = -0.2684 (95% CI -0.3945 to -0.1324), $p<0.0001$ (Figure 4-10). Therefore, it is reasonable to conclude that highly methylated cancers in this group have Endoglin epigenetically silenced, and from this to extrapolate that highly methylated samples within cohort 1 and 2 would be silenced for Endoglin.

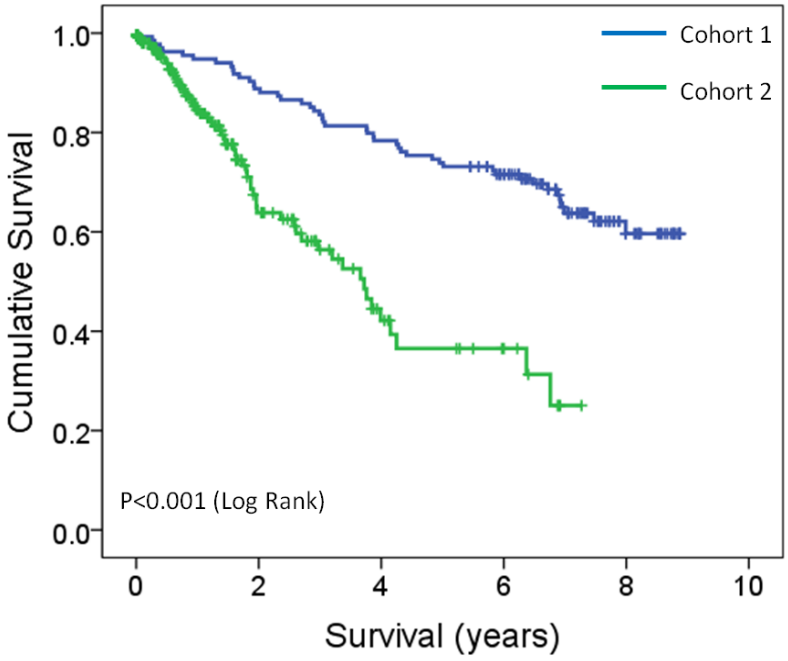
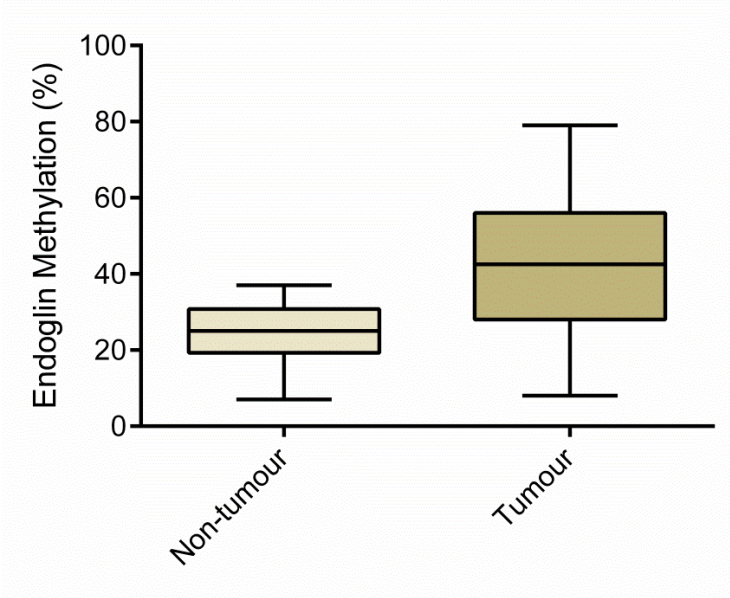


Figure 4-6 Comparison of survival curves for cohort 1 and cohort 2 by Kaplan-Meier analysis. Cohort 2 are a mixed population of early and late stage cancers, while cohort 1 are early stage cancers. As expected, the survival curves of these two cohorts are significantly different (Log Rank $p<0.001$).

Basic patient characteristics Cohort 1 (N=128)				
	Number	**Methylated (n=56)**	**Unmethylated (n=72)**	***P* –value**
Age at diagnosis				
Median: 69 (36-84)				
<69	65	27 (48%)	38 (53%)	
>69	63	29 (52%)	34 (47%)	0.652
Sex				
Female	32	9 (16%)	23 (32%)	
Male	96	47 (84%)	49 (68%)	0.128
Diagnosis				
Squamous	65	31 (55%)	34 (48%)	
Adenocarcinoma	35	10 (18%)	25 (35%)	
Other	14	7 (13%)	7 (10%)	0.233
Type of surgery				
Segmentectomy	4	3 (6%)	1 (1.5%)	
Lobectomy/Bilobectomy	106	42 (84%)	64 (97%)	
Pneumonectomy	6	5 (10%)	1 (1.5%)	0.072
Post-operative stage				
Stage IA	54	19 (38%)	35 (55%)	
Stage IB	58	29 (58%)	29 (45%)	
Stage II	2	2 (4%)	-	0.401
Tumour Grade				
G1	14	9 (19%)	5 (8%)	
G2	57	20 (42%)	37 (57%)	
G3	42	19 (39%)	23 (35%)	0.170
Smoking history				
Smoker	69	35 (76%)	34 (60%)	
Ex-smoker	23	8 (17%)	15 (26%)	
Never-smoker	11	3 (7%)	8 (14%)	0.503

Table 4-1 Patient characteristics of cohort 1 and clinicopathologic correlation of Endoglin methylation in NSCLC. Columns show number of individuals with methylated and unmethylated Endoglin for each characteristic, with percentages referring to proportion of the total methylated or unmethylated population. P-values relate to differences between methylated and unmethylated populations in relation to the clinical characteristics listed. P-values are calculated using the independent samples Mann-Whitney U test (for variables with two groups) or the Kruskal-Wallis test (for variables with more than two groups) for one-way analysis of variance.

A



B

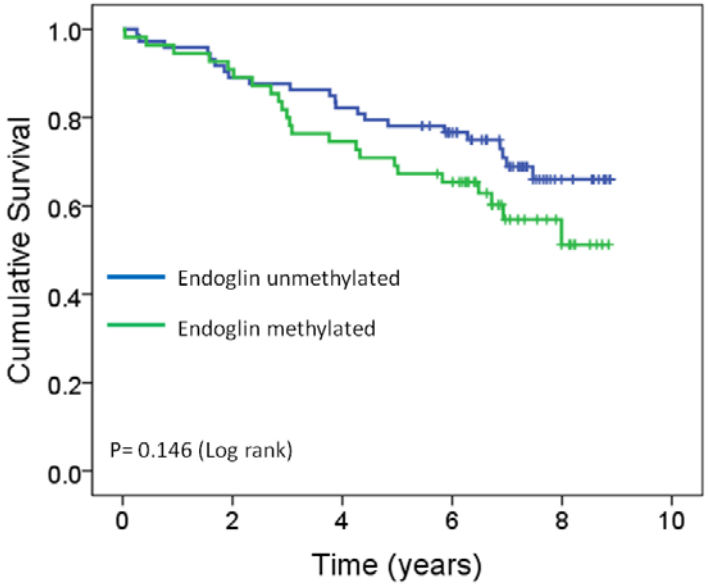
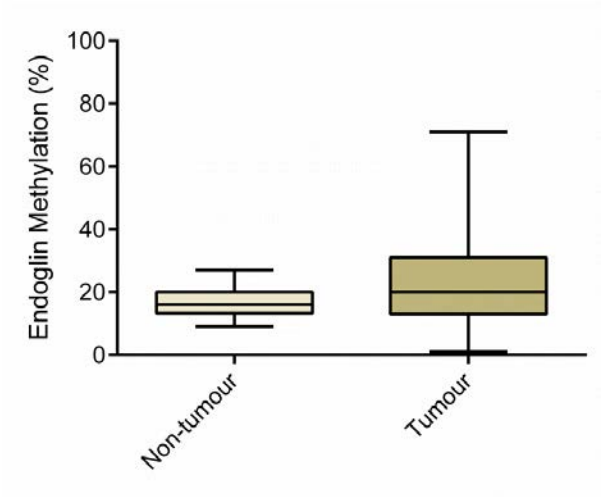


Figure 4-7 Analysis of Endoglin methylation in cohort 1 by pyrosequencing. (A) Distribution of Endoglin methylation in primary lung tumour tissues compared to control tissues in cohort 1. Control tissues are histologically normal, tumour adjacent lung tissues (n=19) and lung tissues from cancer-free emphysema patients (n=9). (B) Kaplan-Meier survival curve showing patients from cohort 1 stratified by Endoglin methylation. Methylated samples are defined as those with methylation value greater than the cut-off of 47%. This cut-off was generated by calculating the mean methylation of control tissues plus three standard deviations. There is a trend towards shorter survival times for patients with methylated Endoglin. Using univariate Cox regression analysis, the hazard ratio (HR)= 1.538; 95% CI, 0.857-2.763, p=0.149).

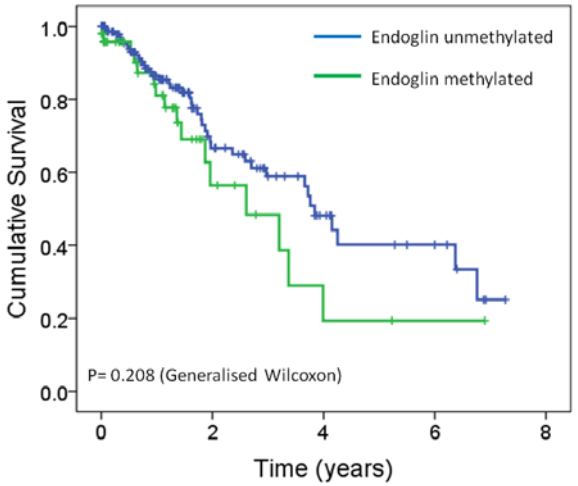
Basic patient characteristics Cohort 2 (N=208)				
	Number	**Methylated (n=49)**	**Unmethylated (n=159)**	***P* –value**
Age at diagnosis				
Median: 65 (23-89)				0.836
<65	102	26 (53%)	76 (48%)	
≥65	106	23 (47%)	83 (52%)	
Sex				
Female	105	24 (49%)	81 (51%)	0.664
Male	103	25 (51%)	78 (49%)	
Diagnosis				
Squamous	49	16 (33%)	33 (21%)	<0.001*
Adenocarcinoma	128	15 (31%)	113 (71%)	
Other	31	18 (37%)	13 (13%)	
Type of surgery				
Wedge	46	11 (22%)	35 (22%)	0.582
Lobectomy/Bilobectomy	158	36 (73%)	122 (77%)	
Pneumonectomy	4	2 (4%)	2 (1%)	
Post-operative stage				0.275
Stage IA	69	14 (29%)	55 (35%)	
Stage IB	46	13 (27%)	33 (21%)	
Stage IIA	40	7 (14%)	33 (21%)	
Stage IIB	20	3 (6%)	17 (11%)	
Stage III, IV	28	10 (20%)	18 (11%)	
Tumour Grade				
G1	41	15 (31%)	26 (16%)	0.146
G2	103	20 (41%)	83 (52%)	
G3	63	13 (27%)	50 (31%)	

Table 4-2 Patient characteristics of cohort 2 and clinicopathologic correlation of Endoglin methylation in NSCLC. Columns show number of individuals with methylated and unmethylated Endoglin for each characteristic, with percentages referring to proportion of the total methylated or unmethylated population. P-values relate to differences between methylated and unmethylated populations in relation to the clinical characteristics listed. P-values are calculated using the independent samples Mann-Whitney U test (for variables with two groups) or the Kruskal-Wallis test (for variables with more than two groups) for one-way analysis of variance.

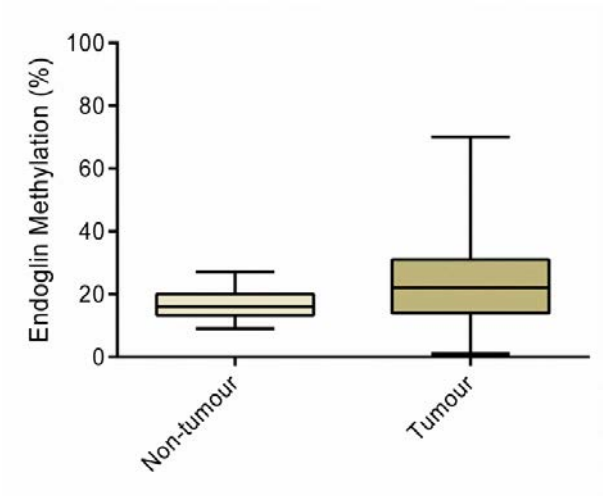
A


Endoglin Methylation (%)
100
80
60
40
20
0
Non-tumour
Tumour

B


Endoglin unmethylated
Endoglin methylated
Cumulative Survival
1.0
0.8
0.6
0.4
0.2
0.0
P= 0.208 (Generalised Wilcoxon)
0
2
4
6
8
Time (years)

C


Endoglin Methylation (%)
100
80
60
40
20
0
Non-tumour
Tumour

D

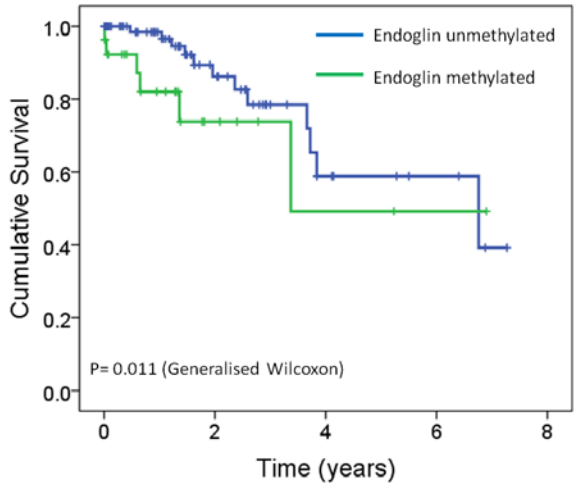

Endoglin unmethylated
Endoglin methylated
Cumulative Survival
1.0
0.8
0.6
0.4
0.2
0.0
P= 0.011 (Generalised Wilcoxon)
0
2
4
6
8
Time (years)

Figure 4-8 Analysis of Endoglin methylation in cohort 2 by pyrosequencing. (A) Distribution of Endoglin methylation in primary lung tumour tissues compared to control tissues in cohort 2. Control tissues are histologically normal, tumour adjacent lung tissues (n=20). (B) Kaplan-Meier survival curve showing patients from cohort 2 stratified by Endoglin methylation. Methylated samples are defined as those with methylation value greater than the cut-off of 32%. This cut-off was generated by calculating the mean methylation of control tissues plus three standard deviations. There is a trend towards shorter survival times for patients with methylated Endoglin. Using univariate Cox regression analysis, the hazard ratio (HR) = 1.522; 95% CI, 0.853-2.715, p=0.155. (C, D) Survival analysis was carried out on stage 1 patients only within this cohort. (D) Comparison of methylated and unmethylated survival functions by Kaplan-Meier analysis was found to be significant in this group of early stage cancers (generalised Wilcoxon p=0.011). However the hazard ratio failed to reach significance levels in a univariate Cox regression analysis. HR = 2.046; 95% CI 0.765-5.475, p=0.154.

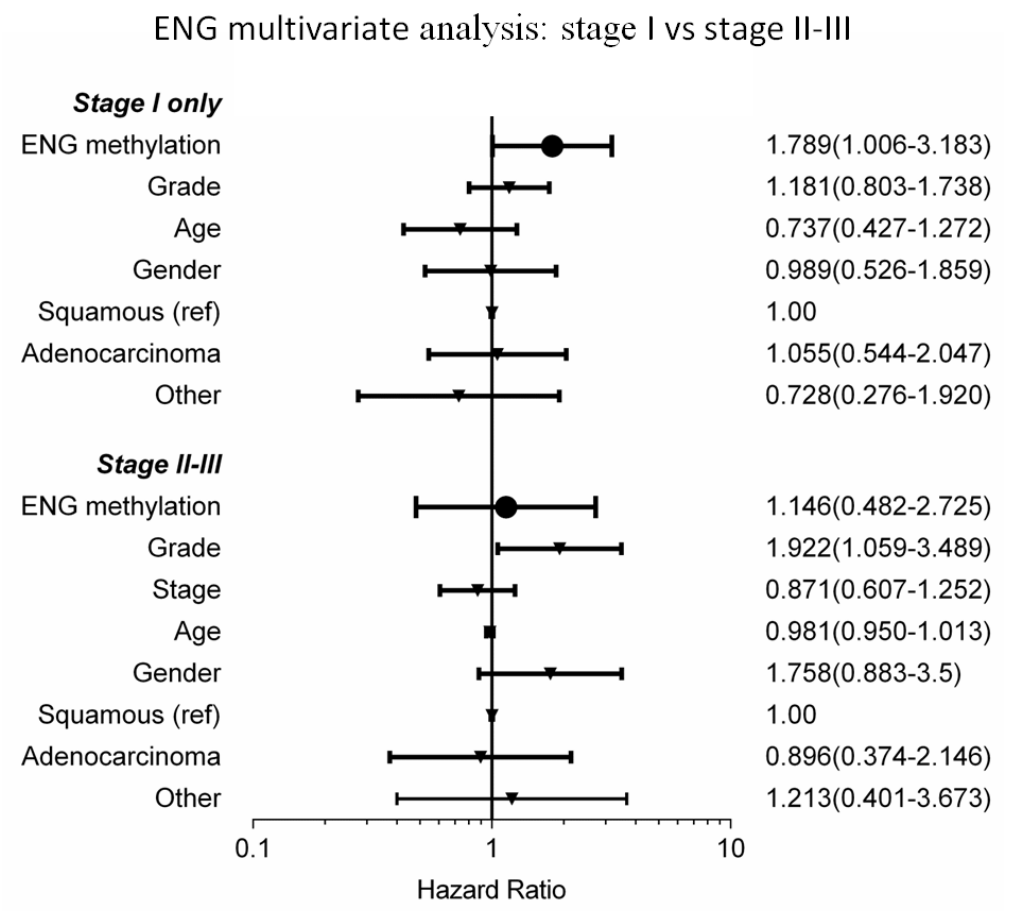
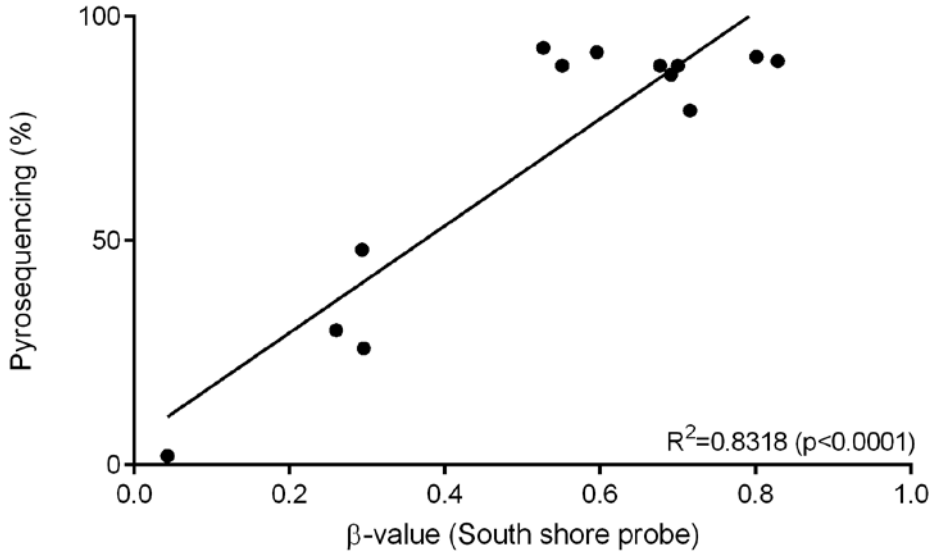


Figure 4-9 Endoglin is an independent predictor of outcome in stage I NSCLC. A multivariate Cox proportional hazards analysis was carried out on stage I patients from both cohorts (n=227) and stage II-III patients (n=90). Forest plot shows hazard ratios and 95% confidence intervals for ENG methylation and potential confounding factors grade, stage, age, gender and histology. For stage I patients, ENG methylation was the only significant predictor of outcome. In the stage II-III patient group, ENG methylation had no prognostic value and only grade was an independent predictor of outcome.

A



B

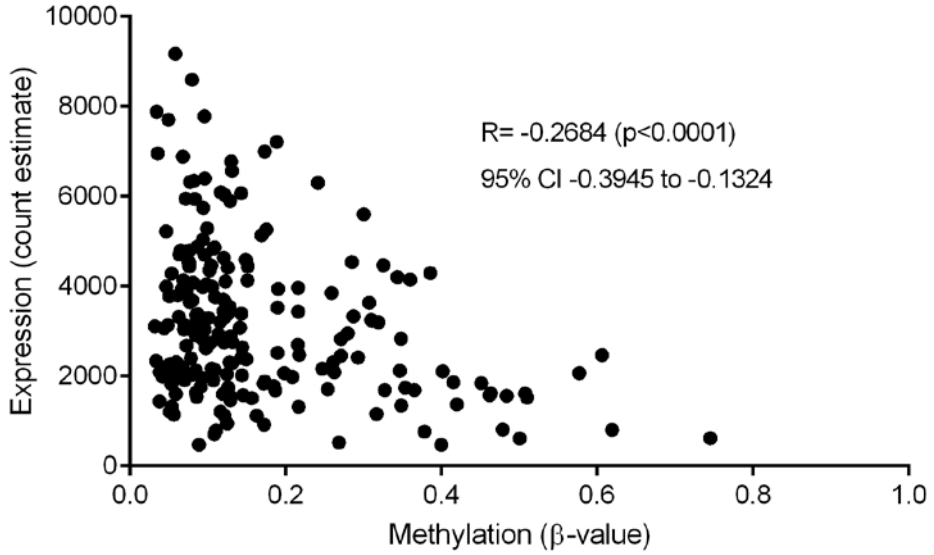


Figure 4-10 *In silico* analysis was carried out on publically available microarray data. Methylation values for the probe located within the Endoglin CpG island were filtered by pre-processing, so data for the nearest probe, which was 279bp upstream within the south shore region, were plotted. In our cell line panel, methylation beta values (Illumina Human 450K Methylation BeadChip) from this upstream probe correlate with % methylation within the island as measured by pyrosequencing according to linear regression analysis; R^2=0.8318, p<0.0001 (A). Therefore methylation beta values (Illumina Human 450K Methylation BeadChip) were plotted against mRNA count estimates (Illumina HiSeq 2000 RNA sequencing) for a series of lung adenocarcinoma primary tissues (N=205) downloaded from the TCGA (The Cancer Genome Alas) database. There was a significant correlation between methylation and expression of Endoglin in these primary tissues according to Spearman's rank correlation coefficient: R= -0.2684 (95% CI -0.3945 to -0.1324), p<0.0001.

4.2.2 Functional analysis of Endoglin in lung cancer cell lines

It has been shown thus far that epigenetic silencing of Endoglin is a common event in NSCLC cell lines and primary tissues, and that methylation of Endoglin may indicate a worse prognosis in early stage cancers. Based on the role of Endoglin in the TGF-β signalling pathway, it was hypothesised that the difference in survival may be due to the effect of Endoglin silencing on signalling within this pathway. Since the TGF-β pathway involves a delicate balance of pro- and anti-invasion, migration and proliferation signals, assays to quantify these functional traits were carried out. The following cell lines were analysed: the highly methylated and Endoglin-silenced NSCLC cell lines HCC95, H23, A549, H322M, MOR after transfecting with a pcDNA3.1 plasmid to ectopically express the gene, as well as the unmethylated and Endoglin expressing cell line Hop62, after transient silencing of Endoglin with an siRNA. Plasmid-transfected cells were selected with G418 at the following concentrations: HCC95, H23-300µg/ml, A549, H322M, and MOR-500µg/ml. Cells were seeded as follows: for the tumour spheroid-based invasion assay- $1x10^4$cells/ml for Hop62, MOR, H322M, $0.1x10^4$cells/ml for HCC95, $0.5x10^4$cells/ml for A549.

Two distinct behavioural profiles were identified within the panel of overexpressing cell lines. In the first group (which will be referred to as group 1), Endoglin expression has no effect on migration, invasion or proliferation either with or without TGF-β and includes the cell lines MOR (Figure 4-11), H23 (Figure 4-12), A549 (Figure 4-13) and Hop62 (Figure 4-16). In these cell lines there was no difference between vector-transfected (VEC) and Endoglin-transfected (ENG) cells.

The second group respond to TGF-β stimulation and this response is dependent on Endoglin expression. This is clearly shown by the cell lines H322M and HCC95. In

these cells, presence of TGF-β leads to increased invasion in cells lacking Endoglin, but Endoglin-expressing cells are protected from this pro-invasive effect. The H322M VEC cells are stimulated by TGF-β to form protrusions which invade the surrounding Matrigel, but ENG cells are protected from this effect (Figure 4-14). HCC95 cells could not be induced to form spheres in 3D culture therefore invasion of these cells was analysed using a Matrigel-coated transwell plate assay (BD Biosciences). Again, Endoglin silenced cells were more invasive compared to Endoglin-expressing cells, this time in a TGF-β independent manner (Figure 4-15). Migration and proliferation of these cells is also affected by Endoglin expression, although in the opposite manner; cells lacking Endoglin migrate and proliferate less than those expressing the gene in the presence of TGF-β. This was unexpected as invasion and migration are accepted to be complementary, rather than opposing characteristics of cancer cells, which will be discussed later in this chapter.

In order to ascertain whether this decreased migration of cells showing increased invasion may be simply a consequence of slower proliferation under normal monolayer culture conditions, proliferation was tested over 5 days using MTT assay. Cells were seeded at the following densities into 95-well plates: MOR: $5x10^3$/well, H23: $3x10^3$/well, A549: $3x10^3$/well, H322M: $6x10^3$/well, HCC95: $1x10^3$/well, Hop62: $1.83x10^3$/well. Indeed, cells with this second behavioural phenotype (i.e. H322M and HCC95) showed a clear trend towards less proliferation of VEC cells compared with ENG cells. No such trend was observed in cells of the first group (Figure 4-17 A, B). It has been shown in the literature that invasive cells may down-regulate proliferation by means of the cell cycle inhibitory protein p16 (Palmqvist *et al.* 2000; Svensson *et al.* 2003). Therefore p16 mRNA expression was quantified by qRT-PCR in the cell line panel to ascertain this was the mechanism by which H322M and HCC95 cells

lacking Endoglin down-regulate proliferation. However in there was no difference in p16 expression between Endoglin-expressing and Endoglin-silenced cells (Figure 4-17 C).

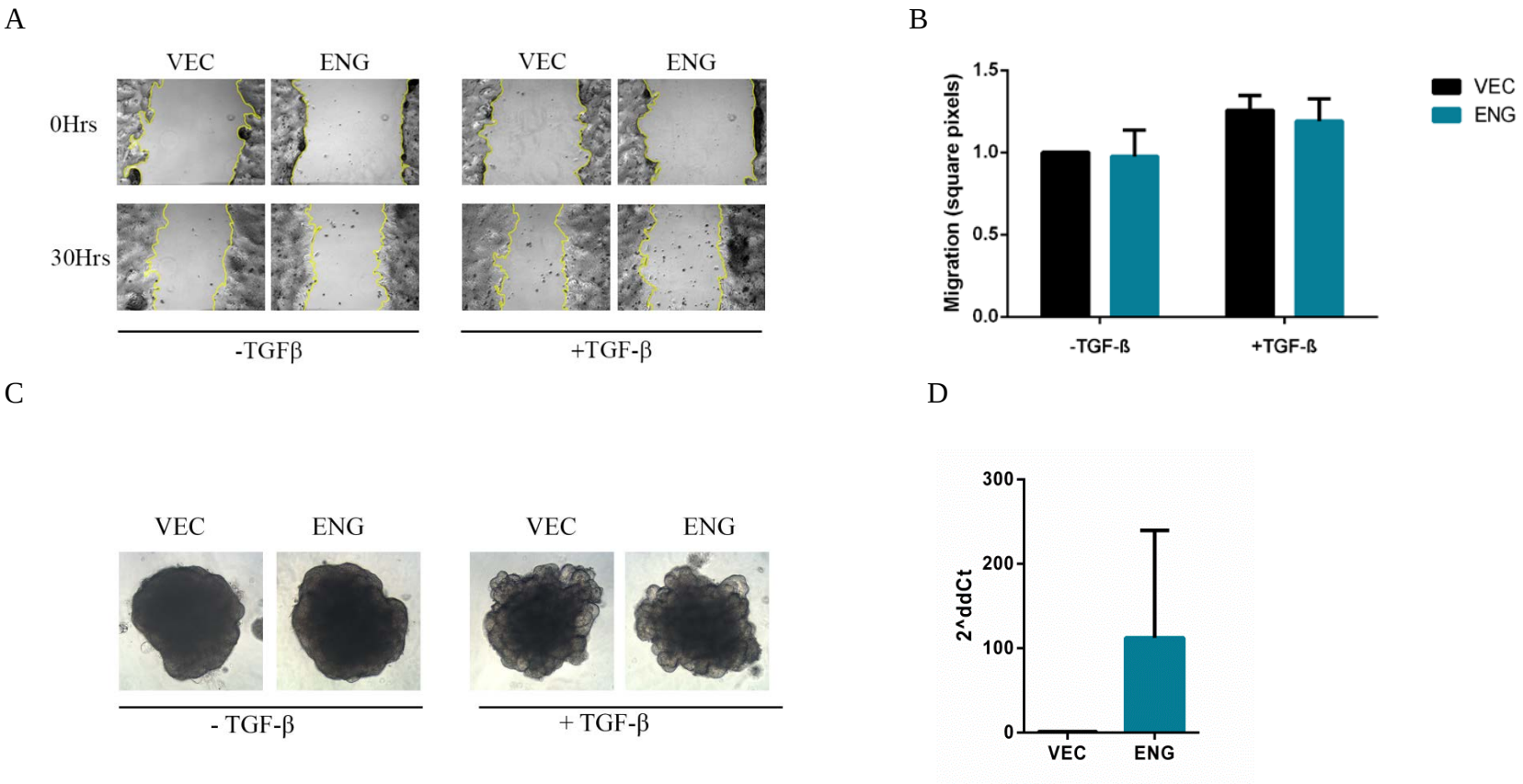


Figure 4-11 Analysis of ENG function in the cell line MOR. This cell line is methylated in the promoter region of Endoglin and has lost expression of the gene. Cells were transfected with the pcDNA3.1 plasmid containing the Endoglin cDNA, or an empty pcDNA3.1 plasmid. Where indicated, cells were treated with TGF-β1 at 10ng/ml. In a scratch wound-healing assay (A, B), there was no difference in migration between Endoglin expressing cells (ENG) and vector control cells (VEC), either in the presence or absence of TGF-β over a 24-hour period, although TGF-β did stimulate migration for both cell types. In a three-dimensional invasion assay (C), cells were induced to form three-dimensional spheroids and encouraged to invade with the addition of Matrigel. Under these conditions a difference in growth was observed in the presence of TGF-β over 5 days, but no difference between ENG and VEC cells was observed. Re-expression of Endoglin in ENG pcDNA3.1 cells was confirmed by RT-qPCR (D). All graphs show mean ±SD of three biological replicates, normalised to the untreated vector control.

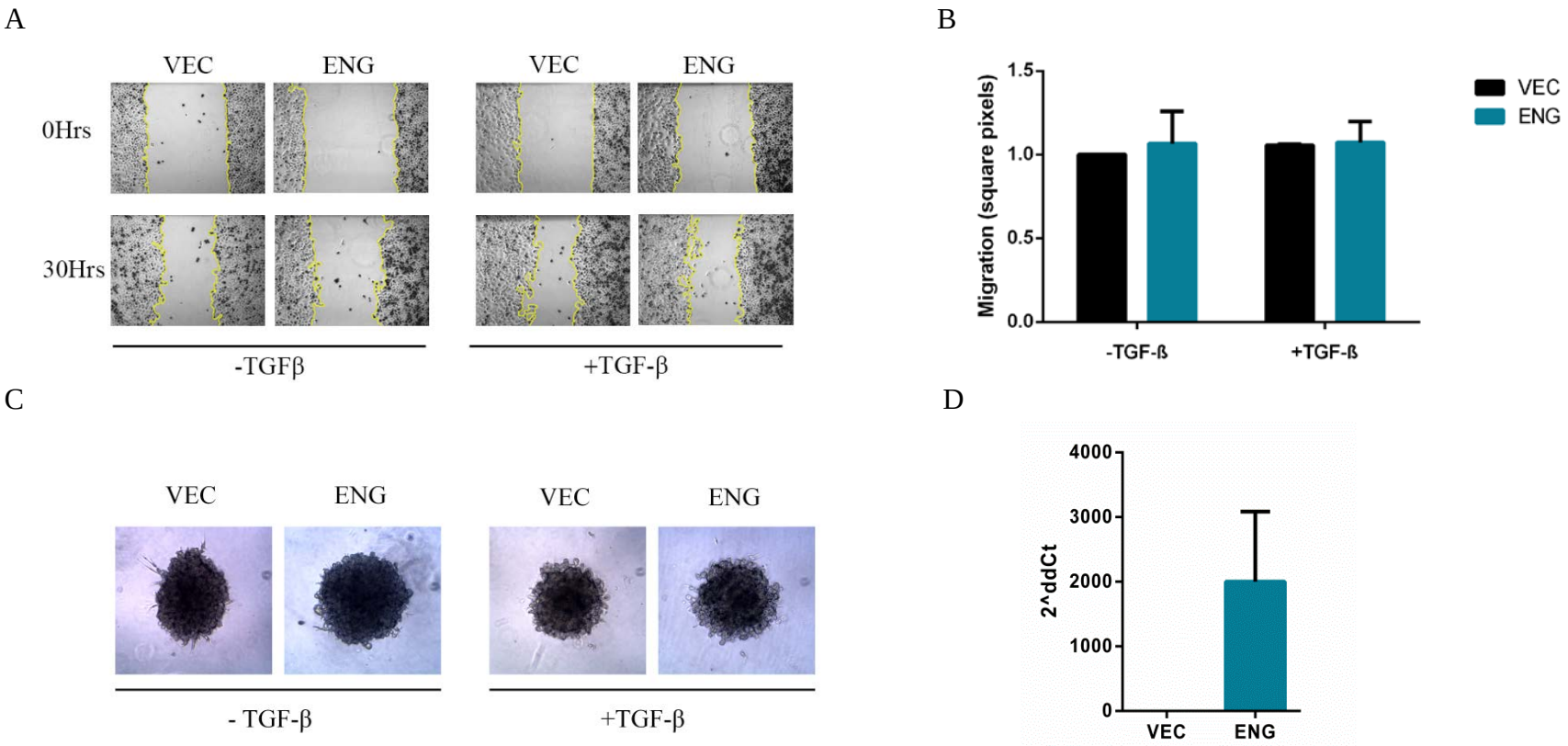


Figure 4-12 Analysis of ENG function in the cell line H23. This cell line is methylated in the promoter region of Endoglin and has lost expression of the gene. Cells were transfected with the pcDNA3.1 plasmid containing the Endoglin cDNA, or an empty pcDNA3.1 plasmid. Where indicated, cells were treated with TGF-β1 at 10ng/ml. In a scratch wound-healing assay (A, B), there was no difference in migration between Endoglin expressing cells (ENG) and vector control cells (VEC), either in the presence or absence of TGF-β over a 24-hour period. In a three-dimensional invasion assay (C), cells were induced to form three-dimensional spheroids and encouraged to invade with the addition of Matrigel. There was no difference in the growth pattern of ENG and VEC cells either in the presence or absence of TGF-β over 5 days. Re-expression of Endoglin in ENG pcDNA3.1 cells was confirmed by RT-qPCR (D). All graphs show mean ±SD of three biological replicates, normalised to the untreated vector control.

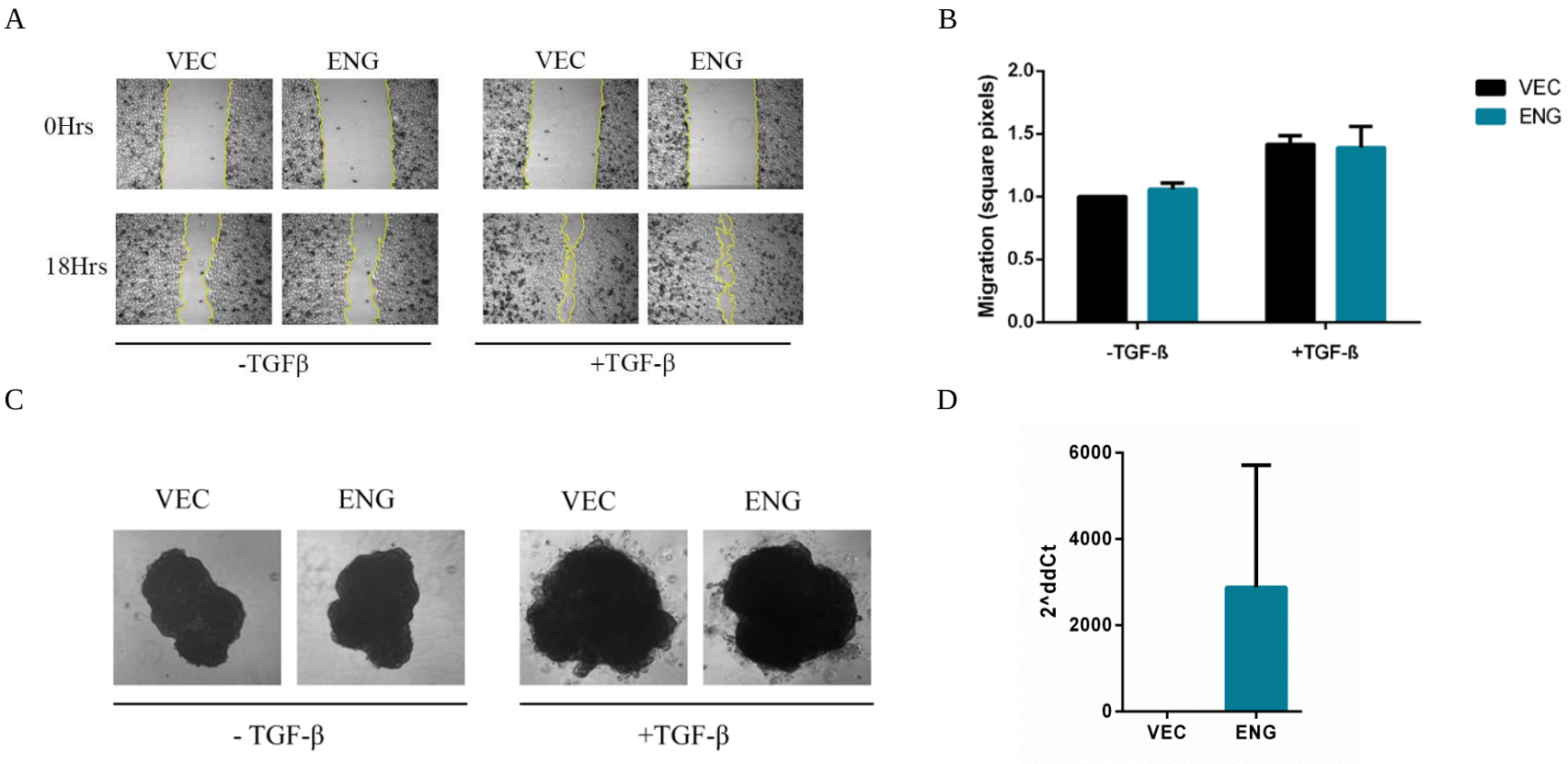


Figure 4-13 Analysis of ENG function in the cell line A549. This cell line is methylated in the promoter region of Endoglin and has lost expression of the gene. Cells were transfected with the pcDNA3.1 plasmid containing the Endoglin cDNA, or an empty pcDNA3.1 plasmid. Where indicated, cells were treated with TGF-β1 at 10ng/ml. In a scratch wound-healing assay (A, B), there was no difference in migration between Endoglin expressing cells (ENG) and vector control cells (VEC), either in the presence or absence of TGF-β over a 24-hour period, although TGF-β did stimulate migration for both cell types. In a three-dimensional invasion assay (C), cells were induced to form three-dimensional spheroids and encouraged to invade with the addition of Matrigel. There was no difference in the growth pattern of ENG and VEC cells either in the presence or absence of TGF-β over 5 days. Re-expression of Endoglin in ENG pcDNA3.1 cells was confirmed by RT-qPCR (D). All graphs show mean ±SD of three biological replicates, normalised to the untreated vector control.

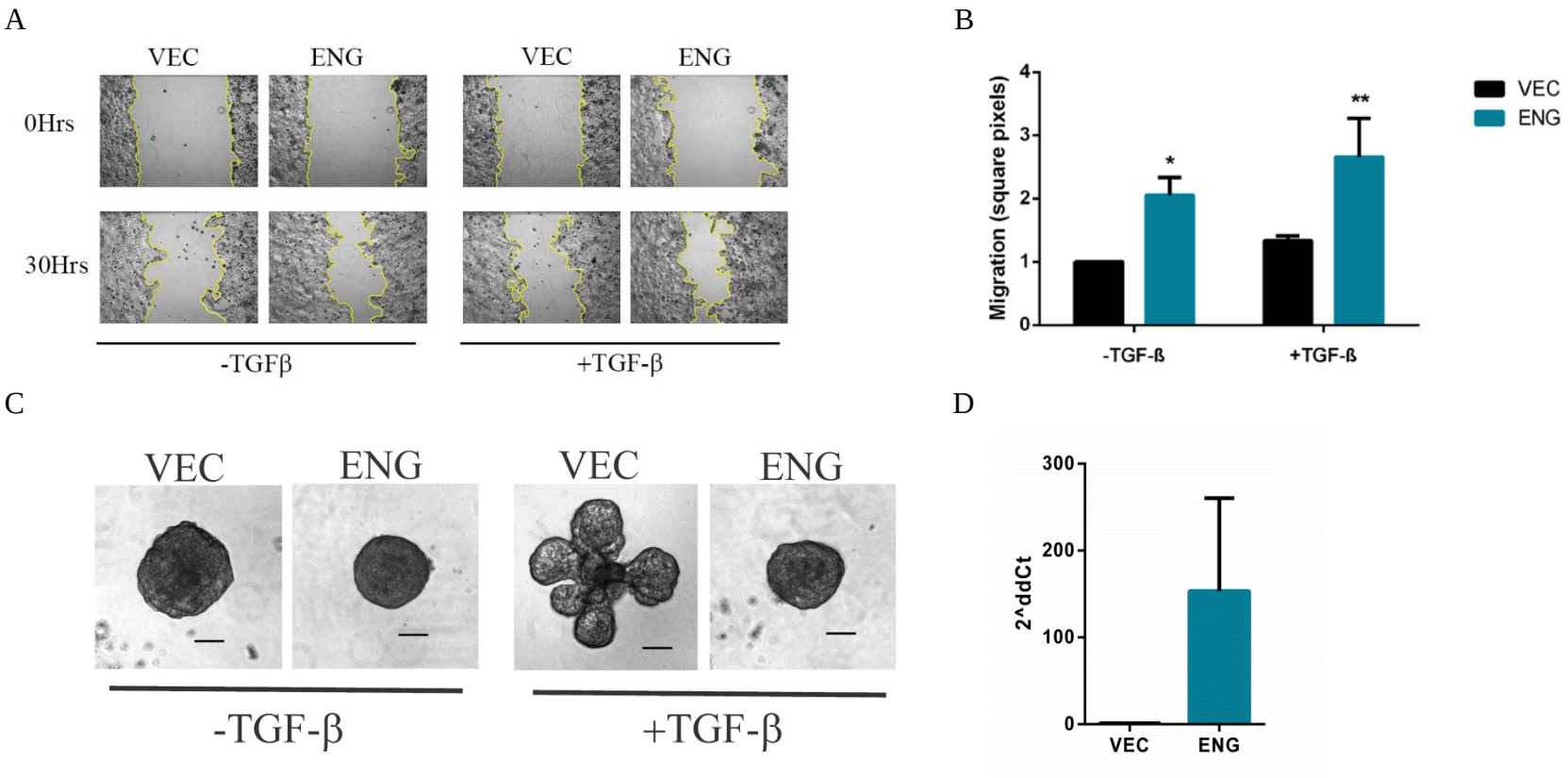


Figure 4-14 Analysis of ENG function in the cell line H322M. This cell line is methylated in the promoter region of Endoglin and has lost expression of the gene. Cells were transfected with the pcDNA3.1 plasmid containing Endoglin cDNA, or an empty pcDNA3.1 plasmid. Where indicated, cells were treated with TGF-β1 at 10ng/ml. In a scratch wound-healing assay (A, B), Endoglin expressing cells (ENG) migrated further than vector control cells (VEC), with and without TGF-β over a 24-hour period. In a three-dimensional invasion assay (C), cells were induced to form three-dimensional spheroids and encouraged to invade with the addition of Matrigel. VEC cells were able to invade into the extracellular matrix in the presence of TGF-β, whereas ENG cells were not. Re-expression of Endoglin in ENG pcDNA3.1 cells was confirmed by RT-qPCR (D). All graphs show mean ±SD of three biological replicates, normalised to the untreated vector control. Scale bar 100µm.

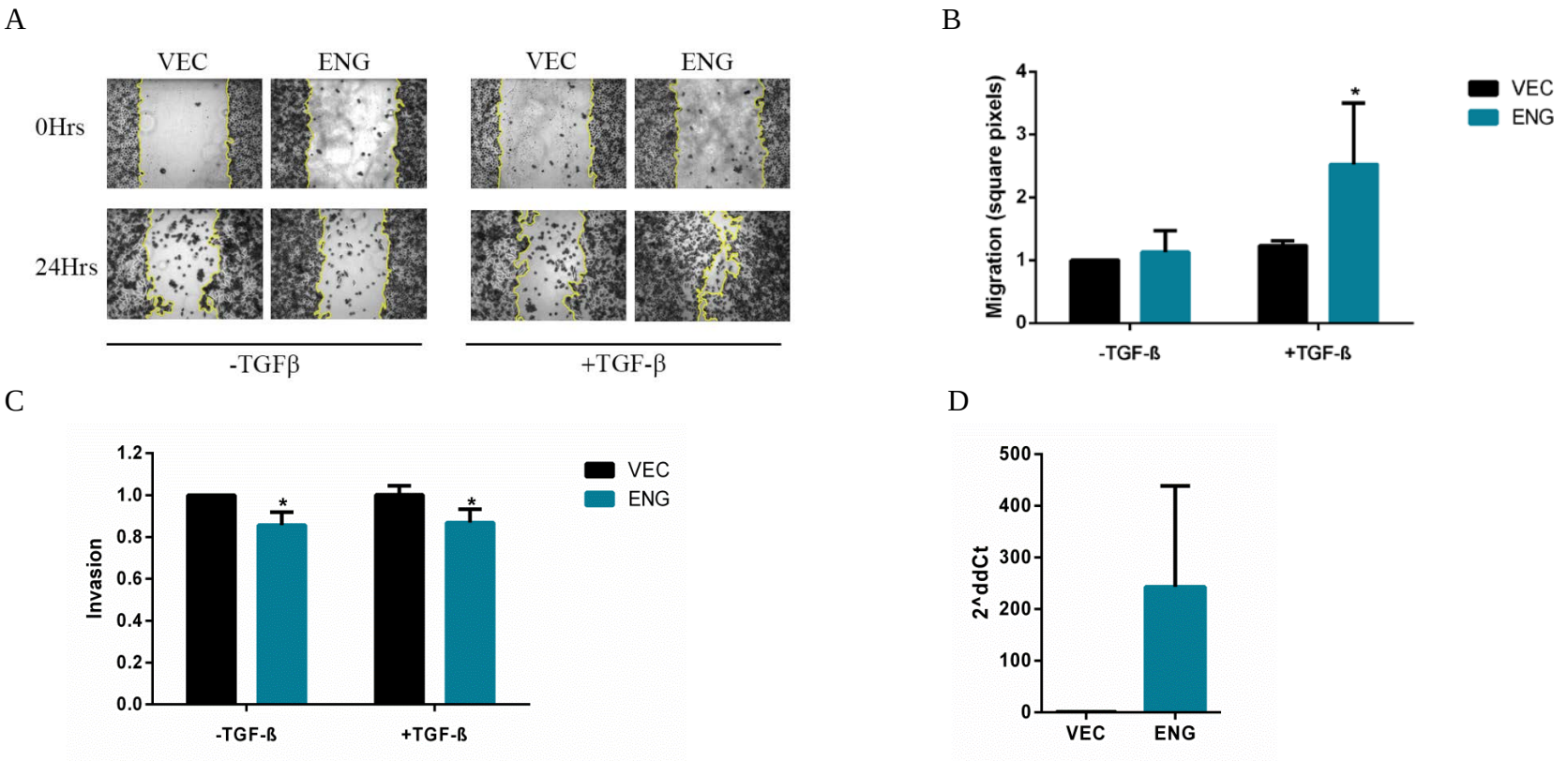


Figure 4-15 Analysis of ENG function in the cell line HCC95. This cell line is methylated in the promoter region of Endoglin and has lost expression of the gene. Cells were transfected with the pcDNA3.1 plasmid containing Endoglin cDNA, or an empty pcDNA3.1 plasmid. Where indicated, cells were treated with TGF-β1 at 10ng/ml. In a scratch wound-healing assay (A, B), Endoglin expressing cells (ENG) migrated further than vector control cells (VEC), in the presence of TGF-β over a 24-hour period. In a transwell invasion assay (C), fluorescently stained cells were seeded into Matrigel-coated plate inserts and allowed to invade towards a chemoattractant (FBS) over 24 hours. VEC cells were able to invade more than ENG cells, both in the presence and absence of TGF-β. Re-expression of Endoglin in ENG pcDNA3.1 cells was confirmed by RT-qPCR (D). All graphs show mean ±SD of three biological replicates, normalised to the untreated vector control.

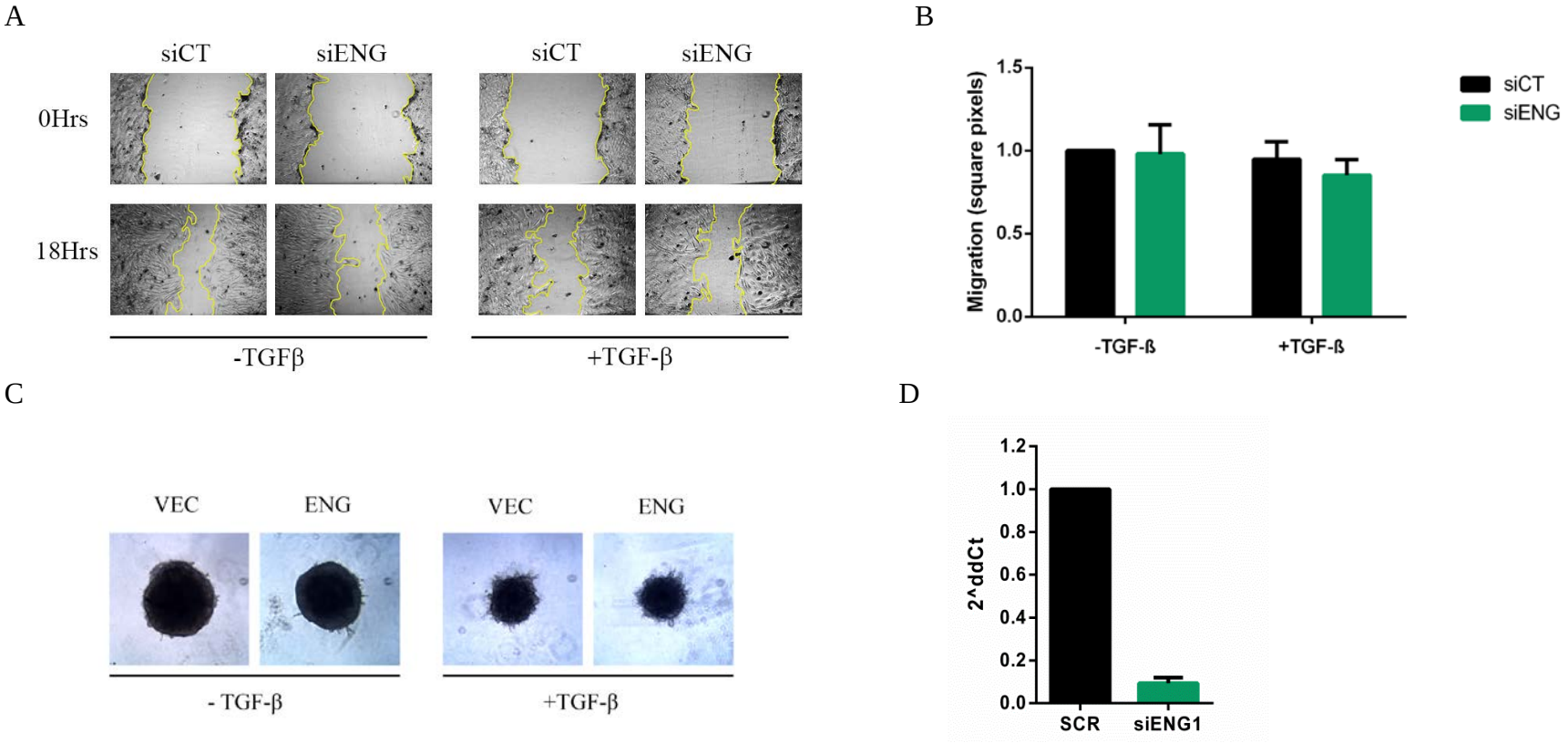
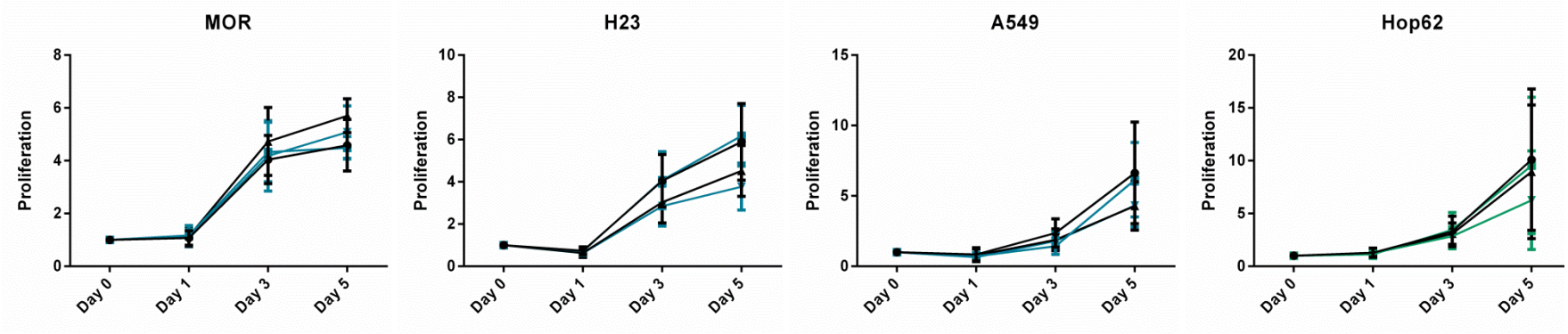
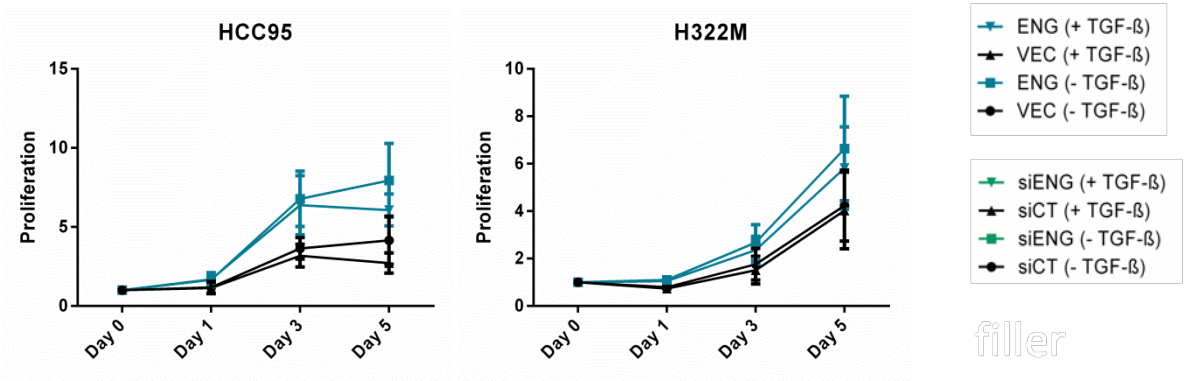


Figure 4-16 Analysis of ENG function in the cell line Hop62. This cell line is not methylated in the Endoglin promoter region and therefore expresses the gene. Cells were transfected with an Endoglin-targeting siRNA, or a non-silencing siRNA control. Where indicated, cells were treated with TGF-β1 at 10ng/ml. In a scratch wound-healing assay (A), there was no difference between Endoglin expressing cells (siCT) and Endoglin-silenced cells (siENG) with or without TGF-β over a 24-hour period. In a three-dimensional invasion assay (C), cells were induced to form three-dimensional spheroids and encouraged to invade with the addition of Matrigel. Under these conditions a difference in growth was observed in the presence of TGF-β over 5 days, but no difference between siCT and siENG cells was observed. Silencing of Endoglin by siRNA was confirmed by RT-qPCR (D). All graphs show mean ±SD of three biological replicates, normalised to the untreated vector control.

A


MOR
H23
A549
Hop62
Proliferation
Day 0
Day 1
Day 3
Day 5

B


HCC95
H322M
Proliferation
Day 0
Day 1
Day 3
Day 5

ENG (+ TGF-ß)
VEC (+ TGF-ß)
ENG (- TGF-ß)
VEC (- TGF-ß)
siENG (+ TGF-ß)
siCT (+ TGF-ß)
siENG (- TGF-ß)
siCT (- TGF-ß)

filler

C

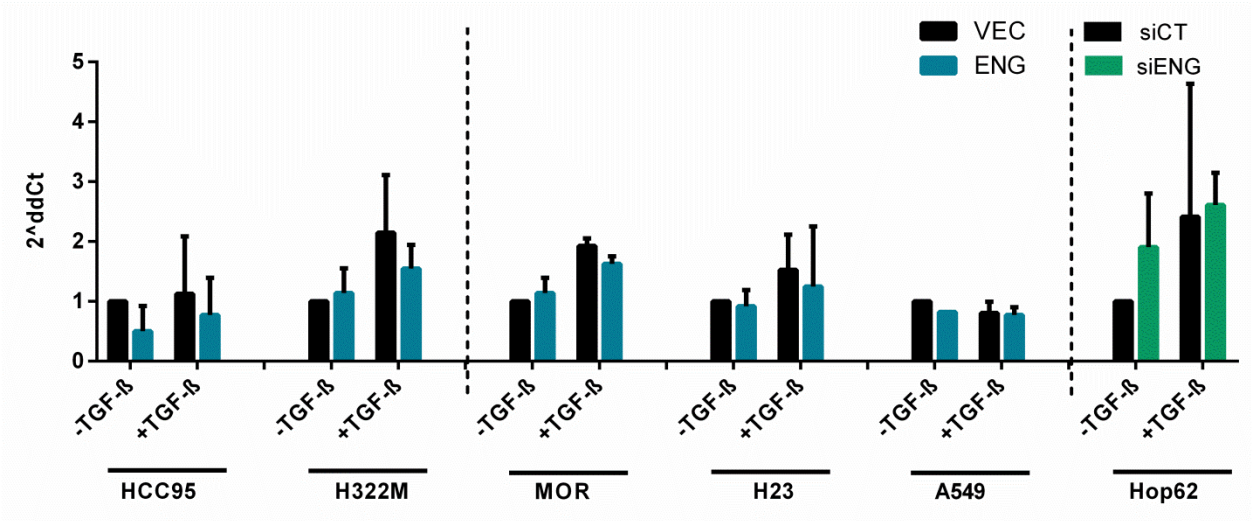


Figure 4-17 Effect of Endoglin expression on proliferation of NSCLC cell lines. Proliferation was measured in Endoglin transfected (ENG) and empty vector transfected (VEC) cell lines which were methylated and had therefore lost expression of the gene, as well as the unmethylated cell line Hop62 which had the highest baseline Endoglin expression, after transfection with an Endoglin-targeting siRNA (siENG) and a scrambled siRNA control (siCT). Proliferation was measured by MTT on indicated days. Data shown are mean ±SD of three independent experiments in which cells were treated with TGF-β1 at 10ng/ml for 5 days. Each sample is normalised to baseline value at day 0. (A) The cell lines MOR, H23, A549 and Hop62, which did not show any Endoglin-dependant difference in invasion or migration, do not show a difference in proliferation. (B) The cell lines HCC95 and H322M, which showed increased invasion but decreased migration of Endoglin-silenced (VEC) cells, show decreased proliferation compared to Endoglin-expressing (ENG) cells. (C) P16 (CDKN2A) expression was measured by RT-qPCR in all of the above cell lines to ascertain whether the decreased proliferation of HCC95 and H322M VEC cells was due to up-regulation of p16. Cells were treated with TGF-β1 at 10ng/ml for 24 hours. There was no difference in expression of p16 between ENG and VEC cells, either with or without TGF-β.

It was hypothesised that TGF-β may be driving epithelial to mesenchymal transition (EMT), in cells lacking Endoglin, while those expressing it are protected by the effect of Endoglin on TGF-β signalling. Vimentin expression, a widely-used mesenchymal marker, was quantified by RT-qPCR in all transfected cell lines to assess whether the behavioural profiles described above may be related to this particular pathological process of acquiring invasiveness. Vimentin expression was found to reflect cell behaviour. Cells in group 1 showed no difference in Vimentin expression. However, the cell lines H322M and HCC95 (group 2) showed up-regulation of this mesenchymal marker in cells lacking Endoglin, in the presence of TGF-β (Figure 4-18).

In order to address the question of why the cell lines H322M and HCC95 behave differently according to Endoglin expression levels while MOR, H23, A549 and Hop62 do not, these cell lines were analysed using an optimised EMT expression panel (Walter *et al.* 2012), to classify the EMT status of each one. According to this panel of markers, 3/4 cell lines of group 2 (H23, A549 and Hop62) are highly mesenchymal-like in profile, while MOR is epithelial-like. Interestingly, group 1 cell lines (H322M, HCC95) are both epithelial (Figure 4-19). For these cell lines, levels of individual EMT markers in Endoglin-expressing compared to Endoglin-silenced cells suggests that loss of Endoglin may promote EMT progression (Figure 4-20B, C), while having no effect on the mesenchymal-like H23 (Figure 4-20D, E).

Together these results suggest that in a proportion of NSCLC cell lines, Endoglin silencing may drive EMT progression and allow cells to acquire a more invasive phenotype. Therefore, all of this reveals a tumour-suppressor effect of Endoglin and a tumour-promoting effect of Endoglin methylation, which may explain the trend towards a worse prognosis for patients with methylated Endoglin.

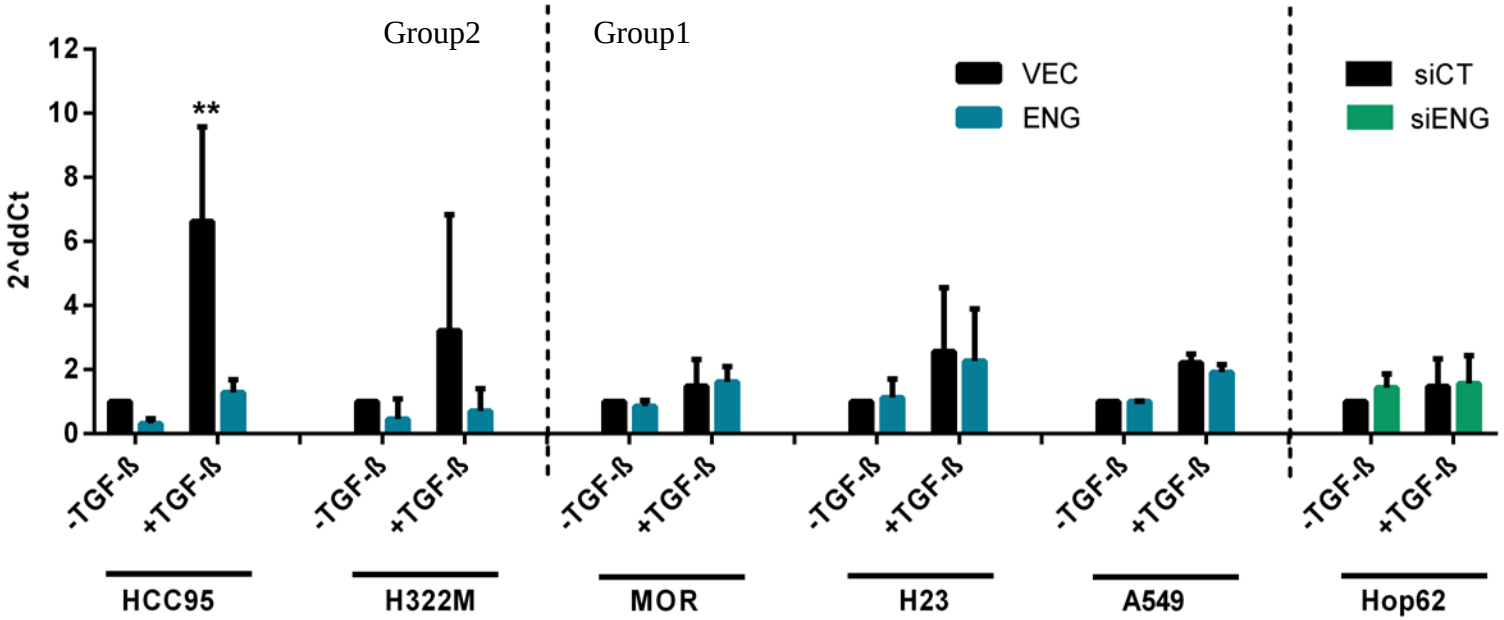
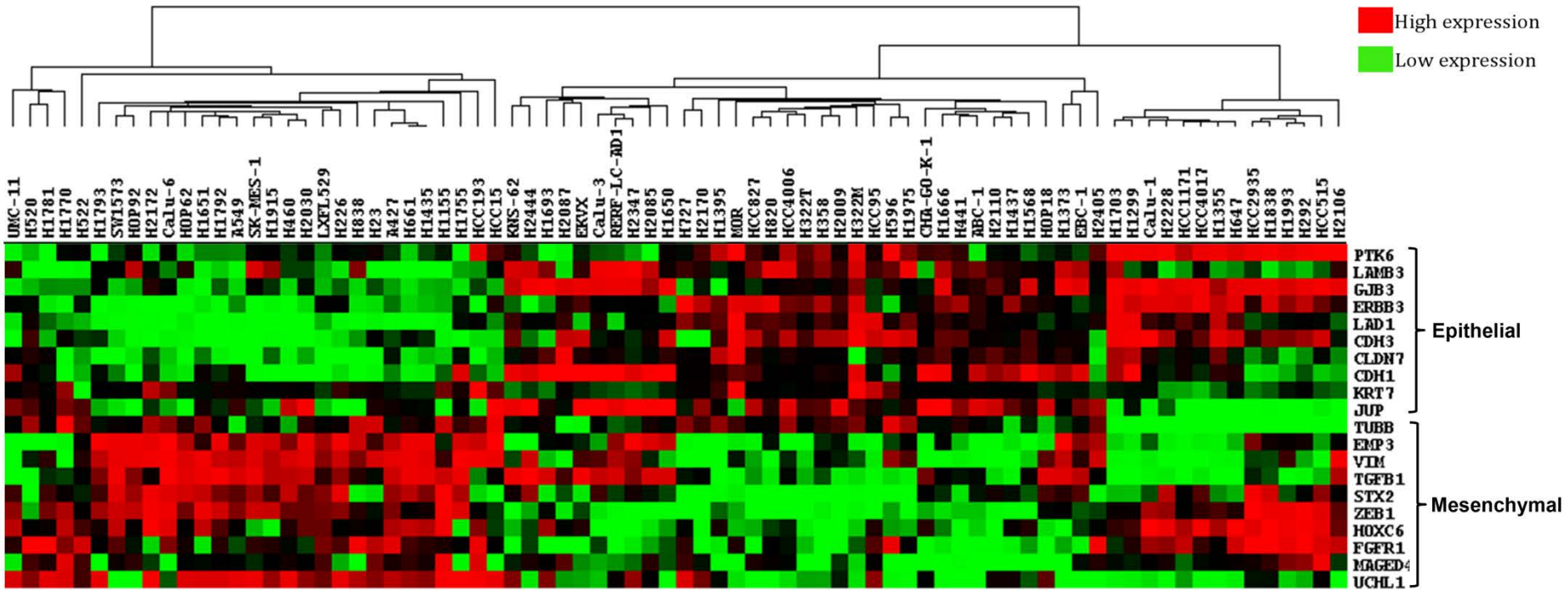


Figure 4-18 Effect of Endoglin on Vimentin expression in NSCLC cell lines. Vimentin expression was measured in all Endoglin transfected (ENG) and empty vector transfected (VEC) cell lines which were methylated and had therefore lost expression of the gene. Vimentin was also measured in the unmethylated cell line Hop62 (separated by the dotted line to the right of the graph) which had the highest baseline Endoglin expression, after transfection with an Endoglin-targeting siRNA (siENG) and a scrambled siRNA control (siCT). Data shown are mean ±SD of three independent experiments in which cells were treated with TGF-β1 at 10ng/ml for 24 hours. The dotted line on the left of the graph separates cell lines for which there is a significant difference between VEC and ENG cells in the presence of TGF-β, and all other cell lines for which there was no difference. HCC95 and H322M were also the only cell lines which showed a difference between VEC and ENG cells in migration and invasion assays.

A


High expression
Low expression
UMC-11
H520
H1781
H1770
H522
H1793
SW1573
HOP92
H2172
Calu-6
HOP62
H1651
H1792
A549
SK-MES-1
H1915
H460
H2030
LXFL529
H226
H838
H23
A427
H661
H1435
H1155
H1755
HCC193
HCC15
KNS-62
H2444
H1693
H2087
EKVX
Calu-3
RERF-LC-AD1
H2347
H2085
H1650
H727
H2170
H1395
MOR
HCC827
H820
HCC4006
H322T
H358
H2009
H322M
HCC95
H596
H1975
CHA-GO-K-1
H1666
H441
ABC-1
H2110
H1437
H1568
HOP18
H1373
EBC-1
H2405
H1703
H1299
Calu-1
H2228
HCC1171
HCC4017
H1355
H647
HCC2935
H1838
H1993
H292
HCC515
H2106
PTK6
LAMB3
GJB3
ERBB3
LAD1
CDH3
CLDN7
CDH1
KRT7
JUP
Epithelial
TUBB
EMP3
VIM
TGFB1
STX2
ZEB1
HOXC6
FGFR1
MAGED4
UCHL1
Mesenchymal

B

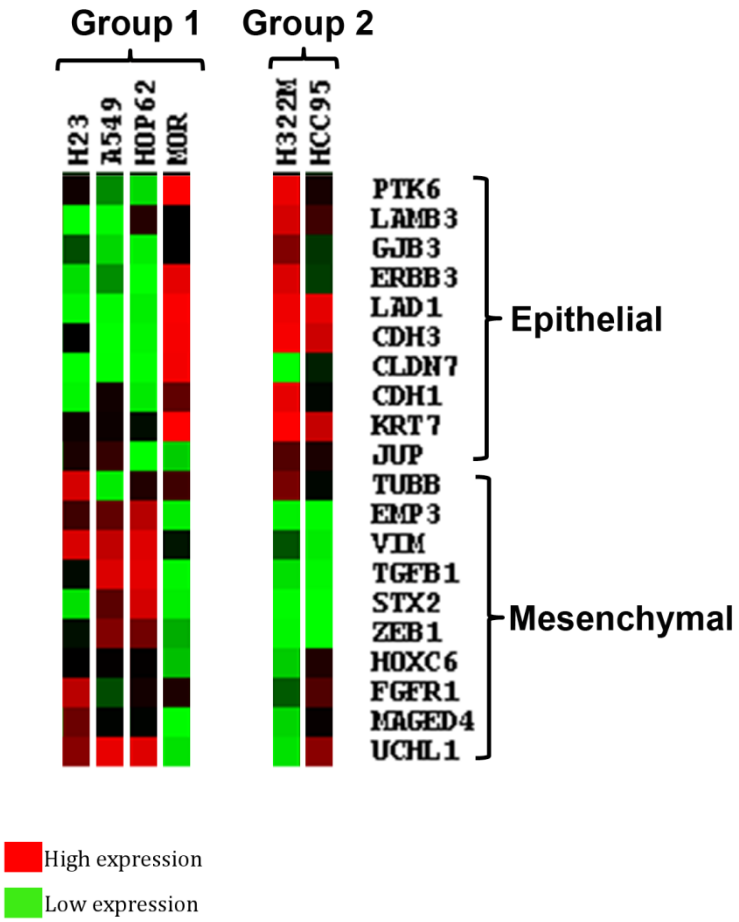
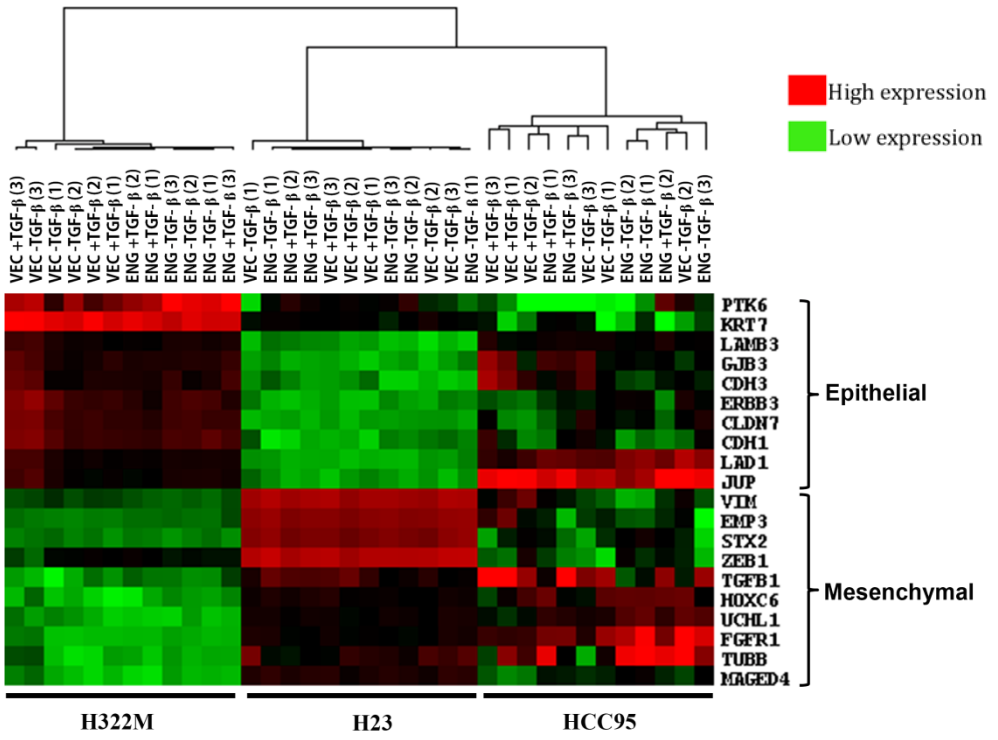
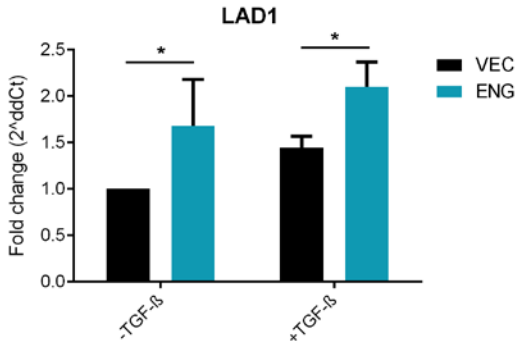


Figure 4-19 Classification of EMT status of NSCLC cell lines. (A) A 20-gene EMT expression panel has been used to classify a panel of 82 NSCLC cell lines as epithelial-like or mesenchymal-like (Walter et al. 2012). The panel included 3 of the 6 cell lines functionally analysed in this study. The remaining three, H23, A549 and Hop62 were analysed for expression of the 20 marker genes and were clustered along with the full panel to classify their baseline EMT status. (B) Group 1 cell lines are those for which there was no effect of Endoglin expression on invasive behaviour or Vimentin expression. Of these, H23, A549 and Hop62 can be classified as mesenchymal-like, while MOR is epithelial-like. Group 2 cell lines are those which show a protective effect of Endoglin expression against the pro-invasive effects of TGF-β. Within this group, the cell lines H322M and HCC95 can be classified as epithelial-like.

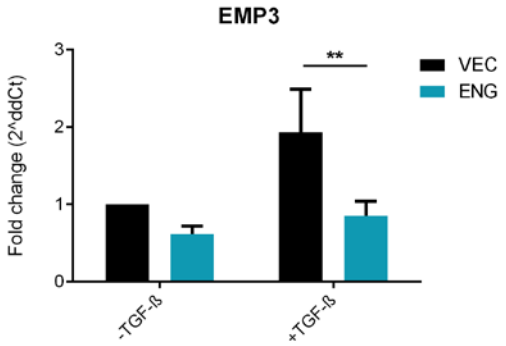
A


High expression
Low expression
VEC +TGF-β (3)
VEC -TGF-β (3)
VEC -TGF-β (1)
VEC -TGF-β (2)
VEC +TGF-β (2)
VEC +TGF-β (1)
ENG +TGF- β (2)
ENG +TGF- β (1)
ENG -TGF- β (3)
ENG -TGF- β (2)
ENG -TGF- β (1)
ENG +TGF- β (3)
VEC -TGF-β (1)
ENG -TGF- β (1)
ENG +TGF- β (2)
ENG +TGF- β (3)
VEC +TGF-β (3)
VEC +TGF-β (2)
VEC +TGF-β (1)
ENG -TGF- β (3)
ENG -TGF- β (2)
VEC -TGF-β (2)
VEC -TGF-β (3)
ENG -TGF- ß (1)
VEC +TGF-β (3)
VEC +TGF-β (1)
VEC +TGF-β (2)
ENG +TGF- β (1)
ENG +TGF- β (3)
VEC -TGF-β (3)
VEC -TGF-β (1)
ENG -TGF- β (2)
ENG -TGF- β (1)
ENG +TGF- β (2)
VEC -TGF-β (2)
ENG -TGF- β (3)
PTK6
KRT7
LAMB3
GJB3
CDH3
ERBB3
CLDN7
CDH1
LAD1
JUP
Epithelial
VIM
EMP3
STX2
ZEB1
TGFB1
HOXC6
UCHL1
FGFR1
TUBB
MAGED4
Mesenchymal
H322M
H23
HCC95

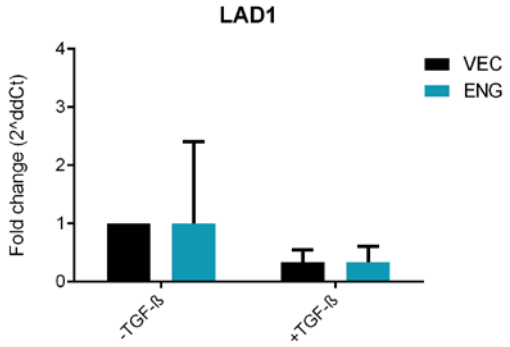
B


LAD1
Fold change (2^ddCt)
2.5
2.0
1.5
1.0
0.5
0.0
*
*
VEC
ENG
-TGF-ß
+TGF-ß

C


EMP3
Fold change (2^ddCt)
3
2
1
0
**
VEC
ENG
-TGF-ß
+TGF-ß

D


LAD1
Fold change (2^ddCt)
4
3
2
1
0
VEC
ENG
-TGF-ß
+TGF-ß

E

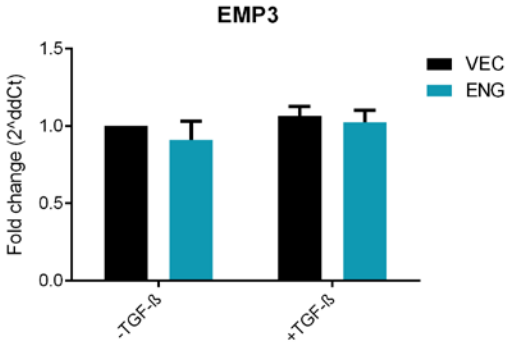

EMP3
Fold change (2^ddCt)
1.5
1.0
0.5
0.0
VEC
ENG
-TGF-ß
+TGF-ß

Figure 4-20 Effect of Endoglin on EMT status of NSCLC cell lines. The cells lines H322M, HCC95 and H23 which are methylated and have lost expression of Endoglin were transfected with empty vector (VEC) or an Endoglin-containing plasmid (ENG) and treated with TGF-β1 at 10ng/ml for 24 hours. The H23 cell line belongs to group 1 in which there was no observed functional effect of Endoglin expression. H322M and HCC95 belong to group 2, in which a functional effect of Endoglin was observed. (A) cDNA from experimental samples was analysed for expression of 20 EMT marker genes. Samples were clustered using –dCt values of marker genes. Differences between VEC and ENG cells were too subtle to be detected by cluster analysis. However differences were found at the level of individual genes for the cell lines H322M and HCC95. (B) The cell line H322M shows significant up-regulation of the epithelial marker LAD1 in ENG compared to VEC cells. (C) The cell line HCC95 shows significant up-regulation of the mesenchymal marker EMP3 both with and without TGF-β in VEC compared to ENG cells. (D, E) The group 1 cell line H23 showed no difference in EMT marker expression in ENG compared to VEC cells.

4.3 Discussion

I have shown here for the first time that Endoglin is epigenetically silenced in NSCLC cell lines and in two independent NSCLC primary tissue cohorts at frequencies of 43% and 24%, respectively. Furthermore, Endoglin methylation is of prognostic relevance in early stage cancers. The epithelial-like cell lines HCC95 and H322M demonstrate increased invasiveness and EMT progression in cells lacking Endoglin, which may explain the observed survival difference in patients with stage I cancers which may not have undergone EMT.

In primary tissues, 43% and 24% of cohort 1 and cohort 2, respectively, were found to be hypermethylated for Endoglin. This is a higher proportion of methylation than what has been reported in other studies on solid tumours; Henry, Johnson *et al.* (2011) reported 17% (n=129) of breast cancer tumours hypermethylated for Endoglin and in a colon cancer study 9% (n=34) of tumours were hypermethylated (Mori *et al.* 2006). In contrast to this, our data show a smaller proportion of methylated samples than the only other study to have looked at methylation in lung, where 69% of samples were methylated (Dammann *et al.* 2005) . However it is important to note that only 16 primary tumour samples were analysed for Endoglin promoter methylation in this study.

I have reported here a trend towards shorter survival for patients with Endoglin promoter methylation. This is in agreement with the study by Henry, Johnson *et al.* (2011) on breast cancer. The authors reported significantly lower metastasis-free and overall survival in patients lacking Endoglin expression, which would corroborate our findings of shorter survival with Endoglin methylation. While the trend is the same, we are the first to show a stage-dependent effect of Endoglin.

The failure to reach significance levels here in both cohorts individually may be due to a number of reasons, mainly involving the nature of the sample material. While FFPE tissues provide an invaluable and large-scale source of primary tissues for analysis, they are likely to be degraded to a certain degree, leading to lower amplification efficiency by PCR and an inherently higher variability for this reason (Gillio-Tos *et al.* 2007). This means that the false positive rate for methylation classification may be higher with this type of sample than with fresh frozen tissue. Despite these issues we see a definite survival trend that is replicated in two cohorts and reaches significance in a subset of one (stage 1 patients of cohort 2, generalised Wilcoxon p=0.011, Figure 4-8D). It would be of interest to analyse Endoglin methylation in a collection of fresh-frozen tissue samples to see if the improved sample quality would yield a similar but more significant result.

Another important issue which will affect significance of the data considerably is the specific choice of cut-off for methylation. The optimal cut-off value for methylation is the one that is most informative and accurate regarding outcome. Large-scale quantitative analysis of methylation has only been made possible in recent years and there has yet to be a consensus agreement or a gold standard as to how to generate a cut-off value for positivity. Published methods vary considerably and range from simply testing multiple cut-offs and choosing the most significant according to the log-rank test (Kang *et al.* 2013), defining the median methylation of tumour tissues as the cut-off, with everything above this classed as hypermethylated (Sigalotti *et al.*), using random groupings such as low<33%, medium 33-66%, high>66% (Al-Moundhri *et al.* 2010) or using mean of control tissues with or without standard deviations (Jover *et al.* 2011; Tanaka *et al.* 2011). I decided here that methylation levels of control tissues are highly informative and should be taken into account for

cut-off generation, but that a generous margin of 3 standard deviations should be used due to the higher margin of error with FFPE samples.

It appears that Endoglin methylation has a higher prognostic value in early stage lung cancers. This is not surprising given that it has been shown here how Endoglin methylation may drive EMT in these cells. Nor is it surprising given the recent findings of Jin *et al.*(2013) who found that Endoglin methylation represents a field defect in esophageal cancer, whereby frequency of methylation increases sequentially during progression from normal tissue to the pre-malignant states of Barrett metaplasia and dysplasia, to cancer. It would be worth investigating whether Endoglin is more frequently methylated in high risk or precancerous lung tissues compared to normal specimens. This is the first study to examine the effect of Endoglin methylation on survival of NSCLC patients, and information relating to other cancers is limited to the aforementioned breast cancer study by Henry *et al.* (2011). Therefore further research is warranted to clarify the prognostic value of this gene in lung cancer and other tumour types.

Reported here is a functional role for Endoglin which supports the survival trend seen in primary tissues. In certain cells, Endoglin expression protects against the pro-invasive effects of TGF-β. This is in agreement with the findings of Craft *et al.* (2007), Liu *et al.* (2002), Henry *et al.* (2011), Mano *et al.* (2011) and Perez-Gomez *et al.* (2007) on Endoglin function in prostate cancer, breast cancer, invasive trophoblasts and mouse keratinocytes, respectively. In 2/5 cell lines, absence of Endoglin led to increased invasion in a three-dimensional cell culture model (Figure 4-14C), and a transwell invasion model (Figure 4-15C), as well as a transition towards a more mesenchymal phenotype, as evidenced by expression levels EMT marker genes (Figure 4-20B, C). In contrast to both of these findings was the opposite effect

seen in wound healing scratch assays for migration, whereby the very cells which were more invasive and expressing markers of EMT progression were migrating less, and vice versa. Since the scratch assay does not involve a barrier such as an extracellular matrix, it was shown that this lowered migratory capacity of invasive cells may simply be a reflection of a lower proliferation rate, rather than a reduced capacity for one of the key determinants of metastatic behaviour. Cell lines in group 2 (H322M, HCC95) which show an effect of Endoglin expression on invasion and EMT marker expression, are the only within our panel that show a reduced proliferation rate in Endoglin-silenced cells compared to Endoglin-positive cells (Figure 4-17 A, B).

While this is the first report that such a phenotype can be induced by loss of Endoglin expression, reduced proliferation with acquired invasiveness has been reported elsewhere in the literature. In basal cell carcinoma primary tissues, infiltrative cells at the invasive front of the tumour showed a lack of proliferation mediated by an up-regulation of the cell cycle inhibitory protein p16 (Svensson *et al.* 2003). A study on primary colorectal cancer tissues showed the same result and again implicated $p16^{INK4a}$ and downstream cyclin D1/Rb signalling (Palmqvist *et al.* 2000). A similar phenotype was observed in a mouse *in vivo* model of glioblastoma, with the difference that in this case the phenotype was shown to be mediated by a decrease in Erk and an increase in Akt activation, leading to acquired stem-cell-like properties of invasive cells (Molina *et al.* 2010). These invasive cells with decreased proliferation have the capacity to resume growth upon migrating to a new environment, and during the transient non-proliferative stage, may be refractory to therapeutic interventions which target dividing cells (Bolteus *et al.* 2001). In this case, p16 was ruled out as the mediator of growth inhibition in Endoglin-silenced cells (Figure 4-17C). Investigation

of a wider panel of cell cycle regulators would be of interest, but was beyond the scope of the current study.

It is probably no coincidence that the above studies which describe the non-proliferative, stem-like phenotype of actively invading cells have all done so via analysis of either primary tissues or *in vivo* models. This highlights the need for more biologically relevant assays which mimic as closely as possible the environmental conditions *in vivo.* The contrasting results presented here from the monolayer scratch assay versus the three-dimensional culture model further emphasize the importance of this. It is now widely recognised that three dimensional culture models using extracellular matrix derived from Engelbreth-Holm-Swarm mouse sarcoma cells (in this case Matrigel from BD Biosciences) recapitulate as closely as possible the *in vivo* tumour environment and are valuable tools for analysis of tumour behaviour. Specific 3D in vitro models continue to be developed, optimised and adapted to high throughput analysis (Lee *et al.* 2007; Vinci *et al.* 2012; Yu and Machesky 2012). The specific 3D assay used here was adapted from that of Vinci *et al.* (2012) for a number of reasons. It allows invasion to occur in three dimensions, rather than simply through a layer on top of which cells are seeded. This protocol allows monitoring of invasion over a period of days rather than hours, and progress can be visualised continually. Most importantly, this assay involves cell growth in a manner which replicates as closely as possible the tumour microenvironment *in vivo*, complete with cell-cell interactions, extracellular matrix, and nutrient and oxygen gradients (Vinci *et al.* 2012).

Since we did not see an effect of Endoglin in all cell lines analysed, it would seem that Endoglin alone is not sufficient to drive invasion and EMT. Other factors are at play here and it may be that the observed functional role of Endoglin is dependent

upon the existence of a specific molecular context. In breast cancer cells, it was found that Endoglin only had an effect on invasion in the presence of activated ERBB2 (Henry *et al.* 2011). This EGFR family member can be overexpressed in lung cancers (Heinmoller *et al.* 2003), but it is not classed as an oncogenic driver in these cancers. EGFR (ERBB1) however plays a crucial role in lung cancer progression as described earlier. However interrogation of microarray data for expression of all EGFR members 1-4 in our panel of cell lines does not support a case for cooperation of these with Endoglin in NSCLC (Supplementary Figure S- 1).

Data presented here suggest that EMT status of the tumour may be a factor in determining whether or not Endoglin methylation has a functional and prognostic effect. No functional effect of Endoglin expression was observed in cell lines which were classified as mesenchymal according to the EMT expression panel. However the non-mesenchymal cell lines H322M and HCC95 showed an Endoglin-dependent alteration in invasiveness and expression of the invasion-associated marker Vimentin. The process of EMT occurs along a spectrum, whereby there is a gradual reduction in expression of epithelial markers and an increase in mesenchymal markers (Kalluri and Weinberg 2009). Therefore it is reasonable to hypothesise that since loss of Endoglin may drive EMT, cells already mesenchymal in their molecular profile would not be affected by Endoglin expression levels. Conversely, cells which have an epithelial profile could be induced to progress towards a more mesenchymal phenotype upon loss of Endoglin expression. In this case we would expect to have seen an effect of Endoglin expression in the cell line MOR since this is classified as epithelial-like by cluster analysis (Figure 4-19B). Since Endoglin acts as an accessory receptor for BMP-2, BMP-7 and Activin A in addition to TGF-β (Barbara *et al.* 1999), it may be that an Endoglin-binding ligand other than TGF-β is required for a functional effect in

this cell line. Alternatively, the effect of Endoglin may be blocked in this cell line by inactivation of essential downstream effectors such as SMAD proteins. This hypothesis is worthy of further investigation in an extended cell line panel. If the additional factors determining response to Endoglin silencing in NSCLC could be determined, the combination of these may be a much more effective marker of prognosis than Endoglin methylation alone, identifying patients for whom adjuvant therapy may be worthwhile.

In summary, I report here for the first time epigenetic silencing of Endoglin as a regular event in NSCLC and a trend towards reduced survival for patients with Endoglin methylation, possibly due to the increased invasiveness of cells lacking Endoglin expression. Functional assays demonstrate that cells lacking Endoglin can adopt a phenotype of increased invasiveness, lowered proliferation capability and a shift from expression of epithelial markers to mesenchymal markers. This suggests a vitally important role for TGF-β signalling in NSCLC. Further work to reveal the additional co-factors necessary for development of this cellular phenotype would illuminate vital pathways by which NSCLC cells acquire metastatic potential, and may help identify stage I patients who are more likely to relapse with metastatic disease.

5 Evaluation of ZNF655 as an epigenetically regulated candidate tumour suppressor gene.

5.1 Introduction

ZNF655 was first identified in 2005 from of a yeast two-hybrid screen to identify binding partners of the proto-oncogene protein Vav1. The most abundant splice variant of the ZNF655 gene, known as Vik-1 (Vav-interacting Kruppel-like protein-1), was isolated and characterised. Several additional putative splice variants exist, but only Vik-1 was shown to be abundantly expressed in a variety of tissues. A larger isoform known as Vik-2 was only detected in peripheral blood lymphocytes, while a third, smaller isoform known as Vik-3 was detected in all analysed tissues, albeit very weakly (Houlard *et al.* 2005).

Vik-1 belongs to the kruppel-like zinc finger (KLF) family of proteins (Houlard *et al.* 2005). These are the classical family of zinc finger proteins (ZNFs), which can bind DNA, RNA and protein (Matthews and Sunde 2002) . The amino acids necessary for DNA binding are very well conserved between KLF family members, as are the linker sequences in between ZNFs, and bear a similarity to the Drosophila Kruppel gene. Intron-exon organisation of these genes is not well conserved (Swamynathan 2010). At least three zinc fingers of the specific C_2H_2 cysteine-histidine format are required for DNA binding (Matthews and Sunde 2002), and ZNF655 has six of these (Houlard *et al.* 2005). On the other hand, protein-protein interactions can be mediated through a single zinc finger (Matthews and Sunde 2002).

Of the three splice variants of ZNF655 described above, Vik-1 and Vik-2 both possess the zinc finger domain encoded by the gene. The slightly larger Vik-2 differs only in that it contains an additional KRAB B domain. The smallest variant, Vik-3 only

shares the first 135bp (exon 1) with the other isoforms and has no zinc finger or KRAB B domain, but contains a proline-rich domain and a KRAB A domain, both of which are absent in Vik-1 and Vik-2. It also contains an unconserved region without sequence similarity to any known functional domains (Houlard *et al.* 2005), (Figure 5-1). Both KRAB A and KRAB B domains act as strong transcriptional repressors (Urrutia 2003). Proline-rich regions are capable of rapid but relatively unspecific binding to substrates. Although they don't bind DNA, they are found in many transcription factors and may be necessary for transcriptional activation (Williamson 1994).

Owing to its zinc finger structure, ZNF655 has the ability to bind DNA, RNA and protein. Thus far, only its protein-binding abilities have been explored in any detail. The ubiquitously expressed splice variant Vik-1 has been shown to directly bind Vav1 and CDK4 (Houlard *et al.* 2005). Vav1 was first identified as a proto-oncogene, whose transforming capacity resulted from production of a truncated protein and whose expression was exclusively restricted to cells of the haematopoietic lineage (Katzav *et al.* 1989). Vav1 is a guanine nucleotide exchange factor (GEF) for the Rho/Rac family of GTPases, which plays an important role in T cell signalling activities including Ca^{2+} flux, cytoskeletal reorganisation and transcriptional activation via nuclear factor of activated T cells (NFAT) (Hornstein *et al.* 2004; Katzav 2004). In their characterisation of Vik-1, Houlard *et al.* (2005) found that, when expressed separately, Vik-1 and Vav1 both negatively regulate cell cycle progression at the G1 phase. However, when the two proteins were co-expressed, the effect was reversed and an accelerated progression through G1 and into S phase was observed.

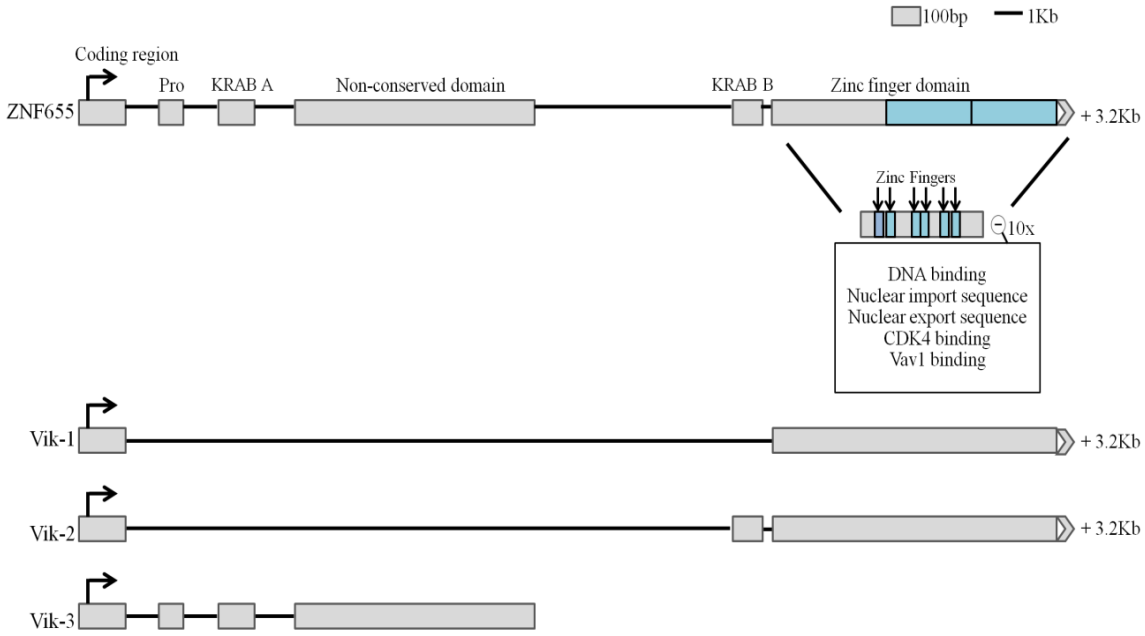


Figure 5-1 Representation of the ZNF655 gene and alternative splicing of the three isoforms Vik-1, Vik-2 and Vik-3. Adapted from (Houlard *et al.* 2005).

CDK4, the other binding partner of Vik-1, is a cyclin-dependant kinase (CDK) which acts in combination with Cyclin D to phosphorylate downstream proteins such as Rb, leading to transcription of additional factors necessary for DNA replication in S phase. CDKs are known to be stably expressed in cells, while CDK inhibitors (CKIs) which regulate their activation show variable but tightly regulated expression (Vermeulen *et al.* 2003). Several CKIs and CDKs regulate the cell cycle in a tissue-specific manner, the details of which are still being unravelled. However, it has been shown in a rat model that CDK4 is expressed in lung tissue and expression is maintained throughout pre- and postnatal development and into adulthood (Breslin *et al.* 1993).

The existence of multiple CKIs with the ability to also act on CDK4/cyclin D at the G1 stage would suggest a redundant function of Vik-1 and the likelihood of compensatory mechanisms. Indeed, it has been shown that CKIs do act in a

collaborative and tissue-specific manner to regulate cell cycle progression, leading to functional redundancy (Franklin *et al.* 2000). However, many CKIs are commonly mutated or deleted in various cancers. In particular, p16 which inhibits CDK4/6 during G1 phase, is lost in up to 50% of gliomas, mesotheliomas, nasopharyngeal, pancreatic and bilary tract tumours and up to 30% of acute lymphoblastic leukaemias (Vermeulen *et al.* 2003), 90% of SCLC and 15% of NSCLC (Cooper *et al.* 2013). Therefore, if compensatory mechanisms are already disrupted then an alternative hypothesis presents itself, whereby loss of Vik-1 through methylation of the ZNF655 gene may have an effect on proliferation of cancer cells by allowing unrestricted progression through G1 into S phase. In addition, the over-activation of CDK4 presents a possible therapeutic target in the form of CDK4 inhibitors.

The role of Vik-1 in cell cycle regulation suggests the possibility that loss of expression by epigenetic mechanisms might have an effect on cell proliferation and tumourigenesis. This has been shown to be the case for the CDK4 inhibitor p16 which is strongly associated with poor prognosis in NSCLC (Lou-Qian *et al.* 2013). The aim of this study was to first ascertain whether ZNF655 is indeed epigenetically regulated in NSCLC cell lines. Secondly, whether methylation occurs in primary lung cancer tissues and has any effect on survival. Finally, the functional implications of epigenetic loss of ZNF655 will be evaluated.

5.2 Results

5.2.1 Validation of ZNF655 as an epigenetically regulated gene in lung cancer cell lines

Methylation of ZNF655 was confirmed in an extended panel of cell lines using methylation-specific PCR (MSP) and pyrosequencing. Two MSP primer sets were used, each targeting a distinct region of the CpG island, in order to cover as wide an area of the island as possible (Figure 5-3 A). Pyrosequencing primers were designed on regions without any CpG dinucleotides so that amplification would occur irrespective of methylation status (Figure 5-3 A). Pyrosequencing primers are also designed to generate as small an amplicon as possible while still covering several CpG sites, so that they will be suitable for later use on FFPE tissues. Methylation status as measured by MSP and pyrosequencing correlated for all cell lines (Figure 5-3 B, C). No further methylated cell lines were identified in an extended panel of 18 cell lines, which included the 13 cell lines analysed by microarray plus 5 others. Corresponding to microarray data, methylation levels for NSCLC cell lines were either <10% or > 80%, with no intermediate methylation values. All cell lines with ZNF655 methylation <10% by pyrosequencing were only amplified by MSP primers targeting unmethylated sequence, while HCC95 and MOR, which had >80% methylation by pyrosequencing were only amplified by MSP primers targeting methylated sequence (Figure 5-3 B, C). Methylation status as measured by pyrosequencing and MSP correlated with methylation array results for all cell lines as indicated by linear regression analysis: R^2= 0.998, p<0.0001 (Figure 5-2).

Expression of ZNF655 mRNA was analysed by RT-qPCR. Only the highly methylated cell lines HCC95 and MOR had no detectable mRNA, suggesting a

relationship between methylation and transcriptional silencing. All other cell lines expressed the gene to varying degrees (Figure 5-4 A). Protein levels were then analysed by Western blot (Figure 5-4 B), however these were not found to correlate with either methylation or mRNA levels, with all cell lines showing protein expression.

In order to validate methylation as a direct cause of transcriptional silencing, and to ascertain whether the existing baseline expression of ZNF655 in methylated cell lines could be increased by inhibiting the epigenetic mechanisms, we treated the highly methylated cell lines MOR and HCC95 with the demethylating agent 5-aza-2'-deoxycytidine (AZA) and the HDAC inhibitor Trichostatin A (TSA). The unmethylated cell lines H322M and Hop62 were also analysed as controls (Figure 5-5). Methylated cell lines were shown to significantly up-regulate ZNF655 expression following treatment with AZA alone and AZA with TSA, although no synergistic effect of the combined treatment was observed. Surprisingly, protein expression was not up-regulated following either treatment.

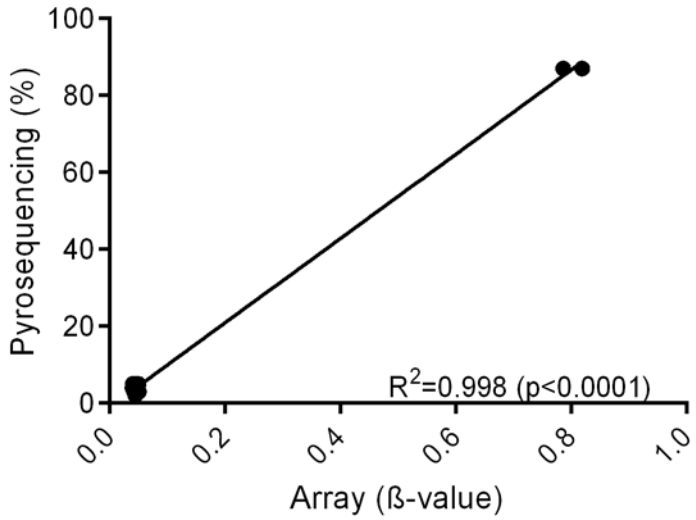
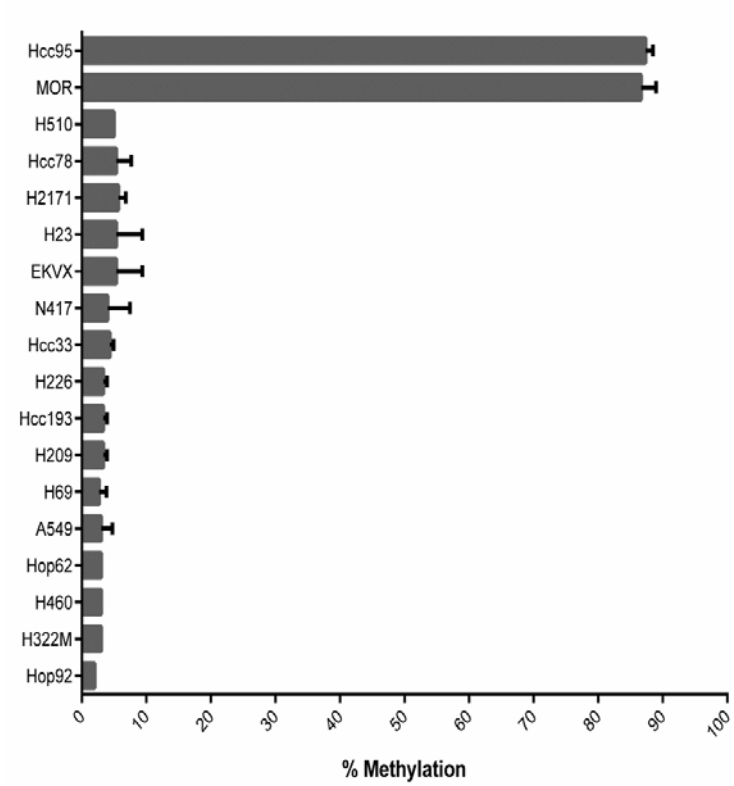


Figure 5-2 Methylation of ZNF655 in cell lines as measured by pyrosequencing and methylation array (Illumina Human 450K Methylation BeadChip). The x-axis shows the average of all beta values across the ZNF655 CpG island for each cell line. Pyrosequencing methylation values on the y-axis are the mean of biological triplicates for each cell line. Each one is calculated as the average methylation across the 8 CpG sites covered by the pyrosequencing primers. Using linear regression analysis, R square=0.998 ($p<0.0001$).

A

B



C

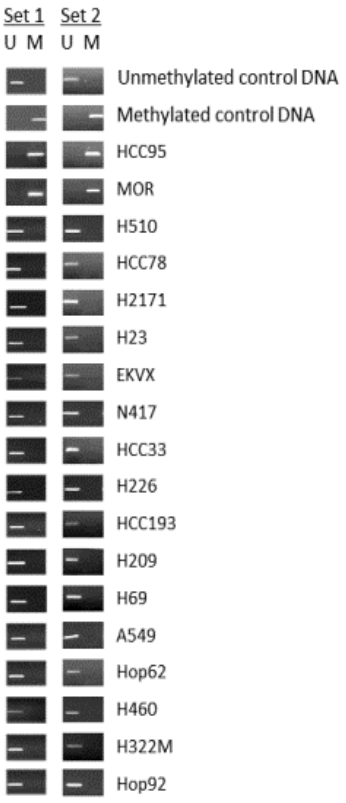
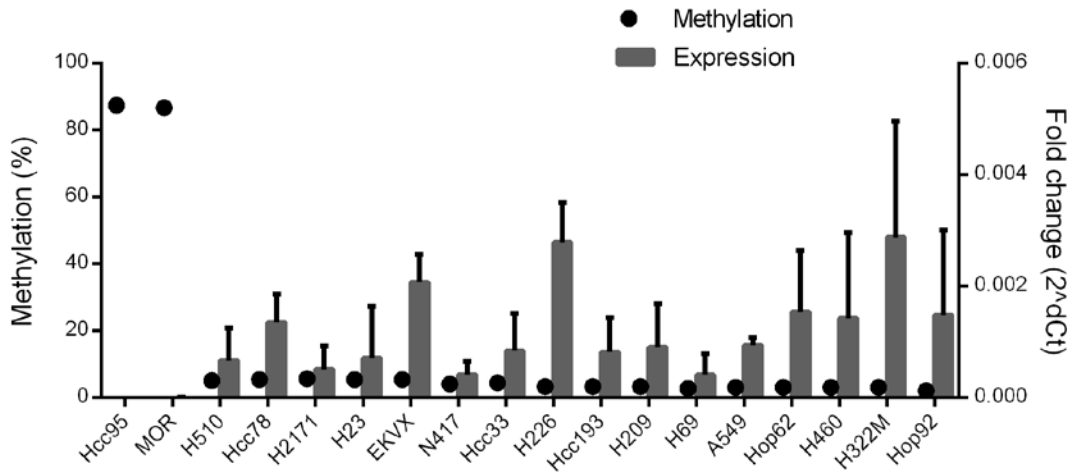


Figure 5-3 Methylation analysis of the ZNF655 promoter region. (A) Diagrammatic representation of the Endoglin gene with CpG island and primer locations. The island overlaps the 5'UTR of the gene. Individual CpG sites within the island are represented by vertical lines. Forward and reverse primers are represented by horizontal arrows. (B) ZNF655 methylation as measured by pyrosequencing. Methylation values are calculated as the average methylation across the 8 CpG sites covered by the pyrosequencing primers. Histogram bars show mean of biological triplicates with standard deviation. (C) ZNF655 methylation measured by MSP. M and U refer to PCR reactions carried out with primers targeting methylated and unmethylated sequence, respectively. Results are shown as PCR products of both reactions for each cell line.

A



B

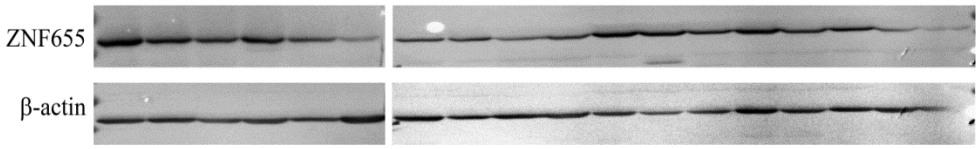


Figure 5-4 ZNF655 methylation corresponds to loss of mRNA expression. (A) ZNF655 mRNA was measured by qPCR. Histogram shows the mean 2^dCt value of 3 biological replicates with standard deviation for each cell line (right Y axis). Cell lines are shown in order of decreasing methylation, as measured by pyrosequencing and represented by black dots (left Y axis). (B) ZNF655 protein was measured by Western blot. Blot shows cell lines in same order as histogram above, with β-actin as a reference for each sample.

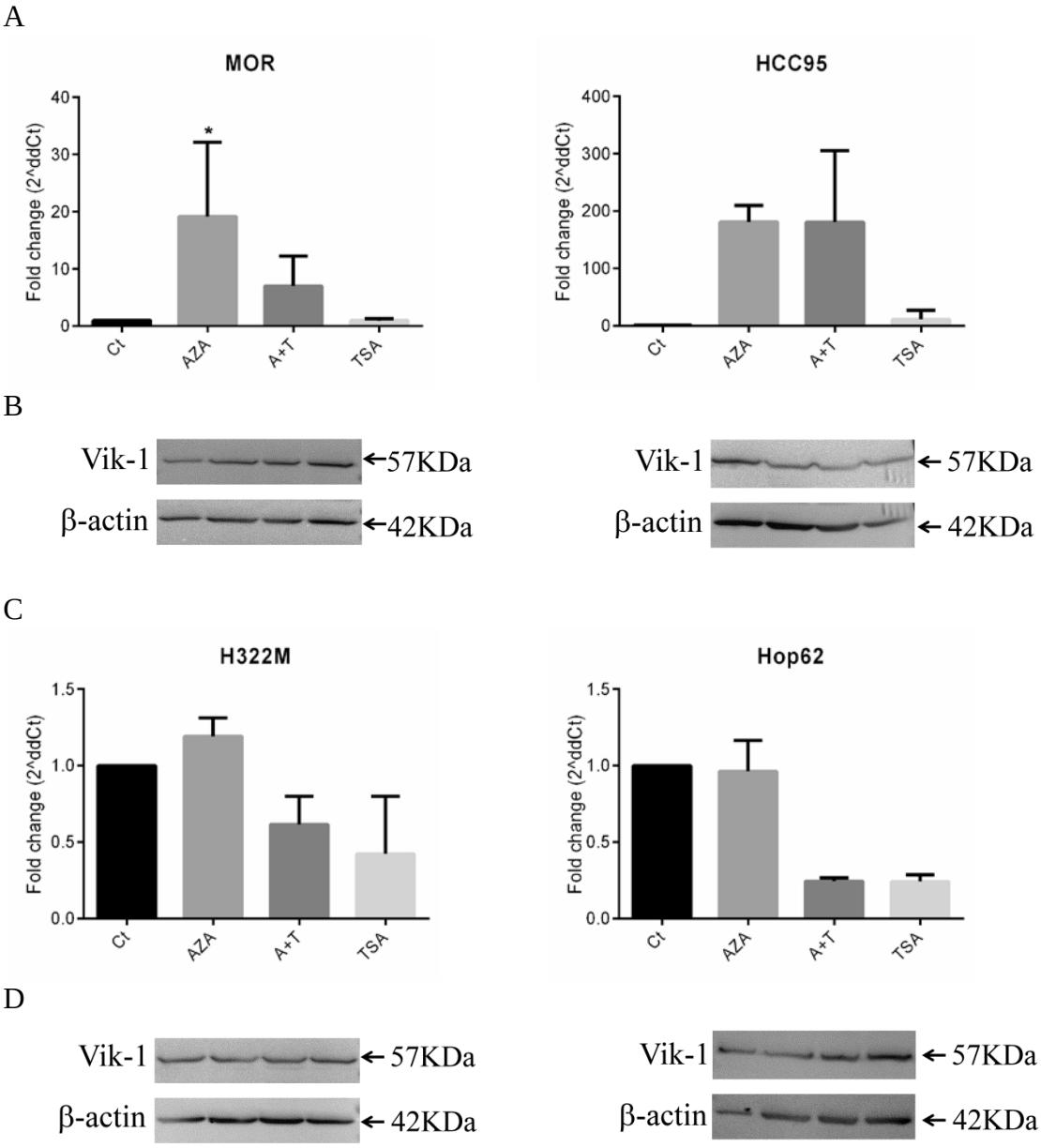


Figure 5-5 Demethylation treatment results in re-expression of ZNF655 by RT-qPCR. (A) The cell lines HCC95 and MOR, which are methylated within the ZNF655 promoter, show up-regulation of expression following treatment with AZA. No effect of TSA on expression was observed. (C) No increase in ZNF655 protein expression was observed following treatment with AZA and/or TSA in the unmethylated cell lines H322M and Hop62, suggesting a direct effect of methylation on Endoglin expression. (B, D) No increase in ZNF655 protein level was observed either in the context of methylated, re-expressing cell lines or unmethylated cell lines with unaffected RNA levels. Samples are in the same order as histograms above.

5.2.2 ZNF655 methylation in primary tissues

ZNF655 methylation was also analysed in both the NSCLC primary tissue cohorts in order to ascertain whether transcription of the gene is epigenetically controlled in these tissues. As described for Endoglin, two independent cohorts of formalin-fixed, paraffin-embedded (FFPE) tissues from biopsy or surgical resection were analysed for methylation using the pyrosequencing assay validated in our cell line panel. The characteristics of the patient cohorts are summarised in Table 5-1 (cohort 1) and Table 5-2 (cohort 2). Using the Shapiro-Wilk test for normality (Shapiro and Wilk 1965; Ghasemi and Zahediasl 2012), ZNF655 methylation data were found not to be normally distributed (cohort 1 $p=0.000$, cohort 2 $p=0.004$). Therefore non-parametric tests were used for all further analyses.

As with cell lines, methylation of primary tissues was calculated as the average across 8 CpG sites within the amplicon generated by the pyrosequencing primers. A panel of non-tumour lung tissue samples were used as controls with which to define a cut-off for methylation in tumour samples. For each cohort, 20 histologically normal, tumour-adjacent lung tissues were used as controls. Distribution of ZNF655 methylation in control tissues compared to tumour tissues for cohort 1 and cohort 2 are shown in Figure 5-6 and Figure 5-7, respectively. For each cohort, the cut-off was defined in the same way used for the Endoglin gene, i.e. as the average of control tissues plus 3 standard deviations. Applying this to our panel of tumour samples, cut-off values of 35% and 29% methylation were used for cohort 1 and cohort 2, respectively. This led to 36% of cohort 1 but only 2.5% of cohort 2 being classified as methylated for ZNF655.

To explore the relationship between clinical characteristics of lung cancer patients and ZNF655 methylation, statistical analyses were carried out to examine methylation

differences in patients according to age, gender, diagnosis, type of surgery, stage, grade and smoking status (only available for cohort 1). For cohort 1, there was no association between clinical characteristics and ZNF655 methylation (Table 5-1). For cohort 2, the association of stage (p=0.032) and grade (p=0.019) with ZNF655 methylation fell just inside significance levels (Table 5-2). This was unexpected due to the fact that there was no difference between distributions of ZNF655 methylation in tumour compared to non-tumour samples and only five samples were classified as methylated based on the cut-off value (Figure 5-7).

Patients were stratified according to ZNF655 methylation status and survival times were plotted in Kaplan-Meier curves (Figure 5-6, Figure 5-7). Although there is a clear separation of the survival curves for cohort 1 up until year 8 of follow-up, there was no statistical difference between survival curves of methylated versus unmethylated populations (generalised Wilcoxon p=0.356). Using univariate Cox regression analysis, hazard ratio (HR) = 1.218; 95% CI, 0.638-2.326, p=0.549. Since there were only 5 samples in the methylated group in cohort 2, further statistical analysis could not be carried out on this group.

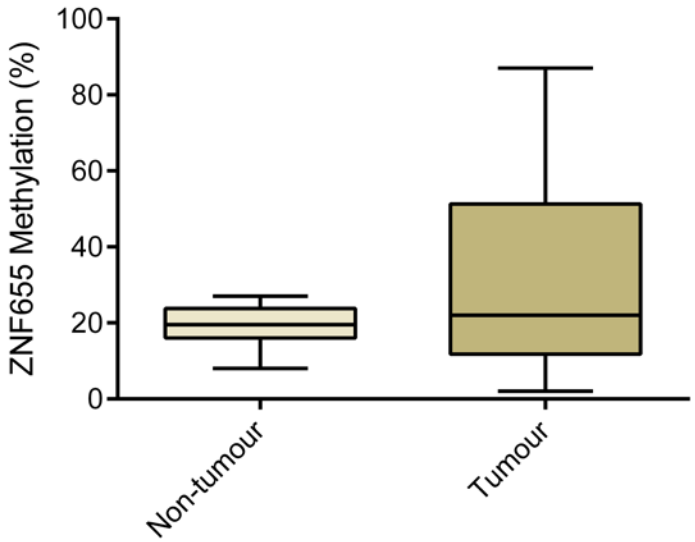
To confirm that the relationship between methylation and gene silencing observed in cell lines also exists in primary tissues, and to address the discrepancy in methylation patterns between our two cohorts, an *in silico* analysis was carried out using data from the TCGA (The Cancer Genome Atlas) database. Matched methylation (Illumina Human 450K Methylation BeadChip) and expression (Illumina HiSeq 2000 RNA sequencing) data were available for 205 lung adenocarcinoma primary tissues. Methylation values for the 13 array probes located within the ZNF655 CpG island were averaged and plotted against expression values. Expression values are normalised RNA-Seq by Expectation Maximization (RSEM) (Li and Dewey 2011)

count estimates for sequence reads mapping to the ZNF655 gene. There was a significant correlation between methylation of this probe and expression of the ZNF655 gene in the TCGA dataset; Spearman's rank correlation coefficient (r) = -0.2535 (95% CI -0.3809 to -0.1166), p=0.0002. It is also important to note that the pattern of methylation in this dataset is similar to that of cohort 2, with distribution very much skewed towards the lower range of methylation values, and only 1/205 with a beta value above 0.5. It would be of interest to link methylation and clinical data for this cohort, however clinical data are not currently available.

Basic patient characteristics Cohort 1 (N=118)				
	Number	**Methylated (n=43)**	**Unmethylated (n=75)**	***P* -value**
Age at diagnosis				
Median: 68 (36-84)				0.233
<68	54	18 (42%)	36 (48%)	
>68	64	25 (58%)	39 (52%)	
Sex				0.071
Female	30	7 (16%)	23 (31%)	
Male	88	36 (84%)	52 (69%)	
Diagnosis				0.276
Squamous	56	16 (37%)	40 (53%)	
Adenocarcinoma	37	17 (4%)	20 (27%)	
Other	15	5 (12%)	10 (13%)	
Type of surgery				0.068
Segmentectomy	4	1 (2%)	3 (4%)	
Lobectomy/Bilobectomy	102	35 (81%)	67 (89%)	
Pneumonectomy	4	3 (7%)	1 (1%)	
Post-operative stage				0.691
Stage IA	50	19 (44%)	31 (41%)	
Stage IB	56	19 (44%)	37 (49%)	
Stage II	2	1 (2%)	1 (1%)	
Tumour Grade				0.667
G1	13	4 (9%)	9 (12)	
G2	53	15 (35%)	38 (51%)	
G3	39	12 (28%)	27 (36%)	
Smoking history				0.644
Smoker	67	23 (53%)	44 (59%)	
Ex-smoker	21	8 (19%)	13 (17%)	
	11	4 (9%)	7 (9%)	

Table 5-1 Patient characteristics of cohort 1and clinicopathologic correlation of ZNF655 methylation in NSCLC. Columns show number of individuals with methylated and unmethylated ZNF655 for each characteristic, with percentages referring to proportion of the total methylated or unmethylated population. P-values relate to differences in distribution of ZNF655 methylation according to the various clinical characteristics. P-values are calculated using the independent samples Mann-Whitney U test (for variables with two groups) or the Kruskal-Wallis test (for variables with more than two groups) for one-way analysis of variance.

A



B

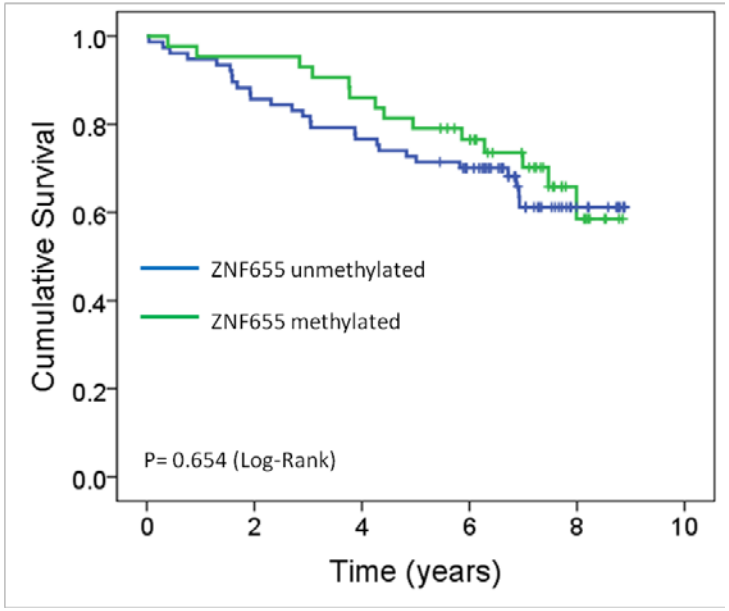
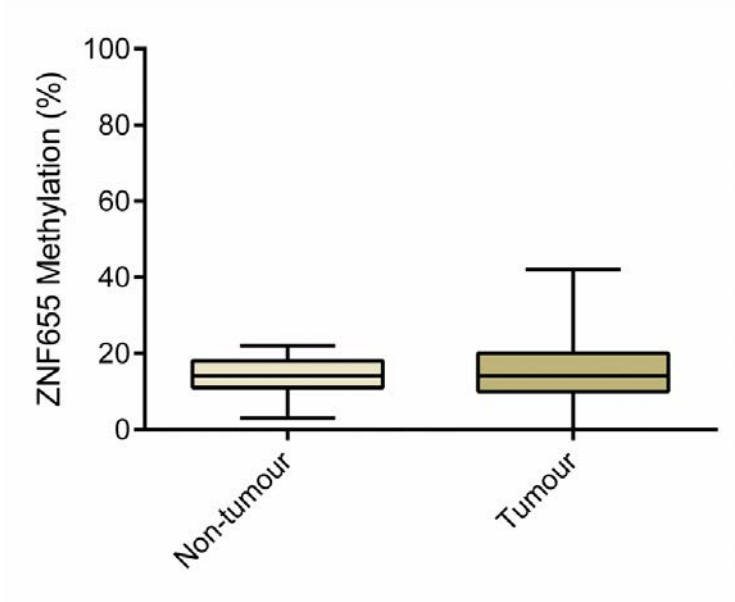


Figure 5-6 Analysis of ZNF655 methylation in cohort 1 by pyrosequencing. (A) Boxplot showing distribution of ZNF655 methylation in primary lung tumour tissues compared to control tissues in cohort 1. Control tissues are histologically normal, tumour adjacent lung tissues (n=20). (B) Kaplan-Meier survival curve showing patients from cohort 1 stratified by ZNF655 methylation. Methylated samples are defined as those with methylation value greater than the cut-off of 35%. This cut-off was generated by calculating the mean methylation of control tissues plus three standard deviations. Using this cut-off, there is no statistical difference in survival according to ZNF655 methylation (p=0.356). Using a univariate Cox proportional hazards model, hazard ratio (HR) = 1.218; 95% CI, 0.638-2.326, p=0.549.

Basic patient characteristics Cohort 2 (N=197)		
	Number	***P* –value**
Age at diagnosis		
Median: 65 (23-89)		0.382
<65	97	
≥65	100	
Sex		0.131
Female	99	
Male	98	
Diagnosis		0.218
Squamous	44	
Adenocarcinoma	123	
Other	30	
Type of surgery		0.906
Wedge	44	
Lobe/Bilobectomy	149	
Pneumonectomy	4	
Post-operative stage		0.032*
Stage IA	67	
Stage IB	43	
Stage IIA	39	
Stage IIB	19	
Stage III, IV	26	
Tumour Grade		0.019*
G1	41	
G2	97	
G3	58	

Table 5-2 Patient characteristics of cohort 2 and clinicopathologic correlation of ZNF655 methylation in NSCLC. Columns show number of individuals in each category. P-values relate to the relationship between each variable and methylation of ZNF655 and are calculated using the independent samples Mann-Whitney U test for one-way analysis of variance.

A



B

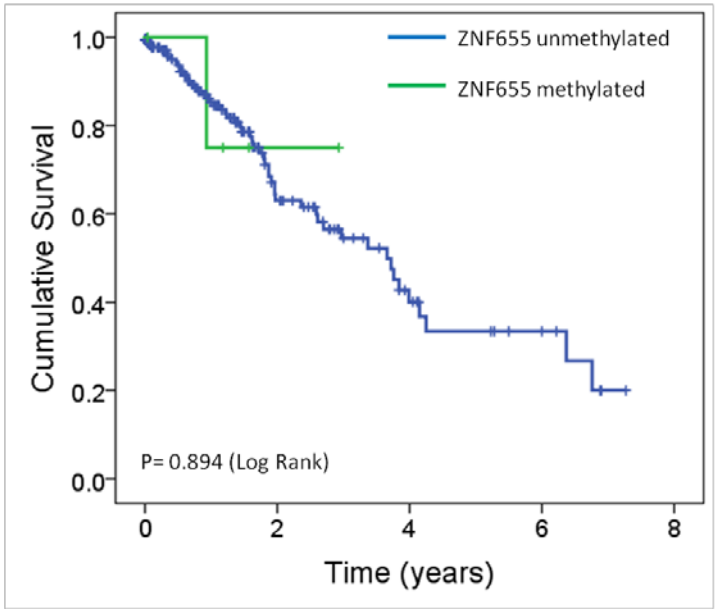


Figure 5-7 Analysis of ZNF655 methylation in cohort 2 by pyrosequencing. (A) Box plot showing distribution of ZNF655 methylation in primary lung tumour tissues compared to control tissues in cohort 2. Control tissues are histologically normal, tumour adjacent lung tissues (n=19). An independent-samples Mann-Whitney U test found there no difference in distribution of ZNF655 methylation between control and tumour samples (p=0.847). (B) Kaplan-Meier survival curve showing patients from cohort 2 stratified by ZNF655 methylation. Methylated samples are defined as those with methylation value greater than the cut-off of 29%. This cut-off was generated by calculating the mean methylation of control tissues plus three standard deviations. The lack of difference in ZNF655 methylation between the tumour and control samples, and therefore the extremely small size of the methylated population (n=5) prevented any meaningful survival analysis on this cohort.

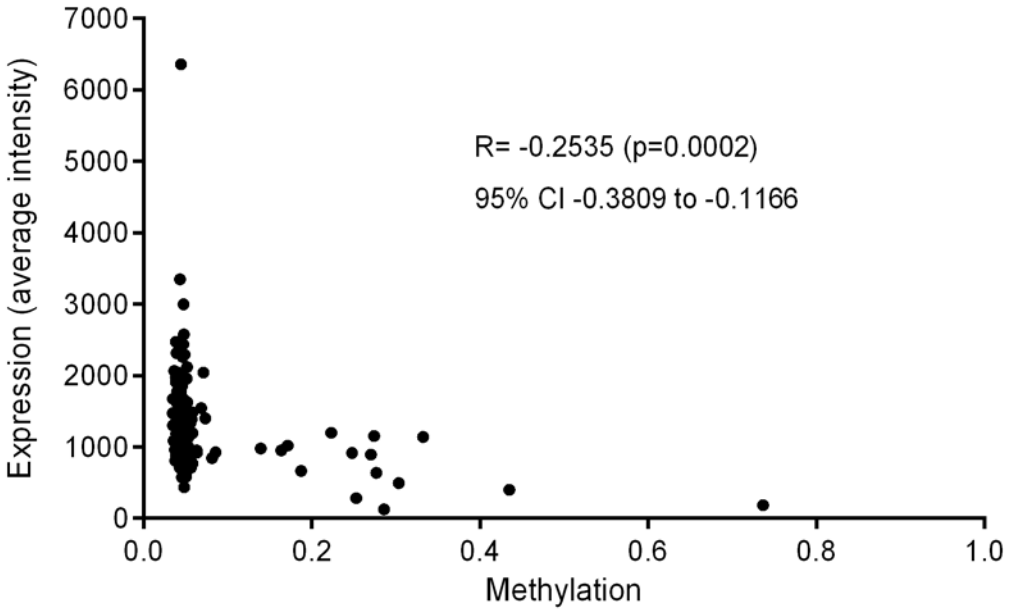


Figure 5-8 *In silico* analysis was carried out on publically available microarray data. The graph shows methylation values (Illumina Human 450K Methylation BeadChip) versus expression values (Illumina HiSeq 2000 RNA sequencing) for ZNF655 in a series of lung adenocarcinoma primary tissues (N=205) downloaded from the TCGA (The Cancer Genome Alas) database. Methylation values are the mean of all probes located within the ZNF655 CpG island (n= 13). Expression values are normalised RNA-Seq by Expectation Maximization (RSEM) count estimates for sequence reads mapping to the ZNF655 gene. There was a significant correlation between methylation and expression of ZNF655 in these primary tissues according to Spearman's rank correlation coefficient: R= -0.2684; 95% CI, -0.3945 to -0.1324, p=0.0002.

5.2.1 Functional analysis of ZNF655 in lung cancer cell lines

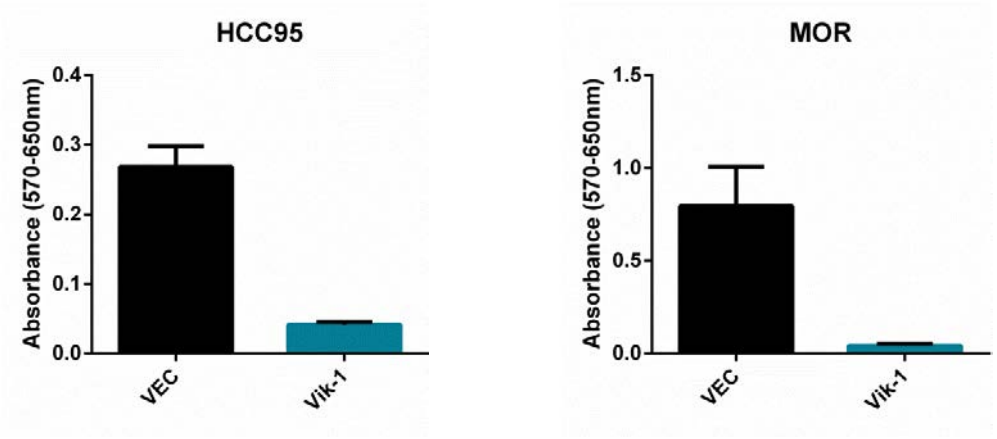
Epigenetic regulation of ZNF655 at the RNA level has been confirmed, and methylation has been shown to affect mRNA expression in primary tissues. Given the role of the most abundant transcript, Vik-1 in cell cycle regulation, it was hypothesised that ectopic expression of the Vik-1 mRNA in methylated cell lines might have an effect on cell proliferation. In order to investigate this, both methylated cell lines, HCC95 and MOR were transfected with a pcDNA3.1 plasmid containing the full Vik-1 cDNA sequence for expression from a CMV promoter.

For both methylated cell lines, cells ectopically expressing Vik-1 did not survive, while those transfected with pcDNA3.1 empty vector grew normally following stable transfection and selection (Figure 5-9). In order to ascertain whether this effect might be exclusive to cells in which there is a background of ZNF655 methylation, we transfected the unmethylated cell line, H322M. However the response was similar to methylated cell lines, demonstrating that over-expression of this cell cycle regulatory protein will not be tolerated by cells, regardless of background expression of the gene. Due to the lack of surviving stably transfected cells, further functional characterisation was not possible at this point.

Finally, since the separate roles of Vik-1 and Vav1 in cell cycle regulation have been reported to change upon their co-expression in cells (Houlard *et al.* 2005), and abnormal expression of the haematopoietically-restricted Vav1 has been reported in several cancers (discussed in section 5.3), cells were analysed for Vav1 expression by RT-qPCR. 12/18 cell lines showed detectable levels of Vav1 mRNA, to varying degrees (Figure 5-10). This is in agreement with other reports of ectopic expression of Vav1 in solid tumours. These results suggest the need for further investigation into the

effect of Vav1 ectopic expression in the context of Vik-1 methylation in NSCLC cell lines and primary tissues.

A



B

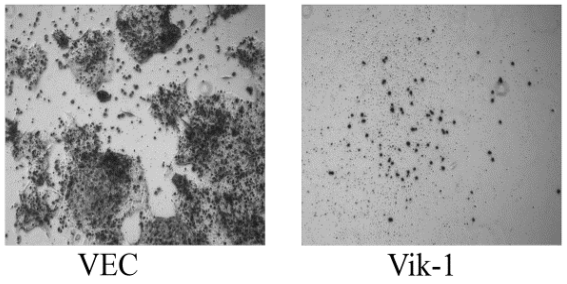


Figure 5-9 Ectopic expression of Vik-1 is lethal to cells. (A) Cells were transfected with an empty pcDNA3.1 vector (VEC) and with a pcDNA3.1 plasmid containing the Vik-1 cDNA. After 10 days of treatment with the selection marker G418, number of live cells was quantified by MTT assay. Histograms show mean corrected absorbance (750-650nm) of three independent experiments with standard deviation. (B) Representative images of MOR VEC and Vik-1 transfected cells, respectively, after selection in G418 for 10 days.

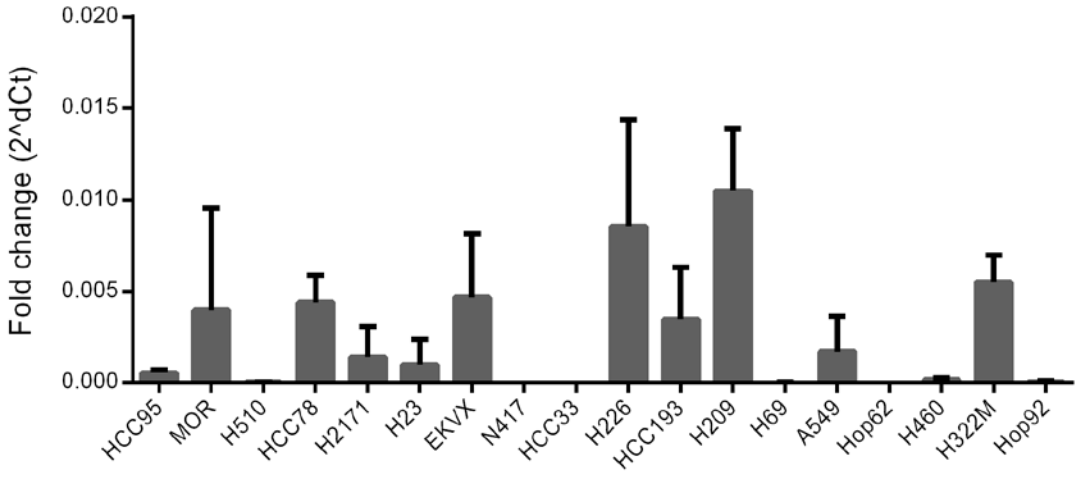


Figure 5-10 Expression of Vav1 in NSCLC cell lines. Vav1 mRNA was measured by RT-qPCR. Histogram shows the mean 2^dCt value of 3 biological replicates with standard deviation for each cell line. The normally haematopoietic-restricted Vav1 is expressed to varying degrees in these NSCLC cell lines.

5.3 Discussion

Although transcription of ZNF655 is clearly epigenetically regulated (Figure 5-4, Figure 5-5), protein levels do not correlate with mRNA levels. This prompts the immediate question: if methylation means lack of transcription and no detectable RNA, how can there be protein present in the methylated cells? One possible explanation is constitutive baseline expression of a stable, possibly inactive form of the protein. This would explain why mRNA levels may be below the detectable threshold by RT-qPCR, but the protein, having a long half-life, is readily detected by blotting. In this case, inducible up-regulation which requires binding of specific transcription factors may be inhibited my methylation of the gene promoter region. The 5'UTR of the ZNF655 gene, which overlapped is by the CpG island, contains a predicted conserved transcription factor binding site for STAT3, according to the UCSC genome browser (hg19/human assembly). STAT3 is a transcription factor whose activity can promote either growth or apoptosis in a cell-type-specific manner (Levy and Lee 2002). Methylation of the ZNF655 CpG island may prevent binding of transcription factors such as STAT3, thereby preventing inducible up-regulation. Presence of a constant baseline protein level followed by transient peaks in synthesis is an expression pattern exhibited by other cell cycle proteins, such as the CDK4-activating protein Cyclin D (Obaya and Sedivy 2002). It is therefore conceivable that ZNF655, a CDK4-inhibiting protein, may have a similar pattern of expression. However, the only published western blot of ZNF655 shows alteration of protein levels of the Vik-1 isoform along different stages of the cell cycle, which argues against constant expression of an inactive form (Houlard *et al.* 2005). In this study the authors generated their own antibody. Unfortunately there are no published Western blots of ZNF655 using commercially available antibodies.

In unmethylated cell lines, RNA and protein are both expressed but there is no correlation between the two, and significant up-regulation of mRNA following AZA treatment is not mirrored by protein levels. This is not surprising given that a recent study on over 1000 gene products in 23 cell lines showed that a significant correlation between RNA and protein existed for only one third of these genes (Gry *et al.* 2009). Differences in stability between RNA and protein as well as specific regulatory processes at each level may all be responsible for the discrepancy. For example, the 3'UTR of a gene is highly important in regulation of translation. This is due to the presence of miRNA-binding sites, as well as other factors such as the addition of a poly(A) tail, whose length is an important determinant of RNA stability and translation initiation (Bava *et al.*) (Doran 2008). Generally, a longer 3'UTR means reduced translation (Bava *et al.*), and average 3'UTR length is ~750 nucleotides for human genes (Hesketh 2004). The 3'UTR of ZNF655 is therefore much longer than average at >2.5kb, implying that this mRNA molecule might be tightly restricted in its translation.

In cohort 1, 36% of the population have ZNF655 methylation levels above the cut-off. While there is no significant difference between survival times of methylated and unmethylated populations, there are a number of interesting points to note. Firstly, in the initial 5 years of follow-up there is a clear separation of the two survival curves. Secondly, there are no censored subjects until after this point, making data over this 0-5 year period the most reliable. Thirdly, the trend observed over this time period is not as expected. ZNF655 is unmethylated in non-tumour control samples (Figure 5-6) and has been shown to be ubiquitously expressed in a variety of non-diseased tissues (Houlard *et al.* 2005). Given the previously described role of ZNF655 in negative regulation of the cell cycle, it was expected that abnormal methylation and

transcriptional silencing might give cancer cells a growth advantage and have a negative prognostic effect; instead the opposite trend is observed.

This may be due to aberrant expression of the normally haematopoietic-restricted Vav1. Co-expression with Vav1 reverses the inhibitory effect of ZNF655 splice variant Vik-1 on movement through G1/S phase, leading to rapid cell cycle progression (Houlard *et al.* 2005). Although analysis of Vav1 expression in primary tissues was outside the scope of this project, Vav1 is expressed in many of our NSCLC cell line panel (Figure 5-10). Others have also shown expression of Vav1 in lung cancer cell lines and primary lung cancer tissues, as well as a reduction in proliferation following Vav-1 silencing (Lazer *et al.* 2009). Aberrant expression of Vav1 has been demonstrated in cancers of other tissue types such as breast (Sebban *et al.* 2013), brain (Hornstein *et al.* 2003; Garcia *et al.* 2012) and pancreas, where its expression was found to have growth stimulatory effects (Fernandez-Zapico *et al.* 2005) . Importantly, in the latter study, normal primary pancreatic tissue was found to express neither Vav1 mRNA nor protein. Vav1 was also shown to be increased 10-20 fold in breast cancer cell lines compared with a non-cancer breast cell line (Gjerstorff *et al.* 2006). The evidence described above points to a possible scenario whereby abnormal co-expression of Vav1 and Vik-1 in lung cancer cells may drive cell proliferation, and therefore loss of Vik-1 by means of epigenetic silencing would confer a survival advantage.

There is a clear lack of concordance in the frequency of ZNF655 methylation between cohort 1 and cohort 2, while the distribution of methylation in cohort 2 is similar to that of the TCGA dataset. There are a number of possible explanations for this discrepancy, some of which can be immediately ruled out in this case. For instance age is significantly related to inter-individual variability in methylation patterns.

Genome-wide increases in CpG island hypermethylation and non-CpG island hypomethylation with advancing age are hallmarks of all healthy tissue types (Christensen *et al.* 2009). However our two cohorts are adequately age-matched and so this cannot explain the variation in ZNF655 methylation. However it has been shown that DNA methylation is highly divergent among populations, due to complex gene-gene and gene-environment interactions (Fraser *et al.* 2012). For example, population-specific environmental pollutants, drugs and pathogens can all contribute to methylation patterns and disease development (Javierre *et al.* 2011), as can exposure to tobacco smoke, both pre- and post-natally (Breton *et al.* 2011).

As discussed in chapter 1, many common genetic aberrations of NSCLC are more commonly or even exclusively found in adenocarcinomas, such as KRAS and EGFR mutations. A larger cohort of patients would be needed to ascertain whether ZNF655 methylation is more common to a one histological sub-type than another. However, since cohort 2 is predominantly made up of adenocarcinomas and cohort 1 has a larger proportion of squamous cell carcinomas, this may be another explanation for the difference in methylation observed between the two cohorts.

Perhaps most important is the variability which may be introduced due to the processing of these 2 cohorts from separate institutions, as well as differences in age of the samples (cohort 1 tissues were collected between 2003-2006, cohort 2 between 2004-2011). In addition, the PCR reaction to amplify the ZNF655 region of interest prior to pyrosequencing generates a 212bp amplicon, much larger than the 77bp amplicon generated by the Endoglin assay. Stability issues surrounding FFPE tissue-derived DNA could introduce an amplification bias at the PCR stage which may lead to differences between cohorts if they are not matched in terms of DNA quality. It would be of interest to examine methylation of this region in fresh frozen tissues, as

this would greatly reduce the problems associated with DNA quality. Or alternatively, multiple pyrosequencing assays may be designed which would target smaller regions within the ZNF655 promoter region. Although costly, this would also address the issue of a PCR amplification bias.

Due to the lack of cells surviving ectopic expression of Vik-1, further functional analysis was not possible on these cells. Expression of Vik-1 in cell lines using a GFP-tagged viral expression vector would allow for sorting of Vik-1 expressing cells within hours, rather than days of transfection. In this way, cell cycle analysis of Vik-1-positive compared to Vik-1-negative cells may be possible before cell death occurs. While this was outside the scope of the current study, such experiments are planned for the future.

In conclusion, it has been shown here for the first time that ZNF655 is epigenetically regulated at the transcriptional level in NSCLC. Methylation in primary tissues has been shown to exist, but may vary between populations. Due to the low frequency of high methylation levels in primary tissues, further work will be necessary on larger sample numbers to determine whether there is any effect on survival in NSCLC patients.

6 Summary and conclusions

This study has validated the use of a high-throughput microarray screening strategy to identify novel epigenetic tumour suppressor genes in lung cancer. Identification of novel and functionally relevant candidate genes has been made possible by the use of a robust experimental design involving both gene expression and methylation arrays, together with rigorous post-array validation. In addition to this, optimisation of pyrosequencing assays for use on DNA from FFPE tissues has allowed for acquisition of reliable quantitative methylation data from valuable archival specimens.

Endoglin has been implicated as both a tumour suppressor and tumour promoter in solid cancers, but previous research into the role of Endoglin in lung cancer has been limited to the context of angiogenesis. I have shown here for the first time that Endoglin is commonly silenced by epigenetic means in lung cancer cell lines and in primary tissues at the population level, and that this silencing may allow increased invasion and EMT. Since an effect of Endoglin was not seen in all cell lines analysed, it would seem that other co-factors are necessary for this invasive phenotype. Future research should focus on deciphering these additional factors, starting with the other ligands for which Endoglin is an accessory receptor, such as BMP-2, BMP-7, activin A (Barbara *et al.* 1999).

The TGF-β pathway has long been recognised for its role in lung cancer (Gazdar *et al.* 2004; Lin *et al.* 2011; Imai *et al.* 2013; Vazquez *et al.* 2013), and small molecule TGF-β inhibitors (such as the one used here in the invasion model) are being explored as therapeutic targets (Jeon and Jen 2010; Hawinkels and Ten Dijke 2011). Therefore defining molecular signatures of cancers in which this pathway is acting as an

oncogenic driver may be of therapeutic benefit, and this study suggests that Endoglin methylation may be one such marker.

ZNF655 not well characterised and the little that is known of its function suggests an important role in cell cycle regulation (Houlard *et al.* 2005). I have shown here for the first time that ZNF655 is epigenetically regulated in lung cancer cell lines and that methylation is also present in primary tissues. The two primary tissue cohorts analysed here showed very different patterns of ZNF655 methylation. This, combined with the fact that a detailed functional analysis was prevented by the lack of cells surviving Vik-1 transfection, means that it remains to be seen whether methylation of the gene has an effect on lung cancer progression. However I have demonstrated here that loss of the cell cycle inhibitor Vik-1 may be a common event in lung cancer cell lines and primary tissues, and that Vav1, whose co-expression with Vik-1 has been shown to drive cell cycle progression, is abnormally expressed in lung cancer cell lines. Together these results suggest that the dynamics of Vik-1 and Vav-1 expression in lung cancer tissues may have contrasting effects on cell cycle regulation and warrant further investigation.

7 Supplementary Data

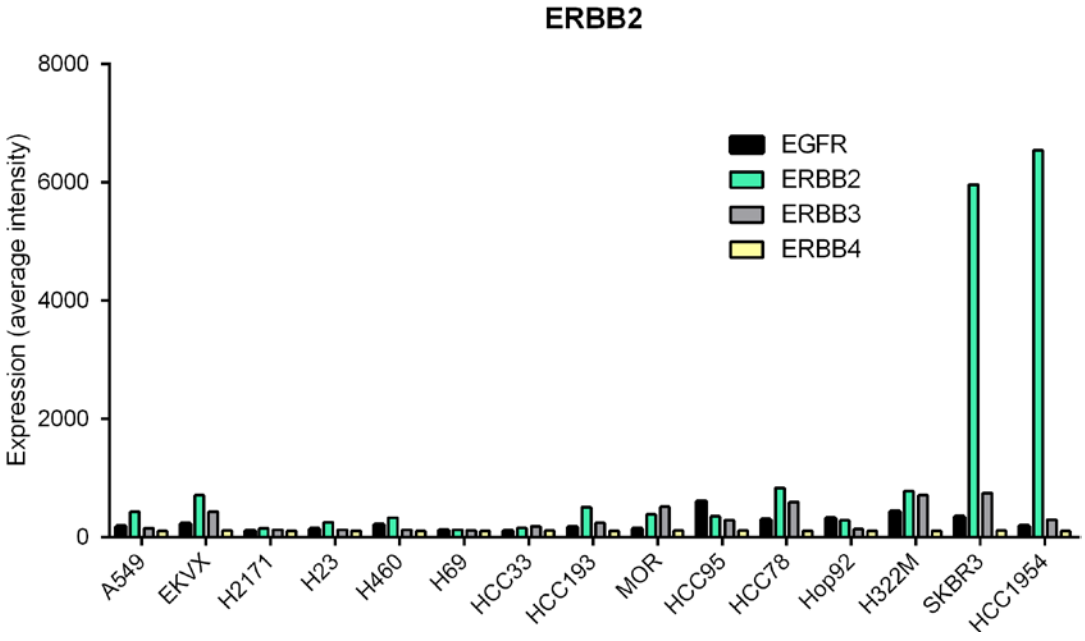


Figure S- 1 Expression of the EGFR family of proteins in 13 NSCLC cell lines, plus the two ERBB" over-expressing breast cancer cell lines SKBR3 and HCC1954 by microarray. Data shown are intensity values post-quantile normalisation, shown as the average of biological duplicates for each sample (Illumina Human HT-12 v4 Expression BeadChip). Since the breast cancer cell lines classed as overexpressing ERBB2 show intensity values of ≥3000, therefore none of this lung cancer cell line panel can be classed as overexpressing this gene or other EGFR family members.

8 References

Abdalla, S. A. and M. Letarte (2006). "Hereditary haemorrhagic telangiectasia: current views on genetics and mechanisms of disease." J Med Genet **43**(2): 97-110.

Abramoff, M., P. Magalhaes and S. Ram (2004). "Image processing with ImageJ." Biophotonics International **11**(7): 36-43.

Al-Moundhri, M. S., M. Al-Nabhani, L. Tarantini, A. Baccarelli and J. A. Rusiecki (2010). "The prognostic significance of whole blood global and specific DNA methylation levels in gastric adenocarcinoma." PLoS One **5**(12): e15585.

Anglim, P. P., T. A. Alonzo and I. A. Laird-Offringa (2008). "DNA methylation-based biomarkers for early detection of non-small cell lung cancer: an update." Mol Cancer **7**: 81.

Applied_Biosystems (2007). "Methylation analysis by bisulfite sequencing: Chemistry, products and protocols from Applied Biosystems. ."

Barbara, N. P., J. L. Wrana and M. Letarte (1999). "Endoglin is an accessory protein that interacts with the signaling receptor complex of multiple members of the transforming growth factor-beta superfamily." J Biol Chem **274**(2): 584-94.

Bartram, U. and C. P. Speer (2004). "The role of transforming growth factor beta in lung development and disease." Chest **125**(2): 754-65.

Bava, F. A., C. Eliscovich, P. G. Ferreira, B. Minana, C. Ben-Dov, R. Guigo, J. Valcarcel and R. Mendez "CPEB1 coordinates alternative 3'-UTR formation with translational regulation." Nature **495**(7439): 121-5.

Baylin, S. B. and J. G. Herman (2000). "DNA hypermethylation in tumorigenesis: epigenetics joins genetics." Trends Genet **16**(4): 168-74.

Bellon, T., A. Corbi, P. Lastres, C. Cales, M. Cebrian, S. Vera, S. Cheifetz, J. Massague, M. Letarte and C. Bernabeu (1993). "Identification and expression of two forms of the human transforming growth factor-beta-binding protein endoglin with distinct cytoplasmic regions." Eur J Immunol **23**(9): 2340-5.

Bellone, G., C. Gramigni, B. Vizio, F. A. Mauri, A. Prati, D. Solerio, L. Dughera, E. Ruffini, G. Gasparri and M. Camandona (2010). "Abnormal expression of Endoglin and its receptor complex (TGF-beta1 and TGF-beta receptor II) as early angiogenic switch indicator in premalignant lesions of the colon mucosa." Int J Oncol **37**(5): 1153-65.

Bibikova, M., B. Barnes, C. Tsan, V. Ho, B. Klotzle, J. M. Le, D. Delano, L. Zhang, G. P. Schroth, K. L. Gunderson, J. B. Fan and R. Shen (2011). "High density DNA methylation array with single CpG site resolution." Genomics **98**(4): 288-95.

Blanco, F. J., J. F. Santibanez, M. Guerrero-Esteo, C. Langa, C. P. Vary and C. Bernabeu (2005). "Interaction and functional interplay between endoglin and ALK-1, two components of the endothelial transforming growth factor-beta receptor complex." J Cell Physiol **204**(2): 574-84.

Bolstad, B. M., R. A. Irizarry, M. Astrand and T. P. Speed (2003). "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias." Bioinformatics **19**(2): 185-93.

Bolteus, A. J., M. E. Berens and G. J. Pilkington (2001). "Migration and invasion in brain neoplasms." Curr Neurol Neurosci Rep **1**(3): 225-32.

Bowden, G. T., B. Schneider, R. Domann and M. Kulesz-Martin (1994). "Oncogene activation and tumor suppressor gene inactivation during multistage mouse skin carcinogenesis." Cancer Res **54**(7 Suppl): 1882s-1885s.

Brena, R. M., C. Morrison, S. Liyanarachchi, D. Jarjoura, R. V. Davuluri, G. A. Otterson, D. Reisman, S. Glaros, L. J. Rush and C. Plass (2007). "Aberrant DNA methylation of OLIG1, a novel prognostic factor in non-small cell lung cancer." PLoS Med **4**(3): e108.

Breslin, J. S., K. S. Phillips and T. E. Weaver (1993). "Expression of the cyclin-dependent kinase cdk4 in perinatal and adult rat lung." Am J Respir Cell Mol Biol **9**(5): 533-40.

Breton, C. V., M. T. Salam and F. D. Gilliland (2011). "Heritability and role for the environment in DNA methylation in AXL receptor tyrosine kinase." Epigenetics **6**(7): 895-8.

Brewer, C. A., J. J. Setterdahl, M. J. Li, J. M. Johnston, J. L. Mann and M. E. McAsey (2000). "Endoglin expression as a measure of microvessel density in cervical cancer." Obstet Gynecol **96**(2): 224-8.

Cameron, E. E., K. E. Bachman, S. Myohanen, J. G. Herman and S. B. Baylin (1999). "Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer." Nat Genet **21**(1): 103-7.

Campling, B. G. and W. S. el-Deiry (2003). "Clinical implications of p53 mutations in lung cancer." Methods Mol Med **75**: 53-77.

Caren, H., A. Djos, M. Nethander, R. M. Sjoberg, P. Kogner, C. Enstrom, S. Nilsson and T. Martinsson (2011). "Identification of epigenetically regulated genes that predict patient outcome in neuroblastoma." BMC Cancer **11**: 66.

Cavazzoni, A., R. R. Alfieri, D. Cretella, F. Saccani, L. Ampollini, M. Galetti, F. Quaini, G. Graiani, D. Madeddu, P. Mozzoni, E. Galvani, S. La Monica, M. Bonelli, C. Fumarola, A. Mutti, P. Carbognani, M. Tiseo, E. Barocelli, P. G. Petronini and A. Ardizzoni (2012). "Combined use of anti-ErbB monoclonal antibodies and erlotinib enhances antibody-dependent cellular cytotoxicity of wild-type erlotinib-sensitive NSCLC cell lines." Mol Cancer **11**: 91.

Chi, A., S. Remick and W. Tse (2013). "EGFR inhibition in non-small cell lung cancer: current evidence and future directions." Biomark Res **1**(1): 2.

Chien, C. Y., C. Y. Su, C. F. Hwang, H. C. Chuang, Y. C. Hsiao, S. L. Wu and C. C. Huang (2006). "Clinicopathologic significance of CD105 expression in squamous cell carcinoma of the hypopharynx." Head Neck **28**(5): 441-6.

Christensen, B. C., E. A. Houseman, C. J. Marsit, S. Zheng, M. R. Wrensch, J. L. Wiemels, H. H. Nelson, M. R. Karagas, J. F. Padbury, R. Bueno, D. J. Sugarbaker, R. F. Yeh, J. K. Wiencke and K. T. Kelsey (2009). "Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context." PLoS Genet **5**(8): e1000602.

Cole, B. B., R. W. Smith, K. M. Jenkins, B. B. Graham, P. R. Reynolds and S. D. Reynolds (2010). "Tracheal Basal cells: a facultative progenitor cell pool." Am J Pathol **177**(1): 362-76.

Cooper, W. A., D. C. Lam, S. A. O'Toole and J. D. Minna (2013). "Molecular biology of lung cancer." J Thorac Dis **5**(Suppl 5): S479-S490.

Costello, J. F., M. C. Fruhwald, D. J. Smiraglia, L. J. Rush, G. P. Robertson, X. Gao, F. A. Wright, J. D. Feramisco, P. Peltomaki, J. C. Lang, D. E. Schuller, L. Yu, C. D. Bloomfield, M. A. Caligiuri, A. Yates, R. Nishikawa, H. Su Huang, N. J. Petrelli, X. Zhang, M. S. O'Dorisio, W. A. Held, W. K. Cavenee and C. Plass

(2000). "Aberrant CpG-island methylation has non-random and tumour-type-specific patterns." Nat Genet **24**(2): 132-8.

Craft, C. S., D. Romero, C. P. Vary and R. C. Bergan (2007). "Endoglin inhibits prostate cancer motility via activation of the ALK2-Smad1 pathway." Oncogene **26**(51): 7240-50.

Crino, L., W. Weder, J. van Meerbeeck and E. Felip (2010). "Early stage and locally advanced (non-metastatic) non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up." Ann Oncol **21 Suppl 5**: v103-15.

Dammann, R., M. Strunnikova, U. Schagdarsurengin, M. Rastetter, M. Papritz, U. E. Hattenhorst, H. S. Hofmann, R. E. Silber, S. Burdach and G. Hansen (2005). "CpG island methylation and expression of tumour-associated genes in lung carcinoma." Eur J Cancer **41**(8): 1223-36.

Davidson, M. R., A. F. Gazdar and B. E. Clarke (2013). "The pivotal role of pathology in the management of lung cancer." J Thorac Dis **5**(Suppl 5): S463-S478.

Dempke, W. C., T. Suto and M. Reck (2010). "Targeted therapies for non-small cell lung cancer." Lung Cancer **67**(3): 257-74.

Diederich, S. (2011). "Lung cancer screening: rationale and background." Cancer Imaging **11 Spec No A**: S75-8.

Doran, G. (2008). "The short and the long of UTRs." J RNAi Gene Silencing **4**(1): 264-5.

Erdem, O., C. Taskiran, M. A. Onan, M. Erdem, H. Guner and O. Ataoglu (2006). "CD105 expression is an independent predictor of survival in patients with endometrial cancer." Gynecol Oncol **103**(3): 1007-11.

Esteller, M. (2007). "Cancer epigenomics: DNA methylomes and histone-modification maps." Nat Rev Genet **8**(4): 286-98.

Esteller, M. (2007). "Epigenetic gene silencing in cancer: the DNA hypermethylome." Hum Mol Genet **16 Spec No 1**: R50-9.

Ferlay, J., E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J. W. Coebergh, H. Comber, D. Forman and F. Bray (2013). "Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012." Eur J Cancer **49**(6): 1374-403.

Fernandez-Ruiz, E., S. St-Jacques, T. Bellon, M. Letarte and C. Bernabeu (1993). "Assignment of the human endoglin gene (END) to 9q34-->qter." Cytogenet Cell Genet **64**(3-4): 204-7.

Fernandez-Zapico, M. E., N. C. Gonzalez-Paz, E. Weiss, D. N. Savoy, J. R. Molina, R. Fonseca, T. C. Smyrk, S. T. Chari, R. Urrutia and D. D. Billadeau (2005). "Ectopic expression of VAV1 reveals an unexpected role in pancreatic cancer tumorigenesis." Cancer Cell **7**(1): 39-49.

Finnson, K. W., W. L. Parker, Y. Chi, C. D. Hoemann, M. B. Goldring, J. Antoniou and A. Philip (2010). "Endoglin differentially regulates TGF-beta-induced Smad2/3 and Smad1/5 signalling and its expression correlates with extracellular matrix production and cellular differentiation state in human chondrocytes." Osteoarthritis Cartilage **18**(11): 1518-27.

Franklin, D. S., V. L. Godfrey, D. A. O'Brien, C. Deng and Y. Xiong (2000). "Functional collaboration between different cyclin-dependent kinase inhibitors suppresses tumor growth with distinct tissue specificity." Mol Cell Biol **20**(16): 6147-58.

Fraser, H. B., L. L. Lam, S. M. Neumann and M. S. Kobor (2012). "Population-specificity of human DNA methylation." Genome Biol **13**(2): R8.

Frommer, M., L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy and C. L. Paul (1992). "A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands." Proc Natl Acad Sci U S A **89**(5): 1827-31.

Garcia, J. L., J. Couceiro, J. A. Gomez-Moreta, J. M. Gonzalez Valero, A. S. Briz, V. Sauzeau, E. Lumbreras, M. Delgado, C. Robledo, M. L. Almunia, X. R. Bustelo and J. M. Hernandez (2012). "Expression of VAV1 in the tumour microenvironment of glioblastoma multiforme." J Neurooncol **110**(1): 69-77.

Garinis, G. A., G. P. Patrinos, N. E. Spanakis and P. G. Menounos (2002). "DNA hypermethylation: when tumour suppressor genes go silent." Hum Genet **111**(2): 115-27.

Gazdar, A. F., K. Miyajima, J. Reddy, U. G. Sathyanarayana, H. Shigematsu, M. Suzuki, T. Takahashi and N. Shivapurkar (2004). "Molecular targets for cancer therapy and prevention." Chest **125**(5 Suppl): 97S-101S.

Ghasemi, A. and S. Zahediasl (2012). "Normality tests for statistical analysis: a guide for non-statisticians." Int J Endocrinol Metab **10**(2): 486-9.

Gillio-Tos, A., L. De Marco, V. Fiano, F. Garcia-Bragado, R. Dikshit, P. Boffetta and F. Merletti (2007). "Efficient DNA extraction from 25-year-old paraffin-embedded tissues: study of 365 samples." Pathology **39**(3): 345-8.

Gjerstorff, M. F., V. M. Benoit, A. V. Laenkholm, O. Nielsen, L. E. Johansen and H. J. Ditzel (2006). "Identification of genes with altered expression in medullary breast cancer vs. ductal breast cancer and normal breast epithelia." Int J Oncol **28**(6): 1327-35.

Gluz, O., P. Wild, C. Liedtke, R. Kates, H. Mendrik, E. Ehm, V. Artinger, R. Diallo-Danebrock, E. Ting, S. Mohrmann, C. Poremba, N. Harbeck, U. Nitz, A. Hartmann and A. Gaumann (2011). "Tumor angiogenesis as prognostic and predictive marker for chemotherapy dose-intensification efficacy in high-risk breast cancer patients within the WSG AM-01 trial." Breast Cancer Res Treat **126**(3): 643-51.

Goodwin, G., J. H. Shaper, M. D. Abeloff, G. Mendelsohn and S. B. Baylin (1983). "Analysis of cell surface proteins delineates a differentiation pathway linking endocrine and nonendocrine human lung cancers." Proc Natl Acad Sci U S A **80**(12): 3807-11.

Gougos, A. and M. Letarte (1990). "Primary structure of endoglin, an RGD-containing glycoprotein of human endothelial cells." J Biol Chem **265**(15): 8361-4.

Gronbaek, K., C. Hother and P. A. Jones (2007). "Epigenetic changes in cancer." APMIS **115**(10): 1039-59.

Gry, M., R. Rimini, S. Stromberg, A. Asplund, F. Ponten, M. Uhlen and P. Nilsson (2009). "Correlations between RNA and protein expression profiles in 23 human cell lines." BMC Genomics **10**: 365.

Guerrero-Esteo, M., T. Sanchez-Elsner, A. Letamendia and C. Bernabeu (2002). "Extracellular and cytoplasmic domains of endoglin interact with the transforming growth factor-beta receptors I and II." J Biol Chem **277**(32): 29197-209.

Hammerschmidt, S. and H. Wirtz (2009). "Lung cancer: current diagnosis and treatment." Dtsch Arztebl Int **106**(49): 809-18; quiz 819-20.

Hanna, J. M. and M. W. Onaitis (2013). "Cell of origin of lung cancer." J Carcinog **12**: 6.

Hawinkels, L. J. and P. Ten Dijke (2011). "Exploring anti-TGF-beta therapies in cancer and fibrosis." Growth Factors.

Heinmoller, P., C. Gross, K. Beyser, C. Schmidtgen, G. Maass, M. Pedrocchi and J. Ruschoff (2003). "HER2 status in non-small cell lung cancer: results from patient screening for enrollment to a phase II study of herceptin." Clin Cancer Res **9**(14): 5238-43.

Henry, L. A., D. A. Johnson, D. Sarrio, S. Lee, P. R. Quinlan, T. Crook, A. M. Thompson, J. S. Reis-Filho and C. M. Isacke (2011). "Endoglin expression in breast tumor cells suppresses invasion and metastasis and correlates with improved clinical outcome." Oncogene.

Herman, J. G., J. R. Graff, S. Myohanen, B. D. Nelkin and S. B. Baylin (1996). "Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands." Proc Natl Acad Sci U S A **93**(18): 9821-6.

Hesketh, J. (2004). "3'-Untranslated regions are important in mRNA localization and translation: lessons from selenium and metallothionein." Biochem Soc Trans **32**(Pt 6): 990-3.

Hirsch, F. R., P. A. Janne, W. E. Eberhardt, F. Cappuzzo, N. Thatcher, R. Pirker, H. Choy, E. S. Kim, L. Paz-Ares, D. R. Gandara, Y. L. Wu, M. J. Ahn, T. Mitsudomi, F. A. Shepherd and T. S. Mok (2013). "Epidermal growth factor receptor inhibition in lung cancer: status 2012." J Thorac Oncol **8**(3): 373-84.

Holliday, R. (2006). "Epigenetics: a historical overview." Epigenetics **1**(2): 76-80.

Hong, K. U., S. D. Reynolds, S. Watkins, E. Fuchs and B. R. Stripp (2004). "In vivo differentiation potential of tracheal basal cells: evidence for multipotent and unipotent subpopulations." Am J Physiol Lung Cell Mol Physiol **286**(4): L643-9.

Hornstein, I., A. Alcover and S. Katzav (2004). "Vav proteins, masters of the world of cytoskeleton organization." Cell Signal **16**(1): 1-11.

Hornstein, I., E. Pikarsky, M. Groysman, G. Amir, N. Peylan-Ramu and S. Katzav (2003). "The haematopoietic specific signal transducer Vav1 is expressed in a subset of human neuroblastomas." J Pathol **199**(4): 526-33.

Houlard, M., F. Romero-Portillo, A. Germani, A. Depaux, F. Regnier-Ricard, S. Gisselbrecht and N. Varin-Blank (2005). "Characterization of VIK-1: a new Vav-interacting Kruppel-like protein." Oncogene **24**(1): 28-38.

Imai, K., Y. Minamiya, A. Goto, H. Nanjo, H. Saito, S. Motoyama, Y. Sato, S. Kudo, S. Takashima, Y. Kawaharada, N. Kurihara, K. Orino and J. Ogawa (2013). "Bronchioloalveolar invasion in non-small cell lung cancer is associated with expression of transforming growth factor-beta1." World J Surg Oncol **11**: 113.

Iyen-Omofoman, B., R. B. Hubbard, C. J. Smith, E. Sparks, E. Bradley, A. Bourke and L. J. Tata (2011). "The distribution of lung cancer across sectors of society in the United Kingdom: a study using national primary care data." BMC Public Health **11**: 857.

Jackson, E. L., N. Willis, K. Mercer, R. T. Bronson, D. Crowley, R. Montoya, T. Jacks and D. A. Tuveson (2001). "Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras." Genes Dev **15**(24): 3243-8.

Jakopovic, M., A. Thomas, S. Balasubramaniam, D. Schrump, G. Giaccone and S. E. Bates (2013). "Targeting the Epigenome in Lung Cancer: Expanding Approaches to Epigenetic Therapy." Front Oncol **3**: 261.

Javierre, B. M., H. Hernando and E. Ballestar (2011). "Environmental triggers and epigenetic deregulation in autoimmune disease." Discov Med **12**(67): 535-45.

Jemal, A., F. Bray, M. M. Center, J. Ferlay, E. Ward and D. Forman (2011). "Global cancer statistics." CA Cancer J Clin **61**(2): 69-90.

Jemal, A., R. Siegel, E. Ward, Y. Hao, J. Xu and M. J. Thun (2009). "Cancer statistics, 2009." CA Cancer J Clin **59**(4): 225-49.

Jeon, H. S. and J. Jen (2010). "TGF-beta signaling and the role of inhibitory Smads in non-small cell lung cancer." J Thorac Oncol **5**(4): 417-9.

Jin, Z., Z. Zhao, Y. Cheng, M. Dong, X. Zhang, L. Wang, X. Fan, X. Feng, Y. Mori and S. J. Meltzer (2013). "Endoglin promoter hypermethylation identifies a field defect in human primary esophageal cancer." Cancer **119**(20): 3604-9.

Jonker, L. and H. M. Arthur (2002). "Endoglin expression in early development is associated with vasculogenesis and angiogenesis." Mech Dev **110**(1-2): 193-6.

Jover, R., T. P. Nguyen, L. Perez-Carbonell, P. Zapater, A. Paya, C. Alenda, E. Rojas, J. Cubiella, F. Balaguer, J. D. Morillas, J. Clofent, L. Bujanda, J. M. Rene, X. Bessa, R. M. Xicola, D. Nicolas-Perez, A. Castells, M. Andreu, X. Llor, C. R. Boland and A. Goel (2011). "5-Fluorouracil adjuvant chemotherapy does not increase survival in patients with CpG island methylator phenotype colorectal cancer." Gastroenterology **140**(4): 1174-81.

Kalluri, R. and R. A. Weinberg (2009). "The basics of epithelial-mesenchymal transition." J Clin Invest **119**(6): 1420-8.

Kang, W. S., S. B. Cho, J. S. Park, M. Y. Lee, S. C. Myung, W. Y. Kim, S. H. Lee, D. H. Kim and E. J. Lee (2013). "Clinico-epigenetic combination including quantitative methylation value of DKK3 augments survival prediction of the patient with cervical cancer." J Cancer Res Clin Oncol **139**(1): 97-106.

Kassouf, W., H. R. Ismail, A. G. Aprikian and S. Chevalier (2004). "Whole-mount prostate sections reveal differential endoglin expression in stromal, epithelial, and endothelial cells with the development of prostate cancer." Prostate Cancer Prostatic Dis **7**(2): 105-10.

Katzav, S. (2004). "Vav1: an oncogene that regulates specific transcriptional activation of T cells." Blood **103**(7): 2443-51.

Katzav, S., D. Martin-Zanca and M. Barbacid (1989). "vav, a novel human oncogene derived from a locus ubiquitously expressed in hematopoietic cells." EMBO J **8**(8): 2283-90.

Kim, C. F., E. L. Jackson, A. E. Woolfenden, S. Lawrence, I. Babar, S. Vogel, D. Crowley, R. T. Bronson and T. Jacks (2005). "Identification of bronchioalveolar stem cells in normal lung and lung cancer." Cell **121**(6): 823-35.

Knudson, A. (2001). "Alfred Knudson and his two-hit hypothesis. (Interview by Ezzie Hutchinson)." Lancet Oncol **2**(10): 642-5.

Kopnin, B. P. (2000). "Targets of oncogenes and tumor suppressors: key for understanding basic mechanisms of carcinogenesis." Biochemistry (Mosc) **65**(1): 2-27.

Kusakabe, M., T. Kutomi, K. Watanabe, N. Emoto, N. Aki, H. Kage, E. Hamano, H. Kitagawa, T. Nagase, A. Sano, Y. Yoshida, T. Fukami, T. Murakawa, J. Nakajima, S. Takamoto, S. Ota, M. Fukayama, Y. Yatomi, N. Ohishi and D. Takai (2010). "Identification of G0S2 as a gene frequently methylated in squamous lung cancer by combination of in silico and experimental approaches." Int J Cancer **126**(8): 1895-902.

Landt, S., M. Wehling, H. Heidecke, S. Jeschke, S. Korlach, F. Stoblen, P. Schmid, J. U. Blohmer, W. Lichtenegger, J. Sehouli and S. Kummel (2011). "Prognostic significance of angiogenic factors in uterine cervical cancer." Anticancer Res **31**(8): 2589-95.

Lazer, G., Y. Idelchuk, V. Schapira, E. Pikarsky and S. Katzav (2009). "The haematopoietic specific signal transducer Vav1 is aberrantly expressed in lung cancer and plays a role in tumourigenesis." J Pathol **219**(1): 25-34.

Lebrin, F., M. J. Goumans, L. Jonker, R. L. Carvalho, G. Valdimarsdottir, M. Thorikay, C. Mummery, H. M. Arthur and P. ten Dijke (2004). "Endoglin promotes endothelial cell proliferation and TGF-beta/ALK1 signal transduction." EMBO J **23**(20): 4018-28.

Lee, G. Y., P. A. Kenny, E. H. Lee and M. J. Bissell (2007). "Three-dimensional culture models of normal and malignant breast epithelial cells." Nat Methods **4**(4): 359-65.

Lee, N. Y. and G. C. Blobe (2007). "The interaction of endoglin with beta-arrestin2 regulates transforming growth factor-beta-mediated ERK activation and migration in endothelial cells." J Biol Chem **282**(29): 21507-17.

Lee, N. Y., B. Ray, T. How and G. C. Blobe (2008). "Endoglin promotes transforming growth factor beta-mediated Smad 1/5/8 signaling and inhibits endothelial cell migration through its association with GIPC." J Biol Chem **283**(47): 32527-33.

Lee, S. Y., Y. D. Hong, P. M. Felipe, M. S. Pyun and S. J. Choi (2009). "Radiolabeling of monoclonal anti-CD105 with (177)Lu for potential use in radioimmunotherapy." Appl Radiat Isot **67**(7-8): 1366-9.

Levy, D. E. and C. K. Lee (2002). "What does Stat3 do?" J Clin Invest **109**(9): 1143-8.

Li, B. and C. N. Dewey (2011). "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome." BMC Bioinformatics **12**: 323.

Li, C., B. Guo, P. B. Wilson, A. Stewart, G. Byrne, N. Bundred and S. Kumar (2000). "Plasma levels of soluble CD105 correlate with metastasis in patients with breast cancer." Int J Cancer **89**(2): 122-6.

Li, C. G., P. B. Wilson, C. Bernabeu, U. Raab, J. M. Wang and S. Kumar (1998). "Immunodetection and characterisation of soluble CD105-TGFbeta complexes." J Immunol Methods **218**(1-2): 85-93.

Li, D. Y., L. K. Sorensen, B. S. Brooke, L. D. Urness, E. C. Davis, D. G. Taylor, B. B. Boak and D. P. Wendel (1999). "Defective angiogenesis in mice lacking endoglin." Science **284**(5419): 1534-7.

Li, S. L., D. L. Gao, Z. H. Zhao, Z. W. Liu, Q. M. Zhao, J. X. Yu, K. S. Chen and Y. H. Zhang (2007). "Correlation of matrix metalloproteinase suppressor genes RECK, VEGF, and CD105 with angiogenesis and biological behavior in esophageal squamous cell carcinoma." World J Gastroenterol **13**(45): 6076-81.

Lin, M., D. J. Stewart, M. R. Spitz, M. A. Hildebrandt, C. Lu, J. Lin, J. Gu, M. Huang, S. M. Lippman and X. Wu (2011). "Genetic variations in the transforming growth factor-beta pathway as predictors of survival in advanced non-small cell lung cancer." Carcinogenesis **32**(7): 1050-6.

Liu, S. V., M. Fabbri, B. J. Gitlitz and I. A. Laird-Offringa (2013). "Epigenetic therapy in lung cancer." Front Oncol **3**: 135.

Liu, Y., B. Jovanovic, M. Pins, C. Lee and R. C. Bergan (2002). "Over expression of endoglin in human prostate cancer suppresses cell detachment, migration and invasion." Oncogene **21**(54): 8272-81.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-8.

Lou-Qian, Z., Y. Rong, L. Ming, Y. Xin, J. Feng and X. Lin (2013). "The prognostic value of epigenetic silencing of p16 gene in NSCLC patients: a systematic review and meta-analysis." PLoS One **8**(1): e54970.

Luchtenborg, M., S. P. Riaz, E. Lim, R. Page, D. R. Baldwin, E. Jakobsen, P. Vedsted, M. Lind, M. D. Peake, A. Mellemgaard, J. Spicer, L. Lang-Lazdunski and H. Moller (2013). "Survival of patients with small cell lung cancer undergoing lung resection in England, 1998-2009." Thorax.

Ma, P. C. (2012). "Personalized targeted therapy in advanced non-small cell lung cancer." Cleve Clin J Med **79 Electronic Suppl 1**: eS56-60.

MacKinnon, A. C., J. Kopatz and T. Sethi (2010). "The molecular and cellular biology of lung cancer: identifying novel therapeutic strategies." Br Med Bull **95**: 47-61.

Mano, Y., T. Kotani, K. Shibata, H. Matsumura, H. Tsuda, S. Sumigama, E. Yamamoto, A. Iwase, T. Senga and F. Kikkawa (2011). "The loss of endoglin promotes the invasion of extravillous trophoblasts." Endocrinology **152**(11): 4386-94.

Marshall, H. M., S. C. Leong, R. V. Bowman, I. A. Yang and K. M. Fong (2012). "The science behind the 7th edition Tumour, Node, Metastasis staging system for lung cancer." Respirology **17**(2): 247-60.

Massague, J. (1998). "TGF-beta signal transduction." Annu Rev Biochem **67**: 753-91.

Matsumura, S., I. Imoto, K. Kozaki, T. Matsui, T. Muramatsu, M. Furuta, S. Tanaka, M. Sakamoto, S. Arii and J. Inazawa (2012). "Integrative array-based approach identifies MZB1 as a frequently methylated putative tumor suppressor in hepatocellular carcinoma." Clin Cancer Res **18**(13): 3541-51.

Matthews, J. M. and M. Sunde (2002). "Zinc fingers--folds for many occasions." IUBMB Life **54**(6): 351-5.

Medetoglu, B., M. Z. Gunluoglu, A. Demir, H. Melek, N. Buyukpinarbasili, N. Fener and S. I. Dincer (2010). "Tumor angiogenesis in predicting the survival of patients with stage I lung cancer." J Thorac Cardiovasc Surg **140**(5): 996-1000.

Miller, D. W., W. Graulich, B. Karges, S. Stahl, M. Ernst, A. Ramaswamy, H. H. Sedlacek, R. Muller and J. Adamkiewicz (1999). "Elevated expression of endoglin, a component of the TGF-beta-receptor complex, correlates with proliferation of tumor endothelial cells." Int J Cancer **81**(4): 568-72.

Minhajat, R., D. Mori, F. Yamasaki, Y. Sugita, T. Satoh and O. Tokunaga (2006). "Organ-specific endoglin (CD105) expression in the angiogenesis of human cancers." Pathol Int **56**(12): 717-23.

Molina, J. R., Y. Hayashi, C. Stephens and M. M. Georgescu (2010). "Invasive glioblastoma cells acquire stemness and increased Akt activation." Neoplasia **12**(6): 453-63.

Mori, Y., K. Cai, Y. Cheng, S. Wang, B. Paun, J. P. Hamilton, Z. Jin, F. Sato, A. T. Berki, T. Kan, T. Ito, C. Mantzur, J. M. Abraham and S. J. Meltzer (2006). "A genome-wide search identifies epigenetic silencing of somatostatin, tachykinin-1, and 5 other genes in colon cancer." Gastroenterology **131**(3): 797-808.

Nakamura, H. and H. Saji (2013). "A worldwide trend of increasing primary adenocarcinoma of the lung." Surg Today.

Nebbioso, A., V. Carafa, R. Benedetti and L. Altucci (2012). "Trials with 'epigenetic' drugs: an update." Mol Oncol **6**(6): 657-82.

Obaya, A. J. and J. M. Sedivy (2002). "Regulation of cyclin-Cdk activity in mammalian cells." Cell Mol Life Sci **59**(1): 126-42.

Oxnard, G. R., M. E. Arcila, J. Chmielecki, M. Ladanyi, V. A. Miller and W. Pao (2011). "New strategies in overcoming acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors in lung cancer." Clin Cancer Res **17**(17): 5530-7.

Palmisano, W. A., K. K. Divine, G. Saccomanno, F. D. Gilliland, S. B. Baylin, J. G. Herman and S. A. Belinsky (2000). "Predicting lung cancer by detecting aberrant promoter methylation in sputum." Cancer Res **60**(21): 5954-8.

Palmqvist, R., J. N. Rutegard, B. Bozoky, G. Landberg and R. Stenling (2000). "Human colorectal cancers with an intact p16/cyclin D1/pRb pathway have up-regulated p16 expression and decreased proliferation in small invasive tumor clusters." Am J Pathol **157**(6): 1947-53.

Park, K. S., M. C. Liang, D. M. Raiser, R. Zamponi, R. R. Roach, S. J. Curtis, Z. Walton, B. E. Schaffer, C. M. Roake, A. F. Zmoos, C. Kriegel, K. K. Wong, J. Sage and C. F. Kim (2011). "Characterization of the cell of origin for small cell lung cancer." Cell Cycle **10**(16): 2806-15.

Park, S., T. A. Dimaio, W. Liu, S. Wang, C. M. Sorenson and N. Sheibani (2013). "Endoglin regulates the activation and quiescence of endothelium by participating in canonical and non-canonical TGF-beta signaling pathways." J Cell Sci **126**(Pt 6): 1392-405.

Perez-Gomez, E., N. Eleno, J. M. Lopez-Novoa, J. R. Ramirez, B. Velasco, M. Letarte, C. Bernabeu and M. Quintanilla (2005). "Characterization of murine S-endoglin isoform and its effects on tumor development." Oncogene **24**(27): 4450-61.

Perez-Gomez, E., M. Villa-Morales, J. Santos, J. Fernandez-Piqueras, C. Gamallo, J. Dotor, C. Bernabeu and M. Quintanilla (2007). "A role for endoglin as a suppressor of malignancy during mouse skin carcinogenesis." Cancer Res **67**(21): 10268-77.

Peters, S., A. A. Adjei, C. Gridelli, M. Reck, K. Kerr and E. Felip (2012). "Metastatic non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up." Ann Oncol **23 Suppl 7**: vii56-64.

Petitjean, A., E. Mathe, S. Kato, C. Ishioka, S. V. Tavtigian, P. Hainaut and M. Olivier (2007). "Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database." Hum Mutat **28**(6): 622-9.

Postiglione, L., G. Di Domenico, M. Caraglia, M. Marra, G. Giuberti, L. Del Vecchio, S. Montagnani, M. Macri, E. M. Bruno, A. Abbruzzese and G. Rossi (2005). "Differential expression and cytoplasm/membrane distribution of endoglin (CD105) in human tumour cell lines: Implications in the modulation of cell proliferation." Int J Oncol **26**(5): 1193-201.

Qiu, J. (2006). "Epigenetics: unfinished symphony." Nature **441**(7090): 143-5.

Qiu, X., H. Wu and R. Hu (2013). "The impact of quantile and rank normalization procedures on the testing power of gene differential expression analysis." BMC Bioinformatics **14**: 124.

Raab, U., P. Lastres, M. A. Arevalo, J. M. Lopez-Novoa, C. Cabanas, E. J. de la Rosa and C. Bernabeu (1999). "Endoglin is expressed in the chicken vasculature and is involved in angiogenesis." FEBS Lett **459**(2): 249-54.

Reya, T., S. J. Morrison, M. F. Clarke and I. L. Weissman (2001). "Stem cells, cancer, and cancer stem cells." Nature **414**(6859): 105-11.

Rock, J. R., M. W. Onaitis, E. L. Rawlins, Y. Lu, C. P. Clark, Y. Xue, S. H. Randell and B. L. Hogan (2009). "Basal cells as stem cells of the mouse trachea and human airway epithelium." Proc Natl Acad Sci U S A **106**(31): 12771-5.

Romani, A. A., A. F. Borghetti, P. Del Rio, M. Sianesi and P. Soliani (2006). "The risk of developing metastatic disease in colorectal cancer is related to CD105-positive vessel count." J Surg Oncol **93**(6): 446-55.

Ronaghi, M., M. Uhlen and P. Nyren (1998). "A sequencing method based on real-time pyrophosphate." Science **281**(5375): 363, 365.

Rossi, G., M. C. Mengoli, A. Cavazza, D. Nicoli, M. Barbareschi, C. Cantaloni, M. Papotti, A. Tironi, P. Graziano, M. Paci, A. Stefani, M. Migaldi, G. Sartori and G. Pelosi (2013). "Large cell carcinoma of the lung: clinically oriented classification integrating immunohistochemistry and molecular biology." Virchows Arch.

Saad, R. S., Y. L. Liu, G. Nathan, J. Celebrezze, D. Medich and J. F. Silverman (2004). "Endoglin (CD105) and vascular endothelial growth factor as prognostic markers in colorectal cancer." Mod Pathol **17**(2): 197-203.

Saintigny, P. and J. A. Burger (2012). "Recent advances in non-small cell lung cancer biology and clinical management." Discov Med **13**(71): 287-97.

Schuebel, K. E., W. Chen, L. Cope, S. C. Glockner, H. Suzuki, J. M. Yi, T. A. Chan, L. Van Neste, W. Van Criekinge, S. van den Bosch, M. van Engeland, A. H. Ting, K. Jair, W. Yu, M. Toyota, K. Imai, N. Ahuja, J. G. Herman and S. B. Baylin (2007). "Comparing the DNA hypermethylome with gene mutations in human colorectal cancer." PLoS Genet **3**(9): 1709-23.

Sebban, S., M. Farago, D. Gashai, L. Ilan, E. Pikarsky, I. Ben-Porath and S. Katzav (2013). "Vav1 fine tunes p53 control of apoptosis versus proliferation in breast cancer." PLoS One **8**(1): e54321.

Sechler, M., A. D. Cizmic, S. Avasarala, M. Van Scoyk, C. Brzezinski, N. Kelley, R. K. Bikkavilli and R. A. Winn (2013). "Non-small-cell lung cancer: molecular targeted therapy and personalized medicine - drug resistance, mechanisms, and strategies." Pharmgenomics Pers Med **6**: 25-36.

Seon, B. K., A. Haba, F. Matsuno, N. Takahashi, M. Tsujie, X. She, N. Harada, S. Uneda, T. Tsujie, H. Toi, H. Tsai and Y. Haruta (2011). "Endoglin-targeted cancer therapy." Curr Drug Deliv **8**(1): 135-43.

Shapiro, S. and M. Wilk (1965). "An analysis of variance test for normality (complete samples)." Biometrika **52**(3-4): 591-611.

Sharma, S. V., D. W. Bell, J. Settleman and D. A. Haber (2007). "Epidermal growth factor receptor mutations in lung cancer." Nat Rev Cancer **7**(3): 169-81.

Shaw, A. T., D. W. Kim, K. Nakagawa, T. Seto, L. Crino, M. J. Ahn, T. De Pas, B. Besse, B. J. Solomon, F. Blackhall, Y. L. Wu, M. Thomas, K. J. O'Byrne, D. Moro-Sibilot, D. R. Camidge, T. Mok, V. Hirsh, G. J. Riely, S. Iyer, V. Tassell, A. Polli, K. D. Wilner and P. A. Janne (2013). "Crizotinib versus chemotherapy in advanced ALK-positive lung cancer." N Engl J Med **368**(25): 2385-94.

Shi, W., J. Xu and D. Warburton (2009). "Development, repair and fibrosis: what is common and why it matters." Respirology **14**(5): 656-65.

Siegel, R., D. Naishadham and A. Jemal (2012). "Cancer statistics, 2012." CA Cancer J Clin **62**(1): 10-29.

Sigalotti, L., E. Fratta, E. Bidoli, A. Covre, G. Parisi, F. Colizzi, S. Coral, S. Massarut, J. M. Kirkwood and M. Maio "Methylation levels of the "long interspersed nucleotide element-1" repetitive sequences predict survival of melanoma patients." J Transl Med **9**: 78.

Song, H., E. Yao, C. Lin, R. Gacayan, M. H. Chen and P. T. Chuang (2012). "Functional characterization of pulmonary neuroendocrine cells in lung development, injury, and tumorigenesis." Proc Natl Acad Sci U S A **109**(43): 17531-6.

Sorensen, M., M. Pijls-Johannesma and E. Felip (2010). "Small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up." Ann Oncol **21 Suppl 5**: v120-5.

Sutherland, K. D. and A. Berns (2010). "Cell of origin of lung cancer." Mol Oncol **4**(5): 397-403.

Svensson, S., K. Nilsson, A. Ringberg and G. Landberg (2003). "Invade or proliferate? Two contrasting events in malignant behavior governed by p16(INK4a) and an intact Rb pathway illustrated by a model system of basal cell carcinoma." Cancer Res **63**(8): 1737-42.

Swamynathan, S. K. (2010). "Kruppel-like factors: three fingers in control." Hum Genomics **4**(4): 263-70.

Taby, R. and J. P. Issa (2010). "Cancer epigenetics." CA Cancer J Clin **60**(6): 376-92.

Takahashi, N., R. Kawanishi-Tabata, A. Haba, M. Tabata, Y. Haruta, H. Tsai and B. K. Seon (2001). "Association of serum endoglin with metastasis in patients with colorectal, breast, and other solid tumors, and suppressive effect of chemotherapy on the serum endoglin." Clin Cancer Res **7**(3): 524-32.

Takamochi, K., S. Oh and K. Suzuki (2013). "Differences in and mutation spectra in lung adenocarcinoma of never and heavy smokers." Oncol Lett **6**(5): 1207-1212.

Takase, Y., K. Kai, M. Masuda, M. Akashi and O. Tokunaga (2010). "Endoglin (CD105) expression and angiogenesis status in small cell lung cancer." Pathol Res Pract.

Tanaka, F., Y. Otake, K. Yanagihara, Y. Kawano, R. Miyahara, M. Li, T. Yamada, N. Hanaoka, K. Inui and H. Wada (2001). "Evaluation of angiogenesis in non-small cell lung cancer: comparison between anti-CD34 antibody and anti-CD105 antibody." Clin Cancer Res **7**(11): 3410-5.

Tanaka, M., P. Chang, Y. Li, D. Li, M. Overman, D. M. Maru, S. Sethi, J. Phillips, G. L. Bland, J. L. Abbruzzese and C. Eng (2011). "Association of CHFR promoter methylation with disease recurrence in locally advanced colon cancer." Clin Cancer Res **17**(13): 4531-40.

Taskiran, C., O. Erdem, A. Onan, O. Arisoy, A. Acar, C. Vural, M. Erdem, O. Ataoglu and H. Guner (2006). "The prognostic value of endoglin (CD105) expression in ovarian carcinoma." Int J Gynecol Cancer **16**(5): 1789-93.

Thomas, L., L. A. Doyle and M. J. Edelman (2005). "Lung cancer in women: emerging differences in epidemiology, biology, and therapy." Chest **128**(1): 370-81.

Tost, J. (2009). "DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker." Methods Mol Biol **507**: 3-20.

Tost, J. and I. G. Gut (2007). "DNA methylation analysis by pyrosequencing." Nat Protoc **2**(9): 2265-75.

Travis, W., E. Brambilla, H. Muller-Hermelink and C. Harris (2004). "Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart. WHO/IASLC classification of lung and pleural tumours." IARC Press:Lyon

Travis, W. D., E. Brambilla, M. Noguchi, A. G. Nicholson, K. R. Geisinger, Y. Yatabe, D. G. Beer, C. A. Powell, G. J. Riely, P. E. Van Schil, K. Garg, J. H. Austin, H. Asamura, V. W. Rusch, F. R. Hirsch, G. Scagliotti, T. Mitsudomi, R. M. Huber, Y. Ishikawa, J. Jett, M. Sanchez-Cespedes, J. P. Sculier, T. Takahashi, M. Tsuboi, J. Vansteenkiste, I. Wistuba, P. C. Yang, D. Aberle, C. Brambilla, D. Flieder, W. Franklin, A. Gazdar, M. Gould, P. Hasleton, D. Henderson, B. Johnson, D. Johnson, K. Kerr, K. Kuriyama, J. S. Lee, V. A. Miller, I. Petersen, V. Roggli, R. Rosell, N. Saijo, E. Thunnissen, M. Tsao and D. Yankelewitz (2011). "International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma." J Thorac Oncol **6**(2): 244-85.

Uneda, S., H. Toi, T. Tsujie, M. Tsujie, N. Harada, H. Tsai and B. K. Seon (2009). "Anti-endoglin monoclonal antibodies are effective for suppressing metastasis and the primary tumors by targeting tumor vasculature." Int J Cancer **125**(6): 1446-53.

Urrutia, R. (2003). "KRAB-containing zinc-finger repressor proteins." Genome Biol **4**(10): 231.

Vaissiere, T., R. J. Hung, D. Zaridze, A. Moukeria, C. Cuenin, V. Fasolo, G. Ferro, A. Paliwal, P. Hainaut, P. Brennan, J. Tost, P. Boffetta and Z. Herceg (2009). "Quantitative analysis of DNA methylation profiles in lung cancer identifies aberrant DNA methylation of specific genes and its association with gender and cancer risk factors." Cancer Res **69**(1): 243-52.

Van Lommel, A. (2001). "Pulmonary neuroendocrine cells (PNEC) and neuroepithelial bodies (NEB): chemoreceptors and regulators of lung development." Paediatr Respir Rev **2**(2): 171-6.

Vansteenkiste, J., C. Dooms and P. De Leyn (2010). "Early stage non-small-cell lung cancer: challenges in staging and adjuvant treatment: evidence-based staging." Ann Oncol **21 Suppl 7**: vii189-95.

Vazquez, P. F., M. J. Carlini, M. C. Daroqui, L. Colombo, M. L. Dalurzo, D. E. Smith, J. Grasselli, M. G. Pallotta, M. Ehrlich, E. D. Bal de Kier Joffe and L. Puricelli (2013). "TGF-beta specifically enhances the metastatic attributes of murine lung adenocarcinoma: implications for human non-small cell lung cancer." Clin Exp Metastasis **30**(8): 993-1007.

Vermeulen, K., D. R. Van Bockstaele and Z. N. Berneman (2003). "The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer." Cell Prolif **36**(3): 131-49.

Vinci, M., S. Gowan, F. Boxall, L. Patterson, M. Zimmermann, W. Court, C. Lomas, M. Mendiola, D. Hardisson and S. A. Eccles (2012). "Advances in establishment and analysis of three-dimensional tumor spheroid-based functional assays for target validation and drug evaluation." BMC Biol **10**: 29.

Walter, K., T. Holcomb, T. Januario, P. Du, M. Evangelista, N. Kartha, L. Iniguez, R. Soriano, L. Huw, H. Stern, Z. Modrusan, S. Seshagiri, G. M. Hampton, L. C. Amler, R. Bourgon, R. L. Yauch and D. S. Shames (2012). "DNA methylation profiling defines clinically relevant biological subsets of non-small cell lung cancer." Clin Cancer Res **18**(8): 2360-73.

Weinstein, J. N., E. A. Collisson, G. B. Mills, K. R. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander and J. M. Stuart (2013). "The Cancer Genome Atlas Pan-Cancer analysis project." Nat Genet **45**(10): 1113-20.

Welti, J., S. Loges, S. Dimmeler and P. Carmeliet (2013). "Recent molecular discoveries in angiogenesis and antiangiogenic therapies in cancer." J Clin Invest **123**(8): 3190-200.

Wendt, M. K., M. Tian and W. P. Schiemann (2011). "Deconstructing the mechanisms and consequences of TGF-beta-induced EMT during cancer progression." Cell Tissue Res.

Wikstrom, P., I. F. Lissbrant, P. Stattin, L. Egevad and A. Bergh (2002). "Endoglin (CD105) is expressed on immature blood vessels and is a marker for survival in prostate cancer." Prostate **51**(4): 268-75.

Williamson, M. P. (1994). "The structure and function of proline-rich regions in proteins." Biochem J **297 (Pt 2)**: 249-60.

Wine, J. J. and N. S. Joo (2004). "Submucosal glands and airway defense." Proc Am Thorac Soc **1**(1): 47-53.

Wistuba, II, A. F. Gazdar and J. D. Minna (2001). "Molecular genetics of small cell lung carcinoma." Semin Oncol **28**(2 Suppl 4): 3-13.

Wojdacz, T. K. and A. Dobrovic (2007). "Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation." Nucleic Acids Res **35**(6): e41.

Wong, V. C., P. L. Chan, C. Bernabeu, S. Law, L. D. Wang, J. L. Li, S. W. Tsao, G. Srivastava and M. L. Lung (2008). "Identification of an invasion and tumor-suppressing gene, Endoglin (ENG), silenced by both epigenetic inactivation and allelic loss in esophageal squamous cell carcinoma." Int J Cancer **123**(12): 2816-23.

Wu, M. C., B. R. Joubert, P. F. Kuan, S. E. Haberg, W. Nystad, S. D. Peddada and S. J. London (2013). "A systematic assessment of normalization approaches for the Infinium 450k methylation platform." Epigenetics **9**(2).

Xu, C. C., L. M. Wu, W. Sun, N. Zhang, W. S. Chen and X. N. Fu (2011). "Effects of TGF-beta signaling blockade on human A549 lung adenocarcinoma cell lines." Mol Med Report.

Yagi, K., K. Akagi, H. Hayashi, G. Nagae, S. Tsuji, T. Isagawa, Y. Midorikawa, Y. Nishimura, H. Sakamoto, Y. Seto, H. Aburatani and A. Kaneda (2010). "Three DNA methylation epigenotypes in human colorectal cancer." Clin Cancer Res **16**(1): 21-33.

Yoshitomi, H., S. Kobayashi, M. Ohtsuka, F. Kimura, H. Shimizu, H. Yoshidome and M. Miyazaki (2008). "Specific expression of endoglin (CD105) in endothelial cells of intratumoral blood and lymphatic vessels in pancreatic cancer." Pancreas **37**(3): 275-81.

Yu, X. and L. M. Machesky (2012). "Cells assemble invadopodia-like structures and invade into matrigel in a matrix metalloprotease dependent manner in the circular invasion assay." PLoS One **7**(2): e30605.

Ziegler, A., U. Zangemeister-Wittke and R. A. Stahel (2002). "Circulating DNA: a new diagnostic gold mine?" Cancer Treat Rev **28**(5): 255-71.

Zijlmans, H. J., G. J. Fleuren, S. Hazelbag, C. F. Sier, E. J. Dreef, G. G. Kenter and A. Gorter (2009). "Expression of endoglin (CD105) in cervical cancer." Br J Cancer **100**(10): 1617-26.