Investigating translation dysregulation in epithelial-mesenchymal transition (EMT) in human primary lung adenocarcinoma

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Samantha SINNADURAI BSc MSc

Department of Genetics and Genome Biology

Leicester Cancer Research Centre

University of Leicester

April 2021

Abstract

Investigating translation dysregulation in epithelial-mesenchymal transition in human primary lung adenocarcinoma

By Samantha SINNADURAI

In solid tumours, metastasis is the immediate cause of most cancer-related death and is orchestrated by epithelial-to-mesenchymal transition (EMT) of tumour cells. EMT manifests as the dedifferentiation of epithelial cells into mesenchymal cells, and is typified by loss of E-cadherin and cytokeratin and gain of TWIST, N-cadherin and vimentin expression. While EMT is well-characterised in *in vitro* models, classical morphological EMT evidence (i.e. morphological dedifferentiation/sarcomatoid change) is rare in primary tumours. mRNA translation is a highly regulated process which has been linked to EMT in cell-based studies but not in primary tumours. This project aims to investigate whether I can find evidence of EMT and of translational dysregulation in EMT in primary lung adenocarcinoma. To answer these questions, we applied histological techniques such as immunohistochemistry and *in situ* hybridisation methods to quantify markers of EMT and translational control and perform statistical analysis to 942 human primary lung adenocarcinoma cases to evaluate association of EMT with tumour growth pattern, physio-pathological factors and translation dysregulation. I have shown that partial EMT state is detected in all tumour growth patterns but there is an association between partial EMT and primary solid lung adenocarcinoma. EMT can be variable within whole tumour section. EMT markers are strongly associated with physio-pathological measures of invasiveness namely lymph nodes metastasis. Furthermore, there are significant associations between EMT markers and some translation factors, particularly with the phosphorylation of eukaryotic initiation factor (eIF) 4E and the α subunit of eIF2. These data suggest that partial EMT is related to solid tumour growth pattern, and that alterations in the translation initiation machinery may play a role in activating invasive and metastatic program simultaneously in human primary lung adenocarcinoma.

Table of contents

List of Tables

List of Figures

List of abbreviations

4EBP	eIF4E Binding Protein
AAH	Atypical Adenomatous Hyperplasia
ACS	American Cancer Society
AE1/3	Cytokeratin AE1/AE3
AIS	Adenocarcinoma *In Situ*
AJCC	American Joint Committee on Cancer
ALK	Anaplastic Lymphoma Kinase
APA	Acinar Predominant Adenocarcinoma
ATF4	Activating Transcription Factor 4
ATM	Ataxia-Telangiectasia-Mutated
ATR	Ataxia-Telangiectasia and Rad3 related protein
ATS	American Thoracic Society
BRAF	v-Raf murine sarcoma viral oncogene homolog B
CCN2	Cellular Communication Network Factor 2
CCNE1	Cyclin E1
CDKN2A	Cyclin-Dependent Kinase Inhibitor 2A
CREBBP	CREB-Binding Protein
DNA	Deoxyribonucleic Acid
EGF	Epidermal Growth factor
EGFR	Epidermal growth factor receptor
eEF1/2/3	eukaryotic Elongation Factor 1/2/3
eIF	eukaryotic Initiation Factor
eIF1/1A	eukaryotic Initiation Factor 1/1A
eIF2	eukaryotic Initiation Factor 2
eIF2B	eukaryotic Initiation Factor 2b
eIF3	eukaryotic Initiation Factor 3
eIF4A1	eukaryotic Initiation Factor 4A1
eIF4A2	eukaryotic Initiation Factor 4A2
eIF4B	eukaryotic Initiation Factor 4B
eIF4E	eukaryotic Initiation Factor 4E

eIF4F	eukaryotic Initiation Factor 4F
eIF4G/4G1	eukaryotic Initiation Factor 4G/4G1
eIF4H	eukaryotic Initiation Factor 4H
eIF5/5A1/5A2	eukaryotic Initiation Factor 5/5A1/5A2
EMT	Epithelial-mesenchymal transition
EMT-TF	Epithelial-mesenchymal transition transcription factor
ER	Endoplasmic Reticulum
ERS	European Respiratory Society
ERK	Extracellular Signal-Regulated Kinase
FFPE	Formalin Fixed Paraffin Embedded
FGFR1	Fibroblast Growth Factor Receptor 1
H&E	Haematoxylin and Eosin
HRI	Heme-Regulated Inhibitor
GCN2	General control Non derepressible 2
GTP	Guanosine Triphosphate
GDP	Guanosine Diphosphate
IASLC	International Association for the study of lung cancer
IF	Immunofluorescence
IHC	Immunohistochemistry
KRAS	Kirsten Rat Sarcoma viral oncogene homolog
LPA	Lepidic Predominant Adenocarcinoma
MAPK	Mitogen Activated Protein Kinase
MDM2	Mouse double minute 2
MET	Mesenchymal-Epithelial Transition
MEK4/MKK3-6	Mitogen-Activated Protein Kinase Kinase 4/3-6
MIA	Minimally Invasive Adenocarcinoma
MMP3	Matrix Metalloproteinase-3
MPPA	Micropapillary Predominant Adenocarcinoma
MNK1/2	MAPK-Interacting Kinases 1/2
mRNA	messenger Ribonucleic Acid
MSC	Mesenchymal stem cell
mTOR	Mechanistic Target of rapamycin

N-CAM	Neural Cell-Adhesion Molecule
NF-1	Neurofibromim
NGS	Next Generation Sequencing
NICE	National Institute for Health and Care Excellence
NSCLC	Non-Small Cell Lung Cancer
ORF	Open Reading Frame
p14ARF	ARF tumour suppressor, p14ARF
p53	Tumour suppressor protein p53
p73	Tumour suppressor protein p73
PABP	Poly(A)-Binding Protein
PDAC	Pancreatic Ductal Carcinoma
PERK	Protein Kinase RNA-activated Endoplasmic Reticulum Kinase
PI3K	Phosphatidylinositol-3 Kinase
PIK3CA	Phosphatidylinositol-4, 5-biphosphate 3-Kinase
PKR	Protein Kinase RNA-activated
PTEN	Phosphatase and Tensin homolog
Raf	Rapidly Accelerated Fibrosarcoma protein kinase
Ras	Rat Sarcoma proto-oncogene protein
RASSF1	Ras Association domain-containing protein 1
RNA	Ribonucleic acid
RNAi	RNA interference
Phospho-eIF4E	Phosphorylated eukaryotic initiation factor 4E
Phospho-eIF2α	Phosphorylated eukaryotic initiation factor 2α
PPA	Papillary Predominant Adenocarcinoma
PTEN	Phosphatase and Tensin homolog
RB	Retinoblastoma gene
RET	Ref proto-oncogene
RhoA	Ras Homolog gene family member A
ROS1	ROS proto-oncogene 1
RSK	Ribosomal S6 Kinase
SCLC	Small Cell Lung Cancer
SPA	Solid Predominant Adenocarcinoma

SOX2	SRY (sex determining region Y)-box 2
TMA	Tissue Microarray
TNM	Tumour, Node, Metastasis
TP53	Tumour Protein p53
TP63	Tumour Protein p63
TRACERx	Tracking Cancer Evolution through therapy Rx
tRNA	Transfer Ribonucleic Acid
TTF-1	Thyroid Transcription Factor 1
uORF	Upstream Open Reading Frame
UTR	Untranslated Region
WHO	World Health Organisation

Chapter 1.   Introduction

Chapter 1: Introduction

1.1   Project introduction

Cancer is the leading cause of death worldwide, and for solid tumours this is mostly due to their ability to invade into normal tissues and metastasise. It is believed that epithelial-mesenchymal transition (EMT) generally underpins invasion and metastasis. As a pathological process, EMT enables cancer cells to invade and migrate to distant secondary sites in the body, and it is well characterised in *in vitro* models. However, the activation of EMT in human solid tumours remains elusive and is rarely seen by pathologists. mRNA translation is a highly regulated process which has been linked to EMT in numerous cell-based studies. This study aims to quantify markers of EMT in a common and highly lethal malignancy, and to answer several key questions:

- i) to what extent does EMT occur in lung adenocarcinoma?

- ii) what phenotypic features is EMT related to?

- iii) does EMT contribute to the virulence of this disease, and if so, how?

- iv) do I see evidence that EMT is accompanied by alterations in the regulation of mRNA translation?

1.2   Cancer

Cancer is multi-step disease characterised by natural selection for genomic and epigenetic events that favour tumour cell proliferation and survival within the host organism.

Hanahan and Weinberg originally identified six biological behaviour enabling tumorigenesis, namely self-sufficiency in growth signals, insensitivity to growth inhibitory signals, evasion of programmed cell death (i.e. apoptosis), limitless replicative potential, sustained angiogenesis and tissue invasion and metastasis (Hanahan and Weinberg, 2000). More understanding of the molecular mechanism leading to tumorigenesis has enabled Hanahan and Weinberg to describe further traits of tumour cells: the capacity of tumour cells to evade detection by immune system and the ability to promote pro-tumorigenic inflammation (Figure 1.1) (Hanahan and Weinberg, 2011).
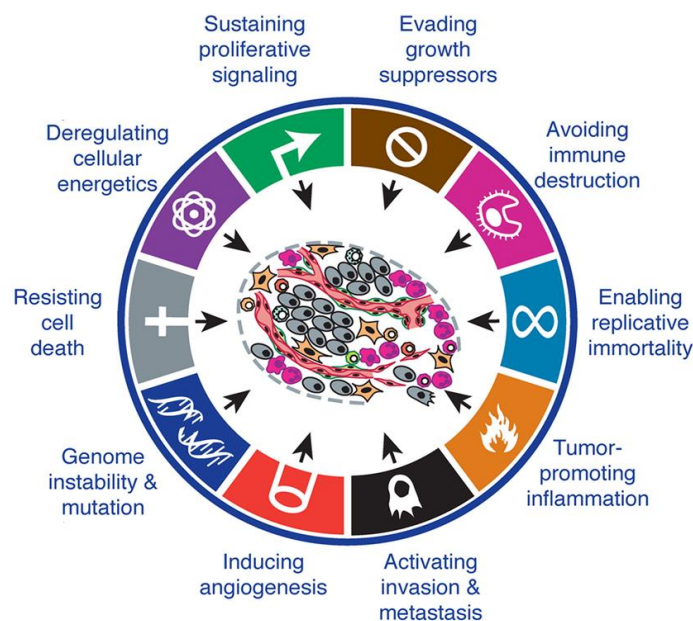
Figure 1.1. Hallmarks of cancer.
Adapted from (Hanahan and Weinberg, 2011).

Carcinomas (i.e. epithelial malignant tumours) acquire the ability to invade, crossing the basement membrane on which they normally lie, and spreading to other parts of the body resulting in distant metastatic tumour formation.

1.3   Cancer metastasis

Cancer metastasis is the immediate cause of cancer mortality in most cancer patients, accounting for 90% of cancer deaths (Seyfried and Huysentruyt, 2013). Cancer metastasis involves the dissemination of tumour cells from the primary tumour to distant sites forming secondary tumour via three main metastatic routes: the primary tumour can directly invade into adjacent tissues; it can spread via lymphatic vessels leading to metastasis in lymph nodes, or via the blood circulation resulting in metastases in distant organs. The metastatic sequence consists of several steps starting with invasion through the basement membrane, followed by the detachment of tumour cells from the primary tumour, intravasation into the lymphatic or blood vessels, extravasation at distant site while evading the immune system, and invasion and proliferation in the second organ forming the secondary tumour.

Most patients die of metastatic malignancy (Asamura *et al.*, 2015; Eberhardt *et al.*, 2015) (Figure 1.2). Lymph node metastasis in non-small cell lung cancer patients is associated with a progressive decrease in overall patient survival outcome with increasing involvement of the lymph nodes. When the tumour has not metastasized in the lymph nodes (pN0), the overall 5- year survival is about 75% but when it spreads to regional lymph nodes increasingly distant from the primary site (pN1-pN3), the 5-year survival rate reduces drastically to 49% for pN1, 36% for pN2 and 20% for pN3 (Figure 1.2A) (Asamura *et al.*, 2015). The effect of distant non-lymphatic metastasis is depicted in Figure 1.2B: in this study, non-small cell lung cancer patients with no metastasis show a 5-year survival rate of 50%. However, 85% of patients at M1a stage, which reflects patients with intra-thoracic metastases, dies after five years post-surgery. No patients at stage M1b or M1c where the tumour has metastasized in one or several extra-thoracic distant organs respectively survived longer than five years after surgery. Therefore, understanding molecular mechanism leading to cancer progression and developing therapeutic strategies targeting this event is essential.

A: Pathological nodal stage

	Events / N	MST	60 Month
N0	5694 / 22938	NR	75%
N1	1842 / 3811	57.5	49%
N2	2674 / 4522	35.0	36%
N3	113 / 155	19.1	20%

B: Pathological metastatic stage
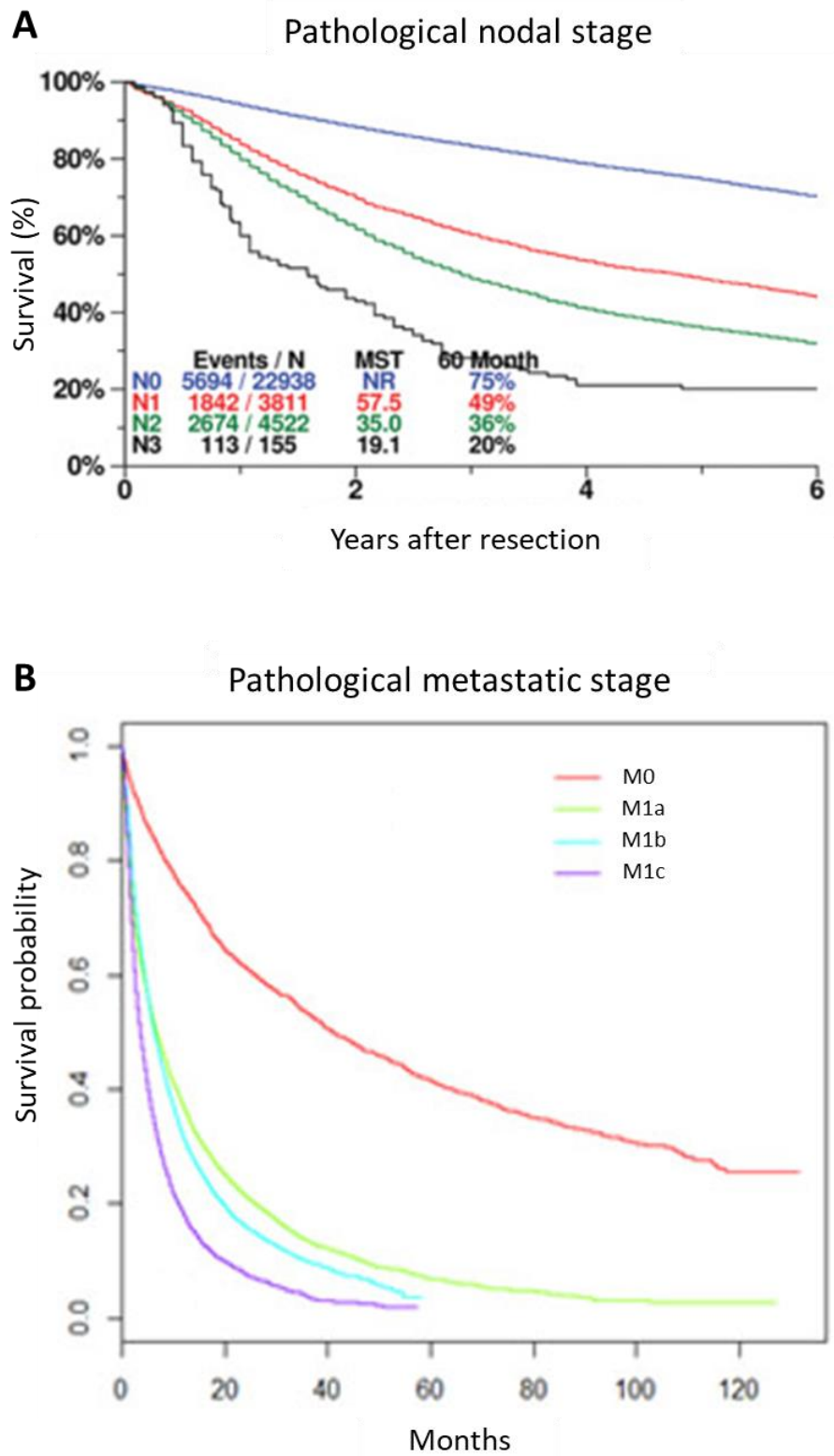
Figure 1.2. Overall survival of pathological nodal stage and metastatic stage in non-small cell lung cancer patients.
A: Kaplan-Meier of overall survival of pathological nodal stage in NSCLC patients. B: Kaplan-Meier of overall survival of pathological nodal stage NSCLC patients.
P <0.001 (Adapted from (Asamura *et al.*, 2015; Yang *et al.*, 2017)).

Several factors have been hypothesised as contributing to tumour metastasis such as genetic alterations, epithelial mesenchymal transition (EMT), cancer stem cells (Popper, 2016; Seyfried and Huysentruyt, 2013) and circulating tumour cells. *In vitro* studies have demonstrated that TWIST upregulation and E-cadherin loss (EMT markers) could promote metastasis in breast cancer and squamous cell carcinoma (Onder *et al.*, 2008; Tsai *et al.*, 2012; Yang *et al.*, 2004). The role of cancer stem cells as metastatic neoplasm development was proposed in the late 1970s, when mice were intravenously injected with B16 melanoma cells showing that only 0.1% of injected tumour cells were able to generate metastases due to their unique properties of survival and ability to reconstitute tumours (Fidler and Hart, 1982).

Although there have been many significant breakthroughs in the understanding of invasion and metastasis over the last few decades, these have generally used cell culture and animal models, and a detailed mechanistic understanding of how these events occur in primary human tumours remains unclear. In this study, I aim to better understand how tumour metastasis occurs in human primary lung tumour tissue.

1.4 Lung cancer

1.4.1 Introduction

Lung cancer remains the leading cause of cancer related death among both sexes worldwide according to the World Health Organisation (WHO) in 2015 (Travis *et al.*, 2015). 1 in 15 men and 1 in 17 women will develop the disease in their lifetime according to the American Cancer Society (ACS, 2020). Lung cancer claims 1.3 million lives every year (Imielinski *et al.*, 2012). The worldwide average 5-year survival rate remains no more than 15%, largely due to the late diagnosis of the disease when the tumour has spread to nearby or distant organs (Popper, 2016; Wood *et al.*, 2015). Lung cancer can metastasize directly by invading through the pleural surface of the lung into adjacent organs, by spreading to regional lymph nodes via lymphatic circulation, or by metastasizing to distant organs such as the brain, adrenal glands or liver through blood vessels (Filosso *et al.*, 2016; Popper, 2016; Travis, 2014). Potentially curative treatments for advanced stage patients are usually not available due to metastatic spread. When the tumour is known to have metastasized beyond N1, the risks of surgery outweighs the benefits. Furthermore, thoracic surgery is hazardous by nature compared to many other sites (e.g. the abdomen for colon cancer, or the breast in the case of breast cancer). This also helps to explain the low survival rate in lung cancer.

The main cause of lung cancer is tobacco consumption (Alberg *et al.*, 2007). In the UK, Doll and Hill demonstrated the strong and significant association between smoking and lung carcinoma in 1950, comparing lung cancer patients and control patients aged from 25 to 74 years (Doll and Hill, 1950). Although tobacco smoking has been recognized as a cause of lung tumorigenesis by health authorities, people still consume all kind of tobacco products.

Chemical carcinogens detectable in tobacco smoke such as polycyclic aromatic hydrocarbons or nitrosamines can cause DNA adducts which subsequently generates point mutations in genes involved in lung cancer formation. In the polycyclic aromatic hydrocarbons family, benzo[a]pyrene diol-epoxide-guanine adducts formation were found in smokers which led to G to T transversion, accounting for, for example, the

commonest driving mutation found in lung adenocarcinoma (*KRAS* G12C) as well as many oncogenic changes in the *TP53* gene (Vineis and Caporaso, 1995).

Other contributors to lung cancer can be the atmosphere that lung cancer patients are exposed to such as outdoor pollution, second-hand smoke or asbestos, radon, nickel as well as inherited genetic alterations (e.g. inherited EGFR T790M missense mutation), and chronic lung diseases such as pulmonary fibrosis, chronic obstructive pulmonary disease or tuberculosis (Travis, 2014).

For many decades, pathologists used to merely distinguish small cell lung cancer (SCLC) from non-small cell lung cancer (NSCLC) in small biopsies and cytology of lung cancer due to the difference in tumour behaviour (Travis *et al.*, 2010). There was not any differentiation within NSCLC until the WHO classification of lung tumours formally established the histological types of these lung tumours in 1967.

1.4.2 Types of lung cancer

Lungs are divided into lobes (left and right lobes having two and three lobes respectively). The trachea branches out into left and right bronchi, which bring air to the lungs. These bronchi further branch out into bronchioles and end with tiny air sacs called alveoli (Chaudhry R, 2021).

In the lung, bronchi are covered on the surface by ciliated pseudostratified columnar epithelium. Although the epithelium is characterised as pseudostratified in microscopic sections (i.e. multi layers of epithelial cells) layer, it is in reality a single layer of epithelium. The pseudostratification name is due to the presence of various cell types and variable nucleus height (Camelo *et al.*, 2014; Tam *et al.*, 2011). Bronchioles are lined by ciliated single cuboidal epithelium (Empey, 1978) (Figure 1.3 A, B and C).

Lung alveolar epithelium is composed of a monolayer of non-ciliated squamous epithelial cells, named type I alveolar epithelial cells (AT1) or type I pneumocyte, with intermittent single cuboidal non-ciliated epithelial cell, named type II alveolar epithelial cells (AT2) or type II pneumocytes (Figure 1.3 A and D). Type I pneumocytes represent 95% of the total alveolar surface and enable oxygen and carbon dioxide gas exchange. Type II cells are secretory cells that synthesise and secrete surfactant responsible for

the control of alveolus surface tension. Their main role is to prevent alveolar space from collapsing during exhalation (Knudsen and Ochs, 2018). They also contribute to the alveolar epithelium homeostasis by renewing the alveolar epithelium or repairing damaged type I alveolar cells and therefore playing a role as epithelial stem cells (Nabhan *et al.*, 2018).
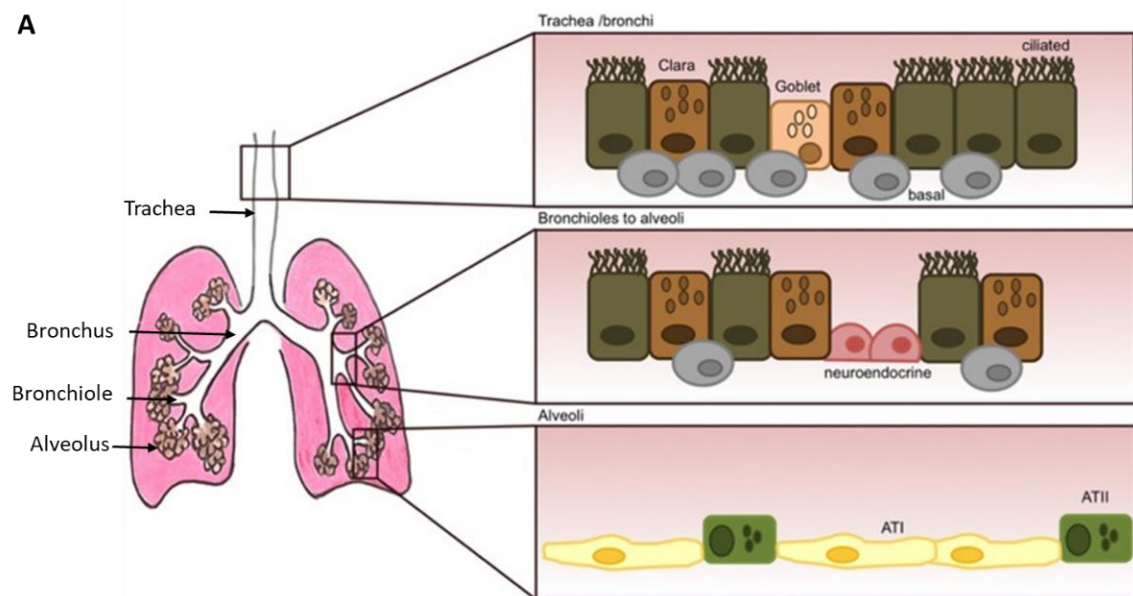
Figure 1.3. Schematic representation of the human normal lung.
Schematic representation of the lung with the different type of cells constituting the lung (Camelo *et al*., 2014).

Lung cancer can be divided into two categories: SCLC and NSCLC. SCLC accounts for approximately 15% of all lung cancer worldwide and NSCLC accounts for 85% of all lung cancer (Zappa and Mousa, 2016). NSCLC can be subcategorised into three main classes: squamous cell carcinoma, large cell carcinoma and adenocarcinoma.
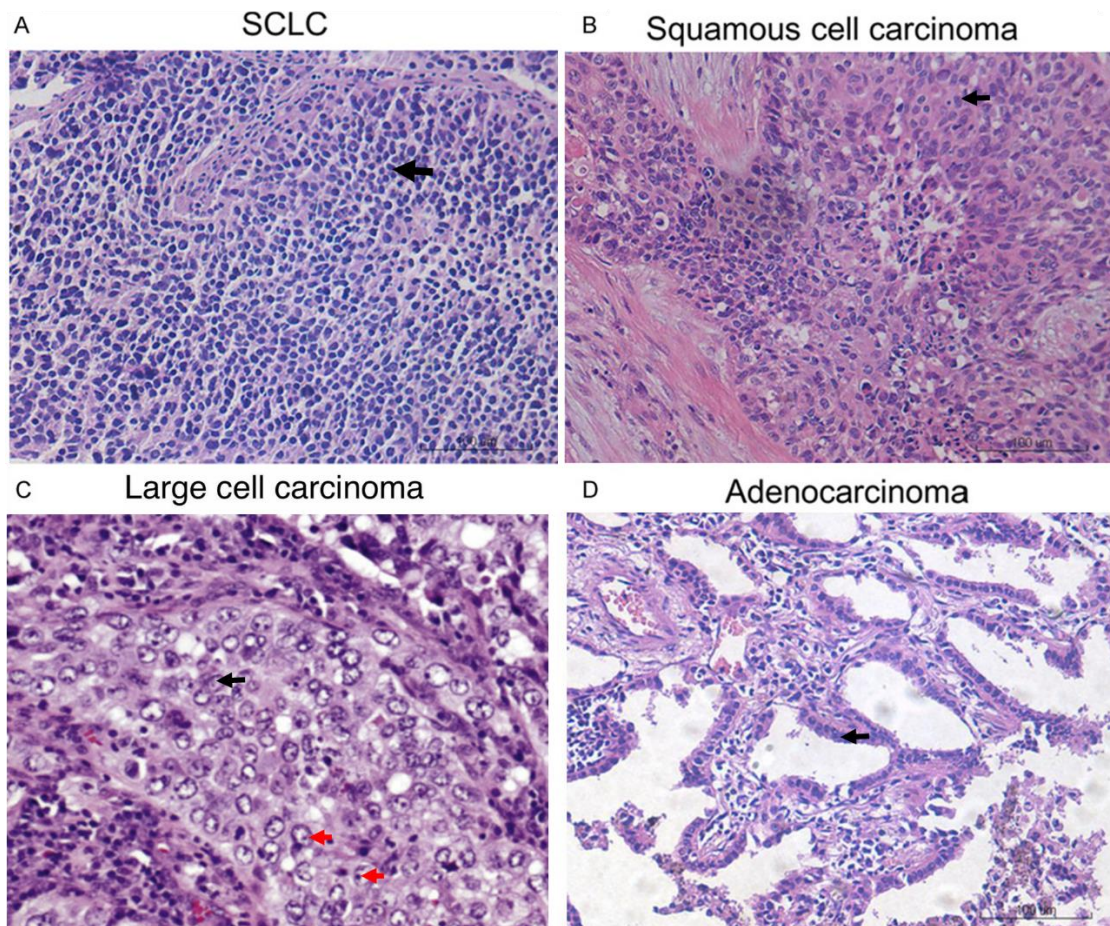
Figure 1.4. Subtypes of lung cancer.

H&E images of subtypes of lung cancer. A: H&E image of small cell lung carcinoma (SCLC). Cytoplasm is stained in pink and nuclei are stained in purple. Dense sheets of small crowded malignant tumour cells. B: H&E image of squamous cell carcinoma. Cytoplasm is stained in pink and nuclei are stained in purple C: H&E image of large cell carcinoma. Sheet of large polygonal cells with moderate cytoplasm and vesicular nuclei that show prominent nucleoli (red arrow). Adapted from (Wistuba *et al.*, 2018)). D: Adenocarcinoma. Carcinoma showing glandular differentiation. Cytoplasm is stained in light purple and nuclei are stained in dark purple. Black arrows in A, B, C and D show tumour cells. Red arrows in C point out nucleoli. A, B and D are adapted from (Ren *et al.*, 2015).

Lung cancer subtypes		Percentage of lung cancer	Morphology	Anatomical location in the lung	Immunohistochemical marker	Prominent genetic alterations	Cell of origin	5-year overall survival
SCLC		15%	Small malignant cell; Scant cytoplasm; Granular nuclear chromatin; Undetectable nucleoli	Central or peripheral	AE1/3; TTF1; NCAM; CD56; Synaptophysin	Mutations: RB, TP53; Amplification: MYC	Unknown	7%
NSCLC	Adenocarcinoma	40%	Glandular differentiation of tumour with potential production of mucin.	Peripheral and central	TTF1; Napsin A	Mutation EGFR, KRAS, TP53	Broncho-alveolar stem cell	25%
	Squamous cell carcinoma	25-30%	Keratinisation; Pearl formation; Intercellular bridges — Non keratinisation	Central	P40; P63; Cytokeratin 5; Cytokeratin 5/6	Mutation: TP53; Amplification: FGFR	Tracheal basal progenitor cells	
	Large cell undifferentiated carcinoma	5-10%	Sheet of large polygonal cell; Vesicular nucleus; Prominent nucleolus; Moderate cytoplasm	Peripheral	Negative for markers above	Mutation: EGFR, KRAS, TP53; Deletion: CDKN2A; Amplification: MYC, CCNE1	Unknown	

Table 1.1. Differences between SCLC and NSCLC.
(ACS, 2021; Borthwick *et al.*, 2001; Hanna and Onaitis, 2013; Kim *et al.*, 2005; Travis, 2014; Zappa and Mousa, 2016).

Table 1.1 summarises the main different features of SCLC and NSCLC.

1.4.2.1 Small cell lung cancer

SCLC has a poor prognosis compared to NSCLC with 6% 5-year survival rate while NSCLC 5-year survival is 24% according to the American Cancer Society (ACS, 2021). SCLC is the most aggressive common form of lung cancer. 60-70% of patients with SCLC are diagnosed when the tumour has already metastasised (Carter *et al.*, 2014).

SCLC accounts for approximately 15% of all lung cancer worldwide (Zappa and Mousa, 2016). It is a poorly differentiated high-grade disease characterised by dense sheets of malignant tumour cells with neuroendocrine features (Travis, 2014). The tumour consists of small crowded cells (Figure 1.4A) containing scant cytoplasm, with a nucleus formed of finely granular and uniform chromatin hardly detectable nucleoli.

SCLC develops mainly centrally in the major airways but can also occur in the periphery of the lung in 5% of cases.

The cells are immunohistochemically positive for neuroendocrine markers such as NCAM/CD56. They also often expresses TTF-1, a lineage marker for alveolar epithelial cells.

Predominant genetic alterations observed in SCLC are inactivation of *TP53* and *RB* genes leading to loss of functional p53 and RB expression and amplification of *MYC* gene conducting to overexpression of myc transcription factors. Several other genetic and epigenetic aberrations changes occur in SCLC such as mutations in *CREBBP* or *PTEN* genes as well as amplification of *FGFR1* and *CCN2* genes, and hypermethylation of the *RASSF1* gene's promoter leading to inactivation.

The cell of origin of SCLC is unclear. It is believed that this disease can origin from neuroendocrine progenitor, or may sometimes share the same origin as adenocarcinoma since combined tumours are not uncommon (Hanna and Onaitis, 2013).

1.4.2.2 Squamous cell carcinoma

25-30% of all lung cancer are squamous cell carcinoma. This is an aggressive carcinoma that can demonstrate keratinisation as well as keratin whorl or pearl formation with visible intracellular bridges (i.e. tight junctions) between cells. It arises from the bronchial epithelium. It can also appear as morphologically undifferentiated NSCLC that are only identifiable by expression of protein markers specific to squamous cell differentiation (Figure 1.4B). Squamous cell carcinoma was the predominant type of lung cancer until the 1950s but a switch has been observed since 1960 where squamous cell carcinoma frequency regressed and adenocarcinoma frequency increased in smokers (Travis, 2014). It mainly occurs in tobacco consumers and tumours are commonly developed in the central lung, in the main or lobar bronchus.

Squamous cell carcinomas are typically immunohistochemically positive for p40, p63, cytokeratin 5 or cytokeratin 5/6 and negative for TTF-1. Genomic alterations characteristics of squamous cell carcinoma are mainly gain/amplification of *SOX2*, *TP63*, *EGFR* and *FGFR1* genes leading to overexpression of these genes, mutation of *TP53* gene, and deletion of *CDKN2A*, leading to loss of p14ARF and p16 expression from this locus.

It is believed that squamous cell carcinoma are arisen from tracheal basal progenitor cells.

1.4.2.3 Large cell carcinoma

Large cell carcinoma represents 5-10% of all lung cancer and is defined as an undifferentiated malignant epithelial tumour that does not exhibit any cytological, architectural or immunohistochemical features of small cell carcinoma, adenocarcinoma or squamous cell carcinoma. It is characterised by sheet of large polygonal cells with moderate cytoplasm and vesicular nuclei that show prominent nucleoli, where the ribosome biogenesis occurs (Figure 1.4C). The main cause of large cell carcinoma is smoking and this type of NSCLC is negative for all the markers specific to squamous cell carcinoma and adenocarcinoma.

Large cell carcinoma usually arises as a mass at the periphery of the lung.

Genetic aberrations commonly occurring in large cell carcinoma are *EGFR*, *KRAS*, *TP53* mutations, *CDKN2A* deletions and *MYC* and *CCNE1* gene amplification. Epithelial-mesenchymal transition related genes were also frequently expressed in this tumour growth.

The cell type(s) origin of undifferentiated large cell carcinoma is as yet unknown.

1.4.2.4 Adenocarcinoma

Adenocarcinoma is by definition a carcinoma showing glandular differentiation or mucin production with diverse subtype of histological growth (Figure 1.3D). Adenocarcinoma can exhibit histopathologically diverse growth patterns (explained in detail in the section 1.4.5). This subcategory of NSCLC has become the most common type of all lung cancer since 1960 accounting for 40% of lung cancer nowadays. One of the reason for the progressive increase of developing adenocarcinoma tumours is believed to be the changes in cigarettes composition. In particular, the introduction of filters vent in cigarettes in the 1950s enabled smokers to perform deeper smoke inhalation allowing the carcinogens constituting the tobacco to reach the periphery of lung (Furrukh, 2013).

Adenocarcinoma mostly arises at the periphery of the lung which may explain the late diagnosis of the disease as some patients can present no symptoms or symptom that might appear quite negligible, such as cough. As the tumour grows, the symptoms accentuate and by the time of diagnosis, the tumour may have already done severe damage. Some adenocarcinomas can grow centrally.

Adenocarcinoma mainly express TTF-1 and napsin A which are useful as diagnostic immunohistochemical markers.

Prominent genetic aberrations implicated in the development of the disease in the Caucasian population are *EGFR* (5-15%) and *KRAS* (20-30%) followed by ALK gene fusion (3-6%) and others including *TP53* genes mutation (Figure 1.5).

It is thought that adenocarcinoma arise from broncho-alveolar stem cells.
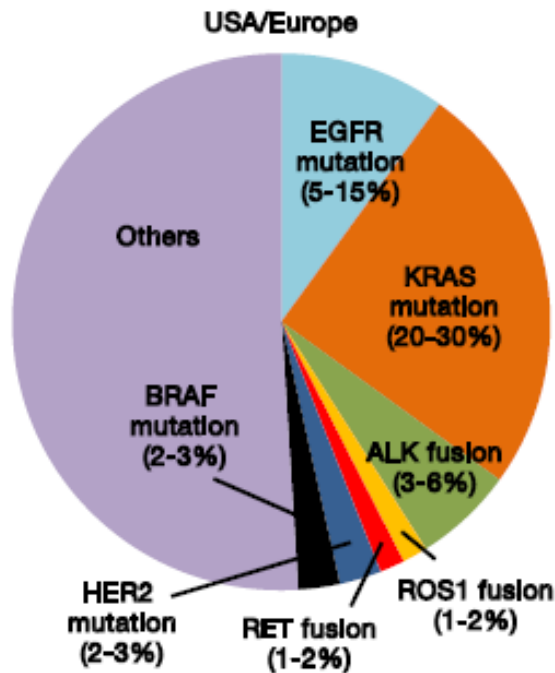
Figure 1.5. Pie chart of the frequencies of common genomic changes in lung adenocarcinoma in USA/Europe.
Adapted from (Kohno *et al.*, 2015).

1.4.3   TNM staging of lung cancer

TNM classification of malignant tumour stage is defined by the analysis of the anatomical extent of three factors i.e. the primary tumour, the lymph nodes and metastases (CRUK, 2020). One of the many purposes of developing an international tumour staging classification is to establish an accurate diagnosis of the disease in order to provide efficient treatment for the patient. NSCLC and SCLC both follow the TNM classification of lung cancer.

The current TNM stage is the eighth edition of TNM but the lung adenocarcinoma patients were classified according to the seventh TNM edition (AJCC, 2010) in our study, which is presented below.

Primary tumour stage	Characteristics		
TX	Impossibility of primary tumour assessement		
T0	No evidence of primary tumour		
Tis	Carcinoma *in situ*		
T1	Tumour ≤ 3 cm in dimension surrounded by lung or visceral pleural without evidence of invasion in the main bronchus		
	T1a	Tumour ≤ 2 cm in size	
	T1b	Tumour > 2 cm but ≤ 3 cm in size	
T2	Tumour > 3 but ≤ 7 cm in dimension or		
	Tumour involves main bronchus 2 cm or more distal to the carina, or tumour invades visceral pleural or tumour is associated with atelectasis or obstructive pneumonitis that extends to the hilar region, involving part or all the lung:		
	T2a	Tumour > 3 cm but ≤ 5 cm in size	
	T2b	Tumour >5 cm but ≤ 7 cm in size	
T3	Tumour > 7 cm in dimension or		
	Tumour directly invades parietal pleural, chest wall, phrenic nerve, parietal pericardium, mediastinal pleura, or invade main bronchus , but without involvement of the carina		
	Tumour is associated with atelectasis or obstructive pneumonitis that extends to the hilar region, involving part or all the lung		
	Presence of separate tumour nodule(s) in the same lobe as the primary tumour		
T4	Any size tumour that has invaded the mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, oesophagus,vertebral body or carina		
	Presence of separate tumour nodule(s) in a different ipsilateral lobe to that of the primary tumour		

Table 1.2. Primary tumour staging criteria of the lung according to AJCC.

Tumour stage assesses the size and extent of the primary tumour site and is divided into four categories. TX means that the primary tumour cannot be evaluated. T0 defines the histological absence of primary tumour. Tis signifies carcinoma *in situ*. T1-4 refers to the increase in size of the primary tumour (Table 1.2).

Nodal stage	Characteristics
NX	Impossibility of evaluating the metastasis of lymph nodes
N0	No evidence of lymph nodes metastasis
N1	Presence of metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension
N2	Presence of metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Presence of metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)

Table 1.3. Nodal staging criteria of the lung according to AJCC.

Nodal stage evaluates the spread of tumour to lymph nodes and is classified into three distinct groups. NX designates tumour where the regional lymph nodes cannot be assessed. N0 refers to the absence of tumour in the regional lymph nodes. N1 to N3 defines the increasing involvement of regional lymph nodes (Table 1.3).

Metastatic stage	Characteristics	
M0	No evidence of distant metastasis	
M1	Presence of distant metastasis :	
	M1a	Presence of separate tumour nodule(s) in a contralateral lobe. Tumour with pleural nodules or malignant pleural or pericardial effusion
	M1b	Presence of metastasis beyond the thoracic cage

Table 1.4. Tumour metastasis staging criteria of the lung according to AJCC.

Metastases assess the presence or the absence of metastasis and is categorised into two groups. M0 refers to the absence of distant metastasis. M1 designates the detection of distant metastasis (Table 1.4).

Overall stage	Tumour stage	Nodal stage	Metastatic stage
Stage 0	Tis	N0	M0
Stage IA	T1a	N0	M0
	T1b	N0	M0
Stage IB	T2a	N0	M0
Stage IIA	T2b	N0	M0
	T1a,b ; T2a	N1	M0
Stage IIB	T2b	N1	M0
	T3	N0	M0
Stage IIIA	T1a,b ; T2a,b	N2	M0
	T3	N1, N2	M0
	T4	N0, N1	M0
Stage IIIB	T1a,b ; T2a,b	N3	M0
	T3, T4	N2	M0
	T4	N3	M0
Stage IV	Any T stage	Any N stage	M1a
	Any T stage	Any N stage	M1b

Table 1.5. Overall staging of lung cancer according to AJCC.

Overall stage refers to the TNM classification of primary tumour staging combining all these three parameters. Stage I is characterised as the size of the primary tumour not exceeding four cm without any detection of intra or extra-thoracic metastasis. Stage II refers to tumour found nearby lymph nodes. Stage III of NSCLC is diagnosed in patients when the primary tumour has metastasised to the lymph nodes in the middle of the chest. Stage IV is defined by the detection of extra-thoracic metastasis (Table 1.5).

1.4.4 Current treatment for NSCLC

As stated previously, lung cancer is the commonest cause of lung cancer death worldwide. Despite improvements in lung cancer patient care, the survival rate still remains low compared to other developed countries (Walters *et al.*, 2015). One of the main reasons for the poor outcome in lung cancer is the late diagnosis of the disease, when the tumour has already spread to the regional lymph nodes and beyond. Approximately 62% of lung cancer patients are diagnosed at advanced stage of the disease in the UK (Jones and Baldwin, 2018). Treatment for NSCLC is mainly based on the staging of the tumour and the performance status of the patient. There are three main branches for NSCLC treatment with curative intent that consists of surgery, chemotherapy and radiotherapy.

Thoracic surgery with curative intent is provided to patients with early stage disease that are fit enough (NICE, 2020). Thoracic surgery is mostly performed on fit patients that are at stage I, II and IIIA of NSCLC (Zappa and Mousa, 2016). Surgery is usually anatomically complete lobectomy with regional lymph node sampling but can also be less radical (wedge resections or segmentectomy).

Radiotherapy treatment aims to eliminate cancer cells by radiation. Fit patients that are at stage I, II and III of NSCLC who are unwell or unwilling to undergo surgery are offered a radiotherapy treatment with curative intent (NICE, 2020). It can also be used as an adjuvant treatment that reduces the risk of tumour recurrence after surgery (Zappa and Mousa, 2016).

Chemotherapy is not curative by itself but its aim is to extend lifespan (Zappa and Mousa, 2016). It can be used in advanced disease or as an adjuvant treatment aiming to prevent recurrence after surgery.

Tyrosine kinase inhibitors are therapies targeted to specific mutations, which inhibit activated proto-oncogenes, for instance activated mutated EGFR or ALK proteins. This treatment is currently reserved for advanced stage patients or those with recurrent disease. Although these treatments impact tumour growth and can improve lung cancer

symptoms, most clinical trials report little improvement in overall survival (Jones and Baldwin, 2018).

Immunotherapy is the newest treatment modality and has the potential for induction of long-term remission in some NSCLC patients. It depends upon the body's own natural immune system to target cancer cells. One of the first immunotherapy drug used for advanced stage cancer patients is pembrolizumab, which is given to patients within whom tumour cells show high levels of PD-L1 protein expression by immunohistochemistry. PD-L1 is a protein expressed on the surface of tumour cell, which allows the latter to evade the immune system detection by binding to the PD-1 receptor of the cytotoxic T-cell and thereby deactivating its cytotoxic effect. Pembrolizumab prevents PD-L1 binding to its PD-1 receptor on the T-cell (Jones and Baldwin, 2018).

Keynote 024 was the first randomised phase III trial proving the benefits of pembrolizumab on NSCLC patient with PD-L1 expression of at least 50%. In this trial, 305 untreated advanced staged (stage IV) NSCLC patients expressing at least 50% of PD-L1 were randomly divided into two groups: the first group constituted of 154 patients received the pembrolizumab while the second group (151 participants) received a platinum-based chemotherapy.
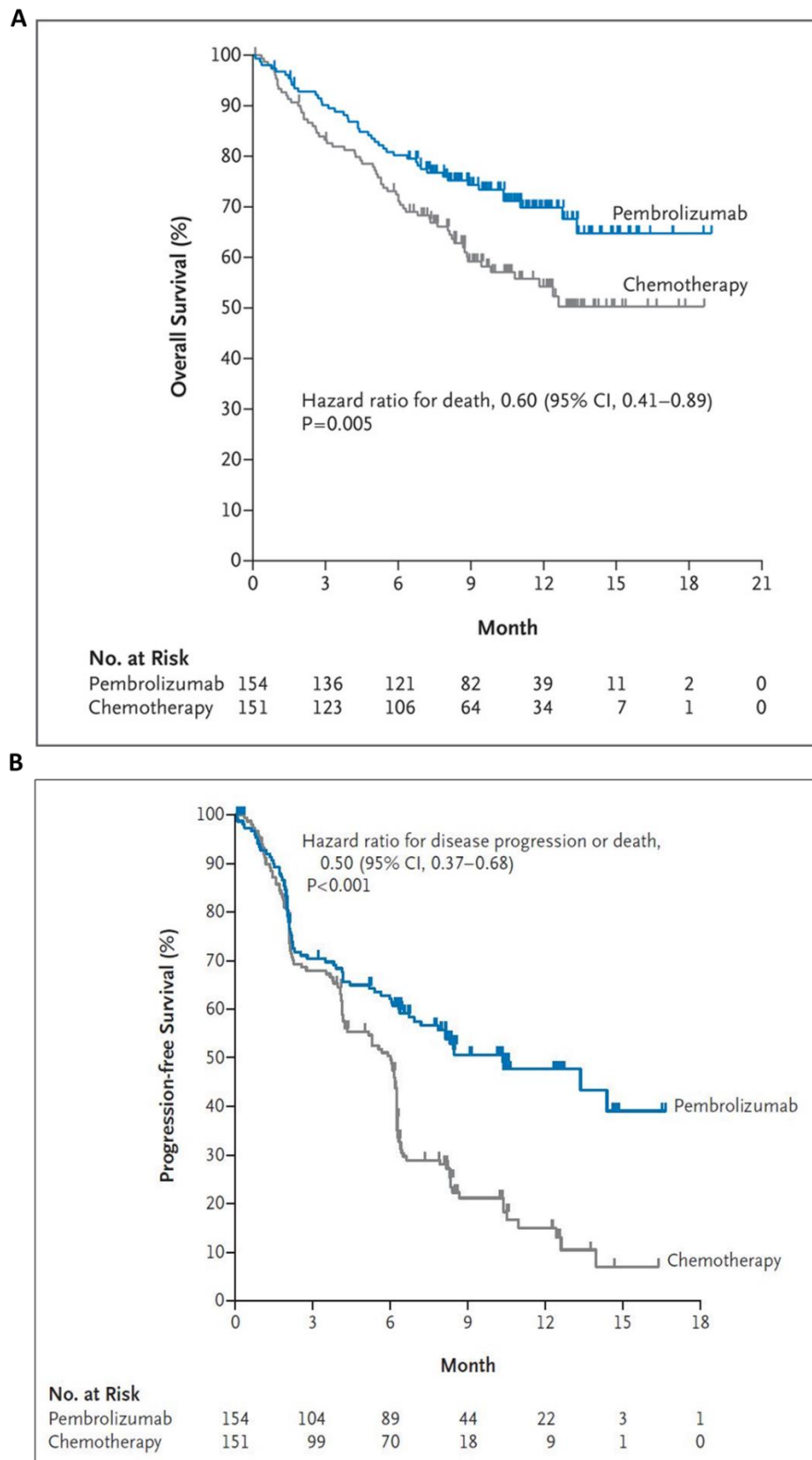
Figure 1.6. Effect of pembrolizumab on overall and recurrence-free survivals of stage IV NSCLC patients.
A: Kaplan-Meier of overall survival of pembrolizumab and chemotherapy in metastatic NSCLC patients.
B: Kaplan-Meier of progression-free survival of pembrolizumab and chemotherapy in metastatic NSCLC patients.

The results show that pembrolizumab significantly gives better overall survival compared to chemotherapy treated group. There is a significant decrease of 40% (hazard ratio = 0.60, p = 0.005) in risk of death with the pembrolizumab treatment (Figure 1.6A).

The median progression-free survival rate significantly differ between the two groups (10 months for pembrolizumab group versus about 6 months for chemotherapy group, p < 0.001). There is a significant decrease of 50% in risk of relapse of the disease in the metastatic NSCLC patients treated with pembrolizumab compared to chemotherapy (Figure 1.6B). Taken together, this data demonstrates the efficiency of pembrolizumab drug treatment versus the chemotherapy treatment. The drug can extend the overall and disease free survival in stage IV NSCLC patients (Reck *et al.*, 2016). Nowadays, this is recommended for metastatic lung cancer patients (NICE, 2021) and PD-L1 immunohistochemistry is considered as predictive marker.

Other immunotherapies have been developed such as immunotherapy through tumour cell vaccine using genetically modified tumour cells and antigen-specific vaccines (Zappa and Mousa, 2016).

1.4.5   Lung Adenocarcinoma

1.4.5.1   Characteristics of lung adenocarcinoma tumour growth patterns

Adenocarcinoma is the most common type in NSCLC and accounts for 40% of all lung cancers (Zappa and Mousa, 2016). With progress in cancer research and therapy, its classification has been refined and revisited over the years. Adenocarcinomas display several distinct growth patterns, which can be mixed together in a single lesion. Adenocarcinoma of the lung can exhibit six different types of tumour growth pattern: lepidic, acinar, cribriform (a variant of acinar pattern which does not define a tumour type in the current WHO classification) papillary, micropapillary and solid.

Lepidic tumour pattern consists of proliferation of type II pneumocyte or Clara cells along the surface of alveolar structures. This is a non-invasive, i.e. *in situ* tumour growth, and the normal lung architecture is preserved. Its presence is a defining feature of several WHO tumour subtypes: adenocarcinoma *in situ*, minimally invasive adenocarcinoma and lepidic predominant adenocarcinoma (Figure 1.7).
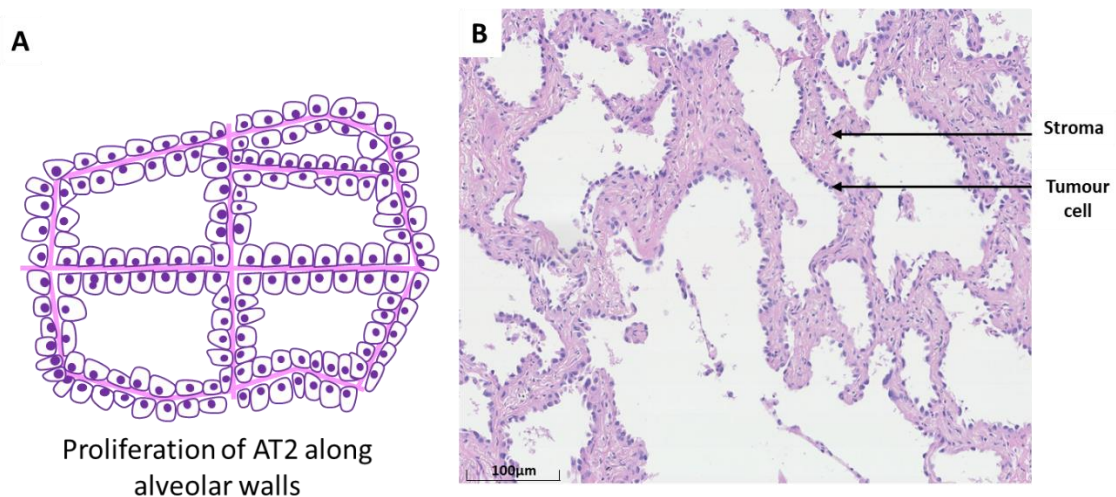
Figure 1.7. Lepidic tumour growth pattern.
A: Schematic representation of lepidic tumour growth pattern. B: H&E staining of lepidic tumour growth pattern. Nuclei are stained in purple (haematoxylin staining) and cytoplasm is stained in pink (eosin staining). Magnification: 20X. Scale bar: 100µm**.**
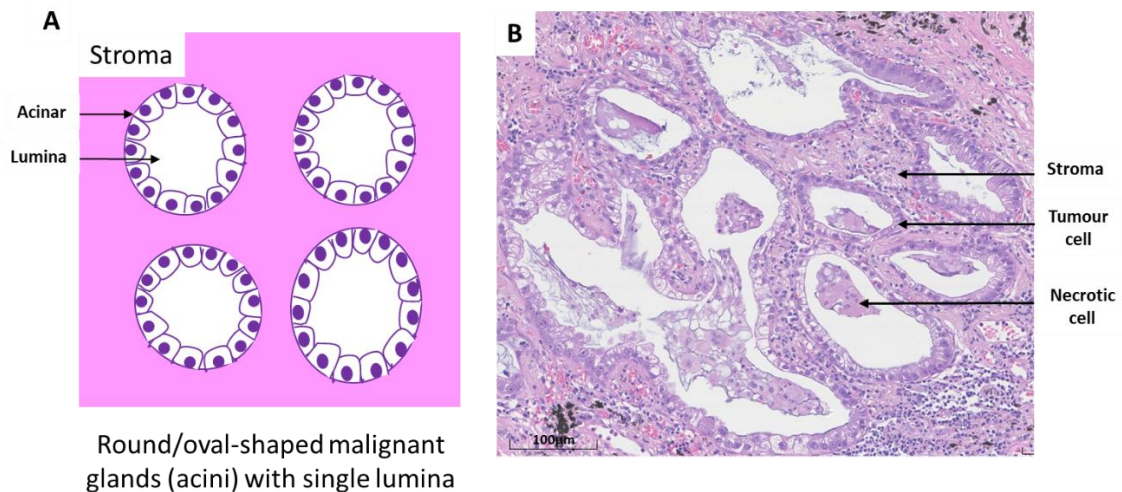
Figure 1.8. Acinar tumour growth pattern.
A: Schematic representation of acinar tumour growth pattern. B: H&E staining of acinar tumour growth pattern. Nuclei are stained in purple (haematoxylin staining) and cytoplasm is stained in pink (eosin staining). Magnification: 20X. Scale bar: 100µm**.**

Acinar pattern is composed of round or oval-shaped malignant glands or acini invading the stroma consisting of single luminal space surrounded by neoplastic cells, which may contain mucin. The alveolar architecture is lost (Figure 1.8).
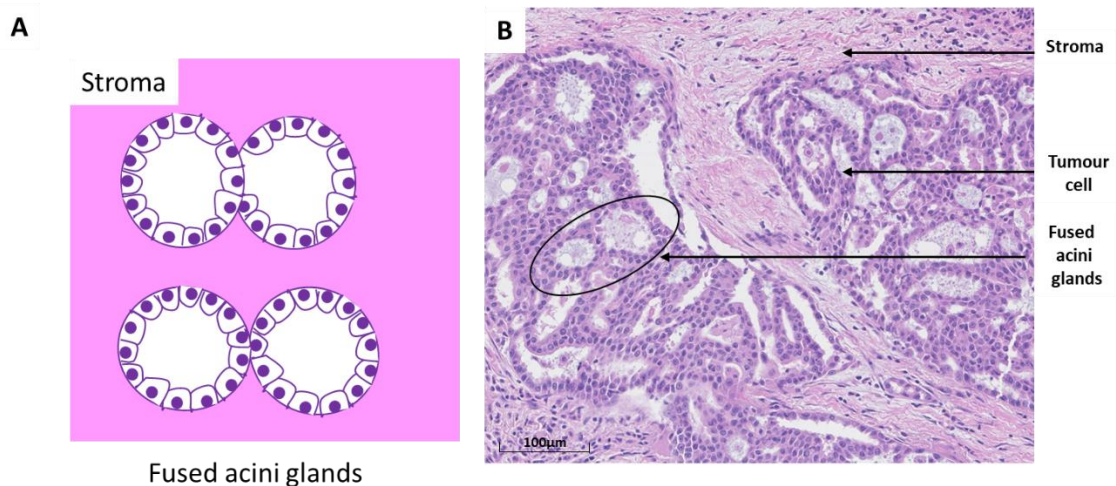
Figure 1.9. Cribriform tumour growth pattern.
A: Schematic representation of cribriform tumour growth pattern. B: H&E staining of cribriform tumour growth pattern. Nuclei are stained in purple (haematoxylin staining) and cytoplasm is stained in pink (eosin staining). Magnification: 20X. Scale bar: 100µm**.**

Cribriform pattern is defined as fused acinar glands with poorly formed glandular spaces and no stroma connection between these glands (Figure 1.9).
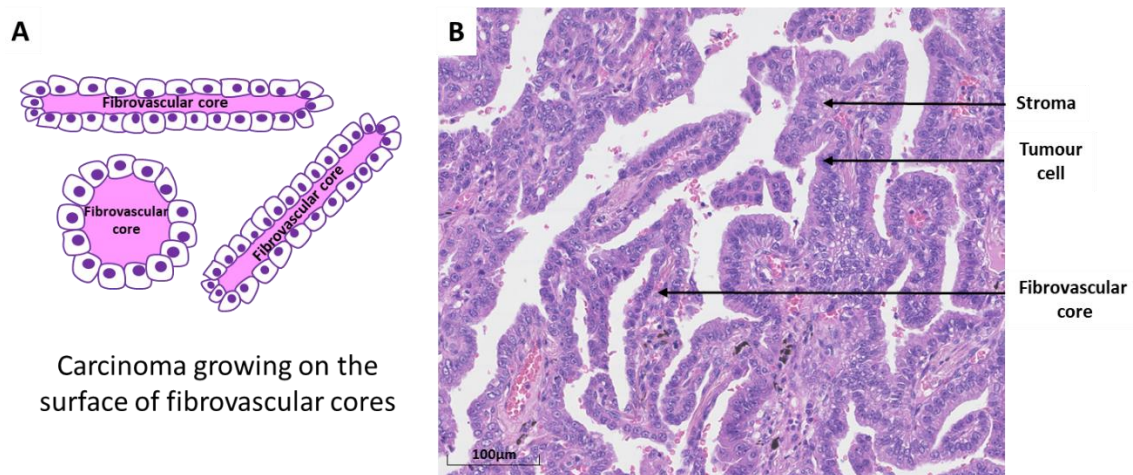
Figure 1.10. Papillary tumour growth pattern.
A: Schematic representation of papillary tumour growth pattern. B: H&E staining of papillary tumour growth pattern. Nuclei are stained in purple (haematoxylin staining) and cytoplasm is stained in pink (eosin staining). Magnification: 20X. Scale bar: 100μm.

Papillary reflects a tumour growth of columnar or cuboidal malignant epithelial tumour on the surface of fibrovascular cores constituted of stroma components (Figure 1.10).
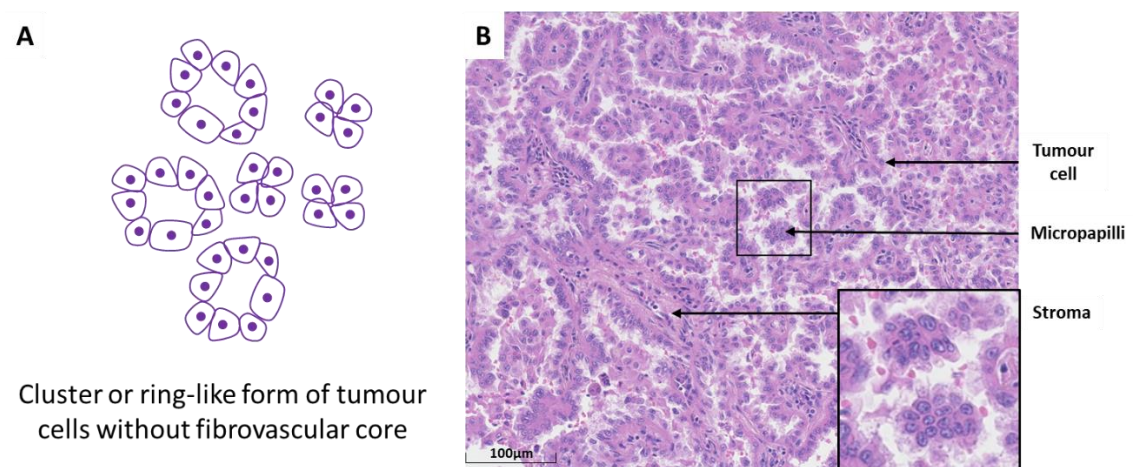
Figure 1.11. Micropapillary tumour growth pattern.
A: Schematic representation of micropapillary tumour growth pattern. B: H&E staining of micropapillary tumour growth pattern. Inset square represents a zoom in micropapillae structure. Nuclei are stained in purple (haematoxylin staining) and cytoplasm is stained in pink (eosin staining). Magnification: 20X. Scale bar: 100μm.

Micropapillary tumour pattern is characterised by clusters or ring-like forms of tumour cells (micropapillae) which lack the fibrovascular cores and occupy spaces within areas of usually papillary or acinar appearance (Figure 1.11).
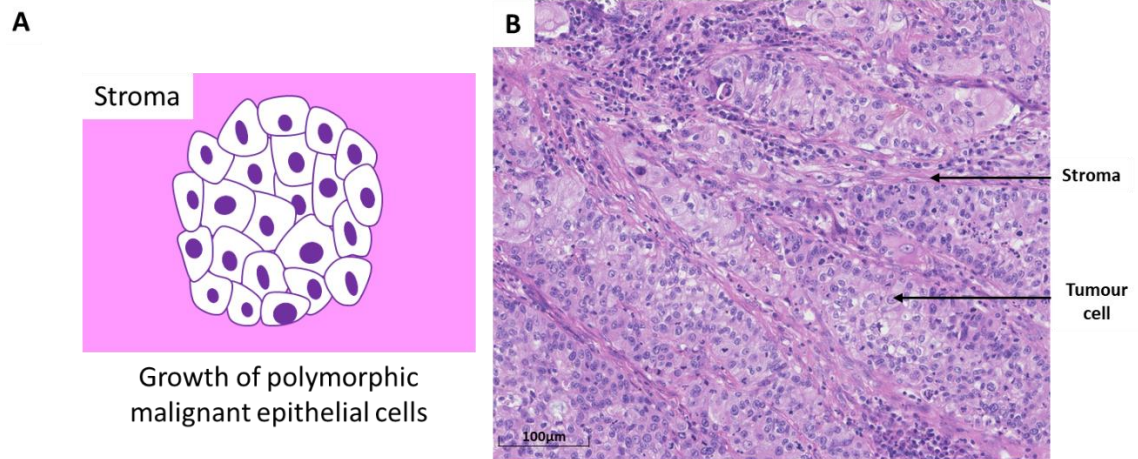
Figure 1.12. Solid tumour growth pattern.
A: Schematic representation of solid tumour growth pattern. B: H&E staining of solid tumour growth pattern. Nuclei are stained in purple (haematoxylin staining) and cytoplasm is stained in pink (eosin staining). Magnification: 20X. Scale bar: 100µm**.**

As depicted in Figure 1.12, solid tumour pattern is represented by sheets of polymorphic invasive epithelial tumour cells which lacks other recognizable tumour growth pattern of adenocarcinoma, namely as lepidic acinar, papillary and micropapillary growths.

Predominant adenocarcinoma sub-category	Tumour differentiation	Invasion
Adenocarcinoma *in situ*	Well	No
Minimally invasive adenocarcinoma	Well	No
Lepidic/*in situ*	Well	No
Acinar	Moderate	Yes
Cribriform	Moderate	Yes
Papillary	Moderate	Yes
Solid	Poor	Yes
Micropapillary	Poor	Yes

Table 1.6. Features of adenocarcinoma groups.
(Kadota *et al.*, 2014; Travis, 2014)

1.4.5.2   *WHO classification of lung adenocarcinoma*

The WHO, which publishes international guidelines for the classification of all malignancies, divides adenocarcinoma of the lung into the following types based upon their histopathological appearances: adenocarcinoma *in situ* (AIS), minimally invasive adenocarcinoma (MIA), lepidic predominant adenocarcinoma (LPA), acinar predominant adenocarcinoma (APA) (including cribriform growth pattern), papillary predominant adenocarcinoma (PPA), solid predominant adenocarcinoma (SPA), micropapillary predominant adenocarcinoma (MPPA).

AIS is well-differentiated, i.e. structurally similar to normal lung structure, small solitary adenocarcinoma, i.e. this ≤ 3 cm in diameter, with lepidic growth and no evidence of invasive component (Table 1.6).

Minimally invasive adenocarcinomas (MIA) are well-differentiated tumours ≤ 3 cm in diameter that predominantly display lepidic growth and a single invasive region ≤ 0.5 cm (Table 1.6).

LPA shows normal lung structure and therefore is defined as a partially well-differentiated lepidic growth with at least one invasive focus greater than 5 mm of diameter (Table 1.6).

APA is mainly composed of acinar pattern and is considered as moderately-differentiated invasive tumour growth (Table 1.6). It starts to lose the normal structure of the lung and become invasive.

Cribriform is considered as a subtype of acinar predominant adenocarcinoma. It is not regarded as a separate growth pattern in formal WHO criteria although it can predict poor prognosis compared to acinar growth pattern. Kadota *et al.* have demonstrated that the 5-year probability of recurrence of the disease was worse in the cribriform predominant pattern compared to acinar and papillary patterns in stage I lung adenocarcinoma patients but comparable to solid and micropapillary, which are associated with poor prognosis (Kadota *et al.*, 2014). Cribriform can be regarded as moderately differentiated invasive histological pattern (Table 1.6).

PPA predominantly exhibits papillary growth and is a moderately-differentiated invasive tumour growth (Table 1.6).

Predominant micropapillary tumour growth is classified as MPPA and has been described as an invasive and poorly differentiated tumour growth, i.e. MPPA has completely lost the normal lung architecture (no more alveolar walls) (Table 1.6).

SPA is composed of solid tumour growth and is poorly differentiated tumour growth (Table 1.6) (Zhang *et al.*, 2014). Similarly to MPPA, SPA has also lost the normal lung architecture.

1.4.5.3   Lung adenocarcinoma patterns and patient outcome

Predominant histological patterns of tumours have been shown to be related to different survival outcome after surgery with curative intent (Hung *et al.*, 2014b; Russell *et al.*, 2011; Warth *et al.*, 2012).
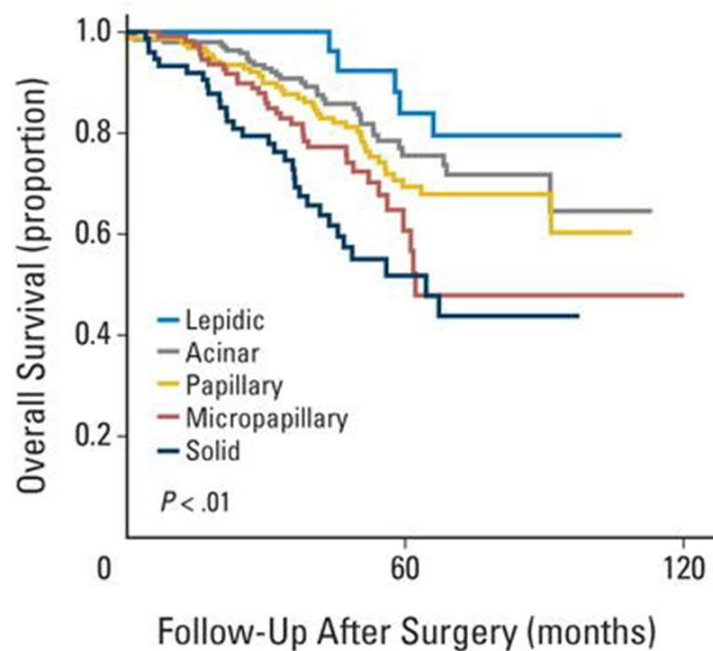
Figure 1.13. Histological pattern features and its impact on overall patient survival.
The Kaplan-Meier graphs describe survival of lung adenocarcinoma patient cohort according to the predominant growth pattern over 10 years follow-up after surgery. Kaplan-Meier of overall survival outcome for lepidic acinar, papillary, solid and micropapillary predominant patterns are shown from Hung *et al.* study (Hung *et al.*, 2014b).

AIS and MIA tumours presents the best overall survival in lung adenocarcinoma patients with nearly 100% of patient alive five years after surgery (Wilshire *et al.*, 2016).

LPA also predicts better outcome in lung adenocarcinoma patients with a 5-year overall survival rate of about 90%. APA and PPA show similar prognostic percentage with about 81% and nearly 80 %of patients alive 60 months after the surgery respectively. SPA and MPPA tumour patterns are related to poor outcome in lung adenocarcinoma with only about 53% and 50% of patients surviving five years post-surgery (Figure 1.13).

This correlation provides a possible grading system, which is based on the predominant growth pattern architecture. There is no formal grading of lung cancers at the moment. However, growth pattern seems to be the best and most reliable source to define the grading system. One has now been proposed. The International Association for the Study of Lung Cancer (IASLC) pathology committee elaborated a tumour grading system based on the predominant and high-grade tumour pattern for invasive pulmonary adenocarcinoma using one training cohort and two independent test cohorts. They found that dividing the tumour grading system into grade 1 (i.e. lepidic predominant tumour), grade 2 (acinar and papillary predominant tumours with no or less than 20% of high-grade histological patterns) and grade 3 (tumour with at least 20% of high-grade pattern, i.e. solid, micropapillary or more advanced tumour) gave a reproducible and significant outcome in invasive pulmonary adenocarcinoma with grade 1 related to better outcome and grade 3 associated to poor survival. This grading system offers a new prognostic value based on the histological pattern of the tumour (Moreira *et al.*, 2020).

1.4.5.4   *Pulmonary sarcomatoid carcinoma*

Sarcomatoid carcinomas are a poorly differentiated subtype of NSCLC. This group accounts for less than 1% of all lung cancer. This category consists of the following WHO tumour types: pleomorphic carcinoma, carcinosarcoma, spindle cell carcinoma, giant cell carcinoma and pulmonary blastoma. Of these, pleomorphic carcinoma is the most common.

Pleomorphic carcinoma is a rare and poorly differentiated squamous cell carcinoma or adenocarcinoma or undifferentiated NSCLC. It consists of a mixture of both epithelial and mesenchymal elements, exhibiting at least 10% of giant and/or spindle cells in addition to adenocarcinoma and squamous cell carcinoma elements. Giant tumour cells show large, irregular and multilobated nuclei and are often multinucleated, while spindle cells are elongated with malignant nuclear features and resemble malignant mesenchymal cells, morphologically suggestive of EMT.

More than 90% of lung cancer patients with pleomorphic carcinoma are heavy smokers however it can also develop in never-smokers due to asbestos or chemicals exposure (Travis, 2014). Pleomorphic carcinoma usually grows at the periphery of the lung and more precisely in the upper lobes. The differentiated region of the pleomorphic carcinoma can express cytokeratins, naspin A, TTF1, p63, CK5/6. The spindle or giant cell component can be positive for vimentin. Chromosomal gains and *TP53* mutations are common. Outcomes are poor even when diagnosed at an early stage of the disease (Travis, 2014).

1.4.6   Progression from *in situ* to invasive adenocarcinoma

Surgically resected lung adenocarcinoma can present both *in situ* and invasive tumour growth. Nonetheless, the continuum of *in situ* to invasive disease is poorly understood.

In a landmark study, Noguchi *et al.* showed that some early stage lung adenocarcinomas have limited metastatic potential. Tumours that are less than 20mm in size were classified into six groups based on their morphological differences. These categories are designated as A-F: type A tumours are wholly *in situ*, type B are tumours being biologically *in situ* with areas of stromal alterations ("pseudo-invasion"), type C tumours present *in situ* and truly invasive patterns and types D, E and F are tumours that are wholly invasive. Types A and B showed good pronosis with 100% 5-year survival rate. The 5-year survival of type C tumours were 74,8%. Type D tumours were found to have the worst prognosis with 52,4% patients alive five years post-surgery (Figure 1.14) (Noguchi *et al.*, 1995).
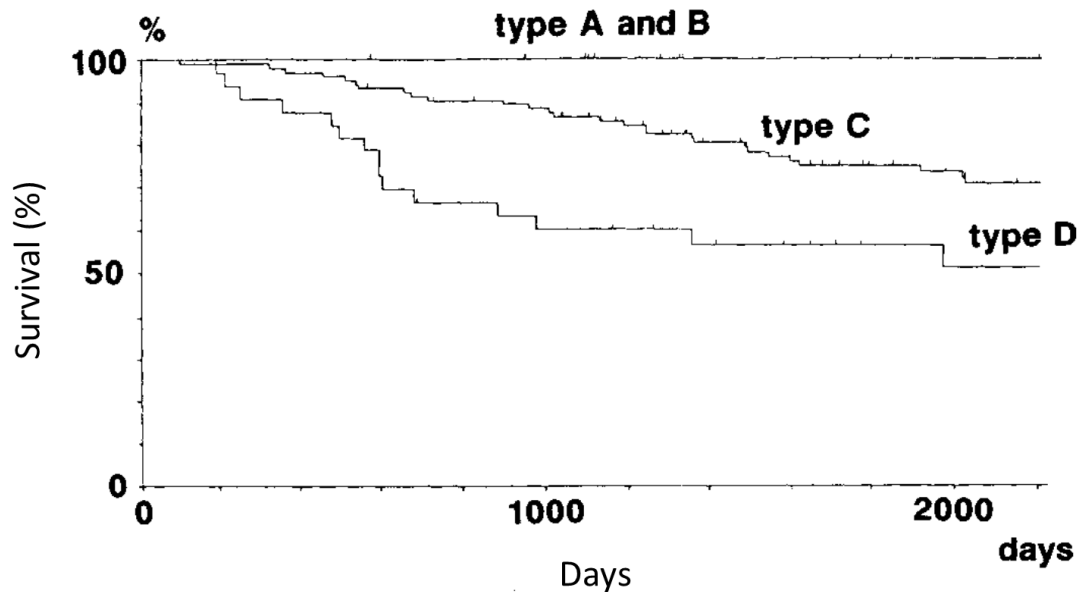
Figure 1.14. Survival rate of type A, B, C and D tumours in lung adenocarcinoma.
Adapted from (Noguchi *et al.*, 1995). Type E and F survival graphs are not shown in this paper due to the insufficient case numbers.

This suggests three important groupings based on biological invasiveness. The transition to invasion may not occur at all (e.g, Noguchi's type A tumours and B, equivalent to the WHO AIS category) where the tumour solely presents *in situ* growth with possible pseudoinvasion. Alternatively, the transition to invasion may occur very early in tumour growth leading to the formation of small but wholly invasive adenocarcinoma (Noguchi's type D tumours). Finally, invasion may be partial (Noguchi's type C tumours) with the tumour showing both *in situ* and invasive growth patterns.

Numerous studies have shown that genetic alteration is one of the main contributors to lung adenocarcinoma progression.

Yoo *et al.* examined 20 precursor lesions (atypical adenomatous hyperplasia, AAH), 43 lepidic predominant adenocarcinomas and 47 invasive-predominant adenocarcinomas. They found out that the frequency of *EGFR* mutations were quite consistent in precursor, *in situ* and invasive tumour growths while overexpression of p53 was more frequently observed in invasive adenocarcinoma. This suggested that *EGFR* mutation occurred at early stage of lung adenocarcinoma and p53 overexpression occurred at

late event in the disease development and was involved in the progression to invasiveness in lung adenocarcinoma (Yoo *et al.*, 2010).

Chen *et al.* investigated the genomic and immune profiles of lung adenocarcinoma using whole-exome sequencing in 98 AIS/MIA patients and 99 invasive adenocarcinoma patients. They found eight commonly mutated genes in the pre-invasive tumour growth, which includes *EGFR*, *KRAS*, *BRAF* and *TP53* mutations. Furthermore, they demonstrated that *TP53* gene aberrations were more commonly detected in the invasive adenocarcinoma and therefore were regarded as mediator of invasive growth in the progression of lung adenocarcinoma disease (Chen *et al.*, 2019).

However, these studies focused on comparing separate *in situ* and invasive tumour lesions from different lung adenocarcinoma patients, i.e. with different genomic backgrounds and not areas of the same tumour. This cannot illustrate the progression from *in situ* to invasive disease within a single lesion and will be confounded by differences between patients (e.g. different driving mutations, ethnicity, smoking history). It is essential to understand what drives the transition to invasiveness within individual tumours in the evolution of lung adenocarcinoma to understand the mechanism by which invasive behaviour is acquired. A few studies have investigated the molecular causes in the continuum progression from *in situ* disease to invasive adenocarcinoma within individual tumours.

Yatabe *et al*. analysed topographical distribution of mutations in three different regions of lung cancer tumours. They inspected in total 107 samples (17 AAH pre-invasive lesions, 21 AIS tumours, 23 MIA tumours and 46 invasive adenocarcinomas) from 48 patients. Using Taq Man-based gene dosage analysis and fluorescent *in situ* hybridization technology, they showed that *EGFR* gene amplification was observed in invasive areas of early adenocarcinoma but not in the *in situ* components: 1 out of 20 AIS tumours presented *EGFR* gene amplification while none of the AAH lesions did. Therefore, *EGFR* gene amplification was associated with lung adenocarcinoma progression (Yatabe *et al.*, 2008).

Murphy *et al.* discovered that progression from *in situ* disease to invasive disease can be due to increased genomic rearrangement such as *EML4-ALK* gene fusion or intragenic *ALK* gene rearrangement in the invasive tumour using next generation DNA-sequencing in 14 cases of lung adenocarcinomas with *in situ* and invasion lesions within individual tumours (Murphy *et al.*, 2014).

The Sidransky group also explored 25 AAH (*in situ*) lesions and their matched invasive adenocarcinomas from 6 patients and different areas of histological progression of lung adenocarcinoma within AISs and MIAs tumours from 5 AIS and 5 MIA using targeted next-generation sequencing. They found that *EGFR* and *KRAS* could be detected in early step of lung adenocarcinoma formation while *TP53* genes alterations were moslty found in invasive areas witihn MIA and AIS tumours, suggesting that *TP53* gene mutations favor the progression to invasive tumour growth in lung adenocarcinoma (Izumchenko *et al.*, 2015).

A recent study from our own laboratory incorporating data from this thesis has demonstrated that in mixed-type tumour presenting *in situ* and invasive components, i.e. type C tumours of Noguchi, the progression of *in situ* to invasive disease can be driven by *TP53* mutations while the common drivers *EGFR* and *KRAS* gene mutations were present in both *in situ* and invasive areas. Nonetheless, many tumours had no detectable additional driver mutation in invasive areas, which leads us to suspect other genetics or epigenetics events such as amplification, or translational dysregulation may also drive invasiveness (Moore *et al.*, 2019).

The concept of clonality has arisen to explain the progression to invasiveness in lung adenocarcinoma. Tumour clones evolve by acquiring driver mutations, i.e. mutation enabling the cancer cell to grow and survive in the tumour microenvironment, and generate progeny (subclones) that have acquired further mutations over time and manage to survive in the microenvironment. Identifying these driver mutations and their order of generation (i.e. whether genes is mutated early or late in the progression in lung adenocarcinoma) is crucial to provide accurate and efficient treatment (Govindan, 2014). The TRACERx study (TRAcking non-small cell lung Cancer Evolution through therapy [Rx]) is an ongoing multicentre study in UK that investigates the tumour

spatio-temporal evolution and the importance of clonal heterogeneity upon therapeutic and survival outcome by sampling multiple regions of primary and metastatic tumour, although it does not separately examine *in situ* and invasive disease. This study demonstrated that most classic driver mutations such as *EGFR*, *KRAS* and *TP53* gene alterations occurred at early stage of lung adenocarcinoma and patients with high amount of copy-number alterations in subclones significantly predicted poor outcome in NSCLC patients (Jamal-Hanjani *et al.*, 2017).
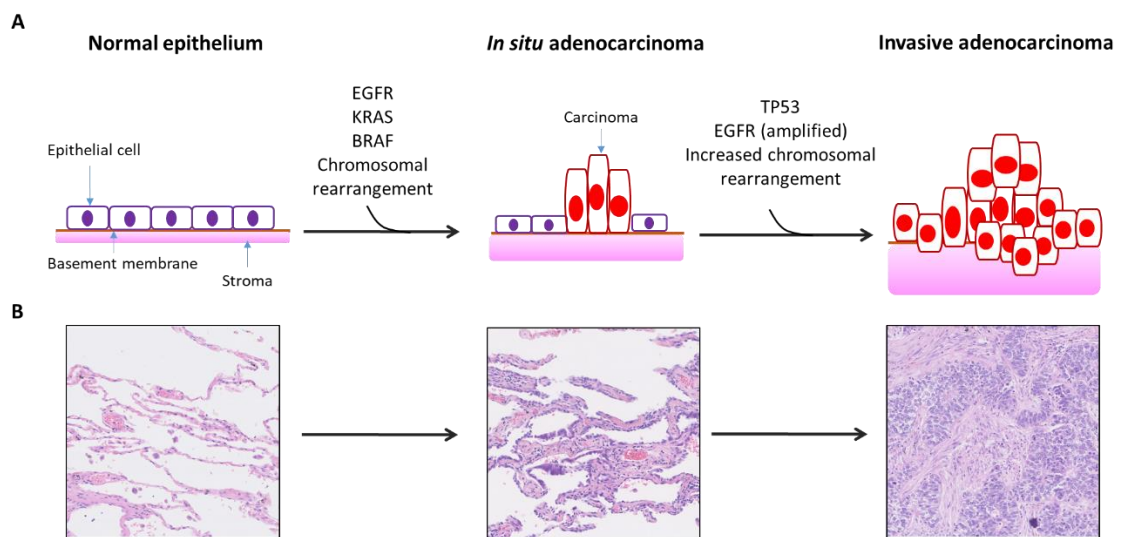
Figure 1.15. Simplified Vogelgram of lung adenocarcinoma disease progression.
A: Simplified schematic representation of commonly foung genetic aleterations in the continuum from *in situ* to invasive lesions in lung adenocarcinoma. B: H&E images of normal and tumorigenic lung illustrated the Vogelgram of lung adenocarcinoma progression. Normal lung epithelium progress to *in situ* disease, which can lead to the formation of invasive adenocarcinoma. Cytoplasm and nuclei are stained in pink (eosin staining) and purple (haematoxylin staining) respectively.

All these findings from published literature are summarised in a "Vogelgram" to illustrate the most common alterations found in the continuum from *in situ* to invasive disease (Figure 1.15). A normal alveolar epithelial cell acquires genetic alterations such as *EGFR*, *BRAF* and/or *KRAS* gene mutations and chromosomal rearrangements leading to *in situ* adenocarcinoma growth. At this stage, the carcinoma can further accumulate genetic changes such as *TP53* gene mutations, *EGFR* gene amplification or increased chromosomal rearrangement, which seems to coincide with the formation of invasive adenocarcinoma with destruction of normal lung architecture (Figure 1.15).

To summarise, these studies have demonstrated that most lung adenocarcinoma tumour present intratumoural heterogeneity and have introduced the notion of spatio-temporal diversity in tumour clone evolution to understand the continuum of lung cancer progression.

Although genomic alteration is believed to drive invasiveness in lung adenocarcinoma, the phenotypic events underpinning this event in human tumours are still very poorly understood.

1.5 Epithelial-mesenchymal transition (EMT)

1.5.1 Epithelial and mesenchymal cells features

1.5.1.1 Epithelial cells

In metazoan organisms, the epithelia are tissues made up of layers of cytokeratin-expressing cells, which cover or line the surface of many organs such as the lung, secretory glands, gut or skin. They play multiple roles, including secretion, gas exchange, and as a barriers to harmful or pathogenic molecules. When malignantly transformed, they give rise to carcinomas.

Epithelia can be classified as squamoid (thin flat cells), columnar or cuboidal, and can be arranged in one layer (simple epithelium) or several layers (stratified epithelium) (Ye and Weinberg, 2015). Tissue epithelia are composed of differentiated non-motile cells presenting a range of types of intercellular junctions: tight and *adherens* junctions at the subapical lateral surface of epithelial cell membrane, followed by desmosomes and gap junctions at the basolateral epithelial membrane (Lamouille *et al.*, 2014). Normal epithelial cells exhibit apical-basal polarity and are anchored to a basement membrane (Figure 1.16A). Intercellular junctions and polarity contribute to the integrity of epithelial cells that protect metazoan organisms from external environment (Garcia *et al.*, 2018).

Lung epithelial cells contain all the classical types of intercellular junctions, described above, to allow their optimal and efficient activities (Figure 1.16C).

Normal epithelium

A

Cilia
Apical
Epithelial cell
Tight junction
Adherens junction
Nucleus
Desmosomes
Gap junction
Basal
Basement membrane
Extracellular matrix

B

Lung alveolus normal epithelium

Apical
Alveolar epithelial cell type II
Alveolar epithelial cell type I
Basal
Extracellular matrix

C

Claudin
Connexins
E-cadherin
Desmosomal cadherin

D

Erythrocytes
Type 2 pneumocyte
Type 1 pneumocytes

Figure 1.16. Simplified Schematic representation of a normal epithelium in general and in lung alveoli. A: Schematic representation of the normal epithelium. B and C: Schematic representation of the lung alveolus epithelium with a zoom on the cell junctions present between to AT1 cells. D: H&E of type 1 and 2 pneumocytes (i.e AT1 and AT2 respectively) and erythrocytes of the human normal lung alveolus. Dark and light purple stain the nuclei and cytoplasm of AT1 and 2 respectively. Pink stains erythrocytes (Atlas).

In the lung alveoli, epithelial cells possess tight junctions generally positioned at the apical membrane structure of epithelial cells, which are linked to the actin cytoskeleton. AT1 cells and AT2 cells both express claudin, which interacts with occludin and scaffold proteins ZO-1 and ZO-2 to regulate the barrier functions of tight junctions and control paracellular permeability (Figure 1.16). They principally regulate the movement of ions and solutes between cells and maintain homeostasis across the epithelium via paracellular diffusion (Hartsock and Nelson, 2008; Itoh and Bissell, 2003).

Adherens junctions or intermediate junctions are mostly positioned beneath the tight junctions at the lateral surface of the cell membrane. The transmembrane protein E-cadherin is the major component of these junctions and interacts with adjacent cells and the intracellular protein p120 catenin in most epithelium, including in the alveoli epithelium (Chignalia *et al.*, 2015; Hartsock and Nelson, 2008). *Adherens* junctions maintain apico-basal polarity (Nawijn *et al.*, 2011), stabilise cell-cell adhesion and control the actin cytoskeleton amongst several others functions (Hartsock and Nelson, 2008).

Gap junctions are intercellular communicant canals formed of connexins. Their essential role is to allow direct diffusion of small molecules, metabolites and ions between cells, including in alveolar epithelial cells (Goodenough and Paul, 2009). In the alveoli, gap junctions are mostly expressed between AT1 cells in close proximity to tight junctions (Koval, 2002).

Desmosomes are intercellular junctions that are positioned at the basolateral of the membrane. They are comprised of proteins such as desmosomal cadherins, desmosomal proteins and armadillo family proteins. In the lung alveoli, they form a connection between intermediate filaments of the cytoskeletons of the adjacent cells (Paine *et al.*, 1995). Their functions are to maintain a strong cell-cell contact between cells and resist to mechanical stress (Kowalczyk and Green, 2013).

Epithelial cells exhibit an apical-basal polarity which is controlled by the polarity complexes that interact with cell junction architecture known as Crumbs, Scribble and PAR complexes (Lamouille *et al.*, 2014). Epithelial cells polarity contributes to the

maintenance of epithelial tissue organisation. The apical and basolateral plasma membranes domain possess different composition of proteins and lipid allowing them to execute directional transport of nutrients and waste between the two sides of the epithelium and therefore maintain the homeostasis in the epithelial cell (Rodriguez-Boulan and Macara, 2014). Apical-basal polarity also enables epithelial cells to regulate the cell division asymmetry and maintain the apical cell junctions (the tight and *adherens* junctions) (Royer and Lu, 2011).

The cytoskeleton of epithelial cells is composed of intermediate filaments such as cytokeratins amongst other filaments, which contribute to the mechanical assistance to the cell and maintain the shape and intracellular organisation (Lamouille *et al.*, 2014).

Epithelial cells are attached to the basement membrane through the interaction of integrins with extracellular matrix proteins present in the basement membrane (Lamouille *et al.*, 2014). The basement membrane separates the underlying cells from adjacent tissues and acts as mechanical support to the cells as well as a barrier to invasion (Liotta *et al.*, 1980).

1.5.1.2 Mesenchymal cells

Mesenchymal stem cells (MSC) are multipotent stem cells that can self-renew and differentiate into mesenchymal lineage cells in adult organism. They give rise to mesenchymal tissues namely bones (osteocytes), cartilage (chondrocytes), muscle (myotube), marrow (stromal cells), adipose tissue (adipocytes) or tendon/ligament (fibroblast) (Caplan, 2015).

The International Society for Cellular Therapy defines MSC as cells that exhibit plastic adherent features, express a specific cluster of differentiated cell surface and can differentiate into adipocyte, chondroblast and osteoblast *in vitro* (Dominici *et al.*, 2006). They are spindle fibroblast-like cell that commonly express CD73, CD90 and CD105 and other tissue- specific surface markers and are negative for CD45, CD34, CD14 or CD11b, CD79a or CD19, and human Leucocyte antigen-DR, i.e. hematopoietic markers (Ullah *et al.*, 2015).

MSC are derived from perivascular cells, named pericytes and therefore can be found in many tissues of the adult organism. Pericytes are fibroblast-like cells that are present at intervals around along the wall of capillaries and venules (Attwell *et al.*, 2016).They are mainly activated in response to injury/inflammation. Briefly, upon injury, pericyte is released in the injury site and differentiated into a MSC. MSC is activated and witnesses the dynamic changes in the microenvironment. It then reacts accordingly by secreting immuno-mediators and anti-bacterial peptides to protect injured microenvironment against pathogens and by producing trophic molecules to initiate regenerative tissue repair process (Caplan, 2015).

In the context of epithelial-mesenchymal transition, mesenchymal cells are phenotypically and functionally opposite to epithelial cells. They are undifferentiated spindle-like shaped cells, losing the apical-basal polarity. They do not form tight and stable cell-cell junctions crucially due to the switch of the E-cadherin to N-cadherin expression. They are motile and invasive. The intermediate filament cytokeratin is replaced by intermediate filament vimentin. They possess the characteristics of stem cells as they are capable of differentiating into epithelial cells (Ye and Weinberg, 2015).

1.5.2   Types of EMT

As discussed earlier, cancer metastasis is the main cause of cancer mortality worldwide. In order to metastasize to distant sites, cancer cells need to detach from the primary tumour, intravasate and extravasate the vascular or lymphatic circulations, invade their new host tissue, and survive and proliferate to form a distant secondary tumour (Seyfried and Huysentruyt, 2013). One of the early crucial events in this series is believed to be epithelial-mesenchymal transition (EMT).

EMT is a process, which manifests as the dedifferentiation of epithelial cells into mesenchymal-like cells accompanied by molecular changes. They lose organisation of cell-cell junctions and cell polarity and gain in motility and invasive potential. These morphological changes are made possible due to the plasticity acquired by epithelial cells.

EMT is categorised into three distinct biological subtypes. Type 1 EMTs are required during embryogenesis and organogenesis. In this scenario, EMT plays a physiological role. Briefly, EMT process is involved in several stages in embryonic development but gastrulation and neural crest formation are the best studied stages for EMT process. The gastrulation stage is responsible for the production of three fundamental germ layer of every metazoan: ectoderm, mesoderm and endoderm. During this stage, EMT intervenes in the generation of mesoderm and endoderm from subset epithelial cells from the epiblast that gives rise to the ectoderm, at the primitive streak (Figure 1.17A). Through several rounds of EMT and its reverse process MET (mesenchymal-epithelial transition), the mesoderm and endoderm will ultimately give rise to many precursors cell types, leading to formation of tissues and organs of the metazoan adult organism. During the neural crest migration, neural epithelial cells in the dorsal neural tube undergo EMT to migrate as mesenchymal cells to their final destination and differentiate to ultimately generate neurons of the peripheral nervous systems, bone, melanocytes, amongst other components of the human body (Acloque *et al.*, 2009). Epithelial-mesenchymal transition is essential in embryonic development. Defective EMT process during migratory neural crest cells or gastrulation can result in congenital malformations or even embryonic lethality (Acloque *et al.*, 2009) (Figure 1.17B).
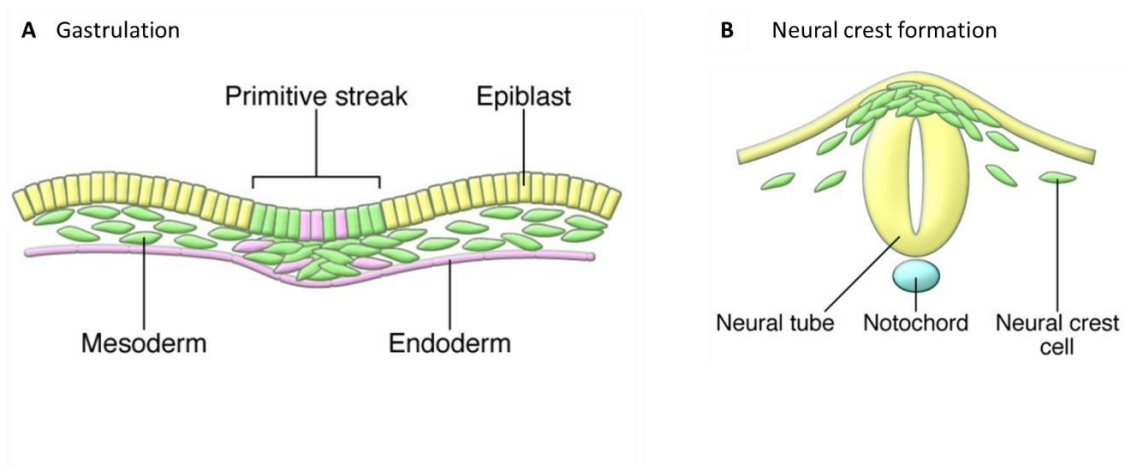
Figure 1.17. Cartoon of type 1 EMT.
A and B: Schematic representations of EMT during gastrulation and neural crest formation. Adapted from (Acloque *et al.*, 2009).

Type 2 EMT is implicated in wound healing, tissue repair/regeneration and fibrosis where EMT again has a physiological role (Figure 1.18). Succinctly, during wound healing

and tissue repair/regeneration process, EMT is involved in reconstructing the damaged tissue in response to trauma or inflammatory injury. Transient EMT is activated allowing epithelial cell migration to re-epithelize the wounded region and myofibroblast differentiation to remodel the extracellular matrix in order to repair wounded tissue (Stone *et al.*, 2016). Notwithstanding, sustained EMT program activation due to ongoing inflammation, resulting in excessive production of myofibroblasts, can lead to organ fibrosis such as pulmonary fibrosis (Hill *et al.*, 2019; Königshoff *et al.*, 2009), renal fibrosis (Iwano *et al.*, 2002) or hepatic fibrosis (Zhao *et al.*, 2016), with deposition of extracellular matrix molecules (Stone *et al.*, 2016).
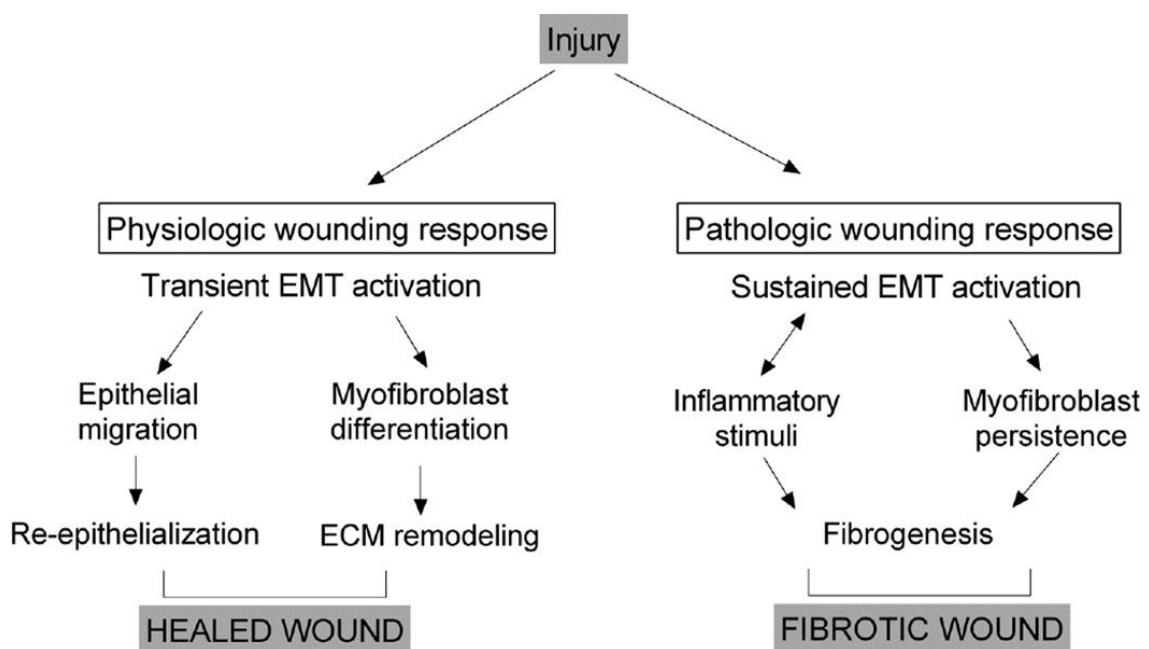
Figure 1.18. Type 2 EMT.
Schematic representation of EMT activation upon injury (Stone *et al.*, 2016).

Type 3 EMTs contribute to cancer progression by enabling carcinomas to become invasive and metastatic (cf section 1.5.4). Although the EMT program can be activated under different physiological contexts, the molecular mechanisms underpinning these events are largely conserved (Kalluri and Weinberg, 2009; Lamouille *et al.*, 2014; Zeisberg and Neilson, 2009).

1.5.3   Regulators of EMT

The well-known EMT transcription factors SNAIL, TWIST and ZEB families drive EMT through the downregulation and upregulation of epithelial and mesenchymal promoting genes respectively. The name of SNAIL and TWIST comes from their discovery in *Drosophila melanogaster*. SNAIL and TWIST play a crucial role in the formation of the mesoderm and the neural crest in Drosophila (Alberga *et al.*, 1991; Leptin, 1991). As illustrated in Figure 1.19, the expression of these transcription factors is induced at an early stage of EMT by several pathways such as hypoxia, inflammation and integrins from the extracellular matrix and secreted soluble proteins such as transforming growth factor-β, or tyrosine kinase, Wnt, Notch and hedgehog signalling pathways. Interestingly, several growth factors such epidermal growth factor, hepatocyte growth factor and fibroblast growth factor are ligands of receptor tyrosine kinase which leads to a cascade of activation of downstream factor involved in the translation control, i.e. the phosphatidylinositol-3-kinase (PI3K)/AKT pathway (Lamouille *et al.*, 2014) and therefore might be implicated in the translational reprogramming in the tumour cells when *EGFR* gene is mutated.

SNAIL proteins family consists of three SNAIL1 aka SNAIL, SNAIL2 aka SLUG, and SNAIL3 aka SMUC. These transcription factors repress E-cadherin and cytokeratin mRNA transcription by binding to specific sequences of E-cadherin and cytokeratin promoters and inducing epigenetic modifications such as histone methylation which prevent proper transcriptions of the epithelial genes. They also induce mesenchymal gene transcription by binding to their promoter regions (Lamouille *et al.*, 2014).

TWIST genes encode homodimeric or heterodimeric basic helix–loop–helix proteins. TWIST1 and TWIST2 act similarly and independently to SNAIL proteins. TWIST1 recruits set8 methyltransferase, which methylates both cadherins' promoters leading to the repression of E-cadherin and activation of N-cadherin transcription (Lamouille *et al.*, 2014).

ZEB proteins are zinc finger transcription factors and two homologous of ZEB proteins are expressed in the vertebrates: ZEB1 and ZEB2. ZEB can be induced by SNAIL or TWIST

(Lamouille *et al.*, 2014). ZEB1/dEF1 and ZEB2/SIP1 complexes can independently repress epithelial markers expression by directly binding to the 5'-CACCT sequences located in the promoter of the gene. They also contribute to the induction of mesenchymal phenotype by activating the transcription of N-cadherin and vimentin transcriptions (Vandewalle *et al.*, 2009).
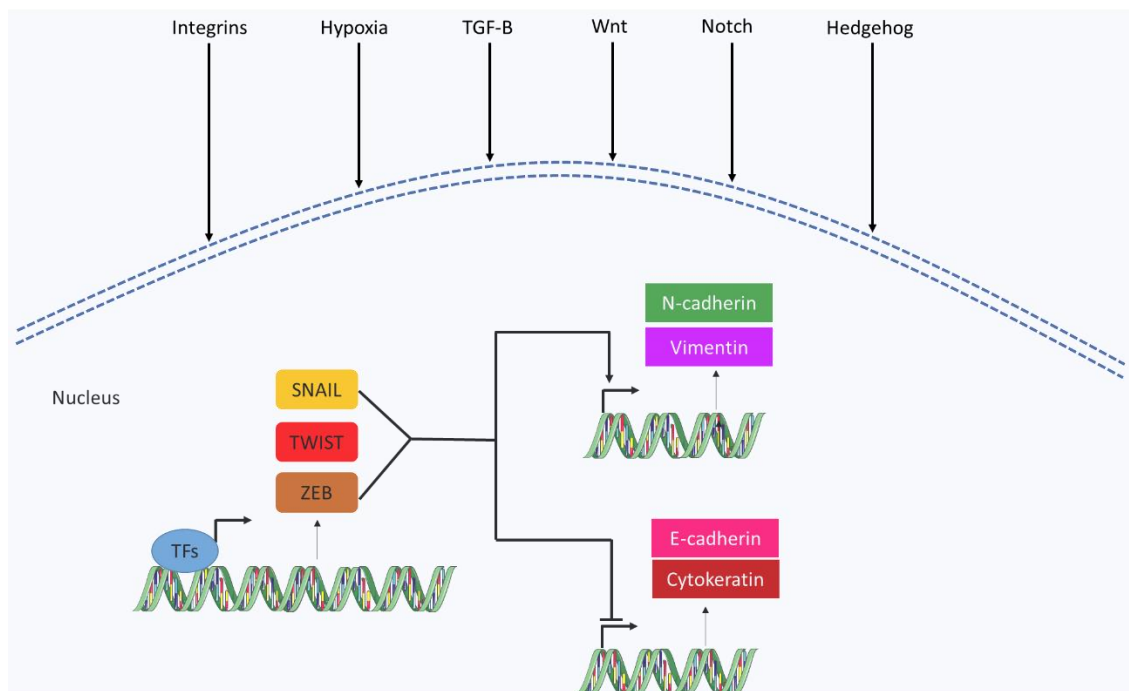
Figure 1.19. Signaling pathways of EMT activation.
Different pathways such as integrins, hypoxia, WNT, Notch and Hedgehog can activate EMT program. This triggers the transcription and translation of EMT-related transcriptions factors SNAIL, TWIST or ZEB by TFs (transcription factors) which subsequently leads to the activation of transcription of N-cadherin and vimentin and the repression of E-cadherin and cytokeratin transcription.

Studies have shown the importance of microRNA as regulator of EMT as well. For instance, overexpression of miR-10b is implicated in the tumour invasion and metastasis in breast cancer (Ma *et al.*, 2007)*.* MiR-200 family and miR-205 prevent EMT activation by repressing ZEB family transcription factors (Gregory *et al.*, 2008; Park *et al.*, 2008).

Although EMT is widely accepted as a mechanism central to tumour biology in *in vitro* studies, classical morphologically dedifferentiated mesenchymal cells are rarely seen in resected primary tumours; when it is seen, this is recognised as sarcomatoid differentiation (i.e. tumour presenting mesenchymal differentiation such as spindle

cells), and is associated with very poor outcomes, indicative of rapid spread and tumour aggressiveness. One key aim of this project is to resolve this apparent paradox.

1.5.4   EMT in cancer progression

1.5.4.1   Description of EMT program

Studies have investigated the role EMT in cancer progression and metastasis using *in vitro* experiments and mice models (Onder *et al.*, 2008).

Figure 1.20 describes the currently generally accepted model of epithelial-mesenchymal transition in the context of cancer progression.
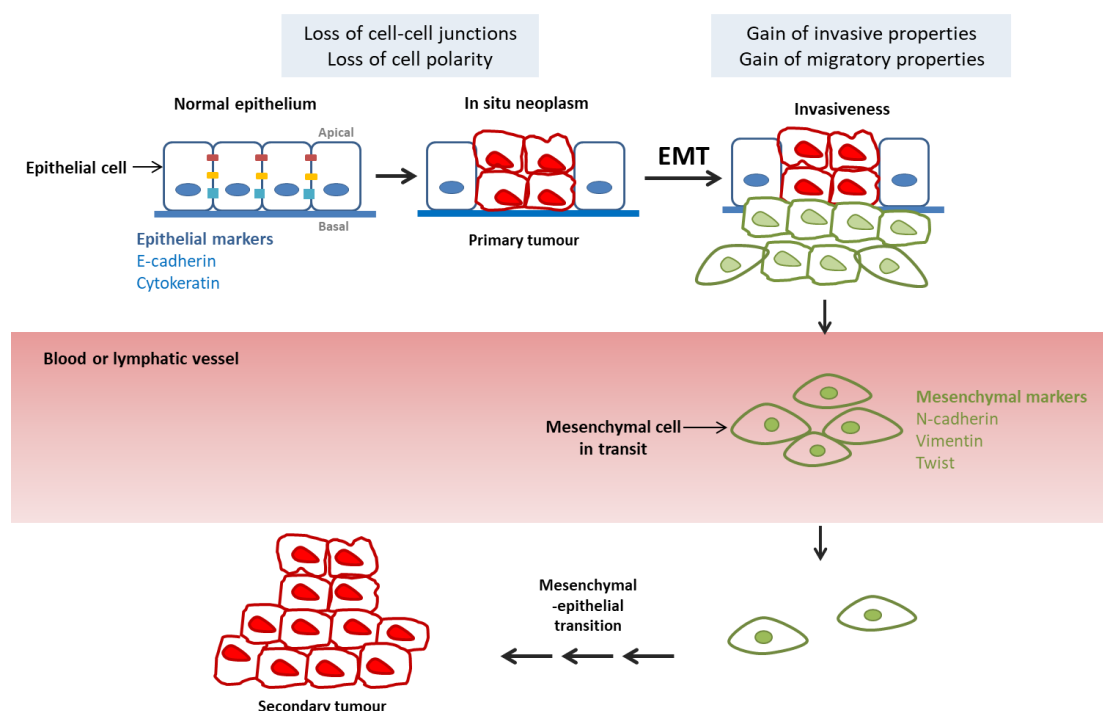
Figure 1.20. Idealised schematic model of EMT process.
In a normal epithelium, cells contain organised diverse types of cellular junctions and exhibit an apical-basal polarity. The basal layer adheres to the basement membrane. Loss of cell junctions and polarity is part of the progression from normal epithelium to *in situ* neoplasm. These transformed epithelial cancer cells then undergo cytoskeletal reorganisation, altered morphology, and acquire the ability to degrade or pierce the basement membrane. They become functionally mesenchymal and gain invasive and migratory properties. This represents EMT. The invasive cells may then enter and exit the blood or lymphatic circulation by intravasation and extravasation. They then undergo the reverse process of EMT called mesenchymal–epithelial transition and form secondary tumour in a distant site.

Epithelial cells undergo genetic alterations and become transformed cancer cells within a carcinoma. The important steps in EMT are believed to be the degradation of cell junctions, the loss of cell polarity, restructuring of the cytoskeleton, modification in cell shape, and degradation of the basement membrane, as a result of the repression of epithelial gene expression and activation of mesenchymal related gene expression. The decrease of E-cadherin and increase of N-cadherin expression, known as the 'cadherin switch' and the remodelling of the cytoskeleton by the replacement of cytokeratin filaments with vimentin filament are the major hallmarks of EMT (Lamouille *et al.*, 2014; Ye and Weinberg, 2015).

The transition of epithelial tumour cells starts with the dissolution of cell junctions. This event is defined by the down-regulation of proteins implicated in cell junction. The foremost downregulation observed during EMT is the E-cadherin expression while N-cadherin is induced which weakens the cell-cell adhesion. The degradation of these junctions subsequently leads to the disappearance of the apical-basal polarity since the polarity complex proteins such as Crumbs or Scribble are unable to interact with proteins constituting the junctions (Lamouille *et al.*, 2014; Yilmaz and Christofori, 2009).

The epithelial cytoskeletal changes that occur are represented by the replacement of cytokeratin to vimentin intermediate filament (Onder *et al.*, 2008; Tsai *et al.*, 2012). The cell motility is acquired by reorganising and reinforcing actin expression which allows the epithelial cell to form membrane projections to promote cell movement (Lamouille *et al.*, 2014).

The mesenchymal-like tumour cells may subsequently intravasate into lymphatic or blood circulation, extravasate at distant site, re-epithelize via mesenchymal-epithelial transition, proliferate and form secondary tumour in distant organ (Tsai *et al.*, 2012).

This is how EMT program is believed to play a major role in cancer progression and metastasis. Loss of epithelial markers, i.e. E-cadherin and cytokeratin and gain of mesenchymal markers, i.e. EMT-related transcription factors SNAIL, TWIST and ZEB, N-cadherin and cytokeratin can predict patient outcome.

1.5.4.2 Molecular EMT is associated with patient outcome

Invasion and metastasis is one of the hallmarks of cancer. The status of EMT as being essential for cancer progression and metastasis has been subject for debate for several decades due to the fact that morphological EMT is often not observed in primary tumour tissue (Tarin, 2005).

However, several studies have witnessed correlations between molecular EMT and cancer progression and patient outcome in solid tumours. In breast cancer, Li *et al*. have performed a meta-analysis on 7353 patients from thirty-three studies and have demonstrated that reduced E-cadherin expression was associated with increased tumour size and histological grade, lymph node positivity as well as poor overall and decreased disease-free survivals (Li *et al.*, 2017). Vora *et al*. have demonstrated that gain of vimentin expression was related to lymph node metastasis and recurrence-free disease while loss of cytokeratin was mainly noticed in stage IV of breast cancer patient and was associated with poor outcome (Vora *et al.*, 2009). In colorectal cancer, studies have shown that loss of E-cadherin is correlated with infiltrative tumour histological pattern (Kim *et al.*, 2016), lymph nodes metastasis (Elzagheid *et al.*, 2012; Kim *et al.*, 2016; Yun *et al.*, 2014) and poor prognosis (Elzagheid *et al.*, 2012). Increased expressions of vimentin and SLUG (aka SNAIL2) were correlated with poor prognosis and short remission period of the disease (Du *et al.*, 2018; Toiyama *et al.*, 2013) as well as TNM stage and lymph nodes metastasis (Du *et al.*, 2018; Toiyama *et al.*, 2013). In prostate cancer, Umbas *et al*. have demonstrated a significant relationship between E-cadherin loss and worse patient outcome (Umbas *et al.*, 1994).

Furthermore, in NSCLC, Sulzer *et al.* have demonstrated that decreased E-cadherin expression was implicated in the progression of NSCLC by being associated with increased lymph node metastasis and indicator of poor outcome in NSCLC patients (Sulzer *et al.*, 1998). In line with this, Zhang *et al.* have further investigated in lung squamous cell carcinoma and indicated that reduction in E-cadherin expression and augmentation in vimentin expression were related to lymph node metastasis and poor prognostic (Zhang *et al.*, 2013).

However, the relationship between EMT and clinic pathological factors is still unclear in lung adenocarcinoma. Two of studies have shown that up-regulation of SLUG (Shih *et al.*, 2005) or reduced E-cadherin in combination with increased SM1004 expression (mesenchymal marker) is significantly related with poor outcome in lung adenocarcinoma (Miyazaki *et al.*, 2006). TWIST1 was associated with poor prognosis in stage IV lung adenocarcinoma patients (Liu *et al.*, 2018).

One of the key aims of this project is to investigate whether EMT program, i.e. loss of both E-cadherin and cytokeratin and increase of N-cadherin and vimentin expression simultaneously, can be detectable in human primary lung adenocarcinoma and whether it is related to physio-pathological factors.

Mechanisms of EMT activation in human primary lung adenocarcinoma remains poorly described. Although several studies have been investigating the role of EMT-related transcription factors in influencing the EMT program reactivation in primary tissue, very little work has explored the potential role of translation deregulation leading to activation of EMT-TFs in human primary lung adenocarcinoma. *In vitro* and *in vivo* studies have indeed shown the importance of translational reprogramming in cancer biology and in the activation of some of the hallmarks of cancer (Eberle *et al.*, 1997; Silvera *et al.*, 2009; Wendel *et al.*, 2004). Therefore, I wanted to investigate whether translation dysregulation can favour the invasion and metastasis of human primary lung adenocarcinoma by instituting an EMT program.

1.6 Control of mRNA translation

1.6.1 Mechanism of mRNA translation in eukaryotes

Gene expression is tightly regulated by many mechanisms, but in general the most control over protein levels is exerted by the regulation of mRNA abundance (transcription/mRNA stability) and translational efficiency (Figure 1.21).
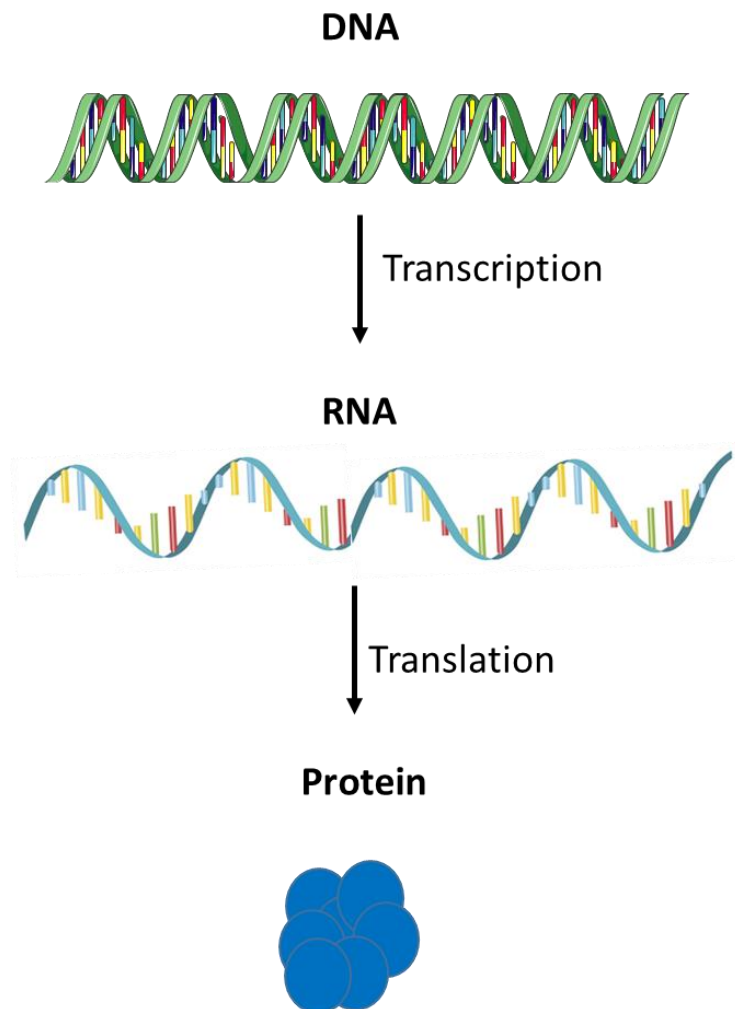
Figure 1.21. Central dogma of gene expression.
Schematic representation of the central dogma of gene expression. DNA is transcribed into mRNA which is subsequently translated into a functional protein.

Several studies have shown that translational control is the single most influential regulator of protein levels (Sonenberg and Hinnebusch, 2009), and translational programming has been shown to influence many aspects of cellular biology (Bhat *et al.*, 2015).

Protein synthesis is an energy-consuming process which relies on the coordinated interaction of transfer RNAs (tRNAs), ribosomes, translation factors and mRNA (Bhat *et al.*, 2015). Translation consists of four phases: initiation, elongation, termination and ribosome recycling (Pelletier *et al.*, 2015).

The initiation phase is often where the most control is exerted (Sonenberg and Hinnebusch, 2009). Most mRNAs are translated by a mechanism, which involves recognition of the modified 5' G nucleotide at the 5' terminus ("cap-dependent" initiation). It involves binding of the eukaryotic initiation factor 4F complex (eIF4F) to the 5' mRNA cap of the mRNA. This facilitates the recruitment of the 40S ribosomal subunit, the ternary complex and eukaryotic initiation auxiliary factors. 40S subunit scanning of 5' untranslated region (UTR) of the mRNA then occurs, followed by 60S ribosomal subunit joining leading to the formation of 80S ribosome (Sonenberg and Hinnebusch, 2009).

The elongation phase begins with an elongation-competent 80S ribosome, which contains the anticodon initiator methionyl-tRNA hybridized to the start codon AUG of the mRNA and the successive codon recognition of the mRNA coding sequence by the aminoacyl-tRNA coupled to the synthesis of the nascent polypeptide chain. This process requires the help of the eukaryotic translation elongation factors eEF1, eEF2 and eEF3.

The termination phase is defined by the recognition of the stop codon (UAA, UGA, UAG) by eukaryotic release factors and the liberation of the nascent polypeptide chain.

The ribosome recycling is finally performed wherein the ribosomal subunits are dissociated from the mRNA with the help of protein factors (Dever and Green, 2012).

1.6.2 Translation initiation

Most of translational control is operated at the translation initiation level. Thus, it is thought to be the crucial step in translation and the regulation of gene expression.
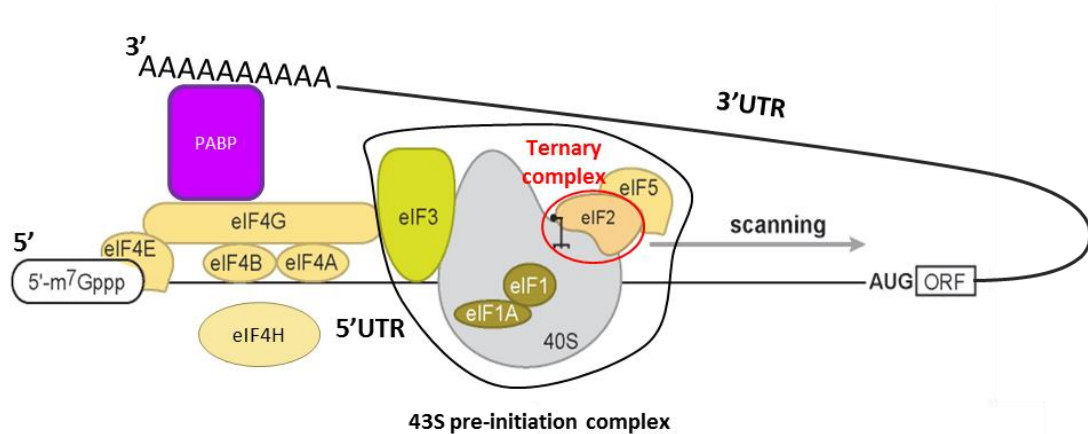
Cap-dependent translation of an mRNA

Figure 1.22. Schematic representation of translation initiation.
In eukaryotes, translation initiation is generally mRNA cap-dependent and involves the eIF4F complex composed of eIF4E, eIF4A and eIF4G, poly (A) binding protein (PABP), the ternary complex comprised of eIF2-GTP- initiator methionyl-tRNA. Firstly, the 43S pre-initiation complex is recruited to the mRNA through the interaction between eIF3 and eIF4G, which is mediated by the RNA helicase activity of eIF4A. The 43S complex then scans the 5' UTR of the mRNA until it reaches the start codon. eIF4H and eIF4B are co-activators of eIF4A RNA helicase activity**.**

As depicted in Figure 1.22, the binding of eIF4F complex to the 5' 7-methylguanylate mRNA cap triggers the recruitment of the 43S pre-initiation complex, which is facilitated through eIF3 via its interaction with eIF4F. This complex, via the unwinding activity of eIF4A in conjunction with its co-factors eIF4B and eIF4H, is able to scan the 5'UTR of the mRNA for the first initiation codon (Jackson *et al.*, 2010). AUG start codon recognition engenders the hydrolysis of the GTP bound to eIF2 into GDP, the release of the eukaryotic initiation factors and joining of the larger 60S ribosomal subunit. The formation of the active 80S ribosome (40S and 60S subunits) marks the end of initiation and the beginning of elongation.

In this study, I will focus on the eIF4F complex and eIF2 ternary complex, as these two complexes have been most implicated in the regulation of translation initiation.

1.6.2.1 eIF4F complex

eIF4F is a heterotrimeric complex. It is composed of the cap binding protein eIF4E, the scaffolding protein eIF4G and the DEAD-box RNA helicase eIF4A. Each component of the complex plays an essential role in translation initiation.

eIF4E binds the 5' 7-methylguanylate mRNA cap of the mRNA (Sonenberg and Hinnebusch, 2009).

eIF4A possesses an ATP-dependent helicase activity which unwinds the secondary structures encountered in the 5'UTR of the mRNA. Its activity is stimulated by its co-activators eIF4B and eIF4H (Rogers *et al.*, 2001). Three paralogues of eIF4A exist in the human genome, two of which are cytoplasmically expressed and are involved in translation: eIF4A1 and eIF4A2. Despite the sequence similarities of these two proteins, they are currently believed to be functionally distinct (Lu *et al.*, 2014). eIF4A1 is abundant in proliferative cells whereas eIF4A2 is up-regulated in differentiated cells (Raza *et al.*, 2015; Williams-Hill *et al.*, 1997).

eIF4G carries binding sites for eIF4E, eIF4A, eIF4B, eIF3 and poly (A) tail binding protein (PABP), hence its attribution as scaffold protein. By interacting with eIF4G, PABP helps to stabilise the mRNA in a circular shape to facilitate translation (Derry *et al.*, 2006).

1.6.2.2 Ternary complex

The ternary complex is composed of eIF2 bound to GTP and the initiator methionyl-tRNA. Its function is to deliver the initiator methionyl-tRNA to the ribosome. The activity of this complex is mainly dictated by the status of eIF2 and eIF2-GTP. AUG recognition provokes the hydrolysis of eIF2-GTP into eIF2-GDP, the inactivated form of the ternary complex. eIF2-GDP is recycled by the guanine nucleotide exchange factor eIF2B to form eIF2-GTP *de novo* and allow the next round of initiation.

Consequently, for optimal mRNA translation within a cell, eIF4F and the ternary complexes need to be available. This is controlled by several mechanisms.

1.6.3   Translational control

Translational control is mainly dependent on the availability of eIF4F and the ternary complexes, which are regulated by signalling pathways.

1.6.3.1   *eIF4F complex availability*

eIF4F availability is fundamental to initiate cap dependent translation. This is dictated by the accessibility of eIF4E, which is the least abundant initiation factor in the cell.

eIF4E activity is principally controlled by the phosphorylation of eIF4E through mitogen-activated protein kinases (MAPK) pathways and phosphorylated status of eIF4E binding proteins (4EBPs) which are downstream targets of the mammalian/mechanistic target of rapamycin (mTOR) (Gao and Roux, 2015).
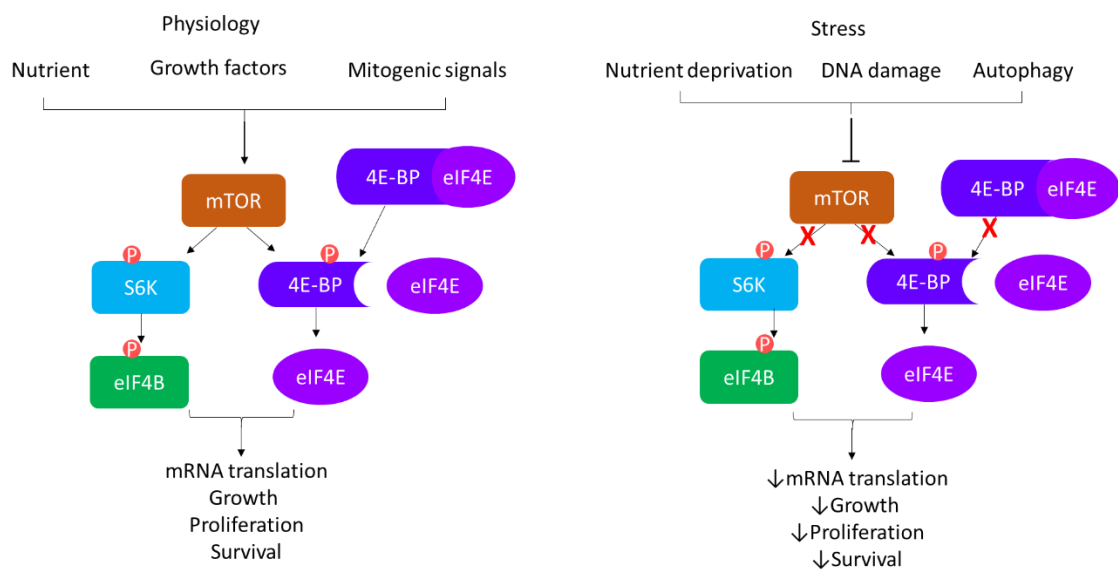
Figure 1.23. Simplified regulation of mTOR signalling pathway under physiological and stress conditions.
Upon physiological stimulus, mTOR (mammalian/mechanistic target of rapamycin) is activated and phosphorylates S6K (S6 kinase) and 4EBP (eIF4E Binding Protein) which respectively leads to the phosphorylation of eIF4B and the liberation of eIF4E. This leads to global translation of mRNA, cell growth, proliferation and survival. Following a cellular stress such as nutrient deprivation, DNA damage or autophagy, mTOR pathway is inactivated and leads to the reduction of global mRNA translation as well as the decrease in cell proliferation, growth and survival.

mTOR is one of the pivotal regulators of protein synthesis and essential to control cell growth, proliferation and cell survival and its main downstream effectors are protein S6

kinase and 4EBP. When 4EBP is hypo-phosphorylated, it sequesters eIF4E thereafter limiting eIF4F complex formation by competing with eIF4G for the binding site on eIF4E (Marcotrigiano *et al.*, 1999). When 4EBP is hyper-phosphorylated by activated mTOR, it liberates eIF4E, which can subsequently stimulate the translation of capped mRNAs. Upon physiological stimuli (growth factors, nutrients availability, mitogenic signals), mTOR phosphorylates S6 kinase and 4EBP leading to phosphorylation of eIF4B by S6 kinase and the release of eIF4E from 4EBP to promote mRNA translation, cell growth, proliferation and cell survival. Under stress condition such as genomic or hypoxic stress, nutrient deprivation or autophagy, mTOR is inhibited leading to inactivation of S6 kinase and the sequestration of eIF4E by hypo-phosphorylated 4EBP, preventing eIF4F complex formation. This results in the reduction for mRNA translation, proliferation, cell growth and survival (Figure 1.23) (Gao and Roux, 2015).
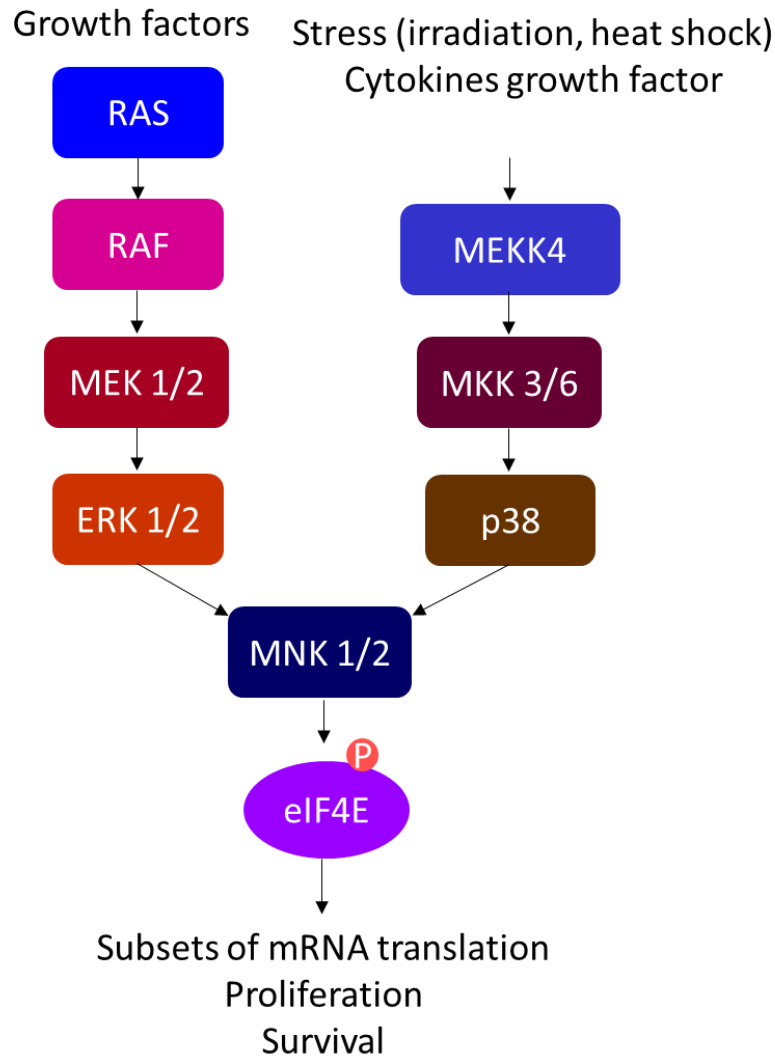
Figure 1.24. MAPK pathways involved in translational machinery.
Upon growth factors stimuli, Ras pathway activation leads to Raf-Mek1/2-ERK1/2-MNK1/2 activation which subsequently results in eIF4E phosphorylation. Under stress conditions such as irradiation or heat shock, Ras activates the MEKK4/MKK3-6/p38 pathway which can also subsequently phosphorylates eIF4E. These two pathways activation leads to translation of subsets of mRNA, proliferation and survival.

MAPKs are stimulated upon mitogenic and stress stimuli and regulate many physiological functions such as differentiation, survival or normal cell proliferation. MAPKs comprise the RSKs and the MAPK-interacting kinases 1 and 2 (MNK1 and MNK2). MNKs can be activated via MEKK4/MKK3-6/p38 or Ras/Raf/ERK MAPK pathways upon growth factors and stress stimuli. MNK1 and MNK2 interact with eIF4G and directly phosphorylate eIF4E at serine 209 which can increase its oncogenic potential (Figure 1.24) (Gao and Roux, 2015; Roberts and Der, 2007). eIF4E phosphorylation has been

associated with cancer development in prostate cancer and melanoma (Carter *et al.*, 2016; Furic *et al.*, 2010).

1.6.3.2   *Ternary complex availability*

Ternary complex availability depends on the phosphorylation of the α subunit of eIF2 (eIF2α). Under normal circumstance, eIF2α is not phosphorylated and can properly function as provider of initiator methionyl-tRNA to the ribosome.
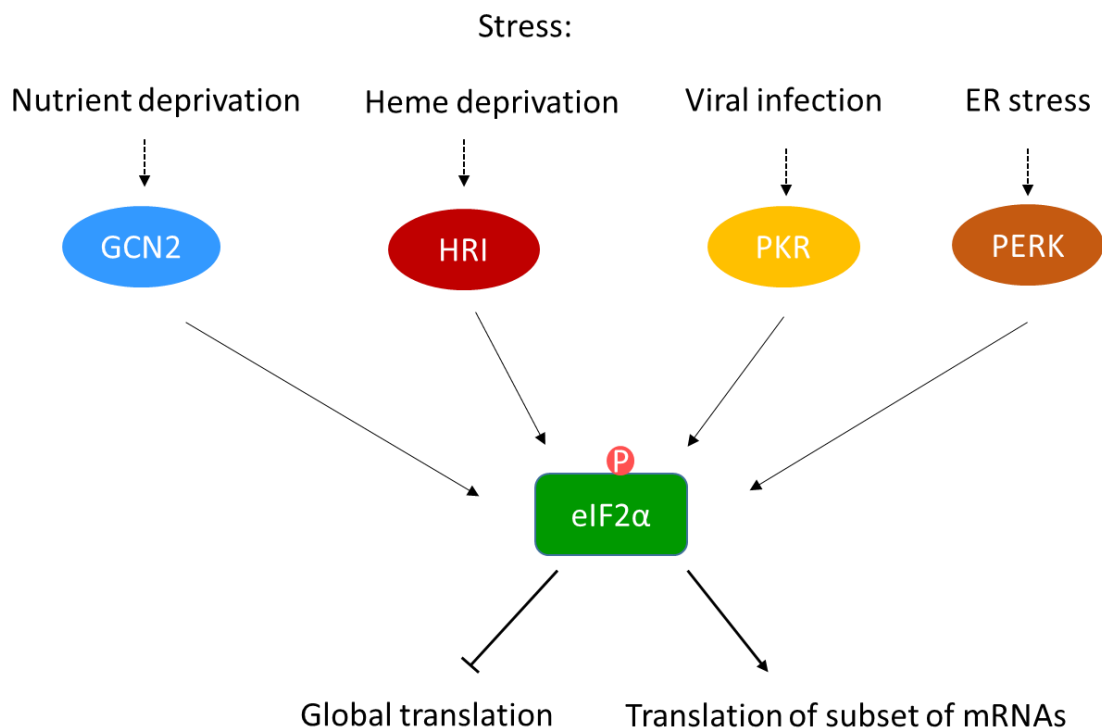
Figure 1.25. Simplified schematic cartoon of phospho-eIF2α activation pathway.
Under stress conditions such as nutrient deprivation, heme deprivation, viral infection or ER (Endoplasmic Reticulum) stress, four different kinases namely GCN2(,General control Non derepressible 2) HRI (Heme-Regulated Inhibitor), PKR (Protein Kinase RNA-activated) or PERK (Protein Kinase RNA-activated Endoplasmic Reticulum Kinase) can be activated depending on the stimulus. This leads to the phosphorylation of the α subunit of eIF2 and subsequently reduces global translation and favours translation of subset of mRNAs.

Upon cellular stress such as nutrient or heme deprivation, viral infection or endoplasmic reticulum (ER) stress, eIF2 is phosphorylated on serine 51 of its α subunit by GCN2, HRI, PKR or PERK kinases respectively depending on the stress stimulus. This phosphorylation prevents eIF2B from exchanging the GDP to GTP which leads to a decrease in ternary complex assembly thereby reducing the translation initiation as well

as upregulating subset of mRNAs essential to respond to the stress occurred (Figure 1.25) (Baird and Wek, 2012).

Given its great influence on protein levels and high energy cost, it is unsurprising that translation is highly regulated and there is emerging evidence that cancer cells require enhanced levels of global translation for their continued growth and survival, as well as specifically upregulating subsets of pro-oncogenic mRNAs. The dysregulation of translation is therefore an emerging hallmark of cancer (Bhat *et al.*, 2015; Silvera *et al.*, 2010). Alterations in eIF4F formation via deregulation of mTOR or MAPK signalling pathways or the ternary complex formation through eIF2α phosphorylation play a critical part in cancer pathogenesis (Silvera *et al.*, 2010).

1.7   Translation dysregulation in cancer

Accumulating evidence supports the substantial implication of translation dysregulation in cancer. Increased levels of multiple translation initiation factors, particularly those involved in the eIF4F complex, and dysregulation in the ternary complex formation are strongly correlated to the development of tumours. It is becoming accepted that translation dysregulation is an emerging hallmark of cancer.

1.7.1   eIF4E/4EBP

Deregulation in mTOR and MNK signalling pathways can favour tumour formation. Increased unbound status level of eIF4E via the phosphorylation of 4EBPs has been correlated with poor clinical outcome in multiples human malignancies, including breast and lung cancer (Coleman *et al.*, 2009; Wang *et al.*, 2009a). Overexpression of eIF4E phosphorylation by MNK pathways is associated with oncogenesis in human lung, neck and head, colon and stomach tumour tissues (Fan *et al.*, 2009; Furic *et al.*, 2010).

It has also been shown that higher levels of eIF4E, which implies higher level of functional eIF4F complex, stimulate the translation of mRNAs with highly structured 5'UTRs that encodes protein involved in cancer promotion (Koromilas *et al.*, 1992).

1.7.2 eIF4A

Dysregulation in eIF4A1 expression, a paralogue of eIF4A, has been shown to drive the malignant phenotype in breast cancer (Modelska *et al.*, 2015). The relationship between eIF4A2, another paralogue of eIF4A, and cancer is not clear. One study has shown that downregulation of eIF4A2 was correlated with poor outcome in NSCLC (Shaoyan *et al.*, 2013) while another study has pointed out the correlation of overexpressed eIF4A2 with poor prognosis in colorectal cancer (Chen *et al.*, 2018). This may suggest that eIF4A2 expression is tissue-specific and it can have an oncogenic property in some tissue.

1.7.3 eIF4B

eIF4B is known as a co-activator of eIF4A in translation mechanism and helps eIF4A to unwind the secondary structure in the 5'UTR. Its activity is regulated by the mTOR and MAPK pathways (Shahbazian *et al.*, 2010). Increased eIF4B expression has been linked to diffuse large B-cell and T-cell lymphoma development as well as breast cancer (Horvilleur *et al.*, 2014; Lin and Aplan, 2007; Modelska *et al.*, 2015).

1.7.4 eIF4G

Overexpression of the scaffold protein eIF4G paralogue eIF4G1 drives tumorigenesis in NSCLC and squamous cell lung cancer (Comtesse *et al.*, 2007).

1.7.5 eIF2

As mentioned earlier, eIF2 activity depends on the phosphorylation status of its α subunit. Phosphorylation of eIF2 subunit α occurs under different physiological stress (viral infection, hypoxia, nutrient deprivation, ER stress) and subsequently leads to generalised translation inhibition, preventing cell proliferation, but upregulated translation of a subset of stress-related mRNAs, many of which contain upstream open reading frame (uORF), such as ATF4 (Baird and Wek, 2012; Young and Wek, 2016). Therefore, phospho-eIF2alpha (phospho-eIF2α) can be considered as a stress marker (Baird and Wek, 2012; Koumenis *et al.*, 2002). Consequently, I would expect that stress-activated phospho-eIF2α would have an anti-tumoural effect in oncogenesis. This was observed in some *in vitro* studies where they showed that inactivation of the PKR-

phospho-eIF2α led to tumorigenesis (Donzé *et al.*, 1995; Meurs *et al.*, 1993), suggesting its tumour suppressor property in cancer formation.

However, more recent studies have demonstrated that phopsho-eIF2α could also promote invasion (Falletta *et al.*, 2017) and tumorigenesis (Falletta *et al.*, 2017; Ye *et al.*, 2010), supporting an oncogenic property in tumour development. By this model, alteration of this component of the translational machinery favours invasion and tumorigenesis by reprogramming translation in cancer cells and favouring expression of pro-invasive gene (perhaps EMT-related genes) and pro-oncogenic genes.

1.8   Translational dysregulation and EMT

MAPK kinase and mTOR pathways can act simultaneously, integrating growth factors stimuli and favouring tumour development.
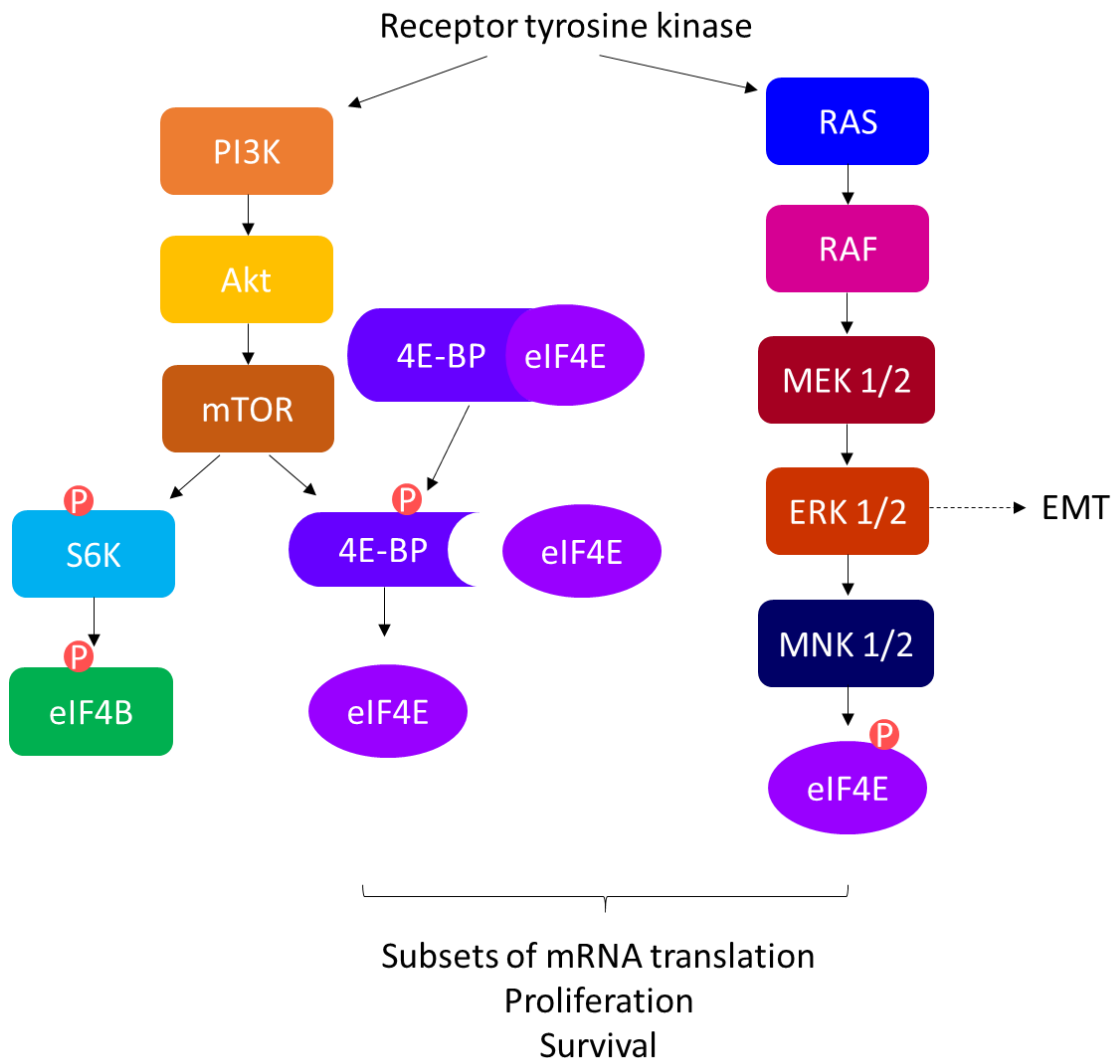
Figure 1.26. Receptor tyrosine kinase pathway.
Depending on the stimulus the receptor tyrosine kinase receives, it activates PI3K and/or Ras pathways. PI3K pathway activates Akt which subsequently activates mTOR. This latter phosphorylates S6K and 4EBP which respectively leads to the phosphorylation of eIF4B and the liberation of eIF4E. Ras pathway activation leads to Raf-Mek1/2-ERK1/2-MNK1/2 activation which subsequently results in eIF4E phosphorylation. These two pathways activation leads to translation of subsets of mRNA, proliferation and survival. Ras pathway activation can also leads to EMT activation through ERK1/2 activation.

Upon stimulation of the receptor tyrosine kinase such as EGFR by growth factor, a cascade of PI3K/AKT (upstream of mTOR pathway) and Ras/Raf/ERK pathways activation is triggered promoting translation of specific mRNAs, proliferation and cell

survival. It is well known that *KRAS* (part of the Ras family) and *EGFR* genes are mutated in lung adenocarcinoma, which can lead to sustained activation of PI3K and Ras/Raf/ERK pathway (Roberts and Der, 2007). Furthermore, Ras/Raf/ERK pathway has also been known to promote EMT (Figure 1.26) (Lamouille *et al.*, 2014). Therefore, it would be interesting to investigate whether phospho-eIF4E can facilitate translation of mRNAs implicated in EMT program and therefore activating EMT in tumour cells. Few studies have examined this using either *in vitro* or *in vivo* models. Furic *et al.* have shown the importance of eIF4E phosphorylation in promoting cancer progression in prostate cancer using a knock-in non-phosphorylable eIF4E mouse model. They have demonstrated a strong correlation between phospho-eIF4E expression and phospho-ERK and phospho-AKT (activator of mTOR) and extracellular matrix metalloprotease MMP-3, a pro-EMT marker, suggesting its potential function in EMT activation (Furic *et al.*, 2010). Consistent with this idea, Robichaud *et al.* further discovered that phospho-eIF4E could promote EMT and metastasis through the translational control of SNAIL and the MMP-3 in mammary cells (Robichaud *et al.*, 2015). In line with this notion, Attar-Schneider *et al.* have demonstrated that upregulated eIF4G1 and eIF4E favour EMT and migration in NSCLC cell lines (Attar-Schneider *et al.*, 2015).

Some studies have also highlighted the importance of eIF5A, particularly the isoform eIF5A2 in facilitating invasion and migration through EMT activation in tumour cells. eIF5A is a translation initiation factor involved in initiation, elongation and termination steps of the translation (Schuller *et al.*, 2017; Sonenberg and Hinnebusch, 2009). Studies have shown that overexpression of eIF5A2 can induce EMT by downregulating E-cadherin and upregulating vimentin expression in bladder cancer (Wei *et al.*, 2014), in gastric cancer (Meng *et al.*, 2015) or in NSCLC (Xu *et al.*, 2014). However, the function of translation dysregulation in EMT in primary lung adenocarcinoma remains a mystery.

It is believed that lung adenocarcinoma follows multistep progression from precursor lesion giving rise to *in situ* neoplasm, which subsequently can progress, into invasive disease. Although several studies have highlighted the importance of genetic aberrations in the progression from *in situ* to invasive disease (Moore *et al.*, 2019; Murphy *et al.*, 2014; Yatabe *et al.*, 2008), no one has investigated the potential contribution of translational dysregulation in promoting invasion in primary lung

adenocarcinoma. Can genetic alterations be a consequence of translation dysregulation, which subsequently promotes EMT in lung adenocarcinoma? Which other translation factors favour EMT besides phospho-eIF4E, eIF4G1 and eIF5A2? Is EMT activation a contributor of the progression from *in situ* to invasive disease in lung adenocarcinoma?

1.9   Aims of the thesis

My project aims to quantify EMT in human lung adenocarcinoma, and to investigate how translational reprogramming may affect it in primary human lung tumour tissue:

1) Quantify expression of EMT-associated genes expression in 1025 lung adenocarcinoma cases, using the LATTICE-A (Leicester Archival Thoracic Tumour Investigator Cohort – Adenocarcinoma) collection of human lung adenocarcinoma TMAs and immunohistochemistry.

2) Analyse the relationships between these markers and physio-pathological indicators (survival, growth pattern, proliferation etc.).

3) Determine the extent and microscopic localisation of EMT in whole tissue sections by performing multiplex *in situ* assays to link EMT to tumour micro-anatomy and to local variables such as stress conditions.

4) Analyse the expression of translation factors and EMT markers at a molecular level and seek evidence of mechanistic links between EMT and translation dysregulation in primary tissue.

Chapter 2.   Material and methods

Chapter 2: Materials and methods

2.1 Patient cohort selection

Tissue microarrays and associated data had previously been assembled by the Le Quesne group. Briefly, cases of lung adenocarcinoma were identified from the University Hospitals Leicester NHS Trust department of histopathology. All cases of primary pulmonary adenocarcinoma were included. Recurrences of previously diagnosed lung cancer were excluded, as well as metastatic tumours. Data on pathological staging, pleural and vascular invasion and patient survival data were collected from histopathology and pathology reports and local records respectively by the research data clerks.

All blocks of interest containing tumour material were dearchived by the research data clerks Claire Smith and Marco Sereno and further used for tissue microarrays construction.

In this study, 1025 adenocarcinoma cases were collected. Invasive mucinous adenocarcinoma, colloid and enteric adenocarcinoma cases were excluded from the analysis, which ultimately resulted in 942 cases, due to their biological difference. Tissues and data were collected under ethics agreement 14/EM/1159, and all data, images and tissues were anonymised and rendered non-identifiable to the research team.

2.2 Tissue microarrays (TMAs)

TMAs were previously constructed using donor paraffin blocks containing patient lung adenocarcinoma resections by Dr Madhumita Das from the MRC histology core facility. A TMA map was designed to permit identification of each case. Three tumour cores of 1mm of diameter each per case were sampled from the donor blocks and inserted into a blank recipient paraffin block in concordance with the TMA map. Cores were selected in order to capture variance in growth patterns. The constructed recipient block was annealed at 37°C to melt the wax and stick the core in the wax block and ready for experiments (Figure 2.1).
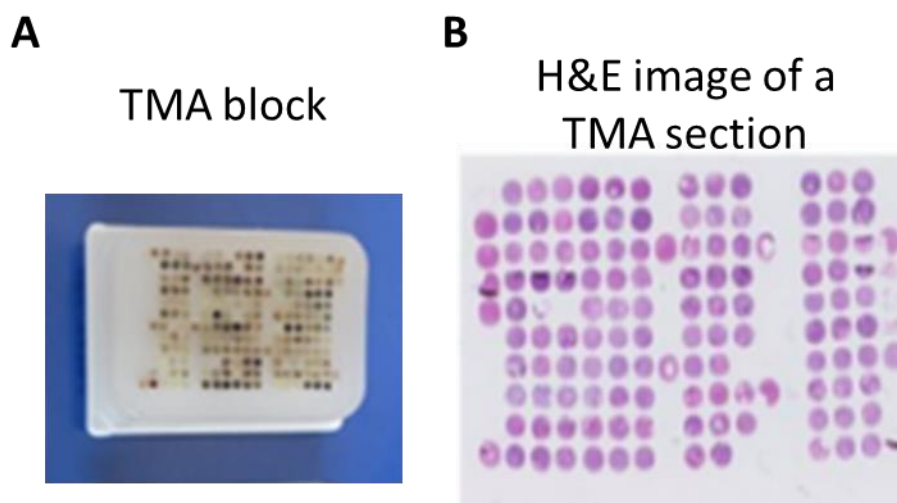
Figure 2.1. Image of TMA block and H&E stained of a TMA section.
A: Representative image of a TMA block. TMA block is made by made by sampling three tumour cores of 1mm of diameter each per tumour block from the donor blocks and inserted into a blank recipient paraffin block. B: Representative HE image of a different TMA section (different from the TMA block image showed in A). Nuclei are stained in purple (haematoxylin staining) and cytoplasm is stained in pink (eosin staining).

In this study, 5 TMAs containing 60 cases each and 18 TMAs with 40 cases each were produced with a total of 23 TMAs blocks.

2.3   Immunohistochemistry (IHC)

Immunohistochemistry (IHC) is a method used to detect antigens of interest based on antigen-antibody complex formation in biological tissue. Primary antibody binds specifically to its target antigen. The secondary antibody coupled with and enzyme, usually the horseradish peroxidase (HRP), binds to the primary antibody. HRP reacts with its substrate diaminobenzidine in presence of hydrogen peroxide, which leads to the production and deposition of brown chromogen on the tissue.

IHC was performed on 5µm thick section of TMAs blocks for 23 TMAs. Sections were processed, cut and stained with different antibodies. For the purpose of this study, antibodies for the EMT markers E-cadherin, N-cadherin, cytokeratin, vimentin and translation markers eIF4A1/2, eIF4G, eIF4E, eIF4B, phospho-eIF4E and phospho-eIF2α were used.

2.3.1 Antibody optimisation

Prior to applying staining on TMAs, antibodies were optimised on tissue biopsy or validation blocks, which are small TMA arrays containing all types of lung adenocarcinoma growth pattern. Three concentrations of antibodies: the recommended concentration from the antibody's manufacturer, a higher and lower concentration were tested with low and high pH antigen retrieval solution separately alongside a negative control without the primary antibody and a positive control known to express the target protein when necessary. Cytokeratin and vimentin antibodies are used in clinical and E-cadherin, N-cadherin, eIF4B and eIF4E antibodies have been validated through many publications. Their expression in the appropriate tissue compartment was confirmed during assay optimisation, which involved concentration ranges and antigen retrieval conditions as mentioned above.

2.3.2 Material

The table below describes the reagent used for IHC on the VENTANA® Discovery Ultra and Agilent link 48 stainers respectively.

Reagents for VENTANA Discovery Ultra autostainer / Company	Reference number for VENTANA reagent Company /	Reagents for Agilent auto-stainer	Machine	Reference number for Agilent reagent / Company
Antigen retrieval: VENTANA® Cell Conditioning 1 (CC1)	950-124 Roche Tissue Diagnostics	EnVision™ FLEX target retrieval solution low pH	Agilent PT link	K8005 Agilent
Peroxidase activity inhibitor : VENTANA® DISCOVERY inhibitor	760-4840 Roche Tissue Diagnostics	EnVision™ FLEX target retrieval solution high pH	Agilent PT link	K8004 Agilent
Blocking solution: VENTANA® DISCOVERY Goat Ig Block	760-6008 Roche Tissue Diagnostics	EnVision FLEX Peroxidase-Blocking Reagent	Agilent link 48 stainer	SM801 Agilent
VENTANA® DISCOVERY DAB Map Kit	760-124 Roche Tissue Diagnostics	EnVision FLEX /HRP	Agilent link 48 stainer	SM802 Agilent
Hematoxylin II	790-2208 Roche Tissue Diagnostics	EnVision™ FLEX DAB + Chromogen	Agilent link 48 stainer	DM827 Agilent
Bluing Reagent	760-2037 Roche Tissue Diagnostics	EnVision FLEX Hematoxylin	Agilent link 48 stainer	K8008 Agilent

Table 2.1. Table of reagents utilised for IHC on VENTANA® Discovery Ultra and Agilent autostainer.
All the reagents used on VENTANA® discovery Ultra and Agilent (PT and 48 links) autostainers are from Roche Tissue Diagnostics and Agilent Technologies respectively.

2.3.3   Protocol

Staining for vimentin was performed on Agilent Autostainer Link 48 by the histopathology facility at Leicester Royal Infirmary. Staining for eIF4E and eIF4B was performed on Agilent® Autostainer Link 48 (the deparaffinisation and retrieval step was performed in the PT link) and staining for the remaining markers was performed on VENTANA® Discovery Ultra auto-stainer by the MRC histology core facility.

2.3.3.1   Protocol on VENTANA Discovery Ultra

Reagents and antibodies used in this protocol are summarised in Table 2.1 and 2.2.

Slides were deparaffinised at 69°C for 24 minutes. They were incubated in the antigen retrieval solution, pH 6.0 or 9.0, for 32 minutes at 95°C. They were blocked for endogenous peroxidase activity in a buffer containing peroxidase inhibitor and blocked for non-specific antibody binding with goat Immunoglobulin for 12 minutes at 37°C. Slides were incubated with the primary antibody at concentration and for time indicated in Table 2.2 at 37°C. They were then incubated with the horseradish peroxidase (HRP) conjugated secondary antibody for time indicated in Table 2 at 37°C. HRP was reacted with the diaminobenzidine substrate in the presence of hydrogen peroxidase for 12 minutes at 37°C. Slides were counterstained with haematoxylin for 8 min at 37°C and bluing reagent for 8 min at 37°C. Slides were washed in 100% IMS and xylene to dehydrate the section. Finally, they were mounted with Leica CV mount medium (reference number: 14046430011) and covered using a glass coverslip (Menzel-Gläzer, reference number: 15797582).

2.3.3.2   Protocol on Agilent link 48 Autostainer

Reagents and antibodies used in this protocol are summarised in Table 2.1 and 2.2.

Slides were deparaffinised at 65°C for 20 minutes and incubated in the antigen retrieval solution, pH 6.1 or 9.0, for 20 minutes at 95°C in the Agilent PT link. They were blocked for endogenous peroxidase activity for 5 minutes and blocked for non-specific antibody binding with blocking reagent containing the goat Immunoglobulin for 10 minutes at room temperature. Slides were then incubated with the primary antibody at

concentration and for time indicated in Table 2.2 at room temperature. They were incubated with the HRP conjugated secondary antibody for time indicated in Table 2.2 at room temperature. HRP reacts with the diaminobenzidine substrate in the presence of hydrogen peroxide and deposits brown chromogen on tissue. Slides were counterstained with haematoxylin for 1 min at room temperature (Table 2.1 and 2.2). Slides were washed in 100% IMS and xylene to dehydrate the section. Finally, they were mounted with Leica CV mount medium (ref 14046430011) and covered using a glass coverslip (Menzel-Gläzer, ref: 15797582).

Primary antibody	Company/ Reference number	Source	Antigen retrieval	Primary antibody dilution (Incubation time in minutes)	Ready to use secondary antibody (Incubation time in minutes)	Company/ Reference number	Subcellular localisation
E-cadherin	Abcam ab40772	Rabbit	pH 9.0	1/1500 (20)	Omni-map anti-rabbit (12)	Roche Tissue Diagnostics 760-4311	Membranous
N-cadherin	Abcam ab76011	Rabbit	pH 9.0	1/100 (36)	Omni-map anti-rabbit (16)	Roche Tissue Diagnostics 760-4311	Membranous
Cytokeratin	Leica NCL-L-AE1/AE3	Mouse	pH 9.0	1/250 (28)	Omni-map anti-mouse (12)	Roche Tissue Diagnostics 760-4310	Cytoplasmic
Vimentin	Dako IR630	Mouse	pH 9.0	Ready to use (20)	Anti-mouse (20)	Agilent K8002	Cytoplasmic
eIF4B	Abcam Ab68474	Rabbit	pH 6.0	1/500 (20)	Anti-rabbit (20)	Agilent K8009	Cytoplasmic
eIF4E	Santa Cruz/ sc-9976	Mouse	pH 9.0	1/250 (50)	Anti-mouse (20)	Agilent K8021	Cytoplasmic

Table 2.2. Summary of antibodies used in IHC in the study.
Antibodies and their respective origin, antigen retrieval solution, dilution and incubation time and subcellular localisation are listed here. DAB: diaminobenzidine.

2.3.4   Image analysis

Stained TMA slides were scanned using the Nanozoomer XR C12000 digital slide scanner at 40X resolution. Digital slides were uploading into Path XL which is a digital pathology image analysis platform by Philips Pathology. TMAs were dearrayed by adding the corresponding TMA map for each TMA. Scoring system for intensity and subcellular compartment (intensity of cytoplasmic staining or intensity of membranous staining scores) for each protein of interest were added onto PathXL platform. Protein

expression levels were evaluated semi-quantitatively by an Allred scoring system. This scoring system is based on the intensity of the staining, scored from 0 to 3 and the percentage of tumour cells stained, measured from 0-5 (cf Appendix: Figure 8.1).

The staining intensity was scored from 0 to 3 as:

- 0 : "negative",
- 1: "weak intensity"
- 2: "moderate intensity"
- 3: "strong intensity".

The percentage of tumours cells stained was assessed as:

- 0: 0% "No stained tumour cells"
- 1: <1% "less than 1 % of tumour cells were stained"
- 2: 1-10% "1-10% of tumours cells are stained"
- 3: 11-33%: "11-33% of tumour cells were stained"
- 4: 34-66% "34-66% of tumour cells were stained"
- 5: 67-100% "67-100% of tumour cells were stained".

I used the cut-offs which are used in Allred scoring of ER expression in breast cancer (Allred *et al.*, 1998). For this study, as EMT is a focal event, I did not consider proportion scoring in my analysis.

The expression levels of proteins of interest in C1 and C2 tumours were also scored manually using H-score system. This is semi-quantitative method assessing the intensity of the staining and the percentage of positive for its respective intensity. The intenisty of the staining and percentage of positive cells are respectively scored from 0-3 and 0-100%. The H-score is obtained by the following formula:

H-score = (intensity 0 x percentage of negative tumour cells) + (intensity 1 x percentage of positive tumour cells with intensity 1) + (intenisty 2 x percentage of positive tumour cells with intensity 2) + (intenisty 3 x percentage of positive tumour cells with intensity 3). H-score values vary from 0-300.

2.4   Co-*in situ* hybridization (co-ISH)/IHC

In situ hybridization (ISH) is a technique that enables the detection of specific nucleic acid molecules on cells and tissue by using a complementary probe against its target. RNAscope® technology is an *in situ* hybridization of probe detecting single RNA molecule in cells or tissue. This technology simultaneously improves sensitivity and suppresses noise background by utilising signal amplification and unique probe design strategies (Wang *et al.*, 2012).

RNAscope® probes are designed in Z shape. The lower part of the Z is composed of 18-25 oligonucleotides that is complementary to the target mRNA. The upper region of the Z possesses a 14-base tail sequence, which corresponds to half of the binding site of the pre-amplifier sequence. These two components of the probes are linked together with a spacer sequence. The probe has to be paired, named as double Z, for the pre-amplifier to hybridize to the probe (Figure 2.2). A minimal of three paired-probes aligned next to each other and bound to target mRNA is required for signal detection (ACD, 2014).
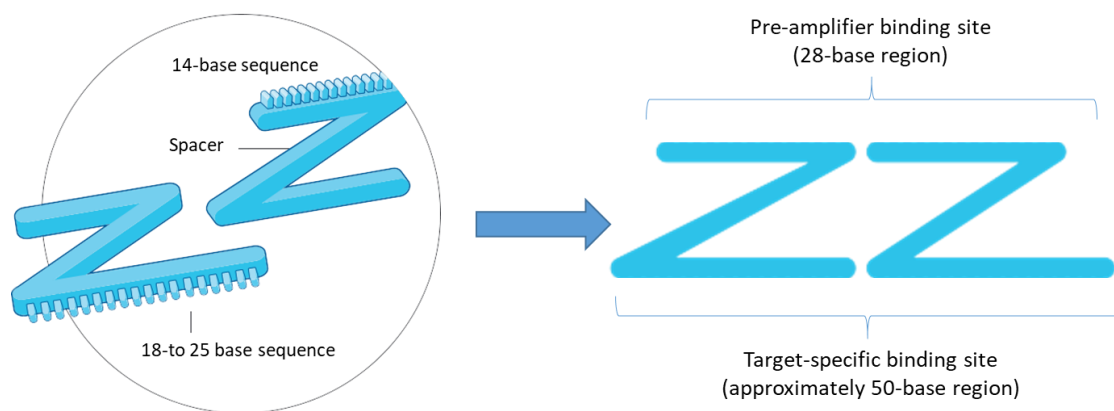
Figure 2.2. Schematic representation of Z probes.
(ACD, 2021).

For the purpose of this study, I used RNAscope® technology to assay *TWIST* mRNA expression in our TMA cohort. A negative and positive control probes for the RNAscope® were run with the assay to check the quality control. This is coupled with an immunohistochemistry assay that stains the tumour cells to permit the analysis of data in tumour cells specifically.

2.4.1   Material

Reagents utilised for the TWIST RNAscope® and the IHC for Cytokeratin are summarized in the table below.

Reagents For RNAscope®	Reference number For RNAscope® reagent / Company	Reagents for IHC	Reference number for IHC reagent / Company
Dewax and mRNA antigen retrieval	323740 Advanced Cell Diagnostics Inc (ACD)	Denaturation: VENTANA® CC2 buffer	950-124 Roche Tissue Diagnostics
Human-TWIST probe	310439 ACD	Blocking solution: VENTANA® DISCOVERY Goat Ig Block	760-6008 Roche Tissue Diagnostics
mRNA Protease and amplification kit	323210 ACD	VENTANA® DISCOVERY UltraMap anti-Rb HRP	760-4315 Roche Tissue Diagnostics
VENTANA® DISCOVERY mRNA Purple HRP Detection Kit	760-255 Roche Tissue Diagnostics	VENTANA® DISCOVERY Yellow Kit (RUO)	760-239 Roche Tissue Diagnostics

Table 2.3. List of reagents used for RNAscope® and IHC.

2.4.2   Protocol

The co-ISH/IHC starts with ISH for TWIST molecule followed by IHC for cytokeratin. Fixed specimen were dewaxed for 20 minutes at 60°C and for 4 minutes at 69°C in the dewax solution. They were pre-treated with target mRNA retrieval buffer for 32 minutes at 97°C followed by mRNA protease pre-treatment for 16 minutes at 37°C to uncover the target mRNA of interest and permeabilize cells. As depicted in Figure 2.3, double Z *TWIST* mRNA probes hybridized with target *TWIST* mRNA molecules for 2 hours at 43°C. This was followed by a sequential hybridization of pre-amplifiers to the double Z probes, amplifiers to the pre-amplifiers and HRP conjugated labelled probes to the amplifiers (Figure 2.3). This latter reacts with the purple substrate in the presence of hydrogen peroxide for 40 minutes at 37°C and stains the *TWIST* mRNA molecules in purple chromogen on tissue.

Then, the cytokeratin immunostaining was performed. Slides were denatured at 100°C for 8 minutes with CC2 buffer. Previously bound HRP-conjugates were neutralised with

peroxidase inhibitor solution at 40°C for 20 minutes. Slides were blocked with the goat Immunoglobulin for 12 minutes at 37°C. They were probed with anti-cytokeratin, diluted in 1/250, for 28 minutes at 37°C and incubated for 12 minutes with the secondary antibody coupled with alkaline phosphatase. They were then incubated with the yellow substrate that reacts with alkaline phosphatase in the presence of the yellow buffer for 32 minutes at 37°C (Table 2.3). This reaction allows the deposition of a unique and stable yellow chromogen in tissue. The yellow kit from Roche Tissue diagnostics is an alkaline phosphatase driven kit, thus the utilisation of a secondary antibody conjugated to alkaline phosphatase (Table 2.3). Slides were finally counterstained with haematoxylin and bluing solution (Table 2.1). They were washed in 100% IMS and xylene on Leica ST5020, mounted with Leica CV mount medium (Ref: 14046430011,) and covered on Leca CV5030 using a glass coverslip (Menzel-Gläzer, ref: 15797582) (Officer *et al.*, 2020).
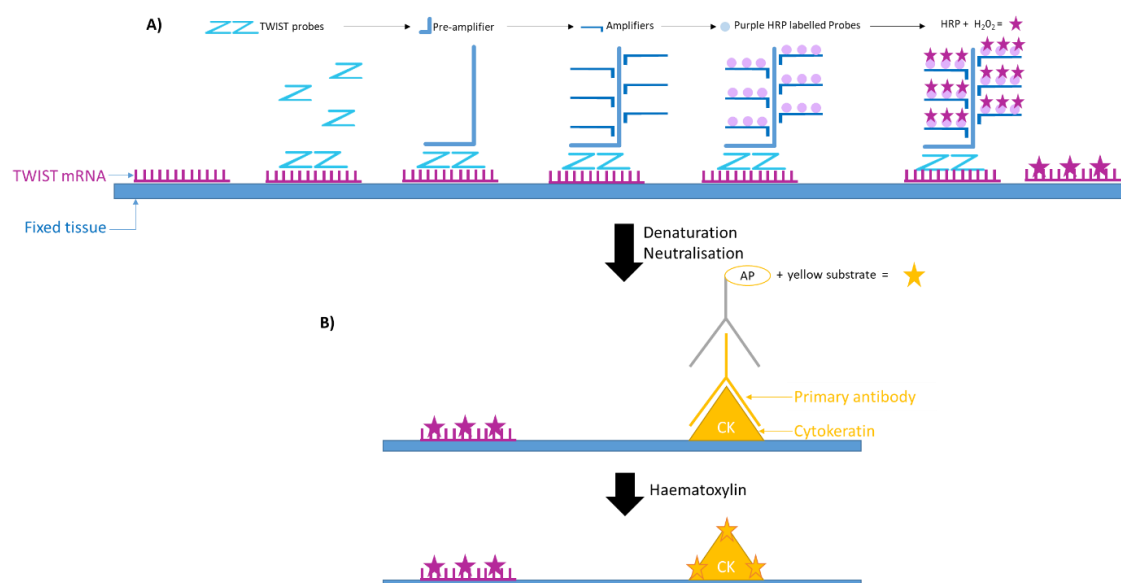
Figure 2.3. Schematic representation of *TWIST* mRNA co-ISH (RNAscope®)/IHC.
Illustration of *TWIST* RNAscope® hybridization (A) followed by cytokeratin IHC (B). In A), a probe designed against *TWIST* mRNA (light blue Z probes) hybridizes to *TWIST* mRNA on fixed tissue, followed by series of hybridizations of pre-amplifiers, amplifiers and HRP labelled probes to yield purple chromogen (purple stars) in the presence of hydrogen peroxide (H_2O_2). B: Immunohistochemistry for cytokeratin: incubation with primary antibody against cytokeratin antigen (yellow triangle) is followed by the secondary antibody coupled with alkaline phosphatase, which produces yellow chromogen that stains the tissue (yellow stars). CK: Cytokeratin, AP: Alkaline phosphatase.

2.4.3 Image analysis

Stained TMA slides were scanned at 40X using the Nanozoomer XR C12000 digital slide scanner. An automated scoring system, Visiopharm®2019.07 software, was applied to analyse *TWIST* mRNA expression. A first algorithm was created by the histopathology core facility and consisted of outlining cytokeratin positive tumour regions as a region of interest. Cytokeratin negative tumours were manually annotated. The second algorithm aimed to quantify the purple dots illustrating *TWIST* mRNA in the tumour region. Therefore, the tumour detection algorithm was run first followed by the *TWIST* mRNA detection algorithm.

To create the *TWIST* mRNA detection algorithm, several steps are completed before the actual analysis in order to extract as much information as possible (cf Appendix, Figure 8.2). I finally generate the output calculation used for the actual analysis, which is explained further down in the paragraph.
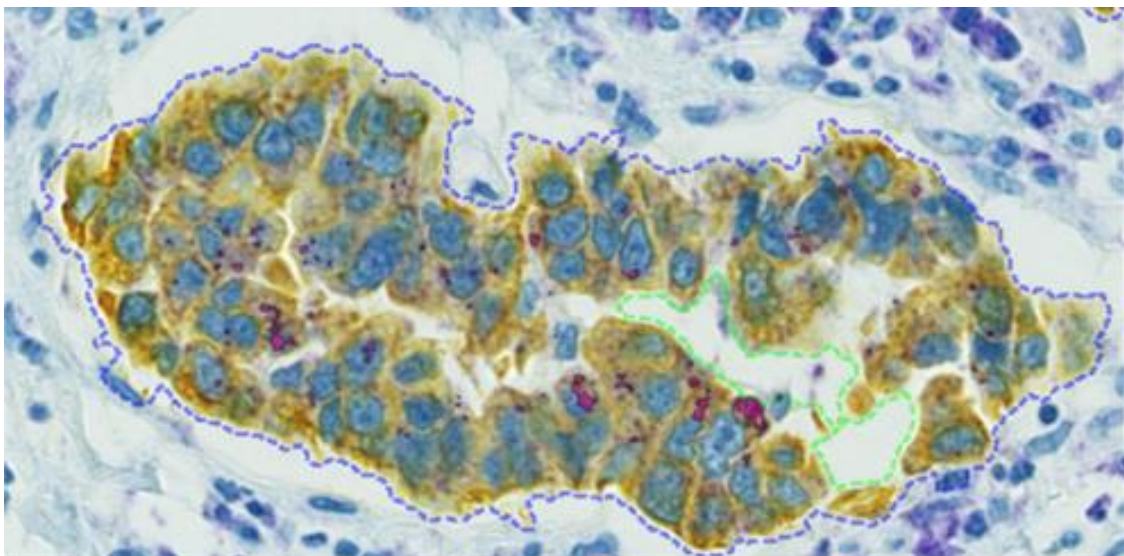
Figure 2.4. *TWIST* mRNA detection in Visiopharm®.
An image of cytokeratin positive (yellow staining) tumour region delimited by blue dots lines containing mRNA spots in purple (region of interest) and non-tumour area delimited by green dot lines. Blue staining (haematoxylin and bluing reagent) represents nuclei. Magnification = 32.49X.

In order to create the *TWIST* mRNA image analysis algorithm, I firstly selected couple of cores showing diverse tumour growth pattern and trained these images by using labels, "label 008" that detects the purple dots. Then, I determined the image magnification. A magnification of 40X was chosen to analyse the images in order to have an accurate

mRNA spot detection. The region of interest selected was ROI 002 as in the tumour detection algorithm, the cytokeratin positive tumour detection area was named ROI 002. I then proceeded to the classification step, which enables to choose the right segmentation method of images. The appropriate classification method for *TWIST* mRNA detection was the threshold method as it can provide a simpler and faster mRNA quantification. Label 008 features are then defined: I first determined the colour channel that enhances the staining of interest. The purple channel was selected as it provided a well distinguishable high and low contrast of the mRNA dots on the image. I then added filters to refine the mRNA detection. I chose the absolute and polynomial smoothing filters to take the absolute pixel intensity and to identify easily mRNA dots respectively. Then, I determined the cut-off between negative and positive staining by looking at the pixel intensity values. The higher the pixel intensity is (255 representing negative staining, i.e. white pixel), the weaker the staining (0 represented strongly positive staining, i.e. black pixel). This is executed by minimal intensity filter. This filter reposed on finding a threshold between negative and positive staining. The cut-off chosen for *TWIST* mRNA detection is 220. Any pixel value below 220 is considered as positive staining and quantify mRNA spots. I then perform a post processing step where I remove all other objects except label 008. Region of interest (ROI) is determined to analyse. I finally generate four output variables:

- New area probe: area of positive staining inside the tumour cells compartment.

- New area tumour: elimination of non-tumour area within the tumour area, i.e stroma components, emptiness and gaps between tumour cells.

-Total tumour area = new area probe + new area tumour: total tumour area without the gaps and stroma components.

- Approximate tumour spot count = total number of spots in the tumour area/ 0.75. 0.75 is the median size of a spot. This median value was obtained by measuring the size of several single dots and clusters of dots and taking the average of them.

The tumour detection followed by the mRNA detection algorithms were run on the entire set of TMAs. The obtained data and images were reviewed and reanalysed if required before exporting and further analysing in STATA SE 14.

2.5 Immunofluorescent multiplex assay

Multiplex fluorescent staining is an immunoassay that consists of detection of multiple antigens on a single tissue section. It allows visualisation of the levels of expression and spatial co-localisation of several markers at the same time on the same tissue. Briefly, this method consists of sequential detection of each antigen by application of the primary antibody followed by its respective HRP conjugated secondary antibody. A fluorescent dye conjugated to an inactive tyramide is added to the primary and secondary antibody complex. In the presence of hydrogen peroxide, the inactive tyramide becomes active and stains the tissue.

OPAL fluorescent dyes from Akoya Biosciences® were used in this study. They are tyramide driven dyes, which allow signal amplification of the target molecule (Faget and Hnasko, 2015; PerkinElmer, 2007).
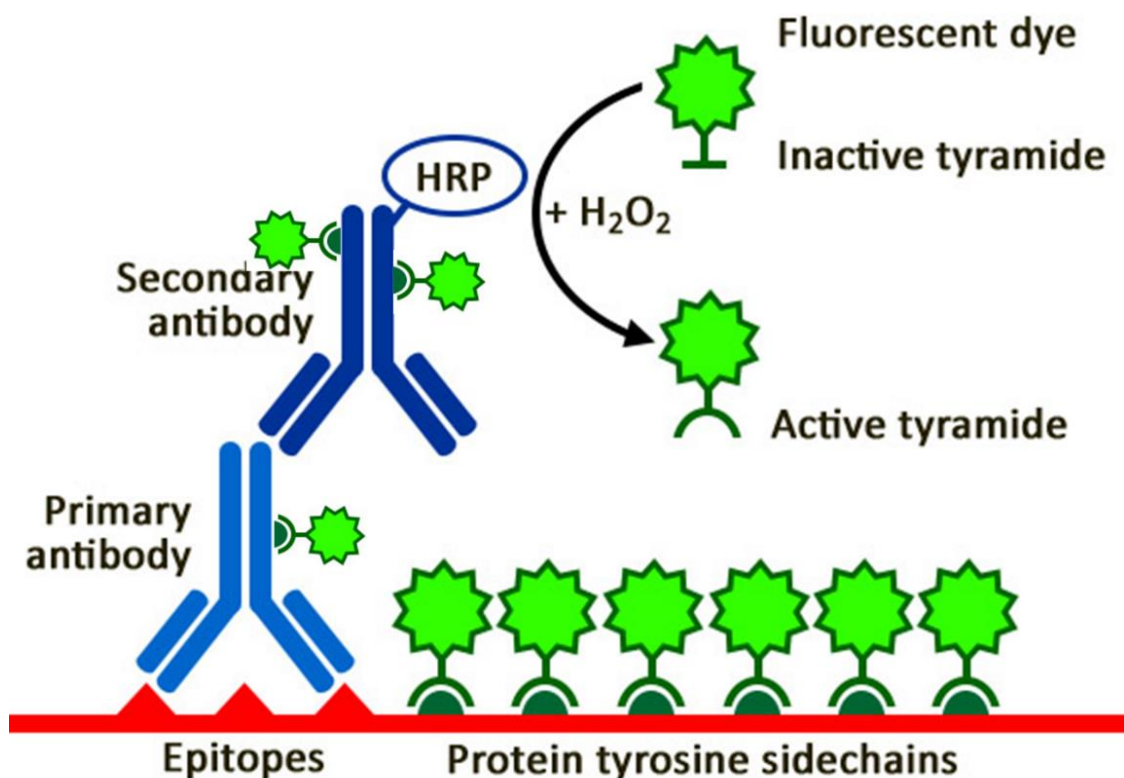
Figure 2.5. Schematic representation of tyramide signal amplification.
Primary antibody binds to the epitopes followed by HRP conjugated secondary antibody. In the presence of H_2O_2, the fluorescent dye coupled with an inactive tyramide reacts with the HRP. This reaction allows the formation of an active tyramide component that covalently binds on and nearby to protein tyrosine sidechains. Adapted from a Biotum's website (Biotum, 2015).

The principle of tyramide signal amplification is the following: OPAL dye is labelled with an inactive form of tyramide. The HRP catalyses the inactive tyramide to an active tyramide radical in the presence of hydrogen peroxide. The active tyramide radical form covalent bonds on and to nearby tyrosine residues resulting in the deposition of active tyramide molecules in the close vicinity of the target molecule, thus amplifying the signal (Figure 2.5) (Faget and Hnasko, 2015). Unbound tyramide form dimers that are washed away.

2.5.1 Material

Table 2.4 describes the reagents used for the fluorescent multiplex assay.

Reagents for multiplex IHC	Reference number / Company
Antigen retrieval: VENTANA® Cell Conditioning 1(CC1)	950-124 Roche Tissue Diagnostics
Denaturation: VENTANA® Cell Conditioning 2(CC2)	950-123 Roche Tissue Diagnostics
Peroxidase activity inhibitor VENTANA® DISCOVERY inhibitor	760-4840 Roche Tissue Diagnostics
Blocking solution: VENTANA®DISCOVERY Goat Ig Block	760-6008 Roche Tissue Diagnostics
DAPI	760-4196 Roche Tissue Diagnostics
OPAL® fluorophore diluent	FP1492 Akoya Biosciences®
OPAL® 520 fluorophore	FP1487 Akoya Biosciences®
OPAL® 570 fluorophore	FP1488 Akoya Biosciences®
OPAL® 650 fluorophore	FP1496 Akoya Biosciences®
OPAL® 690 fluorophore	FP1497 Akoya Biosciences®

Table 2.4. Summary of reagents used for the multiplex assay.

2.5.2 Multiplex optimisation

Several validation steps occurred before building the optimal multiplex assay. Each protein of interest was initially chromogenically validated with DAB standard protocol. The DAB monoplex slides served as reference for staining accuracy throughout the validation procedure. Then, a monoplex fluorescent assay was performed for each marker using OPAL® fluorescent dyes. A different OPAL fluorophore was chosen for each protein of interest. The fluorophore was selected in a way to avoid spectral cross-talk depending on the co-localisation of the proteins expected to be expressed in the same cellular compartment. The concentration of each OPAL® fluorophore was validated using the multispectral image exposure time on Vectra imaging system. The optimal exposure time is between 50 and 200 ms. Below an exposure time of 50ms, the

fluorophore concentration was considered to high and above 200ms it was considered too low. Therefore, the fluorophore's concentration was adjusted accordingly. After the monoplex fluorescent optimisation, the multiplex assay was developed by testing each marker with their respective fluorophore in every available protocol sequence to verify for loss of signal. Once the optimal position with best performance of every marker was determined, snapshots were taken to built the spectral library in inForm®. The multiplex assay was validated against the monoplex DAB and fluorescent slides and applied in samples (Officer L et al., 2019).

In this study, the slides were stained with cytokeratin, N-cadherin, phospho-eIF2α and eIF4B.

2.5.3 Protocol

Staining was performed on VENTANA® Discovery Ultra. Slides were deparaffinised at 69°C for 24 minutes. They were incubated in the antigen retrieval solution with CC1 buffer for 32 minutes at 95°C. They were blocked with the inhibitor of peroxidase for 12 minutes at 37°C and blocked afterwards with the goat Immunoglobulin for 12 minutes at 37°C. Slides were then incubated with the primary antibody and the corresponding secondary antibody one by one at 37°C in the order illustrated in Figure 2.6.
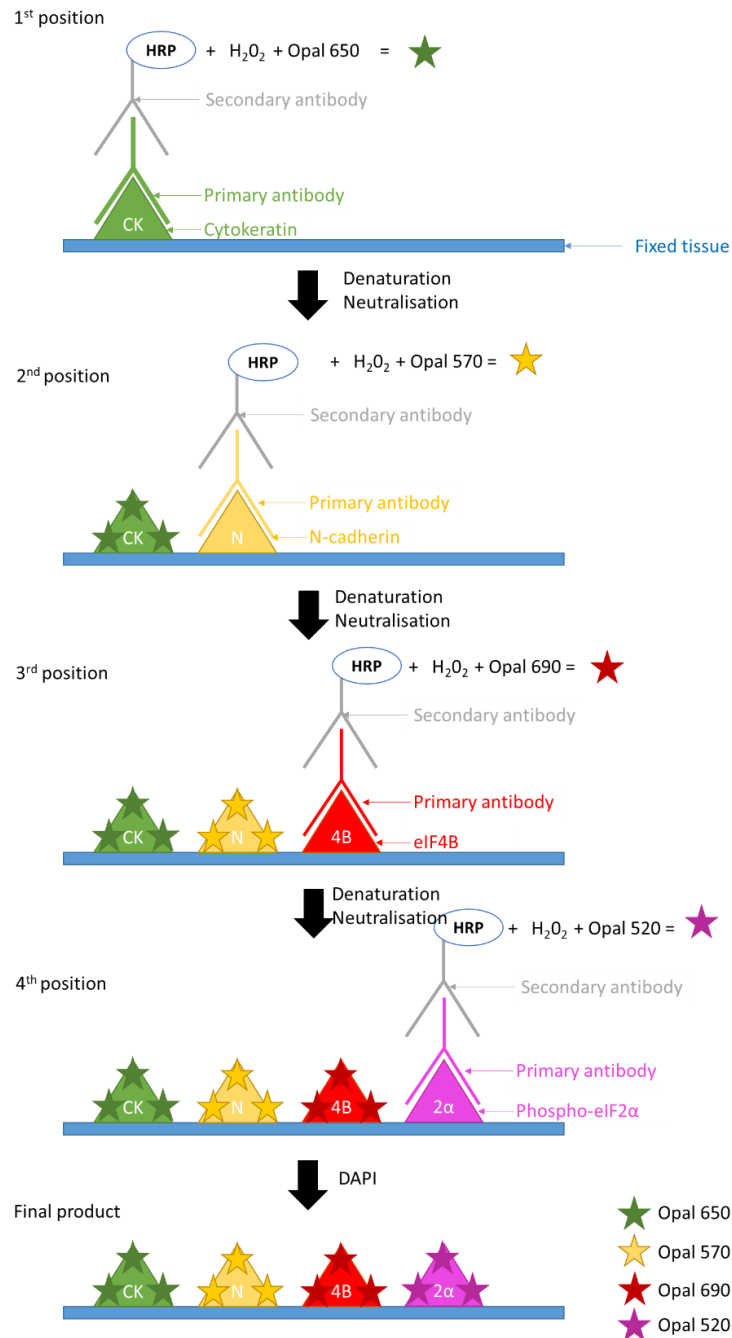
Figure 2.6. Schematic representation of the chronological order of staining in the multiplex assay. Illustration of the multiplex staining assay. The chosen fluorophore colours are shown as they are represented in the fluorescent image (refer to Figure 2.7B). They do not reflect the real fluorophore emission colour. Primary antibody's incubation begins with the incubation of the primary antibody against cytokeratin antigen (green triangle) followed by incubation of secondary antibody coupled with HRP and OPAL® 650 fluorophore, which in presence of hydrogen peroxide (H_2O_2) produces green fluorescent signal (green stars) on the tissue. Subsequently, incubation of primary antibody against N-cadherin antigen (yellow triangle) is continued by the incubation of HRP-secondary antibody and OPAL® 570 fluorophore, which yields yellow fluorescent staining (yellow stars) in presence of H_2O_2. Then, incubation of primary antibody against eIF4B antigen (red triangle) is ensued by the incubation of HRP-secondary antibody and OPAL® 690 fluorophore which in presence of H_2O_2 generates red fluorescent staining (red stars). Next, incubation of primary antibody against phospho-eIF2α antigen (magenta triangle) is followed by incubation of HRP-secondary antibody OPAL® 520 fluorophore which engenders magenta fluorescent staining (magenta stars) in presence of H_2O_2. Finally, I obtain the product with all the markers staining on the tissue. CK: cytokeratin, N: N-cadherin, 4B: eIF4B, 2α: phospho-eIF2α.

98

Primary antibody against cytokeratin antigen was probed at first on tissue at 37°C for a period as indicated in Table 2.5. This is followed by the anti-mouse secondary antibody incubation, conjugated with horseradish peroxidase (HRP), and the fluorophore Opal® 650 incubation at 37°C for a time as indicated in Table 2.5. HRP reacts with Opal® 650 in the presence of hydrogen peroxide. Slides were denatured with CC2 buffer at 100°C for 8 minutes and neutralised with inhibitor of peroxidase buffer to remove previously bound HRP at 40°C for 20 minutes. Then, the anti-N-cadherin was incubated and subsequently its respective secondary antibody couple with HRP, which reacts with fluorophore Opal® 570 in presence of hydrogen peroxide at 37°C (Table 2.5). After the denaturation and neutralisation step, anti-eIF4B was probed on the tissue, pursed by its secondary antibody, which reacts with fluorophore Opal® 690 in presence of hydrogen peroxide at 37°C (Table 2.5). Slides were probed with the anti-phospho-eIF2α and its corresponding HRP driven secondary antibody that reacts with the fluorophore Opal® 520 in presence of hydrogen peroxide at 37°C (Table 2.5). Slides were finally counterstained with DAPI. They were mounted with ProLong™ Diamond Antifade Mountant medium (Ref: PT36970, ThermoFisher®) and covered using a glass coverslip.

Primary antibody position order	Primary antibody Company/ Reference number	Source	Antigen retrieval	Primary Antibody dilution (Incubation in minutes)	Ready to use secondary antibody HRP conjugated (Incubation in minutes)	Secondary antibody Company/ Reference number	Fluorophore (dilution)	Fluorophore Incubation (minutes)	Subcellular localisation
Cytokeratin	Leica NCL-L-AE1/AE3	Mouse	Tris pH 9.0	1/250 (28)	Ultra Map anti-mouse (16)	Roche Tissue Diagnostics 760-4313	Opal 650 (1:300)	8	Cytoplasmic
N-cadherin	Abcam ab76011	Rabbit	Tris pH 9.0	1/100 (36)	Omni Map anti-rabbit (16)	Roche Tissue Diagnostics 760-4311	Opal 570 (1:300)	8	Membranous
eIF4B	Abcam Ab68474	Rabbit	Tris pH 9.0	1/125 (24)	Omni Map anti-rabbit (12)	Roche Tissue Diagnostics 760-4311	Opal 690 (1:100)	8	Cytoplasmic
Phospho-eIF2α	Cell Signaling Technology #3398	Rabbit	Tris pH 9.0	1/25 (360)	Ultra Map anti-rabbit (28)	Roche Tissue Diagnostics 760-4315	Opal 520 (1:300)	8	Cytoplasmic

Table 2.5. Summary of antibodies used in the multiplex assay.

2.5.4   Imaging

Stained TMA slides were scanned using Vectra™ Automated Quantitative Pathology Imaging System (PerkinElmer®). Regions of interest were selected based on N-cadherin expression to investigate the presence of EMT and the role of translation factors in the EMT program. The images were analysed using inForm® 2.4.8 software.

In order to obtain unmixed fluorescent images, a spectral library was built from snapshots taken of optimised single stained slides (no counterstain). The chosen colour of fluorophore emission spectrum are shown as they are represented in the fluorescent image (Figure 2.7A and B). They do not reflect the real fluorophore emission spectrum.

As shown in Figure 2.7, each fluorophore's emission wavelength is within the expected channel [DAPI at 480-500nm (DAPI channel on InForm®), Opal® 520 at 540 nm (FITC channel on InForm®), Opal® 570 at 590nm (Cy3 channel on InForm®), Opal® 650 at 660nm (Texas red channel on InForm®) and Opal® 690 at 700nm (Cy5 channel on InForm®)]. Importantly, there is no overlap among the fluorophore peaks, which means there is no cross talk. Each single fluorescent image was also checked for cross talk using DAB slides.
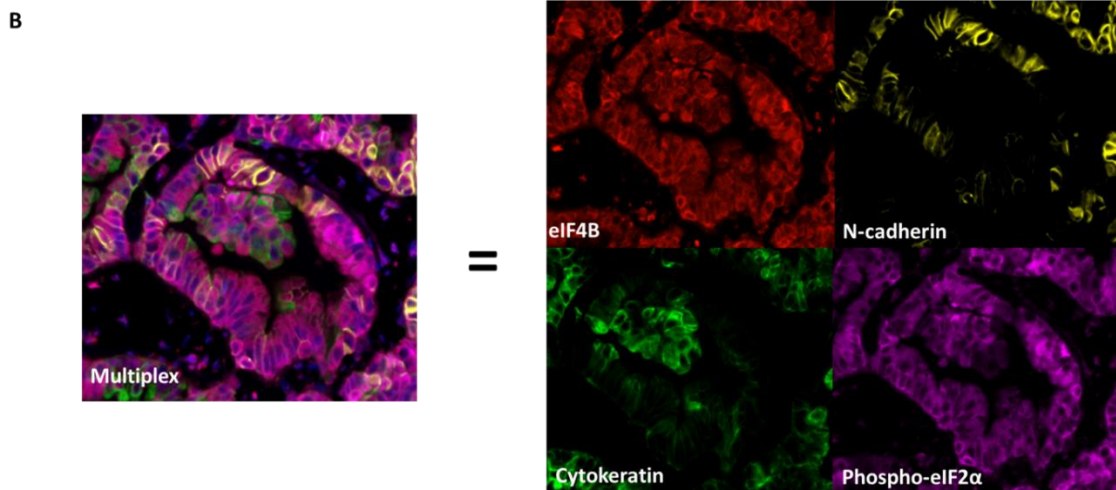
Figure 2.7. Spectral library for unmixing four channels of the multiplex.
A: Spectral library for the 4-plex assay showing the emission peaks of each fluorophore. B: Composite image of the 4-plex in primary lung adenocarcinoma. On the left is shown multiplex image with the four markers. Nuclei are stained with DAPI in blue. On the right, a deconvolution of eIF4B, N-cadherin, cytokeratin and phospho-eIF2α and fluorescent staining is shown in red, in yellow, in green and magenta respectively.

In inForm®, when I train images, I need to provide images showing a dynamic range of expression of each marker and displaying different morphologies of epithelial tumour cells (squamous, cuboid, columnar cells etc.) due to diverse subtypes of lung adenocarcinoma. An auto-fluorescent image is also loaded as lung tissue contains biological structures such as blood cells and collagen, which can emit natural light and cause problem in detecting the true staining of the proteins of interest. This image was acquired from a snapshot of a field from the auto-fluorescent control slide that shows high auto-fluorescent signal. The auto-fluorescent slide was only dewaxed and incubated with antigen retrieval solution. I finally provided the staining information into the software. I assigned artificial colours for all the markers: green for cytokeratin (Opal® 650), yellow for N-cadherin (Opal® 570), red for eIF4B (Opal® 690) and magenta for phospho-eIF2α (Opal® 520). Images are prepared to eliminate the auto-fluorescence. Once images are prepared, I obtained a composite image with all the markers. Cytokeratin, eIF4B and phospho-eIF2α are expressed in the cytoplasm while N-cadherin is expressed in the membrane (Figure 2.7).

I then progressed on to the cell segmentation, which comprises the three compartments of a cell: nucleus, cytoplasm and membrane.
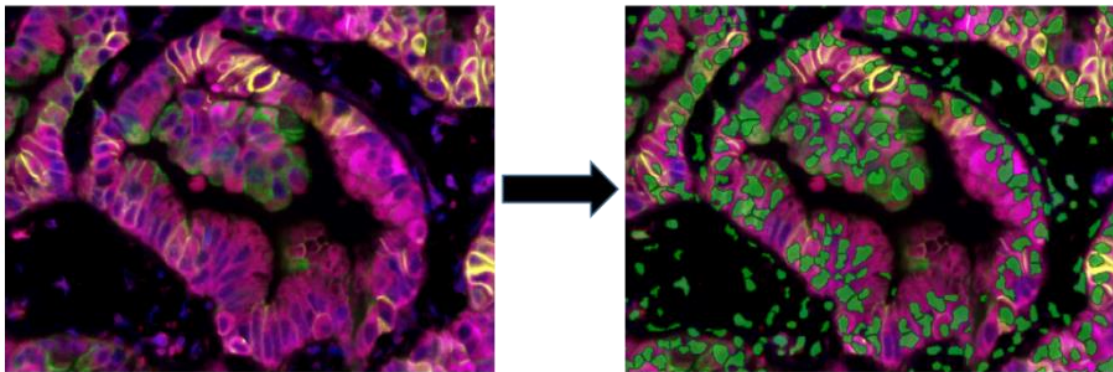
Figure 2.8. Nuclei segmentation in inForm®.
Composite image of the 4-plex in primary lung adenocarcinoma. On the left is shown multiplex image with the four markers. eIF4B, N-cadherin, cytokeratin and phospho-eIF2α and fluorescent staining is shown in red, in yellow, in green and magenta respectively. Nuclei are stained with DAPI in blue.On the right hand side, nuclei are detected in green by the software.
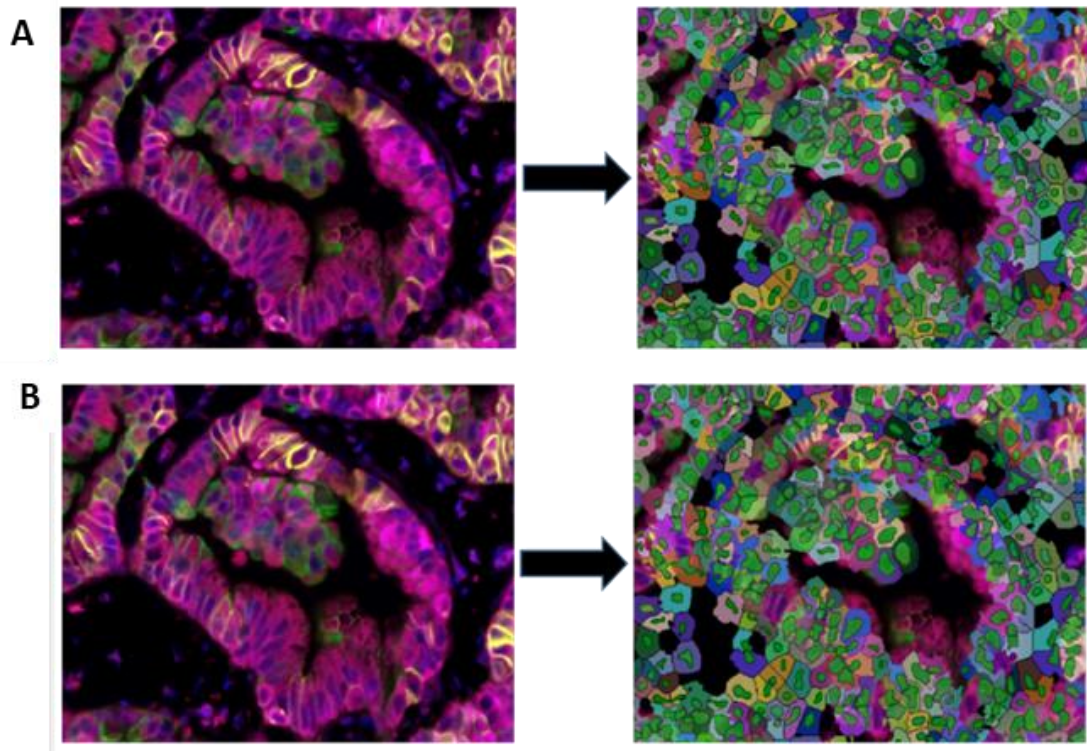
Figure 2.9. Cytoplasm and membrane segmentation in inForm®.
Composite image of the 4-plex in primary lung adenocarcinoma. On the left is shown multiplex image with the four markers and on the right are shown the detection of the cytoplasm (A) and membrane (B) by the software. eIF4B, N-cadherin, cytokeratin and phospho-eIF2α and fluorescent staining is shown in red, in yellow, in green and magenta respectively. Nuclei are stained with DAPI in blue.

I started the cell segmentation by defining the nucleus. The nucleus segmentation was based on the size, the roundness and the component DAPI (Figure 2.8, Appendix: Figure 8.3).

After the nucleus segmentation, the cytoplasm was then proceeded to segmentation. The cytoplasm region is segmented around each detected nuclei. This is built on the distance to the nucleus and the component delimiting the cytoplasm namely as cytokeratin, eIF4B and phospho-eIF2α staining (Figure 2.9A, Appendix: Figure 8.3).

N-cadherin expression was used to delimit the cell membrane. The membrane segmentation is illustrated in red on the image (Figure 2.9B, Appendix: Figure 8.3).

Once an accurate cell segmentation is obtained, I finally performed the cell phenotyping.
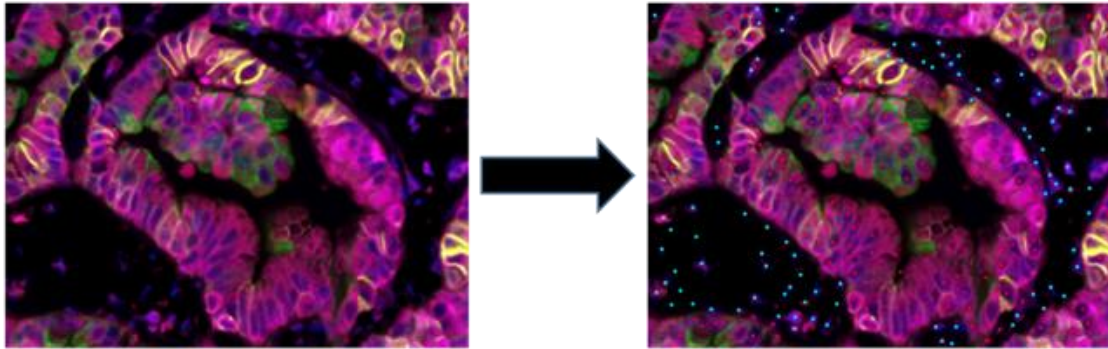
Figure 2.10. Cell phenotyping step in inForm®.
Composite image of the 4-plex in primary lung adenocarcinoma. On the left is shown multiplex image with the four markers. eIF4B, N-cadherin, cytokeratin and phospho-eIF2α and fluorescent staining is shown in red, in yellow, in green and magenta respectively. Nuclei are stained with DAPI in blue. On the right hand side, non-tumour and tumour cells are detected in blue and red respectively by the software.

Phenotyping step consists of classifying different cell populations identifiable in the tumour tissue. As I am only interested in tumour cells, two classes of phenotypes were created: tumour and non-tumour cell. A couple of tumour and no-tumour cells were identified and trained in the image. Tumour and non-tumour cells are represented in red and cyan respectively (Figure 2.10, Appendix: Figure 8.3).

Once the phenotype process was completed, I performed a batch analysis where the entire selected regions of interest was loaded, exported and performed statistical analyses of the data.

2.6   Statistical analyses

All statistical analyses were conducted using the STATA SE 14/16 software. Associations between EMT markers and translation factors and clinicopathologic features were analysed using pairwise Spearman's rank correlation. To determine the existance of significance between EMT markers/translation factors expression and tumour growth pattern, a Kruskal-Wallis tests followed by a Dunn's test with Bonferroni adjustment were performed.

Survival by subgroups was visualised by Kaplan-Meier methods, and significance of difference between groups was assessed by the log-rank test. Overall patient survival was defined between the time of surgery and the time of death from all causes. Cancer-

specific survival was defined between the time of surgery and the time of death from lung cancer cause. Recurrence-free survival was defined the time of surgery and the time of the first reappearance of the disease as judged from radiology reports and clinical data. In cases of death from lung cancer without recurrence date, the date of death was used instead.

Univariate/multivariate survival models were constructed by Cox proportional hazards regression. The proportional hazards assumption was tested by examination of log-log plots.

Comparisons between C1 and C2 tumours and between *in situ* and invasive components of C1 and C2 tumours were performed using the Mann-Whitney test and Wilcoxon signed-rank test respectively.

Chapter 3.   Adenocarcinoma growth patterns show differing degrees of

partial molecular EMT

Chapter 3 Adenocarcinoma growth patterns show differing degrees of partial molecular EMT

3.1 Introduction

One of the central hallmarks of cancer is the ability of tumour cells to metastasize by gaining invasive and migratory properties. Cancer metastasis remains the major cause of cancer related death worldwide. In lung cancer, the metastatic mechanism necessitates the detachment and propagation of cancer cells from the primary tumour site to a secondary distant site either by direct invasion, or via the blood or lymphatic circulations. These are all dependent upon invasion achieved by the epithelial-mesenchymal transtion (EMT).

Lung adenocarcinoma can be suddivided into six tumour histological patterns: lepidic, acinar, cribriform- an aggressive variant of acinar pattern with multiple lumina (Kadota *et al.*, 2014)- papillary, solid and micropapillary (Travis, 2014).

Although EMT is widely accepted as a mechanism central to tumour biology, classical morphologically dedifferentiated mesenchymal cells are rarely seen in resected primary tumours; when it is seen, this is described as sarcomatoid differentiation, and is associated with very poor outcomes, due to rapid spread and tumour aggressiveness. One key aim of this project is to resolve this apparent paradox; does EMT occur in the absence of clear morphological changes?

3.2 Aims and objectives

In this chapter, I aim to answer these following questions:

- i) Can I find molecular evidence of EMT in primary human lung adenocarcinoma?

- ii) Is the degree of molecular EMT-related to tumour growth pattern?

To achieve this aim, 942 non-mucinous lung adenocarcinomas represented in triplicate 1 mm cores on TMAs sections were immunohistochemically stained for E-cadherin, N-

cadherin, cytokeratin, and vimentin antigens, and probed for *TWIST* mRNA by *in situ* hybridisation. The proteins were manually semi-quantified according to their expression intensity while the mRNA was automatically quantified using Visiopharm® software. To investigate heterogeneity between cases and to understand the relationships between EMT markers and predominant tumour growth pattern (information given per patient from the pathology report), median intensity scores values of the three cores were considered. To explore heterogeneity of EMT markers within cores, intensity and proportion scores (i.e. percentage of stained positive tumour cell scores) of EMT markers for all three cores were considered. For correlations between EMT markers and regional tumour histological pattern, core patterns were assessed by Dr Juvenal Dario Baena-Acevedo, pathologist in our lab, and intensity scores of all three cores were used. To assess heterogeneity between cores within cases, the variance was calculated from the intensity scores of the three cores per patient for each EMT marker. Variance (i.e. variance = (standard deviation)^2) defines the average degree of difference of each point from the mean value per case.

3.3 Results

3.3.1 Lung adenocarcinoma growth pattern correlates with patient outcome

The clinico-pathological factors for 942 lung adenocarcinoma patients are listed in Table 1.

Characteristics	Total n = 942 (100.00%)
Gender	
Male	442 (46.9%)
Female	500 (53.1%)
Age (years)	
Range (Min-Max)	31-89 years
Mean	68 years
Smoking history	
Current	287 (30.5%)
Ex-Smoker	476 (50.5%)
Never	70 (7.4%)
Missing	109 (11.6%)
Performance status	
0	394 (41.8%)
1	326 (34.6%)
2	73 (7.8%)
3	11 (1.2%)
4	1 (0.1%)
Missing	137 (14.5%)
TNM staging system	
Tumour stage	
T1	283 (30.0%)
T2	443 (47.0%)
T3	125 (13.3%)
T4	27 (2.9%)
Missing	64 (6.8%)
Nodal stage	
N0	515 (54.7%)
N1	142 (15.1%)
N2	180 (19.1%)
N3	1 (0.1%)
Missing	104 (11.0%)
Overall Stage	
IA	190 (20.2%)
IB	188 (20.0%)
IIA	127 (13.5%)
IIB	87 (9.2%)
IIIA	134 (14.2%)
IIIB	14 (1.5%)
Missing	202 (21.4%)
Pleural Invasion	
PL0 – No pleural invasion	441 (46.8%)
PL1 – Pleura elastic layer invasion	298 (31.6%)
PL2 – Visceral pleural surface invasion	88 (9.3%)
PL3 – Parietal pleura invasion	76 (8.1%)
Missing	39 (4.1%)
Vascular Invasion	
Absent	499 (53.0%)
Present	435 (46.2%)
Missing	8 (0.8%)
Predominant growth pattern	
Lepidic	162 (17.2%)
Acinar	411 (43.6%)
Cribriform	31 (3.3%)
Papillary	104 (11.0%)
Solid	210 (22.3%)
Micropapillary	10 (1.1%)
Missing	14 (1.5%)

Table 3.1: Characteristics of lung adenocarcinoma patients.

Lung adenocarcinoma is a heterogeneous disease that comprises six types of histological growth pattern. Most of the tumours present several types of growth patterns, indicative of intratumoural heterogeneity in terms of growth pattern in lung adenocarcinoma. The histological growth pattern that is predominantly observed in the whole tumour determines the predominant growth pattern. TMAs cores represent regional growth patterns which may vary within the whole tumour. Cores were selected in such a way to maximise representation of the different growth patterns present in the tumour.

Amongst 942 cases of lung adenocarcinoma, 162 cases present lepidic-predominant tumour growth respectively (Figure 3.1).

411, 104 and 31 cases show predominantly acinar, papillary and cribriform tumour growth respectively in the cohort (Figure 3.1).

210 cases are of solid tumour growth, 10 cases present micropapillary growth pattern (Figure 3.1).
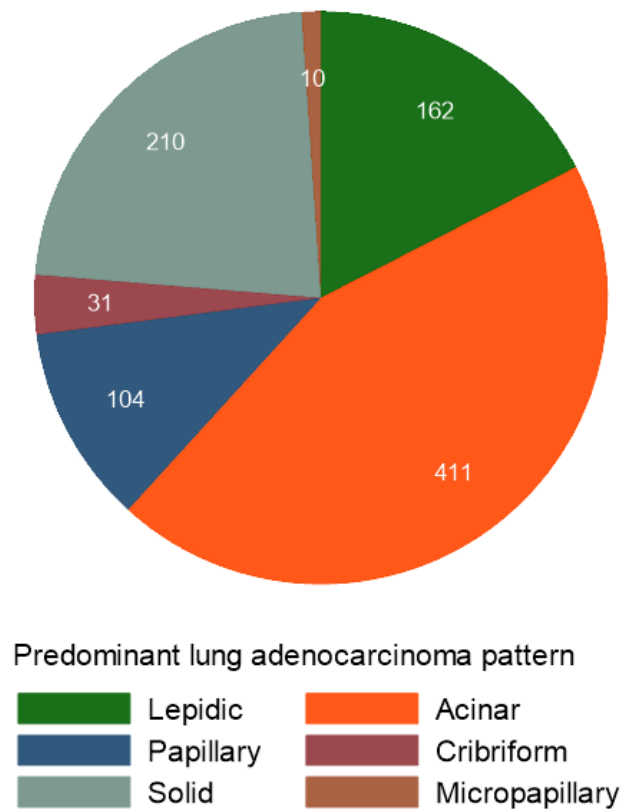
Figure 3.1. Predominant growth pattern features and distribution in our cohort.
Pie-chart representing the number of cases in each predominant tumour growth pattern, n = 928.

To assess the magnitude of the additional risk associated with tumour growth pattern and grade, univariate Cox proportional hazards regression models were created. Each sub-category of tumour growth pattern and tumour grade were compared to lepidic and low-grade respectively. The total number of patients for the tumour grade and growth pattern survival analysis was 926 lung adenocarcinomas. The total number of events for each survival analysis over a period of 5-year after surgery are indicated in the figure legends. Statistically significant differences are highlighted in bold in the Cox regression model tables.

Univariate Cox models for overall survival show that acinar, solid and micropapillary tumour growth patterns are significantly associated with poor overall patient survival compared to lepidic-predominant tumours. When comparing the overall tumour grade based on the predominant histological growth pattern, high-grade tumours are significantly linked the worst prognosis (Figure 3.2) (Table 3.1).
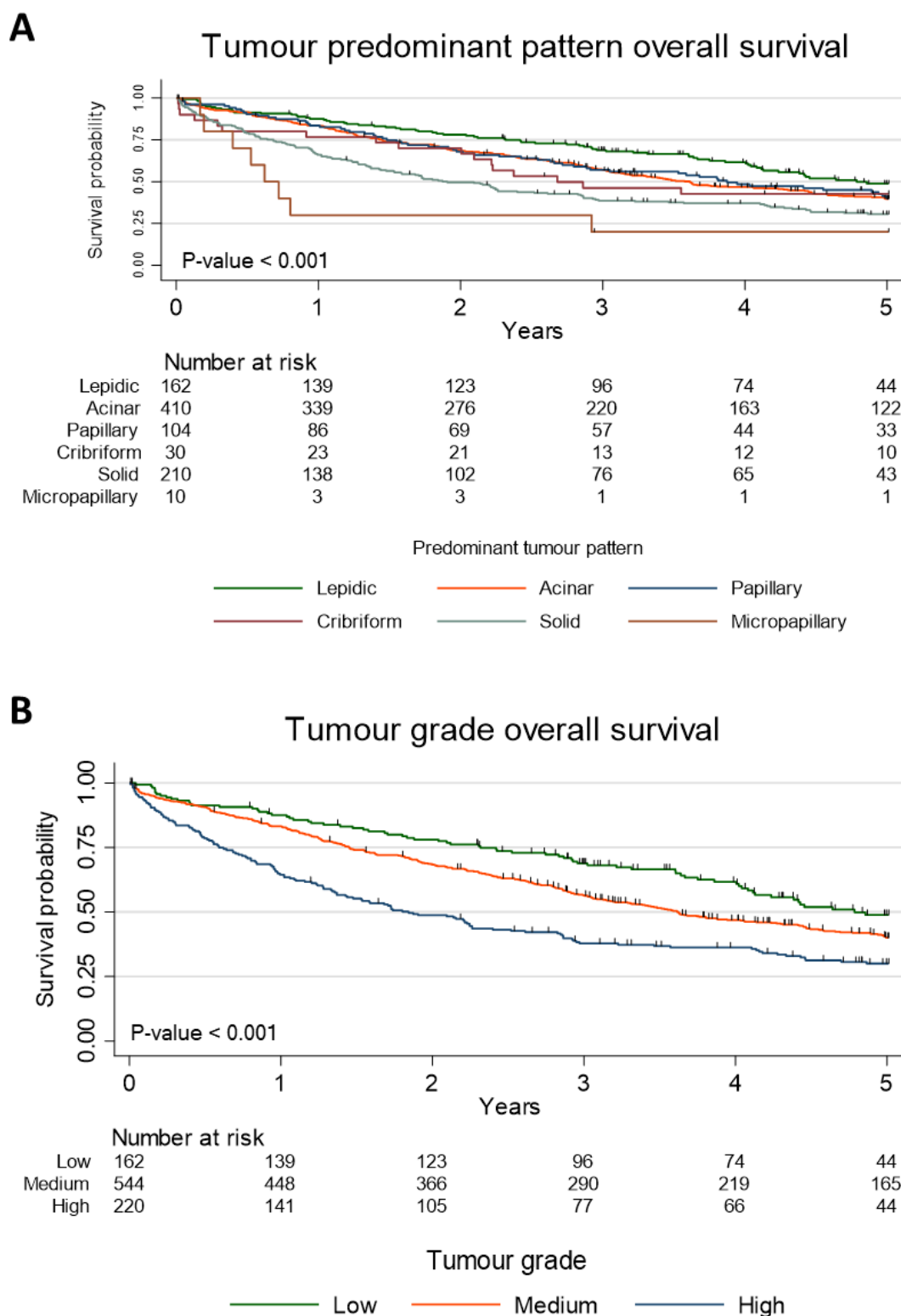
Figure 3.2. Predominant tumour growth patterns/tumour grades and overall patient survival outcome.
A and B: Kaplan-Meier graphs of tumour pattern and tumour grade over 5-year period post-surgery in lung adenocarcinoma patients. Numbers at risk are shown under the graph. P-value refers to the log-rank test, which denotes the existence of significant differences between groups.

Median survival (years)		Overall survival univariate Cox regression models			
		Variables	Hazard ratio	95% CI	P-value
Tumour growth pattern	Lepidic = 4.8				
	Acinar = 3.6	**Acinar vs lepidic**	**1.37**	1.05 - 1.79	**0.019**
	Papillary = 3.9	Papillary vs lepidic	1.33	0.94 - 1.88	0.103
	Cribriform = 2.7	Cribriform vs lepidic	1.45	0.86 - 2.47	0.166
	Solid = 1.9	**Solid vs lepidic**	**2.04**	1.53 - 2.71	**< 0.001**
	Micropapillary = 0.6	**Micropapillary vs lepidic**	**3.82**	1.84 - 7.95	**< 0.001**
Tumour grade	Low = 4.8	**Medium vs low**	**1.37**	1.06 - 1.77	**0.016**
	Medium = 3.6				
	High = 1.8	**High vs low**	**2.10**	1.58 - 2.77	**< 0.001**

Table 3.1. Correlations between predominant growth patterns/tumour grades and overall patient survival outcome.
Table summarising the median survival in years and univariate Cox proportional hazards models of tumour growth pattern and tumour grade in lung adenocarcinoma patients. Total number of events over a 5-year period after surgery = 531 deaths. CI = confidence interval.

The results for the lung cancer-specific and recurrence-free survival end-points are similar. Acinar, papillary tumour growth patterns and high-grade tumours (i.e. solid and micropapillary) are significantly related to poor prognosis in lung cancer-specific and recurrence-free survivals (Figures 3.3-3.4) (Tables 3.2-3.3).

Taking all the data together, it clearly appears that solid and micropapillary tumour growths correlate with poor overall and lung cancer-specific survival as well as an early relapse of the disease in lung adenocarcinoma patients, with micropapillary predominant tumour growth being the worst prognosis. This is in line with previous studies (Hung *et al.*, 2014b; Warth *et al.*, 2012).
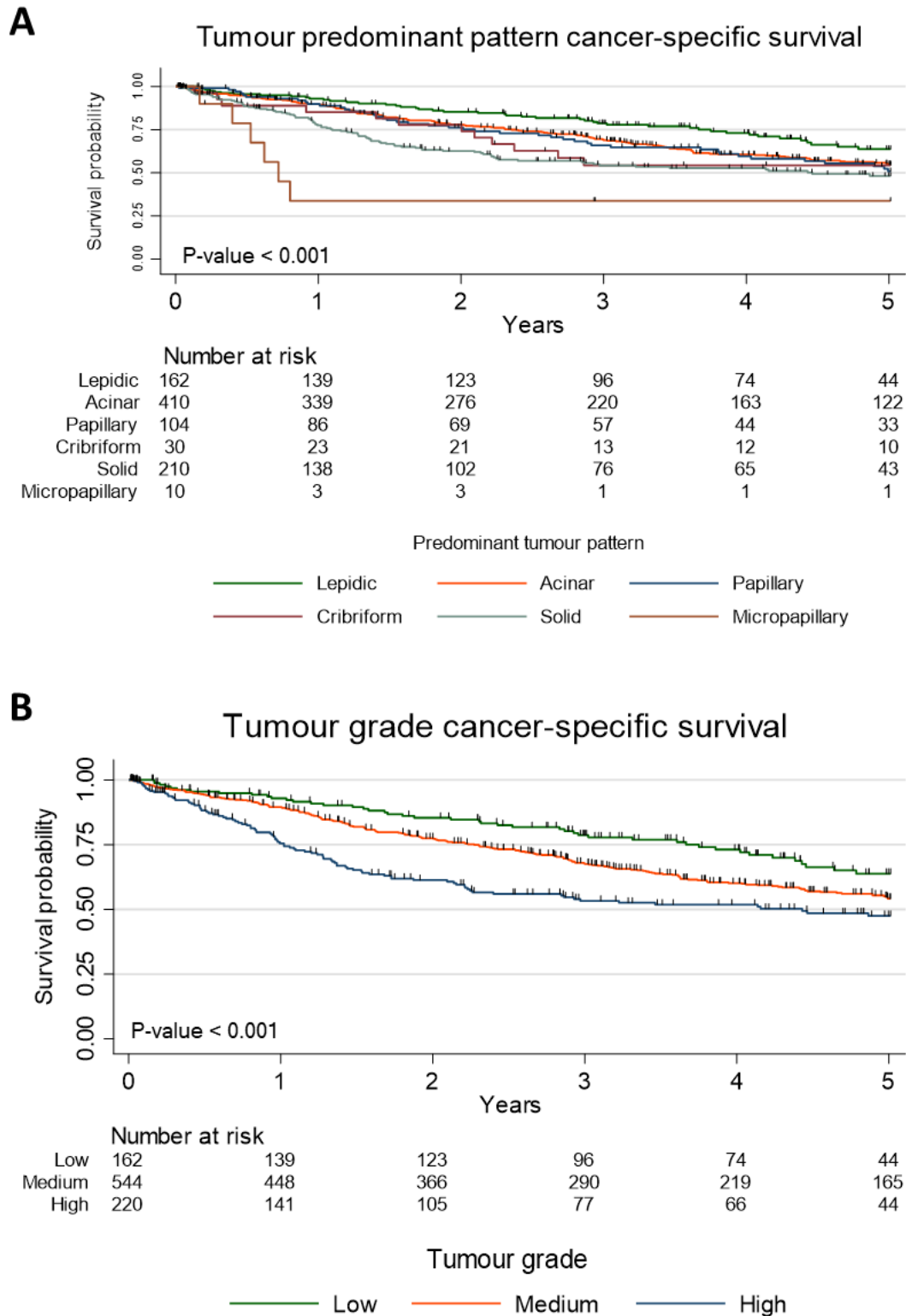
Figure 3.3. Predominant tumour growth patterns/tumour grades and cancer-specific patient survival outcome.

A and B: Kaplan-Meier graphs of tumour predominant pattern and tumour grade over a 5-year period post-surgery in lung adenocarcinoma patients. Numbers at risk are shown under the graph. P-value refers to the log-rank test, which denotes the existence of significant differences between groups.

Median survival (years)		Cancer-specific survival univariate Cox regression models			
Tumour growth pattern	Lepidic = -	Variables	Hazard ratio	95% CI	P-value
	Acinar = -	**Acinar vs lepidic**	**1.43**	1.02 - 1.99	0.035
	Papillary = -	**Papillary vs lepidic**	**1.56**	1.02 – 2.36	0.038
	Cribriform = -	Cribriform vs lepidic	1.64	0.87 – 3.10	0.127
	Solid = 4.5	**Solid vs lepidic**	**2.06**	1.43 – 2.95	**< 0.001**
	Micropapillary = 0.7	**Micropapillary vs lepidic**	**4.82**	2.05 – 11.33	**< 0.001**
Tumour grade	Low = -	**Medium vs low**	**1.46**	1.06 – 2.02	**0.020**
	Medium = -				
	High = 4.4	**High vs low**	**2.14**	1.50 – 3.05	**< 0.001**

Table 3.2. Correlations between predominant growth patterns/tumour grades and cancer-specific survival outcome.

Table summarising the median survival in years and univariate Cox proportional hazards models of tumour growth pattern and tumour grade over 5-year period after surgery in lung adenocarcinoma patients. Total number of events over a 5-year period after surgery = 346 deaths. CI = confidence interval.
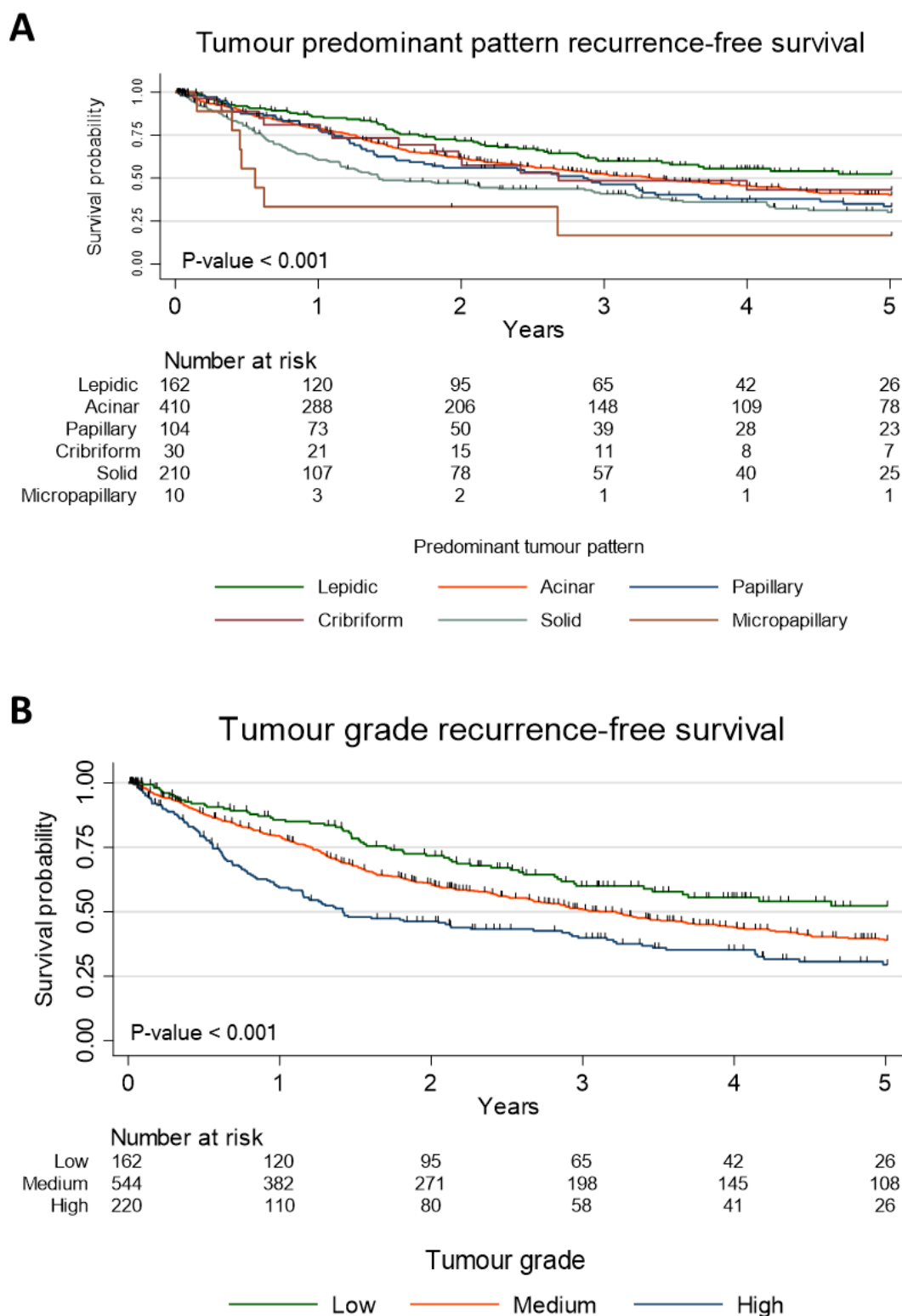
Figure 3.4. Predominant tumour growth patterns/tumour grades and recurrence-free survival outcome.
A and B: Kaplan-Meier graphs of tumour predominant pattern and tumour grade over 5-year period post-surgery in lung adenocarcinoma patients. Numbers at risk are shown under the graph. P-value refers to the log-rank test.

Median survival (years)		Recurrence-free survival univariate Cox regression models			
Tumour growth pattern	Lepidic = -	Variables	Hazard ratio	95% CI	P-value
	Acinar = 3.3	**Acinar vs lepidic**	**1.39**	1.04 - 1.86	**0.024**
	Papillary = 2.9	**Papillary vs lepidic**	**1.64**	1.15 – 2.35	**0.007**
	Cribriform = 2.7	Cribriform vs lepidic	1.36	0.76 - 2.43	0.302
	Solid = 1.4	**Solid vs lepidic**	**2.07**	1.52 - 2.83	**< 0.001**
	Micropapillary = 0.6	**Micropapillary vs lepidic**	**3.62**	1.65 - 7.93	**0.001**
Tumour grade	Low = -	**Medium vs low**	**1.44**	1.09 - 1.90	**0.011**
	Medium = 3.2				
	High = 1.4	**High vs low**	**2.13**	1.56 - 2.90	**< 0.001**

Table 3.3. Correlations between predominant growth patterns/tumour grades and recurrence-free survival outcome.
Table summarising the median survival in years and univariate Cox proportional hazards models of tumour growth pattern and tumour grade over 5-year period after surgery in lung adenocarcinoma patients. Total number of events over a 5-year period after surgery = 454 patients with recurrence of the disease. CI = confidence interval.

3.3.2   Optimisation of EMT markers by immunohistochemistry (IHC)

EMT markers were optimised in primary lung adenocarcinoma validation TMA blocks containing all types of lung adenocarcinoma histological growth pattern. All the antibodies used in this study are well established for IHC in published literature. E/N-cadherins and cytokeratin antibodies were previously optimised and validated in a DAKO autostainer and subsequently re-validated with the same concentration on VENTANA autostainer. Examples of the optimised assays on control tissues are shown below.
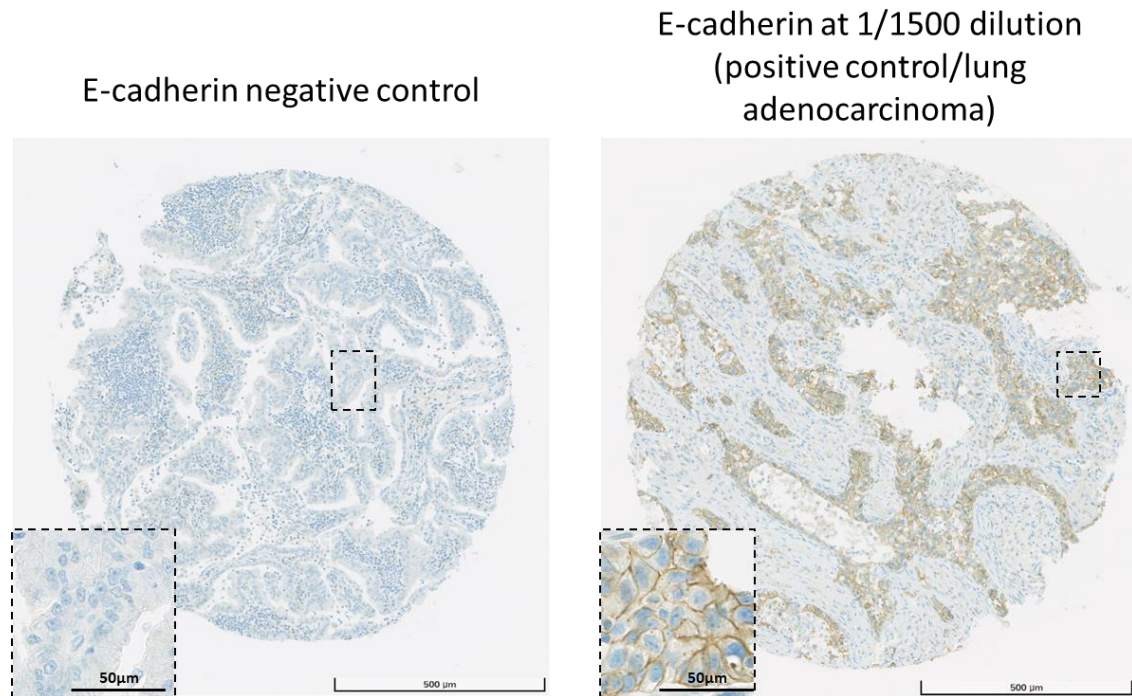
Figure 3.5. IHC validation of E-cadherin antibody.
On the left hand side, negative control with no primary E-cadherin antibody. On the right hand side, positive control and validation on lung adenocarcinoma tissue for E-cadherin at 1/1500 dilution. Blue staining (haematoxylin and bluing reagent) represents the nucleus. DAB staining is represented by brown membranous staining and represents E-cadherin expression in tumour cells. Core image: Magnification = 5x. Scale bar = 500 μm. Inset squares represent high power (40x) of a portion of the cores. Scale bar = 50 μm.

E-cadherin negative control with no primary antibody show no expression of E-cadherin as expected. E-cadherin is known to be expressed in lung adenocarcinoma according to the manufacturer datasheet. Thus, a lung adenocarcinoma validation block was used as positive control as well as the test tissue. E-cadherin staining exhibits a membranous staining as expected as E-cadherin is a transmembrane protein with membrane function (Figure 3.5).

N-cadherin negative control

N-cadherin 1/100 positive control (liver)

N-cadherin at 1/100 dilution
(lung adenocarcinoma)
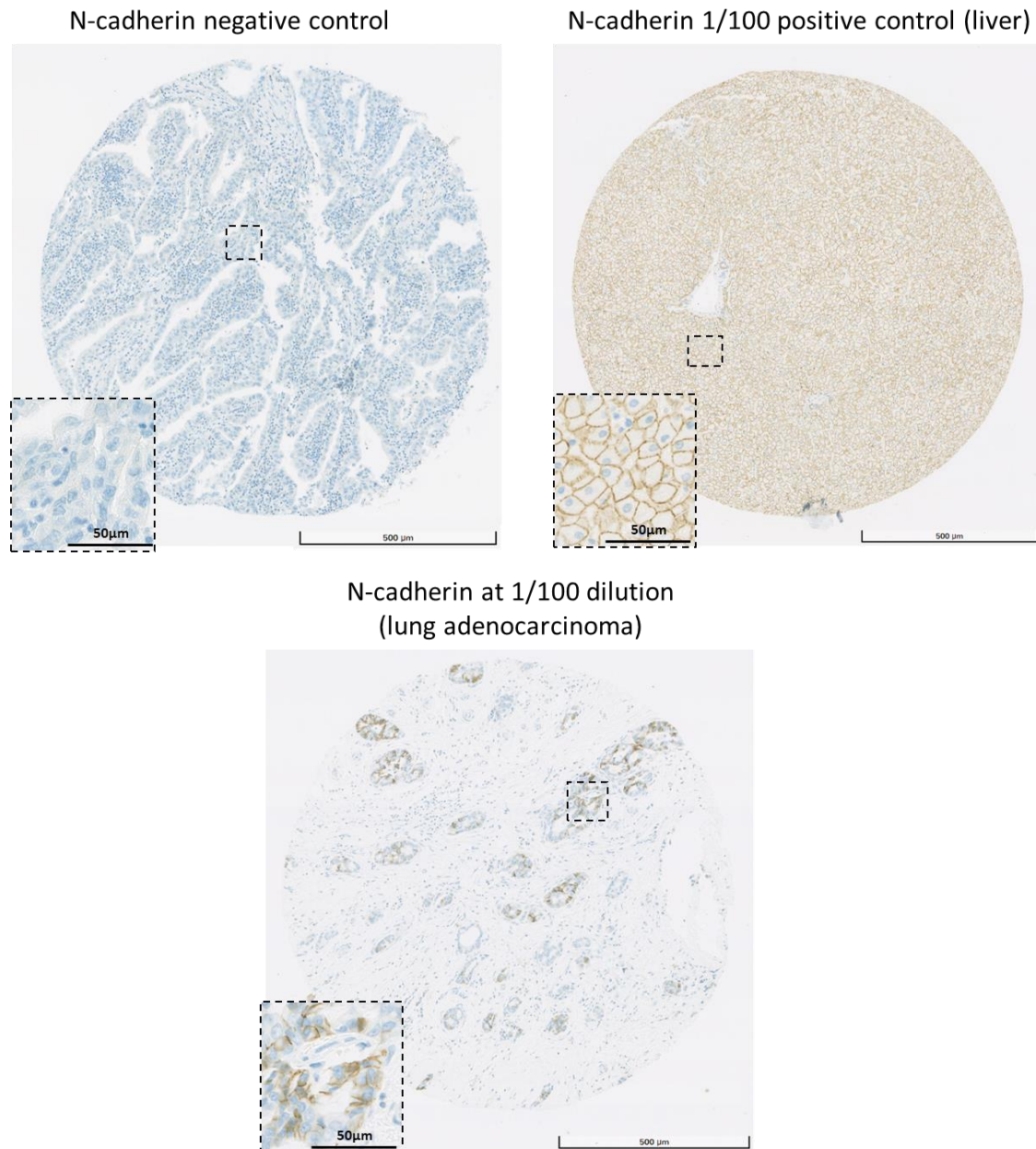
Figure 3.6. IHC validation of N-cadherin antibody.
On the top left hand side, negative control with no primary N-cadherin antibody. In the top right hand side, positive control of N-cadherin image on normal liver tissue at 1/100 dilution as liver tissue is a positive control for N-cadherin IHC according to the antibody manufacturer datasheet. Blue staining (haematoxylin and bluing reagent) represents the nucleus. DAB staining is represented by brown membranous staining and represents N-cadherin expression. On the bottom, validation of N-cadherin antibody on lung adenocarcinoma tissue. Core image: Magnification = 5x. Scale bar = 500 µm. Inset squares represent high power (40x) of a portion of the cores. Scale bar = 50 µm.

N-cadherin negative control image without primary antibody incubation does not exhibit any N-cadherin staining. N-cadherin is known to be expressed in the normal liver tissue and N-cadherin has stained the membrane of hepatocytes as expected as it is also a transmembrane protein. This assay was then validated in lung adenocarcinoma

(Figure 3.6) where it showed also a membranous staining using the same antibody concentration.
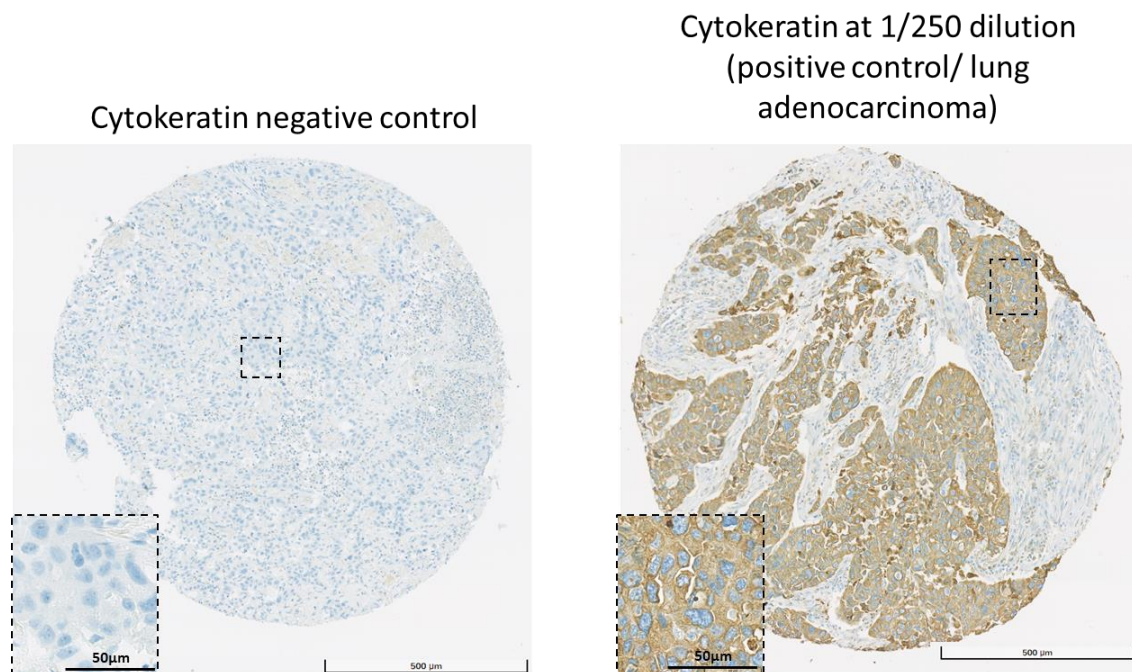
Figure 3.7. IHC validation of cytokeratin antibody.
On the left hand side, negative control with no primary cytokeratin antibody. On the right hand side, positive control and validation on lung adenocarcinoma tissue for cytokeratin at 1/250 dilution as lung adenocarcainoma tissue is a positive control for cytokeratin IHC. Blue staining (haematoxylin and bluing reagent) represents the nucleus. DAB staining is represented by brown cytoplasmic staining and represents cytokeratin expression in tumour cells. Core image: Magnification = 5x. Scale bar = 500 µm. Inset squares represent high power (40x) of a portion of the cores. Scale bar = 50 µm.

Negative control for cytokeratin with no primary antibody does not manifest any positivity. Lung adenocarcinoma tissue was used as positive control and validation test tissue for cytokeratin since it is a marker of epithelial cells. Cytokeratin is stained in the cytoplasm of the epithelial cell as it is a cytoskeletal intermediate filament (Figure 3.7).

Vimentin negative control

Vimentin (positive stroma cell control/lung adenocarcinoma)

Vimentin (positive tumour cell control/lung adenocarcinoma)
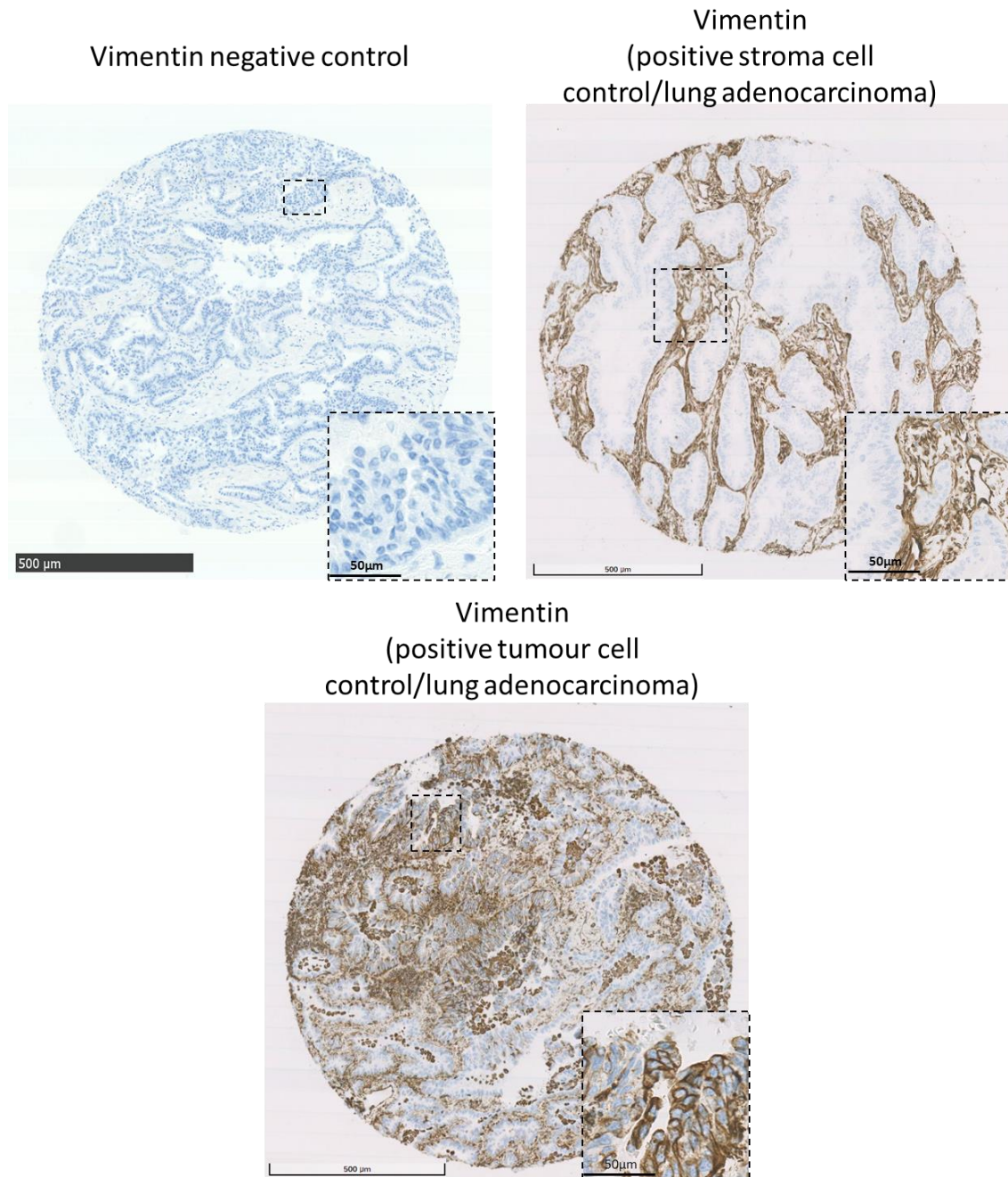
Figure 3.8. IHC validation of vimentin antibody.
On the top left hand side, negative control with no primary vimentin antibody on lung adenocarcainoma tissue. On the top right hand side, positive control of vimentin only in the stroma. On the bottom, positive control and validation on lung adenocarcinoma tissue for vimentin. Blue staining (haematoxylin and bluing reagent) represents the nucleus. DAB staining is represented by brown cytoplasmic staining and represents vimentin expression. Core image: Magnification = 5x. Scale bar =500 μm. Inset squares represent high power (40x) of a portion of the cores. Scale bar = 50 μm.

Vimentin negative control does not show vimentin staining. Lung adenocarcinoma tissue was used as a positive control. Vimentin can stain both the stromal and tumour compartments. To illustrate its stroma staining from cancer cells staining, I display two

images in which the middle image shows stain only in the stroma compartment of the tumour while the right core image illustrates staining in the cytoplasm of tumour cells (Figure 3.8).

dapB mRNA probe (negative control) · *UBC* mRNA probe (positive control)

TWIST mRNA probe (lung adenocarcinoma)
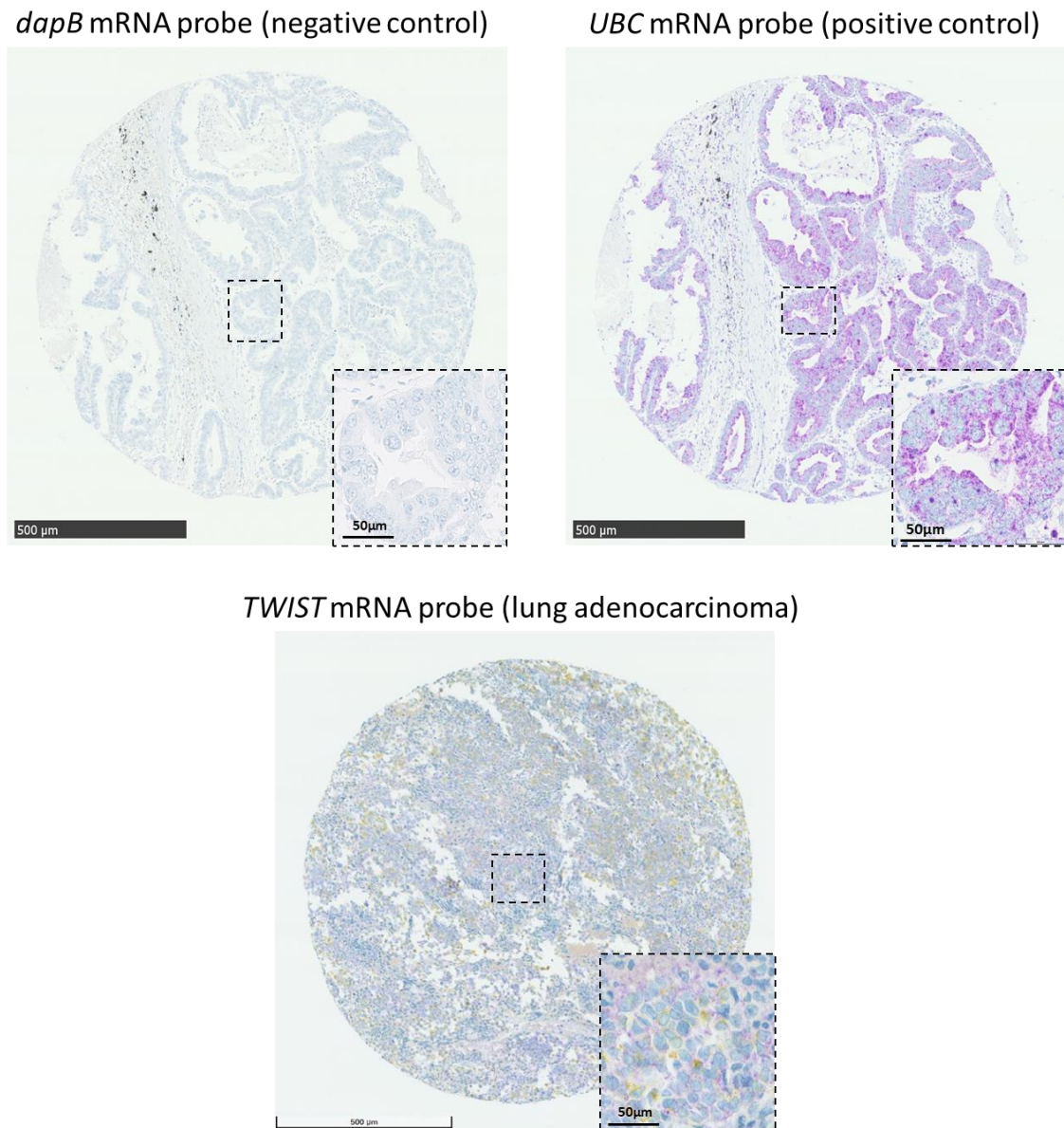
Figure 3.9. ISH validation of *TWIST* mRNA probe.
On the top left hand side, negative control of the probe with probe on lung adenocarcainoma tissue. On the right hand side, positive control of the probe with UBC probe staining on lung adenocarcinoma tissue as UBC probe is a positive control for mRNA ISH representing by purple staining according to the probe manufacturer datasheet. On the bottom, validation of *TWIST* mRNA probe (purple dots) on lung adenocarcinoma tissue. Blue staining (haematoxylin and bluing reagent) represents nuclei and yellow staining shows cytokeratin expression. Core image: Magnification = 5x. Scale bar =500 μm. Inset squares represent high power (40x) of a portion of the cores. Scale bar = 50 μm.

As the expression of epithelial and mesenchymal EMT markers are mainly regulated by the EMT-related transcription factors (EMT-TFs) SNAIL, TWIST and ZEB, I wanted to also quantify an EMT-TF at protein level but the IHC assays were non-specific. We tried to detect SNAIL, TWIST and ZEB proteins but SNAIL and TWIST were undetectable in their respective positive control while ZEB was non-specific in the positive control (data not shown). Therefore, we opted to detect one of the transcription factor, i.e. TWIST, at mRNA level using RNAscope® technology. The manufacturer provides a negative and positive control which are the *dapB* mRNA probe (dihydrodipicolinate reductase which is only expressed in bacteria) and *UBC* mRNA probe (encodes for ubiquitin C protein which is ubiquitously expressed in human cells) respectively to test alongside the test probe, namely *TWIST* mRNA probe. As expected, the negative probe does not show any staining represented by purple dots on lung adenocarcinoma tissue and the positive control probe is expressed in all the cells of lung adenocarcinoma tissue. *TWIST* mRNA probe stains the cell cytoplasm and nucleus in lung adenocarcinoma since it is an mRNA that can be in the transcription progression step in the nucleus or ready for translation or degradation in the cytoplasm (Figure 3.9).

The validated EMT markers were subsequently applied on the whole set of TMAs and quantified.

3.3.3 Characterisation of EMT in primary lung adenocarcinoma

To check whether EMT is quantifiable in primary lung tissue, EMT markers were stained on 23 TMAs, containing 2826 primary lung adenocarcinoma cores (three cores per case to capture variance in tumour growth patterns from 942 cases with several normal lung cores as controls), and scored according to their membranous or cytoplasmic expression intensities with the exception of TWIST, which was scored according to the number of *TWIST* mRNA spots in the tumour regions. Normal lung tissue cores were used as a negative control and were not scored in the study. Cores with too little tumour material (fewer than hundred malignant cells) were not scored as well. Inset squares represent a high magnification (80X) of a portion of the core.

Protein expression levels were evaluated semi-quantitatively by an Allred scoring system. This scoring system is based on the intensity of the staining, scored from 0 to 3 and the percentage of tumour cells stained, measured from 0-5 (Table 3.1).

Allred scoring system				
Intensity scores		**Percentage of tumour cell positivity scores**		
		Scores	**Meaning**	
		0	0%	No stained tumour cells
Scores	**Meaning**	1	<1%	Less than 1 % of tumour cells were stained
0	Negative	2	1-10%	1-10% of tumours cells are stained
1	Weak	3	11-33%	11-33% of tumour cells were stained
2	Moderate	4	34-66%	34-66% of tumour cells were stained
3	Strong	5	67-100%	67-100% of tumour cells were stained

Table 3.1. Allred scoring system.
Allred score system is based on the intensity scores, which goes from 0 for negative up to 3, and the percentage of stained tumour cells (positive tumour cells) which are scored from 0-5.

TWIST mRNA was scored automatically, thus this Allred scoring system was not applied but *TWIST* mRNA expression was then categorised into four groups referring to the four levels of intensity 0, 1, 2 and 3 according to the number of spot counts.

The inter-observer reliability test was performed to validate the scoring system used by the second observer (expert histopathologist with 15 years of experience in image analysis) and myself (with no experience in image analysis) using weighted kappa statistics as the markers of interest are ordinal variables. The inter-rater agreement was not assessed for eIF4B/E intensity and percentage of tumour cells stained scorings, *TWIST* mRNA scoring and N-cadherin percentage of tumour cells stained scoring. A TMA containing as much as different tumour growth pattern and as many scorable cores was chosen to validate the scoring system. There were 120 cores scored in total for the inter-observer test. I performed the kappa weighted inter-observer reliability test. Table 2.3 displays that there is a fair to excellent agreement regarding the scoring system between these two observers.

Inter-observer reliability between two observers		
Protein of interest scoring system validation	**Kappa value**	**P-value**
Cytokeratin intensity	0.80	P < 0.001
Cytokeratin positivity percentage	0.98	P < 0.001
E-cadherin intensity	0.61	P < 0.001
E-cadherin positivity percentage	0.73	P < 0.001
Vimentin intensity	0.77	P < 0.001
Vimentin positivity percentage	0.79	P < 0.001
N-cadherin intensity	0.57	P < 0.001

Table 3.4. Inter-observer reliability test.
Inter-observer reliability test is performed between two observers and provides a kappa value. A kappa value below 0.40 means that there is poor agreement regarding the scoring system between these two observers. A kappa value between 0.40 and 0.75 may be taken to represent fair to good agreement. A kappa value above 0.75 may be taken to represent an excellent agreement regarding the scoring system between these two observers.

I then analysed the distribution of each EMT marker in our cohort using the intensity and proportion score values of the three cores. This illustrates the heterogeneity of each EMT marker within and between tumour cores. Then, I analysed the frequency of each EMT marker and calculated their variance across 942 lung adenocarcinoma cases using the median scoring value of the three cores of the EMT marker to demonstrate the heterogeneity of each marker expression between cases and between cores within cases respectively. The proportion scoring was not considered in this study when analysing the distribution and expression of EMT markers as EMT is a focal event. By considering Allred scores and analysing both intensity and proportion scoring, I would lose information about EMT. Therefore I decided to only analyse the expression intensity of EMT markers.

The epithelial marker E-cadherin immunohistochemical staining shows diverse range of intensity in the cohort. E-cadherin is scored according to the membranous intensity staining. E-cadherin is weakly expressed in the attenuated type 1 epithelial cells in the normal lung. Cores can express weakly, moderately strongly or cannot express the epithelial marker (Figure 3.10).
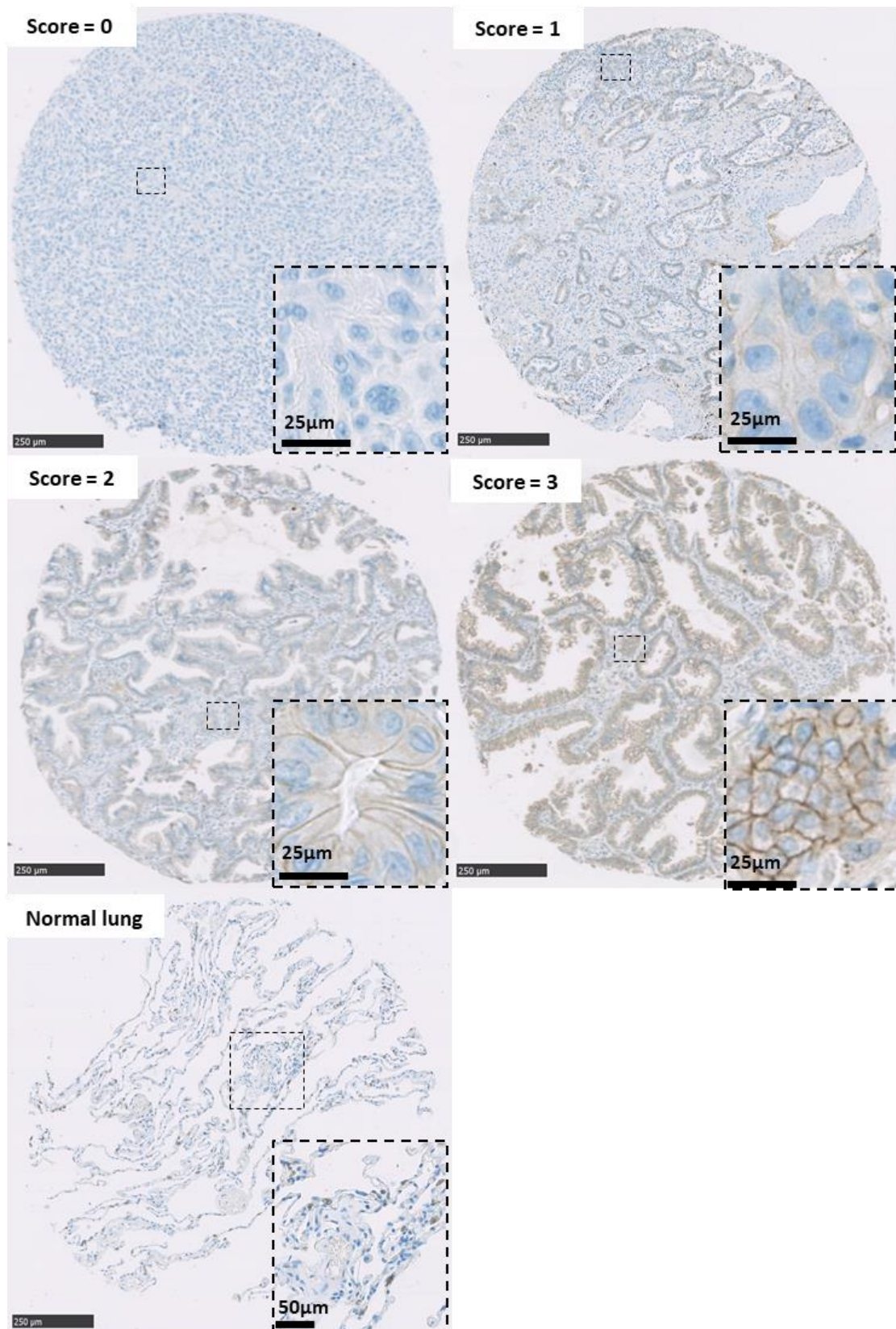
Figure 3.10. Dynamic range of E-cadherin expression.
Images of the epithelial marker E-cadherin immunostaining scored according to their expression intensities. Blue staining (haematoxylin and bluing reagent) represents the nucleus. DAB staining is represented by brown E-cadherin membranous staining. Core image magnification: 7x., scale bar = 250µm. Inset square magnification of the normal lung = 40X, scale bar 50 µm .Scale par for the normal lung Scale bar of the high magnification (80X) of a portion of the core, scale bar = 25 µm.

127

The majority of the cores express the epithelial marker quite homogenously at various intensity (Figure 3.11A). We have to keep in mind that these are small cores of 1mm of diameter and heterogeneity may not be detectable at this scale, despite being apparent across the tumour as a whole.

E-cadherin expression intensity (score 1 to score 3) is more or less equally distributed in all cases with 300 cases per score on average but we observe heterogeneity in E-cadherin expression between cases. Only about 20 cases do not express E-cadherin (Figure 3.11B). The variation of E-cadherin expression between cores within the same primary tumours is generally very low, indicating an homogenous expression of the epithelial marker between cores in most cases. Only about a quarter of patients show a variation in E-cadherin expression between cores of the same primary tumours (Figure 3.11C).

A E-cadherin expression per core

B Median E-cadherin expression per case

C Variation of E-cadherin expression between cores per patient
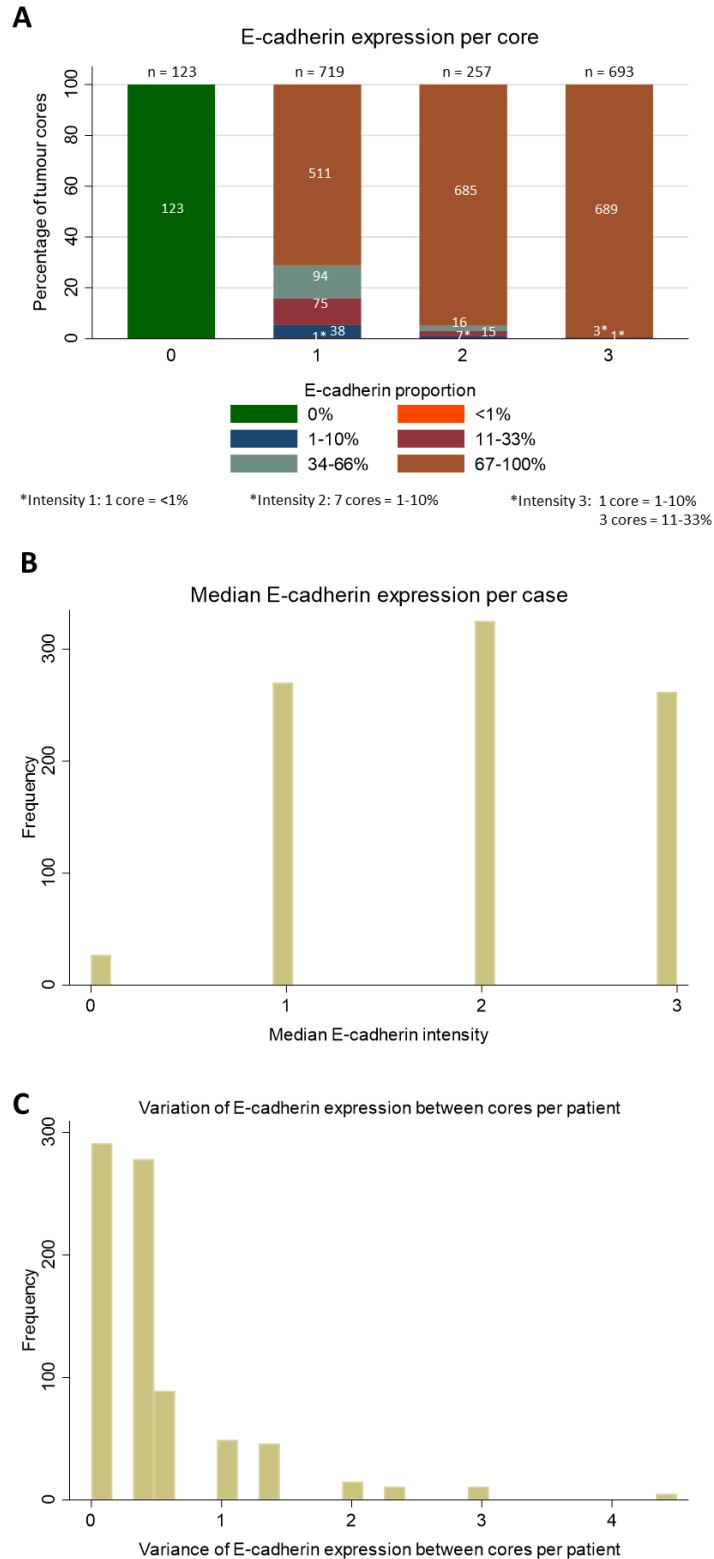
Figure 3.11. Distribution of E-cadherin intensity and proportion scores per core and frequency of E-cadherin expression per case in the cohort.
A: Percentage of tumour cores expressing E-cadherin at different proportion for each intensity in lung adenocarcinoma cores. n = number of cores B: Frequency of median E-cadherin intensity expression according to the median intensity score value in 885 lung adenocarcinoma cases. C: Frequency of the variation of E-cadherin expression between cores within cases in 795 lung adenocarcinoma patients.
Unit of measurements: 1 unit = 1 scoring value.

129

The mesenchymal marker N-cadherin is detected at different levels of expression in the cohort. N-cadherin is undetectable in normal lung. The transmembrane protein is scored according to the membranous intensity staining and can be expressed weakly, moderately or strongly (Figure 3.12).
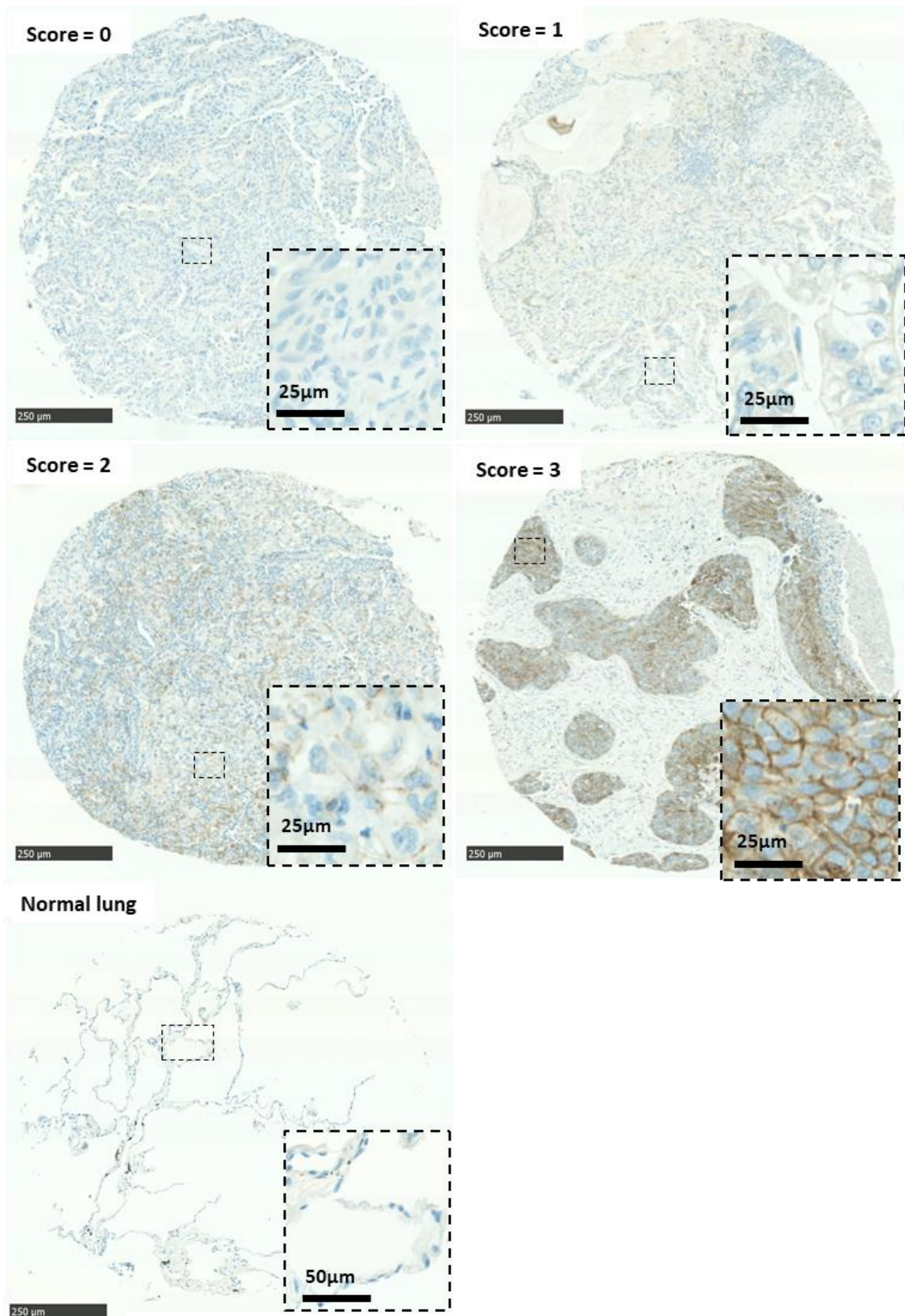
Figure 3.12. Dynamic range of N-cadherin expression.
Images of the mesenchymal marker N-cadherin immunostaining scored according to their expression intensities. Blue staining (haematoxylin and bluing reagent) represents nuclei. DAB staining is represented by brown membranous N-cadherin staining. Core image magnification = 7x, scale bar = 250 µm. Inset square magnification of the normal lung = 40X, scale bar 50 µm. Scale bar of the high magnification (40X) of a portion of the core, scale bar = 25 µm.

N-cadherin is heteregenously and focally expressed within cores when it is scored as 2 or 3, suggesting that EMT may be a focal event (Figure 3.13A).

As observed in Figure 3.14B, N-cadherin expression distribution is skewed to the right. N-cadherin is mainly absent in tumours as more than 600 primary lung adenocarcinoma cases are scored 0. About 230 cases weakly express N-cadherin. About 50 cases and 70 cases show moderate and strong expression of the mesenchymal marker respectively. This illustrates that N-cadherin is heterogenously expressed between cases (Figure 3.13B). The mesenchymal marker is mainly homogenously expressed between cores within the same primary tumours. Few cases demonstrates difference in N-cadherin expression between cores within cases (Figure 3.13C).

Thus, N-cadherin is heteregenously expressed within and between cases but homogenously present between cores within cases.
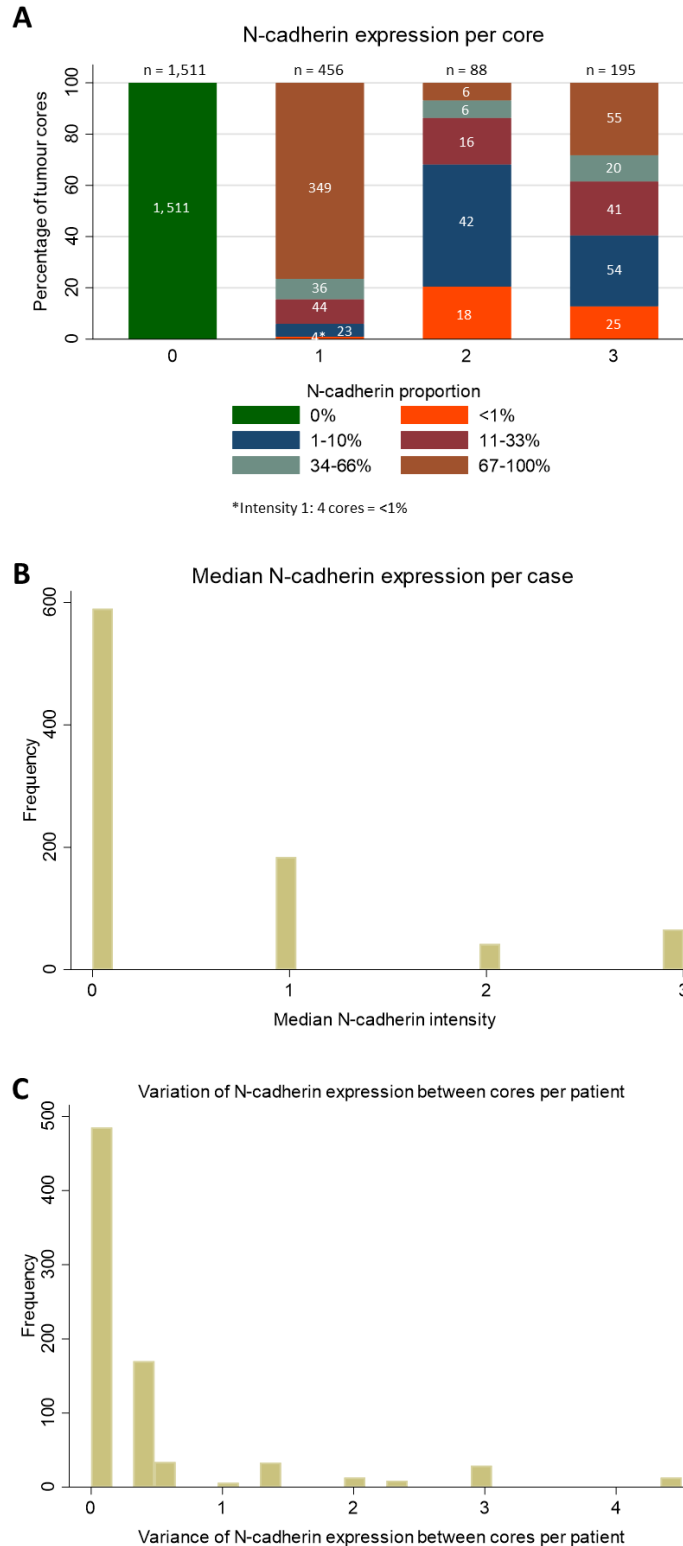
Figure 3.13. Distribution of N-cadherin intensity and proportion scores per core and frequency of N-cadherin expression per case in the cohort.
A: Percentage of tumour cores expressing N-cadherin at different proportion for each intensity in lung adenocarcinoma cores. n = number of cores. B: Frequency of median N-cadherin intensity expression according to the median intensity score in 883 lung adenocarcinoma cases. C: Frequency of the variation of N-cadherin expression between cores within cases in 792 lung adenocarcinoma patients.
Unit of measurements: 1 unit = 1 scoring value.

The immunohistochemical staining of the epithelial cytoskeletal marker cytokeratin presents expression at different levels of intensity in 942 cases. Cytokeratin is expressed in normal lung epithelial cells (Figure 3.14).
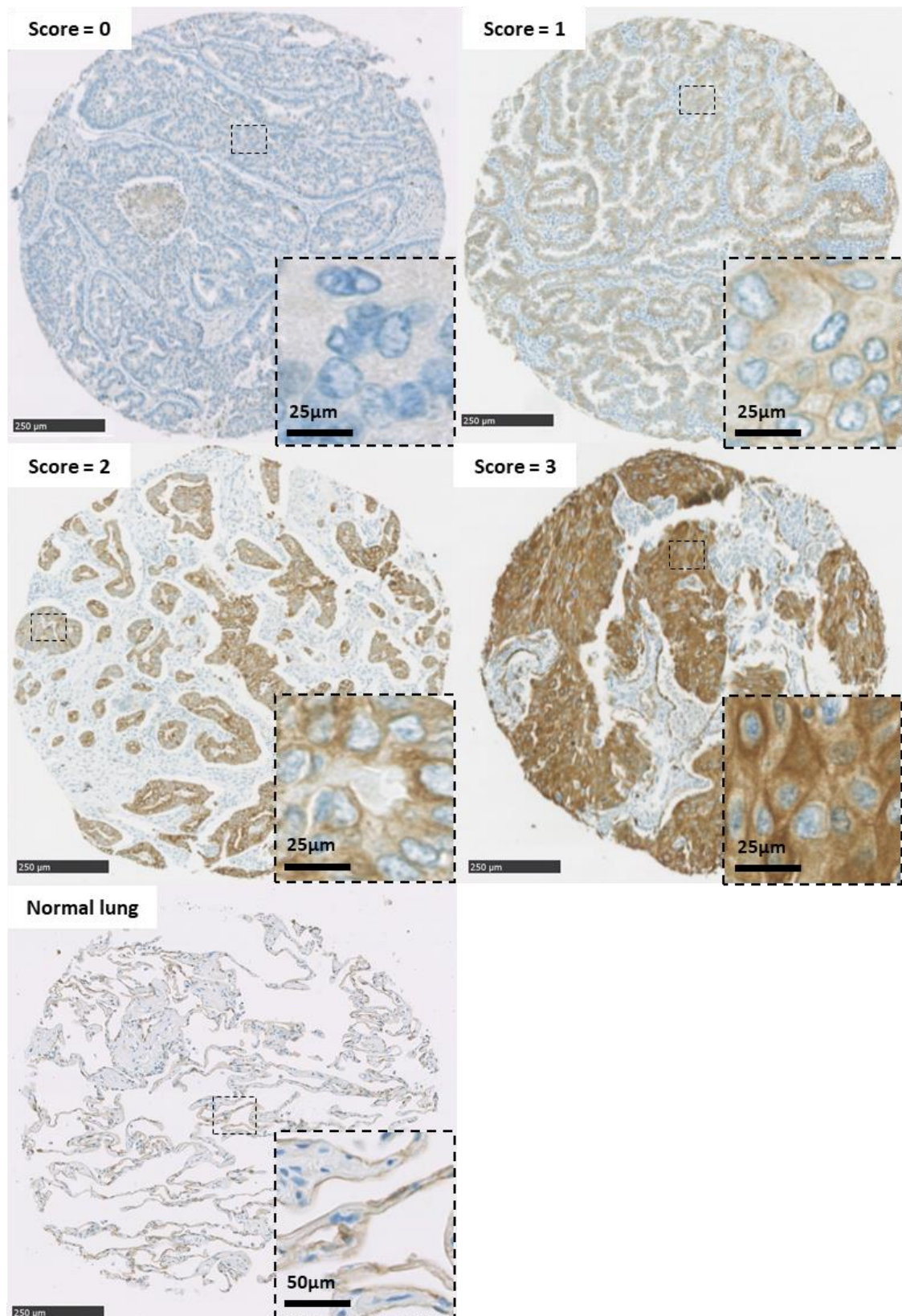
Figure 3.14. Dynamic range of cytokeratin expression.
Images of the epithelial marker cytokeratin immunostaining scored according to their expression intensities. Blue staining (haematoxylin and bluing reagent) represents nuclei. DAB staining is represented by brown cytokeratin cytoplasmic staining. Core image magnification = 7X, scale bar = 250µm. Inset square magnification of the normal lung = 40X, scale bar 50 µm. Scale bar of the high magnification (80x) of a portion of the core = 25 µm.

As shown in Figure 3.15A, cytokeratin is either homogenously expressed or not within cores.

Similarily to E-cadherin expression, the vast majority of primary lung adcenocarcinoma cases express cytokeratin at varying intensities. Only about 10 cases are completely cytokeratin negative. Nearly 300 cases exhibit weak expression of cytokeratin, about 220 cases and 350 cases express moderately and strongly the expithelial marker (Figure 3.15B). Cytokeratin expression does not differ from the mean intensity values in most cases. Few cases show a variation in the epithelial marker expression between cores per primary tumour (Figure 3.15C).

Hence, cytokeratin is mainly homogenously detected within tumour cores and between cores within cases but heterogenously expressed between cases.

A

Cytokeratin expression per core

*Intensity 1: 3 cores = <1%
9 cores = 11-33%
22 cores = 34-66%

*Intensity 2: 3 cores = 1-10%
4 cores= 11-33%

*Intensity 3: 1 core = <1%
3 cores = 1-10%
2 cores = 11-33%
1 core = 34-66%

B

Median cytokeratin expression per case

C

Variation of cytokeratin expression between cores per patient
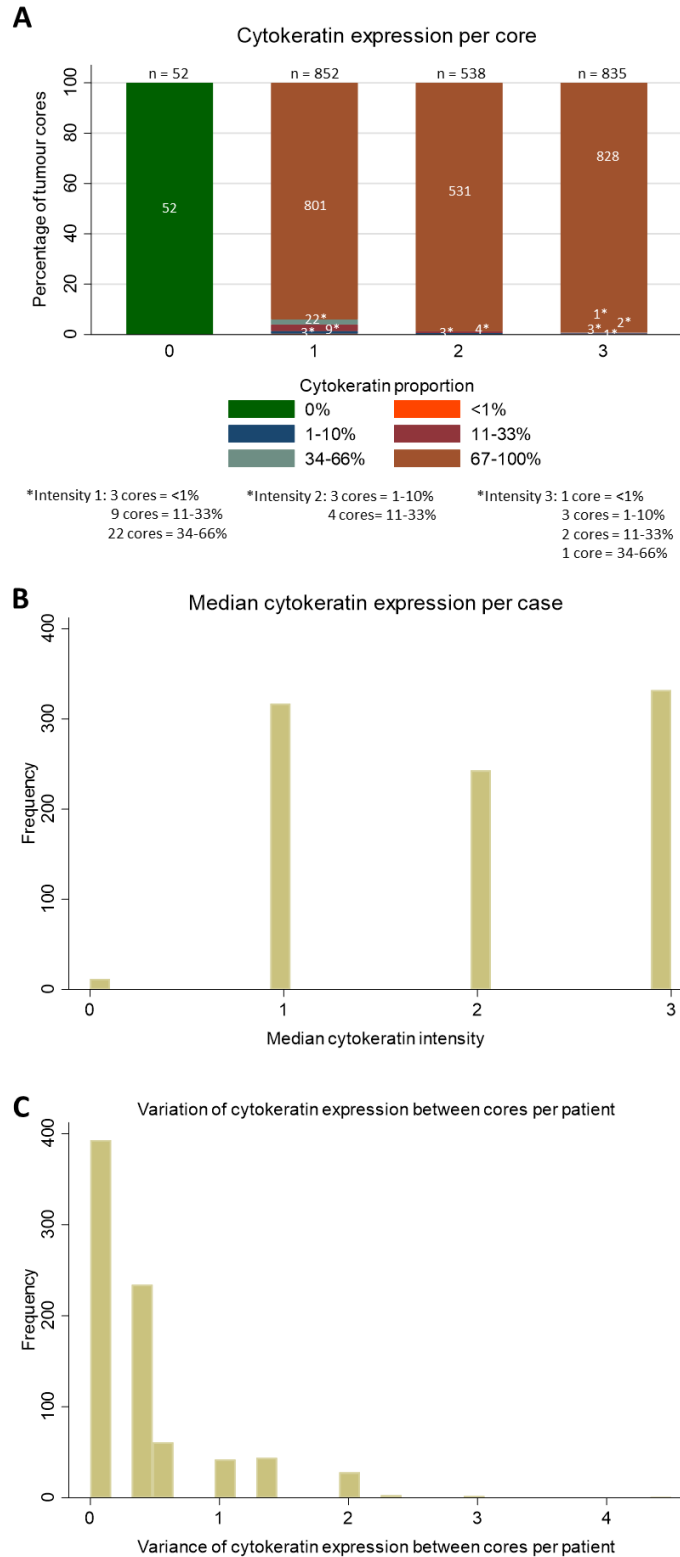
Figure 3.15. Distribution of cytokeratin intensity and proportion scores per core and frequency of cytokeratin expression per case in the cohort.
A: Percentage of tumour cores expressing cytokeratin at different proportion for each intensity in lung adenocarcinoma cores. n = number of cores. B: Frequency of median cytokeratin intensity expression according to the median intensity score in 905 lung adenocarcinoma cases. C: Frequency of the variation of cytokeratin expression between cores within cases in 808 lung adenocarcinoma patients. Unit of measurements: 1 unit = 1 scoring value.

137

The immunohistochemical staining of the mesenchymal cytoskeletal marker vimentin show a variety of expression level. Thus this marker is scored according to the cytoplasmic intensity staining. Vimentin is expressed in the interstitium in the normal lung tissue (Figure 3.16).
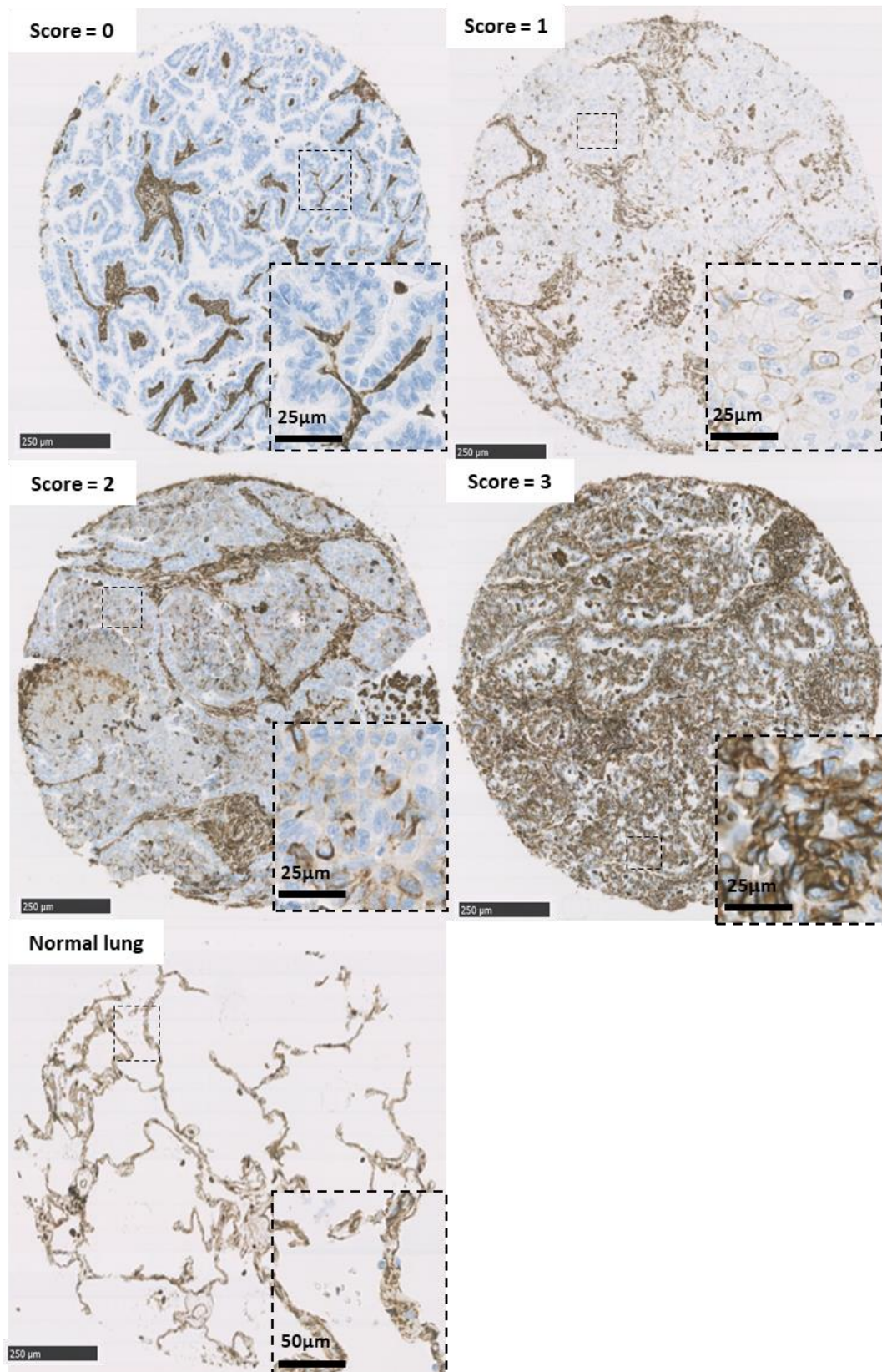
Figure 3.16. Dynamic range of vimentin expression.
Images of the mesenchymal marker vimentin immunostaining scored according to their expression intensities. Blue staining (haematoxylin and bluing reagent) represents nuclei. DAB staining is represented by brown vimentin cytoplasmic staining. Core image magnification = 7x, scale bar = 250 μm. Inset square magnification of the normal lung = 40X, scale bar 50 μm. Scale bar of the high magnification (80X) of a portion of the core, scale bar = 25 μm**.**

139

There is quite high heterogeneity within and between cores in cases which are positive for vimentin (Figure 3.17A).

Approximately 600 cases are negative for vimentin. About 10 cases weakly express vimentin, 50 cases show moderate expression of vimentin and 250 cases strongly express the cytoplasmic mesenchymal marker respectively (Figure 3.17B). There is no difference in vimentin expression between cores within the same primary tumour in most cases. About a third of the cases displays a variation in the mesenchymal marker between cores within cases (Figure 3.17C).

Therefore, vimentin is hetereogenously detected within cores and between cases but is mainly homogenously expressed between cores within cases.
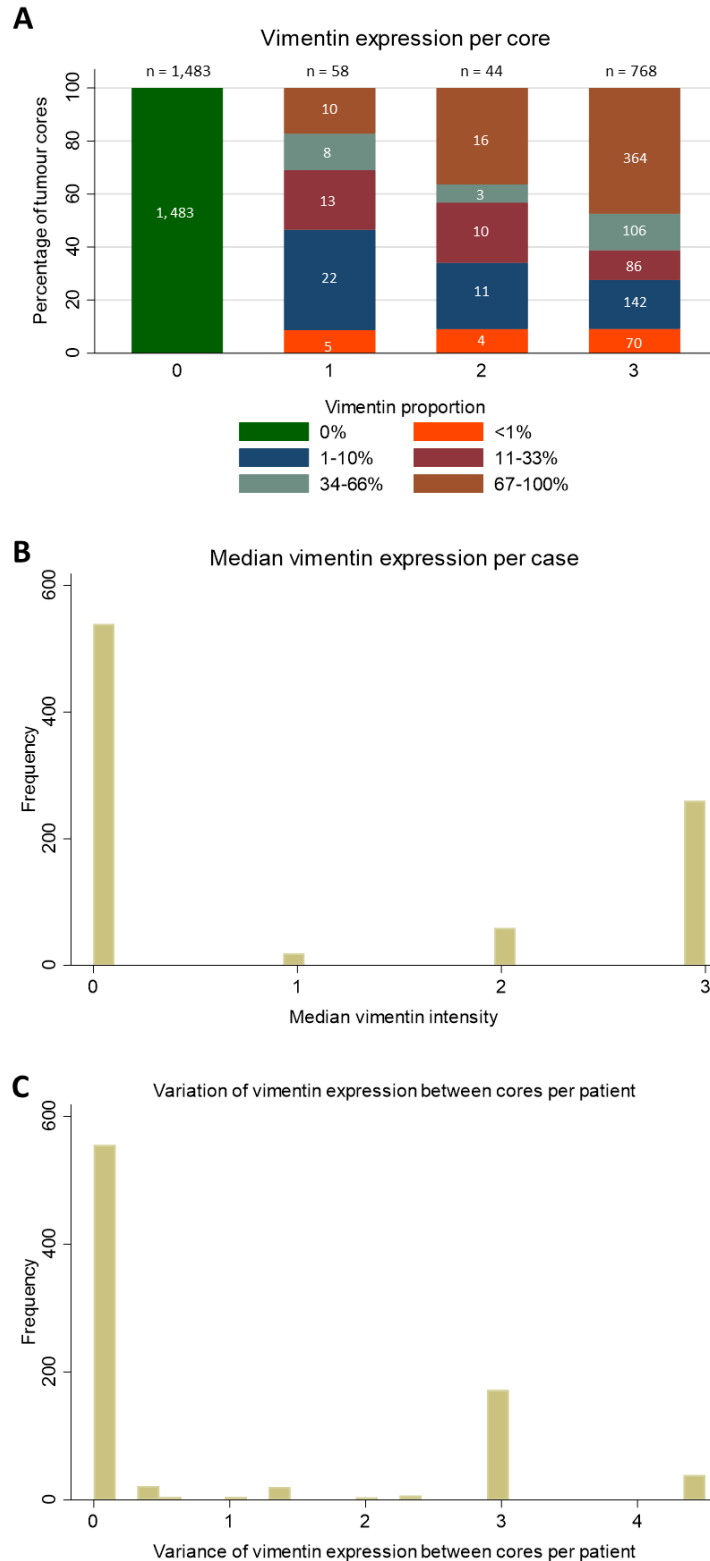
Figure 3.17. Distribution of vimentin intensity and proportion scores per core and frequency of vimentin expression per case in the cohort.
A: Percentage of tumour cores expressing vimentin at different proportion for each intensity in lung adenocarcinoma cores. n = number of cores. B: Frequency of median vimentin intensity expression according to the median intensity score in 878 lung adenocarcinoma cases. C: Frequency of the variation of vimentin expression between cores within cases in 829 lung adenocarcinoma patients.
Unit of measurements: 1 unit = 1 scoring value.

TWIST mRNA expression is illustrated by purple dots, one dot represents one molecule of *TWIST* mRNA. *TWIST* mRNA can also be expressed at different expression level (Figure 3.18).
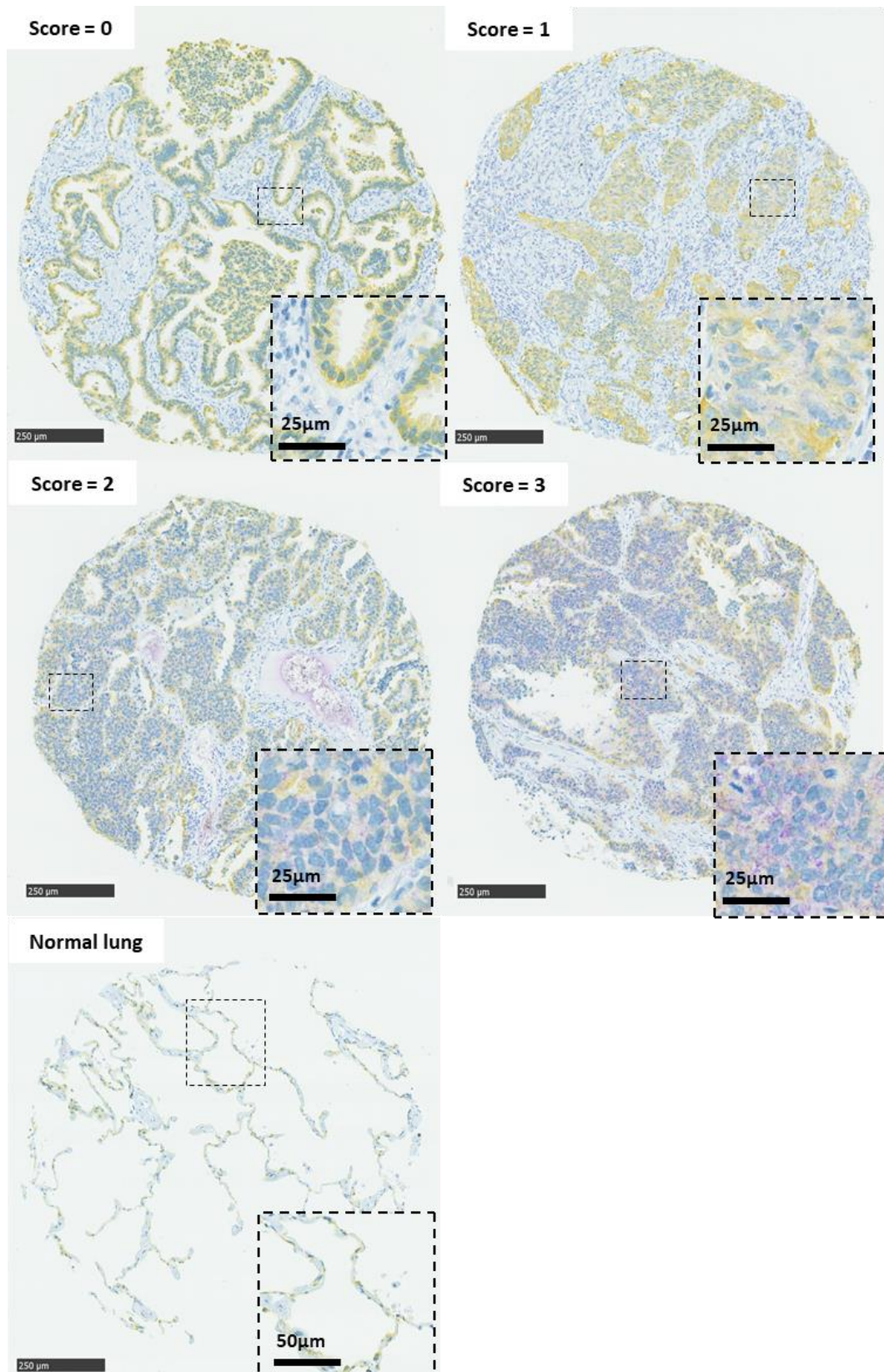
Figure 3.18. *TWIST* mRNA expression level.
Images of the mesenchymal marker *TWIST* mRNA immunostaining scored according to their expression. Blue staining (haematoxylin and bluing reagent) represents the nucleus. *TWIST* mRNA dot is represented by purple staining. Yellow staining represent cytokeratin expression. Core image magnification = 7x, scale bar = 250 μm. Scale bar of the high magnification (80X) of a portion of the core, scale bar = 25 μm. The normal lung high magnification is 40x as I used a different image analysis software (Visiopharm®).

143

The majority of cores does not express *TWIST* mRNA (Figure 3.19A). It is difficult to assess heterogeneity within cores as I count the number of *TWIST* mRNA spots in the tumour core. *TWIST* mRNA cut off between negative and positive staining was selected according to the image visualisation; a core was classified as positive when the scorer can by eye detect spots in the whole tumour core. *TWIST* mRNA spot counts was divided into 4 groups, similarly to Allred scoring system (score 0 for absence of the *TWIST* mRNA up to 3 for high number of *TWIST* mRNA spot counts) based on the number of spots counts.

TWIST mRNA is mainly absent in the cohort. About 250 cases weakly express *TWIST* mRNA. Around 10 cases express moderately or strongly the EMT transcription factor mRNA (Figure 3.19B). *TWIST* mRNA expression does not vary between cores within primary tumours (Figure 3.19C). Thus *TWIST* mRNA is heteregenously quantified between cases but homogenously expressed between cores within cases. Interestingly, the variability of *TWIST* mRNA expression is lower than the variability of the other EMT markers.
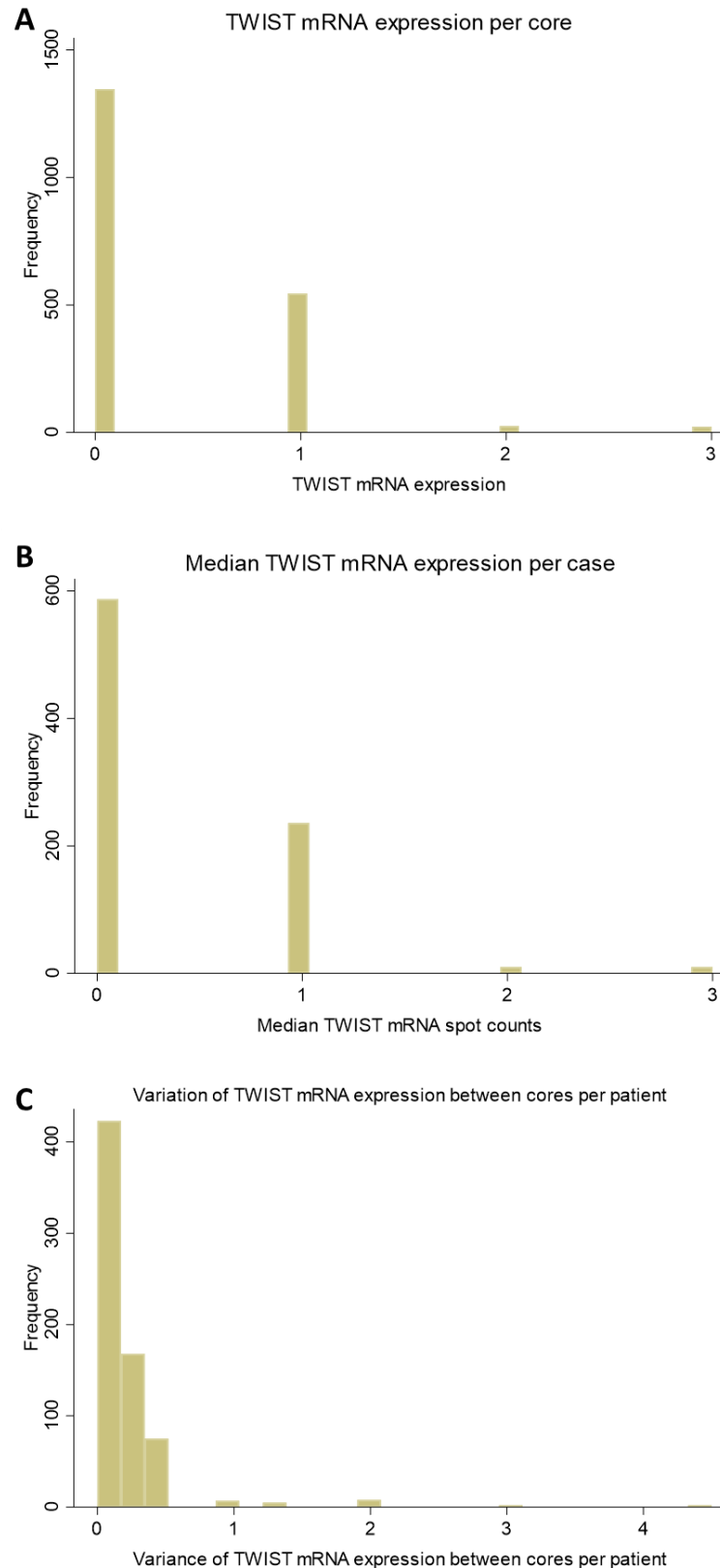
Figure 3.19. *TWIST* mRNA expression level and its distribution in the cohort.
A: Frequency of *TWIST* mRNA expression per core. B: Frequency of *TWIST* mRNA expression per case according to median value in 844 lung adenocarcinoma cases. C: Frequency of the variation of *TWIST* mRNA expression between cores within cases in 690 lung adenocarcinoma patients.
Unit of measurements: 1 unit = 1 scoring value.

Taken together, these data show that most of the primary lung adenocarcinoma show expression of E-cadherin and cytokeratin while approximately 30% of cases express at least one mesenchymal marker N-cadherin, vimentin or *TWIST* mRNA, evocative of a partial EMT process.

Once all the EMT markers were scored, their expression distributions were then analysed across growth patterns using statistical analyses.

3.3.4   EMT and growth pattern

As shown in Figures 3.20 to 3.24, the stacked bar graphs exhibit the classification of the degree of the protein or mRNA of interest in the different tumour growth pattern of lung adenocarcinoma. Lepidic tumour growth is an *in situ* growth pattern while the remaining tumour growth patterns are invasive with solid and micropapillary tumour growth patterns being highly invasive tumour growth (cf Chapter 1, section 1.4.5).

Figure 3.20A indicates the distribution of membranous E-cadherin expression across the predominant growth pattern per case. While E-cadherin is expressed in nearly 100% of lepidic predominant cases, its expression is completely absent in 6.5% (n=13) of solid predominant tumours. However, the difference in E-cadherin expression by the predominant growth pattern is not statistically significant. At the core level, there is only a significant difference in E-cadherin expression between lepidic and cribriform growth pattern but not with the solid pattern.

A

Median E-cadherin expression by predominant tumour growth pattern

B

E-cadherin expression by TMA core growth pattern
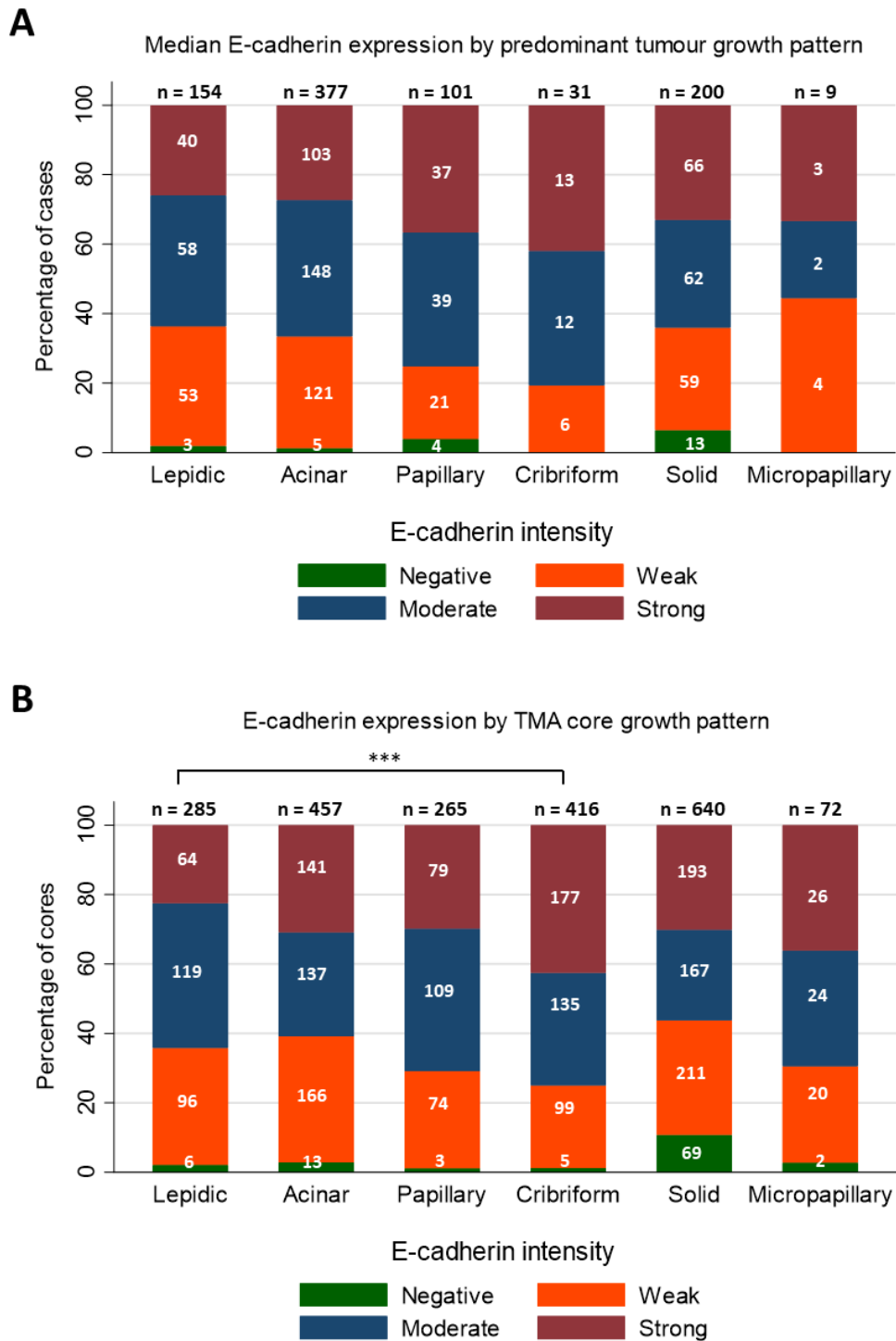
Figure 3.20. Distribution of E-cadherin expression by lung adenocarcinoma tumour growth pattern. Each bar displays the tumour growth pattern of lung adenocarcinoma and each segment of the bar represents the intensity of the EMT marker. Total number of growth pattern cases is indicated at the top of each bar and the number of cases displaying the different degree of the EMT marker expression is shown within each section of the bar. To determine the existance of significance between EMT markers/translation factors expression and tumour growth pattern, a Kruskal-Wallis tests followed by a Dunn's test with Bonferroni adjustment were performed. Degree of significance are indicated in the graphs by bars with stars above them (*** = $p < 0.001$). A: Percentage of median E-cadherin expression intensity by predominant growth pattern at case level, n = number of cases. B: Percentage of E-cadherin expression intensity by growth pattern at tumour core level, n = number of cores.

The proportion of cases strongly expressing the epithelial marker cytokeratin significantly diminishes with invasive tumour growth with about 25% of solid predominant tumour growth patterns being strongly positive for cytokeratin. Interestingly, cytokeratin is absent in about 2% of the solid cores whereas it is present in 100% of lepidic tumour growth pattern (Figure 3.21A). TMA core level analysis exhibits the same trend in the expression of cytokeratin between *in situ* and invasive tumour cores (Figure 3.21B).

This suggests that cytokeratin loss is correlated with invasive properties, especially in the solid tumour pattern.

Interestingly, the high-grade micropapillary predominant tumours express both epithelial markers (Figures 3.20-3.21).
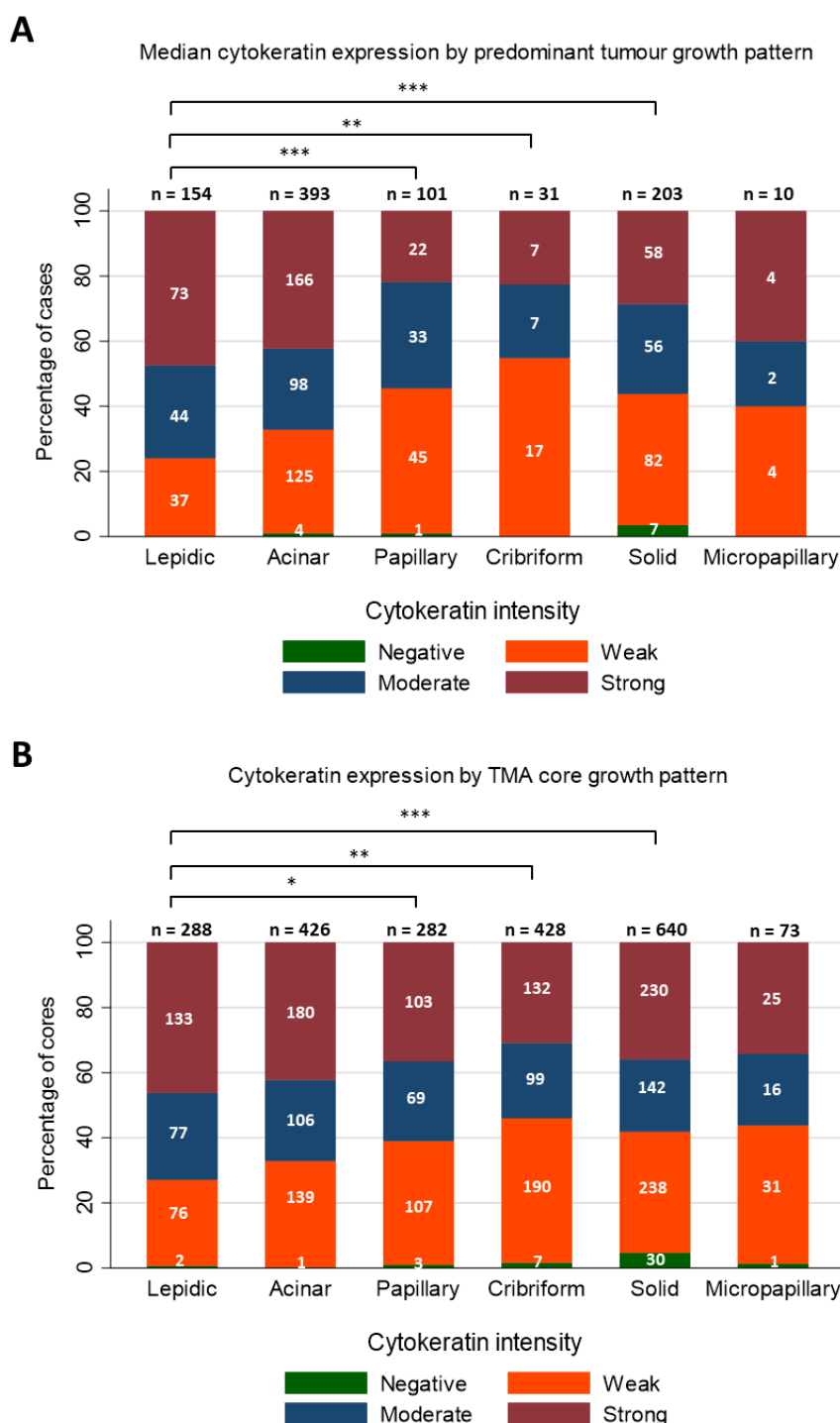
Figure 3.21. Distribution of cytokeratin expression by lung adenocarcinoma tumour growth pattern.
Each bar displays the tumour growth pattern of lung adenocarcinoma and each segment of the bar represents the intensity of the EMT marker. Total number of growth pattern cases is indicated at the top of each bar and the number of cases displaying the different degree of the EMT marker expression is shown within each section of the bar. To determine the existance of significance between EMT markers/translation factors expression and tumour growth pattern, a Kruskal-Wallis tests followed by a Dunn's test with Bonferroni adjustment were performed. Degree of significance are indicated in the graphs by bars with stars above them (* = p < 0.05; ** = p <0.01; *** = p < 0.001). A: Percentage of median cytokeratin expression intensity by predominant growth pattern at case level, n = number of cases. B: Percentage of cytokeratin expression intensity by growth pattern at tumour core level, n = number of cores.

As opposed to the expression of E-cadherin, the mesenchymal marker N-cadherin is only detectable in about 35% on average of all predominant tumour growth pattern cases. N-cadherin is expressed in only about 19% of lepidic tumour predominant pattern but is expressed in 50% of solid areas and the difference in N-cadherin expression between this two predominant tumour growth pattern is statistically significant (Figure 3.22A). Similar observations are present at the core level analysis (Figure 3.22B).

A

Median N-cadherin expression by predominant tumour growth pattern

B

N-cadherin expression by TMA core growth pattern

Lepidic : 5 cores = Moderate   *Papillary : 1 core = Moderate
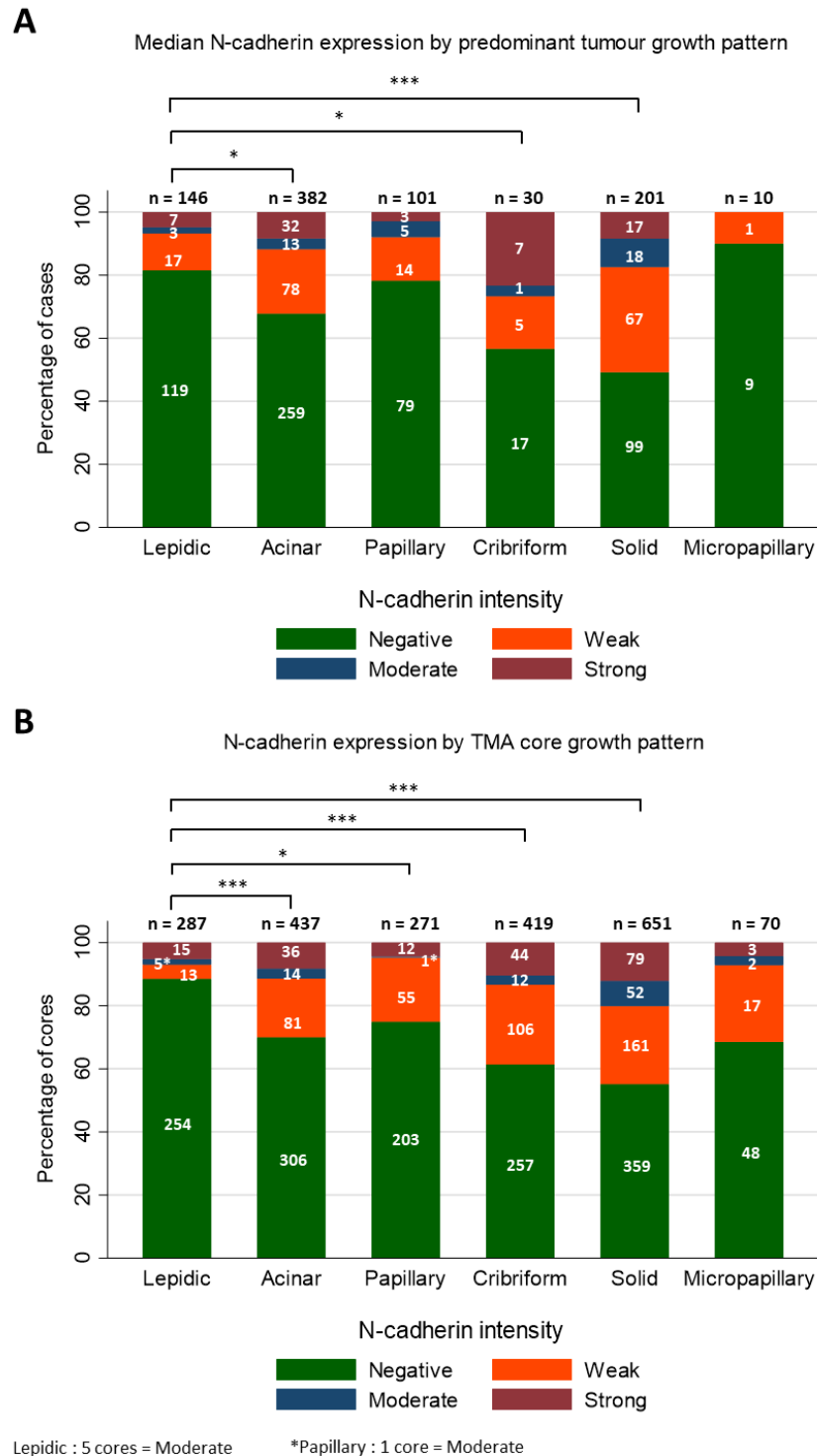
Figure 3.22. Distribution of N-cadherin expression by lung adenocarcinoma tumour growth pattern.
Each bar displays the tumour growth pattern of lung adenocarcinoma and each segment of the bar represents the intensity of the EMT marker. Total number of growth pattern cases is indicated at the top of each bar and the number of cases displaying the different degree of the EMT marker expression is shown within each section of the bar. To determine the existance of significance between EMT markers/translation factors expression and tumour growth pattern, a Kruskal-Wallis tests followed by a Dunn's test with Bonferroni adjustment were performed. Degree of significance are indicated in the graphs by bars with stars above them (* = $p < 0.05$; *** = $p < 0.001$). A: Percentage of median N-cadherin expression intensity by predominant growth pattern at case level, n = number of cases. B: Percentage of N-cadherin expression intensity by growth pattern at tumour core level, n = number of cores.

151

In contrast to cytokeratin expression, the total percentage of tumour expressing vimentin is low in the lepidic predominant tumour growth pattern with approximately 35% of vimentin positive lepidic tumours whereas this proportion increases in solid predominant tumour pattern for up to nearly 50% and the difference in vimentin expression between the well-differentiated and poorly differentiated tumour predominant growth pattern is statistically significant (Figure 3.23A). The same conclusion can be drawn from the TMA core pattern analysis (Figure 3.23B).

A

Median vimentin expression by predominant tumour growth pattern

Vimentin intensity

- Negative
- Moderate
- Weak
- Strong

Acinar : 8 cores = Weak *Papillary : 1 core = Weak *Solid : 3 cores = Weak

B

Vimentin expression by TMA core growth pattern

Vimentin intensity

- Negative
- Moderate
- Weak
- Strong

Lepidic : 8 cores = Weak * Acinar : 11 cores = Weak *Papillary : 6 cores = Weak
3 cores = Moderate 5 cores = Moderate 3 cores = Moderate

Figure 3.23. Distribution of vimentin expression by lung adenocarcinoma tumour growth pattern.
Each bar displays the tumour growth pattern of lung adenocarcinoma and each segment of the bar represents the intensity of the EMT marker. Total number of growth pattern cases is indicated at the top of each bar and the number of cases displaying the different degree of the EMT marker expression is shown within each section of the bar. To determine the existance of significance between EMT markers/translation factors expression and tumour growth pattern, a Kruskal-Wallis tests followed by a Dunn's test with Bonferroni adjustment were performed. Degree of significance are indicated in the graphs by bars with stars above them (** = p <0.01). A: Percentage of median vimentin expression intensity by predominant growth pattern at case level, n = number of cases. B: Percentage of vimentin expression intensity by growth pattern at tumour core level, n = number of cores.

153

Regarding the *TWIST* mRNA expression in the different growth pattern, nearly 40% of the solid tumour pattern express *TWIST* mRNA as opposed to about 25% on average in the lepidic tumour pattern but the difference in the expression of the EMT transcription factor across growth pattern is only statistically significant at the TMA core level (Figure 3.24).

This suggests that molecular EMT can be detected in primary solid predominant tumour growth pattern. Interestingly, the highly lethal micropapillary predominant tumour growth pattern exhibits little evidence of molecular EMT.

A

Median TWIST mRNA expression by predominant tumour growth pattern

B

TWIST mRNA expression by TMA core growth pattern
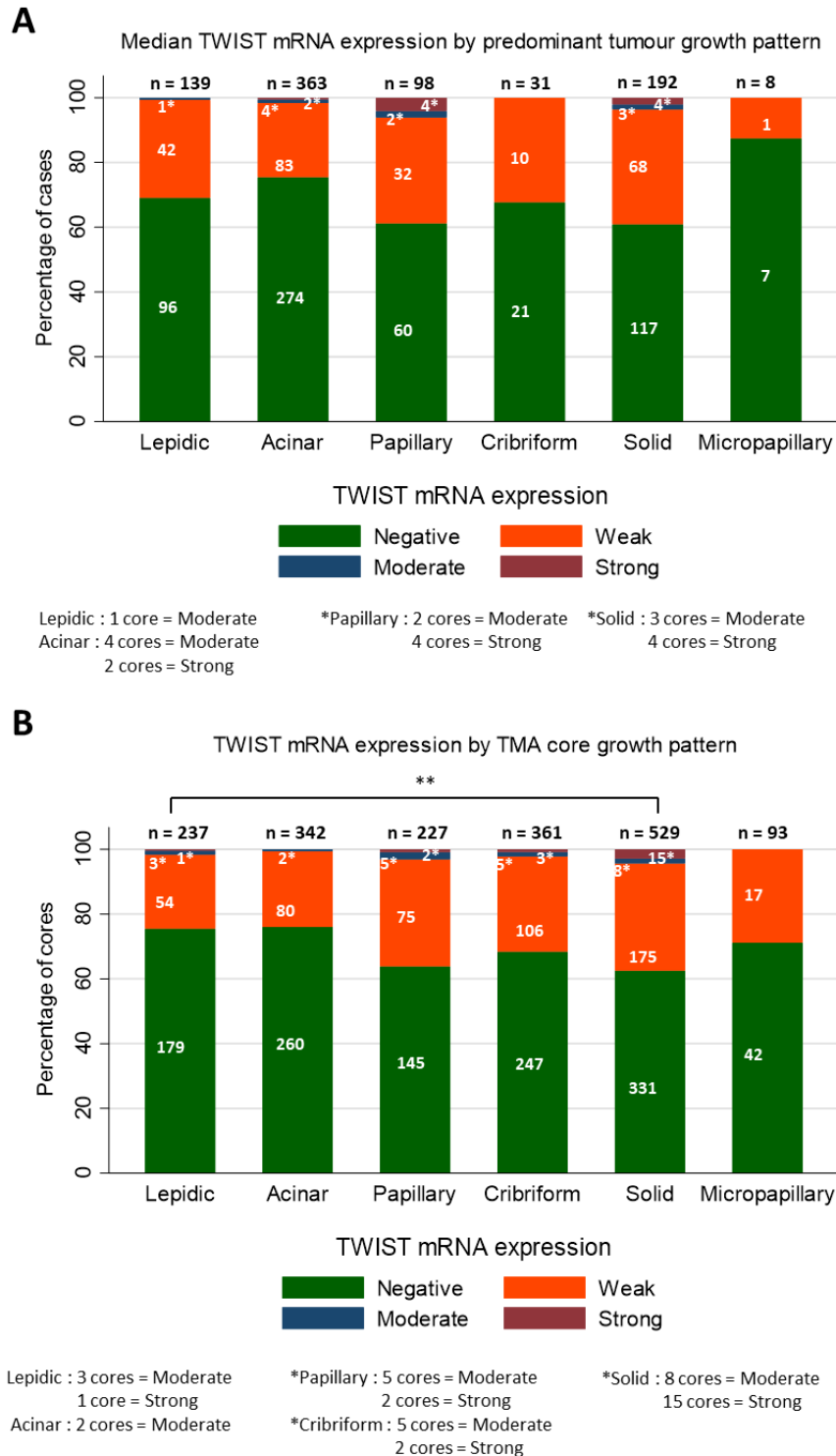
Figure 3.24. *TWIST* mRNA expression by lung adenocarcinoma tumour growth pattern.
Each bar displays the tumour growth pattern of lung adenocarcinoma and each segment of the bar represents the intensity of the EMT marker. Total number of growth pattern cases is indicated at the top of each bar and the number of cases displaying the different degree of the EMT marker expression is shown within each section of the bar. To determine the existance of significance between EMT markers/translation factors expression and tumour growth pattern, a Kruskal-Wallis tests followed by a Dunn's test with Bonferroni adjustment were performed. Degree of significance are indicated in the graphs by bars with stars above them (** = p <0.01). A: Percentage of median *TWIST* mRNA expression across predominant growth pattern at case level, n = number of cases. B: Percentage of *TWIST* mRNA expression intensity across growth pattern at tumour core level, n = number of cores.

I then combined the scores from all EMT markers to make maximum use of available information. I created a new variable called "EMT score" which results from the addition of each EMT marker. To do so, all scoring values from the three cores per case were considered and binarised as 0 "No EMT" and 1 "EMT" for each EMT marker. The loss of the epithelial (i.e. cores that were scored 0 for E-cadherin and cytokeratin) and the gain of the mesenchymal markers expression (cores that were assessed as score 1, 2 or 3 for N-cadherin, vimentin and *TWIST* mRNA) was scored as 1. All the binarised individual EMT marker were summed together creating the "EMT score" variable. This latter contains values from 0 for "No EMT" up to 5 for "Complete EMT" and was then correlated with tumour histological pattern.

To identify cores that present molecular evidence of EMT, I prepared a Venn diagram. In this study, there are five variables. Each variable represents one molecular EMT event, i.e. any loss/gain of epithelial/mesenchymal markers respectively. When two or more molecular EMT groups overlap between each other, it means that the cores display those particular molecular EMT events.

The Venn diagram indicates all the combinations of molecular EMT occurring in our cohort. Amongst 1420 cores, 910 cores exhibit some evidence of molecular EMT while 510 cores do not. Two third of positive EMT cores show one molecular EMT event: 202 cores are only *TWIST* mRNA positive; 319 cores are vimentin positive; 61 cores are N-cadherin positive; 19 and 7 cores have respectively lost membranous E-cadherin and cytoskeletal cytokeratin expression. Only 4 cores exhibit a complete molecular EMT program, i.e. loss of epithelial markers and gain of the mesenchymal markers (Figure 3.25A).

Figure 3.25B depicts a stacked bar graph representing the distribution of growth patterns by the increasing numbers of molecular EMT markers. It shows that additional indicators of molecular EMT incrementally increase the likelihood of solid growth pattern, and diminish the likelihood of all other *in situ* and invasive growth patterns (from about 25% of cores expressing one molecular evidence of EMT to 100% of the cores undergoing full molecular EMT). Furthermore there is a significant difference in EMT score between solid tumour cores and all other growth pattern cores (Figure

3.25C), suggesting that this demonstrates evidence of molecular EMT in human primary solid adenocarcinoma.

I then wanted to check whether I can identify biologically meaningful groups in the data based on similarity of EMT markers. To do so, I performed an unsupervised cluster analysis with the help of the Jeffrey.

Vimentin↑ N-cadherin↑

319 61
23 43
33 1 3
1 2 19
1 2 2
2 2 14 E-cadherin↓
139 4 2
1 2
9 6 1 0 0
2 2 3 2
Twist↑ 202 4 7
Cytokeratin↓

Total = 1 420 cores :
- No EMT= 510 cores
- EMT= 910 cores

B

Solid pattern shows molecular EMT

	n = 485	n = 568	n = 214	n = 51	n = 12	n = 4

Percentage of growth pattern (y-axis, 0–100)

No EMT: 71, 115, 75, 102, 99, 23
1 EMT marker: 75, 105, 88, 125, 160, 15
2 EMT markers: 27, 27, 29, 55, 70, 6
3 EMT markers: 1, 6, 3, 10, 31
4 EMT markers: 12
Full EMT: 4

Growth pattern

Legend: Lepidic, Acinar, Papillary, Cribriform, Solid, Micropapillary

C

Dunn's pairwise comparison test	Lepidic	Acinar	Papillary	Cribriform	Solid	Micropapillary
Lepidic	-	NS	NS	NS	P < 0.001	NS
Acinar	NS	-	NS	P = 0.020	P < 0.001	NS
Papillary	NS	NS	-	NS	P < 0.001	NS
Cribriform	NS	P = 0.020	NS	-	P = 0.007	NS
Solid	P < 0.001	P < 0.001	P < 0.001	P = 0.007	-	P = 0.007
Micropapillary	NS	NS	NS	NS	P = 0.007	-

Figure 3.25. Distribution of number of molecular EMT by tumour growth patterns at TMA core level.
A: Venn diagram of the total scorable cores of the cohort. This diagram shows cores that undergo molecular EMT: arrows going up mean the gain of expression of the marker and those that are pointing down show loss of expression of the marker. B: Percentage of growth pattern by increasing number of molecular EMT markers, n = number of cores. Each bar displays the number of molecular EMT markers and each segment of the bar represents tumour growth patterns of lung adenocarcinoma. Total number of cores is indicated at the top of each bar and the number of cores displaying tumour growth pattern is shown within each section of the bar. C: Table presenting p-values results of Dunn's pairwise comparison in the number of molecular EMT (i.e. EMT score) between tumour growth patterns at core level. NS = not significant.

The cluster analysis further reveals that clusters presenting at least three molecular evidence of EMT, i.e. clusters 4, 5 and 6, are closely related to solid growth pattern (Figure 3.26).

Cluster 1 (green): E-cadherin + ; Cytokeratin + ; Vimentin + ; N-cadherin - ; TWIST -
Cluster 2 (red): E-cadherin + ; Cytokeratin + ; Vimentin - ; N-cadherin - ; TWIST -
Cluster 3 (purple): E-cadherin + ; Cytokeratin + ; Vimentin - ; N-cadherin + ; TWIST -
Cluster 4 (orange): E-cadherin +/- ; Cytokeratin +/- ; Vimentin + ; N-cadherin + ; TWIST+
Cluster 5 (blue): E-cadherin - ; Cytokeratin - ; Vimentin + ; N-cadherin + ; TWIST -
Cluster 6 (brown): E-cadherin - ; Cytokeratin - ; Vimentin - ; N-cadherin + ; TWIST +

Figure 3.26. Unsupervised cluster analysis of EMT markers.
A: Unsupervised cluster analysis of EMT markers among TMA core patterns. Six clusters with different EMT patterns were identified in the analysis. LEP: Lepidic . AC: Acinar. CRIB: Cribriform. PAP: Papillary. MPAP: Micropapillary. B: Pie-charts of TMA core patterns for each cluster. n = number of TMA cores.

Taken together, solid tumours exhibit loss of epithelial markers and gain of mesenchymal markers, showing molecular evidence of EMT despite having epithelioid morphology (i.e. non-sarcomatoid). Surprisingly, about 15% of lepidic cases in average do show at least one EMT molecular indicator, indicating acquisition of partial EMT by cancer cells in an *in situ* tumour growth. Micropapillary tumour growth is associated with poor prognosis as the solid pattern but shows comparatively less molecular evidence of EMT.

3.3.5   Expression of cytoplasmic E-cadherin is associated with invasive tumour growth pattern in primary lung adenocarcinoma

A further notable observation was made while scoring E-cadherin expression. Despite its known membrane function as cell-cell adhesion molecule, E-cadherin was found expressed in the cytoplasm in some tumour regions (Figure 3.27).

Membranous E-cadherin < Cytoplasmic E-cadherin
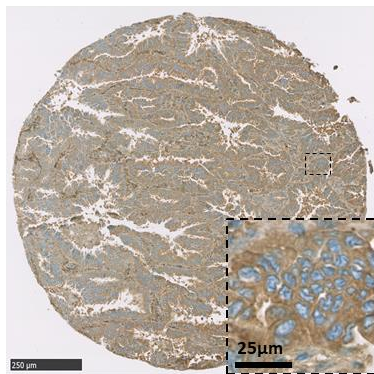
Membranous E-cadherin = Cytoplasmic E-cadherin
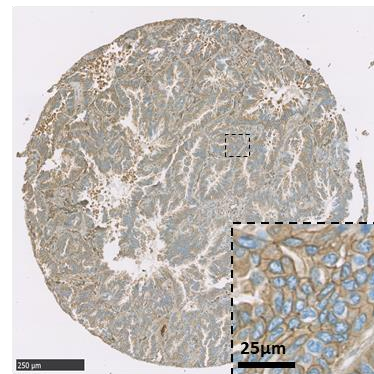
Membranous E-cadherin > Cytoplasmic E-cadherin
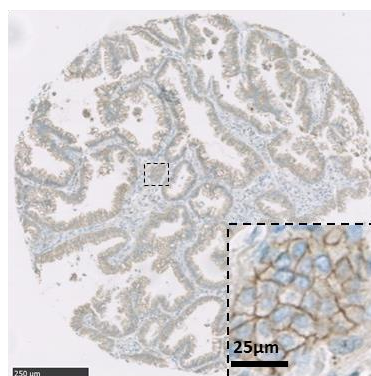
Figure 3.27. IHC images of cytoplasmic/membranous E-cadherin expression difference.
On the top left hand, image shows high expression of E-cadherin in the cytoplasm and low E-cadherin expression in the membrane. On the top right had side, E-cadherin is equally expressed in cytoplasm and membrane. On the bottom image, membranous E-cadherin is more pronounced than cytoplasmic E-cadherin expression. Blue staining (haematoxylin and bluing reagent) represents the nucleus. DAB staining is represented by brown cytoplasmic/membranous E-cadherin staining. Core image magnification = 5x, scale bar = 500 μm. Inset squares represent high resolution (80X) of a portion of the cores, scale bar = 25 μm.

I suspected that mislocalization of E-cadherin in the cytoplasm might contribute to invasion as E-cadherin in this cellular compartment would be ineffectual in cellular adhesion. I therefore analysed the ratio of membranous/cytoplasmic E-cadherin expression by TMA core growth pattern to check whether subcellular localisation of E-cadherin can be related to invasive tumour histological pattern. I excluded tumour cores which were negative for both membranous and cytoplasmic E-cadherin.

Cytoplasmic E-cadherin expression increases with invasive growth pattern, with about 28% of cores expressing prominantly cytoplasmic E-cadherin in the lepidic/*in situ* cores versus around 38% in the solid pattern cores. In contrast, membranous E-cadherin

expression decreases with invasive pattern, about 58% of *in situ* cores predominantly expressing E-cadherin versus 42% in solid cores. These differences are statistically significant, suggesting that cytoplasmic retention of E-cadherin is related to invasiveness (Figure 3.28).
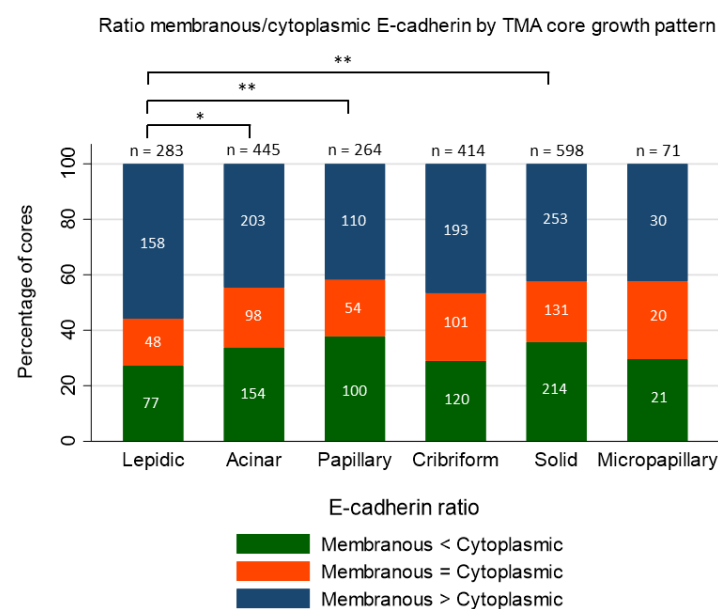
Figure 3.28. Ratio membranous/cytoplasmic E-cadherin expression by TMA tumour core growth pattern.
Each bar displays the tumour growth pattern of lung adenocarcinoma and each segment of the bar represents E-cadherin ratio. Total number of growth pattern cases is indicated at the top of each bar and the number of cases displaying different E-cadherin ration is shown within each section of the bar. Degree of significance are indicated in the graphs by bars with stars above them (* = p < 0.05; ** = p < 0.01). n = number of cores. Total number of cores = 2,085 tumour cores.

3.4   Discussion

The first purpose of this chapter is to explore the relationship between predominant growth pattern and lung adenocarcinoma patient outcome.

In the first section, I have shown that solid and micropapillary predominant tumours growth are significantly related to poor overall, lung cancer-specific and recurrence-free survivals in lung adenocarcinoma patients. Furthermore, 50% patients developing these high-grade tumours have the probability to relapse just a year post-surgery (Figures 3.1-3.3). This is consistent with previous studies highlighting the association of micropapillary and solid subtypes with poor prognosis and early recurrence of the disease (Hung *et al.*, 2014a; Makimoto *et al.*, 2005; Zhang *et al.*, 2014). Consequently, solid and micropapillary are indicators of worst outcome in lung adenocarcinoma patients. Is this related to EMT?

The second purpose of this chapter is to determine whether EMT was detectable in human lung primary tumour tissue and the extent of variation between cases and cores.

When analysing heterogeneity in EMT markers expression, two different scales of heterogeneity were identifiable: microscopic (i.e. within cores) and regional (i.e. between cores within the primary tumour) heterogeneity. Heterogeneity at a microscopic scale was observed in the expression of mesenchymal markers N-cadherin (Figure 3.13A) and vimentin (Figure 3.17A). This evidence of microscopic heterogeneity in partial EMT might be explained by extremely focal epigenetic changes, or micro-environmental cues such as local hypoxia or alterations in the immune microenvironment. Regional heterogeneity, i.e. variance in EMT marker between cores (i.e. across the tumour as a whole) was not marked (Figures 3.11C, 3.13C, 3.15C, 3.17C and 3.19C). When it does occur, this may be due to heterogeneity between spatially separated clones.

The results from Figures 3.20 to 3.24 have shown that loss of epithelial markers E-cadherin and cytokeratin expression and increased expression of mesenchymal markers *TWIST* mRNA, N-cadherin and vimentin are most associated with the solid pattern of tumour growth in lung adenocarcinoma. There is an association between the number

of EMT markers and solid growth pattern (Figure 3.25B and C), rather than a strong link with a particular combination of EMT markers (Figure 3.26), suggesting that the number of molecular EMT (i.e. gain or loss of mesenchymal or epithelial markers respectively) matters the most rather than any combination of molecular EMT. It may also imply that there are many routes to activate EMT as there are several combinations of partial EMT detected in the cluster analysis (Figure 3.26). These data convey the notion that solid pattern can be thought of as representing a partial EMT state despite the morphologically epithelioid appearance of cells in the primary tumour.
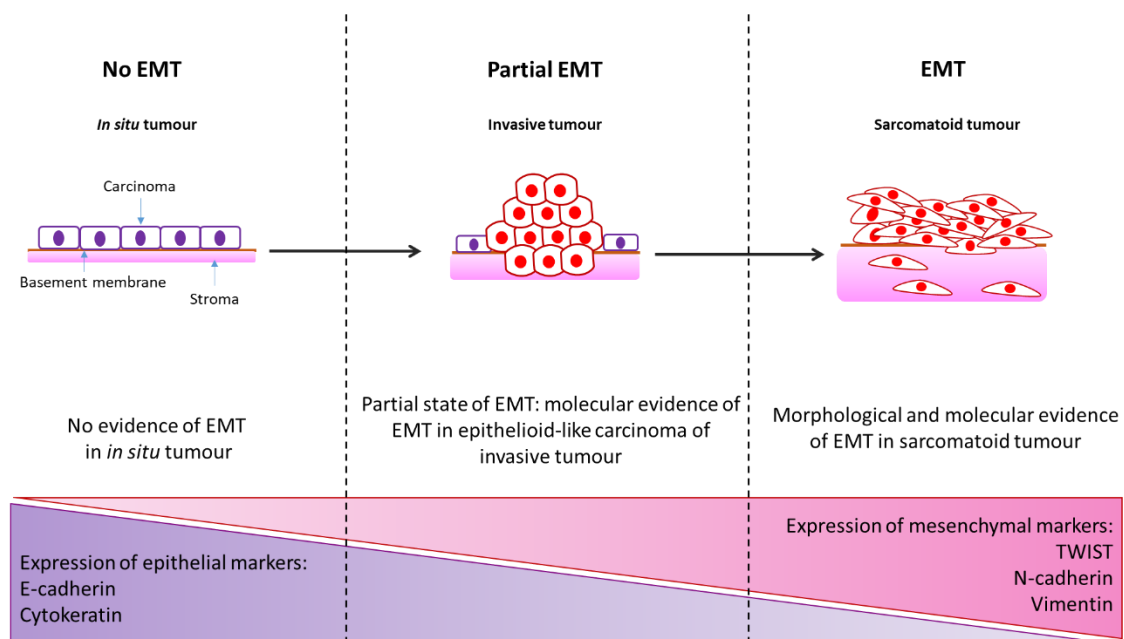
Figure 3.29. Schematic representation of molecular evidence of EMT in tumour growth in primary lung adenocarcinoma.
Epithelial malignant cells (carcinoma) constituting the *in situ* tumour are laid on the basement membrane (brown) which is in contact with the stroma (pink). They are positive for the epithelial markers E-cadherin and cytokeratin and negative for the mesenchymal markers N-cadherin, vimentin and TWIST. Upon internal and/ or external stimuli, these tumour cells grow into invasive tumour pattern and acquire the potentiality to become ultimately mesenchymal. They are in partial state where carcinoma express both epithelial and mesenchymal markers while retaining epithelioid morphology. Finally, carcinoma can transform into fully functional mesenchymal cell with loss of the molecular and morphological epithelial features and gain of mesenchymal characteristics, i.e. spindle-shaped cells expressing the mesenchymal markers. This is observed in sarcomatoid tumour growth.

The notion of partial EMT refers to the fact that tumour cells are identified with concomitant expression both epithelial and mesenchymal markers while maintaining the epithelioid morphology, which may represent the intermediate stages of EMT, called "partial EMT" (Kalluri and Weinberg, 2009) but not a complete sarcomatoid

transformation (Figure 3.29). EMT program is not considered as a binary event but rather a continuum process with several steps. Therefore, I have demonstrated the presence of partial EMT in solid lung adenocarcinoma. By inducing EMT in solid adenocarcinoma tumour, tumour cells express their need to invade and migrate from the primary tumour. Several reasons can explain this phenomenon as several signalling pathways lead to EMT induction such as Wnt/β-catenin, hypoxia or EGFR pathways (Lamouille *et al.*, 2014). It would be interesting to understand what factors (e.g. tumour microenvironment, genetic alterations) activate EMT in lung adenocarcinoma.

Despite the micropapillary tumour growth being lethal and correlated with tumour metastasis (Miyoshi *et al.*, 2003; Yoshizawa *et al.*, 2011), it does not in general show molecular evidence of EMT in primary lung adenocarcinoma. Our data has shown that micropapillary tumours were generally E-cadherin and cytokeratin positive and negative for the mesenchymal markers, suggesting that micropapillary tumours mostly retained their epithelial features. Thus, micropapillary growth has acquired invasive and metastatic potential seemingly without undergoing widespread molecular EMT and while retaining morphological and molecular epithelial features. The next chapter attempts to further understand the formation of micropapillary tumour to explain its poor prognosis.

Interestingly, lepidic/*in situ* tumour growth pattern, known to be low-grade tumour with non-invasive characteristics and associated with good prognosis, often manifests evidence of partial molecular EMT, suggestive of acquisition of invasive behaviour by cancer cells in the *in situ* tumour growth. This suggests that either the lepidic growth region may be colonised by invasive-competent tumour cells or this may represent early steps towards invasive growth formation in lung adenocarcinoma. These findings are further explored in the next chapter.

Finally, I found evidence of cytoplasmic E-cadherin being more predominant in invasive tumour cores, suggesting a redistribution of the membranous E-cadherin expression to the cytoplasm. Stanczak *et al.* have demonstrated that the switch in E-cadherin expression from the membrane to the cytoplasm is necessary for the formation of mesenchymal phenotype in advanced colorectal cancer (Stanczak *et al.*, 2011).

Cytoplasmic E-cadherin expression was also observed in lung cancer (Böhm *et al.*, 1994; Liu *et al.*, 2009; Nawrocki *et al.*, 1998). It would be interesting to explore what triggered E-cadherin cytoplasmic relocalisation in lung cancer. One explanation for cytoplasmic E-cadherin redistribution has been provided by Nawrocki *et al*. They have demonstrated that down-regulation of the tyrosine phosphorylation of E-cadherin was associated with membranous E-cadherin redistribution to the cytoplasm in bronchopulmonary carcinomas, which may result in the dissociation of E-cadherin-catenin complex and loss of membranous E-cadherin expression (Nawrocki *et al.*, 1998).

Our data brought up interesting and unexpected findings such as molecular evidence of EMT in *in situ* tumour growth, which is known to predict better outcome or absence of molecular EMT in general in micropapillary tumour growth, known to be invasive and associated with poor prognosis. The next chapter aims to further investigate on these findings.

Chapter 4.   Molecular EMT is highly variable in lung adenocarcinoma and is

often related to local invasiveness

Chapter 4 Molecular EMT is highly variable in lung adenocarcinoma and is often related to local invasiveness

4.1 Introduction

This chapter focuses on four localised patterns of partial EMT, observed while scoring the 2826 cores for each EMT marker, which each illustrate different aspects of how EMT may arise locally and contribute to tumour virulence. Molecular EMT signifies the detection of any molecular evidence of EMT, i.e. any loss/gain of epithelial and mesenchymal markers respectively. The 'molecular EMT' I observe is almost always partial, i.e. unaccompanied by morphological evidence of mesenchymal differentiation.

The first observed phenonemon was how molecular EMT accompanies the transition to invasiveness in mixed type tumours. More than 90% of primary lung adenocarcinoma are classified as mixed subtype, i.e. showing both *in situ* and invasive growth patterns (Travis *et al.*, 2011). As EMT is generally believed to underlie invasive behaviour in epithelial tumour cells, I wanted to explore whether the transition to invasiveness in mixed *in situ* and invasive tumours is also accompanied by molecular EMT.

The second observed phenomenon is the finding of molecular evidence of EMT detected in areas of lepidic/*in situ* tumour growth. This surprising finding may show us how invasive behaviours can arise in a predominantly *in situ* precursor lesion.

Thirdly, sarcomatoid tumours with spindled areas are the rare examples of EMT observeable by haematoxylin and eosin (H&E) histology. I therefore went on to examine EMT markers in such tumours. Non-small cell lung cancers such as these, displaying mesenchymal changes alongside an adenocarcinoma component are classified by the WHO as "pleomorphic carcinoma". I had speculated whether the pleomorphic carcinoma component would have acquired more evidence of molecular EMT compared to the adenocarcinoma component.

Finally, I have already shown that micropapillary growth pattern shows an interesting negative association with markers of molecular EMT. Nonetheless, micropapillary growth pattern strongly predicts poor prognosis in primary lung adenocarcinoma

patients. How is this distinctive morphology formed, and why is it so lethal in patients? Might EMT play a role in the formation or virulence of micropapillary tumours?

4.2   Aim and objectives

The main purpose of this chapter is to further investigate the relationships between focal/partial EMT processes and tumour growth:

> i)   Can molecular markers of EMT highlight focal invasion within morphologically heterogenous tumours?
>
> ii)   How does focal invasiveness contribute to the establishment of tumour morpholgy as judged by growth patterns?

4.3   Results

4.3.1   Genomic progression from *in situ* to invasive growth is associated with progression in molecular EMT

4.3.1.1   Introduction

As described in the introduction, in a landmark study Noguchi *et al.* demonstrated that six distinct classes of early lung adenocarcinoma could be identified based on their morphology, one of which is the "type C" tumours which are tumours with both *in situ* and truly invasive (i.e. not pseudo-invasive) growth (Noguchi *et al.*, 1995).

A recent study from our laboratory has showed that the Noguchi type C tumours can further be categorised into two subtypes based on the appearance of the *in situ* component of the tumour, namely C1 and C2 tumours (Moore *et al.*, 2019). These tumours are biologically and morphologically different.

Invasion in C1 tumours represents a step change in cell behaviour from *in situ* to invasive growth, and additional genomic alterations in invasive areas are often found in these tumours. These tumours typically contain asymmetric foci of invasion suggesting a

random genetic or epigenetic event, they often show a higher nuclear grade and proliferation rate in the invasive area than the *in situ* region (Figure 4.1A and B), and also often exhibit evidence of genomic progression. In our study, 5 of 27 cases of morphologically C1 tumours, invasive cells had acquired genomic alterations in p53. Furthermore, C1 tumours are associated with good prognostic outcome relative to C2 (Moore *et al.*, 2019).

Contrary to C1 tumours, in C2 tumours the *in situ* growth area is thought to represent a lepidic surface outgrowth of invasion-competent cells. These tumours do not show genomic difference between the *in situ* and invasive areas of the tumours. They typically exhibit an invasive region surrounded by a constant-thickness "penumbra" of lepidic peripheral outgrowth with cells displaying the same high-grade nuclear features in both invasive and *in situ* compartments (Figure 4.1A and B). C2 tumours are highly proliferative compared to C1 tumours but there is no difference in the proliferation rate between both *in situ* and invasive compartments. When present, p53 mutations are shared both between *in situ* and invasive tumour regions. C2 tumours are associated with relatively poor prognostic outcome (Moore *et al.*, 2019).

Thus C1 and C2 tumours are morphologically, biologically and genomically different. On one hand, two populations of tumour cells co-exist in C1 tumours: *in situ* precursor cells with no invasive ability, and an emerging subpopulation of invasive cells that has often gained invasive potential by acquiring genomic changes, accompanied by an increased proliferation rate. On the other hand, C2 tumours are essentially composed of invasive-competent cells that outgrow lepidically and give rise to cytologically high-grade *in situ* cells (Moore *et al.*, 2019). As time passes and the tumour grows in size, I believe that this lepidic zone gradually becomes invasive, perhaps as a consequence of accumulated alterations to the stroma somehow caused by the malignant cells on the surface, and the *in situ* zone transitions outward as malignant cells spread to colonise adjacent normal alveoli.

I were interested in understanding to what extent molecular EMT may contribute to invasiveness in C1 and C2 tumours respectively.
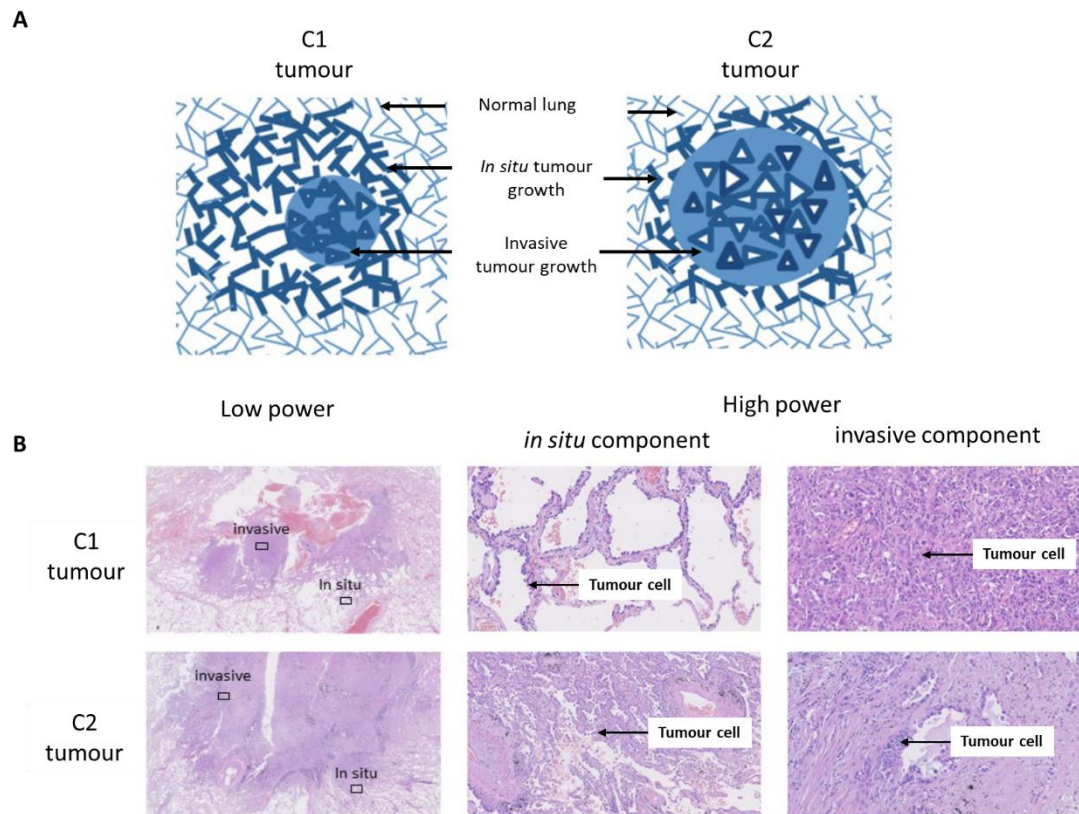
Figure 4.1. C1 and C2 tumours attributes.
A: Cartoon of the morphological details of C1 vs C2 tumours (Moore *et al.*, 2019). B: H&E images of representative C1 and C2 tumours. Nuclei are stained in purple (haematoxylin staining) and cytoplasm is stained in pink (eosin staining). Low-power views show tumour architecture alongside high power regions from the *in situ* and invasive component of both tumours (Moore *et al.*, 2019).

4.3.1.2   Aim and objectives

The aim of this section is to determine whether EMT markers can further demonstrate underlying biology in partially invasive tumours displaying "C1" and "C2" characteristics.

To investigate EMT process as a mechanism enabling the transition to invasiveness, two TMAs were constructed containing 34 C1 and 17 C2 tumours. For each tumour, 2 cores of 1 mm of diameter were taken from both the *in situ* and the invasive areas of the tumour, giving a total of 204 cores in the whole set. Cases for which after examining stained sections the tumour cores contained only *in situ* or invasive patterns were excluded from the analysis, which resulted in 30 C1 and 17 C2 tumour cases (188 cores in total)*.* TMAs were immunohistochemically stained for E-cadherin, cytokeratin, N-

cadherin, vimentin and ISH was performed for *TWIST* mRNA. Digital images of TMA sections were then manually scored for their expression using a H-score system.

4.3.1.3   EMT markers distribution in C1 and C2 tumours

Figures 4.2 to 4.6 illustrate representative images of the expression of EMT markers in *in situ* and invasive components of C1 and C2 tumours.
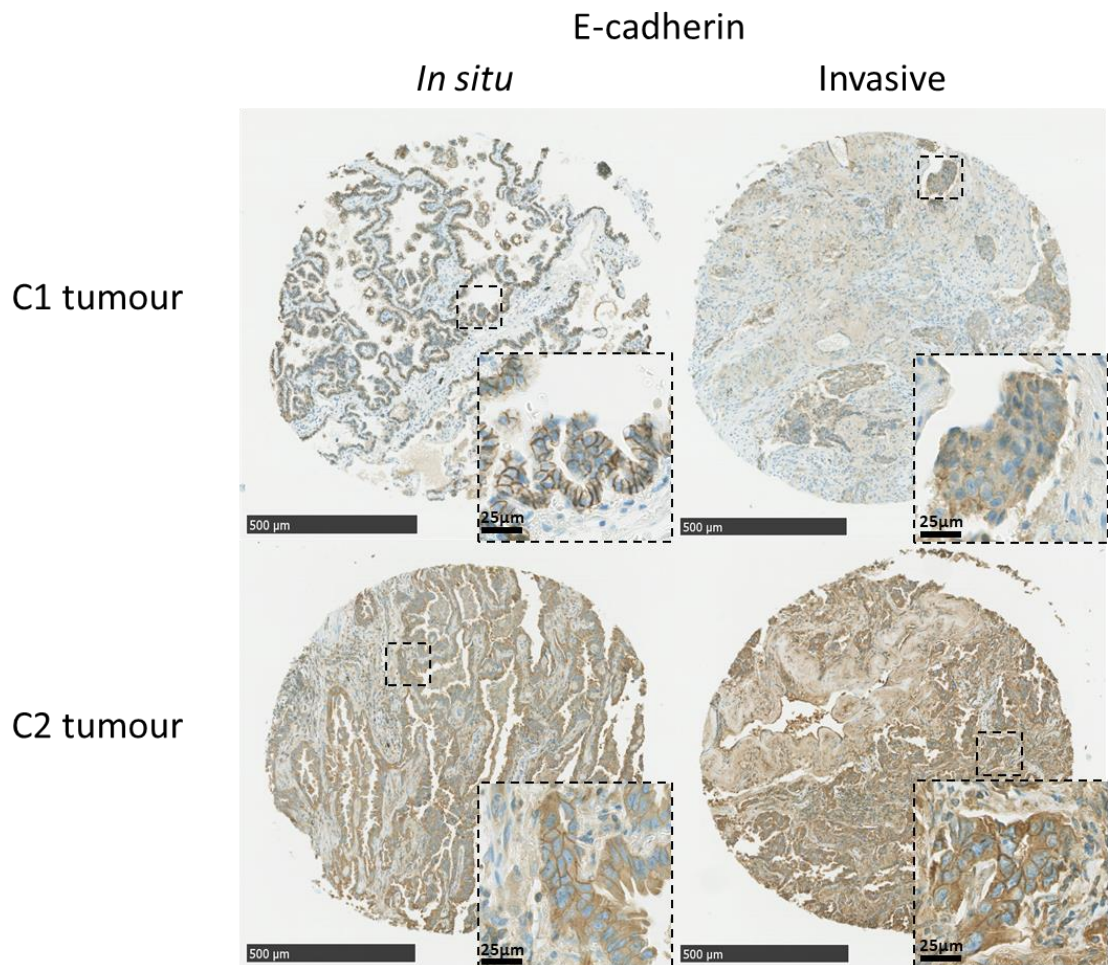
Figure 4.2. Representative immunohistochemical staining of E-cadherin in C1 and C2 tumours.
Core image of E-cadherin expression in C1 and C2 tumours. Top and bottom rows show E-cadherin expression in tumour cells in *in situ* (left column) and invasive (right column) areas of C1 and C2 tumours respectively. Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents E-cadherin expression. Core image magnification = 5X. Scale bar = 500 μm. Inset squares represent high power (80x) of a portion of the core. Scale bar = 25 μm.

In this example, E-cadherin is expressed membranously in the C1 *in situ* tumour area while its expression is mostly cytoplasmic in the invasive component. In the C2 tumour,

E-cadherin appears to be quite strongly expressed in both the cytoplasm and the membrane of both the *in situ* and invasive areas.

As previously stated in chapter 3, functional E-cadherin is a transmembrane protein and so specifically its membranous expression was scored for the purpose of this study.
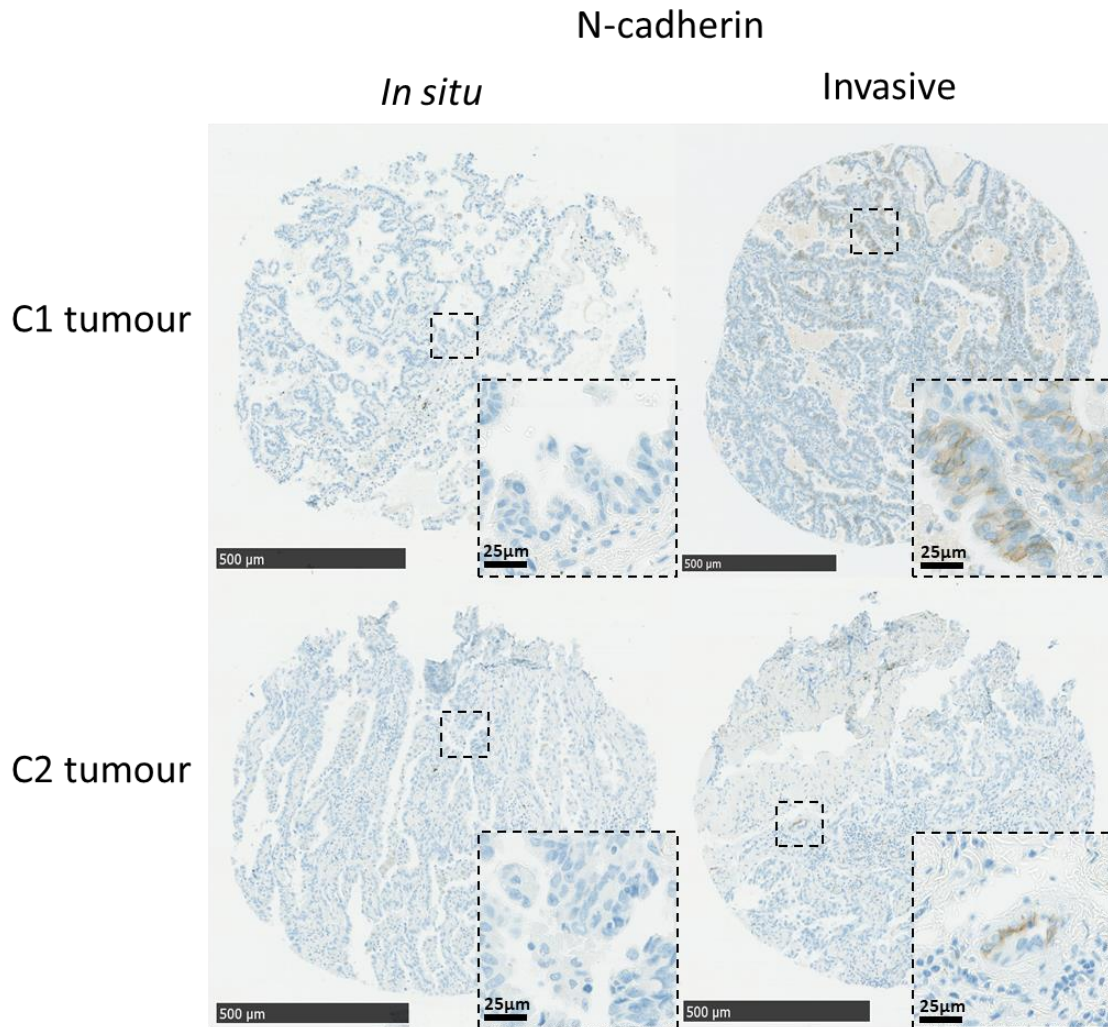
Figure 4.3. Representative immunohistochemical staining of N-cadherin expression in C1 and C2 tumours.
Core image of N-cadherin expression in C1 and C2 tumours. Top and bottom rows show N-cadherin expression in tumour cells in *in situ* (left column) and invasive (right column) areas of C1 and C2 tumours respectively. Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents N-cadherin expression. Core image magnification = 5X. Scale bar = 500 μm. Inset squares represent high power (80x) of a portion of the core. Scale bar = 25 μm.

In the same core, the membranous mesenchymal marker N-cadherin is absent in *in situ* areas. There is multifocal weak expression in the invasive region of the C1 core but none in the C2 core, although expression is detected in endothelial cells.

As for E-cadherin, membranous N-cadherin expression was analysed as it is a transmembrane protein.
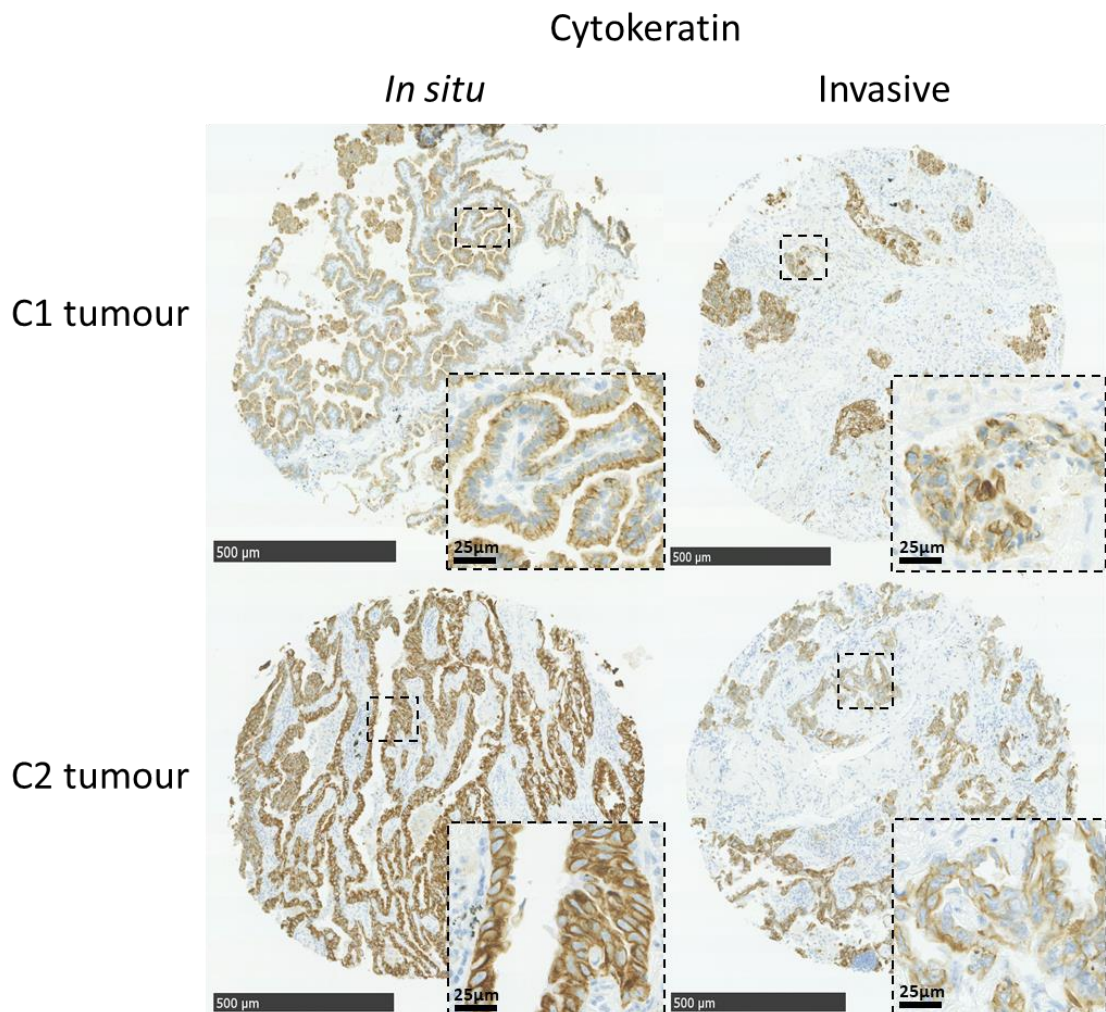
Figure 4.4. Representative immunohistochemical staining of cytokeratin expression in C1 and C2 tumours.
Core image of cytokeratin expression in C1 and C2 tumours. Top and bottom rows show cytokeratin expression in tumour cells in *in situ* (left column) and invasive (right column) areas of C1 and C2 tumours respectively. Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents cytokeratin expression. Core image magnification = 5X. Scale bar = 500 μm. Inset squares represent high power (80x) of a portion of the core. Scale bar = 25 μm.

The cytoskeletal epithelial marker cytokeratin is strongly expressed in *in situ* regions and shows partial loss in the invasive areas of these representative tumours.
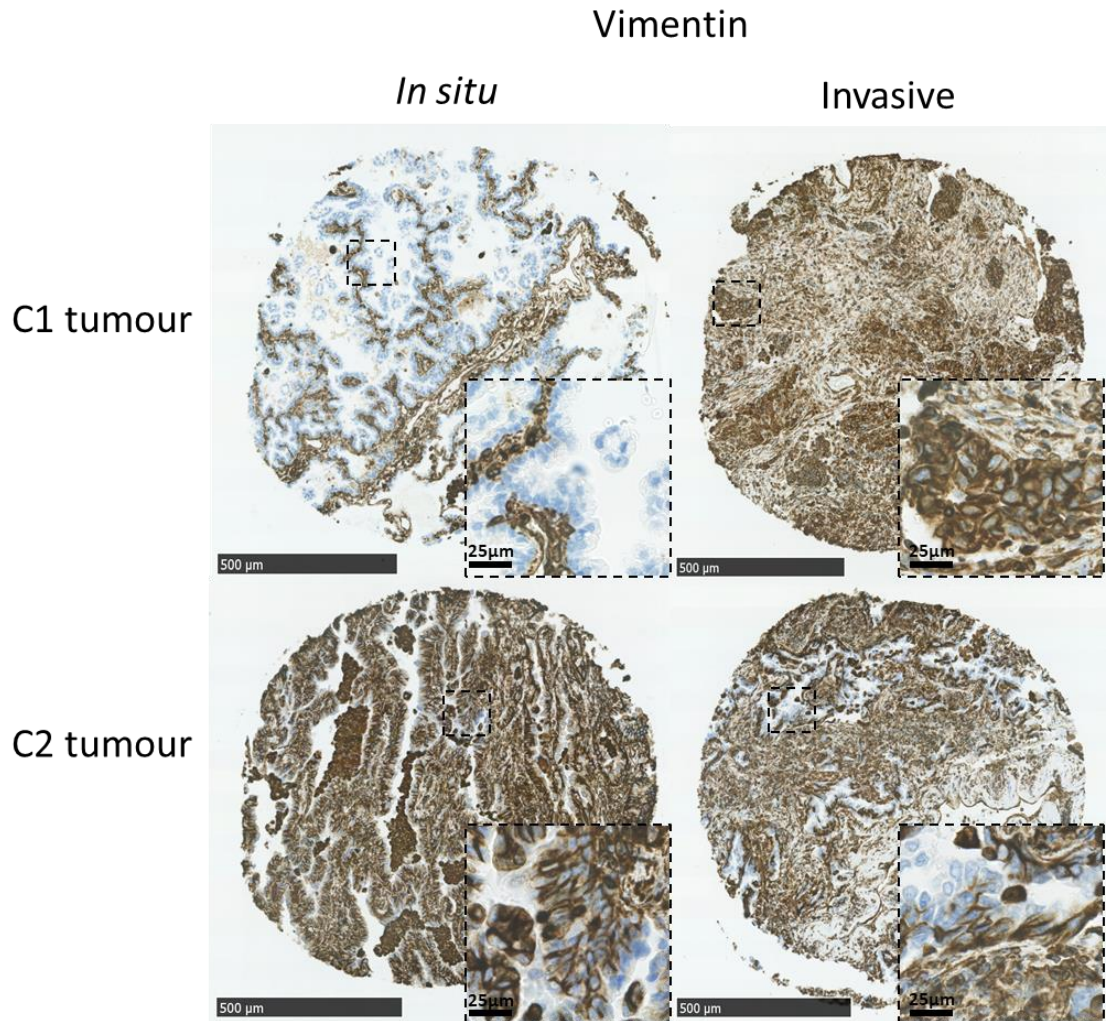
Figure 4.5. Representative immunohistochemical staining of vimentin expression in C1 and C2 tumours. Core image of vimentin expression in C1 and C2 tumours. Top and bottom rows show vimentin expression in tumour cells in *in situ* (left column) and invasive (right column) areas of C1 and C2 tumours respectively. Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents vimentin expression. Core image magnification = 5X. Scale bar = 500 μm. Inset squares represent high power (80x) of a portion of the core. Scale bar = 25 μm.

In the C1 tumour, the mesenchymal marker vimentin is not expressed at all in the *in situ* component of the tumour whereas the invasive area displays strong vimentin expression.

In contrast, vimentin is almost as equally as expressed in *in situ* and invasive zones in the C2 tumour.
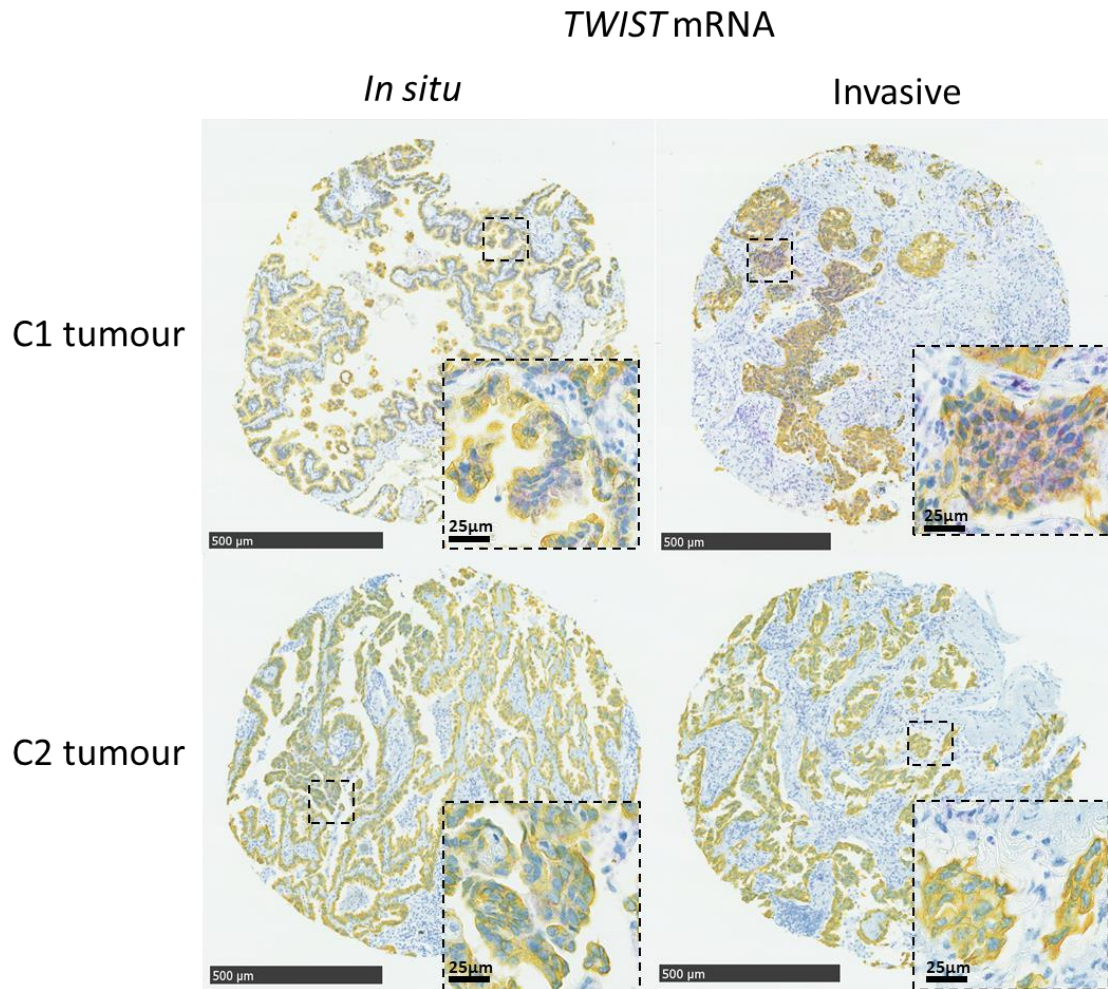
Figure 4.6. Representative immunohistochemical staining of *TWIST* mRNA expression in C1 and C2 tumours.
Core image of *TWIST* mRNA expression in C1 and C2 tumours. Top and bottom rows show *TWIST* mRNA expression in tumour cells in *in situ* (left column) and invasive (right column) areas of C1 and C2 tumours respectively. Blue staining (haematoxylin and bluing reagent) represents nuclei. Purple staining is *TWIST* mRNA staining. Yellow staining represents cytokeratin expression. Core image magnification = 5X. Scale bar = 500 μm. Inset squares represent high power (80x) of a portion of the core. Scale bar = 25 μm.

TWIST mRNA shows patchy moderately strong expression in the invasive area of the C1 tumour with weak focal expression in the *in situ* region of the C1 tumour. Expression is not detected in this C2 tumour.

I next determined the distribution of expression levels of each EMT marker in the entire set of C1 and C2 tumours.

4.3.1.4 Partial molecular EMT is associated with the transition to invasiveness in C1 tumours in primary lung adenocarcinoma

Figures 4.7 to 4.11 represent the distribution of EMT markers in *in situ* and invasive components of C1 and C2 tumours (box plots). The slope graphs illustrate matched paired *in situ* and invasive cores per patient. The medians of the two *in situ* cores and the two invasive cores were calculated for each EMT markers. To compare the difference in EMT markers' expression between C1 and C2 tumours and *in situ* and invasive components of the same tumour, Mann-Whitney and Wilcoxon signed-rank tests were performed respectively.

E-cadherin expression is statistically somewhat lower in C2 compared to C1 tumours, consistent with a more invasive phenotype overall in C2 tumours (p = 0.013) (Fig 4.7A). However, E-cadherin expression does not significantly differ from the *in situ* to invasive compartments in either C1 or C2 tumours.
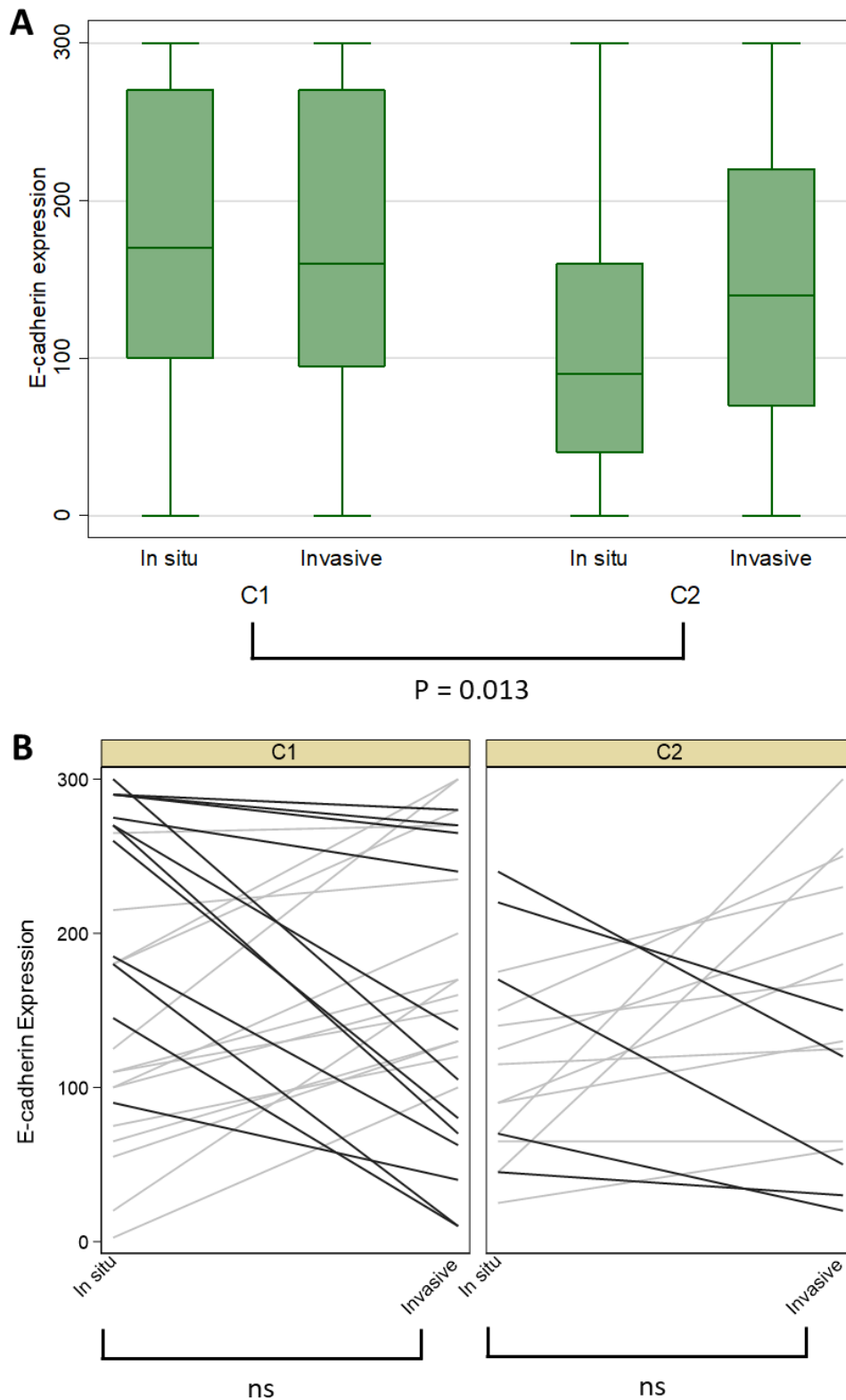
Figure 4.7. Distrubution of E-cadherin expression in C1 and C2 tumours.
A: Box plot of E-cadherin expression in *in situ* and invasive compartments of C1 and C2 tumours at the core level, (n = 188 cores). P = p-value of Mann-Whitney test. B: Slope graphs representing E-cadherin expression in *in situ* and invasive compartments of C1 and C2 tumours at the case level, (n = 47 cases). Wilcoxon signed-rank test was performed: ns = not significant.

N-cadherin is generally indetectable in C1 and C2 tumours (Figure 4.8A), although surprisingly the expression is significantly higher overall in C1 tumours (p = 0.003), although n is small for C2 cases (n = 15). The invasive region shows an increase in expression of N-cadherin compared to the *in situ* zone in C1, showing near nominal significance (p = 0.055). There is no statistical difference in N-cadherin expression between the *in situ* and invasive region of C2 tumours (Figure 4.8B).

This augmentation of N-cadherin expression is suggestive of a gain of invasiveness in C1 tumours which may occur upon activation of partial EMT.
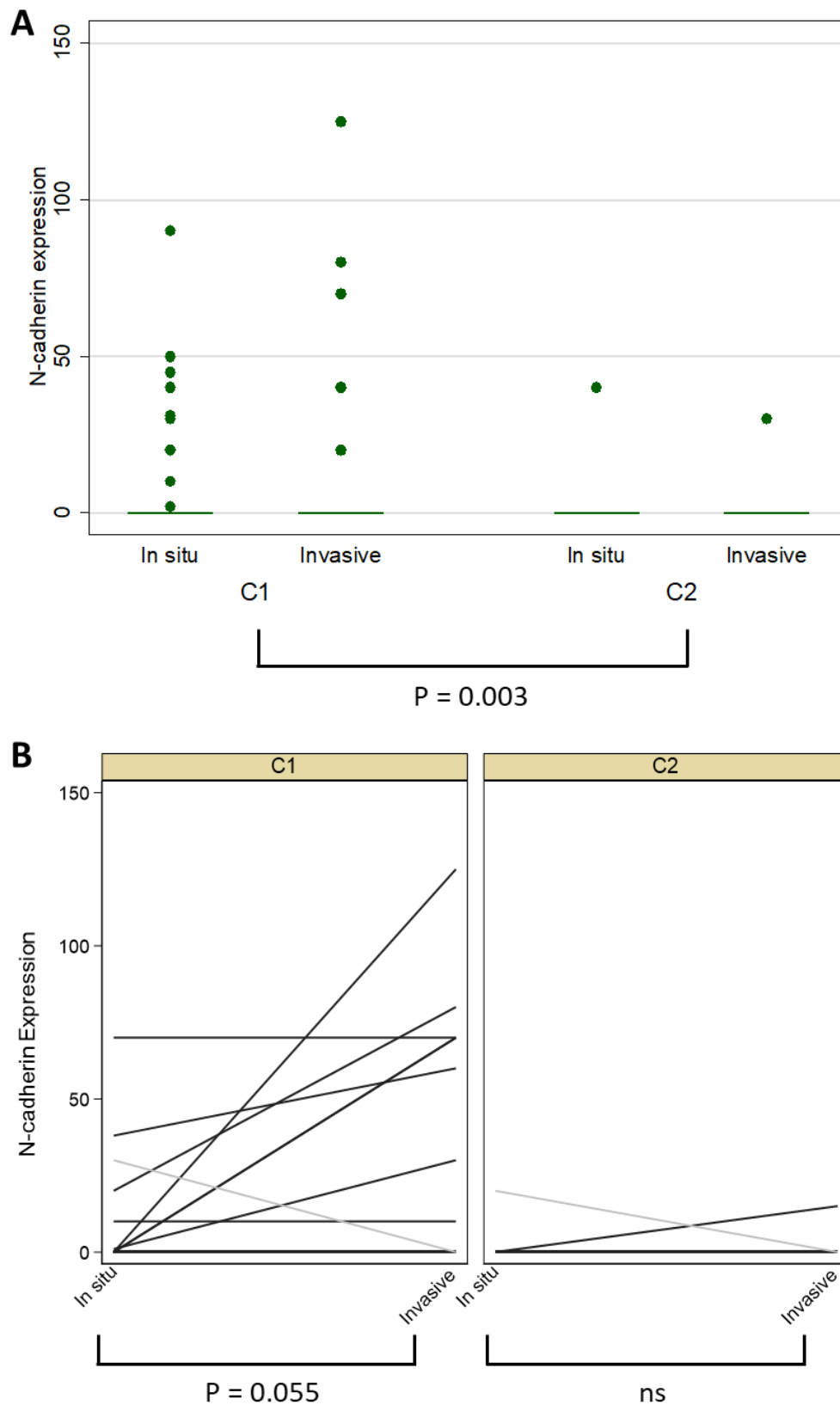
Figure 4.8. Distribution of N-cadherin expression in C1 and C2 tumours.
A: Box plot of N-cadherin expression in *in situ* and invasive compartments of C1 and C2 tumours at the core level, (n = 188 cores). P = p-value of Mann-Whitney test. B: Slope graphs representing N-cadherin expression in *in situ* and invasive compartments of C1 and C2 tumours at the case level, (n = 45 cases). Wilcoxon signed-rank test was performed: P = p-value ns = not significant.

As shown in Figure 4.9A, the epithelial marker cytokeratin AE1/3 is significantly more strongly expressed in C2 compared to C1 tumours overall (p = 0.029) (Figure 4.9A), but there are no significant differences between *in situ* and invasive regions in either tumour type (Figure 4.9B).
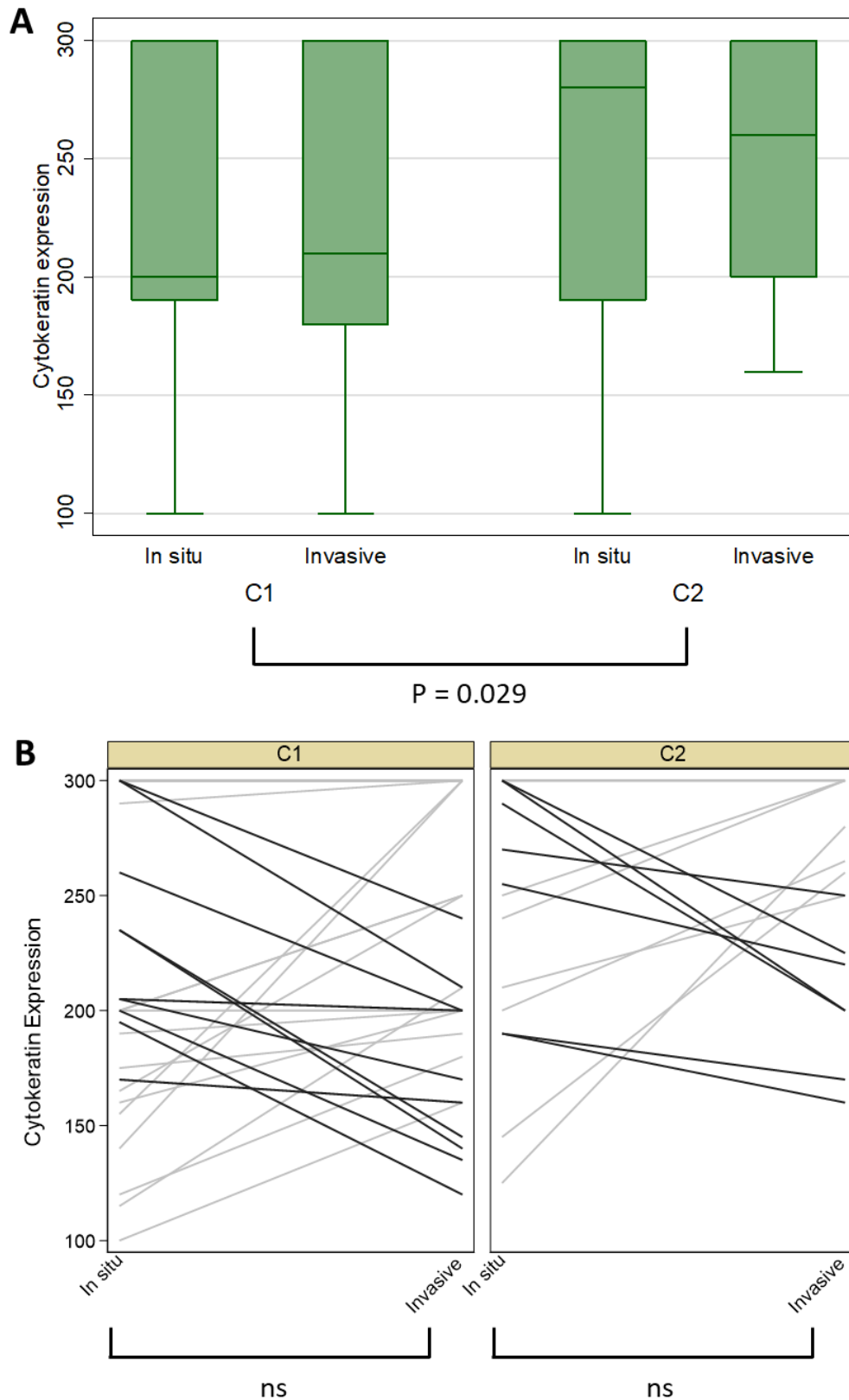
Figure 4.9. Distribution of cytokeratin expression in C1 and C2 tumours.
A: Box plot of cytokeratin expression in *in situ* and invasive compartments of C1 and C2 tumours at the core level, (n = 188 cores). P = p-value of Mann-Whitney test. B: Slope graphs representing cytokeratin expression in *in situ* and invasive compartments of C1 and C2 tumours per patient, (n = 47 cases). Wilcoxon signed-rank test was performed: ns = not significant.

The mesenchymal marker vimentin is statistically significantly higher in C2 as compared to C1 tumours (Figure 4.10A). Furthermore, it is markedly elevated in the transition from the *in situ* to invasive component in C1 tumours (p = 0.001) while there is no statistically significant difference in vimentin expression between the *in situ* and invasive regions of C2 tumours (Figure 4.10B).
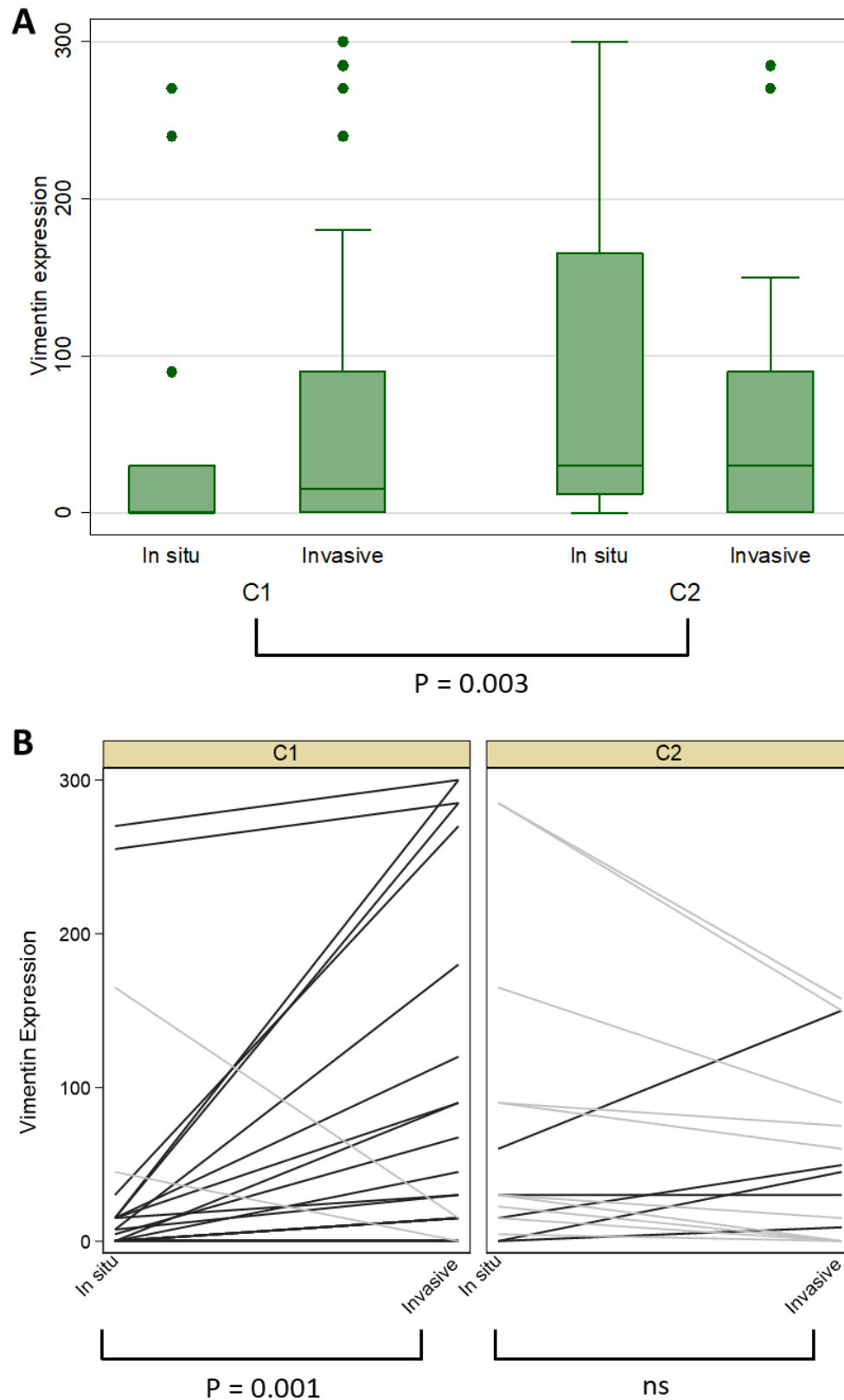
Figure 4.10. Distribution of vimentin expression in the C1 and C2 tumours.
A: Box plot of vimentin expression in *in situ* and invasive compartments of C1 and C2 tumours at the core level, (n = 188 cores). P = p-value of Mann-Whitney test. B: Slope graphs representing vimentin expression in in situ and invasive compartments of C1 and C2 tumours per patient, (n = 47 cases). Wilcoxon signed-rank test was performed: P = p-value, ns = not significant.

Finally, the EMT transcription factor TWIST is expressed at similar low levels in C1 and C2 tumours and there is no statistically significant difference between these groups (Figure 4.11A). Also there is no statistical difference in *TWIST* mRNA expression between the *in situ* and invasive regions of C1 tumours. There is a nominally significant but small relative decrease in *TWIST* mRNA expression in invasive as compared to *in situ* areas in C2 tumours (p = 0.030) (Figure 4.11A and B). This is of uncertain biological significance.

On balance these data suggest that C1 tumours, which frequently acquire genomic changes in p53 and/or proto-oncogenes in invasive areas, show similar changes in expression of EMT markers, as both N-cadherin and vimentin expression are increased in the invasive areas. Our interpretation is that this probably represents a partial EMT programme associated with genomic progression from *in situ* to invasive growth. Even though the cells remain epithelioid in appearance, they have gained invasive properties as well as expression of vimentin and a trend toward increased expression of N-cadherin. Crucially, this suggests that partial EMT, as shown by molecular changes and revealed by immunohistochemistry, may be sufficient to drive invasive behaviour without complete mesenchymal differentiation. Furthermore, this partial EMT is likely, at least sometimes, to be related to subclonal genomic alterations.

In contrast, C2 tumours, where peripheral lepidic growth is believed to arise from the outgrowth of invasion-competent cells, do not demonstrate significant alteration in immunohistochemical markers of invasiveness in invasive vs *in situ* areas. Surprisingly ISH measurement of *TWIST* mRNA expression reveals a significant small relative upregulation in *in situ* regions of C2 tumours, which might imply that EMT-related motility is involved in the expansion of these tumours along the alveolar surfaces. However this is not confirmed by the EMT protein markers. Further investigations would be necessary to clarify this possibility.
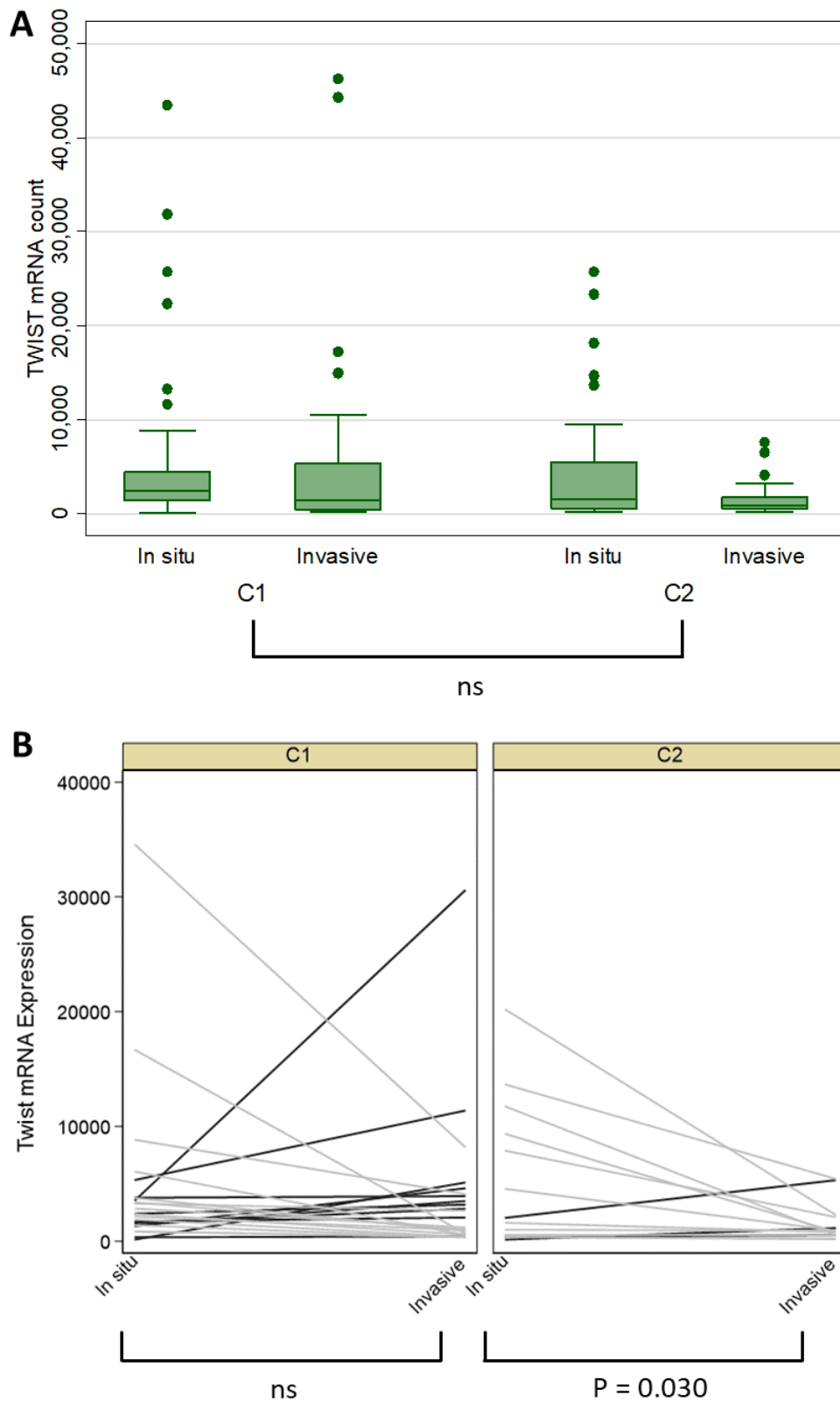
Figure 4.11. Distribution of *TWIST* mRNA expression in the C1 and C2 tumours.
A: Box plot of *TWIST* mRNA expression in *in situ* and invasive compartments of C1 and C2 tumours at the core level, (n = 188 cores). , Mann-Whitney test was performed respectively ns = not sgnificant. B: Slope graphs representing vimentin expression in *in situ* and invasive compartments of C1 and C2 tumours at the case level, (n = 47 cases). Wilcoxon signed-rank test was performed respectively: P = p-value, ns = not significant.

4.3.2 Morphologically *in situ* tumour growth can contain microscopic foci of invasion accompanied by partial EMT

4.3.2.1 Introduction

As discussed in chapter 3, evidence of molecular EMT can often be detected in areas of lepidic/*in situ* growth pattern (Figures 3.20-3.24 and 3.25B). This observation is somewhat challenging, as it suggests either that partial EMT is occurring without invasion, or that invasion may occur subtly in areas of apparently *in situ* tumour growth. Therefore I wanted to explore EMT in *in situ* lung adenocarcinoma.

4.3.2.2 Aim and objective

To further investigate the role of EMT in early invasion in *in situ* growth pattern, I looked at whole tumour sections from 18 tumours containing predominantly *in situ* growth. I selected two lepidic predominant tumour pattern sections with focal invasive areas and immunohistochemically stained the whole section for N-cadherin, E-cadherin, cytokeratin and vimentin. Between the two predominantly *in situ* tumour sections, only one showed molecular evidence of EMT and this tumour is illustrated in this section.

4.3.2.3 Results

The haemotoxylin/eosin staining of lepidic tumour pattern shows low-grade carcinoma of apparently *in situ* growth pattern, with mostly well preserved alveolar architecture with thickened and fibrotic interstitium (Figure 4.12).

However, close examination reveals that multiple foci of N-cadherin positivity are observed in *in situ* tumour whereas it is not present in normal lung (Figure 4.14). However, all the other EMT markers did not show any molecular evidence of EMT (data not shown).

Interestingly, N-cadherin positive cells are generally situated at the tips of 'promontories', microscopic papillary structures covered in transformed cells. Furthermore, they are generally paired, with N-cadherin-positive promontory tips either side of a gap in the alveolar wall (Figure 4.13).

My interpretation is that these gaps may represent very early instances of invasiveness in this tumour, where epithelial cells locally are seen to be acquiring invasive properties at the same time as demonstrating partial molecular EMT.

Furthermore, invasive areas of this predominantly *in situ* tumour reveal N-cadherin positive cells, showing that N-cadherin positivity does indeed signify invasiveness in this tumour (Figures 4.15-4.16).

Together, these findings suggest a possible pattern of microscopic multifocal invasion in areas of morphologically *in situ* tumour growth.
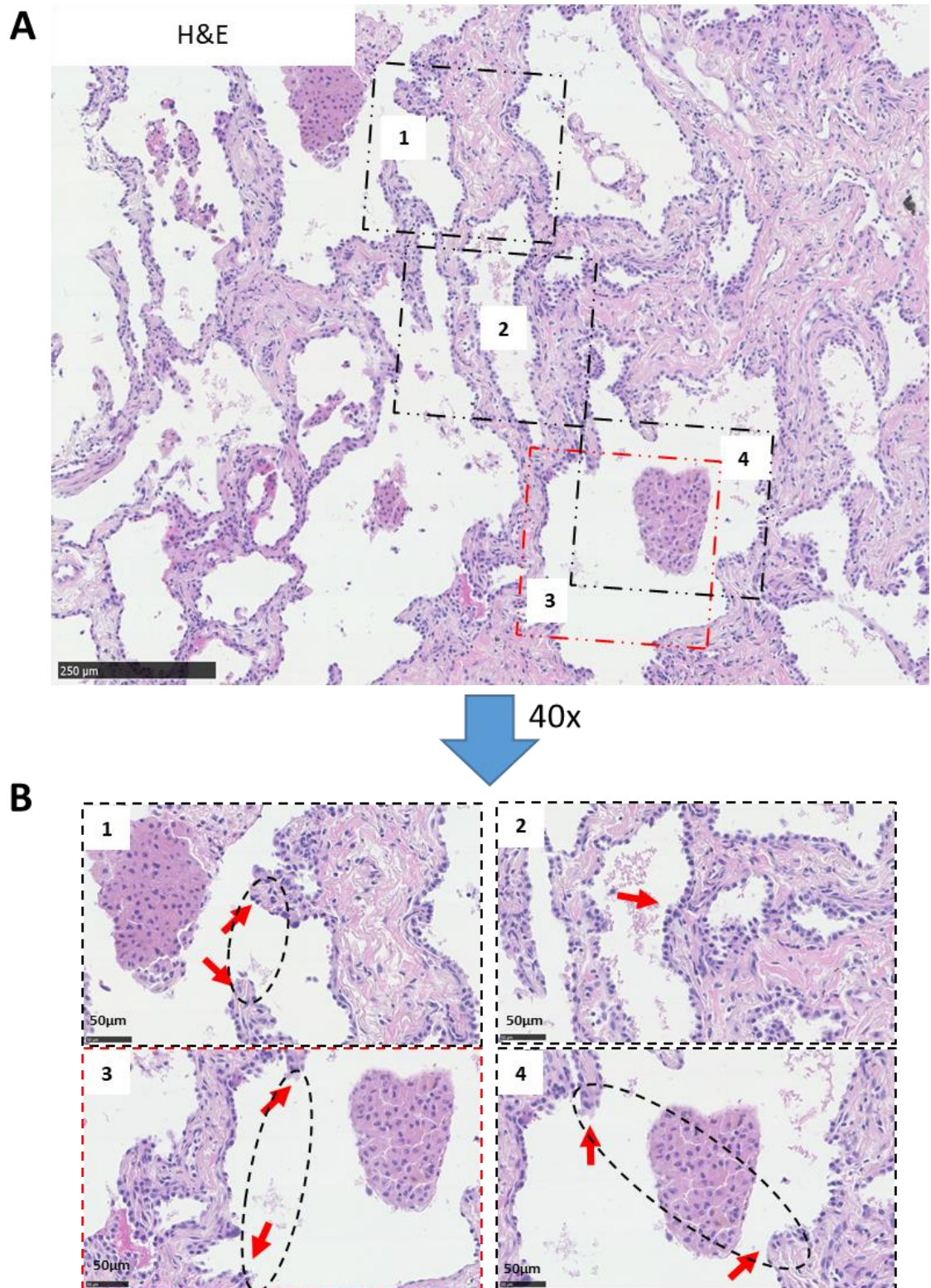
Figure 4.12. Images showing evidence of molecular EMT accompanied by promontory formation in areas of *in situ* tumour in minimally invasive lung adenocarcinoma.

A: H&E staining of *in situ* tumour growth. Red rectangles show the promotory regions in lepidic tumour growth. Magnification = 10X. Scale bar = 250 μm. B: High magnification (40x) of molecular evidence of EMT in *in situ* tumour growth. Red arrows point out promotory regions in the *in situ* growth pattern and dotted circle represents the gap in the alveolar wall. Scale bar = 50 μm. Nuclei are stained in purple (haematoxylin staining) and cytoplasm is stained in pink (eosin staining).
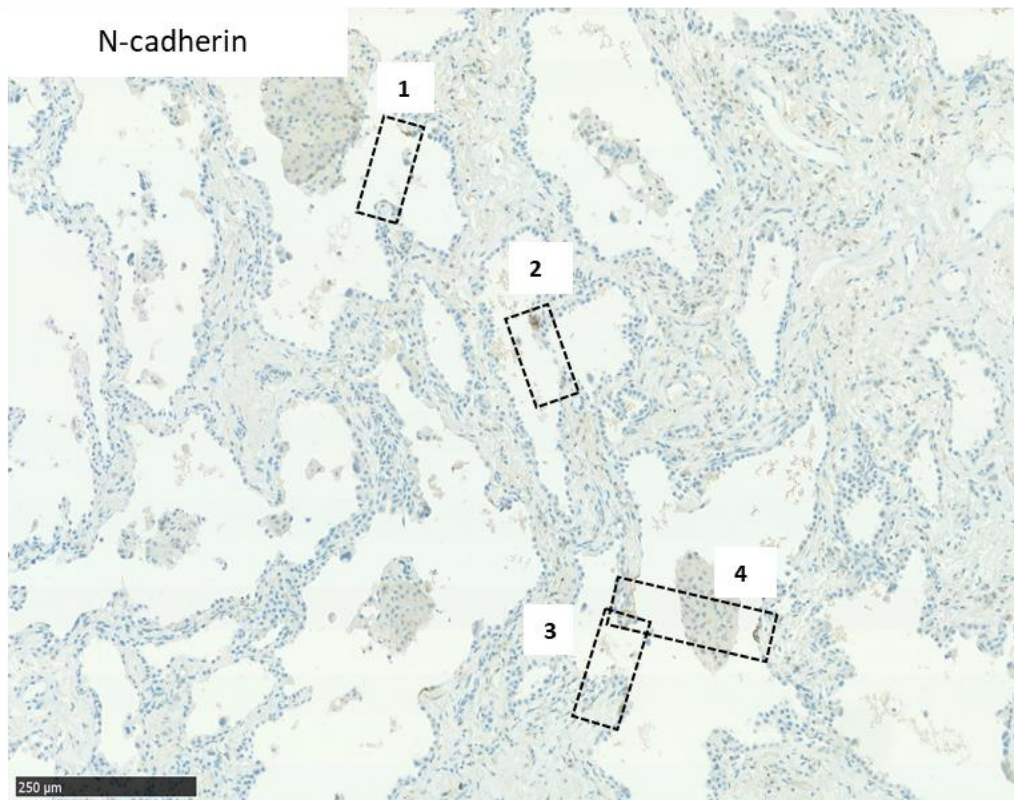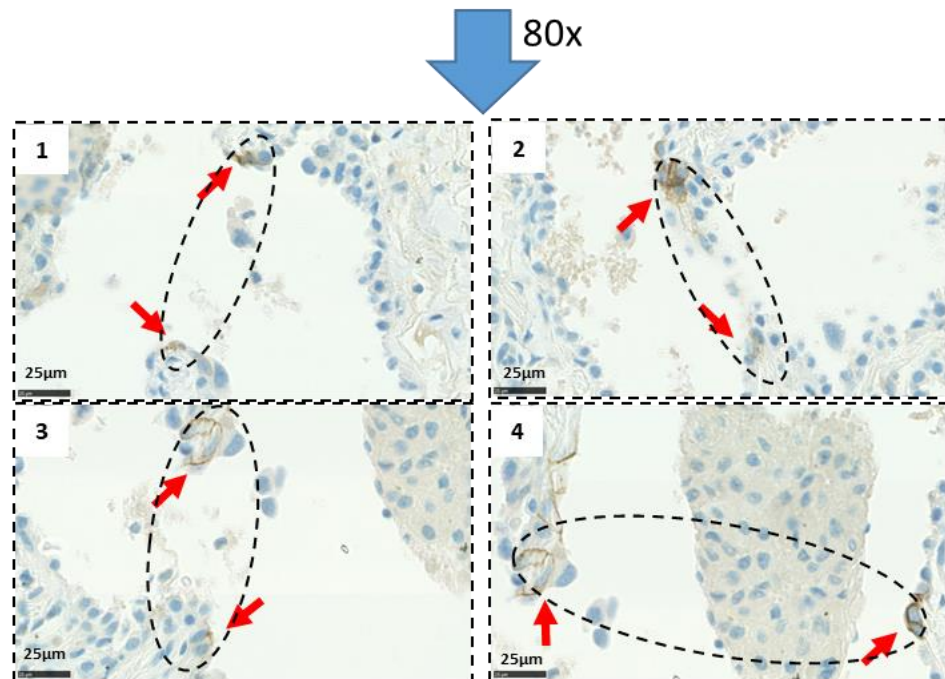
Figure 4.13. Images showing evidence of molecular EMT accompanied by promontory formation in areas of _in situ_ tumour in minimally invasive lung adenocarcinoma.

A: Immunohistochemical staining of N-cadherin on _in situ_ tumour growth (black rectangles show the promotory regions in lepidic tumour growth). Nuclei are stained in blue (haematoxylin and bluing reagent) and membranous brown staining is DAB and represents N-cadherin expression. Magnification = 10X. Scale bar = 250 µm. B: High magnification (80x) of molecular evidence of EMT in _in situ_ tumour growth. Scale = 25 µm. Dotted circle represents the gap in the alveolar wall. Red arrows point out N-cadheirn positive cells.
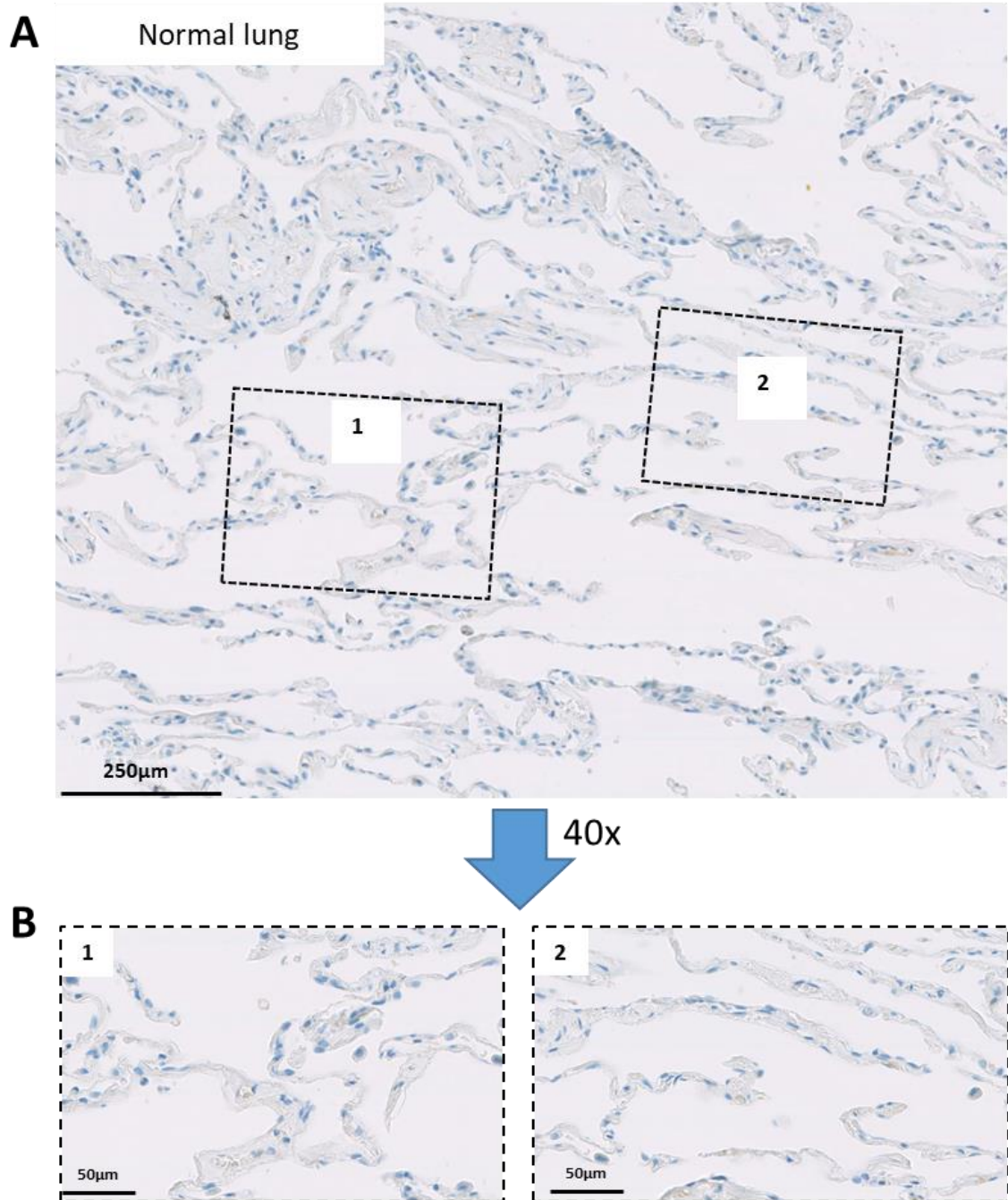
Figure 4.14. Images showing evidence of molecular EMT accompanied by promontory formation in areas of *in situ* tumour in minimally invasive lung adenocarcinoma.

A Immunohistochemical staining of N-cadherin on normal lung with no N-cadherin positive promontories. Magnification = 10X. Scale bar = 250 µm. Nucleui are stained in blue (haematoxylin and bluing reagent). Magnification = 10X. Scale bar = 250 µm. B: High magnification (40x) of no molecular evidence of EMT in normal lung. Scale bar = 50 µm.
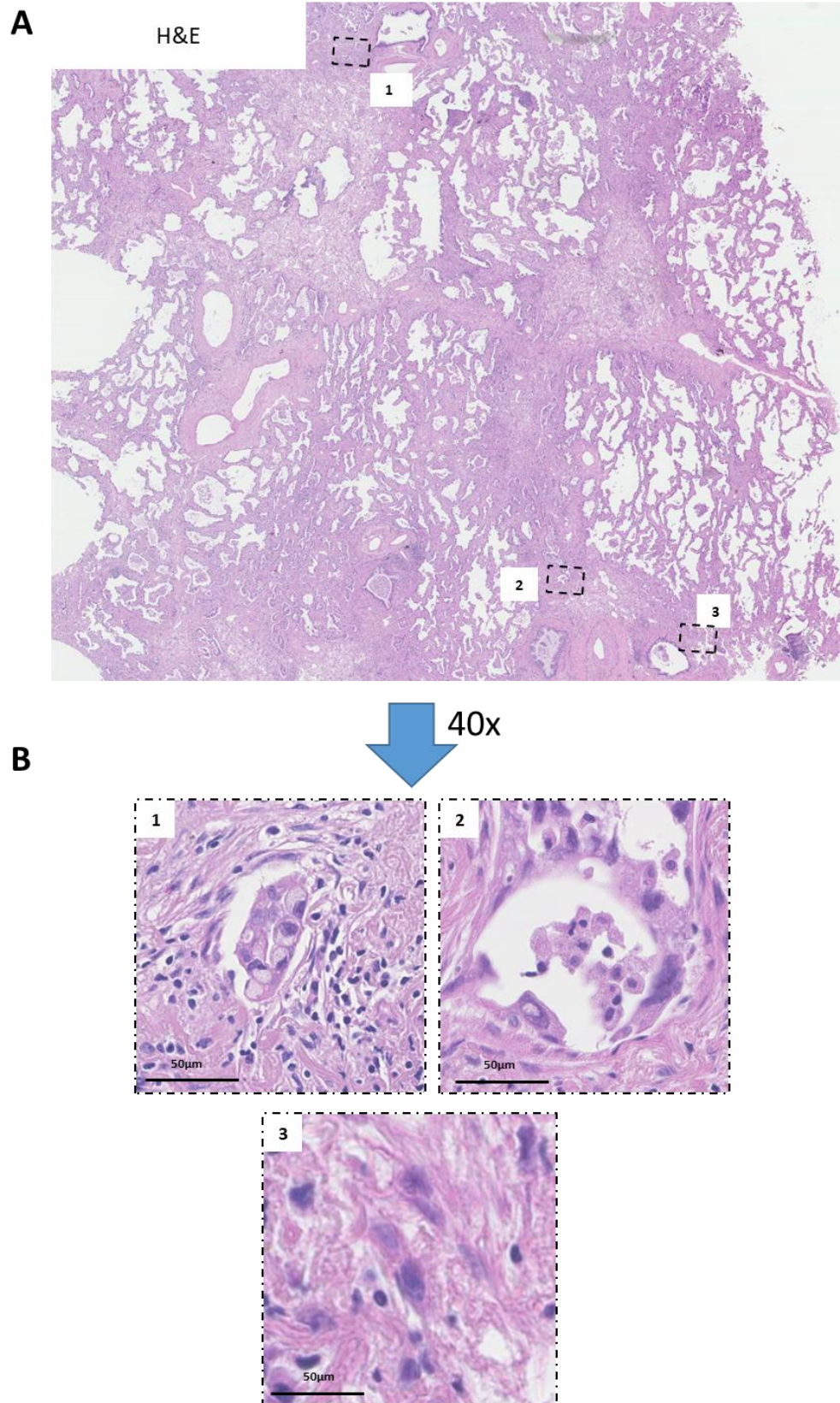
Figure 4.15. H&E of images showing evidence of molecular EMT in areas of invasive tumour in minimally invasive lung adenocarcinoma.
A: H&E staining of invasive tumour growth. Magnification = 1.5X. Scale bar = 1 mm. B: High magnification (40x) of molecular evidence of EMT in invasive tumour growth. Scale bar = 50 μm. Nuclei are stained in purple (haematoxylin staining) and cytoplasm is stained in pink (eosin staining).
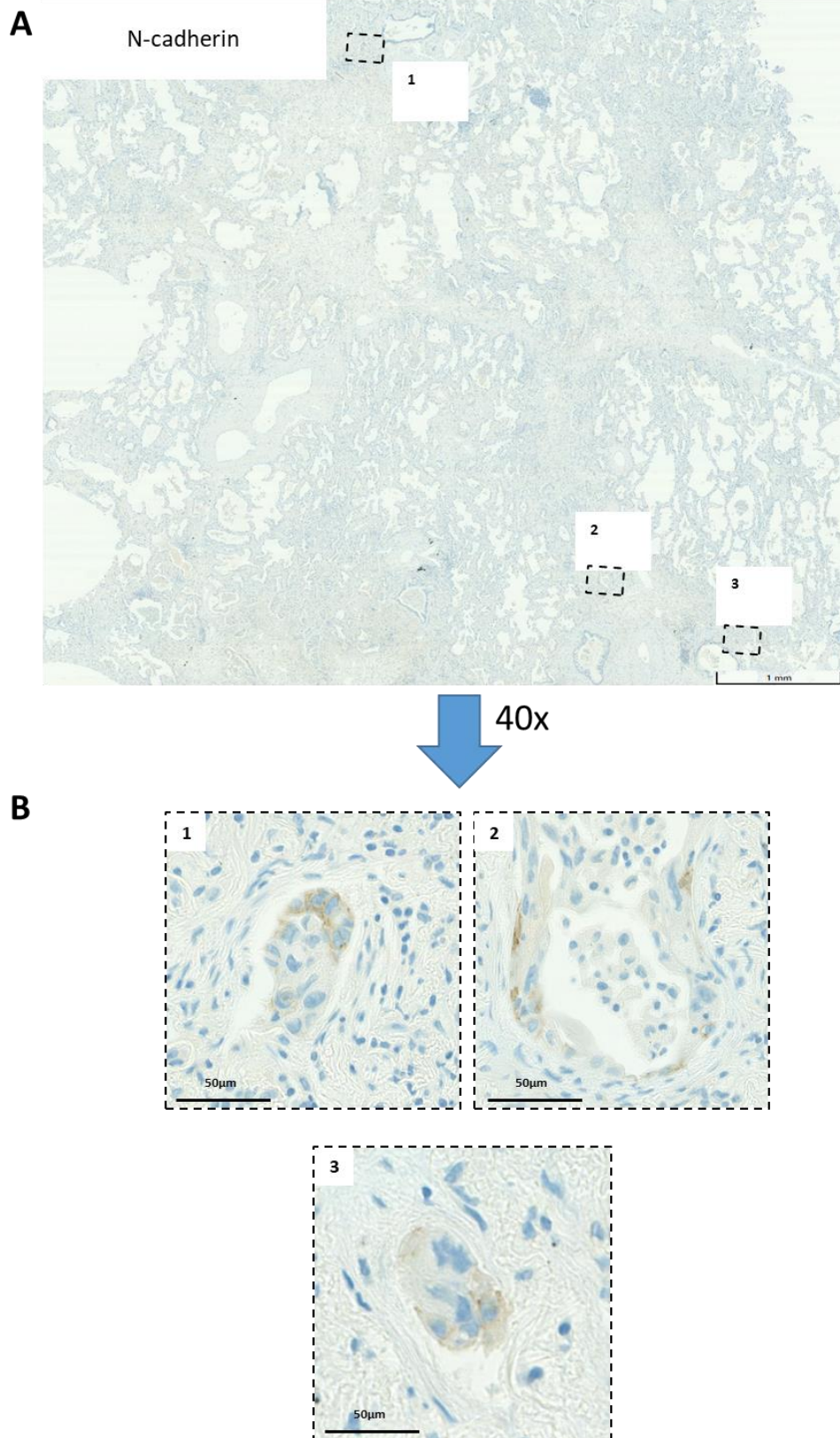
Figure 4.16. Images showing evidence of molecular EMT in areas of invasive tumour in minimally invasive lung adenocarcinoma.

A: Immunohistochemical staining of N-cadherin on invasive tumour growth. Magnification = 1.5X. Scale bar = 1 mm. B: High magnification (40x) of molecular evidence of EMT in invasive tumour growth with N-cadherin positive cells. Scale bar = 50 μm. Nuclei are stained in blue (haematoxylin and bluing reagent) and brown staining is DAB representing N-cadherin expression.

4.3.3 Mesenchymal transformation in pleomorphic carcinoma is associated with molecular and morphological EMT

4.3.3.1 Introduction

Pleomorphic carcinoma, a subgroup of sarcomatoid carcinoma, is a rare highly malignant subtype of non–small cell lung carcinomas. This tumour growth is composed of undifferentiated or poorly differentiated non-small cell carcinoma, i.e squamous cell carcinoma or adenocarcinoma, in conjunction with a mixture of at least 10% of spindled or giant cells (Figure 4.14).
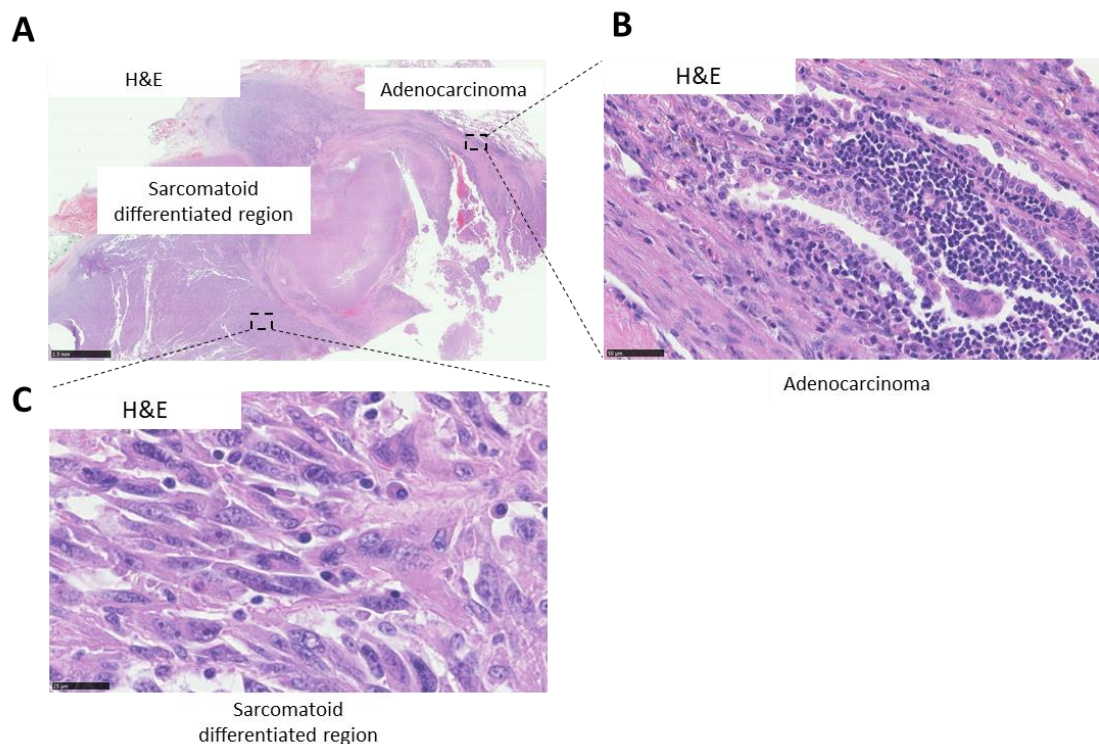
Figure 4.17. H&E images of a pleormorphic carcinoma tumour displaying a sarcomatoid differentiated zone and adenocarcinoma area.
A: H&E image of whole pleomorphic carcinoma tumour tissue comprising adenocarcinoma and sarcomatoid differentiated regions. Magnification = 0.83X. Scale bar = 2.5 mm. B: H&E image of the high resolution (40x) of the adenocarcinoma component of the pleomorphic carcinoma tumour. Scale bar = 50 μm. C: H&E image of the high resolution (80x)of the sarcomatoid differentiated region of the pleomorphic carcinoma tumour. Scale bar = 25 μm.Nuclei are stained in purple (haematoxylin staining) and cytoplasm is stained in pink (eosin staining).

It has been suggested that this type of tumour growth arises from adenocarcinoma elements that have dedifferentiated into mesenchymal phenotype through an EMT

program (Pelosi *et al.*, 2009). I aimed to investigate this, testing if the dedifferentiated area gains EMT markers compared to the adenocarcinoma component.

4.3.3.2   Aim and objectives

The aim of this sub-study was to answer the following question:

Does increasing molecular EMT accompany the spindling of cells cells observed in pleomorphic carcinoma?

To answer this question, five pleomorphic carcinoma with adenocarcinoma component (PCAC) tumours were stained for E-cadherin, cytokeratin, N-cadherin and vimentin. The tumour illustrated below is a representative tumour.

4.3.3.3   Sarcomatoid differentiated region shows molecular and morphological evidence of EMT

In the adenocarcinoma region of pleomorphic carcinoma, illustated by the acinar tumour growth, tumour cells express both epithelial markers E-cadherin and cytokeratin and are negative for the mesenchymal markers N-cadherin and vimentin. In this case epithelial/adenocarcinoma component of pleomorphic carcinoma does not show any molecular evidence of EMT (Figure 4.15A).

The sarcomatoid differentiated component of the pleomorphic carcinoma, illustrated by spindle-shaped cells, exhibits on the contrary a loss of membranous E-cadherin and cytokeratin expression and an increase of the mesenchymal markers N-cadherin and vimentin. This clearly demonstrates that the sarcomatoid area of this pleomorphic carcinoma exhibits molecular evidence of the epithelial-mesenchymal transition in concomittance with morphological EMT (Figure 4.15B).

These observations indicate that the sarcomatoid region of the PCAC undergoes a complete EMT (i.e. molecular and morphological changes) while the adenocarcinoma component does not. This further validates our EMT markers and additionally supports that sarcomatoid transformation is the consequence of an EMT programme.
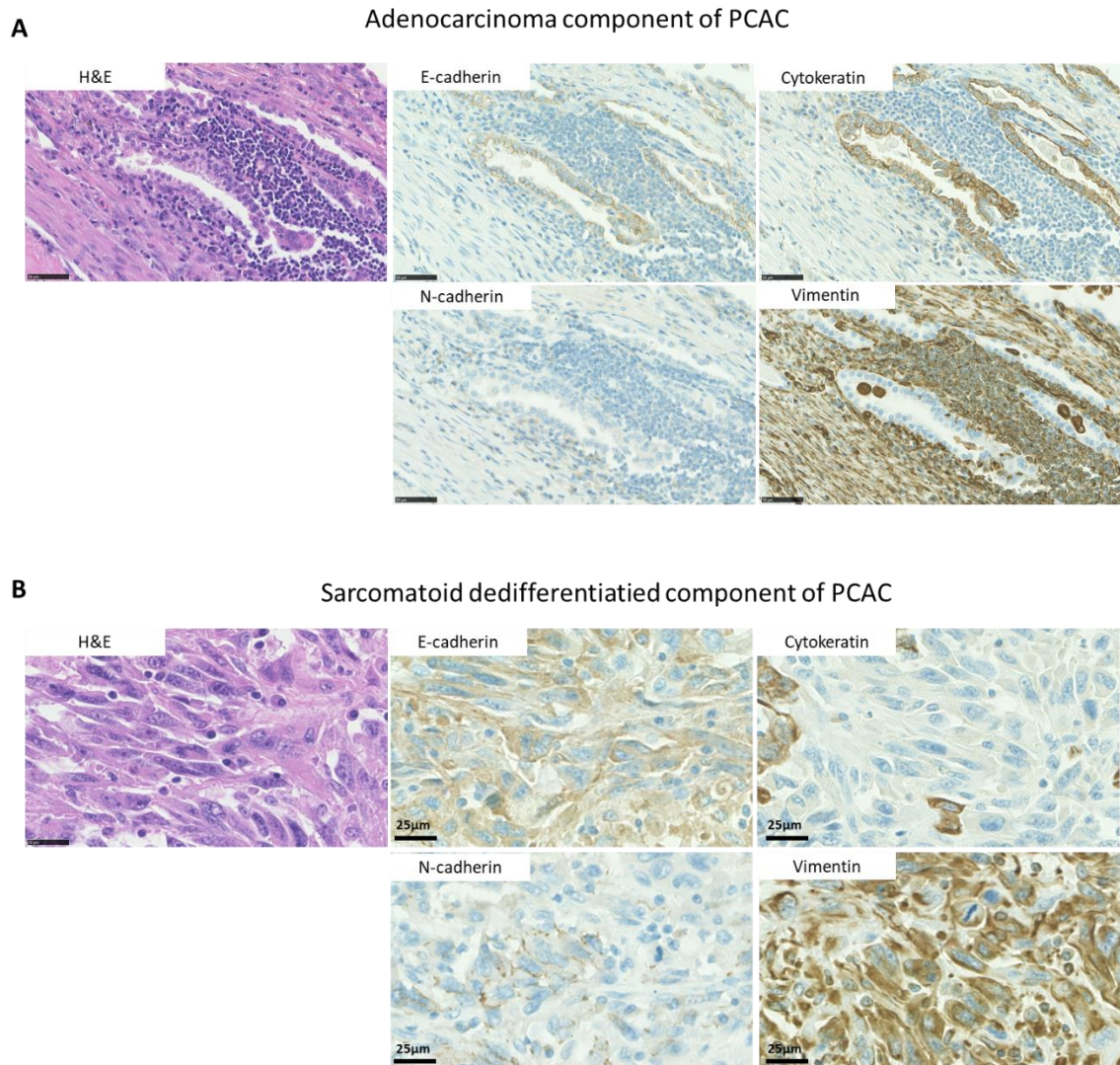
Figure 4.18. Representative pleomorphic carcinoma with adenocarcinoma component tumour.
A: H&E and EMT markers stained images of the adenocarcinoma component of the pleomorphic carcinoma tumour. Nuclei are stained in purple (haematoxylin and bluing reagent) and cytoplasm is stained in pink (eosin) in H&E images. Brown staining represents DAB and blue staining (haematoxylin and bluing reagent) represents nuclei in EMT makers stained images. Magnification = 40X. Scale bar = 50 µm. B: Images of H&E and immunohistochemical staining for EMT markers of sarcomatoid differentiated region of the pleomorphic carcinoma tumour. Nuclei are stained in purple (haematoxylin and bluing reagent) and cytoplasm is stained in pink (eosin). Magnification = 80X. Scale bar = 25 µm.

4.3.4 Micropapillary tumour growth pattern is morphologically and molecularly heterogenous

4.3.4.1 Introduction

Micropapillary tumour growth is a histologic subtype of lung adenocarcinoma. In 2011, this growth pattern was introduced in the lung adenocarcinoma classification in addition to lepidic, acinar, papillary and solid tumour growth by the International Association for the Study of Lung Cancer (IASLC), American Thoracic Society (ATS) and European Respiratory Society (ERS) (Travis *et al.*, 2011). According to the IASLC/ATS/ERS, the micropapillary histological pattern is morphologically defined as a growth of carcinoma in papillary "tufts" which may be seen attached and/or detached from the alveolar wall in 2-dimensional sections and which are bereft of fibrovascular cores. Tumour cells composing the micropapillary structure are small cuboid epithelial cells. Micropapillary adenocarcinoma, when present as either a predominant or minority pattern in adenocarcinoma, is associated with poor outcome (chapter 3, Figures 3.2-3.4) (Travis *et al.*, 2011; Yoshizawa *et al.*, 2011) and has been shown to be significantly related to lymph node metastasis and visceral pleural invasion (Miyoshi *et al.*, 2003). Therefore micropapillary patterns are regarded as high-grade tumour growth in lung adenocarcinoma.

Recently, Emoto *et al.* have identified a new pattern associated with micropapillary growth, called filigree micropapillary pattern. They described this pattern as narrow and columnar tumour cells growth attached to the sessile part of the tumour without fibrovascular cores. They investigated the frequencies of the filigree pattern in 1468 stage I invasive adenocarcinoma and discovered the presence of this pattern in 16 acinar, 37 papillary and 4 solid tumour growth. They then decided to reclassify those 57 tumours as filigree micropapillary in addition to the 87 classical micropapillary tumours to obtain a total of 144 micropapillary predominant adenocarcinoma (MPPA). They subsequently identified 78 filigree predominant micropapillary tumours within the 144 MPPA tumours and demonstrated that the filigree predominant micropapillary pattern significantly correlated with higher pathologic stage, pleural and lympho-vascular invasion and early recurrence of the disease (Emoto *et al.*, 2019).

This suggests a continuum of morphology between flat tumour cells usually constituting the acinar/papillary structure that I labelled "sessile" tumour cells, which can sometimes grow in tufted "filigree" pattern, i.e. short micropapillae, which can when sufficiently long and seen in cross-section give rise to the clusters of tumour cells observed in 2D sections called micropapillae (Figure 4.19).

When exploring EMT in primary lung adenocarcinoma in chapter 3 (Figures 3.20-3.24), cores with micropapillary growth pattern did mostly not show molecular evidence of EMT, and retained expression of epithelial markers. Interestingly the one published study in this area (Nakashima *et al.*, 2015) which demonstrated the upregulation of vimentin expression in micropapillae tumours in lung adenocarcinoma. Therefore I decided to investigate this inconsistency and to further examine the role of molecular EMT in the development of this type of tumour growth.
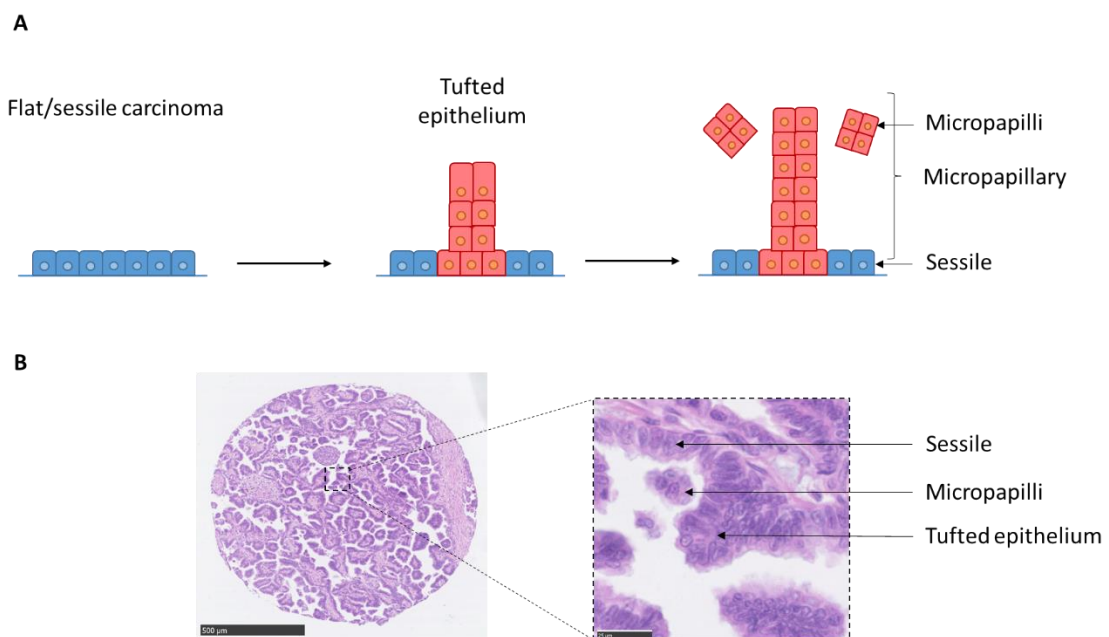
Figure 4.19. Architecture of micropapilli and sessile tumour pattern.
A: Cartoon representation of the continuum of growth. A: Flat sessile sheet of carcinoma cells growing on the stromal surface, followed by a formation of a tufted epithelium. This tumour pattern constitutes the micropapillary growth pattern with micropapillae seen in cross section surrounded by invasive sessile tumour cells. B: Illustration of micropapillary and sessile tumour growth. On the left hand side is an H&E image of a micropapillary tumour growth. Nuclei are stained in purple (haematoxylin and bluing reagent) and cytoplasm is stained in pink (eosin). Magnification = 5X. Scale bar = 500 μm. On the right hand side is a high magnification (80X) of the core portion showing micropapilli surrounded by sessile carcinoma. Scale bar = 25 μm.

I initially aimed to further investigate how EMT occurs in areas of micropapillary tumour growth. Our data show that this growth pattern has little evidence of molecular EMT and generally maintains epithelial molecular features (Chapter 3, Figures 3.20-3.24, 3.25B). However, very close examination indicated that tumour cells lining the acinar or papillary structures surrounding micropapillae, i.e. the "sessile" tumour cells, were often cytokeratin-negative or they weakly express the marker compared to the strongly cytokeratin-positive micropapillae, suggesting a mismatch in partial EMT activation between these two components (Figure 4.20). I therefore wanted to further investigate this phenomenon and designed an experiment with focussed quantification of 20 cores from 20 different cases from the 23 TMAs showing at least five micropapillae structures with sessile tumour cells nearby.

→ Micropapillae
→ Sessile part
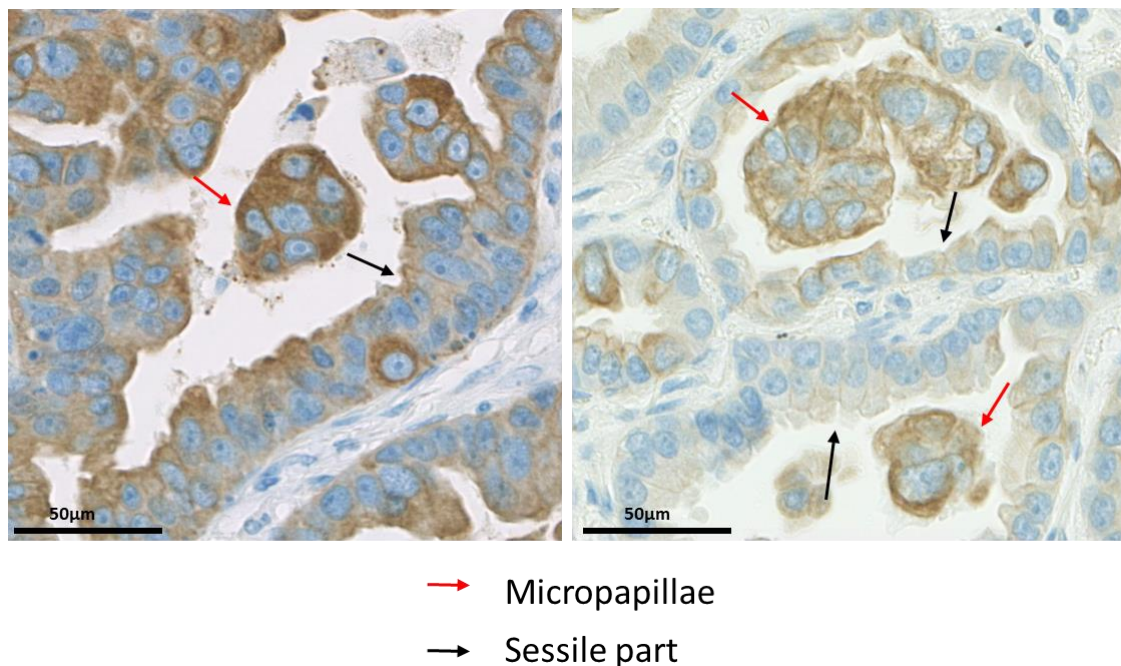
Figure 4.20. Cytokeratin expression variation between micropapillae/sessile complex.
Images of cytokeratin expression in the micropapillae (red arrows) and sessile part of the tumour (black arrows) in a micropapillary tumour growth. Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents cytokeratin expression. Magnification = 40X, scale bar = 50 μm.

The cut-off of five micropapillae was chosen in order to have enough micropapillae and sessile tumour cells to score. These TMAs were already stained for all the EMT markers

from previous experiments (chapter 3). Micropapillae and sessile tumour cells were scored using the H-score system. However, *TWIST* mRNA was not analysed in this study as the TMAs sections stained for this EMT transcription factor did not show much micropapillae structure due to non-sequential sections. They were quite further apart from the rest of the EMT markers stained sections.

To compare the difference in the expression of EMT markers between the micropapillary and the sessile parts, a Wilcoxon signed-rank test was performed.

4.3.4.3 Evidence of focal partial EMT are detected in sessile areas of micropapillary growth pattern

Figure 4.21 illustrates representative images of the expression of EMT markers in the micropapillae/sessile complex in four different tumours cores. Both epithelial markers E-cadherin and cytokeratin exhibit a higher expression in the micropapillae (red arrows) than in the sessile parts (black arrows). In contrast to that, the mesenchymal marker N-cadherin and vimentin are absent in micropapillae with possible very weak focal expression in the sessile part.
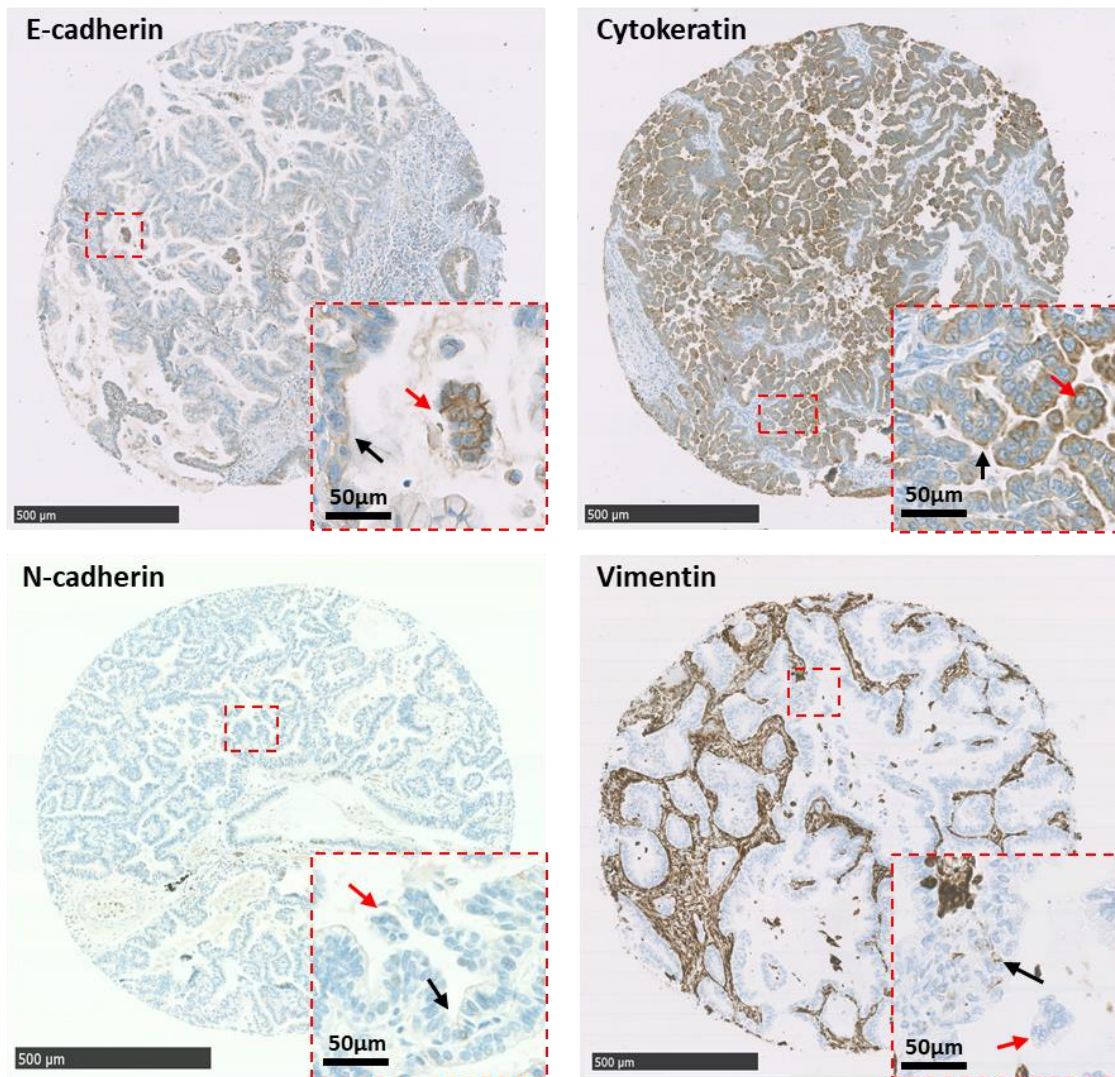
Figure 4.21. Evidence of EMT markers expression in micropapillae/sessile complex.
Core images of E-cadherin, cytokeratin, N-cadherin and vimentin stains. Magnification = 5X. Scale bar = 500 μm. Inset squares represent high magnification (40X) of the core showing micropapillae (represented by red arrows) surrounded by sessile tumour cells (illustrated by black arrows). Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining. Scale bar = 50 μm.

I next quantitated the expression of EMT markers in the micropapillae and sessile tumour cells from the 20 cores.

Figure 4.22A box plot shows that E-cadherin displays a lower expression in the sessile part than the micropapillae. Futhermore, the difference in E-cadherin expression between the two groups is statistically significant with p-value equal to 0.001.

Likewise, cytokeratin expression is half as great in the sessile part than in the micropapillae. The decrease in the sessile tumor cells is significantly more pronounced

than E-cadherin expression (p < 0.001), as the median H-score value of cytokeratin expression in the sessile is around 120 compared to a H-score of 240 in the micropapillae (Figure 4. 22B).

The reduction in both epithelial markers expression may suggest that sessile tumour cells may start undergoing molecular EMT compared to cells in micropapillary tufts.
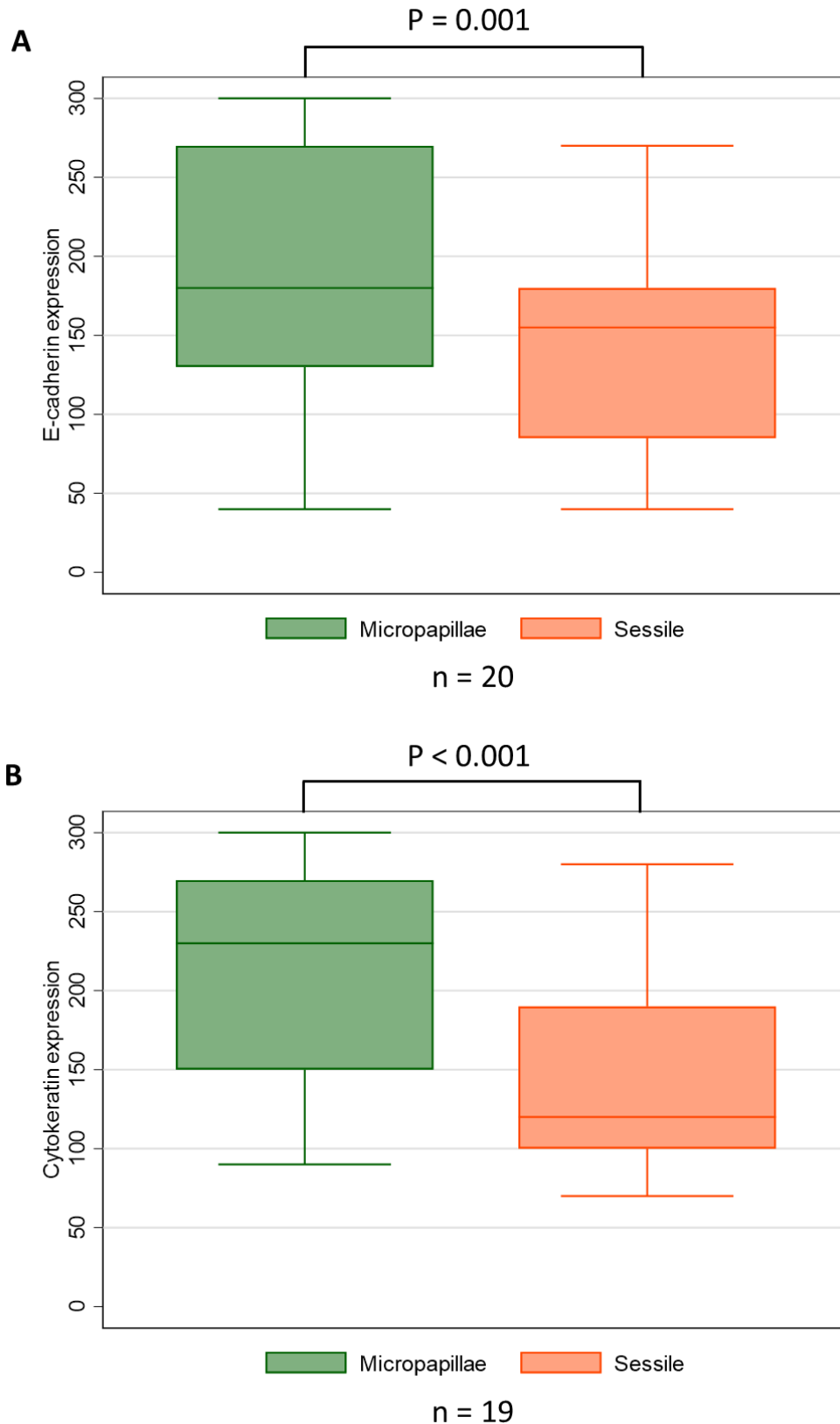
Figure 4.22. Epithelial markers distribution in the micropapillae and sessile complex.
A:Box plot of E-cadherin expression distribution in the micropapillae and sessile part surronding the micropapillae, (n = 20 cores). P = p-value of Wilcoxon signed-rank test. B: Box plot of cytokeratin expression distribution in the micropapillae and sessile part surronding the micropapillae, (n = 19 cores). P = p-value of Wilcoxon signed-rank test.

The mesenchymal markers show an opposite trend to the epithelial markers although without statistical significance. N-cadherin and vimentin are barely detected in the micropapillae while vimentin is very weakly present in the sessile tumour cells (Figure 4.23).

These preliminary data suggest that there is a significantly greater degree of EMT occurring in the sessile parts of micropapillary tumours than in the micropapillary tufts themselves. This is an interesting finding as it seems that micropapillary tumour pattern can present two modes of growth. In tufts/micropapillae, the cells have no stromal contact and are generally highly epithelial in molecular phenotype, while sessile tumours cells in contact with the stroma frequently manifest molecular evidence of EMT, suggesting a gain of invasive properties by the sessile tumours cells.
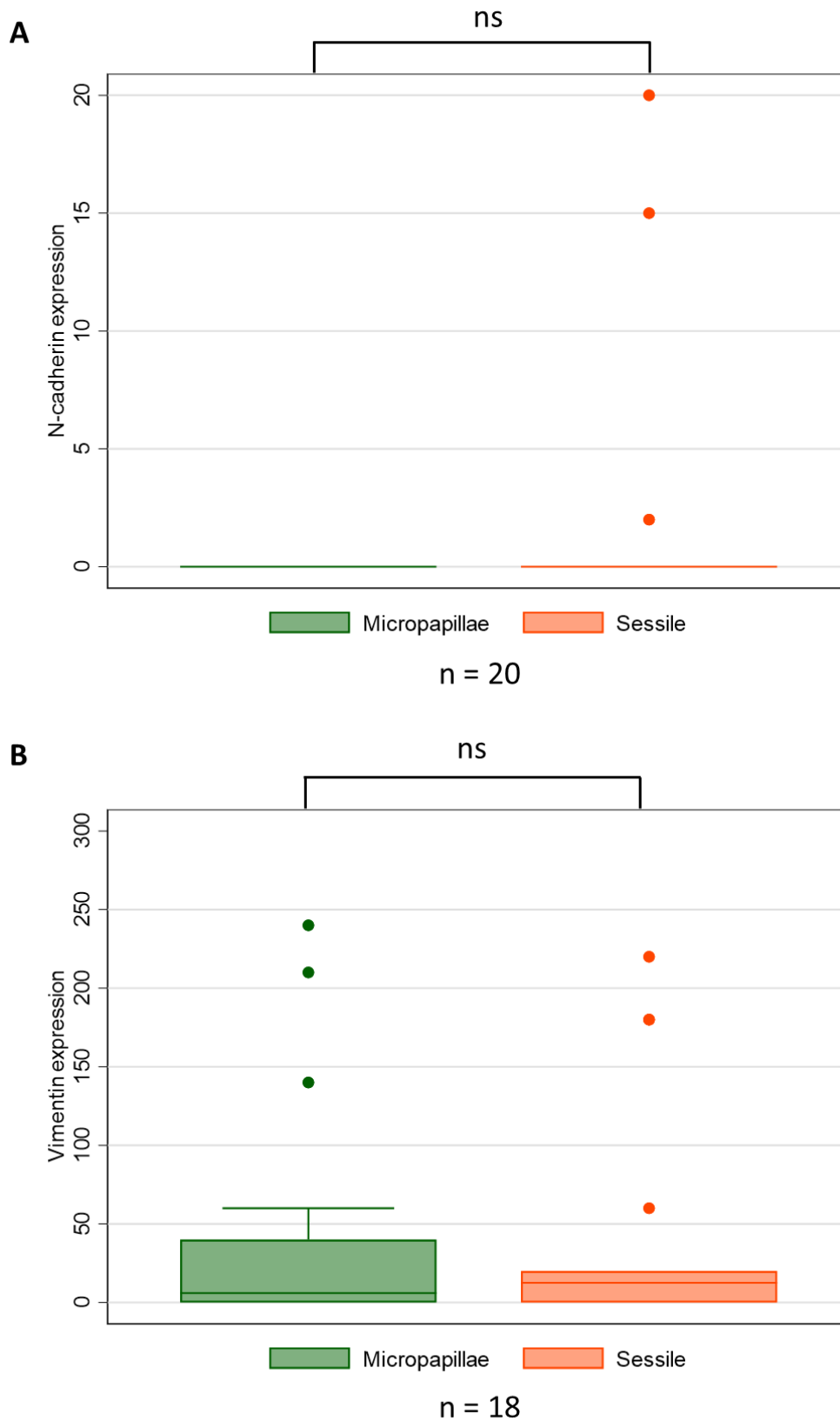
Figure 4.23. Mesenchymal markers distribution in the micropapillae and sessile complex.
A: Box plot of N-cadherin expression distribution in the micropapillae and sessile part surronding the micropapillae, (n = 20 cores). Wilcoxon signed-rank test was performed: ns = not significant. B: Box plot of vimentin expression distribution in the micropapillae and sessile part surronding the micropapillae, (n = 18 cores). Wilcoxon signed-rank test was performed: ns = non significant.

4.4   Discussion

This chapter has explored the role of EMT in different patterns of tumour growth in primary lung adenocarcinomas.

In the first section, I have described the identification of two subgroups of Noguchi's type C tumours. C1 and C2 tumours represent genomically and biologically different types of mixed *in situ*/invasive tumours growth. C1 tumours often represent genomic progression with acquisition of additional mutations in the invasive region. I have detected evidence of molecular EMT with an enhanced mesenchymal markers N-cadherin and vimentin expression in the invasive area of C1 tumours indicating partial EMT in the invasive foci. However, 15 of 18 sequenced C1 tumours in the published study did not show subclonal mutation in the invasive regions (Moore *et al.*, 2019). Given that the step progression to invasiveness is often accompanied by partial EMT, and that not all cases show evidence of genomic progression, I speculate that epigenetic mechanisms may also enable this transition. It would be interesting to correlate the presence of molecular EMT with proven genomic progression; unfortunately I did not have sufficient genomic data on the histologically examined cases to make a meaningful assessment of this. One candidate epigenetic mechanism is dysregulation of mRNA translation, which I investigate further in chapter 6.

The most common mutation seen accompanying the acquisition of invasiveness is mutation of p53. Several studies have demonstrated that modifications in p53 expression can influence metastasis by promoting EMT. For example, p53 mutation, leading to gain-of-function of p53 mutants, is associated with EMT program activation and subsequently an acquisition of metastatic potential by tumour cells (Dong *et al.*, 2013; Lenfert *et al.*, 2015). Wild-type p53 was shown to inhibit cancer invasion by inducing a MDM2-mediated degradation of EMT-transcription factor SNAI2 in NSCLC (Wang *et al.*, 2009b). Wild-type p53 was also demonstrated to prevent the activation EMT programme by regulating micro-RNA expression in breast cancer (Chang *et al.*, 2011). Our data suggest that partial EMT is sufficent to drive invasiveness in C1 tumours. It would be particularly interesting to look specifically at cases with p53 mutations to see if their invasion is indeed related to EMT.

C2 tumours are composed of an invasive central area surrounded by a zone of invasion-competent cells that grow out in a lepidic manner, colonising the alveolar surface. Our data indicate that within C2 tumours *TWIST* mRNA expression is slightly increased and membranous E-cadherin is decreased in this area of *in situ*/lepidic tumour growth relative to the invasive centre. This surprising finding may indicate a potential partial EMT activation enabling invasive cells to colonise the *in situ*/lepidic surface of the lung but further investigation is required for meaningful conclusion. This constant-thickness zone of lepidic growth may be the result of other processes such as proliferation of invasive cells at the tumour periphery rather than motility. Crucially, however, invasiveness in these lesions is not related enhanced EMT; other mechanisms resulting in gradual remodelling of the alveolar architecture appear to be responsible for the transition to invasiveness in these tumours.

I see evidence of one such subtle transition in second part of the chapter, which is focussed on the early detection of invasion in a morphologically *in situ* tumour. I have observed that in this one interesting example, multiple very early foci of apparent invasive behaviour by low-grade cells could be observed. The N-cadherin positive invasive cells seem to be able to damage the alveolar walls, creating gaps with N-cadherin positive edges (or in 2D sections 'promontories'). This may suggest that the invasion can be a subtle and gradual process in lung adenocarcinoma, but it is important to note that I only observed this particular pattern once. Nonetheless, further investigation on *in situ* tumours, preferably with 3D visualisation of such early invasive areas, needs to be carried out to check whether this truly represents an early pattern of invasion.

Next I investigated the potential role of EMT in pleomorphic carcinoma. As mentioned earlier, pleomorphic carcinomas represent a small percentage of lung adenocarcinoma. However this high-grade tumour type predicts poor outcome in lung cancer patients (Travis *et al.*, 2015). Our data have indicated that adenocarcinoma component does not exhibit any molecular evidence of EMT while the regions containing spindle cells show evidence of molecular EMT with loss of membranous E-cadherin and cytokeratin expression and gain of membranous N-cadherin and cytoskeletal vimentin expression. This demonstrates that the spindle-shaped cells underwent complete EMT and acquired

potentially metastatic characteristics. Some studies find that dedifferentiation into a mesenchymal state to create the sarcomatoid region in the pleomorphic adenocarcinoma is often not accompanied by DNA mutation. Ansari-Lari *et al.* examined epithelial and the matched spindle components in 4 out of 20 biphasic sarcomatoid tumours and revealed that *TP53* gene status was identical in both compartments of the sarcomatoid tumour (Ansari-Lari *et al.*, 2002). A case study from Saitoh *et al.* demonstrated that the *EGFR* gene point mutation (EGFR L858R) was found in both epithelial and mesenchymal components of the pleomorphic tumours (Saitoh *et al.*, 2011), confirming a common ancestor for both components. I have not examined the genetic alterations in our pleomorphic carcinoma cases but our data suggests that the sarcomatoid transformation in pleomorphic carcinoma often represents complete EMT (i.e. both molecular and morphological epithelial to mesenchymal changes) in lung adenocarcinoma, and if this is not accompanied by genetic changes, it would be interesting to identify the underlying epigenetic event which promotes EMT in lung pleomorphic carcinoma.

In the last section of chapter 4, I examined the potential role of EMT in developing the micropapillary tumour growth pattern. Micropapillary tumour growth is considered to be high-grade tumour growth and is associated with poor outcome in lung adenocarcinoma patients (Travis *et al.*, 2015). This aggressive tumour pattern also correlates with pleural, haematogenous and lymphatic invasion in addition to micrometastasis to lymph nodes (Cao *et al.*, 2015; Zhang *et al.*, 2011). However the molecular understanding of this tumour growth still remains elusive.

The only other published study to assess EMT in this tumour type was carried out by Nakashima *et al.* in which they examined 101 adenocarcinomas with micropapillary component and showed that micropapillary areas in the tumours exhibited molecular evidence of EMT, with an up-regulation of vimentin expression, suggesting a dedifferentiation and gain of invasiveness of micropapillary tumours (Nakashima *et al.*, 2015). This is inconsistent with our findings and requires explanation. Firstly, our preliminary data (Figures 4.22 and 4.23) suggest that micropapillary growth areas show relative reversal of EMT with an increase of E-cadherin and cytokeratin expression, while the surrounding sessile tumour cells show a greater degree of partial EMT.

Nakashima's work only looked at the expression of vimentin markers in these tumours. I used the same vimentin antibody as Nakashima's study and occasional micropapillary tumours do strongly express vimentin in the micropapillae but I think it is relatively rare. One significant difference between their study and my work is that I separated the micropapillary and sessile part in these tumours and they do not, therefore I are able to comment on intratumoral events in these tumours.

Our study also shows that I can identify two phenotypes within micropapillary areas with relatively more EMT in sessile cells and epithelial phenotypes in micropapillae. How can this help explaining how micropapillary growth contributes to lethality? I suggest that the two components may co-operate. Sessile cells are invasive and facilitate vascular entry, while cells in micropapillae are 'tough' (due to their proven survival without stromal contact) and therefore 'ready to travel'. If micropapillae cannot themselves invade, I speculate that their ability to travel well once they are inside a vessel, with the help of invasive sessile tumour cells to reach lymph nodes or distant sites may explain micropapillary tumour lethality. Further investigation is ongoing to understand the mechanisms underpinning the migration and survival of micropapillae tumours in lung cancer.

In conclusion, EMT can be focal and highly variable within primary lung adenocarcinomas. Partial EMT can accompany the transition to invasiveness in lung adenocarcinoma, either as part of a step change seen as an invasive subclone emerges, or multifocally in apparently morphologically *in situ* tumour growth, or within the sessile parts of highly lethal micropapillary tumours. Finally, complete EMT can accompany the sarcomatoid transformation in pleomorphic carcinoma. EMT can sometimes appear to be accompanied by genomic changes. Can I see evidence of translational control mechanisms regulating this?

Chapter 5.   Molecular EMT is associated with clinical invasiveness and poor lung adenocarcinoma patient outcome

Chapter 5: Molecular EMT is associated with clinical invasiveness and poor lung adenocarcinoma patient outcome

5.1   Introduction

In previous chapters, I investigated the role of EMT in tumour growth pattern of lung adenocarcinoma and I found that partial EMT is associated with solid tumour growth, and that EMT can be highly variable within the whole tumour. I wanted to further understand EMT in primary lung adenocarcinoma. This chapter assesses the relationships between EMT and patients clinicopathological information including patient outcome.

5.2   Aims and objectives

I aim to investigate the relationships between EMT measures and patient clinicopathological factors and outcome using statistical analyses of the immunohistochemical data. I also wanted to determine whether EMT measures could independently predict patient outcome in lung adenocarcinoma patients using multivariate Cox regression models respectively. For the correlation and survival analysis, EMT score was calculated using the median scoring value of each single EMT marker as clinicopathological and survival information are provided per patient rather than per core.

5.3   Results

5.3.1   EMT is associated with physio-pathological factors indicative of invasiveness

I investigated the association between EMT markers themselves and several physio-pathological factors collected from the pathology reports including sex, age and smoking history of patients, tumour staging, pleural and vascular invasion per patient using a Spearman's rank correlation test (Table 5.1 - 5.2). Table 5.2 indicates that all coefficients of correlation between EMT markers and clinicopathological variables were below 0.2 which demonstrates that only weak correlations are observed between these variables.

Spearman's rank correlation	E-cadherin	N-cadherin	Cytokeratin	Vimentin	Twist	EMT score
E-cadherin	1.00	0.18*	- 0.06	- 0.13*	-	- 0.13*
N-cadherin	0.18*	1.00	- 0.16*	- 0.06	- 0.06	0.19*
Cytokeratin	- 0.06	- 0.16*	1.00	-	- 0.06	- 0.12*
Vimentin	- 0.13*	- 0.06	-	1.00	0.10*	0.63*
Twist	-	- 0.06	- 0.06	0.10*	1.00	0.54*
EMT	- 0.13*	0.19*	- 0.12*	0.63*	0.54*	1.00

Table 5.1. Summary of correlations between individual EMT markers and combined EMT score.
The table shows the coefficient of correlations rho between each EMT marker and EMT score. All correlation shown are pairwise and statistically significant, $p < 0.05$.
* denotes the retention of significance after Bonferroni correction. Maximum number of core = 1510 cores.

There is an unexpected association between N- and E-cadherin expression, but that otherwise the observed associations between individual EMT markers are as expected.

Spearman's rank correlation	E-cadherin	N-cadherin	Cytokeratin	Vimentin	Twist	EMT score
Sex	-	-	-	-	-	-
Performance	-	-	-	-	-	-
Smoking	-	-	-	-	-	-
Pleural invasion	0.08	0.09	-	-	-	0.11
Vascular invasion	-	-	-	0.08	-	-
Tumour stage	0.09	0.10	- 0.08	-	-	0.09
Nodal stage	-	0.09	-	0.09	0.09	0.13
Overall stage	-	0.12	-	-	-	0.11

Table 5.2. Summary of correlations between individual EMT markers/score and clinico-pathological factors.
The table shows the coefficient of correlations rho between each EMT marker/score and physio-pathological indicators. All correlation present in the table are pairwise and statistically significant, $p < 0.05$.

Pleural invasion and nodal stage are positively associated with N-cadherin expression and EMT score while nodal stage is also positively linked to vimentin and *TWIST* mRNA expression. Loss of cytokeratin shows no relationship with these clinico-pathological variables.

Tumour stage and overall stage are associated with N-cadherin and EMT score but tumour stage is also negatively related to cytokeratin expression (Table 5.2).

Taken together, these data show that the mesenchymal markers N-cadherin, vimentins and TWIST are statistically and possibly mechanistically related to lymph nodes

metastasis. Although this association is not significant after Bonferroni correction, it is biologically plausible that there may be a causative relationship as cancer cells showing EMT might invade and gain access to lymphatic vessels.

I next sought to assess whether molecular EMT is related to patient survival.

5.3.2 Molecular EMT scores associates with patient outcome

I assessed the relationship between overall, cancer-specific and recurrence-free patient survivals and each EMT marker (E-cadherin, cytokeratin, *TWIST* mRNA, N-cadherin and vimentin) as well as the EMT score created in chapter 3. For the survival analysis, Kaplan-Meier (KM) graphs describe the percentage of survival of 942 cases according to EMT markers scoring values and EMT score over a period of 5-year follow-up after surgery.

Expression levels of all EMT markers were regrouped into two groups except for EMT score: negative and positive groups. The positive group gathers the weak, moderate and strong categories. EMT score variable which combines all EMT markers was regrouped into four categories: the "No EMT" group characterised the patients whose tumours did not present molecular evidence of EMT; the "1 EMT marker", "2 EMT markers" and "≥ 3 EMT markers" groups which designated tumours showing one, two and three or more molecular evidence of EMT.

I then carried out a log-rank test to assess the existence of statistically significant differences between groups. P-value of this test is indicated on the KM graphs while the number at risk is shown under the graphs. Univariate Cox proportional-hazards model was finally performed to evaluate the strenght of the difference between the groups. Each variable is compared to its respective negative group. The results is collected in a table which also indicates the median survival value for each EMT marker and EMT score. EMT markers that show a statistically significant contribution on patient survival outcome are highlighted in bold in the table.
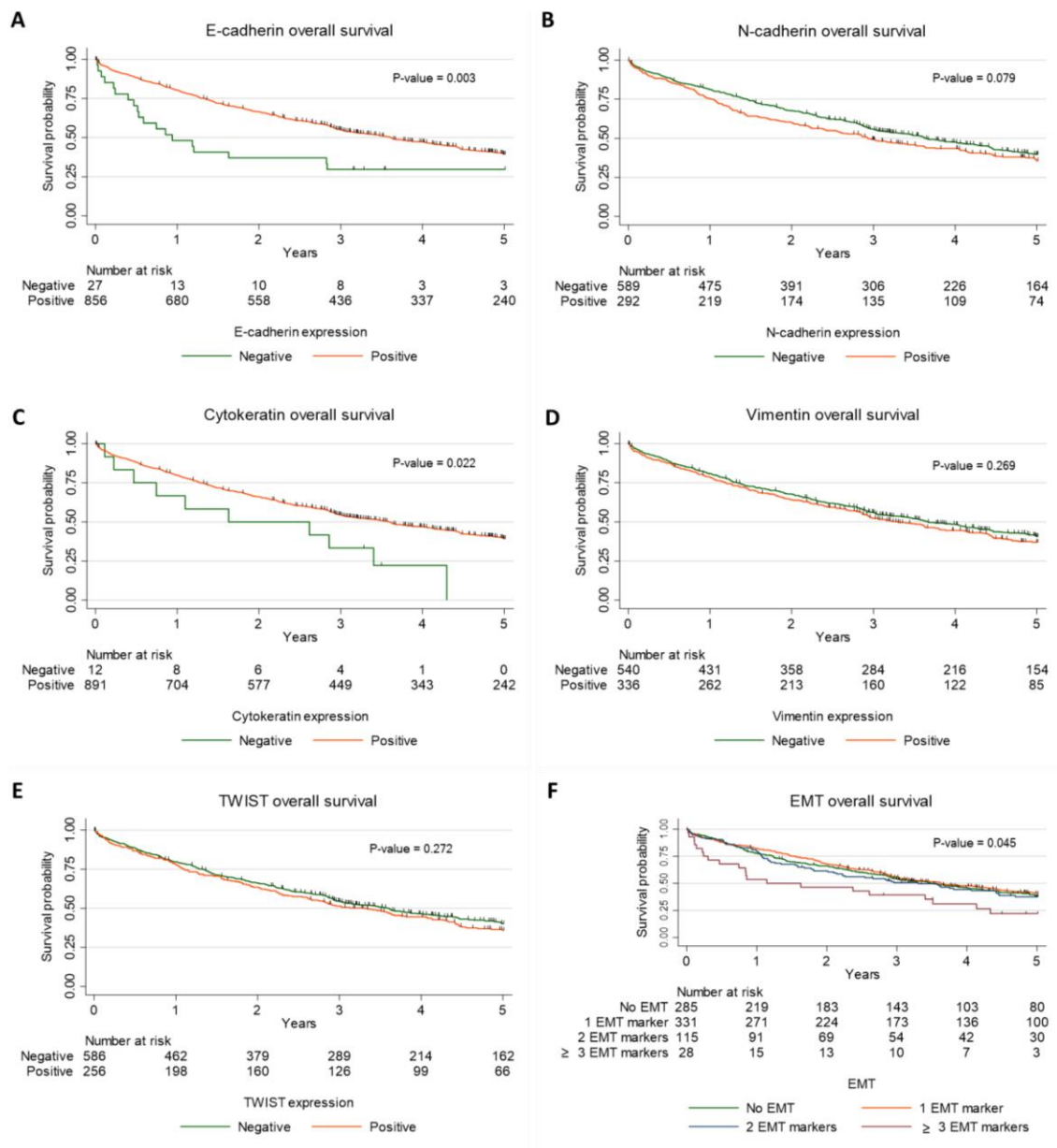
Figure 5.1. EMT markers expression and overall patient survival over a 5-year period post-surgery.
A, B C, D, E and F represent the KM graphs of overall survival of E-cadherin, N-cadherin, cytokeratin, vimentin, *TWIST* mRNA and EMT score respectively over 5-year follow-up after surgery in lung adenocarcinoma patients. P-value of the log-rank test.

EMT markers median survival (years)		Overall survival univariate Cox regression model				
		Variables	Number of events (total number of observations)	Hazard ratio	95% CI	P-value
E-cadherin	Negative = 0.9 Positive = 3.6	E-cadherin Positive vs Negative	508 (883)	0.50	0.32 - 0.79	0.003
N-cadherin	Negative = 3.6 Positive = 2.9	N-cadherin Positive vs Negative	513 (881)	1.17	0.98 - 1.41	0.079
Cytokeratin	Negative = 1.6 Positive = 3.6	Cytokeratin Positive vs Negative	518 (903)	0.49	0.26 - 0.91	0.025
Vimentin	Negative = 3.7 Positive = 3.3	Vimentin Positive vs Negative	501 (876)	1.11	0.91 - 1.30	0.269
TWIST	Negative = 3.6 Positive = 3.3	TWIST Positive vs Negative	484 (842)	1.11	0.92 - 1.36	0.272
EMT score	No EMT = 3.5 1 EMT marker = 3.6 2 EMT markers = 3.3 ≥ 3 EMT markers = 1.1	EMT 1 EMT marker vs No EMT 2 EMT markers vs No EMT ≥ 3 EMT markers vs No EMT	441 (759)	0.93 1.05 1.76	0.75 - 1.15 0.79 - 1.40 1.12 - 2.77	0.498 0.719 0.015

Table 5.3. Correlations between EMT markers and overall survival over a 5-year period post-surgery. The table shows the median survival and the results from the univariate Cox proportional hazards regression over a 5-year period post-surgery in lung adenocarcinoma patients using STATA. 95% CI: 95% confidence interval. Statistically significant associations are highlighted in bold.

The negative E-cadherin group presents a worse outcome with half of the patient surviving within a year (median survival = 0.9 year) whereas median survival in the positive E-cadherin group is four times longer (log-rank p = 0.003). In a Cox proportional hazards regression model positive E-cadherin presents a hazard ratio (HR) equal to 0.50 compared to E-cadherin loss (Figure 5.1A, Table 5.3).

Although the survival trend of positive N-cadherin expression group is poorer compared to the negative group this is without nominal significance (Figure 5.1B, Table 5.3).

Patients presenting with cytokeratin-negative tumours are significantly more likely to die earlier than those with cytokeratin positive tumours (median survival of 1.6 years for the negative group vs 3.6 years positive, log rank test p = 0.022). In the Cox model, cytokeratin-positive tumours show less than 51% of risk of death compared to cytokeratin-negative (HR = 0.49, p = 0.025) (Table 5.3).

Vimentin and *TWIST* mRNA survival graphs reveal that the positive groups show slightly poorer outcome compared to negative one but the difference between these groups is not significantly associated with overall patient survival (Figure 5.1D and E, Table 5.3).

Regarding the composite molecular EMT score, tumours showing one or two molecular evidence of EMT markers demonstrates similar survival tendency between them in

general while tumour presenting three or more molecular EMT indicates a significant poor outcome with a median survival of only 1.1 years. The overall difference between the four groups is significant (p = 0.045). The univariate Cox regression model indicates that tumour with three or more molecular EMT markers has a HR of 1.76 (Figure 5.1F, Table 5.3).

Taken together, loss of E-cadherin or cytokeratin expression on its own correlates with poor overall prognosis, and overall EMT score (combining all the EMT markers favoring EMT process) divulges that patients with their primary tumour exhibiting at least three molecular evidence of EMT are significantly more prone to early death in primary lung adenocarcinoma.
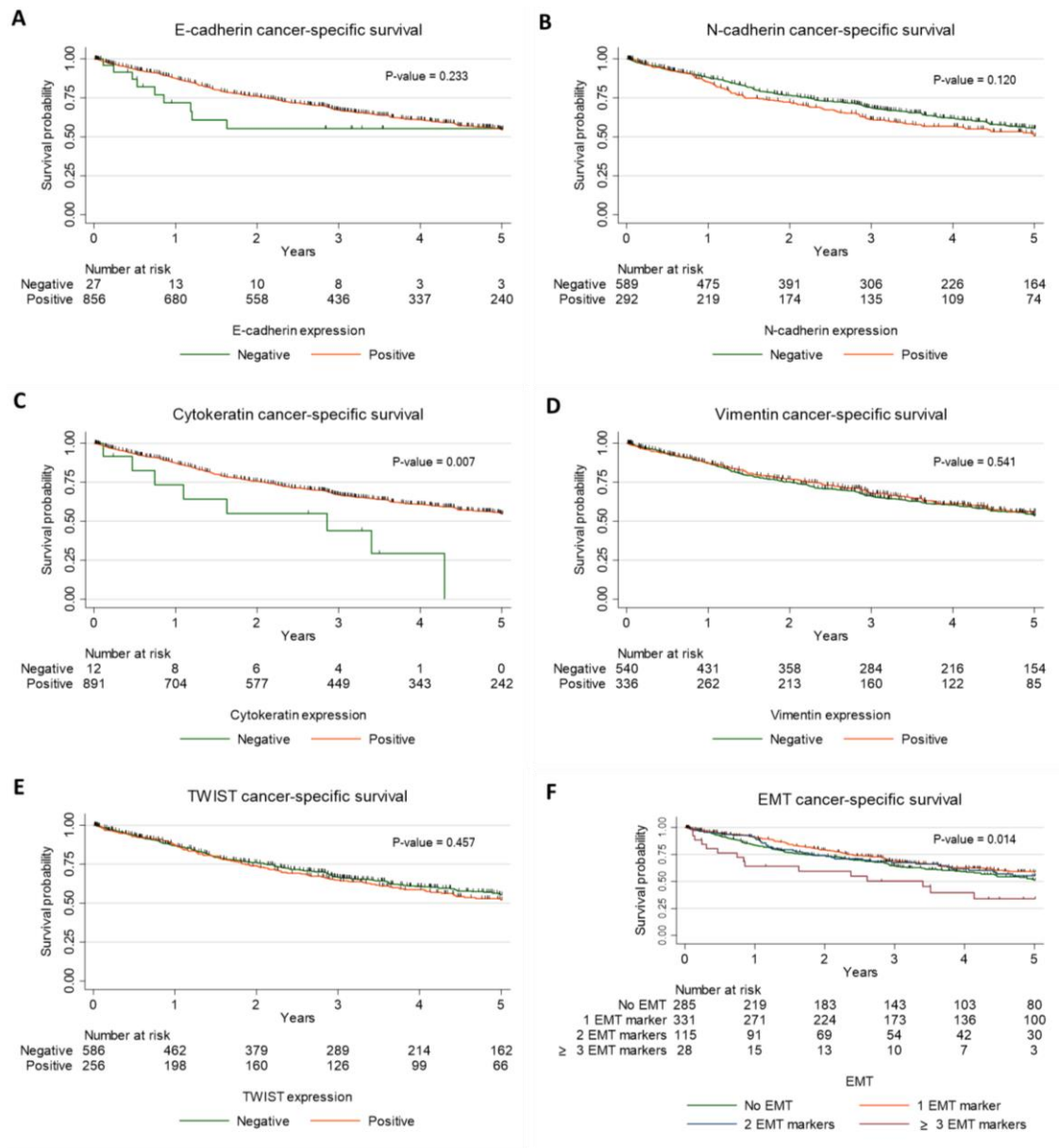
Figure 5.2. EMT markers expression and cancer-specific survival over a 5-year period post-surgery.
A, B C, D, E and F represent the KM graphs of cancer-specific survival of E-cadherin, N-cadherin, cytokeratin, vimentin, *TWIST* mRNA and EMT score respectively over 5-year follow-up after surgery in lung adenocarcinoma patients. P-value of the log-rank test.

.

EMT markers median survival (years)		Cancer-specific survival univariate Cox regression model				
		Variables	Number of events (total number of observations)	Hazard ratio	95% CI	P-value
E-cadherin	Negative = - Positive = -	**E-cadherin** Positive vs Negative	327 (883)	0.67	0.35 - 1.30	0.236
N-cadherin	Negative = - Positive = -	**N-cadherin** Positive vs Negative	331 (881)	1.20	0.95 - 1.50	0.120
Cytokeratin	Negative = 2.9 Positive = -	**Cytokeratin** Positive vs Negative	336 (903)	**0.39**	0.19 - 0.79	**0.009**
Vimentin	Negative = - Positive = -	**Vimentin** Positive vs Negative	325 (876)	0.93	0.74 - 1.17	0.541
TWIST	Negative = - Positive = -	**TWIST** Positive vs Negative	312 (842)	1.09	0.86 - 1.39	0.457
EMT score	No EMT = - 1 EMT marker = - 2 EMT markers = - ≥ 3 EMT markers = 3.4	**EMT** 1 EMT marker vs No EMT 21 EMT markers vs No EMT **≥ 3 EMT markers vs No EMT**	282 (759)	0.80 0.89 **1.83**	0.61 - 1.03 0.62 - 1.28 1.07 - 3.13	0.087 0.529 **0.028**

Table 5.4. Correlations between EMT markers and cancer-specific survival over a 5-year period post-surgery.
The table shows the median survival and the results from the univariate Cox proportional hazards regression over 5-year follow-up after surgery in lung adenocarcinoma patients using STATA. 95% CI: 95% confidence interval. Statistically significant associations are highlighted in bold.

The results for the two next endpoints (i.e. cancer-specific and recurrence-free survivals) are quite similar but interestingly, gain of *TWIST* mRNA expression is associated with poor recurrence-free survival in lung adenocarcinoma patients (Figures 5.2 and 5.3) (Tables 5.4-5.5).

In summary, loss of E-cadherin and cytokeratin expression are generally associated with poor prognosis in all endpoints. When all molecular EMT markers are combined together in one variable, EMT score, it indicates that lung adenocarcinoma exhibiting at least three molecular evidence of EMT, whether it is three gains of mesenchymal factors expression or any combination of loss/gain of epithelial/mesenchymal markers, is related to poor overall and lung cancer-specific survivals as well as an early relapse of the disease in lung primary adenocarcinoma patients.
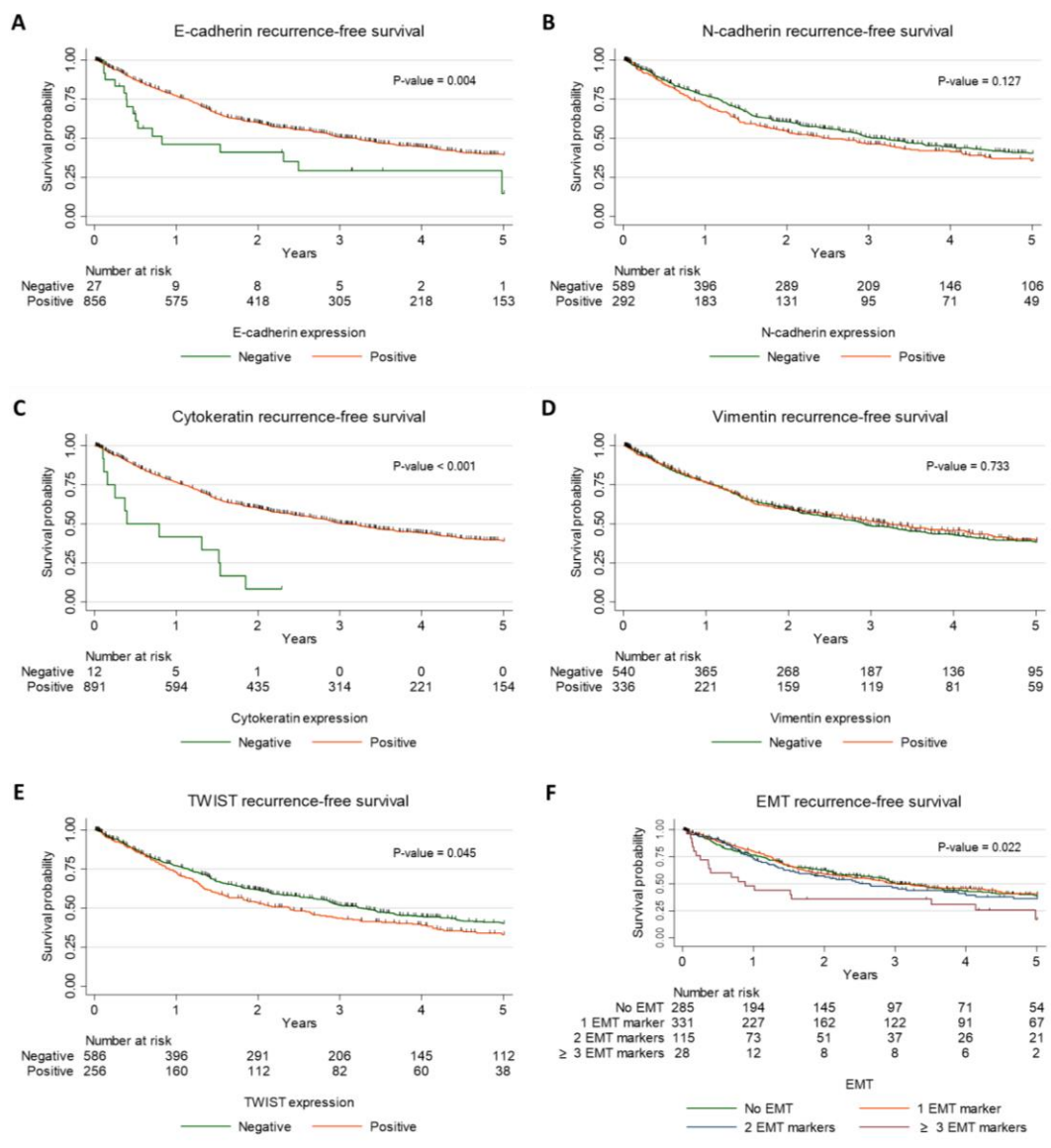
Figure 5.3. EMT markers expression and recurrence free survival analyses over a 5-year period post-surgery.
A, B C, D, E and F represent the recurrence-free KM survival graphs of E-cadherin, N-cadherin, cytokeratin, vimentin, *TWIST* mRNA and EMT score respectively over 5-year follow-up after surgery in lung adenocarcinoma patients. P-value of the log-rank test.

EMT markers median survival (years)		Recurrence-free survival univariate cox regression model				
		Variables	Number of events (total number of observations)	Hazard ratio	95% CI	P-value
E-cadherin	Negative = 0.8 Positive = 3.2	**E-cadherin** Positive vs Negative	433 (883)	**0.49**	0.29 - 0.80	**0.005**
N-cadherin	Negative = 3.1 Positive = 2.5	**N-cadherin** Positive vs Negative	436 (881)	1.17	0.96 - 1.42	0.128
Cytokeratin	Negative = 0.4 Positive = 3.0	**Cytokeratin** Positive vs Negative	446 (903)	**0.26**	0.14 - 0.47	**< 0.001**
Vimentin	Negative = 2.9 Positive = 3.2	**Vimentin** Positive vs Negative	432 (876)	0.97	0.79 - 1.18	0.733
TWIST	Negative = 3.3 Positive = 2.4	**TWIST** Positive vs Negative	418 (842)	**1.23**	1.00 - 1.51	**0.045**
EMT score	No EMT = 3.0 1 EMT marker = 3.2 2 EMT markers = 2.6 ≥ 3 EMT markers = 0.9	**EMT** 1 EMT marker vs No EMT 21 EMT markers vs No EMT **≥ 3 EMT markers vs No EMT**	377 (759)	0.97 1.04 **2.00**	0.77 - 1.22 0.81 - 1.49 1.24 - 3.22	0.790 0.813 **0.005**

Table 5.5. Correlations between EMT markers and recurrence-free survival over a 5-year period post-surgery.
The table shows the median survival and the results from the univariate Cox proportional hazards regression over 5-year follow-up after surgery in lung adenocarcinoma patients using STATA. 95% CI: 95% confidence interval. Statistically significant associations are highlighted in bold.

5.3.3 EMT is not independent of other clinico-pathological variables in prediction of outcome in lung adenocarcinoma patients

I then focused on multivariate Cox models analyses to further investigate whether EMT can independently predict survival outcome in primary lung adenocarcinoma patients.

I first wanted to check whether in combination EMT markers can independently predict overall, lung cancer-specific and recurrence-free survivals outcome in lung adenocarcinoma patients. The total number of patients with complete observations was 759.

	Multivariate Cox regression model											
EMT variables	Overall survival				Cancer-specific survival				Recurrence-free survival			
	Number of events	HR	95% CI	P-value	Number of events	HR	95% CI	P-value	Number of events	HR	95% CI	P-value
E-cadherin Positive vs Negative	411	**0.50**	0.30-0.82	**0.006**	282	0.58	0.29-1.14	0.115	377	**0.54**	0.30-0.96	**0.035**
Cytokeratin Positive vs Negative	411	0.49	0.24-1.02	0.056	282	**0.38**	0.16-0.88	**0.024**	377	**0.26**	0.13-0.53	**< 0.001**
N-cadherin Positive vs Negative	411	**1.23**	1.01-1.50	**0.037**	282	1.21	0.95-1.55	0.127	377	1.22	0.98-1.50	0.069
Vimentin Positive vs Negative	411	1.08	0.89-1.3	0.418	282	0.88	0.68-1.12	0.282	377	0.91	0.74-1.12	0.372
TWIST Negative vs Positive	411	1.09	0.89-1.33	0.406	282	1.14	0.89-1.46	0.302	377	**1.25**	1.01-1.55	**0.040**

Table 5.6. Summary of the multivariate Cox proportional hazards regression models over a 5-year period after surgery.
HR = Hazard ratio. 95% CI: 95% confidence interval. Statistically significant associations are highlighted in bold.

While E-cadherin and cytokeratin loss retain independent significance in overall and cancer-specific survival models, N-cadherin is only significant in overall survival and TWIST is only significant in recurrence-free survival, and that furthermore TWIST barely even shows a trend in the other two endpoints.The lung cancer-specific multivariate Cox model indicates that cytokeratin marker is the only EMT marker to display a significant correlation with patient outcome (HR = 0.38, p = 0.024) in lung adenocarcinoma patients (Table 5.6).

This shows that primary tumours with most evidence of molecular EMT are more likely to result in early recurrence and/or death, presumably due to the presence of occult micro-metastasis at the time of surgery. It also suggests that molecular EMT is more strongly related to recurrence compared to survival endpoints (i.e. overall and cancer-specidic survivals) probably because EMT is more immediately causatively linked to tumour recurrence, i.e. invasion and migration of malignant tumour cells which are likely to form secondary tumour another site in the organism, than patient death, which will be influenced by number of additional tumour factors (i.e. other tumour hallmarks such as increased proliferation, treatment resistance).

I next investigated whether EMT score could predict patient outcome in lung adenocarcinoma independently of known prognostic variables. I performed univariate

and mutivariate Cox regression models for overall and recurrence-free survivals with all key clinical-pathological factors (Tables 5.7-5.8).

High EMT score is not an independent predictor of poor outcome in both multivariate models. Surprisingly a single marker is associated with relatively good overall survival but it is only just significant and of uncertain biological or clinical meaning.

This loss of significance may well be due to the fact that EMT contributes to variables, which are themselves highly predictive, e.g. growth pattern. It is reasonable therefore to stratify the patients by growth pattern.

Overall, EMT score is not statistically significant in the multivariate Cox models for overall and reccurence-free survivals (also for lung cancer-specific, data not shown) therefore EMT score does not predict outcome independently of sex, smoking history, performance status of patients, tumour staging, nodal status, vascular and pleural invasion or lung adenocarcinoma tumour predominant growth pattern in lung adenocarcinoma patients.

	Variables	Univariate overall survival Cox regression model				Multivariate overall survival Cox regression model			
		Number of events	HR	95% CI	P-value	Number of events	HR	95% CI	P-value
EMT score	1 EMT marker vs No EMT	441 (759)*	0.93	0.75 - 1.15	0.498	250 (494)*	0.75	0.57 - 1.00	0.047
	2 EMT markers vs No EMT	441 (759)*	1.05	0.79 - 1.40	0.719	250 (494)*	0.71	0.48 - 1.05	0.087
	≥ 3 EMT markers vs No EMT	441 (759)*	1.76	1.12 - 2.77	0.015	250 (494)*	0.70	0.32 - 1.57	0.393
Sex	Female vs Male	441 (759)*	0.60	0.50 - 0.73	< 0.001	250 (494)*	0.62	0.48 - 0.82	0.001
Smoking	Ex-smoker vs Never smoked	360 (661)*	1.35	0.89 - 2.07	0.161	250 (494)*	1.27	0.74 - 2.17	0.383
	Current smoker vs Never smoked	360 (661)*	1.58	1.02 - 2.44	0.039	250 (494)*	1.36	0.78 - 2.36	0.279
Performance	1 vs 0	339 (638)*	1.15	0.92 - 1.45	0.218	250 (494)*	1.15	0.87 - 1.51	0.317
	2+ vs 0	339 (638)*	1.71	1.21 - 2.43	0.002	250 (494)*	2.16	1.43 - 3.27	< 0.001
Vascular invasion	Present vs Absent	437 (754)*	1.66	1.38 - 2.01	< 0.001	250 (494)*	1.57	1.20 - 2.07	0.001
Pleural invasion	PL1 vs PL0	421 (728)*	1.65	1.33 - 2.05	< 0.001	250 (494)*	1.19	0.85 - 1.67	0.310
	PL2 vs PL0	421 (728)*	2.34	1.72 - 3.20	< 0.001	250 (494)*	1.46	0.89 - 2.38	0.132
	PL3 vs PL0	421 (728)*	3.20	2.29 - 4.47	< 0.001	250 (494)*	1.61	0.90 - 2.86	0.107
Overall stage	Stage IB vs IA	327 (598)*	1.62	1.12 - 2.32	0.010	250 (494)*	1.42	0.89 - 2.26	0.141
	Stage IIA vs IA	327 (598)*	2.72	1.88 - 3.94	< 0.001	250 (494)*	1.99	1.24 - 3.20	0.004
	Stage IIB vs IA	327 (598)*	3.38	2.27 - 5.03	< 0.001	250 (494)*	2.39	1.38 - 4.17	0.002
	Stage IIIA vs IA	327 (598)*	3.54	2.46 - 5.10	< 0.001	250 (494)*	2.76	1.70 - 4.50	< 0.001
	Stage IIIB vs IA	327 (598)*	2.99	1.35 - 6.61	0.007	250 (494)*	2.46	0.95 - 6.35	0.063
Tumour predominant pattern	Acinar vs Lepidic	435 (747)*	1.34	1.00 - 1.80	0.51	250 (494)*	0.81	0.55 - 1.21	0.312
	Cribriform vs Lepidic	435 (747)*	1.35	0.93 - 1.96	0.109	250 (494)*	0.77	0.45 - 1.31	0.335
	Papillary vs Lepidic	435 (747)*	1.31	0.74 - 2.32	0.351	250 (494)*	0.89	0.41 - 1.95	0.773
	Solid vs Lepidic	435 (747)*	2.09	1.52 - 2.87	< 0.001	250 (494)*	1.23	0.79 - 1.91	0.358
	Micropapillary vs Lepidic	435 (747)*	3.92	1.69 - 9.09	0.001	250 (494)*	3.23	1.10 - 9.48	0.033

Table 5.7. Summary of the univariate and multivariate Cox proportional hazards regression models over a 5-year period after surgery.

A and B: Table of univariate and multivariate Cox proportional hazards regression models for overall and recurrence-free survivals respectively in lung adenocarcinoma patients. HR = Hazard ratio. 95% CI: 95% confidence interval.

* indicates the total number of patients.

Variables		Univariate recurrence-free survival Cox regression model				Multivariate recurrence-free survival Cox regression model			
		Number of events	HR	95% CI	P-value	Number of events	HR	95% CI	P-value
EMT score	1 EMT marker vs No EMT	377 (759)*	0.97	0.77 - 1.22	0.790	219 (494)*	0.88	0.66 - 1.20	0.414
	2 EMT markers vs No EMT	377 (759)*	1.10	0.81 - 1.49	0.528	219 (494)*	0.87	0.58 - 1.30	0.487
	≥ 3 EMT markers vs No EMT	377 (759)*	2.00	1.24 - 3.22	0.005	219 (494)*	0.82	0.37 - 1.85	0.638
Sex	Female vs Male	377 (759)*	0.71	0.58 - 0.87	0.001	219 (494)*	0.73	0.55 - 0.96	0.027
Smoking	Ex-smoker vs Never smoked	318 (661)*	1.21	0.80 - 1.85	0.354	219 (494)*	0.98	0.58 - 1.65	0.935
	Current smoker vs Never smoked	318 (661)*	1.20	0.78 - 1.86	0.409	219 (494)*	0.90	0.52 - 1.53	0.674
Performance	1 vs 0	304 (638)*	1.09	0.86 - 1.39	0.482	219 (494)*	1.06	0.79 - 1.42	0.702
	≥2 vs 0	304 (638)*	1.61	1.11 - 2.33	0.013	219 (494)*	1.73	1.08 - 2.78	0.023
Vascular invasion	Present vs Absent	374 (754)*	1.75	1.43 - 2.15	< 0.001	219 (494)*	1.59	1.18 - 2.14	0.002
Pleural invasion	PL1 vs PL0	356 (728)*	1.52	1.20 - 1.92	0.001	219 (494)*	1.20	0.84 - 1.72	0.317
	PL2 vs PL0	356 (728)*	1.87	1.32 - 2.65	< 0.001	219 (494)*	1.01	0.58 - 1.77	0.970
	PL3 vs PL0	356 (728)*	3.74	2.63 - 5.32	< 0.001	219 (494)*	2.05	1.14 - 3.66	0.016
Overall stage	Stage IB vs IA	285 (598)*	1.14	0.77 - 1.69	0.509	219 (494)*	1.05	0.64 - 1.74	0.844
	Stage IIA vs IA	285 (598)*	2.59	1.78 - 3.79	< 0.001	219 (494)*	2.06	1.27 - 3.33	0.003
	Stage IIB vs IA	285 (598)*	3.78	2.51 - 5.70	< 0.001	219 (494)*	2.34	1.30 - 4.21	0.005
	Stage IIIA vs IA	285 (598)*	3.29	2.26 - 4.80	< 0.001	219 (494)*	2.55	1.54 - 4.23	< 0.001
	Stage IIIB vs IA	285 (598)*	3.39	1.53 - 7.51	0.003	219 (494)*	2.73	1.04 - 7.17	0.042
Tumour predominant pattern	Acinar vs Lepidic	372 (747)*	1.37	0.99 - 1.89	0.061	219 (494)*	0.97	0.62 - 1.50	0.882
	Cribriform vs Lepidic	372 (747)*	1.66	1.12 - 2.45	0.012	219 (494)*	0.88	0.49 - 1.58	0.670
	Papillary vs Lepidic	372 (747)*	1.23	0.65 - 2.31	0.530	219 (494)*	0.95	0.41 - 2.22	0.911
	Solid vs Lepidic	372 (747)*	2.25	1.60 - 3.19	< 0.001	219 (494)*	1.37	0.85 - 2.23	0.197
	Micropapillary vs Lepidic	372 (747)*	3.45	1.36 - 8.67	0.009	219 (494)*	2.41	0.70 - 8.27	0.161

Table 5.8. Summary of the univariate and multivariate Cox proportional hazards regression models over a 5-year period after surgery.
A and B: Table of univariate and multivariate Cox proportional hazards regression models for overall and recurrence-free survivals respectively in lung adenocarcinoma patients. HR = Hazard ratio. 95% CI: 95% confidence interval.
* indicates the total number of patients.

If EMT partially explains growth pattern, this could explain why EMT is not statistically independent of pattern in models of outcome. Is it possible that EMT is predictive within 'grade' strata defined by growth pattern?

The survival graphs show that EMT score does not significantly predict outcome in low, medium or high-grade tumours in overall and recurrence-free survivals. Interestingly, high EMT score group best predicts poor outcome in high-grade tumours, despite the lack of nominal statistical significance (Figures 5.4 and 5.5) (Tables 5.9-5.10).

A

EMT overall survival in low-grade tumour

P-value = 0.438

Number at risk

No EMT	51	45	40	33	25	19
1 EMT marker	54	45	39	29	23	12
2 EMT markers	18	17	15	12	7	3
≥ 3 EMT markers	2	1	1	0	0	0

EMT

No EMT — 1 EMT marker — 2 EMT markers — ≥ 3 EMT markers

B

EMT overall survival in medium-grade tumour

P-value = 0.099

Number at risk

No EMT	178	138	119	93	65	51
1 EMT marker	195	170	141	115	91	71
2 EMT markers	65	50	37	29	22	18
≥ 3 EMT markers	9	8	6	4	2	1

EMT

No EMT — 1 EMT marker — 2 EMT markers — ≥ 3 EMT markers

C

EMT overall survival in high-grade tumour

P-value = 0.195

Number at risk

No EMT	53	33	21	16	13	10
1 EMT marker	73	49	37	24	22	17
2 EMT markers	32	24	17	13	13	9
≥ 3 EMT markers	17	6	6	6	5	2

EMT

No EMT — 1 EMT marker — 2 EMT markers — ≥ 3 EMT markers

Figure 5.4. EMT score and overall survival analyses in each tumour grade group over a 5-year period after surgery in lung adenocarcinoma.
A, B and C represent the overall KM survival graphs of EMT score in low-, medium- and high-grade tumour respectively over 5-year follow-up after surgery in lung adenocarcinoma patients. P-values of log-rank tests are displayed on the survival graphs.

225

EMT score median survival (years)	Overall survival univariate Cox regression model				
	Variables	Number of events (total number of observations)	Hazard ratio	95% CI	P-value
EMT score in low-grade tumour No EMT = - 1 EMT marker = 4.1 2 EMT markers = 4.4 ≥ 3 EMT markers = 0.8	**EMT score** 1 EMT marker vs No EMT 2 EMT markers vs No EMT ≥ 3 EMT markers vs No EMT	57 (125)	1.50 1.22 2.95	0.85 - 2.67 0.54 - 2.77 0.39 - 22.3	0.164 0.636 0.295
EMT score in medium-grade tumour No EMT = 3.5 1 EMT marker = 4.3 2 EMT markers = 2.7 ≥ 3 EMT markers = 2.6	**EMT score** 1 EMT marker vs No EMT 2 EMT markers vs No EMT ≥ 3 EMT markers vs No EMT	256 (447)	0.82 1.23 1.43	0.63 - 1.08 0.86 - 1.76 0.66 - 3.07	0.164 0.264 0.361
EMT score in high-grade tumour No EMT = 1.5 1 EMT marker = 2.2 2 EMT markers = 2.3 ≥ 3 EMT markers = 0.5	**EMT score** 1 EMT marker vs No EMT 2 EMT markers vs No EMT ≥ 3 EMT markers vs No EMT	122 (175)	0.87 0.63 1.32	0.58 - 1.32 0.36 - 1.10 0.70 - 2.47	0.513 0.107 0.388

Table 5.9. Correlations between EMT scores and overall survival in each tumour grade group over a 5-year period post-surgery.
The table shows the median survival and the results from the univariate Cox proportional hazards regression in lung adenocarcinoma patients using STATA. 95% CI = 95% confidence interval.

A

EMT recurrence-free survival in low-grade tumour

P-value = 0.840

Survival probability (1.00, 0.75, 0.50, 0.25, 0.00) vs Years (0, 1, 2, 3, 4, 5)

Number at risk

No EMT	51	39	33	23	16	12
1 EMT marker	54	40	29	21	14	7
2 EMT markers	18	13	12	7	4	2
≥ 3 EMT markers	2	1	0	0	0	0

EMT
— No EMT — 1 EMT marker
— 2 EMT markers — ≥ 3 EMT markers

B

EMT recurrence-free survival in medium-grade tumour

P-value = 0.239

Survival probability (1.00, 0.75, 0.50, 0.25, 0.00) vs Years (0, 1, 2, 3, 4, 5)

Number at risk

No EMT	178	124	93	61	47	35
1 EMT marker	195	146	104	80	63	50
2 EMT markers	65	43	27	20	15	13
≥ 3 EMT markers	9	6	3	3	1	1

EMT
— No EMT — 1 EMT marker
— 2 EMT markers — ≥ 3 EMT markers

C

EMT recurrence-free survival in high-grade tumour

P-value = 0.145

Survival probability (1.00, 0.75, 0.50, 0.25, 0.00) vs Years (0, 1, 2, 3, 4, 5)

Number at risk

No EMT	53	28	18	12	8	7
1 EMT marker	73	36	25	18	14	10
2 EMT markers	32	17	12	10	7	6
≥ 3 EMT markers	17	5	5	5	5	1

EMT
— No EMT — 1 EMT marker
— 2 EMT markers — ≥ 3 EMT markers

Figure 5.5. EMT score and recurrence-free survival analyses in each tumour grade group over a 5-year period after surgery in lung adenocarcinoma.
A, B and C represent the recurrence-free KM survival graphs of EMT score in low-, medium- and high-grade tumour respectively over 5-year follow-up after surgery in lung adenocarcinoma patients. P-values of log-rank tests are displayed on the survival graphs.

EMT score median survival (years)		Recurrence free-survival univariate Cox regression model				
		Variables	Number of events (total number of observations)	Hazard ratio	95% CI	P-value
EMT score in low-grade tumour	No EMT = - 1 EMT marker = - 2 EMT markers = - ≥ 3 EMT markers = 0.8	EMT score 1 EMT marker vs No EMT 2 EMT markers vs No EMT ≥ 3 EMT markers vs No EMT	47 (125)	1.00 1.09 2.45	0.54 - 1.89 0.46 - 2.59 0.32 - 18.6	0.981 0.840 0.385
EMT score in medium-grade tumour	No EMT = 3.3 1 EMT marker = 3.4 2 EMT markers = 2.4 ≥ 3 EMT markers = 1.5	EMT score 1 EMT marker vs No EMT 2 EMT markers vs No EMT ≥ 3 EMT markers vs No EMT	222 (447)	0.95 1.31 1.67	0.71 - 1.27 0.89 - 1.92 0.73 - 3.83	0.732 0.176 0.223
EMT score in high-grade tumour	No EMT = 1.4 1 EMT marker = 1.4 2 EMT markers = 3.2 ≥ 3 EMT markers = 0.4	EMT score 1 EMT marker vs No EMT 2 EMT markers vs No EMT ≥ 3 EMT markers vs No EMT	103 (175)	0.86 0.61 1.44	0.55 - 1.35 0.34 - 1.13 0.75 - 2.79	0.511 0.115 0.278

Table 5.10. Correlations between EMT score and recurrence-free survival in each tumour grade over a 5-year period post-surgery.
The table shows the median survival and the results from the univariate Cox proportional hazards regression in lung adenocarcinoma patients using STATA. 95% CI = 95% confidence interval.

5.4   Discussion

This chapter focuses on investigating the relationship of EMT and patient physio-pathological factors as well as the survival outcome.

Studies have demonstrated positive relationship between N-cadherin or vimentin expression and nodal stage in breast cancer (Elmoneim and Zaghloul, 2011; Vora *et al.*, 2009) and colorectal cancer (Toiyama *et al.*, 2013; Yan *et al.*, 2015) but not in lung adenocarcinoma. Our data show for the first time that nodal stage positively correlate with N-cadherin, vimentin, TWIST and EMT score.

The results from univariate and multivariate Cox analyses (Figures 5.1-5.3 and table 5.3) indicate that loss of E-cadherin can predict overall survival independently of other EMT markers, while loss of cytokeratin is independently associated with poor lung cancer-specific survival. Moreover, loss of either epithelial marker or gain of TWIST expression independently associate with early recurrence of the disease in lung adenocarcinoma. Reduced/loss of E-cadherin expression has been linked to poor survival in colorectal, breast or prostate cancer (Elzagheid *et al.*, 2012; Li *et al.*, 2017; Umbas *et al.*, 1994) or even in NSCLC as a whole (Sulzer *et al.*, 1998) but not specifically in lung adenocarcinoma. Studies that show an association between this epithelial marker and patient outcome in lung adenocarcinoma mostly examined E-cadherin expression but I was the first one to demonstrate that cytokeratin loss independently correlated with poor lung cancer-specific and recurrence-free survivals in primary lung adenocarcinoma. High EMT score is linked to overall, lung cancer-specific and recurrence-free outcome in univariate models but EMT score cannot independently predict outcome, suggesting that other patho-physiological indicators in our model may contain the same prognostic information, such as tumour growth pattern. Our data show a model in which EMT favours invasion of the lymphatic vessels. This may also explain the association of high EMT score with poor outcome in univariate Cox models for all three endpoints.

EMT is most predictive in high-grade tumours. This may be due to the fact that molecular EMT is less meaningful in tumours which are less invasive, but in high-grade

tumours, the great majority of which are solid-pattern, EMT can subsequently favour tumours to invade regional lymph nodes and lead to poor outcome in lung adenocarcinoma patients.

Chapter 6.   Associations between translational dysregulation and EMT in

primary pulmonary adenocarcinoma

Chapter 6 Associations between translational dysregulation and EMT in primary pulmonary adenocarcinoma

6.1 Introduction

Translation is a crucial stage for the regulation of gene expression, and mRNA translation reprogramming is regarded an emerging hallmark of cancer. Deregulation of translation initiation factors has been asssociated with EMT in cell culture however, its association in primary tissues had not been investigated in primary lung cancer tissue. This chapter aims to target this issue.

6.2 Aims and objectives

The main aim of this chapter is to explore the potential role of translation dysregulation in EMT:

- Can I find evidence of translation dysregulation favoring activation of EMT programmes in primary lung adenocarcinoma?

To realise this aim, 23 TMAs containing 942 non-mucinous lung adenocarcinomas samples were immunohistochemically stained for the translation factors eIF4B, eIF4E, eIF4A1, eFI4A2, eIF4G, phospho-eIF4E and phospho-eIF2α. I generated data on expression of eIF4B and eIF4E in primary lung adenocarcinoma while Bethany scored phospho-eIF2α and the MRC histology core facility scored the remaining translation factors. These two proteins were manually semi-quantified according to their expression intensity. To investigate heterogeneity between cases and to understand the relationships between eIF4B and eIF4E and predominant tumour growth pattern (information given per patient from the pathology report), median intensity scores values of the three cores were considered. To explore the correlation between these two translation markers and tumour histological pattern, intensity scores of all three cores and the histological tumour patterns per TMA core evaluated by Dr Juvenal Baena-Acevedo, were considered. To assess heterogeneity of eIF4B within cores, intensity and proportion scores (i.e. percentage of positive tumour cell stained scores) of eIF4B for all three cores were considered. Strikingly, eIF4E expression was near-

homogenous within cores and so I did not perform formal proportion scoring of this marker. To evaluate the variation of each translation factor expression between cores within cases, the variance per patient was calculated for each translation factor.

6.3   Results

6.3.1   Optimisation of the translation initiation factors by IHC

The lung adenocarcinoma validation block, described in the material and methods chapter, was used as positive control as well as the test tissue. Three concentrations of antibodies were used: the recommended concentration from the antibody's manufacturer, a higher and lower concentration tested with low and high pH antigen retrieval solution separately. eIF4B staining exhibit a cytoplasmic staining as expected. As shown in Figure 6.1, eIF4B staining demonstrates background in the high pH antigen retrieval solution and eIF4B at dilution 1/100 and 1/250 in the low pH antigen retrieval. Therefore, the chosen antibody dilution and antigen retrieval solution was anti-eIF4B at 1/500 in low pH antigen retrieval solution. Once the opimised antibody concentration was selected, the negative control without the primary antibody was run alongside the positive/validation specimen. eIF4B negative control with no primary antibody show no expression the translation factor as expected.
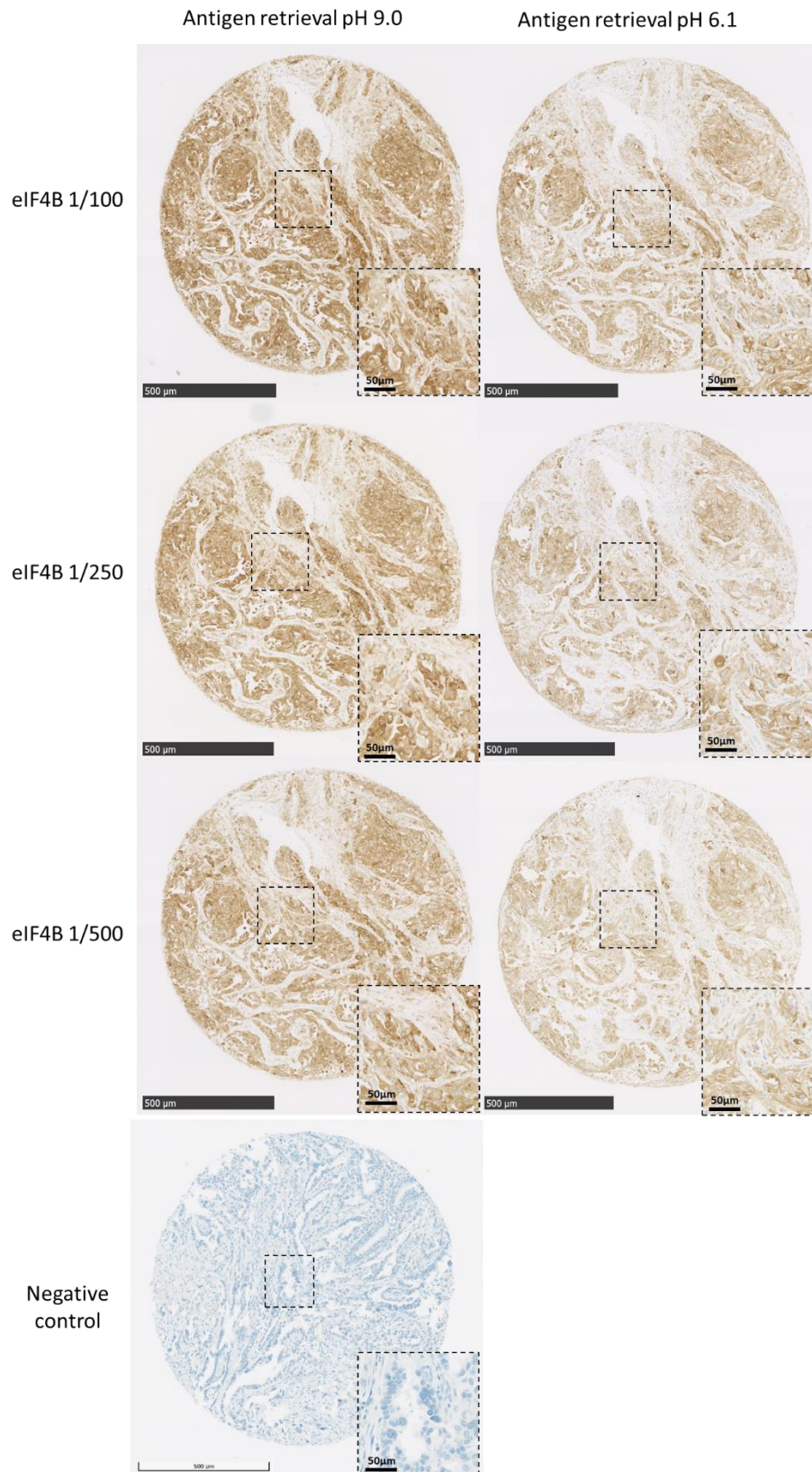
Figure 6.1. IHC validation of eIF4B antibody.
On the left hand side, three eIF4B antibody concentrations in antigen retrieval pH9 in lung adenocarcinoma. The bottom left is the negative control of eIF4B. On the right hand side, three eIF4B antibody concentrations in antigen retrieval pH 6.1. Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents eIF4B expression.Core image: Magnification = 5x. Scale bar = 500 µm.Inset high resolution of core section = 40x. Scale bar = 50 µm.

Similarly to eIF4B antibody optimisation, eIF4E validation was carried out on lung adenocarcinoma tissue. eIF4E is barely detected in low pH antigen retrieval buffer. eIF4E diluted at 1/100 and 1/500 high pH antigen retrieval show strong and weak expression respectively. Therefore the validated eIF4E antibody concentration is 1/250 in antigen retrieval pH 9. The negative control run alongside the positive/validated lung tissue illustrates no staining (Figure 6.2).
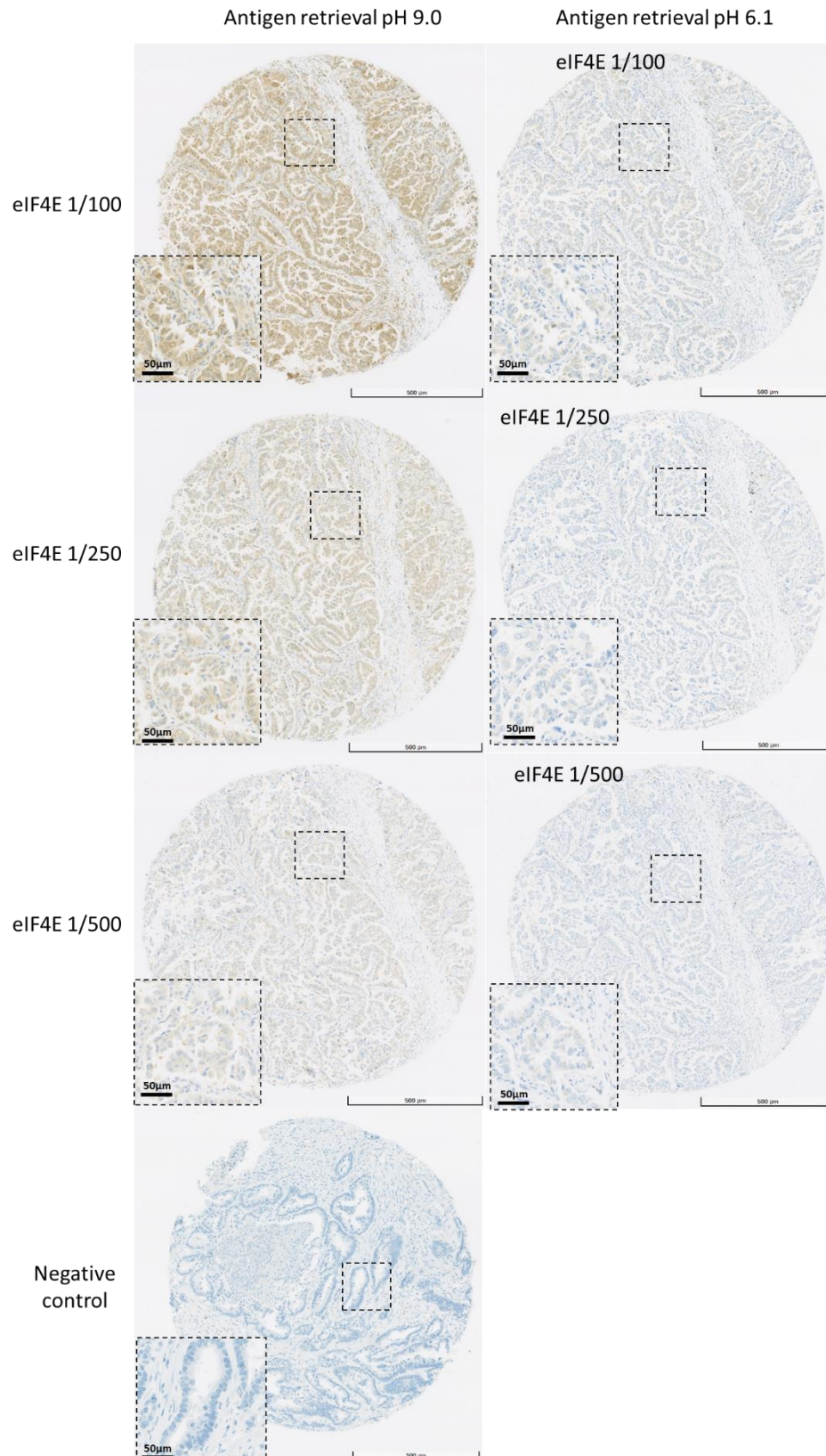
Figure 6.2. IHC validation of eIF4E antibody.
On the left hand side, three eIF4E antibody concentrations in antigen retrieval pH9 in lung adenocarcinoma tissue. The bottom left is the negative control of eIF4E. On the right hand side, three eIF4E antibody concentrations in antigen retrieval pH 6.1. Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents eIF4E expression. Core image: Magnification = 5x. Scale bar = 500 μm. Inset high resolution of core section = 40x. Scale bar = 50 μm.

6.3.2 Characterisation of eIF4B and eIF4E in lung adenocarcinoma

eIF4B immunohistochemical staining shows diverse range of intensity in the cohort. eIF4B is scored according to the cytoplasmic intensity staining. eIF4B is expressed in the normal lung. Tumour cores are either negative or express weakly, moderately or strongly the translation factor (Figure 6.3).

The majority of the cores express eIF4B quite homogenously within the tumour core but at various intensities between cores (Figure 6.4A). Only about 120 cases do not detectably express the protein, which may due to the age of the block and the antigen preservation. The translation factor shows a weak expression in about 370 cases, a moderate and strong expression intensity in about 200 cases each, also indicating heterogeneity between cases (Figure 6.4B). Most of the cases do not show variation in eIF4B expression between tumour cores per patient. Aproximately 200 cases demonstrate a heterogeneity in eIF4B expression between cores of the same primary tumour. Therefore eIF4B expression is mostly homogeneous between cores within the primary tumours of patients (Figure 6.4C).
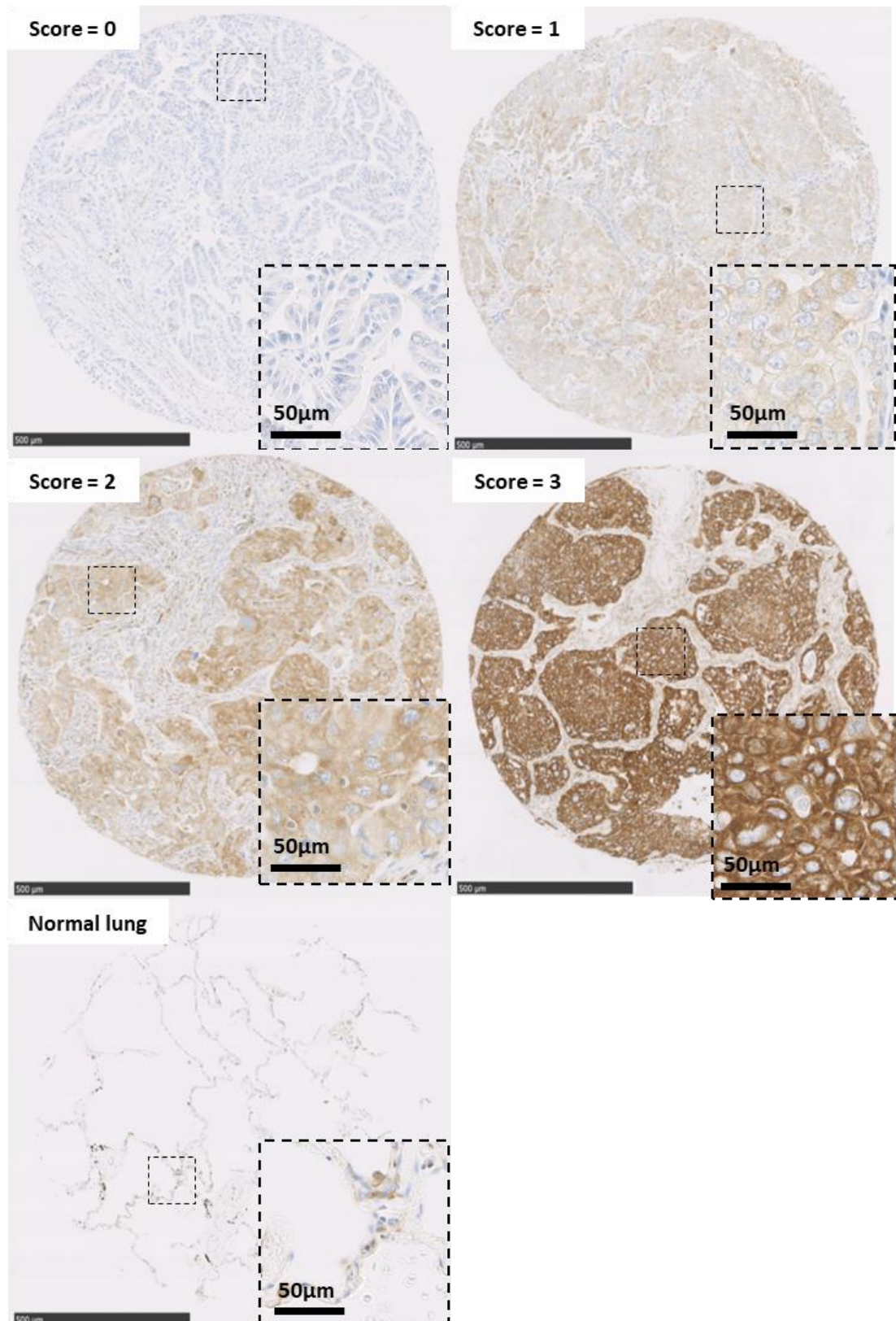
Figure 6.3. Dynamic range of eIF4B intensity.
Images of translation iniation factor eIF4B immunostaining scored according to their expression intensities Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents eIF4B expression. Core image magnification: 5x, scale bar = 500µm. Inset squares represent high magnification (40x) of a portion of the core, scale bar = 50 µm. eIF4B is expressed in type 1 and 2 pneumocytes but not in erytrocytes present in the space(white round cells) in the normal lung.

A: eIF4B intensity per core

B: Median eIF4B expression per case

C: Variation of eIF4B expression between cores per patient
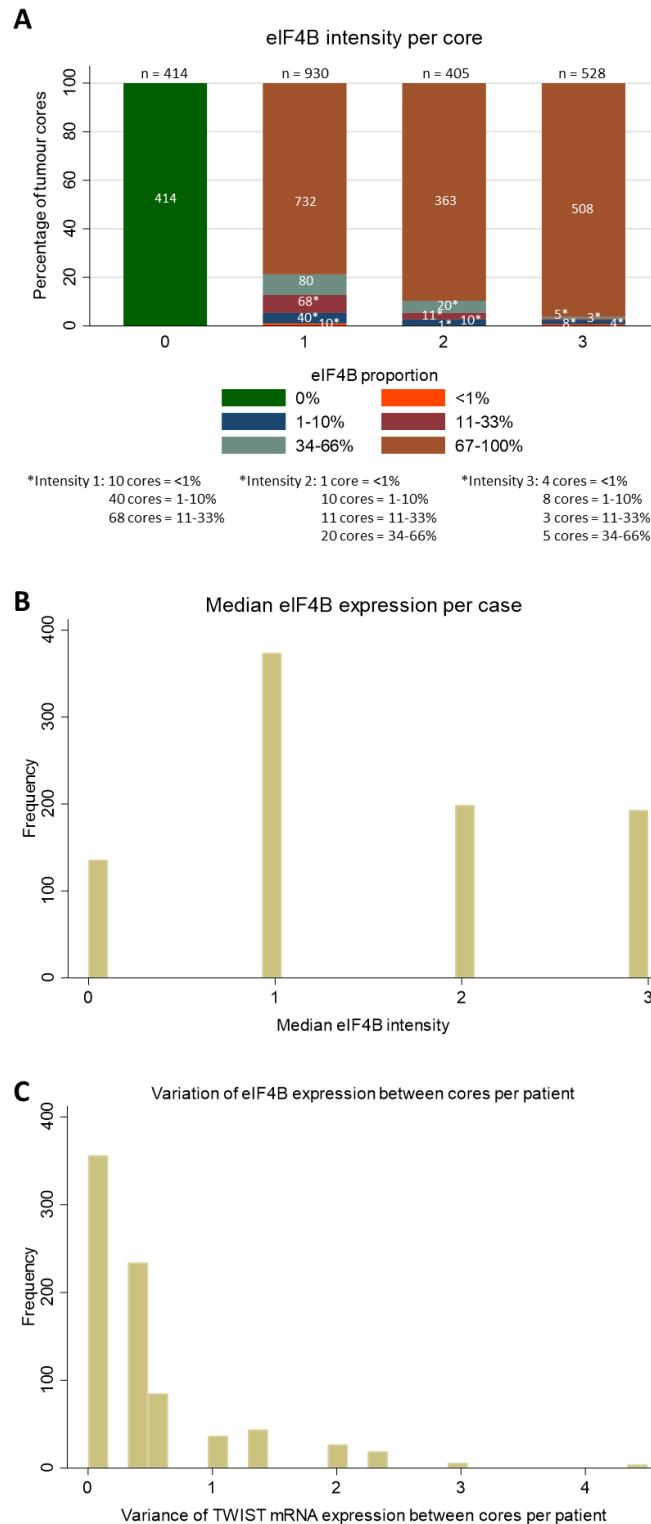
Figure 6.4. Distribution of eIF4B intensity and proportion scores per core and its frequency expression per case in the cohort.

A: Percentage of tumour cores expressing eIF4B at different proportion for each intensity in lung adenocarcinoma cores. n = number of cores B: Frequency of eIF4B expression according to the median intensity score value in 903 lung adenocarcinoma cases. C: Variance of eIF4B expression between cores within cases in 812 lung adenocarcinoma patients. Variance is calculated as the square of standard deviation.

Unit of measurements: 1 unit = 1 scoring value.

eIF4E immunohistochemical staining also shows a dynamic range of expression in tumour cells but is not detectable in the normal lung using this assay. These representative immunohistochemical stainings of eIF4E display that eIF4E is homogenously expressed in tumour cores (Figure 6.5) and this has been previously shown in breast cancer (Ramon y Cajal *et al.*, 2014).
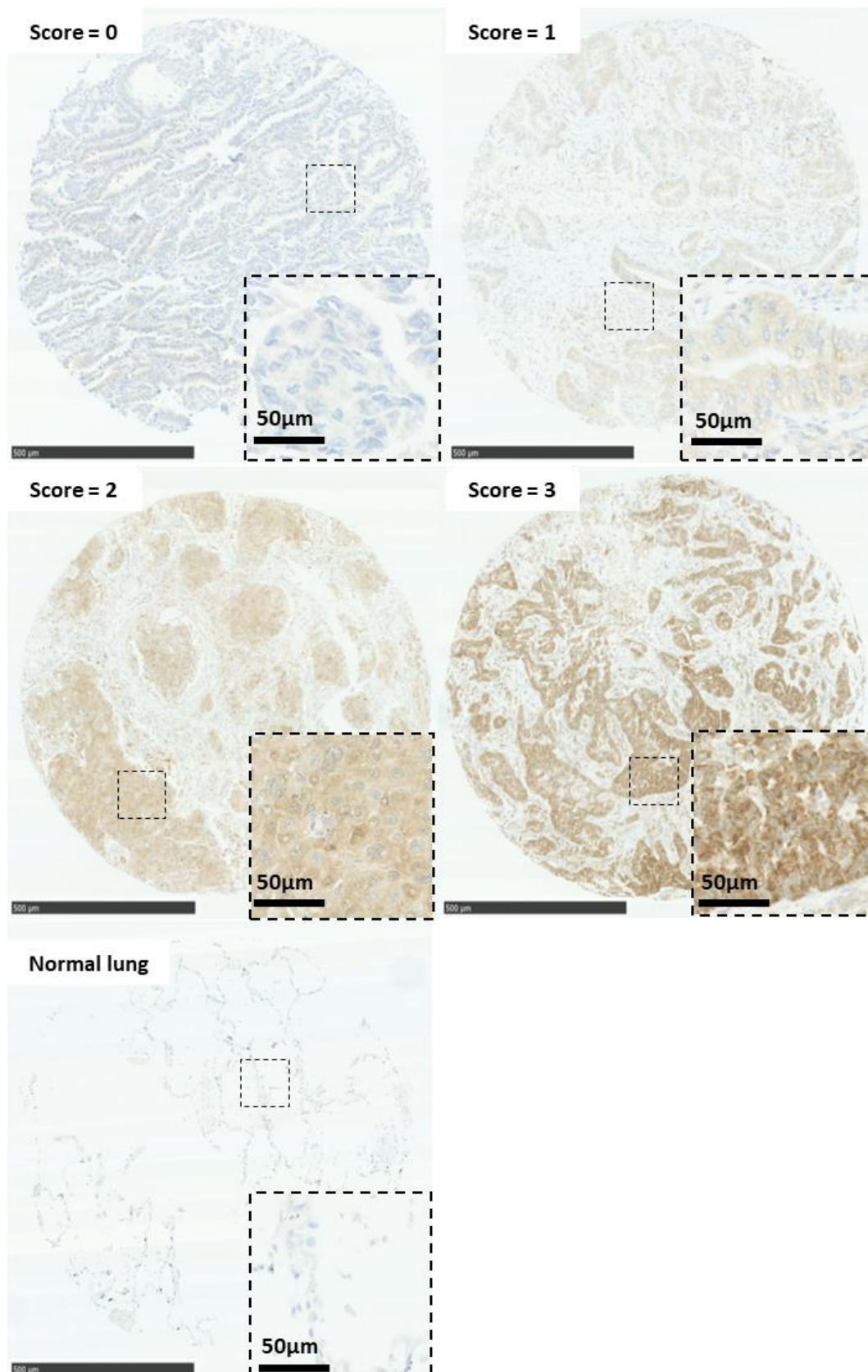
Figure 6.5. Dynamic range of eIF4E intensity.
Images of translation iniation factor eIF4E immunostaining scored according to their expression intensities Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents eIF4E expression. Core image magnification: 5x, scale bar = 500µm. Inset squares represent high magnification (40x) of a portion of the core, scale bar = 50 µm.

eIF4E is heterogenously expressed between cases in primary lung adenocarcinoma. eIF4E is almost expressed in all cases with different intensity while about 30 cases are negative for eIF4E. The majority of the cases demonstrate weak, moderate and strong expression respectively of this translation initiation factor (Figure 6.6A). Furthermore most cases show no difference in eIF4E expression between tumour cores from the same primary tumours of patients, indicating a generally homogenous expression of the mRNA cap-binding protein between cores within cases (Figure 6.6B).
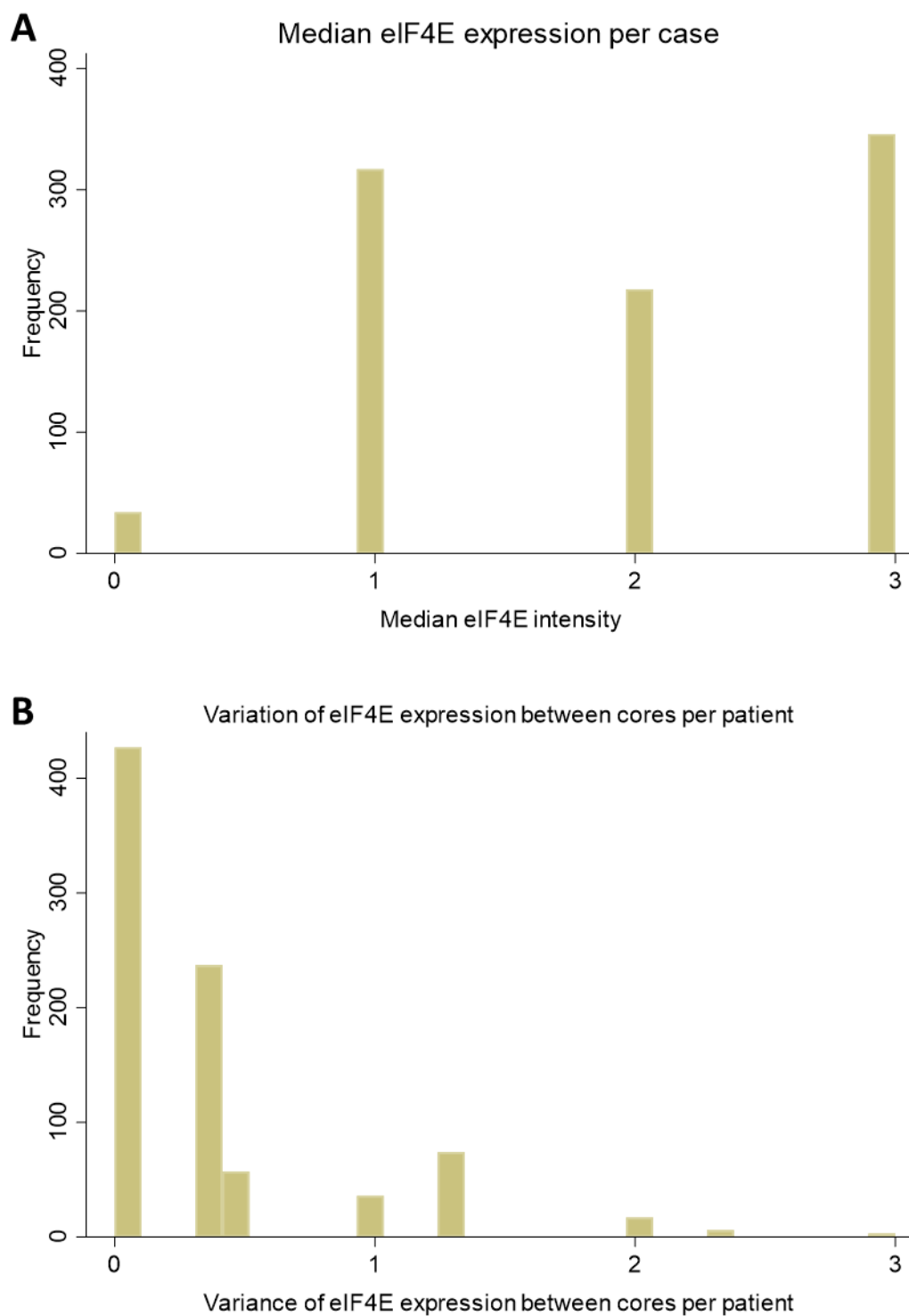
Figure 6.6. Frequency of eIF4E expression per case in the cohort.
A: Frequency of eIF4E expression according to the median intensity score values in 916 lung adenocarcinoma cases. B: Variance of eIF4E expression between cores within cases in 857 lung adenocarcinoma patients. Variance is calculated as the square of standard deviation.
Unit of measurements: 1 unit = 1 scoring value.

6.3.3 eIF4B expression is related to invasive tumour growth pattern

Similarly to chapter 3, I next sought to explore the relatiosnhip between eIF4B and eIF4E expression and tumour growth pattern. As shown in Figures 6.7 and 6.8, the stacked bar graphs exhibit the classification of the degree of the protein of interest in the different tumour histological pattern of lung adenocarcinoma.

Solid pattern exhibits high-grade invasive disease and shows evidence of partial EMT. Therefore I decided to examine the difference in the translation factors between solid tumour growth patterns and the other tumour growth patterns.

The proportion of eIF4B negative cores represents just over 20% of the lepidic predominant tumour growth pattern and this percentage is less in more aggressive growth patterns, being as low as about 4% in the solid predominant tumour growth pattern. Of note, while the percentage of weak and moderate expression of eIF4B seems quite constant across tumour predominant pattern, the proportion of strong staining varies considerably (7% in lepidic predominant tumours vs. 33% in solid predominant tumours) and accounts for the significant difference in eIF4B total expression (Figure 6.7A). Tumour core analysis divulges similar relationship trends in eIF4B expression across tumour pattern as the predominant tumour growth pattern analysis with additional significant difference in the translation factor expression between tumour pattern (Figure 6.7B).

A

Median eIF4B expression by predominant tumour growth pattern

B

eIF4B expression by TMA core growth pattern
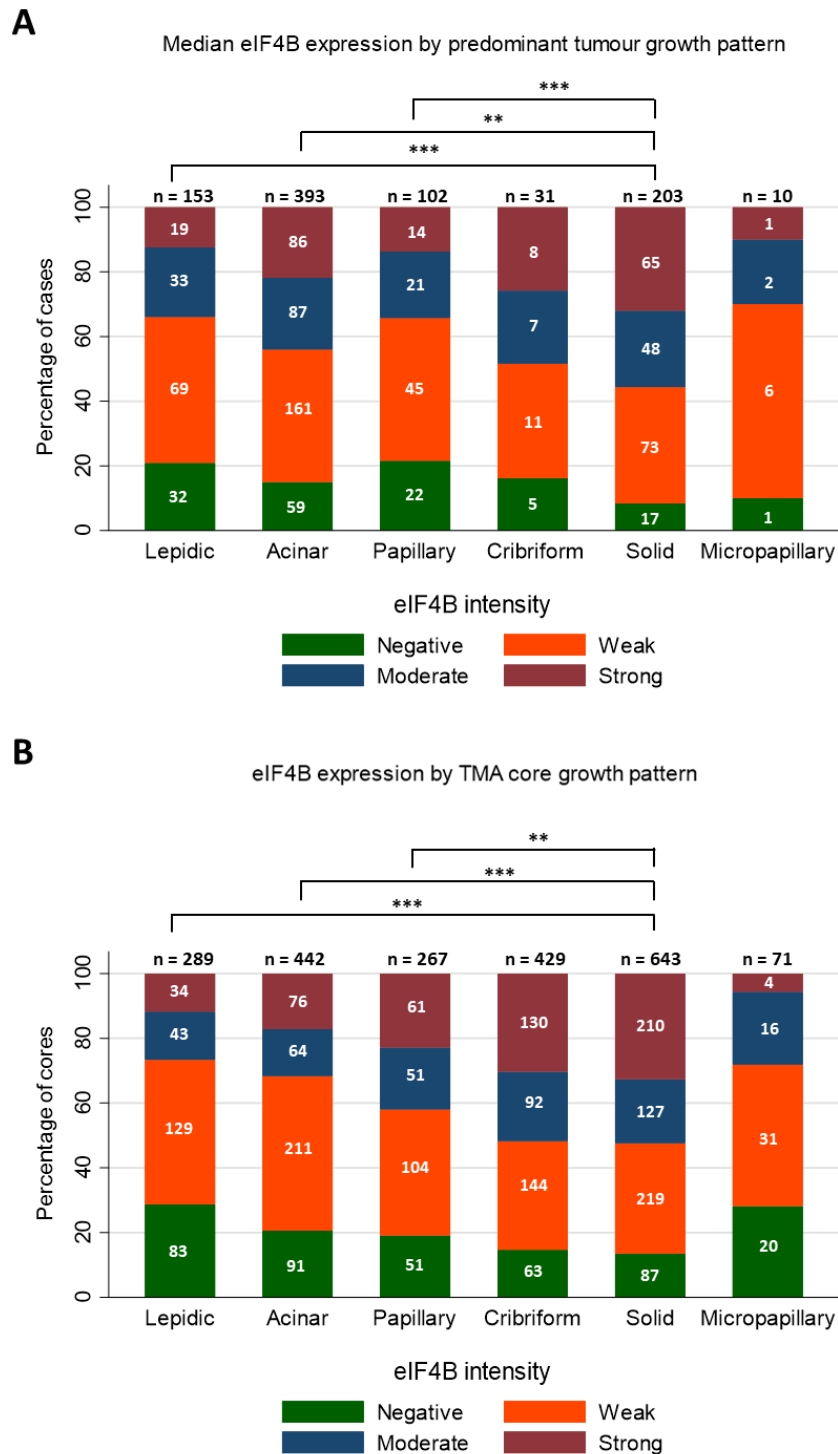
Figure 6.7. Distribution of eIF4B expression by lung adenocarcinoma tumour growth pattern.
Each bar displays the tumour growth pattern of lung adenocarcinoma and each segment of the bar represents the intensity of eIF4B. Total number of growth pattern cases is indicated at the top of each bar and the number of cases displaying the different degree of eIF4B expression is shown within each section of the bar. Degree of significance are indicated in the graphs by bars with stars above them (** = p <0.01; *** = p < 0.001). A: Percentage of median eIF4B expression intensity by predominant tumour growth pattern per case, n= number of case. B: Percentage of eIF4B expression intensity by TMA core growth pattern, n= number of core. Degree of significance are indicated on top of bars in the graphs by stars (** = p <0.01; *** = p < 0.001).

There is a significant decrease in the strongly expressed eIF4E group with increased invasive pattern (42% in lepidic predominant tumour growth pattern vs 30 % in solid predominant tumour growth pattern) (Figure 6.8A), which is somewhat surprising, given that overexpression of eIF4E has been linked to tumorigenesis (Hsieh and Ruggero 2010).
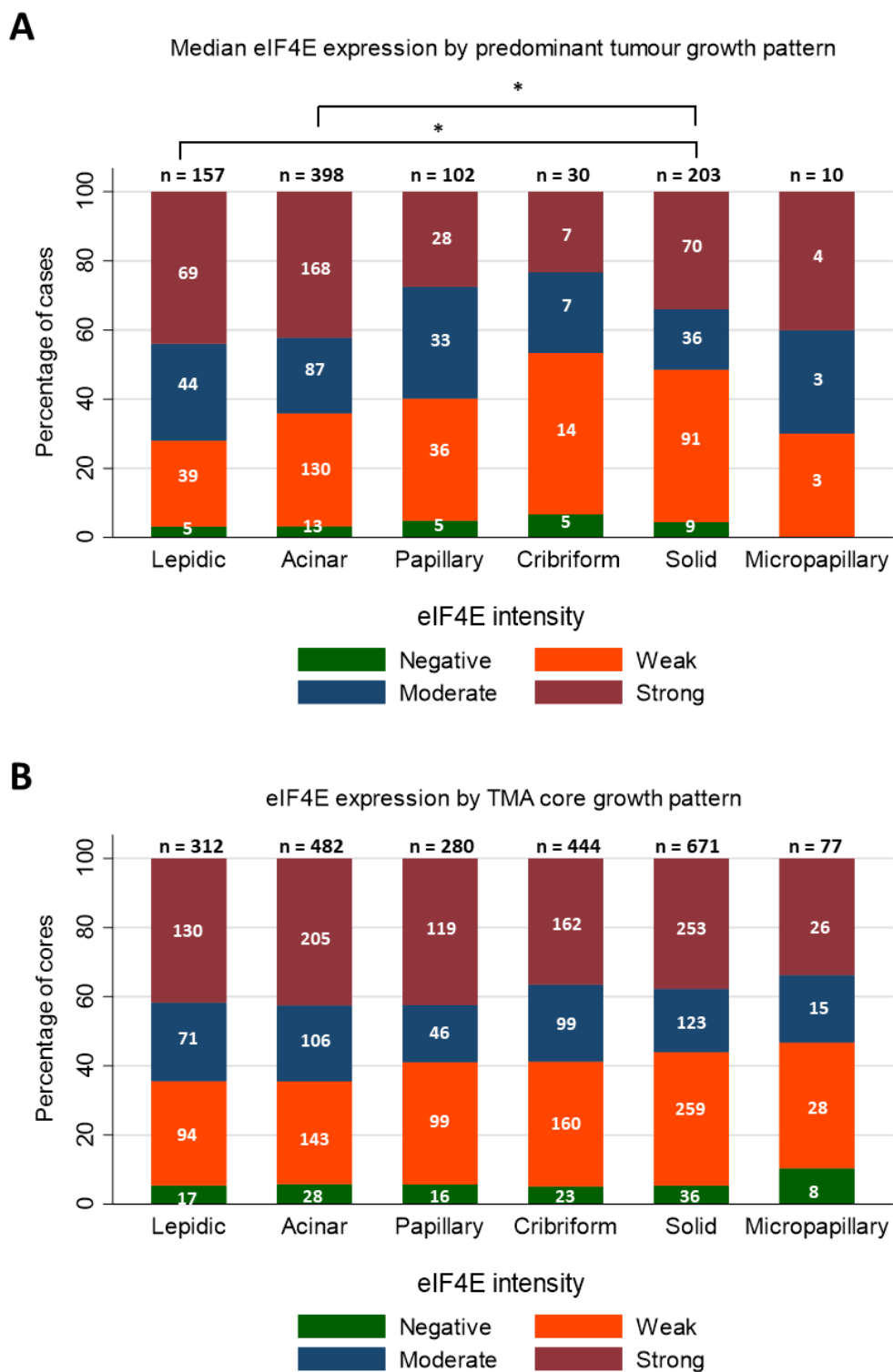
Figure 6.8. Distribution of eIF4E expression by lung adenocarcinoma tumour growth pattern.
Each bar displays the tumour growth pattern of lung adenocarcinoma and each segment of the bar represents the intensity of eIF4E. Total number of growth pattern cases is indicated at the top of each bar and the number of cases displaying the different degree of eIF4E expression is shown within each section of the bar. Degree of significance are indicated in the graphs by bars with stars above them (* = $p < 0.05$). A: Percentage of median eIF4E expression intensity by predominant tumour growth pattern per case, n= number of case. B: Percentage of eIF4E expression intensity by TMA core growth pattern, n= number of core.

I identified an important confounding factor affecting eIF4E immunohistochemistry. eIF4E negative cases were much more commonly found in old tumour blocks aging before 2009 (Figure 6.9). This suggests that detectable eIF4E antigen is degraded over time within the block, and block age is therefore a confounding factor. Thus, I decided not to include eIF4E analyses further. I further assessed whether the age of the block correlated with expression of proteins for all the other antibodies in this study and observed no correlation with the age of the block (data not shown).
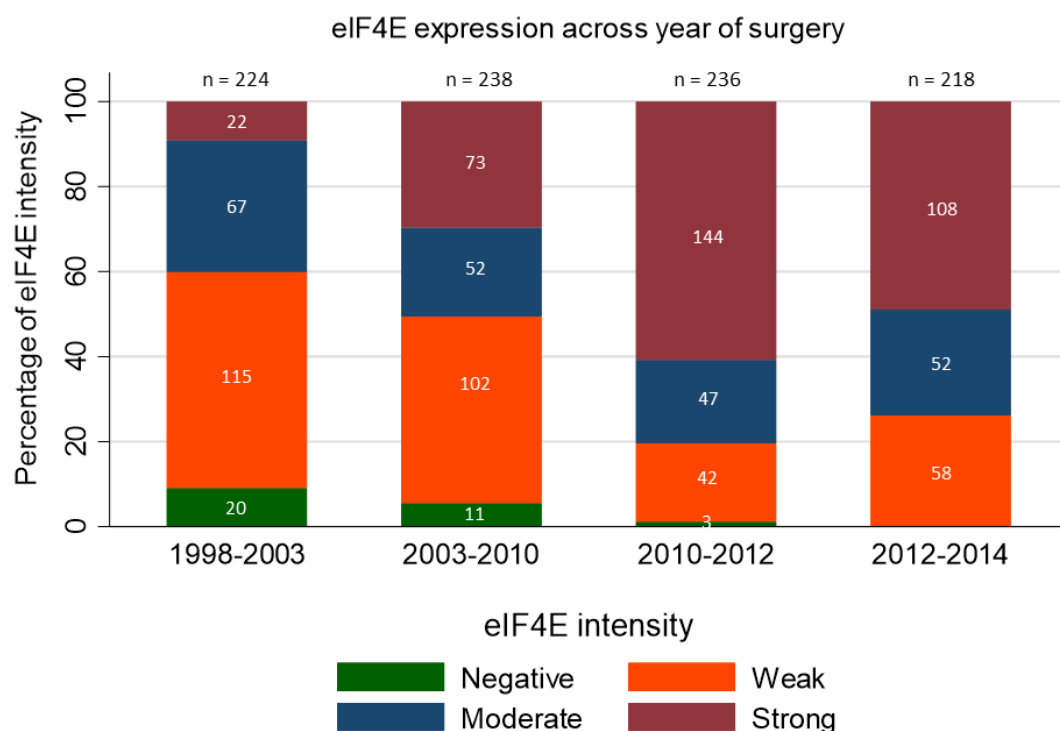
Figure 6.9. Donor age block is a confounding factor for eIF4E expression in lung adenocarcinoma.
Percentage of eIF4E expression across the year of surgery. Each bar displays the year of surgery divided in 4 groups and each segment of the bar represents the intensity of eIF4E. Total number of donor age block cases is indicated at the top of each bar and the number of cases displaying the different degree of eIF4E expression is shown within each section of the bar.

6.3.4 Association between translational control and EMT markers

6.3.4.1 *eIF4B correlation with lung adenocarcinoma patient outcome*

I assessed relationships between overall, cancer-specific and recurrence-free patient survivals and eIF4B expression. Similarly to chapter 5, for the survival analysis, Kaplan-Meier (KM) graphs describe the percentage of survival of 942 cases according to eIF4B expression over a period of 5-year follow-up after surgery. Expression levels of the translation factor was regrouped into two groups: negative and positive groups (positive group gathers weak, moderate and strong eIF4B groups).

The positive eIF4B group presents a worse significant outcome with 36% increase in risk of death when eIF4B is overexpressed in primary lung adenocarcinoma (Figure 6.10).

	Median survival	**Overall survival univariate Cox regression model**				
	(years)	**Variables**	**Number of events** (total number of observations)	**Hazard ratio**	**95% CI**	**P-value**
eIF4B	Negative = 4.5 Positive = 3.2	**eIF4B** Positive vs Negative	518 (901)	1.36	1.05 - 1.75	0.018
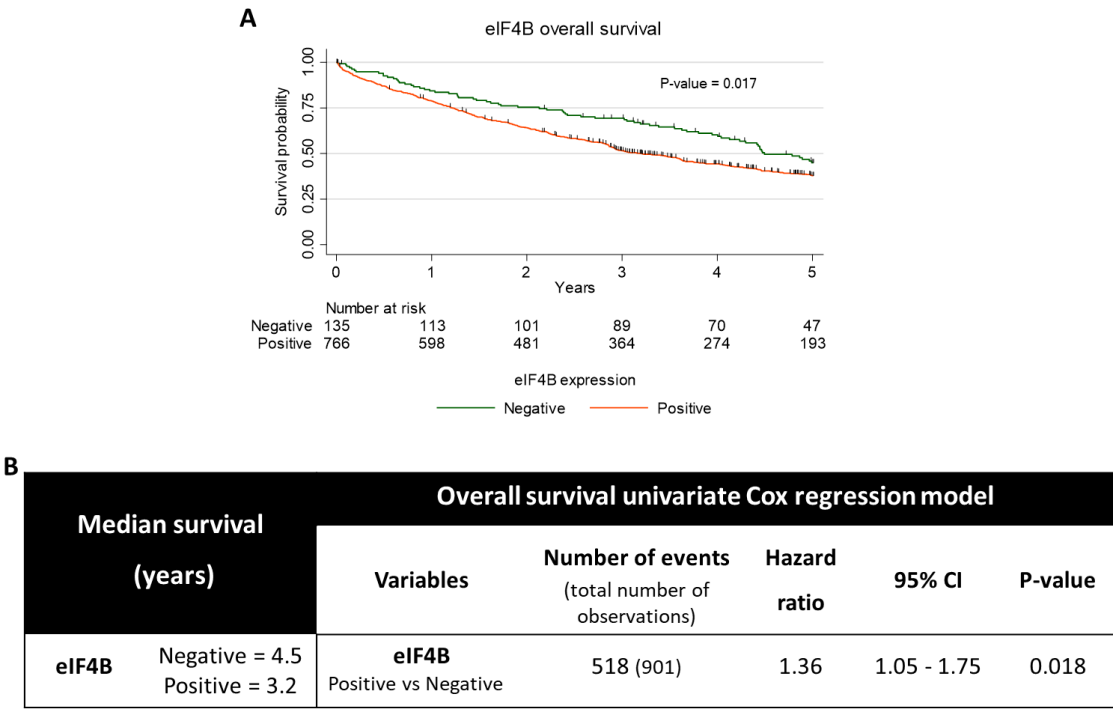
Figure 6.10. eIF4B expression and overall patient survival over a 5-year period post-surgery.
A: Overall KM survival graphs of eIF4B over 5-year follow-up after surgery in lung adenocarcinoma patients. P-values of log-rank tests are displayed on the survival graphs. B: The table shows the median survival and the results from the univariate Cox proportional hazards regression in lung adenocarcinoma patients using STATA. 95% CI: 95% confidence interval.

The results for the two next endpoints are quite similar but eIF4B overexpression is not significantly associated with poor prognosis in cancer-specific and recurrence-free survival models in lung adenocarcinoma patients (Figure 6.11).

A

eIF4B cancer-specific survival

P-value = 0.115

Number at risk

Negative	135	113	101	89	70	47
Positive	766	598	481	364	274	193

eIF4B expression

——— Negative   ——— Positive

B

Median survival (years)		Cancer-specific survival univariate Cox regression model				
		Variables	Number of events (total number of observations)	Hazard ratio	95% CI	P-value
eIF4B	Negative = - Positive = -	eIF4B Positive vs Negative	335 (901)	1.28	0.94 - 1.74	0.116

C

eIF4B recurrence-free survival

P-value = 0.140

Number at risk

Negative	135	102	81	62	50	33
Positive	766	496	353	249	168	120

eIF4B expression

——— Negative   ——— Positive

D

Median survival (years)		Recurrence-free survival univariate Cox regression model				
		Variables	Number of events (total number of observations)	Hazard ratio	95% CI	P-value
eIF4B	Negative = 4.1 Positive = 2.9	eIF4B Positive vs Negative	444 (901)	1.22	0.94 - 1.58	0.137

Figure 6.11. eIF4B expression and cancer-specific and recurrence free survivals over a 5-year period post-surgery.

A and C: Cancer-specific and recurrence-free KM survival graphs of eIF4B respectively over a 5-year follow-up after surgery in lung adenocarcinoma patients. P-values of log-rank tests are displayed on the survival graphs. B and D: Table showing the median survival and the results from the univariate Cox proportional hazards regression in lung adenocarcinoma patients using STATA. 95% CI: 95% confidence interval.

6.3.4.2 Association between translation factors and EMT

I next tested the association between EMT markers and translation factors such as the eIF4F complex (eIF4E, eIF4A1, eIF4A2 and eIF4G), eIF4B, phospho-eIF2α and phospho-eIF4E.

Spearman's rank correlation	E-cadherin	N-cadherin	Cytokeratin	Vimentin	Twist	EMT
eIF4A1	0.12*	-	- 0.15*	0.06	0.34*	0.20*
eIF4A2	-	-	-	-	-	-
eIF4B	0.12*	0.08	-	-	0.16*	0.11*
eIF4G	0.13*	-	-	0.13*	0.30*	0.23*
Phospho-eIF2α	0.09*	0.21*	- 0.13*	-	0.18*	0.17*
Phospho-eIF4E	0.14*	- 0.08	- 0.07	0.11*	0.24*	0.14*

Table 6.1. Correlation of EMT markers with translation factors factors.
The table shows the coefficient of correlations rho between translation factors and each EMT marker as well as EMT score. All correlation present in the table are pairwise and statistically significant, p < 0.05. * denotes the retention of correlations after Bonferroni correction. Maximun core number = 1,389 cores.

As I saw in chapter 5, associations with E-cadherin are difficult to interpret, as E-cadherin is associated with numerous EMT markers despite its biological function. The Spearman's rank correlation reveals that phospho-eIF2α is the only translation factor that indicates strong and significant correlation after Bonferroni correction with all individual EMT markers (except vimentin) and EMT score. Of note, phospho-eIF4E also demonstrates positive association with vimentin, TWIST as well as EMT score and negative link with cytokeratin (despite the diseappearance of this relationship after Bonferonni correction). This supports a model in which phospho-eIF2α and phospho-eIF4E may both promote EMT in lung adenocarcinoma by directly or indirectly facilitating expression of mesenchymal proteins and repressing expression of cytokeratin.

6.3.5 Multiplex immunofluorescence shows micro-anatomical asscociations between EMT and translational dysregulation

I wanted to investigate whether associations observed between EMT and translational dysregulation could be also confirmed at single cell level. To achieve this, the translation initiation factors showing the most significant associations with EMT markers were selected. Phospho-eIF2α disclosed significant strong correlation with cytokeratin and N-cadherin and eIF4B also showed association with N-cadherin (phospho-eIF4E data were unavailable at that time). Therefore I chose cytokeratin, N-cadherin, phospho-eIF2α and eIF4B as the basis for a fluorescent multiplex assay.

The MRC histology core facility and I performed a fluorescent multiplex assay where we simultaneously stained for cytokeratin, N-cadherin, phospho-eIF2α and eIF4B in whole sections of six primary lung adenocarcinomas, representing each predominant growth pattern. After scanning the whole tumour sections, we selected regions with evidence of active partial EMT (i.e. up-regulation of N-cadherin expression). Selected tumour regions were trained to obtain single cell data. Cells were segmented into three compartments using our protein of interest expression. DAPI staining helped to identify the nucleus; cytokeratin, eIF4B and phospho-eIF2α expression helped to delimit the cytoplasm compartment and N-cadherin expression was used to delineate the cellular membrane. Tumour cells were then identified using these markers and phenotyped as tumour cells (all the other cell types present in the regions were non-tumour cells and phenotyped as 'stroma' cells). Selected regions were analysed; single cell data were obtained and statistical analyses were performed. In one case only I found clear evidence of focal EMT affecting some cells but not others; this was an adenocarcinoma of mixed papillary/micropapillary growth pattern and I present these findings.

Figure 6.12. Immunofluorescent multiplex of translational reprogramming in active focal partial EMT region.
A: Composite image of the translation factors and EMT markers. Nuclei are stained with DAPI in blue. B: Deconvolutions of eIF4B, N-cadherin, cytokeratin and phospho-eIF2α and fluorescent staining are shown in red, in yellow, in green and in magenta respectively. Scale bar = 250 μm.

As shown in Figure 6.12A, I detect focal EMT phenomenon with epithelial tumour cells expressing cytokeratin and/or N-cadherin expression. Phospho-eIF2α shows heterogeneity in expression, while eIF4B is homogeneously expressed in this portion of the tumour, suggesting either a constant basal level of eIF4B expression, or that our assay is unable to detect meaningful differences in eIF4B expression which may be present.

Figure 6.13. Correlations between translation factors and EMT markers in active focal partial EMT region from single cell data.
A, B, C, D and E are two-ways scatter plots of the mean of N-cadherin versus the mean of cytokeratin expression, the mean of phospho-eIF2α versus the mean of cytokeratin expression, the mean of eIF4B versus the mean of cytokeratin expression, the mean of phospho-eIF2α versus the mean of N-cadherin expression and the mean of eIF4B versus the mean of N-cadherin expression respectively from the single-cell data. Spearman's rank correlation test was performed. Rho and p-values of this test are displayed on the plots. Number of tumour cells analysed = 3,709 cells.

As I have picked a region showing partial EMT to the eye, I wanted to see whether our quantitative method can detect a negative association between N-cadherin and cytokeratin. Surprisingly, a positive relationship is observed between the mean cellular cytoplasmic pixel intensities of cytokeratin and N-cadherin in this region (Figure 6.13A), even though the scatter plot superficially suggests a negative relationship. This may be due to the fact that most cells are clustered near 0 expression, and have a very low background signal with some positive correlation, such that the truly inversely related signals from high-expressing tumour cells are drowned out. I therefore decided to focus on an even smaller area with marked divergence and very few stromal cells:
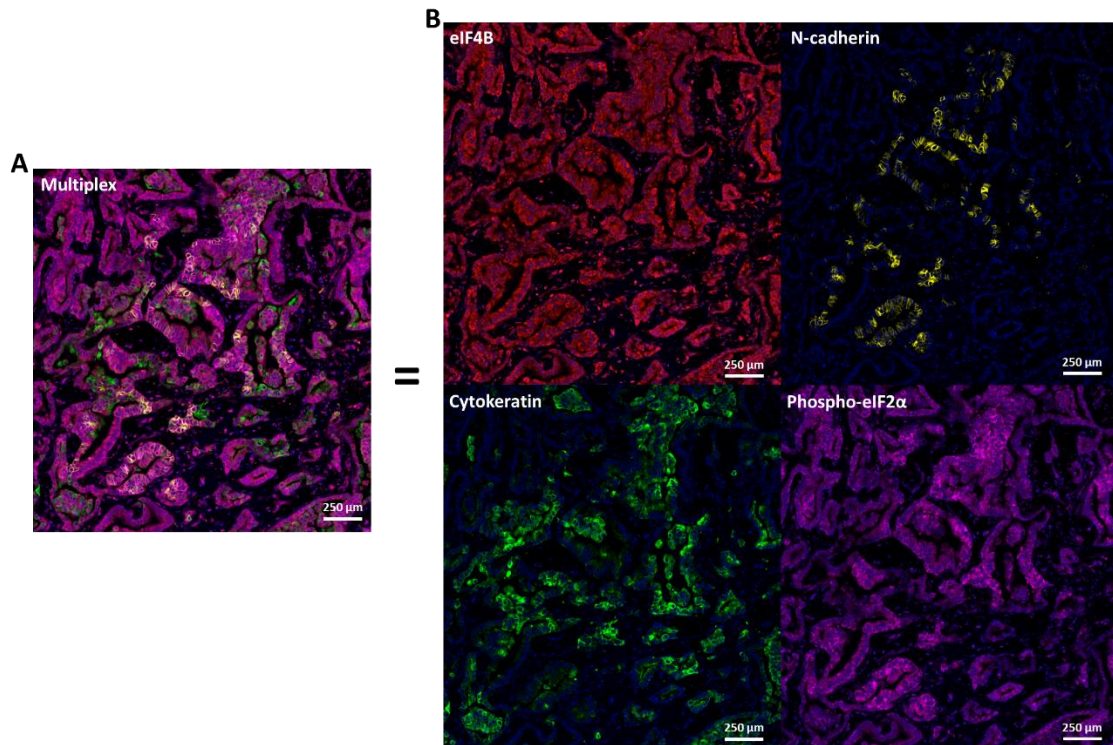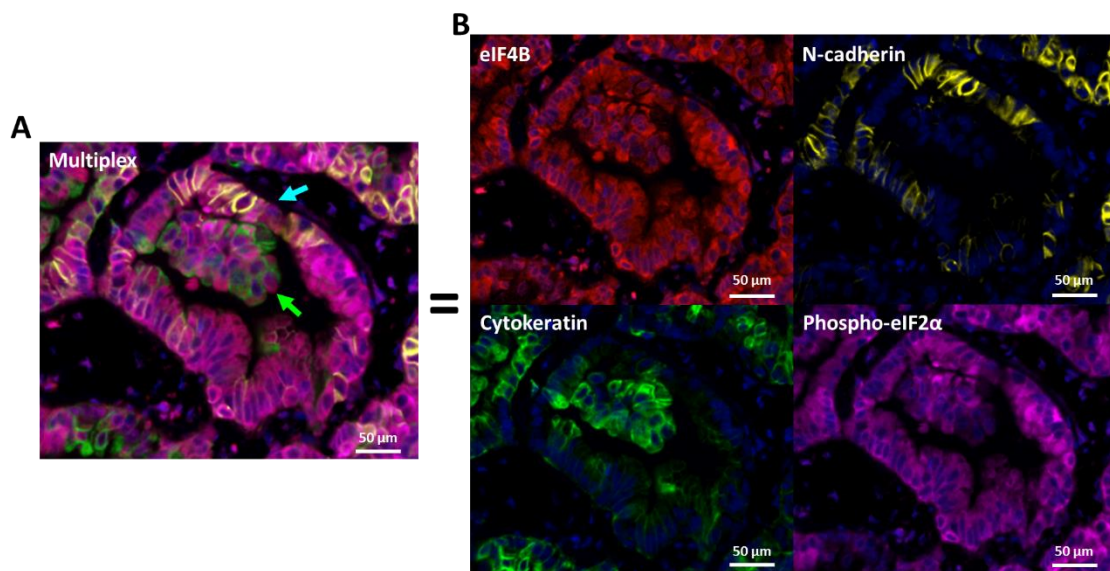

Figure 6.14. Immunofluorescent multiplex of translational reprogramming in a smaller active focal partial EMT region.
A: Composite image of the translation factors and EMT markers. Nuclei are stained with DAPI in blue. Green and blue arrows point out areas of the micropapilli and sessile growths respectively. B: Deconvolutions of eIF4B, N-cadherin, cytokeratin and phospho-eIF2α and fluorescent staining are shown in red, in yellow, in green and magenta respectively. Scale bar = 50 μm.

In this smaller area of active partial EMT, eIF4B continues to be homogenously expressed. When N-cadherin is expressed, cytokeratin expression is either reduced or lost while phospho-eIF2α is often elevated. Of note, there is evidence of partial EMT accompanied by elevated phospho-eIF2α in areas of sessile growth. In contrast, micropapillae show a retained epithelial phenotype (as explored in chapter 4 section 4.3.4), which is here accompanied by absence of detectable phospho-eIF2α (Figure 6.14).
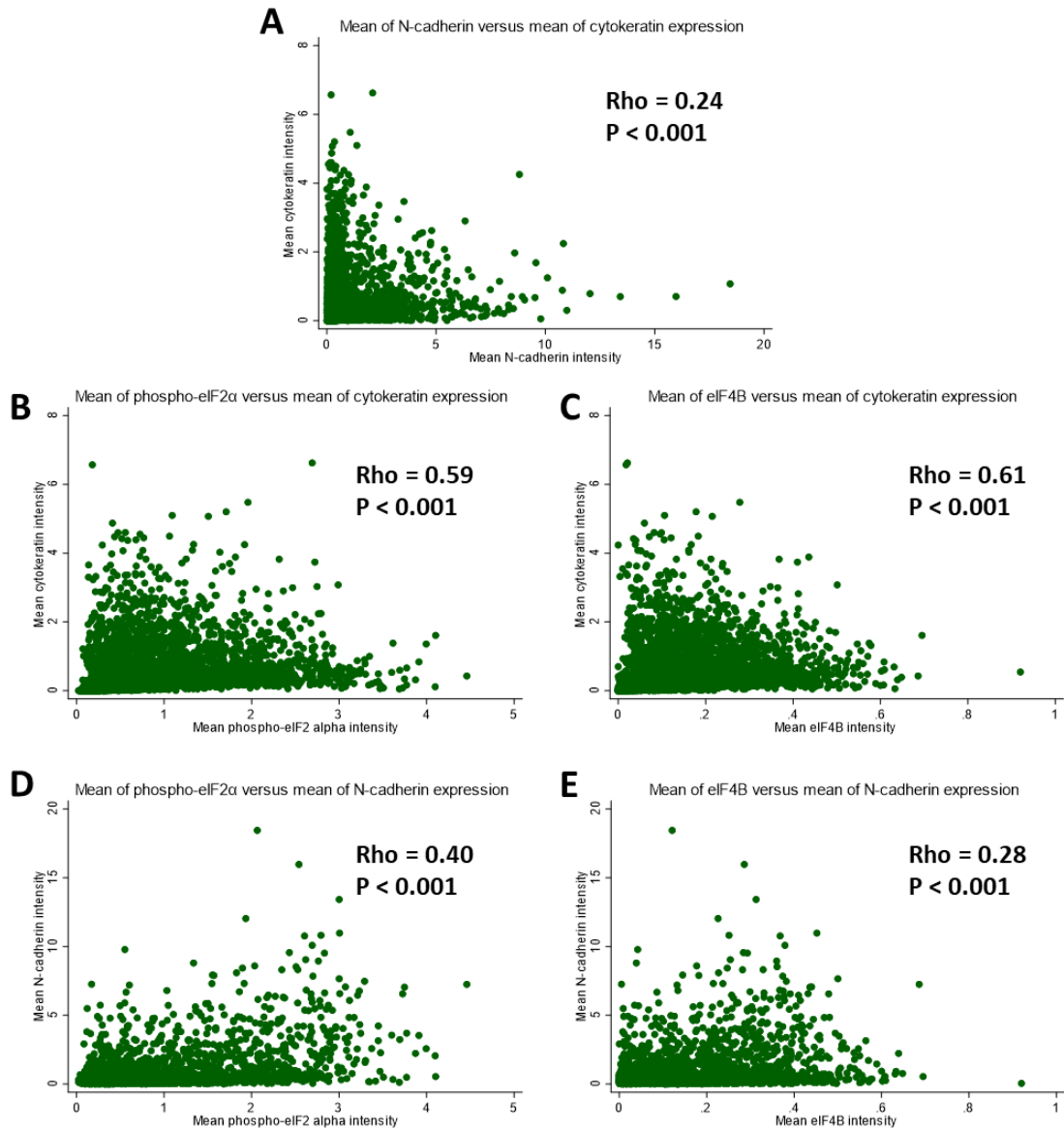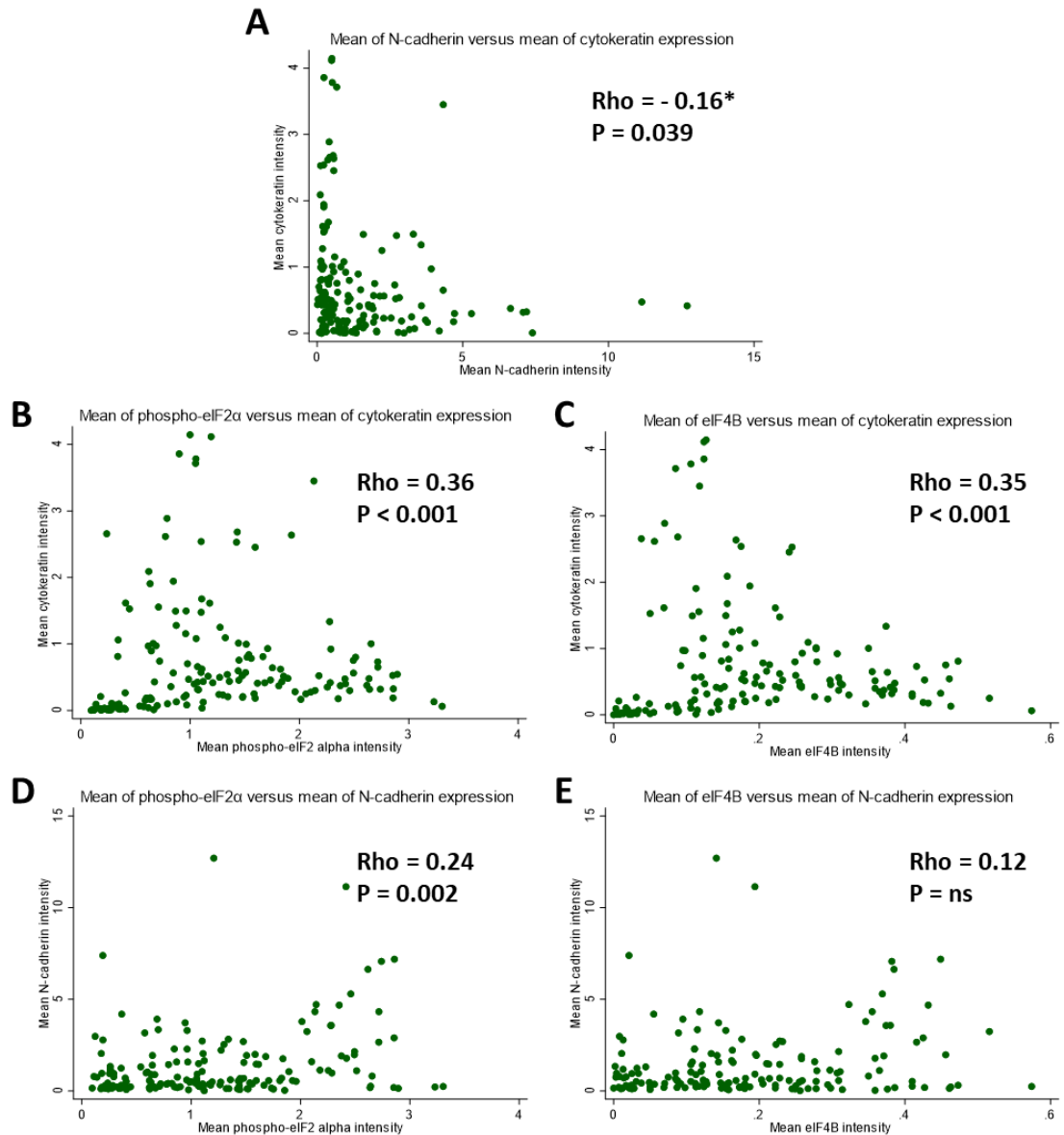
Figure 6.15. Correlations between translation factors and EMT markers in active focal partial EMT region from single cell data.
A, B, C, D and E are two-ways scatter plots of the mean of N-cadherin versus the mean of cytokeratin expression, the mean of phospho-eIF2α versus the mean of cytokeratin expression, the mean of eIF4B versus the mean of cytokeratin expression, the mean of phospho-eIF2α versus the mean of N-cadherin expression and the mean of eIF4B versus the mean of N-cadherin. Spearman's rank correlation with test was performed. Rho and p-values of this test are displayed on the plots. Number of tumour cells analysed = 161 cells. *denotes the disappearance of correlation after Bonferroni correction.

In this smaller field, there is a detectable significant negative association between N-cadherin and cytokeratin, indicating detectable active partial EMT (Figure 6.15A).

eIF4B and phospho-eIF2α are significantly positively associated with cytokeratin, which is unexpected as cytokeratin showed negative correlation with phospho-eIF2α in Table

6.1. N-cadherin expression demonstrates a significant positive link with phospho-eIF2α, similarly as observed in Table 6.1, suggesting that phospho-eIF2α may contribute to N-cadherin expression (Figure 6.15). The single cell data analysis does not show evidence of translational dysregulation favouring EMT activation in lung adenocarcinoma. Further analysis would be necessary to confirm this hypothesis.

6.3.6   Re-assessment of local EMT models with translation markers

C1 and C2 tumours represent genomically and biologically different types of mixed *in situ*/invasive tumours growth. Invasion in C1 tumours represents genomic progression with acquisition of additional mutations in the invasive part. C2 tumours are composed of an invasive central area surrounded by a zone of invasion-competent cells that grow out in a lepidic manner, colonising the alveolar surface. Tumour cells of both *in situ* and invasive components of C2 tumours are genomically identical. In chapter 4, our data has demonstrated that C1 tumours displayed significant evidence of additional partial EMT in the invasive region of these tumours while C2 tumours did not, further supporting biological and genomic progression in these lesions.

I wanted to explore whether changes in translation factor expression may contribute focal invasiveness in these tumours. To realise this, the two TMAs containing 34 C1 and 17 C2 tumours from chapter 4 were immunohistochemically stained for eIF4A1/2, eIF4B, eIF4G and phospho-eIF2α. Digital images of TMA sections were then manually scored for their expression using an H-score system.

6.3.6.1 Translation initiation factors distribution in C1 and C2 tumours

Figure 6.16. Representative immunohistochemical staining of eIF4A1 expression in C1 and C2 tumours. Core images of eIF4A1 expression in C1 and C2 tumours. Top and bottom rows show eIF4A1 expression in *in situ* (left column) and invasive (right column) areas of C1 and C2 tumours respectively. Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents eIF4A1 expression. Core image magnification = 5X. Scale bar = 500 µm. Inset squares represent high power (40x) of a portion of the core. Scale bar = 50 µm.

In this example, eIF4A1 is equally expressed in the C1 in *situ* and invasive tumour areas. In this C2 tumour, eIF4A1 appears to be quite strongly expressed in the *in situ* compared to the invasive regions (Figure 6.16).

Figure 6.17. Representative immunohistochemical staining of eIF4A2 expression in C1 and C2 tumours. Core images of eIF4A2 expression in C1 and C2 tumours. Top and bottom rows show eIF4A2 expression in *in situ* (left column) and invasive (right column) areas of C1 and C2 tumours respectively. Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents eIF4A2 expression. Core image magnification = 5X. Scale bar = 500 μm. Inset squares represent high power (40x) of a portion of the core. Scale bar = 50 μm.

eIF4A2 is equally distributed in *in situ* and invasive areas in this C1 tumour while it is slightly overexpressed in *in situ* related to invasive component of C2 tumour in this example (Figure 6.17).
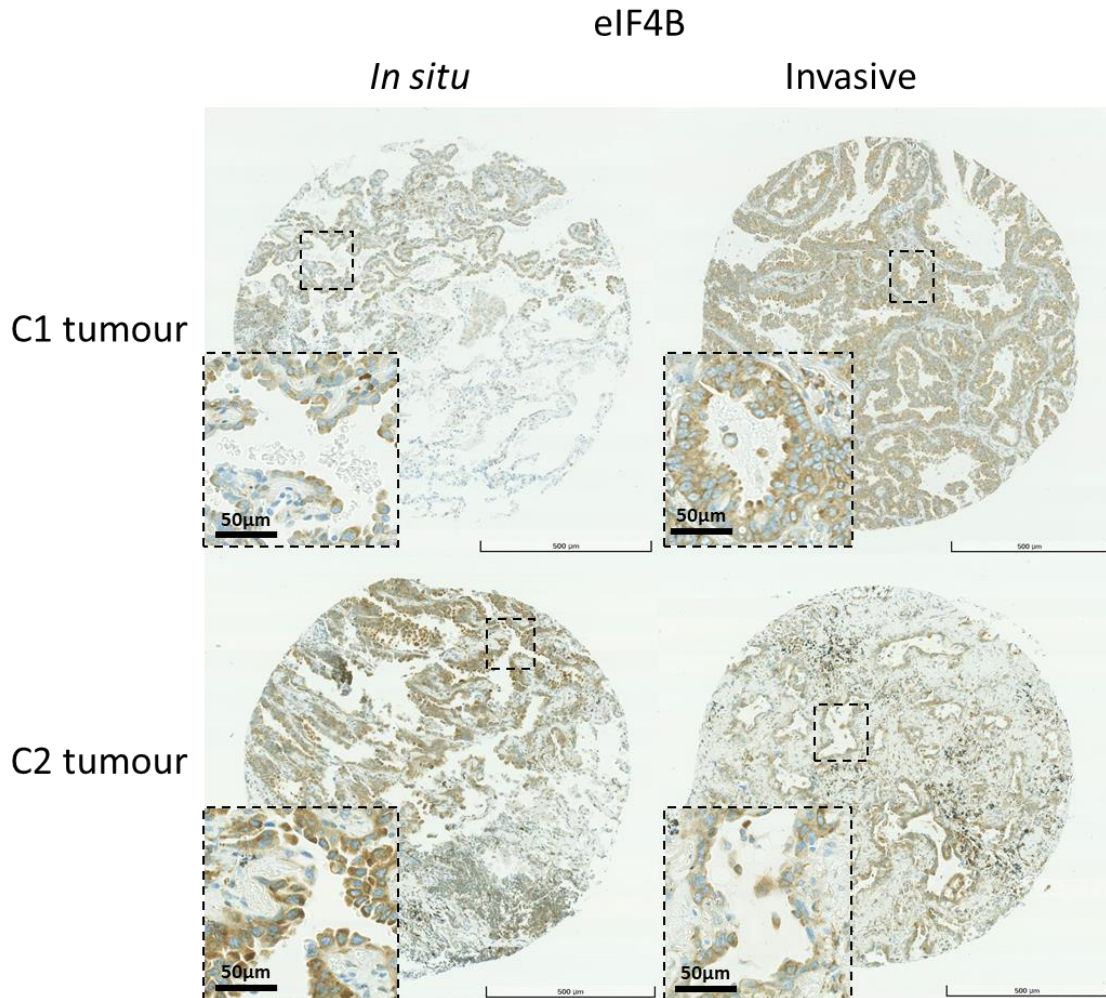
Figure 6.18. Representative immunohistochemical staining of eIF4B expression in C1 and C2 tumours. Core images of eIF4B expression in C1 and C2 tumours. Top and bottom rows show eIF4B expression in *in situ* (left column) and invasive (right column) areas of C1 and C2 tumours respectively. Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents eIF4B expression. Core image magnification = 5X. Scale bar = 500 μm. Inset squares represent high power (40x) of a portion of the core. Scale bar = 50 μm.

This representative immunohistochemical staining of eIF4B reveals that this translation factor is more expressed in the invasive than the *in situ* areas of C1 tumour whereas it is the opposite in C2 tumour (Figure 6.18).

eIF4G

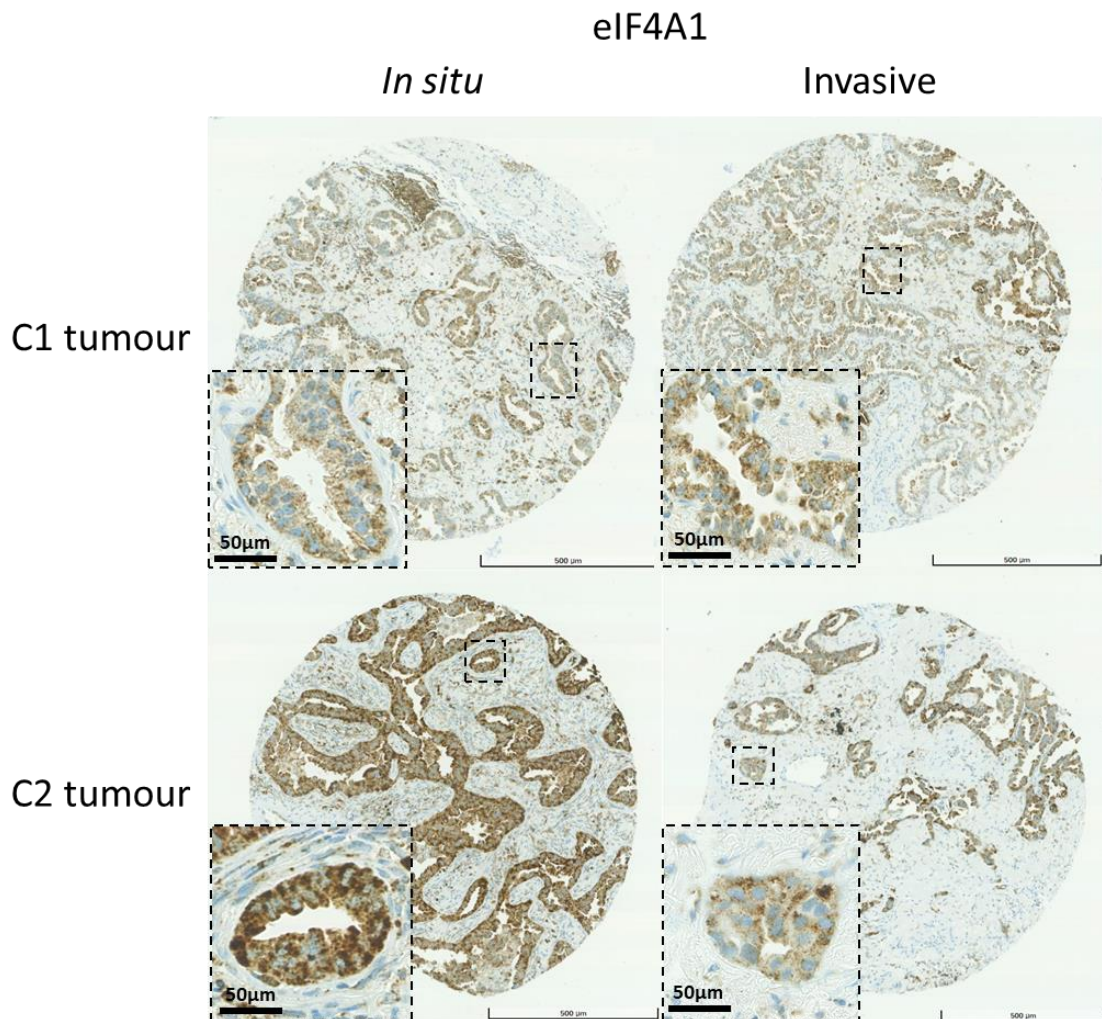*In situ*   Invasive
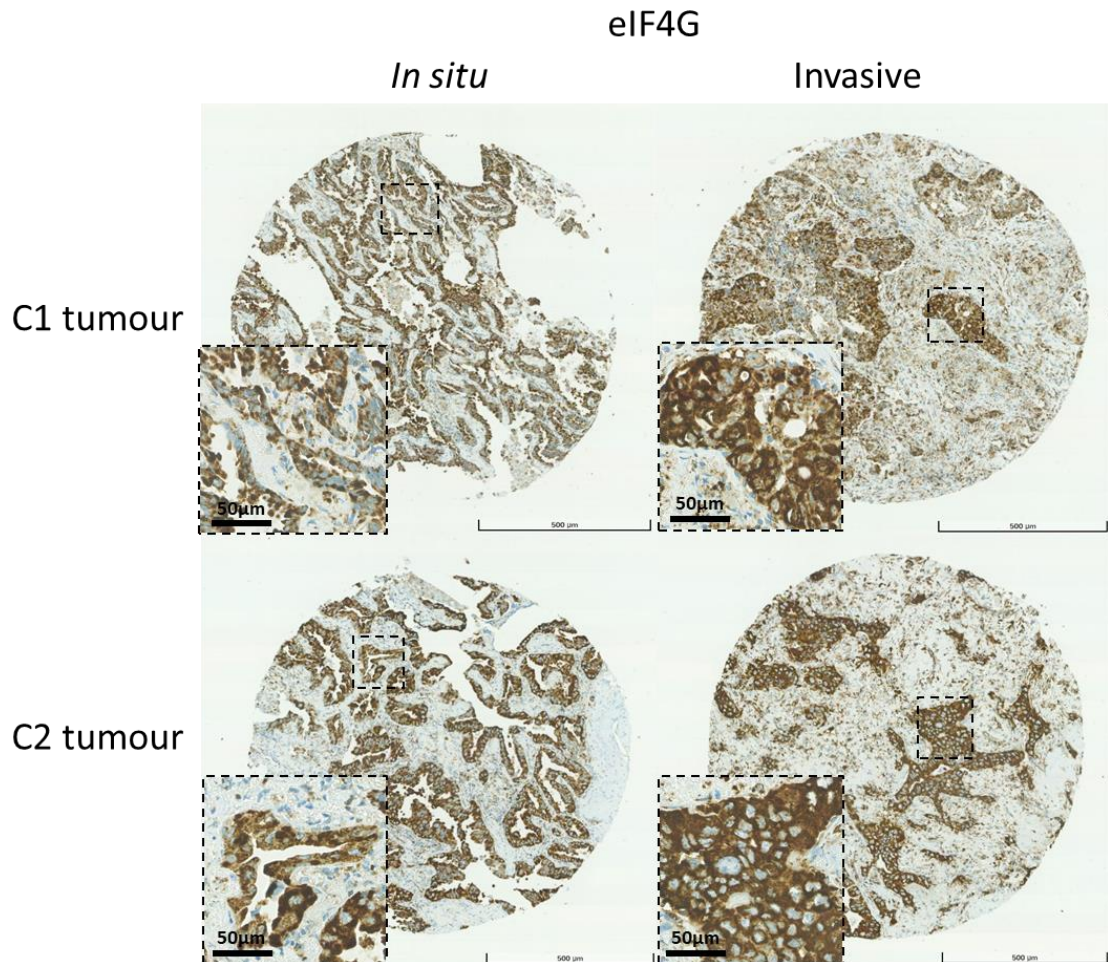
C1 tumour

C2 tumour

Figure 6.19. Representative immunohistochemical staining of eIF4G expression in C1 and C2 tumours.
Core images of eIF4G expression in C1 and C2 tumours. Top and bottom rows show eIF4G in *in situ* (left column) and invasive (right column) areas of C1 and C2 tumours respectively. Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents eIF4G expression. Core image magnification = 5X. Scale bar = 500 μm. Inset squares represent high power (40x) of a portion of the core. Scale bar = 50 μm.

eIF4G is generally strongly expressed in invasive component compared to the *in situ* areas in these C1 and C2 tumours (Figure 6.19).
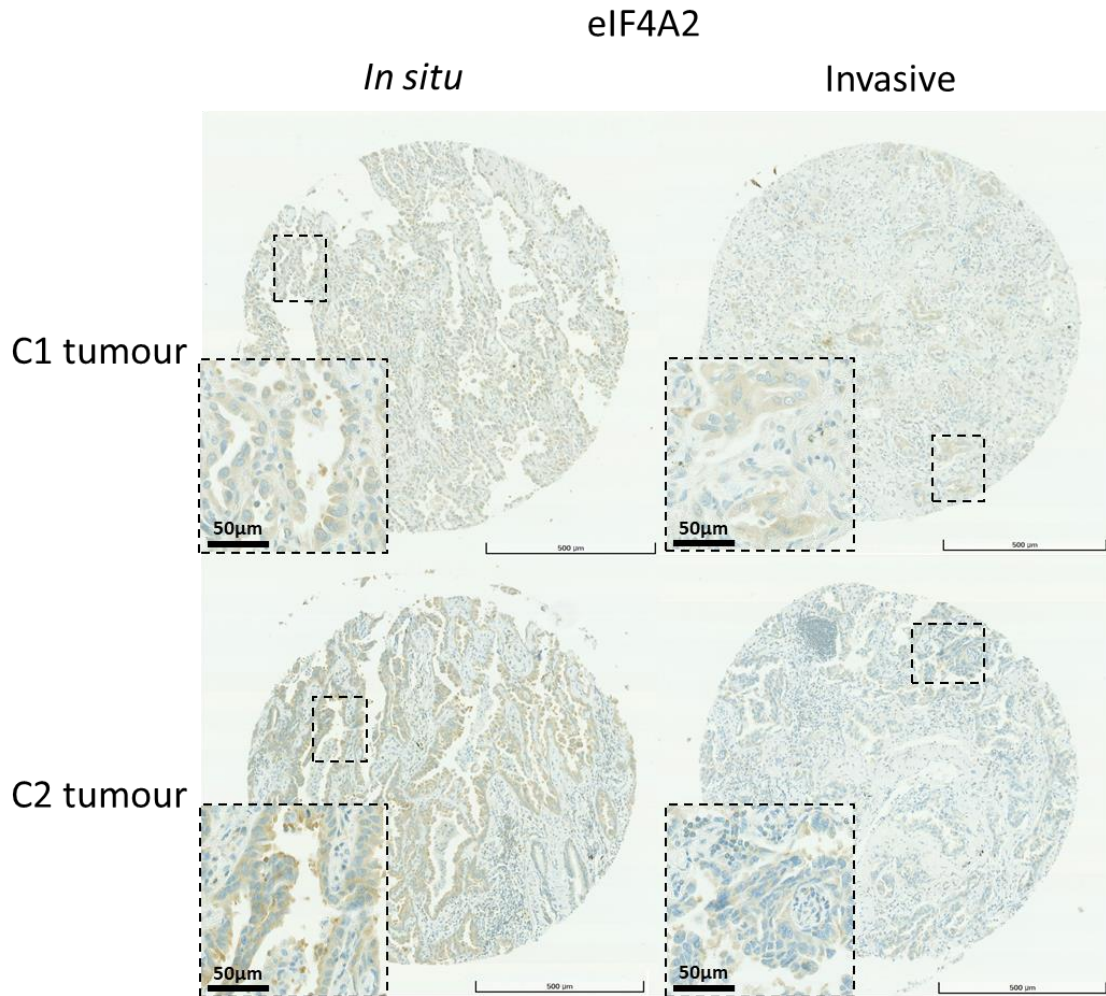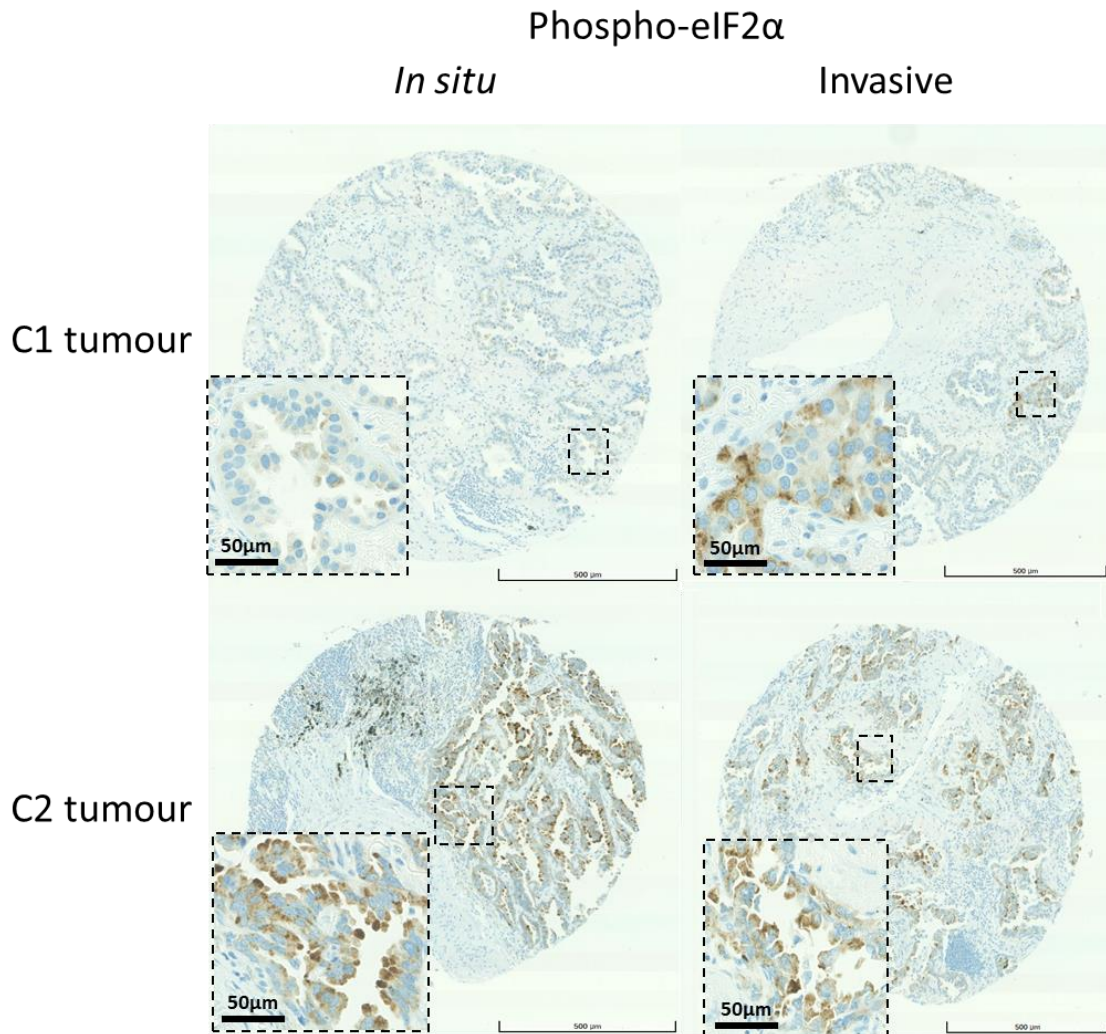
Figure 6.20. Representative immunohistochemical staining of phospho-eIF2α expression in C1 and C2 tumours.

Core image of phospho-eIF2α expression in C1 and C2 tumours. Top and bottom rows show phospho-eIF2α expression in *in situ* (left column) and invasive (right column) areas of C1 and C2 tumours respectively. Blue staining (haematoxylin and bluing reagent) represents nuclei. Brown staining is DAB staining and represents phospho-eIF2α expression. Core image magnification = 5X. Scale bar = 500 μm. Inset squares represent high power (40x) of a portion of the core. Scale bar = 50 μm.

In this representative staining of the translation factor, phospho-eIF2α is barely detectable in the *in situ* region of C1 tumour while it is weakly expressed in the invasive component. In this C2 tumour, phospho-eIF2α is equally detectable in both regions (Figure 6.20).

6.3.6.2 Evidence of translation dysregulation between C1 and C2 tumours

Similarly to chapter 4, Figures 6.21- 6.25 represent the distribution of translation factors in *in situ* and invasive components of C1 and C2 tumours (box plots). The slope graphs illustrate matched paired *in situ* and invasive cores per patient. The medians of the two *in situ* cores and the two invasive cores were calculated for each translation factors. To compare the difference in the expression of translation factors between C1 and C2 tumours and *in situ* and invasive components of the same tumour, Mann-Whitney and Wilcoxon signed-rank tests were performed respectively.

eIF4A1 expression is significantly higher in C2 compared to C1 tumours, consistent with a more invasive phenotype in C2 tumours ($p = 0.010$) (Figure 6.21A). However, eIF4A1 expression does not significantly differ from the *in situ* to invasive compartments in either C1 or C2 tumours (Figure 6.21B).

Figure 6.21. Distribution of eIF4A1 expression in C1 and C2 tumours.
A: Box plot of eIF4A1 expression in *in situ* and invasive compartments of C1 and C2 tumours at the core level, (n = 188 cores). P = p-value of Mann-Whitney test. B: Slope graphs representing eIF4A1 expression in *in situ* and invasive compartments of C1 and C2 tumours at the case level, (n = 47 cases). Wilcoxon signed-rank test was performed: ns = not significant

eIF4A1's paralogue eIF4A2 is slightly more highly expressed in *in situ* portions of C2 tumours but the difference in its expression between C1 and C2 tumours is not statistically significant (Figure 6.22A and B).
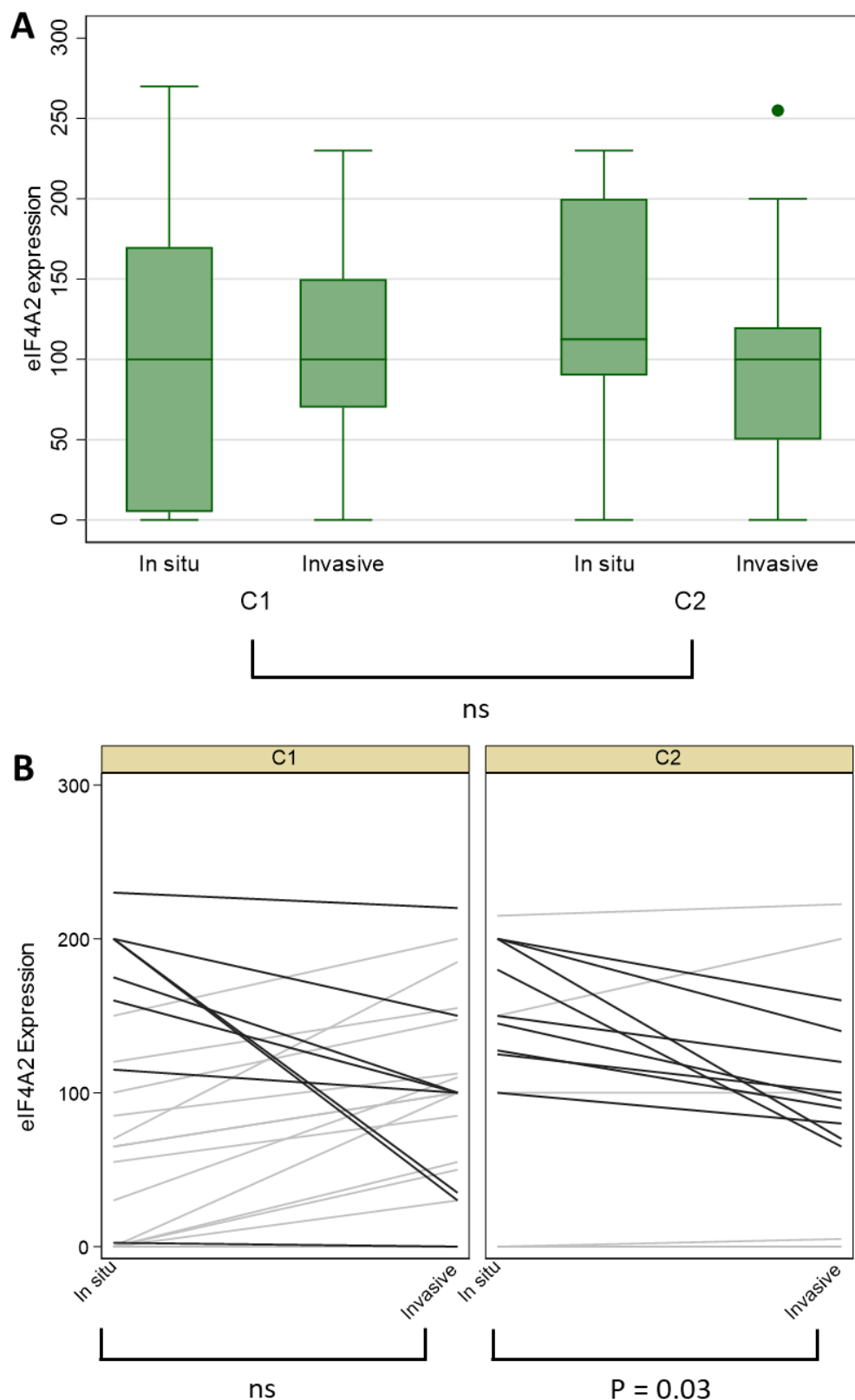
Figure 6.22. Distribution of eIF4A2 expression in C1 and C2 tumours.
A: Box plot of eIF4A2 expression in *in situ* and invasive compartments of C1 and C2 tumours at the core level, (n = 188 cores). Mann-Whitney test was performed: ns = not significant. B: Slope graphs representing eIF4A2 expression in *in situ* and invasive compartments of C1 and C2 tumours at the case level, (n = 47 cases). Wilcoxon signed-rank test was performed: ns = not significant.

eIF4B is overall more detected in C2 tumours than C1 tumours but there is no statistically significant variation in its expression between C1 and C2 tumours (Figure 6.23A) and *in situ* and invasive components of the same tumour (Figure 6.23B).
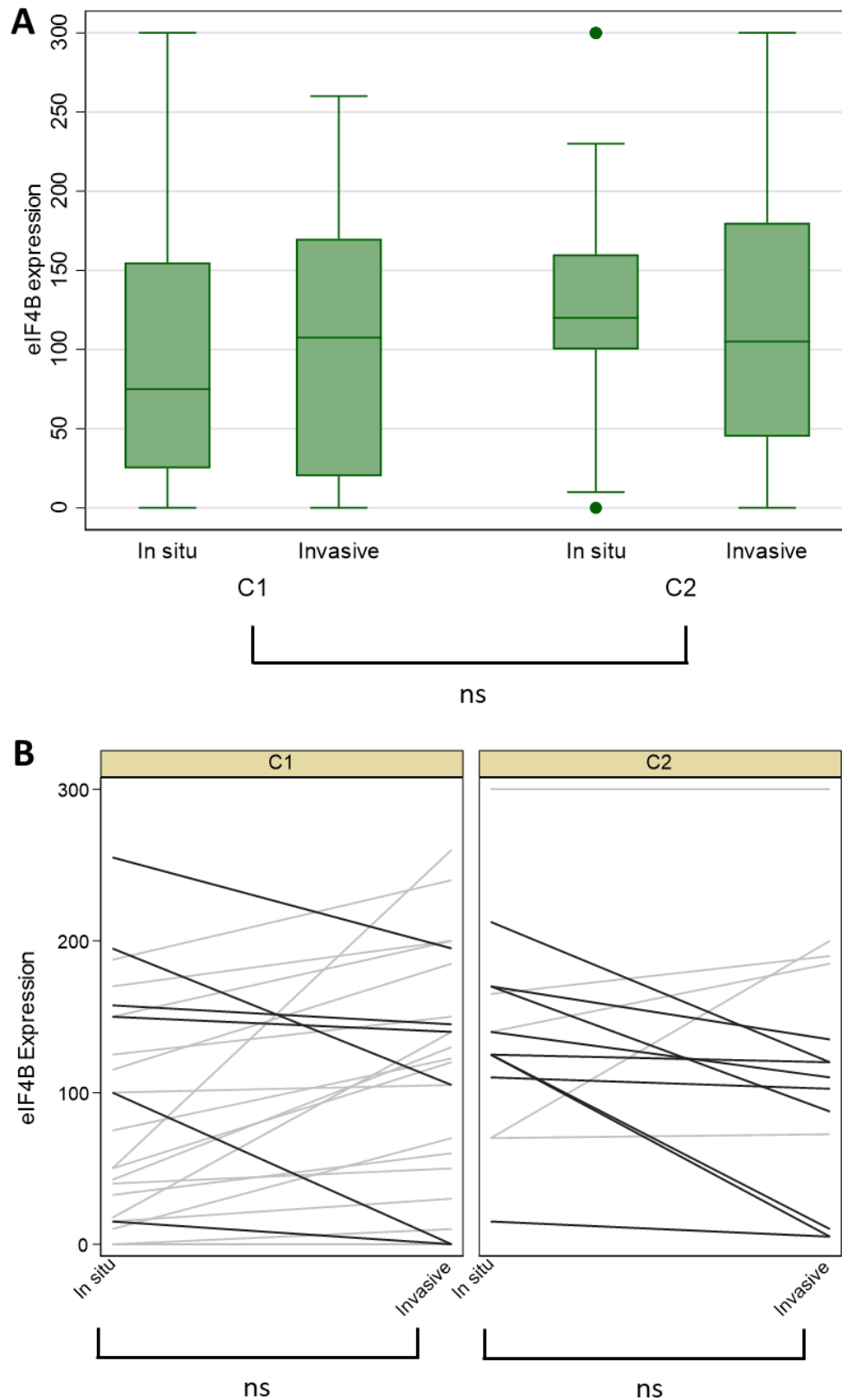
Figure 6.23. Distribution of eIF4B expression in C1 and C2 tumours.
A: Box plot of eIF4B expression in *in situ* and invasive compartments of C1 and C2 tumours at the core level, (n = 188 cores). Mann-Whitney test was performed: ns = not significant. B: Slope graphs representing eIF4B expression in *in situ* and invasive compartments of C1 and C2 tumours at the case level, (n = 47 cases). Wilcoxon signed-rank test was performed: ns = not significant.

eIF4G is significantly highly expressed in C2 tumours in comparison with C1 tumours (p = 0.005) (Figure 6.24A). However, there is no statistically signficant difference in its expression between *in situ* and invasive components of C1 and C2 tumours (Figure 6.24B).
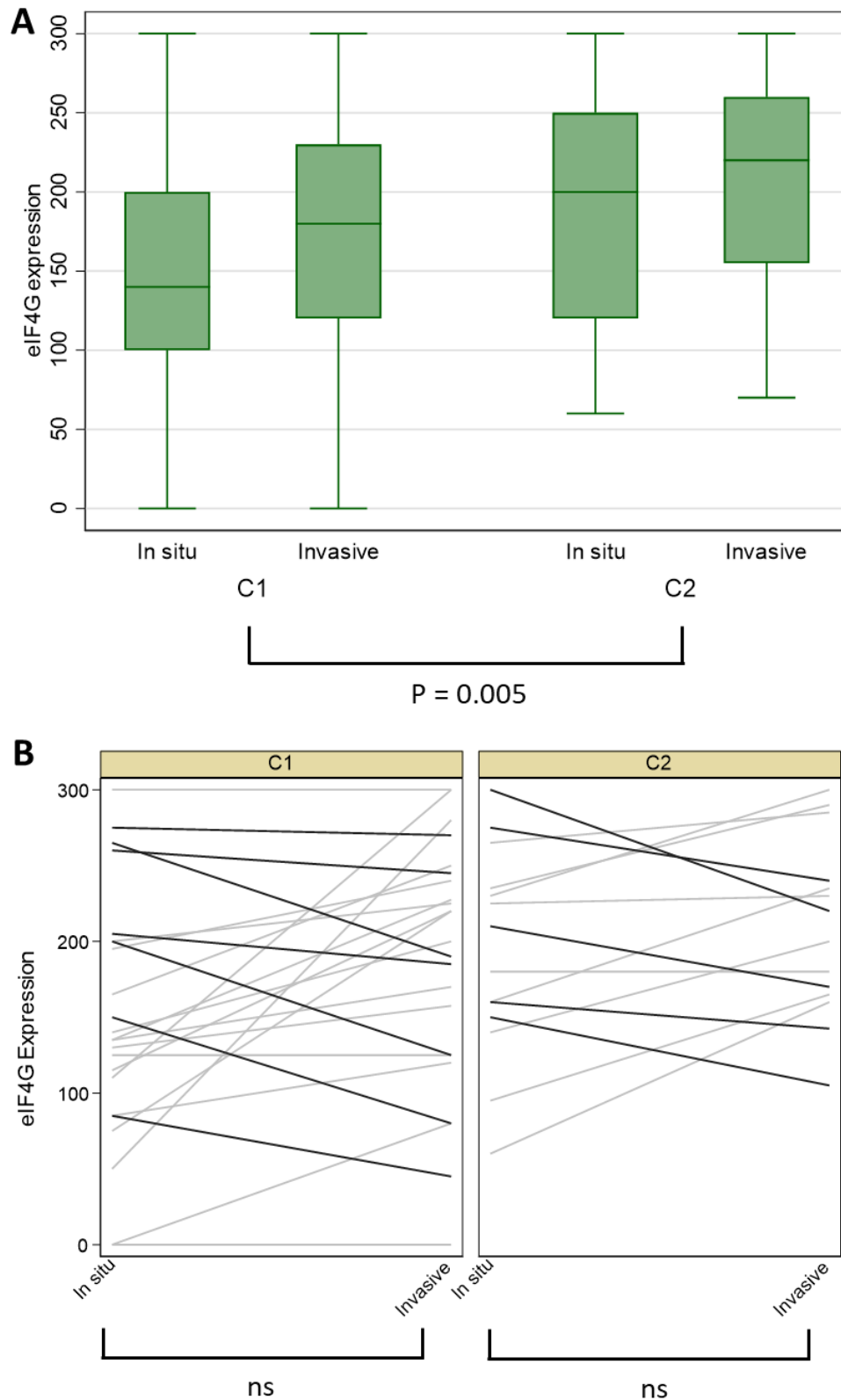
Figure 6.24. Distribution of eIF4G expression in C1 and C2 tumours.
A: Box plot of eIF4G expression in *in situ* and invasive compartments of C1 and C2 tumours at the core level, (n = 188 cores). P = p-value of Mann-Whitney test. B: Slope graphs representing eIF4G expression in *in situ* and invasive compartments of C1 and C2 tumours at the case level, (n = 47 cases). Wilcoxon signed-rank test was performed: ns = not significant.

Phospho-eIF2α expression significantly differ between C1 and C2 tumours, consistent with a more invasive phenotype in C2 tumours (Figure 6.25A). However the difference between the *in situ* and invasive regions is not statistically significant (Figure 6.25B).
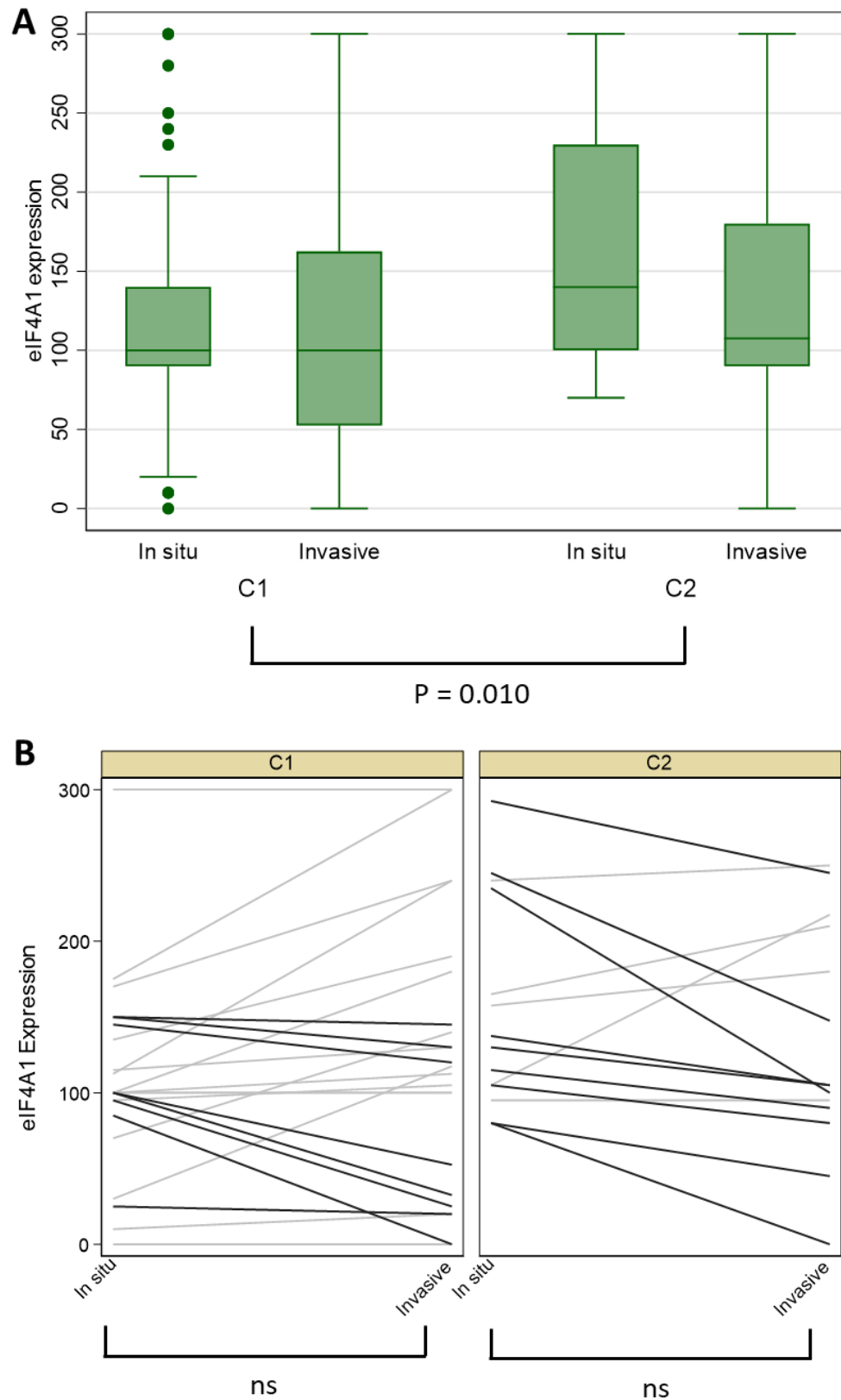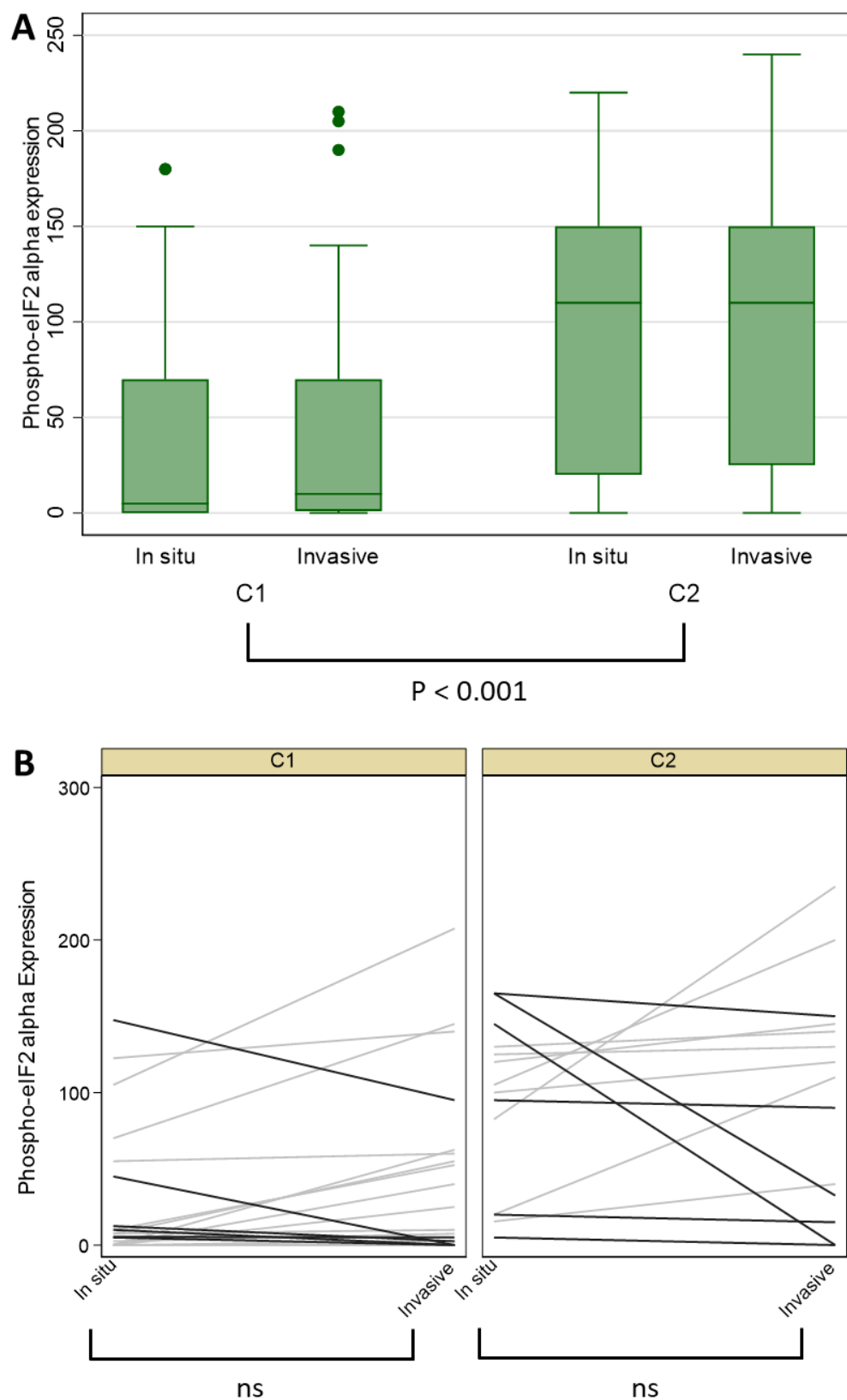
Figure 6.25. Distribution of phospho-eIF2α expression in C1 and C2 tumours.
A: Box plot of phospho-eIF2α expression in *in situ* and invasive compartments of C1 and C2 tumours at the core level, (n = 188 cores). P = p-value of Mann-Whitney test. B: Slope graphs representing phospho-eIF2α expression in *in situ* and invasive compartments of C1 and C2 tumours at the case level, (n = 47 cases). Wilcoxon signed-rank test was performed: ns = not significant.

These data show that in general all the translation initiation factors analysed in this study are more expressed in C2 in contrast to C1 tumours. Phospho-eIF2α, eIF4A1 and eIF4G are significantly up-regulated in C2 tumours which may suggest a translational reprogramming favouring proliferation and invasion in C2 tumours as these tumour cells are believed to be more proliferative and to have acquired invasive-competent behaviour than C1 tumours (Moore et al., 2019). Nonetheless, their expression does not significantly vary between *in situ* and invasive parts of C1 and C2 tumours. The significant downgregulation of eIF4A2 from the invasive to lepidic outgrowth is of uncertain biological significance.

6.4 Discussion

This chapter focuses on evidence of relationships between EMT and translation dysregulation in human primary lung adenocarcinoma.

The results from Figures 6.6 to 6.7 show that increased eIF4B expression is detectable in the solid tumour growth of lung adenocarcinoma, but eIF4E expression was difficult to interpret due to the age of the tissue block confounding the assay (Figure 6.9). eIF4B is a co-activator of the RNA helicase eIF4A and is involved in the mTOR and MAPK pathways, which are pathways promoting proliferation and survival of tumour cells (Shahbazian *et al.*, 2010). eIF4A helicase activity plays an important part in promoting pro-oncogenenic mRNA translation (Ji *et al.*, 2003; Modelska *et al.*, 2015; Raza *et al.*, 2015). Therefore, this suggests that solid tumour pattern may present translational reprogramming with increased in translation of messenger RNAs encoding proliferation and survival of tumour cells.

I find that overexpression of eIF4B predicts poor overall outcome in lung adenocarcinoma, which is a novel finding. Previous studies have also associated eIF4B up-regulation with poor prognosis in diffuse large B-cell and T-cell lymphomas as well as breast cancer (Horvilleur *et al.*, 2014; Lin and Aplan, 2007; Modelska *et al.*, 2015).

Table 6.1 shows that amongst the investigated translation initiation factors, phospho-eIF4E and phospho-eIF2α are the translation factors displaying the most significant associations with EMT-related markers in primary lung adenocarcinoma at TMA core level. Phospho-eIF4E shows a significant positive relationship with vimentin, TWIST and EMT score and a negative link with cytokeratin. Robichaud *et al*. have shown that phosphorylation of eIF4E can promote EMT by stimulating translation of matrix metalloproteinase-3 and EMT-related transcription factor SNAIL in prostate cancer cells (Robichaud *et al.*, 2015). This is consistent with our finding and may imply the importance of phospho-eIF4E in regulating the translation of EMT-related proteins in lung primary adenocarcinoma.

Phospho-eIF2α is significantly negatively associated with cytokeratin and positively related to N-cadherin, TWIST and EMT score. Phosphorylation of the α subunit of eIF2

is the consequence of several cellular stresses, and blocks generalised translation initiation while reprogramming cellular translation to survive the stressed state. It regulates the translation of a particular set of mRNAs containing an upstream open reading frame (uORF) in their 5'UTR sequence in response to cellular stress (Baird and Wek, 2012; Young and Wek, 2016). However its role in oncogenesis remains elusive. While some studies pointed out its tumour suppressor property in cancer formation (Donzé *et al.*, 1995; Meurs *et al.*, 1993), recent studies have demonstrated its oncogenic property in tumour formation and tumour invasion (Falletta *et al.*, 2017; Ye *et al.*, 2010). Our data suggests that alteration of phospho-eIF2α may favour invasion by reprogramming translation in cancer cells and regulating expression of pro-invasive genes, i.e. EMT-related genes. It would be interested to test this hypothesis in *in vitro* models by inducing EMT and check whether phospho-eIF2α is upregulated or by blocking the expression of those translation factors and verify whether EMT markers are expressed upon EMT induction at protein level.

Using multiplex immunofluorescence I were able to examine markers of EMT and translation initiation factor levels at the single cell level. The positive association of cytokeratin with both eIF4B and phospho-eIF2α was unexpected as I would expected the opposite in the smaller EMT active tumour region. This may be due to technical challenges, inability to see dynamic processes, efficacy of signal unmixing, which introduce problems. The positive link between N-cadherin and phospho-eIF2α at single cell level may reinforce the notion that phospho-eIF2α may contribute to promote EMT-related genes translation, particularly N-cadherin encoding mRNA. However, I cannot confirm the correlation observed between translation dysregulation and EMT in Table 6.1 at single cell level. Further investigation is required to understand whether translational dysregulation can promote EMT in lung adenocarcinoma. It may be worth to filter out weakly expressing cells in some other way or to improve cell segmentation/phenotyping methods.

The last section of this chapter targeted the role of translational reprogramming in EMT models identified in chapter 4. The results from Figures 6.21-6.25 display that eIF4A1, eIF4G and phospho-eIF2α are up-regulated in C2 tumours compared to C1 tumours, suggesting a more active translation process facilitating proliferation and invasion in C2

tumours as these tumours are thought to be more proliferative and invasive than C1 tumours (Moore *et al.*, 2019). This underlines that C1 and C2 tumours are biologically different, in particular in the control of mRNA translation. Nevertheless, there is no significant evidence of translational reprogramming in the transition to invasiveness in C1 tumours, which may be due to a small number of observations or inability of seeing dynamic process.

In conclusion, our data suggests that I cannot confirm that there is evidence of translational reprogramming in order to promote EMT and invasion in primary lung adenocarcinoma. Further investigation is required to understand whether translation dysregulation can favour EMT in lung adenocarcinoma.

Chapter 7.   Discussion

Chapter 7 Discussion

Lung cancer is the leading cause of cancer-related death worldwide. This is mainly due to the late diagnosis of disease when the tumour has already spread to regional sites. Metastasis is immediate cause of cancer mortality in most cancer patients and EMT is believed to be a major contributor of cancer metastasis. Deregulation of translation initiation factors has been asssociated with EMT *in vitro*, however, its association in primary tissues had not been investigated in primary lung tumour. The aim of this project was to determine whether I can quantify EMT in primary lung adenocarcinoma and find evidence of translational dysregulation promoting EMT in primary lung adenocarcinoma.

7.1 Characterisation of molecular EMT and its clinical significance in primary lung adenocarcinoma

Pathologists rarely witness morphological evidence of EMT in primary tumours. Our study has shown that partial molecular EMT (i.e. any loss or gain of epithelial or mesenchymal markers respectively) is commonly detectable in primary lung adenocarcinoma and is associated with patient physio-pathological factors and outcome. Our data show that solid and micropapillary tumour growth patterns predict poor outcome in all three end-points in primary lung adenocarcinoma, with micropapillary tumour growth being the worst prognosis. Is this related to EMT? The results from chapter 3 show that partial EMT is significantly related to solid tumour growth pattern.

When analysing the clinical significance of markers of molecular EMT in primary lung adenocarcinoma in chapter 5, loss of E-cadherin predicts overall and recurrence-free survivals independently of other EMT markers, while loss of cytokeratin or gain of TWIST expression are independently associated with a shortened remission period. Combining all EMT markers into one variable, an EMT score, is not significantly an independent predictor of poor outcome but it is associated with lymph node metastasis in lung adenocarcinoma patients. This may suggest that EMT score can favour regional lymph

node invasion in solid tumours and subsequently lead to poor outcome in lung adenocarcinoma patients.

The results from chapter 3 further show that there is evidence of different scale of heterogeneity in EMT markers expression. N-cadherin and vimentin have displayed heterogeneity at microscopic scale (i.e. within cores) while regional heterogeneity of EMT markers (i.e. between cores across the same tumour) was mostly not detected in primary tumours. In future studies, it would be fascinating to carry out genomic sequencing on EMT positive tumour TMA cores to verify whether genetic alteration promoting tumour invasion such as *TP53* gene mutations can favour EMT activation. Interestingly, micropapillary tumour growth shows comparatively much less evidence of molecular EMT. Close examination of this tumour growth indeed reveals that while the micropapillae do not show evidence of molecular EMT, tumour cells lining in the acinar or papillary structures, i.e. the sessile tumour cells, display evidence of partial EMT. The next step would be to increase the sample size and firmly establish if there is evidence of partial EMT in the sessile tumour cells surrounding the micropapillae in micropapillary tumour growth (Chapter 4). It would be also interesting to investigate the reason of focal EMT activation in sessile tumour cells. Is it due to micro-environmental cues or due to alterations in the immune response? Although micropapillary tumours show little evidence of molecular EMT and maintain epithelial molecular features, they are associated with the worst prognosis in lung adenocarcinoma patients. Investigations are ongoing in the laboratory to understand why this tumour growth pattern is so lethal in lung adenocarcinoma patients.

7.2   EMT is variable and related to local invasiveness

Chapter 4 further illustrates the features of EMT in primary tumours and show how EMT can be variable within tumour tissue. A recent study from our lab has identified subtypes of Noguchi type C tumours, i.e. mixed tumours with both *in situ* and invasive growths, based on the appearance of the *in situ* component of the tumour, namely C1 and C2 tumours (Moore *et al.,* 2019). C1 and C2 tumours are biologically and genomically distinct: in general there are additional genetic alterations in the invasive area of C1 tumour (especially *TP53* gene mutation) while there is no difference in

genomic profiling between *in situ* and invasive parts of C2 tumours. Data show that EMT is implicated in the transition to invasiveness in C1 tumours, which may itself be caused by *TP53* mutations, but there is no detectable evidence of translation dysregulation involved in this transition. In future studies, it would be of interest to expand this experiment to a larger cohort of C1/C2 cases to more firmly conclude if EMT is involved in the transition to invasiveness in C1 tumours, and explore whether genetic alteration on its own plays a role in activating EMT in cancer cells or there are others epigenetics/micro-environmental changes, in addition to genetic mutations, which can also promote EMT activation.

Multiple early foci of partial EMT (represented by N-cadherin expression) can be detected in one *in situ*/lepidic tumour, indicating early invasion of lung adenocarcinoma. Next phase would be to confirm this result by increasing the number of predominant lepidic tumour samples. It would be worthwhile exploring why and how invasive behaviours are arisen in low-grade tumour cells (Is it due to epigenetic changes or due to micro-environment cues?).

Pleormorphic carcinoma with adenocarcinoma component (PCAC) tumour is composed of adenocarcinoma or squamous cell carcinoma with a mixture of at least 10% of spindled or giant cells. It has been suggested that this type of tumour growth arises from adenocarcinoma elements that have dedifferentiated into mesenchymal phenotype through an EMT program (Pelosi *et al.*, 2010). Our data show that there is evidence of molecular and morphological EMT in the sarcomatoid region of PCAC tumours while the adenocarcinoma component does not exhibit any evidence of EMT. It would be of interest to investigate what triggers the dedifferentiation of adenocarcinoma tumour cells into mesenchymal phenotype: Is there any difference in the genetic profiling between tumour cells constituting the adenocarcinoma and the sarcomatoid components (i.e. presence of clonal heterogeneity)? Does the micro-environment influence carcinoma to activate EMT?

Of note, cancer invasion can also occur in a non-EMT manner by collective or amoeboid single cell migration. In both cases, collective or single tumour cells form protrusions to invade and migrate to another site. During collective migration, malignant carcinoma

maintain its epithelial features and degrade the extracellular matrix while cells undergoing amoeboid migration are deformable, show weak adhesion to the extracellular matrix and have reduced proteolytic activity (Krakhmal *et al.*, 2015).

Once EMT was characterised in primary lung adenocarcinoma, I focused on the main aim of the project: can I find evidence of correlation between translation dysregulation and EMT in primary lung adenocarcinoma?

7.3   EMT is associated with measure of translation dysregulation

Amongst the examined translation initiation factors, phospho-eIF4E and phospho-eIF2α are the translation factors displaying the strongest associations with EMT-related markers in primary lung adenocarcinoma. However, single cell data analysis from multiplex study could not led to a conclusive association between EMT and translation dysregulation, which may be due to technical issues such as inaccurate cell segmentation and phenotyping, or inability to see dynamic process. The next step would be to improve multiplex optimisation (improve cell segmentation, cell phenotyping, and staining quality). In parallel, it would be worth to confirm the correlation between EMT and translation dysregulation in *in vitro* studies, for instance by inducing EMT and check the expression level of translation factors or by blocking phospho-eIF2α activity and explore whether tumours cells undergo EMT upon EMT induction in lung adenocarcinoma cell lines.

In conclusion, partial EMT is detectable in primary lung adenocarcinoma. More complete molecular EMT is related to less differentiated growth pattern, i.e. solid growth pattern, and poor outcome. Furthermore, I find circumstantial evidence of association between molecular EMT and translation dysregulation. This suggests that targeting mRNA translation may be clinically valuable to stop EMT. Further investigations need to be carry out to understand the role of translation dysregulation in EMT in primary lung adenocarcinoma.

References

ACD (2014) *RNAscope® manual solutions biomarker analysis RNA in situ hybridization* 2014. Available online: https://acdbio.com/manual-assays-rnascope [Accessed.

ACD (2021) *RNAscope How It Works*, 2021. Available online: https://acdbio.com/science/how-it-works [Accessed.

Acloque, H., Adams, M. S., Fishwick, K., Bronner-Fraser, M. and Nieto, M. A. (2009). "Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease". *The Journal of clinical investigation*, 119(6), 1438-1449.

ACS (2020) *Lifetime chance of getting lung cancer*, 2020. Available online: https://www.cancer.org/cancer/lung-cancer/about/key-statistics.html [Accessed.

ACS (2021) *Lung Cancer Survival Rates*, 2021. Available online: https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/survival-rates.html [Accessed.

AJCC (2010) *Cancer staging manual Seventh Edition*, 2010. Available online: https://cancerstaging.org/references-tools/deskreferences/Documents/AJCC%207th%20Ed%20Cancer%20Staging%20Manual.pdf [Accessed.

Alberg, A. J., Ford, J. G. and Samet, J. M. (2007). "Epidemiology of Lung Cancer: ACCP Evidence-Based Clinical Practice Guidelines (2nd Edition)". *Chest*, 132(3, Supplement), 29S-55S.

Alberga, A., Boulay, J. L., Kempe, E., Dennefeld, C. and Haenlin, M. (1991). "The snail gene required for mesoderm formation in Drosophila is expressed dynamically in derivatives of all three germ layers". *Development*, 111(4), 983-92.

Allred, D. C., Harvey, J. M., Berardo, M. and Clark, G. M. (1998). "Prognostic and predictive factors in breast cancer by immunohistochemical analysis". *Mod Pathol*, 11(2), 155-68.

Ansari-Lari, M. A., Hoque, M. O., Califano, J. and Westra, W. H. (2002). "Immunohistochemical p53 Expression Patterns in Sarcomatoid Carcinomas of the Upper Respiratory Tract". *The American Journal of Surgical Pathology*, 26(8), 1024-1031.

Asamura, H., Chansky, K., Crowley, J., Goldstraw, P., Rusch, V. W., Vansteenkiste, J. F., Watanabe, H., Wu, Y.-L., Zielinski, M., Ball, D. and Rami-Porta, R. (2015). "The International Association for the Study of Lung Cancer Lung Cancer Staging Project: Proposals for the Revision of the N Descriptors in the Forthcoming 8th Edition of the TNM Classification for Lung Cancer". *Journal of Thoracic Oncology*, 10(12), 1675-1684.

Atlas, H. P. *The Human Protein Atlas* Available online: https://v15.proteinatlas.org/learn/dictionary/normal/lung/detail+1/magnification+1 [Accessed.

Attar-Schneider, O., Pasmanik-Chor, M., Tartakover-Matalon, S., Drucker, L. and Lishner, M. (2015). "eIF4E and eIF4GI have distinct and differential imprints on multiple myeloma's proteome and signaling". *Oncotarget*, 6(6), 4315-4329.

Attwell, D., Mishra, A., Hall, C. N., O'Farrell, F. M. and Dalkara, T. (2016). "What is a pericyte?". *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, 36(2), 451-455.

Baird, T. D. and Wek, R. C. (2012). "Eukaryotic initiation factor 2 phosphorylation and translational control in metabolism". *Adv Nutr*, 3(3), 307-21.

Bhat, M., Robichaud, N., Hulea, L., Sonenberg, N., Pelletier, J. and Topisirovic, I. (2015). "Targeting the translation machinery in cancer". *Nature Reviews Drug Discovery*, 14(4), 261-278.

Biotum (2015), 2015. Available online: https://biotium.com/technology/primary-secondary-antibody-conjugates/tyramide-signal-amplification/ [Accessed.

Böhm, M., Totzeck, B. and Wieland, I. (1994). "Differences of E-cadherin expression levels and patterns in human lung cancer". *Annals of Hematology*, 68(2), 81-83.

Borthwick, D. W., Shahbazian, M., Krantz, Q. T., Dorin, J. R. and Randell, S. H. (2001). "Evidence for stem-cell niches in the tracheal epithelium". *Am J Respir Cell Mol Biol*, 24(6), 662-70.

Camelo, A., Dunmore, R., Sleeman, M. A. and Clarke, D. L. (2014). "The epithelium in idiopathic pulmonary fibrosis: breaking the barrier". *Frontiers in pharmacology*, 4, 173-173.

Cao, Y., Zhu, L.-Z., Jiang, M.-J. and Yuan, Y. (2015). "Clinical impacts of a micropapillary pattern in lung adenocarcinoma: a review". *OncoTargets and therapy*, 9, 149-158.

Caplan, A. I. (2015). "Adult Mesenchymal Stem Cells: When, Where, and How". *Stem cells international*, 2015, 628767-628767.

Carter, B. W., Glisson, B. S., Truong, M. T. and Erasmus, J. J. (2014). "Small Cell Lung Carcinoma: Staging, Imaging, and Treatment Considerations". *RadioGraphics*, 34(6), 1707-1721.

Carter, J. H., Deddens, J. A., Spaulding, N. R. I. V., Lucas, D., Colligan, B. M., Lewis, T. G., Hawkins, E., Jones, J., Pemberton, J. O., Douglass, L. E. and Graff, J. R. (2016). "Phosphorylation of eIF4E

serine 209 is associated with tumour progression and reduced survival in malignant melanoma". *British journal of cancer*, 114(4), 444-453.

Chang, C.-J., Chao, C.-H., Xia, W., Yang, J.-Y., Xiong, Y., Li, C.-W., Yu, W.-H., Rehman, S. K., Hsu, J. L., Lee, H.-H., Liu, M., Chen, C.-T., Yu, D. and Hung, M.-C. (2011). "p53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs". *Nature cell biology*, 13(3), 317-323.

Chaudhry R, B. B. (2021) *Anatomy, Thorax, Lungs*, 2021. Available online: https://www.ncbi.nlm.nih.gov/books/NBK470197/ [Accessed.

Chen, H., Carrot-Zhang, J., Zhao, Y., Hu, H., Freeman, S. S., Yu, S., Ha, G., Taylor, A. M., Berger, A. C., Westlake, L., Zheng, Y., Zhang, J., Ramachandran, A., Zheng, Q., Pan, Y., Zheng, D., Zheng, S., Cheng, C., Kuang, M., Zhou, X., Zhang, Y., Li, H., Ye, T., Ma, Y., Gao, Z., Tao, X., Han, H., Shang, J., Yu, Y., Bao, D., Huang, Y., Li, X., Zhang, Y., Xiang, J., Sun, Y., Li, Y., Cherniack, A. D., Campbell, J. D., Shi, L. and Meyerson, M. (2019). "Genomic and immune profiling of pre-invasive lung adenocarcinoma". *Nature Communications*, 10(1), 5472.

Chen, Z.-H., Wu, Q., Lu, J., Chen, L.-z., Wu, X.-y., Wang, Y., Ju, H., Lu, Y.-x., Chen, Y., Wang, F. and Xu, R.-h. (2018). "Eukaryotic initiation factor 4A2 (EIF4A2) expression in colorectal cancer and prediction of prognosis". *Journal of Clinical Oncology*, 36(4_suppl), 664-664.

Chignalia, A. Z., Vogel, S. M., Reynolds, A. B., Mehta, D., Dull, R. O., Minshall, R. D., Malik, A. B. and Liu, Y. (2015). "p120-Catenin Expressed in Alveolar Type II Cells Is Essential for the Regulation of Lung Innate Immune Response". *The American Journal of Pathology*, 185(5), 1251-1263.

Coleman, L. J., Peter, M. B., Teall, T. J., Brannan, R. A., Hanby, A. M., Honarpisheh, H., Shaaban, A. M., Smith, L., Speirs, V., Verghese, E. T., McElwaine, J. N. and Hughes, T. A. (2009). "Combined analysis of eIF4E and 4E-binding protein expression predicts breast cancer survival and estimates eIF4E activity". *British journal of cancer*, 100(9), 1393-1399.

Comtesse, N., Keller, A., Diesinger, I., Bauer, C., Kayser, K., Huwer, H., Lenhof, H. P. and Meese, E. (2007). "Frequent overexpression of the genes FXR1, CLAPM1 and EIF4G located on amplicon 3q26-27 in squamous cell carcinoma of the lung". *Int J Cancer*, 120(12), 2538-44.

CRUK (2020) *TNM staging*, 2020. Available online: https://www.cancerresearchuk.org/about-cancer/lung-cancer/stages-types-grades/tnm-staging#:~:text=The%20TNM%20staging%20system%20is,with%20stages%201%20to%204. [Accessed.

Derry, M. C., Yanagiya, A., Martineau, Y. and Sonenberg, N. (2006). "Regulation of poly(A)-binding protein through PABP-interacting proteins". *Cold Spring Harb Symp Quant Biol*, 71, 537-43.

Dever, T. E. and Green, R. (2012). "The elongation, termination, and recycling phases of translation in eukaryotes". *Cold Spring Harbor perspectives in biology*, 4(7), a013706-a013706.

Doll, R. and Hill, A. B. (1950). "Smoking and carcinoma of the lung; preliminary report". *British medical journal*, 2(4682), 739-748.

Dominici, M., Le Blanc, K., Mueller, I., Slaper-Cortenbach, I., Marini, F., Krause, D., Deans, R., Keating, A., Prockop, D. and Horwitz, E. (2006). "Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement". *Cytotherapy*, 8(4), 315-7.

Dong, P., Karaayvaz, M., Jia, N., Kaneuchi, M., Hamada, J., Watari, H., Sudo, S., Ju, J. and Sakuragi, N. (2013). "Mutant p53 gain-of-function induces epithelial-mesenchymal transition through modulation of the miR-130b-ZEB1 axis". *Oncogene*, 32(27), 3286-3295.

Donzé, O., Jagus, R., Koromilas, A. E., Hershey, J. W. and Sonenberg, N. (1995). "Abrogation of translation initiation factor eIF-2 phosphorylation causes malignant transformation of NIH 3T3 cells". *The EMBO journal*, 14(15), 3828-3834.

Du, L., Li, J., Lei, L., He, H., Chen, E., Dong, J. and Yang, J. (2018). "High Vimentin Expression Predicts a Poor Prognosis and Progression in Colorectal Cancer: A Study with Meta-Analysis and TCGA Database". *Biomed Res Int*, 2018, 6387810.

Eberhardt, W. E. E., Mitchell, A., Crowley, J., Kondo, H., Kim, Y. T., Turrisi, A., Goldstraw, P. and Rami-Porta, R. (2015). "The IASLC Lung Cancer Staging Project: Proposals for the Revision of the M Descriptors in the Forthcoming Eighth Edition of the TNM Classification of Lung Cancer". *Journal of Thoracic Oncology*, 10(11), 1515-1522.

Eberle, J., Krasagakis, K. and Orfanos, C. E. (1997). "Translation initiation factor eIF-4A1 mRNA is consistently overexpressed in human melanoma cells in vitro". *International Journal of Cancer*, 71(3), 396-401.

Elmoneim, H. M. and Zaghloul, N. M. (2011). "Expression of E-cadherin, N-cadherin and snail and their correlation with clinicopathological variants: an immunohistochemical study of 132 invasive ductal breast carcinomas in Egypt". *Clinics (Sao Paulo)*, 66(10), 1765-71.

Elzagheid, A., Buhmeida, A., Laato, M., El-Faitori, O., Syrjänen, K., Collan, Y. and Pyrhönen, S. (2012). "Loss of E-cadherin expression predicts disease recurrence and shorter survival in colorectal carcinoma". *APMIS*, 120(7), 539-548.

Emoto, K., Eguchi, T., Tan, K. S., Takahashi, Y., Aly, R. G., Rekhtman, N., Travis, W. D. and Adusumilli, P. S. (2019). "Expansion of the Concept of Micropapillary Adenocarcinoma to Include

a Newly Recognized Filigree Pattern as Well as the Classical Pattern Based on 1468 Stage I Lung Adenocarcinomas". *J Thorac Oncol*, 14(11), 1948-1961.

Empey, D. W. (1978). "Diseases of the respiratory system. Introduction: structure and function of the lungs". *British medical journal*, 1(6113), 631-633.

Faget, L. and Hnasko, T. S. (2015) Tyramide Signal Amplification for Immunofluorescent Enhancement, in Hnasko, R. (ed), *ELISA: Methods and Protocols*. New York, NY: Springer New York, 161-172.

Falletta, P., Sanchez-Del-Campo, L., Chauhan, J., Effern, M., Kenyon, A., Kershaw, C. J., Siddaway, R., Lisle, R., Freter, R., Daniels, M. J., Lu, X., Tüting, T., Middleton, M., Buffa, F. M., Willis, A. E., Pavitt, G., Ronai, Z. e. A., Sauka-Spengler, T., Hölzel, M. and Goding, C. R. (2017). "Translation reprogramming is an evolutionarily conserved driver of phenotypic plasticity and therapeutic resistance in melanoma". *Genes & development*, 31(1), 18-33.

Fan, S., Ramalingam, S. S., Kauh, J., Xu, Z., Khuri, F. R. and Sun, S.-Y. (2009). "Phosphorylated eukaryotic translation initiation factor 4 (eIF4E) is elevated in human cancer tissues". *Cancer biology & therapy*, 8(15), 1463-1469.

Fidler, I. and Hart, I. (1982). "Biological diversity in metastatic neoplasms: origins and implications". *Science*, 217(4564), 998-1003.

Filosso, P. L., Sandri, A., Guerrera, F., Solidoro, P., Bora, G., Lyberis, P., Ruffini, E. and Oliaro, A. (2016). "Primary lung tumors invading the chest wall". *Journal of thoracic disease*, 8(Suppl 11), S855-S862.

Furic, L., Rong, L., Larsson, O., Koumakpayi, I. H., Yoshida, K., Brueschke, A., Petroulakis, E., Robichaud, N., Pollak, M., Gaboury, L. A., Pandolfi, P. P., Saad, F. and Sonenberg, N. (2010). "eIF4E phosphorylation promotes tumorigenesis and is associated with prostate cancer progression". *Proceedings of the National Academy of Sciences*, 107(32), 14134-14139.

Furrukh, M. (2013). "Tobacco Smoking and Lung Cancer: Perception-changing facts". *Sultan Qaboos University medical journal*, 13(3), 345-358.

Gao, B. and Roux, P. P. (2015). "Translational control by oncogenic signaling pathways". *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1849(7), 753-765.

Garcia, M. A., Nelson, W. J. and Chavez, N. (2018). "Cell-Cell Junctions Organize Structural and Signaling Networks". *Cold Spring Harbor perspectives in biology*, 10(4), a029181.

Govindan, R. (2014). "Attack of the clones". *Science*, 346(6206), 169-170.

Gregory, P. A., Bert, A. G., Paterson, E. L., Barry, S. C., Tsykin, A., Farshid, G., Vadas, M. A., Khew-Goodall, Y. and Goodall, G. J. (2008). "The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1". *Nature Cell Biology*, 10(5), 593-601.

Hanahan, D. and Weinberg, R. A. (2000). "The Hallmarks of Cancer". *Cell*, 100(1), 57-70.

Hanahan, D. and Weinberg, Robert A. (2011). "Hallmarks of Cancer: The Next Generation". *Cell*, 144(5), 646-674.

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

Hartsock, A. and Nelson, W. J. (2008). "Adherens and tight junctions: structure, function and connections to the actin cytoskeleton". *Biochimica et biophysica acta*, 1778(3), 660-669.

Hill, C., Jones, M. G., Davies, D. E. and Wang, Y. (2019). "Epithelial-mesenchymal transition contributes to pulmonary fibrosis via aberrant epithelial/fibroblastic cross-talk". *Journal of lung health and diseases*, 3(2), 31-35.

Horvilleur, E., Sbarrato, T., Hill, K., Spriggs, R. V., Screen, M., Goodrem, P. J., Sawicka, K., Chaplin, L. C., Touriol, C., Packham, G., Potter, K. N., Dirnhofer, S., Tzankov, A., Dyer, M. J. S., Bushell, M., MacFarlane, M. and Willis, A. E. (2014). "A role for eukaryotic initiation factor 4B overexpression in the pathogenesis of diffuse large B-cell lymphoma". *Leukemia*, 28(5), 1092-1102.

Hung, J.-J., Yeh, Y.-C., Jeng, W.-J., Wu, K.-J., Huang, B.-S., Wu, Y.-C., Chou, T.-Y. and Hsu, W.-H. (2014a). "Predictive Value of the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society Classification of Lung Adenocarcinoma in Tumor Recurrence and Patient Survival". *Journal of Clinical Oncology*, 32(22), 2357-2364.

Hung, J. J., Yeh, Y. C., Jeng, W. J., Wu, K. J., Huang, B. S., Wu, Y. C., Chou, T. Y. and Hsu, W. H. (2014b). "Predictive value of the international association for the study of lung cancer/American Thoracic Society/European Respiratory Society classification of lung adenocarcinoma in tumor recurrence and patient survival". *J Clin Oncol*, 32(22), 2357-64.

Imielinski, M., Berger, A. H., Hammerman, P. S., Hernandez, B., Pugh, T. J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A., Sougnez, C., Auclair, D., Lawrence, M. S., Stojanov, P., Cibulskis, K., Choi, K., de Waal, L., Sharifnia, T., Brooks, A., Greulich, H., Banerji, S., Zander, T., Seidel, D., Leenders, F., Ansén, S., Ludwig, C., Engel-Riedel, W., Stoelben, E., Wolf, J., Goparju, C., Thompson, K., Winckler, W., Kwiatkowski, D., Johnson, B. E., Jänne, P. A., Miller, V. A., Pao, W., Travis, W. D., Pass, H. I., Gabriel, S. B., Lander, E. S., Thomas, R. K., Garraway, L. A., Getz, G. and Meyerson, M. (2012). "Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing". *Cell*, 150(6), 1107-1120.

Itoh, M. and Bissell, M. J. (2003). "The organization of tight junctions in epithelia: implications for mammary gland biology and breast tumorigenesis". *Journal of mammary gland biology and neoplasia*, 8(4), 449-462.

Iwano, M., Plieth, D., Danoff, T. M., Xue, C., Okada, H. and Neilson, E. G. (2002). "Evidence that fibroblasts derive from epithelium during tissue fibrosis". *The Journal of clinical investigation*, 110(3), 341-350.

Izumchenko, E., Chang, X., Brait, M., Fertig, E., Kagohara, L. T., Bedi, A., Marchionni, L., Agrawal, N., Ravi, R., Jones, S., Hoque, M. O., Westra, W. H. and Sidransky, D. (2015). "Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA". *Nature communications*, 6, 8258-8258.

Jackson, R. J., Hellen, C. U. T. and Pestova, T. V. (2010). "The mechanism of eukaryotic translation initiation and principles of its regulation". *Nature reviews. Molecular cell biology*, 11(2), 113-127.

Jamal-Hanjani, M., Wilson, G. A., McGranahan, N., Birkbak, N. J., Watkins, T. B. K., Veeriah, S., Shafi, S., Johnson, D. H., Mitter, R., Rosenthal, R., Salm, M., Horswell, S., Escudero, M., Matthews, N., Rowan, A., Chambers, T., Moore, D. A., Turajlic, S., Xu, H., Lee, S. M., Forster, M. D., Ahmad, T., Hiley, C. T., Abbosh, C., Falzon, M., Borg, E., Marafioti, T., Lawrence, D., Hayward, M., Kolvekar, S., Panagiotopoulos, N., Janes, S. M., Thakrar, R., Ahmed, A., Blackhall, F., Summers, Y., Shah, R., Joseph, L., Quinn, A. M., Crosbie, P. A., Naidu, B., Middleton, G., Langman, G., Trotter, S., Nicolson, M., Remmen, H., Kerr, K., Chetty, M., Gomersall, L., Fennell, D. A., Nakas, A., Rathinam, S., Anand, G., Khan, S., Russell, P., Ezhil, V., Ismail, B., Irvin-Sellers, M., Prakash, V., Lester, J. F., Kornaszewska, M., Attanoos, R., Adams, H., Davies, H., Dentro, S., Taniere, P., O'Sullivan, B., Lowe, H. L., Hartley, J. A., Iles, N., Bell, H., Ngai, Y., Shaw, J. A., Herrero, J., Szallasi, Z., Schwarz, R. F., Stewart, A., Quezada, S. A., Le Quesne, J., Van Loo, P., Dive, C., Hackshaw, A. and Swanton, C. (2017). "Tracking the Evolution of Non-Small-Cell Lung Cancer". *N Engl J Med*, 376(22), 2109-2121.

Ji, P., Diederichs, S., Wang, W., Böing, S., Metzger, R., Schneider, P. M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., Thomas, M., Berdel, W. E., Serve, H. and Müller-Tidow, C. (2003). "MALAT-1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early-stage non-small cell lung cancer". *Oncogene*, 22(39), 8031-8041.

Jones, G. S. and Baldwin, D. R. (2018). "Recent advances in the management of lung cancer". *Clinical medicine (London, England)*, 18(Suppl 2), s41-s46.

Kadota, K., Yeh, Y.-C., Sima, C. S., Rusch, V. W., Moreira, A. L., Adusumilli, P. S. and Travis, W. D. (2014). "The cribriform pattern identifies a subset of acinar predominant tumors with poor prognosis in patients with stage I lung adenocarcinoma: a conceptual proposal to classify cribriform predominant tumors as a distinct histologic subtype". *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 27(5), 690-700.

Kalluri, R. and Weinberg, R. A. (2009). "The basics of epithelial-mesenchymal transition". *The Journal of Clinical Investigation*, 119(6), 1420-1428.

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

Kim, S. A., Inamura, K., Yamauchi, M., Nishihara, R., Mima, K., Sukawa, Y., Li, T., Yasunari, M., Morikawa, T., Fitzgerald, K. C., Fuchs, C. S., Wu, K., Chan, A. T., Zhang, X., Ogino, S. and Qian, Z. R. (2016). "Loss of CDH1 (E-cadherin) expression is associated with infiltrative tumour growth and lymph node metastasis". *British Journal of Cancer*, 114(2), 199-206.

Knudsen, L. and Ochs, M. (2018). "The micromechanics of lung alveoli: structure and function of surfactant and tissue components". *Histochemistry and cell biology*, 150(6), 661-676.

Kohno, T., Nakaoku, T., Tsuta, K., Tsuchihara, K., Matsumoto, S., Yoh, K. and Goto, K. (2015). "Beyond ALK-RET, ROS1 and other oncogene fusions in lung cancer". *Translational Lung Cancer Research*, 4(2), 156-164.

Königshoff, M., Kramer, M., Balsara, N., Wilhelm, J., Amarie, O. V., Jahn, A., Rose, F., Fink, L., Seeger, W., Schaefer, L., Günther, A. and Eickelberg, O. (2009). "WNT1-inducible signaling protein-1 mediates pulmonary fibrosis in mice and is upregulated in humans with idiopathic pulmonary fibrosis". *The Journal of clinical investigation*, 119(4), 772-787.

Koromilas, A. E., Lazaris-Karatzas, A. and Sonenberg, N. (1992). "mRNAs containing extensive secondary structure in their 5' non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E". *Embo j*, 11(11), 4153-8.

Koumenis, C., Naczki, C., Koritzinsky, M., Rastani, S., Diehl, A., Sonenberg, N., Koromilas, A. and Wouters, B. G. (2002). "Regulation of protein synthesis by hypoxia via activation of the endoplasmic reticulum kinase PERK and phosphorylation of the translation initiation factor eIF2alpha". *Mol Cell Biol*, 22(21), 7405-16.

Koval, M. (2002). "Sharing signals: connecting lung epithelial cells with gap junction channels". *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 283(5), L875-L893.

Kowalczyk, A. P. and Green, K. J. (2013). "Structure, function, and regulation of desmosomes". *Progress in molecular biology and translational science*, 116, 95-118.

Krakhmal, N. V., Zavyalova, M. V., Denisov, E. V., Vtorushin, S. V. and Perelmuter, V. M. (2015). "Cancer Invasion: Patterns and Mechanisms". *Acta naturae*, 7(2), 17-28.

Lamouille, S., Xu, J. and Derynck, R. (2014). "Molecular mechanisms of epithelial-mesenchymal transition". *Nat Rev Mol Cell Biol*, 15(3), 178-96.

Lenfert, E., Maenz, C., Heinlein, C., Jannasch, K., Schumacher, U., Pantel, K., Tolstonog, G. V., Deppert, W. and Wegwitz, F. (2015). "Mutant p53 promotes epithelial-mesenchymal plasticity and enhances metastasis in mammary carcinomas of WAP-T mice". *International Journal of Cancer*, 136(6), E521-E533.

Leptin, M. (1991). "twist and snail as positive and negative regulators during Drosophila mesoderm development". *Genes Dev*, 5(9), 1568-76.

Li, Z., Yin, S., Zhang, L., Liu, W. and Chen, B. (2017). "Prognostic value of reduced E-cadherin expression in breast cancer: a meta-analysis". *Oncotarget*, 8(10), 16445-16455.

Lin, Y. W. and Aplan, P. D. (2007). "Gene expression profiling of precursor T-cell lymphoblastic leukemia/lymphoma identifies oncogenic pathways that are potential therapeutic targets". *Leukemia*, 21(6), 1276-1284.

Liotta, L. A., Tryggvason, K., Garbisa, S., Hart, I., Foltz, C. M. and Shafie, S. (1980). "Metastatic potential correlates with enzymatic degradation of basement membrane collagen". *Nature*, 284(5751), 67-68.

Liu, C., Luo, J., Zhao, Y.-T., Wang, Z.-Y., Zhou, J., Huang, S., Huang, J.-N., Long, H.-X. and Zhu, B. (2018). "TWIST1 upregulates miR-214 to promote epithelial-to-mesenchymal transition and metastasis in lung adenocarcinoma". *Int J Mol Med*, 42(1), 461-470.

Liu, Y., Wang, Y., Zhang, Y., Miao, Y., Zhao, Y., Zhang, P.-X., Jiang, G.-Y., Zhang, J.-Y., Han, Y., Lin, X.-Y., Yang, L.-H., Li, Q.-C., Zhao, C. and Wang, E.-H. (2009). "Abnormal expression of p120-catenin, E-cadherin, and small GTPases is significantly associated with malignant phenotype of human lung cancer". *Lung Cancer*, 63(3), 375-382.

Lu, W. T., Wilczynska, A., Smith, E. and Bushell, M. (2014). "The diverse roles of the eIF4A family: you are the company you keep". *Biochem Soc Trans*, 42(1), 166-72.

Ma, L., Teruya-Feldstein, J. and Weinberg, R. A. (2007). "Tumour invasion and metastasis initiated by microRNA-10b in breast cancer". *Nature*, 449(7163), 682-688.

Makimoto, Y., Nabeshima, K., Iwasaki, H., Miyoshi, T., Enatsu, S., Shiraishi, T., Iwasaki, A., Shirakusa, T. and Kikuchi, M. (2005). "Micropapillary pattern: a distinct pathological marker to subclassify tumours with a significantly poor prognosis within small peripheral lung adenocarcinoma (</=20 mm) with mixed bronchioloalveolar and invasive subtypes (Noguchi's type C tumours)". *Histopathology*, 46(6), 677-84.

Marcotrigiano, J., Gingras, A. C., Sonenberg, N. and Burley, S. K. (1999). "Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G". *Mol Cell*, 3(6), 707-16.

Meng, Q.-B., Kang, W.-M., Yu, J.-C., Liu, Y.-Q., Ma, Z.-Q., Zhou, L., Cui, Q.-C. and Zhou, W.-X. (2015). "Overexpression of eukaryotic translation initiation factor 5A2 (EIF5A2) correlates with cell aggressiveness and poor survival in gastric cancer". *PloS one*, 10(3), e0119229-e0119229.

Meurs, E. F., Galabru, J., Barber, G. N., Katze, M. G. and Hovanessian, A. G. (1993). "Tumor suppressor function of the interferon-induced double-stranded RNA-activated protein kinase". *Proceedings of the National Academy of Sciences*, 90(1), 232-236.

Miyazaki, N., Abe, Y., Oida, Y., Suemizu, H., Nishi, M., Yamazaki, H., Iwasaki, M., Inoue, H., Ueyama, Y. and Nakamura, M. (2006). "Poor outcome of patients with pulmonary adenocarcinoma showing decreased E-cadherin combined with increased S100A4 expression". *Int J Oncol*, 28(6), 1369-1374.

Miyoshi, T., Satoh, Y., Okumura, S., Nakagawa, K., Shirakusa, T., Tsuchiya, E. and Ishikawa, Y. (2003). "Early-Stage Lung Adenocarcinomas With a Micropapillary Pattern, a Distinct Pathologic Marker for a Significantly Poor Prognosis". *The American Journal of Surgical Pathology*, 27(1), 101-109.

Modelska, A., Turro, E., Russell, R., Beaton, J., Sbarrato, T., Spriggs, K., Miller, J., Gräf, S., Provenzano, E., Blows, F., Pharoah, P., Caldas, C. and Le Quesne, J. (2015). "The malignant phenotype in breast cancer is driven by eIF4A1-mediated changes in the translational landscape". *Cell death & disease*, 6(1), e1603-e1603.

Moore, D. A., Sereno, M., Das, M., Baena Acevedo, J. D., Sinnadurai, S., Smith, C., McSweeney, A., Su, X., Officer, L., Jones, C., Dudek, K., Guttery, D., Taniere, P., Spriggs, R. V. and Le Quesne, J. (2019). "In situ growth in early lung adenocarcinoma may represent precursor growth or invasive clone outgrowth-a clinically relevant distinction". *Mod Pathol*, 32(8), 1095-1105.

Moreira, A. L., Ocampo, P. S. S., Xia, Y., Zhong, H., Russell, P. A., Minami, Y., Cooper, W. A., Yoshida, A., Bubendorf, L., Papotti, M., Pelosi, G., Lopez-Rios, F., Kunitoki, K., Ferrari-Light, D., Sholl, L. M., Beasley, M. B., Borczuk, A., Botling, J., Brambilla, E., Chen, G., Chou, T.-Y., Chung, J.-H., Dacic, S., Jain, D., Hirsch, F. R., Hwang, D., Lantuejoul, S., Lin, D., Longshore, J. W., Motoi, N., Noguchi, M., Poleri, C., Rekhtman, N., Tsao, M.-S., Thunnissen, E., Travis, W. D., Yatabe, Y., Roden, A. C., Daigneault, J. B., Wistuba, I. I., Kerr, K. M., Pass, H., Nicholson, A. G. and Mino-Kenudson, M. (2020). "A Grading system for invasive pulmonary adenocarcinoma: a proposal from the IASLC pathology committee". *Journal of Thoracic Oncology*.

Murphy, S. J., Wigle, D. A., Lima, J. F., Harris, F. R., Johnson, S. H., Halling, G., Asiedu, M. K., Seto, C. T., Terra, S., Kosari, F., Peikert, T., Yang, P., Aubry, M.-C. and Vasmatzis, G. (2014). "Genomic rearrangements define lineage relationships between adjacent lepidic and invasive components in lung adenocarcinoma". *Cancer research*, 74(11), 3157-3167.

Nabhan, A. N., Brownfield, D. G., Harbury, P. B., Krasnow, M. A. and Desai, T. J. (2018). "Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells". *Science*, 359(6380), 1118-1123.

Nakashima, H., Jiang, S.-X., Sato, Y., Hoshi, K., Matsumoto, T., Nagashio, R., Kobayashi, M., Matsuo, Y., Shiomi, K., Hayakawa, K., Saegusa, M. and Satoh, Y. (2015). "Prevalent and up-regulated vimentin expression in micropapillary components of lung adenocarcinomas and its adverse prognostic significance". *Pathology International*, 65(4), 183-192.

Nawijn, M. C., Hackett, T. L., Postma, D. S., van Oosterhout, A. J. M. and Heijink, I. H. (2011). "E-cadherin: gatekeeper of airway mucosa and allergic sensitization". *Trends in Immunology*, 32(6), 248-255.

Nawrocki, B., Polette, M., Van Hengel, J., Tournier, J. M., Van Roy, F. and Birembault, P. (1998). "Cytoplasmic redistribution of E-cadherin-catenin adhesion complex is associated with down-regulated tyrosine phosphorylation of E-cadherin in human bronchopulmonary carcinomas". *The American journal of pathology*, 153(5), 1521-1530.

NICE (2020) *Non-small-cell lung cancer: treatments with curative intent*, 2020. Available online: https://pathways.nice.org.uk/pathways/lung-cancer#path=view%3A/pathways/lung-cancer/non-small-cell-lung-cancer-treatments-with-curative-intent.xml&content=view-node%3Anodes-surgery [Accessed.

NICE (2021) *Lung cancer: diagnosis and management*, 2021. Available online: https://www.nice.org.uk/guidance/ng122/chapter/Recommendations#combination-treatment-for-non-small-cell-lung-cancer [Accessed.

Noguchi, M., Morikawa, A., Kawasaki, M., Matsuno, Y., Yamada, T., Hirohashi, S., Kondo, H. and Shimosato, Y. (1995). "Small adenocarcinoma of the lung. Histologic characteristics and prognosis". *Cancer*, 75(12), 2844-2852.

Officer, L. K., Andreou, K. E., Teodósio, A. V., He, Z. and Le Quesne, J. P. (2020). "Automated Co-in Situ Hybridization and Immunofluorescence Using Archival Tumor Tissue". *Methods Mol Biol*, 2148, 245-256.

Onder, T. T., Gupta, P. B., Mani, S. A., Yang, J., Lander, E. S. and Weinberg, R. A. (2008). "Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways". *Cancer Res*, 68(10), 3645-54.

Paine, R., Gaposchkin, D., Kelly, C. and Wilcoxen, S. E. (1995). "Regulation of cytokeratin expression in rat lung alveolar epithelial cells in vitro". *Am J Physiol*, 269(4 Pt 1), L536-44.

Park, S.-M., Gaur, A. B., Lengyel, E. and Peter, M. E. (2008). "The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2". *Genes & development*, 22(7), 894-907.

Pelletier, J., Graff, J., Ruggero, D. and Sonenberg, N. (2015). "Targeting the eIF4F translation initiation complex: a critical nexus for cancer development". *Cancer research*, 75(2), 250-263.

Pelosi, G., Sonzogni, A., De Pas, T., Galetta, D., Veronesi, G., Spaggiari, L., Manzotti, M., Fumagalli, C., Bresaola, E., Nappi, O., Viale, G. and Rosai, J. (2009). "Review Article: Pulmonary Sarcomatoid Carcinomas: A Practical Overview". *International Journal of Surgical Pathology*, 18(2), 103-120.

PerkinElmer (2007) *TSA Signal Amplification (TSA) Systems*, 2007. Available online: https://www.perkinelmer.com/lab-solutions/resources/docs/BRO_tsasignalamplificationsystems.pdf [Accessed.

Popper, H. H. (2016). "Progression and metastasis of lung cancer". *Cancer metastasis reviews*, 35(1), 75-91.

Ramon y Cajal, S., De Mattos-Arruda, L., Sonenberg, N., Cortes, J. and Peg, V. (2014). "The intra-tumor heterogeneity of cell signaling factors in breast cancer: p4E-BP1 and peIF4E are diffusely expressed and are real potential targets". *Clinical and Translational Oncology*, 16(11), 937-941.

Raza, F., Waldron, J. A. and Quesne, J. L. (2015). "Translational dysregulation in cancer: eIF4A isoforms and sequence determinants of eIF4A dependence". *Biochem Soc Trans*, 43(6), 1227-33.

Reck, M., Rodríguez-Abreu, D., Robinson, A. G., Hui, R., Csőszi, T., Fülöp, A., Gottfried, M., Peled, N., Tafreshi, A., Cuffe, S., O'Brien, M., Rao, S., Hotta, K., Leiby, M. A., Lubiniecki, G. M., Shentu, Y., Rangwala, R. and Brahmer, J. R. (2016). "Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer". *N Engl J Med*, 375(19), 1823-1833.

Ren, Y., Hou, J., Xu, A. and Pan, Y. (2015). "Diagnostic utility of PAX2 and PAX5 in distinguishing non-small cell lung cancer from small cell lung cancer". *International journal of clinical and experimental pathology*, 8(11), 14709-14716.

Roberts, P. J. and Der, C. J. (2007). "Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer". *Oncogene*, 26(22), 3291-310.

Robichaud, N., del Rincon, S. V., Huor, B., Alain, T., Petruccelli, L. A., Hearnden, J., Goncalves, C., Grotegut, S., Spruck, C. H., Furic, L., Larsson, O., Muller, W. J., Miller, W. H. and Sonenberg, N. (2015). "Phosphorylation of eIF4E promotes EMT and metastasis via translational control of SNAIL and MMP-3". *Oncogene*, 34(16), 2032-2042.

Rodriguez-Boulan, E. and Macara, I. G. (2014). "Organization and execution of the epithelial polarity programme". *Nature Reviews Molecular Cell Biology*, 15(4), 225-242.

Rogers, G. W., Jr., Richter, N. J., Lima, W. F. and Merrick, W. C. (2001). "Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F". *J Biol Chem*, 276(33), 30914-22.

Royer, C. and Lu, X. (2011). "Epithelial cell polarity: a major gatekeeper against cancer?". *Cell Death & Differentiation*, 18(9), 1470-1477.

Russell, P. A., Wainer, Z., Wright, G. M., Daniels, M., Conron, M. and Williams, R. A. (2011). "Does Lung Adenocarcinoma Subtype Predict Patient Survival?: A Clinicopathologic Study Based on the New International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Lung Adenocarcinoma Classification". *Journal of Thoracic Oncology*, 6(9), 1496-1504.

Saitoh, M., Niijima, M., Takiguchi, Y., Hiroshima, K., Fujita, Y., Nishio, K. and Tatsumi, K. (2011). "An early event of EGFR mutation in pleomorphic carcinoma of the lung". *International Journal of Clinical Oncology*, 16(6), 770-773.

Schuller, A. P., Wu, C. C.-C., Dever, T. E., Buskirk, A. R. and Green, R. (2017). "eIF5A Functions Globally in Translation Elongation and Termination". *Molecular cell*, 66(2), 194-205.e5.

Seyfried, T. N. and Huysentruyt, L. C. (2013). "On the origin of cancer metastasis". *Critical reviews in oncogenesis*, 18(1-2), 43-73.

Shahbazian, D., Parsyan, A., Petroulakis, E., Hershey, J. and Sonenberg, N. (2010). "eIF4B controls survival and proliferation and is regulated by proto-oncogenic signaling pathways". *Cell cycle (Georgetown, Tex.)*, 9(20), 4106-4109.

Shaoyan, X., Juanjuan, Y., Yalan, T., Ping, H., Jianzhong, L. and Qinian, W. (2013). "Downregulation of EIF4A2 in non-small-cell lung cancer associates with poor prognosis". *Clin Lung Cancer*, 14(6), 658-65.

Shih, J.-Y., Tsai, M.-F., Chang, T.-H., Chang, Y.-L., Yuan, A., Yu, C.-J., Lin, S.-B., Liou, G.-Y., Lee, M.-L., Chen, J. J. W., Hong, T.-M., Yang, S.-C., Su, J.-L., Lee, Y.-C. and Yang, P.-C. (2005). "Transcription Repressor Slug Promotes Carcinoma Invasion and Predicts Outcome of Patients with Lung Adenocarcinoma". *Clinical Cancer Research*, 11(22), 8070-8078.

Silvera, D., Arju, R., Darvishian, F., Levine, P. H., Zolfaghari, L., Goldberg, J., Hochman, T., Formenti, S. C. and Schneider, R. J. (2009). "Essential role for eIF4GI overexpression in the pathogenesis of inflammatory breast cancer". *Nature Cell Biology*, 11(7), 903-908.

Silvera, D., Formenti, S. C. and Schneider, R. J. (2010). "Translational control in cancer". *Nat Rev Cancer*, 10(4), 254-66.

Sonenberg, N. and Hinnebusch, A. G. (2009). "Regulation of translation initiation in eukaryotes: mechanisms and biological targets". *Cell*, 136(4), 731-745.

Stanczak, A., Stec, R., Bodnar, L., Olszewski, W., Cichowicz, M., Kozlowski, W., Szczylik, C., Pietrucha, T., Wieczorek, M. and Lamparska-Przybysz, M. (2011). "Prognostic significance of Wnt-1, β-catenin and E-cadherin expression in advanced colorectal carcinoma". *Pathology oncology research : POR*, 17(4), 955-963.

Stone, R. C., Pastar, I., Ojeh, N., Chen, V., Liu, S., Garzon, K. I. and Tomic-Canic, M. (2016). "Epithelial-mesenchymal transition in tissue repair and fibrosis". *Cell and tissue research*, 365(3), 495-506.

Sulzer, M. A., Leers, M. P., van Noord, J. A., Bollen, E. C. and Theunissen, P. H. (1998). "Reduced E-cadherin expression is associated with increased lymph node metastasis and unfavorable prognosis in non-small cell lung cancer". *Am J Respir Crit Care Med*, 157(4 Pt 1), 1319-23.

Tam, A., Wadsworth, S., Dorscheid, D., Man, S. F. and Sin, D. D. (2011). "The airway epithelium: more than just a structural barrier". *Ther Adv Respir Dis*, 5(4), 255-73.

Tarin, D. (2005). "The Fallacy of Epithelial Mesenchymal Transition in Neoplasia". *Cancer Research*, 65(14), 5996-6001.

Toiyama, Y., Yasuda, H., Saigusa, S., Tanaka, K., Inoue, Y., Goel, A. and Kusunoki, M. (2013). "Increased expression of Slug and Vimentin as novel predictive biomarkers for lymph node metastasis and poor prognosis in colorectal cancer". *Carcinogenesis*, 34(11), 2548-2557.

Travis, W. D., Brambilla, E., Nicholson, A. G., Yatabe, Y., Austin, J. H. M., Beasley, M. B., Chirieac, L. R., Dacic, S., Duhig, E., Flieder, D. B., Geisinger, K., Hirsch, F. R., Ishikawa, Y., Kerr, K. M., Noguchi, M., Pelosi, G., Powell, C. A., Tsao, M. S. and Wistuba, I. (2015). "The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification". *Journal of Thoracic Oncology*, 10(9), 1243-1260.

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

oncology : official publication of the International Association for the Study of Lung Cancer, 6(2), 244-285.

Travis, W. D., Rekhtman, N., Riley, G. J., Geisinger, K. R., Asamura, H., Brambilla, E., Garg, K., Hirsch, F. R., Noguchi, M., Powell, C. A., Rusch, V. W., Scagliotti, G. and Yatabe, Y. (2010). "Pathologic Diagnosis of Advanced Lung Cancer Based on Small Biopsies and Cytology: A Paradigm Shift". *Journal of Thoracic Oncology*, 5(4), 411-414.

Travis, W. D. B., Elisabeth; Burke Allen P.; Marx Alexander; Nicholson Andrew G. (2014). *WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart*, 4th Edition edition. World Health Organisation.

Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S. and Yang, J. (2012). "Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis". *Cancer cell*, 22(6), 725-736.

Ullah, I., Subbarao, R. B. and Rho, G. J. (2015). "Human mesenchymal stem cells - current trends and future prospective". *Bioscience reports*, 35(2), e00191.

Umbas, R., Isaacs, W. B., Bringuier, P. P., Schaafsma, H. E., Karthaus, H. F. M., Oosterhof, G. O. N., Debruyne, F. M. J. and Schalken, J. A. (1994). "Decreased E-Cadherin Expression Is Associated with Poor Prognosis in Patients with Prostate Cancer". *Cancer Research*, 54(14), 3929-3933.

Vandewalle, C., Van Roy, F. and Berx, G. (2009). "The role of the ZEB family of transcription factors in development and disease". *Cellular and Molecular Life Sciences*, 66(5), 773-787.

Vineis, P. and Caporaso, N. (1995). "Tobacco and cancer: epidemiology and the laboratory". *Environmental health perspectives*, 103(2), 156-160.

Vora, H. H., Patel, N. A., Rajvik, K. N., Mehta, S. V., Brahmbhatt, B. V., Shah, M. J., Shukla, S. N. and Shah, P. M. (2009). "Cytokeratin and Vimentin Expression in Breast Cancer". *The International Journal of Biological Markers*, 24(1), 38-46.

Walters, S., Benitez-Majano, S., Muller, P., Coleman, M. P., Allemani, C., Butler, J., Peake, M., Guren, M. G., Glimelius, B., Bergström, S., Påhlman, L. and Rachet, B. (2015). "Is England closing the international gap in cancer survival?". *British journal of cancer*, 113(5), 848-860.

Wang, F., Flanagan, J., Su, N., Wang, L.-C., Bui, S., Nielson, A., Wu, X., Vo, H.-T., Ma, X.-J. and Luo, Y. (2012). "RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues". *The Journal of molecular diagnostics : JMD*, 14(1), 22-29.

Wang, R., Geng, J., Wang, J.-h., Chu, X.-y., Geng, H.-c. and Chen, L.-b. (2009a). "Overexpression of eukaryotic initiation factor 4E (eIF4E) and its clinical significance in lung adenocarcinoma". *Lung Cancer*, 66(2), 237-244.

Wang, S.-P., Wang, W.-L., Chang, Y.-L., Wu, C.-T., Chao, Y.-C., Kao, S.-H., Yuan, A., Lin, C.-W., Yang, S.-C., Chan, W.-K., Li, K.-C., Hong, T.-M. and Yang, P.-C. (2009b). "p53 controls cancer cell invasion by inducing the MDM2-mediated degradation of Slug". *Nature Cell Biology*, 11(6), 694-704.

Warth, A., Muley, T., Meister, M., Stenzinger, A., Thomas, M., Schirmacher, P., Schnabel, P. A., Budczies, J., Hoffmann, H. and Weichert, W. (2012). "The novel histologic International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification system of lung adenocarcinoma is a stage-independent predictor of survival". *J Clin Oncol*, 30(13), 1438-46.

Wei, J. H., Cao, J. Z., Zhang, D., Liao, B., Zhong, W. M., Lu, J., Zhao, H. W., Zhang, J. X., Tong, Z. T., Fan, S., Liang, C. Z., Liao, Y. B., Pang, J., Wu, R. H., Fang, Y., Chen, Z. H., Li, B., Xie, D., Chen, W. and Luo, J. H. (2014). "EIF5A2 predicts outcome in localised invasive bladder cancer and promotes bladder cancer cell aggressiveness in vitro and in vivo". *British journal of cancer*, 110(7), 1767-1777.

Wendel, H.-G., Stanchina, E. d., Fridman, J. S., Malina, A., Ray, S., Kogan, S., Cordon-Cardo, C., Pelletier, J. and Lowe, S. W. (2004). "Survival signalling by Akt and eIF4E in oncogenesis and cancer therapy". *Nature*, 428(6980), 332-337.

Williams-Hill, D. M., Duncan, R. F., Nielsen, P. J. and Tahara, S. M. (1997). "Differential Expression of the Murine Eukaryotic Translation Initiation Factor Isogenes eIF4AIand eIF4AIIs Dependent upon Cellular Growth Status". *Archives of Biochemistry and Biophysics*, 338(1), 111-120.

Wilshire, C. L., Louie, B. E., Horton, M. P., Castiglioni, M., Aye, R. W., Farivar, A. S., West, H. L., Gorden, J. A. and Vallières, E. (2016). "Comparison of outcomes for patients with lepidic pulmonary adenocarcinoma defined by 2 staging systems: A North American experience". *The Journal of Thoracic and Cardiovascular Surgery*, 151(6), 1561-1568.

Wistuba, I. I., Brambilla, E. and Noguchi, M. (2018) 17 - Classic Anatomic Pathology and Lung Cancer, in Pass, H. I., Ball, D. & Scagliotti, G. V. (eds), *IASLC Thoracic Oncology (Second Edition)*. Philadelphia: Elsevier, 143-163.e4.

Wood, S. L., Pernemalm, M., Crosbie, P. A. and Whetton, A. D. (2015). "Molecular histology of lung cancer: From targets to treatments". *Cancer Treatment Reviews*, 41(4), 361-375.

Xu, G., Yu, H., Shi, X., Sun, L., Zhou, Q., Zheng, D., Shi, H., Li, N., Zhang, X. and Shao, G. (2014). "Cisplatin sensitivity is enhanced in non-small cell lung cancer cells by regulating epithelial-

mesenchymal transition through inhibition of eukaryotic translation initiation factor 5A2". *BMC Pulmonary Medicine*, 14(1), 174.

Yan, X., Yan, L., Liu, S., Shan, Z., Tian, Y. and Jin, Z. (2015). "N-cadherin, a novel prognostic biomarker, drives malignant progression of colorectal cancer". *Mol Med Rep*, 12(2), 2999-3006.

Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A. and Weinberg, R. A. (2004). "Twist, a Master Regulator of Morphogenesis, Plays an Essential Role in Tumor Metastasis". *Cell*, 117(7), 927-939.

Yang, L., Wang, S., Zhou, Y., Lai, S., Xiao, G., Gazdar, A. and Xie, Y. (2017). "Evaluation of the 7(th) and 8(th) editions of the AJCC/UICC TNM staging systems for lung cancer in a large North American cohort". *Oncotarget*, 8(40), 66784-66795.

Yatabe, Y., Takahashi, T. and Mitsudomi, T. (2008). "Epidermal growth factor receptor gene amplification is acquired in association with tumor progression of EGFR-mutated lung cancer". *Cancer Res*, 68(7), 2106-11.

Ye, J., Kumanova, M., Hart, L. S., Sloane, K., Zhang, H., De Panis, D. N., Bobrovnikova-Marjon, E., Diehl, J. A., Ron, D. and Koumenis, C. (2010). "The GCN2-ATF4 pathway is critical for tumour cell survival and proliferation in response to nutrient deprivation". *The EMBO Journal*, 29(12), 2082-2096.

Ye, X. and Weinberg, R. A. (2015). "Epithelial-Mesenchymal Plasticity: A Central Regulator of Cancer Progression". *Trends Cell Biol*, 25(11), 675-686.

Yilmaz, M. and Christofori, G. (2009). "EMT, the cytoskeleton, and cancer cell invasion". *Cancer Metastasis Rev*, 28(1-2), 15-33.

Yoo, S. B., Chung, J. H., Lee, H. J., Lee, C. T., Jheon, S. and Sung, S. W. (2010). "Epidermal growth factor receptor mutation and p53 overexpression during the multistage progression of small adenocarcinoma of the lung". *J Thorac Oncol*, 5(7), 964-9.

Yoshizawa, A., Motoi, N., Riely, G. J., Sima, C. S., Gerald, W. L., Kris, M. G., Park, B. J., Rusch, V. W. and Travis, W. D. (2011). "Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases". *Modern Pathology*, 24(5), 653-664.

Young, S. K. and Wek, R. C. (2016). "Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response". *The Journal of biological chemistry*, 291(33), 16927-16935.

Yun, J. A., Kim, S. H., Hong, H. K., Yun, S. H., Kim, H. C., Chun, H. K., Cho, Y. B. and Lee, W. Y. (2014). "Loss of E-Cadherin expression is associated with a poor prognosis in stage III colorectal cancer". *Oncology*, 86(5-6), 318-28.

Zappa, C. and Mousa, S. A. (2016). "Non-small cell lung cancer: current treatment and future advances". *Translational lung cancer research*, 5(3), 288-300.

Zeisberg, M. and Neilson, E. G. (2009). "Biomarkers for epithelial-mesenchymal transitions". *The Journal of clinical investigation*, 119(6), 1429-1437.

Zhang, H., Liu, J., Yue, D., Gao, L., Wang, D., Zhang, H. and Wang, C. (2013). "Clinical significance of E-cadherin, β-catenin, vimentin and S100A4 expression in completely resected squamous cell lung carcinoma". *Journal of Clinical Pathology*, 66(11), 937-945.

Zhang, J., Liang, Z., Gao, J., Luo, Y. and Liu, T. (2011). "Pulmonary adenocarcinoma with a micropapillary pattern: a clinicopathological, immunophenotypic and molecular analysis". *Histopathology*, 59(6), 1204-1214.

Zhang, Y., Li, J., Wang, R., Li, Y., Pan, Y., Cai, D., Hu, H., Li, H., Ye, T., Luo, X., Zhang, Y., Li, B., Shen, L., Sun, Y. and Chen, H. (2014). "The prognostic and predictive value of solid subtype in invasive lung adenocarcinoma". *Scientific Reports*, 4(1), 7163.

Zhao, Y.-L., Zhu, R.-T. and Sun, Y.-L. (2016). "Epithelial-mesenchymal transition in liver fibrosis". *Biomedical reports*, 4(3), 269-274.

Chapter 8.   Appendix

Chapter 8 Appendix

Chapter 2 –Material and methods
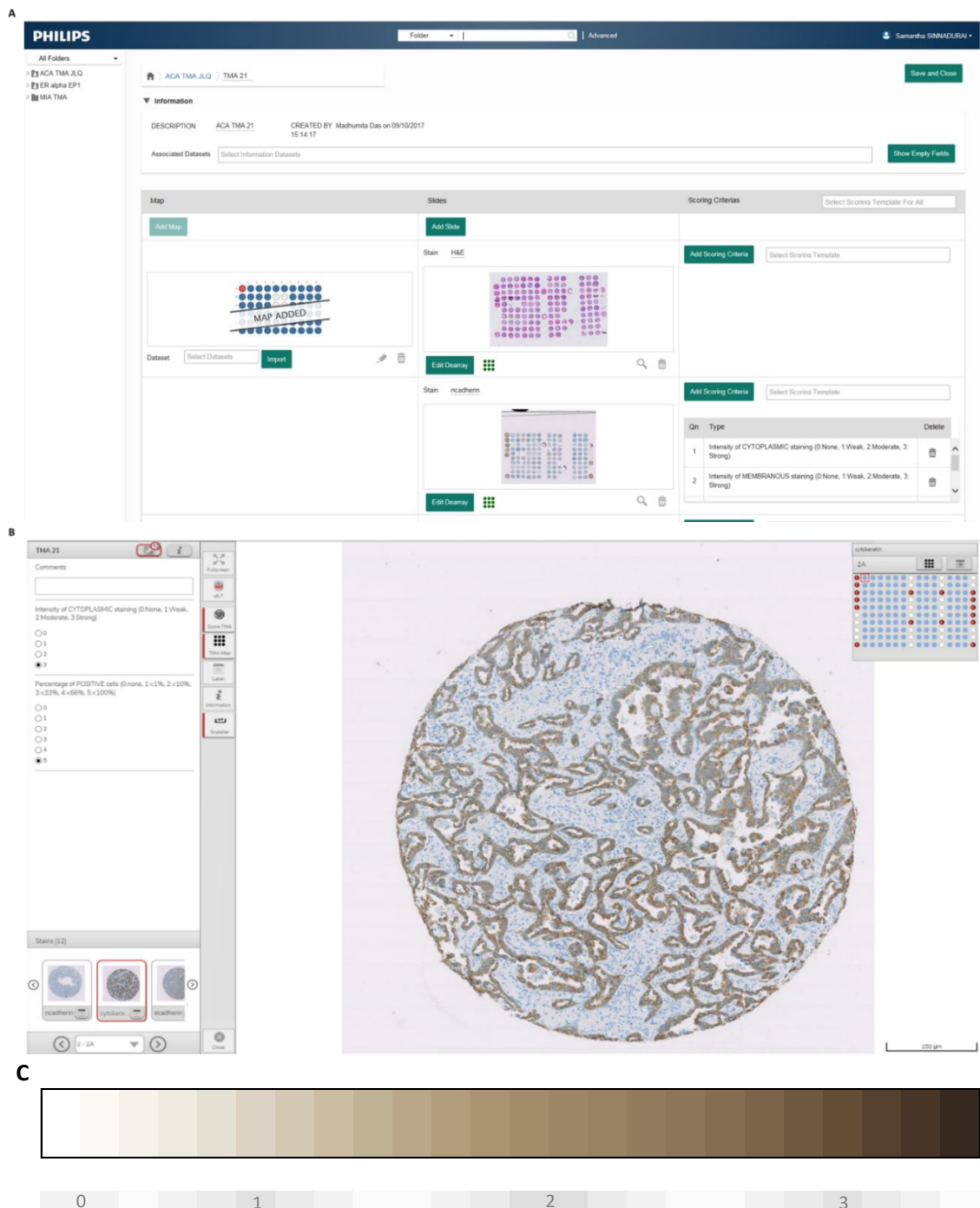
2.3.4. Image analysis

Figure 8.1. Path XL platform.

A: Slide uploading and scoring criteria choosing platform. B: Protein scoring platform. C: Scoring criteria of DAB stained protein. Brown staining is the DAB staining and blue staining represents the nucleus.

2.4.3. Image analysis

Figure 8.2. *TWIST* mRNA detection algorithm in Visiopharm®.

2.5.4. Image analysis

A
Nucleus segmentation parameters

B
Cell segmentation parameters

C
Membrane segmentation parameters

D
Phenotyping parameters

Figure 8.3. Cell segmentation and phenotyoing parameters in inForm®.