THE UNIVERSITY
of EDINBURGH

Prognosis of resected, early-stage, lung adenocarcinoma patients

Kathy Walsh

PhD – The University of Edinburgh - 2017

The University of Edinburgh

Thesis Declaration

See the Postgraduate Assessment Regulations for Research Degrees available via: www.ed.ac.uk/schools-departments/academic-services/policies-regulations/regulations/assessment

Name of Candidate:	Kathryn Walsh	*UUN*	S1066029
University email:	S1066029@exseed.ed.ac.uk		

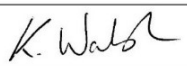
Degree Sought:	PhD	*No. of words in the main text of Thesis:*	36,660
Title of Thesis:	Prognosis of resected, early-stage, adenocarcinoma patients		

I certify:

(a) that the thesis has been composed by me, and

(b) either that the work is my own, or, where I have been a member of a research group, that I have made a substantial contribution to the work, such contribution being clearly indicated, and

(c) that the work has not been submitted for any other degree or professional qualification except as specified.

Signature:	

Abstract

Lung cancer is the leading cause of cancer related death worldwide; despite recent treatment developments survival rates remain poor and are closely related to the patient's clinical stage. Even among patients with early-stage lung cancer, which is amenable to surgical resection, prognosis is highly variable; some go on to live disease-free for many years whereas others quickly recur. Although post-operative chemotherapy is available it has associated morbidities and it is unclear which patients would benefit; therefore, there is a need for more effective stratification of patients. The adenocarcinoma sub-type of lung cancer is known to be morphologically heterogeneous however the majority of observed growth patterns, assessed by light microscopy, can be characterised into one of five formations: lepidic, papillary, acinar, solid and micropapillary. The morphology of each tumour has been proposed as a marker of prognosis and several studies have published a link between the most prevalent growth pattern and prognosis; suggesting those with predominantly solid or micropapillary tumours to have the least favourable outcomes. Indeed, it is now recommended that the proportion of each growth pattern and the predominant growth pattern should be reported for all resected lung adenocarcinomas; although no differential treatments have been recommended based on this assessment.

The aim of this study was to determine whether combining the analysis of clinicopathological; morphological; and candidate protein, molecular genetic and transcriptomic characteristics in a single cohort of 208 early-stage, resected, adenocarcinomas with clinical follow-up could be used to identify a subset of patients at high risk of recurrence. Comprehensive morphological analysis was carried out including the presence, proportion and number of individual growth patterns; the predominant growth pattern as well as features previously associated with tumour grade (the presence of large numbers of mitotic figures, apoptotic bodies, inflammatory cells, prominent nucleoli, pleomorphic tumour cells, dyscohesive tumour cells and large amounts of necrosis and scar tissue within the tumour). In addition, gene expression was assessed using a panel of 31 cell-cycle related genes, *EGFR* and *KRAS* mutation status was determined, and EGFR and TTF1 protein expression investigated.

In this study the predominant growth pattern defined by histopathology showed no ability to identify a group of patients with a poorer prognosis either in univariable or multivariable analysis. Univariable analysis identified nodal status [hazard ratio of N1 compared to N0 was 2.16 (95% CI 1.48 to 3.16, p< 0.0005)], clinical stage [hazard ratios of stage IIa and IIb compared to stage Ia were 3.15 (95% CI 1.73 to 5.73, p< 0.0005) and 2.22 (95% CI 1.10 to 4.48, p= 0.025) respectively], the presence of a significant amount of the papillary growth pattern [the hazard ratio of those with less than 8.5% papillary pattern was 0.657 (95% CI 0.44 to 0.98, p= 0.035)], and overall tumour grade score (including an assessment of necrosis, mitosis, apoptosis, nucleoli, scar tissue and inflammatory cells) [hazard ratio 1.71 (95% CI 1.14 to 2.56, p= 0.008)] as significantly associated with prognosis. Multivariable analysis using Cox's proportional hazards model identified clinical stage (p< 0.0005), the presence of a significant amount of the papillary growth pattern (p= 0.048) and the presence of large numbers of mitotic figures (p=0.029) and apoptotic bodies (p= 0.015) as independently associated with disease specific survival; although after correction for type I errors only clinical stage remained significantly associated with prognosis with patients with stage Ia disease having a significantly better outcomes [hazard ratio 0.418 (95% CI 0.20 to 0.86)]. Classification and regression tree analysis (CART) was used to further explore the data and to develop decision trees for the prognostication of early-stage lung adenocarcinoma patients. Receiver operating characteristic analysis based on 5-year survival showed a minimal improvement in the area under the curve between a model utilizing currently available clinicopathologic characteristics only [nodal status and lesion size, (area under the curve 0.704, 95% CI 0.631 to 0.777)] and one including growth pattern characteristics [area under the curve 0.725, 95% CI 0.654 to 0.796]. The greatest improvement in prognostic accuracy was observed when gene expression analysis was included in the analysis [area under the curve 0.749, 95% CI 0.673 to 0.825]; however even this showed very little impact compared to routinely used clinicopathologic variables.

This analysis suggests that the recommended characterisation of lung adenocarcinoma histology is not a robust predictor of patient outcomes; even a broader model which also included indicators of tumour grade and molecular characteristics was unable to identify a model sufficiently robust to implement into clinical practice and thereby

potentially alter patient treatment. Currently routinely collected clinical characteristics; including nodal status, size and clinical stage; continue to provide the most robust method of prognostication and detailed and time-consuming morphological analysis offers no significant benefit to the patient.

Lay summary

Lung cancer is the leading cause of cancer related death worldwide; despite recent treatment developments survival rates remain poor with patients diagnosed with advanced disease having the least favourable outcomes. Even among patients with early-stage disease, which may be removed by surgery, the prognosis is highly variable; some go on to live disease-free for many years whereas others quickly recur. Although post-operative chemotherapy is available its known side-effects mean there is a need to more effectively select patients who are likely to benefit. One major sub-type of lung cancer (adenocarcinoma) is known have a wide range of growth patterns when assessed by pathologists. Recent studies have suggested that the observed growth patterns can be used to indicate which patients may have the poorest prognoses. Indeed, it is now recommended that the proportion of each growth pattern and the predominant growth pattern should be reported for all resected lung adenocarcinomas; although no different treatment options have been recommended based on this assessment.

The aim of this study was to determine whether combining the analysis of routinely collected clinical and laboratory data (such as tumour size, spread and disease stage) with more detailed analysis of tumour morphology and other candidate biomarkers in a single cohort of 208 early-stage, resected, adenocarcinomas with clinical follow-up could be used to identify a subset of patients at high risk of recurrence.

In this study the predominant growth pattern showed no ability to identify a group of patients with a poorer prognosis. After taking all factors into account only the patient's disease stage, a currently used clinical measure, was found to be significantly associated with prognosis; those with very early stage disease (stage Ia) had a significantly better prognosis.

This study suggests that the recommended characterisation of growth patterns within this group of lung cancers is not a robust predictor of patient outcomes; even an analysis including all available biomarkers was unable to identify a method of selecting patients sufficiently robust to implement into clinical practice and thereby potentially alter a patient's treatment. Currently routinely collected clinical

characteristics continue to provide the most robust method of prognostication and these additional, time-consuming analyses offer no significant benefit to patients.

Acknowledgements

I'd like to thank my supervisors, Professor David Harrison and Professor William Wallace, for their advice and support throughout the duration of this project. I'd also like to thank my husband, Richard, for his continued support and patience.

Contents

Table of Figures

Abbreviations

AKT1	AKT serine/threonine kinase 1
ALK	Anaplastic lymphoma kinase
AR	Amphiregulin
ASF1B	Anti-silencing function 1B histone chaperone
ASPM	Abnormal spindle microtubule assembly
ATP	Adenosine triphosphate
ATRX	ATRX chromatin remodeler
ATS	American Thoracic Society
BIRC5	Baculoviral IAP repeat containing 5
BRAF	B-Raf proto-oncogene, serine/threonine kinase
BTC	Betacellulin
BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B
CCP	Cell-cycle progression
CDC2	Cyclin dependent kinase 1
CDC20	Cell division cycle 20
CDCA3	Cell division cycle associated 3
CDCA8	Cell division cycle associated 8
CDKN2A	Cyclin dependent kinase inhibitor 2A
CDKN3	Cyclin dependent kinase inhibitor 3
CDNK2A	Cyclin dependent kinase inhibitor 2A
CENPF	Centromere protein F
CENPM	Centromere protein M
CEP55	Centrosomal protein 55
DFS	Disease free survival
DSS	Disease specific survival (overall survival censored for those not dying as a result of lung cancer)
DLGAP5	DLG associated protein 5
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EML4	Echinoderm microtubule associated protein like 4
EPGN	Epigen
ERBB2	Erb-b2 receptor tyrosine kinase 2
ERBB3	Erb-b2 receptor tyrosine kinase 3
ERBB4	Erb-b2 receptor tyrosine kinase 4
EREG	Epiregulin
ERS	European Respiratory Society
FGFR4	Fibroblast growth factor receptor 4
FOXM1	Forkhead box M1
GDP	Guanine nucleotide diphosphate
GRB2	Growth factor receptor bound protein 2
GTP	Guanine nucleotide triphosphate
HB-EGF	Heparin binding EGF-like-growth factor
HR	Hazards ratio

IASLC	International Association for the Study of Lung Cancer
IGF-R1	Insulin-like growth factor 1 receptor
IHC	Immunohistochemistry
INSR	Insulin receptor
KIF11	Kinesin family member 11
KIF20A	Kinesin family member 20A
KRAS	KRAS proto-oncogene GTPase
MALDI	Matrix-assisted laser desorption ionization
MAP2K1	Mitogen-activated protein kinase 1
MCM10	Minichromosome maintenance 10 replication initiation factor
MET	MET proto-oncogene, receptor tyrosine kinase
mPS	Molecular prognostic score
NRG1	Neuregulin 1
NSCLC	Non-small cell lung carcinoma
NTRK1	Neurotrophic receptor tyrosine kinase 1
NUSAP1	Nucleolar and spindle associated protein 1
ORC6L	Origin recognition complex subunit 6
OR	Odds ratio
OS	Overall survival
p40	DeltaNp63
p63	Transformation-related protein p63
PBK	PDZ binding kinase
PCLAF	PCNA clamp associated factor
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PFS	Progression free survival
PLK1	Polo like kinase 1
PRC1	Protein regulator of cytokinesis 1
PTEN	Phosphatase and tensin homolog
PTTG1	Pituitary tumor-transforming 1
RAD51	RAD51 recombinase
RB1	RB transcriptional corepressor 1
RET	Ret proto-oncogene
ROS1	ROS proto-oncogene 1
RRM2	Ribonucleotide reductase regulatory subunit M2
SHC1	SHC adaptor protein 1
SKA1	Spindle and kinetochore associated complex subunit 1
SOS1	SOS Ras/Rac guanine nucleotide exchange factor 1
STK11	Serine/threonine kinase 11
TGFα	Transforming growth factor α
TIL	Tumour infiltrating lymphocytes
TK1	Thymidine kinase 1
TKI	Tyrosine kinase inhibitor
TNFSF13B	Tumor necrosis factor superfamily member 13b
TOP2A	Topoisomerase (DNA) II alpha
TP53	Tumour protein 53

TTP	Time to progression
X^2	Chi squared
X^2T	Chi squared test for trend

Chapter 1 Introduction

Lung cancer is the most commonly diagnosed cancer in the world and has the highest mortality rate; although the incidence in females is lower than that in males.[1] Improvements in cancer patient survival observed in recent years, for example in colon and breast cancer, have not been mirrored in lung cancer; the most recent estimates of 5-year survival rates in the UK are approximately 9.5 to 11.4% in males and 12% to 15.5% in females.[2, 3] This is due, at least in part, to the frequently advanced stage of disease at diagnosis at which point surgery would be unlikely to be curative.[4] The TNM classification of tumours; based on lesion size and location and involved structures, the location of involved lymph nodes and the presence and location of metastases; has been routinely used for many years for prognostication and clinical decision-making in lung cancer patients.[5, 6] Patients with clinical stage Ia disease, indicating those with non-metastatic lesions less than 3cm in diameter with no involved lymph nodes, have been shown to have the most favourable prognoses with an estimated 5-year survival of 50%, however this reduces markedly for patients with more advanced disease (43%, 36%, 25%, 19%, 7% and 2% for clinical stages Ib, IIa, IIb, IIIa, IIIb, and IV respectively).[7] Surgical resection of lung cancer is relatively rare, only being carried out in approximately 10% of cases, and dependent on the patient's performance status; the location of the lesion and the patient's clinical stage at diagnosis.[6, 8] In the UK, patients with stage I and II disease would, assuming no other contraindications, be offered surgical resection resulting in a 5-year survival rate of approximately 58% and 28% for stage I and II respectively.[6, 9] Those with stage II disease would also be eligible for treatment with post-operative platinum-based systemic anticancer therapy.[6] However, even within this group of early stage patients, a proportion quickly develop recurrent disease. If additional markers to identify patients with a poor prognosis were developed more aggressive treatment in this group, for example chemotherapy, may help improve survival rates. For patients where surgery with curative intent is not possible treatment options include radiotherapy, cisplatin with pemetrexed (for those with non-squamous histology), cisplatin/carboplatin with docetaxel, gemcitabine, paclitaxel or vinorelbine (squamous

histology) and targeted therapy for patients predicted likely to respond by molecular analysis.[6]

1.1 Morphology

Tumours thought to be derived from epithelial cells form the majority of lung malignancies which are grouped based on their histopathological and clinical characteristics into non-small cell lung carcinomas (NSCLC) including adenocarcinoma, squamous cell carcinoma, large cell carcinoma, and a number of less common tumours (for example adenosquamous carcinoma, pleomorphic carcinoma, spindle cell carcinoma, giant cell carcinoma and carcinosarcoma) and small cell lung carcinoma.[10] NSCLC accounts for approximately 80% of lung cancer cases.[11] The two most common types of NSCLC, adenocarcinoma and squamous cell carcinoma, are both associated with smoking tobacco and other environmental factors including exposure to asbestos, indoor pollution, a pre-existing chronic lung disease and a family history of lung cancer.[10] Until recently lung cancer was far more common in males than females, which may reflect differing smoking habits.[10] Initially squamous cell carcinoma was the most commonly diagnosed type of lung cancer; however, in recent years the number of squamous cell carcinomas (which commonly occurs in centralised locations) has been overtaken by peripherally occurring adenocarcinomas.[10, 11] This shift may have been caused by developments in the design of cigarettes including the widespread introduction of filters and the composition of the tobacco used, leading smokers to inhale more deeply into the peripheral regions of the lung.[10] Adenocarcinomas typically express mucin or pneumocyte markers, including thyroid transcription factor 1 (TTF1, now known as NK2 homeobox 1) and napsin A, which are frequently used to aid histopathological diagnosis by special stains and immunohistochemistry (IHC).[10] Approximately 75% of adenocarcinomas show positive expression of TTF1 by IHC.[10] Patients with lung adenocarcinoma frequently present at a late stage (greater than 40% of cases are diagnosed with stage IV disease);[12] which may be a result of their often peripheral location which can be asymptomatic in earlier stages.[10] Adenocarcinomas invade the pleura and chest wall in approximately 15% of cases and up to 20% of patients present with distant metastases; most commonly in the brain, bone, liver and adrenal glands.[13] Survival of

late stage lung adenocarcinoma (stage IIIB/IV) is approximately 12 months on average.[12]

It has long been recognised that adenocarcinomas are morphologically heterogeneous showing a wide range of growth patterns, not only between lesions from different individuals but also within tumours. Few lesions are composed of only one tumour architecture; the majority, 80% to 95%, are composed of more than one morphological pattern transforming from one pattern to the next in a continuum rather than with a defined border;[10, 14] which until recently had been classified in a single group as 'mixed subtype'.[11, 15] The most commonly observed morphologies in adenocarcinomas have been classified into 5 growth patterns; lepidic, papillary, acinar, solid and micropapillary; although less common variants have been noted including invasive mucinous adenocarcinoma, colloid adenocarcinoma, fetal adenocarcinoma and enteric adenocarcinoma.[10, 16] The lepidic growth pattern consists of relatively bland neoplastic cells, typically columnar or cuboidal with an appearance similar to club cells or type II pneumocytes, which grow along the surface of alveolar walls without invading into the stroma.[11, 16] The acinar growth pattern shows tumour cells (often mucin-producing) surrounding a central lumen forming round or oval acini or tubules.[11, 16] The papillary growth pattern consists of papillae with secondary and tertiary structures and a fibrovascular core surrounded by cuboidal to columnar tumour cells and may again produce mucin.[11, 16] The more recently identified micropapillary pattern is frequently composed of small and cuboidal tumour cells with little cytological atypia, growing in papillae without a fibrovascular core; free-floating in air spaces or as tufts attached to alveolar walls.[16] The solid growth pattern is composed of sheets of mucin producing polygonal tumour cells with no lepidic, acinar, papillary and micropapillary structures.[11, 16]

As yet no method of tumour grading in lung cancer has been established;[10] but in recent years, there has been a great deal of interest in additional markers of prognosis in lung cancer patients in order to identify those who may benefit from additional treatment and potentially improve the overall outcomes for patients. The morphological heterogenetiy of lung adenocarcinomas offers a rapid and, requiring only diagnostic stained tissue sections, inexpensive method for additional stratification. Many groups

have studied the disease characteristics and outcomes of patients with lug adenocarcinoma based on the observed morphological growth patterns and several have shown the presence of the micropapillary growth pattern to be associated with an increased likelihood of involved lymph nodes,[17-24] lymphatic invasion,[17, 19-21, 23, 24] venous invasion,[17, 19-24] pleural invasion,[18-20, 24] higher stage disease[17, 20-22, 24] and poorer prognosis.[17-21, 23-25] Although, an association with patient outcomes was not confirmed in all studies.[22, 26] In addition, the threshold applied to determine the presence of the micropapillary pattern was in some studies any micropapillary growth,[17, 19, 21, 24] in others greater than or equal to 5% of the total area[18, 20, 22, 23, 26] and in others greater than or equal to 10%.[25] Cross-cutting of other growth patterns could be mistaken for small areas of the micropapillary pattern and, depending on the cut-off used, could lead to false positives. Although frequently not the most prevalent pattern in a tumour, the micropapillary growth pattern has been found to be present in 30 to 50% of lung adenocarcinomas.[18, 19, 27] There is some evidence to suggest increasing proportions of the micropapillary pattern are associated with poorer outcomes for patients;[17, 19] although this has not been supported by all groups.[20] Importantly, the majority of studies which reported a clinical utility in the identification of the micropapillary pattern failed to use multivariable analysis and therefore possible confounding effects of multiple potentially significant variables cannot be assessed.[18-21, 24, 25] Of the few that did, most did not adjust the type I error level which would have rendered the effect of the presence of the micropapillary pattern not significant.[22, 23] Although correction for type I errors is not considered strictly necessary in exploratory studies, this fact does highlight weaknesses in the quality of data and potential flaws in the conclusions drawn from these studies. Some reports have also proposed the presence of the solid growth pattern as an indicator of poor prognosis;[26, 28-30] although not all studies made adjustments of the type I error level to account for multiplicity in statistical analyses.[28-30] The presence of the lepidic pattern has been shown to be associated with favourable prognoses, particularly when present at a large proportion of the total tumour area;[28, 31, 32] although, again, adjustments to the type I error level have not always been performed.[28, 31]

In 2011, after a review of lung adenocarcinoma classification, Travis *et al*. proposed that histopathologists apply a more detailed analysis of resected lung adenocarcinomas whereby the proportion of each growth pattern should be estimated in 5% increments and a predominant growth pattern assigned; although this was at the time classified as a weak recommendation with low quality evidence.[16] The association between predominant growth pattern and patient outcomes was supported by the research of several groups;[15, 33-46] and significant correlations were noted between predominant pattern and nodal status,[33] lesion size,[33] pleural invasion[38] and clinical stage.[33] Although in some studies these growth patterns were grouped together in order to show a significant relationship.[39, 44, 46, 47] The majority of studies suggested that tumours with predominantly lepidic growth showed a good prognosis;[15, 29, 33, 34, 38-40, 48] whereas predominantly micropapillary[15, 33, 35, 36, 38-40, 43] and solid tumours[29, 35, 36, 38-40, 43, 49] were associated with poor patient outcomes. In addition, it has been reported that the solid and micropapillary growth patterns were present in metastases even when a different predominant pattern was observed in the primary tumour which may indicate an increased tendency for cells in these patterns to metastasize.[50] However; even in studies with large cohorts, where the association between predominant growth pattern and outcomes would be expected to be strongest, there were discrepancies between the significance of the acinar, solid, papillary and micropapillary patterns. Yoshizawa *et al*. found patients with predominantly papillary and acinar tumours to have an intermediate prognosis with solid and micropapillary showing poor prognosis;[15, 39] Warth *et al*. showed predominantly papillary tumours to indicate a poor disease specific survival similar to solid and micropapillary tumours;[33] whereas Tsuta *et al*. , with a cohort of 904 patients, were only able to identify a prognostic difference in patients in predominantly lepidic tumours with all other patterns grouping closely together.[34] In addition, some groups failed to find any significant associations between predominant growth pattern and patient outcomes.[26, 51-54] These discrepancies may be a result of the reliance on resected tumours for complete morphological assessment, meaning that studies have primarily been carried out in patients with early stage disease with, therefore, a low number of events for survival analysis. In addition, as with other variables, several studies did not adjust the type I error level, the most commonly used of which, Bonferroni's procedure, would render their correlations not

significant.[15, 29, 33, 36, 38, 40, 43, 46, 48] The clinical utility of this classification is unclear as no differential treatments based on this stratification are currently available; although there is some indication that patients with predominantly solid or micropapillary tumours may benefit from adjuvant chemotherapy or radiotherapy whereas other tumours would not.[33, 52, 55] As yet this system of classification based on predominant pattern has not been in place long enough to judge the prognostic capacity of predominant growth pattern in a real-world situation, incorporating the morphological assessment of multiple histopathologists across many different centres. In 2015, this system of classification based on morphology was further refined to provide guidance on micropapillary structures present within acinar or lepidic structures and highlighting the poor prognosis of the cribiform pattern, which was previously classified as acinar.[14] Most recently it has been proposed that the papillary subtype may be further stratified into 3 distinct morphologies based on their association with overall survival (OS) and disease-free survival (DFS); however with ever increasing stratifications comes increased complexity and a reduced number of patients in each group, making this unlikely to be of benefit for clinical prognostication without considerably more research in large patient cohorts.[56] One study has shown that using a model incorporating both predominant and secondary growth patterns can further refine the stratification of patients;[57] if supported, this could indicate that predominant patterns are an important factor but consideration of additional factors may be necessary to more accurately predict the pathway of a patient's disease.

In its most recent edition, the WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart suggested future grading systems may be more complex, involving tumour architecture and nuclear features.[10] In other malignancies tumour grade, including cytological features, has been shown to aid prognostication; for example in breast cancer the commonly applied Nottingham Prognostic Index includes an assessment of nuclear atypia and the frequency of mitotic figures.[58] In NSCLC a high number of mitotic figures have been associated with a poor prognosis[59] and greater likelihood of distant metastasis,[60] however these conclusions were based on data from relatively small cohorts and in one case included a high proportion of squamous cell carcinoma.[60] High numbers of mitotic figures and apoptotic bodies have frequently

been identified in the same tumours[60] and although one group has suggested the presence of large numbers of apoptotic bodies was associated with poorer overall survival;[61] the analysis could only be carried out on a small number of patients and several alternative studies have found no significant association.[60, 62, 63] Several groups have shown the presence of extensive necrosis within a lung tumour to be associated with a poor prognosis,[15, 64, 65] although others have found no association.[63, 66] Univariable analysis has also proposed nuclear pleomorphism[63, 67] and the presence of scar tissue[67] as predictors of survival; however multivariable analysis, where performed, was unable to demonstrate and independent association.[67] A comprehensive analysis of all available histological parameters in a single cohort may help elucidate these associations.

There has been a great deal of interest in the effect of the body's immune response to malignancy and many studies have suggested that the presence of tumour infiltrating lymphocytes (TIL) can be used to predict prognosis in patients with NSCLC.[68-74] However, the available literature shows a complex relationship; the presence of large numbers of CD3,[68, 69] CD8[70-73] and CD4[69, 70, 72] expressing T-cells associated with a tumour may indicate a favourable prognosis; however, not all groups support these conclusions,[73, 75, 76] one study showed conflicting results indicating that high numbers of CD8 T-cells was associated with a poor prognosis.[74] There is evidence to suggest that the relationship between TIL and patient outcomes may be strongest in squamous cell carcinoma rather than adenocarcinoma.[72, 75, 77] In addition, the location of TIL, whether in the stroma surrounding the tumour or in the tumour nests themselves, may indicate differing outcomes.[69-71] Two meta-analyses carried out in lung cancer have shown large numbers of CD3,[78, 79] CD8[78, 79] and CD4[79] T-cells were associated with a favourable prognosis whereas large numbers of FOXP3 T-cells were associated with poorer outcomes.[78, 79]

1.2 Molecular pathology

In recent years; the belief that carcinogenesis is driven primarily by the accumulation of somatic mutations in key genes leading to features of cancer cells (including unlimited cell proliferation independent of normal stimulation and unchecked by

regulatory mechanisms, and the ability to metastasize and induce angiogenesis)[80, 81] has been widely accepted by many cancer researchers. It is certainly true that malignant cells contain many mutations not found in corresponding normal cells and common chromosome rearrangements and mutations have been identified across many different cancer types,[82] frequently affecting genes involved in tumour suppression, cell cycle control, DNA repair, and cell adhesion.[81] Frequent mechanisms of carcinogenesis include inactivation of tumour suppressor pathways (including mutation of tumour protein p53 encoded by the TP53 gene), loss of expression of the retinoblastoma protein, silencing of cyclin dependent kinase inhibitor 2A (CDNK2A), loss of heterozygosity and overexpression of cyclin D1 or the gain of proliferation-inducing driver mutations; including those in the B-Raf proto-oncogene serine/threonine kinase (*BRAF*), mitogen-activated protein kinase kinase 1 (*MAP2K1*), fibroblast growth factor receptor 4 (*FGFR4*), phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (*PIK3CA*), erb-b2 receptor tyrosine kinase 2 (*ERBB2 or HER2*), epidermal growth factor receptor (*EGFR*), KRAS proto-oncogene GTPase (*KRAS*) and anaplastic lymphoma kinase (*ALK*) genes.[11, 83] Between cancer types the range of genomic alterations differs; in lung adenocarcinoma the most commonly mutated genes include *TP53*, *KRAS*, *STK11*, *EGFR*, *BRAF*, *ERBB2*, *ALK*, *ROS1*, ret proto-oncogene (*RET*), neurotrophic receptor tyrosine kinase 1 (*NTRK1*) and neuregulin 1 (*NRG1*).[10, 84, 85] The most frequently occurring mutations, those in TP53 and KRAS are thought to be associated with smoking due to the large proportion of G to T transversion mutations.[10] In the majority of lung adenocarcinomas driver mutations are thought to be mutually exclusive although some may be concomitantly mutated in a small number of cases.[51, 85] Copy number changes are also common in lung adenocarcinoma most frequently gain of chromosome 5p in 18% of cases (leading to amplification of the telomerase reverse transcriptase (*TERT*) gene and gain of chromosome 3q which encodes the telomerase RNA component (TERC) gene (both of which are thought to be important in immortality of cancer cells). Other common copy number alterations include 14q13.3 encoding the *NKX2–1* gene encoding TTF1; amplification of *MYC, EGFR*, *MET*, *KRAS*, *ERBB2*, *MDM2*, and deletion of *LRP1B*, *PTPRD*, and *CDKN2A*.[86] These alterations most commonly affect proliferation

(RTK/RAS/RAF, mTOR, JAK-STAT), DNA repair and cell-cycle regulation pathways.[86]

The complexity of biological pathways is such that many different molecular mechanisms may lead to clinically identical disease; contributing, at least in part, to a variable response to treatment.[87] This complexity has lead researchers to hypothesise that no single treatment will be effective in all patients with a particular cancer type and that, for the most effective disease management, therapy must be targeted to the specific characteristics of each patient's tumour.[88] This has led to the development of targeted therapies for the treatment of lung adenocarcinoma which selectively inhibit the EGFR and ALK proteins, and it is hoped that further therapies targeting more driver mutations will become available in the coming years.

EGFR

The *EGFR* gene encodes a 170kDalton transmembrane receptor tyrosine kinase which is expressed on the plasma membrane of many cells;[89] IHC using antibodies with affinity to the EGFR protein shows particularly high levels of expression in the basal regions of stratified and squamous epithelium.[90] EGFR is one of a family of tyrosine kinase receptors which also includes ERBB2, erb-b2 receptor tyrosine kinase 3 (ERBB3) and erb-b2 receptor tyrosine kinase 4 (ERBB4).[91] The activation of the EGFR tyrosine kinase is thought to be mediated by the binding of ligands (including epidermal growth factor (EGF), transforming growth factor α (TGFα), amphiregulin (AR), heparin binding EGF-like-growth factor (HB-EGF), betacellulin (BTC) epiregulin (EREG), and epigen (EPGN))[92] to the extracellular domain of EGFR creating allosteric changes resulting in homo or heterodimerisation of EGFR with other members of the EGFR family of proteins.[91] This dimerisation causes autophosphorylation of amino acids (including Tyr 1173, 1148, 1086, 1068, and 992) leading to conformational changes to the intracellular tyrosine kinase domain which increases affinity to adenosine triphosphate (ATP) and allows the binding of substrates including SHC adaptor protein 1 (SHC1), growth factor receptor bound protein 2 (GRB2) and SOS Ras/Rac guanine nucleotide exchange factor 1 (SOS1).[89, 91] Activated EGFR catalyses the hydrolysis of bound ATP and the transfer of phosphates to substrate proteins triggering several downstream signalling transduction cascades

leading to cell proliferation; [91] uncontrolled activation is thought to result in unchecked proliferation, avoidance of apoptosis and ultimately cancer.[93] Ligand-bound EGFR is eventually deactivated by incorporation into lysosomes and degradation.[89, 90]

Two EGFR tyrosine kinase inhibitors (TKI); erlotinib (Tarceva®, Roche) and gefitinib (Iressa®, Astra Zeneca UK Ltd), which represented the first generation of their class; have shown good responses in a subset of patients including younger patients, non-smokers, women, patients of East Asian origin and those with adenocarcinoma histological sub-type.[94-98] Gefitinib is a low molecular weight, synthetic anilinoquinazoline which has been shown to reversibly inhibit EGFR by competing with ATP at its binding site and, therefore, prevent autophosphorylation.[99] Erlotinib hydrochloride is a quinazoline derivative reversible inhibitor of *EGFR*; both treatments are thought to have similar binding[100] and efficacies.[101]

Some early reports suggested that EGFR protein expression, determined by IHC with antibodies with affinity to the extracellular region of the protein, was a predictor of benefit from EGFR TKIs;[102] however many studies failed to find a significant predictive role for EGFR IHC.[103-105] The interpretation of IHC is known to be subject to inter-observer variation and in addition analysis of EGFR IHC in matched biopsies and resected samples has shown intratumour heterogeneity and/or variation in tissue fixation; both of which make the validation of a significant predictive relationship more difficult.[106, 107] However, even after efforts to standardise IHC protocols and scoring systems[108] the current consensus is that EGFR protein expression levels, measured by IHC, are unable to adequately predict response to first generation EGFR TKIs. A more recent publication has reported the development of a new antibody with affinity to the intracellular region of EGFR which showed a stronger association with progression-free survival (PFS) and OS in patients treated with gefitinib,[90] although this study was carried out in a relatively small cohort and the predictive capacity of this intracellular region would require further validation in additional cohorts.

Approximately 9 to 11% of NSCLC from western populations have shown increased *EGFR* gene copy numbers, in the majority of cases this is caused by disomy, trisomy and polysomy rather than gene amplification.[109] Many studies found increased *EGFR*

gene copy number to be a predictor of response to EGFR TKIs and has been associated with OS,[102, 105, 110-112] PFS,[104, 105, 113] and time to progression (TTP).[111, 114] The role of *EGFR* gene copy number as a predictive marker of benefit from EGFR TKIs was supported in a meta-analysis of 22 studies;[115] however other studies have found no predictive value.[116, 117]

In 2004 somatic mutations in the *EGFR* gene were reported in tumours responsive to gefitinib which were not found in non-sensitive tumours;[118-121] the sensitivity of clones bearing mutations was confirmed in cell culture experiments.[118-120, 122, 123] *EGFR* mutations were located in exons 18 to 21, the area coding for the active cleft of the tyrosine kinase domain, and were thought to destabilise the inactive conformation of the protein and favour the activated state leading to ligand-independent activation of the EGFR pathway.[100] The most frequent *EGFR* mutations were small deletions in exon 19 (61.9%) followed by mutations in exon 21 (33.1%) and exon 18 (4.2%).[124-126] The p.(E746_A750del) deletion in exon 19 has been shown to be the most common single mutation representing 35.6% of all *EGFR* mutations.[124] The frequency of *EGFR* mutations has been shown to increase from pre-invasive lesions to invasive adenocarcinoma, indicating that *EGFR* mutations may not be the primary cause for cell proliferation but may be a marker of invasiveness.[127] *EGFR* mutations were more common in tumours with adenocarcinoma histology,[124, 128-130] in non-smoking patients[124, 128-130] and in patients of East Asian ethnicity; [85] matching the clinicopathological characteristics of patients who respond to *EGFR* TKIs. The genetic profile of lung adenocarcinomas has also been shown to vary considerably in different ethnic groups, the incidence of *EGFR* mutations in lung adenocarcinomas in East Asian populations is high compared to Caucasian populations (approximately 48% vs 19%).[85] Even within Caucasian populations there may be differences in EGFR mutation frequency with a meta-analysis demonstrating higher mutation frequency in North American (21.6%) than European populations (13.0%) (supplementary data Dearden *et al.* 2013).[85] The reason for these differences is unclear but may be related to differing exposure to known carcinogens, for example smoking or environmental factors, or unknown genetic factors. Although previous studies have suggested *EGFR* mutations were associated with female gender,[124, 128] not all studies support this and

smoking habits may be a confounding factor.[129, 131-133] Lung adenocarcinomas which harbour mutations in the *EGFR* gene express the TTF1 protein in the majority of cases.[124, 134-136]

Many studies have supported the value of *EGFR* mutations for predicting benefit from EGFR TKIs;[104, 110, 111] with only a small number of studies finding no predictive value.[102, 112, 116] Although, final analysis of OS in large scale phase III trials, including the IPASS and SATURN trials, failed to show an association with *EGFR* mutations after treatment with gefitinib;[104, 137-140] a finding which has been further supported in meta-analyses.[141] The large scale treatment cross-over in these trials may explain these disappointing results. Patients with wild-type *EGFR* tumours have been shown to perform better on standard chemotherapy regimens than on TKIs leading to the recommendation for routine analysis of *EGFR* mutation status before prescription of EGFR TKI.[140, 142, 143]

The more recently developed treatment afatinib (Giotrif, Boehringer-Ingelheim Pharma GmbH, Ingelheim, Germany) is a second generation small-molecule irreversible inhibitor of the ErbB-family of proteins which blocks signalling from homodimers and heterodimers involving EGFR, HER2 (ErbB2), ErbB3, and ErbB4.[144] Treatment with afatinib has demonstrated a longer progression free survival than chemotherapy or placebo in patients with *EGFR* mutations[145, 146] and has shown improved PFS compared with gefitinib.[147]

Some *EGFR* mutations, including exon 20 insertions and the p.T790M single nucleotide variant, have been associated with resistance to TKIs.[148-150] The incidence of de novo p.T790M mutations in a primary lesion is low and frequently leads to rapid progression when treated with first generation EGFR TKIs.[151] Even within known sensitising mutations, different *EGFR* variants may not confer the same level of sensitivity to EGFR TKIs. It has been reported that lesions with deletions in exon 19 demonstrate better outcomes than those with exon 21 mutations;[152-155] although the same finding was not evident in all studies.[139, 156, 157] Tumours bearing p.L861Q and p.S768I have been reported to be less sensitive to erlotinib and gefitinib compared to

tumours with deletions in exon 19 or p.L858R.[152, 153, 158] Even the specific exon 19 deletion has been reported to associate with differing PFS.[159]

The association between the presence of *EGFR* mutations and treatment benefit is far from perfect; 17 to 38% of lung adenocarcinoma patients with *EGFR* mutations in their tumours do not respond to first line EGFR TKIs. Reasons for this may include presence of a resistance mutation, or activation of non-EGFR dependent pathways indicating that carcinogenesis in these patients may not be reliant on the EGFR pathway alone.[160] Some patients with *EGFR* mutated tumours have additional mutations in the PI3K/Akt/mTOR pathway which have been associated with primary resistance to EGFR TKI; including *TP53*, *FGFR*, *AKT* serine/threonine kinase 1 (*AKT1*), *PIK3CA*, serine/threonine kinase 11 (*STK11*) and *PTEN*.[161] Alterations to E-cadherin, Beta-catenin and PTEN and mitogen inducible gene 6 protein expression have also been suggested to be mechanisms of primary resistance to EGFR TKIs in *EGFR* mutated patients.[162, 163] Inherited polymorphisms may also have an influence on a patients response to treatment with EGFR TKIs.[164, 165]

A small proportion of NSCLC patients without *EGFR* mutations (1 to 7%) do respond to EGFR TKIs,[166] the reasons for this are unknown but may involve: amplification of the *EGFR* gene, mutations in the *EGFR* promoter region, expression of associated proteins such as the EGFR ligand amphiregulin or other unknown mechanisms.[167] Meaning that somatic *EGFR* mutation status alone may be insufficient to accurately predict response to treatment.

Although targeted therapies have shown notable responses in NSCLC, acquired resistance is almost inevitable. The mechanisms of resistance are only partially understood, differ from patient to patient, and may involve: further mutations in the target gene (development of an additional *EGFR* p.T790M mutation occurs in approximately half of patients with acquired resistance to EGFR TKI),[168, 169] upregulation or activation of alternative pathways (for example MET proto-oncogene receptor tyrosine kinase (MET), HER2 or PI3K-AKT),[163, 168, 169] transformation to a small-cell histology[168] and selecting for pre-existing cells which are inherently resistant to the drug.[170, 171] It has been hypothesised that a minor component of

neoplastic cells containing the p.T790M mutation is frequently present in tumours prior to EGFR TKI therapy and that their presence has been shown to be associated with poorer PFS and OS when treated with EGFR TKI.[172]

A third generation of EGFR TKI, osimertinib (Tagrisso™, AstraZeneca, UK), which has demonstrated activity against tumours with the EGFR p.T790M resistance mutation (whether in the primary lesion or acquired as a result of treatment) has recently been released to the market.[173] However, this has raised the issue of obtaining representative tumour tissue samples from recurrent late stage NSCLC patients in order to detect the p.T790M mutation. As many of these patients are too unwell to tolerate a re-biopsy the prescription of this novel treatment may rely upon the detection of *EGFR* mutations in circulating cell-free tumour DNA released into the bloodstream, although this method has a lower analytical sensitivity than tissue based analysis.[174]

Some studies have suggested a link between histological tumour growth patterns and the frequency of *EGFR* mutations. It has been reported that lepidic,[175] acinar[47, 49, 175, 176] and papillary[47, 175, 176] predominant tumours had an increased incidence of *EGFR* mutations. However, these conclusions were based on relatively small cohorts and, since *EGFR* mutations can also be found in solid and micropapillary predominant tumours, the use of tumour morphology as a predictor of *EGFR* mutations has limited clinical benefit.

ALK

Crizotinib (Xalkori, Pfizer) is a small molecule inhibitor of ALK, MET and ROS proto-oncogene 1 receptor tyrosine kinase (ROS1) which has shown response in NSCLC patients with *ALK* rearrangements (65% response rate), *ROS1* rearrangements (72%), *MET* amplification (60%) or *MET* exon 14 skipping mutations (75%); although the number of treated patients with *MET* aberrations was very small.[177-182] Clinically relevant rearrangements of the *ALK* gene, most commonly with echinoderm microtubule associated protein like 4 (*EML4*), create a fusion protein causing the aberrant expression of the ALK protein; these rearrangements have been detected in 1-5% of lung adenocarcinoma patients.[177, 183, 184] *ALK* rearrangements are thought to be largely mutually exclusive to *EGFR* and *KRAS* mutations,[183] however they have

been reported to co-occur in small numbers of cases.[185, 186] Rearrangements in the *ROS1* gene, with any of several translocation partners, cause the production and expression of the ROS1 fusion protein; these rearrangements have been detected in 1.7% of lung adenocarcinomas.[184] *MET* amplifications and *MET* exon 14 skipping mutations have been identified in 2.2% and 4.3% of lung adenocarcinomas respectively.[184] Unfortunately although patients who respond to treatment have shown remarkable reduction in tumour size, acquired resistance to crizotinib can develop quickly and has been attributed to further mutations in the *ALK* gene, *ALK* gene amplification and activation of alternative pathways.[187]

Ceritnib (Zykadia; Novartis) is a second generation small molecule inhibitor of ALK, ROS1, insulin-like growth factor 1 receptor (IGF-R1) and insulin receptor (INSR) but not MET; this treatment has shown activity against some tumours with acquired resistance to crizotinib allowing a second line of treatment for patients with *ALK* rearranged NSCLC.[187] Several other next generation ALK inhibitors are currently in development.[187]

KRAS

KRAS is a guanosine nucleoside diphosphate (GDP) and triphosphate (GTP) binding protein which acts as a signal transducing molecule in many pathways including the EGFR-ERK proliferation pathway.[188] In normal cells KRAS is activated by the binding of GTP which renders it able to bind and activate downstream targets (including BRAF) by phosphorylation.[188] KRAS cleaves bound GTP by hydrolysis to produce GDP and is once more rendered inactive.[188] However, mutations in hotspot regions (including codons 12, 13, 61, 117 and 146) of KRAS mean it is unable to hydrolyse GTP leaving the protein in a constitutively activated state.[188] Activating mutations in *KRAS* codons 12, 13 or 61 are common in pulmonary adenocarcinomas and have been correlated with exposure to cigarette smoke;[39, 129, 189, 190] most commonly resulting in transversion (G>T) mutations.[132, 191, 192] Unlike somatic alterations in the *EGFR* gene, *KRAS* mutations have been shown to be more common in western than East Asian populations; being present in approximately 26.1 to 35% of Caucasian cohorts.[85, 190] *KRAS* and *EGFR* mutations are considered mutually exclusive; [193, 194] they have been reported to co-occur, albeit rarely[85] and in these cases

it is not known if these two mutations are present in the same tumour cells or in different populations of cells within a single lesion. *KRAS* mutations have been shown to be a poor prognostic factor in NSCLC in meta-analyses [195, 196] and in single cohort studies;[104] although not all studies have confirmed this.[197] Solid predominant tumours have been reported to have less favourable responses to EGFR TKI;[198] which corresponds to an increased incidence of *KRAS* mutations, rather than *EGFR* mutations, in these tumours.[151]

Intratumour heterogeneity

All cell populations harbour some level of heterogeneity; therefore, it is reasonable to assume that tumour populations are no different. In her well cited 1984 review, Heppner describes tumours as multiple clones with different genotypes, and pointed to evidence of heterogeneity in treatment response, protein expression and metastatic potential;[199] these clones are likely to interact with each other, influencing the behaviour of the tumour as a whole.[200] Even cell lines, often considered a homogenous population, can be shown to contain sub-populations of cells which respond differently to perturbations.[201] Minor sub-populations may significantly affect the behaviour of a tumour and average measurements (for example gene expression or mutation status) taken across whole tissue sections are likely to miss small differences within the population and are unlikely to be truly representative of the whole tumour.[202]

Medical research has a bias towards genetic aberrations as the cause of cancer; however, this ignores non-heritable mechanisms as a potential source of variability. Variation in protein expression levels; caused by miRNA sequence, miRNA expression levels, promoter methylation or histone variants; could account for the survival of sub-populations of cells after treatment.[203-205] Differences in growth patterns are likely to be the result of distinct molecular characteristics and may respond differently to perturbations.[202]

IHC has been used to demonstrate intratumour heretogeneity of protein expression; including EGFR,[107] MMP9,[107] Ki-67,[206] P53[206] and BCL2[206]; although if a positive/negative scoring procedure is applied this heterogeneity may have little effect on the result.[207] IHC is known to be affected by poor tissue fixation, which may explain

some apparent heterogeneity in large tissue specimens; however the same cannot be said for small biopsies whose rapid fixation time make this an unlikely explanation.[208]

Somatic mutations have also been found to show intratumour heterogeneity with spatial variability in *EGFR* mutation status detected by laser capture microdissection.[209-213] In addition, the p.T790M mutation has been shown to exist as a minor clone in tumours with other *EGFR* mutations.[170, 214] Next generation sequencing has shown sub-populations of sequencing reads with a different deletion to that found in the majority of reads; the long term effect of these tumour sub-populations and their response to treatment with EGFR TKIs is as yet unknown.[215] However, intratumour heterogeneity of *EGFR* mutations is not supported by all researchers who believe that, as an early event in lung cancer,[216] *EGFR* mutations are likely to be present in all tumour cells in a lesion and discordances are more likely to be a result of false negatives.[134, 217-219]

Antibodies with affinity to EGFR protein bearing the common p.L858R or p.E746_A750del mutations have been developed for use in IHC.[220] The overall sensitivity of these IHCs is considered to be low; 66% to 86% for deletions in exon 19 and 76% to 83% for p.L858R although specificity was high (98% and 96% respectively).[221, 222] For deletions the low sensitivity could be explained by the presence of non-p.E746_A750del deletion mutations; however the reason for low sensitivity of p.L858R IHC is less clear. These IHC assays have been reported to demonstrate intratumoural heterogeneity in mutation status in 46% of known positive samples;[223] however as no control IHC was used varying protein expression levels, rather than mutation status, cannot be ruled out.

To treat intratumour heterogeneity as "experimental noise" rather than an indicator of the complex mechanisms underlying carcinogenesis is likely to result in over simplification and potentially less effective treatment.[201] Further research to explore discordance in *EGFR* mutation status was recommended by the 2010 European *EGFR* mutation consensus workshop.[143]

Biomarkers of prognosis

Determining prognosis in lung cancer patients may help to stratify patients into treatment regimens with the most aggressive therapy targeted at those with the poorest predicted outcomes. The goal of prognostication has been pursued for many years and an enormous number of journal articles have been published claiming prognostic utility of methods using molecular and immunohistochemical techniques. In 2016 alone, papers have been published suggesting expression levels of miRNAs,[224] long non-coding RNAs,[225] mRNAs,[226-228] and proteins[229-231] were effective markers of prognostication; either as single target tests [226, 231] or as panels.[224, 227, 228] However few have been followed up with validation studies in large clinical cohorts. Gene expression signatures have been particularly pursued, one major advantage being their objectivity. However, gene expression can differ greatly between subjects and has shown variation depending on tissue treatment post sample collection[232] and intratumour heterogeneity.[233] In addition; many prognostic gene expression signatures were developed for analysis of fresh frozen tissues, rather than routinely collected FFPE tissue, making their implementation into clinical practice unlikely.

The commercially available Oncotype DX assay (Genomic Health®, USA) examines a 21-gene expression signature based on analysis of RNA isolated from routinely collected FFPE breast cancer tissue. This assay results in a recurrence risk score for breast cancer patients and can be used to identify patients who would be likely to benefit from adjuvant chemotherapy.[234] In NSCLC, Veristrat® (Biodesix®, USA) is a commercially available assay using matrix-assisted laser desorption ionization (MALDI) mass spectrometry analysis of serum or plasma, which has shown value in prognostication.[235] However, analysis using this blood-based assay is not possible in retrospective studies of patients with only archived FFPE material available.

In 2011 the expression levels of a panel of genes, which fluctuated during the normal cell-cycle and therefore indicate tumour cell proliferation, was proposed as a useful prognostic marker in prostate cancer.[236] This panel of 31 cell cycle-related genes included: abnormal spindle microtubule assembly (*ASPM*), anti-silencing function 1B histone chaperone (*ASF1B*), baculoviral IAP repeat containing 5 (*BIRC5*), BUB1 mitotic checkpoint serine/threonine kinase B (*BUB1B*), cell division cycle associated

3 (*CDCA3*), cell division cycle associated 8 (*CDCA8*), cyclin dependent kinase 1 (*CDC2*), cell division cycle 20 (*CDC20*), cyclin dependent kinase inhibitor 3 (*CDKN3*), centromere protein F (*CENPF*), centromere protein M (*CENPM*), centrosomal protein 55 (*CEP55*), DLG associated protein 5 (*DLGAP5*), forkhead box M1 (*FOXM1*), kinesin family member 11 (*KIF11*), kinesin family member 20A (*KIF20A*), minichromosome maintenance 10 replication initiation factor (*MCM10*), nucleolar and spindle associated protein 1 (*NUSAP1*), PCNA clamp associated factor (*PCLAF* also known as *KIAA0101*), PDZ binding kinase (*PBK*), polo like kinase 1 (*PLK1*), protein regulator of cytokinesis 1 (*PRC1*), pituitary tumor-transforming 1 (*PTTG1*), RAD51 recombinase (*RAD51*), ATRX, chromatin remodeler (*ATRX* also known as *RAD54L*), ribonucleotide reductase regulatory subunit M2 (*RRM2*), spindle and kinetochore associated complex subunit 1 (*SKA1* also known as *C18orf24*), origin recognition complex subunit 6 (*ORC6L*), thymidine kinase 1 (*TK1*), tumour necrosis factor superfamily member 13b (*TNFSF13B* also known as *DTL*) and topoisomerase (DNA) II alpha (*TOP2A*).[236] The average expression levels of these genes, normalised against 15 housekeeping genes, was used to generate a cell-cycle progression (CCP) score.[236] A high CCP score showed a strong association with a high risk of prostate cancer recurrence.[236] Analysis of the same 31-gene panel in several published microarray datasets also showed reproducible prognostic utility in bladder; although in lung carcinoma only 5 of the 8 cohorts studied showed a significant correlation between CCP and survival.[237] Interestingly, the same study showed no association between CCP score and prognosis in either lung or head and neck squamous cell carcinoma, although these cohorts were relatively small.[23] In 2013; quantitative RT-PCR analysis was used to determine the CCP score in early stage (stage I and II) lung adenocarcinomas using RNA derived from FFPE tissue and was able to identify patients in two cohorts with a poor prognosis measured by cancer specific survival.[238] Indeed CCP score was considered to be a stronger indicator of prognosis than clinical stage.[238] More recently in a large cohort of stage I lung adenocarcinomas the CCP score and a molecular prognostic score, which incorporated CCP and clinical stage, were both associated with 5-year lung cancer specific mortality.[239] This panel has the advantage of being robust enough to produce results even on RNA isolated from FFPE tissue.

There have been several attempts to produce models predicting prognosis in lung cancer patients which incorporated variables from more than one scientific discipline. Analysis of NSCLC at a population level has shown that a model including clinical stage, tumour grade, patient age, ethnicity and gender was able to improve on clinical stage alone for prognostication.[12] However; these very large population based studies do not, as yet, include more specific details such as growth pattern, cytological features and molecular variables and therefore the effect of these features cannot be examined in the model. One study of adenocarcinoma patients included clinical parameters as well as morphological classification; *EGFR*, *KRAS* and *TP53* mutation status and TTF1 protein expression.[240] However; their morphological classification did not align with the, now recommended, International Association for the Study of Lung Cancer, American Thoracic Society, and European Respiratory Society (IASLC/ATS/ERS) multidisciplinary lung adenocarcinoma classification.[240] It has also been shown that a model including age, gender, staging, and tumour grade and clinical measurements (cardiac comorbidities, lung function and maximum standardised uptake value from positron emission tomography imaging) gave improved prediction of prognosis in NSCLC.[241] No laboratory-based measurements were included in this study which may have further improved the model. A further model based on a cohort of predominantly adenocarcinoma patients found tumour location, stage, histologic grade (including IASLC/ATS/ERS subtype classification), differentiation and lymphovascular invasion were independently associated with recurrence-free survival. However, the analysis was based on a relatively small number of events and included no correction for type 1 errors which would have been likely to alter the conclusions made.[45] A relatively small study (n=102) examined clinical factors and markers of cell proliferation, cell cycle control, apoptosis and angiogenesis and found gender, pathological stage, VEGF and nuclear cyclin D1 expression to be independent markers of post-operative recurrence.[242] Although, this analysis was dependent on only 25 recurrences, included bith adenocarcinoma and squamous cell carcinoma and no further morphological variables were included.[242] One group reported that; as well as T and N stage; lepidic and predominant subtypes, mitotic rate and overall grade (incorporating nuclear size and mitotic rate) were independently associated with prognosis of adenocarcinoma patients; although they made no attempt to produce a model for clinical use.[243]

Recursive partitioning analysis including multiple staging, clinical, molecular (IHC) and functional variables produced a decision tree which included: the presence of ipsilobar nodules, tumour size and invasion, comorbidities, age, performance status, smoking history, previous tumour, haemoglobin levels, and IHC markers (phospho-ACC, Ki67, p63, E-cadherin, phosphor-mTOR, p27 and NF-kB). [244] The resulting decision tree was complex but the most significant variables were the presence of ipsilobar nodules, followed by tumour size and performance status or direct invasion into the diaphragm.[244] Although this model showed an increased ability to predict prognosis the cohort used was biased towards squamous cell carcinoma (only 23% adenocarcinoma); in addition, the authors themselves suggest there is room for improvement and the addition of morphological subtypes and molecular markers may further refine this model.[244]

1.3 Project objectives

The hypothesis of this study is that the analysis of multi-factor data; including histology but also candidate molecular genetic, transcriptomic and proteomic biomarkers; can be used to improve prognostication in early stage resected non-small cell lung carcinoma patients. This study has been designed to address: the clinical significance adenocarcinoma sub-type classification in a Scottish early stage lung adenocarcinoma cohort; the relevance of the proportion of growth patterns detected in each tumour specimen; the clinical significance of the presence of the micropapillary growth pattern as an indicator of poor prognosis and intratumour heterogeneity in *EGFR* mutation status as a possible mechanism of EGFR TKI resistance.

Clinical audits provide invaluable information to explore factors which may add to our understanding of disease in a real-world situation, rather than the heavily selected cohorts used in many clinical trials. The majority of clinical audits studying *EGFR*, *KRAS* and *ALK* mutation frequency in NSCLC patients have been carried out in East Asian Populations and support the comparatively high incidence of *EGFR* mutations in these cohorts.[151, 245] Although studies in Caucasian populations have been carried out,[132, 246] little is known of incidence and clinicopathological associations in a Scottish population. One published abstract of a clinical audit has shown an *EGFR* mutation incidence of 7% in a North of England cohort (including 95% ex- or current-smokers);

this same audit reported an increased frequency in female patients but little information is available on the statistical analyses carried out.[246]

NHS Lothian has carried out molecular pathology testing for lung cancer patients since the recommendation of erlotinib for second-line treatment of locally advanced or metastatic NSCLC in Scotland in 2011, which was later expanded for use in a first line setting.[247] Initially only *EGFR* mutation analysis was considered clinically relevant however after the recommendation of crizotinib for the second line treatment of *ALK* rearranged advanced NSCLC[248] *KRAS* mutation analysis and *ALK* rearrangement analysis were added to the test repertoire. In a clinical setting; where the only available samples may be sub-optimal, containing only a small area of invasive disease or a high proportion of non-neoplastic cells, the detection of a *KRAS* mutation can give increased confidence in the validity of the sampling strategy. By carrying out a clinical audit of molecular pathology testing in our centre the incidence of clinically relevant mutations, associations with clinicopathological factors and the potential benefit to patients of molecular pathology testing can be explored.

As described above, there have been many reports claiming a clinical prognostic utility of the detailed morphological analysis to identify the growth patterns present and the predominant pattern in each tumour. However, many of these publications can be criticised on the basis of the statistical analyses carried out and that they do not include molecular variables. When analysing the data from a single variable the assumption of statistical significance if the p value is less than 0.05 has been widely accepted indicating a 5% probability of the finding occurring by chance (false positive or a type I error). However, when analysing multiple variables, for example 10, in the same sample increases the probability of finding at least one false positive to 50%. It is recommended that analysis of multiple variables should correct for type I errors.[249] One such method, Bonferroni's procedure, simply requires that variables in multivariable analysis should only be considered significant if the p value is less than or equal to 0.05 divided by the number of variables. However, many consider Bonferroni's correction to be overly conservative and increases the risk of type II error (false negative) especially where large numbers of variables have been considered, other less stringent statistical techniques have been developed, although for

implementation in a clinical setting a more conservative approach is appropriate.[250, 251] In addition in exploratory studies, where the number of variables is often large, it is not considered necessary to correct for type I error; although clinical trials carried out prior to the implementation of any intervention should include this correction.[250, 251]

Decision tree analysis may provide a solution to the analysis of many variables in a single dataset and can be of use in identifying potentially informative biomarkers in exploratory research. This method divides individuals in the cohort into groups at multiple decision points, or nodes, based on the status of that variable.[252] In the first node the whole dataset is divided into two branches using the variable with the greatest influence at predicting the outcome.[252] The data in each branch is then divided using the next most influential variable for that subgroup. Variables with more than one category may be grouped and potentially re-used in a later node.[252] Decision trees may be highly complex and by examining the usefulness of a variable within each specific subset of patients the need for adjusting the type I error rate is negated.[252] Multiple methods are now available for decision tree analysis including classification and regression tree (CART) analysis which has the advantage of being able to use continuous or categorical variables as input and outcome measures.[253] CART has been shown to have good performance at model production compared to logistic regression[254] and other machine learning methods.[255] In lung cancer decision tree analysis has been applied to aid accurate diagnosis to subtype,[256, 257] risk of metastasis[258] and clinical outcome.[259] In addition, as CART can be used on continuous variables, this technique can be used to determine threshold levels for binary analysis.[244, 260] However, decision tree analysis may be prone to overfitting, especially when the number of samples in each group is small [252, 253] and strong correlation between variables may make decisions tree analysis unsuitable.[253]

In addition to evaluation of morphological growth patterns as recommended by Travis *et al*. a more detailed examination of cytological features, observed by light microscopy of H&E stained sections, in the same cohort allows for a more comprehensive investigation including features of tumour grade. Objective analysis using molecular pathology techniques in the same cohort of patient samples allows for a direct comparison of morphology with tumour molecular genetics, transcriptomics

and protein characteristics and also allows for the construction of a comprehensive multi-discipline model to optimise the prognostication of early stage NSCLC patients.

In addition, this study aimed to investigate the reported claims of intratumour heterogeneity in *EGFR* mutation status which may provide an alternative mechanism of primary resistance to EGFR TKIs. The development of EGFR mutation specific IHC assays allows for within-cell analysis of the presence of EGFR mutations in a cost-effective manner; and the inclusion of antibodies to control for EGFR expression levels and fixation artefact enables more thorough investigation of this hypothesis.

Chapter 2 Methods

2.1 Review of relevant literature

A systematic review was not possible due to the lack of suitably robust controlled studies. The following literature searches were carried out using Web of Science (https://apps.webofknowledge.com) as the primary literature search tool. English language proceedings, papers or abstracts containing the topics "lung cancer" AND (morphology, histology, acinar, lepidic, papillary, solid, micropapillary, necrosis, mitosis, apoptosis, lymphocytes, pleomorphism, nucleoli, scar or dyscohesion) AND (prognosis or overall survival); results were filtered for with the terms (lung) OR (pulmonary) in the title (last updated 25/04/2017); only studies with at least 100 patients were considered relevant. English language proceedings, papers or abstracts containing the topics (EGFR tyrosine kinase inhibitor) AND (lung cancer) AND (protein OR FISH OR mutation) AND (survival OR response) (last updated 25/04/2017); only studies with at least 100 patients examining patient benefit from erlotinib, gefitinib or afatinib or osimertinib were considered relevant. English language proceedings, papers or abstracts containing the topics (EGFR) AND (mutation) AND (subtype) (last updated 25/04/2017) only studies with over 100 subjects were considered relevant. English language articles containing the topics; (NSCLC) AND (KRAS) AND (TKI) (last updated 08/03/2016); (KRAS) AND (smoking) AND (NSCLC) (last updated 10/03/2016); (TTF1) AND (EGFR mutations) AND (lung cancer) (last updated 10/03/2016); (TTF1) AND (EGFR mutations) (last updated 10/03/2016); (TTF1) AND (EGFR) (last updated 10/03/2016), (molecular pathology) AND (clinical) AND (lung cancer) (last updated 17/02/2016). English language proceedings, papers or abstracts containing the topics (lung cancer) AND (gene expression) AND (prognosis) AND (lung OR pulmonary) (last updated 19/12/2016) only studies with at least 100 patients were considered relevant. English language proceedings, papers or abstracts containing the topics (decision trees) AND (cancer) (last updated 05/12/16). No study has attempted to apply the methodology used in this research.

2.2 Clinical audit

Permission to carry out this clinical audit was granted by the NHS Lothian Information Governance team. Lung cancer cases requested for molecular pathology analysis between January 2011 and March 2014 were retrieved from the NHS Lothian molecular pathology clinical testing database including the following information: patient's surname, first name, date of birth, CHI number and sample identification number. Patient and sample identifiers were used to search NHS Lothian's laboratory LIMS system (iLaboratory) and the patients' electronic records (TRAKcare) to identify the following where available: smoking history, pack-years, smoke-free years, age at diagnosis, pathology diagnosis, tumour differentiation, clinical stage (including TNM), TTF1 IHC result, P63 IHC result, treatment given, age at death, *EGFR* mutation status, *KRAS* mutation status, ALK IHC results and ALK FISH results (undertaken in collaboration with Ms. Yuen Chun Kheng, University of Edinburgh as part of an undergraduate special study). Pack years was defined as the average number of cigarettes smoked per day, divided by 20 then multiplied by the number of years smoked. Histological/cytological diagnoses, derived from the diagnostic pathology report were classified into three groups as follows (in collaboration with Ms.Yuen Chun Kheng and Dr Anca Oniscu):

- [adenocarcinoma] or [non-small cell carcinoma favouring adenocarcinoma].
- NSCLC or carcinoma favouring a lung primary (including metastatic carcinoma, poorly differentiated pleomorphic carcinoma, non-small cell carcinoma showing neuroendocrine differentiation, non-small cell lung cancer showing mucoepidermoid differentiation, large cell undifferentiated carcinoma, metastatic poorly differentiated carcinoma, pleomorphic carcinoma).
- other diagnoses (including squamous cell carcinoma, mixed adenocarcinoma and large cell neuroendocrine carcinoma, non-small cell adenosquamous carcinoma, undifferentiated high grade probable sarcoma, large cell neuroendocrine, large cell undifferentiated tumour, basaloid version favouring sarcomatoid carcinoma, poorly differentiated NSCLC favouring squamous cell carcinoma, neuroendocrine carcinoma, metastatic large cell carcinoma either

poorly differentiated adenocarcinoma or large cell undifferentiated carcinoma, metastatic poorly differentiated carcinoma with neuroendocrine differentiation, mixed poorly differentiated squamous and large cell neuroendocrine carcinoma, low grade carcinoma, PNET and large cell neuroendocrine tumour).

Duplicate samples, tested from the same patient, were identified using Microsoft Excel 2016 and were removed for initial analyses between gender or smoking status and TNM, disease stage and diagnosis. Where a conflict occurred between clinicopathological parameters in duplicate samples the data from the metastasis or later sample was favoured. For all other analyses multiple samples from the same patient were included. Where possible, for cases where no clinical disease stage was specified, this was derived from the T, N, M according to the 7th Edition TNM staging system for lung cancer.[5] Smoking status was categorised as non-smoker, ex-smoker or current smoker. Smoking status was assumed to be ex-smoker where the number of smoke free years was available. Smoking status was also alternatively grouped as never smokers (non-smokers) or ever smokers (ex- and current smokers) and as not currently smoking (non- and ex-smokers) or currently smoking. Patients were also grouped as having smoked less than or equal to 5 pack-years (including non-smokers) vs greater than 5 pack-years; and less than or equal to 15 pack-years (including non-smokers) vs greater than 15 pack-years. Differentiation was as quoted, where available, from the histopathologists final diagnosis and was restricted to resection and wedge specimens only; cases were categorised as well differentiated, moderately differentiated, moderate to poorly differentiated or poorly differentiated; and alternatively as well to moderately differentiated or moderate to poor and poorly differentiated. *EGFR* and *KRAS* mutation status were categorised by the specific mutation and by the presence or absence (no mutation detected) of a mutation. *KRAS* mutations were also grouped as transversions versus transitions; and as G>T mutations versus other *KRAS* mutations. *ALK* status was defined as follows: ALK IHC negative and *ALK* FISH negative were considered negative; ALK IHC positive or equivocal and *ALK* FISH negative were considered *ALK* negative; ALK IHC positive and *ALK* FISH positive were *ALK* positive. An additional molecular variable was created

combining *EGFR, KRAS* and *ALK* status indicating a targeted treatment was available (sensitizing *EGFR* mutation or *ALK* rearrangement detected) or no targeted treatment was available (*KRAS* mutation, no *EGFR* mutation or resistance mutation, no *ALK* rearrangement). Samples with focal TTF1 positivity by IHC were categorised as positive and equivocal TTF1 or P63 IHC were removed from the analysis. The treatment given to a patient, where available, was categorised as no therapy, non-targeted therapy (including surgery, chemotherapy, pemetrexed, radiotherapy), EGFR tyrosine kinase inhibitors (first or second line) or ALK inhibitor. The dataset was anonymised and coded.

2.2.1 Cost modelling

The anonymised data from all patients diagnosed with lung cancer in NHS Lothian, NHS Borders, NHS Fife and NHS Dumfries and Galloway between April 2014 and March 2015 was supplied by the South East Scotland Cancer Network (SCAN). Patients were considered eligible for molecular pathology analysis according to 4 models as follows:

1. For the reflex model any patients with a histological or cytological diagnosis of adenocarcinoma or non-squamous NSCLC would be eligible for molecular pathology testing. *EGFR*, *KRAS* and *ALK* analysis was assumed to be carried out simultaneously.
2. Under the request model any patients with histopathology or cytology specimens diagnosed with stage III or IV disease were considered candidates for molecular testing. In addition, an estimate of the number of patients initially diagnosed with early stage cancer, whose disease would have progressed to stage III or IV, was calculated from the results of the clinical audit and included in the total number of eligible patients. *EGFR*, *KRAS* and *ALK* was assumed to be carried out simultaneously.
3. The serial model would include the same patient cohort as tested under the request model; however, the molecular analysis would be carried out sequentially. *KRAS* mutation analysis would be carried out first and only samples with no *KRAS* mutation detected would have *EGFR* mutation analysis,

similarly only samples with no *EGFR* mutation would progress to *ALK* analysis.

4. In the TTF1-serial model TTF1 IHC would be used as an initial screen and only TTF1 positive samples would be tested for *KRAS* mutations and if no mutation was detected would have *EGFR* analysis. All samples tested under the TTF1-serial model would have *ALK* analysis.

EGFR, *KRAS* and *ALK* analysis was assumed to be carried out as described in section 2.4. Reagent and consumables costs for *EGFR* and *KRAS* testing were calculated from orders placed between April 2014 and March 2015. The cost of staff input and laboratory overheads (estimated at 85%) was calculated from data supplied in the CMD ImPACT business case tool developed by the Royal College of Pathologists, Cancer Research UK and the Association of the British Pharmaceutical Industry (https://www.rcpath.org/cmd-impact.html). The cost of *ALK* IHC and ALK FISH, including consumables, staff and overheads were obtained from Pathology financial speradsheets (Department of Laboratory Medicine, NHS Lothian). A repeat rate of 0.5% was assumed for all analyses based on Molecular Pathology quality performance indicators. Equipment depreciation and maintenance have not been included in the costings and were assumed to be constant across the different models.

2.3 Morphological assessment

Permission for this study was granted by the NHS Lothian BioResource Tissue Governance Unit. The NHS Lothian iLaboratory database was searched to identify cases of stage I or stage II lung adenocarcinoma which were considered resection complete (R0) and were resected between 1996 and 2006. A cohort of 208 archived lung adenocarcinoma patient samples was identified and reviewed. Cases were restaged according to the 7th edition of the TNM classification of malignant tumours by Dr William Wallace (WW) (Consultant Histopathologist, NHS Lothian).[5] Demographic and clinical data including T stage, tumour size, N stage (nodal status), pleural involvement, disease stage, date sample received, site of recurrence (if applicable), date of recurrence (if applicable), date of death and cause of death were

collated by the Tissue Governance Unit. Where more than one lesion was present the size of the tumour was considered to be the largest.

All available tumour containing H&E stained sections from each of the 208 cases were examined by light microscopy. The approximate proportions of non-tumour tissue (including normal and scar tissue), and each morphological growth pattern were estimated for every section. The following descriptions, derived from Travis *et al*. were used to identify each growth pattern:[261]

- *Lepidic pattern*- cells growing along alveolar structures, alveolar walls may be thickened. Cells may be mucinous or non-mucinous, and be columnar, cuboidal or dome shaped with pale eosinophilic or clear to foamy cytoplasm.
- *Acinar pattern*- composed of cuboidal or columnar cells and forming acini or tubules and may be mucin producing. The cribiform pattern was also classified as acinar.
- *Papillary pattern*- forms secondary and tertiary structures, cells may be mucin-producing and cuboidal to columnar in shape.
- *Micropapillary pattern*- appearing free floating or tufts of tumour cells which lack a fibrovascular core.
- *Solid pattern*- sheets of cells with none of the above described structures.

Where micropapillary growth was observed in airspaces surrounded by acinar, lepidic or papillary growth the proportion of each pattern was estimated. The predominant growth pattern was determined for each case according to 4 different methods:

- A – Each pattern was expressed as a proportion of the whole section, including non-tumour tissue. The pattern with the largest total proportion, across all sections, was considered the predominant pattern.
- B – Each pattern was expressed as a proportion of the tumour in the section (not including non-tumour tissue). The pattern with the largest total proportion, across all sections, was considered the predominant pattern.

- C – The predominant pattern was determined for each section. The pattern predominant in the greatest number of sections was considered the predominant pattern for the case.
- D – The predominant pattern in a single representative section was considered the predominant pattern for the case.

If two growth patterns were equally predominant in a case the pattern with the highest morphological grade, as defined by Travis et al., was considered the predominant pattern; whereby predominantly lepidic tumours were considered low grade, predominantly papillary and acinar tumours were intermediate grade, and predominantly solid or micropapillary tumours high grade.[14] Cases were also grouped by predominant morphological grade.[14]

A growth pattern was considered to contribute to the total number of growth patterns if it was present at greater than 5% by area of the tumour in any section of the case. The presence of the micropapillary pattern at any proportion was also noted. The proportion of each growth pattern in the primary lesion, excluding nodal metastases, was calculated. In order to assess inter-observer agreement 10 cases, comprising 52 slides, were also assessed for the proportion of each growth pattern by a second observer (WW). Cohen's kappa statistic was used to assess inter-observer agreement between predominant patterns. The proportion of each growth pattern in every slide was compared between observers using Spearman's Rank Correlation.

Each case was also given a qualitative description detailing notable morphological features with particular reference to numbers of inflammatory cells, mitotic figures, apoptotic bodies, prominent nucleoli, cytological pleomorphism, necrosis, scar tissue within the tumour and evidence of dyscohesive tumour cells. For each case the presence of significant numbers of these features was categorised as follows. For mitotic figures and apoptotic bodies descriptive terms such as not seen, rare, quite rare and few were classified as "few" whereas terms including fairly common, common, and many were classified as "many". Descriptions of inflammatory infiltrate aimed to describe tumours with large numbers of lymphocytes in tissue surrounding tumour cell nests and terms including high inflammatory infiltrate were classified as

"inflammatory infiltrate present" whereas cases with no mention of inflammatory cells were classified as "inflammatory infiltrate not present". For necrosis, pleomorphism, scar tissue and dyscohesive tumour cells the features were classified as "present" if highlighted in the description and not present if no mention was made. Where different areas of the tumour showed varying levels of these features the highest classification was chosen to represent the tumour.

2.4 Molecular pathology

All molecular analyses, whether performed as part of routine NHS Lothian service and included in the audit or on samples used for research, were carried out as follows.

2.4.1 DNA isolation

All lung cancer specimens underwent histopathological assessment to determine the suitability of each assay against its validated limits of detection, evaluate the need for microdissection and to estimate the proportion of neoplastic cells. A minimum of 15% tumour cell content was considered necessary for successful mutation analysis. Three to 6 10µm tissue sections were cut from each tissue block using a microtome and fresh blade. Where macrodissection was required sections were mounted onto slides and using a marked haematoxylin and eosin (H&E) stained section as a guide, the area of interest was scraped from the slide using a fresh disposable scalpel and transferred to a 1.5ml microfuge tube. Otherwise rolled cut tissue sections were placed directly into a 1.5ml microfuge tube. The microtome was cleaned after each tissue block with isopropanol impregnated wipes. DNA isolation was carried out using the QIAamp DNA FFPE Tissue Kit (Qiagen) as follows. Buffer ATL (180µl) and 20µl of Proteinase K were added to the tissue, mixed and incubated overnight at 56°C in a water-bath. The lysate was incubated at 90°C for 1 hour before briefly centrifuging, after which 200µl of buffer AL and 200µl of absolute ethanol (VWRI) was added and mixed by pipetting. The lysate was transferred to a QIAamp MinElute® column and centrifuged for 1 minute at 13,000g. The column was washed with 500µl of Buffer AW1 and 500µl of Buffer AW2 before drying by centrifuging at 13,000 for 3 min. DNA was eluted in 100µl of Buffer ATE. For clinically tested samples and the validation and experimental cohorts *EGFR* and *KRAS* mutation analysis was carried

out as described below. For analysis of the 208 lung adenocarcinoma cohort only cases with metastatic (N1) disease were selected for mutational analysis for which nodal deposits were favoured over the primary lesion.

2.4.2 *EGFR* mutation analysis

EGFR mutation analysis was carried out using the *therascreen*® EGFR RGQ PCR kit (Qiagen) following manufacturer's instructions; 5µl of isolated DNA was added to each of 8 PCR mastermixes. This method was designed to detect 29 common somatic mutations in the *EGFR* gene. The limit of detection for each *EGFR* variant varied from 0.5% mutated alleles in a background of wild-type DNA for p.L861Q to 7.02% for p.T790M.

Where necessary, Sanger sequencing of the *EGFR* gene was also carried out using PCR primers designed to amplify exons 18, 19, 20, and 21 (adapted from Do *et al.*) (Table 2.1).[262]

Table 2.1: The details of PCR primers used for Sanger sequencing of the EGFR gene, each PCR primer included a 5` universal sequencing tag.

Primer name	Exon	Function	Sequence (5′ to 3′)
EGFR18_F	18	PCR	UNI-F-CATGGTGAGGGCTGAGGTGA
EGFR18_R	18	PCR	UNI-R-CCCCACCAGACCATGAGAGG
EGFR19_F	19	PCR	UNI-F-GTGCATCGCTGGTAACATCCA
EGFR19_R	19	PCR	UNI-R-GGAGATGAGCAGGGTCTAGAGCA
EGFR20_F	20	PCR	UNI-F-CGCATTCATGCGTCTTCACC
EGFR20_R	20	PCR	UNI-R-CTATCCCAGGAGCGCAGACC
EGFR21_F	21	PCR	UNI-F-TGGCATGAACATGACCCTGAA
EGFR21_R	21	PCR	UNI-R-CAGCCTGGTCCCTGGTGTC
UNI_F	-	Sequencing	GTAGCGCGACGGCCAGT
UNI_R	-	Sequencing	CAGGGCGCAGCGATGA

To 1µl of DNA was added 10µl of HotStarTaq *Plus* 2x Mastermix (Qiagen), 10pmol of each forward and reverse primer, 1µl of dimethyl sulfoxide (Sigma-Aldrich) and nuclease free water (NFW) to 20µl. Reactions were incubated at 95°C for 5 minutes followed by 40 cycles of 94°C for 15 seconds, 60°C for 1 minute and 72°C for 1 minute; followed 72°C for 10 minutes. Unincorporated primers and nucleotides were removed by treatment of 5µl of each PCR product with 2µl of Illustra ExoProStar (GE Healthcare Life Sciences) and incubated at 37°C for 15 mins followed by 80°C for 15

mins. To 1µl of treated PCR product was added 1µl (3.3pmol) of sequencing primer (UNI_F or UNI_R), 1µl of sequencing reaction buffer, 0.5µl of BigDye Terminator (Applied Biosystems®) and 7.5µl of NFW. Each PCR product was sequenced in the forward and reverse directions in separate wells using UNI_F and UNI_R primers respectively. The sequencing reaction was incubated at 96°C for 1 minute, followed by 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes before cooling to 4°C. Products were precipitated with 2µl of 3M sodium acetate (Sigma-Aldrch) and 28µl of absolute ethanol (VWRI), followed by centrifugation at 1650g for 30 minutes. The supernatant was discarded and the pellet washed with 75µl of 70% ethanol before centrifuging at 1650g for 10 minutes. The supernatant was removed and the pellet dried before resuspending in 10µl of Hi-Di Formamide™ (Applied Biosystems®). Purified DNA was denatured at 96°C for 5 minutes before cooling to 4°C. The products were then run on an ABI 3130XL Genetic Analyser (Applied Biosystems). Resulting sequences were analysed using Mutation Surveyor (Soft Genetics) to detect variants compared to the Genbank reference sequence for *EGFR* (NM_005228.3).

2.4.3 *KRAS* mutation analysis

KRAS mutation analysis was carried out by PCR followed by pyrosequencing using a method designed to detect mutations in codons 12, 13 and 61 (reference sequence NM_004985.4). Two PCR reactions were carried out for each specimen, one for *KRAS* codons 12 and 13, one for codon 61. To 5µl of isolated DNA was added 12.5µl HotStarTaq *plus* 2x mastermix, 2.5µl Coral Load (Qiagen, UK), 10pmol of each primer (table 2.2) and nuclease free water to a total of 25µl.

Table 2.2: PCR and pyrosequencing primer sequences for the KRAS mutation analysis assay.

Primer name	Function	Sequence (5' to 3')
KRAS 1213F	PCR forward primer	GGCCTGCTGAAAATGACTG
KRAS 1213R	PCR reverse primer	Biotin-GCTGTATCGTCAAGGCACTCT
KRAS 1213 Seq	Pyrosequencing primer	TTGTGGTAGTTGGAGCT
KRAS 61 F1	PCR forward primer	Biotin-TGGAGAAACCTGTCTCTTGGATAT

KRAS 61 R1	PCR reverse primer	CTGGTCCCTCATTGCACTGTACTC
KRAS 61 Seq	Pyrosequencing primer	CCTCATTGCACTGTACTC

Cycling conditions were as follows: 95°C for 5 minutes followed by 35 cycles of 94°C for 30s, 58°C for 30s and 72°C for 30s; and a final incubation of 72°C for 10 minutes. The amplified PCR product (10µl) was sequenced using a PyroMark Q24 pyrosequencer (Qiagen), according to manufacturer's instructions, with the following dispensation orders: codons 12 and 13 5`-TACGACTCAGATCGTAG, codon 61 5`-GCTCAGTCAGACT. The results of pyrosequencing were analysed with PyroMark Q24 software using the KRAS mutation analysis plug-in (Qiagen, UK). The limit of detection of this assay is 5% mutant alleles in a background of wild-type DNA for codons 12 and 13 and 6.5% for codon 61.

2.4.4 *ALK* rearrangement analysis

Samples were screened for increased ALK protein expression by IHC using a 1 in 20 dilution of the D5F3 antibody (Cell Signalling Technology, USA) on a Bond-III system (Leica, UK). Tumours with positive staining for ALK were tested for the presence of *ALK* gene rearrangements using the Vysis Break-Apart probe (Abbott Molecular) following manufacturer's instructions. Fifty tumour cells were scored and a tumour was considered positive for the rearrangement if at least 15% of the cells showed a rearrangement.

2.4.5 Transcriptomics

Gene expression analysis was performed by Myriad Genomics Inc. (Salt Lake City, USA) and the data returned in an excel spreadsheet. The method was as described in Bueno *et al*. [263], in brief: a single representative FFPE tumour block was selected for gene expression analysis. After macrodissection total RNA was extracted from 5 to 10µm sections and DNase-treated. Following reverse transcription and preamplification with a multiplex of all PCR primer sets, custom Taqman Low Density arrays were used to amplify 15 housekeeping genes (*RPL38*, *UBA52*, *RPL4*, *RPS29*, *SLC25A3*, *CLTC*, *RPL37*, *PSMA1*, *RPL8*, *PPP2CA*, *TXNL1*, *MMADHC*, *PSMC1*,

RPL13A and *MRFAP1*) and 31 cell-cycle related genes (*ASPM, CDCA8, MCM10, FOXM1, CDC20, CDKN3, BIRC5, DLGAP5, KIF20A, BUB1B, PRC1, TK1, CEP55, PBK, RAD54L, NUSAP1, RRM2, KIAA0101, ORC6L, RAD51, CENPM, SKA1, CENPF, KIF11, PTTG1, CDC2, DTL, PLK1, CDCA3, ASF1B* and *TOP2A*). The average normalised quantity of all cell-cycle related genes was calculated and the CCP score returned. The mPS was calculated as 20×((0.33×CCP)+(0.52×Stage))+15 where clinical stage Ia= 1, Ib= 2, IIa= 3, IIb= 4.

2.4.6 Immunohistochemistry

Three commercially available antibodies were validated for use in IHC with archived lung adenocarcinoma FFPE tissue (Pathology, Dept. Laboratory Medicine, NHS Lothian). The first antibody, 31G7, was a mouse monoclonal designed to bind to the extracellular domain of all EGFR protein (CellPath Ltd., UK); referred to below as tEGFR. The second, D6B6, was a rabbit monoclonal antibody designed to bind to EGFR protein bearing a p.E746_A750del mutation but not to normal EGFR (Cell Signalling Technology, USA); referred to as DEL. The third (43B2) was a rabbit monoclonal antibody designed to bind to EGFR with a p.L858R mutation, but not to normal EGFR (Cell Signalling Technology, USA); referred to below as L858R. Validation tissue was selected from the results of the clinical audit; samples identified with deletions in exon 19 of *EGFR* were further characterised by Sanger sequencing. Cases were selected for validation on the basis of amount of tissue available, the estimated proportion of neoplastic cells and the age of the specimen (very recent specimens were avoided). On this basis 6 samples were selected with the p.E746_A750del mutation, 6 samples with p.L858R mutations and 5 cases with no *EGFR* mutation. After initial development, the protocols for the IHC assays were as follows. tEGFR was diluted 1 in 14, with epitope retrieval for 10 minutes in buffer E. DEL was diluted to 1 in 100, with retrieval 40 minutes in buffer ER2. L858R was diluted to 1 in 75 with 40 minutes retrieval in solution ER2. All IHC was carried out on 3µm tissue sections on a Bond III automated stainer (Leica).

All *EGFR* mutated cases were included in IHC analysis. Five 3µm sections were cut from each tumour-containing tissue block, sections were stained with tEGFR, DEL,

L858R, the routinely performed TTF1 and H&E (Pathology, Dept. Laboratory Medicine, NHS Lothian). All stained slides were scanned at 20x resolution with a Zeiss Axio Scan Z1. No formal scoring criteria were used, each image was visually examined for evidence of intratumour heterogeneity and compared to corresponding areas stained with tEGFR and TTF1.

2.5 Statistical analysis

All statistical analyses were carried out in IBM SPSS Statistics Version 20. For correlations between categorical variables Pearson's χ^2 test, and χ^2 test for trend (χ^2T) where appropriate, were used unless otherwise stated; Fisher's exact test was used when the expected number in any group was less than 5.

Continuous variables (lesion size, pack-years, smoke-free years, age at diagnosis and proportion of each growth pattern) were tested for normality of distribution using the Shapiro-Wilk test. Associations between continuous and categorical variables were investigated using the non-parametric Mann-Whitney U test or Kruskal Wallis test, depending on the number of outcome groups. Spearman's rank correlation was used to investigate relationships between continuous variables. Associations between the following pairs of variables were not analysed due to their known confounding effect: age at diagnosis and pack-years, age at diagnosis and smoke-free years, pack-years and smoke-free years, disease stage and type of sample, diagnosis and type of sample, diagnosis and stage, diagnosis and TTF1 or P63. Repeated measures ANOVA was used to investigate the difference in the proportion of each growth pattern between primary lesions and nodal metastases. Associations between individual variables and Disease specific survival (DSS) were explored using the Kaplan Meier method with significance assessed by the Log-Rank (Mantel-Cox) method. Cases of patients who died from non-lung cancer causes were censored.

Multivariable analysis for categorical outcome measures was carried out using binary logistic regression and for continuous outcome measures using Cox's proportional hazards model; Bonferroni's correction was applied in both analyses.

Decision tree analysis was carried out in SPSS version 20 using Classification and Regression Trees (CART) with 10-fold cross validation limited to 20 and 10 cases in each parent and child node respectively. DSS was calculated from the date of resection to date of death from lung cancer related causes and was limited to a maximum of 10 years. All patients without a date of death, who died after 10 years (whether lung cancer related or not), were given an OS of 10 years. A CART decision tree was created using currently available clinicopathological parameters only and compared to those which also included new morphological variables and additional molecular variables. CART analysis was also used to identify potential thresholds of growth patterns by including the proportion of each growth pattern individually as a predictor of DSS. The performance of prognostic models was tested by carrying out receiver operating characteristic analysis (ROC) and comparing the resulting area under the curve (c-statistic).

Chapter 3 Current molecular pathology practice is of benefit in a small number of cases

3.1 Background

The clinical implementation of targeted therapies for the treatment of non-squamous NSCLC has led to the need for molecular pathology analysis in samples from eligible patients. Analysis of mutations in the *EGFR* and *ALK* genes is now routinely used to identify patients most likely to benefit from treatments targeting key proteins in aberrantly activated signalling pathways.

The presence of mutations in the *EGFR* gene have been detected in approximately 10% of NSCLC from European Caucasian populations and have been associated with adenocarcinoma histology (9.9% to 12.8% *EGFR* mutation frequency) and non-smoking history.[124, 129, 264] *EGFR* mutations have also been reported to be more common in females;[124, 129, 264] however, this has not been supported in all studies and differing smoking habits between men and women may be a confounding factor.[132] Rearrangements in the anaplastic lymphoma kinase (*ALK*) gene, which predict response to ALK inhibitors, have been reported in 3-5% of NSCLCs and have been reported to be associated with adenocarcinoma histology and non-smoking history.[177, 183] Activating hotspot mutations in the *KRAS* gene have been detected in approximately 35% of non-squamous NSCLC occurring in western populations[132] and are associated with a smoking history, most notably G to T transversions.[190, 192] The presence of *KRAS* mutations has been shown to be a poor prognostic factor in NSCLC[196, 265] and there are currently no therapies available which directly target *KRAS* mutated NSCLC. *EGFR*, *ALK* and *KRAS* have been reported to be mutually exclusive in the vast majority of cases.[183]

There is currently little information published on the incidence of *EGFR*, *ALK* and *KRAS* mutations in a Scottish population with robust statistical analysis to investigate clinicopathologic associations. This audit has been carried out on non-squamous NSCLC cases referred for analysis to NHS Lothian Molecular Pathology service over a 38-month period, and included patients from Lothian, Dumfries and Galloway, Fife

and the Borders. The data collected was used to explore the relationship between molecular characteristics and known clinicopathologic variables and to compare the prevalence of *EGFR*, *ALK* and *KRAS* mutations in a Scottish cohort with those previously published. In addition, based on the results of this audit, a cost modelling exercise has been carried out to explore more cost-effective algorithms for delivering molecular pathology testing of lung cancers in a routine healthcare setting.

3.2 Results

From Jan 2011 to March 2014 lung cancer molecular pathology testing was requested on 676 patients from a total of 710 specimens. Thirty-one patients had more than one sample tested because:

- an initial sample failed or was insufficient for testing for one or more molecular markers (14 patients).
- 2 or more samples were available and both were considered clinically relevant; for example, patients with synchronous tumours (16 patients).
- *EGFR* and *ALK* analysis was performed at different times, the later test being performed on a more recent sample (1 patient).

The majority of specimens tested were biopsies (62.1%), 23.4% were cell blocks prepared from cell suspensions (pleural fluid and endobronchial ultrasound fine needle aspirates), only 14.6% of samples were from surgical resections. The clinicopathologic characteristics of patients in the cohort are summarised in table 3.1.

Table 3.1: Clinicopathologic characteristics of lung cancer patients tested by NHS Lothian Molecular Pathology service between January 2011 and March 2014.

		Number	Percentage
Gender			
	Female	363	53.7%
	Male	313	46.3%
Age at diagnosis			
	Median	67	
	Mean (95% CI)	65.8 (65.0 to 66.6)	
	Range	30 to 90	
T stage			
	T1a	41	6.1%
	T1b	68	10.1%
	T2a	135	20.0%
	T2b	79	11.7%
	T3	120	17.8%
	T4	167	24.7%
	TX	66	9.8%
N stage			
	N0	173	25.6%
	N1	57	8.4%
	N2	193	28.6%
	N3	220	32.5%
	NX	33	4.9%
M stage			
	M0	214	31.7%
	M1a	120	17.8%
	M1b	320	47.3%
	MX	22	3.3%
Clinical stage			
	Ia or Ib	49	7.2%
	IIa or IIb	38	5.6%
	IIIa or IIIb	135	20.0%
	IV	448	66.3%
	Unknown	6	0.9%
Diagnosis			
	NSCLC NOS	74	10.9%
	Adenocarcinoma/ NSCLC favouring	581	85.9%
	Other	20	3.0%
	Unknown	1	0.1%
Smoking status			
	Non-smoker	80	11.8%
	Ex-smoker	261	38.6%
	Smoker	292	43.2%
	Unknown	43	6.4%
Total		676	100%

3.2.1 Correlation between clinicopathologic variables

Gender

Smoking status was available for 93.7% of the cohort; the majority of lung cancer patients with known smoking status were smokers (41.1%) or ex-smokers (46.4%), only 12.5% were non-smokers. The proportion of smokers and ex-smokers were similar between the genders; however, a significantly higher proportion of non-smokers were female (74.7%) ($p< 0.0005$) (table 3.2).

Table 3.2: The smoking status of male and female lung cancer patients referred for molecular pathology testing. Non-smokers were significantly more likely to be female than male ($p< 0.0005$). Significance was calculated using the χ^2 test.

	Non-smoker	Ex-smoker	Smoker	Total	Significance
Gender					p< 0.0005
Female	60 (75%)	126 (48.3%)	155 (53.1%)	341	
Male	20 (25%)	135 (51.7%)	137 (46.9%)	292	
Total	80 (100%)	261 (100%)	292 (100%)	633	

Male current and former smokers had smoked significantly more pack-years than females ($p= 0.01$) (figure 3.1). The number of smoke-free years in the ex-smoking population was not correlated with gender.

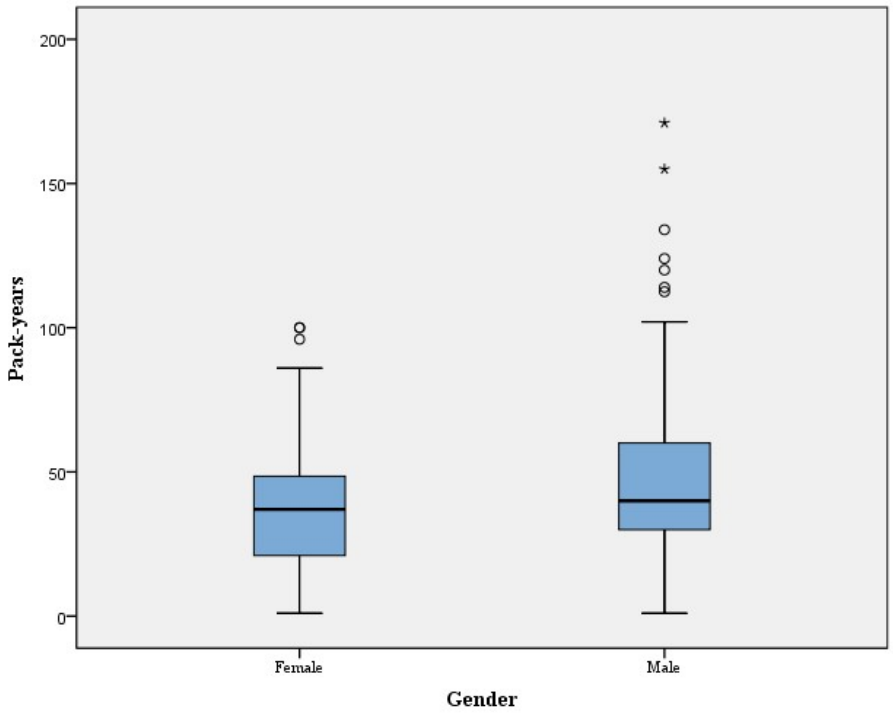


Figure 3.1: Box-plot showing the number of pack-years smoked by female and male ex- and current smokers. Males smoked significantly more pack-years (Mann-Whitney U test p= 0.01).

Metastatic status was available for 654 patients; female patients were more likely than males to have no distant metastases (M0) and whilst the proportion of female and male patients with thoracic metastases was similar, male patients were more likely to have extra-thoracic metastases (M1b) (p= 0.033). This association remained significant when patients were grouped into those with and without metastatic disease (p= 0.021) (table 3.3).

Table 3.3: The association between gender and metastatic status in patients referred for molecular pathology testing. The percentage of female and male patients with each M stage are shown in brackets. Significance was calculated using the χ^2 test.

		Female	Male	Total	Significance
M Stage					p= 0.033
	M0	128 (36.6%)	86 (28.2%)	214	
	M1a	66 (18.9%)	54 (17.7%)	120	
	M1b	155 (44.4%)	165 (54.1%)	320	
M Stage grouped					p= 0.021
	No metastases	128 (36.7%)	86 (28.2%)	214	
	Metastases	221 (63.3%)	219 (71.8%)	440	
Total		349 (100%)	305 (100%)	654	

Smoking history

Smokers were diagnosed with lung cancer at an earlier age than non- or ex-smokers ($p < 0.0005$) and there was no significant difference between non- and ex-smokers (figure 3.2).

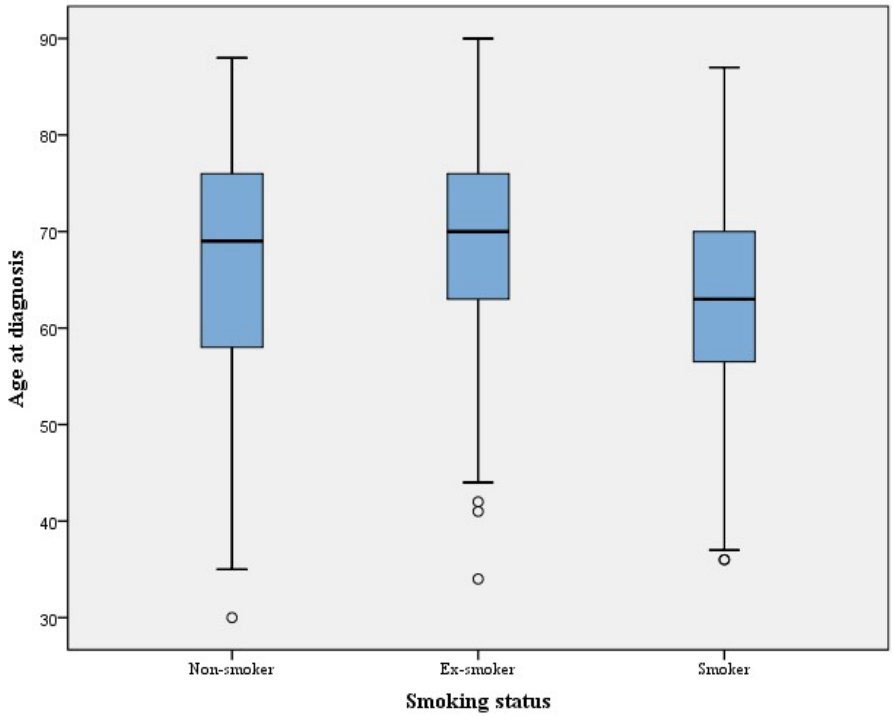


Figure 3.2: Box-plot showing the age at diagnosis of lung cancer patients grouped by smoking status. Smokers were diagnosed with lung cancer at a significantly earlier age than non- or ex-smokers (Kruskall Wallis test $p < 0.0005$).

There was no significant difference between the presence or absence of metastatic disease at diagnosis in patients grouped by their smoking status ($p = 0.763$); however, when metastatic status was classified as no metastases (M0), thoracic metastases only (M1a) or extra-thoracic metastases (M1b) the association with smoking status was statistically significant ($p < 0.0005$). Smokers with metastatic disease were more likely to have extra-thoracic metastases (table 3.4); there was, however, there was no significant difference between the number of pack-years smoked in different metastatic groups ($p = 0.535$).

Table 3.4: The relationship between smoking status and the presence of metastatic disease at diagnosis in lung cancer patients referred for molecular pathology testing. The percentage of non-smokers, ex-smokers and smokers with each M stage is shown in brackets. Significance was calculated using the χ^2 test.

		Non-smoker	Ex-smoker	Smoker	Total	Significance
M Stage						p< 0.0005
	M0	24 (30.4%)	80 (31.5%)	97 (33.9%)	201	
	M1a	21 (26.6%)	65 (25.6%)	31 (10.8%)	117	
	M1b	34 (43.0%)	109 (42.9%)	158 (55.2%)	301	
M Stage grouped						p= 0.763
	No metastases	24 (30.4%)	80 (31.5%)	97 (33.9%)	201	
	Metastatic disease	55 (69.6%)	174 (68.5%)	189 (66.1%)	418	
Total		79 (100%)	254 (100%)	286 (100%)	619	

Ex-smoking patients with a higher nodal status had significantly lower smoke-free years than those with lower nodal status (p= 0.023) (figure 3.3).

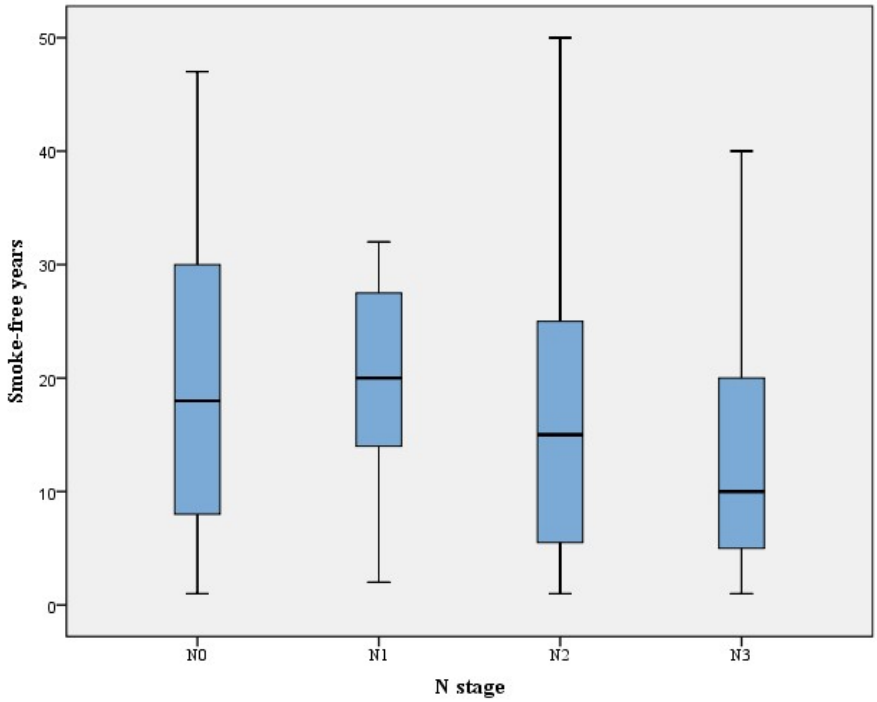


Figure 3.3: Box plot showing the difference in smoke-free years in ex-smoking patients grouped by N stage. The number of patients in the N0, N1, N2 and N3 groups was 52, 16, 63 and 54 respectively (Kruskall Wallis test p= 0.023).

Although there was no significant association between clinical stage and smoking status or pack-years, the number of smoke-free years was correlated with clinical stage (p= 0.004) (figure 3.4); suggesting that patients who gave up smoking a long time ago

presented with earlier stage disease than those who stopped more recently. However, the number of patients with early stage disease was small with only 13 patients each in the stage I and stage II groups.

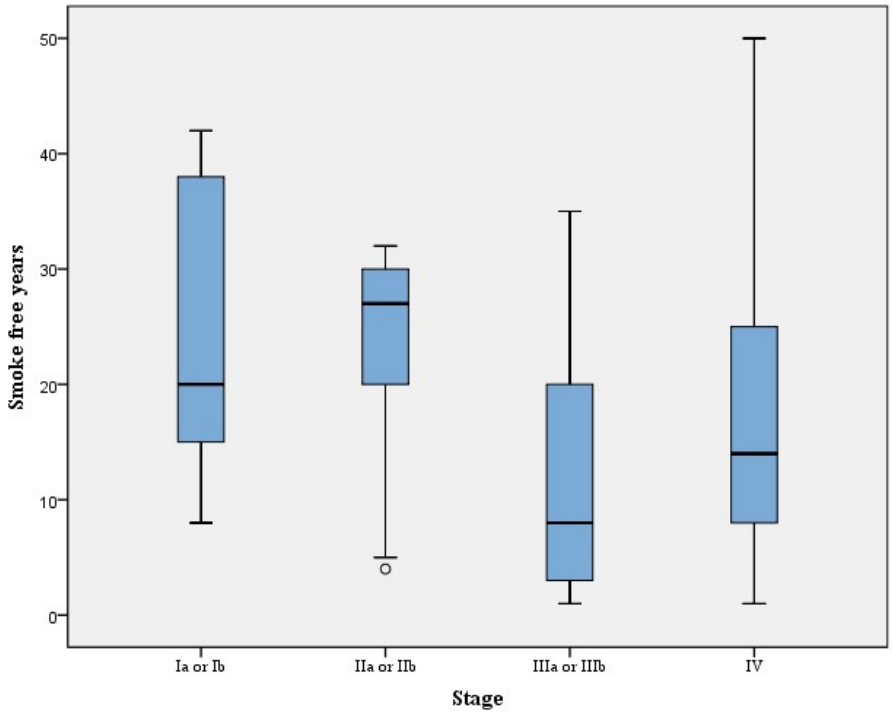


Figure 3.4: Box plot showing the number of smoke-free years in ex-smokers grouped by clinical stage (Kruskall Wallis test p= 0.004). The stage Ia or Ib group included 13 patients, the stage IIa or IIb group 13 patients, the stage IIIa or IIIb group 37 patients and stage IV group 138 patients.

3.2.2 Correlations between molecular genotype and clinicopathological characteristics

The results of molecular analysis of the 710 specimens are summarised in table 3.5. In total, 706 samples were tested for *EGFR* mutations, of which 50 samples failed to yield satisfactory results either due to failed analysis, insufficient neoplastic cell content or insufficient DNA yield. *EGFR* mutations were detected in 10.52% of successfully tested samples. The exon 21 p.(L858R) mutation and deletions in exon 19 comprised 44.9% and 37.7% of detected *EGFR* mutations respectively.

KRAS testing was successfully performed on 244 samples of which 36.5% of patients were found to have a *KRAS* mutation. The most common *KRAS* mutation was c.34G>T p.(G12C) (31.5% of *KRAS* mutations) followed by c.35G>T p.(G12V) (22.5%),

c.35G>A p.(G12D) (16.9%), and c.35G>C p.(G12A) (10.1%) with c.37G>T p.(G13C), c.183A>C p.(Q61H), c.183A>T p.(Q61H), c.38G>A p.(G13D), c.182A>T p.(Q61L), c.34G>A p.(G12S) and c.34_35delGGinsTT p.(G12F) each making up less than 5% of *KRAS* mutations.

Of the 309 patients tested for the presence of *ALK* gene rearrangements, using a combination of IHC and/or FISH, only 7 patients (2.26%) were found to have a rearrangement. *EGFR* mutations, *KRAS* mutations and *ALK* gene rearrangements were mutually exclusive.

Table 3.5: Summary of the results of molecular pathology analysis on specimens from lung cancer patients referred for molecular pathology analysis. Significance was calculated using the χ^2 test unless otherwise indicated.

Mutation status		Gender		Total	Significance
		Female	Male		
EGFR					p= 0.001
	No Mutation detected	300 (85.7%)	287 (93.8%)	587	
	Mutation detected	50 (14.3%)	19 (6.2%)	69	
	Deletion in exon 19	*18*	*8*	*26*	
	p.(L858R)	*23*	*8*	*31*	
	p.(L861Q)	*2*	*0*	*2*	
	p.(G719X)	*2*	*1*	*3*	
	Insertion in exon 20	*2*	*0*	*2*	
	p.(L747P)	*0*	*2*	*2*	
	p.(L858R) and p.(T790M)	*2*	*0*	*2*	
	p.(G719X) and p.(L861Q)	*1*	*0*	*1*	
Total		350 (100%)	306 (100%)	656	
KRAS					p= 0.150
	No mutation detected	74 (59.2%)	81 (68.1%)	155	
	Mutation detected	51 (40.8%)	38 (31.9%)	89	
	p.(G12C)	*16*	*12*	*28*	
	p.(G12V)	*15*	*5*	*20*	
	p.(G12D)	*7*	*8*	*15*	
	p.(G12A)	*6*	*3*	*9*	
	p.(G13C)	*0*	*4*	*4*	
	p.(Q61H)	*2*	*2*	*4*	
	p.(Q13D)	*1*	*2*	*3*	
	p.(Q61L)	*3*	*0*	*3*	
	p.(G12S)	*0*	*2*	*2*	
	p.(G12F)	*1*	*0*	*1*	
Total		125 (100%)	119 (100%)	244	
ALK					p= 0.509[a]
	Negative	151 (98.1%)	151 (97.4%)	302	
	ALK rearrangement	3 (1.9%)	4 (2.6%)	7	
Total		154 (100%)	155 (100%)	309	
TTF1					p= 0.037
	Negative	52 (19.5%)	64 (27.5%)	116	
	Positive	214 (80.5%)	169 (72.5%)	383	
Total		266 (100%)	233 (100%)	499	

aFisher's exact test

Gender

EGFR mutations were detected in 14.3% of female patients but only 6.2% of male patients, this difference was found to be significant (p= 0.001). There was no evidence to suggest a significant correlation between the presence of *KRAS* mutations or *ALK* rearrangements and gender alone. TTF1 IHC was carried out in 499 specimens of which 76.8% showed positive staining; significantly more females (80.4%) than males (72.5%) were TTF1 positive (p= 0.037). Although it is likely that the increased proportion of adenocarcinomas in females, which may in turn be linked to differing smoking habits, would increase the frequency of TTF1 positivity in this group.

TNM Staging

The presence of *EGFR* mutations was significantly more common in patients with lower T stages (p= 0.036 and p= 0.015 for individual and grouped T stage respectively) (table 3.6).

Table 3.6: The association between T stage and the presence of EGFR mutations. The proportion of each T stage group with and without EGFR mutations is shown in brackets. Significance was calculated using the χ^2 test and χ^2 test for trend as indicated.

		EGFR		Total	Significance
		No mutation	Mutation		
T stage					p= 0.036 (χ^2 test for trend p= 0.034)
	T1a	32 (86.5%%)	5 (13.5%)	37 (100%)	
	T1b	51 (79.7%)	13 (20.3%)	64 (100%)	
	T2a	121 (87.7%)	17 (12.3%)	138 (100%)	
	T2b	68 (89.5%)	8 (1.3%)	76 (100%)	
	T3	115 (95.8%)	5 (4.2%)	120 (100%)	
	T4	141 (89.2%)	17 (10.8%)	158 (100%)	
T stage grouped					p= 0.015
	T1a and T1b	83 (82.2%)	18 (17.8%)	101 (100%)	
	T2a and higher	445 (90.4%)	47 (9.6%)	492 (100%)	
Total		528	65	593	

Fisher's exact test showed that the presence of *ALK* gene rearrangements was associated with nodal status; in this cohort, all *ALK* rearrangements were detected in patients with N3 disease. However, given that only 7 *ALK* positive samples were detected, no reliable conclusions can be drawn (table 3.7).

Table 3.7: The incidence of ALK gene rearrangements grouped by the patient's nodal status. Significance was calculated using Fisher's exact test.

		ALK rearrangement		Total	Significance
		No rearrangement	Rearrangement		
Nodal status					p= 0.003
	N0	79	0	79	
	N1	31	0	31	
	N2	86	0	86	
	N3	88	7	95	
Total		284	7	291	

Smoking history

There was a statistically significant association between smoking status and the presence of *EGFR* mutations (p< 0.005); 34.6% of non-smokers were found to have an *EGFR* mutation compared to 9.1% of ex-smokers and 5.61% of smokers (table 3.8). Although the difference between non-smokers and ex-smokers was significant (p< 0.001) there was no difference between ex-smokers and smokers (p= 0.121).

Table 3.8: The number (and proportion) of non-, ex- and current smokers with, and without mutations in EGFR, KRAS and ALK. Significance was calculated using the χ^2 test unless otherwise indicated.

		Smoking status			Total	Significance
		Non-smoker	Ex-smoker	Smoker		
EGFR mutations						p< 0.0005
	No mutation	51 (65.4%)	230 (90.9%)	269 (94.4%)	550	
	Mutation	27 (34.6%)	23 (9.1%)	16 (5.6%)	66	
Total		78 (100%)	253 (100%)	285 (100%)	616	
KRAS mutations						p= 0.001
	No mutation	31 (88.6%)	57 (60%)	60 (58.3%)	148	
	Mutation	4 (11.4%)	38 (40%)	43 (41.7%)	85	
Total		35 (100%)	95 (100%)	103 (100%)	233	
ALK rearrangement						p< 0.0005[a]
	No rearrangement	29 (82.9%)	126 (100%)	131 (99.2%)	286	
	Rearrangement	6 (17.1%)	0 (0%)	1 (0.8%)	7	
Total		35 (100%)	126 (100%)	132 (100%)	293	

[a]*Fisher's exact test*

Patients with *EGFR* mutations were found to have smoked fewer pack-years than the non-mutated group (p< 0.0005) (figure 3.5). There was also a trend towards a greater number of smoke-free years in ex-smoking patients with *EGFR* mutations, although this did not reach statistical significance, (p= 0.053) (figure 3.6).

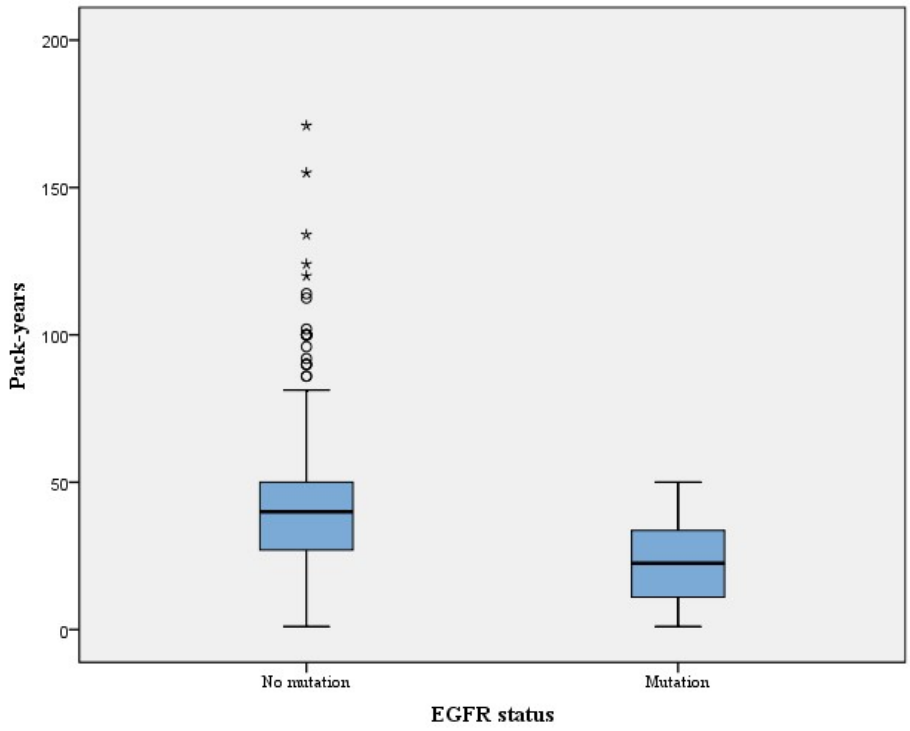


Figure 3.5: Box plot of pack-years smoked, in current and ex-smokers, grouped by EGFR mutation status; patients with EGFR mutations smoked significantly fewer cigarettes than those without EGFR mutations (Mann-Whitney U test p< 0.0005). No mutation detected n= 266, mutation detected n= 20.

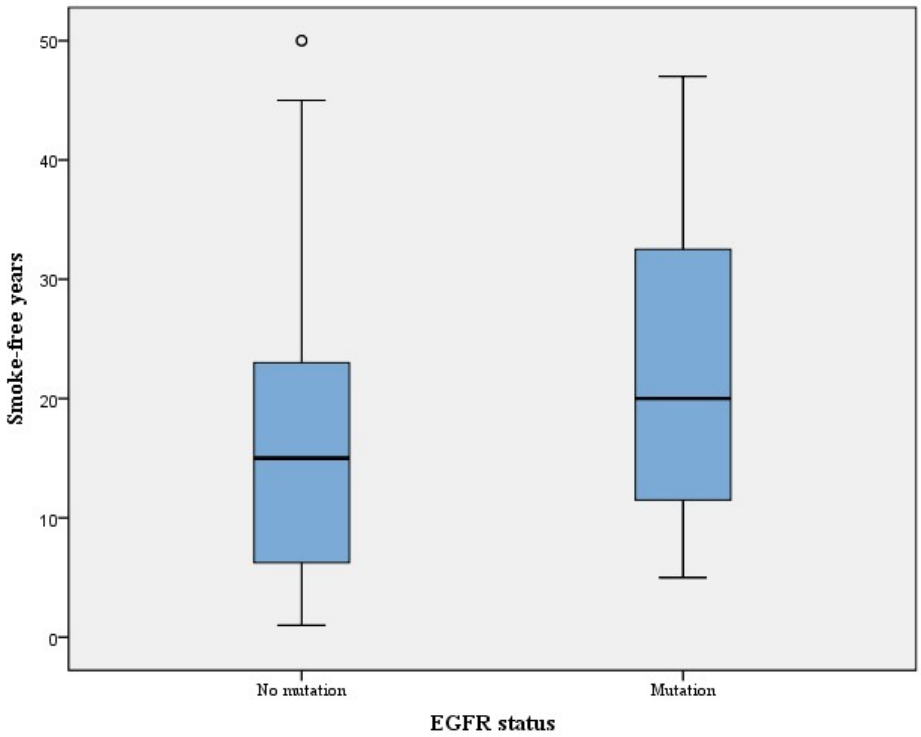


Figure 3.6: Box plot of the number of smoke-free years in ex-smokers grouped by EGFR mutation status. Patients with mutations tended to have been smoke-free for longer than those without EGFR mutations although this did not reach statistical significance (Mann-Whitney U test p= 0.053). No mutation detected n= 171, mutation detected n= 16.

KRAS mutations were detected in 11.43% of non-smokers, 40% of ex-smokers and 41.75% of current smokers; this difference was statistically significant (p= 0.001) (table 3.8). There was no significant difference in the frequency of *KRAS* mutations between ex- and current smokers and no significant difference in pack-years or smoke-free years between patients with or without a *KRAS* mutation. There was no significant association between individual *KRAS* mutations, type of *KRAS* mutation (transversion vs transition) or nucleotide change (*KRAS* G>T mutations vs other *KRAS* mutations) and the smoking status, pack-years or smoke-free years.

Although the number of *ALK* rearranged samples in the cohort was very limited (n= 7) Fisher's exact test showed a significant association with smoking status (p< 0.0005); 6 of the 7 *ALK* positive patients were non-smokers (table 3.8). The number of pack-years was not available for the smoking patient with an ALK rearrangement.

Age at diagnosis

There was no significant difference in age at diagnosis between patients with or without *EGFR* mutations, *KRAS* mutations or *ALK* rearrangements (p= 0.413, p= 0.854 and p= 0.100 respectively).

Protein expression

A total of 461 patient samples were analysed for both TTF1 expression using IHC and *EGFR* mutations (table 3.9). Although a positive TTF1 result was not a good predictor of the presence of *EGFR* mutations (positive predictive value 13.4%), no samples with *EGFR* mutations were found to be negative for TTF1 (negative predictive value 100%). The presence of *KRAS* mutations was not significantly associated with TTF1 expression (p= 0.214). Although all *ALK* rearranged tumours were TTF1 positive the small number of rearranged samples did not allow for robust conclusions to be drawn.

Table 3.9: The results of TTF1 IHC in patients with and without EGFR, KRAS and ALK mutations. Significance was calculated using the χ^2 test unless otherwise indicated.

		TTF1 IHC			Significance
		Negative	Positive	Total	
EGFR mutations					p< 0.0005
	No mutation	109 (100%)	305 (86.6%)	414	
	Mutation	0 (0%)	47 (13.4%)	47	
Total		109 (100%)	352 (100%)	461	
KRAS mutations					p= 0.214
	No mutation	28 (71.8%)	81 (60.9%)	109	
	Mutation	11 (28.2%)	52 (33.1%)	63	
Total		39 (100%)	133 (100%)	172	
ALK rearrangements					p= 0.312[a]
	No rearrangement	44	165	209	
	Rearrangement	0	5	5	
Total		44	170	214	

[a] *Fisher's exact test*

3.2.3 Multivariable analysis of factors associated with molecular characteristics

Both smoking status and gender were identified in univariable analysis as significantly associated with the patient's metastatic status, multinomial logistic regression was

carried out to investigate the independence of these variables. Non-smokers (odds ratio (OR) 2.995, 95% CI 1.527 to 5.874 p= 0.001) and ex-smokers (OR 3.706, 95% CI 1.878 to 5.0405 p< 0.0005) were more likely than smokers to have M1a rather than M1b disease, indicating that smokers were more likely to have extra-thoracic metastases at the time of diagnosis. Women were more likely than men to have no metastases compared to extra-thoracic metastases (OR 1.607, 95% CI 1.114 to 2.320 p=0.011).

Although both gender, smoking status and T stage were shown to be individually correlated with the presence of *EGFR* mutations, binary logistic regression showed only smoking status was independently associated with *EGFR* status. The OR of non-smokers having an *EGFR* mutation compared to current smokers was 10.35 (p= 0.012, CI 2.22 to 48.30) whereas the OR of ex-smokers having an *EGFR* mutation was not statistically different to that of smokers (p=0.068). In order to fully explore the relationship between gender and the presence of *EGFR* mutations, binary logistic regression was repeated without including T stage in the model; only smoking status was found to be independently associated with *EGFR* mutation status.

In univariable analysis the presence of *KRAS* mutations was significantly associated with smoking but not gender. However, 11.5% of female non-smokers had a *KRAS* mutation, compared to 45.8% of female ex-smokers and 52.2% of female smokers (p= 0.002); whereas in males 11.1% of non-smokers, 34.0% of ex-smokers and 33.3% of smokers had *KRAS* mutations (p= 0.457). This may suggest that female ex- or current smokers were more likely than their male counterparts to have a *KRAS* mutation, even though the number of pack-years is higher in males (figure 3.1).

3.2.4 Treatment of lung cancer based on molecular pathology analysis

The treatments given were available for 65 patients with *EGFR* mutations and 7 patients with *ALK* rearrangements. Fifty-one patients with *EGFR* mutated tumours were treated with EGFR targeted therapy (erlotinib, gefitinib and/or afatinib); 1 patient was not treated and 13 patients were given other treatments (including surgery, chemotherapy, pemetrexed and radiotherapy). Of the patients who received other

treatments, two had mutations associated with resistance to EGFR TKIs (insertions in *EGFR* exon 20). One patient with a p.L747P mutation is known to have chosen to not have treatment.[266] Of the remaining 10 patients 8 had tumours harbouring deletions in exon 19 and 2 had double mutations (1 with p.G719X + p.L861Q and 1 with p.L858R + p.T790M). Of the 7 ALK positive patients 2 did not receive ALK inhibitors and were treated with other therapy. In total 80% of patients with a mutation predicting response to either EGFR TKIs or ALK inhibitors were treated with a targeted therapy (table 3.10).

Table 3.10: Summary of treatments given to patients with mutations predicting response to EGFR or ALK targeting therapies.

	Combined molecular characteristics		Total
	Predictive of response to targeted therapies	Predictive of resistance to targeted therapies	
Treated with targeted therapy	56 (80%)	8 (1.4%)	64
Not treated with targeted therapy	14 (20%)	550 (98.6%)	564
Total	70 (100%)	558 (100%)	628

3.2.5 Molecular pathology cost modelling

The estimated annual costs of lung cancer molecular testing were calculated based on 4 proposed algorithms as follows (summarised in table 3.11):

5. The reflex model - all patients diagnosed with non-squamous NSCLC would be would be tested for *EGFR*, *KRAS* and *ALK* mutations simultaneously.
6. The request model - all patients diagnosed with stage III or IV non-squamous NSCLC would be tested for *EGFR*, *KRAS* and *ALK* mutations simultaneously.
7. The serial model - all patients diagnosed with stage III or IV non-squamous NSCLC would be tested sequentially; only samples with no *KRAS* mutation detected would have *EGFR* mutation analysis and only samples with no *EGFR* mutation would progress to *ALK* analysis.
8. The TTF1-serial - all patients diagnosed with stage III or IV non-squamous NSCLC would be tested for TTF1 and ALK rearrangement, only TTF1 positive samples would be tested sequentially for KRAS and EGFR mutations.

Table 3.11: Overview of the investigated testing models.

Model	Cohort	Analyses
Reflex	All non-squamous non-small cell lung carcinoma	Simultaneous *EGFR*, *KRAS*, *ALK*
Request	Metastatic or locally advanced non-squamous non-small cell lung carcinoma (stage III and IV)	
Serial		*KRAS* on all, *KRAS*-wt tested for *EGFR*, *EGFR*-wt tested for *ALK*
TTF1-serial		TTF1 and *ALK* on all, TTF1+ tested for *KRAS*, *KRAS*-wt tested for *EGFR*

NB: KRAS-wt= no KRAS mutation detected, EGFR-wt= no EGFR mutation detected, TTF1+= positive for TTF1 protein expression by IHC.

Data supplied by SCAN showed that 400 patients in the network were diagnosed, with histological or cytopathological specimens, with non-squamous NSCLC in a 12-month period between April 2014 to March 2015 and would be eligible for testing by the reflex model. Of these, 370 had stage III or IV disease and would qualify for testing by the request model.

From the results of the clinical audit the frequencies of *EGFR* and *KRAS* mutations, *ALK* rearrangements and TTF1 expression were used to estimate the number of analyses that would be carried out under each model (table 3.12). In addition, in the Lothian cohort 14.8% of specimens that were tested by ALK IHC also had *ALK* FISH (data not shown).

Table 3.12: Estimated number of tests required when testing according to each model.

Biomarker	No. samples tested in each model			
	Reflex	Request	Serial	TTF1-serial
TTF1	Diagnostic purposes only			370
KRAS	400	370	370	285
EGFR	400	370	236	182
ALK IHC	400	370	212	370
ALK FISH	60	55	32	55

The estimated annual cost of lung cancer testing using the reflex model including reagents, staff time and overheads was £166,021 (£415.05 per sample); whereas the

annual cost of the request model was estimated to be £153,569, a 7.5% reduction. Utilizing the mutually exclusive relationship between *KRAS* and *EGFR* mutations and *ALK* rearrangements in the serial model was estimated to cost £111,662 (£301.79 per sample), a 32.7% reduction compared to the reflex model. If samples were screened for TTF1 expression prior to the serial model an additional £7,941 could be saved leading to a total reduction of 37.5% compared to the reflex model (table 3.13).

Table 3.13: Estimated annual cost of molecular pathology testing models

	Reflex	Request	Serial	TTF1-serial
Number of cases	400	370	370	370
TTF1 IHC	-	-	-	£5,491
Sample selection and DNA isolation	£16,979	£15,706	£15,706	£14,955
EGFR mutation analysis	£93,675	£86,649	£55,039	£42,160
KRAS mutation analysis	£18,900	£17,482	£17,482	£13,391
ALK IHC	£13,558	£12,541	£7,128	£6,534
ALK FISH	£12,231	£11,314	£6,430	£11,313
Reporting time	£10,678	£9,877	£9,877	£9,877
TOTAL	**£166,021**	**£153,569**	**£111,662**	**£103,721**

3.3 Discussion

An audit has been carried out into lung cancer patients who were referred for molecular pathology testing to NHS Lothian over a 38-month period. It should be noted that since the NHS Lothian service employs a largely request based system the resulting cohort was biased towards patients with stage III and IV disease, reflecting those eligible for treatment with EGFR and ALK targeting therapies. This audit considered patient characteristics, smoking behaviour, histopathology (type of sample, differentiation, diagnosis, T, N, M, stage), immunohistochemistry and molecular characteristics (*EGFR*, *KRAS* and *ALK* mutation status).

No statistically significant correlations were identified between histopathological and molecular characteristics; however, morphological variables should be analysed with caution as these qualitative factors are subject to inter-observer variation. In addition; tumour grade (estimated here according to differentiation level) cannot be reliably

assessed on small biopsies or cell blocks and was therefore restricted to resection samples, greatly reducing the number of useable data points.

Not surprisingly there was evidence that smoking history was associated with the development and progression of lung cancer. Current smokers were diagnosed with lung cancer at an earlier age than non- and ex-smokers; interestingly age at diagnosis was very similar between non- and ex-smokers suggesting a benefit to giving up smoking- a finding also reported by Zheng *et al*.[267] In addition, ex-smokers who ceased smoking a long time ago presented with lower nodal status and earlier stage disease than those who stopped more recently; however, the number of patients with early stage disease was very limited. Smokers were also more likely to have extra-thoracic metastases. Combined this data supports the theory that lung cancer patients with a significant smoking history present at an earlier age and with more advanced disease than non-smokers; which may reflect the poor prognosis of smoking associated NSCLC.[265]

The proportion of *EGFR*, and *KRAS* gene mutations found in this cohort was consistent with previously published studies in European Caucasian populations.[124, 264] However, the frequency of *ALK* gene rearrangements in this Scottish cohort is lower than previously published estimates;[177, 183] although, informal communication with clinical molecular pathology colleagues has confirmed that a lower incidence (1-2%) is in line with other UK populations (personal communication Prof. Manuel Salto-Tellez, Queen's University Belfast). The presence of *EGFR*, *KRAS* and *ALK* gene aberrations were mutually exclusive. In our cohort EGFR p.L858R mutations were more common than exon 19 deletions; whereas previous studies, carried out in larger Caucasian cohorts, showed exon 19 deletions to be more common than p.L858R.[124, 132] This variance may be a result of differing testing strategies. The *therascreen*® kit employed in our dataset uses allele specific PCR which will only detect very specific mutations and is, therefore, likely to miss some of the rarer deletions in exon 19. However, improved limits of detection for this method would be likely to detect the p.L858R single base pair substitution at lower allele frequencies than the Sanger sequencing or PCR restriction fragment length polymorphism methods used in these studies. [124, 132, 268] Prior reports have suggested a link between patient age and the presence of *EGFR*

and *ALK* mutations[151] which was not supported in this dataset or other analyses carried out in large Caucasian cohorts.[124, 264, 269] Higher frequencies of *EGFR* mutations have been reported in females and non-smokers;[124, 245, 264] however, multivariate analysis to assess independent associations was not used. In our cohort, when smoking and gender (or smoking, gender and T stage) were taken into account only smoking status was found to be independently associated with the presence of *EGFR* mutations; a similar lack of independent association with gender has been reported by other groups.[129, 131, 269]

A large number of patients in this Scottish cohort harboured *KRAS* mutations in their tumours (36.5%) which is higher than in previously published datasets (approximately 26%); although, this difference is likely to reflect the much higher proportion of current and ex-smokers in this population.[132, 264] The KRAS G to T transversion mutations p.G12C and p.G12V were the most commonly detected (31.5% and 22.5% of KRAS mutations respectively) with proportions very similar to those previously reported in large cohorts;[132, 264, 270] this high incidence of transversions are likely to be associated with smoking.[132, 191] As has been previously reported, patients with a history of smoking were more likely than non-smokers to have *KRAS* mutations.[132, 264, 270] Although the number of pack-years was lower, female smokers had a higher incidence of *KRAS* mutations than male smokers indicating a possible increased susceptibility to tobacco smoke in women; an association that was also reported by Dogan *et al*.[132] In contrast to age at diagnosis, the molecular profile of ex-smokers was very similar to smokers indicating that stopping smoking had little effect on the incidence of *EGFR* and *KRAS* mutations, a finding also noted by Dogan *et al*. [132] The differing molecular characteristics of tumours from patients with and without a history of smoking is likely to signify different oncogenic drivers and different mechanisms of pathogenesis between these two groups of patients.

A fifth of patients with mutations predicting response to targeted therapy were not treated with either EGFR or ALK inhibitors; however, there are likely to be justifiable clinical reasons for this. A patient may refuse treatment or have contraindications for targeted treatment, such as severe hepatic or renal impairment.[249] In addition; within NHS Scotland until recently ALK inhibitors could only be prescribed to previously

treated patients, given the advanced stage of the patients in this cohort it is likely that some may not survive to receive ALK inhibitors. The same was also true of EGFR inhibitors until 2012.

In the NHS EGFR tyrosine kinase inhibitors and ALK inhibitors can only be prescribed to patients with locally advanced or metastatic disease. Several clinical laboratories favour testing all cytology and histology samples diagnosed as non-squamous NSCLC regardless of eligibility for treatment. This reflex testing strategy would be likely to lead to increased costs and workload and the unnecessary waste of precious tissue samples. Simply switching to a request based system of only testing patients considered appropriate for treatment could deliver a 7.5% reduction in annual laboratory costs. However; in some oncology departments delaying a request for testing until after the multidisciplinary discussion of patient management may lead to unacceptable delays in the time taken to receive the molecular pathology report. This is likely to be of particular concern to those referring testing to external laboratories where the time to transfer a sample needs to be taken into account. Utilizing the well documented mutually exclusive relationship between *EGFR* and *KRAS* mutations and *ALK* rearrangements in a serial testing model could deliver a 32.5% reduction in lung cancer molecular pathology costs.[271] Where clinical laboratories have adopted next generation sequencing (NGS) as their primary technique for somatic mutation detection serial testing offers no advantages. However; many laboratories, particularly smaller centres whose lower sample numbers make the cost-effective implementation of NGS difficult, still carry out single gene testing and would benefit from this suggested change in strategy. A serial testing model also has the potential to increase turn-around times for the reporting of results and should be managed carefully to avoid a detrimental impact on patient care. Any major change to molecular pathology testing strategies should only be implemented after a full discussion between the Molecular Pathology, Histopathology and Oncology professionals involved and particular attention should be paid to total turn-around times, from identifying a need for testing to the receipt of results, to ensure that results are available in time for the patient's oncology appointment.

Lack of TTF1 expression by IHC was shown to be a good negative predictor for *EGFR* mutations, in this audit no *EGFR* mutations were found in TTF1 negative samples. Analysis of additional samples, including more recently requested samples not included in the cohort, has identified a total of 118 *EGFR* mutated samples with TTF1 status; of these 117 were positive for nuclear staining and one sample showed only cytoplasmic staining which was therefore considered negative. Interestingly this TTF1 negative sample was taken post chemo-radiation; a pre-treatment sample from the same patient, taken several years previously, was TTF1 positive.[271] Whether these samples represent a recurrence of the same disease or two separate primary tumours, with differing molecular characteristics, cannot be established. The high NPV of TTF1 IHC in this cohort is consistent with several previously published studies[264, 269, 272-276] in which the NPV ranged from 93% to 100% (table 3.14). This would mean that, under the TTF1-serial model, up to 7% of patients could be erroneously denied treatment with EGFR tyrosine kinase inhibitors; although in the majority of these publications the method of assessing TTF1 positivity was not defined. If fully validated for use as a screening tool prior to *EGFR* analysis any staining in the tumour cells, even when focal or atypical as in this case, should be considered sufficient for a specimen to proceed to molecular analysis.

Table 3.14: Summary of studies reporting both TTF1 and EGFR analyses.

Study	Total no. samples	% TTF1 positive	No. *EGFR* mutated TTF1 negative samples	NPV	Population (geographic)
Vallee *et al.* [272]	1038	79.0%	3	98.6%	France
Vincenten *et al.* [273]	797	67.9%	9	96.5%	Netherlands
Krawczyk *et al.* [274]	727	80.4%	10	93.0%	Poland
Chatziandreou *et al.* [264]	595	70.4%	2	98.9%	Greece
Sheffield *et al.* [275]	306	77.1%	4	94.3%	Canada
Somaiah *et al.* [276]	301	90%	2	93.3%	America
(pilot and validation)	131	72.5%	1	97.2%	America
Zhang *et al.*[277]	1042	78.6%	50	62.4%	China
Shanzhi *et al*.[278]	660	98.5%	1	90%	China
Chung *et al*.[279]	496	89.3%	17	67.3%	Taiwan
Sun *et al*.[134]	190	79.5%	6	84.6%	Korea

NB: NPV= Negative predictive value

The association between TTF1 and *EGFR* mutations may not be as strong in East Asian populations which have shown an NPV as low as 62.4%.[134, 277-279] It is not

known if this is due to inherent differences in the biology of the disease between these populations or if this is a factor of chance caused by the increased proportion of *EGFR* mutated patients. Any attempts to use TTF1 as a negative predictor of the presence of *EGFR* mutations should proceed with caution and would require extensive validation. However, utilising this widely available and inexpensive test to target molecular pathology analysis appropriately has the potential to save 37.5% in laboratory costs compared to the reflex model. In some cases, where a diagnostic sample is unsuitable for molecular analysis, it may be appropriate to consider the TTF1 status before attempting to perform a re-biopsy procedure which is likely to be detrimental to the health of the patient. Although TTF1 has also been suggested to be a negative predictor of ALK rearrangements in lung cancer there is, as yet, too little data to confidently use TTF1 IHC as a screen prior to *ALK* analysis.[280] Future lung cancer biomarkers should be assessed for significant correlation with TTF1 and other cost effective markers in order to develop efficient, affordable and accurate molecular pathology testing algorithms.

The low incidence of clinically actionable mutations in a Caucasian population indicates that although EGFR and ALK targeted therapies can offer great benefits to a small number of patients there are few options for the treatment of almost 90% of patients with non-squamous NSCLC. Therapies to target KRAS activated NSCLC, for example MEK inhibitors, which could benefit a large proportion of patients are not currently available for the treatment of lung cancer. In lieu of the development of further targeted therapies stratification of patients based on prognostic factors could help to optimise currently available treatment strategies. This may be of particular benefit for early stage patients post resection to identify those who would benefit from adjuvant treatment.

Chapter 4 Review of the study group dataset

The following chapter describes the comprehensive morphological characterisation of a cohort of 208 resected lung adenocarcinomas including proportion of each growth pattern, predominant growth pattern and the presence of significant levels of necrosis, mitosis, apoptosis, inflammatory infiltrate, pleomorphism, prominent nucleoli and scar tissue.

Clinicopathological variables for the 208 resected lung adenocarcinomas (including T stage, nodal status, lesion size (mm), pleural involvement, clinical stage, resection date and date and cause of death) were supplied by the NHS Lothian tissue governance unit (table 4.1). Each of these cases was reviewed and restaged to comply with the TNM 7th Edition.[5]

Table 4.1: Summary of histopathological characteristics of the lung adenocarcinoma cohort.

		Number	%
Lesion size (mm)			
	Median	32.0	
	Mean (95% CI)	36.3 (33.8 to 38.9)	
	Range	6 to 115	
T stage			
	T1a	33	15.9%
	T1b	32	15.4%
	T2a	96	46.2%
	T2b	30	14.4%
	T3	17	8.2%
N stage			
	N0	138	66.3%
	N1	70	33.7%
Pleural involvement			
	No pleural invasion	134	64.4%
	Pleural invasion	74	35.6%
Clinical stage			
	Ia	46	22.1%
	Ib	61	29.3%
	IIa	69	33.2%
	IIb	32	15.4%
Total		208	100%

4.1 Morphology

Two hundred and eight resected lung adenocarcinomas stage I and II were identified from the NHS Lothian histopathology archives, of these one had no slides available for morphological assessment. In total 907 tumour-containing haematoxylin and eosin (H&E) stained sections underwent comprehensive morphological assessment. In some cases, the growth patterns observed were easily identified and closely resembled previously published images and conformed to descriptions proposed in the literature (figure 4.1).[11, 261]

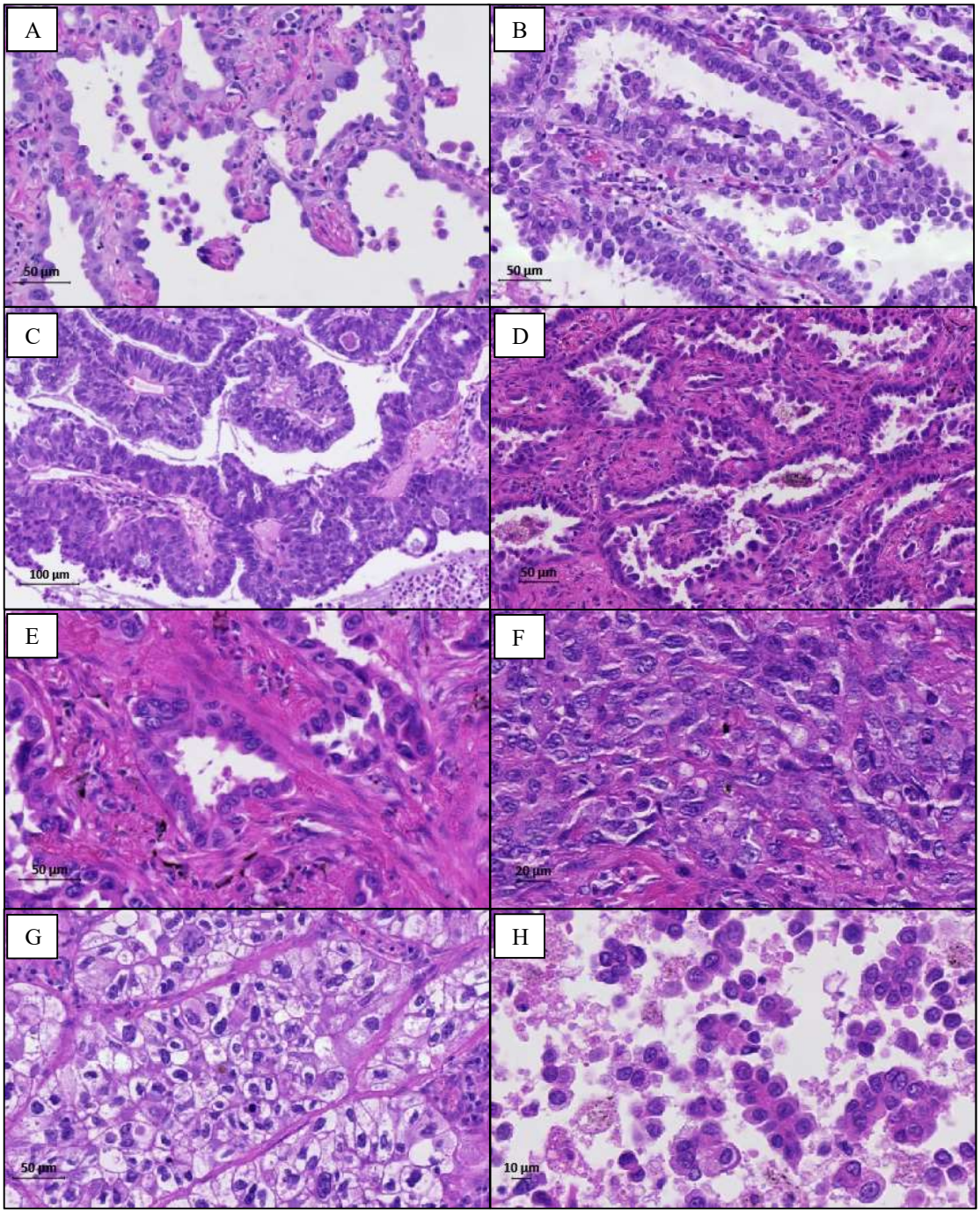


Figure 4.1: Photographs of H&E stained sections with examples of tumour growth patterns showing classical morphology: lepidic (A and B), papillary (C), acinar (D and E), solid (F and G) and micropapillary (H).

However, in a large number of cases the growth patterns present in each case did not form classical structures and were more difficult to distinguish. Micropapillary structures were frequently found apparently growing from the surface of papillary,

lepidic or acinar areas of tumour. Micropapillary tumour clusters were also found 'floating' within the lumen formed at the centre of acinar structures. In some tumours, areas of necrosis or poor preservation of tissue architecture made the identification of growth patterns more challenging. In addition, some tumours produced patterns not classified by Travis *et al*.; for example, small clusters of tumour cells within fibrotic tissue (figure 4.2).[261]

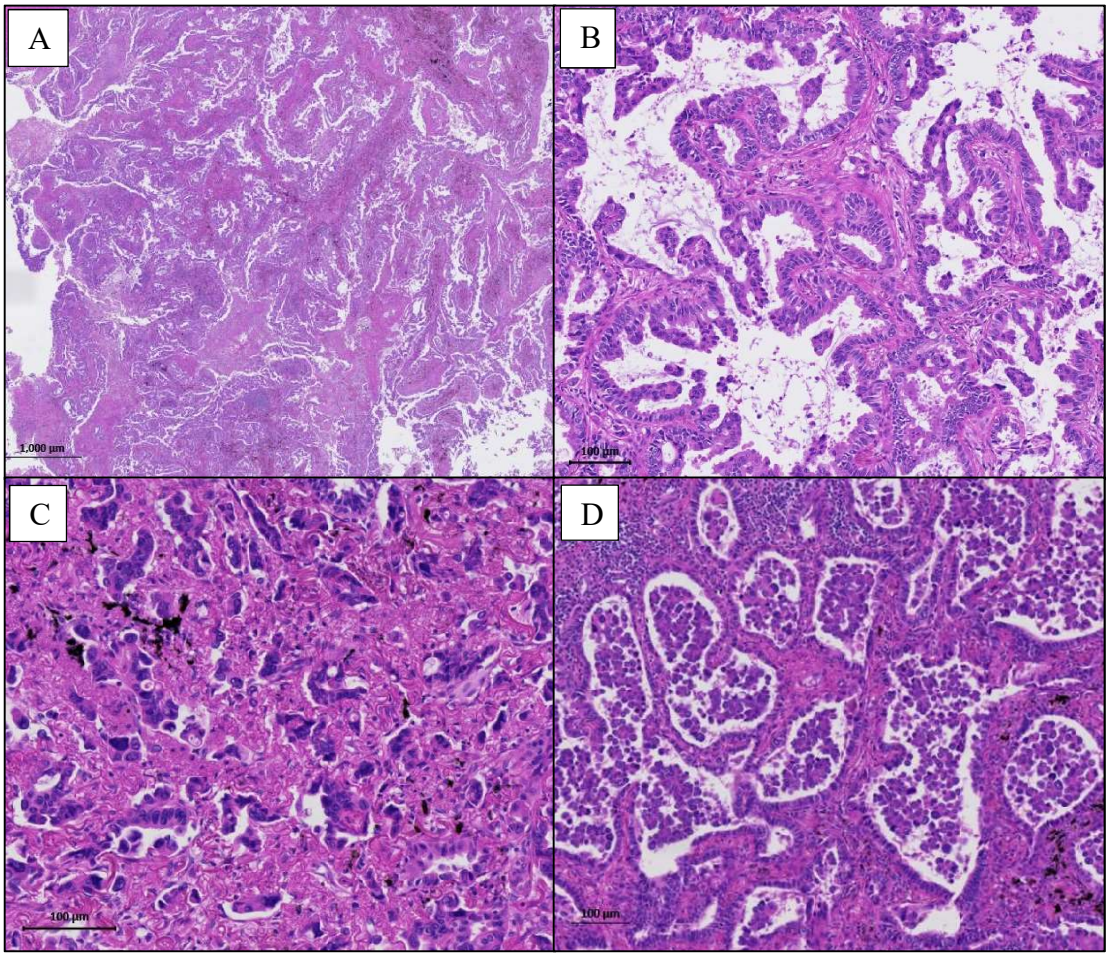


Figure 4.2: Photographs of H&E stained sections. In some cases growth patterns were more difficult to define due to poor tissue preservation (A), the presence of mixed patterns (for example micropapillary structures associated with papillary or acinar growth) (B and D) or not conforming to classical definitions (C).

Many cases showed more than one growth pattern, very few of which had easily defined boundaries between neighbouring patterns; more usually tumour growth showed a continuum of morphological change from one growth pattern to the next. This was most apparent between lepidic and papillary, lepidic and acinar, and acinar and papillary patterns. In these tumours identifying the exact point at which the growth

pattern changed was more challenging making any estimate of the proportion of each pattern likely to be inaccurate (figure 4.3).

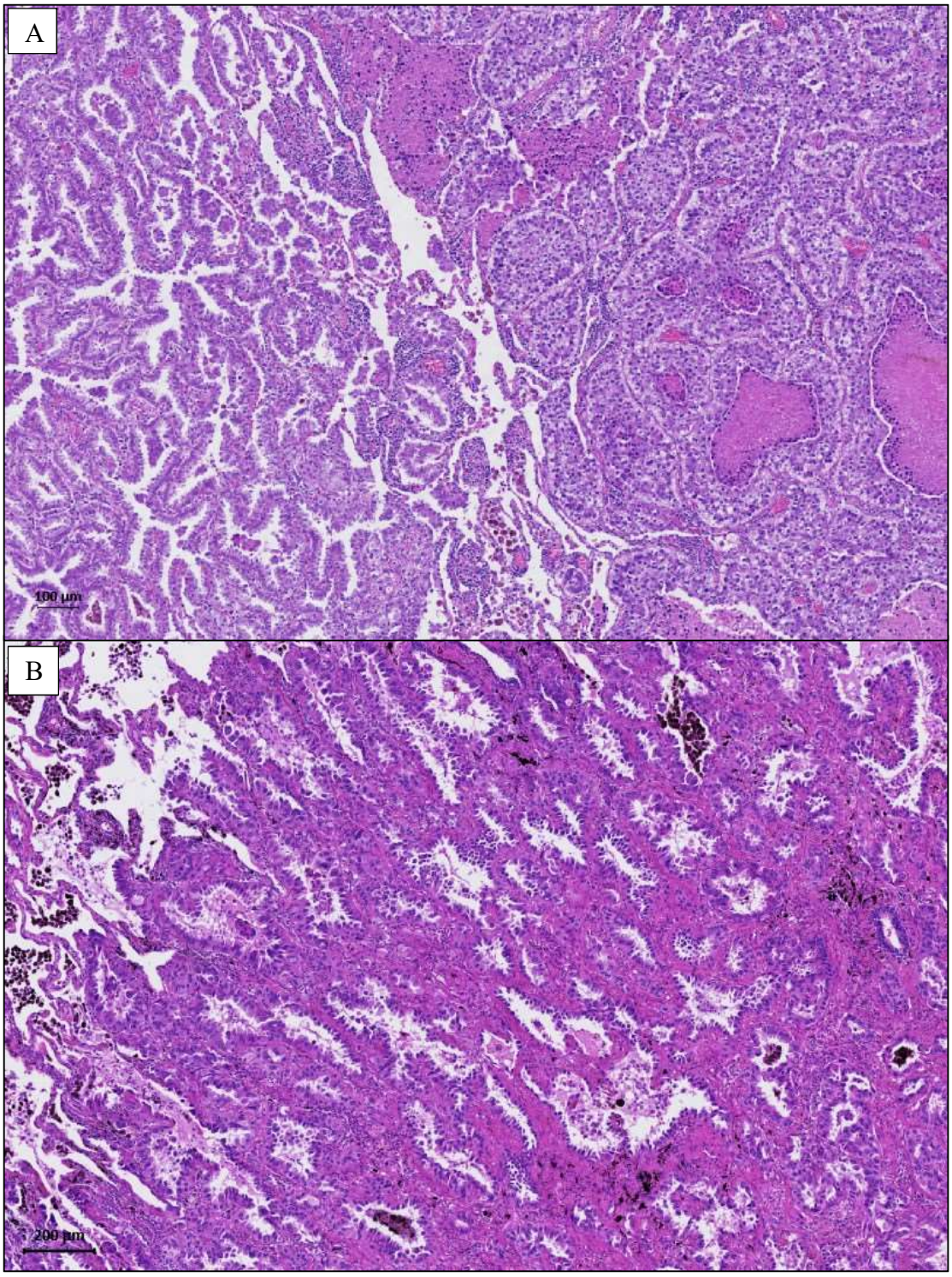


Figure 4.3: Photographs of H&E stained sections from 2 tumours; 1 showing a very marked boundary between 2 growth patterns (A), the other with a more gradual change of tumour growth from lepidic to acinar (B).

In some cases, a single section could show elements of all 5 growth patterns further increasing the difficulty in accurately estimating the contribution of each pattern (figure 4.4).

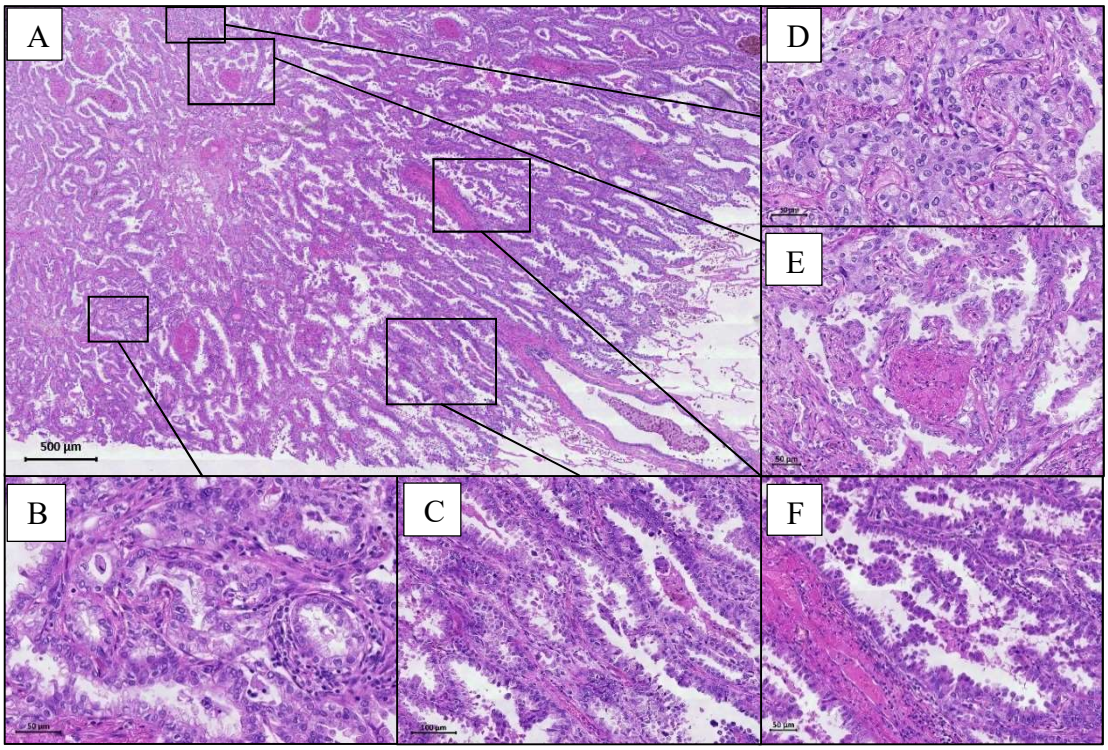


Figure 4.4: Photographs of a single H&E stained section taken at low (A) and high power (B to F). Although at low power this tumour appears largely lepidic there are elements of acinar (B), lepidic (C), solid (D), papillary (E) and micropapillary (F) growth patterns.

In some areas dissociation of clumps of tumour cells from lepidic, acinar, papillary and solid structures, likely to be caused by sub-optimal tissue fixation, could be mistaken for the micropapillary pattern (figure 4.5).

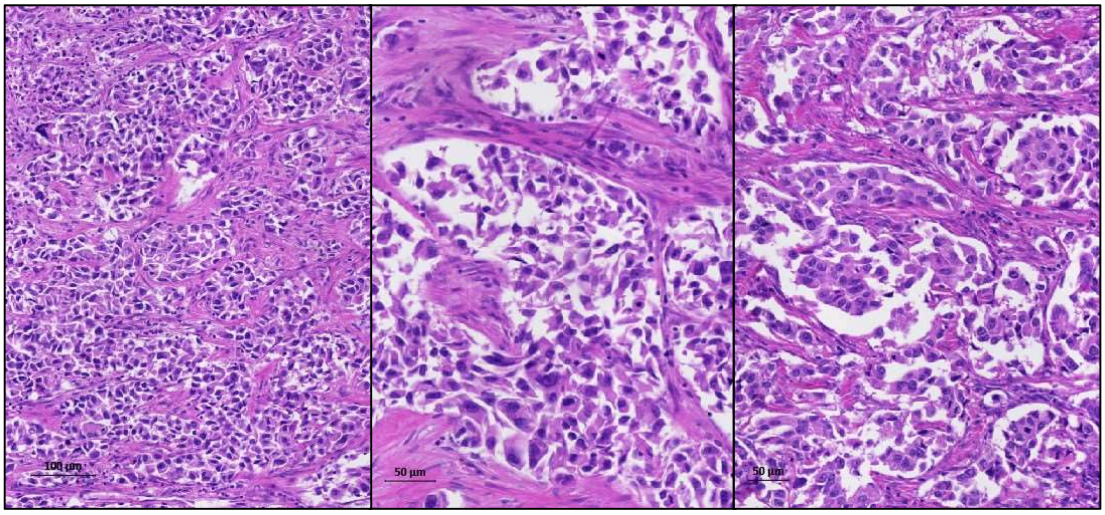


Figure 4.5: Photographs of a single H&E stained section containing predominantly solid growth however poor tissue fixation could have been the cause of micropapillary-like structures.

In the assessment of 207 stage I and II lung adenocarcinomas the acinar growth pattern was the most frequently observed, being present in 89.9% of cases; the papillary pattern was the least commonly seen (55.6% of cases). On average the acinar growth pattern, when present, was found in the highest proportions followed by the solid pattern; the proportions of lepidic, papillary and micropapillary patterns were very similar (Friedman's 2-way analysis of variance by ranks $p < 0.005$) (table 4.2).

Table 4.2: The incidence and proportion of the lepidic, papillary, acinar, solid and micropapillary growth patterns observed at any proportion and at ≥ 5%.

Growth Pattern	Present (at any proportion)	Present (≥5% of tumour)	Median proportion	Mean proportion (95% Confidence Interval)
Lepidic	125 (60.6%)	99 (47.8%)	3.0%	13.4% (10.6% to 16.1%)
Papillary	115 (55.6%)	95 (45.9%)	2.0%	12.7% (10.2% to 15.3%)
Acinar	186 (89.9%)	175 (84.5%)	27.0%	33.5% (29.8% to 37.2%)
Solid	136 (65.7%)	116 (56.0%)	12.0%	27.6% (22.3% to 32.2%)
Micropapillary	132 (63.8%)	106 (51.2%)	6.0%	12.9% (10.5% to 15.4%)

Only growth patterns present at greater than or equal to 5% of the area of each tissue section were included in initial analyses in order to eliminate false positives caused by

cutting artefact. The majority of cases showed multiple growth patterns and only 15 (7.2%) showed only one pattern; whereas, 43 (20.8%) cases showed 2 patterns, 60 (28.8%) showed 3, 59 (28.5%) showed 4 and 30 (14.5%) showed all 5 growth patterns. In the cases where only 1 growth pattern was present the majority (86.7%) had solid growth. The presence of the solid pattern was negatively correlated with the presence of the lepidic ($p< 0.0005$), papillary ($p< 0.0005$) and acinar patterns ($p< 0.0005$). Whereas the micropapillary pattern was positively associated with the papillary ($p< 0.0005$), acinar ($p= 0.004$) and solid patterns ($p= 0.001$). In addition, the lepidic pattern was more commonly found with the papillary pattern ($p= 0.017$) and the presence of the papillary pattern was associated with the presence of the acinar pattern ($p= 0.01$) (tables 4.3 and 4.4).

Table 4.3: The co-occurrence of growth patterns at greater than 5% of each tumour. Significance was calculated using the χ^2 test.

		<5% Lepidic	≥5% Lepidic	Total	Significance
Papillary					p= 0.017
	<5% papillary	67 (62.0%)	45 (45.5%)	112	
	≥5% papillary	41 (38.0%)	54 (54.5%)	95	
Solid					p< 0. 0005
	<5% solid	29 (26.9%)	62 (62.6%)	91	
	≥5% solid	79 (73.1%)	37 (37.4%)	116	
Total		108 (100%)	99 (100%)	207	
		<5% Papillary	≥5% Papillary	Total	Significance
Micropapillary					p< 0.0005
	<5% micropapillary	57 (50.9%)	20 (21.1%)	77	
	≥5% micropapillary	55 (49.1%)	75 (78.9%)	130	
Acinar					p= 0.01
	<5% acinar	24 (21.4%)	8 (8.4%)	32	
	≥5% acinar	88 (78.6%)	87 (91.6%)	175	
Solid					p< 0.0005
	<5% solid	35 (31.3%)	56 (58.9%)	91	
	≥5% solid	77 (68.8%)	39 (41.1%)	116	
Total		112 (100%)	95 (100%)	207	
		<5% acinar	≥5% acinar	Total	Significance
Micropapillary					p= 0.004
	<5% Micropapillary	23 (71.9%)	78 (44.6%)	101	
	≥5% Micropapillary	9 (28.1%)	97 (55.4%)	106	
Solid					p< 0.0005
	<5% solid	5 (15.6%)	86 (49.1%)	91	
	≥5% solid	27 (84.4%)	89 (50.9%)	116	
Total		32 (100%)	175 (100%)	207	
		<5% solid	≥5% solid	Total	Significance
Micropapillary					p= 0.001
	<5% Micropapillary	33 (36.3%)	68 (58.6%)	91	
	≥5% Micropapillary	58 (63.7%)	48 (41.4%)	116	
Total		91 (100%)	116 (100%)	207	

Table 4.4: Summary of the statistically significant associations between the presence of each growth pattern at greater than or equal to 5% of each tumour. Significance was calculated using the χ^2 test.

		Present at ≥5% of the tumour			
		Papillary	Acinar	Solid	Micropapillary
Present at ≥5% of the tumour	Lepidic	+ correlation p=0.017	n. sig.	- correlation p< 0.0005	n. sig.
	Papillary	-	+ correlation p= 0.01	- correlation p< 0.0005	+ correlation p< 0.0005
	Acinar	-	-	- correlation p< 0.0005	+ correlation p= 0.004
	Solid	-	-	-	+ correlation p= 0.001

NB. += positive, -= negative, n. sig.= not significant

In order to assess the clinical significance of further easily available and cost effective histopathological data additional morphological features, derived from descriptive data collected at the time of morphological assessment, were assessed. These additional tumour grading variables included the presence of:

- a large number of mitotic figures (figure 4.6).
- a large number of apoptotic bodies (figure 4.6).
- a high inflammatory infiltrate (figure 4.7).
- a large amount of necrosis (figure 4.8).
- large numbers of prominent nucleoli (figure 4.7).
- a high degree of nuclear pleomorphism (figure 4.7).
- scar tissue (figure 4.8).
- dyscohesive tumour cells (figure 4.7).

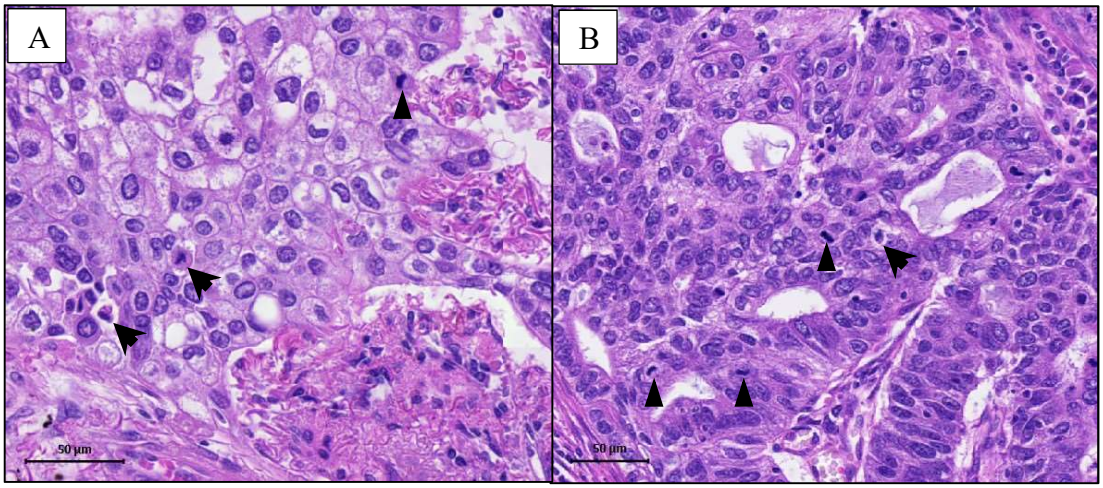


Figure 4.6: Photographs of H&E stained sections from two tumours with relatively large numbers of apoptotic bodies (arrows) and mitotic figures (arrowheads).

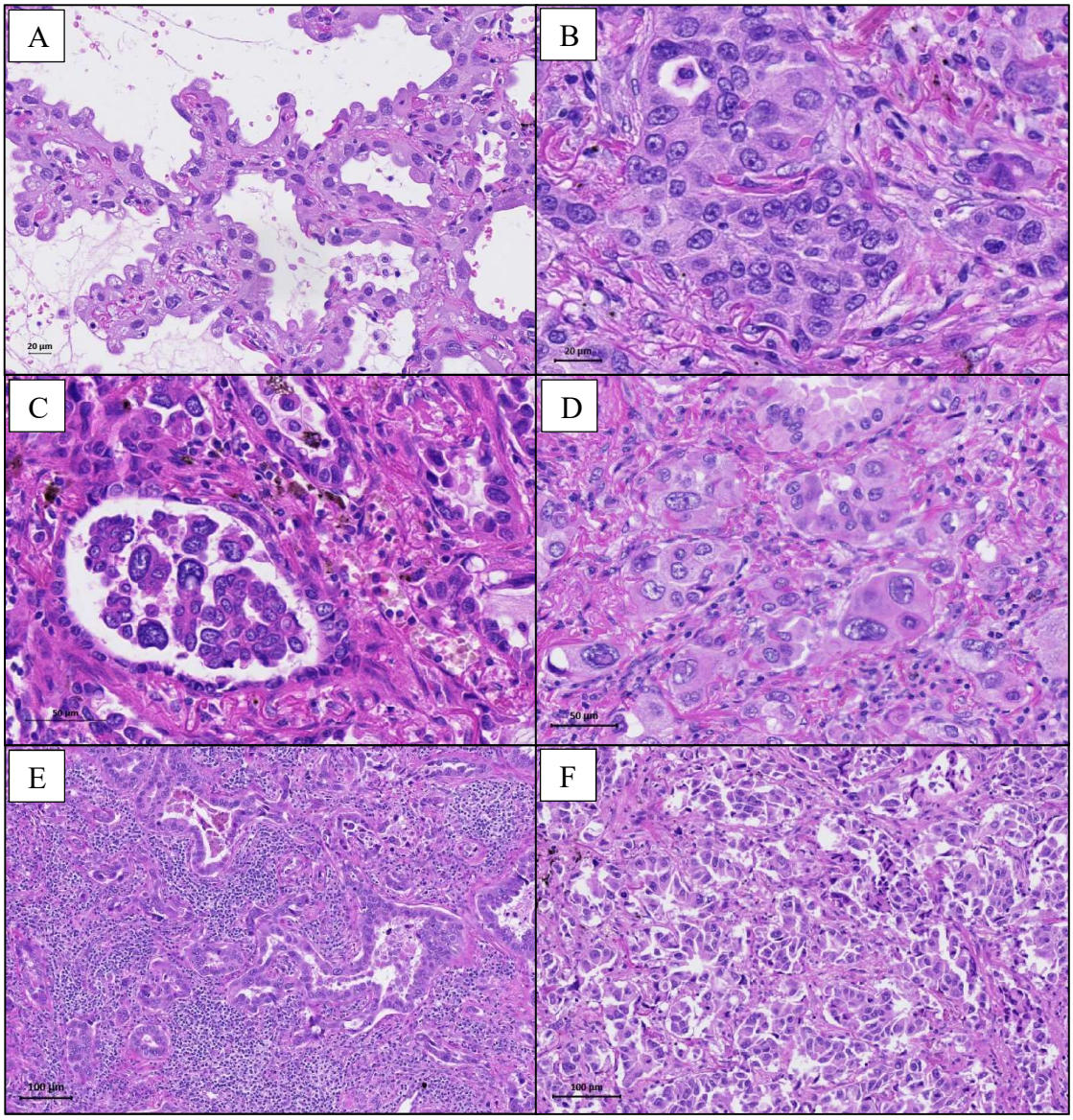


Figure 4.7: Photographs of H&E stained examples sections from tumours showing inconspicuous nucleoli (A), prominent nucleoli (B), pleomorphic nuclei (C and D), a high inflammatory infiltrate (E) and dyscohesive between tumour cells (F).

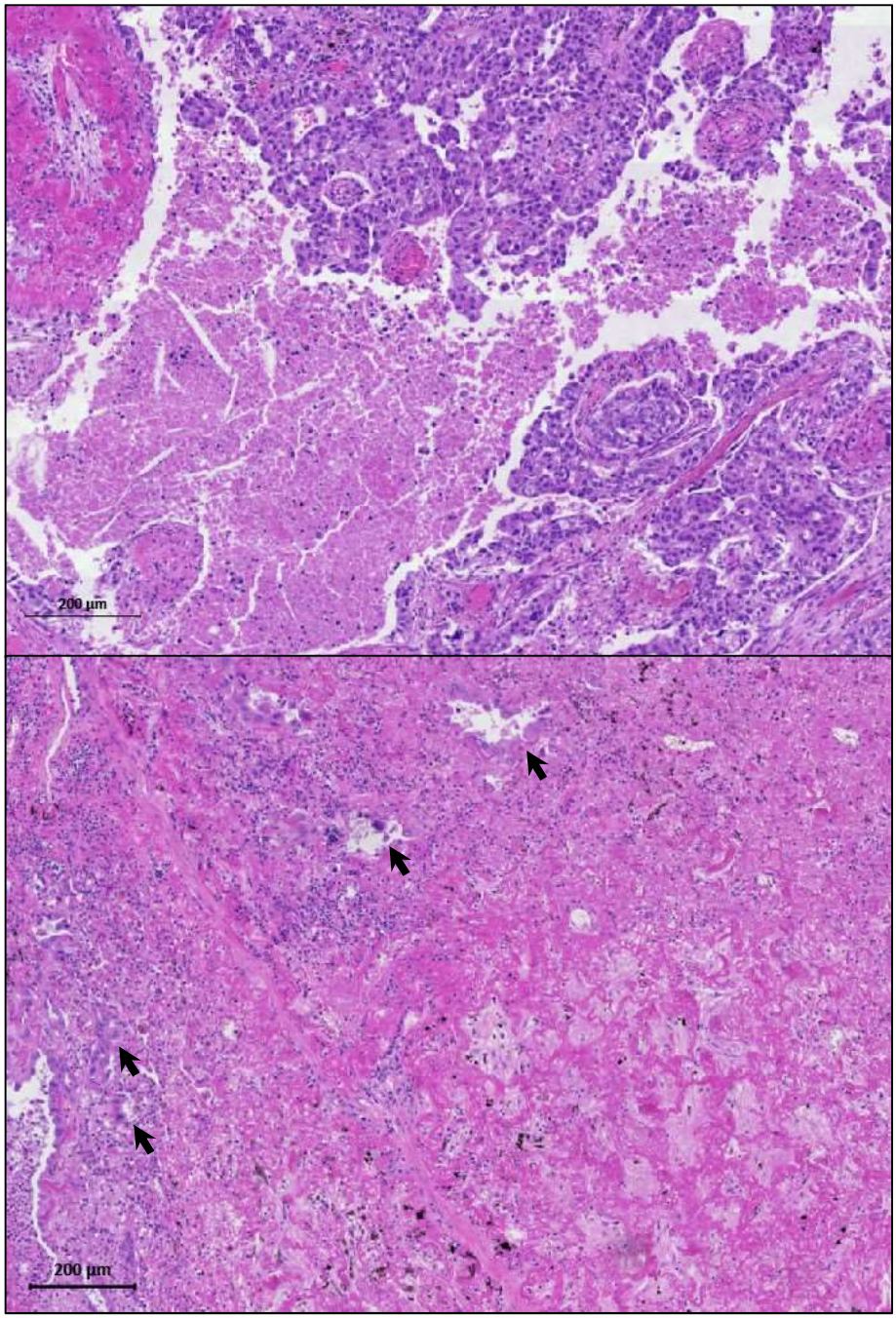


Figure 4.8: Photographs of H&E stained sections from 2 tumours, 1 showing large amounts of necrosis (A) and 1 with scar tissue (B) associated with small tumour nests (arrows).

Of the 207 cases available for morphological assessment, 15 had staining quality too poor to assess tumour grading features leaving a total of 192 cases; due to morphological characteristics one further case could not be assessed for the presence of mitotic figures. Cytological pleomorphism and prominent nucleoli were commonly observed (48.4% and 46.9% of cases respectively) whereas dyscohesion between tumour cells was relatively rare (8.3%) (table 4.5).

Table 4.5: Summary of the results of the assessment of qualitative morphological characteristics.

Morphological characteristic	Not present	Present	Total
Large areas on necrosis	134	58 (30.2%)	192
High degree of inflammatory infiltrate	137	55 (28.6%)	192
Large number of mitotic figures	85	106 (55.5%)	191
Large number of apoptotic bodies	157	35 (18.2%)	192
Numerous prominent nucleoli	102	90 (46.9%)	192
Significant cytological pleomorphism	99	93 (48.4%)	192
Presence of scar tissue	156	36 (18.8%)	192
Areas with dyscohesive tumour cells	176	16 (8.3%)	192

Statistically significant correlations identified between tumour grading variables are detailed in table 4.6. Tumours with large areas of necrosis were more likely to have many mitotic figures and apoptotic bodies and pleomorphic nuclei ($p= 0.004$, $p= 0.009$ and $p= 0.03$ respectively). In addition, tumours with large numbers of mitotic figures were associated with the presence of large numbers of apoptotic bodies and prominent nucleoli ($p< 0.0005$ and $p= 0.026$ respectively)

Table 4.5: Summary of statistically significant associations between tumour grading characteristics. Significance was calculated using the χ^2 test.

		Little or no necrosis	Necrosis	Total	Significance
Mitosis					p= 0.004
	Few mitotic figures	69 (51.5%)	16 (28.1%)	85	
	Many mitotic figures	65 (48.5%)	41 (71.9%)	106	
Total		134 (100%)	57 (100%)	191	
Apoptosis					p= 0.009
	Few apoptotic bodies	116 (86.6%)	41 (70.7%)	157	
	Many apoptotic bodies	18 (13.4%)	17 (29.3%)	35	
Total		134 (100%)	58 (100%)	192	
Pleomorphism					p= 0.03
	Not pleomorphic	76 (56.7%)	23 (39.7%)	99	
	Pleomorphic	58 (43.3%)	35 (60.3%)	93	
Total		134 (100%)	58 (100%)	192	
		Few mitotic figures	Many mitotic figures		
Apoptosis					p< 0.0005
	Few apoptotic bodies	79 (92.9%)	77 (72.6%)	156	
	Many apoptotic bodies	6 (7.1%)	29 (27.4%)	35	
Total		85 (100%)	106 (100%)	191	
Nucleoli					p= 0.026
	Not prominent	53 (62.4%)	49 (46.2%)	102	
	Prominent	32 (37.6%)	57 (53.8%)	189	
Total		85 (100%)	106 (100%)	191	

4.1.1 Presence of each growth pattern

Lepidic

Tumours with greater than or equal to 5% lepidic pattern were associated with a lack of large amounts of necrosis (table 4.7), which was further confirmed by the lower proportions of lepidic pattern in necrotic tumours (p< 0.0005) (figure 4.9). Lepidic positive tumours were also unlikely to have many apoptotic bodies and mitotic figures (p= 0.031 and p= 0.003 respectively). The association with mitotic figures was supported by analysis of the average proportion of lepidic pattern in each case where cases with few mitotic figures had a higher proportion of the lepidic pattern (p= 0.001) (figure 4.9).

Table 4.7: Summary of the statistically significant associations between morphological variables and the presence of greater than or equal to 5% lepidic growth. Significance was calculated using the χ^2 test.

		<5% Lepidic	≥5% Lepidic	Total	Significance
Necrosis					p< 0.0005
	Little or none	55 (55%)	79 (85.9%)	134	
	Large areas	45 (45%)	13 (14.1%)	58	
Total		100 (100%)	92 (100%)	192	
Mitosis					p= 0.003
	Few mitotic figures	34 (34.3%)	51 (55.4%)	85	
	Many mitotic figures	65 (65.7%)	41 (44.6%)	106	
Total		99 (100%)	92 100%)	191	
Apoptosis					p= 0.031
	Few apoptotic bodies	76 (76%)	81 (88.0%)	157	
	Many apoptotic bodies	24 (24%)	11 (12.0%)	35	
Total		100 (100%)	92 (100%)	192	

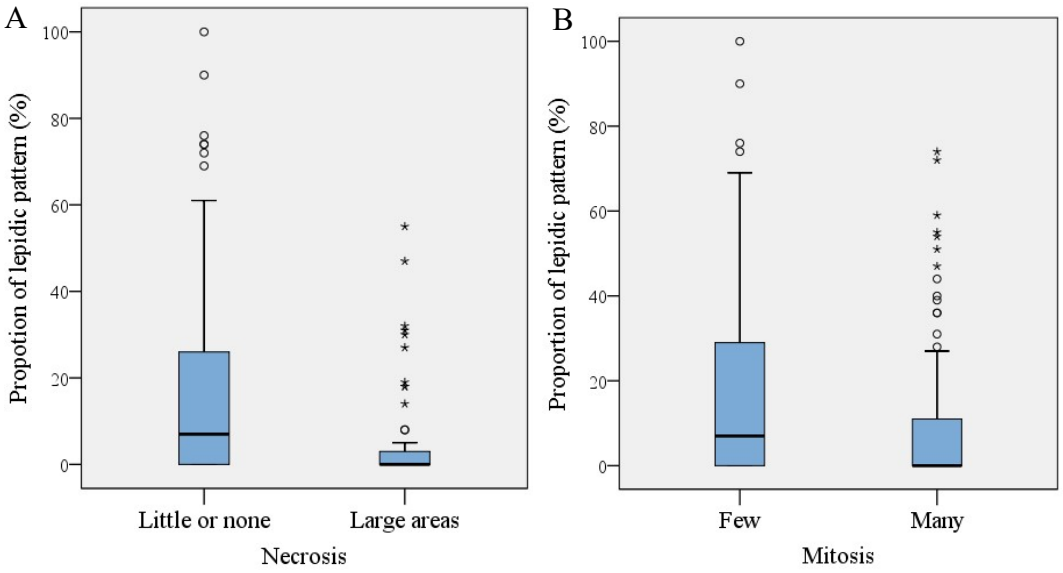


Figure 4.9: Boxplots showing statistically significant associations between the proportion of lepidic growth and the presence of large areas of necrosis (A) (p< 0.0005) and the presence of frequent mitotic figures (B) (Mann-Whitney U test p= 0.001).

Papillary

The presence, and proportion, of the papillary pattern was associated with the lack of large areas of necrosis (p= 0.007 and p= 0.01 respectively) (table 4.8 and figure 4.10). In addition, papillary positive tumours were more likely to have prominent nucleoli (p= 0.050).

Table 4.8: Summary of the statistically significant associations between tumours with greater than or equal to 5% papillary growth and morphological variables. Significance was calculated using the χ^2 test.

		<5% Papillary	≥5% Papillary	Total	Significance
Necrosis					p= 0.007
	No necrosis	64 (61.5%)	70 (79.5%)	134	
	Necrosis	40 (38.5%)	18 (20.5%)	58	
Total		104 (100%)	88 (100%)	192	
Nucleoli					p= 0.050
	Inconspicuous	62 (59.6%)	40 (45.5%)	102	
	Prominent	42 (40.4%)	48 (54.5%)	90	
Total		104 (100%)	88 (100%)	192	

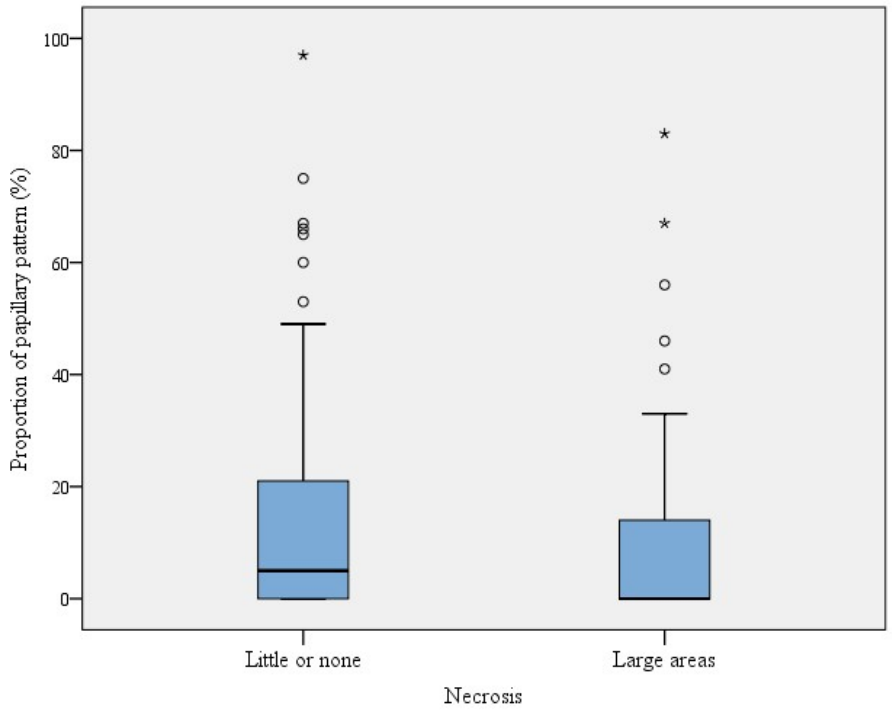


Figure 4.10: Box plot showing the proportion of papillary growth pattern in tumours with and without large areas of necrosis (Mann-Whitney U test p= 0.01).

Acinar

The presence, and proportion, of the acinar pattern was associated with the lack of large areas of necrosis (p< 0.0005 and p= 0.004 respectively) (table 4.9 and figure 4.11). Apoptotic bodies were less common in acinar positive tumours (p= 0.01); this association is further supported by the fact that tumours with few apoptotic bodies had a lower proportion of acinar growth pattern (p< 0.0005) (figure 4.11).

Table 4.9: *Summary of the statistically significant associations between the presence or absence of acinar growth and morphological variables. Significance was calculated using the χ^2 test.*

		<5% Acinar	≥5% Acinar	Total	Significance
Necrosis					p< 0.0005
	No necrosis	13 (40.6%)	121 (75.6%)	134	
	Necrosis	19 (59.4%)	39 (24.4%)	58	
Total		32 (100%)	160 (100%)	192	
Apoptosis					p= 0.01
	Few apoptotic bodies	21 (65.6%)	136 (85%)	157	
	Many apoptotic bodies	11 (34.4%)	24 (15%)	35	
Total		32 (100%)	160 (100%)	192	

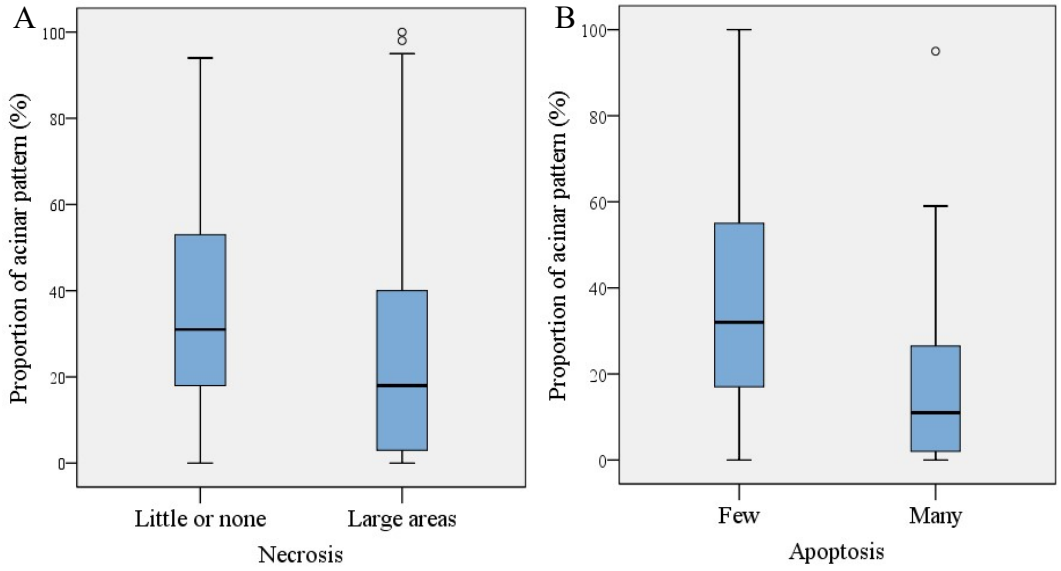


Figure 4.11: Boxplots showing statistically significant associations between the proportion of acinar growth pattern in tumours with and without large areas of necrosis (A) (p= 0.004) and large numbers of apoptotic bodies (B) (Mann-Whitney U test p< 0.0005).

Solid

Solid growth was the only pattern significantly associated with the presence of necrosis (p< 0.0005) (table 4.10 and figure 4.12). Tumours with greater than or equal to 5% solid growth pattern had more mitotic figures and apoptotic bodies (p= 0.001 and p= 0.003 respectively) which was also confirmed by the increased proportion of solid pattern in tumours with many mitotic figures or apoptotic bodies (p< 0.0005 for both

variables) (figure 4.13). In addition, these tumours were less likely to have large areas of scar tissue (p= 0.013).

Table 4.10: *Summary of the statistically significant associations between morphological variables and tumours with greater than or equal to 5% solid growth. Significance was calculated using the χ^2 test.*

		<5% Solid	≥5% Solid	Total	Significance
Necrosis					p< 0.0005
	No necrosis	70 (85.4%)	64 (58.2%)	134	
	Necrosis	12 (14.6%)	46 (41.8%)	58	
Total		82 (100%)	110 (100%)	192	
Mitosis					p= 0.001
	Few mitotic figures	47 (58.0%)	38 (34.5%)	85	
	Many mitotic figures	34 (42.0%)	72 (65.5%)	106	
Total		81 (100%)	110 (100%)	191	
Apoptosis					p= 0.003
	Few apoptotic bodies	75 (91.5%)	82 (74.5%)	157	
	Many apoptotic bodies	7 (8.5%)	28 (25.5%)	35	
Total		82 (100%)	110 (100%)	192	
Scar tissue					p= 0.013
	Not present	60 (73.2%)	96 (87.3%)	156	
	Present	22 (26.8%)	14 (12.7%)	36	
Total		82 (100%)	110 (100%)	192	

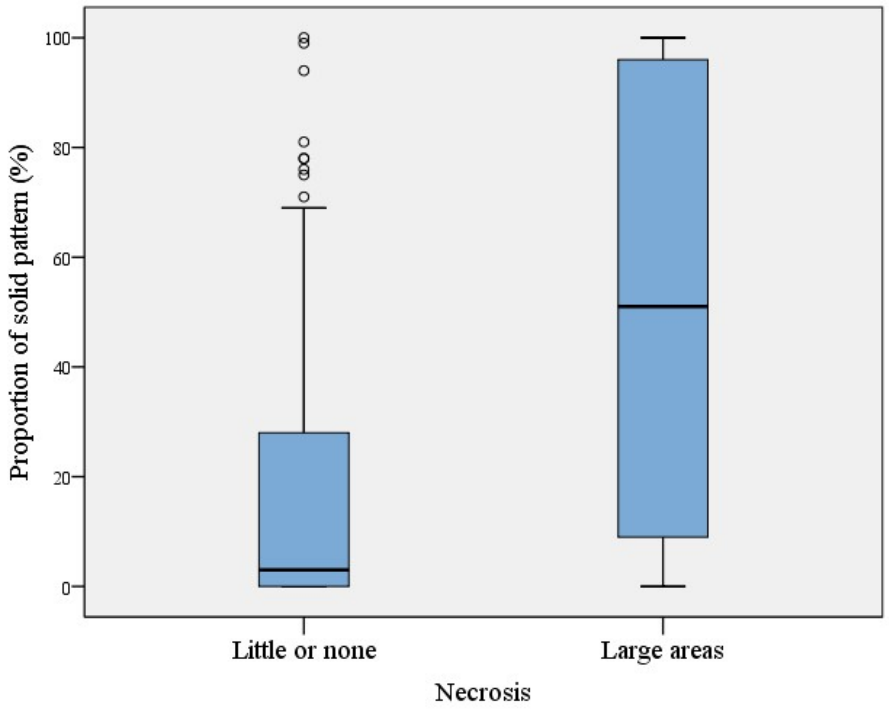


Figure 4.12: Box plot showing the association between the proportion of solid growth and the presence of large areas of necrosis (Mann-Whitey U test p> 0.0005).

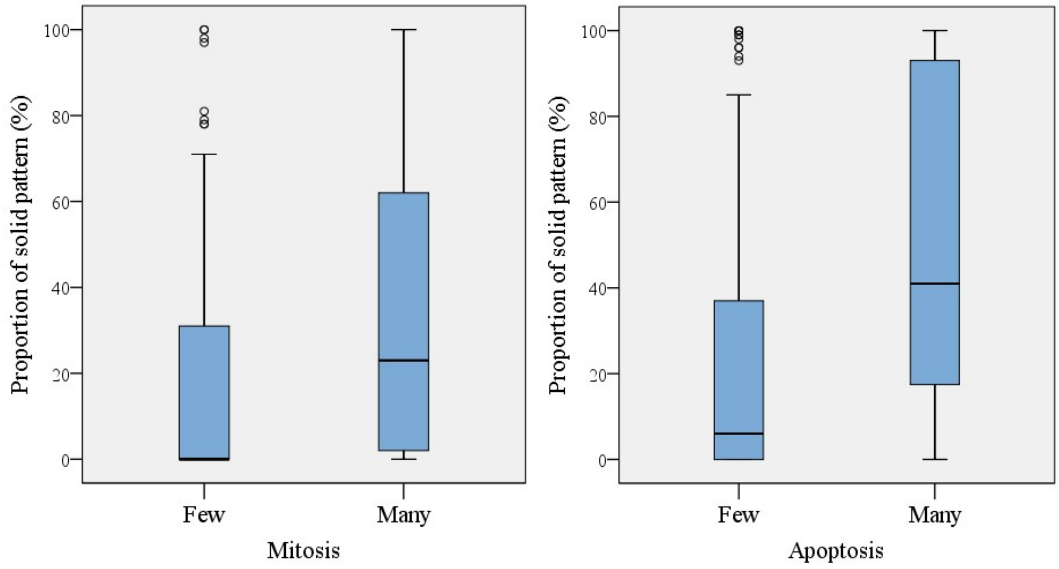


Figure 4.13: Box plot showing the statistically significant associations between the proportion of solid growth and the number of many mitotic figures (A) (p< 0.0005) and apoptotic bodies (B) (p< 0.0005) (Mann-Whitney U test).

Micropapillary

The presence, and proportion of the micropapillary growth pattern was significantly associated with the lack of large areas of necrosis (p= 0.002 and p< 0.0005

respectively) (table 4.11 and figure 4.14). Tumours with prominent nucleoli were more common in micropapillary positive tumours (p= 0.013) and were likely to have higher proportions of the micropapillary growth pattern (p= 0.008) (figure 4.14).

Table 4.11: Summary of the statistically significant associations between tumours with greater than or equal to 5% micropapillary growth and morphological variables. Significance was calculated using the χ^2 test.

		<5% Micropapillary	≥5% Micropapillary	Total	Significance
Necrosis					p= 0.002
	Not present	55 (59.1%)	79 (79.8%)	134	
	Present	38 (40.9%)	20 (20.2%)	58	
Total		93 (100%)	99 (100%)	192	
Nucleoli					p= 0.013
	Inconspicuous	58 (62.4%)	44 (44.4%)	102	
	Prominent	35 (37.6%)	55 (55.6%)	90	
Total		93 (100%)	99 (100%)	192	

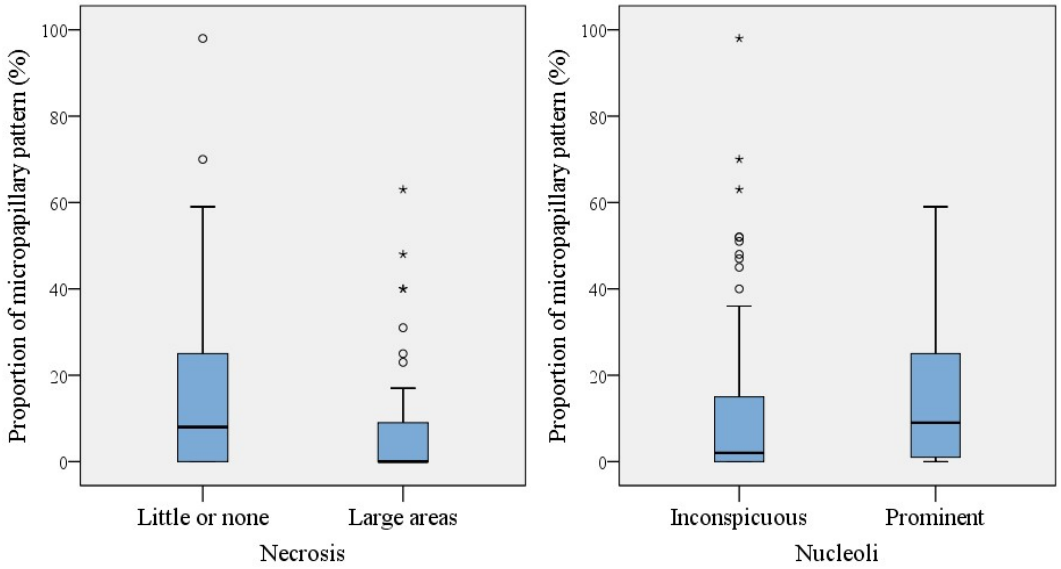


Figure 4.14: Boxplots showing statistically significant associations between the proportion of micropapillary growth pattern and the presence of large areas of necrosis (A) (p< 0.0005) and prominent nucleoli (B) (p= 0.008) (Mann-Whitney U test).

4.1.2 Predominant growth pattern

From the estimated proportions of non-tumour, lepidic, acinar, papillary, solid and micropapillary growth patterns (excluding lymph nodes); 4 different methods were used to determine the predominant growth pattern as described in section 2.3. Method

A (whereby the proportion of individual growth patterns were expressed as a proportion of the whole section, including non-tumour tissue, and the predominant pattern was that with the largest total proportion across all sections of the primary tumour) showed the greatest inter-observer agreement and high concordance with other methods indicating it to be the most robust.

The number and proportion of cases with each predominant growth pattern, calculated by the 4 methods, are detailed in table 4.12. With 3 of the methods (A, B or C) the acinar growth pattern was the most common predominant pattern followed by solid, papillary and micropapillary; predominantly lepidic lesions were rarer. With method D the solid growth pattern was most frequently the predominant pattern followed by acinar, papillary, micropapillary then lepidic.

Table 4.12: The number (and percentage) of cases with predominant lepidic, papillary, acinar, solid and micropapillary growth patterns as calculated using 4 different methods.

Predominant Pattern	Method A	Method B	Method C	Method D
Lepidic	20 (9.7%)	25 (12.1%)	25 (10.1%)	18 (8.78%)
Papillary	26 (12.6%)	27 (13.0%)	27 (12.1%)	31 (15.12%)
Acinar	76 (37.8%)	69 (33.3%)	69 (34.8%)	64 (31.22%)
Solid	60 (30.0%)	60 (29.0%)	60 (31.4%)	66 (32.2%)
Micropapillary	25 (12.1%)	26 (12.6%)	26 (11.6%)	26 (12.68%)
Total	207 (100%)	207 (100%)	207 (100%)	205 (100%)

Cohen's kappa showed that there was strong agreement between the different methods of calculating predominant growth pattern. However; method D, which assessed only a single section of the case, showed less agreement with the other methods (Table 4.13).

Table 4.13: Summary of the agreement between predominant pattern estimated by four different calculation methods calculated using Cohen's kappa (κ).

	Calculation B	Calculation C	Calculation D
Calculation A	κ 0.916 p< 0.0005	κ 0.896 p< 0.0005	κ 0.707 p< 0.0005
Calculation B	-	κ 0.891 p< 0.0005	κ 0.677 p< 0.0005
Calculation C	-	-	κ 0.707 p< 0.0005

Spearman's Rank correlation was used to assess the inter-observer correlation between the proportion of each growth pattern in 52 sections from 10 cases. There was strongest agreement in the proportion of solid (Rho= 0.943) and lepidic (Rho= 0.906) growth patterns, whereas the proportion of papillary pattern showed greater variation between observers (Rho= 0.687) (table 4.14).

Table 4.14: Inter-observer agreement in the proportion of each growth pattern per section assessed using Spearman's Rank correlation.

	% Lepidic	% Papillary	% Acinar	% Solid	% Micropapillary
Rho statistic	0.906	0.687	0.752	0.943	0.822
Significance	p< 0.0005	p= 0.014	p < 0.0005	p< 0.0005	p< 0.0005

Analysis of inter-observer agreement of predominant pattern using method A showed there was disagreement in the predominant pattern in 1 case, between the lepidic and acinar patterns. Using method B and C there was disagreement in 2 cases, one between acinar and micropapillary and one between acinar and lepidic. Using method D there was disagreement in 1 case, between lepidic and papillary. Methods A and D had the highest inter-observer agreement with kappa values of 0.870 (p< 0.0005) (table 4.15).

Table 4.15: Inter-observer agreement for the predominant growth pattern in each case using the 4 calculation methods, assessed using Cohen's kappa.

	Method A	Method B	Method C	Method D
kappa value	0.870	0.752	0.752	0.870
Significance	p< 0.0005	p < 0.0005	p < 0.0005	p< 0.0005

Due to the high inter-observer agreement and strong agreement with other methods method A was selected for estimating predominant growth pattern.

Predominant pattern, calculated using method A, was analysed for associations with tumour grading variables (table 4.16). The presence of necrosis was significantly associated with predominant growth pattern (p<0.0005) and was most common in predominantly solid tumours (55.4%), followed by acinar (23.9%) and papillary (22.7%); but was rare in predominantly lepidic tumours (5.6%). The presence of large numbers of apoptotic bodies was also associated with predominant growth pattern where apoptosis was most common in solid tumours (p= 0.022). The predominant

pattern was also associated with the presence of scar tissue and was most common in predominantly lepidic tumours lepidic, there was no scar tissue noted in any predominantly papillary tumours (p= 0.032).

Table 4.16: Statistically significant correlations between predominant growth pattern (method A) and tumour grading variables. Significance was calculated using the χ^2 test unless otherwise stated.

		Predominant pattern					Total	Significance
		Lepidic	Papillary	Acinar	Solid	Micropapillary		
Necrosis								p<0.0005
	Little or none	17 (94.4%)	17 (77.3%)	54 (76.1%)	25 (44.6%)	21 (84%)	134	
	Large areas	1 (5.6%)	5 (22.7%)	17 (23.9%)	31 (55.4%)	4 (16%)	58	
Apoptosis								p= 0.022[a]
	Few	17 (94.4%)	18 (81.1%)	64 (90.1%)	39 (69.6%)	19 (76%)	157	
	Many	1 (5.6%)	4 (18.2%)	7 (9.9%)	17 (30.4%)	6 (24%)	35	
Scar tissue								p= 0.032
	None	11 (61.1%)	22 (100%)	56 (78.9%)	47 (83.9%)	20 (80.0%)	156	
	Large areas	7 (38.9%)	0 (0%)	15 (21.1%)	9 (16.1%)	5 (20.0%)	36	
Total		18 (100%)	22 (100%)	71 (100%)	56 (100%)	25 (100%)	192	18 (100%)

[a]*Fisher's exact test*

4.1.3 Predominant growth pattern in primary tumours vs metastases

Repeated measures ANOVA was used to explore the difference between growth patterns in primary lesions and their associated nodal metastases. Only the solid growth pattern showed a significant increase in proportion in lymph nodes compared to their associated primary tumours across the whole cohort (F=4.548, p= 0.038) (figure 6). Acinar, papillary and micropapillary showed no significant trends (p= 0.657, 0.777 and 0.816 respectively) (figure 4.15).

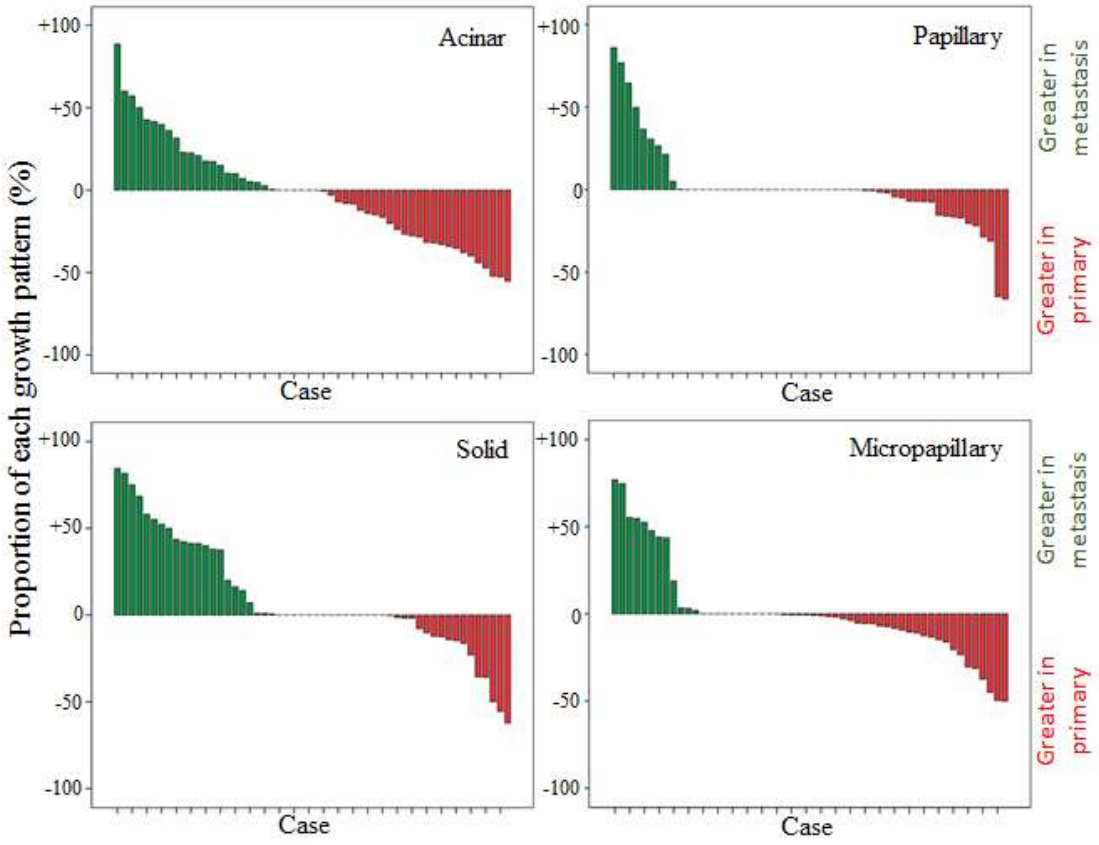


Figure 4.15: Waterfall plots showing the difference between the proportions of each growth pattern in primary lesions and nodal metastases. Cases with positive values (green) have a greater proportion in the metastasis whereas those with negative values (red) have a greater proportion in the primary lesion.

4.2 Nucleic acid analysis of lung adenocarcinomas

4.2.1 Transcriptomics

The cell-cycle progression score (CCP), calculated from the expression of 31 cell cycle genes, was used together with clinical stage to calculate mPS.[263] Of 206 tumour

samples available for molecular analysis 16 failed analysis (7.8%) (summarised in table 4.17).

Table 4.17: Summary of gene expression CCP score and mPS for the cohort of 206 lung adenocarcinomas.

	CCP score	mPS
Number of samples	206	206
Failed analysis	16	16
Mean (95% CI)	0.11 (-0.025 to 2.43)	30.49 (28.62 to 32.37)
Median	0.20	32.25
Range	-2.8 to 2.4	-3.5 to 54.8

The CCP score was found to be normally distributed, whereas mPS was not (Shapiro-Wilk test p= 0.069 and 0.021 respectively); however, since both CCP and mPS shows skewing and the datasets contain outliers, non-parametric tests were used to investigate relationships between CCP, mPS and morphological and molecular characteristics. The CCP score was significantly higher in tumours with solid pattern present ($p< 0.0005$) than in tumours without this growth; whereas tumours with the lepidic or papillary patterns had lower CCP scores ($p< 0.0005$ and $p= 0.007$ respectively). Tumours with the micropapillary pattern also had lower CCP score but the effect was less pronounced ($p= 0.034$). mPS was only found to be significantly correlated with the presence of the lepidic and solid patterns ($p<0.0005$ and $p= 0.025$ respectively) (figure 4.16).

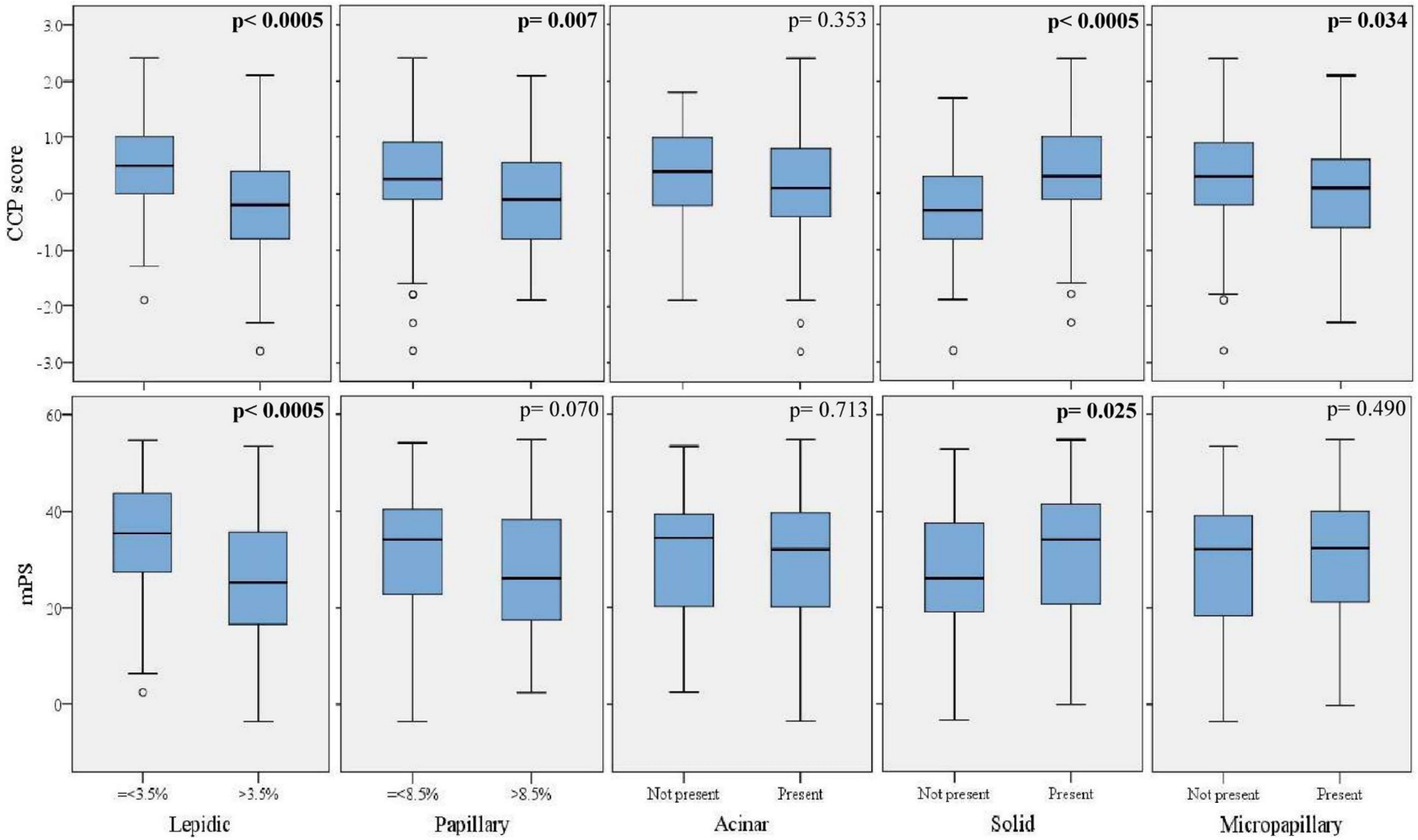


Figure 4.16: Boxplots showing the relationships between CCP score (top) and mPS (bottom) and the presence of different growth patterns. The significance of these relationships, calculated using the Mann-Whitney U test, is shown in the top right of each graph.

Both CCP score and mPS showed predominantly lepidic pattern tumours to have the lowest CCP score and, therefore mPS ($p< 0.0005$ and $p= 0.041$ respectively) (figure 4.17). The same was also true when grouped according to predominant morphological grade whereby CCP score and therefore mPS increased from low, to intermediate, to high grade tumours ($p< 0.0005$ and $p= 0.012$ respectively) (figure 4.18). However, there was no significant relationship between second most predominant pattern and CCP score ($p= 0.082$); although mPS did show an association ($p= 0.016$) which may be a result of the patient's clinical stage which was used to calculate this variable.

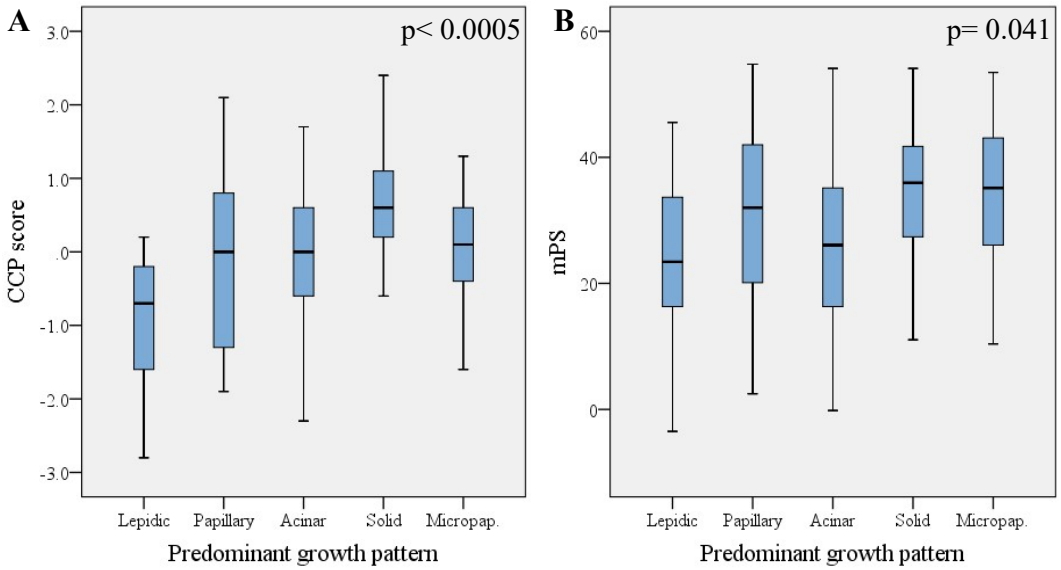


Figure 4.17: Boxplots showing the relationship between CCP score (A) and mPS (B) in tumours with different predominant growth patterns. The statistical significance of each relationship, calculated using the Kruskall-Wallis test is shown in the top right of each graph.

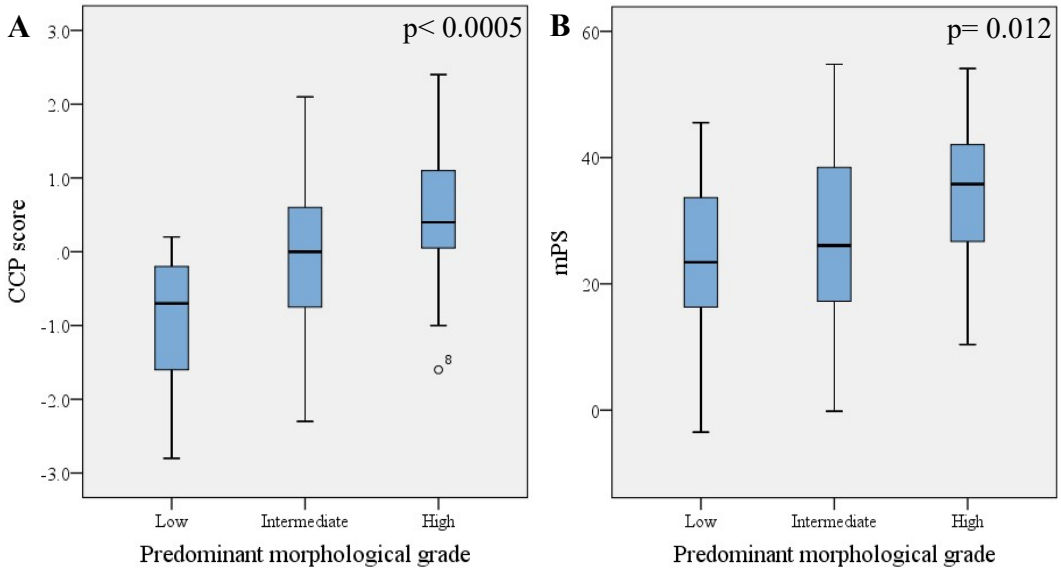


Figure 4.18: Boxplots showing the relationship between CCP score (A) and mPS (B) in tumours with predominant morphological grade. The statistical significance of each relationship, calculated using the Kruskall-Wallis test is shown in the top right of each graph.

The CCP score was significantly higher in tumours which had necrosis (p< 0.0005), mitosis (p< 0.0005) or apoptosis (p= 0.002) (figure 4.19). However, only the presence of necrosis and mitosis were significantly correlated with mPS (p< 0.0005 and p= 0.002 respectively). Although CCP score was not significantly associated with the presence of prominent nucleoli, or scar tissue mPS was (p= 0.014 and p= 0.016 respectively).

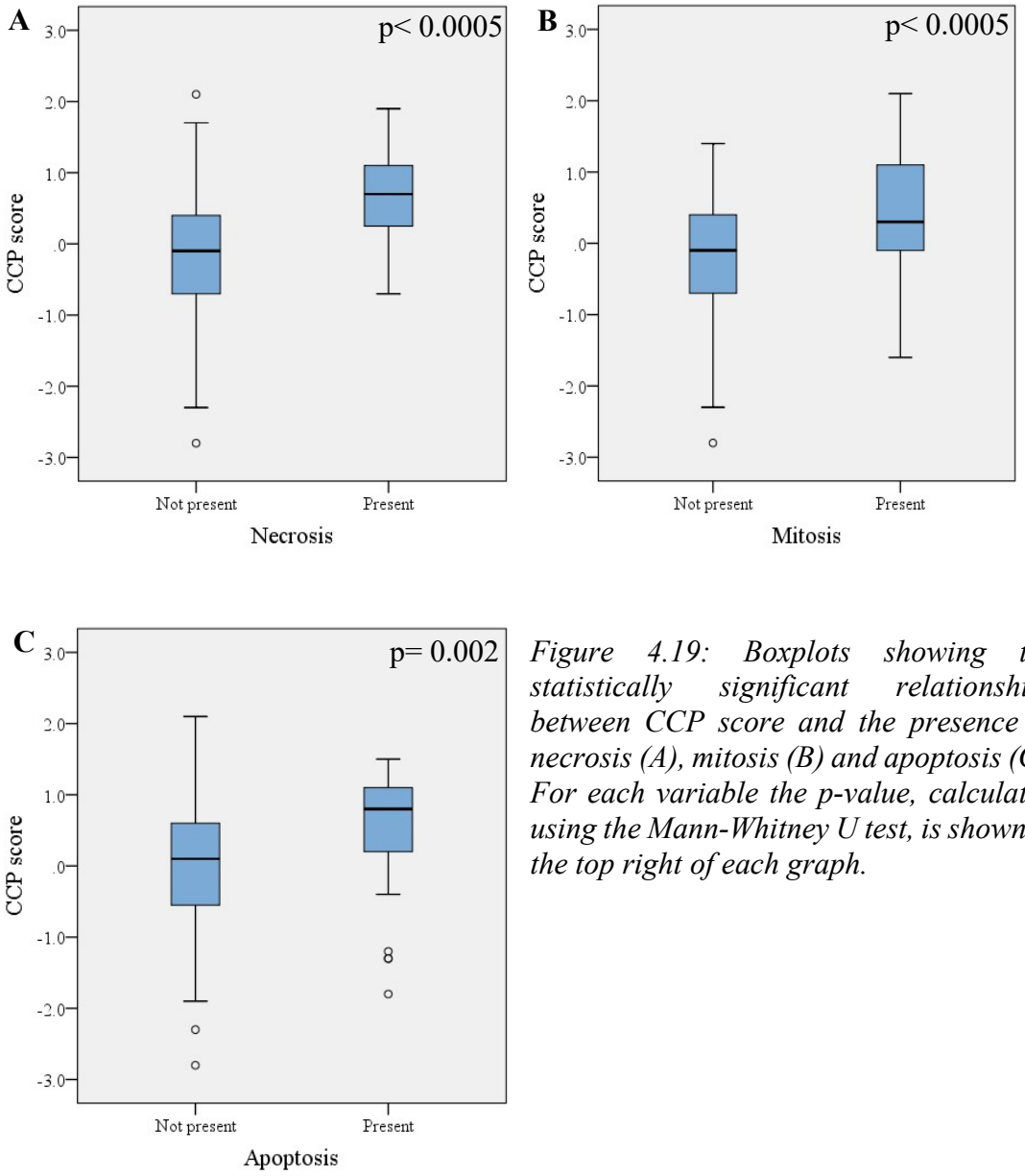


Figure 4.19: Boxplots showing the statistically significant relationships between CCP score and the presence of necrosis (A), mitosis (B) and apoptosis (C). For each variable the p-value, calculated using the Mann-Whitney U test, is shown in the top right of each graph.

The Mann-Whitney U test showed that CCP score and mPS were significantly associated with the overall tumour grade (derived from necrosis, mitosis and prominent nucleoli scar tissue, apoptosis and inflammatory cells) (p= 0.008 and p>0.0005 respectively) (figure 4.20).

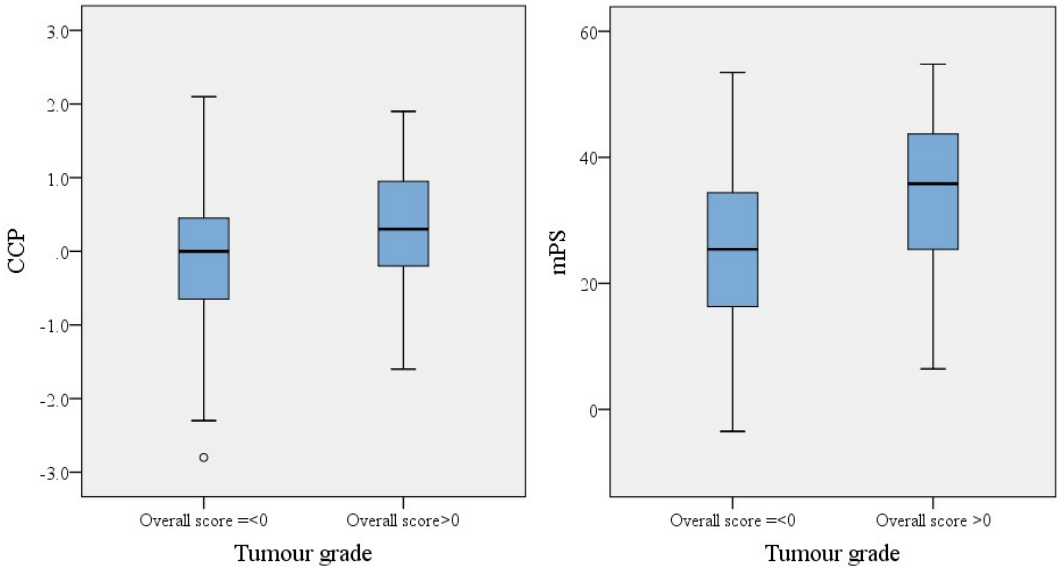


Figure 4.20: Boxplots showing the difference in CCP and mPS scores between tumours whose cytological features indicate low (overall score less than or equal to 0) or high (overall score greater than 0) grade (Mann-Whitney U test p= 0.008 and p>0.0005 respectively).

4.2.2 Molecular genetics of tumours from patients with N1 disease

From the cohort of 206 lung adenocarcinoma cases 70 had N1 disease, of which 3 cases had no tissue available for further testing and one case had a neoplastic cell content too low to allow molecular analysis. Sixty-six cases were available for *EGFR* and *KRAS* hotspot mutation analysis. Sixteen cases were found to harbour *EGFR* mutations using the *therascreen*® EGFR RGQ PCR kit; an additional case, with an insertion in exon 20, was identified by Sanger sequencing. One case, which was identified as having both a deletion in exon 19 and a p.(G719X) mutation using the *therascreen*® assay, was later characterised by Sanger sequencing as having a c.2240T>C p.(L747S) mutation rather than the expected deletion in exon 19. In total 25.8% of the N1 cohort which could be assessed had *EGFR* mutations in their tumours; p.(Leu858Arg) mutations were most common followed by p.(G719X) mutations. One p.(L858R) and 2 p.(G719X) mutations were reproducibly detected at a level below the threshold for the assay; meaning that in a clinical setting these tumours would have been considered to have no detectable mutation. *KRAS* mutation analysis of codons

12, 13 and 61 detected 33 (50%) cases with KRAS mutations; p.(G12V) was the most common followed by p.(G12C) (table 4.18).

Table 4.18: Summary of the results of EGFR and KRAS mutation detection in tumours from patients with N1 disease.

Mutation status		Total	%
EGFR			
	No Mutation detected	50	74.2%
	Mutation detected	17	25.8%
	Deletion in exon 19	*3*	
	p.(L858R)	*5**	
	p.(L861Q)	*1*	
	p.(G719X)	*4**	
	Insertion in exon 20	*2*	
	p.(G719X) and p.(L747S)	*1*	
	p.(G719X) and p.(S768I)	*1*	
KRAS			
	No mutation detected	33	50%
	Mutation detected	33	50%
	p.(G12C)	*10*	
	p.(G12V)	*11*	
	p.(G12D)	*5*	
	p.(G12A)	*2*	
	p.(Q61H)	*4*	
	p.(Q61L)	*1*	
Total		66	100%

** One tumour had a p.(L858R) mutation below the assay cut off as did 2 tumours with p.(G719X) mutations.*

In this cohort of patients with metastatic (N1 disease) there were no significant associations between the presence of *EGFR* or *KRAS* mutations and morphological characteristics. In addition, *EGFR* and *KRAS* mutation status was not associated with CCP score or mPS.

4.3 Candidate protein biomarkers

4.3.1 Validation of immunohistochemistry

Three commercially available antibodies were selected for use in IHC with archived lung adenocarcinoma FFPE tissue. The first, 31G7, was designed to bind to the extracellular domain of all EGFR protein; referred to below as tEGFR. The second,

D6B6, was designed to bind to EGFR protein bearing a p.(E746_A750del) mutation but not to normal EGFR; referred to as DEL. The third (43B2) was designed to bind to EGFR with a p.(L858R) mutation, but not to normal EGFR); referred to below as L858R.

Cases for validation of the mutation specific antibodies were identified from the NHS Lothian lung cancer clinical audit. Validation samples were selected on the basis of the *EGFR* mutation status, amount of tissue available and the age of the specimen. Tumours with a deletion in exon 19 of the *EGFR* gene were further characterised by Sanger sequencing to determine the specific mutation present. Seventeen validation samples were identified including: 6 samples with a p.(E746_A750) mutation, 6 samples with a p.(L858R) mutation and 5 with no *EGFR* mutation detected.

As expected, the tEGFR IHC showed basally located staining in normal bronchial epithelium and in normal glandular structures; whereas in tumour cells staining was membranous and cytoplasmic. Neither of the DEL or L858R IHC showed staining in non-tumour cells (figure 4.21).

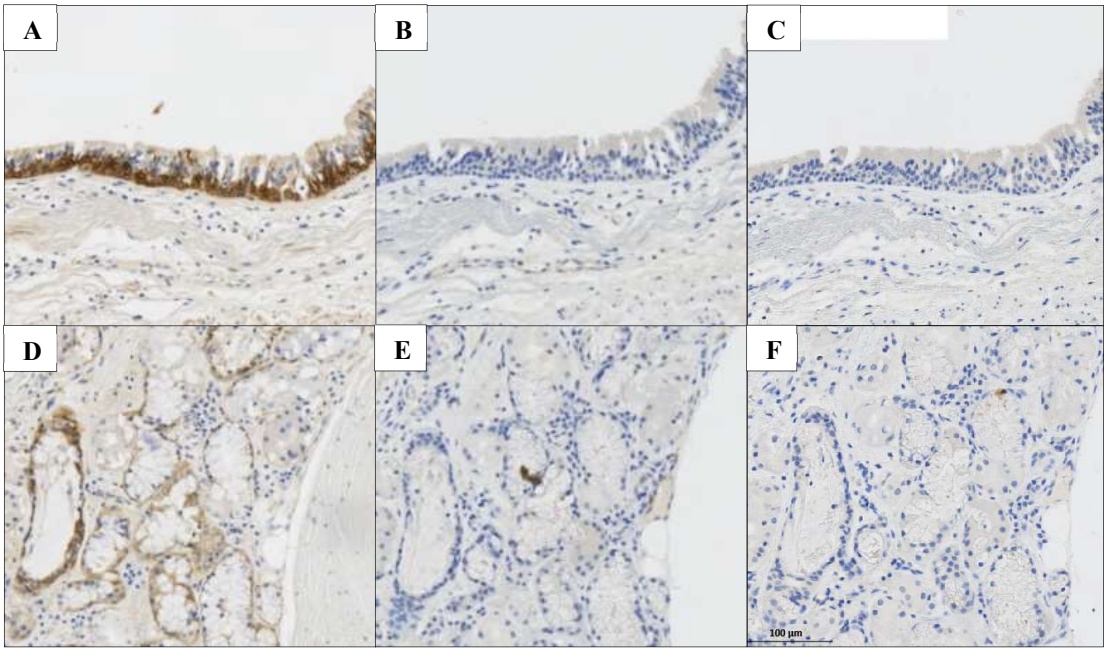


Figure 4.21: Validation of the mutation specific IHCs. Representative areas of normal respiratory epithelium (A, B and C) and submucosal glands (D, E and F) stained with tEGFR (A and D), DEL IHC (B and E) and L858R (C and F). The tEGFR stains the basal regions of respiratory epithelium and glandular structures which are unstained with DEL and L858R.

Two samples in the validation cohort gave unexpected results by IHC (table 4.19). Sample V11 was shown by *therascreen*® to have a c.2573T>G mutation indicating a p.(L858R) amino acid change; however, this sample was negative by L858R IHC. On further investigation, there was a small amount of staining with the L858R IHC but this was not sufficient to be considered positive. In order to rule out a false positive result by *therascreen*® assay the presence of an *EGFR* c.2573T>G p.(L858R) mutation in this sample was confirmed by Sanger sequencing. Sample V14 was classified by *therascreen*® as "no mutation detected" but showed positive staining with the L858R IHC. On review the staining was found to be nuclear as well as cytoplasmic and membranous; the cytoplasmic staining was often granular. Sanger sequencing also did not detect a mutation in this sample and the high estimated neoplastic cell content (80% data not shown) would have made a false negative unlikely.

*Table 4.19: Summary of the validation of mutation specific IHCs. DEL and L858R mutation specific IHC assays were compared to the expected mutation as identified by DNA mutation analysis using the therascreen EGFR RGQ PCR kit and/or Sanger sequencing. Two samples showing discordant results are highlighted in **bold**.*

Sample no.	*EGFR* mutation status	Mutation characterised by Sanger	DEL IHC	L858R IHC
V1	Deletion (exon 19)	c.2236_2250del p.(E746_A750del)	+	-
V2	Deletion (exon 19)	c.2235_2249del p.(E746_A750del)	+	-
V3	Deletion (exon 19)	c.2236_2250del p.(E746_A750del)	+	-
V4	Deletion (exon 19)	c.2236_2250del p.(E746_A750del)	+	-
V5	Deletion (exon 19)	c.2236_2250del p.(E746_A750del)	+	-
V6	Deletion (exon 19)	c.2235_2249del p.(E746_A750del)	+	-
V7	p.(L858R)	-	-	+
V8	p.(L858R)	-	-	+
V9	p.(L858R)	-	-	+
V10	p.(L858R)	-	-	+
V11	**p.(L858R)**	**c.2573T>G p.(L858R)**	**-**	**-**
V12	p.(L858R)	-	-	+
V13	No mutation	-	-	-
V14	**No mutation**	**No mutation**	**-**	**+**
V15	No mutation	-	-	-
V16	No mutation	-	-	-
V17	No mutation	-	-	-

Within this limited validation cohort, the DEL IHC had a sensitivity of 100% and a specificity of 100% whereas the L858R IHC had a sensitivity and specificity of 83.3% and 90.9% respectively.

4.3.2 Mutation specific IHC

The 17 EGFR mutated N1 lung adenocarcinomas were stained with the DEL, L858R and tEGFR IHCs as well as the routinely-used lung marker thyroid transcription factor (TTF1) (table 4.20). All tumours showed positive staining with the TTF1 and tEGFR IHC, although in 7 of the 17 cases the quality of staining was very poor; this was assumed to be caused by the quality of tissue fixation or processing. The tumours known to harbour p.G719X, p.L861Q, p.S768I, p.L747S mutations or insertions in exon 20 showed no positive staining with the DEL or L858R IHCs. Of the 3 cases with deletions in exon 19 of EGFR, only 2 showed positive staining with DEL IHC; all

three were negative with L858R IHC. The DEL negative sample was characterised by Sanger sequencing as c.3337_2255delinsT p.(E746_S752delELREATSinsV) rather than the p.E746_A750del epitope the antibody was designed to detect. Four of the 5 cases known to have p.(L858R) mutations showed positive staining with the L858R IHC. One case with a suspected low level p.(L858R) mutation (detected below the threshold for the *therascreen*® assay) was negative with the L858R IHC. Sanger sequencing was unable to confirm the presence of this, or any other, mutation. All of the 5 p.(L858R) cases were negative with DEL IHC (table 4.20).

Table 4.20: Summary of the results of IHC testing on EGFR mutation positive N1 lung adenocarcinomas.

Case	*EGFR* mutation	TTF1	tEGFR	DEL	L858R
9	p.2156G>C p.G719A	+	+	-	-
10	p.G719C and p.L747S	+	+	-	-
23	Deletion in exon 19	+	+	+	-
27	p.L858R	+	+	-	+
38	Low level p.L858R	+	+	-	-
48	p.L858R	+	+	-	+
49	p.G719X	+	+	-	-
51	p.L858R	+	+	-	+
53	Deletion in exon 19	+	+	+	-
63	p.G719X	+	+	-	-
99	Insertion in exon 20	+	+	-	-
107	Low level p.G719X	+	+	-	-
118	p.G719X and p.S768I	+	+	-	-
192	c.2311delinsGGGG p.N771delinsGD	+	+	-	-
196	p.L861Gln	+	+	-	-
203	p.L858R	+	+	-	+
208	c.3337_2255delinsT p.Glu746_Ser752delinsV	+	+	-	-

4.3.3 Intratumour heterogeneity of EGFR

In the 6 cases which stained positively with either the DEL or L858R IHCs the pattern of staining intensity across each tumour using the mutation specific assays was matched by the pattern produced with tEGFR IHC; indicating that any intratumour heterogeneity was likely to be caused by a variation in the expression levels of the EGFR protein rather than intratumour heterogeneity of mutation status. There was no evidence to suggest intratumour heterogeneity of mutation status.

Eight of the 17 cases showed heterogeneous staining intensity across the tumour with the tEGFR IHC, and where appropriate the mutation specific assays. This heterogeneity was not observed in the TTF1 IHC, indicating that this variation may be related to the quantity of EGFR protein in different areas of the tumour rather than an artefact of fixation (figures 4.22 to 4.27).

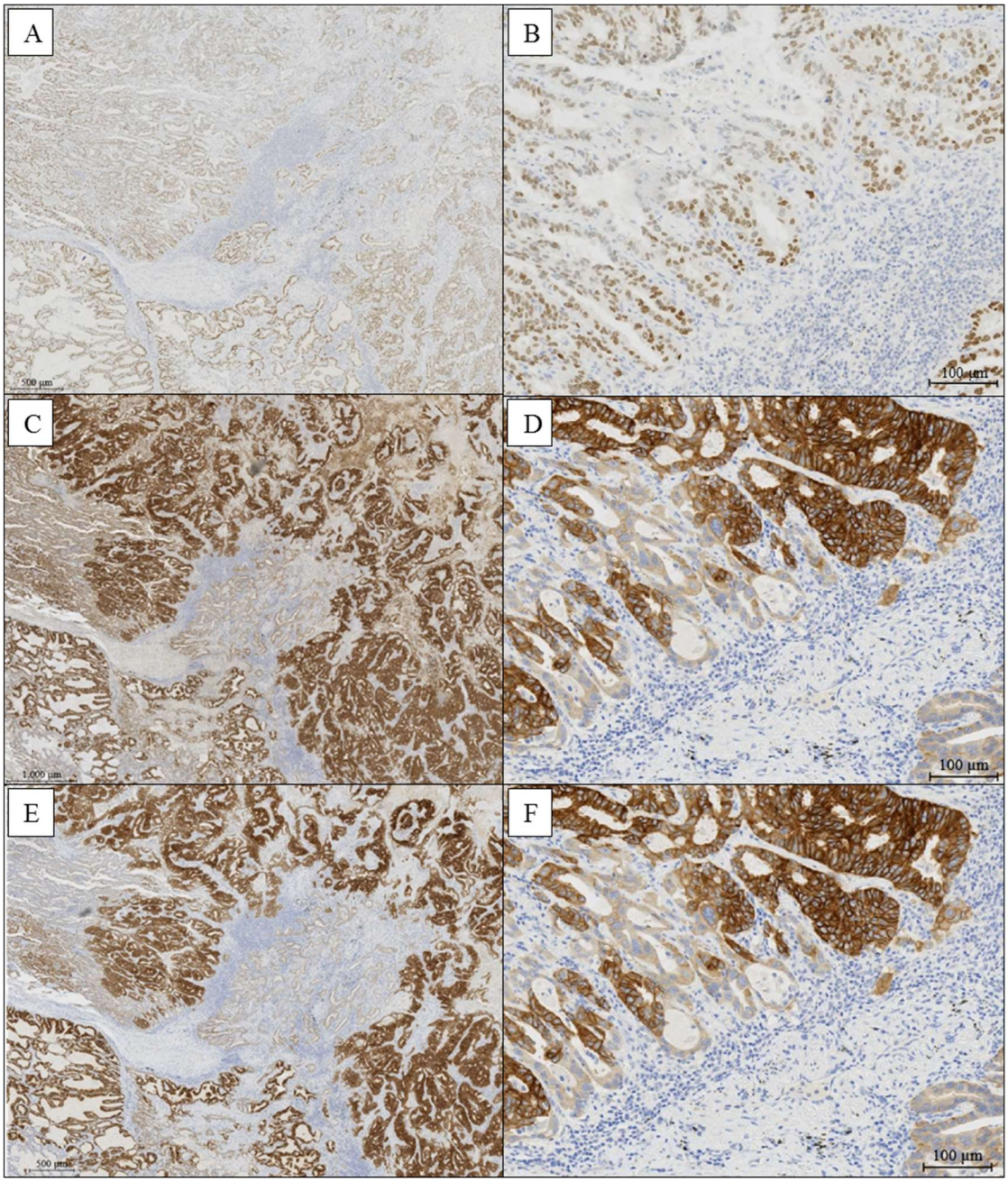


Figure 4.22: Photograph of TTF1 (A and B), tEGFR (C and D) and L858R IHC (E and F) on case 48 at low (A, C and E) and high power (B, D and F); intratumour variation in staining intensity with the L858R and tEGFR IHCs was not matched by similar variation in TTF1.

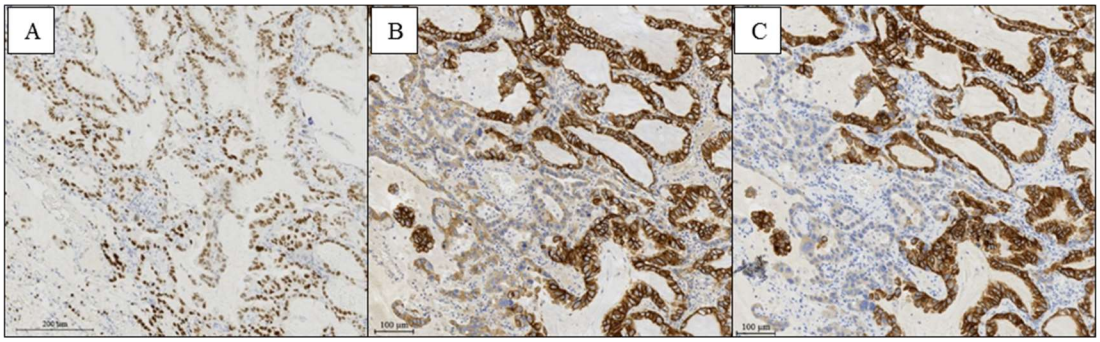


Figure 4.23: Photograph of TTF1 (A), tEGFR (B) and L858R IHC (C) on case 48 showing further intratumour variation in staining intensity with the L858R and tEGFR IHCs not matched by similar variation in TTF1.

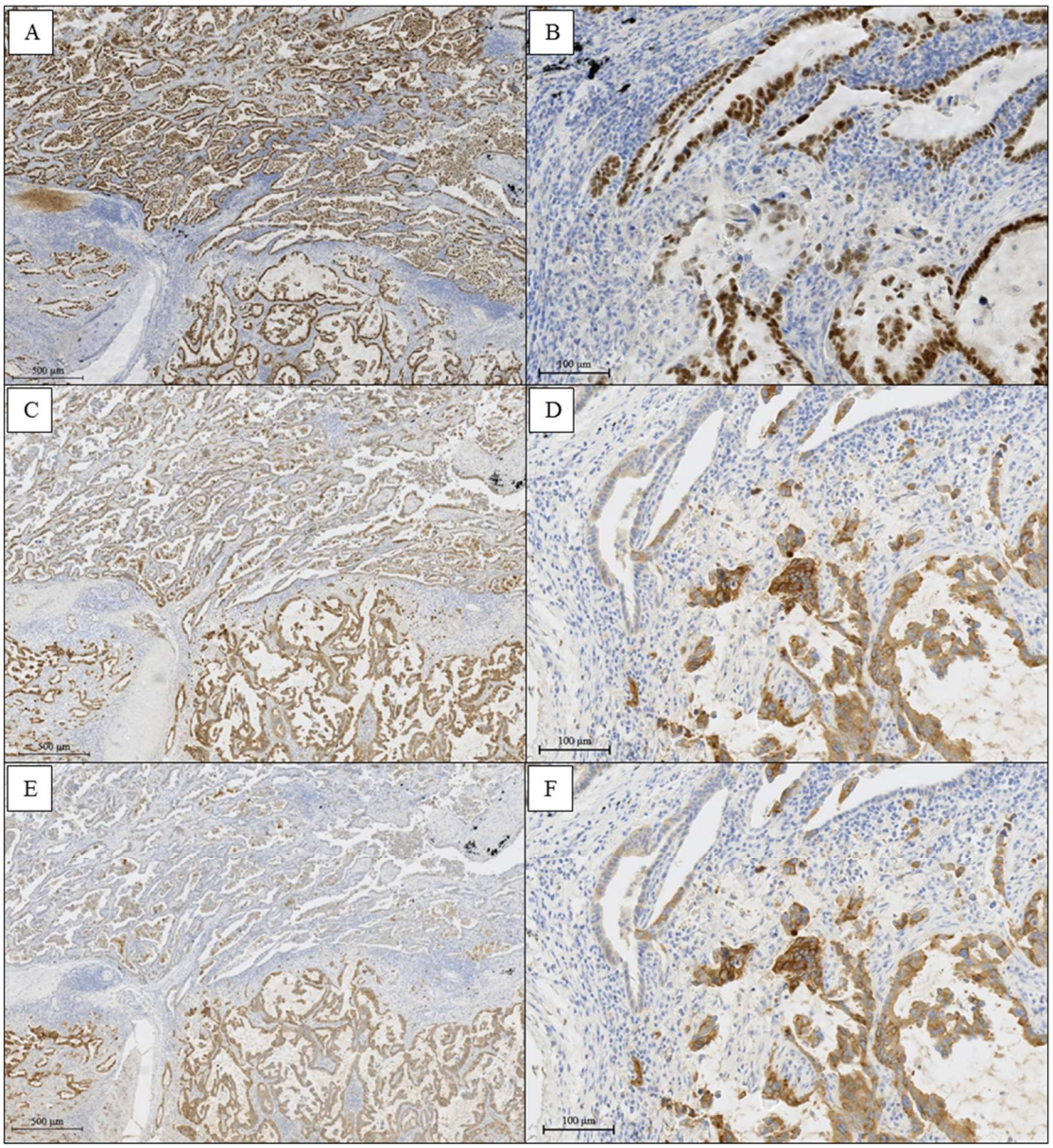


Figure 4.24: Photograph of TTF1 (A and B), tEGFR (C and D) and L858R IHC (E and F) on case 27 at low (A, C and E) and high power (B, D and F); intratumour variation in staining intensity with the L858R and tEGFR IHCs was not matched by similar variation in TTF1.

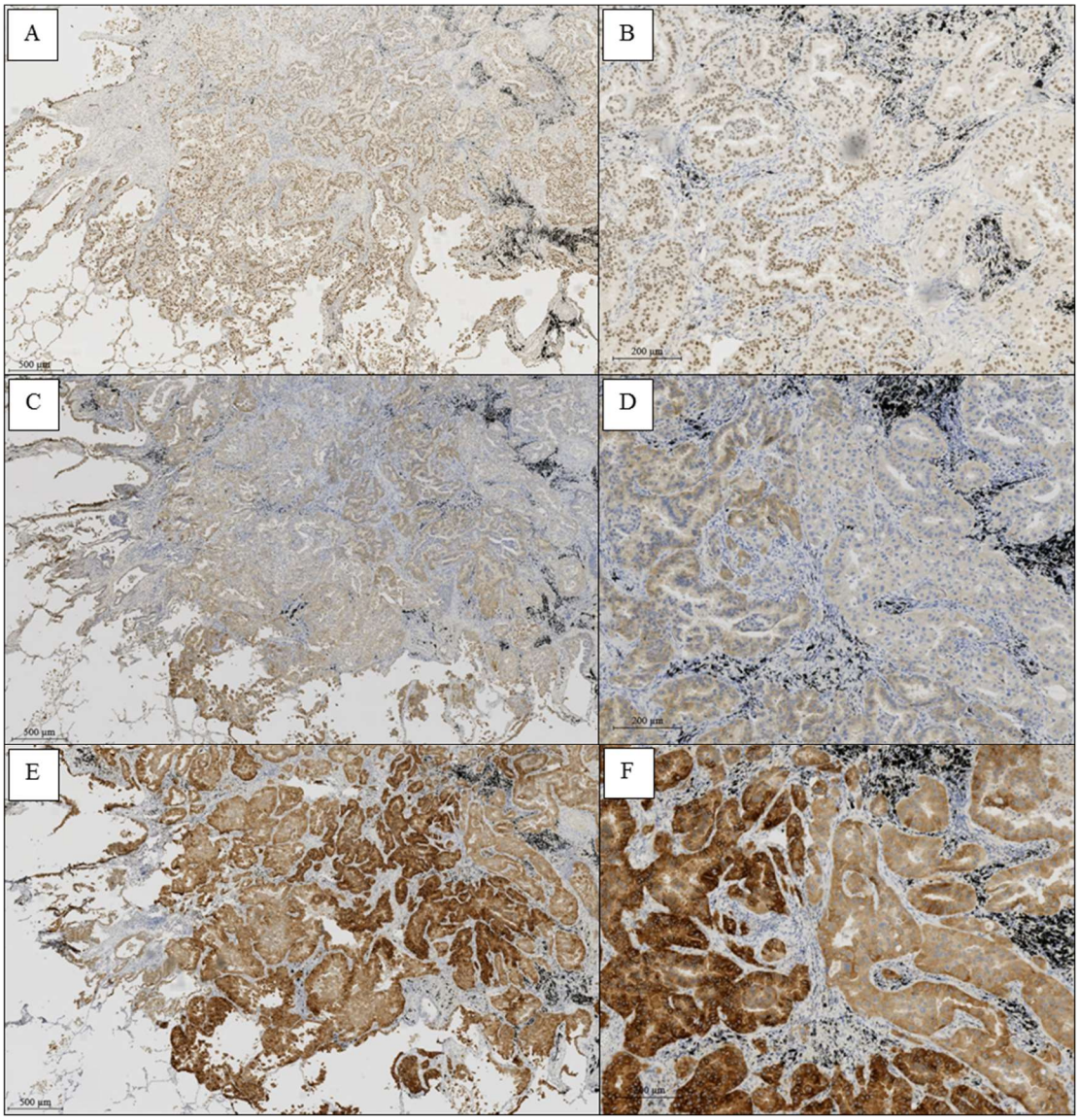


Figure 4.25: Photograph of TTF1 (A and B), tEGFR (C and D) and L858R IHC (E and F) on case 53 at low (A, C and E) and high power (B, D and F); intratumour variation in staining intensity with the L858R and tEGFR IHCs was not matched by similar variation in TTF1.

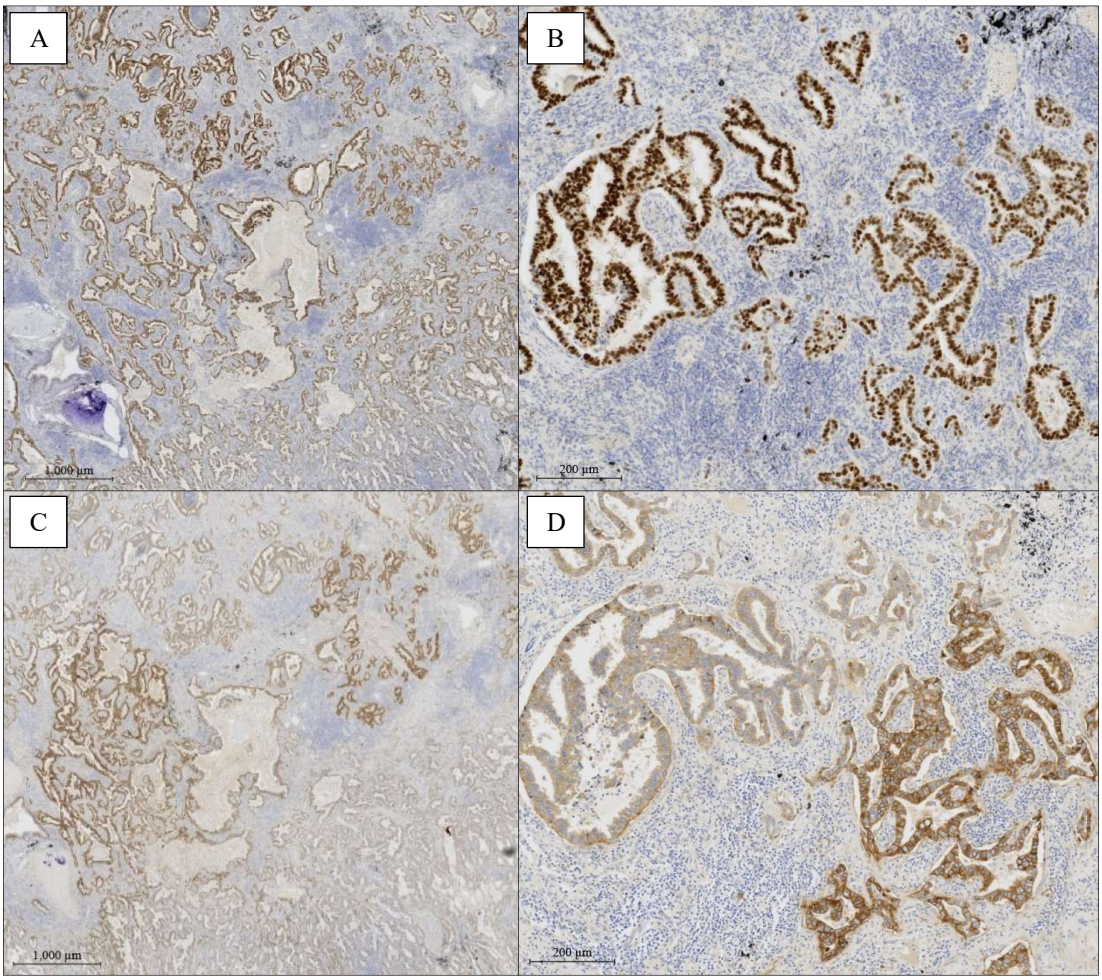


Figure 4.26: Photograph of TTF1 (A and B) and tEGFR (C and D) on case 99 at low (A and C) and high power (B and D); intratumour variation in staining intensity with the tEGFR IHC was not matched by similar variation in TTF1.

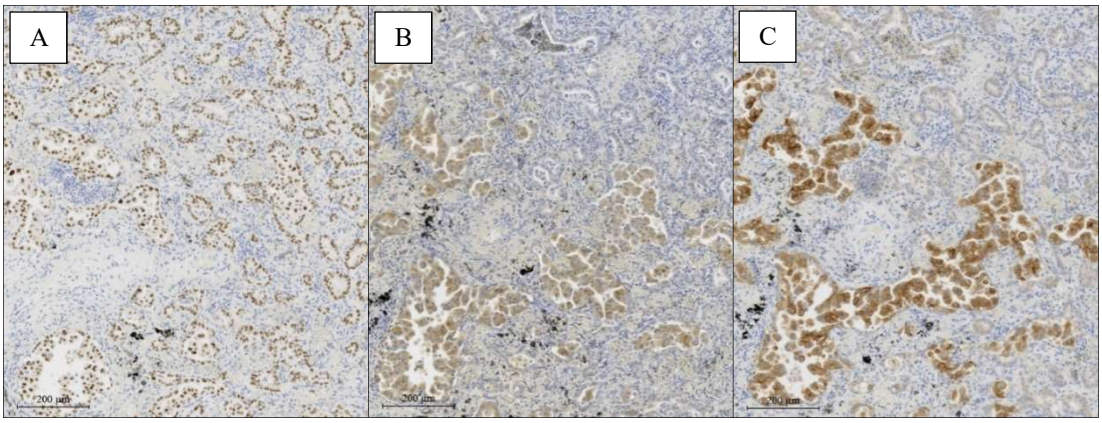


Figure 4.27: Photograph of TTF1 (A), tEGFR (B) and L858R IHC (C) on case 203; intratumour variation in staining intensity with the L858R and tEGFR IHCs was not matched by similar variation in TTF1.

In some areas the staining intensity could be markedly different and in very closely associated regions of the tumour (figure 4.28).

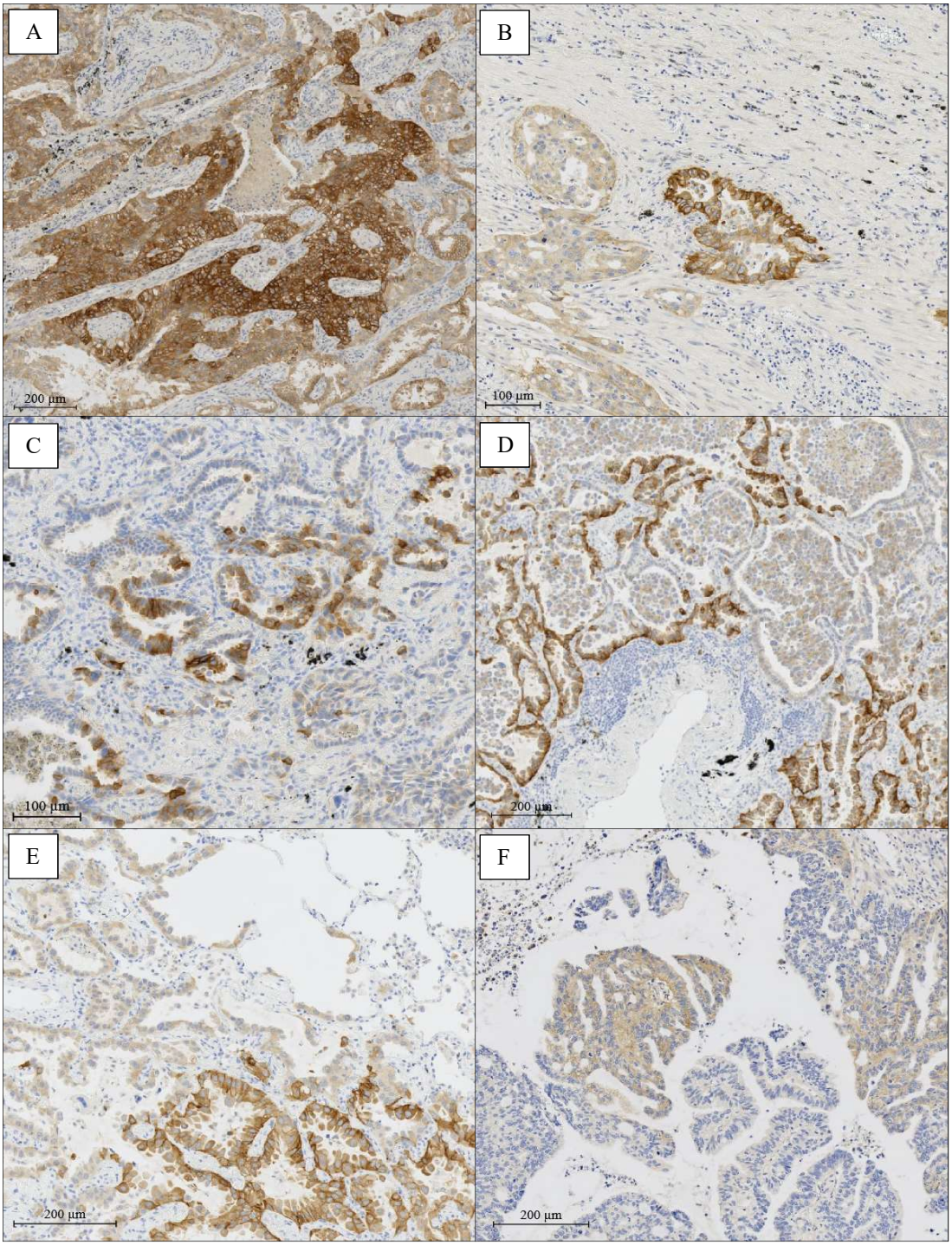


Figure 4.28: Photographs of DEL (A), L858R (C, D and E) and tEGFR (B, F) IHC showing differences in staining intensity of closely associated areas of tumour.

In one case the localisation of EGFR protein showed marked differences between different areas of the tumour, basal localisation was seen in areas of the tumour with

lepidic growth pattern but closely associated areas with a solid growth pattern showed membranous and cytoplasmic staining (figure 4.29).

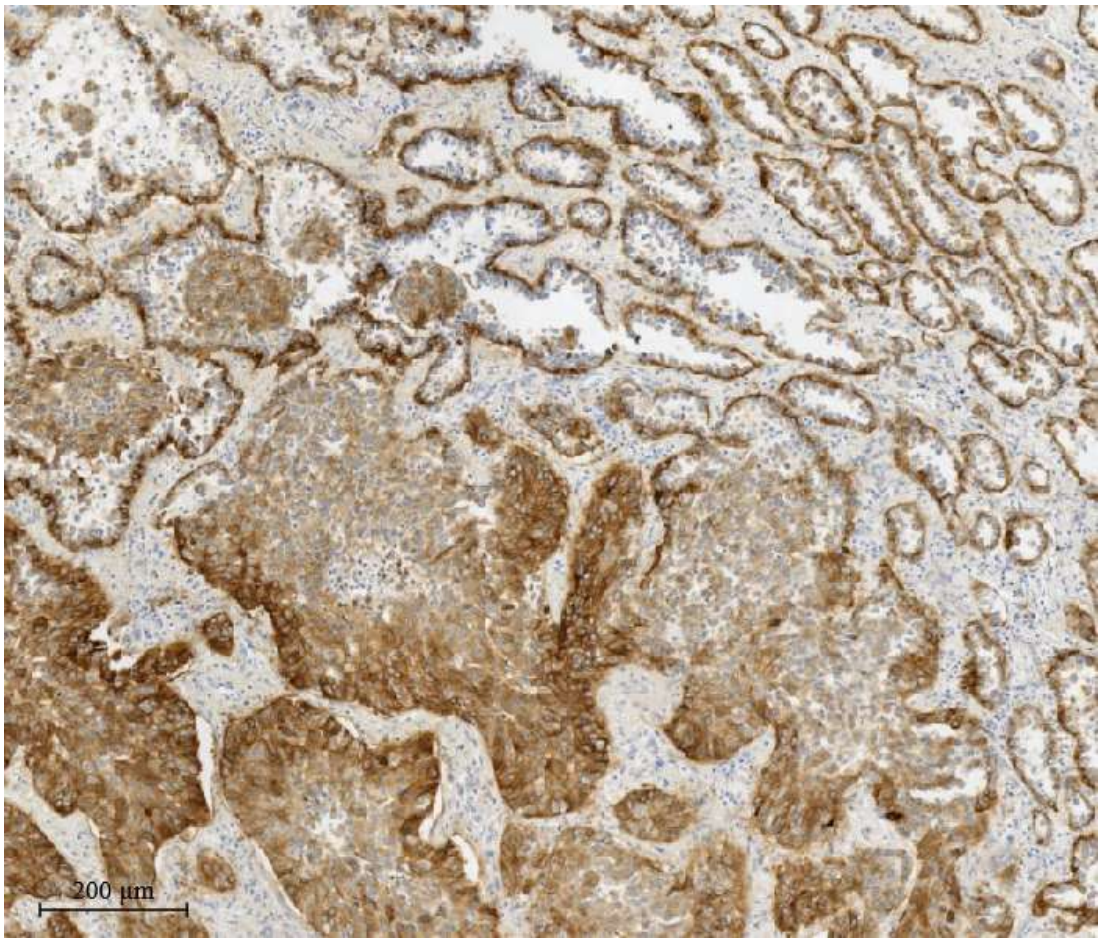


Figure 4.29: Photograph of case 9 stained with tEGFR showing heterogeneity in the localisation of the EGFR protein between areas with different growth patterns; lepidic growth shows staining in the basal regions whereas in solid growth staining is membranous and cytoplasmic.

4.4 Discussion

Full histopathological assessment has been carried out on a total of 207 resected lung adenocarcinomas comprising 907 tumour containing sections. The proportion of lepidic, papillary, acinar, solid and micropapillary growth patterns were estimated in 5% increments as a percentage of each lesion as advised by Travis *et al*.[261] However; since the publication of these recommendations, and the assessment of our cohort, the classification of lung adenocarcinomas has been further refined.[14] It is now suggested that areas showing micropapillary growth within acinar structures should be classified as micropapillary; in the presented dataset efforts were made to account for both the acinar and micropapillary patterns in these areas.[14] The cribriform pattern, which has in this study and previously been considered a variant of the acinar pattern,[40] has been reported to have a poor prognosis and may be categorised as an additional pattern.[14, 59, 281-283] The presence of tumour budding, described as single cells or small clusters within fibrotic stroma, may also represent a poor prognosis and should, potentially, be classified accordingly.[284] In addition, recently it has been suggested that the papillary growth pattern should be sub-divided into 3 categories based on an association between specific morphological features and patient outcomes.[56]

It commonly accepted that lung adenocarcinomas are frequently a heterogeneous mix of 2 or more growth patterns which change from one to the next in a continuum rather than with a defined boundary.[14, 285] In this cohort of 207 lung adenocarcinomas over 92% of tumours contained 2 or more growth patterns and 14.5% showed elements of all 5. The incidence of each growth pattern was initially investigated using a 5% cut off, below which it was considered that false positives could be created by cutting artefact. The acinar growth pattern was the most commonly observed and was found in highest proportions, followed by the solid pattern. In tumours with only 1 growth pattern this was, in the majority of cases, the solid pattern. The reasons for this are unknown but may suggest, in these cases at least, a difference in the aetiology of this form of tumour growth. Common co-occurrences were found between the papillary and acinar or lepidic, as well as between the micropapillary and the papillary, acinar or solid patterns. Whereas; when the solid pattern was observed the lepidic, papillary and acinar patterns were less likely to be present. Whether these co-occurrences,

particularly between papillary and lepidic or acinar, truly reflect the morphology of the tumours or represent the difficulty in differentiating some growth patterns is unknown.

Large studies evaluating the prognostic value of predominant growth patterns have not, as yet, detailed their method of assessment; although a smaller study has attempted to define a protocol.[59] It is assumed that histopathologists should make approximate, 'by eye', estimates over the whole case which can include many tumour containing sections. In the current study an attempt was made to define and investigate different methods of estimating the predominant growth pattern; however, each method still relied upon subjective estimates of the proportions in each section. A calculation for predominant pattern which incorporated the relative size of the tumour tissue in each section proved the most reproducible between observers and had the highest level of agreement with other calculations.

An obvious criticism of the above analysis of 207 lung adenocarcinomas is that the primary assessment, upon which the majority of analyses were based, was carried out by a non-histopathologist. However, an attempt was made to assess the discordance between the primary researcher and a consultant histopathologist specialising in respiratory pathology. Indeed, studies examining inter-observer agreement between histopathologists in the assessment of predominant growth pattern showed similar levels of discordance.[286] Considerable difficulties have been highlighted in differentiating between lepidic and papillary, lepidic and acinar and papillary and micropapillary patterns; whereas the identification of the solid pattern was more reliable.[286-289] Similarly, in the current study there was greatest inter-observer agreement in identifying predominantly solid tumours whereas the papillary predominant tumours had the least agreement.

The presence of multiple growth patterns and the ill-defined boundary between them, as well as poor tissue preservation, is likely to make the estimation of the proportions of each pattern very inaccurate. Given these difficulties in estimating the proportion of growth pattern it is not surprising that the number of each predominant growth pattern in published cohorts varies considerably; although predominantly micropapillary

tumours were considered relatively rare in most studies (table 4.21). In addition, the effect of environmental, behavioural and genetic factors on the incidence of each predominant growth pattern is not known.

Table 4.21: Summary of studies of predominant growth pattern in large cohorts of lung adenocarcinomas.

Study	No. cases	Lepidic	Papillary	Acinar	Solid	Micropapillary
Warth et al. 2012 [33]	500	8.4%	4.7%	42.5%	37.6%	4.7%
Russell et al. 2011[40]	210	5%	12%	40%	23%	7%
Yoshizawa et al. 2010[15]	483	6.0%	29.6%	48.0%	13.9%	2.5%
Sumiyoshi et al. 2013[20]	373	9.7%	48%	16.4%	20.9%	5.1%
Tsuta et al 2013[34]	757	18.0%	44.6%	12.9%	16.4%	8.1%
Zhang et al. 2014[45]	218	21.6%	13.3%	57.3%	6.4%	1.4%
Current study	207	9.7%	12.6%	37.8%	30.0%	12.1%

In addition to tumour growth patterns, tumour grading characteristics were included in this analysis; including the presence of large amounts of: necrosis, inflammatory infiltrate, mitotic figures, apoptotic bodies, prominent nucleoli, cytological pleomorphism, scar tissue and dyscohesive tumour cells. Additional investigations, for example immunohistochemistry for the identification of lymphocyte types or the presence of macrophages, would represent a considerable increase to the routine running costs of a clinical histopathology service and were, therefore, not carried out. All included variables could be assessed from H&E stained sections with minimal added workload or expense. However; it should be noted that these qualitative variables were post hoc analyses from descriptive data captured during morphological assessment; no formal scoring criteria were defined, although all cases were assessed by the same researcher; intra-observer variation has also not been investigated. Tumours with the lepidic pattern were associated with fewer mitotic figures, apoptotic bodies, necrosis and scar tissue. Papillary tumours frequently had prominent nucleoli and necrosis was rare. Acinar tumours commonly had few apoptotic bodies and were associated with the absence of large areas of necrosis. Micropapillary tumours regularly had prominent nucleoli and were associated with a lack of necrosis. Solid tumours more commonly had large areas of necrosis and were associated with many more mitotic figures and apoptotic bodies; however, as these tumours typically have

large amounts of relatively clear cytoplasm it is possible that these features are easier to identify in the solid tumour growth than in other patterns.

There was some evidence to suggest that over the whole cohort the solid growth pattern was likely to increase in proportion in nodal metastases compared to primary tumours. Whether this means that the solid pattern was most likely to metastasize or that the altered environment of a lymph node favoured the solid growth pattern is unknown.

There are significant challenges to the routine use of comprehensive morphological assessment in a clinical setting including the still-evolving categories, standardisation of scoring methods and minimisation of inter and intra-observer discordance. It is hoped that the following in-depth analysis of this cohort of lung adenocarcinomas will help inform future practices to the benefit of lung cancer patients.

Gene expression analysis failed in 7.8% of samples which, considering its dependence on RNA isolated from archived FFPE samples of resection specimens, shows this analysis to be relatively robust. The associated CCP score was found to be linked to predominant morphological grade, being lowest in predominantly lepidic tumours. In addition, CCP score was higher when necrosis, mitosis or apoptosis was found in a tumour.

Three commercially available antibodies were validated for use in IHC on lung adenocarcinoma tissues: one with affinity to the extracellular domain of the EGFR protein (tEGFR), one to EGFR which bears a deletion on exon 19 (DEL) and one to p.L858R mutated EGFR (L858R). As expected, the tEGFR IHC showed basally located staining in normal bronchial epithelium and glandular structures whereas in tumour cells staining was membranous and cytoplasmic.[290] A meta-analysis of studies investigating the diagnostic utility of mutation specific IHC has previously shown the overall sensitivity to be low (66% and 76% respectively) although specificity was found to be high (98% and 96%).[221] For the DEL IHC this is likely to be, at least in part, caused by the presence of less common deletion mutations which would be likely to create a very different epitope.[220, 221] In this validation the DEL IHC had a sensitivity and specificity of 100%; although, only cases classified as having a

p.Glu746_Ala750del mutation were included. The L858R IHC had a sensitivity of 83.3% and a specificity of 90.9% caused by one false negative and one false positive result, in both cases Sanger sequencing confirmed the original molecular analysis and the reasons for these non-concordant results are not known. It is possible that, in cases showing a false negative result by IHC, the mutated *EGFR* allele was not expressed or that a second mutation was present within the epitope. The false positive result could have been caused by a mutation present at a low allele frequency undetectable by molecular methods; although the relatively high neoplastic cell content of this sample makes this explanation unlikely. The unusual staining pattern observed in this sample may favour non-specific binding to another epitope as the most likely explanation. Other groups have suggested that mutation specific IHCs have diagnostic utility as a screen prior to nucleic acid-based mutation analysis whereby any patient whose sample was found to be positive by IHC would proceed to EGFR targeted therapy without molecular mutation analysis. However, these studies were carried out in East Asian laboratories where the proportion of lung cancers with EGFR mutations is much higher. In a Caucasian population, with an EGFR mutation rate of 10.5% (chapter 3), there would be minimal cost savings from this algorithm (data not shown). In addition, the low specificity of the L858R IHC would mean patients with false positive results would wrongly receive therapy with EGFR TKI even though, as EGFR wild-type patients, they would be more likely to benefit from chemotherapy.[95] Although this validation showed that the mutation specific IHCs would be unsuitable for clinical use; they provide a powerful tool to examine the heterogeneity of EGFR mutations between individual cells within a tumour which is as-yet unmatched by less cost effective molecular techniques.

Mutation specific IHC assays have been used to show that intratumour heterogeneity of EGFR mutation status, whereby a mutation is not present in all tumour cells, was present in a large proportion of lung adenocarcinomas and correlated with poorer response to EGFR TKIs.[223, 291-293] In contrast one group found mutated EGFR protein to be homogeneously expressed across the tumours in their cohort.[294] Without direct cell-to-cell comparison with an IHC which stains all EGFR protein, whether mutated or not, it is impossible to determine if observed heterogeneity is a result of a difference

in mutation status or in the quantity of EGFR protein in each cell. In our cohort, the pattern of staining intensity with mutation-specific IHCs across each tumour was matched by the pattern produced with tEGFR IHC indicating that any intratumour heterogeneity was likely to be caused by a variation in the expression levels of the EGFR protein rather than intratumour heterogeneity of mutation status. Although; in this small cohort, and with a limited set of mutation specific antibodies, it is not possible to rule out intratumour heterogeneity of EGFR mutation status. In addition, in cases with heterogeneous staining, the positive control (TTF1) was largely uniform across the tumours and therefore the heterogeneity of EGFR protein expression was assumed to be independent of tissue quality. It is possible that the routinely used TTF1 IHC was more robust than the experimental EGFR targeting assays and would be less affected by poor tissue quality. However; variation in staining intensity was, in some cases, very pronounced and observed in very closely associated tumour cells; suggesting that this heterogeneity was unlikely to be an artefact of tissue fixation. The clinical effect of variation in expression levels is largely unknown but it is conceivable that the quantity of mutated EGFR protein would affect the effective dose of EGFR TKIs. One case showed marked differences in the localisation of EGFR protein across the tumour an observation that has been reported by other groups.[295, 296] The clinical effect of differing localisation of EGFR is unknown.

This study has provided evidence for the presence of intratumour heterogeneity in staining intensity suggesting heterogeneity in EGFR protein expression levels and localisation within the cell in NSCLCs. This heterogeneity may have an effect on a patient's response to EGFR tyrosine kinase inhibitors and may help to refine predictive models. Further work is required in a larger retrospective cohort of patients who have been treated with EGFR targeted therapies in order to explore the relationship between IHC staining and response to therapy and a quantitative measure of EGFR protein expression would aid this research. In addition; development of further mutation specific antibodies, particularly to the p.T790M mutation, would be a powerful tool in investigating mechanisms of acquired resistance.

Chapter 5 Prognosis

5.1 Background

Many studies have reported that tumour growth patterns identified in lung adenocarcinomas can help classify patients according to their likely prognosis; with the lepidic pattern thought to be associated with a good prognosis and the solid and micropapillary patterns with poorer outcomes.[15, 17, 21, 25, 33, 34, 42, 48] It has been proposed that, as part of the routine reporting of resected lung adenocarcinomas, histopathologists should estimate the proportion of lepidic, papillary, acinar and micropapillary growth patterns in 5% increments and determine the predominant growth pattern.[261, 285] However, it is widely known that lung adenocarcinomas are highly morphologically heterogeneous and few details of the method used to make this judgement have been published.

Unlike other malignancies, there is no widely accepted system of tumour grading in lung adenocarcinomas. However, some morphological features have been proposed as predictors of patient outcomes including: the presence of necrosis,[15, 59, 64, 65] mitotic figures,[59, 283] apoptotic bodies,[61] large numbers of inflammatory cells,[69, 71-74, 76-79] cytological pleomorphism[59, 63, 67] and scar tissue.[67]

In order to investigate the prognostic significance of morphological variability in patients with pulmonary adenocarcinomas a cohort of 207 resected stage I and II lung adenocarcinomas underwent comprehensive morphological analysis as described in chapters 2 and 4. Variables derived from this analysis included: the presence of each growth pattern, the proportion of each growth pattern, the predominant growth pattern and morphological grade, the second-most predominant growth pattern, the number of growth patterns and the presence of necrosis, mitosis, apoptosis, nucleoli, pleomorphism, scar tissue, inflammatory cells and dyscohesive tumour cells. In addition, 1 section from each tumour was sent to Myriad Genetics Inc. for gene expression analysis of 31 cell-cycle related genes.[297] The resulting CCP score was used together with the clinical stage to calculate a molecular prognostic score (mPS, as defined in chapter 2) which has been previously reported to be an indicator of prognosis in lung adenocarcinoma patients.[239, 263, 298] Univariable and multivariable

analyses were carried out on all available variables using Cox's proportional hazards to identify characteristics predictive of prognosis. However, this form of multivariable analysis is not without its limitations as assessment of interdependence is laborious in datasets with a large number of variables and strategies to correct for type I error make significant p-values difficult to achieve. Classification and Regression Tree (CART) analysis is not affected by these considerations and has been used in this study as an additional method of multivariable analysis and to design possible decision trees for the prognostication of lung adenocarcinoma patients.

5.2 The prognostic value of morphological, and molecular characteristics

From the cohort of 208 patients, 158 (76%) died within the timeframe studied although 47 were not considered to be as a result of their lung cancer and were therefore censored in survival analysis. The disease specific survival (DSS) was calculated from the date of resection to the date of death where relevant and censored for non-lung cancer related death. The log rank test (Mantel-Cox) showed that nodal status (N stage), and clinical stage were significantly associated with DSS (0.0005 and 0.001 respectively). Compared to N0 patients those with N1 disease had an HR of 2.158 (95% CI 1.476 to 3.156, $p< 0.005$). Compared to clinical stage Ia patients there was a trend towards worse outcomes for stage Ib patients which did not reach statistical significance ($p= 0.064$). The differences between stage Ia and stage IIa and IIb patients were more marked with HRs of 3.150 (95% CI 1.730 to 5.733, $p< 0.0005$) and 2.223 (95% CI 1.104 to 4.477, $p= 0.025$) respectively. T stage showed a trend towards association with DSS ($p= 0.084$). and univariable analysis with Cox's proportional hazards model shows that T2a and T2b patients had significantly poorer DSS than T1a patients; HR for T2a was 2.157 (95% CI 1.129 to 4.123 $p= 0.020$) and for T2b 2.617 (95% CI 1.253 to 5.68 $p= 0.010$). T3 patients did not have significantly poorer outcomes which may have been affected by the small number of patients in this group (n=16). Pleural involvement showed no association with DSS ($p= 0.875$) (figure 6.1).

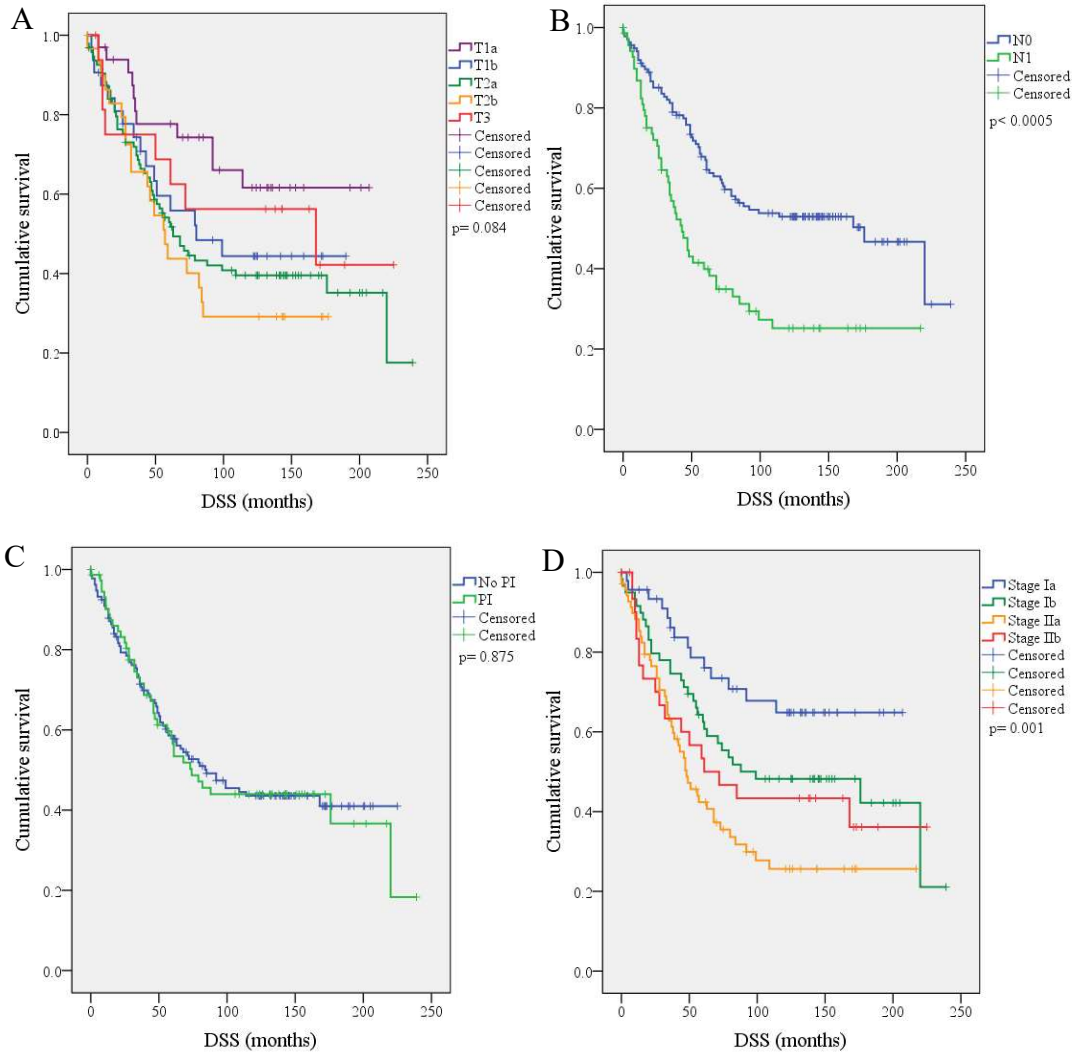


Figure 6.1: Kaplan-Meier's survival curves showing the associations between disease specific survival and T stage (A), N stage (B), pleural involvement (PI) (C) and clinical stage (D). Significance, calculated using the Mantel-Cox log-rank test, are shown to the right of each graph.

Spearman's rank correlation showed that the lesion size was inversely correlated with DSS; indicating, not surprisingly, that as the size of the tumour increased the length of survival decreased (rho -0.192, p= 0.006).

5.2.1 Presence of individual growth patterns

Lepidic

CART analysis identified a threshold of 3.5% lepidic growth across the whole primary lesion; tumours with less than or equal to 3.5% lepidic growth were classified as negative for the lepidic pattern. Use of the 3.5% threshold, compared to any proportion of lepidic growth, was associated with improved significance in associations with

routine pathological characteristics (table 6.1). The presence of greater than 3.5% lepidic pattern was associated with smaller lesion size (p= 0.018) (figure 6.2), lower T stage (χ^2 test for trend (χ^2T) p= 0.016) and clinical stage (χ^2T p= 0.001) and N0 disease (p= 0.045). There was also a trend toward a lower frequency of pleural involvement in lepidic positive tumours (p= 0.066) and the proportion of lepidic growth was significantly lower in tumours with pleural involvement (p= 0.034) (figure 6.2).

Table 6.1: Summary of the associations between the presence of the lepidic growth pattern (at any proportion or using a 3.5% threshold) and T stage, N stage, pleural involvement and clinical stage. Significance was estimated using χ^2 (and χ^2T).

		Lepidic pattern (at any proportion)			Lepidic pattern (3.5% threshold)			
		Not present	Present	Significance	≤3.5%	>3.5%	Significance	Total
T stage				p= 0.271 (χ^2T p= 0.067)			p= 0.072 (χ^2T p= 0.016)	
	T1a	12 (14.6%)	21 (16.8%)		14 (13.5%)	19 (18.4%)		33
	T1b	9 (11.0%)	23 (18.4%)		12 (11.5%)	20 (19.4%)		32
	T2a	37 (45.1%)	59 (47.2%)		47 (45.2%)	49 (47.6%)		96
	T2b	15 (18.3%)	15 (12.0%)		21 (20.2%)	9 (8.7%)		30
	T3	9 (11.0%)	7 (5.6%)		10 (9.6%)	6 (5.8%)		16
N stage				p= 0.703			p= 0.045	
	N0	53 (64.6%)	84 (67.2%)		62 (59.6%)	75 (72.8%)		137
	N1	29 (35.4%)	41 (32.8%)		42 (40.4%)	28 (27.2%)		70
Pleural involvement				p= 0.070			p= 0.066	
	Not involved	47 (57.3%)	87 (69.6%)		61 (58.7%)	73 (70.9%)		134
	Involved	35 (42.7%)	38 (30.4%)		43 (41.3%)	30 (29.1%)		73
Clinical stage				p= 0.070 (χ^2T p= 0.029)			p= 0.013 (χ^2T p= 0.001)	
	Ia	13 (15.9%)	33 (26.4%)		16 (15.4%)	30 (29.1%)		46
	Ib	25 (30.5%)	36 (28.8%)		27 (26.0%)	34 (33.0%)		61
	IIa	26 (31.7%)	43 (34.4%)		40 (38.5%)	29 (28.2%)		69
	IIb	18 (22.0%)	13 (10.4%)		21 (20.2%)	10 (9.7%)		31
Total		82 (100%)	125 (100%)		104 (100%)	103 (100%)		207

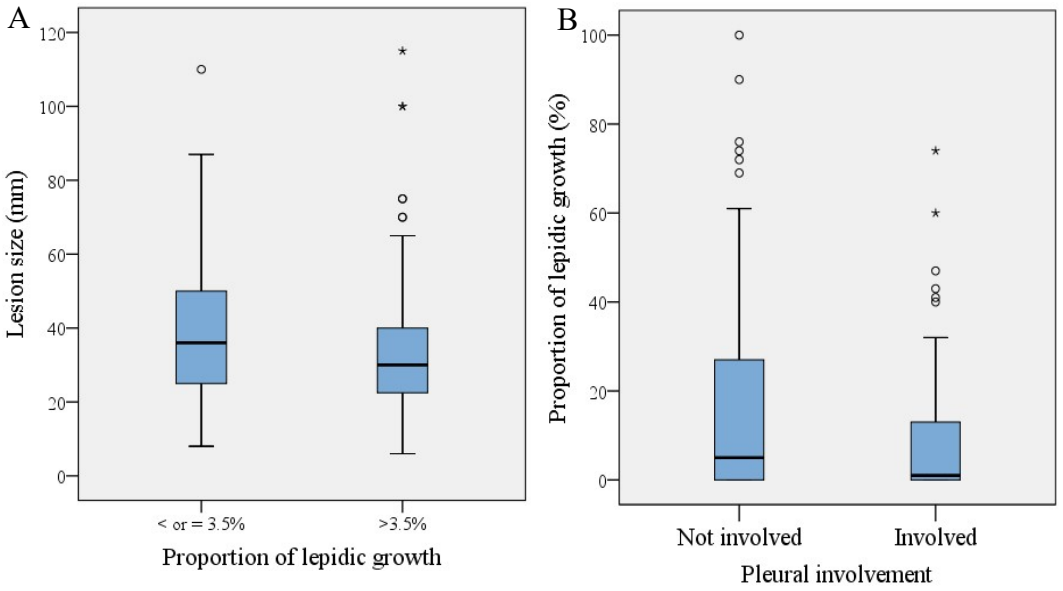


Figure 6.2: Boxplots showing statistically significant associations between the presence of the lepidic pattern (using a threshold of 3.5%) (p= 0.018) and lesion size (A) and the proportion of lepidic growth and pleural involvement (p= 0.034) (B). Significance was measured using the Mann-Whitney U test.

There was, however, no association between DSS and the presence of the lepidic pattern at any proportion or with a threshold of 3.5% (p= 0.864 and 0.687 respectively) (figure 6.3).

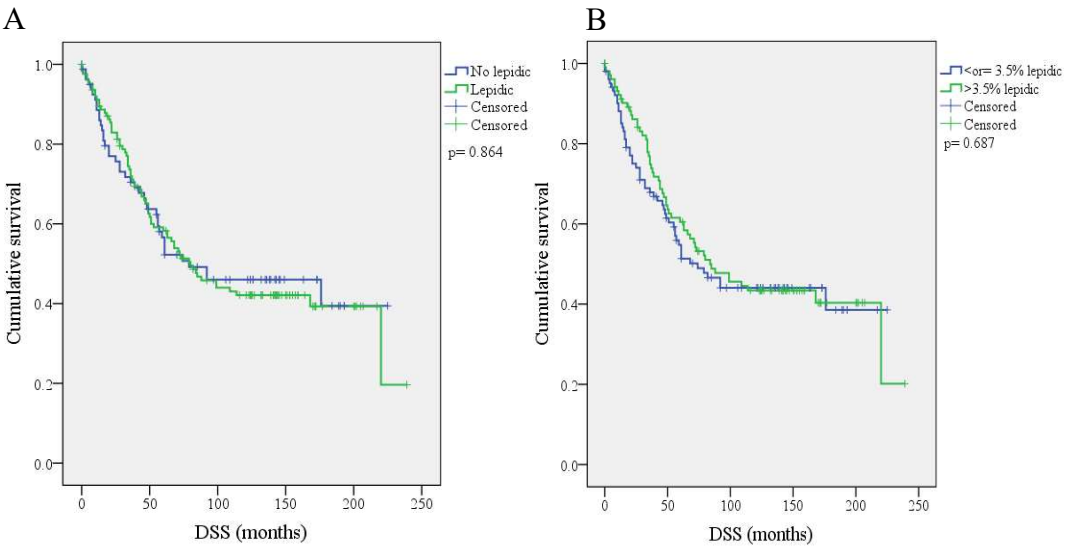


Figure 6.3: Kaplan-Meier's survival curves showing no significant difference in DSS between groups with and without the lepidic pattern at any proportion (A) or using 3.5% as a threshold (B). The statistical significance was assessed by the Mantel-Cox log rank test, and is shown to the right of each graph.

Papillary

CART analysis identified 8.5% papillary growth as a threshold above which lung adenocarcinoma patients had a more favourable prognosis. Use of the 8.5% papillary pattern cut-off led to small improvements in the significance of associations with nodal status and pleural involvement, but a decrease in the association with clinical stage and T stage; indicating that the 8.5% threshold for papillary growth may not have clinical utility (table 6.2). Tumours containing any amount of the papillary growth pattern were more likely to be a higher T stage (χ^2T p= 0.002) and larger size (p= 0.041) (table 6.2 and figure 6.4).

Table 6.2: Summary of the associations between the presence of the papillary growth pattern (at any level or with 8.5% as a threshold) and T stage, N stage, pleural involvement and clinical stage. Significance was estimated using χ^2 (and χ^2T).

		Papillary pattern (at any proportion)			Papillary pattern (8.5% cut off)			
		Not present	Present	Significance	≤8.5%	>8.5%	Significance	Total
T stage				p= 0.006 (χ^2T p= 0.002)			p= 0.228 (χ^2T p= 0.124)	
	1a	24 (26.1%)	9 (7.8%)		26 (20.5%)	7 (8.8%)		33
	1b	12 (13.0%)	20 (17.4%)		17 (13.4%)	15 (18.8%)		32
	2a	40 (43.5%)	56 (48.7%)		58 (45.7%)	38 (47.5%)		96
	2b	12 (13.0%)	18 (15.7%)		17 (13.4%)	13 (16.2%)		30
	3	4 (4.3%)	12 (10.4%)		9 (7.1%)	7 (8.8%)		16
N stage				p= 0.393			p= 0.127	
	N0	58 (63.0%)	79 (68.7%)		79 (62.2%)	58 (72.5%)		137
	N1	34 (37.0%)	36 (31.3%)		48 (37.8%)	22 (27.5%)		70
Pleural involvement				p= 0.672			p= 0.593	
	Not involved	61 (66.3%)	73 (63.5%)		84 (66.1%)	50 (62.5%)		134
	Involved	31 (33.7%)	42 (36.5%)		43 (33.9%)	30 (37.5%)		73
Clinical stage				p= 0.622 (χ^2T p= 0.341)			p= 0.729 (χ^2T p= 0.791)	
	Ia	23 (25.0%)	23 (20.0%)		29 (22.8%)	17 (21.2%)		46
	Ib	26 (28.3%)	35 (30.4%)		34 (26.8%)	27 (33.8%)		61
	IIa	32 (34.8%)	37 (32.2%)		45 (35.4%)	24 (30.0%)		69
	IIb	11 (12.0%)	20 (17.4%)		19 (15.0%)	12 (15.0%)		31
Total		92 (100%)	115 (100%)		127 (100%)	80 (100%)		207

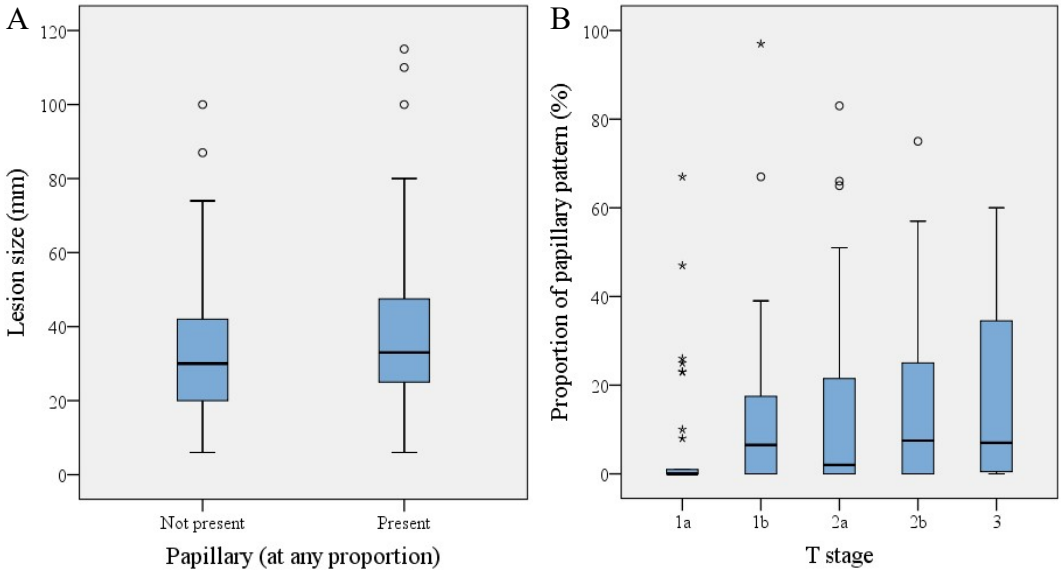


Figure 6.4: Box plots showing the relationship between the presence of the papillary pattern, at any proportion, and lesion size (Mann-Whitney U test p= 0.041) (A) and the proportion of papillary pattern and T stage (Kruskall Wallis test p= 0.030) (B).

In survival analysis; the use of an 8.5% threshold, compared to any proportion, for the presence of the papillary pattern showed an improved association with DSS with longer survival in patients with >8.5% papillary pattern (p= 0.035); HR, calculated by univariable Cox's proportional hazards analysis, was 0.657 (95% CI 0.443 to 0.976 p= 0.037) (figure 6.5).

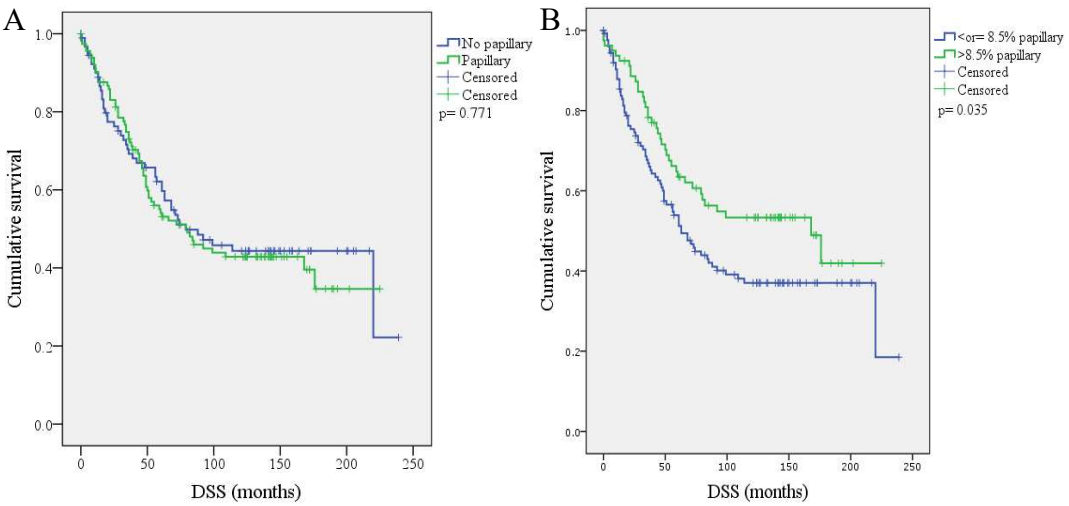


Figure 6.5: Kaplan-Meier's survival curves showing the relationship between disease specific survival and presence of any (A), or >8.5% (B), papillary growth pattern. The statistical significance of the relationship, as assessed by the Mantel-Cox log rank test, is shown to the right of each graph.

Acinar

CART analysis was unable to identify a clinically relevant threshold which could be applied to the proportion of acinar growth pattern based on DSS. The presence of the acinar growth pattern at any proportion showed no significant associations with T stage, N stage, pleural involvement or clinical stage (table 6.3).

Table 6.3: Summary of the relationships between the presence of the acinar growth pattern at any proportion and T stage, N stage, pleural involvement and clinical stage. Significance was calculated using the χ^2 test (and χ^2T) unless otherwise indicated.

		Acinar pattern (at any proportion)			
		Not present	Present	Total	Significance
T stage					p= 0.926[a] (χ^2T p= 1.0)
	1a	3 (14.3%)	30 (16.1%)	33	
	1b	4 (19.0%)	28 (15.1%)	32	
	2a	9 (42.9%)	87 (46.8%)	96	
	2b	4 (19.0%)	26 (14.0%)	30	
	3	1 (4.8%)	15 (8.1%)	16	
N stage					p= 0.662
	N0	13 (61.9%)	124 (66.7%)	137	
	N1	8 (38.1%)	62 (33.3%)	70	
Pleural involvement					p= 0.246
	Not involved	16 (76.2%)	118 (63.4%)	134	
	Involved	5 (23.8%)	68 (36.6%)	73	
Clinical stage					p= 0.892[a] (χ^2T p= 0.810)
	Ia	4 (19.0%)	42 (22.6%)	46	
	Ib	7 (33.3%)	54 (29.0%)	61	
	IIa	6 (28.6%)	63 (33.9%)	69	
	IIb	4 (19.0%)	27 (14.5%)	31	
Total		21 (100%)	186 (100%)	207	

[a]*Fisher's exact test estimated using the Monte Carlo simulation.*

Although these characteristics showed no correlation with the presence of the acinar pattern, survival analysis showed that patients with any acinar growth pattern in their tumours tended to have poorer DSS, however this fell short of statistical significance (p= 0.054) (figure 6.6); which may be due to the small number of tumours without the acinar pattern (n=21).

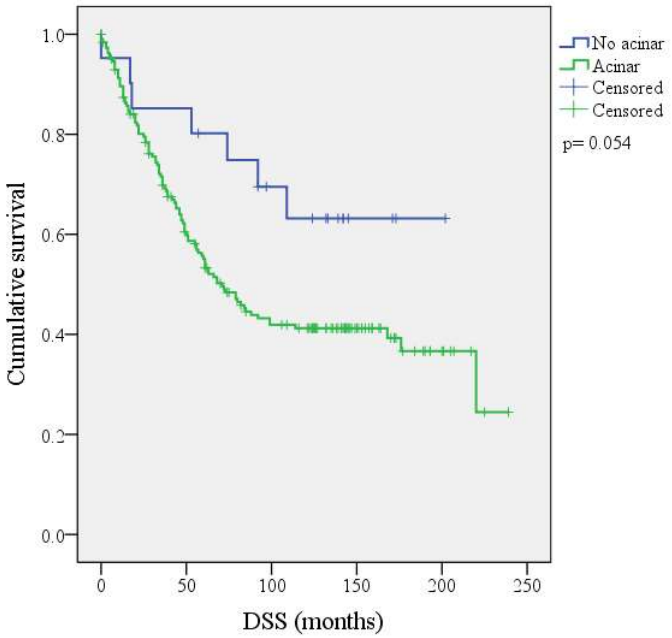


Figure 6.6: Kaplan-Meier's survival curves showing a trend towards an association between DSS and the presence of the acinar growth pattern at any proportion. Statistical significance, calculated using the Mantel-Cox log rank test, is shown to the right of the graph.

Solid

CART analysis revealed no viable threshold that could be applied to the proportion of the solid growth pattern in a patient's tumour based on DSS. There were no statistically significant correlations between the presence of any solid growth and lesion size, T stage, N stage, pleural involvement or clinical stage (table 6.4).

Table 6.4: Summary of the relationships between the presence of the solid growth pattern at any proportion and T stage, N stage, pleural involvement and clinical stage. Significance was calculated using the χ^2 test (and χ^2T) unless otherwise indicated.

		Solid pattern (at any proportion)			
		Not present	Present	Total	Significance
T stage					p= 0.695 (χ^2T p= 0.657)
	1a	12 (16.9%)	21 (15.4%)	33	
	1b	11 (15.5%)	21 (15.4%)	32	
	2a	30 (42.3%)	66 (48.5%)	96	
	2b	10 (14.1%)	20 (14.7%)	30	
	3	8 (11.3%)	8 (5.9%)	16	
N stage					p= 0.534
	N0	49 (69.0%)	88 (64.7%)	137	
	N1	22 (31.0%)	48 (35.3%)	70	
Pleural involvement					p= 0.750
	Not involved	47 (66.2%)	87 (64.0%)	134	
	Involved	24 (33.8%)	49 (36.0%)	73	
Clinical stage					p= 0.792 (χ^2T p= 0.786)
	Ia	17 (23.9%)	29 (21.3%)	46	
	Ib	18 (25.4%)	43 (31.6%)	61	
	IIa	24 (33.8%)	45 (33.1%)	69	
	IIb	12 (16.9%)	19 (14.0%)	31	
Total		71 (100%)	136 (100%)	207	

Kaplan Meier's survival analysis showed no associations between the presence of the solid growth pattern, at any proportion, and DSS (p= 0.575) (fig. 6.7).

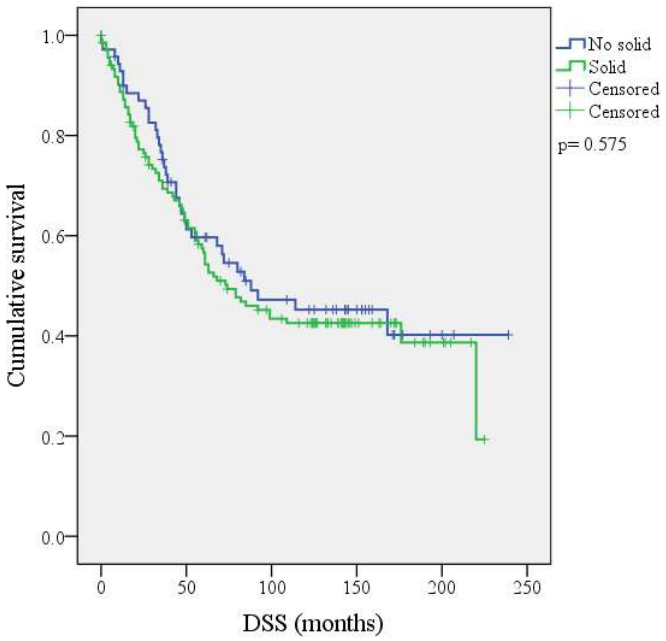


Figure 6.7: Kaplan-Meier's survival curves grouped on the basis of presence or absence of the solid growth pattern. Statistical significance, calculated using the Mantel-Cox log rank test, is shown to the right of the graph.

Micropapillary pattern

CART analysis was unable to define a threshold for the proportion of micropapillary pattern which could group patients according to their DSS. However, the presence of the micropapillary pattern (at any proportion) was significantly associated with a higher T stage (χ^2T p= 0.042), clinical stage (χ^2T p= 0.024), N1 disease (p= 0.015) and pleural involvement (p= 0.014) (table 6.5). The association with nodal status was further confirmed by the increased proportion of micropapillary pattern in patients with N1 disease (p= 0.004) (figure 6.8).

Table 6.5: Showing associations between the presence of the micropapillary pattern at any proportion with T stage, N stage, pleural involvement and clinical stage. The significance was assessed using the χ^2 test (and χ^2 T).

		Micropapillary pattern (at any proportion)			
		Not present	Present	Total	Significance
T stage					p= 0.031 (χ^2T p= 0.042)
	1a	20 (26.0%)	13 (10%)	33	
	1b	9 (11.7%)	23 (17.7%)	32	
	2a	32 (41.6%)	64 (49.2%)	96	
	2b	12 (15.7%)	18 (13.8%)	30	
	3	4 (5.2%)	12 (9.2%)	16	
N stage					p= 0.015
	N0	59 (76.6%)	78 (60%)	137	
	N1	18 (23.4%)	52 (40%)	70	
Pleural involvement					p= 0.014
	No pleural invasion	58 (75.3%)	76 (58.5%)	64	
	Pleural invasion	19 (24.7%)	54 (41.5%)	73	
Clinical stage					p= 0.161 (χ^2T p= 0.024)
	Ia	22 (28.6%)	24 (18.5%)	46	
	Ib	25 (32.5%)	36 (27.7%)	61	
	IIa	22 (28.6%)	47 (36.2%)	69	
	IIb	8 (10.4%)	23 (17.7%)	31	
Total		77 (100%)	130 (100%)	207	

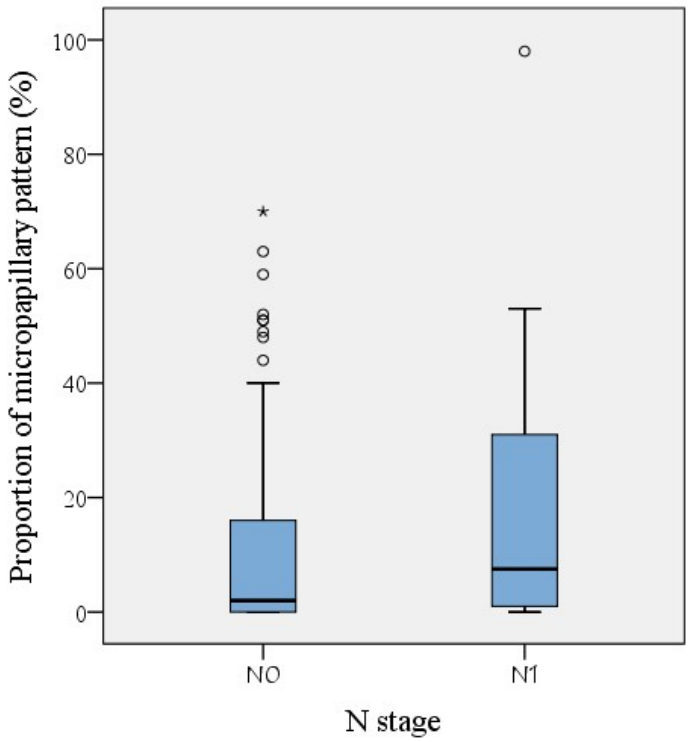


Figure 6.8: Box plot showing the association between nodal status and the proportion of micropapillary growth pattern. Statistical significance was calculated using the Mann-Whitney U test (p= 0.004).

Despite these seemingly prognostic associations, there was no significant difference in DSS between patients with and without the micropapillary growth pattern in their tumours (p= 0.240) (figure 6.9).

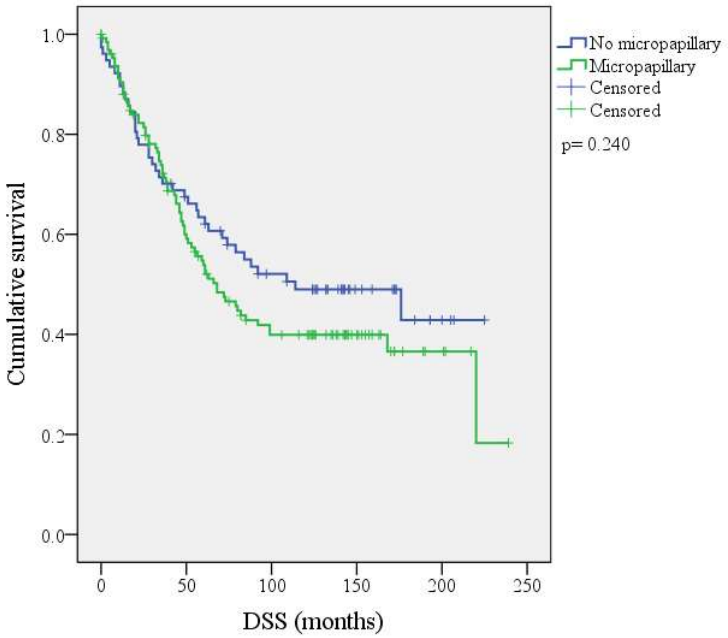


Figure 6.9: Kaplan Meier's survival curves for patients with or without the micropapillary growth pattern in their tumours (at any proportion). Statistical significance, calculated using the Mantel-Cox log rank test, is shown to the right of the graph.

5.2.2 Predominant growth pattern

Predominant growth pattern was significantly associated with nodal status (p= 0.021); whereas, there was no significant association between predominant growth pattern and T stage, pleural involvement or clinical stage (χ^2T p= 0.159, χ2 0.208 and χ^2T 0.710 respectively) (table 6.6).

Table 6.6: Details of the relationships between predominant growth pattern and T stage, N stage, pleural involvement and clinical stage. Significance was calculated using the χ^2 test (and χ^2T) unless otherwise indicated.

		Predominant pattern					Total	Significance
		Lepidic	Papillary	Acinar	Solid	Micropapillary		
T stage								p= 0.479[a] (χ^2T p= 0.159)
	1a	4 (20.0%)	1 (3.8%)	15 (19.7%)	9 (15.0%)	4 (16.0%)	33	
	1b	3 (15.0%)	3 (11.5%)	10 (13.2%)	8 (13.3%)	8 (32.0%)	32	
	2a	9 (45.0%)	14 (53.8%)	31 (40.8%)	33 (55.0%)	9 (36.0%)	96	
	2b	2 (10.0%)	4 (15.4%)	13 (17.1%)	8 (13.3%)	3 (12.0%)	30	
	3	2 (10.0%)	4 (15.4%)	7 (9.2%)	2 (3.3%)	1 (4.0%)	16	
N stage								p= 0.021
	N0	14 (70.0%)	20 (76.9%)	56 (73.7%)	37 (61.7%)	10 (40.0%)	137	
	N1	6 (30.0%)	6 (23.1%)	20 (26.3%)	23 (38.3%)	15 (60.0%)	70	
Pleural involvement								p= 0.208
	Not involved	16 (80.0%)	16 (61.5%)	46 (60.5%)	36 (60.0%)	20 (80.0%)	134	
	Involved	4 (20.0%)	10 (38.5%)	30 (39.5%)	24 (40.0%)	5 (20.0%)	73	
Clinical stage								p= 0.167 (χ^2T p= 0.710)
	Ia	5 (25.0%)	3 (11.5%)	23 (30.3%)	8 (13.3%)	7 (28.0%)	46	
	Ib	6 (30.0%)	10 (38.5%)	19 (25.0%)	24 (40.0%)	2 (8.0%)	61	
	IIa	6 (30.0%)	8 (30.8%)	22 (28.9%)	21 (35.0%)	12 (48.0%)	69	
	IIb	3 (15.0%)	5 (19.2%)	12 (15.8%)	7 (11.7%)	4 (16.0%)	31	
Total		20 (100%)	26 (100%)	76 (100%)	60 (100%)	25 (100%)	207	

[a]*Fisher's exact test approximated using the Monte Carlo simulation*

Overall there was no significant difference in DSS based on predominant growth pattern (p= 0.285) (figure 6.10). However, univariable Cox's proportional hazards analysis showed that patients with predominantly papillary tumours tended to have a lower DSS than those with predominantly micropapillary tumours; with an HR of 0.457 (95% CI 0.200 to 1.045 p= 0.064). There was no evidence of a significant difference between patients with predominantly lepidic, acinar, solid or micropapillary tumours. Patients with predominantly lepidic tumours were shown to have unexpectedly poor prognoses; however, of the 20 tumours classified as predominantly lepidic, 7 lived for at least 10 years and 1 died 0 months after their resection suggesting this was as a result of complications post-surgery rather than recurrence of their lung cancer. Of the remaining 12 patients, 8 had involved lymph nodes and/or pleura which would indicate an increased chance of incomplete removal of the disease and therefore lead to a poorer prognosis post resection.

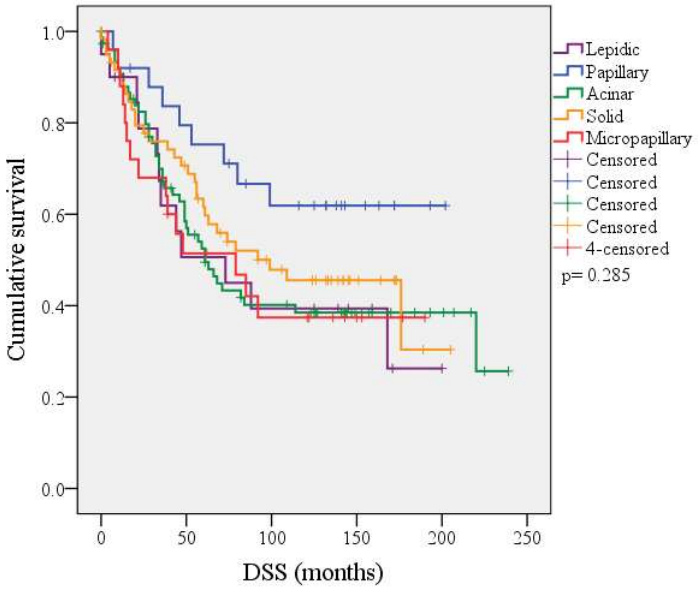


Figure 6.10: Kaplan Meier's curves for patients with different predominant growth patterns in their tumours. Statistical significance, calculated using the Mantel-Cox log rank test, is shown to the right of the graph.

Given the relatively small size of the cohort it is possible that the number of patients in each group was too small to show significant trends. Therefore; each lesion was also classified by predominant morphological grade whereby predominantly lepidic tumours were considered low grade, predominantly acinar or papillary tumours were considered intermediate grade and solid or micropapillary predominant tumours were

considered high grade.[14] Predominantly high grade lesions were more likely than low or intermediate grade to have involved lymph nodes (p= 0.020). However, there was no significant association between predominant morphological grade and lesion size, T stage, pleural involvement or clinical stage (table 6.7).

Table 6.7: The relationship between predominant morphological grade and T stage, N stage, pleural involvement and clinical stage. Significance was calculated using χ^2 test (and χ^2T).

		Morphological grade			Total	Significance
		Low	Intermediate	High		
T stage						p= 0.643 (χ^2T p= 0.404)
	1a	4 (20.0%)	16 (15.7%)	13 (15.3%)	33 (15.9%)	
	1b	3 (15.0%)	13 (12.7%)	16 (18.8%)	32 (15.5%)	
	2a	9 (45.0%)	45 (44.1%)	42 (49.4%)	96 (46.4%)	
	2b	2 (10.0%)	17 (16.7%)	11 (12.9%)	30 (14.5%)	
	3	2 (10.0%)	11 (10.8%)	3 (3.5%)	16 (7.7%)	
N stage						p= 0.020
	N0	14 (70.0%)	76 (74.5%)	47 (55.3%)	137 (66.2%)	
	N1	6 (30.0%)	26 (25.5%)	38 (44.7%)	70 (33.8%)	
Pleural involvement						p= 0.248
	Not involved	16 (80.0%)	62 (60.8%)	56 (65.9%)	134 (64.7%)	
	Involved	4 (20.0%)	40 (39.2%)	29 (34.1%)	73 (35.3%)	
Clinical stage						p= 0.787 (χ^2T p= 0.491)
	Ia	5 (25.0%)	26 (25.5%)	15 (17.6%)	46 (22.2%)	
	Ib	6 (30.0%)	29 (28.4%)	26 (30.6%)	61 (29.5%)	
	IIa	6 (30.0%)	30 (29.4%)	33 (38.8%)	69 (33.3%)	
	IIb	3 (15.0%)	17 (16.7%)	11 (12.9%)	31 (15.0%)	
Total		20 (100%)	102 (100%)	85 (100%)	207	

There was also no significant association between the predominant morphological grade and DSS (p= 0.714) (figure 6.11).

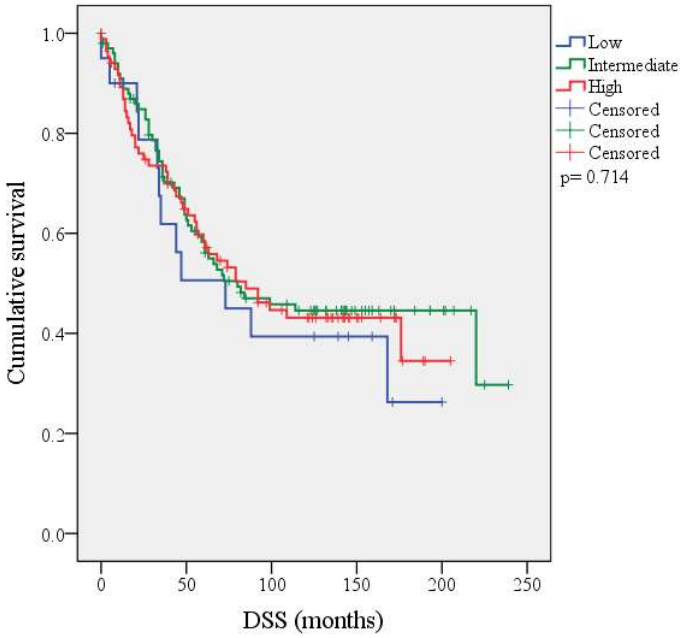


Figure 6.11: Kaplan-Meier's curves showing the relationship between the morphological grade of the predominant growth pattern and DSS; predominantly lepidic tumours were considered low grade, predominantly papillary or acinar tumours were intermediate grade and predominantly solid or micropapillary tumours were high grade. Statistical significance, calculated using the Mantel-Cox log rank test, is shown to the right of the graph.

The second most predominant growth pattern was also calculated and assessed for its utility as a predictor of clinical outcomes. Patients whose tumours had acinar or micropapillary as the second most common pattern were more likely to have involved lymph nodes (p= 0.047) and higher clinical stage disease (χ^2T p= 0.010) (table 6.8). However, survival analysis showed no significant relationship with DSS (p= 0.998) (figure 6.12).

Table 6.8: The relationship between second predominant pattern in each tumour and T stage, N stage, pleural involvement and clinical stage. Significance was calculated using χ^2 test (and χ^2T) unless otherwise indicated.

		Second predominant pattern					Total	Significance
		Lepidic	Papillary	Acinar	Solid	Micropapillary		
T stage								p= 0.195[a] (χ^2T p= 0.104)
	1a	7 (19.4%)	3 (10.3%)	10 (14.1%)	8 (23.5%)	3 (11.1%)	31 (15.7%)	
	1b	8 (22.2%)	5 (17.2%)	12 (16.9%)	5 (14.7%)	1 (3.7%)	31 (15.7%)	
	2a	16 (44.4%)	16 (55.2%)	37 (52.1%)	13 (38.2%)	11 (40.7%)	93 (47.2%)	
	2b	2 (5.6%)	2 (6.9%)	7 (9.9%)	6 (17.6%)	10 (37.0%)	27 (13.7%)	
	3	3 (8.3%)	3 (10.3%)	5 (7.0%)	2 (5.9%)	2 (7.4%)	15 (7.6%)	
N stage								p= 0.047
	N0	30 (83.3%)	22 (75.9%)	41 (57.7%)	23 (67.6%)	15 (55.6%)	131	
	N1	6 (16.7%)	7 (24.1%)	30 (42.3%)	11 (32.4%)	12 (44.4%)	66	
Pleural involvement								p= 0.491
	Not involved	25 (69.4%)	19 (65.5%)	46 (64.8%)	22 (64.7%)	13 (48.1%)	125 (63.5%)	
	Involved	11 (30.6%)	10 (34.5%)	25 (35.2%)	12 (35.3%)	14 (51.9%)	72 (36.5%)	
Clinical stage								p= 0.033 (χ^2T p= 0.010)
	Ia	11 (30.6%)	7 (24.1%)	12 (16.9%)	12 (35.3%)	3 (11.1%)	45	
	Ib	14 (38.9%)	11 (37.9%)	20 (28.2%)	6 (17.6%)	7 (25.9%)	58	
	IIa	8 (22.2%)	7 (24.1%)	31 (43.7%)	12 (35.3%)	8 (29.6%)	66	
	IIb	3 (8.3%)	4 (13.8%)	8 (11.3%)	4 (11.8%)	9 (33.3%)	28	
Total		36 (100%)	29 (100%)	71 (100%)	34 (100%)	27 (100%)	197	

[a]*Fisher's exact test estimated using the Monte Carlo simulation.*

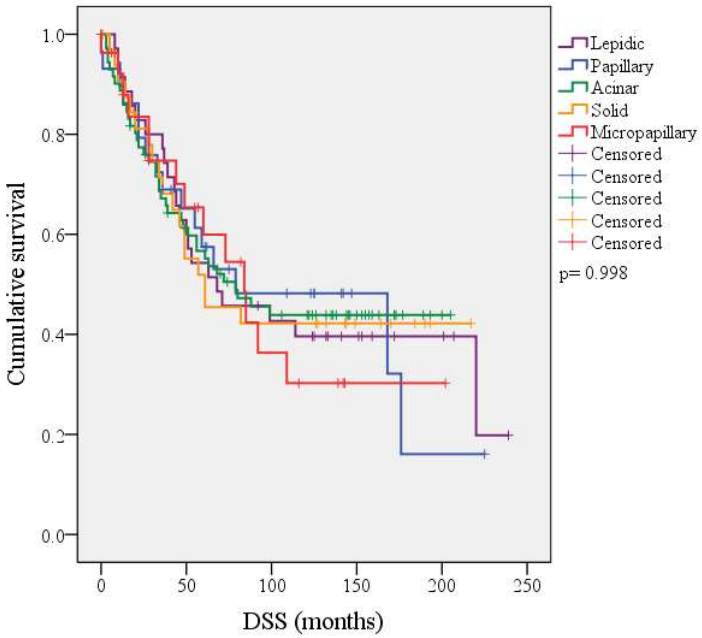


Figure 6.12: Kaplan-Meier's survival curves showing the relationship between DSS and the second predominant pattern. Statistical significance, calculated using the Mantel-Cox log rank test, is shown to the right of the graph.

There was a trend towards poorer outcomes in patients whose tumours had 3 or more growth patterns, but this did not reach statistical significance (p= 0.082). The proportion of each growth pattern did not show any significant correlations with DSS using Spearman's rank correlation (lepidic p= 0.393, acinar p= 0.459, papillary p= 0.253, solid p= 0.672, micropapillary p= 0.253).

5.2.3 Features relating to tumour grade

Necrosis

Large amounts of necrotic tissue were noted in 58 (30.2%) of the 192 cases which could be adequately assessed. Necrosis was more common in larger tumours (Mann-Whitney U test p= 0.002), those with a higher T stage (χ^2T p= 0.004) and those with a higher clinical stage (χ^2T p= 0.007); although, both T stage and clinical stage are related to tumour size (table 6.9).

Table 6.9: Associations between the presence of large amounts of necrosis and T stage and clinical stage. Significance was assessed using χ^2 test (and χ^2T).

		Necrosis			
		Not present	Present	Total	Significance
T stage					p= 0.016 (χ^2T p= 0.004)
	1a	29 (21.6%)	3 (5.2%)	32	
	1b	20 (14.9%)	11 (8.2%)	31	
	2a	66 (49.3%)	27 (46.6%)	93	
	2b	10 (7.5%)	10 (17.2%)	20	
	3	9 (6.7%)	7 (12.1%)	16	
Clinical stage					p= 0.022 (χ^2T p= 0.007)
	Ia	38 (28.4%)	7 (12.1%)	45	
	Ib	39 (29.1%)	19 (32.8%)	58	
	IIa	41 (30.6%)	17 (29.3%)	58	
	IIb	16 (11.9%)	15 (25.9%)	31	
Total		134 (100%)	58 (100%)	192	

Mitosis

A high mitotic rate was more common in tumours with higher T stage (χ^2T p= 0.047). Although the χ^2 test indicated a significant difference in clinical stage between tumours with and without frequent mitosis (p= 0.044), χ^2 test for trend indicated no consistent relationship with increasing clinical stage and therefore no significant association (p= 0.099) (table 6.10).

Table 6.10: Association between the presence of large numbers of mitotic figures and overall disease stage. Significance was assessed using the χ^2 test (and χ^2T).

		Few mitotic figures	Many mitotic figures	Total	Significance
T stage					p= 0.349 (χ^2T p= 0.047)
	T1a	18 (21.2%)	14 (13.2%)	32	
	T1b	15 (17.6%)	16 (15.1%)	31	
	T2a	41 (48.2%)	52 (49.1%)	93	
	T2b	6 (7.1%)	14 (13.2%)	20	
	T3	5 (5.9%)	10 (9.4%)	15	
Clinical stage					p= 0.044 (χ^2T p= 0.099)
	Ia	26 (30.6%)	19 (17.9%)	45	
	Ib	21 (24.7%)	37 (34.9%)	58	
	IIa	29 (34.1%)	29 (27.4%)	58	
	IIb	9 (10.6%)	21 (19.8%)	30	
Total		85 (100%)	106 (100%)	191	

Prominent nucleoli

The presence of large numbers of prominent nucleoli was more common in later clinical stage tumours (χ^2 test for trend p= 0.005) (table 6.11).

Table 6.11: Association between the presence of large numbers of prominent nucleoli, disease stage and the presence of the micropapillary pattern. Significance was assessed using χ^2 test (and χ^2T).

		Nucleoli			
		Inconspicuous	Prominent	Total	Significance
Clinical stage					p= 0.029 (χ^2T p= 0.005)
	Ia	32 (31.4%)	13 (14.4%)	45	
	Ib	30 (29.4%)	28 (31.1%)	58	
	IIa	28 (27.5%)	30 (33.3%)	58	
	IIb	12 (11.8%)	19 (31.7%)	31	
Total		102 (100%)	90 (100%)	192	

Presence of scar tissue

The presence of scar tissue in the tumours was more common in tumours from patients with N0 disease than those with involved lymph nodes (p= 0.022) and increasing clinical stage was associated with reduced incidence of tumour scar tissue (χ^2T p= 0.004) (table 6.12).

Table 6.12: Correlation of the presence of tumour scar tissue with nodal status, disease stage and predominant pattern. Significance was assessed using χ^2 test (and χ^2T).

		Scar tissue			
		Not present	Present	Total	Significance
N stage					p= 0.022
	N0	94 (60.3%)	29 (80.6%)	123	
	N1	62 (39.7%)	7 (24.1%)	69	
Disease stage					p= 0.014 (χ^2T p= 0.004)
	Ia	33 (21.2%)	12 (33.3%)	45	
	Ib	42 (26.9%)	16 (44.4%)	58	
	IIa	52 (33.3%)	6 (16.7%)	58	
	IIb	29 (18.6%)	2 (5.6%)	31	
Total		156 (100%)	36 100%)	192	

There were no statistically significant associations between the presence of large numbers of apoptotic bodies, cytological pleomorphism, or inflammatory infiltrate and T stage, N stage, pleural involvement or clinical stage.

Although there were no significant correlations between any tumour grading variables and disease specific survival, there was a trend towards poorer prognosis for tumours with large numbers of mitotic figures (HR 1.481, 95% CI 0.991 to 2.212 p= 0.055) (figures 6.13.1 and 6.13.2). In addition, the survival curves formed by Kaplan-Meier's analysis suggested a possible link with longer survival in patients whose tumours had no necrosis, inconspicuous nucleoli and a large number of apoptotic bodies.

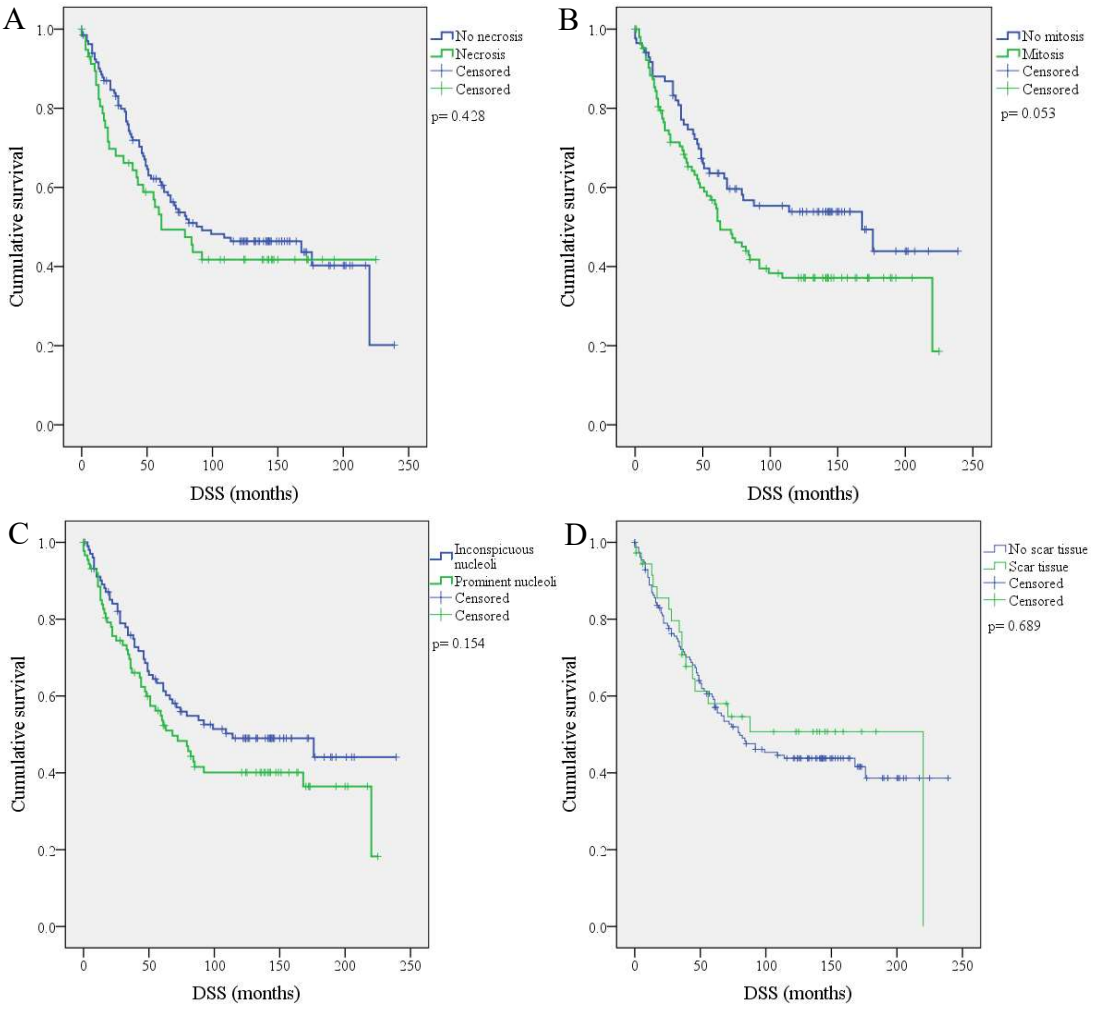


Figure 6.13.1: Kaplan-Meier's survival curves for features related to tumour grade: necrosis (A), mitosis (B), nucleoli (C) and the presence of scar tissue (D). The significance of each variable was assessed using the Mantel-Cox log rank test and shown to the right of each graph.

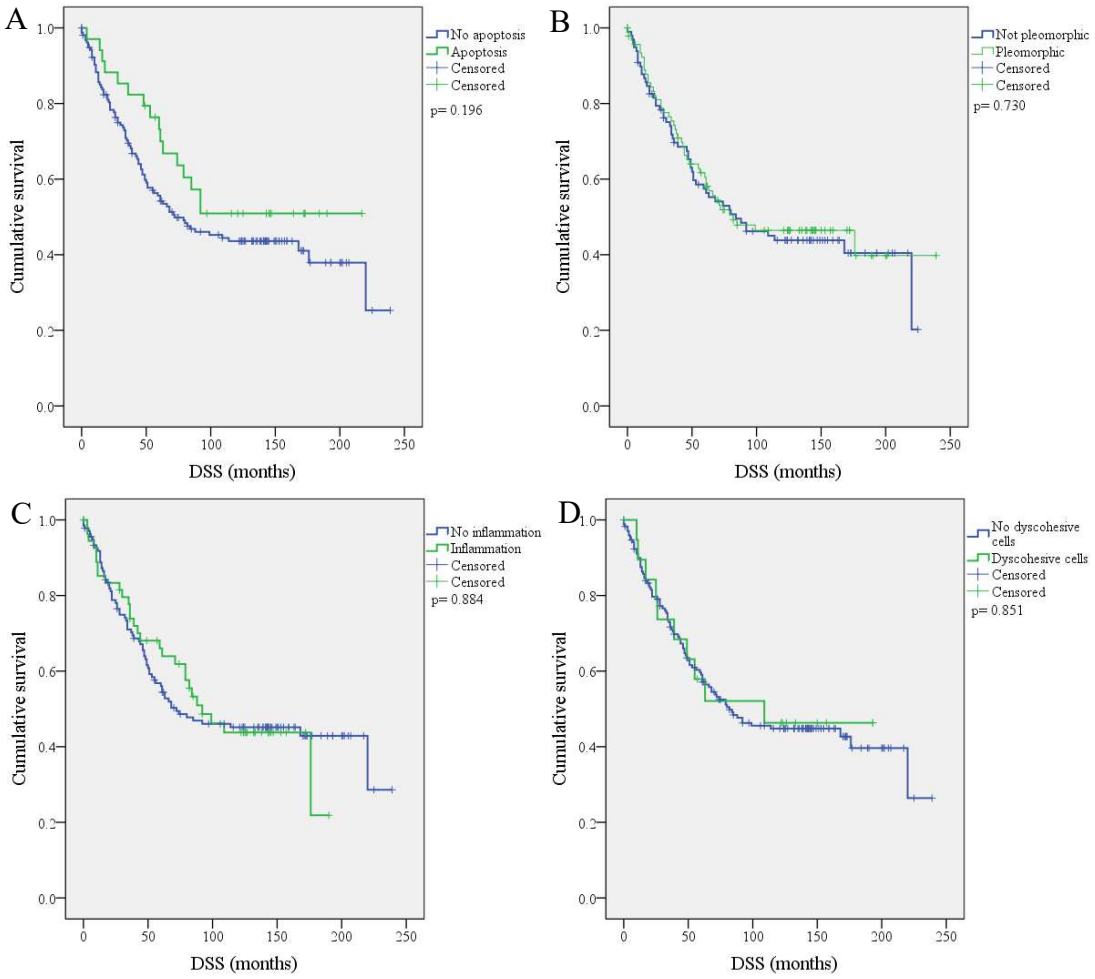


Figure 6.13.2: Kaplan-Meier's survival curves for features related to tumour grade: apoptosis (A), pleomorphism (B), inflammatory infiltrate (C) and dyscohesive tumour cells (D). The significance of each variable was assessed using the Mantel-Cox log rank test and shown to the right of each graph.

The combined results of categorical and Kaplan-Meier's survival analysis suggested that the presence of necrosis, mitosis and prominent nucleoli may have some association with tumours with poor prognostic features; whereas the presence of scar tissue, many apoptotic bodies or a high degree of inflammation may be more favourable characteristics. For each tumour the presence of favourable and poor prognostic traits was given a score of +1 and -1 respectively and the sum of scores in each tumour was calculated. Those with an overall tumour grade score greater than or equal to 0 were found to be associated with longer DSS (p= 0.008) HR 1.712 (95% CI 1.144 to 2.564 p= 0.009) (figure 6.14).

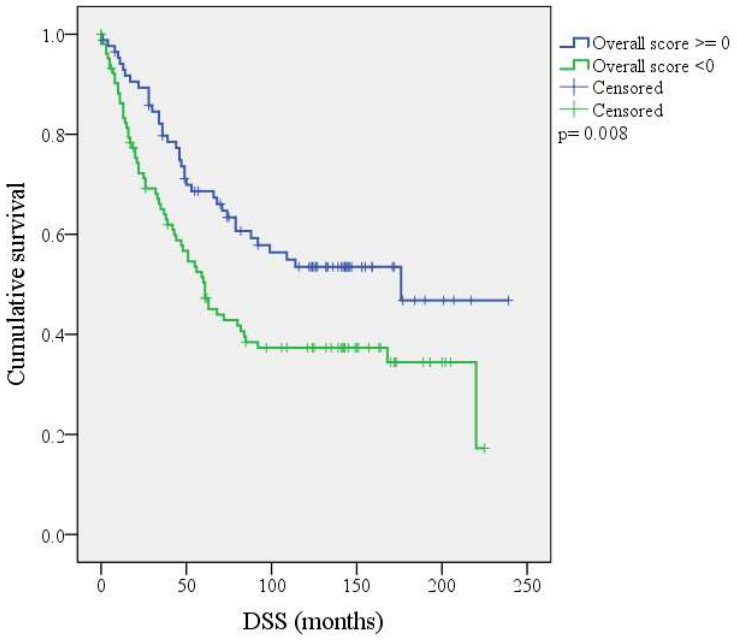


Figure 6.14: Kaplan-Meier's survival curves showing the relationship between the overall tumour grade score and DSS. Significance was assessed using the Mantel-Cox log rank test and shown to the right of the graph.

5.2.4 Molecular characteristics

EGFR and *KRAS* mutation status was only carried out in N1 cases and where the tissue was amenable to molecular analysis (n= 67); in this subset of patients neither *EGFR* or *KRAS* mutation status were significantly associated with lesion size, T stage, N stage, pleural involvement or clinical stage (tables 6.13 and 6.14).

Table 6.13: Summary of the associations between EGFR mutation status and routinely available clinicopathological variables. Significance was estimated using the χ^2 test (and χ^2T) unless otherwise indicated.

		EGFR mutation status			
		No mutation	Mutation	Total	Significance
T stage					p= 0.971 (χ^2T p= 0.887)[a]
	1a	6 (12.0%)	2 (11.8%)	8	
	1b	7 (14.0%)	3 (17.6%)	10	
	2a	26 (52.0%)	9 (52.9%)	35	
	2b	11 (22.0%)	3 (17.6%)	14	
	3	0	0	0	
Pleural involvement					p= 0.728
	No pleural invasion	33 (66.0%)	12 (70.6%)	45	
	Pleural invasion	17 (34.0%)	5 (29.4%)	22	
Clinical stage					p= 1.0 (χ^2T p= 0.750)[a]
	Ia	0	0	0	
	Ib	0	0	0	
	IIa	39 (78.0%)	14 (82.4%)	53	
	IIb	11 (22.0%)	3 (17.6%)	14	
Total		50 (100%)	17 (100%)	67	

[a]*Calculated using Fisher's exact test with the Monte Carlo simulation*

Table 6.14 Summary of the associations between KRAS mutation status and routinely available clinicopathological variables. Significance was estimated using the χ^2 test (and χ^2T) unless otherwise indicated.

		KRAS mutation status			
		No mutation	Mutation	Total	Significance
T stage					p= 0.951 (χ^2T p= 0.696)[a]
	1a	5 (14.7%)	3 (9.1%)	8	
	1b	5 (14.7%)	5 (15.2%)	10	
	2a	17 (50.0%)	18 (54.5%)	35	
	2b	7 (20.6%)	7 (21.2%)	14	
	3	0	0	0	
Pleural involvement					p= 0.339
	No pleural invasion	21 (61.8%)	24 (72.7%)	45	
	Pleural invasion	13 (38.2%)	9 (27.3%)	22	
Clinical stage					p= 0.950 (χ^2T p= 0.950)
	Ia	0	0	0	
	Ib	0	0	0	
	IIa	27 (79.4%)	26 (78.8%)	53	
	IIb	7 (20.6%)	7 (21.2%)	14	
Total		34 (100%)	33 (100%)	67	

[a]*Calculated using Fisher's exact test with the Monte Carlo simulation*

Survival analysis showed no significant relationship with *EGFR* or *KRAS* mutation status (p= 0.798 and 0.119 respectively); however, the number of patients with known mutation status was limited (n= 67) and none of the patients in the cohort were treated with EGFR TKIs (figure 6.15).

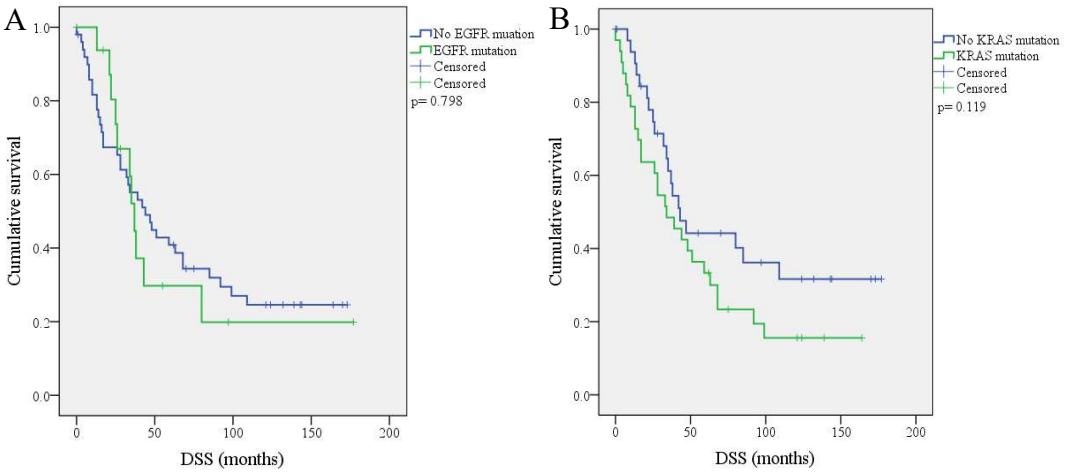


Figure 6.15: Kaplan-Meier's survival curves showing the relationship between EGFR (A) and KRAS (B) mutation status and DSS. Significance was assessed using the Mantel-Cox log-rank test.

Sections from a representative tumour block from each case were sent to Myriad Genetics Inc. for gene expression analysis of 31 cell cycle genes (see chapter 2); and the resulting CCP score and clinical stage used to calculate mPS. Of the 208 samples in the cohort, tissue blocks were unavailable for 2 samples and a further 16 samples failed gene expression analysis. In total 190 samples successfully produced a CCP score. Since mPS is a function of clinical stage, and therefore T and N stage, associations between these variables were not investigated. The CCP score was significantly associated with nodal status (Mann-Whitney U test p= 0.021) and clinical stage (Kruskall Wallis test p= 0.025) and there was a trend towards a relationship with T stage (Kruskall Wallis test p= 0.096) (table 6.15 and figure 6.16). Since pleural involvement was not significantly associated with CCP; the observed statistical significance between mPS and pleural involvement was likely to be a result of an association between pleural involvement and clinical stage.

Table 6.15: Details of CCP score and mPS for the cohort statistical significance was assessed using the non-parametric Kruskall-Wallis[a] or Mann-U Whitney[b] tests as appropriate. The mean (and 95% CI for the mean) are shown for each T stage, N stage, pleural involvement status and clinical stage as appropriate.

		CCP score	mPS
Number of patients		190	190
Median		0.20	32.3
Mean (95% CI)		0.11 (-0.03 to 0.24)	30.5 (28.6 to 32.4)
Range		-2.8 to 2.4	-3.5 to 54.8
T stage		p= 0.096[a]	n/a
	1a	-0.338 (-0.713 to 0.037)	
	1b	0.048 (-0.235 to 0.332	
	2a	0.216 (0.030 to 0.402)	
	2b	0.381 (-0.015 to 0.776)	
	3a	-0.007 (-0.657 to 0.644)	
N stage		p= 0.021[b]	n/a
	N0	-0.022 (95% -0.204 to 0.160)	
	N1	0.355 (0.185 to 0.524)	
Pleural involvement		p= 0.133[b]	p= 0.009[b]
	Not involved	0.006 (-0.171 to 0.182)	28.36 (25.96 to 30.76)
	Involved	0.290 (0.090 to 0.490)	34.25 (31.41 to 37.08)
Clinical stage		p= 0.025[a]	n/a
	Ia	-0.288 (-0.592 to 0.016)	
	Ib	0.089 (-0.159 to 0.338)	
	IIa	0.340 (0.124 to 0.555)	
	IIb	0.207 (-0.149 to 0.563)	

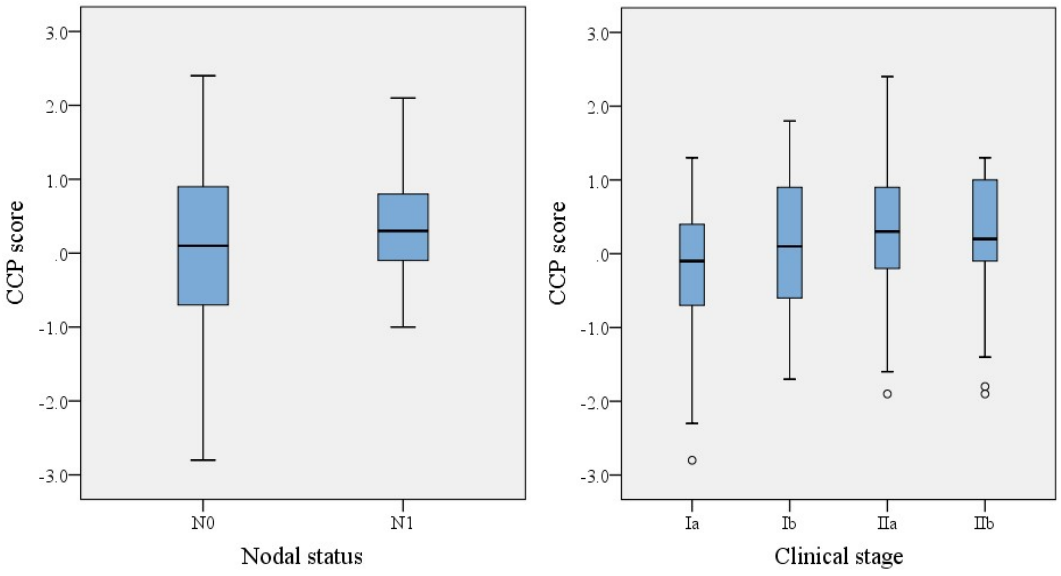


Figure 6.16: Box plots showing the difference in CCP score between tumours based on N stage (A) (Mann-U Whitey test p= 0.021) and clinical stage (B) (Kruskall-Wallis test p=0.025).

Multivariable analysis

Multivariable analysis of survival was carried out using Cox's proportional hazards model; since clinical stage is a factor of T stage and N stage only clinical stage was included in the analysis. All variables with a p-value of <0.05 were removed from the model and the analysis repeated. The final analysis included clinical stage ($p< 0.005$), the presence of the papillary pattern (with a cut-off of 8.5%) ($p= 0.048$) and the presence of large numbers of mitotic figures ($p= 0.029$) and apoptotic bodies ($p= 0.015$); however, after applying Bonferroni's correction for type I errors only clinical stage was considered to have a significant association with DSS.

Compared to patients with clinical stage IIb disease at resection, those with stage Ia had an HR of 0.418 (95% CI 0.204 to 0.855, $p= 0.017$), However there was no significant difference in stage Ib, IIa and IIb, which was previously demonstrated in the univariable analysis (figure 6.1).

5.3 Development of prognostic models

Decision tree analysis was carried out using classification and regression trees (CART) (with 10-fold cross validation with minimum parent and child node sizes of 20 and 10 respectively) to develop models to predict prognosis in lung adenocarcinoma patients. The initial model (the baseline model) was generated using only currently routinely available parameters (lesion size, T stage, N stage, pleural involvement and clinical stage). The decision tree created using only routinely available pathological characteristics (the baseline model) had, not surprisingly, strong similarities with the 7th TNM staging system and approximately separated clinical stage Ia from higher stages in the N0 branch; and stage IIa from higher stages in the N1 branch (figure 6.17).[5] The addition of *EGFR* and *KRAS* mutation status did not improve the model produced by CART analysis and these variables were not incorporated into the decision tree.

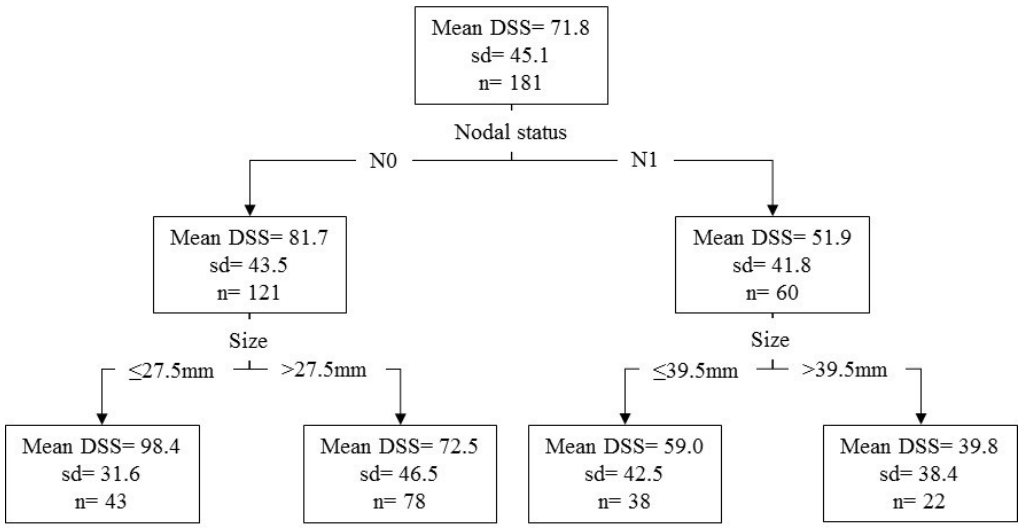


Figure 6.17: The baseline model. Decision tree produced using classification and regression tree analysis with lesion size, T stage, N stage, pleural involvement and clinical stage as potential predictive variables of DSS (months).

In a second iteration of CART analysis a decision tree was constructed using routinely available histopathological parameters with the addition of: the presence or absence of each growth pattern (with thresholds of 3.5% and 8.5% for lepidic and papillary respectively and any proportion for acinar, solid and micropapillary growth), the proportion of each growth pattern, predominant growth pattern, predominant morphological grade, second predominant growth pattern and the total number of growth patterns per case. N stage and lesion size continued to contribute significantly to the model; however, the addition of morphological variables enabled further refinement (figure 6.18) (The growth pattern model). In node positive patients a large proportion of solid tumour growth was associated with a better prognosis. In node negative patients with smaller tumours (≤27.5mm); those with lepidic, papillary or acinar growth as the second most predominant pattern had longer DSS. Other variables, including predominant growth pattern, were not found to contribute significantly to the model.

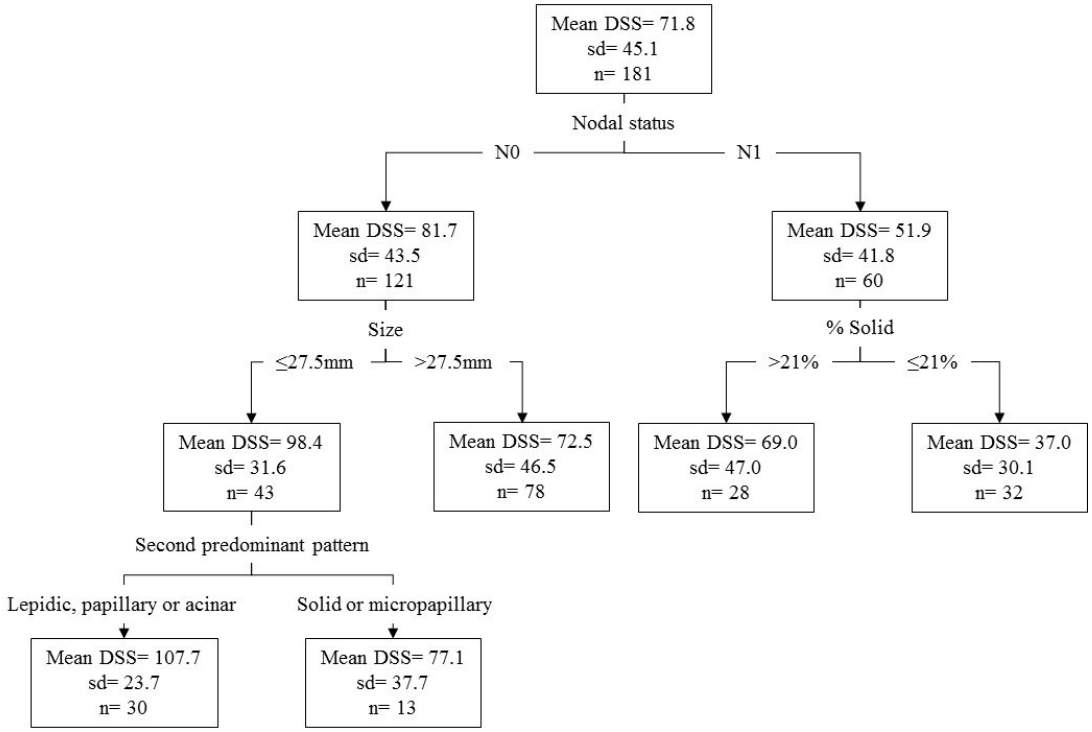


Figure 6.18: The growth pattern model. Decision tree created by CART analysis using routinely available parameters as well as additional morphological variables identified after growth pattern analysis (predominant growth pattern; predominant morphological grade; the second predominant pattern; the presence or absence of the lepidic, papillary, acinar, solid or micropapillary patterns; the proportion of each growth pattern and the number of patterns in each case).

CART analysis was then repeated with the addition of further morphological features which reflect tumour grade including the presence of: necrosis, scar tissue, cytological pleomorphism, dyscohesion between the tumour cells and large numbers of mitotic figures, apoptotic bodies, prominent nucleoli and inflammatory cells; as well as the overall tumour grade score (section 6.2.3) (the growth pattern plus tumour grade model). The resulting tree remained very similar to the Growth pattern model with only the presence of large numbers of apoptotic bodies incorporated which identified a small subset of node positive patients with an improved DSS (figure 6.19).

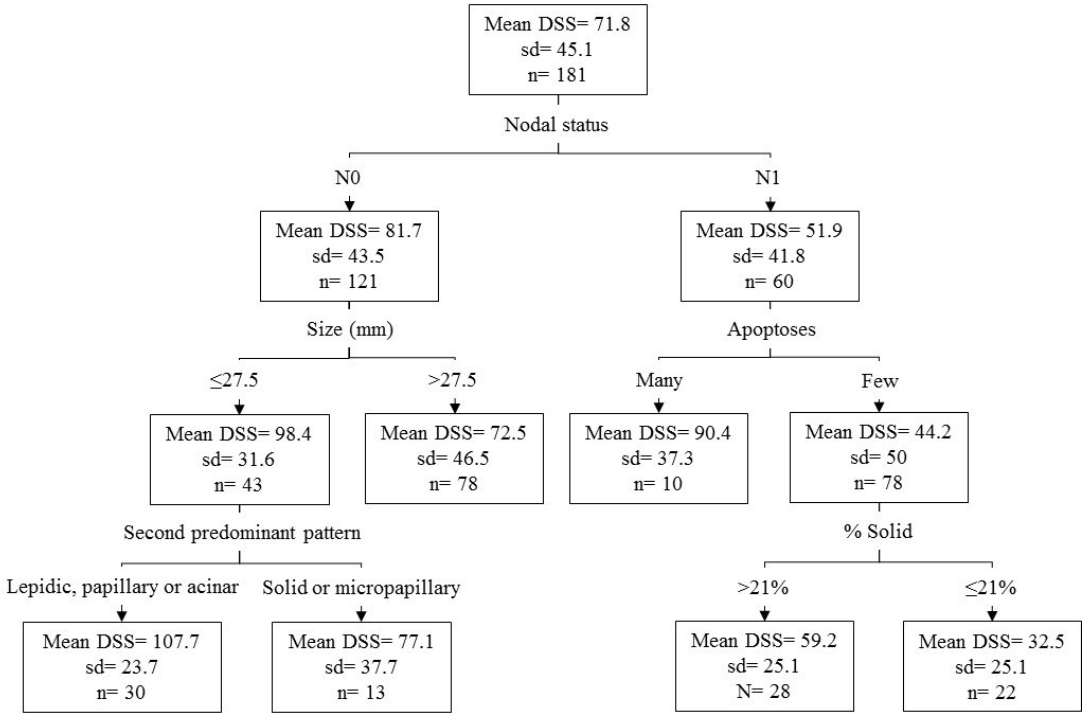


Figure 6.19: Growth pattern plus grade model. Decision tree created by CART analysis using routinely available parameters, morphological variables identified after growth pattern analysis and features related to tumour grade (necrosis, scar tissue, cytological pleomorphism, dyscohesion between the tumour cells and large numbers of mitotic figures, apoptotic bodies, prominent nucleoli and inflammatory cells; as well as the overall tumour grade score).

Finally, CART analysis was carried out with all available variables including the proliferation score mPS which greatly altered the resulting decision tree (figure 6.20). Nodal status was replaced by mPS to form the first branch of the tree with a threshold 26.39. Lesion size did not significantly contribute to the model. For tumours with larger mPS values the absence of large numbers of apoptotic bodies remained a poor prognostic group within which tumours with a large proportion of the micropapillary pattern (greater than 30.5%) had the poorest outcomes. Tumours with lower mPS values were further divided based on the second predominant pattern whereby those with lepidic, acinar or micropapillary pattern had improved DSS compared to those with solid or papillary pattern as their second most common pattern. However, this grouping of second predominant patterns conflicts with that produced in previous decision trees.

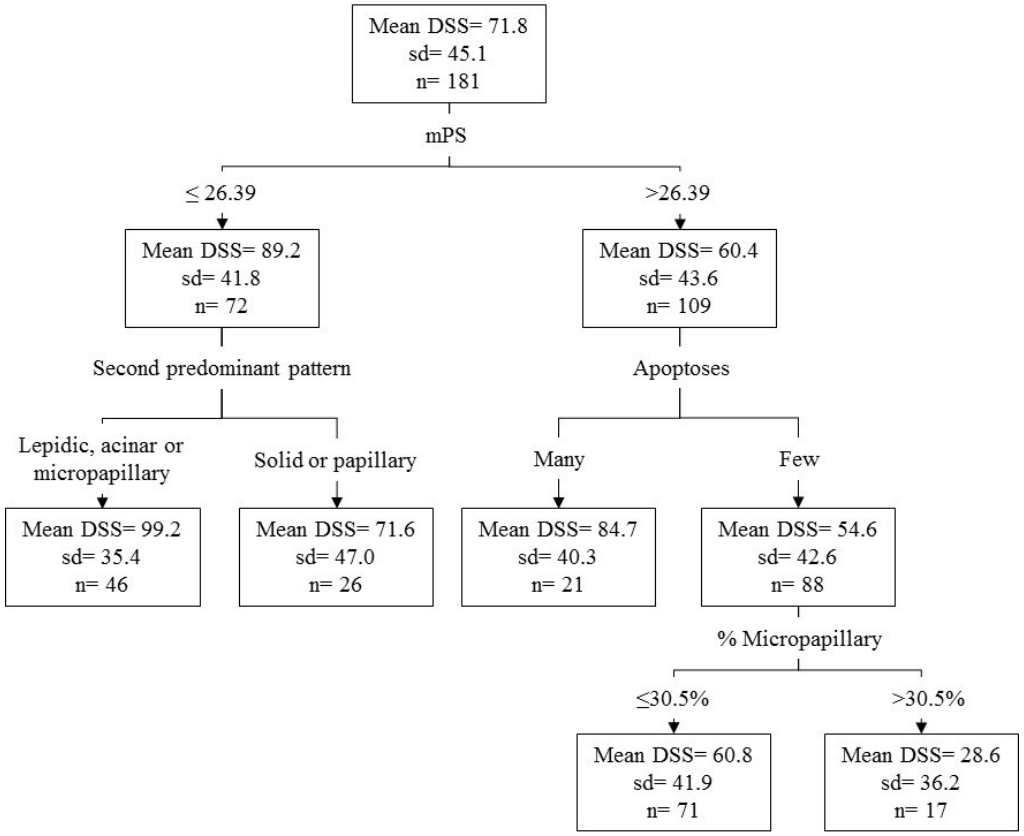


Figure 6.20: mPS model. Decision tree created by CART analysis using all available pathological, morphological and molecular characteristics including mPS.

In order to compare the performance of each model for predicting poor prognosis; the lung cancer related 1-, 2- and 5-year survival of patients was used to calculate the area under the ROC curve (the probability of accurately predicting prognosis - otherwise known as c-statistic) with each decision tree. The c-statistic was also calculated for the currently used classification based on clinical stage whereby patients with stage IIa and IIb, those who would be eligible for adjuvant chemotherapy, were considered the worst prognostic group. The use of 10-year survival reduced the c-statistic for all models and this was, therefore, not included in the analysis (table 6.15).

In this cohort, clinical stage IIa or IIb alone was a poor predictor of prognosis and only 5-year survival data was able to produce a probability significantly better than chance. Compared to clinical stage alone the baseline model, which utilized only nodal status and lesion size showed an improvement in the probability of successfully identifying a group of patients with the poorest prognoses. The inclusion of growth pattern or

grading data produced very little improvement in this probability for 1- and 2-year survival compared to baseline. The mPS model, which included transcriptomics data showed the greatest improvement being the only model classified as good (c-statistic > 0.7) for the prediction of patients who would not survive 1- or 2-years. None of the models would be classified as strong predictors (c-statistic >0.8).

Table 6.16: The area under the ROC curve (c-statistic) of each model in predicting 1-, 2- and 5-year disease specific survival.

		Model				
		Current clinical model	Baseline	Growth pattern	Growth pattern plus grade	mPS
		Clinical stage II	N1 and size >39.5mm	N1, <21% solid	N1, few apoptotic bodies, <21% solid	mPS>26.39, few apoptotic bodies, >30.5% micropapillary
Mean DSS (sd)		59.9 (43.0)	39.8 (38.4)	37.0 (30.1)	32.5 (25.1)	28.6 (36.2)
1-year	Area under the curve	0.611	0.647	0.650	0.663	0.747
	95% CI	0.482 to 0.741	0.543 to 0.751	0.544 to 0.755	0.555 to 0.770	0.625 to 0.868
2-year	Area under the curve	0.581	0.677	0.681	0.688	0.753
	95% CI	0.482 to 0.680	0.597 to 0.758	0.600 to 0.763	0.607 to 0.770	0.664 to 0.842
5-year	Area under the curve	0.631	0.704	0.725	0.736	0.749
	95% CI	0.552 to 0.711	0.631 to 0.777	0.654 to 0.796	0.666 to 0.806	0.673 to 0.825

Given that CART analysis replaced nodal status with mPS, indicating it to be a better predictor of survival, the c-statistic of N stage and mPS alone (N1 versus mPS>26.39) were also compared (table 6.16). The mPS was a better predictor of 1- and 2-year survival than nodal status, however these was no difference in 5-year survival.

Table 6.17: The area under the ROC curve (c-statistic) of nodal status and mPS as single predictors of prognosis based on 1-, 2- and 5-year survival.

		Nodal status	mPS (threshold 26.39)
1-year survival	Area under the curve	0.577	0.642
	95% CI	0.438 to 0.715	0.512 to 0.771
2-year survival	Area under the curve	0.595	0.615
	95% CI	0.493 to 0.696	0.515 to 0.714
5-year survival	Area under the curve	0.637	0.637
	95% CI	0.557 to 0.717	0.555 to 0.720

5.4 Discussion

The lung cancer related 5-year survival rate for this cohort of resected stage I and II patients was 57%; 9% of patients died from their disease within 1 year of their resection. In order to improve these survival rates it is likely that additional treatment, such as adjuvant chemotherapy, may be required for a proportion of patients with more aggressive disease. However patient stratification tools are required to identify those who would be likely to benefit from further therapy. Morphological analysis has been suggested as a method of prognostication and represents an inexpensive, easily obtained, source of additional data. Several published reports have suggested that the presence of the solid or micropapillary growth patterns were associated with poorer outcomes for lung adenocarcinoma patients; [17-21, 23-26, 28-30] however, not all studies included multivariable analysis and therefore the effect of confounding variables is not known.[18-21, 24, 25, 28-30] In addition, some studies failed to find any significant relationships.[22, 26] The prognostic value of predominant growth pattern has been supported in many published articles with predominantly lepidic tumours associated with favourable patient prognoses.[15, 29, 33, 34, 38-40, 48] whereas predominantly solid and micropapillary tumours have generally been considered to have a poor prognosis.[15, 29, 33, 35, 36, 38-40, 43, 49] Although the robustness of statistical analyses carried out in many of

these studies can be criticised,[15, 29, 33, 36, 38, 40, 43, 46, 48] and there are discrepancies between studies in the main conclusions drawn.[26, 29, 59] Some reports have grouped tumours into low, intermediate or high grade based on their morphology alone in order to show a significant association with prognosis.[15, 42, 44, 45, 48, 53] However there is even inconsistency in this grouping strategy as several reports classified predominantly lepidic, acinar and papillary as an intermediate group with solid and micropapillary as high grade;[15, 34, 41, 42, 48, 53] whereas others have classified lepidic as low grade, papillary and acinar as intermediate and solid and micropapillary as high grade.[40, 239]

Detailed morphological and molecular analysis of 208 resected clinical stage I and II lung adenocarcinomas has been used to investigate relationships between pathological (size, nodal status, clinical stage), molecular genetic and transcriptomic parameters with survival. Not surprisingly; lesion size, N stage and clinical stage were strongly associated with disease specific survival. These fundamental characteristics have formed the basis of clinical decision making for lung cancer patients for many years and their relationship to prognosis is widely accepted.[5] However; pleural involvement, which has previously been shown to correlate with poorer prognoses, showed no value as a prognostic marker in this cohort.[67, 299-301] The presence of individual growth patterns showed some association with these pathological characteristics; the lepidic pattern was more common in smaller tumours and in patients with earlier stage disease and the papillary pattern in larger tumours and the micropapillary pattern in more advanced disease. Only the presence of the papillary pattern was shown to be associated with disease specific survival in univariable analysis; however, even this fell short of significance in multivariable analysis after correction for type I errors. Similarly, the predominant growth pattern, whether grouped into morphological grade or not, showed a limited association with pathological features, with micropapillary or solid tumours more likely to have involved lymph nodes; however, there was no association with survival. It is noted that predominantly lepidic tumours in this cohort had a poor prognosis compared to previously published data; [15, 29, 33, 34, 38-40, 48] most likely as a result of small group size (only 20 patients) and a high proportion of patients with additional poor prognostic factors such as involved lymph nodes and pleura.[14, 34, 301] This suggests that, although in large cohorts trends in survival may be observed,

simply relying on one tumour characteristic to predict individual patient outcomes is likely to be an over simplification. Indeed; there is evidence to suggest that the contribution of non-predominant, but aggressive, growth patterns also has an influence on survival.[57] In our cohort the second predominant growth pattern in each tumour was associated with lymph node involvement and clinical stage, with the micropapillary and acinar patterns associated with N1 disease and a higher clinical stage. After multivariable analysis with Cox's proportional hazards model only clinical stage was independently associated with survival and neither the presence of individual growth patterns nor the predominant growth pattern were associated with prognosis.

Unlike other malignancies, there is no accepted single method of incorporating different histological features into a tumour grading in lung adenocarcinomas.[10] Several studies have suggested that the presence of necrosis in lung adenocarcinomas was a marker of poor prognosis,[15, 64, 65, 283] however this has not been supported by all groups.[59, 63, 66] In the reported cohort necrosis was more commonly found in larger tumours and those with higher clinical stage; although this may be a result of a decrease in blood supply to the centre of large tumours rather than a predictor of prognosis and no link was found with disease specific survival. Mitotic rate is well established in other tumour types, including breast cancer and soft tissue sarcomas, as an indicator tumour grade and therefore prognosis.[302, 303] A large study carried out in NSCLC patients showed mitotic index to be associated with survival;[67] however, other studies in lung cancer have failed to find a significant association.[59, 63] In the current study large numbers of mitotic figures were more commonly observed in larger tumours and there was a tendency for poorer prognoses in univariable and multivariable analyses, which may indicate more aggressive growth in tumours with a high mitotic rate. The presence of large numbers of apoptotic bodies was associated with longer survival in multivariable analysis but fell short of significance after correction for type I errors. Prominent nucleoli and the presence of scar tissue within the tumour both showed some association with pathological parameters but were not linked to survival; any prognostic capacity of prominent nucleoli has also been refuted in a previous cohort.[59] Other tumour characteristics, including cytological pleomorphism and the presence of large numbers of inflammatory cells, have previously been associated with prognosis

but failed to show a significant relationship with survival in this cohort.[63, 67, 76, 78, 79] An overall tumour grading score; which assumed a negative prognostic effect for the presence of necrosis, mitosis and prominent nucleoli and a positive effect for the presence of many apoptotic bodies, scar tissue or large numbers of inflammatory cells; was significantly associated with disease specific survival in univariable analysis but was not an independent predictor of prognosis. However, the assessment of each of these features is very subjective and in order to truly determine the prognostic significance of these characteristics a strict scoring criteria should be established.

CART analysis was carried out as an alternative method of multivariable analysis and to explore models of prognostication based on morphological analysis. Including the results of growth pattern analysis and features of tumour grade in the analysis produced a decision tree with nodal status and lesion size as two of the most influential factors for grouping patients based on disease specific survival. Interestingly; factors commonly reported to be prognostic, such as predominant pattern and presence of the micropapillary pattern, were not incorporated into the decision tree. Although, second predominant pattern was used to further discriminate small N0 tumours and the presence of few apoptotic bodies, followed by the proportion of the solid growth pattern were included to further classify N1 tumours. This classification of tumours with few apoptotic bodies as a poor prognosis group is contradictory to some published reports which have shown the presence of large numbers of apoptotic bodies to be associated with worse outcomes and more aggressive disease;[60, 61] although other groups have shown no significant association with survival.[62, 63] Decision tree models which included only morphological and tumour grade characteristics had very little effect on the ability to predict a poor prognosis group of patients compared to the routinely available parameters; demonstrating that, in this cohort at least, this additional morphological data was of little benefit to the patient.

The mPS score, calculated from clinical stage and the CCP score (cell-cycle related gene expression signature), has been previously shown to be an independent marker of prognosis in lung adenocarcinomas; however, the involvement of lymph nodes was not included in these analyses and its confounding status for these cohorts cannot be assessed.[238, 239, 263, 298] The addition of the mPS score into CART analysis caused the

greatest change to the model and replaced nodal status as the first decision point, indicating it to be a more informative predictor of survival than N stage; in this dataset at least. This decision tree showed the greatest improvement in prediction of prognosis compared to routinely available parameters; however, even this did not produce a model with a strong ability to predict poor prognosis in stage I and II lung adenocarcinoma patients. Although the mPS score is an objective measure with, seemingly, some link to prognosis it requires relatively expensive molecular biological techniques and given the minimal improvement compared to N stage is unlikely to be cost effective in a clinical setting. In addition, the mPS model contradicts previous models in the grouping based on second predominant pattern which suggests solid and papillary as poor prognostic groups and micropapillary as a good prognostic. These differences may suggest over fitting in one or more of the models.

The detailed morphological analysis of this cohort of stage I and II lung adenocarcinomas was unable to confirm the previously published claims of the prognostic value of growth patterns. Given the highly subjective nature of morphological analysis of lung adenocarcinomas any tool using these characteristics for the prognostication of patients is liable to inter, and indeed intra-observer heterogeneity. The majority of published reports showing an association between predominant pattern and survival have been carried out in a single centre with small numbers of histopathologists. In order to robustly test these additional morphological characteristics large scale multi-centre trials, involving many contributing histopathologists, should be carried out; but this is likely to be difficult to achieve in practice. Recently developed digital pathology and automated scoring algorithms have shown considerable promise;[304] this technology offers a future potential to standardise scoring and provide a truly objective classification of a patient's tumour and may lead to better stratification of patients who would benefit from more aggressive treatment strategies.

Chapter 6 Conclusions

This study set out to examine the hypothesis that the analysis of multi-factor data; including histology and candidate molecular genetic, transcriptomic and protein biomarkers; could be used to improve prognostication in patients with early-stage, resected, adenocarcinoma. The dataset generated from this cohort, in multiple biomarkers, broadly reflects those previously published indicating this cohort to be a reasonable representation of this population.

The results of the clinical audit have shown that a Scottish cohort of NSCLC patients who underwent molecular pathology analysis show comparable characteristics to those of other Caucasian cohorts. The incidence of *EGFR*, *KRAS* mutations and ALK rearrangements was approximately 10.5%, 36.5% and 2% respectively and, as has been reported for a population dominated by ex- or current smokers, the majority of *KRAS* mutations were transversions. This audit also yielded some noteworthy results. Contrary to some published studies, there was no evidence to suggest that *EGFR* and *ALK* mutations occurred in younger patients and after taking smoking status into account gender was not independently associated with the presence of *EGFR* mutations. However, the incidence of *KRAS* mutations in females with a smoking history was higher than that in males, even though the females had, on average, smoked less (fewer pack-years); indicating a possible increased susceptibility to tobacco in females which would require further research to confirm. In addition, although ex-smokers show increased age at diagnosis compared to current-smokers the molecular profile of their tumours remains similar to current smokers.

There has been a huge investment in implementing molecular pathology analysis in NSCLC patients in order to predict response to targeted treatments; however, this audit highlights the fact that these tests are only of benefit to a minority of NSCLC patients and the treatment of approximately 90% of a Caucasian population is reliant on clinical and pathological markers only to guide treatment. Taking this into account, the acceptance of the suggested cost saving measures; such as switching to a request based algorithm and, where possible, using serial testing strategies to take advantage of the

mutually exclusive relationship between clinically relevant mutations; should be seriously considered by funding bodies.

Survival of NSCLC patients remains poor and even in early stage patients, who undergo resection with curative intent, a proportion of patients quickly recur and eventually die from their disease; whilst others are effectively cured. In order to improve long-term outcomes for those who do recur, it is essential to base their treatment following resection on the features of their disease. The Scottish Chief Medical Officer's 2016 'Realistic Medicine' report encouraged healthcare practitioners in Scotland to develop a personalised approach, reduce unnecessary variation and improve outcomes by innovation.[305] Contributing to this national healthcare goal this thesis represents, to the best of our knowledge, the only prognostic model generated from the analysis of DNA, RNA, protein expression and morphology (including features of tumour grading) on a single cohort of lesions from lung adenocarcinoma patients. However, as with all research, criticisms can be levelled at this study. The cohort analysed included adenocarcinoma specimens from 208 patients; given the number of groups for some variables, for example predominant growth pattern, a larger cohort may have identified more statistically significant associations. In addition, mutation analysis and therefore EGFR IHC was only carried out on a subset of the samples; however this analysis was applied, as it would be in clinical practice, to patients with involved lymph nodes - representing those with potentially metastatic disease. The use of CART analysis to create decision trees is of particular benefit to data with a large number of variables as it enables the analysis of significant contributors within specific subsets and negates the multiplicity issues seen with more tradition statistical techniques. The resulting models included lymph node involvement, tumour size, variables generated from analysis of growth patterns (second predominant pattern and the proportion of solid or micropapillary growth), cytological features (apoptosis) and gene expression signatures. However, use of morphological data alone gave little improvement to the identification of patients with a particularly poor prognosis. Only models including the molecular proliferation score (mPS) determined by mRNA expression signatures showed an enhanced ability to identify poor prognostic groups; although the improvement was relatively small and

the increase in costs and complexity would make it challenging to implement into clinical practice. CART analysis is known to be prone to overfitting and although 10-fold cross validation was applied in order to minimise this there are some findings which appear biologically implausible; for example, the identification of tumours with a large number of apoptotic bodies as an indicator of favourable prognosis.

The analysis of growth patterns and cytological features are subjective histopathological measurements and are known to be prone to inter-observer variation. If a quantitative estimate of morphological features were required for the accurate prediction of prognosis then this would be difficult to implement in a robust and reproducible manner, across many different countries each with their own local practices and training methods. Most histopathologists would recognise that quantitative assessment of tumour features by manual light microscopy alone would be inaccurate meaning that only very strong prognostic features are likely to be robust enough to prove their utility in current clinical practice. Automated histology offers hope for a system free of inter-observer variation. Raman spectroscopy has shown utility in the diagnosis of lung cancers and pre-malignant lesions.[306] More recently, automated analysis of high-power scanned images of stained histopathology sections has been shown to identify poorly differentiated regions of colorectal cancer which was significantly associated with recurrence.[307] One group has used automated histology to identify regions of a lung adenocarcinomas with the solid and micropapillary growth patterns, the far more accurate data which could be provided with this technology may help identify clinically relevant associations.[308] Automated analysis of digital pathology images has also been used to develop an algorithm which could identify a group of lung adenocarcinoma patients with a poor prognosis which out-performed both stage and tumour grade in its prognostic capacity.[304] Many histopathology departments are now investigating digital pathology for implementation in routine clinical practice. The images collected will allow for automated analysis on a very large scale with the aim of improving diagnosis, prediction and prognostication in a wide range of cancer types. In the future, it is likely that automated analysis of digital pathology will be a powerful tool to improve

outcomes in lung cancer patients and these techniques should be keenly pursued to remove inter-observer variation and standardise prognostic classifications.

Although progress has been made in the treatment of adenocarcinoma, for example in the analysis of *EGFR* mutation status, it is clear that such single analyte measures are far from efficient at predicting a patient's disease pathway and response to treatment. It seems obvious that the measurement of the mutation status of a single gene is unlikely to accurately predict tumour behaviour which is dictated by a complex array of interconnecting pathways as well as cell-to-cell and cell-to-environment interactions. This study has shown by IHC analysis of mutated and total EGFR protein in *EGFR* mutated tumours that there is a high degree of intratumour heterogeneity in protein expression, and in one case protein localisation; however, in this cohort at least, there was no evidence of intratumour heterogeneity of mutation status. The effect of this variation in EGFR protein expression is unknown and would require further investigation in a much larger cohort of EGFR TKI treated patients. The current reliance of large areas of cancer research on molecular genetics alone may need to alter before significant progress can be made in the treatment of NSCLC. Likewise relying on a single additional morphological measurement, such as predominant growth pattern, is unlikely to sufficiently represent the complex interactions between tumour and patient genetics, proteomics and environment in order to predict disease progression. In future, tumour analysis may involve a complex analysis of many different variables including the patient's clinical features, morphological analysis, mutation status, and protein expression and function; but this level of multiple analyte assessment would be difficult to achieve in a routine healthcare setting in the near future. One barrier to its implementation is the current lack of systems which are able to analyse DNA, RNA and protein level analytes in the same sample; especially considering the small size of samples routinely available from the majority of lung cancer patients. In order to make considerable improvements in patient outcomes it is likely that any measurements should be made at a level as close to the functional activity of the tumour as possible. A comprehensive analysis of expression level and structure of proteins, for example, may yield more relevant results than genomic DNA sequences which may never be expressed. Several groups, including commercial

companies, are working on machine learning techniques for the analysis of large datasets- such as gene expression, mutations and imaging- in combination with clinical factors. As yet, published studies using machine learning in lung adenocarcinomas have been restricted to a single data source.[309] However once systems including multisource data are developed, these unbiased algorithms offer the best hope of providing objective stratification of lung adenocarcinoma patients which can only improve outcomes in this highly heterogeneous disease.

It is well known that the characteristics of each tumour can change over time, particularly in response to treatment; acquired resistance in patients treated with targeted therapies is frequently mediated by additional mutations or upregulation of alternative pathways.[168, 187] Since the development of a therapy designed to be active in EGFR TKI-treated patients with acquired p.T790M mutations the re-analysis of the *EGFR* gene in samples taken post treatment has become more common. As more treatments are developed targeting specific subgroups of the population the need for repeat analysis is likely to increase. The recently developed ability to interrogate circulating cell-free tumour DNA has aided the implementation of this longitudinal analysis and is of particular value in lung cancer patients for whom obtaining a tissue sample is frequently both difficult and harmful.[174] It is hoped that similar minimally invasive procedures will be developed to analyse a wider range of biomarkers in lung cancer patients over time. However, these strategies would require a further large investment in translational research and a huge shift in our current practices.

Chapter 7 References

1. Ferlay J, Soerjomataram I, Ervik M *et al.* Globocan 2012 v1.0, cancer incidence and mortality worldwide: IARC cancerbase no. 11 [internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr, accessed on 21/01/2014., 2012.
2. ISD. Cancer survival in Scotland 1987-2011. 2015.
3. Magadi W, Exarchakou A, Rachet B *et al.* Cancer survival in England: Patients diagnosed between 2010 and 2014 and followed up to 2015 London, 2016.
4. Gasent Blesa JM, Esteban González E, Alberola Candel V. Screening and chemoprevention in lung cancer. *Clin Transl Oncol* 2008;**10**;274-280.
5. TNM classification of malignant tumours. 7 ed. Chichester UK: Wiley-Blackwell, 2009.
6. Network SIG. Management of lung cancer: a national clinical guideline. In Baxter G, Brown K, Herriot R, James R, Sidek N, Twaddle S eds. Edinburgh: Scottish Intercollegiate Guidelines Network, 2014.
7. Goldstraw P, Crowley J, Chansky K *et al.* The IASLC lung cancer staging project: Proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM classification of malignant tumours. *Journal of Thoracic Oncology* 2007;**2**;706-714.
8. Kerr KM. Pulmonary adenocarcinomas: Classification and reporting. *Histopathology* 2009;**54**;12-27.
9. Strand TE, Rostad H, Moller B, Norstein J. Survival after resection for primary lung cancer: A population based study of 3211 resected patients. *Thorax* 2006;**61**;710-715.
10. WHO classification of tumours of the lung, pleura, thymus and heart (4th edition). 4 ed. Lyon: International Agency for Research on Cancer, 2015.
11. Pathology and genetics of tumours of the lung, pleura, thymus and heart. 3 ed. Lyon, France: International Agency for Research on Cancer, 2004;344.
12. Putila J, Remick SC, Guo NL. Combining clinical, pathological, and demographic factors refines prognosis of lung cancer: A population-based study. *Plos One* 2011;**6**;13.
13. Quint LE, Tummala S, Brisson LJ *et al.* Distribution of distant metastases from newly diagnosed non-small cell lung cancer. *Annals of Thoracic Surgery* 1996;**62**;246-250.
14. Travis WD, Brambilla E, Nicholson AG *et al.* The 2015 World Health Organization classification of lung tumors: Impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* 2015;**10**;1243-1260.
15. Yoshizawa A, Motoi N, Riely GJ *et al.* 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. *Mod Pathol* 2011;**24**;653-664.
16. Travis WD, Brambilla E, Noguchi M *et al.* International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* 2011;**6**;244-285.

17. Zhang J, Liang ZY, Gao J, Luo YF, Liu TH. Pulmonary adenocarcinoma with a micropapillary pattern: A clinicopathological, immunophenotypic and molecular analysis. *Histopathology* 2011;**59**;1204-1214.
18. Miyoshi T, Satoh Y, Okumura S *et al.* Early-stage lung adenocarcinomas with a micropapillary pattern, a distinct pathologic marker for a significantly poor prognosis. *Am J Surg Pathol* 2003;**27**;101-109.
19. Kamiya K, Hayashi Y, Douguchi J *et al.* Histopathological features and prognostic significance of the micropapillary pattern in lung adenocarcinoma. *Mod Pathol* 2008;**21**;992-1001.
20. Sumiyoshi S, Yoshizawa A, Sonobe M *et al.* Pulmonary adenocarcinomas with micropapillary component significantly correlate with recurrence, but can be well controlled with EGFR tyrosine kinase inhibitors in the early stages. *Lung Cancer* 2013;**81**;53-59.
21. Tsutsumida H, Nomoto M, Goto M *et al.* A micropapillary pattern is predictive of a poor prognosis in lung adenocarcinoma, and reduced surfactant apoprotein a expression in the micropapillary pattern is an excellent indicator of a poor prognosis. *Mod Pathol* 2007;**20**;638-647.
22. Satoh Y, Hoshi R, Horai T *et al.* Association of cytologic micropapillary clusters in cytology samples with lymphatic spread in clinical stage I lung adenocarcinomas. *Lung Cancer* 2009;**64**;277-281.
23. Tsubokawa N, Mimae T, Sasada S *et al.* Negative prognostic influence of micropapillary pattern in stage Ia lung adenocarcinoma. *European Journal of Cardio-Thoracic Surgery* 2016;**49**;293-299.
24. Nagano T, Ishii G, Nagai K *et al.* Structural and biological properties of a papillary component generating a micropapillary component in lung adenocarcinoma. *Lung Cancer* 2010;**67**;282-289.
25. Makimoto Y, Nabeshima K, Iwasaki H *et al.* 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* 2005;**46**;677-684.
26. Xu ST, Xi JJ, Jiang W, Lu SH, Wang Q. Solid component and tumor size correlate with prognosis of stage Ib lung adenocarcinoma. *Annals of Thoracic Surgery* 2015;**99**;961-967.
27. Sánchez-Mora N, Presmanes MC, Monroy V *et al.* Micropapillary lung adenocarcinoma: A distinctive histologic subtype with prognostic significance. Case series. *Hum Pathol* 2008;**39**;324-330.
28. Solis LM, Behrens C, Raso MG *et al.* Histologic patterns and molecular characteristics of lung adenocarcinoma associated with clinical outcome. *Cancer* 2012;**118**;2889-2899.
29. Ohtaki Y, Yoshida J, Ishii G *et al.* Prognostic significance of a solid component in pulmonary adenocarcinoma. *Annals of Thoracic Surgery* 2011;**91**;1051-1057.
30. Riquet M, Foucault C, Berna P, Assouad J, Dujon A, Danel C. Prognostic value of histology in resected lung cancer with emphasis on the relevance of the adenocarcinoma subtyping. *Ann Thorac Surg* 2006;**81**;1988-1995.
31. Takahashi Y, Ishii G, Aokage K, Hishida T, Yoshida J, Nagai K. Distinctive histopathological features of lepidic growth predominant node-negative adenocarcinomas 3-5 cm in size. *Lung Cancer* 2013;**79**;118-124.

32. Moon Y, Sung SW, Lee KY, Kim YK, Park JK. The importance of the lepidic component as a prognostic factor in stage I pulmonary adenocarcinoma. *World Journal of Surgical Oncology* 2016;**14**;8.
33. Warth A, Muley T, Meister M *et al.* 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* 2012;**30**;1438-1446.
34. Tsuta K, Kawago M, Inoue E *et al.* The utility of the proposed IASLC/ATS/ERS lung adenocarcinoma subtypes for disease prognosis and correlation of driver gene alterations. *Lung Cancer* 2013;**81**;371-376.
35. Nakamura H, Saji H, Shinmyo T *et al.* Association of IASLC/ATS/ERS histologic subtypes of lung adenocarcinoma with epidermal growth factor receptor mutations in 320 resected cases. *Clin Lung Cancer* 2014.
36. Zhang HB, Lu C, Lu YJ, Yu B, Lv FZ, Zhu ZH. The predictive and prognostic values of factors associated with visceral pleural involvement in resected lung adenocarcinomas. *Oncotargets and Therapy* 2016;**9**;2337-2348.
37. Zhang J, Wu J, Tan Q, Zhu L, Gao W. Why do pathological stage Ia lung adenocarcinomas vary from prognosis?: A clinicopathologic study of 176 patients with pathological stage Ia lung adenocarcinoma based on the IASLC/ATS/ERS classification. *J Thorac Oncol* 2013;**8**;1196-1202.
38. Xu CH, Wang W, Wei Y *et al.* Prognostic value of the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification in stage Ib lung adenocarcinoma. *EJSO* 2015;**41**;1430-1436.
39. Yoshizawa A, Sumiyoshi S, Sonobe M *et al.* Validation of the IASLC/ATS/ERS lung adenocarcinoma classification for prognosis and association with EGFR and KRAS gene mutations: Analysis of 440Jjapanese patients. *J Thorac Oncol* 2013;**8**;52-61.
40. Russell PA, Wainer Z, Wright GM, Daniels M, Conron M, Williams RA. 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* 2011;**6**;1496-1504.
41. Woo T, Okudela K, Mitsui H *et al.* Prognostic value of the IASLC/ATS/ERS classification of lung adenocarcinoma in stage I disease of Japanese cases. *Pathology International* 2012;**62**;785-791.
42. Hung JJ, Yeh YC, Jeng WJ *et al.* 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* 2014;**32**;2357-U2244.
43. Song Z, Zhu H, Guo Z, Wu W, Sun W, Zhang Y. Prognostic value of the IASLC/ATS/ERS classification in stage I lung adenocarcinoma patients-based on a hospital study in China. *Ejso* 2013;**39**;1262-1268.
44. Yanagawa N, Shiono S, Abiko M, Ogata S, Sato T, Tamura G. The correlation of the International Association for the Study of Lung Cancer (IASLC)/American Thoracic Society (ATS)/European Respiratory Society (ERS) classification with prognosis and EGFR mutation in lung adenocarcinoma. *Annals of Thoracic Surgery* 2014;**98**;453-458.

45. Zhang Y, Sun YH, Xiang JQ, Zhang YW, Hu H, Chen HQ. A clinicopathologic prediction model for postoperative recurrence in stage Ia non-small cell lung cancer. *Journal of Thoracic and Cardiovascular Surgery* 2014;**148**;1193-1199.
46. Campos-Parra AD, Avilés A, Contreras-Reyes S *et al.* Relevance of the novel IASLC/ATS/ERS classification of lung adenocarcinoma in advanced disease. *Eur Respir J* 2014;**43**;1439-1447.
47. Nakamura H, Saji H, Shinmyo T *et al.* Association of IASLC/ATS/ERS histologic subtypes of lung adenocarcinoma with epidermal growth factor receptor mutations in 320 resected cases. *Clinical Lung Cancer* 2015;**16**;209-215.
48. Hung JJ, Jeng WJ, Chou TY *et al.* Prognostic value of the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society lung adenocarcinoma classification on death and recurrence in completely resected stage I lung adenocarcinoma. *Annals of Surgery* 2013;**258**;1079-1086.
49. De Melo AC, de Sa VK, Sternberg C *et al.* Mutational profile and new IASLC/ATS/ERS classification provide additional prognostic information about lung adenocarcinoma: A study of 125 patients from Brazil. *Oncology* 2015;**89**;175-186.
50. Russell PA, Barnett SA, Walkiewicz M *et al.* Correlation of mutation status and survival with predominant histologic subtype according to the new IASLC/ATS/ERS lung adenocarcinoma classification in stage III (N2) patients. *J Thorac Oncol* 2013;**8**;461-468.
51. Hu H, Pan Y, Li Y *et al.* Oncogenic mutations are associated with histological subtypes but do not have an independent prognostic value in lung adenocarcinoma. *Onco Targets Ther* 2014;**7**;1423-1437.
52. Tsao MS, Marguet S, Le Teuff G *et al.* Subtype classification of lung adenocarcinoma predicts benefit from adjuvant chemotherapy in patients undergoing complete resection. *Journal of Clinical Oncology* 2015;**33**;3439-+.
53. Urer HN, Kocaturk CI, Gunluoglu MZ *et al.* Relationship between lung adenocarcinoma histological subtype and patient prognosis. *Annals of Thoracic and Cardiovascular Surgery* 2014;**20**;12-18.
54. Westaway DD, Toon CW, Farzin M *et al.* The International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society grading system has limited prognostic significance in advanced resected pulmonary adenocarcinoma. *Pathology* 2013;**45**;553-558.
55. Luo JZ, Huang QY, Wang R *et al.* Prognostic and predictive value of the novel classification of lung adenocarcinoma in patients with stage Ib. *Journal of Cancer Research and Clinical Oncology* 2016;**142**;2031-2040.
56. Warth A, Muley T, Harms A *et al.* Clinical relevance of different papillary growth patterns of pulmonary adenocarcinoma. *American Journal of Surgical Pathology* 2016;**40**;818-826.
57. Zhao ZR, Xi SY, Li W *et al.* Prognostic impact of pattern-based grading system by the new IASLC/ATS/ERS classification in asian patients with stage I lung adenocarcinoma. *Lung Cancer* 2015;**90**;604-609.
58. Elston CW, Ellis IO. Pathological prognostic factors in breast-cancer. The value of histological grade in breast-cancer - experience from a large study with long-term follow-up. *Histopathology* 1991;**19**;403-410.
59. Duhig EE, Dettrick A, Godbolt DB *et al.* Mitosis trumps T stage and proposed International Association for the Study of Lung Cancer/American Thoracic

Society/European Respiratory Society classification for prognostic value in resected stage 1 lung adenocarcinoma. *Journal of Thoracic Oncology* 2015;**10**;673-681.
60. Komaki R, Fujii T, Perkins P *et al.* Apoptosis and mitosis as prognostic factors in pathologically staged N1 nonsmall cell lung cancer. *International Journal of Radiation Oncology Biology Physics* 1996;**36**;601-605.
61. Langendijk H, Thunnissen E, Arends JW *et al.* Cell proliferation and apoptosis in stage III inoperable non-small cell lung carcinoma treated by radiotherapy. *Radiotherapy and Oncology* 2000;**56**;197-207.
62. Dworakowska D, Jassem E, Jassem J *et al.* Prognostic value of the apoptotic index analysed jointly with selected cell cycle regulators and proliferation markers in non-small cell lung cancer. *Lung Cancer* 2009;**66**;127-133.
63. Petersen I, Kotb W, Friedrich KH, Schluns K, Bocking A, Dietel M. Core classification of lung cancer: Correlating nuclear size and mitoses with ploidy and clinicopathological parameters. *Lung Cancer* 2009;**65**;312-318.
64. Swinson DEB, Jones JL, Richardson D, Cox G, Edwards JG, O'Byrne KJ. Tumour necrosis is an independent prognostic marker in non-small cell lung cancer: Correlation with biological variables. *Lung Cancer* 2002;**37**;235-240.
65. Park SY, Lee HS, Jang HJ, Lee GK, Chung KY, Zo JI. Tumor necrosis as a prognostic factor for stage Ia non-small cell lung cancer. *Annals of Thoracic Surgery* 2011;**91**;1668-1673.
66. Nia PS, Van Marck E, Weyler J, Van Schil P. Prognostic value of a biologic classification of non-small-cell lung cancer into the growth patterns along with other clinical, pathological and immunohistochemical factors. *European Journal of Cardio-Thoracic Surgery* 2010;**38**;628-636.
67. Shimizu K, Yoshida J, Nagai K *et al.* Visceral pleural invasion is an invasive and aggressive indicator of non-small cell lung cancer. *Journal of Thoracic and Cardiovascular Surgery* 2005;**130**;160-165.
68. Johnson SK, Kerr KM, Chapman AD *et al.* Immune cell infiltrates and prognosis in primary carcinoma of the lung. *Lung Cancer* 2000;**27**;27-35.
69. Kayser G, Schulte-Uentrop L, Sienel W *et al.* Stromal CD4/CD25 positive T-cells are a strong and independent prognostic factor in non-small cell lung cancer patients, especially with adenocarcinomas. *Lung Cancer* 2012;**76**;445-451.
70. Hiraoka K, Miyamoto M, Cho Y *et al.* Concurrent infiltration by CD8(+) T cells and CD4(+) T cells is a favourable prognostic factor in non-small-cell lung carcinoma. *British Journal of Cancer* 2006;**94**;275-280.
71. Kawai O, Ishii G, Kubota K *et al.* Predominant infiltration of macrophages and CD8(+) T cells in cancer nests is a significant predictor of survival in stage IV nonsmall cell lung cancer. *Cancer* 2008;**113**;1387-1395.
72. Al-Shibli KI, Donnem T, Al-Saad S, Persson M, Bremnes RM, Busund LT. Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. *Clinical Cancer Research* 2008;**14**;5220-5227.
73. Schalper KA, Brown J, Carvajal-Hausdorf D *et al.* Objective measurement and clinical significance of TILS in non-small cell lung cancer. *JNCI-Journal of the National Cancer Institute* 2015;**107**;9.
74. Tian CT, Lu SX, Fan QX *et al.* Prognostic significance of tumor-infiltrating CD8(+) or CD3(+) T lymphocytes and interleukin-2 expression in radically resected non-small cell lung cancer. *Chinese Medical Journal* 2015;**128**;105-110.

75. Toomey D, Smyth G, Condron C *et al.* Infiltrating immune cells, but not tumour cells, express FASL in non-small cell lung cancer: No association with prognosis identified in 3-year follow-up. *International Journal of Cancer* 2003;**103**;408-412.
76. Horne ZD, Jack R, Gray ZT *et al.* Increased levels of tumor-infiltrating lymphocytes are associated with improved recurrence-free survival in stage 1a non-small-cell lung cancer. *Journal of Surgical Research* 2011;**171**;1-5.
77. Feng W, Li Y, Shen L *et al.* Prognostic value of tumor-infiltrating lymphocytes for patients with completely resected stage IIIA(N2) non-small cell lung cancer. *Oncotarget* 2016;**7**;7227-7240.
78. Geng YT, Shao YJ, He WT *et al.* Prognostic role of tumor-infiltrating lymphocytes in lung cancer: A meta-analysis. *Cellular Physiology and Biochemistry* 2015;**37**;1560-1571.
79. Zeng DQ, Yu YF, Ou QY *et al.* Prognostic and predictive value of tumor-infiltrating lymphocytes for clinical therapeutic research in patients with non-small cell lung cancer. *Oncotarget* 2016;**7**;13765-13781.
80. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;**100**;57-70.
81. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell* 2011;**144**;646-674.
82. Jia PL, Pao W, Zhao ZM. Patterns and processes of somatic mutations in nine major cancers. *Bmc Medical Genomics* 2014;**7**;11.
83. Stricker T, Catenacci DV, Seiwert TY. Molecular profiling of cancer--the future of personalized cancer medicine: A primer on cancer biology and the tools necessary to bring molecular testing to the clinic. *Seminars in oncology* 2011;**38**;173-185.
84. Ding L, Getz G, Wheeler DA *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008;**455**;1069-1075.
85. Dearden S, Stevens J, Wu YL, Blowers D. Mutation incidence and coincidence in non small-cell lung cancer: Meta-analyses by ethnicity and histology (MUTmap). *Ann Oncol* 2013;**24**;2371-2376.
86. Devarakonda S, Morgensztern D, Govindan R. Genomic alterations in lung adenocarcinoma. *Lancet Oncology* 2015;**16**;E342-E351.
87. Michor F, Liphardt J, Ferrari M, Widom J. What does physics have to do with cancer? *Nat Rev Cancer* 2011;**11**;657-670.
88. Langreth R, Waldholz M. New era of personalized medicine: Targeting drugs for each unique genetic profile. *Oncologist* 1999;**4**;426-427.
89. Voldborg BR, Damstrup L, Spang-Thomsen M, Poulsen HS. Epidermal growth factor receptor (EGFR) and EGFR mutations, function and possible role in clinical trials. *Ann Oncol* 1997;**8**;1197-1206.
90. Mascaux C, Wynes MW, Kato Y *et al.* EGFR protein expression in non-small cell lung cancer predicts response to an EGFR tyrosine kinase inhibitor-a novel antibody for immunohistochemistry or aqua technology. *Clinical Cancer Research* 2011;**17**;7796-7807.
91. Jorissen RN, Walker F, Pouliot N, Garrett TPJ, Ward CW, Burgess AW. Epidermal growth factor receptor: Mechanisms of activation and signalling. *Experimental Cell Research* 2003;**284**;31-53.
92. Schneider MR, Wolf E. The epidermal growth factor receptor ligands at a glance. *Journal of Cellular Physiology* 2009;**218**;460-466.

93. Herbst RS. Review of epidermal growth factor receptor biology. *Int J Radiat Oncol Biol Phys* 2004;**59**;21-26.
94. Zhou C, Wu YL, Chen G *et al.* Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): A multicentre, open-label, randomised, phase 3 study. *Lancet Oncol* 2011;**12**;735-742.
95. Mok TS, Wu YL, Thongprasert S *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *The New England journal of medicine* 2009;**361**;947-957.
96. Thatcher N, Chang A, Parikh P *et al.* Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: Results from a randomised, placebo-controlled, multicentre study (IRESSA survival evaluation in lung cancer). *Lancet* 2005;**366**;1527-1537.
97. Shepherd FA, Rodrigues Pereira J, Ciuleanu T *et al.* Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med* 2005;**353**;123-132.
98. Chang A, Parikh P, Thongprasert S *et al.* Gefitinib (IRESSA) in patients of Asian origin with refractory advanced non-small cell lung cancer: Subset analysis from the ISEL study. *J Thorac Oncol* 2006;**1**;847-855.
99. Baselga J, Averbuch SD. ZD1839 ('IRESSA')(1,2) as an anticancer agent. *Drugs* 2000;**60**;33-40.
100. Yun CH, Boggon TJ, Li Y *et al.* Structures of lung cancer-derived EGFR mutants and inhibitor complexes: Mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* 2007;**11**;217-227.
101. Lim SH, Lee JY, Sun JM, Ahn JS, Park K, Ahn MJ. Comparison of clinical outcomes following gefitinib and erlotinib treatment in non-small-cell lung cancer patients harboring an epidermal growth factor receptor mutation in either exon 19 or 21. *Journal of Thoracic Oncology* 2014;**9**;506-511.
102. Hirsch FR, Varella-Garcia M, Cappuzzo F *et al.* Combination of EGFR gene copy number and protein expression predicts outcome for advanced non-small-cell lung cancer patients treated with gefitinib. *Ann Oncol* 2007;**18**;752-760.
103. Clark GM, Zborowski DM, Culbertson JL *et al.* Clinical utility of epidermal growth factor receptor expression for selecting patients with advanced non-small cell lung cancer for treatment with erlotinib. *Journal of Thoracic Oncology* 2006;**1**;837-846.
104. Brugger W, Triller N, Blasinska-Morawiec M *et al.* Prospective molecular marker analyses of EGFR and KRAS from a randomized, placebo-controlled study of erlotinib maintenance therapy in advanced non-small-cell lung cancer. *Journal of Clinical Oncology* 2011;**29**;4113-4120.
105. Schneider CP, Heigener D, Schott-von-Römer K *et al.* Epidermal growth factor receptor-related tumor markers and clinical outcomes with erlotinib in non-small cell lung cancer: An analysis of patients from German centers in the trust study. *J Thorac Oncol* 2008;**3**;1446-1453.
106. Taillade L, Penault-Llorca F, Boulet T *et al.* Immunohistochemichal expression of biomarkers: A comparative study between diagnostic bronchial biopsies and surgical specimens of non-small-cell lung cancer. *Ann Oncol* 2007;**18**;1043-1050.
107. Ferrigan L, Wallace WA. Predicting non-small cell lung cancer expression of epidermal growth factor receptor and matrix metalloproteinase 9 from

immunohistochemical staining of diagnostic biopsy samples. *Eur J Cancer* 2004;**40**;1589-1592.
108. Hirsch FR, Dziadziuszko R, Thatcher N *et al.* Epidermal growth factor receptor immunohistochemistry - comparison of antibodies and cutoff points to predict benefit from gefitinib in a phase 3 placebo-controlled study in advanced nonsmall-cell lung cancer. *Cancer* 2008;**112**;1114-1121.
109. Hirsch FR, Varella-Garcia M, Bunn PA *et al.* Epidermal growth factor receptor in non-small-cell lung carcinomas: Correlation between gene copy number and protein expression and impact on prognosis. *J Clin Oncol* 2003;**21**;3798-3807.
110. Hirsch FR, Varella-Garcia M, Bunn PA, Jr. *et al.* Molecular predictors of outcome with gefitinib in a phase III placebo-controlled study in advanced non-small-cell lung cancer. *Journal of Clinical Oncology* 2006;**24**;5034-5042.
111. Cappuzzo F, Hirsch FR, Rossi E *et al.* Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer. *Journal of the National Cancer Institute* 2005;**97**;643-655.
112. Zhu CQ, da Cunha Santos G, Ding K *et al.* Role of KRAS and EGFR as biomarkers of response to erlotinib in national cancer institute of canada clinical trials group study BR.21. *J Clin Oncol* 2008;**26**;4268-4275.
113. Goss G, Ferry D, Wierzbicki R *et al.* Randomized phase II study of gefitinib compared with placebo in chemotherapy-naive patients with advanced non-small-cell lung cancer and poor performance status. *J Clin Oncol* 2009;**27**;2253-2260.
114. Hirsch FR, Varella-Garcia M, Dziadziuszko R *et al.* Fluorescence in situ hybridization subgroup analysis of tribute, a phase III trial of erlotinib plus carboplatin and paclitaxel in non-small cell lung cancer. *Clin Cancer Res* 2008;**14**;6317-6323.
115. Dahabreh IJ, Linardou H, Kosmidis P, Bafaloukos D, Murray S. EGFR gene copy number as a predictive biomarker for patients receiving tyrosine kinase inhibitor treatment: A systematic review and meta-analysis in non-small-cell lung cancer. *Ann Oncol* 2011;**22**;545-552.
116. Tsao MS, Sakurada A, Cutz JC *et al.* Erlotinib in lung cancer - molecular and clinical predictors of outcome. *N Engl J Med* 2005;**353**;133-144.
117. Kim ES, Hirsh V, Mok T *et al.* Gefitinib versus docetaxel in previously treated non-small-cell lung cancer (INTEREST): A randomised phase III trial. *Lancet* 2008;**372**;1809-1818.
118. Pao W, Miller V, Zakowski M *et al.* EGF receptor gene mutations are common in lung cancers from "Never smokers" And are associated with sensitivity of tumors to gefitinib and erlotinib. *Proceedings of the National Academy of Sciences of the United States of America* 2004;**101**;13306-13311.
119. Lynch TJ, Bell DW, Sordella R *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004;**350**;2129-2139.
120. Paez JG, Jänne PA, Lee JC *et al.* Egfr mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science* 2004;**304**;1497-1500.
121. Huang SF, Liu HP, Li LH *et al.* High frequency of epidermal growth factor receptor mutations with complex patterns in non-small cell lung cancers related to gefitinib responsiveness in Taiwan. *Clin Cancer Res* 2004;**10**;8195-8203.
122. Tracy S, Mukohara T, Hansen M, Meyerson M, Johnson BE, Jänne PA. Gefitinib induces apoptosis in the EGFRL858R non-small-cell lung cancer cell line H3255. *Cancer Res* 2004;**64**;7241-7244.

123. Gandhi J, Zhang J, Xie Y *et al.* Alterations in genes of the EGFR signaling pathway and their relationship to EGFR tyrosine kinase inhibitor sensitivity in lung cancer cell lines. *PLoS One* 2009;**4**;e4576.
124. Gahr S, Stoehr R, Geissinger E *et al.* Egfr mutational status in a large series of Caucasian European NSCLC patients: Data from daily practice. *Br J Cancer* 2013;**109**;1821-1828.
125. Gazdar AF. Activating and resistance mutations of EGFR in non-small-cell lung cancer: Role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene* 2009;**28 Suppl 1**;S24-31.
126. Sharma SV, Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. *Nature reviews.Cancer* 2007;**7**;169-181.
127. Yoshida Y, Shibata T, Kokubu A *et al.* Mutations of the epidermal growth factor receptor gene in atypical adenomatous hyperplasia and bronchioloalveolar carcinoma of the lung. *Lung Cancer* 2005;**50**;1-8.
128. Ninomiya H, Hiramatsu M, Inamura K *et al.* Correlation between morphology and EGFR mutations in lung adenocarcinomas significance of the micropapillary pattern and the hobnail cell type. *Lung Cancer* 2009;**63**;235-240.
129. Kosaka T, Yatabe Y, Endoh H, Kuwano H, Takahashi T, Mitsudomi T. Mutations of the epidermal growth factor receptor gene in lung cancer: Biological and clinical implications. *Cancer Research* 2004;**64**;8919-8923.
130. Tokumo M, Toyooka S, Kiura K *et al.* The relationship between epidermal growth factor receptor mutations and clinicopathologic features in non-small cell lung cancers. *Clin Cancer Res* 2005;**11**;1167-1173.
131. Hsiao SH, Lin SE, Chou YT *et al.* Histological subtype and smoking status, but not gender, are associated with epidermal growth factor receptor mutations in non-small-cell lung cancer. *Mol Clin Oncol* 2014;**2**;252-258.
132. Dogan S, Shen R, Ang DC *et al.* Molecular epidemiology of EGFR and KRAS mutations in 3,026 lung adenocarcinomas: Higher susceptibility of women to smoking-related KRAS-mutant cancers. *Clin Cancer Res* 2012;**18**;6169-6177.
133. Kim YT, Seong YW, Jung YJ *et al.* The presence of mutations in epidermal growth factor receptor gene is not a prognostic factor for long-term outcome after surgical resection of non-small-cell lung cancer. *Journal of Thoracic Oncology* 2013;**8**;171-178.
134. Sun PL, Seol H, Lee HJ *et al.* High incidence of EGFR mutations in Korean men smokers with no intratumoral heterogeneity of lung adenocarcinomas: Correlation with histologic subtypes, EGFR/TTF-1 expressions, and clinical features. *J Thorac Oncol* 2012;**7**;323-330.
135. Jie-Liu, Li XY, Zhao YQ *et al.* Genotype-phenotype correlation in chinese patients with pulmonary mixed type adenocarcinoma: Relationship between histologic subtypes, TITF-1/SP-A expressions and EGFR mutations. *Pathol Res Pract* 2014;**210**;176-181.
136. Yatabe Y, Kosaka T, Takahashi T, Mitsudomi T. EGFR mutation is specific for terminal respiratory unit type adenocarcinoma. *Am J Surg Pathol* 2005;**29**;633-639.
137. Fukuoka M, Wu YL, Thongprasert S *et al.* Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients with advanced non-small-cell lung cancer in Asia (IPASS). *J Clin Oncol* 2011;**29**;2866-2874.

138. Inoue Y, Ikegami Y, Sano K *et al.* Gefitinib enhances the antitumor activity of CPT-11 in vitro and in vivo by inhibiting ABCG2 but not ABCB1: A new clue to circumvent gastrointestinal toxicity risk. *Chemotherapy* 2013;**59**;260-272.
139. Zhou C, Wu YL, Chen G *et al.* Final overall survival results from a randomised, phase III study of erlotinib versus chemotherapy as first-line treatment of EGFR mutation-positive advanced non-small-cell lung cancer (OPTIMAL, CTONG-0802). *Annals of Oncology* 2015;**26**;1877-1883.
140. Li GF, Gao SJ, Sheng ZX, Li B. The efficacy of single-agent epidermal growth factor receptor tyrosine kinase inhibitor therapy in biologically selected patients with non-small-cell lung cancer: A meta-analysis of 19 randomized controlled trials. *Chemotherapy* 2015;**61**;179-189.
141. Lee CK, Brown C, Gralla RJ *et al.* Impact of EGFR inhibitor in non-small cell lung cancer on progression-free and overall survival: A meta-analysis. *J Natl Cancer Inst* 2013;**105**;595-605.
142. Keedy VL, Temin S, Somerfield MR *et al.* American society of clinical oncology provisional clinical opinion: Epidermal growth factor receptor (EGFR) mutation testing for patients with advanced non-small-cell lung cancer considering first-line EGFR tyrosine kinase inhibitor therapy. *Journal of Clinical Oncology : official journal of the American Society of Clinical Oncology* 2011.
143. Pirker R, Herth FJ, Kerr KM *et al.* Consensus for EGFR mutation testing in non-small cell lung cancer: Results from a european workshop. *J Thorac Oncol* 2010;**5**;1706-1713.
144. Yang JCH, Shih JY, Su WC *et al.* Afatinib for patients with lung adenocarcinoma and epidermal growth factor receptor mutations (LUX-LUNG 2): A phase 2 trial. *Lancet Oncology* 2012;**13**;539-548.
145. Sequist LV, Yang JC, Yamamoto N *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol* 2013;**31**;3327-3334.
146. Miller VA, Hirsh V, Cadranel J *et al.* Afatinib versus placebo for patients with advanced, metastatic non-small-cell lung cancer after failure of erlotinib, gefitinib, or both, and one or two lines of chemotherapy (LUX-LUNG 1): A phase 2b/3 randomised trial. *Lancet Oncology* 2012;**13**;528-538.
147. Park K, Tan EH, O'Byrne K *et al.* Afatinib versus gefitinib as first-line treatment of patients with EGFR mutation-positive non-small-cell lung cancer (LUX-LUNG 7): A phase 2b, open-label, randomised controlled trial. *Lancet Oncology* 2016;**17**;577-589.
148. Ding D, Yu YF, Li ZM, Niu XM, Lu S. The predictive role of pretreatment epidermal growth factor receptor T790M mutation on the progression-free survival of tyrosine-kinase inhibitor-treated non-small cell lung cancer patients: A meta-analysis. *Oncotargets and Therapy* 2014;**7**;387-393.
149. Woo HS, Ahn HK, Lee HY *et al.* Epidermal growth factor receptor (EGFR) exon 20 mutations in non-small-cell lung cancer and resistance to EGFR-tyrosine kinase inhibitors. *Investigational New Drugs* 2014;**32**;1311-1315.
150. Yu HA, Arcila ME, Hellmann MD, Kris MG, Ladanyi M, Riely GJ. Poor response to erlotinib in patients with tumors containing baseline EGFR T790M mutations found by routine clinical molecular testing. *Annals of Oncology* 2014;**25**;423-428.

151. Lee B, Lee T, Lee SH, Choi YL, Han J. Clinicopathologic characteristics of EGFR, KRAS, and ALK alterations in 6,595 lung cancers. *Oncotarget* 2016.
152. Takano T, Ohe Y, Tsuta K *et al.* Epidermal growth factor receptor mutation detection using high-resolution melting analysis predicts outcomes in patients with advanced non small cell lung cancer treated with gefitinib. *Clin Cancer Res* 2007;**13**;5385-5390.
153. Jackman DM, Yeap BY, Sequist LV *et al.* Exon 19 deletion mutations of epidermal growth factor receptor are associated with prolonged survival in non-small cell lung cancer patients treated with gefitinib or erlotinib. *Clin Cancer Res* 2006;**12**;3908-3914.
154. Sheng MM, Wang F, Zhao YG *et al.* Comparison of clinical outcomes of patients with non-small-cell lung cancer harbouring epidermal growth factor receptor exon 19 or exon 21 mutations after tyrosine kinase inhibitors treatment: A meta-analysis. *European Journal of Clinical Pharmacology* 2016;**72**;1-11.
155. Kim HR, Lee JC, Kim YC *et al.* Clinical characteristics of non-small cell lung cancer patients who experienced acquired resistance during gefitinib treatment. *Lung Cancer* 2014;**83**;252-258.
156. Maemondo M, Inoue A, Kobayashi K *et al.* Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *New England Journal of Medicine* 2010;**362**;2380-2388.
157. Inoue A, Kobayashi K, Maemondo M *et al.* Updated overall survival results from a randomized phase III trial comparing gefitinib with carboplatin-paclitaxel for chemo-naive non-small cell lung cancer with sensitive EGFR gene mutations (NEJ002). *Annals of Oncology* 2013;**24**;54-59.
158. Kancha RK, von Bubnoff N, Peschel C, Duyster J. Functional analysis of epidermal growth factor receptor (EGFR) mutations and potential implications for EGFR targeted therapy. *Clin Cancer Res* 2009;**15**;460-467.
159. Kaneda T, Hata A, Tomioka H *et al.* Possible differential EGFR-TKI efficacy among exon 19 deletional locations in EGFR-mutant non-small cell lung cancer. *Lung Cancer* 2014;**86**;213-218.
160. Kim GW, Song JS, Choi CM *et al.* Multiple resistant factors in lung cancer with primary resistance to EGFR-TK inhibitors confer poor survival. *Lung Cancer* 2015;**88**;139-146.
161. Lim SM, Kim HR, Cho EK *et al.* Targeted sequencing identifies genetic alterations that confer primary resistance to EGFR tyrosine kinase inhibitor (korean lung cancer consortium). *Oncotarget* 2016;**7**;36311-36320.
162. Yoo SB, Kim YJ, Kim H *et al.* Alteration of the e-cadherin/beta-catenin complex predicts poor response to epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) treatment. *Annals of Surgical Oncology* 2013;**20**;S545-S552.
163. Chang XF, Izumchenko E, Solis LM *et al.* The relative expression of MIG6 and EGFR is associated with resistance to EGFR kinase inhibitors. *Plos One* 2013;**8**;11.
164. Huang WF, Liu AH, Zhao HJ, Dong HM, Liu LY, Cai SX. BIM gene polymorphism lowers the efficacy of EGFR-TKIs in advanced nonsmall cell lung cancer with sensitive EGFR mutations a systematic review and meta-analysis. *Medicine* 2015;**94**;6.

165. Larsen AW, Nissen PH, Meldgaard P, Weber B, Sorensen BS. EGFR CA repeat polymorphism predict clinical outcome in EGFR mutation positive NSCLC patients treated with erlotinib. *Lung Cancer* 2014;**85**;435-441.
166. Riely GJ, Politi KA, Miller VA, Pao W. Update on epidermal growth factor receptor mutations in non-small cell lung cancer. *Clin Cancer Res* 2006;**12**;7232-7241.
167. Chang MH, Ahn HK, Lee J *et al.* Clinical impact of amphiregulin expression in patients with epidermal growth factor receptor (EGFR) wild-type nonsmall cell lung cancer treated with EGFR-tyrosine kinase inhibitors. *Cancer* 2011;**117**;143-151.
168. Yu HA, Arcila ME, Rekhtman N *et al.* Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. *Clin Cancer Res* 2013;**19**;2240-2247.
169. Ji W, Choi CM, Rho JK *et al.* Mechanisms of acquired resistance to EGFR-tyrosine kinase inhibitor in Korean patients with lung cancer. *Bmc Cancer* 2013;**13**;8.
170. Inukai M, Toyooka S, Ito S *et al.* Presence of epidermal growth factor receptor gene T790M mutation as a minor clone in non-small cell lung cancer. *Cancer Res* 2006;**66**;7854-7858.
171. Turke AB, Zejnullahu K, Wu YL *et al.* Preexistence and clonal selection of met amplification in EGFR mutant NSCLC. *Cancer Cell* 2010;**17**;77-88.
172. Lee Y, Lee GK, Lee YS *et al.* Clinical outcome according to the level of preexisting epidermal growth factor receptor T790M mutation in patients with lung cancer harboring sensitive epidermal growth factor receptor mutations. *Cancer* 2014;**120**;2090-2098.
173. Janne PA, Yang JCH, Kim DW *et al.* AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *New England Journal of Medicine* 2015;**372**;1689-1699.
174. Tan DSW, Yom SS, Tsao MS *et al.* The international association for the study of lung cancer consensus statement on optimizing management of EGFR mutation-positive non-small cell lung cancer: Status in 2016. *Journal of Thoracic Oncology* 2016;**11**;946-963.
175. Lu FL, Li SL, Dong B, Zhang SY, Lv C, Yang Y. Identification of lung adenocarcinoma mutation status based on histologic subtype: Retrospective analysis of 269 patients. *Thoracic Cancer* 2016;**7**;17-23.
176. Dong YJ, Cai YR, Zhou LJ *et al.* Association between the histological subtype of lung adenocarcinoma, EGFR/KRAS mutation status and the ALK rearrangement according to the novel IASLC/ATS/ERS classification. *Oncology Letters* 2016;**11**;2552-2558.
177. Sasaki T, Rodig SJ, Chirieac LR, Janne PA. The biology and treatment of EML4-ALK non-small cell lung cancer. *European journal of cancer (Oxford, England : 1990)* 2010;**46**;1773-1780.
178. Shaw AT, Kim DW, Nakagawa K *et al.* Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *New England Journal of Medicine* 2013;**368**;2385-2394.
179. Shaw AT, Ou SHI, Bang YJ *et al.* Crizotinib in ROS1-rearranged non-small-cell lung cancer. *New England Journal of Medicine* 2014;**371**;1963-1971.
180. Jorge SE, Schulman S, Freed JA *et al.* Responses to the multitargeted MET/ALK/ROS1 inhibitor crizotinib and co-occurring mutations in lung adenocarcinomas with MET amplification or MET exon 14 skipping mutation. *Lung Cancer* 2015;**90**;369-374.

181. Paik PK, Drilon A, Fan PD *et al.* Response to MET inhibitors in patients with stage IV lung adenocarcinomas harboring MET mutations causing exon 14 skipping. *Cancer Discov* 2015;**5**;842-849.
182. Camidge DR, Ou SHI, Shapiro G *et al.* Efficacy and safety of crizotinib in patients with advanced c-MET-amplified non-small cell lung cancer (NSCLC). *Journal of Clinical Oncology* 2014;**32**;1.
183. Gainor JF, Varghese AM, Ou SH *et al.* ALK rearrangements are mutually exclusive with mutations in EGFR or KRAS: An analysis of 1,683 patients with non-small cell lung cancer. *Clin Cancer Res* 2013;**19**;4273-4281.
184. Collisson EA, Campbell JD, Brooks AN *et al.* Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014;**511**;543-550.
185. Tanaka H, Hayashi A, Morimoto T *et al.* A case of lung adenocarcinoma harboring EGFR mutation and EML4-ALK fusion gene. *BMC Cancer* 2012;**12**;558.
186. Dai Z, Kelly JC, Meloni-Ehrig A *et al.* Incidence and patterns of ALK fish abnormalities seen in a large unselected series of lung carcinomas. *Mol Cytogenet* 2012;**5**;44.
187. Sullivan I, Planchard D. ALK inhibitors in non-small cell lung cancer: The latest evidence and developments. *Therapeutic Advances in Medical Oncology* 2016;**8**;32-47.
188. Dhillon AS, Hagan S, Rath O, Kolch W. MAP kinase signalling pathways in cancer. *Oncogene* 2007;**26**;3279-3290.
189. Li H, Pan Y, Li Y *et al.* Frequency of well-identified oncogenic driver mutations in lung adenocarcinoma of smokers varies with histological subtypes and graduated smoking dose. *Lung Cancer* 2013;**79**;8-13.
190. Rekhtman N, Ang DC, Riely GJ, Ladanyi M, Moreira AL. KRAS mutations are associated with solid growth pattern and tumor-infiltrating leukocytes in lung adenocarcinoma. *Mod Pathol* 2013;**26**;1307-1319.
191. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* 2002;**21**;7435-7451.
192. Riely GJ, Kris MG, Rosenbaum D *et al.* Frequency and distinctive spectrum of KRAS mutations in never smokers with lung adenocarcinoma. *Clin Cancer Res* 2008;**14**;5731-5734.
193. Arcila ME, Chaft JE, Nafa K *et al.* Prevalence, clinicopathologic associations, and molecular spectrum of ERBB2 (HER2) tyrosine kinase mutations in lung adenocarcinomas. *Clinical Cancer Research* 2012;**18**;4910-4918.
194. Wang R, Zhang Y, Pan YJ *et al.* Comprehensive investigation of oncogenic driver mutations in chinese non-small cell lung cancer patients. *Oncotarget* 2015;**6**;34300-34308.
195. Huncharek M, Muscat J, Geschwind JF. K-RAS oncogene mutation as a prognostic marker in non-small cell lung cancer: A combined analysis of 881 cases. *Carcinogenesis* 1999;**20**;1507-1510.
196. Mascaux C, Iannino N, Martin B *et al.* The role of RAS oncogene in survival of patients with lung cancer: A systematic review of the literature with meta-analysis. *Br J Cancer* 2005;**92**;131-139.
197. Kosaka T, Yatabe Y, Onozato R, Kuwano H, Mitsudomi T. Prognostic implication of EGFR, KRAS, and TP53 gene mutations in a large cohort of Japanese patients with surgically treated lung adenocarcinoma. *J Thorac Oncol* 2009;**4**;22-29.

198. Yoshida T, Ishii G, Goto K *et al.* Solid predominant histology predicts EGFR tyrosine kinase inhibitor response in patients with EGFR mutation-positive lung adenocarcinoma. *J Cancer Res Clin Oncol* 2013;**139**;1691-1700.
199. Heppner GH. Tumor heterogeneity. *Cancer Res* 1984;**44**;2259-2265.
200. Heppner GH. Cancer cell societies and tumor progression. *Stem Cells* 1993;**11**;199-203.
201. Slack MD, Martinez ED, Wu LF, Altschuler SJ. Characterizing heterogeneous cellular responses to perturbations. *Proc Natl Acad Sci U S A* 2008;**105**;19306-19311.
202. Sardari Nia P, Van Marck E, Van Schil P. The prospect of biologic staging of non-small-cell lung cancer. *Clin Lung Cancer* 2005;**6**;217-224.
203. Brock A, Chang H, Huang S. Non-genetic heterogeneity--a mutation-independent driving force for the somatic evolution of tumours. *Nat Rev Genet* 2009;**10**;336-342.
204. Voorhoeve PM. Micrornas: Oncogenes, tumor suppressors or master regulators of cancer heterogeneity? *Biochim Biophys Acta* 2010;**1805**;72-86.
205. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* 2011;**11**;726-734.
206. Macdonald C, Michael A, Colston K, Mansi J. Heterogeneity of immunostaining for tumour markers in non-small cell lung carcinoma. *European Journal of Cancer* 2004;**40**;461-466.
207. Bozzetti C, Franciosi V, Crafa P *et al.* Biological variables in non-small cell lung cancer: Comparison between immunocytochemical determination on fine needle aspirates from surgical specimens and immunohistochemical determination on tissue sections. *Lung Cancer* 2000;**29**;33-41.
208. Werner M, Chott A, Fabiano A, Battifora H. Effect of formalin tissue fixation and processing on immunohistochemistry. *American Journal of Surgical Pathology* 2000;**24**;1016-1019.
209. Taniguchi K, Okami J, Kodama K, Higashiyama M, Kato K. Intratumor heterogeneity of epidermal growth factor receptor mutations in lung cancer and its correlation to the response to gefitinib. *Cancer Sci* 2008;**99**;929-935.
210. Jiang SX, Yamashita K, Yamamoto M *et al.* EGFR genetic heterogeneity of nonsmall cell lung cancers contributing to acquired gefitinib resistance. *Int J Cancer* 2008;**123**;2480-2486.
211. Nakano H, Soda H, Takasu M *et al.* Heterogeneity of epidermal growth factor receptor mutations within a mixed adenocarcinoma lung nodule. *Lung Cancer* 2008;**60**;136-140.
212. Bai H, Wang Z, Wang Y *et al.* Detection and clinical significance of intratumoral EGFR mutational heterogeneity in chinese patients with advanced non-small cell lung cancer. *PLoS One* 2013;**8**;e54170.
213. Tomonaga N, Nakamura Y, Yamaguchi H *et al.* Analysis of intratumor heterogeneity of EGFR mutations in mixed type lung adenocarcinoma. *Clin Lung Cancer* 2013;**14**;521-526.
214. Fujita Y, Suda K, Kimura H *et al.* Highly sensitive detection of EGFR T790M mutation using colony hybridization predicts favorable prognosis of patients with lung cancer harboring activating EGFR mutation. *J Thorac Oncol* 2012;**7**;1640-1644.
215. Marchetti A, Del Grammastro M, Filice G *et al.* Complex mutations & subpopulations of deletions at exon 19 of EGFR in NSCLC revealed by next generation sequencing: Potential clinical implications. *PLoS One* 2012;**7**;e42164.

216. Yoo SB, Chung JH, Lee HJ, Lee CT, Jheon S, Sung SW. Epidermal growth factor receptor mutation and p53 overexpression during the multistage progression of small adenocarcinoma of the lung. *J Thorac Oncol* 2010;**5**;964-969.
217. Jakobsen JN, Sorensen JB. Intratumor heterogeneity and chemotherapy-induced changes in EGFR status in non-small cell lung cancer. *Cancer Chemotherapy and Pharmacology* 2012;**69**;289-299.
218. Yatabe Y, Matsuo K, Mitsudomi T. Heterogeneous distribution of EGFR mutations is extremely rare in lung adenocarcinoma. *J Clin Oncol* 2011;**29**;2972-2977.
219. Mattsson JS, Imgenberg-Kreuz J, Edlund K, Botling J, Micke P. Consistent mutation status within histologically heterogeneous lung cancer lesions. *Histopathology* 2012;**61**;744-748.
220. Yu J, Kane S, Wu J *et al.* Mutation-specific antibodies for the detection of EGFR mutations in non-small-cell lung cancer. *Clinical Cancer Research* 2009;**15**;3023-3028.
221. Chen Z, Liu H-b, Yu C-h, Wang Y, Wang L, Song Y. Diagnostic value of mutation-specific antibodies for immunohistochemical detection of epidermal growth factor receptor mutations in non-small cell lung cancer: A meta-analysis. *Plos One* 2014;**9**.
222. Zhang RG, Li Y, Nie X, Dong XR, Wu G. Prognostic implications of immunohistochemistry markers for EGFR-TKI therapy in Chinese patients with advanced lung adenocarcinoma harboring EGFR mutations. *Oncotargets and Therapy* 2016;**9**;355-366.
223. Kitamura A, Hosoda W, Sasaki E, Mitsudomi T, Yatabe Y. Immunohistochemical detection of EGFR mutation using mutation-specific antibodies in lung cancer. *Clinical Cancer Research* 2010;**16**;3349-3355.
224. Cinegaglia NC, Andrade SCS, Tokar T *et al.* Integrative transcriptome analysis identifies deregulated microrna-transcription factor networks in lung adenocarcinoma. *Oncotarget* 2016;**7**;28920-28934.
225. Wang L, Chen ZH, An L *et al.* Analysis of long non-coding RNA expression profiles in non-small cell lung cancer. *Cellular Physiology and Biochemistry* 2016;**38**;2389-2400.
226. Feng AL, Tu ZB, Yin BJ. The effect of HMGB1 on the clinicopathological and prognostic features of non-small cell lung cancer. *Oncotarget* 2016;**7**;20507-20519.
227. Bao LM, Zhang Y, Wang J *et al.* Variations of chromosome 2 gene expressions among patients with lung cancer or non-cancer. *Cell Biology and Toxicology* 2016;**32**;419-435.
228. Girard L, Rodriguez-Canales J, Behrens C *et al.* An expression signature as an aid to the histologic classification of non-small cell lung cancer. *Clinical Cancer Research* 2016;**22**;4880-4889.
229. Shen C, Wang YQ, Wei P, Du X. BRCA1-associated protein 1 deficiency in lung adenocarcinoma predicts poor outcome and increased tumor invasion. *Bmc Cancer* 2016;**16**;13.
230. Jiang W, Fan H, Qian C, Ding JY, Wang Q, Pang XG. Prognostic value of high FOXC2 expression in resectable non-small cell lung cancer, alone or in combination with e-cadherin expression. *Bmc Cancer* 2016;**16**;9.
231. Sun LC, Wang YP, Yuan HB *et al.* CPA4 is a novel diagnostic and prognostic marker for human non-small-cell lung cancer. *Journal of Cancer* 2016;**7**;1197-1204.

232. Freidin MB, Bhudia N, Lim E, Nicholson AG, Cookson WO, Moffatt MF. Impact of collection and storage of lung tumor tissue on whole genome expression profiling. *Journal of Molecular Diagnostics* 2012;**14**;140-148.
233. Blackhall FH, Pintilie M, Wigle DA *et al.* Stability and heterogeneity of expression profiles in lung cancer specimens harvested following surgical resection. *Neoplasia* 2004;**6**;761-767.
234. Toi M, Iwata H, Yamanaka T *et al.* Clinical significance of the 21-gene signature (Oncotype DX) in hormone receptor-positive early stage primary breast cancer in the Japanese population. *Cancer* 2010;**116**;3112-3118.
235. Carbone DP, Ding K, Roder H *et al.* Prognostic and predictive role of the Veristrat plasma test in patients with advanced non-small-cell lung cancer treated with erlotinib or placebo in the NCIC clinical trials group BR.21 trial. *Journal of Thoracic Oncology* 2012;**7**;1653-1660.
236. Cuzick J, Swanson GP, Fisher G *et al.* Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: A retrospective study. *Lancet Oncology* 2011;**12**;245-255.
237. Dancik GM, Theodorescu D. Robust prognostic gene expression signatures in bladder cancer and lung adenocarcinoma depend on cell cycle related genes. *Plos One* 2014;**9**;9.
238. Wistuba II, Behrens C, Lombardi F *et al.* Validation of a proliferation-based expression signature as prognostic marker in early stage lung adenocarcinoma. *Clin Cancer Res* 2013;**19**;6261-6271.
239. Eguchi T, Kadota K, Chaft J *et al.* Cell cycle progression score is a marker for five-year lung cancer-specific mortality risk in patients with resected stage I lung adenocarcinoma. *Oncotarget* 2016;**7**;35241-35256.
240. Okada A, Shimmyo T, Hashimoto T *et al.* Predictive advantage of a cell type classification for pulmonary adenocarcinoma coupled with data for p53, K-RAS and EGFR alterations. *Cancer Science* 2010;**101**;1745-1753.
241. Guerrera F, Errico L, Evangelista A *et al.* Exploring stage I non-small-cell lung cancer: Development of a prognostic model predicting 5-year survival after surgical resection. *European Journal of Cardio-Thoracic Surgery* 2015;**47**;1037-1043.
242. Starnes SL, Pathrose P, Wang J *et al.* Clinical and molecular predictors of recurrence in stage I non-small cell lung cancer. *Annals of Thoracic Surgery* 2012;**93**;1606-1613.
243. von der Thusen JH, Tham YS, Pattenden H *et al.* Prognostic significance of predominant histologic pattern and nuclear grade in resected adenocarcinoma of the lung potential parameters for a grading system. *Journal of Thoracic Oncology* 2013;**8**;37-44.
244. Lopez-Encuentra A, Lopez-Rios F, Conde E *et al.* Composite anatomical-clinical-molecular prognostic model in nonsmall cell lung cancer. *European Respiratory Journal* 2011;**37**;136-142.
245. Yatabe Y, Kerr KM, Utomo A *et al.* EGFR mutation testing practices within the Asia pacific region: Results of a multicenter diagnostic survey. *J Thorac Oncol* 2015;**10**;438-445.
246. Chamberlin NE, Ward AC. EGFR mutation prevalence in patients with non-small cell lung cancer: An audit of testing within the North of England cancer network. *Thorax* 2011;**66**;A134-A134.

247. Consortium SM. Detailed advice: erlotinib (Tarceva®) for the first-line treatment of patients with locally advanced or metastatic non-small cell lung cancer (NSCLC) with epidermal growth factor receptor (EGFR) activating mutations. Glasgow, 2011.
248. Consortium SM. Detailed advice: crizotinib (Xalkori) for the treatment of adults with previously treated anaplastic lymphoma kinase (ALK)-positive advanced non-small cell lung cancer (NSCLC). Glasgow: Scottish Medicines Consortium, 2013.
249. EMA. Points to consider on multiplicity issues in clinical trials. London: European Medicines Agency, 2002.
250. Ranstam J. Multiple p-values and bonferroni correction. *Osteoarthritis and Cartilage* 2016;**24**;763-764.
251. Bender R, Lange S. Adjusting for multiple testing - when and how? *Journal of Clinical Epidemiology* 2001;**54**;343-349.
252. Dasgupta A, Sun YV, Konig IR, Bailey-Wilson JE, Malley JD. Brief review of regression-based and machine learning methods in genetic epidemiology: The genetic analysis workshop 17 experience. *Genetic Epidemiology* 2011;**35**;S5-S11.
253. Song Y-y, Lu Y. Decision tree methods: Applications for classification and prediction *Shanghai Archives of Psychiatry* 2015;**27**;130-135.
254. Wang ZY, Wen XT, Lu YH, Yao Y, Zhao H. Exploiting machine learning for predicting skeletal-related events in cancer patients with bone metastases. *Oncotarget* 2016;**7**;12612-12622.
255. Saleh DT, Attia A, Shaker O, Ieee. Studying combined breast cancer biomarkers using machine learning techniques. *2016 Ieee 14th International Symposium on Applied Machine Intelligence and Informatics (Sami)* 2016;247-251.
256. Gamez-Pozo A, Sanchez-Navarro I, Nistal M *et al.* Maldi profiling of human lung cancer subtypes. *Plos One* 2009;**4**;6.
257. Wang CW, Yu CP. Automated morphological classification of lung cancer subtypes using H&E tissue images. *Machine Vision and Applications* 2013;**24**;1383-1391.
258. Liu Y, Lin DM, Xiao T *et al.* An immunohistochemical analysis-based decision tree model for estimating the risk of lymphatic metastasis in pN0 squamous cell carcinomas of the lung. *Histopathology* 2011;**59**;882-891.
259. Chen HY, Yu SL, Chen CH *et al.* A five-gene signature and clinical outcome in non-small-cell lung cancer. *New England Journal of Medicine* 2007;**356**;11-20.
260. Nitadori J, Bograd AJ, Kadota K *et al.* Impact of micropapillary histologic subtype in selecting limited resection vs lobectomy for lung adenocarcinoma of 2cm or smaller. *Jnci-Journal of the National Cancer Institute* 2013;**105**;1212-1220.
261. Travis WD, Brambilla E, Noguchi M *et al.* International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society: International multidisciplinary classification of lung adenocarcinoma: Executive summary. *Proc Am Thorac Soc* 2011;**8**;381-385.
262. Do H, Krypuy M, Mitchell PL, Fox SB, Dobrovic A. High resolution melting analysis for rapid and sensitive EGFR and KRAS mutation detection in formalin fixed paraffin embedded biopsies. *BMC Cancer* 2008;**8**;142.
263. Bueno R, Hughes E, Wagner S *et al.* Validation of a molecular and pathological model for five-year mortality risk in patients with early stage lung adenocarcinoma. *J Thorac Oncol* 2015;**10**;67-73.

264. Chatziandreou I, Tsioli P, Sakellariou S *et al.* Comprehensive molecular analysis of NSCLC; clinicopathological associations. *PLoS One* 2015;**10**;e0133859.
265. Pan W, Yang Y, Zhu HC, Zhang YC, Zhou RP, Sun XC. KRAS mutation is a weak, but valid predictor for poor prognosis and treatment outcomes in NSCLC: A meta-analysis of 41 studies. *Oncotarget* 2016;**7**;8373-8388.
266. Walsh K, Wallace WA, Butler R *et al.* A cautionary lesson on the use of targeted methods for egfr mutation analysis: A case report. *J Clin Pathol* 2014;**67**;734-735.
267. Skov BG, Hogdall E, Clementsen P e al. The prevelence of EGFR muations in non-small cell lung cancer in an unselected Caucasion population. APMIS 2015;123;108-115.
268. Walsh K, Wallace WA. Molecular pathology in lung cancer: A guide to the techniques used in clinical practice. *Histopathology* 2014;**65**;731-741.
269. Skov BG, Høgdall E, Clementsen P *et al.* The prevalence of EGFR mutations in non-small cell lung cancer in an unselected caucasian population. *APMIS* 2015;**123**;108-115.
270. Paik PK, Johnson ML, D'Angelo SP *et al.* Driver mutations determine survival in smokers and never-smokers with stage IIIB/IV lung adenocarcinomas. *Cancer* 2012;**118**;5840-5847.
271. Walsh K, Kheng YC, Oniscu A, Harrison DJ, Wallace WA. Could molecular pathology testing in lung cancer be more cost-effective? *J Clin Pathol* 2016.
272. Vallee A, Sagan C, Le Loupp AG, Bach K, Dejoie T, Denis MG. Detection of EGFR gene mutations in non-small cell lung cancer: Lessons from a single-institution routine analysis of 1,403 tumor samples. *Int J Oncol* 2013;**43**;1045-1051.
273. Vincenten J, Smit EF, Vos W *et al.* Negative NKX2-1 (TTF-1) as temporary surrogate marker for treatment selection during EGFR-mutation analysis in patients with non-small-cell lung cancer. *J Thorac Oncol* 2012;**7**;1522-1527.
274. Krawczyk P, Ramlau R, Chorostowska-Wynimko J *et al.* The efficacy of EGFR gene mutation testing in various samples from non-small cell lung cancer patients: A multicenter retrospective study. *J Cancer Res Clin Oncol* 2015;**141**;61-68.
275. Sheffield BS, Bosdet IE, Ali RH *et al.* Relationship of thyroid transcription factor 1 to EGFR status in non-small-cell lung cancer. *Curr Oncol* 2014;**21**;305-308.
276. Somaiah N, Fidler MJ, Garrett-Mayer E *et al.* Epidermal growth factor receptor (EGFR) mutations are exceptionally rare in thyroid transcription factor (TTF-1)-negative adenocarcinomas of the lung. *Oncoscience* 2014;**1**;522-528.
277. Zhang Y, Wang R, Li Y *et al.* Negative thyroid transcription factor 1 expression defines an unfavorable subgroup of lung adenocarcinomas. *J Thorac Oncol* 2015;**10**;1444-1450.
278. Shanzhi W, Yiping H, Ling H, Jianming Z, Qiang L. The relationship between TTF-1 expression and EGFR mutations in lung adenocarcinomas. *PLoS One* 2014;**9**;e95479.
279. Chung KP, Huang YT, Chang YL *et al.* Clinical significance of thyroid transcription factor-1 in advanced lung adenocarcinoma under epidermal growth factor receptor tyrosine kinase inhibitor treatment. *Chest* 2012;**141**;420-428.
280. Koh Y, Kim DW, Kim TM *et al.* Clinicopathologic characteristics and outcomes of patients with anaplastic lymphoma kinase-positive advanced pulmonary adenocarcinoma: Suggestion for an effective screening strategy for these tumors. *J Thorac Oncol* 2011;**6**;905-912.

281. Kadota K, Yeh YC, Sima CS *et al.* 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* 2014;**27**;690-700.
282. Warth A, Muley T, Kossakowski C *et al.* Prognostic impact and clinicopathological correlations of the cribriform pattern in pulmonary adenocarcinoma. *Journal of Thoracic Oncology* 2015;**10**;638-644.
283. Xu LR, Tavora F, Burke A. Histologic features associated with metastatic potential in invasive adenocarcinomas of the lung. *American Journal of Surgical Pathology* 2013;**37**;1100-1108.
284. Yamaguchi Y, Ishii G, Kojima M *et al.* Histopathologic features of the tumor budding in adenocarcinoma of the lung tumor budding as an index to predict the potential aggressiveness. *Journal of Thoracic Oncology* 2010;**5**;1361-1368.
285. Travis WD, Brambilla E, Riely GJ. New pathologic classification of lung cancer: Relevance for clinical practice and clinical trials. *Journal of Clinical Oncology* 2013;**31**;992-1001.
286. Warth A, Stenzinger A, von Brünneck AC *et al.* Interobserver variability in the application of the novel IASLC/ATS/ERS classification for pulmonary adenocarcinomas. *Eur Respir J* 2012;**40**;1221-1227.
287. Makinen JM, Laitakari K, Johnson S *et al.* Nonpredominant lepidic pattern correlates with better outcome in invasive lung adenocarcinoma. *Lung Cancer* 2015;**90**;568-574.
288. Warth A, Cortis J, Fink L *et al.* Training increases concordance in classifying pulmonary adenocarcinomas according to the novel IASLC/ATS/ERS classification. *Virchows Archiv* 2012;**461**;185-193.
289. Hwang I, Park KU, Kwon KY. Modified histologic classification as a prognostic factor in pulmonary adenocarcinoma. *International Journal of Surgical Pathology* 2014;**22**;212-220.
290. Aida S, Tamai S, Sekiguchi S, Shimizu N. Distribution of epidermal growth factor and epidermal growth factor receptor in human lung: Immunohistochemical and immunoelectron-microscopic studies. *Respiration* 1994;**61**;161-166.
291. Kozu Y, Tsuta K, Kohno T *et al.* The usefulness of mutation-specific antibodies in detecting epidermal growth factor receptor mutations and in predicting response to tyrosine kinase inhibitor therapy in lung adenocarcinoma. *Lung Cancer* 2011;**73**;45-50.
292. Azuma K, Okamoto I, Kawahara A *et al.* Association of the expression of mutant epidermal growth factor receptor protein as determined with mutation-specific antibodies in non-small cell lung cancer with progression-free survival after gefitinib treatment. *Journal of Thoracic Oncology* 2012;**7**;122-127.
293. Zhao J, Wang X, Xue L *et al.* The use of mutation-specific antibodies in predicting the effect of EGFR-TKIS in patients with non-small-cell lung cancer. *Journal of Cancer Research and Clinical Oncology* 2014;**140**;849-857.
294. Kim CH, Kim SH, Park SY, Yoo J, Kim SK, Kim HK. Identification of EGFR mutations by immunohistochemistry with EGFR mutation-specific antibodies in biopsy and resection specimens from pulmonary adenocarcinoma. *Cancer Res Treat* 2015;**47**;653-660.

295. Xiong Y, Bai Y, Leong N *et al.* Immunohistochemical detection of mutations in the epidermal growth factor receptor gene in lung adenocarcinomas using mutation-specific antibodies. *Diagnostic Pathology* 2013;**8**.
296. Kato Y, Peled N, Wynes MW *et al.* Novel epidermal growth factor receptor mutation-specific antibodies for non-small cell lung cancer immunohistochemistry as a possible screening method for epidermal growth factor receptor mutations. *Journal of Thoracic Oncology* 2010;**5**;1551-1558.
297. Cuzick J. Prognostic value of a cell cycle progression score for men with prostate cancer. *Recent Results Cancer Res* 2014;**202**;133-140.
298. Rakha E, Pajares MJ, Ilie M *et al.* Stratification of resectable lung adenocarcinoma by molecular and pathological risk estimators. *Eur J Cancer* 2015;**51**;1897-1903.
299. Manac'h D, Riquet M, Medioni J, Le Pimpec-Barthes F, Dujon A, Danel C. Visceral pleura invasion by non-small cell lung cancer: An underrated bad prognostic factor. *Annals of Thoracic Surgery* 2001;**71**;1088-1093.
300. Kang JH, Kim KD, Chung KY. Prognostic value of visceral pleura invasion in non-small cell lung cancer. *European Journal of Cardio-Thoracic Surgery* 2003;**23**;865-869.
301. Yilmaz A, Duyar SS, Cakir E *et al.* Clinical impact of visceral pleural, lymphovascular and perineural invasion in completely resected non-small cell lung cancer. *European Journal of Cardio-Thoracic Surgery* 2011;**40**;664-670.
302. Lynch J, Pattekar R, Barnes DM *et al.* Mitotic counts provide additional prognostic information in grade II mammary carcinoma. *J Pathol* 2002;**196**;275-279.
303. Coindre JM. Grading of soft tissue sarcomas: Review and update. *Arch Pathol Lab Med* 2006;**130**;1448-1453.
304. Yu KH, Zhang C, Berry GJ *et al.* Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun* 2016;**7**;12474.
305. Chief medical officer's annual report 2014-15: Realistic medicine. Edinburgh, 2016.
306. Magee ND, Beattie JR, Carland C *et al.* Raman microscopy in the diagnosis and prognosis of surgically resected nonsmall cell lung cancer. *Journal of Biomedical Optics* 2010;**15**;8.
307. Caie PD, Zhou Y, Turnbull AK, Oniscu A, Harrison DJ. Novel histopathologic feature identified through image analysis augments stage II colorectal cancer clinical reporting. *Oncotarget* 2016;**7**;44381-44394.
308. Ing N, Salman S, Ma ZX, Walts A, Knudsen B, Gertych A. Machine learning can reliably distinguish histological patterns of micropapillary and solid lung adenocarcinomas. In Pietka E, Badura P, Kawa J, Wieclawek W eds. *Information technologies in medicine*. Berlin: Springer-Verlag Berlin, 2016;193-206.
309. Frey L, Edgerton ME, Fisher DH, Tang L, Chen Z. Using prior knowledge and rule induction methods to discover molecular markers of prognosis in lung cancer. *AMIA Annual Symposium proceedings. AMIA Symposium* 2005;256-260.