**Molecular biology of lung cancer**

Wendy A. Cooper\(^1\), David C. L. Lam\(^3\), Sandra A. O'Toole\(^1\),\(^4\), John D. Minna\(^6\)

\(^1\)Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Camperdown, Sydney, NSW, Australia; \(^2\)School of Medicine, University of Western Sydney, NSW, Australia; \(^3\)Department of Medicine, University of Hong Kong, Hong Kong SAR, China; \(^4\)Kinghorn Cancer Centre and Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW, Australia; \(^5\)Sydney Medical School, University of Sydney, NSW, Australia; \(^6\)Hamon Center for Therapeutic Oncology Research, University of Texas Southwestern Medical Center at Dallas, Texas, USA

**ABSTRACT**

Lung cancers are characterised by abundant genetic diversity with relatively few recurrent mutations occurring at high frequency. However, the genetic alterations often affect a common group of oncogenic signalling pathways. There have been vast improvements in our understanding of the molecular biology that underpins lung cancer in recent years and this has led to a revolution in the diagnosis and treatment of lung adenocarcinomas (ADC) based on the genotype of an individual’s tumour. New technologies are identifying key and potentially targetable genetic aberrations not only in adenocarcinoma but also in squamous cell carcinoma (SCC) of the lung. Lung cancer mutations have been identified in v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS), epidermal growth factor receptor (EGFR), BRAF and the parallel phosphatidylinositol 3-kinase (PI3K) pathway oncogenes and more recently in MEK and HER2 while structural rearrangements in ALK, ROS1 and possibly rearranged during transfection (RET) provide new therapeutic targets. Amplification is another mechanism of activation of oncogenes such as MET in adenocarcinoma, fibroblast growth factor receptor 1 (FGFR1) and discoidin domain receptor 2 (DDR2) in SCC. Intriguingly, many of these genetic alternations are associated with smoking status and with particular racial and gender differences, which may provide insight into the mechanisms of carcinogenesis and role of host factors in lung cancer development and progression. The role of tumour suppressor genes is increasingly recognised with aberrations reported in TP53, PTEN, RB1, LKB1 and p16/CDKN2A. Identification of biologically significant genetic alterations in lung cancer that lead to activation of oncogenes and inactivation of tumour suppressor genes has the potential to provide further therapeutic opportunities. It is hoped that these discoveries may make a major contribution to improving outcome for patients with this poor prognosis disease.

**KEY WORDS**

Lung cancer; mutation; molecular pathology; oncogene; tumour suppressor gene

---

**Introduction**

The molecular basis of lung cancer is complex and heterogeneous. Improvements in our understanding of molecular alterations at multiple levels (genetic, epigenetic, protein expression) and their functional significance have the potential to impact lung cancer diagnosis, prognostication and treatment. Lung cancers develop through a multistep process involving development of multiple genetic and epigenetic alterations, particularly activation of growth promoting pathways and inhibition of tumour suppressor pathways. Greater understanding of the multiple biochemical pathways involved in the molecular pathogenesis of lung cancer is crucial to the development of treatment strategies that can target molecular aberrations and their downstream activated pathways (1). Specific molecular alterations that drive tumour growth and provide targets for therapy have been best defined in adenocarcinomas (ADC) but there is increasing interest in the molecular landscape of squamous cell carcinoma (SCC) highlighting new potential therapeutic targets. In lung cancer as in other malignancies, tumourigenesis relates to activation of growth promoting proteins [e.g., v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS), epidermal growth factor receptor (EGFR), BRAF, MEK-1, HER2, MET, ALK and rearranged during transfection (RET)] as well as inactivation of tumour suppressor genes [e.g., P53,
phosphatase with tensin homology (PTEN), LKB-1) (1). Activation of growth promoting oncogenes can occur by gene amplification or other genetic alterations including point mutations and structural rearrangements leading to uncontrolled signalling through oncogenic pathways. “Oncogene addiction” results when cell survival depends on continued activation of the aberrant signalling (2,3) making them ideal candidates for targeted therapies. Oncogenic driver mutations have been identified in over 50% of lung ADC and are almost always exclusive of other driver mutations (4,5). Signalling pathways regulated by oncogenes and tumour suppressor genes are often interconnected with cross-talk between pathways involved in carcinogenesis. Added to the complexity is the occurrence of mutational evolution of tumours over time during the natural course of disease progression and in response to selection pressure exerted by therapy.

There is great genetic diversity in lung cancer and they harbour among the greatest numbers of genetic aberrations of all tumours (1). Understanding of the molecular biology of lung cancer has been revolutionised by next-generation sequencing technologies that provide a comprehensive means of identifying somatic alterations in entire cancer genomes or exomes. Lung cancers have highly complex genomes with a recent large-scale exome sequencing study of 31 non-small cell lung cancer (NSCLC) identifying 727 mutated genes not previously reported in the literature or in the COSMIC database (6). Genomic studies have confirmed previously well known alterations in lung cancer such as KRAS, EGFR and BRAF and have also identified low frequency but recurrent mutations that are novel in lung cancer (6-8) including potentially targetable alterations in JAK2, ERBB4 (8), RET (9-11), fibroblast growth factor receptor 1 (FGFR1) (12) and discoidin domain receptor 2 (DDR2) (13). While these studies provide a comprehensive portrait of genetic alterations in lung cancers, the challenge remains of identifying biologically relevant driver mutations from the vast majority of passenger mutations. The relative paucity of high frequency recurrent mutations highlights the heterogeneity and complexity of the molecular biology of lung cancer with common pathways affected by a range of different genetic alterations that poses a challenge for providing personalised medicine.

In this review, we discuss the most commonly altered and most clinically relevant oncogenes and tumour suppressor genes in lung cancer as improved understanding of the molecular pathology of lung cancer is crucial for advancements in treatment strategies.

**KRAS**

KRAS is part of the RAS family of proto-oncogenes (KRAS, NRAS and HRAS occurring in humans) and encodes a G-protein with a critical role in controlling signal transduction pathways which regulate cell proliferation, differentiation and survival (14). Ras proteins are guanosine diphosphate (GDP) bound and inactive in normal quiescent cells. There is a switch to the activated guanosine triphosphate (GTP) bound form following activation of upstream growth factor receptors. The activated Ras-GTP subsequently binds and activates a number of downstream pathways including mitogen-activated protein kinase (MAPK), RAS/RAF/MEK/MAPK pathway and the PI3-K [PI3K/AKT/mammalian target of rapamycin (mTOR)] pathways (15). KRAS plays a critical role in downstream signal transduction induced by a variety of growth factor receptors including EGFR and constitutive activation of the protein circumvents the need for growth factor mediated signalling. Activating mutations alter the GTPase activity of the protein hindering inactivation of the active RAS-GTP to GDP leading to increased signalling through multiple downstream growth promoting pathways (15). The RAS/RAF/MEK/MAPK signal transduction cascade plays a central role in many lung cancers with at least one mutation in the pathway identified in 132 of 188 tumours (7), of which the most common are mutations in KRAS.

Activating mutations in the KRAS oncogene are the commonest oncogenic alteration in lung ADC occurring in about 25-40% of cases (4,5,7,16-18) while HRAS and NRAS mutations are very rare (17). Differences in the prevalence of KRAS mutations in lung ADC most likely relate to different patient populations as KRAS mutations are more common in Western populations compared to Asian populations (19-22) and are more frequent in males and smokers (7,18,22). ADC in never smokers have been reported to harbour KRAS mutations in between 0-15% of cases (16,23). In addition, KRAS mutations are very rare or absent in SCCs and small cell cancer (17,24). Comprehensive genomic analysis of 188 SCCs identified only 1 KRAS mutation in codon 61 (12). KRAS mutations in lung adenocarcinoma consist of single amino acid substitutions in hotspots located mostly in codon 12 but also more rarely in codons 13 and 61 (14,17). The commonest mutations in KRAS are G to T transversions (~84%) in smokers while never smokers are more likely to harbour G to A transitions (16).

In keeping with the role of KRAS alterations as driver mutations, they do not occur in association with EGFR mutations (5,7,21,22), although rare exceptions do occur (18). A meta-analysis has shown KRAS mutant tumours are resistant to EGFR tyrosine kinase inhibitors (TKIs) (25), as KRAS mutations lead to constitutive activation of pathways downstream of EGFR. There is evidence that different KRAS mutant proteins have differing clinical significance. Interestingly, using data from the BATTLE trial (prospective phase II Biomarker-integrated Approaches of Targeted Therapy for Lung cancer Elimination), either G12C or G12V mutant KRAS predicted shorter progression free survival compared to other KRAS mutations or wild type KRAS (26). Furthermore, different
Amino acid substitutions were associated with activation of different pathways (PI3-K and MEK with Gly12Asp and Ras with Gly12Cys or mutant Gly12Val) resulting from divergent protein conformations from specific mutations leading to altered ability to associate with downstream protein mediators (26). This highlights that appropriate use of targeted therapies and clinical trial design needs to carefully evaluate the clinical and therapeutic significance of specific genetic alterations in lung cancer. The high frequency of KRAS mutations in lung cancer makes it an ideal therapeutic target but unfortunately clinical trials of targeted agents have generally been disappointing.

**EGFR**

Alterations of EGFR are involved in the pathogenesis of many tumours including NSCLC. EGFR encodes a transmembrane tyrosine kinase with an extracellular ligand-binding domain and an intracellular component including a tyrosine kinase domain (27). Binding of the ligand epidermal growth factor leads to receptor homo or heterodimerisation with other members of the EGFR family and activation of the tyrosine kinase domain (28,29). Signal transduction stimulated by EGFR occurs through the PI3K/AKT/mTOR, RAS/RAF//MAPK and JAK/STAT signalling pathways (28-30). EGFR is involved in regulation of numerous oncogenic functions such as cell proliferation, survival, differentiation, neovascularisation, invasion and metastasis (29,30). Activating mutations in EGFR lead to constitutive tyrosine kinase activation (30,31) and oncogenic transformation of lung epithelial cells *in vitro* (31). A transgenic mouse model with inducible expression of the commonest EGFR mutations showed development of multiple lung ADC that were sensitive to small molecule inhibition (32). Other mechanisms of increased EGFR signalling include increased protein expression or increased gene copy number (33,34).

Activating mutations of EGFR have been reported in 10-15% of unselected Western patients (5,21,35,36) and 30-40% of Asian populations (19,37,38). Differences in the reported prevalence rates of various mutations may in part relate to different patient populations, but also depends on the sensitivity of mutation analysis techniques utilised in different studies. In NSCLC, EGFR mutations occur in the first four exons of the intracellular tyrosine kinase domain, most commonly exon 19 in frame deletions (~45%), of which there are over 20 variants, the commonest being delE746-A750. The next commonest EGFR mutations are missense mutations, particularly L858R, a single nucleotide point mutation in exon 21 leading to a single amino acid change from leucine to arginine at codon 858 (~40%). However, we found in an Australian population that exon 18 activating mutations constituted 14% of EGFR mutations in patients with early stage lung cancer and L858R mutations comprised only 29% of EGFR mutations present in this cohort (5).

There are also a range of less common mutations including in frame duplications or insertions in exon 20 (~5-10%), of which there are many variants that are often associated with resistance to EGFR TKIs (22,39).

In lung cancer, almost all EGFR mutations occur in ADC (19,21,40,41) although they may also be seen in adenosquamous carcinomas. Mutations in EGFR are more commonly but not exclusively found in patients who are female, younger and with no history of smoking (7,19,21,22,37,40). EGFR mutations occur only very rarely, in histologically well sampled pure SCCs (24,42). However, comprehensive genomic analysis of 188 SCCs identified EGFR mutations in 2 cases, both with L861G mutations (12). While EGFR mutations are very rare in SCCs, variant-III mutations involving the extracellular domain of EGFR, copy-number gains and protein overexpression are more common in SCCs than in ADCs (43).

Secondary mutations in EGFR develop or are clonally selected in patients that develop resistance to EGFR TKIs, the commonest being the T790M activating point mutation in exon 20 which substitutes a ‘bulkier’ methionine for threonine (44) that interferes with binding of reversible TKIs. T790M is found in about 50% of tumours from patients who develop acquired TKI resistance (41,44). Intriguingly, we observed that exon 20 mutations including T790M mutations associated with therapeutic resistance to EGFR TKI were seen in 29% of patients with EGFR mutations in a therapy naïve cohort (5). Activation of downstream pathways that bypass EGFR inhibition can also contribute to EGFR-TKI resistance including activation of PI3K pathway through amplification of MET (45).

**BRAF**

BRAF encodes a serine/threonine protein kinase that is the downstream effector protein of KRAS and activates the MAPK signal transduction pathway involved in regulation of cell proliferation and survival (46). Upon activation, BRAF phosphorylates downstream mediators MEK1 and MEK2 which subsequently activate ERK1 and ERK2, involved in regulation of growth regulating proteins such as c-JUN and ELK1 (14). Activating mutations in BRAF lead to increased kinase activity that exhibit transforming activity *in vitro* (46).

While activating BRAF mutations are common in melanoma (46), they occur in only about 3% of NSCLC (18,46-50). The mutations in NSCLC differ to those in melanoma and colorectal carcinoma with a lower proportion of V600E mutations that affect the kinase domain of the protein. In lung ADC, V600E mutations in exon 15 account for up to about 50% of BRAF mutations followed by G469A in exon 11 and D594G in exon 15 (48,50). Some of the BRAF mutations in NSCLC occur in the kinase domain (such as V600E, D594G and L596R) while others occur in the G-loop of the activation domain of the
gene (such as G46SV and G468A) (46). As *BRAF* and *KRAS* genes are part of the signalling pathway mediated by *EGFR*, it is not surprising that mutations in these genes are almost always mutually exclusive, in keeping with a common downstream pathway to transformation. *BRAF* mutations in lung cancer occur almost always in ADC (48). Non-V600E *BRAF* mutations have been associated with current or former smokers while V600E mutations appear to be more common in female never smokers (48,50). While uncommon, *BRAF* mutations represent an important therapeutic target due to the availability of targeted therapies already in clinical use for melanoma although there is only limited data about the clinical response to this approach in NSCLC (51).

**MEK**

MEK1 (also known as MAPK1) is a serine-threonine kinase that has an important function as a downstream target of RAS activation. MEK1 activates MAPK2 and MAPK3 downstream of *BRAF* (14). Rare cases of somatic mutations of MEK1 have been reported in NSCLC with 2 of 107 lung ADC found to have an activating mutation in exon 2 that did not involve the kinase domain (52). The mutations were exclusive of other driver mutations and were associated with gain of function *in vitro* (52).

**MET**

The proto-oncogene *MET* located on chromosome 7q21-q31 encodes a membrane tyrosine kinase receptor that is also known as *hepatocyte growth factor receptor* (53). Upon binding of its ligand hepatocyte growth factor, there is receptor homodimerisation, kinase activation and signalling through downstream pathways including RAS/RAF/MEK/MAPK, PI3K/AKT and c-SRC kinase pathways (53). In NSCLC, *MET* is altered by gene amplification in about 1-7% of treatment naive patients (54-57) but in one study amplification was found in 21% of patients (58). Increased *MET* copy number may be more common in SCC than ADC (57) and is mutually exclusive with *KRAS* mutations (56,58). *MET* amplification results in overexpression of MET protein and activation of downstream signal transduction pathways. The oncogenic activity of *MET* has been demonstrated *in vitro* with evidence of gene amplification associated with constitutive receptor phosphorylation, activation of the PI3K/AKT pathway and sensitivity to MET inhibition (45,59). Amplification of *MET* is a known mechanism of secondary EGFR-TKI resistance with this kinase switch occurring in approximately 20% of patients with acquired resistance (45,54,55). In this scenario, *MET* amplification drives and maintains the PI3K/AKT pathway bypassing EGFR blockade by TKIs (45), suggesting concomitant MET inhibition may be a means of overcoming TKI resistance. Mutations of *MET* also occur uncommonly in about 3-5% of ADC (7,56).

**HER2**

The human epidermal growth factor receptor 2 (*HER2/ERBB2*) gene encodes a membrane bound receptor tyrosine kinase that is a member of the ERBB family of receptors, along with EGFR. Unlike other ERBB receptors, it does not bind ligand directly but can form heterodimers with other ligand-bound members of the receptor family (60). Activation leads to signalling through a variety of signal transduction pathways including PI3K, MAPK and JAK/STAT pathways (61). Activation of HER2 occurs in a small proportion of lung cancers with overexpression in approximately 20% of cases, gene amplification in 2% (62) and activating mutations in 1.6-4% of NSCLC (63-65). Activating mutations of HER2 are exon 20 in frame insertions of 3 to 12 base pairs in length (63). There is *in vivo* evidence of the oncogenic activity of HER2 with multiple adenocarcinomas developing in a transgenic mouse model expressingmutant HER2 and exhibiting susceptibility to small molecule inhibition (66). Alterations of HER2 occur mostly in ADC (63-65) and mutations occur in tumours that are wild-type for *EGFR* and *KRAS* (63,64) and in some studies, are associated with female gender, Asian ethnicity and non-smoking status (63,65), similar to the clinical profile of *EGFR* mutant tumours.

**PI3K/AKT/mTOR**

The PI3K/AKT/mTOR pathway is an important signal transduction pathway involved in regulation of cell proliferation, survival, differentiation adhesion and motility (67,68). Alterations of this pathway have been implicated in both NSCLC and small cell carcinoma (69,70). The pathway is activated through activation of a variety of membrane tyrosine kinase receptors including EGFR, HER2, insulin-like growth factor receptor, vascular endothelial growth factor receptor and platelet derived growth factor receptor (71,72). Activated receptor tyrosine kinases recruit PI3K to the cell membrane where it phosphorylates PIP2 to PIP3 [phosphatidylinositol 4,5-bisphosphate (PIP2) to phosphatidylinositol 3,4,5-trisphosphate]. PIP3 in turn recruits the serine/threonine kinase AKT to the membrane where it is phosphorylated by 3-phosphoinositide-dependent kinase 1 (P13 kinase) and mTOR. mTOR is a serine/threonine kinase that is a downstream target of AKT (72). Activated AKT in turn activates multiple targets including tuberous sclerosis 2 and Bcl-2 associated death promoter leading to cell proliferation and survival [reviewed in (71)]. There is also interaction with other pathways including RAS/RAF/MEK (Rat sarcoma/rapidly accelerated fibrosarcoma/MAPK or Erk kinase) with RAS having the capacity to directly activate PI3K (72).
The PI3K/AKT/mTOR pathway is frequently deregulated in many tumours including 50-70% of NSCLC (7,71). Significant alterations involving the PI3K pathway were identified in 47% of SCCs in the Cancer Genome Atlas project (12). Pathway activation in lung carcinogenesis occurs through a variety of mechanisms including activating mutations in EGFR, KRAS, PI3K or AKT (68,71) as well as PIK3CA amplification, or loss of negative regulation by the tumour suppressor gene PTEN (72).

The PI3K protein family (phosphatidylinositol 3-kinases) are intracellular lipid kinases and the main catalytic subunit, the p110alpha isoform, is encoded by the PIK3CA gene (71). Activating mutations and amplification of PIK3CA cause constitutive ligand-independent pathway activation (73,74). PIK3CA mutations mostly involve the catalytic domain and have been identified in approximately 1-3% of NSCLCs (7,73,75) and are more common in SCC than ADCs (4,75). Unlike most oncogenic driver mutations, PIK3CA mutations may occur in association with EGFR or KRAS mutations (5,73,75) suggesting they may not represent true driver mutations. However, in vitro studies of lung cancer cell lines with PIK3CA mutations or copy number gains show increased PI3 kinase activity sensitive to small molecule inhibition (73) and in vivo mouse models with PIK3CA mutation expression develop numerous ADC, suggesting oncogenic activity (74). PIK3CA may also be amplified in NSCLC, especially in SCCs (73,76) and increased copy number of PIK3CA has been reported in ~5% of small cell carcinoma cell lines (73). Although rare, PI3K/AKT/mTOR pathway activation can also occur through AKT mutations which have been reported in 0.5-2% of NSCLC (5,7,77), particularly SCCs (77).

**ALK**

Rearrangements of the receptor tyrosine kinase ALK resulting most commonly in fusions of the intracellular kinase domain with the amino terminal end of echinoderm microtubule associated protein-like 4 (EML4) occur in a subset of lung cancers (78-80). The rearrangement results from a short inversion in chromosome 2p, whereby in the commonest variant, intron 13 of EML4 is fused to intron 19 of ALK (ALK [inv (2) (p21; p23)]) (79). Numerous variants of EML4-ALK fusions have been identified differing lengths of EML4, the commonest being exons 1-13 of EML4 joining to exons 20-29 of ALK (78,81,82). More recently, different partner genes have been identified in a small subset of ALK rearrangements (<1% of cases) including KIF5B (kinesin family member 5b), TFG (TRK -fused gene) and KLC-1 (kinesin light chain1) (83,84). The oncogenic EML4-ALK fusion protein has a constitutively activated kinase and has gain of function activity in vitro (80) and in vivo mouse models expressing EML4-ALK develop multiple lung ADC that are susceptible to pharmacologic ALK inhibition (85). Activation of ALK is linked to cell proliferation and inhibition of apoptosis mediated through the RAS/RAF/MAPK1, PI3K/AKT and JAK3-STAT3 signalling pathways (82).

ALK rearrangements have been identified in approximately 4% of unselected NSCLC (86) although some studies have found a slightly lower prevalence (5,87). They are more commonly found in ADC from younger patients who are never smokers or light smokers (78,87-91) and almost always occur in ADCs (90). While ALK rearrangements are usually mutually exclusive with EGFR and KRAS mutations (5,87,91,92) cases of coexistent EGFR mutations have been reported and provide a mechanism for TKI resistance (78,93-95). While ALK inhibition with the tyrosine kinase inhibitor crizotinib produces profound responses, drug resistance develops with evidence of secondary ALK point mutations and activation of EGFR signalling implicated in some cases (81,93).

**ROS1**

ROS1 is a proto-oncogene located on chromosome 6q22 which encodes a transmembrane tyrosine kinase receptor that has high homology with ALK in its protein kinase domain (96). ROS1 activation leads to signalling through the PI3K/AKT/mTOR, STAT3 and RAS/MAPK/ERK pathways (96). In 2007, a large scale phosphoproteomic screen for tyrosine kinase activity in lung cancer identified ROS1 fusion in a NSCLC cell line (1 of 41) and a patient sample (1 of 150) (SLC34A2-ROS1 and CD74-ROS1 respectively) (83). Subsequently, a novel KDELRS2-ROS1 in-frame fusion was identified in adenosarcoma from a non-smoker using whole genome and transcriptome sequencing (8). In two large studies using FISH, ROS1 rearrangements were found in 18 of 694 ADCs (2.6%) (97) and 13 of 1,116 ADCs (1.2%) (98). A variety of 5' fusion partners have been identified in ROS1 gene rearrangements (including FIG, KDELRS2, TPM3, SDC4, LRIG3, EZR, SLC34A2 and CD74) and it is uncertain what role, if any, the partner plays in the oncogenic function of the fusion kinase (8,83,98). Interestingly, ROS1 rearrangements appear to be more common in patients who are younger, never smokers or of Asian ethnicity (97) similar to ALK rearrangements (90). Furthermore, there is in vitro and early clinical evidence that lung cancers with ROS1 rearrangements are sensitive to kinase inhibitors including the ALK/MET inhibitor crizotinib (97).

**RET**

RET is located on chromosome 10q11.2 and encodes a receptor tyrosine kinase involved in neural crest development. Alterations of RET have long been known to play a role in papillary and medullary thyroid carcinoma (99) but it was not until recently that activation of RET through chromosomal rearrangement...
has been identified in a small proportion of lung cancers (9-11). The translocation fuses the functional RET kinase domain from exons 12-20 to KIF5B (kinesin family SB gene), that is 10Mb from RET on chromosome 10 and encodes a coiled coil domain involved in organelle trafficking (9,10). KIF5B-RET fusions have been identified in 1-2% of lung ADC using massively parallel sequencing technologies (10,11) and to date have been found to be mutually exclusive of other driver mutations involving EGFR, KRAS or ALK. In a highly selected cohort of lung ADC from never smokers or light smokers known to be wild type for other driver mutations (EGFR, KRAS, ALK, HER2, BRAF and ROS1), 10 of 159 (6.3%) harboured RET rearrangements (11). Similar to ALK and ROS1, rearrangements of RET also appear to be associated with ADC from never smokers (9-11). Importantly, there are several multi-kinase inhibitors that are effective against RET and there is in vitro evidence that cell lines expressing KIF5B-RET fusions are sensitive to RET inhibition (10,11).

**FGFR1**

Somatic gene amplifications have been found in SCCs in a number of genes including SOX, PDGFRa (12) and FGFR1 (12,100). FGFR1 is a membrane receptor tyrosine kinase that regulates cell proliferation through activation of the MAPK and PI3K pathways (101). Amplification of FGFR1 has an oncogenic effect on NSCLC cell lines in vitro that is sensitive to small molecule inhibition (102). About 20% of SCCs have been shown to harbour FGFR1 amplifications but the abnormality is uncommon in ADCs (100,102).

**DDR2**

Recently, a sequencing screen including the entire tyrosine kinase om was undertaken in SCCs and mutations were identified in DDR2 in 3.8% of cases (13). DDR2 encodes a membrane-bound receptor tyrosine kinase that binds collagen and is involved in regulation of cell proliferation and survival (103). Mutations of DDR2 are associated with oncogenic activity in vitro that is sensitive to inhibition with dasatinib (13).

**Tumour suppressor genes**

Tumour suppressor genes are crucial negative regulators of normal cell growth. Loss of tumour suppressor gene (TSG) function is an important mechanism of carcinogenesis and requires inactivation of both gene alleles, as outlined in Knudson’s two hit hypothesis (104). In one allele, the individual gene is often inactivated by mutation, epigenetic silencing or other aberrations, while the second allele is often inactivated through loss of heterozygosity (LOH) whereby a region of the chromosome is lost by deletion, nonreciprocal translocation or mitotic recombination. In lung cancer, TSGs that are frequently inactivated include TP53, retinoblastoma 1 (RB1), serine-threonine kinase 11 (STK11), CDKN2A, FHIT, RASSF1A and PTEN (1,7,105) and these genes map to chromosomal regions commonly identified in LOH studies. For example, regions frequently exhibiting allelic loss in lung cancer involve known TSGs such as TP53 (17p13), RB (13q12), p16 (9p21), and PTEN (10q22) (105). In a study by Ding et al. (7), mutations were identified in several TSGs not previously known to play a significant role in lung adenocarcinoma including the TSG NF1 (involved in neurofibromatosis type 1), that was mutated in 13 tumours and the TP53 regulator ATM in 13 patients.

**TP53**

TP53 located on chromosome 17p13 encodes a nuclear phosphoprotein of 53 kDa that identifies and binds to regions of damaged DNA (106) and acts as a transcription factor controlling the expression of a multitude of different genes. Damaged DNA or carcinogenic stress induces TP53 leading to cell cycle arrest by inducing expression of cyclin dependent kinase inhibitors to enable DNA repair or apoptosis. TP53 inactivation is one of the most significant genetic abnormalities in lung cancer with hemizygous deletion of 17p13, containing the locus of TP53, occurring in 90% of small cell carcinomas and about 65% of NSCLC (107). Inactivating mutations in TP53 (mostly missense mutations within the DNA-binding domain) have been reported in 80-100% of small cell lung carcinomas (108). By contrast, a meta-analysis of TP53 in over 4,000 NSCLC found alterations by mutation or protein accumulation in only 46.8% of cases (109), more commonly in SCC than ADC and associated with higher tumour stage, grade and male gender. Mutations of TP53 were found in at least 81% of SCCs that underwent comprehensive genomic analysis as part of The Cancer Genome Atlas (TCGA) project (12). Ding et al. (7) found TP53 mutations in 85 of 188 ADC (45%). In NSCLC, TP53 mutations are associated with a positive smoking history or exposure to environmental tobacco smoke (19,110). The mutational spectrum of different types of TP53 mutations also differs between smokers and non-smokers with smoking related cancers having a significantly higher frequency of G to T transversions compared to G to C transversions (thought to be induced by polycyclic aromatic hydrocarbons in tobacco smoke) and G to A transitions at CpG dinucleotides more commonly seen in never smokers (110,111). A meta-analysis of 74 studies showed that aberrant p53 detected by protein expression or mutational analysis is an unfavourable prognostic factor in lung NSCLC (112). Genetic alterations of TP53 have also been associated with treatment resistance (106). TP53 gene mutations can occur in association with EGFR and KRAS mutations (19).
PTEN encodes a lipid and protein phosphatase on chromosome 10 that inhibits the PI3K/AKT/mTOR signalling pathway by dephosphorylating PI-(3,4,5)-triphosphate (68). Inactivation of the TSG function of PTEN leads to unrestricted activation of AKT/protein kinase B independent of ligand binding (68). Mutations of PTEN occur only rarely in about 5% of NSCLC (113) being more common in SCC than ADC (10.2% vs. 1.7%) and associated with a history of smoking. By contrast, reduced protein expression has been reported in about 75% of NSCLC (114).

LKB1 (STK11)

LKB1 (also known as STK11) is a TSG located on chromosome 19p13 that encodes a serine-threonine kinase that inhibits mTOR and has been implicated in a range of biological processes including regulation of the cell cycle, chromatin remodelling, cell polarity, and energy metabolism (115,116). Deregression of mTOR pathway components (not including KRAS mutations) has been reported in 30% of ADCs (7). Germline mutations of LKB1/STK11 occur in patients with Peutz-Jeghers syndrome (115). In lung cancer, LKB1 may be inhibited by a variety of somatic mutations or deletions that produce truncated proteins with inactivation of LKB1 occurring in about 11-30% of lung ADC (7,117-119), making it the third commonest genetic aberration in lung ADC after TP53 and KRAS. LKB1 inactivation is more common in lung ADC compared to SCCs (117,119). There is some evidence of an association between LKB1 mutations and a history of smoking (117) in men (118,120) and a correlation with KRAS mutations has also been reported (117,118).

The p16INK4a-cyclin D1-CDK4-RB pathway

The p16INK4a/RB pathway regulates cell cycle progression from G1 to S phase. RB1 is a tumour suppressor gene that encodes RB protein which regulates cell cycle G1/S transition by binding the transcription factor E2F1. RB1 was the first TSG described in lung cancer (121) and is inactivated in about 90% of small cell lung carcinomas but only about 10-15% of NSCLC (1). In NSCLC, the pathway is mostly switched off through alterations of cyclin D1, CDK4 and the cyclin dependent kinase inhibitor p16 (CDKN2A) (105). p16INK4a inhibits cyclin D1 dependent phosphorylation of RB protein, thereby preventing cell cycle transition through the G1/S checkpoint (122). p16INK4a is inactivated in about 80% of NSCLC (123,124) and was altered in 72% of lung SCCs examined by TCGA, mostly through homozygous deletion, methylation or inactivating mutations (12). In addition, there is overexpression of cyclin D1 through gene amplification or other mechanisms in about 40% of NSCLC (123).

Molecular targeting in NSCLC

The presence of these molecular targets as described above now defines the characteristics of NSCLC, with EGFR mutation and ALK rearrangements being the most clinically relevant at present (125). The prevalence of these mutations varies in lung cancer arising from patient in different regions (126). Activating EGFR mutations were found in up to 20% of Caucasians while in the Asian populations these EGFR mutations can be present in up to 40% of patients with NSCLC (127). These ethnic difference in NSCLC properties appears to be not limited to the presence of activating EGFR mutations but is also evident in other driver oncogenic mutation profiles (including ALK, KRAS, MET etc.), histology and hence tumour response to targeted therapy treatment (63,126,128). The presence of these driver mutations is generally found to be mutually exclusive to others in the same tumour (126). In lung ADC among Asians, ALK rearrangement is seen in up to 7% of patients with lung ADC (79). Lung tumours bearing EML4-ALK rearrangement are non-responsive to conventional chemotherapy or EGFR-tyrosine kinase inhibitors but are sensitive to a specific tyrosine kinase inhibitor named crizotinib (129). Based on our current understanding of therapeutic molecular targets of EGFR mutation and ALK gene rearrangement in NSCLC and the availability of corresponding targeted agents, an algorithm of testing for molecular targets in NSCLC is proposed as in Figure 1, which represents a stepwise approach to testing for individual targets, beginning with EGFR then, if negative, ALK fusion gene or other potential targets if appropriate.

Among NSCLC, adenocarcinoma accounts for up to 80% of histological subtypes (130). There are previous reports of correlations between histological subtypes of ADC demonstrating micropapillary features with presence of activating EGFR mutations, leading to the suggestions that the presence of specific mutations in NSCLC actually represent heterogeneity in cancer biology and also response to therapy (131). Given the heterogeneity of lung cancer histology, however, histological subtypes are difficult to be used as the sole reliable marker for guidance to molecular phenotyping and selection of targeted therapy (132,133).

Targeting therapeutic oncogenic mutations like EGFR and ALK can give dramatic initial treatment response or at least an initial stable clinical disease. The response rate is up to 70% in lung ADC bearing favourable activating EGFR mutations (134). The median progression free survival is usually quoted as 9-11 months with different tyrosine kinase inhibitors (135,136), after which most patients with EGFR mutations will experience disease progression and drug resistance. A proportion of such drug resistance is attributed to the development of a second mutation, usually T790M at exon 20 (137). It is hard to explain the eventual loss of drug sensitivity in tumours bearing those
Cooper et al. Molecular changes in lung cancer

favourable EGFR mutations (exon 19 deletions and L858R) even without the acquisition of secondary mutations like T790M or the presence of other uncommon or less favourable EGFR mutations. This could reflect suboptimal therapeutic targeting and better understanding on the biology of EGFR-related tumour signalling and other oncogenic mutations will improve drug targeting and give patients better prediction of therapeutic response and prognostication.

Conclusions

The identification of driver mutations in EGFR and ALK heralded a new era of targeted therapy in lung adenocarcinoma and advanced sequencing technologies are providing even more sophisticated insights into the molecular aberrations in oncogenes and tumour suppressor genes underlying lung cancer (12,138-142). These studies have identified a range of potentially targetable genetic aberrations in lung cancer but have also highlighted a troubling complexity and heterogeneity which poses significant challenges for molecular diagnosis and targeted treatment. Greater knowledge of the molecular biology and genomic landscape of lung cancer offers promise for the future. Improvements in outcome from lung cancer will almost certainly require the identification of increasing numbers of ever rarer driver mutations, and diagnostic approaches that can identify multiple therapeutic targets offer significant advantages. However, the identification of driver genomic aberrations also requires the parallel development of effective targeted therapies and for many of these changes (such as KRAS) such therapies are not yet available. Resistance to targeted therapeutics is an increasingly recognised issue into which genomic analyses may provide important mechanistic insights underlying future rational therapeutic approaches.

Acknowledgements

Funding sources: Sydney Foundation for Medical Research; Hong Kong SK Yee Medical Foundation; Cancer Institute NSW Clinical Research Fellowship 10/1/07; Sydney Breast Cancer Foundation; Lifehouse at Royal Prince Alfred Hospital Grant; Lung Cancer SPORE P50CA70907.

Disclosure: WC has received honoraria from Pfizer Oncology.
References

40. 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)


51. Graus-Porta D, Beerli RR, Daly JM, et al. ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. EMBO J 1997;16:1647-55.


91. Wong DW, Leung EL, So KK, et al. The EML4-ALK fusion gene is involved in various histologic types of lung cancers from nonsmokers with wild-type EGFR and KRAS. Cancer 2009;115:1723-33.