Molecular biomarkers for lung adenocarcinoma

Olivier Calvayrac1, Anne Pradines1,2,3, Elvire Pons1,3, Julien Mazières1,2,4 and Nicolas Guibert1,2,4

Affiliations:
1Inserm, Centre de Recherche en Cancérologie de Toulouse, CRCT UMR-1037, Toulouse, France.
2Institut Universitaire du Cancer, Toulouse, France.
3Institut Claudius Regaud, IUCT-Oncopole, Laboratoire de Biologie Médicale Oncologique, Toulouse, France.
4Hôpital Larrey, Centre Hospitalier Universitaire, Université Paul Sabatier, Toulouse, France.

Correspondence: Julien Mazières, Thoracic Oncology Unit, Respiratory Disease Department, Hôpital Larrey, CHU Toulouse, Chemin de Pouvourville, 31059 Toulouse Cedex, France. E-mail: mazieres.j@chu-toulouse.fr

The epidemiology and impacts on prognosis and treatment responses of NSCLC strongly depend on molecular profile http://ow.ly/QCGP3094HLZ


ABSTRACT The identification of oncogenic driver alterations that underlie sensitivity to small inhibitors has led to growing interest in identifying additional targetable oncogenes in nonsmall cell lung cancer. Although the therapeutic impact of the discovery of these alterations has now been widely demonstrated, the epidemiological data associated with each of these biomarkers remain insufficiently studied. In this review, we discuss the techniques used to discover each of these candidate oncogenes, their prevalence in nonsmall cell lung cancer, and briefly outline the epidemiological features of the major oncogenes and ways in which their identification can determine therapeutic strategies.
Introduction
Mechanisms of oncogenesis in lung cancer have been largely deciphered over the past 20 years. The concept of "oncogene addiction" refers to tumour-cell dependence on the specific activity of an activated or overexpressed oncogene. The main oncogenic drivers in the field of thoracic oncology are mutations of \textit{EGFR}, \textit{KRAS} and \textit{ALK} rearrangements. They are most often reported in adenocarcinomas. However, new molecular targets have been highlighted recently: \textit{i.e.} \textit{BRAF} mutations, \textit{HER2} and \textit{PIK3CA}, and new translocations, such as \textit{ROS1} and \textit{RET}. Therapeutic strategies have been designed to inhibit these signalling pathways, among which are monoclonal antibodies and tyrosine-kinase inhibitors.

Tumours are now classified according to their molecular profile, which is itself associated with new demographic data. Lung adenocarcinoma can now be considered as a cluster of discrete molecular subtypes, the majority of which are defined by a single alteration to the oncogenic driver. Multiplex genotyping and high-throughput genomic profiling by next-generation sequencing (NGS) has been increasingly refining molecular diagnoses.

In this review, we discuss the techniques used to discover each of these candidate oncogenes, their prevalence in nonsmall cell lung cancer (NSCLC), their associated epidemiological features, and ways in which their identification can inform therapeutic strategies.

Methods of tumour sampling
There is a current paradox between the need to obtain significant samples for multiple analyses for a growing number of molecular biomarkers and the development of minimally invasive or noninvasive techniques, resulting in small tissue samples with very small amounts of DNA. Various techniques (described below) are used for the initial molecular profiling of DNA extracted from formalin-fixed and paraffin-embedded (FFPE) tumour samples.

Cytological samples are often considered insufficient for exhaustive molecular examination. Yet, recent data have demonstrated that very limited amounts of tissue can be sufficient for this type of analysis. Only 10 ng of DNA is sufficient to analyse 22 genes using NGS [1]. Likewise, LASER microdissection on 50 cells can facilitate the analysis of \textit{EGFR} and \textit{KRAS} using pyrosequencing [2]. Combining new sampling procedures with higher sensitivity tests helps the pulmonologist to obtain a molecular profile from various samples. The main advantages and pitfalls of each technique are reported in table 1.

Endobronchial ultrasound and transbronchial needle aspiration
Multiple studies have demonstrated the usefulness of cytological samples obtained by ultrasound-guided cytopuncture of lymph nodes through echobronchoscopy and molecular testing in NSCLC. In the study by BOULANGER et al. [3], screens for \textit{KRAS} and \textit{EGFR} mutations were possible in 95.1% and 97.6% of cases, respectively (n=82). More recently, CASADIO et al. [4] tested 306 samples of NSCLC for \textit{EGFR} and \textit{KRAS}

<table>
<thead>
<tr>
<th>Sample</th>
<th>Diagnostic accuracy</th>
<th>Advantages</th>
<th>Pitfalls</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBUS-TBNA</td>
<td>Sensitivity &gt;95%</td>
<td>Histologic diagnostic, staging and molecular profile at the same time</td>
<td>Relatively invasive</td>
</tr>
<tr>
<td>cfDNA</td>
<td>Good sensitivities &gt;95% [\textit{EGFR}, \textit{KRAS}, \textit{BRAF}]</td>
<td>Noninvasive Rapid, simple monitoring Early detection or acquired resistance [\textit{EGFR} T790M]</td>
<td>Low number of mutated alleles among wild-type alleles</td>
</tr>
<tr>
<td>CTCs</td>
<td>Sensitivity 78% [\textit{KRAS}] Sensitivity 92% [\textit{EGFR}]</td>
<td>Noninvasive Possibility of: - Cytomorphological analysis - FISH [\textit{ALK}] - ICC [\textit{ALK}]</td>
<td>Expensive and laborious Lack of standardisation Multiplicity of methods</td>
</tr>
<tr>
<td>Pleural fluid</td>
<td>88% sensitivity</td>
<td>Possibility of multiplexed molecular testing if previously centrifuged</td>
<td>Contamination by haematopoietic cell DNA</td>
</tr>
<tr>
<td>Bronchoalveolar lavage</td>
<td>Sensitivity 16% [\textit{Sanger}] to 81% [\textit{NGS}]</td>
<td>Good sensitivity with NGS</td>
<td>Low number of tumour cells Poor sensitivity with conventional sequencing methods</td>
</tr>
</tbody>
</table>

EBUS-TBNA: endobronchial ultrasound transbronchial needle aspiration; cfDNA: circulating free DNA; CTCs: circulating tumour cells; FISH: fluorescence in situ hybridisation; ICC: immunocytochemistry; NGS: next-generation sequencing.

https://doi.org/10.1183/13993003.01734-2016
(Sanger and RT-PCR) and ALK rearrangements (FISH), with very good diagnostic accuracy (96.9% of samples were suitable for molecular profiling, with very similar results compared to tissue data). In this context, pyrosequencing and high-resolution melting produce equally excellent results: 97% and 93% of cases, respectively [5].

**Transthoracic fine-needle aspiration**

SCARPA et al. [1] demonstrated the feasibility of obtaining the mutational status of the main genes potentially involved in thoracic oncogenesis using just a small amount of DNA (10 ng) and NGS. In this study, 22 genes were examined in cytological samples obtained by transthoracic fine-needle aspiration. A mutation of at least one of these genes was found in 67% of patients, which included 28% KRAS mutations and 16% EGFR mutations.

**Biological fluids**

**Bronchoalveolar lavage and bronchial-smear rinse fluid**

Few studies have assessed the feasibility of molecular analysis on alveolar-lavage fluid. The PCR-denaturing gradient gel-electrophoresis (DGGE) technique offers low sensitivity (43%) and only 33% consistency with the primary tumour (n=36) for KRAS mutations [6]. In a recent study, sequencing cells collected by bronchoscopic brushing enabled mutations to be found in only one of the 77 patients sampled [7]. Among 36 positive bronchoalveolar lavages (0.3–9% tumour cells) from patients with a known EGFR mutation, the same mutation was detected in only 16% of cases using Sanger sequencing, but in 81% of cases using NGS. NGS could even detect an EGFR mutation in 42% of cytological samples without evidence of a tumour [8]. EGFR, KRAS and ALK statuses can also be established from endobronchial ultrasound (EBUS)-guided brushing of peripheral nodules with good diagnostic accuracy (95.2%) and concordance with other samples [9].

**Pleural fluids**

Testing for mutations in pleural fluid is also possible [2], but is complicated by the presence of DNA of haematopoietic cells. Prior centrifugation could facilitate this procedure, and would concentrate and isolate the tumour cells from leukocytes [10]. Multiplexed molecular testing (EGFR, KRAS, BRAF, PIK3CA, NRAS, MEKI, AKT1, PTEN and HER2 mutations, and ALK, ROS1 and RET fusion genes, by pyrosequencing and RT-PCR) was shown to be feasible in 102 pleural fluid samples from 84 Japanese patients with NSCLC. EGFR mutations were detected in 29% of patients and ALK rearrangement in 4%, with a concordance rate of 88% with FFPE samples [11].

**Circulating tumour cells**

Various methods allow the isolation of circulating tumour cells, among which Cellsearch, an indirect method based on cytokeratin staining, and ISET (isolation by size of epithelial tumour cells), a direct method based on cell size, are the most widely used. Beyond its prognostic value and the possibility of cytomorphological analysis, ISET can detect some molecular alterations in NSCLC, such as EGFR mutations by PCR [12], or ALK rearrangement by FISH and immunocytochemistry, with good concordance to primitive tissue [13].

One study reported the feasibility of detection of KRAS mutations using PCR in circulating tumour cells: 100 patients carrying a KRAS mutation on the primary tumour were included (32 colorectal cancers, 51 breast cancers and 17 lung cancers). The study reported 90.2% sensitivity and 94.9% specificity [14]. More recently, sensitivity of 78% for KRAS analysis was reported with COLD-PCR (co-amplification at lower denaturation temperature-PCR)/high-resolution melting for DNA extracted from circulating tumour cells isolated in a size-selected assay (ScreenCell) [15].

**Circulating free DNA in plasma**

Circulating tumour DNA (ctDNA) corresponds to the fragmented DNA released by tumour cells either actively or passively, during apoptosis or necrosis [16]. Cell-free DNA is increased in lung cancer patients [17]. A high level is correlated with a poor prognosis [18] and its variations during treatment could help monitor tumour burden [19]. Nevertheless, cell-free DNA has some limitations because of its low specificity, as normal cells or multiple inflammatory of infectious benign disorders can increase its level [20]. This problem can be avoided by targeting mutated tumour-specific DNA. Thus, detection of several genomic alterations, including those of EGFR, BRAF and KRAS in circulating DNA is possible with good diagnostic accuracy and excellent specificity and sensitivity, especially when using digital droplet PCR (ddPCR) [21–23].

We have recently shown that monitoring mutated ctDNA with ddPCR is a noninvasive way to follow-up the mutant clone in mutated NSCLC, as is evident with BRAF-mutated lung adenocarcinoma [24].
Another very exciting application is the early detection of the mechanism of resistance during targeted therapy. The detection of EGFR-T790M-mutated ctDNA, for example, precedes the radiological progression of EGFR-mutated lung adenocarcinoma [21]. NGS allows the detection of very small amounts (0.4%) of mutated DNA (KRAS, EGFR, BRAF, etc.), as well as other genomic alterations (HER2 insertions, cMET amplifications and ALK, ROS1 and RET rearrangements) in plasma with 100% specificity and sensitivity, similar to ddPCR (77%) [25].

Decision tree algorithm for tumour sampling technique
For initial genotyping, tissue should still be the first choice for molecular profiling of the tumour. A biopsy is indeed required for the histological diagnosis and could be used in the first instance. However, less invasive techniques are should be preferred when possible. Conventional bronchoscopic biopsies are prefered in cases of proximal endobronchial tumours and linear EBUS-TBNA is preferred for peribronchial tumours or mediastinal lymph nodes [3–5].

New tools have been developed to access distal tumours and avoid the complications associated with computed tomography (CT)-guided transthoracic biopsies (haemorrhage and pneumothorax) [26]. A radial EBUS miniprobe (rEBUS-MP) can be introduced through the involved bronchus after careful analysis of the CT scan. When the minibreipe reaches the tumour, the normal lung "snowstorm" appearance is replaced by focal circumferential hypo-echogenicity, and the probe is then replaced by biopsy forceps. A meta-analysis has reported a pooled sensitivity of 78% for nodules >20 mm and 56% for nodules <20 mm in diameter [27]. However, this procedure must be reserved for patients showing the bronchus sign on a CT scan [28].

Virtual bronchoscopic navigation uses a magnetic field, a magnetic sensor probe and three-dimensional integration to enable CT scan reconstruction and indicate bronchoscope position. The diagnostic yield of this technology is good (73.8%), even for lesions <20 mm in diameter (67.4%). This procedure is expansive and seems to be particularly suited for the upper lobes and peripheral third of the lung field [29].

However, if molecular characterisation cannot be assessed on these samples, a liquid biopsy is preferred to avoid a second invasive sampling procedure. Sequencing of ctDNA is the only technique approved for the analysis of EGFR mutational status and should be used as a first-line method in such circumstances. Its specificity is strong, facilitating the prescription of epidermal growth-factor receptor tyrosine kinase inhibitors (EGFR-TKIs) based on the results. However, the sensitivity of this method is imperfect [30] and a negative result must be confirmed by a tissue biopsy, which will enable wider genotyping including rearrangements (ALK, ROS1, etc.) [31].

At progression, a liquid biopsy constitutes the best choice to analyse mechanisms of resistance. In cases of EGFR-mutated tumours, ctDNA constitutes a very useful tool for detection of the T790M mutation, which occurs in more than half of all patients. The specificity of this technique is high. Furthermore, because of tumour heterogeneity and the improved detection sensitivity of the techniques used, the T790M mutation can even be detected in patients considered T790M-negative based on tissue analysis, making ctDNA a complementary tool. The US Food and Drug Administration (FDA) proposed a new paradigm for detection of the T790M mutation at progression. Plasma genotyping should be proposed as a first-line technique, as its high specificity facilitates the prescription of a third-generation EGFR-TKI. Nevertheless, its sensitivity is imperfect and a negative result must be confirmed by tissue analysis, which enables the detection of other mechanisms of resistance (small-cell carcinoma transformation, MET or HER2 amplification, PIK3CA mutations, etc.) [32].

The decision tree algorithm is presented in figure 1.

Methods used to detect molecular alterations
Different methods are available to identify the diverse somatic genomic aberrations that can be encountered in NSCLC (point mutations, insertions or deletions and gene rearrangement) obtained from FFPE samples or other available sources of material, such as cytology specimens or circulating DNA. Some technologies allow mutations to be screened and sequenced in the region of interest, whereas others analyse specific and known mutations.

Owing to the cellular heterogeneity of solid cancers, the primary technical challenge is the detection of somatic variants in tumour biopsies. Somatic mutations can be present in low numbers within an elevated background of wild type sequences, and more sensitive assays are therefore needed than those used for germline variants. Moreover, the challenge has been to develop more sensitive techniques for small biopsies and liquid biopsies. In particular, circulating mutant DNA represents only a very small fraction of the total circulating DNA [33]. Following the introduction of the first-generation capillary Sanger
sequencer (in 1995), increasing numbers of technologies have now been made available, including real-time PCR platforms, digital PCR and NGS.

**Screening assays**

**Sanger sequencing**

Until recently, Sanger sequencing was considered the gold standard for identification of all mutations. This technique uses sequencing according to termination performed on PCR products, which generates fluorescently labelled DNA fragments that are separated by size using capillary electrophoresis [34]. However, Sanger sequencing has very low sensitivity. The mutant variant should represent at least 15–20% of the total tumour DNA to be detected [35].

**Pyrosequencing**

Pyrosequencing is sequencing by synthesis, and it relies on the detection of pyrophosphate release during the incorporation of nucleotides into the DNA chain using bioluminescence through a series of enzymatic reactions. Pyrosequencing can identify individual bases or short stretches of nucleic-acid sequences at predetermined positions. The limit of detection of an allele is 5% [36].

**High-resolution melting**

High-resolution melting is used to screen for mutations in genes of interest prior to sequencing analysis. It is a post-PCR method based on the detection of small differences in PCR melting curve patterns, generated by the transition from a double to single strand in the presence of fluorescent dyes that intercalate with the double-stranded DNA and differentiate wild type sequences from homozygote or heterozygote variants. To improve the detection limit of this assay from a range of 2–10% to 0.1–1% of mutants in mixtures with wild type DNA co-amplification at lower denaturation temperatures, PCR (COLD-PCR) that allows preferential amplification of minority alleles could be combined with high-resolution melting [36].

**Next-generation sequencing**

NGS is a high-throughput sequencing method that detects somatic mutations in tumour samples at levels as low as 5%. Several chemical and analytical methods are available. The Illumina platform is based on the
preparation of DNA fragments using bridge PCR on a solid support to form clusters, which are then sequenced through repeated cycles on a single-base extension using a mixture of four fluorescently labelled nucleotides, and then imaged. In contrast, the ion-torrent platform relies on template preparation using emulsion PCR. The incorporation of a particular nucleotide is detected following the release of hydrogen ions. The number of theranostic biomarkers and associated targeted therapies is rapidly increasing, and tissue samples can often be limited to very small amounts of DNA. NGS can rapidly evaluate the entire genomic landscape of the tumour simultaneously. At initial diagnosis NGS on tissue samples facilitates identification of the genotype of the tumour. Targeted panels including genes involved in the oncogenesis of lung cancer are preferred. A whole exome approach would significantly increase the turnaround, bioinformatic analysis and cost, as well as the risk of non-targetable genomic aberrations. When genotyping is not possible on tissue, plasma targeted NGS could be of great benefit [25] even if this approach has not yet been validated for routine use. The technique of circulating free DNA genotyping can overcome the dual problem of intra-tumour heterogeneity that exists between different regions of the same tumour (spatial heterogeneity), and between the primary tumour and local or distant recurrences in the same patient (temporal heterogeneity). NGS on plasma samples could also be an appealing technique for follow-up analysis of tumour biology under targeted therapy. If a targeted assay like ddPCR is the preferred method in EGFR-TKI resistant patients, in whom T790M resistance mutation is the main mechanism of resistance to monitor, a multiplex assay should be of great benefit in other TKI failure situations. For example, resistance in anaplastic lymphoma kinase (ALK) patients being treated with crizotinib tends to be more heterogeneous. Numerous ALK resistance mutations have been reported and the spectra of activity associated with different second and third-generation ALK tyrosine kinase inhibitors (ALK-TKIs) are not similar [37]. A plasma multiplexed NGS assay would thus be very useful in such circumstances.

Targeted assays
Most of the methods used to detect a specific mutation are based on multiplex real-time PCR using allele-specific PCR probes; however, they can only detect known mutations. In cases of lung adenocarcinoma, this targeted approach seems particularly suitable for detection of frequent and targetable alterations, like EGFR del19, L858R, KRAS G12X or EGFR T790M mutations.

Competitive allele-specific TaqMan® PCR
Competitive Allele-Specific TaqMan® PCR (castPCR™) technology (Life Technologies) combines allele-specific TaqMan® qPCR with allele-specific minor groove binder blockers to suppress nonspecific amplification from wild type alleles. CastPCR™ technology provides high sensitivity and specificity in the detection of rare mutant alleles among large quantities of wild type DNA (0.1%).

Therascreen® mutation kits
Therascreen® mutation kits (Qiagen) are real-time PCR assays that combine the amplification refractory mutation system (ARMS), which is an allele-specific amplification method, with Scorpion PCR primers, which are linked to a probe (a fluorophore and a quencher).

Cobas® kit
The Cobas® kit (Roche Molecular Systems) uses specific TaqMan probes (each with a different fluorescence intensity, generated by a fluorophore quencher pair) in real-time PCR to amplify and analyse mutations.

Peptide nucleic-acid clamp technology
Peptide nucleic-acid clamp technology consists of selective amplification of mutant alleles using specific peptide nucleic acids that suppress wild type sequence amplification.

BEAMing
BEAMing (beads, emulsion, amplification and magnets) is a process based on emulsion digital PCR that binds DNA to magnetic beads, to facilitate separation and detection using flow cytometry [38].

Digital PCR
Digital PCR can detect and quantify very small amounts of mutant DNA, by partitioning a DNA sample into several thousand replicates. This partitioning facilitates the amplification and detection of single spots, which substantially increases sensitivity (0.005–0.01%) [39]. This method is thus particularly suitable for plasma genotyping and is now validated for the detection of common EGFR (deletion 19, L858R) and KRAS alterations (G12X, G13X) and for early detection of T790M in TKI-resistant patients [40].
ddPCR quantitative analysis of the ctDNA could also be a useful tool to monitor the response to targeted or conventional treatments [21–24].

**Molecular methods used to detect gene rearrangements**

*Fluorescence in situ hybridisation*

Fluorescence in situ hybridisation (FISH) is the gold standard for detection of ALK or ROS1 rearrangements, but does not identify the fusion partner. This cytogenetic technique uses two small DNA strands coupled to fluorescent molecules that are complementary to the 5′ and 3′ ends of the gene. As observed under fluorescence microscopy, the two probes in wild type cells are very close together and their fluorescence can fuse; whereas in mutant cells, the two colours are well separated, or only one is detected. Several certified ALK and ROS1 break-apart fluorescence in-situ hybridisation assays are available [41]. FISH analysis is considered the gold standard for ALK NSCLC mutation testing. In 2011, the FDA approved the Abbot Vysis ALK Break Apart FISH Probe Kit for molecular diagnostic testing.

*Immunohistochemistry*

Immunohistochemistry (IHC) of ALK or ROS 1 assumes that the normal protein is absent; whereas the fusion proteins are overexpressed, as previously observed in anaplastic large-cell lymphoma [42]. In NSCLC, the ALK protein can be easily detected and some studies compare reported results between IHC and FISH for ALK rearrangements [42]. Moreover, IHC has been recommended for the detection of RET and ROS rearrangements [43].

*Reverse transcriptase PCR*

Multiplex reverse transcriptase PCR (RT-PCR) can discriminate between different fusion transcripts, but requires prior knowledge of the possible partners in order to design specific primer sets for each translocation [44, 45].

**Biomarkers for targeted therapies**

The incidence and characteristics of the main biomarkers of NSCLC are summarised in table 2. Their geographic distribution is presented in figure 2.

**EGFR**

*Epidemiology*

Epidermal growth-factor receptor (EGFR; erbB-1; HER1) belongs to the ERBB family of tyrosine-kinase receptors [53]. The EGFR gene is located on chromosome 7 (7p11.2) and encodes a 170 kDa transmembrane glycoprotein with tyrosine-kinase activity. Upon binding to a specific ligand, EGFR undergoes conformational

### Table 2 Incidence and characteristics of the main biomarkers for nonsmall cell lung cancer

<table>
<thead>
<tr>
<th>Target</th>
<th>Biology</th>
<th>Caucasian patients %</th>
<th>Approved treatments</th>
<th>Clinical trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFR</td>
<td>Mutation</td>
<td>10–15</td>
<td>Gefitinib, erlotinib, afatinib osimertinib (second-line if T790M)</td>
<td>Rociletinib (second-line if T790M)</td>
</tr>
<tr>
<td>ALK</td>
<td>Translocation</td>
<td>3–5</td>
<td>Crizotinib</td>
<td>Ceritinib, alecetinib, brigatinib, lorlatinib</td>
</tr>
<tr>
<td>BRAF</td>
<td>Mutation</td>
<td>2</td>
<td>NA</td>
<td>Vemurafenib, dabrafenib, dabrafenib+trametinib</td>
</tr>
<tr>
<td>ROS1</td>
<td>Translocation</td>
<td>1</td>
<td>Crizotinib</td>
<td>Ceritinib, Lorlatinib</td>
</tr>
<tr>
<td>HER2</td>
<td>Mutation</td>
<td>1</td>
<td>NA</td>
<td>Trastuzumab, afatinib, neratinib</td>
</tr>
<tr>
<td>KRAS</td>
<td>Mutation</td>
<td>20–25</td>
<td>NA</td>
<td>Trametinib, selumetinib, abemaciclib</td>
</tr>
<tr>
<td>PI3K</td>
<td>Mutation</td>
<td>2</td>
<td>NA</td>
<td>PI3K inhibitors, mTOR inhibitors</td>
</tr>
<tr>
<td>MET</td>
<td>Amplification mutation</td>
<td>2–5</td>
<td>NA</td>
<td>Crizotinib, INC280, tepotinib</td>
</tr>
<tr>
<td>RET</td>
<td>Translocation</td>
<td>1–2</td>
<td>NA</td>
<td>Cabozantinib, sorafenib, vandetanib</td>
</tr>
</tbody>
</table>

NA: not applicable.
change and phosphorylation, leading to activation of several pathways involved in cell survival and proliferation, such as PI3K/AKT, RAS/ERK and JAK/STAT, among others [54]. Since the 1980s, several clinical studies have reported increased levels of EGFR expression or increased gene-copy numbers in lung cancer, particularly in squamous-cell carcinoma, adenocarcinoma and large-cell carcinoma [55–58]. Although the prognostic association of EGFR is still discussed because of inconsistencies in clinical reports [57–61], the development of selective inhibitors to EGFR tyrosine-kinase activity (e.g. gefitinib and erlotinib) has allowed the identification (in 2004) of specific mutations in the EGFR gene that are harboured in a subgroup of patients who present with a dramatic clinical response to these inhibitors [61–63]. Since 2004, much effort has been made to identify and characterise these mutations, and the large quantity of data thus generated now permits a relatively accurate overview of the prevalence of EGFR mutations among this population.

In NSCLC, the oncogenic mutations of EGFR are found within exons 18 to 21, which encode for part of the tyrosine-kinase domain located around the ATP-binding pocket of the enzyme, which is also the binding site of EGFR-TKi. The most common EGFR mutations are the in-frame deletions of exon 19 (mostly delE746-A750), which account for ~45% of all EGFR mutations, followed by L858R substitution in exon 21, found in 40–45% of all EGFR mutations [64]. For the remaining 10%, other rare EGFR mutations have been reported that include insertions in exons 19 and 20 (1% and 4%, respectively) and point mutations in exon 21 (L861, mostly L861Q: 1–2%) and exon 18 (mostly G719X: 3%) [65–68]. The choice of treatment might differ for patients harbouring common and rare mutations, as not all EGFR mutations confer the same degree of sensitivity to EGFR-TKIs (see below).

EGFR mutations are almost exclusively associated with adenocarcinoma or a bronchioloalveolar histological subtype, and are largely mutually exclusive with mutations in KRAS or ALK rearrangements (described below) [69]. They are more frequently observed in nonsmokers and women. The frequency of EGFR mutations is much higher in individuals of Asian origin (45–50%) than in individuals from Western Europe (10–15%) or North America (15–20%). It is noteworthy that high variability has been reported in the prevalence of EGFR mutations within the Asian-Pacific population, with a higher frequency in Vietnam (64.2%), Thailand (53.8%) and China (50.2%) than in India (22.2%) or Bangladesh (23%) [70, 71]. This variability has also been observed in South America, with a higher proportion of EGFR mutations among the
Amerindian subpopulation (principally Peru, 67%, but also Mexico, 31.2%, and Colombia, 24.8%) compared to Caucasian populations (Argentina, 19.3%) [72]. A world map representing EGFR mutation frequency in patients with NSCLC by country has been recently constructed from 151 studies worldwide [73].

Treatment

Targeted therapy remains the best strategy to treat lung cancer patients who harbour EGFR-activating mutations. Three drugs are currently available in the US and Europe (gefitinib, erlotinib and afatinib). Sensitivity to EGFR-TKIs seems to be highly dependent on the type of EGFR mutation. The best response rates (RR) are observed in patients harbouring the two main EGFR alterations (exon 19 deletions and the L858R mutation), with a higher susceptibility to treatment observed in patients with an exon 19 deletion (RR of 70–100% versus 20–67% for those with an L858R mutation) [65, 66, 74, 75]. Although rare EGFR mutations are often associated with poorer efficiency of EGFR-TKI as compared to common mutations, patients with L861Q or G719X mutations could benefit from this type of treatment, whereas insertions in exon 20 are more likely to confer resistance to TKI therapy [61, 74]. This was confirmed in a French multicentre study that analysed 10117 samples from which 1047 (10%) were EGFR-mutated (102 with rare mutations) and exon 20 insertions were associated with EGFR-TKI resistance, whereas exon 18 mutations showed better sensitivity to treatment [66]. The choice of EGFR-TKI to treat patients with uncommon mutations could also be important; for instance, afatinib should be favoured in patients with G719X and L861Q mutations [64, 76]. For patients with very rare (other than G719X and L861Q) or complex mutations (with more than one mutation) in the EGFR gene, the efficiency of EGFR-TKI therapy has not been clearly demonstrated because of the small number of patients. However, the proportion of EGFR mutations that confers sensitivity to EGFR-TKI therapy seems to be greater than that which confers resistance [74, 77]. We believe that first-line EGFR-TKI therapy should be the most suitable strategy to treat patients with a specific EGFR mutation, unless the data indicate a clear disadvantage for targeted therapy in this particular context.

Despite a high global response rate (~60–70%), all patients usually relapse within a median delay of 12 months. In the majority of cases, this is due to the existence of the T790M gatekeeper mutation in 50–60% of patients, and much less frequently caused by amplification of HER2 and the MET proto-oncogene, the epithelial to mesenchymal transition, or more rarely a small cell histologic transformation [78–81]. Several third-generation EGFR inhibitors have been recently developed (e.g. osimertinib and rociletinib) that show clinical efficacy in patients harbouring the T790M mutation [82, 83]. Osimertinib (AZD9291, Tagrisso) has recently received accelerated approval by the FDA for the treatment of patients with a confirmed T790M mutation. However, adaptive resistance mechanisms have already been described in a small subset of patients who relapsed after third-generation EGFR inhibitor treatment that included those with the EGFR C797S mutation [84]. Although the primary EGFR-activating mutation (mostly exon-19 deletion and a L858R point mutation) is usually conserved in resistant clones arising from erlotinib or gefitinib treatment, resistant clones derived from osimertinib- or rociletinib-treated T790M-positive tumours are rarely positive for the T790M mutation. This is because T790M-resistant tumours are often accompanied by T790M-negative resistant clones, which are likely to confer resistance to third-generation EGFR inhibitors [85]. This intra-tumoural heterogeneity is thus considered the Achilles heel of targeted therapy, and studying the origin of such clonal diversity could provide clues to promote more durable remissions.

ALK

ALK, a member of the insulin-receptor tyrosine-kinase family [86], is encoded by the ALK gene on chromosome 2. ALK was first identified as part of the NPM-ALK (nucleophosmin-anaplastic lymphoma kinase) oncopgenic fusion protein that results from translocation between chromosomes 2 and 5 (t(2;5) [p23;q35]) and is associated with anaplastic large-cell lymphoma [87]. The same translocation has also been described in Hodgkin’s lymphoma [88]. Subsequently, a small inversion within chromosome 2p results in the formation of a fusion gene comprising portions of the echinoderm microtubule-associated protein-like 4 (EML4) gene and the ALK gene, which were identified in a resected adenocarcinoma specimen from a 62-year-old male smoker [89]. Several variants of ALK-EML4 have been identified [90] and other fusion partners have been reported in NSCLC, including KIF5B [91].

ALK rearrangements occur in 3–5% of patients with NSCLC, and are more common among younger patients with a light smoking history, adenocarcinoma histology, and in tumours that are wild type EGFR and KRAS [92, 93]. These factors could help clinicians to identify high-risk populations that can undergo ALK testing, even if these criteria are not strong enough to select patients based on epidemiological characteristics. Consistent with these findings, the International Association for the Study of Lung Cancer
and the European Society for Medical Oncology guidelines both propose screening all patients with advanced-stage lung adenocarcinoma, irrespective of the clinical characteristics.

Crizotinib was the first drug to be approved for ALK+ NSCLC. It is an oral, small-molecule inhibitor that targets ALK, ROS1 and MET tyrosine kinases [94–96]. It has shown significant (~60%) overall response rates in a single-arm phase-I study [95], leading to its approval in the USA. The results of recent phase-III trials have confirmed the superiority of crizotinib compared to chemotherapy in second-line [97] and first-line [98] settings for ALK+ NSCLC. Unfortunately, the majority of patients also develop resistance within the first 12 months of therapy [99].

Developments in the understanding of the molecular biology and resistance mechanisms of ALK+ NSCLC have been made over the past few years. Alectinib and ceritinib, both second-generation ALK inhibitors, with higher intracranial activity than crizotinib, and used to treat leptomeningeal carcinomatosis, have recently gained approval as a form of therapy in patients who are refractory, or intolerant to crizotinib. Sequential use of ALK inhibitors has led to increased overall survival, with some ALK+ patients reaching an overall survival of 5 years.

Lorlatinib (PF06463922), a third-generation ALK/ROS1 inhibitor that can overcome certain resistant ALK mutations (with the exception of ALK L1198F), is now in a phase-II clinical trial that includes ALK+ and ROS1+ NSCLC patients. ALK L1171X and ALK F1174X resistance mutations have different sensitivities to alectinib and ceritinib, which might determine which of the two inhibitors should be selected. Other resistant ALK mutations have been reported. In one case report, Silaw et al. [100] recently reported that C1156Y–L1198F mutations could confer resistance to lorlatinib, ceritinib, alectinib and brigatinib, but re-sensitise cells to crizotinib.

**KRAS**

Mutations in the KRAS gene (v-Ki-ras2 Kirsten rat sarcoma viral-oncogene homologue) are characterised by high frequency and specific epidemiology. Alterations to the KRAS gene are most frequently localised on codon 12 and more rarely, on codons 13 and 61 [101]. Alterations to the KRAS gene cause a loss of GTPase activity, and thus affect its feedback regulation. KRAS then continuously promotes cell proliferation via the MAP-kinase (BRAF/MEK/ERK) and phosphoinositide-3-kinase (PI3K) pathways [102]. G12C substitution is the most frequent alteration, representing 52% of all KRAS mutations. G12V, G12D, G12A and G12S mutations can be found in 11.6%, 7.2%, 2.9% and 1.5% of patients, respectively. In never-smokers who have adenocarcinoma, the G12D substitution is found in 50% of all cases [103, 104]. KRAS gene mutations can be found in 18–32% of adenocarcinomas, 12.8% of large-cell carcinomas, 10% of adenosquamous carcinomas and 1.6–7.1% of squamous-cell carcinomas in Caucasian patients [49, 105–108]. The association between KRAS mutations and tobacco is undeniable [108–110], and has been clearly identified in two meta-analyses [110, 111]. Nevertheless, KRAS mutations can also be found in 5–15% of nonsmokers [112–114]. KRAS mutations are more frequent in African-American patients compared to Caucasians (OR 2.4; p=0.048) [115], and are far less frequent in Asian patients, with an incidence of 11.2% [110].

KRAS gene mutations are usually described as an unfavourable biomarker in patients with resected pulmonary adenocarcinoma [116, 117], with a hazard ratio of 1.40 (n=5216, p=0.01) in a meta-analysis by Massaux et al. [105]. Nevertheless, the prognostic implications of KRAS mutations are still under debate. In a study by Villaruz et al. [106], which focused on 318/988 cases of adenocarcinoma harbouring a KRAS mutation, the multivariate analysis (with adjustment for other prognostic factors, particularly smoking status) showed no prognostic impact of KRAS. In another recent study involving a cohort of 1935 patients, KRAS mutations appeared to correlate with a shorter overall survival, but did not influence progression-free survival (PFS) [111].

Contradictory findings are reported regarding the impact of different subtypes of substitutions; Ilie et al. [104] reported the adverse effects of G12V, whereas Cser Episcopal et al. [109] observed better responses to chemotherapy and longer PFS in similar patients. No specific subtype of KRAS mutations showed any prognostic value in another study that evaluated adjuvant chemotherapy [118].

Previous retrospective studies also report detrimental effects [119–121], which seem to occur primarily in patients with a mutation located on codon 13 [118, 122]. Detection of a KRAS mutation in ctDNA is associated with a poor response to platinum-based chemotherapy [23]. Nevertheless, these results have not been consistently observed [122–124]. Because KRAS mutations cause constant activation of previously mentioned signalling pathways, independently of EGFR activation, the inefficiency of EGFR-TKI seems a logical inference. Most retrospective analyses are consistent with these results, with response rates usually <3% [125–130]. Nevertheless, a few clinical trials were unable to show any difference in survival [131, 132]. Unlike the case of colorectal cancer, KRAS mutational status does not affect the response to monoclonal anti-EGFR antibodies in pulmonary adenocarcinoma [133].
**HER2**

Human epidermal growth factor 2 (HER2, erbB-2/neu) is a member of the erbB receptor tyrosine-kinase family. The ERBB2 gene, which encodes HER2, is a major proliferative driver that activates downstream signalling through the PI3K-AKT and MEK-ERK pathways. No ligand has been described for this receptor, family. The human epidermal growth factor 2 (HER2, erbB-2/neu) is a member of the erbB receptor tyrosine-kinase family.

**BRAF**

After activation by the RAS protein, BRAF induces phosphorylation of two other kinase proteins, MEK and ERK. The activated form of ERK can translocate into the nucleus and phosphorylate transcription factors to regulate differentiation, proliferation, angiogenesis and apoptosis [134]. The majority of BRAF mutations localise to the kinase domain and increase kinase activity of BRAF toward MEK. BRAF mutations are found in 1.6% and 1.8% of cases of NSCLC in smokers and nonsmokers, respectively [49].

A thymidine to adenosine transversion at nucleotide T1799A at exon 15, which results in a valine to glutamate substitution at codon 600 (V600E), is the most common mutation, representing half of all cases. The other two most frequently observed mutations are the G469A mutation (39% of cases) and the D594G mutation (11% of cases) [135]. However, many of the non-V600E mutations show only intermediate and low kinase activity; thus, their roles as driver oncogenes remain unclear [136]. Among 916 NSCLC patients, BRAF mutations were identified in 1.9%, which comprised mostly adenocarcinomas (88%), female patients (64.7%) and smokers (70.6%) [137]. Retrospective analysis of 1046 patients revealed a prevalence of 4.9% and 0.3% in adenocarcinomas and squamous-cell carcinomas, respectively, with V600E representing 56.8% [138].

The prognostic impact of the BRAF mutation remains controversial in NSCLC. Paik et al. [135] and Cardarelli et al. [139] demonstrated no significant impact of BRAF mutations on survival in advanced-stage patients. Nevertheless, the BRAF V600E subtype is associated with aggressive histology (micropapillary features in 80% of patients), and decreased disease-free survival and overall survival [139, 140]. Of the 35 patients receiving BRAF-targeted therapy (vemurafenib, dabrafenib or sorafenib) for BRAF-mutated lung adenocarcinoma in the European cohort, 83% had a V600E mutation, which appeared to be associated with a better prognosis than other subtypes of mutations (median survival 25.3 months versus 11.8 months, respectively) [141]. Another study on a broad series of 63 treated patients found a favourable prognosis of the V600 mutation compared to non-V600 mutations (3-year survival of 24% versus 0%, respectively). Smoking appeared to be more strongly associated with non-V600 mutations (11% of nonsmokers versus 42% of smokers with a V600 mutation) [142].

**PI3K/AKT/MTOR pathway**

The PI3K/AKT/mTOR signalling pathway has a diverse array of functions, including those of regulation of cellular survival, differentiation and display of stem-cell-like properties, growth, proliferation, metabolism, migration and angiogenesis [143, 144]. It plays a key role in downstream signalling by phosphorylation and activation of several effectors, including the serine/threonine kinase, AKT and mTOR (mammalian target of rapamycin), a master regulator of cell growth and metabolism. Signalling through the PI3K/AKT pathway is negatively regulated by the tumour-suppressor gene PTEN (phosphatase and tensin homologue) [145].

Although aberrations in the PI3K pathway are particularly prevalent in some cancers (including invasive breast cancer, head and neck cancer, colon cancer, cervical and ovarian cancer), they are rare in lung cancer. The most frequent alterations are mutations of PIK3CA and the loss of PTEN function (mutations or deletions). These alterations are more common in squamous-cell carcinoma (10–16% PIK3CA mutations, 15% loss of PTEN) than in adenocarcinoma (3–4% PIK3CA mutations, 3% loss of PTEN, frequently associated with KRAS or EGFR mutations) of the lung [146–149]. PI3K aberrations are associated with an unfavourable prognosis, higher burden of metastatic disease and higher incidence of brain metastases [150].

The effects of PIK3CA mutations on prognosis and response to standard therapies remain unknown. A retrospective study reported that a concomitant PIK3CA mutation resulted in decreased overall survival in EGFR-mutant lung adenocarcinoma (18 versus 33 months, p=0.006), although the authors found no evidence that a concurrent PIK3CA mutation affects the outcome of EGFR-TKI therapy [151–155].

**HER2**

HER2 mutations consist of in-frame insertions into exon 20, leading to constitutive activation of the receptor and downstream AKT and MEK pathways. Inducible expression of a HER2 mutant (HER2YVMA) in the lung epithelium of mice, resulted in the emergence of invasive adenosquamous carcinomas, with tumour maintenance requiring continuous expression of the driver, as evidenced in EGFR-driven cancer [156]. HER2 mutations have been identified in about 2–4% of NSCLC patients [50, 157], but in less than 1% of the French national database [85].
This mutation is predominantly observed in females, nonsmokers and in the adenocarcinoma subtype, which is similar to EGFR-mutated NSCLC [50, 157]. HER2 mutations could be more relevant to lung carcinogenesis than HER2 amplification or overexpression, which reflects conditions that are opposite to those of breast cancer. Recent studies suggest that HER2 mutations might be predictive for HER2-targeted therapies in lung cancer, such as HER2–TKI [158] and HER2-antibodies [159].

**ROS1**
The c-ros oncogene 1 (ROS1) is a relatively recent target for lung cancer. It encodes a tyrosine-kinase receptor from the insulin-receptor family. Chromosomal rearrangements involving the ROS1 gene were originally described in glioblastomas, where ROS1 (chromosome 6q22) was fused to the FIG gene and results have shown transformation in transgenic mice [160, 161]. In NSCLC cell lines and in primary tumours, ROS1 fusion has been identified as a driver mutation [162]. ROS1 fusion partners include SLC34A2, CD74, TPM3, SDC4, EZR, LRG3, KDELIR2 and Ccdc6 [163]. ROS1 rearrangement occurs in 1–2% of NSCLCs [94]. The kinase domain is always fully retained on the ROS1 fusion protein, and the junction point at the mRNA level always occurs at the 5′ end of exons 32–36 [164]. The ROS1-kinase domain has significant homology with the ALK-kinase domain. ROS1-positive patients share similar characteristics with ALK-positive patients, such as adenocarcinoma histology, histomorphology, younger age and a high prevalence of nonsmokers [94]. Four new fusion partners (CLTC, LIMA1, MSN and TMEM106B) of ROS1 were identified in 2014 and 2015, and the incidence and clinicopathologic characteristics of ROS1+ NSCLC patients have been described in a comprehensive meta-analysis [165]. Crizotinib is associated with a very good response rate and PFS [51, 166]. Other drugs are currently being evaluated.

**Emerging molecular targets**
The RET gene is a proto-oncogene [167, 168]. Oncogenic activation can occur via its mutation or rearrangement. In nonsmall cell lung cancers, RET rearrangements occur in 1–2% of unselected cases. These are commonly found in adenocarcinomas of never-smokers [169]. In contrast to thyroid cancer, in which Ccdc6 and NCOA4 are more common upstream partner genes, KIF5B is the most common upstream fusion partner of RET in NSCLC [170]. Cabozantinib and vandetanib showed efficiency in patients with RET-rearranged lung cancers in three phase II trials [171–173]. Other RET inhibitors, including cabozantinib, vandetanib, sorafenib and lenvatinib that are approved for the treatment of advanced thyroid cancers, and ponatinib, alectinib and sunitinib that are approved for other indications, have not been yet tested in clinical trials.

**MET** exon 14 mutations represent other new target molecular alterations in 3% of non-squamous NSCLC. These mutations are more likely detected in nonsmokers, and the response to MET inhibitors must be assessed in clinical trials, as a few case reports and limited series have demonstrated favourable outcomes [174, 175].

**Conclusion**
In conclusion, the discovery of many theranostic molecular biomarkers in NSCLC has greatly changed the classification of lung cancer, and thus its management. The epidemiology, impacts on prognosis and responses to conventional and targeted treatments are very varied, depending on genomic alteration.

**References**


