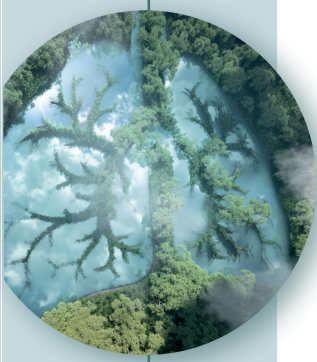
1

MOLECULAR PATHOLOGY: STRATEGIES TO CHARACTERIZE EGFR MUTATED TUMORS



Angela Listì, Fabio Turco,
Antonio Ungaro, Umberto Malapelle

THE MUTATIONAL LANDSCAPE OF EGFR MUTATED LUNG CANCERS

Lung cancer is the most common cause of cancer-related death in both sexes worldwide.[1] The two major subtypes are small-cell lung carcinoma and non-small-cell lung carcinoma (NSCLC), accounting for 15% and 85% of all lung cancer cases, respectively. NSCLC is further classified into three main morphological histotypes, including squamous-cell carcinoma, adenocarcinoma, and large-cell carcinoma, accounting for 30%, 50% and 20% of cases, respectively.[2] Over the past few decades, the identification of an increasing number of oncogenic drivers able to predict clinical response to targeted therapies has determined a radical shift from histological to molecular subtyping of lung adenocarcinoma, establishing the new paradigm of precision medicine.[3, 4] Tumor molecular profiling is now considered as a crucial step of the diagnostic and therapeutic process of advanced NSCLC patients, in order to select them to targeted therapies.[4, 5] For patients with advanced non-squamous NSCLC, the AIOM (Italian Association of Medical Oncology) guidelines recommend the systematic testing of Epidermal growth factor receptor (*EGFR*) and B-Raf proto-oncogene, serine/threonine kinase (*BRAF*) mutations, analysis of Anaplastic Lymphoma Kinase (*ALK*), ROS proto-oncogene 1 (*ROS1*) and Neurotrophic Receptor Tyrosine Kinase (*NTRK*) rearrangements, as well as the immunohistochemical evaluation of Programmed Death-Ligand 1 (PD-L1) tissue expression. According with ESMO (European Society for Medical Oncology) guidelines, the systematic testing of "emerging biomarkers", including RET rearrangements, MET exon 14 skipping alteration, KRASpG12C as well as HER2 mutations is also suggested in order to candidate oncogene-addicted patients to targeted treatments available in the context of clinical trials or compassionate use/expanded access programs.[6, 7]

EGFR was the first predictive biomarker identified in NSCLC: in 2004, Lynch *et al.* and Paez *et al.* reported for the first time that mutations involving the tyrosine kinase domain of EGFR protein may predict responsiveness to the first-generation tyrosine kinase inhibitors (TKI) gefitinib.[8, 9] Epidermal growth factor (EGF) was originally isolated by Stanley Cohen in 1962 while the presence of a specific binding site for EGF was confirmed in 1975.[10] Implication in cancer was described in 1980, when the avian erythroblastosis tumour virus was found to encode an aberrant form of the EGFR.[11] EGFR is a 170kDa tyrosine kinase receptor belonging to the ERbB family, encoded by the *EGFR* gene, that is located in the short arm of chromosome 7 (7p11.2).[12] *EGFR* gene encodes for a polypeptide precursor of 1210 amino acids that, after cleavage of the N-terminal sequence, determines the generation of mature EGFR protein, a 1186-residue protein, localized on the cell membrane.[13, 14] There are four members of the ErbB family: EGFR (also termed ErbB1/HER1), ErbB2/Neu/HER2, ErbB3/HER3 and ErbB4/HER4.[14] EGFR protein is composed by an extracellular ligand-binding domain (or ectodomain) containing: L1 (leucine-rich repeats 1), CR1(-cysteine-rich 1), L2 (leucine-rich repeats 2), CR2(cysteine-rich 2), alternatively indicated as Domains I–IV, a short transmembrane sequence, a cytoplasmic region that incorporates a tyrosine kinase and a C-terminal phosphorylation domain.[15] Each ErbB receptor binds a panel of activating ligands except ErbB-2/HER-2, which has no known ligand and acts primarily as a dimerization partner for other ErbB receptors.[16] Conversely, Epidermal Growth Factor (EGF), Transforming Growth Factor-α (TGF-α) and amphiregulin uniquely bind EGFR.[15] Crystal structures of the EGFR ectodomain with EGF or TGF-α demonstrate that binding of ligand to the L1 and L2 domains leads to a conformational change in which the receptor takes on an extended form that exposes this dimerization loop and allows for interaction of receptor ectodomains.[15] The two Cys-rich domains are extended repeats of seven small disulphide-containing modules. A beta-hairpin loop extends from the first Cys-rich domain to contact the C-terminal portion of the second Cys-rich domain, creating a large pore structure.[17] Dimerization stimulates the intrinsic tyrosine kinase activity of the receptors and triggers autophosphorylation of specific tyrosine residues within the cytoplasmic domain. The pattern of phosphorylation creates docking sites for the recruitment of diverse effector proteins, leading to the generation of intracellular signal transduction cascades and a variety of cellular responses, including Rat sarcoma (Ras) / Mitogen-Activated Protein Kinase (MAPK), Phospholipase C, Gamma1 (PLCγ1) / Protein kinase C (PKC), phosphatidylinositol-3-kinase (PI3K) / Protein kinase B (Akt), and signal transducer and activator of transcription (STAT) which are involved in cell proliferation and survival.[7]

ErbB receptors play fundamental roles in development, proliferation and differentiation across different epithelial, mesenchymal and neuronal tissues. Aberrant activation can promote abnormal cell proliferation by transmitting mitogenic signals.[10] EGFR signaling is frequently altered in several human cancers due to *EGFR* gene amplification and/or mutations as well as protein overexpression.[18] Oncogenic *EGFR* mutations often determines an alteration in receptor endocytosis, which contributes to increased signaling properties. In some cases, mutations directly affect the intracellular domain (*i.e.*,

EGFRvIV and EGFRvV mutants), thus influencing receptor ubiquitination and lysosomal degradation. In other cases, mutations are found in the extracellular domain (*i.e.*, EGFRvIII), which leads to the ligand-independent receptor activation.[18] *EGFR* activating mutations drive tumorigenesis of a significant fraction of non-squamous NSCLC, and their frequency is higher in South-East Asian ethnicity (40-60%) compared to Caucasian one (10-15%), as well as in never or light smokers, young, female, adenocarcinoma patients.[12] The vast majority of *EGFR* activating/sensitizing mutations were reported among exons codifying for the tyrosine kinase (TK) domain.[19] The two most common/classical activating mutations are in-frame deletions involving exon 19 (amino acid residues 746–750, about 45%) and a missense substitution at position 858 (exon 21 p.L858R, about 40%) where leucine is replaced by arginine, resulting in a constitutive activation of the receptor in absence of ligand binding.[3, 20] Classical mutations have been included in the majority of clinical trials testing EGFR-TKI as first-line treatment in advanced NSCLC patients, therefore have been historically considered reliable predictors of clinical response to all generation EGFR-TKI currently available for clinical use.[16] Conversely, *EGFR*-mutated NSCLCs are associated with uninflamed phenotype (increased T regulatory cells, decreased tumor infiltrating lymphocytes, down-regulation of major histocompatibility complex) as well as weak immunogenicity in the tumor microenvironment [lower PD-L1 expression and tumor mutational burden (TMB)],[21] making these tumors less responsive to immune-checkpoint inhibitors.

The remaining 15% of *EGFR* "uncommon mutations" occur within exons 18-21 with clinically variable responses to targeted drugs. It is complex to assess the efficacy of TKIs in patients with uncommon *EGFR* mutations since most phase 3 studies did not allow enrollment of these patients.[5] In addition, detection techniques for uncommon mutations were not as advanced at the time of the original registration trials.[22] More recently, the development of sensitive next-generation sequencing techniques and multiplex polymerase chain reaction assays have increased the detection rate of uncommon mutations in the clinical practice. As a consequence of these technological developments, more than 600 *EGFR* variants are currently being described in the Catalogue Of Somatic Mutations in Cancer database, even if either the biological significance or responsiveness to TKIs remains unknown in most cases.[21]

Similarly to common *EGFR* mutations, TMB has been reported to be lower in these patients than in *EGFR* wild-type NSCLC patients, likely reflecting non-smoking habits.[23] However, an intriguing association with male sex and smoking history has been reported in a subgroup of NSCLC patients harboring exon 18 mutations.[24] Furthermore, patients with uncommon mutations, high PD-L1 expression and smoking history seem to benefit from immune-checkpoint inhibitors administration, as reported in small case series.[4]

Exon 18 mutations collectively account for 3-4% of all *EGFR* mutations, more commonly including point mutations that encompass a glycine change to serine, alanine or cysteine (p.G719S/A/C in the 97% of cases) within the codon 719, and less frequently involving the codon 709 (p.E709K/A/G/V, 0.3% of all *EGFR* mutations).[24] p.E709X mutations account for

<0.5% of all *EGFR* mutations. Rare exon 18 mutations include: p.V689M, p.S720P/F, p.P699S, p.N700D, p.G721A, p.V740A, and p.L718P.[25] Regarding exon 18 deletions, p.E709_T710delinsD is the most common, even though it can be missed when using diagnostic commercially available kits.[24] In addition to the frequent classical exon 19 deletions (including up to 30 alterations), exon 19 harbors many other, less investigated, molecular alterations. Of note, exon 19 deletions may involve the entire exon (codons 746-761) and, in a non-negligible percentage of cases (>50%), may be associated with additional insertions (indels) leading to replacement of the deleted amino acids with a non-native residue (such as the p.L747-A750delinsP, where a proline residue is introduced in substitution).[24] Exon 19 insertions are a relatively uncommon subset of *EGFR* alterations with a reported frequency of <0.5%.[25] Uncommon exon 19 mutations also include rare substitution mutations, such as p.L747P/S, as well as insertions, such as p.I744_K745KIPVAI.[26] Finally, several other rare exon 19 mutations have been sporadically reported, and their sensitivity to first-/second-generation EGFR TKIs is variable, including sensitive (p.L747F, p.P733L, p.K757R, p.E746G, and p.V742X) and resistant mutations (p.D761Y, p.E746V, and p.L747S).[25]

Exon 20 insertions are the most prevalent and heterogeneous group of *EGFR* molecular aberrations.[26] These insertions encompass residues from 762 to 775 (spatially located after the C-helix of the EGFR kinase domain) including over 64 unique variants described to date, with an estimated incidence of 3-8% of all *EGFR* mutations. From an epidemiological point of view, similar to other *EGFR* mutations, these alterations showed a higher incidence in female, non-smoking patients with adenocarcinoma subtype while no significant differences have been reported among different ethnicities.[23] The *EGFR* insertions consist of in-frame insertions or duplications across 15 amino acids that encompass residues from 761 to 775 and include p.A763_Y764insFQEA, p.A767_V769dupASV, p.V769_D770insASV, p.H773_V774insNPH, p.V774_C775insHV, p.D770_N771insSVD. Interestingly, the *EGFR* exon 20 p.V769_D770insASV seems to occur more frequently in older patients (≥65 years) compared to the *EGFR* exon 20 p.A763_Y764insFQEA (<65 years).[23]

The majority of *EGFR* exon 20 insertions occurred near to the C-helix domain, generating a wedge at the end of the C-helix and determining a constitutive activation of the tyrosine kinase activity. It has been demonstrated that *EGFR* exon 20 insertions result in a decreased affinity to the currently available EGFR-TKI, particularly of first- and second-generation. Regarding exon 20-point mutations, the single p.S768I mutation represents approximately 1-2% of all *EGFR* mutations, even if often existing as compound mutation. In addition to p.S768I, a number of other rare exon point mutations have been described including p.L774X, p.R776X, p.V786M, and p.Q787Q.[26] *EGFR* exon 20 p.T790M point mutation is reported in about half of the acquired resistant cases to first- and second-generation EGFR TKI.[19] This mutation affects the "gatekeeper" threonine residue leading to kinase-targeted drug resistance. *EGFR* exon 20 p.T790M point mutation is present also in about 1-2% of EGFR-TKI-naïve NSCLC, sometimes co-occurring with classical exon 19 or 21 mutations. Interestingly, rare germline mutations encoding *EGFR* exon 20

p.T790M (about 2%) appear to cause inherited susceptibility to lung cancer.[23] The most commonly identified *EGFR* mutation occurring within exon 21 includes the point mutation which substitutes arginine for leucine at codon 858 (p.L858R). However, a heterogeneous group of other non-classical *EGFR* point mutations may occur besides p.L858R. The p.L861Q mutation represents the second most frequent exon 21 mutation, accounting for approximately 1-2% of all *EGFR* mutations and not rarely occurring as complex mutation. Other rarer mutations may be identified in the exon 21, but their EGFR-TKI sensitivity is generally low (p.L861R, p.L862V, p.V851X, p.A859X) or uncertain (p.E866K, p.H825L, p.P848L, p.H870Y/R, and p.G836S) although the limited data available to date do not allow definite conclusions.[4] It is also becoming increasingly apparent that many tumors have "subclonal" *EGFR* mutations with a low variant allele frequency, which may occur in isolation or coexisting with an independent common or uncommon *EGFR* mutation (termed a "complex" or "compound" mutation). As different mutations have different effects on the tertiary structure of the EGFR protein, it is unsurprising that their sensitivities to EGFR TKI vary widely.[22] With the development of sensitive detection technologies, it seems that up to 25% of patients with *EGFR* mutation-positive NSCLC harbor compound mutations. These generally comprise a clonal driver along with a subclonal mutation. This aberration may expand under treatment owing to selective pressure and may contribute to acquired resistance to TKI. The sensitivity or resistance of compound mutations to EGFR TKI seems to be largely influenced by the accompanying mutation.[4]

Which technique: Pyrosequencing, Sanger, Real-time PCR or Next-Generation Sequencing?

Molecular testing for *EGFR* mutations is currently considered as an integral part of the standard of care in advanced NSCLC patients. Until 2015, Sanger sequencing (SS) and/or pyrosequencing (PS) were the most widespread technique used to analyze *EGFR* mutational status in advanced NSCLC patients.[27] SS was considered for a long time the gold standard approach for *EGFR* mutations detection because it enables the evaluation of the entire gene sequence and the identification of unknown mutations.[28] Also known as "chain termination method," it was initially developed by Frederick Sanger and his colleagues in 1977 and was subsequently updated to an "automatized" approach, to determine the sequence of nucleotide bases commonly 800-1,000 bp in length. Automated sequencing instruments combine sequencing with fluorescently labelled primers or dideoxy chain terminators with polyacrylamide gel electrophoresis and computer data capture. This kind of sequencing is estimated as 99.99% base accuracy and was used by the Human Genome Project to determine the sequences of relatively small fragments of human DNA (900 bp or less).[29] The major limitations of this method are high costs, laboratory intensiveness, and low sensitivity. The sensitivity of this platform required almost a 40-50% tumour cellularity within the tested sample (20-25% mutated allele assuming heterozygosity at the targeted chromosomal site).[28, 29] Thus, a shift to Pyrosequencing has been performed by different laboratories worldwide to perform *EGFR* mutational analysis, because PS have an inferior limit of detection (~5% *versus* ~20% for Sanger) of

mutant alleles. Pyrosequencing technology was developed at the KTH Royal Institute of Technology and considered as the first alternative to the conventional Sanger method for *de novo* DNA sequencing. PS use a bioluminescence technology where the pyrophosphate released during the nucleotide incorporation into a growing DNA chain produces light through a series of enzymatic reactions.[30] PS can identify individual bases or short stretches of nucleic acid sequences at predetermined positions. Furthermore, the commercially available pyrosequencing kits properly identifies the most common *EGFR* exons 18-21 mutations, but is not able to cover all *EGFR* molecular alterations.[31] Quantitative real-time polymerase chain reaction (qRT-PCR) methodology is based on targeted amplification by using fluorescent probes to identify known mutation in target sequence.[37] In NSCLC patients, this technology was largely adopted to test the most common *EGFR* mutations in exons 19 and 21.[19] RT-PCR is useful to identify either "common" or "uncommon" *EGFR* mutations, including *EGFR* exon 20 insertions.[26] However a recent work revealed that current commercially available PCR kit methods may miss around 50% of *EGFR* exon 20 insertion variants compared to next generation sequencing (NGS) analysis.[32] Despite the high specificity, an important limitation of all targeted-based approaches is the ability to detect only known and well characterized mutations. Thus, RT-PCR technology may be limited when low frequent and heterogeneous alterations are taken into account. In this scenario, NGS assays, able to detect several hotspot gene mutations for different patients simultaneously, are a highly sensitive and specific tools for molecular assessment of less frequent gene mutations.[33] In addition, different from targeted-based approaches, NGS is able to identify either known or unknown mutations within gene panel reference range, ensuring higher diagnostic accuracy, faster turnaround time for low sample volumes, and lower costs.[34] To date, several NGS panels are commercially available enabling the simultaneous analysis of a plethora of clinically relevant hotspots in target genes, including *EGFR*.[35, 36]

Dall'Olio *et al.* retrospectively compared the molecular analysis results obtained by conventional approaches [RT-PCR, pyrosequencing and immunohistochemistry (IHC)] with the NGS Oncomine focus assay. Overall, a similar rate of alterations in *EGFR*, *ALK* and *KRAS* (14.3 % *vs.* 16.5 %, 6.3 % *vs.* 6.3 % and 33.5 % *vs.* 36.0 %, respectively) was reached. Only for *ROS1* translocations a significant difference between the two diagnostic algorithms [4.7 % fluorescent *in situ* hybridization (FISH) *vs.* 0.7 % NGS] was reported.[19] Similar evidences were reported by Yu *et al.*, that demonstrated the technical feasibility of NGS approach on a broad cohort of specimens, mainly represented by small biopsies (70.5 %), with an higher successful analysis rate for NGS (88.4 %) compared to single test sequential approaches (76.6 %).[19] Based on these and other scientific evidences, ESMO has recently recommended NGS as standard approach to routinely profile advanced stage NSCLC patients with non-squamous histology.[37] An Italian survey, addressing 51 referral institutions for thoracic malignancies management, showed that RT-PCR approach still represents the most adopted technique for *EGFR* mutation testing in clinical practice, while only 30% of molecular pathology laboratories routinely use NGS to profile advanced NSCLC patients (**TABLE 1.I**).[19]

TABLE 1.I.
Diagnostic platforms for EGFR molecular testing in NSCLC patients.

Platform	Diagnostic Practice 2021	Diagnostics Aims	Properties
Pyrosequencing	No	Targeted methods	• Common • Limited number of DNA sequences (DNA fragment 100-400 max pb)
Sanger sequencing	Rarely	Screening methods	• Common or uncommon alteration • High input • High neoplastic cell percentage • High cost and TAT
NGS analysis	Yes	Screening methods	• Common or uncommon • High throughput test • Tissue and/or blood monitoring in ultra-deep • Low cost and rapid results
qRT-PCR	Yes	Targeted methods	• Rapid results • Cost effectiveness • Limited number of DNA sequences • Allele frequency of at least 1% for some assays (Cobas® EGFR Mutation Test v2, Therascreen)
ddPCR	Yes	Methods	• High sensitivity (detection of low allele frequency, 0.005-0.1%) • Limited number of DNA sequences

Methods were assigned into one of two broad categories: "screening methods" (detecting all mutations, including novel variants in exons 18-21); "targeted methods" (detecting specific, known mutations). The advantages and disadvantages of screening and targeted testing methods are summarized in the properties column.

Liquid biopsy: diagnosis and treatment monitoring

In patients with oncogene-addicted NSCLC, liquid biopsy is emerging as not only complementary to tissue-based analysis but also acceptable as the initial approach ("plasma-first") for biomarkers evaluation at the time of diagnosis, as well as for monitoring the efficacy of targeted therapies. Finally, a plasma-first approach is appropriate for identification of acquired resistance mechanisms to targeted therapies in many clinical settings.[38] The molecular characterization on tumor tissue represents the current "gold standard" and it is the widely adopted approach to detect both *EGFR* and other predictive biomarkers at the time of lung cancer diagnosis in clinical practice.[38] However, the increasing number of targetable biomarkers has required a larger amount of tumor tissue to molecular investigations and about 30% of NSCLC have no adequate tissue specimens at the time of diagnosis.[39] Although the advent of NGS techniques allowed to perform adequate molecular screening even in the presence of limited amount of tissue, there is still a significant fraction of patients who cannot be candidate to invasive diagnostic procedures or with inconclusive results at histological examination.[39] In these cases, liquid biopsy by circulating tumor (ct) DNA

analysis has definitely proven to be a useful alternative for tumor genotyping, since fractions of DNA originating from tumor cells carry tumor-related alterations detectable by NGS or PCR based techniques.[40] A series of diagnostic accuracy studies have recently investigated the comprehensive evaluation of tumor molecular profiling by highly sensitive NGS technique for ctDNA assessment in advanced NSCLC patients, showing high concordance with tissue genotyping, thus confirming ctDNA-based molecular analysis as a reliable alternative to the standard tumor tissue genomic profiling.[41, 42] This evidence has recently led to the approval of the first ctDNA NGS diagnostic assay in United States, while in Italy liquid biopsy application at the time of diagnosis is currently limited to the detection of *EGFR* sensitizing mutations in the clinical setting.[19] Different studies and meta-analysis consistently demonstrated high specificity (over 90%) for *EGFR* mutation testing by ctDNA analysis, while lower sensitivity, ranging from 50% to 85% depending on the used technology.[43] Indeed, not all metastatic tumors shed sufficient amounts of DNA into peripheral circulation, and particularly treatment-naive patients with indolent, slow-growing, intrathoracic disease may be at higher risk of false-negative findings by ctDNA analysis.[44]

Since the approval of the third generation EGFR-TKI osimertinib in the second line treatment of *EGFR*-mutant and exon 20 p.T790M-positive NSCLC patients who failed prior EGFR-TKIs, liquid biopsy has been considered the new standard approach to detect exon 20 p.T790M resistance mutation by ctDNA analysis. Similarly, to the baseline setting, different studies showed appropriate specificity but low sensitivity of ctDNA-based exon 20 p.T790M testing, with false negative rate ranging from 20% to 50% according to the different used detection technologies (RT-PCR, digital-PCR, NGS) as well as disease burden.[45] Therefore, those patients with negative results on plasma genotyping should be always candidate to tissue re-biopsy, where clinically feasible, in order to confirm exon 20 p.T790M mutation status. The subsequent shifting of osimertinib in the upfront setting has dramatically changed the molecular landscape of resistance mutations occurring under the third generation EGFR-TKI treatment, including both *EGFR*-dependent mutations, like exon 20 p.C797X, and *EGFR*-independent alterations, as *MET/HER2* amplifications, *BRAF/PI3KCA* mutations, as well as the morphological switching to SCLC, among the most common reported.[46] Although ctDNA analysis has shown to accurately detect the majority of osimertinib-related mechanisms of resistance in the FLAURA trial,[47] however its application is still to be considered experimental, while several ongoing trials are evaluating the molecular profile of patients progressing to first-line osimertinib. Among them, the phase II ELIOS study (NCT03239340) aims to assess the concordance between tumour tissue and ctDNA-based NGS analysis for the detection of the mechanisms of resistance to first-line osimertinib in EGFR-mutant NSCLC patients. Nevertheless, it should be borne in mind that ctDNA analysis was not able to identify other resistance mechanisms, such as histologic transitions, that should be evaluated by morphologic evaluation.[38, 42]

To this end, a variety of liquid biopsy assays have been developed to complement routine tissue-based diagnostics and/or identifying acquired resistance mechanisms. RT-PCR is the most widely performed technique for