Molecular diagnostics of lung cancer in the clinic

Lynette Sholl

Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Correspondence to: Lynette Sholl, MD. Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. Email: lmsholl@bwh.harvard.edu.

Abstract: According to current practice guidelines, all patients with advanced non-small cell lung cancer (NSCLC) should undergo predictive biomarker testing. For squamous cell carcinoma patients, PD-L1 immunohistochemistry is indicated to select patients for immunotherapy in the first line. For lung adenocarcinoma, all patients with advanced disease should undergo testing for epidermal growth factor receptor (EGFR) mutations, ALK and ROS1 rearrangements, and PD-L1 expression to predict response to EGFR, ALK, or ROS1 targeted inhibitors or immunotherapy, respectively. Besides these, a number of other biomarkers are under clinical investigation as predictors of response to targeted therapies, including BRAF, ERBB2, MET splice mutations and amplification, and RET rearrangements. Successful testing for this complex array of molecular targets demands careful coordination between proceduralists, pathologists and molecular laboratories to ensure proper tumor tissue handling following biopsy as well as judicious use of diagnostic immunohistochemistry. Even so, sample failure rates due to inadequate tumor tissue are high in practice, particularly when using sequential testing methods. Use of next generation sequencing (NGS) in clinical practice can enable detection of multiple targets and multiple alteration types (mutation, gene copy change, and rearrangement) simultaneously even with small amounts of input nucleic acids, thus increasing molecular testing success rates. In patients with an established lung cancer diagnosis but with prohibitively limited amounts of tumor tissue or who are experiencing relapse, analyses of circulating tumor DNA (ctDNA) from the plasma can serve as an alternate testing substrate, however the more limited clinical sensitivity of this approach must be taken into account. This review will explore the indications for and pitfalls of routine NGS and plasma genotyping in the clinic, including the intersection of these technologies.

Keywords: Epidermal growth factor receptor (EGFR); ALK; ROS1; lung adenocarcinoma; next generation sequencing (NGS); plasma genotyping; circulating tumor DNA (ctDNA)

Submitted May 04, 2017. Accepted for publication Jun 01, 2017.
doi: 10.21037/tlcr.2017.08.03

View this article at: http://dx.doi.org/10.21037/tlcr.2017.08.03

Introduction

Lung cancer is the number one cause of cancer deaths globally (1). Approximately 225,000 new cases of lung cancer are diagnosed annually in the United States (2). In current practice, the most common subtype—nearly half of all lung cancers—is adenocarcinoma (3). Adenocarcinoma is more common in smokers, however it is also the most common lung cancer subtype diagnosed in nonsmokers (4). Adenocarcinomas typically arise in the lung periphery and may be asymptomatic in their early stages. As a result, the majority of patients come to clinical attention with locally advanced or distantly metastatic disease; in the latter situation, fewer than 5% of patients live to 5 years despite therapy (1).

Key biomarkers in the clinic

The genetics of adenocarcinoma are diverse, but heavily influenced by the patient’s smoking history (5). The earliest recognized mutations in non-small cell lung cancers (NSCLC) were identified in KRAS and TP53. However the first clinically impactful genomic discovery
was made in 2004, when mutations in the kinase domain of the epidermal growth factor receptor (EGFR) were described specifically in the tumors of lung cancer patients who responded to tyrosine kinase inhibitor (TKI) therapy directed at EGFR (6,7). These EGFR kinase domain mutations, >90% of which occur in exon 21 (L858R) and exon 19 (small insertions-deletions affecting the ELREA motif), lead to constitutive activation of downstream pro-
growth, oncogenic signaling pathways. Fortuitously, these mutations also sensitize the tumor cells to EGFR TKIs and predict response to a broad spectrum of EGFR TKIs, such as first generation inhibitors erlotinib and gefitinib (8). EGFR mutations are identified almost exclusively in lung adenocarcinomas, occur more commonly in light or never smokers and are enriched in women and individuals of Asian ethnicity (9). EGFR is the second most commonly mutated driver oncogene in lung adenocarcinoma after KRAS in the United States—about 15% in Caucasians and African Americans—and is the most commonly mutated oncogene in lung adenocarcinoma in Asian populations (~60%) (10,11).

Rearrangements involving ALK and ROS1 were first described in lung adenocarcinoma in 2007 (12,13). Crizotinib, a commercially available inhibitor originally designed to target Met, proved effective against lung cancers harboring either ALK or ROS1 alterations (14,15) and has been approved for treatment of lung cancers with proven rearrangements. Both of these alterations are rare (<5% of lung cancers) but are enriched among light to never smokers and are seen almost exclusively in adenocarcinomas (16,17). Despite these clinicopathologic correlations that have been seen in EGFR, ALK and ROS1-altered lung tumors, it is clear that clinical features are neither highly sensitive nor specific for selecting patients for targeted inhibitors (18). Therefore, all patients with advanced lung adenocarcinoma should undergo testing for EGFR mutations and ALK and ROS1 rearrangements, irrespective of smoking status. In general, this testing is not indicated in patients with a diagnosis of squamous cell carcinoma or small cell carcinoma, however there are rare reports of, for example, de novo EGFR-mutated small cell carcinoma or ALK-rearranged squamous cell carcinoma in never smokers (19,20). Therefore, molecular testing is advised in patients with a histologic diagnosis that is out of keeping with their smoking history.

Relapse following targeted therapy is almost inevitable, and tends to occur after about a year of therapy on EGFR TKIs and after a median of 8 and 19 months, respectively, following first-line targeted therapy in the setting of ALK and ROS1 rearrangements (15,21). The mechanisms of resistance are relatively well defined. For EGFR, 50–60% of patients acquire the EGFR T790M mutation at the time of relapse (22). T790M reduces the efficacy of first generation EGFR inhibitors, but third generation inhibitors can overcome this resistance mutation, and one, osimertinib, has been FDA approved specifically for patients with a proven T790M mutation in the relapse setting (23). Other less common mechanisms of resistance include MET amplification, PIK3CA pathway activation, and small cell transformation (22). In ALK-rearranged patients, crizotinib resistance most commonly takes the form of a wide variety of secondary mutations occurring in the ALK kinase domain. Second and third generation ALK inhibitors can variably overcome these secondary mutations. While some authors have advocated for routine biopsy at relapse to define the mechanism of ALK inhibitor resistance (24), this practice is not widely employed, and alternative inhibitors are typically used empirically. Mechanisms of crizotinib resistance in the setting of ROS1 rearrangement are less-well defined, however mutations in ROS1 at codons 2032 and 2033 have been reported in individual cases (25,26).

Most recently, immune checkpoint blockade, or immunotherapy, has proven effective in a variety of tumor types, including lung cancers. Immunotherapeutics target some component of a regulatory network that keeps T cell response in check in inflammatory states; tumors can effectively hijack this network by, for example, upregulating surface PD-L1 expression to evade T cell-mediated anti-
tumor responses. Approved immunotherapies in the lung include anti-programmed death-1 (PD-1) and anti programmed-death ligand-1 (PD-L1) antibodies (27). Biomarker analyses have shown that as a class the PD-1 and PD-L1 inhibitors appear to have greater efficacy in patients whose tumors express PD-L1, with a generally positive correlation between extent of expression and likelihood of response to treatment (28). Several drugs have been approved for NSCLC (both adenocarcinoma and squamous cell carcinoma) irrespective of the PD-L1 status (27), however the anti-PD-1 antibody, pembrolizumab, was approved for first line use only in patients with 50% or greater tumor cell expression of PD-L1 based on results of randomized controlled trials (29).

As a result, in the United States, there are four possible therapeutic options in the first line setting for patients with advanced (stage IIIb or IV) adenocarcinoma (Figure 1), and biomarker testing is required before deciding on
the course of therapy. Retrospective analyses of PD-1 inhibitor trials have demonstrated that EGFR mutation and ALK rearrangement predict inferior response to immunotherapy (30); pembrolizumab is, therefore, approved only for patients with EGFR or ALK negative tumors. Data on immunotherapy efficacy in ROS1-rearranged tumors is limited at this time.

A number of other important oncogenic alterations are recognized in lung adenocarcinoma. These include uncommon but potentially targetable alterations in BRAF, ERBB2, MET and RET as well as common but difficult to target hotspot mutations in KRAS. In the vast majority of cancers, these alterations occur in a mutually exclusive fashion and represent independent oncogenic driver events (Figure 2) (31-33). A number of early clinical trials and case series support a role for routine testing of these targets (Table 1). In addition, there is increasing recognition of the prognostic importance of tumor suppressor genes, including adverse impact of TP53 mutations in patients with EGFR-mutated adenocarcinomas as well as dual RB1 and TP53 mutations as predictors of small cell transformation following EGFR TKI therapy (45,46). Other alterations, such as loss of SMARCA4, have been more broadly implicated as negative prognostic factors (47,48), whereas still others, such as loss function of STK11/LKB1, may predict specific therapeutic vulnerabilities among KRAS mutated tumors (49,50). Tumor mutational burden as defined by whole exome sequencing may predict response to immunotherapies (51), and some early evidence suggests that even smaller targeted panels (~300 genes) can generate informative mutation burden data (52).

The large number of clinically relevant genomic
biomarkers in lung cancers presents a practical challenge to laboratories: sequential testing of up to eight different genes and ten or more targets within these genes is costly, time consuming, and demands significant amounts of tumor nucleic acid. In many cases the amount of available tumor tissue is less than needed to complete multiple molecular assays, as well as FISH and IHC (32). The current demands relating to biomarker testing demand a carefully coordinated effort between the oncologist, proceduralist, pathologist, and molecular laboratory to ensure proper specimen handling. For one, the indication for the biopsy should be clearly communicated down the line, so that the tissue is handled optimally for molecular testing, including minimized ischemic time, use of 10% buffered formalin as a standard fixative, and avoidance of fixatives that degrade nucleic acids (hydrochloric acid-based decalcification solution) or inhibit PCR (heavy metals such as B+). Secondly, the receiving pathologist should be informed if and when a diagnosis is already established in order to minimize diagnostic immunohistochemistry (18). In general, judicious use of immunohistochemistry is advised, with the core panel of diagnostic markers restricted to TTF-1 and p40, followed by other stains only if these are uninformative and the clinical picture is uncertain.

### Next generation sequencing (NGS)

Advances in sequencing technology have yielded relatively cost-effective clinical testing platforms that leverage massively parallel, or NGS technologies; these platforms allow for multiplexing of gene targets over several orders of magnitude and can generate reliable results with anywhere from a few to several hundred nanograms of DNA (53). The choice of NGS platform will be informed by cost, anticipated test volume, breadth of genomic targets desired, and sensitivity desired. This latter variable dictates the depth of sequencing required. In general, the costs go up with number of targets (i.e., costs of whole genome > whole exome > targeted exome). Reagents, including chips for library preparation and flow cells for sequencing are expensive; these costs can be reduced by increasing the numbers of cases run at once (batching)—a step that is facilitated by tagging individual specimens with molecular barcodes to permit sample deconvolution during bioinformatic processing (53). Choice of platform also dictates the types of alterations that can be detected. In general, amplicon sequencing is preferred for targeted panels that are optimized for read depth and assay sensitivity, whereas hybrid capture sequencing is the approach of choice when breadth is desired (whole genome, whole exome, large targeted panels).

Both amplicon and hybrid capture assays can detect relative copy number alterations (i.e., amplification or gene deletion events), however hybrid capture sequencing is generally employed for this purpose, especially when more global copy change data is desired. In addition, the hybrid capture approach should be employed when translocation

---

**Table 1** Targetable alterations in lung adenocarcinoma under clinical investigation

<table>
<thead>
<tr>
<th>Target</th>
<th>Variant type</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRAF</td>
<td>V600E mutations</td>
<td>Planchard et al. (34,35)</td>
</tr>
<tr>
<td>ERBB2</td>
<td>Exon 20 insertion mutations</td>
<td>De Grève et al. (36), Gandhi et al. (37), Mazières et al. (38)</td>
</tr>
<tr>
<td>MET</td>
<td>Exon 14 splice mutations</td>
<td>Awad et al. (39), Paik et al. (40), Liu et al. (41)</td>
</tr>
<tr>
<td>RET</td>
<td>Rearrangement</td>
<td>Drilon et al. (42-44)</td>
</tr>
</tbody>
</table>

**Figure 2** An oncoprint of the five most common oncogenic driver mutations in lung adenocarcinoma, not including rearrangements, derived from the American Association for Cancer Research GENIE registry.
detection is desired, as DNA-based rearrangement detection requires that the assay capture large regions of intronic sequence, where breakpoints most commonly occur. Hybrid capture sequencing is sensitive and specific for translocation detection when the breakpoints and partner genes are well-defined for a given target (54). The sensitivity of this approach, however, is often limited for specimens with low tumor content relative to normal contaminating stromal cells. In addition, rearrangement detection is a thorny problem for bioinformatics, as short sequencing reads that capture two segments of the genome at once and support the presence of a rearrangement event may be discarded by some algorithms as “poorly mapped” (55). Algorithms must draw a distinction between poorly mapped reads that represent rearrangement, versus those that represent low quality sequencing, and their ability to do this defines their sensitivity and specificity and ultimately the reliability of the sequencing-based approach to rearrangement detection (56). Copy number calls, which are calculated based on a ratio of the read depth at a given locus in the sample relative to a known diploid normal sample, are also vulnerable to false negative and positive results. The former occurs in the context of low tumor content, the latter may occur when the DNA is heavily degraded. Given these limitations, parallel methods such as FISH should be available to confirm or refute unexpected or low quality findings, or for use in specimens with tumor content that is too low to generate a reliable translocation or copy number result by sequencing. IHC for protein overexpression, when available, may also help to confirm the significance of novel rearrangement breakpoints or partners. Labs may also consider use of RNA-based sequencing assays for translocation detection; significant technical advances, including the introduction of anchored multiplex PCR, now allow for fusion sequence detection from the short fragments of RNA present within formalin fixed tissues (57). Practically speaking, labs may then consider introducing a single large comprehensive hybrid-capture based panel, multiple smaller amplicon-sequencing based panels, or one smaller DNA based panel for mutation detection and another RNA-based panel for fusion detection, all supported by selected standalone assays for selected critical biomarkers (e.g., ALK, ROS1 FISH and/or IHC).

In contrast to traditional (Sanger) sequencing methods, where the readout appears as an averaging of the nucleotide content at any given position, in NGS each individual DNA molecule is separated in space and sequenced in a parallel fashion; bioinformatics tools then assemble the individual sequences, typically in relation to a reference genome (55). Using any of a variety of publicly available genome viewing tools (Integrated Genome Viewer; Genome Viewer, both from the Broad Institute, Cambridge, MA, USA) the read “pile-ups” can be directly visualized and rare events easily observed. The ability to visualize the data in such a granular fashion can be hazardous, however, as it presents the risk of over-interpreting very low level mutational events as clinically significant when they may in fact represent technical artifact (Figure 3) (55). As a result, bioinformatics pipelines should be employed to generate automated calls; these are validated to optimize the assay sensitivity and specificity within the confines of the particular sequencing platform. Laboratorians should approach manual review of NGS data with caution and avoid enthusiastic endorsement of low level reads that fall below the limit of assay detection determined in the course of validation.

**Liquid biopsy**

So-called liquid biopsy refers to detection of tumor components in bodily fluids. These components may be viable cells (circulating tumor cells) or tumor DNA [circulating tumor DNA (ctDNA)]. Fluids that may contain these components include blood (plasma), urine, saliva, cerebrospinal fluid and liquid cytology specimens (58,59). This review will focus on the practice of ctDNA analysis within the cell free DNA (cfDNA) population in the plasma, as this is the area of most intense study and clinical assay development in lung cancer patients. Plasma testing is hugely attractive from a clinical perspective because tissue biopsies are often insufficient for molecular testing, thus necessitating repeat biopsy (32,54). Use of plasma testing may (I) allow a patient to forego additional invasive biopsy procedures; (II) reduce turnaround time by eliminating the need to schedule a procedure and to process the pathology specimen; (III) reduce sequencing artifact introduced by formalin fixation of tissue specimens; and (IV) be less biased by local tumor heterogeneity.

All individuals have some level of cfDNA detectable in the plasma—this is a product of normal cell apoptosis, with release of ~160 base pair, nucleosome-bound fragments of DNA into the circulation (58). Tumor cells also release their contents into the circulation, and the amount of detectable ctDNA is proportional to the overall burden of disease. Studies using droplet digital PCR (ddPCR) have shown that known tumor driver mutations are detectable in about 60% of lung cancer patients with a single metastatic site...
and 100% of those with four or more metastatic sites (60). Patients with early stage disease commonly do not have detectable levels of ctDNA using most currently available technologies (61,62). In general, ctDNA release from lung cancer is detectable at levels of 0.1% to 5% of total cfDNA (63). Therefore, highly sensitive techniques are needed in most patients to detect tumor specific alterations within the plasma.

Published approaches to ctDNA analysis in lung cancer have largely focused on EGFR mutations, including detection of activating hotspot mutations and the EGFR TKI resistance mutation T790M. Methods in routine use include real time PCR, ddPCR, and NGS. Indeed, the FDA has approved the cobas EGFR mutation test v2 (Roche Diagnostics, Indianapolis, IN, USA) using plasma as a substrate as a companion diagnostic for treatment with erlotinib in the first line setting and with osimertinib in the relapse setting. In acknowledgement of the limited sensitivity of plasma testing, the FDA approval advised that patients who are negative for these mutations in the plasma should undergo routine biopsy and repeat testing on the tumor tissue. Detection of EGFR mutations in the plasma predicts response to EGFR TKIs to the same degree as detecting mutations in tissue. Failure to clear the EGFR mutation in the blood after 8 weeks of combined platinum-based therapy plus erlotinib treatment is associated with an adverse prognosis (64). In the relapse setting, changes in the levels of T790M mutation in the plasma can be
detected upon treatment with third generation inhibitors and in most cases these changes mirror the clinical status according to traditional radiographic staging (60,65). Scenarios in which the plasma levels of T790M decline but the patient experiences radiographic progression have been reported (65). This may reflect heterogeneity of resistance mechanisms, such as combined T790M mutation acquisition and small cell transformation.

Given the lower sensitivity of plasma-based testing relative to tissue, plasma genotyping assays should be designed to maximize the positive predictive value of the assay. This approach will minimize the risk of false positive results that might reduce confidence in a result and confound a clinician’s ability to decide on therapy (66). When a plasma assay is negative, however, follow-up plasma testing or tissue biopsy should be pursued to obtain tumor genotyping information.

Single gene assays can be powerful when employed from plasma specimens, as directed clinical questions in a population with a high pre-test probability (i.e., patients with relapsed EGFR-mutated lung cancer) can be answered rapidly and with confidence following a minimally-invasive blood draw. However, the use of single-gene assays limits the number of indications for which the technology can be used; most available assays are not amenable to large-scale multiplexing that might be necessary to provide more comprehensive information for a broader range of patients. Efforts to develop NGS-based approaches from plasma have led to a variety of promising academic and commercial assays. In principle, these are assays designed to maximize depth, rather than breadth of sequencing, so as to optimize assay sensitivity and clinical actionability. Focused bait design, modifications to the library preparation chemistry, sequencing to thousands-fold depth of coverage, and molecular barcoding to detect and suppress PCR errors have all been proposed as enhancements to NGS design for use in plasma specimens (62,67). Translocations and copy number alterations are also detectable in the plasma, given the appropriate design (68).

**Conclusions**

In routine practice, all patients with a diagnosis of advanced stage lung adenocarcinoma should undergo (I) EGFR mutational analysis that assesses, at a minimum, the L858 codon and the exon 19 deletion hotspot; (II) ALK rearrangement testing by immunohistochemistry, fluorescence in situ hybridization (FISH), RT-PCR or sequencing; (III) ROS1 rearrangement testing by FISH or molecular methods, with an option for prescreening by IHC; and (IV) IHC for PD-L1. When possible, use of a NGS assay that captures EGFR, ALK, and ROS1 as well as other less common but targetable (ERBB2, BRAF, MET, RET) and common but difficult to target (KRAS) oncogenes should be encouraged. Broader panels can detect a range of alterations in tumor suppressor genes such as TP53, STK11 and SMARCA4 that may have prognostic implications. Tumor mutation burden, whether derived from whole exome or large panel sequencing data (52), may predict response to immunotherapy (51). Use of plasma ctDNA in lieu of tumor tissue for molecular analysis allows a subset of patients to forego repeat biopsy due to insufficient tissue at diagnosis or at relapse. Plasma genotyping may also provide additional prognostic information as well as a mechanism to monitor solid tumor patients for response to therapy.

**Acknowledgements**

The authors would like to acknowledge the American Association for Cancer Research and it financial and material support in the development of the AACR Project GENIE registry, as well as members of the consortium for their commitment to data sharing. Interpretations are the responsibility of the author.

**Footnote**

Conflicts of Interest: The author has no conflicts of interest to declare.

**References**


Cite this article as: Sholl L. Molecular diagnostics of lung cancer in the clinic. Transl Lung Cancer Res 2017;6(5):560-569. doi: 10.21037/tlcr.2017.08.03