Gene rearrangement detection by next-generation sequencing in patients with non-small cell lung carcinoma

Aneta Brisudova, Jozef Skarda

Non-small cell lung carcinoma (NSCLC) is the leading cause of cancer-related deaths worldwide. Various molecular markers in NSCLC patients have been developed, including gene rearrangements, currently used in therapeutic strategies. With increasing number of these molecular biomarkers of NSCLC, there is a demand for highly efficient methods for detecting mutations and translocations in treatable targets. Those currently available U.S. Food and Drug Administration (FDA) approved approaches, for example immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH), are inadequate, due to sufficient quantity of material and long time duration. Next-generation massive parallel sequencing (NGS), with the ability to perform and capture data from millions of sequencing reactions simultaneously could resolve the problem. Thanks to gradual NGS introduction into clinical laboratories, screening time should be considerably shorter, which is very important for patients with advanced NSCLC. Moreover, only a minimum sample input is needed for achieving adequate results. NGS was compared to the current detection methods of ALK, ROS1, c-RET and c-MET rearrangements in NSCLC and a significant match, between IHC, FISH and NGS results, was found. Recent available researches have been carried out on a small numbers of patients. Verifying these results on larger patients cohort is important. This review summarizes the literature on this subject and compares current possibilities of predictive gene rearrangements detection in patients with NSCLC.

Key words: non-small cell lung carcinoma, gene rearrangement, next-generation sequencing, immunohistochemistry, fluorescence in situ hybridization

INTRODUCTION

Lung carcinoma is a heterogenous disease and a leading cause of death in both men and women worldwide. It can be divided into two groups. The first is a small cell lung carcinoma (SCLC) and the second is non-small cell lung carcinoma (NSCLC). Non-small cell lung carcinoma accounts for approximately 85-90% of all lung cancer. It is divided into adenocarcinomas (40-50%), squamous carcinomas (20-30%), large cell carcinomas and other uncommon carcinomas. Adenocarcinoma is the most common subtype diagnosed in smokers as well as non-smokers. SCLC patients are more sensitive to systemic chemotherapy and radiotherapy than NSCLC patients in the first months following diagnosis. However, there is later disease relapse. For this reason, the used therapy is targeted and based on tyrosine kinase inhibitors (TKI) or monoclonal antibody. However, despite the current treatment approaches, the overall survival rate over 5 years is only 15%. Early diagnosis together with the detection of treatable targets however significantly contribute to prolongation of patient life span. Lung cancer is often diagnosed in the advanced stages. At least 40% of these cases are diagnosed with advanced distant metastases. For this reason, it is crucial to implement new diagnostic methods.

Clinical biomarkers currently used in therapeutic strategies for NSCLC

With the expansion of modern genetic and molecular techniques, various diagnostic and therapeutic procedures have been progressively developed. Basic diagnostic molecular targets in NSCLC include the first discovered somatic mutation in the human epidermal grow factor receptor (EGFR) (ref.10,11), anaplastic lymphoma kinase (ALK), reactive oxygen species 1 (ROS1) proto-oncogene, tyrosine-protein kinase receptor Ret (c-RET) and mesenchymal epithelial transition factor/hepatocyte growth factor receptor (c-MET) belonging to tyrosine kinase receptors. Discovery of these markers has led to the expansion of targeted therapy using tyrosine kinase inhibitors.

The epidermal growth factor receptor gene is located on chromosome 7 (7p11.2) and encodes a type I transmembrane growth factor tyrosine kinase receptor. Tyrosine kinase activity may be dysregulated by oncogenic gene mutations. In NSCLC, gain of function or activating mutations of EGFR cause overexpression and constitutive kinase activity. These mutations are detected by polymerase chain reaction-amplified genomic DNA in the first step of NSCLC molecular characterization.

The anaplastic lymphoma kinase gene is located on chromosome 2 (2p23) and identified in the nervous system during embryogenesis but its expression decreases postnatally. ALK was originally identified in 1994 as an
oncogene in anaplastic large cell lymphomas (ALCL). The C-terminal cytoplasmatic domain of ALK fusion to the N-terminal of nucleophosmin (NPM) has been identified in 40-60% of ALCLs. The fusion generates a gene product p80(NPMALK), a protein that plays a key role in protein shuttling. ALK rearrangement was later associated with 5-6% of NSCLCs. It is more common in younger patients (median age 55 years), never or light smokers (70%) diagnosed with adenocarcinoma. More than 30 fusion partners have been identified, e.g. KIF5B (kinesin family member 5B), CUX1 (cut-like homeobox 1) (ref.20), STRN (striatin) (ref.21), DCTN1 (dynactin-1), SQSTM1 (sequestosome-1) (ref.22), etc. (ref.23). The most common partner is EML4 (Echinoderm microtubule-associated protein-like 4). EML4-ALK fusion, found in 2007, causes constitutive anaplastic lymphoma kinase activity leading to the oncogenic pathway in NSCLC. This ALK fusion protein is an appropriate therapeutic target of TKI (ref.24, 26). Either alone or in combination with systemic therapy, patients have very favorable responses. The FDA have approved tyrosine kinase ALK inhibitors, crizotinib, alec-tinib and ceritinib. Crizotinib, registered in 2010, was the first drug for ALK positive fusion treatment27-29.

The success of Crizotinib in ALK-positive patients initiated the effort to find new oncogenic fusions in NSCLC. This led to the identification of new oncogenic fusions involving ROS1, c-RET and c-MET. ROS1 is located on chromosome 6 (6q22), whose structure was identified in 2003. Rearrangement of the ROS1 gene was first discovered in 2007 (ref.30). Using diverse genotyping techniques, ROS1 mutations were identified in 1-2% of NSCLC (ref.31). Detection of ROS1 rearrangement is critical for optimal targeted therapy. Initially it was identified in a glioblastoma cell line32 where intrachromosomal deletion on chromosome 6 fused the 5’ region of FIG (Fused in glioblastoma) gene to the 3’ region of ROS1 (ref.33). Recently, more fusion partners have been found, e.g. EZR (Ezrin) (ref.34), SDC4 (Syndecan-4) (ref.35), CCDC6 (Coiled-Coil Domain Containing 6) (ref.36) and CD74 (Cluster of differentiation 74) which is the most common one in NSCLC (ref.37).

Another treatable target is c-RET (rearranged during transfection) proto-oncogene. Its fusion is found in about 1-2% of NSCLC (ref.38). Initially it was discovered in thyroid carcinoma39. Later, in 2012, it was identified in lung40. The c-RET gene is located on chromosome 10 (10q11.2) and encodes a single pass transmembrane protein with a typical intracellular tyrosine kinase domain. It consists of three domains and shares 37% amino acid sequence homology with the ALK kinase domain41. It plays an important role in organogenesis and development of the enteric nervous system. At least 7 fusion partners were identified in adenocarcinomas, including the best characterized KIF5B (Kinesin family member 5B) with CCDC6 and NCOA4 (Nuclear receptor coactivator 4) (ref.42,43). Wang44 examined RET positive, younger, never smoker patients with early lymph node metastases and poorly differentiated tumors. As well as other driving mutations, RET rearrangement appears to be specifically associated with NSCLC and may be a targetable oncogenic driver45.

The last mentioned biomarker is mesenchymal epithelial transition factor tyrosine kinase. It is located on chromosome 7 (7q31) and is important for embryonic development and organogenesis42-46. c-MET and its natural ligand hepatocyte growth factor (HGF) are involved in MET-expressing epithelial cells in an endocrine or paracrine fashion. c-MET dysregulation is usually based on gene amplification (overexpression) and MET exon 14 splice site mutation. The c-MET gene mutation is an important mechanism and is detectable in 5-22% of NSCLCs (ref.47). Constitutively activated MET, receptor tyrosine kinase, promotes tumor angiogenesis, cell invasion and metastasis propagation in NSCLC (ref.48). MET acts as intracellular transducer by recruiting and activating several effectors, including PI3K (phosphatidylinositol 3-kinase), RAS (rat sarcoma), Gab1 (GRB2 [Growth factor receptor-bound protein 2]-associated-binding protein 1) and STAT3 (Signal transducer and activator of transcription 3) (ref.49,53).

Next-generation sequencing

NGS, sometimes called massive parallel sequencing, is becoming an important part of diagnostic and therapeutic practice in NSCLC. This is due to the quality and quantity of the available tumor biopsy or cytology material, that is not always sufficient for performing currently approved immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) approaches. Due to the importance of molecular testing of lung cancer patients, pathologists have to remember to save material for subsequent analyses. Greater samples quantities are thus needed to analyze a required number of molecular markers than is usually available. In addition, the least invasive biopsy should be performed and only small samples with few cells are obtained. Biopsies are invasive examinations, technically difficult, very risky and painful for patients. NGS is an emerging technology with the potential to overcome these limitations. More recent studies indicate that only a minimum sample input, about 10ng DNA/RNA corresponding to approx. 1000-1500 human cells, is necessary for achieving adequate results44. Thanks to these possibilities, rare fusion partners of ALK have been found. With the introduction of NGS to molecular genetics and molecular diagnostics by allowing formalin-fixed paraffin embedded (FFPE) tissues to be screened, there are new efficiency and time duration options available. This technology is being introduced at new laboratories requiring accurate and timely results which guide patient therapy decisions20,22-25,39. NGS for SNVs, INDELs and CNVs detection is based on DNA sequencing, in contrast to RNA-seq developed mainly for detection of gene fusions. Apart from fusions, it is capable of detecting an aberrant splice variant of exon 14 skipped MET (ref.29). Both DNA and RNA sequencing approaches have indisputable advantages. While DNA sequencing in preferable for testing mutations from non-transcribed regions, such as in a promoter, RNA sequencing is more suitable, for example, for intron-intron break testing. This should change the transcription product due to aberrant fusion and/or enhance gene expression. Currently, there is more
Gene rearrangement detection in NSCLC patients

FISH assay with dual-labeled break apart probes for selecting patients for TKI therapy was used recently. Later, carefully validated IHC was found to be an appropriate method for ALK and ROS1 rearrangement. A number of studies have shown that ALK FISH and IHC results did not match, but patients with FISH negative and IHC positive results may also benefit from ALK-TKI therapy. Moreover, IHC is relatively inexpensive and is performed routinely in most diagnostic laboratories. However, a serious disadvantage of both FISH and IHC results is the time duration and large sample size requirement. This is surmountable using next-generation sequencing. NGS is technology with the ability to perform and capture data from millions of sequencing reactions simultaneously. All NGS platforms are able to catch the individual sequence of hundreds of millions of molecules at the genome or transcriptome level. One obstacle to initiating NGS in routine practice is the high initial costs as it method is currently more expensive than the methods above. With increasing numbers of patients tested by NGS, there is an assumption that costs will be sustainable. The cost of an NGS gene panel approach the cost of single gene testing by IHC/FISH owing to the testing time of each target separately and trained staff. Implementation of NGS would create inter laboratory collaboration, sharing of results, automation and improved quality of genetic predictive and diagnostic medicine. This reasoning underlies the rational for believing why the costs will drop in the near future for laboratories with NGS in clinical predictive practice.

Velizheva describe fusion detection assays (FISH, IHC, NGS) of EML4-ALK and EZR-ROS1 in NSCLC and demonstrate the reliability of NGS in comparison to previously FDA approved FISH and IHC. NGS results of ALK fusion had 100% match with IHC assay as expected. NGS+ FISH- patients received crizotinib treatment with durable progression-free survival providing evidence for the validity of the NGS results. Further studies confirmed the high agreement (over 98%) of NGS testing with IHC (Tab. 1) (ref.63). IHC is still the most used and remains the gold standard in ALK rearrangement detection. NGS appears to be a reliable technique for assessing ROS1 rearrangement, offering the above advantages. NGS has 100% match to FISH, which is routinely used in diagnostic laboratories.

Discovery of fusion gene KIF5B-RET from a large-scale sequencing led to the discovery of a novel therapeutic target for adenocarcinomas. Although IHC is an effective detection method for ALK and ROS1, for RET detection it has been disappointing on account of false positive/negative results. FISH is highly sensitive and represents the gold standard of RET rearrangement detection, but it is non specific with respect to RET fusion partners. Their main limitation is confounding of re-

### Table 1. Current comparison of NGS with IHC, FISH and RT-qPCR in detection of gene rearrangement in NSCLC.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Number of patients</th>
<th>Method</th>
<th>Comparison</th>
<th>Match (%)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALK</td>
<td>51</td>
<td>NGS</td>
<td>IHC</td>
<td>98</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>NGS</td>
<td>IHC</td>
<td>100</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>NGS</td>
<td>IHC</td>
<td>98</td>
<td>67</td>
</tr>
<tr>
<td>ROS1</td>
<td>52</td>
<td>NGS</td>
<td>IHC, FISH</td>
<td>100</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>NGS</td>
<td>FISH</td>
<td>100</td>
<td>66</td>
</tr>
<tr>
<td>c-RET</td>
<td>1</td>
<td>NGS</td>
<td>FISH</td>
<td>100</td>
<td>65</td>
</tr>
<tr>
<td>c-MET</td>
<td>51</td>
<td>NGS</td>
<td>RT-qPCR</td>
<td>97.4</td>
<td>74</td>
</tr>
</tbody>
</table>

**ALK** – Anaplastic lymphoma kinase, **c-MET** – Metenchymal epithelial transition factor, **c-RET** – Tyrosine-protein kinase receptor Ret FISH – Fluorescence in situ hybridization, IHC – Immunohistochemistry, NGS – Next-generation sequencing, **ROS1** – Reactive oxygen species 1, RT-qPCR – Quantitative polymerase chain reaction with reverse transcription
sults from FFPE samples with degraded RNA. Finally, NGS provides accurate and sensitive profiling of the RET gene when there is a lack of tumor material. It is possible to catch mutations of <0.1% mutant allele fraction with >99.9999% specificity. On the other hand, complexity, difficulties in interpretation and high cost are still problematic. The frequency of MET exon 14 skipping (METex14) is comparable to other gene rearrangements of interest. Recent research on METex14 detection compared RT-qPCR with Sanger sequencing and NGS for future diagnostic use. The Sanger sequencing assay was proven to be less specific mainly to false negative results, while a METex14 mutation was found by NGS (Tab 1). NGS shows high specificity and sensitivity, as well as RT-qPCR, which is currently routinely used in clinical practice. These recent studies are focused on improving NGS as diagnostic method in NSCLC screening. Analysis of gene rearrangement with very small sample input, only 10ng, was carried out. However, NGS has not been studied in sufficient number of patients and optimized for current conditions and needs of clinical laboratories, which is inevitable for introduction into clinical practice.

CONCLUSION

The importance of recognizing molecular subtypes is paramount for targeted therapy of ALK, ROSI, c-RET and c-MET by specific tyrosine kinase inhibitors. With increasing numbers of molecular biomarkers in NSCLC for targeted therapy determination, provision of sufficient amount of tumour tissue is often problematic. Moreover, there is a requirement to diagnose the patient with advanced lung carcinoma as soon as possible. Current diagnostic methods do not meet the characteristic of present time because the processing of results takes approximately 15 days. Therefore, there is a constant effort to increase diagnostic yield, to reduce the time duration, and to minimalize the sample DNA/RNA input. In addition, new fusion partners should be detected by NGS. It is known, that different ALK-fusion partners can affect diverse ALK-TKI efficiency. Patients with some variants have significantly longer PFS with TKI. Moreover, some ALK-fusion partners led to primary resistance to crizotinib. Identification of new ALK, ROSI, c-RET and c-MET fusion partners can define personalized treatment more precisely.

This work summarizes currently available detection assays in NSCLC and compares individual methods among themselves. The results show, that NGS represents the future possibility of rapid diagnostic approach. However, comparison of NGS with FISH, IHC and RT-qPCR has been carried out on a small numbers of patients (Tab 1). Therefore, it is important to verify these results on larger patients cohorts to demonstrate the sensitivity, specificity and reliability of next-generation sequencing for diagnostic purposes.

Search strategy and selection criteria

Our research strategy was aimed at evaluating studies on the detection methods of selected gene rearrangement in NSCLC patients to determine the optimal method for overcoming the existing failures. Scientific articles from 1985 to 2019 were searched using the PubMed and Google scholar. The search terms used were, “ALK NSCLC”, “ROS1 NSCLC”, “c-RET NSCLC”, “c-MET NSCLC”, “IHC, FISH, NGS in NSCLC” and “gene rearrangement detection in NSCLC”.

ABBREVIATIONS

ALCL, Anaplastic large cell lymphomas; ALK, Anaplastic lymphoma kinase; CCDC6, Coiled-coil domain-containing protein 6; CD74, Cluster of differentiation 74; c-MET, Mesenchymal epithelial transition factor; c-RET, Tyrosine-protein kinase receptor Ret; CUX1, Cut-like homeobox 1; DCTN1, Dynactin-1; EGFR, Epidermal grow factor receptor; EML4, Echinoderm microtubule-associated protein-like 4; EZR, Ezrin; FDA, U.S. Food and Drug Administration; FFPE, Formalin-fixed paraffin embedded; FIG, Fused in glioblastoma; FISH, Fluorescence in situ hybridization; Gab1, GRB2 [Growth factor receptor-bound protein 2]-associated-binding protein 1; HGF, Hepatocyte growth factor; ICGC, International cancer genome consortium; IHC, Immunohistochemistry; KIF5B, Kinesin family member 5B; NCOA4, Nuclear receptor coactivator 4; NPM, Nucleophosmin; NSCLC, Non-small cell lung carcinoma; NGS, Next-generation sequencing; PI3K, Phosphatidylinositol 3-kinase; RAS, Rat sarcoma; ROS1, Reactive oxygen species 1 proto-oncogene; SLC, Small cell lung carcinoma; RT-qPCR, Quantitative polymerase chain reaction with reverse transcription; SDC4, Syndecan-4; SQSTM1, Sequestome-1; STAT3, Signal transducer and activator of transcription 3; STRN, Striatin; TCGA, The Cancer Genome Atlas; TKI, Tyrosine kinase inhibitors.

Author contributions: AB: literature search and manuscript writing; JS: corrections and final approval. Conflict of interest statement: None declared.

REFERENCES

Tumor molecular profiling of NSCLC patients using next generation sequencing.

Kakolyris S, Pentheroudakis G


Sorber L, Zwaenpoel K, Deschoolmeester V, Van Schii PE, Van

Ross K, Pailler E, Faugeroux V, Taylor M, Oulhen M, Auger N, Planchard

Califano R, Morgillo F, De Mello RA, Mountzios G. Role of mesenchy

Organ SL, Tsao MS. An overview of the c-MET signaling pathway.

Fixman ED, Fournier TM, Kamikura DM, Naujokas MA, Park M.


