Update on biomarkers for the detection of lung cancer

Eloisa Jantus-Lewintre¹
Marta Usó¹
Elena Sanmartín¹
Carlos Camps¹⁻³
¹Molecular Oncology Laboratory, Fundación para la Investigación del Hospital General Universitario, Valencia, Spain; ²Department of Medical Oncology, Consorcio Hospital General Universitario, Valencia, Spain; ³Department of Medicine, Universitat de València, Valencia, Spain

Abstract: Patients at risk for lung cancer may have subclinical disease for years before presentation. The diagnosis of this disease is primarily based on symptoms, and detection often occurs after curative intervention is no longer possible. At present, no lung cancer early-detection biomarker is clinically available. This study reviews the most recent advances in early detection and molecular diagnostic biomarkers for the detection of lung cancer. This review includes an overview of the various biological specimens and matrices in which these biomarkers could be analyzed, as well as the diverse strategies and approaches for identifying new biomarkers that are currently being explored. Several novel and attractive biomarker candidates for the early detection of lung cancer exist. A remarkable shift is taking place from research based on single markers to analyzing signatures that are more complex in order to take advantage of new high-throughput technologies. However, it is still necessary to validate the most promising markers and the standardization of procedures that will lead to specific clinical applications.

Keywords: biomarker, detection, lung cancer, diagnosis

Introduction

Lung cancer is a global public health concern. The average five-year lung cancer survival rate in Europe is approximately 15%.¹ More people die from lung cancer than from breast, colon, and prostate cancer combined.² Nevertheless, the five-year survival rate in early-stage, operable, non-small cell lung cancer (NSCLC) is approximately 50%–70%. The five-year survival rate drops to 2%–5% for patients whose cancers are diagnosed after their tumors have spread distantly.³ At present, the diagnosis of lung cancer is primarily based on symptoms (eg, cough, chest pain, hemoptysis, shortness of breath), and detection often occurs when curative intervention (ie, surgery) is no longer possible.

Although early detection can save lives, screening tests for high-risk individuals using sputum cytology or chest X-rays have not resulted in improvements in disease-specific survival.⁴⁵ Patients at risk for lung cancer may have subclinical disease for years before presentation, and until now, no validated model that includes molecular biomarkers of cancer risk have been developed. The clinical challenge, therefore, is to develop a non-invasive approach based on molecular biomarkers that may significantly improve the assessment of risk, early diagnosis, and application of chemoprevention for lung cancer.

This review presents and discusses the most recent advances in early detection and molecular diagnostic biomarkers for the detection of lung cancer, including an overview of the different matrices with which these biomarkers could be analyzed.
Early detection of lung cancer: the role of CT screening

Low-dose chest computed tomography (CT) scans are proven to be effective tools for the detection of early-stage resectable diseases. Several Phase 2 non-randomized trials of CT screening of high-risk individuals (current or former smokers) have shown that CT screening detects small size lung cancers of 12–15 mm in diameter. When these results are compared with those of a validated control group, CT scans detected three times more lung cancer than would be expected.6,8

The results from the National Lung Screening Trial (NLST, the largest randomized trial) were published recently. This study enrolled 53,454 persons at high risk for lung cancer in order to determine the real impact of screening with low-dose CT screening on lung cancer mortality at the population level. The participants were randomly assigned to undergo three annual screenings with either low-dose CT or single-view chest X-ray. The results of the NLST show a 20.0% decrease in mortality from lung cancer in the low-dose CT compared with the chest X-ray. Additionally, a 6.7% overall mortality reduction was observed in the CT group.9

Conversely, low-dose CT screening was associated with a high rate of false positive results. Two other harmful effects of CT screening that should be considered are over-diagnosis and the association of low-dose CT with the development of radiation induced cancers (effects that could not be measured directly and must be assessed in future analyses). Given these results, the potential use of CT as a lung-cancer screening tool is still a matter of debate.

Due to the contentious results of the image-based screening programs, there is increasing interest in the development of risk prediction models that incorporate genetic and molecular biomarkers for the precise and accurate estimation of risks and are focused on the use of noninvasive or minimally invasive samples, such as blood, sputum, or exhaled breath.

Molecular biomarkers

A biomarker is defined as a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic/pharmacodynamic responses to a therapeutic intervention.10

An individual marker may serve more than one purpose and, therefore, can fall into more than one category. In addition, a single biomarker may have different categorizations across tumor types and/or stages of the disease.11 The ideal cancer biomarker must be useful for all purposes (diagnosis, prognosis, or the prediction of pharmacological response) and should also be easy to obtain, inexpensive, easy to perform, accurate, and highly informative.

New technologies, including pharmacogenomics, proteomics, metabolomics, and molecular imaging have been introduced in clinical settings and may potentially offer early diagnosis, accurate characterization of the histological subtype of NSCLC, and insight into disease prognosis and response to therapy. However, these techniques are in early development and considerable efforts to standardize them are necessary. Recently, Hanansh et al reviewed the emerging lung cancer serum biomarkers including proteins, DNA methylation patterns, autoantibodies, microRNAs, circulating tumor cells (CTCs), cell-free DNA, and several metabolites.12 Table 1 summarizes some relevant studies in the field of biomarkers for lung cancer detection and shows the wide range of specimens, type of biomarkers and detection methods that have been used. An overview of the most recent publications in the area of early diagnosis biomarkers shows that the focus in cancer biomarkers has shifted from searching for single biomarkers to analyzing combined biomarkers, which seems to be more effective in decreasing false positives and increasing detection rates. However, at this time, there are no single or combined validated molecular biomarkers for the early detection of lung cancer despite the considerable number of investigations that have been carried out with this purpose (Table 1).

Selection of the analytical matrix: the biological specimen

An important pre-analytic issue is choosing the “correct” analytical matrix in which to find new biomarkers. A good biomarker must be detectable in clinically accessible material; therefore, there is an increasing interest in the development of molecular-based minimally- or non-invasive screening methods. Fluid biomarkers include a variety of components found in blood, urine, or other fluids that may reflect the presence of a tumor in the body.13

The aim of early lung cancer diagnosis is to detect molecular biomarkers in non-invasive biological specimens before symptoms arise.

Tissue bronchoscopy

Bronchoscopy, particularly with the help of techniques such as auto-fluorescence bronchoscopy (AFB) and narrow-band imaging, is a suitable technique for detecting preinvasive lesions and early-stage lung cancer in the central airways that cannot be detected by CT technology.14 In addition,
several studies have demonstrated that the use of fluorescence in situ hybridization (FISH) testing for the detection of chromosomal abnormalities is significantly more sensitive than conventional cytology for detecting lung cancer in bronchoscopically obtained bronchial brushing specimens; however, this result was achieved with an increase in false positive diagnoses.\(^\text{15–18}\) Other studies show similar results that demonstrate how equivocal cytology in lung cancer diagnosis can be overcome with additional methods on the same specimen (ie, slide). Adjuvant FISH or DNA-image cytometry in cytologically equivocal diagnoses improves diagnostic accuracy at comparable rates.\(^\text{19}\) Spira et al have identified an airway gene-expression biomarker that may be useful as a screening tool for the evaluation of smokers with suspected lung cancer. They performed gene-expression profiling on histologically normal upper airway epithelial cells obtained at the time of bronchoscopy, with sufficient quantities of high quality RNA for microarray studies from 152 of 208 samples. Their data suggest that combining cytopathology with the gene-expression biomarker improves the diagnostic sensitivity of the overall bronchoscopy procedure from 53% to 95%.\(^\text{20}\)

AFB biopsies were also useful for analyzing microRNA expression by qPCR, allowing to assess different microRNA patterns during the successive steps of bronchial carcinogenesis.\(^\text{21}\)

**Blood samples**

The rich content of diverse molecular and cellular elements in blood, which gives information about the health status of an individual, makes it an ideal compartment in which to develop non-invasive diagnostics for cancer. Hanash et al recently reviewed the role of blood-borne biomarkers in cancer screening; this role could be extended from general population risk assessment to treatment response evaluation and recurrence monitoring.\(^\text{12}\)

Blood samples comprise a complex matrix composed of a cellular fraction and serum or plasma. CTCs and macromolecules such as lipids, proteins, RNA, microRNA, and DNA that originate from tumor cells can be found in blood. In addition to CTCs, other circulating cells are currently being explored as sources of biomarkers such as immune, stromal, and endothelial cells.\(^\text{13}\)

The first detection of cell-free nucleic acids originating from tumor cells and circulating in the serum or plasma of cancer patients, including lung cancer, was reported in 1997.\(^\text{22}\) Several case-control studies have reported higher levels of circulating DNA in patients’ blood than in normal healthy controls, showing that DNA may be an important risk factor for the presence of illness and a prognostic index in the follow-up procedure (reviewed by Bremnes et al).\(^\text{23}\) Interestingly, Sozzi et al were able to discriminate NSCLC patients from age-, sex-, and smoking-matched controls according to the amount of circulating plasma DNA, which was determined through the use of real-time quantitative polymerase chain reaction (PCR) amplification of the human telomerase reverse transcriptase gene (hTERT).\(^\text{24}\) Since then, several studies have assessed varying concentrations of free circulating DNA using diverse data analysis methods or laboratory procedures that have resulted in significant differences between lung cancer patients and controls.\(^\text{24–27}\)

In contrast to these studies, Tamkovich et al found no significant differences between the amount of DNA in plasma samples of lung cancer patients and healthy subjects.\(^\text{28}\)

In general, the vast majority of studies show that higher levels of circulating DNA can be detected in patients with lung cancer compared with disease-free smokers and control subjects, which suggests that the analysis of this biomarker may represent a new non-invasive approach for the early detection of lung cancer.

In terms of circulating nucleic acids, it is also worth considering the value of the isolation and characterization of circulating tumor DNA because a correlation between the detection of molecular alterations (genetic and epigenetic) in the circulating nucleic acids and lung tumor DNA has been demonstrated.\(^\text{28–31}\)

Circulating cell-free RNA and miRNA as well as exosomal miRNA content has also been identified as a source of biomarkers for early detection in lung cancer patients.\(^\text{32–34}\) The usefulness of circulating nucleic acid detection in plasma or serum for the early detection of lung cancer has been extensively assessed, but the results are still inconsistent due to the high rate of false negative results (the absence of circulating nucleic acids does not correlate with the absence of cancer).\(^\text{25}\)

Recent microarray studies have identified additional early detection signatures using samples from peripheral blood. Shoew et al demonstrated the feasibility of using the PBMC gene expression signature to identify early-stage NSCLC in at-risk populations.\(^\text{36}\)

The use of blood samples for tumor protein quantification is another point to be addressed in the development of biomarkers for the early detection of lung cancer. Some proteomics biomarkers of lung cancer, such as carcinoembryonic antigen (CEA), cytokeratin-19 fragment (CYFRA), or neuron-specific enolase (NSE) are clinically available, while others (serum amyloid A–SAA–, haptoglobin-α2, or Apo A-1)\(^\text{37}\) still need
Exhaled breath condensate

Exhaled breath condensate (EBC) is a fluid from the airways that can be collected by means of an easy, completely non-invasive, and repeatable procedure that is well accepted by patients. EBC contains both volatile and non-volatile compounds, including cytokines, proteins, and DNA. The analysis of exhaled air for volatile organic compounds (VOCs) shows that there is a pattern of VOCs that distinguishes lung cancer patients. Even when most VOCs in the cancer patients are higher at baseline compared to the controls, this difference is still evident after surgery. Therefore, exhaled VOCs cannot be considered specific biomarkers for lung cancer, as they do not seem to be solely produced by tumor tissue, but rather by an epiphenomenon that accompanies lung cancer development.

Various studies have reported genetic changes in DNA found in EBC from lung cancer patients. Carpagnano et al demonstrated that the microsatellite alterations in DNA from the EBC and tumor tissue in NSCLC patients had an overlapping profile consisting of loss of heterozygosity and microsatellite instability. In another study by Gessner et al, mutations were present in the EBC of patients with NSCLC, while no mutations were found in the EBC of the control subjects. All these findings support the suggestion that EBC may be an appropriate matrix for the non-invasive identification of early markers of NCSLC. On the other hand, a weakness of the use of EBC is that the collection devices can significantly affect the levels of biomarkers obtained; therefore, the values acquired from different devices may not be directly comparable.

Sputum

Sputum is another potentially attractive non-invasive source of biomarkers for lung cancer. Although it is unlikely that early-stage lung cancers will exfoliate detectable cancer cells in sputum, it is hypothesized that because of field tumorigenesis process, genetic or epigenetic changes in exfoliated cells may reflect an increased risk of lung cancer. Consequently, many research groups have examined sputum as a matrix in which biomarkers can be determined. Byers et al studied cytologic atypia in a large cohort of smokers and former smokers with chronic obstructive lung diseases. They reported that sputum cytologic atypia is a marker for increased lung cancer risk and examined whether it might complement radiologic imaging in a combined approach to the early detection of lung cancer. With the intention of improving the diagnostic accuracy of cytologic examination, additional molecular approaches have been investigated, such as gene-promoter methylation analysis and FISH for chromosomal aneusomy using a four-target DNA FISH probe including EGFR, MYC, 5p15, and CEP6.

Using sputum as starting material, Xie et al reported aberrant miRNA expression in this type of sample, which could be used as a marker for the early diagnosis of NSCLC, especially in adenocarcinomas. However, there are currently no definitive data indicating that sputum biomarkers are efficient for routine clinical applications to reliably predict individuals at risk of developing lung cancer or who are in the early stages of the disease.

Strategies for the identification of early-detection lung cancer biomarkers

Genomic/genetic approaches

Since human genome sequencing was completed, DNA has been widely studied for the detection of new biomarkers for many diseases. In lung cancer, genomic instability is frequently observed, which includes allelic loss and instability at loci on 3p (fragile histidine triad), 9p (p16INK4A), and 17p. By analyzing a wide collection of lung adenocarcinomas through high resolution (250k) SNP arrays, Weir et al...
Lung Cancer: Targets and Therapy 2012:3

identified 57 significantly recurrent events in this type of tumor, but concluded that many of the genes involved in lung adenocarcinoma are yet to be discovered. Twenty-six of the most prevalent somatic mutations in lung adenocarcinomas have recently been categorized. The most frequently mutated genes include tyrosine kinases, among them the EGFR homologue ERBB4; multiple ephrin receptor genes, notably EPHA3; vascular endothelial growth factor receptor, KDR; and NTRK genes. The detection of these changes in DNA is considered a potential diagnostic tool, but due to the low frequency of alterations in the specific genes found in biological samples, most of these studies obtained limited sensitivity and specificity.

Over last 5 years, next generation sequencing (NGS) methods have changed genomic analyses, thus providing a more accurate and sensitive detection technology for somatic mutations in tumors. NGS chemistries vary, but the results are similar: millions of sequencing reads of DNA templates are generated in parallel, gathering information of up to 200 Gb, while the read lengths for each DNA template are very short (35–500 bp) compared to the traditional Sanger sequencing method. This technique can be applied to a

---

### Table 1 Summary of relevant studies in the field of biomarkers for lung cancer detection

<table>
<thead>
<tr>
<th>Reference</th>
<th>Sample type</th>
<th>Biomarker type</th>
<th>Detection method</th>
<th>Number of cases</th>
<th>Sensitivity/specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halling et al</td>
<td>Bronquial brushing</td>
<td>Cytology and chromosomal abnormalities</td>
<td>Combined cytology and FISH</td>
<td>Patients = 89</td>
<td>75/83 Controls = 48*</td>
</tr>
<tr>
<td>Spira et al</td>
<td>Bronquial brushing</td>
<td>Gene expression</td>
<td>Microarrays</td>
<td>Patients = 60</td>
<td>80/84 Controls = 69* Validation dataset Patients = 18</td>
</tr>
<tr>
<td>Sozzi et al</td>
<td>Blood</td>
<td>DNA</td>
<td>PCR</td>
<td>Patients = 100</td>
<td>90/86* Controls = 100</td>
</tr>
<tr>
<td>Tanaka et al</td>
<td>Blood</td>
<td>CTC</td>
<td>CTC count</td>
<td>Patients = 125</td>
<td>30.4/88.0 Controls = 25* Validation dataset Patients = 18</td>
</tr>
<tr>
<td>Patz et al</td>
<td>Serum</td>
<td>Proteins</td>
<td>ELISA</td>
<td>Patients = 50</td>
<td>89.3/84.7 Controls = 50</td>
</tr>
<tr>
<td>Miura et al</td>
<td>Serum</td>
<td>Gene expression</td>
<td>RT-PCR</td>
<td>Patients = 112</td>
<td>89/72* Controls = 80</td>
</tr>
<tr>
<td>Zheng et al</td>
<td>Plasma</td>
<td>microRNAs</td>
<td>RT-PCR</td>
<td>Patients = 74</td>
<td>81.33/86.76 Controls = 68</td>
</tr>
<tr>
<td>Jantus-Lewintre et al</td>
<td>Plasma</td>
<td>Proteins</td>
<td>ELISA</td>
<td>Patients = 432</td>
<td>74.5/85.7 Controls = 89</td>
</tr>
<tr>
<td>Showe et al</td>
<td>PBMC</td>
<td>Gene expression</td>
<td>Arrays</td>
<td>Patients = 137</td>
<td>91/80 Controls = 91* Validation dataset Patients = 38 Controls = 17*</td>
</tr>
<tr>
<td>Yu et al</td>
<td>Sputum</td>
<td>microRNAs</td>
<td>RT-PCR</td>
<td>Patients = 36</td>
<td>80.6/91.7 Controls = 36 * Validation dataset Patients = 64 Controls = 58</td>
</tr>
<tr>
<td>Xie et al</td>
<td>Sputum</td>
<td>microRNAs</td>
<td>RT-PCR</td>
<td>Patients = 23</td>
<td>69.7/100 Controls = 17*</td>
</tr>
</tbody>
</table>

**Notes:**  
\*Control group consisted of patients without lung cancer; \*Highest value obtained with a DNA concentration value of 9 ng/mL; \*Cut-off CTC count ≥ 1; \*Obtained with hTERT; \*Obtained with EGFR.

**Abbreviations:** CTC, circulating tumor cell; EGFR, epidermal growth factor receptor; ELISA, enzyme-linked immunosorbent assay; FISH, fluorescence in situ hybridization; hTERT, human telomerase reverse transcriptase gene; PBMC, peripheral blood mononuclear cells; PCR, polymerase chain reaction; RT-PCR, real-time polymerase chain reaction.
Elevated levels of methylated genes in plasma have been observed for a distinction between patients with adenocarcinoma and normal subjects. Moreover, these miRNA could be detected in sputum samples by qPCR with high sensitivity and specificity.

The presence of miRNAs in plasma and serum samples has been widely studied. Endogenous plasma miRNAs are highly stable. They are resistant to incubation at 37°C, freeze-thaw cycles, and even to RNase activity. The mechanisms by which miRNAs are protected are still unknown, although it has been hypothesized that they may be packaged inside exosomes or associated with other protective molecules. Several studies carried out on plasma and serum have identified patterns of differential expression of miRNAs that allow lung cancer patients to be distinguished from cancer-free control subjects, even in patients with Stage I lung cancer. In sum, these results support the hypothesis that
miRNA expression patterns may serve as biomarkers for the early diagnosis of lung cancer.

**Proteomics**

Proteins are the real players in biomarking processes; therefore, they may have a significant advantage over genes in terms of biomarker discovery. Proteomics is becoming a useful tool for biomarker identification in different types of tumors due to the development of high-throughput platforms, which allow for the simultaneous detection of multiple proteins. Two approaches – the identification of individual proteins and protein expression profiling – have been suggested for the application of proteomics in biomarker detection. The role of some plasma proteins as diagnostic biomarkers has been further studied, and some are currently available in the clinical setting (e.g., CEA, CYFRA, and NSE). The problem lies in the low specificity and/or sensitivity of these biomarkers, which means they are not recommended for clinical use.\(^\text{76–78}\) Many additional proteins have recently been studied as potential diagnostic biomarkers in lung cancer. Patz et al\(^\text{79}\) tested a panel of four serum proteins (CEA, retinol-binding protein, \(\alpha\)-1-antitrypsin, and squamous cell carcinoma antigen) in a matched case-control study including lung cancer patients, in which considerable specificity and sensitivity was obtained.

The value of other serum proteins, such as amyloid A, haptoglobin-\(\alpha\)2, and a fragment of apolipoprotein A-1, as biomarkers has also been explored, although more clinical validation is needed.\(^\text{80–82}\) Due to the importance of the angiogenic process in cancer, our group recently studied the diagnostic and prognostic implications of vascular endothelial growth factor (VEGF) and its receptor VEGFR-2 in patients with advanced NSCLC, where we observed that both angiogenic markers exhibited higher plasma concentrations in patients compared to the age- and sex-matched controls, in which considerable specificity and sensitivity was obtained.

Comparative protein profiling is another promising method for the detection of protein expression patterns. Advances in new analytical methods such as two-dimensional gel electrophoresis (2DE) and mass spectrometry have made the simultaneous detection of multiple protein changes possible.\(^\text{82,83}\) Tyan et al characterized a list of 1400 proteins in pleural effusion samples from lung adenocarcinoma patients by separating peptides and HPLC-ESI-MS/MS.\(^\text{86}\) Furthermore, Soltermann et al isolated 170 glycoproteins that were also from pleural effusions, including several known lung cancer biomarkers such as CA125, CD44, and TTF.\(^\text{87}\) These studies have some limitations due to their lack of independent validation. Moreover, they are not superior to other lung cancer-diagnosis methods, despite the recent discovery of new biomarkers.

**Conclusion**

Cancer biomarker research has recently experienced a remarkable growth in technology, which has resulted in biomarker discoveries from various “omics” including genomics, transcriptomics, proteomics, and metabolomics. Several examples of novel candidate biomarkers for the early detection of lung cancer have emerged following this growth in technology. In conjunction with the use of non-invasively obtained samples, such as blood, sputum, or EBC, this technological growth represents a significant advance in the field of diagnostic biomarkers for lung cancer. The tasks ahead are to select the most promising markers (single or a panel) for the early detection of lung cancer in order to choose the correct specimen and novel methodology and to optimize and standardize procedures for the implementation of successful validation studies that will lead to specific clinical applications.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


