New biomarkers for lung cancer

Massimiliano Paci†, Cristian Rapicetta & Sally Maramotti
Division of Thoracic Surgery, Azienda Santa Maria Nuova di Reggio Emilia,
Viale Risorgimento 80, 42100 Reggio Emilia, Italy

**Importance of the field:** Despite many efforts to improve early detection, lung cancer remains the leading cause of cancer deaths. Stage is the main determinant of prognosis and the basis for deciding treatment options. Screening tests for lung cancer have not been successful so far.

**Areas covered in the review:** The article reviews the available literature related to biomarkers in use at present and those that could be used for early diagnosis, staging, prognosis, response to therapy and prediction of recurrence. The single biomarkers are analysed, divided according to the technological methods used and the locations of sampling.

**What the reader will gain:** The reader will gain knowledge on biomarkers in use and those now under study. The reader will also gain insights into the difficulties pertaining to the development of biomarkers, results reproducibility and clinical application.

**Take home message:** Although some markers seem to be promising, at present there is no consensus on the proven value of their clinical use in lung cancer. The future lies probably in a panel of biomarkers instead of individual assays, or in predictive models derived from the integration of clinical variables and gene expression profiles.

**Keywords:** early detection, genomics, lung cancer, metabolomics, molecular biomarkers, prognosis, proteomics, screening

1. Introduction

Primary lung cancer is considered one of the major epidemiological issues worldwide; it is the most frequent malignancy in the male population, and in the last few years it has been rising among women and remains the leading cause of cancer-related deaths in both genders. Despite many efforts to improve lung cancer outcome, long-term survival has not improved significantly over the last 20 years, with a 5-year cumulative survival rate that remains at only 15% [1]. One of the main reasons is that most lung cancers are diagnosed at an advanced stage, owing to low clinical signs and symptoms of the disease in early stages [2]. Only 15% of the patients show a form located in stages I and II. At the time of diagnosis, 30% of patients have a regional lymph node concern and >55% have a distance metastatic diffusion [3].

When lung cancer is detected in stage Ia, however, survival rate can reach a remarkable 80% [4]; the relevance of time bias owing to anticipation of diagnosis is still controversial, or, more precisely, it is discussed if the natural history of the disease is positively affected by the current therapies. However, it is clear that surgery with curative intent in early stage lung cancer ensures the best chance of long-term survival.

This means that the development of a screening tool for early detection is needed. Therefore, screening for lung cancer has been addressed in several studies focused on cost/effectiveness and feasibility of various techniques. An effective screening test should detect disease at an early stage, when the patient is asymptomatic and cure may be possible. The benefits of a screening campaign...
New biomarkers for lung cancer

2. Definition of biomarker

A biomarker could be considered every tool for the assessment of biological homeostasis and distinction of anomaly through qualitative or quantitative measurements. Although the last two characteristics are applicable to imaging techniques as well, the authors meant to focus attention on molecular rather than 'physical' tumour-induced alterations, meaning that data such as dimensions, density and doubling-growth time of an indeterminate pulmonary nodule were not considered as 'biomarkers' for lung cancer.

The ideal biomarker should be produced by malignant cells or only in response to them, it should not be present in healthy tissue or in benign disease, it should be detectable when the tumour is in a subclinical phase, readily detectable in accessible biological material, and when modulated it should directly correlate with the bulk of the tumour, prognosis, disease response to therapy or recurrence. It should be sensitive, specific, simple and cost-effective [7].

Most of the studies report sensitivity and specificity for each marker: these parameters are inversely related and their values diverge according to cutoff value adopted, as well as negative and positive predictive value (NPV and PPV). The accuracy of a biomarker is then better represented by a receiver operating characteristics (ROC) curve, which links sensitivity and specificity, and a calculation of area under curve (AUC), which should be as near unity (or 100%) as possible. The prognostic significance of a marker is estimated by log-rank univariate test or Cox-proportional hazard model, using overall survival or disease-free survival as the main end points.

Biomarkers have potentially a wide range of application: they could be used for risk stratification in early diagnosis, staging, prognosis, proper treatment selection (i.e., 'targeted therapies') and follow-up for response to treatment and early recurrence detection (Figure 1) [8].

3. Methods

PubMed was searched with the MeSh terms ‘lung cancer’ and ‘biomarkers’, with the search limited to ‘humans’ and ‘English language’. To limit the scope of this review, only markers that have appeared in reviews are discussed in detail.

4. Current available biomarkers in clinical practice

Although their use is not actually recommended or encouraged in lung cancer screening, diagnosis and follow-up, the following biomarkers are frequently used in clinical practice.

Carcinoembryonic antigen (CEA) is a glycoprotein involved in cellular adhesion: its production is normally limited to the fetal period, although it can be found in heavy smokers. Increased production by cancer cells may be related to de-repression of CEA-encoding genes that belong to a...
superfamily of immunoglobulins and is formed by 29 genes whose 18 are normally expressed in adults. CEA is one of the most used ‘historical’ cancer biomarkers but it has low sensitivity and specificity (sometimes elevated in non-neoplastic diseases such as pancreatitis, cirrhosis and ulcerative colitis); moreover, it is not lung specific because it is elevated in colorectal, gastric and breast cancer. Higher levels of CEA are related to worse prognosis [9]. Moreover, preoperative CEA levels seem to be good predictors of the pathological stage in clinical stage I [10] even though other studies report no difference in terms of 5-year survival rate in early stage non-small-cell lung cancer (NSCLC) [11]. Clinical significance of preoperative CEA and higher preoperative values, even if within normal range (0 – 5 ng/ml), are related to worse prognosis [12]. A limited use for monitoring of recurrences after treatment is the only recommended use.

Cytokeratin fragments (CYFRA) 21.1 is a valid measure of presence of cytokeratin 19 fragments in serum, which is expressed by epithelial cancer cells and tracheobronchial cells: although it could be elevated in some benign lung diseases. The sensitivity ranges from 23 to 70% [13]; it is quite specific for NSCLC, especially squamous cell carcinoma (SCC), in which it reaches a remarkable sensitivity of 84.6% and shows correlation with T and N stage [14,15]. CYFRA 21.1 can be evaluated also in induced sputum of NSCLC patients, where it is 7 times greater than in chronic obstructive pulmonary disease (COPD) patients, although sensitivity and specificity (86 and 75%, respectively) are not satisfactory enough to suggest routine use in early diagnosis [16], so clinical value is limited again to prognostication and recurrence monitoring.

The tissue polypeptide-specific antigen (TPA) is another marker of cytokeratines (cytokeratins 8, 18 and 19) tested as a predictor of tumour response to chemotherapy but it is less lung-specific than CYFRA 21.1 and is no longer used in lung cancer [17].

Squamous cell carcinoma antigen (SCC) is a structural cytoplasmic cell protein whose circulating form is particularly elevated in squamous cell carcinomas, not specifically of the lung. Overall sensitivity ranges from 15 to 55%. Its values may correlate with metastatic potential [13].

Neuron-specific enolase (NSE) is a glycolytic enzyme present in neurons of the central and peripheral nervous system and in malignant cells of neuroectodermic origin, such as those of medullary-thyroid cancer and small-cell lung cancer (SCLC), in which it reaches a sensitivity of 74% [17] and has a prognostic value for survival [18].

Progastrin-releasing peptide (ProGRP) is a precursor of gastrin-release peptide, produced by gastrointestinal and tracheobronchial neuroendocrine cells, so it is relatively specific for SCLC, although it is increased in renal failure and in non-malignant lung lesions. Sensitivity is 47 – 86% for SCLC [13,20].

Tumour M2-pyruvate is an isofrom of the glycosile enzyme pyruvate kinase, existing as an active dimer and less active tetramer. The dimeric isoform is typical of tumour cells and its level can be measured in blood. Choosing a cutoff value at 95% of specificity, sensitivity was 78% of the patients with SCLC and 81% of patients with NSCLC [21]. Levels correlate well with tumour progression and remission, making it a valuable tool in disease monitoring, independently from histological subset [22].

C-reactive protein (CRP) is a serum peptide increased in many acute and chronic inflammatory diseases, included malignancies where it is probably related to immune response to cancer cells. It has been hypothesised that elevated and low values of CRP may reflect chronic inflammation and impaired immune response and both of these conditions may be associated to increased risk of cancer [23].

Carbohydrate antigen 125 is a glycoprotein produced by mesothelial cells in the fetal period and in some malignant
New biomarkers for lung cancer

diseases (mainly ovarian carcinoma but also adenocarcinoma and large-cell carcinoma of the lung), with sensitivity of 39 – 53.6% [24], but sometimes it can be elevated in benign inflammatory diseases involving pleura and peritoneum with similar serum levels as in malignant disease.

All these markers have proved to have a lack of sensitivity and therefore are of little clinical value for screening if used alone. Overall specificity is satisfactory for most of the biomarkers, ranging from 71 to 99%: false positive cases are usually related to the presence of chronic kidney or liver disease rather than etiology of lung disease. Several authors have tried to study the value of various combinations of biomarkers to improve overall sensitivity and specificity, especially in screening and early diagnosis: because their level usually correlates with extension of disease, for early stage lung cancers lower cutoff values should be chosen to maintain satisfactory sensitivity, leading to an overlapping with healthy subjects and consequently loss of specificity. Molina and colleagues, analysing various combinations of serum tumour markers including CEA, SCC, CA 125, CYFRA 21.1, NSE and ProGRP, found that profile expression of biomarkers could help in determining histology when biopsy is not feasible [25]. NSE and SCC are the key tumour markers in histological differentiation of SCLC and NSCLC; also, CEA, CA 125 and CA 15.3 showed higher concentration in NSCLC but their relationship with histology remains unclear [26].

Bekci et al. found that three tumour markers (CEA, CA 125 and CA 15.3) were significantly higher in solitary pulmonary nodules, and when considered together it was possible to predict malignancy in 66% (PPV) and exclude it if all the three markers were below the cutoff values (NPV: 100%) [27].

Finally, at high elevated concentrations of cytokeratin 19 fragment, tissue polypeptide antigen and squamous cell carcinoma antigen in squamous cell lung cancer, carcino-embryonic antigen and cancer antigen 125 in adenocarcinoma, as well as progastrin-releasing peptide and neuron-specific enolase in small cell lung cancer outlines the best suggestive panel of markers available (Table 1).

5. Genomic approach

The growth of lung tumours is a result of multiple and sequential accumulation of genetic anomalies because of exposure to environmental mutagen factors or due to genetic anomalies that are inherited. Each alteration can be classified in the following essential steps: i) the acquisition of autonomous or self-sufficient growth signals; ii) insensitivity to growth inhibitory signals; and iii) inhibition of apoptosis [28]. Several redundant sub-pathways concur to each of these steps, making the process very complex: nevertheless, some particular aberrations are more likely to predict the biological behaviour of a cancer regarding risk of metastasis and response to therapy.

Therefore, analysis of genetic and epigenetic alterations in tumour cells as gene promoter methylation, gene expression regulation, DNA repair mechanisms alteration and genomic instability represent areas of active research and innovation that can potentially produce lots of new biomarkers that must be evaluated in clinical settings [29].

5.1 Genomic technologies

The genomic approach to detect the tumour cell alterations consists of various techniques, among which fluorescence in situ hybridisation (FISH) [30], gene expression microarrays and polymerase chain reaction in real time (real-time PCR) [31] play the most important role.

5.1.1 FISH, microarrays and real-time PCR

Large alterations are routinely analysed by DNA techniques in cytogenetics (FISH). This technique uses fluorescent labelled DNA probes to identify or confirm gene abnormalities or chromosome abnormalities that usually cannot be detected with other routine methods. In a recent study, Yendamuri et al. detected chromosomal deletions at 3p22.1 and 10q22.3 by FISH and examined their distribution in different areas of the airway of patients with NSCLC. The FISH analysis of bronchoscopic brushes could identify patients with a high risk of developing lung cancer and also could be useful for prognostic evaluation [30].

Microarrays together with clustering analysis are largely used to study and make comparisons among genome-wide expression patterns in biological systems. A new taxonomy with prognostic and therapeutic relevance can be envisaged applying microarrays to the evaluation of the molecular diversity among cancers. This technique, when applied in studies of lung tumours, provides important information regarding prognosis and survival and helps the identification of potential therapeutic targets by offering key insights into processes such as lung tumorigenesis and metastasis [32].

Together with the development of microarray technologies, advances in real-time quantitative PCR have taken place and nowadays slab gels, radioactivity and sample manipulation are no longer needed. This technique will play an increasingly important role in clinical testing because it can provide information about gene loss, amplification and expression together with the detection of small alterations (i.e., point mutations). The applications of these techniques in the search for new biomarkers for lung cancer are described in the following sections.

5.1.2 Gene expression profiling

The use of gene expression for molecular staging may enhance the sensitivity of clinical and pathologic methods for staging tumours, improving treatment decisions and outcomes for lung cancer patients. This signature could also be useful in stratifying patients according to risk in trials of adjuvant treatment of the disease.

The expression levels of thousands of genes are quantified by microarray or quantitative polymerase chain reaction...
Table 1. Tumour markers criteria for histological classification.

<table>
<thead>
<tr>
<th>Criteria for NSCLC</th>
<th>Correct diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCC &gt;2 ng/ml and NSE &lt;45 ng/ml</td>
<td>98.6%</td>
</tr>
<tr>
<td>Stages I – III</td>
<td></td>
</tr>
<tr>
<td>CEA &gt;5 ng/ml and NSE &lt;35 ng/ml</td>
<td>97.2%</td>
</tr>
<tr>
<td>CYFRA &gt;3.3 ng/ml or CA 15.3 &gt;35 ng/ml and NSE &lt;35 ng/ml</td>
<td>98.1%</td>
</tr>
<tr>
<td>Stage IV</td>
<td></td>
</tr>
<tr>
<td>CEA &gt;8 ng/ml and NSE &lt;45 ng/ml</td>
<td>97.8%</td>
</tr>
<tr>
<td>CYFRA &gt;4.9 ng/ml or CA 15.3 &gt;35 ng/ml and NSE &lt;45 ng/ml</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criteria for SCLC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stages I – III</td>
<td></td>
</tr>
<tr>
<td>NSE &gt;35 ng/ml and SCC &lt;2 ng/ml</td>
<td>100%</td>
</tr>
<tr>
<td>Stage IV</td>
<td></td>
</tr>
<tr>
<td>NSE &gt;45 ng/ml and SCC &lt;2 ng/ml</td>
<td>97.7%</td>
</tr>
</tbody>
</table>

Reproduced with permission from [24].

techniques and the data obtained are then processed, normalised and possibly filtered. The expression data are then combined and grouped by clustering, risk score generation, or other means, to generate a gene signature that correlates with a clinical outcome, usually survival. Statistical analyses are performed using the metagene construction and binary prediction tree analysis and the signature is finally validated in data sets of independent cohorts [33].

Initial applications of high-throughput gene expression technology in lung cancer were to explore whether or not differences in gene expression could be identified between the different histological subtypes of lung tumours. These studies demonstrated that gene expression patterns could distinguish between the histological subtypes of lung cancer and found that adenocarcinomas had the greatest heterogeneity [34,35]. High-throughput gene expression technologies have also been used in the diagnostic evaluation of smokers to diagnose the disease at an early stage [36], and to characterise tumour stage [35,37,38]. Moreover, correlations between molecular sets and prognosis have been found, particularly in adenocarcinomas [34,35,39,40].

Metagene prediction models of predicting disease recurrence have been also found [41]. The model had a higher accuracy than models containing clinical data alone (age, sex, tumour diameter, stage of disease, histological subtype and smoking history), or both clinical and gene expression [32].

Owing to the availability of large lung cancer data sets, several statistical research groups have performed meta-analyses in which they searched for commonalities among large expression studies [35,39,42,43]. Lu et al. performed a meta-analysis that demonstrated the feasibility of combining different DNA microarrays to increase sample size and predictive power and identify a robust gene expression signature predictive of survival [44].

The use of microarrays in clinical practice, however, is limited by the large number of genes in the analysis, complicated methods, lack of reproducibility, independent validation of the results and the need for fresh-frozen tissue; moreover, the gene expression profile can vary according to the microarray platform, the analytic strategy used and the samples. These variances may be derived from variation in printing or processing of chips, hybridisation or scanning, sample preparation, or probes. Real-time PCR involving a small number of genes may be a more useful method to determine the gene expression in small amounts paraffin-embedded specimens [32,45,46]. So far, none of the signatures has performed significantly better than the others, and it remains unclear not only which genomic predictor of prognosis is the best but also whether specific genes or entire signatures are most important in predicting outcome.

In addition, large longitudinal studies measuring gene expression as well as routine clinical, biochemical and pathologic measures are needed to demonstrate gene expression is a better predictor of outcome than more routine measures. The ultimate barrier to adoption of these markers in the clinic is the need for more of them to be validated in prospective multi-centre studies to demonstrate their reproducibility and accuracy across multiple sites and operators. The goal is to construct a lung cancer-specific gene chip for prospective testing in clinical trials. To create a prognostic model that could reasonably be used in routine clinical practice, a microarray technology to use in clinical and regulatory settings by examining repeatability of data generated within a particular site, across multiple sites and between seven different microarray platforms has been evaluated by the FDA (the Microarray Quality Control [MAQC] project) [47]. The reproducibility of gene expression measurements between sites and across platforms demonstrated by these studies is a critical milestone in the development of gene expression biomarkers that can be routinely used in the clinic.

5.2 Potential lung cancer genetic biomarkers

5.2.1 Self-sufficient or autonomous growth signals

5.2.1.1 Epidermal growth factor receptor

Among the great number of cellular biology alterations, epidermal growth factor receptor (EGFR) has received particular attention owing to its relevance in predicting response to new ‘targeted’ therapies: EGFR is a family of transmembrane receptor tyrosine kinases. In a review of 16 studies published between 1989 and 2001, the overall expression rate in NSCLC was 51%, whereas it reached 82.6% in squamous cell carcinoma [48]. EGFR mutation is expressed predominantly in Asian, non-smoker females with adenocarcinoma and its occurrence is mutually exclusive with K-ras mutation in virtually all patients, probably because K-ras is a downstream signalling pathway of EGFR [49].
New biomarkers for lung cancer

Mutation of EGFR failed to demonstrate a usefulness in diagnosis and prognosis but it is actually one of the best markers predictive of response to therapy addressed for targeted therapies: drugs available at present that act in this pathway are represented by inhibitors of tyrosine kinase (TKi) intracellular domain of EGFR, such as gefitinib and erlotinib, or monoclonal antibodies against extracellular domain, such as cetuximab. In a meta-analysis of Gupta et al., most of the studies, despite some bias in their design, provided a statistically significant value of EGFR assessed by immunohistochemistry (IHC) (cell expression), FISH (gene amplification) or PCR (gene mutation) in predicting response to TKi therapy in lung adenocarcinoma [58]. However, a wide variability among the studies was reported in terms of PPV (6.5 – 100%) and NPV (50 – 99%), even for the ‘best’ and largest double-blind randomised clinical trial: a 95% NPV was, however, reported for IHC and FISH, making EGFR-negative patients unlikely to respond to TKi-based therapies unless EGFR amplification (not mutation) is present [51]. Some authors thought that EGFR mutations were mainly prognostic and consequently that they identify a subset of patients with good prognosis irrespective of treatment, but also IPASS study (a Phase III trial comparing gefitinib with standard chemotherapy in non-smokers Asian patients with adenocarcinoma) confirmed higher progression-free survival in the gefitinib group: the lack of statistical significance was probably a result of high rate of treatment crossover (change of treatment between the two arms after progression of disease) and administration of TKi as a first-line drug even in EGFR-negative patients [52].

5.2.1.2 HER-2
Overall expression of HER-2 in NSCLC accounts for 35% and is significantly different in histological subtypes [53]: it is expressed more frequently in adenocarcinoma (ADK) (38%) where it seems to be associated to poor 3- and 5-year survival rate. This prognostic value, however, has not yet been confirmed.

5.2.1.3 KRAS
RAS is a small G-protein with intrinsic GTPase activity: it is a critical regulator of signalling downstream of cell surface receptors and is mutated in various types of human cancer, so much interest is focused on targeted therapies for this pathway. In lung adenocarcinoma, the oncogene KRAS is mutated in ~ 20% of cases [54], mainly in ADK or large-cell carcinoma (LCC). The most frequent ras mutation discovered in lung cancer patients is located in codon 12 and is known to be related to pre-cancerous lesions [55]. Reports regarding prognostic significance of K-ras mutation are contrasting: some studies reported a lower survival in patients having a mutation [56], whereas others did not find any difference in patients with NSCLC [57]. Inhibition of KRAS farnesylation (essential for membrane localisation) was not efficient alone and concomitant inhibition of alternative geranyl-geranyl transferase was excessively toxic [49]. On the other hand KRAS inhibition alone was ineffective at killing all cells in vitro, meaning that only 50% of RAS-mutated cells were RAS-addicted regarding their carcinogenesis [58]. One alternative approach could be the combined inhibition of ERK (extracellular signal-regulated kinase) and PI3K (phosphatidylinositol 3 kinase), the two main downstream of RAS-RAF-MAP kinase pathways: combination of MEK and PI3K inhibitors in particular led to almost complete tumour regression in the KRAS G12D rat model [59]. Another approach included HSP90 inhibitors, based on evidence that KRAS confers higher heat-shock protein 90 dependency, and more recently the use of TBK1, an upstream regulator of NF-κB [60].

5.2.1.4 c-MET
Met codifies for a tyrosine kinase receptor that binds hepatocyte growth factor (HGF, a polypeptidyl cytokine involved in cell proliferation, motility and angiogenesis. HGF-Met signalling is usually paracrine, but lung cancer cells can express both ligand and receptor, making the signal system autocrine [61]. Overexpression of c-MET is reported in 40 – 60% of NSCLC [62] and is associated with elevated levels of the serum circulating form [63], and seems to be related with N stage and probability of early recurrence. However, prognostic significance of Met/HGF in NSCLC remains uncertain and needs to be validated by further larger studies.

5.2.1.5 CYCLIN D1
Cyclins (CCN) and associated kinases control progression through phases of cell cycle and their constitutive high expression may result in uncontrolled proliferation. CCND1 is the most studied cyclin in NSCLC, where it is overexpressed in ~ 50% [64]. Overexpression has been noted also in smokers, but it has not been studied extensively as a marker for screening [65]. The prognostic value remains uncertain: some reports showed a negative impact on survival [66], in others it was associated with better survival [67], whereas others found no correlation [66].

5.2.1.6 Proliferating cell nuclear antigen
Proliferating cell nuclear antigen is the subunit of DNA polymerase responsible for its proofreading activity during DNA replication: in most studies no prognostic value is actually reported in lung cancer [68].

5.2.1.7 Ki-67
Ki-67 is another nuclear antigen that is expressed only in proliferating cells. A recent meta-analysis showed that only 41% of the studies reported a negative effect of Ki-67 overexpression on prognosis for patients with NSCLC. The aggregated survival data show that Ki-67 immunoreactivity is associated with poorer survival (hazard ratio [HR] 1.55) and 95% confidence interval (CI).

Ki-67 is commonly considered to be a marker of cellular proliferation; in a meta-analysis on 16 studies it demonstrated
an association with poorer prognosis with a hazard ratio of 1.55 (95% CI: 1.34 – 1.78) [69].

5.2.2 Insensitivity to growth inhibitory signals

5.2.2.1 Transforming growth factor-β
Transforming growth factor-β (TGF-β) is an antiproliferative factor for epithelial cells. Few studies have provided contradictory results in the prognostic significance of TGF-β expression [70].

5.2.2.2 P21, p27, p57
These are a family of genes (CIP/KIP) that inhibit CDKs (promoters of cell cycle progression) whose loss leads to tumour progression. Data on p21 are discordant with significant differences in the studies [71,72]. Loss of p27 expression has been reported in ~ 30% of NSCLC; some studies have reported an adverse survival effect on the loss of p27 protein expression [73,74].

5.2.2.3 P16, p15, p18, p19
This is another family genes (INK4) that inhibit CDKs (promoters of cell cycle progression). P16 is inactivated in ~ 50% of NSCLC; otherwise, only two studies showed a prognostic significance in multivariate analysis [75,76].

5.2.2.4 Retinoblastoma
Retinoblastoma (Rb) gene mutations have been found in > 90% of cases of SCLC and in ~ 15% of NSCLC [54]. Mutations and dysregulated phosphorylation of the gene could cause a loss of genetic function, which results in an unfavourable (poor) prognosis in patients with lung tumours [77]. Also, the absence of Rb protein seems to cause a negative prognostic rate [78], but several studies have reported controversial data and the prognostic significance of Rb in NSCLC has to be resolved [79].

5.2.3 Inhibition of apoptosis

5.2.3.1 p53, MDM2, p14
P53 is the most frequently mutated gene in human cancers. Its function is the activation of Bax, a proapoptotic factor for the mitochondria pathway. P53 is inactivated in 50% of NSCLC (particularly in squamous cell carcinoma) and in 90% of SCLC [80,81]. P53 mutations have been found in 10 – 30% of patients with lung cancer compared with minimal detection in the healthy control groups. One study found p53 mutations in 41% of lung tumours, with the identical mutation identified in the plasma of 73% of them [82,83].

Despite many studies focused on p53, its prognostic significance in lung cancer remains uncertain [84]; actually it is rather considered a marker of increased risk of cancer, even if non-lung-specific.

The MDM2 gene codifies for the E3 ubiquitin ligase that is involved in p53 nuclear export and degradation. MDM2 and p53 form a feedback loop modulated by p14, a product of the INK4 gene. P14 binds MDM2 and prevents its binding with p53 MDM2 gene amplification is associated with poor prognosis [85].

5.2.3.2 Bcl2-Bax
Bcl2-Bax are factors with antiapoptotic (Bcl2) and proapoptotic (Bax) function. The expression of Bcl-2 is present in ~ 35% of NSCLC (61% in adenocarcinoma, 32% in epidermoid). A meta-analysis showed a better survival outcome for patients with Bcl-2-positive tumours (HR 0.72, 95% CI 0.64 – 0.82) [86]. No prognostic significance exists for Bax in NSCLC [87-89].

5.2.4 Gene promoter methylation
In the carcinogenesis process methylation of CpG islands within the promoter sequences of several tumour-suppressor genes is an early event and inhibits transcription, causing gene silencing [90]. More than 80 genes have been found to be methylated in lung cancer [91].

One of most investigated alterations is mutilation of onco-suppressor gene promoter p16, present in 50% of NSCLC [92]; this has been proposed as a biomarker for early detection of lung cancer and monitoring of prevention trials [93]. In a recent paper, Belinsky et al. analysed promoter methylation of 14 genes in 3259 subjects and discovered that 6 genes (p16, MGMT, DAPK, RASSF1A, PEX5 b and GATA5) were individually associated with a 50% increased risk of lung cancer. Methylation of two or more of these genes carried a 64% sensitivity and specificity in predicting cancer disease [94]; this level of accuracy is not yet high enough for prospective screening studies, but called for more evaluation of candidate gene panels as technology improves.

Han-Shui Hsu et al. showed that multiple epigenetic markers in plasma, especially the p16 and RASSF1A genes, can be used for lung cancer detection [95]. Hypermethylation of the p16 promoter was reported in lung cancer induced by the inhalation of cigarette smoke in F344/N rats and in the bronchial epithelium and sputum of smokers [96]. This methylation marker panel should improve the detection of cancer or the risk assessment for lung cancer in combination with conventional diagnostic tools.

Changes in the methylation pattern of specific promoter regions have been shown to help discriminate histological types of lung cancer and to correlate with worse prognosis [97-99]. One reports shows that adenocarcinoma and squamous cell carcinoma typically show two different patterns of methylation of gene OLIG-1 [100]. Another study demonstrated that in patients with tumours known to be hypermethylated at a specific location, 63% were detectible in the BAL fluid of those patients [101].

Methylation of a tumour-suppressor gene, however, does not necessarily indicate that it is tumour-specific, because CpG island hypermethylation of some tumour-suppressor genes occurs after the onset of neoplastic evolution, and others become hypermethylated initially in normal epithelial cells by environmental factors such as exposure to tobacco and ageing.
New biomarkers for lung cancer

Thus, such factors could be the source of false positives in the study of tumour-specific methylation.

5.2.5 Gene expression regulation
5.2.5.1 MicroRNA
MicroRNAs (miRNAs) are a class of single-strand endogenous non-coding RNA that act as gene expression regulators at post-transcriptional level. One miRNA can have also thousands of possible targets and ~74 ~ 92% of gene transcription products are likely under miRNA control [102]. miRNAs come from long hairpin precursors spliced by a cellular nuclease (Drosha) and exported in the cytoplasm, where they are cleaved further by Dicer enzyme in 17 ~ 24 nt filaments. miRNAs probably have a wide range of biological functions, including cell proliferation, apoptosis and stress resistance.

Since the first miRNA discovered (lin-4) in 1984, >8000 mature miRNA have been counted in primates, birds, fish, plants and viruses. Genes encoding miRNAs are frequently located at fragile sites, as well as regions of loss of heterozygosity, amplification and common breakpoint regions, suggesting that they could be considered as a new class of genes involved in human carcinogenesis [103]. Interestingly, miRNAs not only control the expression of known protein-coding oncogenes and tumour suppressors, but also act as oncogenes and tumour suppressors directly.

Several studies have demonstrated that miRNAs could be mutated or over/underexpressed in tumours and that patterns of expression could have utility in diagnosis, progression and prognosis staging, and respond to the therapy. Besides, they could be an important tool for the differential diagnosis between primary lung cancer and metastasis [104-106]. Moreover, the tumour-derived miRNA are resistant to RNase digestion; these data suggest that circulating miRNAs could be clinical biomarkers for blood-based detection of human cancer, reflecting profile expression of solid tissues [107]. By comparing profile expression of miRNA between lung cancer tissue versus the corresponding non-cancerous lung tissue, a variable number of miRNA (35 ~ 45) showed statistical significant difference. Unfortunately not so many RNAs present the same pattern of expression in lung cancers, some being upregulated and others being downregulated in contrasting reports [104,108]. A remarkable aspect is the altered profile expression of miRNAs consisting in downregulation of those with onco-suppressor activity as demonstrated in smoke rat lung [109], reflecting a very early phase of carcinogenesis.

Six miRNAs (miR-205, miR-99b, miR-203, miR-202, miR-102, miR-204) were found to be expressed differently in adenocarcinoma and squamous cell carcinoma, whereas five miRNAs (miR-21, miR-191, miR-155, miR-210, miR-126) shared both histological types of NSCLC [110]. The first test utilising unique miRNA profiles to differentially diagnose squamous from non-squamous was developed by Rosetta Genomics and was able to classify squamous from non-squamous NSCLC at 90% of specificity and 96% of sensitivity. Recently, a single miRNA was proven to be discriminating in the diagnosis of squamous cell lung cancer [111].

A recent study identified a five miRNA-signature profile (let-7a, miR-221, miR-137, miR-371, miR-182) closely associated with overall and disease-free survival and cancer relapse in NSCLC patients independently from histology and stage subgroups [112]. Low expression of let-7a and overexpression of miR-155 were associated with shortened postoperative survival, and this prognostic impact maintained significance after adjusting for age, sex, disease stage and smoking habits in multivariate analysis [110]; on the contrary, reduction of let-7 expression was not correlated with the prognosis of BAC [113]. Further studies identified eight miRNA to be related to prognosis of patients with adenocarcinoma: high expression of miR-155, miR-17, miR-3p, miR-106a, miR-93, miR-21 and low expression of let-7a-2, 7b, miR-145 were significantly associated with worse prognosis [102]. Let-7 was the first miRNA downregulated in lung tumour to be discovered and it is one of the most examined because of its potential anticancer activity (it is able to inhibit lung cancer cell growth in vitro). Downregulation of Let-7c is detected by northern analysis in 25 ~ 75% of lung cancers and ~ 30% for Let-7g [114]. Let-7 is a negative regulator of Ras family and MYC gene pathways (N-Ras, H-Ras, K-Ras) acting as a tumour-suppressor gene [115] and probably also as an antiangiogenetic factor [116]. Moreover, it is hypothesised that it could regulate levels of dicer (DCR) and consequently the equilibrium of various miRNAs [117]. The association of miRNAs and tumour invasiveness is not well established yet, although transfection of three miRNAs (miR-137, miR-182, miR-372) conferred higher invasive potential to a low-invasive lung cancer line (CL1-0) [112].

On the contrary, the miR-17-92 cluster (containing miR-17-18-19-20-92) is highly overexpressed in lung cancer [118] and also in normal lung during embryonic development: transcription is upregulated by the c-Myc oncoprotein while targets include both oncogenes and tumour-suppressor genes.

MicroRNAs are potential new biomarkers in lung cancer, especially for early diagnosis and prognosis (i.e., biological behaviour) [119-121]. Their role as oncogenes or onco-suppressor genes is very relevant in pharmacetics, as each endogenous oncogenic miRNA could be inhibited by single-stranded antisense nucleotides (siRNA), whereas overexpressing onco-suppressor miRNAs by viral vector could be induced to block cancer cells growth both in vivo and in vitro. A few examples of both categories have already been proposed and tested in Phase I/II trials with promising results: the advantages of miRNA-based drugs should be in vivo stability, long-term activity and low toxicity, although caution should be taken in systemic administration of these drugs because of uncontrolled gene-silencing (siRNA are relatively more specific) [102].
5.2.6 DNA repair mechanism alteration

5.2.6.1 ERCC1/RRM1

Although cisplatin-based adjuvant chemotherapy showed a benefit in survival of surgically resected patients, the overall impact is low: a biomarker of sensitivity/resistance to cisplatin would allow the selection of responsive patients in an adjuvant setting, avoiding exposure to useless adverse side effects. Resistance to platinum agents depends on DNA repair mechanisms.

ERCC1 (Excision cross complementation group 1) and RRM1 (Regulatory subunit of ribonucleotide reductase) both belong to the nucleotide excision repair pathway (NER), which is responsible for the repair of various DNA errors caused by tobacco carcinogens and cisplatin adducts. NER was first studied in xeroderma pigmentosum, but appears to affect response to both surgical and medical treatment of NSCLC. High levels of RRM1 are associated with better prognosis and longer disease-free survival [122].

The most consistent evidence about the role of ERCC1 and RRM1 in predicting response to chemotherapy comes from the International Adjuvant Lung Cancer Trial (IALT): patients who received chemotherapy had a better survival in the presence of ERCC1-negative tumours, whereas patients who did not receive chemotherapy showed a longer survival in the presence of ERCC1-positive tumours [123]. To explain these contrasting results it has been hypothesised that ERCC1 plays a positive role in early untreated patients, preventing further DNA mutations that could lead to increased cell proliferation and invasiveness, but the same mechanism could contrast the cisplatin mechanism of action by repairing DNA mutations from cisplatin-adducted products themselves. The introduction of this rationale in choosing drugs in an adjuvant setting demonstrated a favourable impact on prognosis, improving survival from 39.3 to 50% [124].

5.2.7 Genomic instability

Microsatellite instability (MSI) refers to the insertion of repeating nucleotide units within the genome: these insertions result in frameshift mutations and aberrant protein expression. The most widely used method for MSI study is a PCR-based method. Studies have reported a high percentage of MSI frequency in patients with lung cancers [125].

Examination of chromosomes for loss of heterozygosity (LOH) is used as an indicator for the presence of a tumour-suppressor gene locus. LOH can indicate an early stage of cancer or simply the susceptibility to develop it [126]. In lung cancer the loss of heterozygosity of chromosomal regions on 3p, 5q, 9p and 12p is well documented [127,128]. Wistuba et al. performed high-resolution chromosome allelo-typing using a panel of 28 3p markers and showed that 3p losses were found in 96% of lung cancers and 78% of preneoplastic/preinvasive lesions [129].

Another potential lung genetic biomarker is BJ-TSA-9. It is a tumour-specific gene highly expressed in lung cancer tissues. In the study of Li et al., the BJ-TSA-9 mRNA was expressed in 52.5% (21 of 40) of human lung cancer tissues and was especially high in lung adenocarcinoma (68.8%). The overall positive detection rate was 34.3% (24 of 70) in peripheral blood mononuclear cells (PBMCs) of patients with various types of lung cancer and was 53.6% (15 of 28) in PBMCs of lung adenocarcinoma patients [130]. In combination with two other tumour markers, squamous cell carcinoma antigen (SCC) and lung-specific X protein (LUNX), whose mRNA expression is strictly limited to lung tissues, the detection rate was increased to 81.4% [131]. These genes may serve as a marker for lung cancer diagnosis, and in combination as a valid marker for prediction of recurrence and prognosis.

6. Proteomic approach

Despite several advantages of the genomic-based biomarkers, some limitations exist. First, accessibility to DNA from tissues, sputum or blood is limited to samples in which a sufficient amount of DNA is present. Second, reverse transcription (RT-PCR) or microarray data delineate the level of RNA present, which may not always indicate the level of protein present because many events occur during translation and post-transcriptional modification processes.

Proteins are easily obtained non-invasively from a variety of sources, including blood, sputum and exhaled breath condensate; the differential expression of proteins in malignant cells allows us to see exactly how the genetic changes are translated. Genomic analysis in fact is not able to recognise post-translational modifications such as proteolysis, glycosylation, phosphorylation and acetylation, which are essential for biological activity [132].

Proteomic-based biomarkers can include proteins involved in inflammation or protein fragments generated by the aberrant tumour environment, proteins that act on tumour generation in response to the action of oncogenes and tumour-suppressor genes, as well as proteins involved in the acquisition of unlimited proliferation potential, sustained angiogenesis, and capacity of tissue invasion and metastatization. Biomarker proteins can be detectable in the circulation as the free, shed proteins or as new autoantibodies to such proteins, the latter indicating that the host immune system can be exploited as a biosensor of the disease.

6.1 Proteomics technologies

Several proteomic methodologies, such as two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), surface-enhanced laser desorption/ionisation time-of-flight (SELDI-TOF), matrix-assisted laser desorption/ionisation (MALDI-TOF), protein arrays, iTRAQ and multidimensional protein identification technology (MudPIT) have been used to analyse different biological samples, including serum, plasma and cancer tissue, in order to identify new tumoural biomarkers.
New biomarkers for lung cancer

6.1.1 Two-dimensional electrophoresis
Since 1974, two-dimensional electrophoresis has been the most widely utilised technique in the proteomic field to study the protein expression profile and to research biomarkers in patients with a cancer [133,134]. Since then, several reports concerning the study of the proteomic profile and the discovery of molecular markers in lung cancer using this technique have been published [135-137]. A recent discovery considered the identification of annexin A1 in tumour microvascular endothelial cells as a possible molecular candidate for diagnostic imaging and therapeutic targeting [138]. Compared with 2D-PAGE, 2D-DIGE contemporaneously allows the study of two different proteic patterns by using fluorescent tags. This technique presents a high reproducibility and allows an accurate quantification of protein expression differences [139].

6.1.2 MALDI-TOF
Mass spectrometry allowed the identification of proteins present in the clinical samples by the technique of matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI TOF-MS) combined with the use of a database [140]. Yanagisawa et al. demonstrated that this technique is 100% accurate at distinguishing the healthy from the tumoural, and the histological type of the tumour [141]. Using two-dimensional electrophoresis coupled with MALDI-TOF-MS, Huang et al. identified 14 human proteins from the conditioned media of a non-small cell lung cancer cell line A549 [142]. Moreover, Alfonso et al. analysed 12 surgically resected lung cancers with a two-dimensional gel combined with MALDI MS and identified several proteins reported previously (such as annexin II, cathepsin D, HSP27, stathmin and MnSOD), confirming the validity of this technique in the identification of candidate biomarkers [143].

6.1.3 SELDI-TOF-MS
In the recent years, a shift has been seen from SELDI-TOF discovery, to direct spectral or spectral abundance-based comparison, over to label-based methods. A small amount of sample of interest is loaded onto ProteinChip arrays that selectively bind different subsets of proteins in crude samples according to their surface chemistries. Biomarker Wizard software analyses the spectral map and detects differentially expressed protein/peptides with statistical significance. Applications of SELDI-TOF have been demonstrated for the early detection of lung cancer [144]. However, there is some controversy over this technology [145].

6.1.4 Protein array technologies
Reverse phase arrays is a more suitable method for biomarker screening or validation than discovery of new proteins; hundreds of patient specimens can be spotted onto the same array, allowing a large number of samples to be compared simultaneously under the same conditions [146]. Each array is incubated with one particular antibody, and signal intensity proportional to the amount of analyte in the sample spot is generated [147]. Signal detection is commonly performed by fluorescence, chemiluminescence or colorimetric methods and the results are quantified by scanning and analysed by software.

A recent report has shown that cell lysate from a lung cancer cell line was fractionated and arrayed onto a nitrocellulose-coated glass slide. Sera samples from 14 lung cancer patients, colon cancer patients and normal subjects were then incubated with the array slides. A total of eight fractions of the cell lysate were found to be recognised by sera from four patients, whereas none of the sera from normal individuals was positive [148]. The application of protein arrays in biomedical research is limited because of the production cost and the low specificity of antibodies used for detecting proteins in the sample.

6.1.5 iTRAQ and MudPIT
To perform differential labelling of clinical samples with iTRAQ, individual samples are trypsinised and each sample is labelled with a different isobaric label. Differentially labelled peptide mixtures are then combined at equimolar concentrations and further separated by two-dimensional chromatography, tandem mass spectrometry and database searching [149,150]. MudPIT is a non-gel approach that uses multidimensional high-pressure liquid chromatography (LC/LC) separation, tandem mass spectrometry and database searching [151]. With shotgun proteomics combined with MudPIT applied to the plasma of patients with lung cancer, 120 proteins have been shown to be expressed exclusively in the plasma of patients with lung adenocarcinoma [152,153]. A remarkable new orthogonal MS-based clinical assay, multiple-reaction-monitoring (MRM-MS), is expected to accelerate the discovery of cancer biomarkers, the verification of the biomarkers, and also their clinical translations [154]. Modification of this tool using isotope-coded antibody capture technology or by multidimensional fractionation will increase the sensitivity to nanograms per millilitre levels, which can make it possible to diagnose direct biological samples targeting specific cancer biomarkers [155-157].

6.2 Lung cancer protein biomarkers: potential
It is remarkable that some potential high-protein biomarkers for lung cancer exist that have not been used in clinical practice and operated in angiogenesis pathway, in invasion and metastasis pathway and in unlimited growth potential pathway. None of markers that have been investigated has shown consistent studies. CCNE, VEGF-A, p16, p27, betacatenin and E-cadherin have shown correlation with poor prognosis in a large number of studies.

6.2.1 Angiogenesis pathway
Vascular endothelial growth factor (VEGF) is a cytokine involved in angiogenesis in both physiological and pathologi- cal conditions; it is overexpressed in 60% of NSCLC and its presence (evaluated by immunoassay) correlates well with
tumour vascularisation, progression of disease and poor prognosis [158], although there is not complete consensus on that [159,160].

Connective tissue-activating peptide III (CTAP III) and neutrophil-activating protein-2 (NAP-2) belong to the sub-family of ELR-CXC chemokines that are powerful promoters of angiogenesis, tumorigenesis and metastases. Blood levels do not correlate with stage of disease but only with prognosis and risk of recurrence: in particular, levels remain higher even after surgery in patients that will develop local or distant disease. These two biomarkers have both been demonstrated to be able to detect lung cancer in stage 0/Ia and to improve prediction of lung cancer when considered in a model with age, smoke and lung function (AUC 0.84 versus 0.80) [161].

IL-8 is an effective angiogenic cytokine found to be upregulated in cancer cells (up to 270-fold) by interaction with tumour-infiltrating macrophages, resulting in increased metastatic potential in vitro and worse prognosis [162].

Epithelial neutrophil-activating peptide (ENA-78) is another marker of tumour vascularisation in human specimens and also cancer growth rate in rat models expressing human cancer cells lines [163].

6.2.2 Invasion and metastasis pathway
Adhesions or junctional molecules (integrins, cadherins, selectins, immunoglobulin super gene family [IgSF], CD44) play an important role in tumour progression. Reduced expression of catenins (reduction of beta-catenin is found in ~30% of NSCLC) and E-cadherin, two partner components of adherence junction cellular proteins, has been found to affect prognosis adversely, probably through higher metastatic potential [164,165]. A favourable prognosis is also reported for NSCLC with normal E-cadherin expression [166-168].

Matrix metalloproteinases (MMP) are a family of enzymes that can degrade fibrillar collagen and are also involved in metastatic spread. Results from most of the studies reported that high expression of MMP-2 and MMP-9 are poor prognostic markers for NSCLC, especially the expression of MMP-2 [169]. Most of the studies focused on tumour expression of MMP rather than stromal expression, which seems to be related to prognosis [170].

Nectin-4 oncoprotein is an immunoglobulin-like molecule overexpressed in NSCLC cells by transactivation of gene with a sensitivity and specificity are 53.7 and 97.7%, respectively. It increases the invasive ability of cells through activation of small GTPase Rac1 and it should be a new candidate as both serum/tissue biomarker and therapeutic target, as small RNA interfering probes against Nectin-4 suppress cell growth [171].

6.2.3 Unlimited growth potential pathway
The protein catalytic subunit human telomerase reverse transcriptase (hTERT) is responsible for genomic stability. hTERT is frequently expressed in NSCLC (50 – 90%); it has been correlated with telomerase activity without any predictive value [172].

Otherwise, the existence of a large number of markers is notable, which have been less investigated for their putative diagnostic and prognostic value for lung cancer with promising or controversial results. These include the following proteins. A fragment of serum amyloid A (SAA, molecular mass 11.6 kDa) is significantly elevated in lung cancer patients. Its level increases with the clinical stage and is higher in squamous cell carcinoma subtype [173].

The secretion of dihydrodiol dehydrogenase (DDH) protein has been studied in lines of pulmonary adenocarcinoma A549 and the levels of proteic expression and of miRNA in NSCLC cancer tissue have been evaluated. The protein has subsequently been researched in serum of patients with NSCLC and healthy patients, showing a significant increase in them [142,174].

Recently, a serum glycoprotein, plasma kallikrein B1 (KLKB1), has been identified as a potentially useful biomarker in lung adenocarcinoma B7-H3 [175], which belongs to a family of immuno-modulating molecules (inhibitors of natural killer and T cells), and may have a role in escaping immune response by lung and other solid cancers.

Aberrant expression of a membrane and soluble form of B7-H3 was assessed in 68 patients with NSCLC and was demonstrated to be a valuable marker with comparable specificity (90 – 98%) but much higher sensitivity (48 – 71.4%, depending on cutoff value), with an AUC of 0.8 (versus patients with COPD) and 0.87 (versus healthy subjects), which is much better than any other marker used at present [176].

Another protein biomarker is PGP 9.5, a neurospecific polypeptide, which has been studied in the past and is overexpressed in tumourial tissue [177]. Recently, some direct antibodies against PGP in patient serum with lung cancer have been highlighted [178]. Also, carbonic anhydrase IX [179], whose positive expression is associated with poorer prognosis and fragile histidine triad (FHIT), which is associated with better outcome [180].

Finally, Haptoglobin [181], fragment of apolipoprotein A-1 [142] and heterogeneous ribonuclear proteins A2/B1 (hnRNP A2/B1) [182] are also mentioned as lung cancer biomarkers with controversial results.

6.2.4 Autoantibodies
One potential group of serum biomarkers are autoantibodies that target specific tumour-associated antigens. The discovery of panels of tumour-associated antigens and autoantibody signatures with high sensitivity and specificity would aid the development of diagnostics, prognostics and therapeutics for cancer patients. The production of autoantibodies as a result of cancer immunosurveilance has been found to precede manifestations of clinical signs of tumorigenesis by several months to years [183]. The tumour-associated antigens are proteins that become immunogenetic after overexpression, mutations, incorrect degradation or wrong post-translational modifications [184]; the presence of these abnormalities could induce an immune response, with the production of...
New biomarkers for lung cancer

autoantibodies that could be noticed in blood and be considered biomarkers of the presence of a tumour [185]. Autoantibodies against p53 have been identified in several kinds of tumour, including lung cancer, and seem to be associated with high-grade tumours and poor survival [186]. Although a single autoantigen would lack adequate sensitivity and specificity, a panel of tumour-associated antigens may overcome this problem by enabling multiple autoantibodies to be detected simultaneously [187].

Zhong et al. reported the use of combined autoantibodies as markers for early detection of non-small-cell lung cancer. They achieved >90% sensitivity and specificity in diagnosing stage I non-small-cell lung cancer using 5 antibody markers in serum samples [188]. Another recent study identified direct autoantibodies of serum against 14-3-3 theta protein and, moreover, tested the serum against a panel of 14-3-3 theta, annexin I and PGF 9.5 proteins. This panel gave a sensitivity of 55% and specificity of 95% in identifying lung cancer at the preclinical stage [189]. More than discriminating healthy individuals from patients with lung cancer, autoantibodies can be useful in distinguishing several subtypes of lung cancer [190]. Tureci et al. identified direct autoantibodies against NY-ESO-1, useful in distinguishing patients with SCLC and NSCLC [191]. Personalised profiles of tumour associated antigens (TAAs) and autoantibodies could be used to identify therapeutic targets to develop new vaccines for targeted immunotherapy against cancer. For example, constant and strong expression of neural cell adhesion molecule (NCAM) in small cell lung cancer supports the development of anti-NCAM immunotherapy for this cancer [192].

7. Metabolomic approach

Metabolomics is rapidly becoming another large area of study in cancer biomarker discovery. Metabolomics refers to the study of small-molecule metabolite profiles in specific times and under specific conditions, which represent the metabolic state of the cell or tissue. Metabolites include any metabolic intermediates, hormones and other components of signalling pathways [193].

Cancer cells are known to possess a highly unique metabolic phenotype and the analysis of metabolites may be used in the development of tumour-specific biomarkers because it gives a complete picture of metabolic changes that result in the malignant phenotype [194].

Current knowledge of lung cancer metabolism is limited. At this early stage of application to lung metabolism, a few metabolomics studies have demonstrated the utility of NMR and MS in providing global metabolite profiles in lung cells and tissues, bronchoalveolar lavage fluids, and urine from model animals grafted with lung cancers [195]. One study was conducted on lung cancer tissues resected from patients after administering the stable isotope tracer [U-13C]-glucose. The metabolic changes in human lung cancer patients were investigated by infusing labelled 13C-glucose and in particular showed an altered but full Krebs cycle activity in lung tumour tissue and activation of anaplerotic pyruvate carboxylation [196].

8. Biomarkers in biological fluids

Several specimens have been investigated in the search for a useful and reliable source of biomarkers for early diagnosis of lung cancer. These include naturally produced or induced sputum, BAL, EBC, pleural fluid and blood (serum/plasma) (Table 2).

8.1 Bronchoalveolar lavage

Many biomarkers for the precocious diagnosis have been researched in BAL. The bronchoalveolar lavage is conducted by reaspirating the sterile saline solution infused into the airways of a patient undergoing bronchoscopy. DNA methylation and genomic instability are analysed in BAL specimens and cell-free lavage supernatants. The DNA methylation has been detected in BAL from lung cancer patients and healthy controls [197-201]. Kim et al. also reported a good correlation between methylation in tumours and BAL, ranging from 39 to 61% for the 5 loci they analysed [197]. Schmiemann et al. applied a marker panel (APC, CDKN2A/p16, and RASSF1) to detect cancer in 247 patients and reported 53% sensitivity and, in cases without a previous history of cancer, >99% specificity [201]. Also, the genomic instability can be identified in bronchoalveolar lavage of patients with cancer: a study showed a sensitivity of 73.9% and specificity of 76.5% for a panel of markers [202].

Unluckily, nowadays potential biomarkers studied in the material from BAL present a high specificity but a low sensitivity to be used as markers of the presence of lung tumour. Analysis performed on DNA from cell-free lavage supernatants was more promising, for example, qualitative and quantitative analysis of extracellular mRNA gives interesting results [203].

8.2 Sputum

The advantages of sputum as a screening tool include its non-invasive procurement, and the fact that it contains cells from the lungs and lower respiratory tract. Sputum is produced by increased bronchial secretions and is commonly found in smokers, hence it can be used to screen high-risk populations. As the cytological evaluation of sputum samples presents a low sensitivity, several studies have addressed the identification of molecular biomarkers. DNA mutation, promoter hypermethylation and microsatellite alteration in sputum samples may be suitable biomarkers for the detection of lung cancer [94,204,205]. Abnormal methylation in the cytologically negative sputum samples was displayed by 64% of HOX A9, 50% of MAGE A1, 41% of MAGE B2 and 27% of p16 promoters, and 95.5% of the negative sputum samples from NSCLC patients showed abnormal methylation in at least one gene tested [206]. Nuclear image analysis appears...
highly promising for improving or refining diagnosis beyond the use of a conventional sputum cytology examination. Palcic et al. analysed malignancy-associated changes in non-malignant sputum cells by means of semiquantitative nuclear function; these characteristics make it an ideal biological fluid in invasive procedure, unrelated in any way to patient airway.

8.3 Exhaled breath condensate

EBC collection is an easy, repeatable and totally non-invasive procedure, unrelated in any way to patient airway function; these characteristics make it an ideal biological fluid for an eventual screening procedure and precocious diagnosis of lung cancer [209]. However, standard procedures concerning sample collection and results presentation need to be established. In exhaled breath there are numerous biologic molecules that can be able to reflect the status of lung, in fact, breath condensate is composed not only by water vapour but it also contains lipids, proteins, DNA and products of oxidation derived from the fluid lining the respiratory tract. The quantity of these molecules depends on the state of health of the patients in the analysis [210]. Carpagnano et al. studied the levels of ET1-1 in EBC of patients with lung cancer and found significant differences between healthy controls and NSCLC patients, and between stage I-III and stage IV patients [211]. The same group demonstrated that the number of microsatellite alterations on chromosome locus 3p in DNA present in EBC of patients with NSCLC and healthy patients was significantly higher in the NSCLC patients [212], and recently they found that the number of 3p microsatellite alterations found in the exhaled breath condensate DNA shows a remarkable correlation with patient’s survival [213]. In a recent case/control study, they found a high quantity of three analytes (II-2, TNF-α and leptin) in patients, particularly in an advanced phase of the disease [214].

Some researchers reported that unique volatile organic compounds (VOCs) are exhaled in the breath of patients with lung cancer, and this may have potential diagnostic importance [215]. Since then new techniques using VOCs have been developed. Mazzone et al. used a colorimetric sensor array in 143 subjects: 49 had NSCLC, 73 had various chronic lung diseases such as COPD, idiopathic pulmonary fibrosis, pulmonary arterial hypertension and sarcoidosis, and there were 21 controls. Once a prediction model had been developed using 70% of the subjects, it could be used to predict the presence of lung cancer in the remaining 30% with a sensitivity of 73.3% and specificity of 72.4%. Although this suggests only moderate accuracy of the diagnosis, the study does prove the potential of this technique [216].

Finally, by mass spectrometry Giannazza et al. identified some cytokeratins in EBC of smokers, demonstrating that...
the exposure to toxic chemicals may cause structural alterations in lung, resulting in a diagnostic signature [217].

8.4 Plasma and serum
Sampling blood and its derived plasma or serum from a patient is among the least invasive techniques in medical practice [150], therefore extensive research projects are now being conducted to evaluate in detail the diagnostic value of multiple biomarkers measured in the peripheral blood.

Quantitative measurement by real-time PCR of free DNA in serum/plasma might be considered a highly promising and very cost-effective biomarker for lung cancer screening and detection. The authors and others reported that cancer patients have a higher level of circulating DNA than non-cancer cases [218,219]. Sozzi et al. applied quantitative real-time PCR of the hTERT gene to measure cancer-derived DNA and to discriminate lung cancer patients from healthy smoking and non-smoking controls. Very satisfactory sensitivity (90%) and specificity (86%) were demonstrated, with a positive predictive value of 90% and negative predictive value of 90%. Sixty-nine of 100 evaluated NSCLC patients and 2 of 100 controls showed elevated circulating DNA levels, which were 8 times higher in the plasma of patients than controls (24.3 versus 3.1 ng/ml) [220]. The authors’ report confirmed literature data, establishing a lower cutoff value (2 ng/ml) and defining that more elevated concentrations of DNA are correlated with the status of the disease [218].

Genetic and epigenetic alterations can be detected in circulating DNA [221,222]. There are multiple reports of DNA methylation in blood but the data referring to sensitivity and specificity are discordant. Recently a report examining the methylation of CDH13, CDKN2A/p16, FHIT, RARB, RASSF1A and ZMYND10 (BLU) in which methylation of any 2 loci in plasma was considered cancer positive [228], and results were enriched by Potti et al., who added useful information about inverse sensitivity to docetaxel-etoposide and docetaxel-PI3-kinase inhibitor, candidate them as second-line therapies [229]. Using the same data, Hsu et al. found inverse correlation between cisplatin and docetaxel, pemetrexed and abraxane [230], whereas Gemma et al. did not demonstrate correlation of gemcitabine with seven commonly used drugs in lung cancer therapy, suggesting it for combinations therapy regimens [231]. Despite the gene expression profile showing that information collected from cell lines could discover predictors of drug sensitivity in lung cancer patients, meaning that they can represent more accurately the complexity of cellular biology in both healthy and cancer cells, all these studies, for the known above-mentioned limitations, need further validation in larger prospective clinical trials.

Apart from the plasma DNA studies, interesting reports concerning tumour-related circulating mRNA and miRNA in the peripheral blood of lung cancer patients should also be mentioned. Koperski et al. showed the possibility of extraction of 5T4 mRNA, a glycoprotein frequently overexpressed in epithelial malignancies, from the serum of patients with breast and lung cancer. This potentially permits the identification of patients who might benefit from 5T4-direct therapy [225]. Lodes et al. showed that sufficient miRNAs are present in 1 ml of serum to detect miRNA expression patterns, without the need for amplification techniques. Besides, the patterns obtained seem to be able to discriminate between normal and cancer patients samples [226]. The stability of miRNAs in serum makes them potentially new clinical biomarkers for diagnosis and prognosis [102,227]. The specific description of expression profiles of miRNAs in lung cancer has been described in section 5.2.5.1 dedicated to miRNA.

9. Biomarkers and therapy
Proper selection of patients and therapies is key in lung cancer treatment because a remarkable percentage of patients do not respond to commonly used drugs. Furthermore, developing resistance to chemotherapy is a significant problem in patients with lung cancer, contributing to disease relapse, progression and death: tailoring therapies to every subject represents an attractive aim that would avoid useless toxic effects reaching the higher therapeutic responses. The biomarkers underlying the multistep phases of carcinogenesis promise to have a potential role as either a prognostic factor (i.e., determining the effect of tumour on the patient) or a predictive factor (determining the effect of treatments on the tumour). Identifying the subset of patients who could benefit from therapies even in the early stages of lung cancer and at the same time predicting drug responsiveness of such patients seem to be the major key points to reduce lung cancer mortality independently from diagnostic anticipation. Some of the prognostic and predictive factors in NSCLC, as some of the current tailored-targeted therapies available, are summarised in Tables 3 and 4.

Initial studies using consolidated technologies such as IHC have led to the discovery of the value of EGFR mutation in predicting response to TKi therapy, whose clinical relevance has been described above. Since 2001 attention has focused even more on gene expression profile using DNA-microarray. A first study conducted by Staunton et al. in a collection of 60 human cancer cell lines demonstrated that transcriptional profiling was able to predict drug sensitivity in an independent test set [228], and results were enriched by Potti et al., who added useful information about inverse sensitivity to docetaxel-etoposide and docetaxel-PI3-kinase inhibitor, candidate them as second-line therapies [229]. Using the same data, Hsu et al. found inverse correlation between cisplatin and docetaxel, pemetrexed and abraxane [230], whereas Gemma et al. did not demonstrate correlation of gemcitabine with seven commonly used drugs in lung cancer therapy, suggesting it for combinations therapy regimens [231]. Despite the gene expression profile showing that information collected from cell lines could discover predictors of drug sensitivity in lung cancer patients, meaning that they can represent more accurately the complexity of cellular biology in both healthy and cancer cells, all these studies, for the known above-mentioned limitations, need further validation in larger prospective clinical trials.

At least two considerations should be pointed out. First, no data are actually available about change of molecular profile of a tumour at the first diagnosis and after surgery or chemotherapy: this implies that repeated biopsy could be necessary in the presence of a metastasis or local recurrence to define the validity of specific drug targets. Second, technical issues should be considered in the calculation of cost/effectiveness
of a biomarker because most of the sophisticated molecular techniques reported and described (apart from immunohistochemistry) are not available in a larger number of routine clinical laboratories [232].

10. Conclusion

Despite many efforts, lung cancer remains a fatal disease with poor prognosis. The disease typically presents as a highly aggressive neoplasm with frequent occurrence of lymph node and distant metastases. The high mortality is a result of late diagnosis. Most patients with lung cancer are still diagnosed at an advanced stage and at present no screening tests are available. Early diagnosis is crucial to decreasing lung cancer mortality, and personalised therapy based on genetic or functional markers may improve response to treatment.

Significant advances in our understanding of the molecular and genetic changes involved in lung carcinogenesis have provided the basis for many investigations of possible lung cancer biomarkers. Lung cancer biomarkers may allow us to identify populations who would benefit from computed tomography screening, may differentiate individuals with benign pulmonary nodules from those with early malignancies and may allow for personalisation of lung cancer treatment based on tumour characteristics.

Several biomarkers have been studied exhaustively. Despite some markers being promising none of them has been proved to be sufficiently useful for clinical use; the data are often retrospective, troubled by small size, a lack of reproducibility or inconsistent controls between studies, thus confirmation with prospective and larger studies is needed. Moreover, it remains unclear not only which genomic predictor of prognosis is the best but also whether specific genes or entire signatures are most important in predicting outcome. The future lies probably in a panel of biomarkers instead of individual assays.

So, at present, there is no consensus on the proven value of the clinical use of biomarkers for diagnosis, staging, prognosis and monitoring for disease relapse or treatment response; however, there has been considerable progress in elucidating the role of these biomarkers in activating key intracellular signalling pathways, and this may lead to improved outcomes.

11. Expert opinion

As the literature shows, researchers are engaged to identify biomarkers that have the characteristics of ideal biomarkers. However, most studies lack consistency of cases and methodological correctness to reach valid conclusions, therefore results still remain controversial, and although some markers seem...
to be promising it could easily be concluded that so far no useful biomarkers for clinical use in lung cancer exist. As lung cancer results from the acquisition of multiple somatic mutations, it would be unlikely that a single gene expression pattern could effectively predict the clinical behaviour of the disease.

Probably, in parallel to an increase and use of new technologies and research of new biomarkers, it would be useful to concentrate efforts on identifying the best panel between the most studied markers, or to integrate various forms of data, including clinical variables and multiple gene expression profiles that would allow solid predictive models to be built for the single patient. Combining them with imaging technologies could improve cumulative accuracy further.

Note that actually only a limited number of cancer biomarkers have been approved by the FDA and are now in use for cancer detection, monitoring, prognosis or therapy selection; so far no cancer biomarkers have been approved for early diagnosis. Although lung cancer is a deadly disease, none of the lung cancer biomarkers is approved by the FDA at present. Nevertheless, the working field is not as rewarding in terms of results. Every effort made in improving technologies, methods of validation, approaches for development, in better understanding of the disease, which could bring a small contribution to improvement of clinical management and prognosis of these patients, has to be taken into account.

Validation by clinical trials in large cohorts of patients, high sensitivity and specificity technology platforms, as well as stringent statistic/bioinformatics tools are necessary before new molecules can be translated into the clinic as reliable biomarkers.

Precise sample handling (collection procedures, type of containers used, preservatives added to the sample, stability, processing and storage conditions) is essential for analytical reliability and reproducibility. Concerning this matter, two reviews describe the standard operating procedures for biological sample collection and processing for molecular epidemiological studies [233,234]. A certification regarding procedures for sample collection and handling is also available from the Clinical Laboratory Improvement Amendments [235].

Finally, for biomarker development, intensive verification and validation processes are necessary. The Early Detection Research Network (National Cancer Institute, Division of Cancer Prevention [236]) has proposed a stepwise method for evaluating new biomarkers, and for identifying people at risk [237]. Moreover, in 2005 the Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics published the reporting recommendations for tumour marker prognostic studies (REMARK) guidelines. The data that were collected during this review [238] clearly highlight the multiple study design factors that have contributed to the inconsistencies and contradictions. A rigorous implementation of such a multiphase approach to studies on prognostic markers can improve the chances of identifying true prognostic markers that may be reliably applied in clinical practice.

In the future, certainly, combinations of biomarkers will lead to non-invasive, cost-effective screening tools to improve outcomes, however many of the above-mentioned issues require serious discussion by the scientific community.

Acknowledgements

The authors are grateful to S De Franco and R Iori for their technical-scientific support in the bibliographical research. Many thanks also to G Mazzi for English revision. Moreover, the authors wish to thank the Executive Council of Vittorio Lodini Association for Surgery Research.

Declaration of interest

The authors state no conflict of interest and have received no payment in preparation of this manuscript.
Bibliography


New biomarkers for lung cancer


64. Ratschiller D, Heighway J, Gugger M, et al. Cyclin D1 overexpression in bronchial epithelia of patients with lung cancer is associated with smoking and...


New biomarkers for lung cancer


129. Tainsky MA. Genomic and proteomic opportunities. Biochim Biophys Acta 2009;1796:176-93

130. Wright GL Jr. Two-dimensional acrylamide gel electrophoresis of


140. Cox J, Mann M. Is proteomics the new genomics? Cell 2007;130:395-98


New biomarkers for lung cancer


190. Fan TW, Lane AN, Higashi RM. The promise of metabolomics in cancer molecular therapeutics. Curr Opin Mol Ther 2004;6:584-92


205. Olaussen KA, Soria JC, Park YW, et al. Copy number gains on 5p15, 6p11-q11, 7p12, and 8q24 are rare in sputum from patients with NSCLC. Lung Cancer 2000;60:5954-8


New biomarkers for lung cancer


Affiliation
Massimiliano Paci†, Cristian Rapicetta & Sally Maromotti
†Author for correspondence
1Division of Thoracic Surgery, Azienda Santa Maria Nuova di Reggio Emilia, Viale Risorgimento 80, 42100 Reggio Emilia, Italy
Tel: +39 0522 296029; Fax: +39 0522 296191; E-mail: paci.massimiliano@asmn.re.it
2Department of Clinical Pathology, Azienda Ospedaliera Santa Maria Nuova, 42100 Reggio Emilia, Italy