Prognosis of Different Types of Non-Small Cell Lung Cancer Progression: Current State and Perspectives

Anastasia A. Schegoleva a, Anna A. Khozyainova a, Anton A. Fedorov a, Tatiana S. Geraschenko a, Evgeny O. Rodionov b, Evgeny B. Topolnitsky c, Nikolay A. Shefer d, Olga V. Pankova a, Anastasia A. Durova f, Marina V. Zavyalova e,f, Vladimir M. Perelmuter e, Evgeny V. Denisov a

aLaboratory of Cancer Progression Biology, Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences, Tomsk, Russia, bDepartment of Thoracic Oncology, Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences, Tomsk, Russia, cDepartment of Surgery with a Course of Mobilization Training and Medicine of Disasters, Siberian State Medical University, Tomsk, Russia, dSurgery Department, Tomsk Regional Oncology Center, Tomsk, Russia, eDepartment of General and Molecular Pathology, Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences, Tomsk, Russia, fDepartment of Pathological Anatomy, Siberian State Medical University, Tomsk, Russia

Key Words
Non-small cell lung cancer • Progression • Recurrence • Metastasis • Marker

Abstract
Despite advances in diagnostics and therapy of non-small cell lung cancer (NSCLC), the problem of prognosis and prevention of tumor progression is still highly important. Even if NSCLC is diagnosed in the early stages, almost a quarter of patients develop relapse; most of them die from recurrent disease. A large number of different markers have been proposed to predict the risk of NSCLC progression; however, none of them are used in clinical practice. It is obvious that this situation is related to the economic and methodological complexity of the proposed markers and/or their insufficient efficiency due to a lack of effective study models and tumor heterogeneity. Another reason may be that potential markers are developed for NSCLC progression in general, which is represented by at least four pathogenetically-distinct processes: synchronous lymph node metastasis, local, regional, and distant recurrence. In this review, we summarize data from published literature on clinicopathological, genetic, and molecular factors associated with different types of NSCLC progression and emphasize challenges and approaches to developing prognostic factors. In conclusion, we highlight the importance of further studies to reveal molecular mechanisms of NSCLC progression and the need for differential analysis of markers of local, regional, and distant recurrences for disease prognosis.
Introduction

Despite significant advances in diagnostics and therapy of non-small cell lung cancer (NSCLC), there is still no effective treatment approaches to prevent tumor progression. Even if NSCLC is diagnosed at stages I and II, almost a quarter of patients develop postoperative relapse; most of them die from recurrent disease [1, 2]. This issue is complicated by the fact that in the vast majority of patients relapses are inoperable [3]. As a result, the 5-year relative survival rate of patients with advanced NSCLC is only 6.1%. However, in patients with regional involvement and localized NSCLC, the 5-year relative survival rate is 34.5% and 61.4%, respectively, which indicate the importance of early diagnosis, relapse prognosis, and the development of approaches to prevent the recurrent disease [4].

Many studies described potential markers of NSCLC recurrence including the expression of genes, proteins, and non-coding RNAs, as well as genetic and epigenetic alterations, and circulating molecules. For example, a signature of miR-210, miR-214, and miR-15a has been proposed to predict the probability of lung adenocarcinoma (LUAD) metastasis into the brain (90.4% accuracy) [5]. In another study, a panel of CSF1, EGFR, and CA9 (CA-IX) genes has been developed to predict the risk of distant recurrence in patients with lung squamous cell carcinoma (LUSC) (85% sensitivity and 60% specificity) [6]. Some researches have shown an association of somatic mutations in the ALK, EGFR, and KRAS genes with the risk of NSCLC recurrence [7-13].

Despite a large number of studies, the proposed markers and prognostic signatures are mainly applicable to LUAD and significantly less often to LUSC. Moreover, they have a number of significant limitations, such as insufficient predictive efficiency, economic and methodological complexity of analysis. Because of this, none of the markers are approved to be used in clinical practice. Among other reasons, this may be due to the heterogeneous nature of NSCLC progression that is represented by lymph node metastasis synchronous with the primary tumor, as well as local, regional, and distant recurrence. In present, there is some confusion in the literature regarding definitions of local, regional, and distant recurrence of NSCLC. In this review, we settled on the following definitions. Local recurrence means that cancer reappears in the same lung or at the bronchial stump [14-18]. Regional recurrence is identified as a recurring disease in the ipsilateral hilar, mediastinal, and supraclavicular lymph nodes detected during follow-up after treatment [9, 14, 19]. All other sites of recurrence are referred to as distant recurrence or metastasis [14]. It is interesting to note that metastasis to the pleura is referred to as a form of distant recurrence [9, 10]. However, there is an opinion that pleural metastasis can be an independent type of cancer progression [19]. The probability of distinct types of NSCLC progression and their contribution to the disease outcome are different. It was found that in surgically resected NSCLC patients are more likely to develop distant relapses than local (67.8-79.5% vs. 20.5%-28.8%). The simultaneous development of local and distant relapses is quite rare (3.4-6.4%) [3]. The most favorable outcome is observed in patients with locoregional relapses [20].

Previously, we showed that the probability of locoregional and distant recurrence of NSCLC is associated with the presence of premalignant lesions in the small bronchi distant from the primary tumor. The isolated basal cell hyperplasia (BCH) was related to increased distant metastasis and a decreased locoregional recurrence [21]. On the contrary, the co-presence of BCH and squamous metaplasia (SM) was associated with a high probability of locoregional recurrence and a low risk of distant metastasis [22, 23]. Probably, epithelial-stromal interactions in the small bronchi may reflect constitutive features of the inflammation and can be indirectly related to tumor-stromal interactions and cancer progression. In other words, some individuals are predisposed to locoregional recurrence, while others are predisposed to distant metastasis.

Thus, the biology and clinical significance of NSCLC progression significantly vary, and it is necessary to develop prognostic markers specific to synchronous lymph node metastasis, local, regional, and distant recurrence. Unfortunately, most studies reported the markers that are associated with the NSCLC recurrence in general. The aim of the current review is
to systematize the available literature data on clinicopathological parameters, germline and somatic genetic variations, gene and protein expression, non-coding RNAs, and circulating molecules that can be potential markers of synchronous lymph node metastasis, as well as local, regional, and distant recurrence of NSCLC. This review also emphasizes the main challenges and approaches to developing prognostic factors of NSCLC and the importance of further studies of molecular mechanisms of different types of NSCLC progression.

**Synchronous lymph node metastasis**

*Clinicopathological markers*

Surgical pathology has an important role in the diagnosis and treatment of lung cancer [24]. Despite the remarkable progress in molecular pathology related mainly to the development of targeted therapies specific for histological types of NSCLC, routine clinical and morphological parameters still remain of significant prognostic value. In particular, it is known that non-upper lobe located NSCLC and micropapillary-predominant LUAD have an increased lymph node metastasis [25]. In addition, tumor size correlates with lymph node involvement [26-29]. For example, tumor size more than 3 cm and central localization of the tumor are associated with lymph node metastasis [30]. Lymph node lesions are also related to tumor grade and pleural involvement [31].

*Genetic markers*

*Germline variations.* Single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) form the genetic landscape of the individual that controls cellular processes and determines predisposition to different diseases [32]. Many studies demonstrated the contribution of certain SNPs and CNVs to an increased risk of NSCLC [33]. However, only a few studies showed that the probability of lymph node metastasis of NSCLC can be determined by germline variations. For example, a high risk of lymph node involvement is associated with the rs9876 AA genotype of the *TNS3* gene [34]. In smokers, lymph node metastasis is related to the *MMP1-MMP3* 1G/5A haplotype [35].

*Somatic alterations.* Transformation of normal cells to tumor cells occurs due to acquiring a series of mutations over time. Tumors can harbor tens to hundreds and thousands of different mutations some of whom are critically important for carcinogenesis and called “drivers” whereas other alterations, “passengers”, are considered to be neutral [36]. The mutational landscape of NSCLC is well studied and driver genes have been identified to contribute to carcinogenesis in the respiratory epithelium. LUAD is believed to be associated mainly with alterations in the *KRAS, EGFR, BRAF, and MET* genes, [37] whereas the development of lung LUSC is predominantly related to changes in the *TP53, CDKN2A, PTEN, PIK3CA, KEAP1, MLL2, HLA-A, and NFE2L2* genes [38]. Single studies showed that driver and some other genes can be involved in lymph node metastasis of NSCLC. LUAD patients harboring both a known driver (*KRAS, EGFR, NRAS,* etc.) or candidate mutation, and *TP53* mutations have high rates of lymph node metastasis [39]. ALK-rearranged NSCLC tended to show more frequent lymph node involvement, even in patients with low T stage [11].

*Molecular markers*

*Gene and protein expression.* Genes, particularly encoded proteins, are the main players in cellular processes in normal physiology and cancer. There is a lot of data demonstrating the association between the expression of genes and proteins and synchronous lymph node metastasis of NSCLC. In general, lymph node involvement of NSCLC is associated with overexpression of *hMTH1, SPD, ITGA11, COL11A1, MACC1, S100A4, MT2A, CXCR4* and *CPSF3* genes, as well as a high level of PRMT5, Hsp90-beta, annexin A1 (ANXA1), MACC1, integrin αvβ5 and TRIM44 proteins [40-49] (Table 1). In LUAD, a high incidence of lymph node metastases is related to overexpression of RACK1, CD147 and underexpression of RASAL2 proteins [50, 51].
**Table 1.** Molecular markers associated with synchronous lymph node metastasis of NSCLC. DFS, disease-free survival; LncRNAs, long non-coding RNAs; LUAD, adenocarcinoma; LUSC, lung squamous cell carcinoma; NSCLC, non-small cell lung cancer; PFS, progression-free survival; OS, overall survival; RFS, recurrence/relapse-free survival; “-”, not available. The numbers in brackets indicate the size of validation cohorts.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Level</th>
<th>Survival</th>
<th>Cohort</th>
<th>Histology</th>
<th>Stage</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hMTH1, SPO, ITGA11, COL1A1</td>
<td>Up</td>
<td>OS</td>
<td>70</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[40]</td>
</tr>
<tr>
<td>MACC1</td>
<td>Up</td>
<td>DFS, OS</td>
<td>180</td>
<td>NSCLC</td>
<td>I-III</td>
<td>[49]</td>
</tr>
<tr>
<td>MT2</td>
<td>Up</td>
<td>DFS</td>
<td>25 (51)</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[47]</td>
</tr>
<tr>
<td>CPSF1</td>
<td>Down</td>
<td>OS, PFS</td>
<td>706</td>
<td>LUAD</td>
<td>-</td>
<td>[43]</td>
</tr>
<tr>
<td>CXCR4</td>
<td>Up</td>
<td>OS</td>
<td>2037 meta-analysis</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[41]</td>
</tr>
<tr>
<td>S100A4</td>
<td>Up</td>
<td>-</td>
<td>41</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[44]</td>
</tr>
<tr>
<td>RACK1, CD147</td>
<td>Up</td>
<td>OS</td>
<td>180</td>
<td>LUAD</td>
<td>I-II</td>
<td>[50]</td>
</tr>
<tr>
<td>Hsp90-beta, ANXA1</td>
<td>Up</td>
<td>DFS</td>
<td>96</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[45]</td>
</tr>
<tr>
<td><strong>Proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>integrin avβ5</td>
<td>Up</td>
<td>OS</td>
<td>147</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[46]</td>
</tr>
<tr>
<td>MACC1</td>
<td>Up</td>
<td>DFS, OS</td>
<td>180</td>
<td>NSCLC</td>
<td>I-III</td>
<td>[49]</td>
</tr>
<tr>
<td>PRMT5</td>
<td>Up</td>
<td>OS</td>
<td>85 (209)</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[42]</td>
</tr>
<tr>
<td>TRIM44</td>
<td>Up</td>
<td>DFS, OS</td>
<td>331</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[48]</td>
</tr>
<tr>
<td><strong>miRNAs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>miR-451</td>
<td>Down</td>
<td>OS</td>
<td>23</td>
<td>NSCLC</td>
<td>-</td>
<td>[53]</td>
</tr>
<tr>
<td>miR-26a</td>
<td>Up</td>
<td>-</td>
<td>10</td>
<td>NSCLC</td>
<td>-</td>
<td>[62]</td>
</tr>
<tr>
<td>miR-449a</td>
<td>Down</td>
<td>OS</td>
<td>84</td>
<td>NSCLC</td>
<td>I-III</td>
<td>[54]</td>
</tr>
<tr>
<td>miR-486-5p</td>
<td>Down</td>
<td>-</td>
<td>76 (33)</td>
<td>NSCLC</td>
<td>I-III</td>
<td>[55]</td>
</tr>
<tr>
<td>miR-9</td>
<td>Up</td>
<td>PFS, OS</td>
<td>116</td>
<td>NSCLC</td>
<td>I-III</td>
<td>[63]</td>
</tr>
<tr>
<td>miR-944</td>
<td>Up</td>
<td>-</td>
<td>60 (40)</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[64]</td>
</tr>
<tr>
<td>miR-206</td>
<td>Down</td>
<td>-</td>
<td>35</td>
<td>LUAD</td>
<td>I-II</td>
<td>[56]</td>
</tr>
<tr>
<td>miR-139</td>
<td>Down</td>
<td>-</td>
<td>75</td>
<td>NSCLC</td>
<td>I-III</td>
<td>[57]</td>
</tr>
<tr>
<td>miR-187-5p</td>
<td>Down</td>
<td>OS</td>
<td>64</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[58]</td>
</tr>
<tr>
<td>miR-212</td>
<td>Down</td>
<td>OS</td>
<td>115</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[59]</td>
</tr>
<tr>
<td>miR-138</td>
<td>Down</td>
<td>OS</td>
<td>45</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[60]</td>
</tr>
<tr>
<td>miR-422a</td>
<td>Up</td>
<td>-</td>
<td>5</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[65]</td>
</tr>
<tr>
<td>miR-101</td>
<td>Down</td>
<td>OS</td>
<td>20</td>
<td>NSCLC</td>
<td>-</td>
<td>[61]</td>
</tr>
<tr>
<td>miR-1290</td>
<td>Up</td>
<td>OS</td>
<td>33</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[66]</td>
</tr>
<tr>
<td><strong>LncRNAs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAS5-AS1</td>
<td>Down</td>
<td>OS</td>
<td>50</td>
<td>NSCLC</td>
<td>-</td>
<td>[68]</td>
</tr>
<tr>
<td>BANCR</td>
<td>Down</td>
<td>PFS, OS</td>
<td>113</td>
<td>NSCLC</td>
<td>I-III</td>
<td>[74]</td>
</tr>
<tr>
<td>CAT2</td>
<td>Up</td>
<td>-</td>
<td>57</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[73]</td>
</tr>
<tr>
<td>ANRIL</td>
<td>Up</td>
<td>OS</td>
<td>87</td>
<td>NSCLC</td>
<td>I-III</td>
<td>[56]</td>
</tr>
<tr>
<td>GASS-AS1</td>
<td>Down</td>
<td>-</td>
<td>48</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[75]</td>
</tr>
<tr>
<td>HOTAIR</td>
<td>Up</td>
<td>OS</td>
<td>42</td>
<td>NSCLC</td>
<td>II-V</td>
<td>[72]</td>
</tr>
<tr>
<td><strong>Circulating molecules</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>miR-1290</td>
<td>Up</td>
<td>OS</td>
<td>73</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[66]</td>
</tr>
<tr>
<td>lct-7c, miR-152</td>
<td>Down</td>
<td>-</td>
<td>120 (360)</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[78]</td>
</tr>
</tbody>
</table>
| **Non-coding RNAs.** Non-coding RNAs are crucial regulators of cellular processes and represented by three main types: translation-related RNAs (rRNAs and tRNAs), small non-coding RNAs (siRNA, miRNAs, and piRNAs), and long non-coding RNAs [52]. Many studies showed the association of miRNAs and long non-coding RNAs with synchronous lymph node metastasis of NSCLC. In general, lymph node involvement is associated with underexpression of miR-451, miR-449a, miR-486-5p, miR-206, miR-139, miR-187-5p, miR-212, miR-138, and miR-101 and overexpression of miR-9, miR-26a, miR-422a, miR-944 and miR-1290 (Table 1) [53-66]. miR-449, miR-422a, miR-187-5p, and miR-101 microRNAs are known to promote apoptosis, whereas miR-451, miR-486-5p, miR-187-5p, and miR-101 inhibit proliferation of lung cancer cells [53-55, 58, 61, 67]. Other miRNAs, such as miR-212, miR-486-5p,
miR-206, miR-139, miR-187-5p, miR-212, miR-138, miR-1290, and miR-101, suppress epithelial–mesenchymal transition (EMT) and decrease migration, invasion, and metastasis of lung cancer cells [55-60, 68-70]. In contrast, miR-944, miR-9, and miR-26a increase migration and invasion of lung cancer cells and promote metastasis [62, 64, 71]. Thus, miR-138, miR-486-5p, miR-212, miR-449a, miR-451, miR-206, miR-139, miR-101, and miR-187-5p work as tumor suppressors, whereas miR-944, miR-422a, miR-1290, miR-9, and miR-26a possess oncogenic functions.

Underexpression of long non-coding RNAs GAS5-AS1, GAS6-AS1, and BANCR and overexpression of CCAT2, ANRIL, and HOTAIR are associated with lymph node metastasis of NSCLC [56, 68, 72-76]. GAS5-AS1, GAS6-AS1, and BANCR are known to be involved in the suppression of EMT, migration, invasion, and metastasis of lung cancer cells, while HOTAIR, CCAT2, and ANRIL, on the contrary, induce cell proliferation, migration, and invasion [56, 68, 72-75].

Circulating molecules. Blood contains different molecules including circulating DNA, RNA, and proteins. They can be used for the diagnosis of tumors, prediction of treatment response, cancer prognosis, and detection of recurrence [77]. In NSCLC, high level of circulating miR-1290, miR-422a, mRNA of the MACC1 gene, and carcinoembryonic antigen (CEA), and low level of miRNA let-7c and miR-152 in blood plasma were found to be associated with lymph node metastasis (Table 1) [25, 26, 65, 66, 78, 79]. miR-1290, miR-422a, let-7c, and miR-152 are known to inhibit migration and invasion of lung cancer cells, while miR-422a induces apoptosis [67, 70, 80, 81].

Locoregional recurrence

Clinicopathological markers

Visceral pleural invasion (VPI), lymphovascular invasion (LVI), and blood vessel invasion (BVI) were shown to be the main risk factors for locoregional recurrence [82-84]. Besides VPI, in patients with early-stage NSCLC, locoregional recurrence is also related to wedge resection, microscopic positive surgical margin, lymphatic permeation, and large tumor size [18, 85]. Moreover, wedge resection is frequently associated with margins less than 1 cm and, thus, a high risk of locoregional recurrence [86]. Segmentectomy is also an independent risk factor for locoregional recurrence. Superior segmentectomies have significantly lower locoregional recurrence rates and excellent recurrence-free survival (RFS) comparable to lower lobectomies [15].

Locoregional relapses are also more often observed in NSCLC patients with the co-presence of BCH and SM in the small bronchi distant from the primary tumor [22, 23]. Spread through air spaces (STAS) is associated with an increased frequency of locoregional recurrence in patients with LUSC [87]. In LUAD, STAS-positive I stage patients, who underwent sublobar resection or limited resection comparing to lobectomy, have a high risk of locoregional recurrence regardless of margin-to-tumor ratio [16, 17, 88]. Among STAS-negative patients, locoregional recurrence is related to a margin-to-tumor ratio of less than 1 [88]. Interestingly, STAS was more frequently observed in LUAD patients with a cribriform component which is associated with increased risk of locoregional recurrence. Moreover, in patients with the cribriform component, STAS-positive cases had a significantly inferior outcome as compared to STAS-negative cases [89].

Genetic markers

The frequency of locoregional recurrence depends on the C609T polymorphism of the NQO1 gene. Relapse is more often observed in patients with the T/T genotype compared to C/C and C/T genotype carriers [90]. Locoregional recurrence is also more frequent in NSCLC patients with missense mutations in exons 18 and 21 and exon 19 deletion of the EGFR gene [8]. Interestingly, ALK rearrangements are associated with greater tendency to develop locoregional recurrence than EGFR and KRAS mutations [9].
Despite the presence of studies focusing on locoregional recurrence of NSCLC in general, we believe that the distinguishing local and regional relapse separately is needed due to their different nature and underlying mechanisms. Below, we provide the literature data regarding potential markers of local and regional recurrence.

**Local recurrence**

VPI, LVI, and STAS mentioned above in case of locoregional recurrence are also associated with local relapse [83, 91, 92]. Besides, local recurrence is related to other clinical and pathological parameters such as younger age, smoking index more than 20 pack-years, central tumor localization, squamous/large cell histology, stage > IA, sublobar resections, lobectomy with negative surgical margins, segmentectomies in the right upper lobe and of basal segments, tumor margin less than 1.0 cm, lower radiation dose, fewer medical comorbidities, and higher body mass index [15, 83, 92-95]. The recent study showed that local recurrence is also more often observed in NSCLC patients that have circulating tumor cells with small and irregular nuclei [96].

In LUAD patients, VEGFR-2 -906C>T and -271G>A variants are associated with tumor size and local recurrence. In particular, -906 CC and -271 GG genotypes significantly correlate with high frequency and early appearance of local relapse [97]. Local recurrence is also more often observed in NSCLC with NRF2/KEAP1 gene mutations compared to wild-type tumors [98] and high expression of MMP12 [99].

**Regional recurrence**

There are only a few studies regarding potential markers of regional recurrence of NSCLC. In the postoperative period, lymph node involvement is associated with intratumoral LVI (L1-status) [30]. Regional recurrence of NSCLC is also related to a low level of insulin-like growth factor binding protein (IGFBP5 and IGFBP7) in blood plasma [100]. In LUAD, regional recurrence is associated with overexpression of the CUB domain-containing protein (CDCP1) [101].

**Distant recurrence**

*Clinicopathological markers*

Non-squamous cell histology, pneumonectomy, and lymphatic invasion have been shown to be associated with distant metastasis in NSCLC patients [85, 95, 102]. Similar to locoregional recurrence, distant recurrence correlates to BVI, large tumor size, LVI, VPI, and STAS [16, 83-85, 87, 103-105]. In contrast to locoregional recurrence, distant metastases are associated with isolated BCH in the small bronchi distant from the tumor [21]. STAS is associated with a high risk of recurrence in LUSC patients undergoing lobectomy, but not sublobar resection; while in LUAD patients, STAS is a prognostic factor of recurrence in sublobar resection group [87]. Interestingly, another study reported that STAS is tightly linked to nodal and distant metastasis in LUAD patients, most of whom underwent lobectomy [106]. In addition, hematogenous metastases of LUAD are related to LVI and the cribriform component in the tumor [89, 107]. Similar to local recurrence, distant metastasis is more often observed in NSCLC patients that harbor circulating tumor cells with small and irregular nuclei [96].
Genetic markers

Germline variations. A high risk of distant recurrence is specific for NSCLC patients-carriers of the RETN rs3219175 AG and AG+AA genotypes [108]. Pleural metastasis is more frequently observed in LUAD patients with the EGFR rs712829 -216G/T and -216T/T genotypes [109]. Bone metastasis is associated with the CD44 rs187115 and the OPN 443 C/T polymorphisms, whereas the TGFBI rs1800469, ATG10 rs10036653, ATG12 rs26532, and ATG16L1 rs2241880 variants contribute to brain metastasis [110-113]. For example, a high risk of brain metastases was observed in rs26532 AC/CC carriers, while the rs10036653 AT/TT and rs2241880 AG/GG genotypes had a protective effect [111]. Moreover, the transfection of NSCLC cells with ATG16L1 rs2241880 -300T (threonine) encoded by A allele increases their metastatic potential to the brain in a mouse model [111].

Somatic alterations. The data regarding the association of genetic alterations with distant recurrence of NSCLC are contradictory. Mutations in the EGFR gene are associated with a high risk of distant recurrence in general [10] and metastasis to the brain [12, 13], liver [7], and pleura [9]. ALK gene rearrangements are associated with metastasis to the pleura, liver, and pericardium [7], as well as to the brain similar to the ROS1 and RET gene alterations [13, 114]. Mutations in the KRAS gene tend to be frequently detected in NSCLC patients with brain metastases [9]. At the same time, no association was found between metastases to the brain and ALK [7, 9], EGFR [7, 9, 115], and KRAS mutations [7, 115], liver metastases and ALK/EGFR alterations [9], and pleural metastases and aberrations in the EGFR [7] and ALK [9] genes.

Molecular markers

Gene and protein expression. Different genes and proteins whose expression is associated with the risk of distant recurrence and organotropic metastasis have been described (Table 2). Distant metastases are more frequently detected in NSCLC patients with overexpression of thymosin β4 (TMSB4X), eukaryotic translation initiation factor 4A1 (eIF4A1), chemokine receptors CXCR4 and CXCR7, EMT-regulators MC2S1 and SPOCK1, cyclin E (CCNE1), α smooth muscle actin (ACTA2), tyrosine kinase receptor (IR, NTRK1, EGFR, ERBB2, ERBB3, PDGFR-β, FGFR1, and LTK), and underexpression of ribosomal S6 kinase 1 (RSK1) and huntingtin-interacting protein 1 (HIP1) [41, 116-124]. NSCLC metastasis to the brain is associated with the expression of the GAP43 gene probably through the promotion of migration of cancer cells by Rac1 activation and prevention of F-actin depolymerization [125].

Several multigenic signatures have been proposed to predict the risk of NSCLC distant metastasis. Skrzypski and colleagues developed a panel of CSF1, EGFR, and CA9 (CA-IX) genes to predict the probability of LUSC metastasis (sensitivity 85%, specificity 60%) [6]. Grinberg-Rashi et al. showed that the expression of CDH2, KIFC1, and FALZ genes is associated with NSCLC metastasis to the brain [126]. Fregni and colleagues described a signature of 4 genes of mesenchymal stem cells (GREM1, LOXL2, ADAMTS12, and ITGA11), whose expression is associated with NSCLC metastasis [127].

Non-coding RNAs. A large number of microRNAs and long non-coding RNAs, whose expression correlates with distant recurrence of NSCLC and organotropic metastasis, have been identified (Table 2). In LUSC, increased risk of distant metastases is associated with expression of miR-10b, miR-662, miR-502-3p, miR-192*, miR-192, and miR-128 [128]. It is known that these microRNAs are involved in the regulation of proliferation (miR-10b and miR-128), apoptosis (miR-10b and miR-128), migration, and invasion of lung cancer cells (miR-10b, miR-192, and miR-662) as well as angiogenesis and lymphangiogenesis (miR-128) [129-133].

Brain metastasis of NSCLC is associated with low level of miR-590, miR-375, miR-145-5p, and miR-1280 and high expression of miR-219-2-3p, miR-219-5p, miR-124, miR-9*, miR-128, and miR-338-3p [134-136]. In LUAD, brain metastasis is also related to miR-214, miR-145, and miR-23a underexpression and overexpression of miR-9*, miR-1471, miR-718, miR-3656, miR-720, miR-423-5p, miR-184, and miR-197 [137-139]. It is known that miR-145,
miR-590, miR-718, miR-124, miR-338-3p, and miR-184 [136, 140-144] inhibit proliferation, migration, and invasion of lung cancer cells, while miR-423-5p induces tumor progression [138]. miR-23a and miR-145-5p regulate EMT, miR-9* controls cell cycle, and miR-718 and miR-197 induce apoptosis of lung cancer cells [144-148]. The contribution of miR-375 to the NSCLC prognosis is probably related to VEGF and MMP9 overexpression [134].

Bone metastasis correlates with overexpression of miR-21 and hsv2-miR-H9-5p [149, 150]. miR-21 is known to activate proliferation and inhibit apoptosis through suppression of cytochrome c oxidase (COX)-assembly protein COX-19, and hsv2-miR-H9-5p increases survival, migration, and invasion of lung cancer cells [149, 150].

The expression of lncRNA HOTAIR is significantly higher in brain metastases than in primary NSCLC and associated with low postoperative disease-free survival [151]. It is known that HOTAIR promotes proliferation, survival, invasion, metastasis, and drug resistance of lung cancer cells [152]. Overexpression of lncRNAs MALAT1 and AWPPH is associated with NSCLC metastasis [117, 153]. There is also data that the MALAT1 level is significantly higher in brain metastases of NSCLC [154]. Probably, MALAT1 promotes lung cancer brain metastasis by inducing EMT [154]. In addition, MALAT1 expression displays the strongest

<table>
<thead>
<tr>
<th>Marker</th>
<th>Level</th>
<th>Survival</th>
<th>Cohort</th>
<th>Histology</th>
<th>Stage</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCNE1</td>
<td>Up</td>
<td>OS</td>
<td>70</td>
<td>NSCLC</td>
<td>I-II</td>
<td>[123]</td>
</tr>
<tr>
<td>TMBS4X, eIF4A1</td>
<td>Up</td>
<td>OS</td>
<td>70</td>
<td>NSCLC</td>
<td>I</td>
<td>[117]</td>
</tr>
<tr>
<td>IR, NTRK1, EGFR, ERBB2, ERBB3, PDGFR-β, FGFR1, LTK</td>
<td>Up</td>
<td>OS</td>
<td>70</td>
<td>NSCLC</td>
<td>I-III</td>
<td>[122]</td>
</tr>
<tr>
<td>CXCR4, CXCR7</td>
<td>Up</td>
<td>DFS</td>
<td>127</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[116]</td>
</tr>
<tr>
<td>SPOCK1</td>
<td>Up</td>
<td>DFS</td>
<td>40</td>
<td>NSCLC</td>
<td>-</td>
<td>[121]</td>
</tr>
<tr>
<td>MCPS1</td>
<td>Up</td>
<td>-</td>
<td>25</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[120]</td>
</tr>
<tr>
<td>GAP43</td>
<td>Up</td>
<td>OS, PFS</td>
<td>70</td>
<td>NSCLC</td>
<td>I-II</td>
<td>[125]</td>
</tr>
<tr>
<td>RSK1</td>
<td>Down</td>
<td>OS</td>
<td>100</td>
<td>LUAD</td>
<td>I-II</td>
<td>[118]</td>
</tr>
<tr>
<td>ACTA2</td>
<td>Up</td>
<td>MFS</td>
<td>263</td>
<td>LUAD</td>
<td>I-V</td>
<td>[119]</td>
</tr>
<tr>
<td>HIP1</td>
<td>Down</td>
<td>DFS</td>
<td>121</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[124]</td>
</tr>
<tr>
<td>miR-9*, 1471, 718, 3656, 720</td>
<td>Up</td>
<td>-</td>
<td>8 (43)</td>
<td>LUAD</td>
<td>II-III</td>
<td>[137]</td>
</tr>
<tr>
<td>miR-214, 145, 23a</td>
<td>Down</td>
<td>OS</td>
<td>18 (20)</td>
<td>NSCLC</td>
<td>-</td>
<td>[150]</td>
</tr>
<tr>
<td>miR-10b, 662, 502-3p, 192*, 192, 128</td>
<td>Down</td>
<td>OS</td>
<td>17 (54)</td>
<td>NSCLC</td>
<td>-</td>
<td>[139]</td>
</tr>
<tr>
<td>miR-423-5p, 1280</td>
<td>Up</td>
<td>-</td>
<td>29</td>
<td>NSCLC</td>
<td>-</td>
<td>[135]</td>
</tr>
<tr>
<td>LncRNAs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALAT1</td>
<td>Up</td>
<td>OS</td>
<td>70</td>
<td>NSCLC</td>
<td>I</td>
<td>[117]</td>
</tr>
<tr>
<td>HOTAIR</td>
<td>Up</td>
<td>DFS</td>
<td>77</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[151]</td>
</tr>
<tr>
<td>AWPPH</td>
<td>Up</td>
<td>-</td>
<td>128</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[153]</td>
</tr>
<tr>
<td>Circulating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEA</td>
<td>Up</td>
<td>OS</td>
<td>293</td>
<td>NSCLC</td>
<td>III-IV</td>
<td>[158]</td>
</tr>
<tr>
<td>miR-375</td>
<td>Down</td>
<td>OS</td>
<td>164 (53)</td>
<td>NSCLC</td>
<td>I-V</td>
<td>[157]</td>
</tr>
<tr>
<td>miR-139-5p</td>
<td>Down</td>
<td>OS</td>
<td>50</td>
<td>LUAD</td>
<td>IV</td>
<td>[160]</td>
</tr>
</tbody>
</table>
association with genes involved in cancer cell growth, movement, proliferation, signaling, and immune regulation [155]. AWPPH overexpression is known to upregulate TGF-β1 and promote NSCLC cell migration and invasion [153].

Based on microRNA profiling data, two signatures were proposed to predict the probability of LUAD metastasis to the brain: 1. miR-450b-3p, miR-29c, miR-145, miR-148a, miR-1, miR-30d, miR-187, miR-218, miR-708, and miR-375 [156] and 2. miR-210, miR-214, and miR-15a [5]. Also, Skrzypski and coauthors developed a signature of miR-10b, miR-662, miR-502-3p, miR-192*, miR-192, and miR-128 to predict LUSC metastasis risk and distant recurrence-free survival (RFS) [128].

Circulating molecules. Several serum proteins and circulating microRNAs have been shown to be associated with distant recurrence of NSCLC (Table 2). Metastases were more often detected in patients with a low level of miR-375 in blood plasma [157]. The high content of CEA in the plasma of NSCLC patients is associated with brain metastasis [158, 159]. Probably, CEA-positive tumor cells could bond to brain vasculature, favoring central nervous system metastasis, similar to leukocyte transendothelial arrest and migration through the blood-brain barrier [158]. The target blockade of CEA with antibodies inhibits the adhesion, migration, and invasion of several tumor cell lines in vitro and in vivo [158]. Bone metastasis is associated with a low plasma level of miR-139-5p in LUAD patients [160]. miR-139-5p is known to inhibit lung cancer cell invasion [57].

Differences in the markers of NSCLC progression types

It is not surprising that the different types of NSCLC progression are distinct in the biology and clinical significance. However, in most studies, they, especially local, regional, and distant recurrences, are considered together. Here, we conducted a comparative analysis of the above-reviewed literature data to select markers specific for each type of NSCLC progression (Fig. 1).

Most of the clinicopathological parameters including LVI, STAS, tumor size, and some others are common for different types of NSCLC progression. Nevertheless, there are clinicopathological factors specific to certain types of NSCLC progression. In particular, local recurrence is more common in LUSC, whereas distant metastases – in non-squamous cell histology, particularly in LUAD [93, 95]. The underlying mechanisms are unclear; however, it is suggested that tobacco smoke being the main cause of LUSC may have a field effect predisposing to local disease recurrence [93]. In addition, the distant metastasis of NSCLC is related to isolated BCH in the small bronchi distant from the tumor [21], whereas the co-presence of BCH and SM is associated with a high probability of locoregional recurrence [22, 23]. The mechanisms of association between premalignant bronchial lesions and NSCLC progression remain unclear. Probably, individual features of immune-inflammatory reactions, on the one hand, can predispose to isolated BCH or the development of BCH and SM, whereas, on another hand, can promote NSCLC distant metastasis or locoregional recurrence. Distant recurrence of NSCLC is also specifically related to BVI that is considered as a step of tumor cell escape from the primary site to distant organs [83, 103, 105, 161].

In contrast to clinicopathological parameters, genetic and molecular markers common for different types of NSCLC progression are rare and include only EGFR, KRAS, and ALK mutations as well as expression of CXCR4, HOTAIR, and miR-9 and level of circulating CEA. Almost all molecular markers are specifically associated with one or another type of NSCLC progression. Nevertheless, most of these unique markers are related to the regulation of similar processes such as proliferation, cell growth and differentiation, apoptosis, EMT, migration, invasion, and angiogenesis. For example, all types of NSCLC progression are related to molecules involved in the regulation of cell migration: lymph node metastasis – miR-206, miR-101, ITGA11, S100A4, and RACK1 [40, 44, 50, 56, 61], local recurrence – MMP12 [99], regional recurrence – IGFBP5 [100], and distant recurrence – miR-590, miR-375, MALAT1, ACTA2, CD44, TMSB4X, GAP43, and RSK1 [112, 117-119, 125, 134, 136, 137, 154].
Thus, lymph node metastasis, local, regional, and distant recurrences of NSCLC differ in the spectrum of clinicopathological and molecular markers that once again emphasize their attitude to pathogenetically distinct types of cancer progression.

**Challenges and approaches to developing NSCLC progression markers**

A number of clinicopathological parameters, genes, non-coding RNAs, and proteins have been described to be associated with the risk of NSCLC recurrence. However, the proposed markers are applicable mainly to LUAD and much less often to LUSC (Tables 1 and 2). Moreover, most of these markers have a number of significant limitations, such as insufficient predictive accuracy and economic and methodological complexity of analysis that prevents their validation in clinical practice.

Nevertheless, several prognostic markers, such as STAS, LVI, CEA, and mutations in the **EGFR** and **ALK** genes, described in multiple studies would seem to be reliable and valid. However, their translation into clinical practice is limited by several factors including lack of...
uniformity in the definition of STAS [162] and standardization in the methods used for STAS [163, 164] and LVI [165, 166] analysis, non-specificity of CEA for lung cancer [167, 168], its more sensitivity for LUAD, and unclear significance for non-adenocarcinomas [169], as well as contradictory results regarding the prognostic role of EGFR and ALK mutations in distant recurrence [7, 9, 12, 13, 114, 115]. Another problem is that all of these markers are associated with NSCLC progression, in general, that is heterogeneous and represented by clinically and molecularly distinct processes, particularly local, regional, and distant recurrences. In our opinion, the NSCLC prognosis should be built on the basis of markers specific to individual forms of tumor progression. Unfortunately, most of the proposed markers (Fig. 1) have been described in single studies, and additional research is required to validate them. The exception is BVI which has been reported in many studies to be associated predominantly with distant recurrence of NSCLC [83, 103, 105, 161]. However, there are many variations of BVI assessment methods and their standardization and quality control are needed [170].

The intermingling of different types of NSCLC progression and their presentation as cancer recurrence, in general, can be a possible explanation for the low effectiveness of suggested approaches to predict the risk of metastatic and recurrent disease. In this case, there also is some confusion regarding definitions of local, regional, and distant recurrence of NSCLC. For example, the definition of local recurrence, on which we relied, implies that cancer reappears in the same lung or at the bronchial stump [14-18]. However, according to other authors, local recurrence also includes recurring disease in ipsilateral hemithorax and mediastinum, pleural cavity, and lymph node regions [10, 95, 102, 103]. It is important to note that the term “regional recurrence” is often not used and the clinical manifestations of this type of progression are referred to as local recurrence [10, 95, 102, 103]. Regional recurrence is defined in this review as a recurring disease in the ipsilateral hilar, mediastinal, and supraclavicular lymph nodes detected during follow-up after treatment [9, 14, 19]. All other sites of recurrence are referred to as distant recurrence or metastasis [14]. Nevertheless, some studies refer relapse to second ipsilateral lobe or stump, as well as ipsilateral pleural recurrence, to regional recurrence, and metastases to supraclavicular lymph nodes are described as distant recurrence [16, 84, 87].

Also, the reason for the insufficient predictive power of the proposed markers may be tumor heterogeneity and the use of inadequate study models. Besides these challenges, the situation is complicated by the absence of highly effective methods to prevent NSCLC progression. Different therapeutics with anti-metastatic and anti-relapse effects have been suggested [171-173]; however, none of them are still used in clinical practice.

In our opinion, the use of the prognostic value of premalignant bronchial lesions can be one of the potential strategies to overcome the above-mentioned challenges. Previously, we showed that the probability of locoregional and distant recurrence of NSCLC is associated with the presence of premalignant lesions in the small bronchi distant from the primary tumor. In particular, the isolated BCH is related to increased distant metastasis and a decreased locoregional recurrence [21], whereas the co-presence of BCH and SM is associated with a high probability of locoregional recurrence and a low risk of distant metastasis [22, 23]. Importantly, neoadjuvant chemotherapy combined with intraoperative radiation shows high efficacy in the prevention of locoregional recurrence in high-risk NSCLC patients [23]. In this regard, we suggest that the analysis of premalignant changes in the bronchial epithelium at a distance from the primary tumor can be an effective instrument to assess the risk of NSCLC progression, whereas high- and low-risk patients may represent potential populations for the identification of highly reliable markers of cancer metastasis/recurrence. The analysis of the constitutive features of these patients, for example, germline variations and immune system parameters, and the investigation of the mutational landscape of their tumors can be also used to find potential molecular targets for the development of novel therapeutics aimed at the prevention of NSCLC progression.
Conclusion

The problem of prognosis and prevention of the risk of NSCLC progression is highly important. Despite advances in diagnostics and treatment, NSCLC continues to be diagnosed at the late stages. Even if NSCLC is detected early, the probability of recurrence is high. A huge amount of prognostic markers have been proposed; however, none of them are validated and recommended for the prediction of NSCLC progression in clinics. Moreover, there are no effective approaches to prevent cancer metastasis/recurrence. These problems are mainly related to the heterogeneous nature of NSCLC progression. Further studies should be conducted to investigate the mechanisms of different types of NSCLC progression including synchronous lymph node metastasis, local, regional, and distant recurrence, and to develop prognostic markers and potential therapeutic targets.

Abbreviations

BCH (basal cell hyperplasia); BVI (blood vessel invasion); LUAD (lung adenocarcinoma); LUSC (lung squamous cell carcinoma); LVI (lymphovascular invasion); NSCLC (non-small cell lung cancer); SM (squamous metaplasia); STAS (spread through air spaces); VPI (visceral pleural invasion).

Acknowledgements

We thank Ms. Ekaterina Khitrinskaya for the preparation of the figure.

Author Contributions

AAS and AAK wrote the text regarding molecular markers of different types of NSCLC progression. AAF and TSG focused on genetic markers. EOR, EBT, NAS, OVP, and AAD prepared subsections related to clinicopathological markers. MVZ, VMP, and EVD contributed to the conception and design and revised the review article.

Funding Source

This work was supported by the Russian Science Foundation [grant number 20-75-10060].

Statement of Ethics

The authors have no ethical conflicts to disclose.

Disclosure Statement

The authors have no conflicts of interest to declare.

References


120 Liu MX, Zhou KC, Cao Y: MCRS1 overexpression, which is specifically inhibited by miR-129*, promotes the epithelial-mesenchymal transition and metastasis in non-small cell lung cancer. Mol Cancer 2014;13:245.


