Genomic profiling and metastatic risk in early-stage non--small cell lung cancer

Cameron N. Fick, MD, a Elizabeth G. Dunne, MD, a Manendra B. Lankadasari, PhD, a,b Brooke Mastrogiagcomo, MS, a,c Tetsuhiko Asao, MD, PhD, a Stijn Vanstraelen, MD, a Yuan Liu, MD, PhD, a,b Francisco Sanchez-Vega, PhD, c and David R. Jones, MD a,b

Next-generation sequencing (NGS) and its applications have provided tremendous insight into the biologic landscape of a variety of solid tumors, including non--small cell lung cancer (NSCLC). The genomic characterization of NSCLC has not only altered treatment paradigms but has also aided investigations into tumor biology and the risk of developing metastatic disease. Complete surgical resection with or without chemotherapy and radiation has been the cornerstone of treatment for stage I-III NSCLC; however, up to 50% of patients with operable, early-stage NSCLC are at risk of metastasis after surgical resection.1,2 Thus, recent investigations have applied the knowledge gained in the stage IV NSCLC setting to earlier-stage disease. Studies are now focused on leveraging tumor genomic data (DNA and RNA) to elucidate the biologic mechanisms associated with disease recurrence and to potentially identify patients with a higher risk of recurrence for additional adjuvant therapies.

Diagnostic molecular pathologic analysis and related tumor genomic and transcriptomic analyses are presently being applied to several solid tumors. An obvious example is the use of Oncotype Dx, a 21-gene expression assay for patients with ER-positive/HER2-negative stage I-IIIA breast cancer. Oncotype Dx tests for genes related to cell proliferation, metastasis, HER2 expression, and sex hormone production.3,4 Current National Comprehensive Cancer Network (NCCN) guidelines recommend Oncotype Dx as the preferred panel for prognosis and prediction of benefit from adjuvant chemotherapy.5 The broad applicability of gene assays such as Oncotype Dx in NSCLC has been limited by the substantial histologic and molecular heterogeneity of NSCLC. In this report, we outline the current role of genomics in the management of NSCLC, including risk-stratification of patients, and provide foundational knowledge for thoracic surgeons caring for patients with NSCLC.

NSCLC HETEROGENEITY CONTRIBUTES TO METASTATIC RISK

Recent publications by the TRACERx Lung consortium have highlighted the presence of significant intratumoral subclonal somatic copy number alterations (SCNA) and mutations using whole-exome sequencing data on surgically resected NSCLC.6,7 In a cohort of treatment-naïve patients with stage I-III NSCLC who underwent complete surgical resection, tumors with a greater proportion of subclonal mutations were associated with an increased risk of recurrence or death.8 Subclonal whole-genome doubling, recent subclonal expansion, and a high level of SCNA intratumor heterogeneity (ITH) were also predictive of relapse; however, only high SCNA ITH was independently prognostic of early (<12 months after surgery) and extrathoracic recurrences.7 Wu and colleagues8 demonstrated similar findings in a spatial analysis of the ITH of lung adenocarcinoma in which they identified 2 distinct patterns—clustered...
and random geographic diversification—using both proteomic and genomic data. When clinicopathologic features were controlled for, the random proteomic geographic diversification pattern was associated with a greater risk of recurrence or death. These studies, which used NGS, offer important therapeutic insights into the mechanisms of resistance and vulnerability in NSCLC tumors, and future studies will continue to improve our appreciation of the diverse intratumoral environment.

In addition to molecular heterogeneity, NSCLC displays considerable histologic heterogeneity. The most common histologic type of NSCLC is adenocarcinoma, and further division into predominant subtypes (ie, lepidic, acinar, papillary, micropapillary, and solid) has provided unique genomic insights into the risk of metastasis for these patients. Caso and colleagues described changes at the gene and pathway levels that are associated with histologic subtype. A greater proportion of lepidic-predominant tumors had alterations in \(\text{EGFR}, \text{RBM10}, \text{and TERT}\), compared with other subtypes, whereas the more-aggressive histologic subtypes (micropapillary and solid) had enrichment of \(\text{BRAF}, \text{TP53}, \text{SETD2}, \text{MGA},\) and \(\text{SMARCA4}\) mutations. The presence of an aggressive histologic subtype was also associated with greater chromosomal instability, and in the TRACERx cohort, greater chromosomal instability was associated with shorter disease-free survival (DFS) and an increased risk of metastasis. Oncogenic pathway alterations also vary according to histologic subtype, with less-invasive lepidic-predominant tumors associated with RTK/RAS pathway changes and more-invasive micropapillary- and solid-predominant tumors associated with Myc, p53, and Wnt pathway changes.

### Genomic Changes Associated with Lung Adenocarcinoma Metastasis

Several genes and oncogenic pathways have been associated with metastasis in early-stage lung adenocarcinoma. Lengel and colleagues performed NGS on 766 primary lung adenocarcinoma samples and compared surgically resected primary tumors that metastasized with those that did not (Figure 1). Alterations in several genes, including \(\text{TP53}, \text{KEAP1}, \text{CDKN2A}, \text{MDM2}, \text{PIK3CA}, \text{NKX2-1}, \text{RB1}, \text{MYC}, \text{SMARCA4},\) and \(\text{FOXA1}\), were more common in primary lung adenocarcinomas that metastasized. Other studies have confirmed some of these findings in similar cohorts of patients. Conversely, \(\text{RBM10}\) is more common in lung adenocarcinoma that does not metastasize. In surgically resected tumors that metastasize, specific genomic alterations also influence survival; \(\text{MDM2}, \text{MYC}, \text{SMARCA4},\) and \(\text{TP53}\) are associated with shorter metastasis-free survival, whereas \(\text{EGFR}\) and \(\text{NF1}\) are associated with longer survival after relapse.

Certain pathway alterations also increase a patient’s risk of NSCLC metastasis, including alterations in the p53, PI3K, cell cycle, and transforming growth factor-beta pathways.
Changes in the Wnt signaling pathway have also been associated with poor outcomes in lung adenocarcinoma without a clear oncogenic driver. Cui and colleagues showed that differential expression of 4 Wnt pathway genes (CTNNB1, SOB9, DVL3, and WNT2B) can be used to partition patients into low- and high-risk groups with respect to overall survival. This finding is supported by the work of Kim and colleagues, who identified CTNNB1 as an independent predictor of recurrence in early-stage lung adenocarcinoma. Finally, the total number of oncogenic pathways altered is also important. An increased number of pathways altered is not only associated with high-risk clinicopathologic features—such as maximum standardized uptake value on positron emission tomography, aggressive histologic subtype, and lymph node involvement—but is also an independent predictor of worse DFS.

There has been a focus on driver alterations in NSCLC, and targeted therapies for an increasing number of these mutations have been developed. However, activating mutations in EGFR and their association with prognosis in early-stage lung adenocarcinoma is unclear. Deng and colleagues reported that the presence of an EGFR mutation portended a better prognosis, whereas others have found no association between EGFR mutation and recurrence or absence of metastasis. Deng and colleagues reported that EGFR mutations were associated with metastatic organotropism to the brain and bone in a cohort of 1531 patients who underwent NGS; however, Lengel and colleagues did not find a link between EGFR alterations and metastatic organotropism. The diverging results on the association between EGFR and metastasis may be attributable to the differing frequencies of EGFR mutations in Asian and North American populations. KRAS mutations have also been implicated in recurrence of early-stage NSCLC. KRAS G12C alterations are associated with shorter recurrence-free survival (RFS), compared with other KRAS mutations, in surgically resected lung adenocarcinoma. Furthermore, co-mutation of any KRAS mutation with STK11, ATM, or LRRQ3 was associated with shorter RFS, compared with KRAS mutation alone. KRAS alterations have also been associated with early metastasis, which is defined as a recurrence within 2 years of surgical resection.

Finally, individual genomic alterations are not the sole determinant of prognosis in lung adenocarcinoma, as other genomic parameters play a major role. Multiple studies, including a recent report of TRACERx data, found that high tumor mutation burden, fraction of genome altered, and whole-genome doubling were all more common in primary lung adenocarcinoma tumors that metastasized than in those that did not. Tumor mutation burden is an important biomarker, as studies have demonstrated its ability to predict response to immunotherapy in NSCLC. These observations strongly suggest that NGS panels should investigate changes at the chromosomal level in addition to individual gene alterations.

### VALIDATED MODELS PREDICTING RECURRENCE IN LUNG ADENOCARCINOMA

The identification of significant clinical, pathologic, and genomic features that influence metastasis and survival is an important aspect of properly risk-stratifying patients and provides useful information for decisions regarding treatment. Yet these factors in isolation are often inadequate, and models incorporating multiple variables to predict patient outcomes are more clinically applicable. In a cohort of patients with surgically resected early-stage lung adenocarcinoma (75% pathologic stage I), a multivariable analysis including clinicopathologic and genomic variables showed that fraction of genome altered and alterations in SMARCA4 and TP53 were associated with worse RFS. Based on association analyses, Jones and colleagues created a recurrence prediction model, PRecur, using a machine-learning framework (Table 1). When PRecur was used to stratify patients on the basis of predicted RFS, it outperformed the standard prognostic tumor, node and metastasis model (concordance probability estimate, 0.73 vs 0.61; \( P < .001 \) ) (Figure 2). Interestingly, the PRecur model correctly classified 83% of patients with stage I who developed a recurrence.

In a similar cohort of patients with nonsquamous NSCLC who underwent surgical resection, a smaller, 14-gene assay (Encore Clinical) identified patients at high risk of recurrence (Table 1). This assay has been validated in both North American and Asian populations, and it outperformed tumor, node and metastasis staging and the NCCN...
high-risk guidelines for prediction of recurrence.\textsuperscript{24,25} Furthermore, in a prospective cohort of 100 patients with stage I-IIA disease, 5-year DFS was 92\% among molecularly high-risk patients who received adjuvant chemotherapy, compared with 49\% among molecularly high-risk patients who did not receive adjuvant

\textbf{FIGURE 2.} Computational machine-learning prediction model (PRecur) for relapse-free survival applied to 2 patient scenarios. A, Patient with a small, 1.8-cm tumor (pT1bN0M0, stage IA). Three-year RFS curves were predicted by PRecur versus the TNM model for all patients with pT1bN0M0 in the study cohort (n = 136). B, Patient with a large, 5.1-cm tumor (pT3N0M0, stage IIB). Three-year RFS curves were predicted by PRecur versus the TNM model for all patients with pT3N0M0 in the study cohort (n = 53). SUV\textsubscript{max}, Maximum standardized uptake value; TMB, tumor mutation burden; FGA, fraction of genome altered; LVI, lymphovascular invasion; RFS, recurrence-free survival; TNM, tumor, node and metastasis. (From Jones and colleagues.\textsuperscript{22} Reprinted with permission).
Given the complexity of ITH in NSCLC, Biswas and colleagues tested previously published RNA sequencing (RNA-seq)-based and microarray-based prognostic assays on multiple tumor regions in 48 patients. They recognized that whether a patient was classified as low risk or high risk on RNA-seq prognostic models was dependent on the tumor region analyzed for a given patient, with discordance rates of 29% to 43%. When microarray-based prognostic models were used, median discordance was 50% among 9 models, which means that half of patients could be misclassified by these models, depending on the area of the tumor selected for analysis. Therefore, using RNA-seq data to identify genes associated with intratumoral homogeneity but high variability between tumors, the authors developed a 23-gene prognostic signature, ORACLE (Table 1). ORACLE had a dramatically lower rate of discordance of a 23-gene prognostic signature, ORACLE (Table 1).

Last, recent work by Sorin and colleagues has highlighted the predictive role that the tumor immune microenvironment plays with respect to patient outcomes. They used highly multiplexed imaging mass cytometry to perform spatial analysis of immune cells and their activation states and, using deep learning, identified patients who will have recurrence after surgical resection. While this method is not widely available at this time, it is a tremendous advancement that could serve as a valuable tool for future investigations.

**GENOMIC PROFILING IN LUNG SQUAMOUS CELL CARCINOMA**

NGS has identified far fewer distinct genomic profiles in lung squamous cell carcinoma than in lung adenocarcinoma. Genomic alterations in lung squamous cell carcinoma appear to resemble alterations in squamous cell carcinomas of other solid organs, rather than alterations in lung adenocarcinoma. Sanchez-Vega and colleagues analyzed the oncogenic signaling pathways altered in 502 lung adenocarcinoma and 464 lung squamous cell carcinoma samples from The Cancer Genome Atlas (TCGA). Squamous cell carcinoma samples more frequently had alterations in the p53 (86% vs 61%) and PI3K (68% vs 38%) pathways and less frequently had alterations in the RTK/RAS pathway (54% vs 77%), compared with adenocarcinoma samples. Compared with genomic alterations in lung adenocarcinoma, alterations in squamous cell carcinoma have been investigated less frequently; however, several of these commonly altered genes have been associated with recurrence in patients with squamous cell carcinoma. Alterations in *TP53* have been associated with an increased risk of recurrence, whereas *PI3KCA* mutations exhibit increased time to recurrence and better overall survival. *BCL6* and *ARID1A* alterations have also been shown to be associated with shorter DFS and overall survival, respectively, in patients with lung squamous cell carcinoma.

The tumor immune microenvironment appears to play a significant role in recurrence of early-stage squamous cell carcinoma, as Fan and colleagues uncovered differentially expressed immune-related genes. They developed a signature comprising 17 immune-related genes to predict overall survival in patients with early-stage lung squamous cell carcinoma; this gene signature has been externally validated. A summary of genes associated with both lung adenocarcinoma and squamous cell carcinoma is shown in Table 2.

### USING CIRCULATING TUMOR DNA (ctDNA) TO ASSESS METASTATIC RISK

When considering metastasis in NSCLC, providers face the challenge of risk-stratifying patients with early-stage NSCLC and how to accurately surveil them postoperatively. At present, the NCCN guidelines recommend that all patients undergo a history, physical examination, and computed tomography of the chest with or without contrast every 6 months for 2 to 3 years, and a low-dose noncontrast computed tomography of the chest annually thereafter. One recent development that provides an assessment of a patient’s status in real time is ctDNA. LUNGCA-1, a prospective multicenter study of perioperative ctDNA in patients with NSCLC undergoing surgery, demonstrated that a high concentration of ctDNA before resection was a negative predictor of RFS. Conversely, in the CheckMate 816 cohort, ctDNA clearance before the last cycle of nivolumab and chemotherapy was associated with pathologic complete response, compared with those who remained ctDNA positive (46% vs 0%); pathologic complete response was associated with better event-free survival.

### Summary

- NGS has identified far fewer distinct genomic profiles in lung squamous cell carcinoma than in lung adenocarcinoma.
- Genomic alterations in lung squamous cell carcinoma appear to resemble alterations in squamous cell carcinomas of other solid organs, rather than alterations in lung adenocarcinoma.
- Genes associated with worse prognosis and improved prognosis for both lung adenocarcinoma and squamous cell carcinoma are summarized in Table 2.
- Using circulating tumor DNA (ctDNA) to assess metastatic risk is a promising area of research, with recent studies showing its potential to provide real-time assessment of patient status.

---

**Table 2: Genes associated with recurrence in non–small cell lung cancer**

<table>
<thead>
<tr>
<th>Histologic type</th>
<th>Genes associated with worse prognosis</th>
<th>Genes associated with improved prognosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenocarcinoma</td>
<td><em>TP53, KEAP1, CDKN2A, MDM2, PIK3CA, NKX2-1, RB1, MYC, SMARCA4, FOXA1, CTNNB1, ALK/RET/ROS1, KRAS G12C</em></td>
<td><em>RBM10</em></td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td><em>TP53, BCL6, ARID1A</em></td>
<td><em>PIK3CA</em></td>
</tr>
</tbody>
</table>

---

*Given the complexity of ITH in NSCLC, Biswas and colleagues tested previously published RNA sequencing (RNA-seq)-based and microarray-based prognostic assays on multiple tumor regions in 48 patients. They recognized that whether a patient was classified as low risk or high risk on RNA-seq prognostic models was dependent on the tumor region analyzed for a given patient, with discordance rates of 29% to 43%. When microarray-based prognostic models were used, median discordance was 50% among 9 models, which means that half of patients could be misclassified by these models, depending on the area of the tumor selected for analysis. Therefore, using RNA-seq data to identify genes associated with intratumoral homogeneity but high variability between tumors, the authors developed a 23-gene prognostic signature, ORACLE (Table 1). ORACLE had a dramatically lower rate of discordance of 11%. Furthermore, this model has been predictive of overall survival in several external validation cohorts.*

*Last, recent work by Sorin and colleagues has highlighted the predictive role that the tumor immune microenvironment plays with respect to patient outcomes. They used highly multiplexed imaging mass cytometry to perform spatial analysis of immune cells and their activation states and, using deep learning, identified patients who will have recurrence after surgical resection. While this method is not widely available at this time, it is a tremendous advancement that could serve as a valuable tool for future investigations.*

*NGS has identified far fewer distinct genomic profiles in lung squamous cell carcinoma than in lung adenocarcinoma. Genomic alterations in lung squamous cell carcinoma appear to resemble alterations in squamous cell carcinomas of other solid organs, rather than alterations in lung adenocarcinoma. Sanchez-Vega and colleagues analyzed the oncogenic signaling pathways altered in 502 lung adenocarcinoma and 464 lung squamous cell carcinoma samples from The Cancer Genome Atlas (TCGA). Squamous cell carcinoma samples more frequently had alterations in the p53 (86% vs 61%) and PI3K (68% vs 38%) pathways and less frequently had alterations in the RTK/RAS pathway (54% vs 77%), compared with adenocarcinoma samples. Compared with genomic alterations in lung adenocarcinoma, alterations in squamous cell carcinoma have been investigated less frequently; however, several of these commonly altered genes have been associated with recurrence in patients with squamous cell carcinoma. Alterations in *TP53* have been associated with an increased risk of recurrence, whereas *PI3KCA* mutations exhibit increased time to recurrence and better overall survival. *BCL6* and *ARID1A* alterations have also been shown to be associated with shorter DFS and overall survival, respectively, in patients with lung squamous cell carcinoma.*

*The tumor immune microenvironment appears to play a significant role in recurrence of early-stage squamous cell carcinoma, as Fan and colleagues uncovered differentially expressed immune-related genes. They developed a signature comprising 17 immune-related genes to predict overall survival in patients with early-stage lung squamous cell carcinoma; this gene signature has been externally validated. A summary of genes associated with both lung adenocarcinoma and squamous cell carcinoma is shown in Table 2.*

*Using circulating tumor DNA (ctDNA) to assess metastatic risk is a promising area of research, with recent studies showing its potential to provide real-time assessment of patient status.*
The presence of preoperative ctDNA was more common in patients with nonadenocarcinoma lung cancer (compared with lung adenocarcinoma [92% vs 42%]), more smoking pack-years, and clinically occult mediastinal disease in lung adenocarcinoma. A landmark analysis of these patients 120 days postoperatively revealed that 93% of patients with minimal residual disease (MRD) detected by the presence of ctDNA developed recurrence. They also used ctDNA to determine that clonal expansion predicted metastasis.

Currently, there is no standard ctDNA assay, and the basis for detection in these assays can be either tumor-specific or tumor-independent (ie, epigenetic features), with variable limits of detection. While tumor-independent methods have the benefit of lower costs and do not require tissue acquisition, tumor-specific approaches have better limits of detection and sensitivity. In the MRD setting, sensitivity remains an issue, as a landmark analysis of a variety of ctDNA assays across multiple solid tumors revealed a median sensitivity of only 56%. Chaudhuri and colleagues improved on the previous methodology with the use of a tumor-informed NGS platform (CAPP-seq) that detects commonly mutated lung cancer genes at a lower limit of detection of 0.002%. In a cohort that included patients with stage I-III NSCLC treated with curative intent and healthy controls, receiver operating characteristic analysis revealed an area under the curve of 0.97 and maximal sensitivity and specificity of 93% and 96%, respectively. Similarly, Kurtz and colleagues greatly improved the limit of detection of ctDNA and thus assay sensitivity using an alternative method called phased variant enrichment and detection sequencing (PhasED-seq), which detects multiple somatic mutations in individual DNA fragments. Although these assays are not widely available in clinical practice, these 2 studies offer the most promising ctDNA assays that should be incorporated in future trials.

Concerns have also been raised regarding the predictive value of MRD ctDNA assays in the context of recurrence. The IMpower010 trial, which investigated atezolizumab versus best supportive care (BSC) after chemotherapy in patients with resected stage IB-IIIA NSCLC, used a different commercially available ctDNA assay (Signatera; Natera) to detect MRD postoperatively. Patients who were ctDNA-negative had better DFS than ctDNA-positive patients, regardless of which treatment they received (atezolizumab vs BSC). These results demonstrate the strong positive predictive value of these assays. However, in the BSC arm, nearly two-thirds of recurrences occurred in patients who were ctDNA negative, highlighting the low negative predictive value of the current ctDNA platforms. The IMpower010 authors also found that ctDNA positivity increased with stage (IB = 9%, II = 14%, IIIA = 29%). The ctDNA assays presented here are only three of several platforms available. Thus, ctDNA is an exciting development in the fight against metastasis in NSCLC, and subsequent studies will need to build on the current assays to improve the applicability of this biomarker in the clinical setting.

CONCLUSIONS

Tumor genomic profiling is now a standard-of-care tool for early-stage NSCLC and has several clinical implications. Models incorporating tumor genomic alterations serve as important predictors of metastasis and assist in our ability to risk-stratify patients. However, genomic analyses have also shown that patients with NSCLC are a divergent group who cannot be separated in neatly organized silos. Currently, clinicians caring for these patients face the challenge of synthesizing all the data that genomics has provided to make treatment decisions. Strategies incorporating the complexity of ITH in NSCLC into these prediction models will strengthen their utility and clinical applicability. As researchers better understand genetic alterations in NSCLC, additional treatments tailored to the individual patient will become available. Further, ctDNA assays that detect ctDNA preoperatively or the presence of MRD postoperatively represent a promising development for the identification of patients with early-stage disease who may benefit from neoadjuvant or adjuvant therapy. Overall, genomic profiling has the potential to move us away from a one-size-fits-all treatment approach to a personalized approach that offers patients a greater benefit to survival.

Conflict of Interest Statement

David R. Jones serves as a consultant or speaker for or has received grant support from Merck, AstraZeneca, Genentech, More Health, and JAVAA Oncology. All other authors reported no conflicts of interest.

The Journal policy requires editors and reviewers to disclose conflicts of interest and to decline handling or reviewing manuscripts for which they may have a conflict of interest. The editors and reviewers of this article have no conflicts of interest.

References


---

**Key Words:** early-stage non–small cell lung cancer, genomics, metastasis, tumor heterogeneity, ctDNA
E-References


