Lung cancer is the leading cause of cancer death in the United States, with 157,000 deaths annually, and in the world, with 1,375,000 deaths annually. The overall 5-year survival for lung cancer remains at 16%, which is significantly lower than the mortality rate for other common cancers, including colon, breast, and prostate cancer. The discrepancy in the mortality rate of lung cancer compared with other cancers has several explanations. First, lung cancers are typically detected at an advanced stage. The anticipated implementation of population-based screening for lung cancer is expected to decrease mortality. Second, compared with the other common cancers that share adenocarcinoma as the histology, lung cancer is heterogenous histologically and biologically. Until recently, lung cancer was treated as a homogenous disease with all non-small cell lung cancers (NSCLCs) treated identically, solely on the basis of clinical stage. Based on bench and clinical research, the treatment of lung cancer has been refined, with treatments allocated according to histology and specific molecular features. For example, targeting mutations such as epidermal growth factor receptor (EGFR) with tyrosine kinase inhibitors has been particularly successful as a treatment modality, demonstrating response rates in selected patients with adenocarcinoma tumors harboring EGFR mutations that are significantly higher than those for conventional chemotherapy. However, the development of new targeted therapies is, in part, highly dependent on an improved understanding of the molecular underpinnings of tumor initiation and progression, knowledge of the role of molecular aberrations in disease progression, and the development of highly reproducible platforms for high-throughput biomarker discovery and testing. In this article, we review clinically relevant research directed toward understanding the biology of lung cancer. The clinical purposes of this research are (1) to identify susceptibility variants and field molecular alterations that will promote the early detection of tumors and (2) to identify tumor molecular alterations that serve as therapeutic targets, prognostic biomarkers, or predictors of tumor response. We focus on research developments in the understanding of lung cancer somatic DNA mutations, chromosomal aberrations, epigenetics, and the tumor microenvironment, and how they can advance diagnostics and therapeutics.
and the development of highly reproducible platforms for high-throughput biomarker discovery and testing.

In this article, we review the clinically relevant research directed toward the molecular biology of lung cancer. The clinical purposes of this research are (1) to identify susceptibility variants and field molecular alterations that will promote the early detection of tumors and (2) to identify tumor molecular alterations that serve as therapeutic targets, prognostic biomarkers, or predictors of tumor response. The literature review was restricted to English-language papers from 2007 to 2012 that addressed the following six topics: somatic mutations in NSCLC; lung developmental and differentiation pathways in adenocarcinoma and squamous cell carcinoma (SCC); copy number alterations and gene rearrangements in NSCLC; epigenetic changes in promoter hypermethylation and microRNA (miRNA) expression in NSCLC; lung tumor microenvironment; and EGFR, K-ras, and echinoderm microtubule-associated protein like-4 (EML4)/anaplastic lymphoma kinase (ALK) alterations treated with TKIs of ALK inhibitors.

**1.0 LUNG CARCINOGENESIS**

The processes of tumor initiation and progression involve a multitude of orchestrated steps that encompass DNA damage and mutations induced by cigarette smoke carcinogens, to field carcinogenesis and preneoplasia in the airway epithelium, to the development of tumors and (2) to identify tumor molecular alterations that are important for cancer development. In the following sections, we outline several of these cancer hallmarks as they apply specifically to lung cancer.

### 1.1 Somatic DNA Mutations

Multiple somatic mutations contribute to malignant transformation in lung cancer. Mutations within tumor suppressors, oncogenes, and DNA repair-related genes may function in concert to confer a biologic advantage to tumors. Mutations at the chromosomal level (3p, 9p21, 12p, 17p13, 13q) in the form of loss of heterozygosity, translocations, deletions, and point mutations have all been well described in lung cancer.

These mutations, in turn, may be used as diagnostic and prognostic biomarkers. The development of rapid and sensitive sequencing technologies has advanced the characterization of the multitude of somatic mutations inherent in many tumors.

Early studies using sequencing technologies were focused on individual target genes such as p53 and p16. Traditionally, Sanger sequencing has served as the primary modality for exon-based sequencing, but it has limitations that include decreased sensitivity for the detection of low-frequency mutations. The majority of sequencing-based investigations have been designed to target specific tumor suppressors and oncogenes. The Tumor Sequencing Project, conducted in 188 cases of lung cancer, identified 26 genes that harbored mutations significantly above the baseline rate in uninvolved adjacent tissue, including several commonly identified mutations in EGFR, p53, and K-ras, among others. More advanced techniques that survey the genome now allow for the simultaneous detection of multiple mutations. A recent demonstration of a comprehensive approach to whole-genome sequencing in a patient with adenocarcinoma demonstrated not only previously identified mutations within K-ras, TP53, and CDK4, but numerous less well-described mutations within genes such as NEK9 and MUC16. In addition, investigators were able to identify mutations within nonprotein coding regions. This study and others confirm the concept that single tumors may harbor several genomic aberrations ranging from those at the chromosomal level down to single nucleotide variations that all must be taken into consideration with prognosis and treatment response in the development of the tumor. Each of these critical steps remains the focus of intense clinical and laboratory investigation. Although tobacco consumption remains the primary risk factor for the development of lung cancer, only 10% of smokers develop lung cancer. Thus, interindividual differences in susceptibility as modulated by genetic and epigenetic factors play an important role in lung carcinogenesis. In reviews, Hanahan and Weinberg outlined the crucial molecular alterations that are important for cancer development.
correlation of mutations. Su et al described two assays termed SNaPshot and polymerase chain reaction (PCR)-based sizing assays that may also be used to detect select somatic mutations in genes commonly implicated in cancer. High-resolution melting represents another PCR-based platform for initial mutation screening that holds some promise. In the future, it is likely that exon-specific or whole-genome-based sequencing will supplant existing technologies for screening tumor genome alterations.

1.2 Lung Developmental and Differentiation Pathways

Current paradigms suggest that lung carcinomas arise from pluripotent stem and progenitor cells capable of differentiation into one or several histologic cell types. These paradigms suggest that lung tumor cell ontology is determined by the consequences of gene transcriptional activation and/or repression events that recapitulate embryonic lung development. The hypothesis that lung cancer arises from the aberrant expression of genes involved in lung development is supported by gene expression studies demonstrating similarities between signatures obtained from human lung tumors and signatures characteristic of normal lung development. Genes associated with early lung development are more often expressed by small cell lung carcinoma and by other aggressive tumors with poor prognosis. Taken together, these observations suggest that poor differentiation is linked to the molecular parameters of early development representing lung stem and progenitor programs, and that the gene signatures of these phenotypes are important for lung cancer progression and clinical outcome.

The importance of identifying the NSCLC histologic subtype has increased recently because of the demonstrated success of histology-specific chemotherapeutic regimens. Until recently, all patients with NSCLC were treated without regard for histologic subtype. The first data suggesting the importance of histologic dependence were shown in the phase I-II trials of the vascular endothelial growth factor antibody bevacizumab. These results indicated that the drug was effective, but that toxicity was increased in patients with squamous histology. Subsequent trials (eg, Eastern Cooperative Oncology Group 4599) were restricted to patients with non-squamous histology, and current US Food and Drug Administration (FDA) approval is applicable to these patients only.

More recent clinical studies have shown a major treatment-by-histology interaction in the responsiveness of advanced stage IV NSCLC to pemetrexed. Scagliotti and colleagues performed an interaction analysis of the three pivotal phase 3 studies leading to the approval of pemetrexed in the first-line, second-line, and maintenance settings. The results demonstrated clear-cut and very significant treatment-by-histology interactions with regard to both progression-free and overall survival, thus confirming the superior efficacy of pemetrexed in patients with nonsquamous NSCLC. In addition, mutations of EGFR and K-ras and translocations of EML4/ALK are largely confined to adenocarcinoma. These findings have prompted a major shift toward targeted therapy in lung cancer management, as detailed by Socinski et al in the article, “Treatment of Stage IV Non-small Cell Lung Cancer,” in the American College of Chest Physicians Lung Cancer Guidelines, thus mandating that diagnostic specimens be acquired and processed in a fashion that permits the histologic subtyping of NSCLC.

In studies using immunostains to subtype NSCLC in cytologic specimens, Loo et al and Nicholson et al showed that thyroid transcription factor-1 and nucin vs p63 stains are the best stains for detecting adenocarcinoma vs SCC, respectively. However, none of the antibodies used in immunohistochemistry is 100% specific and sensitive for this differential diagnosis. Further details on clinical immunophenotyping of lung cancer are found in the article by Schwartz and Rezaei, “Diagnostic Surgical Pathology in Lung Cancer,” in the American College of Chest Physicians Lung Cancer Guidelines.

Identification of putative progenitor cells and stem cells of lung cancer is an active area of investigation. The objective is to target these cells to prevent cancer development and to treat established lung cancer. To date, no cell that strictly exemplifies the hallmarks of a stem cell has been confirmed as a lung cancer stem cell. Putative stem and progenitor cells include basal cells expressing p63 and cytokeratins 5 and 14, epithelial cells expressing the five-transmembrane glycoprotein CD133, isolated cancer cells with high activity of aldehyde dehydrogenase1, and bronchioloalveolar stem cells identified at the bronchioloalveolar duct junction in a murine model of lung cancer.

1.3 DNA Copy Number Alterations and Gene Rearrangements

Chromosomal abnormalities including amplifications, deletions, and rearrangements have been investigated extensively in lung cancer. One commonly used assay, fluorescent in situ hybridization (FISH), relies on fluorescent labeled DNA segments to detect regions with chromosomal aberrations. Early studies demonstrated that FISH could identify chromosomal aneuploidy in established lung cancer samples ranging from primary tumors, to effusions, to bronchial washings. Sokolova et al showed that FISH carried a sensitivity of 84% and a specificity of 82% in the...
detection of lung cancer in bronchial washings. The results from several studies suggest that FISH could complement the diagnostic accuracy of cytology. For example, Nakamura et al. compared FISH with conventional cytology and determined that FISH had a specificity of 100% and a sensitivity of 87%. Halling et al. detected a sensitivity of FISH of 71% in bronchial brushing specimens and 49% in bronchial washing specimens. Select FISH probes have been used as a diagnostic assay for the detection of lung cancer.

For example, in a cohort of patients with lung cancer and a control group with nonneoplastic disease, FISH for MYC (8q24), EGFR (7p12), and chromosomes 5 and 6 in bronchial brushings showed increased sensitivity compared with conventional cytology. Varella-Garcia et al. recently showed that chromosomal aneusomy in sputum as detected by FISH may be useful in the early detection of lung cancer.

An emerging application for FISH technology has been to determine EGFR status and, thus, TKI sensitivity. Mutational analysis for EGFR is increasingly becoming standard in the assessment of adenocarcinomas, given the significant response rates to EGFR/TKI inhibitors among patients harboring mutations. The role of the EGFR copy number in determining therapy or response remains unclear. Some studies suggest that the EGFR copy number correlates with mutational status and, to a lesser extent, with drug responsiveness. Another study confirmed that mutational status is the most valuable predictor of responsiveness, independent of copy number or protein expression. Several phase 2 and 3 trials have been conducted to answer this question, but the results have been inconsistent. Two meta-analyses confirmed that EGFR mutational analysis, not FISH, is the assay of choice to predict response to TKI therapy.

The EML4/ALK protein represents a fusion between ALK and EML4 that is present in approximately 4% of NSCLC. The rearrangement is associated with response to the targeted ALK inhibitor crizotinib. A comparison of different methodologies for dysregulated ALK detection concluded that ALK break-apart FISH could accurately detect rearrangements. Thus, the break-apart assay is now FDA approved to determine treatment with crizotinib. However, studies are ongoing to determine if a complementary set of assays, including immunohistochemistry, FISH, and reverse transcription-PCR, may be helpful in clinical decision making.

SOX2, located at chromosome 3q26.33, is a transcription factor gene involved in foregut development and pluripotency that has been shown to be amplified in SCCs, including lung. In fact, SOX2 carries a specificity of 95% in distinguishing SCC from adenocarcinoma. Several studies have shown that SOX2 has oncogenic properties and that direct targeting can suppress both metastases and growth. Interestingly, elevated SOX2 may correlate with a favorable prognosis in SCC but an unfavorable prognosis in stage 1 adenocarcinoma.

Fibroblast growth factor receptor (FGFR) is selectively amplified in SCC (22%). Reports from in vivo and in vitro studies indicate that FGFR1 amplification signifies sensitivity to treatment of SCC with FGFR inhibitors. Additional studies suggest that metastatic SCC may be particularly susceptible to anti-FGFR agents. TKIs including ponatinib (AP24534), which have potent anti-FGFR properties, are currently being investigated in lung cancer in phase 1 trials.

### 2.0 Epigenetics

#### 2.1 Methylation

Over the past several years, investigators have determined that epigenetic changes within the human genome at the level of chromatin remodeling, histone modifications, and DNA methylation can alter the expression patterns of critical tumor suppressors and oncogenes to influence malignant transformation. Aberrant hyper- and hypomethylation of CpG islands represent the most common finding, ranging from 15% to 80% of lung cancers. Several genes that have been identified, including CDKN2A, FHIT, APC, p16, and RASSF1A, appear to be commonly methylated.

Methylated genes may be detected using several techniques. The majority of early studies used technologies such as methylation-specific PCR and restriction landmark genomic scanning, which are targeted to specific genomic regions and candidates. These techniques have proven successful in identifying patterns of methylation that distinguish tumor from normal tissues and also in identifying clinical subgroups.

High-throughput profiling allows for global epigenetic analysis for candidate genes within promoter CpG islands; untranslated regions; and noncoding regions of the genome, including miRNA, long noncoding RNAs, and ultraconserved regions.

Both global and gene-specific epigenetic changes are being tested as noninvasive biomarkers for early detection and prognosis of lung cancer. Systematic examination of the “methylome” can be used to further subclassify molecular and clinically distinct subgroups of lung cancer and to potentially guide therapeutic decisions. Brock et al. examined the prognostic value of methylation patterns for seven genes (p16, MGMT, DAPK, RASSF1A, CDH13, ASC, and APC) in the setting of resected stage 1 NSCLC. The authors demonstrated that pairs of gene combinations of p16, CDH13, APC, and RASSF1A served as risk factors for recurrence. Methylation analysis for specific genes is now being applied to biomarker...
discovery efforts using noninvasive specimens. Belinsky et al\textsuperscript{73} demonstrated in sputum that the analysis of methylation patterns in a panel of predetermined genes had predictive value for the eventual development of lung cancer among high-risk smokers. An independent study concluded that methylation patterns within sputum correlated much better with primary tumor than did methylation within serum.\textsuperscript{74}

Epigenetics analysis is being integrated into clinical trials. De Fraipont et al\textsuperscript{75} examined tumor samples for methylation of RASSF1a and DAPKI within the French Intergroup (IFCT-0002) phase 3 trial of neoadjuvant treatment of stage I to II NSCLC and determined that methylation of both genes showed prognostic efficacy.

2.2 microRNA

During the past 2 decades, investigators have determined that large noncoding portions of the human genome may harbor functional capacity. These regions of noncoding RNA are now known to have gene regulatory function. miRNAs (or miRs) represent perhaps the best studied of this group of noncoding RNAs.\textsuperscript{76,77} Currently, there are >1,200 human miRNAs. Based on in silico analysis, miRNAs may regulate up to two-thirds of the human genome.\textsuperscript{78} miRNAs have been implicated in all the fundamental hallmarks of cancer initiation and progression. Profiling studies have demonstrated global differences in miRNA expression between lung tumor and noninvolved adjacent lung tissues.\textsuperscript{79} These initial observations have led to research focused on functional in vitro and in vivo validation of single-miRNA single-target relationships and on linking patterns of miRNA expression to clinical parameters such as prognosis and therapeutic response.

Several studies have examined miRNAs as prognostic biomarkers in lung cancer. The earliest studies in stage 1 adenocarcinomas demonstrated that increased miR-155 and decreased Let-7a2 correlated with poor survival.\textsuperscript{70} Landi et al\textsuperscript{80} identified a 5-miRNA signature that correlated with survival. Saito et al\textsuperscript{81} determined that miR-21 consistently correlated with poor outcomes among early lung stage cancers from three independent cohorts. Another study suggested that patterns of plasma miRNA may be used for risk stratification in lung cancer screening programs.\textsuperscript{82} Other studies have been unsuccessful in correlating miRNA to outcome. A study among patients undergoing surgical resection and adjuvant chemotherapy found no association between select miRNAs and survival or response to therapy.\textsuperscript{83} A few select miRNAs have emerged as having the capacity to distinguish histologic subtypes. Several studies have determined that miR-205, which is overexpressed in SCC compared with adenocarcinoma, can be used to distinguish the histologies.\textsuperscript{84-86} The role of miRNA in the clinical setting remains unclear, awaiting large, prospective profiling studies that demonstrate reproducibility.

3.0 Tumor Microenvironment

The past few decades have seen a new appreciation for the role of the tumor stroma in cancer initiation and progression.\textsuperscript{87} Signaling between the stroma and epithelium, via direct contact and secreted mediators, is essential for tumor initiation, growth, differentiation, progression, and metastasis.\textsuperscript{88} The tumor stroma is a complex system composed of fibroblasts (cancer-associated fibroblasts), macrophages (tumor-associated macrophages), other immune cells, vasculature, and extracellular matrix.\textsuperscript{89} Cancer cells modulate the stromal microenvironment via secretion of a variety of growth factors, including transforming growth factor-β,\textsuperscript{90} platelet-derived growth factor, vascular endothelial growth factor, and basic fibroblast growth factor. In turn, the stroma can produce antiangiogenic factors and recruit an antitumor immune response. Alternatively, the stromal response can facilitate tumor progression via the release of growth factors that feed back to tumor cells, participate in new vessel growth, and promote tumor growth and invasion.\textsuperscript{91} The stroma can have powerful tumor-promoting effects. For example, culturing cancer-associated fibroblasts with initiated prostatic epithelial cells enhances tumor growth and invasion,\textsuperscript{91} and in a murine model, loss of TβRII in fibroblasts can promote tumor initiation in the adjacent epithelium.\textsuperscript{92} In addition to its roles in lung tumor initiation and progression, the stroma, in cooperation with tumor cells, can establish a niche within the lung that is conducive to the establishment and growth of metastatic foci from other primary tumors.\textsuperscript{93,94} Several researchers have directed research toward understanding the prognostic significance of immune cell composition within and surrounding the lung tumor. Various immune cell compartments have been examined, including macrophages (M1 vs M2), T-helper (CD4 and CD8) cells, cytotoxic T cells, T-regulatory cells, B cells, natural killer cells, neutrophils, and myeloid-derived suppressor cells.\textsuperscript{95} To date, there is no validated immune cell biomarker of prognosis. The fields of cancer immunology and immunotherapy are evolving; thus, it is possible that further research will provide clinically important biomarkers of prognosis and predictors of response to immunotherapy.

4.0 Oncogene Addiction: EGFR, K-ras, EML4/ALK Alterations

The oncogene addiction hypothesis states that precise targeting of discrete fundamental genetic alterations in tumors will kill tumor cells and result in
clinical response. The successful application of this principle has been demonstrated in lung adenocarcinoma using agents targeted toward EGFR and ALK.

EGFR is a 170-kDa tyrosine kinase receptor that is overexpressed in 40% to 80% of NSCLC. It can be targeted with small-molecule TKIs, such as erlotinib and gefitinib, which bind to the ATP-binding pocket of the receptor. The sequencing of tumor samples from responding patients in clinical trials of TKIs led to the discovery of somatic mutations in the EGFR tyrosine kinase domain. The activating and oncogenic mutations reported to date are in exons 18 to 21, affecting the tyrosine kinase portion of EGFR. The majority of mutations are either small deletions of exon 19 affecting a 3-amino acid sequence or a point mutation L858R on exon 21. Mutations affecting exon 20, such as small insertions in T790M, account for 3% to 5% of all EGFR mutations, and although they are activating, they are also associated with resistance to EGFR TKIs. The prevalence of EGFR mutations varies by ethnicity, from 5% to 20% in whites, to 20% to 40% in the Asian population; the mutations are less common in black patients. Clinical trials from Japan with Maemondo and colleagues and from the West Japan Oncology Group, have shown that TKI therapy in EGFR mutant adenocarcinoma results in response rates of >70% and longer progression-free survival when compared with conventional chemotherapy in advanced lung adenocarcinoma. Together, these studies strongly support the role of molecular testing in lung adenocarcinoma for identification of EGFR mutant cancers and suggest that TKI treatment is the preferred first-line therapy for advanced EGFR mutant tumors. The American Cancer Society has issued a provisional clinical opinion that patients with NSCLC who are being considered for first-line therapy in response rates of

5.0 Molecular Profiling in Early-Stage Lung Cancer

To discriminate a poor prognosis from a good prognosis in early-stage lung cancers, a multitude of gene-profiling studies have been conducted on early-stage NSCLC, each proposing a signature of prognostic genomic information that correlates with patient
outcomes, and may subsequently guide decisions regarding adjuvant chemotherapy.\textsuperscript{109-116} However, the lack of reproducibility of signatures among independent studies raises concerns for implementation of this technology in routine clinical practice. Variations in patient selection, platforms, and statistical interpretation among studies and a lack of evaluation in prospective clinical trials have all been implicated as contributing factors to this issue.\textsuperscript{111}

Studies integrating gene signatures into clinical trials are beginning to demonstrate promise. Zhu et al\textsuperscript{112} demonstrated that a 15-gene tumor signature derived from patients enrolled in the JBR.10 adjuvant chemotherapy trial (National Cancer Institute of Canada Clinical Trials Group) could successfully identify high-risk patients who would benefit from adjuvant therapy.\textsuperscript{112}

Gene signatures in noninvasive specimens continue to be examined for the purpose of early diagnosis of lung cancer. However, several issues remain in the search for a noninvasive biomarker, including selection of proper source (eg, blood vs sputum), selection of the ideal profiling method, and analysis of data. Thus, gene expression profiling of sputum, airway, and blood shows promise but will require a extensive validation and prospective evaluation prior to clinical application.\textsuperscript{113-116}

6.0 CONCLUSION

Despite the small improvements in lung cancer 5-year survival over the past 3 decades, there is cause to be optimistic for a future with fewer lung cancer deaths. This future will deploy molecular techniques to diagnose lung cancer earlier and will continue to translate laboratory-based investigations of lung cancer into diagnostics and novel therapeutics that will have a significant impact on improving lung cancer survival. Recent technologic developments\textsuperscript{117} promise a preponderance of high-throughput data and novel disease insights that will continue to propel diagnostics and therapeutics toward a future with reduced lung cancer morbidity and mortality.

ACKNOWLEDGMENTS

Author contributions: Dr Powell had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Dr Nana-Sinkam: contributed to the literature search and drafting and revision of the manuscript. Dr Powell: contributed to the literature search and drafting and revision of the manuscript.

Financial/nonfinancial disclosures: The authors have reported to CHEST that no potential conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.

Role of Sponsors: The American College of Chest Physicians was solely responsible for the development of these guidelines. The remaining supporters played no role in the development process. External supporting organizations cannot recommend panelists or topics, nor are they allowed prepublication access to the manuscripts and recommendations. Further details on the Conflict of Interest Policy are available online at http://chestnet.org.

Endorsements: This guideline is endorsed by the European Society of Thoracic Surgeons, Oncology Nursing Society, American Association for Bronchology and Interventional Pulmonology, and the Society of Thoracic Surgeons.

REFERENCES


