

# Early Events in the Molecular Pathogenesis of Lung Cancer

Humam Kadara<sup>1,2</sup>, Paul Scheet<sup>2,3</sup>, Ignacio I. Wistuba<sup>1</sup>, and Avrum E. Spira<sup>4</sup>

## Abstract

The majority of cancer-related deaths in the United States and worldwide are attributed to lung cancer. There are more than 90 million smokers in the United States who represent a significant population at elevated risk for lung malignancy. In other epithelial tumors, it has been shown that if neoplastic lesions can be detected and treated at their intraepithelial stage, patient prognosis is significantly improved. Thus, new strategies to detect and treat lung preinvasive lesions are urgently needed in order to decrease the overwhelming public health burden of lung cancer. Limiting these advances is a poor knowledge of the earliest events

that underlie lung cancer development and that would constitute markers and targets for early detection and prevention. This review summarizes the state of knowledge of human lung cancer pathogenesis and the molecular pathology of premalignant lung lesions, with a focus on the molecular premalignant field that associates with lung cancer development. Lastly, we highlight new approaches and models to study genome-wide alterations in human lung premalignancy in order to facilitate the discovery of new markers for early detection and prevention of this fatal disease. *Cancer Prev Res*; 9(7); 518–27. ©2016 AACR.

## Introduction

Lung cancer is the leading cause of cancer-related deaths in the United States and worldwide (1, 2). The 5-year survival rates for advanced late-stage lung cancer remain poor (~5%; refs. 3, 4), leading to the supposition that early diagnosis of lung cancer or its treatment in premalignant stages will reduce the public health burden of this fatal malignancy (5–8). The vast majority (~85%) of lung cancers are causally linked to tobacco carcinogen exposure (3, 4). Heavy smokers as well as patients who have survived a prior aerodigestive cancer comprise a high-risk population that may be targeted for early detection and chemoprevention efforts (7). Also, while the risk of developing lung cancer decreases after smoking cessation, the risk never returns to baseline (9). Therefore, novel approaches for early detection and treatment of lung cancer among former smokers are urgently needed. Curtailing these key advances is a limited understanding of the earliest events in lung cancer pathogenesis.

Lung cancer consists of several histologic types, including small-cell lung carcinoma (SCLC) and non-small cell lung carcinoma (NSCLC), with the latter comprising the majority of diagnosed (~85%) lung tumors (4, 10). NSCLC is composed of the following three major histologic subtypes: lung adeno-

carcinomas (LUAD), lung squamous cell carcinomas (LUSC), and large-cell lung carcinomas (LCLC; refs. 3, 4, 10). The different lung cancer subtypes are thought to develop through diverse and unique molecular pathways. This supposition is supported by the divergent anatomical locations and suggested cells of origin of the different lung cancer subtypes. For example, compared with LUSCs and SCLCs that typically arise from the major bronchi and are centrally located, LUADs are thought to develop from small bronchi, bronchioles or alveolar epithelial cells and are typically peripherally located as reviewed elsewhere (10, 11).

Accumulating evidence suggests that malignant lung tumors develop through progressive pathologic changes known as preneoplastic or premalignant lesions (12, 13). Limited molecular aberrations (discussed below) have been described in lung premalignant lesions. Studies have shown that cytologically normal airway epithelial cells lining the entire respiratory tract and that have been exposed to smoking carry molecular alterations that may signify the onset of lung cancers (14, 15), a paradigm commonly referred to as the "airway field of injury" (refs. 14, 15; Fig. 1A). Additional studies have also pointed to molecular alterations (e.g., tumor-associated gene expression patterns) that are shared between NSCLCs and adjacent cytologically normal airway epithelia (referred to as molecular and adjacent airway field of cancerization; ref. 14; Fig. 1B). These premalignant fields can provide biologic insights into the development of lung tumors (refs. 14, 15; Fig. 1). Particularly, the airway field of injury, because it is widespread in the lung and precedes the development of tumors and premalignant lesions, also provides potential clinical opportunities for early detection and chemoprevention in high-risk smokers and in patients who have survived a cancer of the upper aerodigestive tract (14, 16).

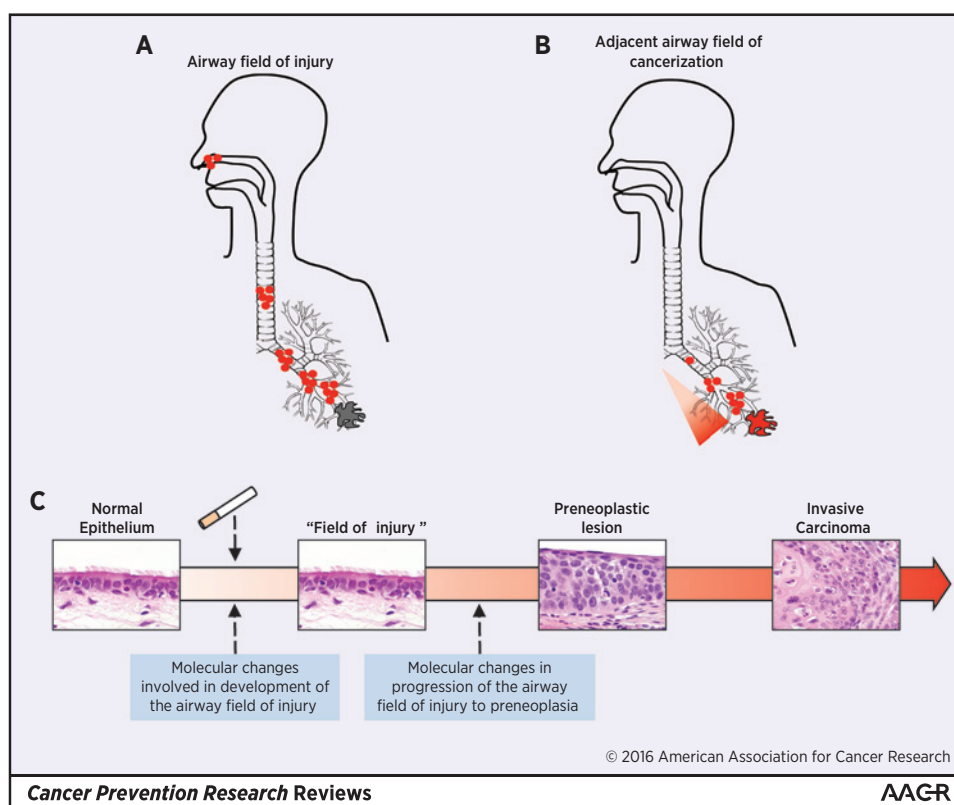
It is clear that early detection of lung cancer or intraepithelial lesions and effective prevention are crucial to reducing lung cancer mortality. However, and to date, large-scale lung cancer chemoprevention trials have been negative with neutral, or

<sup>1</sup>Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas. <sup>2</sup>The University of Texas Graduate School of Biomedical Sciences, Houston, Texas. <sup>3</sup>Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas. <sup>4</sup>Section of Computational Biomedicine, Boston University School of Medicine, Boston University, Boston, Massachusetts.

**Corresponding Author:** Humam Kadara, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 2951, Houston, TX 77030. Phone: 713-745-3186; Fax: 713-834-6082; E-mail: hkadara@mdanderson.org

doi: 10.1158/1940-6207.CAPR-15-0400

©2016 American Association for Cancer Research.



**Figure 1.**

**Premalignant airway fields in the molecular pathogenesis of lung cancer.** **A**, smoking induces widespread molecular alterations, such as gene expression changes (small red circles), in exposed epithelia throughout the respiratory tract. This airway field of injury has been described in smokers with or without lung cancer and is highly relevant for the identification of markers for minimally invasive and early detection of lung cancer. **B**, the adjacent airway field of cancerization represents the field in normal-appearing airways adjacent to lung tumors. It has been suggested that in this adjacent field of cancerization, there is closer molecular genealogy between lung cancers and airways that are in closest proximity to the tumors compared with airways that are more distant from the tumors [illustrated in the figure by the decreased number of red circles (i.e., aberrations) in airways with greater spatial distance from the tumor]. **C**, the progression of the molecular airway field of injury to preneoplasia and lung malignancy is still not clear. Understanding molecular changes involved in the development of the airway field of injury and changes mediating progression of this field to lung preneoplasia would enable the identification of early markers for lung cancer detection and chemoprevention.

even harmful, results (6, 17). Additionally, early discovery and diagnosis of lung cancers is still suboptimal (5). These limitations are largely attributed to the scarcity of reliable molecular targets (e.g., genomic loci or genes) on which to assess the efficacy of agents in clinical trials (6, 16, 17) and markers for early detection of lesions (5, 18). Identification of these markers and targets is thought to heavily rest on comprehending early lung oncogenesis. In spite of this urgency, we still do not know the molecular mechanisms by which the airway field of injury evolves to a lung cancer (Fig. 1C) and we are unable to identify the individuals at higher risk (e.g., heavy smokers) who are more likely to develop these lesions. Further and in particular, we still have no effective means to predict which premalignant lesions will go on to progress to malignant lung cancer and, if so, how we can prevent or block this process.

Here, we review the current state of knowledge of the molecular pathology of lung preneoplasia and oncogenesis. The review also describes the airway field of injury phenomenon and how molecular alterations in this epithelial field are relevant to the biology of lung cancer and to early detection and prevention of the disease. We will also consider new approaches to better understand

genome-wide changes in lung premalignant phases as potential means to identify new markers that can guide strategies for early detection and prevention of lung cancer.

## Lung Cancer Pathogenesis

Multistep tumorigenesis, in which tumors develop through a series of progressive pathologic changes (preneoplastic or precursor lesions) with corresponding genetic and epigenetic aberrations, has been demonstrated in various organs, including the skin, lung, and colon (19–23). Hyperplasia, squamous metaplasia, squamous dysplasia, and carcinoma *in situ* (CIS) comprise changes in the large airways that precede or accompany invasive LUSC (10, 12, 13). Dysplastic squamous lesions themselves are typically categorized by different intensities—mild, moderate and severe—based on a continuum of cytologic and histologic atypical changes (10, 12, 13). The sequential preneoplastic pathologic and molecular changes associated with LUSCs point to a multistage manner in the development of this lung cancer subtype from the respiratory mucosa (22). On the other hand, studies have suggested that the accumulation of

molecular abnormalities beyond a certain threshold, rather than the sequence of alterations, is the basis for development of the malignant phenotype (14). For example, sequential premalignant changes have been poorly documented for LUADs, neuroendocrine tumors, and SCLCs. Atypical adenomatous hyperplasia (AAH) is the only sequence of morphologic change identified so far for the development of invasive LUADs (12, 13, 24, 25). Diffuse idiopathic pulmonary neuroendocrine cell hyperplasia (DIPNECH), a rare lesion that includes local extraluminal proliferation in the form of tumorlets, is considered to be the precursor lesion for neuroendocrine carcinoids of the lung (26). Notably, there are no known preneoplastic lesions for the most common type of neuroendocrine lung tumors, SCLCs. Increasing our understanding of early events in lung cancer pathogenesis would allow us to better pinpoint the manner (multistep vs. linear) in which the malignant lung phenotype arises (27). Development of the lung malignant phenotype appears to be both due to stepwise, sequence-specific and multistage molecular pathogenesis and due to accumulation and combination of genetic and epigenetic abnormalities.

#### Molecular pathogenesis of LUAD

A large body of evidence suggests that at least two molecular pathways are involved in the development of LUAD, the Kirsten rat sarcoma viral oncogene (*KRAS*) and epidermal growth factor receptor (*EGFR*) pathways in smoker and non-smoker patients, respectively. Mutations in *EGFR*, particularly in-frame deletions of exon 19 and L858R and L861Q variants in exon 21, are strongly associated with never-smoking status, female gender, and East Asian ethnicity (28–31). On the other hand, mutations in *KRAS* are strongly associated with development of LUADs linked to tobacco consumption (4, 28, 32–35). Other commonly occurring point mutations and copy-number alterations (CNA) have been described in LUAD (32, 36, 37) and are reviewed elsewhere (38).

As mentioned before, AAH is the only sequence of morphologic change identified so far for the development of LUADs (12, 13, 24, 25) and there is consensus that the pathogenesis of many adenocarcinomas is largely unknown. Earlier studies have proposed that both Clara cells and alveolar type II (AT2) cells are the progenitor cells of LUADs (10, 39). However, accumulating evidence now strongly suggests that surfactant protein C (SPC) expressing AT2 cells, and not Clara (also known as club) cells, are the cell of origin of LUADs (40, 41). Due to the difficulty in distinction between AAHs and bronchioalveolar carcinoma (BAC), new classification guidelines for LUAD have discontinued the BAC terminology, which is now replaced with adenocarcinoma *in situ* (AIS; for LUADs with pure lepidic growth) or minimally invasive adenocarcinoma (MIA; predominant lepidic growth with less than 5-mm invasion; ref. 42). It is worthwhile to note that AAHs are thought to be involved in the linear progression of cells of the "terminal respiratory unit" (TRU) to AIS and subsequently invasive LUADs (11, 27, 43) due to the expression of common genes between the TRU and AAH. These studies in East Asian LUAD patients have postulated that most, if not all, peripheral LUADs progress from alveoli through AAH as a preneoplastic lesion. On the other hand, other studies suggested that LUADs may arise from bronchiolar epithelium and small bronchi based on the finding of driver *EGFR* mutations (e.g., *EGFR* p.L858R) in bronchioles adjacent to *EGFR*-mutant LUADs (44, 45).

There are shared molecular alterations between LUADs and AAHs and they provide further evidence that AAHs represent preneoplastic lesions in the course of LUAD pathogenesis (46). Previous studies have reported that ~30% to 40% of AAHs display *KRAS* (codon 12) mutations (25). In addition, *EGFR* mutations are shared at similar frequencies between AAHs, AIS lesions, and LUADs, particularly in sequence of lesions preceding development of the TRU subtype of LUADs (11). It is worthy to note that *EGFR* mutations were detected in adjacent normal-appearing peripheral respiratory epithelium in 43% of *EGFR*-mutant LUAD but not in patients wild-type for the oncogene (44), suggesting that there may be different cells of origin (bronchiolar epithelium and AT2 cells) for *EGFR*-mutant LUADs. Other molecular aberrations that were identified in AAHs include overexpression of Cyclin D1, survivin, and ERBB2 oncoproteins (24, 47, 48), loss-of-heterozygosity (LOH) in chromosomes 3p (18%), 9p (*CDKN2A*), 9q (tuberous sclerosis complex 1/*TSC1*), 17q, and 17p (*TP53*; refs. 10, 49) and reduced expression of the tumor suppressor serine threonine kinase 11 (*STK11* also known as *LKB1*; ref. 50). Epigenetic alterations including DNA methylation of *CDKN2A* and *PTPRN2* were also reported in AAHs (51). Studies aimed at interrogating differential gene expression and copy-number profiles between low-grade lesions (e.g., AAH and AIS) and LUADs found that amplification of the *EGFR* oncogene was the predominant differential molecular feature between early lesions and LUADs (27).

Studies have also shown that AAH and AIS lesions express high levels of *NKX2-1* (also known as *TTF-1*; ref. 43), a homeodomain-containing transactivating factor predominantly expressed in the terminal lung bronchioles and lung periphery (52, 53). In addition, *NKX2-1* is crucial for branching morphogenesis during normal lung development (52–54) and transactivates the expression of surfactant proteins important for the differentiation of AT2 cells (55). *NKX2-1* is commonly gained or amplified (14q13.3 locus) in LUAD pointing to a cell-lineage specific oncogenic function for this transcriptional factor in LUADs (32, 36, 56, 57). It is important to mention that while various reports have pointed to a growth-promoting role for *NKX2-1* in human LUAD cells (56–58), studies in *Kras*(LSL-G12D/+);*p53*(flox/flox) mice demonstrated that *Nkx2-1* suppresses murine lung tumor growth and metastatic potential *in vivo* (59).

#### Molecular pathogenesis of LUSC

It has been postulated that basal cells in the large airways exhibit pluripotent capacity following cigarette smoke exposure, giving rise to metaplastic and dysplastic squamous cells, which in turn function as precursors of squamous cell carcinomas (10, 12, 13). Earlier studies aimed at examining sequential molecular abnormalities in the pathogenesis of LUSC indicated that genetic abnormalities commence in histologically normal epithelium and augment with progression of histopathologic changes (22, 60, 61). Allelic losses at multiple 3p (3p21, 3p14, 3p22–24 and 3p12) chromosome sites and 9p21 (*CDKN2A*) in normal-appearing bronchial epithelia were described as the earliest detected changes in the sequence of pathogenesis of LUSCs (22, 61). Allelic imbalance in 8p21–23, 13q14 (*RB1*) and 17p13 (*TP53*) were detected in squamous preinvasive lesions (22, 61, 62). DNA methylation of *CDKN2A* is also found in early squamous preinvasive lesions with

frequencies that increase with histopathologic progression (24% in squamous metaplasia and 50% in CIS; ref. 63). Vascular endothelial growth factor (*VEGF*) isoforms and *VEGF* receptors (*VEGFR*) have been found to be elevated in bronchial squamous dysplastic lesions compared with normal bronchial epithelia (64), supporting the notion that angiogenesis develops early in lung carcinogenesis and that these abnormalities provide a rationale for the development of targeted antiangiogenic chemoprevention strategies. Further, angiogenic squamous dysplasias (ASD), a subset of squamous dysplasias, exhibit vascular budding in the subepithelial tissues with elevated microvessel density pointing to an architectural rearrangement of the capillary microvasculature (65). Additionally, the presence of this lesion in high-risk smokers suggests that aberrant patterns of microvascularization may occur at an early stage of bronchial carcinogenesis (65, 66). Besides altered angiogenesis, other pathways, including fatty acids and retinoic acid signaling, are also shared between squamous preinvasive lesions and LUSCs and are reviewed elsewhere (67, 68).

The *SOX2* oncogene, demonstrated to be amplified (3q26.3) in LUSCs, promotes survival of tumors with amplification of the gene (69, 70). Additionally, *SOX2* was found to foster the growth of lung tumor cells, particularly cells of squamous histology, and to mediate stemness of lung cancer stem cells (69, 71, 72). Alterations in *SOX2* are also evident in premalignant phases of LUSC development. Yuan and colleagues demonstrated that *SOX2* protein expression was completely absent in lung adenocarcinoma pathogenesis and highly expressed in preinvasive squamous lesions and in LUSCs (73). McCaughan and colleagues demonstrated, by analysis of the 3q chromosomal region, that high-grade bronchial dysplasias but not low-grade lesions exhibited amplification of *SOX2* (74). Notably, *SOX2* gain was associated with clinical progression of high-grade preinvasive squamous lesions (74). In a manner analogous to *NKX2-1* in LUADs, it is plausible to surmise that *SOX2* functions as a lineage-restricted oncogene in the early pathogenesis of LUSCs (73, 74).

In contrast to studies in malignant tumors, genome-wide analyses of lung premalignancy have been extremely rare. It is conceivable that molecular profiling of premalignant lung lesions will substantially increase our understanding of early events in lung cancer pathogenesis. An important step in this direction was the study by Nakachi and colleagues in which single-nucleotide polymorphism (SNP) arrays were used to characterize novel chromosomal alterations in preinvasive dysplastic lesions (75). The study revealed known and novel recurrent chromosomal alterations in dysplastic lesions (75). These chromosomal regions included losses of putative tumor suppressor genes *RNF20* and *SSBP2* and amplification of the *RASGRP3* oncogene (75). Notably, some of the chromosomal alterations were found in multiple premalignant lesions (75), indicative of widespread field effects (discussed further below). In another study by Ooi and colleagues, laser-capture microdissected and paired normal basal cells, premalignant lesions and LUSCs were analyzed by RNA-sequencing (RNA-seq; ref. 76). This study pinpointed gene expression profiles modulated early on between normal cells and premalignant lesions as well as profiles that changed between preneoplastic tissues and the LUSCs (76). Using pathways and gene network analyses, these expression changes were found to be indicative of *TP53* inhibition and *MYC* activation (76). The study by Ooi

and colleagues provides important clues into early mechanisms of LUSC pathogenesis and potential biomarkers for chemoprevention. It is important to note that the findings were limited by a small number of LUSC patients ( $n = 4$ ; ref. 76), thus warranting future studies with larger sample sizes. In addition, similar studies that comprise orthogonal (e.g., DNA sequencing) analyses will allow the identification of drivers in the development of squamous preinvasive lesions and in their progression to malignant lung tumors.

### SCLC pathogenesis

To date, there is no phenotypically identifiable preneoplastic lesion for SCLC. Compared with NSCLC, very little is known about early events in the molecular pathogenesis of SCLC. A study assessing LOH at several chromosomal sites and microsatellite instability in histologically normal and hyperplastic bronchial epithelia adjacent to lung tumors demonstrated a significantly higher incidence of alterations in epithelia nearby SCLCs than those adjacent to NSCLCs (77). These earlier findings point to widespread and extensive molecular damage in normal tissues surrounding SCLCs and suggest that SCLC may develop directly from histologically normal or mildly abnormal epithelium without passing through a more complex histologic sequence. It is noteworthy that studies using genetically engineered mouse (GEM) models demonstrated that SCLC development is driven by inactivation of *TP53* and *RB1* (78) as well as by activation of the hedgehog pathway (79, 80). In a study using integrative genomic analyses, human SCLCs were found to exhibit known alterations in *TP53* and *RB1*, recurrent mutations in the histone modifiers *CREBBP*, *EP300*, and *MLL* as well as amplifications in the *FGFR1* oncogene (81). More recently, proteomic profiling studies have identified alterations in SCLC that comprise therapeutically viable targets, including PARP1 and DNA repair proteins (82, 83). Also, Mohammad and colleagues described the histone demethylase lysine demethylase 1 (*LMD1*) as a therapeutic target for SCLC (84). It is noteworthy that the roles of these aforementioned alterations and targets in normal and premalignant phases of SCLC development are still poorly comprehended.

### Field of Injury and Field Carcinogenesis

Earlier work by Danely Slaughter assessing oral premalignant and malignant lesions revealed widespread histologic abnormalities in epithelial tissues that were distant to the oral cavity (e.g., in esophagus and the lung) and at the margins of invasive cancers (85). The studies by Slaughter and colleagues point to a "field cancerization," an effect in which the entire carcinogen-exposed epithelial field is primed for cancerization and tumor development (85). Field carcinogenesis has been described in various epithelial cell malignancies, including gastric, esophageal, hepatic, cervical, skin, and lung cancers (14, 86–88). In 1961, Auerbach and colleagues suggested that cigarette smoke induces extensive histologic changes in the bronchial epithelia of smokers, leading to widespread premalignant lesions throughout the respiratory epithelium, suggestive of a field effect (89). The advent of molecular high-throughput assays allowed the extension of Slaughter's concept to molecular changes, a phenomenon termed airway field of injury (15). It is now thought that the injured (e.g., by tobacco) airway epithelium exhibits various molecular alterations (e.g.,

aberrant gene expression, LOH) that precede and thus inform of the onset of primary lung cancer (15, 90–95). Because airways of smokers invariably exhibit some degree of inflammation and inflammatory-related damage, the field of injury in the lung may also be explained by both direct effect of tobacco carcinogens and initiation of the inflammatory response, and this concept has been extensively reviewed elsewhere (15).

Various studies have suggested that adjacent normal-appearing airway epithelia exhibit molecular alterations that are characteristic of the nearby lung tumor, an effect referred to by our group and others as the molecular and adjacent airway field of cancerization (14, 96, 97). As mentioned before, detailed analysis of histologically normal, premalignant, and malignant epithelia from LUSC patients revealed that multiple, sequentially occurring allelic imbalance events such as LOH commence in clonally independent foci early in the multistage pathogenesis of LUSC (22, 61). Notably, 31% of histologically normal epithelium adjacent to lung tumors exhibited LOH at chromosomal regions 3p and 9p (22). LOH was also described in normal bronchial epithelia of smokers without cancer and, importantly, appeared to have persisted for many years after smoking cessation (92, 95). Activating mutations in the *KRAS* and *EGFR* oncogene have also been described in histologically normal tissue adjacent to lung tumors (44, 98) and widespread mutations in *TP53* were reported in bronchial epithelia of cancer-free smokers (99). Epigenetic alterations are also present in the airway field of injury (63, 100–102). Belinsky and colleagues demonstrated aberrant methylation of *CDKN2A* and *DAPK* in the normal-appearing bronchial epithelium of lung cancer cases and cancer-free controls (100). The same study also revealed high level of concordance in *CDKN2A* methylation status between lung tumors and adjacent normal bronchial epithelium (100). Other genes that were described as methylated in the airway field of injury in cancer-free smokers include *RAR-β2*, *RASSF1A*, and *GSTP1* (101, 102).

Studies that used high-throughput approaches, such as gene expression profiling, to interrogate the airway field of injury that is widespread throughout the lung have provided new insights into early events in lung cancer pathogenesis (103). Hackett and colleagues analyzed expression of 44 anti-oxidant-related and demonstrated significant upregulation of 16 anti-oxidant genes in the airways of smokers relative to non-smokers (91). Normal-appearing bronchial epithelia of healthy cancer-free smokers were shown to comprise global changes in gene expression when compared with epithelium in non-smokers (93). Importantly, irreversible changes in airway gene expression following several years of smoking cessation were described and were proposed to be associated with the persistent risk former smokers exhibit for developing lung cancer (90, 93). An important milestone in field of injury studies was the development of a gene-expression signature in cytologically normal mainstem bronchus epithelium that can distinguish smokers with lung cancer from those with benign disease (94). Later, Gustafson and colleagues derived a *PIK3CA* pathway activation signature, which was found to be elevated in the cytologically normal bronchial airway of smokers with lung cancer or with premalignant lesions relative to cancer-free controls. Importantly, among smokers with premalignant lesions, this *PIK3CA* signature decreased in the field of injury following chemoprevention with the *PIK3CA* inhibitor myo-inositol, and changes in this airway gene-expression signature

were associated with clinical response as measured by regression of the dysplastic lesion (104). More recently, Shaykhiiev and colleagues demonstrated that airway basal cells in healthy smokers undergo reprogramming toward an embryonic stem cell-like phenotype and postulated that this process represents an early event in the development of lung carcinomas in smokers (105). It is worthwhile to mention that expression profiling studies also demonstrated widespread extent of spread of the molecular field of injury, revealing common alterations in bronchial, nasal, and buccal epithelia of smokers (106). Also, 119 genes were found to be modulated similarly by smoking in both bronchial and nasal epithelium (107), although the bronchial and nasal compartments also exhibited notable differences in gene expression.

Field carcinogenesis has also been shown to comprise aberrant modulation of microRNAs (miRNA; ref. 108). Global alterations in miRNAs were found in the airways of smokers relative to non-smokers (109). More recently, a study using microRNA-sequencing (miRNA-seq) strategy by Perdomo and colleagues identified a novel miRNA, miR-4423, that exhibited lineage-specific properties because it was found to be exclusively restricted in expression to the airway epithelium (110). This study also revealed that miR-4423 was decreased in lung tumors and inhibited features of the lung malignant phenotype *in vitro* and *in vivo* (110), suggesting that miRNAs also lie at the intersection between field carcinogenesis and early events in lung cancer pathogenesis.

The spatiotemporal dynamics of the airway field of injury and of the adjacent (to lung tumor) molecular field of cancerization were analyzed in recent studies by assessment of multiple field samples per patient. A study analyzing airway gene expression at several time points (up to 3 years) following surgery in definitively treated early-stage smoker NSCLC patients, who were enrolled in a phase II surveillance trial, provided novel insights into spatiotemporal characteristics of the airway field of injury (96, 111). Although this study was limited by lack of airway samples prior to surgery and when the tumor was still *in situ*, it identified candidate pathways and markers that were differentially modulated with time in the field (96, 111). These included increased bronchial expression of phosphorylated forms of the *AKT1* and *ERK1/2* oncogenes (111). These findings suggest that the airway field of injury exhibits temporal molecular changes that may be associated with lung cancer recurrence in surgically treated early stage patients (96). In another study by the same group, paired NSCLCs, uninvolved normal lung tissues and multiple normal-appearing airways in the adjacent field of cancerization were analyzed by whole-transcriptome profiling (97). Importantly, the study demonstrated that gene expression patterns in the adjacent airway field of cancerization were enriched with a signature that can distinguish smokers with lung cancer from smokers with benign disease (97). This study also identified notable spatial properties in the airway field of cancerization in which gene profiles of adjacent but not distant airway cells closely embodied expression patterns of the nearby NSCLC (97). One of these genes, *LAPTM4B*, a lysosome-associated oncogene, was later shown to be a novel activator of the *NRF2* stress response pathway in lung cancer cells (112). Like the report by Gustafson and colleagues (104), these studies (97, 111, 112) suggest that airway field of injury and the relatively more adjacent or local field of cancerization comprise

activation of therapeutically pliable signaling pathways, thus creating an opportunity for the development of early targeted chemoprevention strategies.

The above studies in field carcinogenesis hold the potential to directly impact clinical care by providing novel strategies for early detection and diagnosis (7). Although bronchoscopy is routinely used as an initial diagnostic tool in smokers with a suspicious lesion found on chest imaging, it is frequently non-diagnostic, often resulting in additional invasive testing, including surgical lung biopsy even when the lesion is benign (113, 114). In order to address this clinical unmet need, Whiteny and colleagues extended the initial observations of a cancer-specific gene-expression signature in the mainstem bronchus (94) by developing a 17-gene classifier, derived from bronchial epithelial cells collected during bronchoscopy, able to discern current or former smokers with lung cancer from those without the malignancy (115). In a recent translational study by Silvestri and colleagues, the performance of the 17-gene classifier as a bronchial biomarker for lung cancer detection was prospectively validated in 639 suspect smokers undergoing bronchoscopy in two multicenter trials, Airway Epithelial Gene Expression in the Diagnosis of Lung Cancer (AEGIS) 1 and 2 trials (116). It is worthwhile to note that the classifier and bronchoscopy exhibited a combined sensitivity of 97% for diagnosing lung cancer in both cohorts, with high sensitivity for the classifier irrespective of size, location, cell type, or stage of the cancer (116). Notably, among smokers with a non-diagnostic bronchoscopy and an intermediate pre-test probability of malignancy (where the physician is most uncertain about cancer status), the negative predictive value of the classifier was 91%, potentially impacting clinical decision making by allowing these patients to be followed with CT surveillance as opposed to invasive biopsies in patients without lung cancer (116).

## Emerging Concepts and Approaches to Study Lung Cancer Pathogenesis

The application of standard genomic technologies (e.g., arrays and sequencing) to survey genome-wide alterations has largely enhanced our understanding of the biology of malignant lung tumors. In sharp contrast, our understanding of the molecular pathology of the airway fields of injury and cancerization and of lung premalignant lesions is extremely limited. Understanding the landscape of genomic changes in normal or premalignant phases in lung oncogenesis may provide valuable insights into the earliest events that drive the pathogenesis of lung cancer.

### Deep next-generation sequencing

Next-generation sequencing (NGS) technology, through whole-genome, whole-exome, and whole-transcriptome approaches, holds great promise for providing invaluable insights into lung cancer biology, diagnosis, prevention, and therapy (117). NGS enables the sequencing of expressed genes, exons, and complete genomes, providing data on levels of expression, with a substantially larger dynamic range compared with array technology, sequence alterations, single-nucleotide variations as well as structural genomic aberrations (117). While various reports have used NGS to interrogate the molecular pathology of malignant lung tumors (32, 36–38, 118, 119),

similar studies in premalignant lung tissues have been very limited.

Beane and colleagues interrogated the transcriptome of airway epithelia from a small number of smokers with and without lung cancer by RNA-seq (120). This study identified markers and operative signaling cues, in the airway field of injury, that appear pertinent to lung oncogenesis (120). More recently, using deep-sequencing technology, Perdomo and colleagues identified a novel lineage-specific miRNA, miR-4423, that is primarily expressed in normal airway epithelium (110). Following RNA-seq, the study then demonstrated that miR-4423 regulated airway epithelial cell differentiation *in vitro* and suppressed features of the lung malignant phenotype *in vitro* and *in vivo* (110). Recently and as described above, RNA-seq was used to interrogate genes that are differentially expressed between paired normal epithelia, preinvasion squamous lesions, and LUSCs (76). This study identified genes modulated "early" between normal epithelia and preneoplasia as well as markers modulated later in the pathogenesis of the lesions to LUSCs, although findings were limited by the small cohort size (76). Adequately powered NGS studies of the airway field and lung preneoplasia will undoubtedly expand our understanding of the biology of lung cancer pathogenesis.

### Intra-lesional heterogeneity

Concepts in evolutionarily biology have provided important insights into the evolution and progression of cancer (121, 122). Earlier work by Nowell suggested that tumor cells acquire mutations some of which provide survival advantages to selective clones of cells (123). Accumulating evidence points to a branched evolution pattern for tumor pathogenesis in which subclones expand independently, acquiring different mutations over time (121, 122). In addition, these multiple subclonal populations can coexist together (121). Evidence of heterogeneity among tumor cells was first observed by molecular pathologists through cytogenetic analysis and examination of karyotypes (121, 124, 125). Heterogeneity was also previously genetically discerned by the identification of differential copy-number signals among different cells in the same tumors (reviewed in ref. 126). With the advent of NGS technologies, intratumoral heterogeneity (ITH) can now be assessed genome wide at the single-nucleotide level (121). Genomic ITH has now been described in various malignancies, including lung cancer (121, 127, 128). Zhang and colleagues used whole-exome sequencing to survey the spectrum of ITH in 11 early-stage LUADs (128). Notably, this study alluded to an association between the extent of ITH and post-surgical relapse, although findings were limited by the relatively small cohort and warrant validation in additional studies (128). It is tempting to speculate that paired analysis of molecular spectrum of ITH in the matched normal-appearing airway epithelia, premalignant lesions and primary NSCLCs will shed light on shared genealogy among the different stages of lung oncogenesis. It is conceivable to surmise that smoking-exposed airway epithelia in the fields of injury/cancerization or premalignant lesions may exhibit cellular subclones with different mutational patterns, and thus variable capacity to progress to NSCLC. Resolving these features will allow us to better understand the spatiotemporal dynamics of lung cancer evolution.

### Identification of subtle alterations in lung premalignancy

Airway epithelia with field carcinogenesis or premalignant lesions present unique challenges when profiling their molecular changes via standard genomic technologies. These technologies were originally designed for scenarios of higher purity of the targeted tissue, or at least more balanced contributions of components of interest, i.e., the cells of the premalignant lesion harboring somatic mutation (129, 130). When the proportion of aberrant cells in the targeted tissue is low, analogous to low tumor purities ("cellularity," or proportion of DNA harboring an aberration), the task of identifying alterations becomes exceedingly difficult. For the normal-appearing airway fields of injury/cancerization and preinvasive lesions, genomic alterations will certainly be subtle, and in some cases rare (within samples and across individuals).

There are two main sources of information when assessing evidence for CNAs: (i) total copy-number gains and losses from allelic intensities (from SNP microarrays) or read coverage (from NGS); and (ii) imbalance in allele-specific signals (B-allele frequencies; BAF) at germline heterozygous loci (131). Generally, standard computational methods may discern BAF dispersions when the aberration proportions are above 10% to 15% (132–135). However, methods that model the dispersion in BAFs induced by a CNA ignore critical information in subtle shifts of the BAF values (136). Because it is likely that a single chromosomal event, such as a CNA (i.e., a loss, gain, or cn-LOH), induced the allelic imbalance in a particular contiguous region on a chromosome, the direction the BAFs move from the canonical value for a heterozygote (one-half) is not random (136). They, instead, move according to the allelic composition of an inherited chromosome, also known as a haplotype (136). Thus, methods that model the distribution of BAF values jointly can gain considerable power, particularly at low cell fractions, over standard algorithms that ignore this information and instead consider only the magnitude of their shifts but not their direction (136). Precancerous lesions present greater challenges than characterizing cancer genomes, and incorporation of haplotype information in this space is even more critical.

Recently, computational methods have been developed that incorporate this information and deal directly with uncertainty in haplotype estimation, hapLOH (136), and J-LOH (137). These highly sensitive haplotype-based computational methods can deal with low-cellularity samples, or situations where the aberrant cell fraction is below 5% by assessing consistency between BAF values and inherited chromosomes, estimated statistically (138, 139), capturing the direction of the BAF values (136, 137, 140, 141). This extra level of specificity is critical when attempting to characterize phenomena that are rare (infrequent along the genome or in few samples) and/or subtle (occurring in a low proportion of the cells from which the DNA was obtained), settings anticipated in normal-appearing the airway field of injury or in preneoplasia. When applied to study chromosomal alterations in the airway field of cancerization, these haplotype-based methods may be able to discern

subtle events of allelic imbalance in the adjacent normal-appearing airway that are also present in the nearby NSCLC as well as within different multiregion intratumoral biopsies within the same early-stage NSCLC tumor. Applying these sensitive methods to infer the order of somatic events should lend valuable insights into studies of field carcinogenesis and preneoplasia and offer windows into the earliest genomic changes associated with lung cancer pathogenesis and, thus, targets for chemoprevention.

### Concluding Remarks

Despite numerous efforts that have centered on characterizing molecular profiles in lung cancer, this malignancy still accounts for the biggest fraction of cancer-related deaths in the United States and worldwide (1, 2, 7). Despite encouraging findings from the National Lung Screening Trial (142, 143), early detection of lung cancer is challenging due to the lack of biomarkers for the early diagnosis of the disease and due to the presence of multiple neoplastic molecular pathways that mediate lung carcinogenesis (5, 18, 103, 114). Moreover, little progress has been made in personalized and targeted chemoprevention strategies for this fatal malignancy (6, 17). Preneoplastic changes have been used as surrogate endpoints for chemopreventive studies. However, it was suggested that this "shooting in the dark" approach may explain the reasons behind the general failures of clinical chemoprevention studies (16). Therefore, novel approaches to identify the best population to be targeted for early detection and chemoprevention should be devised, and risk factors for lung cancer development or relapse need to be better defined. Achieving this rests heavily on understanding early events (e.g., field carcinogenesis) in the molecular pathogenesis of lung cancer. This notion is exemplified by the recent and encouraging findings of the prospective AEGIS-1 and -2 clinical trials demonstrating sensitive and specific detection of lung cancer based on a molecular gene classifier derived from relatively accessible normal airway cells of suspect smokers (116). Comprehensive analysis of early molecular events in lung cancer pathogenesis will undoubtedly unravel biomarkers that can aid early detection and prevention efforts deliver their longstanding promise to oppose lung cancer.

### Disclosure of Potential Conflicts of Interest

A.E. Spira is a consultant/advisory board member for Veracyte Inc. No potential conflicts of interest were disclosed by the other authors.

### Grant Support

This study was supported in part by Departments of Defense (DoD) grant W81XWH-10-1-1007 (H. Kadara, I.I. Wistuba, and A. Spira) and W81XWH-11-2-0161 (A. Spira), NIH grant R01HG005859 (P. Scheet), and Cancer Prevention and Research Institute of Texas (CPRIT) award RP150079 (P. Scheet and H. Kadara).

Received November 23, 2015; revised February 29, 2016; accepted March 1, 2016; published OnlineFirst March 22, 2016.

### References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA Cancer J Clin* 2015;65:5–29.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;65:87–108.

3. Goldstraw P, Ball D, Jett JR, Le Chevalier T, Lim E, Nicholson AG, et al. Non-small-cell lung cancer. *Lancet* 2011;378:1727–40.
4. Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med* 2008;359:1367–80.
5. Kanodra NM, Silvestri GA, Tanner NT. Screening and early detection efforts in lung cancer. *Cancer* 2015;121:1347–56.
6. Keith RL, Miller YE. Lung cancer chemoprevention: current status and future prospects. *Nat Rev Clin Oncol* 2013;10:334–43.
7. Spira A, Halmos B, Powell CA. Update in Lung Cancer 2014. *Am J Respir Crit Care Med* 2015;192:283–94.
8. Szabo E, Mao JT, Lam S, Reid ME, Keith RL. Chemoprevention of lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143:e40S–60S.
9. Karam-Hage M, Cinciripini PM, Gritz ER. Tobacco use and cessation for cancer survivors: an overview for clinicians. *CA Cancer J Clin* 2014;64:272–90.
10. Wistuba II, Gazdar AF. Lung cancer preneoplasia. *Annu Rev Pathol* 2006;1:331–48.
11. Yatabe Y. EGFR mutations and the terminal respiratory unit. *Cancer Metastasis Rev* 2010;29:23–36.
12. Colby TV, Wistuba II, Gazdar A. Precursors to pulmonary neoplasia. *Adv Anat Pathol* 1998;5:205–15.
13. Kerr KM. Pulmonary preinvasive neoplasia. *J Clin Pathol* 2001;54:257–71.
14. Kadara H, Wistuba II. Field cancerization in non-small cell lung cancer: implications in disease pathogenesis. *Proc Am Thorac Soc* 2012;9:38–42.
15. Steiling K, Ryan J, Brody JS, Spira A. The field of tissue injury in the lung and airway. *Cancer Prev Res* 2008;1:396–403.
16. Gold KA, Kim ES, Lee JJ, Wistuba II, Farhangfar CJ, Hong WK. The BATTLE to personalize lung cancer prevention through reverse migration. *Cancer Prev Res* 2011;4:962–72.
17. Greenberg AK, Tsay JC, Tchou-Wong KM, Jorgensen A, Rom WN. Chemoprevention of lung cancer: prospects and disappointments in human clinical trials. *Cancers* 2013;5:131–48.
18. Brothers JF, Hijazi K, Mascoux C, El-Zein RA, Spitz MR, Spira A. Bridging the clinical gaps: genetic, epigenetic and transcriptomic biomarkers for the early detection of lung cancer in the post-National Lung Screening Trial era. *BMC Med* 2013;11:168.
19. Cho KR, Vogelstein B. Suppressor gene alterations in the colorectal adenoma-carcinoma sequence. *J Cell Biochem Suppl* 1992;16C:137–41.
20. Cho KR, Vogelstein B. Genetic alterations in the adenoma-carcinoma sequence. *Cancer* 1992;70:1727–31.
21. Vogelstein B, Kinzler KW. The multistep nature of cancer. *Trends Genet* 1993;9:138–41.
22. Wistuba II, Behrens C, Milchgrub S, Bryant D, Hung J, Minna JD, et al. Sequential molecular abnormalities are involved in the multistage development of squamous cell lung carcinoma. *Oncogene* 1999;18:643–50.
23. Yuspa SH. The pathogenesis of squamous cell cancer: lessons learned from studies of skin carcinogenesis—thirty-third G. H. A. Clowes Memorial Award Lecture. *Cancer Res* 1994;54:1178–89.
24. Westra WH. Early glandular neoplasia of the lung. *Respir Res* 2000;1:163–9.
25. Westra WH, Baas IO, Hruban RH, Askin FB, Wilson K, Offerhaus GJ, et al. K-ras oncogene activation in atypical alveolar hyperplasias of the human lung. *Cancer Res* 1996;56:2224–8.
26. Carr LL, Chung JH, Duarte Achcar R, Lesic Z, Rho JY, Yagihashi K, et al. The clinical course of diffuse idiopathic pulmonary neuroendocrine cell hyperplasia. *Chest* 2015;147:415–22.
27. Yatabe Y, Borczuk AC, Powell CA. Do all lung adenocarcinomas follow a stepwise progression? *Lung Cancer* 2011;74:7–11.
28. Mounawar M, Mukeria A, Le Calvez F, Hung RJ, Renard H, Cortot A, et al. Patterns of EGFR, HER2, TP53, and KRAS mutations of p14arf expression in non-small cell lung cancers in relation to smoking history. *Cancer Res* 2007;67:5667–72.
29. Rudin CM, Avila-Tang E, Harris CC, Herman JG, Hirsch FR, Pao W, et al. Lung cancer in never smokers: molecular profiles and therapeutic implications. *Clin Cancer Res* 2009;15:5646–61.
30. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst* 2005;97:339–46.
31. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers—a different disease. *Nat Rev Cancer* 2007;7:778–90.
32. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 2012;150:1107–20.
33. Le Calvez F, Mukeria A, Hunt JD, Kelm O, Hung RJ, Taniere P, et al. TP53 and KRAS mutation load and types in lung cancers in relation to tobacco smoke: distinct patterns in never, former, and current smokers. *Cancer Res* 2005;65:5076–83.
34. Ridanpaa M, Karjalainen A, Anttila S, Vainio H, Husgafvelpursiainen K. Genetic alterations in p53 and k-ras in lung-cancer in relation to histopathology of the tumor and smoking history of the patient. *Int J Oncol* 1994;5:1109–17.
35. Rodenhuis S, Slebos RJ, Boot AJ, Evers SG, Mooi WJ, Wagenaar SS, et al. Incidence and possible clinical significance of K-ras oncogene activation in adenocarcinoma of the human lung. *Cancer Res* 1988;48:5738–41.
36. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014;511:543–50.
37. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* 2012;150:1121–34.
38. Devarakonda S, Morgensztern D, Govindan R. Genomic alterations in lung adenocarcinoma. *Lancet Oncol* 2015;16:e342–51.
39. Kim CF, Jackson EL, Woolfenden AE, Lawrence S, Babar I, Vogel S, et al. Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* 2005;121:823–35.
40. Desai TJ, Brownfield DC, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* 2014;507:190–4.
41. Mainardi S, Mijimolle N, Francoz S, Vicente-Duenas C, Sanchez-Garcia I, Barbacid M. Identification of cancer initiating cells in K-Ras driven lung adenocarcinoma. *Proc Natl Acad Sci U S A* 2014;111:255–60.
42. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* 2011;6:244–85.
43. Yatabe Y, Mitsudomi T, Takahashi T. TTF-1 expression in pulmonary adenocarcinomas. *Am J Surg Pathol* 2002;26:767–73.
44. Tang X, Shigematsu H, Bekele BN, Roth JA, Minna JD, Hong WK, et al. EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. *Cancer Res* 2005;65:7568–72.
45. Tang X, Varella-Garcia M, Xavier AC, Massarelli E, Ozburn N, Moran C, et al. Epidermal growth factor receptor abnormalities in the pathogenesis and progression of lung adenocarcinomas. *Cancer Prev Res* 2008;1:192–200.
46. Kitamura H, Kameda Y, Ito T, Hayashi H. Atypical adenomatous hyperplasia of the lung. Implications for the pathogenesis of peripheral lung adenocarcinoma. *Am J Clin Pathol* 1999;111:610–22.
47. Nakanishi K, Kawai T, Kumaki F, Hiroi S, Mukai M, Ikeda E. Survivin expression in atypical adenomatous hyperplasia of the lung. *Am J Clin Pathol* 2003;120:712–9.
48. Tominaga M, Sueoka N, Irie K, Iwanaga K, Tokunaga O, Hayashi S, et al. Detection and discrimination of preneoplastic and early stages of lung adenocarcinoma using hnRNP B1 combined with the cell cycle-related markers p16, cyclin D1, and Ki-67. *Lung Cancer* 2003;40:45–53.
49. Takamochi K, Ogura T, Suzuki K, Kawasaki H, Kurashima Y, Yokose T, et al. Loss of heterozygosity on chromosomes 9q and 16p in atypical adenomatous hyperplasia concomitant with adenocarcinoma of the lung. *Am J Pathol* 2001;159:1941–8.
50. Ghaffar H, Sahin F, Sanchez-Cepedes M, Su GH, Zahurak M, Sidransky D, et al. LKB1 protein expression in the evolution of glandular neoplasia of the lung. *Clin Cancer Res* 2003;9:2998–3003.
51. Selamat SA, Galler JS, Joshi AD, Fyfe MN, Campan M, Siegmund KD, et al. DNA methylation changes in atypical adenomatous hyperplasia, adenocarcinoma in situ, and lung adenocarcinoma. *PLoS One* 2011;6:e21443.
52. Mino P, Su G, Drum H, Bringas P, Kimura S. Defects in tracheoesophageal and lung morphogenesis in Nkx2.1(-/-) mouse embryos. *Dev Biol* 1999;209:60–71.



53. Yuan B, Li C, Kimura S, Engelhardt RT, Smith BR, Minoo P. Inhibition of distal lung morphogenesis in Nkx2.1(-/-) embryos. *Dev Dyn* 2000;217:180-90.
54. Kimura S, Hara Y, Pineau T, Fernandez-Salguero P, Fox CH, Ward JM, et al. The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev* 1996;10:60-9.
55. Ikeda K, Clark JC, Shaw-White JR, Stahlman MT, Boutell CJ, Whitsett JA. Gene structure and expression of human thyroid transcription factor-1 in respiratory epithelial cells. *J Biol Chem* 1995;270:8108-14.
56. Kwei KA, Kim YH, Girard L, Kao J, Pacyna-Gengelbach M, Salari K, et al. Genomic profiling identifies TTF1 as a lineage-specific oncogene amplified in lung cancer. *Oncogene* 2008;27:3635-40.
57. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhi R, et al. Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007;450:893-8.
58. Tanaka H, Yanagisawa K, Shinjo K, Taguchi A, Maeno K, Tomida S, et al. Lineage-specific dependency of lung adenocarcinomas on the lung development regulator TTF-1. *Cancer Res* 2007;67:6007-11.
59. Winslow MM, Dayton TL, Verhaak RG, Kim-Kiselak C, Snyder EL, Feldser DM, et al. Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* 2011;473:101-4.
60. Park IW, Wistuba II, Maitra A, Milchgrub S, Virmani AK, Minna JD, et al. Multiple clonal abnormalities in the bronchial epithelium of patients with lung cancer. *J Natl Cancer Inst* 1999;91:1863-8.
61. Wistuba II, Behrens C, Virmani AK, Mele G, Milchgrub S, Girard L, et al. High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. *Cancer Res* 2000;60:1949-60.
62. Wistuba II, Behrens C, Virmani AK, Milchgrub S, Syed S, Lam S, et al. Allelic losses at chromosome 8p21-23 are early and frequent events in the pathogenesis of lung cancer. *Cancer Res* 1999;59:1973-9.
63. Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, et al. Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc Natl Acad Sci U S A* 1998;95:11891-6.
64. Merrick DT, Haney J, Petrunich S, Sugita M, Miller YE, Keith RL, et al. Overexpression of vascular endothelial growth factor and its receptors in bronchial dysplasia demonstrated by quantitative RT-PCR analysis. *Lung Cancer* 2005;48:31-45.
65. Keith RL, Miller YE, Gemmill RM, Drabkin HA, Dempsey EC, Kennedy TC, et al. Angiogenic squamous dysplasia in bronchi of individuals at high risk for lung cancer. *Clin Cancer Res* 2000;6:1616-25.
66. Hirsch FR, Prindiville SA, Miller YE, Franklin WA, Dempsey EC, Murphy JR, et al. Fluorescence versus white-light bronchoscopy for detection of preneoplastic lesions: a randomized study. *J Natl Cancer Inst* 2001;93:1385-91.
67. Hirsch FR, Lippman SM. Advances in the biology of lung cancer chemoprevention. *J Clin Oncol* 2005;23:3186-97.
68. Khuri FR, Cohen V. Molecularly targeted approaches to the chemoprevention of lung cancer. *Clin Cancer Res* 2004;10:4249s-53s.
69. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet* 2009;41:1238-42.
70. Hussenet T, Dali S, Exinger J, Monga B, Jost B, Dembele D, et al. SOX2 is an oncogene activated by recurrent 3q26.3 amplifications in human lung squamous cell carcinomas. *PLoS One* 2010;5:e8960.
71. Lu Y, Futtner C, Rock JR, Xu X, Whitworth W, Hogan BL, et al. Evidence that SOX2 overexpression is oncogenic in the lung. *PLoS One* 2010;5:e11022.
72. Xiang R, Liao D, Cheng T, Zhou H, Shi Q, Chuang TS, et al. Down-regulation of transcription factor SOX2 in cancer stem cells suppresses growth and metastasis of lung cancer. *Br J Cancer* 2011;104:1410-7.
73. Yuan P, Kadara H, Behrens C, Tang X, Woods D, Solis LM, et al. Sex determining region Y-Box 2 (SOX2) is a potential cell-lineage gene highly expressed in the pathogenesis of squamous cell carcinomas of the lung. *PLoS One* 2010;5:e9112.
74. McCaughan F, Pole JC, Bankier AT, Konfortov BA, Carroll B, Falzon M, et al. Progressive 3q amplification consistently targets SOX2 in preinvasive squamous lung cancer. *Am J Respir Crit Care Med* 2010;182:83-91.
75. Nakachi I, Rice JL, Coldren CD, Edwards MC, Stearman RS, Glidewell SC, et al. Application of SNP microarrays to the genome-wide analysis of chromosomal instability in premalignant airway lesions. *Cancer Prev Res* 2014;7:255-65.
76. Ooi AT, Gower AC, Zhang KX, Vick JL, Hong L, Nagao B, et al. Molecular profiling of premalignant lesions in lung squamous cell carcinomas identifies mechanisms involved in stepwise carcinogenesis. *Cancer Prev Res* 2014;7:487-95.
77. Wistuba II, Berry J, Behrens C, Maitra A, Shivapurkar N, Milchgrub S, et al. Molecular changes in the bronchial epithelium of patients with small cell lung cancer. *Clin Cancer Res* 2000;6:2604-10.
78. Meuwissen R, Linn SC, Linnoila RI, Zevenhoven J, Mooi WJ, Berns A. Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model. *Cancer Cell* 2003;4:181-9.
79. Park KS, Martelotto LG, Peifer M, Sos ML, Karnezis AN, Mahjoub MR, et al. A crucial requirement for Hedgehog signaling in small cell lung cancer. *Nat Med* 2011;17:1504-8.
80. Watkins DN, Berman DM, Burkholder SG, Wang B, Beachy PA, Baylin SB. Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. *Nature* 2003;422:313-7.
81. Peifer M, Fernandez-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, et al. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat Genet* 2012;44:1104-10.
82. Byers LA, Wang J, Nilsson MB, Fujimoto J, Saintigny P, Yordy J, et al. Proteomic profiling identifies dysregulated pathways in small cell lung cancer and novel therapeutic targets including PARP1. *Cancer Discov* 2012;2:798-811.
83. Cardnell RJ, Feng Y, Diao L, Fan YH, Masrorpour F, Wang J, et al. Proteomic markers of DNA repair and PI3K pathway activation predict response to the PARP inhibitor BMN 673 in small cell lung cancer. *Clin Cancer Res* 2013;19:6322-8.
84. Mohammad HP, Smitheman KN, Kamat CD, Soong D, Federowicz KE, Van Aller GS, et al. A DNA hypomethylation signature predicts antitumor activity of LSD1 inhibitors in SCLC. *Cancer Cell* 2015;28:57-69.
85. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* 1953;6:963-8.
86. Braakhuis BJ, Tabor MP, Kummer JA, Leemans CR, Brakenhoff RH. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. *Cancer Res* 2003;63:1727-30.
87. Gomperts BN, Spira A, Massion PP, Walser TC, Wistuba II, Minna JD, et al. Evolving concepts in lung carcinogenesis. *Sem Respir Crit Care Med* 2011;32:32-43.
88. Graham TA, McDonald SA, Wright NA. Field cancerization in the GI tract. *Future Oncol* 2011;7:981-93.
89. Auerbach O, Stout AP, Hammond EC, Garfinkel L. Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. *N Engl J Med* 1961;265:253-67.
90. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol* 2007;8:R201.
91. Hackett NR, Heguy A, Harvey BG, O'Connor TP, Luettich K, Flieder DB, et al. Variability of antioxidant-related gene expression in the airway epithelium of cigarette smokers. *Am J Respir Cell Mol Biol* 2003;29:331-43.
92. Mao L, Lee JS, Kurie JM, Fan YH, Lippman SM, Lee JJ, et al. Clonal genetic alterations in the lungs of current and former smokers. *J Natl Cancer Inst* 1997;89:857-62.
93. Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, et al. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci U S A* 2004;101:10143-8.
94. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 2007;13:361-6.
95. Wistuba II, Lam S, Behrens C, Virmani AK, Fong KM, LeRiche J, et al. Molecular damage in the bronchial epithelium of current and former smokers. *J Natl Cancer Inst* 1997;89:1366-73.
96. Gomperts BN, Walser TC, Spira A, Dubinett SM. Enriching the molecular definition of the airway "field of cancerization:" establishing new paradigms for the patient at risk for lung cancer. *Cancer Prev Res* 2013;6:4-7.

97. Kadara H, Fujimoto J, Yoo SY, Maki Y, Gower AC, Kabbout M, et al. Transcriptomic architecture of the adjacent airway field cancerization in non-small cell lung cancer. *J Natl Cancer Inst* 2014;106:dju004.
98. Nelson MA, Wymer J, Clements N, Jr. Detection of K-ras gene mutations in non-neoplastic lung tissue and lung cancers. *Cancer Lett* 1996;103:115–21.
99. Franklin WA, Gazdar AF, Haney J, Wistuba II, La Rosa FG, Kennedy T, et al. Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. *J Clin Invest* 1997;100:2133–7.
100. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Res* 2002;62:2370–7.
101. Soria JC, Rodriguez M, Liu DD, Lee JJ, Hong WK, Mao L. Aberrant promoter methylation of multiple genes in bronchial brush samples from former cigarette smokers. *Cancer Res* 2002;62:351–5.
102. Zochbauer-Muller S, Lam S, Toyooka S, Virmani AK, Toyooka KO, Seidl S, et al. Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. *Int J Cancer* 2003;107:612–6.
103. Gesthalter YB, Vick J, Steiling K, Spira A. Translating the transcriptome into tools for the early detection and prevention of lung cancer. *Thorax* 2015;70:476–81.
104. Gustafson AM, Soldi R, Anderlind C, Scholand MB, Qian J, Zhang X, et al. Airway PI3K pathway activation is an early and reversible event in lung cancer development. *Sci Transl Med* 2010;2:26ra5.
105. Shaykhiev R, Wang R, Zwick RK, Hackett NR, Leung R, Moore MA, et al. Airway basal cells of healthy smokers express an embryonic stem cell signature relevant to lung cancer. *Stem cells* 2013;31:1992–2002.
106. Sridhar S, Schembri F, Zeskind J, Shah V, Gustafson AM, Steiling K, et al. Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* 2008;9:259.
107. Zhang X, Sebastiani P, Liu G, Schembri F, Dumas YM, Langer EM, et al. Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium. *Physiol Genomics* 2010;41:1–8.
108. Perdomo C, Spira A, Schembri F. MiRNAs as regulators of the response to inhaled environmental toxins and airway carcinogenesis. *Mutat Res* 2011;717:32–7.
109. Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, et al. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. *Proc Natl Acad Sci U S A* 2009;106:2319–24.
110. Perdomo C, Campbell JD, Gerrein J, Tellez CS, Garrison CB, Walser TC, et al. MicroRNA 4423 is a primate-specific regulator of airway epithelial cell differentiation and lung carcinogenesis. *Proc Natl Acad Sci U S A* 2013;110:18946–51.
111. Kadara H, Shen L, Fujimoto J, Saintigny P, Chow CW, Lang W, et al. Characterizing the molecular spatial and temporal field of injury in early-stage smoker non-small cell lung cancer patients after definitive surgery by expression profiling. *Cancer Prev Res* 2013;6:8–17.
112. Maki Y, Fujimoto J, Lang W, Xu L, Behrens C, Wistuba II, et al. LAPTM4B is associated with poor prognosis in NSCLC and promotes the NRF2-mediated stress response pathway in lung cancer cells. *Sci Rep* 2015;5:13846.
113. Rivera MP, Mehta AC, Wahidi MM. Establishing the diagnosis of lung cancer: diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest* 2013;143:e142S–65S.
114. Wang Memoli JS, Nietert PJ, Silvestri GA. Meta-analysis of guided bronchoscopy for the evaluation of the pulmonary nodule. *Chest* 2012;142:385–93.
115. Whitney DH, Elashoff MR, Porta-Smith K, Gower AC, Vachani A, Ferguson JS, et al. Derivation of a bronchial genomic classifier for lung cancer in a prospective study of patients undergoing diagnostic bronchoscopy. *BMC Med Genomics* 2015;8:18.
116. Silvestri GA, Vachani A, Whitney D, Elashoff M, Porta Smith K, Ferguson JS, et al. A bronchial genomic classifier for the diagnostic evaluation of lung cancer. *N Engl J Med* 2015;373:243–51.
117. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 2010;11:685–96.
118. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012;489:519–25.
119. Kim Y, Hammerman PS, Kim J, Yoon JA, Lee Y, Sun JM, et al. Integrative and comparative genomic analysis of lung squamous cell carcinomas in East Asian patients. *J Clin Oncol* 2014;32:121–8.
120. Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, et al. Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. *Cancer Prev Res* 2011;4:803–17.
121. de Bruin EC, Taylor TB, Swanton C. Intra-tumor heterogeneity: lessons from microbial evolution and clinical implications. *Genome Med* 2013;5:101.
122. Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet* 2012;13:795–806.
123. Nowell PC. The clonal evolution of tumor cell populations. *Science* 1976;194:23–8.
124. Gorunova L, Hoglund M, Andren-Sandberg A, Dawiskiba S, Jin Y, Mitelman F, et al. Cytogenetic analysis of pancreatic carcinomas: intratumor heterogeneity and nonrandom pattern of chromosome aberrations. *Genes Chromosomes Cancer* 1998;23:81–99.
125. Teixeira MR, Heim S. Cytogenetic analysis of tumor clonality. *Adv Cancer Res* 2011;112:127–49.
126. Navin NE, Hicks J. Tracing the tumor lineage. *Mol Oncol* 2010;4:267–83.
127. Parker JS, Perou CM. Tumor heterogeneity: focus on the leaves, the trees, or the forest? *Cancer Cell* 2015;28:149–50.
128. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multi-region sequencing. *Science* 2014;346:256–9.
129. LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res* 2009;37:4181–93.
130. Reuter JA, Spacek DV, Snyder MP. High-throughput sequencing technologies. *Molecular cell* 2015;58:586–97.
131. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. *Nat Rev Genet* 2013;14:307–20.
132. Assie G, LaFramboise T, Platzer P, Bertherat J, Stratakis CA, Eng C. SNP arrays in heterogeneous tissue: highly accurate collection of both germline and somatic genetic information from unpaired single tumor samples. *Am J Hum Genet* 2008;82:903–15.
133. Beroukhim R, Lin M, Park Y, Hao K, Zhao X, Garraway LA, et al. Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. *PLoS Comput Biol* 2006;2:e41.
134. Li A, Liu Z, Lezon-Geyda K, Sarkar S, Lannin D, Schulz V, et al. GPHMM: an integrated hidden Markov model for identification of copy number alteration and loss of heterozygosity in complex tumor samples using whole genome SNP arrays. *Nucleic Acids Res* 2011;39:4928–41.
135. Staaf J, Lindgren D, Vallon-Christersson J, Isaksson A, Goransson H, Juliusson G, et al. Segmentation-based detection of allelic imbalance and loss-of-heterozygosity in cancer cells using whole genome SNP arrays. *Genome Biol* 2008;9:R136.
136. Vattathil S, Scheet P. Haplotype-based profiling of subtle allelic imbalance with SNP arrays. *Genome Res* 2013;23:152–8.
137. Xia R, Vattathil S, Scheet P. Identification of allelic imbalance with a statistical model for subtle genomic mosaicism. *PLoS Comput Biol* 2014;10:e1003765.
138. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995;12:921–7.
139. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68:978–89.
140. Baugher JD, Baugher BD, Shirley MD, Pevsner J. Sensitive and specific detection of mosaic chromosomal abnormalities using the parent-of-origin-based detection (POD) method. *BMC Genomics* 2013;14:367.
141. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. *Cell* 2012;149:994–1007.
142. Aberle DR, Abtin F, Brown K. Computed tomography screening for lung cancer: has it finally arrived? Implications of the national lung screening trial. *J Clin Oncol* 2013;31:1002–8.
143. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.