

REVIEW

Open Access

Bridging the clinical gaps: genetic, epigenetic and transcriptomic biomarkers for the early detection of lung cancer in the post-National Lung Screening Trial era

John F Brothers^{1,2*†}, Kahkeshan Hijazi^{1,2†}, Celine Mascaux³, Randa A El-Zein⁴, Margaret R Spitz⁴ and Avrum Spira^{1,2}

Abstract

Lung cancer is the leading cause of cancer death worldwide in part due to our inability to identify which smokers are at highest risk and the lack of effective tools to detect the disease at its earliest and potentially curable stage. Recent results from the National Lung Screening Trial have shown that annual screening of high-risk smokers with low-dose helical computed tomography of the chest can reduce lung cancer mortality. However, molecular biomarkers are needed to identify which current and former smokers would benefit most from annual computed tomography scan screening in order to reduce the costs and morbidity associated with this procedure. Additionally, there is an urgent clinical need to develop biomarkers that can distinguish benign from malignant lesions found on computed tomography of the chest given its very high false positive rate. This review highlights recent genetic, transcriptomic and epigenomic biomarkers that are emerging as tools for the early detection of lung cancer both in the diagnostic and screening setting.

Keywords: Biomarker, Diagnostics, Early detection, Epigenetics, Genetics, Lung cancer, Screening, Transcriptomics

Review

Introduction

Lung cancer is the leading cause of cancer death in both men and women in the US and the world, causing more than 1 million deaths per year [1-4]. The global cancer burden in annual cases is projected to double by 2050, and lung cancer is expected to remain the leading cause of all cancer deaths during that time. Cigarette smoke remains the main risk factor for lung cancer, with 85% to 90% percent of lung cancer cases in the US occurring in current or former smokers. However, only 10% to 20% of heavy smokers develop lung cancer [5]. While smoking cessation gradually reduces the risk of lung cancer, the majority of new lung cancer cases occur in former smokers. The high mortality in patients with

lung cancer (80% to 85% in five years) results in part from our inability to predict which of the 100 million current and former smokers in the US are at greatest risk for developing lung cancer, and from the lack of effective tools to diagnose the disease at an early stage [6].

Recent results published from the National Lung Screening Trial have shown that screening high-risk smokers (based on age and cumulative exposure to tobacco smoke) with low-dose helical computed tomography (CT) can lead to a reduction in both lung cancer mortality (by 20.0%) and all-cause mortality (by 6.7%) compared to standard radiographic screening. While this landmark study is already impacting CT screening guidelines and practices across the US, there were a number of important caveats to the study, including the age (55 to 75 years old) and smoking exposure cutoffs (>30 pack-years) chosen for patient inclusion, and the duration of annual CT screening (three years). Importantly, 39.1% of all participants in the low-dose CT arm of the trial had at least one positive screen for lung cancer, and 96.4% of these initial positive screenings represented

* Correspondence: jb2@bu.edu

†Equal contributors

¹Bioinformatics Program, Boston University, Boston, MA, USA

²Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA

Full list of author information is available at the end of the article

false positives for lung cancer [7]. This overabundance of false positives could lead to higher screening costs and unnecessary invasive procedures on many smokers who do not actually have lung cancer. Thus, there is a critical need to develop biomarkers that can determine which of the frequently detected lung nodules on CT scan are malignant (that is, diagnostic markers), and to further define the large high-risk population that would be eligible for screening by CT to increase the efficacy of screening and to reduce the cost and morbidity associated with it (that is, screening markers; Figure 1).

The sequencing of the human genome together with the technological advances that enabled this accomplishment have ushered in a new era of molecular biomarker development that promises to help address these unmet needs. This review will summarize recent genetic, transcriptomic and epigenomic biomarkers that are emerging as tools for the early detection of lung cancer (Figure 2), both in the diagnostic and the screening setting (prognostic and predictive biomarkers will not be covered). The review will focus on genome-wide studies in clinical biospecimens (no animal models or cell line studies) that leverage these emerging high-throughput technologies, and will review commonality of variants between lung cancer and chronic obstructive airways disease. Although there are a number of promising metabolic and proteomic biomarkers for early lung cancer detection, these fall outside the scope of this review [8].

Genome-wide association studies to identify genetic risk factors for lung cancer

Initial genome-wide associations in lung cancer robustly implicated SNPs (Table 1) spanning the chromosome 15q25 region encoding the gene cluster of nicotinic receptors, *CHRNA3/A5/B4* [9-12]. Subsequent multi-

investigator consortia analyses confirmed the association of SNPs spanning this region with heavy smoking, nicotine dependence, craving and related endophenotypes [11,13,14]. Saccone *et al.* [13] conducted a meta-analysis across 34 datasets of European-ancestry participants (Table 1), including a diverse group of 38,617 smokers, and demonstrated that rs16969968, a nonsynonymous coding polymorphism of the *CHRNA5* gene, correlated highly significantly with smoking behavior (odds ratio = 1.33, $P = 5.96 \times 10^{-31}$). Three other large smoking genetics consortia confirmed this locus as that most associated with smoking quantity [11,14,15].

Therefore, the challenging question was the degree to which the associations between these chromosome 15q25 variants and lung cancer were due to their effects on smoking intensity, rather than a direct carcinogenic effect. The lung cancer association, though statistically robust, and initially not altered by adjustment for smoking, increasingly appears to be mediated through smoking. However, there is still uncertainty as to the degree with which the association for lung cancer is mediated through genetic risk beyond that attributed to smoking intensity. Saccone *et al.* [13] showed that locus 1 was associated with lung cancer even when controlling for amount smoked per day (odds ratio = 1.31, $P = 1.99 \times 10^{-21}$), suggesting possible direct genetic effects of locus 1 on this cancer, at least in the presence of smoking. Spitz *et al.* [16] noted that the lung cancer risk associated with the variant genotype was highest in the lightest smokers (<20 cigarettes per day) and younger patients (<61 years), arguing a role for genetic susceptibility in these lesser-exposed groups. Furthermore, they [16] were not able to implicate this locus as a risk factor in other smoking-related cancers (bladder and renal), suggesting genetic effects on both smoking behavior and lung cancer risk.

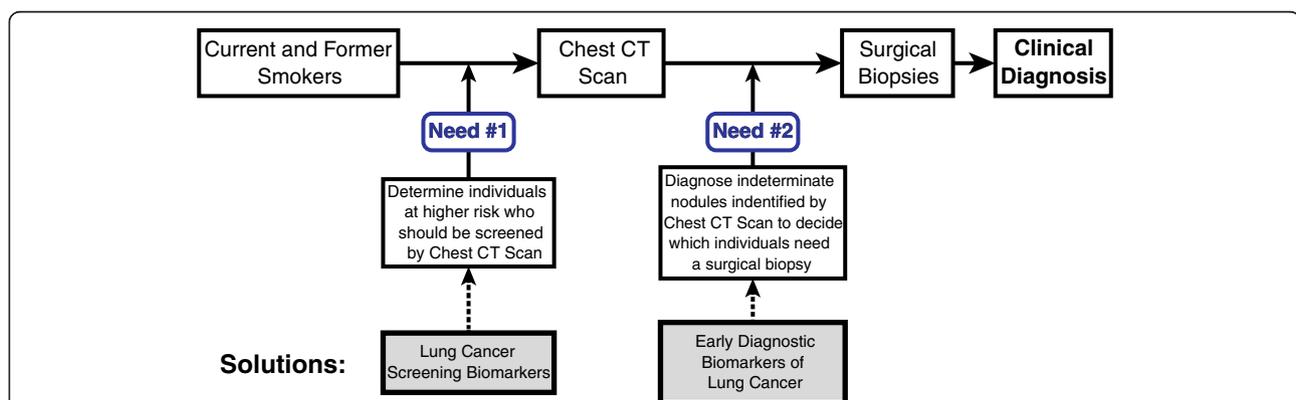


Figure 1 An overview of clinically unmet needs that exist following the National Lung Screening Trial. While there is a reduction in both lung cancer mortality and all-cause mortality when using low-dose CT, there are still two major unmet needs highlighted by the trial. The first is the need to limit the number of people who are screened with low-dose CT to those with the highest risks. Genetic, transcriptomic and epigenetic screening biomarkers could meet this need by identifying smokers with the highest likelihood of developing lung cancer. The second unmet need comes from the high number of nodules identified by CT, which are false positives for lung cancer. Early diagnostic biomarkers could play a key role in identifying which nodules are likely to be cancerous before sending patients into surgery.

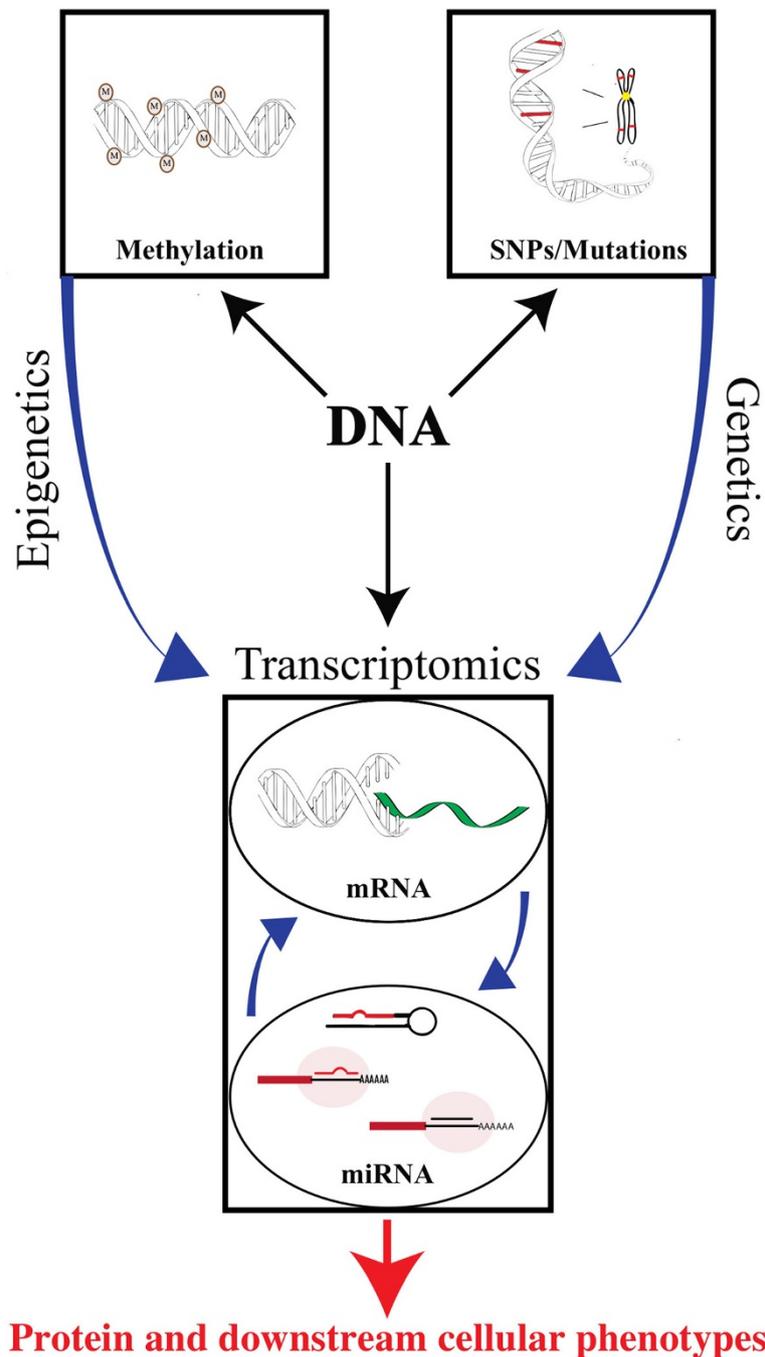


Figure 2 Biological rationale for addressing clinical issues by using upstream early events that ultimately lead to lung cancer phenotypes as genomic biomarkers. The diagram highlights early upstream markers for diagnosing or screening of lung cancer far in advance of the development of clinically evident invasive carcinomas, which are mainly driven by genetic, epigenetic and transcriptomic damage.

Wang *et al.* [17] demonstrated that each copy of chromosome 15q risk alleles was associated with increased cigarette consumption of 1.0 cigarette per day at rs12914385 and 0.9 cigarettes per day at rs8042374 with, and concluded that these modest differences in smoking behavior were sufficient to account for the 15q25 association with lung cancer risk. However, it could also be

argued that cigarettes per day is not a sufficient proxy for carcinogen exposure [18].

Truong *et al.* [19] used data from 21 case-control studies (nine in North America, eight in Europe and four in Asia) and replicated the association between chromosome 15q25 SNPs and lung cancer risk (Table 1) in white ever smokers (odds ratio = 1.26, CI:1.21-1.32,

Table 1 Regions and genes associated with lung cancer and/or chronic obstructive pulmonary disease

Reference	Gene	Chromosome region	Population size	Platform
[9]	<i>CHRNA5</i> , <i>CHRNA3</i> and <i>CHRN4</i> , <i>HTERT</i> , <i>CLPTM1L</i>	15q24-25.1 5p15.33	Discovery 1,154 Cases 1,137 Replication two sets- Texas 711/632 and UK 2,013/3,062 Caucasian	Illumina Hapmap 300
[10]	<i>CHRNA5</i> , <i>CHRNA3</i> and <i>CHRN4</i> <i>HTERT</i> , <i>CLPTM1L</i>	15q24-25.1 5p15.33	Discovery 1,989/2,513/4,752 Caucasian	Illumina Hapmap 300
[11]	<i>CHRNA5</i> , <i>CHRNA3</i> and <i>CHRN4</i>	15q24-25.1	1,024/32,244 Caucasian	Illumina (Human Hap300 and Human Hap300-duo + Bead Arrays, Illumina)
[12]	<i>CHRNA5</i> , <i>CHRNA3</i> and <i>CHRN4</i> <i>HTERT</i> , <i>CLPTM1L</i>	15q25, 5p15, and 6p21	5,739/5,848; Meta-analysis 7,561/13,818 Caucasian	550 K, 610QUAD4, 317 K + 540S
[32]	<i>RAD52</i>	12p13.33	5,355/4,344 replication 3,359 squamous cell /9,100 Caucasian	Variety of platforms 550, 300, Infinium AB7900 7,700/5,914
[13]	<i>CHRNA5</i> , <i>CHRNA3</i> and <i>CHRN4</i>	15q25	7,700/5,914 Caucasian	PCR 7,700/5,914
[19]	<i>CHRNA5</i> , <i>CHRNA3</i> and <i>CHRN4</i> <i>HTERT</i> , <i>CLPTM1L</i>	15q25 5p15	11,645/14,954 Caucasian and Asian	Illumina Omni1-Quad and OmniExpress chips
[27]	<i>CHRNA5</i> , <i>CHRNA3</i> and <i>CHRN4</i> <i>HTERT</i> , <i>CLPTM1L</i>	15q25 5p15	1,094/1,100 Korean	PCR 1,094/1,100
[29]	<i>TP63</i> <i>TERT-CLPTM1L</i> <i>MIPEP-TNFRSF19</i> <i>MTMR3-HORMAD2-LIF</i>	3q285 p15.33 13q12.12 22q12.2	Discovery 2,331/3,077 Replication 6,313/ 6,409 Chinese	--
[30]	<i>GATA3</i> <i>CYP24A1</i> <i>PPP2R2B-STK32A-DPYSL3</i> <i>IL3-CSF2-P4HA2</i> <i>AJAP1-NPHP4</i>	10p14 20q13.2 5q32 5q31.1 1p36.32	Discovery 2,331/3,077 Validation 7,436/ 7,483 Chinese	Affymetrix SNP Array 6.0 TaqMan, iPLEX Sequenom MassARRAY
[31]	<i>LRFN2</i> <i>SP4</i> and <i>DNAH11</i>	6p21.17p15.3	Discovery 2,331/4,006 Replication 2,665/11,436	Affymetrix SNP Array 6.0 chips

P-trend = $2 \times 10^{(-26)}$ and also confirmed that this association was higher at younger age of onset (P -trend = 0.002), whereas no association was found in never smokers or in Asian participants. Spitz *et al.* [16] found no elevated risk associated with these variants in over 547 lifetime never smoking patients with lung cancer. Subsequent meta-analyses of never smokers with lung cancer (Galvan and Dragani [20] in >1,000 never smokers and >1,800 controls; and Wang *et al.* [17] in 2,405 patients and 7,622 controls) replicated the lack of any statistically significant association with this locus in never smokers.

Other top hits identified in the GWAS have also been replicated. A number of well-designed GWAS and meta-analyses have implicated variants at the 5p15.33 locus in cancer risk at several different sites, including lung cancer in both white and Asian patients [21]. Truong *et al.*

[19] confirmed the significant association in white patients for rs2736100 in the chromosome 5p15 locus. Both Truong *et al.* [19] and Landi *et al.* [12] noted a histology-specific role of rs2736100 in adenocarcinoma. This locus was also recently implicated in lung cancer risk in African American patients [22]. There is biologic plausibility for this finding because mean relative telomere length has been associated with four genetic variants of the *hTERT* gene, including rs2736100 [23], and *TERT* gene amplification is responsible for *TERT* mRNA overexpression in a majority of lung adenocarcinomas [24]. Cleft lip and palate transmembrane protein 1-like (*CLPTM1L*) gene also resides in this region of chromosome 5 for which copy number gain has been found to be the most frequent genetic event in early stages of non-small cell lung cancer. James *et al.* [25] demonstrated

increased CLPTM1L expression in lung adenocarcinomas and protection from genotoxic stress-induced apoptosis and concluded that anti-apoptotic CLPTM1L function could be another mechanism of susceptibility to lung tumorigenesis. A third region implicated by GWAS in susceptibility to lung cancer in Caucasians is the human leukocyte antigen region at chromosome 6p21 [10,26].

The association with SNPs in the 5p15 and 15q25 regions was confirmed in a Korean population with similar magnitude of effect as reported for other ethnic groups, but there was no association with the 6p locus [27]. Likewise, the effect of the 5p15 SNP was significant only for adenocarcinoma. Truong *et al.* [19] noted no effect for the chromosome 15q locus, but replicated the association with the 5p locus in an Asian population. A Japanese study [28] confirmed the finding at 5p15.33. There have been several GWAS in Chinese populations. Hu *et al.* [29] replicated findings of significance in both 3q28 (*TP53*) and at the 5p13 locus (Table 1). They also reported significance at two additional loci, 12q12 and 22q12 (Table 1). In an attempt to identify additional susceptibility loci in Chinese patients with lung cancer, Dong *et al.* [30] reported genome-wide significance for three additional lung cancer susceptibility loci at 10p14 (close to *GATA3*), 5q32 in *PPP2R2B-STK32A-DPYSL3*, and 20q13.2 in *CYP24A1*. They also found additional associations for rs247008 at 5q31.1 (*IL3-CSF2-P4HA2*), and rs9439519 at 1p36.32 (*AJAPI-NPHP4*). There was suggestive evidence for interactions with smoking dose. Jin *et al.* [31] noted that genetic variants at 6p21.1 and 7p15.3 were associated with risk of multiple cancers in Han Chinese patients, including lung cancer. Finally, Shi *et al.* [32] reported that a locus on *RAD52*, involved in DNA double-strand break repair and homologous recombination, influenced risk of squamous cell lung cancer but not other cell types.

It is likely that many more common variants can be anticipated to contribute to lung cancer risk, although with effect sizes too small to reach significance in genome-wide analyses. It has been argued that there are diminishing returns in predicting disease risk from common marker SNPs, and greater effort should be spent investigating functional relevance of the GWAS findings. For example, evaluating the effect that SNP variation has upon expression and activity of nicotinic receptors can be explored by taking advantage of animal and cellular models of *CHRNA3* and *CHRNA5* knock-out animals [33,34]. Studies of cell lines and primary lung cancers can provide insights into the effects of these variants on proliferation and apoptosis; one such study suggested a role of a proteasome gene in this region beyond the effects of nicotinic receptors [35]. Emerging metabolomic markers may provide useful biomarker dosimeters of smoking damage relative to carcinogenesis. Certainly, multiple strategies are needed to further tease apart these complex relationships [18].

Overlap in genetic risk factors for lung cancer and chronic obstructive pulmonary disease

Lung cancer and chronic obstructive pulmonary disease (COPD) result from the combined effects of smoking exposure and genetic susceptibility. Tobacco smoke exposure has been responsible for 80% of lung cancers, however only 15% to 20% of chronic smokers develop lung cancer or COPD. Approximately 50% to 90% of smokers with lung cancer also have COPD. Studies have shown that COPD is an independent risk factor for lung cancer among Caucasians and African Americans, conferring a four- to six-fold increased risk. Over the past few years, several lung cancer risk models have been developed [36-40], some of which included pulmonary diseases such as COPD and pneumonia. Consistently, the inclusion of COPD in the models leads to improvement of the discriminatory power and good calibration [41]. The model with the highest discriminatory power reported to date is the extended Prostate, Lung, Colorectal and Ovarian lung cancer risk model [37], which also includes COPD. This dual susceptibility indicates a link between the processes that induce COPD and lung cancer.

Results from recent GWAS suggest a possible overlap in the genetic risk factors predisposing smokers to lung cancer and COPD. Several regions in the genome associated with lung cancer and/or COPD have been identified, including chromosome 1q21, 4q22, 4q24, 4q31, 5p15, 5q32, 6p21, 6q24, 15q25 and 19q13 [9,10,41-49]. Several important genes mapping to those regions have also been identified as significant players in the pathogenesis of lung cancer and/or COPD (Table 1), and many of these loci overlap. For example, a variant in the *FAM13A* gene has been reported to have a protective effect in COPD and lung cancer [49]. *CHRNA3/5* (15q25) was reported to be associated with both COPD and lung cancer [10,48,49] through its effects on both smoking exposure and COPD. Using mediation analysis, Wang *et al.* [50] reported that COPD is a mediating phenotype that could partially explain the effect of smoking exposure on lung cancer. These findings suggest the presence of shared susceptibility mechanisms for these two smoking-related diseases. Such susceptibility may also be mediated through receptors expressed on the bronchial epithelium that implicate molecular pathways underlying both COPD and lung cancer [51]. To date, most of the lung cancer and COPD genetic studies have been conducted independent of each other, which has contributed to the mediating effect of one disease over the other being overlooked [52].

Epigenetic screening and diagnostic markers for lung cancer

Epigenetics is classically defined as the study of changes in downstream phenotypes or gene expression that cannot

be attributed to changes in DNA and is heritable. Another refined definition is that epigenetics concerns structural changes in chromosomal regions that are not related to changes in DNA that mark altered activity states [53]. Two major types of epigenetic regulation are DNA methylation and histone modification, both of which are known to modulate gene expression. Given that the abundance of molecular biomarkers in this field have been DNA methylation-based, this section will focus on DNA methylation studies that hold the potential to impact early lung cancer detection (Table 2a).

DNA methylation is an epigenetic mechanism marked by the joining of a methyl group to a cytosine base to form 5-methylcytosine, typically at a CpG dinucleotide near or within a CpG island. When CpG dinucleotides are methylated to a high degree in the promoter region of a gene, that gene's expression is usually down-regulated as a result. This is one way that cells can regulate which genes are expressed (Figure 2) and is a mechanism utilized during cell and tissue differentiation during development [54]. Aberrant hypermethylation of oncogenes or hypomethylation of tumor suppressor genes (Table 2a) is one way that transcriptional regulation can spiral out of control in cancer cells [55].

Genome-wide methylation profiling has been used to identify altered methylation patterns in lung cancer tissue (including genes such as *CDKN2A*, *RASSF1A*, *ARHI*, *MGMT* and *RARβ*) [56,57], but so far only one larger scale study has shown the possibilities of identifying methylation biomarkers for the diagnostic or screening setting in noninvasive biospecimens utilizing microarray-based technologies. In this study, nine CpGs were able to discriminate between lung cancer cases and controls with an area under the receiver operator characteristic curve (AUC) of 0.86 [58]. The vast majority of current methylation studies that could be useful for screening and diagnostic tests remain at a candidate gene or gene panel level analysis (Table 2a).

Belinsky et al. [59] originally identified the hypermethylation of *CDKN2A* in lung tumors but within the same study also examined the sputum of 33 people who smoked. In this small initial study, eight patients had sputum with methylated *CDKN2A* detected by methylation-specific polymerase chain reaction (MSP). Of those, three were diagnosed with lung cancer at the time of sputum collection and one other would develop lung cancer a year later [59]. Work on identifying *CDKN2A*, as well as *MGMT*, as a measure of cancer risk and diagnosis was expanded in a 21-patient study of matched sputum and squamous cell carcinoma (SCC) samples as well as sputum samples from 32 patients evaluated for possible lung cancer. This study was able to significantly improve cancer detection and risk using the methylation status of the two genes compared to cytology alone, with

100% of patients with SCC displaying methylation of one or both of these genes. More importantly, these genes were aberrantly methylated up to three years before diagnosis [60]. By looking at the sputum of lung cancer-surviving smokers, cancer-free smokers and never smokers, then adjusting for age and smoking duration, *MGMT*, *RASSF1A*, *DAPK* and *PAX5α* were also identified as being significantly differently methylated in the lung cancer survivors. This indicates that aberrant methylation of a panel of candidate genes could identify patients with higher risk of lung cancer (lung cancer-surviving smokers had a 6.2-fold higher odds of having three or more of these genes methylated in sputum) [61]. Other genes that have been identified in the sputum with aberrant methylation associated with increased risk for lung cancer include *ASC/TMS1* (increased odds in cancer patients from 7.2 to 28.6) [62], *GATA4*, *GATA5* and *PAX5β* (6.5-fold increase in cancer risk with methylation of three or more genes) [63]. Recently, a larger panel of 31 genes in the sputum was used to identify signatures of stage I lung cancer. It had >70% accuracy and could predict which smokers had cancer between 3 and 18 months before clinical diagnosis (AUC of 0.71 and 0.77 for the two cohorts in the study) [64].

Other potential distal sites for assessing lung cancer risk using methylation markers include the serum, plasma and blood leukocytes. Based on evidence that DNA from tumor cells can be found freely in circulating serum [65], Esteller et al. [66] examined the serum, normal lung tissue and tumor tissue from 22 patients with non-small cell lung cancer (NSCLC). They found that 73% of patients had serum DNA that reflected hypermethylation events found in their tumors. Specifically using MSP, they looked at methylation of *CDKN2A*, *MGMT*, *DAPK* and *GSTP1*, genes whose aberrant methylation profiles have already been shown to associate with lung cancer risk or diagnosis [66]. A larger study with a cross-section case-control design looked at the serum from 200 patients, 91 of whom had lung cancer, 100 had non-malignant lung disease, and nine had some other malignant disease. *RARβ*, *CDKN2A*, *DAPK*, *RASSF1A* and *MGMT* were examined, and the analysis showed that a patient having methylation of just one gene had an odds ratio of 5.08, meaning they were approximately five times likely to have lung cancer than patients without any methylated genes. This odds ratio increased in patients with two or more genes being aberrantly methylated [67]. Overall, just looking at this limited candidate gene list, almost 50% of patients with lung cancer displayed at least one case of aberrant methylation in their serum. Other genes with aberrant methylation in serum DNA have been found to associate with lung cancer risk, including *TMEFF2* [68], *RUNX3* [69] and *CDH13* [70], suggesting that many genes in the serum could signify

Table 2 Methylation-, gene-expression- and miRNA-based biomarkers for risks and early detection of lung cancer

Reference	Sample type	Genetics/ genomics platform	Clinical settings	Key findings
a) Epigenetic biomarkers				
[59]	Sputum, lung tissue, biopsies	MSP	Lung tissue, precursor lesions and bronchial biopsies from patients with SCC and sputum from individuals with suspicion of lung cancer	CDKN21 hypermethylation more often observed in patients with cancer than with no cancer
[66]	Paired serum and lung tissue	MSP	Lung tissue and serum from patients with NSCLC and control	73% of patients had serum DNA that reflected aberrant methylation in their tumors, specifically in CDKN2A, MGMT, DAPK, GSTP1
[60]	Paired sputum and lung tissue	MSP	Lung tissue and sputum from smokers with SCC	CDKN2A and MGMT were hypermethylated in both sputum and tumor of patients at time of diagnosis
[72]	Bronchial epithelial cells, blood lymphocytes, lung tissue	MSP	Paired blood and bronchial epithelial samples from smokers/non-smokers with pre-neoplastic lesions and neoplastic lesions from individuals with NSCL versus controls	ECAD and DAPK more likely to be methylated in smokers' peripheral lymphocytes or bronchial epithelium and never methylated in non-smokers
[58]	Peripheral blood leukocytes	Illumina Beadchip and Pyrosequencing	Smokers with recently diagnosed SCLC and controls	Forty-three CpG sites were differentially methylated between SCLC and controls, and nine of these, validated by pyrosequencing, could discriminate SCLC with AUC of 0.86
[71]	Paired serum and lung tissue	MSP	Paired serum and lung tissue samples from individuals with lung cancer and controls	Six-gene serum panel that discriminated patients with lung cancer with 75% sensitivity and 73% specificity
b) Transcriptomics biomarkers				
[76]	Bronchial brushing, large airway epithelium	Affymetrix array	Bronchial brushings of cytologically normal large airway epithelium obtained from smokers undergoing bronchoscopy for suspicion of lung cancer	Eighty gene airway biomarker with >80% diagnostic sensitivity and specificity, and 95% sensitivity and negative predictive value when biomarker is combined with cytology collected at bronchoscopy
[78]	Bronchial brushings from normal airway bronchial epithelial cells	(StaRT)-PCR	Normal bronchial epithelial cells of patients with lung cancer and non-lung cancer controls	Fourteen gene airway biomarkers of antioxidant, DNA repair and transcription factor genes with performance in a test AUC >0.84 and an accuracy of 80%
[86]	Peripheral blood mononuclear cells	cDNA array	Blood collection from smokers with newly diagnosed lung cancer confirmed by histopathology	twenty-nine-gene blood signature with >80% sensitivity and specificity
[79]	Bronchial brushing from airway epithelium	Affymetrix array	Bronchial airway brushings of cytologically normal epithelium from smokers with and without lung cancer or premalignancy	Gene-expression signature of PI3K signaling pathway activation was differentially expressed in airways of smokers with lung cancer or dysplasia and was reversible with chemopreventive therapy
[88]	Whole blood	Sentrix whole genome bead chips WG6 (Illumina)	PAX gene-stabilized blood samples from three independent groups consisting of patients with NSCLC and controls	Genes differently expressed in whole blood of patients with NSCLC and controls were used to build a diagnostic classifier with AUC >0.82
[85]	Saliva	Affymetrix array	Whole saliva collected from untreated patients with lung cancer with matched cancer-free controls	Seven highly discriminatory transcriptomic salivary biomarker with AUC = 0.925 with >82% sensitivity and specificity
c) MicroRNA biomarkers				
[93]	Sputum	RT-qPCR	Sputum from patients with squamous lung cancer and healthy controls	Three miRNA diagnosed stage I squamous cell lung cancer with AUC = 0.87
[94]	Sputum	RT-qPCR	Sputum from patients with lung adenocarcinoma and healthy controls	Four miRNA diagnosed stage I lung adenocarcinoma with AUC = 0.90
[107]	Serum	Genoexplorer microRNA expression system	Serum from patients with lung cancer versus healthy controls	Two miRNA discriminated individuals with early stages NSCLC with AUC = 0.77

Table 2 Methylation-, gene-expression- and miRNA-based biomarkers for risks and early detection of lung cancer (Continued)

[103]	Serum	Taqman Low Density Arrays RT-qPCR	Serum from asymptomatic patients with NSCLC and healthy smokers. Patients were screened by low-dose CT and sera were collected at the time of diagnosis before the surgery	Thirty-two miRNA predicted risk of developing lung cancer in asymptomatic high-risk individuals with an accuracy of 80%
[109]	Plasma	Taqman Low Density Arrays RT-qPCR	Multiple plasma samples were collected before and at the time of disease, from two independent spiral CT-screening trials	Fifteen miRNA predicted the risk of lung cancer with AUC = 0.85 and 13 miRNA diagnosed lung cancer in undetermined CT nodules with AUC = 0.88
[110]	Plasma	RT-qPCR	Plasma from patients with lung cancer versus healthy controls	Four miRNAs discriminated patients with NSCLC with AUC = 0.93
[105]	Serum	RT-qPCR	Serum from patients with lung cancer versus healthy controls	Ten miRNAs discriminated patients with NSCLC with AUC = 0.97

lung cancer risk and that a larger profile of aberrant methylation could produce a more accurate biomarker for lung cancer risk. The work by Begum *et al.* [71], who looked at methylation profiles of a slightly larger set of 15 genes and then selected the six most sensitive and specific genes for predicting lung cancer risk (*APC*, *CDHI*, *MGMT*, *DCC*, *RASSF1A* and *AIM1*), clearly shows evidence that a more global methylome approach could lead to a more sensitive (75%) and specific (73%) biomarker of lung cancer risk from serum DNA [71]. Methylation events in plasma, specifically in *CDKN2A*, *MGMT* and *RASSF1A* [61], as well as in peripheral blood leukocytes [58] and lymphocytes [72,73], are promising less invasive sites for assessing lung cancer risk through measuring DNA methylation differences.

Transcriptomic biomarkers for screening and diagnosing lung cancer

Gene-expression profiling or transcriptomics has been used to delineate disease classification, improve diagnostic accuracy, identify new molecular targets for drugs and provide new biological insights into lung cancer. High-throughput technologies, such as microarray, and sequencing platforms allow the measurement of thousands of genes simultaneously, to look for different pattern changes across subsets that help characterize a particular physiological state or clinical phenotype. In this section, we will review the diagnostic and screening transcriptomic biomarkers that have been developed in the airway and blood of at-risk smokers (Table 2b).

Airway-based transcriptomic biomarkers for early detection of lung cancer

A number of transcriptomic biomarkers for the early detection of lung cancer have leveraged the so-called field cancerization or field effect paradigm in which abnormalities in gene expression in the normal bronchial mucosa are shared with those found in the tumor. Two

genome-wide gene-expression profiling studies identified transcriptomic alterations related to smoking that were found both in the cancer and in the normal lung tissue [74,75]. The first study analyzed both lung SCC compared to the normal epithelium of the bronchi and adenocarcinoma as compared to the normal alveolar lung tissue [74]. The second study focused on SCC and normal bronchial epithelium [75]. Abnormalities in the normal bronchial tissue that were similar to those identified in the tumor were seen in tumor suppressor genes and oncogenes, as well as different functions such as xenobiotic metabolism and redox stress, matrix degradation, and cell differentiation.

Based on these studies, a number of groups have been using a comparatively easily available specimen, airway epithelial cells through bronchial brushings, to measure the changes in gene expression associated with lung cancer. An 80 gene-expression-based biomarker was developed in mainstem bronchial airway epithelial cells that can serve as a sensitive and a specific biomarker for diagnosing lung cancer among smokers undergoing bronchoscopy for suspected disease [76]. Importantly, combining the gene-expression biomarker with cytology obtained at bronchoscopy resulted in 95% sensitivity and 95% negative predictive value, enabling the physician to avoid unnecessary further invasive procedures in those smokers without lung cancer. Furthermore, the biomarker was shown to be associated with lung cancer diagnosis independent of clinical and radiographic risk factors for disease, although the study was limited in terms of the clinical and radiographic risk factors that were modeled (for example, COPD positron emission tomography scan results not included) [77]. Later, Blomquist *et al.* also reported that a pattern of antioxidant and DNA repair gene expression in normal airway epithelium was associated with lung cancer [78]. They identified a signature of 14 genes that discriminates cases versus controls with an AUC of 0.84 and an accuracy of 80%.

Beyond diagnosing lung cancer, airway gene expression has also been used to identify molecular pathways that are deregulated in the bronchial airway of smokers with or at risk for lung cancer [79]. A gene-expression signature of phosphoinositide-3-kinase signaling pathway was differentially activated in the cytologically normal bronchial airway of both smokers with lung cancer and smokers with pre-malignant airway lesions [76]. Furthermore, that study found that the PI3K pathway gene-expression signature reverses back to baseline in those patients whose dysplastic lesions regress upon treatment with the candidate lung cancer chemoprophylaxis agent myo-inositol. As airway epithelial cell dysplasia is a pre-neoplastic event in lung carcinogenesis, these data suggest both that PI3K pathway activation is an early and reversible event during lung carcinogenesis and, more broadly, that bronchial airway epithelial cell gene expression reflects carcinogenic processes that precede the development of frank malignancy [79]. This suggests that alterations in airway gene expression are an early and potentially reversible event in the process of lung carcinogenesis that could potentially be used to guide personalized approaches to lung cancer chemoprevention.

Leveraging the microarray dataset of airway epithelium from smokers with and without lung cancer [76], Wang *et al.* [80] provided additional insight into the molecular pathways altered in the airway of smokers with lung cancer. They identified that the antioxidant response pathway, regulated by the transcription factor nuclear factor erythroid-derived 2-like 2, was down-regulated in the airway of smokers with lung cancer. Furthermore, they identified potential polymorphisms in the promoter regions of the antioxidant genes that may associate with decreased airway gene expression in response to tobacco smoke.

With the emergence of next-generation sequencing as a more robust tool for transcriptomic profiling, Beane *et al.* sequenced the RNA from bronchial airway epithelial cell brushings obtained during bronchoscopy from healthy never smokers, current smokers and smokers with and without lung cancer undergoing lung nodule resection surgery [81]. There was a significant correlation between the RNA-sequencing gene-expression data and Affymetrix microarray data generated from the same samples ($P < 0.001$), although the RNA-sequencing data detected additional smoking- and cancer-related transcripts whose expression was not found to be significantly altered when using microarrays.

Over the past several years, a number of studies have attempted to move transcriptomic profiling of the airway in at-risk smokers to biosamples that are less invasive and more easily collected in population-based studies. Two separate groups have demonstrated that the buccal mucosa gene-expression response to smoking mirrors

that seen in the bronchial airway (one study using punch biopsies of the cheek [82] and the second using buccal scrapings [83]). Both studies were limited to healthy smokers and did not assess the relationship of bronchial and buccal gene expression within the same individual. More recently, Zhang *et al.* [84] demonstrated a strongly concordant gene-expression response to smoking in matched nasal and bronchial samples from active smokers. These studies raise the exciting possibility that buccal and nasal swabs could be used as a surrogate to bronchial brushings for a relatively noninvasive screening or diagnostic tool for individual susceptibility to smoking-induced lung diseases. Additionally, Zhang *et al.* [85] profiled salivary transcriptomes of recently diagnosed and untreated smoker and non-smoker patients with lung cancer and matched cancer-free controls. The study led to the discovery of seven highly discriminatory transcriptomic salivary biomarkers with 93.75% sensitivity and 82.81% specificity in the pre-validation sample set. Data suggest that lung cancer transcriptomic biomarker signatures are present in human saliva, which could be clinically used to discriminate patients with lung cancer from cancer-free controls.

Blood-based transcriptomic biomarkers for early detection of lung cancer

Although the development of a gene-expression biomarker in blood that can be collected in a noninvasive manner is highly attractive, studies have been relatively limited by the degradation of circulating mRNA in serum and plasma. However, gene-expression alterations identified in lung tumors have been identified in circulating white blood cells by a number of groups. Showe *et al.* analyzed gene expression in peripheral blood mononuclear cell samples of current or former smokers with histologically diagnosed NSCLC tumors [86]. They identified a 29-gene signature that separates patients with and without lung cancer with 86% accuracy (91% sensitivity, 80% specificity). Accuracy in an independent validation set was 78% (sensitivity of 76% and specificity of 82%). Rotunno *et al.* analyzed gene expression of lung tissue and peripheral whole blood collected using PAXgene blood RNA tubes from patients with adenocarcinoma and controls to identify dysregulated lung cancer genes that could be tested in blood to improve identification of at-risk patients in the future [87]. Zander *et al.* further investigated the validity of whole-blood-based gene-expression profiling for the detection of patients with lung cancer among smokers from three different datasets. They showed that RNA-stabilized whole-blood samples can indeed be used to develop a gene-expression-based classifier that can be used as a biomarker to discriminate between NSCLC patients and controls [88].

miRNA biomarkers for the early detection of lung cancer

MicroRNAs are recently discovered small molecules that play an important role in regulating gene expression. These noncoding RNAs, in their final active form, are usually 22 nucleotides in length and target specific parts or mRNA sequences, usually found in the 3' untranslated regions of mRNA, which either prevent translation or promote mRNA degradation, and lead to down-regulation of specific genes [89]. Because miRNA are relatively more stable than mRNA [90], any miRNA profiles of lung cancer risk or diagnosis are likely to be more accurate when moving from the bench to the clinic. This review will focus on large-scale miRNA studies that have been performed in airway, sputum and blood for early lung cancer detection (Table 2c).

In bronchial tissue

By global profiling of miRNA in pre-malignant airway lesions, 69 miRNA were found to evolve in high-risk patients from a pre-invasive stage to a higher stage in the multistep process of lung carcinogenesis. The expression profiles of 30 and 15 miRNAs were able to discriminate low-grade lesions from high-grade ones including or not invasive carcinoma [91]. While this data suggests that airway miRNA expression may serve as an early detection biomarker, this study was limited to bronchial biopsies of pre-malignant airway lesions, which are relatively invasive. As with the gene-expression studies outlined above, more microRNA profiles in airway epithelial brushings are needed to advance the field.

In sputum

Given the relative stability of miRNA in biological specimens, a number of groups have explored the utility of miRNA-based biomarkers in sputum samples. Xie *et al.* [92] showed that miRNA profiles in the sputum could be used to identify NSCLC. More recently, two studies were also able to identify and distinguish miRNA profiles that could do early detection of SCC [93] or adenocarcinoma [94]. Both studies included a test set and a validation set. A SCC signature of three miRNAs diagnosed the presence of a stage I SCC in patients' sputum with a sensitivity of 73%, a specificity of 96% and an AUC of 0.87 in the test set [93]. The adenocarcinoma signature composed of four miRNA detected patients with stage I adenocarcinoma with a specificity of 81%, a sensitivity of 92% and an AUC of 0.90 [94]. There was no overlap between the two signatures in sputum. In total, seven different miRNAs were identified in these two signatures and these miRNAs could be risk factors for lung cancer and be used to diagnose lung cancer.

In blood

The relative stability of miRNA has prompted numerous groups to explore the potential utility of a blood-based miRNA biomarker for early detection of lung cancer. Ten of these have been specifically looking for circulating miRNA in plasma or serum, whereas five studies have examined miRNA expression profiles in whole blood [95-99].

Among the whole-blood miRNA studies, one study took a candidate approach by analyzing the expression of let-7a in the blood of patients with NSCLC [97]. The other four studies screened the expression of larger panels of miRNA in a small number of patients (range of 10 to 28 patients per group), not including any validation set [95,96,98,99]. These studies identified an miRNA signature discriminating between patients with lung cancer and healthy controls with a sensitivity and specificity ranging from 86% to 98% and from 88% to 100%, respectively, using cross-validation within training set. Keller *et al.* [96] have applied next-generation miRNA-sequencing to whole blood to identify miRNAs associated with lung cancer. Using ultra-deep (approximately 25 million reads per sample of small RNA) sequencing of blood samples from 10 patients with NSCLC and 10 healthy individuals, they were able to identify seven entirely novel miRNAs (not annotated in miRBase at the time) that were significantly altered in patients with cancer [96]. This relatively small study demonstrates the potential resolution that miRNA-sequencing could provide in discovering entirely new biomarkers for lung cancer.

Seven studies analyzed miRNA expression in serum [100-106] and three in plasma [107-109]. Six out of the ten studies included a validation set and four of the same six studies described the performance of the test, that is, sensitivity, specificity and/or AUC [100,101,103,105,107,108]. Notably, only three studies included samples at earlier time points than diagnosis [101,102,107], which is required for evaluating miRNAs as a risk or screening biomarker. Boeri *et al.* identified miRNA signatures that predict lung cancer development and prognosis [107]. They analyzed miRNA expression in 38 patients with lung cancer from the INT-IEO cohort (training set) and 53 from the MILD trial (validation set). With a signature composed of a ratio of 15 miRNAs, they could predict risk of lung cancer in patients with nodules in the CT screening with a sensitivity of 80%, a specificity of 90% and an AUC of 0.85. A signature composed of a ratio of 13 miRNAs was able to diagnose lung cancer in undermined CT-screened lung nodules with a sensitivity of 75%, a specificity of 100% and an AUC of 0.88. The study of Boeri *et al.* [107] is the only work so far directly addressing the role of biomarkers for the work-up of CT-screened nodules. In addition to requiring further

prospective validation, this study might be too complex to apply in practice. Another more recent study by Bianchi *et al.* [101] identified a 34-miRNA profile that could predict which asymptomatic high-risk individuals were likely to develop a lung cancer with an accuracy of 80%. Among the 5,203 high-risk individuals studied, 93 went on to be diagnosed with NSCLC in the first two years of screening. Serum was collected before surgery from 59 of these 93 patients, and serum was also collected from 69 matched control patients who were enrolled in the same study. Using a training set and test set, they were able to identify a 34-miRNA biomarker, one which can better identify lung cancer risk and be more properly used as a screening test [101].

Free circulating DNA biomarkers

Circulating cell-free DNA (cf-DNA) is a known marker of cancer cell death and an increase in the prevalence of cf-DNA in the blood has been used as a marker to distinguish patients with cancer patients from patients with no cancer [110,111]. In this section, we will review recent efforts to utilize cf-DNA as a diagnostic and screening biomarker for differentiating patients with lung cancer from those without, especially some studies that have been able to identify changes in cf-DNA that can distinguish patients with early stage lung cancer from patients with no cancer.

Utilizing polymerase chain reaction (PCR), Sozzi *et al.* [112] examined the levels of plasma DNA in 84 patients with NSCLC and 43 healthy blood donor controls, and showed over a range of cutoff points that healthy controls could be distinguished from patients with lung tumors with an AUC of 0.844. They showed that, even in patients with stage 1a cancer, the amount of circulating plasma DNA was significantly higher than in the control patients [112]. Although this study showed the use of quantifying cf-DNA to distinguish patients with cancer from healthy controls, only 8 of the 43 controls were smokers, so smoking may have been a confounding effect in the cancer diagnosis. Sozzi *et al.* [113] addressed this by matching 93 control individuals who smoked with 100 patients with NSCLC, and were able to discriminate the patients from controls by concentration of cf-DNA with an AUC of 0.94. They also quantified the risk such that a unit increase in plasma DNA led to a 21% increase in NSCLC risk [113]. Other studies have shown that cf-DNA can distinguish patients with benign lung disease from lung cancer with an AUC of 0.73 [114]. However, in a screening cohort of >1,000 higher risk smoking volunteers, Sozzi *et al.* [115] found that quantification of cf-DNA could not be used to distinguish the individuals who would develop lung cancer from those who did not [115]. Other studies have also confirmed this finding [116].

In terms of the emerging clinical needs (Figure 1), these data argue that cf-DNA may not be an effective marker for screening high-risk smokers, but based on these and many other cf-DNA studies [117-120], it could still play a role in diagnosing whether nodules identified by low-dose CT are either benign or malignant. The field is also progressing towards the identification of screening- or diagnostic-specific markers within lung tumor circulating cf-DNA including methylation markers [68] and genetic mutations such as epidermal growth factor receptor mutations [121-123]. Although these developments are relatively new, the Dawson *et al.* study [124], in which genetic alterations could be identified in cf-DNA that corresponded dynamically with metastatic breast cancer, shows the potential for identifying novel genomic and genetic biomarkers within cf-DNA to better stratify patients [124].

Conclusions

As CT screening programs for lung cancer proliferate in the post-National Lung Screening Trial era, there is an urgent and growing need to develop and validate biomarkers that can both help identify those smokers at highest risk who are most likely to benefit from screening and help distinguish benign from malignant lesions found on chest imaging. The recent advances in genetics and genomics have ushered in an era of genome-wide studies aimed at identifying molecular biomarkers for diagnosis and risk for lung cancer. While a number of promising genetic, transcriptomic and epigenomic markers have been identified as detailed above, we have yet to see translation from biomarker discovery to clinical application.

A review of these studies reveals several important limitations that will need to be addressed in the coming years if the field is to advance and have a clinical impact. First, molecular biomarkers discussed in this review will need to be validated in multicenter trials on independent cohorts to demonstrate the validity and generalizability of the biomarker. Importantly, the biomarkers will need to be validated in the clinical setting in which they will be applied. This latter caveat is best addressed at the biomarker development stage, where molecular markers are identified among clinical specimens that reflect the ultimate clinical application (for example, for diagnostic markers, using specimens collected prior to lung cancer diagnosis among patient and controls who present with suspicion of disease). To have clinical utility, these molecular markers will need to demonstrate performance metrics that would alter clinical decision making (for example, having a very high negative predictive value in the diagnostic setting). They will further need to demonstrate that they provide information about cancer risk and/or diagnosis that is independent of clinical and

radiographic risk factors that have been well established for disease. The ultimate translation to the clinic, however, will require transitioning to analytical platforms that can be readily applied in the clinic to facilitate physician adoption as part of their standard of care.

Abbreviations

AUC: Area under the receiver operator characteristic curve; COPD: Chronic obstructive pulmonary disease; CT: Computed tomography; GWAS: Genome-wide association study; MSP: Methylation-specific PCR; RT-qPCR: Reverse transcription quantitative polymerase chain reaction; SCC: Squamous cell carcinoma; NSCLC: Non-small cell lung cancer; SCLC: Small cell lung cancer; SNP: Single nucleotide polymorphism.

Competing interests

AS is a founder and consultant to AllegroDx Inc. All other authors declare that they have no competing interests.

Authors' contributions

All authors contributed to the writing of this manuscript. JB, CM, KH and AS wrote the transcriptomics, epigenetics and miRNA sections. RE and MS wrote the genetics section. JB and KH were responsible for the final design of the figures and tables. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the National Institutes for Health/National Cancer Institute's Early Detection Research Network via U01CA152751 (AS) as well as support from the Department of Defense via CDMRP W81XWH-11-2-0161 (AS).

Author details

¹Bioinformatics Program, Boston University, Boston, MA, USA. ²Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA. ³Department of Medical Oncology, Princess Margaret Cancer Centre, University of Health Network, University of Toronto, Toronto, Ontario, Canada. ⁴Dan L Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA.

Received: 11 March 2013 Accepted: 20 June 2013

Published: 19 July 2013

References

1. Patel JD, Bach PB, Kris MG: **Lung cancer in US women: a contemporary epidemic.** *JAMA* 2004, **291**:1763–1768.
2. Patel JD: **Lung cancer in women.** *J Clin Oncol* 2005, **23**:3212–3218.
3. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: **Global cancer statistics.** *CA Cancer J Clin* 2011, **61**:69–90.
4. Jemal A, Center MM, DeSantis C, Ward EM: **Global patterns of cancer incidence and mortality rates and trends.** *Cancer Epidemiol Biomarkers Prev* 2010, **19**:1893–1907.
5. Collins LG, Haines C, Perkel R, Enck RE: **Lung cancer: diagnosis and management.** *Am Fam Physician* 2007, **75**:56–63.
6. Shopland DR, Burns DM: **Changes in cigarette-related disease risks and their implications for prevention and control.** In *Volume 8 of Smoking and Tobacco Control Monographs*. Bethesda, MD: National Institutes of Health/National Cancer Institute; 1997.
7. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD: **Reduced lung-cancer mortality with low-dose computed tomographic screening.** *N Engl J Med* 2011, **365**:395–409.
8. Hassanein M, Callison JC, Callaway-Lane C, Aldrich MC, Grogan EL, Massion PP: **The state of molecular biomarkers for the early detection of lung cancer.** *Cancer Prev Res (Phila)* 2012, **5**:992–1006.
9. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijaykrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai YY, Chen WW, Shete S, Spitz MR, Houlston RS: **Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1.** *Nat Genet* 2008, **40**:616–622.
10. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Chen C, Goodman G, Field JK, Liloglou T, Xinarianos G, Cassidy A, McLaughlin J, Liu G, Narod S, Krokan HE, Skorpen F, Elvestad MB, Hveem K, Vatten L, Linseisen J, et al: **A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25.** *Nature* 2008, **452**:633–637.
11. Thorgerirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, et al: **Sequence variants at CHRNA3-CHRNA6 and CYP2A6 affect smoking behavior.** *Nat Genet* 2010, **42**:448–453.
12. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, et al: **A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma.** *Am J Hum Genet* 2009, **85**:679–691.
13. Saccone NL, Culverhouse RC, Schwantes-An TH, Cannon DS, Chen X, Cichon S, et al: **Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD.** *PLoS Genet* 2010, **6**:e1001053.
14. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, et al: **Meta-analysis and imputation refines the association of 15q25 with smoking quantity.** *Nat Genet* 2010, **42**:436–440.
15. Tobacco and Genetics Consortium: **Genome-wide meta-analyses identify multiple loci associated with smoking behavior.** *Nat Genet* 2010, **42**:441–447.
16. Spitz MR, Amos CI, Dong Q, Lin J, Wu X: **The CHRNA5-A3 region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer.** *J Natl Cancer Inst* 2008, **100**:1552–1556.
17. Wang Y, Broderick P, Matakidou A, Eisen T, Houlston RS: **Chromosome 15q25 (CHRNA3-CHRNA5) variation impacts indirectly on lung cancer risk.** *PLoS One* 2011, **6**:e19085.
18. Spitz MR, Amos CI, Bierut LJ, Caporaso NE: **Cotinine conundrum—a step forward but questions remain.** *J Natl Cancer Inst* 2012, **104**:720–722.
19. Truong T, Hung RJ, Amos CI, Wu X, Bickeboller H, Rosenberger A, Sauter W, Illig T, Wichmann HE, Risch A, Dienemann H, Kaaks R, Yang P, Jiang R, Wiencke JK, Wrensch M, Hansen H, Kelsey KT, Matsuo K, Tajima K, Schwartz AG, Wenzlaff A, Seow A, Ying C, Staratschek-Jox A, Nürnberg P, Stoelben E, Wolf J, Lazarus P, Muscat JE, et al: **Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium.** *J Natl Cancer Inst* 2010, **102**:959–971.
20. Galvan A, Dragani TA: **Nicotine dependence may link the 15q25 locus to lung cancer risk.** *Carcinogenesis* 2010, **31**:331–333.
21. Zou P, Gu A, Ji G, Zhao L, Zhao P, Lu A: **The TERT rs2736100 polymorphism and cancer risk: a meta-analysis based on 25 case-control studies.** *BMC Cancer* 2012, **12**:7.
22. Spitz MR, Amos CI, Land S, Wu X, Dong Q, Wenzlaff AS, Schwartz AG: **Role of select genetic variants in lung cancer risk in African Americans.** *J Thorac Oncol* 2013, **8**:391–397.
23. Melin BS, Nordfjall K, Andersson U, Roos G: **hTERT cancer risk genotypes are associated with telomere length.** *Genet Epidemiol* 2012, **36**:368–372.
24. Zhu CQ, Cutz JC, Liu N, Lau D, Shepherd FA, Squire JA, Tsao MS: **Amplification of telomerase (hTERT) gene is a poor prognostic marker in non-small-cell lung cancer.** *Br J Cancer* 2006, **94**:1452–1459.
25. James MA, Wen W, Wang Y, Byers LA, Heymach JV, Coombes KR, Girard L, Minna J, You M: **Functional characterization of CLPTM1L as a lung cancer risk candidate gene in the 5p15.33 locus.** *PLoS One* 2012, **7**:e36116.
26. Wang Y, Broderick P, Webb E, Wu X, Vijaykrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WW, Spitz MR, Eisen T, Amos CI, Houlston RS: **Common 5p15.33 and 6p21.33 variants influence lung cancer risk.** *Nat Genet* 2008, **40**:1407–1409.
27. Bae EY, Lee SY, Kang BK, Lee EJ, Choi YY, Kang HG, Choi JE, Jeon HS, Lee WK, Kam S, Shin KM, Jin G, Yoo SS, Lee J, Cha SI, Kim CH, Jung TH, Park JY: **Replication of results of genome-wide association studies on lung cancer susceptibility loci in a Korean population.** *Respirology* 2012, **17**:699–706.
28. Shiraishi K, Kunitoh H, Daigo Y, Takahashi A, Goto K, Sakamoto H, Ohnami S, Shimada Y, Ashikawa K, Saito A, Watanabe S, Tsuta K, Kamatani N, Yoshida T, Nakamura Y, Yokota J, Kubo M, Kohno T: **A genome-wide association study identifies two new susceptibility loci for lung adenocarcinoma in the Japanese population.** *Nat Genet* 2012, **44**:900–903.
29. Hu Z, Wu C, Shi Y, Guo H, Zhao X, Yin Z, Yang L, Dai J, Hu L, Tan W, Li Z, Deng Q, Wang J, Wu W, Jin G, Jiang Y, Yu D, Zhou G, Chen H, Guan P, Chen Y, Shu Y, Xu L, Liu X, Liu L, Xu P, Han B, Bai C, Zhao Y, Zhang H, et al:

- A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese.** *Nat Genet* 2011, **43**:792–796.
30. Dong J, Hu Z, Wu C, Guo H, Zhou B, Lv J, Lu D, Chen K, Shi Y, Chu M, Wang C, Zhang R, Dai J, Jiang Y, Cao S, Qin Z, Yu D, Ma H, Jin G, Gong J, Sun C, Zhao X, Yin Z, Yang L, Li Z, Deng Q, Wang J, Wu W, Zheng H, Zhou G, et al: **Association analyses identify multiple new lung cancer susceptibility loci and their interactions with smoking in the Chinese population.** *Nat Genet* 2012, **44**:895–899.
31. Jin G, Ma H, Wu C, Dai J, Zhang R, Shi Y, Lu J, Miao X, Wang M, Zhou Y, Chen J, Li H, Pan S, Chu M, Lu F, Yu D, Jiang Y, Dong J, Hu L, Chen Y, Xu L, Shu Y, Pan S, Tan W, Zhou B, Lu D, Wu T, Zhang Z, Chen F, Wang X, et al: **Genetic variants at 6p21.1 and 7p15.3 are associated with risk of multiple cancers in Han Chinese.** *Am J Hum Genet* 2012, **91**:928–934.
32. Shi J, Chatterjee N, Rotunno M, Wang Y, Pesatori AC, Consonni D, Li P, Wheeler W, Broderick P, Henrich M, Eisen T, Wang Z, Chen W, Dong Q, Albanes D, Thun M, Spitz MR, Bertazzi PA, Caporaso NE, Chanock SJ, Amos CI, Houlston RS, Landi MT: **Inherited variation at chromosome 12p13.33, including RAD52, influences the risk of squamous cell lung carcinoma.** *Cancer Discov* 2012, **2**:131–139.
33. Tammimaki A, Horton WJ, Stitzel JA: **Recent advances in gene manipulation and nicotinic acetylcholine receptor biology.** *Biochem Pharmacol* 2011, **82**:808–819.
34. De BM, Dani JA: **Reward, addiction, withdrawal to nicotine.** *Annu Rev Neurosci* 2011, **34**:105–130.
35. Liu Y, Liu P, Wen W, James MA, Wang Y, Bailey-Wilson JE, Amos CI, Pinney SM, Yang P, de Andrade M, Petersen GM, Wiest JS, Fain PR, Schwartz AG, Gazdar A, Gaba C, Rothschild H, Mandal D, Kupert E, Lee J, Seminara D, Minna J, Anderson MW, You M: **Haplotype and cell proliferation analyses of candidate lung cancer susceptibility genes on chromosome 15q24-25.1.** *Cancer Res* 2009, **69**:7844–7850.
36. Maisonneuve P, Bagnardi V, Bellomi M, Spaggiari L, Pelosi G, Rampinelli C, Bertolotti R, Rotmensz N, Field JK, Decensi A, Veronesi G: **Lung cancer risk prediction to select smokers for screening CT—a model based on the Italian COSMOS trial.** *Cancer Prev Res (Phila)* 2011, **4**:1778–1789.
37. Tammemagi CM, Pinsky PF, Caporaso NE, Kvale PA, Hocking WG, Church TR, Riley TL, Commins J, Oken MM, Berg CD, Prorok PC: **Lung cancer risk prediction: Prostate, Lung, Colorectal And Ovarian Cancer Screening Trial models and validation.** *J Natl Cancer Inst* 2011, **103**:1058–1068.
38. Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, Field JK: **The LLP risk model: an individual risk prediction model for lung cancer.** *Br J Cancer* 2008, **98**:270–276.
39. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, Shete S, Etzel CJ: **A risk model for prediction of lung cancer.** *J Natl Cancer Inst* 2007, **99**:715–726.
40. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, Hsieh LJ, Begg CB: **Variations in lung cancer risk among smokers.** *J Natl Cancer Inst* 2003, **95**:470–478.
41. Young RP, Hopkins RJ, Whittington CF, Hay BA, Epton MJ, Gamble GD: **Individual and cumulative effects of GWAS susceptibility loci in lung cancer: associations after sub-phenotyping for COPD.** *PLoS One* 2011, **6**:e16476.
42. Wilk JB, Walter RE, Laramie JM, Gottlieb DJ, O'Connor GT: **Framingham Heart Study genome-wide association: results for pulmonary function measures.** *BMC Med Genet* 2007, **8**:S8.
43. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marcianti KD, Franceschini N, van Durme YM, Chen TH, Barr RG, Schabath MB, Couper DJ, Brusselle GG, Psaty BM, van Duijn CM, Rotter JJ, Uitterlinden AG, Hofman A, Punjabi NM, Rivadeneira F, Morrison AC, Enright PL, North KE, Heckbert SR, Lumley T, Stricker BH, O'Connor GT, London SJ: **Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function.** *Nat Genet* 2010, **42**:45–52.
44. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V, Huffman JE, Igl W, Albrecht E, Deloukas P, Henderson J, Graneli R, McArdle WL, Rudnicka AR, Wellcome Trust Case Control C, Barroso I, Loos RJ, Wareham NJ, Mustelin L, Rantanen T, Surakka I, Imboden M, Wichmann HE, Grkovic I, Jankovic S, Zgaga L, et al: **Genome-wide association study identifies five loci associated with lung function.** *Nat Genet* 2010, **42**:36–44.
45. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, Ruppert A, Lørdrup Carlsen KC, Roses A, Anderson W, Rennard SJ, Lomas DA, Silverman EK, Goldstein DB, ICGN Investigators: **A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci.** *PLoS Genet* 2009, **5**:e1000421.
46. Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST, O'Connor GT: **A genome-wide association study of pulmonary function measures in the Framingham Heart Study.** *PLoS Genet* 2009, **5**:e1000429.
47. Broderick P, Wang Y, Vijayakrishnan J, Matakidou A, Spitz MR, Eisen T, Amos CI, Houlston RS: **Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study.** *Cancer Res* 2009, **69**:6633–6641.
48. Thorgerirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, Stacey SN, Bergthorsson JT, Thorlacius S, Gudmundsson J, Jonsson T, Jakobsdottir M, Saemundsdottir J, Olafsdottir O, Gudmundsson LJ, Bjornsdottir G, Kristjansson K, Skuladottir H, Isaksson HJ, Gudbjartsson T, Jones GT, Mueller T, Gottsäter A, Flex A, Aben KK, de Vegt F, et al: **A variant associated with nicotine dependence, lung cancer and peripheral arterial disease.** *Nature* 2008, **452**:638–642.
49. Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, Himes BE, Sylvia JS, Klanderman BJ, Ziniti JP, Lange C, Litonjua AA, Sparrow D, Regan EA, Make BJ, Hokanson JE, Murray T, Hetmanski JB, Pillai SG, Kong X, Anderson WH, Tal-Singer R, Lomas DA, Coxson HO, Edwards LD, MacNee W, Vestbo J, Yates JC, Agusti A, Calverley PM, et al: **A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13.** *Hum Mol Genet* 2012, **21**:947–957.
50. Wang J, Spitz MR, Amos CI, Wilkinson AV, Wu X, Shete S: **Mediating effects of smoking and chronic obstructive pulmonary disease on the relation between the CHRNA5-A3 genetic locus and lung cancer risk.** *Cancer* 2010, **116**:3458–3462.
51. Young RP, Hopkins RJ: **How the genetics of lung cancer may overlap with COPD.** *Respirology* 2011, **16**:1047–1055.
52. El-Zein RA, Young RP, Hopkins RJ, Etzel CJ: **Genetic predisposition to chronic obstructive pulmonary disease and/or lung cancer: important considerations when evaluating risk.** *Cancer Prev Res (Phila)* 2012, **5**:522–527.
53. Bird A: **Perceptions of epigenetics.** *Nature* 2007, **447**:396–398.
54. Bird A: **DNA methylation patterns and epigenetic memory.** *Genes Dev* 2002, **16**:6–21.
55. Robertson KD: **DNA methylation and human disease.** *Nat Rev Genet* 2005, **6**:597–610.
56. Shames DS, Girard L, Gao B, Sato M, Lewis CM, Shivapurkar N, Jiang A, Perou CM, Kim YH, Pollack JR, Fong KM, Lam CL, Wong M, Shyr Y, Nanda R, Olopade OI, Gerald W, Euhus DM, Shay JW, Gazdar AF, Minna JD: **A genome-wide screen for promoter methylation in lung cancer identifies novel methylation markers for multiple malignancies.** *PLoS Med* 2006, **3**:e486.
57. Field JK, Liloglou T, Warrak S, Burger M, Becker E, Berlin K, Nimrich I, Maier S: **Methylation discriminators in NSCLC identified by a microarray based approach.** *Int J Oncol* 2005, **27**:105–111.
58. Wang L, Aakre JA, Jiang R, Marks RS, Wu Y, Chen J, Thibodeau SN, Pankratz VS, Yang P: **Methylation markers for small cell lung cancer in peripheral blood leukocyte DNA.** *J Thorac Oncol* 2010, **5**:778–785.
59. Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, Baylin SB, Herman JG: **Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis.** *Proc Natl Acad Sci USA* 1998, **95**:11891–11896.
60. Palmisano WA, Divine KK, Saccomanno G, Gilliland FD, Baylin SB, Herman JG, Belinsky SA: **Predicting lung cancer by detecting aberrant promoter methylation in sputum.** *Cancer Res* 2000, **60**:5954–5958.
61. Belinsky SA, Klinge DM, Dekker JD, Smith MW, Bocklage TJ, Gilliland FD, Crowell RE, Karp DD, Stidley CA, Picchi MA: **Gene promoter methylation in plasma and sputum increases with lung cancer risk.** *Clin Cancer Res* 2005, **11**:6505–6511.
62. Machida EO, Brock MV, Hooker CM, Nakayama J, Ishida A, Amano J, Picchi MA, Belinsky SA, Herman JG, Taniguchi S, Baylin SB: **Hypermethylation of ASC/TMS1 is a sputum marker for late-stage lung cancer.** *Cancer Res* 2006, **66**:6210–6218.
63. Belinsky SA, Liechty KC, Gentry FD, Wolf HJ, Rogers J, Vu K, Haney J, Kennedy TC, Hirsch FR, Miller Y, Franklin WA, Herman JG, Baylin SB,

- Bunn PA, Byers T: Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res* 2006, **66**:3338–3344.
64. Leng S, Do K, Yingling CM, Picchi MA, Wolf HJ, Kennedy TC, Feser WJ, Baron AE, Franklin WA, Brock MV, Herman JG, Baylin SB, Byers T, Stidley CA, Belinsky SA: **Defining a gene promoter methylation signature in sputum for lung cancer risk assessment.** *Clin Cancer Res* 2012, **18**:3387–3395.
65. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ: **Free DNA in the serum of cancer patients and the effect of therapy.** *Cancer Res* 1977, **37**:646–650.
66. Esteller M, Sanchez-Cespedes M, Rosell R, Sidransky D, Baylin SB, Herman JG: **Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients.** *Cancer Res* 1999, **59**:67–70.
67. Fujiwara K, Fujimoto N, Tabata M, Nishii K, Matsuo K, Hotta K, Kozuki T, Aoe M, Kiura K, Ueoka H, Tanimoto M: **Identification of epigenetic aberrant promoter methylation in serum DNA is useful for early detection of lung cancer.** *Clin Cancer Res* 2005, **11**:1219–1225.
68. Lee SM, Park JY, Kim DS: **Methylation of TMEFF2 gene in tissue and serum DNA from patients with non-small cell lung cancer.** *Mol Cells* 2012, **34**:171–176.
69. Tan SH, Ida H, Lau QC, Goh BC, Chieng WS, Loh M, Ito Y: **Detection of promoter hypermethylation in serum samples of cancer patients by methylation-specific polymerase chain reaction for tumour suppressor genes including RUNX3.** *Oncol Rep* 2007, **18**:1225–1230.
70. Ullivi P, Zoli W, Calistri D, Fabbri F, Tessei A, Rosetti M, Mengozzi M, Amadori D: **p16INK4A and CDH13 hypermethylation in tumor and serum of non-small cell lung cancer patients.** *J Cell Physiol* 2006, **206**:611–615.
71. Begum S, Brait M, Dasgupta S, Ostrow KL, Zahurak M, Carvalho AL, Califano JA, Goodman SN, Sidransky D, Thiagalingam S: **An epigenetic marker panel for detection of lung cancer using cell-free serum DNA.** *Clin Cancer Res* 2011, **17**:4494–4503.
72. Russo AL, Thiagalingam A, Pan H, Califano J, Cheng KH, Ponte JF, Chinnappan D, Nemani P, Sidransky D, Thiagalingam S: **Differential DNA hypermethylation of critical genes mediates the stage-specific tobacco smoke-induced neoplastic progression of lung cancer.** *Clin Cancer Res* 2005, **11**:2466–2470.
73. Yang P, Ma J, Zhang B, Duan H, He Z, Zeng J, Zeng X, Li D, Wang Q, Xiao Y, Liu C, Xiao Q, Chen L, Zhu X, Xing X, Li Z, Zhang S, Zhang Z, Ma L, Wang E, Zhuang Z, Zheng Y, Chen W: **CpG site-specific hypermethylation of p16INK4alpha in peripheral blood lymphocytes of PAH-exposed workers.** *Cancer Epidemiol Biomarkers Prev* 2012, **21**:182–190.
74. Woenckhaus M, Klein-Hitpass L, Grepmeier U, Merk J, Pfeifer M, Wild P, Bettstetter M, Wuensch P, Blaszyk H, Hartmann A, Hofstaedter F, Dietmaier W: **Smoking and cancer-related gene expression in bronchial epithelium and non-small-cell lung cancers.** *J Pathol* 2006, **210**:192–204.
75. Boelens MC, van den Berg A, Fehrmann RS, Geerlings M, de Jong WK, te Meerman GJ, Sietsma H, Timens W, Postma DS, Groen HJ: **Current smoking-specific gene expression signature in normal bronchial epithelium is enhanced in squamous cell lung cancer.** *J Pathol* 2009, **218**:182–191.
76. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, Gilman S, Dumas YM, Calner P, Sebastiani P, Sridhar S, Beamis J, Lamb C, Anderson T, Gerry N, Keane J, Lenburg ME, Brody JS: **Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer.** *Nat Med* 2007, **13**:361–366.
77. Beane J, Sebastiani P, Whitfield TH, Steiling K, Dumas YM, Lenburg ME, Spira A: **A prediction model for lung cancer diagnosis that integrates genomic and clinical features.** *Cancer Prev Res (Phila)* 2008, **1**:56–64.
78. Blomquist T, Crawford EL, Mullins D, Yoon Y, Hernandez DA, Khuder S, Ruppel PL, Peters E, Oldfield DJ, Austermiller B, Anders JC, Willey JC: **Pattern of antioxidant and DNA repair gene expression in normal airway epithelium associated with lung cancer diagnosis.** *Cancer Res* 2009, **69**:8629–8635.
79. Gustafson AM, Soldi R, Anderlind C, Scholand MB, Qian J, Zhang X, Cooper K, Walker D, McWilliams A, Liu G, Szabo E, Brody J, Massion PP, Lenburg ME, Lam S, Bild AH, Spira A: **Airway PI3K pathway activation is an early and reversible event in lung cancer development.** *Sci Transl Med* 2010, **2**:26ra25.
80. Wang X, Chorley BN, Pittman GS, Kleeberger SR, Brothers J 2nd, Liu G, Spira A, Bell DA: **Genetic variation and antioxidant response gene expression in the bronchial airway epithelium of smokers at risk for lung cancer.** *PLoS One* 2010, **5**:e11934.
81. Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, Luo L, Zhang XH, Xiao J, Alekseyev YO, Wang S, Levy S, Massion PP, Lenburg M, Spira A: **Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq.** *Cancer Prev Res (Phila)* 2011, **4**:803–817.
82. Boyle JO, Gümüs ZH, Kacker A, Choksi VL, Bocker JM, Zhou XK, Yantiss RK, Hughes DB, Du B, Judson BL, Subbaramaiah K, Dannenberg AJ: **Effects of cigarette smoke on the human oral mucosal transcriptome.** *Cancer Prev Res (Phila)* 2010, **3**:266–278.
83. Sridhar S, Schembri F, Zeskind J, Shah V, Gustafson AM, Steiling K, Liu G, Dumas YM, Zhang X, Brody JS, Lenburg ME, Spira A: **Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium.** *BMC Genomics* 2008, **9**:259.
84. Zhang X, Sebastiani P, Liu G, Schembri F, Zhang X, Dumas YM, Langer EM, Alekseyev Y, O'Connor GT, Brooks DR, Lenburg ME, Spira A: **Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium.** *Physiol Genomics* 2010, **41**:1–8.
85. Zhang L, Xiao H, Zhou H, Santiago S, Lee JM, Garon EB, Yang J, Brinkmann O, Yan X, Akin D, Chia D, Elashoff D, Park NH, Wong DT: **Development of transcriptomic biomarker signature in human saliva to detect lung cancer.** *Cell Mol Life Sci* 2012, **69**:3341–3350.
86. Showe MK, Vachani A, Kossenkov AV, Yousef M, Nichols C, Nikonova EV, Chang C, Kucharczuk J, Tran B, Wakeam E, Yie TA, Speicher D, Rom WN, Albelda S, Showe LC: **Gene expression profiles in peripheral blood mononuclear cells can distinguish patients with non-small cell lung cancer from patients with nonmalignant lung disease.** *Cancer Res* 2009, **69**:9202–9210.
87. Rotunno M, Hu N, Su H, Wang C, Goldstein AM, Bergen AW, Consonni D, Pesatori AC, Bertazzi PA, Wacholder S, Shih J, Caporaso NE, Taylor PR, Landi MT: **A gene expression signature from peripheral whole blood for stage I lung adenocarcinoma.** *Cancer Prev Res (Phila)* 2011, **4**:1599–1608.
88. Zander T, Hofmann A, Staratschek-Jox A, Classen S, Debey-Pascher S, Maisel D, Ansén S, Hahn M, Beyer M, Thomas RK, Gathof B, Mauch C, Delank KS, Engel-Riedel W, Wichmann HE, Stoelben E, Schultze JL, Wolf J: **Blood-based gene expression signatures in non-small cell lung cancer.** *Clin Cancer Res* 2011, **17**:3360–3367.
89. Kato M, Slack FJ: **microRNAs: small molecules with big roles - C. elegans to human cancer.** *Biol Cell* 2008, **100**:71–81.
90. Eszlinger M, Krohn K, Hauptmann S, Dralle H, Giordano TJ, Paschke R: **Perspectives for improved and more accurate classification of thyroid epithelial tumors.** *J Clin Endocrinol Metab* 2008, **93**:3286–3294.
91. Mascaux C, Laes JF, Anthoine G, Haller A, Ninane V, Burny A, Sculier JP: **Evolution of microRNA expression during human bronchial squamous carcinogenesis.** *Eur Respir J* 2009, **33**:352–359.
92. Xie Y, Todd NW, Liu Z, Zhan M, Fang H, Peng H, Alattar M, Deepak J, Stass SA, Jiang F: **Altered miRNA expression in sputum for diagnosis of non-small cell lung cancer.** *Lung Cancer* 2010, **67**:170–176.
93. Xing L, Todd NW, Yu L, Fang H, Jiang F: **Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers.** *Mod Pathol* 2010, **23**:1157–1164.
94. Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z, Fang H, Zhang J, Katz RL, Jiang F: **Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers.** *Int J Cancer* 2010, **127**:2870–2878.
95. Patnaik SK, Yendamuri S, Kannisto E, Kucharczuk JC, Singhal S, Vachani A: **MicroRNA expression profiles of whole blood in lung adenocarcinoma.** *PLoS One* 2012, **7**:e46045.
96. Keller A, Backes C, Leidinger P, Kefer N, Boisguerin V, Barbacioru C, Vogel B, Matzas M, Huwer H, Katus HA, Stähler C, Meder B, Meese E: **Next-generation sequencing identifies novel microRNAs in peripheral blood of lung cancer patients.** *Mol Biosyst* 2011, **7**:3187–3199.
97. Jeong HC, Kim EK, Lee JH, Lee JM, Yoo HN, Kim JK: **Aberrant expression of let-7a miRNA in the blood of non-small cell lung cancer patients.** *Mol Med Report* 2011, **4**:383–387.
98. Leidinger P, Keller A, Borries A, Huwer H, Rohling M, Huebers J, Lenhof HP, Meese E: **Specific peripheral miRNA profiles for distinguishing lung cancer from COPD.** *Lung Cancer* 2011, **74**:41–47.
99. Keller A, Leidinger P, Borries A, Wendschlag A, Wucherpfennig F, Scheffler M, Huwer H, Lenhof HP, Meese E: **miRNAs in lung cancer - studying complex fingerprints in patient's blood cells by microarray experiments.** *BMC Cancer* 2009, **9**:353.

100. Hennessey PT, Sanford T, Choudhary A, Mydlarz WW, Brown D, Adai AT, Ochs MF, Ahrendt SA, Mambo E, Califano JA: **Serum microRNA biomarkers for detection of non-small cell lung cancer.** *PLoS One* 2012, **7**:e32307.
101. Bianchi F, Nicassio F, Marzi M, Belloni E, Dall'olio V, Bernard L, Pelosi G, Maisonneuve P, Veronesi G, Di Fiore PP: **A serum circulating miRNA diagnostic test to identify asymptomatic high-risk individuals with early stage lung cancer.** *EMBO Mol Med* 2011, **3**:495–503.
102. Keller A, Leidinger P, Gislefoss R, Haugen A, Langseth H, Staehler P, Lenhof HP, Meese E: **Stable serum miRNA profiles as potential tool for non-invasive lung cancer diagnosis.** *RNA Biol* 2011, **8**:506–516.
103. Chen X, Hu Z, Wang W, Ba Y, Ma L, Zhang C, Wang C, Ren Z, Zhao Y, Wu S, Zhuang R, Zhang Y, Hu H, Liu C, Xu L, Wang J, Shen H, Zhang J, Zen K, Zhang CY: **Identification of ten serum microRNAs from a genome-wide serum microRNA expression profile as novel noninvasive biomarkers for nonsmall cell lung cancer diagnosis.** *Int J Cancer* 2012, **130**:1620–1628.
104. Roth C, Kasimir-Bauer S, Pantel K, Schwarzenbach H: **Screening for circulating nucleic acids and caspase activity in the peripheral blood as potential diagnostic tools in lung cancer.** *Mol Oncol* 2011, **5**:281–291.
105. Foss KM, Sima C, Ugolini D, Neri M, Allen KE, Weiss GJ: **miR-1254 and miR-574-5p: serum-based microRNA biomarkers for early-stage non-small cell lung cancer.** *J Thorac Oncol* 2011, **6**:482–488.
106. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, Guo J, Zhang Y, Chen J, Guo X, Li Q, Li X, Wang W, Zhang Y, Wang J, Jiang X, Xiang Y, Xu C, Zheng P, Zhang J, Li R, Zhang H, Shang X, Gong T, Ning G, Wang J, Zen K, Zhang J, Zhang CY: **Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases.** *Cell Res* 2008, **18**:997–1006.
107. Boeri M, Verri C, Conte D, Roz L, Modena P, Facchinetti F, Calabrò E, Croce CM, Pastorino U, Sozzi G: **MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer.** *Proc Natl Acad Sci USA* 2011, **108**:3713–3718.
108. Shen J, Todd NW, Zhang H, Yu L, Lingxiao X, Mei Y, Guarnera M, Liao J, Chou A, Lu CL, Jiang Z, Fang H, Katz RL, Jiang F: **Plasma microRNAs as potential biomarkers for non-small-cell lung cancer.** *Lab Invest* 2011, **91**:579–587.
109. Rabinowits G, Gercel-Taylor C, Day JM, Taylor DD, Kloecker GH: **Exosomal microRNA: a diagnostic marker for lung cancer.** *Clin Lung Cancer* 2009, **10**:42–46.
110. Stroun M, Anker P, Lyautey J, Lederrey C, Maurice PA: **Isolation and characterization of DNA from the plasma of cancer patients.** *Eur J Cancer Clin Oncol* 1987, **23**:707–712.
111. Fournié GJ, Courtin JP, Laval F, Chalé JJ, Pourrat JP, Pujazon MC, Lauque D, Carles P: **Plasma DNA as a marker of cancerous cell death. Investigations in patients suffering from lung cancer and in nude mice bearing human tumours.** *Cancer Lett* 1995, **91**:221–227.
112. Sozzi G, Conte D, Mariani L, Lo Vullo S, Roz L, Lombardo C, Pierotti MA, Tavecchio L: **Analysis of circulating tumor DNA in plasma at diagnosis and during follow-up of lung cancer patients.** *Cancer Res* 2001, **61**:4675–4678.
113. Sozzi G, Conte D, Leon M, Ciricione R, Roz L, Ratcliffe C, Roz E, Cirenei N, Bellomi M, Pelosi G, Pierotti MA, Pastorino U: **Quantification of free circulating DNA as a diagnostic marker in lung cancer.** *J Clin Oncol* 2003, **21**:3902–3908.
114. Xie GS, Hou AR, Li LY, Gao YN, Cheng SJ: **Quantification of plasma DNA as a screening tool for lung cancer.** *Chin Med J (Engl)* 2004, **117**:1485–1488.
115. Sozzi G, Roz L, Conte D, Mariani L, Andriani F, Lo Vullo S, Verri C, Pastorino U: **Plasma DNA quantification in lung cancer computed tomography screening: five-year results of a prospective study.** *Am J Respir Crit Care Med* 2009, **179**:69–74.
116. Zhang R, Shao F, Wu X, Ying K: **Value of quantitative analysis of circulating cell free DNA as a screening tool for lung cancer: a meta-analysis.** *Lung Cancer* 2010, **69**:225–231.
117. Szepehinski A, Dancewicz M, Kopinski P, Kowalewski J, Chorostowska-Wynimko J: **Real-time PCR quantification of plasma DNA in non-small cell lung cancer patients and healthy controls.** *Eur J Med Res* 2009, **14**:237–240.
118. Kumar S, Guleria R, Singh V, Bharti AC, Mohan A, Das BC: **Efficacy of circulating plasma DNA as a diagnostic tool for advanced non-small cell lung cancer and its predictive utility for survival and response to chemotherapy.** *Lung Cancer* 2010, **70**:211–217.
119. Yoon KA, Park S, Lee SH, Kim JH, Lee JS: **Comparison of circulating plasma DNA levels between lung cancer patients and healthy controls.** *J Mol Diagn* 2009, **11**:182–185.
120. Paci M, Maramotti S, Bellesia E, Formisano D, Albertazzi L, Ricchetti T, Ferrari G, Annessi V, Lasagni D, Carbonelli C, De Franco S, Brini M, Sgarbi G, Lodi R: **Circulating plasma DNA as diagnostic biomarker in non-small cell lung cancer.** *Lung Cancer* 2009, **64**:92–97.
121. Akca H, Demiray A, Yaren A, Bir F, Koseler A, Iwakawa R, Bagci G, Yokota J: **Utility of serum DNA and pyrosequencing for the detection of EGFR mutations in non-small cell lung cancer.** *Cancer Genet* 2013, **206**:73–80.
122. Nakamura T, Sueoka-Aragane N, Iwanaga K, Sato A, Komiya K, Kobayashi N, Hayashi S, Hosomi T, Hirai M, Sueoka E, Kimura S: **Application of a highly sensitive detection system for epidermal growth factor receptor mutations in plasma DNA.** *J Thorac Oncol* 2012, **7**:1369–1381.
123. Isobe K, Hata Y, Kobayashi K, Hirota N, Sato K, Sano G, Sugino K, Sakamoto S, Takai Y, Shibuya K, Takagi K, Homma S: **Clinical significance of circulating tumor cells and free DNA in non-small cell lung cancer.** *Anticancer Res* 2012, **32**:3339–3344.
124. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, Dunning MJ, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C, Rosenfeld N: **Analysis of circulating tumor DNA to monitor metastatic breast cancer.** *N Eng J Med* 2013, **368**:1199–1209.

doi:10.1186/1741-7015-11-168

Cite this article as: Brothers et al.: 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 Medicine* 2013 **11**:168.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

