

Translating the transcriptome into tools for the early detection and prevention of lung cancer

Yaron B Gesthalter,^{1,2,3} Jessica Vick,³ Katrina Steiling,^{1,3,4} Avrum Spira^{1,3,4}

¹The Pulmonary Center, Boston University School of Medicine, Boston, Massachusetts, USA

²Division of Pulmonary, Allergy, and Critical Care Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

³Division of Computational Biomedicine, Boston University School of Medicine, Boston, Massachusetts, USA

⁴Bioinformatics Program, Boston University, Boston, Massachusetts, USA

Correspondence to

Dr Yaron B Gesthalter, The Pulmonary Center, Boston University School of Medicine, 72 East Concord Street, Boston, MA 02118, USA; ygesty@gmail.edu

Received 24 November 2014

Revised 8 January 2015

Accepted 9 January 2015

Published Online First

27 January 2015

ABSTRACT

Despite advances in the management of lung cancer, this disease remains a significant global health burden with survival rates that have not significantly improved in decades. The mortality reduction achieved by low-dose helical CT (LDCT) screening of select high-risk patients is challenged by the high false positive rate of this screening modality and the potential for morbidity associated with follow-up diagnostic evaluation in patients with high risk for iatrogenic complications. The diagnostic dilemma of the indeterminate nodule incidentally identified on diagnostic or screening CT has created a need for reliable biomarkers capable of distinguishing benign from malignant disease. Furthermore, there is an urgent need to develop molecular biomarkers to supplement clinical risk models in order to identify patients at highest risk for having an early stage lung cancer that may derive the greatest benefit from LDCT screening, as well as identifying patients at high-risk for developing lung cancer that may be candidates for emerging chemopreventive strategies. Evolving bioinformatic techniques and the application of these algorithms to analyse the transcriptomic changes associated with lung cancer promise translational discoveries that can bridge these large clinical gaps. The identification of lung cancer associated transcriptomic alterations in readily accessible tissue sampling sites offers the potential to develop early diagnostic and risk stratification strategies applicable to large populations. This review summarises the challenges associated with the early detection, screening and chemoprevention of lung cancer with an emphasis on how genomic information encapsulated by the transcriptome can facilitate future innovations in these clinical settings.

THE CLINICAL NEED FOR EARLY DETECTION AND RISK STRATIFICATION FOR LUNG CANCER

Lung cancer mortality has proven a challenge to curtail, with an overall 5-year survival of only 16.8%.¹ The overall individual and public health burden of lung cancer has led to the embrace of innovative efforts to meet the challenges this disease poses. While tremendous strides have been made to identify effective and targeted pharmacologic therapies for later stages of this disease, strategies for identifying the disease at earlier stages may further impact disease outcomes. The National Lung Screening Trial (NLST) demonstrated a reduction in overall and lung cancer-specific mortality using low-dose helical CT (LDCT) screening of high-risk patients.² This has given credence to the concept of early diagnosis in asymptomatic patients as a means to alter the natural progression of this disease. However, the widespread implementation of this strategy remains challenged by the risk of subjecting screened individuals to false

positive tests and subsequent follow-up studies.³ Additional areas of concern regarding imaging-based screening approaches lie in the potential for overdiagnosis of indolent disease, and the large proportion of individuals with other risk factors for lung cancer (eg, family history, occupational exposures, obstructive lung disease) who do not meet the current screening criteria. These observations emphasise the need for effective diagnostic tools capable of both distinguishing malignant from benign disease and indolent from aggressive lesions. Furthermore, there is an unmet clinical need for improving the lung cancer risk quantification to better tailor screening and chemoprevention strategies.

High-throughput molecular profiling technologies coupled with emerging computational algorithms have the potential to address this clinical challenge through the development of biomarkers for lung cancer detection, risk stratification and chemoprevention. One challenge to this approach is untangling the heterogeneity of lung carcinogenesis, as there are a multitude of genetic, epigenetic, transcriptomic and proteomic alterations that occur in response to smoking and in lung tumours.⁴ The integration of transcriptomic biomarkers that reflect host-specific and gene-by-environment molecular perturbations related to lung cancer and lung cancer risk with clinical factors may help to overcome these obstacles by accurately predicting early stage disease, identifying those at risk for developing lung cancer and more precisely targeting chemopreventive therapies based on an each individual's molecular susceptibility pattern.

The ultimate clinical applicability of lung cancer early detection and risk stratification biomarkers are dependent on the ability to safely and readily collect disease-relevant biospecimens. As lung tissue is relatively inaccessible except through invasive procedures, less-invasive biospecimens such as urine and exhaled breath condensate are attractive alternatives.⁴ However, the relationship of metabolic and molecular changes in these sites to alterations in the lung tumour itself is unclear, and other systemic diseases might influence the molecular alterations in these compartments. Sputum and saliva are particularly of interest given the direct tobacco exposure, the direct relation to the respiratory tract of these biofluids and their readiness of sampling. Peripheral blood is also a potential biospecimen for biomarker development, as it circulates throughout the entire body including the lung and may reflect lung cancer-associated patterns, although the sensitivity of relatively dilute changes in this compartment combined with the lack of specificity for the lung remain significant challenges. By contrast, the airway epithelium, which can be



To cite: Gesthalter YB, Vick J, Steiling K, et al. *Thorax* 2015;**70**:476–481.



sampled by bronchoscopic or nasal brushing, represents a relatively non-invasive biosample that can potentially provide molecular information about an individual's physiological response to and damage from tobacco smoke and other inhaled toxins that are causally linked to lung carcinogenesis.⁵

While there are a large number of genetic, genomic and proteomic biomarkers, from these types of biosamples, being developed for the early detection of lung cancer,^{4 6 7} few have been validated in large prospective studies in the clinical setting in which the test will be applied. Proteomic profiling is an attractive avenue for deriving biomarkers, as proteins are the molecules that effect phenotype; however, technologies for accurately measuring them in a high-throughput manner are limited compared with those for nucleic acids. Genetic profiling of lung tumour tissue has led to the identification of somatic mutations with prognostic and therapeutic implications, although these markers have yet to be shown to improve early detection of lung cancer, identification of at-risk patients or stratification of patients for chemopreventive therapy. The measurement of RNA expression provides a number of advantages over other methods for identifying clinically useful biomarkers including (1) the availability of robust, cost-effective and reproducible profiling platforms which can be used in a clinical setting; (2) that gene expression represents an intermediate biological endpoint between underlying genetic or epigenetic alterations driving carcinogenesis and clinical phenotype and (3) the vast amount of publicly available gene expression microarray data sets that can be leveraged. This review will focus on transcriptomic profiling as clinically useful biomarkers for lung cancer early detection, risk stratification and chemoprevention in patients with smoking as a primary risk factor for lung cancer development (figure 1). While there is clearly a growing unmet need to apply these same paradigms to never smokers who develop lung cancer, there has been relatively little transcriptomic biomarker work in this space and thus falls beyond the scope of this review.

EARLY DETECTION BIOMARKERS FOR LUNG CANCER

The impact of tobacco smoke on the airway transcriptome

Several studies have described the transcriptomic effects of tobacco smoke, the largest risk factor for lung cancer, on the cytologically normal airway epithelium of the intra- and extra-thoracic airways.^{8–11} Active cigarette smokers demonstrate distinct gene expression patterns in the bronchial airway epithelium compared with lifelong non-smokers.^{8 11} These alterations are enriched in genes that function in oxidative stress, apoptosis, profibrotic, mucin, anti-oxidants, anti-proteases and immune-related processes.¹¹ Smoking-induced alterations in airway epithelial gene expression are similarly changed in both large and small bronchial airways, and may provide insight into the molecular events occurring in the distal airways through proximal airway genomic profiling.¹² Furthermore, microRNA, which are short RNA molecules that regulate gene expression, display distinct expression patterns in the bronchial epithelium of active cigarette smokers relative to never smokers, suggesting that microRNAs may regulate part of the gene expression response to cigarette smoking.¹³

The majority of smoking-induced alterations in airway gene expression rapidly revert towards normal among former smokers.⁹ These rapidly reversible genes are enriched in detoxification oxidoreductase functions. Despite decades of smoking cessation, a small subset of tobacco-induced gene expression alterations in the airway epithelium do not revert towards levels observed in lifelong non-smokers.⁹ This observation suggests

that gene expression profiling of the cytologically normal airway epithelium might reflect molecular alterations that persist in former smokers and that may precede the development of lung cancer. Furthermore, these data support the hypothesis that the airway epithelium might be leveraged to develop sensitive and specific biomarkers for identifying lung cancer in both current and former smokers.

While gene expression profiling of the bronchial airway epithelium provides information reflecting the individual response to tobacco smoke in the airway, biospecimens from extra-thoracic sites would further broaden the capability to study lung cancer-associated processes in larger populations. Based on the unified airway field of injury concept, which suggests that cigarette smoking creates injury throughout the respiratory tract, prior studies have shown that tobacco-related gene expression alterations in the extra-thoracic epithelium of the nose and mouth similarly reflect the host response to cigarette smoke in the distal airways.^{14–16} Other studies have shown that a targeted subset of smoking-associated gene expression changes in the buccal mucosa are similarly altered in distal lung tissue.¹⁰ Furthermore, both the nasal and bronchial epithelium demonstrate similar patterns of gene expression changes in response to cigarette smoking,¹⁶ suggesting that the nose may be a surrogate biospecimen for monitoring the individual response to smoking and potentially the risk for lung cancer.

The airway transcriptome as a biomarker for the early detection of lung cancer

The substantially lower mortality for localised lung cancer (54%) as compared with distantly spread disease (4%) illustrates the urgent need for sensitive and specific biomarkers for detecting lung cancer early.¹ While there are clinical guidelines regarding the management of pulmonary nodules,¹⁷ many of these cases represent indeterminate lesions in which the clinician is uncertain as to whether an invasive work-up is warranted. The diagnostic evaluation of lung cancer is limited by the need for direct tissue sampling with varying diagnostic yields depending on procedure, lesion location and size,¹⁸ and the potential for invasive diagnostic procedures to cause excess morbidity in patients with low pulmonary reserve.¹⁹ However, the identification of sensitive and specific biomarkers from readily accessible biospecimens has the potential to facilitate the detection of lung cancer at early stages where therapeutic intervention is more likely to be curative.

Based on the airway field of injury described above, cytologically normal airway epithelial brushings obtained from the mainstem bronchus of participants undergoing bronchoscopy for suspect lung cancer have been used to develop a gene expression signature that distinguishes patients with lung cancer from those without lung cancer.²⁰ This airway gene expression biomarker for lung cancer diagnosis performs independently of other clinical risk factors,²¹ and demonstrates improved performance when combined with clinical risk factors. Importantly, a recently completed large prospective multicentre trial confirmed that bronchial airway gene expression is a highly sensitive biomarker for lung cancer with a high negative predictive value when applied in the clinical setting in which the classifier would be used (Whitney D *et al*, as abstract American Thoracic Society 2014).²² These findings support that airway epithelial gene expression can address the clinical challenge of ruling out lung cancer in patients with an intermediate clinical suspicion of disease, and has the potential to reduce unnecessary invasive follow-up procedures (following a non-diagnostic bronchoscopy) in patients with non-malignant lung disease.

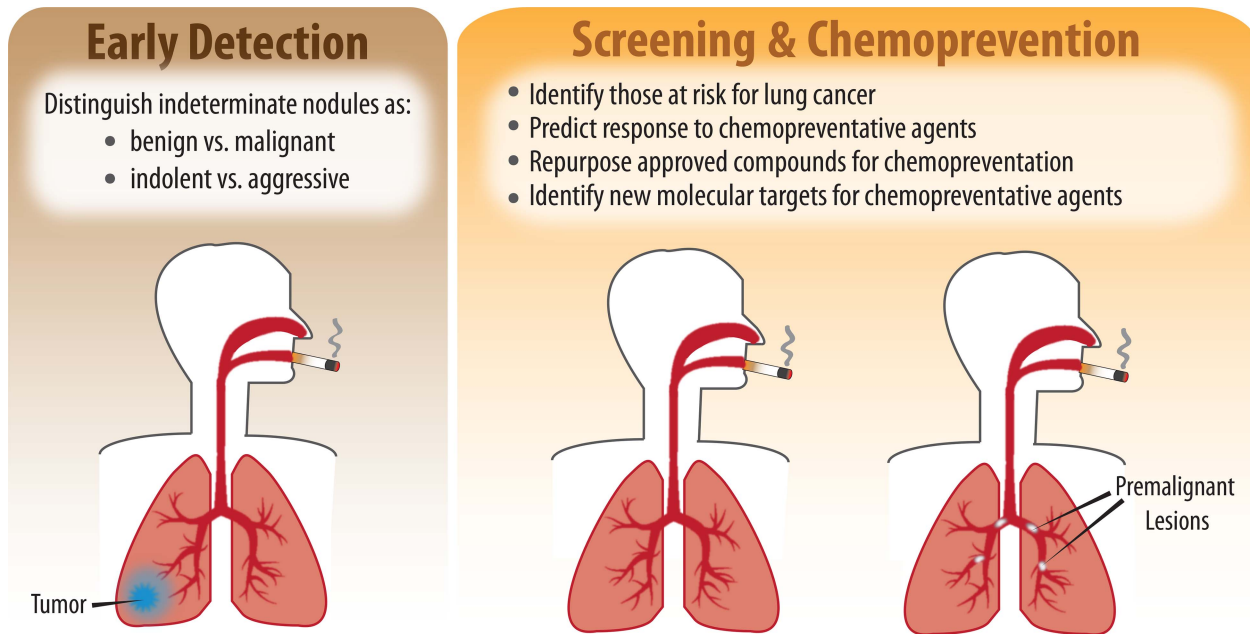


Figure 1 Schematic overview of strategic intervention points for lung cancer transcriptomic biomarkers. Gene expression biomarkers derived from readily obtainable biospecimens may facilitate the interception of lung cancer at earlier stages and even before the disease develops. Diagnostic biomarkers, which can detect the presence of early stage lung cancer (left panel), can serve as an adjunct to clinical factors to stratify patients towards or away from invasive diagnostics. Prior to lung cancer genesis (right panel), transcriptomic alterations that precede the development of lung cancer might improve the accurate identification of high-risk patients that might derive greatest benefit from low-dose helical CT scan screening and chemoprevention. Further profiling may offer insight into novel potential chemopreventive agents and personalise chemopreventive strategies, which would alter the course before disease manifestation.

The concept that bronchial airway gene expression can serve as a biomarker for lung cancer has been strengthened by the recent characterisation of the ‘local’ field of injury that occurs in the small airways adjacent to lung tumours.²³ Importantly, this study demonstrated two categories of transcriptomic alterations throughout the small airways: (1) genes that change in a gradient-like fashion as anatomic distance from the tumour increases, and which may reflect a local response to the tumour microenvironment; and (2) genes that are consistently altered throughout the small airway regardless of the anatomic location relative to the tumour. This latter set of genes, which were consistently altered throughout the lung cancer-associated field of injury, was enriched among the genes previously described as being altered in the mainstem bronchial airway epithelium of smokers with lung cancer.^{20, 23} These findings suggest that gene expression alterations observed in the large airway of smokers with lung cancer reflect, in part, the altered transcriptome observed in lung tumours and small airways adjacent to these tumours, supporting the concept that large airway gene expression can serve as tool to detect the presence of lung cancer within the lung parenchyma.⁵

Given that microRNA modulate the airway gene expression response to smoking,¹³ there is a potential role for microRNA in regulating the airway gene expression changes associated with lung cancer. By sequencing small RNA from cytologically normal bronchial airway epithelial cells, our group identified miR-4423 as a novel primate-specific miRNA almost exclusively expressed in the bronchial airway epithelium.²⁴ miR-4423 exhibited lower expression levels in both lung tumour tissue and in the cytologically normal epithelial cells obtained from the mainstem bronchi of smokers with lung cancer. Further, the predicted targets of miR-4423 were enriched among genes with altered expression levels in the bronchial airway of smokers with lung cancer, suggesting that miR-4423 regulates lung

cancer-associated gene expression throughout the airway field of injury. Importantly, this microRNA was shown to regulate airway epithelial cell differentiation, and overexpression of this microRNA suppresses growth of lung cancer cell line growth both in vitro and in vivo.²⁴ Beyond serving as diagnostic biomarkers, these findings suggest microRNAs may also potentially serve as novel therapeutic targets for lung cancer.

Extending transcriptomic biomarker development beyond the intrathoracic airways

Although bronchial airway epithelium show promise for developing more accurate tools to achieve early lung cancer diagnosis, this approach may not be suitable for patients in whom bronchoscopy is not indicated as part of their clinical evaluation. For example, in patients with small pulmonary nodules (<8 mm) such as those expected in the LDCT-screening era, there is an urgent need for developing minimally invasive biomarkers to distinguish malignant lung nodules from indolent or benign nodules. Potential sites for the collection of minimally invasive tissue biosamples that might harbour cancer related gene expression changes range from the extra-thoracic respiratory tract (nasal and oral epithelium) as an extension of the airway field of injury, and to peripheral biofluids such as blood, sputum and saliva.

The buccal mucosa is an anatomic compartment of particular interest given its readily accessible location, its direct exposure to mainstream cigarette smoke, and its relationship to other smoking-induced gene expression changes in intrathoracic and extra-thoracic airway specimens.^{14, 15} This is consistent with previous descriptions of the relationship between nasal and buccal mucosa to distal airway molecular changes related to smoking. Of note, gene expression changes in the nasal epithelium have been suggested as more pronounced relative to buccal mucosa, likely reflecting issues with RNA quality and extraction.¹⁴

Together, this suggests that the upper airway epithelium may serve as an easily accessible sample-type, which can be used to profile the transcriptomic alterations related to smoking and lung cancer in large numbers of patients that are otherwise unable to undergo bronchoscopy.

Saliva is another potential biospecimen for developing minimally invasive and easily obtainable early detection biomarkers. In a proof of concept trial, Zhang *et al*²⁵ used whole genome arrays to describe a five gene expression signature using saliva supernatant that distinguished subjects with lung cancer from those without lung cancer, although this signature remains to be validated in an independent prospective cohort. While saliva remains an appealing biospecimen given the ease of collection, this approach may ultimately be limited by the presence of RNases, which degrade mRNA and challenge gene expression profiling of the saliva due to low RNA quantity and quality. Additional work has demonstrated the potential use of microRNA detected in sputum as a predictor of early adenocarcinoma and squamous cell carcinoma, however, these studies are equally limited by the relative expense of the quantitative analyses involved and have yet to undergo large-scale validation.^{26 27}

Gene expression profiling of peripheral blood is another approach for early detection biomarker development given that blood circulates throughout all anatomic compartments of the body including the lung. Given that chemokines and cytokines may be released by malignant cells and thereby induce tumour-specific signatures in normal immune cells, one study derived a 29 gene expression signature from peripheral mononuclear cells (PBMC), which distinguished patients with benign lung nodules from those with malignant nodules in an independent validation cohort.²⁸ The higher diagnostic accuracy of this biomarker in patients with more advanced lung cancer, in addition to the degradation of this biomarker signal after tumour resection, underscores the potential relationship between disease burden and peripheral mononuclear cell gene expression. The application of PBMC gene expression profiling as a method to develop early detection biomarkers may be limited by the technical requirements for cell sorting from whole blood. Gene expression profiling of peripheral whole blood (PWB), which requires less laboratory preprocessing, may also have the utility for developing lung cancer biomarkers. For example, the comparison of PWB gene expression profiles obtained from subjects with confirmed adenocarcinoma to healthy controls without lung cancer in the Environment and Genetics in Lung cancer Etiology (EAGLE) study produced an eight gene expression signature that distinguished subjects with and without lung cancer.²⁹ Importantly, this signature was validated in an cohort of subjects with Stage I lung cancer, suggesting the feasibility of using gene expression profiling of PWB to develop early detection biomarkers.

A related approach to biomarker development in peripheral blood leverages the regulatory role of microRNAs in modulating gene expression and their relative resistance to degradation. MicroRNA expression profiling of plasma samples collected from subjects undergoing LDCT for lung cancer screening as part of the Multicenter Italian Lung Detection (MILD) trial derived a microRNA expression signature that demonstrated a higher sensitivity compared with LDCT screening alone, and identified a higher proportion of true cancers.³⁰ These findings suggest that microRNA profiling of peripheral blood or other lung cancer-relevant biosamples may ultimately reduce the false positive rate associated with LDCT. Similarly, in the Continuing Observation of Smoking Subjects (COSMOS) trial, a serum circulating 34 miRNA expression signature distinguishing early

stage Non Small Cell Cancer (NSCLC) in asymptomatic subjects from healthy controls.³¹ While biomarkers derived from circulating microRNA in the peripheral blood have the benefit of being readily available, there are several potential challenges translating circulating microRNA to widespread clinical use including (1) the technical restraints in isolating sufficient amounts and quality of microRNA; (2) the limited reproducibility across different studies measuring different components of blood and (3) the question of whether these circulating microRNA biomarker profiles reflect tumour-related changes or constituents of the peripheral circulation such as inflammatory cells.³²

USING THE TRANSCRIPTOME TO DEVELOP BIOMARKERS FOR LUNG CANCER SCREENING AND CHEMOPREVENTION Identification of patients at high risk for lung cancer who may benefit from LDCT screening

The early identification of malignancy in high-risk asymptomatic patients is the cornerstone principle of screening. An ideal screening test is characterised by high sensitivity and specificity, while conferring minimal risk of morbidity and mortality to tested individuals. However, even stringent selection of high-risk patients based on established clinical risk models for LDCT screening is limited by a ~25% positive screening tests of which ~95% were ultimately determined to be non-cancerous.² Additionally, approximately half of lung cancers in the USA occur in individuals who do not meet NLST criteria, which further challenges the concept of clinical risk stratification for lung cancer screening. Transcriptomic profiling has the potential to augment existing clinical risk models with molecular input that reflects lung cancer risk.³³

Beyond identifying patients that have clinical and molecular risk factors that warrant screening examinations to detect early stage lung cancer, transcriptomic profiling also has the potential to identify patients at a higher risk for future development of lung cancer well before the disease is clinically detectable. In a proof of concept study, gene expression profiling by RNA sequencing of normal and dysplastic bronchial cells and tumour tissue from four subjects undergoing the resection of squamous cell carcinoma delineated gene expression patterns that could be categorised as 'early' changes (those that are similarly altered in tumour and dysplastic cells, but different in normal epithelial cells), and 'late' changes (those that are similarly expressed in normal and dysplastic cells, but differently in tumour cells).³⁴ Early events related to protein ubiquitination and cell cycle progression, whereas late events were enriched in functions related to cellular migration and transformation. While limited by a relatively small sample size, this study illustrates a definable 'stepwise' chain of events in the molecular progression of lung cancer from initiation to actual cancer development. Transcriptomic alterations associated with this premalignant disease progression may ultimately serve as novel screening biomarkers to identify those smokers at highest risk to develop lung cancer, ultimately allowing the identification of patients who might benefit from increased lung cancer surveillance or targeted chemoprevention.

Beyond profiling the premalignant lesion itself, gene expression studies of the cytologically normal airway epithelium within the field of injury of high-risk smokers holds the potential to identify screening biomarkers. This approach is based on the paradigm that the airway gene expression response to smoking reflects a gene-by-environment interaction that could identify that subset of smokers that are at highest risk for future lung cancer development. As proof of concept, Blomquist *et al*³⁵ found that variability in the expression level of 14 antioxidant

and DNA repair genes in the cytologically normal bronchial epithelium could distinguish smokers with and without lung cancer, with an area under the curve of 0.87 in an independent cross-sectional cohort of 40 patients. While these findings suggest that an aberrant airway gene expression response to smoking associates with lung cancer, these studies have yet to be conducted in prospective studies of high-risk smokers where airway gene expression changes are observed prior to lung cancer development.

As was the case in the diagnostic setting, there is a need to move screening biomarkers to more readily accessible extrathoracic biospecimens. One recent promising study demonstrated the potential feasibility of utilising serum microRNA expression as a screening biomarkers in a prospective European LDCT trial.³⁶ This study generated a microRNA signature that distinguished patients undergoing lung cancer screening that were ultimately diagnosed with lung cancer from healthy control subjects. This signature had a sensitivity and specificity of 80% for the likelihood of developing lung cancer in an independent validation cohort.³⁶ While this report demonstrates early changes in plasma which is appealing in its potential use to further earlier intervention in a minimally invasive fashion, microRNA-based platform may prove logistically challenging given technical challenges in isolating microRNA from serum.

Personalising chemoprevention: the airway transcriptome as an intermediate marker of therapeutic efficacy

Beyond the early identification of individuals at risk for developing lung cancer, there is also a need for effective pharmacologic interventions to reduce this risk. The lack of approved chemopreventive agents for lung cancer is a testimony to the challenges this field faces.³³ These hurdles stem, in part, from the heterogeneity of lung cancer, multiple molecular drivers of lung carcinogenesis and long follow-up times needed in clinical studies of chemoprevention to identify incident lung cancer and medication effectiveness. Transcriptomic biomarkers may help to surmount these challenges by providing quantifiable surrogate markers for important endpoints, and facilitate the identification of novel approaches to preventive pharmacotherapy.

As an example of the promise of this approach, an *in vitro* gene expression signature of the oncogenic pathway PI3K was shown to be activated in the cytologically normal bronchial airway epithelium obtained from both patients with dysplastic airway lesions who did not have lung cancer, and from patients with lung cancer.³⁷ This gene expression pathway signature reversed in subjects with airway dysplasia that regressed after treatment with myoinositol, a medication that targets the PI3K pathway.³⁷ These findings suggest that gene expression changes in the airway epithelial field of injury occur prior to the development of lung cancer, and are key markers of early lung carcinogenesis that can be leveraged to both guide which chemopreventive agent to use as well as measure response to that intervention.

Leveraging the transcriptome to identify new chemopreventive agents via drug repurposing

Drug development, including the development of effective chemopreventive agents for lung cancer, is a costly process fuelled by the need for large studies with long follow-up periods, in addition to the need to demonstrate safety and efficacy. The repositioning of Food and Drug Administration (FDA)-approved therapeutics for novel indications holds the promise to overcome these burdens and also expedite the translation to

intercepting and preventing lung cancer given the established safety profiles of these agents.

Transcriptomic profiling offers a unique opportunity to reposition existing drugs for chemoprevention given the public availability of microarray data sets of cell line response to large numbers of FDA-approved drugs and compounds. The feasibility of this approach has been demonstrated in several studies, including using gene expression signatures of disease to identify the tripeptide GHK as an agent that reverses an emphysema lung tissue signature,³⁸ and the antiseizure medication valproic acid as a novel therapy for triple-negative breast cancer that has been validated in a xenograph mode.³⁹ This approach has also been successful in identifying potential novel therapeutics for lung cancer, including the identification of cimetidine as a potential treatment for lung adenocarcinoma,⁴⁰ and the tricyclic imipramine and the antihistamine promethazine as predicated therapies for small cell lung cancer.⁴¹ Taken together, these combined models that leverage *in silico* data to predict *in vivo* response illustrate the potential to identify and prioritise existing compounds for lung cancer chemoprevention trials.

THE FUTURE OF TRANSCRIPTOMIC BIOMARKERS IN LUNG CANCER EARLY DETECTION, SCREENING, AND CHEMOPREVENTION

While recent studies have produced a number of promising transcriptomic biomarkers in the lung cancer diagnostic setting, bringing these biomarkers to the bedside requires stringent validation through large multicentre prospective studies in which the test is applied in the clinical setting in which it will be used. Specifically, these markers should demonstrate clinical utility among subjects with intermediate risk for disease (eg, indeterminate pulmonary nodules) where they can impact the clinical management (ie, whether to biopsy the nodule vs follow non-invasively with serial imaging). There is a pressing need to expand these transcriptomic markers into the screening setting, studies that are currently hindered by the lack of longitudinal screening cohorts where biosamples are collected/banked for RNA expression. Importantly, the transcriptomic profiling of premalignant lesions may help identify novel biomarkers that can serve as tools for identifying smokers at highest risk for disease as well as serve as intermediate biomarkers of therapeutic efficacy in chemoprevention trials. Finally, gene expression profiling offers the possibility of repositioning existing drugs into the chemoprevention space and potentially personalise lung cancer prevention by matching the right patient to the right drug. These approaches can be further facilitated by the collection and storage of RNA from biosamples in ongoing and future screening and chemoprevention trials with the ultimate goal of developing precise genomic tools to guide clinical decisions.

Contributors KS and AS contributed equally as cosenior authors; YBG, KS and AS reviewed the literature. YBG, JV, KS and AS structured and drafted the manuscript. YG is responsible for the overall content.

Funding AS is a consultant to Veracyte.

Competing interests None.

Provenance and peer review Commissioned; internally peer reviewed.

REFERENCES

- 1 Cancer of the Lung and Bronchus—SEER Stat Fact Sheets [Internet] (cited 18 November 2014). <http://seer.cancer.gov/statfacts/html/lungb.html>
- 2 Aberle DR, Adams AM, Berg CD, *et al*; National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.

- 3 Croswell JM, Baker SG, Marcus PM, *et al.* Cumulative incidence of false-positive test results in lung cancer screening: a randomized trial. *Ann Intern Med* 2010;152:505–12, W176–80.
- 4 Hassanein M, Callison JC, Callaway-Lane C, *et al.* The state of molecular biomarkers for the early detection of lung cancer. *Cancer Prev Res Phila Pa* 2012;5:992–1006.
- 5 Steiling K, Ryan J, Brody JS, *et al.* The field of tissue injury in the lung and airway. *Cancer Prev Res Phila Pa* 2008;1:396–403.
- 6 Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012;489:519–25.
- 7 Collisson EA, Campbell JD, Brooks AN, *et al.* Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014;511:543–50.
- 8 Spira A, Beane J, Shah V, *et al.* Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci USA* 2004;101:10143–8.
- 9 Beane J, Sebastiani P, Liu G, *et al.* Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol* 2007;8:R201.
- 10 Spivack SD, Hurteau GJ, Jain R, *et al.* Gene-environment interaction signatures by quantitative mRNA profiling in exfoliated buccal mucosa cells. *Cancer Res* 2004;64:6805–13.
- 11 Harvey BG, Heguy A, Leopold PL, *et al.* Modification of gene expression of the small airway epithelium in response to cigarette smoking. *J Mol Med* 2007;85:39–53.
- 12 Turetz ML, O'Connor TP, Tilley AE, *et al.* Trachea epithelium as a "Canary" for cigarette smoking-induced biologic phenotype of the small airway epithelium. *Clin Transl Sci* 2009;2:260–72.
- 13 Schembri F, Sridhar S, Perdomo C, *et al.* MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. *Proc Natl Acad Sci USA* 2009;106:2319–24.
- 14 Sridhar S, Schembri F, Zeskind J, *et al.* Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* 2008;9:259.
- 15 Boyle JO, Gümüš ZH, Kacker A, *et al.* Effects of cigarette smoke on the human oral mucosal transcriptome. *Cancer Prev Res Phila Pa* 2010;3:266–78.
- 16 Zhang X, Sebastiani P, Liu G, *et al.* Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium. *Physiol Genomics* 2010;41:1–8.
- 17 MacMahon H, Austin JHM, Gamsu G, *et al.* Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner society¹. *Radiology* 2005;237:395–400.
- 18 Schreiber G, McCrory DC. Performance characteristics of different modalities for diagnosis of suspected lung cancer: summary of published evidence. *Chest* 2003;123(1 Suppl):115S–28S.
- 19 Wiener RS, Schwartz LM, Woloshin S, *et al.* Population-based risk for complications after transthoracic needle lung biopsy of a pulmonary nodule: an analysis of discharge records. *Ann Intern Med* 2011;155:137.
- 20 Spira A, Beane JE, Shah V, *et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 2007;13:361–6.
- 21 Beane J, Sebastiani P, Whitfield TH, *et al.* A prediction model for lung cancer diagnosis that integrates genomic and clinical features. *Cancer Prev Res Phila Pa* 2008;1:56–64.
- 22 Validation Of A Bronchial Airway Multi-Gene Test To Improve Diagnostic Yield In Suspect Lung Cancer Cases: Results Of The Prospective AEGIS I Trial—ajrccm-conference.2014.189.1_MeetingAbstracts.A6567 [Internet] (cited 18 November 2014). http://www.atsjournals.org/doi/pdf/10.1164/ajrccm-conference.2014.189.1_MeetingAbstracts.A6567
- 23 Kadara H, Fujimoto J, Yoo S-Y, *et al.* Transcriptomic architecture of the adjacent airway field cancerization in non-small cell lung cancer. *JNCI J Natl Cancer Inst* 2014;106:dju004.
- 24 Perdomo C, Campbell JD, Gerrein J, *et al.* MicroRNA 4423 is a primate-specific regulator of airway epithelial cell differentiation and lung carcinogenesis. *Proc Natl Acad Sci USA* 2013;110:18946–51.
- 25 Zhang L, Xiao H, Zhou H, *et al.* Development of transcriptomic biomarker signature in human saliva to detect lung cancer. *Cell Mol Life Sci CMLS* 2012;69:3341–50.
- 26 Xing L, Todd NW, Yu L, *et al.* Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers. *Mod Pathol* 2010;23:1157–64.
- 27 Yu L, Todd NW, Xing L, *et al.* Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. *Int J Cancer J Int Cancer* 2010;127:2870–8.
- 28 Showe MK, Vachani A, Kossenkov AV, *et al.* 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;69:9202–10.
- 29 Rotunno M, Hu N, Su H, *et al.* A gene expression signature from peripheral whole blood for stage I lung adenocarcinoma. *Cancer Prev Res (Phila Pa)* 2011;4:1599–608.
- 30 Sozzi G, Boeri M, Rossi M, *et al.* Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: a correlative MILD trial study. *J Clin Oncol Off J Am Soc Clin Oncol* 2014;32:768–73.
- 31 Bianchi F, Nicassio F, Marzi M, *et al.* A serum circulating miRNA diagnostic test to identify asymptomatic high-risk individuals with early stage lung cancer: serum miRNAs to diagnose asymptomatic NSCLC. *EMBO Mol Med* 2011;3:495–503.
- 32 Pritchard CC, Kroh E, Wood B, *et al.* Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer Prev Res (Phila Pa)* 2012;5:492–7.
- 33 Kathuria H, Gesthalter Y, Spira A, *et al.* Updates and controversies in the rapidly evolving field of lung cancer screening, early detection, and chemoprevention. *Cancers* 2014;6:1157–79.
- 34 Ooi AT, Gower AC, Zhang KX, *et al.* Molecular profiling of premalignant lesions in lung squamous cell carcinomas identifies mechanisms involved in stepwise carcinogenesis. *Cancer Prev Res Phila Pa* 2014;7:487–95.
- 35 Blomquist T, Crawford EL, Mullins D, *et al.* Pattern of antioxidant and DNA repair gene expression in normal airway epithelium associated with lung cancer diagnosis. *Cancer Res* 2009;69:8629–35.
- 36 Boeri M, Verri C, Conte D, *et al.* MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. *Proc Natl Acad Sci USA* 2011;108:3713–18.
- 37 Gustafson AM, Soldi R, Anderlind C, *et al.* Airway PI3K pathway activation is an early and reversible event in lung cancer development. *Sci Transl Med* 2010;2:26ra25.
- 38 Campbell JD, McDonough JE, Zeskind JE, *et al.* A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. *Genome Med* 2012;4:67.
- 39 Cohen AL, Soldi R, Zhang H, *et al.* A pharmacogenomic method for individualized prediction of drug sensitivity. *Mol Syst Biol* 2011;7:513.
- 40 Sirota M, Dudley JT, Kim J, *et al.* Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med* 2011;3:96ra77.
- 41 Jahchan NS, Dudley JT, Mazur PK, *et al.* A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors. *Cancer Discov* 2013;3:1364–77.