Revised manuscript CAPR 11-0441R-2

The state of molecular biomarkers for the early detection of lung cancer

Mohamed Hassanein[1], J. Clay Callison[1], Carol Callaway-Lane[2], Melinda C. Aldrich[3,4], Eric L. Grogan[2,3,], Pierre P. Massion[1,2*]

[1]Division of Allergy, Pulmonary and Critical Care Medicine, Vanderbilt-Ingram Cancer Center, Tennessee; [2]Veterans Affairs, Tennessee Valley Healthcare System, Nashville Campus, [3]Department of Thoracic Surgery and [4]Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN.

*Correspondence to: Pierre P. Massion, M.D. Email: pierre.massion@vanderbilt.edu.

Running title: Lung cancer molecular biomarker candidates

Key words: Biomarkers, early detection, diagnosis, lung cancer.

Author address: Pierre P. Massion, MD Vanderbilt Ingram Cancer Center, PRB 640, 2220 Pierce Avenue, Nashville TN 37232

There are no conflicts of interest to declare for any of the authors.

Abstract

Using biomarkers to select the most at-risk population, to detect the disease while measurable and yet not clinically apparent has been the goal of many investigations. Recent advances in molecular strategies and analytical platforms including genomics, epigenomics, proteomics, and metabolomics, have identified increasing numbers of potential biomarkers in the blood, urine, exhaled breath condensate, bronchial specimens, saliva, and sputum, but none have yet moved to the clinical setting. Therefore, there is a recognized gap between the promise and the product delivery in the cancer biomarker field. In this review, we define clinical contexts where risk and diagnostic biomarkers may have utility in the management of lung cancer, identify the most relevant candidate biomarkers of early detection, provide their state of development, and finally discuss critical aspects of study design in molecular biomarkers for early detection of lung cancer.

1. Introduction

Lung cancer is the leading cause of cancer-related death in the United States (1). More than 60% of patients are diagnosed at advanced stages when a cure is unlikely (2). Five-year survival for patients with advanced disease is less than 10%, while 5-year survival is greater than 70% in patients with stage 1 disease (3). The annual mortality rate for lung cancer exceeds the annual rate for breast, prostate, and colon cancer combined, all of which have successful clinical screening tools for the detection of early stage disease (4). For this reason, the search for diagnostic strategies for early lung cancer detection has intensified.

Until recently, the case for early detection in lung cancer was extrapolated from other cancers such as colon and breast. Clinicians and scientists continued to hypothesize that the earlier lung cancer is diagnosed, the opportunity for improved survival increases. Historically, lung cancer has been difficult to detect early and thus survival advantages were difficult to ascertain. In the 1970s and 1980s, chest x-ray and sputum cytology were tested in screening trials for lung cancer. Although this approach increase the number of lung cancers diagnosed, it did not improve lung cancer-specific mortality (5, 6). The Early Lung Cancer Action Program (ELCAP) was a large lung cancer screening trial started in the 1990's utilizing chest CT imaging (7). It showed an improved detection rate and survival of early stage lung cancers, which prompted the design of a large randomized National Lung Screening Trial (NLST). Exciting results of the recently completed study showed a 20% reduction in lung cancer-specific mortality using low dose computed tomography (CT) screening for patients at high risk for lung cancer after a median follow up of 6.5-years, compared with chest x-ray (8, 9). This is the first large randomized screening study of lung cancer by low-dose chest CT to show an improvement in

overall survival, thus giving new hope in the survival for this cancer. Extrapolating from the NLST results, a screening method that reduces lung cancer specific mortality by 20% could save an estimated 11,074 lives annually in the U.S., which is far greater than 2,303, the number currently estimated to be saved with adjuvant chemotherapy (see supplement) therefore providing a strong rationale to pursue efforts in early detection.

How do we define early detection? Early detection involves a high-risk population, a screening test, and a testing schedule. Within this context, one must distinguish populations of individuals at risk before or after the disease becomes measurable (Figure 1).

What clinical endpoints do the biomarker candidates of early detection address? A distinction is made between risk biomarkers to assess the risk of developing lung cancer (individuals at risk but with no measurable disease), and diagnostic biomarkers to determine whether cancer is present (individuals at risk with measurable asymptomatic disease such as lung nodules). Prognostic biomarkers in patients with early stage disease can identify individuals with an aggressive phenotype and shorter survival regardless of the type of treatment provided, and may help select populations who may benefit from adjuvant therapy. The biomarkers of risk of developing lung cancer in the absence of measurable disease are only discussed should the biomarker be developed originally as a diagnostic biomarker. The literature on biomarkers of risk of developing lung cancer based on proteins or SNPs and including GWAS studies is beyond the scope of this review. Likewise, prognostic biomarkers will not be discussed further.

What are the benchmarks for clinical utility? To be useful in the clinical setting, biomarkers go through careful phases of development as discussed below and should respond to specific criteria. The biomarkers should (1) be quantifiable and reproducible, (2) have good

testing performance (with good positive predictive value and negative predictive value), (3) be measurable in accessible material, in small amounts and with little preparation, (4) indicate a disease state, (5) have proven clinical utility, (6) be adopted by the community-at-large to take advantages of the benefits testing affords, (7) be cost effective and (8) reimbursed by health insurers.

2. The clinical context of early detection:

To be successful at improving lung cancer detection, biomarkers must address a specific clinical question. Two pressing clinical needs are identified, biomarkers that will address the risk of developing lung cancer, and others that are diagnostic in nature and will distinguish malignant from benign nodules.

The risk of developing lung cancer: Biomarkers of risk for lung cancer have the potential to improve early detection beyond the use of CT scans that suffer from lack of sensitivity (particularly among never smokers), specificity (high false positive rate), and from high cost. Several published models exist that predict an individual's risk of developing lung cancer (10-15). These models were developed to select patients who may benefit from additional radiographic screening. Identifying a risk biomarker for developing lung cancer would further define the at-risk population, decrease the overall number of screening CTs performed, and ultimately limit the downstream consequences of discovering these "false positive" nodules.

Distinguishing benign from malignant lung nodules: Diagnostic biomarkers that may assist in distinguishing a benign nodule from a malignant one would be invaluable. Depending on geographic location, up to 30% of indeterminate pulmonary nodules are ultimately found to

have benign pathology when surgically resected (16). In the NLST, 24% of patients who underwent a diagnostic operation (mediastinoscopy, thoracoscopy, or thoracotomy) had benign disease (17). Thus, additional testing with a biomarker could decrease the number of surgical resections for ultimately benign disease. Current guidelines recommend that providers use models to assist with determining the likelihood that a particular nodule identified by CT scan is malignant and thus should be resected (18). However, as the use of low dose CT for lung cancer screening evolves, better predictive models that incorporate biomarkers would assist the clinical provider in determining which patients have lung cancer. We recently validated a blood-based proteomic signature for lung cancer diagnosis and demonstrated that it may provide added value to the clinical and imaging assessment of indeterminate lung nodules (19). Hopefully as biomarkers are developed, they will assist in identifying not only those individuals without malignancy, but also help determine those that are malignant and amenable to surgical resection.

3. Current status of early detection biomarkers for lung cancer

We will review the most recent advances made to date in the field of molecular biomarkers for risk assessment and diagnosis of lung cancer, as well as discuss the clinical utility and limitations of different approaches.

In an effort to identify the most relevant lung cancer biomarkers of early detection, we selected published reports from PubMed based on the key words biomarkers, risk, diagnosis, early detection and lung cancer. To narrow our search we further applied the following two filters. First, the proposed marker, or panel of markers must be quantitatively measurable and its performance tested in at least one sample set of clinically relevant specimens. Second, the report

adhered to the PRoBE biomarker validation guidelines discussed later in this review. As a result we have selected original reports summarized in Tables 1, 2, and 3 of this review. We recognize the limitation of selection, outcome reporting and publication biases.

We have organized our report based on specimen types, either tissue-based (Table 1) or biofluids-based markers. We have further subcategorized biofluids-based biomarkers into blood-based (Table 2), sputum, WBC and peripheral blood cells (Table 3). The phases of biomarker development are assessed following the Early Detection Research Network (EDRN) classification (20). This classification was designed for biomarkers of early detection in the context of screening and therefore may not directly address phases of development of diagnostic biomarkers. These tables also include efforts from investigators to integrate biomarkers in models of risk prediction or diagnosis. The validation sets reported in the tables correspond to an attempt to test the biomarker (or signature) in a true independent population, also described as clinical validation (4), to evaluate the performance of the test.

3.a. Tissue-based candidate biomarkers

Numerous studies have adapted large scale analytical approaches to profile the full spectrum of molecular aberrations associated with lung cancer malignancy in tumors. These studies have yielded valuable information that has unraveled several key molecular events of lung cancer tumorigenesis, including mapping the genomic loci associated with high risk of developing lung cancer, hypermethylation of a number of tumor suppressor genes (21-25), regions of chromosomal amplification (26, 27), mRNA expression variation (28-31), the

differential expression of several micro RNAs (32) and the proteomic signature of invasive (33, 34) and preinvasive lesions in lung tissues (35, 36).

Several research groups, including ours, have been testing genetic and proteomic alterations in surrogate tissues such as bronchial brushings and biopsies to determine the probability of having lung cancer (25-27, 36, 37), see Table 1. While this approach requires bronchoscopy, molecular markers obtained from the airways may pair with the recently proven CT screening to provide additional benefit when evaluating individuals at high risk for lung cancer. Although much of the early biomarker discovery efforts have used fresh frozen samples as a primary source, acquiring these specimens is costly and laborious. Because surgical pathology specimens stored as formalin fixed paraffin embedded (FFPE) block are widely available, many researchers are attempting to prolife genomic and proteomic aberrations in such specimens. Some of these aberrations include hypermethylation of genes (38) and microRNA expression (39), which can be successfully extracted as candidate biomarkers.

MicroRNAs (miRNAs) are a class of small noncoding RNA genes that are thought to regulate gene expression. They are abnormally expressed in several types of cancer (40, 41) and involved in a variety of biologic and pathologic processes with tissue specificity (41, 42) with the potential for clinical application (43). Another advantage of microRNA is that it is well preserved in formalin-fixed tissue, making it ideal for use in routinely processed material (44). Previous studies have identified differences in microRNA expression between squamous cell carcinoma and adenocarcinoma in lung cancer (32), as well as in other cancers (42, 43, 45). The current trend towards utilizing FFPE samples will allow for a greater number of available samples, and thus will increase statistical power and generalization of results.

Tissue-based biomarkers that reflect the molecular changes associated with specific histological subtypes of non-small cell lung cancer (NSCLC) may provide the means to differentiate tumors originating in the lung from metastases from other organ sites. Furthermore, using immunohistochemical profiling of lung cancer tissue markers in conjunction with well-established histological examination can provide more accurate sub-classification of lung malignancies, and thus may directly impact the clinical decision making of anti-tumor therapy.

The molecular changes associated with progression from normal to malignant tissue may lead to the discovery of novel markers that can be detected in circulation or other biofluids. Limited access to early stage tumor tissue samples, tumor heterogeneity combined with the complexity of the genome and the proteome, and the low abundance of potential biomarkers represent some of the challenges that translational researchers face when attempting to bring these biomarker candidates from the bench to the bedside.

Therefore, the potential utility of tissue-based biomarkers is highly dependent on the accessibility of the specimens and the robustness of the assay offered. FFPE samples are preferred by scientists because of their availability but their molecular analyses remain more challenging. Although non-invasive diagnostic approaches are also preferred, it may take additional time to refine an airway epithelium-based biomarker versus one that can derive the same information from a less invasive sample. For example developing surrogate biomarkers of early stage disease from tissues in the field of cancerization (bronchial brushings or biopsies) may require testing in more proximal and less invasive samples (e.g. nasal epithelium). This problem may be less acute for prognostic biomarkers and biomarkers predictive of response to therapy because tumor samples will be generally available for analysis. Although molecular

analysis of lung tumor tissues holds great promise to revolutionize our understanding of the disease development and progression, tissue-based biomarkers from the bronchial airway have significant limitations related to tissue acquisition that may be overcome by the translation of that knowledge to more accessible specimens and by guiding the development of biofluids-based early detection strategies.

3.b. Biofluids-based markers

The underlying premise of biofluids-based biomarker research is that molecular alterations of tumor cells lead to the synthesis of distinct molecular species that can be detected in biofluids. Biofluids-based detection strategies are an attractive approach for screening, namely due to their ease of acquisition. Biofluids including peripheral blood and its components (circulating cells, plasma, and serum), exhaled breath condensate (EBC), urine, and sputum offer non-invasive access to large quantities of samples available for analysis. These alterations can lead to the generation of disease specific molecular species such as altered or methylated DNA, overexpressed mRNA, microRNA or proteins that can potentially be released into the extracellular microenvironment. Therefore, molecular analyses of early stage lung cancer-related biofluids represent an attractive choice for the discovery and validation of diagnostic biomarkers (46, 47).

3.b.1 Blood-based markers

Blood is a complex and dynamic medium whose components can reflect various physiological or pathological states such as the presence of some cancers. Detectable moieties of the blood are

currently the subject of many investigations and include cellular elements such as circulating tumor cells, cell-free DNA and RNA, proteins, peptides, and metabolites. Changes of the cell-free genomic components of the blood including DNA methylation (48, 49), DNA amplification, and gene expression (50) have been reported in the circulation of lung cancer patients. These candidates are reported in Table 2.

microRNAs. More recently, microRNAs have also been identified in the blood of lung cancer patients (51, 52). In an effort to test the validity of miRNA as biomarkers able to predict lung tumor development, diagnosis and prognosis, an extensive microRNA profiling was performed in paired lung tumor and normal lung tissue and in plasma collected at the time of diagnosis by spiral CT. A signature of 15 microRNAs present in the blood was able to identify subjects at high risk of developing lung cancer in two independent cohorts of patient with (80% sensitivity) and (90% specificity) (53). These results suggest that microRNA expression ratios may be molecular predictors of lung cancer development and aggressiveness, and may have clinical implication for lung cancer management in the future. In a separate study a test included 34 serum miRNAs, that could identify patients with early stage non-small cell lung carcinomas (NSCLCs) in a population of asymptomatic high-risk individuals with 80% accuracy (54) . These provocative results will have to be validated in independent cohorts.

Proteomic profiles. Recent proteomic studies have focused on rapid proteomic profiling of blood with minimal sample preparation. One of these approaches uses matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI TOF MS) patterns of abundant proteins or peptide fragments that correlate with early disease stage. Several other studies utilized MALDI MS to identify proteins and peptides in serum. For example, Patz and

colleagues were able to identify four differentially expressed serum proteins (transferrin, retinol-binding protein, antitrypsin and haptoglobin) that discriminate between NSCLC and controls (55). Using the same MALDI MS approaches, several other groups including ours (56) have reported serum protein expression profiles that distinguish patients with various cancers from control subjects (57). Recently, we validated that proteomic signature in two prospective cohorts of patients with lung nodules and demonstrated that it may provide added value to the clinical and imaging assessment of indeterminate lung nodules (19).

Autoantibodies. Other promising developments in the blood biomarkers field is the discovery of autoantibodies directed against tumor proteins. Alterations of the protein production in cancer cells by either overexpression, mutations, misfolding, truncation, or proteolysis break immunologic tolerance and generate tumor specific antigens which in turn elicit a host immune response (58). Autoantibodies generated against these tumor associated antigens (TAA) during the course of disease progression can be further amplified by the immune network, making them attractive candidates for the early detection and diagnosis of cancer. Many TAA targets have been identified in patient sera in several immunologic diseases and malignancies using high throughput screening platforms, such as cDNA expression libraries, phage display, and protein microarrays (59). For example, two separate groups identified several potential immunoreactive peptides for autoantibodies using a T-7 cDNA-based phage library to screen the sera of patients with NSCLC (60, 61). Using similar techniques, Chen *et al* also identified and validated ubiquilin 1 peptides as a potential autoantibody target in lung adenocarinoma from sera of patients with early stage lung cancer (62). More recently, Wu *et al* reported the identification of 6 peptide clones discriminatory of NSCLC using phage display

techniques, but only one protein has been confirmed (63). Recent improvements in blood fractionation techniques and liquid chromatography led to the identification of several other autoantibodies (47). Autoantibodies against known lung cancer associated proteins such as autoantibodies against p53, c-Myc, HER2, MUC1, CAGE, GBU4-5, NY-ESO-1 (64) or annexin I, PGP9.5, and 14-3-3 theta, LAMAR1 (65), or IMPDH, PGAM1 and ANXA2 (66) have also been reported as independent signatures. These recent autoantibody studies are particularly provocative because some allow the detection of cancer specific markers in the preclinical phase of lung cancer progression. Similar to other circulating protein markers, the low abundance of autoantibodies and the complexity of the blood proteome are still substantial challenges facing these discovery efforts.

Circulating tumor cells (CTCs). The ability to capture and study CTCs is an emerging and interesting development in the field that carries the potential to become a noninvasive tool for early detection and diagnosis of cancer, measuring response to therapy, as well as for understanding the basic biology of cancer progression and metastasis (67-71). CTCs are rare cells that originate from a malignancy and circulate freely in the peripheral blood. CTCs are usually captured by immobilized antiepithelial-cell-adhesion-molecule (EpCAM, also known as TACSTD1) antibodies either in chip or bead platforms (72, 73). The technology of rare CTCs capture is still in the early phase of development and requires more specific surface markers to increase its specificity for circulating lung cancer cells.

In summary, peripheral blood is a rich medium for cancer-specific markers from small molecules such as microRNAs to whole cells, all of which represent a great opportunity for developing a minimally invasive diagnostic test of lung cancer. Significant challenges are still

preventing the clinical success of blood markers, including the extreme complexity of the blood matrix, the scarce quantity of any given marker, and the lack of sensitive, reproducible, and high throughput verification modalities, in particular in proteomics research. New and innovative fractionation techniques, more sensitive and specific detection reagents, and well validated assays will increase our chances of capturing blood-based biomarkers.

3.b.2 Exhaled-breath condensate

The analysis of exhaled breath condensate (EBC) represents another non-invasive method of diagnosing lung cancer. The analysis of volatile organic compounds (VOCs) that are linked to cancer is likely to provide a novel opportunity for the identification of diagnostic cancer biomarkers because such a large volume of sample can be collected easily and inexpensively (74-76). The underlying rationale of this approach is based on the observation that tumor cell growth is accompanied by the alteration of protein expression pattern that may lead to peroxidation of the cell membrane, and thus to the emission of VOCs (76). Several recent studies have utilized gas chromatography combined with mass spectrometric analysis (GC-MS) of VOCs as both discovery and validation platforms (77-81). Other groups have utilized the analytical power of GC-MS and the sensitivity of custom designed nanosensors in which changes in electrical resistance from organic compounds contained in exhaled breath of patients can be detected by these sensors and recorded. For example in a study by Peng *et al,* a VOC signature that distinguished patients with lung, colorectal, and breast cancers from healthy individuals was recently identified from exhaled alveolar breath (82). These candidates are reported in Table 3. Other studies attempted to identify volatile proteins and peptides present in

EBC and used them as potential markers for the early detection of lung cancer (83-85). The results of these studies provide evidence for feasibility of this strategy to isolate and identify proteins useful for early detection of lung cancer. Further studies are still needed to standardize a collection device, to further demonstrate specificity of any test, and to determine the utility of this approach in clinical practice.

3.b.3 Sputum and urine

Cigarette smoking leads to the increased production of sputum with glycoproteins, inflammatory cells and exfoliated cells from the bronchial tree. Because sputum is so readily available, particularly in current and ex-smokers, its molecular analysis has been an active area of research for lung cancer biomarkers (21). Although detecting lung cancer using sputum cytology alone has low sensitivity (86), several studies showed that combining cytology with analysis of genetic abnormalities improves diagnosis accuracy. Several types of genetic abnormalities have been detected in the sputum of patients with lung cancer, such as deletions of *HYAL2, FHIT*, and *SFTPC* (87), chromosomal aneusomy (88, 89), DNA methylation (90, 91), and microRNA (92, 93). These candidates are reported in Table 3. Also recently, measurements of genomic aneuploidy when combined with pulmonary function can significantly improves lung cancer risk prediction (94). The performance of most of these potential markers has not been tested in large scale validation studies and whether these markers will add value to standardized sputum cytology remains to be seen.

Urine, much like blood, EBC, and sputum, is another easily accessible biofluid that could be an important source of cancer specific markers. Some recent proof of principle studies have

attempted to profile the molecular changes of urine using mass spectrometric analysis. The molecular species that were detected in urine include VOCs previously identified in an animal model of lung cancer (95) and their investigation in patients with lung cancer is just beginning (96).

4. Study design for early detection biomarkers validation

Appropriate study design is crucial for the successful validation of a promising biomarker for clinical use. Validation of a biomarker useful for lung cancer screening should be conducted using a nested case-control study design within a prospective longitudinal cohort following the PRoBE design (97). Specifically, random sampling of cases and controls identified from within a well-defined cohort population allows both cases and controls to be sampled from the same source population, thus providing validity to the case-control design. Matching strategies may be considered, such as using incidence density sampling to sample controls at the same time each case occurs so that cases and controls are matched on time. While there are advantages to matching the potential pitfalls of matching should be carefully considered prior to implementation (97, 98).

Generalizability of biomarkers to the appropriate clinical setting and populations is requisite for a clinically useful biomarker. The prospective cohort from whom the cases and controls are sampled must be representative of the targeted clinical population to which the biomarker will be applied. Thus the cohort study population should comprise individuals with conditions found in the target population, such as inflammatory disease, granulomas, or benign

tumors so that false positives can be minimized and individuals developing lung cancer can be differentiated from those not developing the disease (99).

Biospecimens necessary for biomarker development should be collected at the initiation of the prospective cohort study, prior to ascertainment of lung cancer status (97), and potentially over multiple time points if the biomarker changes with age and with progression to disease (99). These biospecimens are then evaluated in patients who develop biopsy proven cancer (cases) and those who do not (controls) to develop a biomarker for clinical use as a cost efficient approach. Importantly, the outcome should be clearly defined (100) and the biomarker assay development should be blinded to case-control status to avoid information bias (97). To validate the usefulness of a biomarker for early detection of lung cancer, diagnostic validation of the biomarker should be conducted in a different population than the one in which the biomarker was developed. Finally, this should be followed by early diagnosis validation using a screening trial with lung cancer mortality as the endpoint (101).

Assessing whether a biomarker has clinical validity requires estimation of sensitivity and specificity, which can be summarized with the receiver operator characteristic curve (ROC) (102). Two additional clinically relevant measures that can be measured by ROC include negative predictive value (NPV) and positive predictive value (PPV), which are estimated using sensitivity and specificity. These clinically important indices describe the probability of developing or having disease given a positive test. Estimates of PPV and NPV are influenced by the prevalence of the disease and consequently will vary by patient age, target population, and disease stage. Merely targeting the screening to a high-risk population based on demographic factors can alter the screening test performance characteristics (103). Thus for a biomarker to be

clinically valid and generalizable, the biomarker validation process must be applied to multiple populations having different demographic characteristics for determining the clinical validity and utility of a biomarker.

The use of lung cancer diagnosis prediction models will grow as the models accuracy improves. When a patient presents in the clinic and undergoes imaging, e.g. computed tomography resulting in a detected pulmonary nodule, current predictive models for assessing lung cancer malignancy include those developed by Cummings (104), Gurney (105), Swensen (106) and Gould (18). However, these models suffer from relatively poor accuracy in particular for indeterminate pulmonary nodules, and do not provide accurate prediction of malignancy among patients referred for surgery (107). While these models may include predictors such as patient age, smoking status, duration and intensity of smoking, cancer history, gender, race, asbestos exposure, COPD/emphysema, and pulmonary nodule characteristics by CT (size, shape, density, and location), the addition of biomarkers may provide additional classification accuracy to the current lung cancer malignancy prediction models. Recent interest has focused on the potential for molecular tools to improve models predicting lung cancer diagnosis yet most studies have shown little improvement with added gene expression profile in cytologically normal large airway epithelium obtained via bronchoscopic brushings (28), or a serum proteomic profile in patients presenting with pulmonary nodules (19). While molecular markers are not yet fully incorporated into lung cancer malignancy prediction models, it is likely that a profile of molecular markers will be necessary to be clinically useful as biomarkers for early detection of lung cancer (108). Future development of predictive models should incorporate previously identified predictors and newly identified biomarkers (100).

Current Challenges in Lung Cancer Biomarker Development and Implementation

One of the main objectives of molecular medicine in lung cancer is to identify biomarkers that discriminate between low versus high risk individuals and between benign and malignant lung tumors. Ultimately these biomarkers can potentially be translated to noninvasive, simple, and reliable diagnostic tests for early detection of the disease. The underlying assumption behind these efforts is that tumor-specific or overexpressed proteins can be detected simply and accurately in complex clinical samples such as surrogate tissues and biofluids. The intensive research in genomics and proteomics aimed at identifying these biomarkers has yielded a large number of potential diagnostic biomarkers, although few have progressed to the level of FDA-approval for diagnostics (109).

This disappointingly slow pace of lung cancer biomarkers discovery and validation is attributed to a host of technological and methodological factors. The gap between promise and product can partially be explained by the fact that the current discovery methods are neither reliable nor efficient. One reason is that the current analytical technologies still suffer from the limited power to detect low-abundant cancer markers against a high background of high-abundance molecular species such as proteins in very complex matrices such as plasma or serum. These low-abundance markers in biofluids may be the most promising cancer biomarkers. Consequently, many of the best candidates may thus be missed during the discovery phase.

Another quandary is the limited capacity to verify and validate analytically existing candidate markers in a high-throughput manner. This is particularly true in proteomics research. The lack of available quality reagents such as antibodies, or methodologies to translate the discovery of candidates in tissue specimens and measure their concentration in the circulation

remains an enormous challenge. Therefore it is possible that biomarkers have already been "discovered," but not yet validated. Furthermore, once a long list of candidate biomarkers is compiled, no current standardized method exists for selecting those that are most promising for systematic validation. In addition, the reproducibility of biomarker data has been flawed because of the poor design (e.g. underrepresentation of studies using a nested case-control design (97)), model over-fitting, and the lack of cross-validation and independent validation. Changing technology, low concentration of signals combined with very few prospective studies, and a low incidence disease, make the area of biomarker research challenging.

Conclusions and Future Clinical Implications

The molecular analysis of a variety of biospecimens has allowed the discovery of relevant candidate biomarkers and consequently the identification of novel proteins that may have a role in the development of lung cancer. A high volume of data from multiple high-throughput biochemical analyses of clinical material from "-omics" sources has been accumulating at an exponential rate in the last few years, generating large number of biomarker candidates. None of the published candidate biomarkers of risk or of lung cancer diagnosis are ready for clinical use, and few have moved to phase III of biomarker development. Lung cancer is recognized as a complex and heterogeneous disease, not only at the biochemical level (genes, proteins, metabolites), but also at the tissue, organism, and population level. There is a need for incorporating findings from multiple discovery platforms into a mathematical framework that can improve our level of understanding of the disease process. A biofluids-based molecular test may improve the selection of high risk individuals for CT screening, distinguish those with

malignant nodules from benign lesions, and identify patients with particularly aggressive cancer. Clinical benefit could include further reductions in mortality and thus provide significant cost-savings to the healthcare system.

Acknowledgment: We would like to acknowledge the following funding source for support of this work: National Institutes of Health (U01CA152662 and RO1 CA102353) awarded to PPM, Veterans Health Administration (VA-CDA2) awarded to ELG. A Vanderbilt Clinical and Translational Research Scholars Award awarded to MCA. We would like to thank Mr. William Alborn for his insightful comments and suggestions. We thank Judith Roberts from the Vanderbilt Tumor Registry for her help in providing data to support the estimates of lives saved from adjuvant chemotherapy in the US.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61:69-90.

2. Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev. 2010;19:1893-907.

3. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. Lancet. 2000;355:479-85.

4. Brenner DE, Normolle DP. Biomarkers for cancer risk, early detection, and prognosis: the validation conundrum. Cancer Epidemiol Biomarkers Prev. 2007;16:1918-20.

5. Fontana RS, Sanderson DR, Taylor WF, Woolner LB, Miller WE, Muhm JR, et al. Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Mayo Clinic study. Am Rev Respir Dis. 1984;130:561-5.

6. Stitik FP, Tockman MS. Radiographic screening in the early detection of lung cancer. Radiol Clin North Am. 1978;16:347-66.

7. Henschke CI, McCauley DI, Yankelevitz DF, Naidich DP, McGuinness G, Miettinen OS, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet. 1999;354:99-105.

8. Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, Galen B, et al. The National Lung Screening Trial: overview and study design. Radiology. 2011;258:243-53.

9. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. The New England journal of medicine. 2011;365:395-409.

10. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, et al. Variations in lung cancer risk among smokers. J Natl Cancer Inst. 2003;95:470-8.

11. Cassidy A, Duffy SW, Myles JP, Liloglou T, Field JK. Lung cancer risk prediction: A tool for early detection. International Journal of Cancer. 2007;120:1-6.

12. Peto R, Darby S, Deo H, Silcocks P, Whitley E, Doll R. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. Bmj. 2000;321:323-9.

13. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, et al. A risk model for prediction of lung cancer. J Natl Cancer Inst. 2007;99:715-26.

14. D'Amelio AM, Jr., Cassidy A, Asomaning K, Raji OY, Duffy SW, Field JK, et al. Comparison of discriminatory power and accuracy of three lung cancer risk models. Br J Cancer. 2010;103:423-9.

15. Tammemagi CM, Pinsky PF, Caporaso NE, Kvale PA, Hocking WG, Church TR, et al. Lung cancer risk prediction: prostate, lung, colorectal and ovarian cancer screening trial models and validation. Journal of the National Cancer Institute. 2011;103:1058-68.

16. Isbell JM, Deppen S, Putnam Jr JB, Nesbitt JC, Lambright ES, Dawes A, et al. Existing General Population Models Inaccurately Predict Lung Cancer Risk in Patients Referred for Surgical Evaluation. The Annals of Thoracic Surgery. 2011;91:227-33.

17. Berg ea. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. The New England journal of medicine. 2011.

18. Gould MK, Ananth L, Barnett PG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. Chest. 2007;131:383-8.

19. Pecot CV, Li M, Zhang XJ, Rajanbabu R, Calitri C, Bungum A, et al. Added Value of a Serum Proteomic Signature in the Diagnostic Evaluation of Lung Nodules. Cancer Epidemiol Biomarkers Prev. 2012.

20. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst. 2001;93:1054-61.

21. Belinsky SA. Gene-promoter hypermethylation as a biomarker in lung cancer. Nat Rev Cancer. 2004;4:707-17.

22. Anglim PP, Galler JS, Koss MN, Hagen JA, Turla S, Campan M, et al. Identification of a panel of sensitive and specific DNA methylation markers for squamous cell lung cancer. Molecular Cancer. 2008;7:62.

23. Castro M, Grau L, Puerta P, Gimenez L, Venditti J, Quadrelli S, et al. Multiplexed methylation profiles of tumor suppressor genes and clinical outcome in lung cancer. Journal of Translational Medicine. 2010;8:86.

24. Richards KL, Zhang B, Sun M, Dong W, Churchill J, Bachinski LL, et al. Methylation of the candidate biomarker TCF21 is very frequent across a spectrum of early-stage nonsmall cell lung cancers. Cancer. 2011;117:606-17.

25. Schmidt B, Liebenberg V, Dietrich D, Schlegel T, Kneip C, Seegebarth A, et al. SHOX2 DNA Methylation is a Biomarker for the diagnosis of lung cancer based on bronchial aspirates. BMC Cancer. 2010;10:600.

26. Halling KC, Rickman OB, Kipp BR, Harwood AR, Doerr CH, Jett JR. A comparison of cytology and fluorescence in situ hybridization for the detection of lung cancer in bronchoscopic specimens. Chest. 2006;130:694-701.

27. Massion PP, Zou Y, Uner H, Kiatsimkul P, Wolf HJ, Baron AE, et al. Recurrent genomic gains in preinvasive lesions as a biomarker of risk for lung cancer. PLoS One. 2009;4:e5611.

28. Beane J, Sebastiani P, Whitfield TH, Steiling K, Dumas Y-M, Lenburg ME, et al. A Prediction Model for Lung Cancer Diagnosis that Integrates Genomic and Clinical Features. Cancer Prev Res. 2008:1940-6207.CAPR-08-0011.

29. Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med. 2002;8:816-24.

30. Blomquist T, Crawford EL, Mullins D, Yoon Y, Hernandez DA, Khuder S, 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.

31. Wilkerson MD, Yin X, Hoadley KA, Liu Y, Hayward MC, Cabanski CR, et al. Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. Clin Cancer Res. 2010;16:4864-75.

32. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell. 2006;9:189-98.

33. Yanagisawa K, Shyr Y, Xu BJ, Massion PP, Larsen PH, White BC, et al. Proteomic patterns of tumour subsets in non-small-cell lung cancer. Lancet. 2003;362:433-9.

34. Yanagisawa K, Tomida S, Shimada Y, Yatabe Y, Mitsudomi T, Takahashi T. A 25-signal proteomic signature and outcome for patients with resected non-small-cell lung cancer. J Natl Cancer Inst. 2007;99:858-67.

35. Rahman SM, Shyr Y, Yildiz PB, Gonzalez AL, Li H, Zhang X, et al. Proteomic patterns of preinvasive bronchial lesions. Am J Respir Crit Care Med. 2005;172:1556-62.

36. Rahman SMJ, Gonzalez AL, Li M, Seeley EH, Zimmerman LJ, Zhang XJ, et al. Lung Cancer Diagnosis from Proteomic Analysis of Preinvasive Lesions. Cancer Research. 2011;71:3009-17.

37. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. Nat Med. 2007;13:361-6.

38. Feng Q, Hawes SE, Stern JE, Wiens L, Lu H, Dong ZM, et al. DNA Methylation in Tumor and Matched Normal Tissues from Non-Small Cell Lung Cancer Patients. Cancer Epidemiology Biomarkers & Prevention. 2008;17:645-54.

39. Lebanony D, Benjamin H, Gilad S, Ezagouri M, Dov A, Ashkenazi K, et al. Diagnostic Assay Based on hsa-miR-205 Expression Distinguishes Squamous From Nonsquamous Non-Small-Cell Lung Carcinoma. Journal of Clinical Oncology. 2009;27:2030-7.

40. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A. 2006;103:2257-61.

41. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435:834-8.

42. Roldo C, Missiaglia E, Hagan JP, Falconi M, Capelli P, Bersani S, et al. MicroRNA expression abnormalities in pancreatic endocrine and acinar tumors are associated with distinctive pathologic features and clinical behavior. J Clin Oncol. 2006;24:4677-84.

43. Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, Zepeniuk M, et al. MicroRNAs accurately identify cancer tissue origin. Nat Biotechnol. 2008;26:462-9.

44. Li J, Smyth P, Flavin R, Cahill S, Denning K, Aherne S, et al. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. BMC Biotechnol. 2007;7:36.

45. Feber A, Xi L, Luketich JD, Pennathur A, Landreneau RJ, Wu M, et al. MicroRNA expression profiles of esophageal cancer. J Thorac Cardiovasc Surg. 2008;135:255-60; discussion 60.

46. Hanash S. Harnessing immunity for cancer marker discovery. Nat Biotechnol. 2003;21:37-8.

47. Hanash SM, Pitteri SJ, Faca VM. Mining the plasma proteome for cancer biomarkers. Nature. 2008;452:571-9.

48. Greenberg AK, Rimal B, Felner K, Zafar S, Hung J, Eylers E, et al. S-adenosylmethionine as a biomarker for the early detection of lung cancer. Chest. 2007;132:1247-52.

49. Begum S, Brait M, Dasgupta S, Ostrow KL, Zahurak M, Carvalho AL, et al. An Epigenetic Marker Panel for Detection of Lung Cancer Using Cell-Free Serum DNA. Clinical Cancer Research. 2011;17:4494-503.

50. Showe MK, Vachani A, Kossenkov AV, Yousef M, Nichols C, Nikonova EV, 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 Research. 2009;69:9202-10.

51. Lai CY, Yu SL, Hsieh MH, Chen CH, Chen HY, Wen CC, et al. MicroRNA Expression Aberration as Potential Peripheral Blood Biomarkers for Schizophrenia. PloS one. 2011;6:e21635.

52. Li S, Zhu J, Zhang W, Chen Y, Zhang K, Popescu LM, et al. Signature microRNA Expression Profile of Essential Hypertension and Its Novel Link to Human Cytomegalovirus Infection. Circulation. 2011;124:175-84.

53. Boeri M, Verri C, Conte D, Roz L, Modena P, Facchinetti F, et al. MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. Proceedings of the National Academy of Sciences. 2011;108:3713-8.

54. Bianchi F, Nicassio F, Marzi M, Belloni E, Dall'olio V, Bernard L, et al. A serum circulating miRNA diagnostic test to identify asymptomatic high-risk individuals with early stage lung cancer. EMBO Mol Med. 2011;3:495-503.

55. Patz EF, Jr., Campa MJ, Gottlin EB, Kusmartseva I, Guan XR, Herndon JE, 2nd. Panel of serum biomarkers for the diagnosis of lung cancer. J Clin Oncol. 2007;25:5578-83.

56. Yildiz PB, Shyr Y, Rahman JS, Wardwell NR, Zimmerman LJ, Shakhtour B, et al. Diagnostic accuracy of MALDI mass spectrometric analysis of unfractionated serum in lung cancer. J Thorac Oncol. 2007;2:893-901.

57. Ocak S, Chaurand P, Massion PP. Mass spectrometry-based proteomic profiling of lung cancer. Proc Am Thorac Soc. 2009;6:159-70.

58. Caron M, Choquet-Kastylevsky G, Joubert-Caron R. Cancer immunomics using autoantibody signatures for biomarker discovery. Mol Cell Proteomics. 2007;6:1115-22.

59. Feng Z, Prentice R, Srivastava S. Research issues and strategies for genomic and proteomic biomarker discovery and validation: a statistical perspective. Pharmacogenomics. 2004;5:709-19.

60. Zhong L, Coe SP, Stromberg AJ, Khattar NH, Jett JR, Hirschowitz EA. Profiling tumor-associated antibodies for early detection of non-small cell lung cancer. J Thorac Oncol. 2006;1:513-9.

61. Khattar NH, Coe-Atkinson SP, Stromberg AJ, Jett JR, Hirschowitz EA. Lung cancer-associated auto-antibodies measured using seven amino acid peptides in a diagnostic blood test for lung cancer. Cancer Biol Ther. 2010;10.

62. Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. The New England journal of medicine. 2007;356:11-20.

63. Wu L, Chang W, Zhao J, Yu Y, Tan X, Su T, et al. Development of autoantibody signatures as novel diagnostic biomarkers of non-small cell lung cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010;16:3760-8.

64. Chapman CJ, Murray A, McElveen JE, Sahin U, Luxemburger U, Tureci O, et al. Autoantibodies in lung cancer: possibilities for early detection and subsequent cure. Thorax. 2008;63:228-33.

65. Qiu J, Choi G, Li L, Wang H, Pitteri SJ, Pereira-Faca SR, et al. Occurrence of autoantibodies to annexin I, 14-3-3 theta and LAMR1 in prediagnostic lung cancer sera. J Clin Oncol. 2008;26:5060-6.

66. Farlow EC, Vercillo MS, Coon JS, Basu S, Kim AW, Faber LP, et al. A multi-analyte serum test for the detection of non-small cell lung cancer. Br J Cancer. 2010;103:1221-8.

67. Stahel RA, Mabry M, Skarin AT, Speak J, Bernal SD. Detection of bone marrow metastasis in small-cell lung cancer by monoclonal antibody. J Clin Oncol. 1985;3:455-61.

68. Peck K, Sher YP, Shih JY, Roffler SR, Wu CW, Yang PC. Detection and quantitation of circulating cancer cells in the peripheral blood of lung cancer patients. Cancer Res. 1998;58:2761-5.

69. Pachmann K, Camara O, Kavallaris A, Schneider U, Schunemann S, Hoffken K. Quantification of the response of circulating epithelial cells to neodadjuvant treatment for breast cancer: a new tool for therapy monitoring. Breast Cancer Res. 2005;7:R975-9.

70. Pachmann K, Heiss P, Demel U, Tilz G. Detection and quantification of small numbers of circulating tumour cells in peripheral blood using laser scanning cytometer (LSC). Clin Chem Lab Med. 2001;39:811-7.

71. Rolle A, Gunzel R, Pachmann U, Willen B, Hoffken K, Pachmann K. Increase in number of circulating disseminated epithelial cells after surgery for non-small cell lung cancer monitored by MAINTRAC(R) is a predictor for relapse: A preliminary report. World J Surg Oncol. 2005;3:18.

72. Nagrath S, Sequist LV, Maheswaran S, Bell DW, Irimia D, Ulkus L, et al. Isolation of rare circulating tumour cells in cancer patients by microchip technology. Nature. 2007;450:1235-9.

73. Wu C, Hao H, Li L, Zhou X, Guo Z, Zhang L, et al. Preliminary investigation of the clinical significance of detecting circulating tumor cells enriched from lung cancer patients. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer. 2009;4:30-6.

74. Amann A, Spanel P, Smith D. Breath analysis: the approach towards clinical applications. Mini Rev Med Chem. 2007;7:115-29.

75. Mazzone P. Progress in the development of a diagnostic test for lung cancer through the analysis of breath volatiles. J Breath Res. 2008;2:037014.

76. Mazzone PJ. Analysis of volatile organic compounds in the exhaled breath for the diagnosis of lung cancer. J Thorac Oncol. 2008;3:774-80.

77. Phillips M, Gleeson K, Hughes JMB, Greenberg J, Cataneo RN, Baker L, et al. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. The Lancet. 1999;353:1930-3.

78. Fuchs P, Loeseken C, Schubert JK, Miekisch W. Breath gas aldehydes as biomarkers of lung cancer. International Journal of Cancer. 2009:NA-NA.

79. Bajtarevic A, Ager C, Pienz M, Klieber M, Schwarz K, Ligor M, et al. Noninvasive detection of lung cancer by analysis of exhaled breath. BMC Cancer. 2009;9:348.

80. Ligor M, Ligor T, Bajtarevic A, Ager C, Pienz M, Klieber M, et al. Determination of volatile organic compounds in exhaled breath of patients with lung cancer using solid phase microextraction and gas chromatography mass spectrometry. Clin Chem Lab Med. 2009;47:550-60.

81. Mazzone PJ, Hammel J, Dweik R, Na J, Czich C, Laskowski D, et al. Diagnosis of lung cancer by the analysis of exhaled breath with a colorimetric sensor array. Thorax. 2007;62:565-8.

82. Peng G, Hakim M, Broza YY, Billan S, Abdah-Bortnyak R, Kuten A, et al. Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. British Journal of Cancer. 2010;103:542-51.

83. Kurova VS, Kononikhin AS, Sakharov DA, Popov IA, Larina IM, Tonevitskii AG, et al. [Exogenic proteins in the human exhaled breath condensate]. Bioorg Khim. 2011;37:55-60.

84. Gessner C, Rechner B, Hammerschmidt S, Kuhn H, Hoheisel G, Sack U, et al. Angiogenic markers in breath condensate identify non-small cell lung cancer. Lung Cancer. 2010;68:177-84.

85. Kurova VS, Anaev EC, Kononikhin AS, Fedorchenko KY, Popov IA, Kalupov TL, et al. Proteomics of exhaled breath: methodological nuances and pitfalls. Clin Chem Lab Med. 2009;47:706-12.

86. Melamed MR. Lung cancer screening results in the National Cancer Institute New York study. Cancer. 2000;89:2356-62.

87. Li R, Todd NW, Qiu Q, Fan T, Zhao RY, Rodgers WH, et al. Genetic deletions in sputum as diagnostic markers for early detection of stage I non-small cell lung cancer. Clin Cancer Res. 2007;13:482-7.

88. Varella-Garcia M, Kittelson J, Schulte AP, Vu KO, Wolf HJ, Zeng C, et al. Multi-target interphase fluorescence in situ hybridization assay increases sensitivity of sputum cytology as a predictor of lung cancer. Cancer Detect Prev. 2004;28:244-51.

89. Katz RL, Zaidi TM, Fernandez RL, Zhang J, He W, Acosta C, et al. Automated detection of genetic abnormalities combined with cytology in sputum is a sensitive predictor of lung cancer. Mod Pathol. 2008;21:950-60.

90. Palmisano WA, Divine KK, Saccomanno G, Gilliland FD, Baylin SB, Herman JG, et al. Predicting lung cancer by detecting aberrant promoter methylation in sputum. Cancer Res. 2000;60:5954-8.

91. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer Res. 2002;62:2370-7.

92. Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z, et al. Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. International Journal of Cancer. 2010;127:2870-8.

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;23:1157-64.

94. Tammemagi MC, Lam SC, McWilliams AM, Sin DD. Incremental value of pulmonary function and sputum DNA image cytometry in lung cancer risk prediction. Cancer Prev Res (Phila). 2011;4:552-61.

95. Matsumura K, Opiekun M, Oka H, Vachani A, Albelda SM, Yamazaki K, et al. Urinary volatile compounds as biomarkers for lung cancer: a proof of principle study using odor signatures in mouse models of lung cancer. PLoS One. 2010;5:e8819.

96. Li Y, Zhang Y, Qiu F, Qiu Z. Proteomic identification of exosomal LRG1: A potential urinary biomarker for detecting NSCLC. Electrophoresis. 2011;32:1976-83.

97. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. J Natl Cancer Inst. 2008;100:1432-8.

98. Janes H, Pepe MS. Matching in studies of classification accuracy: implications for analysis, efficiency, and assessment of incremental value. Biometrics. 2008;64:1-9.

99. Baker SG, Kramer BS, Srivastava S. Markers for early detection of cancer: Statistical guidelines for nested case-control studies. BMC Med Res Methodol. 2002;2:4.

100. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ. 2009;338:b606.

101. Baker SG. Improving the biomarker pipeline to develop and evaluate cancer screening tests. Journal of the National Cancer Institute. 2009;101:1116-9.

102. Taylor JM, Ankerst DP, Andridge RR. Validation of biomarker-based risk prediction models. Clin Cancer Res. 2008;14:5977-83.

103. Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. Clin Chem. 2004;50:473-6.

104. Cummings SR, Lillington GA, Richard RJ. Estimating the probability of malignancy in solitary pulmonary nodules. A Bayesian approach. Am Rev Respir Dis. 1986;134:449-52.

105. Gurney JW, Lyddon DM, McKay JA. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part II. Application. Radiology. 1993;186:415-22.

106. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. Arch Intern Med. 1997;157:849-55.

107. Isbell JM, Deppen S, Putnam JB, Jr., Nesbitt JC, Lambright ES, Dawes A, et al. Existing general population models inaccurately predict lung cancer risk in patients referred for surgical evaluation. Ann Thorac Surg. 2011;91:227-33; discussion 33.

108. Pepe MS, Gu JW, Morris DE. The potential of genes and other markers to inform about risk. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2010;19:655-65.

109. Anderson JE, Hansen LL, Mooren FC, Post M, Hug H, Zuse A, et al. Methods and biomarkers for the diagnosis and prognosis of cancer and other diseases: towards personalized medicine. Drug Resist Updat. 2006;9:198-210.

110. Hassanein M, Rahman JS, Chaurand P, Massion PP. Advances in proteomic strategies toward the early detection of lung cancer. Proc Am Thorac Soc. 2011;8:183-8.

111. Kim B, Lee HJ, Choi HY, Shin Y, Nam S, Seo G, et al. Clinical Validity of the Lung Cancer Biomarkers Identified by Bioinformatics Analysis of Public Expression Data. Cancer Research. 2007;67:7431-8.

112. Farlow EC, Patel K, Basu S, Lee BS, Kim AW, Coon JS, et al. Development of a multiplexed tumor-associated autoantibody-based blood test for the detection of non-small cell lung cancer. Clin Cancer Res. 2010;16:3452-62.

113. Boyle P, Chapman CJ, Holdenrieder S, Murray A, Robertson C, Wood WC, et al. Clinical validation of an autoantibody test for lung cancer. Ann Oncol. 2011;22:383-9.

114. Chen X, Hu Z, Wang W, Ba Y, Ma L, Zhang C, et al. 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. 2011.

115. Kulpa J, Wojcik E, Reinfuss M, Kolodziejski L. Carcinoembryonic antigen, squamous cell carcinoma antigen, CYFRA 21-1, and neuron-specific enolase in squamous cell lung cancer patients. Clin Chem. 2002;48:1931-7.

116. Takano A, Ishikawa N, Nishino R, Masuda K, Yasui W, Inai K, et al. Identification of Nectin-4 Oncoprotein as a Diagnostic and Therapeutic Target for Lung Cancer. Cancer Research. 2009;69:6694-703.

117. Diamandis EP, Goodglick L, Planque C, Thornquist MD. Pentraxin-3 Is a Novel Biomarker of Lung Carcinoma. Clinical Cancer Research. 2011;17:2395-9.

118. Ostroff RM, Bigbee WL, Franklin W, Gold L, Mehan M, Miller YE, et al. Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer. PLoS ONE. 2010;5:e15003.

119. Zhong L, Hidalgo GE, Stromberg AJ, Khattar NH, Jett JR, Hirschowitz EA. Using protein microarray as a diagnostic assay for non-small cell lung cancer. Am J Respir Crit Care Med. 2005;172:1308-14.

120. Kneip C, Schmidt B, Seegebarth A, Weickmann S, Fleischhacker M, Liebenberg V, et al. SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. J Thorac Oncol. 2011;6:1632-8.

121. Shen J, Todd NW, Zhang H, Yu L, Lingxiao X, Mei Y, et al. Plasma microRNAs as potential biomarkers for non-small-cell lung cancer. Laboratory Investigation. 2010;91:579-87.

122. Wei J, Gao W, Zhu CJ, Liu YQ, Mei Z, Cheng T, et al. Identification of plasma microRNA-21 as a biomarker for early detection and chemosensitivity of non-small cell lung cancer. Chin J Cancer. 2011;30:407-14.

123. Taguchi A, Politi K, Pitteri SJ, Lockwood WW, Faca VM, Kelly-Spratt K, et al. Lung cancer signatures in plasma based on proteome profiling of mouse tumor models. Cancer Cell. 2011;20:289-99.

124. Palmisano WA, Crume KP, Grimes MJ, Winters SA, Toyota M, Esteller M, et al. Aberrant promoter methylation of the transcription factor genes PAX5 alpha and beta in human cancers. Cancer Res. 2003;63:4620-5.

125. Varella-Garcia M, Akduman B, Sunpaweravong P, Di Maria MV, Crawford ED. The UroVysion fluorescence in situ hybridization assay is an effective tool for monitoring recurrence of bladder cancer. Urol Oncol. 2004;22:16-9.

126. Tanaka F, Yoneda K, Kondo N, Hashimoto M, Takuwa T, Matsumoto S, et al. Circulating Tumor Cell as a Diagnostic Marker in Primary Lung Cancer. Clinical Cancer Research. 2009;15:6980-6.

Figure legends

Figure 1: Clinical contexts for biomarker development in early detection of lung cancer. This diagram illustrates four clinical contexts within 4 windows of time. The period during which lung cancer is non-measurable and precedes the diagnosis characterizes the context of *risk assessment*. It represents a long window of time during which the disease develops and corresponds to an opportunity for chemoprevention. When the disease becomes measurable but remains asymptomatic, we enter the context of *early diagnosis*. Two other clinical contexts relate to *clinical diagnosis,* i.e. when the disease is measurable and patients symptomatic, and to *detection of recurrence*. These windows of time correspond to the different contexts for which different biomarker targets can be developed. Adapted from Hassanein et al., 2011(110). Adapted with permission of the American Thoracic Society. Copyright © 2012 American Thoracic Society.

Table 1. Characteristics and performance of most recent tissue-based candidate biomarkers for the early detection of lung cancer. (data organized by year of publication and type of marker considered)

References	Specimens	Type of marker	Analyte	Clinical Purpose	# markers	Pathological subtype	Assay platform	Preclinical Samples	BM Dev. Phase	Training set	Validation set	Sensitivity	Specificity	AUC
Halling, 2006 (26)	Bronchial Specimens	DNA	5p15, 7p12 (EGFR), 8q24 (C-MYC), CEP6	Diagnosis	4	NSCLC	FISH + Cytology	n/a	I	n/a	137	61-75*	83-100*	n/a
Massion, 2009 (27)	Bronchial biopsies	DNA	*TP63, MYC, CEP3, CEP6* + sputum cytology + demographics	Diagnosis	4	NSCLC	FISH	n/a	II	n/a	70	n/a	n/a	92.6
Feng , 2008 (38)	Tumors and normal tissues	DNA methylation	*RARB, BVES, CDKN2A, KCNH5, RASSF1, CDH13, RUNX , CDH1*	Diagnosis	8	NSCLC	Methylation array	n/a	I	49	n/a	n/a	n/a	n/a
Anglim, 2008 (22)	Tumors and normal tissues	DNA methylation	*GDNF, MTHFR, OPCML, TNFRSF25, TCF21, PAX8, PTPRN2 and PITX2*	Diagnosis	8	SCC	Methylation array	n/a	I	43	n/a	95.6*	95.6*	n/a
Schmidt, 2010 (25)	Bronchial aspirates	DNA methylation	SHOX2	Diagnosis	1	NSCLC	PCR	n/a	II	n/a	523	68	95	86
Richards, 2011 (24)	Tumors and normal tissues	DNA methylation	*TCF21*	Diagnosis	1	NSCLC	PCR	n/a	II	42	63	76	98*	n/a
Spira, 2007 (13)	airway epithelium	mRNA	gene expression signature	Diagnosis	80	NSCLC	Affy array	n/a	II	77	52	80	84	n/a
Beane, 2008 (28)	airway epithelium	mRNA	gene expression signature + clinical factors	Diagnosis	80	NSCLC & SCLC	Affy array	n/a	II	76	62	100	91	97
Kim, 2007 (111)	Tumors and normal tissues	mRNA	*CBLC, CYP24A1, ALDH3A1, AKR1B10, S100P, PLUNC, LOC147*	Diagnosis	7	NSCLC	q RT-PCR	n/a	II	32	36**	n/a	n/a	n/a
Blomquist, 2009 (30)	Tumors and normal tissues	mRNA	*CAT, CEBPG, E2F1, ERCC4, ERCC5, GPX1, GPX3, GSTM3, GSTP1, GSTT1, GSTZ1, MGST1, SOD1, XRCC1*	Diagnosis	14	NSCLC	RT-PCR	n/a	II	n/a	49;40	n/a	n/a	82-87
Rahman, 2011 (36)	Bronchial biopsies	MALDI signature	*TMLS4, ACBP, CSTA, cyto C, MIF*, ubiquitin, *ACBP*, Des-ubiquitin	Diagnosis	9	NSCLC	MALDI MS	n/a	II	51	60	66	88	77

Abbreviations. AUC= area under the curve, FISH= fluorescence in situ hybridization, MALDI-MS= matrix-assisted laser desorption/ionization mass spectrometry, miRNA= microRNA, NSCLC= non-small cell lung cancer, SCC= squamous cell carcinoma, SCLC= small cell lung cancer, RT-PCR= reverse transcriptase polymerase chain reaction,- q-PCR= quantitative (real time PCR), lung cancer, SNP= single nucleotide polymorphism, SCC= squamous cell carcinoma, Note***=** values derived from training set only; ** = validation and training sets overlap; BM Dev Phase= Biomarker Development Phase, n/a= not available.

Table 2. Characteristics and performance of most recent blood-based candidate biomarkers for the early detection of lung cancer. (data organized by year of publication, specimen type and type of marker considered)

References	Specimens	Type of marker	Analyte	Clinical Purpose	# markers	Pathological subtype	Assay platform	Preclinical Samples	BM Dev. Phase	Training set	Validation set	Sensitivity	Specificity	AUC
Zhong, 2006 (60)	Serum	AutoAB	phage peptide clones	Diagnosis	5	Lung cancer	ELISA	n/a	II	46	56	91*	91*	99*
Chapman, 2008 (64)	Serum	AutoAB	p53, cmyc, *HER2, NY-ESO-1, CAGE, MUC1,GBU4-5*	Diagnosis	7	Lung cancer	ELISA	n/a	I	154	n/a	n/a	n/a	n/a
Qiu, 2008 (65)	Serum	AutoAB	annexin I, 14-3-3 theta, LAMR1	Diagnosis	3	NSCLC	Protein-array	170	III		170	51	82	73
Wu, 2010 (63)	Serum	AutoAB	phage peptide clones	Diagnosis	6	NSCLC	ELISA	n/a	II	20	180	92	92	96
Farlow, 2010 (112)	Serum	AutoAB	*IMPDH, PGAM1*, ubiquillin, *ANXA1, ANXA2, HSP70-9B*	Diagnosis	6	NSCLC	ELISA	n/a	II	196	n/a	94.8*	91.1*	96.4*
Boyle, 2010 (113)	Serum	AutoAB	*p53, NY-ESO-1, CAGE, GBU4-5, Annexin 1* and *SOX2*	Diagnosis	6	NSCLC	ELISA	n/a	II	241	255	32	91	64
Greenberg, 2007 (48)	Serum	DNA methylation	S-Adenosylmethionine	Diagnosis	1	Lung cancer	HPLC	n/a	I	68	n/a	92-100*	91-97*	94-99*
Begum, 2011 (49)	Serum	DNA methylation	*APC, CDH1, MGMT, DCC, RASSF1A, AIM*	Diagnosis	6	NSCLC	q-PCR	n/a	II	32-639	106	84	57	n/a
Chen, 2011 (114)	Serum	micro RNA	micro RNA signature	Diagnosis	10	NSCLC	qRT-PCR	n/a	II	310	310	93	90	97
Bianchi, 2011 (54)	Serum	micro RNA	miRNA signature	Diagnosis	34	NSCLC	qRT-PCR	n/a	II	64	64	71	90	89
Kulpa , 2002 (115)	Serum	Protein	CEA, CYFRA 21-1, SCC-Ag, NSE	Diagnosis	4	SCC	ELISA	n/a	II		420	20-62	95	71-90
Patz, 2007 (55)	Serum	Protein	CEA,RBP4, hAAT, SCCA	Diagnosis	4	Lung cancer	ELISA	n/a	II	100	97	78	75	n/a
Takano, 2009 (116)	Serum	Protein	Nectin-4	Diagnosis	1	NSCLC	ELISA	n/a	II		295	54	98	n/a
Yildiz, 2007 (56)	Serum	Protein	MALDI MS signature	Diagnosis	7	NSCLC	MALDI MS	n/a	II	185	106	58	85.7	82
Pecot, 2012 (19)	Serum	Protein	Model: MALDI MS signature + clinical and imaging data	Diagnosis	7	Indeterm. Lung nodule	MALDI MS	n/a	II		100	n/a	n/a	72
Diamandis, 2011 (117)	Serum	Protein	Penatraxin-3	Diagnosis	1	Lung cancer	ELISA	n/a	I		426	37-48	80-90	60-74
Ostroff, 2010 (118)	Serum	Aptamers	*cadherin-1, CD30 ligand, endostatin, HSP90a, LRIG3,*	Diagnosis	6	NSCLC	Aptamers	n/a	II	985	341	89	83	90

			MIP-4, pleiotrophin, PRKCI, RGM-C, SCF-sR, sL-selectin, and YES											
Zhong, 2005 (119)	Plasma	AutoAB	TAA signature	Diagnosis	5	NSCLC	Protein microarray	5	I	81	n/a	90*	95*	n/a
Kneip, 2011 (120)	Plasma	DNA Methylation	SHOX2	Diagnosis	1	NSCLC	qPCR	n/a	II	40	371	60	90	78
Shen, 2010 (121)	Plasma	micro RNA	miRNA-21, -126, -210, 486-5p	Diagnosis	4	NSCLC	qRT-PCR	n/a	II	28	87	86	97	93
Wei, 2011 (122)	Plasma	micro RNA	miR-21	Diagnosis	1	NSCLC	qRT-PCR	n/a	I	93	n/a	76*	70*	77.5*
Taguchi, 2011 (123)	Plasma	Protein, 2 panels	*EGFR, SFTPB, WFDC2, ANGPTL3, ANXA1, YWHAQ, Lmr1*	Diagnosis	7	NSLCL	ELISA	52	III		n/a	n/a	n/a	89
Boeri, 2011 (53)	Plasma/ tissues	micro RNA	micro RNA signature	Diagnosis	13	Lung cancer	miRNA array & RT-PCR	n/a	II	19	22	75	100	88
Boeri 2011 (53)	Plasma/ tissues	micro RNA	micro RNA signature	Diagnosis	15	Lung cancer	miRNA array & RT-PCR	25	III	20	25	80	90	85

Abbreviations. AutoAB= autoantibody, AUC= area under the curve, ADC= adenocarcinoma, ELISA= enzyme linked immunosorbant assay, FISH= fluorescence in situ hybridization, HPLC= high performance liquid chromatography, MALDI-MS= matrix-assisted laser desorption/ionization mass spectrometry, miRNA= microRNA, NSCLC= non-small cell lung cancer, SCC+ squamous cell carcinoma, RT-PCR= reverse transcriptase polymerase chain reaction,- q-PCR= quantitative (real time PCR), SNP= single nucleotide polymorphism, SCC= squamous cell carcinoma, TAA= tumor associated antigen, Note*****= values derived from training set only; ** = validation and training sets overlap; BM Dev Phase= Biomarker Development Phase, n/a= not available.

Table 3. Characteristics and performance of most recent sputum, EBC, peripheral blood cells candidate biomarkers for the early detection of lung cancer. (data organized by year of publication, specimen type and type of marker considered)

References	Specimens	Type of marker	Analyte	Clinical Purpose	# markers	Pathological subtype	Assay platform	Preclinical Samples	BM Dev. Phase	Training set	Validation set	Sensitivity	Specificity	AUC
Palmisano, 2000 (124)	Sputum	DNA methylation	p16, MGMT	Diagnosis	2	SCC	PCR	11	III	144	n/a	n/a	n/a	n/a
Belinsky, 2002 (91)	Sputum	DNA methylation	*p16, MGMT, DAP, RASSFIA*	Diagnosis	4	NSCLC	PCR	n/a	I	141	n/a	n/a	n/a	n/a
Garcia, 2004 (125)	Sputum	DNA	Chromosomal aneusomy + cytology	Diagnosis	4	Lung cancer	FISH + Cytology	36	III	66	n/a	83*	80*	n/a
Li, 2007 (87)	Sputum	DNA	*HYAL2, FHIT, SFTPC*	Diagnosis	3	NSCLC	FISH	n/a	I	102	n/a	76*	92*	n/a
Yu, 2010 (92)	Sputum	micro RNA	miR-21, miR-486, miR-375, miR-200b	Diagnosis	4	NSCLC	qRT-PCR	n/a	II	72	122	70	80	84
Showe, 2009 (50)	Peripheral blood cells	mRNA	gene expression signature	Diagnosis	29	NSCLC	cDNA-array	n/a	II	228	55	76	82	n/a
Tanaka, 2009 (126)	Peripheral blood cells	CTC	Circulating tumor cells (CTCs)	Diagnosis	1	NSCLC	Cell search-system	n/a	I	150	n/a	30*	88*	60*
Philips, 1999 (77)	EBC	VOCs	VOCs signature	Diagnosis	22	NSCLC	GC-MS	n/a	I	108	n/a	100*	81*	n/a
Bajtarevic, 2009 (79)	EBC	VOCs	VOCs signature	Diagnosis	50	Lung cancer	GC-MS	n/a	I	96		52-80*	100*	n/a
Gessner, 2010 (84)	EBC	Protein	VEGF, bFGF, ANG,TNF-alpha, IL-8	Diagnosis	5	NSCLC	ELISA	n/a	I		74	n/a	n/a	99-100

Abbreviations. AutoAB= autoantibody, AUC= area under the curve, ADC= adenocarcinoma, EBC= exhaled breath condensate, CTCs= circulating tumor cells, ELISA= enzyme linked immunosorbant assay, FISH= fluorescence in situ hybridization, GC-MS= gas chromatography mass spectrometry, miRNA= microRNA, NSCLC= non-small cell lung cancer, q-PCR= quantitative (real time PCR), SCLC= small cell lung cancer, SNP= single nucleotide polymorphism, SCC= squamous cell carcinoma, VOCs= volatile organic compounds. Note*= values derived from training set only; ** = validation and training sets overlap; BM Dev Phase= Biomarker Development Phase, n/a= not available.

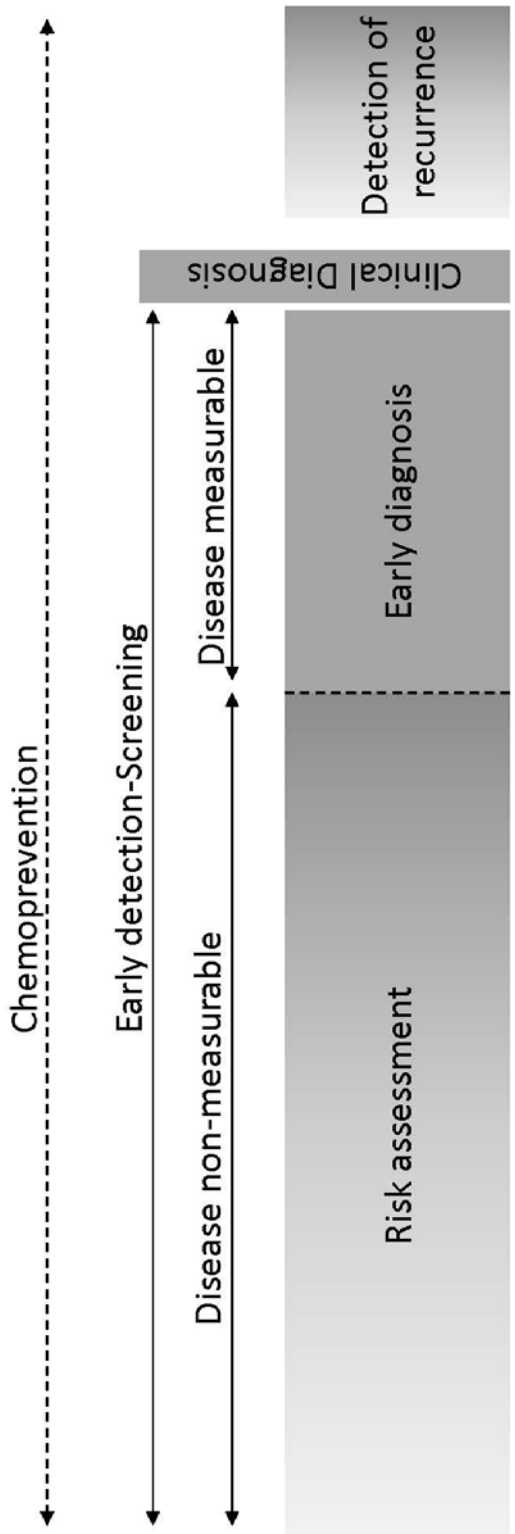

Figure 1

AACR American Association for Cancer Research

Cancer Prevention Research

The state of molecular biomarkers for the early detection of lung cancer

Mohamed Hassanein, J. Clay Callison, Carol Callaway-Lane, et al.

Cancer Prev Res Published OnlineFirst June 11, 2012.