Biomarkers in lung cancer screening: achievements, promises and challenges

Luis M. Seijo[a, b], Nir Peled[c], Daniel Ajona[d, e, f, g], Mattia Boeri[h], John K. Field[i], Gabriella Sozzi[h], Ruben Pio[d, e, f, g], Javier J Zulueta[j,k], Avrum Spira[l], Pierre P. Massion[m], Peter J. Mazzone[n], Luis M. Montuenga[d, e, f, ñ]

Corresponding author: Luis M. Montuenga

Affiliations
[a]L.S. Clinica Universidad de Navarra, Pamplona, Spain.
[b]Pulmonary Department, Instituto de Investigación Sanitaria, Fundación Jiménez Díaz, CIBERES, Madrid, Spain.
[c]Soroka Cancer Center, Ben-Gurion University, Beer Sheva, Israel.
[d]Solid Tumors Program, Centro de Investigación Médica Aplicada (CIMA), Pamplona, Spain.
[e]IdiSNA (Navarra Institute for Health Research), Pamplona, Spain.
[f]CIBERONC (Centro de Investigación Biomédica en Red de Cáncer), Spain.
[g]Department of Biochemistry and Genetics, School of Sciences, University of Navarra, Pamplona, Spain.
[h]Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.
[i]The Roy Castle Lung Cancer Research Programme, Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, UK.
[j]Department of Pulmonology, Clinica Universidad de Navarra, Pamplona, Spain.
[k]Visiongate Inc
[l]BU-BMC Cancer Center and Division of Computational Biomedicine, Boston University School of Medicine, Boston, MA, USA.
[m]Vanderbilt Ingram Cancer Center, Nashville, TN, USA.
[n]Cleveland Clinic, Cleveland, OH, United States.
[ñ]Department of Histology and Pathology, School of Medicine, University of Navarra, Pamplona, Spain.

Key words: Biomarkers; Lung Cancer; Screening; Genomics; Proteomics; Diagnosis; Early Detection; test validation

Acknowledgements

We thank Heber Longas (Fundamentium medical and scientific illustrator) for his contribution to the graphics of the present review, in particular Fig 2. L.M.M. and RP work is funded by FIMA, Spanish Ministry of Economy and Innovation and Fondo de Investigación Sanitaria-Fondo Europeo de Desarrollo Regional (FIS projects PI16/01821, PI17/00411 and CIBERONC), and AECC Scientific Foundation (GCB14-2170 and PROYE16091). G.S. work by investigator grants No.18812 and 5x1000 No. 12162 from the Italian Association for Cancer Research; M.B. was supported by a Fondazione Pezcoller Fellowship. P.P.M. work was supported by NCI CA152662.

Conflicts of interest:
DA, RP and LMM are listed as co-inventors of two patents related to complement fragments use as biomarkers. G.S and M.B. are co-inventors for three patent applications regarding the miRNA signature MSC. These patents were licensed to a private company, Gensignia Life Science, under regulations of Fondazione IRCCS Istituto Nazionale dei Tumori of Milan.

1. LUNG CANCER SCREENING CHALLENGES. IS IMAGING SUFFICIENT FOR SUCCESSFUL SCREENING? THE UNMET NEEDS

Current evidence supports screening of subjects who fit the NCI sponsored National Lung Screening Trial (NLST) criteria. The NLST compared annual low-dose computed tomography (LDCT) screening with conventional chest radiography and found that LDCT screening achieved a 20% lung cancer mortality reduction after only three rounds of screening[1]. Prospective data accrued by the International Early Lung Cancer Action Program (iELCAP) from thousands of screening rounds, also support annual screening with LDCT (Henschke et al 2006), while data from the European NELSON trial, a Dutch-Belgian initiative, are also favourable to screening[2]. Consequently, recommendations from prominent scientific societies support screening with LDCT despite concerns regarding false positive findings, the risk of overdiagnosis, logistical challenges, and differences in selection criteria[3–6]. Chief among these concerns is the widespread adoption of NLST age (55-74) and tobacco exposure (Pack-Years≥30) inclusion criteria by screening guidelines. Current evidence suggests that such criteria may preclude screening of many individuals at risk[7–11]. It is becoming clearer that a more sophisticated risk-based strategy, taking emphysema into account, for example, may be better than the current NLST criteria[7,12–17]

The advantage of the LDCT based protocol is its simplicity and its high sensitivity. Refined criteria defining positive findings, largely based on nodule size and/or volume, reduce false positive rates. That notwithstanding, there is a need for evidence-based biomarkers to support pre- and post-test (LDCT) risk assessment[18]. Ideally, robust biomarkers would optimize image-based screening in two ways. First, they would allow further refinement of screening selection criteria, independent of age and tobacco exposure, in order to limit the costs of lung cancer screening. This risk management biomarker strategy would be a welcome addition to current screening practice. For example, a number of single nucleotide polymorphisms (SNPs) have been proposed in this regard as potential biomarkers of constitutive genomic risk for a given individual (see paragraph 4.4). Such biomarkers are the focus of ongoing research when integrated with current clinical-epidemiological risk models for lung cancer[19].

Secondly, a validated panel of biomarkers may provide a post-test risk assessment capable of informing clinical decision making in the management of indeterminate pulmonary nodules (IPNs). Current management of IPNs is largely based on watchful waiting and may imply a risk of dissemination. Nodules found on annual screening, often so small that they are out of reach of current biopsy techniques, may benefit from a biomarker-based risk assessment. In particular biomarkers may be helpful in the case of patients with nodules that need sooner surveillance or a decision for biopsy. Patients with multiple nodules or those subject to frequent interval scans during screening might also benefit. Finally, biomarkers might also inform decisions regarding screening intervals, personalized follow up of survivors of screen detected early stage lung cancer, outcome prediction, or response to adjuvant therapy for those at high-risk of recurrence. In the present review we will refer only to the biomarkers intended for the first two unmet needs (risk management and IPN characterization). Other recent articles have dealt with early lung cancer prognostic biomarkers[20,21]. We will focus initially on biomarkers that are non-invasive, reproducible and validated, and conclude the current review with other promising technologies which are being developed in the context of early detection.

2. WHAT IS A GOOD BIOMARKER?

The National Institutes of Health define a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention"[22]. A useful biomarker must influence clinical decision making in a manner that leads to improved patient care. The benefit of clinical decisions based on true test results must outweigh the harms of decisions based on false positives or negatives. In the risk management setting, a biomarker should minimize harm and expense without leading to an increase in lung cancer deaths. When used for IPN characterization, a biomarker should anticipate the diagnosis of malignancy without substantially increasing the number of diagnostic procedures performed for benign nodules or delaying therapeutic procedures for malignant nodules[23].

2.1 Optimizing lung cancer biomarker discovery

We believe that validation of a clinically useful biomarker should adhere to the following principles summarized in table 1. **Study design**. Much is gained from the careful selection of the molecular approach chosen and should be guided by the intended use, where the biomarker would find potential clinical utility[24]. All too often investigators focus on versatility, seeking biomarkers which address multiple clinical needs such as risk assessment, diagnosis, or response to therapy. Much merit lies in limiting scope by addressing specific clinical needs. **Biomarker stability**. Information about the stability of the analyte over time, including changes in temperature, pH, enzymatic or oxidative stress is critical[25]. **Analytical validation**. Biomarker measurements should follow a well-defined strategy, and be accurate, precise and robust. Validation should include testing reproducibility against larger sources of variability such as biospecimen collection (e.g., sample processing, freeze thaw cycles, duration of storage, etc.), operator characteristics, laboratory environment, and quality control (standard curves, SOPs). Some variability is inherent to the technology itself (energy source, enzymatic activity, temperature control)[26]. Metrics of success include coefficient of variance, Z statistic, limits of detection, and quantitation. **Clinical validation**. The ideal diagnostic biomarker is both sensitive and specific, with diagnostic likelihood ratios independent of known predictors of the disease (e.g. age, smoking history or COPD). Validation should be performed in the clinical context of intended use. Case control studies are discouraged while prospective cohort studies and observational registries are preferable. The biomarker will be tested in multiple cohorts with similar prevalence of disease. Biomarkers rarely perform well across a large range of disease prevalence and their performance characteristics are often susceptible to changes in simple variables such as age or disease stage. **Clinical utility**. The biomarker should be tested for clinical utility in larger studies in a pragmatic setting, not disrupting the clinical workflow. The goal of any biomarker is to achieve superior performance compared to standard of care approaches and eventually reduce cost and harm of testing while limiting false negative rates[23]. Study design is challenging due to randomization and the need to impact clinical management (see below a potential trial design). Ultimately, implementation of the biomarker in routine practice will determine its true value for clinical decision making.

2.2. The metrics of success

Biomarker performance and accuracy are dependent on the intended use and current alternatives. A successful biomarker must supersede the current standard of care[27,28], be cost effective, welcomed by the community, and eventually demonstrate cancer control if early detection is the goal (Figure 1), or promote personalized medicine by identifying candidates for targeted therapies[1]. Understanding traditional metrics of success in this context is key. Sensitivity and specificity, for example, are often unstable over multiple variables such as age or disease stage. Positive and negative predictive values (PPV and NPV) are dependent on the prevalence of disease. ROC curves, i.e. true-positive vs false-positive rates (TPR and FPR), are helpful, but complicate decision making because of the need for dichotomous biomarker cut-off values. Reclassification indices have a role in testing a biomarker's ability to accurately reclassify cases and controls and therefore influence clinical decision making.

3. CURRENT PROMISING MOLECULAR BIOMARKERS

3.1 Molecular biomarkers for lung cancer screening

Blood is an obvious first choice as the source of biomarker candidates for lung cancer screening. Blood-based biomarkers provide an overview of the whole patient body, including the primary tumor, metastatic disease, immune response, and peri-tumoral stroma. However, sputum, bronchial lavage or aspirate samples, exhaled breath, or airway epithelium sampling are unique to lung and other respiratory tract cancers as potential sources of alternative biomarkers. These may provide information regarding molecular changes which may be anatomically closer to the tumor cells and their microenvironment and therefore potentially more relevant to clinical decision making in screened patients with early stage disease (Figure 2). Urine or saliva have also been collected as potential sources of biomarkers. The former is particularly useful in a metabolomics-based approach.

A concise review of the most prominent molecular biomarkers for lung cancer screening includes examples of molecular candidates for both risk management

and IPN characterization in diverse stages of validation. We have included those which we consider most promising. We are well aware of the risk of omitting potential candidates. Table 2 includes a list of biomarkers which have reached different levels of validation.

3.1.1. Autoantibodies

Autoantibodies (AAbs) develop in response to an abnormal tumor antigen in some patients with lung cancer, often in the pre-clinical phase well before symptoms appear or imaging-based detection is possible. Autoantibodies have been identified in all histologic types and stages of lung cancer. They are usually absent or found in low titers in those without cancer, but also in many patients with the disease. Autoantibody panels are therefore likely to be specific but not sensitive. A well validated autoantibody panel has been studied in different screening cohorts as a lung cancer risk management approach[29–36]. In a clinical validation study including all lung cancer histologies and stages the panel performed well with 93% specificity, but only 40% sensitivity[37]. Similarly, a practice audit of 1699 patients (61 with lung cancer and 1/3 in stage I) found that the panel had robust specificity (91%) but low sensitivity (37%)[33]. Autoantibodies may find a place in clinical practice by improving the overall test accuracy of hybrid panels featuring diverse biomarkers[38].

3.1.2 Complement fragments

Lung cancer can activate the complement cascade via the classical complement pathway[39]. Concentrations of a downstream split product of this pathway, C4d, are increased in biological fluids from lung cancer patients. Plasma C4d levels have been linked to increased lung cancer risk in a cohort of 190 asymptomatic individuals, including 32 patients with screening detected cancer, enrolled in the iELCAP cohort (odds ratio = 4.38; 95% CI = 1.61 to 11.93). In that study, potential confounders such as emphysema and COPD did not appear to affect C4d plasma levels[39]. Unfortunately, its use as a marker for the selection of risk patients could not be validated using samples from the MILD CT-screening trial[40]. Nevertheless, results on its use for the management of indeterminate pulmonary nodules are more promising. Plasma samples from patients from two independent cohorts with malignant nodules, presented

significantly higher levels of C4d than those with benign nodules. In selected patients with intermediate-sized pulmonary nodules (8-30 mm), C4d plasma levels identified benign lung nodules with an 84% NPV and a 54% PPV. Once again the test enjoyed high specificity (89%), at the expense of low sensitivity (44%)[40].

3.1.3 miRNAs

Circulating microRNAs (miRNAs) reflecting tumor-host interactions, have emerged as potential biomarkers for cancer diagnosis and prognosis irrespective of tumor stage and mutational bourden36 (Verri et al, JTO 2017). The role of miRNA-based liquid biopsies has been assessed in the context of screening with LDCT in two large Italian retrospective validation studies37,38. Use of the miRNA signature classifier (MSC) and the miR-Test resulted in a five- and four-fold reduction in the LDCT-false positive rate with comparable specificity (81-75%) and sensitivity (87-78%). In post-surgical plasma samples, the MSC showed good performance in monitoring disease relapse (Sestini et al. Oncotarget 2015). The two tests are now undergoing prospective validation in three independent screening trials enrolling a total of 16,000 high-risk subjects.

3.1.4 Circulating tumor DNA

The value of circulating tumor DNA (ctDNA) as a biomarker in advanced tumor stages is well established[41,42]. However, its role in early lung cancer detection is still uncertain[43]. Abbosh et al. reported 48% sensitivity overall, setting a threshold of 2 single-nucleotide variants (SNVs) in 96 stage I-III NSCLC patients. Sensitivity ranged from 15% for stage I adenocarcinomas to 100% for stage II-III squamous cell carcinomas[44]. Current efforts to develop Next Generation Sequencing (NGS) technologies in order to study ctDNA in the context of early detection may improve sensitivity in this context (see below).

3.1.5 Serum DNA methylation

Tumor tissue is characterized by a global DNA hypomethylation status together with hypermethylation of specific CpG islands in the promoter region of tumor-suppressor genes[45]. Hypermethylation of at least one of four studied genes was detected twenty years ago in 15 of 22 (68%) NSCLC tumors but not in any paired normal lung tissue. In these primary tumors with methylation, 11 of 15

(73%) samples also had abnormal methylated DNA in the matched serum sample[46]. More recently, a 64-qPCR-assay was studied in 204 serum samples from 33 lung cancer, 68 fibrotic ILD and 42 COPD patients, as well as 61 healthy controls. The test had 88% sensitivity, 90% specificity when compared to controls, and 88% specificity when compared to COPD and ILD patients[47]. In 2017, Ooki et al. reported that a 6-gene panel correctly classified 43 stage IA and 42 control subjects with 72% sensitivity and 71% specificity[48]. Hulbert et al. recently described a three-gene model discriminating subjects with suspicious nodules on CT imaging (150 stages I-II NSCLC and 60 controls) with 98% sensitivity and 71% specificity in sputum and 93% sensitivity and 62% specificity in plasma samples[49].

3.1.6 Serum protein profiling

Many studies have identified measurable serum antigens in lung cancer patients. Panels of serum cancer antigens have been developed to improve diagnostic accuracy. One panel of 3 serum proteins (CEA, CA-125, CYFRA 21-1) and an AAb (NY-ESO-1) performed well in a high-risk cohort with 71% sensitivity and 88% specificity for lung cancer[38]. Clinical validation was performed in a separate high-risk cohort (based on age and smoking history) with lower sensitivity (49%) but higher specificity (96%). The incorporation of clinical variables improved accuracy[50]. A different panel of cancer antigens (CEA, CA15.3, SCC, CYFRA 21-1, NSE, ProGRP) increased the AUC of a clinical prediction model based on nodule size, age, and smoking status from 0.85 to 0.93 in a series of 3144 symptomatic individuals, including 1828 with lung cancer (52% stage IV) (Molina et al 2016). A two-protein biomarker ratio combined with a lung nodule clinical risk predictor had a sensitivity of 97% and a specificity of 44% for malignant disease in a series of 178 patients with suspicious lung nodules. This integrated classifier could have led to a 40% relative reduction in invasive testing for patients with benign nodules (10% ARR) while potentially delaying the management of 3% of malignant nodules[50,51].

3.1.7 RNA airway and nasal signature

Based on the "field of injury" paradigm[52], airway epithelial gene-expression has been developed as a diagnostic biomarker for lung cancer. Initial studies focused on bronchial airway epithelial cells obtained via endobronchial brushings of the mainstem bronchus[53,54]. A 23-gene biomarker measured in bronchial epithelial cells has been tested as an adjuvant diagnostic biomarker for patients undergoing bronchoscopy for suspected lung cancer[55]. This biomarker underwent clinical validation in 2 independent prospective cohorts, demonstrating a sensitivity of 88-89% and a specificity of 48%. The biomarker was particularly helpful in patients with an intermediate (10-60%) pre-test risk of lung cancer (91% NPV). Patients with inconclusive bronchoscopy results could have benefited from the biomarker's negative predictive value by avoiding further invasive testing, suggesting they could be followed safely with serial imaging studies[28]. Following analytic validation[56] and other clinical studies[57,58], the test received a favourable Medicare coverage decision in 2017.

The same "field of injury" concept may be useful in samples of nasal epithelial cells. This approach has obvious advantages as a minimally invasive diagnostic alternative for those not undergoing bronchoscopy as part of their clinical work up. A 30-gene nasal expression panel has been developed for diagnosing lung cancer among ever smokers with suspected disease, demonstrating improvement in AUC, sensitivity, and NPV when combined with clinical risk models[59].

3.2 Current trials in which biomarkers are considered or included.

Clinical validation study results have been published for a handful of biomarkers. Other biomarkers linger at various stages of development, while a few have entered formal clinical testing. The aforementioned panel of autoantibodies[29] is currently being assessed as part of a Scottish NHS funded randomized controlled screening study enrolling 12,000 subjects (the ECLS study; NCT01925625). A bronchial gene-expression classifier that could improve the diagnostic performance of bronchoscopy is being tested in a large registry. The combination of the plasma MSC with LDCT results informs screening intervals in 4119 at risk subjects in the bioMILD screening trial (NCT02247453)[60]. Plasma samples prospectively collected during the

COSMOS II screening trial have been profiled to set-up and validate the clinical utility of the miR-Test (Lococo F et al, Lung 2015). The DECAMP-1 and -2 prospective observational trials (NCT01785342 and NCT02504697) have been designed to examine a variety of existing biomarkers for lung cancer diagnosis as well as new biomarkers discovered specifically in this clinical setting. DECAMP-1 seeks to improve follow-up of patients with IPNs by determining whether analyzed biomarkers are able to distinguish incidentally detected malignant from benign pulmonary nodules in high-risk smokers, while DECAMP-2 will test biomarkers of risk in asymptomatic high-risk screened individuals.

An exciting amount of high-quality discovery and clinical validation work is ongoing. Some companies are in the process of planning true clinical utility studies for lung nodule management. The lack of an established trade-off regarding the consequences of true and false biomarker results is a challenge that every biomarker developer will face. It would behoove the clinical community to provide guidance regarding acceptable trade-offs both in the screening and the lung nodule management settings[23].

4. EMERGING BIOMARKERS, NEW TECHNOLOGIES TO FOLLOW AND FUTURE DIRECTIONS. THE POWER OF INTEGRATION

The aforementioned biomarkers have been the object of intense research and a number of them are being assessed in a risk management strategy to recommend screening or aiming to characterize IPNs. We will now discuss promising new technologies with potential, including integrated approaches to biomarker development in lung cancer screening.

4.1 Exhaled breath biomarkers

There is growing evidence to support the use of exhaled breath (EB), including exhaled breath condensate (EBC) for diagnostic purposes, including cancer detection. The EBC, which includes cells and DNA fragments, may even support detection of resistant clones of EGFR cells[61]. The volatile fragments of the EB are sensitive biomarkers of lung cancer. Volatile organic compounds (VOCs) can be captured and analysed by a wide range of technologies,

including GC-MS, nano-sensors, colorimetric sensors, and other methods.[62,63] An artificially intelligent nano-array sensor has been used in the diagnosis and classification of 17 different diseases from breath samples of 1404 subjects, with 86% accuracy. Some studies suggest that such an array may discriminate benign from malignant pulmonary nodules[64], or predict response to therapy and recurrence[65]. It may also distinguish histologic type[66] or predict molecular analysis results[67]. Interestingly, it apparently may also discriminate between different types of cancer (lung, breast, colorectal, and prostate)[68].

4.2 Sputum cell based image analysis

While sputum cytology has traditionally failed to yield either adequate or useful samples for lung cancer screening, the advent of "enhanced cytology", in which sophisticated image analysis algorithms are combined with artificial intelligence, may yet prove sputum useful in this context. A newly developed test can identify abnormal cells in sputum samples of screened patients[69]. This test may be used as a primary screening modality with a reported sensitivity of 90% when 800 or more bronchial cells are available for analysis, or integrated with LDCT in the context of a conventional CT based screening program for IPN characterization[70]. In the latter case, fewer cells may be needed, since the clinician can integrate clinical, molecular or conventional sputum cytology data together with imaging results for greater diagnostic accuracy.

4.3 Metabolomics

A range of different analytical platforms and methodologies have been applied to identify metabolic biomarkers of lung cancer[71]. Metabolomics provides a direct functional readout of the phenotypic changes associated with the development of lung tumors and their microenvironment. Metabolomics has several advantages when compared to other omics, including a reduced number of metabolites and a wide range of biological samples that can be tested. Changes in lung cancer metabolites include those involved in glycolysis, the citric acid cycle, amino acid metabolism and cell membrane synthesis[71,72]. Metabolomics can differentiate between histological subtypes or genetic backgrounds[73,74]. A panel of metabolites excreted in the urine, including creatine riboside (CR) and N-acetylneuraminic acid (NANA), have been

associated with lung cancer risk prior to clinically detectable disease[75,76]. Panels as well as individual markers in blood, sputum, or exhaled breath condensate have also been proposed to identify high-risk candidates for screening or to discriminate between benign and malignant IPNs[77–82]. Finally, other -omics, such as microbiomics, are providing us with novel diagnostic markers that merit a closer look[83] .

4.4 Genetic predisposition to lung cancer

The advent of GWAS analysis potentially could provide the lung cancer community with strong evidence of genetic susceptibility genes, which may be included in lung cancer risk prediction models[12,13,84]. Current evidence from a major review in 2017 of over 1000 candidate association studies, identified 22 variants in 21 genes, which had strong cumulative epidemiological evidence of significant associations with lung cancer risk[14].The OncoArray consortium research programme[19] has provided recent new insights and a new set of susceptibility genes[85], however, it still needs to be demonstrated that they make a significant contribution to risk prediction models used in lung cancer screening trials, over and above the patient's epidemiological and clinical information[86]. However, these susceptibility genes do provide an insight into the biological process and association with specific pathologies, which are relevant to lung cancer aetiology[87,88]. The question which has to be asked, is how should we utilize the state of art mathematical and statistical approaches which can incorporate very large numbers of SNPs within risk models, through artificial intelligence and supervised machine learning approaches. Clearly there is a wealth of information captured within the current GWAS data sets. We just need to find the next generation of tools to release it[89,90].

4.5 Integrating molecular biomarkers with radiomics and artificial intelligence

The current scientific field of radiomics, a term first used by Dutch researcher Philippe Lambin in 2012, is a newcomer in search of biomarkers among the seemingly limitless supply of data related to lung cancer imaging based phenotypes and tumor microenvironment[91,92]. The accumulation of detailed

imaging data in the current era of artificial intelligence has set the stage for much progress in this field. Deep-learning architectures, for example, can be useful in lung nodule characterization[93,94]. Current research in the field is centered on robust identification of the region of interest in time, direct spatio-temporal phenotypic characterization of tumor microenvironments, the integration of multiscale information at the local (nodule), regional (lobe), and organ levels, as well as the integration of imaging, clinical and -omics data in end to end learning architectures.

The combination of imaging based deep learning with molecular biomarkers may be very powerful in the characterization of IPNs. Radiomics can identify EGFR and KRAS mutated tumors[95,96]. Imaging signatures based of quantitative analysis of imaging data can also predict survival[97]. Some studies have shown that the integration of plasma biomarkers and radiological characteristics is a better predictor of lung cancer in patients with IPNs[98]. Prediction models integrating serum biomarkers with clinical characteristics and radiographic features of suspicious nodules correctly identified malignant nodules in several studies[27,99]. The integrated models outperform the use of serum biomarkers alone and overall represent a very promising approach for the future of early lung cancer detection, especially if artificial intelligence is incorporated[100–102].

4.6 Integration of multiple next generation sequencing (NGS) analysis in circulating DNA.

We have already alluded to the use of NGS of ctDNA as a promising strategy for early lung cancer detection. The biggest technical challenge is sensitivity. In an attempt to overcome this limitation, a recently reported test for pan-cancer early detection combined the NGS analysis of ctDNA in blood with a large panel of protein biomarkers in 1,005 stage I-III pan-cancer patients and 812 cancer-free controls[103]. While specificity was > 99%, sensitivity ranged from 33% for breast cancer to 98% for ovarian cancer. The sensitivity for lung cancer was 59% in 104 patients. Although promising, the study had some important limitations including the fact that most cancer patients were symptomatic, and control subjects had no comorbidities which could have acted as confounding variables.

The scientific community is also awaiting results of the Circulating Cell-free Genome Atlas Study (CCGA) for early cancer detection, enrolling 15,000 participants including cancer-free controls in the U.S. and Canada. Plasma samples collected at baseline and during 5 years of follow-up will be analyzed by whole-genome sequencing (WGS) for copy number variation (CNV), targeted DNA sequencing (507-gene panel), and whole genome methylome profiling. Preliminary results in an observational case-control setting include 95% specificity, high sensitivity for advanced lung cancer in 54 patients (85% for targeted sequencing, 91% for CNV WGS, and 93% for methylome profiling), and modest sensitivity for 63 patients with stage I-III lung cancer (48% for targeted NGS, 54% for CNV WGS, and 56% for methylome profiling)[104]. The generalizability of these findings to the screening setting is uncertain. A recent review by Aravanis, Lee and Klausner[105] addressed the challenges NGS faces in early cancer detection. The authors suggested that a successful pan-cancer screening NGS-based blood test would have to test up to 1000 genes, and the ctDNA limit of detection (sensitivity) would have to improve ten-fold from the current 0.1% to less than 0.01%. More importantly, a validation trial would have to enroll between 10,000 and 100,000 individuals. Despite these seemingly long odds, an observational trial (NCT02889978) investigating the discriminating power of the CCGA test is already under way.

5. FUTURE RESEARCH CHALLENGES

Table 3 summarizes the research challenges faced by biomarker development in the context of lung cancer screening. The interaction between genetics and environment is multidimensional and hard to control. Samples need to be carefully collected, processed using standard operating procedures, and annotated using clinical variables reliably collected from patients and/or electronic medical records[106]**.** Informed consent is essential to preserve confidentiality (CoC). Researchers with a CoC may only disclose identifiable, sensitive information if the subject consents, while anyone conducting research as a sub-awardee or receiving a copy of identifiable sensitive information must also comply with and understand disclosure restrictions. Even though samples

may be anonymized, genetic fingerprints may reveal a subject's identity rendering us vulnerable to the misuse of our most personal information. On the other hand, genetic privacy acts can hinder progress in this field[107].

Because so many biomarkers are approaching clinical validation, the field is in great need of standardized metrics of clinical utility. In the context of lung cancer screening, we can envision a study design (Figure 3) that would test the clinical utility of a biomarker-based risk assessment strategy. Because many patients with lung cancer do not meet NLST screening criteria, the study would test the value of a candidate biomarker as a predictor of risk independent of age and tobacco exposure and therefore justify annual screening with LDCT for 5 years (the duration of the trial). Conceivably, patients not meeting the U.S. Preventive Services Task Force (USPSTF) or other formally accepted screening criteria could be prospectively enrolled based on modeling outcomes. Indeed, the selection criteria could also include those used in other settings, such as the PanCAN and UKLS, with used risk-based prediction models[19,85]. The study would perform biomarker testing using a clinically validated biomarker(s) and assign patients to management strategies based on the results of combined testing. Those identified by the biomarker as having a lung cancer risk akin to those meeting USPSTF, PLCO2012, LLPv2, or other accepted criteria would be offered LDCT. Those identified as having a lower risk profile would be followed without LDCT. All subjects would sign an informed consent and undergo biomarker testing (or a series of tests). The primary outcome of this hypothetical trial would be risk prediction accuracy. Nodule management would follow current clinical guidelines. Biomarker test results would be shared with the patient and his/her provider, who would in turn decide in light of the results whether LDCT is warranted or not. We would recommend testing patient-reported outcomes based on expected risks and benefits of getting tested, the way the test results are communicated, anxiety related to test results, smoking habits, and willingness to undergo further testing based on biomarker results. We would also recommend determining the accuracy of the risk assessment before and after biomarker testing as well as outcome values. The best sequence (annual vs biannual) and combination of tests to offer should also be tested. Such a study would pave the way for a biomarker of risk

driven strategy for lung cancer screening. An alternative trial designed to validate the clinical utility of diagnostic biomarkers in the context of IPNs found incidentally could also be undertaken.

6. THE FUTURE OF MOLECULAR BIOMARKERS IN THE CONTEXT OF LUNG CANCER SCREENING

Despite the vast potential of existing candidates and methodologies, no single lung cancer molecular biomarker is currently being used in routine clinical practice. The clinical validation and utility steps are critical, but much more demanding, resource needy, and time consuming than the initial discovery and retrospective validation. That notwithstanding, the unmet clinical needs remain. Individual risk needs to be refined, and screening criteria modified in order to impact lung cancer related mortality. Orphan images of IPNs stand to improve our success differentiating benign from malignant with a robust biomarker at our disposal. There is also a clear unmet need for prognostic molecular and clinical markers for patients with screening detected early stage tumors. Although some believe that testing a new biomarker would be comparable to the gargantuan effort embodied by the NLST, we believe that less complex and more affordable validation is possible in the setting of established lung cancer screening programs.

There is plenty of room for improvement. We need to promote studies integrating promising candidate biomarkers, including molecular and image-based, and the use of artificial intelligence technologies to help in the selection of the most appropriate combinations. Head to head comparisons of biomarkers in specific clinical scenarios would also be welcome. Deep mining of the troves of data provided by ongoing screening efforts with new mathematical and computational models based on machine learning will surely help. This will require a systematic collection of patient samples in the context of screening. Finally, ways to prove cost-effectiveness of the new tests as well as to overcome the potential hurdles to get the approval by regulatory agencies need to be considered in the list of challenges that we face ahead in the development of molecular biomarkers in screening. Although the logistics and expense of such an effort may seem daunting at first, we believe the long-term outcome may prove highly efficient.

7. REFERENCES

1. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395-409. doi:10.1056/NEJMoa1102873.
2. Yousaf-Khan U, Van Der Aalst C, De Jong PA, et al. Final screening round of the NELSON lung cancer screening trial: The effect of a 2.5-year screening interval. *Thorax*. 2017;72(1):48-56.
3. Oudkerk M, Devaraj A, Vliegenthart R, et al. European position statement on lung cancer screening. *Lancet Oncol*. 2017;18(12):e754-e766.
4. Moyer VA. Screening for lung cancer: U.S. preventive services task force recommendation statement. *Ann Intern Med*. 2014;160(5):330-338.
5. Roberts H, Walker-Dilks C, Sivjee K, et al. Screening High-Risk Populations for Lung Cancer: Guideline Recommendations. *J Thorac Oncol*. 2013;8(10):1232-1237.
6. Mazzone PJ, Silvestri GA, Patel S, et al. Screening for Lung Cancer: CHEST Guideline and Expert Panel Report. *Chest*. 2018;153(4):954-985.
7. Sanchez-Salcedo P, Wilson DO, De-Torres JP, et al. Improving selection criteria for lung cancer screening: The potential role of emphysema. *Am J Respir Crit Care Med*. 2015;191(8):924-931.
8. Pinsky PF, Berg CD. Applying the National Lung Screening Trial eligibility criteria to the US population: What percent of the population and of incident lung cancers would be covered? *J Med Screen*. 2012;19(3):154-156.
9. Tammemägi MC, Katki HA, Hocking WG, et al. Selection Criteria for Lung-Cancer Screening. *N Engl J Med*. 2013;368(8):728-736.
10. Tanner NT, Dai L, Bade BC, Gebregziabher M, Silvestri GA. Assessing the generalizability of the national lung screening trial: Comparison of patients with stage 1 disease. *Am J Respir Crit Care Med*. 2017;196(5):602-608.
11. Wood DE. POINT: Should Lung Cancer Screening Be Expanded to Persons Who Don't Currently Meet Accepted Criteria Set Forth by the CHEST Guidelines on Lung Cancer Screening? Yes. *Chest*. 2018;153(6):1299-1302.
12. Tammemagi CM, Pinsky PF, Caporaso NE, et al. Lung cancer risk

prediction: Prostate, lung, colorectal and ovarian cancer screening trial models and validation. *J Natl Cancer Inst*. 2011;103(13):1058-68.

13. ten Haaf K, Jeon J, Tammemägi MC, et al. Risk prediction models for selection of lung cancer screening candidates: A retrospective validation study. *PLoS Med*. 2017;14(4).
14. Wang J, Liu Q, Yuan S, et al. Genetic predisposition to lung cancer: Comprehensive literature integration, meta-analysis, and multiple evidence assessment of candidate-gene association studies. *Sci Rep*. 2017;7(1).
15. De-Torres JP, Wilson DO, Sanchez-Salcedo P, et al. Lung cancer in patients with chronic obstructive pulmonary disease: Development and validation of the COPD lung cancer screening score. *Am J Respir Crit Care Med*. 2015;191(3):285-291.
16. Van Riel SJ, Ciompi F, Winkler Wille MM, et al. Malignancy risk estimation of pulmonary nodules in screening CTs: Comparison between a computer model and human observers. *PLoS One*. 2017;12(11).
17. Tammemagi MC, Schmidt H, Martel S, et al. Participant selection for lung cancer screening by risk modelling (the Pan-Canadian Early Detection of Lung Cancer [PanCan] study): A single-arm, prospective study. *Lancet Oncol*. 2017;18(11):1523-1531.
18. Peled N, Ilouze M. Screening for lung cancer: What comes next? *J Clin Oncol*. 2015;33(33):3847-3848.
19. Amos CI, Dennis J, Wang Z, et al. The oncoarray consortium: A network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomarkers Prev*. 2017;26(1):126-135.
20. Feldman R, Kim ES. Prognostic and predictive biomarkers post curative intent therapy. *Ann Transl Med*. 2017;5(18):374.
21. Martínez-Terroba E, Behrens C, de Miguel FJ, et al. A novel protein-based prognostic signature improves risk stratification to guide clinical management in early lung adenocarcinoma patients. *J Pathol*. 2018;245(4):421-432.
22. Atkinson AJ, Colburn WA, DeGruttola VG, et al. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin Pharmacol Ther*. 2001;69(3):89-95.

23. Mazzone PJ, Sears CR, Arenberg DA, et al. Evaluating molecular biomarkers for the early detection of lung cancer: When is a biomarker ready for clinical use? An official American Thoracic Society Policy Statement. *Am J Respir Crit Care Med*. 2017;196(7).
24. Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst*. 2001;93(14):1054-1061.
25. Atwater T, Cook C, Massion P. The Pursuit of Noninvasive Diagnosis of Lung Cancer. *Semin Respir Crit Care Med*. 2016;37(5):670-680.
26. Rodriguez H, Težak Ž, Mesri M, et al. Analytical validation of protein-based multiplex assays: A workshop report by the NCI-FDA interagency oncology task force on molecular diagnostics. *Clin Chem*. 2010;56(2):237-243.
27. Pecot C V, Li M, Zhang XJ, et al. Added Value of a Serum Proteomic Signature in the Diagnostic Evaluation of Lung Nodules. *Cancer Epidemiol Biomarkers Prev*. 2012;21(5):786-792.
28. Silvestri GA, Vachani A, Whitney D, et al. A Bronchial Genomic Classifier for the Diagnostic Evaluation of Lung Cancer. *N Engl J Med*. 2015;373(3):243-251. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=25981554.
29. Lam S, Boyle P, Healey GF, et al. EarlyCDT-Lung: An immunobiomarker test as an aid to early detection of lung cancer. *Cancer Prev Res*. 2011;4(7):1126-1134.
30. Macdonald IK, Murray A, Healey GF, et al. Application of a High Throughput Method of Biomarker Discovery to Improvement of the EarlyCDT®-Lung Test. *PLoS One*. 2012;7(12).
31. Chapman CJ, Healey GF, Murray A, et al. EarlyCDT®-Lung test: Improved clinical utility through additional autoantibody assays. *Tumor Biol*. 2012;33(5):1319-1326.
32. Healey GF, Lam S, Boyle P, Hamilton-fairley G, Peek LJ, Robertson JFR. Signal stratification of autoantibody levels in serum samples and its application to the early detection of lung cancer. *J Thorac Dis*. 2013;5(5):618-625.
33. Jett JR, Peek LJ, Fredericks L, Jewell W, Pingleton WW, Robertson JFR.

Audit of the autoantibody test, EarlyCDT®-Lung, in 1600 patients: An evaluation of its performance in routine clinical practice. *Lung Cancer*. 2014;83(1):51-55.

34. Massion PP, Healey GF, Peek LJ, et al. Autoantibody Signature Enhances the Positive Predictive Power of Computed Tomography and Nodule-Based Risk Models for Detection of Lung Cancer. *J Thorac Oncol*. 2017;12(3):578-584.
35. Sullivan FM, Farmer E, Mair FS, et al. Detection in blood of autoantibodies to tumour antigens as a case-finding method in lung cancer using the EarlyCDT®-Lung Test (ECLS): Study protocol for a randomized controlled trial. *BMC Cancer*. 2017;17(1).
36. Edelsberg J, Weycker D, Atwood M, Hamilton-fairley G, Jett R. Cost-effectiveness of an autoantibody test (Early CDT-Lung) as an aid to early diagnosis of lung cancer in patients with incidentally detected pulmonary nodules. *Oncoimmunology*. 2018;442:1-14.
37. Boyle P, Chapman CJ, Holdenrieder S, et al. Clinical validation of an autoantibody test for lung cancer. *Ann Oncol*. 2011;22(2):383-389.
38. Doseeva V, Colpitts T, Gao G, Woodcock J, Knezevic V. Performance of a multiplexed dual analyte immunoassay for the early detection of non-small cell lung cancer. *J Transl Med*. 2015;13(55).
39. Ajona D, Pajares MJ, Corrales L, et al. Investigation of complement activation product C4d as a diagnostic and prognostic biomarker for lung cancer. *J Natl Cancer Inst*. 2013;105(18).
40. Ajona D, Okrój M, Pajares MJ, et al. Complement C4d-specific antibodies for the diagnosis of lung cancer. *Oncotarget*. 2018;9(5).
41. Jenkins S, Yang JCH, Ramalingam SS, et al. Plasma ctDNA Analysis for Detection of the EGFR T790M Mutation in Patients with Advanced Non–Small Cell Lung Cancer. *J Thorac Oncol*. 2017;12(7):1061-1070.
42. Giroux Leprieur E, Herbretau G, Dumenil C, et al. Circulating tumor DNA evaluated by Next-Generation Sequencing is predictive of tumor response and prolonged clinical benefit with nivolumab in advanced non-small cell lung cancer. *Oncoimmunology*. 2018;7(5).
43. Merker JD, Oxnard GR, Compton C, et al. Circulating Tumor DNA Analysis in Patients With Cancer: American Society of Clinical Oncology

and College of American Pathologists Joint Review. *J Clin Oncol*. 2018;36(16):1631-1641.

44. Abbosh C, Birkbak NJ, Wilson GA, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*. 2017;545(7655):446-451.
45. Ehrlich M. DNA hypomethylation in cancer cells. *Epigenomics*. 2009;1(2):239-259.
46. Esteller M, Sanchez-Cespedes M, Resell R, Sidransky D, Baylin SB, Herman JG. Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. *Cancer Res*. 1999;59(1):67-70.
47. Wielscher M, Vierlinger K, Kegler U, Ziesche R, Gsur A, Weinhäusel A. Diagnostic Performance of Plasma DNA Methylation Profiles in Lung Cancer, Pulmonary Fibrosis and COPD. *EBioMedicine*. 2015;2(8):929-936.
48. Ooki A, Maleki Z, Tsay JCJ, et al. A panel of novel detection and prognostic methylated DNA markers in primary non–small cell lung cancer and serum DNA. *Clin Cancer Res*. 2017;23(22):7141-7152.
49. Hulbert A, Jusue Torres I, Stark A, et al. Early detection of lung cancer using DNA promoter hypermethylation in plasma and sputum. *Clin Cancer Res*. 2016;23(8):1998-2005.
50. Mazzone PJ, Wang X, Han X, et al. Evaluation of a Serum Lung Cancer Biomarker Panel. *Biomark Insights*. 2018;13.
51. Silvestri GA, Tanner NT, Kearney P, et al. Assessment of Plasma Proteomics Biomarker's Ability to Distinguish Benign From Malignant Lung Nodules: Results of the PANOPTIC (PulmonAry NOdule Plasma proTeomIc Classifier) trial. *Chest*. 2018.
52. Billatos E, Vick JL, Lenburg ME, Spira AE. The Airway Transcriptome as a Biomarker for Early Lung Cancer Detection. *Clin Cancer Res*. 2018;24(13):2984-2992.
53. 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(3):361-366.
54. 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(22):8629-8635.

55. Whitney DH, Elashoff MR, Porta-Smith K, et al. Derivation of a bronchial genomic classifier for lung cancer in a prospective study of patients undergoing diagnostic bronchoscopy. *BMC Med Genomics*. 2015;8(18).
56. Hu Z, Whitney D, Anderson JR, et al. Analytical performance of a bronchial genomic classifier. *BMC Cancer*. 2016;16(161).
57. Vachani A, Whitney DH, Parsons EC, et al. Clinical Utility of a Bronchial Genomic Classifier in Patients With Suspected Lung Cancer. *Chest*. 2016;150(1):210-218.
58. DK. Hogarth, TL. Dotson HL et al. PulmonaryThe Percepta® Registry: A Prospective Registry to Evaluate Percepta Bronchial Genomic Classifier Patient Data. In: *CHEST Annual Meeting*. ; 2016.
59. JF. Perez-Rogers, J. Gerrein CA et al. Shared Gene Expression Alterations in Nasal and Bronchial Epithelium for Lung Cancer Detection. *J Natl Cancer Inst*. 2017;109(7).
60. Boeri M, Sestini S, Fortunato O, et al. Recent advances of microRNA-based molecular diagnostics to reduce false-positive lung cancer imaging. *Expert Rev Mol Diagn*. 2015;15(6):801-813.
61. Smyth RJ1, Toomey SM2, Sartori A3, Hanrahan EO4, Cuffe SD5, Breathnach OS6, Morgan RK6 HB. Brief report on the detection of the EGFR-T790M mutation in exhaled breath condensate from lung cancer patients. *J Thorac Oncol*. 2018;13(8):1213-1216.
62. Mazzone PJ, Wang XF, Lim S, et al. Progress in the development of volatile exhaled breath signatures of lung cancer. *Ann Am Thorac Soc*. 2015;12(5):752-757.
63. Nakhleh MK, Amal H, Jeries R, et al. Diagnosis and Classification of 17 Diseases from 1404 Subjects via Pattern Analysis of Exhaled Molecules. *ACS Nano*. 2017;11(1):112-125.
64. Peled N, Hakim M, Bunn PA, et al. Non-invasive breath analysis of pulmonary nodules. *J Thorac Oncol*. 2012;7(10):1528-1533.
65. Nardi-Agmon I, Abud-Hawa M, Liran O, et al. Exhaled breath analysis for monitoring response to treatment in advanced lung cancer. *J Thorac Oncol*. 2016;11(6):827-837.

66. Hakim M, Broza YY, Barash O, et al. Volatile organic compounds of lung cancer and possible biochemical pathways. *Chem Rev*. 2012;112(11):5949-5966.
67. Peled N, Barash O, Tisch U, et al. Volatile fingerprints of cancer specific genetic mutations. *Nanomedicine Nanotechnology, Biol Med*. 2013;9(6):758-766.
68. Peng G, Hakim M, Broza YY, et al. Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. *Br J Cancer*. 2010;103(4):542-551.
69. Meyer MG, Hayenga JW, Neumann T, et al. The Cell-CT 3-dimensional cell imaging technology platform enables the detection of lung cancer using the noninvasive LuCED sputum test. *Cancer Cytopathol*. 2015;123(9):512-523.
70. Nelson A, Meyer M, Katdare R, et al. Early detection of lung cancer based on three-dimensional, morphometric analysis of cells from sputum. In: *Journal of Clinical Oncology*. Vol 32. ; 2014:7547.
71. Yu L, Li K, Zhang X. Next-generation metabolomics in lung cancer diagnosis, treatment and precision medicine: mini review. *Oncotarget*. 2017;8(70):115774-115786.
72. Puchades-Carrasco L, Jantus-Lewintre E, Pérez-Rambla C, et al. Serum metabolomic profiling facilitates the non-invasive identification of metabolic biomarkers associated with the onset and progression of non-small cell lung cancer. *Oncotarget*. 2016;7(11):12904-12916.
73. Rezola A, Pey J, Rubio Á, Planes FJ. In-Silico prediction of key metabolic differences between two non-small cell lung cancer subtypes. *PLoS One*. 2014;9(8).
74. Caiola E, Brunelli L, Marabese M, Broggini M, Lupi M, Pastorelli R. Different metabolic responses to PI3K inhibition in NSCLC cells harboring wild-type and G12C mutant KRAS. *Oncotarget*. 2016;7(32):51462-51472.
75. Mathé EA, Patterson AD, Haznadar M, et al. Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. *Cancer Res*. 2014;74(12):3259-3270.
76. Haznadar M, Cai Q, Krausz KW, et al. Urinary Metabolite Risk Biomarkers of Lung Cancer: A Prospective Cohort Study. *Cancer*

Epidemiol Biomarkers Prev. 2016;25(6):978-986.
77. Roś-Mazurczyk M, Wojakowska A, Marczak Ł, et al. Panel of serum metabolites discriminates cancer patients and healthy participants of lung cancer screening - A pilot study. *Acta Biochim Pol*. 2017;64(3):513-518.
78. Wikoff WR, Hanash S, DeFelice B, et al. Diacetylspermine is a novel prediagnostic serum biomarker for non-small-cell lung cancer and has additive performance with pro-surfactant protein B. *J Clin Oncol*. 2015;33(33):3880-3886.
79. Wen CP, Zhang F, Liang D, et al. The ability of bilirubin in identifying smokers with higher risk of lung cancer: A large cohort study in conjunction with global metabolomic profiling. *Clin Cancer Res*. 2015;21(1):193-200.
80. Fahrmann JF, Grapov D, Defelice BC, et al. Serum phosphatidylethanolamine levels distinguish benign from malignant solitary pulmonary nodules and represent a potential diagnostic biomarker for lung cancer. *Cancer Biomarkers*. 2016;16(4):609-617.
81. Cameron SJS, Lewis KE, Beckmann M, et al. The metabolomic detection of lung cancer biomarkers in sputum. *Lung Cancer*. 2016;94:88-95.
82. Peralbo-Molina A, Calderón-Santiago M, Priego-Capote F, Jurado-Gámez B, Luque De Castro MD. Identification of metabolomics panels for potential lung cancer screening by analysis of exhaled breath condensate. *J Breath Res*. 2016.
83. Lee SH, Sung JY, Yong D, et al. Characterization of microbiome in bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions. *Lung Cancer*. 2016;102:89-95.
84. Cassidy A, Myles JP, Van Tongeren M, et al. The LLP risk model: An individual risk prediction model for lung cancer. *Br J Cancer*. 2008;98(2):270-276.
85. McKay JD, Hung RJ, Han Y, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet*. 2017;49(7):1126-1132.
86. Field JK, Duffy SW, Baldwin DR, et al. The UK lung cancer screening trial: A pilot randomised controlled trial of low-dose computed tomography

screening for the early detection of lung cancer. *Health Technol Assess (Rockv)*. 2016;20(40):1-146.

87. Ji X, Bossé Y, Landi MT, et al. Identification of susceptibility pathways for the role of chromosome 15q25.1 in modifying lung cancer risk. *Nat Commun*. 2018;9(1).
88. Fusco JP, Pita G, Pajares MJ, et al. Genomic characterization of individuals presenting extreme phenotypes of high and low risk to develop tobacco-induced lung cancer. *Cancer Med*. 2018.
89. Mieth B, Kloft M, Rodríguez JA, et al. Combining multiple hypothesis testing with machine learning increases the statistical power of genome-wide association studies. *Sci Rep*. 2016;6.
90. Schrider DR, Kern AD. Supervised Machine Learning for Population Genetics: A New Paradigm. *Trends Genet*. 2018;34(4):301-312.
91. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441-446.
92. Chen B, Zhang R, Gan Y, Yang L, Li W. Development and clinical application of radiomics in lung cancer. *Radiat Oncol*. 2017;12(1).
93. Shen W, Zhou M, Yang F, Yang C, Tian J. Multi-scale convolutional neural networks for lung nodule classification. *Inf Process Med Imaging*. 2015;24:588-599.
94. Arindra A, Setio A, Ciompi F, et al. Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. *IEEE Trans Med Imaging*. 2016;35(5):1160-1169.
95. Liu Y, Kim J, Balagurunathan Y, et al. Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas. *Clin Lung Cancer*. 2016;17(5):441-448.
96. Rios Velazquez E, Parmar C, Liu Y, et al. Somatic mutations drive distinct imaging phenotypes in lung cancer. *Cancer Res*. 2017;77(14):3922-3930.
97. Lee J, Li B, Cui Y, et al. A quantitative CT imaging signature predicts survival and complements established prognosticators in stage I non-small cell lung cancer. *Int J Radiat Oncol*. 2018.
98. Lin Y, Leng Q, Jiang Z, et al. A classifier integrating plasma biomarkers and radiological characteristics for distinguishing malignant from benign

pulmonary nodules. *Int J Cancer*. 2017;141(6):1240-1248.
99. Ma J, Guarnera MA, Zhou W, Fang H, Jiang F. A Prediction Model Based on Biomarkers and Clinical Characteristics for Detection of Lung Cancer in Pulmonary Nodules. *Transl Oncol*. 2017;10(1):40-45.
100. Jiang R, Dong X, Zhu W, et al. Combining PET/CT with serum tumor markers to improve the evaluation of histological type of suspicious lung cancers. *PLoS One*. 2017;12(9).
101. Jiang F, Todd NW, Qiu Q, Liu Z, Katz RL, Stass SA. Combined genetic analysis of sputum and computed tomography for noninvasive diagnosis of non-small-cell lung cancer. *Lung Cancer*. 2009;66(1):58-63.
102. Grossmann P, Stringfield O, El-Hachem N, et al. Defining the biological basis of radiomic phenotypes in lung cancer. *Elife*. 2017.
103. Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science (80-)*. 2018;359(6378):926-930.
104. GR. Oxnard T. Maddala E. Hubbell et al. Genome-wide sequencing for early stage lung cancer detection from plasma cell-free DNA (cfDNA): The Circulating Cancer Genome Atlas (CCGA) study. In: *2018 ASCO Annual Meeting*. ; 2018.
105. Aravanis AM, Lee M, Klausner RD. Next-Generation Sequencing of Circulating Tumor DNA for Early Cancer Detection. *Cell*. 2017;168(4):571-574.
106. Tuck MK, Chan DW, Chia D, et al. Standard operating procedures for serum and plasma collection: Early detection research network consensus statement standard operating procedure integration working group. *J Proteome Res*. 2009;8(1):113-117.
107. Minari J, Brothers KB, Morrison M. Tensions in ethics and policy created by National Precision Medicine Programs. *Hum Genomics*. 2018;12(1).
108. Molina R, Marrades RM, Augé JM, et al. Assessment of a combined panel of six serum tumor markers for lung cancer. *Am J Respir Crit Care Med*. 2016;193(4):427-437.
109. Guida F, Sun N, Bantis LE, et al. Assessment of Lung Cancer Risk on the Basis of a Biomarker Panel of Circulating Proteins. *JAMA Oncol*. 2018.
110. 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*. 2011;108(9):3713-3718.

111. 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*. 2014;32(8):768-773.
112. A. Vachani, M. Atalay R, Al B et al. A Multi-Gene Classifier for the Diagnosis of Benign Versus Malignant Pulmonary Nodules. In: *American Thoracic Society International Conference*. ; 2017.
113. Yanaihara N, Caplen N, Bowman E, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*. 2006;9(3).
114. A. Schlegel, O. Hasinger SE et al. Biomarkers for Lung Cancer. In: *AACR Annual Meeting*. ; 2016.
115. Weiss G, Schlegel A, Kottwitz D, König T, Tetzner R. Validation of the SHOX2/PTGER4 DNA Methylation Marker Panel for Plasma-Based Discrimination between Patients with Malignant and Nonmalignant Lung Disease. *J Thorac Oncol*. 2017;12(1):77-84.
116. Newman AM, Bratman S V., To J, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med*. 2014;20(5):548-554.
117. Guo N, Lou F, Ma Y, et al. Circulating tumor DNA detection in lung cancer patients before and after surgery. *Sci Rep*. 2016;6(33519).
118. Phallen J, Sausen M, Adleff V, et al. Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med*. 2017;9(403).
119. Showe MK, Vachani A, Kossenkov A V., 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(24):9202-9210.
120. 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(3):694-701.
121. Nichols JAA, Grob P, de Lusignan S, Kite W, Williams P. Genetic test to stop smoking (GeTSS) trial protocol: Randomised controlled trial of a genetic test (Respiragene) and Auckland formula to assess lung cancer

risk. *BMC Pulm Med*. 2014;14(77).

122. Zhang C, Yu W, Wang L, et al. DNA methylation analysis of the SHOX2 and RASSF1A panel in bronchoalveolar lavage fluid for lung cancer diagnosis. *J Cancer*. 2017;8(17):3585-3591.
123. Xing L, Su J, Guarnera MA, et al. Sputum microRNA biomarkers for identifying lung cancer in indeterminate solitary pulmonary nodules. *Clin Cancer Res*. 2015;21(2):484-489.
124. Subramanian H, Viswanathan P, Cherkezyan L, et al. Procedures for risk-stratification of lung cancer using buccal nanocytology. *Clin Cancer Cancer Res Acta Cytol Clin Cancer Res J Thorac Oncol J Cancer Endosc Ann Intern Med Ann Intern Med Biomed Opt EXPRESS J*. 2016;7(9):3795-3810. doi:10.1364/BOE.7.003795.
125. Patriquin L, Merrick DT, Hill D, et al. Early Detection of Lung Cancer with Meso Tetra (4-Carboxyphenyl) Porphyrin-Labeled Sputum. *J Thorac Oncol*. 2015;10(9):1311-1318.

TABLES AND FIGURES

Table 1: Principles to optimize the research on lung cancer biomarker development

- **Principle 1:** Selection of the molecular approach guided by the intended use.
- **Principle 2:** Look for stable analytes, minimally dependent on storage time, temperature, pH, enzymatic or oxidative stress**.**
- **Principle 3:** The analyte should be measured with accuracy, precision and robustness. Thoroughly test for reproducibility across sources of variability, laboratories, conditions, etc.
- **Principle 4:** Test the biomarker in multiple cohorts in the clinical context it will be considered for use (screening, nodule management). Case control studies are discouraged, while prospective cohort studies and eventually observational registries are favored though less convenient.
- **Principle 5:** Tests in larger cohorts to demonstrate superiority over standard of care, reduction of cost and false positive and negative rates reduction

Table 2: Candidate biomarkers for lung cancer early detection and phase of development

Candidates	Biomarker	Target	Phase 1 Discovery, prediction	Phase 2 Assay validation	Phase 3 Retrolongitudinal	Phase 4 Clinical validation*	Phase 5 Clinical utility	References	Tr
SERUM/PLASMA									
Specific proteins/autoantibodies	Three proteins (CEA, CA-125, and CYFRA 21–1) and 1 AAb (NY-ESO-1)	RMS						[38]	
	Two proteins (LG3BP and C163A) and clinical features	DIPN						[51]	NCT01
	Seven AAbs (p53, NY-ESO-1, CAGE, GBU4-5, SOX2, HuD, and MAGE A4)	RMS						[31,33]	NCT01
		DIPN						[36]	
	Six proteins (CEA, CA-125, CEA 15-3, SCC, CYFRA 21–1, NSE, and proGRP)	DIPN						[109,110]	
	Complement fragment C4d	RMS						[39]	
		DIPN						[40]	
MiRNA	Ratios among 24 miRNAs	RMS						[110,111]	NCT02
		DIPN							
	Signature of 13 microRNA + 6 for normalization	RMS						[112]	COSMO trial
		DIPN							
	Signature of 2 microRNA	DIPN						[113]	
DNA methylation	SOX2 and PTGER4 methylation	RMS						[114,115]	
		DIPN							
Circulating tumor nucleic acids	Circulating tumor DNA; NGS technology	RMS						[104]	NCT02
	Circulating tumor DNA; NGS technology	DIPN						[116]	
	Circulating tumor DNA; Ion Torrent DNA Sequencing technology	DIPN						[117]	
	Circulating tumor DNA; TEC-Seq technology	RMS						[118]	
	Signature of 29 genes (RNA)	DIPN						[119]	
	ctDNA mutation and proteins (CA-125, CEA, CA19-9, PRL, HGF, OPN, MPO, and TIMP-1)	DIPN						[103]	
TUMOR/AIRWAY EPITHELIUM									
Chromosome aberrations	Chromosome regions copy number or fusions (FISH)	DIPN						[120]	
mRNA gene expression classifier	Twenty three gene classifier	DIPN							NCT01 NCT00
SNPs	12 SNPs for COPD and clinical features	RMS						[100,121]	
SPUTUM, BREATH AND URINE									
DNA methylation	SHOX2 and RASSF1A methylation	RMS						[122]	
MiRNA	Signature of 3 microRNA	DIPN						[123]	
Exhaled breath	VOC- Nanoparticle Biometric Tagging (NBT)	DIPN							
	VOC- Field Asymmetric Ion Mobility Spectrometry (FAIMS)								NCT02
Tumor cells	>700 morphological features (by Cell CT)	RMS							

		DIPN						
	Buccal nanocytology	RMS					[124]	
	Porphyrin differential uptake by tumor cells	RMS					[125]	
Unrine markers	Metabolites	RMS					[76]	

RMS: risk management in screening context; DIPN: diagnosis of indeterminate pulmonary nodules; *DECAMP-1 and DECAMP-2 trials (NCT01785342 and NCT02504697) are currently recruiting patients in order to test some of these biomarkersAAB Autoantibody.

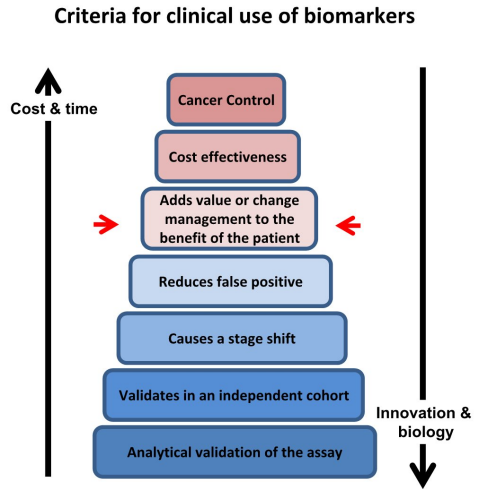

Figure 1. Metric for evaluation of clinical use of candidate biomarkers. Proposed metric for success is suggested for the level at or above the red arrows. From Atwater et al. Semin Respir Crit Care Med 2016;37:670–680.

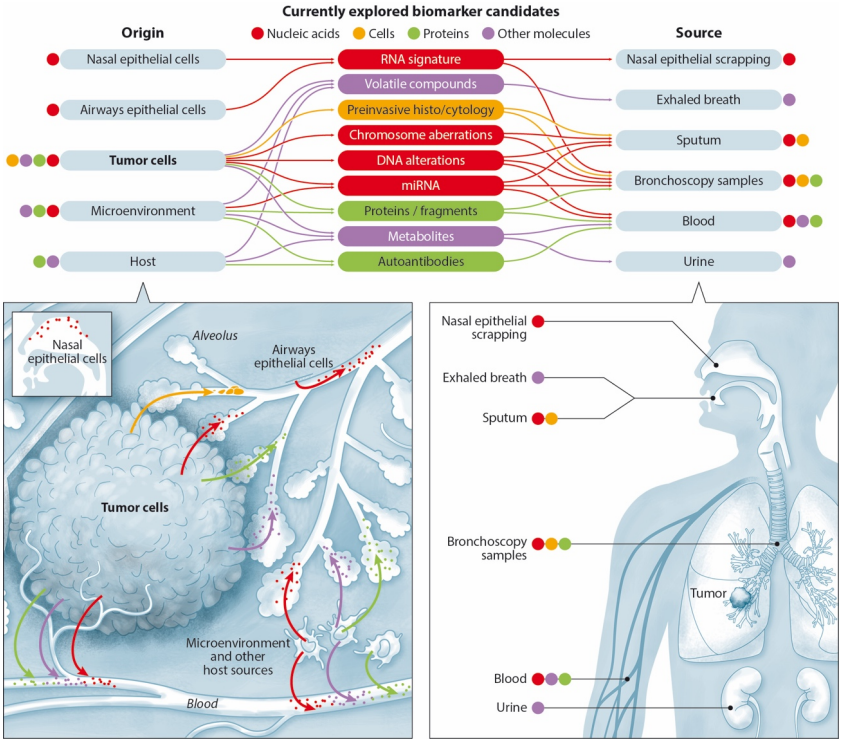

Figure 2 Sources of biomarkers in lung cancer screening

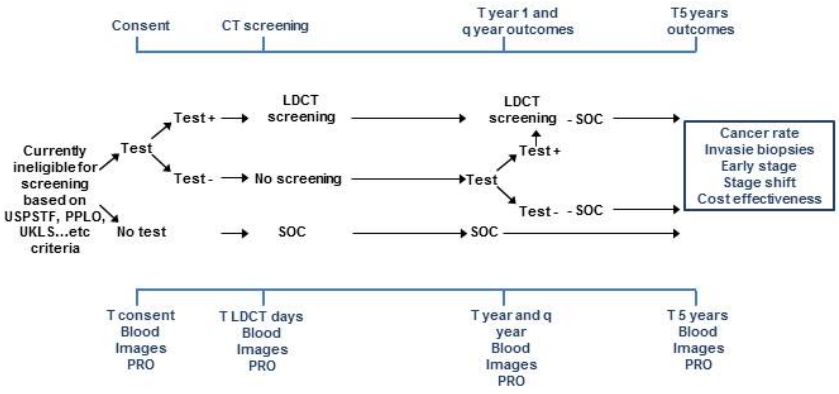

Fig 3 Potential design for testing the efficacy of a biomarker based test to optimize a LDCT lung cancer screening program

Table 3: Challenges faced by the research on lung cancer biomarker development

-
- **Challenge 1:** Need for deeper knowledge of lung carcinogenesis, tumor molecular and cellular landscape, gene-environment relationship, etc.
- **Challenge 2:** Need for careful sample collection; processing using standard operating procedures and properly annotated clinical data in the intended use type of patient (screening cohorts).
- **Challenge 3:** Need to obtain the samples from individuals following informed consent and comply with all rules, regulations and policies regarding human subjects' research.
- **Challenge 4:** Need to establish robust consensus criteria for the selection of the single or integrated combined biomarkers to be tested
- **Challenge 5:** Need to design and approve new mechanisms to show clinical utility of care, reduction of cost, false positive/negative rates reduction and aceptable ratios of true and false results.
- **Challenge 6:** Need to further convince stakeholders and research promoters and funders of the relevance of developing single and integrated biomarkers to optimize the efficacy of current lung cancer screening protocols.
- **Challenge 7**: Need to analyze, determine the causes and try to overcome potentially unnecessary hurdles to approval even after utility testing is complete.

Commented [MMP1]: There may be hurdles to approval even after utility testing is complete.

Commented [LM2]: Please look at this new challenge to answer Peter Mazzones previous comment.