



Developing technologies and areas of interest in lung cancer screening adjuncts

Ju Ae Park¹, Duy Pham², Sriya Yalamanchili¹, Shaina Twardus², Kei Suzuki¹

¹Department of Surgery, Inova Fairfax Medical Campus, Falls Church, VA, USA; ²University of Virginia School of Medicine, Charlottesville, VA, USA

Contributions: (I) Conception and design: K Suzuki, JA Park; (II) Administrative support: K Suzuki; (III) Provision of study materials or patients: K Suzuki, JA Park; (IV) Collection and assembly of data: JA Park, D Pham, S Yalamanchili, S Twardus; (V) Data analysis and interpretation: JA Park, D Pham, S Yalamanchili, S Twardus; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Kei Suzuki, MD, MS. Department of Surgery, Inova Fairfax Medical Campus, 8081 Innovation Park Drive, Building B, 3rd Floor Clinical Administration, Fairfax, VA 22031, USA. Email: kei.suzuki@inova.org.

Abstract: Lung cancer remains the leading cause of cancer mortality. Screening guidelines have been implemented in the past decade to aid in earlier detection of at-risk groups. Nevertheless, computed tomography (CT) scans, the principal screening modality in use today, are still low yield, with 3.6% of lung cancer confirmed amongst 39.1% of lesions detected over a 3-year period. They also carry relatively high false positive rates, between 9% and 27%, which can bear unnecessary financial and emotional costs to patients. As such, research efforts have been dedicated to the development of lung cancer screening adjuncts to improve detection reliability. We herein review several emerging technologies in this specific arena and their efficacy. These include plasma markers (microDNA, DNA methylation, and tumor-associated antibodies), breath/sputum biomarkers [volatile organic compounds (VOCs) and exhaled breath condensate (EBC)], proteomics, metabolomics, and machine learning, such as radiomics technology. We find that, across the board, they offer promising results in terms of non-invasive diagnostics, genetic sequencing for higher-risk individuals, and accessibility for a diverse cohort of patients. While these screening adjuncts are unlikely to completely replace the current standard of care at the moment, continued research into these technologies is crucial to improve and personalize the identification, treatment, and outcome of lung cancer patients in the near future.

Keywords: Lung cancer screening; developing technology; adjunctive studies; biomarkers; machine learning

Submitted Aug 24, 2023. Accepted for publication Dec 22, 2023. Published online Feb 27, 2024.

doi: 10.21037/jtd-23-1326

View this article at: <https://dx.doi.org/10.21037/jtd-23-1326>

Introduction

Lung cancer is the number one cause of cancer mortality worldwide. Screening protocols for cancer such as breast and colon have been well established and pervasive within the medical culture since the 1960s. However, lung cancer has not had an established screening method until recently. Studies of chest X-ray (CXR) as a possible screening method began in the 1960s and the 1970s. These first investigations included randomized control trials to compare CXR alone *vs.* CXR with sputum cytology as a potential screening mode.

The randomized trials did not demonstrate a benefit of sputum cytology compared to CXR alone. However, even CXR alone was not deemed a suitable screening method as meta-analyses of CXR screening also did not find any advantage of CXR to all-cause mortality from lung cancer (1).

With the advent of computed tomography (CT) in 1972 and subsequent further development of the technology, consideration for lung cancer screening with CT imaging was established. Conventional CT scan was not ideal due to the high radiation dosage per imaging. However, a low-dose CT (LDCT) could detect nodules of 0.5–1 cm

in size and was comparable in sensitivity to conventional CT. After early studies found CT superior to CXR, the largest randomized control trial, The National Lung Screening Trial (NLST), determined that CT scans were able to identify early-stage lung cancer at a higher rate and demonstrated a 20% reduction in mortality in high-risk patients (1). The updated standard for lung cancer screening is an LDCT scan starting at the age of 50 years for any current or former smokers (those who quit within 15 years) with a 20-pack-year smoking history. However, there remains a false positive rate of 9–27% with LDCT, and the current screening standard may not be adequate to capture patients at higher risk in certain populations, such as minorities and women (2). Although the racial disparity gap for African American patients has decreased due to the revised U.S. Preventive Service Task Force (USPSTF) guidelines, a cross-sectional survey found that Hispanics and African Americans were less likely to be eligible for screening compared to Caucasians (3). Similar gaps have been seen in women even though some studies have indicated an increased susceptibility to lung carcinogens in women than men.

In order to improve current screening methods, there have been multiple studies of adjunct tests that could be applied to screening (*Figure 1*). Compared to general factors such as age and smoking, these novel methods look to personalize the risk of lung cancer development. We herein explore current technologies in progress that could improve the future of lung cancer screening.

Plasma markers

Plasma remains an active area of research due to the ease of collection and feasibility of studies with plasma markers for cancer. Several subject areas within plasma markers are currently developing, with microRNA (miRNA), cell-free DNA, and auto-antibodies being highly pursued topics in lung cancer.

MiRNA

MiRNA are non-protein coding sequences that regulate gene expression. They primarily promote gene silencing by binding to messenger RNA (mRNA) molecules and are released by cells in a stable form. MiRNAs have been found to be involved in the pathogenesis of human cancers by negative regulation of tumor suppressor genes and effects on oncogenes, as studies have noted decreased expression

of miRNAs leading to an abundance of expression in tumor oncogenes (4). These activities have been observed in lung cancer with examples of specific miRNA effects on promoting tumor growth. For example, an analysis of 60 pairs of human non-small cell lung cancer (NSCLC) tissue samples compared to non-cancerous lung tissue by Shi *et al.* was able to demonstrate the inverse relationship of a specific miRNA molecule expression with the histological grade of the tumor and confirm significantly decreased levels of the molecule in cancerous tissue compared to the normal tissue (5). Studies such as these have led to exploring miRNA as a screening adjunct in lung cancer. A three-miRNA marker panel was confirmed across multiple independent studies for increased sensitivity in identifying lung cancer.

Further studies discovered various such panels. A study by Boeri *et al.* identifies miRNA signatures in plasma that predicted cancer development and prognosis in samples collected 1–2 years before disease onset amongst 53 patients of all-stage lung cancers (6). Currently, two different miRNA markers are in an advanced phase of development. The miRNA signature classifier (MSC) is based on 24 miRNA molecules. Patients are categorized into three levels of risk of cancer (low, intermediate, and high) by determining the molecules' four different expression ratio signatures. The intermediate- and high-risk categories resulted in 87% sensitivity and 81% specificity in correctly classifying patients with lung cancer. MSC was able to significantly differentiate survival at 3 years—low 100%, intermediate 97%, and high 77%—speaking to the possibility of identifying the aggressiveness of tumor with miRNA. The study also demonstrated a five-fold reduction of the false positive rate of LDCT to 3.7% when LDCT and MSC were utilized together as a screening tool (7). The second miRNA marker, miR-Test, is a 34 miRNA signature that demonstrated an accuracy of 74.9%, a sensitivity of 77.8%, and a specificity of 74.8% in a validation cohort of the COSMOS trial. With this encouraging result, the prospective COSMOS II trial is currently ongoing (8).

DNA methylation

DNA methylation is one of the primary epigenetic mechanisms responsible for gene regulation. Methylation of most CpG islands is seen in normal cells with certain regions which are hypomethylated. Aberrant hypermethylation of CpG islands on tumor-suppressor genes plays a role in cancer development, and these changes

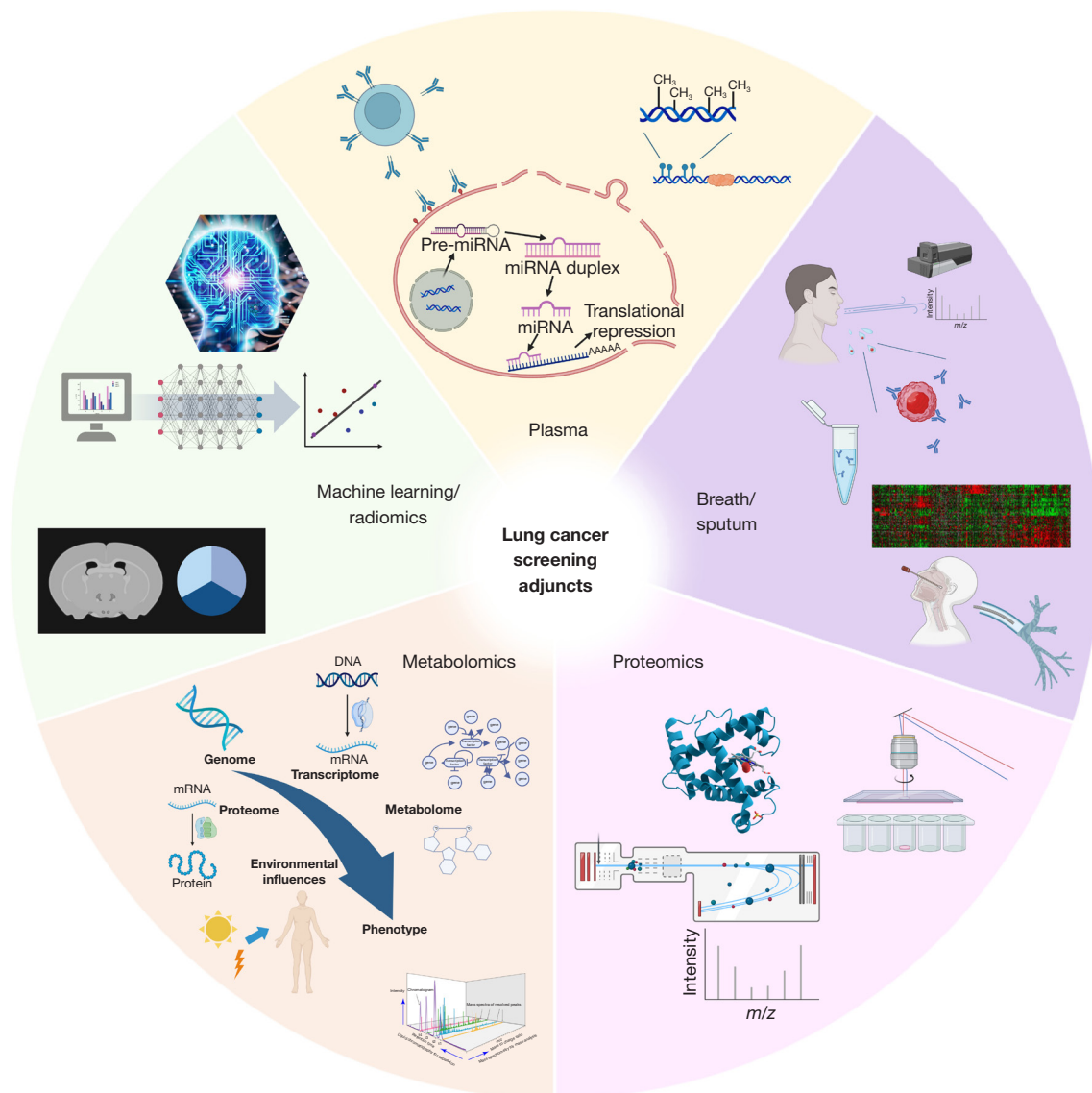


Figure 1 Lung cancer screening adjuncts. Figure adapted from: Thoracic Surgery Clinics. Permission for this reproduction is obtained. Pre-miRNA, precursor microRNA; mRNA, messenger RNA.

can also be noted in serum samples of patients. Such changes have also been found in lung cancer patients. In an epigenome-wide analysis by Ooki *et al.*, stage I NSCLC samples were analyzed with the recognition of 30 genes of differentially methylated regions across lung cancer specimens. A six-gene methylation panel was developed from this selection with 62.5–87.5% sensitivity for lung cancer (9). Lung EpiCheck is a polymerase chain reaction (PCR)-based six-methylation marker assay that utilizes plasma samples. The assay was validated in two different populations—European and Chinese—testing the feasibility,

specificity, and sensitivity of the test. Lung EpiCheck detected 70–85% of early-stage NSCLC in these validation experiments. Such results have been promising in affirming the feasibility of clinical use in early lung cancer detection. This test can potentially be used to improve the sensitivity of identifying the high-risk population from our current demographic and exposure factors standard. Not only was Lung EpiCheck able to provide high discrimination for lung cancer at 94.2% when combined with risk factors, there was no significant difference in the strong relationship between lung cancer cases and the test result regardless of

the presence of risk factors (10).

Tumor-associated autoantibodies (TAAbs)

TAAbs may be more easily detected as they circulate in the blood longer than the antigens. These autoantibodies are generated locally, and the immune response amplifies early, even before the tumor is clinically detectable. Studies have found the presence of autoantibodies in serum samples of lung cancer patients up to 5 years before obtaining screening CT scans (11). A meta-analysis of 31 articles revealed a diagnostic accuracy of 78.4% (range, 67.5–88.8%) for a panel of seven TAAbs with an overall area under the curve (AUC) of 0.90 (0.87–0.93) in patients at all stages of lung cancer (12). EarlyCDT lung is currently the only commercially available test. It has been validated in a multinational population and demonstrated the ability to detect early- and late-stage disease with up to 40% sensitivity and 91% specificity. An audit study on the clinical utility of EarlyCDT lung also demonstrated a 41% sensitivity with 57% of the detected cancers found in early stage (I/II) (13). However, a recent systematic review of studies utilizing EarlyCDT lung alone in five different cohorts revealed a decreased sensitivity of 22% at 92% specificity, compared to the estimate by the manufacturer (14). This result demonstrates the limitations of EarlyCDT lung as a stand-alone tool. However, combining EarlyCDT with LDCT and other biomarkers could potentially decrease false positive results within select populations and enhance the current screening options.

Breath/sputum biomarkers

Airway and sputum biomarkers are another area of study that benefits from a lower cost of collection and noninvasiveness. In addition, the ease of collecting breath and sputum samples makes this field of study especially attractive. The recently broadened screening criteria by the United States Preventive Services Task Force are estimated to almost double the pool of patients eligible for screening (15). Considering this upcoming need, there is a rapidly growing interest in identifying relevant biomarkers in airway epithelium, exhaled breath, and sputum.

Field of injury

The “field of injury” hypothesis postulates that an inhaled carcinogen—such as tobacco smoke—induces cellular

injury and molecular changes throughout all areas of the respiratory tract that it encounters, from the nasal cavity to the bronchi. While bronchial airway epithelium may often appear normal on bronchoscopy, these cellular and molecular changes can contribute to the development of premalignant and frank malignant lesions (16).

AEGIS-1 and AEGIS-2, a set of two multicenter prospective studies conducted by Silvestri *et al.* and part of the overarching AEGIS clinical trials, followed patients undergoing bronchoscopy for suspected lung cancer. The AEGIS bronchial genomic classifier study included 639 patients who were either current or former smokers. This study showed that when bronchoscopy alone was used for screening, the sensitivity of the screening test was 74–76%, but sensitivity rose to 96–98% when a bronchial-airway gene expression classifier was measured in epithelial cells collected from bronchoscopy (17). In a follow-up prospective study, the team collected nasal epithelium samples—representing another section of the respiratory tract—from 505 unique AEGIS-1 and AEGIS-2 patients and profiled them via gene microarrays: 309 of these patients had a confirmed lung cancer diagnosis, while 196 had benign disease. The team discovered 535 genes differentially expressed in the nasal epithelium of patients diagnosed with lung cancer compared to patients with benign disease.

Upregulated genes included those involved in endocytosis and ion transport, while downregulated genes were involved in DNA damage checkpoints, apoptosis regulation, and immune system activation. Moreover, for both the upregulated and downregulated groups of genes, they found strong concordance between nasal and bronchial sample cancer-related changes in gene expression patterns, supporting the “field of injury” model. The team developed a model combining clinical risk factors, genes associated with those risk factors, and genes associated with lung cancer found to have altered expression in nasal epithelium samples. Compared to a model that only used clinical risk factors, the screening test’s sensitivity and negative predictive value significantly increased from 0.79 to 0.91 and from 0.73 to 0.85, respectively (18).

In a more recent study, Qureshi *et al.* used network analysis to identify key genes associated with lung cancer in nasal epithelium, including *EMR3* (implicated in cell and leukocyte migration), *NCF1* (a regulator of reactive oxygen species production), *DYSF* (a regulator of plasma membrane repair), and others (19). Given the above results, the usage of nasal epithelial brushings as a biomarker for lung cancer

screening shows great promise, especially considering its ease of collection. Prospective studies collecting nasal brushings alongside other biological samples to better detect early lung cancer is ongoing, including efforts by the DECAMP consortium, consisting of 15 U.S. military treatment facilities, Veterans Affairs (VA) hospitals, and academic centers (20).

Volatile organic compounds (VOCs)

VOCs are organic chemical compounds that can be exogenous (inhaled or absorbed through the skin) or endogenous (produced via physiological and metabolic processes). Exogenous VOCs are reflective of an individual's exposure to carcinogens. While tumor cells can produce endogenous VOCs that are either usually not found in healthy patients or that become abnormally high in concentration relative to healthy patients. After these tumor cell-produced VOCs are released into the endobronchial cavity, they can be exhaled in the breath—thus ultimately becoming a component of exhaled breath condensate (EBC)—or become distributed within the blood and other bodily fluids (21). A variety of VOCs have been associated with lung cancer. Individual VOCs are typically identified using gas chromatography coupled with mass spectrometry—an expensive and time-intensive process—while cross-reactive sensor arrays, also referred to as “electronic noses” or “e-noses”, can define specific patterns of disease-related VOCs (22). Several e-nose technologies, including the Cyranose 320 used by McWilliams *et al.* and BIONOTE by Rocco *et al.*, have shown promising results in clinical studies (23,24). Shang *et al.* recently developed a novel, portable, and wireless chemiresistive sensor array system utilizing nanotechnology and machine learning. They successfully demonstrated the effective separation of a group of lung cancer-related VOCs in both simulations and experimental conditions. Experiments using human breath samples with and without lung cancer-specific VOCs showed that the system could discriminate between “lung cancer” and “healthy” breath signatures. Along with this specificity, the system can detect compounds with concentrations as low as 6 ppb (25). The precise number and identities of VOCs needed to reliably detect early-stage lung cancer is still an ongoing investigation.

EBC

In addition to VOCs, other molecules have been found in

EBC, such as polypeptides, proteins, DNA, mitochondrial DNA, and miRNAs. Consequently, some early studies have been conducted on the proteomics, genomics, and epigenomics of EBC (26). Through the usage of next-generation sequencing (NGS), Youssef *et al.* found 39 hotspot mutations in the EBC—reflecting DNA from pulmonary tissue—of lung cancer patients and 35 hotspot mutations in controls' EBC, with a higher average mutant allele fraction in lung cancer patients (27). Pérez-Sánchez *et al.* analyzed ECB samples from lung cancer patients using genome-wide miRNA profiling and machine learning analysis. They found nine miRNAs that were significantly upregulated and three miRNAs significantly downregulated compared to the ECB of healthy controls. They also found three miRNAs that were highly sensitive to detection and highly specific to lung cancer; several other miRNAs were highly specific to either lung adenocarcinoma or squamous cell carcinoma in particular (28). A recent study using a panel of 24 miRNAs upregulated in lung cancer demonstrated a modest increase in case-control discrimination—between 1.1% to 2.5%—compared to using only clinical models (29).

Sputum

Sputum is another easily acquired and often entirely non-invasively collected biospecimen for which biomarkers are currently being investigated. One challenge with this approach is that the sputum must exclusively come from the lungs and thus can potentially be contaminated by salivary proteins as it transits through the oral cavity. Additionally, the data must be appropriately normalized, as samples can have significant variations in dilution (30). Nevertheless, several successful proteomic and epigenomic studies have been conducted so far using sputum samples from lung cancer patients. For example, Yu *et al.* demonstrated that protein enolase 1 (ENO1) concentration was roughly four times higher within the sputum of early-stage lung cancer patients *vs.* that of controls, making it a potential sputum biomarker (31). Another study by Ali-Labib *et al.* quantitatively analyzed matrix metalloproteinase-2 (MMP-2), a well-known contributor to tumor invasion and metastasis, in sputum using an enzyme-linked immunosorbent assay (ELISA) assay. They found that MMP-2 was significantly increased in patients with lung cancer compared to benign and control groups, and its concentration increased exponentially as the severity of lung cancer phenotype increased (32). Finally, Su *et al.*, who

had previously identified three sputum miRNA biomarkers with 82.9% sensitivity and 87.8% specificity for early-stage lung cancer, as well as two sputum small nucleolar RNA (snoRNA) biomarkers with 74.6% sensitivity and 83.6% specificity, found that when they integrated these biomarkers, both the sensitivity and specificity increased to 89%, reflecting a synergistic effect (33).

The study of VOCs, EBC, and sputum biomarkers is still a relatively nascent field. However, it also reflects an immense potential for developing non-invasive, affordable, highly sensitive, and specific tests for detecting early-stage lung cancer.

Tissue studies

Proteomics

The proteome is a dynamic and highly complex system that reflects the final stage of biological information from a genome. Thus, the information garnered from proteomics is highly specific and theorized to be closer to the final tissue product than information acquired through genomics. Proteomics differs from singular protein studies in that it allows for studying protein expression and interaction amongst proteins specific to a tissue sample (34). Proteomics is an emerging field in the study of cancer that can be utilized to identify and compare the protein content between normal and tumor tissues. Quantitative proteomics provides information regarding important molecular interactions as well as biomarker identification. This approach has become more feasible in recent years due to the development of technology such as multiple reaction monitoring and the improvement of other mass spectrometry-based techniques. Studies utilizing liquid biopsies and/or cancerous tissues have demonstrated the potential of utilizing proteomics to improve earlier lung cancer detection and to discover possible therapeutic targets.

Serum studies are easily collected and are highly reproducible. An early study by Song *et al.* was able to fractionate serum proteins from lung cancer patients, tuberculosis patients, and healthy controls utilizing a magnetic bead-based surface-enhanced laser desorption/ionization time-of-flight mass spectrometry. This analysis identified a four-peak model representing four separate proteins (chaperonin, hemoglobin subunit beta, serum amyloid A, and an unknown protein), which could discriminate lung cancer patients from the remainder of the cohort with 93.3% sensitivity and 90.5% specificity in the

training set (35). Another study by Kim *et al.* included 198 patients, half with lung cancer and the other half with non-cancerous lung diseases (tuberculosis, pneumonia, and non-cancerous nodules) were studied. Serum from these patients was collected, and nano-flow liquid chromatography with multiple-reaction monitoring mass spectrometry was utilized to obtain proteomic profiles. They demonstrated a significant change in a single protein subtype—serpin peptidase inhibitor, clade A, member 4 (SERPINA4) between patients with lung cancer *vs.* other lung diseases ($P < 0.001$). The study looked to further improve the diagnostic power and was able to determine a meta-marker comprised of serum paraoxonase 1 (PON1), SERPINA4, and age, which was noted to be a potential combined marker for lung cancer utilizing logistic regression (36).

A more recent case-control pilot study by Gasparri *et al.* studied 87 subjects with stage I lung cancer and healthy controls. Serum samples were obtained from these patients, and proteomes within circulating microvesicles were analyzed via liquid chromatography with tandem mass spectrometry (LC-MS/MS). They were able to identify 33 proteins with significant differential expression across the serum samples, which included arylsulfatase A (ARSA) and protein kinase α -type (PRKCA)—these two proteins in altered levels have previously been identified to be associated with lung cancer in separate studies (37). This study demonstrated that despite the fluid nature of serum composition, a fractionation strategy utilizing microvesicles could recognize potential proteomic biomarkers for early detection of lung cancer.

Although many proteomics studies involving lung cancer utilized plasma specimens, some studies have used tissue samples. One such study by Hsu *et al.* utilized quantitative proteomic analyses of lung cancer tissues. Frozen lung adenocarcinoma tissue and adjacent normal lung tissue were studied with a discovery phase of fourteen paired tissues, followed by a validation cohort of 48 patients. A pathway analysis of stage I lung cancer tissues determined that 133 proteins were upregulated in cancer tissues compared to adjacent normal tissues via quantitative proteomics. These proteins were primarily upregulated in translation, elongation termination, and protein folding. Subsequent validation of these findings by immunohistochemistry staining and western blot was also performed in order to narrow down to six potential biomarkers (38). By taking this discovery and expanding upon it, future studies can be postulated to detect these specific biomarkers in plasma, breath, or biopsy specimens so that early detection of these

markers can lead to increased diligence in screening and surveillance for disease.

These studies demonstrate the feasibility of applying proteomic technology to discover alternative candidate biomarkers for lung cancer detection. Utilization of readily available plasma and tissue samples to obtain more conclusive biomarker results could potentially lead to studies that utilize pre-resection biopsy specimens to increase methods of early detection of lung cancer.

Metabolomics

Metabolomics is an emerging application in diagnosing, detecting, and treating cancers. The field examines differences between the metabolomes of normal and cancer tissues via the detection of changes in metabolites, and it has been able to determine promising markers that can aid in the early detection of lung cancer (39).

Overview of Warburg effect and glycolysis-associated pathways

One hallmark of cancer energy metabolism is the preferential utilization of glycolysis, regardless of the availability of oxygen (40). First described by Otto Warburg in the 20th century, the phenomenon has been proven to be clinically advantageous in developing tumor detection modalities, one of which is fluorodeoxyglucose positron emission tomography (FDG-PET) (41). True to this principle, lung cancer drastically reprograms metabolic pathways compared to healthy tissues. Elevated glucose transporters (GLUTs) levels have been implicated in increasing glucose intake in lung cancer tissues (42). More specifically, GLUT1 has been shown to be overexpressed in primary lung adenocarcinoma and associated with *KRAS* mutation (43). Kurata *et al.* have also reported that GLUT3 and GLUT5 are upregulated in metastatic lesions compared to primary lung lesions (44). Furthermore, important rate-limiting glycolysis enzymes such as hexokinase and phosphofructokinase are upregulated in lung cancer tissues, as evidenced by the results of various research efforts (45).

Lung cancers have also shown the ability to modify other glycolysis-associated metabolic pathways, the metabolites of which serve some potential for early screening. In 2016, Hensley *et al.* demonstrated that lactate, thought previously to be only a by-product of cancerous growth, plays a role in serving as a respiratory substrate in tissues of patients with NSCLC (46). Monocarboxylate transporters (MCTs), proteins responsible for lactate intake, and lactase

dehydrogenase isoform B (LDHB), enzymes that favor tissues using lactate as fuel, are both upregulated in NSCLC (47,48). Moreover, Faubert *et al.* have provided data indicating that NSCLC can use lactate as a fuel *in vivo*, with concomitant rates of FDG and lactate consumption (49). In the same study, the authors noted that a PET lactate probe exists for cardiac metabolism imaging. A similar tool could potentially be developed to further assess the role of lactate in human NSCLC (49).

Amino acid metabolism pathways

In addition to glycolysis, a shift towards glutamine metabolism is also considered a differentiating characteristic of cancer growth (50). In cancer cells, pyruvate, the end-product of glycolysis, is understood to be preferentially converted to lactic acid rather than to be used as the starting point for the tricarboxylic acid cycle (TCA) (42). Thus, cancer cells replenish the metabolites for the TCA cycle by enabling glutamine anaplerosis (51). In line with this understanding, lung cancers have been shown to significantly increase glutamine uptake (52). Solute-linked carrier family 1 member A5 (SLC1A5), a high-affinity transporter of glutamine, is reported to be upregulated in tissues of stage I lung adenocarcinoma and squamous cell carcinoma (53). Furthermore, Zhang *et al.* show in a targeted metabolic profiling study that glutamate and aspartic acid, two products of glutamine metabolism, are upregulated in serum of stage I lung cancer patients relative to their healthy counterparts (54).

Another known metabolite to lung cancer pathogenesis is serine, the precursors for which are products of glycolysis and glutaminolysis (42). In normal cells, the production of serine leads to the crucial carbon donation to one-carbon metabolic pathways, and this process has been implicated with various oncogenic genetic markers in the development of different types of tumors (42,55). Using lung cancer patients-derived tissues, Yao *et al.* show that expression of methylenetetrahydrofolate dehydrogenase 2 (MTHFD2), an enzyme in one-carbon metabolism, is particularly associated with adenocarcinoma proliferation (56). MTHFD2 is also associated with a worse prognosis in lung cancer patients as well (57). As MTHFD2 is not typically expressed in healthy adult tissues, its presence can potentially serve as a novel predictive marker (56,58).

Lipid metabolism pathways

The increased rates of glycolysis and availability of energy sources in cancer cells are subsequently coupled with

heightened lipid metabolic activity (42). Some key enzymes in lipid metabolism have been identified to be upregulated in lung cancer patients. Using surgically resected tissue samples from 106 patients of various lung cancer subtypes, Visca *et al.* reported the overexpression of fatty acid synthase (FAS), an enzyme responsible for the synthesis of long-chain fatty acids, in cancerous tissue (59). The rate-limiting enzyme of fatty acid synthesis, acetyl-CoA carboxylase, has similarly been found to be upregulated in NSCLC tissues in a cohort of 63 patients (60).

Other products of the fatty acid synthesis pathway have also been found to be associated with lung cancer. Mitchell *et al.* conducted a paired analysis of stage I or IIA primary NSCLC and found that cancer tissues had higher sterol and sphingolipid levels relative to healthy cells (61). In the case of the higher sterol content, this finding concurs with a previous study demonstrating that the risk of lung cancer is increased by abnormally high or low blood cholesterol levels (62). Considering the widespread challenges in maintaining optimal cholesterol levels, particularly in developed countries, sterol markers are likely to serve only in an adjunctive role to other more differentiating metabolite markers. Similarly, the sphingolipid metabolism pathway, with the molecule ceramide at the center, is believed to be highly involved in lung cancer progression (42). In pairwise comparisons of serum lipid profiles, Smolarz *et al.* showed that ceramide is elevated in screening-detected lung cancer patients relative to patients with benign lung nodules and healthy controls (63). In terms of other sphingolipid-associated metabolites, Ni *et al.* specifically identified D-erythro-sphingosine 1-phosphate and palmitoyl sphingomyelin as being upregulated in patients with early-stage NSCLC (64). These detectable changes in lipid enzyme profiles and serum lipid markers can hopefully provide another tool for more sensitive and earlier lung cancer screening tools.

Machine learning

Machine learning has emerged as a promising technology for improving the accuracy of lung cancer diagnosis. Its application in imaging diagnostics increases the sensitivity of lung cancer screening, which decreases morbidity and mortality linked with lung cancer. Furthermore, it can lessen the burden of diagnoses on radiologists. Machine learning models can employ large datasets quickly and effectively to identify patterns that may not be clear to

human observers (65).

Machine learning in lung cancer diagnosis

There are several types of machine learning algorithms, including K-means clustering and random forest. These algorithms can be trained using large datasets of medical images (radiomics), clinical data, and other patient information useful in identifying patterns that could indicate lung cancer (66). With the support of machine learning models, clinicians can make more accurate and timely diagnoses, improving their patients' survival rates (65).

Deep learning algorithms are a subset of machine learning algorithms that focus on recognizing complex patterns in the data to produce accurate insights and predictions. They are based on artificial neural networks (ANNs), in which multiple processing layers extract progressively higher-level features from data. These algorithms learn by example and improve their function by imitating how humans think and learn.

Early-stage detection of lung cancer through screening has led to greater survival rates. A non-invasive, accurate, and quick method, which can be accomplished with machine learning, is ideal for screening for early-stage lung cancer. Deep learning models and algorithms like convolutional neural networks (CNNs) have been shown to be effective at analyzing radiometric data and identifying patterns that may indicate the presence of lung cancer. For example, a study by Protonotarios *et al.* demonstrated that a deep learning algorithm was able to accurately identify lung cancer lesion segmentation on PET/CT scans with an average accuracy rate of 98.9% for PET/CT scans together and a high average of 99.1% for the Co-learning method (67).

Furthermore, deep learning can also differentiate between benign and malignant nodules, which can aid in selecting appropriate treatment. Wang *et al.* created a model based on retrospective non-contrast thin-layer CT scans to distinguish between benign and malignant solid lung nodules. The model's accuracy was 98.9%, which was higher than other models' accuracies (68). Schwyzer *et al.* used ANNs to assess the detection of lung cancer based on varying doses of PET (69). For the standard-dose PET, the ANN model had a sensitivity of 95.9% and a specificity of 98.1%. Contrarily, for the ultralow dose PET (3.3%), the model had a sensitivity of 91.5% and a specificity of 94.2%. Huang *et al.* aimed to create a breath test to detect lung cancer using a chemical sensor array and machine learning

algorithms (70). Studying prospective lung cancer cases between 2016 to 2018, they analyzed alveolar air samples using carbon nanotube sensor arrays. As a result, the model had an AUC of 0.91 using the linear discriminant analysis and 0.90 by the support vector machine algorithm.

Predicting patient outcomes

In addition to improving the accuracy of lung cancer diagnosis, screening tools combined with machine learning may also predict future lung cancer risk. Machine learning algorithms can identify factors that may impact a patient's diagnosis by analyzing large datasets of patient information. For instance, the PLCO (Prostate, Lung Colorectal, and Ovarian Cancer Screening) model's top five variables in increasing importance for risk prediction are family history of lung cancer, smoking years, quit time, cigarettes per day, and age (71).

Sybil is another deep learning model by Mikhael *et al.* that uses radiometric data (LDCTs) to predict future lung cancer risk 1–6 years after screening (72). The model was trained using the National Lung Screening Trial (NLST) dataset and was validated using the scans from Massachusetts General Hospital (MGH) and Chang Gung Memorial Hospital (CGMH). They attained an AUC of 0.92 for the first year's prediction. Sybil does not require clinical data or radiologist annotation and can run in real-time in the background, which allows for faster responses and greater efficiency in its use.

Limitations of machine learning in lung cancer diagnosis

Although machine learning has shown great promise as an adjunct to lung cancer diagnosis, several limitations need to be considered with its use. One of the primary limitations is its need for large amounts of high-quality data. Creating or obtaining these datasets may be challenging in countries or regions where access to medical equipment is limited. Furthermore, the quality of the data (occurrence of incomplete or incorrect data) can impact the algorithm's accuracy.

Typically, deep learning modeling is managed through "black box" development, meaning that the models reach a conclusion or produce a result without explaining how they did so. This lack of interpretability in machine learning models can be problematic regarding validity and fairness. If the model's decisions cannot be demonstrated, it is difficult to justify its results (73). Unfairness is also a limitation of

machine learning, known as algorithmic bias. This bias can lead to inaccurate diagnoses or treatment recommendations, especially for minority populations that may be underrepresented in the dataset. Addressing algorithmic bias requires ongoing monitoring of the algorithm to ensure that it remains accurate and unbiased (74). Lastly, it must be noted that the accuracy of machine learning algorithms is often limited to the specific dataset used to train it. The algorithm may not perform as well on new datasets or clinical settings. Therefore, validating the accuracy of machine learning algorithms in diverse patient populations and clinical settings before they are widely implemented is critical (75).

Machine learning is an encouraging technology for improving the accuracy of lung cancer diagnosis and predicting patient outcomes. Deep learning algorithms have effectively analyzed radiometric data and identified patterns suggestive of lung cancer. However, there are some limitations, such as the need for high-quality data, the chance for algorithmic bias, and limited generalizability. Therefore, machine learning can be used as a supportive tool to clinicians for diagnosis and treatment options to improve patients' survival rates.

Conclusions

Improvement in the accuracy of lung cancer screening and in identifying further individuals who may benefit from screening is an active area of research. A wide breadth of technologies is being developed, from plasma-related biomarker testing to radiomic imaging and machine learning models. Several promising prospects are currently reviewed in this chapter, which explores various ways to make incremental improvements to the current standards of screening. The survey of multiple smaller powered studies via many different institutions and populations as well as the incomplete nature of the developing technologies present as limitations of this review, however, this provides a foundation for understanding the future directions for improvement. So far, the most advanced studies in markers, such as plasma/breath/sputum biomarkers, have demonstrated higher sensitivity and specificities that could enhance the current screening process with CT scans. Areas requiring more extensive research, such as proteomics and metabolomics, show potential for markers or possibly combined marker panels that could apply to improve screening technology. Machine learning is a newer technology that can be expanded upon with

multiple learning iterations amongst different populations, potentially allowing for early detection of lung cancer using imaging alone. Lastly, given the different limitations of these studies, it may be prudent to formulate a combined modality of lung cancer screening technology that can utilize the best features of each technology to enhance the sensitivity and applicability of lung cancer screening. Further work on these new technologies is needed, but the current outcomes are promising.

Acknowledgments

We would like to thank Wayne Pereanu (affiliation: Equestria Developers) for his assistance in figure formulation and review.

Funding: None.

Footnote

Peer Review File: Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-23-1326/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-23-1326/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Sharma D, Newman TG, Aronow WS. Lung cancer screening: history, current perspectives, and future directions. *Arch Med Sci* 2015;11:1033-43.
- Hammer MM, Byrne SC, Kong CY. Factors Influencing the False Positive Rate in CT Lung Cancer Screening. *Acad Radiol* 2022;29 Suppl 2:S18-22.
- Narayan AK, Chowdhry DN, Fintelmann FJ, et al. Racial and Ethnic Disparities in Lung Cancer Screening Eligibility. *Radiology* 2021;301:712-20.
- Iorio MV, Croce CM. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. *EMBO Mol Med* 2012;4:143-59.
- Shi ZM, Wang L, Shen H, et al. Downregulation of miR-218 contributes to epithelial-mesenchymal transition and tumor metastasis in lung cancer by targeting Slug/ZEB2 signaling. *Oncogene* 2017;36:2577-88.
- 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 U S A* 2011;108:3713-8.
- 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;32:768-73.
- Montani F, Marzi MJ, Dezi F, et al. miR-Test: a blood test for lung cancer early detection. *J Natl Cancer Inst* 2015;107:djv063.
- Oooki A, Maleki Z, Tsay JJ, 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;23:7141-52.
- Gaga M, Chorostowska-Wynimko J, Horváth I, et al. Validation of Lung EpiCheck, a novel methylation-based blood assay, for the detection of lung cancer in European and Chinese high-risk individuals. *Eur Respir J* 2021;57:2002682.
- Desmetz C, Mange A, Maudelonde T, et al. Autoantibody signatures: progress and perspectives for early cancer detection. *J Cell Mol Med* 2011;15:2013-24.
- Tang ZM, Ling ZG, Wang CM, et al. Serum tumor-associated autoantibodies as diagnostic biomarkers for lung cancer: A systematic review and meta-analysis. *PLoS One* 2017;12:e0182117.
- Jett JR, Peek LJ, Fredericks L, et al. Audit of the autoantibody test, EarlyCDT®-lung, in 1600 patients: an evaluation of its performance in routine clinical practice. *Lung Cancer* 2014;83:51-5.
- Duarte A, Corbett M, Melton H, et al. EarlyCDT Lung blood test for risk classification of solid pulmonary nodules: systematic review and economic evaluation. *Health Technol Assess* 2022;26:1-184.
- Marshall RC, Tiglaio SM, Thiel D. Updated USPSTF

- screening guidelines may reduce lung cancer deaths. *J Fam Pract* 2021;70:347-9.
16. Steiling K, Ryan J, Brody JS, et al. The field of tissue injury in the lung and airway. *Cancer Prev Res (Phila)* 2008;1:396-403.
 17. Silvestri GA, Vachani A, Whitney D, et al. A Bronchial Genomic Classifier for the Diagnostic Evaluation of Lung Cancer. *N Engl J Med* 2015;373:243-51.
 18. Shared Gene Expression Alterations in Nasal and Bronchial Epithelium for Lung Cancer Detection. *J Natl Cancer Inst* 2017;109:djw327.
 19. Qureshi N, Chi J, Qian Y, et al. Looking for the Genes Related to Lung Cancer From Nasal Epithelial Cells by Network and Pathway Analysis. *Front Genet* 2022;13:942864.
 20. Billatos E, Duan F, Moses E, et al. Detection of early lung cancer among military personnel (DECAMP) consortium: study protocols. *BMC Pulm Med* 2019;19:59.
 21. Haick H, Broza YY, Mochalski P, et al. Assessment, origin, and implementation of breath volatile cancer markers. *Chem Soc Rev* 2014;43:1423-49.
 22. Rocco G, Pennazza G, Santonico M, et al. Breathprinting and Early Diagnosis of Lung Cancer. *J Thorac Oncol* 2018;13:883-94.
 23. McWilliams A, Beigi P, Srinidhi A, et al. Sex and Smoking Status Effects on the Early Detection of Early Lung Cancer in High-Risk Smokers Using an Electronic Nose. *IEEE Trans Biomed Eng* 2015;62:2044-54.
 24. Rocco R, Incalzi RA, Pennazza G, et al. BIONOTE e-nose technology may reduce false positives in lung cancer screening programmes†. *Eur J Cardiothorac Surg* 2016;49:1112-7; discussion 1117.
 25. Shang G, Dinh D, Mercer T, et al. Chemiresistive Sensor Array with Nanostructured Interfaces for Detection of Human Breaths with Simulated Lung Cancer Breath VOCs. *ACS Sens* 2023;8:1328-38.
 26. Campanella A, De Summa S, Tommasi S. Exhaled breath condensate biomarkers for lung cancer. *J Breath Res* 2019;13:044002.
 27. Youssef O, Knuutila A, Piirilä P, et al. Hotspot Mutations Detectable by Next-generation Sequencing in Exhaled Breath Condensates from Patients with Lung Cancer. *Anticancer Res* 2018;38:5627-34.
 28. Pérez-Sánchez C, Barbarroja N, Pantaleão LC, et al. Clinical Utility of microRNAs in Exhaled Breath Condensate as Biomarkers for Lung Cancer. *J Pers Med* 2021;11:111.
 29. Shi M, Han W, Loudig O, et al. Initial development and testing of an exhaled microRNA detection strategy for lung cancer case-control discrimination. *Sci Rep* 2023;13:6620.
 30. D'Amato M, Iadarola P, Viglio S. Proteomic Analysis of Human Sputum for the Diagnosis of Lung Disorders: Where Are We Today? *Int J Mol Sci* 2022;23:5692.
 31. Yu L, Shen J, Mannoor K, et al. Identification of ENO1 as a potential sputum biomarker for early-stage lung cancer by shotgun proteomics. *Clin Lung Cancer* 2014;15:372-378.e1.
 32. Ali-Labib R, Louka ML, Galal IH, et al. Evaluation of matrix metalloproteinase-2 in lung cancer. *Proteomics Clin Appl* 2014;8:251-7.
 33. Su Y, Guarnera MA, Fang H, et al. Small non-coding RNA biomarkers in sputum for lung cancer diagnosis. *Mol Cancer* 2016;15:36.
 34. Harpole DH, Meyerson SL. Lung cancer staging: proteomics. *Thorac Surg Clin* 2006;16:339-43.
 35. Song QB, Hu WG, Wang P, et al. Identification of serum biomarkers for lung cancer using magnetic bead-based SELDI-TOF-MS. *Acta Pharmacol Sin* 2011;32:1537-42.
 36. Kim YI, Ahn JM, Sung HJ, et al. Meta-markers for the differential diagnosis of lung cancer and lung disease. *J Proteomics* 2016;148:36-43.
 37. Gasparri R, Noberini R, Cuomo A, et al. Serum proteomics profiling identifies a preliminary signature for the diagnosis of early-stage lung cancer. *Proteomics Clin Appl* 2023;17:e2200093.
 38. Hsu CH, Hsu CW, Hsueh C, et al. Identification and Characterization of Potential Biomarkers by Quantitative Tissue Proteomics of Primary Lung Adenocarcinoma. *Mol Cell Proteomics* 2016;15:2396-410.
 39. Tang Y, Li Z, Lazar L, et al. Metabolomics workflow for lung cancer: Discovery of biomarkers. *Clin Chim Acta* 2019;495:436-45.
 40. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646-74.
 41. Hsu PP, Sabatini DM. Cancer cell metabolism: Warburg and beyond. *Cell* 2008;134:703-7.
 42. Kannampuzha S, Mukherjee AG, Wanjari UR, et al. A Systematic Role of Metabolomics, Metabolic Pathways, and Chemical Metabolism in Lung Cancer. *Vaccines (Basel)* 2023;11:381.
 43. Pezzuto A, D'Ascanio M, Ricci A, et al. Expression and role of p16 and GLUT1 in malignant diseases and lung cancer: A review. *Thorac Cancer* 2020;11:3060-70.
 44. Kurata T, Oguri T, Isobe T, et al. Differential expression of facilitative glucose transporter (GLUT) genes in primary

- lung cancers and their liver metastases. *Jpn J Cancer Res* 1999;90:1238-43.
45. Li XB, Gu JD, Zhou QH. Review of aerobic glycolysis and its key enzymes - new targets for lung cancer therapy. *Thorac Cancer* 2015;6:17-24.
 46. Hensley CT, Faubert B, Yuan Q, et al. Metabolic Heterogeneity in Human Lung Tumors. *Cell* 2016;164:681-94.
 47. Eilertsen M, Andersen S, Al-Saad S, et al. Monocarboxylate transporters 1-4 in NSCLC: MCT1 is an independent prognostic marker for survival. *PLoS One* 2014;9:e105038.
 48. McClelland ML, Adler AS, Deming L, et al. Lactate dehydrogenase B is required for the growth of KRAS-dependent lung adenocarcinomas. *Clin Cancer Res* 2013;19:773-84.
 49. Faubert B, Li KY, Cai L, et al. Lactate Metabolism in Human Lung Tumors. *Cell* 2017;171:358-371.e9.
 50. DeBerardinis RJ, Mancuso A, Daikhin E, et al. Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc Natl Acad Sci U S A* 2007;104:19345-50.
 51. Yang L, Venneti S, Nagrath D. Glutaminolysis: A Hallmark of Cancer Metabolism. *Annu Rev Biomed Eng* 2017;19:163-94.
 52. Matés JM, Di Paola FJ, Campos-Sandoval JA, et al. Therapeutic targeting of glutaminolysis as an essential strategy to combat cancer. *Semin Cell Dev Biol* 2020;98:34-43.
 53. Hassanein M, Hoeksema MD, Shiota M, et al. SLC1A5 mediates glutamine transport required for lung cancer cell growth and survival. *Clin Cancer Res* 2013;19:560-70.
 54. Zhang X, Zhu X, Wang C, et al. Non-targeted and targeted metabolomics approaches to diagnosing lung cancer and predicting patient prognosis. *Oncotarget* 2016;7:63437-48.
 55. Yang M, Vousden KH. Serine and one-carbon metabolism in cancer. *Nat Rev Cancer* 2016;16:650-62.
 56. Yao S, Peng L, Elakad O, et al. One carbon metabolism in human lung cancer. *Transl Lung Cancer Res* 2021;10:2523-38.
 57. Zhang WC, Shyh-Chang N, Yang H, et al. Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell* 2012;148:259-72.
 58. Nilsson R, Jain M, Madhusudhan N, et al. Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. *Nat Commun* 2014;5:3128.
 59. Visca P, Sebastiani V, Botti C, et al. Fatty acid synthase (FAS) is a marker of increased risk of recurrence in lung carcinoma. *Anticancer Res* 2004;24:4169-73.
 60. Li EQ, Zhao W, Zhang C, et al. Synthesis and anti-cancer activity of ND-646 and its derivatives as acetyl-CoA carboxylase 1 inhibitors. *Eur J Pharm Sci* 2019;137:105010.
 61. Mitchell JM, Flight RM, Moseley HNB. Untargeted Lipidomics of Non-Small Cell Lung Carcinoma Demonstrates Differentially Abundant Lipid Classes in Cancer vs. Non-Cancer Tissue. *Metabolites* 2021;11:740.
 62. Chen Q, Pan Z, Zhao M, et al. High cholesterol in lipid rafts reduces the sensitivity to EGFR-TKI therapy in non-small cell lung cancer. *J Cell Physiol* 2018;233:6722-32.
 63. Smolarz M, Kurczyk A, Jelonek K, et al. The Lipid Composition of Serum-Derived Small Extracellular Vesicles in Participants of a Lung Cancer Screening Study. *Cancers (Basel)* 2021;13:3414.
 64. Ni B, Kong X, Yan Y, et al. Combined analysis of gut microbiome and serum metabolomics reveals novel biomarkers in patients with early-stage non-small cell lung cancer. *Front Cell Infect Microbiol* 2023;13:1091825.
 65. Joy Mathew C, David AM, Joy Mathew CM. Artificial Intelligence and its future potential in lung cancer screening. *EXCLI J* 2020;19:1552-62.
 66. Kumar Y, Koul A, Singla R, et al. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J Ambient Intell Humaniz Comput* 2023;14:8459-86.
 67. Protonotarios NE, Katsamenis I, Sykiotis S, et al. A few-shot U-Net deep learning model for lung cancer lesion segmentation via PET/CT imaging. *Biomed Phys Eng Express* 2022. doi: 10.1088/2057-1976/ac53bd.
 68. Wang S, Zhou L, Li X, et al. A Novel Deep Learning Model to Distinguish Malignant Versus Benign Solid Lung Nodules. *Med Sci Monit* 2022;28:e936830.
 69. Schwyzer M, Ferraro DA, Muehlemaier UJ, et al. Automated detection of lung cancer at ultralow dose PET/CT by deep neural networks - Initial results. *Lung Cancer* 2018;126:170-3.
 70. Huang CH, Zeng C, Wang YC, et al. A Study of Diagnostic Accuracy Using a Chemical Sensor Array and a Machine Learning Technique to Detect Lung Cancer. *Sensors (Basel)* 2018;18:2845.
 71. Kobylńska K, Orłowski T, Adamek M, et al. Explainable machine learning for lung cancer screening models.

- Applied Sciences 2022;12:1926.
72. Mikhael PG, Wohlwend J, Yala A, et al. Sybil: A Validated Deep Learning Model to Predict Future Lung Cancer Risk From a Single Low-Dose Chest Computed Tomography. *J Clin Oncol* 2023;41:2191-200.
 73. Petch J, Di S, Nelson W. Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology. *Can J Cardiol* 2022;38:204-13.
 74. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. *J Glob Health* 2019;9:010318.
 75. Yu AC, Mohajer B, Eng J. External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. *Radiol Artif Intell* 2022;4:e210064.

Cite this article as: Park JA, Pham D, Yalamanchili S, Twardus S, Suzuki K. Developing technologies and areas of interest in lung cancer screening adjuncts. *J Thorac Dis* 2024;16(2):1552-1564. doi: 10.21037/jtd-23-1326