Biomarkers of lung cancer for screening and in never-smokers—a narrative review

Edward K. H. Stephens1#, Jazmin Guayco Sigcha1#, Kenneth Lopez-Loo1, Ian A. Yang1,2#, Henry M. Marshall1,2#, Kwun M. Fong1,2#

1UQ Thoracic Research Centre, Faculty of Medicine, The University of Queensland, Brisbane, Australia; 2Department of Thoracic Medicine, The Prince Charles Hospital, Brisbane, Australia

Contributions: (I) Conception and design: All authors; (II) Administrative support: IA Yang, HM Marshall, KM Fong; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: EKH Stephens, J Guayco Sigcha, K Lopez-Loo; (V) Data analysis and interpretation: EKH Stephens, J Guayco Sigcha, K Lopez-Loo; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work as co-first authors.

Correspondence to: Edward K. H. Stephens, BBiomed Sc, (Hons). UQ Thoracic Research Centre, Faculty of Medicine, The University of Queensland, Brisbane, Australia; Department of Thoracic Medicine, The Prince Charles Hospital, 627 Rode Road, Chermside QLD 4032, Brisbane, Australia. Email: edward.stephens@uq.net.au.

Background and Objective: Lung cancer is the leading cause of cancer-related mortality worldwide, partially attributed to late-stage diagnoses. In order to mitigate this, lung cancer screening (LCS) of high-risk patients is performed using low dose computed tomography (CT) scans, however this method is burdened by high false-positive rates and radiation exposure for patients. Further, screening programs focus on individuals with heavy smoking histories, and as such, never-smokers who may otherwise be at risk of lung cancer are often overlooked. To resolve these limitations, biomarkers have been posited as potential supplements or replacements to low-dose CT, and as such, a large body of research in this area has been produced. However, comparatively little information exists on their clinical efficacy and how this compares to current LCS strategies.

Methods: Here we conduct a search and narrative review of current literature surrounding biomarkers of lung cancer to supplement LCS, and biomarkers of lung cancer in never-smokers (LCINS).

Key Content and Findings: Many potential biomarkers of lung cancer have been identified with varying levels of sensitivity, specificity, clinical efficacy, and supporting evidence. Of the markers identified, multi-target panels of circulating microRNAs, lipids, and metabolites are likely the most clinically efficacious markers to aid current screening programs, as these provide the highest sensitivity and specificity for lung cancer detection. However, circulating lipid and metabolite levels are known to vary in numerous systemic pathologies, highlighting the need for further validation in large cohort randomised studies.

Conclusions: Lung cancer biomarkers is a fast-expanding area of research and numerous biomarkers with potential clinical applications have been identified. However, in all cases the level of evidence supporting clinical efficacy is not yet at a level at which it can be translated to clinical practice. The priority now should be to validate existing candidate markers in appropriate clinical contexts and work to integrating these into clinical practice.

Keywords: Non-small cell lung cancer (NSCLC); screening; biomarkers; never-smokers

Introduction

Lung cancer

Lung cancer is the leading cause of cancer-related mortality worldwide (1). In 2018, lung cancer accounted for an estimated 1,761,000 deaths, 18.4% of all cancer deaths worldwide (2). While overall cancer survival has improved in the last 30 years, with 70% of all individuals with cancer surviving 5 years after diagnosis, only 20% of individuals with lung cancer survive past 5 years following diagnosis (3). This high mortality rate can be attributed to late-stage diagnosis, with 75% of lung cancer diagnoses presenting at stages III and IV (4). In order to successfully treat individuals with lung cancer and reduce disease mortality, improvements must be made in early disease detection, screening, and prevention.

Overview of biomarkers

Biomarkers are biologically significant molecules that reflect the homeostatic state of the host tissue and can be used for identifying and characterising disease states, notably cancer (5). Biomarkers have a wide range of applications from disease diagnosis and screening, to predicting patient outcomes and responses to treatment (6). Biologically significant markers of disease have long been assessed for their applicability in disease management. Biomarkers can be classified as nucleic acid markers, protein markers, or cellular markers (Figure 1), referring to the functional level at which they operate (7). Recent interest in biomarkers of lung cancer has seen the identification of a wide range of markers with significant clinical applications, most prominently in cancer characterisation, with the identification of a range of targetable driver mutations which have significantly improved lung cancer treatment (8-10). Focus is now beginning to shift towards markers for early detection that work to either complement or potentially replace current diagnostic and screening methods with the aim of identifying lung cancer in its early stages while curative treatment options are still viable (11).

Biomarkers of lung cancer in never-smokers (LCINS)

Lung cancer occurs most commonly as a result of cigarette smoking (19). As such, the focus on screening individuals with a significant smoking history for lung cancer is justified,
however those with negligible smoking histories who may still be at risk fall through the gaps of screening efforts (20). LCINS represents a growing subgroup of individuals, up to 25% of lung cancer diagnoses worldwide, and is a significant contributor to cancer-related mortality (21-23). Common risk factors other than cigarette smoke include exposure to coal dust, asbestos, indoor and outdoor pollutants, environmental radon, nickel, chromium, and ionising radiation, as well as pre-existing respiratory conditions and genetic abnormalities (22,24-27). Radon exposure is considered the leading cause of LCINS—and the second leading cause of lung cancer overall—with individuals becoming exposed to high levels of radon when living or working in buildings with poor ventilation in areas of high environmental radon (28,29). While reported rates vary, 60–90% of lung cancers in people who have never smoked are adenocarcinomas (30,31). Although mortality rates greatly vary in LCINS, it is characterized by longer survival times than smoking-related lung cancer, despite typically being diagnosed at later disease stage (30,32). To reduce the high disease burden and mortality rate associated with LCINS, improved methods of cancer screening and early detection are required, and numerous biomarkers of LCINS have been identified and investigated over the last decade.

**Figure 1** Body fluid sources for lung cancer biomarkers, and the subtypes of biomarkers that can be derived from these sources. Created with BioRender.com.

**Focussed aims**

Although lung cancer biomarkers are the subject of intense research worldwide, there remains comparatively little information on their clinical efficacy and how this compares to current LCS strategies. Further, LCINS has been identified as molecularly, histologically, and pathologically distinct from tobacco smoking-related lung cancers and as such, requires distinctive characterisation (33,34). In this review we provide a focussed up to date summary on:

(I) Body-fluid biomarker research for LCS and how it may supplement screening strategies, and

(II) Circulating biomarkers of LCINS lending attention to applications for lung cancer detection.

We present this article in accordance with the Narrative Review reporting checklist (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-23-291/rc).

**Methods**

**Search strategy**

The search strategy is presented in Table 1. Briefly, in August 2022 we performed two systematic literature searches of English-language publications using the terms...
Table 1  Search strategies

<table>
<thead>
<tr>
<th>Items</th>
<th>Lung cancer screening</th>
<th>Lung cancer in never-smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of search</td>
<td>18th–24th of August 2022</td>
<td>18th–24th of August 2022</td>
</tr>
<tr>
<td>Databases and other sources searched</td>
<td>MEDLINE, CINAHL, Embase, and Web of Science databases</td>
<td>MEDLINE, CINAHL, Embase, and Web of Science databases</td>
</tr>
<tr>
<td>Search terms used</td>
<td>Detailed example on Table S1</td>
<td>Detailed example on Table S2</td>
</tr>
<tr>
<td>Timeframe</td>
<td>Studies published since 2010</td>
<td>Studies published since 2012</td>
</tr>
<tr>
<td>Inclusion and exclusion criteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inclusion criteria:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• People undergoing lung cancer screening by computed tomography</td>
<td>• Never-smokers diagnosed with lung cancer</td>
<td></td>
</tr>
<tr>
<td>• Biomarkers in body fluids used as intervention</td>
<td>• Biomarker in body fluids and breath used as intervention</td>
<td></td>
</tr>
<tr>
<td>• Improved participant selection for lung cancer screening, or improved nodule management after screen detection</td>
<td>• Early-stage lung cancer detection</td>
<td></td>
</tr>
<tr>
<td>• English-language articles</td>
<td>• English-language articles</td>
<td></td>
</tr>
<tr>
<td>Exclusion criteria:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Non-human studies</td>
<td>• Non-human studies</td>
<td></td>
</tr>
<tr>
<td>• Metastatic lung cancers</td>
<td>• Metastatic lung cancers</td>
<td></td>
</tr>
<tr>
<td>• Editorial, reviews, and case reports</td>
<td>• Editorial, reviews, and case reports</td>
<td></td>
</tr>
<tr>
<td>• No full text available</td>
<td>• No full-text available</td>
<td></td>
</tr>
<tr>
<td>Selection process</td>
<td>Studies obtained from our search strategy were imported into systematic review management software “Covidence”. After removing duplicates, one researcher (Jazmin Guayco Sigcha) evaluated relevance of all titles and abstract following the inclusion criteria for this review</td>
<td>Studies obtained from our search strategy were imported into systematic review management software “Covidence”. After removing duplicates, two researchers (Edward Stephens and Kenneth Lopez-Loo) independently evaluated relevance of all titles and abstract following the inclusion criteria for this review</td>
</tr>
</tbody>
</table>

“lung cancer”, “biomarker”, and “CT screening”, as well as “never-smoking”, “lung cancer”, and “biomarkers” on MEDLINE, CINAHL, Embase, and Web of Science databases.

Eligibility criteria and study selection

Identified publications were de-duplicated and assessed for relevance based on title and abstract screening. Full texts of potentially relevant articles were retrieved for final eligibility check followed by data extraction (Covidence systematic review software, Melbourne). The NHMRC Evidence Hierarchy and the NIH Early Detection Research Network Five-Phase Approach were used to rank the level of evidence and development phase of the identified biomarkers. Biomarkers from all non- and minimally-invasive biological sources were eligible (Figure 1).

Results

Our literature search yielded 1,433 and 958 articles for LCS and LCINS, respectively. For LCS, our literature search yielded 1,433 articles: 569 from MEDLINE, 141 from CINAHL, 436 from Embase and 287 from Web of Science. For LCINS, our literature resulted on 958 articles: 457 from MEDLINE, 54 from CINAHL, 215 from Embase and 232 from Web of Science. After removing duplicates, articles that do not cover the topic of body-fluid biomarkers for LCS or circulating biomarkers in never-smokers were excluded. Following screening, these were reduced to 22 and 91 total studies, respectively (Figures S1,S2). Table 2 comprises a list of the biomarkers included in this review.
Table 2 Summary of advantages and disadvantages of biomarkers for LCS and in never-smokers

<table>
<thead>
<tr>
<th>Biomarker Source</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Development phase</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleic acid markers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA markers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microsatellite instability/loss of heterozygosity and plasma DNA</td>
<td>Sputum/plasma</td>
<td>High sensitivity, minimally invasive</td>
<td>Moderate specificity</td>
<td>Phase 3 (35)</td>
</tr>
<tr>
<td>DNA methylation</td>
<td>Sputum</td>
<td>High sensitivity, minimally invasive</td>
<td>Low sensitivity</td>
<td>Phase 3 (36)</td>
</tr>
<tr>
<td>MicroRNA markers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MicroRNA-155</td>
<td>Plasma</td>
<td>High sensitivity and specificity, minimally invasive, detects NSCLC at early-stage</td>
<td>No large-scale validation</td>
<td>Phase 3 (37-41)</td>
</tr>
<tr>
<td>Multiple microRNA panels</td>
<td>Plasma/serum</td>
<td>High sensitivity, stable in serum in plasma, specific to lung cancer, minimally invasive</td>
<td>No large-scale validation</td>
<td>Phase 4 (42-45)</td>
</tr>
<tr>
<td>Protein markers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteopontin</td>
<td>Plasma</td>
<td>Can be used for both risk and nodule stratification, minimally invasive</td>
<td>Not unique to lung cancer, evidence is still in early stage (control-case)</td>
<td>Phase 2 (46)</td>
</tr>
<tr>
<td>Carcinoembryonic antigen</td>
<td>Serum</td>
<td>Minimally invasive, high sensitivity</td>
<td>Low sensitivity</td>
<td>Phase 2 (47,48)</td>
</tr>
<tr>
<td>Combination proteins</td>
<td>Serum</td>
<td>Can be used for both risk and nodule stratification, cost-effective, minimally invasive</td>
<td>Overlap with other cancers and inflammatory diseases</td>
<td>Phase 2 (49,50)</td>
</tr>
<tr>
<td>Complement fragments</td>
<td>Plasma</td>
<td>Minimally invasive</td>
<td>Discrepancies in results between studies, quantification of C4d is not standardised</td>
<td>Phase 3 (51-53)</td>
</tr>
<tr>
<td>T-cell receptors</td>
<td>Peripheral blood</td>
<td>Stable in pathological subtypes, minimally invasive</td>
<td>Only studied in stage I lung cancer, limited to Asian population</td>
<td>Phase 3 (54)</td>
</tr>
<tr>
<td>Autoantibodies</td>
<td>Peripheral blood</td>
<td>Panel compensate for cancer heterogeneity, assessed in randomized control trial, minimally invasive</td>
<td>Moderate sensitivity</td>
<td>Phase 5 (55)</td>
</tr>
<tr>
<td>Metabolites</td>
<td>Serum</td>
<td>Panel has high sensitivity and specificity, minimally invasive</td>
<td>Biomarkers are still in exploratory stage</td>
<td>Phase 4 (56-58)</td>
</tr>
<tr>
<td>Circulating lipids</td>
<td>Serum</td>
<td>High sensitivity and specificity, minimally invasive, detects NSCLC at early-stage</td>
<td>Confounded by systemic pathology, validated in female only cohort</td>
<td>Phase 3 (59)</td>
</tr>
<tr>
<td>Cellular markers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circulating tumour cells</td>
<td>Bloodstream</td>
<td>High detection rate by CellCollector, minimally invasive</td>
<td>Confounded by systemic pathology, non-specific</td>
<td>Phase 4 (60,61)</td>
</tr>
<tr>
<td>Circulating inflammatory cells</td>
<td>Bloodstream</td>
<td>Minimally invasive</td>
<td>Confounded by systemic pathology, non-specific</td>
<td>Phase 3 (62,63)</td>
</tr>
</tbody>
</table>

The NIH Early Detection Research Network Five-Phase Approach was used to rank biomarker development. NSCLC, non-small cell lung cancer; LCS, lung cancer screening.
Although we searched for studies without discrimination based on sample type (Table 1), our review yielded blood- and sputum-borne markers only. We discuss our findings in the following paragraphs according to biomarker type.

**Nucleic acid markers of lung cancer**

Gradual accumulation of genetic and epigenetic changes in the cell nucleus can be used to detect lung cancer formation, progression, and metastasis (64). While the disease is primarily driven by somatic alterations, typically linked to smoking exposure, germline mutations could also predispose individuals to lung cancer development (65,66). Emerging as promising biomarkers, nucleic acid markers for lung cancer such as cell-free DNA (cfDNA) and circulating RNA are significantly advancing lung cancer diagnosis through the immense potential of liquid biopsy detection methods. The advent of real-time polymerase chain reaction (PCR) and next-generation sequencing has enhanced the sensitivity and specificity of circulating nucleic acid analysis, making it a valuable asset in the arsenal of lung cancer detection methods (67,68).

**DNA**

**Microsatellite instability/loss of heterozygosity and cfDNA**

Lung cancer contains highly altered chromosomal regions, strongly influenced by smoking habits, although not all are specific to lung cancer (69-71). Furthermore, cfDNA concentration in individuals with lung cancer is usually several times higher than in healthy individuals, most likely due to necrosis/apoptosis of cancerous tissue or circulating cancer cells (72). The ITALUNG biomarker panel measured microsatellite instability/loss of heterozygosity and cfDNA in plasma and retrospectively discriminated lung cancer with high sensitivity (90%) and moderate specificity (62%) in 154 individuals with screen-detected cancers and 486 screening controls (35).

When circulating DNA (cirDNA) is detected in the blood, it typically exhibits a distinctive size profile due to its high degree of fragmentation (73,74). This emerging field of research is known as fragmentomics and has recently gained attention for its applicability in pan-cancer screening, and involves the analysis of both transcriptional and topological features of cirDNA (75). Corroborating the utility of this approach, a study using a five circulating cfDNA feature fragmentomic model successfully distinguished between healthy participants and patients with lung cancer across three validation cohorts with high sensitivity (91.4%, 84.7%, and 92.5%) and specificity (95.7%, 98.6%, and 94.2%) in all three groups; additionally, the model showed sensitivity of 83.2% when identifying stage I lung cancer (76). While promising, to further build on these findings, validation in asymptomatic screening cohorts is required.

A different approach to cancer screening are multi-cancer early detection (MCED) tests, which aim to detect different types of cancer at an early stage for further diagnosis. Numerous MCED tests have successfully discriminated multiple cancers from healthy individuals, but most of these findings have been reported in clinical rather than screening population (77). In an exclusively female screening program, the CancerSEEK test—which used a multi-analyte gene mutation and protein panel—demonstrated high specificity (99.6%) when discriminating 9 targeted types of cancer, including lung cancer, from healthy controls when used in conjunction with positron emission tomography (PET)-CT (78). Still, participants enrolled on MCED studies are not reflective of LCS populations. Furthermore, the inability to determine the type of cancer in patients who have tested positive for MCED may lead to difficult clinical follow-ups. Future studies should focus on determining the feasibility of these tests in LCS populations alone by assessing their diagnostic performance in asymptomatic individuals at a high risk of developing lung cancer.

**Methylation**

Among epigenetic changes, DNA hypomethylation and hypermethylation of specific 5'-C-phosphate-G-3 (CpG)-rich regions in the promoter region of tumour suppressor genes are early events in carcinogenesis, making them markers of interest for early lung cancer detection (79-81). In the context of LCS, a retrospective study analysing DNA hypermethylation in sputum found that a panel of Ras association domain-containing protein 1 (RASSF1), 3-O-sulphotransferase 2 and PR/SET Domain 14 hypermethylation was able to discriminate high-risk screened individuals with (n=65) and without (n=99) lung cancer with specificity of 90%, although it showed a sensitivity of just 28%. When focusing on individual biomarkers, RASSF1 hypermethylation demonstrated the best diagnostic performance with 93% specificity but only 17% sensitivity (36). Further, MCED biomarkers assessed in the PanSeer screening program—which analysed 11,787 CpG sites across 595 regions in the genome from plasma DNA—successfully discriminated 95% [confidence interval (CI): 89–98%] of patients with eight different cancers, including lung cancer, from healthy controls (82).
Nevertheless, longitudinal studies are required to validate these results and, as previously stated, future studies need to take into account the specific characteristics of individuals at a high risk of developing lung cancer in order to be of relevance for LCS.

DNA methylation has been linked to environmental exposures and comorbidities other than lung cancer, adding a layer of complexity to its potential use as a biomarker for lung cancer detection. Research has found correlations between methylation patterns and exposure to traffic-related pollutants, polycyclic aromatic hydrocarbons, and particulate matter rich in metals, affecting genes associated with immune responses and other processes (83-85). Moreover, methylation has been tied to medical conditions such as osteoporosis, obesity and chronic obstructive pulmonary disease (COPD), with studies showing distinct methylation profiles in individuals with these diseases compared to healthy controls (86-88). Obesity, in particular, has been associated with alterations in DNA methylation, influencing the likelihood of developing diseases like type 2 diabetes (87). Factors such as the intrauterine environment, physical activity, and diet can also impact both obesity and DNA methylation (89). Therefore, when investigating methylation as a potential biomarker for lung cancer, it is crucial to consider these additional influences.

**MicroRNAs (miRNAs)**

miRNAs are small noncoding RNA molecules that modulate gene activity and are aberrantly expressed in cancer (90). Protected from degradation by encapsulation in exosomes, circulating RNA remains stable in cancer patients, even amid an increase in ribonuclease activity (91). This stability facilitates cancer detection and characterization through microarray technologies or quantitative PCR (92). As a result of their small size and stability, miRNAs can be detected in fluids such as plasma and serum (93) and have been shown to successfully differentiate prostate, colon, and lung cancer (94-96). A retrospective study of a panel consisting of 24 plasma-derived miRNAs, the miRNA signature classifier, showed high sensitivity (87%) and specificity (81%) in 58 screen-detected lung cancers and 594 screening controls (42). Likewise, a prospective study of a panel consisting of 34 serum-derived miRNAs demonstrated high sensitivity (71%) and specificity (90%) when discriminating 34 screen-detected lung cancers from 30 screening controls (43). Refinement of this panel resulted in a panel of 13-miRNAs with high sensitivity (77.8%) and specificity (74.8%) when discriminating 48 screen-detected cancers from 1,067 screening controls (44). Further, a signature composed of 15 miRNAs was not only able to discriminate 16 screen-detected cancers from 10 screening controls with high sensitivity (80%) and specificity (90%), but could also predict risk of lung cancer, up to 8–9 months before conventional diagnostic methods (45). High sensitivity and specificity shown in these studies illustrate the potential of miRNAs in LCS, but future studies need to validate the diagnostic performance in different populations.

**miR-155**

While the aforementioned miRNA panels successfully discriminated screen-detected cancers from healthy controls, never-smokers were omitted from these cohorts. A potential miRNA candidate for early cancer detection in never-smokers is miR-155, an miRNA known to function as a mediator of immune response and the immune system, however, when dysregulated, is widely implicated in immune-centric diseases including chronic inflammation, auto-immunity, and cancer, among others (37). A 2014 study identified the up-regulation of miR-155 in plasma to be predictive of early stage (stage I and II) LCINS (n=37) compared to healthy controls (n=60) with a sensitivity of 91%, a specificity of 93% (38). Further, studies investigating the role of miR-155 in non-small cell lung cancer (NSCLC) and chronic obstructive pulmonary disease (COPD), with studies showing distinct methylation profiles in individuals with these diseases compared to healthy controls (86-88). Obesity, in particular, has been associated with alterations in DNA methylation, influencing the likelihood of developing diseases like type 2 diabetes (87). Factors such as the intrauterine environment, physical activity, and diet can also impact both obesity and DNA methylation (89). Therefore, when investigating methylation as a potential biomarker for lung cancer, it is crucial to consider these additional influences.

**Protein markers of lung cancer**

Proteins mediate homeostatic as well as pathological processes and are potential biomarkers (97). Cancer-induced aberrations result in a marked difference in protein translation and expression between normal and cancer cells (98,99). Furthermore, monitoring of cancer-related metabolites in blood has been an emerging approach to detect different malignancies in recent years (100).

**Osteopontin (OPN)**

The secreted phosphoprotein 1 gene (SPP1) encodes the protein OPN. SPP1 is upregulated in the cancer tissue of early-stage and relapsed NSCLC individuals as well as in LCINS (101-105). OPN is of interest as a lung cancer biomarker as its activity leads to numerous downstream
processes associated with cancer progression and cellular transformation (106) by promoting cell migration, invasion, and adhesion, as well as playing an important role in immune cell recruitment, wound healing, and tissue remodelling (107-109). In a case-control study, change-over time of plasma OPN levels (OPN velocity or OPNV) were linked to a statistically significant increase in lung cancer risk [area under the curve (AUC) of 0.88] in 10 screen-detected incident cancers matched to 1–4 screening controls each. Furthermore, when analysis was limited to cancer cases and controls presenting ground glass opacities or stable solitary nodules, OPNV could differentiate between malignant and benign nodules with an AUC of 0.91 (46). Still, validation studies using larger LCS cohorts are necessary to confirm the diagnostic value of plasma OPN.

**Carcinoembryonic antigen (CEA) and combination markers**

CEA is a glycoprotein produced during foetal development and is usually absent in healthy adults (110); however, it is widely reported as a marker of lung cancer with potential as a predictor of disease prognosis. A South-East Asian study investigating the relationship between serum biomarkers and residential radon levels in never and former (>15 years) smokers, described a significant increase in serum CEA and cytokeratin 19 fragment (CYFRA21-1) in individuals with lung cancer, compared to healthy controls with high and low radon exposure. Interestingly, an increase in CEA (P=0.009) and CYFRA21-1 (P=0.0031) was also observed in healthy controls with high radon exposure when compared to low, potentially indicating high serum CEA as a biomarker for lung cancer development in never-smokers. Receiver operating characteristic analyses of CEA and CYFRA21-1 for diagnosing lung cancer illustrate high specificity (98% and 94% respectively) but inadequate sensitivity (57.3% and 58.6% respectively), which similarly has been reported in other studies investigating CEA as a biomarker of NSCLC and mutational status (48,111,112).

CEA has also been assessed in a panel with other protein markers such as CYFRA21-1, progastrin-releasing peptide (ProGRP), and squamous cell carcinoma (SCC) antigen. CYFRA21-1 is abundant in pulmonary tissue, with concentrations particularly elevated in SCC where it correlates with tumour size, lymph node status and cancer stage (113,114). In a prospective cohort of older smokers (n=634), participants with a positive or indeterminate CT scan result (n=92) were tested for CEA and CYFRA21-1. In the 17 cases with screened-lung cancer, positivity in either biomarker yielded a higher AUC than CT alone (CEA 0.75; CEA/CYFRA21-1 0.76; CT 0.68) (50). In a different prospective case-control study, ProGRP, SCC, CEA, and CYFRA21-1 showed acceptable discrimination (AUC =0.719) between screen-detected lung cancers and screening controls (n=715; China). However, when analysis was limited to cancer cases and controls presenting benign nodules, discrimination was very limited in the validation data set (AUC =0.5836) (49). When considering never-smokers, the combination of CEA and CYFRA21-1 may be more clinically informative than when assessed independently, however further research is required to evaluate the efficacy of this in a clinical setting (47,111,115-117).

**Immunological markers**

The complement system is a central component of innate immunity and plays an essential role in immune surveillance and homeostasis. The C4d split-product, is a passive indicator of complement pathway activation that is found in higher concentrations in biological fluids of individuals with lung cancer (118). In an independent cohort study, plasma C4d levels were linked to increased lung cancer risk in 32 screen-detected incident cancers compared to 158 screening controls (AUC =0.735) (51), however, its use as a marker for LCS was not validated in a study of 20 screen-detected incident cancers matched to two screening controls each [odds ratio (OR) =1.53; 95% CI: 0.93–2.51; P=0.079] (52). Unlike C4d, C4c is released to the extracellular milieu upon fragmentation during activation of the classical pathway of complement, thus making it more detectable in bodily fluids, such as plasma (119). In a case-control study the panel of plasma C4c plus CYFRA 21-1 and C-reactive protein discriminated screen-detected cancers (n=32) from screening controls (n=93) with a sensitivity of 73% and specificity of 70% (53).

**T-cell receptors (TCRs)**

A prior investigation found that TCRs present in the bloodstream can differentiate between individuals with lung cancer and healthy controls (120). In the context of LCS, a retrospective study indicated that lung cancer-associated TCRs in peripheral blood could distinguish stage I screened-cancers (n=52) from screening controls (n=94) with high sensitivity (72%) and specificity (91%). In addition, the sensitivity was stable in pathological subtypes, being 73% in lung SCC and 71% in lung adenocarcinoma (54).
Autoantibodies
Autoantibodies develop in response to an abnormal cancer antigen in individuals with lung cancer, often well before symptoms arise or screening detection is possible (121,122). Autoantibodies are usually absent or found in low levels not only in those without cancer but also in many individuals with the disease, thus they are likely to be specific but not sensitive (123). A randomised trial evaluating the EarlyCDT-Lung test for up to 2 years indicated a high specificity (90.4%) but low sensitivity (32.1%) for detecting lung cancer in 127 screen-detected cancers and 11,610 controls (55). Tumour-induced suppression of immune responses can induce a reduction of autoantibody production and detection, which would explain the low sensitivity of the test at 2 years (124). Still, EarlyCDT-lung test could be used in combination with CT to ensure a high detection rate of stage I/II lung cancer cases. A limitation of this study is the absence of CT scans for EarlyCDT-Lung test negative participants and control arm participants, necessary to evaluate the test’s effectiveness in LCS.

Metabolites
Serum and plasma sampling has revealed metabolites with the power to discriminate individuals with lung cancer from healthy individuals and individuals with benign cancers (125-128). When investigating metabolites used in LCS, one prospective study showed significantly elevated levels of serum metabolites, particularly, L-(+)-glucose, cysteinyl-glutamine, phosphatidylethanolamine (PE) [22:2(13Z,16Z)/15:0] and threonine-glutamine, in 34 individuals with ground glass opacities compared to 39 healthy controls (56). Similarly, a cross-sectional study demonstrated that PE (34:2), PE (36:2) and PE (38:4) have modest accuracy (69%, 71% and 67%, respectively) when discriminating screen-detected lung cancers (n=29) from screen-detected benign nodules (n=25); though additional analyses are required to validate the diagnostic value of the compounds discovered in this study (57). When examining panels, a classifier including nine serum metabolites allowed the discrimination of 31 screen-detected lung cancers from 92 matched screening controls with 100% sensitivity and 95% specificity (58).

Lipid profiling
Carcinogenesis disturbs normal lipid metabolism, thus making lipidomics a promising tool for early identification of lung and other cancers (59,129). A proposed marker for the early detection of LCINS is a combinational serum marker of three lipid products of fatty acid (FA) metabolism; FA (20:4), FA (22:0), and lysophosphatidylethanolamine (20:4). Together, these markers successfully distinguished never-smoking females with NSCLC (discovery set n=39, validation set n=25) from healthy controls (discovery set n=46, validation set n=17) with high sensitivity (discovery set 0.949, validation set 1.000) and specificity (discovery set 1.000, validation set 1.000), including those with early-stage disease (59). In addition to this, lipid metabolites have also been identified as potential markers for lung cancer detection in smoking-related NSCLC however these findings have also not been validated in larger cohort studies (130,131).

Cellular markers of lung cancer
Cellular changes promoted by cancer development can be assessed using various markers routinely measured in common blood tests or as ratios derived from these measurements (132). The association between blood cell ratios related to systemic inflammation and cancer risk shows potential as a biomarker for earlier identification of the disease.

Circulating tumour cells (CTCs)
CTCs are metastatic tumour cells that have emerged from the primary tumour site into the bloodstream to form secondary tumours at distinct sites (133). These cells present notable advantages for use as biomarkers as they can be sourced minimally-invasively and provide diagnostic and prognostic information (134,135). CTC collection is not standardised but uses minimally- or non-invasive methods that avoid tissue biopsy (136,137). The isolation by size of epithelial tumour cells (ISET) test uses fixed blood samples and vertical filtration to capture rare cells and CTCs, thus preserving the integrity of the cells for subsequent analysis (138); as a result, many clinical investigations have opted to employ ISET technology for CTC isolation (139). However, a prospective cohort study evaluating the performance of CTCs isolated by ISET test in a high-risk population with COPD found the sensitivity (26.3%) was too low for LCS (61). This study deliberately chose individuals with COPD, as this population is considered at a high risk of developing lung cancer independent of cigarette smoking and so it should be of interest for LCS (140,141). On the other hand, CellCollector, an in-vivo isolation method based on epithelial-cell adhesion molecule (EpCAM) recognition, was found to be able to
Lymphocyte, and platelet counts are nonspecific parameters for lung cancer and may be influenced by concurrent comorbidities (152). Furthermore, considering that both COPD and the extent of airway obstruction are linked to elevated levels of systemic inflammation, future research must take these variables into account while evaluating the differences in inflammation markers between cancer and non-cancer individuals (153,154).

Conclusions

Lung cancer is the leading cause of cancer mortality worldwide largely as a result of late-stage diagnoses. To address this, improvements must be made to current screening measures to identify cancer early, as to allow for the best possible outcomes for the patient. Biomarkers are playing an emerging role in the early detection and management of lung cancer, with applications in screening and detection.

We identified numerous biomarkers for LCS with varying levels of supporting evidence. Metabolites have shown the highest sensitivity and specificity to lung cancer when compared to healthy controls, however, abnormal circulating metabolites are known to be indicative of systemic pathologies and may not be lung cancer specific. Similarly, cellular markers of disease such PLR and SII also occur as result of comorbid disease. While valuable, these markers are likely not applicable for specific LCS strategies for early cancer detection. As such, it is likely that a panel of tumour-specific circulating proteins or nucleic acids would be the most viable for targeted screening strategies. We described recent evidence around miRNA panels for LCS; while the evidence for these markers was not as strong, these markers are less likely to be confounded by other systemic pathologies and, as such, are likely would make more effective markers for LCS. Further, fragmentomics is an emerging area with great potential for biomarker discovery due to its ability to identify numerous tumour-derived changes in cirDNA. However, current detection models lack clinical validation. Despite this, circulating cfDNA remains an attractive target for early lung cancer detection.

LCINS is becoming increasingly prevalent, appearing to be clinicopathologically distinct to smoking-related lung cancers, and therefore unique approaches to its management may be required. Of the markers identified, lipid panels demonstrated the highest sensitivity and specificity when discriminating LCINS compared to healthy controls, however, abnormal lipid levels are known to be indicative of systemic pathologies and may not be lung cancer specific.
Thus, nucleic acids are likely the most viable biomarker for LCINS, the most efficacious of which is miR-155. Evidence suggests miR-155 may be a valid pan-NSCLC marker however confirmation in large, randomised studies is required.

Several markers in various stages of development are currently available for LCS and LCINS, and further advancement in terms of external validation and impact assessment is in progress. Randomised trials are considered to be the gold standard for external validation (155). However, proving ultimate evidence of mortality benefits is challenging and may take a significant amount of time. As a result, more time- and cost-effective models are increasingly being used to complement clinical decision-making with the aim of improving patient outcomes (156,157). Such models have already been implemented to compare the effectiveness of certain biomarkers in LCS (158).

To conclude, significant advancements have been made in the field of lung cancer biomarker research, with numerous biomarkers of lung cancer displaying varying levels of clinical efficacy and showing improvement in diagnostic accuracy over standard clinical workflow in LCS and LCINS. The priority now should be the validation of existing candidate markers in appropriate clinical contexts to integrate these into clinical practice. To do this, randomized controlled trials or similar methods of validation should be designed to test the efficacy of these biomarkers. This will positively impact lung cancer diagnosis and treatment, and help to reduced lung cancer mortality worldwide.

**Acknowledgments**

We would like to acknowledge The Prince Charles Hospital Library team for performing the literature search of MEDLINE, CINAHL, Embase, and Web of Science databases.

**Funding:** None.

**Footnote**

**Reporting Checklist:** The authors have completed the Narrative Review reporting checklist. Available at https://tlcr.amgroups.com/article/view/10.21037/tlcr-23-291/rc

**Peer Review File:** Available at https://tlcr.amgroups.com/article/view/10.21037/tlcr-23-291/prf

**Conflicts of Interest:** All authors have completed the ICMJE uniform disclosure form (available at https://tlcr.amgroups.com/article/view/10.21037/tlcr-23-291/coif). KMF serves as an unpaid editorial board member of *Translational Lung Cancer Research* from August 2023 to July 2025. KMF has received competitive Research Grant Funding from MRFF and NHMRC in the past 36 months, honoraria or royalties from UpToDate and Cochrane Clinical Answers, travel support funding from WCLC, ATS, PCCP, and in kind equipment funding from Olympus and Mevis Veolity. KMF is the Former President of the Asia Pacific Society of Respirology. The other 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**


65. Roland M, Rudd RM. Genetics and pulmonary medicine.


106. Dall'Olio FG, Abbatì F, Facchinetti F, et al. CEA and


## Table S1 MEDLINE complete (EbscoHOST) search terms and results for biomarkers in LCS

<table>
<thead>
<tr>
<th>#</th>
<th>Query</th>
<th>Limiters/expanders</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>S10</td>
<td>S3 AND S6 AND S9</td>
<td>Limiters—English language; expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>569</td>
</tr>
<tr>
<td>S9</td>
<td>S7 OR S8</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>580,144</td>
</tr>
<tr>
<td>S8</td>
<td>TI (&quot;computed tomography&quot; OR &quot;ct&quot; OR cat) N2 (screen* OR imag* OR scan*)) OR &quot;pet<em>ct&quot;) OR AB (&quot;computed tomography&quot; OR &quot;ct&quot; OR cat) N2 (screen</em> OR imag* OR scan*)) OR &quot;pet*ct&quot;)</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>217,001</td>
</tr>
<tr>
<td>S7</td>
<td>(MH &quot;Tomography Scanners, X-Ray Computed&quot;) OR (MH &quot;Tomography, X-Ray Computed+&quot;)</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>481,270</td>
</tr>
<tr>
<td>S6</td>
<td>S4 OR S5</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>872,063</td>
</tr>
<tr>
<td>S5</td>
<td>TI (marker* OR biomarker* OR genom*)</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>427,584</td>
</tr>
<tr>
<td>S4</td>
<td>(MM &quot;Biomarkers, Tumor&quot;) OR (MM &quot;Genome+&quot;) OR (MM &quot;Whole Genome Sequencing+&quot;) OR (MM &quot;Genome-Wide Association Study&quot;)</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>562,561</td>
</tr>
<tr>
<td>S3</td>
<td>S1 OR S2</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>263,304</td>
</tr>
<tr>
<td>S2</td>
<td>TI (&quot;lung OR pulmonary OR bronch&quot;) N2 (cancer* OR tumor* OR tumour* OR lymphoma* OR sarcoma* OR neoplasm* OR carcinoma* OR adenocarcinoma* OR adenoma* OR blastoma* OR malignanc*)) OR mesothelioma)</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>194,619</td>
</tr>
<tr>
<td>S1</td>
<td>(MM &quot;Lung Neoplasms+&quot;)</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>213,399</td>
</tr>
<tr>
<td>#</td>
<td>Query</td>
<td>Limiters/expanders</td>
<td>Results</td>
</tr>
<tr>
<td>----</td>
<td>-----------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>S11</td>
<td>S3 AND S6 AND S9</td>
<td>Limiters—English language; expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>457</td>
</tr>
<tr>
<td>S10</td>
<td>S3 AND S6 AND S9</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>479</td>
</tr>
<tr>
<td>S9</td>
<td>S7 OR S8</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>41,445</td>
</tr>
<tr>
<td>S8</td>
<td>TI (nonsmoker* or “non-smoker”* or “never smoker”<em>) OR AB (nonsmoker</em> or “non-smoker”* or “never smoker”*)</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>41,346</td>
</tr>
<tr>
<td>S7</td>
<td>(MH “Non-Smokers”)</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>634</td>
</tr>
<tr>
<td>S6</td>
<td>S4 OR S5</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>871,839</td>
</tr>
<tr>
<td>S5</td>
<td>TI (marker* OR biomarker* OR genom*)</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>427,393</td>
</tr>
<tr>
<td>S4</td>
<td>(MM “Biomarkers, Tumor”) OR (MM “Genome+”) OR (MM “Whole Genome Sequencing+”) OR (MM “Genome-Wide Association Study”)</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>562,504</td>
</tr>
<tr>
<td>S3</td>
<td>S1 OR S2</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>263,179</td>
</tr>
<tr>
<td>S2</td>
<td>TI (((lung OR pulmonary OR branch*) N2 (cancer* OR tumor* OR tumour* OR lymphoma* OR sarcoma* OR neoplasm* OR carcinoma* OR adenocarcinoma* OR adenoma* OR blastoma* OR malignance*)) OR mesothelioma)</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>194,532</td>
</tr>
<tr>
<td>S1</td>
<td>(MM “Lung Neoplasms+”)</td>
<td>Expanders—apply related words, apply equivalent subjects; search modes—find all my search terms</td>
<td>213,269</td>
</tr>
</tbody>
</table>
**Figure S1** Flow chart indicating lung cancer screening article inclusion and exclusion with reasons for exclusion.

**Figure S2** Flow chart indicating lung cancer in never-smoker article inclusion and exclusion with reasons for exclusion.