Identification of biomarkers for the early detection of non-small cell lung cancer: a systematic review and meta-analysis

Eithar Mohamed[1], Daniel J. García Martínez[2], Mohammad-Salar Hosseini[3], Si Qi Yoong[4], Simon Hart[5] & Barbara-ann Guinn[1*]

[1]Centre for Biomedicine, Hull York Medical School, University of Hull, HU6 7RX, UK

[2]Department of Biotechnology, Pozuelo de Alarcón, University Francisco De Vitoria, Madrid, 28223, Spain.

[3]Research Centre for Evidence-Based Medicine, Tabriz University of Medical Sciences, Tabriz, Iran.
[4]Alice Lee Centre for Nursing Studies, Yong Yoo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore.
[5]Respiratory Medicine, Hull York Medical School, University of Hull, HU6 7RX, UK

*Correspondence: Dr Barbara Guinn, Centre for Biomedicine, Hull York Medical School, Cottingham Road, University of Hull, Hull, HU6 7RX, UK; e-mail: Barbara.Guinn@hyms.ac.uk;

Highlights:

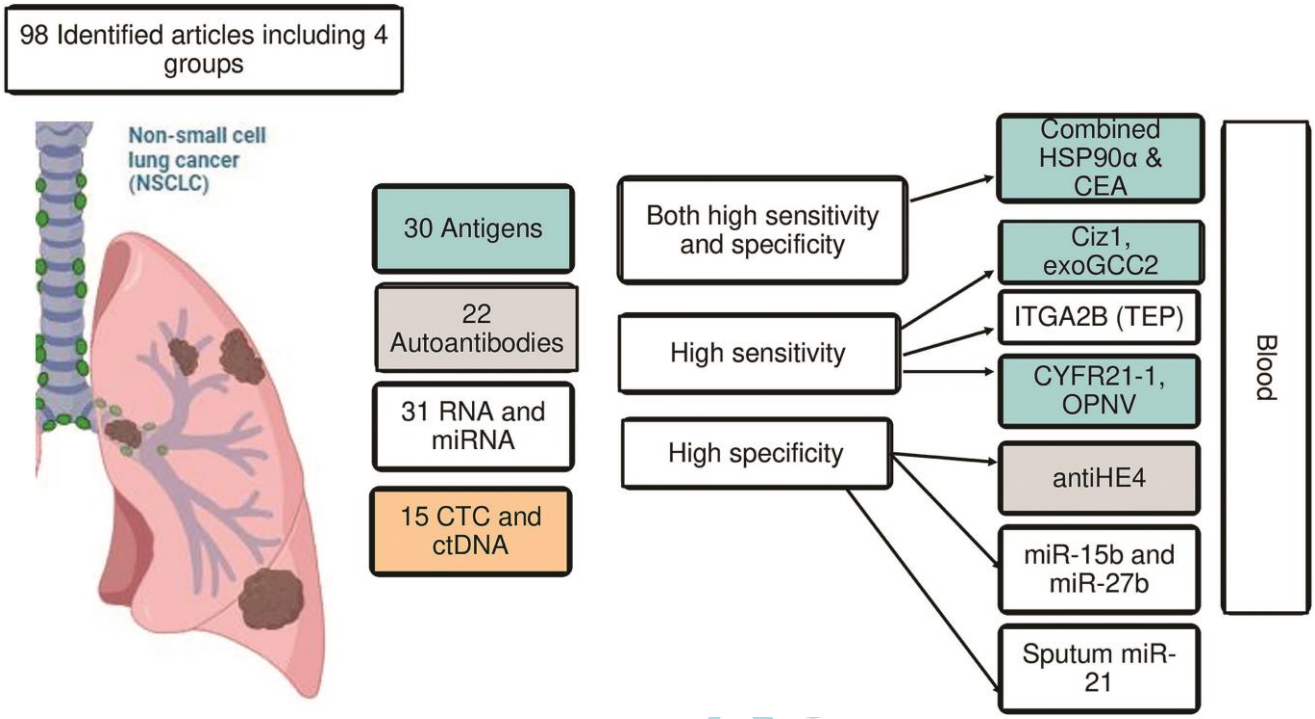
- Identification of 98 articles that found biomarkers for early lung cancer.
- Pooled area under curve of 0.86 indicated an excellent diagnostic performance.
- Four types of biomarkers were identified - antigens, autoantibodies, RNAs and circulating DNA.
- Biomarkers with high sensitivities/specificities can improve early detection.

ABSTRACT

Lung cancer (LC) causes few symptoms in the earliest stages, leading to one of the highest mortality rates among cancers. Low-dose computerised tomography (LDCT) is used to screen high-risk individuals, reducing the mortality rate by 20%. However, LDCT results in a high number of false positives and is associated with unnecessary follow-up and cost. Biomarkers with high sensitivities and specificities could assist in the early detection of LC, especially in patients with high-risk features. Carcinoembryonic antigen (CEA), cytokeratin 19 fragments, and cancer antigen 125 have been found to be highly expressed during the later stages of LC but have low sensitivity in the earliest stages. We determined the best biomarkers for the early diagnosis of LC, using a systematic review of eight databases. We identified 98 articles that focused on the identification and assessment of diagnostic biomarkers and achieved a pooled area under curve of 0.85 (95% CI 0.82-0.088), indicating that the diagnostic performance of these biomarkers when combined was excellent. Of the studies, 30 focussed on single/antigen panels, 22 on autoantibodies, 31 on miRNA and RNA panels, and 15 suggested the use of circulating DNA combined with CEA or NSE for early LC detection. Verification of blood biomarkers with high sensitivities (Ciz1, exoGCC2, ITGA2B), high specificities (CYFR21-1, antiHE4, OPNV), or both (HSP90α, CEA) along with miR-15b and miR-27b/miR-21 from sputum may improve early lung cancer detection. Further assessment is needed using appropriate sample sizes, control groups that include patients with non-malignant conditions, and standardised cut-off levels for each biomarker.

Keywords: non-small cell lung cancer (NSCLC), blood/serum/plasma biomarkers, early diagnosis, systematic literature review

Graphical abstract



Abbreviations:

ADC	adenocarcinoma
AUC	area under curve
CEA	carcinoembryonic antigen
CI	confidence interval
CIZ1	CDKN1A Interacting Zinc Finger Protein 1
COPD	chronic obstructive pulmonary disease
CRP	c-reactive protein
CT	computer tomography
CTA	cancer-testis antigen
CTC	circulating tumour cells
CYFRA21-1	cytokeratin 19 fragments
GCC2	GRIP And Coiled-Coil Domain Containing 2 protein
ITGA2B	integrin alpha 2b (tumour-educated platelets, TEP)
LC	lung cancer
LDCT	low dose computer tomography
MAGE	melanoma-associated antigen gene
NSCLC	non-small cell lung carcinoma
NSE	neuron-specific enolase
OPNV	osteopontin velocity
PET	positron emission tomography
PRISMA	preferred reporting items for systematic review
PROSPERO	prospective register of systematic review
ROB	risk of bias
SCC	squamous cell carcinoma
TA	tumour antigens
TAAb	tumour-associated antibodies
QUADAS	quality assessment of diagnostic accuracy studies

1. BACKGROUND

Lung cancer (LC) is one of leading causes of cancer-related mortality. It is the second most common cancer, accounting for 18.6% of all tumours [1, 2] and affecting both genders with an annual incidence of two million worldwide. There are two main types of LC: small cell LC (SCLC) and non-small cell LC (NSCLC), the latter accounting for >80% of all LC cases. Two-thirds of patients are diagnosed at an advanced stage of disease, when surgical options are not recommended. Smoking is a major risk factor for LC and is associated with 80% of LC-associated mortality [3]. Thus, late detection and poor prognosis in LC makes disease management challenging. The 5-year survival rate can be raised to 80% if diagnosed at an early stage, but it has remained stubbornly low at around 15% [1].

The common diagnostic methods of LC [4] include medical imaging computer tomography (CT), Positron Emission Tomography (PET), and fluorescence bronchoscopy as well as many biochemical and histological assays such as sputum and pleural cytology, and polymerase chain reaction (PCR). Although low dose (LD) CT reduces LC mortality, it increases the number of nodules identified which are associated with a high frequency of false-positives (up to 95%) due to the limited indicators of their propensity to become malignant. This time and cost implications including patients' psychological distress and exposure to radiation associated with imaging techniques.

A number of serological markers have been investigated for their ability to provide a LC diagnosis including cytokeratin 19 fragments (CYFRA21-1), carcinoembryonic antigen (CEA) and neuron-specific enolase (NSE) [5-7]. CYFRA21-1 is the most sensitive biomarker for NSCLC, especially squamous cell tumours. However, CYFRA21-1 is also highly expressed in gastrointestinal, urological and gynaecological tumours, and in low amounts in some benign diseases [8, 9], giving it a low specificity for NSCLC. CEA is also expressed in the foetal gastrointestinal epithelium, pancreas and liver in low concentrations [10]. It is already used as a diagnostic tumour marker in colon cancer but is also highly expressed in adenocarcinomas such as gastric and pancreatic cancer [10, 11]. CEA levels are low in the early stages of SCLC but increase in 40-65% of NSCLC patients in the late stages with metastasis [12]. NSE is a tumour marker of SCLC used in diagnosis, follow-up and prognosis, but it has low sensitivity and specificity [13]. Some reports identified NSE expression in approximately 10-20% of NSCLC patients with an expression associated with tumour burden, number of metastatic sites and treatment response [14].

Tumour antigens (TAs) and tumour associated antibodies (TAAbs) formed through immune responses against LC could be identifiable before the onset of symptoms. Thus, TAAbs may be valuable for the early diagnosis of LC as they are stable and persist in serum for a long time compared to TAs. The Early Cancer Detection Test (CDT)-Lung was developed by Oncimmune Inc and focusses on two TAAbs panels. A panel of six or seven AAbs were found to have sensitivities of 83% and 91% and specificities of 46% and 37%, respectively [15]. Yang et al. reviewed the use of autoantibodies as an early detection tool for LC diagnosis [16]. Among those autoantibodies considered, the panel of p53, PGP9.5, SOX2, GAGE7, GBU4-5, CAGE and melanoma-associated antigen gene (MAGE)A1 had a sensitivity of 56.4% for the detection of early LC in 397 patients with lung nodules compared with 74 control individuals. This study recommended the panel could be combined with CT for early LC detection as it achieved high specificity of 95.80 % [17].

This systematic review aims to report the diagnostic biomarkers being considered for use in the early detection of NSCLC and includes the analysis of their sensitivities, specificities and area under curves (AUC) receiver operating characteristics (ROC) to help provide a current prioritisation list. This article considers four main groups of studies focussing on antigens, autoantibodies, miRNAs and circulating DNA in blood and sputum that may provide useful non-invasive biomarkers for the earlier detection of NSCLC.

2. METHODS

This review was performed and reported according to the Preferred Reporting Items for Systematic Review (PRISMA) guidelines [18]. The protocol (Supplementary Table I) was registered on the international prospective register of systematic reviews (PROSPERO: CRD42022336488). Papers which did not acquire informed, written consent from all participants in their study were excluded.

2.1 Search strategy

CINAHL, MEDLINE, PubMed, Scopus, Web of Science, Cochrane library and Clinicaltrial.gov were searched from 1 January 1970 until 21st May 2023. Literature searches were performed using the following terms: (cancer* or tumor* or tumour* or neoplasm* or carcinoma* or malignancy*) AND (lung* or pulmonary) AND (antigen* OR protein* OR RNA* OR ctDNA* OR miRNA* OR cell surface marker* OR inflammatory cell*) AND (early detection OR early diagnosis OR early biomarker OR early marker). The initial search, removal of duplicates, title and abstract screening, and full-text reviews were performed by two authors independently.

2.2 Exclusion and inclusion criteria

Exclusion and inclusion criteria are detailed in the protocol (Supplementary Table I). Briefly the inclusion criteria were primary research articles that had studied human adult LC in at least 10 patients. Both retrospective and prospective studies were eligible, including case-control and cohort studies. Biomarkers included single or biomarker panels found in blood, urine, sputum and pleural fluids for LC diagnosis. Cell line and animal studies, as well as prognostic and predictive biomarkers were excluded. Studies that did not report sensitivity and specificity and/or raw data were excluded as well.

2.3 Data extraction

Data extraction was performed by E.M. and D.G.M. using a pre-piloted data extraction form. Information extracted included author, year, country, population comparison groups, specimen type, name of biomarkers, technique used, sensitivity, specificity, AUC for stage I and II NSCLC.

2.4 Risk of bias (ROB)

ROB was performed by two independent researchers using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2). It was based on four domains: participant selection, index test, reference standard and timing and flow. Each prompt in the domain was assessed as "yes" or "no" or "unclear". The first three domains were also assessed for applicability. The study had low ROB/applicability concerns if all domains were rated low, unclear if there was at least one domain rated unclear, and had high ROB/applicability concerns if at least one domain was rated high [19]. No study was excluded based on ROB. Figures were visualized using Robvis [20].

2.5 Synthesis methods

Due to the wide range and combinations of biomarkers assessed and the limited information provided, a diagnostic test meta-analysis and the evaluation of the diagnostic performance of each biomarker was not possible. Instead, we conducted a random effects meta-analysis by pooling the AUC and the corresponding 95% confidence intervals (CI) using Review Manager Version 5.4. A random-effects model was used as it accounted for between-study heterogeneity. Cochran's Q test and I^2 statistics were used to evaluate heterogeneity. Heterogeneity was considered unimportant when I^2=0–40%, moderate when I^2=30–60%, substantial when I^2=50–90% and considerable when I^2=75–100%. Heterogeneity was significant when $p<0.10$ and I^2 value>50%. If there were more than 10 studies in the meta-analysis, heterogeneity was explored using sensitivity and subgroup analysis. Sensitivity analysis was done by excluding studies one by one. If the results remain consistent, they were robust. If results differed, they were treated with caution [21]. Subgroup analysis was performed based on the type of biomarkers investigated [antigens, autoantibodies, miRNA and RNA, ctDNA and circulating tumour cells (CTC)] and the type of control. A statistically significant subgroup effect was defined as $p<0.1$ [22]. If there were more than 10 studies in a meta-analysis, publication bias was assessed by visual inspection of a funnel plot and conducting Egger's regression and Kendalls tau test [23, 24].

If studies did not report AUC and 95% CI, a narrative synthesis of the diagnostic properties of the biomarkers were conducted, by categorising studies based on the type of biomarker assessed and the findings of each study were summarised (sensitivity, specificity and AUC if reported) [25].

3. RESULTS

The database searches identified 7295 articles in total and 2474 duplicates were removed (Supplementary Table II). 4636 articles were excluded based on title and abstract. After evaluating 185 full-texts, 98 articles were included in this systematic review (Figure 1, Table 1). The sample sizes varied across the included studies, ranging from 18 to 1479 LC cases. 30 studies investigated either single antigens or antigen panels and reported sensitivities ranging from 48-95% and most studies investigated blood biomarkers. 22 studies investigated autoantibodies, 31 studies focused on miRNAs and RNA, and 15 studies explored CTCs and ctDNA in early-stage NSCLC. Six studies (Table 2) identified biomarkers that had a sensitivity and specificity of more than 90% but the average specificity and sensitivity of the biomarkers in each group were determined (Table 3). Antigens had the lowest values for both variables and the Standard Deviation (SD), and were not deemed to be the best option as biomarkers of NSCLC based on the literature examined. ctDNA and CTC had the highest values of sensitivity, and a high specificity, with the lowest Standard Deviation of all of the groups suggesting these were the best options for the early (minimally-invasive) detection of NSCLC.

3.1 ROB

86 studies had high ROB (Figure 2; Supplementary Table III) most commonly due to the use of case-control study designs, causing the "patient selection" domain to have high risk of bias. Three other domains also scored poorly, most notably "flow and timing". Applicability concerns were low for all studies. Different methods of detection were used in each study and could have impacted the robustness of the results obtained. However, ELISA was the most common technique applied to analyses of these early LC biomarkers, mainly used for antibodies and antigens, while RT-PCR was used for miRNA and ctDNA detection.

3.2 Synthesis of the results - Meta-analysis

Thirty-one studies reported adequate data to enable the pooling of AUC, and random effects meta-analysis found that the pooled AUC was 0.85 (95% CI 0.82-0.088), indicating that the diagnostic performance of biomarkers for early NSCLC were excellent [30]. However, the heterogeneity was also considerable (I^2 = 96%, $p < 0.00001$). Sensitivity analysis found that the pooled AUC remained consistent, indicating that the results were robust. Subgroup analysis found that there was no significant subgroup difference (I^2 = 51.8%, p = 0.10) based on the type of biomarker used (Figure 3A). Of the four types of biomarkers, pooled AUC for the autoantibodies subgroup was the lowest (pooled AUC = 0.80, 95% CI 0.72-0.88). Subgroup analysis based on the type of control showed that there was a significant subgroup difference in diagnostic performance (I^2 = 71.7%, p = 0.003), and the biomarkers performed the least accurately in differentiating early NSCLC from benign lung diseases (pooled AUC = 0.74, 95% CI 0.67-0.81) (Figure 3B). There was also no significant subgroup difference based on the source of biomarker (I^2 = 0%, p = 0.95), suggesting that the diagnostic performance of biomarkers was similar regardless of the source of biomarker (Figure 3C). The funnel plot appeared asymmetric. Kendalls tau (p = 0.009) and Egger's test (p = 0.003) were significant, indicating that publication bias may be present (Figure 3D). The findings of this meta-analysis should be interpreted with caution as each study investigated different individual biomarkers. There were limited studies with data suitable for meta-analysis, hence this meta-analysis was not representative of all studies encompassed by the

systematic review. However, we provide preliminary evidence that current NSCLC biomarkers can generally be expected to perform well diagnostically.

Taking all data into consideration, the subgroup of miRNA and RNA biomarkers showed the highest specificity (0.91) (Table 3), followed by antigens (0.86), DNA and CTC (0.84), and finally, autoantibodies (0.77), although, results of DNA and CTC showed more statistical robustness as their I^2 value was lower.

3.2.1 Antigens as biomarkers for NSCLC

The most sensitive and specific antigens biomarkers, with values over 90% (Table 2), were discovered by Farlow et al., [26] when they analysed different combinations of biomarkers based on TNF-α, CYFRA 21, interleukin-1ra, MMP-2, monocyte chemotactic protein-1 and sE-selectin achieving 99% sensitivity and 95% specificity. Yang et al [27], Yuan et al [28] and Gasparri et al. [29] used different antigen combinations; but did not achieve the same sensitivity and specificity as Farlow et al. [26], except for the specificity, 99%, in the combination of HSP90α and CEA described by Yuan et al. [28]. Ma et al. [30] showed that malate dehydrogenase 2 (MDH2) was detected in urine with sensitivities of 70.13%, 68.92% and specificities of 66.11%, 58.22% in training and validation cohorts respectively [30].

Jeong et al. [31] demonstrated that the exosomal GRIP And Coiled-Coil Domain Containing 2 (GCC2) could be a biomarker for early NSCLC. When derived from NSCLC exosomes, GCC2 was upregulated, but it has also been found to be increased in other cancers such as those affecting the liver, which are also associated with poor prognosis [32]. However, the isolation of exosomes from the blood is a time-consuming process and requires high quality controls to ensure the purity of these molecules is of the standard required for further analyses [31]. Higgins et al. [33] focussed on the CDKN1A Interacting Zinc Finger Protein 1 (Ciz1), which is a nuclear matrix protein. Its expression was restricted to tumours and not found in normal tissues with a specificity of 71% for early NSCLC and a false positive rate of 50%. This suggested that Ciz1 had a limited capacity to differentiate NSCLC stage among high-risk sub-groups.

Osteopontin (OPN), a secreted phosphoprotein, was increased in NSCLC, however, it could not differentiate benign lung diseases from pulmonary carcinoma probably due to its roles in wound healing and tissue remodelling. Joseph et al [34] found that OPN concentrations in plasma changed as a function of time (OPN velocity; OPNV) and acted as a biomarker for early NSCLC with 80% sensitivity and 88% specificity. Suggesting that OPNV velocity could be useful for the detection of early NSCLC especially in the context of indeterminate nodules. Zhong et al. [35] examined patients with stage I NSCLC versus those patients in high-risk groups for expression of a biomarker panel that included paxillin (PXN), SEC15L2, BAC clone RP11-499F19, XRCC5 and MALAT1. They demonstrated a sensitivity of 100% and specificity of 95.7% in the training cohort, and a sensitivity and specificity of 91.5% in the validation group [35]. This panel represents a promising approach to complement a CT scan for early NSCLC diagnosis based on a predictive accuracy of 91%. The validation cohort contained 102 patients including 40 patients with indeterminate nodules, 56 patients with autoimmune diseases and six patients with prevalence cancers. Rigorous validation was required including sample analysis during tumour transformation as eight occult cancers were misclassified as normal whilst using this panel. The protein panel and its association with NSCLC requires further investigation using different types of cancers and controls as the proteins did not show significant homology to the complete sequence

and small differences in amino acids may indicate they belong to different parent proteins than they were assigned to in the GenBank database.

Farlow et al. 2010a [26] identified six markers: tumour necrosis factor α (TNF-α), CYFRA 21-1, interleukin-1ra (IL-1ra), matrix met alloproteinase-2 (MMP-2), monocyte chemotactic protein-1 (MCP-1) and sE-selectin. Only CYFRA 21-1 was a well-characterised biomarker for NSCLC diagnosis whilst TNF-α and IL-1ra were known inflammatory mediators that were non-specific to cancer. The sensitivity and specificity of this panel was high for early LC detection, at 99% and 95% respectively, with 66/74 patients being correctly identified stage I and II LC when compared to 43 non-malignant lung conditions that were used as controls. However, the panel had a 47% false positive rate and failed to differentiate between NSCLC and inflammatory diseases such as chronic obstructive pulmonary disorder and pneumonia which share similar inflammatory mediators. To improve the specificity, tumour-associated autoantibodies could also be used [26]. Wang et al [36] used a panel of four-markers including MIC-1, CYFRA21-1, CA125 and CEA and showed high sensitivity and specificity for LC diagnosis. There was a higher sensitivity of 90.4% for adenocarcinomas and 92.1% for squamous cell carcinoma (in NSCLC), compared with 83.9% for SCLC. This panel also had a lower sensitivity for early stages (84.4%) compared to the sensitivity (89.5%) for disease diagnosis at all stages. The panel was tested using an independent group of patients, demonstrating a sensitivity and specificity of 88.4% and 93.1% respectively. However, there was a false positive rate of 56.4% in the blinded samples from patients with benign tumours and tuberculosis [36].

Combining serum ferritin, shown to play a prognostic role in advanced hepatic cancer [37], with NSCLC markers (CA125, CEA, NSE and CYFRA21-1) improved the diagnostic performance of the panel for early NSCLC diagnosis to a sensitivity of 92.97% and specificity of 90% in elderly patients [27]. This panel requires verification in a larger cohort and in younger members of the patient population. Wang et al [38] found that the addition of haptoglobin to the clinical LC biomarkers CEA, NSE and CYFRA21-1 improved their diagnostic performance, especially for CYFRA21-1 and Hp for squamous LC. Ajona et al [39] showed that C4c, CYFRA 21-1 and C-reactive protein (CRP) are potential biomarkers of early NSCLC. However, this panel was derived from a limited selection of circulating proteins using retrospective samples. The standardisation and calibration of assays, especially for C4c, and larger prospective studies with different control groups are essential to test diagnostic accuracy.

3.2.2 Autoantibodies as biomarkers for NSCLC

Autoantibodies biomarkers with the highest sensitivities and specificities were paxillin combined with SEC15L2, BAC clone RP11-499F19, XRCC5 and MALAT1 [35] while Farlow et al., 2010b found that a six-autoantibody panel (IMPDH, phosphoglycerate mutase, ubiquillin, Annexin I, Annexin II, and HSP70-9B) achieved an AUC of 0.964, a sensitivity of 94.8%, and a specificity of 91.1%. The overall misclassification rate was 7% within the patient population analysed (*n* = 196). Further validation of this finding will require the use of an asymptomatic cohort containing a suitable control such as smokers, as well as healthy and cancer patients [40]. Inflammatory conditions such as COPD induce specific autoantibody production that may lead to the misclassification of patients having NSCLC. Whether these autoantibodies are produced during or before carcinogenesis still needs to be explored [40]. Jiang et al. [41] stated their seven-TAAb panel showed the highest sensitivity and specificity for early NSCLC, at 94.4% and 82.7% respectively. A limitation of this study included the small sample number of patients with early NSCLCs (72 patients with stage I and II) and the diagnostic performance of the panel for malignant rather than benign lung nodules will require further assessment.

Pan et al. [42] suggested a panel of six autoantibodies could enable the early detection of NSCLC with sensitivities and specificities of 73.5%/68.2% and >85%/87 for the training and validation cohorts respectively. The six autoantigens were shown to be highly expressed in NSCLC by immunohistochemistry with positive scores of 66.7%, 61.6%, 58.3%, 58.3%, 26.6% and 36.7% for BCL7A, TRIM33, MTERF4, CTAG1A, DDX4 and MAGEC2 respectively [43]. Some of these genes are known to be mutated in NSCLC especially the cancer testis antigens (CTAs) CTAG1A, DDX4 and MAGEC2. Inclusion of CTAs can increase both the specificity and sensitivity of a biomarker panel by virtue of their elevated levels in disease and restricted expression in healthy tissues. One drawback of this study, like most published biomarker studies, is that it is a single-centre study, that would benefit from the utilisation of a larger cohort size. Considering the genetic variation between ethnicities, and the impact of the environment on LC development, we note that in this study, economic progress and air pollution may have impacted the biomarkers that are relevant to LC in East Asia [44].

Chen et al. [45] focused on the autoantibodies that are associated with cancer-stem cell-like (stem) signatures. Only SOX2 expression was associated with tumour stage. MAGEA1 is a CTA that is overexpressed in NSCLC and is associated with necrosis. 70-85% of NSCLC patients have upregulated MAGEA1, A3 and B2 due to global promoter hypomethylation [46]. MAGEA1 and MAGEA3/4 have been found to be expressed in 17% and 44% of NSCLC samples respectively [47]. MAGEA3/4 was found more often in squamous cell carcinoma $P<0.001$ while MAGEA1 was found more frequently in adenocarcinoma. Determining MAGE transcript levels in urine and sputum may be useful for biomarker discovery in LC [48]. However, MAGEA1 levels were also increased in other cancers such as breast and gastric tumours [47]. Thus, this panel is not recommended to identify the type of cancer but could aid in early diagnosis of NSCLC when combined with a CT scan. The presence of autoantibodies against these CTAs may predict poorer survival but further studies are required to validate the utility of this panel for early LC diagnosis with a larger number of participants [45].

Song et al. [49] found that the ratio of anti-CYFRA 21-1 autoantibody immune complex (CIC) and free CYFRA 21-1 had a sensitivity of 76.0%, 80.0%, 76.9% and 50.0% for the detection of stage I, II, III and IV LCs respectively. Therefore, this could be applied to the identification of asymptomatic patients in

a seemingly healthy population. A similar finding was made in colon cancer as CYFRA21-1 is a fragment of cytokine 19 that is overexpressed in epithelial cancers [50].

Studies also combined autoantibodies with antigens as biomarkers for NSCLC. Zang et al. [51] found that combining autoantibodies, with LC antigens, improved the diagnostic performance of this biomarker panel, with a sensitivity of 86.5% and specificity of 82.3%. Both alpha enolase and annexin A1 (autoantibodies investigated in this study) are upregulated in LC and were suggested to be biomarkers for NSCLC staging. This study did not include non-malignant lung diseases and was not externally validated. Doseeva et al. [52] found that a panel of one autoantibody marker and the detection of three Ags had a sensitivity and specificity of 74% and 80% in the training cohort, and a higher sensitivity of 77% but the same specificity (80%) in the validation cohort. Five out of eight cases of false positives were COPD patients but COPD is known as independent risk factor of LC [53]. However, this study also lacked samples from patients with benign nodules [52].

3.2.3 miRNAs and mRNAs as biomarkers for NSCLC

miRNA are short sequences of noncoding RNA of 19-22 nucleotides that are involved in the control of gene transcription. miRNAs have also been described as possible biomarkers for LC with high sensitivity. The highest was miR-15b and miR-27b with a 100% sensitivity and 84% specificity [54] while miR-16-5p, miR-92a-3p and miR-451a can facilitate an early LC diagnosis [55]. Further studies should include large numbers of patients and controls from ethnic groups to which these panels are needed. The specificities of each miRNA were also high, ranging from 72% to 92.5% for all 46 miRNAs examined in 15 selected articles.

Xing et al. [56] identified mRNA from tumour-educated platelets (ITGA2B) with a sensitivity of 92.8% and specificity of 78.6% in the training cohort. In the validation cohort, the specificity decreased significantly to 56%. Li et al. [57] found that CEA and exo-GAS5 also showed a high diagnostic performance for stage I NSCLC with a sensitivity of 84.21% and a specificity of 90%. Growth arrest-specific transcript 5 (GAS5)[58] has been shown to be decreased in both tissue and plasma from NSCLC patients and its expression was associated with NSCLC tumour size [58]. lncRNA GAS5 expression is downregulated in NSCLCs while GAS5 expression in secreted exosomes was upregulated in NSCLCs. The diagnostic significance of exosomic GAS5 was higher in tumour tissue than in circulating GAS5 serum levels but also more difficult and time-consuming to determine.

3.2.4 DNA and CTC as biomarkers for NSCLC

Five different loci from the microsatellite instability (MSI)/loss of heterozygosity (LOH) loci family had a sensitivity of 90% and a specificity of 71% for early NSCLC [59]. In contrast, Yang et al. [60] showed the highest specificity (91%) for eight methylated genes, with a 72% of sensitivity. Chromosome enumeration probe 8 (CEP8) is one of the CTCs produced by NSCLC tissue [61] and it's expression was associated with diagnosis and prognosis of NSCLC with high sensitivity and specificity at 83.3% and 98.6%, respectively [62]. CEP8 combined with CA125 increased the detection of NSCLC from 83% to 100% for sensitivity. Combining CEP8 and NSE achieved a sensitivity and specificity of 83% each. The limitation of this study included the small sample size of 18 solid nodules, the absence a control group and a lack of result validation. CEP8 performance and reference values also need to be established for LC [63].

3.2.5 Sputum biomarkers in lung cancer

Sputum is produced directly by the upper and lower respiratory tract and can serve as a surrogate sample for the diagnosis of lung cancer. Lung cancer tissue can affect the biological components of sputum and detection of overexpressed genes in sputum can help diagnose lung cancer [64, 65]. Single antigen, A proliferation-inducing ligand (APRIL), has high specificity of 97% and sensitivity of 82% following detection by immunocytochemistry [65]. Three sputum TAAbs were developed as a biomarker panel for the diagnosis of lung cancer, regardless of stage, site, and histologic type, with 81% sensitivity and 83% specificity [64]. miR-21 has low sensitivity of 48% but an absolute 100% specificity in 23 NSCLC samples compared to 17 controls [66]. Panels of miR-205, miR-210 and miR-708 had a diagnostic sensitivity of 73% and specificity of 96% of distinguishing patients with squamous cell carcinoma from healthy controls [67]. Su et al. [68] found that three miRNAs combined with two small nucleolar RNA had an AUC-ROC of 0.94 when distinguishing NSCLC patients from cancer-free subjects. Subsequently Su et al. [69] found that a combination of two miRs and the methylation of two genes had sensitivity of 87% and specificity of 90% when identifying stage I NSCLC compared with controls.

4. DISCUSSION

Our review found that well-performing single biomarkers for early NSCLC diagnosis included Ciz1 (sensitivity: 95%) and exosomal GCC2 (sensitivity: 90%) with a slightly lower specificity of 71% for CIZ1 and 75% for exosomal GCC2 respectively. Tumour-educated blood platelets (ITGA2B) also had high sensitivities in both the training (92.8%) and validation cohorts (91.2%) but low specificity. In contrast, CYFRA 21-1 and anti-HE4 had high specificity for LC (95% each). OPNV had a sensitivity of 80% and a specificity of 88% as a biomarker for early LC. Biomarker panels (Table 1) had high sensitivity and specificity (greater than 90%). Combining biomarkers was more likely to facilitate the early detection of NSCLC, especially when antigens or autoantibodies were combined with miRNAs. Early detection is the holy grail of NSCLC diagnoses as it offers the opportunity to significantly increase survival rates, aid in the management of the disease, and reduce overall healthcare costs. CYFRA21-1 is a prognostic biomarker for advanced NSCLC, predominately found in lung tissues that correlates with tumour size, lymph node involvement and the stage of the disease. Lower baseline levels of CYFRA21-1 were associated with both longer overall survival and failure free survival ($p<0.0001$ and $p=0.0003$)[70]. Wang et al. [71] found that serological levels CYFRA21-1 combined with other markers delivered different sensitivities and specificities, depending on the sample size. A commonly used combination of CYFRA21-1, CEA and NSE for NSCLC detection revealed a very low sensitivity of 31% in contrast to its very high specificity of 96%. These assays have the drawback of low sensitivity especially in the early stages of NSCLC, whilst the same panel of biomarkers have shown high sensitivity at advanced stages. Therefore, this combination could not be recommended for use in early detection in clinical practice. However, the high concentration of these biomarkers in body fluids/levels in tissues are poor prognostic indicators. Thus, these three biomarkers could be used to predict relapse before the onset of clinical symptoms as their concentration can be used to monitor therapy response/resistance.

In the early stages of NSCLC, patients are primarily asymptomatic but a low tumour presence can elicit TAs/TAAbs that are detectable at higher levels in liquid biopsies, compared to samples from healthy

individuals [72]. Thus, changes in the immune response can be detected in blood before clinical symptoms appear. Different molecules such as proteins and miRNA have been shown to be very sensitive biomarkers, which are cost effective and safe compared to imaging techniques such as CT scans that are associated with an increased risk of cancer due to radiation, require specialist training and are expensive compared to blood tests. miRNAs in blood have a higher sensitivity for LC diagnosis compared to miRNAs in sputum as the oral cavity contains many enzymes such as those that degrade these small molecules [73]. Although sputum represents a source for LC biomarkers and is considered a non-invasive technique [64], it requires patient co-operation and capacity to deliver spontaneous samples. Sputum collection and analysis can be highly variable, dependent on factors such as the quality of the collected sputum and the techniques used in the laboratory which can impact the reliability and reproducibility. miRNAs lack the specificity needed for early LC diagnosis as they are expressed in many cancers and healthy tissues. Due to this and their reduced stability, miRNAs are not suggested for clinical use [73]. In contrast, antigens are frequently used as markers for disease diagnosis with the aid of imaging techniques [11, 74].

Biomarkers can be diagnostic, predictive or indicative [75, 76]. Biomarker discovery is largely dependent on an analytic validation for measuring biomarkers in body fluids. Blood is mostly used to detect molecular changes associated with LC after depleting the abundant proteins; leaving the biomarkers of interest that are usually present in very low concentrations. The stability of biomarkers is a crucial factor as it affects reproducibility and analytic validation procedures [77, 78]. In addition to study design and population selection should also be considered. Sample size should be statistically valid as a low number of participants exaggerates the diagnostic performance of biomarkers. The required sample sizes should be calculated to achieve 95% confidence levels and 80% power for purpose of testing the validity of the biomarker [79]. Moreover, age and gender matched controls should be considered. Ideally, biomarkers must be highly sensitive and specific for cancer diagnosis. However, there is no marker in clinical practice that possesses both 100% sensitivity and specificity. The use of biomarkers have been proposed in addition to imaging techniques, which would have a greater benefit-to-risk ratio compared to markers or imaging alone [80].

Biomarker research should be optimised by developing a common workflow. Identifying the optimal cut-off point of biomarkers is required for their application in the clinical setting. Most biomarkers in this review utilised retrospective designs and samples from tissue banks. Ideal biomarker studies should have a prospective design such as randomised controlled trials, with a large sample size ensuring that the study is able to achieve adequate precision following the Standards for Reporting of Diagnostic Accuracy Studies (STARD) guidelines and examine populations with disease and compare them to age- and sex-matched controls [81]. This would reduce false positives associated with CT results and thus overtreatment and side effects from unnecessary interventions [82]. Although a change in biomarker expression may not reflect true clinical benefit, it has been associated with pathway modulation [83]. Biomarker translation into clinical practice is a challenging mission and even with approved markers such as CEA for colon cancer diagnosis, it's sensitivity is not ideal as it is expressed in other cancers and non-malignant conditions [84].

Sensitivities and specificities are dependent on the biomarker selected and the LC types studied. Biomarker assays require both robustness and reproducibility to be applied for clinical use [85]. Studies with validation cohorts are more robust than studies with only a testing group [86]. For

example, Xing et al [56] showed that the variations in results were due to a difference in the number of participants and controls with a range of non-malignant conditions being used to determine the specificity of the biomarkers. Goebel et al [87] examined 21 candidate biomarkers including antigens and cytokines using a multiplex immunoassay but many were excluded even with >80% sensitivity and >95% specificity as the assay lacked reproducibility and was difficult to perform using such a large number of biomarkers. Developing an optimal multiplex test is required to validate the findings of this study and to examine its functionality and clinical use [87].

This systematic review has several limitations. We only included articles in English and some quantitative studies could not be included as they did not adequately report the diagnostic performance of the biomarkers investigated. There was also considerable variability across studies in terms of timing, participants and control groups, sampling, and biomarker detection methods. Included studies assessed a combination of biomarkers, which commonly were not validated in multi-centre studies hence we were unable to make firm conclusions on their diagnostic accuracy, nor conduct a meta-analysis for each biomarker. Future studies should report their findings adequately, following the STARD guidelines for the construction of 2-by-2 tables for diagnostic meta-analysis, and minimally also including the 95% CI of diagnostic effect measures [81]. Future research of NSCLC biomarker diagnosis should emphasise the validation of biomarkers so that they can be translated into clinical use and impact patient treatment and care.

Acknowledgements

We would like to thank the librarians of the Brynmor Jones Library at the University of Hull for their guidance.

Availability of data and materials

All data generated or analysed during this study are included in this published article while the systematic literature review process is available.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported the University of Hull cluster PhD studentship (E.M.)

CONTRIBUTION: E.M., S.H. and B.G. designed the study. E.M., D.M. and S.Q.Y. performed the Systematic Literature Review, analysed the data and made the figures. E.M., S.Q.Y. and B.G. wrote the paper. S.Q.Y. and M.H. assessed the risk of bias. All authors read and approved the final manuscript.

REFERENCES

1. Prabhakar, B., P. Shende, and S. Augustine, *Current trends and emerging diagnostic techniques for lung cancer.* Biomedicine & Pharmacotherapy, 2018. **106**: p. 1586-1599.
2. Schabath, M.B. and M.L. Cote, *Cancer Progress and Priorities: Lung Cancer.* Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 2019. **28**(10): p. 1563-1579.
3. Zappa, C. and S.A. Mousa, *Non-small cell lung cancer: current treatment and future advances.* Translational lung cancer research, 2016. **5**(3): p. 288-300.
4. Woodman, C., et al., *Applications and strategies in nanodiagnosis and nanotherapy in lung cancer.* Seminars in Cancer Biology, 2021. **69**: p. 349-364.
5. Patz Jr, E.F., et al., *Panel of serum biomarkers for the diagnosis of lung cancer.* Journal of Clinical Oncology, 2007. **25**(35): p. 5578-5583.
6. Schneider, J., *Tumor markers in detection of lung cancer.* Advances in clinical chemistry, 2006. **42**: p. 1-41.
7. Hanagiri, T., et al., *Preoperative CYFRA 21-1 and CEA as prognostic factors in patients with stage I non-small cell lung cancer.* Lung cancer, 2011. **74**(1): p. 112-117.
8. Kulpa, J., et al., *Carcinoembryonic antigen, squamous cell carcinoma antigen, CYFRA 21-1, and neuron-specific enolase in squamous cell lung cancer patients.* Clinical chemistry, 2002. **48**(11): p. 1931-1937.
9. Schneider, J., et al., *Pro-gastrin-releasing peptide (ProGRP), neuron specific enolase (NSE), carcinoembryonic antigen (CEA) and cytokeratin 19-fragments (CYFRA 21-1) in patients with lung cancer in comparison to other lung diseases.* Anticancer research, 2003. **23**(2A): p. 885.
10. Grunnet, M. and J. Sorensen, *Carcinoembryonic antigen (CEA) as tumor marker in lung cancer.* Lung cancer, 2012. **76**(2): p. 138-143.
11. Thomas, S.N., et al., *Carcinoembryonic antigen and CD44 variant isoforms cooperate to mediate colon carcinoma cell adhesion to E-and L-selectin in shear flow.* Journal of biological chemistry, 2008. **283**(23): p. 15647-15655.
12. Scott, A. and R. Salgia, *Biomarkers in lung cancer: from early detection to novel therapeutics and decision making.* Biomarkers in medicine, 2008. **2**(6): p. 577-586.
13. Yang, B., et al., *Measuring serum human epididymis secretory protein autoantibody as an early biomarker of lung cancer.* Transl Cancer Res, 2020. **9**: p. 735-41.
14. Altintas, Z. and I. Tothill, *Biomarkers and biosensors for the early diagnosis of lung cancer.* Sensors and Actuators B: Chemical, 2013. **188**: p. 988-998.
15. Chapman, C.J., et al., *EarlyCDT®-Lung test: improved clinical utility through additional autoantibody assays.* Tumor Biology, 2012. **33**(5): p. 1319-1326.
16. Yang, B., et al., *Autoantibodies as diagnostic biomarkers for lung cancer: A systematic review.* Cell Death Discovery, 2019. **5**(1): p. 126.
17. Du, Q., et al., *Significance of tumor‐associated autoantibodies in the early diagnosis of lung cancer.* The Clinical Respiratory Journal, 2018. **12**(6): p. 2020-2028.
18. Page, M.J., et al., *PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews.* bmj, 2021. **372**.

19. Whiting, P.F., et al., *QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies.* Annals of internal medicine, 2011. **155**(8): p. 529-536.
20. McGuinness, L.A. and J.P. Higgins, *Risk‐of‐bias VISualization (robvis): an R package and Shiny web app for visualizing risk‐of‐bias assessments.* Research synthesis methods, 2021. **12**(1): p. 55-61.
21. McKenzie, J.E., et al., *Summarizing study characteristics and preparing for synthesis*, in *Cochrane Handbook for Systematic Reviews of Interventions*. 2019. p. 229-240.
22. Richardson, M., P. Garner, and S. Donegan, *Interpretation of subgroup analyses in systematic reviews: a tutorial.* Clinical Epidemiology and Global Health, 2019. **7**(2): p. 192-198.
23. Begg, C.B. and M. Mazumdar, *Operating characteristics of a rank correlation test for publication bias.* Biometrics, 1994: p. 1088-1101.
24. Egger, M., et al., *Bias in meta-analysis detected by a simple, graphical test.* Bmj, 1997. **315**(7109): p. 629-634.
25. Campbell, M., et al., *Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline.* bmj, 2020. **368**.
26. Farlow, E., et al., *A multi-analyte serum test for the detection of non-small cell lung cancer.* British journal of cancer, 2010. **103**(8): p. 1221-1228.
27. Yang, J., et al., *Early diagnosis of lung cancer in the elderly using four tumor markers and serum ferritin for better surgical management.* Asian journal of surgery, 2020. **43**(11): p. 1088-1089.
28. Yuan, Z., et al., *Diagnostic value of HSP90α and related markers in lung cancer.* J Clin Lab Anal, 2022. **36**(6): p. e24462.
29. Gasparri, R., et al., *Serum proteomics profiling identifies a preliminary signature for the diagnosis of early-stage lung cancer.* PROTEOMICS – Clinical Applications, 2023. **17**(2): p. 2200093.
30. Ma, Y.C., et al., *Urinary malate dehydrogenase 2 is a new biomarker for early detection of non‐small‐cell lung cancer.* Cancer Science, 2021. **112**(6): p. 2349-2360.
31. Jeong, H., et al., *GCC2 as a New Early Diagnostic Biomarker for Non-Small Cell Lung Cancer.* Cancers, 2021. **13**(21): p. 5482.
32. Uhlén, M., et al., *Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
33. Higgins, G., et al., *Variant Ciz1 is a circulating biomarker for early-stage lung cancer.* Proceedings of the National Academy of Sciences, 2012. **109**(45): p. E3128-E3135.
34. Joseph, S., et al., *Plasma osteopontin velocity differentiates lung cancers from controls in a CT screening population.* Cancer biomarkers : section A of Disease markers, 2012. **12**(4): p. 177-184.
35. Zhong, L., et al., *Profiling Tumor-Associated Antibodies for Early Detection of Non-small Cell Lung Cancer.* Journal of Thoracic Oncology, 2006. **1**(6): p. 513-519.
36. Wang, X., et al., *A novel serum based biomarker panel has complementary ability to preclude presence of early lung cancer for low dose CT (LDCT).* Oncotarget, 2017. **8**(28).
37. Song, A., et al., *Significance of serum ferritin as a prognostic factor in advanced hepatobiliary cancer patients treated with Korean medicine: A retrospective cohort study.* BMC Complementary and Alternative Medicine, 2018. **18**(1).

38. Wang, R., et al., *Clinical evaluation and cost-effectiveness analysis of serum tumor markers in lung cancer.* BioMed research international, 2013. **2013**.
39. Ajona, D., et al., *A model based on the quantification of complement C4c, CYFRA 21-1 and CRP exhibits high specificity for the early diagnosis of lung cancer.* Translational Research, 2021. **233**: p. 77-91.
40. Farlow, E.C., et al., *Development of a multiplexed tumor-associated autoantibody-based blood test for the detection of non-small cell lung cancer.* Clin Cancer Res, 2010. **16**(13): p. 3452-62.
41. Jiang, D., et al., *Discovering Panel of Autoantibodies for Early Detection of Lung Cancer Based on Focused Protein Array.* Frontiers in immunology, 2021. **12**: p. 658922.
42. Pan, J., et al., *Integration of IgA and IgG Autoantigens Improves Performance of Biomarker Panels for Early Diagnosis of Lung Cancer.* Molecular & Cellular Proteomics, 2020. **19**(3): p. 490-500.
43. Jung, Y.J., et al., *Development of a protein biomarker panel to detect non–small-cell lung cancer in Korea.* Clinical Lung Cancer, 2017. **18**(2): p. e99-e107.
44. Lam, S., et al., *Early CDT-Lung: an immunobiomarker test as an aid to early detection of lung cancer.* Cancer prevention research, 2011. **4**(7): p. 1126-1134.
45. Chen, S.-S., et al., *Stem signatures associated antibodies yield early diagnosis and precise prognosis predication of patients with non-small cell lung cancer.* Journal of Cancer Research and Clinical Oncology, 2021. **147**(1): p. 223-233.
46. Chapman, C.J., et al., *Autoantibodies in lung cancer: possibilities for early detection and subsequent cure.* Thorax, 2008. **63**(3): p. 228-233.
47. Grah, J.J., et al., *Clinical significance of immunohistochemical expression of cancer/testis tumor-associated antigens (MAGE-A1, MAGE-A3/4, NY-ESO-1) in patients with non-small cell lung cancer.* Tumori Journal, 2014. **100**(1): p. 60-68.
48. Sugita, M., et al., *Combined use of oligonucleotide and tissue microarrays identifies cancer/testis antigens as biomarkers in lung carcinoma.* Cancer Research, 2002. **62**(14): p. 3971-3979.
49. Song, K.-S., et al., *Quantification of CYFRA 21-1 and a CYFRA 21-1–anti-CYFRA 21-1 autoantibody immune complex for detection of early stage lung cancer.* Chemical Communications, 2019. **55**(68): p. 10060-10063.
50. Hong Woo, C., et al., *Diagnostic Significance of the Ratio of Plasma CYFRA 21-1 Autoantibody Immune Complex to Free CYFRA 21-1 in Patients with Colon Cancer.* Lab Med Qual Assur, 2020. **42**(4): p. 218-223.
51. Zang, R., et al., *Enhancement of diagnostic performance in lung cancers by combining CEA and CA125 with autoantibodies detection.* Oncoimmunology, 2019. **8**(10): p. e1625689.
52. Doseeva, V., et al., *Performance of a multiplexed dual analyte immunoassay for the early detection of non-small cell lung cancer.* Journal of translational medicine, 2015. **13**: p. 55-55.
53. Loganathan, R.S., et al., *Prevalence of COPD in women compared to men around the time of diagnosis of primary lung cancer.* Chest, 2006. **129**(5): p. 1305-1312.
54. Hennessey, P.T., et al., *Serum microRNA biomarkers for detection of non-small cell lung cancer.* PloS one, 2012. **7**(2): p. e32307.

55. Reis, P.P., et al., *Circulating miR-16-5p, miR-92a-3p, and miR-451a in Plasma from Lung Cancer Patients: Potential Application in Early Detection and a Regulatory Role in Tumorigenesis Pathways.* Cancers, 2020. **12**(8).
56. Xing, S., et al., *Development and validation of tumor-educated blood platelets integrin alpha 2b (ITGA2B) RNA for diagnosis and prognosis of non-small-cell lung cancer through RNA-seq.* International journal of biological sciences, 2019. **15**(9): p. 1977.
57. Li, C., et al., *Tumor‐derived exosomal lncRNA GAS5 as a biomarker for early‐stage non‐small‐cell lung cancer diagnosis.* Journal of cellular physiology, 2019. **234**(11): p. 20721-20727.
58. Ma, C., et al., *The growth arrest-specific transcript 5 (GAS5): a pivotal tumor suppressor long noncoding RNA in human cancers.* Tumour Biol, 2016. **37**(2): p. 1437-44.
59. Carozzi, F.M., et al., *Multimodal lung cancer screening using the ITALUNG biomarker panel and low dose computed tomography. Results of the ITALUNG biomarker study.* International Journal of Cancer, 2017. **141**(1): p. 94-101.
60. Yang, Z., et al., *DNA methylation analysis of selected genes for the detection of early-stage lung cancer using circulating cell-free DNA.* Advances in Clinical and Experimental Medicine, 2019. **28**(3): p. 355-360.
61. Atasoy, S., et al., *Analysis of Chromosome 3, 7 and 8 Centromeric Regions in Bronchial Lavage Specimens by FISH.* Turk Thorac J, 2016. **17**(4): p. 141-147.
62. Chen, Q., et al., *Lung cancer circulating tumor cells isolated by the EpCAM-independent enrichment strategy correlate with Cytokeratin 19-derived CYFRA21-1 and pathological staging.* Clin Chim Acta, 2013. **419**: p. 57-61.
63. Zhong, M., et al., *Clinical Utility of Circulating Tumor Cells in the Early Detection of Lung Cancer in Patients with a Solitary Pulmonary Nodule.* Technology in Cancer Research & Treatment, 2021. **20**.
64. Li, N., et al., *Autoantibodies against tumor-associated antigens in sputum as biomarkers for lung cancer.* Translational Oncology, 2021. **14**(2): p. 100991.
65. Sun, B., et al., *A proliferation-inducing ligand: a new biomarker for non-small cell lung cancer.* Experimental lung research, 2009. **35**(6): p. 486-500.
66. Xie, Y., et al., *Altered miRNA expression in sputum for diagnosis of non-small cell lung cancer.* Lung Cancer, 2010. **67**(2): p. 170-176.
67. Xing, L., et al., *Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers.* Modern Pathology, 2010. **23**(8): p. 1157-1164.
68. Su, Y., et al., *Small non-coding RNA biomarkers in sputum for lung cancer diagnosis.* Molecular Cancer, 2016. **15**(1): p. 36.
69. Su, Y., H. Fang, and F. Jiang, *Integrating DNA methylation and microRNA biomarkers in sputum for lung cancer detection.* Clinical Epigenetics, 2016. **8**(1): p. 109.
70. Edelman, M.J., et al., *CYFRA 21-1 as a prognostic and predictive marker in advanced non-small-cell lung cancer in a prospective trial: CALGB 150304.* J Thorac Oncol, 2012. **7**(4): p. 649-54.
71. Wang, J., et al., *Increased CYFRA 21-1, CEA and NSE are Prognostic of Poor Outcome for Locally Advanced Squamous Cell Carcinoma in Lung: A Nomogram and Recursive Partitioning Risk Stratification Analysis.* Translational Oncology, 2018. **11**(4): p. 999-1006.

72. Guibert, N., et al., *Current and future applications of liquid biopsy in nonsmall cell lung cancer from early to advanced stages.* European Respiratory Review, 2020. **29**(155).
73. Kammer, M.N. and P.P. Massion, *Noninvasive biomarkers for lung cancer diagnosis, where do we stand?* Journal of Thoracic Disease, 2020. **12**(6): p. 3317-3330.
74. Thomas, S.N., et al., *Identification, characterization and utilization of tumor cell selectin ligands in the design of colon cancer diagnostics.* Biorheology, 2009. **46**(3): p. 207-225.
75. Nalejska, E., E. Mączyńska, and M.A. Lewandowska, *Prognostic and predictive biomarkers: tools in personalized oncology.* Molecular diagnosis & therapy, 2014. **18**(3): p. 273-284.
76. Voon, P.J. and H.L. Kong, *Tumour genetics and genomics to personalise cancer treatment.* Annals of the Academy of Medicine-Singapore, 2011. **40**(8): p. 362.
77. Fleming, T.R., *Surrogate endpoints and FDA's accelerated approval process.* Health affairs, 2005. **24**(1): p. 67-78.
78. Schatzkin, A. and M. Gail, *The promise and peril of surrogate end points in cancer research.* Nature Reviews Cancer, 2002. **2**(1): p. 19-27.
79. Hajian-Tilaki, K., *Sample size estimation in diagnostic test studies of biomedical informatics.* Journal of Biomedical Informatics, 2014. **48**: p. 193-204.
80. Koscielny, S., *Why most gene expression signatures of tumors have not been useful in the clinic.* Sci Transl Med, 2010. **2**(14): p. 14ps2.
81. Cohen, J., et al., *STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016 Nov 14; 6 (11): e012799. doi: 10.1136/bmjopen-2016-012799*. 2016.
82. Mazzone, P.J., et al., *Evaluating Molecular Biomarkers for the Early Detection of Lung Cancer: When Is a Biomarker Ready for Clinical Use? An Official American Thoracic Society Policy Statement.* American journal of respiratory and critical care medicine, 2017. **196**(7): p. e15-e29.
83. Dunn, B.K., K. Jegalian, and P. Greenwald, *Biomarkers for early detection and as surrogate endpoints in cancer prevention trials: issues and opportunities.* Clinical Cancer Prevention, 2010: p. 21-47.
84. Ransohoff, D.F. and M.L. Gourlay, *Sources of bias in specimens for research about molecular markers for cancer.* J Clin Oncol, 2010. **28**(4): p. 698-704.
85. Pass, H.I., et al., *Biomarkers and molecular testing for early detection, diagnosis, and therapeutic prediction of lung cancer.* Thoracic surgery clinics, 2013. **23**(2): p. 211-224.
86. Mehan, M.R., et al., *Validation of a blood protein signature for non-small cell lung cancer.* Clinical proteomics, 2014. **11**(1): p. 32-32.
87. Goebel, C., et al., *Diagnosis of non-small cell lung cancer for early stage asymptomatic patients.* Cancer Genomics & Proteomics, 2019. **16**(4): p. 229-244.
88. Bigbee, W.L., et al., *A multiplexed serum biomarker immunoassay panel discriminates clinical lung cancer patients from high-risk individuals found to be cancer-free by CT screening.* Journal of thoracic oncology, 2012. **7**(4): p. 698-708.
89. Fahrmann, J.F., et al., *Blood-Based Biomarker Panel for Personalized Lung Cancer Risk Assessment.* J Clin Oncol, 2022. **40**(8): p. 876-883.
90. Kupert, E., et al., *Plasma secretory phospholipase A2-IIa as a potential biomarker for lung cancer in patients with solitary pulmonary nodules.* Bmc Cancer, 2011. **11**(1): p. 1-10.

91. Lai, Y., et al., *Identification and Validation of Serum CST1 as a Diagnostic Marker for Differentiating Early-Stage Non-Small Cell Lung Cancer from Pulmonary Benign Nodules.* Cancer Control, 2022. **29**: p. 10732748221104661.
92. Li, J., et al., *Secreted proteins MDK, WFDC2, and CXCL14 as candidate biomarkers for early diagnosis of lung adenocarcinoma.* BMC Cancer, 2023. **23**(1): p. 110.
93. Li, G., et al., *Serum Markers CA125, CA153, and CEA along with Inflammatory Cytokines in the Early Detection of Lung Cancer in High-Risk Populations.* Biomed Res Int, 2022. **2022**: p. 1394042.
94. Meng, F., et al., *Ratiometric electrochemical OR gate assay for NSCLC-derived exosomes.* J Nanobiotechnology, 2023. **21**(1): p. 104.
95. Nolen, B.M., et al., *Serum biomarker profiles as diagnostic tools in lung cancer.* Cancer Biomark, 2011. **10**(1): p. 3-12.
96. Pakvisal, N., et al., *Differential expression of immune-regulatory proteins C5AR1, CLEC4A and NLRP3 on peripheral blood mononuclear cells in early-stage non-small cell lung cancer patients.* Sci Rep, 2022. **12**(1): p. 18439.
97. Pio, R., et al., *Complement factor H is elevated in bronchoalveolar lavage fluid and sputum from patients with lung cancer.* Cancer epidemiology, biomarkers & prevention, 2010. **19**(10): p. 2665-2672.
98. Sun, N., et al., *Utility of isocitrate dehydrogenase 1 as a serum protein biomarker for the early detection of non‐small‐cell lung cancer: A multicenter in vitro diagnostic clinical trial.* Cancer science, 2020. **111**(5): p. 1739-1749.
99. Song, K.S., et al., *Detection and Quantification of Tp53 and p53-Anti-p53 Autoantibody Immune Complex: Promising Biomarkers in Early Stage Lung Cancer Diagnosis.* Biosensors (Basel), 2022. **12**(2).
100. Wieskopf, B., et al., *Cyfra 21-1 as a biologic marker of non-small cell lung cancer: evaluation of sensitivity, specificity, and prognostic role.* Chest, 1995. **108**(1): p. 163-169.
101. Wu, H.-Y., et al., *Assessment of Serological Early Biomarker Candidates for Lung Adenocarcinoma by using Multiple Reaction Monitoring-Mass Spectrometry.* Proteomics. Clinical applications, 2020. **14**(4): p. e1900095.
102. Yu, W., et al., *Combination of serum ACSL4 levels and low-dose 256-slice spiral CT exhibits the potential in the early screening of lung cancer.* Medicine (Baltimore), 2023. **102**(5): p. e32733.
103. Zhang, X., et al., *Identification of serum MiRNAs as candidate biomarkers for non-small cell lung cancer diagnosis.* BMC Pulm Med, 2022. **22**(1): p. 479.
104. Ezzatifar, F., et al., *Detection of Novel Autoantibodies to Nucleolin's RNA-binding Domains as a Serum Tumor Biomarker Through ELISA.* Iran J Allergy Asthma Immunol, 2022. **21**(6): p. 616-625.
105. Hua, Y., et al., *Autoantibody panel on small extracellular vesicles for the early detection of lung cancer.* Clinical Immunology, 2022. **245**: p. 109175.
106. Huo, Y., et al., *Case study of an autoantibody panel for early detection of lung cancer and ground-glass nodules.* Journal of cancer research and clinical oncology, 2020. **146**(12): p. 3349-3357.
107. Lastwika, K.J., et al., *Tumor-derived autoantibodies identify malignant pulmonary nodules.* American Journal of Respiratory and Critical Care Medicine, 2019. **199**(10): p. 1257-1266.

108. Liu, S., et al., *Detection of circulating natural antibodies against CD25, MUC1, and VEGFR1 for early diagnosis of non‐small cell lung cancer.* FEBS Open bio, 2020. **10**(7): p. 1288-1294.
109. Lowe, F.J., et al., *A novel autoantibody test for the detection of pre-neoplastic lung lesions.* Mol Cancer, 2014. **13**: p. 78.
110. Mu, Y., et al., *Efficacy of autoantibodies combined with tumor markers in the detection of lung cancer.* J Clin Lab Anal, 2022. **36**(8): p. e24504.
111. Ouyang, R., et al., *Clinical value of tumor-associated antigens and autoantibody panel combination detection in the early diagnostic of lung cancer.* Cancer Biomark, 2021. **32**(3): p. 401-409.
112. Ren, S., et al., *Early detection of lung cancer by using an autoantibody panel in Chinese population.* Oncoimmunology, 2018. **7**(2): p. e1384108.
113. Zhang, X., et al., *A Diagnostic Model With IgM Autoantibodies and Carcinoembryonic Antigen for Early Detection of Lung Adenocarcinoma.* Frontiers in immunology, 2022. **12**: p. 728853.
114. Cazzoli, R., et al., *microRNAs derived from circulating exosomes as noninvasive biomarkers for screening and diagnosing lung cancer.* Journal of thoracic oncology, 2013. **8**(9): p. 1156-1162.
115. D'Ambrosi, S., et al., *Combinatorial Blood Platelets-Derived circRNA and mRNA Signature for Early-Stage Lung Cancer Detection.* Int J Mol Sci, 2023. **24**(5).
116. Dong, X., et al., *Plasma miR-1247-5p, miR-301b-3p and miR-105-5p as potential biomarkers for early diagnosis of non-small cell lung cancer.* Thoracic Cancer, 2021. **12**(4): p. 539-548.
117. Dong, X., et al., *Tumor-educated platelet SNORD55 as a potential biomarker for the early diagnosis of non-small cell lung cancer.* Thoracic Cancer, 2021. **12**(5): p. 659-666.
118. Dou, Y., et al., *Plasma small ncRNA pair panels as novel biomarkers for early-stage lung adenocarcinoma screening.* BMC genomics, 2018. **19**(1): p. 1-10.
119. Duan, X., et al., *Circulating miRNAs in Serum as Biomarkers for Early Diagnosis of Non-small Cell Lung Cancer.* Frontiers in Genetics, 2021. **12**.
120. Fan, L., et al., *Evaluation of serum paired microRNA ratios for differential diagnosis of non-small cell lung cancer and benign pulmonary diseases.* Molecular Diagnosis & Therapy, 2018. **22**(4): p. 493-502.
121. Gupta, C., et al., *Sputum long non-coding RNA biomarkers for diagnosis of lung cancer.* Cancer Biomarkers, 2019. **26**: p. 219-227.
122. Jiang, Y.F., et al., *Evaluation of circulating small extracellular vesicle-derived miRNAs as diagnostic biomarkers for differentiating between different pathological types of early lung cancer.* Sci Rep, 2022. **12**(1): p. 17201.
123. Kim, J.O., et al., *Non-small cell lung cancer detection using microRNA expression profiling of bronchoalveolar lavage fluid and sputum.* Anticancer Res, 2015. **35**(4): p. 1873-80.
124. Li, N., et al., *Digital PCR quantification of miRNAs in sputum for diagnosis of lung cancer.* Journal of Cancer Research and Clinical Oncology, 2014. **140**(1): p. 145-150.
125. Lin, Y., et al., *A classifier integrating plasma biomarkers and radiological characteristics for distinguishing malignant from benign pulmonary nodules.* International journal of cancer, 2017. **141**(6): p. 1240-1248.

126. Ma, J., et al., *A prediction model based on biomarkers and clinical characteristics for detection of lung cancer in pulmonary nodules.* Translational oncology, 2017. **10**(1): p. 40-45.
127. Razzak, R., et al. *MicroRNA Expression Profiling of Sputum for the Detection of Early and Locally Advanced Non-Small-Cell Lung Cancer: A Prospective Case–Control Study*. Current Oncology, 2016. **23**, 86-94 DOI: 10.3747/co.23.2830.
128. Roa, W.H., et al., *Sputum microRNA profiling: a novel approach for the early detection of non-small cell lung cancer.* Clinical and Investigative Medicine, 2012: p. E271-E281.
129. Tulinsky, L., et al., *Overexpression of the miR-143/145 and reduced expression of the let-7 and miR-126 for early lung cancer diagnosis.* J Appl Biomed, 2022. **20**(1): p. 1-6.
130. Wang, K., et al., *Plasma SNORD42B and SNORD111 as potential biomarkers for early diagnosis of non-small cell lung cancer.* J Clin Lab Anal, 2022. **36**(11): p. e24740.
131. Wang, W., et al., *Early Detection of Non-Small Cell Lung Cancer by Using a 12-microRNA Panel and a Nomogram for Assistant Diagnosis.* Frontiers in Oncology, 2020. **10**.
132. Wu, Y., H. Jing, and J. Zhang, *MicroRNA-340 and MicroRNA-450b-5p: Plasma Biomarkers for Detection of Non-Small-Cell Lung Cancer.* J Environ Public Health, 2022. **2022**: p. 8024700.
133. Wu, Q., et al., *Combination of serum miRNAs with serum exosomal miRNAs in early diagnosis for non-small-cell lung cancer.* Cancer management and research, 2020. **12**: p. 485.
134. Xing, L., et al., *Sputum microRNA biomarkers for identifying lung cancer in indeterminate solitary pulmonary nodules.* Clinical Cancer Research, 2015. **21**(2): p. 484-489.
135. Yu, L., et al., *Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers.* International Journal of Cancer, 2010. **127**(12): p. 2870-2878.
136. Zhou, H., et al., *Identification of Small Nucleolar RNA SNORD60 as a Potential Biomarker and Its Clinical Significance in Lung Adenocarcinoma.* Biomed Res Int, 2022. **2022**: p. 5501171.
137. Abou-Zeid, A., et al., *HOXA9 gene promotor methylation and copy number variation of SOX2 and HV2 genes in cell free DNA: A potential diagnostic panel for non-small cell lung cancer.* BMC Cancer, 2023. **23**(1): p. 329.
138. Chen, C., et al., *Ultrasensitive DNA hypermethylation detection using plasma for early detection of NSCLC: a study in Chinese patients with very small nodules.* Clinical Epigenetics, 2020. **12**: p. 1-11.
139. Chen, L., et al., *Combined use of EpCAM and FRα enables the high-efficiency capture of circulating tumor cells in non-small cell lung cancer.* Scientific Reports, 2018. **8**(1): p. 1188.
140. Gao, L., et al., *Methylated APC and RASSF1A in multiple specimens contribute to the differential diagnosis of patients with undetermined solitary pulmonary nodules.* Journal of thoracic disease, 2015. **7**(3): p. 422.
141. Leung, M., et al., *Blood‐based circulating tumor DNA mutations as a diagnostic and prognostic biomarker for lung cancer.* Cancer, 2020. **126**(8): p. 1804-1809.
142. Hulbert, A., et al., *Early detection of lung cancer using DNA promoter hypermethylation in plasma and sputum.* Clinical Cancer Research, 2017. **23**(8): p. 1998-2005.

143. Hubers, A., et al., *DNA hypermethylation analysis in sputum for the diagnosis of lung cancer: training validation set approach.* British journal of cancer, 2015. **112**(6): p. 1105-1113.
144. Paci, M., et al., *Circulating plasma DNA as diagnostic biomarker in non-small cell lung cancer.* Lung Cancer, 2009. **64**(1): p. 92-97.
145. Wan, L., et al., *Circulating tumor cell and metabolites as novel biomarkers for early-stage lung cancer diagnosis.* Frontiers in Oncology, 2021. **11**: p. 630672.
146. Xue, Y., et al., *Folate-receptor-positive circulating tumor cells as an efficacious biomarker for the diagnosis of small pulmonary nodules.* Journal of cancer research and therapeutics, 2018. **14**(7): p. 1620.

Table 1 Characteristics of included studies

Author	Country	Sample size	Comparative group	Specimen	Name of protein(s) evaluated	Method of detection	Sensitivity %	Specificity %	AUC 95% CI
Antigens									
Ajona et al., 2021 [39]	Spain	78	NSCLC/indeterminate nodules	P	C4c, CYFRA 21-1, and CRP	ELISA	82	95	0.9
Bigbee et al., 2012 [88]	USA	56/30	NSCLC/indeterminate nodule	S	Prolactin, transthyretin, thrombospondin-1, E-selectin, C-C motif chemokine 5, macrophage migration inhibitory factor, plasminogen activator inhibitor, receptor tyrosine-protein kinase, erbb-2, cytokeratin fragment 21.1 and serum amyloid A	Luminex xMAP immunoassay	77.10	76.2	NG
Fahrmann et al., 2022 [89]	USA	1,299	High risk	S	A four-marker protein panel (4MP) consisting of CA125, CEA, SPA, CYFRA21-1	bead-based immunoassays	91.5	45.4	0.79
Farlow et al., 2010a [26]	USA	90/43	NSCLC/non-cancer	S	TNF-α, CYFRA 21-1, IL-1ra , MMP-2,	Luminex xMAP immunoassay	99	95	0.979

					MCP-1 and sE-selectin				
Gasparri et al., 2023 [29]	Italy	46/41	NSCLC/high-risk	S	ARSA, PRKCA, ACTR3B, and CD59	MS	94.83	93.56	0.98
Goebel et al., 2019 [87]	Multiple -UK, Russia, Ukraine	1,479	NSCLC/HC	P	CA-125, CEA, CYFRA21-1, EGFR/HER1/ErBB 1, Gro-Pan, HGF, IL-10, IL-12p70, IL-16, IL-2,IL-4, IL-5, IL-7, IL-8, IL-9, Leptin, LIF, MCP-1, MIF, MIG, MMP7,MP9, MPO, NSE, PDGF-BB, Rantes, Resistin, sFasL, SAA,sCD40-ligand, sICAM-1, TNFRI and sTNFRII.	Multiplex immunoassay platform	80	95	0.963
Higgins et al., 2012 [33]	UK & USA	35/170/160	LC/inflammatory diseases	T/P	Ciz1	SDS-PAGE	95	74	0.958
Jeong et al., 2021 [31]	Korea	70/16	NSCLC/HC	P	exosomal GCC2	ELISA	90	75	0.844
Joseph et al., 2012 [34]	USA	1,182	NSCLC/nodules	P	OPNV	ELISA	80	88	0.88
Jung et al., 2017 [43]	Korea	200/150	LC/control group	S	EGFR1, MMP7, CA6, KIT, CRP, C9 and SERPINA3	Proteomic	75	91.70	0.82/0.77
Kupert et al., 2011 [90]	USA	145	NSCLC/BN/HC	P	secretory phospholipase A2-IIa	ELISA	48-67	86	0.68-0.86

Lai et al., 2022 [91]	China	201/112/94	NSCLC/HC/Nodules	S	CEA, Cyfra21-1, CST1	ELISA	88.4	89.1	0.92
Li et al., 2023 [92]	China	37/11	NSCLC/HC	P	MDK, WFDC2, and CXCL14	Luminex technology	NG	NG	0.96
Li et al., 2022 [93]	China	98/100	NSCLC/BLD	S	CA153 + CA125 + CEA + TNF − alpha + hs − CRP	Immunoluminescence analyser	66.82	93.51	NG
Ma et al., 2021 [30]	China	318/239	NSCLC stage I/HC	urine	MDH2	ELISA	70.13	66.11	0.768
		769/493					68.92	58.22	0.723
Meng et al., 2023 [94]	China	60/15	NSCLC/HC	S	EpCAM and CEA	Ratiometric biosensor for exosome	93.3	86.7	0.916
Nolen et al., 2011 [95]	USA	172	LC/high risk	S	MIF, TTR, THSP, sVCAM-1 and tPAI-1	Multiplexed bead-based immunoassays	70/74	90/93	0.85/0.894
Pakvisal et al., 2022 [96]	Thailand	76/12/53	NSCLC/BLD/ HC	S	C5AR1, CLEC4A and NLRP3 specific to CD3	Flow cytometry	71.5	70	NG
Pio et al., 2010 [97]	Spain	56/22	NSCLC/BLD	Sputum/ BAL	Complement factor H	ELISA	80-sputum 82-BAL	88-sputum 77-BAL	NG
Sun et al., 2020 [98]	China	1223	NSCLC/BPC/ OC/ HC	S	IDH1	ELISA	63.3/55	86.8/86.3	0.907/0.788
Sun et al., 2009 [65]	China	71/62	NSCLC/BLD	Sputum	A proliferation-inducing ligand (APRIL)	Immunocytochemistry	82	97	NG
Song et al., 2022 [99]	Korea	30/15	NSCLC/HC	P	p53-anti-p53-autoantibody complex	Labelled immunoassay	81.6	93.3	NG
Wang et al., 2017 [36]	China	350/411	NSCLC/control (BLD, HC)	S	MIC-1, Cyfra21-1, CA125 and CEA	Immunoassay/ELISA	84.40	90	0.957
Wang et al., 2013 [38]	China	132/48/92	LC/BLD/HC	S	NSE + CEA + CYFRA21-1	Electrochemical luminescence	75.76	88.57	0.63

Wieskopf et al., 1995 [100]	France	161/97	LC/BD	S	CYFRA 21-1	Immunoradiometric assay	59	94	0.85
Wu et al., 2020a [101]	Taiwan	102/84	ADC/HC	P	Beta-1,4-galactosyltransferase 1, CD44 antigen, eukaryotic initiation factor 4A-I, galectin-1, mucin-16, protein disulfide-isomerase A3, and vimentin	LC-MRM-MS assay	97.2	61	0.76
Yang et al., 2020b [27]	China	370/110	NSCLC stage I/BLD	S	Ferritin, CA125, CEA, NSE and CYFRA21-1	Electrochemiluminescence	92.97	90	0.95
Yu et al., 2023 [102]	China	513	Nodules	S	ACSL4	ELISA	65.1	90.2	0.762
Yuan et al., 2022 [28]	China	175/160	LC/HC	P	HSP90α, CEA	ELISA	95.63	99.97	0.996
Zhang et al., 2022a [103]	China	78/44	NSCLC/BLD	S	CEA, Cyfra21-1, miR3149 and miR-4769.3p	Flow fluorescence immunoanalyser, qPCR	88.46	81.82	0.898
Autoantibodies									
Chen et al., 2021 [45]	China	458	NSCLC/non-malignant nodules/HC	S	MAGEA1, PGP9.5, SOX2, and TP53	ELISA	71.8	89	0.89
Doseeva, et al., 2015 [52]	USA	230/150	NSCLC/BLD	S	One autoantibody marker (NY-ESO-1) and three Ags (CEA,	Luminex xMAP technology	74/77	80/80	0.81/0.85

					CA-125, and CYFRA 21–1)				
Du et al., 2018 [17]	China	397	LC/nodules	S	Seven TAAs (p53, PGP9.5, SOX2, GAGE7, GBU4-5, CAGE and MAGEA1)	ELISA	56.53	91.60	NG
Ezzatifar et al., 2022 [104]	Iran	190/30	NSCLC/HC	S	Nucleolin	ELISA	85	96.67	0.948
Farlow et al., 2010b [40]	USA	16/196	NSCLC/COPD/non malignant nodules/NC	S	IMPDH, phosphoglycerate mutase, ubiquillin, Annexin I, Annexin II, and HSP70-9B	Proteomic/Luminex-based "direct-capture" immunobead assays	94.8	91.1	0.964
Hua et al., 2022 [105]	China	83/26	NSCLC/BLD	S	7-TAAbs (P53, PGP9.5, SOX2, GAGE7, GBU4–5, MAGEA1 and CAGE)	ELISA	55.44	87.5	0.65
Huo et al., 2020 [106]	China	121/34/100	NSCLC/HC/nodules	S	7AAb (GAGE7, CAGE, MAGEA1, SOX2, GBU4-5, PGP9.5, and p53)	ELISA	45.5	85.3	0.66
Lastwika, et al., 2019 [107]	USA	20/10/250	LC/nodules	T/P	IgG: EPB41L3, ANKRD36B, FGCR2A, and	ELISA	50	70	0.74/0.78

					LINGO1; IgM: S100A7L2				
Li et al., 2021 [64]	USA	30/30	NSCLC/control	Sputum	DDX6, ENO1, and 14–3-3 ζ (protein zeta)	Array/ELISA	81	83	0.87
Liu et al., 2020 [108]	China	211/200	NSCLC/HC	P	CD25-MUC1-VEGFR1	ELISA	49.6	95	0.883
Lowe, et al., 2014 [109]	USA	600	AAH & SCD	S	AAH: LTBP1*, BMI1*, GAGE7*, AGBL5 HES1*	SEREX	86	78	0.81/0.88
Jiang et al., 2021 [41]	China	150	LC /HC/BLD	S	7AAb (TP53, NPM1, FGFR2, PIK3CA, GNA11, HIST1H3B, and TSC1)	Protein array/ELISA	94.4	84.9	0.897
		744					89.4	78.2	0.838
Mu et al., 2022 [110]	China	633/147	NSCLC/BLD	S	7-TAAbs+SCCA+CYFRA21-1	Chemiluminescence immunoassay	37.76	81.84	0.648
Ouyang et al., 2021 [111]	China	443	NSCLC /HC/BLD	S	7 AAB, CEA, CYFRA 21-1	ELISA	52.26	77.46	0.686
		569					44.02	83	0.668
Pan et al., 2020 [42]	China	69/30/25	NSCLC stage I/HC/BLD	S	IgA autoantigens (i.e. BCL7A, and TRIM33 and MTERF4) and three IgG autoantigens (i.e. CTAG1A, DDX4 and MAGEC2)	ELISA	73.5	>85	0.503
		88/36/18					68.2	87	0.673

Ren, et al., 2018 [112]	China	2008	LC/patients (GGNs) and/or solid nodules	S	p53, GAGE7, PGP9.5, CAGE, MAGEA1, SOX2 and GBU4-5	ELISA	59/62	90	0.781
Song et al., 2019 [49]	Korea	170	NSCLC/HC	P	CYFRA 21-1–anti-CYFRA 21-1 autoantibody immune complex (CIC) and free CYFRA 21-1	Labelled immunoassay	76	87.5	NG
Yang et al., 2020a [13]	China	(42)61/24/29	LC/BLD/HC	S	HE4	ELISA	54.76	96.23	0.848
Zang et al., 2019 [51]	China	176/140	LC/HC	S	CEA, CA125, Annexin A1-Ab, and Alpha enolase-Ab	Multiplexed serum immunoassays	86.5	82.3	0.897
Zhang et al., 2022b [113]	China	68/68	ADC/HC	S	CEA, 5 IgM AAB (TSHR, ERBB2, survivin, PIK3CA, and JAK2)	ELISA	56.63	93.98	0.744
Zhong et al., 2006 [35]	USA	46 102	Stage I NSCLC & risk-matched control	P	PXN, SEC15L2, BAC clone RP11-499F19, XRCC5, and MALAT1	Phage library/Affymetric array	100 91.3	95.7 91.3	0.99
miRNAs & RNA									
Cazzoli, et al., 2013 [114]	USA	30/105	LC/BD/HC	P	miR-151a-5p, miR-30a-3p, miR-200b-5p, miR-629, miR-	RT-PCR	97.5/96	72/60	0.76

					100, and miR-154-3p				
D'Ambrosi et al., 2023 [115]	Netherlands	30/27/3	NSCLC/HC/Nodules	B	2 circRNAs (circSLC8A1 and circCHD9) and 3 mRNAs (PSMB9, RUNX1, and LILRB1)	RNAseq	85	86	0.96
Dong et al., 2021a [116]	China	300	NSCLC/HC	P	CEA, miR-1247-5p, miR-301b-3p, and miR-105-5p	RT-PCR	88.4	64.7	0.815
Dong et al., 2021b [117]	China	290/105	NSCLC/HC	P	CEA, TEP SNORD55	RT-PCR	66.3	90	0.828
Dou et al., 2018 [118]	USA	50/35/29 44/32/51	ADC I,II/BLD/HC	P	hsa-miR-101-3p/hsa-miR-126-5p	Sequencing/PCR	81.1/70.4	78.1/72.7	0.82/0.742
Duan et al., 2021 [119]	China	12/120	NSCLC/HC	S	miR-492, miR-590-3p, and miR-631	RT-PCR	86.7	71.7	0.828
Fan, et al., 2018 [120]	China	128/193	NSCLC/BPD	S	Five miRNA ratios-miR-15b-5p/miR-146b-3p, miR-20a-5p/miR-146b-3p, miR-19a-3p/miR-146b-3p, miR-92a-3p/miR-146b-3p, and miR-16-	RT-PCR	70	90	0.79

					5p/miR-146b-3p)				
Gupta et al., 2019 [121]	USA	67/65 59/60	NSCLC/HC	Sputum	Three lncRNAs (SNHG1, H19, and HOTAIR)	qRT-PCR	82.09 81.36	89.23 88.33	0.80
Hennessey, et al., 2011 [54]	USA	50/130	NSCLC/HC	S	miR-15b and miR-27b	RT-PCR	100	84	0.98
Jiang et al., 2022 [122]	China	35/15	NSCLC/HC	P	miR-152-3p and miR-1277-5p	qRT–PCR	73.3	86.7	0.79
Kim et al., 2015 [123]	Canada	21/10	NSCLC/HC	BAL/sputum	5 miRNAs (miR-21, miR-143, miR-155, miR-210, and miR-372)	qRT-PCR	85.7 BAL 67.8 sputum	100-BAL 90-sputum	NG
Li et al., 2019 [57]	China	64/40	NSCLC /HC	S	CEA+Exo-GAS5	RT-PCR	89.06	90.00	0.919
Li et al., 2014 [124]	USA	35/40	NSCLC stage I/HC	Sputum	miR-31 and miR-210	Digital PCR	80.6	91.7	0.89
Lin, et al., 2017 [125]	USA	135 126	Indeterminate nodules	P	11 (miR-21–5p miR-103a-3p miR-126–3p miR-135a-5p miR-145–5p miR-141–3p miR-193b-3p miR-200b-3p miR-205–5p)	Microarray and droplet digital PCR (ddPCR)	89.90 73.5	90.90 75.5	0.91
Ma, et al., 2017 [126]	USA	1272 111	Indeterminate nodules	B	miRs-19b-3p and -29b-3p	qRT-PCR	80.30 72.6	89.40 81.9	0.91

Razzak et al., 2016 [127]	Canada	21/10	NSCLC/HC	Sputum	miR-21, miR-210, and miR-372	qRT-PCR	67	90	NG
Reis et al., 2020 [55]	Canada	54/40	Early NSCLC/HC	P	miR-16-5p, miR-92a-3p, miR-451a	RT-PCR	84	100	0.87
Roa et al., 2013 [128]	Canada	24/6	NSCLC/HC	Sputum	miR-21, miR-155, miR-210, miR-143, miR-372	qRT-PCR	83.3	100	NG
Su et al., 2016 a [69]	China	117/174	NSCLC stage I/control (PN)	Sputum	2 miRNAs (miR-31 and miR-210) and methylation of 2 genes (RASSF1A and 3OST2)	qRT-PCR	87.3	90.3	0.93
Su et al., 2016 b [68]	USA	117/103	NSCLC stage I/control (PN)	Sputum	miRs-21, 31, and 210 + small nucleolar RNA (snoRDs-66 and 78)	qRT-PCR	89	89	0.94
Tulinský et al., 2022 [129]	Czech	60/60	NSCLC/HC	P	miR-126, miR-143, miR-145, let-7a and let7g	qRT-PCR	75-85	75-85	0.90-0.93
Wang et al., 2022 [130]	China	165/118	NSCLC/HC	P	SNORD42B and SNORD111	qRT-PCR	61.8	77.1	0.719
Wang et al., 2020 [131]	China	82	pulmonary nodules both benign and malignant	P	miRNA-17, miRNA-146a, miRNA-200b, miRNA-182, miRNA-155, miRNA-221,	RT-PCR	50	92.9	0.896

					miRNA-205, miRNA-126, miRNA-7, miRNA-21, miRNA-145, and miRNA-210				
Wu et al., 2022 [132]	China	100/100	NSCLC/HC	P	miR-340 and miR-450b-5p	qRT-PCR	78.33	77.5	0.862
Wu, et al., 2020b [133]	China	48/48/32	NSCLC I/II/HC/ LBL	S	Four serum miRNAs including miR-21-5p, miR-141-3p, miR222-3p, and miR-486-5p, and 2 serum exosomal miR-146a-5p and miR-486-5p	qRT-PCR	85.42	92.50	0.96
Xie et al., 2010 [66]	USA	23/17	NSCLC/cancer free	Sputum	mir-21	qRT-PCR	47.82	100	NG
Xing et al., 2010 [67]	USA	67/55	SCC/HC	Sputum	miR-205, miR-210 and miR-708	Microarray/ qRT-PCR	73	96	0.87
Xing, et al., 2015 [134]	USA	122/136/ 155	Indeterminate solid nodules	Sputum	miR205/miR708/ miR375/miR200b/ miR182/ miR155/ miR372 miR143(miRs21, 31, and 210)	RT-PCR	82.93/82.09/ 80.52	87.84/88.41/86.08	0.919

Xing, et al., 2019 [56]	China	17/534	NSCLC/control (BN/HC)	S	ITGA2B belongs TEP	RNA-seq/ q-PCR and ddPCR	92.8/91.2	78.6/56	0.892
Yu et al., 2010 [135]	USA	64/58	NSCLC/HC	Sputum	miR-21, miR-486, miR-375 and miR-200b	qRT-PCR	69.22	81.7	0.83
Zhou et al., 2022 [136]	China	15	ADC	P	SNORD60	qRT-PCR	74.2	75.3	0.828
ctDNA & CTC									
Abou-Zeid et al., 2023 [137]	Egypt	25/25	NSCLC/HC	P	HOXA9, SOX2, HV2	qRT-PCR	88	100	0.958
Carozzi et al., 2017 [59]	Italy	1356	LC/smokers/ ex-smokers	P/sputum	MSI/LOHs loci, with the loci 1 to 5 (3p14.2, 3p21-p23, 3p26.1, 3p13, 5q15) and 7 to 9 (9p22-p23, 9p21, 13q12.3)	PCR	90	71	NG
Chen et al., 2020 [138]	China	161	Nodules	P	CDO1, SOX17 and HOXA7	QMSP	90	71	NG
Chen et al., 2018 [139]	China	41/10	NSCLC/HC	B	EpCAM and Folate receptor alpha (FRα)	Immunomagnetic separation method	75.61	90	NG
Gao et al., [140]	China	89	Nodules	S/P/T	APC, RASSF1A	QMSP	56.9	90.3	0.81
Leung et al., 2020 [141]	UK	211	NSCLC/HR	P/T	ctDNA (EGFR, KRAS, and TP53 mutation)	RTPCR	75	89	NG
Hulbert et al., 2017 [142]	Netherlands	150/60	NSCLC stage I IIA/BLD	P/sputum	P-CDO1, TAC1, and SOX17	SMART-MSP	93-P 98-sputum	62-P 71-sputum	0.77-P 0.89

					Sputum-(TAC1, HOXA17, and SOX17)				
Hubers et al., 2015 [143]	Netherlands	159/154	NSCLC stage I IIA/BLD	Sputum	TAC1, HOXA7, SOX17	QMSP	67.1 42.5	89.5 96.5	0.69
Paci et al., 2009 [144]	Italy	151/79	NSCLC/HC	B	Amplification of hTERT	qRT-PCR	85.8	46.8	0.79
Su et al., 2016 a [69]	China	117/174	NSCLC stage I/control (PN)	Sputum	RASSF1A, 3OST2 and PRDM14	QMSP	82.9 45.3	76.4 86.2	0.79 0.68
Wan et al., 2021 [145]	China	48	NSCLC	B	NOTCH1, IGF2, EGFR, and PTCH1	CellCollector® in vivo CTC capture	65.85	62.5	NG
Xue, et al., 2018 [146]	USA	(31)72/26	NSCLC/control	P	FR+CTC	immunomagnetic leukocyte depletion/PCR	74.19	73.08	0.8221
Yang, et al., 2019 [60]	China	50	Nodules	P	methylation of 8 genes (CDH13, WT1, CDKN2A, HOXA9, PITX2, CALCA, RASSF1A, and DLEC1)	QMSP	72	91	NG
Zhong et al., 2021 [63]	China	18	Solid nodules	S	(CEP8) CTC, CA125 or NSE	electrochemiluminescence /FISH	83	100/83 with NSE	NG

AAH: Atypical adenomatous hyperplasia, ARSA: Arysulfatase A, AUC: area under curve, B: blood, BAL: bronchioalveolar lavage, BN: benign nodules, BPC: benign pulmonary condition, BLD: benign lung diseases, Bmi-1: B-lymphoma Moloney murine leukaemia virus insertion region-1, C4c : complement-derived fragment, CI: confidence interval,

CRP: C-reactive protein, CTC: circulating tumour cells, Ciz1: nuclear matrix-associated DNA replication factor, CXCL14: C- X-C motif chemokine ligand 14, ddPCR: droplet digital PCR, FR: folate receptor, FOXL2: fork-head box L2 gene, HC: healthy control, HE4: Human epididymis secretory protein 4, HES1: mammalian hairy and Enhancer-of-split homologues 1, IDH1: isocitrate dehydrogenase 1, ILra: interleukin-1ra, LTBP1: Latent Growth Factor Beta Binding Protein, MCP1: monocyte chemotactic protein-1, MDH2: malate dehydrogenase 2, MDK: Midkine, MIC-1: Macrophage inhibitory cytokine-1, MIF: macrophage migration inhibitory factor, MMP2 : matrix metalloproteinase-2, MSI/LOH: genomic instability loss of heterozygosity/ microsatellite instability, ncRNA: non-coding RNA, NG: not given, NSCLC: non-small cell lung cancer, OC: other cancers, OPV: OPN velocity, P: plasma, PTGER4: prostaglandin E receptor 4 gene, q-PCR: quantitative real time PCR, QMSP: real-time quantitative methylation-specific polymerase chain reaction, SCD: squamous cell dysplasia, S: serum, SHOX2: methylation of short stature homeobox 2 gene, sVCAM-1: soluble vascular cell adhesion molecule, T: tissue, TC: Training cohort, TB: tuberculosis, TEP: tumour-educated platelets, THSP: thrombospondin, TNF-α: tumour necrosis factor α, tPAI-1: tissue plasminogen activator inhibitor 1, TTR: transthyretin, VC: Validation cohort, WFDC2: WAP four-disulphide core domain 2. The number between () represents number of NSCLC patients with stages I and II.

Table 2. Studies with sensitivity and specificity > 90%

Group	Study	Sensitivity TC/VC	Specificity TC/VC	AUC	Biomarker panel
Antigens	Farlow et al., 2010a [26]	99	95	0.979	TNF-α, CYFRA 21-1, interleukin-1ra, MMP-2, monocyte chemotactic protein-1 and sE-selectin
	Yang et al., 2020b [27]	92.97	90	0.95	Ferritin, CA125, CEA, NSE and CYFRA21-1
	Yuan et al., 2022 [28]	95.63	99.97	0.996	HSP90α and CEA
	Gasparri et al., 2023 [29]	94.83	93.56	0.98	ARSA, PRKCA, ACTR3B and CD59
Auto antibodi	Zhong et al., 2006 [35]	100/ 91.3	95.7/ 91.3	0.99	Paxillin, SEC15L2, BAC clone RP11-499F19, XRCC5 and MALAT1
	Farlow et al., 2010b [40]	94.8	91.1	0. 964	IMPDH, phosphoglycerate mutase, ubiquillin, Annexin I, Annexin II and HSP70-9B

TC: Training cohort; VC: Validation cohort

Table 3. Averages of the sensitivity and specificity of all possible lung-cancer biomarkers

	Sensitivity % ±SD	**Specificity %±SD**
Antigens	77.2 ± 10.1	86.08 ± 17.5
Antibodies	79.4 ± 15.2	77.33 ± 5.8
miRNA	79.83 ± 8.9	90.33 ± 12.5
ctDNA & CTC	81.43 ± 5.7	84.15 ± 4.8

Figure 1

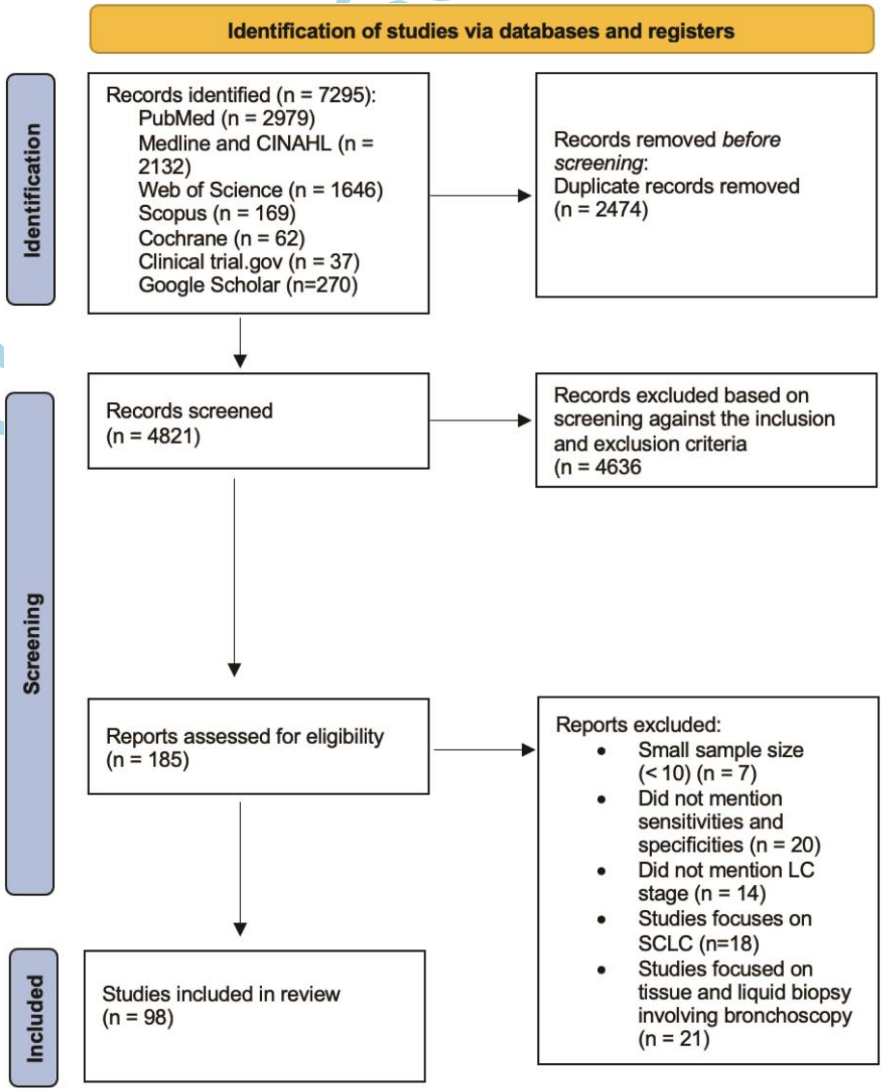


Figure 2a

Figure 2b

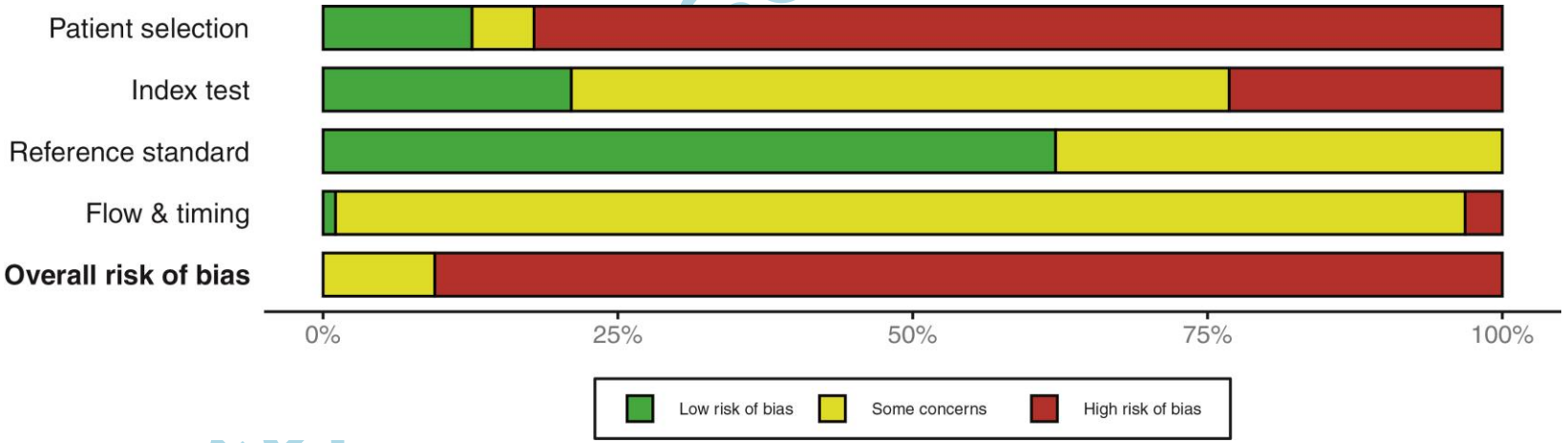


Figure 3a

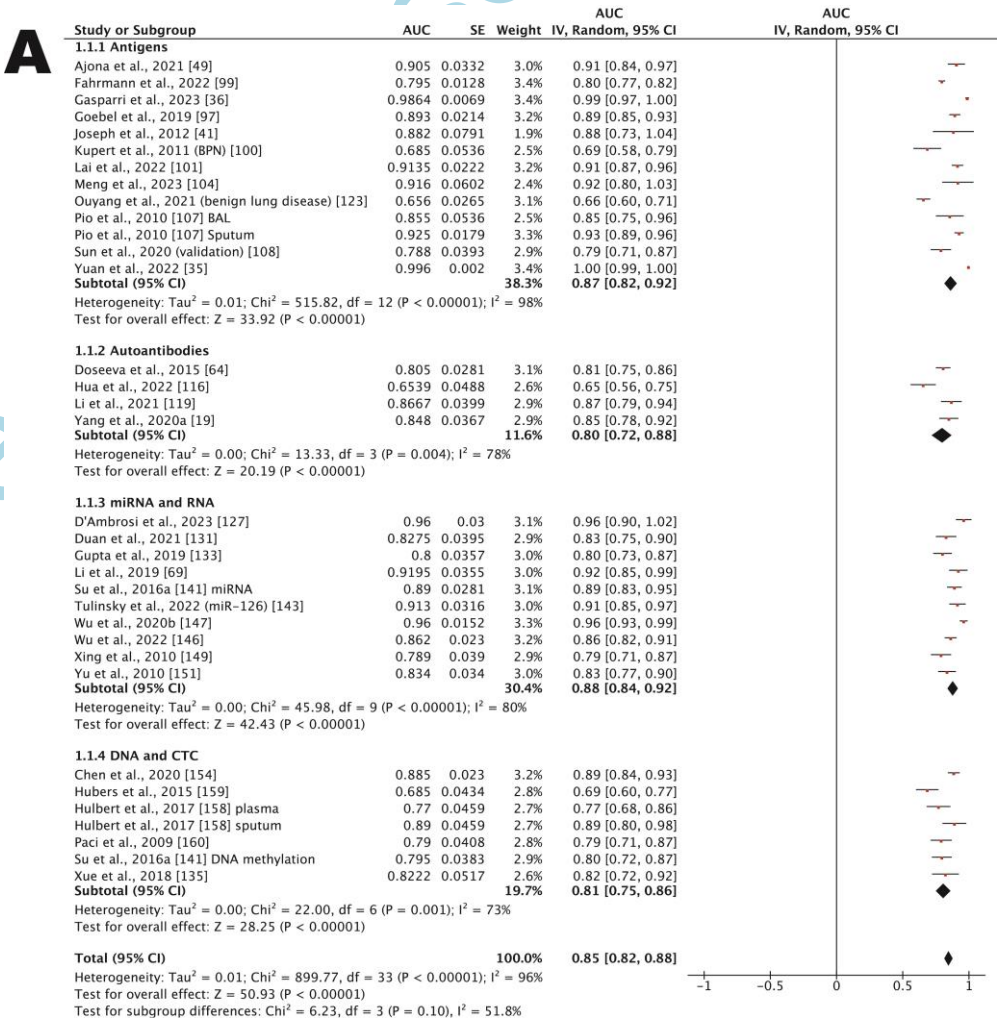


Figure 3b

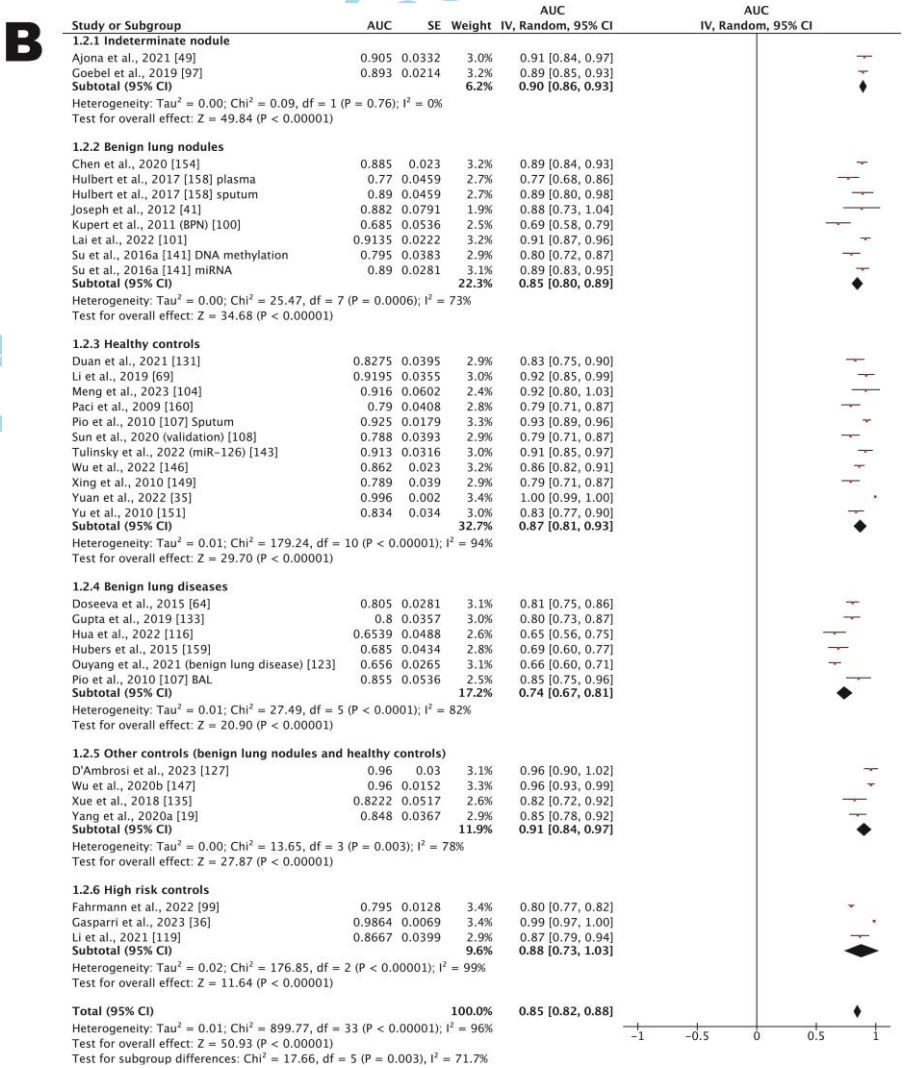


Figure 3c

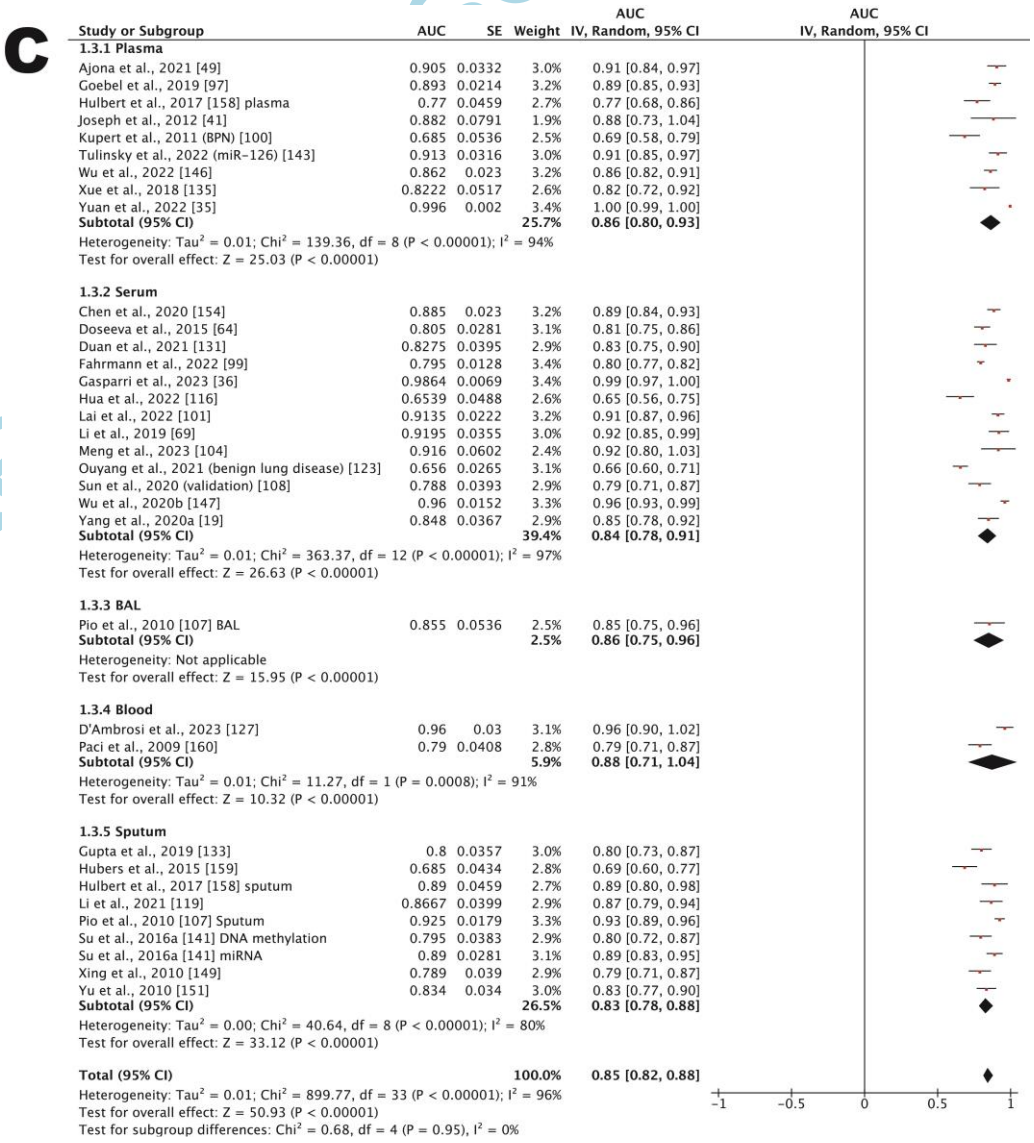


Figure 3d