Biomarkers for early diagnosis, prognosis, prediction, and recurrence monitoring of non-small cell lung cancer

Abstract: Despite advances in the management of non-small cell lung cancer, it remains to be the leading cause of cancer-related deaths worldwide primarily because of diagnosis at a late stage with an overall 5-year survival rate of 17%. A reduction in mortality was achieved by low-dose computed tomography screening of high-risk patients. However, the benefit was later challenged by the high false positive rate, resulting in unnecessary follow-ups, thus entailing a burden on both the health care system and the individual. The diagnostic dilemma imposed by imaging modalities has created a need for the development of biomarkers capable of differentiating benign nodules from malignant ones. In the past decade, with the advancements in high-throughput profiling technologies, a huge amount of work has been done to derive biomarkers to supplement clinical diagnosis. However, only a few of them have efficient sensitivity and specificity to be utilized in clinical settings. Therefore, there is an urgent need for the development of sensitive and specific means to detect and diagnose lung cancers at an early stage, when curative interventions are still possible. Due to the invasiveness of tissue biopsies and inability to capture tumor heterogeneity, nowadays enormous efforts have been invested in the development of technologies and biomarkers that enable sensitive and cost-effective testing using substrates that can be obtained in a noninvasive manner. This review, primarily focusing on liquid biopsy, summarizes all documented potential biomarkers for diagnosis, monitoring recurrence treatment response.

Keywords: liquid biopsy, circulating-tumor DNA, circulating tumor cells, non-small cell lung cancer, exosomes, biomarkers

Introduction

Despite the progress made in the treatment of non-small cell lung cancer (NSCLC) over the past few decades, the 5-year survival rate has not improved much primarily because of initial diagnosis at a late stage. Currently, its detection primarily relies on computed tomography (CT) scans in combination with sputum cytology test. The 5-year survival rate, ranging from 50% for stage IA to 2% for stage IV, improves significantly when diagnosed at an early stage.1 However, only one third of patients with NSCLC are diagnosed at an early stage when the disease is still localized and potentially curable. Most cases are diagnosed with locally advanced or metastatic disease.2

Lung cancer screening aiming to detect malignant nodules at an early stage has been applied to improve the survival rate. Recent implementation of low-dose computed tomography (LDCT) screening has been shown to diagnose lung cancer at an earlier stage with survival benefits, and therefore, it is recommended to perform on healthy individuals with high risk, such as current and former smokers.3 LDCT has
been shown to double the early detection rate (from 30% to 60%) and reduce mortality rate by 20% comparing to chest radiography.\textsuperscript{3–5} However, the reduction in mortality was challenged by a significant increase in false positives, resulting in unnecessary follow-ups, thus entailing a burden on the health care system and the individual.\textsuperscript{6–8} National lung screening trial\textsuperscript{9} revealed that in order to prevent one case of lung cancer-related mortality, 320 individuals with high risk need to undergo LDCT. Therefore, there is an urgent need to derive biomarkers to better stratify high-risk individuals, thus maximizing the benefits from LDCT screening.

In addition to LDCT, many biomarkers, especially molecular biomarkers, were developed to supplement clinical diagnosis. In the past few decades, researchers explored tumor-specific alterations primarily by tissue biopsies to interrogate the genome and epigenome of NSCLC to derive biomarkers for diagnosis, prognosis, and monitoring. However, tumor biopsy is not only invasive but also biased with limited ability to reflect tumor heterogeneity, owing to its temporal and spatial snapshot nature. Liquid biopsy has received enormous attention in the past few years owing to obvious clinical advantages.\textsuperscript{9,10} Circulating tumor DNA (ctDNA), composed of small fragments of nucleic acid released from apoptotic or necrotic tumor cells, can reflect the genetic profile of a tumor.\textsuperscript{11,12} Numerous studies have shown that ctDNA can be used as a surrogate for patient stratification, diagnosis, and disease monitoring.\textsuperscript{13} Nowadays, with improved understanding of molecular features associated with NSCLC and the availability of new technologies that have enabled sensitive and specific biomarker analysis using liquid biopsy samples, it is becoming promising to use biomarkers detected in blood and/ or other types of body fluids for NSCLC diagnosis, treatment guidance, prognosis prediction, and recurrent monitoring.

A large variety of substrates in the circulation can be obtained from liquid biopsy samples and used for biomarker analysis, such as circulating tumor cells (CTCs), ctDNA, microRNA (miRNA), and serum proteins. With the efficacies achieved by immune checkpoint inhibitors, immune profiling of peripheral blood would also provide important biomarkers to predict and monitor patient response to immune modulators. This article reviews all documented biomarkers for detection and monitoring of NSCLC using blood samples.

Circulating tumor cell
The presence of CTCs was first discovered in 1896. During tumor formation and growth, cancerous cells are released into the blood stream. Such cells provide unprecedented insights into both primary and metastatic tumors. Numerous studies have shown that CTCs can serve as a biomarker for predicting disease progression and survival.\textsuperscript{14} High CTC numbers are often associated with aggressive disease, increased metastasis, and shorter disease-free survival (DFS).\textsuperscript{15} Collecting CTCs from blood is noninvasive; therefore, CTCs can be used for “real-time” evaluation of tumor dynamics. In addition, CTCs have the potential to guide treatment, reflect therapeutic efficacy, and reveal drug resistances.\textsuperscript{16}

A number of CTC capturing techniques have been developed to concentrate and isolate the relatively rare cell population from whole blood. Two main principles have been employed based on either immunoselection in a fluid flow chamber or size separation by a filtration device. The former usually incorporates epithelial cell adhesion molecule (EpCAM)-positive selection system in which CTCs are selected by anti-EpCAM antibodies.\textsuperscript{17,18} To date, the CellSearch system based on this principle is the only USA Food and Drug Administration-approved CTC detection method to numerate CTC for limited tumor prognosis applications, but this system is not yet approved for NSCLC diagnosis.\textsuperscript{19–21} Incidence of CTCs in NSCLC was lower compared with other tumors as reported for the CellSearch method.\textsuperscript{19} In addition, EpCAM-positive CTCs are less frequently detected in NSCLC comparing to other epithelial tumors using different CTC detection methods,\textsuperscript{22} primarily because of epithelial-to-mesenchymal transition (EMT). Epithelial proteins have been shown to be downregulated in CTCs; therefore, these EMT-associated CTCs can be missed by EpCAM-based enrichment technologies.\textsuperscript{23} Isolation by size of epithelial tumor cells (ISET) is another method to separate CTC from leukocytes. Cancer cells are generally larger than normal cells, and even small-cell lung cancer cells are larger than circulating lymphocytes or large monocytes. Rarecells and ScreenCell systems are two devices available based on size difference, and both of them use a filter with particular pore size to capture individual and viable CTC.\textsuperscript{24} CTCs can be detected in 20%–30% advanced NSCLC patients with metastasis, and the detection rate is associated with a positive lymph node status.\textsuperscript{25–28} However, the detection rate of CTCs in nonmetastatic patients is very low, suggesting that alternative methods should be utilized for early-stage patients. Multiple markers may also need to be incorporated in CTC capturing to increase the sensitivity of CTC detection in NSCLC patients.

CTC quantification as a biomarker for prognosis, prediction, and treatment response assessment
A study using the CellSearch method showed that 21 out of 101 late-stage treatment-naïve patients (stages III–IV) had at
least 2 CTCs detected. The number of CTCs detected prior to and during treatment was correlated with overall survival (OS) and stage of disease.\textsuperscript{29} Hofman et al reported a study using ISET. All enrolled NSCLC patients underwent surgery and all stages were included, in which 36\% were considered to be CTC-positive. Patients with high level (≥50 cells, 31\% of all samples) of suspicious cells had a worse prognosis.\textsuperscript{30}

Studies have shown that patients with an increased number of CTCs during treatment are often associated with radiographic tumor progression; in contrast, patients with a reduction in number of CTCs are associated with radiographic response.\textsuperscript{31} In patients with stage IV NSCLC, a correlation between changes in number of CTC and fluordeoxyglucose-positron emission tomography or RECIST (Response Evaluation Criteria in Solid Tumors) response was observed.\textsuperscript{32} Patients harboring EGFR mutations were more likely to be CTC-positive compared to patients with wild-type EGFR.\textsuperscript{32}

Molecular changes in CTCs are thought to play important roles in response assessment. Currently, analysis of isolated CTCs can stratify patients according to their mutation types, such as \textit{EGFR}, \textit{HER2}, and \textit{KRAS} mutations.\textsuperscript{33–35} In subsets of patients, CTCs were found to express genes involved in resistance to therapy such as \textit{HER3} and \textit{MET}. Another study reported EGFR T790M detection in CTCs in 2 out of 6 (33\%) NSCLC patients who responded to tyrosine kinase inhibitor therapy compared with 9 out of 14 (64\%) with clinical progression.\textsuperscript{31}

### Genomic rearrangement identified in CTCs

CTC is potentially a noninvasive alternative to tissue biopsies and a surrogate for the entire tumor genome. Current progress in developing whole-genome amplification and whole transcriptome library construction made it possible for the sequencing-based analyses of CTCs to be scaled down to single-cell level.

Whole-genome analysis for CTCs was published by many labs. One NSCLC study isolated CTCs from seven lung adenocarcinoma patients using CellSearch System and performed whole-exome sequencing on each single CTC to detect single-nucleotide variations (SNVs) and copy number variations. The study reported that SNVs detected in single CTCs are highly heterogeneous, and that the important mutations associated with drug target (\textit{EGFR}), drug resistance (\textit{PIK3CA}), and phenotypic transition (\textit{TP53}, \textit{RB1}) can all be detected in CTCs.\textsuperscript{36}

CTCs can be separated according to different surface markers and physical properties using several methods, including flow cytometry and micromanipulation. However, due to their rarity in patient blood, especially for NSCLC, the number of CTCs is usually not sufficient to provide enough DNA or RNA for next-generation sequencing. Although the significant clinical value of CTC genomic analyses is obvious, the use of CTCs to obtain tumor genomic information needs further improvement in the identification of CTC, efficient isolation and capturing of viable cells, automated method for subsequent assessment of the retained cells, and standardized pipeline for sequencing data analysis before clinical implementation in NSCLC.

### Tumor-derived nuclear acids

Tremendous efforts have been invested to explore the potential of ctDNA in diagnosing and monitoring treatment responses. Several studies have shown that increased levels of DNA fragments are found in the blood of cancer patients.\textsuperscript{37,38} Circulating free DNA (cfDNA) are small double-stranded fragments of DNA found in plasma. cfDNA is originating from different tissues, and hematopoietic cells are the main source in healthy subjects.\textsuperscript{39} Although with great variations among individuals, the concentration of cfDNA is usually very low – often <10 ng per mL of plasma.\textsuperscript{40,41} Increased level of cfDNA was often detected as a result of tissue trauma, inflammation, or diseases such as cancer. In the context of cancer, ctDNA, constituting a substantial part of cfDNA, primarily derives from apoptosis and necrosis of the cancer cells.\textsuperscript{37} Some studies have also reported alternative sources of ctDNA, such as CTCs\textsuperscript{42,43} and exosomes, which can be actively secreted by tumor cells.\textsuperscript{44–46} As the concentration of ctDNA is significantly affected by tumor volume, localization and vascularization as well as a patient’s response to treatments, and hepatic and renal clearance function, the abundance of ctDNA in the blood may reflect tumor burden and provide an earlier indication of drug treatment response compared to imaging monitoring.\textsuperscript{47} The concordance rate between the genetic profiles of ctDNA and the corresponding tumor tissue has been demonstrated by numerous studies.\textsuperscript{47–49} Therefore, ctDNA is a valuable tool for mutation detection in cancer patients who cannot undergo tissue biopsy. However, challenges remain because ctDNA only represents a very small fraction of cfDNA,\textsuperscript{50,51} thus requiring methods with very high sensitivity.

### ctDNA as prognosis predictive marker

Numerous studies have demonstrated the prognostic and predictive value of ctDNA in a number of cancers, including but not limited to NSCLC, colon cancer, breast cancer, and other malignancies.\textsuperscript{13,32,57} A pooled analysis including
Methylated DNA

DNA methylation is rising as a promising marker for early detection, prognosis, and real-time follow-up of tumor dynamics. Hypermethylation of the CpG islands in promoter regions of tumor-suppressor genes has been shown to contribute to carcinogenesis. A number of loci have been identified in lung cancer that show significantly elevated DNA methylation in tumor tissue. In the plasma of lung cancer patients, DNA hypermethylation was also reported, including p16INK4A, DAPK11, RASSF1A, RAR, MGMT, GSTP1, and APC. However, only 49% sensitivity and 85% specificity were achieved in identifying lung cancer patients from nonmalignant pulmonary disease. Several studies have reported the potential of investigating tumor-specific methylations in blood for screening and diagnosis of lung cancer. For example, various gene promoters were found to be differentially methylated in ctDNA between patients with lung cancer and controls, including short stature homeobox 2 (SOX2), doublecortin like kinase 1 (DCLK1), septin9 (SEPT9), ras association domain family 1 isoform A (RASSF1A), and retinoic acid receptor B2 (RARB2).

DNA methylation can also be used as a prognostic biomarker. Detection of methylated breast cancer metastasis suppressor-1 (BRMS1) and (sex determining region Y)-box 17 (SOX17) in operable and advanced NSCLC was shown to have a negative impact on survival. In contrast, SFN (14–3–3 Sigma) promoter methylation was correlated with a reduced risk of death. Interestingly, in addition to its diagnostic value, DCLK1 methylation was also associated with shorter survival.

Several studies have reported the use of tumor-specific methylation for tracking a patient’s response to therapy. Elevated level of APC and RASSF1A promoter methylation in ctDNA within 24 h after cisplatin-based therapy was reported, consistent with chemotherapy-induced cell death. Methylation of SHOX2, RASSF1A, and RARB2 has shown potential to monitor disease recurrence after surgery and chemotherapy. Patients with unmethylated checkpoint for kinase and ring finger domains (CHFR) promoter survived longer when receiving EGFR tyrosine kinase inhibitors as second-line treatment, compared to conventional chemotherapy.

MicroRNA

It has been demonstrated that circulating miRNAs’ levels in human plasma remain stable even when plasma is subjected to prolonged incubation at room temperature. Circulating miRNAs are also shown resistant to multiple freeze–thaw cycles. The noninvasiveness and stability make circulating miRNAs a potential tool to identify diagnostic markers in oncology. Indeed miRNA level changes were identified in certain tumor mouse models and lung cancer patients. It is, therefore, possible that cell-free miRNA signatures can be used to diagnose cancers.

Ever since miRNAs were identified in human plasma, many studies have been aiming at identifying a reliable diagnostic tool by investigating the differential expression of circulating miRNAs among NSCLC patients, healthy controls, and patients harboring benign tumors. It is found that compared to healthy donors, the expression of miR-25 is fivefold higher and miR-223 is threefold higher in serum of NSCLC patients. It has been demonstrated that a predictive model consisting of serum miR-15b-5p, miR-16-5p, and miR-20a-5p can be used to differentiate early-stage NSCLC cases from healthy subjects. Downregulation of miR-486-5p was observed in both primary tumors and the serum of NSCLC patients by several independent research groups, which suggests its potential as a tumor suppressor involved in multiple signaling circuits in NSCLC. Furthermore, a 34-miRNA panel can identify patients with early-stage NSCLC in a group of asymptomatic high-risk subjects with up to 80% accuracy. The signature could possibly serve as a noninvasive screening tool to stratify individuals at high risk. Most importantly, such signature is lung cancer specific. Moreover, an eight-miRNA panel was identified using three different analytical methods, which can be used for differentiating cancerous tissues from noncancerous ones in NSCLC patients. However, only three of these miRNAs (miR-30e-5p, miR-21, and miR-10a) are differentially expressed in NSCLC plasma specimens compared to plasma obtained from healthy individuals.
Some miRNAs have even shown great potential as a therapeutic target of lung cancer. For instance, miR-21 is commonly overexpressed in cancers including lung, breast, and colorectal cancer (CRC).\textsuperscript{77} Higher expression of miR-21 and miR-155 in both tumor tissues and serum can predict recurrence and poor survival in NSCLC, which may suggest that overexpression of circulating miR-21 has great prognostic potential in NSCLC.\textsuperscript{78} miR-21 affects response to chemotherapy in several tumor types and thus can also act as a therapeutic target for overcoming drug resistance in cancers. Indeed, recent study has revealed that antisense inhibition of miR-21 or miR-221 sensitizes the effects of gemcitabine, a chemotherapeutic treatment of pancreatic cancer.\textsuperscript{79} Besides miR-21, let-7 also acts as a promising therapeutic target for lung cancer. Systemic delivery of synthetic let-7 mimic or miR-34a into lung tumor-bearing mice results in significant tumor regression compared to the delivery of an miRNA control.\textsuperscript{79} As most studies of miR-21 and let-7 are focusing on tumor tissues, the application of such circulating miRNA as a biomarker is still at research stage. Further validations are needed before applying them as biomarkers for the early diagnosis of cancers including NSCLC.

However, the inconsistency and even contradictions shown in different studies of miRNA evaluation make this method less applicable in clinical settings. This may be contributed by three factors: 1) the internal control gene used for the normalization of miRNA expression may vary between different research groups. Unfortunately, the field has not reached a uniform and reliable internal control for data analysis. 2) Circulating miRNAs do not exist only in exosomes, and most of them are shown to be associated with protein complexes; therefore, different materials such as plasma/serum, whole blood, or exosomes used for cell-free miRNAs isolation make the quality and quantity of circulating miRNAs isolated by different research groups vary. 3) In some studies, patient number is very limited, which may cause the results to be less reliable.

**Exosome as surrogate for mutation detection**

Exosomes are nanovesicles of 50–100 nm diameters that are released from most viable cells. Current knowledge suggests that exosomes form by membrane invagination of late endosomes, resulting in vesicles that contain cytosolic components and expose the extracellular domain of some plasma membrane receptors at their surface.\textsuperscript{80,81} Because of their cellular origin, exosomes bear specific protein markers of the endosomal pathway, such as tetraspanins (CD63, CD9, and CD81), heat shock proteins (HSP70), and proteins from the Rab family, Tsg101 and Alix, which are not found in other types of vesicles of similar size.\textsuperscript{82,83} Exosomes have been found in various body fluids of healthy individuals and patients with malignancies as cargos containing a wide variety of components such as single-stranded RNA, long noncoding RNA, miRNA, and double-stranded DNA.\textsuperscript{84} Profiling of exosomes provided tailored information of individual patients and may be crucial for a precise prognosis or a differential diagnosis.

Numerous studies have investigated and confirmed the potential of exosomes as surrogate for mutation detection. Double-stranded genomic DNA with mutated KRAS and p53 were identified in the serum exosomes of patients with pancreatic cancer.\textsuperscript{85} In addition, using whole-genome sequencing revealed that serum exosomes from patients with pancreatic cancer contain genomic DNA spanning all chromosomes. These results indicated that serum-derived exosomes can be used to determine genomic DNA mutations for cancer prediction, treatment, and therapy resistance. In an independent study, high prevalence of mutant KRAS in circulating exosome-derived DNA (exoDNA) was detected in plasma of early-stage pancreatic cancer patients, by droplet digital PCR on exoDNA and cfDNA.\textsuperscript{86} KRAS mutations in exoDNA were identified in 7.4%, 66.7%, 80%, and 85% of age-matched controls, localized, locally advanced, and metastatic pancreatic ductal adenocarcinoma (PDAC) patients, respectively. Although a higher percentage of patients with localized PDAC exhibited detectable KRAS mutations in exoDNA than previously reported for cfDNA, a substantial minority of healthy samples demonstrated mutant KRAS in circulation, dictating careful consideration and application of liquid biopsy findings. In CRC, genomic DNA extracted from the tumor tissues of 35 patients with histologically confirmed CRC and exosomal mRNA obtained from peripheral blood of the corresponding patients prior to surgery are studied together.\textsuperscript{87} The KRAS mutation rates in tumor tissues and the matched serum exosomes were 57.6% and 42.4%, respectively; BRAF mutation was 24.2% and 18.2%, respectively. There was no significant difference between the rate detected by tissue and serum exosomes. The total consistency rate was 94.9% and 93.9% for KRAS and BRAF mutations, respectively. These results suggested that serum exosomal mRNA may be used as a novel source for the rapid and noninvasive genotyping of cancer patients.\textsuperscript{88,89} Recently, EGFR T790M resistance mutation in lung cancer patients is successfully detected using exosomal RNA and the sensitivity is improved compared to analysis of ctDNA alone.\textsuperscript{90}
Recent studies also suggest that plasma exosomes might be valuable diagnostic indicators in lung cancer. Profiling of the exosome proteins from the plasma of 219 suspected lung cancer patients with 37 antibodies targeting lung-cancer-related proteins identified a 30-marker model classifying correctly 75% of patients (sensitivity of 0.75 and specificity of 0.76).91 Similarly, analysis of the plasma exosomes from 581 patients (431 with lung cancer and 150 control individuals) using a custom array (49 antibodies). The authors demonstrated that CDC151, CDC171, and tetraspanin 8 were the strongest discriminators of malignancy, compared with healthy controls, and proposed a 10-marker model as a diagnostic tool.92

Serum protein markers
NSCLC is often associated with the differential expression of several proteins, which may be potential biomarkers for lung cancer. For instance, carcinoembryonic antigen, CYFRA 21–1 (cytokeratine 19 fragment), neuron-specific enolase, progastrin-releasing peptide, and squamous cell carcinoma antigen are reported to be elevated in serum of a fraction of lung cancer patients. However, these protein markers are not sensitive or specific enough to detect lung cancer in a clinically relevant way or to have enough value as biomarker for the diagnosis of asymptomatic patients with lung cancer.93,94

Summary
Diagnosing NSCLC at an early stage remains to be a hurdle to improve the OS rate, which has only increased by a few percent with all the advancements in targeted therapies. Liquid biopsy, providing an opportunity for interrogating the cancer genome and epigenome noninvasively, allows for real-time tracking of tumor dynamics. The above-mentioned substrates in the blood have the potential to serve as early diagnosis, treatment stratification, and recurrent monitoring tools. While recent studies have produced numerous promising biomarkers, translating such markers to the bedside necessitates rigorous validations through large multicenter prospective studies.

Acknowledgments
This work was supported by the Science and Technology Program of Guangzhou, China (grant number 201607010117). We thank Dr Flora Han-Zhang from Burning Rock Biotech for editing and valuable discussion.

Disclosure
The authors report no conflicts of interest in this work.

References


