Lung cancer biomarkers: State of the art

Sangeetha Subramaniam, Ram Krishna Thakur¹, Vinod Kumar Yadav¹, Ranjan Nanda, Shantanu Chowdhury¹, Anurag Agrawal¹*

International Center for Genetic Engineering and Biotechnology, ¹CSIR Institute of Genomics and Integrative Biology, Delhi, India

E‑mail: a.agrawal@igib.in
*Corresponding author

Published: 28 February, 2013  Received: 10 July, 2012
Journal of Carcinogenesis 2013,12:3  Accepted: 10 August, 2013
This article is available from: http://www.carcinogenesis.com/content/12/1/3
© 2013 Subramaniam

Abstract

Lung cancer is one of the deadliest cancers worldwide, with the highest incidence and mortality amongst all cancers. While the prognosis of lung cancer is generally grim, with 5-year survival rates of only 15%, there is hope, and evidence, that early detection of lung cancer can reduce mortality. Today, only computed tomography screening has shown to lead to early detection and reduction in mortality, but is limited by being anatomic in nature, unable to differentiate between inflammatory and neoplastic pathways, and therefore, susceptible to false positives. There is increasing interest in biomarkers for lung cancer, especially those that predict metastatic risk. Some biomarkers like DNA mutations and epigenetic changes potentially require tissue from the at-risk site; some like serum proteins and miRNAs are minimally invasive, but may not be specific to the lung. In comparison, emerging biomarkers from exhaled breath, like volatile organic compounds (VOC), and exhaled breath condensate, e.g., small molecules and nucleic acids, have the potential to combine the best of both. This mini review is intended to provide an overview of the field, briefly discussing the potential of what is known and highlighting the exciting recent developments, particularly with miRNAs and VOCs.

Keywords: Exhaled breath, lung cancer, miRNA, volatile organic compounds

MINIREVIEW

Lung cancer is one of the deadliest cancers worldwide, with the highest incidence and mortality amongst all cancers[1]. While the prognosis of lung cancer is generally grim, with 5-year survival rates of only 15%, there is hope, and evidence, that early detection of lung cancer can reduce mortality. Reduced mortality has recently been shown as a consequence of low dose computed tomography (LD‑CT) screening, but the effect size was modest (about 20% reduction in relative risk) and false positive rates were 96%.[2] The low efficacy is attributable in part to the fact that radiological visibility of tumors requires substantial cell mass and fast growing tumors may grow substantially between consecutive CT screening tests, which were kept 1 year apart to minimize radiation exposure and costs. It is intuitively obvious that small collections of tumor cells cannot be differentiated from similar masses of inflammatory cells using anatomic imaging approaches and probably not even by metabolic imaging methods like Fluoro-deoxy-glucose positron emission tomography (FDG-PET). Based on the presumption that early detection will lead to higher likelihood of cure, alternative cellular and molecular detection strategies such as sputum cytology and molecular biomarkers in various biological samples, have been tested for potential value in the early detection of lung cancer or prognostic and therapeutic guidance.

A number of approaches have resulted in a diverse set of molecular biomarkers for lung carcinoma, particularly in recent years with the advent of next generation sequencing.
technologies combined with advances in bronchoscopic and imaging techniques. These have classically been genetic, such as mutations in p53, K-ras, Rb and myc genes;[3-6] epigenetic (abnormal methylation of APC, TMS1, RASSF1, p16INK4a, DAPK; or chromosomal changes (deletion in chromosome 3p which harbors several tumor suppressor genes).[7-9] Such DNA based changes are robust and not susceptible to degradation or handling related changes, unlike RNA expression. Unfortunately, the applicability of DNA markers is often limited by the need for sufficient tumor tissue, as well as limited sensitivity and specificity. Further, where DNA markers are used to guide therapy, such as epidermal growth factor receptor (EGFR) mutations, cancer cells may further mutate during treatment and require repeated tumor sampling. Recent work, where circulating tumor cells (CTC) were trapped, has provided proof-of-concept for non-invasive monitoring of tumor DNA and may be more widely used in coming years.[10,11] However, this would not be sufficient to diagnose early limited tumors, which are unlikely to be associated with CTC. Recent evidences also show the possible role of small non-coding RNAs such mi(cro) RNA as biomarkers or targets, e.g., miR17 and miR21 clusters or let-7 as a suppressor of RAS in lung cancer.[12,13] Unlike mRNA, miRNAs are exceptionally stable and their potential as biomarkers will be discussed further in subsequent sections.

Protein biomarkers for detection in lung cancer are typically those measurable in sera (ones that are currently clinically useful include, tissue polypeptideantigen (TSA), CYFRA-21-1 and carcinoembryonic antigen for non-small cell lung carcinoma (NSCLC), and neuron-specific enolase and progastrin-releasing peptide (ProGRP) for neuroendocrine lung carcinoma).[14,15] Other potential ones include, plasma kalikrein B1 (KLKB1), serum amyloid A, haptoglobin-alpha-2. Tumor protein expression is important for prognostic usage and therapeutic guidance. For example, expression of (EGFR) is important in guiding therapeutic use of several drugs that are specific for the EGFR receptor.[16,17] Many of these and several other molecules are at various levels of testing in clinical/pre-clinical settings for indications ranging from diagnosis to guidance of therapy.[15] Some biomarkers can even directly lead to therapeutic advances, as seen for the oncogenic EML4-ALK fusion gene in NSCLC.[16-21]

While detection of tumors is an important area for lung cancer biomarker discovery, it is equally important to determine, which tumors are likely to have metastasized, and which are likely to be limited. Detection of CTC is one such approach, but remains challenging and not clinically validated. Given that it is not the primary tumor that kills but its secondary spread, it is surprising that metastasis suppressor genes (MSG) have not received much attention as biomarkers.[22] An emerging body of work aims to fill this lacuna. To establish metastases successfully tumor cells must successfully negotiate several steps, including detachment from the primary tumor, intravasation, survival and extravasation from blood vessels into the secondary site followed by colonization of the secondary site.[22] Therefore, discovery of the first gene (NM23) that could ‘single-handedly’ mitigate the metastatic process was enigmatic.[23] Several other MSGs have been identified since, including, Breast cancer Metastasis Suppressor 1 (BRMS1), KiSS, KAI1, and Rho GDP dissociation inhibitor-2 (RhoGDI2) in different cancer types.[24-26]

In order to test association of MSG levels with lung tumor progression, we conducted large-scale meta-analysis of 23 MSGs in lung tumor clinical samples available from expression project for oncology (expO) database (GSE2109; expO is hosted by International Genomics Consortium (IGC, USA, www.intgen.org)). Interestingly, investigation of 93 patient derived lung tumor transcriptomes grouped stage wise showed statistically significant depletion (stages III and IV vs. stages I and II) in only three classic MSGs (NM23 H2, ARHGDIB, and RECK), and one candidate MSG (PTPN11) [Figure 1]. Importantly, NM23 H2, a member of the Non Metastatic 23 (NM23) family showed very significant change in expression (P < 0.005). In order to ascertain the extent of deregulation, NM23 H2 expression was specifically examined in above clinical profiles; significant decrease in NM23 H2 was found in advanced stages [Figure 1; P < 0.005]. Decreased expression of NM23 H2 in metastatic NSCLC has been probed earlier, however, its prognostic significance has been unclear, and mechanisms of NM23 H2 action as a metastasis suppressor are poorly understood. A previous study of NM23 H2 in lung carcinoma derived A549 cells indicated its gene expression regulatory potential[27] suggesting a possible transcriptional role in progression of lung carcinoma. Though several studies have demonstrated NM23 H2 association with promoters, pointed out amino acids involved in interaction with DNA, and shown independence of regulatory function from enzymatic activity.[28] the precise contribution of NM23 H2 in suppression of metastasis has remained poorly understood. Notably, in our lab, NM23 H2 expression showed profound changes in transcriptome profile in lung cancer cells. Specifically, genes related to cell adhesion, epithelial organization and related signaling pathways were perturbed, supporting anti-metastatic action of NM23 H2. Since a causative role for NM23 H2 or the other MSG could lead to a new therapeutic approach in prevention of lung cancer metastasis, it is currently an active area of investigation.

In the preceding sections, a brief overview of DNA and
protein biomarkers of lung cancer was provided, with examples of how each aspect of lung cancer can have different sets of biomarkers, e.g., risk of cancer, presence of cancer, type of cancer, likelihood of metastasis, probability of response to targeted therapy, and prognosis. Additional molecular types (miRNAs, small molecules, volatile organic compounds) from a larger range of biological substrates (breath, breath condensate, urine, saliva) are being reported. This explosion is attributable to advances in molecular strategies and analytical platforms, including genomics, epigenomics, proteomics, and metabolomics, but has not yet changed clinical management. Since the full fabric of this busy field cannot be captured in a minireview, we have chosen to focus upon two important emerging threads, which may have a large clinical impact in coming years.

**EXHALED BREATH BIOMARKERS IN LUNG CANCER**

Exhaled breath contains volatile organic compounds (VOC) from the environment and the lungs, as well as aerosolized airway lining fluid. An exhaled breath condensate (EBC) can be obtained by passing exhaled breath over a very cold surface. Linus Pauling first attempted characterization of VOC in human breath in 1971. Subsequently, more than 3000 exogenous and endogenous VOC have been detected in normal human breath, contributed by environmental inhalation or by physiological end-products. The rationale behind the study of VOC biomarkers for lung cancer is that in the cancerous state, altered metabolic and biochemical pathways may produce and process VOCs differently from normal cells. The ease of use and high degree of lung-specificity make this an attractive proposition, if validated. Promising data has now emerged from several gas chromatography/mass spectrometry (GC/MS) based studies that have reported such biomarkers, but unfortunately, none has reached clinical usability so far. Potential marker compounds are alcohols, aldehydes, ketones and hydrocarbons, some of which are differentially found in breath of normal subjects and cancer patients. Specific sensor array based methods for the detection of volatile biomarkers have also reached significant advancement. A recent study used solid-phase microextraction and GC/MS to identify 42 VOC that represent lung cancer biomarkers. Four of these were used to train and optimize the gold nanoparticle based sensors, demonstrating good agreement between patient and simulated breath samples, showing that a simple e-nose for cancer may not be far away.

However, given the relative lack of clinical successes so far, an emerging direction in this field is the application of systems biology methods to biomarker datasets. This is a natural offshoot of the realization that no single molecule can sufficiently discriminate complex processes, without context. Recent studies that used statistical approaches incorporating patient data, such as smoking status, have found that different sets of VOC were discriminatory for cancer in non-smokers and smokers. Such mathematical models not only aid in the identification of multi-component markers with high discriminatory power, but also may enhance mechanistic understanding of the disease mechanisms. However, the level of computational analysis depends on the experimental data in hand, which so far is not much.

We speculate that an ideal methodology of volatile-based biomarkers should include system biology approaches in addition to chemometrics. To illustrate this point, results of a pilot-scale study conducted in an author’s (Ranjan Nanda) lab are shown, where exhaled breath collected over
packed polymer Tenax tubes was analyzed by GC/MS. We identified 140 VOCs present in the breath of 9 lung cancer and 18 healthy subjects. Multivariate analysis of this data was conducted using Partial Least Squares Discriminant Analysis (PLS-DA) to establish a preliminary classification model and evaluate the relation between compounds and classes. Of 140 metabolites that were identified, 18 were significantly altered in lung cancer [Figure 2b], which can then be visualized as networks for studying interrelationships [Figure 2c]. The dataset shown is for illustrative purposes only, and larger studies are needed for reliable biomarker discovery.

EBC is also a useful substrate for small-molecule biomarker discovery, although much less convenient. Nuclear Magnetic Resonance spectroscopy based metabolomic studies of EBC have been found to discriminate between healthy and asthmatic subjects. A limitation of EBC is lack of effective data normalization strategies and currently complex statistical strategies are required for multi-parametric data.

**MIRNAS AS BIOMARKERS FOR DIAGNOSIS AND TREATMENT OF LUNG CANCER**

miRNAs are ultrashort (18-25 nucleotides), non-coding RNA molecules that can specifically bind to target mRNA sequences, usually resulting in mRNA degradation. They are highly stable, being relatively resistant to nucleases, and can be found in most bodily fluids. miRNA expression profiles are highly dynamic, yet regulated, which makes them very suitable as biomarkers. However, normalization of miRNA expression is difficult because of the lack of a “housekeeping” stably expressed miRNA. Bianchi et al., have reported that asymptomatic lung cancer detected by CT screening is also associated with changes in circulating miRNA profile. A multivariate risk-predictor algorithm based on the weighted linear combination of the 34-miRNA expression levels, was sufficient to identify patients with early stage NSCLCs in a
population of asymptomatic high-risk individuals with 80% accuracy.[43] This relatively simple test, requiring only 1 ml blood and quantitative-polymerase chain reaction (Q-PCR), shows the potential to discriminate between benign and malignant CT lesions. Independent validation of this important report is awaited. Tumor miRNA profiles have also been shown to predict recurrences of localized stage I non-small cell lung cancer after surgical resection.[44] Whether circulating miRNA profiles can be used instead, remains to be seen.

In summary, while much progress has been made in biomarker discovery for lung cancer, much more needs to be done. For early detection, tests using exhaled breath or circulating miRNA appear to be particularly promising, although both fields are still young. So far, prognostic and therapeutic markers rely upon isolation of cancerous tissue, whether from lung or from circulating tumor cells. While miRNAs appear to be promising, it is unlikely that VOCs would be useful in this regard. Unpublished work from our lab (Anurag Agrawal) shows that it is possible to measure miRNAs in exhaled breath condensate by Q-PCR. This has the potential to create simple non-invasive breath based tests, incorporating the best of all. The field of miRNA based biomarkers in lung cancer is starting to bloom and further investigations are warranted.

REFERENCES


Source and Support: Gates Foundation (Ranjan Nanda) and Council of Scientific and Industrial Research (Anurag Agrawal and Shantanu Chowdhury).

Conflict of Interest: None declared.