Lung Cancer 107 (2017) 50–58

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

Integration of multiple "OMIC" biomarkers: A precision medicine strategy for lung cancer☆

Ana I. Robles*, Curtis C. Harris

Laboratory of Human Carcinogenesis, National Cancer Institute, NIH, Bethesda, MD 20892, USA

ARTICLE INFO

Article history:
Received 9 March 2016
Received in revised form 7 June 2016
Accepted 10 June 2016

Keywords:
Low-dose computed tomography
Cytokine
Metabolomics
Methylation
Gene expression
MicroRNA

ABSTRACT

More than half of all new lung cancer diagnoses are made in patients with locally advanced or metastatic disease, at which point therapeutic options are scarce. It is anticipated, however, that the widespread use of Low-Dose Computed Tomography (LDCT) screening, will lead to a greater proportion of lung cancers being diagnosed at an early, operable, stage. Still, the overall rate of recurrence for surgically treated Stage I lung cancer patients is up to 30% within 5 years of diagnosis. Thus, the identification and clinical application of biomarkers of early stage lung cancer are a pressing medical need. The integrative analysis of "omic," clinical and epidemiological data for single patients is a core principle of precision medicine. Through rigorous bioinformatics and statistical analyses we have identified biomarkers of early-stage lung cancer based on DNA methylation, expression of mRNA and miRNA, inflammatory cytokines, and urinary metabolites. Beyond a more comprehensive understanding of the molecular taxonomy of lung cancer, these biomarkers can have very practical implications in the context of unmet clinical needs of early stage lung cancer patients: First, current guidelines for LDCT screening broadly include individuals based on age and history of heavy smoking. Tumor-derived circulating biomarkers in the blood and urine associated with lung cancer risk could narrow and prioritize individuals for LDCT screening. Second, a high number of nodules are identified by LDCT, of which fewer than 5% are finally diagnosed as lung cancer. Biomarkers may help discriminate malignant nodules from benign or indolent lesions. Third, the expected rise in the numbers of lung cancer patients diagnosed at an early stage will necessitate new treatment options. Circulating, urinary and tissue-based biomarkers that molecularly categorize Stage I patients after tumor resection can help identify high-risk patients who may benefit from adjuvant chemotherapy or innovative immunotherapy regimens.

Published by Elsevier Ireland Ltd.

1. Recent advances in lung cancer detection reveal unmet clinical needs

Lung cancer remains the leading cause of cancer-associated deaths worldwide, despite a slow but continuous decline in incidence and mortality in the US and other Western countries over the past two decades [1,2]. Global variations in lung cancer incidence largely follow historical patterns of smoking, and incidence and mortality rates are still on the rise in Asia and some countries in Latin America and Africa, where the smoking epidemic began later [2]. Most lung cancer patients are diagnosed with locally advanced or metastatic disease, with few therapeutic options and a dismal survival rate. The introduction of innovative therapies targeted to specific molecular alterations continues to improve outcomes for a subset of patients with advanced stage lung cancer [3]. In addition, recent promising results of T cell-based immunotherapy associated high mutational burden in lung cancer patients suggest that exome-guided neoantigen identification may improve treatment responses [4].

When diagnosed at an early, operable stage, the 5-year survival rate from lung cancer climbs above 50% [1]. Cigarette smoking is the major risk factor for lung cancer and other smoking-related diseases [5]. Even as this risk gradually decreases after smoking cessation, former smokers account for most new lung cancer diagnoses. Thus, lung cancer screening efforts have focused on older individuals with a history of heavy smoking. The results of the landmark National Lung Screening Trial (NLST) published in 2011 [6], demonstrated a statistically significant mortality benefit of low-dose computed tomography (LDCT) over chest radiography (CXR) screening in high-risk individuals (defined by age and history of

☆ The authors are supported by the Intramural Research Program of the National Cancer Institute, NIH.

* Correspondence to: 37 Convent Dr Rm 3060D, MSC 4258, Bethesda, MD 20892-4258, USA.
E-mail address: Ana_Robles@nih.gov (A.I. Robles).

http://dx.doi.org/10.1016/j.lungcan.2016.06.003
0169-5002/Published by Elsevier Ireland Ltd.

heavy smoking). This evidence and a systematic review of LDCT screening studies [7], prompted the American Cancer Society and other health care organizations to issue guidelines for clinicians to discuss lung cancer screening by LDCT with older patients with a history of heavy smoking, along with smoking-cessation counseling [8,9].

In early 2015, the Centers for Medicare and Medicaid Services (CMS) made public a decision to cover the cost of lung cancer screening with LDCT for patients at high-risk according to guidelines similar to NSLT eligibility criteria [10]. As a result, the use of LDCT screening is expanding, creating a need for prioritization of the over 9 million individuals who would be eligible for screening under the current guidelines [11]. Non-invasive circulating or urinary biomarkers associated with lung cancer risk, i.e., tumor-derived metabolites, could help prioritize individuals for LDCT screening among those at large high-risk to increase the efficacy of screening and to reduce the cost and morbidity associated with it (Risk biomarkers, Fig. 1). The magnitude of the task is compounded by the fact that LDCT scanning identifies a high number of nodules that prompt further, invasive testing but do not result in a lung cancer diagnosis. In the NLST, 96.4% of initial positive screenings were deemed non-cancer on further testing [6]. A recent retrospective study of the clinical management of patients with Intermediate Pulmonary Nodules (IPNs, 8–20 mm) in community practice, found wide variation in nodule management that led to a high number of unnecessary invasive procedures [12]. Thus, there is a need for non-invasive biomarkers that can help discriminate malignant nodules from benign or indolent lesions (Diagnostic biomarkers, Fig. 1). Up to 60% of lung cancers detected by LDCT in the NLST were Stage I, primarily adenocarcinoma histology [6]. With the projected rise in LDCT screening [13], a shift in the stage at diagnosis, towards early, operable disease is expected. The recommended treatment for patients with Stage I Non-small-cell lung cancer (NSCLC) is surgery, which may be followed by chemotherapy in patients with pathologically high-risk, margin-negative Stage IB tumors [14]. Still, up to 30% surgically-treated Stage I patients will die from recurrent disease [15]. Non-invasive or tissue-based biomarkers that molecularly categorize Stage I patients after tumor resection and identify those at high-risk for recurrence could lead to improved clinical management (Prognostic biomarkers, Fig. 1). High-risk patients may benefit from adjuvant chemotherapy or innovative checkpoint immunotherapy, while low-risk patients might safely be spared further treatment and instead be followed by surveillance LDCT. In summary, we will discuss three critical needs in early stage lung cancer: (1) To prioritize high-risk individuals for screening by LDCT (screening); (2) To assess the malignant potential of IPNs to reduce overdiagnosis and unnecessary surgery (diagnosis); (3) To identify Stage I patients at high risk of recurrence (prognosis).

2. Biomarkers in precision medicine

Biomarker discovery and validation are main components of the precision medicine strategy (Fig. 2). The precision medicine approach, first outlined in the 2011 report from the National Research Council [16], includes four basic premises. First, a disease Information Commons is populated with comprehensive measurements of various types of molecules from individual patients (collectively referred to as "-omic" data). This multi-layer reservoir of molecular data may include global analysis of the exposome, genome, epigenome, transcriptome, metabolome, proteome and microbiome, as well as clinical and epidemiological information. Second, these data are integrated into a Knowledge Network that examines the interconnectivity across data layers from the Information Commons. Third, this Knowledge Network is used to develop new molecular classifications of disease. The ultimate goal of these new Taxonomic Classifiers is to refine risk assessment, more precisely diagnose patients, and make informed decisions on therapeutic strategies. Finally, this knowledge is used to inform biomedical research, preventive care, and clinical medicine and to fuel relevant mechanistic and observational studies. Progress is based on iterative process of acquiring information in individuals or cohorts of patients, making improvements in taxonomy and utilizing that knowledge to care for patients and design new studies that further feed the Information Commons. This approach has come to the forefront with the recent announcement of the oncology "precision medicine" research initiative by the National Institutes of Health [17]. The just-launched NCI-MATCH (Molecular Analysis for Therapy Choice) trial is an example of a precision oncology clinical trial that aims to evaluate the extent to which treating cancers according to their molecular abnormalities will be able to improve patient outcome [18]. The Adjuvant Lung Cancer Enrichment Marker Identification and Sequencing Trial (ALCHEMIST) [19] applies this concept to the treatment of patients with early-stage non-squamous NSCLC, while feeding back into the Information Commons through comprehensive genomic analysis.

Although the field of biomarker research has for years operated within a similar framework as outlined by the precision medicine approach, the more recent focus on high-dimensional "-omic" data as a source for disease markers has brought with it a new set of analytical and clinical validation challenges [20]. As the number of features measured greatly exceeds the number of samples, analytical strategies have had to evolve to avoid over-fitting of the models and ensure that biomarkers are widely applicable beyond the sample cohort used to generate them. A set of recommendations has been put forward by the Institute of Medicine to guide the translation of omics-based biomarkers into clinical tests [21]. At the Discovery Phase, biomarkers should be confirmed in a set of samples that is independent from that in which the original discovery was made. The primary data and computational procedures should be fully disclosed and the derived algorithms should be defined precisely. At the Validation Phase, it is recommended that the candidate omics-based test and its intended clinical use be discussed with the US Food and Drug Administration (FDA) and that validations are conducted under Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory standards. Evaluation for clinical utility requires close consultation with FDA and a fully defined, validated and locked-down test [21]. To date, no molecular biomarkers that identify individuals for priority screening or facilitate the assessment of IPNs or prospectively identify lung cancer patients at a high risk of recurrence after surgery have been successfully translated into clinical use, although several are at advanced stages of development [22–33].

3. Strategy to identify OMICS-based biomarkers with potential clinical utility

A strategy for biomarker discovery and validation is outlined in Fig. 3. Briefly, a cohort (nested case-control or case series) of sufficient size and with well-curated epidemiological and clinical data is the starting point for comprehensive profiling studies. Measures of association with presence of disease or disease outcome are utilized to select candidate biomarkers. This rigorous assessment requires evidence for statistically significant risk separation as well as improved predictive value over known risk factors, including age and smoking. A candidate biomarker that passes this threshold is then validated internally in the same patient cohort using a secondary, targeted assay, such as quantitative RT-PCR (qRT-PCR) or pyrosequencing, to avoid platform-specific biases. This step also allows the development of an assay that may be more easily