Journal of Biometrics & Biostatistics
ISSN: 2155-6180

Fenghai, J Biom Biostat 2016, 7:3
DOI: 10.4172/2155-6180.1000299

Opinion Open Access

The Use of Molecular and Imaging Biomarkers in Lung Cancer Risk Prediction

Fenghai D*

Department of Biostatistics and Center for Statistical Sciences, Brown University School of Public Health, USA

***Corresponding author:** Fenghai D, Department of Biostatistics and Center for Statistical Sciences, Brown University School of Public Health, USA, Tel: (401) 863-2175; Fax: (401) 863-9182; E-mail: fduan@stat.brown.edu

Rec date: April 25, 2016; **Acc date:** April 26, 2016; **Pub date:** May 3, 2016

Introduction

High-dimensional genomics, genetics and proteomics techniques have been widely used in cancer research for over two decades. Correspondingly, various genomic, genetic and proteomic signatures have been discovered in the cancer's diagnosis, prognosis and prediction. For instance, over the last decade, considerable effort and resources have been devoted to characterize the genomic, genetic, and proteomic profiles of lung cancers [1-3]. These studies can enable us to have a deep understanding of the molecular heterogeneity of this disease and help create new therapeutic targets that will facilitate personalize targeted therapy. Now, as non-invasive medical image technologies are likely to become routine in screening high-risk populations, the use of imaging features may greatly assist the therapy guidance and the monitoring of development and progression of lung cancer and its response to treatment. Similar to other – omics technologies, radiomics refers to the high-throughput extraction and analysis of a large amount of quantitative features from advanced medical images with the assistance from compute science, and can provide a comprehensive quantification of the tumor phenotype [4-6].

Studies that Collect Molecular and Imaging Features

The National Lung Screening Trial (NLST) was a randomized screening trial that accrued over 53,000 older smokers to compare low-dose helical computed tomography (CT) relative to chest-x-ray screening in reducing lung cancer mortality. Half of accrued participants (about 26,000) underwent at least one CT screen. In addition, about 10,000 participants consented to have their specimens collected for the development of the NLST biorepository for lung cancer biomarker validation research. The NLST study has shown that compared to chest-x-ray, low-dose helical CT can reduce the death from lung cancer by 20% [7].

More recently, combination of the molecular findings with image-based features of lung cancer on chest CT has emerged as new tool that can potentially impact both the diagnostic and prognostic spaces [8,9]. A few prospective studies have been developed in this regards. The Detection of Early lung Cancer Among Military Personnel (DECAMP) consortium is an ongoing multidisciplinary and translational research program that was funded by DoD to study the diagnostic ability of a number of developed molecular biomarkers, including one genomic biomarker measured in bronchial airway brushings, two proteomic biomarkers measured in bronchial airway biopsies or serum, and one cytokine biomarker measured in serum. The consortium aims to enroll 500 heavy smokers with indeterminate pulmonary nodules (ranging from 0.7cm – 3.0cm) on chest CT from 7 VA hospitals and 4 designated Military Treatment Facilities (and also one academic hospital). The research team of the consortium includes several molecular laboratories and the cores of Biostatics, Bioinformatics and Biorepository. In addition to its primary endpoint, an important aim of this study is to develop models that can combine the features from demographic, clinical, radiographic, and molecular sources to predict the risk of lung cancers [10,11].

Recently, a few grants were awarded by the NCI to create a consortium that studies the molecular characterization of screen-detected lesions, including the domain of prostate cancer, lung cancer, breast cancer, and pancreatic cancer. The consortium has seven molecular characterization laboratories (MCLs) and a coordinating center, and is supported by the Division of Cancer Prevention and the Division of Cancer Biology [12]. In the context of lung cancer, the aim is to seek evidence that screening will detect a class of non-aggressive tumors, which is different from the tumors detected in patients with symptoms. For this purpose, the study team will characterize the mutational status, RNA expression profiles, tumor microenvironment, and imaging related features in these screen-detected tumors. One of the key questions is to integrate the feature data from various sources to develop a composite model that can assist the prediction of lung cancer risk.

Method of Integrating Molecular and Imaging Biomarkers

In these studies, various types of biomarkers will be collected from various platforms, e.g., demographics, clinical practice, molecular assay, imaging modality, and so on. Many methods can be used to analyze and integrate these biomarker data. Unsupervised clustering analysis can be conducted to group biomarkers in discrimination analysis and allow us to obtain an assessment of the overall relationship among them. Specially, clustering analysis can be used to determine the possible clusters formed from these platforms and then characterize each cluster based on different biomarkers [13-15]. Dependent on the types of outcomes, logistic regression and Cox-proportional hazards regression will usually be used to model these biomarkers in the integration analysis. When there are too many biomarkers, robust regression techniques (Such as LASSO) are often used to reduce dimensionality [16].

Challenging Issues in the Integration of Molecular and Imaging Biomarkers

One essential aim of the risk predictive modeling is to predict the outcome of new subjects. For this, the biggest challenging issue is how to avoid overfitting, i.e., that the data fit the training set well, but perform poorly in the validation set. This is particular true when building a complex risk prediction modeling with the inclusion of too many biomarkers. Overfitting causes optimism about a model's

performance in new subjects and will greatly limit the model's capacity of generalization. Here, we recommend the bootstrap resampling to evaluate a model's optimism-corrected performance, which repeatedly draws samples with replacement from the original sample to fit the model and then evaluate the model's performance in the original sample. Detailed procedure of the calculation can be found in the chapter 5 of Steyerberg's book [17]. Of course, the best approach in evaluating a risk prediction model's performance is to design a new prospective study and test the model's performance there, e.g., the ongoing DECAMP study. Then the model's accuracy can be independently assessed by sensitivity, specificity and AUC in the ROC approach.

References

1. Cancer Genome Atlas Research Network (2012) Comprehensive genomic characterization of squamous cell lung cancers. Nature 489: 519-525.
2. Cancer Genome Atlas Research Network (2014) Comprehensive molecular profiling of lung adenocarcinoma. Nature 511: 543-550.
3. Silvestri GA, Vachani A, Whitney D, Elashoff M, Porta Smith K, et al. (2015) A bronchial genomic classifier for the diagnostic evaluation of lung cancer. New England Journal of Medicine 373: 243-251.
4. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, et al. (2012) Radiomics: the process and the challenges. Magnetic resonance imaging 30: 1234-1248.
5. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, et al. (2012) Radiomics: extracting more information from medical images using advanced feature analysis. European Journal of Cancer 48: 441-446.
6. http://www.radiomics.org/
7. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, et al. (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 365: 395-409.
8. Gevaert O1, Xu J, Hoang CD, Leung AN, Xu Y, et al. (2012) Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data--methods and preliminary results. Radiology 264: 387-396.
9. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, et al. (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature communications.
10. Spira AE, Maple E (2012) Detection of early lung cancer among military personnel (decamp). Boston University, USA.
11. https://clinicaltrials.gov/ct2/show/NCT01785342
12. http://prevention.cancer.gov/news-and-events/news/consortium-molecular
13. Monti S, Tamayo P, Mesirov J, Golub T (2003) Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Machine learning 52: 91-118.
14. Cancer Genome Atlas Research Network (2011) Integrated genomic analyses of ovarian carcinoma. Nature 474: 609-615.
15. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, et al. (2014) Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell 158: 929-944.
16. Tibshirani R (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society 1: 267-88.
17. Steyerberg E (2008) Clinical prediction models: a practical approach to development, validation, and updating. Springer Science & Business Media.