If there is Online Only content that cannot be converted to a Word processing format, you may have to click the Supplemental Files icon on the menu bar in your Reviewer Center to access.

Probability of Cancer in High Risk Patients Predicted by the Protein-Based Lung Cancer Biomarker Panel in China: LCBP Study

Journal:	*CHEST*
Manuscript ID	CHEST-16-2258
Article Type:	Original Research
Date Submitted by the Author:	07-Oct-2016
Complete List of Authors:	Yang, Dawei; Zhongshan Hospital, Fudan University, Department of Pulmonary Medicine; Chinese Alliance Against Lung Cancer; Shanghai Respiratory Research Institution Zhang, Xiaoju; Zhongshan Hospital, Fudan University, Pulmonary Medicine; Henan Provincial People's Hospital, Department of Respiratory and Critical Care Medicine; Chinese Alliance Against Lung Cancer Powell, Charles; Mount Sinai School of Medicine, Pulmonary, Critical Care and Sleep Medicine Ni, Jun; Peking Union Medical College Hospital, Department of Pulmonary Medicine; Chinese Alliance Against Lung Cancer Wang, Bin; Xinqiao Hospital, the Third Military Medical University, Department of Pulmonary Medicine; Chinese Alliance Against Lung Cancer Zhang, Jianya; Jinling Hospital, Nanjing University School of Medicine, Department of Respiratory Medicine; Chinese Alliance Against Lung Cancer Zhang, Yafei; the Chest Hospital of Henan Province; Chinese Alliance Against Lung Cancer Wang, Lijie; the Second Military Medical University Xu, Zhi-hong; Ruijin Hospital, Shanghai Jiaotong University, Department of Geriatrics; Chinese Alliance Against Lung Cancer Zhang, Li; Department of Respiratory Disease, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences& Peking Union Medical College; Chinese Alliance Against Lung Cancer Wu, Guoming; Xinqiao Hospital, the Third Military Medical University, Department of Pulmonary Medicine; Chinese Alliance Against Lung Cancer Song, Yong; Jinling Hospital, Nanjing University School of Medicine, Department of Respiratory Medicine; Chinese Alliance Against Lung Cancer Hu, Jiaan; Ruijin Hospital, Shanghai Jiaotong University, Department of Geriatrics Zhang, Yong; Zhongshan Hospital, Fudan University, Department of Pulmonary Medicine; Chinese Alliance Against Lung Cancer; Shanghai Respiratory Research Institution Song, Yuanlin; Zhongshan Hospital, Fudan University, Pulmonary Medicine; Chinese Alliance Against Lung Cancer; Shanghai Respiratory Research Institution

	Zhou, Jian; Department of Pulmonary Medicine, Zhongshan Hospital, Fudan University; Chinese Alliance Against Lung Cancer; Shanghai Respiratory Research Institution Bai, Chunxue; Zhongshan Hospital, Fudan University, Pulmonary Medicine; Chinese Alliance Against Lung Cancer; Shanghai Respiratory Research Institution
Keywords:	LUNG NEOPLASM, BIOMARKERS, DIAGNOSIS, CT, EARLY DETECTION CANCER

Abbreviation List

ACCP	American College of Chest Physician
ADC	adenocarcinoma
AUC	area under the receiver
CAALC	Chinese Alliance Against Lung Cancer
CEA	carcinoembryonic antigen
CMIA	chemiluminescent microparticle immunoassay
CT	computed tomography
CYRFRA21-1	cytokeratin 19 fragment
ECLS	Early Cancer detection test – Lung Cancer Scotland
LCBP	Lung Cancer Biomarker Panel
LDCT	low dose computed tomography
NHS	National Health Service
NLST	National Lung Screening Trial
ProGRP	progastrin-releasing peptide
ROC	receiver-operating-characteristic
SCC	squamous cell carcinoma antigen
SCLC	small cell lung cancer
SD	standard deviation

Probability of Cancer in High Risk Patients Predicted by the Protein-Based Lung Cancer Biomarker Panel in China: LCBP Study

Dawei Yang[1, 10, 11#], Xiaoju Zhang[1, 2, 10#], Charles Powell[3], Jun Ni[4,10], Bin Wang[5, 10], Jianya Zhang[6, 10], Yafei Zhang[7, 10], Lijie Wang[8], Zhihong Xu[9, 10], Li Zhang[4, 10], Guoming Wu[5, 10], Yong Song[6, 10], Jia-an Hu[9, 10], Yong Zhang[1, 10, 11], Yuanlin Song[1, 10, 11], Jian Zhou[1, 10, 11*], Chunxue Bai[1, 10, 11,12*]

[1] Department of Pulmonary Medicine, Zhongshan Hospital, Fudan University, Shanghai, China

[2] Department of Respiratory and Critical Care Medicine, Henan Provincial People's Hospital, Zhengzhou, China

[3] Division of Pulmonary, Critical Care, and Sleep Medicine, Icahn School of Medicine at Mount Sinai, New York, New York

[4] Department of Respiratory Disease, Peking Union Medical College Hospital, Beijing, China

[5] Department of Pulmonary Medicine, Xinqiao Hospital, the Third Military Medical University, Chongqing, China

[6] Department of Pulmonary Medicine, Jinling Hospital, Nanjing University School of Medicine, Nanjing, China

[7] Department of Pulmonary Medicine, the Chest Hospital of Henan Province, Zhengzhou, China

[8] Department of Health Services Management, the Second Military Medical University, Shanghai, China

[9] Department of Geriatrics, RuiJin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[10] Chinese Alliance Against Lung Cancer, China

[11] Shanghai Respiratory Research Institution, Shanghai, China

[12] State Key Laboratory of Respiratory Disease, Guangzhou Medical University, Guangzhou, China.

*Corresponding author: Chunxue Bai, MD & PhD, FCCP, Department of Pulmonary Medicine, Zhongshan Hospital, Fudan University, bai.chunxue@zs-hospital.sh.cn

Co-corresponding author: Jian Zhou, PhD, Department of Pulmonary Medicine, Zhongshan Hospital, Fudan University, zhou.jian@fudan.edu.cn

[#]The first two authors contributed equally to this article.

The work was supported by Shanghai Municipal Leading Academic Discipline Project No. B115, National Basic Research Program of China (973 Program) No.2012CB933304, National Key Scientific and Technology Support Program No.2013BAI09B09, Natural Science Foundation of China No.81500078, No.81400018 and Shanghai Natural Science Foundation No.13ZR1452400. Cost of reagents was partly supported by Abbott Laboratories.

Abstract

RATIONALE: Robust biomarkers have promise to improve efficacy and efficiency of lung cancer early detection programs. We examined the clinical utility of a blood test consisting of four protein-based markers.

METHODS: In this study (NCT01928836), we enrolled 715 participants from five centers in China between October 2012 and February 2014 with high risk of lung cancer. We collected serum biomarker data and clinical information to develop and validate two early diagnosis models. For the patient risk model, we analyzed all cases and for the nodule risk model, we analyzed cases with lung nodules using the logistic regression model. Both models were developed and validated in training and validation data sets.

RESULTS: Serum biomarker (ProGRP, CEA, SCC, CYFRA21-1) were included in the both two models. For the patient risk model, the area under the receiver-operating-characteristic curve (AUC of the ROC curves) was 0.7037 and 0.7190 in training and validation data sets, respectively. For the nodule risk model, the AUC of the ROC curve was 0.9151 and 0.5836 in training and validation data sets. Moreover, compared with the ACCP model, our nodule risk model has a relatively higher AUC (0.9151 vs. 0.8360, P=0.001) in patients with lung nodules.

CONCLUSIONS: We developed and validated two early diagnosis models based on clinical and imaging features combined with a panel of four biomarkers to assist clinical evaluation of lung cancer risk probability. These models are the first serum biomarker based lung cancer early diagnosis models suitable for the Chinese high risk population.

Keywords

Lung neoplasm, biomarker, diagnosis, ct, early detection cancer

Background

Lung cancer has the highest incidence and mortality of all malignant tumors in China[1]. In 2015, it was estimated that there were 610, 200 (432,400 males and 177,800 females) lung cancer deaths in China[2]. Among lung cancer deaths, 87% are related to cigarette smoking[3]. Because the 10-year survival rate of stage I patients is up to 88%, the early diagnosis of lung cancer patients is of key importance[4]. In China, 88.2% of the lung cancer patients are symptomatic when they visit the hospital, and 65.3% of lung cancer patients present with stage III or IV of the disease, at diagnosis[5]. This further underlines the importance of timely detection of early stage lung cancer.

In the National Lung Screening Trial (NLST) (ClinicalTrails.gov number, NCT00047385), low dose computed tomography (LDCT) scan for screening in high-risk population (with a history of cigarette smoking $\geqslant$ 30 pack-years and elder between 55 and 74 years old), reduced lung cancer mortality by 20%[6]. The NELSON trial showed a similar shift more often in cases diagnosed at stage I and less often in those diagnosed at stage IIIB-IV[7]. Final results on lung cancer mortality from NESLSON are awaited. The DANTE trial did not find the test a significant benefit of LDCT screening[8], but the trial was smaller and perhaps underpowered. The potential benefit of screening is diminished by management of false positive results by computed tomography (CT) scan that lead to unnecessary and potential harmful invasive procedures[9]. Currently, in American College of Chest Physician (ACCP) and Chinese Alliance Against Lung Cancer (CAALC) guidelines, the nodule follow-up procedure is mainly dependent on the nodule diameter[10-12].

There are several new blood biomarkers under investigation to help detect early-stage lung cancer[13-14]. Plasma DNA levels, for example, may improve the accuracy of lung cancer CT screening[15]. The efficacy of biomarker panels is increased when combined with clinical parameters[16]. To aid risk assessment of lung cancer risk in Chinese patients and in CT screen detected nodules, we developed a multi-biomarker test in blood. Previously we carried out a preliminary, single center study on a lung cancer biomarker

panel, and found that combined progastrin-releasing peptide (ProGRP), carcinoembryonic antigen (CEA), squamous cell carcinoma antigen (SCC) and cytokeratin 19 fragment (CYRFRA21-1), improved CT scan efficacy for patients with lung cancer[17].

We evaluated this serum biomarker panel using real world data, in the first multi-center Lung Cancer Biomarker Panel (LCBP) study based on Chinese population that was initiated in October 2012, by the Department of Respiratory Diseases, Shanghai Zhongshan Hospital, Fudan University. This study included patients from four other hospitals, including Peking Union Medical College Hospital, Xinqiao Hospital affiliated to the Third Military Medical University, Henan Provincial People's Hospital, and Jinling Hospital, Nanjing University School of Medicine. In this study, we report the training and validation of models and calculators for predicting the probability of lung cancer in (a) high-risk patients, and (b) those with nodules' detected during the course of a CT. This study aims to build a local dataset, which is then used to validate the prediction accuracy of LCBP. The results of this study will help to establish and standardize the practice processes of early diagnosis of lung cancer in China.

Methods

Overall Study Design

This study was designed to evaluate a combination of four serum proteins (ProGRP, CEA, SCC and CYRFRA21-1) and clinical information to predict the occurrence of cancer in lung cancer high risk patients. We developed two models, one for all high risk patients and one for patients with nodules detected by CT. To this end, we developed and validated two prediction models for all high risk patients with or without nodules after CT detection (patient risk model) and for patients with nodules detected by CT (nodule risk model) in a training data set and a validation data set.

The training data set included participants enrolled from Peking Union Medical College Hospital, Henan Provincial People's Hospital and Jinling Hospital, Nanjing University School of Medicine. The validation data set included participants enrolled from Zhongshan Hospital Fudan University and Third Military Medical University.

All patients had blood drawn, and their serum was stored immediately at -80°C in the laboratory repository. Clinical data, including age, sex, race, past medical history, pathological diagnosis, including histology and stage, and imaging findings, including CT data and nodule size, were recorded in a confidential trial database.

Patient Selection

All patients enrolled in this study signed an informed consent form approved by Zhongshan Hospital Fudan University's Institutional Review Board and the four other branch institutions in accordance with an assurance filed with and approved by the Department of Health and Human Services (approval number B2012-069(4)). The registered ClinicalTrails.org number is NCT01928836.

In the training data set, a total of 389 achieved serum samples of lung cancer high risk patients were collected between December 2013 and February 2014 (Table 1,2). Inclusion criteria were: (1) aged 18 to 90 years; (2) male smoker (⩾400 cig/year), female smoker

or non-smoker; (3) no prior history of lung cancer; (4) no currently known extrathoracic malignancy. The nodule risk model had an additional inclusion criterion: a CT of the thorax with an indeterminate pulmonary nodule. Exclusion criteria were (1) absence of histopathological diagnosis; and (2) prior treatment by chemotherapy or surgery. In the validation data set, 326 high risk patients were included between October 2012 and January 2014, with the same inclusion and exclusion criteria (Table 1,2).

Serum Marker Assays

We established a lung cancer biomarker panel (LCBP) that measures the serum levels of ProGRP (1P45, Abbott, U.S.), CEA (7K68, Abbott), SCC (8D18D18, Abbott) and CYFRA21-1 (2P55, Abbott). The serum samples were tested via a commercial chemiluminescent microparticle immunoassay (CMIA) method with ARCHITECT i2000SR, an automated immunoassay analyzer (Abbott Laboratories, Chicago, Ill). All samples in the training set were run in duplicate. For the validation study, all assays were run on automated analyzers per the manufacturer's instructions.

Imaging Studies, Nodule Size, and Pathological Classification

All patients enrolled in either the training or validation arm had a CT study of the thorax. CT studies at five institutions were performed using multiple contiguous sequential axial images through the thorax. The size of the nodule was analyzed by Pneuview (Myrian, Pairs). The maximum dimension on axial CT images was recorded. Histological diagnoses were performed according to the criteria of the 2015 WHO classification[18], and stage classifications followed the 7th edition AJCC TNM staging criteria[19]. Patients without lung cancer had an alternative histologic diagnosis, CT resolution, 2-year stability of the nodule, or clinical observation without evidence of lung cancer for 2 years.

Statistical Methods

Continuous data were presented with mean ± standard deviation (SD) and analyzed with Student's t-test followed by ANOVA analysis, while categorical data were expressed

as n (%). Multivariate logistic regression was prepared to evaluate the risk of lung cancer associated with potential predictors based on the existing knowledge of risk factors for lung cancer, such as domestic information, clinical characteristics and serum levels of LCBP parameters (ProGRP, CEA, SCC, CYFRA21-1).

Two predictive models (patient risk model and nodule risk model) were developed. For the patient risk model, we included the following variables in the model: age, sex, smoke status, history of cancer and serum expression level of ProGRP, SCC, CYFRA21-1, CEA. For the nodule risk model, we included the following variables: age, sex, smoke status, diameter of nodules, spiculation, and serum expression level of ProGRP, SCC, CYFRA21-1, CEA. Prediction models were developed in the training data set and evaluated for their predictive performance in the validation data set, including their discrimination (ability of classify correctly) and calibration (comparability between predicted and observed probabilities). The discrimination was measured with the area under the receiver-operating-characteristic curve (AUC of ROC curve). The sensitivity, specificity and accuracy were calculated for all models. The calibration was evaluated by subtracting the model-estimated probability from observed probability for each subject in this study. According to the ROC curve, patients were further classified into the following risk strata: low-risk, intermediate-risk, and high-risk. For comparison, we applied the Mayo Clinic model suggested by the ACCP in validation set[10]. The prediction model of Mayo Clinic is described by the following equations: Probability of malignancy = e^x / [1 + e^x], x = -6.8272 + [0.0391 × age] + [0.7917 × smoke status] + [1.3388 × cancer history] + [0.1274 × nodule diameter] + [1.0407 × spiculation] + [0.7838 × location]. where e is the base of natural logarithms, age is the patient's age in years, smoke status = 1 if the patient is a current or former smoker (otherwise = 0), cancer history = 1 if the patient has a history of an extrathoracic cancer that was diagnosed > 5 years ago (otherwise = 0), nodule diameter is the diameter of the nodule in millimeters, spiculation = 1 if the edge of the nodule has spicules (otherwise = 0), and location = 1 if the nodule is located in an upper lobe (otherwise = 0). All P values are two-sided, and a P value less

than 0.05 indicates statistical significance. The statistical analysis was performed with SPSS, version 20.0.

Results

Study Population

In the patient risk model study, a total of 715 subjects (389 subjects in training data set and 326 subjects in validation data set) with high risk of lung cancer were enrolled from five centers in China. In the lung nodule risk model study, a total of 342 subjects (163 subjects in training data set and 179 subjects in validation data set) were found to have nodules. The characteristics of subjects in training and validation data set are described in Table 1 and Table 2.

Biomarker Level in the Total Population

Significant higher levels of ProGRP (P<0.001), SCC (P<0.001), CYFRA21-1 (P<0.001) and CEA (P=0.026) were observed in lung cancer patients in comparison with the control group, who had no indication for lung biopsy (Figure 2A). Similar results were obtained when we compared the lung cancer patients (ProGRP, P<0.001; SCC, P<0.001; CYFRA21-1, P<0.001; CEA, P=0.310) with patients with normal biopsy result. CEA levels were not different in cancer patients vs. patients with a non-malignant nodule. As expected, significantly higher level of ProGRP for small cell lung cancer was observed in comparison with both adenocarcinoma (P<0.001) and squamous cell carcinoma (P<0.001). Further, the SCC marker was higher for squamous cell cancers compared with adenocarcinoma and small cell cancer. (Figure 2B).

Figure 2C presents the serum levels of the four biomarkers for lung cancer patients with nodules of diameter less than 20 mm and those with lung nodule or lesion of diameter greater than 20 mm. Significantly higher level of ProGRP (P=0.036) was observed in patients with larger malignant lesion than in those with smaller malignant lesion. Similar results were obtained when we compared the advanced stage (IIIB-IV) patients (CEA, P<0.01) with patients with limited stage (I-IIIA) (Figure 2D).

Patient Risk Model

We analyzed 389 lung cancer high risk participants without using nodule information from the CT scan result to develop the patient risk model. After logistic regression, ProGRP, CEA, SCC, CYFRA21-1, age, smoke status, cancer history, and sex were used to predict the diagnosis of lung cancer in lung cancer high risk population (Table 3). In the validation data set, we analyzed 326 participants with high risk to validate the patient risk model.

According to the logistic regression model, we obtained an equation to calculate the probability of cancer in the high risk population (Probability of malignancy = e^x / [1 + e^x], x = -2.9169 + [0.03 × age] + [1.0721 × smoke status] + [0.306 × cancer history] + [-0.7012 × sex] + [-0.000155 × ProGRP] + [0.0151 × SCC] + [0.1238 × CRFRA21-1] + [-0.00043 × CEA]), where e is the base of natural logarithms, age is the patient's age in years, smoke status = 1 if the patient is a current or former smoker (otherwise = 0), cancer history = 1 if the patient has a history of an extrathoracic cancer that was diagnosed > 5 years ago (otherwise = 0), sex = 1 if the patient is male (otherwise = 0). A patient with a calculated cancer probability of 0.65 (cut-off value 1) or higher was considered to have cancer (high-risk), and one with a probability less than 0.21 (cut-off value 2) was considered not to have cancer (low-risk). The cut-off values were established in the training set and tested in the validation set.

Both in training and validation data sets, the patient risk model also showed excellent discrimination with AUCs of 0.7037 and 0.7190 (Figure 3A and 3B). The overall sensitivity and specificity in the training data set were 94.2% and 94.5%, respectively, and those in the validation data set were 92.9% and 99.5% (Table 5).

Nodule Risk Model

Combining the CT scan data, we analyzed a total of 163 participants with nodules from the training data, and we developed the nodule risk diagnosis model. After logistic

regression, ProGRP, CEA, SCC, CYFRA21-1, age, smoke status, nodule diameter, spiculation and sex were used to predict diagnosis lung cancer in population with nodules (Table 3). In the validation data set, we analyzed 179 participants with nodules to validate the nodule risk model.

According to the logistic regression model, we obtained an equation to calculate the probability of cancer (Probability of malignancy = e^x / [1 + e^x], x = - 5.6017 + [0.0264 × age] + [8.8539 × smoke status] + [0.1859 × nodule diameter] + [3.1865 × spiculation] + [-8.7109 × sex] + [-0.00001 × ProGRP] + [0.0057 × SCC] + [0.1686 × CRFRA21-1] + [-0.00311 × CEA]), where e is the base of natural logarithms, age is the patient's age in years, smoke status = 1 if the patient is a current or former smoker (otherwise = 0), nodule diameter is the diameter of the nodule in millimeters, spiculation = 1 if the edge of the nodule has spicules (otherwise = 0), sex = 1 if the patient is male (otherwise = 0). A patient with a calculated cancer probability of 0.94 (cut-off value 1) or higher was considered to have cancer (high-risk), and one with a probability less than 0.22 (cut-off value 2) was considered not to have cancer (low-risk).

Both in training and validation data sets, the nodule risk model showed excellent discrimination with AUCs of 0.9151 and 0.5836 (Figure 4A and 4B). The maximum sensitivity and specificity in the training data set were 94.6% and 94.2%, respectively, and were 85.6% and 37.5% in the validation data set (Table 5). Table 3 and Table 4 show that only CYFRA21-1 is independently associated with risk of cancer with $P< 0.05$. The other biomarkers are not significant except for CEA in Table 4; and CEA levels are inversely correlated with lung cancer risk in Table 4. In addition, we compared the nodule risk model and ACCP model for the prediction of lung cancer diagnosis in the 163 participants with nodules in the training data set. In this population, the AUC of ACCP model was 0.8360, which was less than that of the nodule risk model (P=0.001) (Figure 5).

Discussion

Many factors increase the delay of diagnosis of lung cancer, such as failure to recognize abnormal imaging results, failure to receive key diagnostic procedures in a short period and errors in understanding of pathological reports[20-21]. Current early diagnosis procedure for lung cancer exists primarily in comprehensive hospitals, which involve expensive examination costs and huge human resources to collaborate. There are two prospective clinical lung cancer screening trials based on blood biomarker panels: one being the Early Cancer detection test – Lung Cancer Scotland (ECLS Study) conducted by National Health Service (NHS)[22], and the other conducted by National Jewish Health hospital[23]. Previously, we performed a preliminary single center study and found combined biomarker panel improved CT scan efficacy for patients with lung cancer[17]. In the current study which is the first multi-center Lung Cancer Biomarker Panel (LCBP) study based on a Chinese population, we evaluated this serum biomarker panel on real world data. It can serve as a useful aide to help clinicians at every level to comprehensively and scientifically evaluate the risk of lung cancer.

A total of 715 participants with high risk factors for lung cancer were enrolled in this study, including 294 patients with malignant tumor. We developed and validated two models for early lung cancer early diagnosis based on clinical information and serum protein biomarkers in a relatively large Chinese population. For both early diagnosis models, we observed the AUCs were excellent with relatively high sensitivity and specificity. In the population with lung nodules, compared with the ACCP model, our nodule risk model showed a better ability for accurate classification (Accuracy 62.6% vs. 82.2%, AUC 0.8360 vs. 0.9151, $P<0.01$).

As multi-slice spiral CT and LDCT scans are widely used to screen the subjects at high risk of lung cancer, the detection of solitary pulmonary nodules is significantly increased, ranging from approximately 8% to 51%[24]. In particular, the solitary pulmonary nodules of diameters less than 1 cm are now detected much more frequently. Only 1.1% to 12.0% of these solitary nodules are malignant[25]. These malignant tumors are difficult to

differentiate from the benign ones due to the limitations of current diagnostic methods[26], especially for solitary pulmonary nodules of diameter less than 8 mm. The lung cancer patients diagnosed at an early stage will have significantly better survival if they are treated by an experienced respiratory physician, rather than by a non-respiratory physician. The nodule risk models can be used to evaluate the risk of lung cancer prior to surgery in patients with or without pulmonary nodules. The data from our study in a high risk Chinese population have shown the sensitivity and specificity of the models to be nearly 90% in identifying lung cancer. Such a screening tool will prove helpful for doctors at different levels, especially the doctors working in community centers and remote areas. Identifying malignant nodules at an early stage is favorable for selecting more appropriate treatment strategies to achieve optimal treatment efficacy.

Conclusions

Lung cancer screening based on low dose CT detected around one third participated with indeterminate abnormalities, which needed further evaluation and created an undesirable burden on the healthcare system. Currently, the guideline recommended multiple sequential CT scan to detect the growth of the pulmonary nodule, or with PET and invasive procedure, which depend on risk stratification based on risk for lung cancer and size of the lesion. Several studies have shown application of tumor-specific, noninvasive biomarkers could modify the paradigm, while most studies are focused on Caucasian population.

The current study is an extension of our previous single center study on a panel of serum markers to be used for the diagnosis of lung cancer in China. In this prospective, case-control and multicenter study, we firstly combined these markers with the most predictive imaging feature of lung cancer, nodule size, to stratify Chinese patients with indeterminate nodules into low- and high-risk groups. We developed and validated two models for the clinical application. The patient risk model is suggested for people with high risk factor of lung cancer living in rural area of Chinese, since it is without nodule information from CT scan. By this biomarker based panel, the local physician could efficiently stratification the population with high risk of lung cancer and transfer them for the further examination. While the nodule risk model is combined with CT scan result, which demonstrated better accuracy on risk predication than ACCP lung nodule risk model. Those individuals assigned to a high-risk category suggested to receive more immediate intervention with a biopsy, PET, or close surveillance. While patients in the low-risk groups could be followed with less frequent sequential CT scan. So, it could be applied to evaluate a patient with lung nodule before and to avoid the unnecessary invasive procedure, such as biopsy or surgery.

We enrolled patients in this study in the content of several tertiary medical centers, additional studies are designed to determine the impact of integrating serum biomarker result with nodule size among a screening program. Currently, we are launching a

large-scale lung cancer screening program based on these two models and it is expected to improve early diagnosis and decrease mortality of lung cancer in China. These studies are expected to define the frequency and use of sequential follow-up imaging or repeat serum marker evaluation for a more efficient and effective approach for high-risk patients with indeterminate pulmonary nodules.

Contributions of authors

Dawei Yang and Xiaoju Zhang had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Dawei Yang contributed to the study design, laboratory work, intellectual discussion of the results, statistical analysis, and writing of the article and approved the submitted article. Xiaoju Zhang contributed to the study design and subject recruitment and approved the submitted article. Jun Ni, Bin Wang, Jianya Zhang, Yafei Zhang, Lijie Wang, Zhihong Xu, Li Zhang, Guoming Wu, Yong Song, Jia-an Hu, Yong Zhang and Yuanlin Song contributed to the subject recruitment and approved the submitted article. Charles Powell contributed to the study design and intellectual discussion of the results and approved the submitted article. Jian Zhou and Chunxue Bai contributed to the application for the study grant, study design, subject recruitment, and intellectual discussion of the results and approved the submitted article.

Acknowledgements

We thank the study participants and their families and are grateful for the collaboration received from the Zhongshan Hospitals, Fudan University, Henan Provincial People's Hospital, Peking Union Medical College Hospital, Xinqiao Hospital, the Third Military Medical University and Jinling Hospital, Nanjing University School of Medicine and the staff.

Reference

1 Yang G, Hu J, Rao KQ, Ma J, Rao C, Lopez AD. et al. Mortality registration and surveillance in China: history, current situation and challenges. *Popul Health Metr.* 2005;3(1):3.

2 Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66(2):115-132.

3 Godtfredsen NS, Prescott E, Osler M. Effect of smoking reduction on lung cancer risk. *JAMA*. 2005;294(12):1505-1510.

4 Overholt RH, Woods FM. Early diagnosis and treatment of cancer of the lung. *N Eng J Med*. 1951;245(15):555-559.

5 Peng H, Han BH, LI XQ, Tao LN. Analysis of clinical characteristics and survival rates in 1279 lung cancer patients. *China Oncology*. 2011;21(5):354-358.

6 National Lung Screeing Trial Research Team, Aberle Dr, Adams Am, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395-409.

7 Horeweg N, van der Aalst CM, Thunnissen E, et al. Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial. *Am J Respir Crit Care Med*. 2013;186(8):848-854.

8 Infante M, Cavuto S, Lutman FR, et al. Long-term follow-up results of the DANTE trial, a randomized study of lung cancer screening with spiral computed tomography. *Am J Respir Crit Care Med*. 2015;191(10):1166-1175.

9 Wilson DO, Weissfeld JL, Fuhrman CR, et al. The Pittsburgh Lung Screening Study (PLuSS) outcomes within 3 years of a first computed tomography scan. *Am J Respir Crit Care Med*. 2008;178(9):956-961.

10 Gould MK, Donington J, Lynch WR, et al. Evaluation of individuals with pulmonary nodules: When is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2013;143(5 Suppl):e93S-e120S.

11 Herder GJ, van Tinteren H, Golding RP, et al. Clinical prediction model to

characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. *Chest*. 2005;128(4):2490-2496.

12 Lung cancer group of the Chinese Medical Association, Chinese Alliance Against Lung CancerC. Diagnosis and treatment of pulmonary nodules in chinese expert consensus. *Chin J Tuberc Respir Dis*. 2015;38(4):249-254.

13 Blasberg JD, Pass HI, Goaraju CM, Flores RM, Lee S, Donington JS. Reduction of Elevated Plasma Osteopontin Levels With Resection of Non–Small-Cell Lung Cancer. *J Clin Oncol*. 2010;28(6):936-941.

14 Qiu J, Choi G, Li L, et al. Occurrence of autoantibodies to annexin I, 14-3-3 theta and LAMR1 in prediagnostic lung cancer sera. *J Clin Oncol*. 2008;26(31):5060-5066.

15 Sozzi G, Roz L, Conte D, et al. Plasma DNA quantification in lung cancer computed tomography screening: five-year results of a prospective study. Am *J Respir Crit Care Med*. 2009;179(1):69-74.

16 Yee J, Sadar MD, Sin DD, et al. Connective tissue-activating peptide III: a novel blood biomarker for early lung cancer detection. *J Clin Oncol*. 2009;27(17): 2787-2792.

17 Yang DW, Zhang Y, Hong QY, et al. Role of a serum-based biomarker panel in the early diagnosis of lung cancer for a cohort of high-risk patients. *Cancer*. 2015; 121(17 Suppl):3113-3121.

18 Travis WD, Brambilla E, Burke AP, Marx A, Nicholson AG. *WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart.* International Agency for Research on Cancer, Lyon;2015.

19 Goldstraw P, Crowley J, Chansky K, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. *J Thorac Oncol*. 2007;2(8):706-714.

20 Singh H, Hirani K, Kadiyala H, et al. Characteristics and predictors of missed opportunities in lung cancer diagnosis: an electronic health record–based

study. *J Clin Oncol*. 2010;28(10):3307-3315.

21 Singh H, Sethi S, Raber M, Petersen LA. Errors in cancer diagnosis: current understanding and future directions. *J Clin Oncol*. 2007;25(31):5009-5018.

22 Jett JR, Peek LJ, Fredericks L, Jewell W, Pingleton WW, Robertson JF. Audit of the autoantibody test, EarlyCDT®-lung, in 1600 patients: an evaluation of its performance in routine clinical practice. *Lung Cancer*. 2014;83(1):51-55.

23 Birse CE, Lagier RJ, FitzHugh W, et al. Blood-based lung cancer biomarkers identified through proteomic discovery in cancer tissues, cell lines and conditioned medium. *Clin Proteomics*. 2015;12(1):18.

24 Brawley OW, Flenaugh EL. Low-dose spiral CT screening and evaluation of the solitary pulmonary nodule. *Oncology*. 2014;28(5):441-446.

25 Shin KE, Lee KS, Yi CA, Chung MJ, Shin MH, Choi YH. Subcentimeter lung nodules stable for 2 years at LDCT: Long-term follow-up using volumetry. *Respirology*.2014;19(6):921-928.

26 Slatore CG, Horeweg N, Jett JR, et al. An Official American Thoracic Society Research Statement: A Research Framework for Pulmonary Nodule Evaluation and Management. *Am J Respir Crit Care Med*. 2015;192(4):500-514.

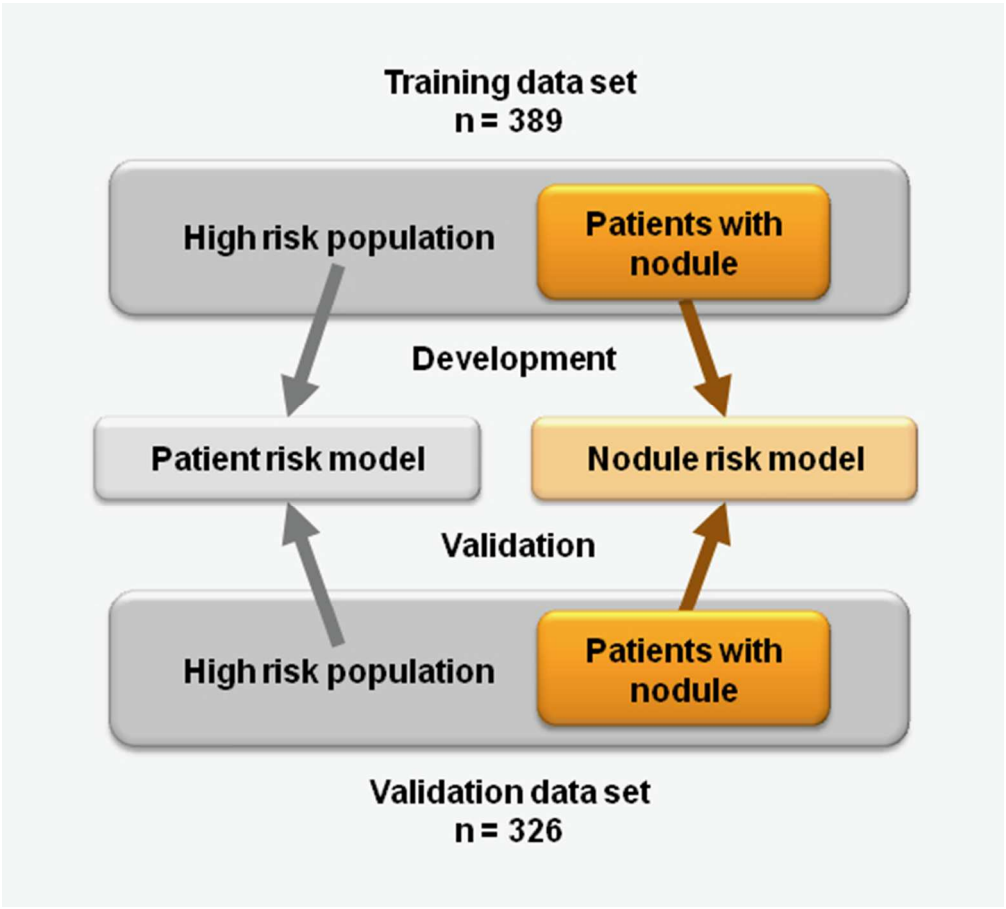

Figure 1. Study strategy schematic diagram.

Figure 1
207x187mm (72 x 72 DPI)

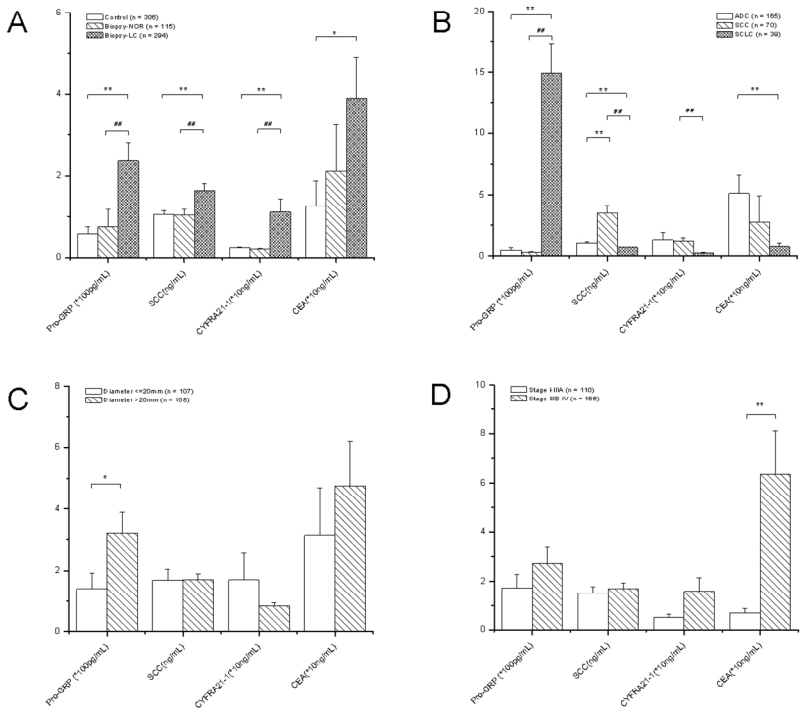

Figure 2. Levels of ProGRP, SCC, CYFRA21-1 and CEA in the total population. A. Biomarker level in participants with and without lung cancer. Control, patients with no indication for lung biopsy. Biopsy-NOR, patients with normal biopsy. Biopsy-LC, patients diagnosed with lung cancer. * P<0.05, ** P<0.01 vs. control population. # P<0.05, ## P<0.01 vs. biopsy-NOR population. B. Biomarker level in participants with different histology. ADC, adenocarcinoma. SCC, squamous cell carcinoma. SCLC, small cell lung cancer. ** P<0.01 vs. patients with adenocarcinoma. ## P<0.01 vs. patients with squamous carcinoma. C. Biomarker level in patients with malignant nodules with different nodule size. * P<0.05 vs. patients with smaller nodules. D: Biomarker level in participants with different tumor stages. ** P<0.01 vs. patients with lower stage.

Figure 2
543x472mm (72 x 72 DPI)

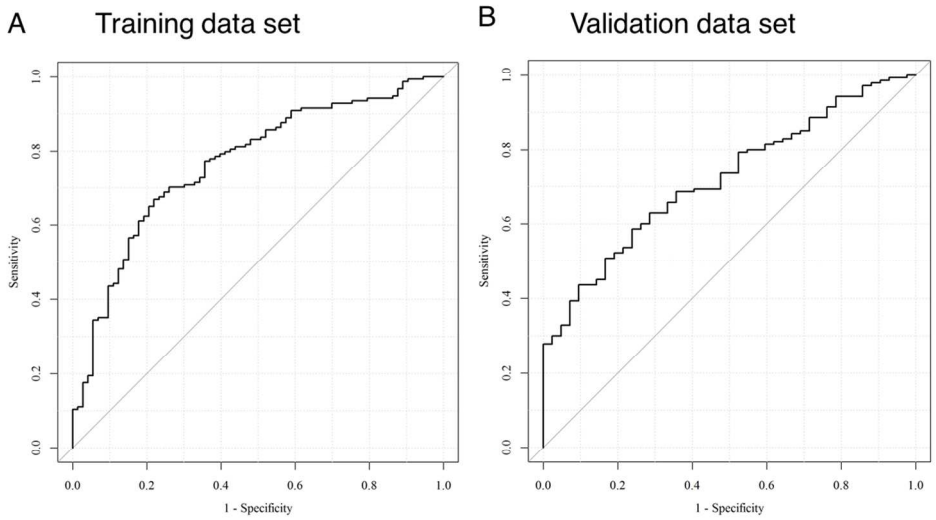

Figure 3. ROC curve of patient risk model in all participants. A. training data set. B. validation data set.
Figure 3

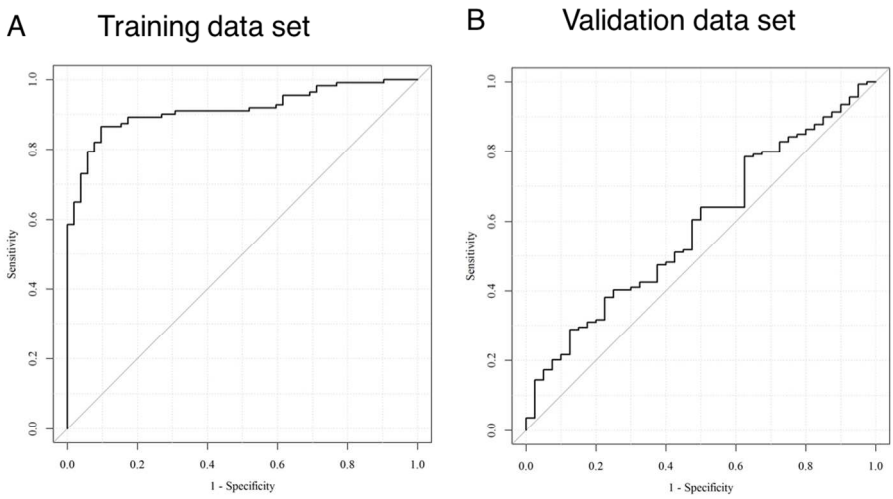

Figure 4. ROC curve of nodule risk model in participants with nodules. A. training data set. B. validation data set.
Figure 4

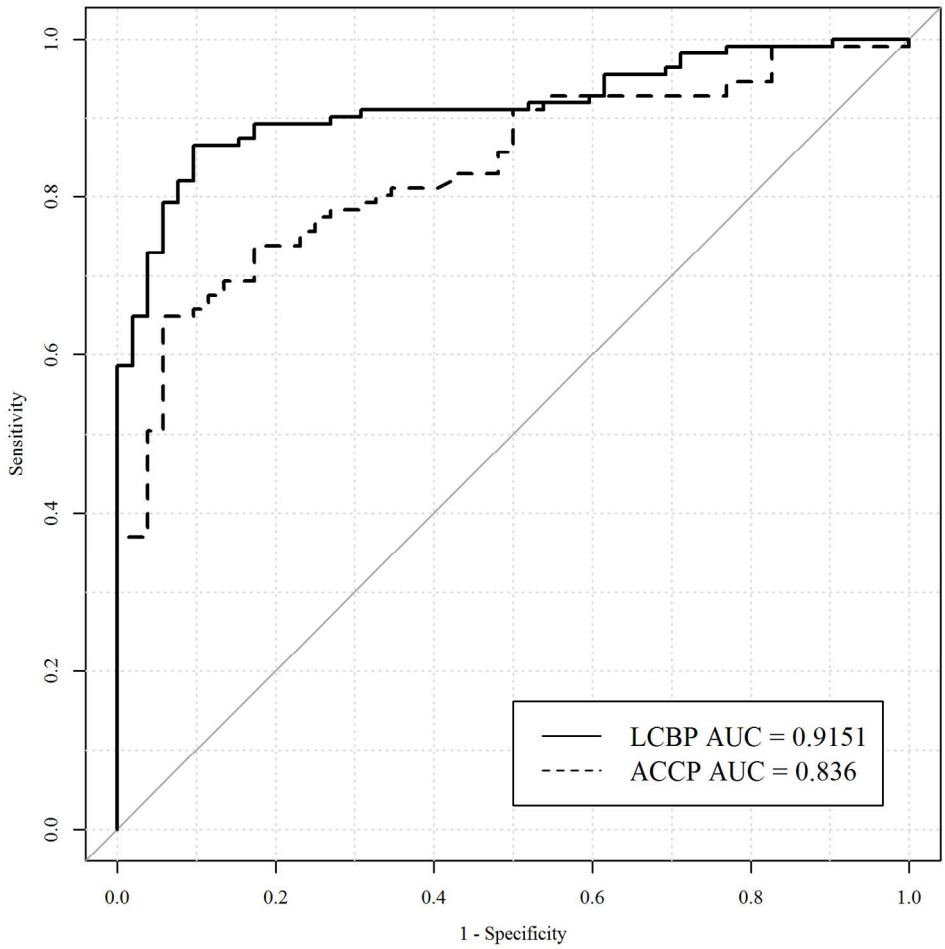

Figure 5. ROC comparison between nodule risk model and ACCP model.
Figure 5
592x594mm (72 x 72 DPI)

Table 1. Training data set: patients' demographic and clinical profiles

Demographic	Cancer (*n = 154*)	Control (*n = 235*)
Age, yr	60.1 ± 8.3	57.1 ± 10.2
Range	40-81	33-83
Sex, number of patients		
Male	92 (59.87)	118 (50.00)
Female	62 (40.12)	117 (50.00)
Smoke, number of patients		
Smoker	94 (61.14)	115 (48.73)
Non-smoker	60 (38.85)	120 (51.27)
Stage, number of patients		
I	23 (15.29)	
II	16 (10.19)	
III	42 (27.39)	
IV	60 (38.85)	
N/A	13 (8.28)	
Histology, number of patients		
Small cell lung cancer	15 (9.55)	
Non-small cell lung cancer	133 (86.62)	
Not specified	6 (3.82)	

Table 2. Validation data set: patients' demographic and clinical profiles

Demographic	Cancer (*n = 140*)	Control (*n = 186*)
Age, yr	59.8 ± 10.3	56.4 ± 11.4
Range	31-85	21-90
Sex, number of patients		
Male	100 (71.42)	115 (61,83)
Female	40 (28.57)	71 (38.17)
Smoke, number of patients		
Smoker	77 (55.00)	97 (52.15)
Non-smoker	63 (45.00)	89 (47.85)
Stage, number of patients		
I	25 (17.86)	
II	11 (7.86)	
III	31 (22.14)	
IV	68 (48.57)	
N/A	5 (3.57)	
Histology, number of patients		
Small cell lung cancer	24 (17.14)	
Non-small cell lung cancer	111 (79.29)	
Not specified	5 (3.57)	

Table 3. Patient risk model for probability of lung cancer participants with high risk of lung cancer

Variable	Odds ratio (95% CI)	P value	Beta Coefficient
Age	1.03 (1.006, 1.055)	0.0127	0.03
Smoke status	2.922 (0.825, 10.345)	0.0965	1.0721
Cancer history	1.358 (0.216, 8.519)	0.7439	0.306
Sex	0.496 (0.14, 1.753)	0.2763	-0.7012
ProGRP	1 (1, 1.001)	0.4453	0.000155
SCC	1.015 (0.916, 1.125)	0.7723	0.0151
CYFRA21-1	1.132 (1.063, 1.205)	<.0001	0.1238
CEA	1 (0.998, 1.001)	0.5882	-0.00043
Model constant			-2.9169

Table 4. Nodule risk model for probability of lung cancer for participants with nodules

Variable	Odds ratio (95% CI)	P value	Beta Coefficient
Age	1.027 (0.973, 1.084)	0.3371	0.0264
Smoke status	>999.999 (7.045, >999.999)	0.0119	8.8539
Nodule diameter	1.204 (1.097, 1.322)	<.0001	0.1859
Spiculation	24.204 (4.206, 139.284)	0.0004	3.1865
Sex	<0.001 (<0.001, 0.175)	0.0143	-8.7109
Pro GRP	1 (0.999, 1.001)	0.9784	-0.00001
SCC	1.006 (0.82, 1.233)	0.9563	0.0057
CYFRA 21-1	1.184 (1.038, 1.35)	0.0121	0.1686
CEA	0.997 (0.994, 1)	0.0489	-0.00311
Model constant			5.6017

Table 5. Statistical efficacy of patient risk model and nodule risk model

			Accuracy	Sensitivity in Percent	Specificity in Percent
Patient Risk Model	CT		73.1	98.8	72.8
	Biomarker	Training	63.0	94.2	94.5
		Validation	69.0	92.9	99.5
Nodule Risk Model	CT		62.6	99.1	96.2
	Biomarker + CT	Training	82.8	94.6	94.2
		Validation	70.9	85.6	37.5