Observational Study

Medicine®

OPEN

Combined use of salivary biomarkers and carcinoembryonic antigen for lung cancer detection in a Chinese population

Xi Shang, MD[a], Huanpu Zi, MD[b], Yangyang Li, MD[a], Yu Gao, MD[a], Changsheng Ge, MD[c], Zhongmei Sun, MD[c], Youmin Zhang, MD[a,*]

Abstract

Blood-based biomarkers, such as carcinoembryonic antigen (CEA), and saliva-based biomarkers, such as mRNA, have emerged as potential liquid biopsies for non-invasive detection of many cancers. However, current tests typically use single type of biomarkers, and their sensitivity and specificity is often unsatisfactory.

In this study, we developed a novel biomarker panel that measures both CEA level in blood and GREB1 and FRS2 levels in saliva to achieve high sensitivity and high specificity in detecting Non-Small Cell Lung Cancer (NSCLC).

In the discovery phase, we achieved sensitivity of 96.67% and specificity of 93.33% for 30 NSCLC patients and 30 healthy controls. To further evaluate the prediction performance of our biomarker panel, we applied it to an independent set of 15 NSCLC cancer patients and 25 healthy controls. The sensitivity and specificity of our test reached 93.33% and 80.00% respectively.

Our study discovered that the combined analysis of CEA and mRNA can be a novel liquid-biopsy technology for non-invasive detection of NSCLC.

Abbreviations: AUC = area under the curve, CEA = carcinoembryonic antigen, NSCLC = non-small cell lung cancer, ROC = receiver operating characteristic curve.

Keywords: cancer, carcinoembryonic antigen, liquid biopsy, lung, non-small cell lung cancer, RNA, saliva

1. Introduction

Lung cancer is the leading cause of cancer death in both sexes combined and in men in the world.[1] In China, lung cancer is the leading cause of cancer death among both men and women.[2] The primary reason for lung cancer is smoking, which accounts for 90% of all cases.[3–5] In China, cigarette consumption has continued to increase since 1960s, and seen a marked increase since 2000.[6] The main cause for the high death rate of lung cancer is that early detection can be very difficult. In the early stage, the 5-year survival rate of lung cancer is 56%, but dramatically drops to 5% in the late stage.[7]

Carcinoembryonic antigen (CEA) is one of the most commonly used cancer biomarkers in blood test for detection of disease recurrence, and response or progression in metastatic disease.[8] Since it is measured in whole peripheral blood, CEA can be used as a non-invasive prognostic biomarker to monitor cancer progression.[9] However, CEA tend to have a high false positive rate, in which healthy individuals might receive positive results due to benign lung diseases (e.g., tuberculosis). Hence, CEA cannot serve as a sole predictor of cancer detection.[10]

Transcriptomic salivary biomarkers have been recently studied as potential non-invasive biomarkers for detection of lung cancer. Five mRNA biomarkers, namely CCNI, FGF19, GREB1, FRS2, EGFR, have been suggested for discrimination of lung cancer patients from healthy controls.[11] Using the aforementioned combination of mRNA biomarkers, Zhang et al achieved the sensitivity of 93.75% and specificity of 82.81% in their pre-validation sample set of 32 patients and 64 healthy controls. However, most clinical applications of salivary RNA biomarkers were currently conducted in populations of the United States.[11–20] Whether or not these findings could be suitable for lung cancer detection in Chinese populations remain largely unknown.

The goal of this study is to explore the possibility of combining a panel of salivary RNA biomarkers with CEA detection in blood to detect non-small cell lung cancer (NSCLC) in a Chinese population. To this end, a total of 30 NSCLC patients and 30 healthy controls were recruited in the initial discovery phase. The expression levels of CCNI, EGFR, FGF19, FRS2, GREB1 in saliva and the CEA level in blood were analyzed. The sensitivity and specificity of our combined analysis of RNAs and CEA reached 96.67% (29/30) and 93.33% (28/30), respectively. In a subsequent validation phase that included 15 NSCLC patients and 25 healthy controls, the sensitivity and specificity of our method remained as high as 93.33% (14/15) and 80.00% (20/25). Our study discovered that by combining the analysis of

Editor: Patricia Severino.

The authors have no conflicts of interest to disclose.

[a] Department of Thoracic Surgery, [b] Department of Oncology Medicine, [c] Department of Respiratory Medicine, Rizhao City Traditional Chinese Medicine Hospital, Rizhao, Shandong, China.

[*] Correspondence: Youmin Zhang, Department of Thoracic Surgery, Rizhao City Traditional Chinese Medicine Hospital, 35 Wanghai Road, Donggang District, Rizhao, Shandong 276800, China (e-mail: ymzhang_rzhtcm@163.com).

Medicine (2019) 98:31(e16511)

Received: 10 April 2019 / Received in final form: 1 June 2019 / Accepted: 24 June 2019

http://dx.doi.org/10.1097/MD.0000000000016511

Table 1

Demographic information of all subjects used in this study.

Demographic variable	Characteristics	Discovery phase		Validation phase	
		NSCLC (n = 30)	Healthy control (n = 30)	NSCLC (n = 15)	Healthy control (n = 25)
Age, y	Mean ± SD	63.03 ± 11.09	63.07 ± 15.34	60.07 ± 10.48	51.09 ± 16.05
Gender, n (%)	Male	20 (66.7%)	23 (76.7%)	8 (53.3%)	14 (56.0%)
	Female	10 (33.3%)	7 (23.3%)	7 (46.7%)	11 (44.0%)
Ethnicity, n (%)	Asian (China)	30 (100%)	30 (100%)	15 (100%)	25 (100%)
Stage, n (%)	I–II	7 (30.4%)	0 (0%)	4 (26.7%)	0 (0%)
	II–IV	23 (69.6%)	0 (0%)	11 (73.3%)	0 (0%)

RNAs in saliva and CEA in blood, accurate and non-invasive detection of NSCLC in Chinese population could be indeed achieved. To our best knowledge, it is the first time that multiple biomarkers from saliva and blood were used for lung cancer detection in Chinese populations.

2. Materials and methods

2.1. Patient cohort

In the discovery phase, we recruited totally 30 NSCLC patients and 30 NSCLC-negative people (i.e., healthy controls) between June 3, 2017 and March 11, 2018. In the independent clinical validation phase, we recruited totally 15 NSCLC patients and 25 NSCLC-negative people (i.e., healthy controls) between April 6, 2018 and January 11, 2019. All the patients and healthy controls were recruited from Rizhao City Traditional Chinese Medicine Hospital (Rizhao, China). Written informed consents and questionnaire data sheets were obtained from all patients who agreed to serve as saliva and blood donors. The information on patient characteristics, such as age and sex, is presented in Table 1.

This study was approved by the institutional review board at the Rizhao City Traditional Chinese Medicine Hospital and conducted in adherence to the Declaration of Helsinki Principles. Inclusion criteria of NSCLC patients consisted of a confirmed diagnosis of NSCLC via histology or cytology. Exclusion criteria of NSCLC patients included chemotherapy/radiotherapy within 1 month, a diagnosis of other malignancies within 5 years prior to blood collection, clotted, or hemolysis patient samples.

2.2. Sample collection

Both whole peripheral blood samples and unstimulated whole saliva samples were blindly collected following a standard procedure between 08:00 and 09:00 in the morning from NSCLC patients and healthy controls. All blood samples (2 mL) were collected using tubes without anti-coagulant for CEA analysis. Serum was separated in 30 minutes after collection, divided into 4 aliquots, and stored at –25 °C until assays. All saliva samples (2 mL) were collect at least 1 hour after the subjects were asked to refrain from eating, drinking, smoking, or oral hygiene procedures. Saliva samples were kept on ice during collection and were then centrifuged at 13,000 × *g* for 10 minutes at 4 °C. For preservation of salivary RNA, the supernatant was removed from the pellet and treated with RNase inhibitor (Superase-In, Ambion Inc., Austin, TX) and stored at –80 °C prior to assay. The samples in the validation phase were stored and measured in a different batch from that of discovery phase.

2.3. Analysis of CEA level in blood samples

CEA concentrations were determined by microparticle enzyme immunoassays using ARCHITECT CEA kit (Abbott, Ireland) and ARCHITECT *i* System (Abbott, Ireland) following the manual. Briefly, serum samples and anti-CEA coated paramagnetic microparticles were combined. CEA present in the sample bound to the anti-CEA coated microparticles. After washing, anti-CEA acridinium-labeled conjugate was added in the next step. Pre-Trigger and Trigger Solutions were then added to the reaction mixture; the resulting chemiluminescent reaction was measured as relative light units (RLUs). A direct relationship existed between the amount of CEA in the sample and the RLUs detected by the ARCHITECT *i* optical system.

2.4. Analysis of RNA in saliva samples

We followed a protocol that was previously published for extracting salivary RNA.[11] In brief, the supernatant of 330 μL saliva supernatant was used and processed. Extraction of saliva RNA was conducted by using RNeasy Protect Saliva Mini Kit (Qiagen, Germany) as instructed by the product manual. After extraction, RNA samples were stored at –80 °C and analyzed within 12 hours of sample preparation. Extracted RNA was linearly amplified with TURBO DNase treatment. Quantitative real-time PCR (qPCR) analysis was conducted to detect the expression levels of selected genes in saliva samples. Amplification was performed using Roche LightCycler 480 (Roche, Switzerland) in 20 μL reactions containing 2 μL cDNA, the primers (Table 2), the probe, and the reaction mix with qPCR. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as an internal control. The primers used in qPCR analysis were listed in Table 2. Raw Ct data were normalized by subtracting GAPDH Ct values

Table 2

Primers used in this study.

Primer name	Sequence
CCNI_F	CTACCGTAAAGGCTCATCCAAAA
CCNI_R	GAAACTGTCTCTTGCCAATACCT
EGFR_F	CCCACTCATGCTCTACAACCC
EGFR_R	TCGCACTTCTTACACTTGCGG
FGF19_F	CGGAGGAAGACTGTGCTTTCG
FGF19_R	CTCGGATCGGTACACATTGTAG
GREB1_F	CTGTACCACAGACGGGTTTTG
GREB1_R	TTCCGTGAAGTAACAGAAGCC
FRS2_F	CCTGCGACGCTATGGCTATG
FRS2_R	ACGGGCACACTTAAAGGCAAA
GAPDH_F	CTGGGCTACACTGAGCACC
GAPDH_R	AAGTGGTCGTTGAGGGCAATG

Table 3

Confusion matrix of combined biomarker analysis in an independent clinical validation phase.

		Actual class		
		Cancer	Normal	Σ
Predicted class	Cancer	14	5	19
	Normal	1	20	21
	Σ	15	25	40
		Specificity = $\frac{TN}{TN+FP}$ = 80.0%⊠($\frac{20}{25}$)		
		Sensitivity = $\frac{TP}{TP+FN}$ = 93.3%⊠($\frac{14}{15}$)		
		Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$ = 85.0%⊠($\frac{34}{40}$)		

from the biomarker Ct values to generate ΔCt. We next followed a reported method[21] to calculate the relative gene expression in various saliva samples: Relative Gene Expression $= 2^{(-\Delta Ct)} \times K$, where K was chosen as 1000 in our study.

2.5. Statistical analysis

CEA level and expression levels of candidate genes were compared using the multiple t test. We considered $P < .05$ as statistically significant. The receiver operating characteristic (ROC) curve and the area under the curve (AUC) value were calculated, using MedCalc software (MedCalc Software, Ostend, Belgium) and MATLAB software (The MathWorks, Inc., Natick, MA).

2.6. Machine learning for sample classification

We selected decision tree algorithm as our machine learning classifier. We used the "fitctree" and "predict" functions in Statistics and Machine Learning Toolbox as implemented in MATLAB software to classify sample classes in the discovery phase. Leave-one-out cross-validation (LOOCV) was performed to evaluate model performance. The sensitivity, specificity, and accuracy of the discovery phase was calculated using one machine-learning model based on the whole training dataset. After training the classifier, we then applied it to the datasets in the independent clinical validation phase, and used RNA and CEA levels as the features for the classifier to predict the

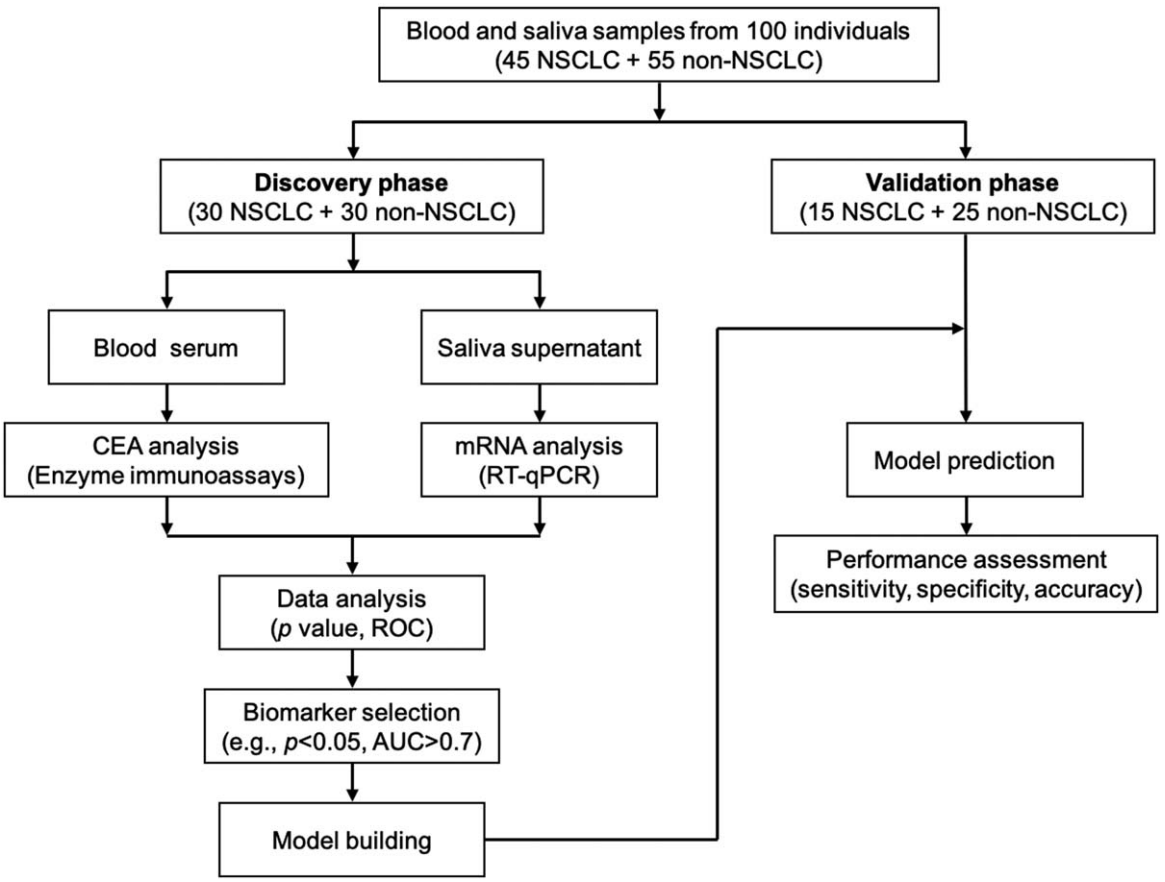

Figure 1. Comparison of CEA levels and RNA expression levels of 5 candiate genes (CCNI, EGFR, FGF19, FRS2, GREB1) between healthy control group (green) and NSCLC patient group (yellow) in the discovery phase (A, B) and validation phase (C, D). In the discovery phase, the healthy control group included 30 normal people; the NSCLC patient group included 30 patients. In the validation phase, the healthy control group included 25 normal people; the NSCLC patient group included 15 patients. * indicates $P < .05$; ** indicates $P < .01$; *** indicates $P < .001$. CEA = carcinoembryonic antigen, NSCLC = non-small cell lung cancer.

occurrence of NSCLC. The predicted classes of the 15 NSCLC patients and 25 healthy controls were listed in Table 3.

3. Results

3.1. Study design

This study contained a discovery phase and a subsequent clinical validation phase (Fig. 1). In the discovery phase, we built a machine-learning model that uses the CEA level and gene expression levels as input to predict the occurrence of NSCLC. In the validation phase, we applied our trained model to a different set of subjects, and used their CEA and gene expression levels to predict if a sample was from a NSCLC patient. We then compared our predictions with pathological classification to evaluate the clinical performance of our method (sensitivity, specificity, and accuracy).

3.2. Analysis of CEA levels and gene expression levels

First, the CEA levels of 30 NSCLC patients and 30 healthy controls were measured in the blood samples in the discovery phase. The mean CEA level was 1.408 ng/mL for healthy controls and 9.326 ng/mL for NSCLC patients (Fig. 2A). The difference of the CEA level between healthy controls and NSCLC patients was not significantly different ($P = .08$), indicating that CEA alone cannot discriminate NSCLC patients from healthy controls.

Next, we measured 5 mRNAs biomarkers CCNI, EGFR, FGF19, FRS2, and GREB1, in saliva samples (Fig. 2B). Four biomarkers, namely EGFR, FGF19, FRS2, and GREB1, were significantly up-regulated in NSCLC patients ($P < .05$). Specifically, FRS2 and GREB1 demonstrated 3.13-fold and 2.06-fold elevated expression in NSCLS patients ($P < .001$). No significant difference was observed for CCNI expression between NSCLC patients and healthy controls ($P = .79$).

3.3. Machine-learning model building and evaluation

We then used a machine-learning model to predict NSCLC occurrence from clinical characteristics. We plotted receiver operating characteristic (ROC) curve for each candidate biomarker and calculated the area under curve (AUC) value (Fig. 3). In general, an AUC value that is >0.70 suggests a decent performance for segregating clinical positives from negatives. GREB1 and FRS2 were ranked as the top 2 biomarkers, with AUC values reaching 0.84 and 0.79, respectively. The AUC value of CEA as well as the rest biomarkers in saliva was below 0.70, indicating unsatisfactory clinical performance when using these biomarkers alone for discriminating NSCLC patients and healthy controls.

Next, we chose 2 biomarkers, CEA and GREB1 as our design parameters. We included CEA because it was commonly use in clinical applications, in spite of its low AUC value (0.62). We used the pairwise data of the design parameters and the corresponding NSCLC occurrence to train a decision-tree algorithm. The model was constructed by following a toolkit developed in MATLAB (i.e., "fitctree" and "predict" in Statistics and Machine Learning Toolbox), which automatically adjusted the nodes and connections of the decision tree to optimize the fitting.[22] We found that our model could indeed discriminate NSCLC patients and healthy controls. The sensitivity and specificity for the combination of CEA and GREB1 was 96.67% and 83.33% for NSCLC prediction (Fig. 4).

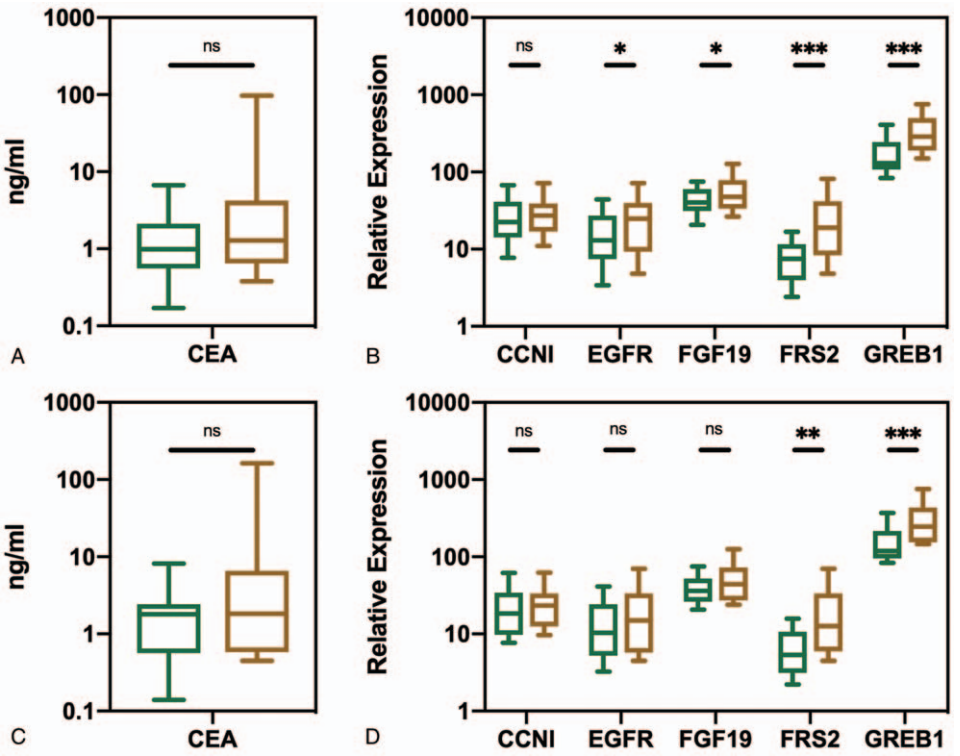

Figure 2. Schematic diagram of the study design and experimental procedures used in this study.

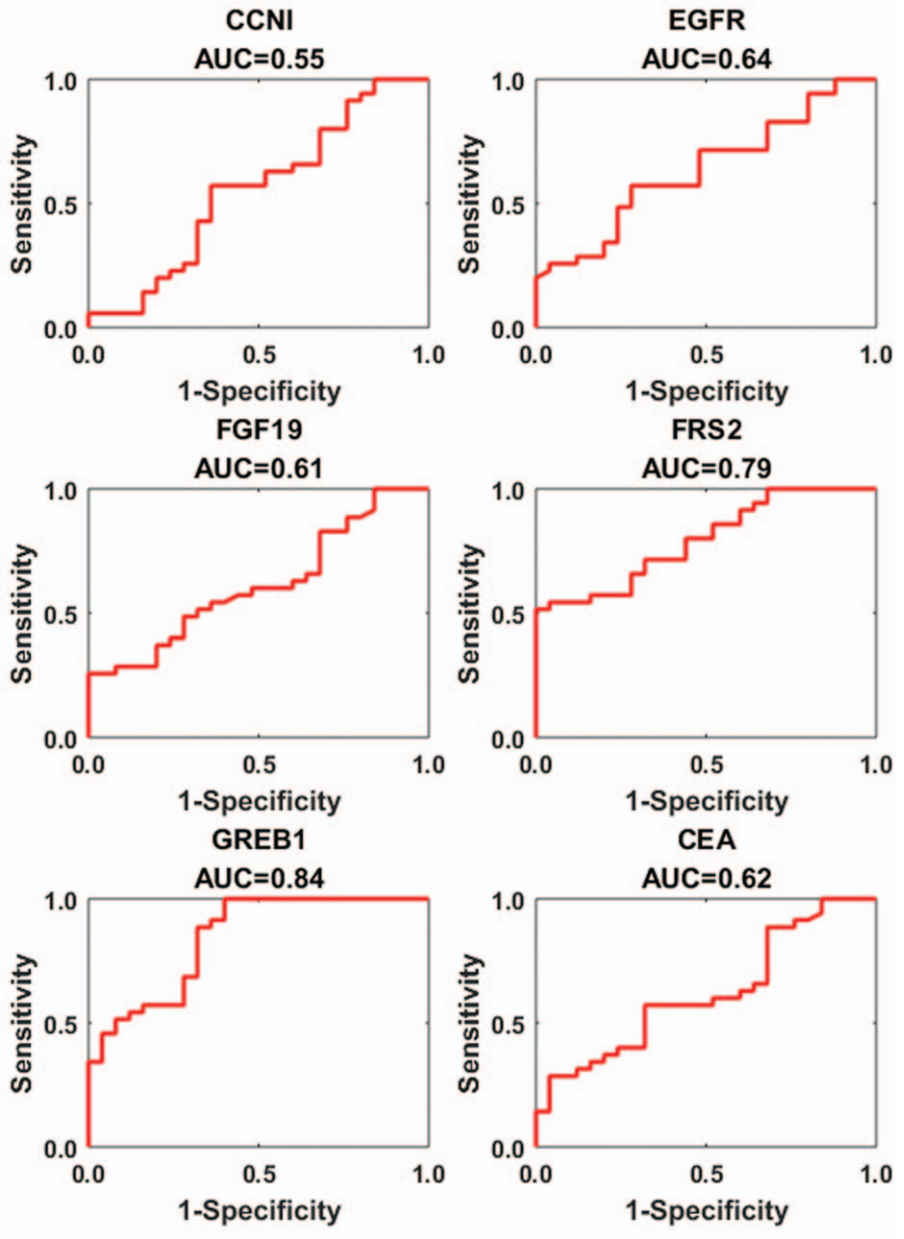

Figure 3. ROC curves of single biomarkers (CCNI, EGFR, FGF19, FRS2, GREB1) in saliva and CEA level in blood in the discovery phase. ROC = receiver operating characteristic curve.

Finally, we added FRS2 in the machine-learning model to see if the prediction performance could be further improved (Fig. 4). We determined the mean accuracies by 10 iterations of LOOCV for each case. We found that by including FRS2, the specificity could be further improved to 93.33%. Considering the high sensitivity (96.67%) and high specificity (93.33%) of this 3-parameter model, we decided to use these design parameters, that is, CEA level in blood, and GREB1 and FRS2 expression levels in saliva, in the machine-learning model for the subsequent clinical validation phase.

3.4. Validation with independent datasets

To find out if our machine-learning algorithm could be generally applied for NSCLC detection, we performed a validation study in an independent dataset. In this independent validation phase, we performed a blinded study of 15 NSCLC patients and 25 healthy controls (Table 1). The blood and saliva samples were collected from a total of 40 subjects, blinded, and analyzed for CEA levels (Fig. 2C) and gene expression levels (Fig. 2D). For both patients and healthy controls, CEA levels showed no significant difference ($P > .05$) between the discovery phase and the validation phase. Similarly, the expression levels of CCNI, EGFR, FGF19, FRS2, and GREB1 in the validation phase were not significantly different ($P > .05$) from those in the discovery phase. In sum, the difference between discovery phase and validation phase was not significant in our study.

We next used the machine-learning model that was trained in the discovery phase to predict if a sample from the validation phase was from a NSCLC patient. In total, we successfully

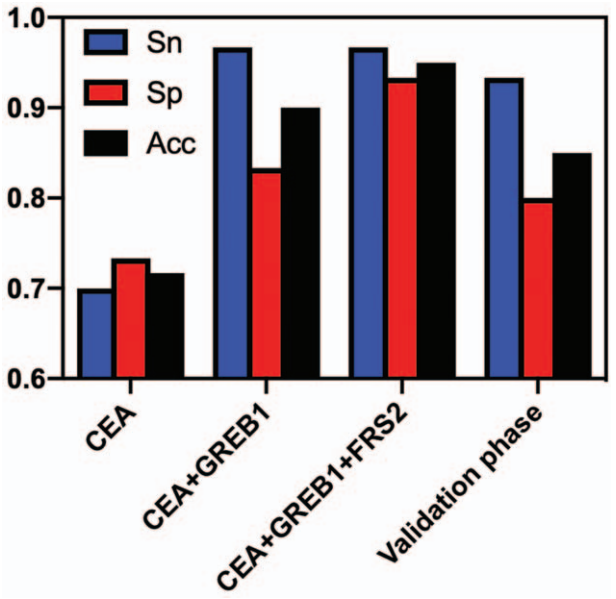

Figure 4. Performance of different panels of biomarkers in the discovery phase and validation phase. Blue color indicates sensitivity (or Sn), red color indicates specificity (or Sp), and black color indicates accuracy (or Acc). The sensitivity, specificity and accuracy of the discovery phase was calculated using one machine-learning model based on the whole training dataset, and the model performance was evaluated by leave-one-out cross-validation (LOOCV).

predicted 20 out of 25 healthy controls, and 14 out of 15 NSCLC patients (Table 3). Our predictions highly matched with pathological classification, with sensitivity and specificity reaching 93.33% and 80.00% respectively (Table 3). Overall, our study discovered that a novel panel of CEA and mRNA could be applied in clinical diagnosis of NSCLC as a non-invasive method.

4. Discussion

We discovered that the combined analysis of gene expressions in saliva and CEA in blood could be complementary to the current use of single type of biomarkers to diagnose NSCLC. By combining RNA and CEA analysis, we can extract more unique bio-signature of cancers from higher dimensions and thus increase both sensitivity and specificity. Combined analysis of multiple analytes for non-invasive detection of cancers is gradually becoming a consensus among researchers. For example, Cohen et al[23] recently developed CancerSEEK, a blood test that analyzes ctDNA mutations and various protein biomarkers to detect multiple cancers such as ovarian, liver, and lung cancer. Their method achieved over 95% sensitivity in detection of ovary and liver cancers. However, when using CancerSEEK for lung cancer detection, the sensitivity was <60%, while our combined analysis of CEA and RNAs led to a sensitivity of 93.33%. Future studies using larger and prospective cohorts will help define the role of this approach in clinical practice.

The key of our method is using a machine-learning modeling approach. Machine learning is advantageous than linear model in solving complex problems in 2 aspects: requiring little a priori knowledge of the system, and capable of resolving complex systems with high non-linearity and multi-dimensionality. Actually, the predictive power of machine learning is now being recognized by biologists, as we have witnessed multiple applications of using machine-learning models to analyze biomarkers from ctDNA,[24] ctRNA,[25] proteomics,[26,27] and metabolomics.[28] Most importantly, in addition of the use of machine-learning, we implemented an independent validation study to evaluate the practical use of our model. Such design is crucial as machine-learning models could face over-fitting problem, that is, a model could over-fit training datasets but suffers low accuracy in other datasets. In this study, our model achieves equally good predictions between the training datasets and the validation datasets, suggesting no over-fitting during model development. It is also worth noticing that the feature selection process used in this study is not entirely data-driven. For example, we included CEA level in our model although its AUC value was low and might not be used as a feature based on machine learning analysis. The reason we included CEA level is that it is commonly used in blood test and has great value in real-world clinical applications. The integration of domain knowledge with machine learning model is important for converting analytical data into clinical insights, which represents an applicable and non-invasive method for lung cancer detection in this study.

Several limitations of our study should be acknowledged. First, the sample size of this study (60 subjects in the discovery phase and 40 subjects in the validation phase) is relatively small compared with that required for implementation in the clinical. To plan for clinical trials, we would first need to estimate the correct sample size with sound scientific justification.[29] Second, our database included NSCLC patients and healthy controls, but left out subjects with benign lung diseases, such as pneumonia or tuberculosis. To achieve generalization of our findings, we will need to improve our model by including more types of patients. Third, the patient cohort in our study was composed of subjects from China only. To discover the biomarkers that are universal for people of all ethnicities, we might need to collect samples from other races (e.g., Caucasian) or include data from previous work that are based on other races. Fourth, we derived our biomarker panel from the biomarkers suggested by Zhang et al[11] (GREB1 and FRS2) to detect NSCLC in China. Both biomarkers were discovered from a database that is primarily Caucasian people, rather than Chinese people. As a result, some region-specific biomarkers that are particularly good for Chinese people might be missing. For example, KRAS, NOTCH1, STRN, and TP53 are potential candidates, because they are classical driver genes for lung cancer.[30–33] We are currently conducting a large-scale clinical test to confirm their discriminatory power in the Chinese population. Finally, smoking, age, and sex were not considered in our model for NSCLC prediction. In particular, smoking is a critical risk factor for lung cancer, and is involved in almost all of the lung cancer cases.[3–5] A more comprehensive clinical test is being conducted to evaluate the role of these factors to test whether or not these factors could further improve the prediction performance of our model.

Our study proves that combining CEA and salivary transcriptomic biomarkers can achieve non-invasive detection of NSCLC in a Chinese population. In future, we plan to endeavor our efforts into detection of multiple cancers, and also evaluate the clinical utility of our study in a large population. We expect our method, together with other liquid biopsies, will eventually provide early, painless, and affordable test for all incident cancer types.

Acknowledgments

The authors thank Rizhao City Traditional Chinese Medicine Hospital for providing funding support.

Author contributions

Conceptualization: Xi Shang, Youmin Zhang.
Data curation: Xi Shang, Huanpu Zi, Yangyang Li, Yu Gao, Changsheng Ge, Zhongmei Sun.
Formal analysis: Xi Shang, Huanpu Zi.
Investigation: Yangyang Li, Yu Gao.
Methodology: Changsheng Ge, Zhongmei Sun.
Supervision: Youmin Zhang.
Writing – original draft: Xi Shang, Huanpu Zi, Youmin Zhang.
Writing – review & editing: Yangyang Li, Yu Gao, Changsheng Ge, Zhongmei Sun.

References

[1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424.
[2] Hong QY, Wu GM, Qian GS, et al. Prevention and management of lung cancer in China. Cancer 2015;121(suppl):3080–8.
[3] Hecht SS. Cigarette smoking and lung cancer: chemical mechanisms and approaches to prevention. Lancet Oncol 2002;3:461–9.
[4] Spira A, Beane J, Shah V, et al. Effects of cigarette smoke on the human airway epithelial cell transcriptome. Proc Natl Acad Sci USA 2004;101:10143–8.
[5] Spira A, Beane J, Shah V, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. Nat Med 2007;13:361–6.
[6] Jha P. Avoidable global cancer deaths and total deaths from smoking. Nat Rev Cancer 2009;9:655–64.
[7] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018;68:7–30.
[8] Henry NL, Hayes DF. Cancer biomarkers. Mol Oncol 2012;6:140–6.
[9] Wielders JP, Bartels CT, Bank CM, et al. The diagnostic value of neuron-specific enolase and carcino-embryonic antigen analyses in patients with carcinoma of the lung. J Clin Chem Clin Biochem 1990;28:225–31.
[10] Litvak A, Cercek A, Segal N, et al. False-positive elevations of carcinoembryonic antigen in patients with a history of resected colorectal cancer. J Natl Compr Cancer Netw 2014;12:907–13.
[11] Zhang L, Xiao H, Zhou H, et al. Development of transcriptomic biomarker signature in human saliva to detect lung cancer. Cell Mol Life Sci 2012;69:3341–50.
[12] Park NJ, Zhou H, Elashoff D, et al. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. Clin Cancer Res 2009;15:5473–7.
[13] Li Y, St John MAR, Zhou X, et al. Salivary transcriptome diagnostics for oral cancer detection. Clin Cancer Res 2004;10:8442–50.
[14] Gleber-Netto FO, Yakob M, Li F, et al. Salivary biomarkers for detection of oral squamous cell carcinoma in a Taiwanese population. Clin Cancer Res 2016;22:3340–7.
[15] Brinkmann O, Kastratovic DA, Dimitrijevic MV, et al. Oral squamous cell carcinoma detection by salivary biomarkers in a Serbian population. Oral Oncol 2011;47:51–5.
[16] Lee YH, Kim JH, Zhou H, et al. Salivary transcriptomic biomarkers for detection of ovarian cancer: for serous papillary adenocarcinoma. J Mol Med (Berl) 2012;90:427–34.
[17] Zhang L, Xiao H, Karlan S, et al. Discovery and preclinical validation of salivary transcriptomic and proteomic biomarkers for the non-invasive detection of breast cancer. PLoS One 2010;5:e15573.
[18] Li F, Yoshizawa JM, Kim KM, et al. Discovery and validation of salivary extracellular RNA biomarkers for noninvasive detection of gastric cancer. Clin Chem 2018;64:1513–21.
[19] Zhang L, Farrell JJ, Zhou H, et al. Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. Gastroenterology 2010;138:949.e7–57.e7.
[20] St John MAR, Li Y, Zhou X, et al. Interleukin 6 and Interleukin 8 as potential biomarkers for oral cavity and oropharyngeal squamous cell carcinoma. Arch Otolaryngol Head Neck Surg 2004;130:929–35.
[21] Ishiba T, Hoffmann AC, Usher J, et al. Frequencies and expression levels of programmed death ligand 1 (PD-L1) in circulating tumor RNA (ctRNA) in various cancer types. Biochem Biophys Res Commun 2018;500:621–5.
[22] Documentation, M. MATLAB R2014b fitctree; 2015. Available at: http://es/.mathworks.com/help/stats/fitctree.html. Accessed February 11, 2019.
[23] Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science 2018;359:926–30.
[24] Chen X, Wang X, He H, et al. Combination of circulating tumor cells with serum carcinoembryonic antigen enhances clinical prediction of non-small cell lung cancer. PLoS One 2015;10:e0126276.
[25] Danaee P, Ghaeini R, Hendrix DA. A deep learning approach for cancer detection and relevant gene identification. pacific symposium on biocomputing. Pac Symp Biocomput 2016;22:219–29.
[26] Kourou K, Exarchos TP, Exarchos KP, et al. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8–17.
[27] Saghapour E, Kermani S, Sehhati M. A novel feature ranking method for prediction of cancer stages using proteomics data. PLoS One 2017;12: e0184203.
[28] Alakwaa FM, Chaudhary K, Garmire LX. Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data. J Proteome Res 2018;17:337–47.
[29] Sakpal TV. Sample size estimation in clinical trial. Perspect Clin Res 2010;1:67–9.
[30] Tomasini P, Walia P, Labbe C, et al. Targeting the KRAS pathway in non-small cell lung cancer. Oncologist 2016;21:1450–60.
[31] Motooka Y, Fujino K, Sato Y, et al. Pathobiology of Notch2 in lung cancer. Pathology 2017;49:486–93.
[32] Nakanishi Y, Masuda S, Iida Y, et al. Case report of non-small cell lung cancer with STRN-ALK translocation: a nonresponder to alectinib. J Thorac Oncol 2017;12:e202–4.
[33] Mogi A, Kuwano H. TP53 mutations in nonsmall cell lung cancer. J Biomed Biotechnol 2011;2011:583929.