EBioMedicine xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Urine Proteome Profiling Predicts Lung Cancer from Control Cases and Other Tumors

Chunchao Zhang [c,1], Wenchuan Leng [a,b,1], Changqing Sun [b], Tianyuan Lu [a], Zhengang Chen [b], Xuebo Men [b], Yi Wang [a,b,c], Guangshun Wang [b,*], Bei Zhen [a,b,*], Jun Qin [a,b,c,*]

[a] State Key Laboratory of Proteomics, National Center for Protein Sciences (The PHOENIX Center, Beijing), Beijing Proteome Research Center, Beijing 102206, China
[b] Joint Center for Translational Medicine, Tianjin, Baodi Hospital, Tianjin 301800, China
[c] Alkek Center for Molecular Discovery, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA

ARTICLE INFO

Article history:
Received 9 January 2018
Received in revised form 1 March 2018
Accepted 9 March 2018
Available online xxxx

Keywords:
Lung cancer
Machine learning
Urinary biomarkers

ABSTRACT

Development of noninvasive, reliable biomarkers for lung cancer diagnosis has many clinical benefits knowing that most of lung cancer patients are diagnosed at the late stage. For this purpose, we conducted proteomic analyses of 231 human urine samples in healthy individuals (n = 33), benign pulmonary diseases (n = 40), lung cancer (n = 33), bladder cancer (n = 17), cervical cancer (n = 25), colorectal cancer (n = 22), esophageal cancer (n = 14), and gastric cancer (n = 47) patients collected from multiple medical centers. By random forest modeling, we nominated a list of urine proteins that could separate lung cancers from other cases. With a feature selection algorithm, we selected a panel of five urinary biomarkers (FTL: Ferritin light chain; MAPK1IP1L: Mitogen-Activated Protein Kinase 1 Interacting Protein 1 Like; FGB: Fibrinogen Beta Chain; RAB33B: RAB33B, Member RAS Oncogene Family; RAB15: RAB15, Member RAS Oncogene Family) and established a combinatorial model that can correctly classify the majority of lung cancer cases both in the training set (n = 46) and the test sets (n = 14–47 per set) with an AUC ranging from 0.8747 to 0.9853. A combination of five urinary biomarkers not only discriminates lung cancer patients from control groups but also differentiates lung cancer from other common tumors. The biomarker panel and the predictive model, when validated by more samples in a multi-center setting, may be used as an auxiliary diagnostic tool along with imaging technology for lung cancer detection.

1. Introduction

Lung cancer is the second most common cancer among males and females worldwide and the most common cancer in China (Torre et al., 2016b, Torre et al., 2016a). It is the leading cause of cancer death in both men and women in the United States (Torre et al., 2016a). In 2012, there were approximately 1.8 million new cases and 1.6 million cancer deaths documented, which highlight a global public health concern (Stewart et al., 2014). Non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) are the two main histologic subtypes of lung cancer with the NSCLC as the most common subtype, accounting for about 83% of all lung cancers (Miller et al., 2016).

Computed tomography (CT) screening is the main test for lung cancer screening but is associated with a high false positive rate (Aberle et al., 2013). Disease stage significantly affects cancer treatment and survivorship. The 5-year survival rate is 55% for patients diagnosed at the early stage and 4% at the advanced stage (Miller et al., 2016). Unfortunately, majority of cases are diagnosed at the advanced stage due to the lack of symptoms and reliable biomarkers at the early stage (Miller et al., 2016).

Searching noninvasive biomarkers for clinical diagnosis is a continuous effort but success has been limited (Zhang and Chan, 2005). Current clinically used tumor markers for lung cancer screening including AFP (alpha fetoprotein), CA 19-9 (carbohydrate antigen 19-9), CA 125 (carcinoma antigen 125), CA 15-3 (carcinoma antigen 15-3), and CEA (carcino-embryonic antigen) lack sensitivity and specificity (Li et al., 2012, Harmsma et al., 2013). Some earlier proteomic studies towards lung cancer diagnosis based on urine or serum specimens have identified a few putative biomarkers, but the specificity against other tumors is poor or has not been investigated (Zhang et al., 2015, Nolen et al., 2015, Patz et al., 2007, Yildiz et al., 2007). In this study, we employed proteomics technology implemented with machine learning statistics to search for sensitive, lung cancer-specific diagnostic biomarkers from patient urines as a commonly used, noninvasive matrix as an alternative to blood.

* Corresponding authors at: Joint Center for Translational Medicine, Tianjin Baodi Hospital, Tianjin 301800, China.
E-mail addresses: wgsTMUBH@163.com (G. Wang), zp1963@sina.com (B. Zhen), jqin1965@126.com (J. Qin).
[1] Equal contributing authors.

https://doi.org/10.1016/j.ebiom.2018.03.009

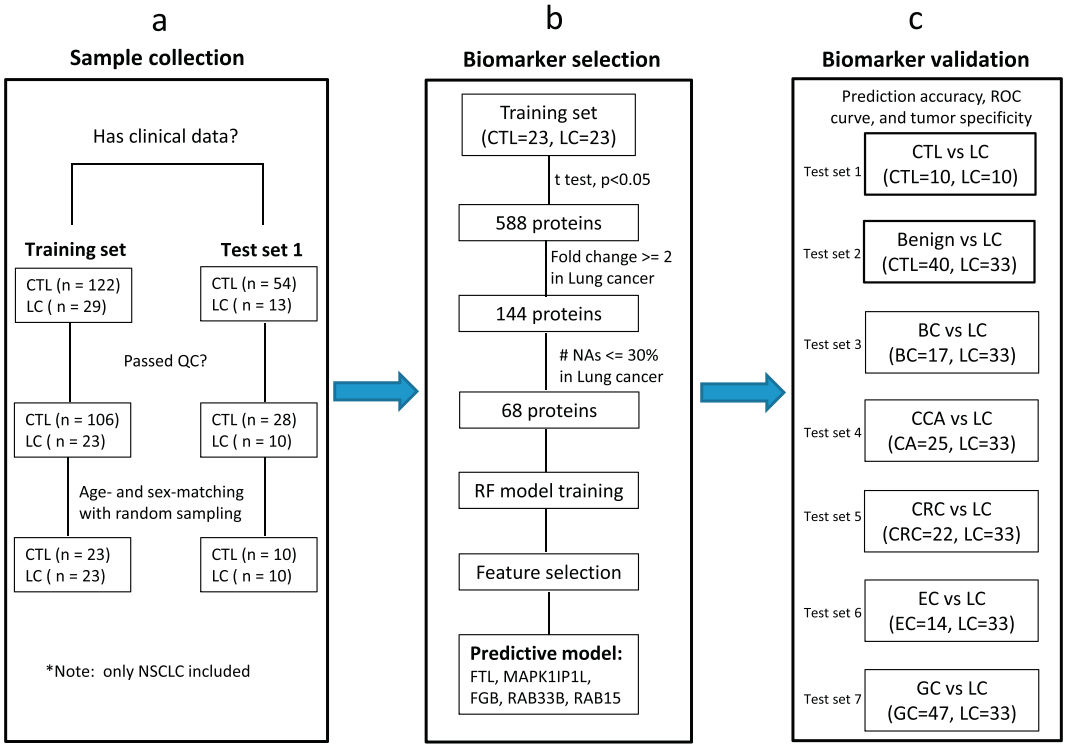

Fig. 1. Flow diagram of lung cancer biomarker study. (a) A total of 218 urine specimens were randomly collected from healthy donors or NSCLC patients. After QC filtering and age/sex-matching, a pair of 23 or 10 case-control urine samples was selected in the training set or test set (test set 1), one for biomarker discovery and the other one for biomarker validation, respectively. (b) Student's *t*-test revealed a total of 588 proteins with a p value <.05 in the training set, 144 were up-regulated with at least 2 folds in the cancer group. Finally, 68 proteins were retained by restricting the number of missing values in <30% of lung cancer cases. A random forest model was developed upon the training set with 68 proteins. By running feature selection algorithm, five biomarkers were selected and incorporated into a predictive model. (c) The biomarker panel and the predictive model were evaluated on 7 independent test sets to determine how well the model can predict lung cancer from healthy individuals and benign lung diseases (test set 1–2) or from other cancers (test set 3–7). Abbreviations: CTL, healthy controls; LC, lung cancer; BC, bladder cancer; CCA, cervical cancer; CRC, colorectal cancer; EC, esophageal cancer; GC, gastric cancer; NSCLC, non-small-cell lung cancer; QC, quality control.

2. Materials and Methods

2.1. Patient Specimens

At the biomarker discovery stage, a total of 46 urine specimens in the training set from healthy controls (CTL, n = 23) and lung cancer patients (LC, n = 23) were collected at Tianjin Baodi Hospital, Tianjin, China. Healthy controls were age- (>50 year) and gender-matched (frequency matching with random sampling) to lung cancer cases (Fig. 1a). Urine samples were collected from Non-small cell lung cancer (NSCLC) patients at the time they were diagnosed with lung cancer and had no anticancer treatment. Urine samples were collected from healthy donors who had no known lung diseases and had negative clinical tumor markers (AFP: alpha fetoprotein, CA 19-9: carbohydrate antigen 19-9, CA 125: carcinoma antigen 125, CA 15-3: carcinoma antigen 15-3, and CEA: carcino-embryonic antigen). A blood test monitored the levels of urea nitrogen, creatinine, and uric acid to exclude any cases that may have renal dysfunction. For validation purposes, an independent case-control test set (10 CTL, 10 LC; Fig. 1a, test set 1) with same criteria was obtained from the same Hospital. In addition to healthy donors, urines from benign pulmonary conditions (pneumonia, n = 23; COPD: Chronic Obstructive Pulmonary Disease,

Table 1
Clinical profiles and demographics of healthy controls and lung cancer patients.

Demographics	Training set		Test set		Benign lung diseases	
	CTL (n = 23)	LC (n = 23)	CTL (n = 10)	LC (n = 10)	COPD (n = 17)	Pneumonia (n = 23)
Age, years	55.61 ± 8.02	65.65 ± 11.2	55.8 ± 3.49	65.7 ± 8.96	73.88 ± 10.07	60.39 ± 22
Sex						
Male	16	16	7	7	13	16
Female	7	7	3	3	4	7
Clinical stage						
1		1		2		
2		4		1		
3		10		3		
4		8		4		
Subtype						
ADC		10		2		
SCC		13		8		

ADC, adenocarcinoma; SCC, squamous cell carcinoma; CTL, healthy controls; LC, lung cancer; COPD, Chronic Obstructive Pulmonary Disease.

Please cite this article as: Zhang, C., et al., Urine Proteome Profiling Predicts Lung Cancer from Control Cases and Other Tumors, EBioMedicine (2018), https://doi.org/10.1016/j.ebiom.2018.03.009