

Development of transcriptomic biomarker signature in human saliva to detect lung cancer

Lei Zhang · Hua Xiao · Hui Zhou · Silverio Santiago · Jay M. Lee · Edward B. Garon · Jieping Yang · Ole Brinkmann · Xinmin Yan · David Akin · David Chia · David Elashoff · No-Hee Park · David T. W. Wong

Received: 19 January 2012/Revised: 23 April 2012/Accepted: 7 May 2012
© Springer Basel AG 2012

Abstract Lung cancer is the leading cause of cancer death for both men and women worldwide. Since most of the symptoms found for lung cancer are nonspecific, diagnosis is mostly done at late and progressed stage with the consecutive poor therapy outcome. Effective early detection techniques are sorely needed. The emerging field of salivary diagnostics could provide scientifically credible,

easy-to-use, non-invasive and cost-effective detection methods. Recent advances have allowed us to develop discriminatory salivary biomarkers for a variety of diseases from oral to systematic diseases. In this study, salivary transcriptomes of lung cancer patients were profiled and led to the discovery and pre-validation of seven highly discriminatory transcriptomic salivary biomarkers (BRAF, CCNI, EGFR, FGF19, FRS2, GREB1, and LZTS1). The logistic regression model combining five of the mRNA biomarkers (CCNI, EGFR, FGF19, FRS2, and GREB1) could differentiate lung cancer patients from normal control subjects, yielding AUC value of 0.925 with 93.75 %

L. Zhang and H. Xiao contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00018-012-1027-0) contains supplementary material, which is available to authorized users.

L. Zhang · H. Xiao (✉) · H. Zhou · J. Yang · O. Brinkmann · X. Yan · D. Akin · N.-H. Park · D. T. W. Wong
School of Dentistry and Dental Research Institute,
University of California, 10833 Le Conte Avenue,
73-024 CHS, Los Angeles, CA 90095, USA
e-mail: xiaohuacn@gmail.com

D. T. W. Wong
e-mail: dtww@ucla.edu

Present Address:

H. Zhou
Genome Sciences Resource, Vanderbilt University Medical
Center, Nashville, TN, USA

D. Chia
Department of Pathology and Laboratory Medicine, David
Geffen School of Medicine, University of California-Los
Angeles, Los Angeles, CA 90095, USA

J. M. Lee · D. Chia · N.-H. Park · D. T. W. Wong
Jonsson Comprehensive Cancer Center, David Geffen School of
Medicine, University of California-Los Angeles, Los Angeles,
CA 90095, USA

D. Elashoff
Department of Medicine, David Geffen School of Medicine,
University of California-Los Angeles, Los Angeles, CA 90095, USA

D. T. W. Wong
Division of Head and Neck Surgery/Otolaryngology, David
Geffen School of Medicine, University of California-Los
Angeles, Los Angeles, CA 90095, USA

D. T. W. Wong
Henry Samueli School of Engineering and Applied Science,
University of California-Los Angeles, Los Angeles, CA 90095,
USA

D. T. W. Wong
Dental Research Institute, School of Dentistry, University of
California-Los Angeles, 10833 Le Conte Avenue, 73-017 CHS,
Los Angeles, CA 90095, USA

S. Santiago · N.-H. Park
David Geffen School of Medicine, University of California-Los
Angeles, Los Angeles, CA 90095, USA

J. M. Lee
Department of Surgery, David Geffen School of Medicine,
University of California-Los Angeles, Los Angeles, CA 90095, USA

E. B. Garon
Department of Medicine, Division of Hematology/Oncology,
David Geffen School of Medicine, University of California-Los
Angeles, Los Angeles, CA 90095, USA

sensitivity and 82.81 % specificity in the pre-validation sample set. These salivary mRNA biomarkers possess the discriminatory power for the detection of lung cancer. This report provides the proof of concept of salivary biomarkers for the non-invasive detection of the systematic disease. These results poised the salivary biomarkers for the initiation of a multi-center validation in a definitive clinical context.

Keywords Human saliva · Transcriptomic · Lung cancer · Biomarker signature · Early detection

Abbreviations

CT	Computerized tomography
MACs	Malignancy-associated changes
ROC	Receiver-operating characteristic
AUC	Calculation of the area under the curve
qPCR	Quantitative polymerase chain reaction
COPD	Chronic obstructive pulmonary disease
EGFR	Epidermal growth factor receptor
BRAF	v-raf murine sarcoma viral oncogene homolog B1
LZTS1	Leucine zipper, putative tumor suppressor 1

Introduction

According to the National Cancer Institute, lung cancer is the leading cause of cancer death for both men and women worldwide. In 2010, 222,520 new lung cancer patients were diagnosed in the United States, and roughly 87 % were non-small-cell lung carcinomas [1]. Smoking, particularly of cigarettes, is by far the main contributor to lung cancer [2–4], which is responsible for 90 % of all lung cancers. In the United States, it is estimated that there are 45 million current smokers and 46 million ex-smokers who are at increased risk for developing lung carcinoma. Over 75 % of lung cancer cases are diagnosed in late stages because there remains no practical way to screen a large number of people at risk. The 5-year relative survival rates of lung and bronchus cancer in 2010 is only 16 % in the United States, slowly increased 3 % from that of 30 years ago [1]. The poor results in lung cancer treatment can be attributed to different factors and the high mortality rate for lung cancer, in part, is from a lack of effective tools to detect the disease at an early stage [3, 4]. Early detection offers the promise of improved cure rates. New strategies to identify high-risk individuals are sorely needed to control the disease.

The first attempts to identify early detection and screening tools with sputum cytology and chest X-ray failed to increase the number of curable cases [5].

Subsequently, computerized tomography (CT) was developed, with increased resolution for detecting small nodules. Broad application of CT screening to high-risk populations has several drawbacks [5, 6]. A recently released study shows that low-dose helical CT might be helpful for lung cancer screening [7]. Besides, flexible bronchoscopy is a relatively non-invasive initial diagnostic test [4]. The sensitivity of bronchoscopy for lung cancer ranges from 30 to 80 %. As a result, most patients require further invasive diagnostic tests, which delay treatment and generate additional costs and risks for complications.

In the last decade, biomarker discovery for lung cancer detection has focused on sputum [5, 8], bronchoalveolar lavage [9], lung condensate/expired air [10] and blood [11], or more invasively on tissue [12], using proteomic, as well as genomic approaches. Sputum might be an ideal source for lung cancer biomarker development, in that it has clinical relevance with the organ. However, the development of detection biomarkers in these sources is limited, mainly because of lack of sensitivity and specificity. Especially, sputum for automated analysis or biomarker analysis has not demonstrated feasibility, since 30 % of the high-risk population of ex-smokers has difficulty producing phlegm after multi-day collections [13]. There is a pressing need for new sources to harness discriminatory biomarkers for early detection of lung cancer, which should carry significant translational potentials.

As a mirror of the body, saliva is readily accessible and harbors diverse components, including protein [14], RNA [15], and exosomes [16], which may serve as biomarkers [17, 18]. There are many advantages to use saliva as a clinical diagnostic biofluid. Sample collection is simple, non-invasive, and causes less anxiety on the part of patients. The use of saliva also offers a cost-effective approach for large-scale screens [17]. Biomarker signatures in human saliva are telltale molecules that could be used to monitor the health status, disease onset, treatment responsiveness, and outcome. Informative biomarkers can further serve as early sentinels of disease, and this has been considered as a promising alternative to classic environmental epidemiology [17, 19]. In the past decade, a number of findings have stimulated interests in the use of saliva as a source of biomarkers [20–22]. Salivary proteomic and genomic biomarkers have been successfully developed by our team for the detection of oral squamous cell carcinoma [20, 21, 23], Sjögren's syndrome [24], pancreatic cancer [15], and breast cancer [25]. Most recently, our lab has successfully discovered and pre-validated three candidate proteomic biomarkers for the detection of lung cancer in human saliva [26].

In this study, human saliva samples were collected from 42 lung cancer patients and 74 healthy control subjects. Salivary transcriptomes were analyzed by gene microarray

and compared in the two groups. We hypothesized that discriminatory lung cancer transcriptomic biomarker signatures are present in human saliva, which could be harnessed and clinically used to discriminate lung cancer patients from cancer-free control subjects.

Patients and methods

Patients and the study design

Recently diagnosed and untreated lung cancer patients and matched cancer-free controls were recruited from the UCLA Medical Center under an approved IRB (UCLA IRB#10-000505). Written informed consent forms and questionnaire data sheets were obtained from all participants. Controls were matched for gender, age, and ethnicity with the cancer group. Their smoking history was matched generally by whether they are non-smokers or smokers (current smokers and ex-smokers). Any participants with a diagnosis of any malignancy within the last 5 years were excluded from the study. Patient demographics and clinical profiles are presented in Table 1.

Unstimulated whole saliva was collected and processed according to previously established protocol [21, 25]. Briefly, saliva samples were kept on ice during collection and were then centrifuged at $2,600 \times g$ for 15 min 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.

This study consisted of two phases, a discovery phase, followed by a pre-validation phase (Fig. 1). Of these collected samples, ten lung cancer samples and ten matched control samples were randomly selected for the biomarker discovery. The transcriptomic approach profiled the saliva supernatant samples from the discovery sample set using

the Affymetrix HG U133 Plus 2.0 Array (Affymetrix, Santa Clara, CA). Biomarker signatures identified from the microarray study were first verified in the discovery sample set ($n = 20$). A pre-validation sample set, including 32 lung cancer samples and 64 matched control samples, was used for further biomarker pre-validation.

Salivary transcriptomic profiling and data analysis

For transcriptome profiling, 330 μ l of the saliva supernatant of each individual sample was used and processed according to the previously described protocol [25]. Briefly, MagMax Viral RNA isolation kit (Ambion, Austin, TX) was used for saliva RNA isolation by using KingFisher mL technology (Thermo Fisher Scientific, Waltham, MA). Following TURBO DNase treatment, extracted RNA was linearly amplified. After purification, complementary DNA (cDNA) was in vitro transcribed and biotinylated using GeneChip Expression 3'-Amplification Reagents for in vitro transcription labeling (Affymetrix, Santa Clara, CA). The Affymetrix Human Genome U133 Plus 2.0 Array, which represents $>54,000$ transcripts and variants, was used for the discovery on the ten randomly selected lung cancer samples and ten matched control samples. Salivary mRNA profiling was performed at the UCLA Microarray Core Facility. Using the minimum information about a microarray experiment (MI-AME) criteria [27], all Affymetrix Human Genome U133 Plus 2.0 Array data generated in this study has been uploaded to the GEO database (<http://www.ncbi.nlm.nih.gov/geo>). The access number is GSE32175.

U133 Plus 2.0 array data analysis

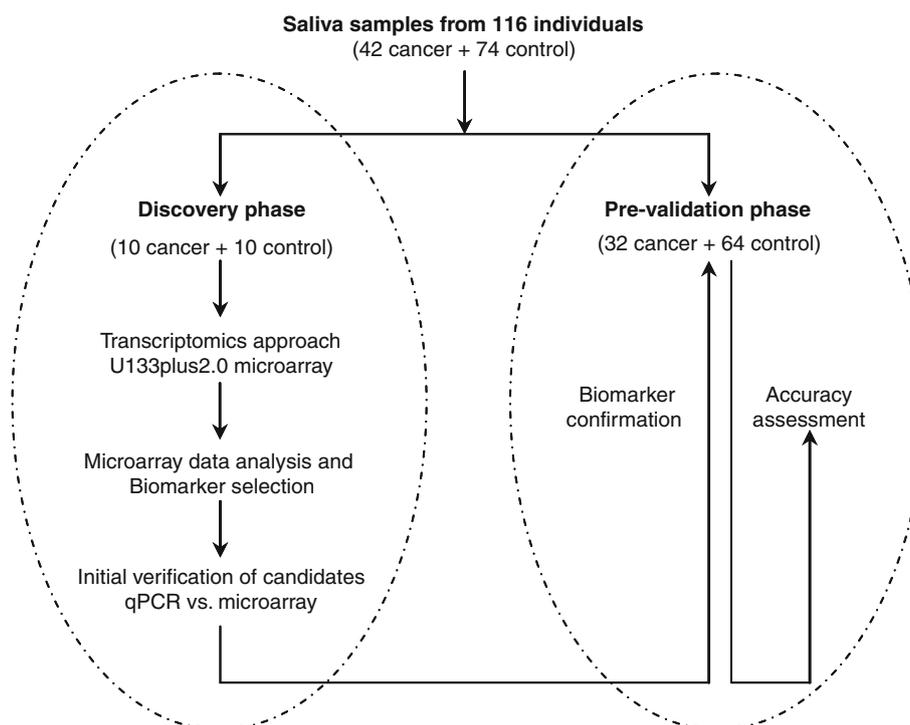
The microarray data analysis was performed using R 2.7.0 (<http://www.r-project.org>). The probe logarithmic intensity

Table 1 Patient demographics and clinical profiles

Demographic	Discovery set		Confirmation set	
	Cancer ($n = 10$)	Control ($n = 10$)	Cancer ($n = 32$)	Control ($n = 64$)
Age, years	59.2 ± 6.36	60.4 ± 4.72	64.28 ± 10.43	61.53 ± 8.11
Range	45–69	45–71	43–90	45–89
Sex				
Male	5	5	16	28
Female	5	5	16	36
Ethnicity (USA)				
Caucasian	8	8	25	51
Others	2	2	7	13
Smoke history				
Yes	6	7	24	40
Lung cancer				
NSCLC	9		28	
SCLC	1		4	

NSCLC Non-small-cell lung carcinoma, SCLC small cell lung carcinoma

Fig. 1 The study design for the two phases salivary mRNA biomarker development for lung cancer



error (PLIER) [28] estimation expression measures were computed after background correction and quantile normalization for each microarray data set. Probe set-level quantile normalization was performed across all samples to make the effect sizes similar in all data sets. Finally, for every probe set, the two-sample t test was applied to identify differential expression between lung cancer and healthy control. After obtaining the estimates and the p values of each probe set, we corrected the p values for false discovery rate.

Microarray candidate identification, logistic regression method, leave-one-out cross-validation, construction of receiver operating characteristic (ROC) curve, calculation of the area under the curve (AUC) and intergroup comparison of quantitative polymerase chain reaction (qPCR) results were performed as previously described [25].

mRNA biomarker verification and pre-validation using qPCR

The selected candidate mRNA biomarkers were first verified with qPCR in the discovery sample set as described previously [24, 29]. qPCR primers were designed using Primer Express 3.0 software (Applied Biosystems, Foster City, CA). Raw Ct values of biomarker candidates were normalized to Ct values of GAPDH. All primers were synthesized by Sigma-Genosys (Woodlands, TX). The amplicons were intron spanning whenever possible. qPCR was carried out in duplicate. Initial verified biomarker signatures were then assayed by qPCR in the 96 saliva

samples of pre-validation set (32 lung cancer cases and 64 matched controls). The Wilcoxon test was used to compare the biomarkers between groups. The gene accession numbers and primer sequences used for transcriptomic biomarker verification and pre-validation are shown in Supplementary Table S1.

Biomarker performance evaluation

As described previously [15], the performance of the individual biomarker and biomarker panel was evaluated by constructing the ROC curve and computed the AUC value by numerical integration of the ROC curve. The sensitivity and specificity for the biomarker combinations were estimated by identifying the cut-off point of the predicted probability that yielded the highest sum of sensitivity and specificity.

Cross-disease comparisons of salivary mRNA biomarkers based on microarray studies

The validated mRNA biomarker signatures for lung cancer detection were checked with other microarray studies that have been conducted in our laboratory on different diseases, including pancreatic cancer, ovarian cancer, breast cancer, type II diabetes, and chronic obstructive pulmonary disease (COPD). Briefly, t test-based p values were calculated for all validated genes of lung cancer study in other microarray studies to check whether they were also significantly varied between cancers and controls in those

diseases. Variation was considered significant if p value was <0.05 .

Results

Salivary transcriptomic biomarker discovery

In the discovery phase (Fig. 1), expression microarrays were used to examine gene expression profiles and levels in saliva samples from lung cancer patients ($n = 10$) and matched cancer-free controls ($n = 10$). The quantity and quality of extracted mRNA in saliva was measured for the 20 samples to ensure the sufficiency and accuracy for microarray profiling. On average, about 150 ng of total RNA was obtained from 330 μ l of saliva supernatant. There was no significant difference between the lung cancer group and the matched control group. GAPDH was consistently used for normalization in all qPCR experiments [29]. After two rounds of RNA amplification, the yield of biotinylated cRNA was about 35 μ g ($n = 20$) on average. There were no significant differences in the yield of cRNA between lung cancer patients and matched controls.

In total, transcriptomic analysis revealed that 2,212 genes exhibited a greater than twofold up-regulation, and 1,114 genes exhibited greater than twofold down-regulation in the saliva of lung cancer patients compared to the matched controls ($n = 20$, $p < 0.05$). Of these, 593 genes showed greater than fivefold up-regulation and 585 genes showed greater than fivefold down-regulation in the saliva of lung cancer patients ($p < 0.05$). These transcripts identified were unlikely to be attributed to chance (χ^2 test, $p < 0.0001$), considering the false-positive rate with $p < 0.05$. The top 99 up-regulated transcripts were selected as the first set of biomarker candidates. Many of them have been shown to be related to lung cancer or other cancers, including epidermal growth factor receptor (EGFR), v-raf murine sarcoma viral oncogene homolog B1 (BRAF) and leucine zipper, putative tumor suppressor 1 (LZTS1).

Identification of mRNA biomarkers for lung cancer

In order to verify the microarray data for further independent validation, the top 16 up-regulated transcripts out of 99 top candidates were selected for evaluation. qPCR was performed to verify the microarray results on the discovery sample set ($n = 20$). The results confirmed that the relative mRNA expression levels of 12 of the 16 up-regulated transcripts were consistent with the microarray data. All 12 biomarkers showed significant differences between lung cancer and the matched controls ($p < 0.05$): GREB1, LZTS1, FGF19, CCNI, EGFR, BRAF, FRS2, GFRA1, HDAC6, TNFRSF25, CTAGE1, and SYK.

qPCR pre-validation

The 12 verified candidates were subjected to an independent pre-validation using a new independent cohort of 32 lung cancer samples and 64 matched control samples. As shown in Table 2, seven of the 12 up-regulated genes were validated, including BRAF, CCNI, EGFR, FGF19, FRS2, GREB1, and LZTS1. These seven mRNA biomarkers all showed significant differences between lung cancer and the matched controls ($p < 0.05$, $n = 96$), yielding AUC values between 0.707 and 0.850 (Table 2). The expression patterns of these mRNA biomarker signatures were consistent with the microarray data (up-regulation and fold change). The heat map of these seven markers in 32 lung cancer samples and 64 control samples is shown in Fig. 2.

Prediction models using the pre-validated salivary mRNA biomarkers

To demonstrate the clinical utility of salivary mRNAs biomarker signatures for lung cancer detection, logistic regression models were built based on different combinations of these biomarkers (Table 3). The logistic regression model with the combination of five mRNA biomarkers (CCNI, EGFR, FGF19, FRS2, and GREB1) could differentiate lung cancer patients from control subjects, yielding an AUC value of 0.925 with 93.75 % sensitivity and 82.81 % specificity).

Table 2 Performance of seven individual pre-validated mRNA markers

mRNA marker	p value	AUC	Sensitivity	Specificity	Reported relation to cancer
BRAF	0.002	0.707	0.66	0.77	[38]
CCNI	<0.001	0.850	0.94	0.66	[51]
EGFR	<0.001	0.764	0.78	0.67	[37]
FGF19	<0.001	0.820	0.75	0.86	[45]
FRS2	<0.001	0.745	0.59	0.84	[46]
GREB1	<0.001	0.806	0.84	0.75	[52]
LZTS1	<0.001	0.755	0.69	0.81	[39]

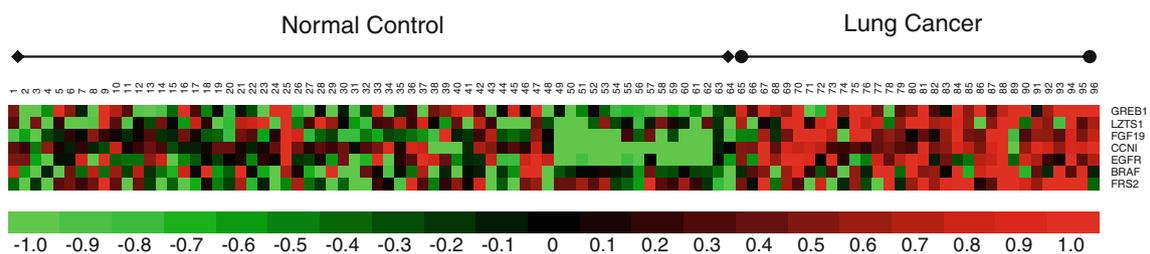


Fig. 2 Heat map of the seven pre-validated genes in the pre-validation sample set ($n = 96$). The pre-validation sample set including 32 lung cancer samples and 64 matched control samples

Table 3 Performance of different biomarker combination in the pre-validation sample set

mRNA biomarker combination	AUC	Sensitivity (%)	Specificity (%)	Sensitivity (%) + specificity (%)
CCNI	0.8501	93.75	65.63	159.38
CCNI + FGF19	0.8782	93.75	75.00	168.75
CCNI + FGF19 + GREB1	0.9146	93.75	78.13	171.88
CCNI + FGF19 + GREB1 + FRS2	0.9165	93.75	76.56	170.31
CCNI + FGF19 + GREB1 + FRS2 + EGFR	0.9253	93.75	82.81	176.56
CCNI + FGF19 + GREB1 + FRS2 + EGFR + LZTS1	0.9241	90.63	85.94	176.57
CCNI + FGF19 + GREB1 + FRS2 + EGFR + LZTS1 + BRAF	0.9224	90.63	85.94	176.57

The biomarker combination in bold is the best combination with the highest AUC value

Figure 3a showed the ROC curve for the combination of the five mRNA biomarkers, and Fig. 3b was the corresponding interactive dot diagram of these biomarkers in the 96 pre-validation samples.

Cross-disease comparisons of salivary mRNA biomarkers

It is possible that the different cancers may have overlapping salivary biomarker signatures, which make it very

important to determine the specific profiles of molecular changes in specific cancer types. We evaluated the specificity of the seven pre-validated salivary lung cancer mRNA biomarkers against other microarray discovery studies that have been performed in our laboratory on diverse diseases, including pancreatic cancer [15], ovarian cancer, breast cancer [25], type II diabetes, and COPD. As shown in Table 4, none of the seven mRNAs/transcripts were significantly altered in other cancer/disease microarray studies ($p > 0.05$). It is worth mentioning that most

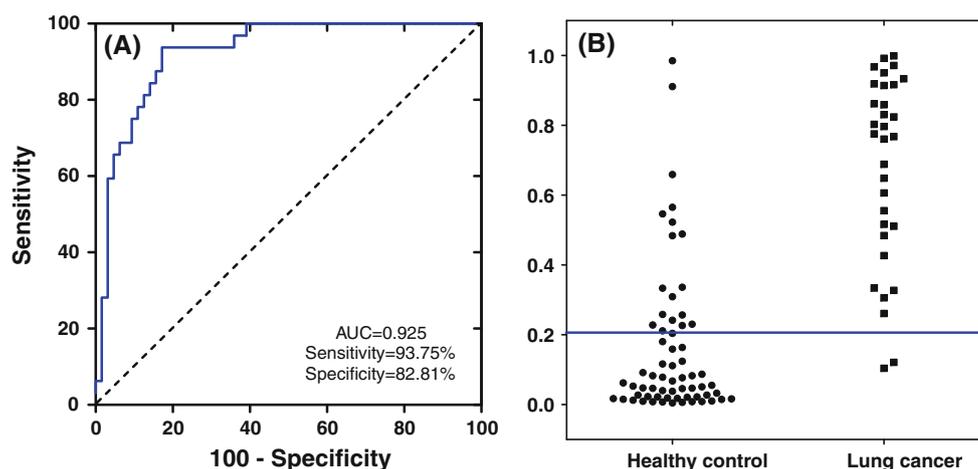


Fig. 3 ROC curve and interactive dot diagram for the logistic regression model. **a** Logistic regression model using five biomarkers (CCNI, FGF19, FRS2, GREB1, and EGFR) in the pre-validation sample set results in AUC

value of 0.925 with 93.75 % sensitivity and 82.81 % specificity (cut-off, 0.2598). **b** Interactive dot diagram based on the qPCR data of the lung cancer group ($n = 32$) and healthy control group ($n = 64$)

Table 4 Cross-disease comparison of microarray profiles of seven validated mRNA biomarkers

mRNA marker	Lung cancer	Pancreatic cancer	Ovarian cancer	Breast cancer	Type II diabetes	COPD
BRAF	0.002	0.684	0.826	0.965	0.987	0.090
CCNI	<0.001	0.925	0.058	0.404	0.095	0.548
EGFR	<0.001	0.832	0.257	0.664	0.855	0.559
FGF19	<0.001	0.071	0.130	0.425	0.341	0.646
FRS2	<0.001	0.343	0.657	0.264	0.108	0.849
GREB1	<0.001	0.850	0.612	0.321	0.334	0.822
LZTS1	<0.001	0.273	0.529	0.466	0.194	0.211

Cancer specificity of the seven validated mRNA biomarkers was evaluated across different microarray discovery studies performed in our laboratory on diverse diseases, including pancreatic cancer, ovarian cancer, breast cancer, type II diabetes, and COPD.

t test-based *p* values were calculated for each transcript between diseases and healthy controls in different microarray studies. All of the mRNAs/transcripts that showed significant variations in this lung cancer study were not significantly altered in other diseases microarray studies ($p > 0.05$)

lung cancer patients will have COPD because both conditions are mainly caused by smoking [30, 31]. Especially, many of the common symptoms of lung cancer are similar to those of COPD, which is very necessary to differentiate it from lung cancer. Our gene microarray data shown that the pre-validated mRNA biomarker could not differentiate COPD from normal control ($p > 0.05$). All these cross-disease comparisons indicated that the validated mRNA biomarkers in saliva were specific for lung cancer.

Influence of smoking

Among the 96 subjects in the pre-validation study, 64 of them were smokers and 32 of them were non-smokers. Twenty-four of the 32 lung cancer patients were smokers while eight were non-smokers. The seven biomarkers' performance were evaluated in non-smokers and smokers (Table 5). The significant difference between each group was labeled in bold. For cancer-free control subjects, smoking history did not change these biomarkers' saliva levels significantly, except for CCNI (row 1). The same was found for lung cancer patients in row 2. When compared to the cancer-free subjects, the expressions for most of these biomarkers in the saliva of lung cancer patients

were significantly different. Especially, for the patients with smoking history, all seven salivary markers were significantly altered (rows 3 and 6).

Discussion

Lung cancer is a highly prevalent disease with a very poor therapeutic outcome. This is mainly because most lung cancers are detected at late stages where the successful cures are unlikely and death rates are high. As in all cancers, early detection would be the key for improved survival rates.

Current blood-based tests for lung cancer, including CA 125, CA 19-9, CA15-3, CEA, SCC, CYFRA 21-1 and NSE, lack of sufficient sensitivity and specificity to be of use in screening of lung cancer [32, 33]. For example, the tumor marker sensitivity for CYFRA 21-1, CA 125, CEA, SCC, and NSE was ranging from 22 to 76 % [33]. Diagnostic value of CYFRA 21-1, CEA CA19-9, CA15-3, and CA 125 assays in pleural effusions have been evaluated and the accuracy of these markers is relatively low (ranging from 40.5 to 85.3 %) [34]. The utility of these biomarkers for discriminating lung cancer from healthy control is limited by their low sensitivity and specificity.

Table 5 Smoking on the performance of salivary mRNA biomarkers

Gene symbol	BRAF	CCNI	EGFR	FGF19	FRS2	GREB1	LZTS1
(smk−)(lc−) versus (smk+)(lc−) ^a	0.82163	0.00089^b	0.28618	0.78227	0.25152	0.28517	0.31999
(smk−)(lc+) versus (smk+)(lc+)	0.35609	0.30198	0.11439	0.83628	0.74682	0.75930	0.35904
(smk−)(lc−) versus (smk+)(lc+)	0.05855	0.01779	0.01189	2.5E−05	0.00111	3.9E−05	0.02560
(smk+)(lc−) versus (smk−)(lc+)	0.07632	0.00026	0.00665	0.12541	0.12215	0.00291	0.00237
(smk−)(lc−) versus (smk−)(lc+)	0.09384	0.03145	0.01532	0.10500	0.05017	0.00063	0.01466
(smk+)(lc−) versus (smk+)(lc+)	0.01103	1.6E−07	0.00015	0.00015	0.00501	6.7E−05	0.00043

^a (smk−), non-smokers; (smk+), smokers; (lc−), healthy control subjects; (lc+), lung cancer patients

^b *t* test-based *p* value is in bold if <0.05

Our study showed that saliva harbors highly discriminatory transcriptomic biomarkers that could be used for effective and non-invasive lung cancer detection. The combination of five transcriptomic markers yielded a sensitivity of 93.75 % and a specificity of 82.81 % in the pre-validation sample set. These values are encouraging and highly promising. The performance of these salivary lung cancer biomarkers are in the same range as the ones found for pre-validated salivary markers for oral cancer [21], Sjögren's syndrome [24], pancreatic cancer [15], and breast cancer [25]. This again highlights the clinical utility of salivary diagnostics for molecular oncology applications. Especially, our lab has successfully developed three candidate proteomic biomarkers for the detection of lung cancer in human saliva [26]. The three proteomic biomarker panel could reach 88.5 % sensitivity and 92.3 % specificity with AUC = 0.90. We foresee that the combination of proteomic and transcriptomic biomarkers could achieve more ideal performance for the detection of lung cancer.

In this study, the profiles of molecular signatures in saliva and their changes between disease and controls have been successfully linked to the detection of lung cancer. Consistent with previous studies, our high-throughput analysis indicated that the mRNA in the saliva supernatant is relatively stable and informative and is a suitable source of disease discriminatory biomarkers [21, 24, 35, 36]. The consistency between different mRNA analysis methods (microarray and qPCR) demonstrated that the alteration of the salivary mRNA signatures between the cancer group and control group could serve as biomarkers for the detection of lung cancer. The known biological functions of these seven validated mRNA biomarkers are shown in supplementary Table S2, mainly linked to nucleotide binding, protein binding, and protein kinase activity. Many of them (EGFR [37], BRAF [38], and LZTS1 [39]) have been shown to be related to lung cancer or other cancers (Table 2).

Of note is that EGFR, a frequently mutated molecular target in lung cancer [40–42], is a discriminatory biomarker in saliva. EGFR was more abundantly expressed in lung carcinoma tissue than in adjacent normal lung, prompting the development of specific pharmacological inhibitors such as gefitinib, which disrupts EGFR kinase activity by binding the adenosine triphosphate pocket within the catalytic domain [41, 42]. Indeed, EGFR mRNA was elevated in the saliva of lung cancer patients (Fig. 2). EGFR amplification was detected in dysplasia, which is associated with lung-cancer risk when detected in the sputum of smokers [43]. Furthermore, BRAF, the downstream factor of EGFR, was mutated with different frequencies in non-small-cell lung cancer and its mutations in tissue might predict clinical response to EGFR inhibitors [44]. FGF19

was overexpressed in lung squamous cell carcinomas relative to normal tissues [45], which is consistent with our finding in saliva. FRS2 β , belongs to the FRS2 family, was a potential prognostic gene for non-small-cell lung cancer, encodes a feedback inhibitor of EGF receptor family members by ERK binding [46].

Smoking, particularly of cigarettes, is by far the main etiological contributor to lung cancer [2, 47, 48]. Smoking's effects on the performance of salivary biomarkers were assessed and compared (Table 5) in this study. In the healthy control subjects group, smoking only elevated salivary level of CCNI significantly ($p < 0.05$). In the lung cancer patients group, none of the seven verified genes were significantly changed because of smoking. In those subjects who have a smoking history and developed lung cancer, all seven salivary mRNAs were significantly elevated. If subjects did not have smoking history but developed lung cancer, four out of the seven salivary mRNAs were significantly elevated, which may due to the limited number of lung cancer patients in this group ($n = 8$). This demonstrated that smoking has minor impact on the performance of these salivary mRNA markers.

Our study is the first systematic study that profiling transcriptome in saliva samples of lung cancer patients. The salivary biomarkers that have been identified and pre-validated are discriminatory for the detection of this cancer, with high sensitivity and specificity. Our findings enhance the prospect of an important role for salivary diagnostics in the detection of systemic diseases. Not only are these saliva-based diagnostic and detection tests for lung cancer simple and non-invasive, they may also represent an improvement in specificity and sensitivity over currently used procedures for lung cancer detection.

We are aware of the lack of mechanistic data that explains the scientific rationale why cancer in an adjacent or distal organ (for example, lung) can result in saliva biomarker profile changes. These mechanistic data are emerging. Using animal models and a systems approach, we have demonstrated that upon development of systemic diseases (lung cancer and melanoma), there are robust and highly discriminatory salivary biomarker profile changes in the mouse model [49]. Furthermore, mediators from distal tumors can alter salivary gland transcription factor activities responsible for the majority of gene and protein expression changes detected in saliva [15, 49].

Conclusions

Through the two phases of biomarker discovery and pre-validation study of salivary mRNA biomarker development for lung cancer detection, our data showed that a panel of pre-validated salivary biomarkers have the potential to

assist current detection strategies for lung cancer. The performance of these pre-validated salivary transcriptomic biomarkers was evaluated, which has demonstrated their discriminatory power for lung cancer detection. To the best of our knowledge, the present study is the first report on de novo salivary transcriptomic biomarker development for lung cancer detection. Although further validation in a larger sample set is necessary for definitive validation, these discovered and confirmed salivary biomarkers have the potential to be used for the non-invasive detection of lung cancer. The results poised the salivary biomarkers for the initiation of a multi-center PRoBE-designed (acronym for prospective specimen collection, retrospective blinded evaluation) validation in a definitive clinical context [50].

Acknowledgments Supported by the University of California Tobacco Research Related Research Program (TRDRP, 20PT-0032) and DOD Lung Cancer Research Program (LC110207). DTWW is also supported by the Felix & Mildred Yip Endowed Professorship.

Conflict of interest HX, LZ, HZ, and DTWW have patent application (US 2011/0207622 A1) pertaining to this work. DTWW is the co-founder of RNAmeTRIX. The other authors disclosed no potential conflicts of interests.

References

- Jemal A, Siegel R, Xu JQ, Ward E (2010) Cancer Statistics, 2010. *CA Cancer J Clin* 60:277–300
- Hecht SS (2002) Cigarette smoking and lung cancer: chemical mechanisms and approaches to prevention. *Lancet Oncol* 3:461–469
- Spira A, Beane J, Shah V, Liu G, Schembri F, Yang XM, Palma J, Brody JS (2004) Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci USA* 101:10143–10148
- Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, Gilman S, Dumas YM, Calner P, Sebastiani P, Sridhar S, Beamis J, Lamb C, Anderson T, Gerry N, Keane J, Lenburg ME, Brody JS (2007) Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 13:361–366
- Petty TL (2000) The early identification of lung carcinoma by sputum cytology. *Cancer* 89:2461–2464
- Gohagan JK, Marcus PM, Fagerstrom RM, Pinsky PF, Kramer BS, Prorok PC, Ascher S, Bailey W, Brewer B, Church T, Engelhard D, Ford M, Fouad M, Freedman M, Gelmann E, Gierada D, Hocking W, Inampudi S, Irons B, Johnson CC, Jones A, Kucera G, Kvale P, Lappe K, Manor W, Moore A, Nath H, Neff S, Oken M, Plunkett M, Price H, Reding D, Riley T, Schwartz M, Spizarny D, Yoffie R, Zylak C (2005) Final results of the Lung Screening Study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer. *Lung Cancer* 47:9–15
- Singh S, Pinsky P, Fineberg NS, Gierada DS, Garg K, Sun YH, Nath PH (2011) Evaluation of reader variability in the interpretation of follow-up CT scans at lung cancer screening. *Radiology* 259:263–270
- Gray RD, Duncan A, Noble D, Imrie M, O'Reilly DSJ, Innes A, Porteous DJ, Greening AP, Boyd AC (2010) Sputum trace metals are biomarkers of inflammatory and suppurative lung disease. *Chest* 137:635–641
- Fliiss MS, Usadel H, Caballero OL, Wu L, Buta MR, Eleff SM, Jen J, Sidransky D (2000) Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. *Science* 287:2017–2019
- Gessner C, Rechner B, Hammerschmidt S, Kuhn H, Hoheisel G, Sack U, Ruschpler P, Wirtz H (2010) Angiogenic markers in breath condensate identify non-small cell lung cancer. *Lung Cancer* 68:177–184
- Lacroix J, Becker HD, Woerner SM, Rittgen W, Drings P, von Knebel Doeberitz M (2001) Sensitive detection of rare cancer cells in sputum and peripheral blood samples of patients with lung cancer by preproGRP-specific RT-PCR. *Int J Cancer* 92:1–8
- Taylor MD, Smith PW, Brix WK, Wick MR, Theodosakis N, Swenson BR, Kozower BD, Lau CL, Jones DR (2009) Fluoro-deoxyglucose positron emission tomography and tumor marker expression in non-small cell lung cancer. *J Thorac Cardiovasc Surg* 137:43–48
- Miller YE, Vu KO, Kennedy TC, Hirsch FR, Petty TL, Bunn PA, Keith RL, Franklin WA, Wolf HJ, Prindiville S, Byers T (2006) Lack of association between sputum atypia and chronic obstructive pulmonary disease mortality. *J Thorac Oncol* 1:302–307
- Xiao H, Wong DT (2011) Proteomics and its applications for biomarker discovery in human saliva. *Bioinformation* 5:294–296
- Zhang L, Farrell JJ, Zhou H, Elashoff D, Akin D, Park NH, Chia D, Wong DT (2010) Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. *Gastroenterology* 138:949–957
- Gonzalez-Begne M, Lu BW, Han XM, Hagen FK, Hand AR, Melvin JE, Yates JR (2009) Proteomic analysis of human parotid gland exosomes by multidimensional protein identification technology (MudPIT). *J Proteome Res* 8:1304–1314
- Wong DT (2006) Salivary diagnostics powered by nanotechnologies, proteomics and genomics. *J Am Dent Assoc* 137:313–321
- Pfaffe T, Cooper-White J, Beyerlein P, Kostner K, Punyadeera C (2008) Diagnostic potential of saliva: current state and future applications. *Clin Chem* 57:675–687
- Zhang L, Xiao H, Wong DT (2009) Salivary biomarkers for clinical applications. *Mol Diagn Ther* 13:245–259
- Hu S, Arellano M, Boontheung P, Wang J, Zhou H, Jiang J, Elashoff D, Wei R, Loo JA, Wong DT (2008) Salivary proteomics for oral cancer biomarker discovery. *Clin Cancer Res* 14:6246–6252
- Li Y, St John MAR, Zhou X, Kim Y, Sinha U, Jordan RCK, Eisele D, Abemayor E, Elashoff D, Park N-H, Wong DT (2004) Salivary transcriptome diagnostics for oral cancer detection. *Clin Cancer Res* 10:8442–8450
- Mager DL, Haffajee AD, Devlin PM, Norris CM, Posner MR, Goodson JM (2005) The salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. *J Transl Med* 3:27
- Wei F, Patel P, Liao W, Chaudhry K, Zhang L, Arellano-Garcia M, Hu S, Elashoff D, Zhou H, Shukla S, Shah F, Ho C-M, Wong DT (2009) Electrochemical sensor for multiplex biomarkers detection. *Clin Cancer Res* 15:4446–4452
- Hu S, Wang J, Meijer J, Jeong S, Xie Y, Yu T, Zhou H, Henry S, Vissink A, Pijpe J, Kallenberg C, Elashoff D, Loo JA, Wong DT (2007) Salivary proteomic and genomic biomarkers for primary Sjögren's syndrome. *Arthritis Rheum* 56:3588–3600
- Zhang L, Xiao H, Karlan S, Zhou H, Gross J, Elashoff D, Akin D, Yan X, Chia D, Karlan B, Wong DT (2010) Discovery and preclinical validation of salivary transcriptomic and proteomic

- biomarkers for the non-invasive detection of breast cancer. *PLoS One* 5:e15573
26. Xiao H, Zhang L, Zhou H, Lee JM, Garon EB, Wong DT (2012) Proteomic analysis of human saliva from lung cancer patients using two-dimensional difference gel electrophoresis and mass spectrometry. *Mol Cell Proteomics* 11:M111.012112
 27. Edgar R, Barrett T (2006) NCBI GEO standards and services for microarray data. *Nat Biotechnol* 24:1471–1472
 28. Therneau TM, Ballman KV (2008) What does PLIER really do? *Cancer Inf* 6:423–431
 29. Hu ZZ, Zimmermann BG, Zhou H, Wang JH, Henson BS, Yu WX, Elashoff D, Krupp G, Wong DT (2008) Exon-level expression profiling: a comprehensive transcriptome analysis of oral fluids. *Clin Chem* 54:824–832
 30. Jessie K, Pang WW, Rahim ZHA, Hashim OH (2010) Proteomic analysis of whole human saliva detects enhanced expression of interleukin-1 receptor antagonist, thioredoxin and lipocalin-1 in cigarette smokers compared to non-smokers. *Int J Mol Sci* 11:4488–4505
 31. Vanni H, Kazeros A, Wang R, Harvey BG, Ferris B, De BP, Carolan BJ, Hubner RH, O'Connor TP, Crystal RG (2009) Cigarette smoking induces overexpression of a fat-depleting gene AZGP1 in the human. *Chest* 135:1197–1208
 32. Molina R, Auge JM, Escudero JM, Marrades R, Vinolas N, Carcereny E, Ramirez J, Filella X (2008) Mucins CA 125, CA 19.9, CA 15.3 and TAG-72.3 as tumor markers in patients with lung cancer: comparison with CYFRA 21-1, CEA, SCC and NSE. *Tumor Biol* 29:371–380
 33. Molina R, Filella X, Auge JM, Fuentes R, Bover I, Rifa J, Moreno V, Canals E, Vinolas N, Marquez A, Barreiro E, Borrás J, Viladiu P (2003) Tumor markers (CEA, CA 125, CYFRA 21-1, SCC and NSE) in patients with non-small cell lung cancer as an aid in histological diagnosis and prognosis—comparison with the main clinical and pathological prognostic factors. *Tumor Biol* 24:209–218
 34. Shitrit D, Zingerman B, Shitrit ABG, Shlomi D, Kramer MR (2005) Diagnostic value of CYFRA 21-1, CEA, CA 19-9, CA 15-3, and CA 125 assays in pleural effusions: analysis of 116 cases and review of the literature. *Oncologist* 10:501–507
 35. Li Y, Zhou X, John M, Wong DTW (2004) RNA profiling of cell-free saliva using microarray technology. *J Dent Res* 83:199–203
 36. Park NJ, Li Y, Yu TW, Brinkman BMN, Wong DT (2006) Characterization of RNA in saliva. *Clin Chem* 52:988–994
 37. Balko JM, Potti A, Saunders C, Stromberg A, Haura EB, Black EP (2006) Gene expression patterns that predict sensitivity to epidermal growth factor receptor tyrosine kinase inhibitors in lung cancer cell lines and human lung tumors. *BMC Genomics* 7:289
 38. Shibata T, Hanada S, Kokubu A, Matsuno Y, Asamura H, Ohta T, Sakamoto M, Hirohashi S (2007) Gene expression profiling of epidermal growth factor receptor/KRAS pathway activation in lung adenocarcinoma. *Cancer Sci* 98:985–991
 39. Nonaka D, Fabbri A, Roz L, Mariani L, Vecchione A, Moore GW, Tavecchio L, Croce CM, Sozzi G (2005) Reduced FEZ1/LZTS1 expression and outcome prediction in lung cancer. *Cancer Res* 65:1207–1212
 40. Kobayashi S, Boggon TJ, Dayaram T, Janne PA, Kocher O, Meyerson M, Johnson BE, Eck MJ, Tenen DG, Halmos B (2005) EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* 352:786–792
 41. Sordella R, Bell DW, Haber DA, Settleman J (2004) Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science* 305:1163–1167
 42. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 304:1497–1500
 43. Herbst RS, Heymach JV, Lippman SM (2008) Molecular origins of cancer: lung cancer. *N Engl J Med* 359:1367–1380
 44. Schmid K, Oehl N, Wrba F, Pirker R, Pirker C, Filipits M (2009) EGFR/KRAS/BRAF mutations in primary lung adenocarcinomas and corresponding locoregional lymph node metastases. *Clin Cancer Res* 15:4554–4560
 45. Desnoyers LR, Pai R, Ferrando RE, Hotzel K, Le T, Ross J, Carano R, D'Souza A, Qing J, Mohtashemi I, Ashkenazi A, French DM (2008) Targeting FGF19 inhibits tumor growth in colon cancer xenograft and FGF19 transgenic hepatocellular carcinoma models. *Oncogene* 27:85–97
 46. Sato T, Gotoh N (2009) The FRS2 family of docking/scaffolding adaptor proteins as therapeutic targets of cancer treatment. *Expert Opin Ther Targets* 13:689–700
 47. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, Davies HR, Ordóñez GR, Mudie LJ, Latimer C, Edkins S, Stebbings L, Chen LN, Jia MM, Leroy C, Marshall J, Menzies A, Butler A, Teague JW, Mangion J, Sun YA, McLaughlin SF, Peckham HE, Tsung EF, Costa GL, Lee CC, Minna JD, Gazdar A, Birney E, Rhodes MD, McKernan KJ, Stratton MR, Futreal PA, Campbell PJ (2010) A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 463:184–190
 48. Church TR, Anderson KE, Caporaso NE, Geisser MS, Le CT, Zhang Y, Benoit AR, Carmella SG, Hecht SS (2009) A prospectively measured serum biomarker for a tobacco-specific carcinogen and lung cancer in smokers. *Cancer Epidemiol Biomarkers Prev* 18:260–266
 49. Gao K, Zhou H, Zhang L, Lee JW, Zhou Q, Hu S, Wolinsky LE, Farrell J, Eibl G, Wong DT (2009) Systemic disease-induced salivary biomarker profiles in mouse models of melanoma and non-small cell lung cancer. *PLoS One* 4:e5875
 50. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD (2008) Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst* 100:1432–1438
 51. Landberg G, Nilsson K, Jirstrom K, Ryden L, Kitching R, Burger AM, Seth A (2005) Cyclin I is expressed in human breast cancer and closely associated with VEGF and KDR expression. *Breast Cancer Res Treat* 89:313–316
 52. Ghosh MG, Thompson DA, Weigel RJ (2000) PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. *Cancer Res* 60:6367–6375