JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Pro–Surfactant Protein B As a Biomarker for Lung Cancer Prediction

Don D. Sin, C. Martin Tammemagi, Stephen Lam, Matt J. Barnett, Xiaobo Duan, Anthony Tam, Heidi Auman, Ziding Feng, Gary E. Goodman, Samir Hanash, and Ayumu Taguchi

Don D. Sin, Stephen Lam, and Anthony Tam, University of British Columbia; Don D. Sin and Anthony Tam, Institute of Heart and Lung Health, James Hogg Research Center, St. Paul's Hospital; Stephen Lam and Xiaobo Duan, British Columbia Cancer Agency, Vancouver, British Columbia; C. Martin Tammemagi, Brock University, St. Catharines, Ontario, Canada; Matt J. Barnett, Ziding Feng, and Gary E. Goodman, Fred Hutchinson Cancer Research Center, Seattle, WA; Heidi Auman, Canary Foundation, Palo Alto, CA; and Ziding Feng, Samir Hanash, and Ayumu Taguchi, University of Texas MD Anderson Cancer Center, Houston, TX.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Ayumu Taguchi, MD, PhD, Department of Clinical Cancer Prevention and Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, 6767 Bertner Ave, Unit 1013 S13.8214, Houston, TX 77030; e-mail: ATaguchi@mdanderson.org.

ABSTRACT

Purpose
Preliminary studies have identified pro–surfactant protein B (pro-SFTPB) to be a promising blood biomarker for non–small-cell lung cancer. We conducted a study to determine the independent predictive potential of pro-SFTPB in identifying individuals who are subsequently diagnosed with lung cancer.

Patients and Methods
Pro-SFTPB levels were measured in 2,485 individuals, who enrolled onto the Pan-Canadian Early Detection of Lung Cancer Study by using plasma sample collected at the baseline visit. Multivariable logistic regression models were used to evaluate the predictive ability of pro-SFTPB in addition to known lung cancer risk factors. Calibration and discrimination were evaluated, the latter by an area under the receiver operating characteristic curve (AUC). External validation was performed with samples collected in the Carotene and Retinol Efficacy Trial (CARET) participants using a case-control study design.

Results
Adjusted for age, sex, body mass index, personal history of cancer, family history of lung cancer, forced expiratory volume in one second percent predicted, average number of cigarettes smoked per day, and smoking duration, pro-SFTPB (log transformed) had an odds ratio of 2.220 (95% CI, 1.727 to 2.853; $P < .001$). The AUCs of the full model with and without pro-SFTPB were 0.741 (95% CI, 0.696 to 0.783) and 0.669 (95% CI, 0.620 to 0.717; difference in AUC $P < .001$). In the CARET Study, the use of pro-SFPTB yielded an AUC of 0.683 (95% CI, 0.604 to 0.761).

Conclusion
Pro-SFTPB in plasma is an independent predictor of lung cancer and may be a valuable addition to existing lung cancer risk prediction models.

INTRODUCTION

Despite reduced smoking rates in the western world, lung cancer remains the leading cause of cancer mortality in the United States and elsewhere. In 2013, it is projected that more than 160,000 Americans will die from lung cancer, which represents 29% of all cancer deaths in men and 26% of all cancer deaths in women.[1] Lung cancer survival is largely dependent on stage at diagnosis. Whereas localized disease (without lymphatic or distant spread) is associated with a 5-year survival greater than 50%, those with distant or regional metastasis have survival measured in weeks to months.[1] Unfortunately, less than 15% of all tumors are found as localized disease. The advent and widespread availability of thoracic computed tomography (CT) scanning has the potential to shift detection to earlier stages and thus improve survival. Data from the National Lung Screening Trial (NLST) suggest that yearly screening with low-dose thoracic CT scans in high-risk current smokers and ex-smokers reduces lung cancer mortality by 20% and total mortality by 7%.[2] However, if these data are generalized and applied to the entire US population, CT screening strategy would cost $1.3 to $2 billion per year.[3]

A potential low-cost solution to this dilemma is to develop a simple blood test that can augment clinical information in risk-stratifying smokers for early detection of lung cancer. Several promising candidate biomarkers have been published previously. Here, we report on the possible role of circulating pro–surfactant protein B (pro-SFTPB) as a biomarker for early detection of non–small-cell lung cancer (NSCLC). SFTPB is synthesized initially as a hydrophilic 42-kD protein by type 2 alveolar pneumocytes and nonciliated bronchiolar cells as pro-SFTPB. On synthesis, pro-SFTPB quickly undergoes

proteolytic cleavage by cysteine proteases in the endoplastic reticulum resulting in the synthesis and secretion of a 9-kD noncollagenous hydrophobic SFTPB, which is the functional mature form of SFTPB.[4] Lung tumor cells (especially adenocarcinomas) have dysregulated SFTPB synthesis leading to the overexpression of pro-SFTPB with muted ability to post-translationally modify the precursor into the mature hydrophobic form.[5,6] We have recently demonstrated that levels of circulating mature SFTPB were increased in patients with resectable NSCLC relative to matched controls.[7] Mass spectrometric identification indicated N-terminal pro-peptide of SFTPB as a potential biomarker for NSCLC. A sandwich pro-SFTPB enzyme-linked immunosorbent assay (ELISA) was then developed and validated in plasma samples obtained at the time of diagnosis from patients with operable NSCLC and from healthy controls. The study aims were thus to determine whether plasma levels of pro-SFTPB are associated with lung cancer independently of known clinical risk factors and improve lung cancer prediction beyond currently existing prediction models in individuals at high risk for lung cancer.

PATIENTS AND METHODS

Study Populations

Pan-Can Study. The initial discovery work was performed on data from the multicenter Pan-Canadian Early Detection of Lung Cancer (Pan-Can) Study (NCT00751660; Screening Methods in Finding Lung Cancer Early in Current or Former Smokers), which enrolled 2,537 individuals with no prior history of lung cancer but with a minimum 2% 3-year risk of lung cancer as predicted by lung cancer risk prediction models.[8,9] Following informed consent, all participants completed a structured epidemiologic questionnaire and had blood samples drawn, processed, and stored in a study biorepository at baseline. The participants also underwent low-dose non–contrast-enhanced thoracic CT scanning and spirometry, according to American Thoracic Society/European Respiratory Society guidelines.[10] All study participants were followed up in person at least every 6 months for at least 2 years or until the date of lung cancer diagnosis, date of death, loss to follow-up, or February 1, 2013, whichever came first. The primary outcome was the occurrence of lung cancer during follow-up. Details of the Pan-Can Study have been presented elsewhere and are summarized in the Data Supplement.[11] The study was approved by the Clinical Research Ethics Board of the University of British Columbia and at each of the participating Pan-Can Study sites.

CARET Study. The validation test samples consisted of sera collected from participants in the Carotene and Retinol Efficacy Trial (CARET). CARET was a multicenter, randomized, double-blind, placebo-controlled study to evaluate the efficacy and safety of daily supplementation with 30 mg of beta-carotene and 25,000 IU retinyl palmitate on primary lung cancer prevention.[12] Eligible participants were men and women age 50 to 69 years who were either current or former smokers (quit within the previous 6 years) and had at least 20 pack-years of cigarette smoking (n = 14,254) or were men age 45 to 69 years who were current or former smokers (quit no more than 15 years prior to the start of the study) and had a substantial history of occupational asbestos exposure (n = 4,060). Participants were enrolled from 1985 to 1994 and were observed for cancer and mortality outcomes until 2005. For this study, we randomly selected 61 current smokers who developed NSCLC during follow-up and analyzed pro-SFTPB in serum samples, which had been collected within 12 months before the diagnosis of NSCLC. For each case, two controls who were free of lung cancer throughout CARET follow-up were selected and were matched for age, sex, smoking history (current smoker), study enrollment cohort, and the date of blood draw. One-to-two case-control matching was possible for all cases except for one patient, leading to 121 control participants. The clinical characteristics of the CARET participants are provided in the Data Supplement. All serum samples were obtained following informed consent and approval by the institutional review board of the Fred Hutchinson Cancer Research Center.

Pro-SFTPB Assay

By using mass spectrometry, we discovered N-terminal and C-terminal pro-peptides of SFTPB in the circulatory systems of mice that harbored lung adenocarcinoma and in the conditioned media of NSCLC cell lines (Data Supplement). Mouse monoclonal antibodies against the N-terminus of pro-SFTPB (Data Supplement) were raised, leading to the development of a sandwich pro-SFTPB ELISA, which specifically reacted with pro-SFTPB and did not react with other surfactant proteins or mature SFTPB. The standards were calibrated according to the absolute mass of the recombinant antigen (details are provided in the Data Supplement). We then validated this assay with plasma samples obtained at the time of diagnosis from patients with operable NSCLC (n = 28) and healthy controls (n = 38). These samples had previously been analyzed for levels of mature SFTPB by ELISA (see the Data Supplement for demographics of the newly diagnosed NSCLC set). Plasma levels of pro-SFTPB were significantly higher in cases compared with controls ($P < .001$ by the Mann-Whitney U test; Data Supplement). The area under the receiver operating characteristic curve (AUC) of pro-SFTPB was superior to that of mature SFTPB[7] (0.793 and 0.646, respectively; Data Supplement).

For the Pan-Can Study, the baseline plasma samples (ie, samples taken at the time of enrollment) were used for the assay. For both the Pan-Can and CARET studies, samples were blinded and analyzed by using pro-SFTPB ELISA (details of the assay are provided in the Data Supplement). For samples whose pro-SFTPB levels were below the level of detection, we assigned a value that was half the detection limit. The median coefficient of variation was 6.1%. Because the Pan-Can Study and the CARET Study used different standards, the absolute levels of pro-SFTPB are not directly comparable.

Statistical Methods

Descriptive comparisons of study variables between groups used Fisher's exact test for categorical data, t test for continuous data, and nonparametric test of trend for ordinal data. Multivariable logistic regression models were used to evaluate whether pro-SFTPB was independently associated with lung cancer. Known risk factors for lung cancer were evaluated in models and included age, sex, body mass index (BMI), personal history of cancer, family history of lung cancer, forced expiratory volume in one second percent predicted (FEV_1 % pred), average number of cigarettes smoked per day, and duration smoked. Pro-SFTPB was right skewed, and log-transformed pro-SFTPB (log-proSFTPB) was used in modeling. Selected interaction terms,
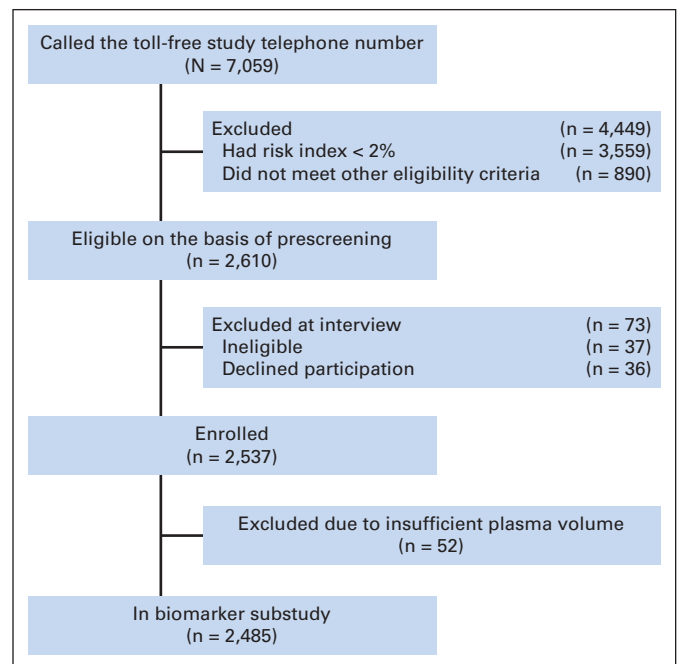

Fig 1. Flow diagram of recruitment to the Pan-Canadian Early Detection of Lung Cancer Study.

Table 1. Baseline Characteristics of Participants in the Pan-Canadian Early Detection of Lung Cancer Study Overall and by Lung Cancer Status

Characteristic	No Cancer		Lung Cancer		Total		$P*$
	No.	%	No.	%	No.	%	
Age, years							.036†
50-54	234	95.90	10	4.10	244	10.38	
55-59	443	96.72	15	3.28	458	19.49	
60-64	726	95.53	34	4.47	760	32.34	
65-69	568	93.57	39	6.43	607	25.83	
≥ 70	266	94.66	15	5.34	281	11.96	
Sex							.020
Male	1,242	96.13	50	3.87	1,292	54.98	
Female	995	94.05	63	5.95	1,058	45.02	
BMI							.016†
Underweight (< 18.5 kg/m^2)	20	95.24	1	4.76	21	0.88	
Normal (18.5-24.9 kg/m^2)	803	93.59	55	6.41	858	36.51	
Overweight (25-29.9 kg/m^2)	1,031	96.18	41	3.82	1,072	45.62	
Obese (≥ 30 kg/m^2)	383	95.99	16	4.01	399	16.98	
Personal history of cancer							.112
No	2,095	95.36	102	4.64	2,197	93.77	
Yes	135	92.47	11	7.53	146	6.23	
Family history of cancer							.353
No	1,458	95.54	68	4.46	1,526	66.12	
Yes	740	94.63	42	5.37	782	33.88	
Pneumonia							.037
No	1,648	95.76	73	4.24	1,721	73.42	
Yes	583	93.58	40	6.42	623	26.58	
Emphysema							.423
No	2,091	95.26	104	4.74	2,195	93.76	
Yes	137	93.84	9	6.16	146	6.24	
Smoking status							.489
Former smoker	871	95.61	40	4.39	911	38.77	
Current smoker	1,366	94.93	73	5.07	1,439	61.23	
Race/ethnicity							.674†
White	2,175	95.14	111	4.86	2,286	97.65	
Asian	22	95.65	1	4.35	23	0.98	
Aboriginal	7	100	0	0	7	0.30	
Black or African Canadian	12	92.31	1	7.69	13	0.56	
Other	12	100	0	0.00	12	0.51	
Education							.936†
8th grade or less	63	94.03	4	5.97	67	2.85	
9th to 11th grade	288	94.12	18	5.88	306	13.02	
High school graduate	585	96.22	23	3.78	608	25.87	
Technical/vocational school	239	94.09	15	5.91	254	10.81	
Associate's degree/some college	428	94.48	25	5.52	453	19.28	
Bachelor's degree	401	96.63	14	3.37	415	17.66	
Advanced degree	233	94.33	14	5.67	247	10.51	
FEV$_1$ ratio of predicted	2,220		111		2,331		.0034‡
Mean		0.82		0.77		0.82	
SD		0.18		0.19		0.18	
Median		0.83		0.77		0.83	
IQR		0.71-0.94		0.67-0.86		0.71-0.94	
Range		0.15-1.68		0.27-1.36		0.15-1.68	
Pro-SFTPB, ng/mL	2,193		112		2,305		< .001‡
Mean		43.5		75.0		45.1	
SD		42.1		64.5		43.9	
Median		30.7		54.2		31.6	
IQR		16.5-54.3		30.8-99.1		16.9-55.7	
Range		1.2-426.9		10.0-344.4		1.2-426.9	
No. of cigarettes smoked per day	2,237		113		2,350		.8943‡
Mean		24.68		24.84		24.69	
SD		10.51		12.51		10.61	

(continued on following page)

Table 1. Baseline Characteristics of Participants in the Pan-Canadian Early Detection of Lung Cancer Study Overall and by Lung Cancer Status (continued)

Characteristic	No Cancer		Lung Cancer		Total		
	No.	%	No.	%	No.	%	P*
Median	25		25		25		
IQR	20-25		20-25		20-25		
Range	1-100		5-100		1-100		
Smoking duration, years	2,235		112		2,347		.0012‡
Mean	43.97		45.91		44.06		
SD	5.87		6.06		5.89		
Median	44		46.5		44		
IQR	40-48		42-50		40-48		
Range	11-69		27-60		11-69		
Time since quitting smoking, years	2,237		113		2,350		.9128‡
Mean	17.85		17.62		17.84		
SD	21.29		21.71		21.31		
Median	0		0		0		
IQR	0-41		0-41		0-41		
Range	0-61		0-55		0-61		

Abbreviations: BMI, body mass index; FEV$_1$ ratio of predicted, forced expiratory volume in one second % predicted; IQR, interquartile range; pro-SFTPB, pro–surfactant protein B; SD, standard deviation.
*Fisher's exact test.
†Nonparametric test.
‡Satterthwaite's unequal variance *t* test.

including main effects and cross-product terms in the model and nonlinear associations between continuous variables and lung cancer, were evaluated by multivariable fractional polynomials.[13] No interactions or nonlinear relationships were found to be significant.

Regarding prediction, improvement in discrimination was assessed by comparing AUC between nested models with and without log-proSFTPB. For AUCs, 95% CIs were prepared by using bootstrap resampling with 1,000 samples.[14] Calibration was assessed by evaluating the mean and 90th percentile absolute errors.[15] For each model, we calculated a Brier score.[16] Optimism or overfit in models was assessed by using a bootstrap method that employed Harrell's RMS package in R (version 3.0.1).[15,17] Bootstrap-optimism–corrected estimates of AUCs and Brier statistics are also presented. For comparative purposes, we produced Cox proportional hazards survival models analogous to our logistic regression models. All analyses, statistics, and figures were prepared by using STATA/MP version 12.1 (STATA, College Station, TX). All presented *P* values are two-sided.

In the CARET Study, pro-SFTPB levels were categorized into quintiles on the basis of the distribution in control participants. Logistic regression was performed to obtain odds ratios (ORs), and adjusted ORs were generated by using multiple logistic regression analyses in which we controlled for matching variables (age, sex, smoking status, enrollment period, and blood draw visit), pack-years, years since quitting smoking, asbestos exposure, and BMI.

RESULTS

Study Populations

For the Pan-Can Study, enrollment began September 24, 2008, and was completed on December 17, 2010 (Fig 1). As of February 1, 2013, 113 of 2,537 individuals had been diagnosed with lung cancer. Pro-SFTPB data were available for 2,485 individuals. The minimum, median, and maximum follow-up durations were 0.02, 3.02, and 4.36 years. During this follow-up period, 187 individuals (7.4%) were lost to follow-up. Loss-to-follow-up status was not associated with pro-SFTPB (*P* = .527), nor were pro-SFTPB levels associated with time to loss-to-follow-up (*P* = .954). The overall cumulative incidence of

lung cancer was 4.45%, and the annual incidence rate of lung cancer was 1.48 per 100 person-years of follow-up. Distributions of study variables by lung cancer status are presented in Table 1.

Pro-SFTPB was measured in nanograms per milliliter; for pro-SFTPB, the mean was 45.32 ng/mL (standard deviation, 44.64 ng/mL), and the median was 31.93 ng/mL (interquartile range, 16.92 to 56.26 ng/mL). Distributions of pro-SFTPB by study variables are presented in Table 2.

Prediction Models

In an unadjusted logistic model of log-proSFTPB predicting lung cancer, the OR was 2.331 (95% CI, 1.837 to 2.958; *P* < .001) and the AUC was 0.690 (95% CI, 0.642 to 0.735). The sensitivity and specificity for log-proSFTPB over the range of model probabilities are presented in the Data Supplement. When the aforementioned model probability for positivity is set to $P \geq .032$, sensitivity is 80.4%, specificity is 40.1%, the positive predictive value is 6.4%, and the negative predictive value is 97.6%. In the unadjusted logistic model of log-proSFTPB for events occurring at least 1 year after baseline blood draw, AUC was 0.655 (95% CI, 0.570 to 0.719).

In the logistic model fully adjusted for lung cancer risk factors, including smoking and nonsmoking predictors, log-proSFTPB was a significant independent predictor of lung cancer (OR, 2.220; 95% CI, 1.727 to 2.853; *P* < .001; Table 3). In the fully adjusted model, when the analysis was limited to lung cancers occurring within the first year, the OR for pro-SFTPB was 2.53 (95% CI, 1.79 to 3.59; *P* < .001). The AUCs for the full logistic models with and without log-proSFTPB were 0.741 (95% CI, 0.696 to 0.783) and 0.669 (95% CI, 0.620 to 0.717; difference in AUC *P* < .001; Fig 2). The respective bootstrap-bias–corrected AUCs were 0.718 and 0.636. When pro-SFTPB concentrations were grouped into quintiles, the univariate OR per one level change was 1.62 (95% CI, 1.39 to 1.89; *P* < .001) with a model AUC of 0.579 (95% CI, 0.526 to .626). In the multivariable model, the OR was

Table 2. Distribution of Pro-SFTPB by Selected Study Variables

Variable	No. of Participants	Mean	SD	Median	IQR	Range	P*
Pro-SFTPB (ng/mL)	2,485	45.32	44.64	31.93	16.92-56.26	1.15-426.86	N/A
Log-proSFTPB	2,485	3.49	0.83	3.49	2.89-4.05	0.77-6.06	N/A
Age, years							.219†
50-54	270	46.26	47.19	33.27	16.21-57.36	1.15-324.35	
55-59	491	45.34	44.13	33.11	17.28-57.67	2.56-426.86	
60-64	791	42.82	42.81	29.82	16.26-53.85	1.42-330.24	
65-69	642	44.10	43.05	30.18	16.72-54.64	2.85-344.38	
≥ 70	291	53.93	50.30	37.98	19.12-66.82	4.29-294.48	
Sex							.0794
Male	1,377	46.73	45.54	33.24	17.59-58.35	1.42-426.86	
Female	1,108	43.58	43.47	30.02	16.31-53.45	1.15-344.38	
BMI, kg/m²							.001†
Underweight (< 18.5)	27	37.28	24.57	28.76	18.95-55.37	7.45-98.90	
Normal (18.5-24.9)	919	47.70	46.39	33.70	18.63-57.67	1.15-426.86	
Overweight (25-29.9)	1,123	45.53	45.17	31.64	16.01-58.68	2.18-372.25	
Obese (≥ 30)	416	40.03	39.62	28.32	15.87-48.71	1.42-282.03	
Personal history of cancer							.4399
No	2,321	45.14	44.75	31.65	16.88-55.85	1.15-426.86	
Yes	156	47.95	43.88	34.90	18.58-59.51	2.18-282.03	
Family history of cancer							.2389
No	1,625	45.90	45.13	32.08	17.29-56.94	1.42-426.86	
Yes	819	43.71	42.59	30.85	16.21-55.79	1.15-330.24	
Pneumonia							.0651
No	1,813	44.29	43.32	31.79	16.88-55.40	1.42-426.86	
Yes	665	48.22	48.15	31.73	16.96-58.71	1.15-324.35	
Emphysema							.0984
No	2,315	44.82	43.73	31.71	16.94-55.80	1.15-426.86	
Yes	160	51.97	53.19	36.25	15.38-64.30	3.14-328.88	
Smoking status							< .001
Former smoker	938	35.51	42.42	21.26	12.13-40.81	1.15-426.86	
Current smoker	1,547	51.28	44.92	38.48	21.77-63.99	2.18-344.38	
Race/ethnicity							.396†
White	2,410	45.50	44.95	31.96	16.93-56.33	1.15-426.86	
Asian	29	40.28	34.98	33.30	15.62-50.70	5.04-160.48	
Aboriginal	8	35.19	26.60	30.89	17.19-47.46	3.92-86.55	
Black or African Canadian	13	45.89	45.79	27.05	18.63-64.41	5.90-172.51	
Other	15	33.53	23.13	30.94	11.95-41.49	6.15-78.02	
Education							< .001†
8th grade or less	74	56.22	42.43	47.60	23.77-75.39	7.22-241.07	
9th to 11th grade	326	53.33	51.04	38.23	18.48-67.05	1.15-372.25	
High school graduate	649	42.99	43.20	29.96	16.68-52.69	1.42-344.38	
Technical/vocational school	264	48.74	48.06	33.62	17.46-64.93	2.15-330.24	
Associate's degree/some college	475	43.66	39.07	32.03	16.68-57.59	2.39-274.28	
Bachelor's degree	426	42.99	46.71	28.99	15.65-51.78	4.15-426.86	
Advanced degree	271	41.57	41.32	26.62	16.54-52.06	2.18-235.45	

Abbreviations: IQR, interquartile range; N/A, not applicable; pro-SFTPB, pro–surfactant protein B; SD, standard deviation.
*P values are for Satterthwaite's unequal variance t test, unless otherwise marked.
†Nonparametric test.

1.59 (95% CI, 1.36 to 1.87; $P < .001$) with a model AUC of 0.730 (95% CI, 0.680 to 0.775). The improvement in magnitude in discrimination attributable to pro-SFTPB is large compared with that of most lung cancer predictors.[18] Of the 113 lung cancers, 96 (85.0%) were stage I or II. When the full model was estimated in these individuals, log-proSFTPB remained a statistically significant predictor (OR, 2.195; 95% CI, 1.679 to 2.870; $P < .001$) and significantly improved the AUC compared with the nested model excluding log-proSFTPB (0.735 v 0.659; $P < .001$).

The mean and 90th percentile absolute error (observed minus predicted probabilities) in the model without log-proSFTPB were

0.005 and 0.007; for the model with log-proSFTPB, they were 0.004 and 0.010. For both models, the mean absolute errors in all deciles of model-predicted risk were less than 1% (Data Supplement). For the full model with log-proSFTPB versus the nested model without log-proSFTPB, the Brier scores were 0.0438 and 0.0448, and the boot-strap-bias–corrected scores were 0.0442 and 0.0450, respectively. These statistics indicate that calibration was excellent in both models but slightly better in the model with log-proSFTPB.

The magnitude of Cox model hazard ratios and CIs was similar to that of the ORs in the logistic models (Data Supplement). When the full Cox model was limited to lung cancers that were diagnosed more

Table 3. Logistic Regression Prediction Model With Log-Transformed Pro-SFTPB Predicting Lung Cancer in the Pan-Canadian Early Detection of Lung Cancer Study (n = 2,233)

Predictor Variables	OR	95% CI	P	Beta Coefficients
Age (per year)	1.023	0.978 to 1.070	.326	0.0226304
Sex (male v female)	0.592	0.391 to 0.897	.013	−.5239236
BMI (kg/m^2)	0.957	0.912 to 1.005	.077	−.0439466
Personal history of cancer (yes v no)	1.379	0.684 to 2.780	.369	0.3215962
Pneumonia (yes v no)	1.341	0.876 to 2.055	.177	0.2936113
Family history of cancer (yes v no)	1.412	0.923 to 2.160	.112	0.344805
FEV$_1$ ratio of predicted	0.270	0.091 to 0.804	.019	−1.310217
Cigarettes smoked per day	1.010	0.991 to 1.030	.292	0.0102629
Smoking duration	1.034	0.989 to 1.082	.142	0.0336636
Log-proSFTPB	2.220	1.727 to 2.853	< .001	0.7975728
Model constant				−6.948646

Abbreviations: BMI, body mass index; FEV$_1$ ratio of predicted, forced expiratory volume in one second % predicted; OR, odds ration; pro-SFTPB, pro–surfactant protein B.

than 1 year and more than 2 years after study entry, the hazard ratios for log-proSFTPB were 1.875 (95% CI, 1.346 to 2.610; $P <$ .001; number of subjects with lung cancer, 53), and 1.650 (95% CI, 1.028 to 2.649; $P =$ .038; number of subjects with lung cancer, 26).

For the CARET Study, pro-SFTPB levels were significantly higher among NSCLC cases compared with controls ($P <$ .001) and receiver operating characteristic analysis yielded an AUC of 0.683 (95% CI, 0.604 to 0.761; Table 4 and Data Supplement for clinical characteristics). In terms of histologic subgroups, pro-SFTPB levels were significantly increased in adenocarcinoma, but not in squamous cell carcinoma compared with matched controls. In multivariable logistic regression analysis, the risk of NSCLC increased along with the pro-SFTPB concentration gradient in the CARET set ($P_{trend} <$ .001, adjusted for matching variables; Table 4). The risk of NSCLC also increased per quintile increase (OR, 1.77 [95% CI, 1.35 to 2.33] adjusted for matching variables; OR, 1.64 [95% CI, 1.22 to 2.20] adjusted for matching variables, pack-years, years since quitting smoking, asbestos exposure, and BMI).
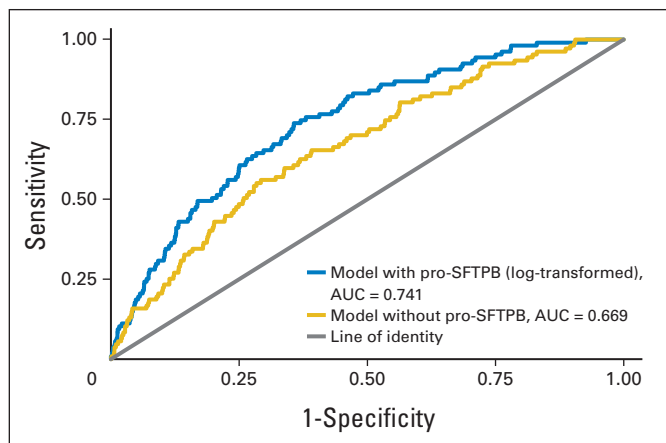

Fig 2. Area under the receiver operating characteristic curves (AUCs) for the full model with and without pro–surfactant protein B (pro-SFTPB) in the Pan-Canadian Early Detection of Lung Cancer Study.

DISCUSSION

Our study demonstrates that plasma pro-SFTPB is significantly and independently associated with lung cancer and adds to lung cancer prediction beyond that contributed by established risk factors in two independent cohorts. Furthermore, pro-SFTPB was associated with early-stage (I and II) lung cancer and with lung cancers diagnosed more than 1 year after plasma collection, suggesting its potential utility in predicting early-stage NSCLC tumors that may be amenable to surgical resection.

The precise mechanism by which circulating pro-SFTPB levels become increased in individuals at risk for NSCLC is unknown. Surfactant proteins are transcriptionally regulated by thyroid transcription factor 1 (TTF-1)/NK2 homeobox 1 (NKX2-1), which plays a crucial role in lung development.[19] TTF-1/NKX2-1 is markedly increased in regions of lung parenchyma undergoing regeneration and repair.[20] In addition, increased expression of TTF-1/NKX2-1 in alveolar type II cells has been found to induce dose-dependent alterations in alveolar morphology, epithelial cell hyperplasia, emphysema, and pulmonary inflammation.[21] It is also possible that circulating pro-SFTPB is a biomarker of lung conditions that predispose to lung cancer such as chronic obstructive pulmonary disease. In the Pan-Can Study, in which lung function data were carefully measured, the inclusion of FEV$_1$ did not materially alter the overall findings, suggesting an alternate explanation. Additional studies will be needed to understand the mechanisms and pathways leading to increased plasma expression of pro-SFTPB in those who develop NSCLC.

Our sample size and number of outcome events were adequate to find statistically significant results regarding the relationship between plasma levels of pro-SFTPB and lung cancer risk, provide effect estimates with precise CIs, and demonstrate significant incremental improvement in prediction. However, because more than 75% of the lung cancer cases diagnosed in the Pan-Can Study were adenocarcinomas, we could not adequately evaluate whether the relationship between pro-SFTPB and lung cancer risk differed across different histologic tumor subtypes. In the CARET Study, which proportionately had more cases of squamous cell

Table 4. Relationship Between Pro-SFTPB and the Risk of NSCLC in the CARET Study According to Quintiles of Serum Pro-SFTPB Concentrations

Variable	Q1			Q2					Q3					Q4					Q5				
	No.	OR*	Adjusted OR†	No.	OR*	95% CI	Adjusted OR†	95% CI	No.	OR*	95% CI	Adjusted OR†	95% CI	No.	OR*	95% CI	Adjusted OR†	95% CI	No.	OR*	95% CI	Adjusted OR†	95% CI
Total	26	1 (ref)	1 (ref)	29	2.64	0.46 to 15.11	2.74	0.42 to 17.12	35	5.61	1.11 to 28.37	6.66	1.15 to 38.32	45	11.63	2.40 to 56.33	12.24	2.22 to 67.43	46	12.34	2.54 to 60.00	9.64	1.74 to 53.34
Controls‡	24			24					24					24					24				
NSCLC	2			5					11					21					22				

Abbreviations: BMI, body mass index; CARET, Carotene and Retinol Efficacy Trial; NSCLC, non–small-cell lung cancer; OR, odds ratio; pro-SFTPB, pro–surfactant protein B; Q, quartile.
*Adjusted for matching variables (age, sex, smoking status, enrollment period, and blood draw visit). $P_{trend} < .001$.
†Adjusted for matching variables (pack-years, years since quitting smoking, asbestos exposure, and BMI). $P_{trend} = .001$.
‡One sample was excluded from this analysis because of missing BMI data.

carcinoma, pro-SFTPB appeared to be less predictive in squamous cell carcinomas than in adenocarcinomas (Data Supplement). These data, however, should be interpreted with caution because of small numbers of study participants. Moreover, our study took place in individuals at high risk for lung cancer. It is unclear how generalizable the findings are to individuals at lower risk. Selection of individuals for lung cancer screening based on high risk rather than the NLST criteria (age 55 to 79 years, $\geq$ 30 pack-years smoked, < 15 years of time since quitting smoking) has been shown to save more lives and to be more efficient.[9] Thus, in the future, candidates selected for lung screening may be similar to those included in this study.

One of the strengths of our study is that it was prospective, with biospecimens collected at study baseline when individuals had no clinical history of lung cancer. One of the primary aims of the study was to evaluate the use of thoracic CT for lung cancer screening. This led to detailed and rigorous diagnostic assessment of individuals for the study outcome, lung cancer. Known predictors of lung cancer from previous risk prediction models were forced into models in this study regardless of statistical significance to attempt to truly test the added incremental value of pro-SFTPB. In addition, we were able to demonstrate the predictive ability of pro-SFTPB in two independent cohorts.

The association between pro-SFTPB and NSCLC and the important ability of pro-SFTPB to predict NSCLC needs to be further validated in different populations in well-designed prospective studies. If current findings are validated, it will be important to integrate pro-SFTPB into a comprehensive NSCLC prediction model based on a large sample with adequate numbers of outcomes. Once pro-SFTPB becomes integrated into a comprehensive lung cancer prediction model, it is expected to make important contributions to clinical and public health practice and may lead to more efficient lung cancer screening by improving enrollment criteria for identifying those who would benefit from screening.

AUTHOR CONTRIBUTIONS

Conception and design: Don D. Sin, C. Martin Tammemagi, Stephen Lam, Heidi Auman, Ziding Feng, Ayumu Taguchi
Financial support: Don D. Sin, Heidi Auman
Provision of study materials or patients: Matt J. Barnett, Gary E. Goodman
Collection and assembly of data: C. Martin Tammemagi, Stephen Lam, Matt J. Barnett, Xiaobo Duan, Anthony Tam, Gary E. Goodman, Ayumu Taguchi
Data analysis and interpretation: Don D. Sin, C. Martin Tammemagi, Stephen Lam, Matt J. Barnett, Samir Hanash, Ayumu Taguchi
Manuscript writing: All authors
Final approval of manuscript: All authors

REFERENCES

1. Siegel R, Naishadham D, Jemal A: Cancer statistics, 2012. CA Cancer J Clin 62:10-29, 2012

2. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al: Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 365:395-409, 2011

3. Goulart BH, Bensink ME, Mummy DG, et al: Lung cancer screening with low-dose computed tomography: Costs, national expenditures, and cost-effectiveness. J Natl Compr Canc Netw 10:267-275, 2012

4. Guttentag S, Robinson L, Zhang P, et al: Cysteine protease activity is required for surfactant protein B processing and lamellar body genesis. Am J Respir Cell Mol Biol 28:69-79, 2003

5. Khoor A, Whitsett JA, Stahlman MT, et al: Utility of surfactant protein B precursor and thyroid transcription factor 1 in differentiating adenocarcinoma of the lung from malignant mesothelioma. Hum Pathol 30:695-700, 1999

6. O'Reilly MA, Gazdar AF, Clark JC, et al: Glucocorticoids regulate surfactant protein synthesis in a pulmonary adenocarcinoma cell line. Am J Physiol 257:L385-L392, 1989

7. Taguchi A, Politi K, Pitteri SJ, et al: Lung cancer signatures in plasma based on proteome profiling of mouse tumor models. Cancer Cell 20:289-299, 2011

8. Tammemagi CM, Pinsky PF, Caporaso NE, et al: Lung cancer risk prediction: Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial models and validation. J Natl Cancer Inst 103:1058-1068, 2011

9. Tammemagi M, Katki HA, Hocking WG, et al: Selection criteria for lung-cancer screening. N Engl J Med 368:728-736, 2013

10. Miller MR, Hankinson J, Brusasco V, et al: Standardisation of spirometry. Eur Respir J 26:319-338, 2005

11. McWilliams A, Tammemagi M, Mayo JR, et al: Probability of cancer in pulmonary nodules detected on first screening CT. N Engl J Med 369:910-919, 2013

12. Goodman GE, Thornquist MD, Balmes J, et al: The Beta-Carotene and Retinol Efficacy Trial: Incidence of lung cancer and cardiovascular disease mortality during 6-year follow-up after stopping beta-carotene and retinol supplements. J Natl Cancer Inst 96:1743-1750, 2004

13. Royston P, Sauerbrei W: Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables. Chichester, West Sussex, United Kingdom, John Wiley & Sons, 2008

14. Pepe M, Longton G, Janes H: Estimation and comparison of receiver operating characteristic curves. Stata J 9:1-16, 2009

15. Harrell FE Jr: Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York, NY, Springer-Verlag, 2001

16. Ikeda M, Itoh S, Ishigaki T, et al: Application of resampling techniques to the statistical analysis of the Brier score. Methods Inf Med 40:259-264, 2001

17. Harrell FE Jr: Regression Modeling Strategies: Package 'rms'. 2010:1-217

18. Lam S, Boyle P, Healey GF, et al: EarlyCDT-Lung: An immunobiomarker test as an aid to early detection of lung cancer. Cancer Prev Res (Phila) 4:1126-1134, 2011

19. Whitsett JA, Glasser SW: Regulation of surfactant protein gene transcription. Biochim Biophys Acta 1408:303-311, 1998

20. Stahlman MT, Gray ME, Whitsett JA: Expression of thyroid transcription factor-1 (TTF-1) in fetal and neonatal human lung. J Histochem Cytochem 44:673-678, 1996

21. Wert SE, Dey CR, Blair PA, et al: Increased expression of thyroid transcription factor-1 (TTF-1) in respiratory epithelial cells inhibits alveolarization and causes pulmonary inflammation. Dev Biol 242:75-87, 2002

Appendix

The following institutions and individuals participated in the Pan-Canadian Early Detection of Lung Cancer Study Group: British Columbia Cancer Agency, Vancouver, BC: Annette McWilliams, MB; Sukhinder Atkar-Khattra, BSc; John R. Mayo, MD; Ken Evans, MD; Richard Finley, MD; John Yee, MD; John English, MD; Vancouver General Hospital, Vancouver, BC, and University Health Network-Princess Margaret Cancer Centre and Toronto General Hospital, Toronto, ON: Ming-Sound Tsao, MD; Heidi Roberts, MD; Geoffrey Liu, MD; Kam Soghrati, MD; Kazuhiro Yasufuku, MD, PhD; Institut universitaire de cardiologie et de pneumologie de Québec, Quebec City, QC: Simon Martel, MD; Francis Laberge, MD; Michel Gingras, MD; John Goffin, MD; Serge Juravinski Hospital and Cancer Center, Hamilton, ON: Serge Puksa, MD; Lori Stewart, MD; Scott Tsai, MD; Dalhousie University, Halifax, NS: Michael R. Johnston, MD; Daria Manos, MD; Ottawa Hospital Cancer Centre, Ottawa, ON: Garth Nicholas, MD; Glenwood D. Goss, MD; Jean M. Seely, MD; Kayvan Amjadi, MD; University of Calgary, Calgary, AB: Alain Tremblay, MD; CM, Paul Burrowes, MD; Paul MacEachern, MD; Memorial University, St. John's, NF: Rick Bhatia, MD.