Biomarkers, 2011; 16(2): 181–191

informa
healthcare

RESEARCH ARTICLE

Proteomic detection of cancer in asbestosis patients using SELDI-TOF discovered serum protein biomarkers

Brian C. Tooker[1], Lee S. Newman[1], Russell P. Bowler[2], Antti Karjalainen[3], Panu Oksa[3], Harri Vainio[3], Eero Pukkala[4], and Paul W. Brandt-Rauf[5]

[1]*Colorado School of Public Health and School of Medicine, University of Colorado Denver, Aurora, CO, USA,* [2]*Department of Medicine, National Jewish Health, Denver, CO, USA,* [3]*Finnish Institute of Occupational Health, Helsinki, Finland,* [4]*Finnish Cancer Registry, Institute for Statistical and Epidemiological Cancer Research, Helsinki, Finland, and* [5]*School of Public Health, University of Illinois at Chicago, Chicago, IL, USA*

Abstract

Objectives: To identify biomarkers for cancer in asbestosis patients.

Methods: SELDI-TOF and CART were used to identify serum biomarker profiles in 35 asbestosis patients who subsequently developed cancer and 35 did not develop cancer.

Results: Three polypeptide peaks (5707.01, 6598.10, and 20,780.70 Da) could predict the development of cancer with 87% sensitivity and 70% specificity. The first two peaks were identified as KIF18A and KIF5A, respectively, and are part of the Kinesin Superfamily of proteins.

Conclusions: We identified two Kinesin proteins that can be potentially used as blood biomarkers to identify asbestosis patients at risk of developing lung cancer.

Keywords: *Disease progression; CART analysis; serum*

Introduction

It is well-established that some people with asbestosis are at a higher risk for developing lung cancer, mesothelioma, and other cancers (Reid et al., 2005). A number of protein biomarkers have been identified and associated with carcinogenesis (Shamma et al., 2000; Mori et al., 2002; Winer et al., 2002; Lea et al., 2004; Cappello et al., 2005). We have previously used the banked serum samples from a cohort of Finnish asbestosis cases who have been followed up for the subsequent development of cancer to validate some of these known cancer-related protein biomarkers in terms of their ability to identify those individuals at highest risk for the occurrence of cancer (Li et al., 2004). In this regard, combinations of these protein biomarkers have been found to have relatively high positive predictive value (0.76) and specificity (0.85) but to be less robust in terms of sensitivity (0.51), limiting their value as clinical biomarkers. This suggested the need to use other approaches for new biomarker discovery that would have increased sensitivity, such as proteomics.

Surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) mass spectrometry is a powerful technique that allows researchers to look for biomarkers within entire proteomes. SELDI-TOF attempts to simplify the proteome by utilizing two distinct selection criteria. An initial separation of the proteome by pH selection yields six unique protein fractions. Each of these serum fractions can then be further simplified by the capacity of individual proteins within the fraction to bind to the different binding chemistries available on various chips (IMAC30, CM10, and H50). This binding simplifies the proteome further so that individual biomarkers can be identified. Depending on the stringency of the washing steps, each Chip/Serum Fraction combination can contain several dozen to several hundred peaks (proteins

Address for Correspondence: Brian C. Tooker, Division of Allergy and Clinical Immunology, University of Colorado Denver, Building: RC2, Mail Stop: B164, 12700 E. 19th Ave, Aurora, CO, 80045. USA. Tel: (303)-724-7208. E-mail: Brian.Tooker@ucdenver.edu

(Received 01 July 2010; revised 23 November 2010; accepted 23 November 2010)

ISSN 1354-750X print/ISSN 1366-5804 online © 2011 Informa UK, Ltd.
DOI: 10.3109/1354750X.2010.543289

and/or polypeptide chains) that are subjected to mass spectrometry for relative quantification.

This study was undertaken in order to test the hypothesis that the technique of SELDI-TOF in conjunction with CART analysis could, using these banked serum samples, distinguish asbestosis patients who would subsequently develop lung cancer or mesothelioma from those who would not. Furthermore, we sought to determine if the SELDI-TOF-derived peaks could lead to the identification of new proteins that may be important in asbestos-related cancer pathogenesis.

Methods and materials

Study population

In 1978–1979, a cohort of 115 Finnish asbestos workers with compensable asbestosis was assembled at the Finnish Institute of Occupational Health (FIOH) in Helsinki, as described previously (Oksa et al., 1997; Li et al., 2004). These were all workers with compensable pneumoconiosis according to Finnish law (i.e. they fulfilled diagnostic criteria to be eligible for workers' compensation for their disease; Husgafvel-Pursiainen et al., 1997) who had been referred to the Institute's Occupational Medicine Clinic, the only such clinic in the country at that time, for further evaluation and follow-up of their disease. Participation in this evaluation was voluntary so not all Finnish asbestosis patients were included, and thus participants may not necessarily have been representative of all Finnish asbestos workers or asbestosis cases. The study protocol was approved by the ethics boards of the University of Illinois at Chicago and Helsinki University Hospital, with necessary permissions from the FIOH and Cancer Registry for data granted by the Ministry of Health and the National Institute for Health and Welfare. On return visits, serum samples were collected from most of the cases and stored frozen at–70°C. Cancer incidence in this group was followed up through December 31, 2007 from the Finnish Cancer Registry (Teppo et al., 1994)

For the purposes of this set of SELDI-TOF experiments, the 35 cancer cases of types most closely related to asbestos exposure (four malignant mesotheliomas and 31 lung cancers with specific histological diagnoses available for 31 of the cases; Husgafvel-Pursiainen et al., 1997) with available serum samples were selected for further study (see Table 1). This cancer subcohort consisted of 32 males and three females with an average age of 61.3 years (range 41–82 years) who were diagnosed with cancer between 1981 and 2002. This subcohort was group-matched for average age, gender, and smoking status at baseline (ever versus never) with 35 asbestosis patients who had not developed cancer during the follow-up period (see Table 1). This noncancer subcohort consisted of 32 males and three females with an average age of 62.0 years (range 45–74 years) who had either died without cancer or were still alive in 2007. In terms of other risk factors, the two groups were matched on the single most significant other cancer risk, for example, smoking status at baseline in those with lung cancer, so the differences detected are more confidently attributed to asbestos-induced carcinogenic changes. For both subcohorts, the most recent available serum sample from each patient was chosen for analysis (e.g. the sample closest, but prior, to diagnosis for the cancer cases or closest to follow-up for the noncancer controls) for a total of 70 samples. For the cancer subcohort, the average time interval between this sample collection and the subsequent diagnosis of their cancer was 3.7 years (range 1–14 years). All serum samples had been collected between 1981 and 1988 and stored frozen as 2 mL aliquots until the time of analysis.

All samples were thawed on ice and 20 μL aliquots added to 96-well plates to avoid freeze/thaw cycles. Serum was fractionated using the Serum Fractionation Kit (Bio-Rad, Hercules, CA). In brief, 30 μL of U9 Buffer (9 M urea, 2% CHAPS, 50 mM Tris–HCl pH 9) was added to serum aliquots (20 μL) and mixed for 20 min using a MicroMix 5 (DPC) on settings 20, 7. The Q ceramic HyperD F plates (Bio-Rad) were kept ready by adding 200 μL of Rehydration Buffer (50 mM Tris–HCl pH 9, 0.05% Germall® Plus [diazolidinyl urea]) to each well and mixed for 60 min using the same mixing settings as above.

Rehydration Buffer was removed from the fractionation plate using a vacuum manifold and 200 μL of Rehydration Buffer was again added to each well and removed by vacuum manifold for a total of three times. Next, the Q ceramic HyperD F wells were equilibrated by adding 200 μL of Buffer U1 (1 M urea, 0.22% CHAPS, and 50 mM Tris–HCl pH 9) and removing it by vacuum for a total of three additions. The equilibrated fractionation plate was loaded with the serum/U9 Buffer mix. Following this addition, 50 μL of Buffer U1 was added to the original sample wells and mixed with any remaining serum by repeated pipetting. This mixture was then added to the Fractionation Plate. The plate was covered with foil tape and incubated for 30 min at room temperature on the MicroMix 5 (20, 7) to facilitate protein binding to the Q ceramic HyperD F beads.

The Fractionation Plate was then placed over a 96-well plate on a vacuum manifold and Fraction #1 was pulled through the Fractionation Plate using vacuum so that the samples were directed to each well of a 96-well plate. The Fractionation Plate was returned to the MicroMix 5 and 100 μL of Wash Buffer #1 (50 mM Tris–HCl, 0.1% OGP pH 9, 0.05% Germall® Plus) added to each well. Following this addition, the plates were mixed for 10 min at the same settings as above. The Fractionation Plate was then placed over the same 96-well plate on a vacuum manifold,

Table 1. Characteristics of study subjects.

Noncancer cases				Cancer cases			
No.	Age	Sex	Status	No.	Age	Sex	Status
1N	70	M	Deceased, 1991	1C	70	M	Lung cancer, 1985
2N	65	M	Deceased, 1999	2C	58	M	Lung cancer, 1994
3N	57	M	Deceased, 1990	3C	65	F	Lung cancer, 1983
4N	60	M	Alive, 2002	4C	60	M	Lung cancer, 1994
5N	64	M	Deceased, 1998	5C	63	M	Lung cancer, 1988
6N	60	M	Deceased, 2005	6C	66	M	Lung cancer, 1984
7N	48	M	Deceased, 2005	7C	72	M	Lung cancer, 1989
8N	54	M	Deceased, 1993	8C	61	M	Lung cancer, 1987
9N	48	M	Deceased, 1989	9C	68	F	Lung cancer, 1996
10N	54	M	Deceased, 1986	10C	41	M	Lung cancer, 1985
11N	65	M	Deceased, 1990	11C	51	M	Lung cancer, 1991
12N	50	M	Deceased, 1991	12C	61	M	Mesothelioma, 1987
13N	72	M	Deceased, 2007	13C	70	M	Lung cancer, 1985
14N	50	M	Alive, 2007	14C	57	M	Lung cancer, 1987
15N	67	F	Deceased, 1999	15C	63	M	Lung cancer, 1985
16N	70	M	Deceased, 1989	16C	54	M	Lung cancer, 2002
17N	60	M	Alive, 2007	17C	55	M	Lung cancer, 1989
18N	67	M	Deceased, 1989	18C	59	M	Lung cancer, 1991
19N	73	M	Deceased, 1993	19C	60	M	Mesothelioma, 1991
20N	52	M	Alive, 2007	20C	49	M	Mesothelioma, 2002
21N	58	M	Alive, 2007	21C	57	M	Lung cancer, 1987
22N	61	M	Alive, 2007	22C	61	M	Lung cancer, 1988
23N	70	M	Deceased, 2000	23C	66	M	Lung cancer, 1988
24N	67	F	Alive, 2007	24C	53	M	Lung cancer, 1999
25N	45	M	Alive, 2007	25C	45	M	Mesothelioma, 1983
26N	60	M	Alive, 2007	26C	63	M	Lung cancer, 1981
27N	65	F	Deceased, 2005	27C	62	M	Lung cancer, 1983
28N	74	M	Deceased, 2000	28C	65	M	Lung cancer, 1984
29N	42	M	Deceased, 1995	29C	78	M	Lung cancer, 1993
30N	55	M	Alive, 2002	30C	82	M	Lung cancer, 1983
31N	70	M	Deceased, 2005	31C	68	M	Lung cancer, 1991
32N	63	M	Deceased, 1989	32C	61	M	Lung cancer, 1983
33N	59	M	Deceased, 2005	33C	61	M	Lung cancer, 1982
34N	46	M	Deceased, 1997	34C	68	F	Lung cancer, 1981
35N	62	M	Alive, 2007	35C	54	M	Lung cancer, 1982

The table contains demographic information on the subjects who donated samples for this study. Included are the ages, sex, and cancer diagnosis for each subject.

and the second part of Fraction #1 was pulled into the Fraction #1 plate using vacuum. This plate containing Fraction #1 was then sealed using foil tape and stored at–70°C. This process of elution was repeated for each of the remaining fractions utilizing increasingly acidic wash buffers for Fractions #2, #3, #4, #5, and #6.

SELDI-TOF biomarker discovery

Initial determinations of SELDI-TOF chip type, serum fraction, and wash stringency were empirically determined (see Figure 1) using seven randomly chosen samples from the total pool of serum regardless of cancer status, according to Bio-Rad protocols (Bio-Rad). Both H50 and IMAC30 copper-coated SELDI-TOF chips offered optimal binding properties for Fractions #3 and #4, respectively (see Figure 1).

IMAC30 chips were pre-incubated in several different solutions prior to the addition of the serum fraction as follows. Initial incubation of 10 min with 50 μL 0.1 M copper sulfate with mixing on a MicroMix 5 (20, 7) was followed with the copper sulfate solution being removed and the chip washed for 60 sec with 250 μL Milli-Q water (Millipore, Billerica, MA). The Milli-Q water was removed

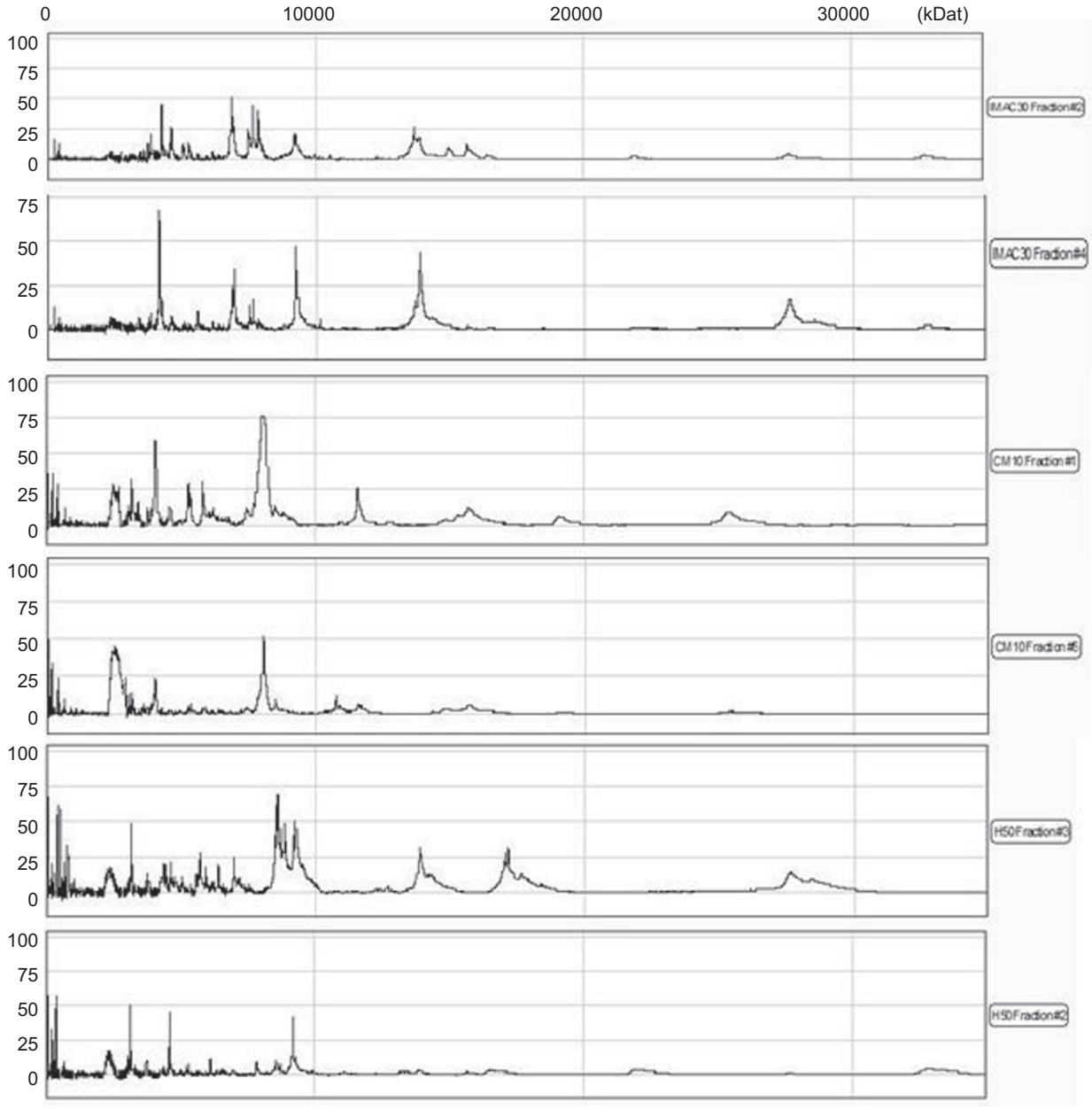

Figure 1. Examples of different Chip chemistries binding various Serum Fractions utilizing a low stringency wash and focusing on lower molecular mass proteins/peptides. On the right of the figure, you will notice the Chip chemistry (IMAC30, H50, or CM10) and the Serum Fraction.

and 250 µL 0.1 M sodium acetate pH 4 buffer added to each chip for 5-min incubation on the MicroMix 5. Following the sodium acetate solution removal, the chip was again washed twice in Milli-Q as above. Binding Buffer (low stringency: 0.1 M sodium phosphate, 0.5 M, NaCl pH 7) was then added to the chips 2× at 250 µL for 5 min of mixing on the MicroMix 5. The chip was then ready for the addition of the serum fraction. H50 chips were incubated in 50 µL of 50% methanol for 5 min at room temperature. The methanol was removed and the incubation repeated. The second methanol incubation was removed, and the chips allowed to air-dry at least 15 min. Following air-drying, 200 µL of binding buffer (low stringency: 10% acetonitrile [ACN], 0.1% trifluoroacetic acid [TFA], high stringency: 50% methanol, 0.1% TFA) was added to the chips and allowed to incubate with shaking on the MicroMix 5 twice for 5 min at room temperature. The chip was then ready for the addition of the serum fraction.

In a blinded fashion, duplicate serum fractions (45 µL) were added along with 100 µL of binding buffer to specific spots on the chips (Fraction #4 to IMAC30 and Fraction #3 to H50) for 60-min incubations at room temperature with shaking on the MicroMix 5. Following incubation, all

chip types were washed 3× with 250 µL of binding buffer. IMAC30 chips were then allowed to dry for 15–20 min. H50 chips were further rinsed 2× with 250 µL of Milli-Q water and then allowed to dry for 5–10 min. When chips were dry, 1 µL of the energy-absorbing molecule (EAM) α-cyano-4-hydroxycinnamic acid (CHCA) dissolved in a 50:50 1% TFA and ACN (v/v) was added to each chip spot and allowed to dry for 5 min. This EAM addition was then repeated. The Bio-Rad All-In-1 Protein Standard II (Bio-Rad) was suspended in the EAM and added to spot "H" on each chip. Chips were held in the dark at room temperature until read on the SELDI-TOF.

All chips were scanned on a Protein Biosystems IIc (PBS IIc) instrument that was calibrated to focus on proteins and polypeptides from 3000 to 40,000 Da (empirically determined as recommended by the manufacturer). Peak sizes were standardized using the All-in-1 Protein Standard II, and peaks were grouped into clusters using the autodetection function of the Bio-Rad ProteinChip Software (Bio-Rad).

Statistics and data analysis

Classification and regression tree analysis (CART) was used to identify a possible panel of biomarkers (Breiman, 1984) derived from peaks observed in the SELDI-TOF data. The target (dependent) variable was group status (cancer or noncancer). All peak intensities were considered as predictor (independent) variables. Since the target variable is categorical, classification trees were used. For building our classification trees, duplicates were considered individually as equal contributors to the CART analysis and we assumed equal prior on the target variable and no penalties on predictor variables. For selecting the optimal tree, we considered the tree with the minimum cost (lowest misclassification rate) and the best prediction rates (sensitivity and specificity) via 10-fold cross-validation. Different splitting functions were tested, but the Gini classification method with no preference on even splits yielded the best cost benefit. We built trees with all the peak intensities as predictor variables with and without using automatic best predictor discovery where best predictor discovery selected different predictors based on predictor importance. Receiver-operating characteristic (ROC) curves were also used to evaluate the results of a given prediction. Mean peak intensities of each member of the biomarker panel are presented in Table 2 along with a repeated measures analysis of variance (ANOVA) to compare intensities between the two groups.

Isolation and identification of serum biomarkers

Isolation of biomarkers discovered utilizing SELDI-TOF was performed as recommended by the manufacturer (Bio-Rad). For each biomarker, serum was fractionated utilizing the serum fractionation protocol above with some modifications. In brief, 150 µL of U9 Buffer was added to 100 µL of serum and mixed by vortex (setting 4.5) for 20 min at room temperature. After mixing, 250 µL of U1 Buffer was added to each serum mixture. While the serum mixtures were mixing, protein-binding beads (Q ceramic HyperD F) were suspended in 200 µL of Rehydration Buffer and incubated for 10 min at room temperature and then placed into Pierce Spin Columns (Pierce, Rockford, IL). Spin Columns were centrifuged at 1000 *g* to remove buffer, and the beads were then equilibrated in 500 µL of U1 Buffer by mixing for 10 min (vortex set at 4.5) at room temperature.

Serum mixtures were next added to the equilibrated beads in the Spin Columns and mixed at room temperature for 30 min by vortex (setting 4.5). Samples were next set into collection vessels and microfuged at 1000 *g* to collect the flow-through (FT) fraction. Wash Buffer #1 (400 µL) was added to each column and mixed for 10 min by vortex (setting 4.5). Columns were then placed into a microfuge and the FT collected as Fraction #1. This process of collection was repeated for each of the remaining fractions utilizing increasingly acidic wash buffers for Fractions #2, #3, #4, #5, and #6. The location of the greatest concentration of putative biomarker was then assayed by SELDI-TOF.

NP20 chips (normal phase chemistry) were incubated in 250 µL of Milli-Q H_2O for 10 min and then the water was discarded. Aliquots of each fraction (45 µL) were added to spots on the NP20 chips with 155 µL of Milli-Q H_2O and incubated for 60 min with mixing on a MicroMix 5. Solutions were removed from NP20 chips and the chips were allowed to air-dry for 10–15 min. EAM and standards were added as above and the chips read on a SELDI-TOF to determine which fraction contained the biomarkers of interest. SELDI-TOF-derived spectra were standardized using the All-in-1 Protein Standard II and protein peaks were grouped using the saved clusters derived above.

Fractions containing the biomarkers of interest were pooled (to increase the overall amount of biomarker) and subjected to hydrophobic fractionation to reduce the complexity of the protein sample. ANC and TFA were added to each pooled sample for final concentrations of 10% and 0.5%, respectively. RPC Poly-Bio Beads (Biosepra, Marlborough, MA) were equilibrated with 500 µL of 10% ACN/0.1% TFA for 30 min at room temperature. Following equilibration, beads were added to each pooled biomarker solution and mixed on a tilt table for 30 min at room temperature. Samples were added to Pierce Spin Columns and microfuged at 2000 *g* for 60 sec. FT was collected for each biomarker and labeled FT. An elution solution of 10% ACN/0.1% TFA was added to each column (400 µL) and mixed by vortex for 10 min. Columns were then centrifuged at 2000 *g* for

Table 2. IMAC30/Fraction #4-derived protein biomarkers.

		Cancer		Noncancer		Repeated ANOVA
Mass (Da)	Chip/Fraction	Mean	SD	Mean	SD	*P*-value
5707.01	IMAC30/4	0.52	0.65	1.50	3.47	0.0772
6598.10	IMAC30/4	2.95	1.89	2.41	1.54	0.2699
20,780.70	IMAC30/4	0.72	0.38	0.90	0.39	0.0227
3283.32	H50/3	7.05	4.52	6.15	3.65	0.2572
5695.31	H50/3	10.16	3.69	12.35	7.60	0.0591
6364.71	H50/3	16.59	7.79	15.22	7.35	0.3010
6559.69	H50/3	18.13	10.28	15.63	8.02	0.1601
39,037.30	H50/3	0.04	0.03	0.06	0.06	0.0505

The table contains the masses and relative abundance of biomarkers in both sample groups, where the biomarkers were found (Chip/Fraction) and the repeated ANOVA *P*-values for each biomarker. These biomarkers were used to generate the classification and regression trees (Figures 2 and 3).

60 sec to collect FT. This FT fraction was then labeled as 10%. These steps were repeated for increasing concentrations of ACN (20%, 30%, 40%, 50%, or 70%). To determine which fractions held the putative biomarkers, NP20 chips were employed as described above.

Hydrophobic fractions containing putative biomarkers were pooled and then dried by vacuum centrifugation. Proteins and/or polypeptides were suspended in 50 µL of 2× gel-loading buffer (100 mM Tris–HCl pH 6.8, 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, and 200 mM 2-mercaptoethanol) and heated at 95°C for 5 min to denature the proteins and/or polypeptides. Protein and/or polypeptides were separated utilizing PAGE through a 4%/10%/16% Tricine/SDS/polyacrylamide gel (Schägger, 2006) at 40 V for 24 h at room temperature. Following electrophoresis, the gel was stained using Gel Code Blue Safe Protein Stain (Pierce) and bands near the size of the putative biomarker excised from the gel using 1.2 mm Harris Unicore punches (Ted Pella, Redding, CA).

Gel punches were incubated two times for 30 min in 200 µL of 50% methanol/10% acetic acid with shaking to remove protein stain. Solvents were removed, and 100 µL of ACN was added to each tube for 15-min incubations with shaking to dehydrate the gel sections. The ACN was then removed and the punches were incubated for 2 h in 70 µL of 50% formic acid, 25% ACN, and 15% isopropanol with vigorous shaking to elute the proteins. Supernatants were then transferred to a new tube, and 1 µL of this supernatant added to a spot on an NP20 chip. EAM and standards were added to the NP20 chips as described above, and the chips were read on a SELDI-TOF to determine if the eluted band held the biomarkers of interest. SELDI-TOF-derived spectra were standardized using the All-in-1 Protein Standard II and peaks were grouped using the saved clusters derived above.

Following SELDI-TOF confirmation of isolated biomarkers' size (i.e. the size was confirmed to be the size of the putative biomarker) biomarkers were vacuum-centrifuged to dryness and suspended in 20 µL of 50 mM ammonium bicarbonate pH 8. Biomarkers were digested by adding Trypsin Gold (Promega, Madison, WI) for a final concentration of 100 ng/µL and incubated for 3 h at 37°C. Peak patterns were evaluated by adding 5 µL of the protein digest to an NP20 chip as described above and visualizing the spectra from 1 to 3 kDa on the SELDI-TOF in single MS mode. Saved spectra were then evaluated on Protein Prospector (UCSF) using the MS-Fit function. If a biomarker's identity could not be elucidated utilizing Protein Prospector, the spectra was subjected to the MASCOT Peptide Mass Fingerprint program (Matrix Direct, San Diego, CA). This program has the ability to match protein sequences with uncharacterized proteins from open-reading frames (ORFs) in genomic DNA.

Results

Biomarker discovery by SELDI-TOF

In this cohort study design, SELDI-TOF data analyzed by CART identified two separate clusters of serum protein/polypeptide peaks that yield high sensitivity and specificity when distinguishing asbestosis cases that developed cancer from those that did not. Classification trees were generated using data from Serum Fraction #4 and IMAC30 copper-coated chips or Serum Fraction #3 and H50 SELDI chips to maximize the sensitivity and specificity while minimizing the Cost statistic, which leads to enhanced specificity and sensitivity. Tree #1 used three protein peaks (5701.01, 6598.10, and 20,780.7 Da) to differentiate between cancer and noncancer (see Figure 2 and Table 2) with a sensitivity of 87% and a specificity of 70% with a 10-fold cross-validation of 81.4% and 64.3%, respectively. The second classification tree was generated from data obtained by incubation of Serum Fraction #3 and H50 SEDLI chips. This tree utilizes five separate protein peaks (3283.32, 5695.31, 6364.71, 6559.69, and 39,037.3 Da) to differentiate between cancer and noncancer within the asbestosis cohort (see Figure 3 and Table 2). The tree yields prediction rates of 83% sensitivity

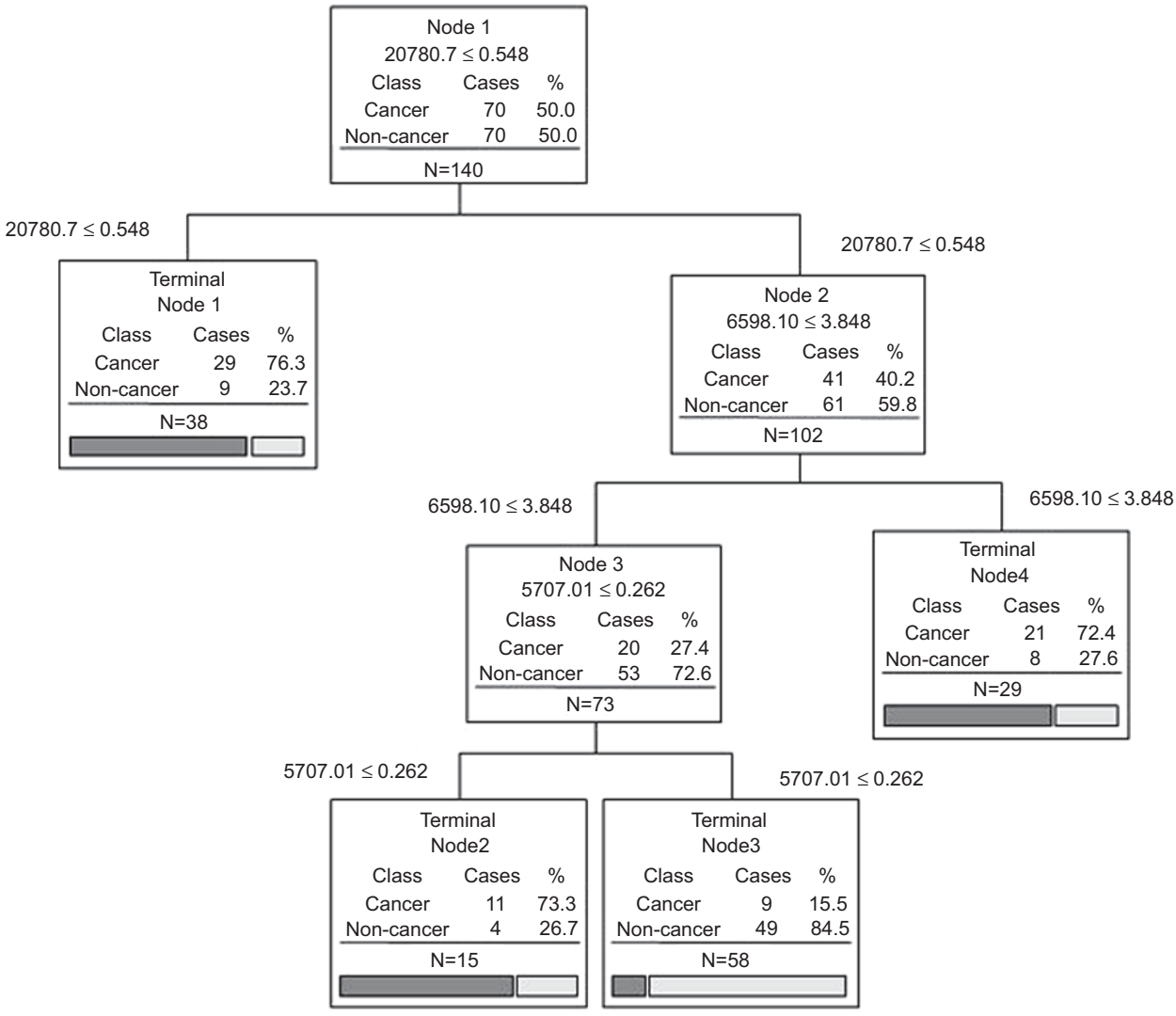

Figure 2. Best classification tree using IMAC30/Fraction 4 data. Three peaks were used to construct the tree (5707.01, 6598.10, and 20,780.7 Da). Each node is sequentially labeled and shows splitting criteria. For instance, 20,780.7 £ 0.548 would mean that subjects with peak intensities of £0.548 at m/z ratio at 20,780.7 Da would move down the left side and all other subjects would move down the right side. Subjects continue down the tree until they reach terminal nodes. A terminal node is classified as cancer if the majority of samples in the terminal node are from cancer subjects (denoted by the dark gray bar being larger than the light gray bar). Otherwise, the terminal node is classified as noncancer (denoted by the light gray bar being larger than the dark gray bar).

and 80% specificity with a 10-fold cross-validation of 64.3% and 61.4%.

Biomarker identification

Utilizing protein enrichment and isolation strategies described above, we were able to isolate five of the eight putative biomarkers that compose Tree #1 and #2 (see Figures 2 and 3). However, we were only able to definitively identify three. Utilizing Protein Prospector, biomarkers 5701.01 and 6598.10 Da (IMAC/Fraction #4) were identified as Kinesin-like protein 18A (KIF18A) and Kinesin heavy chain isoform 5A (KIF5A), respectively (see Table 3). Protein Prospector does not utilize *P*-values but instead uses MOWSE Score (MOlecular Weight SEarch), which is not a true statistical value (Pappin et al., 1993). However, in conjunction with percent coverage, percent total ion chromatography (TIC) and percent of peaks matched can be used to infer relevance with the higher the MOWSE Score the better the protein match (see Table 3). The KIF18A protein identification has a MOWSE Score of 9.48E+13 with 55.1% coverage, 74.2% TIC, and 71% peaks matched, whereas the KIF5A protein identification has a MOWSE of 1.82E+14 with 68.4% coverage, 74.2% TIC, and 74% peaks matched. The H50/Fraction #3 biomarker at 6559.69 Da did not yield a protein identification utilizing Protein Prospector and was subjected to the

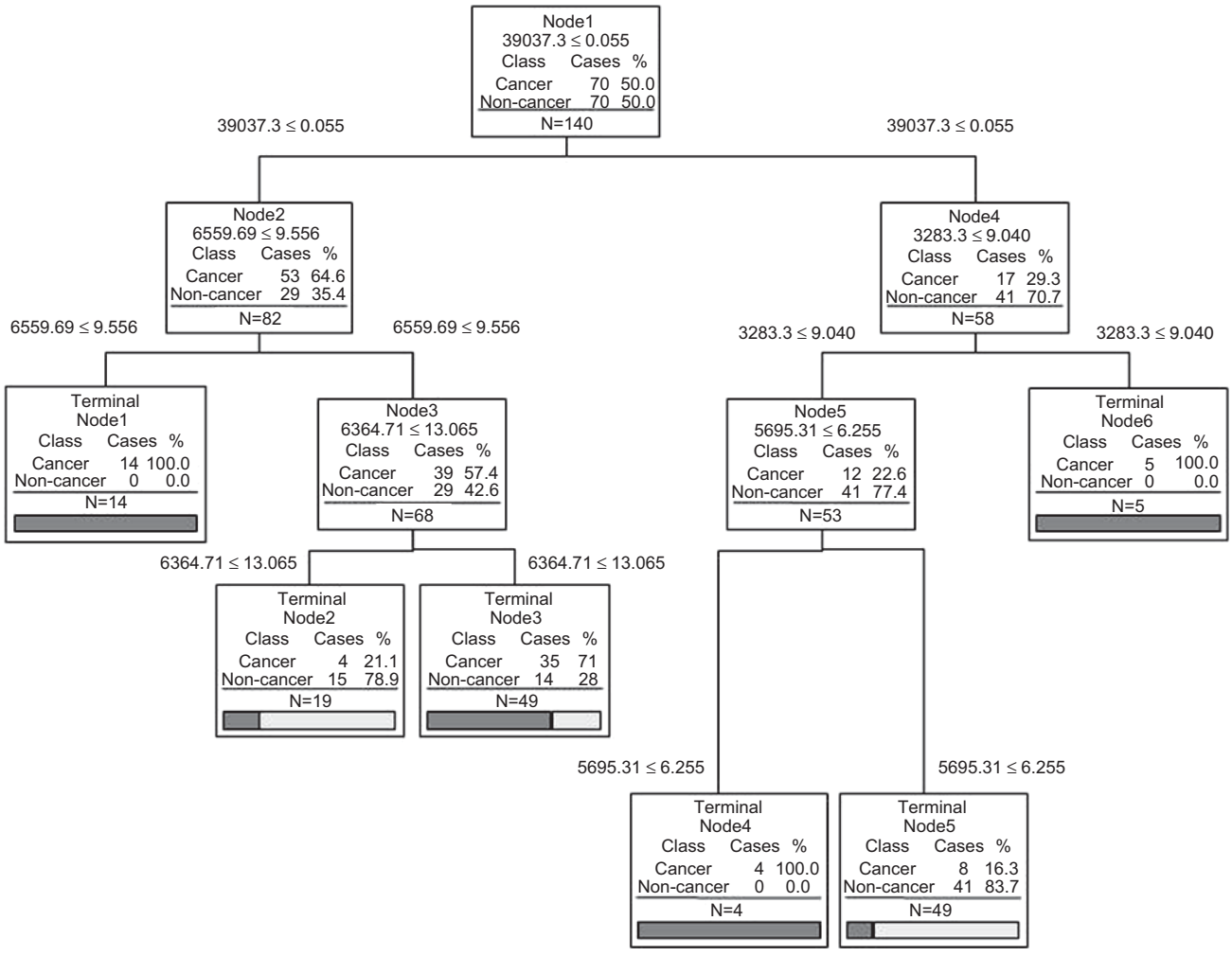

Figure 3. Best classification tree using H50/Fraction 3 data. There are five peaks, which were used to construct the tree (3283.32, 5695.31, 6364.71, 6559.69, and 39,037.3 Da). Each node is sequentially labeled and shows splitting criteria. For instance, 39,039.3 £ 0.055 would mean that subjects with peak intensities of £0.055 at m/z ratio at 39,037.3 Da would move down the left side and all other subjects would move down the right side. Subjects continue down the tree until they reach terminal nodes. A terminal node is classified as cancer if the majority of samples in the terminal node are from cancer subjects (denoted by the dark gray bar being larger than the light gray bar). Otherwise, the terminal node is classified as noncancer (denoted by the light gray bar being larger than the dark gray bar).

MASCOT Search utilizing Matrix Science software that returned an identification for the biomarker as a putative uncharacterized protein C12orf37 (unknown protein on chromosome 12 in ORF 37) with a level of confidence of $P < 0.02$.

Discussion

The goal of this study was 2-fold. First, we sought to utilize SELDI-TOF to distinguish asbestosis patients who have developed cancer from those who have not, and, second, we hoped to identify new and interesting proteins involved in cancer development in asbestosis. We have shown that SELDI-TOF can indeed distinguish between samples from asbestosis patients who developed cancer and patients who did not develop cancer, with specificities of 87% or 70% and sensitivities of 83% or 80%, utilizing serum biomarkers from Tree #1 or #2, respectively (see Figures 2 and 3). It is interesting to note that even though these two biomarker trees are able to distinguish cancer from noncancer with good sensitivities and specificities, other biomarker trees could be generated utilizing different Chip/Serum Fraction combinations. Utilizing the Chip/Serum Fraction combinations presented here, Trees #1 and #2 best-distinguish cancer from noncancer within this asbestosis cohort.

The nature of SELDI-TOF experimentation is to simplify the sample complexity by pH and binding characteristics without regard to previously identified proteins. It is possible and even likely that previously identified protein biomarkers for cancer were not utilized for either biomarker tree because they were excluded by either pH or chip-binding chemistries. Also, since this is a predictive

Table 3. Biomarker identification.

Size (Da)	Chip	ID	MOWSE Score	% Coverage	% TIC	% Peak match	*P*-value
5707.01	IMAC30	KIF18	9.48E+13	55.1	74.2	71.0	NA
6598.10	IMAC30	KIF5A	1.82E+14	68.4	74.2	74.0	NA
6364.71	H50	Putative protein C12orf37	NA	NA	NA	NA	0.02

Utilizing Protein Prospector, MOWSE Scores were generated for the biomarkers that were identified as KIF18A and KIF5A. Additional scoring criteria of % Coverage, % TIC and % Peak Match lend further weight to the relevance of the identifications. The remaining biomarker was not identified by utilizing Protein Prospector but by utilizing MASCOT. This identification has a $P < 0.02$.

model there is the possibility that previously discovered proteins important to asbestosis-derived cancer would not improve the predictive power of the model if they were included. Perhaps, the more interesting result of this study was the identification of two separate Kinesin Superfamily proteins as biomarkers for cancer development in asbestosis.

Kinesins are motor proteins that run along microtubules to transport "packages" within cells. These "packages" can be other proteins, vesicles, organelles, or chromosomes (Miki et al., 2005). Of the 45 identified human kinesins, 12 have been shown to be essential in mitosis (Miki et al., 2005; Zhu et al., 2005) and some have been implicated in cancer (Huszar et al., 2009). KIF18A has been shown to be essential in the accurate alignment of the spindle equator (Mayr et al., 2007), which is critical for cellular mitosis. In fact, KIF18A has been shown to suppress kinetochore movements to control mitotic chromosome alignment (Stumpff et al., 2008). In this study, we found KIF18A protein at lower concentrations in the serum of asbestosis patients who developed cancer when compared with asbestosis patients who had not developed cancer (see Tables 2 and 3). This lower concentration of KIF18A in the serum of asbestosis patients who developed cancer may be indicative of alterations in cellular mitosis that are often observed in various cancers. On the other hand, the neuron-specific kinesin KIF5A has not previously been shown to be related to cancer. This kinesin has been associated with a modest increase in rheumatoid arthritis (RA) susceptibility (Li and Begovich, 2009) or to be weakly protective in juvenile idiopathic arthritis (JIA) (Hinks et al., 2009), and mutations in the motor domain have been shown to cause heredity spastic paraplegia (Ebbing et al., 2008). Nevertheless, a possible relationship to cancer is plausible, since the KIF5A protein does show greater than 69% protein homology to KIF5B, and KIF5B has been shown to be integral to the survival of three cancer-derived cell lines: HeLa, MCF-7, and U2OS (Cardoso et al., 2009). Finally, it should be noted that alterations in kinesins in asbestos-induced carcinogenesis are also quite possible. Asbestos has been long recognized to be able to induce significant mitotic aberrations leading to chromosomal instability that is associated with cancer (Hesterberg and Barrett, 1985). Furthermore, in cell culture experiments, asbestos has been shown to induce these chromosomal effects by binding to proteins that regulate the cell cycle, cytoskeleton, and the mitotic process (MacCorkle et al., 2006). Although KIF18A, KIF5A, or Kinesin Superfamily proteins have not been identified as specific targets in these experiments, it is certainly plausible that they would be secondarily affected by any alterations induced by asbestos exposure produced in the proteins that regulate the mitotic spindle.

Possible limitations of this study should be considered. For example, collection and processing bias are issues in any experiment in which samples are collected and processed over a significant period of time (Marshall et al., 2003; Banks et al., 2005). Our serum samples were collected at one location (FIOH in Helsinki) utilizing a standardized protocol that serves to minimize collection bias. Although it is not possible to eliminate all variation when collecting samples over any period of time, we believe that the utilization of a standardized protocol and a single site for serum collection have reduced sample-to-sample variation to the extent possible in this type of study. Furthermore, in the laboratory analysis serum sample processing and SELDI-TOF methodologies were performed by a single technician utilizing standard protocols. Samples were utilized for SELDI-TOF analysis within days of preparation to reduce variation due to time (Rogers et al., 2003). Although it is true that samples were collected many years prior to analysis, our prior studies have suggested that other protein biomarkers in these samples are quite stable over these extended periods of time under these storage conditions. Thus, we are reasonably confident that biomarker deterioration is not a major consideration. In addition, the samples from the cancer cases and the controls were of the same approximate age, so any minor changes that could have occurred over time might be expected to affect them both in a similar fashion.

Even with standardized protocols to minimize variation, there are protein-binding issues that are inherent with SELDI-TOF. Serum contains upwards of thousands of proteins that are not expressed in equal amounts. In fact, 10 most abundant serum proteins account for ~98% of the total serum protein content (Zolg and Langen, 2004). The abundance of these highly expressed proteins can make it difficult to find meaningful protein biomarkers due to the highly abundant proteins utilizing the binding area on the SELDI-TOF chips (Aebersold and Mann,

2003; Diamandis, 2004; Gillette et al., 2005). To address this issue, we reduced the complexity of the total serum sample by performing a pH-based fractionation of the serum proteins and by using various binding buffers and the specific binding chemistries of the different SELDI-TOF chips to selectively bind proteins with high affinity for the specific binding chip chemistries. Furthermore, we bound each serum fraction to each chip type to determine empirically, which Chip/Serum Fraction combinations maximized protein binding within our target region (3–40 kDa) without crowding the peaks until they were indistinguishable from one another.

One other issue with biomarker identification experiments does not involve SELDI-TOF directly but is a concern over utilizing incorrect or inappropriate referrent groups (Diamandis, 2004; Koomen et al., 2005; Villar-Garea et al., 2007). Since our "disease group" in this case is a progression of those in the asbestosis group to development of cancer and is otherwise similar in terms of age, gender, ethnicity (all Finnish), socioeconomic status (all blue-collar workers), and other risk factors, we feel that this issue with biomarker identification has been largely negated. In addition, as also noted, the generalizability of these results to other asbestos workers is uncertain, since this is a relatively small sample and this cohort is a relatively select group of asbestosis cases that may not be representative of exposed workers as a whole in Finland or elsewhere.

Overall, we believe the limitations of this study to be minimal and the results to be highly promising. These results suggest that it is feasible to use a proteomic approach based on SELDI-TOF and CART to distinguish, among a cohort of patients at high risk for the development of cancer such as these asbestosis cases, those individuals who will subsequently develop cancer from those who will not with a high degree of sensitivity and specificity. Furthermore, several of the specific biomarkers identified by this approach may provide new insights into the critical molecular pathways of asbestos-induced carcinogenesis.

Acknowledgements

The authors would like to thank the patients who have participated in our research.

Declaration of interest

This work was supported in part by grants from the National Institute for Occupational Safety and Health (R01-OH04192, R01-OH07590) to PWB-R and the National Institute of Environmental and Health Sciences (2 P01 ES011810) to LSN. Dr. Brandt-Rauf is currently on the editorial board of *Biomarkers*.

References

Aebersold R, Mann M. (2003). Mass spectrometry-based proteomics. *Nature* 422:198–207.

Banks RE, Stanley AJ, Cairns DA, Barrett JH, Clarke P, Thompson D, Selby PJ. (2005). Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. *Clin Chem* 51:1637–1649.

Breiman L, Friedman JH, Olshen RA, Stone CJ. (1984). *Classification and Regression Trees*. New York: Chapman & Hall.

Cappello F, Di Stefano A, D'Anna SE, Donner CF, Zummo G. (2005). Immunopositivity of heat shock protein 60 as a biomarker of bronchial carcinogenesis. *Lancet Oncol* 6:816.

Cardoso CM, Groth-Pedersen L, Høyer-Hansen M, Kirkegaard T, Corcelle E, Andersen JS, Jäättelä M, Nylandsted J. (2009). Depletion of kinesin 5B affects lysosomal distribution and stability and induces peri-nuclear accumulation of autophagosomes in cancer cells. *PLoS ONE* 4:e4424.

Diamandis EP. (2004). Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J Natl Cancer Inst* 96:353–356.

Ebbing B, Mann K, Starosta A, Jaud J, Schöls L, Schüle R, Woehlke G. (2008). Effect of spastic paraplegia mutations in KIF5A kinesin on transport activity. *Hum Mol Genet* 17:1245–1252.

Gillette MA, Mani DR, Carr SA. (2005). Place of pattern in proteomic biomarker discovery. *J Proteome Res* 4:1143–1154.

Hesterberg TW, Barrett JC. (1985). Induction by asbestos fibers of anaphase abnormalities: mechanism for aneuploidy induction and possibly carcinogenesis. *Carcinogenesis* 6:473–475.

Hinks A, Eyre S, Ke X, Barton A, Martin P, Flynn E, Packham J, Worthington J, Thomson W; Childhood Arthritis Prospective Study (CAPS); UKRAG Consortium; BSPAR Study Group. (2009). Overlap of disease susceptibility loci for rheumatoid arthritis and juvenile idiopathic arthritis. *Ann Rheum Dis* 69:1049–1053.

Husgafvel-Pursiainen K, Kannio A, Oksa P, Suitiala T, Koskinen H, Partanen R, Hemminki K, Smith S, Rosenstock-Leibu R, Brandt-Rauf PW. (1997). Mutations, tissue accumulations, and serum levels of p53 in patients with occupational cancers from asbestos and silica exposure. *Environ Mol Mutagen* 30:224–230.

Huszar D, Theoclitou ME, Skolnik J, Herbst R. (2009). Kinesin motor proteins as targets for cancer therapy. *Cancer Metastasis Rev* 28:197–208.

Koomen JM, Shih LN, Coombes KR, Li D, Xiao LC, Fidler IJ, Abbruzzese JL, Kobayashi R. (2005). Plasma protein profiling for diagnosis of pancreatic cancer reveals the presence of host response proteins. *Clin Cancer Res* 11:1110–1118.

Lea JS, Coleman R, Kurien A, Schorge JO, Miller DS, Minna JD, Muller CY. (2004). Aberrant p16 methylation is a biomarker for tobacco exposure in cervical squamous cell carcinogenesis. *Am J Obstet Gynecol* 190:674–679.

Li L, Sun TD, Zhang X, Lai RN, Li XY, Fan XJ, Morinaga K. (2004). Cohort studies on cancer mortality among workers exposed only to chrysotile asbestos: a meta-analysis. *Biomed Environ Sci* 17:459–468.

Li Y, Begovich AB. (2009). Unraveling the genetics of complex diseases: susceptibility genes for rheumatoid arthritis and psoriasis. *Semin Immunol* 21:318–327.

MacCorkle RA, Slattery SD, Nash DR, Brinkley BR. (2006). Intracellular protein binding to asbestos induces aneuploidy in human lung fibroblasts. *Cell Motil Cytoskeleton* 63:646–657.

Marshall J, Kupchak P, Zhu W, Yantha J, Vrees T, Furesz S, Jacks K, Smith C, Kireeva I, Zhang R, Takahashi M, Stanton E, Jackowski G. (2003). Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. *J Proteome Res* 2:361–372.

Mayr MI, Hümmer S, Bormann J, Grüner T, Adio S, Woehlke G, Mayer TU. (2007). The human kinesin Kif18A is a motile microtubule depolymerase essential for chromosome congression. *Curr Biol* 17:488–498.

Miki H, Okada Y, Hirokawa N. (2005). Analysis of the kinesin superfamily: insights into structure and function. *Trends Cell Biol* 15:467–476.

Mori H, Yamada Y, Hirose Y, Kuno T, Katayama M, Sakata K, Yoshida K, Sugie S, Hara A, Yoshimi N. (2002). Chemoprevention of large bowel carcinogenesis; the role of control of cell proliferation and significance of beta-catenin-accumulated crypts as a new biomarker. *Eur J Cancer Prev* 11 (Suppl 2):S71–S75.

Oksa P, Pukkala E, Karjalainen A, Ojajärvi A, Huuskonen MS. (1997). Cancer incidence and mortality among Finnish asbestos sprayers and in asbestosis and silicosis patients. *Am J Ind Med* 31:693–698.

Pappin DJ, Hojrup P, Bleasby AJ. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* 3:327–332.

Reid A, de Klerk N, Ambrosini GL, Olsen N, Pang SC, Berry G, Musk AW. (2005). The effect of asbestosis on lung cancer risk beyond the dose related effect of asbestos alone. *Occup Environ Med* 62:885–889.

Rogers MA, Clarke P, Noble J, Munro NP, Paul A, Selby PJ, Banks RE. (2003). Proteomic profiling of urinary proteins in renal cancer by surface enhanced laser desorption ionization and neural-network analysis: identification of key issues affecting potential clinical utility. *Cancer Res* 63:6971–6983.

Schägger H. (2006). Tricine-SDS-PAGE. *Nat Protoc* 1:16–22.

Shamma A, Doki Y, Shiozaki H, Tsujinaka T, Yamamoto M, Inoue M, Yano M, Monden M. (2000). Cyclin D1 overexpression in esophageal dysplasia: a possible biomarker for carcinogenesis of esophageal squamous cell carcinoma. *Int J Oncol* 16:261–266.

Stumpff J, von Dassow G, Wagenbach M, Asbury C, Wordeman L. (2008). The kinesin-8 motor Kif18A suppresses kinetochore movements to control mitotic chromosome alignment. *Dev Cell* 14:252–262.

Teppo L, Pukkala E, Lehtonen M. (1994). Data quality and quality control of a population-based cancer registry. Experience in Finland. *Acta Oncol* 33:365–369.

Villar-Garea A, Griese M, Imhof A. (2007). Biomarker discovery from body fluids using mass spectrometry. J *Chromatogr B Analyt Technol Biomed Life Sci* 849:105–114.

Winer I, Normolle DP, Shureiqi I, Sondak VK, Johnson T, Su L, Brenner DE. (2002). Expression of 12-lipoxygenase as a biomarker for melanoma carcinogenesis. *Melanoma Res* 12:429–434.

Zhu C, Zhao J, Bibikova M, Leverson JD, Bossy-Wetzel E, Fan JB, Abraham RT, Jiang W. (2005). Functional analysis of human microtubule-based motor proteins, the kinesins and dyneins, in mitosis/cytokinesis using RNA interference. *Mol Biol Cell* 16:3187–3199.

Zolg JW, Langen H. (2004). How industry is approaching the search for new diagnostic markers and biomarkers. *Mol Cell Proteomics* 3:345–354.