<https://helda.helsinki.fi>

Reactive oxygen species-regulating proteins peroxiredoxin 2 and thioredoxin, and glyceraldehyde-3-phosphate dehydrogenase are differentially abundant in induced sputum from smokers with lung cancer or asbestos exposure.

Rostila, Annina M.

2020-05

Rostila , A M , Anttila , S L , Lalowski , M M , Vuopala , K S , Toljamo , T I , Lindström , I , Baumann , M H & Puustinen , A 2020 , ' Reactive oxygen species-regulating proteins peroxiredoxin 2 and thioredoxin, and glyceraldehyde-3-phosphate dehydrogenase are differentially abundant in induced sputum from smokers with lung cancer or asbestos exposure. ' , European Journal of Cancer Prevention , vol. 29 , no. 3 , pp. 238-247 . <https://doi.org/10.1097/CEJ.0000000000000537>

<http://hdl.handle.net/10138/321380>

<https://doi.org/10.1097/CEJ.0000000000000537>

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

**ROS-regulating proteins
peroxiredoxin 2 and thioredoxin, and
glyceraldehyde-3-phosphate dehydrogenase are
differentially abundant in induced sputum from
smokers with lung cancer or asbestos exposure**

Running head: Sputum biomarkers in risk of lung cancer

**Annina M. Rostila^{1,2}, Sisko L. Anttila³, Maciej M. Lalowski¹, Katri S. Vuopala⁴,
Tuula I. Toljamo⁴, Irmeli Lindström⁵, Marc H. Baumann¹, Anne M. Puustinen⁶.**

**1 Helsinki Institute for Life Science (HiLIFE) and Faculty of Medicine,
Biochemistry and Developmental Biology, Meilahti Clinical Proteomics Core
Facility, University of Helsinki, Helsinki, Finland;**

**2 Helsinki Normal Lyceum, Faculty of Educational Sciences, University of Helsinki;
Helsinki, Finland;**

**3 Department of Pathology, HUSLAB and Helsinki University Hospital, Helsinki
and Espoo, Finland;**

4 Departments of Pathology and Pulmonary Diseases, Lapland Central Hospital, Rovaniemi, Finland;

5 Finnish Institute of Occupational Health (FIOH), Helsinki, Finland;

6 Finnish Institute for Verification of the Chemical Weapons Convention (VERIFIN), Department of Chemistry, University of Helsinki, Helsinki, Finland.

Corresponding author: Marc Baumann

Phone: +358505671980; Email: marc.baumann@helsinki.fi

Conflicts of Interest:

None declared.

Abstract

Lung cancer is a deadly disease, typically caused by known risk factors, such as tobacco smoke, and asbestos exposure. By triggering cellular oxidative stress, and altering the antioxidant pathways eliminating reactive oxygen species (ROS), tobacco smoke and asbestos predispose to cancer. Despite easily recognizable high-risk individuals, lung cancer screening, and its early detection are hampered by poor diagnostic tools including the absence of proper biomarkers. This study aimed to recognize potential lung cancer biomarkers using induced sputum noninvasively collected from the lungs of individuals in risk of contracting lung cancer. Study groups comprised of current and former smokers, who either were significantly asbestos exposed, had lung cancer, or were unexposed and asymptomatic. Screening of potential biomarkers was performed with 52 , and five differentially abundant proteins, peroxiredoxin 2 (PRDX2), thioredoxin (TXN), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), extracellular matrix protein 1 (ECM1), and protein S100 A8 (S100A8), were chosen to undergo validation, for their previously known connection with oxidative stress or cancer. Results from the validation in 123 sputa showed that PRDX2, TXN and GAPDH were differentially abundant in sputa from individuals with lung cancer. TXN had a negative correlation with asbestos exposure, yet a positive correlation with smoking and lung cancer. Thus, tobacco smoking, asbestos exposure, and lung carcinogenesis may disturb the cellular redox state in different ways. A strong correlation was found between PRDX2, TXN, GAPDH and S100A8, suggesting that these proteins may present a diagnostic biomarker panel to aid recognizing individuals at high risk of contracting lung cancer.

Keywords: biomarker, peroxiredoxin 2, thioredoxin, glyceraldehyde-3-phosphatase, induced sputum, asbestos, lung cancer, tobacco smoking, reactive oxygen species

Introduction

Lung cancer, the leading cause of cancer-related deaths worldwide (Jemal *et al.*, 2011), results mostly from tobacco smoking (Ezzati *et al.*, 2005). A small fraction of 5-7 % of lung cancers is caused by exposure to asbestos (Tossavainen, 2004), albeit asbestos accounts for more than half of all occupational cancers (Rushton *et al.*, 2008). Asbestos fibers and carcinogens in tobacco smoke add synergistically to the lung cancer risk (reviewed in Nielsen *et al.*, 2014). Although lung cancer carries known risk factors, it is difficult to diagnose at an early stage, resulting in poor prognosis and survival, despite some advances in computed tomography (CT)-screening of high-risk individuals (Aberle *et al.*, 2011). Thus, biomarkers for early detection of lung cancer are greatly needed.

Reactive oxygen species (ROS), such as hydrogen peroxide (H_2O_2), are normal byproducts of cellular metabolism, and play a key role in signal transduction, cell differentiation and proliferation (reviewed in Lennicke *et al.*, 2015). The amount of ROS in cells is tightly regulated by endogenous antioxidants, such as peroxiredoxin, and thioredoxin, both under regulation of the transcription factor nuclear-factor-erythroid 2 p45-related factor 2 (NRF2). Under normal circumstances, these antioxidant molecules maintain the cellular homeostasis of H_2O_2 reducing it into water and oxygen, thus keeping the cellular concentration of H_2O_2 moderately low. This is in part due to negative regulation of *NRF2* by the kelch-like ECH-associated protein (KEAP1), which, in basal conditions, binds to NRF2 marking it for degradation, and thus limits the potential of NRF2 to activate antioxidant response genes in the nucleus (Menegon *et al.*, 2016). Under oxidative stress, the KEAP1-NRF2 signaling pathway is impaired by inactivation of

KEAP1, leading to NRF2-induced stress response, and fight against free oxygen radicals, such as the hydroxyl radical formed from H₂O₂.

While *NRF2* activation, and antioxidant response are beneficial to normal cells, in cancer cells constant *NRF2* overexpression provides a growth advantage. The antiapoptotic Bcl-2 family proteins are under the regulative control of NRF2, and *NRF2* activation leads to overabundance of Bcl-2 and Bcl-xL, apoptosis resistance, and increased survival of cancer cells (Niture and Jaiswal, 2012 and 2013). Such alterations of *NRF2* or its repressor *KEAP1*, which result in the loss of the regulative control and constant activation of *NRF2*, are common in lung cancer (Shibata *et al.*, 2008; Singh *et al.*, 2006; Wang *et al.*, 2008).

Changes in ROS homeostasis maintaining molecules have been described for instance in breast cancer (Amizi *et al.*, 2017), and lung cancer (Jiang *et al.*, 2014). Asbestos is known to induce cellular ROS production leading to DNA damage (Shukla *et al.*, 2003), and genetic aberrations in *KEAP1* have been associated with exposure to asbestos (Ruosaari *et al.*, 2008). Similarly, many studies have shown the connection between tobacco smoking and ROS induction (Carnevali *et al.*, 2003), and tobacco smoke has been associated with upregulation of NRF2-dependent genes (Rangasamy *et al.*, 2004).

Induced sputum consists of airway secretions with varying numbers of bronchoalveolar macrophages, leucocytes, and airway epithelial cells. In proteomics, sputum is little used to search for lung cancer biomarkers, despite its obvious potential in predicting disease

phenotype or illness severity at the site of disease origin. In other lung diseases, such as chronic obstructive pulmonary disease (COPD) and asthma, the sputum proteome has been extensively researched (reviewed in Paone *et al.*, 2016).

The aim of this study was to compare the effects of tobacco smoking, asbestos exposure, and lung cancer on the sputum proteome, and to explore potential biomarkers for lung cancer from sputum. Two-dimensional difference gel electrophoresis (2D-DIGE) coupled to a LC-MS/MS was used to screen for differences in sputum proteome between six groups consisting of current and former smokers, the asbestos exposed, and individuals with lung cancer, all matched for age and smoking history. Antibodies were used to validate the results in over 100 sputa. Our main findings include 1) peroxiredoxin 2 (PRDX2), thioredoxin (TXN), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) abundance in lung cancer, and 2) TXN paucity in asbestos exposure. These results bring more knowledge on the complex role of oxidative stress in lung cancer, and provide new insights as for the possible use of a set of proteins as disease markers.

Materials and methods

Patient samples

All samples were collected in Finnish health survey studies on tobacco smoking, and asbestos exposure either at the Finnish Institute of Occupational Health (FIOH), or at the Lapland Central Hospital. Sputa from the asbestos exposed came from a follow-up study on prevention and detection of asbestos-related disease program at FIOH in 2007.

Samples from healthy controls and those with lung cancer were collected in a health promotion campaign among tobacco smokers in Lapland between 2003 and 2005. Lung cancer group sputa were from patients diagnosed with primary lung cancer, classified according to the latest WHO principles at the Lapland Central Hospital. Participants were personally interviewed as to their smoking and work history, and gave informed consent to this study. Asbestos exposure evaluation was performed by a pulmonary specialist based on extensive interview on work history. The ethics committees of FIOH, the Hospital Districts of Helsinki and Uusimaa, and the Hospital District of Lapland reviewed and approved the experimental protocols (7/93, Dnro 75/E2/01, 3/9/03).

Samples were divided into two groups based on smoking history: current and former smokers. Both groups were further divided into three groups based on lung cancer and exposure to asbestos: A) current / former smokers with lung cancer, B) current / former smokers with pulmonary fibrosis caused by asbestos fibers, and C) control asymptomatic current / former smokers with no exposure to asbestos. General exclusion criteria were age < 40 years, smoking time < 10 years, pack years < 10 years, and cancer for groups B and C (allowing localized basal cell carcinoma). Individuals with any other cancer than lung cancer were omitted from group A.

Screening of potential biomarkers was done with 52 samples. Twenty sputa were from patients with histologically proven lung cancer, but no history of asbestos exposure (Group A). Sixteen patients suffered from asbestosis (Group B). Sixteen control patients had neither lung cancer nor known exposure to asbestos fibers (Group C). Twenty-four

sputa were from current, and twenty-eight from former smokers. Screening was performed using the 2-D DIGE protocol with CyDye DIGE Fluor minimal dyes in the Ettan DIGE system (GE Healthcare). Differentially abundant protein spots were cut from gels, and in-gel trypsin digested. Proxeon EASY nanoLC 1000 (Thermo Fisher Scientific, MA, USA) coupled to Q Exactive LC-MS/MS (Thermo Fisher Scientific, MA, USA) was used to analyze the obtained peptide mixtures. Mass fragment spectra were analyzed with Proteome Discoverer 1.4 (Thermo Fisher Scientific, MA, USA), and searched against human entries in the SwissProt and NCBI databases.

Validation of screening results was performed with Western blotting from 123 samples using four antibodies: anti-peroxiredoxin 2 (anti-PRDX2), anti-extracellular matrix protein 1 (anti-ECM1), anti-glyceraldehyde-3-phosphate dehydrogenase (anti-GAPDH), and anti-thioredoxin (anti-TXN). ELISA calprotectin (S100A8/A9 complex) detection kit was used to validate the anti-S100 calcium binding protein A8 (anti-S100A8). Of sputa that underwent validation, 27 belonged to group A, 37 to group B, and 59 to group C. Sixty-four were current, and 59 former smokers. Patient demographics and clinical profiles are provided in Table 1. Detailed experimental methods are described in Supplemental Digital Content 1 (see document).

Statistical analyses of antibody-based validation results

Statistical analyses were performed using the IBM SPSS Statistics 24.0 software (IBM Corporation, NY, USA). Normality of data was assessed using Shapiro-Wilk test for normality, and Levene's test for homogeneity of variances, whereafter nonparametric

tests were chosen to test the association between protein abundances and other variables. Nonparametric median test, and Mann-Whitney U-tests were used for pairwise comparisons, whereas Spearman's rho was used to study the correlation between variables. General linear modelling was used to study the effects of different variables on protein abundances.

Results

Identified proteins, and protein selection

For protein identification, 58 differentially abundant protein spots were chosen from the 2D-DIGE gels (see image, Supplemental Digital Content 2, demonstrating picked spots on 2D-DIGE gel). Four spots remained unidentified. Altogether 22 protein spots containing 32 distinctive proteins passed the demanded quality requirements for reliable identification (Table 2). Many of the proteins identified were saliva proteins, and possessed enzymatic activity. Figure 1 summarizes the molecular functions of the 32 identified proteins.

PRDX2 and GAPDH are significantly abundant in lung cancer

Pairwise comparisons of validated proteins across study groups, revealed a significant abundance of PRDX2 and GAPDH in those with lung cancer (Figure 2). TXN was also abundant in the lung cancer group, but statistically significant paucity of TXN was observed in the asbestos exposed (Figure 3). No statistically significant differences were observed in the abundances of ECM1 or S100A8.

PRDX2, TXN, and GAPDH form a correlative protein panel in lung cancer

A strong correlation was observed amidst the abundances of PRDX2, TXN, GAPDH, and S100A8 in the validation of 123 sputa, and the interactions between these proteins were further confirmed with a STRING database query (Figure 4). They all correlated significantly with smoking (Table 3), and all but TXN correlated with increasing age. PRDX2, TXN, and GAPDH correlated significantly with the presence of lung cancer, but no correlation was observed with cancer metastasis or stage. ECM1 did not correlate either with lung cancer or exposure to asbestos but did have a negative correlation with increasing smoking start age (Table 3). Comparison of ECM1 abundance in individuals with adenocarcinoma, or squamous cell carcinoma, revealed paucity in adenocarcinoma (see image Supplemental Digital Content 3 demonstrating pairwise comparisons), whereas all other studied proteins were abundant in adenocarcinomas. However, the difference in ECM1 did not reach statistical significance. GAPDH and S100A8 were significantly more abundant in adenocarcinomas when compared to squamous cell carcinomas.

Discussion

In this study, 2D-DIGE was used to detect differences in protein abundances from induced sputum. The purpose was to study the effects of tobacco smoking, asbestos exposure, and lung cancer on the sputum proteome, and to find potential biomarkers for lung cancer. Mass spectrometry (LC-MS/MS) was used to identify the differentially expressed proteins, from which five proteins were chosen for validation in over 100 sputa.

Four proteins, PRDX2, TXN, GAPDH, and S100A8 were shown to associate with lung cancer. TXN was a negatively correlated marker for asbestos exposure. ECM1 was also validated, but we did not observe any statistically significant differences in its abundance. Although ECM1 has been implicated in many cancers, only recent work links it to lung adenocarcinomas (Zhang *et al.*, 2015).

PRDX2 and TXN are biomarkers for lung cancer

Peroxiredoxins and thioredoxins are endogenous antioxidant molecules, whose role is to regulate the levels of ROS, especially H₂O₂, in cells. We have previously shown abundant plasma PRDX2 in individuals with lung cancer, and its paucity in the asbestos exposed (Rostila *et al.*, 2012). In the same study, we presented an association between PRDX1 and lung cancer. TXN was chosen for validation based on its association with tobacco smoking-induced oxidative stress in smokers with COPD (Lehtonen *et al.*, 2008), and on the paucity of thioredoxin-interacting protein, TXNIP, in asbestos-induced malignant mesothelioma (Thompson *et al.*, 2014).

PRDX2 and TXN are activated upon cellular oxidative stress. The protective and stabilizing role of peroxiredoxins in cells is performed by H₂O₂ reduction into water and oxygen. The peroxiredoxins act in the H₂O₂-mediated signal transduction in three different ways: 1) PRDXs modulate the level of H₂O₂, which directly reacts with a signaling protein; 2) H₂O₂ oxidizes PRDX, which relays oxidation to a signaling protein; or 3) H₂O₂ oxidized PRDX is reduced back by TXN, which becomes oxidized, and relays its oxidation to a signaling protein (Netto and Antunes, 2016). PRDX2 scavenges ROS

in the cytosol, and its gene overexpression has been associated with tumor progression, especially in lung cancer (Lehtonen *et al.*, 2004). However, PRDX2, like PRDX1, is also a known molecular chaperone, and interacts with many proteins, such as PGDF receptor (Choi *et al.*, 2005), which work as upstream sensors for H₂O₂. According to the floodgate hypothesis (Rhee and Woo, 2011), accumulated H₂O₂ in the cells hyperoxidizes PRDX2, and lifts the levels of TXN. *TXN* and *PRDX2* are targets of NRF2, and thus affected by oxidative stress-inactivated KEAP1. High levels of NRF2-regulated proteins in lung cancer patients may be due to genetic alterations of *NRF2* or *KEAP1*, previously reported in lung cancer cells, and resulting in uncontrolled expression of NRF2-driven genes.

GAPDH plays a role in lung cancer metabolism

GAPDH is a glycolytic enzyme that was first recognized pivotal in energy metabolism but has later been associated with other cellular key functions, such as DNA repair (Baxi and Vishwanatha, 1995), and phosphotransferase activity (Kim *et al.*, 2002). *GAPDH* overexpression is associated with poor prognosis in individuals with non-small cell lung cancer (NSCLC) (Puzone *et al.*, 2013), and with *EGFR*-mutated lung cancers (Sasaki *et al.*, 2007). GAPDH is extremely sensitive to oxidative stress, since its cysteine residue can easily be thiolated by H₂O₂ (Schuppe-Koistinen *et al.*, 1994). This leads to blocked glycolysis, and low levels of ATP (Colussi *et al.*, 2000), much due to GAPDH inactivation.

Compared to normal cells, cancer cells exhibit altered metabolism, such as overactive glycolysis (Nakajima and Van Houten, 2013), explaining the high levels of GAPDH in

individuals with lung cancer. Interestingly, Aykin-Burns *et al* (2009) showed that cancer cells under oxidative stress due to superoxide radicals or H₂O₂, are in greater risk of becoming glucose deprived than normal cells in the same conditions. This can explain why in our study, GAPDH correlated strongly with antioxidant proteins PRDX2 and TXN. High levels of PRDX2 and TXN can protect GAPDH from H₂O₂-mediated oxidation, allowing cancer cells to maintain increased levels of glycolysis and ATP, resulting in growth advantage over normal cells.

S100A8 and GAPDH contribute to inflammation and tumor formation

The S100 proteins, such as S100A8, are calcium binding proteins that function as intracellular sensors for calcium concentration. Genetic studies have revealed that mutated S100 genes are involved in many different cancer types, such as oral cancer (Lunde *et al.*, 2014) and NSCLC (Strazisar *et al.*, 2009). Calprotectin, a heterodimer of S100A8 and S100A9, is associated with colorectal cancer (Turvill *et al.*, 2016), and high abundance of calprotectin correlates with extended survival in NSCLC (Kawai, Minamiya and Takahashi, 2011). In human blood cells under oxidative stress, calprotectin controls the nitric oxide synthase-mediated *S*-nitrosylation of GAPDH (Jia *et al.*, 2014). This hampers the function of GAPDH in the interferon (IFN)- γ -activated inhibitor of translation (GAIT) system, leading to activation of pro-inflammatory genes, like interferon- γ (Bresnick *et al.*, 2015). Prolonged, the overactivation of these genes contributes to tumor formation (Coussens and Werb, 2002).

In our present study, we observed a strong correlation between S100A8, age, and increased smoking cessation age. Interestingly, there was also a statistically significant correlation between S100A8, GAPDH, TXN, and PRDX. These proteins are all involved in inflammatory pathways, and of them GAPDH is most tightly connected with S100A8. Although S100A8 does not by itself correlate with lung cancer, or exposure to asbestos, it could be used as a biomarker for lung cancer together with PRDX2, TXN, and GAPDH.

Inflammasome activation in exposure to asbestos involves TXN inactivation

Inhalation of asbestos causes asbestosis (pulmonary fibrosis), malignant mesothelioma, and lung cancer, especially in tobacco smokers (Mossmann and Churg, 1998). Once asbestos fibers are taken into cells by phagocytosis, they can, for instance, damage DNA, and prevent normal cell proliferation by interrupting the mitotic spindle formation. Inhaled fibers evoke inflammation through macrophage recruitment, and lead to apoptosis. In mice, asbestos induced ROS production activates the Nalp3 inflammasome, which leads to IL-1 β -driven inflammation (Dostert *et al.*, 2007). IL-1 β promotes tumor growth and invasiveness in many cancer types (Krelin,*et al.*, 2007).

In our study, we observed a negative correlation between asbestos exposure and TXN abundance. Previously, we have reported a negative correlation with PRDX2 and asbestos exposure in plasma (Rostila *et al.*, 2012). Yet in sputa obtained from individuals with lung cancer, both TXN and PRDX2 were abundant. This suggests that the ROS production in lung cancer, and in asbestos exposure, differ from each other, probably due to the loss of regulative control of NRF2 in lung cancer cells.

Studies in mice (Dostert *et al.*, 2007) show, that macrophages treat internalized asbestos fibers in a similar manner to pathogens: via the inflammasome activation. Tobacco smoke, for instance, does not activate the inflammasome, or lead to IL-1 β secretion. Short asbestos fibers ($\leq 3 \mu\text{m}$) are internalized and cleared from the pleural space easily, causing no inflammation (Donaldson *et al.*, 1989). Long fibers ($> 10 \mu\text{m}$) undergo frustrated phagocytosis, since they cannot be completely engulfed by the cell membrane. This leads to initiation of intracellular stress, and inflammation after fiber exposure. Dostert *et al* (2007) showed that frustrated phagocytosis of asbestos fibers triggers the formation of ROS that contributes to inflammasome activation. They also observed increased Nalp3 inflammasome-activated IL-1 β secretion in TXN deficient cells. Inflammasome activation is mediated by TXNIP, lack of which is present in malignant mesothelioma (Thompson *et al.*, 2014). Albeit, asbestos exposed individuals in our study were not mesothelioma patients, this data together with our results suggests that oxidative stress in asbestos exposed cells is different from cells exposed to tobacco smoke, or from cells under carcinogenesis. Thus, cellular responses to ROS in asbestos exposure are mediated via different molecules.

Concluding remarks

The present study aimed to study the effects of tobacco smoking and asbestos exposure on the sputum proteome, and to identify potential sputum biomarkers for lung cancer. We found that PRDX2, TXN, GAPDH, and S100A8 could serve as a panel of biomarkers for recognizing individuals with increased risk of lung cancer. Especially proteins acting in

ROS-regulation, PRDX2 and TXN, and GAPDH displayed significant association with the presence of lung cancer. We observe that the changes associated with the ROS homeostasis of the cell, are different in the asbestos-exposed than those with lung cancer. Our results affirm that sputum is a promising matrix for biomarker detection in lung cancer, even though further validations are needed to confirm the usefulness of the studied proteins and sputum in screening high-risk individuals for lung cancer.

Acknowledgements

The authors would like to acknowledge Niina Ahonen and Teija Inkinen for excellent technical assistance.

References

Aykin-Burns N, Ahmad IM, Zhu Y, Oberley LW, Spitz DR. Increased levels of superoxide and H₂O₂ mediate the differential susceptibility of cancer cells versus normal cells to glucose deprivation. *Biochem J* 2009;**418**(1):29-37.

Azimi I, Petersen RM, Thompson EW, Roberts-Thomson SJ, Monteith GR. Hypoxia-induced reactive oxygen species mediate N-cadherin and SERPINE1 expression, EGFR signalling and motility in MDA-MB-468 breast cancer cells. *Sci Rep* 2017;**7**(1):15140.

Baxi MD, Vishwanatha JK. Uracil DNA-glycosylase/glyceraldehyde-3-phosphate dehydrogenase is an Ap4A binding protein. *Biochemistry* 1995;**34**(30):9700-9707.

Bresnick AR, Weber DJ, Zimmer DB. S100 proteins in cancer. *Nat Rev Cancer* 2015;**15**(2):96-109.

Carnevali S, Petruzzelli S, Longoni B, Vanacore R, Barale R, Cipollini M, et al. (2003). Cigarette smoke extract induces oxidative stress and apoptosis in human lung fibroblasts. *Am J Physiol Lung Cell Mol Physiol* **284**(6):L955-963.

Choi MH, Lee IK, Kim GW, Kim BU, Han YH, Yu DY, et al. (2005). Regulation of PDGF signalling and vascular remodelling by peroxiredoxin II. *Nature* **435**(7040):347-353.

Colussi C, Albertini MC, Coppola S, Rovidati S, Galli F, Ghibelli L. H₂O₂-induced block of glycolysis as an active ADP-ribosylation reaction protecting cells from apoptosis. *FASEB J* 2000;**14**(14):2266-2276.

Coussens LM, Werb Z. Inflammation and cancer. *Nature* 2002;**420**(6917):860-867.

Donaldson K, Brown GM, Brown DM, Bolton RE, Davis JM. Inflammation generating potential of long and short fibre amosite asbestos samples. *Br J Ind Med* 1989;**46**:271-276.

Dostert C, Pétrilli V, Van Bruggen R, Steele C, Mossman BT, Tschopp J. Innate immune activation through Nalp3 inflammasome sensing of asbestos and silica. *Science* 2008;**320**(5876):674-677.

Ezzati M, Henley SJ, Lopez AD, Thun MJ. Role of smoking in global and regional cancer epidemiology: current patterns and data needs. *Int J Cancer* 2005;**116**(6):963-971.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011;**61**(2):69-90.

Jia J, Arif A, Terenzi F, Willard B, Plow EF, Hazen SL, et al. (2014). Target-selective protein S-nitrosylation by sequence motif recognition. *Cell* **159**(3):623-634.

Jiang H, Wu L, Mishra M, Chawsheen HA, Wei Q. Expression of peroxiredoxin 1 and 4 promotes human lung cancer malignancy. *Am J Cancer Res* 2014;**4**(5):445-460.

Kawai H, Minamiya Y, Takahashi N. Prognostic impact of S100A9 overexpression in non-small cell lung cancer. *Tumour Biol* 2011;**32**(4):641-646.

Kim HJ, Song EJ, Lee KJ. Proteomic analysis of protein phosphorylations in heat shock response and thermotolerance. *J Biol Chem* 2002;**277**(26):23193-23207.

Krelin Y, Voronov E, Dotan S, Elkabets M, Reich E, Fogel M, et al. (2007). Interleukin-1beta-driven inflammation promotes the development and invasiveness of chemical carcinogen-induced tumors. *Cancer Res* **67**:1062.

Lehtonen ST, Svensk AM, Soini Y, Pääkkö P, Hirvikoski P, Kang SW, et al. (2004). Peroxiredoxins, a novel protein family in lung cancer. *Int J Cancer* **111**(4):514-521.

Lehtonen ST, Ohlmeier S, Kaarteenaho-Wiik R, Harju T, Pääkkö P, Soini Y, et al. (2008).

Does the oxidative stress in chronic obstructive pulmonary disease cause thioredoxin/peroxiredoxin oxidation? *Antioxid Redox Signal* **10**(4):813-819.

Lennicke C, Rahn J, Lichtenfels R, Wessjohann LA, Seliger B. Hydrogen peroxide - production, fate and role in redox signaling of tumor cells. *Cell Commun Signal* 2015;**13**:39.

Lunde ML, Roman E, Warnakulasuriya S, Mehrotra R, Laranne J, Vasstrand EN, et al. (2014). Profiling of chromosomal changes in potentially malignant and malignant oral mucosal lesions from south and south-east Asia using array-comparative genomic hybridization. *Cancer Genomics Proteomics* **11**:127-140.

Menegon S, Columbano A, Giordano S. The Dual Roles of NRF2 in Cancer. *Trends Mol Med* 2016;**22**(7):578-593.

Mossman BT, Churg A. Mechanisms in the pathogenesis of asbestosis and silicosis. *Am J Respir Crit Care Med* 1998;**157**:1666.

Nakajima EC, Van Houten B. Metabolic symbiosis in cancer: refocusing the Warburg lens. *Mol Carcinog* 2013;**52**(5):329-337.

National Lung Screening Trial Research Team, Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, Galen B, et al. (2011). The National Lung Screening Trial: overview and study design. *Radiology* **258**(1):243-253.

Nielsen LS1, Bælum J, Rasmussen J, Dahl S, Olsen KE, Albin M, et al. (2014). Occupational asbestos exposure and lung cancer--a systematic review of the literature. *Arch Environ Occup Health* **69**(4):191-206.

Niture SK, Jaiswal AK. Nrf2 protein up-regulates antiapoptotic protein Bcl-2 and prevents cellular apoptosis. *J Biol Chem* 2012;**287**(13):9873-9886.

Niture SK, Jaiswal AK. Nrf2-induced antiapoptotic Bcl-xL protein enhances cell survival and drug resistance. *Free Radic Biol Med* 2013;**57**:119-131.

Netto LE, Antunes F. The Roles of Peroxiredoxin and Thioredoxin in Hydrogen Peroxide Sensing and in Signal Transduction. *Mol Cells* 2016;**39**(1):65-71.

Puzone R, Savarino G, Salvi S, Dal Bello MG, Barletta G, Genova C, et al. Glyceraldehyde-3-phosphate dehydrogenase gene over expression correlates with poor prognosis in non small cell lung cancer patients. *Mol Cancer* 2013;**12**(1):97.

Rangasamy T, Cho CY, Thimmulappa RK, Zhen L, Srisuma SS, Kensler TW, et al. (2004). Genetic ablation of Nrf2 enhances susceptibility to cigarette smoke-induced emphysema in mice. *J Clin Invest* **114**(9):1248-1259.

Rhee SG, Woo HA. Multiple functions of peroxiredoxins: peroxidases, sensors and regulators of the intracellular messenger H₂O₂, and protein chaperones. *Antioxid Redox Signal* 2011;**15**(3):781-794.

Rostila A, Puustinen A, Toljamo T, Vuopala K, Lindström I, Nyman TA, et al. (2012). Peroxiredoxins and tropomyosins as plasma biomarkers for lung cancer and asbestos exposure. *Lung Cancer* **77**(2):450-459.

Ruosaari ST, Nymark PE, Aavikko MM, Kettunen E, Knuutila S, Hollmén J, et al. (2008). Aberrations of chromosome 19 in asbestos-associated lung cancer and in asbestos-induced micronuclei of bronchial epithelial cells in vitro. *Carcinogenesis* **29**(5):913-917.

Rushton L, Hutchings S, Brown T. The burden of cancer at work: estimation as the first step to prevention. *Occup Environ Med* 2008;**65**(12):789-800.

Sasaki H, Okuda K, Endo K, Kawano O, Yukiue H, Yokoyama T, et al. CCND1 messenger RNA expression is correlated with EGFR mutation status in lung cancer. *Clin Lung Cancer* 2007;**8**(8):493-496.

Schuppe-Koistinen I, Moldéus P, Bergman T, Cotgreave IA. S-thiolation of human endothelial cell glyceraldehyde-3-phosphate dehydrogenase after hydrogen peroxide treatment. *Eur J Biochem* 1994;**221**(3):1033-1037.

Shibata T, Ohta T, Tong KI, Kokubu A, Odogawa R, Tsuta K, et al. (2008). Cancer related mutations in NRF2 impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. *Proc Natl Acad Sci U S A* **105**(36):13568-13573.

Shukla A, Gulumian M, Hei TK, Kamp D, Rahman Q, Mossman BT. Multiple roles of oxidants in the pathogenesis of asbestos-induced diseases. *Free Radic Biol Med* 2003;**34**(9):1117-1129.

Singh A, Misra V, Thimmulappa RK, Lee H, Ames S, Hoque MO, et al. (2006). Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer. *PLoS Med* 3(10):e420.

Strazisar M, Rott T, Glavac D. Frequent polymorphic variations but rare tumour specific mutations of the S100A2 on 1q21 in non-small cell lung cancer. *Lung Cancer* 2009;63:354-359.

Thompson JK, Westbom CM, MacPherson MB, Mossman BT, Heintz NH, Spiess P, et al. (2014). Asbestos modulates thioredoxin-thioredoxin interacting protein interaction to regulate inflammasome activation. *Part Fibre Toxicol* 11:24.

Tossavainen A. Global use of asbestos and the incidence of mesothelioma. *Int J Occup Environ Health* 2004;10(1):22-25.

Turvill J, Aghahoseini A, Sivarajasingham N, Abbas K, Choudhry M, Polyzois K, et al. (2016). Faecal calprotectin in patients with suspected colorectal cancer: a diagnostic accuracy study. *Br J Gen Pract* 66(648):e499-506.

Wang R, An J, Ji F, Jiao H, Sun H, Zhou D. Hypermethylation of the Keap1 gene in human lung cancer cell lines and lung cancer tissues. *Biochem Biophys Res Commun* 2008;**373**(1):151-154.

Zhang Y, Wang H, Wang J, Bao L, Wang L, Huo J, et al. (2015). Global analysis of chromosome 1 genes among patients with lung adenocarcinoma, squamous carcinoma, large-cell carcinoma, small-cell carcinoma, or non-cancer. *Cancer Metastasis Rev* **34**(2):249-264.

Tables

Table 1 Summary of study population

	<i>Two dimensional fluorescence difference gel electrophoresis (2D-DIGE)</i>			<i>Western blot / ELISA</i>		
	<i>Screening set (n=52)</i>			<i>Validation set (n=123)</i>		
	<i>(A) lung cancer n=20</i>	<i>(B) asbestos exposed n=16</i>	<i>(C) healthy smokers n=16</i>	<i>(A) lung cancer n=27</i>	<i>(B) asbestos exposed n=37</i>	<i>(C) healthy smokers n=59</i>
Sex						
Male / Female	17 / 3	16 / 0	15 / 1	21 / 6	37 / 0	39 / 20
Age						
Mean $\pm$ SD	65.5 $\pm$ 9.2	66.9 $\pm$ 7.3	65.9 $\pm$ 6.4	66.9 $\pm$ 8.4	66.5 $\pm$ 5.9	63.1 $\pm$ 6.5
Smoker status						
Current / Former	8 / 12	8 / 8	8 / 8	15 / 12	11 / 26	38 / 21
Smoking start age						
Mean $\pm$ SD	16.8 $\pm$ 3.6	17.1 $\pm$ 2.8	18.3 $\pm$ 3.6	16.4 $\pm$ 3.3	16.8 $\pm$ 3.2	18.5 $\pm$ 4.8
Smoking cessation age						
Mean $\pm$ SD	57.4 $\pm$ 10.6	45.3 $\pm$ 12.2	53.4 $\pm$ 9.6	58.9 $\pm$ 10.9	43.8 $\pm$ 12.5	52.7 $\pm$ 8.3
Smoking years						
Mean $\pm$ SD	42.5 $\pm$ 10.6	38.1 $\pm$ 13.2	41.7 $\pm$ 10.0	45.4 $\pm$ 9.7	33.2 $\pm$ 14.7	41.3 $\pm$ 7.8

Pack years						
Mean $\pm$ SD	40.3 $\pm$ 14.8	39.9 $\pm$ 17.8	40.6 $\pm$ 12.6	50.0 $\pm$ 30.5	36.4 $\pm$ 21.8	44.6 $\pm$ 22.1
Tumor histology						
AC/ SCC/	4/ 8/ 5/ 1/ 2	-	-	4/ 15/ 3/ 1/ 4	-	-
SCLC/ LCC /						
other						
Stage*						
I/ II/ III/ IV	2/ 2/ 8/ 3	-	-	4/ 4/ 8/ 8	-	-
Sputum class**						
benign/						
suspicious for						
malignancy	9 / 11	16 / 0	16 / 0	58 / 1	15 / 11	30 / 0

Validation set includes the screening set. The screening set includes four lung cancer samples not present in the validation set. Abbreviations: AC, adenocarcinoma; SCC, squamous cell carcinoma; SCLC, small cell lung cancer; LCC, large cell lung cancer.

*SCLC omitted from staging. **Sputum class unknown from seven asbestos exposed individuals in the validation set.

Table 2 Summary of protein identification data, and relationships between differently abundant proteins in study groups

spot number	protein / abbreviation	UniProt accession number	Theoretical ^{a)}		Sequence coverage (%)	Mascot score	Number. of unique peptides/ all peptides	Direction of change
			MW (kDa)	pI				
1	Elongation factor 2	P13639	95.3	6.8	4.3	61	4 / 13	up A,B
2	Serum albumin	P02768	69.3	6.3	21.4	513	13 / 50	up A
	Extracellular matrix protein 1 / ECM1	Q16610	60.6	6.7	12.8	130	4 / 8	
3	Serpin B3	P29508	44.5	6.8	30.8	327	9 / 31	up B, C
	Alpha-amylase 1	P04745	57.7	6.9	19.0	305	8 / 24	
	Carboxypeptidase A4	Q9UI42	47.3	6.7	7.1	85	2 / 9	
4	Alpha-amylase 1	P04745	57.7	6.9	11.0	197	4 / 13	up C
	Ig alpha-1 chain C region	P01876	37.6	6.5	13.9	88	4 / 8	
5	Annexin A2	P07355	38.6	7.8	21.8	132	6 / 15	up B, C
	Ig alpha-1 chain C region	P01876	37.6	6.5	13.9	82	3 / 5	
	Suprabasin	Q6UWP8	60.5	7.0	4.8	78	2 / 3	

6	Catalase	P04040	59.7	7.4	10.1	74	3 / 12	up C
	Alpha-amylase 1	P04745	57.7	6.9	5.5	60	2 / 6	
	Alpha-amylase 1	P04745	57.7	6.9	17.0	293	6 / 29	
	Ig alpha-1 chain C region	P01876	37.6	6.5	13.9	150	4 / 6	
7	Annexin A2	P07355	38.6	7.8	32.5	235	9 / 28	up B, C
	Alpha-amylase 1	P04745	57.7	6.9	17.0	210	5 / 24	
	Arginase-1	P05089	34.7	7.2	24.8	116	5 / 14	
8	Leukocyte elastase inhibitor	P30740	42.7	6.3	23.0	209	7 / 33	up A, B
	Actin, cytoplasmic 1	P60709	41.7	5.5	27.7	84	6 / 26	
9	Alpha-amylase 1	P04745	57.7	6.9	12.9	211	5 / 18	up B, C
10	Annexin A2	P07355	38.6	7.8	15.9	121	3 / 12	up A
	Glyceraldehyde-3-phosphate dehydrogenase / GAPDH	P04406	36.0	8.5	18.8	155	5 / 32	
	Arginase-1	P05089	34.7	7.2	9.3	87	4 / 15	
11	Glyceraldehyde-3-phosphate dehydrogenase / GAPDH	P04406	36.0	8.5	32.2	353	8 / 40	up A

	Annexin A2	P07355	38.6	7.8	21.2	158	5 / 21	
	Arginase-1	P05089	34.7	7.2	23.6	101	6 / 21	
	F-box only protein 50	Q6ZVX7	30.8	6.6	18.6	50	3 / 9	
	Cathepsin D	P07339	44.5	6.5	9.5	45	2 / 10	
12	Arginase-1	P05089	34.7	7.2	6.8	84	2 / 10	up A
13	Zinc-alpha-2-glycoprotein	P25311	34.2	6.1	9.1	80	2 / 10	up A
14	Zinc-alpha-2-glycoprotein	P25311	34.2	6.1	7.4	63	2 / 4	up A
15	Glyceraldehyde-3-phosphate dehydrogenase / GAPDH	P04406	36.0	8.5	43.0	451	11 / 50	up C
	Arginase-1	P05089	34.7	7.2	33.5	226	9 / 24	
	Proteasome subunit alpha type-6	P60900	27.4	6.8	14.2	156	3 / 15	
	F-box only protein 50	Q6ZVX7	30.8	6.6	8.00	117	2 / 6	
	Kallikrein-7	P49862	27.5	8.5	19.0	74	4 / 13	
	Proteasome subunit alpha type-7	O14818	27.9	8.5	10.1	60	2 / 5	
	Peroxiredoxin-1	Q06830	22.1	8.1	23.6	49	5 / 11	
	Ig gamma-1 chain C region	P01857	36.1	8.2	7.9	41	2 / 2	

16	Peptidyl-prolyl cis-trans isomerase A	P62937	18.0	7.8	20.6	73	2 / 9	up B, C
17	Peroxiredoxin-2 / PRDX2	P32119	21.9	6.0	28.8	155	5 / 17	up A, C
	Gamma-glutamylcyclotransferase	O75223	21.0	5.1	27.7	83	4 / 11	
	Immunoglobulin J chain	P01591	18.1	5.2	11.3	36	2 / 8	
18	Gamma-glutamylcyclotransferase	O75223	21.0	5.1	18.1	83	3 / 6	up A, C
	Cystatin-A	P01040	11.0	5.5	52.0	79	3 / 7	
19	Lysozyme C	P61626	16.5	9.2	12.8	85	2 / 8	up B, C
20	Cystatin-SN	P01037	16.4	7.2	56.7	662	5 / 72	up B, C
	Thioredoxin / TXN	P10599	11.7	4.9	31.4	64	2 / 5	
21	Protein S100-A8 / S100A8	P05109	10.8	7.0	23.7	72	2 / 6	up B, C
	Cystatin-B	P04080	11.1	7.6	24.5	45	2 / 4	
22	Cystatin-SN	P01037	16.4	7.2	44.7	209	5 / 32	up B, C
	Protein S100-A8 / S100A8	P05109	10.8	7.0	23.7	43	2 / 4	

Significant change was determined as average ratio ≥ 1.5 , and as $p \leq 0.05$ from Student's T-test. Validated proteins in bold. Abbreviations:

A, lung cancer; B, asbestos exposure; C, control.

Table 3 Summary of correlations between validated proteins and study population variables

	<i>Increasing age</i>	<i>Continued smoking</i>	<i>Increasing smoking start age</i>	<i>Increasing smoking cessation age</i>	<i>Increasing smoking years</i>	<i>Presence of asbestosis</i>	<i>Intensifying asbestos exposure</i>	<i>Increasing packyears</i>	<i>Presence of lung cancer</i>
PRDX2									
Spearman's correlation coefficient	0,194			0,358	0,212			0,217	0,327
Significance (2-tailed)	0,032			0,005	0,019			0,018	0,000
TXN									
Spearman's correlation coefficient		0,217				-0,327	-0,318	0,261	0,245
Significance (2-tailed)		0,017				0,000	0,000	0,004	0,006
GAPDH									
Spearman's correlation coefficient	0,228								0,297
Significance (2-tailed)	0,011								0,000
ECM1									
Spearman's correlation coefficient			-0,231						
Significance (2-tailed)			0,0105						
S100A8									
Spearman's correlation coefficient	0,197			0,282					
Significance (2-tailed)	0,030			0,030					

Abbreviations: PRDX2, peroxiredoxin 2; TXN, thioredoxin; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; ECM1, extracellular matrix protein 1; S100A8, protein S100-A8.

Captions

Figure 1. Functional classification obtained from Panther gene list analysis tool. Data shows the molecular functions of 28/32 proteins identified from the screening set in the sputa.

Figure 2. Results from the immunological validation of 123 sputa in three study groups. The plots present medians with interquartile range. The asymptomatic p-values were obtained from Mann-Whitney U pairwise 2-tailed testing.

Figure 3. Comparison of the TXN abundance between the asbestos exposed group and the non-exposed group from the immunological validation of 123 sputa. The plots present medians with interquartile range. The asymptomatic p-values were obtained from Mann-Whitney U pairwise 2-tailed testing.

Figure 4. A) Protein interactions of PRDX2, TXN, GAPDH, and S100A8 from STRING database (v10.5). PRDX2, TXN, and GAPDH are co-expressed (purple line), and their linkage is determined both experimentally (pink line) and from textmining (green line). S100A8 connects with GAPDH by textmining, and experimentally. **B)** Protein interactions of PRDX2, TXN, GAPDH, and S100A8 observed in the validation of sputa. Figure represents the Spearman's correlation coefficients/significance (2-tailed).

Supplemental Digital Content 2. 2D-DIGE gel from screening set of sputa. Numbers refer to spot numbers, and identified proteins listed in Table 2.

Supplemental Digital Content 3. Comparison of the results from the immunological validation of 123 sputa in two histological types of lung cancer: adenocarcinoma (n=4), and squamous cell carcinoma (n=15). The plots present medians with interquartile range. The asymptomatic p-values were obtained from Mann-Whitney U pairwise 2-tailed testing.

List of Supplemental content

Supplemental digital content 1.doc

Supplemental digital content 2.ppt

Supplemental digital content 3.ppt

PANTHER GO-Slim Molecular Function
Total # Genes: 32 Total # function hits: 28

No.	Molecular function	No. of proteins
1	translation regulator activity (GO:0045182)	1
2	binding (GO:0005488)	8
3	structural molecule activity (GO:0005198)	1
4	catalytic activity (GO:0003824)	15
5	antioxidant activity (GO:0016209)	3

Figure 1.

Figure 2.

Figure 3.

Figure 4.

Supplemental Digital Content 2.

Supplemental Digital Content 3.

Supplemental Digital Content 1.

Sample processing

Sputum was induced with hypertonic saline, and papanicolaou staining was performed to all induced sputum (IS) smear samples. Sputa from controls, and individuals with asbestos exposure had benign cytology. Sputa from individuals with lung cancer either had benign cytology or were suspicious for containing malignant cells. Cytologic atypia of samples was screened by a cytotechnologist, and evaluated by a pathologist.

After induction, sputa were stored in 10 ml of Saccomanno fixative (2 % polyethyleneglycol / 50 % ethanol), and stored in -20 C degrees. To homogenize the samples, sputolysin (1 g dithiotreitol (DTT) / 100 ml 1M potassiumphosphate buffer solution, pH 7) was added to the sputa. The amount of sputolysin added (1-6 ml), depended on the consistency of the sample. The samples were shaken (125-150 rpm) at room temperature for 45-60 minutes, until filtered with a nylon filtration unit (28-100 μm) to a fresh falcon tube.

Once fully filtered, all samples were centrifuged at 500 g for 10 minutes at +4 C degrees to separate between the cell fraction and the sputum supernatant. The supernatant was then transferred into a fresh tube, and centrifuged at 4000 g for 20 minutes at +4 C degrees. The remaining supernatant was then transferred into a syringe and filtered with a PVDF filtration unit (0.45 μm) to a fresh falcon tube. A pool sample was created from

all filtered samples by taking an equal amount of each sample. Liquefied samples were stored at -80 C degrees.

Two-Dimensional Differential in Gel Electrophoresis (2-D DIGE) and protein identification

To maximize the protein concentration, and to remove excess DTT, 52 sputa (1ml), including a pool sample, were concentrated using ProteoSpin protein concentration spin columns (Norgen Biotek, Ontario, Canada). The pH of the samples was adjusted to 3.5 with 1M phosphoric acid prior to the concentration, according to the manufacturer's instructions. After sample concentration, the protein concentrations were checked with 2-D Quant kit (GE Healthcare, Uppsala, Sweden), and 55 µg of each sample/pool was chosen to undergo the 2-D Clean-Up protocol (GE Healthcare).

Samples underwent the 2-D DIGE protocol with CyDye DIGE Fluor minimal dyes using the Ettan DIGE system (GE Healthcare). A pooled sample from all 52 sputa was used as an internal control on all gels. In the first dimension, proteins (50 µg) from concentrated sputa underwent isoelectric focusing with the IPGphor system, which was carried out with 18 cm IPG strips at 3-10 nonlinear strip gradient (GE Healthcare). In the second dimension, SDS-PAGE was performed in 12% polyacrylamide gels with the Ettan DALTsix system (GE Healthcare). DeCyder 7.0 software (GE Healthcare) was used for spot detection and relative quantitation of proteins. Student's t test was used to determine the P-values for differential protein abundance levels of each spot (fold change $|\geq 1.5|$, $P < 0.05$). Gels were silver stained for spot picking.

Fifty-eight significantly up- or down-regulated protein spots were cut from the gels, reduced with DTT, and alkylated with iodoacetamide (GE Healthcare) in preparation for in-gel digestion with trypsin (modified sequencing grade porcine trypsin, Promega, Madison, WI, USA) for 16 h at +37 °C. After removing supernatants to fresh tubes, the remaining peptides were extracted twice from selected gel punches by using 100 µl of 5% formic acid in 50% acetonitrile (ACN). The combined extracts were then dried in a vacuum centrifuge. Peptide mixtures were analyzed by Proxeon EASY nanoLC 1000 (Thermo Fisher Scientific, MA, USA) coupled to Q Exactive LC-MS/MS (Thermo Fisher Scientific, MA, USA). In reverse phase, peptides were separated using a Dionex Acclaim PepMap100 C18 75 µm x 15 cm column (Thermo Fisher Scientific, MA, USA) at a flow rate of 300 nl / min. Peptides were eluted from the column in 80 minutes with a linear gradient of 5-35 % 0.1 % formic acid in 95 % acetonitrile.

The obtained mass fragment spectra were analyzed with Proteome Discoverer 1.4 (Thermo Fisher Scientific, MA, USA), and searched against human entries in the SwissProt and NCBI databases. One missed cleavage was allowed, and searches were performed with fixed carbamidomethylation of cysteines, and variable oxidation of methionine, histidine, and tryptophan residues. A fragment tolerance of 0.2 Da and parent tolerance of 0.5 Da were used with trypsin as the specified digestion enzyme. Selection criteria for identified proteins included two unique matched peptides, post-error probability ≤ 0.05 , q-value ≤ 0.05 , Mascot score ≥ 20 , and sequence coverage ≥ 4 %, and correct size on 2DE gel.

Antibody-based validations

To validate the screening results, 123 unconcentrated sputa underwent western blot validation with four antibodies, anti-pxoxiredoxin 2 (anti-PRDX2), anti-extracellular matrix protein 1 (anti-ECM1), anti-glyceraldehyde-3-phosphate dehydrogenase (anti-GAPDH), and anti-thioredoxin (anti-TXN). ELISA calprotectin (S100A8/A9 complex) detection kit was used to validate the anti-S100 calcium binding protein A8 (anti-S100A8).

Protein concentrations of the unconcentrated sputa were determined by running them on 12 % Criterion TGX Precast Gels (Bio-Rad), and silver-staining. Band intensities were detected and analyzed with ImageQuant LAS 4000 Mini CCD camera, and ImageQuantTL software (GE Healthcare). Protein concentrations were determined against the BSA standard (2 µg).

Sputa proteins (15 µg) were cleaned from interfering reagents, such as salts and DTT, by acetone precipitation to undergo western blotting. A pooled sample of 52 screening set sputa was used as an internal control on all gels. The precipitated pellets were dissolved in Bolt running buffer, and Bolt 4-12 % Bis-Tris Plus gels were run with 165 V for 35 minutes according to the Bolt gel system protocol (Invitrogen by Thermo Fisher Scientific, MA, USA). Color Protein Prestained Broad Range (11-245 kDa) molecular weight standard (New England BioLabs, MA, USA) was used on all gels. Blotting was

performed with Trans-Blot Turbo transfer system, and RTA Mini PVDF Transfer kit (Bio-Rad, CA, USA) with 25 V / 1.3 A for 7 minutes.

After blocking the membranes with 5% skimmed milk in PBS containing 0.05% Tween (Medicago AB, Uppsala, Sweden) for 30 minutes, they were incubated with commercially available mouse polyclonal primary antibodies anti-PRDX2 (1:2000, Alexis Biochemicals, San Diego, CA, USA), and rabbit polyclonal antibodies anti-ECM1 (1:5000), anti-GAPDH (1:2500), and anti-TXN (1:5000) (Abcam, Cambridge, UK) overnight at +4°C. Immunoblots were stained with horseradish peroxidase (HRP)-conjugated polyclonal goat anti-mouse, and goat anti-rabbit secondary antibodies (Bio-Rad, CA, USA) at 1:2000 dilution for 1 h at room temperature, and visualized with the chemiluminescent Clarity Western ECL substrate (Bio-Rad, CA, USA). Imaging of blots was executed with the ChemiDoc Touch imaging system (Bio-Rad, CA, USA), and analyses of spot intensities between groups were performed using the ImageLab 6.0 software (Bio-Rad, CA, USA).

ELISA was performed to the acetone-precipitated pellets according to the Buhlman fCAL ELISA calprotectin protocol (Buhlman Laboratories AG, Schönenbuch, Switzerland). The absorbances were measured with 450 nm (Perkin Elmer, MA, USA).