

## ORIGINAL ARTICLE

# Cancer-related proteins in serum are altered in workers occupationally exposed to polycyclic aromatic hydrocarbons: a cross-sectional study

Ayman Alhamdow<sup>1</sup>, Håkan Tinnerberg<sup>2</sup>, Christian Lindh<sup>2</sup>, Maria Albin<sup>1,2,3</sup>, Karin Broberg<sup>1,2,\*</sup>

<sup>1</sup>Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, <sup>2</sup>Division of Occupational and Environmental Medicine, Department of Laboratory Medicine, Lund University, Lund, Sweden and <sup>3</sup>Centre for Occupational and Environmental Medicine (Camm), Stockholm County Council, Stockholm, Sweden

\*To whom correspondence should be addressed. Tel: +46 8 524 874 05; Fax: +46 8 33 69 81; Email: [karin.broberg@ki.se](mailto:karin.broberg@ki.se)

## Abstract

Exposure to some polycyclic aromatic hydrocarbons (PAH) increases the risk of cancer and is common particularly for workers in occupations such as chimney sweeping. In exposed workers, screening of early cancer-related markers provides important information to identify individuals at risk. Here, we aimed to elucidate the associations between PAH exposure and serum levels of cancer-related proteins in 118 chimney sweeps and 126 occupationally unexposed controls, all non-smoking males from Sweden. Monohydroxylated metabolites of pyrene, phenanthrene, benzo[a]pyrene and benzo[a]anthracene were measured in urine using liquid chromatography coupled to tandem mass spectrometry and 90 cancer-related proteins were measured in serum using a proximity extension assay. Linear regression analysis adjusted for age and body mass index, and false discovery rate (FDR) identified 17 serum proteins that were differentially expressed (16 upregulated and 1 downregulated) in chimney sweeps compared with controls (FDR < 0.05). Concentrations of the peptidase kallikrein 13 (KLK13) showed significant positive associations with urinary concentrations of the PAH metabolites 3-hydroxybenzo[a]pyrene (3-OH-BaP) [B, 95% confidence interval (CI): 0.042, 0.008–0.076] and 3-hydroxybenzo[a]anthracene (3-OH-BaA) [B, 95% CI: 0.068, 0.002–0.134]. Moreover, dose–response relationships were observed between KLK13 and 3-OH-BaP (trend test  $P = 0.027$ ) and 3-OH-BaA ( $P = 0.035$ ). Pathway and gene ontology analyses showed that cell movement, cell adhesion and cell migration were the predominant molecular functions associated with the top differentially expressed proteins. In conclusion, we found a number of putative cancer-related proteins differentially expressed in workers exposed to PAH. This warrants effective measure to reduce PAH exposure among workers as well as further investigation to confirm these findings.

## Introduction

The World Health Organization has estimated that cancer accounted for about 9.6 million deaths worldwide in 2018 (1). Moreover, cancer accounted for 32% of all work-related deaths worldwide and more than 50% of work-related deaths in developed countries, including the European Union (2). Exposure to carcinogens occurs in multiple industries and workplaces. The first epidemiological evidence of occupational cancer came

from Britain in 1775, when Sir Percivall Pott linked exposure to soot in chimney sweeps to skin cancer of the scrotum (3). More recently, several epidemiological studies have shown higher incidence of cancer and higher mortality among chimney sweeps. Pukkala *et al.* investigated the incidence of occupational cancer among several occupational groups including data from 15 million individuals from Sweden, Finland, Norway, Iceland

Received: September 24, 2018; Revised: January 4, 2019; Accepted: February 7, 2019

© The Author(s) 2019. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

**Abbreviations**

1-OH-PYR	1-hydroxypyrene
3-OH-BaA	3-hydroxybenzo[a]anthracene
3-OH-BaP	3-hydroxybenzo[a]pyrene
AHR	aryl hydrocarbon receptor
ANXA1	annexin A1
BMI	body mass index
CEACAM5	carcinoembryonic antigen-related cell adhesion molecule 5
CI	confidence interval
CXC13	C-X-C motif chemokine 13
DEP	differentially expressed proteins
FADD	FAS-associated death domain protein
FDR	false discovery rate
GO	gene ontology
GPNMB	transmembrane glycoprotein NMB
IARC	International Agency for Research on Cancer
IGF1R	insulin-like growth factor 1 receptor
IPA	ingenuity pathway analysis
IQR	interquartile range
ITGB5	integrin beta-5
KLK13	kallikrein 13
LOD	limit of detection
LYN	tyrosine-protein kinase Lyn
METAP2	methionine aminopeptidase 2
mRNA	messenger RNA
NPX	Normalized Protein eXpression
PAH	polycyclic aromatic hydrocarbons
PCA	principal component analyses
PEA	proximity extension assay
qPCR	quantitative polymerase chain reaction
S100A4	S100 calcium-binding protein A4
S100A11	S100 calcium-binding protein A11
SCAMP3	secretory carrier-associated membrane protein-3
SYND1	syndecan-1
TXLNA	alpha-taxilin
VIM	vimentin

and Denmark. The authors found that the incidence of cancer among chimney sweeps was one of the highest in men; in addition, they showed that lifestyle factors were an unlikely explanation for the increased incidence of colon cancer and pancreatic cancer among chimney sweeps (4). Swedish registry-based studies reported increased incidences of cancers of the lung, bladder, prostate, oesophagus, liver, colon and pleura, as well as hematopoietic cancers (5,6). In addition, the incidence of total cancer and bladder cancer among chimney sweeps showed exposure–response relationships (6). Another study of Swedish chimney sweeps found increased risk of mortality from all cancer forms and site-specific malignancies, e.g. lung, oesophagus, bowel and liver cancers (7). The authors concluded that smoking and alcohol consumption could partially explain the findings, but occupational exposure to soot is the probable cause for the increased cancer risk.

Soot is a black substance produced from incomplete combustion of organic matter and is rich in polycyclic aromatic hydrocarbons (PAH), several of which are carcinogens or probable carcinogens (8,9). Indeed, the International Agency for Research on Cancer (IARC) has classified ‘soot, as found in occupational exposure of chimney sweeps’ as carcinogenic to humans (group 1) (9). Chimney sweeps are mostly exposed to soot while carrying out soot-sweeping

tasks, such as removing soot from chimneys or boilers in private houses or industrial buildings (10). We recently examined PAH exposure in working chimney sweeps, measuring exposure by quantifying monohydroxy metabolites of PAH in urine; median 1-hydroxypyrene (1-OH-PYR) concentrations were seven times higher in chimney sweeps compared with unexposed controls (11). Further, the chimney sweeps demonstrated epigenetic changes related to lung cancer in their peripheral blood (12). However, the relationship between early cancer-related markers and soot or PAH exposure remains widely unexplored. In this study, we explored the profiles of 90 cancer-related proteins in chimney sweeps. We further aimed to elucidate the associations between cancer-related proteins and PAH exposure. Screening of early markers associated with occupational cancer and understanding their relationships with occupational exposure provides important opportunities for reducing cancer morbidity and promoting public health.

**Materials and methods****Study participants**

Participants were current non-smokers selected from the parent study, which included 151 chimney sweeps and 152 control workers. We excluded current smokers because tobacco smoke contains a mixture of carcinogens and is a major source of exposure to PAH (13). We included 118 chimney sweeps and 126 control workers in this study (Supplementary Figure 1, available at *Carcinogenesis* Online). Details of the enrolment process have been published (11). Briefly, male chimney sweeps and matched control workers [in age, body mass index (BMI), gender and tobacco smoking] were recruited from southern Sweden during 2010–15 (more controls were recruited at the beginning than chimney sweeps). The controls were recruited from companies (food storage, waste management, housing and goods transport) and municipalities with no occupational exposure to PAH. We started the enrolment by contacting managers of companies who were instructed to ask their employees about participation. The company response rate was 87% for chimney sweeps and 58% for controls. Non-participating companies did not differ from participating ones with regard to the type of service provided, company size or location.

A trained nurse recorded weight and height measurements, drew blood samples and collected questionnaires and urine samples from all participants. The Regional Ethics Committee in Lund (Sweden) approved the study and all participants gave written informed consent.

**Questionnaire**

Participants filled in questionnaires about age, education, personal history of disease, family history of cancer, prescribed and non-prescribed medicines, level of physical activity, diet (intake of vegetables, fruits and fish), tobacco smoking status, use of snus, passive smoking, area of residence, employment history and exposure to smoke, diesel exhaust or dust from hobbies. The questionnaire for the chimney sweeps was supplemented with questions on occupational details, e.g. use of protective equipment and estimated percentage of working time that involved soot sweeping, such as soot sweeping in private houses, and percentage that involved other non-soot sweeping tasks, such as cleaning ventilation channels. More details are provided in a previous publication (11).

**Sampling and handling of blood and urine**

Blood and urine samples were collected from all participants on Wednesdays or Thursdays after the work shift. Venous blood was collected in BD Vacutainer SST tubes (Becton, Dickinson and Company, Plymouth, UK) following a standard protocol. Blood samples were allowed to clot in room temperature for 10 min and then centrifuged onsite at 1800g for 15 min. Upon separation, serum samples were aliquoted and kept on dry ice for transportation to the laboratory of Division of Occupational and Environmental Medicine at Lund University and then stored at –80°C until they were used for analysis. Post-shift urine samples (100–500 ml) were collected from participants, transported at room temperature to the laboratory and then stored at –20°C until they were used for analysis.

## Measurement of PAH metabolites in urine

Monohydroxylated metabolites of the PAH pyrene (four aromatic rings, 'not classifiable as to its carcinogenicity to humans' according to IARC), benzo[a]pyrene (five aromatic rings, 'carcinogenic to humans' according to IARC), phenanthrene (three aromatic rings, 'not classifiable as to its carcinogenicity to humans' according to IARC) and benzo[a]anthracene (four aromatic rings, 'possibly carcinogenic to humans' according to IARC) were measured in urine. Liquid chromatography coupled to tandem mass spectrometry (QTRAP 5500, AB Sciex, Foster City, CA) was used for analyses. For each PAH, the resulted peaks could be represented by several isomers of monohydroxy metabolites. For example, the peak for monohydroxy phenanthrene (2-OH-PH) represented the sum of several OH-PH isomers. However, the quantification was performed using a pure 2-OH-PH standard. Similarly, pure standards of 1-OH-PYR, 3-hydroxybenzo[a]pyrene (3-OH-BaP) and 3-hydroxybenzo[a]anthracene (3-OH-BaA) were used for quantification of monohydroxy metabolites of respective PAH. Thus, the measurement of 1-OH-PYR, 2-OH-PH, 3-OH-BaP and 3-OH-BaA served as a proxy for exposure to the parent PAH. Details of the measurement were described elsewhere (11). In short, the urine samples (fresh aliquots) were treated with  $\beta$ -glucuronidase (*Escherichia coli* K12) and isotopically labelled internal standards for each compound were added. Thereafter, 5  $\mu$ l of the sample was injected for analysis of 1-OH-PYR and 2-OH-PH, and 20  $\mu$ l of the sample was injected for analysis of 3-OH-BaP and 3-OH-BaA onto a two-dimensional liquid chromatography system with two columns. All samples were analysed in duplicates and the average value was calculated for each duplicate and used for statistical analysis. Urinary creatinine was measured (g/l) and used for adjustment of PAH metabolite concentrations in spot urine samples (14).

## Sample preparation and measurement of protein biomarkers in serum

Serum samples were analysed on two occasions (batch I and batch II). Batch I comprised 59 samples (29 chimney sweeps and 30 controls) and batch II comprised 204 samples (100 chimney sweeps and 104 controls), including 16 samples from batch I. In the end, three samples (2 chimney sweeps and 1 control) failed to pass the quality control, leaving 244 samples (118 chimney sweeps and 126 controls) for final analysis (Supplementary Figure 2, available at *Carcinogenesis* Online).

Aliquots of serum samples (40  $\mu$ l) from chimney sweeps and controls were randomized in 96-well plates (Thermo Fisher Scientific, Waltham, MA) and covered with MicroAmp Clear Adhesive Film (Thermo Fisher Scientific). To calculate intra- and inter-assay coefficients of variation (%CV), a pooled serum sample from randomly selected chimney sweeps and controls was prepared and shipped on dry ice, along with the 96-well plates, to Olink Proteomics (Uppsala, Sweden). All serum samples had undergone one freeze-thaw cycle before the proteomic analysis.

The 92 cancer-related protein biomarkers in serum were measured simultaneously in the same sample using the multiplex immunoassay Oncology II panel (Olink Proteomics). The measurement is based on the proximity extension assay (PEA) technology (15). In brief, 92 protein-specific pairs of antibodies labelled with unique complementary oligonucleotides (PEA probes) are added to 1  $\mu$ l of serum sample. Each pair of antibodies targets one protein. Once the pair of antibodies is attached to its target, the unique PEA probes are in close proximity, allowing hybridization and producing a DNA template that can act as a marker for the protein. The template is extended by DNA polymerase and pre-amplified by quantitative polymerase chain reaction (qPCR) using universal primers. Surplus primers are digested prior to quantification by microfluidic chip (96.96 Dynamic Array IFC, Fluidigm Biomark) run in a qPCR platform (BioMark HD System). More details are available online at (<https://www.olink.com>).

## Quality control and normalization of proteomic analyses

The validity of the proteomic analysis was tested by analysing internal controls: (i) incubation controls 1 and 2 (both non-human proteins) were added to each serum sample to monitor protein-antibody reactions; (ii) an extension control (antibody labelled with PEA probes) was added to each sample to monitor the DNA extension step and (iii) a detection control

(pre-designed double-stranded DNA) was added to monitor detection quality of the qPCR. External controls were also added as follows: (i) an artificial sample containing the 92 proteins (inter-plate control) was run in triplicate in each plate to monitor inter-plate variability and for normalization; (ii) negative controls were run in triplicate in each plate and the background signal was used to calculate the limit of detection (LOD) for each protein (LOD = background value + 3 standard deviations) and (iii) a pooled serum sample was run in duplicate in each plate to calculate intra- and inter-assay %CV. For quality control, if the signal of incubation control 2 (the more sensitive of the incubation controls) and/or the detection control signal deviated by >0.3 from the signal of all samples, then the sample was excluded from further analysis. In total, three samples (two chimney sweeps and one control) did not pass these criteria and therefore were excluded from the analysis. All proteins had intra- and inter-assay %CV <20%, except for XPNPEP2 and CRNN, and therefore, these two proteins were excluded from further analysis (total number of proteins used in downstream analyses  $n = 90$ ). Protein concentration was measured in arbitrary units on a log<sub>2</sub> scale and termed Normalized Protein eXpression (hereafter referred to as NPX). This means for every 1 unit increase in NPX, the protein concentration doubles. Since the NPX value represents a relative quantification, the data for a specific protein can be compared across different samples (individuals), but different proteins cannot be compared in a specific individual. Adjustment for batch effect (batch I and batch II) was performed using the data of the samples that were run in both batches: for each protein (assay), we calculated the median value of NPX for batch I and II. We then calculated the NPX median difference (for each protein) and subtracted it from the NPX values for each sample.

Out of the 90 proteins, three showed signals below LOD in 0.4–0.8% of the samples. NPX values below LOD were replaced with the LOD value of the respective protein assay.

## Pathway and gene ontology enrichment analyses

Ingenuity pathway analysis (IPA) software (Ingenuity IPA - 42012434, Qiagen, Hilden, Germany) was employed to explore relevant molecular networks, upstream regulators, diseases and biological functions associated with the top 17 differentially expressed proteins (DEP) (FDR < 0.05). Log<sub>2</sub> (fold change) was calculated for the proteins and core analysis was created according to expression analysis and expression log ratio.

The webtool WebGestalt (<http://www.webgestalt.org>) was used for gene ontology (GO) enrichment analysis of the 17 DEP. The basic parameters were set as follows: organism (hsapiens), method [OverRepresentation enrichment analysis (ORA)], functional database (geneontology), gene ID type (uniprot\_swissprot) and reference set for enrichment analysis (genome\_protein-coding). Three enrichment categories were checked, i.e. biological process, cellular component and molecular function.

## Statistical analyses

Age was calculated based on date of birth and date of recruitment. BMI was obtained from the ratio weight (kilograms)/height (metres) squared. No current smokers were included in the study; therefore, smoking status was allocated into two categories; party/former smoker and non-smoker. Fisher's Exact Test and Mann-Whitney *U*-test were used to examine differences between chimney sweeps and controls for categorical and continuous variables, respectively. Median and interquartile range (IQR) were calculated for continuous variables and  $n(\%)$  for categorical variables stratified by study group. PAH metabolites were positively skewed, and therefore, log<sub>2</sub>-transformation was applied to improve the distribution. Serum protein concentrations were reported as NPX on a log<sub>2</sub> scale. XPNPEP2 and CRNN proteins were excluded from the analysis as the inter-assay %CV > 20% (Supplementary Table 1, available at *Carcinogenesis* Online).

We have conducted two different principal component analyses (PCA). First, a PCA was performed to test for potential clustering of study groups (chimney sweeps versus controls) including the 17 DEP. Second, principal component regression analysis was carried out to identify potential confounders and covariates influencing the associations between study groups and the 90 serum protein concentrations (16). The top five principal components, which explained 42% of the data variability, were regressed against the following covariates; age, BMI, smoking status (party/former

smoker versus non-smoker), use of snus (yes/no), season of sampling (Spring, Summer, Fall, Winter), plate number (proteomic analysis), passive smoking (yes/no), intake of vegetables (high/low), intake of fruits (high/low), intake of fish (high/low), physical activity (high/low), residential area (big city/small city), level of education (university or higher/high school or lower), history of cardiovascular disease (yes/no), family history of cancer (yes/no), use of prescribed medicines (yes/no) and exposure to smoke from hobbies (yes/no). Thereafter, all covariates that were significantly associated with at least one of the five principal components were considered in the Kolmogorov–Smirnov test (KS test). This test examines the distributions of P-values of the associations between study groups and serum proteins (differential expression analysis) before and after deselecting covariates from the model in a stepwise fashion (dropping one covariate at a time; starting from covariates associated with principal component 1, then 2 and so forth). A  $P > 0.05$  from the KS test indicates that the deselected covariate does not significantly influence the association of the differential expression analysis. None of the covariates was found to be influential by this procedure (Supplementary Table 4 and Supplementary Figure 3, available at Carcinogenesis Online). We fit two general linear models to examine the associations between serum protein concentrations and study groups (differential expression analysis): Model I (unadjusted) and Model II (adjusted for age and BMI, the latter for comparison with other studies). FDR correction was applied with a sample size of 90 considered for multiple testing ( $FDR < 0.05$ ). In order to compare our data with previous studies, we evaluated the effect of several covariates on serum proteins in the controls because PAH exposure in chimney sweeps can mask the potential associations between the covariates and serum proteins. Spearman's correlations combined with FDR adjustment was used for continuous covariates (age and BMI), while Kruskal–Wallis test or Mann–Whitney U-test combined with FDR adjustment was used for categorical covariates (use of snus, physical activity, residential area, sampling season and intake of vegetables, fruits and fish).

We could not adjust for storage time in the differential expression analysis, as storage time was strongly associated with study groups (i.e. in general, older samples were from the controls and recent samples were from chimney sweeps) and therefore adjustment for storage time would conceal the effect of exposure (study group) on serum proteins. Nevertheless, since the sampling of all chimney sweeps and all controls lasted for about 15–16 months, we examined the potential effect of this storage time on serum protein concentrations among the controls and chimney sweeps separately by Spearman's correlations adjusting for FDR.

Correlations between PAH metabolites ( $\log_2$ -transformed) and concentrations (NPX) of the 17 DEP were evaluated by Spearman's correlations for chimney sweeps and controls (all together and per group) and presented as a heatmap. General linear models adjusted for age, BMI and storage time were further employed to explore the associations between PAH metabolites and serum protein concentrations (NPX) among chimney sweeps only. To examine dose–response relationships between serum protein levels and PAH exposure (PAH metabolites), boxplots for protein concentrations across quartiles of PAH metabolite concentrations were plotted and the Jonckheere–Terpstra trend test was applied.

All statistical analyses were performed using SPSS version 23 (IBM SPSS Statistics, NY, USA) and R (version 3.5.1).  $P < 0.05$  was considered for statistical significance.

## Results

### Characteristics of the study participants

Table 1 summarizes the basic characteristics of the 118 chimney sweeps and 126 control participants in this study. Chimney sweeps did not significantly differ from the controls in age, BMI, smoking (non-smokers versus party/ex-smokers), passive smoking, season in which the samples were collected, exposure to PAH from hobbies, consumption of different foods (vegetables, fruits and fish), level of education, personal history of disease and prescribed and non-prescribed medicines ( $P >$

0.05). Chimney sweeps, compared with controls, reported higher physical activity during leisure time ( $P = 0.03$ ), higher use of snus ( $P = 0.003$ ) and resided in smaller cities ( $P = 0.02$ ). Comparison of the characteristics of the participants selected in this study versus the full cohort showed no substantial differences for the majority of characteristics apart from the controls' higher level of education (university or higher): 13.5% (this study) versus 21% (full cohort).

As previously reported (11), median urinary concentrations of 1-OH-PYR, 2-OH-PH, 3-OH-BaP and 3-OH-BaA were up to seven times higher among chimney sweeps compared with controls ( $P < 0.001$ ) (Table 1).

### Correlations between covariates and serum protein concentrations

Principal component regression analysis combined with KS test showed that none of the potential confounders/covariates was influential on the linear regression models comparing chimney sweeps versus controls; largely because the two study groups were more or less similar with regard to age, BMI and lifestyle factors (Table 1 and Supplementary Table 4, available at Carcinogenesis Online). In the controls, use of snus and residential area did not show any significant correlations with serum protein concentrations, evaluated by Mann–Whitney U-test combined with FDR adjustment (Supplementary Table 6, available at Carcinogenesis Online). Physical activity appeared to influence nine proteins, while season of sampling and intake of vegetables, fruits and fish showed no effect on serum protein concentrations. We also evaluated the effect of age and BMI on serum proteins among the controls and found that 18 proteins showed correlations with age, of which 12 were positive, and 15 proteins showed correlations with BMI, of which 13 were positive (Supplementary Table 7, available at Carcinogenesis Online).

We did not observe a profound effect of storage time (up to 16 months) on serum proteins. Only syndecan-1 (SYND1) was significantly affected by storage time, showing increasing concentrations with longer storage time among the controls (Supplementary Table 8, available at Carcinogenesis Online). Similarly, one protein (EGF; pro-epidermal growth factor) was influenced by storage time among chimney sweeps (Supplementary Table 8, available at Carcinogenesis Online).

### Several cancer-related proteins in serum were differentially expressed between chimney sweeps and controls

Seventeen proteins were significantly differentially expressed in chimney sweeps compared with controls, as evaluated by linear regression analysis adjusted for age, BMI and FDR ( $FDR < 0.05$ ; Table 2 and Supplementary Table 3, available at Carcinogenesis Online). Further adjustment for physical activity, use of snus and residential area did not significantly influence the estimates. The 17 DEP included, among others, kallikrein 13 (KLK13), S100 calcium-binding protein A4 (S100A4), FAS-associated death domain protein (FADD), methionine aminopeptidase 2 (METAP2), annexin A1 (ANXA1), S100 calcium-binding protein A11 (S100A11), vimentin (VIM), alpha-taxilin (TXLNA), tyrosine-protein kinase Lyn (LYN) and secretory carrier-associated membrane protein-3 (SCAMP3). All DEP were upregulated in chimney sweeps compared with the controls, apart from SYND1, which showed lower levels in chimney sweeps. Furthermore, PCA showed partial separation between chimney sweeps and

**Table 1.** Characteristics of controls and chimney sweeps

Continuous variables	Controls n = 126	Chimney sweeps n = 118	P <sup>k</sup>
	Median (IQR)	Median (IQR)	
Age (years)	43 (14)	41 (20)	0.908
BMI (kg/m <sup>2</sup> )	27 (4.9)	26.7 (4.9)	0.317
1-OH-PYR (µg/g creatinine)	0.05 (0.04)	0.38 (0.67)	<0.001
2-OH-PH (µg/g creatinine)	0.12 (0.08)	0.53 (0.7)	<0.001
3-OH-BaP (ng/g creatinine)	0.92 (1.9)	3.00 (3.6)	<0.001
3-OH-BaA (ng/g creatinine)	1.56 (1.1)	4.39 (5.1)	<0.001
Categorical variables <sup>*</sup>	n (%) <sup>#</sup>	n (%) <sup>#</sup>	
Sampling season (Spring/Summer/ Fall/Winter)	41 (33)/25 (20)/52 (41)/8 (6)	33 (28)/15 (13)/60 (51)/10 (8)	0.274
Former or party smoker	44 (34.9)	47 (39.8)	0.508
Non-smoker	82 (65.1)	71 (60.2)	
Use of snus	24 (19)	43 (36.4)	0.003
Passive smoking	16 (12.7)	22 (18.6)	0.218
Exposure to smoke from hobbies	5 (4)	9 (7.6)	0.275
Intake of vegetables (≥5 times a week) <sup>a</sup>	82 (65.1)	82 (69.5)	0.584
Intake of fruits (≥5 times a week) <sup>b</sup>	78 (61.9)	65 (55.1)	0.297
Intake of fish (≥ once a week) <sup>c</sup>	60 (47.6)	58 (49.2)	0.898
Physical activity (high) <sup>d</sup>	52 (41.3)	66 (55.9)	0.029
Area of residence (big city) <sup>e</sup>	57 (45.2)	35 (29.7)	0.017
Education (university or higher)	17 (13.5)	21 (17.8)	0.383
Diabetes <sup>f</sup>	4 (3.2)	1 (0.8)	0.371
Cancer <sup>g</sup>	2 (1.6)	6 (5.1)	0.161
Personal history of cardiovascular disease <sup>h</sup>	25 (19.8)	32 (27.1)	0.226
Prescribed medicine <sup>i</sup>	25 (19.8)	34 (28.8)	0.134
Non-prescribed medicine <sup>j</sup>	27 (21.4)	20 (16.9)	0.419

<sup>a</sup>Intake of all kinds of vegetables, legumes and root vegetables (fresh, frozen, canned, stewed, juice, soup, etc.).

<sup>b</sup>Intake of all kinds of fruits and berries (fresh, frozen, canned, juice, jam, etc.).

<sup>c</sup>Intake of all kinds of fish.

<sup>d</sup>Once a week or more of at least 30 min of regular physical activity (running, swimming, tennis, badminton or similar sport activities).

<sup>e</sup>Big city compared with small city (town or a village).

<sup>f</sup>Participants were asked 'Do you have/have you had diabetes?'

<sup>g</sup>Participants were asked 'Do you have/have you had cancer?'

<sup>h</sup>Hypertension, myocardial infarction, angina pectoris, stroke, thrombosis in the arm or leg, arrhythmia or myocarditis.

<sup>i</sup>Predominantly included medication for cardiovascular disease, inflammation, lowering blood lipids, gastroesophageal reflux and asthma.

<sup>j</sup>Vitamins, analgesics and omega 3 fatty acids.

<sup>k</sup>P-value of Mann-Whitney U-test for continuous variables and Fisher's Exact test for categorical variables.

<sup>\*</sup>Variables were categorized into (yes/no), unless otherwise stated.

<sup>#</sup>There was one missing case for some of the categorical variables.

**Table 2.** Top 17 DEP between chimney sweeps and controls (reference group) measured as NPX values and explored by general linear models

Protein	Model I (unadjusted)		Model II (adjusted for age and BMI)	
	P	B (95% CI)	P	B (95% CI)
S100A4	<0.001	0.31 (0.22, 0.39)	<0.001	0.31 (0.22, 0.39)
FADD	<0.001	0.31 (0.21, 0.40)	<0.001	0.30 (0.20, 0.39)
METAP2	<0.001	0.31 (0.21, 0.42)	<0.001	0.33 (0.22, 0.43)
ANXA1	<0.001	0.52 (0.36, 0.67)	<0.001	0.50 (0.35, 0.66)
S100A11	<0.001	0.22 (0.14, 0.30)	<0.001	0.21 (0.13, 0.30)
VIM	<0.001	0.47 (0.28, 0.66)	<0.001	0.46 (0.27, 0.64)
TXLNA	<0.001	0.26 (0.12, 0.40)	0.001	0.26 (0.11, 0.40)
LYN	<0.001	0.12 (0.06, 0.18)	<0.001	0.12 (0.06, 0.18)
SCAMP3	0.001	0.21 (0.09, 0.33)	0.001	0.21 (0.09, 0.33)
SYND1	0.001	-0.17 (-0.28, -0.07)	0.002	-0.17 (-0.27, -0.06)
KLK13	0.004	0.16 (0.05, 0.27)	0.005	0.16 (0.05, 0.26)
IGF1R	0.004	0.11 (0.03, 0.18)	0.004	0.11 (0.03, 0.18)
TCL1A	0.007	0.32 (0.09, 0.55)	0.008	0.31 (0.08, 0.53)
CXCL13	0.004	0.16 (0.05, 0.28)	0.004	0.17 (0.05, 0.28)
ITGB5	0.003	0.09 (0.03, 0.16)	0.002	0.09 (0.03, 0.15)
GPNUMB	0.002	0.06 (0.02, 0.10)	0.003	0.06 (0.02, 0.10)
CEACAM5	0.005	0.26 (0.08, 0.44)	0.007	0.25 (0.07, 0.43)

Effect estimates are presented as B-values and 95% CI. All presented proteins are statistically significant after adjustment for FDR.

controls based on the principal components PC1 and PC2 that explained 37.5 and 18.5% of the variation between the study groups, respectively (Supplementary Figure 4, available at Carcinogenesis Online).

### PAH metabolites in urine were associated with serum protein levels

PAH metabolite concentrations in urine showed weak to moderate correlations with the 17 DEP, as evaluated by Spearman's correlations (Figure 1a). Except for SYND1, the DEP showed positive correlations with PAH metabolite concentrations among chimney sweeps. This is in line with the results from the differential expression analysis in Table 2, which showed upregulation of most of the DEP in chimney sweeps compared with controls. The correlations were less consistent among the controls compared with the correlations for chimney sweeps and for study participants all together (Figure 1a).

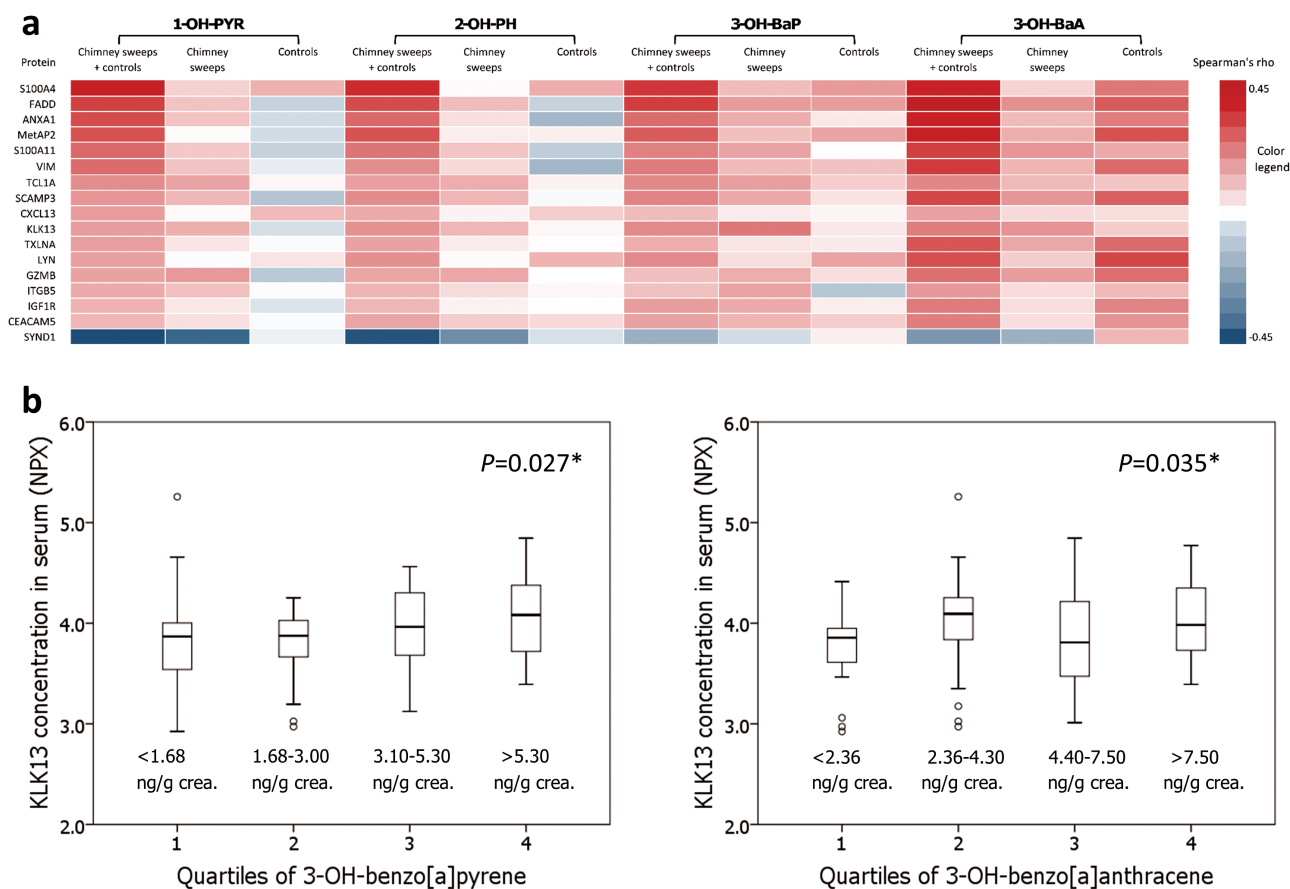
Results from linear regression analysis adjusted for age, BMI and storage time (Table 3) were in line with those from Spearman's correlations. Out of the 17 DEP, 13 (upregulated in chimney sweeps) showed positive associations with all PAH metabolites (Table 3). In addition, SYND1 (downregulated in chimney sweeps) was inversely associated with the concentrations of all PAH metabolites. Notably, the two PAH metabolites (3-OH-BaP and 3-OH-BaA) were significantly

positively associated with serum KLK13 concentrations among chimney sweeps (Table 3 and Supplementary Figure 5, available at Carcinogenesis Online). Further analysis showed a dose-response relationship between KLK13 and PAH metabolite concentrations when comparing KLK13 concentrations across quartiles of 3-OH-BaP and 3-OH-BaA concentrations in urine ( $P < 0.035$ , Jonckheere trend test, Figure 1b); the dose-response was more apparent for 3-OH-BaP. Among the controls, there was no significant trend for PAH metabolites versus KLK13 ( $P > 0.05$ ). FADD, S100A11 and SCAMP3 showed significant trends with 3-OH-BaA ( $P < 0.05$ ) only.

Data from the Comparative Toxicogenomics Database (<http://ctdbase.org>) showed interactions between different PAH and 10 out of 17 of the DEP, including KLK13, which showed increased messenger RNA (mRNA) levels upon exposure to benzo(b)fluoranthene (*in vitro*, mice cells) and decreased mRNA levels upon exposure to benzo(k)fluoranthene (*in vitro*, mice cells) and BaP (*Oikopleura dioica*) (Table 4).

### Pathway and GO enrichment analyses

Pathway analysis for the 17 DEP revealed that interferon gamma was the main upstream regulator with increased predicted activated state (activation z-score  $>2.0$ ) (Supplementary Figure 6, available at Carcinogenesis Online). The top hits of the network analysis included 14 out of 17 DEP, and these proteins showed a



**Figure 1.** Correlations between concentrations of PAH metabolites in urine and serum proteins. (a) Heatmap of Spearman's correlations between the top 17 differentially expressed proteins and PAH metabolite concentrations in urine for chimney sweeps, controls and for both chimney sweeps and controls together. The order of the proteins was set according to the Spearman's rho of 1-OH-PYR versus protein concentrations (for all participants together) in a descending fashion (left-most column). NPX values ( $\log_2$ -transformed) for serum proteins were used. 1-OH-PYR and 2-OH-PH concentrations were  $\mu\text{g/g}$  creatinine, whereas 3-OH-BaP and 3-OH-BaA concentrations were  $\text{ng/g}$  creatinine. (b) Serum concentrations of KLK13 across quartiles of 3-OH-BaP and 3-OH-BaA among chimney sweeps. \* $P$ -value of Jonckheere-Terpstra trend test to examine the hypothesis that KLK13 concentrations in serum are increasing across the quartiles of 3-OH-BaP and 3-OH-BaA.

**Table 3.** Associations between PAH metabolites ( $\log_2$ -transformed) and the top 17 serum proteins (NPX values) for chimney sweeps, explored by general linear model adjusted for age, BMI and storage time

Protein	1-OH-PYR		2-OH-PH		3-OH-BaP		3-OH-BaA	
	P	B (95% CI)	P	B (95% CI)	P	B (95% CI)	P	B (95% CI)
S100A4	0.34	0.018 (-0.019, 0.056)	0.79	0.006 (-0.039, 0.051)	0.15	0.020 (-0.007, 0.046)	0.36	0.024 (-0.028, 0.075)
FADD	0.22	0.026 (-0.015, 0.067)	0.17	0.034 (-0.015, 0.084)	0.24	0.017 (-0.012, 0.046)	0.10	0.046 (-0.009, 0.101)
METAP2	0.99	0.00 (-0.043, 0.043)	0.58	0.015 (-0.037, 0.066)	0.36	0.014 (-0.016, 0.044)	0.07	0.052 (-0.005, 0.110)
ANXA1	0.31	0.033 (-0.031, 0.097)	0.46	0.029 (-0.049, 0.106)	0.60	0.012 (-0.034, 0.058)	0.59	0.024 (-0.064, 0.113)
S100A11	0.19	0.026 (-0.013, 0.065)	0.15	0.035 (-0.012, 0.081)	0.11	0.023 (-0.005, 0.051)	0.07	0.049 (-0.004, 0.102)
VIM	0.27	0.042 (-0.033, 0.117)	0.27	0.050 (-0.040, 0.141)	0.35	0.025 (-0.028, 0.079)	0.31	0.053 (-0.050, 0.156)
TXLNA	0.72	0.011 (-0.048, 0.069)	0.61	0.018 (-0.053, 0.089)	0.56	0.012 (-0.029, 0.054)	0.16	0.057 (-0.023, 0.136)
LYN	0.70	0.006 (-0.026, 0.038)	0.56	0.011 (-0.027, 0.050)	0.18	0.015 (-0.007, 0.038)	0.29	0.023 (-0.020, 0.067)
SCAMP3	0.14	0.040 (-0.014, 0.093)	0.15	0.047 (-0.017, 0.112)	0.14	0.028 (-0.009, 0.066)	0.08	0.063 (-0.008, 0.135)
SYND1	0.03	-0.053 (-0.101, -0.004)	0.15	-0.043 (-0.102, 0.016)	0.64	-0.008 (-0.044, 0.027)	0.41	-0.028 (-0.097, 0.04)
KLK13	0.14	0.037 (-0.012, 0.085)	0.23	0.035 (-0.023, 0.094)	0.02	0.042 (0.008, 0.076)	0.04	0.068 (0.002, 0.134)
IGF1R	0.93	-0.002 (-0.040, 0.037)	0.76	-0.007 (-0.053, 0.039)	0.12	0.022 (-0.006, 0.049)	0.38	0.024 (-0.029, 0.076)
TCL1A	0.14	0.079 (-0.025, 0.182)	0.26	0.066 (-0.054, 0.197)	0.07	0.066 (-0.006, 0.139)	0.22	0.090 (-0.053, 0.233)
CXCL13	0.90	0.004 (-0.053, 0.060)	0.87	0.006 (-0.062, 0.074)	0.54	0.012 (-0.027, 0.052)	0.55	0.023 (-0.054, 0.100)
ITGB5	0.20	0.017 (-0.009, 0.042)	0.33	0.015 (-0.016, 0.046)	0.17	0.012 (-0.005, 0.030)	0.57	0.010 (-0.025, 0.046)
GPNMB	0.95	0.001 (-0.015, 0.017)	0.56	-0.006 (-0.025, 0.014)	0.25	0.007 (-0.005, 0.018)	0.74	0.004 (-0.018, 0.026)
CEACAM5	0.99	-0.001 (-0.09, 0.088)	0.75	0.017 (-0.090, 0.124)	0.50	0.022 (-0.042, 0.086)	0.94	-0.005 (-0.129, 0.118)

Effect estimates are presented as B-values and 95% CI.

clear involvement in cell adhesion and cell invasion, which are associated with cancer development (Supplementary Figure 7, available at *Carcinogenesis* Online). Three molecular and cellular functions, (i) cell movement, (ii) migration of cells and (iii) cell movement of tumour cell lines, showed increased predicted activation states (activation z-score >2.0;  $P < 10 \times 10^{-5}$ ) (Table 5 and Supplementary Figure 8, available at *Carcinogenesis* Online).

GO enrichment analysis for the 17 DEP revealed that cell adhesion and cell migration were the top enriched ontologies for the 'biological process' category, adherens and anchoring junctions were the top ontologies for the 'cellular component' category and protein binding was the top ontology for the 'molecular function' category (Supplementary Table 5, available at *Carcinogenesis* Online).

Similar results were obtained from IPA and GO enrichment analysis after conducting a sensitivity analysis excluding SYND1, as this DEP was influenced by storage time.

## Discussion

It has long been known that chimney sweeps have higher incidences and mortalities related to cancer, but little research has so far been dedicated to the detection of early biomarkers of cancer in this occupational group. In the present study, we found a number of putative cancer-related serum proteins that were differentially expressed between PAH-exposed chimney sweeps and control workers, and the majority of the DEP were upregulated in chimney sweeps. Interestingly, most of the upregulated DEP were positively associated with PAH metabolite concentrations among chimney sweeps. In addition, the results from pathway and GO enrichment analyses indicated that the DEP were involved in metastasis-related cellular functions, i.e. cell movement, cell adhesion and cell migration.

In support of a role in the link between exposure to PAH and alteration of cancer-related proteins, KLK13 showed positive associations and a dose-response relationship with metabolites of two PAH that have a carcinogenic potential in humans, i.e. BaP and BaA. KLK13 belongs to the 15-member kallikrein family and is implicated in degradation of the extracellular matrix and

cancer pathology (17,18). Dysregulated KLK13 expression has been reported in a wide spectrum of cancers (18–20). On the one hand, increased mRNA or/and protein expression of KLK13 has been found in human samples of cancers of relevance for PAH exposure, such as non-small-cell lung cancer (21) and colorectal cancer (22). On the other hand, decreased KLK13 expression was found in bladder cancer (23), primary gastric carcinomas (24), testicular cancer (25) and oral squamous cell carcinoma (26). Few studies have measured KLK13 or other kallikreins in serum or plasma samples from cancer patients. Increased serum levels of KLK13, as well as KLK11 and KLK14, were observed in non-small-cell lung cancer patients ( $n = 51$ ) compared with healthy controls ( $n = 50$ ) (27). Further, *in vitro* and *in vivo* studies have shown that cell invasion and migration were enhanced by KLK13 overexpression in human lung adenocarcinoma cell lines but reduced by KLK13 knockdown (28). In contrast, less invasiveness and less cell motility were observed with overexpression of KLK13 in oral squamous cell carcinoma cell lines (26).

To our knowledge, no studies have examined the association between KLK13 expression and PAH exposure in humans, but a limited number of non-human studies have examined this association. A mouse study showed increased expression of KLK13 mRNA with exposure to the PAH benzo(b)fluoranthene and decreased expression with exposure to benzo(k)fluoranthene (29). Moreover, BaP exposure in the tunicate *O. dioica* resulted in decreased KLK13 mRNA expression (30). Similarly, toxicological studies using *in vitro* and *in vivo* models (mouse and rat) have reported inconsistent associations between PAH exposure and expression of METAP2, S100A11, TXLNA, insulin-like growth factor 1 receptor (IGF1R), C-X-C motif chemokine 13 (CXCL13), integrin beta-5 (ITGB5), transmembrane glycoprotein NMB (GPNMB), carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5) and SYND1; increased and decreased mRNA and/or protein levels were observed in association with exposure to several PAH, including BaP and BaA (Table 4). These inconsistencies might be due to differences in the cell lines, animals, dosage or exposure regimens used in the experiments. In this context, measuring serum proteins among individuals

**Table 4.** PAH that interact with the top DEP between chimney sweeps and controls

N	Protein*	PAH	Interaction	Number of studies	Organism
1	SYND1	Benzo(a)pyrene	Decreased expression of SDC1 (gene encoding SYND1) mRNA	3	Mouse (in vitro + in vivo), rat (in vivo)
			Increased expression of SDC1 mRNA	2	Human (in vitro)
			Increased expression of SDC1 protein	1	Rat (in vivo)
2	METAP2	Benzo(a)pyrene	Increased expression of METAP2 mRNA	2	Mouse (in vitro + in vivo)
			Decreased expression of METAP2 mRNA	1	Mouse (in vivo)
3	S100A11	Benzo(a)pyrene	Decreased expression of S100A11 mRNA	2	Mouse (in vitro + in vivo)
			Increased expression of S100A11 mRNA	4	Human (in vitro)
4	TXLNA	Benz(a)anthracene Co-treatment: Benzo(a)pyrene, Benz(a)anthracene, Benzo(b)fluoranthene, Chrysene	Decreased expression of TXLNA mRNA	1	Mouse (in vivo)
			Decreased expression of TXLNA mRNA	1	Mouse (in vivo)
5	KLK13	Benzo(b)fluoranthene Benzo(k)fluoranthene	Decreased expression of TXLNA mRNA	1	Mouse (in vivo)
			Increased expression of KLK13 mRNA	1	Mouse (in vivo)
6	IGF1R	Benzo(a)pyrene	Decreased expression of KLK13 mRNA	1	Mouse (in vivo)
			Decreased expression of IGF1R mRNA	1	Oikopleura dioica
			Aryl hydrocarbon receptor (AHR) protein promotes the reaction [Benzo(a)pyrene results in decreased expression of IGF1R mRNA]	1	Mouse (in vitro)
			Decreased expression of IGF1R mRNA	4	Human (in vitro)
7	CXCL13	Benzo(a)pyrene	Decreased expression of IGF1R mRNA	2	Mouse (in vitro)
			Increased phosphorylation of IGF1R protein	1	Human (in vitro)
			Decreased expression of CXCL13 mRNA	3	Mouse (in vitro), rat (in vivo), human (in vitro)
8	ITGB5	Co-treatment: Benzo(a)pyrene, Benz(a)anthracene, Benzo(b)fluoranthene, Chrysene Benzo(a)pyrene	Decreased expression of CXCL13 mRNA	1	Mouse (in vivo)
			BaP affects the reaction (AHR protein binds to the ITGB5 promoter)	1	Mouse (in vitro)
9	GPNMB	Benzo(a)pyrene	Decreased expression of ITGB5 mRNA	1	Mouse (in vitro)
			Increased expression of ITGB5 mRNA	1	Mouse (in vivo)
			AHR protein affects the reaction [Benzo(a)pyrene affects the expression of GPNMB mRNA]	1	Mouse (in vivo)
10	CEACAM5	Benzo(a)pyrene Co-treatment: Benzo(a)pyrene and 1-methylphenanthrene Co-treatment: Benzo(a)pyrene and dibenzo(a,l)pyrene	Decreased expression of GPNMB mRNA	1	Mouse (in vitro)
			Increased expression of GPNMB mRNA	2	Mouse (in vivo)
			Increased expression of CEACAM5 protein	1	Rat (in vivo)
			Affects the expression of CEACAM5 mRNA	1	Human (in vitro)
			Affects the expression of CEACAM5 mRNA	1	Human (in vitro)

Data were obtained from <http://ctdbase.org/> [accessed 21 May 2018].

\*No data available for S100A4, FADD, ANXA1, VIM, LYN, SCAMP3 and TCL1A.

occupationally exposed to PAH would give a better indication of the effect of PAH exposure on human proteomic profiles.

The DEP, including KLK13, have been shown to be implicated in tumour development, including metastasis. *In vitro* studies using human cancer cell lines have shown associations between cancer development and dysregulated expression of several proteins, such as S100A4 (31), FADD (32), METAP2 (33), VIM (34) and SYND1 (35). Likewise, *in vivo* mouse studies have shown similar associations for S100A4 (31), FADD (36), ANXA1 (37) and LYN (38). Studies on human cancer tissues have found dysregulated expression of S100A11, TXLNA, SCAMP3 and VIM in various malignancies (34,39–41). The disparate expression

profiles of these proteins in various cancers might be attributed to interaction with other proteins, availability of the molecular substrates, tissue-specific expression or different mechanisms underlying tumorigenesis in different organs (18). Another factor is that these proteins might have different roles (stimulation or suppression of tumour growth) at different stages of cancer development (19). Inflammation, which has been linked to PAH exposure, plays an essential role in tumorigenesis by influencing several stages of cancer development (42–45). Our data did not show signs of inflammation since the category inflammatory response, which was predicted to be increased by pathway analysis, appeared towards the bottom of the list of top functions



**Table 5.** All diseases and functions with activation score >2.0 or <-2.0, predicted from IPA using the top 17 DEP between chimney sweeps and controls

Categories	Diseases or functions annotation	P-value*	Predicted activation state	Activation** z-score	Molecules
Cellular movement	Migration of breast cancer cell lines	4.6E-09	Increased	2.21	ANXA1, IGF1R, ITGB5, S100A11, S100A4, SYND1, VIM
Cellular movement	Cell movement of tumour cell lines	1.3E-08	Increased	2.76	ANXA1, CXCL13, FADD, IGF1R, ITGB5, LYN, S100A11, S100A4, SYND1, VIM
Cellular movement	Migration of tumour cell lines	7.5E-07	Increased	2.61	ANXA1, IGF1R, ITGB5, LYN, S100A11, S100A4, SYND1, VIM
Cellular movement	Cell movement	5.6E-06	Increased	2.73	ANXA1, CXCL13, FADD, GPNMB, IGF1R, ITGB5, LYN, S100A11, S100A4, SYND1
Cellular movement	Migration of cells	1.6E-05	Increased	2.76	ANXA1, CXCL13, GPNMB, IGF1R, ITGB5, LYN, S100A11, S100A4, SYND1, VIM
Cellular movement	Chemotaxis	1.9E-05	Increased	2.36	ANXA1, CXCL13, FADD, IGF1R, LYN, S100A4
Cellular movement	Invasion of cells	3.7E-05	Increased	2.42	ANXA1, IGF1R, LYN, S100A11, S100A4, SYND1, VIM
Cellular movement	Invasion of tumour cell lines	9.6E-05	Increased	2.42	ANXA1, IGF1R, LYN, S100A11, S100A4, VIM
Hematological system development and function, tissue morphology	Quantity of leukocytes	3.2E-04	Increased	2.25	CXCL13, FADD, IGF1R, LYN, SYND1, TCL1A
Cardiovascular system development and function, organismal development	Vasculogenesis	1.3E-03	Increased	2.24	ANXA1, IGF1R, METAP2, S100A4, VIM
Inflammatory response	Inflammatory response	1.7E-03	Increased	2.06	ANXA1, CXCL13, LYN, S100A4, SYND1
Organismal survival	Morbidity or mortality	6.3E-04	Decreased	-2.59	ANXA1, FADD, GPNMB, IGF1R, LYN, METAP2, S100A4, SYND1, VIM
Organismal survival	Organismal death	3.1E-03	Decreased	-2.41	ANXA1, FADD, GPNMB, IGF1R, LYN, METAP2, S100A4, VIM

\*Indicates how significant the enrichment of specific function is (regardless of the direction of the protein changes in the dataset).

\*\*Indicates the overall score of a specific function, based on the direction of protein changes in the dataset and thus can predict the activation state of the function (increased when z-score >2.0 and decreased when z-score <-2.0).

associated with the DEP (Table 5). It is important to mention that the list of 90 proteins we measured contained only a limited number of proteins related to inflammation (around 17 proteins; according to information obtained from public bioinformatics databases). It is thus possible that the role of inflammation is more pronounced than it seems to be due to the potential participation of other unmeasured inflammatory proteins.

We could speculate that, in addition to inflammation and other mechanisms, aberrant expression of the DEP, potentially KLK13, plays an important role in PAH-induced carcinogenesis by promoting cell invasion and migration. These findings inform our understanding of the mechanisms underlying carcinogenesis in workers exposed to PAH; however, some caution should be taken in interpreting these results.

Sampling of the controls and chimney sweeps were not carried out simultaneously, which might have contributed to some of the observed differences in serum protein levels. Nevertheless, the time difference between the two samplings was small and all samples were kept at -80°C during the whole period, meaning that the effect of storage time is probably negligible, as discussed below. Due to the nature of the recruitment process, we could only calculate the company response rate (not individual response rate). The lower response rate (58%) among companies, from which we enrolled

the controls, compared with the response rate of chimney sweeping companies (87%) could have caused a selection bias. However, the probable reason for non-participation was the limited time during work shift and not the health status of the employees. Moreover, the non-participating companies did not differ from participating ones with regard to geographical area, company size, or business model. Taken together, we cannot eliminate the possibility of selection bias, but it is likely minimal.

Another limitation is that the measurement of PAH metabolites in urine provides information about recent exposure, not long-term exposure, owing to the short half-lives of these metabolites (4–36 h) (46,47). Measurement of PAH-DNA or PAH-protein adducts can provide estimates for long-term PAH exposure as the half-lives of these adducts range from several days to a few months; however, it was not possible to measure PAH adducts in this study (48). Proteins' half-lives, in turn, can vary considerably from few minutes to several hundreds of hours (49,50). Therefore, the differences in half-lives between PAH metabolites and serum proteins may contribute to the weak associations between PAH metabolites and serum protein levels.

The main strength of our study is the individual data for four different PAH metabolites (exposure data), two of which are metabolites of carcinogenic PAH and the individual data for

90 different proteins in serum for non-smoking PAH-exposed workers and unexposed controls. All samples were handled homogeneously and randomized prior to analysis, which minimized the bias due to technical errors or batch effects.

Although the season of sampling was previously shown to influence a number of proteins in plasma (51), we observed no effect in our study. Also, food intake before sampling should not be of concern according to previous studies (52,53). Storage time and storage temperature of biological samples can have profound effects on protein stability and hence protein measurement. The serum samples used in our study were all stored at  $-80^{\circ}\text{C}$  and underwent one freeze-thaw cycle, which, according to earlier studies, should have no effect on the analysis (54,55). We evaluated the effect of storage time (around 16 months) among the chimney sweeps and controls separately. One DEP (SYND1) and a non-differentially expressed protein (EGF; pro-epidermal growth factor) appeared to be affected by storage time, showing increased and decreased concentrations with storage time, respectively (Supplementary Table 8, available at *Carcinogenesis* Online). As a result, the lower SYND1 concentrations found in chimney sweeps that were sampled later than the controls need to be cautiously interpreted. Enroth *et al.* have evaluated the effect of storage time (up to 30 years at  $-80^{\circ}\text{C}$ ) on 108 plasma proteins using the same analytical method (PEA, Olink) and showed that only one protein (CA-125 or Mucin 16) was significantly influenced by storage time after adjustment for multiple comparisons (51). Therefore, the short storage time in our study is unlikely to have had a significant effect on protein levels. Chimney sweeps and controls had similar age and BMI, and therefore, the differential expression analysis was not influenced after adjustment for these two covariates. In agreement with previous studies, age and BMI affected 18 and 16 proteins, respectively (including METAP2, CEACAM5 and CXL17), when examining the correlations amongst the controls (51,55–58).

## Conclusion

We found a number of putative cancer-related proteins differentially expressed in chimney sweeps: workers occupationally exposed to PAH. This supports a need for more efficient measures to reduce PAH exposure in chimney sweeps. Also, further investigation of cancer biomarkers is warranted in a longitudinal approach in chimney sweeps and other occupational groups exposed to PAH.

## Supplementary material

Supplementary data are available at *Carcinogenesis* online.

## Funding

This study was funded by the Swedish Research Council for Health, Working Life and Welfare (FORTE), AFA Insurance (AFA Försäkring) and Karolinska Institutet.

## Acknowledgements

We thank all participants in this study; the nurses Pia Tallving, Patrice Milton and Eva Assarsson for recruiting the workers in the study and Simon Kebede Merid for helpful comments and support in bioinformatics. We also thank the trade union (Kommunal) and the employer organization of chimney sweeps (Skorstensfejaremästares Riksförbund).

Conflict of Interest Statement: None declared.

## References

1. WHO (2018) Cancer. Available at: <http://www.who.int/cancer/en>. World Health Organization, Geneva, Switzerland. (18 September 2018, date last accessed).
2. Takala, J. *et al.* (2014) Global estimates of the burden of injury and illness at work in 2012. *J. Occup. Environ. Hyg.*, 11, 326–337.
3. Pott, P. (1775) *Chirurgical Observations: Relative to the Cataract, the Polypus of the Nose, the Cancer of the Scrotum, the Different Kinds of Ruptures, and the Mortification of the Toes and Feet.* L. Hawes, W. Clarke, and R. Collins, London, UK.
4. Pukkala, E. *et al.* (2009) Occupation and cancer - follow-up of 15 million people in five Nordic countries. *Acta Oncol.*, 48, 646–790.
5. Evanoff, B.A. *et al.* (1993) Mortality and incidence of cancer in a cohort of Swedish chimney sweeps: an extended follow up study. *Br. J. Ind. Med.*, 50, 450–459.
6. Hogstedt, C. *et al.* (2013) Cancer incidence in a cohort of Swedish chimney sweeps, 1958–2006. *Am. J. Public Health*, 103, 1708–1714.
7. Jansson, C. *et al.* (2012) Mortality among Swedish chimney sweeps (1952–2006): an extended cohort study. *Occup. Environ. Med.*, 69, 41–47.
8. IARC (2010) *Monographs on the Evaluation of Carcinogenic Risks to Humans. Some non-heterocyclic polycyclic aromatic hydrocarbons and some related exposures.* Vol 92. International Agency for Research on Cancer, Lyon, France. 754–773.
9. IARC (2012) *Chemical agents and related occupations: a review of human carcinogens. Soot, as found in occupational exposure of chimney sweeps.* Vol. 100F. International Agency for Research on Cancer, Lyon, France. 209–214.
10. Alhamdow, A. *et al.* (2017) Chimney sweeps in Sweden: a questionnaire-based assessment of long-term changes in work conditions, and current eye and airway symptoms. *Int. Arch. Occup. Environ. Health*, 90, 207–216.
11. Alhamdow, A. *et al.* (2017) Early markers of cardiovascular disease are associated with occupational exposure to polycyclic aromatic hydrocarbons. *Sci. Rep.*, 7, 9426.
12. Alhamdow, A. *et al.* (2018) DNA methylation of the cancer-related genes F2RL3 and AHRR is associated with occupational exposure to polycyclic aromatic hydrocarbons. *Carcinogenesis*, 39, 869–878.
13. Talhout, R. *et al.* (2011) Hazardous compounds in tobacco smoke. *Int. J. Environ. Res. Public Health*, 8, 613–628.
14. Mazzachi, B.C. *et al.* (2000) Reference range and method comparison studies for enzymatic and Jaffé creatinine assays in plasma and serum and early morning urine. *Clin. Lab.*, 46, 53–55.
15. Assarsson, E. *et al.* (2014) Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One*, 9, e95192.
16. Xu, Z. *et al.* (2016) ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. *Nucleic Acids Res.*, 44, e20.
17. Kapadia, C. *et al.* (2004) Human kallikrein 13 involvement in extracellular matrix degradation. *Biochem. Biophys. Res. Commun.*, 323, 1084–1090.
18. Tailor, P.D. *et al.* (2018) Diagnostic and prognostic biomarker potential of kallikrein family genes in different cancer types. *Oncotarget*, 9, 17876–17888.
19. Borgeño, C.A. *et al.* (2004) The emerging roles of human tissue kallikreins in cancer. *Nat. Rev. Cancer*, 4, 876–890.
20. Kalinska, M. *et al.* (2016) Kallikreins - The melting pot of activity and function. *Biochimie*, 122, 270–282.
21. Planque, C. *et al.* (2008) Quantitative RT-PCR analysis and immunohistochemical localization of the kallikrein-related peptidases 13 and 14 in lung. *Biol. Chem.*, 389, 781–786.
22. Talieri, M. *et al.* (2009) The use of kallikrein-related peptidases as adjuvant prognostic markers in colorectal cancer. *Br. J. Cancer*, 100, 1659–1665.
23. Tokas, T. *et al.* (2017) Downregulated KLK13 expression in bladder cancer highlights tumor aggressiveness and unfavorable patients' prognosis. *J. Cancer Res. Clin. Oncol.*, 143, 521–532.

24. Konstantoudakis, G. et al. (2010) Kallikrein-related peptidase 13 (KLK13) gene expressional status contributes significantly in the prognosis of primary gastric carcinomas. *Clin. Biochem.*, 43, 1205–1211.
25. Luo, L.Y. et al. (2003) Human tissue kallikreins and testicular cancer. *APMIS*, 111, 225–232; discussion 232.
26. Ishige, S. et al. (2014) Decreased expression of kallikrein-related peptidase 13: possible contribution to metastasis of human oral cancer. *Mol. Carcinog.*, 53, 557–565.
27. Planque, C. et al. (2008) A multiparametric serum kallikrein panel for diagnosis of non-small cell lung carcinoma. *Clin. Cancer Res.*, 14, 1355–1362.
28. Chou, R.H. et al. (2011) Epigenetic activation of human kallikrein 13 enhances malignancy of lung adenocarcinoma by promoting N-cadherin expression and laminin degradation. *Biochem. Biophys. Res. Commun.*, 409, 442–447.
29. Labib, S. et al. (2016) Comparative transcriptomic analyses to scrutinize the assumption that genotoxic PAHs exert effects via a common mode of action. *Arch. Toxicol.*, 90, 2461–2480.
30. Yadetie, F. et al. (2012) Conservation and divergence of chemical defense system in the tunicate *Oikopleura dioica* revealed by genome wide response to two xenobiotics. *BMC Genomics*, 13, 55.
31. Fei, F. et al. (2017) S100A4 in cancer progression and metastasis: a systematic review. *Oncotarget*, 8, 73219–73239.
32. Zhang, R. et al. (2017) The role of FADD in pancreatic cancer cell proliferation and drug resistance. *Oncol. Lett.*, 13, 1899–1904.
33. Tucker, L.A. et al. (2008) Ectopic expression of methionine aminopeptidase-2 causes cell transformation and stimulates proliferation. *Oncogene*, 27, 3967–3976.
34. Satelli, A. et al. (2011) Vimentin in cancer and its potential as a molecular target for cancer therapy. *Cell. Mol. Life Sci.*, 68, 3033–3046.
35. Nikolova, V. et al. (2009) Differential roles for membrane-bound and soluble syndecan-1 (CD138) in breast cancer progression. *Carcinogenesis*, 30, 397–407.
36. Tourneur, L. et al. (2005) FADD adaptor in cancer. *Med. Immunol.*, 4, 1.
37. Moraes, L.A. et al. (2017) Annexin-A1 enhances breast cancer growth and migration by promoting alternative macrophage polarization in the tumour microenvironment. *Sci. Rep.*, 7, 17925.
38. Nguyen, P.H. et al. (2016) LYN Kinase in the tumor microenvironment is essential for the progression of chronic lymphocytic leukemia. *Cancer Cell*, 30, 610–622.
39. Cross, S.S. et al. (2005) Expression of S100 proteins in normal human tissues and common cancers using tissue microarrays: S100A6, S100A8, S100A9 and S100A11 are all overexpressed in common cancers. *Histopathology*, 46, 256–269.
40. Mashidori, T. et al. (2011) Increased alpha-taxilin protein expression is associated with the metastatic and invasive potential of renal cell cancer. *Biomed. Res.*, 32, 103–110.
41. Zhang, X. et al. (2017) Overexpression of SCAMP3 is an indicator of poor prognosis in hepatocellular carcinoma. *Oncotarget*, 8, 109247–109257.
42. Coussens, L.M. et al. (2002) Inflammation and cancer. *Nature*, 420, 860–867.
43. Curfs, D.M. et al. (2005) Polycyclic aromatic hydrocarbons induce an inflammatory atherosclerotic plaque phenotype irrespective of their DNA binding properties. *FASEB J.*, 19, 1290–1292.
44. Grivennikov, S.I. et al. (2010) Immunity, inflammation, and cancer. *Cell*, 140, 883–899.
45. Alshaarawy, O. et al. (2013) Polycyclic aromatic hydrocarbon biomarkers and serum markers of inflammation. A positive association that is more evident in men. *Environ. Res.*, 126, 98–104.
46. Buckley, T.J. et al. (1992) An examination of the time course from human dietary exposure to polycyclic aromatic hydrocarbons to urinary elimination of 1-hydroxypyrene. *Br. J. Ind. Med.*, 49, 113–124.
47. Jongeneelen, F.J. et al. (1990) Ambient and biological monitoring of cokeoven workers: determinants of the internal dose of polycyclic aromatic hydrocarbons. *Br. J. Ind. Med.*, 47, 454–461.
48. Castaño-Vinyals, G. et al. (2004) Biomarkers of exposure to polycyclic aromatic hydrocarbons from environmental air pollution. *Occup. Environ. Med.*, 61, e12.
49. Doherty, M.K. et al. (2009) Turnover of the human proteome: determination of protein intracellular stability by dynamic SILAC. *J. Proteome Res.*, 8, 104–112.
50. Bachmair, A. et al. (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science*, 234, 179–186.
51. Enroth, S. et al. (2016) Effects of long-term storage time and original sampling month on biobank plasma protein concentrations. *EBioMedicine*, 12, 309–314.
52. Dencker, M. et al. (2017) Effect of food intake on 92 biomarkers for cardiovascular disease. *PLoS One*, 12, e0178656.
53. Dencker, M. et al. (2017) Effect of food intake on 92 neurological biomarkers in plasma. *Brain Behav.*, 7, e00747.
54. Shen, Q. et al. (2018) Strong impact on plasma protein profiles by precentrifugation delay but not by repeated freeze-thaw cycles, as analyzed using multiplex proximity extension assays. *Clin. Chem. Lab. Med.*, 56, 582–594.
55. Björkstén, J. et al. (2017) Stability of proteins in dried blood spot biobanks. *Mol. Cell. Proteomics*, 16, 1286–1296.
56. Larsson, A. et al. (2015) The effects of age and gender on plasma levels of 63 cytokines. *J. Immunol. Methods*, 425, 58–61.
57. Larsson, A. et al. (2015) The body mass index (BMI) is significantly correlated with levels of cytokines and chemokines in cerebrospinal fluid. *Cytokine*, 76, 514–518.
58. Enroth, S. et al. (2014) Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cutoffs. *Nat. Commun.*, 5, 4684.