

Protein networks in induced sputum from smokers and COPD patients

James N Baraniuk¹

Begona Casado¹

Lewis K Pannell²

Peter B McGarvey³

Piera Boschetto⁴

Maurizio Luisetti^{5,†}

Paolo Iadarola⁶

¹Division of Rheumatology, Immunology and Allergy, Georgetown University, Washington, DC,

²Proteomics and Mass Spectrometry Laboratory, Mitchell Cancer Center, University of South Alabama,

Mobile, AL, ³Innovation Center for Biomedical Informatics, Georgetown University, Washington, DC, USA;

⁴Department of Medical Sciences, University of Ferrara, Ferrara, ⁵SC Pneumologia, Dipartimento Medicina Molecolare, Fondazione IRCCS Policlinico San Matteo, ⁶Lazzaro Spallanzani Department of Biology and Biotechnology, University of Pavia, Pavia, Italy

[†]Maurizio Luisetti passed away on October 20, 2014

Correspondence: James N Baraniuk
Division of Rheumatology, Immunology and Allergy, Georgetown University,
3800 Reservoir Road Northwest,
Washington, DC 20007-2197, USA
Tel +1 202 687 2906
Fax +1 202 687 9886
Email baraniuj@georgetown.edu

Rationale: Subtypes of cigarette smoke-induced disease affect different lung structures and may have distinct pathophysiological mechanisms.

Objective: To determine if proteomic classification of the cellular and vascular origins of sputum proteins can characterize these mechanisms and phenotypes.

Subjects and methods: Individual sputum specimens from lifelong nonsmokers (n=7) and smokers with normal lung function (n=13), mucous hypersecretion with normal lung function (n=11), obstructed airflow without emphysema (n=15), and obstruction plus emphysema (n=10) were assessed with mass spectrometry. Data reduction, logarithmic transformation of spectral counts, and Cytoscape network-interaction analysis were performed. The original 203 proteins were reduced to the most informative 50. Sources were secretory dimeric IgA, submucosal gland serous and mucous cells, goblet and other epithelial cells, and vascular permeability.

Results: Epithelial proteins discriminated nonsmokers from smokers. Mucin 5AC was elevated in healthy smokers and chronic bronchitis, suggesting a continuum with the severity of hypersecretion determined by mechanisms of goblet-cell hyperplasia. Obstructed airflow was correlated with glandular proteins and lower levels of Ig joining chain compared to other groups. Emphysema subjects' sputum was unique, with high plasma proteins and components of neutrophil extracellular traps, such as histones and defensins. In contrast, defensins were correlated with epithelial proteins in all other groups. Protein-network interactions were unique to each group.

Conclusion: The proteomes were interpreted as complex "biosignatures" that suggest distinct pathophysiological mechanisms for mucin 5AC hypersecretion, airflow obstruction, and inflammatory emphysema phenotypes. Proteomic phenotyping may improve genotyping studies by selecting more homogeneous study groups. Each phenotype may require its own mechanistically based diagnostic, risk-assessment, drug- and other treatment algorithms.

Keywords: cigarette smokers, chronic bronchitis, emphysema, proteomics, mucous hypersecretion, mucin 5AC, neutrophil extracellular nets

Introduction

Cigarette smoking leads to several clinical phenotypes, including COPD. Approximately half of smokers develop chronic bronchitis with mucus hypersecretion,¹ while approximately 15% develop obstruction to airflow defined by a forced expiratory volume in 1 second (FEV₁) <80% of predicted and a ratio of FEV₁ to forced vital capacity <0.70.² A proportion of COPD subjects develop the alveolar destruction of emphysema evident on computed tomography (CT) scans.³ These criteria define a phenotypic framework of: 1) nonsmokers (Non), 2) healthy smokers without airflow obstruction (HS), 3) chronic bronchitis subjects with mucus hypersecretion but normal airflow (CB), 4) COPD subjects with airflow obstruction (COPD), and 5) emphysema subjects with airflow obstruction (E&C).

The pathophysiological mechanisms responsible for each clinically relevant phenotype are poorly defined.

The proteins and other mediators in induced sputum⁴ may predict disease phenotype and illness severity. For example, outcomes such as FEV₁/forced vital capacity, sputum eosinophilia, and metalloprotease 9 are more severely altered in the E&C subset of smokers than CB or HS subjects.^{5,6} Specific phenotype-related proteomic patterns may be the result of discrete combinations of airway secretory, exudative, and inflammatory dysfunction.⁷ Immunohistological and functional evidence suggests that discrete patterns of dysfunction may involve: 1) production of secretory IgA (sIgA), 2) submucosal gland serous cell protein exocytosis exemplified by BPIFB1,⁸ 3) epithelial cell secretion, 4) goblet-cell MUC5AC release in small airways and after goblet-cell hyperplasia,⁹ 5) exocytosis of glandular and ductal mucus products (eg, MUC5B) in larger bronchi, 6) extravasation of plasma proteins that originate from distant B lymphocytes (eg, IGHG1) and the liver (eg, ALB, TF, and HP), 7) neutrophil infiltration with phagocytosis or formation of neutrophil extracellular nets (NETs),^{10–12} 8) synthesis of additional heavy, light, and variable Ig chains, and 9) saliva.

sIgA is an example of functional coherence within a group of proteins. Mucosal B cells produce IGHA1 at a 3:1 to 4:1 ratio compared to IGHA2. IgA heavy chains are linked by joining chains (IGJ) to form IgA dimers (IgA1₂–IgJ or IgA2₂–IgJ). Kappa light chains are rearranged before lambda chains, and so more kappa chains are incorporated into antibodies. The dimers bind to PIGR, located on the interstitial surface of submucosal gland serous cells.¹³ PIGR binds these dimers covalently. The pinocytosed complex is shepherded through serous cells for exocytosis with other secreted proteins. This is in contrast to monomeric IgA that lacks IGJ, and the IgG-subclass antibodies that are produced by distant plasma cells. These enter the airways predominantly by plasma exudation along with hepatic proteins.

Recognition of these shared glandular, vascular, and epithelial protein sources, mechanisms of secretion, and

potentially mutual regulatory controls permit a systems-biology approach to identify individual biomarkers and sets of proteins (“biosignatures”) that may be predictive of cigarette smoking and COPD phenotypes.¹⁴

Subjects and methods

This paper represents a further elaboration of previously published data,⁷ where details of methodological issues are described. The main aspects are reported for clarity.

Subjects

Ethical approvals were obtained from the Institutional Review Boards of University of Pavia and Georgetown University. Informed consent was obtained to assess induced sputum,^{5,7} smoking history, and pulmonary function,¹⁵ and CT scans³ classified subjects into phenotypes of: lifelong never-smokers (Non; n=7), healthy smokers (HS; n=13), chronic bronchitic patients (CB; n=11), patients with COPD without significant emphysema (COPD; n=15), and patients with significant emphysema and the airflow obstruction of COPD (E&C; n=10). Demographics, clinical data, and sputum biochemical⁵ and proteomic⁷ outcomes have been described. The HS and CB groups were younger than COPD and E&C. This suggested that more time and exposure was required for HS subjects to transition to airflow obstruction and alveolar inflammation. Age-matched studies are required as follow-up studies.

Sputum

Lyophilized sputum was reconstituted, digested with trypsin, and peptides separated by capillary liquid chromatography–quadrupole–time-of-flight mass spectrometry (Waters Corporation, Milford, MA, USA). Mascot software (Matrix Science, London, UK) and the National Center for Biotechnology Information database matched peptides to proteins (Table 1).⁷ Unmatched peptide sequences with ion scores >20 were assessed using the Protein Information Resource’s “peptide match” feature.^{16,17} RefSeq (Reference Sequence) protein identifiers were converted to unique

Table 1 Demographics

	Non	HS	CB	COPD	E&C
n	7	13	11	15	10
Age (years)	68±2	42±6	42±5	70±9	69±8
Pack-years	0	25.0±8.6	36.3±15.0	33.4±25.8	47.7±35.7
% female	29%	15%	27%	27%	18%
FEV ₁ /FVC (% predicted)	79.5%±6.0%	81.4%±5.0%	75.6%±5.8%	49.5%±14.8%*	37.2%±6.3%*

Notes: *P=0.0002 by two-tailed, unpaired Student’s t-test after significant analysis of variance results between phenotypes. Data expressed as percentages or means ± standard deviation. HS and CB subjects were the youngest in this male-dominated group. COPD and E&C had significantly worse lung function.

Abbreviations: Non, nonsmokers; HS, healthy smokers; CB, chronic bronchitic patients; E&C, patients with significant emphysema and airflow obstruction; FEV₁, forced expiratory volume in 1 sec; FVC, forced vital capacity.

National Center for Biotechnology Information gene-symbol abbreviations. This avoided confusion from redundant coding numbers (albumin), alternative names (BPIFB1 = C20orf114 = LPLUNC1),⁹ and peptide sequences shared by different Ig isotypes,¹⁸ DEFA1&3, ORM1&2, and between other proteins. Proteins detected in fewer than three subjects (eg, S100A9) were excluded to reduce data complexity in this pilot study. Frequencies of protein detection were calculated for each clinical phenotype. Collating proteins by their cellular sources significantly constrained data complexity (see Introduction). Ig variable-chain results were excluded because of idiotype hypervariability and somatic mutation that generated many unique peptide sequences. Subjects with fewer than five remaining proteins were removed from analysis. Factor and multilogistic regression analyses were negative because of this stratified analytical strategy.

Relative protein abundance: spectral counts

Mass-spectrometry data were reanalyzed to determine the number of individual tryptic peptides in each specimen. The number of peptides per protein is the spectral count, and provides a means to quantify relative protein abundances.¹⁹ Spectral counts were normalized by dividing the number of tryptic peptides per protein^{20–22} by the total number of peptides per specimen, then multiplying by protein concentrations to correct for differing amounts of starting materials. Log₁₀ transformation was effective at reducing the impacts of exceptionally high or absent outliers. Mean transformed spectral counts and 95% confidence intervals for each protein and phenotype were compared between groups by analysis of variance (ANOVA) with significance at $P \leq 0.05$, followed by two-tailed unpaired Student *t*-tests.^{23–25} These pilot data were not corrected for multiple comparisons. Our intent was to assess data reduction, parallel analysis, and obtain preliminary trends and differences for future power-analysis calculations.

Spectral counts for each identified protein from all sputum samples were ranked by conversion to percentiles. The percentiles for each protein provided relative levels for comparisons within phenotypes, across all subjects, and between the five phenotypes. The normalized protein percentiles facilitated comparisons of very high- and very low-abundance proteins.

Correlation

Normalized spectral counts for all pairs of proteins detected within each phenotype were correlated. Significant relationships (Pearson's correlations with $P \leq 0.05$)^{23–25} were found for exocytosis from submucosal gland serous and mucous

cells, goblet and other epithelial cells, vascular permeability, and neutrophilic inflammation.

Frequencies of detection

For each protein, the number of subjects who had detectable protein was divided by the number of subjects in that phenotypic group. This provided a measure of the presence of that protein in sputum that was independent of spectral counts. The relative presence or absence of a protein in a phenotypic group allowed inferences about variations between groups, the patterns of change for sputum proteins in Non and smoker phenotypes, and homeostatic and pathophysiological mechanisms involving each protein.

Scattergrams

Scattergrams were used to plot the presence of each protein (frequencies of detection) versus protein abundances (normalized spectral counts converted to percentiles). This allowed visualization of significant positive and negative Pearson's correlations ($P \leq 0.05$) between high- and low-abundance proteins. In general, correlated sets conformed to the predicted cellular origins described in the Introduction. Changes in sources of sputum proteins were traced with the progression of disease from HS to E&C. These results suggested independent modular responses of exocytosis from bronchial glands, epithelial secretion, vascular permeability, and neutrophilic inflammation in different groups of smokers. Unexpected significant deviations like DEFA1&3 in E&C were identified.

Protein network-interaction analysis

Significantly correlated sets of proteins in single phenotypes were analyzed. Functional and regulatory networks were created for sets of epithelial, glandular, inflammatory, and other cell products. Tools and resources included Cytoscape 2.8,^{26,27} the ReactomeFI plug-in^{28–30} with the 2012 FI database supplemented with interactions from the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database,^{31,32} and pathway databases from KEGG release 64.0 (October 1, 2012),³³ BioCarta,³⁴ the Pathway Interaction Database (PID),³⁵ and Reactome.³⁶ Significant interactions were defined by $P \leq 0.05$ and false-discovery rate ≤ 0.01 . Gene networks were constructed using a spectral partition-based network clustering algorithm in the ReactomeFI plug-in.³⁰

Results

Subjects

Subjects were predominantly male (Table 1).⁵ The HS and CB groups were younger than the others. By definition,

lung function was significantly worse for COPD and E&C than the other smokers ($P=0.0002$ by two-tailed, unpaired Student t -test after significant ANOVA). This reflected the differences in age, pack-years, and severity of lung disease.

Proteins detected in induced sputum

The original set of 203 RefSeq proteins⁷ was reduced to 114 gene products using the National Center for Biotechnology Information gene database and improved annotation of Ig sequences.¹⁸ The 50 proteins detected in at least three specimens were analyzed. The 16 with the highest transformed spectral counts were detected in 64%–100% of samples (Table 2). Sputum proteins were organized by tissue of origin. Percentile spectral counts and frequencies of detection assessed relative trends in protein expression between groups (Figure 1).

Secretory IgA

Uniformly high spectral counts and frequencies of detection for IGHA1 and IGHA2 heavy chains created a “ceiling” effect in the five groups (Figure 2). IGJ was lower in COPD than the other groups ($P=0.044$ by t -test after significant ANOVA) (Figure 1D). This may indicate damage or dysregulation in the bronchial mucosa, with decreased plasma cell production of IgA dimers, reduced expression of PIGR or

pinocytosis by serous submucosal gland cells, or a reduction in number of IgA-transporting serous cells in the glands in the COPD group.³⁷

Submucosal gland serous cells

LYZ, LTF, SLPI, and BPIFB1 were the predominant serous cell products (Figure 3). BPIFB1 was detected in 80% or more of all specimens per group. However, counts were lowest in Non and HS ($P=0.014$), suggesting that BPIFB1 was a common protein whose secretion increased as lung function worsened. The BPIFA1 (PLUNC) protein was detected only in the COPD group (not significant). In contrast, BPIFB2 was essentially present only in the Non group ($P=0.000003$). These data were consistent with previous studies showing large increases in messenger RNA (mRNA) levels for BPIFB1 and BPIFA1 in airway cells of smokers,³⁸ and BPIFA1 from bronchial epithelial scrapings.³⁹ LCN1, the namesake for this lipid-sequestering β -barrel LCN protein family, was detected in the HS group ($P=0.027$).

Epithelial proteins

Counts for SCGB1A1 (uteroglobin; Clara cell-specific 10 kD protein), PRR4, and AZGP1 were significantly higher in Non than the smoker groups. AZGP1 showed the largest magnitude drop from Non to HS of all proteins ($P=0.009$), suggesting it may be a marker of nonsmoker status (Figure 4).

Table 2 Proteins with the highest spectral counts in the 56 induced sputum specimens (mean [95% confidence interval])

Gene	Protein	Gene number	Log ₁₀ -normalized spectral counts	Frequency of detection
IGHA1	Ig α 1 constant	3493	1.73 [1.61–1.85]	96%
DMBT1	Deleted in malignant brain tumors 1	1755	1.71 [1.60–1.83]	100%
IGJ	Ig J polypeptide (linker protein for Ig α and μ polypeptides)	3512	1.63 [1.60–1.82]	95%
ALB	Albumin	213	1.53 [1.35–1.71]	91%
LYZ	Lysozyme (renal amyloidosis)	4069	1.44 [1.28–1.59]	95%
BPIFB1	Bacterial permeability-increasing fold-containing family B, member 1 (chromosome 20 open reading frame 114, C20orf114; long palate, lung, nasal epithelium, carcinoma-associated 1 [LPLUNC1]; von Ebner's minor salivary gland protein)	92747	1.33 [1.18–1.47]	91%
IGHG1	Ig γ 1 constant (G1m marker)	3500	1.28 [1.11–1.44]	88%
IGHA2	Ig α 2 constant (A2m marker)	3494	1.24 [1.11–1.37]	95%
IGK@	Ig κ variable region	50802	1.17 [1.01–1.34]	88%
LTF	Lactotransferrin; lactoferrin	4057	1.17 [1.02–1.33]	86%
MUC5B	Mucin 5B, oligomeric mucus/gel-forming	727897	1.12 [0.96–1.28]	84%
SCGB1A1	Secretoglobin, family 1A, member 1 (uteroglobin; Clara cell-specific 10 kD protein)	7356	1.04 [0.86–1.21]	77%
PIGR	Polymeric Ig receptor	5284	1.02 [0.86–1.18]	79%
SLPI	Secretory leukocyte peptidase inhibitor	6590	0.917 [0.749–1.085]	75%
AZGP1	Alpha-2 glycoprotein 1, zinc-binding	563	0.734 [0.563–0.904]	64%
ACTB	Actin- β	60	0.727 [0.576–0.877]	73%

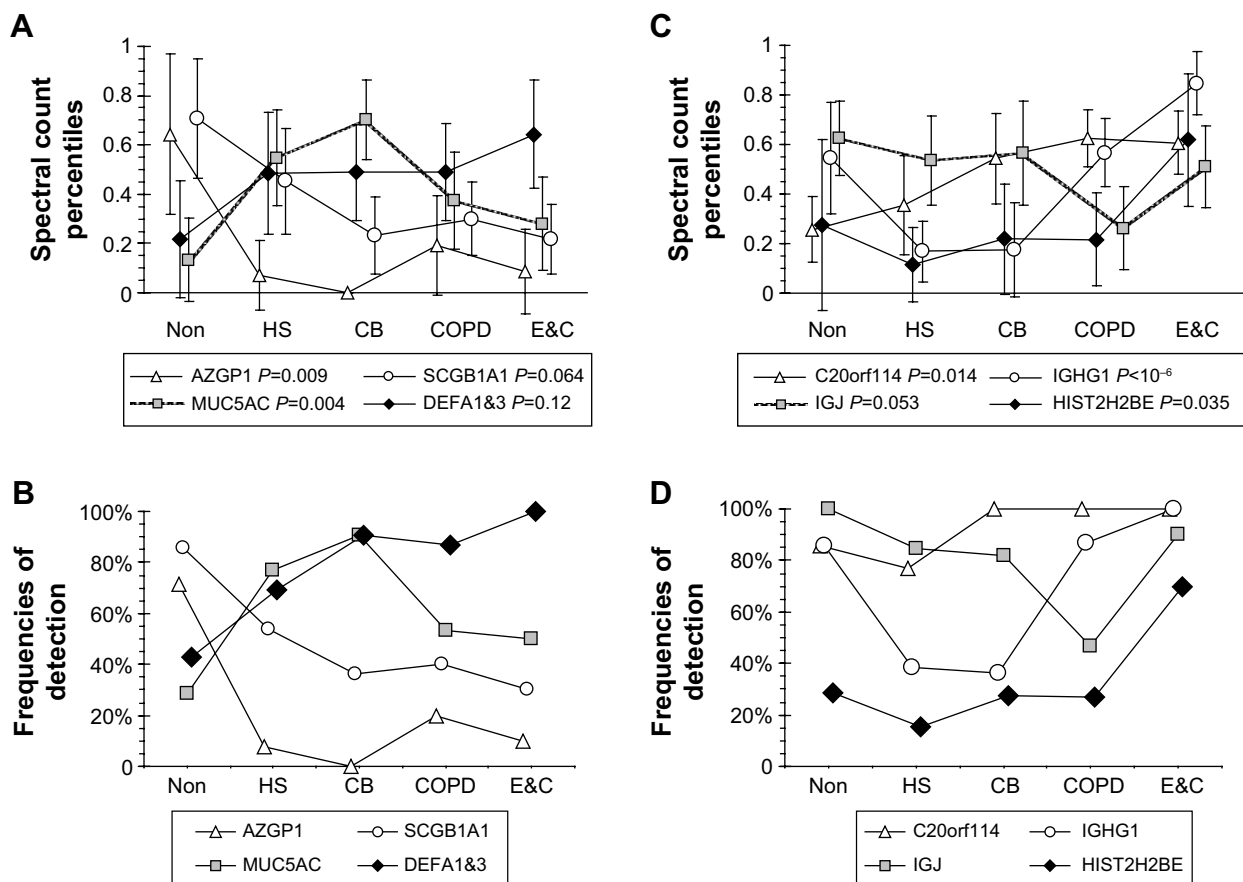


Figure 1 Patterns for relationships between protein spectral count percentiles and frequencies of detection for each phenotype. **Notes:** Trends in spectral count percentiles (**A, C**, ANOVA probabilities) and frequencies of detection (**B, D**) are shown for the proteins with the largest ranges of expression and differences between phenotypic groups. AZGP1 (open triangles) and SCGB1A1 (open circles) were highest in Non and trended downward in smokers (**A, B**). MUC5AC (gray squares) was maximal in CB. DEFA1&3 (black diamonds) increased with disease severity from Non to E&C. Spectral count percentiles (**C**) and frequencies of detection (**D**) are shown for BPIFB1 (C20orf114; open triangles), IGHG1 (open circles), IGJ (gray squares), and HIST2H2BE (black diamonds). **Abbreviations:** Non, nonsmokers; HS, healthy smokers; CB, chronic bronchitic patients; E&C, patients with significant emphysema and airflow obstruction; ANOVA, analysis of variance.

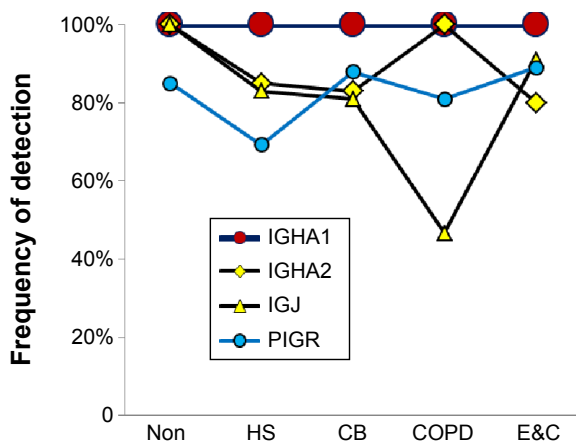


Figure 2 Secretory IgA-related proteins. **Notes:** IGHA1 (red circles, prototype sIgA protein), IGHA2 (yellow diamonds), and IGJ (yellow triangles) were highly expressed, although IGJ was detected less frequently in sputum from COPD subjects. PIGR, the serous cell-binding protein for dimeric IgA, had comparable expression (blue). **Abbreviations:** Non, nonsmokers; HS, healthy smokers; CB, chronic bronchitic patients; E&C, patients with significant emphysema and airflow obstruction.

However, this finding was contrary to Vanni et al who found higher AZGP1 protein expression by Western blot and immunohistochemistry, and mRNA levels in large airway epithelia of healthy smokers compared to healthy nonsmokers.⁴⁰ One potential explanation may be increased glycosylation in smokers, which may allow detection of higher concentrations of AZGP1 epitopes by immune methods, but reduce the efficiency of trypsin digestion and tryptic peptide formation prior to mass spectrometry.

SCGB1A1, the most highly expressed gene in Clara cells (club cells), had its highest proteomic levels in Non, with intermediate levels in HS ($P=0.064$) and lowest levels in CB, COPD, and E&C. Smoking caused large decreases for SCGB1A1 mRNA in bronchial⁴¹ and nasal mucosa⁴² and small airway epithelial cells,³⁸ which was consistent with our proteomic trend. DEFA1&3 levels were lowest in Non, then increased progressively to maximum values in E&C. DEFA1&3 was

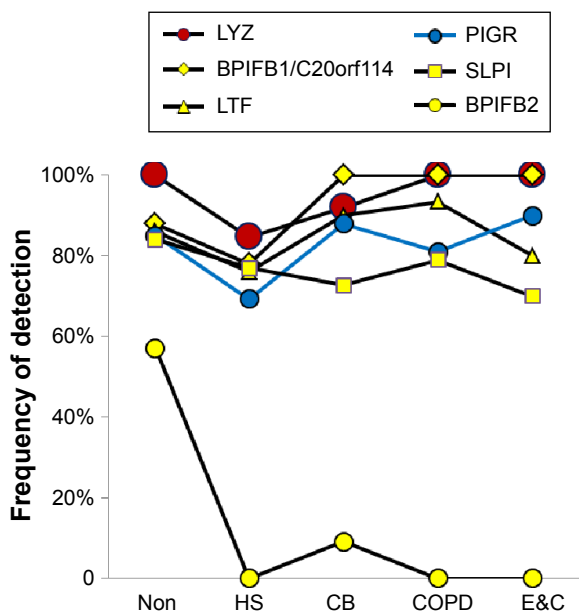


Figure 3 Submucosal gland serous cell proteins.
Notes: LYZ (red circles, prototypical glandular serous cell protein), BPIFB1 (yellow diamonds), LTF (yellow triangles), PIGR (blue), and SLPI (yellow squares) had comparable high levels of expression in each phenotype. PIGR is shown in blue for comparison to sIgA-related proteins. BPIFB2 (yellow circles) was essentially secreted only in Non, and absent in smokers.
Abbreviations: Non, nonsmokers; HS, healthy smokers; CB, chronic bronchitic patients; E&C, patients with significant emphysema and airflow obstruction.

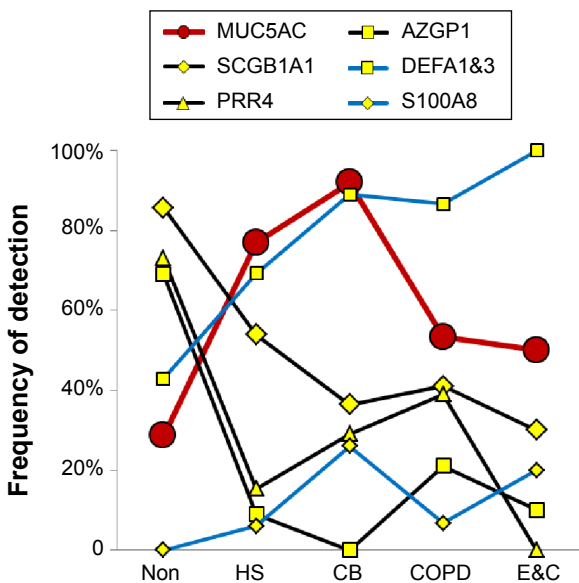


Figure 4 Epithelial proteins.
Notes: Three general patterns were seen: 1) goblet-cell MUC5AC (red circles and line) was increased in HS and CB; 2) SCGB1A1 (yellow diamonds), PRR4 (yellow triangles), and AZGP1 (yellow squares) were highest in Non and tended to decrease between HS and E&C (black lines); and 3) in contrast, DEFA1&3 (yellow squares, blue line) and S100A8 (yellow diamonds, blue line) tended to increase from Non to E&C.
Abbreviations: Non, nonsmokers; HS, healthy smokers; CB, chronic bronchitic patients; E&C, patients with significant emphysema and airflow obstruction.

probably of epithelial cell origin in all groups except E&C, where DEFA1&3 correlated with neutrophil proteins. Future studies of defensin expression in sputum and lung parenchyma are warranted to confirm and extend this finding.

Goblet-cell mucin 5AC

MUC5AC was significantly higher in CB and HS than Non, COPD, and E&C ($P=0.004$), (Figure 1A, B and Figure 4). This was consistent with goblet-cell rather than submucosal gland mucous cell hypersecretion in CB.^{9,43}

Submucosal gland and duct mucous cells

MUC5B and DMBT1 were highly expressed in all groups. MSMB and PIP were detected in fewer specimens (Figure 5). These changes were consistent with data from small airway secretory cells obtained from current smokers compared to nonsmokers³⁸ and after smoking cessation.⁴⁴

Neutrophil and cellular proteins

Two patterns were found (Figure 6). CB had frequent CST1&4 (sequences shared by cystatin SN and cystatin S), GAPDH, and GSN. In contrast, E&C had significantly higher HIST2H2BE ($P=0.035$) and MPO ($P=0.030$) than the other groups (Figure 1C). DEF1&3, LCN2, ACTG1, and CTSG were also more frequent in E&C. Correlations were shown between histones, neutrophil granule proteins, and ACTB. These proteins were characteristic of neutrophil cell death with extrusion of NETs in E&C.¹⁰⁻¹² Actins and histones have been previously detected by proteomic means without mRNA expression.³⁹

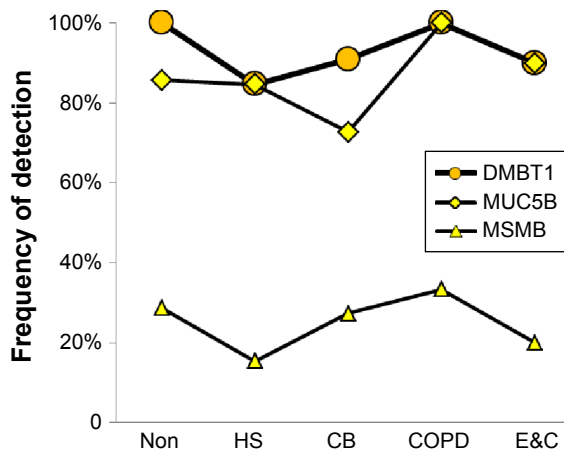


Figure 5 Submucosal gland mucous cell proteins.
Notes: DMBT1 (orange circles), MUC5B (yellow diamonds), and MSMB (yellow triangles) were detected in similar percentages of each phenotypic group.
Abbreviations: Non, nonsmokers; HS, healthy smokers; CB, chronic bronchitic patients; E&C, patients with significant emphysema and airflow obstruction.

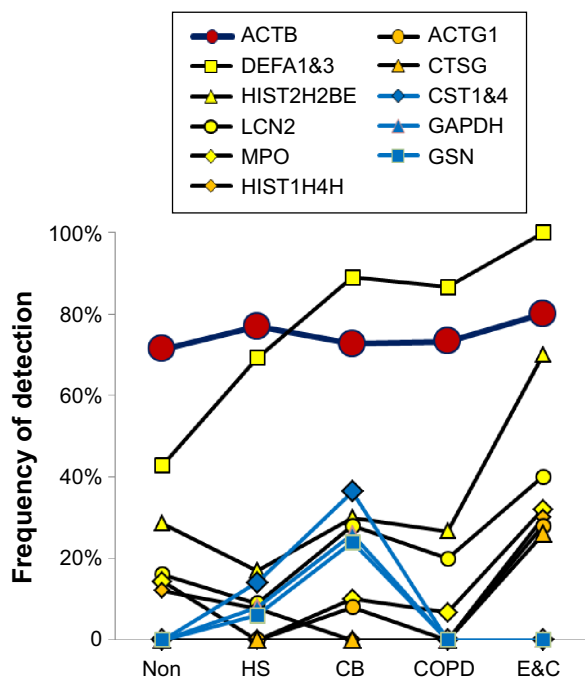


Figure 6 Cellular proteins.

Notes: Three patterns were found: 1) ACTB (red circles) had comparable frequencies of detection in all phenotypes; 2) CST1&4 (blue diamonds and line), GAPDH (blue triangles and line), and GSN (blue squares and line) were most frequently detected in CB; and 3) proteins that trended upward and were highest in E&C included DEFA1&3 (yellow squares), histones (yellow triangles and orange diamonds), LCN2 (yellow circles), MPO (yellow diamonds), ACTG1 (orange circles), and CTSG (orange triangles).

Abbreviations: Non, nonsmokers; HS, healthy smokers; CB, chronic bronchitic patients; E&C, patients with significant emphysema and airflow obstruction.

Plasma proteins

ALB ($P=0.0013$) and IGHG1 ($P<10^{-6}$) were significantly lower in HS and CB compared to the Non, COPD, and E&C groups (Figure 7). This may indicate relatively reduced plasma extravasation in HS and CB. The parallel trends for IGHG1, ALB, and HP in sputum suggested that the bulk of the IgG was synthesized distant from the lung and then delivered by plasma extravasation along with hepatic proteins. In contrast, TF was most frequently detected in HS and CB. This may reflect a need for the antimicrobial, iron-sequestering function of TF.⁴⁵

Immunoglobulin polypeptides of plasma and mucosal B-cell origin

The precise plasma and mucosal Ig origins of IGHG2, IGHG3, IGHG4, IGHM, κ (IGK) and λ (IGL) light chains, and variable regions of heavy (IGH@) and light (IGK@, IGL@) chains could not be identified.¹⁸ Because κ -light chains are translated before λ -chains, the significant elevation of IGL@ in HS and CB suggested extensive B-lymphocyte differentiation in response to chronic microbial infection.

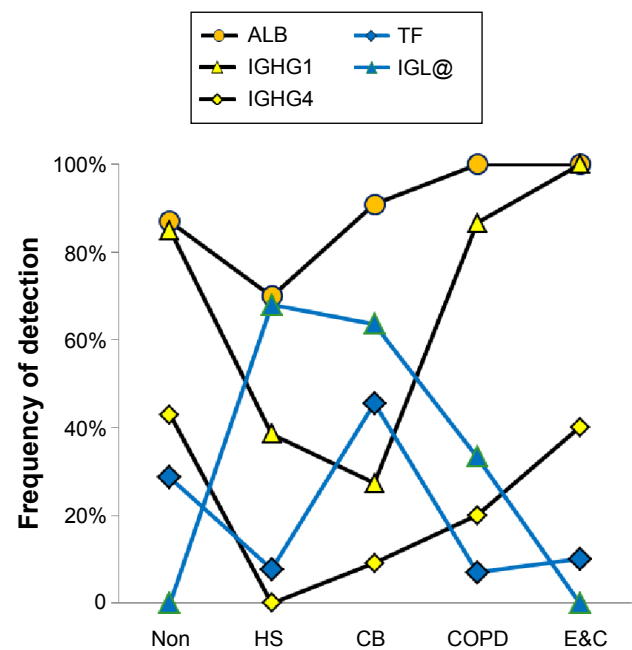


Figure 7 Plasma proteins.

Notes: Three patterns were found: 1) albumin was detected in all phenotypes (orange circles); 2) the Ig heavy chains IGHG1 (yellow triangles) and IGHG4 (yellow diamonds) were detected in Non and decreased in HS and CB, but increased to higher detection rates in COPD and E&C; and 3) the opposite pattern was shown for TF (blue diamonds and line) and Igl light-chain variable regions (IGL@; blue triangles and line), with the highest levels in HS and CB.

Abbreviations: Non, nonsmokers; HS, healthy smokers; CB, chronic bronchitic patients; E&C, patients with significant emphysema and airflow obstruction.

Amylase IA

AMY1A was present with low counts in 30%–57% of samples. There were no differences between groups indicating equivalent sampling and processing of induced sputum and normalization of peptide and protein abundances. AMY1A mRNA may be expressed in small airway epithelia.³⁸

Correlation analysis

The Non group had highly correlating serous (BPIFB1, LYZ, LTF, SLPI, PIGR) and epithelial (SCGB1A1) proteins (Figure 8), suggesting similar regulation of secretion in healthy airways. Their frequencies of detection were 0.86. Separate positive correlations were found for high-abundance plasma-cell IGJ and IGHA2, and lower-abundance epithelial AZGP1 and PRR4. IGJ was negatively correlated with IGHG1. This was consistent with their distinct origins from airway plasma cells, as opposed to vascular permeability, respectively.

The HS scattergram showed lower percentiles for sIgA and serous proteins (Figure 9) compared to Non. The sIgA proteins IGHA1, IGHA2, and PIGR were correlated with serous cell LTF, LYZ, and BPIFB1, goblet-cell MUC5AC, and plasma ALB.

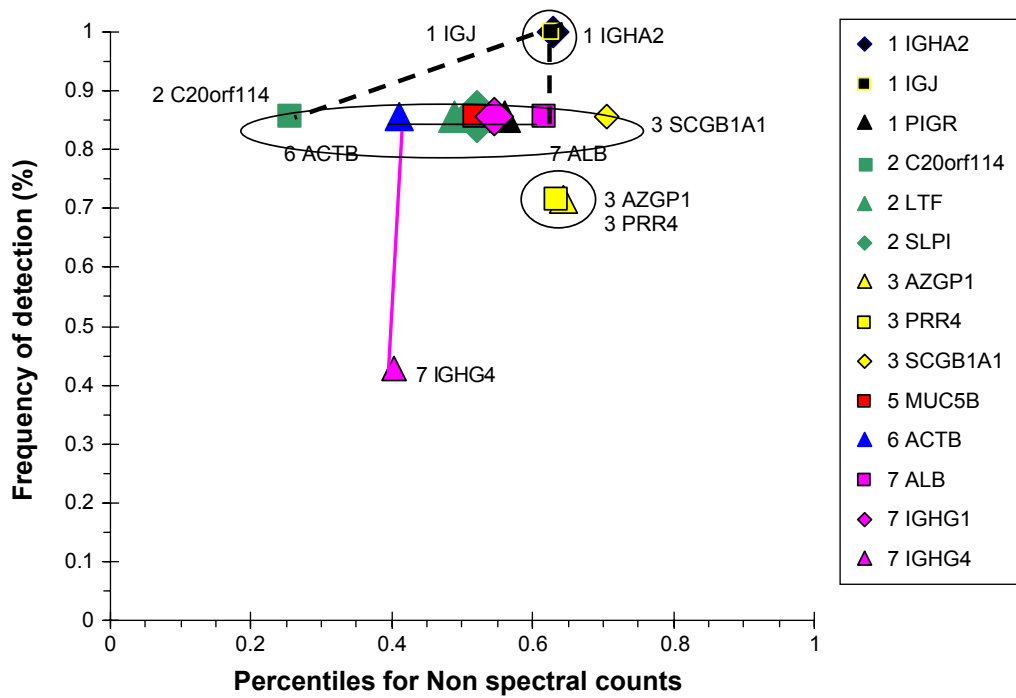


Figure 8 Correlation analysis for Non-group proteins.

Notes: The scattergram-plotted frequencies of detection versus the percentiles of spectral counts for each significantly correlated protein ($P \leq 0.05$). The proteins were numbered and colored by origin in secretory IgA (1, black), serous cells of submucosal glands (2, green), epithelial cells (3, yellow), mucous cells of submucosal glands (5, red), cellular cytoplasm (6, blue), and plasma (7, magenta). The ellipses enclose groups of proteins that were significantly correlated with each. Solid lines connect smaller sets of significantly correlated proteins, such as ACTB, IGHG1, and IGHG4. The dashed lines indicate negative correlations for IGJ with IGHG1 and C20orf114 (BPIFB1). Positively correlated proteins suggested similar cellular origins and mechanisms of release into sputum.

Abbreviation: Non, nonsmokers.

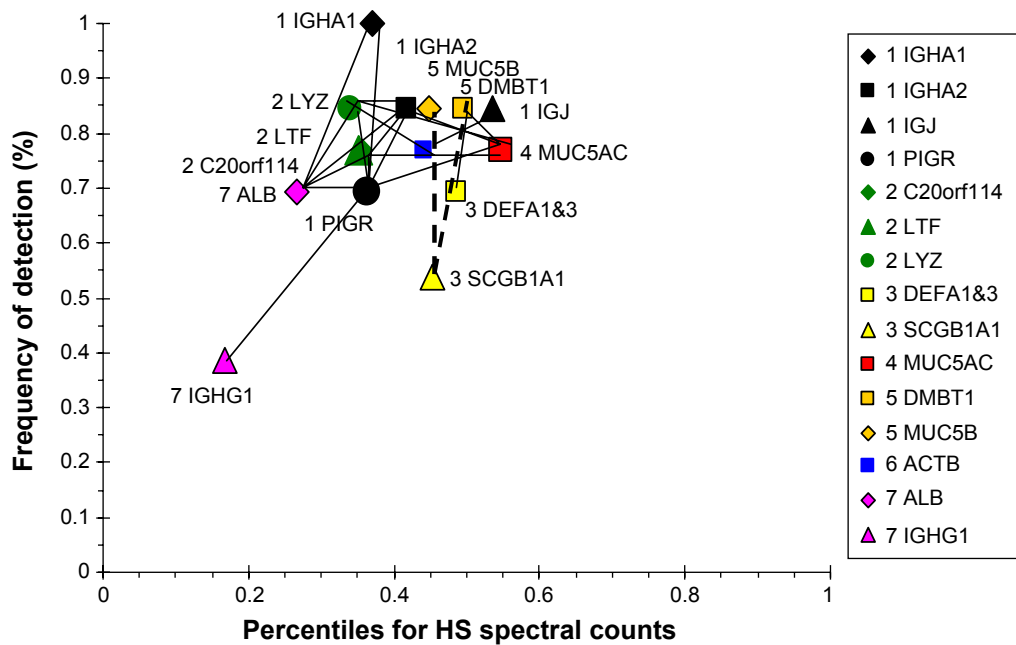


Figure 9 Correlation analysis for HS-group proteins.

Notes: Secretory IgA (1, black), serous (2, green), epithelial (3, yellow) goblet (4, red), mucous (5, gold), and plasma (7, magenta) proteins were positively correlated (solid lines). Submucosal gland mucous cell products and epithelial proteins had different trends. MUC5B and DMBT1 were positively correlated with DEFA1&3, but negatively correlated with SCGB1A1 (dashed lines). ALB and IGHG1 had surprisingly low expression in HS.

Abbreviation: HS, healthy smokers.

BPIFB1 was increased by smoking.³⁸ The mucous cell products MUC5B and DMBT1 were positively correlated with epithelial DEFA1&3, but negatively correlated with SCGB1A1. The inverse relationship of MUC5B and SCGB1A1 with smoking was consistent with previous mRNA data.³⁸ AZP1 and PRR4 were absent, suggesting a major change in epithelial protein production compared to Non. Mucous products were relatively increased, and plasma ALB and IGHG1 were decreased compared to Non in these cigarette smokers.

CB had the highest level of MUC5AC of all of the groups. This was consistent with goblet-cell hyperplasia and mucus hypersecretion. MUC5AC was negatively correlated with LYZ and LTF suggesting inverse effects on goblet and submucosal gland serous cells (Figure 10). MUC5AC had no positive correlations, reinforcing its independent regulation. BPIFB1 was detected in all subjects, but did not correlate with any other major protein. This may indicate BPIFB1 also had independent expression in CB. LTF and LYZ were positively correlated with a network of sIgA, SLPI, mucous (MUC5B, DMBT1), and ALB proteins. In contrast, LTF and LYZ were negatively correlated with a

set of relatively low-abundance but highly intercorrelated cellular proteins: GAPDH, GSN, HIST2H2BE, LCN2, and S100A8. Presumably, the low-abundance proteins were not derived from serous cells. The plasma proteins ALB, TF, and IGHG1 were correlated as a separate group, suggesting shared vascular extravasation.

COPD proteins were clustered as high- and low-abundance proteins arranged along a single narrow axis of expression with no outliers (Figure 11). The intercorrelated high-abundance set consisted of serous cell LYZ, LTF, and BPIFB1, and mucous MUC5B and DMBT1. These correlated with lower-abundance PIGR, ACTB, and MUC5AC. LYZ and LTF were negatively correlated with the plasma proteins ALB and IGHG4, suggesting glandular secretion was relatively more important than vascular permeability as a source of sputum proteins at the time of onset and progression of airway obstruction and COPD. Correlations for the lower-abundance secretory proteins included LCN2 with PRR4, and BPIFA1 with MSMB and MUC5AC. mRNA for BPIFA1 and MSMB were significantly increased in bronchial brushings from smokers

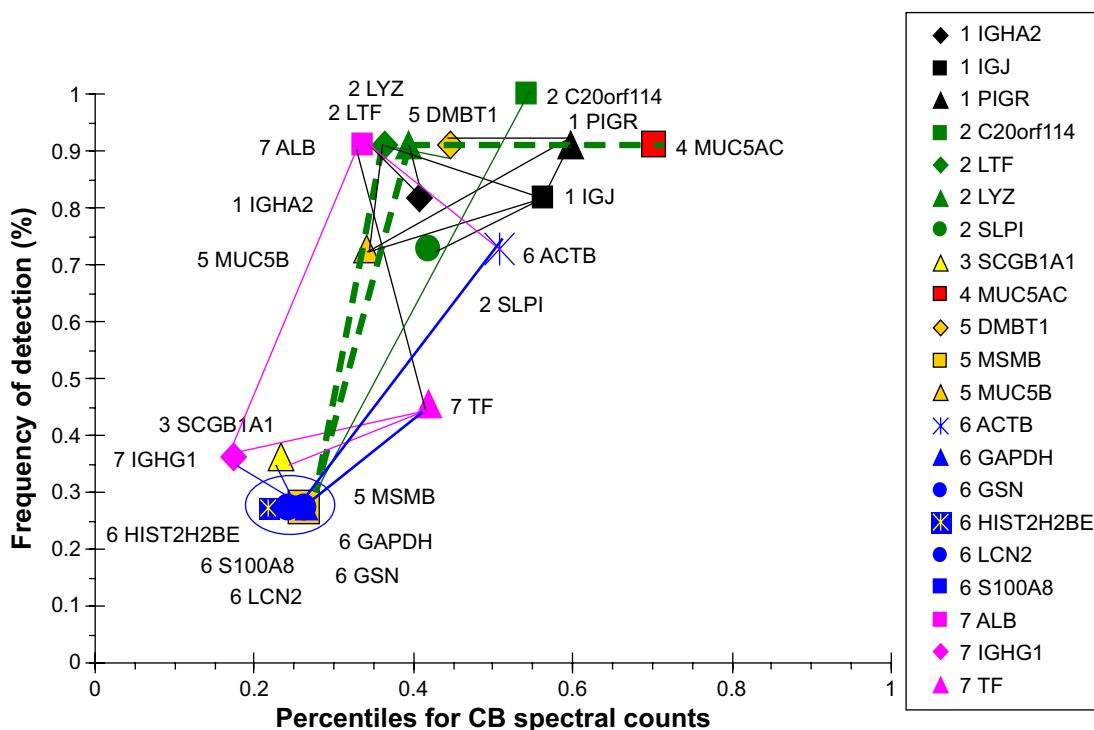


Figure 10 Correlation analysis for CB-group proteins.

Notes: CB proteins were divided into high (upper half of figure)- and low (lower left corner)-abundance subsets. MUC5AC and C20orf114 (BPIFB1) had the highest expression, but few correlations. Secretory IgA (1, black), serous cell LYZ and LTF (2, green), and the mucous gland products MUC5B and DMBT1 (5, yellow) were correlated. Low-abundance cellular proteins (6, blue) were highly clustered and correlated with each other (blue ellipse). MSMB (5, orange) is in the ellipse, but did not correlate with these cellular proteins. LYS and LTF were negatively correlated with MSMB and MUC5AC (dashed green lines). Plasma ALB, IGHG1, and TF were correlated with each other, which was consistent with their entry into sputum via plasma extravasation.

Abbreviation: CB, chronic bronchitic patients.

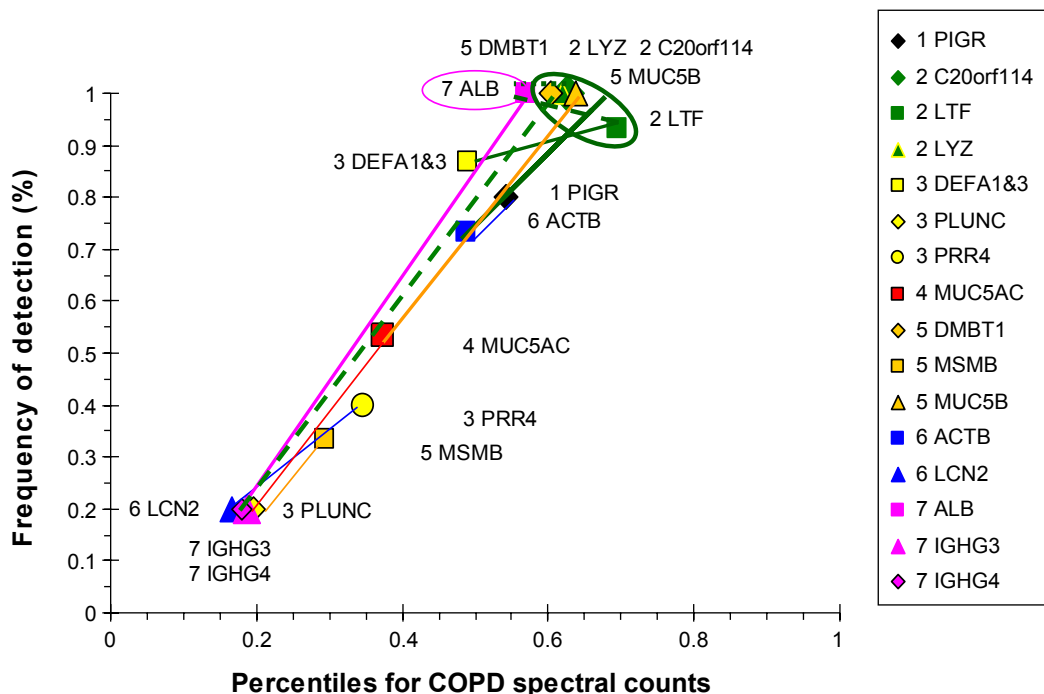


Figure 11 Correlation analysis for COPD-group proteins.

Notes: Serous (LYZ, LTF, BPIFB1; 2, green) and mucous (MUC5B, DMBT1; 5, orange) proteins had high abundances and were highly intercorrelated (green ellipse). These proteins also correlated (solid green lines) with PIGR (1, black), DEFA1&3 (3, yellow square), MUC5AC (4, red square), and ACTB (6, blue square). Serous cell LTF and LYZ were negatively correlated (dashed green lines) with the abundant ALB (7, magenta ellipse) and low-abundance IGHG4 (7, magenta). Low-abundance proteins were clumped into the lower left corner. IGHG3 and IGHG4 were correlated with each other, but not the adjacent LCN2 (6, blue) or PLUNC (BPIFA1; 3, yellow). LCN2 was correlated with PRR4 (3, yellow), while PLUNC (BPIFA1) correlated with MSMB (5, orange) and MUC5AC (4, red).

compared to nonsmokers.³⁸ These relationships suggested that glandular exocytosis predominated in the COPD group and was differently regulated from vascular exudation and epithelial exocytosis.

The E&C group had a unique set of correlated proteins that were distinct from the COPD group (Figure 12). IGHA1, PIGR, LTF, and MUC5B were negatively correlated with ALB. IGJ was negatively correlated with IGHG4. This suggested that sIgA formation and submucosal gland serous and mucous cell secretion were inversely related to vascular permeability in emphysema. ALB percentile count and frequency of detection were higher than tracheobronchial glandular proteins in E&C. The inverse correlation may indicate an inverse relationship between bronchoalveolar vascular permeability and tracheobronchial glandular secretion. Cellular proteins were highly correlated with one another. Unlike the other four phenotypic groups, DEFA1&3 were correlated with cellular MPO, CTSG, ACTG1, HIST2H2BE, and HIST1H4H, but not epithelial proteins, such as SCGB1A1. This raised the hypothesis that DEFA1&3 were predominantly neutrophil products in E&C. Although neutrophil counts were equivalent for E&C (21.2×10^3 cells/mg; 13.2 and 23.8) and COPD (13.3×10^3 cells/mg; 9.0 and

21.1; median [interquartile range]; Mann–Whitney *U*-test),⁵ their mechanisms of activation may have been distinct, with viable neutrophils in COPD but NET formation and neutrophil disintegration in E&C.^{10–12}

Cytoscape network analysis

Non had three separate protein networks. First was an extended grouping of serous cell LYZ, LTF, PIGR, and SLPI with epithelial SCGB1A1 (Figure 13). The linker genes *ETS1* and *NFKB1* were introduced by Cytoscape and STRING based on text-mined data. *ETS1* interacted with LYZ and LTF. *NFKB1* has homeostatic functions for PIGR expression and B-lymphocyte antibody synthesis. LTF was networked to LRP1, then to either LRP2 and SCGB1A1, or via *ELANE* (*ELA2*) to SLPI. SLPI was inserted as a linker protein, since it is an inhibitor of *ELANE*. Because *ELANE* was not significantly detected in Non sputum, this portion of the network may not be relevant to airways of Non subjects. In the second network, *AZGP1* was a central hub associated with *USP53*, *PIP*, *ITGAV*, and *B2M*. The third independent network consisted of the secreted antimicrobial proteins *PRR4* and *MUC7*. These maps infer a central role for *ETS1* in maintaining glandular serous and epithelial Clara-cell

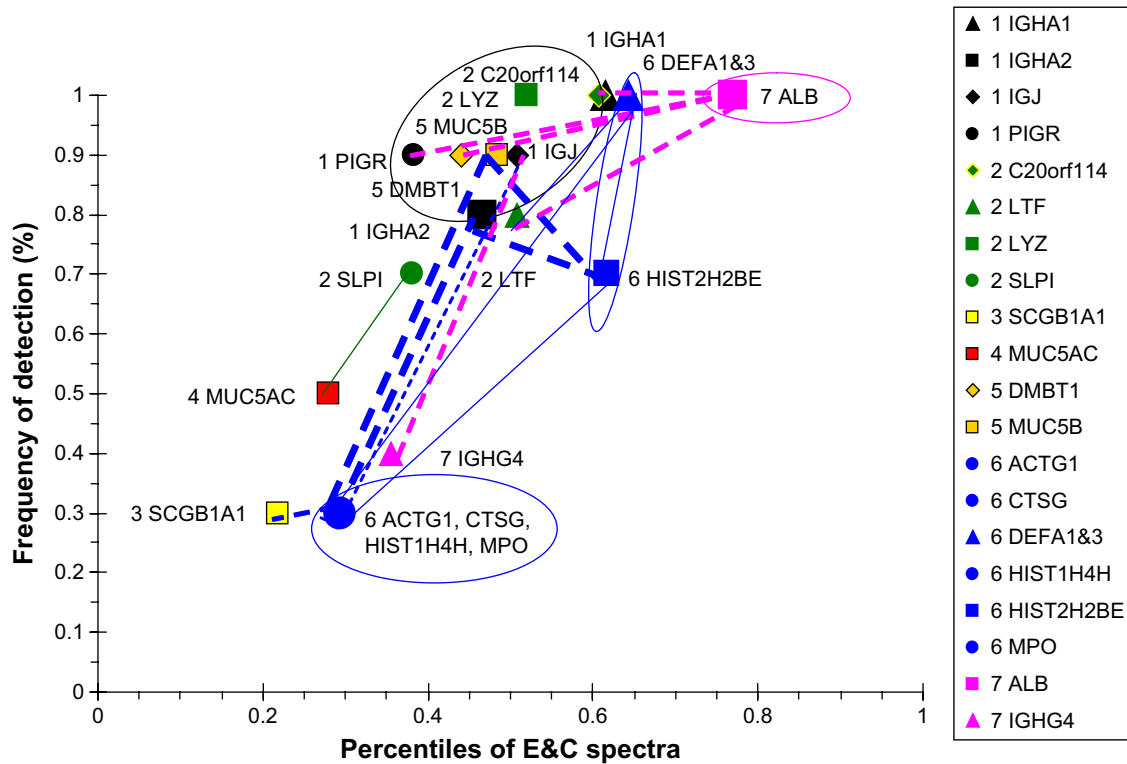


Figure 12 Correlation analysis for E&C-group proteins.

Notes: The black ellipse encloses proteins from the secretory IgA (1), serous (2), and mucous (5) groups that were highly intercorrelated. IGHA1, PIGR, LTF, and MUC5B were negatively correlated with ALB (7, plasma source, dashed magenta lines). IGJ was also negatively correlated with IGHG4. Cellular proteins (6, blue symbols, lines and ellipses) were significantly correlated with each other, and negatively correlated with IGHA2 and MUC5B. The low-frequency cellular proteins at the bottom of the frame were negatively correlated (blue dashed line) with epithelial SCGB1A1 (3, yellow square).

Abbreviation: E&C, patients with significant emphysema and airflow obstruction.

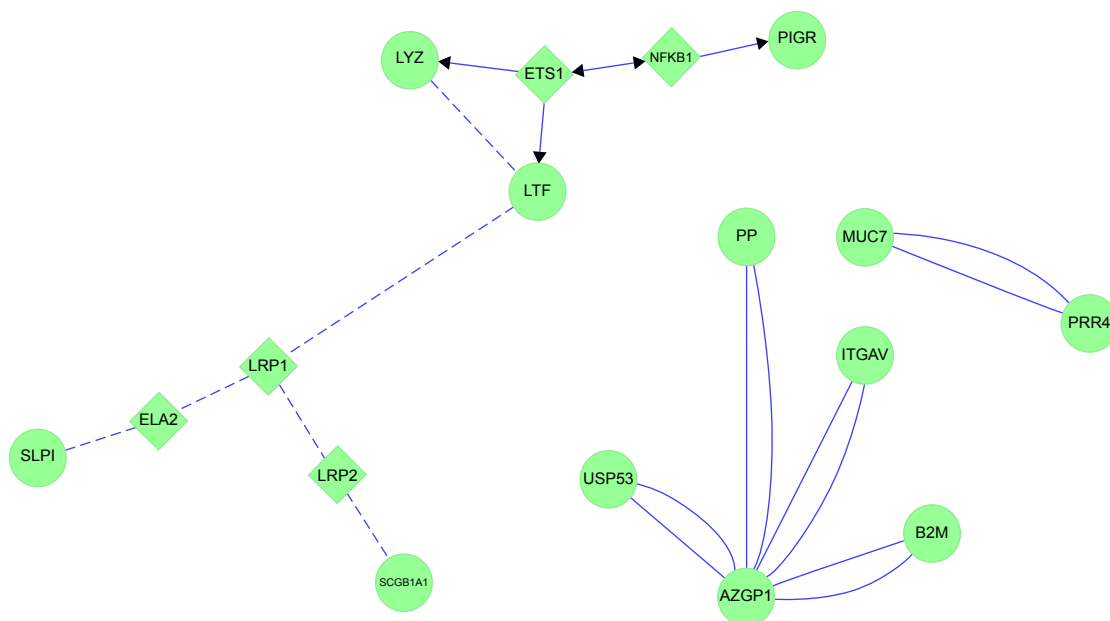


Figure 13 Cytoscape network analysis for nonsmokers.

Notes: Three separate networks were identified between proteomically identified proteins (green circles, gene symbols). Inferred linker proteins (diamonds) were introduced by Cytoscape based on text mining. Previously reported interactions are shown by solid blue lines, and inferred relationships for these proteins by dashed lines.

secretion as predominant pathways for innate immune defense in the airway.

HS showed the impact of cigarette smoking. The exocytosed goblet-cell protein MUC5AC and serous cell LTF, LYZ, SPLI, and PIGR were linked to SP1, SP3, and USF2 transcription factors (Figure 14). These were not among the top 30 transcription factors found using small airway brushings,³⁸ suggesting that they were expressed by epithelial cells in large-diameter bronchi and submucosal gland serous cells, but not smaller-diameter bronchial Clara cells. USF2 regulates epithelial and glandular cell protein expression, and may be an early marker of bronchial dysplasia.⁴⁶

CB generated two complex protein networks (Figure 15). The first was centered on hubs of GAPDH, GSN, and S100A8. GAPDH was clustered with other glycolysis proteins and MYC. GAPDH was linked to GSN through APP. GSN was associated with CASP3 and phosphatidylinositol metabolism. These interactions may have been related to GSN's actin-capping role with ACTB. ACTB was also linked to GAPDH, and through MMP9 to LCN2. The distal tail of this map had tenuous links through BAX and TP53 to S100A8 and histones. MYC, TP53, and GSN provided a logical link between smoking, glandular hypersecretion, reactive oxidant species-mediated chromosomal damage, defective DNA repair, and carcinogenesis.

CEBPA is integral to cell-cycle regulation,^{47,48} and was linked to two sets of proteins. CEBPA was linked through LTF to LYZ and SLPI, which initially suggested serous cell modulation. However, the interconnections included LTF, ELANE, SLPI, and LRP1, which may be neutrophil products in CB. In the opposite direction, CEBPA was linked through a tail to DMBT1. CEPBA mRNA was highly expressed in small airway epithelial brushings,³⁸ Clara cells, type II alveolar cells, and alveolar macrophages. CEPBA is required for airway extracellular matrix repair and remodeling. One outcome of this network-interaction analysis is to hypothesize that small-molecule modulators of CEBPA may be beneficial for CB treatment. However, this will require further testing, since CEBPA may act as a lung-tumor suppressor.

MYB regulates enzymes of aerobic glycolysis, chaperone expression in unfolded protein responses, and serous and neutrophil granule innate immune proteins.⁴⁹ Aerobic glycolysis is relevant to phagocytosis and reactive oxidant production by viable neutrophils in CB.

The COPD proteome formed a highly linked protein network (Figure 16). The transcription factors ETS1, CEBPA, SP1, and NFkB1 formed a central interactive core. CEBPA was again linked to mucous cell DMBT1 and MSMB. SP1 was linked to goblet-cell MUC5AC. SP1, CEBPA, and ETS1 were connected to serous cell LTF and LYZ, and played roles in modulation of extracellular

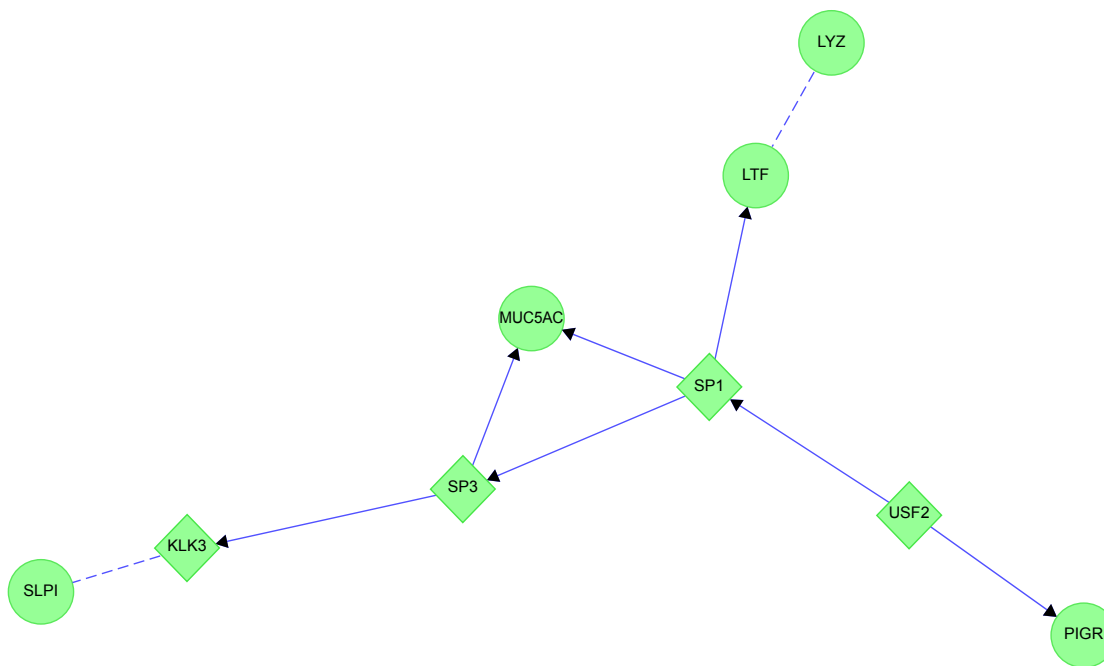


Figure 14 MUC5AC and glandular serous cell proteins LYZ, LTF, SLPI and PIGR in healthy smokers (HS).

Notes: The transcription factors SP1, SP3, and USF2 were linked to goblet-cell MUC5AC and glandular serous cell proteins in sputum from HS subjects.

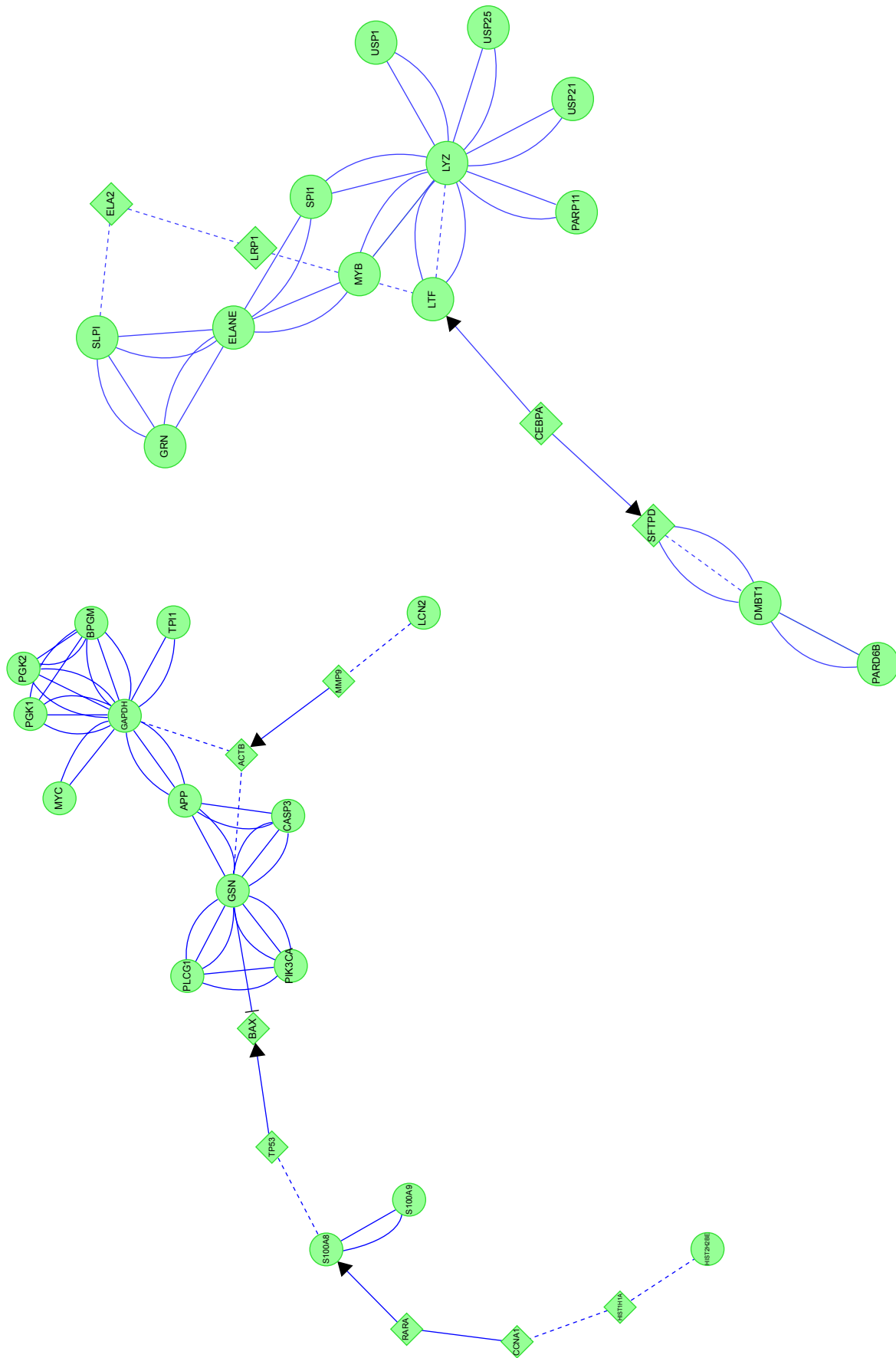


Figure 15 Chronic bronchitis: the two protein interaction networks suggest that distinct pathological mechanisms contribute to bronchial pathology. **Notes:** The map on the left is dominated by connections of GAPDH, ACTB, and GSN. GAPDH is linked to glycolysis proteins. GSN connects phosphoinositides to mechanisms of actin regulation. Links through the long tail to histones include S100A8 and S100A9. On the right, CEBPA is linked to DMBT1 in the tail, and LTF and potential neutrophil or serous cell products in the interconnected head.

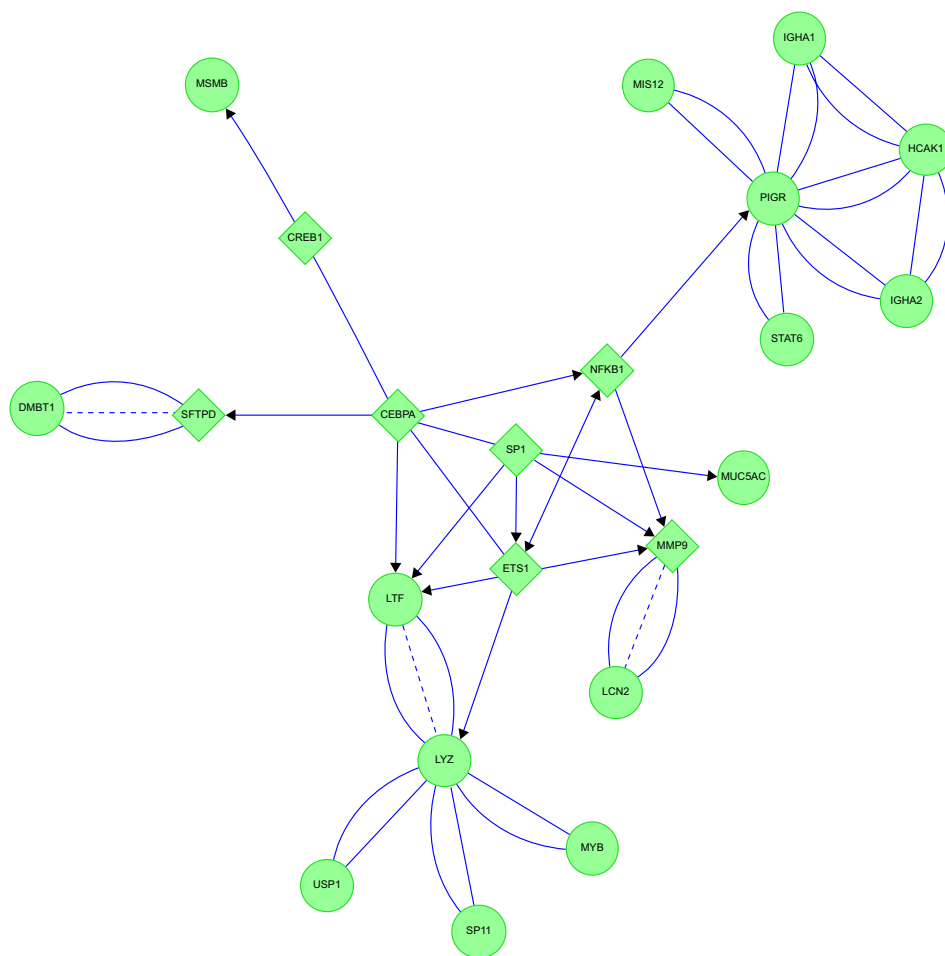


Figure 16 COPD: protein interaction networks suggest central roles for transcription factors in bronchiole pathology.

Notes: The central core of ETS1, CEBPA, SPI, and NFKB1 transcription factors interacted with proteins from serous (LTF, LYZ, PIGR), mucous (DMBT1, MSMB), and goblet (MUC5AC) exocrine cells, and IgA-producing B lymphocytes. This pattern of interactions suggests hypersecretion from the cluster of submucosal gland serous and mucous cells and their neighboring IgA B cells as a potential pathology, bridging chronic bronchitis and COPD. The link of SPI with MUC5AC may indicate continued goblet cell hyperplasia as seen in healthy smokers with COPD. These support two avenues for the transition from healthy smokers to the airflow limitation that defines COPD. One would involve mechanisms of airflow limitation with their roots in healthy smokers, and the other a progression of bronchial wall submucosal gland mucous hypersecretion pathology through chronic bronchitis to airflow limitation and COPD.

matrix components for bronchial repair and metastasis.⁵⁰ As suggested by its origin as a B-lymphocyte transcription factor, NFKB1 regulated serous cell PIGR and IGHA1 and IGHA2 expression.

E&C had four connected features (Figure 17). ITGAM and MMP2 were central links. ITGAM is a component of the macrophage MAC-1/CR3 receptor. ITGAM was linked through PRTN3 to CTSG. CTSG was closely connected to F2R, IGFBP3, other coagulation proteins, and their protease inhibitor SERPINB13. ITGAM was also linked directly to MPO. ITGAM and MMP2 shared linkages to ACTG and its cluster of actin-sequestering proteins. The opposite limb from ITGAM projected through MMP2 to CEBPA, which was coupled separately to DEF1&3 and histones. These interactions implicated CEBPA, neutrophil cytokinesis, protease activation, and innate immune

aspects of complement, coagulation, and NET pathways in emphysema.

Discussion

Each of the smoker and healthy subject phenotypes had unique patterns of proteins in their induced sputum specimens. This suggests that distinct pathophysiological processes were associated with smoking, mucus hypersecretion, airflow obstruction, and emphysema (Figure 18). For example, AZGP1 may be a marker for the healthy epithelium of non-smokers that is inhibited by cigarette-smoke components. Serous cell BPIFB1 had comparable frequencies of detection in each group, but Non had the lowest spectral counts. The higher levels of this glandular protein in sputum from smokers suggest BPIFB1 may be a marker of cigarette smoking and progressive small airway damage, leading to damage

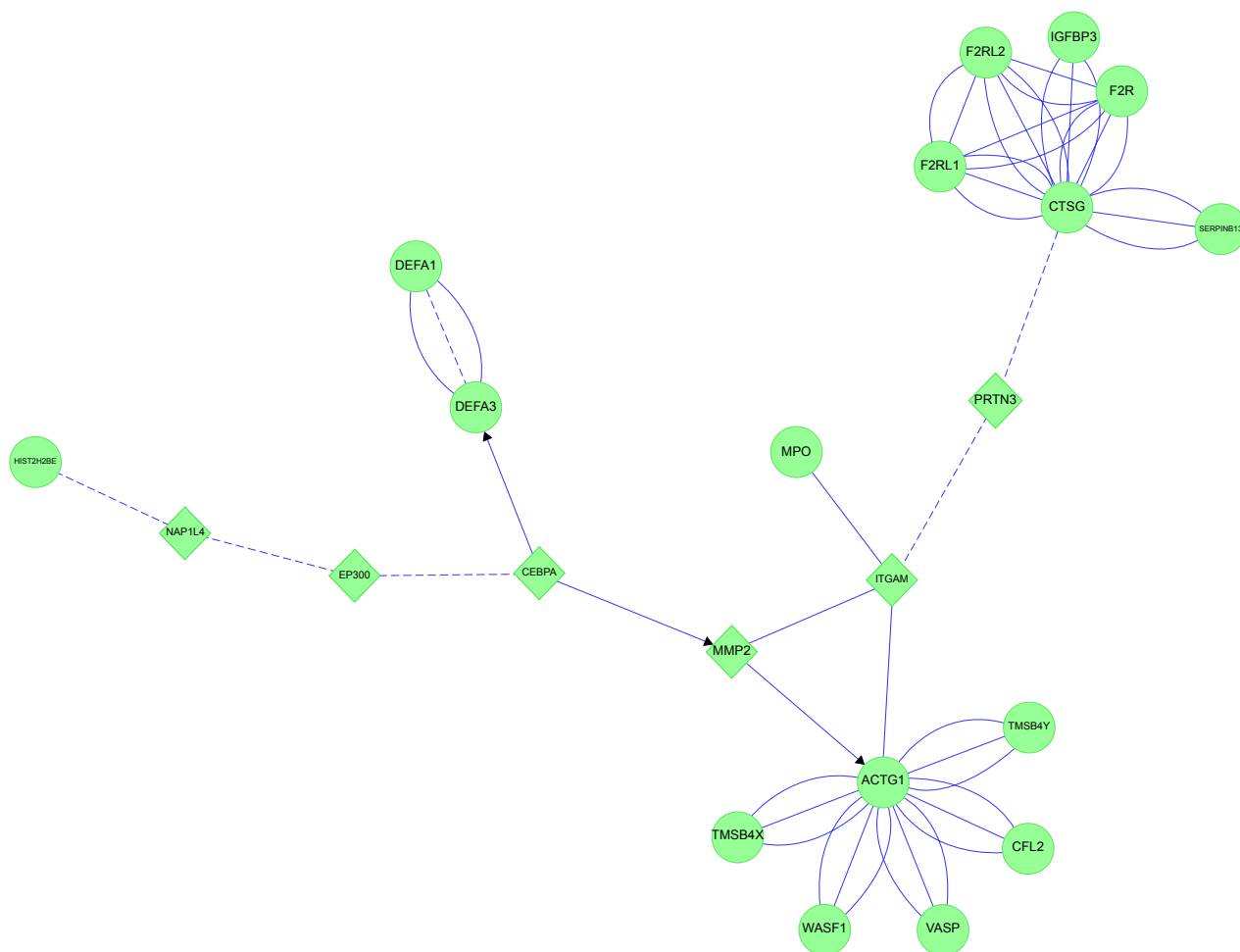


Figure 17 Emphysema: neutrophil and plasma protein inflammatory cascades were inferred from protein interaction networks. **Notes:** ITGAM had a central position. It was connected to MPO, and through PRTN3 to CTSG, coagulation cascade enzymes, and their potential inhibitor, SERPINB13. ITGAM and MMP2 were linked to ACTG and actin-sequestering proteins. MMP2 was linked through CEBPA to defensins and histones. These interactions implicate neutrophil and plasma protease inflammatory cascades such as coagulation in emphysema pathology. Other cascades such as complement may also be involved but were not detected here. Histones, defensins, MPO, ACTG, and the multiple proteases supported neutrophil extracellular nets and proteolysis.

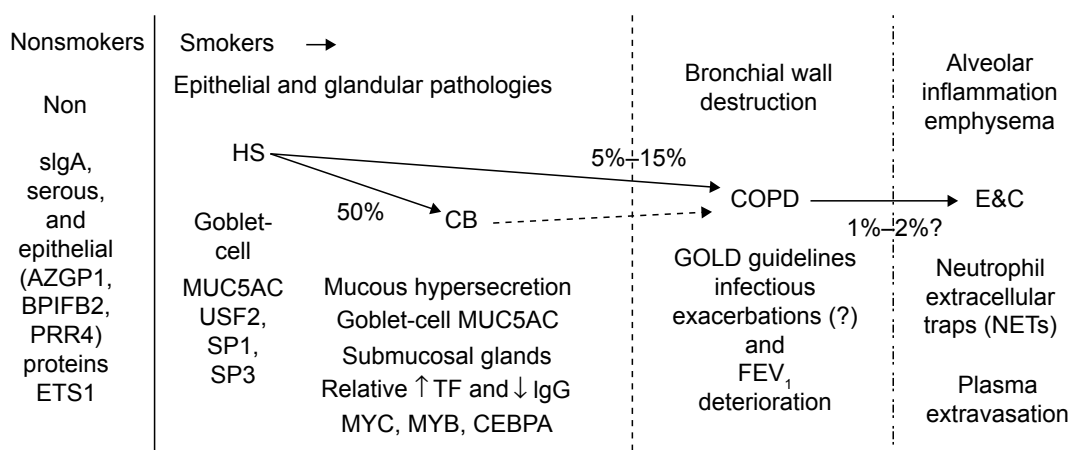


Figure 18 Sputum proteome-derived approach to cigarette smoke-induced lung diseases. **Notes:** All smokers were at risk of lung cancer. Approximately half of HS progressed to mucous hypersecretion without airflow obstruction (CB). As smokers aged, 5%–15% developed accelerated obstruction (COPD). A smaller proportion developed alveolar destruction (E&C). Diagnostics and treatments may be significantly improved by directing them at the dynamic alterations in pathogenic mechanisms. **Abbreviations:** Non, nonsmokers; HS, healthy smokers; CB, chronic bronchitic patients; E&C, patients with significant emphysema and airflow obstruction; sIgA, secretory IgA; GOLD, Global initiative for chronic Obstructive Lung Disease.

of the lower tracheobronchial tree and alveoli. The detection of SCGB1A4 in Non and increased levels of BPIFB1, BPIFA1, MUC5B, MSMB, and GSN were consistent with high-throughput RNA-sequencing results from bronchial brushings of smokers compared to nonsmokers.³⁸

Goblet-cell MUC5AC hypersecretion was the predominant source of mucins in CB.^{9,43} Half of smokers may develop CB.^{1,2} Genetic risk-factor analysis for CB may be beneficial by selecting phenotypes of smokers with and without sputum MUC5AC hypersecretion. Goblet-cell hypersecretion may represent a separate pathophysiological pathway that is not related to the small airway destruction, increased vascular permeability, and decreased airflow (FEV₁) found in COPD and E&C. Separation of emphysema from other COPD subjects may also help focus genotypic analyses of these phenotypes.⁵¹ Differences between Non and CB strongly suggest that smokers without airflow obstruction (HS, CB) should be recognized by international forums so that stage-specific diagnostic and treatment algorithms can be developed to reduce smoking and mechanisms of mucus hypersecretion.¹⁵ Involvement of MYC may provide insights into origins of lung cancer in CB without the airflow limitation required for classification as COPD. In addition, more focused evaluations of the pathophysiologically distinct E&C group are warranted.

Phenotypic changes in vascular permeability were inferred from IGHG1 and other plasma protein results. IGHG1 was detected in 85% of Non, but only approximately half as many HS and CB subjects. This suggested reduced vascular permeability in HS and CB. One consequence may be a relative lack of water in the tenacious sputum of CB. In contrast, IGHG1 was detected in all E&C subjects, suggesting increased plasma flux across damaged capillary–alveolar walls in emphysema.

Only half of COPD expressed IGJ compared to all Non subjects. This suggested a large reduction in endogenous plasma-cell dimeric IgA synthesis as part of the bronchial mucosal injury in COPD. IGJ was more abundant when emphysema was present (E&C).

COPD was the only group to express BPIFA1. This protein may act as a protease inhibitor and regulate epithelial lining fluid volume.⁵² BPIFA1 may be secreted from neutrophil granules,⁵³ but was not correlated with other neutrophil proteins in COPD.

E&C demonstrated neutrophilic inflammation with NETS containing chromatin, HIST2H2BE, proteases, and other cellular components.^{10–12} DEFA1&3 were highly correlated with these neutrophilic proteins. This was in contrast to the predominantly epithelial correlates for defensins in the other phenotypes. Elevated BPIFB1, DEFA1&3, and IGHG1

distinguished E&C from Non, HS and CB. This distinct mechanism argues that new, inexpensive, robust tools are needed for widespread screening to identify emphysema, and to spur research into the differential treatment of the E&C group.

Proteomic studies of sputum and lung diseases have been limited by several factors. Comparisons to other proteomic studies are complicated by marked differences in sample sizes, phenotypic definitions,⁵³ comparison groups,³⁹ mucopurulent versus normal sputum,^{5,39,54,55} bronchoalveolar lavage fluid,^{56,57} brushings,³⁸ laser dissection of tissue sections,⁴⁰ and lung tissue.^{58,59} Studies that rely on pooling specimens from COPD subjects will not show more subtle, phenotype-related differences observed by examining each sample independently. Pooling does not allow for comparisons based on differences between individuals within two groups. The correct matching of peptides to proteins suffers from the redundant entries of the same amino acid sequences into reference-protein databases. Improved protein annotation based on genomic sequences has reduced the number of Ig proteins that may otherwise be designated as hypothetical or unknown proteins. Although specific protein isoforms were often selected in the Mascot peptide–protein matching process, it was not possible to confirm the presence of alternatively spliced or single amino acid replacements from this data set.⁶⁰ Future analysis of posttranslational modifications will provide information about acetylation, glycosylation, and other regulatory alterations that may affect protein function. Glycosylation and other posttranslational modifications may mask (or expose) trypsin-cleavage motifs and lead to different patterns of tryptic peptides detected by mass spectrometry. Other digestive enzymes are likely to provide complementary peptides, and may identify additional relevant sputum proteins.

Data analysis was facilitated by several methodological adaptations. Gene-symbol abbreviations identified individual proteins and eliminated the redundancies of names, annotations, and multiple identification numbers found in protein databases. Spectral counting was used for nonisotopic evaluation of relative peptide and matching protein abundances.^{20–22} Correction for the wide range of original total-protein concentrations was required for these reconstituted, lyophilized sputum samples. Logarithmic transformation reduced the impact of outliers with very high values and specimens with absent counts. Conversion of transformed counts to percentiles made it possible to compare the trends for relative yields of high- and low-abundance proteins for all subjects and between the five phenotypic groups. Correlation analysis verified functional and mechanistic relationships between sets of proteins secreted from different types of cells or derived from plasma. Imposing

these multiple constraints reduced data dimensionality and improved confidence in the outcomes predicted for nonsmoker, smoker, and chronic destructive airway illness phenotypes.

Several caveats are needed for network analysis by such software as Cytoscape. We did not extend our network analysis beyond one linker. As a result, more tenuous protein–protein and gene–gene interactions and associations were excluded. However, some of these may have been of direct relevance to cigarette-smoking diseases. Programs are updated on regular intervals, but may lag behind publication dates by several months. Despite the lag, we have found that their linkers were generally robust. In separate studies, Cytoscape identified links that were not found by text mining of the PubMed database. Selective manual annotation is critical, since some associations may not be relevant to lung biology. These two approaches are complementary for identifying the most reliable and up-to-date network-interaction patterns.

Current criteria and clinical practice focus on smokers with reduced FEV₁ and decrements that may occur following exacerbations of lung disease.¹⁵ Elevated serous cell BPIFB1 secretion and vascular permeability (ALB, IGHG1) separated our COPD group from Non, and HS from CB groups, respectively. Verification of these potential biomarkers by measuring concentrations and mRNA expression in sputum, bronchoalveolar lavage fluid, and epithelial biopsies may better define the COPD subgroup of lung-disease patients. We predict that a different set of induced sputum-biomarker proteins including bronchial microbiome proteins and microRNAs will be found in exacerbations. The chronology and magnitude of these proteomic trends may provide insights into the pathophysiology of these acute events, and subsequent chronic declines in pulmonary status.

Results from the E&C phenotype strongly suggest a neutrophilic proteome and inflammatory cascade in CT scan-proven emphysema. The airflow-obstructed smoker groups were discriminated by higher PRR4 in COPD compared to higher IGJ and IGHG1 in E&C. Evidence of more severe inflammation in E&C⁵ should prompt the development of noninvasive methods to identify and more aggressively treat subjects with significant emphysema (E&C) rather than focus on airflow obstruction alone (COPD).

This proteomic and network analysis suggests a framework for cigarette smoke-induced change in lung proteins, transcription factors, and inflammatory mechanisms. HS showed increased goblet-cell MUC5AC expression that implicated USF2, SP1, and SP3, and a relative decrease in plasma protein extravasation. Approximately half of HS progress to CB. CB had MUC5AC and submucosal gland MUC5B hypersecretion,

with a reduction in Clara-cell SCGB1A1. Network analysis implicated MYC, MYB, and CEBPA. Investigation of pathogenic mechanisms in this majority of smokers requires greater attention. In contrast, pathology of the age-dependent 5%–15% of smokers⁶¹ who develop recurrent severe exacerbations and progressive airflow obstruction (COPD) has been the focus of close scrutiny and guidelines.⁶²

The COPD sputum proteome implicated submucosal gland hypersecretion, with involvement of CEBPA, SP1, ETS1, MYB, and NFKB. The emphasis on COPD may be related to the widespread use of spirometry for disease grading, while quantification of mucus hypersecretion in CB and the alveolar damage of E&C are still limited in scope and utility. The E&C proteome was remarkable for the increased plasma extravasation and proteins linked to NET formation.^{10–12} This form of neutrophil death and toxicity was distinct from CB and COPD based on their sputum proteomic profiles. Understanding the dynamic progression of these pathological mechanisms may lead to significantly improved diagnostic tools and treatments of epithelial, glandular, airway-wall, and alveolar inflammation.

Acknowledgments

This study was supported by US Public Health Service Award RO1 ES015382 from the National Institute of Environmental and Health Sciences (NIEHS) (JNB, BC, LP), Department of Defense (DoD) Congressionally Directed Medical Research Program W81XWH-07-1-0618 (JNB), and Strategic Program 2006, Italian Ministry of Health – BPCO (ML, PI). The project was also supported by Grant M01RR-023942-01 from the National Center for Research Resources (NCR), a component of the National Institutes of Health (NIH). The contents are solely the responsibility of the authors, and do not necessarily represent the official views of the NCR, NIH, or DoD.

Author contributions

All authors contributed toward data analysis, drafting and revising the paper, and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Vestbo J. Chronic mucus hypersecretion, exacerbations and natural history of COPD. *Exp Lung Res.* 2005;31 Suppl 1:63–65.
2. Joos GF. Mechanisms of COPD. *Exp Lung Res.* 2005;31 Suppl 1:66–71.

3. Omori H, Nakashima R, Otsuka N, et al. Emphysema detected by lung cancer screening with low-dose spiral CT: prevalence, and correlation with smoking habits and pulmonary function in Japanese male subjects. *Respirology*. 2006;11:205–210.
4. Nicholas B, Djukanović R. Induced sputum: a window to lung pathology. *Biochem Soc Trans*. 2009;37:868–872.
5. Boschetto P, Quintavalle S, Zeni E, et al. Association between markers of emphysema and more severe chronic obstructive pulmonary disease. *Thorax*. 2006;61:1037–1042.
6. Luisetti M, Ma S, Iadarola P, et al. Desmosine as a biomarker of elastin degradation in COPD: current status and future directions. *Eur Respir J*. 2008;32:1146–1157.
7. Casado B, Iadarola P, Pannell LK, et al. Protein expression in sputum of smokers and chronic obstructive pulmonary disease patients: a pilot study by CapLC-ESI-Q-TOF. *J Proteome Res*. 2007;6:4615–4623.
8. National Center for Biotechnology Information. BPIFB1: BPI fold containing family B, member 1 [*Homo sapiens* (human)]. 2015. Available from: <http://www.ncbi.nlm.nih.gov/gene/92747>. Accessed May 21, 2015.
9. Fahy JV, Dickey BF. Airway mucus function and dysfunction. *N Engl J Med*. 2010;363:2233–2247.
10. Wartha F, Beiter K, Normark S, Henriques-Normark B. Neutrophil extracellular traps: casting the NET over pathogenesis. *Curr Opin Microbiol*. 2007;10:52–56.
11. von Köckritz-Blickwede M, Nizet V. Innate immunity turned inside-out: antimicrobial defense by phagocyte extracellular traps. *J Mol Med (Berl)*. 2009;87:775–783.
12. Fuchs TA, Abed U, Goosmann C, et al. Novel cell death program leads to neutrophil extracellular traps. *J Cell Biol*. 2007;176:231–241.
13. Kaetzel CS. The polymeric immunoglobulin receptor: bridging innate and adaptive immune responses at mucosal surfaces. *Immunol Rev*. 2005;206:83–99.
14. Richens JL, Urbanowicz RA, Lunt EA, et al. Systems biology coupled with label-free high-throughput detection as a novel approach for diagnosis of chronic obstructive pulmonary disease. *Respir Res*. 2009;10:29.
15. Gold PM. The 2007 GOLD Guidelines: a comprehensive care framework. *Respir Care*. 2009;54:1040–1049.
16. Protein Information Resource [homepage on the Internet]. Available from: <http://pir.georgetown.edu/>. Accessed May 21, 2015.
17. Huang H, McGarvey PB, Suzek BE, et al. A comprehensive protein-centric ID mapping service for molecular data integration. *Bioinformatics*. 2011;27:1190–1191.
18. ImMunoGeneTics [homepage on the Internet]. Available from: <http://www.imgt.org>. Accessed May 21, 2015.
19. Li M, Gray W, Zhang H, et al. Comparative shotgun proteomics using spectral count data and quasi-likelihood modeling. *J Proteome Res*. 2010;9:4295–4305.
20. Lui H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*. 2004;76:4193–4201.
21. Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol*. 2005;1:252–262.
22. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*. 2001;19:242–247.
23. Swinscow TD. *Statistics at Square One*. Oxford, UK: BMJ Books; 1980.
24. Campbell MJ. *Statistics at Square Two: Understanding Modern Statistical Applications in Medicine*. London: BMJ Books; 2001.
25. Barnett V, Lewis T. *Outliers in Statistical Data*. New York: John Wiley & Sons; 1994.
26. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2011;27:431–432.
27. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498–2504.
28. Reactome Wiki. Reactome FI Cytoscape Plugin 4. Available from: http://wiki.reactome.org/index.php/Reactome_FI_Cytoscape_Plugin. Accessed May 21, 2015.
29. Wu G, Feng X, Stein L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol*. 2010;11:R53.
30. Newman ME. Modularity and community structure in networks. *Proc Natl Acad Sci U S A*. 2006;103:8577–8582.
31. STRING [homepage on the Internet]. Available from: <http://string-db.org>. Accessed May 21, 2015.
32. Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013;41:D808–D815.
33. KEGG: Kyoto Encyclopedia of Genes and Genomes [homepage on the Internet]. Available from: <http://www.genome.jp/kegg>. Accessed May 21, 2015.
34. BioCarta [homepage on the Internet]. Available from: <http://www.biocarta.com>. Accessed May 21, 2015.
35. Schaefer CF, Anthony K, Krupa S, et al. PID: the Pathway Interaction Database. *Nucleic Acids Res*. 2009;37:D674–D679.
36. Croft D, O’Kelly G, Wu G, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res*. 2011;39:D691–D697.
37. Gohy ST, Detry BR, Lecocq M, et al. Polymeric immunoglobulin receptor down-regulation in chronic obstructive pulmonary disease. Persistence in the cultured epithelium and role of transforming growth factor- β . *Am J Respir Crit Care Med*. 2014;190:509–521.
38. Hackett NR, Butler MW, Shaykhiev R, et al. RNA-Seq quantification of the human small airway epithelium transcriptome. *BMC Genomics*. 2012;13:82.
39. Steiling K, Kadar AY, Bergerat A, et al. Comparison of proteomic and transcriptomic profiles in the bronchial airway epithelium of current and never smokers. *PLoS One*. 2009;4:e5043.
40. Vanni H, Kazeros A, Wang R, et al. Cigarette smoking induces overexpression of a fat-depleting gene AZGP1 in the human. *Chest*. 2009;135:1197–1208.
41. Boelens MC, van den Berg A, Fehrmann RS, et al. Current smoking-specific gene expression signature in normal bronchial epithelium is enhanced in squamous cell lung cancer. *J Pathol*. 2009;218:182–191.
42. Sridhar S, Schembri F, Zeskind J, et al. Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics*. 2008;9:259.
43. Wang G, Xu Z, Wang R, et al. Genes associated with MUC5AC expression in small airway epithelium of human smokers and non-smokers. *BMC Med Genomics*. 2012;5:21.
44. Zhang LI, Lee J, Tang H, et al. Impact of smoking cessation on global gene expression in the bronchial epithelium of chronic smokers. *Cancer Prev Res*. 2008;1:112–118.
45. Barber MF, Elde NC. Nutritional immunity. Escape from bacterial iron piracy through rapid evolution of transferrin. *Science*. 2014;346:1362–1366.
46. Ocejó-García M, Baokbah TA, Ashurst HL, et al. Roles for USF-2 in lung cancer proliferation and bronchial carcinogenesis. *J Pathol*. 2005;206:151–159.
47. Sato A, Xu Y, Whitsett JA, Ikegami M. CCAAT/enhancer binding protein- α regulates the protease/antiprotease balance required for bronchiolar epithelium regeneration. *Am J Respir Cell Mol Biol*. 2012;47:454–463.
48. Sato A, Yamada N, Ogawa Y, Ikegami M. CCAAT/enhancer-binding protein- α suppresses lung tumor development in mice through the p38 α MAP kinase pathway. *PLoS One*. 2013;8:e57013.
49. Hibi K, Liu Q, Beaudry GA, et al. Serial analysis of gene expression in non-small cell lung cancer. *Cancer Res*. 1998;58:5690–5694.

50. Hong JS, Kim SW, Koo JS. Sp1 up-regulates cAMP-response-element-binding protein expression during retinoic acid-induced mucous differentiation of normal human bronchial epithelial cells. *Biochem J*. 2008; 410:49–61.
51. Patel BD, Coxson HO, Pillai SG, et al. Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2008;178:500–505.
52. Garcia-Caballero A, Rasmussen JE, Gaillard E, et al. SPLUNC1 regulates airway surface liquid volume by protecting ENaC from proteolytic cleavage. *Proc Natl Acad Sci U S A*. 2009;106:11412–11417.
53. Bartlett JA, Hicks BJ, Schlomann JM, Ramachandran S, Nauseef WM, McCray PB Jr. PLUNC is a secreted product of neutrophil granules. *J Leukoc Biol*. 2008;83:1201–1206.
54. Gray RD, MacGregor G, Noble D, et al. Sputum proteomics in inflammatory and suppurative respiratory diseases. *Am J Respir Crit Care Med*. 2008;178:444–452.
55. Braido F, Riccio AM, Guerra L, et al. Clara cell 16 protein in COPD sputum: a marker of small airways damage? *Respir Med*. 2007;101:2119–2124.
56. Merkel D, Rist W, Seither P, Weith A, Lenter MC. Proteomic study of human bronchoalveolar lavage fluids from smokers with chronic obstructive pulmonary disease by combining surface-enhanced laser desorption/ionization-mass spectrometry profiling with mass spectrometric protein identification. *Proteomics*. 2005;5:2972–2980.
57. Sepper R, Prikk K. Proteomics: is it an approach to understand the progression of chronic lung disorders? *J Proteome Res*. 2004;3:277–281.
58. Lee EJ, In KH, Kim JH, et al. Proteomic analysis in lung tissue of smokers and COPD patients. *Chest*. 2009;135:344–352.
59. Ohlmeier S, Vuolanto M, Toljamo T, et al. Proteomics of human lung tissue identifies surfactant protein A as a marker of chronic obstructive pulmonary disease. *J Proteome Res*. 2008;7:5125–5132.
60. Roth MJ, Forbes AJ, Boyne MT 2nd, Kim YB, Robinson DE, Kelleher NL. Precise and parallel characterization of coding polymorphisms, alternative splicing, and modifications in human proteins by mass spectrometry. *Mol Cell Proteomics*. 2005;4:1002–1008.
61. Choi SM, Lee J, Park YS, et al. Prevalence and Global Initiative for Chronic Obstructive Lung Disease group distribution of chronic obstructive pulmonary disease detected by preoperative pulmonary function test. *PLoS One*. 2015;10:e0115787.
62. Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187:347–365.

International Journal of COPD

Publish your work in this journal

The International Journal of COPD is an international, peer-reviewed journal of therapeutics and pharmacology focusing on concise rapid reporting of clinical studies and reviews in COPD. Special focus is given to the pathophysiological processes underlying the disease, intervention programs, patient focused education, and self management protocols.

Submit your manuscript here: <http://www.dovepress.com/international-journal-of-chronic-obstructive-pulmonary-disease-journal>

Dovepress

This journal is indexed on PubMed Central, MedLine and CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.