



Metabolic profiling detects biomarkers of protein degradation in COPD patients

Baljit K. Ubhi*, John H. Riley#, Paul A. Shaw[†], David A. Lomas⁺, Ruth Tal-Singer[§], William MacNee[‡], Julian L. Griffin* and Susan C. Connor*

ABSTRACT: There is a paucity of biomarkers for chronic obstructive pulmonary disease (COPD). Metabolomics were applied to a defined COPD patient cohort from the ECLIPSE study (Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points). Results were correlated with accepted biomarkers for the disease.

Baseline control serum (n=66) and Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage II (n=70), III (n=64) and IV (n=44) COPD patients were analysed by proton nuclear magnetic resonance (¹H NMR). Liquid chromatography with tandem mass spectrometry (LC-MS/MS) was used to confirm amino acid changes detected by ¹H NMR. Data were correlated with body composition, emphysema and systemic inflammation.

¹H NMR identified decreased lipoproteins, *N,N*-dimethylglycine, and increased glutamine, phenylalanine, 3-methylhistidine and ketone bodies in COPD patients with decreased branched-chain amino acids (BCAAs) observed in GOLD stage IV patients. BCAAs, their degradation products, 3-methylhistidine, ketone bodies, and triglycerides were correlated negatively with cachexia and positively with systemic inflammation. Emphysema patients also displayed decreased serum creatine, glycine and *N,N*-dimethylglycine. LC-MS/MS confirmed ¹H NMR findings relating to BCAAs, glutamine and 3-methylhistidine in GOLD stage IV patients.

NMR-based metabolomics characterised COPD patients based on systemic effects and lung function parameters. Increased protein turnover occurred in all COPD patients with increased protein degradation in individuals with emphysema and cachexia.

KEYWORDS: Chronic obstructive pulmonary disease, liquid chromatography-mass spectrometry, metabolomics, metabonomics, nuclear magnetic resonance spectroscopy, serum

Chronic obstructive pulmonary disease (COPD) is a major health burden worldwide, affecting 10–15% of the adult population aged ≥40 yrs [1]. 3 million deaths are caused by COPD annually and it is a major cause of morbidity in an estimated 210 million people worldwide [2]. COPD is the result of chronic airway inflammation that manifests from progressive airflow limitation. Characterisation of COPD patients has relied upon the measurement of airflow limitation, as assessed by forced expiratory volume in 1 s (FEV₁) [3], which has been used to define the condition, its severity and subsequent treatment. However, it is now widely recognised that COPD is a heterogeneous disease with features that are not captured by the measurement of FEV₁ [4]. These include systemic inflammation, muscle dysfunction and cardiovascular disease. There is an urgent need for

biomarkers that reflect underlying pathological processes giving rise to different components of the COPD phenotype; these biomarkers may predict patient response to different therapies and outcomes in COPD.

Metabolomics is a functional genomic technology that provides a detailed investigation of changes in the metabolic profiles of biofluids, gases (breath) and semi-solids (e.g. intact tissues) [5]. We have used nuclear magnetic resonance (NMR) spectroscopy-based metabolomics to detect metabolites in serum that are associated with COPD disease manifestations, thereby serving as potential biomarkers. This was undertaken on samples from a well-characterised group of individuals with COPD from the ECLIPSE cohort (Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points) [6]. A targeted analysis using a bespoke liquid chromatography with tandem mass spectrometry

AFFILIATIONS

*Dept of Biochemistry and Cambridge Systems Biology Centre, University of Cambridge,
[†]Dept of Medicine, University of Cambridge, Cambridge Institute for Medical Research, Cambridge,
[#]Respiratory MDC, GlaxoSmithKline, Stevenage,
[‡]Dept of Investigative Preclinical Toxicology, Safety Assessment Division, GlaxoSmithKline R&D, Ware, and
[§]Centre for Inflammation Research, Queen's Medical Research Institute, Edinburgh, UK.
[§]GlaxoSmithKline R&D, King of Prussia, PA, USA.

CORRESPONDENCE

S.C. Connor
Dept of Biochemistry and Cambridge Systems Biology Centre
University of Cambridge
The Sanger Building
80 Tennis Court Road
Cambridge
UK
E-mail: susan.c.connor@googlemail.com

Received:
June 30 2011
Accepted after revision:
Nov 15 2011
First published online:
Dec 19 2011

This article has supplementary material accessible from www.erj.ersjournals.com

(LC-MS/MS) assay of amino acids was also carried out on a subset of the open-profiling metabolomics cohort to analytically validate the NMR results.

MATERIALS AND METHODS

Study design

ECLIPSE is a 3-yr longitudinal observational study in 40–75-yr-old current and ex-smoker COPD patients (current/ex-smokers n=2,756) and control subjects (current/ex-smokers n=343, nonsmokers n=223). A full description of the methods, including the computed tomography (CT) scanning procedures and the cohort, have been published [4, 5] and are summarised in the online supplementary material.

Serum samples for open-profiling metabolomics (n=244) were obtained from pair-matched control *versus* COPD subjects based on age, sex and smoking history. All patients and former smoker controls in the metabolomics cohort were ex-smokers to remove the potentially-confounding effects of current smoking. A subgroup of controls who had never smoked was included for comparison with the former smoker controls. Patients were categorised based using the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria (described in the online supplementary material): GOLD stages II, III and IV subjects in the NMR cohort showed similar FEV₁ to the whole ECLIPSE cohort (table 1). Targeted amino acid analysis used a subset of patients from the original NMR dataset (male smoker control n=30, GOLD IV n=30). Baseline fasting serum samples were obtained from morning blood samples at the initial clinic visit. Diet and over-the-counter pain or prescribed medications were recorded for the previous 24 h.

Open-profiling metabolomics

High-resolution proton Carr–Purcell–Meiboom–Gill NMR spectra were acquired on a Bruker Avance III NMR spectrometer (699.72 MHz; Bruker BioSpin, Rheinstetten, Germany). Methodological details for sample preparation and spectral acquisition are provided in the online supplementary material. Spectra were pre-processed using Bruker software (Topspin version 1.3 and Amix version 3.7.10; Bruker BioSpin, Karlsruhe, Germany) using a 0.005-ppm bucket width, and referenced and scaled to 4,4-dimethyl-4-silapentane-1-ammoniumtrifluoroacetate [7]. Spectra were assigned using two-dimensional NMR experiments (correlation spectroscopy, total correlation spectroscopy, heteronuclear single-quantum correlation and heteronuclear multiple-bond correlation [8]) and with reference to NMR assignment databases: Chemomx NMR Suite version 4.0 (Chemomx Inc., Edmonton, Canada); Human Metabolome Database (Genome Alberta (Alberta, Canada) and Genome Canada (Ottawa, Canada)); and Biological Magnetic Resonance Bank (BMRB; University of Wisconsin System, Madison, WI, USA) (table S1).

Amino acid analysis

A quantitative LC-MS/MS amino acid assay was developed to validate the open-profiling metabolomics results. Chloroformate amino acid derivatives (EZFaast; Phenomenex, Macclesfield, UK) were analysed by positive ion electrospray mass spectrometry (MS) using a Waters Acquity UPLC system (Waters Corp., Millford, MA, USA) and Waters Quattro Premier XE triple quadrupole mass spectrometer (Waters Corp.) (table S2). Additional

TABLE 1 Demographic data for the ECLIPSE subjects included in the metabolomics study and the ECLIPSE cohort

Group	Subjects	Age yrs	M/F	Smoking exposure pack-yrs	Post-bronchodilator FEV ₁ % pred	Reversibility % FEV ₁	Post-bronchodilator FEV ₁ /FVC	BMI kg m ⁻²	Fat-free mass kg	6MWD m	LAA %
Nonsmoker controls	15	61±8	8/7	0±0	111.6±12.3	0.7±6.2	78.7±5.5	27.7±7.6	54.2±12.5	454±113	4.4±6.3
Smoker controls	53	54.1±9	34/19	0.2±1.1	114.8±13.9	2.7±4.5	81.1±5.2	27.7±5.4	57.4±12	405±111	4.1±4.2
GOLD stage		55.4±9		31.6±21.5	108.6±12.0	4.5±5.8	79.2±5.2	26.8±4.6			2.4±3.1
II	69	65±6	46/23	48±34	62.9±7.7	10.7±10.1	51.4±9.4	27.9±5.1	53.5±10.8	405±111	12.9±10.6
III	63	63.7±2	43/20	48.1±27.5	63.1±8.4	11.3±12.6	52.8±8.8	27.4±5.7	51.8±12.8	379±118	12.1±9.5
IV	31	64.6±9	18/13	49.1±25.1	40.3±5.8	10.3±11.6	36.7±6.3	26.1±5.5	50.5±13.9	335±117	20.2±12.2
Non-emphysematous Emphysematous LC-MS/MS cohort	41	63±6	23/18	49±30	25±3.7	8.9±12.7	29.6±6.7	25.8±6.5	52.1±14	331±110	20±11.6
Smoker controls	30	64±6	58/19	49.2±26.7	24.7±3.6	8±13.8	32.1±7.4	25±5.6	51.5±10.7	289±119	24.3±15.1
GOLD stage IV	30	65±6	30/0	51±25	55.7±15.7	9.8±11.8	34.1±7.7	29±5.5	51.5±12.9	440±128	28.2±12.5
					38.6±14.9	9.7±10.6		25.7±5.4		351±124	30.7±8.9
					109.6±15.1	5.1±5.4	81±5.6	29.9±5.0	62.8±14.8	0±0	3.9±4.8
					24.6±4.6	8.4±14.9	30.4±7.4	26.2±5.1	56.7±10.1	330±159	24.9±13.3

Data are presented as n or mean±sd. The data in bold represents the data from the ECLIPSE cohort. ECLIPSE: Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points; M: male; F: female; FEV₁: forced expiratory volume in 1 s; % pred: % predicted; FVC: forced vital capacity; BMI: body mass index; 6MWD: 6 min walking distance; LAA: lung pixels with attenuation values <−950 HU (patients with emphysema were defined as >20% LAA and no emphysema as ≤10% LAA); GOLD: Global Initiative for Chronic Obstructive Lung Disease; LC-MS/MS: liquid chromatography–tandem mass spectrometry.

description of the assay methodology and characteristics are described in the online supplementary material.

Statistical analysis

Outliers and COPD-related effects were investigated using univariate data analysis and multivariate data analysis (MVDA). MVDA methods included unsupervised principal components analysis (PCA), using colour coding to highlight disease and confounding factors (GOLD stage, cachexia, emphysema, diabetes, patient location, age, sex and comorbidities). Using two-thirds of the dataset, supervised partial least squares (PLS) discriminant analysis (DA) were performed to maximise patient group differences (*i.e.* GOLD *versus* control, emphysema *versus* non-emphysema, cachexic *versus* non-cachexic), by rotating PCA components to improve separation among patient groups and highlight key variables. The remaining one-third of the data, "unseen" by the PLS-DA models, was used as independent data to assess the model relevance and error together with any data over-fit.

Statistical significance for individual metabolites was assigned using PLS-DA coefficients or unpaired t-test $\alpha < 0.05$, < 0.1 and < 0.15 , where α was the false discovery rate (FDR) multiplicity adjustment [9] at the 5, 10 and 15% level. The NMR data were

correlated with physiological and biochemical measurements using PLS and Pearson's correlation. Correlation significance was assigned using the t-distribution for FDR $\alpha < 0.05$ and ≤ 0.15 , with raw p-values < 0.01 .

RESULTS

Quality control and outlier detection

ECLIPSE baseline serum NMR spectra ($n=244$) were checked for acquisition and processing errors (fig. 1). PCA of processed NMR data was used to observe the overall variation in the data, technical and biological outliers, and sample groupings. Three poorly controlled diabetics were outliers and excluded from analysis. Current smokers ($n=5$), detected by cotinine assay, were excluded from analysis as low sample numbers precluded adequate analysis of the effects of smoking. Following exclusions, the main variation in principal component (PC) one was from very low-density lipoproteins (VLDLs)/chylomicrons and glucose, probably arising from variable compliance with overnight fasting. The PCA score and loadings plots were dominated by shifting of amino and organic acid resonances in GOLD stage IV COPD patients (fig. S1), consistent with variable serum pH. Further analysis was, therefore, performed on bucketed data summed across individual metabolite regions.

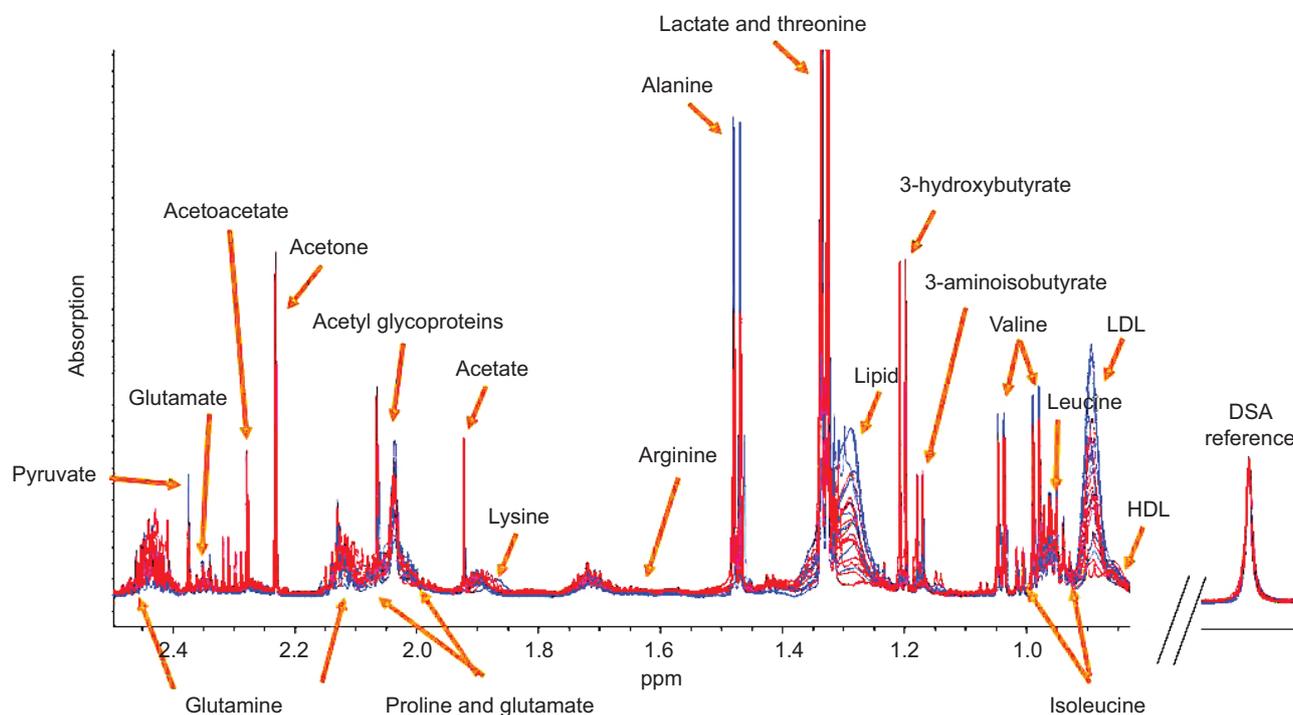


FIGURE 1. Partial 700-MHz ^1H nuclear magnetic resonance (NMR) spectra for Global Initiative for Chronic Obstructive Lung Disease stage IV patients (red) versus control subjects (blue). The individual compounds in the spectra were assigned using additional, more detailed NMR experiments (*i.e.* two-dimensional (2D) NMR: correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), heteronuclear single-quantum correlation (HSQC) and heteronuclear multiple-bond correlation (HMBC)) applied to ECLIPSE (Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points) cohort serum samples. These experiments showed which resonances corresponded to protons in the same molecule and provided additional information regarding neighbouring ^{13}C in the same molecule. Using the chemical shift (resonance position) information for ^1H and ^{13}C in the same molecule it was possible to assign the NMR signals using existing metabolomic databases. 2D NMR ^1H - ^1H COSY and TOCSY experiments highlight interactions between adjacent protons (proton spin-spin coupling). 2D ^1H - ^{13}C HSQC and HMBC experiments showed interactions between protons and their attached ^{13}C or ^{13}C up to three/four bonds away, respectively. Where possible, additional confirmatory evidence was obtained using Chenomx deconvolution software (Chenomx Inc., Edmonton, Canada). LDL: low-density lipoproteins; HDL: high-density lipoproteins; DSA: 4,4-dimethyl-4-silapentane-1-ammonium trifluoroacetate.

Serum metabolite differences between COPD patients and control subjects

NMR spectra of COPD patients were significantly different from controls by PCA, PLS-DA and t-test (fig. 2 and table 2). The PCA PC2 scores for patients with severe (GOLD stage III) and very severe (GOLD stage IV) COPD, and the majority of patients with moderate (GOLD stage II) COPD clustered separately from the controls (fig. 2). Because PCA is unsupervised (*i.e.* no prior group knowledge is used in the calculation) the partial clustering observed for COPD patients demonstrates that the second major source of variation in the data (after 33% of variation due to food intake (PC1)) was due to COPD. The partial COPD-related clustering was caused by increased glutamine, 3-methylhistidine, phenylalanine and dimethylglycine, acetoacetate, and 3-hydroxybutyrate, and decreased unsaturated lipid VLDL. The significant spectral regions containing these metabolites are detailed in the online supplementary material (tables S3–S5).

Supervised MVDA (PLS-DA) was used to maximise this observed difference between groups and the results compared with t-test results. The PLS-DA model ($R^2X=44\%$, $R^2Y=22\%$

and $Q^2=17\%$) showed similar metabolite changes to those observed by PCA plus increased acetate, monoglycerol fatty acids, creatine, ascorbate and aspartate, and decreased *O*-acetylated glycoproteins and glycerol (table 2).

Relationship between serum metabolites and disease severity

No GOLD stage II-related group pattern was observed. PCA of GOLD stages III and IV and control samples showed clustering of patients *versus* controls subjects (fig. 2). Supervised PLS-DA classification of GOLD stages III and IV *versus* control subjects, both separately and jointly, produced significant models. In addition to those metabolites highlighted in the patient *versus* control subjects comparison (table 2), decreased 3-hydroxyisobutyrate, isobutyrate, branched-chain amino acids (BCAAs), methionine, trimethylamine and high-density lipoprotein (HDL) were also observed in GOLD stage IV patients.

The PLS-DA model for moderate-to-severely affected (GOLD stages III and IV) patients *versus* controls subjects was generated from two-thirds of the data so that the remaining one-third

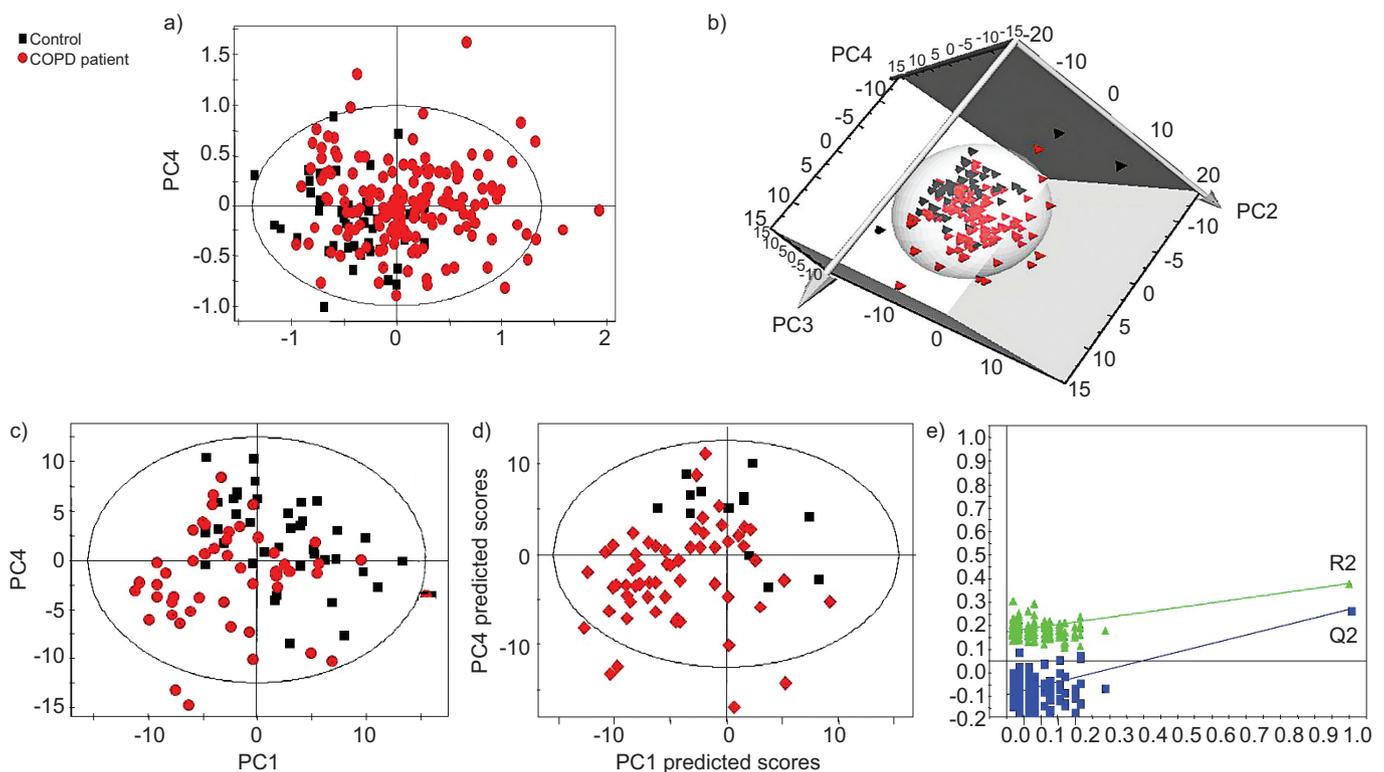


FIGURE 2. Score plots summarising the multivariate analysis of chronic obstructive pulmonary disease (COPD) patients *versus* control subjects. a) Two-dimensional score plot of principal components (PCs) 2–4 for a principal component analysis (PCA) model of all patients: Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages II, III and IV *versus* control subjects. The variance explained by the model (*i.e.* the sum of squares of all the X's explained by the extracted components (R²)) for PC1 was 33.5% and for PCs 2–4 was 34.4%. b) Three-dimensional (3D) PCA score plot of GOLD stages III and IV patients *versus* control subjects; R² for PC1 was 38.8% and for PCs 2–4 was 34.5%. The variation described by the first component for both a) and b) was mainly due to very low-density lipoproteins and glucose. c–e) Partial least squares (PLS) discriminant analysis (DA) predictive modelling of GOLD stage III plus GOLD stage IV COPD patients *versus* ex-smoker control subjects (male and female) showing score plots of c) the training set, d) the test set and e) random permutation test-validation plot. The training dataset used to generate the model consisted of equal numbers of COPD patients and controls. To test the accuracy of the model and degree of over-fit, the classes of independent samples “unseen” by the original model were predicted to calculate prediction error. The test set samples had fewer controls than patients due to the smaller number of controls in the dataset. A PLS-DA model was generated for 90 control subjects and GOLD stages III and IV patients (training set), and the scores predicted for 71 new samples using the model generated for the training set. New samples were predicted with good accuracy (82%; model diagnostics were R²X=52.2%, R²Y=33% and Q²=22.2%).

TABLE 2 Metabolites that significantly distinguish all chronic obstructive pulmonary disease patients (COPD) and Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages III and IV patients from control subjects, and emphysema patients from non-emphysema patients

Metabolic pathway/process	Metabolite [#]	All COPD patients versus control	GOLD III patients versus control	GOLD IV patients versus control	Emphysema versus non-emphysema patients
Microbial trimethylamine metabolism	Trimethylamine ^{+,5}			27.6 ⁺⁺	
Creatine biosynthesis/glycine degradation	Creatine ^{*,5}				-13.2 ⁺⁺
	Glycine ^{+,5,f}				-10.7 ⁺⁺
	<i>N,N</i> -dimethylglycine ^{*,+,5}	-30.6 ⁺⁺		-54 ⁵⁵	-35.4 ⁺⁺
BCAA degradation	3-hydroxyisobutyrate ^f			-17 ⁺⁺	-15.4 ⁺⁺
	Isobutyrate ^{+,5,f}			-14.5 ⁺⁺	
	Isoleucine ^{*,##}			-16.1 ⁺⁺	-11.3 ⁺⁺
	Leucine ^{*,##}				-10.9 ⁺⁺
	Valine ^{*,##}			-24.9 ⁺⁺	-12.3 ⁺⁺
Muscle-protein catabolism	3-methylhistidine ^{+,f}	37 ⁵⁵	43.2 ⁵⁵	53.5 ⁵⁵	34.5 ⁺⁺
Several pathways	Glutamine ^{*,5,##}	30.4 ⁵⁵	38.2 ⁵⁵	52.4 ⁵⁵	17.4 ⁺⁺
Cysteine and methionine metabolism	Methionine ⁵			-12 ⁺⁺	
Phenylalanine and tyrosine metabolism	Phenylalanine ^{*,+,5,f}	24.7 ⁵⁵	29.7 ⁵⁵	25.9 ⁺⁺	17.9 ⁺⁺
Lipid metabolism	HDL ^{*,##}			-15.4 ⁺⁺	-12.3 ⁺⁺
	LDL/VLDL ^{*,##}	-31 ⁺⁺		-31.5 ⁺⁺	-28.3 ⁺⁺
	Polyunsaturated lipid ^{*,+,##}	-24.2 ⁺⁺		-26.8 ⁺⁺	-21.7 ⁺⁺
	Monoglyceride ^{+,5,f}				-11.5 ⁺⁺
Ketone bodies	Glycerol ^{+,5,f}			-20.1 ⁺⁺	-24.6 ⁺⁺
	3-hydroxybutyrate ^{*,+,5,f}	15.3 ⁵⁵	20.3 ⁵⁵	26.6 ⁵⁵	17.1 ⁺⁺
	Acetoacetate ^{*,5}	25.9 ⁺⁺	42.2 ^{*,†}		
Vitamin	Ascorbate ^{f,##}	30.8 ⁺⁺		41.3 ⁺⁺	
Post-translational glycoprotein acetylation	<i>O</i> -acetylated glycoproteins [†]	-14.3 ^{*,†}			-16.2 ⁺⁺
	<i>N</i> -acetylated glycoproteins [†]				-12.7 ⁺⁺

Data presented as % positive or negative mean fold change (MFC). BCAA: branched-chain amino acid; HDL: high-density lipoproteins; LDL: low-density lipoproteins; VLDL: very low-density lipoproteins. [#]: unassigned metabolites also discriminated between each of the patient groups, the largest effect being an MFC of 106% (more information is included in the online supplementary material). [†]: assigned from the literature of TUKIAINEN and co-workers [10, 11]. ⁺: assigned from GlaxoSmithKline database of standards (unpublished data). ⁵: assigned using Chenomx software. ^f: assigned from the Human Metabolome Database. ^{##}: assigned from two-dimensional nuclear magnetic resonance correlation spectroscopy, total correlation spectroscopy, heteronuclear single-quantum correlation, and heteronuclear multiple-bond correlation on selected GOLD stage IV and control subjects' samples. The t-test raw p-value thresholds corresponding to α 0.05, 0.1 and 0.15 for each group were: GOLD III male and female $p < 0.0013$, $p < 0.0048$ and $p = 0.0075$, respectively; GOLD IV male and female $p < 0.0031$, $p < 0.011$ and $p = 0.0335$, respectively; all COPD patients (*i.e.* GOLD II, III, and IV male and female) $p < 0.0013$, $p < 0.0151$ and $p = 0.0266$, respectively; combined GOLD III and IV male and female *versus* control $p < 0.0027$, $p < 0.0110$ and $p = 0.0250$. ^{*,†}: false discovery rate (FDR) $\alpha < 0.15$; ⁺⁺: FDR $\alpha < 0.10$; ⁵⁵: FDR $\alpha < 0.05$.

remained "unseen" by the model. To add confidence that the results would be repeatable the PLS-DA models were tested with the remaining independent "unseen" data. Using the joint training model of GOLD stages III and IV *versus* control ($n=90$), 82% of new independent test samples ($n=71$) were classified correctly into either COPD patient or control subjects (fig. 2).

Relationship between serum metabolites and emphysema

Emphysema was assessed by a CT scan and described as the percentage of lung pixels with low attenuation (LAA) values < 950 HU (see online supplementary material for methodological detail). The relationship of NMR data to emphysema (defined by LAA) was calculated using two different approaches. The first used LAA as a category to split the data into two groups defining the presence (LAA $> 20\%$) or absence (LAA 0–10%) of emphysema, which were then analysed by

PLS-DA and t-tests. The second used LAA as a response variable that measured the extent of emphysema correlated with NMR data using PLS and Pearson's correlation.

Significant PLS-DA models and t-tests showed that BCAAs, aspartate, asparagine, glycine, creatine, *N,N*-dimethylglycine, glycerol, lipids (VLDL, LDL and HDL), *O*- and *N*-acetylglycoproteins, and 3-hydroxyisobutyrate were significantly decreased in emphysema patients, and that 3-methylhistidine, phenylalanine, glutamine and 3-hydroxybutyrate were found to have increased (table 2 and fig. 3). No significant models were obtained when control subjects were excluded, possibly due to the reduced power ($n=121$ *versus* $n=254$ with controls, respectively).

Emphysema severity was defined in PLS models by increasing glutamine, betaine and 3-hydroxybutyrate, and decreasing

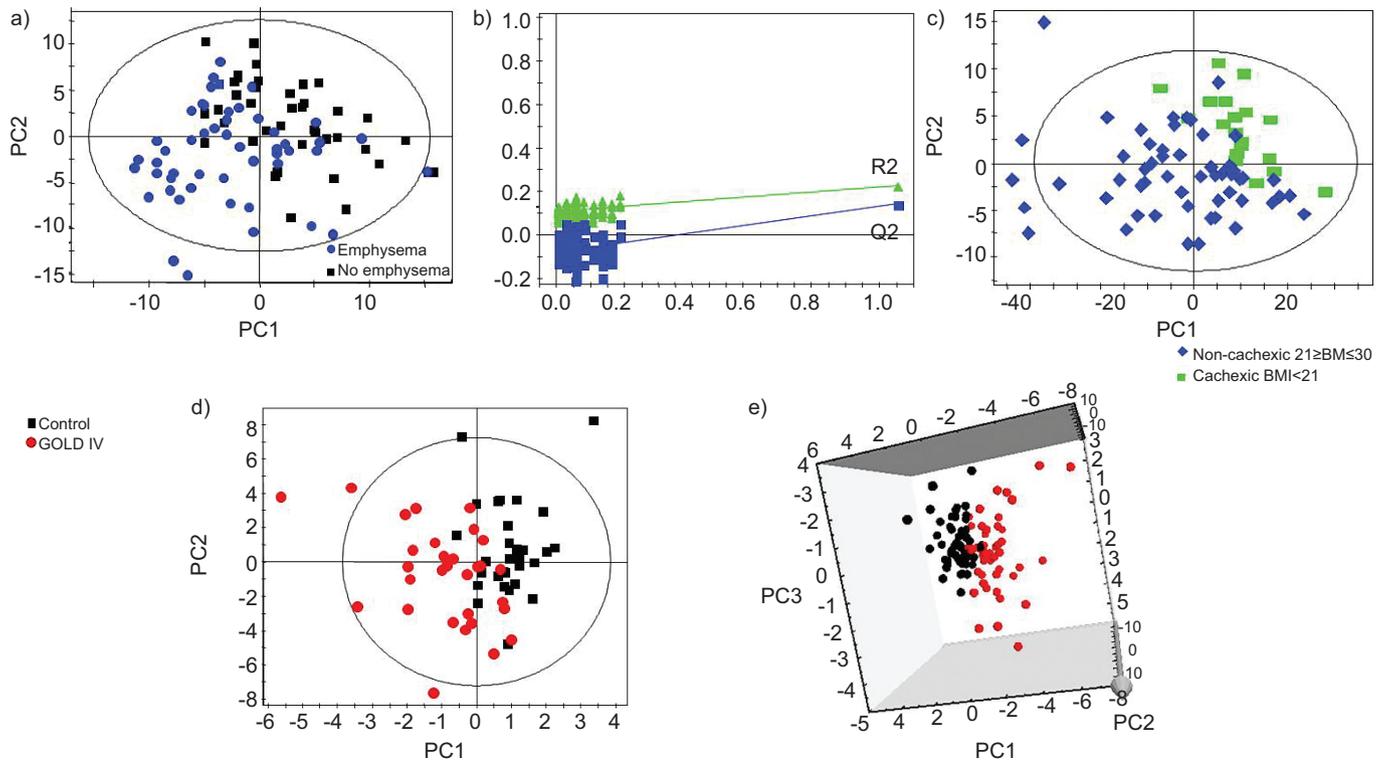


FIGURE 3. Scores and validation plots summarising the partial least squares (PLS) discriminant analysis (DA) of chronic obstructive pulmonary disease (COPD) patients a) with and without emphysema and c) with and without cachexia (excluding control patients). The random permutation validation test results for the emphysema model are shown in b); 70% of independent test data, not previously included in the modelling, were accurately predicted using the PLS-DA emphysema training set. d) and e) Two- and three-dimensional score plots of PLS-DA models for liquid chromatography with tandem mass spectrometry amino acid data; these plots highlights the differences between smoker control and GOLD stage IV; $R^2=45\%$, $Q^2=26\%$. All data were normalised according to serum creatinine levels.

glycerol and *O*-acetylglycoproteins. Significant Pearson's correlation coefficients with emphysema included 3-methylhistidine, glutamine, 3-hydroxybutyrate, α -ketoadipate and scyllo-inositol (positive correlation) and BCAAs, alanine, glycine, guanidinoacetate, dimethylglycine, histidine, lipids, glucose, glycerol, 3-hydroxyisobutyrate, isobutyrate, and *O*- and *N*-acetylated glycoproteins (negative; tables 3 and S3).

Relationship between serum metabolites and systemic inflammation

NMR data showed significant Pearson correlation coefficients with plasma C-reactive protein (CRP) and fibrinogen in patient and control groups. Fibrinogen levels were positively correlated with creatinine and log-transformed CRP levels with dimethylamine, 3-methylhistidine, glycine, creatinine, acetate, 3-aminoisobutyrate, 3-hydroxybutyrate and *cis*-aconitate in the patient group but not in the control group (table 3). PLS multivariate modelling highlighted that phenylalanine, histidine, alanine, *N*-acetylglycoproteins, glutamine, aspartate, 3-hydroxybutyrate, creatine, glycerol, citrate and valine predicted levels of CRP and fibrinogen, and hence are most closely associated with increased systemic inflammation.

Both plasma CRP and fibrinogen were significantly altered between control and the different GOLD stages (ANOVA *p*-values were 2.55×10^{-5} and 4.5×10^{-5} for log CRP and fibrinogen, respectively). *Post hoc* t-test analysis of the individual group

comparisons showed that CRP was significantly different between former smoker controls and GOLD stages II ($p=0.0014$), III ($p=0.0008$) and IV ($p=0.0005$). Fibrinogen levels were significantly different between control and GOLD stages III ($p=0.0003$) and IV ($p=5.71 \times 10^{-5}$), and also between GOLD stages II and III ($p=0.0127$) and IV ($p=0.0021$), but not between control and GOLD stage II ($p=0.0637$).

Relationship between serum metabolites and body mass index

PCA of the full dataset showed a body mass index (BMI)-related pattern in the score plot of PC1 and PC2, with cachexic subjects (BMI 15–21 $\text{kg}\cdot\text{m}^{-2}$) clustering separately from extremely obese subjects (BMI >40 $\text{kg}\cdot\text{m}^{-2}$) due to reduced serum levels of acetylated glycoproteins and BCAAs.

Significant Pearson correlation coefficients of NMR data with BMI and fat-free mass (FFM) were indicative of metabolites that depicted cachexia and body habitus. COPD-specific effects of cachexia, an extreme form of habitus, were determined by comparing correlation coefficients in patients with controls (table 3). Metabolites correlated with FFM for COPD patients but not controls were glycine, proline, glycerol, and creatine (negative), and methylamine, creatinine, *N,N*-dimethylglycine and polyunsaturated lipid (positive). Metabolites that were correlated to BMI in patients but not controls were glutamine, HDL, *N*-methylnicotinate and ascorbate (table 3).

TABLE 3 Metabolites that significantly correlate with markers of emphysema (FV950) and systemic inflammation (C-reactive protein (CRP) and fibrinogen)[#]

Biochemical pathway/process	BMI		FFM		CRP Pos	Fibrinogen Pos	FV950 [†] Pos [‡]
	Pos	Neg	Pos	Neg			
His metabolism, muscle protein catabolism [§]					3MH, His		3MH
BCAA degradation	Val, Leu, Ile, 3HIB, ISB		Val, Leu, Ile		Val, Leu, Ile, ISB, 3AIB		
Cr biosynthesis/Gly degradation [§]	Arg, Guan		Cn, DMG	Gly, Cr	Cr, Cn, Gly, Guan, Bet, Ch	Cn	
Phe and Tyr metabolism [§]	Tyr				Phe, Tyr, PAG (microbial)	Phe	
Ala, Asp+Glu/Gln+Glu metabolism [§]	Ala, Asp, Glu, Gln		Glu		Ala, Asp, Glu		Gln
Arg and Pro metabolism [§]	Arg			Pro			
Lys degradation [§]					Lys, AKA		AKA
Energy metabolism including glycolysis, gluconeogenesis and TCA cycle	Lac, Pyr, Glc				Glc, Cac, Cit, Ac		
Ketone bodies	3HB				3HIB, AcAc, Aco		3HIB
Nicotinate/nicotinamide metabolism		NMA					
Acetylated glycoproteins	OAc, NAc		OAc, NAc		OAc, NAc	NAc	
Vitamins	Asc				Myo	Asc	Scyllo
Lipid metabolism	HDL, LDL/VLDL, TG, MG, Poly, Glc		LDL/VLDL, Poly	MG, Glycl	TG, MG, PC, Glycl		
Microbial metabolism (gut flora); TMA/methane metabolism			TMA		Form		

No negative correlations were observed with CRP or fibrinogen. BMI: body mass index; FFM: fat-free mass; Pos: positive; Neg: negative; BCAA: branched-chain amino acids; Cr: creatine; Gly: glycine; Phe: phenylalanine; Tyr: tyrosine; Ala: alanine; Asp: aspartate; Glu: glutamate; Gln: glutamine; Arg: arginine; Pro: proline; Lys: lysine; TCA: tricarboxylic acid; TMA: trimethylamine; Val: valine; Leu: leucine; Ile: isoleucine; 3HIB: 3-hydroxyisobutyrate; ISB: isobutyrate; Guan: guanidinoacetate; Lac: lactate; Pyr: pyruvate; Glc: glucose; 3HB: 3-hydroxybutyrate; OAc: *O*-acetylglucoprotein; NAc: *N*-acetylglucoprotein; Asc: ascorbate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; VLDL: very low-density lipoprotein; TG: triglyceride; MG: monoglycerides; Poly: polyunsaturated lipid; Glycl: glycerol; NMA: *N*-methylnicotinic acid; Cn: creatinine; DMG: *N,N*-dimethylglycine; 3MH: 3-methylhistidine; His: histidine; 3AIB: 3-aminoisobutyrate; Bet: betaine; Ch: choline; PAG: phenylacetylglucine; AKA: α -ketoacid; Cac: *cis*-aconitate; Cit: citrate; Ac: acetate; AcAc: acetoacetate; Aco: aconitate; Myo: myo-inositol; PC: phosphorylcholine; Form: formate; Scyllo: scyllo-inositol; form: formate. [#]: full data including correlation coefficient values are shown in table S5; [†]: percentage of lung pixels with low attenuation on the whole lung scans with attenuation values of <-950 HU; [‡]: higher FV950 values indicated greater emphysema; [§]: amino acid metabolites generated and utilised during protein synthesis and degradation.

PLS modelling showed that increased VLDL, unsaturated lipid, glucose, valine, lactate, alanine, pyruvate, guanidinoacetate, glutamate and *O*-acetylated glycoproteins, and reduced glycine, glutamine, 3-hydroxybutyrate, phenylalanine, methionine, acetone, acetate, betaine, lysine, proline, aspartate and alanine (table 3) predicted BMI. In addition, increased BCAAs, creatinine, alanine, glutamate, betaine, acetone and *O*-acetylated glycoproteins, and decreased glycine, creatine, HDL, VLDL, glycerol, monoglycerides and acetate predicted FFM (tables 4, S4 and S5).

PLS-DA modelling of cachexic (BMI <21 kg·m⁻²) versus non-cachexic patients (BMI 21–30 kg·m⁻²) showed that in cachexic patients, glutamine, glycine, 3-hydroxybutyrate, acetate,

methionine and 3-aminoisobutyrate were increased, and valine, isoleucine, ascorbate, pyruvate and glucose were decreased.

Preliminary validation of amino acid changes using LC-MS/MS

Open-profiling metabolomics showed that several metabolites differed between control and COPD patient subgroups. In order to validate the amino acid changes analytically using a more sensitive and quantitative approach, targeted metabolomic experiments using LC-MS/MS were developed. The LC-MS/MS amino acid assay was then applied to a subset of the original samples of the most severely affected male patients (GOLD

TABLE 4 Changes in amino acid concentration in Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage IV patients versus control subjects, measured by liquid chromatography with tandem mass spectrometry

	Control subjects	GOLD IV patients
Decreasing amino acid in concentration nM		
ABA	655.8 ± 178.2	543.7 ± 191.9
4-HYP	12258.7 ± 11552.1	8600.4 ± 7001.7
AAA [#]	50.1 ± 19.2	40.9 ± 17.3
Pro [#]	3478.3 ± 490.1	3049.1 ± 965.1
Ile	1963.7 ± 408.1	1814.1 ± 585.2
Leu	3730.0 ± 725.5	3361.6 ± 1110.7
Val	4642.1 ± 1292.4	4190.2 ± 1319.8
Increasing amino acid concentrations nM		
AIB [#]	56.1 ± 22.5	80.3 ± 42.9
1-MH [#]	242.4 ± 54.6	282.5 ± 126.8
3-MH [#]	166.5 ± 182.9	461.7 ± 675.5
Asp [#]	239.3 ± 82.5	296.1 ± 176.3
Glu	8805.3 ± 1158.8	9198.0 ± 2716.4
GABA	9.2 ± 3.2	10.0 ± 3.4

Data are presented as mean ± SD. ABA: α -aminobutyrate; 4-HYP: 4-hydroxyproline; AAA: amino adipate; Pro: proline; Ile: isoleucine; Leu: leucine; Val: valine; AIB: aminoisobutyrate; 1-MH: 1-methylhistidine; 3-MH: 3-methylhistidine; Asp: aspartate; Glu: glutamine; GABA: γ -aminobutyrate. Principal component (PC)1 and PC2 $R^2=37.5\%$ and $Q^2=26.2\%$; PC3–PC5 $R^2=63.9\%$, $Q^2=36.3\%$. #: amino acids with p-values <0.1 (adjusted for using Benjamini–Hochberg test, false discovery rate $\alpha=0.1$).

stage IV; n=30) compared to controls (ex-smoker; n=30). A significant PLS-DA model (fig. 3; PC1 $R^2=37.5\%$ and $Q^2=26.2\%$) highlighted aminoisobutyrate as most predictive of GOLD stage IV versus control (table 4). BCAA coefficients were also high for the model, consistent with NMR findings, although absolute values did not show significant separation using t-tests.

DISCUSSION

Non-targeted metabolomics is well suited for biomarker discovery, allowing investigation of many molecules, with diverse chemical structures, simultaneously. The investigation of multiple markers facilitates the potential discovery of biomarker panels, which are more likely to be necessary than single entities to describe systemic effects of COPD. We have applied metabolomics to ECLIPSE serum samples to assist in the drive to discover biomarkers that may reflect patient outcome more predictively than FEV₁, which is currently used. Because ECLIPSE is the largest study currently attempting to better describe the subtypes of COPD [4], we have had unprecedented access to substantial amounts of patient information related to the characterisation of disease status, body composition, serum biochemistry, medication use and food intake. This has enabled us to put metabolomic data in context with the patient and look for associations that have not been possible previously (n=244, total number of serum samples in the dataset). Our results have confirmed that serum metabolomic profiles of COPD patients correlate with different characteristics of COPD severity assessed by spirometry, body composition, emphysema and systemic inflammation.

The use of independent test sets for PLS-DA models (*i.e.* data not used to generate the model), together with the observation of patient groupings in unsupervised PCA analysis and the application of FDR to t-test results reduce the chance of making a false-positive discovery (*i.e.* type one statistical error), giving

extra confidence in the statistical results. Without these tests the nature of all statistical tests indicates that at the 5–15% significance level, used in these studies, 15% of all the variables tested (*i.e.* ≥ 90) would be assigned as significant, even in a random dataset with no real significant effects.

Serum metabolite differences between COPD patients and control subjects

Because COPD-related effects were greater than variation due to sex, age, geographical location or any other potentially confounding factors, joint male and female models were used in analysis. Metabolomic profiles for all COPD patients depicted reduced food intake (decreased VLDL, LDL and chylomicrons), increased utilisation of ketone bodies as an alternative energy source (3-hydroxybutyrate) and decreased post-translational protein modification/protein synthesis (acetylated glycoproteins). All COPD patients also showed increased concentrations of serum methylhistidine, an amino acid formed during cross-linking of muscle protein synthesis, suggesting increased muscle actin and myosin protein degradation [12], even in patients that were not cachexic.

Relationship between serum metabolites and disease severity

As with many inflammatory markers, particularly markers of systemic inflammation, significant changes did not appear until more severe disease stages (GOLD III and IV) with no correlation or significant models observed for the GOLD stage II group. In severely affected patients (GOLD stage IV), increased degradation of muscle protein (3-methylhistidine) and utilisation of BCAAs as an alternative source of energy for muscle and adipose tissue were indicated.

Several findings are reflective of altered protein metabolism (methylhistidine, BCAAs and their degradation product

3-hydroxyisobutyrate) associated with disease as well as the extent of disease. As well as being correlated with BMI, COPD patients, independent of cachexia, have increased levels of glutamine and ketone bodies, and decreased VLDL/LDL and O-acetylglycoproteins.

Relationship between serum metabolites and emphysema

In individuals with emphysema, increased degradation of muscle protein (3-methylhistidine) and utilisation of BCAAs were indicated. Increased serum methylhistidine in patients with emphysema was not correlated to either BMI or FFM, suggesting that increased muscle protein turnover is a feature of COPD related to emphysema that precedes or is unrelated to the development of cachexia.

Disrupted creatine synthesis and glycine degradation occurred in this group, with lower serum levels of many intermediates (table 2 and 3).

Relationship between serum metabolites and systemic inflammation

For individuals experiencing high systemic inflammation (higher CRP levels) increased degradation of muscle protein (3-methylhistidine) and utilisation of BCAAs were indicated.

Relationship between serum metabolites and BMI

Cachexia, defined as low BMI and FFM, was characterised more extensively by protein muscle breakdown and decreased BCAAs and their degradation products. The metabolomic results can be subdivided into those that reflect general changes in body habitus and those that reflect distinct differences in COPD patients. Correlations between NMR data and BMI and FFM in controls indicate that many cachexia-related changes are also associated with altered general habitus (*i.e.* BMI changes; table S4). These include positive correlation with BCAAs, alanine, creatinine, choline, betaine, VLDL/LDL, ketone bodies, acetate, 3-hydroxyisobutyrate, citrate, lactate, pyruvate, glycerol and acetylated glycoproteins. The PLS-DA model of cachexia *versus* non-cachexia and Pearson correlations in the patient group show that some metabolites are only altered in COPD patients and, therefore, are not reflective of general habitus (HDL, ascorbate, *N*-methylnicotinate, glutamine (negative correlation)).

Preliminary validation of amino acid changes using LC-MS/MS

The LC-MS/MS targeted amino-acid profiling served as a preliminary analytical validation study of the NMR results. The main role of this study was to explore whether there was consistency between both analytical platforms prior to extension of the targeted amino acid study to the full sample set. This was successful using LC-MS/MS of male GOLD stage IV patients *versus* control subjects, where MS results were in agreement with NMR results within the different limitations and detection limits for each platform (tables 2 and S5). The reduced power of the small study meant that whilst trends were observed in the amino acids detected by NMR only one of these was consistent after accounting for repeat measures. A batch and run order effect was also observed in the full sample set; methods of elimination of run order and batch effects for MS data are currently being explored and the analysis of this dataset will be reported separately.

Biochemical role of metabolite changes and possible implication for ECLIPSE patients

Decreased BCAAs have been found in previous studies of COPD patients [13–15]. BCAAs regulate protein turnover and glucose homeostasis [16] with a continuous flux of BCAA to skeletal muscle. Skeletal muscle produces 90% of total glutamine [16], which in turn is a substrate for proliferative cells including immune cells [17]. Muscle proteolysis and transamination of BCAAs *via* branched-chain aminotransferase is part of an essential physiological function to provide carbon for gluconeogenesis during fasting. Cachexic patients who are losing weight demonstrate increased rates of gluconeogenesis not suppressed by glucose [18]; BCAAs may represent a physiological response to weight loss, consistent with the strong correlation of BCAA levels with BMI in data from the ECLIPSE cohort.

Approximately 25% of COPD patients develop cachexia, which is associated with a poor prognosis [19]. Cachexia is characterised by altered BCAA availability, due partly to anorexic effects of COPD, *i.e.* low food intake accompanied by elevated oxidation of BCAAs and gluconeogenesis. Protein degradation to increase BCAAs supply and glucogenic amino acids occur at the same time as suppression of protein synthesis. Changes observed in the NMR data, *i.e.* reduced BCAAs and increased glutamine and alanine are consistent with cachexia/increased protein degradation within this patient group. Increases in other essential amino acids including phenylalanine were consistent with increased protein degradation. Consistently altered 3-methylhistidine suggests increased muscle protein turnover precedes or is independent of the onset of cachexia.

In this study, metabolites that correlated strongly with cachexia markers also correlated strongly with systemic inflammation. In fasting and starving states [20], short-term levels of BCAAs rise in parallel with increased protein degradation. After several weeks, the rate of muscle protein degradation decreases and, therefore, blood concentration of BCAAs decreases as ketone-body production increases. Metabolomic data for COPD patients showed increased ketone bodies and decreased BCAAs, and in emphysema patients there was also a decrease in serum creatine and its precursors, glycine, guanidinoacetate and dimethylglycine (fig. 4). This suggests a reduced mitochondrial function, consistent with recent reports in the early stages of COPD where skeletal muscle mitochondria show an electron transport chain blockade and excessive production of reactive oxygen species [21].

Further work

Serum and urine are available from each time-point in the ECLIPSE study. Expansion of the current work will include investigation of serum samples longitudinally to determine which markers track with disease progression and are most successful at prediction of patient outcome. These studies will include LC-MS/MS amino acid targeted analysis to obtain quantitative values for biomarkers of interest and exploration of any sex bias of significant effects. Urine data have already been acquired for the baseline time-point; open-profiling NMR and LC-MS results will be reported elsewhere. The combination of metabolomics data with other types of gene transcript, genetic and protein data at both the numerical and pathway level is in progress to investigate significant pathways involved with COPD.

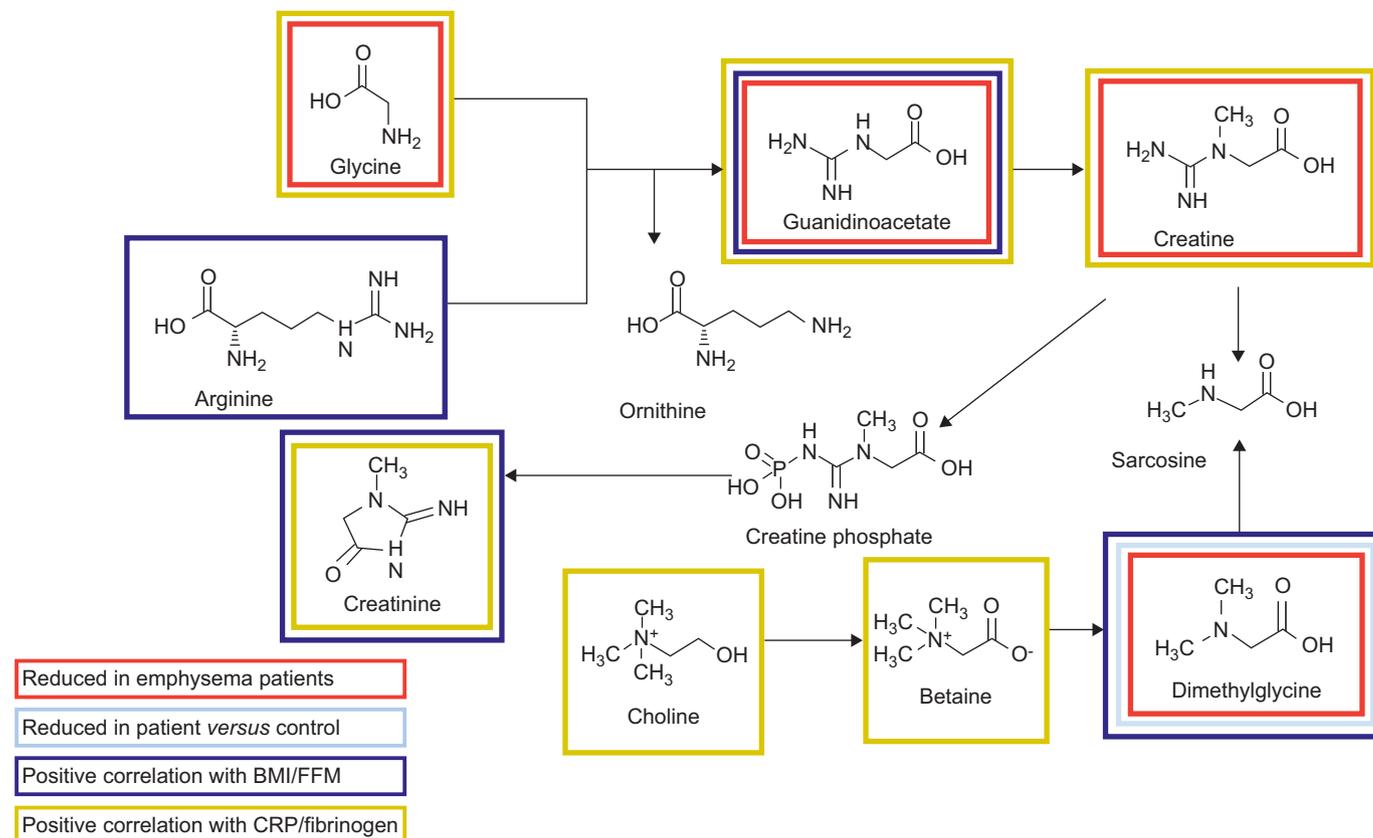


FIGURE 4. Creatine biosynthesis and glycine degradation pathway showing metabolites that were altered in chronic obstructive pulmonary disease patients. Each arrow represents one step in the pathway. BMI: body mass index; FFM: fat-free mass; CRP: C-reactive protein.

Ultimately, biomarker validation studies in other patient cohorts and investigation of normal variation of candidate biomarkers in human epidemiological studies are required. It is also planned to explore the different metabolomics profiles and proposed biomarker panels in relation to different drug treatments.

Conclusions

We have applied metabolomics to the analysis of serum samples from ECLIPSE patients. ECLIPSE is the largest COPD study of its type and, as such, we have had access to patient characterisation on a much larger scale than previous metabolomics studies of COPD. Due to the increased power of the experiments, we have been able to characterise metabolomics profiles characteristic of cachexia, inflammation, emphysema and disease severity, as well as distinguishing spectra from patients and controls. Thus, we have demonstrated that NMR-based metabolomics provides additional phenotypic information in COPD patients based on a combination of their systemic effects and lung-function parameters. The main effects suggest increased protein turnover in all COPD patients together with increased protein degradation in emphysema and cachexic patients. Metabolic evidence of reduced food intake was observed in COPD patients, which correlated with reduced BMI. Evidence of reduced mitochondrial oxidation efficiency was also indicated by the correlation of ketone bodies with clinical parameters. A subset of these markers (*e.g.* BCAAs, methylhistidine, phenylalanine, acetylated glycoproteins, 3-hydroxypyruvate, serum

lipid and ascorbate) could provide a useful biomarker panel by which to assess and stratify COPD patients according to their protein turnover, mitochondrial function and nutritional status.

SUPPORT STATEMENT

This study was funded by GlaxoSmithKline.

STATEMENT OF INTEREST

Statements of interests for B.K. Ubhi, J.H. Riley, P.A. Shaw, D.A. Lomas, R. Tal-Singer, W. MacNee, S.C. Connor and the study itself can be found at www.erj.ersjournals.com/site/misc/statements.xhtml

ACKNOWLEDGEMENTS

The authors would like to thank the Depts of Sample Management, Statistical Sciences, and Investigative Preclinical Toxicology at GlaxoSmithKline (Stevenage and Ware, respectively, UK) for technical assistance and scientific input. We would also like to thank B.C. Sweatman (investigative Preclinical Toxicology Dept, GlaxoSmithKline, Ware, UK) for NMR analysis and S.M. Fox (Biostatistics, Stevenage Medicines Development, GlaxoSmithKline, Stevenage, UK) for statistical advice.

REFERENCES

- 1 Buist AS, McBurnie MA, Vollmer WM, *et al.* International variation in the prevalence of COPD (the BOLD study): a population-based prevalence study. *Lancet* 2007; 370: 741–750.
- 2 Lopez AD, Shibuya K, Rao C, *et al.* Chronic obstructive pulmonary disease: current burden and future projections. *Eur Respir J* 2006; 27: 397–412.

- 3 Barton RH. A decade of advances in metabonomics. *Expert Opin Drug Metab Toxicol* 2011; 7: 129–136.
- 4 Vestbo J, Anderson W, Coxson HO, *et al.* Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J* 2008; 31: 869–873.
- 5 Agusti A, Calverley PMA, Celli B, *et al.* Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respir Res* 2010; 11: 122–136.
- 6 Pauwels RA, Buist AS, Calverley PMA, *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop Summary. *Am J Respir Crit Care Med* 2001; 163: 1256–1276.
- 7 Alum MF, Shaw A, Sweatman BC, *et al.* 4,4-Dimethyl-4-silapentane-1-ammonium trifluoroacetate (DSA), a promising universal internal standard for NMR-based metabolic profiling studies of biofluids, including blood plasma and serum. *Metabolomics* 2008; 4: 2 122–127.
- 8 Simpson JH. Organic Structure Determination using 2-D NMR Spectroscopy: A Problem-Based Approach (Advanced Organic Chemistry). Burlington, Academic Press, 2008.
- 9 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc Series B* 1995; 57: 289–300.
- 10 Tukiainen T, Tynkkynen T, Mäkinen VP, *et al.* A multi-metabolite analysis of serum by 1H NMR spectroscopy: early systemic signs of Alzheimer's disease. *Biochem Biophys Res Commun* 2008; 375: 356–361.
- 11 Nicholson JK, Foxall PJD, Spraul M, *et al.* 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. *Anal Chem* 1995; 67: 793–811.
- 12 Young VR, Munro HN. N τ -methylhistidine (3-methylhistidine) and muscle protein turnover: an overview. *Fed Proc* 1978; 37: 2291–2300.
- 13 Hofford JM, Milakofsky L, Vogel WH, *et al.* The nutritional status in advanced emphysema associated with chronic bronchitis. A study of amino acid and catecholamine levels. *Am Rev Respir Dis* 1990; 141: 902–908.
- 14 Schols AMWJ, Duetz NEP, Mostert R, *et al.* Plasma amino acid levels in patients with chronic obstructive pulmonary disease. *Monaldi Arch Chest Dis* 1993; 48: 546–548.
- 15 Engelen MPKJ, Wouters EFM, Duetz NEP, *et al.* Factors contributing to alterations in skeletal muscle and plasma amino acid profiles in patients with chronic obstructive pulmonary disease. *Am J Clin Nutr* 2000; 72: 1480–1487.
- 16 Layman DL, Walker DA. Branched chain amino acids: metabolism, physiological function and application. *J Nutrition* 2006; 136: 319S.
- 17 Thompson D. Cancer: genes, cachexia and death. In: Newsholme E, Leech T, eds. *Functional Biochemistry in Health and Disease*. Chichester, John Wiley & Sons Ltd, 2010; pp. 485–507.
- 18 Baracos VE, Mackenzie ML. Investigations of branched-chain amino acids and their metabolites in animal models of cancer. *J Nutri* 2006; 136: 237S–242S.
- 19 Wagner PD. Possible mechanisms underlying the development of cachexia in COPD. *Eur Respir J* 2008; 31: 492–501.
- 20 Tom A, Nair S. Assessment of branched-chain amino acids and potential for biomarkers. *J Nutri* 2006; 136: 324S–330S.
- 21 Puente-Maestu L, Pérez-Parra J, Godoy R, *et al.* Abnormal mitochondrial function in locomotor and respiratory muscles of COPD patients. *Eur Respir J* 2009; 33: 1045–1052.