Metabolic profiling detects biomarkers of protein degradation in COPD patients

[1]Baljit K. Ubhi, [2]John H. Riley, [3]Paul A. Shaw, [4]David A. Lomas, [5]Ruth Tal-Singer, [6]William MacNee, [1]Julian L Griffin, [1]Susan C Connor

[1] Department of Biochemistry and Cambridge Systems Biology Centre, University of Cambridge, the Sanger Building, 80 Tennis Court Road, Cambridge, UK

[2] Respiratory MDC, Stevenage, GlaxoSmithKline, UK

[3] Department of Investigative Preclinical Toxicology, Safety Assessment Division, GlaxoSmithKline R & D, Park Road, Ware, Hertfordshire UK

[4] Department of Medicine, University of Cambridge, Cambridge Institute for Medical Research, Cambridge, UK

[5] GlaxoSmithKline, King of Prussia, PA, USA

[6] Centre for Inflammation Research, Queen's Medical Research Institute, Edinburgh Scotland, UK,

Corresponding Author to whom requests for reprints and correspondence should be addressed: Dr Susan C Connor, Department of Biochemistry and Cambridge Systems Biology Centre, University of Cambridge, the Sanger Building, 80 Tennis Court Road, Cambridge, UK. Scc59@cam.ac.uk, susan.c.connor@googlemail.com

Abstract

Rationale

There is a paucity of well-validated biomarkers for chronic obstructive pulmonary disease (COPD). Metabolomics, which identifies novel biomarkers, was applied to a well-defined COPD patient cohort from the ECLIPSE study. Results were correlated with accepted biomarkers of disease.

Methods

Baseline serum from controls (n=66) and GOLD II (n=70), III (n=64) and IV (n=44) COPD patients were analysed by NMR-based metabolomics. LC-MS/MS targeted metabolomics was used to confirm amino acid changes detected by NMR. Data were correlated with body composition, emphysema and systemic inflammation.

Measurements and Main Results

Open profiling metabolomics identified decreased lipoproteins (VLDL/chymicrons; LDL) and N, N-dimethylglycine and increased glutamine, phenylalanine, 3-methylhistidine and ketone bodies in COPD patients with decreased branched chain amino acids (BCAAs) also observed in GOLD 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 had decreased serum creatine, glycine and N, N-dimethylglycine. LC-MS/MS confirmed NMR findings relating to BCAAs, glutamine and 3-methylhistidine in GOLD IV patients.

Conclusion

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.

[2]

[3]

Introduction

Chronic obstructive pulmonary disease (COPD) is a major health burden worldwide, affecting 10 to 15% of the adult population aged 40 years and older (1). Three 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 manifested by progressive airflow limitation. Characterisation of COPD patients has relied on the measurement of airflow limitation, as assessed by the forced expiratory volume in one second (3), which has been used to define the condition, its severity and treatment. However, it is now well recognized that COPD is a heterogeneous disease with features that are not captured by the measurement of FEV_1 (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 characterized group of individuals with COPD from the ECLIPSE cohort (Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints ECLIPSE, SCO104960, ClinicalTrials.gov identifier NCT00292552 (6)). A targeted analysis using a bespoke 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.

[4]

Materials and Methods

Study Design

ECLIPSE is a 3 year longitudinal observational study in 40-75 year-old current and ex-smoker COPD patients (current/ex-smokers n=2756) and control subjects (current/ex-smokers n=343; non-smokers n= 223). A full description of the methods, including CT scanning method, and the cohort is published (4, 5) and summarised in an online data supplement.

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 on GOLD criteria (described in the on-line data supplement): GOLD stage II to IV subjects in the NMR cohort showed similar FEV_1 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 hours.

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

[5]

provided in an online data supplement. Spectra were pre-processed using Bruker software (Topspin (version 1.3); 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 2D NMR experiments (COSY, TOCSY, HSQC and HMBC (8), and with reference to metabolomics databases (9) (Table E1).

Amino acid analysis

A Quantitative LC-MS/MS amino acid assay was developed to validate the open-profiling metabolomics results. Chloroformate amino acid derivatives (EZ: Faast, www.phenomenex.com) were analysed by positive ion electrospray MS using a Waters Acquity UPLC system and Waters Quattro Premier XE triple quadupole mass spectrometer (Waters Corp, Millford, MA, US) (Table E2). Additional description of the assay methodology and characteristics are described in the on-line data supplement.

Statistical Analysis

Outliers and COPD-related effects were investigated using univariate (UVDA t-test; Microsoft excel 2003) and multivariate (MVDA; Umetrics Simca-P version 11.5) data analysis. 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 co-morbidities). Using two thirds of the dataset, supervised partial least squares (PLS) - discriminant analyses (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

[6]

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 t-test alpha (α) <0.05, <0.1 and <0.15 results, where α was the false discovery rate (FDR) multiplicity adjustment (10) 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 (α) <0.05 and up to 0.15, with a raw p value less than 0.01.

Results

Quality control and outlier detection

ECLIPSE baseline serum NMR spectra (n=244) were checked for acquisition and processing errors (Figure 1A). 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 (VLDL)/chylomicrons and glucose, probably arising from variable compliance with overnight fasting. The PCA scores and loadings plots were dominated by shifting of amino and organic acid resonances in GOLD IV COPD patients (Figure 1B), consistent with variable serum pH. Further analysis was therefore performed on bucketed data summed across individual metabolite regions.

[7]

Serum metabolite differences between COPD patients and control subjects

NMR spectra of COPD patients were significantly different from controls by PCA, PLS-DA and Student's t-test (Figure 2; Table 2). The PCA PC2 scores for patients with severe (GOLD III) and very severe (GOLD IV) COPD and the majority of patients with moderate (GOLD II) COPD clustered separately from controls (Figure 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 (PC 1)) 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 on-line supplement (Tables E3-E5).

Supervised MVDA (PLS-DA) was used to maximize this observed difference between groups and the results compared with t-test results. The PLS-DA model (R^2 X 44%, R^2 Y = 22% and Q^2 = 17%) showed similar metabolite changes to those observed by PCA plus increased acetate, monoglyceryl fatty acids, creatine, ascorbate and aspartate, and decreased O-acetylated glycoproteins and glycerol (Table 2).

Relationship between serum metabolites and disease severity (GOLD status)

No GOLD II-related group pattern was observed. PCA of GOLD III, GOLD IV and control samples showed clustering of patients versus controls (Figure 2B). Supervised PLS-DA classification of GOLD III and GOLD IV versus control, both separately and

[8]

jointly, produced significant models. In addition to those metabolites highlighted in the patient versus control comparison (Table 2), decreased 3-hydroxyisobutyrate, isobutyrate, branched chain amino acids (BCAAs), methionine, trimethylamine, and HDL were also observed in GOLD IV patients.

The PLS-DA model for moderate to severely affected (GOLD stage III-IV) patients versus controls was generated on two thirds of the data so that the remaining one third 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 III and GOLD IV versus control (n=90) 82% of new independent test samples (n=71) were classified correctly into either COPD patient or control subjects (Figure 2).

Relationship between serum metabolites and emphysema

Emphysema was assessed by CT scan and described as the percentage lung pixels with low attenuation values < -950 Hounsefield units (%LAA) (see online supplement 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 measuring 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

[9]

emphysema and 3-methylhistidine, phenylalanine, glutamine and 3-hydroxybutyrate were increased (Table 2; Figure 3). No significant models f 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 glycerol and O-acetylglycoproteins. Significant Pearson's correlation coefficients with emphysema included 3-methylhistidine, glutamine, 3-hydroxybutyrate, alpha ketoadipate and scylloinositol (positive correlation) and BCAAs, alanine, glycine, guanidinoacetate, dimethyglycine, histidine, lipids, glucose, glycerol, 3-hydroxyisobutyrate, isobutyrate, O- and N-acetylated glycoproteins (negative; Tables 3 and E3).

Relationship between serum metabolites and systemic inflammation

NMR data showed significant Pearson correlation coefficients with plasma CRP and fibrinogen in patient and control groups. Fibrinogen levels were positively correlated with creatinine and log CRP with dimethylamine, 3-methylhistidine, glycine, creatinine, acetate, 3-aminoisobutyrate, 3-hydroxybutyrate and cis-aconitate in the patient group but not controls (Table 3). PLS multivariate modeling 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 value =2.55E-05 and 4.5E-05 for log CRP and

[10]

fibrinogen respectively). Post hoc ttest analysis of the individual group comparisons showed that CRP was significantly different between former smoker controls and GOLD stage II (p=0.0014), stage III (p=0.0008) and stage IV (p=0.0005). Fibrinogen levels were significantly different between control and GOLD stage III (p=0003) and IV (p=5.71E-05) and also between GOLD stage II and stage III (p=0.0127) and IV (p=0.0021), but not between control and GOLD stage 2 (p=0.0637).

Relationship between serum metabolites and BMI

PCA of the full dataset showed a BMI-related pattern in the scores plot of PC 1 and PC 2 with cachexic subjects (BMI 15 - 21) clustering separately from extremely obese subjects (BMI>40) 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 and proline, glycerol, (negative, opposite sense to controls), 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 (NMA) and ascorbate (Table 3).

PLS modeling showed that increased VLDL, unsaturated lipid and glucose valine, lactate, alanine, pyruvate, guanidinoacetate, glutamate and O-acetylated glycoproteins predicted BMI and reduced glycine, glutamine, 3-hydroxybutyrate, phenylalanine, methionine,

[11]

acetone, acetate, betaine, lysine, proline, aspartate and alanine (Table 3). 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, E4 and E5).

PLS-DA modeling of cachexic (BMI <21) versus non-cachexic patients (BMI 21-30) showed that in cachexic patients glutamine, glycine, 3-hydroxybutyrate, acetate, methionine and 3-aminoisobutyrate were increase 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 liquid chromatography-tandem mass spectrometry (LC-MS/MS) were developed. The LC-MS/MS amino acid assay was then applied to a subset of the original samples of most severely affected male patients (GOLD IV; n=30) compared to controls (former smoker; n=30). A significant PLS-DA model (Figure 3D; PC 1 R2 = 37.5% and Q2 = 26.2%) highlighted aminoisobutyrate (AIB) as most predictive of GOLD 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 of diverse chemical structure, simultaneously. The investigation of

[12]

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 current FEV_1. Because ECLIPSE is the largest study currently attempting to better describe the subtypes of COPD (Vestbo *et al*.) we have had unprecedented access to substantial amounts of patient information related to characterization of disease status, body composition, serum biochemistry, medication use and food intake. This has enabled us to put metabolomics data in context of the patient and look for associations that have not been possible (n=244) previously. 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 1 statistical error), giving extra confidence in the statistical results. Without these tests the nature of all statistical tests indicates at the 5-15 percent significance level used in these studies 15 percent of all the variables tested (ie at least 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 gender, age,

[13]

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 utilization 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 (11), even in patients that were not cachexic.

Relationship between serum metabolites and disease severity (GOLD status)

As with many inflammatory markers, particularly markers of systemic inflammation, significant changes did not appear until more severe COPD (GOLD III and IV) with no correlation or significant models observed for the GOLD II group. In severely affected patients (GOLD IV) increased degradation of muscle protein (3-methylhistidine) and utilization 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

[14]

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

Relationship between serum metabolites and BMI

Cachexia, depicted by low BMI and FFM, was characterized 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; supplementary table E4.). 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

[15]

only altered in COPD patients and therefore are not reflective of general habitus (HDL, ascorbate, NMA, 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 IV versus control patients, where MS results were in agreement with NMR results within the different limitations and detection limits for each platform (Tables 2 and 5). 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 (12-14). BCAAs regulate protein turnover and glucose homeostasis (15) with a continuous flux of BCAA to skeletal muscle. Skeletal muscle produces 90% of total glutamine (15) which in turn is a substrate for proliferative cells including immune cells (16). Muscle proteolysis and transamination of BCAAs via branched chain aminotransferase (BCKDH) is part of

[16]

an essential physiological function to provide carbon for gluconeogenesis during fasting. Weight-losing cachexic patients demonstrate increased rates of gluconeogenesis not suppressed by glucose (17); 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 (18). Cachexia is characterized 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 occurs 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 (19) 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. Metabolomics data for COPD patients shows increased ketone bodies and decreased BCAAs and in emphysema patients there is also decreased serum creatine and its precursors, glycine, guanidinoacetate and dimethylglycine (Figure 4). This suggests reduced mitochondrial function consistent with recent reports that even in the early stages of COPD skeletal muscle mitochondria show

[17]

electron transport chain blockade and excessive production of reactive oxygen species (20).

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 gender bias of significant effects. Urine data has 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.

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 characterization on a much larger scale than previous metabolomics studies of COPD. Due to the increased power of the experiments we have been able to characterize metabolomics

[18]

profiles characteristic of cachexia, inflammation, emphysema and disease severity as well as distinguishing spectra from patients and controls. We have thus 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.

Acknowledgements

References

1. Buist AS, McBurnie MA, Vollmer WM, Gillespie S, Burney P, Mannino DM, Menezes AM, Sullivan SD, Lee TA, Weiss KB, Jensen RL, Marks GB, Gulsvik A, Nizankowska-Mogilnicka E; BOLD Collaborative Research Group.International

[19]

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, Mathers C D, Hansell A L, Held L S, Schmid V, Buist S. 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 Endpoints (ECLIPSE). *Eur Respir J* 2008; 31:869–73.

5. Agusti A, Calverley PMA, Celli B, Coxson HO, Edwards LD, Lomas DA, MacNee W, Miller BE, Rennard S, Silverman EK, Tal-Singer Wouters RE, Yates JC, Vestbo J. Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respiratory Research* 2010; 11:122-136.

6. Pauwels RA, Buist SA, Calverley PMA, Jenkins CR, and HURD SS on behalf of the GOLD Scientific Committee 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 PA, Sweatman BC, Ubhi BK, Haselden JN and Connor SC. 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, J H, Organic Structure Determination using 2-D NMR: A Problem-Based Approach (Advanced Organic Chemistry, 2008, Academic press, Elsevier.

[20]

9. NMR assignment databases Chenomx NMR Suite v4.0 (Chemomx Inc., Edmonton, Canada, HMDB (www.hmdb.ca), and BMRB (www.bmrb.wisc.edu)).

10. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 1995; 57(1):289–300.

11. Young VR, Munro HN. Ntau-methylhistidine (3-methylhistidine) and muscle protein turnover: an overview. *Fed Proc* 1978; 37(9):2291-300.

12. Hofford JM, Milakofsky L, Vogel WH, Sacher RS, Savage GJ, and Pell S. 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.

13. Schols AMWJ, Duetz NEP, Mostert R and Wouters EFM. Plasma amino acid levels in patients with chronic obstructive pulmonary disease. *Monaldi Arch Chest Dis* 1993; 48:546-8.

14. Engelen M, Wouters EFM, Duetz NEP, Menheere PPCA and Schols AMWJ. Factors contributing to alterations in skeletal muscle and plasma amino acid profiles in patients with chronic obstructive pulmonary disease. *Am J Clin Nutr* 2000; 72:1480-7.

15. Layman DL and Walker DA. Branched chain amino acids: metabolism, physiological function and application. *J Nutrition* 2006; 136(1): 319S.

16. Newsholme E and Leech T. Cancer: Genes, Cachexia and Death. In: Wiley-Blackwell. Functional Biochemistry in health and disease. 2010, Chapter 21 p485-507.

17. Baracos and Mackenzie. Investigations of branched chain amino acids and their metabolites in animal models of cancer. *J Nutrition* 2006; 136(1):237(S).

[21]

18. Wagner PD. Possible mechanisms underlying the development of cachexia in COPD. *Eur Respir J* 2008; 31:492-501.

19. Tom A and Nair S. Branched chain amino acids: metabolism, physiological function and applications, *J Nutrition* 2006; 136(1): 324S.

20. Puente-Maestu L, Pérez-Parra J, Godoy R, Moreno N, Tejedor A, González-Aragoneses F, Bravo JL, Alvarez FV, Camaño S, Agustí A. Abnormal mitochondrial function in locomotor and respiratory muscles of COPD patients. *Eur Respir J* 2009; 33(5):1045-52.

[22]

Figure legends

Figure 1: Partial 700 MHz Proton NMR spectra of GOLD IV versus control patients

(A) Scaled spectra between 0 – 2.5 ppm; Red = Gold IV patients, blue= control subjects. The individual compounds in the spectra were assigned using additional, more detailed NMR experiments (2D NMR COSY, TOCSY, HSQC and HMBC) applied to ECLIPSE serum samples. These experiments showed which resonances corresponded to hydrogens in the same molecule and provided additional information regarding neighbouring carbons in the same molecule. Using the chemical shift (resonance position) information for hydrogens (protons) and carbons in the same molecule it was possible to assign the NMR signals using existing metabolomic databases. 2D NMR ^{1}H-^{1}H COSY and TOCSY experiments highlight interactions between adjacent protons that interact (proton spin-spin coupling). 2D ^{1}H-^{13}C HSQC and HMBC experiments show interactions between protons and their attached carbons or carbons up to 3-4 bonds away respectively. Where possible additional confirmatory evidence was obtained using Chenomx deconvolution software.

(B) Bucketed NMR data showing the histidine (right) and tyrosine (left) multiplets; COPD patients (red), control subjects (black), shifted signals (blue box plus arrow).
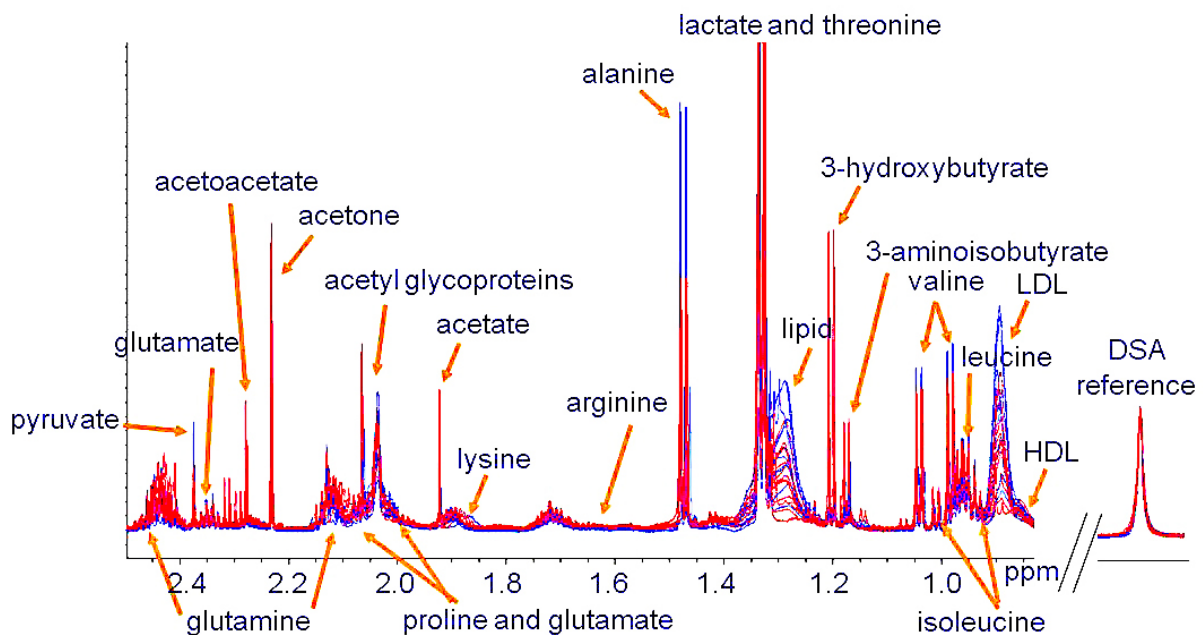

Figure 2: Scores plots summarising the multivariate analysis of COPD patients versus controls

(A) 2D scores plot of PCs 2-4 for a PCA model of all patients (GOLD II, III and IV; red) versus control (black); 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) 3D PCA scores plot of GOLD III and GOLD IV patients (red) versus controls (black); 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 VLDL and glucose

[23]

(C) PLS-DA predictive modelling of GOLD III plus GOLD IV COPD patients versus control (male and female) showing scores plots of the training set (left), test set (middle) and random permutation test validation plot (right); Red= Gold III + GOLD IV, black = former smoker controls. 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 overfit, 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 controls, GOLD 3 and GOLD 4 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%, Q2 22.2).
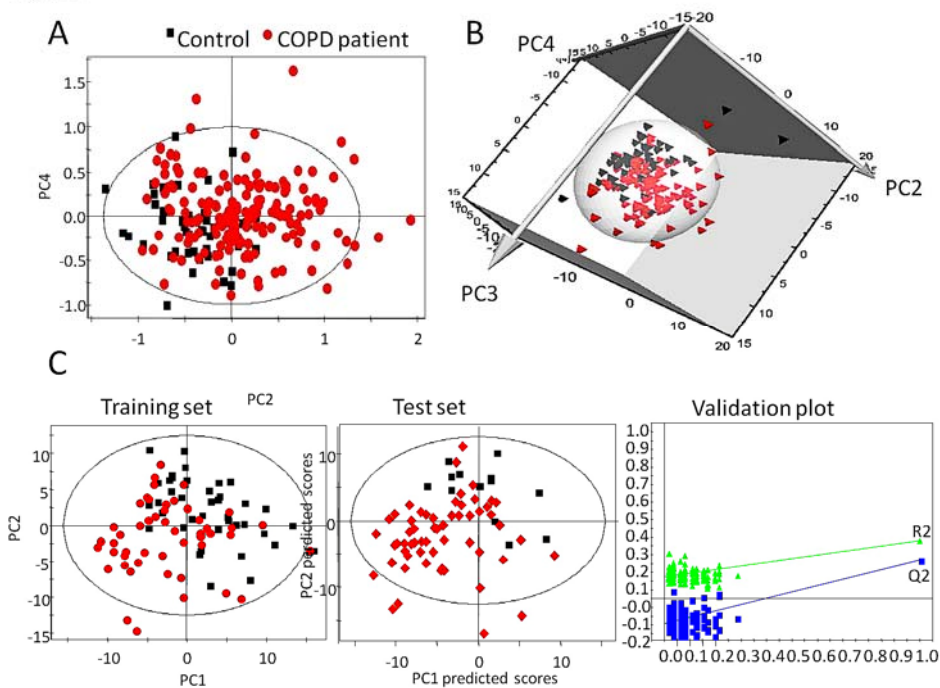

Figure 2

[24]

Figure 3: Scores and validation plots summarising the PLS-DA analysis of COPD patients with and without emphysema (A &B) and with and without cachexia (C) excluding controls and PCA scores plot colour coded according to BMI ranges (D) and loadings plot (E) of the full dataset

- (A) Scores plot for PLS-DA training sets of COPD patients with and without emphysema; Model diagnostics were R2X 52%, R2Y 32%, Q2 22% (2 components); 70% of independent test data was accurately predicted using the PLS-DA training set emphysema model;
- (B) Random permutation test of emphysema PLS-DA model
- (C) Scores plot for PLS-DA training sets of COPD patients with and without cachexia; R2X 56%, R2Y 36%, Q2 17% (2 components)
- (D) 2D and 3D scores plots of PLS-DA modelling for LC-MS/MS amino acid data highlighting differences between smoker control and GOLD IV; R2 = 45.2%, Q2 = 25.7%; all data were normalised according to serum creatinine levels
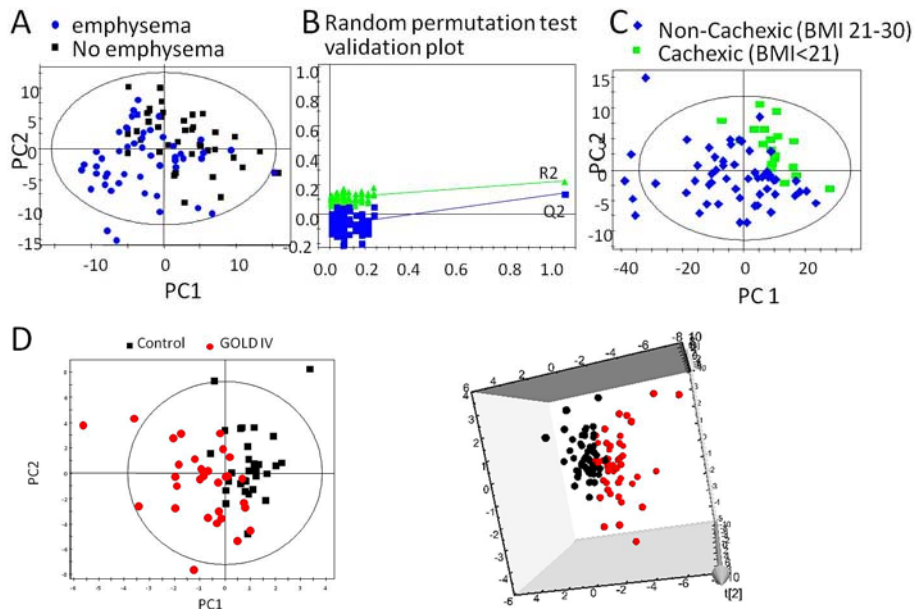

Figure 4: Creatine biosynthesis/glycine degradation pathway showing metabolites that were altered in COPD patients, coloured according to significance (see tables for significance values). Each arrow represents one step in the pathway.
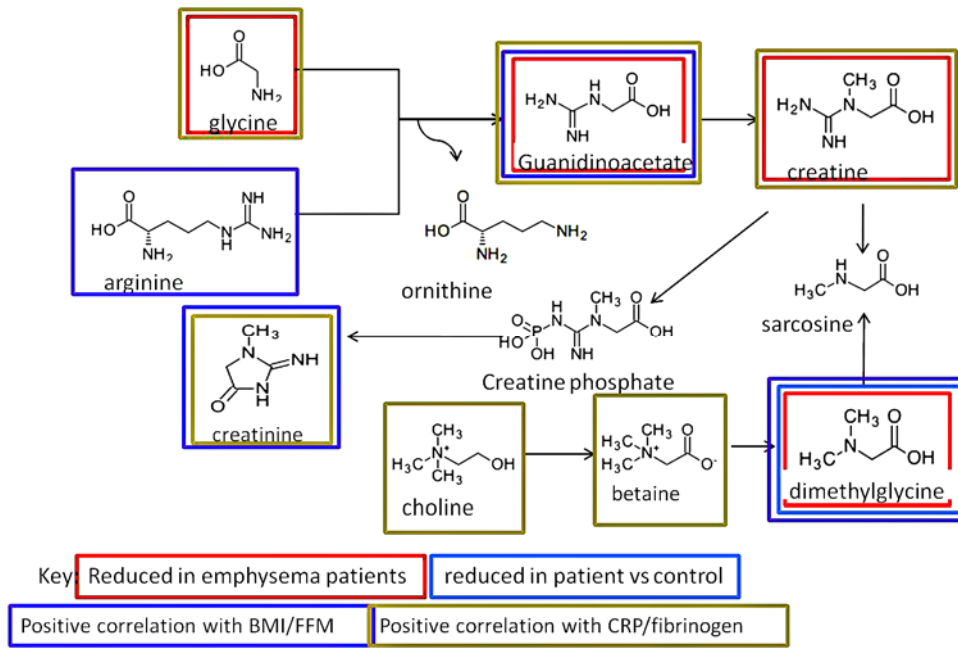
Key:

| Reduced in emphysema patients | reduced in patient vs control |
| Positive correlation with BMI/FFM | Positive correlation with CRP/fibrinogen |

[26]

Table 1: Demographic data of the Eclipse subjects included in metabolomics study *[including values for full ECLIPSE cohort]*

Group	No	Age (years) mean (SD)	M:F	Mean pack years	Post bronchodilator FEV1% predicted mean (SD) (l)	% reversibility FEV1 mean (SD)	Post broncho-dilator mean FEV1/FVC %(SD)	Body mass index (kg/m2) mean(SD)	Fat free mass mean (SD)	6MWD (metres)	%LAA
Non-smoker controls	15	61 (8) *[54.1 (9)]*	8:7	0 (0) *[0.2 (1.1)]*	111.6 (12.3) *[114.8 (13.9)]*	0.7 (6.2) *[2.7 (4.5)]*	78.7 (5.5) *[81.1 (5.2)]*	27.7 (7.6) *[27.7 (5.4)]*	54.2 (12.5)		4.4 (6.3) *[4.1 (4.2)]*
Smoker controls	53	57 (9) *[55.4 (9)]*	34:19	29 (16) *[31.6 (21.5)]*	108 (11.5) *[108.6 (12.0)]*	4.8 (9.7) *[4.5 (5.8)]*	80.6 (5.7) *[79.2 (5.2)]*	28.6 (4.7) *[26.8 (4.6)]*	57.4 (12)		2.9 (3.8) *[2.4 (3.1)]*
Stage II	69	65 (6) *[63 (7.2)]*	46:23	48 (34) *[48.1 (27.5)]*	62.9 (7.7) *[63.1 (8.4)]*	10.7 (10.1) *[11.3 (12.6)]*	51.4 (9.4) *[52.8 (8.8)]*	27.9 (5.1) *[27.4 (5.7)]*	53.5 (10.8)	454 (113) *[405 (111)]*	12.9 (10.6) *[12.1 (9.5)]*
Stage III	63	64 (6) *[64 (6.9)]*	43:20	50 (25) *[49.1 (25.1)]*	39.3 (5.8) *[40.3 (5.8)]*	10.3 (11.6) *[11 (14.6)]*	36.7 (6.3) *[40.5 (8.9)]*	26.9 (5.3) *[26.1 (5.5)]*	51.8 (12.8)	379 (118) *[335 (117)]*	20.2 (12.2) *[20 (11.6)]*
Stage IV	31	63 (6) *[62 (6.9)]*	18:13	49 (30) *[49.2 (26.7)]*	25 (3.7) *[24.7 (3.6)]*	8.9 (12.7) *[8 (13.8)]*	29.6 (6.7) *[32.1 (7.4)]*	25.8 (6.5) *[25 (5.6)]*	50.5 (13.9)	331 (110) *[289 (119)]*	24.3 (15.1) *[28.2 (12.5)]*
non-emphysematous patients	41	64 (6)	23:18	37 (25)	55.7 (15.7)	9.8 (11.8)	51.5 (10.7)	29 (5.5)	52.1 (14)	440 (128)	6 (2.6)
emphysematous patients	77	64 (6)	58:19	51 (25)	38.6 (14.9)	9.7 (10.6)	34.1 (7.7)	25.7 (5.4)	51.5 (12.9)	351 (124)	30.7 (8.9)
LC-MS/MS cohort Smoker controls	30	57 (9)	30:0	29 (18)	109.6 (15.1)	5.1 (5.4)	81 (5.6)	29.9 (5)	62.8 (14.8)	0 (0)	3.9 (4.8)
LC-MS/MS cohort Stage IV	30	65 (6)	30:0	52 (34)	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)

[27]

Table 2: Metabolites that significantly distinguish all patients from controls, GOLD III and GOLD IV patients from control and emphysema from non-emphysema patients (quoted as positive or negative mean fold change (MFC))

Metabolite[7]	metabolic pathway/process	All COPD patients vs control: % MFC	Gold III patients vs control: % MFC	Gold IV patients vs control: % MFC	Emphysema vs no emphysema patients % MFC
trimethylamine (TMA) (4), (5)	Microbial TMA metabolism			-27.6**	
creatine (1), (4)	creatine biosynthesis/glycine degradation[1]				-13.2**
glycine (1), (2), (5)					-10.7**
NN dimethylglycine (1), (4), (5)		-30.6**		-54***	-35.4**
3-hydroxyisobutyrate (2)	Branched chain amino acid degradation			-17**	-15.4**
isobutyrate (1), (2), (5)				-14.5**	
isoleucine (3), (4)				-16.1**	-11.3**
leucine (3), (4)					-10.9**
valine (3), (4)				-24.9**	-12.3**
3-methylhistidine (2), (5)	Muscle protein catabolism[1]	+37***	+43.2***	+53.5***	+34.5**
glutamine (1), (3), (4)	several pathways[1]	+30.4***	+38.2***	+52.4***	+17.4**
Methionine (met) (1)	Cysteine & met metabolism[1]			-12**	
phenylalanine (phe) (1), (2), (4), (5)	phe & tyrosine metabolism[1]	+24.7***	+29.7***	+25.9**	+17.9**
HDL (3) (4)	Lipid metabolism			-15.4**	-12.3**
LDL/VLDL (3) (4)		-31**		-31.5**	-28.3**
polyunsaturated lipid (3) (4) (5)		-24.2**		-26.8**	-21.7**
Monoglyceride (1) (2) (5)					-11.5**
glycerol (1), (2), (5)				-20.1**	-24.6**
3-hydroxybutyrate (1), (2), (4), (5)	Ketone bodies	+15.3***	+20.3***	+26.6***	+17.1**
acetoacetate (1), (4)		+25.9**	+42.2*		
ascorbate (2), (3)	vitamin	+30.8**		+41.3**	
O-acetylated glycoproteins (4)	Post translational glycoproteins acetylation	-14.3*			-16.2**
N-acetylated glycoproteins (4)					-12.7**

(1) assignment using chenomx software (Chenomx NMR Suite v 4.0 (Chenomx, Inc.)
(2) HMDB (Wishart DS, Knox C, Guo AC, et al. HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res. 2009 37(Database issue):D603-610)

[28]

(3) 2D COSY, TOCSY, HMBC and HSQC on selected gold IV and control samples

(4) literature information - (ala korpela, 2008, biochem, biophys res. Com. 375 plus Nicholson references)

(5) In-house GSK database of standards

(6) http://lipidlibrary.aocs.org/nmr/1NMRglyc/index.htm, match with standards

(7) Several unassigned metabolites also discriminated between each of the patient groups, the largest effect being a MFC of +106%; more information is included in the supplementary on-line information

The ttest raw p value thresholds corresponding to alpha, α = 0.05, 0.1 and 0.15 for each group were: GOLD 3 male and female (G3M&F) p <0.0013, <0.0048 and 0.0075; GOLD 4 male and female (G4M&F) p<0.0031, <0.011 and 0.0335; All COPD patients (ie GOLD 2,3 and 4 male and female) (ALL) p <0.0013, <0.0151 and 0.0266; combined Gold 3 and 4 male and female versus control (G3&4) <0.0027, <0.0110 and 0.0250. * FDR α <0.15; . ** α <0.1; . *** α <0.0.05

[29]

Table 3: Metabolites that significantly correlate with markers of emphysema (FV950) and systemic inflammation (CRP and Fibrinogen); no negative correlations were observed to CRP or fibrinogen[1,2]

Biochemical pathway/process	BMI		FFM		CRP	Fibrinogen	FV950[4]
	Positive	Negative	Positive	Negative	positive	positive	positive[5]
His metabolism, muscle protein catabolism[3]					3MH, his		3MH
BCAA degradation[3]	val, leu, ile, 3HIB, ISB		val, leu, ile		val, leu, ile, ISB, 3AIB		
cr biosynthesis/gly degradation[3]	arg, guan		cn, DMG	gly, cr	cr, cn, gly, guan, bet,cho	cn	
phe and tyr metabolism[3]	tyr				phe, tyr, PAG(microbial)	phe	
ala, asp& glu / gln & glu metabolism[3]	ala, asp, glu, gln		glu		ala, asp, glu,		gln
arg and pro metabolism[3]	arg			pro			
lys degradation[3]					lys, AKA		AKA
Energy metabolism including glycolysis, gluconeogenesis and TCA cycle	lac, pyr, gluc				gluc, cac, cit, ac		
ketone bodies	3HB				3HB, acac, aco		3HB
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, glcl		LDL/VLDL, poly	MG, glycerol	TG, MG, PC, glcl		
Microbial metabolism (gut flora); TMA/methane metabolism			TMA		form		

Table 4: Changes in amino acid concentration (nmoles) in GOLD IV patients vs Controls, measured by LC-MS/MS

DECREASING

		ABA	4-HYP	AAA*	PRO*	ILEU	LEU	VAL
Control	Mean	655.8	12258.7	50.1	3478.3	1963.7	3730.0	4642.1
	SD	178.2	11552.1	19.2	490.1	408.1	725.5	1292.4
Gold IV	Mean	543.7	8600.4	40.9	3049.1	1814.1	3361.6	4190.2
	SD	191.9	7001.7	17.3	965.1	585.2	1110.7	1319.8

INCREASING

		AIB*	1-MH*	3-MH*	ASP*	GLU	GABA
Control	Mean	56.1	242.4	166.5	239.3	8805.3	9.2
	SD	22.5	54.6	182.9	82.5	1158.8	3.2
Gold IV	Mean	80.3	282.5	461.7	296.1	9198.0	10.0
	SD	42.9	126.8	675.5	176.3	2716.4	3.4

Footnote: Highlighted (*) amino acids with p values <0.1 (adjusted for using Benjamini-Hochberg test, FDR α = 0.1). ABA = α-Amniobutyrate, 4-HYP = 4-Hydroxyproline, AAA = Aminoadipate, Pro = Proline, ILEU = Isoleucine, LEU = Leusine, VAL = Valine, AIB = Aminoisobutyrate, 1-MH = 1-Methyhistidine, 3-MH = 3-Methylhistidine, ASP = Aspartate, GLU = Glutamine and GABA = γ-Aminobutyrate. PC1&2 R2 = 37.5% and Q2 = 26.2%; PC's3-5 R2 = 63.9%, Q2 = 36.3

[31]