Association of circulating cell-free double-stranded DNA and metabolic derangements in idiopathic pulmonary fibrosis

William Whalen,1 Mustafa Buyukozkan,2 Bethany Moore,3 Jong-Seok Moon,4 Charles S Dela Cruz,5,6 Fernando J Martinez,1 Augustine M K Choi,1 Jan Krumsiek,2 Heather Stout-Delgado,1 Soo Jung Cho 1

ABSTRACT

Idiopathic pulmonary fibrosis (IPF) is a progressive and fatal lung disease with unclear aetiology and poorly understood pathophysiology. Although plasma levels of circulating cell-free DNA (ccf-DNA) and metabolic changes have been reported in IPF, the associations between ccf-DNA levels and metabolic derangements in lung fibrosis are unclear. Here, we demonstrate that ccf-double-stranded DNA (dsDNA) is increased in patients with IPF with rapid progression of disease compared with slow progressors and healthy controls and that ccf-dsDNA associates with amino acid metabolism, energy metabolism and lipid metabolism pathways in patients with IPF.

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive lung disease characterized by epithelial cell damage, proliferation and activation of fibroblasts and extracellular matrix accumulation which portends a poor prognosis with an estimated mean survival of 2–5 years from time of diagnosis. While the etiology of disease progression and exacerbation is unknown, there is growing evidence that double-stranded DNA (dsDNA) may play an important role in IPF pathogenesis.2 Also, we have recently reported that absent in melanoma 2 (AIM2) inflammasome, a multiprotein complex that recognises dsDNA, plays a critical role in IPF pathogenesis and that AIM2 is regulated by glucose metabolism.3 To advance the knowledge of the role of dsDNA in IPF pathogenesis, we examined plasma levels of circulating cell-free dsDNA (ccf-dsDNA) in patients with IPF and identified metabolites and metabolic pathways associated with ccf-dsDNA levels.

METHODS

The detailed method is included in the online supplemental material.

RESULTS

Subjects in this study were patients with IPF from the Correlating Outcomes with Biochemical Markers to Estimate Time to Progression in Idiopathic Pulmonary Fibrosis (COMET-IPF) trial, a multicenter prospective observational study correlating biomarkers with IPF disease progression.2 Rapid progressors were defined as patients who met any part of a composite outcome (death, acute exacerbation of IPF, relative decline in forced vital capacity of at least 10% or carbon monoxide transfer factor of 15%) during the 80-week follow-up. Healthy subjects served as controls (figure 1A). Demographics and clinical characteristics of study subjects are summarised in figure 1B and online supplemental table 1.

To investigate the potential role of ccf-dsDNA in IPF, ccf-dsDNA was isolated from plasma of 98 patients with IPF (39 slow progressors, 59 rapid progressors) and 28 healthy controls (figure 1B). Plasma concentrations of ccf-dsDNA were higher in rapid progressor patients relative to those in slow progressors and healthy controls (figure 1C). ccf-dsDNA levels were not significantly different between the slow progressor and control groups.

Untargeted metabolomics assays run on 799 metabolites revealed that 73 metabolites were positively correlated with ccf-dsDNA concentration, while 59 were negatively correlated (figure 2A). Clear clustering of metabolites by ccf-dsDNA levels was noted. From this analysis, we identified two distinct groups of metabolites (group 1 and group 2). Interestingly, the levels of ccf-dsDNA of group 1 were significantly higher than group 2. Differentially changed metabolites were annotated mapped to KEGG pathways. The 15 pathways (10 amino acid pathways, 2 energy pathways, 2 nucleic acid pathways, 1 lipid pathway) with the highest number of correlating metabolites are highlighted in figure 2B. The complete list of significantly associated pathways is shown in online supplemental table 2 and online supplemental figure 1.

There was notable correlation between amino acid metabolism and ccf-dsDNA, especially in the alanine/aspartate/glutamate, arginine, glutathione and taurine/hypouraine metabolic pathways. Additionally, several lipid metabolic pathways were associated with ccf-dsDNA levels. Key lipid pathways included sphingolipid synthetized and fatty acid metabolic pathway. ccf-dsDNA was also associated with energy metabolism pathways such as glycolysis and tricarboxylic acid (TCA) cycle.

Amino acid metabolism

In addition, specific metabolites pertinent to amino acid pathways were found to correlate with ccf-dsDNA levels (figure 3A). Significant associations existed between ccf-dsDNA and metabolites of the arginine/proline pathway. Arginine is one of the essential amino acids in animal cells, serving as a
precursor for the synthesis not only of proteins but also of nitric oxide, urea, polyamines, proline, glutamate and creatine. In our study, arginine and creatinine were negatively correlated with ccf-dsDNA, whereas glutamate, spermidine and ornithine were positively correlated. Of note, ornithine converts to proline and hydroxyproline for collagen synthesis during fibrogenesis, while spermidine has been shown to serve as an antioxidant in murine fibrosis models.6

We also found associations between ccf-dsDNA and metabolites of the glutathione pathway. L-cysteinyl-glycine, L-glutamate, 5-oxoproline, spermidine and L-ornithine were positively correlated with ccf-dsDNA, while L-gamma-glutamyl-L amino acid was negatively correlated with ccf-dsDNA (figure 3A). Glutathione, the most abundant antioxidant, regulates gene expression, DNA and protein synthesis, cell proliferation and apoptosis, signal transduction, cytokine production and protein glutathionylation. A prior study showed that cysteine, glycine and glutamate, components of glutathione synthesis, are increased in IPF lung compared with controls, consistent with our current observations in plasma.

Like glutathione, taurine is an antioxidant that has diverse cytoprotective function including regulation of antioxidation, energy metabolism, endoplasmic reticulum (ER) stress, neuromodulation, quality control and calcium homeostasis.3 We found that taurine/hypotaurine metabolism is significantly associated with ccf-dsDNA in IPF plasma. Specifically, hypotaurine, taurine, pyruvate, L-alanine, 2-oxo-glutarate and L-glutamate were positively correlated with ccf-dsDNA levels, whereas 3-sulfino-L-alanine was negatively correlated with ccf-dsDNA (figure 3A). This is an important finding as taurine/hypotaurine metabolism has not previously been identified as being altered in lung fibrosis in either animal or human metabolomics studies.

Energy metabolism ccf-dsDNA levels were also found to be associated with energy metabolism pathways (figure 3B), specifically glycolysis and TCA cycle pathways. Glycolysis is the first step in the breakdown of glucose to extract energy for cellular metabolism. In our study, glucose levels negatively correlated with ccf-dsDNA level, while lactate, 3-phosphoglycerate and pyruvate, later-stage glycolysis metabolites, were positively correlated with ccf-dsDNA. The glycolytic byproducts measured in plasma in our study may represent increase in glycolytic activity in IPF lungs, which has previously shown to be a key mediator of lung fibrosis.9 ccf-dsDNA levels were also associated with the TCA cycle pathway which, along with glycolysis, is the key pathway for cellular energy production. We found significant positive correlations between ccf-dsDNA and succinate,
fumarate, malate, 2-oxo-glutamate and pyruvate. Conversely, citrate and aconitate were negatively correlated with ccf-dsDNA. These findings in plasma complement data from prior studies showing that TCA cycle metabolites are increased in IPF lung tissue. Lipid metabolism
Finally, our study identified significant changes in metabolites involved in sphingolipid metabolism, a bioactive lipid pathway that plays pivotal roles in various cellular processes including inflammation, apoptosis and fibrosis. In our patients with IPF, sphinganine, phosphoethanolamine, sphingosine and sphinganine-1-phosphate were positively correlated with ccf-dsDNA levels (figure 3C). This is an interesting finding in light of a previous study showing a reduction of sphingolipid metabolism in IPF lung. The opposing phenomenon observed between plasma and lung tissue may possibly be explained by a surplus of plasma lipid metabolites arising from an underutilisation of these metabolites in lung tissue.

CONCLUSION
We found that ccf-dsDNA levels are elevated in patients with IPF who are rapid progressors compared with healthy controls and slow progressors and are furthermore associated with alterations in multiple central metabolic pathways. This serves as an interesting complement to prior observations that these same metabolic pathways are significantly changed in IPF lung tissue. The source of ccf-dsDNA in patients with IPF is unknown. Most existing data are from patients with cancer whose ccf-dsDNA appears to come from different sources, including the tumour cells that circulate in peripheral blood in addition to normal cell types such as haematopoietic and stromal cells. In patients with IPF, a recent study showed that mitochondria may be a major source of ccf-dsDNA.

The association between ccDNA and various metabolites is
central to our study. Although we did not perform mechanistic studies in our paper, there is evidence to suggest that metabolic reprogramming regulates dsDNA-dependent-immune responses via cytosolic DNA sensors such as AIM2 inflammasome and cGAS (GMP–AMP synthase)–STING (stimulator of interferon genes). As an example, glucose metabolism, sphingolipid

Figure 3 The linear regression graph showing significant metabolites in selected pathways using univariate linear regression. Amino acid metabolism pathways (A), energy metabolism pathways (B) and sphingolipid metabolism pathway (C) are associated with ccf-dsDNA levels. ccf-dsDNA, cell-free double-stranded DNA.

metabolism and fatty acid metabolism all contribute to dsDNA-dependent AIM2 inflammasome activation. Evidence linking cGAS–STING signalling and metabolic dysregulation is scarce, which suggests that further investigation is warranted.

In sum, our results suggest ccf-dsDNA both as a potential biomarker for IPF disease progression and severity and also as a mediator of IPF disease progression by modulation of specific metabolic pathways. While no causal relationship can be determined from this observational study, further research may better elucidate the role of ccf-dsDNA and metabolic derangements both as possible biomarkers and as targets for therapeutics in IPF.

Author affiliations
1Department of Medicine, Pulmonary and Critical Care, Weill Cornell Medicine, New York, New York, USA
2Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Cornell Medicine, New York, New York, USA
3Department of Internal Medicine, Division of Pulmonary and Critical Care Medicine, University of Michigan, Ann Arbor, Michigan, USA
4Department of Integrated Biomedical Science, Soonchunhyang Institute of Medi-bio Science (SIMS), Soonchunhyang University, Asan, Chungcheongnam-do, The Republic of Korea
5Section of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut, USA
6Pulmonary Critical Care, Yale University School of Medicine, New Haven, Connecticut, USA

Twitter Mustafa Buyukozkan @mu2bu

Contributors Conception and design: WW, HS-D and SJC; analysis and interpretation: WW, MB, JK, FJM, AMKC, HS-D and SJC; and drafting the manuscript for important intellectual content: WW, MB, JK, FJM, BM, AMKC, HS-D and SJC.

Funding HWS-D was supported by grants R01AG052530 and R01AG056699 and SJC was supported by grant K08HL138285.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Subjects included in this analysis were a subset of patients who had participated in a prospective observational study correlating biomarkers with disease progression (clinicaltrials.gov, clinical trials ID no. NCT01071707) (Correlating Outcomes with biochemical Markers to Estimate Time progression in Idiopathic Pulmonary Fibrosis—COMET). Written informed consent was obtained from the individual(s) for that was included in this article.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD
Soo Jung Cho http://orcid.org/0000-0003-0759-7295

REFERENCES
Supplemental Methods

Study population

Subjects included in this analysis were a subset of patients who had participated in a prospective observational study correlating biomarkers with disease progression (clinicaltrials.gov, clinical trials ID no. NCT01071707) (Correlating Outcomes with biochemical Markers to Estimate Time-progression in Idiopathic Pulmonary Fibrosis–COMET)(4). Our cohort consisted of 98 patients who had samples available for analysis over at least 3 follow-up time points; this report focuses only on the baseline samples. Inclusion criteria required patients to be aged 35-80 years with a diagnosis of IPF. Exclusion criteria included IPF diagnosis greater than 4 years prior to screening, concomitant collagen-vascular disorder, FEV1 (forced expiratory volume in 1 second)/FVC (forced volume capacity) < 0.6, evidence of active infection at screening, and comorbidities other than IPF likely to result in death within one year. Subject follow-up was for 80 weeks. Rapid progressors were defined by a combined endpoint - free survival defined as time to death, acute exacerbation of idiopathic pulmonary fibrosis(4), lung transplant, or fall in FVC of 10% or greater or diffusion capacity of the lung (DLCO) of 15% or greater. Informed consent had been obtained from all participating patients. The study protocol was reviewed and approved by the institutional review board of each participating center and methods were carried out in accordance with the relevant guidelines and regulations of each institution. Participating centers were University of California Los Angeles (Los Angeles, CA), University of California, San Francisco (San Francisco, CA), National Jewish Medical and Research Center (Denver, CO), University of Chicago (Chicago, IL), University of Michigan (Ann Arbor, MI), Cleveland Clinic Foundation (Cleveland, OH), Temple University (Philadelphia, PA), Brown University (Providence, RI), and Vanderbilt University (Nashville, TN). Patients were enrolled from March 2010 to March 2011. Blood samples and demographic data acquired from healthy human controls (n = 28) were from Yale University (New Haven, CT). IPF diagnosis had been made using a multidisciplinary approach as per published international guidelines(1).
Peripheral blood sampling

Samples were collected in EDTA-containing vacutainers and samples were shipped in cold packs by overnight mail from study centers to the University of Michigan. The samples were centrifuged at 2,500 rpm for 10 minutes and plasma was collected and stored at −80°C in aliquots. Study samples were then shipped to Weill Cornell Medical Center for analysis. Plasma samples from healthy human controls were obtained from Yale University for further analysis.

The ccf-dsDNA was extracted from plasma samples of IPF patients and controls according to manufacturer’s instructions using the Maxwell RSC ccf-DNA Plasma Kit (RSC; Promega, Leiden, the Netherlands). Another set of plasma samples from IPF patients were sent to Metabolon Inc. (Durham, NC, USA) for metabolomics profiling.

Metabolomics profiling
Untargeted metabolite profiling was performed by Metabolon Inc. (Durham, NC, USA) on plasma from IPF patients and controls using ultrahigh-performance liquid-phase chromatography and gas-chromatography separation coupled with tandem mass spectrometry.

**Sample Accessioning:** Following receipt, samples were inventoried and immediately stored at -80°C. Each sample received was accessioned into the Metabolon LIMS system and was assigned by the LIMS a unique identifier that was associated with the original source identifier only. This identifier was used to track all sample handling, tasks, results, etc. The samples (and all derived aliquots) were tracked by the LIMS system. All portions of any sample were automatically assigned their own unique identifiers by the LIMS when a new task was created; the relationship of these samples was also tracked. All samples were maintained at -80°C until processed.

**Sample Preparation:** Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Several recovery standards were added prior to the first step in the extraction process for QC purposes. To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phases (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

**QA/QC:** Several types of controls were analyzed in concert with the experimental samples: a pooled matrix sample generated by taking a small volume of each experimental sample (or
alternatively, use of a pool of well-characterized human plasma) served as a technical replicate throughout the data set; extracted water samples served as process blanks; and a cocktail of QC standards that were carefully chosen not to interfere with the measurement of endogenous compounds were spiked into every analyzed sample, allowed instrument performance monitoring and aided chromatographic alignment. Tables 1 and 2 describe these QC samples and standards. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the pooled matrix samples. Experimental samples were randomized across the platform run with QC samples spaced evenly among the injections.

Data preprocessing and statistical analysis

Probabilistic Quotient normalization was used to correct for varying sample intensities. Metabolites having more than %25 missing values were dropped. Data was subsequently log2-transformed. Missing values were imputed using a k-nearest-neighbor algorithm. The resulting p-values were corrected for multiple hypothesis testing using Benjamini-Hochberg method. The ccf-dsDNA levels are mean ± SEM and analysis of variance (ANOVA) (with post hoc comparisons using Dunnett’s test), using a statistical software package (GraphPad Prism version 9.0, GraphPad Software Inc. (San Diego, CA, USA)) for comparison of multiple groups.

Acknowledgement

We acknowledge members of the Yale School of Medicine and its Center for Pulmonary Infection Research and Treatment (CPIRT) for providing human samples.