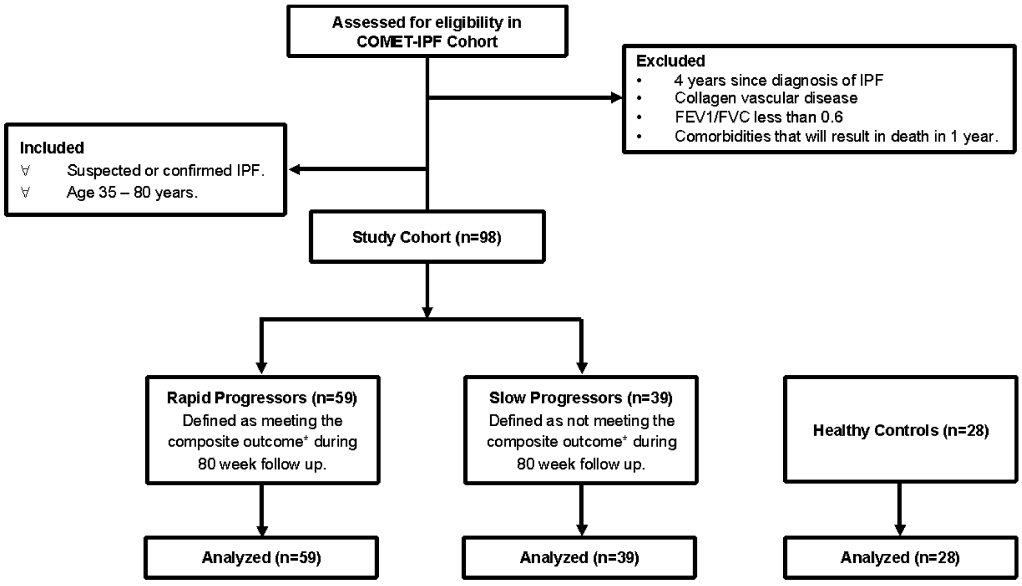
Supplemental Methods

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).



* Rapid progressors defined as composite outcome (death, acute exacerbation of IPF, lung transplant relative decline in FVC of at least 10% or DLCO of 15%) during the 80-week follow-up

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.