**Methodology**

The NIH sponsored multicenter Genetic Epidemiology of COPD (COPDGene ([ClinicalTrials.gov](http://clinicaltrials.gov/) Identifier: NCT01969344) study was approved and reviewed by the institutional review board at all participating centers (1). All study participants provided written informed consent. This study enrolled 10,198 non-Hispanic white (NHW) and African American (AA) individuals from January 2008 until April 2011 (Phase 1) who were aged 45-80 with ≥10 pack-year smoking history and no exacerbations for >30 days. In addition, 465 age and gender matched healthy individuals with no history of smoking were enrolled as controls (mostly at Phase 2). From July 2013 to July 2017, 5,697 subjects returned for an in-person 5-year visit. Each in-person visit included spirometry before and after albuterol, quantitative CT imaging of the chest, and blood sampling. From two clinical centers (National Jewish Health and University of Iowa) 162 subjects at Phase 1 (all NHW) and 1,136 subjects (1,040 NHW, 96 AA) participated in an ancillary study in which they provided fresh frozen plasma collected using an 8.5 ml p100 tube (Becton Dickinson) at Phase 2.

P100 plasma was profiled using the Metabolon (Durham, USA) Global Metabolomics Platform, as described (2-4). Briefly, samples were extracted with methanol under vigorous shaking for two minutes (Glen Mills GenoGrinder 2000) followed by centrifugation to remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase/ultrahigh performance liquid chromatography/tandem mass spectrometry (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 hydrophilic interaction chromatography (HILIC)/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup.

Metabolon has developed peak detection and integration software to generate a list of (mass-to-charge) *m/z* ratios, retention indices (RI), and area under the curve (AUC) values for each detected metabolite, as described in detail (2-4) . User specified criteria for peak detection included thresholds for signal to noise ratio, area and width. Relative standard deviations (RSDs) of peak area were determined for internal and recovery standards to confirm extraction efficiency, instrument performance, column integrity, chromatography and mass calibration. The biological data sets, including QC samples, were chromatographically aligned based on a retention index that utilized internal standards assigned a fixed RI value. The RI of the experimental peak was determined by assuming a linear fit between flanking RI markers whose RI values are set. Peaks were matched against an in-house library of authentic standards and routinely detected unknown compounds specific to the respective method. Identifications were based on retention index values, experimental precursor mass match to the library authentic standard within 10 ppm, and quality of MS/MS match. All proposed identifications were then manually reviewed and curated by an analyst who approved or rejected each identification based on the criteria above. The platform reported 1,392 features, including 1,064 annotated features which were grouped by Metabolon into “super pathways” including: 436 lipids, 261 xenobiotics, 207 amino acids, 40 peptides, 38 cofactors and enzymes, 35 nucleotides, 25 carbohydrates, 11 energy pathway compounds, and 11 partially characterized molecules (**Table S1**). All compounds are further annotated by “sub pathway” (e.g. “sphingomyelins”, “carnitine metabolism”, “lysine metabolism”).

1. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic epidemiology of COPD (COPDGene) study design. COPD. 2010;7(1):32-43.

2. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. Anal Chem. 2009;81(16):6656-67.

3. Dehaven CD, Evans AM, Dai H, Lawton KA. Organization of GC/MS and LC/MS metabolomics data into chemical libraries. J Cheminform. 2010;2(1):9.

4. Miller MJ, Kennedy AD, Eckhart AD, Burrage LC, Wulff JE, Miller LA, et al. Untargeted metabolomic analysis for the clinical screening of inborn errors of metabolism. J Inherit Metab Dis. 2015;38(6):1029-39.