Unique and shared systemic biomarkers for emphysema in Alpha-1 Antitrypsin deficiency and chronic obstructive pulmonary disease

K.A. Serban,[a,b]* K.A. Pratte,[a] C. Strange,[c] R.A. Sandhaus,[a] A.M. Turner,[f] T. Beiko,[c] D.A. Spittle,[g] L. Maier,[a,b] N. Hamzeh,[h] E.K. Silverman,[d,e] B.D. Hobbs,[d,e] C.P. Hersh,[d,e] D.L. DeMeo,[d,e] M.H. Cho,[d,e] and R.P. Bowler[a,b]*

[a]Department of Medicine, Division of Pulmonary, Critical Care, and Sleep Medicine, National Jewish Health, Denver, CO, United States
[b]Department of Medicine, Division of Pulmonary Sciences and Critical Care Medicine, University of Colorado, Aurora, CO, United States
[c]Pulmonary, Critical Care, Allergy and Sleep Medicine, Medical University of South Carolina, Charleston, SC, United States
[d]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States
[e]Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States
[f]Institute for Applied Health Research, University of Birmingham, Birmingham, UK
[g]Institute of Inflammation and Aging, University of Birmingham, UK
[h]Pulmonary, Critical Care, Allergy and Sleep Medicine, University of Iowa, Iowa City, IA, United States

eBioMedicine 2022;84: 104262
Published online xxx
https://doi.org/10.1016/j.ebiom.2022.104262

Summary

Background Alpha-1 Antitrypsin (AAT) deficiency (AATD), the most common genetic cause of emphysema presents with unexplained phenotypic heterogeneity in affected subjects. Our objectives to identify unique and shared AATD plasma biomarkers with chronic obstructive pulmonary disease (COPD) may explain AATD phenotypic heterogeneity.

Methods The plasma or serum of 5,924 subjects from four AATD and COPD cohorts were analyzed on SomaScan V4.0 platform. Using multivariable linear regression, inverse variance random-effects meta-analysis, and Least Absolute Shrinkage and Selection Operator (LASSO) regression we tested the association between 4,720 individual proteins or combined in a protein score with emphysema measured by 15th percentile lung density (PD15) or diffusion capacity (DLCO) in distinct AATD genotypes (Pi*ZZ, Pi*SZ, Pi*MZ) and non-AATD, PiMM COPD subjects. AAT SOMAmer accuracy for identifying AATD was tested using receiver operating characteristic curve analysis.

Findings In PiZZ AATD subjects, 2 unique proteins were associated with PD15 and 98 proteins with DLCO. Of those, 68 were also associated with DLCO in COPD also and enriched for three cellular component pathways: insulin-like growth factor, lipid droplet, and myosin complex. PiMZ AATD subjects shared similar proteins associated with DLCO as COPD subjects. Our emphysema protein score included 262 SOMAmers and predicted emphysema in AATD and COPD subjects. SOMAmer AAT level <7.99 relative fluorescence unit (RFU) had 100% sensitivity and specificity for identifying Pi*ZZ, but it was lower for other AATD genotypes.

Interpretation Using SomaScan, we identified unique and shared plasma biomarkers between AATD and COPD subjects and generated a protein score that strongly associates with emphysema in COPD and AATD. Furthermore, we discovered unique biomarkers associated with DLCO and emphysema in PiZZ AATD.

Funding This work was supported by a grant from the Alpha-1 Foundation to RPB. COPDGene was supported by Award U01 HL089897 and U01 HL089856 from the National Heart, Lung, and Blood Institute. Proteomics for COPDGene was supported by NIH 1R01HL137995. GRADS was supported by Award U01HL112707, U01

Abbreviations: AATD, Alpha-1 Antitrypsin Deficiency; COPD, Chronic Obstructive Pulmonary Disease; PD15, 15th percentile lung density; DLCO, diffusion capacity; GRADS, The Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis; QUANTUM-1, QUANTitative Chest Computed Tomography UnMasking Emphysema Progression in Alpha-1 Antitrypsin Deficiency; COPDGene, Genetic Epidemiology of COPD

*Corresponding authors.
E-mail addresses: serbank@njhealth.org (K.A. Serban), bowlerr@njhealth.org (R.P. Bowler).

One-sentence summary: Systemic biomarkers predict emphysema in AATD and COPD.

HL112695 from the National Heart, Lung, and Blood Institute, and UL1TRR002535 to CCTSI; QUANTUM-1 was supported by the National Heart Lung and Blood Institute, the Office of Rare Diseases through the Rare Lung Disease Clinical Research Network (1 U54 RR019498-01, Trapnell PI), and the Alpha-1 Foundation. COPDGene is also supported by the COPD Foundation through contributions made to an Industry Advisory Board that has included AstraZeneca, Bayer Pharmaceuticals, Boehringer-Ingelheim, Genentech, GlaxoSmithKline, Novartis, Pfizer, and Sunovion.

Keywords: SomaScan; Protein score; Emphysema; Plasma biomarker; Alpha-1 antitrypsin deficiency

Research in context

Evidence before this study

Alpha-1 Antitrypsin deficiency (AATD) is typically diagnosed by Alpha-1 Antitrypsin (AAT) protein measurements and genetic sequencing in individuals presenting with emphsyema on chest computer tomography (CT). AATD carriers (PiMZ), may develop emphysema if secondary risk factors, like cigarette smoking are present, however most individuals with deficient genotypes, PiZZ or PiSZ with mild emphysema remain undiagnosed or have delayed diagnosis due to near normal spirometry or diffusion capacity (DLCO). While CT is diagnostic of emphysema it is associated with radiation exposure which limits ist use as a biomarker. Hence, there is a critical need to develop non-invasive and accessible biomarkers, e.g. plasma biomarkers, that detect and predict emphysema in AATD at risk individuals. Furthermore, there is little known regarding the plasma proteome in AATD emphsyema and whether there is overlap with non-AATD, cigarette smoke-induced chronic obstructive pulmonary disease (COPD)/emphysema.

Added value of this study

Our report comprehensively evaluates the plasma proteome using SomaScan platform in three large AATD cohorts and one non-AATD, COPD cohort. We identified two proteins, betacellulin (BTC) and Cyclin Dependent Kinase-2-associated protein-1 (CDK2AP1) which are uniquely associated with the emphysema in PiZZ individuals. We also created an emphysema severity score using 262 SomaScan proteins. The score predicts emphysema similarly in AATD and COPD individuals. While the protein score performed better in AATD and COPD individuals with less severe emphysema, the protein score was not superior to DLCO at predicting emphysema in AATD inviduals with advanced emphysema or PiZZ individuals treated with augmentation therapy. Lastly, we report that SomaScan is both sensitive and specific for AATD individuals and carriers and would be suitable for identifying previously undiagnosed subjects.

Implications of all the available evidence

Our study supports the use of proteomic platforms for detection of early emphysema diagnosis and severity stratification and also for screening populations for severe- and intermediate-deficient AAT genotypes. Furthermore, SomaScan has excellent diagnostic characteristics for severe- and intermediate-deficient AAT genotypes. Further work is needed to determine whether these individual biomarkers and/or the protein score are useful to assess progression of emphysema and airflow obstruction as well as other comorbidities, such as exacerbations in AATD individuals.

Introduction

Alpha-1 Antitrypsin deficiency (AATD), a genetic disease that accounts for 1-2% of chronic obstructive pulmonary disease (COPD),[1,2] is caused by mutations in *Serpin Family A Member 1* (*SERPINA1*) gene. The normal *SERPINA1* allele is Pi*M, with the most common disease variants referred to as Pi*Z and Pi*S alleles, characterized by low AAT serum levels. The mean allele frequencies in Europe are 0-3% (mean=1.5%) for Pi*Z and 1-13% for Pi*S (mean=3.3%).[3] Mainly the Pi*ZZ and Pi*SZ severe-deficient genotypes account for severe AATD, that presents with clinical significant emphysema and requires Alpha-1 Antitrypsin (AAT) augmentation therapy to prevent emphysema progression.[1] Although the genotype predicts the emphysema risk, AAT levels do not correlate with disease onset or progression within the genotypes, especially in the intermediate-deficient, PiMZ and PiMS subjects.[4,5] Thus, additional AATD modifiers explain disease heterogeneity and biomarkers are needed to stratify AATD individual at risk for emphysema development.

Current clinical measurements of emphysema, including spirometry [forced expiratory volume at 1 second (FEV_1)], diffusion capacity (DLCO), and computer tomography (CT) imaging [visual emphysema score and the 15th percentile lung density (PD15)] typically identify subjects with severe disease. Within the same genotype,

these clinical measurements vary widely,[6,7] are not specific to AATD, and only abnormal when AATD is clinically advanced. Even emerging systemic biomarkers for COPD, such as C reactive protein (CRP), fibrinogen, interleukin-6 (IL-6) and interleukin-8 (IL-8), and soluble receptor for advanced glycation end products (sRAGE)[8–10] may not be shared with AATD-associated emphysema, since AATD subjects were frequently excluded from COPD biomarker studies. Dedicated studies that investigated plasma biomarkers in AATD subjects, using Myriad Discovery MAP and Bio-Rad/Bio-Plex 200 platforms, included small cohorts, and lacked replication or PiMM COPD groups.[11,12]

In the last decade we witnessed rapid advance in unbiased high-throughput proteomic approaches that measure thousands of proteins in a single assay. For example, SomaScan, an aptamer-affinity-based technology, allows for multiplexed measurements of >5,000 proteins.[13,14] Since multiple publications have shown that SomaScan is reproducible[14] and that many SomaScan biomarkers are highly correlated with traditional antibody biomarker-based assays,[15] this platform is increasingly used in clinical biomarker discovery.

In this study we measured SomaScan profiles in four independent cohorts including a large non-AATD population [COPDGene][16] and three AATD populations [The Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS),[17] QUANTitative Chest Computed Tomography UnMasking Emphysema Progression in Alpha-1 Antitrypsin Deficiency (QUANTUM-1),[11] and Birmingham Alpha-1 Antitrypsin registry (Birmingham)]. We hypothesized that our unbiased proteomic approach would identify new plasma AATD biomarkers for emphysema that correlate with clinical outcomes.

Methods

Study populations

The study cohorts included 5,924 subjects enrolled in the COPDGene[16] (N= 5,607), the Alpha-1 Antitrypsin-Deficiency and Sarcoidosis[17] (GRADS) (N=133), the QUANTitative Chest Computed Tomography UnMasking Emphysema Progression in Alpha-1 Antitrypsin Deficiency[11] (QUANTUM-1) studies (N=38), and in the Birmingham Alpha-1 Antitrypsin registry (N=146), Figure 1. All subjects had AAT genotyping, at least one spirometry measurement and one plasma or serum SomaScan proteomic assay. The respective local Institutional Review Boards (IRB) approved all study protocols and informed consent was obtained from all participants [COPDGene: HS 1883 – National Jewish Health; GRADS: Pro00024143 – Medical University of South Carolina; QUANTUM-1: HR17301 - Medical University of South Carolina; Birmingham Alpha-1 Antitrypsin registry: 18/SC/0541 – South Central Oxford C Research Ethics Committee].

Participant baseline characteristics are presented in Table 1.

COPDGene is a multi-center, longitudinal cohort funded by the National Heart, Lung, and Blood Institute (NHLBI) which enrolled >10,000 non-Hispanic White and African Americans adults with a smoking history of >10 pack-years and either with and without COPD.[16] The cohort specifically recruited for genome-wide association studies, but other significant clinical, functional, laboratory, and radiological data were collected at 21 centers across the USA.

GRADS is a multi-center, cross-sectional cohort of adults older than age 35 years with PiZZ or PiMZ Alpha-1 Antitrypsin genotypes. It was designed to conduct state-of-art genomic, microbiomics and phenotypic studies to better understand AATD.

QUANTUM-1 is multicenter, longitudinal cohort funded by the National Institutes of Health Office of Rare Diseases and NHLBI to study radiographic emphysema progression in adults with PiZZ AATD and normal lung function. The cohort recruited 51 AATD patients with normal Forced Expiratory Volume in the first second (FEV_1) ≥ 80% predicted to determine whether baseline CT measurements of emphysema, e.g. lung density, predicted a more rapid decline in FEV_1.

The Birmingham Alpha-1 Antitrypsin registry enrolls patients with all AATD phenotypes that undergo clincially-indicated pulmonary function, chest CT, and laboratory tests in Birmingham, England.

Clinical data and definitions

COPD was defined using spirometric evidence of airflow obstruction: i.e., post-bronchodilator FEV_1/ forced vital capacity (FVC) < 0.70. FEV_1 and FVC maneuvers were recorded per ATS/ERS standards for spirometry in COPDGene, GRADS, and QUANTUM-1. The severity of COPD was based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines using post-bronchodilator FEV_1% as follows: GOLD 1 (≥ 80%); GOLD 2 (≥ 50% and <80%); GOLD 3 (≥ 30% and <50%); or GOLD 4 (<30%). Subjects with FEV_1/FVC ≥ 0.70 and FEV_1% < 80% predicted were defined as having Preserved Ratio Impaired Spirometry (PRISm).[18] Subjects with FEV_1/FVC ≥0.70 were defined as controls (GOLD 0). DLCO percent predicted was based on the GLI (traditional units/min/mmHg) and adjusted for altitude, age, height, and sex.

Emphysema was reported as the 15th percentile adjusted lung density (PD15, g/L) measured as the Hounsfield units (HU) below which the 15% of voxels with the lowest density are distributed, adjusted for the race-corrected total lung capacity and multiplied by 1000 to be expressed as PD15, as previously described.[19] In addition, percent emphysema was defined as the percentage of lung voxels ≤ -950 HU (% low attenuation areas, % LAA) on the full inspiratory scans.

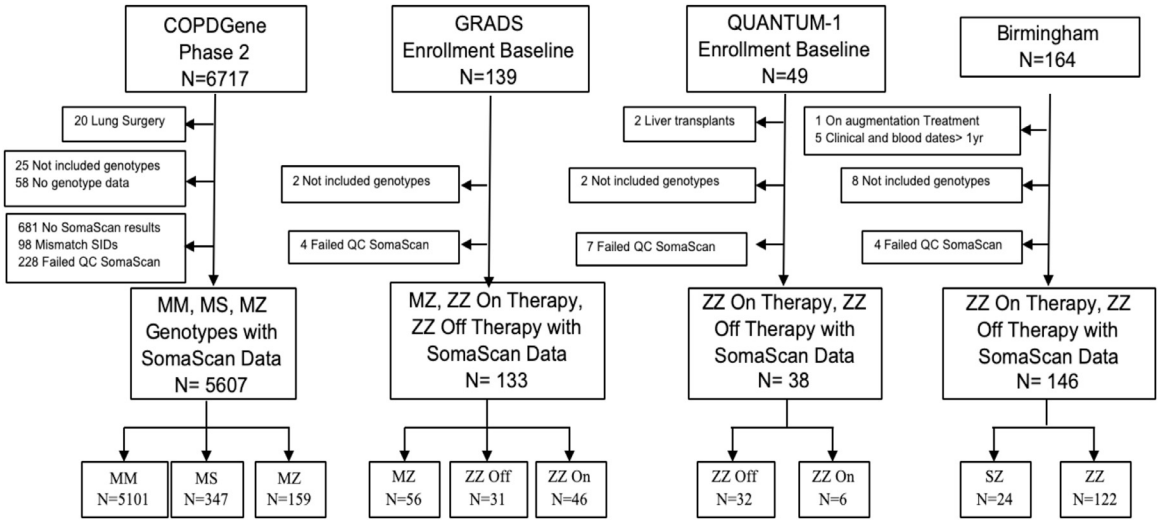


Figure 1. Study consort. Diagram representation of the individuals recruited in the COPDGene, GRADS, QUANTUM, and Birmingham cohorts but excluded in the current study due to absent SomaScan data and exclusion genotypes. Genotypes not included: COPDGene – Pi*FZ, Pi*SZ, Pi*SS; QUANTUM-1 - Pi*$M_{heerlen}$Z and Pi*SZ; GRADS- discordant genotype / phenotype measurements; Birmingham - Pi*M_{Malton}Null, Pi*$M_{procida}$Z, Pi*SS, Pi*ZNull, Pi*FZ, Pi*IZ, Pi*M_{malton}Z, Pi*SP_{Lowell}.

Visual emphysema severity recorded in the electronic medical record was used in the Birmingham cohort. For the purpose of this manuscript visual emphysema was graded as present or absent.

Alpha-1 Antitrypsin testing

- COPDGene subjects were genotyped for *SERPINA1* Z (rs28929474) and S (rs17580) alleles using 5′ to 3′ exonuclease assays (TaqMan assay, Applied Biosystems, Foster City, CA) and known Pi*ZZ, Pi*MZ, and Pi*MS (Coriell Institute for Medical Research, Camden, NJ) control samples. Pi Protein Phenotyping: Isoelectric focusing was performed on plasma samples from subjects with the Pi*ZZ, Pi*SZ, and Pi*MZ genotypes as described.[20] Discordance between the genotyping and protein phenotyping was resolved by medication review for AAT augmentation therapy as well as by repeat genotyping and protein phenotyping.
- GRADS subjects used clinical tests positive for Pi*MZ or Pi*ZZ genotypes.[17] AAT augmentation therapy was documented during the medication reconciliation at the time of study enrollment. Two subjects were excluded because had discordant genotype – phenotype measurements and 4 subjects were dropped from the analysis because the plasma sample did not meet the quality control criteria on the SomaScan (Figure 1)
- QUANTUM-1 subjects had the Pi*ZZ or Pi*Znull genotype confirmed by gene probe analysis. Previous AAT genotype results were acceptable if documented from a Clinical Laboratory Improvement Amendments (CLIA) certified laboratory. Subjects were included if their serum AAT was less than 11uM or 80mg/dL. {https://www.clinicaltrials.gov/ct2/show/NCT00532805}. Two subjects were excluded because they were liver transplant recipients, 2 subjects had exclusion genotypes Pi*SZ and Pi*$M_{Heerlen}$Z, and 7 subjects were dropped from the analysis because the plasma sample did not meet the quality control criteria on the SomaScan (Figure 1).
- Birmingham subjects with clinical testing using a combination of isoelectric focusing and sequencing were included if they were Pi*ZZ or Pi*SZ phenotype, independent of the AAT serum levels. In the Birmingham SomaScan group 1 subject was not considered because was receiving augmentation therapy, 5 subjects were excluded because of > 1year difference between plasma samples and spirometry or CT measurements, 8 subjects had exclusion genotypes Pi*M_{Malton}Null, Pi*$M_{procida}$Z, Pi*SS, Pi*ZNull, Pi*FZ, Pi*IZ, Pi*M_{malton}Z, Pi*SP_{Lowell}, and 4 subjects were dropped from the analysis because the plasma sample did not meet the quality control criteria on the SomaScan. (Figure 1).

Proteomics platform

SomaScan v4.0 uses 4,979 different SOMAmers (aptamers) to quantify 4,776 unique human proteins with 4,720 unique Uniprot numbers.[15] SomaScan signal normalization included plate hybridization to control for variability across array signals, median signal normalization to control for technical variability of

Table 1. Total Population characteristics.	COPDGene			GRADS			QUANTUM-1		Birmingham	
	MM N = 5101	MS N=347	MZ N = 159	MZ N = 56	ZZ N = 31	N = 46	ZZ N = 32	N = 6	SZ N=24	ZZ N = 122
Augmentation Therapy	No	No	No	No	No	Yes	No	Yes	No	No
AAT level										
SOMAscan RFU (Natural Log)	10.63 (0.23)	10.37 (0.21)	9.94 (0.23)	10.08 (0.26)	6.63 (0.22)	10.66(0.55)	6.7 (0.24)	10.9 (0.36)	9.10 (0.16)	6.13 (0.26)
AAT µmol/L							13.0 (3.25)	10.4 (2.0)	14.35 (2.78)	4.56 (1.60)
Demographics										
Age in years mean (SD)	65.0 (8.8)	66.2 (9.0)	67.7 (8.6)	53.8 (10.9)	52.3 (11.2)	60.5 (9.4)	51.4 (8.3)	56.5 (10.7)	52.7 (12.4)	53.1 (12.2)
Male N (%)	2565 (50.3%)	165 (47.6%)	85 (53.5%)	17 (30.4%)	14 (45.1%)	21 (45.7%)	9 (28.1%)	4 (66.7%)	6 (25.0%)	58 (47.5%)
AA N (%)	1548 (30.3%)	37 (10.7%)	11 (6.9%)	0%	0%	0%	0%	0%	0 (%)	0 (%)
NHW N (%)	3553 (69.7%)	310 (89.3%)	148 (93.1%)	56 (100%)	30 (96.8%)	46 (100%)	32 (100%)	6 (100%)	24 (100%)	121 (99.2%)
BMI (kg/m^2)	28.9 (6.3)	28.6 (6.3)	29.4 (6.2)	28.2 (6.0)	26.4 (3.9)	27.8 (5.4)	29.5 (7.8)	25.5 (3.4)	26.5 (5.2)	26.9 (5.3)
Never Smokers N (%)	335 (6.6%)	24 (6.9%)	10 (6.3%)	30 (53.6%)	25 (80.7%)	19 (41.3%)	20 (62.5%)	1 (16.7%)	15 (62.5%)	50 (41.0%)
Current smoking status N (%)	1876 (36.8%)	100 (28.8%)	36 (22.6%)	3 (5.4%)	1 (3.2%)	1 (2.2%)	0%	0%	2 (8.3%)	3 (2.5%)
ATS Pack-years median (IQR)	38.5 (29.0)	39.0 (27.0)	40.0 (32.0)	0.00 (2.5)	0.00 (0.00)	2.1 (18.5)	0.0 (1.8)	6.0 (8.5)	0.00 (17.5)	4.5 (15.0)
Spirometry										
COPD GOLD PRISm	603 (11.8%)	37 (10.7%)	13 (8.2%)	7 (12.7%)	2 (6.5%)	2 (4.4%)	1 (3.1%)	0%	0 (0%)	2 (1.6%)
GOLD 0	2335 (45.8%)	152 (43.8%)	67 (42.1%)	35 (62.5%)	17 (54.8%)	4 (8.7%)	24 (75.0%)	5 (83.3%)	15 (62.5%)	28 (23.0%)
GOLD 1	473 (9.3%)	31 (8.9%)	9 (5.7%)	6 (10.7%)	3 (9.7%)	2 (4.4%)	7 (21.9%)	1 (16.7%)	3 (12.5%)	14 (11.5%)
GOLD 2	951 (18.6%)	63 (18.2%)	29 (18.2%)	6 (10.7%)	6 (19.4%)	20 (43.5%)	0%	0%	2 (8.3%)	33 (27.1%)
GOLD 3	487 (9.6%)	41 (11.8%)	26 (16.4%)	0%	2 (6.5%)	16 (34.8%)	0%	0%	2 (8.3%)	30 (24.6%)
GOLD 4	195 (3.8%)	19 (5.2%)	12 (7.6%)	1 (1.8%)	0%	2 (4.4%)	0%	0%	2 (8.3%)	14 (11.5%)
FEV_1 Percent Predicted mean (SD)	79.7 (24.6)	77.9 (26.4)	73.1 (27.8)	91.3 (20.6)	87.9 (21.5)	59.7 (21.0)	99.3 (11.8)	97.5 (9.0)	87.5 (32.7)	66.5 (30.2)
FEV_1 post BD (Liter) mean (SD)	2.2 (0.9)	2.2 (0.9)	2.1 (1.0)	2.9 (0.9)	3.0 (1.0)	1.9 (0.8)	3.2 (0.8)	3.3 (0.8)	2.6 (1.1)	2.2 (1.1)
FEV_1/FVC post BD mean (SD)	0.68 (0.14)	0.66 (0.16)	0.63 (0.18)	0.74 (0.13)	0.70 (0.16)	0.50 (0.15)	0.77 (0.10)	0.70 (0.10)	0.70 (0.19)	0.53 (0.20)
DLCO percent predicted mean (SD)	78.6 (23.0)	80.7 (24.7)	79.6 (24.2)	99.5 (25.3)	90.3 (23.6)	67.3 (25.5)	96.0 (17.6)	71.0 (8.1)	103.5 (24.5)	74.4 (23.8)
CT Emphysema										
% emphysema (-950 Hu), total lung, median (IQR)	1.5 (4.7)	1.9 (5.5)	3.7 (9.2)	1.3 (3.1)	5.3 (10.0)	17.4 (19.4)	9.5 (8.1)	16.0 (16.1)	NA	NA
Adjusted lung density (g/l) mean (SD)	86.2 (25.0)	84.2 (26.9)	76.3 (26.3)	77.5 (20.2)	68.1 (17.2)	51.6 (19.9)	68.9 (20.8)	44.5 (14.2)	NA	NA
Visual (Yes/No) N (%)									4 (16.7%)	67 (54.9%)
Charlson index	3.4 (1.9)	3.6 (2.0)	3.6 (1.6)	NA	NA	NA	1.6 (1.2)	3.0 (2.2)	1.8 (1.6)	2.1 (1.3)

***Table 1*: Patient characteristics. Data presented as mean ± SD or median ± IQR.**

Abbreviations: COPDGene = Genetic Epidemiology of COPD; GRADS = Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis; QUANTUM = QUANTitative Chest Computed Tomography UnMasking Emphysema Progression in Alpha-1 Antitrypsin Deficiency; AAT = Alpha-1 Antitrypsin; FEV_1 = Forced expiratory volume in 1 second; FEV_1/FVC = Forced expiratory volume in 1 second to forced vital capacity; Dlco = Diffusing capacity for carbon monoxide, post BD = post bronchodilator; SD = standard deviation; IQR = interquartile range; GOLD = Global initiative for chronic Obstructive Lung Disease; PRISm = Preserved Ratio Impaired SpiroMetry; Hu = Hounsfield units.

replicates within a run, and plate scaling and calibration of SOMAmers to control for inter-assay variation between analytes and batch differences between plates. Finally, median normalization to a reference using adaptive normalization by maximum likelihood was applied within the dilution group to quality control replicates and individual samples to remove edge effects and technical variance. Orthogonal data supporting aptamer specificity for target proteins has been previously published.[15]

Statistical analysis

Cross-sectional analysis of proteins – COPD phenotype associations. Natural log transformed SomaScan proteins, pulmonary function [FEV_1, FEV_1 percent predicted (FEV_1pp), FEV_1 /FVC, DLCO], and emphysema measurements [visual emphysema and PD15] were treated as continuous variables. Multivariable linear regression was used to identify proteins significantly associated with pulmonary function and emphysema. The regression model for FEV_1 was adjusted for age, age^2, height, height^2, sex, BMI, pack-years, current smoking and never smoked status. The FEV_1pp model included pack-years, and current smoking or never smoking status. The FEV_1/FVC model included age, sex, BMI, pack-years, current smoking and never smoking status. The DLCO model was adjusted for BMI, pack-years, and current smoking and never smoking status. In the PD15 model we controlled for age, sex, BMI, pack-years, current smoking and never smoking status. In the COPDGene cohort analyses, we further adjusted for race and included clinical center as a random effect. Because the ZZ cohorts were predominantly never smokers with very few current smokers, we only included never smoking status in their analysis. QUANTUM-1 ZZ on therapy with a small sample size of 6 was not included in the analysis. Results across cohorts were combined by genotype and treatment into an inverse variance random-effects meta-analysis. STROBE guidelines for cohort studies reporting were followed.

The PiZZ group on augmentation therapy (N=46, GRADS) was compared with a 3:1 matched PiMM group (N=138, COPDGene). Matching was done using SAS surveyselct procedure; for never smoker status, sex, and age category.

Predictive modeling of PD15. We used supervised regularization methods to select SomaScan proteins and to derive an emphysema predictive score using PD15. Two methods, Least Absolute Shrinkage and Selection Operator (LASSO) and elastic net were evaluated for the best fit on a training dataset. We found no significant difference between models, therefore, we used LASSO because it provided a more parsimonious model, by retaining only one of the collinear variables. In the COPDGene PiMM dataset 4745 observations had complete data for SomaScan and adjusted lung density (PD15). This dataset was randomly split 70/30% to create the training (n=3324) and testing (n=1421) datasets. Using the training dataset, a 10-fold cross-validation on standardized variables was used to estimate the model. Using the test dataset's mean squared errors (MSE), R^2, and Pearson correlations (between observed and predicted adjusted lung density), we evaluated whether to use the LASSO model based on the minimum mean cross-validated error lambda or at 1-standard error from the minimum. We used the lambda based on 1-standard error because it provided a more parsimonious model with minimal effect on model fit. Predicted values for adjusted lung density were calculated for COPDGene PiMZ; GRADS PiMZ, GRADS PiZZ not on and -on therapy; QUANTUM-1 PiZZ not on therapy; and Birmingham PiZZ. MSE, R-squared (R^2), and Pearson correlations between observed and predicted values were calculated for each group, with the exception of the Birmingham group. Birmingham PiZZ had visual emphysema (Yes/No) data only, and logistic regression was used to calculate the AUC for the predicted adjusted lung density association with visual emphysema. To evaluate the emphysema score against another emphysema biomarker, e.g. DLCO we calculated the R^2, Pearson correlation between DLCO and measured PD15.

Pathway enrichment analysis

Using Gene Ontology enRIchment anaLysis and visuaLizAtion tool (GOrilla), we conducted a pathway enrichment analysis on N=98 and N=1,306 SomaScan proteins significantly associated with DLCO in the PiZZ not on therapy and the PiMM groups, respectively. A hypergeometric test was performed to determine significant enrichment of cellular component GO terms. A color-coded trimmed directed acyclic graph (DAG) and a list of all significantly enriched GO terms was generated.[21]

SomaScan diagnostic accuracy. Receiver operating characteristic curve (ROC) analysis, plots and area under the curves (AUC) were estimated for aptamer sequence ID 3580-25 (AAT) using logistic regression in PiZZ not on therapy, PiSZ, PiMZ, PiMS, and PiMM individuals. We excluded PiZZ on therapy because they had AAT levels approaching PiMM AAT serum levels. The Youden J-index was used to select the optimal predicted cut-point for AAT. Due to complete separation between the tested genotypes the cut-point was determined by the midpoint of the separation range. STARD guidelines for diagnostic studies reporting were followed.

Regression analyses, ROC, histograms, and matching were performed with SAS 9.4 (SAS/STAT 15.1). Meta-analysis and forest plots (metafor V3.0-2); LASSO (glmnet V4.1-2); beeswarm, scatter and volcano plots (ggplot2 V3.3.5) were generated with R (V4.1.0). A false discovery rate adjusting for 4,979 SOMAmers (FDR $\leq$ 0.05) was considered significant.[22]

Role of funders
The study design, data collection, data analysis, interpretation, and writing of report are solely the responsibility of the authors and do not necessarily represent the official views of the industry, the Foundations, National Heart, Lung, and Blood Institute or the National Institutes of Health.

Results
There was significant demographic heterogeneity of cohorts (Figure 1, Tables 1 and S1), with AATD cohorts being younger and including more never-smokers than COPDGene ($p<0.0001$, ANOVA 1-way). In COPDGene we identified 11 and 37 African American subjects with intermediate-deficient, Pi*MZ and Pi*MS, respectively; other cohorts recruited predominantly non-Hispanic whites. The GRADS and Birmingham cohorts recruited more subjects with COPD GOLD stages 3 and 4 (Table 1). The PiZZ subjects on AAT therapy in GRADS and QUANTUM-1 had more emphysema compared to those off therapy ($p<0.001$, ANOVA 1-way), while the PiZZ subjects in general had more emphysema and lower DLCO than their PiMM counterparts (Table 1). A third of PiMZ subjects in COPDGene were active smokers, with a corresponding higher number of pack-years, higher emphysema ($p<0.006$, Kruskal-Wallis), and lower DLCO ($p<0.001$, Kruskal-Wallis) than the PiMZ in GRADS, suggesting, as expected, that cigarette smoking is an additive risk factor for disease severity (Table 1). PiMS subjects in COPDGene shared similar demographics, tobacco exposure, and functional characteristics with PiMM individuals. Interestingly, PiSZ subjects in Birmingham cohort, despite significantly lower AAT levels than PiMM and PiMZ subjects, presented with nearly normal FEV_1, DLCO, and GOLD severity at similar age as PiZZ subjects. Lastly, COPDGene subjects had a higher Charlson comorbidity index.

Biomarkers associated with airflow obstruction and DLCO in AATD patients off augmentation therapy
We identified 177, 169, and 216 proteins significantly associated with FEV_1, FEV_1pp, and FEV_1/FVC ratio, respectively (nominal $p \leq 0.05$, but none with FDR $\leq$ 0.05, multivariable linear regression, Table S2). There were 671 proteins associated with DLCO, 98 of which were either positively or negatively associated (FDR $\leq$ 0.05, multivariable linear regression) with DLCO (Figure 2a). Of those positively associated with DLCO, the top SOMAmers were: Cerebellin-4 (CBLN4), Immunoglobulin superfamily DCC subclass member-4 (IGDC4), Hemojuvelin (HFE2), and Insulin-like growth factor-1 (IGF-1). Of those negatively associated with DLCO, the top SOMAmers were: Macrophage scavenger receptor type-1 (MSR1), Transgelin (TAGL), Growth/differentiation factor-15 (GDF15), Macrophage mannose receptor-1 (MRC1), Gremlin-2 (GREM2), CUB domain-containing protein-1 (CDCP1), and Fatty acid-binding protein, adipocyte (FABP4) (Table S2). Of 98 proteins significantly associated with DLCO in PiZZ subjects, 68 were shared proteins associated with DLCO in PiMM subjects as well (Figure 2b and S2), but importantly, 30 proteins were uniquely associated with DLCO only in PiZZ off therapy subjects (Figure 2b). The top 5 unique proteins were: keratin 1, myomesin-2, insulin receptor, insulin-like growth factor, and hydroxymethylbilane synthase; they are depicted in the table inserts (Figure 2b).

We then considered the 671 and 1305 nominally-significant proteins associated with DLCO in PiZZ off therapy and PiMM individuals, respectively for pathway analysis. Using GO*rilla* we found three significantly enriched cellular components pathways in PiZZ off therapy: insulin-like growth factor complex, lipid droplet, and myosin complex (Figure 3a, Table S3), as expected, considering the top nominal proteins we identified above are growth factors (e.g. IGF-1, GDF15), lipid transporters or receptors (e.g. FABP4, MRC1), or proteins derived from skeletal muscle (e.g. myomesin-2, hemojuvelin). These cellular component pathways did not overlap with those enriched in PiMM, of which the top 3 were: extracellular matrix, extracellular organelle / vesicles, and intracellular endoplasmic reticulum lumen (Figure 3b, Table S4).

Biomarkers associated with airflow obstruction, DLCO, and emphysema in AATD subjects on augmentation therapy
We analyzed the AATD subjects on augmentation therapy separately from those off therapy because those on therapy have already established severe emphysema, even though they now have normal levels of AAT achieved through weekly augmentation therapy infusion. AATD subjects on therapy (N=46) were matched on age, sex, smoking history, and GOLD severity with 138 PiMM subjects (COPDGene) 1:3 (Table S5). In the PiZZ on therapy group we identified 373, 255, 111, 100, and 137 proteins nominally associated with PD15, DLCO, FEV1pp, FEV_1/FVC, and FEV_1 respectively (Table S6). Two proteins, endothelin-2 and sterol carrier protein-2 sterol-binding domain-containing protein-1 (SCP2D1) were significantly associated with FEV_1 ($p = 0.0003$ and $p = 0.0004$, FDR = 0.08, multivariable linear regression). In the matched PiMM COPDGene subgroup we identified 265, 517, 227, 341, and 220 proteins significantly associated with PD15, DLCO, FEV1pp, FEV_1/FVC, and FEV_1 respectively, but only 21 proteins associated with DLCO (FDR $\leq$ 0.05, multivariable linear regression, Table S7). The top-50 ranked proteins associated with DLCO for PiZZ subjects on therapy were rather different than similar to the PiMM subgroup, suggesting that PiZZ on therapy plasma

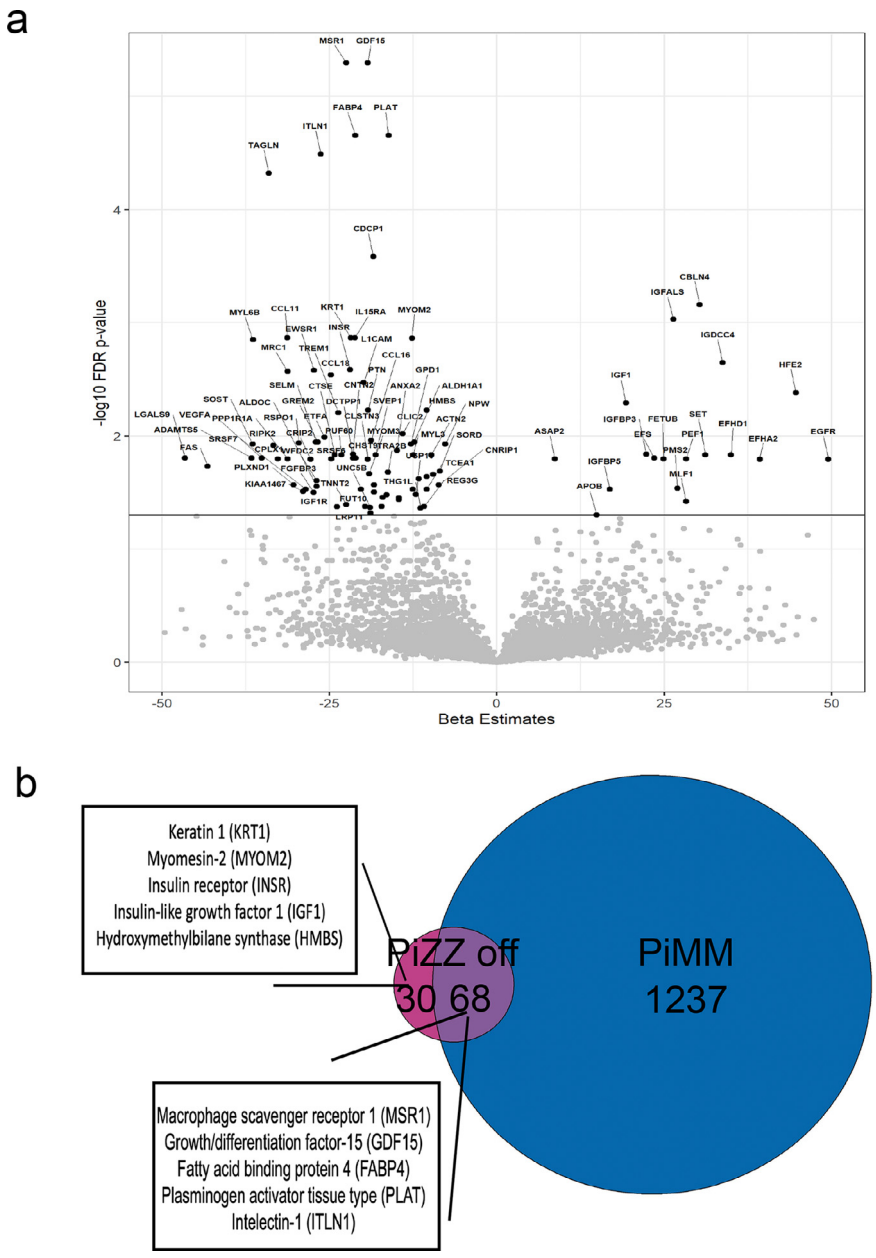


Figure 2. SomaScan proteins association with DLCO in PiZZ subjects off augmentation therapy. a. Volcano plot of SomaScan proteins positively and negatively associated with DLCO (N=98) for PiZZ subjects (N=185) off augmentation therapy from the Birmingham, GRADS, QUANTUM cohorts. SomaScan proteins, labeled with their gene abbreviations, are shown in black if significantly (FDR $\leq$ 0.05, multivariable linear regression) associated with DLCO; non-significant (FDR $>$0.05, multivariable linear regression) SomaScan proteins are shown in grey. **b.** Euler plot of SomaScan proteins associated with DLCO that are specific (magenta circle, N=30) and shared (purple circle, N=68) between PiZZ off therapy and PiMM (blue circle, N=1237) subjects. The top five unique proteins keratin 1 (KRT1), myomesin-2 (MYOM2), insulin receptor (INSR), insulin-like growth factor (IGF-1), and hydroxymethylbilane synthase (HMBS) and shared proteins macrophage scavenger receptor types I and II (MSR1), growth/differentiation factor 15 (GDF15), fatty acid-binding protein, adipocyte (FABP4), tissue-type plasminogen activator (PLAT), intelectin-1 (ITLN1) between PiZZ off therapy and PiMM subjects are shown in the insert tables.

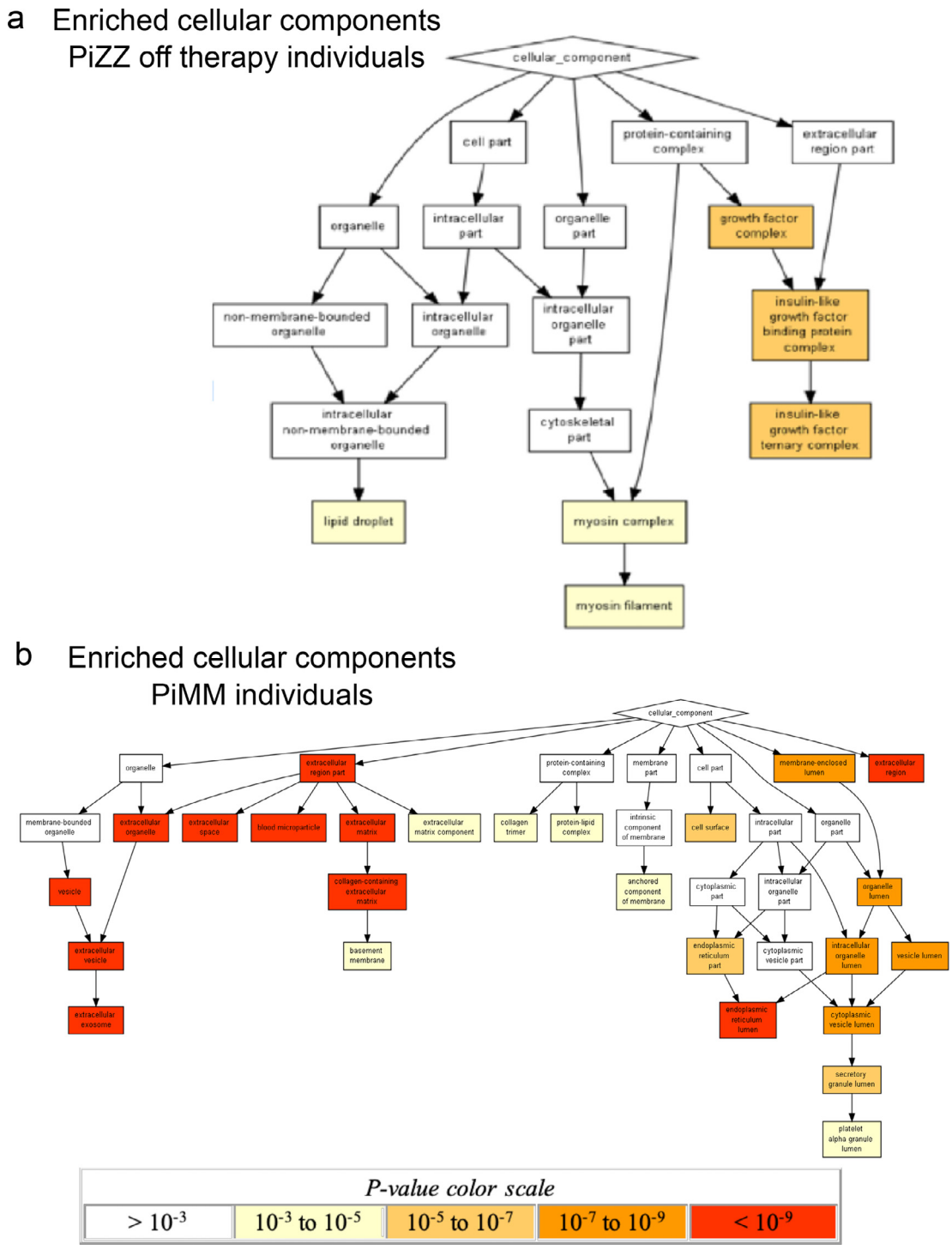


Figure 3. Cellular component pathways enriched in the SomaScan proteins associated with DLCO in PiZZ off augmentation therapy and PiM subjects. a. Cellular component pathways enriched within the 671 nominally-significant SomaScan proteins associated with DLCO in PiZZ subjects off therapy. **b**. Cellular component pathways enriched within the 1305 nominally-significant SomaScan proteins associated with DLCO in PiMM subjects. Individual cellular component pathways are color-coded based on the significance of enrichment p-values depicted at the bottom, light yellow $p<0.00005$, light orange $p<0.000005$, dark orange $p<0.00000005$, red $p<0.000000005$ (Pathway enrichment analysis, GOrilla). The graphical representation of Directed acyclic graph (DAG) was created using GO*rilla*.

proteomic signature is different from the COPD signature, even after accounting for demographics and disease severity (Figure S2). The 2 proteins associated with FEV_1 in PiZZ on therapy subgroup and the 21 proteins associated with DLCO in PiMM subgroup were not among the top-50 proteins of the other's subgroup, highlighting the differences in plasma proteomics between PiZZ and COPD individuals, even after augmentation therapy was instituted (Tables S6 and S7).

Biomarkers associated with airflow obstruction, DLCO, and emphysema in PiMZ AATD subjects

Using individuals in two cohorts, COPDGene and GRADS, we identified 10 proteins that were significantly associated with FEV_1; one, CRP, was significant (FDR $\leq$ 0.05, multivariable linear regression, Figure 4a). There were 142, 164, and 145 proteins nominally associated with FEV_1pp, FEV_1/FVC, and PD15 ($p \leq 0.05$, but none at FDR $\leq$ 0.05, multivariable linear regression, Table S8). There were 280 proteins associated with DLCO; 10 proteins were significant (FDR $\leq$ 0.05, multivariable linear regression, Figure 4b). We had one SomaScan protein positively associated with DLCO, Cyclic AMP-dependent transcription factor ATF-6 alpha (ATF6), while the other 9 proteins were negatively associated: Retinoic acid receptor responder protein-2 (RARRES2), Chordin-like protein-1 (CHRDL-1), R-spondin-1 (RSPO-1), Fibroblast growth factor-binding protein-1 (FGFBP1), Pleiotrophin (PTN), Spondin-1 (SPO-1), C-C motif chemokine-16 (CCL16), and Collagen alpha-1(XXVIII) chain (COL28A1). Most of these associations were also significant in PiMM subjects (Figures S3–S4, Table S9).

Emphysema protein score to boost explanation of measured emphysema variance and assess biomarker similarity between AATD and COPD

Using GRADS and QUANTUM-1 cohorts (PiZZ AATD subjects) that had PD15 measured, we identified two proteins, betacellulin (BTC) and Cyclin Dependent Kinase-2-associated protein-1 (CDK2AP1) positively associated with PD15 at (FDR $\leq$ 0.05, multivariable linear regression, Figure 5a-b).

There were more FDR and nominal significant proteins associated with markers of emphysema (DLCO, PD15) compared to markers of airflow obstruction (FEV_1, FEV_1/FVC) in the AATD cohorts (Figure 2a, Figure 5a-b, Table S2). Although, in PiMM subjects there were many more proteins (N=665) nominally significant associated with PD15, in both PiZZ and PiMM the individual proteins had small effect size. Therefore, we investigated whether a protein score performs better than individual SomaScan proteins at explaining the clinical variance of measured PD15, used as an emphysema marker.

We used LASSO, the regression analysis that allows feature selection and regularization to develop a protein score for emphysema (PD15). The protein score was trained and tested on COPDGene PiMM subjects because of its larger sample size. The best risk score had 262 proteins combined in an emphysema (PD15) protein score showing strong correlations with measured emphysema in PiMM subjects from the testing cohort (R^2=0.44, rho=0.66, LASSO, Figure 5, Table 2). The calculated emphysema protein score was also accurate in PiZZ and PiMZ subjects from AATD cohorts (rho from 0.47 to 0.70, LASSO, Figure 5b-d, Table 2). The lower rho values were associated with GRADS PiMZ and PiZZ -subjects on AAT therapy, which had the mildest and the most severe emphysema, respectively (Figure 5d). The correlations between emphysema protein score and measured emphysema were comparable with DLCO correlations with measured emphysema (R^2 from 0.20 to 0.76, Pearson correlation, Table 2), with the highest correlation seen in PiMZ COPDGene individuals with mild-moderate emphysema. Quantitative CT measurements were not available in the Birmingham cohort; however, the emphysema protein score was associated with measured visual emphysema (AUC=0.74, logistic regression).

Although we used three AATD independent cohorts to identify several unique protein - PD15 associations, there were insufficient AATD subjects in any one cohort to develop a unique AATD protein risk score.

SomaScan identifies AATD subjects

Severe-deficient PiZZ subjects are easily identified based on a low AAT serum level. However, many AATD subjects intermediate-deficient, i.e. PiSZ, PiMS, and PiMZ remain unidentified, despite being at increased risk for COPD,[5,23,24] unless AAT genotyping is performed. We found that the SomaScan could readily identify PiZZ subjects not on AAT therapy (Figure 6a-b). The receiver operating characteristic (ROC) curve for diagnosis of PiZZ versus PiSZ, PiMM, PiMS, and PiMZ, as well as PiSZ versus PIMZ, and PIMM subjects were perfect (AUC 1.00, logistic regression, Figure 6c). PiSZ, PiMS, and PiMZ subjects were also well distinguished (AUCs 0.8-0.9, logistic regression, Figure 6b-c). Thus, a cut-off value <7.99 (RFU, natural log) had a 100% sensitivity/specificity for the diagnosis of PiZZ and values above 10.52 were likely PiMM subjects.

Discussion

Since AATD is only a risk factor for emphysema,[6] not all subjects have clinically significant disease. Although environmental exposures are important, proteomic disease-modifiers may be able to explain part of the emphysema heterogeneity in AATD and COPD.[9,11] Our study has comprehensively tested the AATD proteome

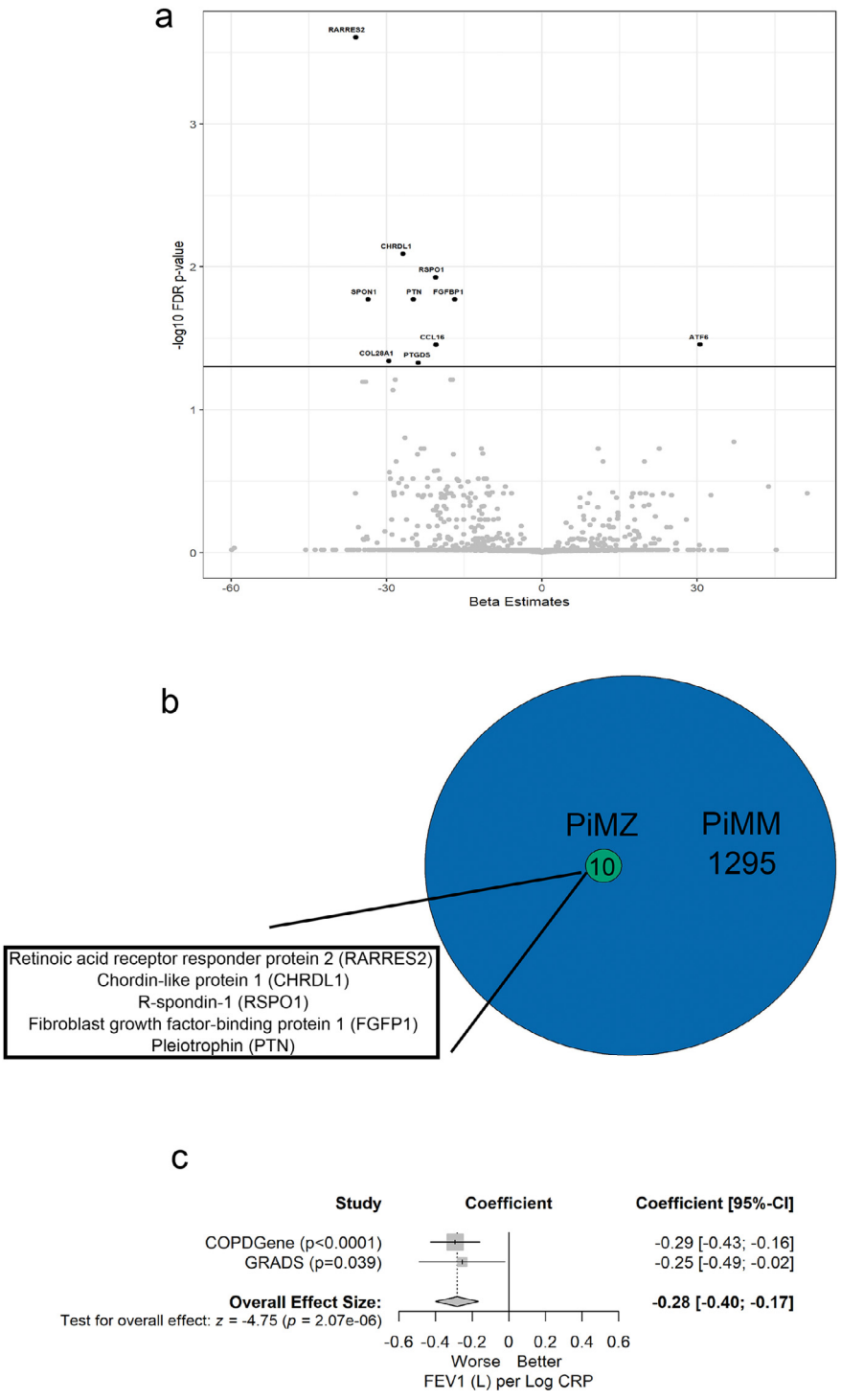


Figure 4. Somascan proteins association with DLCO and spirometry in PiMZ subjects. a. Volcano plot of SomaScan proteins positively and negatively associated with DLCO (N=10) in PiMZ subjects (N=215) from two cohorts, COPDGene and GRADS studies. SomaScan proteins, labeled with their gene abbreviations, are shown in black if significantly (FDR $\leq$ 0.05, multivariable linear regression) associated with DLCO; non-significant (FDR $>$ 0.05, multivariable linear regression) SomaScan proteins are shown in grey. **b**. Euler plot of SomaScan proteins associated with DLCO that are shared (green circle, N=10) between PiMZ and PiMM (blue circle, N=1295) subjects. The top five shared proteins between PiMZ and PiMM subjects are Retinoic acid receptor responder protein 2 (RARRES2), Chordin-like protein 1 (CHRDL1), R-spondin-1 (RSPO1), Fibroblast growth factor-binding protein 1 (FGFP1), and pleiotrophin (PTN). **c**. Forest plot of natural log CRP (RFU, natural log) negative association with FEV_1 in PiMZ subjects (N=215) from two cohorts, COPDGene and GRADS studies.

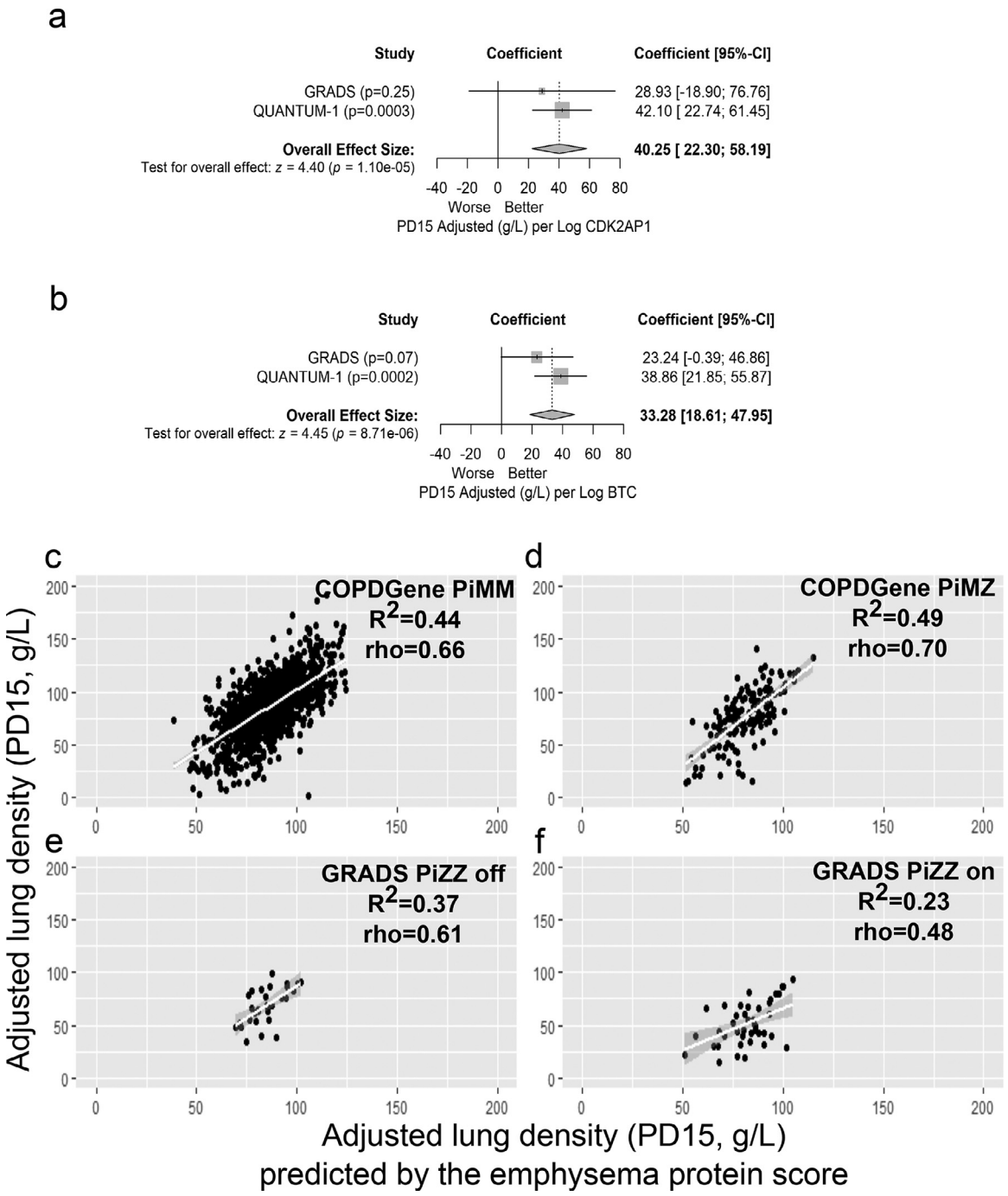


Figure 5. Somascan individual proteins and emphysema protein score association with adjusted lung density. a-b. Forest plots of cyclin-dependent kinase 2-associated protein 1 (CDK2AP1) and betacellulin (RFU, natural log) positive association with adjusted lung density (PD15, g/L) in PiZZ individuals (N=63) off augmentation therapy from two, GRADS and QUANTUM study cohorts. **c-f**. Scatter plots showing the association between measured adjusted lung density (PD15, y axis) vs. PD15 predicted by the emphysema protein score (x axis) in **c**) COPDGene PiMM (R^2=0.44, rho=0.66, LASSO); **d**) COPDGene PiMZ (R^2=0.49, rho=0.70, LASSO); **e**) GRADS PiZZ off therapy (R^2=0.37, rho=0.61, LASSO); and **f**) PiZZ on therapy individuals (R^2=0.23, rho=0.48, LASSO). Emphysema protein score (N=262 proteins) was developed using LASSO in the COPDGene PiMM subjects and the other genotypes were used for validation.

Table 2. Measured adjusted Lung Density (PD15) vs. predicted PD15 by LASSO or vs. measured DLCO

Model	Protein Score (by LASSO)						DLCO		
	Training Sample Size	Testing Sample Size	MSE Testing	R^2	Rho_P	Visual Emphysema Yes/No AUC	R^2	Rho_P	Visual Emphysema Yes/No AUC
COPDGene PiMM	3324	1421	360.9	0.44	0.66	NA	0.15	0.39	NA
Validation									
COPDGene PiMZ		146	409.2	0.49	0.70	NA	0.36	0.60	NA
GRADS PiMZ		54	456.9	0.22	0.47	NA	0.06	0.24	NA
GRADS PiZZ off therapy		30	452.1	0.37	0.61	NA	0.33	0.56	NA
GRADS PiZZ on therapy		44	1234.2	0.23	0.48	NA	0.57	0.76	NA
QUANTUM-1 PiZZ off therapy		30	809.9	0.24	0.49	NA	0.04	0.20	NA
Birmingham		100	NA	NA	NA	0.74	NA	NA	0.78

***Table 2*: Measured adjusted lung density (PD15) vs. predicted PD15 by LASSO emphysema protein score or vs. measured DLCO in the testing cohort, COPDGene (PiMM) and the validation cohorts: COPDGene (PiMZ), GRADS (PiMZ), GRADS (PiZZ off therapy), GRADS (PiZZ on therapy), QUANTUM (PiZZ off therapy). In the Birmingham cohort we report measured visual emphysema vs. predicted emphysema by LASSO protein score. R^2 and Pearson correlation (Rho) were calculated in R. AUC was calculated using logistic regression in R.**
Abbreviations: MSE: mean square error, AUC: area under the curve.

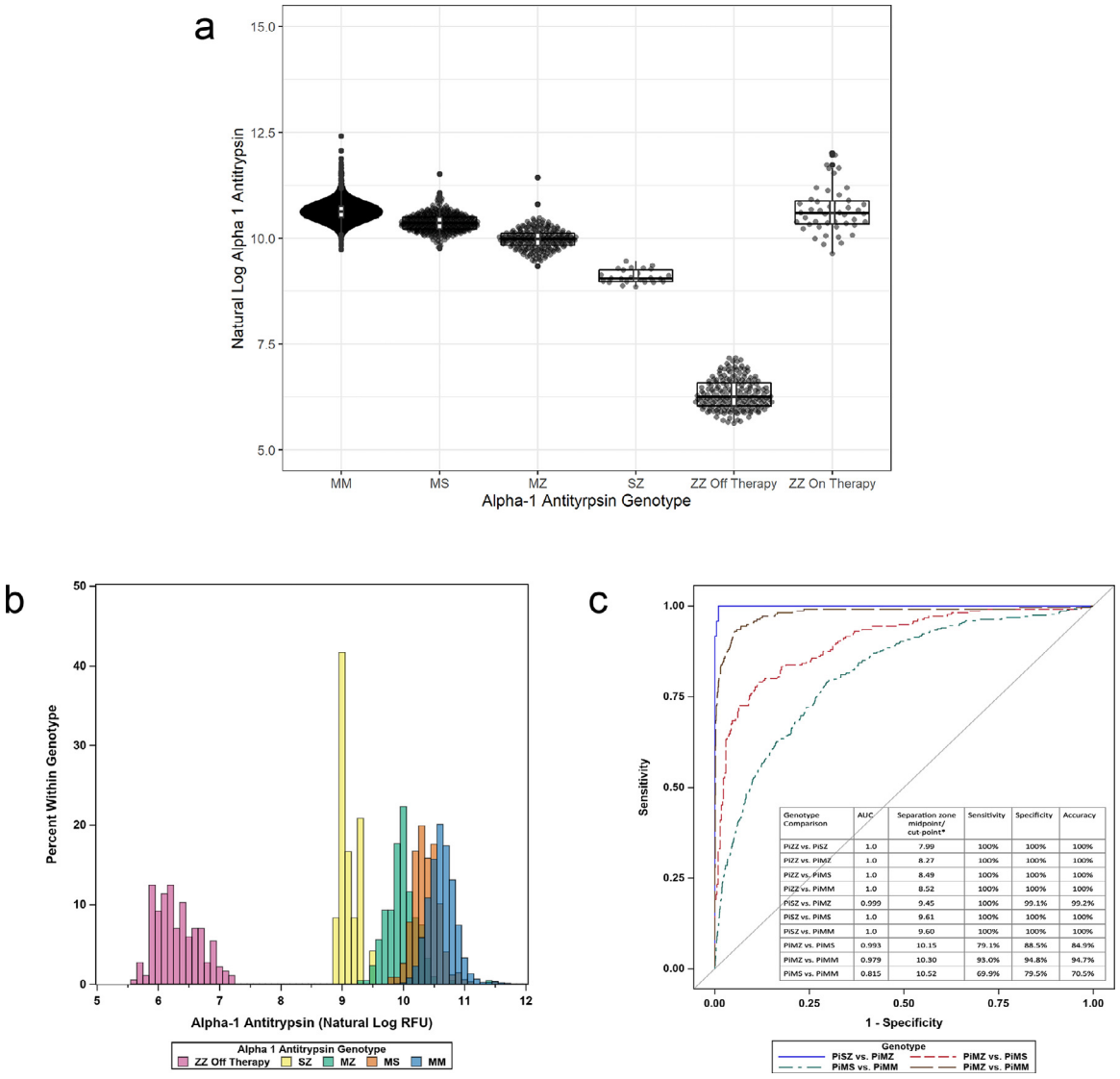
association with markers of emphysema in PiZZ and PiMZ subjects enrolled in multiple independent cohorts, including a large PiMM, COPD reference population. For both PiMM and PiZZ subjects there were many different proteins associated with functional and radiologic markers of emphysema.

Like other recent publications in non-pulmonary cohorts, we found that multiple proteins in combination explained a much higher percentage of the variance of a clinical phenotype compared to individual proteins.[9,25] Interestingly, patients at risk for cardio-vascular events, early death in congestive heart failure, and kidney failure in diabetic subjects were easily identified by SomaScan protein scores.[26–28] In fact, a combination of 262 proteins in our emphysema protein score explained more of measured PD15 variance than DLCO in COPD, PiMZ, and PiZZ subjects off augmentation therapy, outperforming DLCO as a known functional marker of emphysema. The emphysema protein score explained less of measured PD15 variance in PiZZ on augmentation therapy, possibly because the protein score was developed in the COPD patients, where the SomaScan proteins associated with PD15 were rather different than similar to those in PiZZ on therapy.

We identify CDK2AP1 as unique biomarker associated with PD15 in PiZZ off augmentation therapy subjects. CDK2AP1 is the only known inhibitor of cyclin-dependent kinase-2 and a master regulator of cell division cycle. CDK2AP1 is down-regulated in various malignancies[29–31] and upregulated during embryonic development.[32] Its positive association with PD15 suggests tighter control of cell cycle check-points in PiZZ subjects off augmentation therapy and with early emphysema. It remains to be determined in prospective AATD cohorts whether CDK2AP1 is a signal of efficient repair, because we only detect CDK2AP1 association with PD15 in PiZZ subjects with early emphysema and off therapy, and not in PiZZ on therapy or in PiMM with advanced emphysema.

Betacellulin (BTC), a second biomarker specific to PiZZ subjects off therapy, is a ligand of the epidermal growth factor (EGF) superfamily. Similar to transforming growth factor-α, heparin-binding EGF-like growth factor and amphiregulin, EGF mediates BTC downstream signaling and results in airway epithelial reprogramming or epithelial to mesenchymal transition (EMT), mucus hypersecretion and airway obliteration, and possible malignant transformation of large and small airways.[33,34] BTC ranked 14th in a support vector machine classifier used to diagnose and endotype COPD individuals[35] and it was higher in ex-smokers with COPD than without COPD.[36] Our study reports that BTC is significantly associated with emphysema in AATD. The positive association between BTC and PD15 suggests that, in AATD patients off augmentation therapy, emphysema is characterized by an active repair process involving epithelial airway remodeling. More targeted work needs to be done in AATD pre-clinical models to determine BTC's role in EMT, mucus hypersecretion, and remodeling in AATD emphysema.

Our study confirms previous nominal protein - emphysema associations (e.g. CRP, FABP4) reported in 31 AATD patients enrolled in QUANTUM-1 cohort and measured on Myriad discovery panel.[11] Other associations were not confirmed, like leptin, gesolin, and metalloproteinase-3, proteins that were associated with emphysema at baseline and with its progression when measured on the Myriad panel.[11] We do describe the fatty acid binding protein (FABP4) association with emphysema seen also in QUANTUM-1 cohort,[11] but our meta-analysis showed stronger association between FABP4 and lung function (FEV_1 and DLCO). FABP-4 association with DLCO was found in all individual PiZZ cohorts and in the meta-analysis. Our findings suggest



Genotype Comparison	AUC	Separation zone midpoint/ cut-point*	Sensitivity	Specificity	Accuracy
PiZZ vs. PiSZ	1.0	7.99	100%	100%	100%
PiZZ vs. PiMZ	1.0	8.27	100%	100%	100%
PiZZ vs. PiMS	1.0	8.49	100%	100%	100%
PiZZ vs. PiMM	1.0	8.52	100%	100%	100%
PiSZ vs. PiMZ	0.999	9.45	100%	99.1%	99.2%
PiSZ vs. PiMS	1.0	9.61	100%	100%	100%
PiSZ vs. PiMM	1.0	9.60	100%	100%	100%
PiMZ vs. PiMS	0.993	10.15	79.1%	88.5%	84.9%
PiMZ vs. PiMM	0.979	10.30	93.0%	94.8%	94.7%
PiMS vs. PiMM	0.815	10.52	69.9%	79.5%	70.5%

Figure 6. SomaScan Alpha-1 Antitrypsin relative levels in various Alpha-1 genotypes. a. Beeswarm plot of AAT relative fluorescence units (RFU, natural log, y axis) by genotype (x axis). AAT relative levels in pooled PiMM (10.63 ±0.23, N=5101), PiMS (10.37 ± 0.21, N=347), PiMZ (9.98 ± 0.25, N=215), PiSZ (9.10 ± 0.16, N=24), PiZZ off therapy (6.32 ± 0.36, N=185), and PiZZ on AAT augmentation therapy (10.69 ± 0.54, N=52) subjects from all 4 study cohorts. **b**. Histogram showing the percent distribution of PiZZ (N = 185), PiSZ (N=24), PiMZ (N=215), PiMS (N=347), and PiMM (N=5101) subjects from all 4 study cohorts. **c**. Receiving operator curves (ROC) of AAT (RFU, natural log) for the following genotype comparisons are shown: PiSZ vs. PiMZ, PiMZ vs. PiMS, PiMZ vs. PiMM, and PiMS vs. PiMM. The ROCs characteristics for the genotype comparisons not graphically depicted are shown in the table insert. *If AUC is equal to 1.0 it is a midpoint of the range of complete separation. If AUC less than 1.0 the value is the cut-point determined by the Youden Index.

that FABP-4 is a promising biomarker of AATD severity. Our study was not powered to investigate FABP-4 association with PD15 or DLCO in response to augmentation therapy, as we did not see an association with PD15 and DLCO in the PiZZ on augmentation therapy group.

Although we did find unique protein associations in PiZZ subjects, many of the biomarkers for PiZZ off AAT therapy and PiMZ were similar to PiMM subjects. For instance, CRP, retinoic acid receptor responder protein-2, and members of the spondin family identified in our PiMZ cohort are previously reported biomarkers of airflow obstruction and emphysema in PiMM subjects.[8] This suggests that plasma proteome of PiMZ is similar to PiMM subjects and might argue that PiMZ are similar enough to PiMM subjects to be included in general COPD clinical studies; however, often they are excluded.[23,24]

There were differences in our cohorts which may explain our unique biomarker - emphysema associations.

Most of the biomarkers we observed in AATD cohorts were for emphysema measurements (DLCO or PD15) whereas in COPDGene there were more biomarkers associated with airflow limitation (FEV_1 and FEV_1/FVC). This is likely because the three AATD cohorts included in this study had predominantly subjects with normal-to-low FEV_1, but evidence of emphysema on CT scan.

The strength of our data relies on representative number of never smokers PiZZ individuals form North America and Europe with various degrees of airflow limitation but significant emphysema who demonstrated unique and similar protein biomarkers to a large cigarette smoke-related COPD cohort. Overall the emphysema phenotype was better predicted by a common protein score. This score outperformed the more commonly used DLCO biomarker in all individual PiZZ and PiMZ cohorts. These findings strongly suggest that our unique plasma biomarkers and emphysema protein score could be useful in both AATD and COPD research. Indeed, the identification of emphysema in patients without airflow obstruction may be the most important aspect of disease prevention, because AATD subjects with early emphysema may be the most likely to progress and have the most at-risk healthy lung. The generalizability of our newly described plasma biomarkers and emphysema protein score can be further validated on more longitudinal cohorts to assess their ability to prospectively predict emphysema diagnosis and progression. Additionally, the protein score could address the lack a non-invasive, easy to collect instrument that is both diagnostic and prognostic of early emphysema. The instrument would be most useful in higher risk subjects (smokers or never-smokers with abnormal pulmonary function test or hypoxemia) or patients with abnormal Alpha-1 genotypes, but normal spirometry. Additionally the instrument, might play a useful role in therapeutic clinical trials to study emphysema progression in response to therapy.

This study also demonstrates a potential role for identifying clinically missed AATD subjects because many AATD subjects are under-diagnosed and may have unappreciated emphysema. Current AATD testing is a two-step procedure, initially AAT serum levels are measured by radial immunodiffusion or nephelometry with the low levels confirmed by a second test, like genotyping or protein electrophoresis.[37] Unfortunately, this strategy misses subjects in the general population with low clinical suspicion for AATD and subjects with at-risk, intermediate-deficient genotypes, Pi*MZ and Pi*SZ, who may develop emphysema.[38,39] We report that SomaScan identifies the most common clinically-significant AATD genotypes, the severe and intermediate-deficient Pi*ZZ, Pi*SZ, and Pi*MZ genotypes with excellent sensitivity and specificity. This is relevant because SomaScan is frequently used in population-based studies and could be useful for diagnosing AATD. The ability to detect less severe genotypes (Pi*MS) appears lower because they present with AAT levels similar to Pi*MM genotype. The SomaScan assay can also inform on the ability of augmentation therapy to raise AAT to near-physiologic levels as evidenced by AAT levels within PiMM range in PiZZ subjects on therapy. One disadvantage of SomaScan, i.e., that it only reports relative fluorescent, not absolute AAT "units", doesn't appear to be a limitation for identifying clinically-significant genotypes. We can't exclude an inclusion bias that resulted in high sensitivity and specificity in identifying clinically-significant genotypes, as the subjects included in PiZZ cohorts were not selected from the general population, nevertheless COPDGene subjects were. However, the characteristics of the assay suggest utility in identifying undiagnosised AATD subjects.

Limitations to the SomaScan proteomics include the lack of SOMAmers for small molecules such as desmosine,[40] fibrinogen degradation product ($A\alpha$-Val_{360}, a specific product generated by elastase cleavage of fibrinogen),[41] and sphingomyelin,[42] which have been suggested to be emphysema biomarkers in other studies,[43] While our study is the largest biomarker study in AATD subjects, none of the cohorts had enough subjects with very rare genotypes (Pi*IZ or $Pi*SP_{Lowell}$) to achieve adequate power. Finally, only the GRADS cohort had matching plasma – bronchoalveolar lavage samples, therefore our study has not investigated lung-specific protein biomarkers. There are also geographic differences such as the large number of subjects in US-based GRADS and QUANTUM-1 on augmentation therapy versus no patients on augmentation in the Birmingham cohort. GRADS subjects on therapy tended to have much worse emphysema, but we were able to identify two candidate proteins in endothelin-2 and SCP2D1 associated with FEV_1, suggesting that these plasma proteins might be used as biomarkers of disease progression or response to therapy rather than diagnostic biomarkers.

In summary, we demonstrate that the SomaScan proteomic platform helps risk stratify AATD carriers and subjects with early disease who might be at higher risk for emphysema. Also, the SomaScan emphysema protein score enhances our ability to predict emphysema. Furthermore, SomaScan has excellent diagnostic characteristics for severe- and intermediate-deficient AAT genotypes, which are the main clinically actionable AATD phenotypes. Further work is needed to determine whether these same biomarkers are useful to assess progression of emphysema and airflow obstruction as well as other COPD comorbidities such as exacerbations in AATD subjects.

Contributors

KAS - data curation, investigation, methodology, data interpretation, visualization, writing - original draft.
KAP - data curation, investigation, formal analysis, validation, methodology, visualization, writing -original draft.

CS - conceptualization, funding acquisition, investigation, methodology, data interpretation, project administration, supervision, visualization, writing - original draft.
RAS- conceptualization, funding acquisition, project administration.
AMT- conceptualization, funding acquisition, investigation, methodology, data interpretation, project administration, supervision, visualization, writing - original draft.
TB - investigation, visualization.
DAS - data curation, investigation, visualization.
LM - conceptualization, funding acquisition, data interpretation, project administration, visualization, writing - original draft.
NH - conceptualization, funding acquisition, project administration, visualization.
EKS- conceptualization, funding acquisition, investigation, methodology, data interpretation, project administration, supervision, visualization, writing - original draft.
BDH - investigation, methodology, data interpretation, visualization, writing - original draft.
CPH - data interpretation, visualization, writing - original draft.
DLD - data interpretation, visualization, writing - original draft.
MHC - investigation, methodology, data interpretation, visualization, writing - original draft.
RPB- conceptualization, funding acquisition, investigation, methodology, data interpretation, project administration, supervision, visualization, writing - original draft.
KAS, KAP, CS, AMT, DAS, EKS, CPH, DLD, and RPB have verified the underlying data.
All authors have reviewed and approved the final version of the manuscript.

Data sharing statement

Data underlying the results from COPDGene and Quantum cohorts are deposited and will be available per request in dbGaP database (phs000179.v6.p2 for COPDGene, phs000698.v1.p1 for QUANTUM). Data underlying the results from the Birmingham Alpha-1 Antitrypsin registry will be available per request, contact Dr. Alice M Turner.

Declaration of interests

KAS serves on Alpha-1 Foundation Grant Advisory Committee, Alpha-1 Foundation Medical Advisory and Scientific Committee, ATS - RCMB Website Committee, National Jewish Health IBC committee (all unpaid); **KAP** does not report any potential conflict of interest; **CS** reports grants or contracts from Adverum, Arrowhead, AstraZeneca, CSA Medical, Grifols, Nuvaira, Takeda, Vertex; consulting fees from AstraZenca, Dicerna, Glaxo Smith Kline, Inhibrx, Morair, UpToDate, Vertex; has received honoraria for presentations from the American Thoracic Society; has received support for travel from CSL Behring; serves as the Medical Director for AlphaNet; **RAS** reports grants or contracts from Vertex, NIH/NCATS, Alpha-1 Foundation, consulting fees from Grifols, CSL Behring, Vertex, Intellia, Inhibrx, Takeda, Evolve Biologics, has received travel support from CSL Behring; has served on the Advisory Board of Arrowhead Pharmaceuticals, and serves as the Medical Director for AlphaNet and on the board of directors for Global Implementation Solutions, Osteogenesis Imperfecta Foundation, Alpha-1 Foundation, and AlphaNet; **AMT** reports grants or contracts from Vertex, Grifols, and CSl Behring; has received consulting fees from CSL Behring, Inhibrix, Z-factor, and Takeda, has received honoraria for lectures from Glaxo Smith Kline and AstraZeneca; **TB** does not report any potential conflict of interest; DAS does not report any potential conflict of interest; **LM** does not report any potential conflict of interest; **NH** does not report any potential conflict of interest; **EKS** reports grants or contracts from Glaxo Smith Kline, Bayer; **BDH** does not report any potential conflict of interest; **CPH** reports grants or contracts from Alpha-1 Foundation, Bayer, Boehringer-Ingelheim, and Vertex, consulting fees from AstraZeneca and Takeda; **DLD** has received honoraria for lectures from Novartis and financial support from Bayer towards the institution; **MHC** reports grants or contracts from Glaxo Smith Kline, Bayer, consulting fees from AstraZeneca and Genentech, honoraria for presentations from Illumina, **RPB** does not report any potential conflict of interest.

Acknowledgements

SomaLogic, Inc. as the provider of the proteomic data measured using the modified aptamer-based SomaScan® Assay. SomaScan® and SOMAmer® are registered trademarks of SomaLogic, Inc. and are used under license. Funding: This work was supported by a grant from the Alpha-1 Foundation to RPB. COPDGene was supported by Award U01 HL089897 and U01 HL089856 from the National Heart, Lung, and Blood Institute. Proteomics for COPDGene was supported by NIH 1R01HL137995. GRADS was supported by Award U01HL112707, U01 HL112695 from the National Heart, Lung, and Blood Institute, and UL1TR002535 to CCTSI; QUANTUM-1 was supported by the National Heart Lung and Blood Institute, the Office of Rare Diseases through the Rare Lung Disease Clinical Research Network (1 U54 RR019498-01, Trapnell PI), and the Alpha-1 Foundation. COPDGene is also supported by the COPD Foundation through contributions made to an Industry Advisory Board that has included AstraZeneca, Bayer Pharmaceuticals, Boehringer-Ingelheim, Genentech, GlaxoSmithKline, Novartis, Pfizer, and Sunovion.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2022.104262.

References

1 Silverman EK, Sandhaus RA. Clinical practice. Alpha1-Antitrypsin deficiency. *N Engl J Med*. 2009;360(26):2749–2757.
2 Brode SK, Ling SC, Chapman KR. Alpha-1 Antitrypsin deficiency: a commonly overlooked cause of lung disease. *CMAJ*. 2012;184(12):1365–1371.
3 Blanco I, de Serres FJ, Fernandez-Bustillo E, Lara B, Miravitlles M. Estimated numbers and prevalence of Pi*S and Pi*Z alleles of alpha1-antitrypsin deficiency in European countries. *Eur Respir J*. 2006;27(1):77–84.
4 Franciosi AN, Hobbs BD, McElvaney OJ, et al. Clarifying the risk of lung disease in SZ Alpha-1 Antitrypsin deficiency. *Am J Respir Crit Care Med*. 2020;202(1):73–82.
5 Dahl M, Tybjaerg-Hansen A, Lange P, Vestbo J, Nordestgaard BG. Change in lung function and morbidity from chronic obstructive pulmonary disease in alpha1-antitrypsin MZ heterozygotes: a longitudinal study of the general population. *Ann Intern Med*. 2002;136(4):270–279.
6 Silverman EK, Province MA, Campbell EJ, Pierce JA, Rao DC. Family study of alpha 1-antitrypsin deficiency: effects of cigarette smoking, measured genotype, and their interaction on pulmonary function and biochemical traits. *Genet Epidemiol*. 1992;9:317–331.
7 Ghosh AJ, Hobbs BD, Moll M, et al. Alpha-1 Antitrypsin MZ heterozygosity is an endotype of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2022;205(3):313–323.
8 Serban KA, Pratte KA, Bowler RP. Protein biomarkers for COPD outcomes. *Chest*. 2021;159(6):2244–2253.
9 Zemans RL, Jacobson S, Keene J, et al. Multiple biomarkers predict disease severity, progression and mortality in COPD. *Respir Res*. 2017;18(1):117.
10 Pratte KA, Curtis JL, Kechris K, et al. Soluble receptor for advanced glycation end products (sRAGE) as a biomarker of COPD. *Respir Res*. 2021;22(1):127.
11 Beiko T, Janech MG, Alekseyenko AV, et al. Serum proteins associated with emphysema progression in severe Alpha-1 Antitrypsin deficiency. *Chronic Obstr Pulm Dis*. 2017;4(3):204–216.
12 Campos MA, Geraghty P, Holt G, et al. The biological effects of double-dose Alpha-1 Antitrypsin augmentation therapy. A pilot clinical trial. *Am J Respir Crit Care Med*. 2019;200(3):318–326.
13 Kim CH, Tworoger SS, Stampfer MJ, et al. Stability and reproducibility of proteomic profiles measured with an aptamer-based platform. *Sci Rep*. 2018;8(1):8382.
14 Candia J, Cheung F, Kotliarov Y, et al. Assessment of variability in the SOMAscan assay. *Sci Rep*. 2017;7(1):14248.
15 Raffield LM, Dang H, Pratte KA, et al. Comparison of proteomic assessment methods in multiple cohort studies. *Proteomics*. 2020;20(12):e1900278.
16 Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010;7(1):32–43.
17 Strange C, Senior RM, Sciurba F, et al. Rationale and design of the genomic research in Alpha-1 Antitrypsin deficiency and Sarcoidosis study. Alpha-1 protocol. *Ann Am Thorac Soc*. 2015;12(10):1551–1560.
18 Wan ES, Castaldi PJ, Cho MH, et al. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. *Respir Res*. 2014;15:89.
19 Parr DG, Dirksen A, Piitulainen E, Deng C, Wencker M, Stockley RA. Exploring the optimum approach to the use of CT densitometry in a randomised placebo-controlled study of augmentation therapy in alpha 1-antitrypsin deficiency. *Respir Res*. 2009;10:75.
20 Foreman MG, Wilson C, DeMeo DL, et al. Alpha-1 Antitrypsin PiMZ genotype is associated with chronic obstructive pulmonary disease in two racial groups. *Ann Am Thorac Soc*. 2017;14(8):1280–1287.
21 Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinf*. 2009;10:48.
22 Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*. 2001;125(1-2):279–284.
23 Lipson DA, Criner GJ, Lomas DA. Single-Inhaler triple versus dual therapy in patients with COPD. *N Engl J Med*. 2018;379(6):592–593.
24 Strange C, Herth FJ, Kovitz KL, et al. Design of the endobronchial Valve for Emphysema Palliation Trial (VENT): a non-surgical method of lung volume reduction. *BMC Pulm Med*. 2007;7:10.
25 Mastej E, Gillenwater L, Zhuang Y, Pratte KA, Bowler RP, Kechris K. Identifying protein-metabolite networks associated with COPD phenotypes. *Metabolites*. 2020;10(4):124. https://doi.org/10.3390/metabo10040124.
26 Niewczas MA, Pavkov ME, Skupien J, et al. A signature of circulating inflammatory proteins and development of end-stage renal disease in diabetes. *Nat Med*. 2019;25(5):805–813.
27 Cuvelliez M, Vandewalle V, Brunin M, et al. Circulating proteomic signature of early death in heart failure patients with reduced ejection fraction. *Sci Rep*. 2019;9(1):19202.
28 Tanaka T, Biancotto A, Moaddel R, et al. Plasma proteomic signature of age in healthy humans. *Aging Cell*. 2018;17(5):e12799.
29 Choi MG, Sohn TS, Park SB, et al. Decreased expression of p12 is associated with more advanced tumor invasion in human gastric cancer tissues. *Eur Surg Res*. 2009;42(4):223–229.
30 Figueiredo ML, Kim Y, St John MA, Wong DT. p12CDK2-AP1 gene therapy strategy inhibits tumor growth in an in vivo mouse model of head and neck cancer. *Clin Cancer Res*. 2005;11(10):3939–3948.
31 Zolochevska O, Figueiredo ML. Cell-cycle regulators cdk2ap1 and bicalutamide suppress malignant biological interactions between prostate cancer and bone cells. *Prostate*. 2011;71(4):353–367.
32 Kim Y, McBride J, Kimlin L, Pae EK, Deshpande A, Wong DT. Targeted inactivation of p12, CDK2 associating protein 1, leads to early embryonic lethality. *PLoS One*. 2009;4(2):e4518.
33 Leikauf GD, Borchers MT, Prows DR, Simpson LG. Mucin apoprotein expression in COPD. *Chest*. 2002;121(5 suppl):166S–182S.
34 Vallath S, Hynds RE, Succony L, Janes SM, Giangreco A. Targeting EGFR signalling in chronic lung disease: therapeutic challenges and opportunities. *Eur Respir J*. 2014;44(2):513–522.
35 Yao Y, Gu Y, Yang M, Cao D, Wu F. The gene expression biomarkers for chronic obstructive pulmonary disease and interstitial lung disease. *Front Genet*. 2019;10:1154.
36 de Boer WI, Hau CM, van Schadewijk A, Stolk J, van Krieken JH, Hiemstra PS. Expression of epidermal growth factors and their receptors in the bronchial epithelium of subjects with chronic obstructive pulmonary disease. *Am J Clin Pathol*. 2006;125(2):184–192.
37 Sandhaus RA, Turino G, Brantly ML, et al. The diagnosis and management of Alpha-1 Antitrypsin deficiency in the adult. *Chronic Obstr Pulm Dis*. 2016;3(3):668–682.
38 McElvaney GN, Sandhaus RA, Miravitlles M, et al. Clinical considerations in individuals with alpha1-antitrypsin Pi*SZ genotype. *Eur Respir J*. 2020;55(6):1902410. https://doi.org/10.1183/13993003.02410-2019.
39 Molloy K, Hersh CP, Morris VB, et al. Clarification of the risk of chronic obstructive pulmonary disease in alpha1-antitrypsin deficiency PiMZ heterozygotes. *Am J Respir Crit Care Med*. 2014;189(4):419–427.
40 Cantor J, Ochoa A, Ma S, Liu X, Turino G. Free desmosine is a sensitive marker of smoke-induced emphysema. *Lung*. 2018;196(6):659–663.
41 Manon-Jensen T, Langholm LL, Ronnow SR, et al. End-product of fibrinogen is elevated in emphysematous chronic obstructive pulmonary disease and is predictive of mortality in the ECLIPSE cohort. *Respir Med*. 2019;160:105814.
42 Bowler RP, Jacobson S, Cruickshank C, et al. Plasma sphingolipids associated with chronic obstructive pulmonary disease phenotypes. *Am J Respir Crit Care Med*. 2015;191(3):275–284.
43 Rabinovich RA, Miller BE, Wrobel K, et al. Circulating desmosine levels do not predict emphysema progression but are associated with cardiovascular risk and mortality in COPD. *Eur Respir J*. 2016;47(5):1365–1373.