Original research

Gene coexpression networks reveal novel molecular endotypes in alpha-1 antitrypsin deficiency

Jen-hwa Chu [ID],[1,2] Wenlan Zang,[2] Milica Vukmirovic,[1,3] Xiting Yan,[1,2] Taylor Adams,[1] Giuseppe Deluliis [ID],[1] Buqu Hu,[1] Antun Mihaljinec,[1] Jonas C Schupp,[1] Michael J Becich,[4] Harry Hochheiser,[4] Kevin F Gibson,[5] Edward S Chen,[6] Alison Morris,[5] Joseph K Leader,[7] Stephen R Wisniewski,[8] Yingze Zhang [ID],[5] Frank C Sciurba,[5] Ronald G Collman,[9,10] Robert Sandhaus,[11] Erica L Herzog,[1] Karen C Patterson,[10,12] Maor Sauler,[1] Charlie Strange,[13] Naftali Kaminski [ID],[1] on behalf of the GRADS Investigators

ABSTRACT

Background Alpha-1 antitrypsin deficiency (AATD) is a genetic condition that causes early onset pulmonary emphysema and airways obstruction. The complete mechanisms via which AATD causes lung disease are not fully understood. To improve our understanding of the pathogenesis of AATD, we investigated gene expression profiles of bronchoalveolar lavage (BAL) and peripheral blood mononuclear cells (PBMCs) in AATD individuals.

Methods We performed RNA-Seq on RNA extracted from matched BAL and PBMC samples isolated from 89 subjects enrolled in the Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS) study. Subjects were stratified by genotype and augmentation therapy. Supervised and unsupervised differential gene expression analyses were performed using Weighted Gene Co-expression Network Analysis (WGCNA) to identify gene profiles associated with subjects' clinical variables. The genes in the most significant WGCNA module were used to cluster AATD individuals. Gene validation was performed by NanoString nCounter Gene Expression Assay.

Result We observed modest effects of AATD genotype and augmentation therapy on gene expression. When WGCNA was applied to BAL transcriptome, one gene module, ME31 (2312 genes), correlated with the highest number of clinical variables and was functionally enriched with numerous immune T-lymphocyte related pathways. This gene module identified two distinct clusters of AATD individuals with different disease severity and distinct PBMC gene expression patterns.

Conclusions We successfully identified novel clusters of AATD individuals where severity correlated with increased immune response independent of individuals' genotype and augmentation therapy. These findings may suggest the presence of previously unrecognised disease endotypes in AATD that associate with T-lymphocyte immunity and disease severity.

Key messages

What is the key question?
► What are the effects of alpha-1 antitrypsin genotype and augmentation therapy on bronchoalveolar lavage (BAL) and peripheral blood mononuclear cells (PBMCs) gene expression profiles?

What is the bottom line?
► While these effects of genotype and augmentation therapy are not strong, we identified a signature of genes that distinguish two clusters of patients that differ in extent of emphysema, and lymphocytic infiltration, but not in therapy or genotype, potentially reflecting novel endotypes of disease driven by inflammation.

Why read on?
► This is the largest study that describes the transcriptome of BAL or PBMC samples from well-characterised individuals with alpha-1 antitrypsin deficiency and the first study that aims to use PBMC and BAL gene expression to understand the effects of the causal genetic variant and augmentation therapy on their connection to patient clinical characteristics.

INTRODUCTION

Alpha-1 antitrypsin deficiency (AATD) is a genetic condition that causes early-onset airway obstruction, emphysema and liver cirrhosis. It is estimated that 60 000–100 000 individuals in the USA have AATD.[1] AATD is caused by a mutation in the gene that encodes the alpha-1 antitrypsin protein (AAT), serpin peptidase inhibitor, clade A, member 1 (*SERPINA1*). Multiple AATD mutations have been identified that associate with low AAT serum levels, but the Z-allele is the most common point mutation in SERPINA1 (rs28929474) associated with COPD. This mutation results in a substitution of glutamic acid for lysine at position 342 in the AAT protein,[2] and in the homozygous state produces an 85%–90% serum and lung reduction in this antiprotease that neutralises neutrophil elastase.[3] As a consequence, there is an inability of AAT to neutralise neutrophil elastase, and this contributes to tissue destruction and matrix remodelling that underlie the pathogenesis of COPD.[4]

Currently, the only specific treatment for AAT-related lung diseases is augmentation therapy. The rationale for the AAT augmentation therapy is that replenishing AAT in patients deficient for this protein protects the lungs from excessive neutrophil elastase activity.[2] While augmentation therapy has been shown to slow the decline in CT lung tissue density, it is not a cure for this disease.[5–9] Augmentation therapy does not fully reverse accelerated lung function decline and has no proven effect on COPD exacerbations.[5] Therefore, identifying other mechanisms by which AATD causes COPD and emphysema is necessary to develop therapies to better treat this disease[10–12] while taking into consideration highly heterogenous AATD clinical presentations.[13 14]

To determine the extent and diversity of effects that AAT genotype and augmentation therapy have on gene expression, we analysed gene expression profiles of paired bronchoalveolar lavage (BAL) and peripheral blood mononuclear cell (PBMC) samples based on the Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis Study (GRADS). This multicentre cohort study provided clinical data and RNAseq profiling of BAL cells and PBMC on a well-characterised cohort of AATD individuals.[13]

METHODS
Study participants
The GRADS Alpha-1 Study was a prospective, multicentre study of adults older than age 35 years with PiZZ or PiMZ alpha-1 antitrypsin genotypes.[13] The study protocol, including study design, recruitment and measurements of clinical data, has been previously described.[13] Briefly, 130 individuals with AATD were recruited through the Alpha-1 Foundation Research Registry at the Medical University of South Carolina and directly through physician offices. Written informed consent for genomic research

was obtained from all participants. Figure 1 shows the overview of the study design, enrolment process and the computational analysis flow. The final population consists of 89 individuals with matched BAL and PBMC samples,[13] including individuals with PiZZ not receiving augmentation therapy (n=29), individuals with PiZZ receiving augmentation therapy (n=22) and individuals with PiMZ not receiving augmentation therapy (n=38, see table 1). All patients with pulmonary diseases were more than 6 weeks from an exacerbation and 4 weeks from any antibiotics.

RNAseq
Detailed RNA isolation, library construction, sequencing, normalisation and analysis methods are described in an online supplement. After sequencing, reads were mapped to human genome (hg38) as described[15 16] and normalised to address systematic variation that resulted from the sequencing process.[17 18] Differential gene expression analyses were performed to evaluate the association of PiZ genotype and augmentation therapy status on BAL and PBMC gene expression patterns. For the test of augmentation therapy effect, due to possible confounding between treatment and disease severity that requires treatment, we adjusted for age, sex and disease severity as measured by the FEV_1 % predicted. See the online supplemental data for additional details on differential expression analyses.

Weighted Gene Co-expression Network Analysis (WGCNA) and clustering analysis
We performed WGCNA[19] on BAL to identify gene modules and gene networks significantly correlated with clinical phenotypes (demographics, pulmonary function tests, chest CT findings and augmentation therapy effects). Genes expressed in over 5% of total BAL samples, and with coefficients of variation greater than
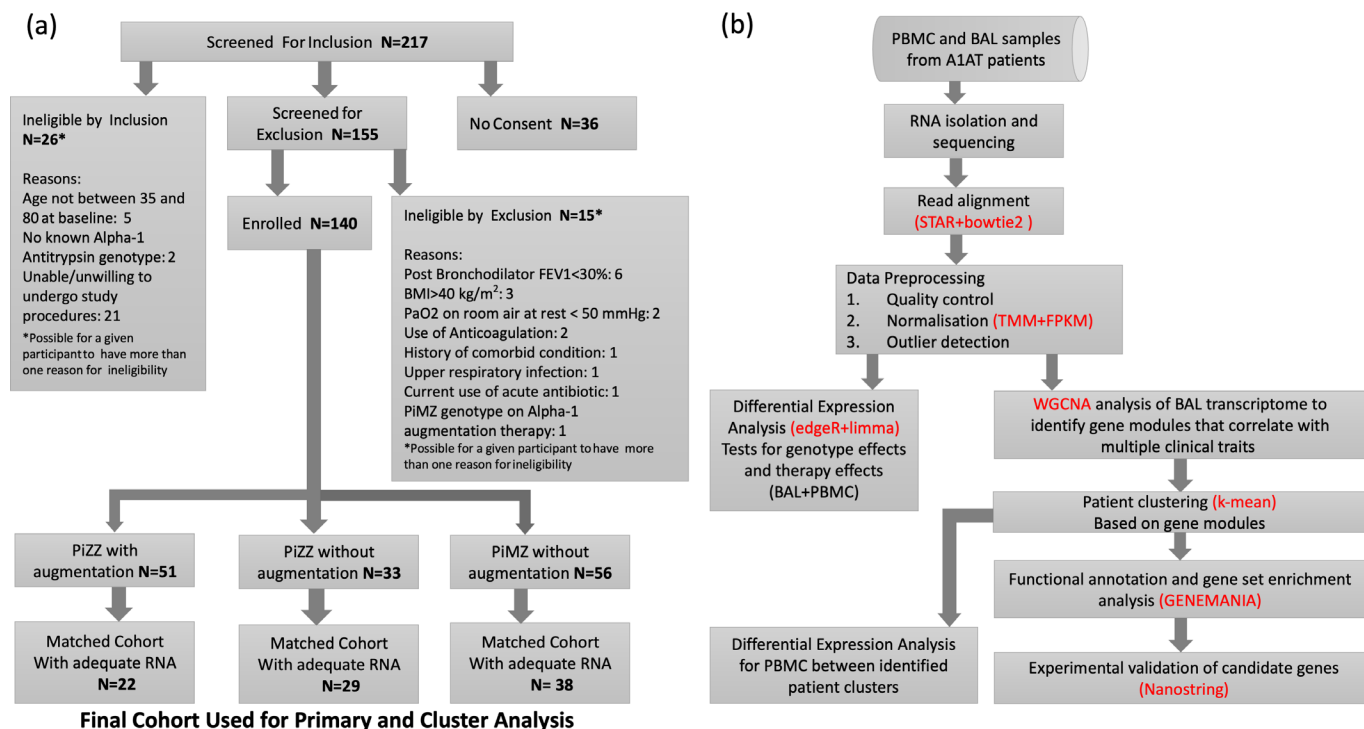
Figure 1 Overview of the study design and computational analysis. (A) The CONSORT Diagram of the GRADS Alpha-1 Study. (B) Diagram of an overview of the computation analysis flow. A1AT, alpha-1 antitrypsin protein; BAL, bronchoalveolar lavage; BMI, body mass index; CONSORT, Consolidated Standards of Reporting Trials; FPKM, fragments per kilobase of transcript per million mapped reads; GRADS, Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis; PBMC, peripheral blood mononuclear cell; TMM, trimmed mean of M-values.

Table 1 GRADS A1AT cohort characteristics

| | PiZZ off therapy group 1 | PiZZ on therapy group 2 | PiMZ group 3 | P values* | | | |
				ANOVA	1 versus 2	1 versus 3	2 versus 3
N	29	22	38				
Age, mean±SD	50.0±9.8	61.8±10.3	52.8±10.0	0.0023	0.0025	0.4587	0.0128
Female (N, %)	16 (55)	16 (73)	27 (71)	0.4033	0.3651	0.4587	1.0000
White (N, %)	28 (97)	22 (100)	38 (100)	0.6048	1.0000	0.6853	1.0000
Ever smoking (N, %)	6 (21)	11 (50)	19 (50)	0.0549	0.0931	0.1012	1.0000
FVC (L), mean±SD	4.13±1.09	3.44±1.18	3.79±1.00	0.1147	0.0931	0.4587	0.2853
FVC % predicted, mean±SD	98.50±14.03	91±17	93±14	0.2768	0.1106	0.7124	0.4004
FEV_1 (L), mean±SD	3.10±1.04	2.01±0.82	2.92±0.82	0.0006	0.0025	0.6644	0.0005
FEV_1% Predicted, mean±SD	89.96±21.25	67±20	91±16	0.0011	0.0069	1.0000	0.0005
DLCO, mean±SD	25.06±9.21	19.33±5.78	25.5±8.7	0.0408	0.0931	1.0000	0.0128
DLCO % predicted, mean±SD	81±23	71.29±16.15	89±19	0.0104	0.0766	0.6276	0.0039
Emphysema presence (N, %)	14 (50)	14 (67)	4 (11)	0.0002	0.4535	0.0133	0.0002
PD15, mean±SD	−931±17	−943±19	−917±20	0.0002	0.0931	0.0285	0.0002
Bronchiectasis (N, %)	21 (75)	15 (71)	36 (97)	0.0252	0.8876	0.1012	0.0169
Airway internal perimeter, mean±SD	1.15±0.24	1.09±0.22	1.13±0.35	0.4363	0.3795	0.4587	1.0000
Airway wall area, mean±SD	40.50±3.36	41.70±4.09	40.76±3.51	0.6048	0.4144	0.7904	0.5555
Alveolar macrophage %, mean±SD	90.38±7.29	85.31±12.06	89.91±7.91	0.3408	0.1881	1.0000	0.2860
BAL eosinophil %, mean±SD	0.31±1.13	1.12±2.98	0.71±1.92	0.2330	0.1106	0.4587	0.6470
BAL lymphocyte %, mean±SD	7.88±6.75	10.00±8.29	8.19±7.72	0.6111	0.4903	0.9288	0.5369
BAL neutrophil %, mean±SD	1.10±1.20	2.88±3.47	0.90±0.86	0.0966	0.1264	0.7124	0.0444

Statistically significant (p<0.05) and borderline signifcant (0.05<p<0.06); p-values are marked in red and yellow, respectively.

*P values are based on non-parametric Kruskal and Wilcoxon tests for numerical variables and χ^2 tests for binary variables. All p values are adjusted for FDR.

A1AT, alpha-1 antitrypsin protein; ANOVA, analysis of variance; BAL, bronchoalveolar lavage; DLCO, diffusing capacity of the lungs for carbon monoxide; GRADS, Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis Study; PD15, Pixel value (HU) at the 15th percentile of HU value histogram.

Table 2 Clinical traits used in WGCNA

Genotype	PiMZ versus PiZZ off therapy	
Augmentation therapy	PiZZ on therapy versus PiZZ off therapy	
Baseline variable	Gender	0 female; 1 male
	Age	(Date enrolled – birthday)/365
Pulmonary function test	FVC	Prebronchodilator FVC
	FVC % PRED	FVC % predicted
	FEV_1	Postbronchodilator FEV_1
	FEV_1 % PRED	Postbronchodilator FEV_1 % predicted
	DLCO	Diffusing capacity of the lungs for carbon monoxide
	DLCO % PRED	Diffusing capacity of the lungs for carbon monoxide % predicted
CT variables	Emphysema presence	Fraction of lung voxels less than −950 HU. Dichotomised as 0 when FRAC950 ≤0.05 and 1 when FRAC950 >0.05.
	PD15	Pixel value (HU) at the 15th percentile of HU value histogram. Dichotomised based on the median value.
	Bronchiectasis	Visual scoring for presence as a categorical variable.
	Airway wall area	Mean airway wall area (mm^2) across all airways automatically detected. Dichotomised based on the median value.
	Airway internal perimeter	Mean airway lumen perimeter (mm) across all airways automatically detected as a continuous variable.
BAL cell differential	Alveolar macrophage %	
	BAL eosinophil %	
	BAL lymphocyte %	
	BAL neutrophil %.	

BAL, bronchoalveolar lavage; WGCNA, Weighted Gene Co-expression Network Analysis.
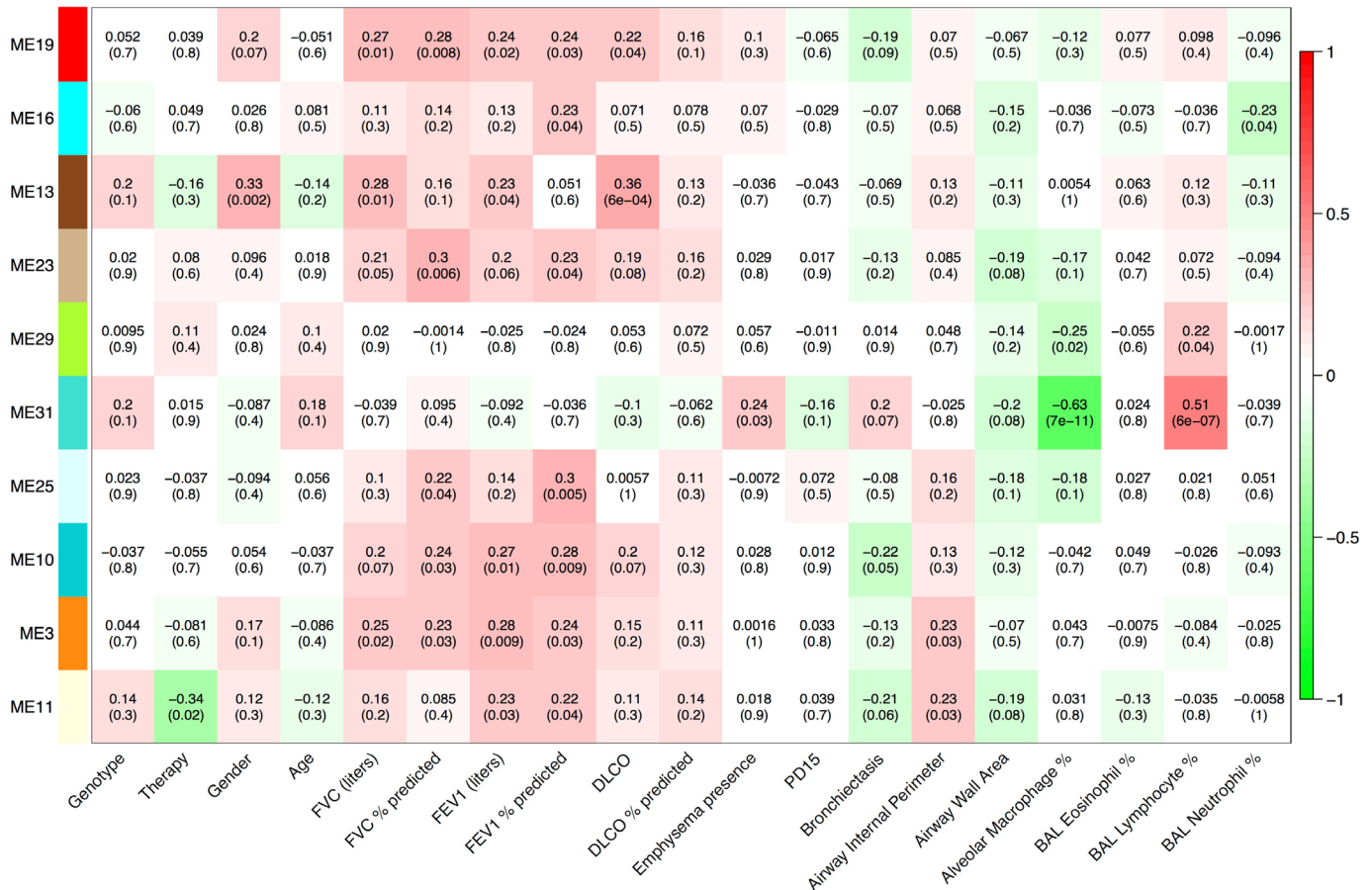
Figure 2 Module–trait relationship from WGCNA for 10 selected modules with the strongest correlations. The numbers in each cell represent the correlation coefficients and p values between each clinical trait and module eigengenes. The module–trait relationship for all 31 modules are presented in online supplemental figure 4). WGCNA, Weighted Gene Co-expression Network Analysis.

1 (n=10 718), were analysed in WGCNA.[19] The default WGCNA setting was applied for data cleaning, outlier detection, network construction and module detection. A Module was defined as a group of co-expressed genes across the samples.[20] Module eigengenes (first principle component of the gene expressions within the module) were taken as the representative of the module and were correlated with AATD health data. Table 2 shows the variables used in WGCNA. Gene module network functional annotation and gene set enrichment analysis were performed using GeneMANIA.[21] GeneMANIA searches large publicly available data sets for connectivity patterns within a module based on coexpression, protein–protein interaction, pathways and other connections that might suggest any functional relations. Gene Ontology pathway enrichment analysis was performed based on FDR-corrected hypergeometric tests. K-means clustering analysis was performed using genes from the chosen WGCNA gene module to create a heatmap and reveal visually distinct clusters. Additional details for our clustering analysis are provided in the online supplemental data.

NanoString validation
To validate the results from WGCNA analysis, we measured the gene expression of 14 genes in BAL by NanoString nCounter Gene Expression Assay, which is generally considered a more accurate technology, in particular for low-quality RNA samples.[22] Among the genes selected were 14 genes with the highest expression level and fold change from WGCNA clustering analysis (CCL5, IL32, LY9, IFITM1, CD3E, PDCD1, ZAP70, LCK,

TGFBR3, FOXP3, IL12RB2, IL18RAP, PRF1 and GZMA) and genes related to genotype (EGR3) and therapy effect (CCDC40, MORN2 and SPA17) in BAL and PBMC samples. Analysis was performed on 77 available samples following manufacturer instructions. nSolver V.3.0 digital analyser software was used to analyse data (see online supplemental material).

RESULTS
Supervised analysis for genotypes and augmentation therapy
The subject demographics are provided in table 1. There were no significant differences between the three groups (PiZZ on therapy, PiZZ off therapy and PiMZ) in basic demographic characteristics other than age (table 1, also see table 2), but individuals not on augmentation therapy (group 1, PiZZ off) were younger, healthier and had better lung function than individuals on AAT augmentation therapy (group 2, PiZZ on). BAL cell counts were similar among three groups, although a trend for PiZZ on having more neutrophils was seen.

We examined the genotype effects in gene expression profiles in BAL and PBMC isolated from individuals with PiZZ off therapy and PiMZ (group 1 vs 3, table 1). Overall, we did not observe significant changes in gene expression profiles. A total of 113 genes in BAL and 181 in PBMC were differentially expressed between PiZZ off therapy and PiMZ (FDR <0.05, see online supplemental tables 2 and 3 and online supplemental figure 1). The 113 BAL genes were enriched for inflammatory, cellular chemotaxis and antiviral responses (online supplemental table
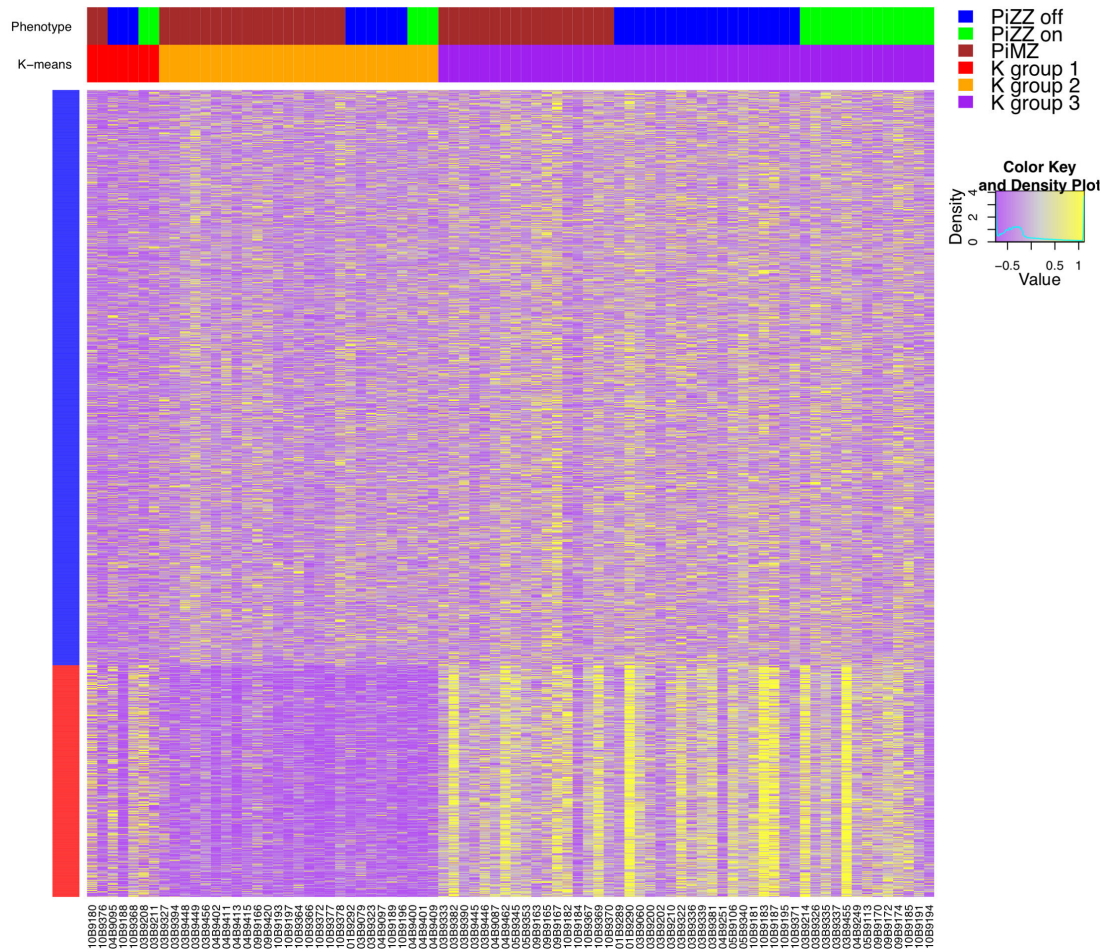
Figure 3 Heatmap for module 31 expression based on K-mean clustering of samples. The colour bar (K-means) indicates the three clusters (1, 2 and 3 from left to right). The colour bar (phenotype) indicates the three groups from the original design (PiZZ on therapy, PiZZ off therapy and PiMZ off therapy). The two subgroups of genes (red and blue bars from the left) are from hierarchical clustering, with the red group visually differentiating clusters 2 and 3.

3). In addition, miR-5010–3 p, miR-6807–3 p and miR-6797–3 p were significantly downregulated in PiMZ group in our BAL data set, while their predicted target genes (RNF11, SUMO2, CYCS and MCMDC2) were significantly upregulated (online supplemental table 4). Downregulated miRs: -let-7b, -9, -30, -34, -3653, -4537, 5196 and 6516 were also identified in PBMC data set. While miR-9 expression is downregulated, its target gene BIK expression was shown upregulated in PiMZ (online supplemental table 5).

We also compared gene profiles of individuals with PiZZ not receiving augmentation therapy (group 1) with PiZZ receiving augmentation therapy (group 2, table 1) the majority of differentially expressed genes were low expressed genes (see online supplemental figure 1 and online supplemental tables 6 and 7). A total of 733 and 64 genes were found to be differentially expressed in BAL and PBMC data set, respectively. These genes were enriched for cilium morphogenesis, chemotaxis and inflammatory response (online supplemental table 8). Many of these genes were predicted to be target genes of following significantly differentially expressed miRs: miR-155-5 p, miR-27-3 p, miR-10b-5p, miR-5001-5 p, miR-1909-5 p, miR-3605-5 p (online supplemental table 9).

To investigate commonality between gene expression profiles in lung and blood within genotypes and therapy effect, we performed comparisons of all expressed genes in BAL and PBMC samples (online supplemental figure 1). While we observed high overall correlation in gene expression between BAL and PBMC (r=0.91, see online supplemental figure 2), the overlap of differentially expressed genes across the two tissues was minimal, with only five and six genes (FPKM >1) significantly differentially expressed in both BAL and PBMC for genotype effect and therapy effect, respectively (online supplemental figure 1 and online supplemental tables 10 and 11).

Weighted Gene Co-expression Network Analysis

Since we observed rather low effect of genotype and therapy on BAL and PBMC gene expression in clinically defined groups, we applied an unsupervised approach to analyse gene expression patterns from BAL, the more disease-relevant tissue compartment in the study of AATD lung diseases. We performed WGCNA analysis on 89 BAL samples to identify gene networks and modules associated with AATD genotypes and augmentation therapy samples. A total of 10 718 genes (CV >1%) were included in the WGCNA. This analysis aimed at identifying modules that are significantly correlated with the measured clinical traits for AATD individuals. Online supplemental figure 3 shows the cluster dendrogram. Thirty-one modules were identified, of which 10 were significantly correlated with multiple clinical variables (figure 2 and online supplemental figure 4).

Table 3 Clinical phenotypes and module 31 sample clustering

		Cluster 1	Cluster 2	Cluster 3	P value* 2 versus 3
Prevalence, n		7	27	48	
Age, year, mean±SD		51.89±7.50	48.43±9.97	57.03±10.40	0.0073
Female sex, n (%)		5 (71)	16 (59)	32 (67)	0.7094
Ever smoker, n (%)		2 (29)	13 (48)	18 (38)	0.6160
Genotype	PiZZ off, n (%)	3 (60)	6 (25)	18 (51)	0.1331
	PiMZ, n (%)	2 (40)	18 (75)	17 (49)	
Therapy	PiZZ on, n (%)	2 (40)	3 (33)	13 (42)	0.7548
	PiZZ off, n (%)	3 (60)	6 (67)	18 (58)	
FVC (L), mean±SD		3.62±1.23	4.06±0.98	3.84±1.11	0.5079
FVC % predicted, mean±SD		89.71±17.49	95.15±13.29	96.55±16.33	0.9373
FEV_1 (L) mean±SD		2.37±1.10	3.21±0.89	2.65±0.90	0.0660
FEV_1 % predicted, mean±SD		71.86±28.70	91.30±16.16	85.51±21.32	0.5447
DLCO, mean±SD		19.60±4.23	27.06±7.41	23.18±8.28	0.0739
DLCO % predicted, mean±SD		74.71±11.57	87.81±17.41	83.28±19.33	0.5447
Emphysema presence, n (%)		4 (57)	2 (7)	22 (48)	0.0020
PD15, mean±SD		−939.00±24.39	−915.89±20.21	−932.11±19.69	0.0204
Bronchiectasis, n (%)		7 (100)	27 (100)	44 (96)	0.6160
Airway internal perimeter, mean±SD		1.07±0.20	1.16±0.36	1.14±0.26	0.6160
Airway wall area, mean±SD		41.40±1.27	41.99±3.99	40.35±3.58	0.1154
Alveolar macrophage %, mean±SD		90.64±5.10	95.92±2.34	85.72±9.29	8.71E-07
BAL eosinophil %, mean±SD		0.36±0.48	0.40±1.08	0.85±2.58	0.7094
BAL lymphocyte %, mean±SD		7.79±5.60	2.79±2.27	11.02±7.56	2.35E-06
BAL neutrophil %, mean±SD		1.21±0.99	0.79±0.64	1.81±2.55	0.2520

*P values are based on non-parametric Kruskal and Wilcoxon tests for numerical variables and χ^2 tests for binary variables. All p values are adjusted for FDR.

BAL, bronchoalveolar lavage; DLCO, diffusing capacity of the lungs for carbon monoxide.

Some of the strongest correlations were identified for the module ME 31 (2312 genes) that correlated with emphysema presence (dichotomised variable; emphysema present defined as greater than 5% of lung voxels less than a −950 HU threshold, r=0.24, p=0.03), bronchiectasis presence as measured by a visual scoring system, r=0.2, p=0.07), alveolar macrophage %, r=−0.6, p=1E-08) and alveolar lymphocyte % (r=0.57, p=6E-08).

Clustering analysis

To examine if the genes in gene module ME31 (2312 genes) could cluster AATD individuals into clinically meaning groups, we performed K-means clustering (figure 3) that distinguished three clusters of AATD individuals. Analysis of individuals' demographics, clinical and CT chest data (table 3) among the three clusters revealed a small group of seven individuals in cluster 1 with slightly increased emphysema. Cluster 2 had younger individuals (p=0.007) compared with cluster 3, as well as slightly higher FEV_1 and diffusing capacity of the lungs for carbon monoxi (DLCO). Cluster 3 had more subjects with a PD15 value above the median and emphysema presence than observed in cluster 2. Cluster 3 had higher BAL lymphocyte % than cluster 2 (11.0±7.6 vs 2.8±2.3) as well as lower macrophage % (85.7±9.3 vs 95.9±2.3), while neutrophil % was not significantly different. There was no significant difference in genotype, therapy, gender or smoking status between clusters 2 and 3 (table 3). Despite the strong associations between these clusters and non-transcriptomic variables, our attempt to cluster patients based on non-transcriptomic variables did not replicate

these clusters (data not shown). These findings all together suggest the gene expression signature from module ME31 identified two well-separated groups of individuals in clusters 2 and 3 with distinct clinical emphysema characteristics that are not primarily driven by either AATD genotype or therapy.

Network analysis and functional enrichment analysis

As the WGCNA module ME 31 (2312 genes) showed significant correlation with key AATD clinical variables and identified two well-defined clusters of AATD individuals, we aimed to further investigate the biological function of these genes. We focused on a subset of 666 genes that were downregulated in cluster 2 and visually differentiated clusters 2 and 3 (marked in red in the lower half of figure 3, online supplemental tables 12 and 13). Figure 4 shows 135 genes in the module that belong to the multiple T cell and immune response enriched pathways (exp: PDCD1, CD3, LCK, ZAP70, CCL5, BTLA, GRAP2 and CD247). Functional enrichment analysis revealed that several immune system pathways were over-represented in this module, including T cell activation, regulation of immune response and antigen receptor-mediated and plasma membrane signalling pathways (see figure 4 and online supplemental table 14). Together these findings suggest an active immune T cell response related to presence of emphysema in AATD individuals.

Figure 4 The gene network and functional enrichment analysis of the module 31. Heatmap for the most enriched pathways (FDR <0.05) and their corresponding 135 genes in module 31, and the GO Term significance score ($-\log_{10}$(FDR)) for each pathway. GO, Gene Ontology.

Differential gene expression for PBMC between newly identified AATD clusters

To better understand biological differences between the individuals in clusters 2 and 3, we performed differential gene expression analysis on PBMC expression profiles using *edgeR*. We found 125 differentially expressed genes between clusters 2 and 3 (see figure 5 and online supplemental table 15). These genes were enriched for plasma lipoproteins, lipid transport and inflammatory responses (online supplemental table 16). Most notably, some of these genes, including CCL18, FABP4, FN1 and miR-9, have been implicated in COPD and were indeed highest in cluster 3 characterised by more severe disease.
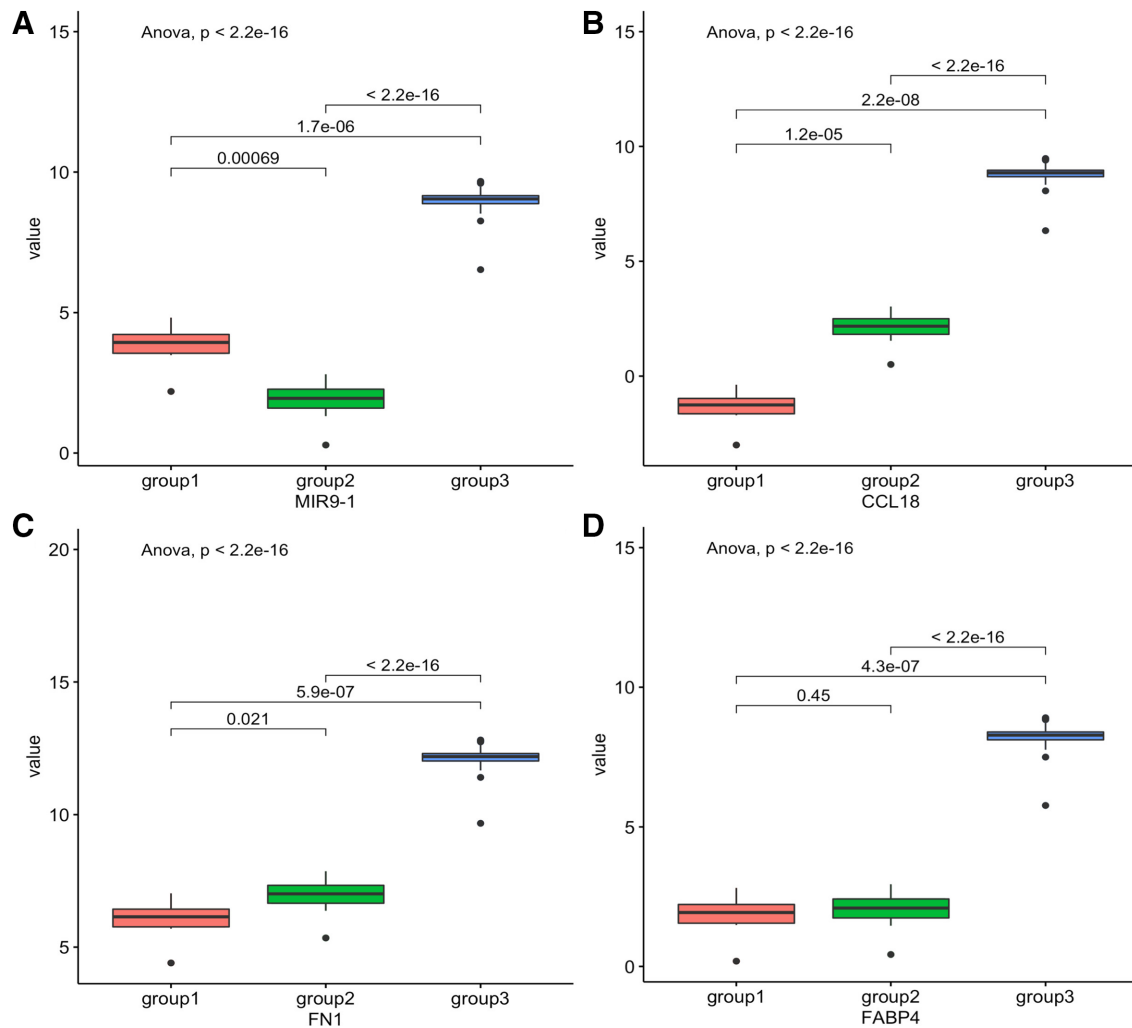
7

Figure 5 Box plots for selected differentially expressed genes in peripheral blood mononuclear cell samples across the clusters characterised by module ME31 (A) MIR9-1, (B) CCL18, (C) FN1 and (D) FABP4.

NanoString validation

Validation by NanoString confirmed that EGR3, a key negative regulator of T cell activation,[23] was upregulated in PiZZ versus PiMZ in both BAL and PBMC (figure 6), and CCDC40, MORN2 and SPA17 genes related to cilia dysfunction in lung disease[24] were upregulated in PiZZ on therapy in BAL (figure 6). In addition, validation of genes associated with emphysema (CCL5[25] and IL32[26]), IFN pathway (IFITM1)[27] and T cell activation and PD1- immune signalling (LY9,[28] CD3E, PDCD1, ZAP70, LCK, TGFBR3, FOXP3, IL12RB2, IL18RAP, PRF1 and GZMA)[29] revealed complete concordance with RNA-Seq data (figure 6).

DISCUSSION

In this study, we applied supervised and unsupervised methods on gene expression profiling of BAL isolated from AATD individuals, and we identified genes associated with genotype and augmentation therapy as well as a gene module enriched for T cell pathways and immune response that correlated with severity of disease independent of genotype and therapy in AATD affected individuals.

Our most impressive finding was a previously unrecognised endotype of individuals with AATD that regardless of genotype or augmentation therapy had more severe respiratory disease and an altered pulmonary inflammatory and immune response driven by lymphocytic influx (cluster 3, table 3). Although this AATD cluster had individuals with increased alveolar lymphocyte %, the cluster was not significantly correlated with AATD therapy or changes in lungs such as: FEV_1 % predicted, DLCO, bronchiectasis or airway wall thickness. Instead, the chest CT analysis using PD15 and emphysema presence were the most informative[30–32] of disease severity.[33] We show that lymphocyte % was elevated in individuals with PD15 lower than the median (p=0.05) and in individuals with emphysema presence (p=0.009, data not shown) suggesting that increase in lymphocyte % is a biological signal of AATD disease severity.

A potential mechanistic clue of AATD lies in the genes that characterised this cluster. Among the most expressed and upregulated BAL genes from cluster 3 were genes implicated in immune responses via T cell activation and PD-1 signalling suggesting the presence of a PD1 immunosuppressive and proinflammatory endotype.[29] Recently, it has been suggested that alpha-1 antitrypsin protein (A1AT) has roles that extend beyond its antiproteolytic effects. For instance, it can regulate inflammatory milieu by inhibiting proliferation of T helper cells and by controlling antigen presentation.[34–36] While our study did not study T cell functions, the identification of an AATD individual cluster characterised by emphysema and T cells suggest the need for detailed immunophenotyping of AATD patients and exploration of specific therapies. This is also supported by our PBMC gene expression analysis. The PBMC of the individuals
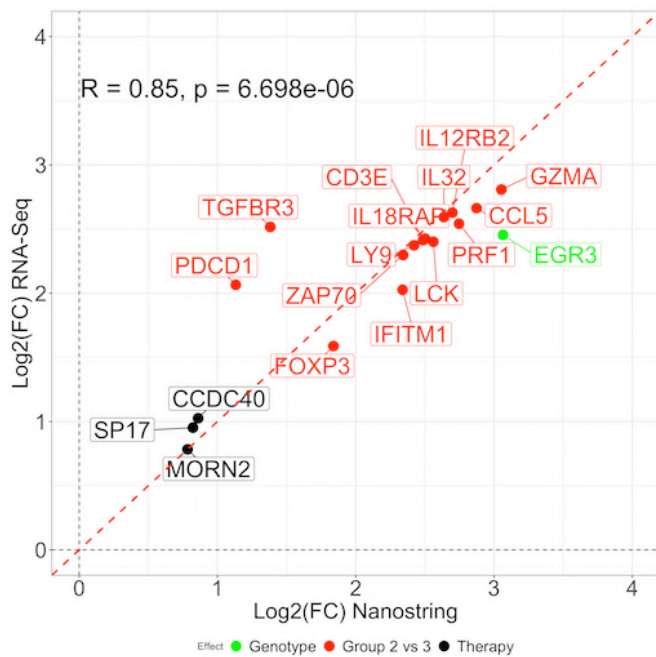
Figure 6 Validation of selected differentially expressed genes using NanoString nCounter. NanoString $\log_2$(FC) between cluster 2 and cluster 3 on x axis and RNA-seq $\log_2$(FC) between cluster 2 and cluster 3 on y axis. Gene names are labelled. FC, fold change.

in cluster 3 exhibited distinct gene expression patterns with increased expression of genes such as CCL18, FABP4, FN1 and miR-9, which were previously associated with COPD. CCL18 was shown to be upregulated in COPD and predict risk for exacerbations.[37] [38] FABP4 was shown increased in COPD and was negatively correlated with lung function.[39] FN1 was also shown to be a gene correlated with progression of COPD.[40] Mir-9 was shown to be associated with the loss of muscle force in patients during an acute exacerbation of COPD.[41] Our unsupervised analysis may have revealed a subgroup of AATD individuals that cluster together based on gene expression and segregates a group of individuals with biological changes in lungs and lung function during development of emphysema that goes beyond clinical phenotyping described previously.[42] The challenge will be to develop and validate diagnostic parameters that characterise these individuals that could be used in the clinic. There also will need to be much work done to exclude environmental stimuli that correlate with these genomic signatures of AATD.

Our study has several limitations that should be considered when interpreting the results. The first is that it contains a relatively small cohort of individuals with AATD, with no independent replication. While this is true, our study in fact represents the first and largest study that describes the BAL and PBMC transcriptome in a carefully phenotyped cohort of AATD individuals. We believe that our findings will encourage others to follow-up on our results and potentially focus on the subgroup we identified to get detailed mechanistic understanding of the mechanisms regulating emphysema in A1AT. Another limitation is the lack of PiMM individuals in the cohort. While we did not find significant difference between gene expression profiles of PiMZ and PiZZ individuals, more difference might have been detected had PiMM individuals been included. Finally, the obvious limitation of bulk RNAseq of BAL cells and PBMC is that we do not necessarily know whether the changes we observed are derived from changes in cell content,

transcriptional regulation or both. For example, it appears that the separation between subphenotypes clusters 2 and 3 is primarily driven by lymphocyte %. However, a comparison with lymphocyte marker genes in single cell RNAseq data from control BAL (data not shown) suggests that there are signals specific to the subphenotypes not explained by lymphocyte %. While gene expression values could have been normalised for cell proportions, doing so would have hidden potentially relevant signals. Therefore, we decided to consider the cell counts as clinical phenotypic attributes. The fact that we found very few differential expression genes for genotype and therapy effects suggests that the changes in gene expression here were not primary driven by differences in cell counts. As detailed immunophenotyping was beyond the scope of our study, we believe that we have shown convincingly that gene expression associated with T cell inflammation is common in patients with emphysema. We hope that this work will motivate other more detailed studies as well as the development of interventions that target the immune aberrations that we observed.

In conclusion, using unsupervised data analysis methods, we identified a subgroup of AATD individuals characterised by more severe disease, and increased T cell inflammation, that is independent of genotype or augmentation therapy. This finding may represent a novel endotype in AATD. Further studies that apply advanced immunophenotyping approaches, including longer follow-up, and focus on emphysema may help to validate this endotype and potentially identify specific interventions.

Author affiliations
[1]Department of Medicine, Yale University School of Medicine, New Haven, Connecticut, USA
[2]Department of Biostatistics, Yale University School of Public Health, New Haven, Connecticut, USA
[3]Department of Medicine, Division of Respirology, McMaster University, Hamilton, Ontario, Canada
[4]Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA
[5]Division of Pulmonary, Allergy and Critical Care Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[6]Department of Medicine, Johns Hopkins University, Baltimore, Maryland, USA
[7]Department of Radiology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA
[8]Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania, USA
[9]Department of Microbiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA
[10]Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA
[11]Department of Medicine, National Jewish Health, Denver, Colorado, USA
[12]Department of Clinical and Experimental Medicine, Brighton and Sussex Medical School, Brigton, UK
[13]Medical University of South Carolina, Charleston, South Carolina, USA

ORCID iDs
Jen-hwa Chu http://orcid.org/0000-0001-7179-9428
Giuseppe Deluliis http://orcid.org/0000-0003-0841-8156
Yingze Zhang http://orcid.org/0000-0001-6947-2901
Naftali Kaminski http://orcid.org/0000-0001-5917-4601

REFERENCES

1 Tuder RM, Janciauskiene SM, Petrache I. Lung disease associated with alpha1-antitrypsin deficiency. *Proc Am Thorac Soc* 2010;7:381–6.
2 American Thoracic Society, European Respiratory Society. American thoracic Society/European respiratory Society statement: standards for the diagnosis and management of individuals with alpha-1 antitrypsin deficiency. *Am J Respir Crit Care Med* 2003;168:818–900.
3 Lomas DA, Evans DL, Finch JT, *et al*. The mechanism of Z alpha 1-antitrypsin accumulation in the liver. *Nature* 1992;357:605–7.
4 Sullivan AL, Stockley RA, Stockley RA, *et al*. *Proteinases and COPD. Chronic obstructive pulmonary disease*. Blackwell Publishing Ltd, 2008: 349–66.
5 Chapman KR, Burdon JGW, Piitulainen E, *et al*. Intravenous augmentation treatment and lung density in severe α1 antitrypsin deficiency (rapid): a randomised, double-blind, placebo-controlled trial. *Lancet* 2015;386:360–8.
6 Stockley RA, Miravitlles M, Vogelmeier C, *et al*. Augmentation therapy for alpha-1 antitrypsin deficiency: towards a personalised approach. *Orphanet J Rare Dis* 2013;8:149.
7 Lawless MW, Greene CM, Mulgrew A, *et al*. Activation of endoplasmic reticulum-specific stress responses associated with the conformational disease Z alpha 1-antitrypsin deficiency. *J Immunol* 2004;172:5722–6.
8 Edgar RG, Patel M, Bayliss S, *et al*. Treatment of lung disease in alpha-1 antitrypsin deficiency: a systematic review. *Int J Chron Obstruct Pulmon Dis* 2017;12:1295–308.
9 Barros-Tizón JC, Torres ML, Blanco I, *et al*. Reduction of severe exacerbations and hospitalization-derived costs in alpha-1-antitrypsin-deficient patients treated with alpha-1 antitrypsin augmentation therapy. *Ther Adv Respir Dis* 2012;6:67–78.
10 Janciauskiene SM, Bals R, Koczulla R, *et al*. The discovery of α1-antitrypsin and its role in health and disease. *Respir Med* 2011;105:1129–39.
11 McCarthy C, Reeves EP, McElvaney NG. The role of neutrophils in alpha-1 antitrypsin deficiency. *Ann Am Thorac Soc* 2016;13(Suppl 4):S297–304.
12 Hassan T, de Santi C, Mooney C, *et al*. Alpha-1 antitrypsin augmentation therapy decreases miR-199a-5p, miR-598 and miR-320a expression in monocytes via inhibition of NFκB. *Sci Rep* 2017;7:13803.
13 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;12:1551–60.
14 Fregonese L, Stolk J. Hereditary alpha-1-antitrypsin deficiency and its clinical consequences. *Orphanet J Rare Dis* 2008;3:16.
15 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.
16 Dobin A, Davis CA, Schlesinger F, *et al*. Star: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
17 Oshlack A, Robinson MD, Young MD. From RNA-seq reads to differential expression results. *Genome Biol* 2010;11:220.
18 Pickrell JK, Marioni JC, Pai AA, *et al*. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 2010;464:768–72.
19 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
20 Modena BD, Bleecker ER, Busse WW, *et al*. Gene expression correlated with severe asthma characteristics reveals heterogeneous mechanisms of severe disease. *Am J Respir Crit Care Med* 2017;195:1449–63.
21 Warde-Farley D, Donaldson SL, Comes O, *et al*. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 2010;38:W214–20.
22 Vukmirovic M, Herazo-Maya JD, Blackmon J, *et al*. Identification and validation of differentially expressed transcripts by RNA-sequencing of formalin-fixed, paraffin-embedded (FFPE) lung tissue from patients with idiopathic pulmonary fibrosis. *BMC Pulm Med* 2017;17:15.
23 Safford M, Collins S, Lutz MA, *et al*. Egr-2 and Egr-3 are negative regulators of T cell activation. *Nat Immunol* 2005;6:472–80.
24 Tilley AE, Walters MS, Shaykhiev R, *et al*. Cilia dysfunction in lung disease. *Annu Rev Physiol* 2015;77:379–406.
25 Di Stefano A, Caramori G, Gnemmi I, *et al*. Association of increased CCL5 and CXCL7 chemokine expression with neutrophil activation in severe stable COPD. *Thorax* 2009;64:968–75.
26 Khawar B, Abbasi MH, Sheikh N. A panoramic spectrum of complex interplay between the immune system and IL-32 during pathogenesis of various systemic infections and inflammation. *Eur J Med Res* 2015;20:7.
27 Pfoertner S, Jeron A, Probst-Kepper M, *et al*. Signatures of human regulatory T cells: an encounter with old friends and new players. *Genome Biol* 2006;7:R54.
28 Sintes J, Cuenca M, Romero X, *et al*. Cutting edge: Ly9 (CD229), a SLAM family receptor, negatively regulates the development of thymic innate memory-like CD8$^+$ T and invariant NKT Cells. *J.i.* 2013;190:21–6.
29 Bhat TA, Panzica L, Kalathil SG, *et al*. Immune dysfunction in patients with chronic obstructive pulmonary disease. *Ann Am Thorac Soc* 2015;12(Suppl 2):S169–75.
30 Parr DG, Dirksen A, Piitulainen E, *et al*. 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.
31 Dawkins PA, Dowson LJ, Guest PJ, *et al*. Predictors of mortality in alpha1-antitrypsin deficiency. *Thorax* 2003;58:1020–6.
32 Cuvelier A, Muir JF, Hellot MF, *et al*. Distribution of alpha(1)-antitrypsin alleles in patients with bronchiectasis. *Chest* 2000;117:415–9.
33 Esquinas C, Janciauskiene S, Gonzalo R, *et al*. Gene and miRNA expression profiles in PBMCs from patients with severe and mild emphysema and PiZZ alpha1-antitrypsin deficiency. *Int J Chron Obstruct Pulmon Dis* 2017;12:3381–90.
34 Bergin DA, Hurley K, McElvaney NG, *et al*. Alpha-1 antitrypsin: a potent anti-inflammatory and potential novel therapeutic agent. *Arch Immunol Ther Exp* 2012;60:81–97.
35 Cosio MG, Bazzan E, Rigobello C, *et al*. Alpha-1 antitrypsin deficiency: beyond the Protease/Antiprotease paradigm. *Ann Am Thorac Soc* 2016;13:S305–10.
36 Polverino F, Baraldo S, Bazzan E, *et al*. A novel insight into adaptive immunity in chronic obstructive pulmonary disease: B cell activating factor belonging to the tumor necrosis factor family. *Am J Respir Crit Care Med* 2010;182:1011–9.
37 Sin DD, Miller BE, Duvoix A, *et al*. Serum PARC/CCL-18 concentrations and health outcomes in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2011;183:1187–92.
38 Dilektasli AG, Demirdogen Cetinoglu E, Uzaslan E, *et al*. Serum CCL-18 level is a risk factor for COPD exacerbations requiring hospitalization. *Int J Chron Obstruct Pulmon Dis* 2017;12:199–208.
39 Zhang X, Li D, Wang H, *et al*. Gender difference in plasma fatty-acid-binding protein 4 levels in patients with chronic obstructive pulmonary disease. *Biosci Rep* 2016;36:e00303.
40 Gosselink JV, Hayashi S, Elliott WM, *et al*. Differential expression of tissue repair genes in the pathogenesis of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010;181:1329–35.
41 Duan Y, Zhou M, Xiao J, *et al*. Prediction of key genes and miRNAs responsible for loss of muscle force in patients during an acute exacerbation of chronic obstructive pulmonary disease. *Int J Mol Med* 2016;38:1450–62.
42 Castaldi PJ, Benet M, Petersen H, *et al*. Do COPD subtypes really exist? COPD heterogeneity and clustering in 10 independent cohorts. *Thorax* 2017;72:998–1006.