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

Jen-hwa Chu 1,2, Wenlan Zang,2 Milica Vukmirovic,1,3 Xiting Yan,1,2 Taylor Adams,1 Giuseppe Delulisi1,1, 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 Zhang9, Frank C Sicurba,5 Ronald G Collman,9,10 Robert Sandhaus,11 Erica L Herzog,1 Karen C Patterson,10,12 Maar Sauer,1 Charlie Strange,13 Naftali Kaminski 1,1 on behalf of the GRADS Investigators

INTRODUCTION

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.

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

Jen-hwa Chu 1,2, Wenlan Zang,2 Milica Vukmirovic,1,3 Xiting Yan,1,2 Taylor Adams,1 Giuseppe Delulisi 1,1, 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 Zhang9, Frank C Sicurba,5 Ronald G Collman,9,10 Robert Sandhaus,11 Erica L Herzog,1 Karen C Patterson,10,12 Maar Sauer,1 Charlie Strange,13 Naftali Kaminski 1,1 on behalf of the GRADS Investigators

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.3 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 neutralise neutrophil elastase, and this contributes to understanding the effects of the causal genetic variant and augmentation therapy on their connection to patient clinical characteristics.
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. While augmentation therapy has been shown to slow the decline in CT lung tissue density, it is not a cure for this disease. Augmentation therapy does not fully reverse accelerated lung function decline and has no proven effect on COPD exacerbations. Therefore, identifying other mechanisms by which AATD causes COPD and emphysema is necessary to develop therapies to better treat this disease while taking into consideration highly heterogenous AATD clinical presentations.

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.

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. The study protocol, including study design, recruitment and measurements of clinical data, has been previously described. 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, 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 and normalised to address systematic variation that resulted from the sequencing process. 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 FEV1 % 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 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...
### Table 1  GRADS A1AT cohort characteristics

<table>
<thead>
<tr>
<th>Age, mean±SD</th>
<th>50.0±9.8</th>
<th>61.8±10.3</th>
<th>52.8±10.0</th>
<th>0.0023</th>
<th>0.0025</th>
<th>0.4587</th>
<th>0.0128</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female (N, %)</td>
<td>16 (55)</td>
<td>16 (73)</td>
<td>27 (71)</td>
<td>0.4033</td>
<td>0.3651</td>
<td>0.7124</td>
<td>0.5555</td>
</tr>
<tr>
<td>White (N, %)</td>
<td>28 (97)</td>
<td>22 (100)</td>
<td>38 (100)</td>
<td>0.6048</td>
<td>1.0000</td>
<td>0.6853</td>
<td>1.0000</td>
</tr>
<tr>
<td>Ever smoking (N, %)</td>
<td>6 (21)</td>
<td>11 (50)</td>
<td>19 (50)</td>
<td>0.0549</td>
<td>0.0931</td>
<td>0.1012</td>
<td>1.0000</td>
</tr>
<tr>
<td>FVC (L), mean±SD</td>
<td>4.13±1.09</td>
<td>3.44±1.18</td>
<td>3.79±1.00</td>
<td>0.1147</td>
<td>0.0931</td>
<td>0.4587</td>
<td>0.2853</td>
</tr>
<tr>
<td>FVC % predicted, mean±SD</td>
<td>98.50±14.03</td>
<td>91±1.17</td>
<td>93±1.14</td>
<td>0.2768</td>
<td>0.1106</td>
<td>0.7124</td>
<td>0.4004</td>
</tr>
<tr>
<td>FEV1 (L), mean±SD</td>
<td>3.10±1.04</td>
<td>2.01±0.82</td>
<td>2.92±0.82</td>
<td>0.0006</td>
<td>0.0025</td>
<td>0.6644</td>
<td>0.0005</td>
</tr>
<tr>
<td>FEV1 % predicted, mean±SD</td>
<td>89.96±21.25</td>
<td>67±2.0</td>
<td>91±1.6</td>
<td>0.0011</td>
<td>0.0069</td>
<td>0.0000</td>
<td>0.0005</td>
</tr>
<tr>
<td>DLCO, mean±SD</td>
<td>25.06±9.21</td>
<td>19.33±5.78</td>
<td>25.5±8.7</td>
<td>0.0408</td>
<td>0.0931</td>
<td>1.0000</td>
<td>0.0128</td>
</tr>
<tr>
<td>DLCO % predicted, mean±SD</td>
<td>81±23</td>
<td>71.29±16.15</td>
<td>89±19</td>
<td>0.0104</td>
<td>0.0766</td>
<td>0.6276</td>
<td>0.0039</td>
</tr>
<tr>
<td>Emphysema presence (N, %)</td>
<td>14 (50)</td>
<td>14 (67)</td>
<td>4 (11)</td>
<td>0.0002</td>
<td>0.4535</td>
<td>0.0133</td>
<td>0.0002</td>
</tr>
<tr>
<td>PD15, mean±SD</td>
<td>−91±1.17</td>
<td>−943±19</td>
<td>−917±20</td>
<td>0.0002</td>
<td>0.0931</td>
<td>0.0285</td>
<td>0.0002</td>
</tr>
<tr>
<td>Bronchiectasis (N, %)</td>
<td>21 (75)</td>
<td>17 (71)</td>
<td>16 (42)</td>
<td>0.252</td>
<td>0.8876</td>
<td>0.1012</td>
<td>0.0169</td>
</tr>
<tr>
<td>Airway internal perimeter, mean±SD</td>
<td>1.15±0.24</td>
<td>1.09±0.22</td>
<td>1.13±0.35</td>
<td>0.4363</td>
<td>0.3795</td>
<td>0.4587</td>
<td>1.0000</td>
</tr>
<tr>
<td>Airway wall area, mean±SD</td>
<td>40.50±3.36</td>
<td>41.70±4.09</td>
<td>40.76±3.51</td>
<td>0.6048</td>
<td>0.4144</td>
<td>0.7904</td>
<td>0.5555</td>
</tr>
<tr>
<td>Alveolar macrophage %, mean±SD</td>
<td>90.38±7.29</td>
<td>85.31±12.06</td>
<td>89.91±7.91</td>
<td>0.3408</td>
<td>0.1881</td>
<td>1.0000</td>
<td>0.2860</td>
</tr>
<tr>
<td>BAL eosinophil %, mean±SD</td>
<td>0.31±1.13</td>
<td>0.11±2.98</td>
<td>0.71±1.92</td>
<td>0.2330</td>
<td>0.1106</td>
<td>0.4587</td>
<td>0.6470</td>
</tr>
<tr>
<td>BAL lymphocyte %, mean±SD</td>
<td>7.88±6.75</td>
<td>10.0±8.29</td>
<td>8.19±7.72</td>
<td>0.6111</td>
<td>0.4903</td>
<td>0.9288</td>
<td>0.5369</td>
</tr>
<tr>
<td>BAL neutrophil %, mean±SD</td>
<td>1.10±1.20</td>
<td>2.88±3.47</td>
<td>0.90±0.86</td>
<td>0.0966</td>
<td>0.1264</td>
<td>0.7124</td>
<td>0.0444</td>
</tr>
</tbody>
</table>

*P values are based on non-parametric Kruskal and Wilcoxon tests for numerical variables and $\chi^2$ 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

<table>
<thead>
<tr>
<th>Genotype</th>
<th>PIMZ versus PIZZ off therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augmentation therapy</td>
<td>PIZZ on therapy versus PIZZ off therapy</td>
</tr>
<tr>
<td>Baseline variable</td>
<td>Gender 0 female; 1 male</td>
</tr>
<tr>
<td>Age</td>
<td>(Date enrolled – birthday)/365</td>
</tr>
<tr>
<td>Pulmonary function test</td>
<td>Prebronchodilator FVC</td>
</tr>
<tr>
<td>FVC</td>
<td>FVC % PRED FVC % predicted</td>
</tr>
<tr>
<td>FEV1</td>
<td>Postbronchodilator FEV1</td>
</tr>
<tr>
<td>FEV1 % PRED</td>
<td>Postbronchodilator FEV1, % predicted</td>
</tr>
<tr>
<td>DLCO</td>
<td>Diffusing capacity of the lungs for carbon monoxide</td>
</tr>
<tr>
<td>DLCO % PRED</td>
<td>Diffusing capacity of the lungs for carbon monoxide % predicted</td>
</tr>
<tr>
<td>Emphysema presence</td>
<td>Fraction of lung voxels less than −950 HU. Dichotomised as 0 when FRAC950 ≤0.05 and 1 when FRA950 &gt;0.05.</td>
</tr>
<tr>
<td>PD15</td>
<td>Pixel value (HU) at the 15th percentile of HU value histogram. Dichotomised based on the median value.</td>
</tr>
<tr>
<td>Bronchiectasis</td>
<td>Visual scoring for presence as a categorical variable.</td>
</tr>
<tr>
<td>Airway wall area</td>
<td>Mean airway wall area (mm2) across all airways automatically detected. Dichotomised based on the median value.</td>
</tr>
<tr>
<td>Airway internal perimeter</td>
<td>Mean airway lumen perimeter (mm) across all airways automatically detected as a continuous variable.</td>
</tr>
<tr>
<td>BAL cell differential</td>
<td>Alveolar macrophage %</td>
</tr>
<tr>
<td>BAL eosinophil %</td>
<td>BAL lymphocyte %</td>
</tr>
<tr>
<td>BAL neutrophil %</td>
<td>BAL, bronchoalveolar lavage; WGCNA, Weighted Gene Co-expression Network Analysis.</td>
</tr>
</tbody>
</table>
Chronic obstructive pulmonary disease

1 (n=10718), 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, TGFB3, FOXP3, IL12RB2, IL18RAP, PRF1 and GZMA) and genes related to genotype (EGR3) and therapy effect (CCDC40, MOR2N 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).
In addition, miR-5010–3p, miR-6807–3p and miR-6797–3p 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-155-5p, miR-27-3p, miR-10b-5p, miR-5001-5p, miR-1909-5p, miR-3605-5p (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

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>P value* 2 versus 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence, n</td>
<td>7</td>
<td>27</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>Age, year, mean±SD</td>
<td>51.89±7.50</td>
<td>48.43±9.97</td>
<td>57.03±10.40</td>
<td>0.0073</td>
</tr>
<tr>
<td>Female sex, n (%)</td>
<td>5 (71)</td>
<td>16 (59)</td>
<td>32 (67)</td>
<td>0.7094</td>
</tr>
<tr>
<td>Ever smoker, n (%)</td>
<td>2 (29)</td>
<td>13 (48)</td>
<td>18 (38)</td>
<td>0.6160</td>
</tr>
<tr>
<td>Genotype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PiZZ off, n (%)</td>
<td>3 (60)</td>
<td>6 (25)</td>
<td>18 (51)</td>
<td>0.1331</td>
</tr>
<tr>
<td>PiMZ, n (%)</td>
<td>2 (40)</td>
<td>18 (75)</td>
<td>17 (49)</td>
<td></td>
</tr>
<tr>
<td>Therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PiZZ on, n (%)</td>
<td>3 (40)</td>
<td>3 (33)</td>
<td>13 (42)</td>
<td>0.7548</td>
</tr>
<tr>
<td>PiZZ off, n (%)</td>
<td>3 (60)</td>
<td>6 (67)</td>
<td>18 (58)</td>
<td></td>
</tr>
<tr>
<td>FVC (L), mean±SD</td>
<td>3.62±1.23</td>
<td>4.06±0.98</td>
<td>3.84±1.11</td>
<td>0.5079</td>
</tr>
<tr>
<td>FVC % predicted, mean±SD</td>
<td>89.71±17.49</td>
<td>95.15±13.29</td>
<td>96.55±16.33</td>
<td>0.9373</td>
</tr>
<tr>
<td>FEV1 (L) mean±SD</td>
<td>2.37±1.10</td>
<td>3.21±0.89</td>
<td>2.65±0.90</td>
<td>0.0660</td>
</tr>
<tr>
<td>FEV1 % predicted, mean±SD</td>
<td>71.86±28.70</td>
<td>91.30±16.16</td>
<td>85.51±21.32</td>
<td>0.5447</td>
</tr>
<tr>
<td>DLCO, mean±SD</td>
<td>19.60±4.23</td>
<td>27.06±7.41</td>
<td>23.18±8.28</td>
<td>0.0739</td>
</tr>
<tr>
<td>DLCO % predicted, mean±SD</td>
<td>74.71±11.57</td>
<td>87.81±17.41</td>
<td>83.28±19.33</td>
<td>0.5447</td>
</tr>
<tr>
<td>Emphysema presence, n (%)</td>
<td>4 (57)</td>
<td>2 (7)</td>
<td>22 (48)</td>
<td>0.0020</td>
</tr>
<tr>
<td>PD15, mean±SD</td>
<td>−939.00±24.39</td>
<td>−915.89±20.21</td>
<td>−932.11±19.69</td>
<td>0.0204</td>
</tr>
<tr>
<td>Bronchiectasis, n (%)</td>
<td>7 (100)</td>
<td>27 (100)</td>
<td>44 (96)</td>
<td>0.6160</td>
</tr>
<tr>
<td>Airway internal perimeter, mean±SD</td>
<td>1.07±0.20</td>
<td>1.16±0.36</td>
<td>1.14±0.26</td>
<td>0.6160</td>
</tr>
<tr>
<td>Airway wall area, mean±SD</td>
<td>41.40±1.27</td>
<td>41.99±3.99</td>
<td>40.35±3.58</td>
<td>0.1154</td>
</tr>
<tr>
<td>Alveolar macrophage %, mean±SD</td>
<td>90.64±5.10</td>
<td>95.92±2.34</td>
<td>85.72±9.29</td>
<td>8.71E-07</td>
</tr>
<tr>
<td>BAL eosinophil %, mean±SD</td>
<td>0.36±0.48</td>
<td>0.40±1.08</td>
<td>0.85±2.58</td>
<td>0.7094</td>
</tr>
<tr>
<td>BAL lymphocyte %, mean±SD</td>
<td>7.79±5.60</td>
<td>2.79±2.27</td>
<td>11.02±7.56</td>
<td>2.35E-06</td>
</tr>
<tr>
<td>BAL neutrophil %, mean±SD</td>
<td>1.21±0.99</td>
<td>0.79±0.64</td>
<td>1.81±2.55</td>
<td>0.2520</td>
</tr>
</tbody>
</table>

*P values are based on non-parametric Kruskal and Wilcoxon tests for numerical variables and χ² tests for binary variables. All p values are adjusted for FDR.

Chronic obstructive pulmonary disease

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=0.03) 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 FEV1, and diffusing capacity of the lungs for carbon monoxide (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 signaling 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.
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.

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.
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 disease24 were upregulated in PiZZ on therapy in BAL (figure 6). In addition, validation of genes associated with emphysema (CCL525 and IL3226), 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: FEV1 % predicted, DLCO, bronchiectasis or airway wall thickness. Instead, the chest CT analysis using PD15 and emphysema presence were the most informative30–32 of disease severity. 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.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...
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 characterize 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 differential gene expression, and focus on emphysema may highlight the immune aberrations that we observed.

In conclusion, using unsupervised data analysis methods, we identified a subgroup of AATD individuals characterized 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.
Chronic obstructive pulmonary disease

Provenance and peer review  Not commissioned; externally peer reviewed.

Data availability statement  Data are available in a public, open access repository. RNAseq data described in this paper has been deposited in the NCBI Gene Expression Omnibus (GEO) under accession code GSE109515.

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

REFERENCES
17. Oshlack A, Robinson MD, Young MD. From RNA-seq reads to differential expression results. Genome Biol 2010;11:120.