OPEN

Transcriptional analysis identifies potential biomarkers and molecular regulators in pneumonia and COPD exacerbation

Wilhelm Bertrams[1], Kathrin Griss[2], Maria Han[2], Kerstin Seidel[1], Andreas Klemmer[3], Alexandra Sittka-Stark[1], Stefan Hippenstiel[2], Norbert Suttorp[2], Florian Finkernagel [ID][4], Jochen Wilhelm[5], Timm Greulich[3], Claus F. Vogelmeier[3], Julio Vera [ID][6] & Bernd Schmeck [ID][1,3,7,8*]

Lower respiratory infections, such as community-acquired pneumonia (CAP), and chronic obstructive pulmonary disease (COPD) rank among the most frequent causes of death worldwide. Improved diagnostics and profound pathophysiological insights are urgent clinical needs. In our cohort, we analysed transcriptional networks of peripheral blood mononuclear cells (PBMCs) to identify central regulators and potential biomarkers. We investigated the mRNA- and miRNA-transcriptome of PBMCs of healthy subjects and patients suffering from CAP or AECOPD by microarray and Taqman Low Density Array. Genes that correlated with PBMC composition were eliminated, and remaining differentially expressed genes were grouped into modules. One selected module (120 genes) was particularly suitable to discriminate AECOPD and CAP and most notably contained a subset of five biologically relevant mRNAs that differentiated between CAP and AECOPD with an AUC of 86.1%. Likewise, we identified several microRNAs, *e.g.* miR-545-3p and miR-519c-3p, which separated AECOPD and CAP. We furthermore retrieved an integrated network of differentially regulated mRNAs and microRNAs and identified HNF4A, MCC and MUC1 as central network regulators or most important discriminatory markers. In summary, transcriptional analysis retrieved potential biomarkers and central molecular features of CAP and AECOPD.

Community acquired pneumonia (CAP) is clinically defined by a sudden onset of severe illness that is accompanied by signs of lower respiratory tract infection, fever, cough and dyspnoea[1]. When left untreated, severe secondary effects such as organ damage and occurrence of bacteria in the blood (bacteremia) can ensue. While subject to variance due to region, season and population characteristics, the incidence of CAP is estimated to lie between 1.5 and 14 cases per 1,000 persons per year, with children under 5 years of age and the elderly of more than 65 years being most strongly affected[2]. Immunocompromised persons also bear a higher risk of CAP contraction.

The leading underlying cause of CAP is infection with the gram-positive bacterium *Streptococcus pneumoniae*, accounting for 30–35% of CAP cases worldwide[3]. The initial colonization of the nasopharynx and the upper respiratory tract often remains asymptomatic. Aspiration into the alveoli can cause severe respiratory or systemic disease, depending on the host immune status and the pneumococcal serotype[1].

As CAP is a multifaceted disease with a host of potential causative agents, the robust identification of microRNAs that are functionally involved in pneumonia depends on the pathogen. In severe Influenza A Virus (H1N1)

[1]Institute for Lung Research, Universities of Giessen and Marburg Lung Center, German Center for Lung Research (DZL), Marburg, Germany. [2]Department of Internal Medicine/Infectious Diseases and Respiratory Medicine, Charité – University Medicine Berlin, Berlin, Germany. [3]Pulmonary and Critical Care Medicine, University Medical Center Giessen and Marburg, German Center for Lung Research (DZL), Marburg, Germany. [4]Institute of Molecular Biology and Tumor Research (IMT), Genomics Core Facility, Philipps-University of Marburg, Marburg, Germany. [5]Justus-Liebig-University, Universities Giessen & Marburg Lung Center, German Center for Lung Research (DZL), Giessen, Germany. [6]Laboratory of Systems Tumor Immunology, Department of Dermatology, Universitätsklinikum Erlangen and Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. [7]Center for Synthetic Microbiology (SYNMIKRO), Philipps-University of Marburg, Marburg, Germany. [8]German Center for Infection Research (DZIF), partner site Giessen-Marburg-Langen, Marburg, Germany. *email: bernd.schmeck@uni-marburg.de

	Healthy	CAP	AECOPD
MicroArray Cohort			
mean age [years $\pm$ SD]	43.8 ± 12.61	75 ± 7.26	62.67 ± 8.84
gender m/f (%)	1/4 (20/80)	4/2 (66.6/33.3)	3/3 (50/50)
pack years $\pm$ SD	0	30 ± 19.75	72.5 ± 57.86
n	5	6	6
TLDA Cohort			
mean age [years $\pm$ SD]	44.9 ± 11.3	65.6 ± 13.6	62.7 ± 9.9
gender m/f (%)	3/4 (42.9/57.1)	4/3 (57.1/42.9)	3/4 (42.9/57.1)
pack years $\pm$ SD	0	14.29 ± 14.25	67.86 ± 54.50
n	7	7	7

Table 1. Basic characteristics of the BioInflame study cohort.

related pneumonia, *e.g.*, miR-150 was significantly up-regulated in serum, whereas the expression of miR-210, miR-126 and miR-222 was decreased, albeit the small sample size precludes definitive conclusions[4].

Chronic obstructive pulmonary disease (COPD) is a poorly reversible condition that is characterized by airflow limitation and subsequent decrease of lung function[5]. It is linked to an abnormal inflammatory response of the respiratory tract and the lung to noxious particles and gases. Disease manifestations are cough, sputum production and dyspnoea. Genetic susceptibility (a1-antitrypsin deficiency), severe respiratory infections, neutrophilic childhood asthma or prolonged exposition to chemicals and dust pose additional risk factors for COPD.

The pathology of COPD is defined by destruction of lung parenchyma (resulting in emphysema) and obstruction of the small airways (resulting in obstructive bronchiolitis). Comorbidities of COPD include cardiovascular diseases, osteoporosis and lung cancer. Acute exacerbations (AE) are episodes of aggravated COPD symptoms and often co-occur with respiratory infection. Furthermore, exacerbations increase the patient's propensity for new exacerbations[6,7]. Exacerbations might require hospitalization, and they increase the COPD mortality rate[8]. This deleterious disease development mandates the search for biomarkers that can identify AECOPD.

The regulation of microRNAs in the lung during COPD or in patients at risk to develop COPD has been described in several studies. Alveolar macrophages[9], induced sputum[10], lung tissue[11] and lung fibroblasts[12] have been analysed. In most cases, these studies found decreased levels of microRNAs. In alveolar macrophages from smokers vs. non-smokers, e.g., the number of down-regulated microRNAs in the smoker cohort was larger than the number of up-regulated microRNAs[9]. Likewise, 27 of 34 differentially expressed microRNAs in induced sputum between smokers and never-smokers were decreased in current smokers[10]. In serum, microRNA let-7c was down-regulated in COPD patients, and miR-34c was decreased in smokers with or without COPD[13], which matched results from sputum[10]. miR-199a-5p has been shown to be down-regulated in regulatory T cells of COPD patients[14]. Among the microRNA candidates that show an upregulation in COPD lung tissue are miR-15b and miR-144 in lung tissue[11]. The same study revealed that in total lung tissue of smokers with and without COPD, 57 microRNAs were up-regulated and 13 microRNAs were down-regulated in patients with COPD compared to healthy smokers, thus contrasting with studies that suggest a comprehensive miRNA downregulation in COPD.

Therefore, we comprehensively analysed both protein coding mRNA and regulatory miRNAs in clinically accessible leukocytes in order to identify markers that not only inform about pathophysiology but may also be suitable as diagnostic biomarkers. We collected PBMCs from healthy donors and patients with either CAP or acute exacerbation of COPD and found very distinct miRNA and mRNA expression pattern, also when compared to healthy controls. Computational analysed retrieved differentially regulated RNA circuits with HNF4A, MCC, MUC1, miR-545-3p, and miR-519c-3p as important parts.

Methods

Patient samples. Patients suffering from community-acquired pneumonia or acutely exacerbated COPD were recruited immediately after hospitalization. In addition, healthy subjects were recruited (Tables 1 and S1). Inclusion criteria of community-acquired pneumonia patients included pulmonary infiltrates on chest x-ray and clinical presentation. Inclusion criteria of COPD patients with an acute exacerbation (AECOPD) were an acute respiratory worsening requiring a hospitalization in pre-diagnosed COPD, but without pulmonary infiltrates on chest x-ray or highly elevated CRP-levels (Table S1). Only one CAP patient presented with a diagnosis of COPD. No other CAP patient showed clinical signs or a medical history of COPD. Immunosuppressed, pregnant and HIV-positive patients were excluded from the study. The BioInflame study was approved by the ethics committee of the Charité - Universitätsmedizin Berlin (EA2/030/09) and the University Medical Center Marburg (55/17). All blood donors were at least 18 years of age and provided written informed consent for use of their blood samples for scientific purposes. PBMCs were isolated by Pancoll gradient centrifugation of one collected Vacutainer EDTA-tube (6 ml whole blood). All methods were performed in accordance with the relevant guidelines and regulations.

Microarray and Taqman Low Density Array. PBMC samples from healthy donors, AECOPD patients and CAP patients were obtained via gradient centrifugation of 6 ml whole blood. Cells were lysed in TriReagent (Invitrogen, Carlsbad USA), and RNA was extracted by Phenol/Chloroform precipitation. Gene expression was measured on an Agilent Sureprint mRNA array as described before[15], using an Agilent SurePrint G3 Human Gene

Expression 8 × 60 K v2 Microarray kit with 200 ng of RNA per sample. Background correction was performed with the normexp model using negative control spots and offset of 1. Spot intensities were quantile normalized between arrays. Array quality was assessed using spatial hybridization homogeneity, per-probe intra-array signal variation and per-probe median-to-mean ratios. Contrasts were fitted using limma. Genes that were significantly regulated (unadjusted p < 0.01) in either of the contrasts AECOPD vs. healthy, CAP vs. healthy or AECOPD vs. CAP were selected (Table S2). Genes whose expression changes could be attributed to changes in cellular composition of the PBMC fractions between donors (Table S3) were removed from the list of significantly regulated genes. To this end, linear modeling was applied using statsmodels[16] ordinary least squares regression (OLS-function). Gene expression was standardized to run from 0 to 1, and modelled as a linear combination of disease states (boolean: AECOPD/CAP) and cellular composition (real valued: neutrophils/eosonophils/monocytes/lymphocytes in %) plus an error term. The data are deposited under the accession number GSE94916. Expression data from the independent reference dataset GSE6269[17] was used as deposited in GEO. Data were obtained via the getGEO function from the GEOquery package[18] in R. Contrasts were fitted using limma. Genes were considered to be differentially expressed when unadjusted p < 0.05.

Taqman Low Density Arrays (TLDAs) were applied to reveal disease-related-regulation of microRNAs as described before[19]. The data are deposited under the accession number GSE136390.

Bioinformatics. The R programming software[20] was used with the packages Weighted Gene Co-expression Network Analysis (WGCNA) v. 1.66[21], Ensemble Feature Selection (EFS) v. 1.0.3[22], gplots v. 3.0.1 and ggplot2 v. 3.1.0.9. WGCNA identifies expression modules in a given dataset by correlating gene expression data across samples[21]. Gene groups (modules) were established on the basis of high adjacency and subsequent gene-tree cutting with a minimum module size of 30 and maximum height of 1. A weighted average for each module (the module eigengene, which is the first principal component of the correlated genes in a given module) was used to compute a measure of module membership (kME) for each gene in the module. kME values represent the correlation between a gene's expression and the module eigengene (ME). WGCNA was used here to find modules that allow discrimination between CAP and AECOPD. Biological connections of identified factors were investigated with Ingenuity Pathway Analysis[23] and visualized with Cytoscape v. 3.7.1[24]. Hub genes from the identified discriminatory module (module I) were ranked by their contribution to disease separation using EFS.

Gene expression intensities of the independent dataset GSE6269[17] were probed for enrichment of module I genes by Gene Set Enrichment Analysis (GSEA)[25], available at http://software.broadinstitute.org/gsea/index.jsp.

Results
Differential gene expression in CAP and AECOPD as compared to healthy donors can be detected in PBMCs.
We probed the PBMC transcriptome of healthy donors, AECOPD patients and CAP patients for potential biomarkers by microarray and Taqman Low Density Array. As illustrated (Fig. 1A), coding transcripts that were significantly regulated (unadjusted p < 0.01) in the contrasts CAP vs. healthy, AECOPD vs. healthy or CAP vs. AECOPD were identified (Table S2). Then, these genes were correlated to the PBMC composition of the individual donors to filter out gene expression changes that were attributable to a change of PBMC cellular composition (Fig. 1B), which removed 362 of 1,983 genes (Table S3). While these genes hold potential as biomarkers, they are not suited for functional network building, and they might be strongly influenced by composition of the drawn blood sample. The remaining coding transcripts (1,621, Table S4) tend to broadly follow similar regulation patterns in CAP and AECOPD when compared to healthy controls (Fig. 1C). All 1,621 DE genes are shown in a z-score representation and were probed for GO term enrichment. To this end, they were separated into genes that were up-regulated in disease or down-regulated in disease. The strongest enrichments ($-\log_{10}$ pvalue > 5) are shown including exemplary genes that were driving this enrichment (Fig. 1D). Detailed z-score heatmaps of the genes from the six populated outer regions of Fig. 1C, sorted by hierarchical clustering, are given in the supplement section "Appendix to Fig. 1".

Correlation analysis yields a subset of genes that can discriminate CAP from AECOPD and is also applicable to an independent PBMC transcriptomic dataset.
In order to test the disease-discriminatory potential of the complete gene set from Fig. 1, we applied PCA analysis which successfully clustered the healthy and disease phenotypes, but failed to properly resolve differences between the CAP and AECOPD samples (Fig. S1). Therefore, we applied supervised Weighted Gene Co-expression Network Analysis (WGCNA) to this subset of 1,621 significantly regulated coding genes. Parameters were iteratively set to yield the largest possible modules while still preserving a given gene set's ability to separate AECOPD from CAP. We tested each module Eigengene for its correlation with disease phenotype (CAP or AECOPD). We finally identified nine modules, the bottom module IX serving as collector for genes that could not be assigned to any of the other modules (see file modules_expression_values.xlsx). Module I with 120 genes (Table S5) showed by far the most significant and clear discrimination between CAP and AECOPD (Fig. S1), and consequently discriminated these disease subtypes in PCA, maximizing the distances to nearest neighbours from different clusters (Fig. 2A and Table S6). In module I, correlation between the individual gene's expression and the module Eigengene was robust for the module genes ($R^2 = 0.609$, Fig. S2), highlighting the overall stability of gene expression within this module and its suitability for disease discrimination. Module hub genes (Fig. 2B) were identified by Ingenuity Pathway Analysis on the basis of their connectivity with other members from module I. In an additional step, all genes with at least two intra-module interacting partners were probed by Ensemble Feature Selection in order to find genes with high importance in separating CAP from AECOPD and biological relevance in the module. We thereby ranked the identified genes by their contribution to the model and identified five out of twelve genes that are the most important discriminators (EFS selected features: MCC, MUC1, PDK4, TSN and HNF4A) and that perform with an AUC of 86.1% in a ROC analysis (Fig. 2C). Furthermore, we tested module I on an independent
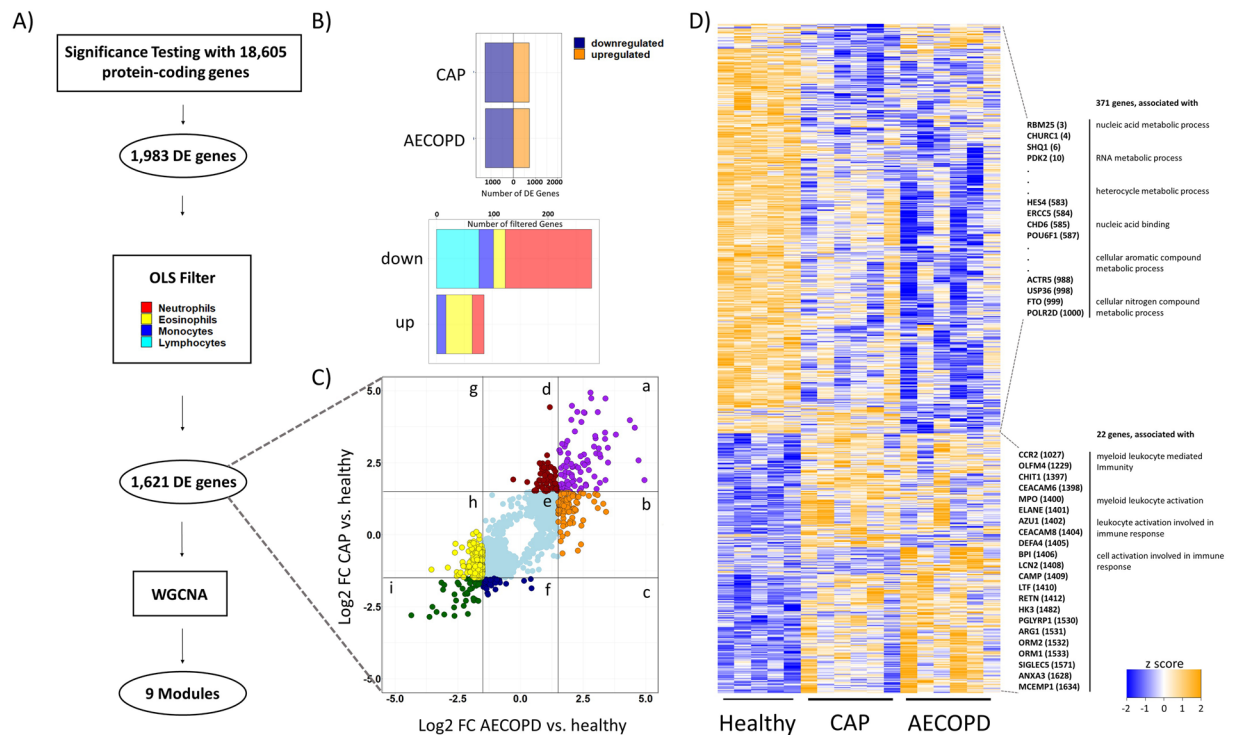
3

Figure 1. CAP and AECOPD have an impact on the PBMC transcriptome. Schematic diagram of the workflow (**A**). From a total number of 1,983 pairwise significantly differentially expressed (DE) genes between the coding transcriptomes of PBMCs from AECOPD or CAP patients or healthy controls, genes whose expression changes were attributable to changes in cell number were removed (OLS filter) (**B**). The remaining 1,621 genes that were significantly differentially expressed are shown by their $\log_2$ fold change vs. healthy control (**C**). All 1,621 are shown in a z-score representation. Enriched GO terms ($-\log_{10}$ pvalue > 5) and exemplary genes that drive this enrichment are indicated with their position (in brackets) in the heatmap. Gene ranking was achieved by hierarchical clustering (**D**).

dataset of blood leukocytes from patients with acute infections, Gene Expression Omnibus accession number GSE6269[17]. We used Gene Set Enrichment Analysis (GSEA)[25] to probe for enrichment of module I genes in this dataset, and we found a marked positive enrichment of these genes when comparing *S. pneumoniae* with *S. aureus* infected patients (Fig. 2D), strengthening the applicability of module I to discriminate diseases. Finally, the full RNA network was constructed with IPA database knowledge. A central node, HFN4A, has been described before as an important transcription factor of COPD-sensitive genes before in ATII cells[26]. The connection of HNF4A to other members of module I has been experimentally established by a protein-chromatin interaction study[27]. Furthermore, it was among the top five EFS-selected features. Another central node of the network was E2F1, a cell cycle-associated transcription factor (Fig. 3), which was also identified as an important feature by EFS.

Module I can be integrated with autologous microRNA expression data. In order to substantiate our findings and extend them to a different class of RNA, we collected the microRNA profile from all groups by Taqman Low Density Array. The microRNAs that passed visual inspection of their amplification curves and that were found to be significantly regulated between CAP and AECOPD were miR-135a-5p, miR-139-5p, miR-141-3p, miR-184, miR-365a-3p, miR-519c-3p, miR-545-3p, and miR-642a-3p. We integrated these microRNAs with the set of 1,621 genes by miRNA/mRNA target filtering into a comprehensive network of miRNA/mRNA interaction (Fig. 4). Furthermore, these miRNAs proved to be suitable discriminators of AECOPD and CAP. The five feature miRNAs as selected by EFS increased the AUC from 57.1% to 87.8% in a ROC analysis (Fig. S3).

Discussion

Pneumonia and AECOPD are among the most common causes for acute hospitalization, and the difficulty of a timely differential diagnosis mandates the search for accessible biomarkers, as currently pneumonia tends to be underdiagnosed in COPD patients[28]. In this study, we identified RNA marker molecules in PBMCs to differentiate between healthy subjects, CAP patients and AECOPD patients, while taking PBMC composition into account. Both disease conditions showed clear differences to healthy controls as illustrated on the basis of all 1,621 coding genes that were significantly regulated in either the CAP vs. healthy or AECOPD vs. healthy contrast. Significant GO term enrichment testing of these genes yielded a strong correlation with metabolic processes and leukocyte activation. Nevertheless, the total of these differentially expressed genes was not suitable to differentiate CAP from AECOPD. The main contribution of this study is therefore the identification of a module of 120 genes (module I) that can differentiate between CAP and AECOPD.
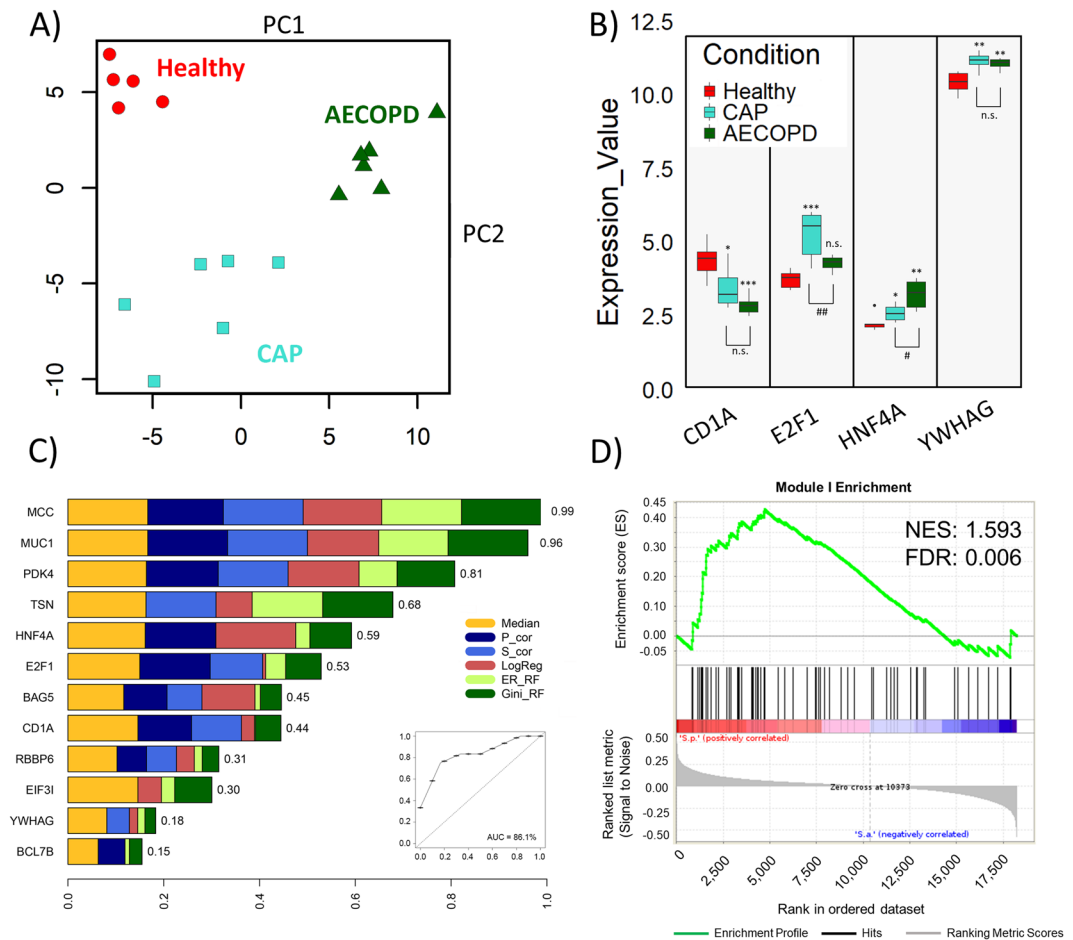
Figure 2. Module I is highly discriminatory for CAP/AECOPD. Module I genes effectively separate healthy donors (red) from CAP (turquoise) and AECOPD patients (green) in a PCA. Data were centred and scaled before analysis (**A**). Microarray expression values of module I hub genes, which were identified as such by Ingenuity Pathway Analysis, show differential expression as a function of disease status. The boxes define the first and third quartile, whiskers extent at most 1.5*IQR (interquartile range) from the hinge. The horizontal bar represents the median (**B**). Hub genes and connecting genes from module I were analysed by Ensemble Feature Selection (EFS) for their potential to differentiate between CAP and AECOPD. We used the default EFS algorithm, which ranks feature importance as an additive value comprised of median comparison (p values from Wilcoxon signed rank test), S_Cor (Spearman's rank correlation test by fast correlation filter), P_Cor (Pearson's product moment correlation test by fast correlation filter), LogReg (beta-Values of logistic regression), ER_RF (Error-rate-based variable importance measure embedded in random forest) and Gini_RF (Gini-index-based variable importance measure embedded in random forest). The plot shows the relative normalized importance for each individual method plus the combined performance of MCC, MUC1, PDK4, TSN and HNF4A in a ROC analysis (**C**). Genes from module I positively correlate with blood leukocyte DE (p < 0.05) mRNA from *Streptococcus Pneumoniae*-infected patients from an independent study (*GSE6269*) when compared to *Staphylococcus Aureus*-infected patients from the same study. NES = Normalized enrichment score. FDR = False Discovery Rate (**D**).

In order to find this gene subgroup with deeper discriminative potential, we applied the WGCNA pipeline. We identified a module (termed module I) containing 120 genes, which was highly performant in discrimination of CAP and AECOPD. While many genes in this module are significantly regulated between CAP and AECOPD, as was to be expected, the cluster also contains genes that fail to pass this significance threshold. We corroborated the identified module by constructing an integrated network that highlights biologically relevant intra-modular interactions. An integral part of this network, linking the main hubs HNF4A and E2F1, is MCC, which we also identified as the most potent discriminator between CAP and AECOPD by EFS. One function of MCC is the negative regulation of canonical Wnt signalling[29], which has been observed in the airways of COPD patients[30]. The second differentiating gene as ranked by EFS is MUC1, a member of the mucin family. MUC1 has been shown to be induced by IFNγ[31] and IFNγ was shown to be protective in a mouse model of pneumococcal pneumonia[32]. Of note, Serum Amyloid A2 (SAA2) is also a part of this network downstream of HNF4A, and was recently described as a COPD-associated IL17 response signature gene in bronchial airway epithelial brushings[33]. We furthermore extended this network with microRNAs that were differentially expressed in CAP or
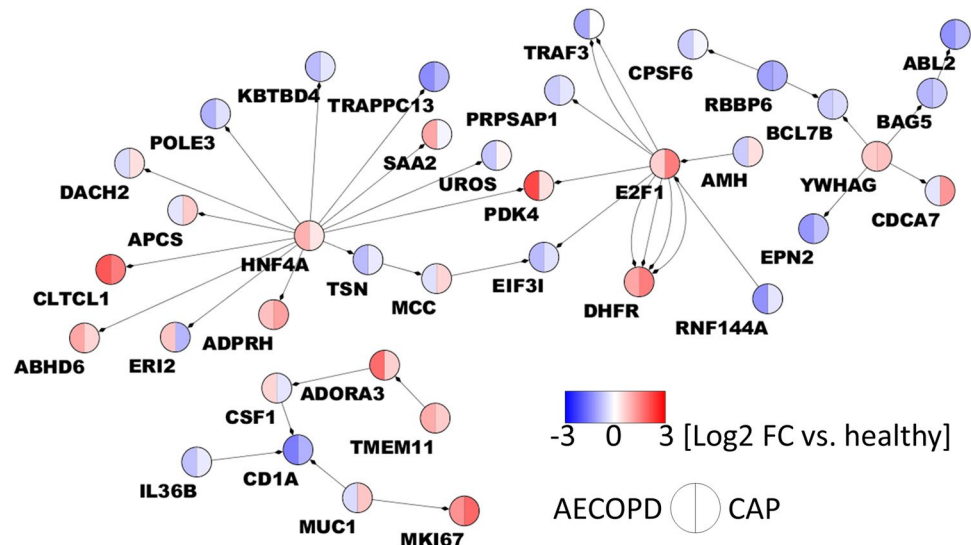
Figure 3. Central network constructed on the basis of genes included in module I. Ingenuity Pathway Analysis was used to integrate module I biologically. Parameters of interaction screening were set to only include experimentally validated candidates from the *human* and *uncategorized* species settings. Orphan nodes and isolated node pairs were deleted. Log$_2$ fold expression changes vs. healthy control are shown for AECOPD (left semicircle) and CAP (right semicircle). The network was visualized with Cytoscape v. 3.7.1.
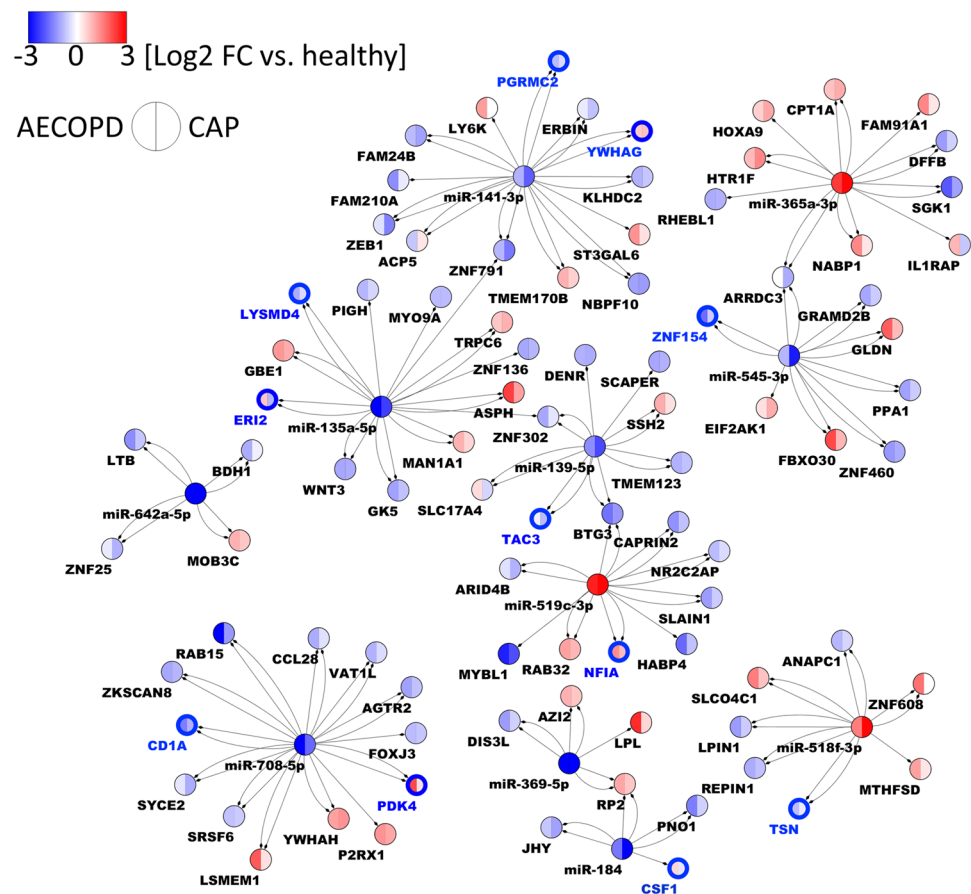
Figure 4. miRNA/mRNA core network of CAP/AECOPD. The entire dataset (1,621 genes) was probed for interaction with the identified miRNA pool by Ingenuity Pathway Analysis. Only experimentally validated interactions or interactions with a high prediction score (binary IPA parameter) are shown. Log$_2$ fold expression changes vs. healthy control are shown for AECOPD (left semicircle) and CAP (right semicircle). Nodes circled in blue are part of module I. The network was visualized with Cytoscape v. 3.7.1.

AECOPD compared to healthy controls, and observed that these microRNAs have targets that are part of module I, strengthening the module's hypothesized biological functionality. Additionally, miR-365 has been described before in the context of COPD[11] and has been indicated as a negative regulator of pro-inflammatory IL-6 gene expression, a pro-inflammatory marker[34]. An IPA database search identified twelve targets of miR-365 that were differentially regulated between CAP and AECOPD in our dataset. Two of these, CPT1A and SGK1, are important factors in COPD pathogenesis. While CPT1A is a cigarette smoke-inducible regulator of fatty acid oxidation[35], SGK1 increases CFTR, which might be a counter-mechanism to the observed down-regulation of CFTR in COPD[36,37]. Another miR-365 target, RHEBL1, activates mTOR signalling[38], which controls cell proliferation via SGK1[39]. Furthermore, we found that miR-708-5p was markedly downregulated in AECOPD and CAP. This microRNA was also downregulated in the bronchial biopsies of patients with COPD after long-term treatment with inhaled corticosteroids[40]. In PBMCs, this miRNA was found to be upregulated in patients with severe vs. mild alpha1 antitrypsin deficiency (AATD), which is a genetic determinant of COPD[41]. In the same study, miR-139-5p was shown to be downregulated in AATD patients, which is a trend we also observed in AECOPD and CAP patients.

One limitation of our study is that it lacks an independent validation cohort. Such a cohort would be required to test whether the chosen approach of module finding is superior to mere significance testing. In support of our model, we can show that it increases the silhouette distance in PCA when compared to gene sets selected on the basis of significance only (Table S6) and that the identified module I is also applicable to an independent dataset. As our patient numbers are small (Table 1), the performance of the module in determining whether a patient has both CAP and AECOPD remains to be assessed.

It has been recently demonstrated that the transcriptomic differences in lung tissue in patients with COPD are largely independent from the gene expression patterns determined in whole blood[42]. Still, reproducible gene co-expression modules have been identified from whole blood of COPD patients which encourages the search for easily reachable biomarkers[43]. We have focused on AECOPD in this study, as it represents a disease manifestation with an acute therapeutical and also diagnostic need, and because it might have the potential to yield biomarkers that are absent in non-AE COPD. In CAP, while procalcitonin is an established biomarker, the need to extend the current panel has been recently stressed[44].

Taken together, we identified a panel of protein coding transcripts in PBMCs that were differentially expressed between healthy subjects, and patients suffering from CAP or AECOPD. We identified a regulatory network of these mRNAs and regulated miRNAs, extracted central nodes of the network and retrieved a small panel of RNA markers for potential clinical application that require further validation in independent cohorts.

References

1. van der Poll, T. & Opal, S. M. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. *Lancet* **374**, 1543–1556, https://doi.org/10.1016/S0140-6736(09)61114-4 (2009).
2. Prina, E. *et al.* Risk factors associated with potentially antibiotic-resistant pathogens in community-acquired pneumonia. *Ann. Am. Thorac. Soc.* **12**, 153–160, https://doi.org/10.1513/AnnalsATS.201407-305OC (2015).
3. Cilloniz, C., Civljak, R., Nicolini, A. & Torres, A. Polymicrobial community-acquired pneumonia: An emerging entity. *Respirology* **21**, 65–75, https://doi.org/10.1111/resp.12663 (2016).
4. Moran, J. *et al.* Circulating levels of miR-150 are associated with poorer outcomes of A/H1N1 infection. *Exp. Mol. Pathol.* **99**, 253–261, https://doi.org/10.1016/j.yexmp.2015.07.001 (2015).
5. Vogelmeier, C. F. *et al.* Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am. J. respiratory Crit. care Med.* **195**, 557–582, https://doi.org/10.1164/rccm.201701-0218PP (2017).
6. Guerrero, M. *et al.* Readmission for Acute Exacerbation within 30 Days of Discharge Is Associated with a Subsequent Progressive Increase in Mortality Risk in COPD Patients: A Long-Term Observational Study. *PLoS One* **11**, e0150737, https://doi.org/10.0151371/journal.pone.0150737. eCollection 0152016 (2016).
7. Wedzicha, J. A., Brill, S. E., Allinson, J. P. & Donaldson, G. C. Mechanisms and impact of the frequent exacerbator phenotype in chronic obstructive pulmonary disease. *BMC Med.* **11**, 181, https://doi.org/10.1186/1741-7015-1111-1181. (2013).
8. Suissa, S., Dell'Aniello, S. & Ernst, P. Long-term natural history of chronic obstructive pulmonary disease: severe exacerbations and mortality. *Thorax.* **67**, 957–963, https://doi.org/10.1136/thoraxjnl-2011-201518. Epub 202012 Jun 201518 (2012).
9. Graff, J. W. *et al.* Cigarette smoking decreases global microRNA expression in human alveolar macrophages. *PLoS one* **7**, e44066, https://doi.org/10.1371/journal.pone.0044066 (2012).
10. Van Pottelberge, G. R. *et al.* MicroRNA expression in induced sputum of smokers and patients with chronic obstructive pulmonary disease. *Am. J. respiratory Crit. care Med.* **183**, 898–906, https://doi.org/10.1164/rccm.201002-0304OC (2011).
11. Ezzie, M. E. *et al.* Gene expression networks in COPD: microRNA and mRNA regulation. *Thorax* **67**, 122–131, https://doi.org/10.1136/thoraxjnl-2011-200089 (2012).
12. Sato, T. *et al.* Reduced miR-146a increases prostaglandin E(2)in chronic obstructive pulmonary disease fibroblasts. *Am. J. respiratory Crit. care Med.* **182**, 1020–1029, https://doi.org/10.1164/rccm.201001-0055OC (2010).
13. Akbas, F., Coskunpinar, E., Aynaci, E., Oltulu, Y. M. & Yildiz, P. Analysis of serum micro-RNAs as potential biomarker in chronic obstructive pulmonary disease. *Exp. lung Res.* **38**, 286–294, https://doi.org/10.3109/01902148.2012.689088 (2012).
14. Chatila, W. M. *et al.* Blunted expression of miR-199a-5p in regulatory T cells of patients with chronic obstructive pulmonary disease compared to unaffected smokers. *Clin. Exp. immunology* **177**, 341–352, https://doi.org/10.1111/cei.12325 (2014).
15. Hoffmann, J. *et al.* Distinct differences in gene expression patterns in pulmonary arteries of patients with chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis with pulmonary hypertension. *Am. J. respiratory Crit. care Med.* **190**, 98–111, https://doi.org/10.1164/rccm.201401-0037OC (2014).
16. Seabold, S. P. J. Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference*, 57–61 (2010).
17. Ramilo, O. *et al.* Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* **109**, 2066–2077, https://doi.org/10.1182/blood-2006-02-002477 (2007).

18. Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847, https://doi.org/10.1093/bioinformatics/btm254 (2007).

19. Scheller, N. *et al.* Pro-viral miRNAs detected in BALF extracellular vesicles of patients with influenza virus-induced ARDS. *The Journal of infectious diseases*, https://doi.org/10.1093/infdis/jiy554 (2018).

20. Core Team, R. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2018).

21. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma.* **9**, 559, https://doi.org/10.1186/1471-2105-9-559 (2008).

22. Neumann, U., Genze, N. & Heider, D. EFS: an ensemble feature selection tool implemented as R-package and web-application. *BioData Min.* **10**, 21, https://doi.org/10.1186/s13040-017-0142-8 (2017).

23. Kramer, A., Green, J., Pollard, J. Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530, https://doi.org/10.1093/bioinformatics/btt703 (2014).

24. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504, https://doi.org/10.1101/gr.1239303 (2003).

25. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. U S Am.* **102**, 15545–15550, https://doi.org/10.1073/pnas.0506580102 (2005).

26. Wei, L. *et al.* Comprehensive analysis of gene-expression profile in chronic obstructive pulmonary disease. *Int. J. chronic Obstr. pulmonary Dis.* **10**, 1103–1109, https://doi.org/10.2147/COPD.S68570 (2015).

27. Odom, D. T. *et al.* Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**, 1378–1381, https://doi.org/10.1126/science.1089769 (2004).

28. Finney, L. J. *et al.* Validity of the diagnosis of pneumonia in hospitalised patients with COPD. *ERJ open research* **5**, https://doi.org/10.1183/23120541.00031-2019 (2019).

29. Pangon, L. *et al.* MCC inhibits beta-catenin transcriptional activity by sequestering DBC1 in the cytoplasm. *Int. J. cancer* **136**, 55–64, https://doi.org/10.1002/ijc.28967 (2015).

30. Wang, R. *et al.* Down-regulation of the canonical Wnt beta-catenin pathway in the airway epithelium of healthy smokers and smokers with COPD. *PLoS one* **6**, e14793, https://doi.org/10.1371/journal.pone.0014793 (2011).

31. Reddy, P. K. *et al.* Interferon-gamma upregulates MUC1 expression in haematopoietic and epithelial cancer cell lines, an effect associated with MUC1 mRNA induction. *Eur. J. Cancer* **39**, 397–404 (2003).

32. Rubins, J. B. & Pomeroy, C. Role of gamma interferon in the pathogenesis of bacteremic pneumococcal pneumonia. *Infect. Immun.* **65**, 2975–2977 (1997).

33. Christenson, S. A. *et al.* An airway epithelial IL-17A response signature identifies a steroid-unresponsive COPD patient subgroup. *J. Clin. investigation* **129**, 169–181, https://doi.org/10.1172/JCI121087 (2019).

34. Xu, Z. *et al.* miR-365, a novel negative regulator of interleukin-6 gene expression, is cooperatively regulated by Sp1 and NF-kappaB. *J. Biol. Chem.* **286**, 21401–21412, https://doi.org/10.1074/jbc.M110.198630 (2011).

35. Jiang, Z. *et al.* Genetic Control of Fatty Acid beta-Oxidation in Chronic Obstructive Pulmonary Disease. *Am. J. respiratory Cell Mol. Biol.* **56**, 738–748, https://doi.org/10.1165/rcmb.2016-0282OC (2017).

36. Koeppen, K., Coutermarsh, B. A., Madden, D. R. & Stanton, B. A. Serum- and glucocorticoid-induced protein kinase 1 (SGK1) increases the cystic fibrosis transmembrane conductance regulator (CFTR) in airway epithelial cells by phosphorylating Shank2E protein. *J. Biol. Chem.* **289**, 17142–17150, https://doi.org/10.1074/jbc.M114.555599 (2014).

37. Rab, A. *et al.* Cigarette smoke and CFTR: implications in the pathogenesis of COPD. *Am. J. Physiol. Lung Cell. Mol. physiology* **305**, L530–541, https://doi.org/10.1152/ajplung.00039.2013 (2013).

38. Tee, A. R., Blenis, J. & Proud, C. G. Analysis of mTOR signaling by the small G-proteins, Rheb and RhebL1. *FEBS Lett.* **579**, 4763–4768, https://doi.org/10.1016/j.febslet.2005.07.054 (2005).

39. Mori, S. *et al.* The mTOR pathway controls cell proliferation by regulating the FoxO3a transcription factor via SGK1 kinase. *PLoS one* **9**, e88891, https://doi.org/10.1371/journal.pone.0088891 (2014).

40. Faiz, A. *et al.* Effect of long-term corticosteroid treatment on microRNA and gene-expression profiles in COPD. *The European respiratory journal* **53**, https://doi.org/10.1183/13993003.01202-2018 (2019).

41. Esquinas, C. *et al.* Gene and miRNA expression profiles in PBMCs from patients with severe and mild emphysema and PiZZ alpha1-antitrypsin deficiency. *Int. J. chronic Obstr. pulmonary Dis.* **12**, 3381–3390, https://doi.org/10.2147/COPD.S145445 (2017).

42. Faner, R. *et al.* Do sputum or circulating blood samples reflect the pulmonary transcriptomic differences of COPD patients? A multi-tissue transcriptomic network META-analysis. *Respiratory Res.* **20**, 5, https://doi.org/10.1186/s12931-018-0965-y (2019).

43. Obeidat, M. *et al.* Network-based analysis reveals novel gene signatures in peripheral blood of patients with chronic obstructive pulmonary disease. *Respiratory Res.* **18**, 72, https://doi.org/10.1186/s12931-017-0558-1 (2017).

44. Sibila, O. & Restrepo, M. I. Biomarkers in community-acquired pneumonia: still searching for the one. *The European respiratory journal* **53**, https://doi.org/10.1183/13993003.02469-2018 (2019).

Acknowledgements

Author contributions

B.S., S.H., N.S. and J.V. conceptualized the study. W.B., K.G., M.H., K.S. and A.S.S. performed the experiments. B.S., C.F.V., T.G. and A.K. provided resources and administered the project. W.B., F.F. and J.W. curated the data and bioinformatically exploited the data. B.S. and J.V. acquired funding. W.B. wrote the original draft and designed the figures. W.B., F.F., J.W., T.G. and B.S. reviewed and edited the original draft. B.S. supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-57108-0.

Correspondence and requests for materials should be addressed to B.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.