


SCIENTIFIC REPORTS



OPEN

Detection of Cystic Fibrosis Serological Biomarkers Using a T7 Phage Display Library

Harvinder Talwar¹, Samer Najeeb Hanoudi², Andreea Geamanu¹, Dana Kissner¹, Sorin Draghici^{2,3} & Lobelia Samavati^{1,4} 

Cystic fibrosis (CF) is an autosomal recessive disorder affecting the cystic fibrosis transmembrane conductance regulator (CFTR). CF is characterized by repeated lung infections leading to respiratory failure. Using a high-throughput method, we developed a T7 phage display cDNA library derived from mRNA isolated from bronchoalveolar lavage (BAL) cells and leukocytes of sarcoidosis patients. This library was biopanned to obtain 1070 potential antigens. A microarray platform was constructed and immunoscreened with sera from healthy (n = 49), lung cancer (LC) (n = 31) and CF (n = 31) subjects. We built 1,000 naïve Bayes models on the training sets. We selected the top 20 frequently significant clones ranked with student t-test discriminating CF antigens from healthy controls and LC at a False Discovery Rate (FDR) < 0.01. The performances of the models were validated on an independent validation set. The mean of the area under the receiver operating characteristic (ROC) curve for the classifiers was 0.973 with a sensitivity of 0.999 and specificity of 0.959. Finally, we identified CF specific clones that correlate highly with sweat chloride test, BMI, and FEV1% predicted values. For the first time, we show that CF specific serological biomarkers can be identified through immunoscreenings of a T7 phage display library with high accuracy, which may have utility in development of molecular therapy.

There is a tremendous need for developing reliable serum based biomarkers in various diseases including proliferative disorder such as cancer, inflammatory diseases and infections as well as genetic disorders such as cystic fibrosis (CF).

Cystic fibrosis is an autosomal recessive disease caused by mutations in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR)¹. Currently, there are more than 1300 various mutations in CFTR gene that is known to cause the CF phenotype. The CF phenotype is characterized by chronic bacterial airway infections, neutrophilic inflammation with mucus in airways, progressive bronchiectasis and advanced cystic fibrosis lung disease. Mutations in the CFTR gene affect the epithelial innate immune function in the lungs, resulting in exaggerated and ineffective airway inflammation that fails to eradicate pulmonary pathogens². Bacterial infections in CF are characterized by organisms that have substantial genetic flexibility to evade phagocytic clearance and develop resistance to multiple antibiotics². Repeated or chronic microbial infections are thought to be the major contributor to excessive inflammation leading to CF lung damage. In addition to chronic lung infections, CF subjects may exhibit exocrine pancreatic insufficiency, diabetes mellitus, and sexual organ dysfunction.

Circulating autoantibodies and autoantigens in CF sera have been widely reported, yet their significance is unknown³⁻⁵. Various proteins and protein degradation products have been explored as candidate biomarkers for clinical outcome, such as neutrophil elastase, IL-8⁶, and degradation products of lung surfactant protein SP-A⁷⁻¹⁰. A variety of proteomic approaches exploited antigenic biomarkers that could provide candidates for the diagnosis of infection, prognostic indicators or vaccine development. Pedersen et al. used antibodies from CF patients to probe a protein array of body fluids prepared by two-dimensional gel electrophoresis for antigenic biomarker detection in *Pseudomonas aeruginosa*⁵. Others identified the outer membrane protein OprL as a seromarker for the initial diagnosis of *Pseudomonas aeruginosa* infection in CF patients¹¹.

¹Department of Medicine, Division of Pulmonary, Critical Care and Sleep Medicine, Wayne State University School of Medicine and Detroit Medical Center, Detroit, MI, 48201, USA. ²Department of Computer Science, Wayne State University, 540 E, Canfield, Detroit, MI, 48201, USA. ³Department of Obstetrics and Gynecology, Wayne State University, 540 E, Canfield, Detroit, MI, 48201, USA. ⁴Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, 540 E, Canfield, Detroit, MI, 48201, USA. Correspondence and requests for materials should be addressed to L.S. (email: lsamavat@med.wayne.edu)

Received: 23 May 2017

Accepted: 19 November 2017

Published online: 18 December 2017

Characteristic	Control Subjects	CF Patients	Lung Cancer Patients
Age (Mean \pm SEM)	45.3 \pm 11.5	31.7 \pm 10.8	62.3 \pm 11.9
Gender, N (%)			
Male	7 (14)	20 (65)	
Female	42 (86)	11 (35)	31 (100)
Race, N (%)			
African American	1 (3)	1 (3)	
Caucasian	25 (81)	30 (97)	31 (100)
Other	5 (16)	NA	
Smoker (>50 packs/yr.), N (%)	15 (30)	NA	6 (20)
BMI (Mean \pm SEM)	28 \pm 3.6	22.76 \pm 0.61	24 \pm 4.6
Sweat Chloride values (mM/L)	NA	103.31 \pm 13.5	NA
PFT Values (Mean \pm SEM)			
FEV1 (% predicted)	NA	59.30 \pm 4.90	NA
FVC (% predicted)	NA	75.27 \pm 4.49	NA
TLC (% predicted)	NA	101.47 \pm 2.42	NA
DLCO (% predicted)	NA	89.77 \pm 3.90	NA
Gene Mutation			
*Homogygous (Double mutation at F508 del)	NA	15 (48)	NA
**Heterogygous (Double mutation one with F508 del)	NA	9 (29)	NA
***Other (Double mutation with none at F508 del)	NA	7 (23)	NA
Bacterial culture results			
Pseudomonas (mucoid & non-mucoid)	NA	21 (67)	NA
Staphylococcus aureus	NA	13 (42)	NA
Aspergillus	NA	3 (1)	NA
Adenocarcinoma of Lung, N (%)	NA	NA	31 (100)

Table 1. Subjects demographics. Age values are presented as means and variability in SD. N = Number of patients and percent shown in parentheses.

Recently, we developed a heterologous cDNA library derived from bronchoalveolar cells (BAL) and total white blood cells (WBC) from sarcoidosis patients and combined it with cultured human monocytes and embryonic lung fibroblasts cDNA libraries to build a complex sarcoidosis library (CSL)^{12,13}. Because the CSL represents a segment of the human lung microbiome, we hypothesize that it contains potential antigens relevant to CF. To test this, we immunoscreened our microarray platform with sera from healthy controls, CF and lung cancer patients using the power of antibody recognition present in human sera to discover potential serological biomarkers in CF.

Results

Complex sarcoidosis library detects unique antigens in the CF sera. A panel of potential antigens was randomly selected from two highly enriched pools of T7 phage cDNA libraries through biopanning of the CSL library¹². A microarray platform was constructed and immunoscreened with 111 sera (49 healthy controls, 31 with CF and 31 with adenocarcinoma (LC) of the lungs). The demographics of the study subjects are shown in Table 1. Among the CF patients, 15 (48%) were genotyped as F508del homozygotes, 9 (29%) were heterozygotes for F508del, and 7 (23%) had various mutations such as G542X or 2789 + 5 GT0A/S489X and others (Table 1). Following immunoreaction, the microarray data were pre-processed and then analyzed. We applied a student *t*-test on 1,000 training sets (FDR < 0.01) between CF vs. healthy controls samples. A total of 599 clones appeared significant at least once. We calculated the frequency of each significant clone and ranked the top 20 clones according to their significance and frequency. Furthermore, we performed an unsupervised PCA for all 1070 clones with data from 111 study subject sera. As shown in Fig. 1a, several LC and healthy controls clustered together with the CF samples. To investigate whether the identified 20 highly significant CF clones can improve class separation of CF samples from LC and healthy controls, we constructed a PCA plot using only those clones (Fig. 1c). Using the 20 highly significant CF clones aided to a class separation of CF samples from LC and healthy controls. Forty nine percent of variance was explained along the PC1.

Next, we performed unsupervised hierarchical clustering with all 1070 clones on 111 samples. We observed that the magenta cluster has a mix of samples and lacks specific sub-clusters of CF samples (Fig. 1b). In contrast, when the clustering algorithm was performed using the 20 highly significant CF clones on all samples, we observed a distinct hierarchical linkage, clearly demarcating CF samples from others (healthy controls and LC) (Fig. 1d). Distinct expression features of 20 highly significant CF clones among study subjects are highlighted in a heatmap plot (Fig. 2).

Next, we applied the classifier model and calculated the AUC values on accumulating numbers of clones (see method section) on test and validation sets. Figure 3a shows the AUC values for the test set. The lowest average AUC values for the test set was 0.956. Figure 3b graphically represents the performance of the classifier model when applied to the validation set. When we applied the classifier model on the validation set, the lowest average

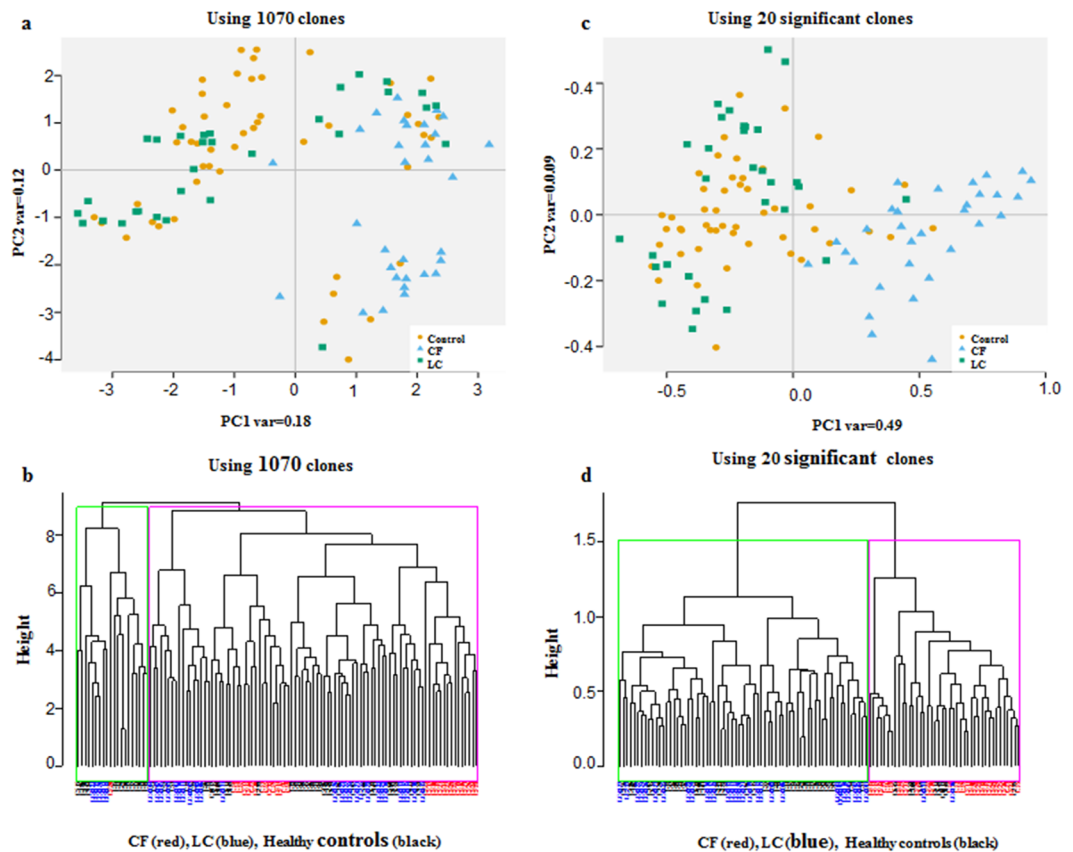


Figure 1. PCA and hierarchical clustering. **(a)** PCA score plots along the PC1 and 2 generated with 1070 clones of three groups: 1) healthy control samples (yellow circle), 2) CF samples (blue triangle) and 3) LC samples (green square). Along the PCA1 explaining a variance of only 0.18 and along the PC2 of 0.12. **(b)** The hierarchical clustering was applied on the healthy controls (black labels), CF patients (red labels) and LC (blue labels) with 1070 clones. **(c)** PCA score plots along the PC1 and 2 results when applied on the highly significant 20 CF clones. The PC1 explained 0.49 of variance, whereas PC2 explained 0.09 of variance. As shown the CF samples are well separated from the healthy controls and LC samples. **(d)** Hierarchical clustering using only the highly significant 20 CF clones. The green cluster includes LC and healthy control samples (no CF samples), the magenta cluster includes all the CF samples, few healthy control and two LC samples. This figure demonstrates better clustering with the highly significant 20 CF clones (panels c and d) when compared with the clustering using all clones (panels a and b).

AUC value was 0.926. These results clearly indicate that the classification model based on the accumulating number of significant clones when applied on the test and the validation sets have a very good classification performance. Finally, to assess if the identified highly significant CF clones provide a sound classification performance, we applied the naïve Bayes classification algorithm with the highly significant CF clones to predict CF samples from healthy controls and LC samples. At the optimal threshold (highest true positivity with lowest false positivity for each of the 1000 runs), we could reliably predict CF from healthy controls and LC samples with a mean specificity of 0.959 (95% CI, 0.11–0.15) and a mean sensitivity of 0.999 (95% CI, 0.18–0.21). The mean AUC under the ROC for the classifier was 0.973 (95% CI, 0.07–0.094) (Fig. 3c).

Characterization of significant CF clones. Based on the results of training and validation sets, we characterized the 20 highly performing clones through sequencing and identified which clones can predict sweat chloride tests, FEV1% predicted and body mass index (BMI). After obtaining the sequences of clones, ExPasy program was used to translate the cDNA sequences to protein sequences¹². Protein blast using algorithms of the BLAST program were applied to identify the highest homology to identified peptides. Additionally, we compared these results with corresponding nucleotide sequences using nucleotide blast and determined the predicted amino acid in frame with phage T7 10 B gene capsid proteins. Among 20 clones four CF reactive antigens comprise relatively large peptides, while 16 CF antigens are coded by the inserted gene fragments leading to out-of-frame-peptides, hereby meeting the definition of mimotopes¹⁴ (Table 2). As CF sera reacted to these out-of-frame-peptides, it is likely that these clones represent CF antigens that are produced as a result of altered reading frames or alternative splicing, as shown in previous studies^{14,15}. Full length of peptides and genes of the top 20 CF clones are shown in Supplementary Table 1. Table 2 shows the 14 most significant CF antigens, gene names, sensitivity, specificity and FDR adjusted p-value. Figure 4a and b show the ROC curves for the 14 CF antigens. Finally, we sought to determine whether any of the biomarkers correlate with sweat chloride test, BMI and FEV1% predicted values.

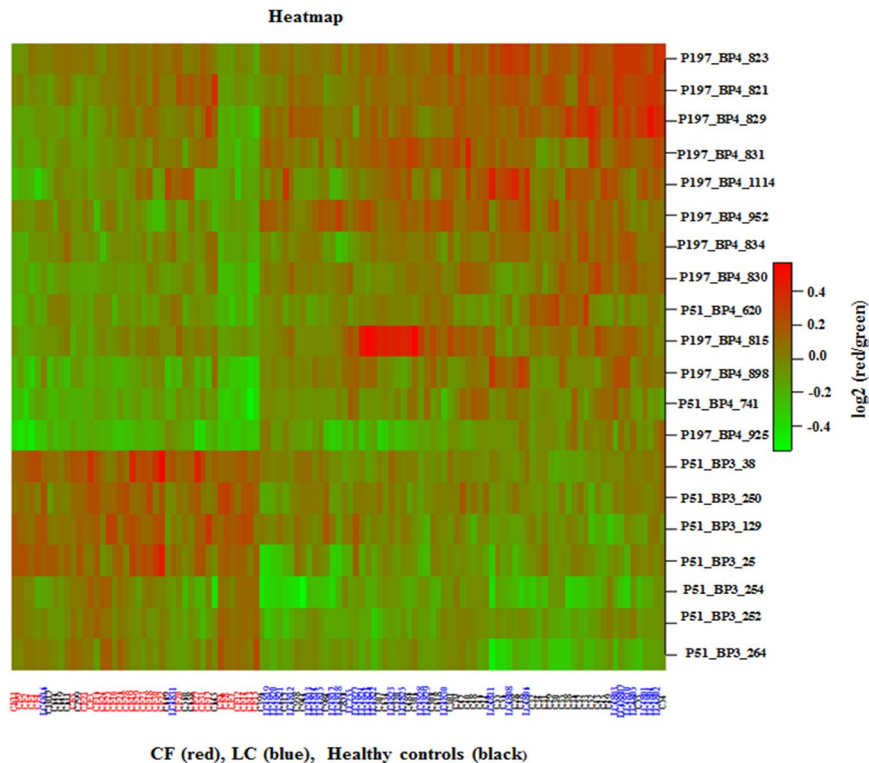


Figure 2. Heatmap generated based on the 20 highly significant CF clones from the data of 111 study subjects (49 healthy controls, 31 with CF and 31 with LC). Each row represents a clone, while each column represents a study subject. As shown in Fig. 2, most CF samples clustered to the left side of the heat map plot, while the LC samples and healthy controls clustered to the right side of the plot indicating different expression profiles.

Sweat chloride test, PFT and BMI values for CF subjects are shown in Table 1. Sweat chloride test is commonly used as screening tool for CF diagnosis¹⁶. We found highest spearman correlation ($r = -0.54$) between sweat chloride values and the clone p51_BP3_113 (GEM_5047) (Fig. 5a). By combining this clone with four additional clones a higher correlation was reached ($r = -0.72$) (Fig. 5b). BMI is an important clinical measure among CF patients to predict exacerbation and decline of lung function testing¹⁷. We found highest spearman correlation ($r = -0.31$) between BMI and the P51_BP3_47 clone (dnaJ homolog) (Fig. 5c). By combining this clone with 4 other clones a higher correlation with BMI was reached ($r = -0.58$) (Fig. 5d). Additionally, we found the highest correlation ($r = -0.42$) between FEV1% predicted and clone P197_BP4_926 (Fig. 5e). The correlation value ($r = -0.6$) improved once we added 4 other clones (Fig. 5f). Table 3 shows the correlation between sweat chloride values, BMI and FEV1% predicted values and significant clones. Seven out of 16 identified clones overlapped with highly specific and sensitive CF clones shown in Table 2. In addition, we identified 6 other clones with significant correlation with sweat chloride test, BMI and FEV1% predicted values. Similar results were observed when we plotted other PFT values including FVC (data not shown).

Discussion

CF is characterized by a self-perpetuating cycle of airway obstruction, chronic bacterial infection, and vigorous inflammation that results in bronchiectasis, progressive obstructive lung disease, and marked shortening of life expectancy. Despite having identical cystic fibrosis transmembrane conductance regulator genotypes, individuals with F508del homozygous CF demonstrate significant variability in severity of pulmonary disease and infection. Non-invasive serological biomarkers that can aid to monitor disease progression or evaluate response to therapy would be extremely valuable. Several groups attempted to identify specific biomarkers to predict inflammation in CF using various biofluids such as sputum, BAL and serum^{10,18}. Most of these methods led to the discovery of a series of markers or expression signatures but failed to be useful in clinical practice¹⁸. In view of this background, we applied a novel high throughput technology to overcome the current gap by constructing phage-protein microarrays in which peptides were derived from a unique sarcoidosis cDNA library and expressed as a phage fusion protein. Through immunoscreening and rigorous statistical analysis, we identified 20 highly significant CF clones as biomarkers that are able to discriminate between CF and healthy controls as well as lung cancer sera. One important issue in biomarker discovery is the validation of biomarkers and sample selection. To overcome this issue, we randomly assigned samples into 1000 training sets instead of using one training set. Then, we compared the healthy controls and CF samples for each pair of such random sets. The ranking of the top 20 clones was based on the significance and frequency of each clone (how many times each clone appears significant at $FDR < 0.01$).

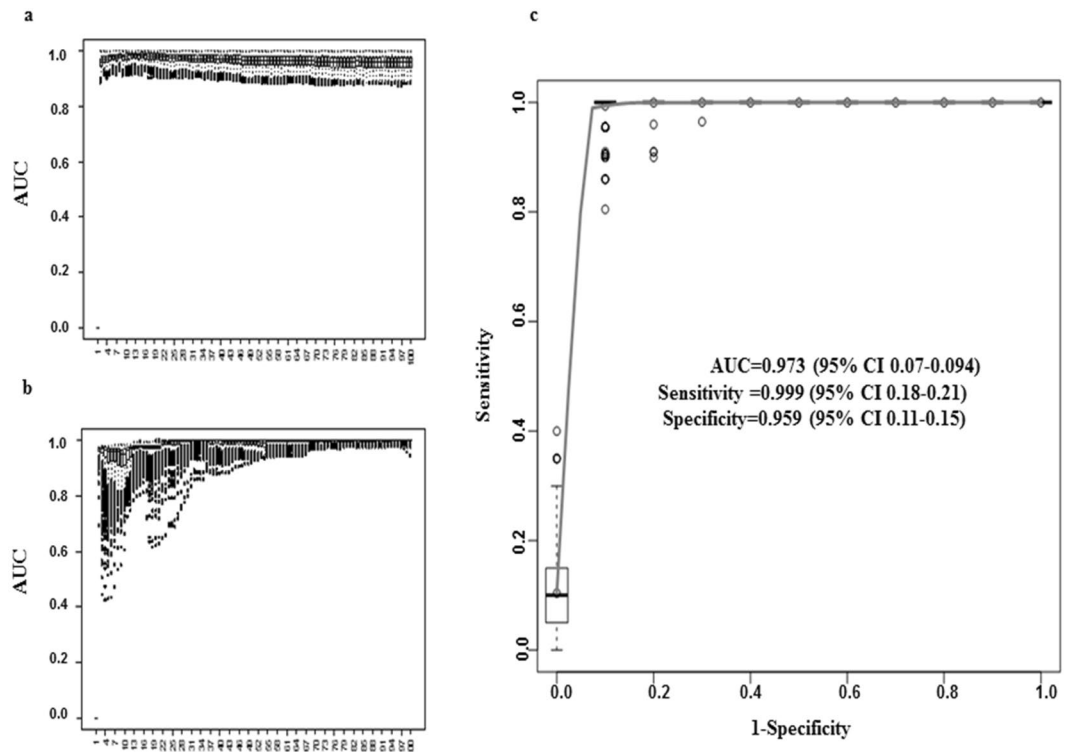


Figure 3. Classification performance of the naïve Bayes classifier. The classifier is to predict CF from LC and healthy control samples. **(a)** Performance of the classifier on the testing sets. Box plots indicate the AUC values (y-axis) when the classifier model was applied on the 1000 test sets. The x-axis is accumulating sets of clones. The accumulation of the clones starts with the most frequent clone and then one clone added at a time to reach 100 clones. **(b)** Performance of the classifier models on the validation set. As indicated the classifier models when they were built using the significant clones shows a high AUC values on the testing sets as well as on the completely independent validating set. **(c)** The ROCs generated from the average of the 1000 runs of the classifier models when applied on the validation set (randomly selected healthy controls, CF and LC) using the 20 highly significant CF clones. The box plot shows the distribution of the sensitivities. The ROC curve demonstrates an excellent classification performance with an average AUC of 0.973 (95% CI: 0.07–0.094) with sensitivity of 0.99 (95% CI: 0.18–0.21) and specificity of 0.959 (95% CI: 0.11–0.15). These results indicate excellent performance of the naïve Bayes classifier on the 20 highly significant CF clones.

Environmental stresses including cigarette smoking, hypoxia, and chronic inflammation have also been implicated in reduced CFTR function^{19,20}. Additionally, subjects with smoking related chronic obstructive lung disease (COPD) can develop a similar clinical phenotype with recurrent respiratory infections, mucus inspissation and airway obstruction that is attributed to acquired CFTR deficiency²¹. In unsupervised HC, we have seen few false positive classification that might have been due to the selection of control groups (lung cancer and healthy controls), who had a significant smoking history. Interestingly, these subjects had more than 50 pack years of smoking history. Because it is known that cigarette smoking affects bacterial clearance²², we speculate that long-term cigarette smoking in these subjects might have led to a similar immunoreactivity to our microarray platform as CF individuals. Therefore, it is probable that if we choose a younger non-smoker group as control subjects, we would have no false positive classification.

Furthermore, we sequenced the top 20 discriminating antigens for CF and identified homologies in a public database. The range length of identified peptides for CF antigens was between 8–213 amino acids (AA). Among the 20 CF specific phage peptides, five out-of-frame peptides and one epitope were increased in sera of CF patients. One epitope (HLA-DR) was three times randomly selected (P51BP3_296, P51BP4_704 and P197_BP4_925), suggesting the importance of HLA-DR in pathology of CF. Recently, studies have demonstrated that the transcript levels of HLA-DR and HLA-DQ are reduced in CF patients²³. Another epitope was DnaJ homologue (Hdj)-1/heat shock protein (Hsp) 40, a protein chaperon, which along with its co-chaperone Hsp70 regulates protein folding and trafficking in the endoplasmic reticulum (ER) and facilitates degradation of misfolded proteins²⁴. It has been shown that Hsp40 and Hsp70 facilitate CFTR assembly²⁵. We found DnaJ homologue was increased in sera of CF patients and had a negative correlation with BMI of CF subjects. Another epitope (Thioredoxin like protein) was decreased in CF patients. Studies have shown that excessive neutrophil elastase activity in the airways of pediatric and adult CF patients resulted in lung damage^{26–28}. Disruption of neutrophil elastase activity by adding exogenous thioredoxin or dihydrolipoic acid in the sputum of CF patients reduced the neutrophil elastase activity²⁹. Another in-frame epitope with relevance to FEV1% predicted was Thymosin β -4

Clone	Increased in Cystic Fibrosis vs Healthy Controls	NCBI Protein Number	p value	FDR corrected p Value	AUC 95% CI	Sensitivity, 95% CI	Specificity, 95% CI
P51_BP3_129	Chain A <i>Pseudomonas Aeruginosa</i> Metap, In Mn Form	4FO8 4FO7 A	0.0069	0.032	0.964 (0.04–0.10)	0.9 (0.21–0.26)	0.97 (0.08–0.14)
P51_BP3_250	Beta-lactamase	GEM_5327 AFQ51711.1	0.0125	0.033	0.908 (0.08–0.14)	0.8 (0.20–0.26)	0.93 (0.10–0.17)
P51_BP3_25	Histidine kinase	narX gb KJS26770.1	0.0017	0.031	0.875 (0.12–0.19)	0.71 (0.22–0.30)	0.95 (0.08–0.17)
P51_BP3_254	Outer membrane Porin	GEM_5047 WP_060349386.1	GEM_5047 0.0056	0.022	0.857 (0.14–0.20)	0.71 (0.26–0.29)	0.93 (0.10–0.16)
P51_BP3_47	dnaJ homolog	DNAJC10 NP_071760.2	0.0006	0.05	0.745 (0.08–0.14)	0.82 (0.20–0.21)	0.74 (0.08–0.14)
P51_BP3_252	γ -glutamyltranspeptidase	PS113–4947 WP_042609239.1	0.0005	0.03	0.727 (0.06–0.11)	0.61 (0.12–0.14)	0.9 (0.11–0.15)
	Decreased in Cystic Fibrosis vs Healthy Controls						
P197_BP4_830	TetR family transcriptional regulator	GEM_1794 WP_043185598.1	3.44E-05	0.032	0.942 (0.06–0.09)	0.9 (0.21–0.22)	0.94 (0.10–0.12)
P197_BP4_898	AraC-family transcriptional regulator	APZ15_34865 WP_059486090.1	0.0001	0.019	0.932 (0.07–0.09)	0.99 (0.20–0.26)	0.71 (0.08–0.16)
P197_BP4_925	HLA-DR alpha	HLA-DR AAO23887.1	0.0076	0.037	0.931 (0.07–0.11)	0.99 (0.20–0.26)	0.81 (0.10–0.15)
P197_BP4_1109	Thioredoxin like protein	TXNL1 CAA09375.1	0.001	0.045	0.924 (0.05–0.11)	0.89 (0.18–0.20)	0.87 (0.10–0.13)
P197_BP4_952	NADH dehydrogenase subunit I	MT-ND1 AFA28546.1	0.0007	0.03	0.895 (0.08–0.20)	0.91 (0.20–0.29)	0.78 (0.15–0.18)
P197_BP4_834	AMP-dependent synthetase	VL15_07170 WP_048244810.1	0.0036	0.034	0.884 (0.10–0.20)	0.99 (0.20–0.26)	0.71 (0.13–0.18)
P197_BP4_1114	Peptide ABC transporter substrate binding protein	135_2059 ALB12327.1	0.0011	0.05	0.845 (0.13–0.17)	0.8 (0.26–0.26)	0.84 (0.14–0.16)
P197_BP4_762	Ketoacyl-ACP reductase	fabG WP_034004244.1	0.0011	0.05	0.826 (0.09–0.14)	0.79 (0.12–0.14)	0.87 (0.11–0.17)

Table 2. Significant Cystic Fibrosis Clones.

(TMSB4X). *In vitro* addition of Thymosin β -4 in the sputum of CF patients decreases the sputum cohesivity by depolymerizing actin³⁰.

Among the 20 sequenced CF specific phage peptides we identified 16 antigens with relatively short out-of-frame peptides meeting the criteria as mimotopes (mimetic sequence of a true epitope)¹⁴. Although the significance of mimotopes is not clear, it has been shown that some out-of-frame peptides can be immunogenic and can activate MHC class I molecules³¹. Due to smaller peptide sequences of mimotopes, they may have homology with diverse proteins. Prior studies using similar techniques have identified out-of-frame peptides^{14,15,32}. We identified two sequenced peptides (narX and barA_4) with similarity to histidine kinases that belong to a large family of membrane-spanning proteins found in many prokaryotes and some eukaryotes. This gene controls the bacterial virulence, growth and biofilm formation in CF patients³³. Similarly, IgG response to *Burkholderia capacia* 80-kDa outer membrane protein has been shown to be significantly higher in patients with CF³⁴. Interestingly, when we explored the correlation of biomarkers with sweat chloride values, we found a good correlation with the outer membrane porin³⁵. Another significant biomarker detected is beta-lactamase. Several studies have shown association between the development of resistance to beta-lactam antibiotics and high beta-lactamase production in CF patients³⁶.

Among 16 mimotopes, we found eight with decreased expression in CF patients (Table 2). Interestingly, one out of eight CF antigens with higher specificity and sensitivity (P197_BP4_830), belongs to repressor transcriptional regulators^{37,38}. One *in-vitro* study showed that *Pseudomonas aeruginosa* toxin regulates TetR family transcriptional regulator and hence regulates CFTR expression through transcriptional repression³⁷. Interestingly, TetR is involved in the regulation of antibiotic resistance and controls the expression of membrane-associated proteins that are involved in antibiotic resistance³⁹. Through immunoscreening, we identified decreased NADPH dehydrogenase subunit I. Similarly, studies have shown that mitochondrial complex I activity is reduced in cells with impaired cystic fibrosis transmembrane conductance regulator⁴⁰. CFTR chloride channels belong to the superfamily of ABC transporter ATPases⁴¹. Interestingly, we identified reduced ABC transporter substrate binding protein expression in CF patients. The ABC transporters are widespread in prokaryotes and eukaryotes containing nucleotide-binding domains (NBD) and two transmembrane domains (TMDs). ATP hydrolysis on the NBD drives conformational changes in the TMD, resulting in alternating access from inside and outside of the cell for unidirectional transport across the lipid bilayer⁴².

To our knowledge no previous study used phage display technology to detect CF serum biomarkers. We detected novel antigens for CF using a heterologous library derived from sarcoidosis subjects. Lungs are highly exposed to numerous bacteria and our library is predominantly derived from sarcoidosis BAL cells and WBCs containing diverse immune cells, including macrophages that were exposed to various pathogens. Hence, we postulate that the CSL represents a segment of the lung microbiome containing diverse antigens including CF specific antigens, sarcoidosis and TB specific antigens^{12,13}. The phage display technology and immunoscreening has utilities not only in identifying of diagnostic biomarkers, but also may enable us to develop a novel targeted therapy utilizing the peptide sequences (mimotopes) as vehicles to deliver specific drugs. For instance, among highly significant clones, we found a sequence peptide homologous to histidine kinase (narX) with high specificity and sensitivity. Bacterial histidine kinases are promising targets for the development of antibacterial therapy. Currently efforts have been made to identify specific compounds targeting the inhibition of histidine kinase as antibacterial therapy⁴³. Additionally, this technology might enable us to discover unknown epitopes targeting specific bacterial

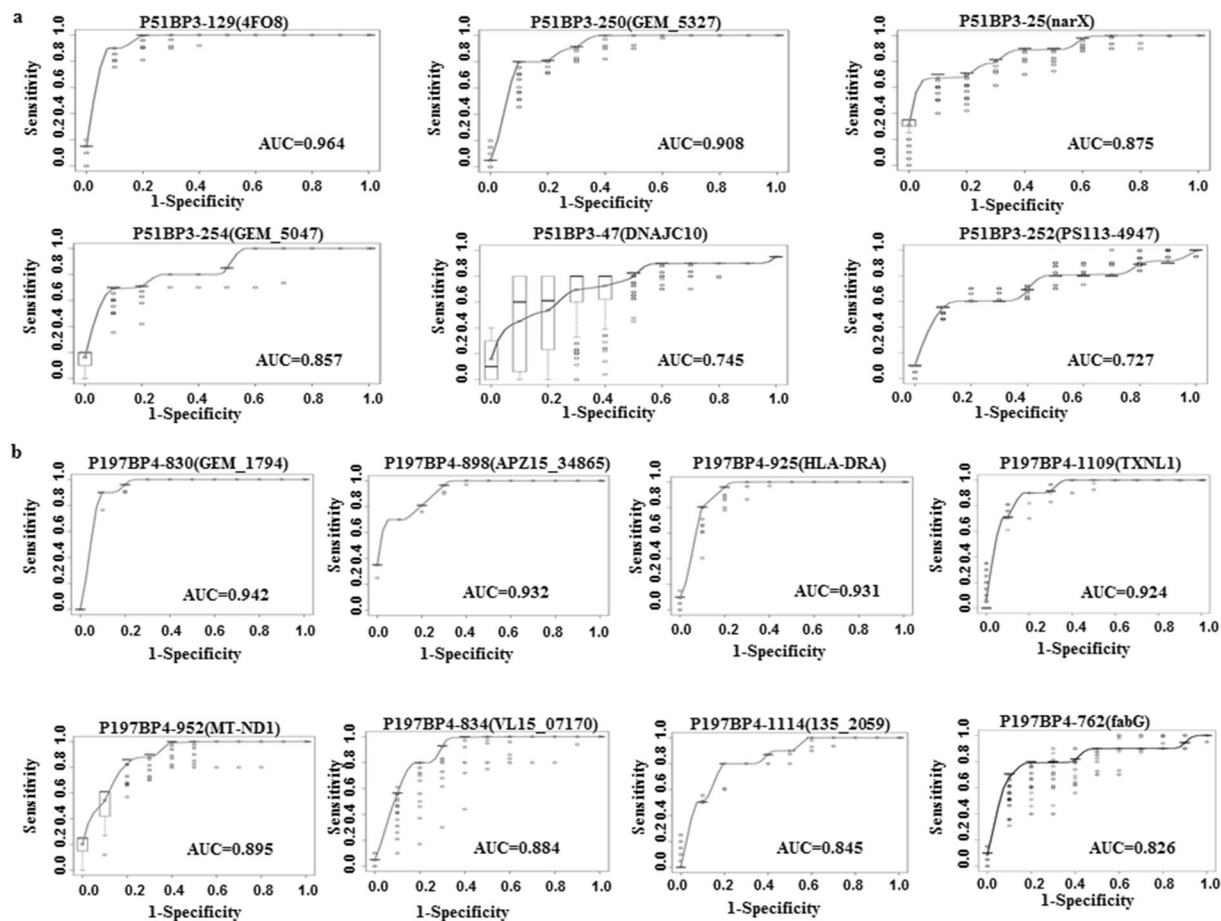


Figure 4. Naïve Bayes classification performance for the top 14 clones. (a) ROCs for the top 6 significant clones that are increased (up-regulated) in CF sera compared to healthy control. (b) ROCs for the top 8 significant clones that are decreased (down-regulated) in CF compared to healthy controls. This figure demonstrates reasonable classification performance when the classification was applied just to one clone.

antigens leading to immunogenicity and antibody production in CF subjects, as well as providing us with a better understanding of host immune defenses in CF subjects. Furthermore, this microarray platform can be hybridized to detect IgA in sera or saliva of CF patients that may have clinical values.

In summary, we have developed a novel T7 phage display library derived from BALs and leukocytes of patients with sarcoidosis that displays a significant segment of the potential antigens that can recognize IgG antibodies in CF sera with high accuracy. Furthermore, we have identified a set of CF clones that highly correlate with clinical measures such as, sweat chloride values, BMI and FEV1. Microarray and immunoscreening has a value in clinical practice in antibody detection as it is non-invasive and requiring a minimal amount of blood. The identified sequences can be used to develop peptide/protein-coated magnetic nanoparticles for clinical testing or for applications in drug delivery⁴⁴. The present study describes a novel approach to identify CF biomarkers. Further studies with a larger cohort group of patients and/or longitudinal studies are needed to investigate the role of these antigens in CF, their mechanism of action and their utilities in drug design and monitoring of therapy.

Materials and Methods

Chemicals. All chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless specified otherwise. LeukoLOCK filters and RNAlater were purchased from Life Technologies (Grand Island, NY). The RNeasy Midi kit was obtained from Qiagen, (Valencia, CA). The T7 mouse monoclonal antibody was purchased from Novagen (San Diego, CA). Alexa Fluor 647 goat anti-human IgG and Alex Fluor goat anti-mouse IgG antibodies were purchased from Life Technologies (Grand Island, NY).

Patient selection. This study was approved by the institutional review board at Wayne State University, the Detroit Medical Center and Cystic Fibrosis Center. Sera collected from 3 groups: 1) healthy volunteers; 2) confirmed CF subjects, and 3) sera from subjects with adenocarcinoma of the lungs. All study subjects signed a written informed consent. All methods were performed in accordance with the human investigation guidelines and regulations by the IRB (protocol Number = 055208MP4E) at Wayne State University.

Pulmonary function tests were performed following ATS guidelines in a licensed laboratory in all patients unless contraindicated⁴⁵. All spirometric studies were performed using a calibrated pneumotachograph and

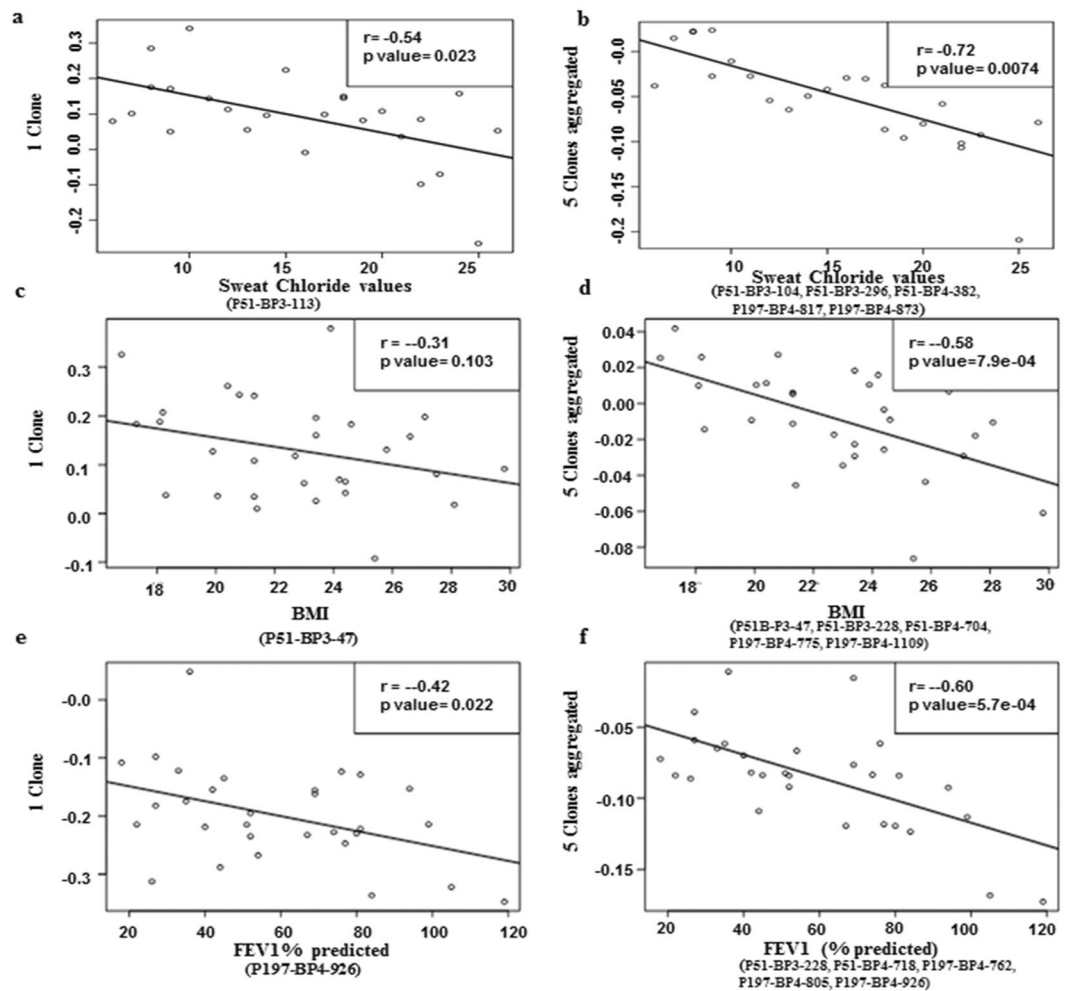


Figure 5. Pearson correlation of identified biomarkers with clinical values. Scatter plots depicted correlation of the sweat chloride values with one clone (a) and aggregated 5 clones (b). Scatter plots depicted the correlation for BMI predicted with one clone (c) and aggregated 5 clones (d). Scatter plots depicted correlation of FEV1% with one clone (e) and aggregated 5 clones (f). The correlation values and p values are shown in the top right of each plot. The names of the clones are shown at the bottom of each plot.

Clinical measure	Single clone	Correlation r	Set of five clones	Gene name	p value	Correlation r
Sweat Chloride values	P51_BP3_113 (p value 0.002, Gene Name: GEM_5047)	-0.54	P51_BP3_104	BamMC406_2945	0.009	-0.72
			P51_BP3_296	HLA-DRA	0.024	
			P51_BP4_382	4FO8	0.017	
			P197_BP4_817	narX	0.00015	
			P197_BP4_873	fabG	0.009	
BMI	P51_BP3_47	-0.31	P51_BP3_47	DNAJC10	0.0006	-0.58
			P51_BP3_228	PA4503	0.0022	
			P51_BP4_704	HLA-DRA	0.004	
			P197_BP4_775	CL-9	0.014	
FEV1 (%Predicted)	P197_BP4_926	-0.42	P197_BP4_1109	TXNL1	0.001	-0.6
			P51_BP3_228	PA4503	0.0022	
			P197_BP4_718	TMSB4X	0.007	
			P197_BP4_762	fabG	0.0011	
			P197_BP4_805	WL94_35745	0.001	
			P197_BP4_926	barA_4	0.015	

Table 3. Correlation of biomarkers with Sweat Chloride test, BMI and FEV1% predicted.

lung volumes were measured in a whole-body plethysmograph (Jaeger Spirometry and SensorMedics Vmax 22, VIASYS Respiratory Care, Inc; Yorba Linda, CA, USA). All CF subjects were ambulatory patients. Sweat chloride test values were obtained from the medical records.

Serum collection. Using standardized phlebotomy procedures blood samples were collected and stored at -80°C ¹².

Construction and Biopanning of T7 phage display cDNA libraries. T7 phage display libraries from BAL, WBC, EL-1 and MRC5 were made to generate a complex sarcoid library (CSL)¹². Differential biopanning for negative selection was performed using sera from healthy controls to remove the non-specific IgG, and sarcoidosis sera for positive enrichment¹².

Microarray construction and immunoscreening. Informative phage clones were randomly picked and amplified after four rounds of biopannings and their lysates were arrayed in quintuplicates onto nitrocellulose FAST slides (Grace Biolabs, OR) using the ProSys 5510TL robot (Cartesian Technologies, CA). The nitrocellulose slides were hybridized with sera and processed as described previously¹².

Sequencing of phage cDNA clones. Individual phage clones were PCR amplified using T7 phage forward primer 5' GTTCTATCCGCAACGTTATGG 3' and reverse primer 5' GGAGGAAAGTCGTTTTTTGGGG 3' and sequenced by Genwiz (South Plainfield, NJ), using T7 phage sequence primer TGCTAAGGACAACGTTATCGG.

Data acquisition and pre-processing. Following the immunoreaction, the microarrays were scanned in an Axon Laboratories 4100 scanner (Palo Alto, CA) using 532 and 647 nm lasers to produce a red (Alexa Fluor 647) and green (Alexa Fluor 532) composite image. Cy5 (red dye) labeled antihuman antibody was used to detect IgGs in human serum that were reactive to peptide clones, and a Cy3 (green dye) labeled antibody was used to detect the phage capsid protein¹². Using the ImaGene 6.0 (Biodiscovery) image analysis software, the binding intensity of each peptide with IgGs in sera was expressed as \log_2 (red/green) fluorescent intensities. These data were pre-processed using the limma package in the R language environment^{46,47} and normexp method was applied to correct the background⁴⁸. Within array normalization was performed using the LOESS method^{48,49}. The scale method was applied to normalize between arrays^{48,49}. Intensity ratio of a clone in CF samples divided by the same clone intensity ratio from healthy control samples were calculated to determine the fold change of a clone.

Statistical analyses. To detect frequently differentially expressed antigens for CF we applied a two-tailed *t*-test. To evaluate the significant CF antigens identified with *t*-test, we applied principal component analysis (PCA), agglomerative hierarchical clustering (HC), heatmap, and naïve Bayes classifier. To avoid the problem of *over-fitting* the classifiers, we randomly split the CF and healthy controls samples into: i) training, ii) test, and iii) validation sets. Out of the 31 CF samples, 21 samples were randomly assigned into training (10 samples) and test (11 samples) sets. We repeated 1000 times random processing to generate 1000 training and test sets. The remaining 10 CF samples were used as an independent validation set. The 1000 training and testing sets for the healthy controls were randomly selected from 33 out of 49 samples (16 training and 17 test set). Therefore, the number of samples for the validation set for healthy controls was 16. While 31 LC samples were randomly split into test (15 samples) and validation (16 samples) sets. For CF clones specific selection, we applied a *t*-test between the 1000 CF training sets vs. 1000 healthy control-training sets. To correct for multiple comparisons, we applied the false discovery rate (FDR) algorithm with a threshold of 0.01 FDR⁵⁰. The frequency of each significant clone (FDR < 0.01) across all 1000 runs was calculated and sorted based on their frequency of occurrence. The top 20 clones were considered highly significant CF clones. We built a naïve Bayes classifier on each of the 1000 training sets and tested the classifier model on the 1000 testing sets. Finally, the classifier model was validated on a complete independent validation set. The range of clones starts with the most frequent clone followed by adding one clone at a time. We constructed the models on training sets and applied the model on testing sets, as well as validation set. Finally, we determined correlation of biomarkers with body mass index (BMI) and % predicted forced expiratory volume (FEV1) of CF patients. We calculated combinations of 5 clones from the top set of markers. For each combination, the aggregated vector was calculated from the mean of 5 clones and Pearson correlation between the aggregated vector and BMI and FEV1% predicted was determined.

References

- Stoltz, D. A., Meyerholz, D. K. & Welsh, M. J. Origins of cystic fibrosis lung disease. *N Engl J Med* **372**, 351–362, <https://doi.org/10.1056/NEJMra1300109> (2015).
- Cohen, T. S. & Prince, A. Cystic fibrosis: a mucosal immunodeficiency syndrome. *Nature medicine* **18**, 509–519, <https://doi.org/10.1038/nm.2715> (2012).
- Carter, C. J. Pathogen and autoantigen homologous regions within the cystic fibrosis transmembrane conductance regulator (CFTR) protein suggest an autoimmune treatable component of cystic fibrosis. *FEMS immunology and medical microbiology* **62**, 197–214, <https://doi.org/10.1111/j.1574-695X.2011.00803.x> (2011).
- Budding, K., van de Graaf, E. A., Hoefnagel, T., Hack, C. E. & Otten, H. G. Anti-BPIFA1/SPLUNC1: a new autoantibody prevalent in patients with endstage cystic fibrosis. *Journal of cystic fibrosis: official journal of the European Cystic Fibrosis Society* **13**, 281–288, <https://doi.org/10.1016/j.jcf.2013.10.005> (2014).
- Pedersen, S. K. *et al.* An immunoproteomic approach for identification of clinical biomarkers for monitoring disease: application to cystic fibrosis. *Molecular & cellular proteomics: MCP* **4**, 1052–1060, <https://doi.org/10.1074/mcp.M400175-MCP200> (2005).
- Mayer-Hamblett, N. *et al.* Association between pulmonary function and sputum biomarkers in cystic fibrosis. *American journal of respiratory and critical care medicine* **175**, 822–828, <https://doi.org/10.1164/rccm.200609-1354OC> (2007).
- von Bredow, C., Birrer, P. & Griese, M. Surfactant protein A and other bronchoalveolar lavage fluid proteins are altered in cystic fibrosis. *Eur Respir J* **17**, 716–722 (2001).

8. Downey, D. G. *et al.* The relationship of clinical and inflammatory markers to outcome in stable patients with cystic fibrosis. *Pediatric pulmonology* **42**, 216–220, <https://doi.org/10.1002/ppul.20553> (2007).
9. Rowe, S. M. *et al.* Potential role of high-mobility group box 1 in cystic fibrosis airway disease. *American journal of respiratory and critical care medicine* **178**, 822–831, <https://doi.org/10.1164/rccm.200712-1894OC> (2008).
10. Sagel, S. D., Chmiel, J. F. & Konstan, M. W. Sputum biomarkers of inflammation in cystic fibrosis lung disease. *Proc Am Thorac Soc* **4**, 406–417, <https://doi.org/10.1513/pats.200703-044BR> (2007).
11. Rao, A. R., Laxova, A., Farrell, P. M. & Barbieri, J. T. Proteomic identification of OprL as a seromarker for initial diagnosis of *Pseudomonas aeruginosa* infection of patients with cystic fibrosis. *Journal of clinical microbiology* **47**, 2483–2488, <https://doi.org/10.1128/JCM.02182-08> (2009).
12. Talwar, H. *et al.* Development of a T7 Phage Display Library to Detect Sarcoidosis and Tuberculosis by a Panel of Novel Antigens. *EBioMedicine* **2**, 341–350, <https://doi.org/10.1016/j.ebiom.2015.03.007> (2015).
13. Talwar, H., Talreja, J. & Samavati, L. T7 Phage Display Library a Promising Strategy to Detect Tuberculosis Specific Biomarkers. *Mycobacterial diseases: tuberculosis & leprosy* **6**, <https://doi.org/10.4172/2161-1068.1000214> (2016).
14. Wang, X. *et al.* Autoantibody signatures in prostate cancer. *N Engl J Med* **353**, 1224–1235, <https://doi.org/10.1056/NEJMoa051931> (2005).
15. Lin, H. S. *et al.* Autoantibody approach for serum-based detection of head and neck cancer. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **16**, 2396–2405, <https://doi.org/10.1158/1055-9965.EPI-07-0318> (2007).
16. Gibson, L. E. & Cooke, R. E. A test for concentration of electrolytes in sweat in cystic fibrosis of the pancreas utilizing pilocarpine by iontophoresis. *Pediatrics* **23**, 545–549 (1959).
17. Sheikh, S., Zemel, B. S., Stallings, V. A., Rubenstein, R. C. & Kelly, A. Body composition and pulmonary function in cystic fibrosis. *Frontiers in pediatrics* **2**, 33, <https://doi.org/10.3389/fped.2014.00033> (2014).
18. Srivastava, M. *et al.* Serum proteomic signature for cystic fibrosis using an antibody microarray platform. *Molecular genetics and metabolism* **87**, 303–310, <https://doi.org/10.1016/j.ymgme.2005.10.021> (2006).
19. Rab, A. *et al.* Cigarette smoke and CFTR: implications in the pathogenesis of COPD. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **305**, L530–L541, <https://doi.org/10.1152/ajplung.00039.2013> (2013).
20. Cantin, A. M. Cystic Fibrosis Transmembrane Conductance Regulator. Implications in Cystic Fibrosis and Chronic Obstructive Pulmonary Disease. *Annals of the American Thoracic Society* **13**(Suppl 2), S150–155, <https://doi.org/10.1513/AnnalsATS.201509-588KV> (2016).
21. Solomon, G. M., Raju, S. V., Dransfield, M. T. & Rowe, S. M. Therapeutic Approaches to Acquired Cystic Fibrosis Transmembrane Conductance Regulator Dysfunction in Chronic Bronchitis. *Annals of the American Thoracic Society* **13**, S169–S176, <https://doi.org/10.1513/AnnalsATS.201509-601KV> (2016).
22. Lovewell, R. R., Patankar, Y. R. & Berwin, B. Mechanisms of phagocytosis and host clearance of *Pseudomonas aeruginosa*. *American journal of physiology. Lung cellular and molecular physiology* **306**, L591–603, <https://doi.org/10.1152/ajplung.00335.2013> (2014).
23. Hofer, T. P. *et al.* Decreased expression of HLA-DQ and HLA-DR on cells of the monocytic lineage in cystic fibrosis. *Journal of molecular medicine* **92**, 1293–1304, <https://doi.org/10.1007/s00109-014-1200-z> (2014).
24. Stolz, A. & Wolf, D. H. Endoplasmic reticulum associated protein degradation: a chaperone assisted journey to hell. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1803**, 694–705, <https://doi.org/10.1016/j.bbamcr.2010.02.005> (2010).
25. Meacham, G. C. *et al.* The Hdj-2/Hsc70 chaperone pair facilitates early steps in CFTR biogenesis. *The EMBO Journal* **18**, 1492–1505, <https://doi.org/10.1093/emboj/18.6.1492> (1999).
26. Sly, P. D. *et al.* Risk factors for bronchiectasis in children with cystic fibrosis. *New England Journal of Medicine* **368**, 1963–1970, <https://doi.org/10.1056/NEJMoa1301725> (2013).
27. DeBoer, E. M. *et al.* Automated CT scan scores of bronchiectasis and air trapping in cystic fibrosis. *CHEST Journal* **145**, 593–603, <https://doi.org/10.1378/chest.13-0588> (2014).
28. Liu, H., Lazarus, S. C., Caughey, G. H. & Fahy, J. V. Neutrophil elastase and elastase-rich cystic fibrosis sputum degranulate human eosinophils *in vitro*. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **276**, L28–L34 (1999).
29. Lee, R. L. *et al.* Thioredoxin and dihydrolipoic acid inhibit elastase activity in cystic fibrosis sputum. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **289**, L875–L882, <https://doi.org/10.1152/ajplung.00103.2005> (2005).
30. Rubin, B. K., Kater, A. P. & Goldstein, A. L. Thymosin β 4 sequesters actin in cystic fibrosis sputum and decreases sputum cohesivity *in vitro*. *CHEST Journal* **130**, 1433–1440, <https://doi.org/10.1378/chest.130.5.1433> (2006).
31. Schirmbeck, R. *et al.* Translation from cryptic reading frames of DNA vaccines generates an extended repertoire of immunogenic, MHC class I-restricted epitopes. *J Immunol* **174**, 4647–4656, <https://doi.org/10.4049/jimmunol.174.8.4647> (2005).
32. Chatterjee, M. *et al.* Diagnostic markers of ovarian cancer by high-throughput antigen cloning and detection on arrays. *Cancer research* **66**, 1181–1190, <https://doi.org/10.1158/0008-5472.CAN-04-2962> (2006).
33. Worthington, R. J., Richards, J. J. & Melander, C. Small molecule control of bacterial biofilms. *Organic & biomolecular chemistry* **10**, 7457–7474, <https://doi.org/10.1039/c2ob25835h> (2012).
34. Lacy, D. E. *et al.* Serum IgG response to an outer membrane porin protein of *Burkholderia cepacia* in patients with cystic fibrosis. *FEMS immunology and medical microbiology* **17**, 87–94 (1997).
35. Aronoff, S. C. Outer membrane permeability in *Pseudomonas cepacia*: diminished porin content in a beta-lactam-resistant mutant and in resistant cystic fibrosis isolates. *Antimicrobial agents and chemotherapy* **32**, 1636–1639 (1988).
36. Ciofu, O. *Pseudomonas aeruginosa* chromosomal beta-lactamase in patients with cystic fibrosis and chronic lung infection. Mechanism of antibiotic resistance and target of the humoral immune response. *APMIS. Supplementum*, 1–47 (2003).
37. MacEachran, D. P., Stanton, B. A. & O'Toole, G. A. Cif is negatively regulated by the TetR family repressor CifR. *Infection and immunity* **76**, 3197–3206, <https://doi.org/10.1128/IAI.00305-08> (2008).
38. Mahenthalingam, E., Simpson, D. A. & Speert, D. P. Identification and characterization of a novel DNA marker associated with epidemic *Burkholderia cepacia* strains recovered from patients with cystic fibrosis. *Journal of clinical microbiology* **35**, 808–816 (1997).
39. Cuthbertson, L. & Nodwell, J. R. The TetR family of regulators. *Microbiology and molecular biology reviews: MMBR* **77**, 440–475, <https://doi.org/10.1128/MMBR.00018-13> (2013).
40. Valdivieso, A. G. *et al.* The mitochondrial complex I activity is reduced in cells with impaired cystic fibrosis transmembrane conductance regulator (CFTR) function. *PLoS One* **7**, e48059, <https://doi.org/10.1371/journal.pone.0048059> (2012).
41. Schneider, E. & Hunke, S. ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS microbiology reviews* **22**, 1–20, <https://doi.org/10.1111/j.1574-6976.1998.tb00358.x> (1998).
42. Gadsby, D. C., Vergani, P. & Csányi, L. The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* **440**, 477–483, <https://doi.org/10.1038/nature04712> (2006).
43. Bem, A. E. *et al.* Bacterial histidine kinases as novel antibacterial drug targets. *ACS chemical biology* **10**, 213–224, <https://doi.org/10.1021/cb5007135> (2014).
44. Rana, S., Bajaj, A., Mout, R. & Rotello, V. M. Monolayer coated gold nanoparticles for delivery applications. *Advanced drug delivery reviews* **64**, 200–216, <https://doi.org/10.1016/j.addr.2011.08.006> (2012).

45. Raghu, G. *et al.* An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *American journal of respiratory and critical care medicine* **183**, 788–824, <https://doi.org/10.1164/rccm.2009-040GL> (2011).
46. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, <https://doi.org/10.1093/nar/gkv007> (2015).
47. R: A language and environment for statistical computing, (R Foundation for Statistical Computing, Vienna, Austria., 2015).
48. Ritchie, M. E. *et al.* A comparison of background correction methods for two-colour microarrays. *Bioinformatics* **23**, 2700–2707, <https://doi.org/10.1093/bioinformatics/btm412> (2007).
49. Yang, Y. H. *et al.* Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic acids research* **30**, e15, <https://doi.org/10.1093/nar/30.4.e15> (2002).
50. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300, [https://doi.org/10.1016/0306-9877\(95\)90228-7](https://doi.org/10.1016/0306-9877(95)90228-7) (1995).

Acknowledgements

We thank all patients and healthy volunteers for their participation in this study. This project was funded by NIH grant R21HL104481-01A1 awarded to L.S. and with the support of the Department of Medicine, Wayne State University. We would like to thank Drs. Michael A. Tainsky and Nancy Levin for providing lung cancer samples as well as healthy controls and for his invaluable assistance in completing this study. This work has been partially supported by the following grants: NIH R01 DK089167, NIH STTR R42GM087013, NSF DBI-0965741 (to Sorin Draghici), and by the Robert J. Sokol Endowment in Systems Biology.

Author Contributions

Harvinder Talwar contributed to the sample processing, conducted the analysis. Samer Hanoudi performed the preprocessing, the processing of the data and the statistical analysis. Andreea Geamanu enrolled the patients, obtained consents and collected the clinical data. Sorin Draghici supervised the data analysis and contributed to the writing of the manuscript. Dana Kissner provided access to patients with CF. Lobelia Samavati conceived and designed the study, participated in all areas of the research such as patient selection and oversaw patient enrollment, data analysis and writing of the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-18041-2>.

Competing Interests: The authors declare that they have no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017