Title

Microbes Are Associated with Host Innate Immune Response in Idiopathic Pulmonary Fibrosis

Journal

American Journal of Respiratory and Critical Care Medicine, 196(2)

ISSN

1073-449X

Authors

Huang, Yong
Ma, Shwu-Fan
Espindola, Milena S
et al.

Publication Date

2017-07-15

DOI

10.1164/rccm.201607-1525oc

Peer reviewed

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Title: Microbes associate with host innate immune response in idiopathic pulmonary fibrosis

Author: Yong Huang[1*], Shwu-Fan Ma[1*], Milena S. Espindola[2], Rekha Vij[1], Justin M. Oldham[3], Gary B. Huffnagle[4], John R. Erb-Downward[4], Kevin R. Flaherty[4], Beth B. Moore[4], Eric S. White[4], Tong Zhou[5], Jianrong Li[6], Yves A. Lussier[6], MeiLan K Han[4], Naftali Kaminski[7], Joe G. N. Garcia[6], Cory M. Hogaboam[2#], Fernando J. Martinez[8#], Imre Noth[1#§], for the COMET Investigators

Affiliation:

1. Section of Pulmonary and Critical Care Medicine, Department of Medicine, University of Chicago, Chicago, IL
2. Pulmonary & Critical Care Medicine, Cedars-Sinai Medical Center, Los Angeles, CA
3. Pulmonary and Critical Care Medicine, University of California at Davis, Sacramento, CA
4. Division of Pulmonary and Critical Care Medicine, University of Michigan Health System, Ann Arbor, MI
5. Department of Physiology and Cell Biology, University of Nevada School of Medicine, Reno, NV
6. University of Arizona Health Sciences at the University of Arizona, Tucson, AZ
7. Pulmonary, Critical Care & Sleep Medicine, Yale University, New Haven, CT
8. Weill Cornell Medical College, Cornell University, New York, NY

*Huang Y. and Ma S. F. contributed equally to the manuscript

#Hogaboam C. M., Martinez F. J., and Noth I. are co-senior authors

§Correspondence and requests for reprints should be addressed to Imre Noth, M.D., Section of Pulmonary and Critical Care Medicine, Department of Medicine, The University of Chicago, 5841 S. Maryland Avenue, MC6076, Chicago, IL 60637. E-mail: inoth@medicine.bsd.uchicago.edu

Author Contributions: Y.H. analyzed the data and wrote the manuscript; S.-F.M. collected data, assisted with data analysis, prepared and wrote the manuscript; M.S.E. performed the CpG-ODN experiments and assisted with manuscript preparation; R.V., and J.M.O. contributed to interpretation of results and manuscript preparation; Z.T., J.R.E., G.B.H., J.L., B.B.M., Y.A.L., and C.M.H. contributed to data analysis; E.S.W., K.R.F., M.K.H., G.B.H., B.B.M., N.K., J.G.N.G., C.M.H., and F.J.M. contributed to conception and study design, interpretation of the results and manuscript preparation; I.N. contributed to conception and study design, interpretation of the results, and wrote the manuscript. All authors reviewed, revised, and approved the manuscript for submission.

Supported by RC2 HL101740 to FJM and IN; HL115618 to BBM； R01HL123899 to CMH; U01HL112696, U01HL125208, P01HL126609 to JGNG

Running title: Microbes and host immune response in IPF

This article has an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Abstract

Rationale: Differences in the lung microbial community influence idiopathic pulmonary fibrosis (IPF) progression. Whether the lung microbiome influences IPF host defense remains unknown.

Objectives: To explore the host immune response and microbial interaction in IPF, as they relate to progression-free survival (PFS), fibroblast function, and leukocyte phenotypes.

Methods: Paired microarray gene expression data from peripheral blood mononuclear cells (PBMCs) and 16s rRNA sequencing data from bronchoalveolar lavage (BAL) obtained as part of the COMET-IPF study were used to conduct association pathway analyses. The responsiveness of paired lung fibroblasts to TLR9 stimulation using CpG-oligodeoxynucleotide (CpG-ODN) was integrated into microbiome-gene expression association analyses for a subset of individuals. The relationship between associated pathways and circulating leukocyte phenotypes by flow cytometry was also explored.

Measurements and Main Results: Down-regulation of immune response pathways, including NOD-, TOLL- and RIG1-like receptor pathways, was associated with worse PFS. Ten of the 11 PFS-associated pathways correlated with microbial diversity and individual genus, with SAC-richness as a hub. Higher SAC-richness was significantly associated with inhibition of NODs and TLRs, while increased abundance of *streptococcus* correlated with increased NOD-like receptor signaling. In a network analysis, expression of up-regulated signaling pathways were strongly associated with decreased abundance of OTU1341 *(Prevotella)* among individuals with fibroblasts

responsive to CpG-ODN stimulation. The expression of TLR signaling pathway was also linked to CpG-ODN responsive fibroblasts, OTU1341 (*Prevotella*), and microbial Shannon's index in a network analysis. Lymphocytes expressing CXCR3 CD8 significantly correlated with OTU1348 (*Staphylococcus*).

Conclusions: These findings suggest that host-microbiome interactions influence PFS and fibroblast responsiveness.

Abstract words count: 250

Keywords: Host immune response and microbial interaction, PBMCs transcriptome, BAL microbiome, CpG-ODN response, Pattern recognition receptors

At a Glance Commentary

Scientific Knowledge on the Subject: Peripheral blood gene expression profiling has identified immune signaling pathways to be associated with outcomes and disease severity in patients with idiopathic pulmonary fibrosis (IPF). Variation in the lower airway microbiome has also been associated with differential outcome risk. Whether the airway microbiome influences host immune response in patients with IPF remains unclear.

What this Study Adds to the Field: This study demonstrates that host defense, as assessed by immune pathway gene expression, may be modulated by variations in the lower airway microbiome. This study also demonstrates that host-microbiome

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

interaction may influence immune-mediated fibroblast responsiveness and the composition of circulating leukocytes.

Total words count: 3432

Introduction (words count 450)

Idiopathic pulmonary fibrosis (IPF) is an interstitial lung disease with high mortality and variable disease progression. While some patients progress rapidly to death, others experience a slower decline with periods of stability (1). The cause(s) of IPF and factors leading to disease progression remain incompletely characterized (2).

IPF pathogenesis involves recurrent injury to the alveolar epithelium with aberrant wound healing that results in fibrosis, rather than normal repair (3-5). Studies have implicated immunologic aberrations to be associated with IPF progression. These include the development of autoimmunity (6) and dysregulation of immune signaling (7-10). *TOLLIP, TLR3, TLR9* and *MUC5B* (11-15), all of which contribute to innate immunity and host defense (16-21), and have been linked to IPF susceptibility and outcomes. The down-regulation of *CD28*, *ICOS*, *LCK*, and *ITK,* genes involved in T-cell signaling, have also been found to predict transplant-free survival in patients with IPF (9), further supporting an immunologic link. Genes involved in the alpha defensin pathway, critical to host defense, also predict disease severity (22). Importantly, sustained *TLR9* activation in lung myofibroblasts can characterize rapidly progressive IPF (15). Altogether, these data suggest a role for modulation of immune mechanisms in IPF disease activity.

While the lung has been historically considered sterile, modern culture-independent techniques show diverse populations of bacteria in the lung (23, 24). Microaspiration along with microbial migration, elimination and relative growth rates determine the composition of the lung microbiome (25, 26). Changes in host microanatomy, cell biology and innate defenses can alter the dynamics of bacterial turnover, leading to colonization by well-recognized bacterial pathogens. In turn, this dynamic can influence lung inflammation (27).

Aberrant host immunity may lead to colonization and proliferation of pathologic organisms in lower airways, potentially contributing to the recurrent alveolar injury characteristic of IPF pathogenesis. We previously demonstrated that the presence of two operational taxonomic units (OTUs), belonging to *Staphylococcus* and *Streptococcus* species, in bronchoalveolar lavage (BAL) fluid predicted differential survival in patients with IPF (28). Other investigators also identified *Streptococcus* to be more abundant in IPF cases compared to controls and identified increased bacterial burden and decreased diversity to be predictive of poor outcomes in IPF (16). Whether the composition of lower airway microbiome influences IPF-relevant peripheral blood gene expression pathways remain unknown.

In this investigation we conduct a comprehensive analysis of host-microbiome interaction by integrating peripheral blood gene expression profiles, lung microbial community and IPF outcomes. Specifically, we hypothesized that the lung microbiome

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

influences innate and adaptive immune responses signaling and that host-microbiome interaction modulates progression-free survival (PFS). We utilized paired clinical, gene expression, circulating leukocytes and microbial data from patients enrolled in the Correlating Outcomes with Biochemical Markers to Estimate time-progression" (COMET-IPF) study (NCT 01071707) to address this global hypothesis. Additionally, we investigate links between identified lung microbes and CpG-ODN responsiveness of lung fibroblasts to innate TLR9 activation.

Methods

Study Population

COMET-IPF participants were diagnosed using ATS/ERS criteria (29) and prospectively enrolled at nine U.S. clinical centers. Inclusion and exclusion criteria, along with trial endpoints have been previously described (10, 28, 30). Description of demographic and baseline pulmonary function test are presented in Online Data Supplement. A composite physiologic index (CPI) (31) was calculated for all participants. Progression free survival (PFS), the primary combination endpoint for COMET-IPF, was defined as the time from study enrollment to death, acute exacerbation, lung transplant, or relative change in forced vital capacity (FVC) ≥10% or diffusion capacity for carbon monoxide (D_LCO) ≥15%.

PBMCs isolation, RNA extraction, microarray hybridization, data processing

RNA was extracted from PBMCs isolated from blood of participants at study entry for gene expression profiling using a microarray platform. Detailed descriptions of methods are in Online Data Supplement.

Bronchoscopic alveolar lavage (BAL) for microbiome 16S rRNA sequencing and data processing

Bronchoscopy was completed at enrollment in patients who were clinically stable and without evidence of active infection, as described previously (28). Descriptions of BAL microbial species determination and data processing are presented in the Online Data Supplement.

Lung fibroblasts culture

Transbronchial biopsy samples were obtained for fibroblast culture in DMEM (Lonza) containing 15% fetal calf serum (Cell Generation), 100 IU penicillin, 100 µg/ml streptomycin (Mediatech), 292 µg/ml L-Glutamine (Mediatech), and 100 µg/ml of Primocin (Invivogen) at 37ºC and 10% CO_2. Fresh medium was added to the fibroblasts every 2-3 days and the cells were passaged when they were 70-90% confluent.

Data analyses and integration

Please see the Online Data Supplement for additional details.

PFS-associated host canonical signaling pathways

Microarray gene expression data from host PBMCs was transformed into a data matrix of Pathways-Samples using the R/Bioconductor package "GSVA", as described in the

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Online Data Supplement. Pathways in the Molecular Signature Database (MSigDB), a collection of annotated gene sets, were correlated with PFS using Cox-PH regression.

Correlation of gene co-expression modules with clinical traits and Operational Taxonomic Units (OTUs)

The gene expression profile of PBMCs were clustered into gene modules, based on their co-expression pattern, using an unsupervised "Weighted gene co-expression network analysis (WGCNA)" package in R (32). A Principle Component Analysis (PCA) was used to calculate an eigengene for each gene module (32, 33). A p-value<0.05 was considered significant for Pearson's correlation among individual gene module eigengenes, clinical traits, and microbial community features.

Functional pathways enrichment analysis

Canonical pathways were analyzed using Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Redwood City, CA). Significant pathways were identified using one-sided Fisher's exact. See Online Data Supplement for details.

Cytospace network analysis

Identification and visualization of interaction networks that incorporate clinical traits, canonical pathways, and microbial features was performed using the Cytospace platform (http://www.cytoscape.org).

Fibroblast responsiveness to TLR9 stimulation using CpG-ODN.

Transbronchial biopsies specimens were available in a subset of COMET-IPF patients (n=27). Lung fibroblasts derived from these transbronchial biopsies were treated for 24 hours with or without CpG-ODN (15). Alpha-SMA expression (i.e. treated vs. untreated) >2 fold was deemed "responsive". Logit regression with univariate and multivariate analysis was conducted to determine microbial genera associated CpG-ODN responsiveness. Support Vector Machines (SVM) followed by Leave-One-Out (LOO) was used to construct a microbial model predictive of CpG-ODN responsiveness.

Correlation of circulating leukocyte phenotypes with microbial measures

Peripheral blood collected from COMET-IPF patients was available in 32 of the 68 cases and was analyzed for leukocyte phenotypes by flow cytometry as previously described (10).

Results

Study Population

Sixty-eight of the 88 subjects enrolled in COMET-IPF possessed both sufficient PBMC RNA and BAL specimens for the paired integration analyses **(Supplemental Figure E1)**. A subset (n=27) of these 68 COMET-IPF had paired CpG-ODN response data to TLR9 stimulation in lung fibroblasts derived from transbronchial biopsies and microarray gene expression data from host PBMCs **(Supplemental Figure E1)**. Demographics and clinical traits of these two subsets are summarized in **Table 1**.

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

PFS-associated canonical signaling pathways are predominantly immunologic

The events comprising the composite endpoint in the COMET-IPF cohort **(Supplemental Table E1)** are summarized with the Kaplan-Meier estimator (**Supplemental Figure E2**) and demonstrate that categorical lung function decline in FVC or D_LCO are the largest constituents of the PFS composite endpoint in this study. Acute exacerbations were only a minor component at 5 total events. In our pathway analysis of PBMC gene expression microarray, 11 canonical pathways with odds ratios < 1 were significantly associated with PFS, indicating inhibition of these pathways was associated with poorer PFS (**Table 2**). Among these pathways, eight involved the immune inflammatory response and pathogen infection, while three involved pattern recognition receptors (PRRs) responding to pathogen associated molecular patterns (PAMPs). These included Toll-like receptors (TLRs), NOD-like receptors (NODs), and RIG 1-like receptor (RIG1) signaling pathways.

Systems biology integration of PFS-associated pathways and lung microbial community

Ten out of 11 PFS-associated pathways (**Figure 1,** grey hexagons) correlated with microbial community features (green circles). Network analysis and visualization revealed a microbial feature (SAC richness, golden circle) as the hub node in this network, connecting to 7 of 10 pathways with significant negative correlation (green edges). The most significant connections were between increased SAC richness and decreased NODs and TLRs signaling pathways (thick green edges). Additionally, OTU1345 (*Streptococcus* genus), found in increased abundance in progressive IPF

patients (34) was negatively (green edges) and significantly correlated with the NODs signaling pathway (**Figure 1)**. Increased abundance of OTU1345 (*Streptococcus*), OTU1274 (*Pseudomonas)*, OTU1249 (*Leptotrichia*), OTU1350 (*Streptococcus*), and OTU1254 (*Actinomyces*) positively (red edges) correlated with PRR pathways (i.e. TLRs, NOD2, and RIG1 signaling pathways), which is opposite to SAC richness.

Gene co-expression modules associate with clinical traits, microbiome diversity and individuals OTUs

A correlation matrix of paired gene expression modules, microbial community features, and clinical traits showed varying degrees of correlation between each of these variables (**Figure 2)**. A list of genes in each module correlated with microbial SAC richness can be found in **Supplemental Table E2**. The magenta module was positively correlated with D_LCO (r=0.32, *p*=0.007, red box) and negatively correlated with CPI (r=-0.33, *p*=0.006, green box). This inverse relationship is expected, as decreasing D_LCO leads to increasing CPI.

The magenta module was also positively correlated with several microbial community features, including SAC-richness (r=0.52, $p<5e\text{-}10^{-6}$), OTU1302 (*Pseudomonadaceae,* r=0.45*, p*<1x-10^{-4}), OTU1256 (*Prevotella,* r=0.3, *p*=0.01), and OTU1291 (*Prevotella,* r=0.26*, p*=0.03). OTU1348 *(Staphylococcus)* and OTU1341 (*Prevotella)* were negatively correlated (i.e. r<0) with the purple, green, yellow and pink gene modules. The full taxonomy classification for each OTU is listed in **Supplemental Table E3.** A gene significance analysis of the magenta module further confirmed a negative correlation

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

with CPI (p=4.7x10^{-6}) and a positive correlation with microbial richness (p=1.2x10^{-27}) (**Supplemental Figure E3A and E3B,** respectively). The canonical pathways within each host gene module are provided in **Supplemental Table E4**.

Pathways enriched (p<0.05) within the magenta module included integrin signaling, Rho family GTPase signaling, granulocyte movement and coagulation system signaling (**Supplemental Figure E4)**. Integrin alpha(v)beta3, commonly expressed in platelets, is involved in Rho GTPases mediated activation of sphingosine-1-phosphate (S1P) endothelial cell chemotaxis (35) and also acts as a receptor for another gene in the set, Von Willebrand Factor. S1P activity has been shown to be involved in the onset of pulmonary fibrosis (36, 37). Interestingly, integrin alpha(v)beta6, also linked to pulmonary fibrosis via Rho-mediated signaling (38-40) is not included in the magenta module. Other genes include carbonic anhydrase 2, which is involved in bacterial killing (41), and PDE5A, which mediates pulmonary hypertension and has been previously targeted in IPF (42).

PBMC canonical pathways associate with microbial community features

Significance Analysis of Microarray (SAM) identified genes significantly associated with selected microbial community features, including indices of Shannon, Inv-Simpson, SAC-richness, and abundance of OTU1348 (*Staphylococcus*), OTU1341 (*Prevotella*), and OTU1302 (*Pseudomonas*) using criterion of FDR<10% (**Supplemental Table E5**). Ingenuity pathways analysis (IPA) revealed that ILK signaling pathway and iCOS-iCOSL signaling in T-helper cells were positively and negatively associated with Shannon

divergence, respectively. Integrin signaling and epithelial adherence junction signaling pathways were associated with SAC-richness index, while natural killer cell signaling was associated with Inv-Simpson index. All genes negatively correlated with OTU1348 (*Staphylococcus*) were overrepresented in IGF-1 and VEGF signaling pathways. All significantly enriched canonical pathways identified by this approach are shown in **Supplemental Table E6**. Both WGCNA and SAM approaches support linkage of diversity indices scores with less expression of immune regulatory functions (ie. iCOS-iCOSL), and more integrin signaling, or cell survival pathways.

Microbial abundance correlates with TLRs gene expression in host PBMCs

In the list of PFS-associated canonical pathways, TLRs signaling was the most prominent feature (**Table 2**). Thus, we correlated microbial OTUs with expression of *TLR1-10* and their inhibitory regulator, toll-interacting protein (*TOLLIP*) (**Supplemental Figure E5**). When correlating TLRs mRNA expression levels with the OTUs in BAL, the most significant positive correlation was observed between *TLR9* and OTU1348 (*Staphylococcus,* r=0.38, adjusted *p*-value (FDR) =0.02) and OTU1341 (*Prevotella*, r=0.48, FDR=0.0003). *TOLLIP* expression was correlated with OTU1274 (*Pseudomonas*, r=0.31, FDR=0.1). OTU1278 (*Bordetella*) was significantly correlated with *TOLLIP* and multiple *TLRs*, including *TLR2* (r=0.37, FDR=0.02).

Fibroblast responsiveness to TLR9 signaling correlates with altered microbial community

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

TLR9-mediated CpG-ODN fibroblast responsiveness was associated with a trend towards differential PFS in this cohort, though log rank p-value did not cross the significance threshold, due in part to the small sample size (**Supplemental Figure E6**). Using an iterative SVM model followed by LOO cross-validation, we prioritized 10 OTUs that demonstrated 75% sensitivity and 82% specificity in prediction of fibroblast CpG-ODN responsiveness (**Supplemental Table E7**). Individual predictors of CpG-ODN responsiveness included OTU1345 (*Staphylococcus)* and OTU1331 (*Veillonella*), which had opposing effect sizes (fold change 4.07 vs 0.085, respectively). Notably, four OTUs representing *Prevotella* were identified by this prediction model (**Supplemental Table E7).** The abundance of OTU1331 *Veillonella* was significantly correlated with CpG-ODN responsiveness (odds ratio=1.14, 95% CI 1.02-1.68; *p*=0.048). The association was maintained after adjustment for gender, age, and CPI (OR 1.2; 95% CI 1.01-1.43; *p*=0.034).

A network of canonical pathways is shared among CpG-ODN responsiveness in fibroblasts, gene co-expression modules in host PBMC and microbial features in BAL

We identified 401 differentially expressed genes from host PBMCs in cases correlating with CpG-ODN responsiveness in fibroblasts of IPF patients. An FDR≦0.01 was selected to reduce the number of pathways for the KEGG pathway analysis. This analysis revealed that the 93 up-regulated genes associated with CpG-ODN responsiveness were enriched in 22 canonical pathways. These, included the immune/inflammatory response, the role of cytokines in mediating communication

between immune cells, IL-10, TLR and IL-6 signaling pathways (**Supplemental Figure E7**).

To look at potential interactions, the 22 CpG-ODN responsive associated pathways (**Supplemental Figure E7**) were integrated to pathways enriched within each of the 10 host gene co-expression modules (**Supplemental Table E4**), and to microbiota associated pathways (**Supplemental Table E6**). We identified 14 overlapping canonical pathways, which were used to construct an interaction network by Cytospace v3.0 software (**Figure 3**). This network demonstrates that patients with CpG-ODN responsive fibroblasts (yellow hexagon) are positively correlated (red lines) with numerous immune response pathways (light blue circles). There was a negative association (green lines) between these immune response pathways and the abundance of OTU1341 (*Prevotella;* purple diamond), and to a lesser degree, OTU1348 (*Stapylococcus*; purple diamond). The top four nodes displaying highest network betweenness-centrality were OTU1341 (*Prevotella*), glucocorticoid receptor (GR) signaling, green gene module, and Shannon's diversity index (**Supplemental Figure E8A**). Three nodes displayed high closeness-centrality, including NF-kB signaling, role of osteoblasts and chondrocytes in rheumatoid arthritis, and acute phase response signaling (**Supplemental Figure E8B**).

Inflammatory leukocyte phenotypes associate with microbial measures

Associations between paired lung microbial data and blood leukocyte phenotypes were assessed in a subset of patients with these data (n=32) (**Supplemental Figure E1**). This demonstrated a significant correlation between CXCR3 CD8 activated T cells and

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

with OTU1348 (*Staphylococcus,* r=0.77, FDR=0.0002) and OTU1283 (*Cronobacter,* r=0.71, FDR=0.001) (**Figure 4A, red box**). The most significant correlation in CD14-expressing cells was found between CD14hiCD16lo CCR2-expressing classical inflammatory macrophages and OTU1281 (*Acinetobacter,* r=0.47, FDR=0.05) **(Figure 4A, red box)**. Both CD4+CD28lo and CD8+CD28lo non-activated T cells significantly correlated with decreased microbial Sac richness, (**Figure 4B**, r=-0.43, FDR=0.06, green box; and r=-0.38, FDR=0.06, green box) and increased SAC richness correlated with total CD4+CD28hi non-activated T cells (**Figure 4B**, r=0.35, FDR=0.07, red box).

Discussion

Microbial pathogenicity, arising either from loss of microbial communal health or loss of an appropriate immune response, has been implicated as a mechanism for variability in IPF disease progression (28). Our comprehensive systems biology approach layered clinical features, including PFS, with multiple biological samples (PBMC gene expression, BAL 16S rRNA sequencing, fibroblast CpG-ODN responses and circulating leukocyte phenotypes) to reveal supportive links to key host immune response pathways. We demonstrate that 1) Reduced expression of PFS-associated canonical pathways in PBMCs was associated with a worse outcome (Table 2); 2) Down-regulation of immune response relevant pathways is associated with changes in the abundance of specific microbial OTUs (Figures 1 and 2); 3) Abundance of particular OTUs correlates with several *TLRs* expression (Supplemental Figure E5); 4) Lung fibroblast CpG-ODN responsiveness to TLR9 activation is positively correlated with

expression of host immune response relevant signaling pathways (Figure 3); 5) Increased expression of these same immune response relevant pathway genes is negatively correlated with the abundance of OTU1341 (*Prevotella*) and OTU1348 (*Staphylococcus*) (Figure 3); and 6) Alterations in the microbial community structure associate with changes in circulating leukocyte phenotypes (Figure 4). These data suggest that PRR and immune response relevant signaling pathways play a central role in PFS-associated host-microbial interactions and may provide a target for therapeutic intervention.

Our PFS-transcriptomic survival analysis identified 11 significant canonical pathways, including major PRR signaling pathways and four other inflammatory response pathways (Table 2). TLR-, RIG1-, and NOD-like receptor signaling are integral to the innate immune response, which constitutes the first line of defense against invading microbial pathogens. PRRs detect distinct evolutionarily conserved structures on pathogens, termed pathogen-associated molecular patterns (PAMPs) (43). Recognition of PAMPs by PRRs rapidly triggers an array of anti-microbial innate immune responses through the induction of various inflammatory cytokines, chemokines and type I interferons to promote host defense.

Nine of the 11, mostly immune, host canonical pathways correlated with both PFS and microbial diversity indices (Figure 1). Most dramatic was increased SAC-richness index and a strong negative correlation with NOD-like receptor and TLR signaling pathways. Furthermore, a reduced immune response represented by decreased NOD-like receptor

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

signaling associated with increased abundance of OTU1345 (*Streptococcus)*, and both correlated with poorer PFS, suggesting a genomic link. Interestingly, recent work suggests *Streptococcus* pneumolysin exacerbates lung fibrosis in murine models (44).

Our data also suggest that the microbial community composition is related to impaired epithelial integrity and severity of lung function abnormality, which is a surrogate of IPF severity. In the WGCNA analysis, bacteria community SAC-richness index, OTU1302 (*Pseudomonadaceae)*, and preserved D_LCO and CPI were linked to the magenta gene module. This module, by IPA, was enriched for integrin and epithelial adherens junction signaling, as well as signaling by Rho Family GTPases and coagulation system, among others (35, 45, 46). All of these are known to be important in epithelial and mesenchymal cell homeostasis (Supplemental Figure E4) (47-50). Conversely, loss of microbial community diversity is associated with a reduced PBMC gene expression in this module. This may be acting as a surrogate for changes of epithelial and mesenchymal cell activity, which are important in IPF. In addition to OTU1348 (*Staphylococcus*), WGCNA identified OTU1341 (*Prevotella*) as strongly linked to multiple gene expression modules (Figure 2), suggesting a potential symbiosis between *Prevotella spp.* and *Staphylococcus ssp.* in reducing PFS.

Examining genes correlated with quantitative traits by SAM for Inv-Simpson (219 genes) and SAC-richness (325 genes), suggest that a more diverse microbial environment leads to an appropriate host immunologic response (Supplemental Table E5). Conversely, OTU1348 (*Staphylococcus*) significantly correlated with decreased

expression of all 464 significant genes identified with overrepresentation in IGF-1 and VEGF signaling pathways, both of which have been implicated in pulmonary fibrosis (51), and are involved in tissue repair (52).

To our knowledge, our study is the first to link blood gene expression with lung microbiome. There are studies that have linked the microbiome to lung tissue gene expression (53), and between the PBMC expression and lung tissue expression (54), but none between the PBMC expression and microbiome directly. Therefore, any comparison to these studies and our current analysis should be taken with caution.

Our findings also suggest the lung fibroblast response to CpG-ODN-mediated TLR9 signaling may be dependent upon the lung microbial community and/or may also reflect the status of immune signaling within the patient as a whole. It is important to note, that in prior studies by our group using fibroblasts from surgical lung biopsies, only IPF cases demonstrate significant activation by CpG-ODN stimulation, while fibroblasts from normal individuals did not (55). However, these historical surgical biopsies differ in anatomical location than the transbronchial biopsies collected by COMET investigators, leaving it unknown whether these transbronchial biopsy specimens approximate those that would be observed in surgical lung biopsies. Both alone and within our LOOCV predictor model, increased presence of OTU1331 (*Veillonella*) correlated with increased CpG-ODN fibroblast responsiveness. *Veillonella* is a commensal, anaerobic gram-negative coccus, normally found in the mouth and gut. It commonly triggers the innate immune response via TLR4 (56). Multiple OTUs representing *Prevotella* and

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Streptococcus, which predicts reduced PFS in IPF, were also found to be significant negative predictor of fibroblast responsiveness. Positive correlations between *TLR9* expression in PBMCs and OTU1348 (*Staphylococcus)* and OTU1341 (*Prevotella*) (Supplemental Figure E5) were also observed. While speculative, impaired innate responses in patients with progressive IPF may allow outgrowth of potentially pathogenic species such as *Staphylococcus,* which may in turn promote lung injury and fibrosis. Alternatively, we may be measuring abundance of microbial OTUs (e.g. *Prevotella)* that are not generally considered pathogenic but may co-segregate with other microbial species which are harmful, but not identified in this analysis.

Our network (Figure 3) shows the inter-relationship between an altered microbial community, CpG-ODN responsiveness and PBMC gene expressions. Our KEGG analysis of CpG-ODN responsiveness supports activation of the immune/inflammatory response, IL-6 signaling and TLRs, among others. Furthermore, the magenta module, which is strongly representative of richness, was notably absent in our network analysis. Our network modeling revealed that CpG-ODN responsiveness was connected with OTU1348 (*Staphylococcus*) via glucocorticoid receptor (GR) signaling, and with OTU1341 (*Prevotella)* via 10 canonical pathways, including GR, TLR, and IL6 signaling pathways (Figure 3). This network suggests that immune responses in PBMCs may be down-regulated in response to accumulation of certain microbial species, depending on fibroblast responsiveness.

Our circulating leukocyte phenotype data provides additional evidence that innate immune responses are aberrant in IPF patients and may be modulated by alterations in the microbial community. The correlation between potentially pathogenic genus (*Staphylococcus*, OTU1348) with accumulation of a CXCR3+ CD8+ T cells involved in Th1 signaling, could be evidence that the immune system is activated in response to certain microbes, potentially causing lung damage or alternatively, that immune cells accumulate to combat the infection. Similarly, we cannot assume that TLR9 signaling is universally beneficial. While TLR9 activation of professional immune cells may help to limit pathogens, we have also reported that higher expression of TLR9 and consequent CpG-ODN-induced myofibroblast differentiation may accelerate disease progression in IPF.

Our results suggest one possible mechanism for why use of immunosuppressants in PANTHER proved to be harmful. One patient participated in both PANTHER and COMET, though the PANTHER treatment assignment is unknown. An additional 5 COMET cases were treated with low-dose prednisone at the time of bronchoscopy and blood collections, one of which was also treated with azathioprine. These small numbers precluded the ability to include such therapies in our analysis.

There are several limitations to this study. Given the study design, our findings are only associative and cannot prove causality. Additionally, the expression of some genes identified in this investigation may be influenced by non-PBMC-derived mRNA. This includes PDE5A and other vascular-derived genes. Another limitation stems from the

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

use of a composite endpoint in this analysis. Categorical decline in lung function constituted the largest number of events in this PFS composite endpoint. This may explain why canonical pathways involving innate immunity were highlighted in this investigation, while those involved in adaptive immunity with T cell signaling were identified as predictive of transplant-free survival previously (9). Limiting the analysis to pulmonary function decline by exclusion of deaths (n=4) and acute exacerbations (n=5) reduced the total number of significant pathways, but did not appreciably change the results (Supplement Table E8).

The size and complexity of this project precluded any reasonable pursuit of a validation cohort. As such, replication of our findings, along with mechanistic confirmation, is needed. Until then caution is advised when focusing on any individual gene identified by this investigation. We attempted to optimize interpretation of our results by employing several supervised and unsupervised approaches, including GSVA, WGCNA, and network modeling, all of which showed similar findings. We also acknowledge that PBMC transcriptomics and circulating leukocyte phenotypes may not reflect the lung immune response. It is possible that activated inflammatory cells are trafficking to lung, leaving behind peripheral leukocytes that have either not been activated or that are unable to be activated. We did not investigate immune cells in the lung or BAL. Therefore, whether or not our findings are indicative of lung pathobiology remains to be determined. Furthermore, while the peripheral blood provides an easily accessible compartment for RNA collection, it remains unclear how representative RNA from peripheral blood approximates that derived from relevant lung tissue.

A larger scale replication for a blood transcriptomics linked to bronchoscopic microbiome and clinical outcomes should be pursued in the future and would allow biomarker development for pragmatic clinical application. Additional high-priority investigations include assessing change over time in the microbiome and transcriptomics in the absence and presence of antimicrobial therapy, which would help address the “chicken and egg” issue for cause and effect. Lastly, ascertaining deeper scale lung microbiome characteristics using whole metagenomic sequencing would allow investigation into microbial functional traits instead of taxa. This has the potential to be significantly more meaningful for pathway analyses and the host immune response.

Conclusion

Our comprehensive approach sheds light on the molecular mechanism underpinning host-microbial interaction. In each instance, microbes with increased abundance and decreased community diversity were associated with decreased PBMC transcriptomic expression of immune pathways and poorer PFS. Our data support an evolving concept of pathophysiological fibrosis-like reactions as an essential process in host immune reaction against potential pathogens. These data support the exploration of therapeutic approaches targeting modulation of the lung microbial community of IPF patients, such as the upcoming NIH sponsored CLEANUP-IPF trial (NCT#02759120) assessing the

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

efficacy of antibiotic therapy in IPF. More advanced metagenomic analyses are required to elucidate the functional role of individual genus and communities in IPF progression.

References

1. Nathan SD, Shlobin OA, Weir N, Ahmad S, Kaldjob JM, Battle E, Sheridan MJ, du Bois RM. Long-term course and prognosis of idiopathic pulmonary fibrosis in the new millennium. *Chest* 2011; 140: 221-229.
2. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA, Lynch DA, Ryu JH, Swigris JJ, Wells AU, Ancochea J, Bouros D, Carvalho C, Costabel U, Ebina M, Hansell DM, Johkoh T, Kim DS, King TE, Jr., Kondoh Y, Myers J, Muller NL, Nicholson AG, Richeldi L, Selman M, Dudden RF, Griss BS, Protzko SL, Schunemann HJ, Fibrosis AEJACoIP. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011; 183: 788-824.
3. Kim KK, Kugler MC, Wolters PJ, Robillard L, Galvez MG, Brumwell AN, Sheppard D, Chapman HA. Alveolar epithelial cell mesenchymal transition develops in vivo during pulmonary fibrosis and is regulated by the extracellular matrix. *Proc Natl Acad Sci U S A* 2006; 103: 13180-13185.
4. Myers JL, Katzenstein AL. Epithelial necrosis and alveolar collapse in the pathogenesis of usual interstitial pneumonia. *Chest* 1988; 94: 1309-1311.
5. Uhal BD, Joshi I, True AL, Mundle S, Raza A, Pardo A, Selman M. Fibroblasts isolated after fibrotic lung injury induce apoptosis of alveolar epithelial cells in vitro. *The American journal of physiology* 1995; 269: L819-828.
6. Wilkes DS, Chew T, Flaherty KR, Frye S, Gibson KF, Kaminski N, Klemsz MJ, Lange W, Noth I, Rothhaar K. Oral immunotherapy with type V collagen in idiopathic pulmonary fibrosis. *Eur Respir J* 2015; 45: 1393-1402.
7. Gilani SR, Vuga LJ, Lindell KO, Gibson KF, Xue J, Kaminski N, Valentine VG, Lindsay EK, George MP, Steele C, Duncan SR. CD28 down-regulation on circulating CD4 T-cells is associated with poor prognoses of patients with idiopathic pulmonary fibrosis. *PLoS One* 2010; 5: e8959.
8. O'Dwyer DN, Ashley SL, Moore BB. Influences of innate immunity, autophagy, and fibroblast activation in the pathogenesis of lung fibrosis. *Am J Physiol Lung Cell Mol Physiol* 2016; 311: L590-601.
9. Herazo-Maya JD, Noth I, Duncan SR, Kim S, Ma SF, Tseng GC, Feingold E, Juan-Guardela BM, Richards TJ, Lussier Y, Huang Y, Vij R, Lindell KO, Xue J, Gibson KF, Shapiro SD, Garcia JG, Kaminski N. Peripheral blood mononuclear cell gene expression profiles predict poor outcome in idiopathic pulmonary fibrosis. *Sci Transl Med* 2013; 5: 205ra136.
10. Moore BB, Fry C, Zhou Y, Murray S, Han MK, Martinez FJ, Flaherty KR, The CI. Inflammatory leukocyte phenotypes correlate with disease progression in idiopathic pulmonary fibrosis. *Front Med* 2014; 1.
11. Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, Broderick SM, Wade MS, Hysi P, Scuirba J, Richards TJ, Juan-Guardela BM, Vij R, Han MK, Martinez FJ, Kossen K, Seiwert SD, Christie JD, Nicolae D, Kaminski N, Garcia JG. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med* 2013; 1: 309-317.

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

12. O'Dwyer DN, Armstrong ME, Trujillo G, Cooke G, Keane MP, Fallon PG, Simpson AJ, Millar AB, McGrath EE, Whyte MK, Hirani N, Hogaboam CM, Donnelly SC. The Toll-like receptor 3 L412F polymorphism and disease progression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2013; 188: 1442-1450.
13. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, Loyd JE, Cosgrove GP, Lynch D, Groshong S, Collard HR, Wolters PJ, Bradford WZ, Kossen K, Seiwert SD, du Bois RM, Garcia CK, Devine MS, Gudmundsson G, Isaksson HJ, Kaminski N, Zhang Y, Gibson KF, Lancaster LH, Cogan JD, Mason WR, Maher TM, Molyneaux PL, Wells AU, Moffatt MF, Selman M, Pardo A, Kim DS, Crapo JD, Make BJ, Regan EA, Walek DS, Daniel JJ, Kamatani Y, Zelenika D, Smith K, McKean D, Pedersen BS, Talbert J, Kidd RN, Markin CR, Beckman KB, Lathrop M, Schwarz MI, Schwartz DA. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013; 45: 613-620.
14. Peljto AL, Zhang Y, Fingerlin TE, Ma SF, Garcia JG, Richards TJ, Silveira LJ, Lindell KO, Steele MP, Loyd JE, Gibson KF, Seibold MA, Brown KK, Talbert JL, Markin C, Kossen K, Seiwert SD, Murphy E, Noth I, Schwarz MI, Kaminski N, Schwartz DA. Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *JAMA* 2013; 309: 2232-2239.
15. Trujillo G, Meneghin A, Flaherty KR, Sholl LM, Myers JL, Kazerooni EA, Gross BH, Oak SR, Coelho AL, Evanoff H, Day E, Toews GB, Joshi AD, Schaller MA, Waters B, Jarai G, Westwick J, Kunkel SL, Martinez FJ, Hogaboam CM. TLR9 differentiates rapidly from slowly progressing forms of idiopathic pulmonary fibrosis. *Sci Transl Med* 2010; 2: 57ra82.
16. Molyneaux PL, Cox MJ, Willis-Owen SA, Mallia P, Russell KE, Russell AM, Murphy E, Johnston SL, Schwartz DA, Wells AU, Cookson WO, Maher TM, Moffatt MF. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014; 190: 906-913.
17. Shah JA, Vary JC, Chau TT, Bang ND, Yen NT, Farrar JJ, Dunstan SJ, Hawn TR. Human TOLLIP regulates TLR2 and TLR4 signaling and its polymorphisms are associated with susceptibility to tuberculosis. *J Immunol* 2012; 189: 1737-1746.
18. Saito T, Yamamoto T, Kazawa T, Gejyo H, Naito M. Expression of toll-like receptor 2 and 4 in lipopolysaccharide-induced lung injury in mouse. *Cell Tissue Res* 2005; 321: 75-88.
19. Janardhan KS, McIsaac M, Fowlie J, Shrivastav A, Caldwell S, Sharma RK, Singh B. Toll like receptor-4 expression in lipopolysaccharide induced lung inflammation. *Histol Histopathol* 2006; 21: 687-696.
20. Roy MG, Livraghi-Butrico A, Fletcher AA, McElwee MM, Evans SE, Boerner RM, Alexander SN, Bellinghausen LK, Song AS, Petrova YM, Tuvim MJ, Adachi R, Romo I, Bordt AS, Bowden MG, Sisson JH, Woodruff PG, Thornton DJ, Rousseau K, De la Garza MM, Moghaddam SJ, Karmouty-Quintana H, Blackburn MR, Drouin SM, Davis CW, Terrell KA, Grubb BR, O'Neal WK, Flores SC, Cota-Gomez A, Lozupone CA, Donnelly JM, Watson AM, Hennessy CE, Keith RC, Yang IV, Barthel L, Henson PM, Janssen WJ, Schwartz DA, Boucher RC, Dickey BF, Evans CM. Muc5b is required for airway defence. *Nature* 2014; 505: 412-416.
21. Oldham JM, Ma SF, Martinez FJ, Anstrom KJ, Raghu G, Schwartz DA, Valenzi E, Witt L, Lee C, Vij R, Huang Y, Strek ME, Noth I, Investigators IP. TOLLIP, MUC5B, and the

Response to N-Acetylcysteine among Individuals with Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2015; 192: 1475-1482.
22. Yang IV, Luna LG, Cotter J, Talbert J, Leach SM, Kidd R, Turner J, Kummer N, Kervitsky D, Brown KK, Boon K, Schwarz MI, Schwartz DA, Steele MP. The peripheral blood transcriptome identifies the presence and extent of disease in idiopathic pulmonary fibrosis. *PLoS One* 2012; 7: e37708.
23. Dickson RP, Erb-Downward JR, Freeman CM, Walker N, Scales BS, Beck JM, Martinez FJ, Curtis JL, Lama VN, Huffnagle GB. Changes in the lung microbiome following lung transplantation include the emergence of two distinct Pseudomonas species with distinct clinical associations. *PLoS One* 2014; 9: e97214.
24. Suau A, Bonnet R, Sutren M, Godon JJ, Gibson GR, Collins MD, Dore J. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Applied and environmental microbiology* 1999; 65: 4799-4807.
25. Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB. The Microbiome and the Respiratory Tract. *Annu Rev Physiol* 2016; 78: 481-504.
26. Dickson RP, Huffnagle GB. The Lung Microbiome: New Principles for Respiratory Bacteriology in Health and Disease. *PLoS pathogens* 2015; 11: e1004923.
27. Huffnagle GB, Dickson RP. The bacterial microbiota in inflammatory lung diseases. *Clinical immunology* 2015; 159: 177-182.
28. Han MK, Zhou Y, Murray S, Tayob N, Noth I, Lama VN, Moore BB, White ES, Flaherty KR, Huffnagle GB, Martinez FJ, Investigators C. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir Med* 2014; 2: 548-556.
29. American Thoracic S, European Respiratory S. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. *Am J Respir Crit Care Med* 2002; 165: 277-304.
30. Naik PK, Bozyk PD, Bentley JK, Popova AP, Birch CM, Wilke CA, Fry CD, White ES, Sisson TH, Tayob N, Carnemolla B, Orecchia P, Flaherty KR, Hershenson MB, Murray S, Martinez FJ, Moore BB, Investigators C. Periostin promotes fibrosis and predicts progression in patients with idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* 2012; 303: L1046-1056.
31. Wells AU, Desai SR, Rubens MB, Goh NS, Cramer D, Nicholson AG, Colby TV, du Bois RM, Hansell DM. Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. *Am J Respir Crit Care Med* 2003; 167: 962-969.
32. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 559.
33. Huang Y, Ma SF, Vij R, Oldham JM, Herazo-Maya J, Broderick SM, Strek ME, White SR, Hogarth DK, Sandbo NK, Lussier YA, Gibson KF, Kaminski N, Garcia JG, Noth I. A functional genomic model for predicting prognosis in idiopathic pulmonary fibrosis. *BMC Pulm Med* 2015; 15: 147.

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

34. Han MK, Huang YJ, Lipuma JJ, Boushey HA, Boucher RC, Cookson WO, Curtis JL, Erb-Downward J, Lynch SV, Sethi S, Toews GB, Young VB, Wolfgang MC, Huffnagle GB, Martinez FJ. Significance of the microbiome in obstructive lung disease. *Thorax* 2012; 67: 456-463.
35. Wang L, Lee JF, Lin CY, Lee MJ. Rho GTPases mediated integrin alpha v beta 3 activation in sphingosine-1-phosphate stimulated chemotaxis of endothelial cells. *Histochemistry and cell biology* 2008; 129: 579-588.
36. Huang LS, Berdyshev E, Mathew B, Fu P, Gorshkova IA, He D, Ma W, Noth I, Ma SF, Pendyala S, Reddy SP, Zhou T, Zhang W, Garzon SA, Garcia JG, Natarajan V. Targeting sphingosine kinase 1 attenuates bleomycin-induced pulmonary fibrosis. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2013; 27: 1749-1760.
37. Huang LS, Berdyshev EV, Tran JT, Xie L, Chen J, Ebenezer DL, Mathew B, Gorshkova I, Zhang W, Reddy SP, Harijith A, Wang G, Feghali-Bostwick C, Noth I, Ma SF, Zhou T, Ma W, Garcia JG, Natarajan V. Sphingosine-1-phosphate lyase is an endogenous suppressor of pulmonary fibrosis: role of S1P signalling and autophagy. *Thorax* 2015; 70: 1138-1148.
38. Jenkins RG, Su X, Su G, Scotton CJ, Camerer E, Laurent GJ, Davis GE, Chambers RC, Matthay MA, Sheppard D. Ligation of protease-activated receptor 1 enhances alpha(v)beta6 integrin-dependent TGF-beta activation and promotes acute lung injury. *J Clin Invest* 2006; 116: 1606-1614.
39. Scotton CJ, Krupiczojc MA, Konigshoff M, Mercer PF, Lee YC, Kaminski N, Morser J, Post JM, Maher TM, Nicholson AG, Moffatt JD, Laurent GJ, Derian CK, Eickelberg O, Chambers RC. Increased local expression of coagulation factor X contributes to the fibrotic response in human and murine lung injury. *J Clin Invest* 2009; 119: 2550-2563.
40. Xu MY, Porte J, Knox AJ, Weinreb PH, Maher TM, Violette SM, McAnulty RJ, Sheppard D, Jenkins G. Lysophosphatidic acid induces alphavbeta6 integrin-mediated TGF-beta activation via the LPA2 receptor and the small G protein G alpha(q). *Am J Pathol* 2009; 174: 1264-1279.
41. Tang XX, Fok KL, Chen H, Chan KS, Tsang LL, Rowlands DK, Zhang XH, Dong JD, Ruan YC, Jiang X, Yu SS, Chung YW, Chan HC. Lymphocyte CFTR promotes epithelial bicarbonate secretion for bacterial killing. *Journal of cellular physiology* 2012; 227: 3887-3894.
42. Idiopathic Pulmonary Fibrosis Clinical Research N, Zisman DA, Schwarz M, Anstrom KJ, Collard HR, Flaherty KR, Hunninghake GW. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N Engl J Med* 2010; 363: 620-628.
43. Wynn TA. Cellular and molecular mechanisms of fibrosis. *J Pathol* 2008; 214: 199-210.
44. Knippenberg S, Ueberberg B, Maus R, Bohling J, Ding N, Tort Tarres M, Hoymann HG, Jonigk D, Izykowski N, Paton JC, Ogunniyi AD, Lindig S, Bauer M, Welte T, Seeger W, Guenther A, Sisson TH, Gauldie J, Kolb M, Maus UA. Streptococcus pneumoniae triggers progression of pulmonary fibrosis through pneumolysin. *Thorax* 2015; 70: 636-646.
45. Kudo M, Melton AC, Chen C, Engler MB, Huang KE, Ren X, Wang Y, Bernstein X, Li JT, Atabai K, Huang X, Sheppard D. IL-17A produced by alphabeta T cells drives airway

hyper-responsiveness in mice and enhances mouse and human airway smooth muscle contraction. *Nat Med* 2012; 18: 547-554.
46. Campus F, Lova P, Bertoni A, Sinigaglia F, Balduini C, Torti M. Thrombopoietin complements G(i)- but not G(q)-dependent pathways for integrin {alpha}(IIb){beta}(3) activation and platelet aggregation. *J Biol Chem* 2005; 280: 24386-24395.
47. Horan GS, Wood S, Ona V, Li DJ, Lukashev ME, Weinreb PH, Simon KJ, Hahm K, Allaire NE, Rinaldi NJ, Goyal J, Feghali-Bostwick CA, Matteson EL, O'Hara C, Lafyatis R, Davis GS, Huang X, Sheppard D, Violette SM. Partial inhibition of integrin alpha(v)beta6 prevents pulmonary fibrosis without exacerbating inflammation. *Am J Respir Crit Care Med* 2008; 177: 56-65.
48. Riches DW, Backos DS, Redente EF. ROCK and Rho: Promising therapeutic targets to ameliorate pulmonary fibrosis. *Am J Pathol* 2015; 185: 909-912.
49. Osborn-Heaford HL, Ryan AJ, Murthy S, Racila AM, He C, Sieren JC, Spitz DR, Carter AB. Mitochondrial Rac1 GTPase import and electron transfer from cytochrome c are required for pulmonary fibrosis. *J Biol Chem* 2012; 287: 3301-3312.
50. Noth I, Anstrom KJ, Calvert SB, de Andrade J, Flaherty KR, Glazer C, Kaner RJ, Olman MA, Idiopathic Pulmonary Fibrosis Clinical Research N. A placebo-controlled randomized trial of warfarin in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2012; 186: 88-95.
51. Farkas L, Farkas D, Ask K, Moller A, Gauldie J, Margetts P, Inman M, Kolb M. VEGF ameliorates pulmonary hypertension through inhibition of endothelial apoptosis in experimental lung fibrosis in rats. *J Clin Invest* 2009; 119: 1298-1311.
52. Lee KS, Park SJ, Kim SR, Min KH, Lee KY, Choe YH, Hong SH, Lee YR, Kim JS, Hong SJ, Lee YC. Inhibition of VEGF blocks TGF-beta1 production through a PI3K/Akt signalling pathway. *Eur Respir J* 2008; 31: 523-531.
53. Sze MA, Dimitriu PA, Suzuki M, McDonough JE, Campbell JD, Brothers JF, Erb-Downward JR, Huffnagle GB, Hayashi S, Elliott WM, Cooper J, Sin DD, Lenburg ME, Spira A, Mohn WW, Hogg JC. Host Response to the Lung Microbiome in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2015; 192: 438-445.
54. Koth LL, Solberg OD, Peng JC, Bhakta NR, Nguyen CP, Woodruff PG. Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. *Am J Respir Crit Care Med* 2011; 184: 1153-1163.
55. Meneghin A, Choi ES, Evanoff HL, Kunkel SL, Martinez FJ, Flaherty KR, Toews GB, Hogaboam CM. TLR9 is expressed in idiopathic interstitial pneumonia and its activation promotes in vitro myofibroblast differentiation. *Histochemistry and cell biology* 2008; 130: 979-992.
56. Matera G, Muto V, Vinci M, Zicca E, Abdollahi-Roodsaz S, van de Veerdonk FL, Kullberg BJ, Liberto MC, van der Meer JW, Foca A, Netea MG, Joosten LA. Receptor recognition of and immune intracellular pathways for Veillonella parvula lipopolysaccharide. *Clin Vaccine Immunol* 2009; 16: 1804-1809.

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Table 1. Baseline characteristics of subsets of COMET-IPF patients

Characteristics	N=68*	N=27‡
Male (%)	66.2	77.8
Age (SD)	64.8 (7.2)	64.8(8.5)
Height (cm, SD)	170.1 (10.3)	173.7 (8.7)
Weight (lb, SD)	89.8 (17.6)	93.1 (16.2)
Progression free survival at 48 wks (%)	43 (62.3)	19/27 (70.4)
CPI (SD)	50.5 (11.6)	49.9 (11.7)
D_LCO % predicted baseline (SD)	43.4 (14.3)	44.7 (14.6)
FVC % predicted baseline (SD)	70.0 (17.4)	71.5 (17.6)
FEV1 % predicted baseline (SD)	73.5 (18.9)	76.2 (19.4)
Gastroesophageal reflux (%)	57	63

*A subset of COMET-IPF patients (n=68) with expression array data from PBMC; 16S rRNA sequencing data from BAL; and clinical demographics.

‡A subset of COMET-IPF patients (n=27) with CpG-ODN mediated TLR9 response data from fibroblast; and expression array data from PBMC.

CPI=composite physiological index; D_LCO= diffusion capacity for carbon monoxide; FVC=forced vital capacity; FEV1=forced expiratory volume in 1 second

Table 2. Progression free survival (PFS)-associated host canonical pathways identified by Cox-PH regression analysis*

PFS-associated signaling pathways	Odds Ratio	Wald *p*-value
TOLL LIKE RECEPTOR SIGNALING PATHWAY	0.046	0.003‡
RIG 1 LIKE RECEPTOR SIGNALING PATHWAY	0.022	0.005†
REGULATION OF AUTOPHAGY	0.027	0.005‡
NOD LIKE RECEPTOR SIGNALING PATHWAY	0.175	0.039†
NEUROTROPHIN SIGNALING PATHWAY	0.049	0.022†
NATURAL KILLER CELL MEDIATED CYTOTOXICITY	0.035	0.009‡
EPITHELIAL SIGNLING IN H. PYLORI INFECTION	0.052	0.005‡
CYTOSOLIC DNA SENSING PATHWAY	0.037	0.017†
RENAL CELL CARCINOMA	0.073	0.015†
ADIPOCYTOKINE SIGNALING PATHWAY	0.038	0.017†
KEGG CHEMOKINE SIGNALING PATHWAY	0.065	0.025†

*Canonical pathways except “KEGG CHEMOKINE SIGNALING PATHWAY” can be found in the Interaction Network illustrated in Figure 2. †*p*-value<0.05; ‡*p*-value<0.01

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Figure Legend

Figure 1. Interaction network between progression free survival (PFS)-associated canonical pathways and microbial diversity indices and Operational Taxonomy Units (OTUs). This network illustrates the correlation interaction of PFS-associated host canonical pathways (listed in Table 2) with grey hexagons for pathways with a Wald *p*-value <0.05 or grey hexagons with blue outline for pathways with a Wald *p*-value <0.01 in PFS analysis, respectively, with microbial community features designated by golden circles for microbial diversity indices; and green circles for OTUs abundance. The diameter of the green circles is proportional to the correlation coefficient. The red edges represent positive correlation and green edges represent negative correlation. The thickness of edges is determined by 1-(*p*-value). All PFS-associated host canonical pathways except KEGG chemokine signaling pathway in Table 2 correlated with microbial community features. Microbial richness (SAC index) is the hub node in this network, connecting to 7 of 10 pathways with significant negative correlation demonstrated by green edges.

Figure 2. Correlation of gene co-expression modules with clinical traits and microbial community features. Gene co-expression modules were constructed using R package “WGCNA”. The number of genes is labeled within each module. Correlation of the module eigengene with each clinical trait and microbial community features was determined by Pearson's correlation algorithm and displayed in corresponding box (coefficient of “r” value on top and *p*-value in parenthesis). The color of each box

represents the directionality of the correlation (red=positive correlation; green=negative correlation). The bar on the right scales the degree of correlation. FVC=forced vital capacity; D_LCO=diffusion capacity of carbon monoxide; CPI=composite physiologic index. Pearson's correlation was used to determine the significance of correlation (p<0.05) between the eigengenes of individual gene module with clinical traits, including race, sex, age, FVC, D_LCO, CPI, and prognosis, as well as microbial diversity and OTU abundance.

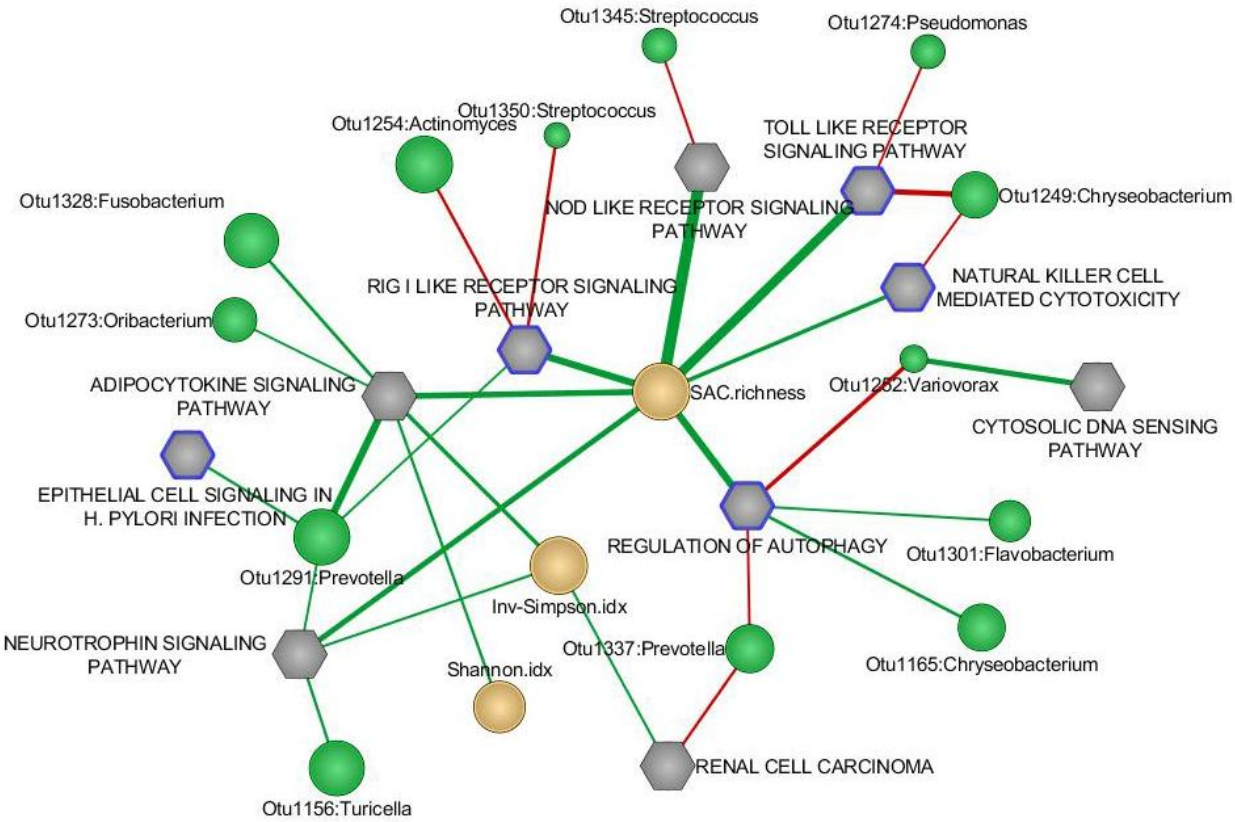
Figure 3. Network of shared canonical signaling pathways associated with CpG-ODN responsiveness, host gene co-expression modules, and microbial community features. An integrated approach was used to identify common canonical pathways (light blue circles) enriched from the following three paired data sets obtained from COMET-IPF patients: (1) fibroblasts response to CpG-ODN; (2) microbial diversity indices and OTUs; and (3) WGCNA host gene co-expression modules. Up-regulated genes between cases with “CpG-ODN response” and “CpG-ODN nonresponse” were subjected to Ingenuity Pathway Analysis (IPA) with criterion of FDR<0.01 to identify 22 significantly CpG-ODN response-associated canonical pathways. These 22 pathways were then matched to the canonical pathways associated with 10 host gene co-expression modules (grey squares) from PBMC microarray and microbial diversity indices and OTUs (purple diamonds) from bacterial 16S rRNA sequencing (Supplemental table S2 and S4, respectively). Fourteen pathways (light blue circles) were shared among CpG-ODN response (yellow hexagon) microbial features (purple diamonds) and host gene co-expression modules (grey squares). The red and blue

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

edges represent positive and negative association between nodes, respectively. Grey edges represent both positive and negative association between nodes. The width of the edge is proportional to the value of -log(*p*-value), i.e. thicker edge is more significant.

Figure 4. Correlation of circulating leukocyte phenotypes with microbial community. The correlation of microbial diversity indices and OTU abundance with diverse circulating leukocytes phenotypes (A) and CD28 (B) determined by Person's correlation algorithm are shown. In order to maintain positive or negative directional correlations, "r" coefficient instead of "r^2" was displayed on top in each box. Criterion for significance is set at coefficient >0.3 (red box) or <-0.3 (green box), and FDR <0.1 after multiple testing correction shown in parenthesis below each coefficient r value.

Figure 1



AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Figure 2

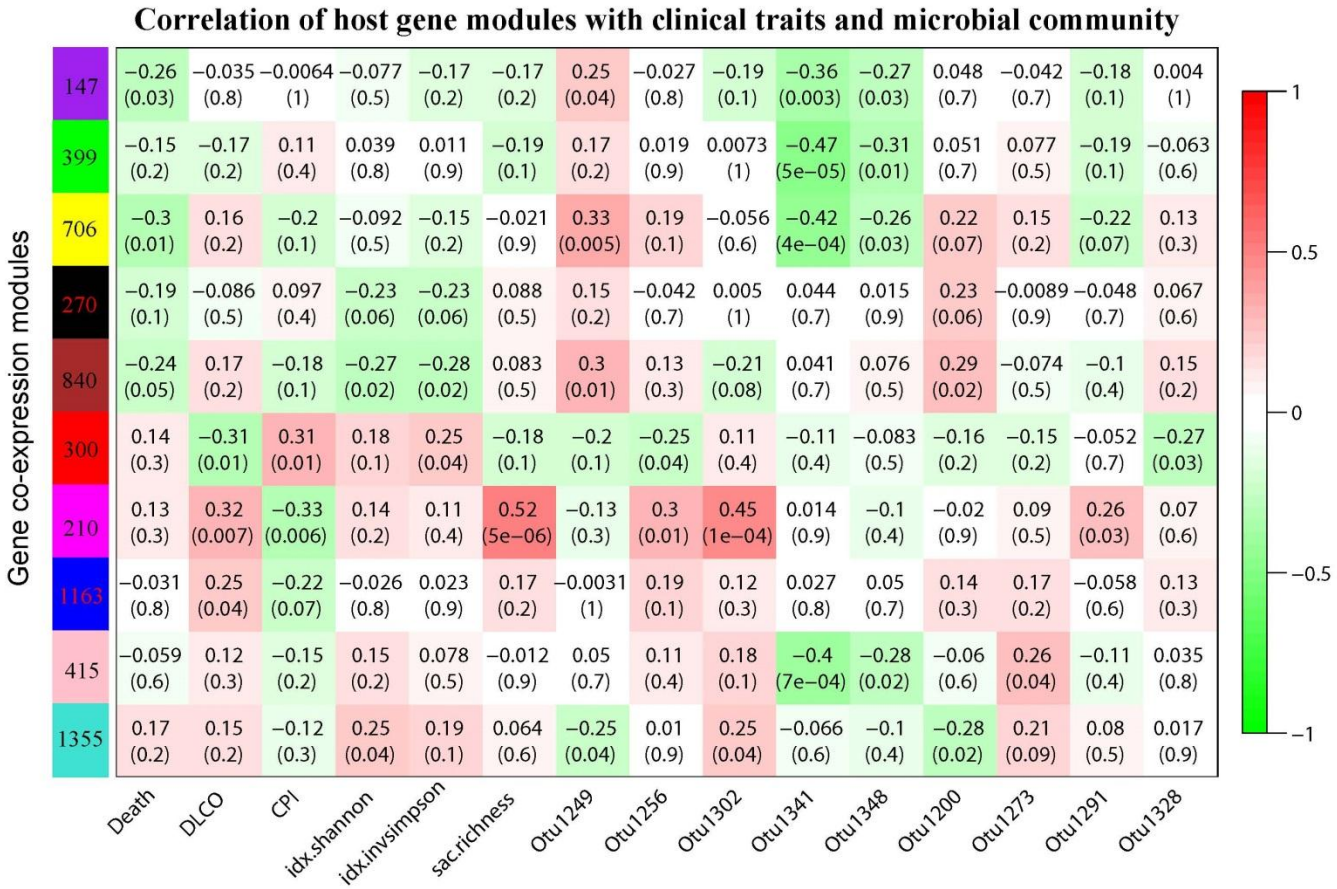
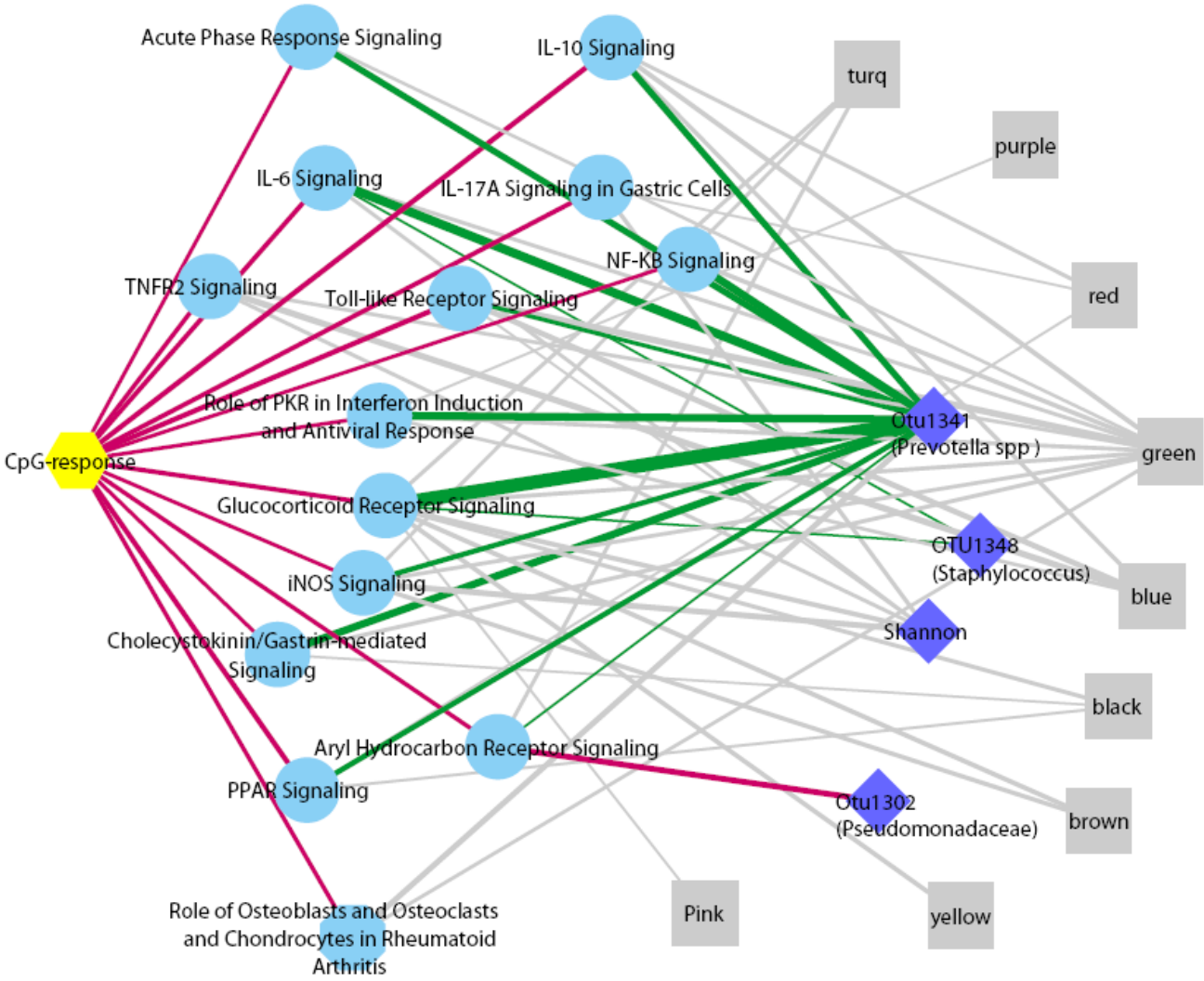


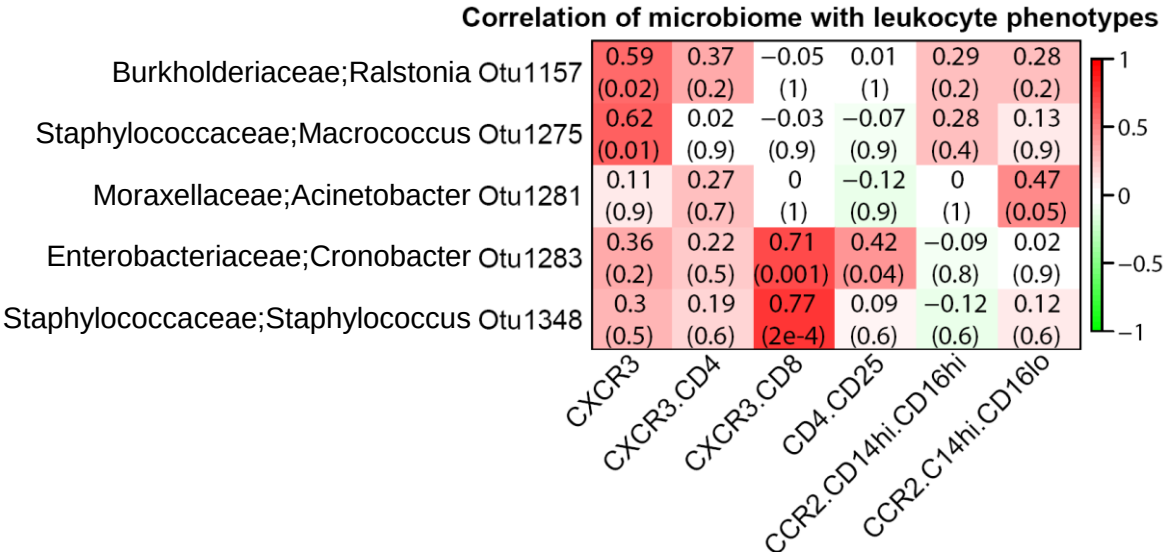
Figure 3



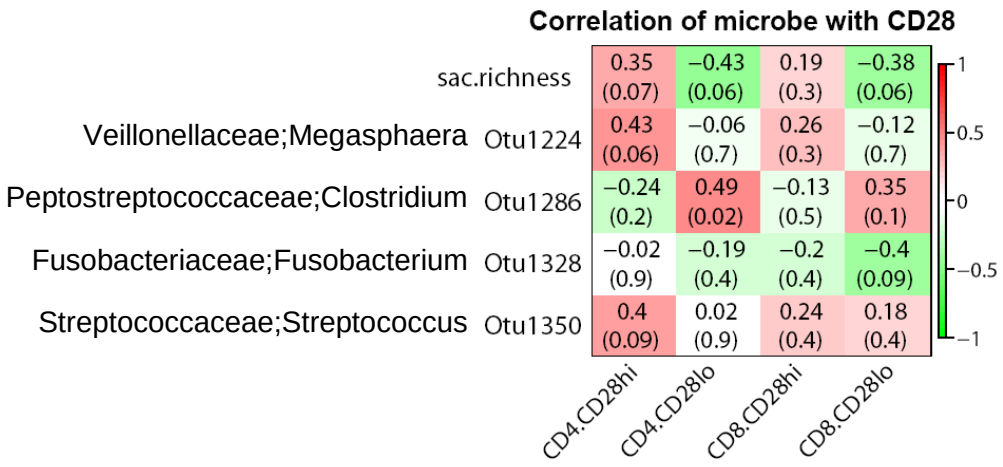
AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Figure 4

A



B



Supplemental Methods

Study Population

Subjects were all prospectively enrolled as a part of the COMET-IPF (Correlating Outcomes with biochemical Markers to Estimate Time-progression in Idiopathic Pulmonary Fibrosis), an observational study correlating biomarkers with disease progression (NCT01071707) (1). This multicenter study recruited subjects at nine clinical centers in the US. Inclusion and exclusion criteria of IPF diagnosis were previously described (2). The local Institutional Review Boards approved the study at each center, and informed consent was provided to all study subjects. COMET-IPF cohort data collection has been reported in three previous studies (1-3). Demographic information, clinical characteristics, and pulmonary function tests were collected from all IPF patients. Spirometry testing, including forced vital capacity (FVC), forced expiratory volume in 1 second (FEV1), and diffusion capacity for carbon monoxide (D_LCO), as well as lung volumes by plethysmography was obtained per ATS guidelines (4-6). The composite physiologic index (CPI) was calculated as described by Wells *et al.,* (7). The primary outcome (combined endpoint) for COMET-IPF was progression free survival (PFS) as determined by time until any of the following: death, acute exacerbation, lung transplant, or relative change in FVC, liters of ≥10% or D_LCO of ≥15% (2).

This study reports microbiome and peripheral blood mononuclear cell (PMBC) gene expression data on all subjects with supporting data on fibroblast

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

CpG oligonucleotide DNA (CpG-OND) responsiveness and circulating leukocyte phenotypes in subset cohorts .

PBMC sample collection, RNA isolation, microarray hybridization, and data processing

EDTA blood samples were collected from IPF patients enrolled in COMET cohort at study entry. PBMCs were isolated by Ficoll-Paque Plus (GE Healthcare Life Science, Pittsburgh, PA) as previously described (8) and lysed with TRIzol reagent (Thermo Fisher Sci., Waltham, MA). RNA was extracted following manufacturer's protocol and was re-precipitated by sodium acetate/ethanol prior to Affymetrix PrimeView™ array assay (Affymetrix, Santa Clara, CA) according to manufacturer's technical manual (http://media.affymetrix.com/support/downloads/manuals/expression_analysis_technical_manual.pdf). RNA quality and integrity were confirmed by Nanodrop (A260/A280 ratios between 1.7 and 2.2) and Bio-Analyzer mini-gel assay, respectively. One hundred fifty ng RNA per sample was reverse transcribed to single stranded cDNA, and then amplified to cRNA using Affymetrix GeneChip WT cDNA Synthesis Kit. Qualities and yields (exceeding 25 μg/ml and 1000 μg/ml, respectively) of cRNA after the first and second amplifications were all satisfactory prior to hybridization and scanning.

The microarray raw data (.cel files) were processed using R/Bioconductor package "affy". Background correction and gene expression intensities were summarized and normalized using "rma" algorithm (9). Microarray data were

compliant with MIAME (Minimum Information About a Microarray Experiment) guidelines and submitted to the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/) with the accession number GSE85268.

Bronchoscopic alveolar lavage (BAL), microbiome 16S assay and data processing

As previously described (2), bronchoscopy was performed at enrollment only in patients who were clinically stable and without evidence of active infection. Four instillations of 50 mL of sterile isotonic saline aliquots for a total of 200 ml were pooled. Genomic DNA was extracted from BAL pellets using Qiagen DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA). The V3-V5 hypervariable regions of the bacterial 16S rRNA gene were sequenced following manufacturer's protocols. To be consistent, we did not reanalyze microbiome data but used the summary data of OTU abundance provided (2) as input data for current study. We used "vegan" package in R which has most basic functions of diversity analysis, community ordination and dissimilarity analysis to compute the diversity index from .share file generated from Muthor pipeline. We used species accumulation model to evaluate species richness (sac.richness). The species accumulation model implements Kindt's exact accumulator that resembles rarefaction (10).

Lung fibroblasts culture

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

IPF diagnostic transbronchial biopsy samples were obtained from COMET study. IPF fibroblasts were cultured in DMEM (Lonza) containing 15% fetal calf serum (FCS) (Cell Generation), 100 IU penicillin, 100 µg/ml streptomycin (Mediatech), 292 µg/ml L-Glutamine (Mediatech), and 100 µg/ml of Primocin (InvivoGen) at 37ºC and 10% CO_2. Fresh medium was added to the fibroblasts every 2-3 days and the cells were passaged when they were 70-90% confluent.

Data analyses, integration, and statistics

PFS-associated host canonical signaling pathways

Normalized microarray gene expression data from PBMCs of COMET-IPF cohort (n=68) was transformed from a “gene by sample matrix” to a “pathway by sample matrix” using R/Bioconductor package "GSVA" (11). Gene Set Variation Analysis (GSVA) is a non-parametric, unsupervised method for estimating variation of gene set enrichment through the samples of a expression data set. GSVA performs a change in coordinate systems, transforming the data from a gene by sample matrix to a gene-set by sample matrix, thereby allowing the evaluation of pathway enrichment for each sample. This new matrix of GSVA enrichment scores facilitates applying standard analytical methods like functional enrichment, survival analysis, clustering, CNV-pathway analysis or cross-tissue pathway analysis, in a pathway-centric manner. Basically, GSVA algorithm performs the following steps: 1. Kernel estimation of the cumulative density function; 2. The expression-level statistic is rank ordered for each sample; 3. For every gene set, the Kolmogorov-Smirnov-like rank statistic is calculated; 4. The GSVA

enrichment score is summarized to represent either the difference between the two sums or the maximum deviation from zero. Canonical pathways in Molecular Signatures Database (MSigDB; http://www.broadinstitute.org/gsea/msigdb/index.jsp), a collection of annotated gene sets, were used for GSVA. PFS-associated pathways were determined using Cox proportional hazard regression (Cox-PH) with criterion of Wald tests *p*-value < 0.05. The significant canonical pathways were further correlated with microbial measures using Pearson's correlation assay. We also conducted a sensitivity analysis by removing 4 death and 5 patients with acute exacerbation or by randomly removing 9 patients from the COMET-IPF cohort.

Correlation of gene co-expression modules with clinical traits and Operational Taxonomy Units (OTUs)

Normalized microarray data were filtered to remove probe sets with minimum variation (i.e. coefficient of variation <0.3 across all samples). For redundant probe sets targeting on the same gene, only the probe set with largest coefficient of variation is used for downstream data analysis. Genes that passed the filtering criteria were clustered into gene modules, based on their co-expression pattern, using an unsupervised "Weighted gene co-expression network analysis (WGCNA)" package in R (12). Principle Component Analysis (PCA) was used to calculate an eigengene for each gene module. Pearson's correlation was used to determine the significance of correlation (p<0.05) between the eigengenes of

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

individual gene module with clinical traits, including race, sex, age, FVC, DLCO, CPI, and prognosis, as well as microbial diversity and OTU abundance.

Functional pathways enrichment analysis

Significant canonical pathways or gene interaction networks were analyzed using Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Redwood City, CA). In general, pathways were considered to be overrepresented when one-sided Fisher's exact test p <0.05. In circumstances of large number of significant pathways, we further refine the criterion by adjustment of p-value using Benjamini-Hochberg False Discovery Rate (FDR).

Statistical analysis of Microarray data

Genes correlated with the OTU abundance were identified using Significant Analysis of Microarray (SAM) software (13). The criteria of significance is set at False Discovery Rate (FDR) <10%. Differentially expressed genes between CpG-ODN responsive and non-responsive patients were identified using empirical Bayes-moderated *t*-statistics test implemented in R/Bioconductor package "limma' (14). The criteria of significance is set at p<0.05.

Cytospace Network analysis

We used Cytospace platform to visualize molecular interaction networks that incorporate host canonical pathways in co-expressed gene modules and microbial community features (http://www.cytoscape.org/). The significantly

enriched pathways, host gene modules, microbial features including OTU and diversity indices, were represented by nodes in the network, while their relationships were represented by edges. The quantitative traits of the nodes and edges were also included in the network construction by Cytospace.

Network statistics and topological parameters were computed using Network Analysis (15). The neighborhood N of a given node is the set of its neighbors and the connectivity of node is unaffected by the existence of self-loops (16). Closeness-centrality uses information about the length of the shortest paths within a network; it uses the sum of the minimal distances of a node to all other nodes. Shortest path betweenness-centrality quantifies the ability of a node to monitor communication between other vertices. Every node that is part of a shortest path between two other nodes monitors communication or flow between them. Central node monitors many communications between other nodes (17).

Fibroblast responsiveness to TLR9 stimulation using CpG-ODN.

Transbronchial biopsies specimens were available in a subset of COMET-IPF patients (n=27). Proliferating fibroblast cells at passage 6-10 were plated in complete media decribed above for overnight and then stimulated with or without 10μM hypomethylated CpG-oligonucleotide DNA (CpG–ODN) for 24h. CpG-ODN containing a phosphotorothioate backbone (GpC-ODN) was used as a negative control for the TLR9 stimulation in our experiments. Nonspecific activation mediated by GpC-ODN was not observed in lung fibroblasts. For gene expression analysis, TRIzol reagent was added to each well and RNA extraction

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

was performed according to the manufacturer's instructions. RNA concentration and purity were determined by Nanodrop. Purified RNA was subsequently reverse-transcribed into cDNA. Alpha smooth muscle actin (α-SMA) gene expression was analyzed by a real-time quantitative RT-PCR procedure using ViiA™ 7 Real-Time PCR System machine (Applied Biosystems, Foster City, CA). Forward and reverse primers' sequence for α-SMA were 5'GCGTGGCTATTCCTTCGTTACT3' and 5'GCTACATAACACAGTTTCTTCTTGATG3', respectively. 18S rRNA was used as an internal control. Primers and probe used for 18S (Hs03928990_g1) were purchased from Applied Biosystems. Gene expression was normalized to 18S and the fold increase was calculated over untreated fibroblasts (18). Since we observed that a greater than 2-fold increase in alpha smooth muscle actin (αSMA) transcript expression is directly correlated with a significant increase in the expression of αSMA protein in the same cells, a statistical analysis was performed at the protein level (measured via Western blot) to validate the selection of the >2 fold increase at the transcript level observed in these cells. α-SMA expression (i.e. treated vs. untreated) > 2 fold was deemed "responsive". Logit regression with univariate and multivariate analysis was conducted to determine taxa associated with CpG-ODN responsiveness. Support Vector Machine (SVM) followed by Leave-One-Out (LOO) was used to construct a multi-taxa microbial model predictive of CpG-ODN responsiveness.

Correlation of circulating leukocyte phenotypes with microbial measures

Peripheral blood of 32 COMET-IPF patients was collected in EDTA-containing vacutainers at study centers and processed as described previously (3). The leukocyte phenotypes were quantified using flow cytometry analyses as previously described (3). Cells were stained at $1x10^6$/well with antibodies: APC Mouse anti-Human CD4, FITC Mouse Anti-Human CD8, PE Mouse Anti-Human CD28, PE-Cy7 Mouse anti-Human CD14, and PE Mouse Anti-Human CD16. All antibodies were purchased from BD Biosciences, Franklin, NJ. Patients with available paired leukocyte phenotypes and microbial community features from 16s rRNA amplicon sequencing data were included in Pearson's correlation analysis. Criterion for significance is set at coefficient>0.3 or <-0.3, and FDR<0.1 after multiple testing adjustment.

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

References

1. Naik PK, Bozyk PD, Bentley JK, Popova AP, Birch CM, Wilke CA, Fry CD, White ES, Sisson TH, Tayob N, Carnemolla B, Orecchia P, Flaherty KR, Hershenson MB, Murray S, Martinez FJ, Moore BB, Investigators C. Periostin promotes fibrosis and predicts progression in patients with idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* 2012; 303: L1046-1056.
2. Han MK, Zhou Y, Murray S, Tayob N, Noth I, Lama VN, Moore BB, White ES, Flaherty KR, Huffnagle GB, Martinez FJ, Investigators C. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir Med* 2014; 2: 548-556.
3. Moore BB, Fry C, Zhou Y, Murray S, Han MK, Martinez FJ, Flaherty KR, The CI. Inflammatory leukocyte phenotypes correlate with disease progression in idiopathic pulmonary fibrosis. *Front Med* 2014; 1.
4. Macintyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CP, Brusasco V, Burgos F, Casaburi R, Coates A, Enright P, Gustafsson P, Hankinson J, Jensen R, McKay R, Miller MR, Navajas D, Pedersen OF, Pellegrino R, Wanger J. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *The European respiratory journal* 2005; 26: 720-735.
5. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, Jensen R, Johnson DC, MacIntyre N, McKay R, Navajas D, Pedersen OF, Pellegrino R, Viegi G, Wanger J. Standardisation of spirometry. *The European respiratory journal* 2005; 26: 319-338.
6. Wanger J, Clausen JL, Coates A, Pedersen OF, Brusasco V, Burgos F, Casaburi R, Crapo R, Enright P, van der Grinten CP, Gustafsson P, Hankinson J, Jensen R, Johnson D, Macintyre N, McKay R, Miller MR, Navajas D, Pellegrino R, Viegi G. Standardisation of the measurement of lung volumes. *The European respiratory journal* 2005; 26: 511-522.
7. Wells AU, Desai SR, Rubens MB, Goh NS, Cramer D, Nicholson AG, Colby TV, du Bois RM, Hansell DM. Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. *Am J Respir Crit Care Med* 2003; 167: 962-969.
8. Herazo-Maya JD, Noth I, Duncan SR, Kim S, Ma SF, Tseng GC, Feingold E, Juan-Guardela BM, Richards TJ, Lussier Y, Huang Y, Vij R, Lindell KO, Xue J, Gibson KF, Shapiro SD, Garcia JG, Kaminski N. Peripheral blood mononuclear cell gene expression profiles predict poor outcome in idiopathic pulmonary fibrosis. *Sci Transl Med* 2013; 5: 205ra136.
9. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; 4: 249-264.
10. Ugland KI, Gray JS, Ellingsen KE. The species-accumulation curve and estimation of species richness. *Journal of Animal Ecology,* 2003; 72: 888-897.
11. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013; 14: 7.

12. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 559.
13. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001; 98: 5116-5121.
14. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004; 3: Article3.
15. Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, Albrecht M. Computing topological parameters of biological networks. *Bioinformatics* 2008; 24: 282-284.
16. Maslov S, Sneppen K. Specificity and stability in topology of protein networks. *Science* 2002; 296: 910-913.
17. Koschutzki D, Schreiber F. Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regul Syst Bio* 2008; 2: 193-201.
18. Trujillo G, Meneghin A, Flaherty KR, Sholl LM, Myers JL, Kazerooni EA, Gross BH, Oak SR, Coelho AL, Evanoff H, Day E, Toews GB, Joshi AD, Schaller MA, Waters B, Jarai G, Westwick J, Kunkel SL, Martinez FJ, Hogaboam CM. TLR9 differentiates rapidly from slowly progressing forms of idiopathic pulmonary fibrosis. *Sci Transl Med* 2010; 2: 57ra82.

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Supplemental Figure Legends

Supplemental Figure E1. Schematic data structures and data analysis pipelines. We integrate the following data according to our analysis pipelines. A subset (n=68) of COMET-IPF patients with paired microarray gene expression data from peripheral blood mononuclear cells (PBMCs), 16s rRNA sequencing data from bronchoscopic alveolar lavage (BAL), and clinical phenotypes. A subset (n=32) of these 68 COMET-IPF patients with paired circulating leukocyte phenotypes data and 16s rRNA sequencing data from BAL. A subset (n=27) of them has paired CpG-oligodeoxynucleotide (CpG-ODN) response to TLR9 stimulation in lung fibroblasts derived from transbronchial biopsies data and microarray gene expression data from host PBMCs .

Supplemental Figure E2. Kaplan-Meyer plot of COMET-IPF patients based on 10% FVC decline, 15% D_LCO decline, death, transplant and acute exacerbation.

Supplemental Figure E3. Correlation of magenta module membership with gene significance. Each circle represents a gene in magenta module. Module Membership (X-axis) was computed as correlation coefficient of gene expression with magenta module eignegene across all samples. Gene Significance (Y-axis) was computed as correlation coefficient of gene expressions with CPI (**E3A**) or microbial richness (**E3B**).

Supplemental Figure E4. Ingenuity pathway analysis of the magenta gene module. Canonical pathways were enriched from genes in magenta module. Significant pathways were determined using one-tailed Fisher's exact test with criterion of FDR < 0.05 (i.e. -log (*FDR*)>1.3).

Supplemental Figure E5. Correlation of Toll-like receptors (TLRs) and toll interacting protein (*TOLLIP*) expression in PBMC with microbial community features. Correlation of the gene expression levels of *TLR1-10* and *TOLLIP* in peripheral blood mononuclear cells (PBMCs) with OTU abundance in BAL is determined with Pearson's correlation with multiple testing corrections. The criteria of significance were set at correlation coefficient “r” value > 0.3 (red box) or < -0.3 (green box) and displayed on top, and FDR < 0.1 displayed in parenthesis in each box. The most significant correlation was found between TLR9 with OTU1341 (*Prevotella*, r=0.48, FDR=0.0003) and OTU1348 (*Staphylococcus*, r=0.038, FDR=0.02).

Supplemental Figure E6. Kaplan-Meier plot of PFS in the CpG-ODN responsive and CpG-OND non-responsive fibroblasts derived from transbronchial biopsies of IPF patients.

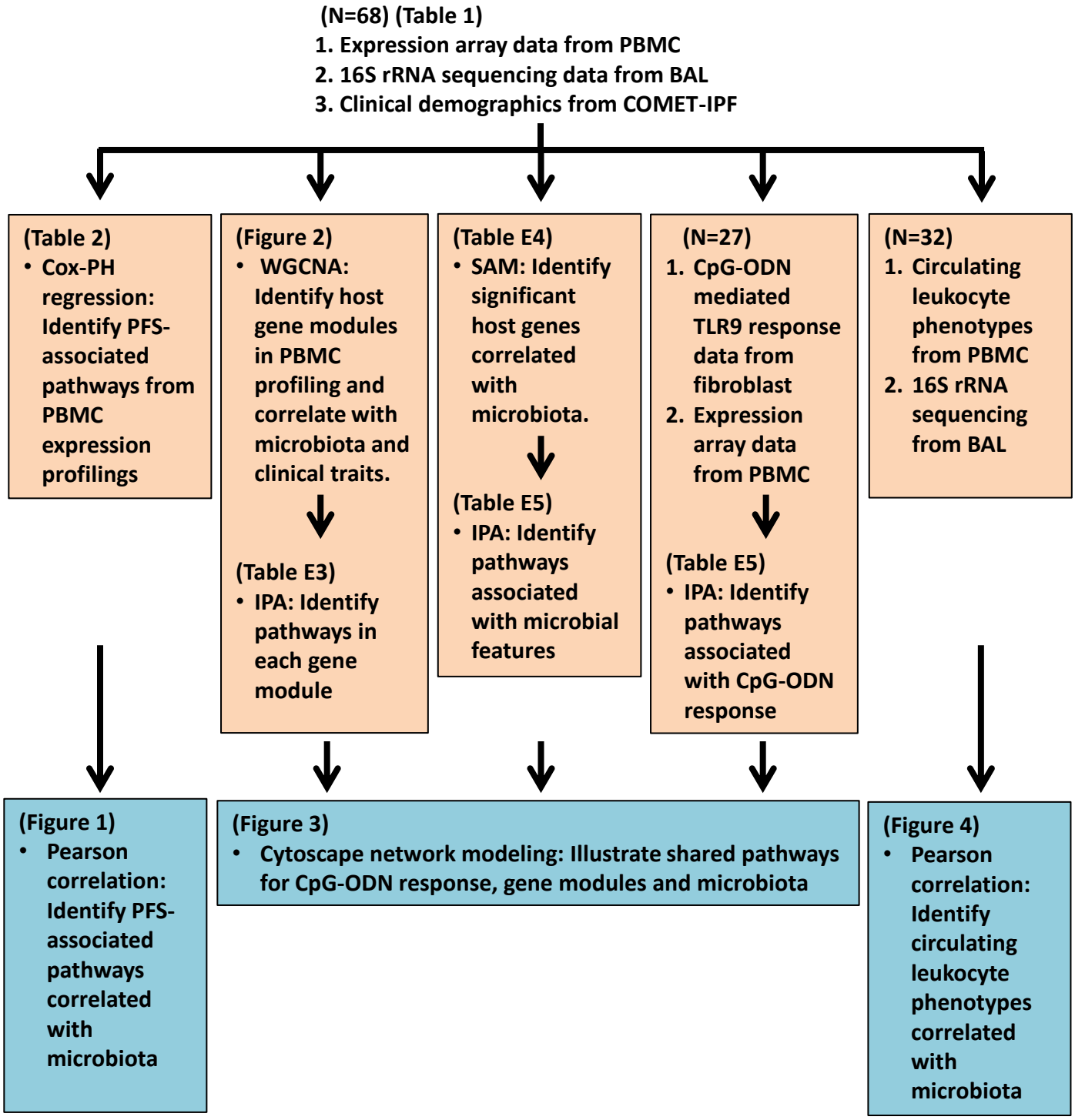
Supplemental Figure E7. Canonical pathways enriched from upregulated genes in PBMCs microarray data comparing CpG-ODN responsive cases versus CpG-ODN nonresponsive cases. A subset of PBMC microarray gene

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

expression data corresponding to 16 CpG-ODN responsive and 11 CpG-ODN nonresponsive cases, respectively, were selected from COMET-IPF microarray dataset (n=68). The CpG-ODN stimulated fibroblasts were compared to their own BAL. Empirical Bayes-moderated *t*-test revealed significantly differentially expressed genes (93 up- and 308 down-regulated) between CpG-ODN responsive and CpG-ODN nonresponsive cases with *p*-value < 0.05. Ingenuity Pathway Analysis of the up-regulated genes revealed 22 significantly CpG-ODN response-associated canonical pathways with one-tailed Fisher's exact FDR ≤0.01 (i.e. -log (FDR)≥2, roundup).

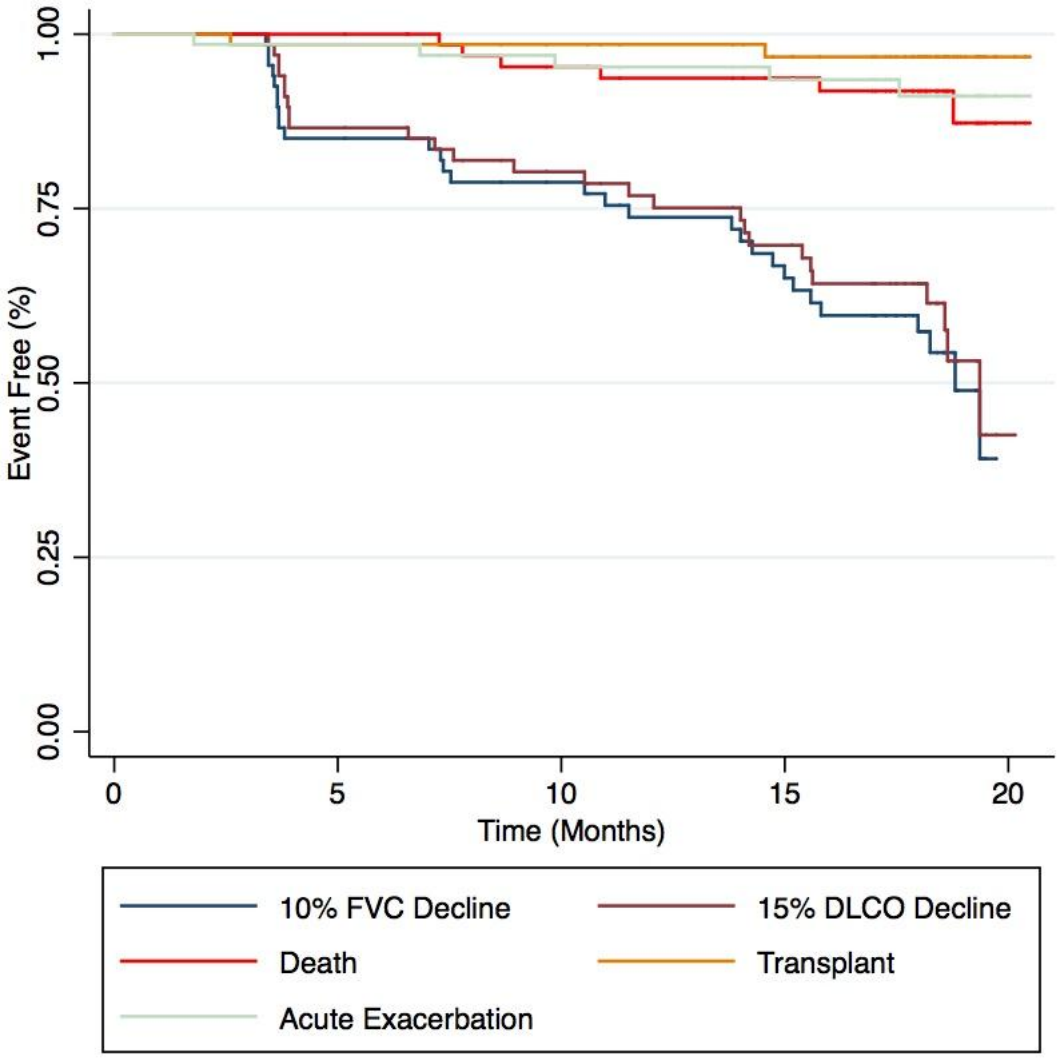
Supplemental Figure E8. Statistical analysis of CpG-ODN response network. Network construction was described in Figure 3. Statistical analysis of the network parameters was performed. **A.** Correlation of betweenness-centrality with number of neighbor nodes. Betweenness centrality measures the number of shortest paths going through a node. Therefore nodes with high betweenness centrality such as glucocorticoid receptor signaling and OTU1341 (*Prevotella*) demonstrate a shortcut of this network. **B.** Correlation of closeness centrality with number of neighbor nodes. Nodes with highest closeness centrality in our network are physically nearest nodes to all nodes.

Supplemental Figure E1

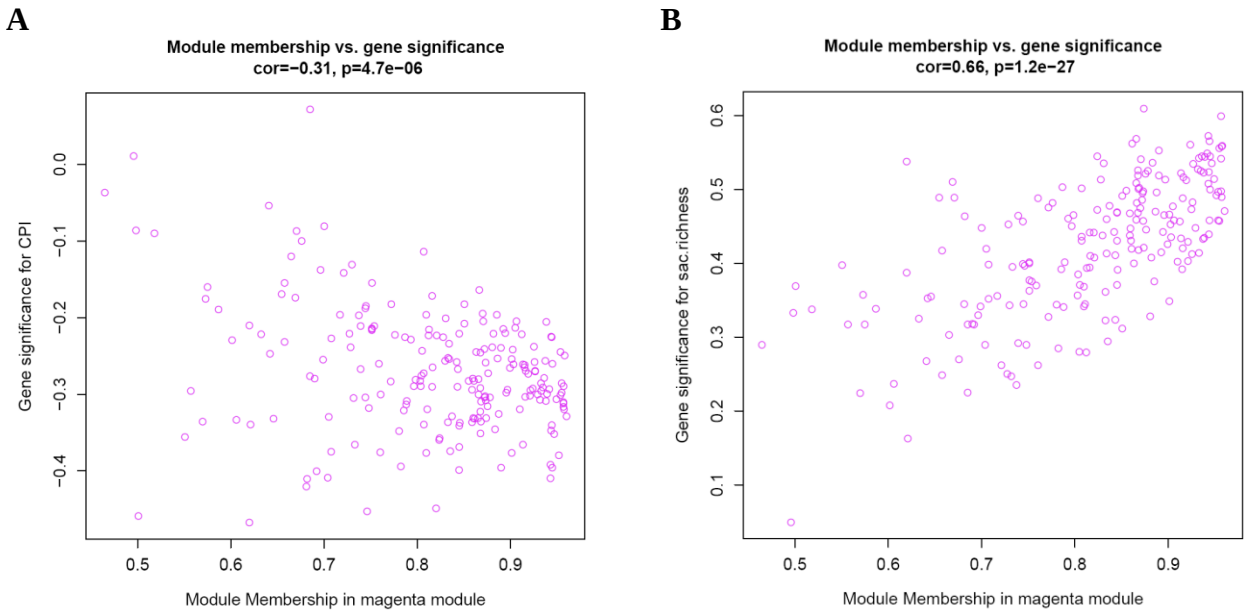


AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Supplemental Figure E2

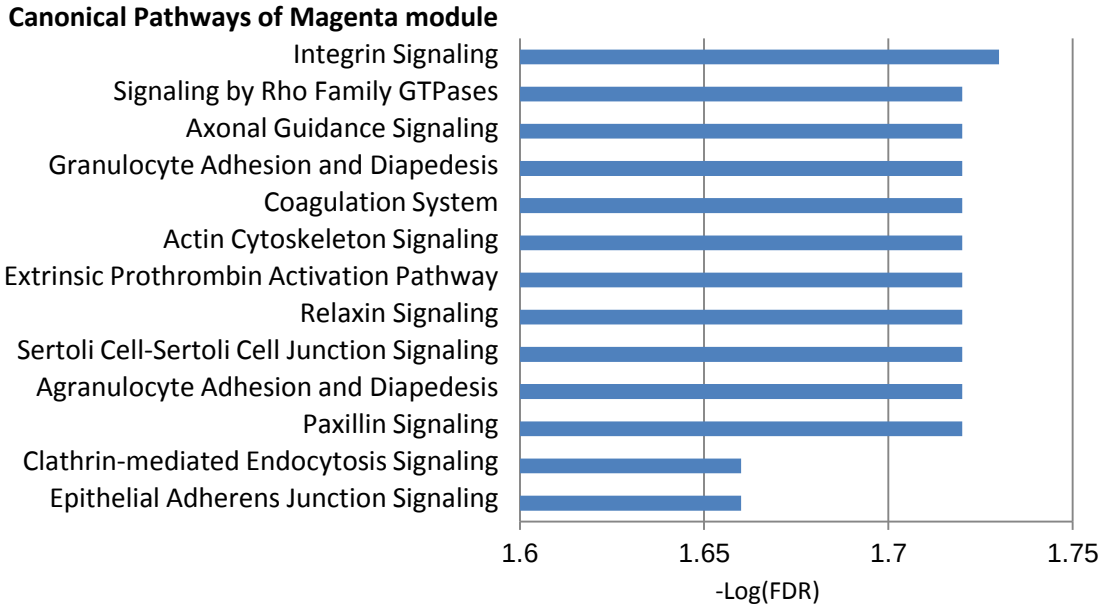


Supplemental Figure E3

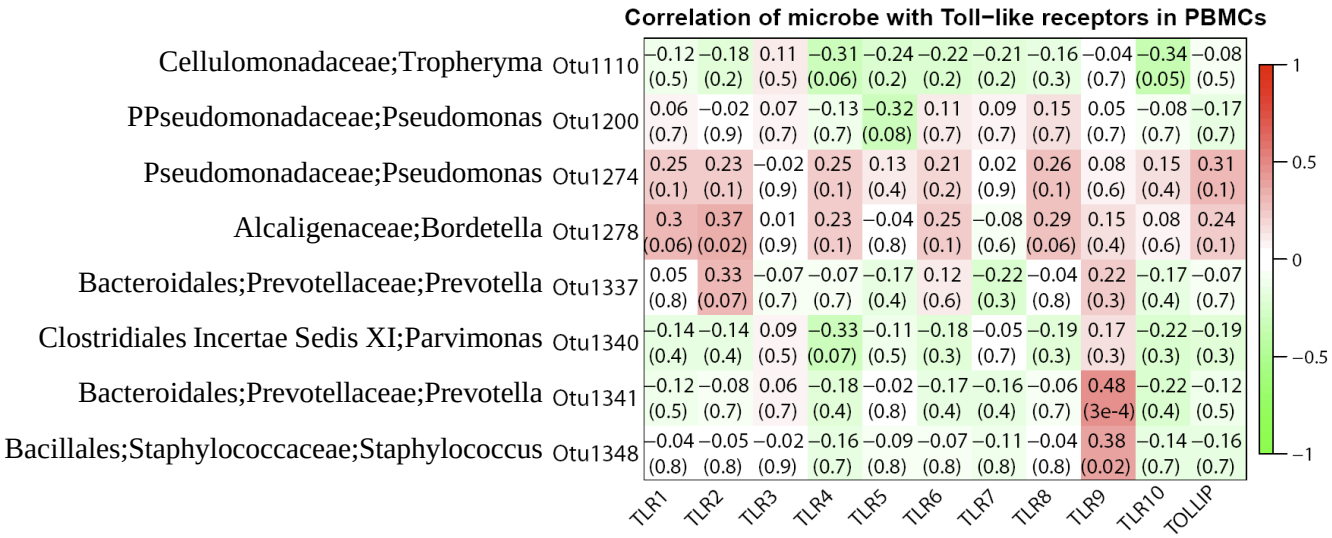


AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Supplemental Figure E4

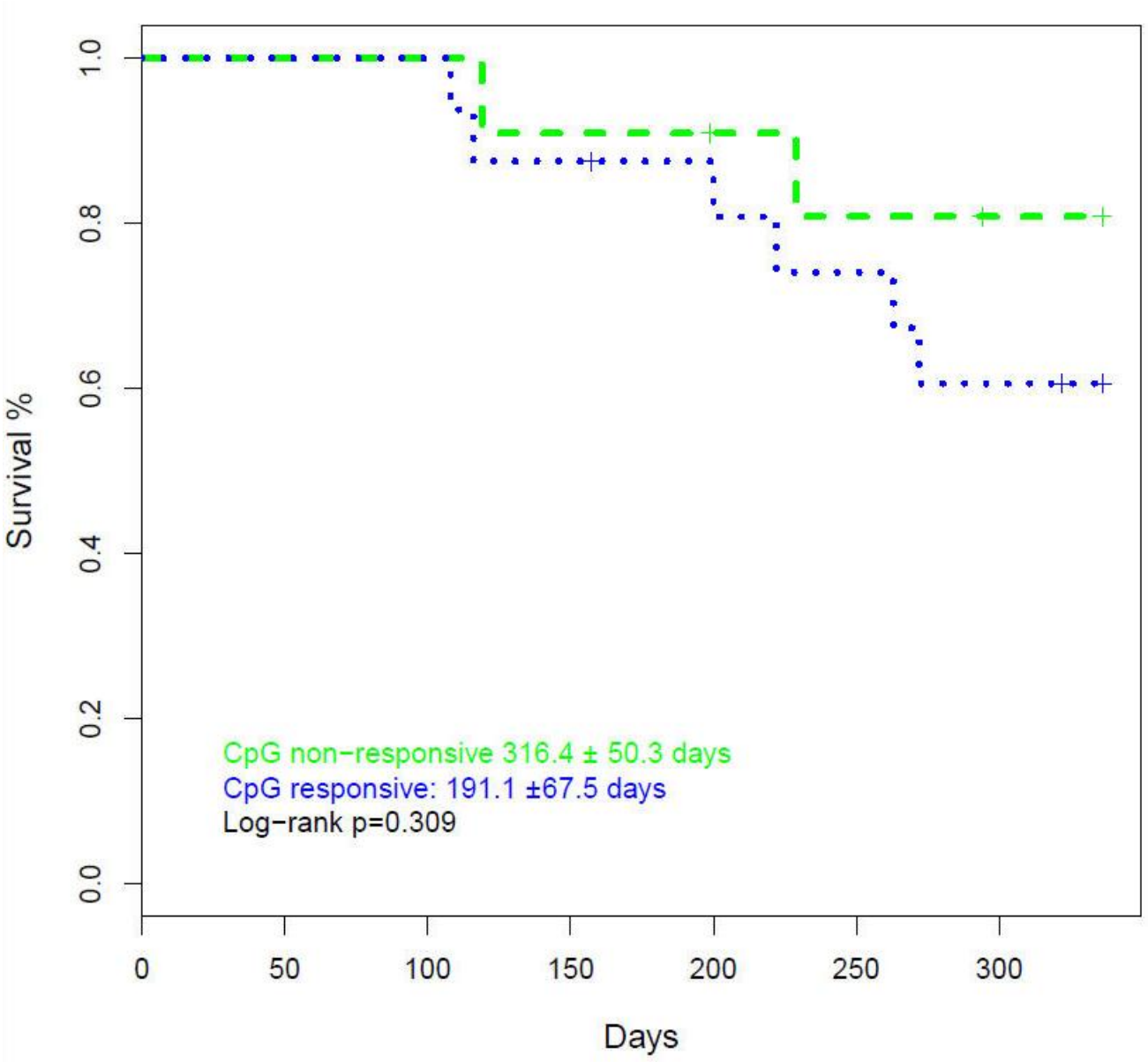


Supplemental Figure E5

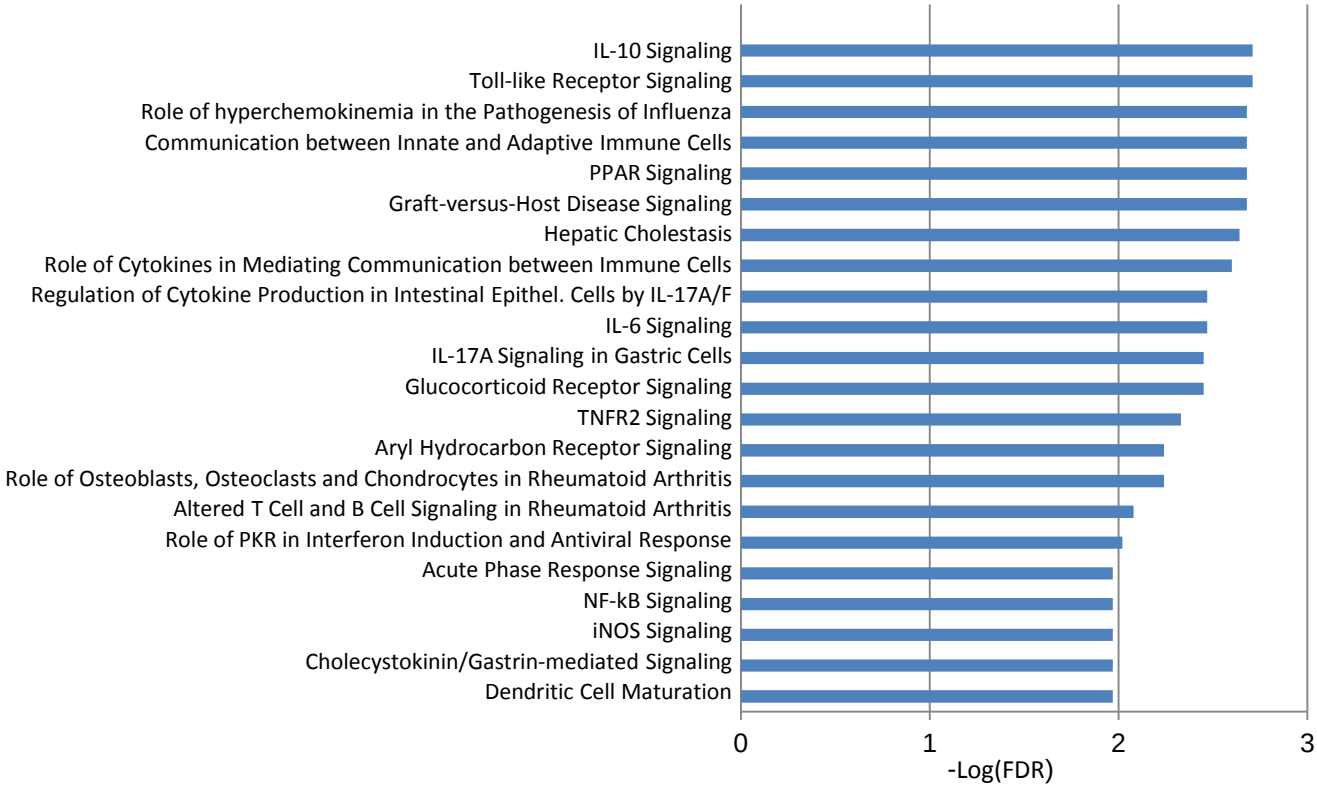


AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Supplemental Figure E6



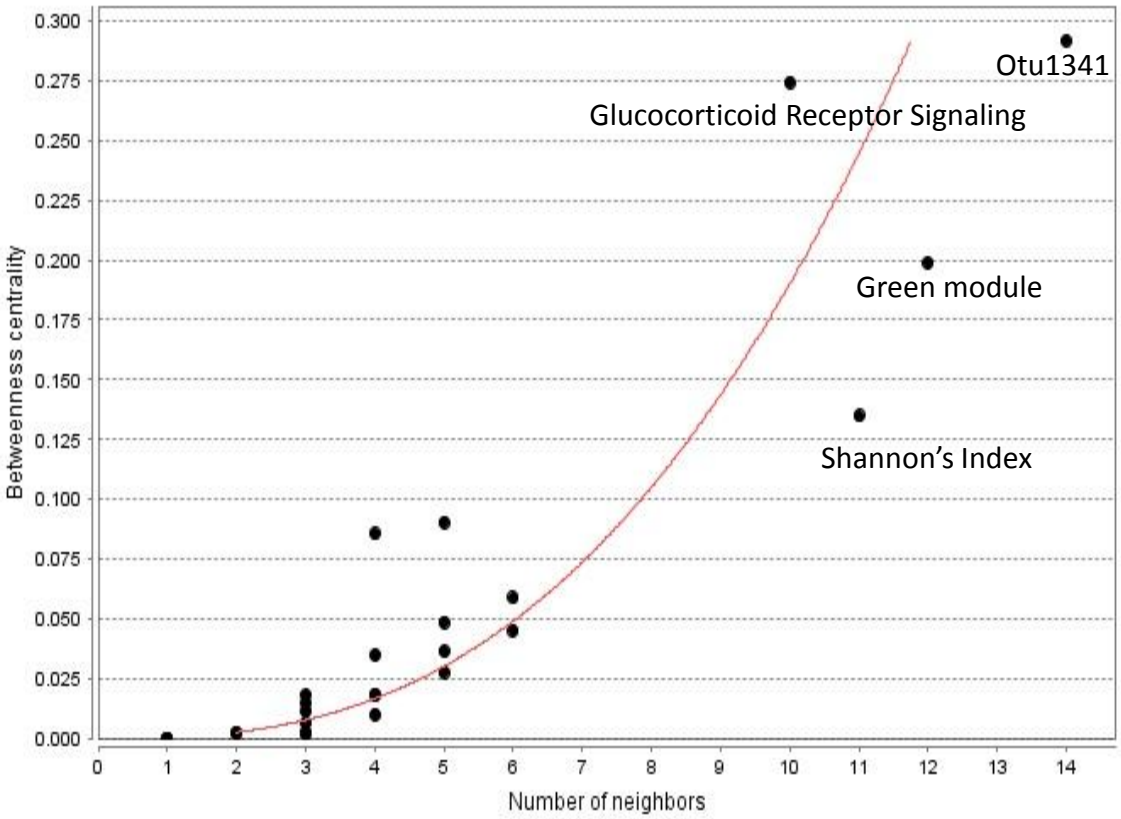
Supplemental Figure E7



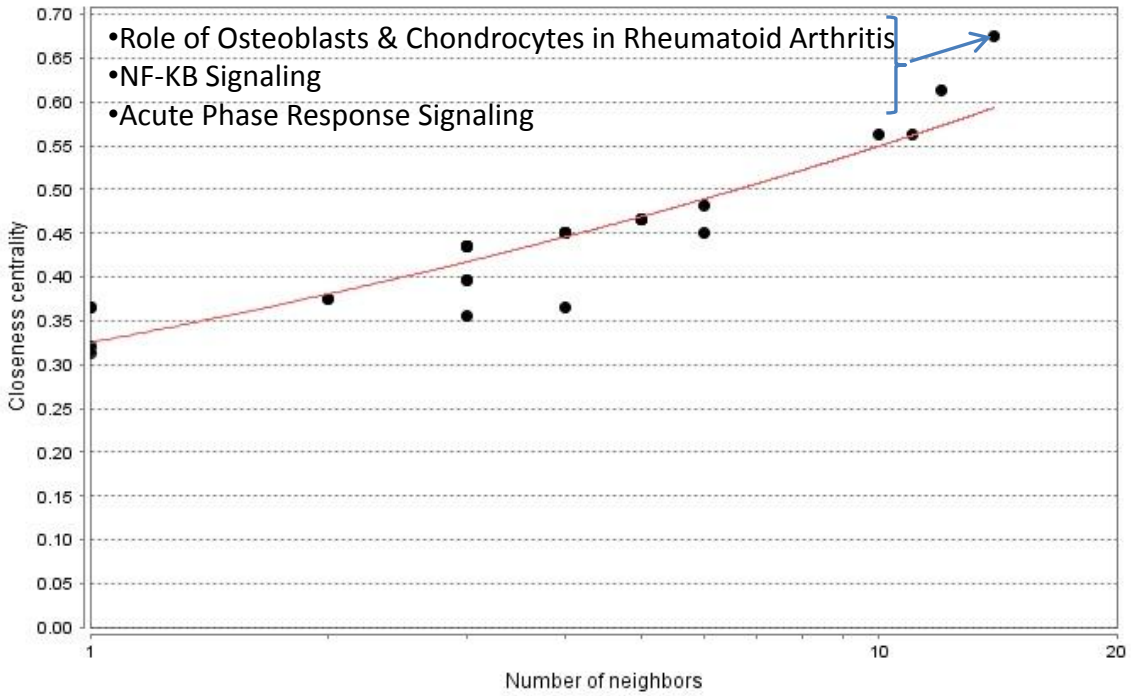
AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Supplemental Figure E8

A



B



Supplemental Table E1. Composite Endpoint Events in COMET cohort (n=68)

Event	n	Median time to event [IQR]
10% FVC Decline	26	9.73 [3.88-15.39]
15% D_LCO Decline	29	10.52 [3.68-14.99]
Death	6	9.77 [7.79-15.78]
Lung Transplant	2	8.58 [2.60-14.56]
Acute Exacerbation	5	9.86 [6.84-14.66]

Abbreviations: FVC = forced vital capacity; D_LCO = diffusion capacity of the lung for carbon monoxide

AJRCCM Articles in Press. Published on 03-February-2017 as 10.1164/rccm.201607-1525OC

Supplemental Table E7. Prediction of CpG-responsiveness by microbial taxa*

Taxa	%CV	Ratio
OTU1331 (*Veillonella*)	100	0.085
OTU1350 (*Streptococcus*)	96	0.15
OTU1273 (*Oribacterium*)	63	0.21
OTU1327 (*Prevotella*)	85	0.36
OTU1260 (*Prevotella*)	85	1.05
OTU1345 (*Streptococcus*)	100	4.07
OTU1224 (*Megasphaera*)	63	3.57
OTU1157 (*Ralstonia*)	56	1.57
OTU1341 (*Prevotella*)	93	3.04
OTU1337 (*Prevotella*)	96	4.46

*Iterative supporting vector machine (SVM) followed by Leave-One-Out (LOO) prioritized 10 OTUs that demonstrated 75% sensitivity and 82% specificity in prediction of CpG-responsiveness

Supplemental Table E8. Progression free survival (PFS)-associated host canonical pathways identified by Cox-PH regression analysis*

	N=68		N=59*	
	Odd Ratios	Wald p-value	Odd Ratios	Wald p-value
TOLL-LIKE RECEPTOR SIGNALING PATHWAY	0.046	0.003	0.037	0.009
REGULATION OF AUTOPHAGY	0.022	0.005	0.018	0.007
EPITHELIAL CELL SIGNALING IN HELICOBACTER PYLORI INFECTION	0.027	0.005	0.042	0.010
RIG-I-LIKE RECEPTOR SIGNALING PATHWAY	0.175	0.039	0.013	0.009
NATURAL KILLER CELL MEDIATED CYTOTOXICITY	0.049	0.022	0.026	0.019
RENAL CELL CARCINOMA	0.035	0.009	0.061	0.026
CYTOSOLIC DNA SENSING PATHWAY	0.052	0.005	0.147	0.227
ADIPOCYTOKINE SIGNALING PATHWAY	0.037	0.017	0.022	0.020
NEUROTROPHIN SIGNALING PATHWAY	0.073	0.015	0.078	0.108
CHEMOKINE SIGNALING PATHWAY	0.038	0.017	0.089	0.091
NOD-LIKE RECEPTOR SIGNALING PATHWAY	0.065	0.025	0.284	0.213

*Sensitivity analysis after removing 4 death and 5 patients with acute exacerbation from the COMET-IPF cohort