Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Microbe. Author manuscript; available in PMC 2024 January 23.

Published in final edited form as:

Microbe. 2023 December ; 1: . doi:10.1016/j.microb.2023.100023.

Integrating airway microbiome and blood proteomics data to identify multi-omic networks associated with response to pulmonary infection

Brenton I.M. Graham[a], **J. Kirk Harris**[b], **Edith T. Zemanick**[b], **Brandie D. Wagner**[a,b,*]

[a]Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[b]Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Abstract

Host response to airway infections can vary widely. Cystic fibrosis (CF) pulmonary exacerbations provide an opportunity to better understand the interplay between respiratory microbes and the host. This study aimed to investigate the observed heterogeneity in airway infection recovery by analyzing microbiome and host response (i.e., blood proteome) data collected during the onset of 33 pulmonary infection events. We used sparse multiple canonical correlation network (SmCCNet) analysis to integrate these two types of -omics data along with a clinical measure of recovery. Four microbe–protein SmCCNet subnetworks at infection onset were identified that strongly correlate with recovery. Our findings support existing knowledge regarding CF airway infections. Additionally, we discovered novel microbe–protein subnetworks that are associated with recovery and merit further investigation.

Keywords

Cystic fibrosis; Pulmonary exacerbation; Microbe-protein networks; Sparse Multiple Canonical Correlation Network; (SmCCNet) Analysis; Host response; Microbiome data integration; Aptamers

1. Introduction

Mucosal barriers are important for maintaining human health, they evolved to co-exist in the presence of commensal bacteria while recognizing and responding to pathogens. Exposure to micro-organisms in a normal respiratory tract can induce an inflammatory response or the offending organisms can often be cleared prior to eliciting a response. Moreover, the

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

*Correspondence to: 13000 E 17th place, B119, Aurora, CO 80045, USA. Brandie.Wagner@cuanschutz.edu (B.D. Wagner).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.microb.2023.100023.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

number of bacteria found in the airway can be large and the resulting host-response can be multi-factorial, complicating the interplay between the airway microbiome and the host inflammatory response.

In a lung disease like cystic fibrosis (CF), ineffective clearance mechanisms and a less regulated inflammatory response results in chronic respiratory infections and periods of acute infections called pulmonary exacerbations (PExs). Individuals who experience frequent and recurring PEx episodes often sustain long-term deterioration in lung function and shortened survival (Amadori et al., 2009; Liou et al., 2001; Sanders et al., 2011). Intravenous (IV) antibiotic treatment is typically administered at the onset of a CF PEx to combat bacterial airway infection. PEx recovery is variable, however, as many patients do not return to baseline health (Sanders et al., 2010a, 2010b). These episodes of acute infection provide an opportunity to evaluate the host-response to airway infections. Blood and airway inflammatory biomarkers have been shown to decrease after treatment of a PEx, suggesting that airway infection in CF results in a robust host immune response (Ordoñez et al., 2003; Colombo et al., 2005; Chiron et al., 2008; Sagel et al., 2015). Further, exploring microbe–protein relationships in the context of PEx recovery may provide novel insights into how the host responds to different bacterial infections and the mechanisms of PEx.

In this study, we use sparse multiple canonical correlation network (SmCCNet) to integrate plasma proteomics and airway microbiome data collected at the onset of PExs (Shi et al., 2019). Specifically, these omics data are integrated into interpretable microbe–protein subnetworks that are associated with PEx recovery. Our aim was to identify microbe–protein subnetworks at PEx onset that are associated with PEx recovery.

2. Materials and methods

2.1. Study design and cohort

This study was conducted between 2010 and 2012 and includes 33 PEx events from a cohort of 29 subjects aged 10–22 years old with a confirmed diagnosis of CF. Participants could re-enroll in the study if PEx events were separated by at least six months. Accordingly, the study population includes four subjects with two PEx events and 26 subjects with one PEx event. Participants were recruited prospectively and enrolled at the time of hospital admission for IV antibiotic therapy of a clinically diagnosed PEx. Study participants were treated with IV antibiotics, selected by the treating physician targeting their individual CF pathogens, and aggressive mucus clearance based on standard CF clinical care guidelines. The study was approved by the Colorado Multiple IRB (COMIRB #07–0365). Written informed consent and HIPPA authorization was obtained from all patients ≥ 18 years or from parents or legal guardians of patients younger than 18 years. Assent was obtained from patients between 10 and 17 years.

2.2. Clinical phenotype definition

The PEx scoring system is a systematic definition of PEx using patient symptoms (e.g., change in exercise tolerance, cough, sputum production, chest congestion, and school or work attendance) and physical examination measures (e.g., change in FEV_1 over the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

preceding month) (Rosenfeld et al., 2001). Higher PEx scores correspond to increased PEx severity. A PEx corresponds to a PEx score ≥ 5 and at least three out of 11 PEx criteria being met, as defined by the CF Foundation (CFF) Consensus Conference guidelines. The PEx score was collected at two time points: hospital admission (i.e., PEx onset, day 0–2) and hospital discharge (i.e., after the IV treatment had been administered, day 4–21). The phenotype of interest in this study is percent change in PEx score (%ΔPExS) between hospital admission (t_1) and hospital discharge (t_2). %ΔPExS is defined in Eq. 1. Negative %ΔPExS values represent an improved PExS between hospital admission and discharge, with larger negative values indicating greater improvement.

$$\%\Delta PExS = \frac{\Delta PExS_{t_2} - \Delta PExS_{t_1}}{\Delta PExS_{t_1}} \times 100\% \quad (1)$$

2.3. Microbiome data

2.3.1. Sputum sample collection—Spontaneously expectorated sputum was collected into a sterile container for microbiologic analysis. Participants unable to spontaneously expectorate underwent sputum induction using a standardized, published protocol with 3% hypertonic saline (Sagel et al., 2001). Quantitative bacterial culture was performed on sputum (expectorated and induced) samples following CFF guidelines (Saiman et al., 2014). Residual sputum samples were frozen at − 70 °C for small subunit rRNA (SSU-rRNA) gene sequencing.

2.3.2. SSU-rRNA sequencing—Bacterial profiles were determined by broad-range amplification and sequence analysis of the SSU-rRNA gene following previously described methods. Amplicons were generated using primers targeting approximately 300 base pairs of the V1/V2 variable region of the SSU-rRNA gene. An Illumina MiSeq v2 reagent kit was used to prepare libraries for 2 × 250 bp paired-end sequencing on the Illumina MiSeq platform.

2.3.3. Analysis of illumina paired-end reads—Quality control procedures were performed on paired-end sequences as described in the online supplement. Assembled sequences were aligned and classified with SINA (1.2.11) using the SILVA 115 database as reference configured to yield the SILVA taxonomy (Quast et al., 2012; Pruesse et al., 2012).

2.3.4. Data preprocessing—Before statistical analysis, microbiome count data were filtered to include only prominent taxa. Taxa were filtered out if they did not exceed 0.1% relative abundance (RA, calculated as count/total) in at least 10% of samples. A total of 446 taxa did not meet the filtering criteria and were not included in the analytic dataset which included 60 taxa. Count data were transformed using the centered log-ratio (CLR) transformation. A pseudocount ($RA_{min}/2$) was applied to exact zero RA entries before CLR transformation was performed. Microbiome features were standardized prior to downstream statistical analysis by subtracting the mean and scaling to unit variance.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2.4. Proteomics data

2.4.1. Blood sample collection—An aliquot of 3.5 mL was collected in an EDTA-collection tube and processed following a standard operating procedure. Blood samples were frozen and batch shipped on dry ice to SomaLogic in Boulder, Colorado for proteomics analysis.

2.4.2. SomaLogic proteomics assay—Proteomics data were measured using the SomaScan multiplex proteomics assay (version 3.2), an aptamer-based quantitative proteomic biomarker discovery platform (Gold et al., 2010). Approximately 4000 aptamers, or single-stranded deoxyribonucleic acid (ssDNA) molecules that bind specific protein targets, were measured in this assay. Aptamers targeted approximately 3600 unique proteins.

2.4.3. Data preprocessing—SomaScan measurements, reported in relative fluorescent units (RFUs), were normalized using internal hybridization controls added to the assay prior to hybridization. Normalization was performed to adjust for inter-sample, inter-plate, and inter-run variation. Normalized RFU values were $\log_2$ transformed to handle skewness and improve normality. Proteomics features were standardized prior to downstream statistical analysis by subtracting the mean and scaling to unit variance.

2.5. Statistical analysis

A complete description of the statistical approach is included in the online supplement. Statistical methods, including multi-omic integration, subnetwork identification, and subnetwork sensitivity analysis, were performed using R version 4.1.1.

2.5.1. Multi-omic integration with SmCCNet—Airway microbiome, plasma proteomics and PEx recovery were integrated using SmCCNet (version 0.99.0). SmCCNet is an extension of canonical correlation analysis (CCA) that incorporates a quantitative phenotype to construct phenotype-related multi-omic subnetworks. Given our omic data types and phenotype of interest, we use SmCCNet to construct microbe–protein subnetworks at PEx onset that are correlated with PEx recovery (%ΔPExS). Penalty parameters to induce sparsity were selected using five-fold cross-validation and a randomized grid search approach.

2.5.2. Subnetwork–phenotype correlations—Subnetworks were summarized using principal component analysis (PCA), whereby each subnetwork's first principal component (PC1) was used for univariate subnetwork representation. Subnetwork-specific taxa and proteins were merged into a common matrix for determining the PC1 of each subnetwork. Subnetwork–phenotype correlation was calculated as the Pearson correlation between a the PC1 of a subnetwork and %ΔPExS. Absolute Pearson correlation values are reported since PCA obscures the interpretability of negative and positive relationships. The percentage of variance explained by each subnetwork PC1 was additionally calculated and reported.

2.5.3. Subnetwork selection—SmCCNet results in subnetworks with varying levels of relevance to the phenotype (i.e., not all subnetworks are strongly correlated with the phenotype). We set the following subnetwork selection criteria to focus our results and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

discussion on subnetworks that are related to %ΔPExS. First, we required subnetworks to be strongly correlated with the phenotype (i.e., $r_{\text{PC1-\%ΔPExS}} \geq 0.35$). Second, we only considered subnetworks with PC1s that explained greater than 50% of the variability observed within the subnetwork. This was enforced to ensure that PC1s were reasonable subnetwork representations. Third, we required at least two nodes of each omic type to be present in selected subnetworks to limit omic type imbalance.

2.5.4. Gene ontology enrichment analysis—GO enrichment analyses were performed on selected subnetworks using subnetwork-specific protein sets. Metascape (v3.5.20230501), a method that is advantageous for identifying non-redundant GO terms, was used for enrichment analysis (Zhou et al., 2019). The aim was to identify over-represented GO terms (specifically biological processes) associated with subnetworks of interest. The full set of proteins targeted by the SomaScan proteomics assay was used as the background protein list (3554 unique proteins after filtering out duplicates). Entrez Gene symbols were used as input for both the sample gene list and background gene list. 56 aptamers in the SomaScan assay were found to target multiple proteins, these repeats were removed prior to performing GO enrichment analyses. Minimum enrichment factor and *p*-value cutoffs of 3 and 0.001 were used to guard against false discoveries (i.e., type I errors). We further required a 3 protein overlap between our subnetwork protein lists and GO term protein lists to consider a given GO term over-represented.

2.5.5. Subnetwork visualization—Network visualizations were created using CytoScape (version 3.9.1) (Shannon et al., 2003). We focus visualizations on one over-represented GO term per selected network, rather than full subnetworks, to preserve the interpretability of visualizations and focus discussion. Subnetwork visualizations include overlapping proteins (i.e., proteins found in the subnetwork and GO term) and subnetwork-specific taxa. Subnetwork visualizations provide information about individual node–phenotype associations, node–node relationships and connectivity, and canonical weight contributions.

2.5.6. Subnetwork sensitivity analysis—A sensitivity analysis was performed after removing the repeat PEx events and only including the first PEx event for each study participant. Node overlaps (i.e., microbe and protein overlaps) between reported subnetworks and sensitivity analysis subnetworks were subsequently assessed using visualizations and Fisher's exact test.

3. Results

3.1. Study cohort

Cohort characteristics and PEx event clinical measures are described in Table 1 (Table S1). The median PEx score at study enrollment was 12 and ranged between 8 and 16. %ΔPExS ranged from – 100–0% (Fig. S1), indicating that PEx score either improved (i.e., decreased) or remained constant from hospital admission to discharge for all exacerbations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3.2. SmCCNet configuration and subnetwork selection

A total of 10 subnetworks were identified using SmCCNet, 4 of which had a strong correlation between the first principal component (PC1) and %ΔPExS ($r_{PC1\text{-}\%\Delta PExS} \geq 0.35$) (Fig. 1). Subnetworks 2, 4, 5, and 6 met criteria for subnetwork selection and are described in depth in subsequent sections.

3.3. Identified subnetworks

3.3.1. Subnetwork 2—Subnetwork 2 is a large subnetwork comprised of the most influential nodes and edges in the data set. Features in this network are associated with large canonical weights relative to other networks, suggesting their importance in maximizing the canonical correlation between the taxa, proteins and %ΔPExS (Fig. 1B). The original subnetwork constructed by SmCCNet contained 364 nodes, including 334 proteins and 30 taxa. The pruned subnetwork used for GO enrichment analysis contains 325 nodes, including 298 proteins and 27 taxa (Table 2). Correlations between individual nodes and %ΔPExS range from – 0.57–0.62 (Fig. 1A). A correlation of 0.43 is observed between PC1 of the full subnetwork and %ΔPExS (p-value = 0.01). The central node of the network is *Atopobium* (Fig. 2), which is strongly associated with %ΔPExS (r = 0.44; p-value = 0.01, Table S2). The largest weighted edge connects *Atopobium* and *Actinomyces* and corresponds to a positive association. The top non-redundant GO biological pathways that are enriched in Subnetwork 2 include GO:0048738 (cardiac muscle tissue development), GO:0042100 (B cell proliferation), and GO:0006703 (estrogen biosynthetic process) (Fig. 3).

3.3.2. Subnetwork 4—Subnetwork 4 contains 209 nodes, including 204 proteins and 5 taxa (Table 2). The subnetwork includes *Stenotrophomonas*, a taxon that is traditionally tracked in CF airways but does not appear to be associated with %ΔPExS in this cohort (r = −0.02; p-value = 0.92). Soluble low-density lipoprotein receptor-related protein 1 (sLRP1), the omic feature most strongly correlated with %ΔPExS, is included in this subnetwork (r = 0.63; p-value < 0.01). Subnetwork 4 is the strongest of the selected subnetworks when considering subnetwork–phenotype correlation (r = 0.54; p-value < 0.01) (Fig. 1A). GO biological processes enriched in Subnetwork 4 include GO:0010498 (proteasomal protein catabolic process), GO:0043949 (regulation of cAMP-mediated signaling), and GO:2000328 (regulation of T-helper 17 cell-lineage commitment) (Fig. 3). GO:2000328 proteins are visualized with Subnetwork 4 proteins in Fig. 4. Interestingly, IL-23 and IL-23 receptor, each found in this subnetwork and GO:2000328, are inversely expressed. *Streptococcus anginosus* shows no connectivity to other nodes in Fig. 4, indicating that the taxon is weakly associated with GO:2000328 proteins (r < 0.2).

3.3.3. Subnetwork 5—Subnetwork 5 is a 208-node subnetwork comprised of 201 proteins and 7 taxa (Table 2). Nodes include taxa traditionally found in CF airways, such as *Pseudomonas aeruginosa (P. aeruginosa*) and *Prevotella* (both at the species and genus level). *P. aeruginosa* abundance at PEx onset is moderately positively associated with %ΔPExS (r = 0.23; p-value = 0.19, Table S3). Individual node–%ΔPExS correlations range from – 0.46–0.47 (Fig. 1A). A correlation of 0.35 is observed between PC1 of the subnetwork and %ΔPExS (p-value = 0.05). Edges are relatively weak in Subnetwork 5 when compared to the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

edges in the other selected subnetworks, both in terms of the edge strength median and range (Fig. 1B). Weaker edges suggest that Subnetwork 5 nodes are less interrelated than nodes in other subnetworks. GO biological processes found within the Subnetwork protein set include GO:0006412 (translation), GO:0006417 (regulation of transcription), GO:0051092 (positive regulation of NF-kappaB transcription factor activity), and GO:0043039 (tRNA aminoacylation) (Fig. 3). Proteins overlapping both GO:0051092 and Subnetwork 5 are visualized with Subnetwork 5 taxa in Fig. 5.

3.3.4. Subnetwork 6—Subnetwork 6 contains 98 nodes, including 96 proteins and 2 taxa (Table 2). Taxa include *Streptococcus speciesIG2* and *Prevotella histicola*. Several nodes within Subnetwork 6 are well-correlated with %ΔPExS, with node–%ΔPExS correlation values ranging from – 0.50–0.58. The PC1 representation of Subnetwork 6 shows an absolute correlation of 0.42 with %ΔPExS (p-value = 0.01). GO enrichment analysis results suggest that GO:0072126 (positive regulation of glomerular mesangial cell proliferation), GO:1901881 (positive regulation of protein depolymerization), and GO:0072524 (pyridine-containing compound metabolic process) are over-represented in the Subnetwork 6 protein set (Fig. 3). Proteins overlapping Subnetwork 6 and GO:0072126 are visualized with *Streptococcus speciesIG2* and *Prevotella histicola* in Fig. 6.

3.4. Subnetwork sensitivity analysis

Subnetwork sensitivity analysis results are visualized in Fig. 7. Of the originally reported subnetworks, Subnetwork 2 is shown to be the most robust to PEx repeat removal. 81% and 84% of the reported Subnetwork 2 taxa and proteins are preserved in the corresponding sensitivity analysis (SA) subnetwork. Interestingly, taxa and proteins from Subnetworks 4, 5 and 6 agglomerate into SA Subnetwork 5. These results suggest that the originally reported Subnetworks 4, 5, and 6 are more sensitive to PEx repeat removal. Over-represented and non-redundant GO terms determined from the protein sets of SA Subnetworks 2 and 5 are shown in Fig. 8. Similar to the reported GO enrichment analysis results for Subnetwork 2, GO:0042100 (B cell proliferation) and GO:0048738 (cardiac muscle tissue development) proteins are enriched in SA Subnetwork 2. GO terms associated with the SA Subnetwork 5 protein set are inconsistent with GO terms reported for Subnetworks 4, 5 and 6. Some of the SA Subnetwork 5 GO terms are interesting in the context of CF, however, such as GO:0031396 (protein ubiquitination; $-\log_{10}p = 3.8$) and GO:19033844 (regulation of cellular response to transforming growth factor beta stimulus; $-\log_{10}p = 2.9$).

4. Discussion

Previous analyses investigating the contribution of the microbiome in CF PExs have shown limited evidence that typical CF pathogens change prior to exacerbations and that these organisms are resilient despite treatment with antibiotics (Bevivino et al., 2019; Price et al., 2013; Raghuvanshi et al., 2020; Widder et al., 2022; Zhao et al., 2012). Furthermore, the role of non-traditional CF pathogens, particularly anaerobic organisms, is less clear (Khanolkar et al., 2020). A paradigm shift has recently been suggested in order to better understand the complex polymicrobial nature of CF airway infections (Bevivino et al., 2019; Khanolkar et al., 2020; O'Toole et al., 2021). Essentially, it has been suggested that there is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

redundancy in the bacterial community, that a set of taxa could perform similar functions, in which case, classifying taxa based on functions or mechanism may yield additional insight over focusing on individual taxa. Similarly, using a hypothesis generating method to elucidate microbe-host interactions may also be used for identifying groups of taxa. In this work, we have used such an approach to identify relationships between taxa and markers of host-response that correspond to sub-networks. A better understanding of the complex relationships is needed before treatments targeted at a taxa group can be explored (Khanolkar et al., 2020).

Several microbe–protein networks from PEx onset associated with PEx recovery were identified in this study. To our knowledge, this is the first study to integrate airway microbiome data and host response blood proteome data within the context of CF and PEx recovery. Further, this is the first known report to use SmCCNet for integration of these omic types. We identified four distinct microbe–protein subnetworks from PEx onset that are associated with %ΔPExS. Certain subnetworks, including Subnetwork 5, align with preexisting knowledge in CF and provide evidence for the usefulness of SmCCNet in this multi-omic application. Novel, hypothesis generating subnetworks additionally emerged from this analysis. These include subnetworks like Subnetwork 2, which include taxa that have not been historically tracked in CF.

Subnetwork 5, a subnetwork containing *P. aeruginosa* and proteins involved in inflammatory response processes such as regulation of nuclear factor kappa beta (NF-κB) transcription factor activity, serves as a traditional CF subnetwork. *P. aeruginosa* is largely considered the hallmark pathogen of CF and its involvement in the excessive inflammatory and immune responses seen in people with CF has been consistently reported. The NF-κB signaling pathway plays significant roles in adaptive and innate immune response and has been shown to induce the expression of pro-inflammatory cytokines and chemokines. These pro-inflammatory markers include cytokines and chemokines previously observed during CF PExs (Liu et al., 2017). Several studies have linked NF-κB transcription factor activity to CF. One study found that NF-κB activation can result in respiratory epithelial cells from the presence of *P. aeruginosa* pilin (DiMango et al., 1998). A separate study aimed to inhibit IL-8 transcription (and subsequently reduce the inflammatory host response) in CF epithelial cells infected with *P. aeruginosa* by interfering with the NF-κB transcription signaling pathway (Finotti et al., 2012).

Subnetwork 2 was found to contain the strongest node connectivity and served as the most robust subnetwork when removing repeat PEx events in the sensitivity analysis. Interestingly, this subnetwork revolves around *Atopobium* and *Actinomyces*; two taxa that have not historically been tracked in CF airways. These nodes are associated with larger canonical weights than all other omic features. They are positively associated with %ΔPExS, suggesting that increased abundance of these taxa at PEx onset is associated with less favorable PEx recovery. GO:0042100, the GO term representative of B cell activation, was found to be enriched in the Subnetwork 2 protein set. B cells play a critical role in humoral, or antibody-mediated, immunity in humans (Althwaiqeb and Bordoni, 2023). While the involvement of B cells in CF is ill-defined, some studies suggest heightened B cell activation in CF airways (Giacalone et al., 2020). One report found that B cell-derived lymphoid

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

aggregates were significantly larger in CF airways as compared to non-CF airways (Hubeau et al., 2002). The authors suggest that increased presence of B cells in the airway contributes to the exaggerated inflammatory process observed in CF. Interleukin 7 (IL-7), a protein that overlaps Subnetwork 2 and GO:0042100 and is known for its growth-promoting effects on B cells progenitors, has specifically been linked to CF. One study found that increased levels of serum IL-7 are associated with worsened lung function in CF, as measured by FEV-1% predicted (Seyfarth et al., 2019). This result is consistent with our findings. We observed a positive association between both IL-7 and the IL-7 receptor subunit alpha and %ΔPExS, suggesting worsened PEx recovery with increased levels of these proteins in the blood at PEx onset ($r = 0.31$ and $r = 0.34$, respectively). An additional Subnetwork 2 protein involved in CF includes interleukin 17 (IL-17), a pro-inflammatory cytokine that promotes the production of airway neutrophilia (Hsu et al., 2016).

In light of these findings, it is important to acknowledge the limitations of this study that should be considered when interpreting the results. First, our methodology assumes independent observations while there are repeated exacerbation events for four study participants. It is possible that results are weighted more heavily toward these individuals, especially if airway communities or host response are similar for these subjects between separate PEx events. We address this limitation by performing a sensitivity analysis that aims to assess the impact of repeat PEx removal. Results should be interpreted holistically, with sensitivity analysis results providing a gauge of confidence with respect to each reported subnetwork. Second, the generalizability of this study is limited since the study cohort is small and since subjects originate from a single site. Third, the host-response was measured in circulating blood which may not be fully representative of what is occurring in the lung, however, given that subjects are hospitalized with acute lung infections, it is likely that the inflammatory responses are originating from the site of infection. Fourth, the use of speciated taxa in this analysis might lead to misleading results for some of the non-speciated taxa. Fifth, and finally, non-linear relationships likely exist between some omic features or between some omic features and the phenotype, violating the linear assumption of CCA.

5. Conclusions

We identified subnetworks of microbe-protein combinations that were associated with recovery during a PEx in subjects with CF. This highly acute infection and inflammatory event provides a useful context in which to study the associations between microbial communities and host-response. The identified networks included both known and established relationships in CF as well as potentially new and novel associations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding Sources

This work was supported by the National Institutes of Health (NIH/NCATS Colorado CTSA Grant Number UM1 TR004399, K23HL114883,1R01HL124103, R01 DA035804 and R01 DA032555), Children’s Hospital Colorado Research Institute and the Cystic Fibrosis Foundation (ZEMANI12A0, ZEMANI16A0, PROMISE-BIO18K0).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Data Availability

Related code and de-identified data used for paper can be found at https://github.com/BrentonGraham/multiomic-thesis-project.

References

Althwaiqeb SA, Bordoni B, 2023. Histology, B Cell Lymphocyte, 2023.

Amadori A,Antonelli A, Balteri I, Schreiber A, Bugiani M, De Rose V, 2009. Recurrent exacerbations affect FEV(1) decline in adult patients with cystic fibrosis. Respir. Med 103, 407–413. 10.1016/j.rmed.2008.09.024. [PubMed: 19027279]

Bevivino A, Bacci G, Drevinek P, Nelson MT, Hoffman L, Mengoni A, 2019. Deciphering the ecology of cystic fibrosis bacterial communities: towards systems-level integration. Trends Mol. Med 25, 1110–1122. 10.1016/j.molmed.2019.07.008. [PubMed: 31439509]

Chiron R, Grumbsach YY, Quynh NVT, Verriere V, Urbach V, 2008. Lipoxin A4 and interleukin-8 levels in cystic fibrosis sputum after antibiotherapy. J. Cyst. Fibros 7, 463–468. 10.1016/j.jcf.2008.04.002. [PubMed: 18541460]

Colombo C, Costantini D, Rocchi A, Cariani L, Garlaschi ML, Tirelli S, Calori G, Copreni E, Conese M, 2005. Cytokine levels in sputum of cystic fibrosis patients before and after antibiotic therapy. Pedia Pulmonol. 40, 15–21. 10.1002/ppul.20237.

DiMango E, Ratner AJ, Bryan R, Tabibi S, Prince A, 1998. Activation of NF-kappaB by adherent Pseudomonas aeruginosa in normal and cystic fibrosis respiratory epithelial cells. J. Clin. Investig 101, 2598–2605. 10.1172/JCI2865. [PubMed: 9616231]

Finotti A, Borgatti M, Bezzerri V, Nicolis E, Lampronti I, Dechecchi M, Mancini I, Cabrini G, Saviano M, Avitabile C, Romanelli A, Gambari R, 2012. Effects of decoy molecules targeting NF-kappaB transcription factors in Cystic fibrosis IB3–1 cells. Artif. DNA PNA XNA 3, 97–104. 10.4161/adna.21061. [PubMed: 22772035]

Giacalone VD, Dobosh BS, Gaggar A, Tirouvanziam R, Margaroli C, 2020. Immunomodulation in Cystic Fibrosis: Why and How? Int. J. Mol. Sci 21, 3331. 10.3390/ijms21093331. [PubMed: 32397175]

Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody E, Carter J, Cunningham V, Dalby A, Eaton B, Fitzwater T, Flather D, Forbes A, Foreman T, Fowler C, Gawande B, Goss M, Gunn M, Gupta S, Halladay D, Heil J, Heilig J,Hicke B, Husar G, Janjic N, Jarvis T, Jennings S, Katilius E, Keeney T, Kim N, Kaske T, Koch T, Kraemer S, Kroiss L, Le N, Levine D, Lindsey W, Lollo B, Mayfield W, Mehan M, Mehler R, Nelson M, Nelson S,Nieuwlandt D, Nikrad M, Ochsner U, Ostroff R, Otis M, Parker T, Pietrasiewicz S, Resnicow D, Rohloff J, Sanders G, Sattin S, Schneider D, Singer B, Stanton M, Sterkel A, Stewart A, Stratford S, Vaught J, Vrkljan M, Walker J, Watrobka M, Waugh S, Weiss A, Wilcox S, Wolfson A, Wolk S, Zhang C, Zichi D, 2010. Aptamer-based multiplexed proteomic technology for biomarker discovery. Nat. Preced 10.1038/npre.2010.4538.l.

Hsu D, Taylor P, Fletcher D, van Heeckeren R, Eastman J, van Heeckeren A, Davis P, Chmiel JF, Pearlman E, Bonfield TL, 2016. Interleukin-17 pathophysiology and therapeutic intervention in cystic fibrosis lung infection and inflammation. Infect. Immun 84, 2410–2421. 10.1128/IAI.00284-16. [PubMed: 27271746]

Hubeau C, Lorenzato M, Couetil JP, Hubert D, Dusser D, Puchelle E, Gaillard D, 2002. Quantitative analysis of inflammatory cells infiltrating the cystic fibrosis airway mucosa. Clin. Exp. Immunol 124, 69–76. 10.1046/j.1365-2249.2001.01456.x.

Khanolkar RA, Clark ST, Wang PW, Hwang DM, Yau YCW, Waters VJ, Guttman DS, 2020. Ecological succession of polymicrobial communities in the cystic fibrosis airways. MSystems 5. 10.1128/mSystems.00809-20.

Liou TG, Adler FR, FitzSimmons SC, Cahill BC, Hibbs JR, Marshall BC, 2001. Predictive 5-Year Survivorship Model of Cystic Fibrosis. Am. J. Epidemiol 153, 345–352. 10.1093/aje/153.4.345. [PubMed: 11207152]

Liu T, Zhang L, Joo D, Sun S-C, 2017. NF-κB signaling in inflammation. Signal Transduct. Target Ther 2, 17023 10.1038/sigtrans.2017.23. [PubMed: 29158945]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

O'Toole GA, Crabbé A, Kümmerli R, LiPuma JJ, Bomberger JM, Davies JC, Limoli D, Phelan VV, Bliska JB, DePas WH, Dietrich LE, Hampton TH, Hunter R, Khursigara CM, Price-Whelan A, Ashare A, Cramer RA,Goldberg JB, Harrison F, Hogan DA, Henson MA, Madden DR, Mayers JR, Nadell C, Newman D, Prince A, Rivett DW, Schwartzman JD, Schultz D, Sheppard DC, Smyth AR, Spero MA, Stanton BA, Turner PE, van der Gast C, Whelan FJ, Whitaker R, Whiteson K, 2021. Model systems to study the chronic, polymicrobial infections in cystic fibrosis: current approaches and exploring future directions. MBio 12. 10.1128/mBio.01763-21.

Ordoñez CL, Henig NR, Mayer-Hamblett N, Accurso FJ, Burns JL, Chmiel JF, Daines CL, Gibson RL, McNamara S, Retsch-Bogart GZ, Zeitlin PL,Aitken ML, 2003. Inflammatory and microbiologic markers in induced sputum after intravenous antibiotics in cystic fibrosis. Am. J. Respir. Crit. Care Med 168, 1471–1475. 10.1164/rccm.200306-7310C. [PubMed: 12969869]

Price KE, Hampton TH, Gifford AH, Dolben EL, Hogan DA, Morrison HG, Sogin ML, O'Toole GA, 2013. Unique microbial communities persist in individual cystic fibrosis patients throughout a clinical exacerbation. Microbiome 1, 27. 10.1186/2049-2618-l-27. [PubMed: 24451123]

Pruesse E, Peplies J, Glöckner FO, 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28, 1823–1829. 10.1093/bioinformatics/bts252. [PubMed: 22556368]

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J,Glockner FO, 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41, D590–D596. 10.1093/nar/gksl219. [PubMed: 23193283]

Raghuvanshi R, Vasco K, Vázquez-Baeza Y, Jiang L, Morton JT, Li D,Gonzalez A, DeRight Goldasich L, Humphrey G, Ackermann G, Swafford AD, Conrad D, Knight R, Dorrestein PC, Quinn RA, 2020. High-resolution longitudinal dynamics of the cystic fibrosis sputum microbiome and metabolome through antibiotic therapy. MSystems 5. 10.1128/mSystems.00292-20.

Rosenfeld M, Emerson J, Williams-Warren J, Pepe M, Smith A, Montgomery AB, Ramsey B, 2001. Defining a pulmonary exacerbation in cystic fibrosis. J. Pedia 139, 359–365. 10.1067/mpd.2001.117288.

Sagel SD, Kapsner R, Osberg I, Sontag MK, Accurso FJ, 2001. Airway inflammation in children with cystic fibrosis and healthy children assessed by sputum induction. Am. J. Respir. Crit. Care Med 164, 1425–1431. 10.1164/ajrccm.164.8.2104075. [PubMed: 11704590]

Sagel SD, Thompson V, Chmiel JF, Montgomery GS, Nasr SZ, Perkett E,Saavedra MT, Slovis B, Anthony MM, Emmett P, Heltshe SL, 2015. Effect of treatment of cystic fibrosis pulmonary exacerbations on systemic inflammation. Ann. Am. Thorac. Soc 12, 708–717. 10.1513/AnnalsATS.201410-4930C. [PubMed: 25714657]

Saiman L, Siegel JD, LiPuma JJ, Brown RF, Bryson EA, Chambers MJ,Downer VS, Fliege J, Hazle LA, Jain M, Marshall BC, O'Malley C, Pattee SR, Potter-Bynoe G, Reid S, Robinson KA, Sabadosa KA, Schmidt HJ,Tullis E, Webber J, Weber DJ, 2014. Infection prevention and control guideline for cystic fibrosis: 2013 Update. Infect. Control Hosp. Epidemiol 35, s1–s67. 10.1086/676882.

Sanders DB, Bittner RCL, Rosenfeld M, Hoffman LR, Redding GJ, Goss CH, 2010a. Failure to recover to baseline pulmonary function after cystic fibrosis pulmonary exacerbation. Am. J. Respir. Crit. Care Med 182, 627–632. 10.1164/rccm.200909-1421OC. [PubMed: 20463179]

Sanders DB, Hoffman LR, Emerson J, Gibson RL, Rosenfeld M, Redding GJ, Goss CH, 2010b. Return of FEV_1 after pulmonary exacerbation in children with cystic fibrosis. Pedia Pulmonol. 45, 127–134. 10.1002/ppul.21117.

Sanders DB, Bittner RCL, Rosenfeld M, Redding GJ, Goss CH, 2011. Pulmonary exacerbations are associated with subsequent FEV_1 decline in both adults and children with cystic fibrosis. Pedia Pulmonol. 46, 393–400. 10.1002/ppul.21374.

Seyfarth J, Sivagurunathan S, Ricken S, Weinreich G, Olbrich L, Taube C, Mayatepek E, Schramm D, Jacobsen M, 2019. Higher Interleukin-7 serum concentrations in patients with cystic fibrosis correlate with impaired lung function. J. Cyst. Fibros 18, 71–77. 10.1016/j.jcf.2018.09.008. [PubMed: 30389600]

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T, 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13, 2498–2504. 10.1101/gr.1239303. [PubMed: 14597658]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Shi WJ, Zhuang Y, Russell PH, Hobbs BD, Parker MM, Castaldi PJ, Rudra P, Vestal B, Hersh CP, Saba LM, Kechris K, 2019. Unsupervised discovery of phenotype-specific multi-omics networks. Bioinformatics 35, 4336–4343. 10.1093/bioinformatics/btz226. [PubMed: 30957844]

Widder S, Zhao J, Carmody LA, Zhang Q, Kalikin LM, Schloss PD, LiPuma JJ, 2022. Association of bacterial community types, functional microbial processes and lung disease in cystic fibrosis airways. ISME J. 16, 905–914. 10.1038/S41396-021-01129-z. [PubMed: 34689185]

Zhao J, Schloss PD, Kalikin LM, Carmody LA, Foster BK, Petrosino JF, Cavalcoli JD, VanDevanter DR, Murray S, Li JZ, Young VB, LiPuma JJ, 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. Proc. Natl. Acad. Sci 109, 5809–5814. 10.1073/pnas.1120577109. [PubMed: 22451929]

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK, 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun 10, 1523 10.1038/S41467-019-09234-6. [PubMed: 30944313]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

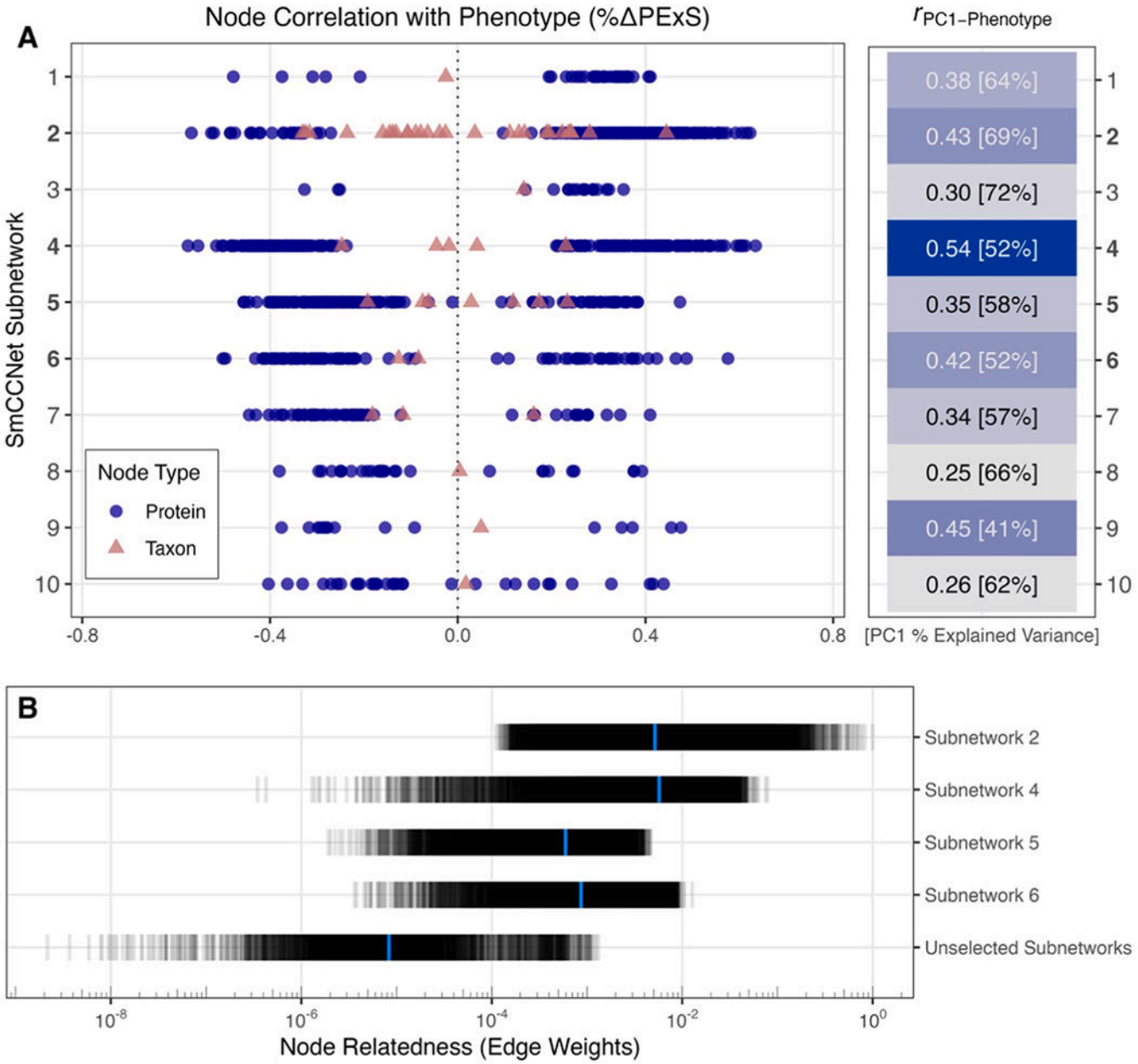

Fig. 1.

SmCCNet subnetwork summarizations. (A) High level summarization of all subnetworks identified by SmCCNet. The scatterplot (left) shows correlations between individual subnetwork nodes and %ΔPExS. Protein nodes are represented as blue circles and taxon nodes are represented as pink triangles. The heatmap (right) reports the absolute correlation observed between PC1 of each subnetwork and %ΔPExS. (B) Shared canonical weight summarization for selected SmCCNet subnetworks. The rug plot shows the distribution of network edge weights (or edge strength) for each subnetwork. Edges for all unselected subnetworks are additionally shown for reference. Vertical blue lines indicate the median edge weight for a given subnetwork. The x-axis is $\log_{10}$transformed to improve resolution of small weights.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

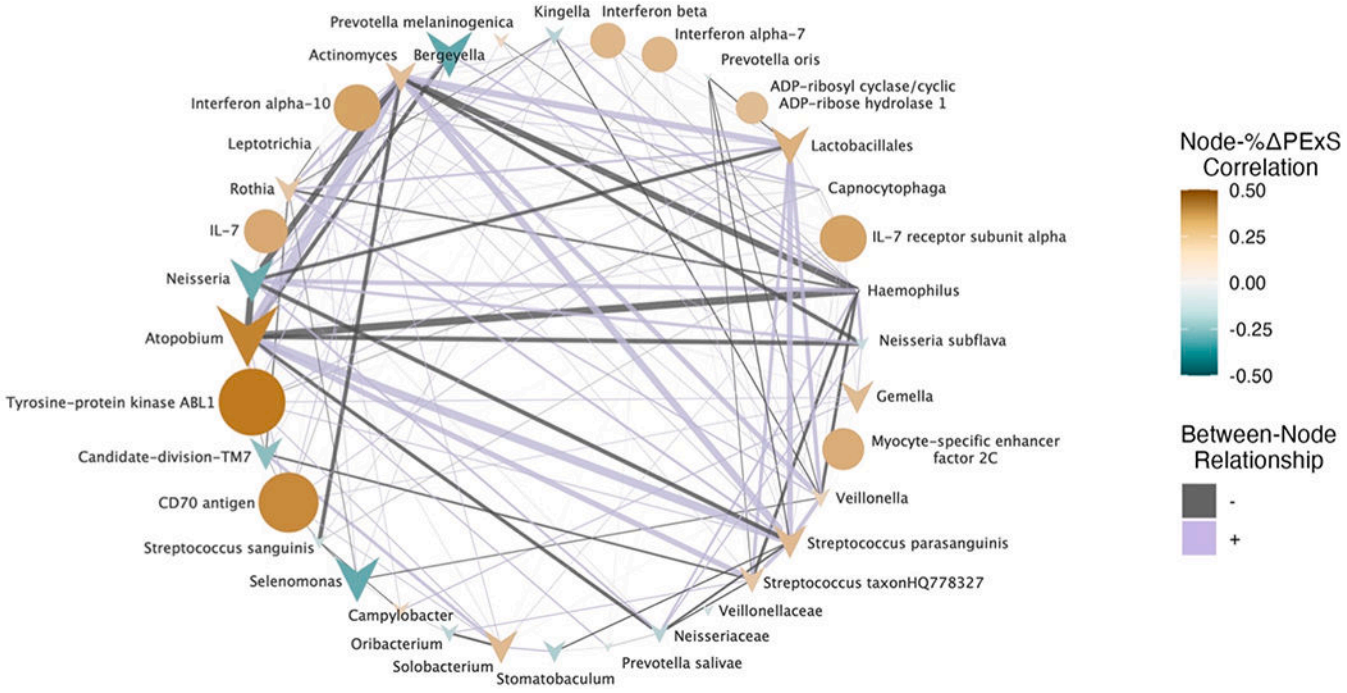

Fig. 2.

Subnetwork 2, GO:0042100 (B cell proliferation) network visualization. Protein nodes are circular and taxon nodes are V-shaped. Node size corresponds to the absolute correlation observed between a given node and %ΔPExS. Edge thickness corresponds to between-node canonical weight-based connectivity. Positively and negatively associated nodes are connected by purple and grey, respectively.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

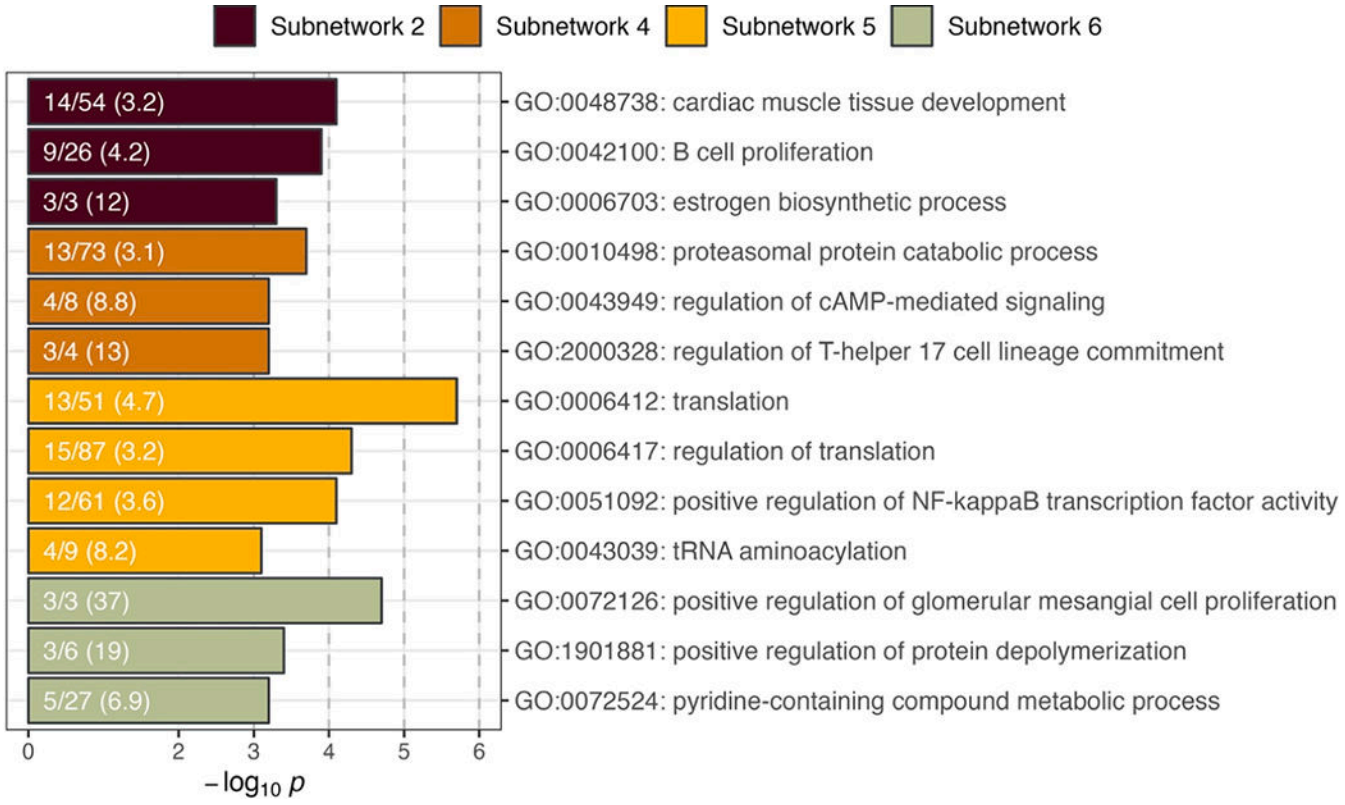

Fig. 3.

GO biological processes over-represented in selected subnetwork protein sets. The bar plot shows the *p*-value associated with each GO biological process, represented as $-\log_{10}p$. Each bar includes information about the number of proteins overlapping a given subnetwork and the GO term (*FreqSample*), the number of proteins annotated to the GO biological process in the background list (*FreqBackground*), and the determined enrichment factor (*Enrichment*). This information is presented in the following format: *FreqSample* / *FreqBackground* (*Enrichment*).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

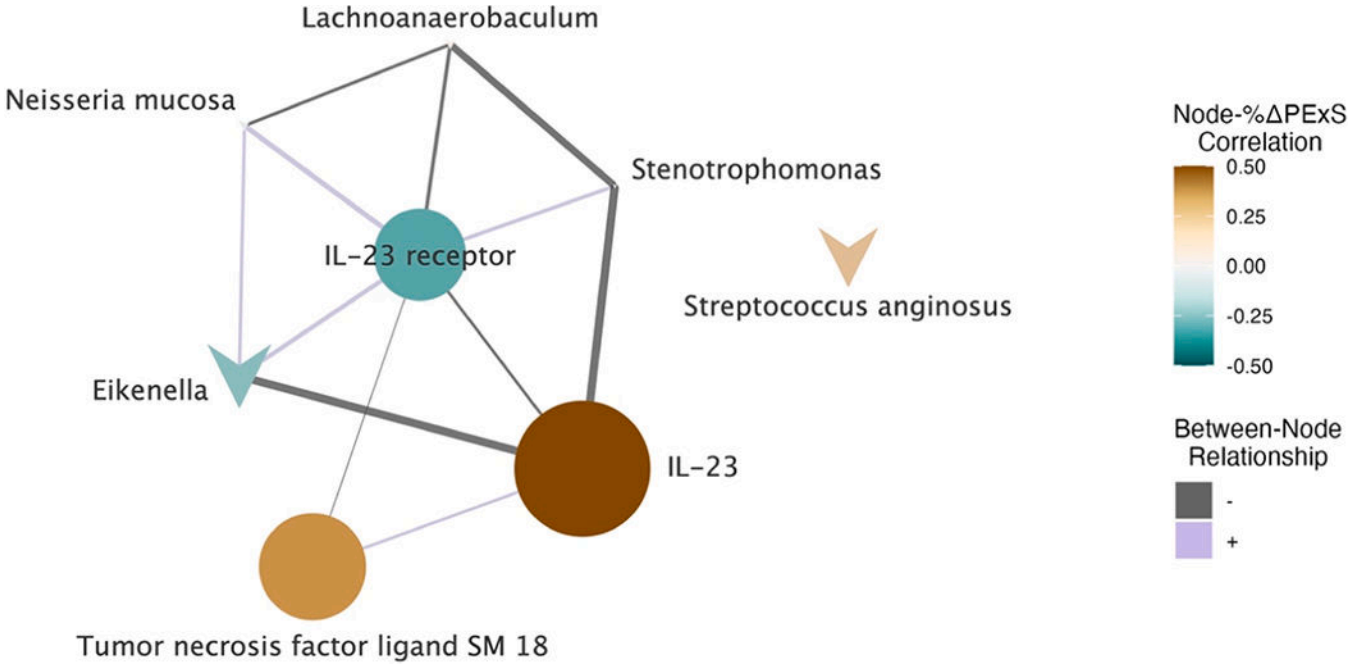

Fig. 4.

Subnetwork 4, GO:2000328 (regulation of T-helper 17 cell-lineage commitment). Protein nodes are circular and taxon nodes are V-shaped. Node size corresponds to the absolute correlation observed between a given node and %ΔPExS. Edge thickness corresponds to between-node canonical weight-based connectivity. Positively and negatively associated nodes are connected by purple and grey, respectively.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

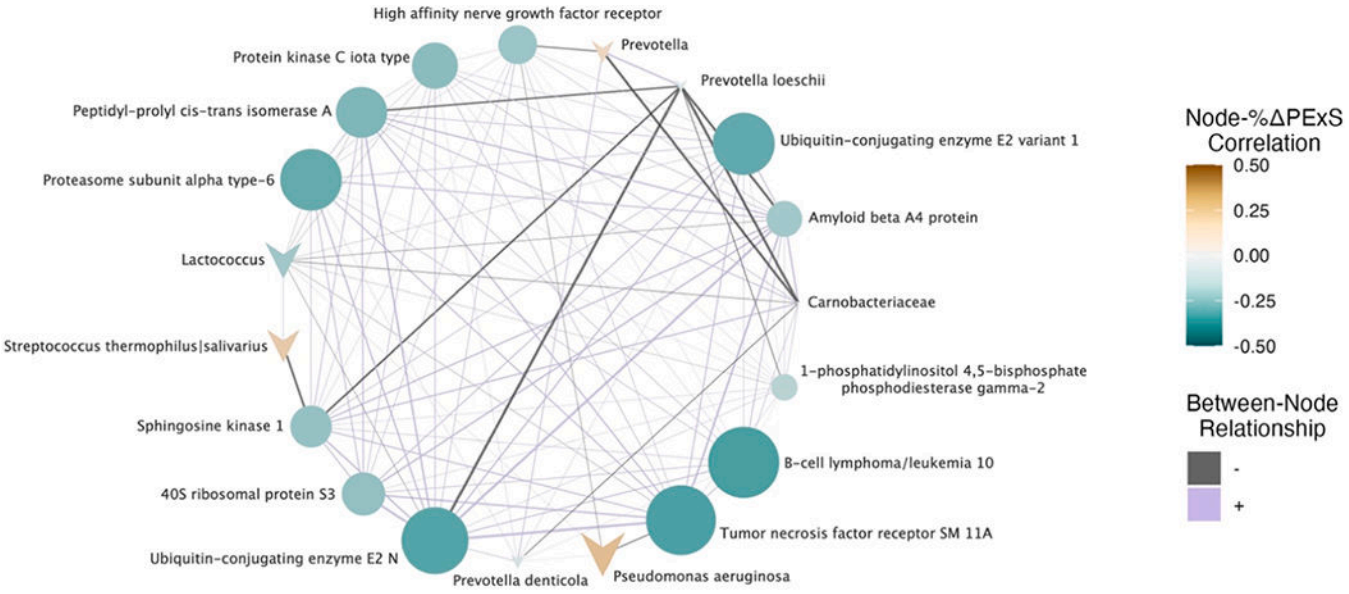

Fig. 5.

Subnetwork 5, GO:0051092 (positive regulation of NF-kappaB transcription factor activity). Protein nodes are circular and taxon nodes are V-shaped. Node size corresponds to the absolute correlation observed between a given node and %ΔPExS. Edge thickness corresponds to between-node canonical weight-based connectivity. Positively and negatively associated nodes are connected by purple and grey, respectively.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

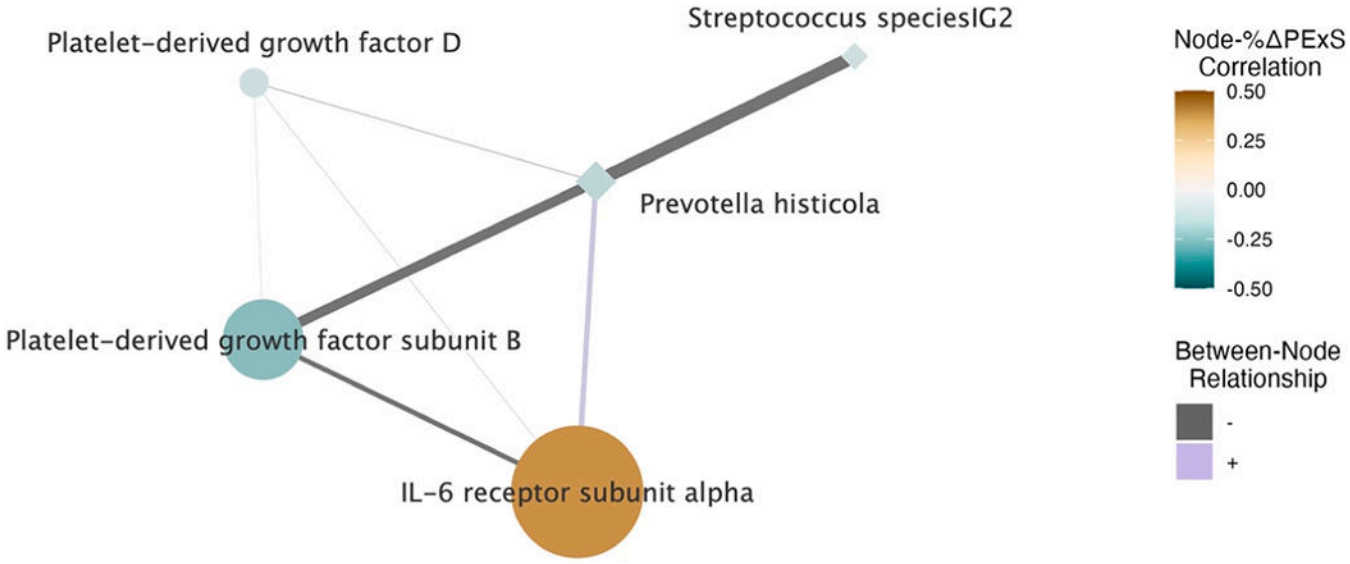

Fig. 6.

Subnetwork 6, GO:0072126 (positive regulation of glomerular mesangial cell proliferation). Protein nodes are circular and taxon nodes are V-shaped. Node size corresponds to the absolute correlation observed between a given node and %ΔPExS. Edge thickness corresponds to between-node canonical weight-based connectivity. Positively and negatively associated nodes are connected by purple and grey, respectively.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

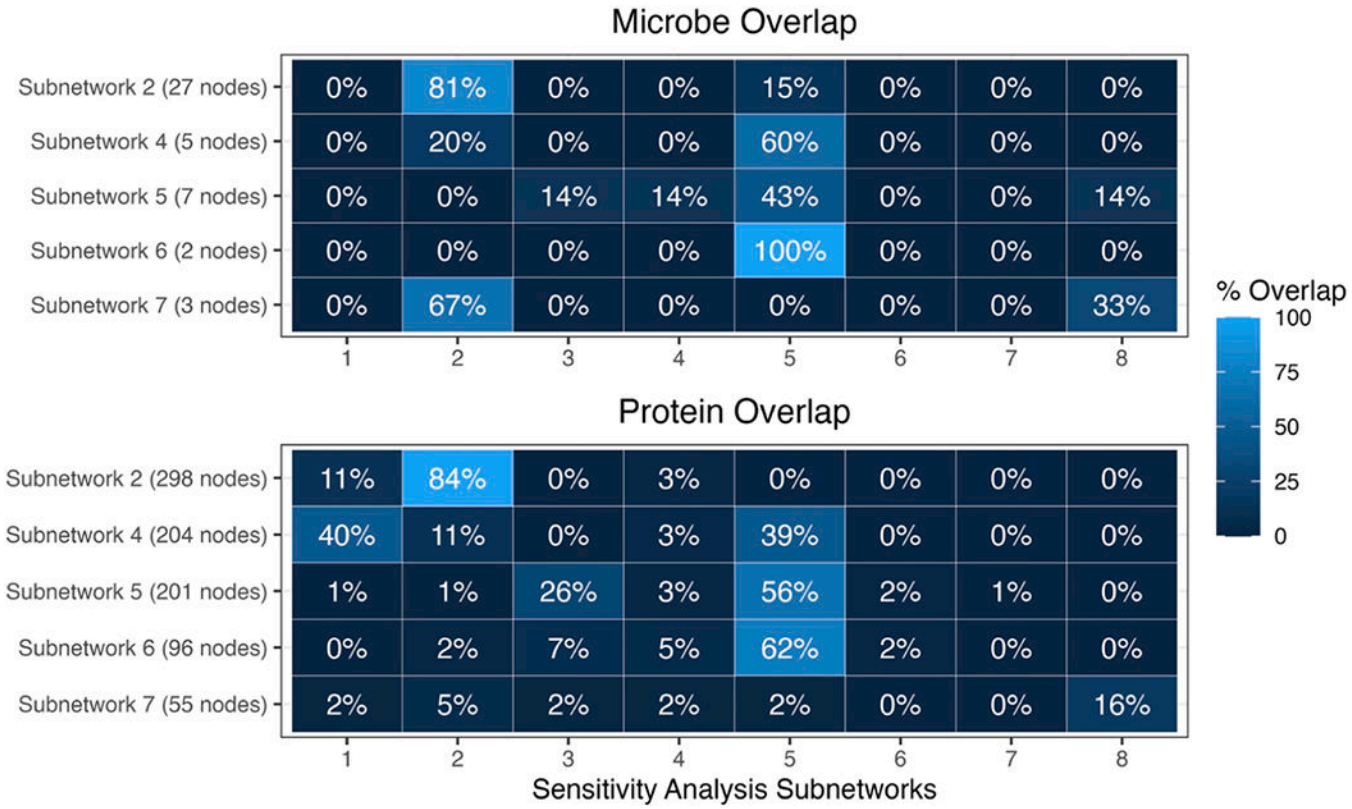

Fig. 7.

Subnetwork sensitivity analysis subnetwork overlap results. Heatmaps show the percentage of originally reported microbes or proteins (y-axis) present in SmCCNet-constructed sensitivity analysis subnetworks (x-axis).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

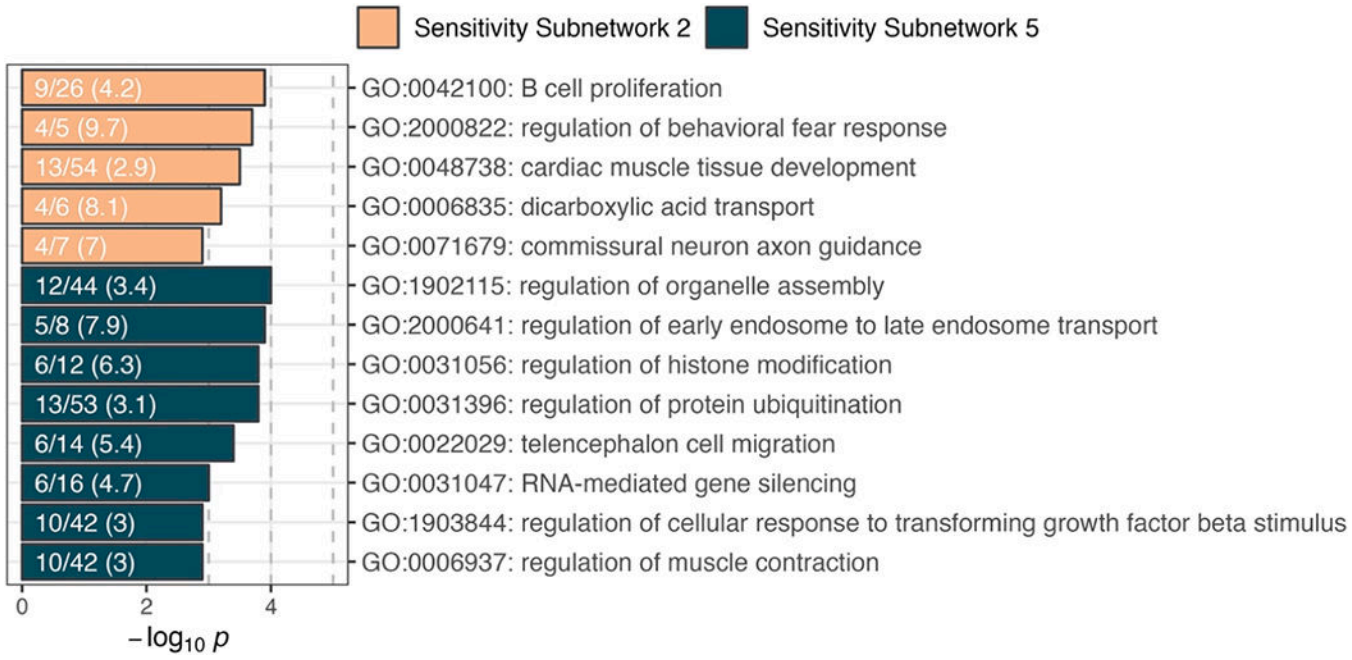

Fig. 8.

Subnetwork sensitivity analysis GO enrichment analysis results. Bar plots indicate GO biological processes over-represented in sensitivity analysis Subnetworks 2 and 5. Each bar includes information about the number of proteins overlapping a given subnetwork and the GO term (*FreqSample*), the number of proteins annotated to the GO biological process in the background list (*FreqBackground*), and the determined enrichment factor (*Enrichment*). This information is presented in the following format: *FreqSample* / *FreqBackground* (*Enrichment*).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Cohort demographics and clinical measures at hospital admission for a PEx. %ΔPExS, the phenotype of interest, is additionally reported. Continuous demographic and clinical information are reported as median [range]. Demographic information is reported at the subject level while clinical information is reported at the PEx event level. Multiple organisms can be detected by culture in the same sample; these groups are not mutually exclusive.

Cohort Subjects (n = 29)	
Age (Years)	15.9 [10.5, 22.1]
Sex (Female)	15 (51.7%)
Genotype (CF Mutations)	
0 F508del	2 (6.9%)
1 F508del	9 (31.0%)
2 F508del	18 (62.1%)
All PEx Events (n = 33)	
FEV-1% Predicted at Admission	81 [30,119]
PEx Score at Admission	12 [8,16]
%ΔPExS	−73 [− 100, 0]
CF Bacteria Culture Detection	
P. aeruginosa	11 (33.3%)
S. aureus	19 (57.6%)
Haemophilus	1 (3.0%)
Stenotrophomonas	5 (15.2%)
Burkholderia	5 (15.2%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

SmCCNet subnetwork summarization metrics for the selected SmCCNet subnetworks (i.e., subnetworks with strong associations to %ΔPExS, or $r_{\text{PC1-\%ΔPExS}} \geq 0.35$). Metrics include subnetwork PC1–phenotype correlation estimates, subnetwork size information, and node–phenotype correlation ranges.

	PC1-Phenotype Correlation		Node Description			Node–Phenotype Correlation Range	
Subnetwork	***r***	***p*-value**	n_{Nodes}	n_{Taxa}	n_{Proteins}	r_{min}	r_{max}
2	0.43	0.01	325	27	298	−0.57	0.62
4	0.54	< 0.01	209	5	204	−0.57	0.63
5	0.35	0.05	208	7	201	−0.46	0.47
6	0.42	0.01	98	2	96	−0.50	0.58