Integrating airway microbiome and plasma proteomics data to identify multi-omic networks associated with cystic fibrosis pulmonary exacerbation recovery

Authors:

**Introduction**

Pulmonary exacerbations (PExs) are the leading cause of morbidity and mortality in cystic fibrosis (CF). Typically characterized by an acute decrease in lung function and quality of life, CF PExs are considered significant life events for people with CF. Individuals who experience frequent and recurring PEx episodes often sustain long-term deterioration in lung function and shortened survival (1-3). Intravenous (IV) antibiotic treatment is typically administered at the onset of a CF PEx to combat bacterial airway infection. Response to treatment is variable, however, as many patients do not return to baseline health (4, 5). There is therefore interest in characterizing the factors that contribute to PEx recovery heterogeneity. Identifying these determinants may improve our understanding of CF PEx disease mechanisms, provide evidence for the effectiveness and inadequacies of current CF therapies, as well as point to new therapeutic targets.

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 (6-9). Identifying associations between specific airway bacteria or bacterial communities and host-response may be critical to understanding the pathogenicity of CF bacteria. Further, exploring microbiota—protein relationships in the context of PEx recovery may provide novel insights into the mechanisms of PEx and the efficacy of PEx treatments.

In this study, we use an extension of canonical correlation analysis (CCA) called sparse multiple canonical correlation network (SmCCNet) to integrate plasma proteomics and airway microbiome data collected at the onset of PEx (i.e., pre-treatment CF PEx samples) (10). Specifically, these omics data are integrated into interpretable microbiota—protein subnetworks that are associated with PEx recovery. We use percent change in PEx score (%ΔPExS) between hospital admission (i.e., PEx onset) and hospital discharge (i.e., late in a PEx, post-treatment) as a clinical measure of PEx recovery. Our aim was to identify potential biomarkers or microbiota—protein subnetworks at PEx onset that are indicative of PEx recovery.

**Materials & Methods**

**Study Design & Population**

This study includes 33 PEx events from a cohort of 29 subjects aged 10 to 22 years old with a confirmed diagnosis of CF. Participants could be reenrolled 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. A PEx was defined by 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. Subjects were treated with IV antibiotics targeting specific CF pathogen(s) detected in the sputum by microbial culture.

The study consists of 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). Study procedures at each visit included a physical, a spirometry test following standard American Thoracic Society guidelines [ref], a standardized PEx score [ref], a validated quality-of-life measure, and specimen collection (blood plasma and sputum).

**Clinical Phenotype Definition**

The phenotype of interest in this study is percent change in PEx score (PExS) between hospital admission (*t1*) and hospital discharge (*t2*) (%ΔPExS). %ΔPExS is defined in Equation 1. A PExS is a standardized measure of PEx severity. The PEx scoring system was originally developed to establish a systematic definition of PEx in the clinic that does not rely on physician discretion (11). PEx scores are measured using patient symptoms (two week change in exercise tolerance, cough, sputum production, chest congestion, school or work attendance, and appetite) and physical examination findings (increased adventitial sounds on auscultation of the chest and change in FEV1 over the preceding month) (11). Increased measures of PExS correspond to increased PEx severity. We interpret %ΔPExS in this study as a clinical measure of PEx recovery. Negative %ΔPExS values represent an improved PExS between hospital admission and discharge, with larger negative values indicating greater improvement.



A black text on a white background

Description automatically generated with low confidence

**Microbiome Data**

*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 [ref]. Quantitative bacterial culture was performed on sputum (expectorated and induced) samples following CFF guidelines [ref]. A respiratory viral panel (Luminex Molecular Diagnostics X-Tag® RVP) was used to detect the presence of virus in sputum samples. Residual sputum samples were frozen at −70°C for 16S rRNA sequencing.

*16S rRNA Sequencing*

Bacterial profiles were determined by broad-range amplification and sequence analysis of 16S rDNA following previously described methods [ref]. Amplicons were generated using primers targeting approximately 300 base pairs of the V1/V2 variable region of the 16S rRNA gene. Illumina paired-end sequencing was performed on the MiSeq platform using a 500 cycle v2 reagent kit.

*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) [ref] using the Silva 111 database [ref] as reference configured to yield the Silva taxonomy. Sorted paired-end sequence data were deposited in the NCBI Short Read Archive under accession number xxx.

*Data Preprocessing*

Microbiome count data were filtered to include only prevalent taxa. Detection and prevalence thresholds were set to 0.1% and 10%, respectively (i.e., taxa were filtered out if they did not exceed 0.1% relative abundance (RA) in at least 10% of samples). Count data were transformed using the centered log-ratio (CLR) transformation. A pseudocount (RAmin/2) was applied to exact zero RA entries before CLR transformation was performed. Microbiome features were standardized prior to downstream statistical analysis by removing the mean and scaling to unit variance.

**Proteomics Data**

*Sample Collection*

3.5 mL aliquots of blood were collected 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.

*SomaLogic Proteomics Assay*

Proteomics data were measured using the SomaScan multiplex proteomics assay, an aptamer-based quantitative proteomic biomarker discovery platform. Approximately 4,000 aptamers, or single-stranded deoxyribonucleic acid (ssDNA) molecules that bind specific protein targets, were measured in this assay. Aptamers targeted approximately 3,600 unique proteins.

*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 removing the mean and scaling to unit variance.

**Statistical Analysis**

*Statistical Software*

All statistical analyses, including multi-omic integration, subnetwork identification, gene ontology (GO) enrichment analysis, and sensitivity analysis, were performed using R version 4.1.1.

*Multi-omic Integration Task*

Airway microbiome (Χ1) and plasma proteomics (Χ2) data from PEx onset were integrated using SmCCNet (version 0.99.0). SmCCNet is an extension of canonical correlation analysis (CCA) that incorporates a quantitative phenotype to construct interpretable multi-omic subnetworks. Given our omic data types and phenotype of interest, we use SmCCNet to construct microbiota—protein subnetworks at PEx onset that are correlated with PEx recovery (%ΔPExS, or Υ).

*Canonical Correlation Analysis*

CCA is an unsupervised learning approach that aims to identify relationships between two multivariable data sets by finding the linear combination of variables that maximizes correlation between the sets. In the general form, CCA seeks to maximize Cor(Χ1*w*1, Χ2*w*2) where *w*1 and *w*2 are canonical weights corresponding to Χ1 and Χ2, respectively. The canonical weight objective function is defined in Equation 2.

****A picture containing text, font, white, algebra

Description automatically generated

*Sparse Multiple CCA*

SmCCNet uses an adaptation of CCA, called sparse multiple CCA (SmCCA), to incorporate a third data type, Υ (i.e., the phenotype of interest), into the integration task. SmCCA accomplishes multimodal integration by accounting for phenotype—omic correlation within the canonical weight objective function. Sparsity is imposed on the canonical weight vectors by further adding penalization terms to Equation 2 (functions *P*1(•) and *P*2(•)). The least absolute shrinkage and selection operation (LASSO) is used within the SmCCA implementation used by SmCCNet. The extended canonical weight objective function that allows for sparse, multivariate integration is defined in Equation 3. In this objective function, *c*1 and *c*2 are the pre-specified penalty constants and *a*, *b*, and *c* are scaling constants that can be used to prioritize specific between-set correlations (i.e., microbiota—protein, microbiome—%ΔPExS, or protein—%ΔPExS correlation). Shi et al suggest that weighted SmCCA (i.e., the case in which *a*, *b*, and *c* are not all equal) might be preferred when the correlation between phenotype and omic data is weaker than the correlation between the omic data sets.

**A black number with a white background

Description automatically generated with low confidenceA picture containing text, font, white, algebra

Description automatically generated**

*SmCCNet*

SmCCNet uses SmCCA and a feature subsampling approach to obtain a set of canonical weight vectors, *w*1 and *w*2. The determined weight vectors are represented as a singular weight vector *u*, where *u* = (*w*1, *w*2). The 1-dimensional weight vector *u* is used to construct a similarity matrix *A* following *A* = *u* ⊗ *u* (i.e., the outer product). A dissimilarity (or distance) matrix, *D*, is then calculated following *D* = 1 − *A*. *D* is used as input for hierarchical clustering and the resultant hierarchical tree is cut to form distinct subnetworks that are associated with the phenotype of interest. 500 feature subsampling iterations and default parameters were used for SmCCNet analysis unless stated otherwise.

*SmCCNet Hyperparameter Tuning*

As mentioned, SmCCNet induces sparsity into the CCA framework by incorporating two LASSO-based penalty parameters (i.e., one penalty parameter for each of Χ1 and Χ2). Penalty parameters were selected using five-fold cross-validation (CV) and a randomized grid search approach. The selected penalty pair corresponds to the penalty pair that minimized the prediction error between the CV training and test sets. Unlike typical LASSO penalties, increasing the value of a penalty parameter weakens the strength of regularization applied to a given omic set in SmCCNet’s implementation. We searched a range of larger values for Χ1’spenalty (ranging from 0.40 to 0.60) and a range of smaller values for Χ2’spenalty (ranging from 0.10 to 0.30) since the dimensionality of Χ1 (60 taxa) after pre-processing was much smaller than the dimensionality of Χ2 (4,001 aptamers). The aim was to impose a greater penalization on Χ2 in order to generate subnetworkswith an improved balance of proteins and taxa. We further used imbalanced feature subsampling proportions (i.e., 0.90 and 0.70 for Χ1 and Χ2, respectively) to additionally account for the feature imbalance between omics sets. Ten feature subsampling iterations were used during CV.

Scaling coefficient *b* (i.e., *b* in the SmCCA objective function, Equation 3) was additionally tuned using a manual hyperparameter tuning process. We explored the *weighted* version of SmCCNet since the observed correlation between Χ1 and Υ (microbiome–%ΔPExS) was noticeably weaker than the other between-set correlations (Supplemental Figure 2). Separate five-fold CVs were performed to select penalty parameters for each of the tested weighting schemes. We considered subnetwork strengths (i.e., subnetwork–phenotype correlations), subnetwork sizes (i.e., number of nodes), and microbiota–protein balance when determining the optimal weighting scheme (Supplemental Table 1). The aim was to generate robust and reasonably sized multi-omic subnetworks that are strongly correlated with PEx recovery (ρ > 0.30).

*Subnetwork Summarization*

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 (Χ1) and proteins (Χ2) were merged into a common matrix for PCA. Subnetwork—phenotype strength was calculated as the correlation between a subnetwork’s PC1 and %ΔPExS. Absolute correlation values are reported since the use of PC1 obscures the interpretability of negative and positive relationships. The percentage of variance explained by each PC1 was additionally calculated and reported.

*Subnetwork Node Pruning*

Signals from gene enrichment analysis are typically too weak to detect if gene lists are large relative to the set of background genes (i.e., approximately 3,600 genes targeted by the SomaScan assay) [ref]. We therefore aimed to limit subnetworks to 325 or less nodes, which represents less than 10% of the total feature space. A 325-node limit was chosen since we expect subnetworks with 325 nodes to contain approximately 300 proteins (genes). The following pruning process was performed to trim large subnetworks to 325 nodes. First, the PageRank algorithm was used to rank nodes by importance. Briefly, the importance of a node *N*  is measured by considering the amount and importance of other nodes connected to *N*. Second, nodes ranked outside the top 325 nodes were removed, or pruned, from the subnetwork.

*Subnetwork Edge Pruning*

Subnetworks often contain a set of insubstantial node-to-node connections (or edges) that might suggest erroneous associations. Our aim was to remove, or prune, these edges to improve biological interpretability. We used between-node pairwise correlations for systematic removal of weak edges, removing edges between nodes that were weakly correlated. A between-node correlation of 0.20 is used as the edge inclusion threshold for subnetwork visualizations unless stated otherwise.

*Gene Ontology Enrichment Analysis*

GO enrichment analyses were performed on selected subnetworks using subnetwork-specific protein sets and Metascape [ref]. The aim was to identify over-represented GO pathways (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 (3,561 unique proteins). UniProt protein IDs were used as input. P-value and minimum enrichment cutoffs of 0.001 and 3 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 biological process protein lists.

*Subnetwork Visualization*

Network visualizations were created using CytoScape (version 3.9.1) [ref]. We decided to focus visualizations on one to two enriched GO biological processes per selected network, rather than full subnetworks, to preserve interpretability and focus discussion. A given subnetwork visualization includes the overlapping proteins (i.e., those proteins that are found in both the subnetwork and GO biological process) and subnetwork-specific taxa. Subnetwork visualizations provide information about individual node—phenotype associations, node—node relationships and connectivity, and canonical weight contributions.

**Results**

*Study Demographics*

Cohort characteristics and PEx event clinical measures are described in Table 1. The distribution of %ΔPExS in our cohort is shown in Supplemental Figure 1. The median PExS at study enrollment was 12 and ranged between 8 and 16. %ΔPExS ranged from -100% to 0%, indicating that PExS either improved (i.e., decreased) or remained constant from hospital admission to discharge for all exacerbations.

**A picture containing text, screenshot, font, number

Description automatically generated**

**Table 1.** Study cohort demographics and clinical measures at hospital admission. %ΔPExS, the phenotype of interest in this study, is reported as well. Continuous demographic and clinical information are reported as the median and range in squared brackets. Demographic information is reported at the subject level while clinical information is reported at the PEx event level.

*Selected SmCCNet Configuration & Subnetworks*

A weighting scheme of (1, 2, 1) (i.e., *a* = c = 1, *b* = 2) with a CV selected penalty parameter pair of (0.15, 0.40) was found to generate the most desirable subnetworks when considering the following subnetwork characteristics: subnetwork–phenotype correlations, subnetwork sizes, and protein–taxon balance (Supplemental Table 1). Ten subnetworks resulted from this SmCCNet configuration including four selected subnetworks with a strong correlation to %ΔPExS (ρ > 0.30) (Figure 1). Selected subnetworks include subnetworks 2, 4, 5, and 6. These subnetworks are described in depth in subsequent sections.

A screenshot of a graph

Description automatically generated with low confidence

**Figure 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. Edge weights correspond to the values calculated in similarity matrix *A* (described in the Methods section). Edges for all unselected subnetworks are additionally shown. 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.

*A picture containing text, receipt, font, screenshot

Description automatically generated*

**Table 2.** SmCCNet subnetwork summarization metrics for the selected SmCCNet subnetworks (i.e., subnetworks with strong associations to %ΔPExS, or ρ > 0.30). Metrics include subnetwork–phenotype correlation estimates, subnetwork size information, and node–phenotype correlation ranges.

*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 Χ1, Χ2, and %ΔPExS (Equation 3, Figure 1B). The original subnetwork constructed by SmCCNet contained 364 nodes, including 334 proteins and 30 taxa. The pruned subnetwork, summarized in Figure 1, contains 325 nodes, including 298 proteins and 27 taxa. Correlations between individual nodes and %ΔPExS range from -0.568 to 0.623 (Figure 1A). A correlation of 0.450 is observed between PC1 of the full subnetwork and %ΔPExS (p = 0.009). The central node of the network is *Atopobium*, which is strongly associated with %ΔPExS (ρ = 0.444; *p* = 0.010). The edge that connects *Atopobium* and *Actinomyces* corresponds to the largest node-to-node weight observed in the integrated data set (i.e., the maximum weight calculated in matrix *A*). GO biological pathways that are enriched in Subnetwork 2 include GO:0048738 (cardiac muscle tissue development), GO:0042100 (B cell proliferation), and GO:1903587 (regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis) (Figure 7). GO:0042100 and GO:1903587 proteins are visualized with Subnetwork 2 taxa in Figure 2 and Figure 3, respectively.

A picture containing screenshot, line, diagram, circle

Description automatically generated

**Figure 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. An edge pruning threshold of ρ > 0.30 was used for improved visualization.

A picture containing screenshot, diagram, line, circle

Description automatically generated

**Figure 3.** Subnetwork 2, GO:1903587 (regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis). 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. An edge pruning threshold of ρ > 0.30 was used for improved visualization.

*Subnetwork 4*

Subnetwork 4 is comprised of 209 nodes, including 204 proteins and 5 taxa. 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 (ρ = -0.018; *p* = 0.919). Soluble low-density lipoprotein receptor-related protein 1 (sLRP1), the omic feature most strongly correlated with %ΔPExS, is included in this subnetwork (ρ = 0.634; *p* < 0.001). Subnetwork 4 is the strongest of the selected subnetworks when considering subnetwork—phenotype correlation (ρ = 0.451; *p* = 0.008) (Figure 1A). GO biological processes enriched in Subnetwork 4 include GO:2000328 (regulation of T-helper 17 cell-lineage commitment), GO:0010498 (proteasomal protein catabolic process), GO:0043161 (proteasome-mediated ubiquitin-dependent protein catabolic process), and GO:0090559 (regulation of membrane permeability). GO:2000328 proteins are visualized with Subnetwork 4 proteins in Figure 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 Figure 4, indicating that the taxon is weakly associated with GO:2000328 proteins (ρ < 0.20).

A diagram of a cell division

Description automatically generated

**Figure 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. An edge pruning threshold of ρ > 0.20 was used for improved visualization.

*Subnetwork 5*

Subnetwork 5 is a 208-node subnetwork comprised of 201 proteins and 7 taxa. 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* abundanceat PEx onset is positively associated with %ΔPExS (ρ = 0.233; *p* = 0.191). Individual node—%ΔPExS correlations range from -0.457 to 0.474 (Figure 1A). A correlation of 0.304 is observed between the subnetwork and %ΔPExS (*p* = 0.086). Edges are relatively weak in Subnetwork 5 when compared to the edges in the other selected subnetworks, both in terms of the edge strength median and range (Figure 1B). Weaker edges suggest that Subnetwork 5 nodes are collectively less influential in maximizing the canonical correlation than the nodes in other selected networks. GO biological processes found within the Subnetwork protein set include GO:0006412 (translation), GO:0043043 (peptide boisynthesis process), GO:0006417 (regulation of transcription), GO:0051092 (positive regulation of NF-kappaB transcription factor activity), GO:0043604 (amide biosynthetic process), GO:0051091 (positive regulation of DNA-binding transcription factor activity, and GO:0006413 (translational initiation). GO:0051092 proteins included in Subnetwork 5 are visualized with Subnetwork 5 taxa in Figure 5.

A picture containing screenshot, diagram, line, circle

Description automatically generated

**Figure 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. An edge pruning threshold of ρ > 0.20 was used for improved visualization.

*Subnetwork 6*

A diagram of a growth factor

Description automatically generated with low confidence

**Figure 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. An edge pruning threshold of ρ > 0.20 was used for improved visualization.

A picture containing text, screenshot

Description automatically generated

**Figure 7.** Enriched GO biological processes associated with each of the selected subnetworks. The bar plot shows the negative log-10 transform of the *p*-value associated with each GO biological process. Each bar includes information about the number of proteins overlapping the subnetwork and the GO biological process (*GenesInGOAndHitList*), the number of proteins in the GO biological process (*GenesInGO*), and the associated enrichment value (*Enrichment*). This information is presented in the following format: *GenesInGOAndHitList* / *GenesInGO* (*Enrichment*).

**Discussion**

* Where is staph aureus?

**Limitations**

* Assuming independent observations even though we have repeats
* Cohort limitations (sample size, generalizability)
* The aptamer approach – difficult to rationalize how to deal with aptamers that hit multiple genes
* Genus vs species level

**Supplementary Materials**

**Supplemental Figures**

Supplemental Figure 1.

A picture containing text, screenshot, line, diagram

Description automatically generated

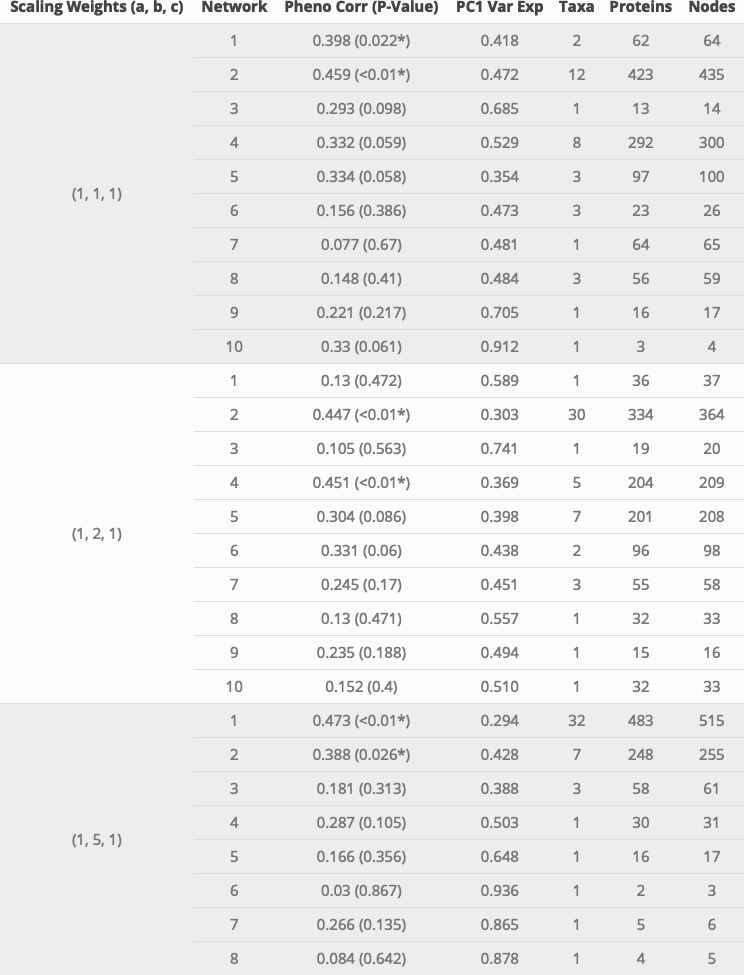
Supplemental Figure 2.

**A picture containing screenshot, diagram, line, plot

Description automatically generated**

**Supplemental Tables**

Supplemental Table 1.

****

**Citations**

1. Amadori A, Antonelli A, Balteri I, Schreiber A, Bugiani M, De R, V. Recurrent exacerbations affect FEV(1) decline in adult patients with cystic fibrosis. Respir Med. 2009 Mar;103(3):407-13
2. Liou TG, Adler FR, Fitzsimmons SC, Cahill BC, Hibbs JR, Marshall BC. Predictive 5-year survivorship model of cystic fibrosis. Am.J.Epidemiol. 2001 Feb 15;153(4):345-52
3. Sanders DB, Bittner RC, Rosenfeld M, Redding GJ, Goss CH. Pulmonary exacerbations are associated with subsequent FEV1 decline in both adults and children with cystic fibrosis. Pediatr.Pulmonol. 2011 Apr;46(4):393-400. PMCID:20967845
4. Sanders DB, Bittner RC, Rosenfeld M, Hoffman LR, Redding GJ, Goss CH. Failure to recover to baseline pulmonary function after cystic fibrosis pulmonary exacerbation. Am.J.Respir.Crit Care Med. 2010 Sep 1;182(5):627-32
5. Sanders DB, Hoffman LR, Emerson J, Gibson RL, Rosenfeld M, Redding GJ, Goss CH. Return of FEV1 after pulmonary exacerbation in children with cystic fibrosis. Pediatr.Pulmonol. 2010 Feb;45(2):127-34
6. Ordonez CL, Henig NR, Mayer-Hamblett N, Accurso FJ, Burns JL, Chmiel JF, Daines CL, Gibson RL, McNamara S, Retsch-Bogart GZ, et al. Inflammatory and microbiologic markers in induced sputum after intravenous antibiotics in cystic fibrosis. Am.J.Respir.Crit Care Med. 2003 Dec 15;168(12):1471-5
7. Colombo C, Costantini D, Rocchi A, Cariani L, Garlaschi ML, Tirelli S, Calori G, Copreni E, Conese M. Cytokine levels in sputum of cystic fibrosis patients before and after antibiotic therapy. Pediatr.Pulmonol. 2005 Jul;40(1):15-21
8. Chiron R, Grumbach YY, Quynh NV, Verriere V, Urbach V. Lipoxin A(4) and interleukin-8 levels in cystic fibrosis sputum after antibiotherapy. J.Cyst.Fibros. 2008 Nov;7(6):463-8
9. Sagel, S. D., Thompson, V., Chmiel, J. F., Montgomery, G. S., Nasr, S. Z., Perkett, E., Saavedra, M. T., Slovis, B., Anthony, M. M., Emmett, P., & Heltshe, S. L. (2015). Effect of treatment of cystic fibrosis pulmonary exacerbations on systemic inflammation. *Annals of the American Thoracic Society*, *12*(5), 708–717. https://doi.org/10.1513/AnnalsATS.201410-493OC
10. Shi, W. J., Zhuang, Y., Russell, P. H., Hobbs, B. D., Parker, M. M., Castaldi, P. J., et al. (2019). Unsupervised Discovery of Phenotype-specific Multi-Omics Networks. Bioinformatics 35, 4336–4343. doi:10.1093/bioinformatics/btz226
11. Rosenfeld M, Emerson J, Williams-Warren J, Pepe M, Smith A, Montgomery AB, Ramsey B. Defining a pulmonary exacerbation in cystic fibrosis. J.Pediatr. 2001 Sep;139(3):359-65