DEBUGGING LUNG DISEASES: APPLYING MATHEMATICAL TECHNIQUES, INVOLVING MODELLING, DATA INTEGRATIONAND MACHINE LEARNING FOR PRECISION MEDICINE

QUALIFYING EXAMINATION REPORT

SUBMITTED BY: JAYANTH KUMAR NARAYANA
MATRICULATION NUMBER : G1902804D

Lee Kong Chian School of Medicine, Singapore

Contents

List of Figures

List of Tables

0.1 Introduction

Introduction of the thesis

Chapter 1

"Integrative microbiomics" reveals a disrupted interactome in bronchiectasis exacerbations

1.1 Introduction

The term microbiome is used to refer to the collection of genes within a community of microbes (including bacteria, fungi, virus, protists and bacteriophages). In the last few years, microbiome research has helped us gained new insights into how microbes shape our human biology and have brought paradigm-shifting implications for translational research and clinical care. The human microbiota is crucial for our body to maintain its homeostasis. Disruption of this can lead to diseases such as obesity, inflammatory bowel disease, malnutrition, Parkinson's, Autism, Asthma, dental caries, bacterial vaginosis, and depression [17]. Currently, microbiome researchers use culture-independent techniques that involve DNA sequencing to derive the microbiome. Broadly, the community taxonomy/microbiome can be identified using two approaches (see Figure1.1) 1) Targeted and 2) Metagenomic. Targeted sequencing approach uses the PCR amplified, target gene markers (16S rRNA in case of bacteria or ITS in case of Fungi) derived from the samples to reference it against gene-marker databases (Silva, Green Genes, etc.). In contrast, the metagenomic sequencing approach directly sequence the whole community DNA and compares it to reference genomes [25].

Present microbiome studies focus on a single profile of the human microbiome in isolation, even though bacteria, fungi and viruses coexist and interact in the body as a community. Thus, it is essential to look at these biological components together in an integrated fashion to understand more holistically the true underlying in vivo state. However, one of the primary reasons for the lack of multi-biomic research is the lack of methods to merge microbiome datasets and integrative analysis. Consequently, I tried addressing some of these challenges in my master's thesis, using microbiome datasets derived from bronchiectasis patients as an example. Bronchiectasis, is a chronic inflammatory respiratory disease associated with progressive, irreversible dilatation of the airway. It is crucial to study bronchiectasis because in most cases it is known to be idiopathic(unknown cause) [8] and it is a significant contributor to lung diseases globally with a substantial four-fold higher predominance in Asian populations [28].

Previously in my master's thesis, I developed weighted similarity network fusion (wSNF) to allow weightage of input datasets during integration, otherwise unaccounted by conventional SNF [31]. Ensemble-based co-occurrence analysis strategy developed by Faust et al. [10] was extended to allow weightage of

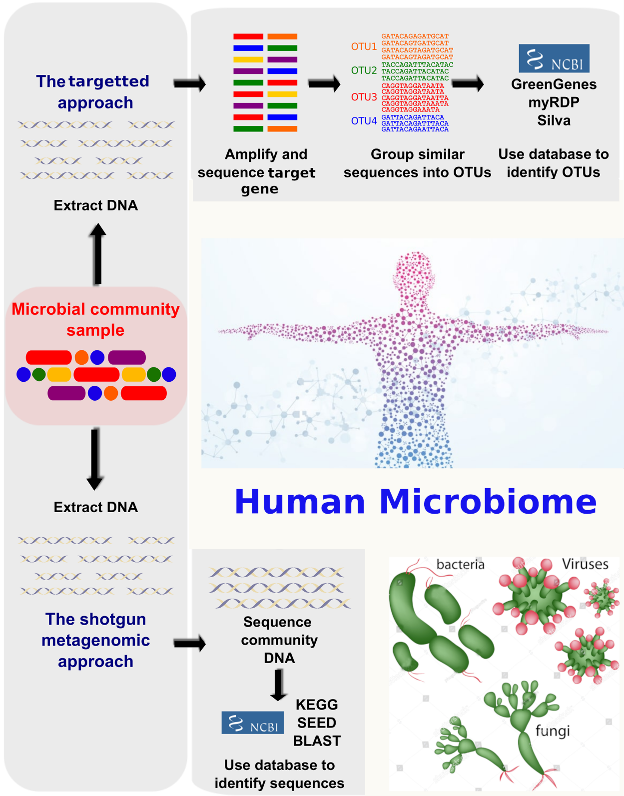


Figure 1.1: A figure illustrating the different sequencing approaches used to derive the human microbiome, consisting of interacting bacteria, fungi and viruses. Adapted from: [25]

individual methods in the ensemble along with other modification to better infer microbial association networks. Microbiome and Mycobiome derived using targeted amplicon sequencing of the 16S and ITS regions from the sputum samples of the CAMEB cohort [22]; virome from qPCR on an extensive panel of 17 respiratory viruses, were used as the example dataset to integrate the microbiomes (Figure1.2a). Multi-biome (Microbiome, Mycobiome and Virome) integration by wSNF identifies a high-risk exacerbation cluster with increased precision (Figure1.2b). Co-occurrence network analysis of this high-risk cluster revealed an elevated antagonistic interactome with reduced alpha-diversity (Figure1.2c) [26].

Having developed the wSNF and shown its increased precision to identify exacerbators (clinical outcomes); here in this chapter of my PhD thesis, I attempt to extend my results further. I aim to develop a web tool to enable users to integrate their microbiome datasets and to illustrate its advantages using publicly available microbiome datasets. The tool would motivate clinicians and microbiome researchers to explore multi-biome strategies for their problem and aid them in integrating their datasets. Secondly, I aim to study exacerbation events, antimicrobial perturbations and "Time to next exacerbation" using the developed "Interactome" framework. Thirdly, I aim to validate the "high-risk" exacerbation cluster of Bronchiectasis patients and its "interactome" as derived in my previous work [26] using an alternate sequencing approach: metagenomics. Further, we also pick an interaction from the interactome of the high and low-risk clusters and validate it experimentally.

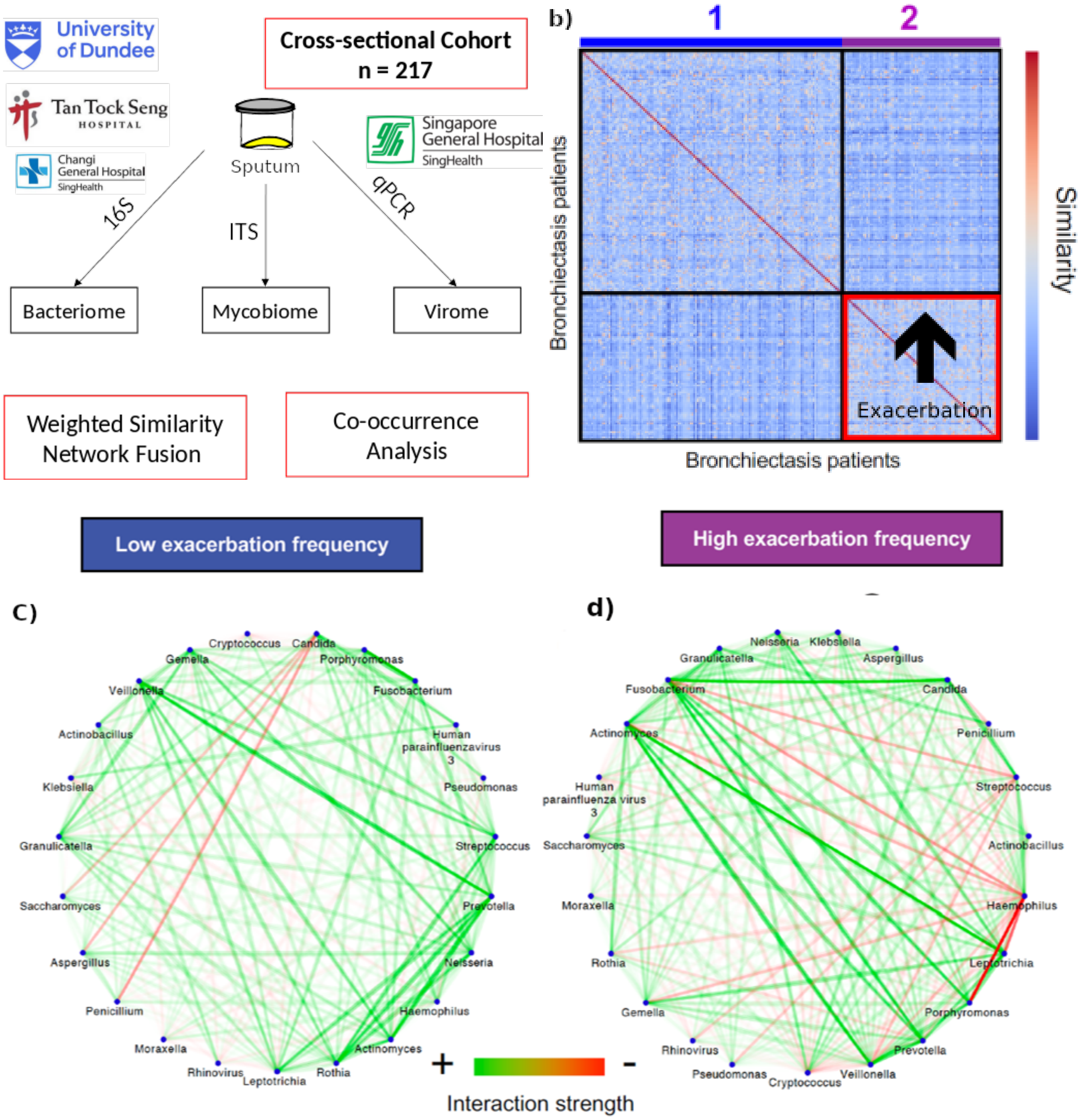


Figure 1.2: (a) A schematic representing, overview of analysis performed on the CAMEB cohort (n=217). Methodologies: Weighted SNF and Co-occurrence analysis were used for microbiome integration and intreactome construction. (b) A patient similarity matrix with each cell representing the integrated similarity between patients. Two clusters of low (black) and high (red) risk patients identified by wSNF are highlighted by boxes. Visualization of the interactome of low (c) and high (d) risk clusters. Interactions between microbes are classified as negative if the sign of the edge weights between them is negative (coloured red) with positive interactions indicated by green colouration. The strength of the interaction is indicated by the colour depth

1.2 Methods

1.2.1 Integrative-microbiomics, a webtool

Given the input microbiome datasets, the tool converts them into patient/sample similarity networks for each view based on the user-specified similarity measure before merging them using the user-specified algorithm. Further, the tool then implements a spectral clustering algorithm to allow cluster analysis on the merged dataset outputting the cluster assignments for each sample/patient. The optimum default number of clusters

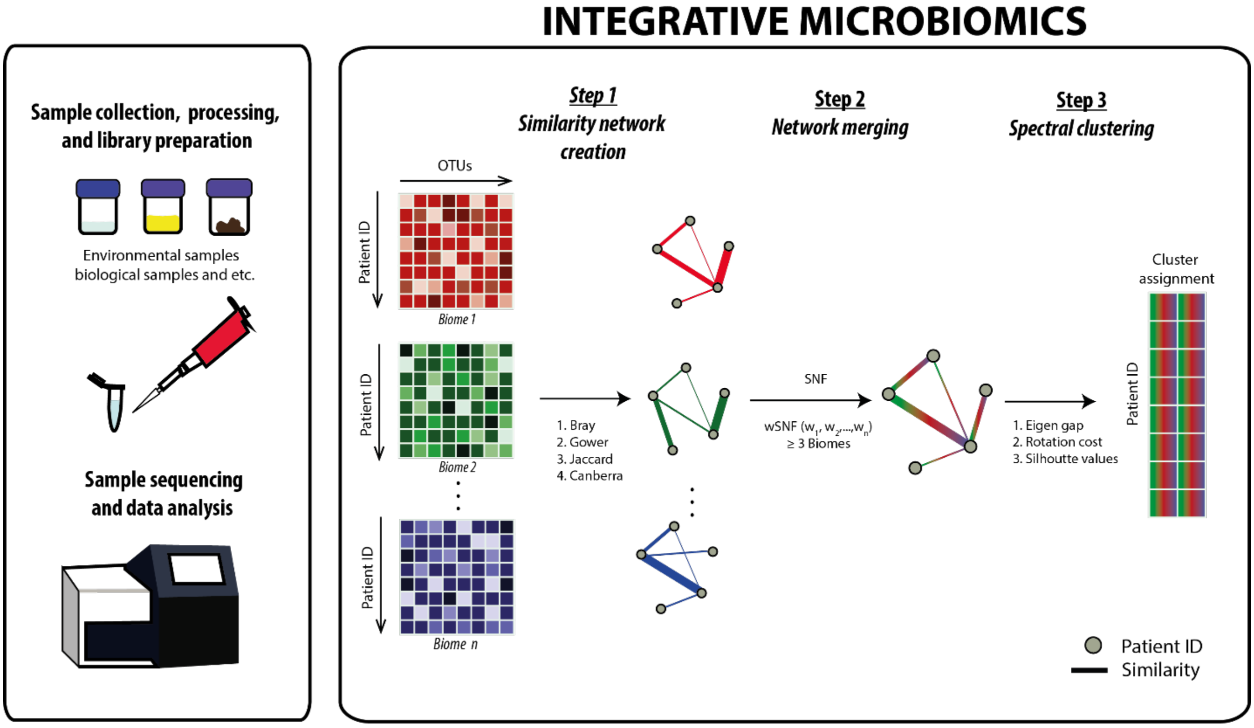


Figure 1.3: A figure describing the workflow of integrative microbiomics. The input microbiome datasets, are converted into patient/sample similarity networks based on the user-specified similarity measures: 1) Bray-Curtis, 2) Gower, 3) Canberra and 4) Jaccard; before merging them using the user-specified algorithm: 1) SNF, 2) wSNF. Further, the tool then implements a spectral clustering algorithm to allow cluster analysis on the merged dataset.

is computed using ensemble-based voting of three differing methodologies: Best Eigen Gap, Rotation cost and average silhouette method (Figure1.3). For a given value of 'k' (the number of clusters), we calculate a score/vote using the below rules

1. If the average silhouette score ¿ 0.7 → Score = Score + 3
2. If 0.5 ¡ average silhouette score ¡ 0.7 → Score = Score + 2
3. If 0.3 ¡ average silhouette score ¡ 0.5 → Score = Score + 1
4. If k equals the first best value as derived from eigen gap method → Score = Score + 3
5. If k equals the second-best value as derived from eigen gap method → Score = Score + 2
6. If k equals the first best value as derived from rotation cost method → Score = Score + 3
7. If k equals the second-best value as derived from rotation cost method → Score = Score + 2

The value of k for which the Score is the highest is chosen as the default optimum number of clusters. In addition, the tool also outputs the integrated similarity matrix which can be used for downstream analysis such as for label propagation and survival analysis [31].

The tool presently provides four similarity measures 1) Bray-Curtis, 2) Gower, 3) Canberra and 4) Jaccard, appropriate for microbiome datasets which is used to construct patient/sample similarity network and two approaches 1) SNF, 2) wSNF to integrate these networks. For the implementation of wSNF the following formula in SNF

$$P^{(v)} = S^{(v)} \times \frac{\sum_{k=v} p^k}{(m-1)} \times (S^{(v)})^T, v = 1, 2, 3, \ldots, m$$

was modified into

$$P^{(v)} = S^{(v)} \times \frac{\sum_{k=v} \omega_k \times p^{(k)}}{\sum_{k\neq v} \omega_k} \times (S^{(v)})$$

$v = 1, 2, 3, \ldots, m$ where ω_k is the weight of the k^{th} dataset, m the total number of views, P the status matrix and S the kernel matrix as defined by Wang et.al [31].

This webtool allows the users to integrate multiple microbiome datasets obtained from different sites in a patient/biological entity or from various methods (targeted sequencing, metagenomics and qPCR) from the same site. For example, the lung microbiome (bacteria) with the gut microbiome (bacteria) or the lung microbiome (bacteria) with lung mycobiome. The tool assumes each input microbiome datasets represent a view of an underlying biological mechanism or a disease. Reliable estimation of each view is assumed when using SNF [16]. However, it may not be always practical to reliably estimate each view, although they play an equal role in the underlying biological process. This is due to the limitations and differing rates of development, in the present technologies and reference databases. In such cases, a weighted SNF approach is preferred, which still assumes the input datasets share an underlying biological mechanism but accounts for inconsistency of the microbiome data based on the user specified weights. The default weights are assigned based on the taxonomical richness (i.e. the number of microbes present) of the datasets.

The interface of the webtool was developed using Rshiny and is available through Shiny Server (Open Source) in confluence with nginx-1.19.1. The tool is powered by custom scripts written in python2.7 and R; and containerized using Docker for ease of offline implementation. The developed webtool can be accessed at `https://integrative-microbiomics.ntu.edu.sg`.

1.2.2 Longitudinal assessment of Exacerbation

A longitudinal cohort of n=17 patients were recruited from two hospitals in the east of Scotland (2016-2017) to study changes in the microbiome during exacerbation and following antibiotic treatment. DNA and RNA extraction were performed on sputum samples obtained from each patient and on a blank sterile PBS (Phosphate buffer solution). The extracted DNA was subjected to targeted amplicon sequencing of the 16S rRNA and ITS2 regions of the genome to derive the Microbiome and Mycobiome, by mapping them to green genes and UNITE databases, respectively. Blank samples contained read counts many orders of magnitude lower than test samples and hence unlikely to have any influence on the observed microbiome. RT-qPCR (real time quantitative polymerase chain reaction) was performed on the cDNA derived from the extracted RNA to quantify the viral burden of the 17 viruses investigated in each patient. α and β diversity of the multi-biome was calculated from the concatenated microbiome and the integrated patient similarity matrix using the vegan package in R.

1.2.3 Antibiotic action simulation

To predict the impact of antibiotics on the interactome, β-lactam antibiotic action was simulated by a 75% reduction in the relative abundance of the microbes targeted by this antibiotic in the baseline (pre-antibiotic) state including the following genera: *Streptococcus, Staphylococcus, Haemophilus, Moraxella, Actinomyces, Arachnia, Bacteroides, Bifidobacterium, Eubacterium, Fusobacterium, Lactobacillus, Leptotrichia, Peptococcus, Peptostreptococcus, Propionibacterium, Selenomonas, Treponema and Veillonella.* In order to remove interactions resulting from random noise at the expense of sensitivity to weak signals and to allow comparison between the derived interactomes, the following abundance and prevalence filters were applied followed by co-occurrence analysis; retention of microbes present at greater than 1% abundance in at least three subjects; in the pre OR post OR modelled antibiotic state.

1.2.4 "Time to next exacerbation" prediction

To predict Time to next exacerbation, Microbiome datasets were CLR (Centred log ratio) transformed before concatenation and microbes that are present in at least 4 patients at an abundance of 1% were considered for further analysis. To derive pairwise microbial interactions for each patient, LIONESS [18], a single patient network inference framework was implemented with General Boosted Linear model (GBLM) as the network inference algorithm. Correlation between the abundance of each microbes and interaction strength with time to next exacerbation was assessed using Spearmans rank correlation with statistical testing. Multivariate adaptive regression spline (MARS) [13], a non-linear regression model was implemented with microbes or interaction strength as the predictor variable to predict time to next exacerbation groups; defined as (Time to exacerbation: <12 weeks and >12weeks). The goodness of the fit of the model was evaluated by computing the R-squared (RSq) and the Generalized R-squared metric (GRsq). A feature importance plot based on Generalized Cross validation score (gcv) was also computed on the feature selected (microbes) by the model. All the above analysis was implemented in R using the following packages 1)Hmisc 2) earth 3)vegan 4)compositions 5)lionessR.

1.2.5 Validation of the interactome

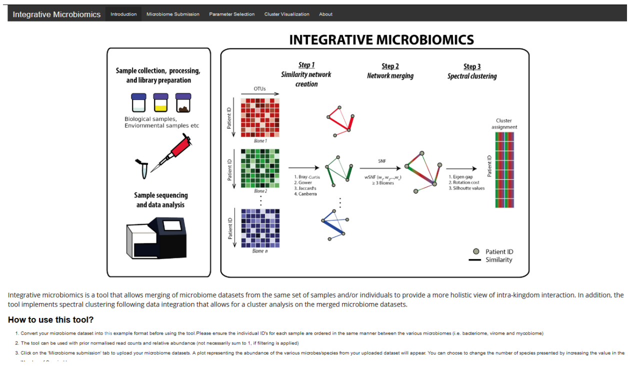
Experimental microbiological validation of Pseudomonas aeruginosa and Aspergillus fumigatus interaction was performed using one strain Aspergillus fumigatus (Af293) and three strains of Pseudomonas aeruginosa: (1) lab strain (PAO1) as control and (2,3) two clinical isolates of Pseudomonas aeruginosa derived from low-risk and high-risk patient clusters. The interaction was investigated using the disk inhibition method as described by Homa et al. [15]. An independent cohort of 166 patients was recruited from 4 sites (3 in Singapore and 1 in Dundee, Scotland) to validate the high-risk cluster and its interactome. DNA extraction was performed on the collected sputum samples of each patient. A shotgun metagenomic sequencing was performed at the NTU core sequencing facility on these samples according to the methods described by Gusareva et al. [14]. Kaiju [24] with default parameters was implemented on the raw sequences after human read removal to estimate the taxonomic composition by referencing against NCBI BLAST nr+euk database. Estimation of the viruses that include prokaryotic phages and eukaryotic viruses was implemented using a custom pipeline that uses Demovir (`https://github.com/feargalr/Demovir`).

1.3 Results

1.3.1 Integrative-microbiomics, a webtool

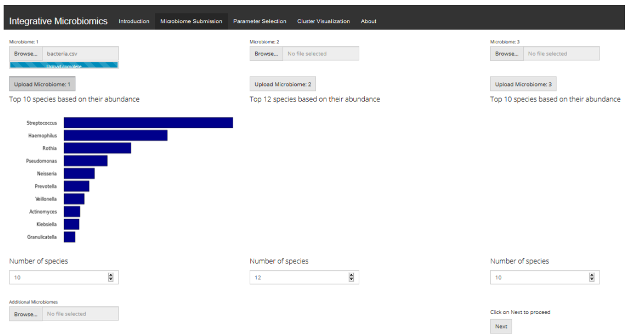
To motivate and enable; researchers and clinicians to opt for an integrative strategy when analysing multi biome datasets, we developed a web tool Integrative Microbiomics available at `https://integrative-microbiomics.ntu.edu.sg`. The tool implements spectral clustering following data integration which allows clustering based on a holistic view of the integrated dataset. The web tool consists of simple design layout with five full page tabs: 1) Introduction, 2) Microbiome submission, 3) Parameter selection, 4) Cluster Visualization and 5) About; providing ease of access and navigation to the users.

Introduction tab



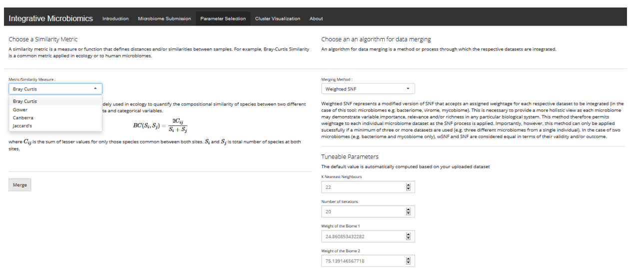
This tab serves as the landing page of the web tool, containing an illustration of the workflow and a section on how to use the webtool. This page also provides an example format (csv) for the users to input their microbiome datasets.

Microbiome submission tab



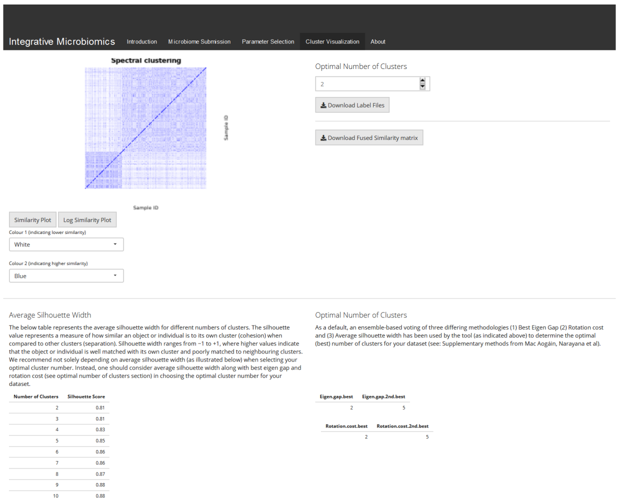
This page enables users to upload microbiome datasets, which they intend to integrate. The tool produces an abundance plot of top 10 microbes (can be modified by a scroll bar) if the microbiomes are successfully uploaded. Users who wish to integrate more than three microbiome datasets can use, the additional microbiomes option to upload any number of microbiome datasets. A maximum file size of upto 30MB is accepted for each microbiome dataset.

Parameter selection tab



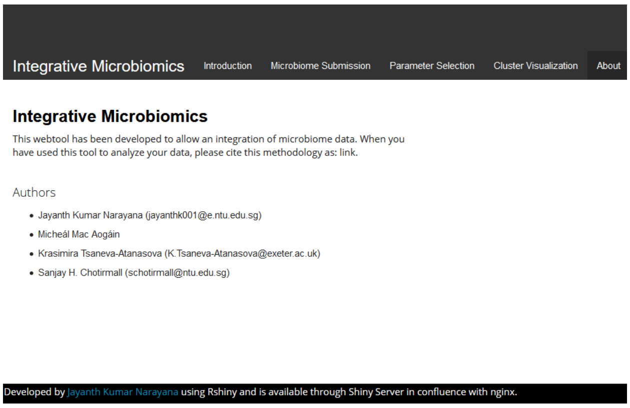
This tab allows the users to select the merging algorithm 1)SNF or 2)Weighted SNF and a similarity measure. It also provides an option to set the model hyper-parameters of the merging algorithm such as number of iterations, the weights and the value of k nearest neighbours. The default values of these hyper parameters are set based on the input dataset.

Cluster visualization tab



This tab allows the users to visualize; the user-specified optimum number of clusters, as a heatmap of the patient/sample similarity matrix. The default value for optimum number of clusters is calculated based on ensemble-based voting of three differing methodologies as described in the methods. Further, this page also outputs the results from the three methodologies to aid the users to select the optimum number of clusters. Additionally, this page provides option for users to download the cluster memberships of samples/patients (as csv) and the integrated patient/sample similarity matrix (as csv).

About tab



This page provides the link to the article, which the users could cite if they had used the tool. This page also provides a list of authors and their contacts, to report any concerns and feedback.

1.3.2 "Integrative Microbiomics" identifies biologically relevant clusters

To evaluate the added advantage of integrating datasets over singular analysis, and superiority of Integrative Microbiomics over conventional concatenation we performed an unsupervised clustering on three datasets (two publicly available datasets and CAMEB derivation/example dataset) and then evaluated the clusters on the available meta-data. Clustering of the individual biomes and the concatenated biomes were performed using spectral clustering and cluster comparisons using a chi-squared or Kruskal-Wallis test wherever applicable. The cluster consistency was assessed using Average Silhouette score. Assessing the results from the three examples (Table 1.1, 1.2 and 3) shows an increased cluster consistency for the clusters derived using integrative microbiomics. Additionally, we observe an increased precision in identifying meaningful clusters, reflected by the decrease in the p-value for the evaluation of the features/meta-data between the clusters.

Example 1: Oral Lichen Planus(OLP) dataset

All paired-end fastq files containing ITS2 and 16S rRNA sequences and the accompanying meta-data of the saliva samples under accession number SRP067603 were retrieved from NCBI SRA as described in [21]. Pre-processing (filtering, trimming, de-replication, merging paired reads, removal of primers and chimeras) and taxonomy profiling (using UNITE 02.02.2019 release for ITS2 and Silva version 132 for 16S rRNA) were carried out using DADA2 package. The resulting datasets of 52 samples were integrated using integrative microbiomics webtool with k=6 and method=SNF.

	Bacteria	**Fungi**	**Concatenated**	**IM**
Number of Clusters	3	3	2	5
Silhouette	0.28	0.69	0.326	0.85
Class: Healthy or Erosion or Reticulate	NS	0.025	NS	0.0385
IL17	NS	NS	NS	0.002
IL23	NS	NS	NS	0.03
Age	0.033	NS	NS	NS

Table 1.1: A table representing the optimal clusters derived from various views of the dataset using Spectral clustering with Bray-Curtis similarity and p-value for the assessment of meta-data on the derived clusters, computed using chiq-squared test or kruskal-wallis test, wherever appropriate. The optimal number of clusters was calculated using the eigen gap method; followed by an assessment of cluster consistency (Average silhouette width). NS- Non-significant (p-value > 0.05). IM: Integrative Microbiomics

Example 2: Ecological (Soil) dataset

Bacterial and Fungal OTU table described in [30] along with their meta-data was downloaded from the supplementary materials. These datasets on 48 samples were then integrated using integrative Microbiomics webtool with k=3 and method=SNF.

	Bacteria	Fungi	Concatenated	Integrative Microbiomics
No. of clusters	2	2	2	2
Silhouette	0.44	0.645	0.51	0.92
Block	NS	NS	NS	NS
Treatment	0.0005	0.0005	0.0005	0.0005
Decom	5.745e-07	1.863e-06	6.784e-07	5.942e-07
N2O	0.03185	0.04419	0.02942	0.01801
grNt	3.455e-05	0.0002598	2.131e-06	1.093e-06
herbNt	0.0001019	0.0004297	1.331e-05	4.315e-08
legNt	0.002282	0.004783	0.0001821	1.106e-05
grPt	9.889e-06	5.978e-06	6.784e-07	5.563e-05
herbPt	3.822e-05	0.0005955	7.879e-06	2.567e-09
legPt	0.0006874	0.003719	0.0001038	4.614e-06
Pleach	0.01	0.008947	0.002958	0.00924
Nleach	0.003096	0.0001295	0.001887	0.002256
aveMF	0.001669	0.0257	0.002037	9.519e-05
pcaMF	5.744e-06	6.273e-05	6.784e-07	9.731e-10

Table 1.2: A table illustrating the optimal clusters derived from various views of the dataset using Spectral clustering with Bray-Curtis similarity and p-values for the evaluation of meta-data using chiq-squared test or kruskal-wallis test, wherever appropriate are reported. The optimal number of clusters was calculated using the eigen gap method; followed by an assessment of cluster consistency (Average silhouette width). NS- Non-significant (p-value < 0.05)

Example 3: CAMEB dataset

Microbiome, mycobiome and virome were derived from 217 patients of the CAMEB cohort [22]. This dataset was used as the example dataset to derive and evaluate Integrative Microbiomics. These datasets were integrated using integrative microbiomics webtool with k=8, method=" Weighted SNF and other parameters were set to default values.

Dataset assessed	Optimum cluster number	Median Exacerbation number			p-value
		Cluster 1	Cluster 2	Cluster 3	
Bacteriome alone (B)	3	2	1	2	0.00021
Mycobiome alone (M)	3	1	1	1	n.s
Virome alone (V)	3	1	1	1	n.s
SNF-network (B + M + V) - unweighted	3	2	1	1	0.039
SNF-network (B+ M+ V) - weighted	2	2	1	-	0.000024

Increased precision with integration of additional biomes and weighted analysis

Figure 1.4: Table 3: A table illustrating the optimal clusters derived from various views of the dataset using Spectral clustering with Bray-Curtis similarity and p-values for the evaluation of meta-data using kruskal-wallis test are reported. The optimal number of clusters was calculated using the eigen gap method. NS- Non-significant (p-value < 0.05)

1.3.3 Increased antagonistic interaction during exacerbation with no difference in microbial diversity, α and β diversity

Next, to study exacerbation events using the interactome framework; we assessed the interactome at baseline (pre-exacerbation), during exacerbation and post-exacerbation. A comparison of the longitudinal multi-biome signatures across time points revealed no significant difference in microbial composition (Fig 1.5 a), α (Fig 1.5 b) and β (Figure 1.5 c) diversity suggesting overall stability of the microbiome during exacerbation. On the contrary, co-occurrence analysis reveals significant changes in the interactome with an increased number and strength of negative interactions during exacerbation as opposed to baseline and post exacerbation.

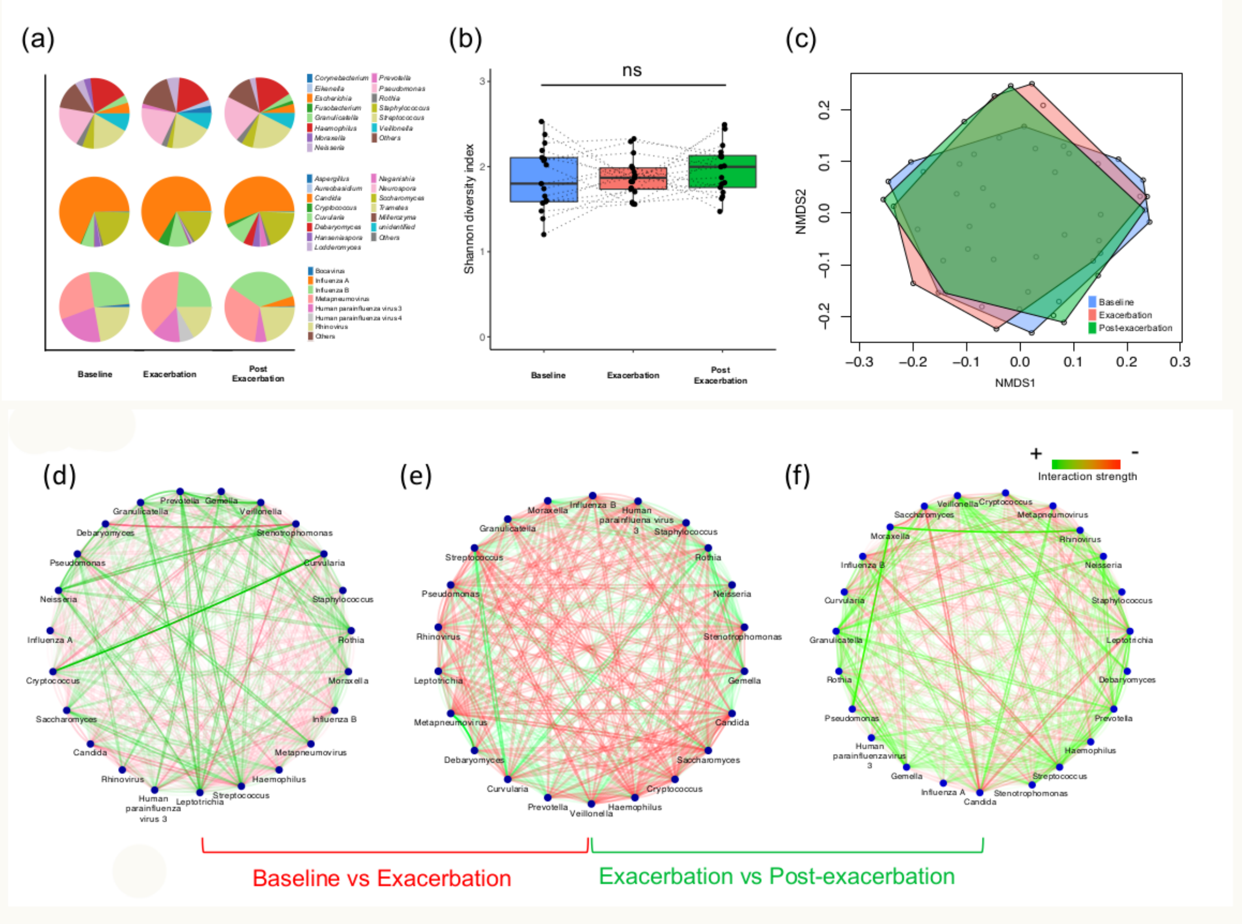


Figure 1.5: Longitudinal analysis of the integrated multi-biome during bronchiectasis exacerbations. (a) Bacterial, fungal and viral community status were assessed longitudinally in n=17 bronchiectasis patients at baseline (pre-exacerbation) (B), during an established pulmonary exacerbation (E) and then post-exacerbation (P) following completion of antibiotic therapy. Pie charts illustrate aggregate microbial composition of the bacterial, fungal and viral community profiles across each time point with the most abundant taxa indicated by the colour legend. (b) Boxplots illustrating comparable α-diversity across baseline (B), exacerbation (E) and post-exacerbation (P) specimens. Dotted lines indicate the longitudinal pattern of each individual patient (n=17). (c) Non-metric Multi-Dimensional Scaling (NMDS) plot illustrating comparable multi-biome β-diversity across baseline (B), exacerbation (E) and post-exacerbation (P) specimens. Samples are grouped according to their respective longitudinal timepoint. (d-f) Visualization of the interactomes positive and negative interactions between the most abundant taxa at (d) baseline (pre-exacerbation), (e) during exacerbation and (f) post-exacerbation. Interactions between microbes are classified as negative if the sign of the edge weights between them is negative (coloured red) with positive interactions indicated by green colouration.

In-depth analysis of changes in the interactome from baseline to post-exacerbation state reveals a reduced number of total interactions in the post-exacerbation state, likely explained by the broad antibiotics usage which eliminates potential interacting microbes.

1.3.4 Simulation of the antibiotic action using the Interactome framework reliably predicts the rank order difference of key microbial taxa

To evaluate the clinical utility of our derived network-based interactomes by predicting the influence of antibiotic exposure on its contained microbiome. As several (n=12) patients of our longitudinal cohort received -lactam antibiotics for treatment of their initial exacerbation, we used the baseline (pre-β-lactam exposure) interactome network (Figure 1.6a) to predict network reconfiguration post β-lactam treatment by artificially reducing the abundance of β-lactam-sensitive microbes by 75% (Figure 1.6b). We then compared

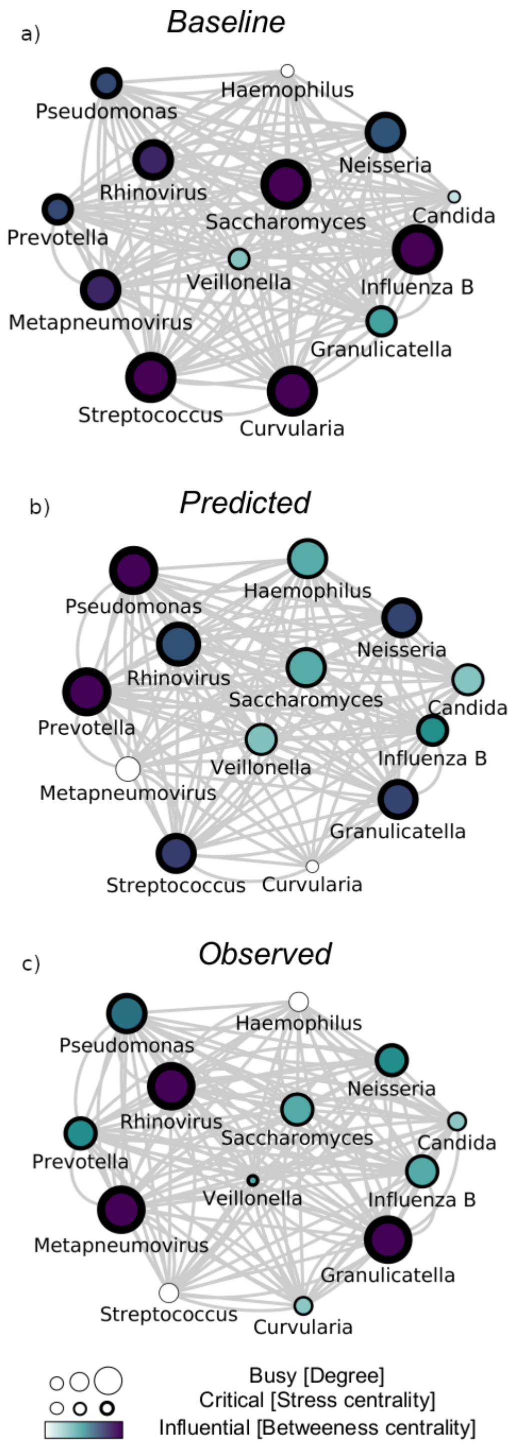


Figure 1.6: (a) Baseline network analysis of bronchiectasis patients who subsequently received β-lactam therapy for treatment of an exacerbation (n=12). (b) a simulated network based on 75% reduction in the abundance of β-lactam-susceptible organisms and calculation of the re-configured network. (c) observed network reconfiguration in patients following β-lactam therapy. Circle size, outline thickness and colour respectively represent node importance based on network metrics; degree, stress centrality and betweenness centrality

our simulated network to that observed among our β-lactam-treated patients following therapy (Figure 1.6c). Our network-based prediction had reliable comparability to the network observed in β-lactam-treated patients with respect to several microbial nodes. Notably, the rank order difference in key microbial taxa post-antibiotic treatment was correctly predicted for 10 out of 13 taxa in our simulated model.

1.3.5 wSNF and Interactome analysis on the validation cohort rediscovers a "high-risk" cluster and validates the interactome.

To assess and validate the previously derived high-risk cluster, we reimplemented the weighted Similarity Network Fusion (wSNF) followed by spectral clustering on the multi-biome derived from a validation cohort (n=166) using a metagenomic sequencing approach, contrary to the targeted sequencing approach used previously. Integrative Microbiomics identified two clusters, with one exhibiting an increased exacerbation phenotype. Thus, validating the high-risk cluster of bronchiectasis patients (Figure 1.7a). Furthermore, Interactome analysis of the high-risk cluster derived from the validation cohort identifies 89.9% of the interactions from the derivation cohort (Figure 1.7b). Hence, validating the interactome signature of the high-risk cluster.

To further assess and validate specific interactions within our derived interactomes, we selected the interaction between P. aeruginosa and A. fumigatus for further interrogation since, it is known that they frequently co-exist in the airway of respiratory disease patients (Cystic Fibrosis and Bronchiectasis). Furthermore, it has been shown that they can inhibit [12, 27] or promote [6, 23] each others growth. These two organisms exhibit opposing interactions from our originally derived clusters (Figure 1.8a): co-exclusion in the low and co-occurrence in the high exacerbation frequency clusters respectively (Figure 1.8a). Comparisons of P. aeruginosa clinical isolates derived from patients belonging to the low and high exacerbation frequency clusters reveal these contrasting interactions (Figure 1.8b). Consistent with the observation from our derived interactomes, the low exacerbation frequency cluster isolate (LEF) exhibited negative interactions with A. fumigatus whereas no such inhibitory effect was observed with the high exacerbation frequency cluster isolate (HEF) (Figure 1.8c). Further analysis of pyoverdine levels reveals a hyperproduction in the HEF isolate exceeding that of the LEF or the PAO1 control. As pyoverdine demonstrates anti-A. fumigatus properties and allows P. aeruginosa to coexist, these findings offer a potential mechanism for our in-vitro observations which most critically are consistent with our in vivo-derived interactomes further validating their accuracy and clinical relevance.

1.3.6 Interactions better predict time to next exacerbation over individual taxa.

To assess, the clinical applicability of the interactome framework we tried to assess the predictability of Time to next exacerbation given the post-exacerbation microbiome by considering individual taxa and interactions (pairs of microbes) as features. A correlation analysis revealed that a greater number of pairwise interactions compared to individual taxa are correlated with Time to next exacerbation (Figure 1.9ab) irrespective of the time point (Baseline, Exacerbation and Post-exacerbation). Evaluation of the accuracy of the Non-linear prediction model fitted to predict Time to next exacerbation reveals that the post-exacerbation microbiome is a better predictive than the baseline and exacerbation microbiomes. Additionally, there is a major increase of predictive capacity (Figure 1.9c) ($0.551 > 0.239$) if interactions are used as features as opposed to individual taxa; suggesting that pairwise interaction of microbes are better markers to study and associate disease outcomes.

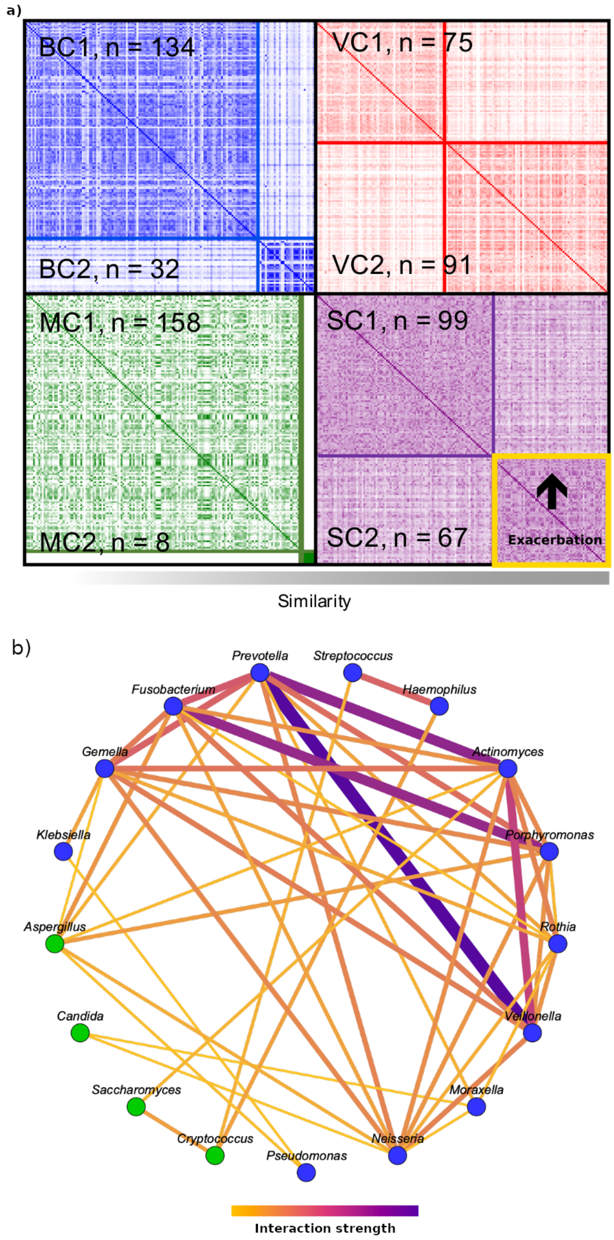


Figure 1.7: (a) Metagenomic integration of microbiomes: A patient similarity matrix illustrating patient clusters derived using spectral clustering of Microbiome (BC), Mycobiome (MC),Virome (VC) and SNF integrated (SC) patient similarity matrices, (b)Interactome signature of the high-risk cluster: an interactome plot with nodes as microbes (common between high-risk cluster of the derivation and validation cohort) and edges as interactions. Node colour indicates bacteria (blue) and fungi (green). Edge width and colour represent the interaction strength.

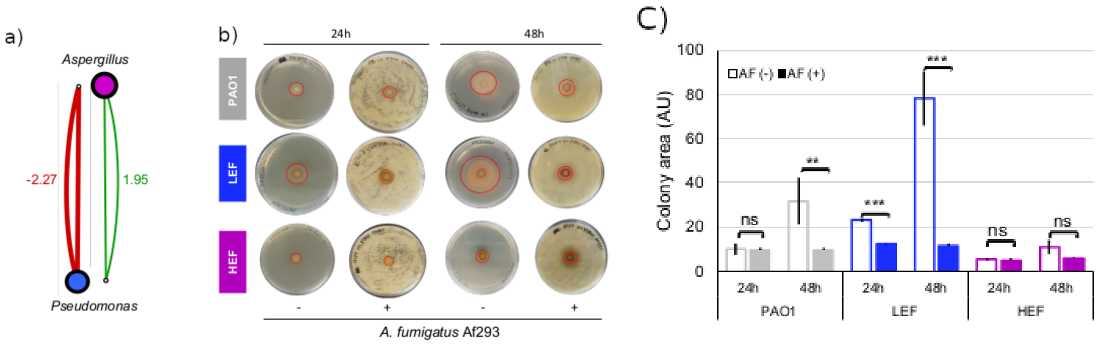
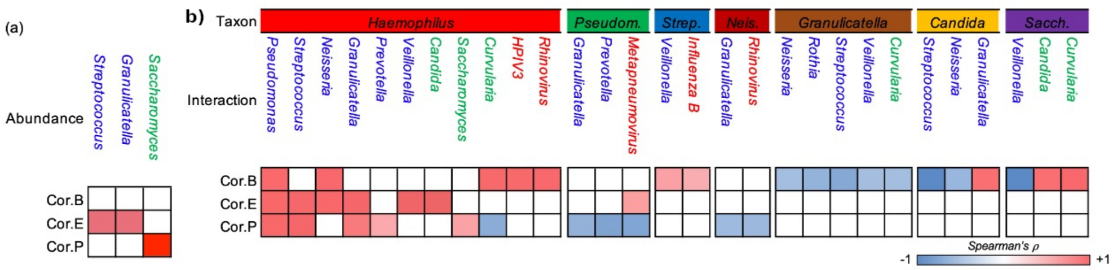


Figure 1.8: (a) Node and edge plots extracted from the LEF and HEF network cluster analysis highlighting opposing interactions between P. aeruginosa and A. fumigatus related to exacerbation frequency. Edges are coloured green or red, reflecting a positive (co-occurrence) or negative (co-exclusion) interaction, respectively. Circle size, outline thickness and colour respectively represent node importance based on network metrics; degree, stress centrality, and betweenness centrality. (b) Demonstration of strain-dependent inter-kingdom interaction between P. aeruginosa and A. fumigatus. Comparison of direct interactions between P. aeruginosa laboratory strain (PAO1; grey) and isolates obtained from patients from the LEF and HEF clusters respectively (LEF; blue, HEF; purple) with A. fumigatus (Af293) by disk inhibition assays. Colony zone diameter is indicated by a red circle for P. aeruginosa strains grown in the presence (+) or absence of (-) Af293 at 24h and 48h timepoints, respectively. (c) Analysis of P. aeruginosa zone diameters observed following co-culture with Af293 following 24h and 48h incubation. Bars are coloured according to the respective P. aeruginosa strain as described above. Open bars indicate zone diameters observed in the absence of A. fumigatus and filled bars indicate zone diameters observed on co-culture. Error bars represent the standard deviation of triplicate determinations. ns: non-significant; **p<0.01; ***p<0.001.



Timepoint		Model	
		Abundance	Interaction
Baseline	Rsq	0.756	0.933
	GRsq	*0.135*	*0.525*
Exacerbation	Rsq	0.368	0.525
	GRsq	*0.113*	*0.333*
Post-exacerbation	Rsq	0.640	0.937
	GRsq	*0.239*	*0.551*

Figure 1.9: A correlogram illustrating the individual taxa (a) and pairwise interactions (b) significantly correlated with time to next exacerbation at various time points: Baseline (Cor.B), Exacerbation (Cor.E) and Post-exacerbation (Cor.P). (c) A table illustrating the evaluation metrics of the MARS, a non-linear regression model when fitted to predict time to next exacerbation using individual taxa abundance and pairwise interaction strength as predictors/features.

1.4 Discussion

In my masters thesis, we presented to the best of our knowledge, the first multi-biome analysis using integrative microbiomics combining bacterial, viral, and fungal communities in individual patients. Using a modified weighted-SNF, we identified frequent exacerbators with high precision and classified microbes within an interactome as busy, influential and/or critical. Frequent exacerbators exhibited antagonistic interactomes. In my present PhD thesis, we extended this further by preforming a longitudinal assessment over an exacerbation. This reveal disrupted interactomes, undetectable by assessing microbial identity alone. By use of simulation followed by confirmatory validation, we demonstrate interactomes clinical relevance for modelling microbiome re-configuration in response to antibiotic exposure. Validation of interactomes was achieved by metagenomics which identifies a cluster that exhibits, a similar high-risk of exacerbation phenotype as identified from the derivation cohort. Further, interactome analysis of the high-risk cluster derived using the metagenomic validation cohort validates 89.9% of the interactions. We also, provide microbiological evidence in support of our interactome approach by demonstrating variable interaction between *P. aeruginosa* and *A. fumigatus* using cluster-specific clinical isolates. We then assessed the clinical applicability of the interactome by modelling time to next exacerbation using interactions and individual taxa as features. Interestingly, we find a major increase in the accuracy of prediction when using interactions, in contrast to individual taxa. Taken together, our findings reveal a novel aspect of the functional microbiome with potential implications for the use of antibiotics in clinical practice.

It is well recognised in bronchiectasis, that patients improve despite receiving antibiotics not necessarily targeting their dominant pathogen. However, the conventional model where targeting bacteria with antibiotics reduces bacterial load, accompanying inflammation and therefore, exacerbation risk, which, in turn, alleviates symptoms and improves clinical outcomes; fails to explain this. If the interactome framework were true, then this could offer explanations of unexplained clinical observations of antibiotic use and help treat exacerbations. Results from this study show that interactions are more predictive than individual taxa of time to next exacerbation and better explain exacerbation, in support of the hypothesis. The airway microbiome (and its accompanying interactome) is likely a critical predictor of antibiotic treatment response and provides a theoretical basis for understanding several phenomena associated with antibiotics that remain unexplained clinically including antimicrobial responses in apparently resistant organisms. Manipulating microbiomes by means other than antibiotics are being explored and the effect of probiotics on the interactome should be considered.

The value of data integration using SNF for multidimensional datasets (such as multi-omics) in airways disease such as COPD has been demonstrated, however, these methods have not been previously applied to microbiome integration [20]. Conventional SNF is not optimized for biological systems such as multi-kingdom microbiomes where dynamism and potential dominance of one kingdom over the others needs to be considered. Employing a weighted SNF approach based on richness, we demonstrate improved patient stratification in bronchiectasis by identifying high frequency exacerbators with accuracy exceeding that of using a single microbial group. Hence to motivate and enable; researchers and clinicians to opt for an integrative strategy when analysing multi biome datasets, we developed a web tool Integrative Microbiomics (`https://integrative-microbiomics.ntu.edu.sg`) capable of implementing both SNF and weighted SNF to integrate microbiome datasets. This webtool also aims to motivate users to obtain multi biome datasets, as integrating datasets would better represent/ bring clarity to the underlying biological process.

Limitations of this work include the cross-section design of the CAMEB cohort, a static dataset which we largely use to predict dynamic interaction [3, 22]. However, this is partially overcome by inclusion of a longitudinal case series to our analysis to better assess temporal dynamics in association to exacerbation and antibiotic treatment. Next, although 16S methodologies are well established, there are inherent limitations, including under-representation of mycobacteria, an important group of organisms in bronchiectasis [29]. Additionally, fungal ITS sequencing approaches are challenged by under-developed reference databases [2]. Our virome analysis, while broad, comprehensive, and informed by established literature, targets a known virus panel and therefore is subject to bias. Nonetheless, this is partially addressed by our metagenomic dataset, which comprehensively assess the virome. While metagenomics potentially represents a less biased alternative approach, it underestimates fungal presence given the significantly higher airway bacterial burden hence obscuring the influence that fungi have on the interactome. We further acknowledge that sputum is an imperfect matrix, and, make no inference about lower airway ecology, noting only the clinical associations between sputum as a surrogate, readily obtainable, non-invasive upper airway sample. Finally, while observational data suggests potential causal association, other factors may drive observed effects. Observed interactions may represent epiphenomena of a selectively operating immune system, for example and our work did not include any assessment of host responses.

1.5 Future works

The developed and validated interactome framework is based on Graph theory, a mathematical theory that study graphs as basic structures to model pairwise relation of nodes as points. Besides, mathematics also provides a generalisation of Graphs through Simplicial complexes from the field of Algebraic topology. A simplicial complex is a mathematical structure that models pairwise relation of simplices (generalisation of nodes), which captures complex relationships as points, lines, triangles and their n-dimensional counterparts. Given that interactions of nodes (microbes) are beneficial than isolated microbes in studying clinical outcomes such as exacerbations and antibiotic action; as my next step, I would like to test the hypothesise that the simplicial complex framework is more beneficial than the interactome framework, to study important clinical outcomes. We have shown that the post-exacerbation (post-antibiotic) interactome is predictive of time to next exacerbation with a GRsq (Generalised R Squared) of 55%, if the proposed hypothesis were true, this could lead to increased accuracy of prediction. Secondly, I also plan on implementing powerful prediction models that are based on machine learning to further increase the accuracy to predict clinical outcomes, advancing the field of precision medicine. Further, I am planning to submit an ERJMethods paper detailing methods such as Similarity Network Fusion (SNF) to integrate datasets.

Chapter 2

Microbial dysregulation of the ‘lung-gut’ axis in high-risk bronchiectasis

2.1 Introduction

The Gut-microbiome is by far the best and widely studied microbial ecosystem of the human anatomy, partly due to the rich microbial environment and partly due to the ease of sample collection (non-invasive) through faeces [7]. On the other hand, healthy lung was long considered to be sterile but with advent high-throughput sequencing techniques this has been proven otherwise [cite]. Extensive research on the gut-microbiota has that shown that gut microbiota is capable of influencing other organs, such as the brain, liver or lungs [5]. This has led to the coining of terms such as the ‘gutbrain axis’ and the ‘lung-gut axis’.

The epithelial surfaces of the gut and lung are exposed to diverse microbes; ingested microorganisms can access both sites and the microbiota from the gut can enter the lungs through processes such as micro-aspiration [7]. Furthermore, the lung and gut can interact thorough the systemic cytokines released by host immune cells in response to microbes or microbes from one-site may secrete metabolites which are absorbed into the blood stream and thus regulate the organs [9]. A study used germ free mice, which lack an appropriately developed immune system and showed mucosal alterations, both of which is restored through colonization with gut microbiota. Thus, supplementing the concept of ‘lung-gut’ axis [7].

Literature survey using the keyword ‘lung-gut axis’ shows that this concept was first introduced in 2004 and increasing work is being done 2.1. This increasing evidence also suggests, a potential existence of lung-gut axis and its effectual role in lung diseases. Although the gutlung axis is only beginning to be understood, emerging evidence indicates that there is potential for manipulation of the gut microbiota in the treatment of lung diseases. Despite this, the influence of microbial gut health in Bronchiectasis lung is poorly studied. Hence, in the second chapter of my PhD thesis

Want to check if microbes interact

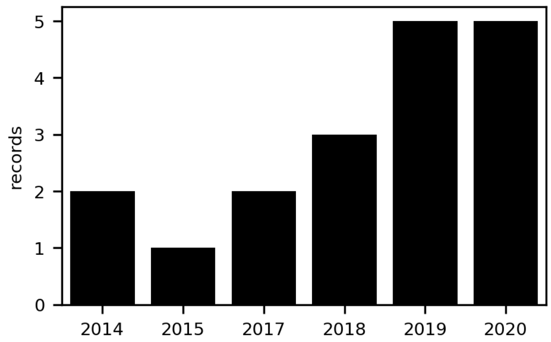


Figure 2.1: A histogram illustrating all available publications (including original articles and perspectives) matching the keyword “lung-gut axis” from 1900 to 2020 in the web of science database.

2.2 Methods

2.2.1 Study population

57 patients with stable bronchiectasis were recruited as a part of this study by our collaborators in Milan, Italy. Recruitment of patients was performed cross-sectionally as a part of the Bronchiectasis Program of Fondazione IRCCS Ca’ Granda Ospedale Maggiore Policlinico, Milan, Italy. Patients were enrolled during their clinical stability (at least one month apart from the last exacerbation and antibiotic course) and underwent clinical, radiological and microbiological evaluation. Patients were asked to provide a sputum and stool sample with a maximum gap of 12hrs. between the sputum and stool sample. DNA was extracted by our lab members from the sputum and stool samples as described previously [22]. The extracted DNA was subjected to targeted amplicon sequencing of the 16S rRNA and ITS2 regions of the genome to derive the Microbiome and Mycobiome, by mapping them to green genes and UNITE databases, respectively.

2.2.2 Data-preprocessing and Statistical analysis

Only microbes present $\geq$1% in at-least 5 patients were considered for all analyses. Read counts of Microbiome and Mycobiome datasets from the sputum and stool samples of the 57 samples were converted into relative abundances for below analysis. Intersection analysis was performed on the columns of the microbiomes (bacteria and fungi from lung and gut) using the ‘intersection’ function in R. Diversity of the individual microbiomes was measured using the Shannon-diversity index computed in R using the ’vegan’ package. Differences between the median diversity between the sites were assessed using the Maan-Whitney U test. A p-value < 0.05 was considered statistically significant.

2.2.3 Co-occurrence analysis

Sequence analysis captures microbial composition on a relative scale, rendering microbiome datasets compositional and sparse. Hence, an absolute increase in the relative abundance of one species is accompanied by a compositional decrease in another, causing the problem of spurious correlations [1]. To address this, Faust et al. developed a novel bootstrap and renormalisation (reboot) approach that mitigates these potential issues by calculating statistical significance thresholds that accounts for similarity due to pure compositionality [10]. Microbial association networks using GBLM in confluence with reboot was implemented as described previously in [26]. This linear method only captures complex linear interactions between the microbes with the assumption that all the microbes can interact with each other.

Sparse Inverse Covariance for Ecological Statistical Inference (spiec-easi) was implemented to estimate microbial association networks from the precision matrix within and between the four compositional microbial read-counts datasets (bacteria-lung, fungi-lung, bacteria-gut and fungi-gut), under the joint sparsity penalty [19]. Speic-easi was implemented using the 'multi.spiec.easi' function of the 'SpiecEasi' package in R with the following parameters: method = glasso, lamda.min.ratio=1e-4, nlambda=200, re.number=100. The resulting networks was appended with edge weights calculated by scaling the inverse covariance matrix of the optimal network into a correlation matrix.

2.2.4 Integrative analysis

Multi Omic Factor Analysis (MOFA) was implemented to perform an unsupervised factor analysis on the integrated multibiome datasets. Broadly, MOFA , tries to infer an interpretable low-dimensional representation of the multibiome datasets in terms of latent factors [4]. Microbial read-counts of the multibiome datasets were centered log ratio (clr) transformed after addition of 0.1 to the read-counts followed by concatenation. This additions is necessary for computational stability, convergence and doesn't affect the overall value. MOFA was implemented on this transformed and concatenated dataset, using the 'MOFA2' package in R and the following parameters: num_factors = 3 and convergence_mode=medium. The explained variance of each of these were calculated and plotted. Further, these factors were evaluated against clinical attributes to check for statistical differences between the median of the factor values between the groups defined by the clinical attributes. A non-parametric, dunn's test with Benjamin-Hochberg FDR correction OR a Maan-Whitney U test was implemented to assess the statistical significance between multiple groups, whenever appropriate. A p-value < 0.05 was considered statistically significant.

Weighted SNF analysis as developed and described, in Chapter1 and [26] was implemented with k-nearest neighbours = 9 and weights equal to the taxonomical richness of the individual microbiomes on the multibiome datasets (bacteria-lung, bacteria-gut, fungi-lung and fungi-gut). Following this integration, spectral clustering was implemented with optimum number of clusters=2 to cluster the patients based on the integrated microbiomes. This analysis was performed using the web-tool developed in the previous chapter and available at `integrative-microbiomics.ntu.edu.sg`. The robustness of our identified clusters was assessed using a bootstrapping approach with 70% of the integrated data being sampled over 100 bootstrap iterations followed by spectral clustering with k (number of clusters) = 2 on this 70% bootstrap sample. The resulting bootstrap clusters (subsamples data, 70%) were compared with the original clusters (100%). A differential abundance analysis to identify discriminant taxa between the derived clusters was implemented using ALDEX2 [11]. ALDEX2, uses a Dirichlet-multinomial model to infer abundance and sample variation from read-counts and calculates the expected false discovery rate given the biological and sampling variation using the given test. ALDEX2 was implemented using test = Welches t-test and the fdr corrected p-values calculated by the Benjamini Hochberg correction was computed. P-value < 0.05 was considered statistically significant and the corresponding taxa as a discriminant taxa. This was implemented in R using the ALDEX2 package with default parameters.

2.3 Results

2.3.1 Significant overlap of fungal communities of lung and gut contrary to bacteria

Intersection analysis of bacterial and fungal communities between sputum and stool samples reveals increased overlap of fungal communities between the lung and gut, contrary to bacteria. Three bacterial genera including *Lactobacillus, Prevotella* and *Streptococcus* compared to six fungal genera including *Candida, Cryptococcus, Curvularia, Debaryomyces, Lodderomyces* and *Saccharomyces* were present in both sputum and stool samples. Interestingly, upon assessment of diversity between the sputum and stool samples, a similar pattern is observed. Overall, mycobiome exhibits a decreased diversity compared to microbiome. Further, an increased diversity of bacteria is found in the gut compared to the lung whereas the fungal diversity doesn't change.

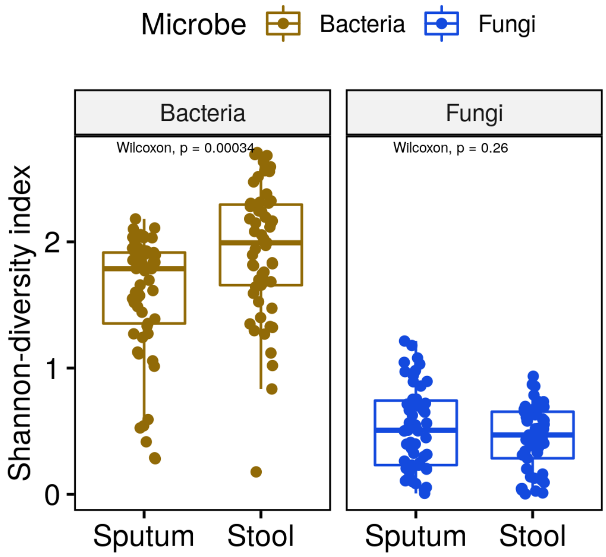


Figure 2.2: A boxplot illustrating the difference in Shannon-diversity index of the Microbiome (Ochre) and Mycobiome (Blue) between the sputum (Lung) and stool (Gut) samples. Statistical significance of these differences were calculated using 'wilcoxon test' and are indicated above as p-values.

2.3.2 Co-occurrence analysis reveals lung gut microbial (bacteria and fungi) interactions suggestive of a potential lung-gut axis.

Assessment of interactions of microbes (bacteria and fungi) between the lung and gut was construed using co-occurrence analysis. Co-occurrence analysis was performed using two methods 1)GBLM: captures complex linear interactions and 2) Spiec-easi: captures non-linear interactions but assumes sparsity, to derive microbial association networks. Microbial associations networks from both these methods shows cross-talk between the microbes of the lung and the gut, indicative of potential existence of the lung-gut axis [Figure2.3]. However, a greater number of inter-axis interactions is observed through the use of GBLM methods as compared to spiec-easi and this is partially due to the assumption of sparsity by spiec-easi. Integrative assessment of the microbiomes was performed using MOFA2. Roughly, MOFA2 works like a PCA to create factors which maximises the explained variance of the integrated microbiome. MOFA2 analysis on the microbiomes (bacteria lung, bacteria gut, fungi lung and fungi gut) reveals a factor (Factor1) associated with exacerbations and Non-tuberculosis mycobacterial(NTM) infections [Figure2.4(c,d)]. Following, assessment of Factor1 in terms of contribution from individual microbiomes reveal gut microbiome as the highest contributor [Figure2.4(b)], possibly explainable by lung-gut axis which further supplements the existence of lung-gut axis. Moreover, upon assessment of the loadings of Factor1 we find four of the six overlapping microbes including *Streptococcus, Saccharomyces, Candida* and *Curvularia* have loadings $\geq 0.5\%$.

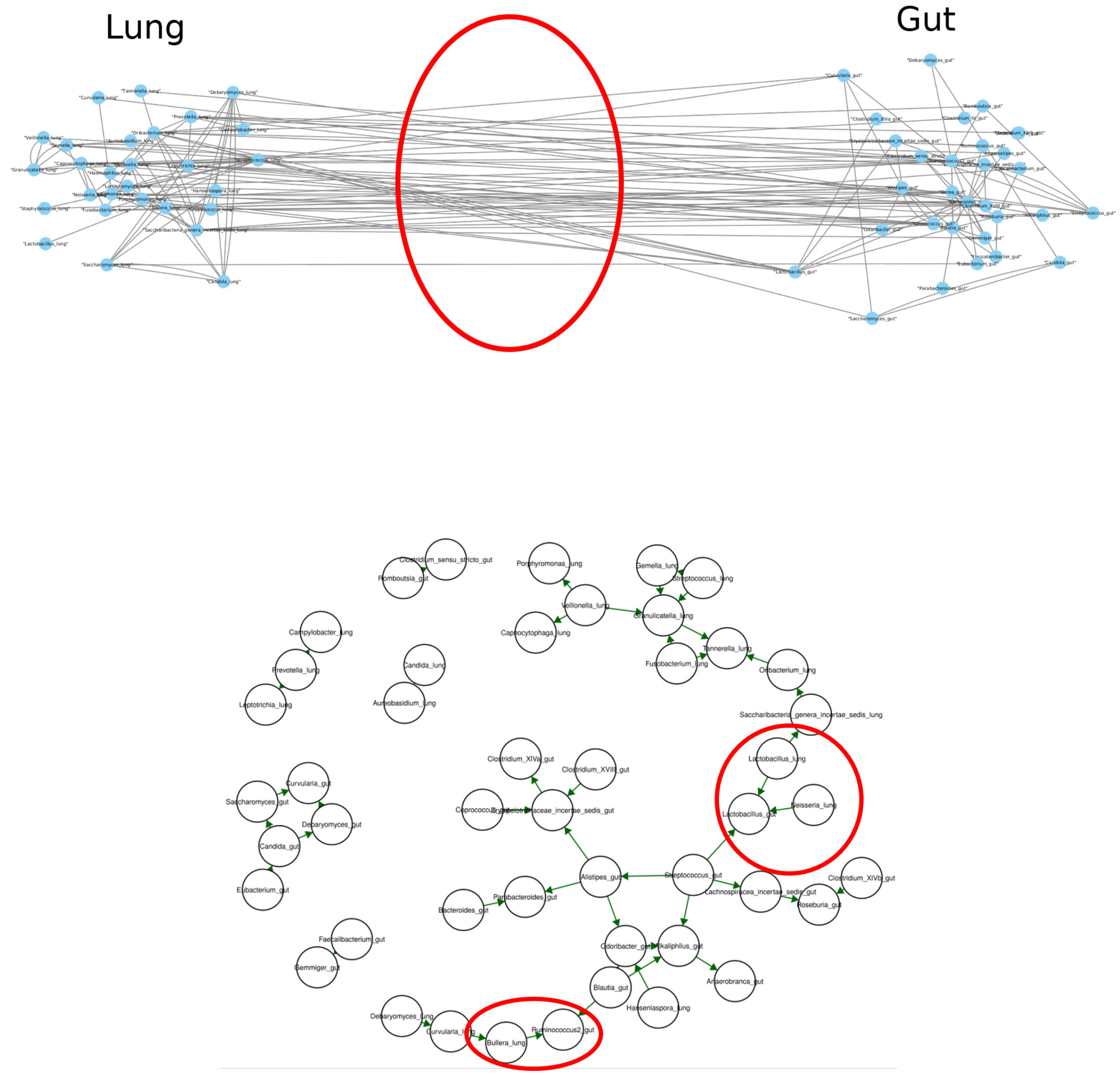
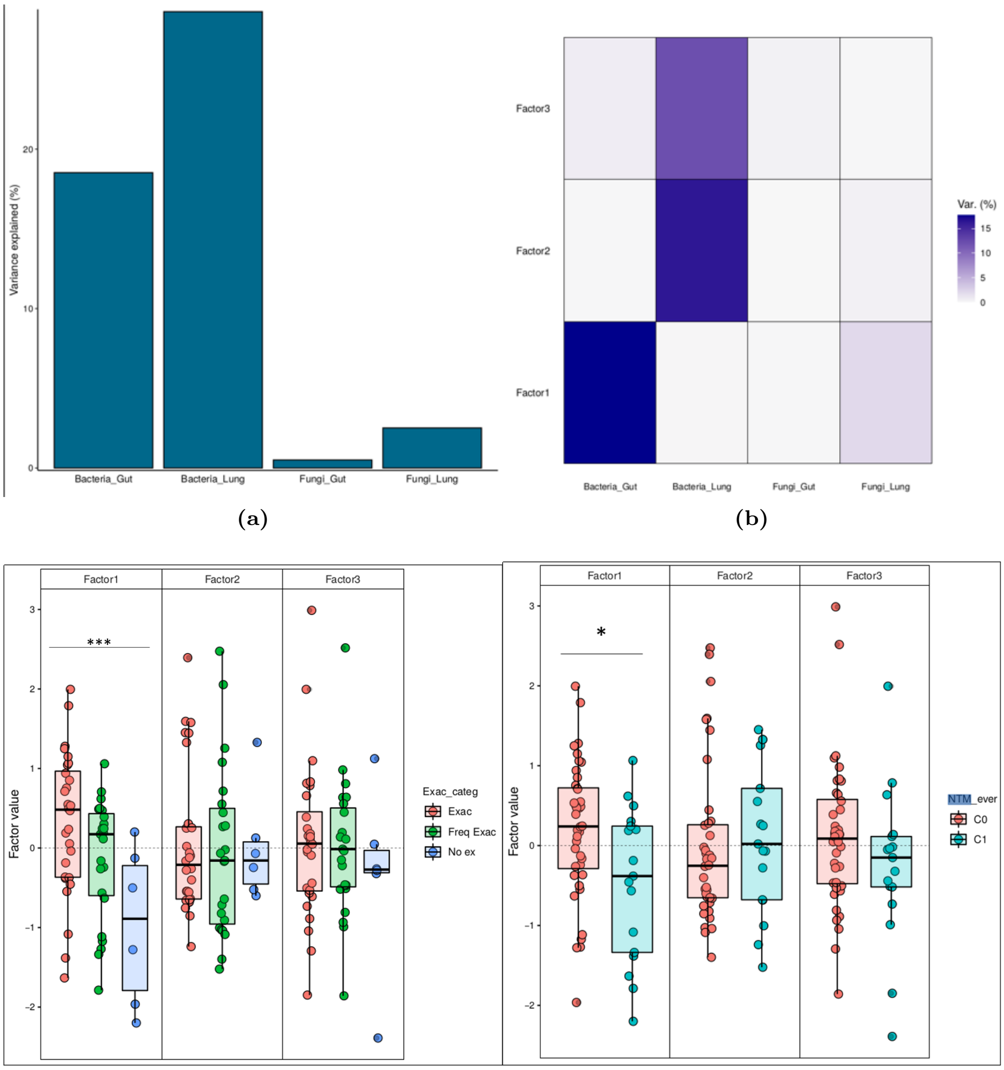


Figure 2.3: Microbial association networks derived using co-occurrence analysis methods 1) GBLM(top) and 2)Spiec-easi(bottom). Nodes represents microbes including bacteria and fungi from both lung(left) and gut(right). Edges illustrate the association/interactions between the microbes derived using the respective methods. Highlighted red circle represents the interactions between the lung and gut microbiome.

2.3.3 Integrated microbiomes identifies a 'high-risk' patient cluster

Integration of microbiomes and mycobiomes from lung and gut using weighted SNF with k=9 was performed, following which spectral clustering was implemented to cluster the patients into two groups. Li *et.al.* in their paper showed that integrating multiple views of the same patient/sample increases the power and compensates for smaller sample sizes[20]. Hence, clustering of the integrated microbiomes of these (n=57) patients is admissible. Cluster robustness was evaluated using a bootstrap approach and was found to be 79.15%. Evaluation of the derived clusters across clinical attributes reveals patients belonging to cluster1



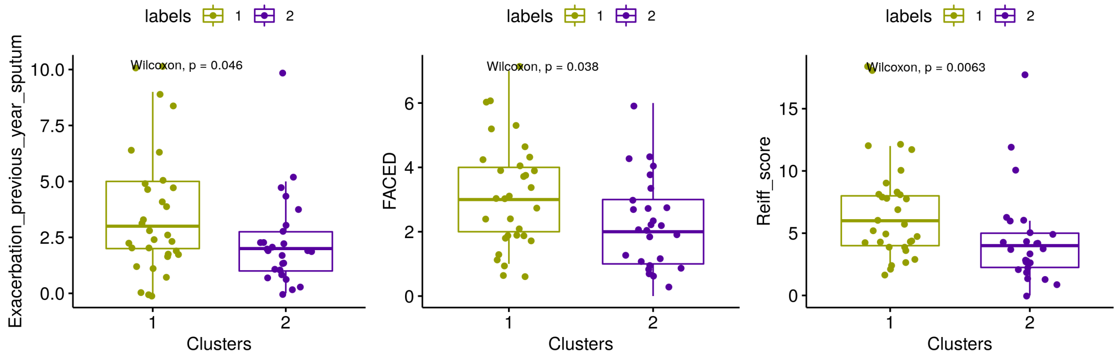
(a) (b) (c) (d)

Figure 2.4: Integrative microbiome data analysis using MOFA2: (a) A bar chart representing the cumulative variance explained by each of the individual integrated biomes. (b) A heat-map illustrating the breakdown for variance explained of the individual biomes across the first three factors. Box plots illustrating the factor values of factors 1,2 and 3 across exacerbation category (c) and NTM_ever groups (d).'*' illustrate the statistical significance of kruskal-wallis test in terms p-values; '*' p-value < 0.05, '**' p-value < 0.001,'***' p-value < 0.0001

have a higher median risk of exacerbation, FACED score and Reiff score [Figure2.5] compared to cluster2. Differential analysis illustrates significantly increased *Candia* in Gut and *Fusobacterium* in Lung, of high-risk patients (cluster1) compared to that of low-risk patients(cluster2).

2.3.4 Dysregulated lung-gut axis in high-risk patients

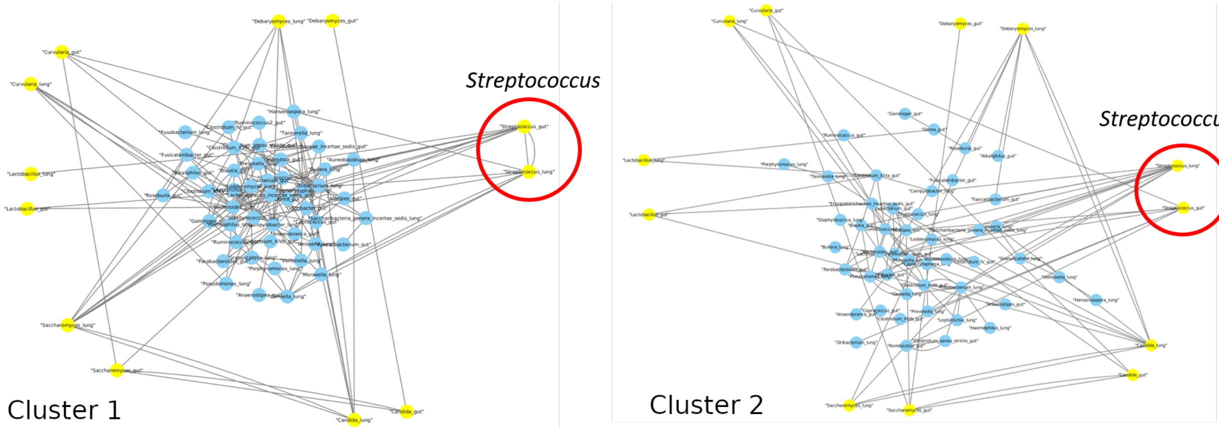
Having shown the existence of lung-gut axis and identifying a sub-group of high-risk bronchiectasis patients using the integrative microbiomics. We next evaluated the changes in lung-gut axis across the two clusters. Microbes between the lung and gut can broadly interact in two ways: 1) Microbes can travel between the sites through mechanisms such as micro-aspiration and 2) Microbes can secrete secondary metabolites and other biomolecules through which they can interact. To assess the first mechanism, a linear correlation analysis between the clr transformed abundance of the overlapping microbes from the lung and gut was performed. *Lactobacillus* lung was found to be significantly correlated with *Lactobacillus* gut in high-risk cluster 1 but not in cluster 2. However, the observed correlation of *Lactobacillus* between the two sites



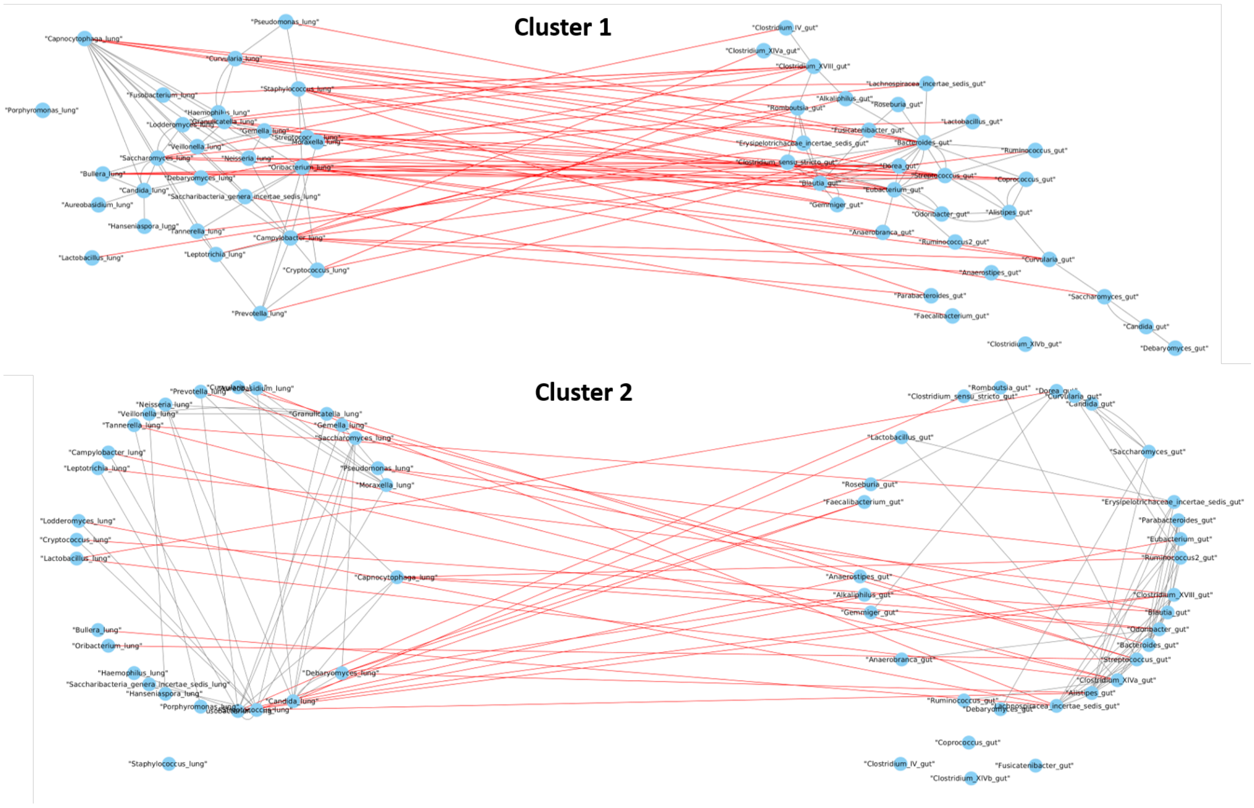
(a) **(b)** **(c)**

Figure 2.5: Boxplots illustrating the differences in Exacerbation (a), FACED (b) and Reiff score (c) between high-risk cluster 1(yellow) and low-risk cluster 2(blue). Statistical significance of these differences were calculated using wilcoxon test and are indicated above as p-values.

may be due to confounding microbes. Therefore, a GBLM analysis was performed to assess the association of overlapping microbes between the two sites given all the other microbes. Interestingly, the correlation between *Lactobacillus* lung and *Lactobacillus* gut disappears. However, a new association between *Streptococcus* lung and *Streptococcus* gut is found in high-risk cluster 1 but not in low-risk patients; suggestive of movement of *Streptococcus* between lung and gut in high-risk patients (probably due to dysregulation of the axis) [Figure2.6(a)]. Assessment of the second mechanism was performed by evaluating the inter-axis microbial interactions between clusters. This reveals an increased lung-gut microbial interaction in high-risk cluster(35%) compared to low-risk cluster(29%) [Figure2.6(b)]; further supplementing the dysregulation of this axis in high-risk patients.



(a)



(b)

Figure 2.6: Microbial co-occurrence network across the clusters, derived using GBLM with nodes as microbes (bacteria and fungi) from both lung and gut, and edges representing the significant (p-value<0.0001) interaction between nodes. (a) Overlapping microbes are highlighted as yellow nodes. (b) Inter lung-gut microbial interactions are highlighted as red edges.

2.4 Discussion

2.5 Future works

Bibliography

[1] Aitchison, J. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological) 44*, 2 (1982), 139–160.

[2] Ali, Aogáin, M., Morales, Tiew, and Chotirmall. Optimisation and benchmarking of targeted amplicon sequencing for mycobiome analysis of respiratory specimens. *International Journal of Molecular Sciences 20*, 20 (oct 2019), 4991.

[3] Aogáin, M. M., Tiew, P. Y., Lim, A. Y. H., Low, T. B., Tan, G. L., Hassan, T., Ong, T. H., Pang, S. L., Lee, Z. Y., Gwee, X. W., Martinus, C., Sio, Y. Y., Matta, S. A., Ong, T. C., Tiong, Y. S., Wong, K. N., Narayanan, S., Au, V. B., Marlier, D., Keir, H. R., Tee, A., Abisheganaden, J. A., Koh, M. S., Wang, D. Y., Connolly, J. E., Chew, F. T., Chalmers, J. D., and Chotirmall, S. H. Distinct "immunoallertypes" of disease and high frequencies of sensitization in non–cystic fibrosis bronchiectasis. *American Journal of Respiratory and Critical Care Medicine 199*, 7 (apr 2019), 842–853.

[4] Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J. C., and Stegle, O. Mofa+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology 21*, 1 (2020), 1–17.

[5] Bell, J., Spencer, J., Yates, R., Yee, S., Jacobs, B., and DeLuca, G. Invited review: From nose to gut–the role of the microbiome in neurological disease. *Neuropathology and applied neurobiology 45*, 3 (2019), 195–215.

[6] Briard, B., Heddergott, C., and Latgé, J.-P. Volatile compounds emitted by pseudomonas aeruginosa stimulate growth of the fungal pathogen aspergillus fumigatus. *mBio 7*, 2 (mar 2016).

[7] Budden, K. F., Gellatly, S. L., Wood, D. L., Cooper, M. A., Morrison, M., Hugenholtz, P., and Hansbro, P. M. Emerging pathogenic links between microbiota and the gut–lung axis. *Nature Reviews Microbiology 15*, 1 (2017), 55–63.

[8] Chalmers, J. D., and Chotirmall, S. H. Bronchiectasis: new therapies and new perspectives. *Lancet Respir Med 6*, 9 (Sep 2018), 715–726.

[9] Dang, A. T., and Marsland, B. J. Microbes, metabolites, and the gut–lung axis. *Mucosal Immunology 12*, 4 (2019), 843–850.

[10] Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., and Huttenhower, C. Microbial co-occurrence relationships in the human microbiome. *PLoS computational biology 8*, 7 (2012), e1002606.

[11] Fernandes, A. D., Reid, J. N., Macklaim, J. M., McMurrough, T. A., Edgell, D. R., and Gloor, G. B. Unifying the analysis of high-throughput sequencing datasets: characterizing rna-seq, 16s rrna gene sequencing and selective growth experiments by compositional data analysis. *Microbiome 2*, 1 (2014), 15.

[12] Ferreira, J. A. G., Penner, J. C., Moss, R. B., Haagensen, J. A. J., Clemons, K. V., Spormann, A. M., Nazik, H., Cohen, K., Banaei, N., Carolino, E., and Stevens, D. A. Inhibition of aspergillus fumigatus and its biofilm by pseudomonas aeruginosa is dependent on the source, phenotype and growth conditions of the bacterium. *PLOS ONE 10*, 8 (aug 2015), e0134692.

[13] Friedman, J. H. Multivariate adaptive regression splines. *The Annals of Statistics 19*, 1 (mar 1991), 1–67.

[14] Gusareva, E. S., Acerbi, E., Lau, K. J. X., Luhung, I., Premkrishnan, B. N. V., Kolundžija, S., Purbojati, R. W., Wong, A., Houghton, J. N. I., Miller, D., Gaultier, N. E., Heinle, C. E., Clare, M. E., Vettath, V. K., Kee, C., Lim, S. B. Y., Chénard, C., Phung, W. J., Kushwaha, K. K., Nee, A. P., Putra, A., Panicker, D., Yanqing, K., Hwee, Y. Z., Lohar, S. R., Kuwata, M., Kim, H. L., Yang, L., Uchida, A., Drautz-Moses, D. I., Junqueira, A. C. M., and Schuster, S. C. Microbial communities in the tropical air ecosystem follow a precise diel cycle. *Proceedings of the National Academy of Sciences 116*, 46 (oct 2019), 23299–23308.

[15] Homa, M., Sándor, A., Tóth, E., Szebenyi, C., Nagy, G., Vágvlgyi, C., and Papp, T. In vitro interactions of pseudomonas aeruginosa with scedosporium species frequently associated with cystic fibrosis. *Frontiers in Microbiology 10* (mar 2019).

[16] Jiang, D., Armour, C. R., Hu, C., Mei, M., Tian, C., Sharpton, T. J., and Jiang, Y. Microbiome multi-omics network analysis: Statistical considerations, limitations, and opportunities. *Frontiers in Genetics 10* (nov 2019).

[17] Knight, R., Callewaert, C., Marotz, C., Hyde, E. R., Debelius, J. W., McDonald, D., and Sogin, M. L. The microbiome and human biology. *Annual Review of Genomics and Human Genetics 18*, 1 (aug 2017), 65–86.

[18] Kuijjer, M. L., Hsieh, P.-H., Quackenbush, J., and Glass, K. lionessR: single sample network inference in r. *BMC Cancer 19*, 1 (oct 2019).

[19] Kurtz, Z. D., Mller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., and Bonneau, R. A. Sparse and compositionally robust inference of microbial ecological networks. *PLOS Computational Biology 11*, 5 (may 2015), e1004226.

[20] Li, C.-X., Wheelock, C. E., Skld, C. M., and Wheelock, Å. M. Integration of multi-omics datasets enables molecular classification of COPD. *European Respiratory Journal 51*, 5 (mar 2018), 1701930.

[21] Li, Y., Wang, K., Zhang, B., Tu, Q., Yao, Y., Cui, B., Ren, B., He, J., Shen, X., Nostrand, J. D. V., Zhou, J., Shi, W., Xiao, L., Lu, C., and Zhou, X. Salivary mycobiome dysbiosis and its potential impact on bacteriome shifts and host immunity in oral lichen planus. *International Journal of Oral Science 11*, 2 (jun 2019).

[22] Mac Aogáin, M., Chandrasekaran, R., Lim Yick Hou, A., Teck Boon, L., Liang Tan, G., Hassan, T., Thun How, O., Hui Qi Ng, A., Bertrand, D., Yu Koh, J., Lei Pang, S., Yang Lee, Z., Wei Gwee, X., Martinus, C., Yie Sio, Y., Anusha Matta, S., Tim Chew, F., Keir, H. R., Connolly, J. E., Arputhan Abisheganaden, J., Siyue Koh, M., Nagarajan, N., Chalmers, J. D., and Chotirmall, S. H. Immunological corollary of the pulmonary mycobiome in bronchiectasis: The cameb study. *European Respiratory Journal* (2018).

[23] Margalit, A., Carolan, J. C., Sheehan, D., and Kavanagh, K. The aspergillus fumigatus secretome alters the proteome of pseudomonas aeruginosa to stimulate bacterial growth: Implications for co-infection. *Molecular & Cellular Proteomics 19*, 8 (may 2020), 1346–1359.

[24] Menzel, P., Ng, K. L., and Krogh, A. Fast and sensitive taxonomic classification for metagenomics with kaiju. *Nature Communications 7*, 1 (apr 2016).

[25] Morgan, X. C., and Huttenhower, C. Human microbiome analysis. *PLoS Comput Biol 8*, 12 (2012), e1002808.

[26] Narayana, J. K. Investigating the respiratory microbiome in bronchiectasis through "integrative micro- biomics". Master's thesis, Indian Institute of Science Education and Research, Pune, 2019.

[27] Reece, E., Doyle, S., Greally, P., Renwick, J., and McClean, S. Aspergillus fumigatus inhibits pseudomonas aeruginosa in co-culture: Implications of a mutually antagonistic relationship on virulence and inflammation in the CF airway. *Frontiers in Microbiology 9* (jun 2018).

[28] Seitz, A. E., Olivier, K. N., Adjemian, J., Holland, S. M., and Prevots, D. R. Trends in bronchiectasis among medicare beneficiaries in the united states, 2000 to 2007. *CHEST 142*, 2 (Aug 2012), 432–439.

[29] Sulaiman, I., Wu, B. G., Li, Y., Scott, A. S., Malecha, P., Scaglione, B., Wang, J., Basavaraj, A., Chung, S., Bantis, K., Carpenito, J., Clemente, J. C., Shen, N., Bessich, J., Rafeq, S., Michaud, G., Donington, J., Naidoo, C., Theron, G., Schattner, G., Garofano, S., Condos, R., Kamelhar, D., Addrizzo-Harris, D., and Segal, L. N. Evaluation of the airway microbiome in nontuberculous mycobacteria disease. *European Respiratory Journal 52*, 4 (aug 2018), 1800810.

[30] Wagg, C., Schlaeppi, K., Banerjee, S., Kuramae, E. E., and van der Heijden, M. G. A. Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nature Communications 10*, 1 (oct 2019).

[31] Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B., and Goldenberg, A. Similarity network fusion for aggregating data types on a genomic scale. *Nature Methods 11* (Jan 2014), 333 EP –. Article.