



# Integrative microbiomics in bronchiectasis exacerbations

Micheál Mac Aogáin<sup>1,12</sup>, Jayanth Kumar Narayana<sup>1,12</sup>, Pei Yee Tiew<sup>1,2</sup>, Nur A'tikah Binte Mohamed Ali<sup>1</sup>, Valerie Fei Lee Yong<sup>1</sup>, Tavleen Kaur Jaggi<sup>1</sup>, Albert Yick Hou Lim<sup>3</sup>, Holly R. Keir<sup>4</sup>, Alison J. Dicker<sup>4</sup>, Kai Xian Thng<sup>1</sup>, Akina Tsang<sup>1</sup>, Fransiskus Xaverius Ivan<sup>1</sup>, Mau Ern Poh<sup>5</sup>, Martina Oriano<sup>6,7</sup>, Stefano Aliberti<sup>6,7</sup>, Francesco Blasi<sup>6,7</sup>, Teck Boon Low<sup>8</sup>, Thun How Ong<sup>2</sup>, Brian Oliver<sup>9,10</sup>, Yan Hui Giam<sup>4</sup>, Augustine Tee<sup>8</sup>, Mariko Siyue Koh<sup>2</sup>, John Arputhan Abisheganaden<sup>3</sup>, Krasimira Tsaneva-Atanasova<sup>11</sup>, James D. Chalmers<sup>4</sup> and Sanjay H. Chotirmall<sup>1</sup>✉

**Bronchiectasis, a progressive chronic airway disease, is characterized by microbial colonization and infection. We present an approach to the multi-biome that integrates bacterial, viral and fungal communities in bronchiectasis through weighted similarity network fusion (<https://integrative-microbiomics.ntu.edu.sg>). Patients at greatest risk of exacerbation have less complex microbial co-occurrence networks, reduced diversity and a higher degree of antagonistic interactions in their airway microbiome. Furthermore, longitudinal interactome dynamics reveals microbial antagonism during exacerbation, which resolves following treatment in an otherwise stable multi-biome. Assessment of the *Pseudomonas* interactome shows that interaction networks, rather than abundance alone, are associated with exacerbation risk, and that incorporation of microbial interaction data improves clinical prediction models. Shotgun metagenomic sequencing of an independent cohort validated the multi-biome interactions detected in targeted analysis and confirmed the association with exacerbation. Integrative microbiomics captures microbial interactions to determine exacerbation risk, which cannot be appreciated by the study of a single microbial group. Antibiotic strategies probably target the interaction networks rather than individual microbes, providing a fresh approach to the understanding of respiratory infection.**

The global burden of bronchiectasis is increasing, and there remains a lack of proven treatment options due to disease heterogeneity. This is, however, being addressed through endophenotyping efforts and the identification of treatable traits<sup>1–4</sup>. Recurrent infection and inflammation result in progressive, irreversible airway dilatation characterized by an altered airway microbiome<sup>2,5,6</sup>. Bacterial, viral and fungal communities in bronchiectasis have been found to be associated with clinical outcomes, including exacerbations<sup>7–9</sup>. Although specific pathogens are implicated in bronchiectasis exacerbations, prior bacterial microbiome studies show minimal actual change during exacerbations-based analyses of dominant taxa or dissimilarity metrics, thereby demonstrating our incomplete understanding of the microbiome's role<sup>10</sup>. Exacerbation occurrence and frequency in bronchiectasis and other respiratory diseases remain a major cause of morbidity and a key driver of mortality. The precise microbial relationships underpinning exacerbations are complex, however, the general consensus is the simplistic model in which single-kingdom bacterial overgrowth causes infection, which is then suppressed by antibiotics<sup>5,10,11</sup>. Disease heterogeneity in bronchiectasis has hindered success in clinical trials, but, given the variability in clinical, immunological and

inflammatory phenotypes, etiologies and therapeutic responses, patient stratification based on the microbiome could provide focused and precision-based therapy<sup>7,12,13</sup>. Microbiome studies to date have considered bacteria, viruses and fungi as separate entities, but the true microbiome consists of all microorganisms and their genes, including bacteria, viruses and fungi. Prior studies are therefore incomplete, and a greater understanding of disease is likely to be gained through holistic and integrated so-called multi-biome analysis that more accurately represents the in vivo state. Here we perform an integrated multi-biome analysis of the bronchiectasis airway combining bacterial, viral and fungal community profiles from individual patients, as well as longitudinal assessment during exacerbations. We demonstrate that integrative microbiomics provides a novel framework for understanding exacerbations and has potential applications across the spectrum of respiratory disease.

## Results

**Multi-biome data integration by weighted similarity network fusion.** To evaluate the bronchiectasis microbiome, we assessed respiratory specimens from 217 patients. These specimens captured bacterial, fungal and viral taxa (three datasets per patient; 651

<sup>1</sup>Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. <sup>2</sup>Department of Respiratory and Critical Care Medicine, Singapore General Hospital, Singapore, Singapore. <sup>3</sup>Department of Respiratory and Critical Care Medicine, Tan Tock Seng Hospital, Singapore, Singapore. <sup>4</sup>School of Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK. <sup>5</sup>Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. <sup>6</sup>Respiratory Unit and Cystic Fibrosis Adult Center, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. <sup>7</sup>Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy. <sup>8</sup>Department of Respiratory and Critical Care Medicine, Changi General Hospital, Singapore, Singapore. <sup>9</sup>Woolcock Institute of Medical Research, University of Sydney, Sydney, New South Wales, Australia. <sup>10</sup>School of Life Sciences, University of Technology Sydney, Sydney, New South Wales, Australia. <sup>11</sup>Department of Mathematics and Living Systems Institute, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK. <sup>12</sup>These authors contributed equally: Micheál Mac Aogáin, Jayanth Kumar Narayana. ✉e-mail: [schotirmall@ntu.edu.sg](mailto:schotirmall@ntu.edu.sg)

biomes in total). Patients were recruited as part of the cross-sectional CAMEB (Cohort of Asian and matched European bronchiectasis) study<sup>8</sup>. Patients had a median age of 68 years (range, 60–74 years) and were equally distributed according to sex (Supplementary Table 1). Most had idiopathic or post-infectious (non-mycobacterial) bronchiectasis and were classified as having moderate to severe disease using a validated scoring system (median bronchiectasis severity index (BSI), 9; interquartile range (IQR), 6–13). Patients were recruited in Asia (Singapore and Kuala Lumpur, Malaysia) and Europe (Dundee, Scotland, UK). For inclusion, patients had confirmed radiological bronchiectasis on high-resolution computed tomography (HRCT) and were recruited during outpatient attendance when clinically stable.

Having previously characterized the fungal mycobiome in the CAMEB cohort<sup>8</sup>, we generated accompanying bacterial and viral microbiome profiles for all of the patients to assess a more holistic microbiome in each individual (Extended Data Fig. 1). Individual analysis of the bacterial microbiome revealed three patient clusters (Extended Data Fig. 1a), characterized by predominance of *Pseudomonas*, *Streptococcus* or *Haemophilus* (Extended Data Fig. 1b). Of the three clusters, the *Pseudomonas*- and *Haemophilus*-dominant clusters had lower  $\alpha$ -diversity and a more positive correlation with clinical features such as exacerbation (Extended Data Fig. 1c–g), which was most apparent in the *Pseudomonas*-dominant cluster. A 17-virus panel targeting respiratory viruses was also assessed in all of the patients, which showed a significantly higher airway viral load in bronchiectasis than in the healthy airway (Extended Data Fig. 1h and Supplementary Table 2). Human parainfluenza virus 3 was most prevalent (57%), with higher occurrence in patients of Asian than European origin (79% versus 39%,  $P < 0.001$ ). Other viruses detected at lower frequency include rhinovirus, bocavirus, human parainfluenza virus 2, human parainfluenza virus 4, influenza A, influenza B, respiratory syncytial virus A, metapneumovirus, human coronavirus 229E and enterovirus (Extended Data Fig. 1i), none of which was associated with any clinical outcomes (data not shown).

Bacterial and viral community profiles were integrated with existing mycobiome data<sup>8</sup> by implementing a weighted similarity network fusion (WSNF) approach, which assumes differential influences of each biome on the overall multi-biome based on individual taxonomic composition and richness (Fig. 1a,b and Extended Data Fig. 2). Weighting was assigned according to the total number of observed taxa present in a particular biome, with filtering based on a prevalence of at least 5% across the patient cohort; that is, bacteriome (62 genera) > mycobiome (52 genera) > virome (four viral species) observed across 217 patients. Based on these observationally stable inter-kingdom taxa ( $n = 118$ ), assigned weights of 53% (62 of 118) for bacteria, 44% for fungi (52 of 118) and 3% for viruses (4 of 118) were applied in the network fusion, consistent with the breadth of information content underlying each network (Fig. 1b). Spectral clustering of the resultant similarity matrix identified two patient clusters (Fig. 1c), with a mean misclassification ratio (based on 100 bootstrap iterations) of 12.4%, indicating a cluster robustness of 87.6%. The WSNF method used here is freely available as an online webtool (<https://integrative-microbiomics.ntu.edu.sg>; Methods). Each cluster contains a range of discriminant bacterial, fungal and viral taxa, which highlights the potential for interactions in the observed clinical states (Fig. 1d). Patients from the larger cluster (cluster 1,  $n = 134$ ) had greater microbial diversity (Fig. 1e) and had better clinical outcomes than those in the smaller cluster (cluster 2,  $n = 83$ ) in terms of exacerbation frequency and symptoms (Fig. 1f,g). Additional geographic and clinical differences between the two clusters include a higher proportion of European patients in the frequent exacerbations cluster (82% versus 22%,  $P < 0.001$ ), with patients in that cluster also having a higher body mass index (20–31 kg m<sup>-2</sup> versus 18–24 kg m<sup>-2</sup>,  $P < 0.001$ ), a

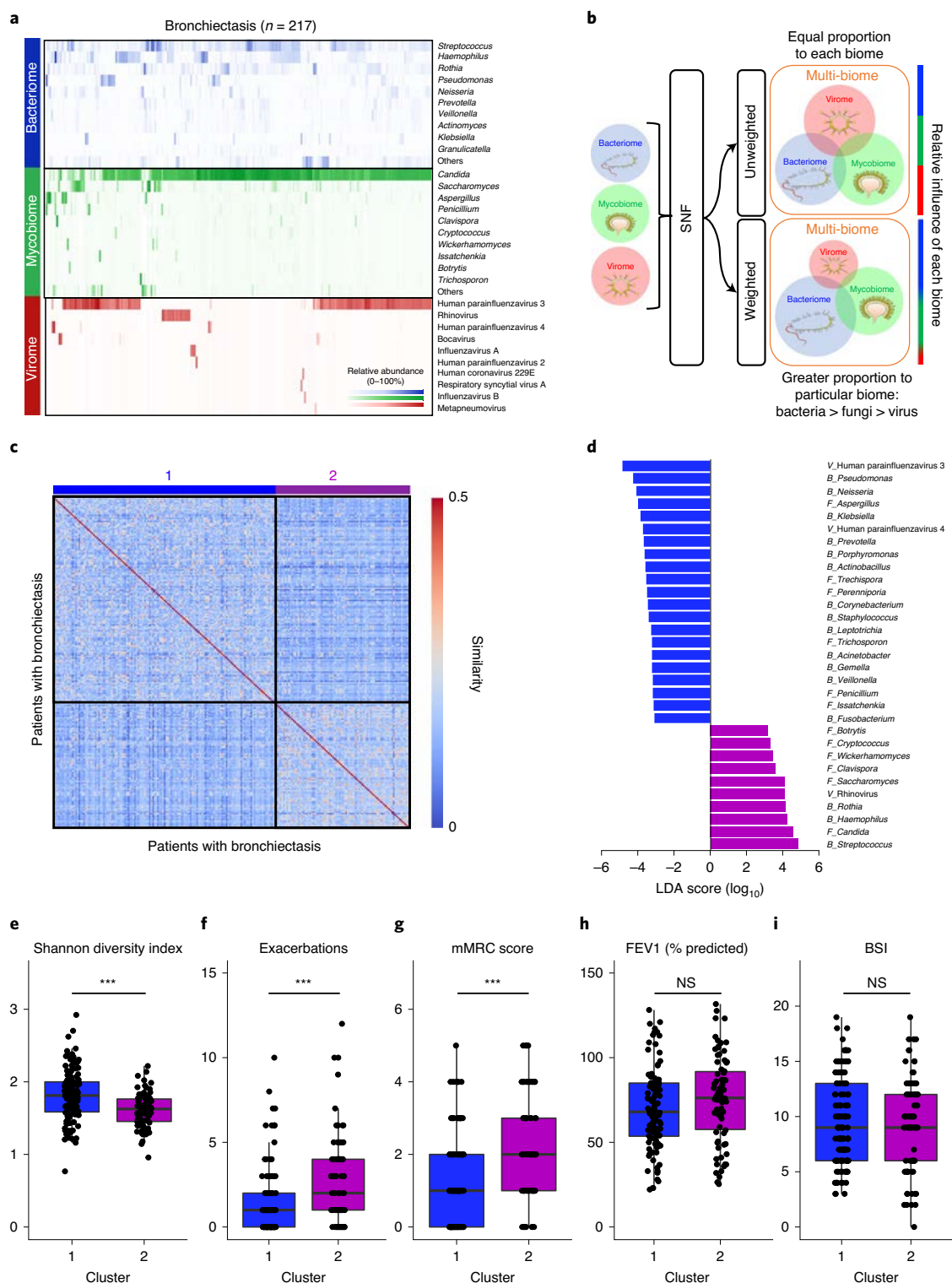
higher prevalence of chronic rhinosinusitis (CRS) (36% versus 18%,  $P < 0.001$ ), greater inhaled corticosteroid (ICS) use (50% versus 26%,  $P < 0.001$ ) and a greater likelihood of a smoking history (41% versus 22%,  $P < 0.005$ ). Although the prevalence of CRS was higher in cluster 2 than in cluster 1, it was not associated with the presence of sinus disease-associated taxa (Extended Data Fig. 3a,b), which suggests that the sampling methodologies were robust with regard to these potential confounding variables<sup>14</sup>. The association of ICS and antibiotic use on microbiome composition was not statistically significant (Extended Data Fig. 3c–e). Patients in cluster 2 carried a relative risk of 2.4 (95% confidence interval, 1.6–3.5,  $P < 0.0001$ ) for the ‘frequent exacerbator’ phenotype, defined as  $\geq 3$  exacerbations annually<sup>1</sup>. Although lung function and disease severity were similar between the clusters (Fig. 1h,i), integrated multi-biome analysis enabled clinically meaningful patient stratification in that the high-frequency exacerbators were able to be identified with greater precision when compared with use of the bacterial microbiome alone (Supplementary Table 3), a demonstration of the clinical utility of the integrative process. In contrast to existing bronchiectasis paradigms<sup>15</sup>, the high-frequency exacerbation cluster had lower prevalence (31% versus 66%,  $P > 0.001$ ) and relative abundance (10.8% versus 1.7%,  $P > 0.001$ ) of *Pseudomonas* species compared with the low-frequency exacerbation cluster.

#### Co-occurrence analyses of the interactome in low- and high-frequency exacerbation clusters.

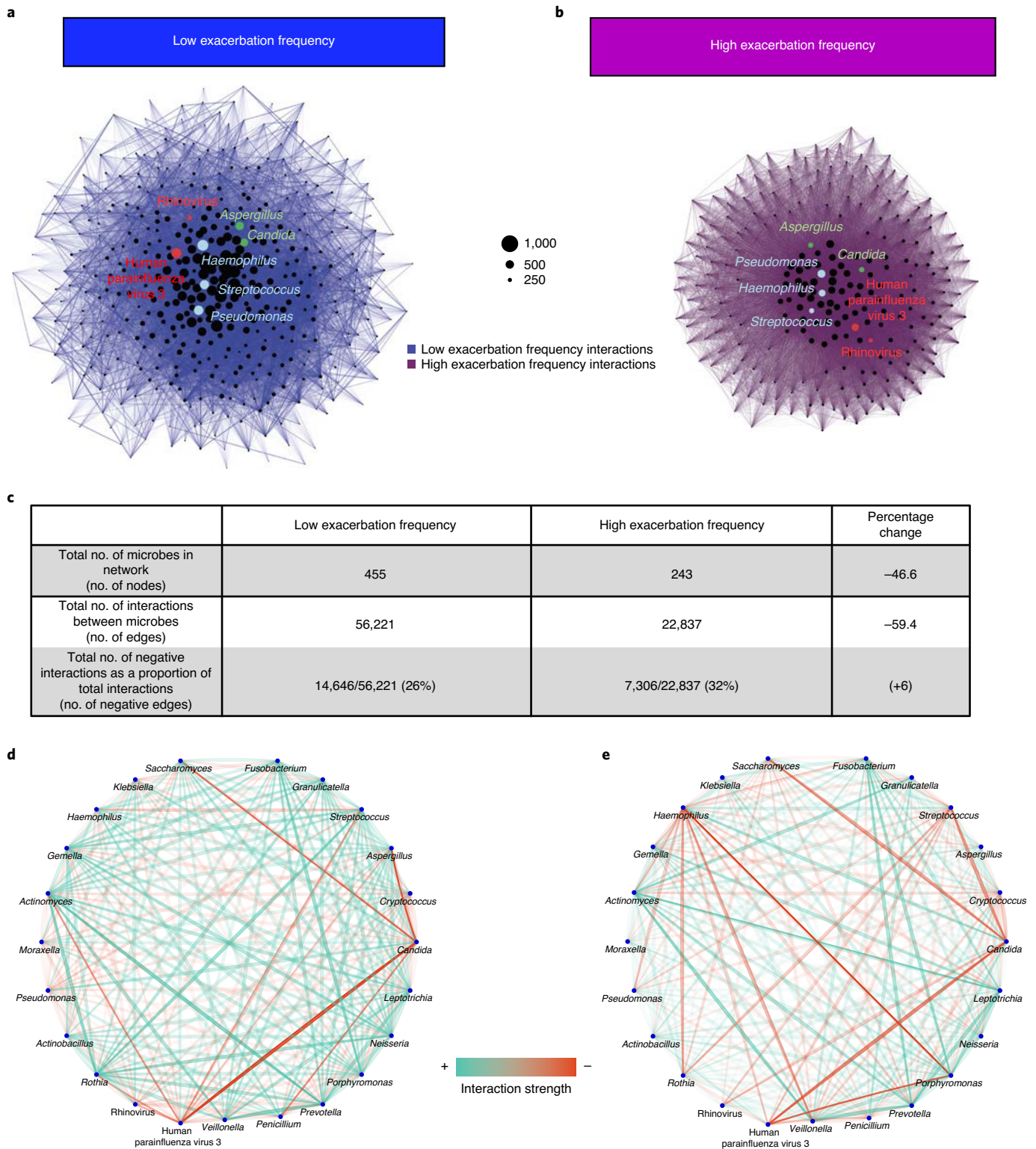
To characterize microbial interactions (the interactome) within each cluster, weighted co-occurrence analysis with an ensemble of similarity measures and regression techniques was used to generate microbial association networks (Fig. 2a,b). Leveraging the methodology of Faust et al.<sup>16</sup> mitigated compositionality of relative abundance data and provided a framework (based on graph theory) in which microbes (described as nodes) can be assessed in the context of their interconnection with (predicted) interacting partners (edges), which can be positively or negatively correlated (see Methods). Therefore, a positive interaction between microbes is defined by the consensus ensemble correlative score, whereby a positive value represents the co-occurrence of microbes and a negative value represents co-exclusion. The low-frequency exacerbation cluster had a higher total number of microbes and microbial interactions than the high-frequency exacerbation cluster, which exhibited lower diversity and a greater proportion of negative interactions between constituent microbes (Fig. 2c). An altered interactome is therefore evident in the high-frequency exacerbation cluster, suggestive of opposing microbial interactions that potentially drive this observed clinical state (Fig. 2d,e).

#### Busy, critical and influential microbes in the interactome.

Adopting network-based approaches permits assessment of alternate metrics to characterize microbiomes for potential clinical applicability<sup>17</sup>. We evaluated the network metrics node degree, stress centrality and betweenness centrality (of the nodes) to describe microbes in a network that we refer to as busy (microbes with an increased number of direct interactions with other microbes), critical (key microbes to maintain network integrity) and influential (microbes influencing other microbes in a network, including indirectly). Using this approach, we identified key taxa of clinical relevance and potential targets for antimicrobial intervention in the clusters (Fig. 3a,b). Using these metrics, a different view of the cluster-specific interactome is appreciated, with *Rothia*, *Streptococcus*, *Candida*, *Actinomyces* and *Haemophilus* representing the highest ranked taxa in the low-frequency exacerbation cluster. Of these taxa, only *Haemophilus* is similarly ranked in the high-frequency exacerbation cluster, alongside *Cryptococcus*, *Leptotrichia*, *Poryphyromonas*, *Prevotella* and *Veillonella* (Fig. 3b and Supplementary Table 4). Although some of the top taxa identified



**Fig. 1 | Integration of multi-biome data through WSNF.** Overview of the integrative microbiomics strategy for analysis of bacteriome, mycobiome and virome datasets by WSNF. **a**, Heatmap illustrating the relative abundance of key taxa within bacterial, fungal and viral communities. Collectively, these datasets represent the airway multi-biome of patients with stable bronchiectasis ( $n = 217$ ). **b**, Schematic comparison of conventional (unweighted) and weighted SNF approaches to assess the airway multi-biome. Weighting is assigned to each biome dataset based on taxonomic richness. WSNF reflects the in vivo state and overcomes the weaknesses of conventional SNF methodologies. **c**, Heatmap illustrating pairwise patient WSNF similarity scores stratified by spectral clustering. **d**, Linear discriminant analysis effect size (LEfSe) analysis of observed clusters illustrating discriminant taxa. Prefixes indicate whether taxa are bacterial (B), fungal (F) or viral (V). **e–i**, Comparison of  $\alpha$ -diversity (as Shannon diversity index,  $P = 1.9 \times 10^{-6}$ ) (**e**) and clinical features such as number of exacerbations in the preceding year ( $P = 2.5 \times 10^{-5}$ ) (**f**), breathlessness (mMRC) score ( $P = 6.3 \times 10^{-6}$ ) (**g**), lung function (as FEV1 % predicted) (**h**) and BSI (**i**) between the two identified patient clusters (cluster 1,  $n = 134$ ; cluster 2,  $n = 83$ ) according to integrated multi-biome profiles, derived from  $n = 217$  biologically independent samples. Box plots reflect median and IQRs, with whiskers bounding non-outlier values. NS, not significant; \*\*\* $P < 0.001$  (Mann–Whitney  $U$ -test).



**Fig. 2 | Co-occurrence analysis of the multi-biome network in high-frequency exacerbators. a,b,** Co-occurrence network maps of low and high exacerbation frequency clusters illustrating identified microbial interactions. Interactions between microbes (nodes) are represented by connecting lines (edges), and the number of interactions for each microbe is reflected by node size (see scale bar). Selected bacterial (light blue), fungal (green) and viral (red) taxa of clinical relevance are indicated by node coloration. **c,** Summary table listing network characteristics of the low and high exacerbation frequency clusters, highlighting the total number of detectable microbes in each network and the total number of interactions, and separating out the number of negative interactions (negative edges) observed as a proportion of overall interactions. **d,e,** Visualization of positive and negative interactions between the most abundant taxa in each respective cluster. Interactions between microbes are classified as negative if the sign of the edge weights between them is negative (negative correlation) and vice versa.

in each respective cluster commonly share the busy, influential and critical network characteristics (for example, *Streptococcus*, *Haemophilus*, *Candida* and *Cryptococcus*), all exhibit markedly different interaction networks when assessed independently within their respective clusters, where they also exhibit variable prevalence (Extended Data Fig. 4). This suggests that a microbe's interactome, rather than the microbe itself, dictates clinical status such as the risk of exacerbations. Accordingly, classification based on the interactome provides superior resolution to that provided by a single microbe (for example, *Pseudomonas* spp.) in the identification of patient populations at risk of adverse clinical outcomes. Conversely, when interaction networks were assessed in a supervised manner according to documented exacerbation frequency (<3 versus  $\geq 3$ ), as opposed to WSNF, the observed network configurations were similar but with the notable appearance of *Pseudomonas* spp. in the top taxa of frequent exacerbators (Extended Data Fig. 5).

Given that *Pseudomonas* spp. detected on culture is strongly associated with exacerbations in bronchiectasis<sup>15,18</sup>, we next specifically assessed *Pseudomonas*-interaction networks (Fig. 3c,d). *Pseudomonas* spp. exhibits distinct interactomes based on a patient's exacerbation frequency, and although an overall trend toward more negative interactions was observed for the high-frequency exacerbation cluster (Fig. 2c–e), the *Pseudomonas*-interaction network in that cluster had a higher number of positive interactions compared with that in the low-frequency exacerbation cluster (Fig. 3c,d). The *Pseudomonas*-interaction network in the low-frequency exacerbation cluster had a greater number of negative interactions than that in the high-frequency cluster ( $n=19$  versus  $n=5$ , Fig. 3d). Several important differences in *Pseudomonas*-interaction networks between clusters were evident, one of which was that *Pseudomonas* exerts a greater negative influence on *Haemophilus* and a lesser negative influence on *Streptococcus* in the high-frequency exacerbation cluster when compared with the low-frequency exacerbation cluster. Additionally, inverse relationships were observed in relation to *Aspergillus*, *Prevotella*, *Veillonella*, *Neisseria* and human parainfluenza virus 3 between clusters, in that positive interactions predominate in the high-frequency exacerbation cluster and negative interactions in the low-frequency exacerbation cluster (Fig. 3d). Therefore, *Pseudomonas* spp. presence alone does not adequately explain the published links between this microorganism and bronchiectasis exacerbations.

**Network analysis of the interactome over the course of a bronchiectasis exacerbation.** Interactomes differ based on exacerbation frequency, therefore we next evaluated the interactome prospectively across the course of exacerbations in an independent bronchiectasis cohort recruited from two hospitals in the east of Scotland. Patients had a median age of 72 years (range, 68–74 years) and were predominantly female (65%). Most had idiopathic bronchiectasis (65%) and were classified as having moderate to severe disease (median BSI, 10; IQR, 6–14) (Methods and Supplementary Table 5). We assessed bacterial, fungal and viral microbiome profiles generated for these 17 patients across three timepoints (total, 153 biomes), and generated interactome

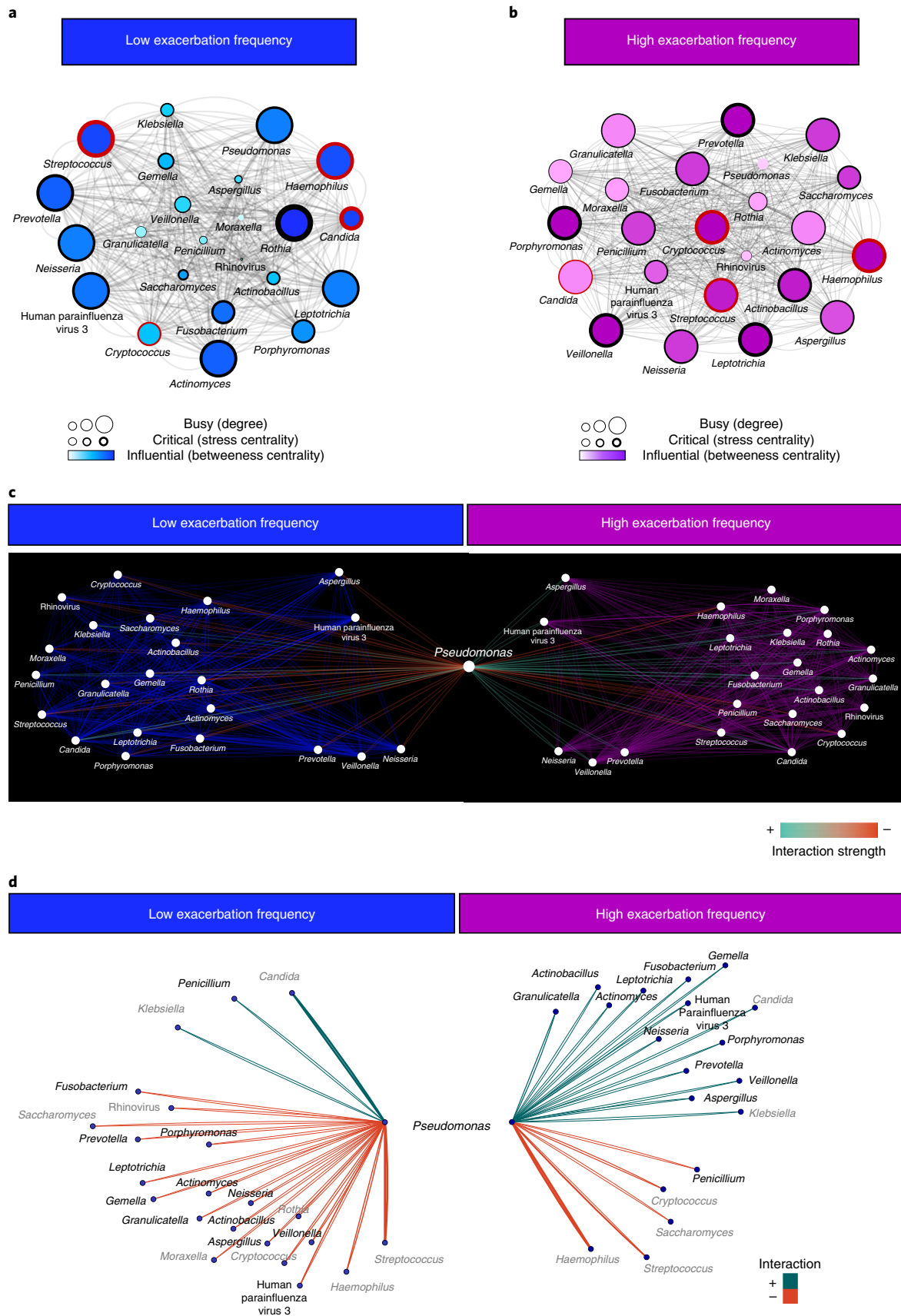
networks at baseline (before exacerbation), during exacerbation and after exacerbation (after 2 weeks of antibiotic therapy). Longitudinal analysis indicated a broad similarity of multi-biome signatures, with no significant differences observed in microbial composition or in  $\alpha$ - and  $\beta$ -diversity, suggesting the overall stability of the microbiome across exacerbation and recovery (Fig. 4a–c and Extended Data Fig. 6). By contrast, co-occurrence analysis highlighted major changes in interactomes, with an increase in the number and strength of negative interactions during exacerbations compared with baseline (before exacerbation) or following treatment (after exacerbation) (Fig. 4d–f). Detailed comparison of changes from baseline to exacerbation and from exacerbation to post-exacerbation status illustrates dynamic shifts towards a new post-exacerbation network (Fig. 4g). Fewer interactions are observed during and after exacerbation compared with baseline, which is probably explained by broad-spectrum antibiotic usage (which eliminates potentially interacting microorganisms), with the greatest overlap observed between exacerbation and post-exacerbation states (Fig. 5a). Importantly, a core interactome of 64 conserved microbial interactions exists, with the strongest interactions noted between *Prevotella*, *Leptotrichia* and *Veillonella* (Fig. 5b). To further assess microbial interactions related specifically to the exacerbation state, differential network analysis was implemented, which illustrates core and ancillary networks (Fig. 5b). The core network remains unaltered by exacerbation or therapy and includes key bacteria such as *Streptococcus*, *Prevotella*, *Veillonella*, *Neisseria*, *Leptotrichia* and *Rothia*. Conserved fungal and viral interactions involve *Cryptococcus* and rhinovirus, respectively. The interactions most susceptible to variability with exacerbation, and therefore treatment (ancillary network), involve more established respiratory pathogens such as *Pseudomonas*, *Haemophilus*, *Stenotrophomonas*, *Moraxella* and *Staphylococcus*, but also *Saccharomyces*, *Candida* (fungi), influenza virus B and metapneumovirus (viruses) (Fig. 5b).

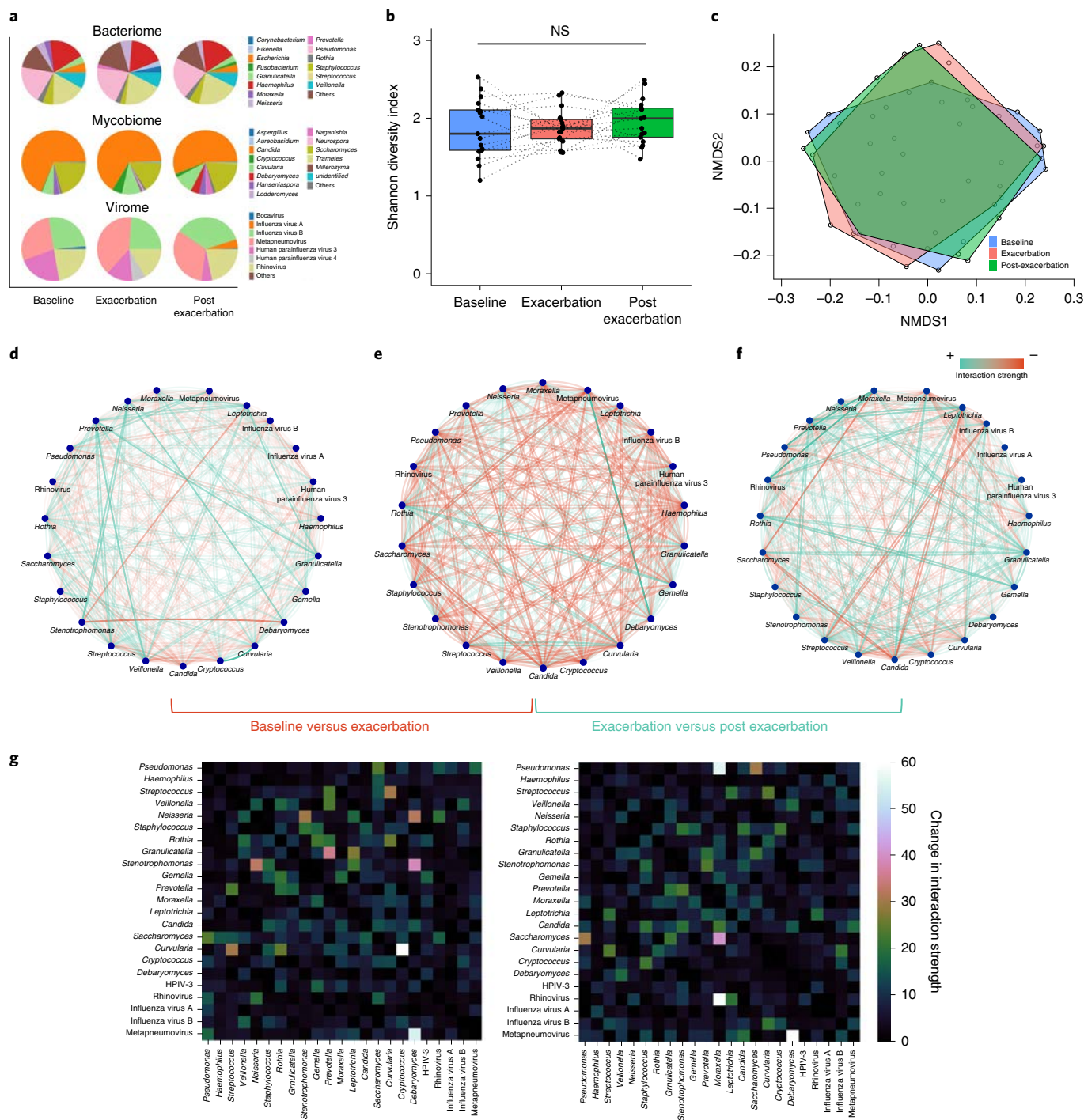
We next considered the potential effects of antibiotic exposure on interactome network dynamics by comparing observed and simulated responses. In the longitudinal study arm, patient therapy was guided by culture-based microbiology, which was generally consistent with our derived 16S ribosomal RNA analysis (Extended Data Fig. 7a). This approach resulted in several patients in the longitudinal study ( $n=12$ ) receiving  $\beta$ -lactam antibiotics for treatment of their initial exacerbation (Supplementary Table 6). We used the baseline (pre- $\beta$ -lactam exposure) interactome network (Fig. 5c) to predict network reconfiguration after  $\beta$ -lactam treatment by artificially reducing the abundance of  $\beta$ -lactam-sensitive microbes by 75% (Fig. 5d; Methods). We then compared the simulated network to that observed in the  $\beta$ -lactam-treated patients following therapy (Fig. 5e). Our network-based prediction had reliable similarity to the network observed for  $\beta$ -lactam-treated patients with respect to several microbial nodes. The rank order difference in key microbial taxa after antibiotic treatment was correctly predicted for 10 out of 13 taxa in the simulation model, further underscoring the clinical relevance and translatability of interactome analysis (Fig. 5c–e and Supplementary Table 7). Finally, as a prognostic indicator, we also

**Fig. 3 | Network characterization of busy, critical and influential microbes (nodes) in patients with frequent exacerbations.** a,b, Network visualization of key taxa in low (a) and high (b) exacerbation frequency clusters. Colored circles represent microbes and gray lines represent their associated interactions. Circle size (degree) reflects the number of direct interactions for a given microbe (termed 'busy'). Circle border thickness represents calculated stress centrality for each microbe (termed 'critical'), while color depth reflects the betweenness centrality (the 'influence') of the microbe in the network. Bacterial (*Streptococcus* and *Haemophilus*) and fungal (*Candida* and *Cryptococcus*) genera exhibiting high calculated network metrics (in both clusters) are indicated by a red border. c, Microbial network graphs in low and high exacerbation frequency clusters centering on the *Pseudomonas* node. Microbes not interacting directly with *Pseudomonas* (that is, not part of the *Pseudomonas*-interaction network) are colored according to their respective cluster. Microbes directly interacting with *Pseudomonas* are colored to reflect positive (green) or negative (red) interactions. Color depth reflects the strength of interaction (edge weight). d, Co-occurrence analysis highlighting the *Pseudomonas*-interaction network in low and high exacerbation frequency clusters. Taxa shown in black represent contrasting interactions between clusters; taxa shown in gray demonstrate no change to the directionality of their interaction between clusters.

found that interactions rather than individual microbial abundance served as a better predictor of time to next exacerbation in patients in the longitudinal arm of the study (Extended Data Fig. 7b–h).

**Shotgun metagenomic analysis of function and microbial interaction in bronchiectasis.** To assess function in identified clusters, we performed metagenomic sequencing on a subset of 20 patients

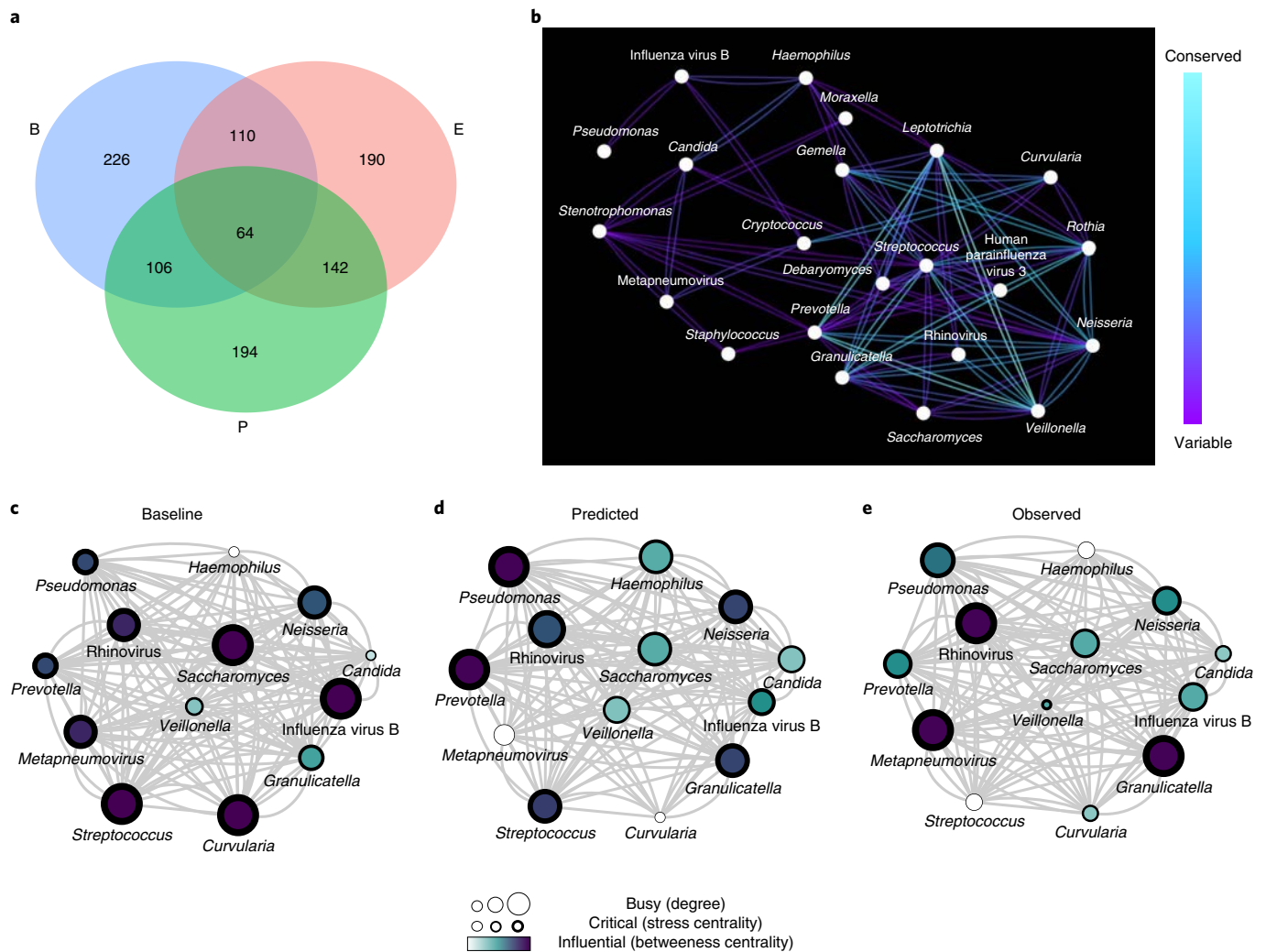




**Fig. 4 | Longitudinal analysis of the bronchiectasis multi-biome during exacerbations.** **a**, Longitudinal analysis of bacterial, fungal and viral community status in  $n=17$  bronchiectasis patients at the baseline (pre-exacerbation), exacerbation (during an established pulmonary exacerbation) and post-exacerbation (after completion of antibiotic therapy) timepoints. Pie charts illustrate aggregate microbial composition of the bacterial, fungal and viral community profiles across each timepoint. **b**, Boxplots of  $\alpha$ -diversity (as Shannon diversity index) across the three timepoints. Plots reflect median and IQRs with whiskers bounding non-outlier values of the distribution. Dotted lines identify longitudinal samples from individual patients ( $n=17$ ). NS, not significant (Mann-Whitney  $U$ -test). **c**, Non-metric multi-dimensional scaling (NMDS) plot illustrating similar multi-biome  $\beta$ -diversity across the three timepoints. Samples are grouped according to their respective longitudinal timepoint, as indicated by colored planes. **d-f**, Visualization of the interactome's positive and negative interactions between the most abundant taxa at baseline (**d**), during exacerbation (**e**) and at the post-exacerbation stage (**f**). **g**, Plots of relative interaction change comparing the changes occurring between baseline and exacerbation, and between the exacerbation and post-exacerbation states. Pairwise matrices indicate the comparative change in interaction observed between individual bacteria, fungi, or viruses. HPIV-3, human parainfluenza virus 3.

from each cluster (total,  $n=40$ ; Supplementary Table 8). The linear discriminant analysis score showed that several genes were significantly enriched in the high exacerbation frequency (HEF) cluster,

highlighting potential genetic components related to exacerbation phenotypes and to observed differences in the corresponding microbial interactome (Fig. 6a). Functional mapping of these



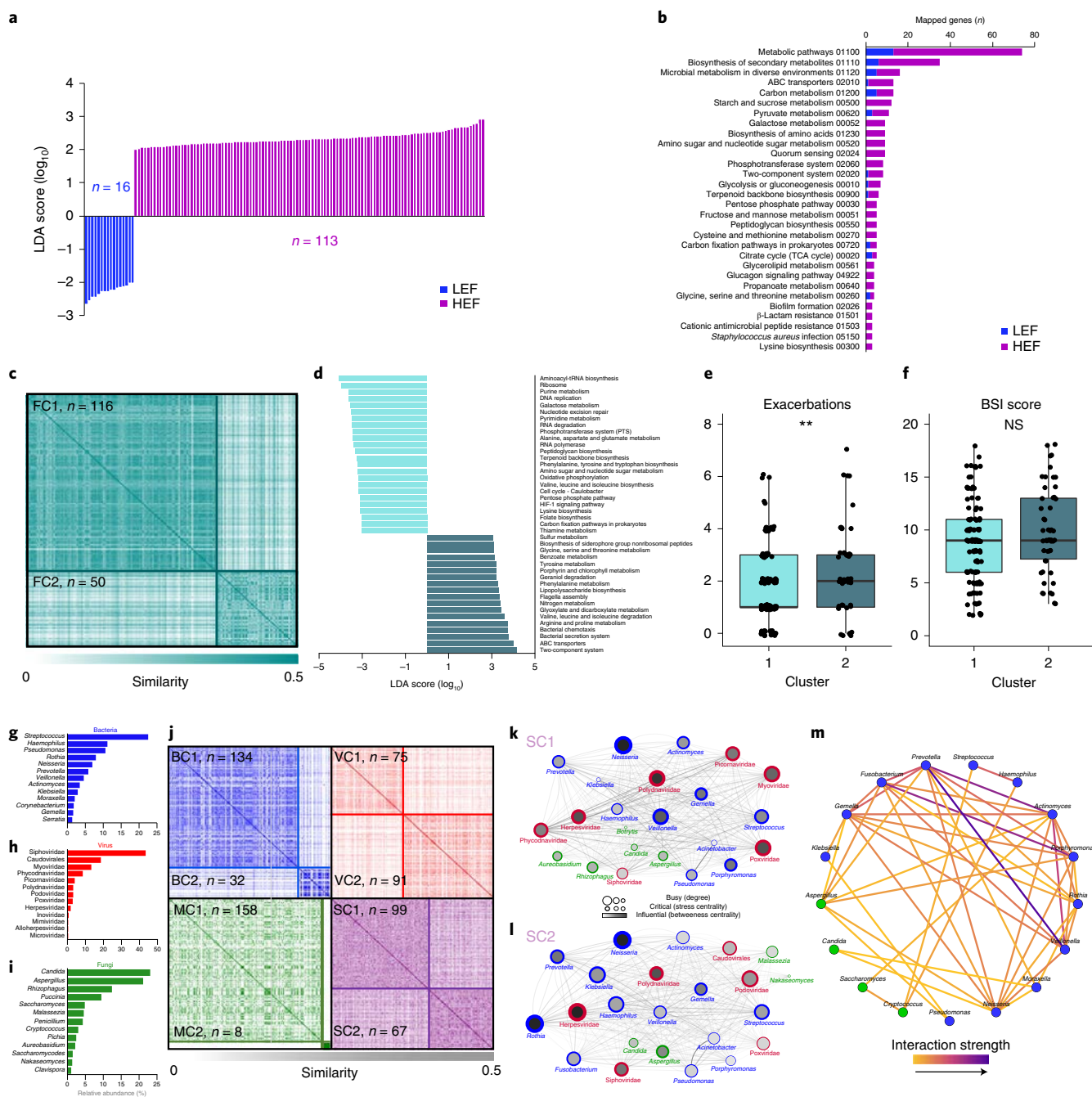
**Fig. 5 | Integrative microbiomics of the multi-biome showing the core and ancillary microbial networks in bronchiectasis exacerbations. a**, Venn diagram summarizing the observed interactions of the multi-biome across longitudinally sampled timepoints (baseline, B; exacerbation, E; and post-exacerbation, P). **b**, Network analysis illustrating the core microbial network and the ancillary microbial network implicated in bronchiectasis exacerbation. The presented network captures the interactions common to the baseline (pre-exacerbation) reference network (shown in blue as highly conserved across the exacerbation) and outlines the condition-specific networks (the interactions that vary (shown in purple) during the exacerbation and post-exacerbation stages). **c**, Baseline network analysis of bronchiectasis patients who subsequently received  $\beta$ -lactam therapy for treatment of an exacerbation ( $n=12$ ). **d,e**, A simulated network based on 75% reduction in the abundance of  $\beta$ -lactam-susceptible organisms from the baseline network (**d**) and the actual observed network reconfiguration in patients following  $\beta$ -lactam therapy (**e**).

genes identified several microbial virulence-related pathways enriched in the high-frequency exacerbation cluster, including functional categories related to quorum sensing, biofilm formation and antibiotic resistance (Fig. 6b). Based on these initial findings, we performed further metagenomic analysis on an independently recruited bronchiectasis cohort from four jurisdictions ( $n=166$ ; Singapore, Malaysia, Scotland and Italy) (Supplementary Table 9). That analysis independently identified two patient clusters based on assessment of gene function, again distinguished by the over-representation of microbial virulence functions including chemotaxis, two-component systems, secretion systems and siderophore production pathways coincident with HEF (Fig. 6c–e). Patient clusters had significant differences in lung function (forced expiratory volume in 1 second (FEV1) % predicted,  $P=0.035$ ), while disease severity (BSI) and symptoms (modified Medical Research Council (mMRC) scores) were similar (Fig. 6f and Extended Data Fig. 8).

We next taxonomically assigned all of the reads and recapitulated the multi-biome based on metagenomic data, thus enabling

comparison with prior analyses derived from targeted amplicon sequencing, and validation of the observed interactomes (Figs. 2 and 3). We performed analysis of the virome (including bacteriophages) to generate a rich viral profile (Extended Data Fig. 9). This revealed broadly similar bacteriophage profiles across geographically distinct cohorts, with a predominance of viruses from the families Siphoviridae, Caudovirales, Myoviridae and Phycodnaviridae, accounting for >70% of identified viral sequences (Extended Data Fig. 9b). Microbiome integration was then achieved with WSNF by applying weights in accordance with taxonomic richness as follows: bacteriome (992 genera) > virome (703 viral contigs) > mycobiome (16 genera) (Fig. 6g–j). Implementation of integrative microbiomics resulted in stratification of patients into two clusters separated by exacerbation risk, but in this case with even greater precision than in the initial functional analysis (Fig. 6j and Supplementary Table 10). In addition to a greater frequency of exacerbations (similarly to the earlier HEF cluster), patients in the metagenomic higher risk cluster, spectral WSNF cluster 2(SC2), had significantly reduced lung





**Fig. 6 | Metagenomic analysis of gene function and microbial interactions associated with exacerbation.** **a**, LefSe analysis of gene function in  $n=20$  patients from each of the low exacerbation frequency (LEF) and HEF clusters. **b**, KEGG pathway mapping of the genes in **a**, indicating enriched functional pathways. ABC, ATP-binding cassette; TCA, tricarboxylic acid. **c**, Heatmap illustrating pairwise patient similarity scores based on functional analysis illustrating distinct clusters (FC1 and FC2). **d**, LefSe analysis of gene functional pathways defining groups FC1 and FC2. Comparison of patient clusters (FC1,  $n=116$ ; FC2,  $n=50$ ) derived from the functional metagenomic profiles of  $n=166$  biologically independent samples recruited from Singapore, Malaysia (Kuala Lumpur), Scotland (Dundee) and Italy (Milan). HIF-1, hypoxia-inducible factor 1; tRNA, transfer RNA. **e, f**, Exacerbation frequency ( $P=0.009$ ) (**e**) and BSI ( $P=0.1112$ ) (**f**) for those two clusters. Box plots reflect median and IQRs with whiskers bounding non-outlier values. NS, not significant;  $**P<0.01$  (Mann-Whitney  $U$ -test). **g–i**, Relative abundance of bacteria (**g**), viruses (**h**) and fungi (**i**) determined by metagenomics across an independently recruited bronchiectasis cohort. **j**, Heatmap of pairwise similarity scores and spectral clustering of individual biome views (bacteria, blue; virus, red; fungi, green) compared with WSNF integration (purple). **k, l**, Network analysis of WSNF clusters SC1 (low exacerbation) (**k**) and SC2 (high exacerbation) (**l**). Bacteriome, virome and mycobiome nodes are indicated, respectively, by blue, red and green borders. **m**, Conserved interactions (bacteria and fungi only) across targeted and metagenomic interactome analyses. Interactions common to both networks are color-coded according to the strength of the conserved interaction.

function (78% versus 72% predicted FEV1;  $P=0.0179$ ), while symptoms (mMRC score) and disease severity (BSI) were similar between the clusters (SC1 versus SC2). Patients from the SC2 (high exacerbation)

cluster also had distinct bacteriophage profiles, with noticeable decreases in the relative abundance of Siphoviridae, Caudovirales and Myoviridae, and increases in Picornaviridae, Polydnaviridae

and Herpesviridae (Extended Data Fig. 9b–d). BLAST (Basic Local Alignment Search Tool) analysis enabled the assignment of approximately 60% of contigs, and revealed increased abundance of specific bacteriophage sequences related to the *Dickeya* phage, the *Pseudomonas* bacteriophage Pfl and the *Streptococcus* phages Javan367, phiARI0131-1/phiARI0131-2, Javan116 and Javan 536 (Extended Data Fig. 9e). Although an increased prevalence of CRS was identified and assessed in our initial analysis (Extended Data Fig. 3), this was not replicated in metagenomic analysis (26.9% versus 36.2%,  $P=0.557$ ). Metagenomic validation did not detect significant differences in ICS use between clusters (43.3% versus 40.6%,  $P=0.850$ ) or in antibiotic therapy (52.6% versus 62.3%,  $P=0.058$ ). Antimicrobial resistance genes were increased in abundance in the SC2 cluster, coincident with the observed altered bacteriophage profiles (Extended Data Fig. 9f). The key difference between clusters (that is, exacerbation frequency) remained consistent even when alternate sequencing approaches were used, representing a strong validation of the clinical relevance and reproducibility of interactomes and their associated microbial networks (Fig. 6k–m).

Co-occurrence analysis of the high-risk cluster defined on metagenomic sequencing identifies interaction networks and keystone taxa including bacteriophages, which themselves exhibit a marked shift in phage profile between clusters (Extended Data Fig. 9a–e). Clear differences in overall network configuration are observed between the low-exacerbation (SC1) and high-exacerbation (SC2) clusters, consistent with prior targeted amplicon sequencing analysis, while metagenomics also indicates an increased abundance of antimicrobial resistance determinants in the SC2 cluster (Fig. 6k–m and Extended Data Fig. 9f).

The derivation of network configurations from two independent bronchiectasis cohorts, using two sequencing approaches, facilitated direct comparison between the interactomes generated by each method, with regard to the high exacerbation risk clusters (that is, HEF versus SC2) (Fig. 6m). Direct comparison between microbes detected in both approaches indicated that 89.9% of interactions (that is, 267 interactions between 18 microbes) were common between the HEF and SC2 clusters, which strongly validates the associations of these networks with clinical exacerbation risk and confirms the overall importance and clinical relevance of interactome analysis.

## Discussion

Here we present a multi-biome framework that uses integrative microbiomics to assess bacterial, viral and fungal communities in individual patients to reveal features of the microbiome, the findings of which have implications for the use of antibiotics in clinical practice.

Infection is central to bronchiectasis pathogenesis and is based upon conceptual frameworks such as the ‘vicious cycle’ and ‘vicious vortex’<sup>2,3</sup>. Targeting bacteria with antibiotics reduces bacterial load, the accompanying inflammation and therefore exacerbation risk, which in turn alleviates symptoms and improves clinical outcomes<sup>19,20</sup>. However, the role of coexisting, even commensal or pathobiont microorganisms, is not considered in this model. Furthermore, this model fails to explain improved outcomes in those receiving antibiotics not necessarily targeting their dominant pathogen, exemplified by macrolide use in *Pseudomonas* infection<sup>21</sup>. Patients receiving amoxicillin or a macrolide in the presence of *Pseudomonas* improve, and this is thought to be due to the drug’s anti-inflammatory properties or the presence of co-infection<sup>22</sup>. Here interactomes provide an alternative conceptual framework to better understand antibiotic use and the treatment of exacerbations in bronchiectasis. This is supported by our longitudinal analysis in which prescribed antibiotics both influence the interactome and confer clinical benefit. Therefore, modulation of *Pseudomonas*-related interactions, instead of directly targeting the organism itself, might

restrict its pathogenic potential, and explains the observed improved clinical outcomes. Additionally, the observed clinical benefit gained by the use of antibiotics to which a target organism may be resistant (as established in cystic fibrosis) is potentially also explained by the influence of antibiotics on the interactome, in that the effects on susceptible microorganisms within the interaction network of the target pathogen indirectly modulate its virulence<sup>23</sup>. The reported absence of detectable change in respiratory microbiomes during bronchiectasis exacerbations, even following antibiotics, further suggests that microbial abundance alone provides an incomplete view of airway microbial ecology<sup>10,24</sup>. Our validation cohort, including the metagenomic survey of bacteriophages, uncovered a striking change in bacteriophage profiles between the clinical exacerbation clusters. A higher burden of antimicrobial resistance genes, coincident with altered bacteriophage abundance, was observed despite an absence of significant difference in antimicrobial therapy between clusters. Data on the relationships between the resident airway microbes and the increased bacterial load during exacerbations, and on the mechanisms driving the evolution from stability to exacerbation, are lacking, and an improved understanding of interactomes (including bacteriophages) will provide key insights into the in vivo state.

The value of data integration with SNF for multi-dimensional datasets in transcriptomics, proteomics and metabolomics has been demonstrated in airway diseases such as chronic obstructive pulmonary disease (COPD)<sup>25</sup>, but these methods have not been previously applied to microbiome integration. Conventional SNF is not optimized for biological systems such as multi-kingdom microbiomes, in which dynamism and potential dominance of one kingdom over the others need to be considered. Using a WSNF approach based on richness, we have demonstrated improved patient stratification in bronchiectasis by identifying high-frequency exacerbators with an accuracy exceeding that obtained using a single microbial group. The methods described here have been made accessible to the research community through our online webtool (<https://integrative-microbiomics.ntu.edu.sg>).

Traditionally, exacerbations are considered to occur when an increased bacterial load or acquisition of a new virus ensues, however analysis of a single microbial group by bacterial abundance or viral polymerase chain reaction (PCR) has been shown to be inadequate to discriminate between the stable and exacerbation states in bronchiectasis<sup>9,10</sup>. Interactome analysis goes deeper by identifying changing inter-kingdom interactions during an exacerbation. Despite identifying clinically relevant patient clusters tied to exacerbation frequency, key bronchiectasis pathogens, including *Pseudomonas* and *Aspergillus*<sup>8,26</sup>, were detectable in the low-frequency exacerbator group, suggesting that reliance on detecting the presence or abundance of a particular organism alone is sub-optimal for predicting exacerbation risk. To better understand this, we used network analysis, which provides insights into the overall community dynamics rather than the occurrence or abundance alone<sup>16,17</sup>. The application of this approach to the airway microbiome demonstrates that bronchiectasis patients at highest risk of exacerbations have an interactome dominated by antagonistic interaction between microbial kingdoms, explaining their lower  $\alpha$ -diversity, in which microbes compete rather than cooperate with one another.

Assessment of the interactome as a network of busy, critical and influential microorganisms in an airway ecosystem highlights the relevance of established bronchiectasis pathogens such as *Haemophilus*. However, particularly in the high-frequency exacerbation cluster, relationships with other bacteria such as anaerobes (*Prevotella* and *Veillonella*) or with other kingdoms such as fungi (*Cryptococcus*) are novel and are previously unrecognized in bronchiectasis. The uncovered relationship to anaerobes is particularly interesting given that anaerobes are detected at high frequencies in the cystic fibrosis airway, with conflicting results in attempts to

link them with disease outcomes or exacerbations<sup>27,28</sup>. However, *Veillonella* has been associated with bronchiectasis exacerbations in the BLESS cohort<sup>7</sup>. Key pathogenic taxa such as *Pseudomonas*, with established links to bronchiectasis exacerbations, demonstrate contrasting interactomes between the low- and high-frequency exacerbators and confirm the relevance of integrative microbiomics in precision microbiology.

To further validate our findings, we assessed the time course of the interactome in a prospective longitudinal bronchiectasis cohort experiencing exacerbations. This first confirmed the findings of prior microbiome studies in bronchiectasis, which indicated the stability of the microbiome through the course of the exacerbation, and then, following treatment, little change in microbial composition,  $\alpha$ - or  $\beta$ -diversity<sup>5,10</sup>. However, we detected a changed interactome, unassessed in prior studies<sup>5,10</sup>, which did not use integrative approaches. A clear shift toward antagonistic microbial relationships during exacerbation was evident, similar to that observed in our high-frequency exacerbation cluster, a finding unexplained by linear increases in pathogen dominance as reflected by similar diversity indices. These findings, further validated by the core and ancillary interactomes observed during exacerbation, underscore the advantages conferred by network analysis, which reveals relationships that are undetectable by microbial abundance or identity assessment alone. Our approach highlights the relevance of inter-kingdom interactomes that vary during exacerbations and which offer deeper insight into potential triggers of microbial virulence. The interactome provides new and previously unrecognized targets for antimicrobial therapy that could be considered as an alternative to, or used in combination with, established regimens to increase efficacy. We demonstrate and confirm that simulated microbial networks can be reconfigured in response to antibiotic therapy, highlighting the clinical potential and applicability of the interactome approach as a model for the prediction of therapy-induced microbial dynamics. The benefits of targeting busy, critical or influential microbes in an interactome, however, remain unknown and unaddressed by this work, and should be the focus of future studies.

Our study demonstrates an integrative microbiomics approach to the study of the multi-biome in chronic airway disease, however, it has some limitations. First, the patients were recruited from the established CAMEB cohort, which by design is cross-sectional, hence we use largely static data to predict dynamic interaction<sup>8,12</sup>. This is partially overcome by the inclusion of a longitudinal cohort in our analysis, to better assess the temporal dynamics in relation to exacerbation and antibiotic treatment. Next, although 16S methodologies are well-established, there are inherent limitations, including the underrepresentation of mycobacteria, an important group of organisms in bronchiectasis<sup>29</sup>. The European arm of the CAMEB cohort was wholly recruited from Scotland, and therefore may not fully capture the diversity in the European subcontinent. Additionally, fungal internal transcribed spacer (ITS) sequencing approaches are challenged by the underdevelopment of available reference databases<sup>30</sup>. Our initial virome analysis, although broad, comprehensive and informed by established literature, targets a known virus panel and therefore is subject to bias. We attempted to overcome this, at least partially, through the use of a metagenomics validation approach with the inclusion of bacteriophage analysis. Future work and alternative approaches to the assessment of viromes, such as RNA sequencing, may yield different results and be more comprehensive, thereby enabling greater weighting of the viral contribution to the overall integrated microbiome, an important area of future exploration given the relatively poorly defined role of viruses in bronchiectasis. Only young healthy controls recruited from Asian countries were evaluated in our comparison of viral loads with bronchiectasis, and, additional older controls, in the age groups afflicted by bronchiectasis may have been of value. Furthermore, although networks were weighted based on species

richness, their true influence on the microbiome is not necessarily captured by richness alone, but is instead a function of functional genes, competition, substrate utilization and energy flux through the ecosystem, traits that cannot be comprehensively assessed by sequencing alone. Although metagenomics potentially represents a less biased alternative approach, which we have used as the validation procedure, it may itself underestimate fungal presence given the technical challenges associated with mycobion evaluation and the relatively higher airway bacterial burden, which together potentially obscure the influence that fungi may have on the interactome. Furthermore, we 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 sample that is intermediate between the upper and lower airway. Finally, although observational data suggest a 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; this is another avenue for future work.

The airway microbiome and its accompanying interactome are likely to be critical predictors of antibiotic treatment response, and to provide a theoretical basis for understanding several phenomena associated with antibiotics that remain unexplained. Manipulation of microbiomes by means other than antibiotics is being explored<sup>31,32</sup>, and the effect of probiotics and manipulation of the host response on the interactome should be considered. Holistic analytical approaches that reflect the *in vivo* state, which go beyond microbial identity alone, and which consider the complexity of the inter-kingdom interactions demonstrated by integrative microbiomics, may improve patient stratification, clinical trial design and the therapeutic outcomes in bronchiectasis and other respiratory diseases.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-021-01289-7>.

Received: 30 April 2020; Accepted: 16 February 2021;

Published online: 05 April 2021

## References

- Chalmers, J. D. et al. Characterization of the 'frequent exacerbator phenotype' in bronchiectasis. *Am. J. Respir. Crit. Care Med.* **197**, 1410–1420 (2018).
- Flume, P. A., Chalmers, J. D. & Olivier, K. N. Advances in bronchiectasis: endotyping, genetics, microbiome, and disease heterogeneity. *Lancet* **392**, 880–890 (2018).
- Chalmers, J. D., Chang, A. B., Chotirmall, S. H., Dhar, R. & McShane, P. J. Bronchiectasis. *Nat. Rev. Dis. Primers* **4**, 45 (2018).
- Chotirmall, S. H. & McShane, P. J. Time to acknowledge, address, and take action against bronchiectasis. *Lancet Glob. Health* **7**, e1162–e1163 (2019).
- Woo, T. E. et al. A longitudinal characterization of the non-cystic fibrosis bronchiectasis airway microbiome. *Sci. Rep.* **9**, 6871 (2019).
- Rogers, G. B., Bruce, K. D., Martin, M. L., Burr, L. D. & Serisier, D. J. The effect of long-term macrolide treatment on respiratory microbiota composition in non-cystic fibrosis bronchiectasis: an analysis from the randomised, double-blind, placebo-controlled BLESS trial. *Lancet Respir. Med.* **2**, 988–996 (2014).
- Rogers, G. B. et al. A novel microbiota stratification system predicts future exacerbations in bronchiectasis. *Ann. Am. Thorac. Soc.* **11**, 496–503 (2014).
- Mac Aogáin, M. et al. Immunological corollary of the pulmonary microbiome in bronchiectasis: the CAMEB study. *Eur. Respir. J.* **52**, 1800766 (2018).
- Mitchell, A. B. et al. Viruses in bronchiectasis: a pilot study to explore the presence of community acquired respiratory viruses in stable patients and during acute exacerbations. *BMC Pulm. Med.* **18**, 84 (2018).
- Cox, M. J. et al. Longitudinal assessment of sputum microbiome by sequencing of the 16S rRNA gene in non-cystic fibrosis bronchiectasis patients. *PLoS One* **12**, e0170622 (2017).

11. Crichton, M. L., Aliberti, S. & Chalmers, J. D. A systematic review of pharmacotherapeutic clinical trial end-points for bronchiectasis in adults. *Eur. Respir. Rev.* **28**, 180108 (2019).
12. Mac Aogáin, M. et al. Distinct ‘immunoallertypes’ of disease and high frequencies of sensitization in non-cystic fibrosis bronchiectasis. *Am. J. Respir. Crit. Care Med.* **199**, 842–853 (2019).
13. Chalmers, J. D. & Chotirmall, S. H. Bronchiectasis: new therapies and new perspectives. *Lancet Respir. Med.* **6**, 715–726 (2018).
14. Kumpitsch, C., Koskinen, K., Schöpf, V. & Moissl-Eichinger, C. The microbiome of the upper respiratory tract in health and disease. *BMC Biol.* **17**, 87 (2019).
15. Araujo, D. et al. The independent contribution of *Pseudomonas aeruginosa* infection to long-term clinical outcomes in bronchiectasis. *Eur. Respir. J.* **51**, 1701953 (2018).
16. Faust, K. et al. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* **8**, e1002606 (2012).
17. Layeghifard, M., Hwang, D. M. & Guttman, D. S. Disentangling interactions in the microbiome: a network perspective. *Trends Microbiol.* **25**, 217–228 (2017).
18. Finch, S., McDonnell, M. J., Abo-Leyah, H., Aliberti, S. & Chalmers, J. D. A comprehensive analysis of the impact of *Pseudomonas aeruginosa* colonization on prognosis in adult bronchiectasis. *Ann. Am. Thorac. Soc.* **12**, 1602–1611 (2015).
19. Laska, I. F., Crichton, M. L., Shoemark, A. & Chalmers, J. D. The efficacy and safety of inhaled antibiotics for the treatment of bronchiectasis in adults: a systematic review and meta-analysis. *Lancet Respir. Med.* **7**, 855–869 (2019).
20. Sibila, O. et al. Airway bacterial load and inhaled antibiotic response in bronchiectasis. *Am. J. Respir. Crit. Care Med.* **200**, 33–41 (2019).
21. Chalmers, J. D. et al. Long-term macrolide antibiotics for the treatment of bronchiectasis in adults: an individual participant data meta-analysis. *Lancet Respir. Med.* **7**, 845–854 (2019).
22. Dickson, R. P. & Morris, A. Macrolides, inflammation and the lung microbiome: untangling the web of causality. *Thorax* **72**, 10–12 (2017).
23. Vandeplasseche, E., Tavernier, S., Coenye, T. & Crabbé, A. Influence of the lung microbiome on antibiotic susceptibility of cystic fibrosis pathogens. *Eur. Respir. Rev.* **28**, 190041 (2019).
24. Budden, K. F. et al. Functional effects of the microbiota in chronic respiratory disease. *Lancet Respir. Med.* **7**, 907–920 (2019).
25. Li, C. X., Wheelock, C. E., Skold, C. M. & Wheelock, A. M. Integration of multi-omics datasets enables molecular classification of COPD. *Eur. Respir. J.* **51**, 1701930 (2018).
26. Aliberti, S. et al. Clinical phenotypes in adult patients with bronchiectasis. *Eur. Respir. J.* **47**, 1113–1122 (2016).
27. Tunney, M. M. et al. Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **177**, 995–1001 (2008).
28. Muhlebach, M. S. et al. Anaerobic bacteria cultured from cystic fibrosis airways correlate to milder disease: a multisite study. *Eur. Respir. J.* **52**, 1800242 (2018).
29. Sulaiman, I. et al. Evaluation of the airway microbiome in nontuberculous mycobacteria disease. *Eur. Respir. J.* **52**, 1800599 (2018).
30. Ali, N., Mac Aogáin, M., Morales, R. F., Tiew, P. Y. & Chotirmall, S. H. Optimisation and benchmarking of targeted amplicon sequencing for mycobion analysis of respiratory specimens. *Int. J. Mol. Sci.* **20**, 4991 (2019).
31. Panigrahi, P. et al. A randomized synbiotic trial to prevent sepsis among infants in rural India. *Nature* **548**, 407–412 (2017).
32. Durack, J. et al. Delayed gut microbiota development in high-risk for asthma infants is temporarily modifiable by *Lactobacillus* supplementation. *Nat. Commun.* **9**, 707 (2018).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

## Methods

**Study population.** Patients with stable bronchiectasis were recruited as part of the CAMEB study<sup>8</sup>. Recruitment was carried out at three sites in Singapore (Singapore General Hospital, Changi General Hospital and Tan Tock Seng Hospital) and one Malaysian site (UKM Medical Centre, Kuala Lumpur), while an age-, sex- and disease severity-matched group (based on BSI) was recruited from a single European site (Ninewells Hospital, Dundee, UK) to control for confounding factors across different geographic locations<sup>8,35</sup>. At screening, patients for inclusion had confi med radiological bronchiectasis on HRCT. Patients were recruited during outpatient attendance and were clinically stable, which was defi ed as the absence of new symptoms and no change to bronchiectasis therapy in the preceding 4 weeks. Patients were excluded if they had a concurrent major respiratory disease as their primary diagnosis (asthma or COPD)<sup>34,35</sup>, were pregnant or breastfeeding, had active mycobacterial disease or were on chemotherapy. Patients with active infection (necessitating acute use of antibiotics) or who were taking systemic corticosteroids in the 4 weeks preceding recruitment were also excluded. Exacerbations in the preceding year were defi ed in accordance with established consensus<sup>36</sup>. To study changes to the integrated microbiome during exacerbations and after antibiotic treatment, a cohort of patients was recruited from two hospitals in the east of Scotland (2016–2017). Patients were enrolled during clinical stability using the inclusion criteria described above and were asked to provide a spontaneous sputum sample at baseline, with repeat sputum sampling performed at the onset of exacerbation ( $\leq 24$  hours after commencing antibiotic therapy), and after 14 days of antibiotic treatment. The sfi al sample was taken following cessation of antibiotic treatment on day 14 once clinical recovery had been achieved. The presence of an exacerbation was defi ed in accordance with established consensus<sup>36</sup>. Thi ty controls, with no active or past history of respiratory or other medical disease and normal spirometry, were recruited in Singapore. These individuals were recruited as community volunteers through Nanyang Technological University, Singapore. Patient demographics are listed in Supplementary Tables 1 and 5. The study was approved by the institutional review boards (IRBs) of all of the participating institutes, and all of the patients gave written informed consent. For metagenomic analysis and validation, an independent cohort of 166 bronchiectasis patients was recruited from clinical sites described above in Singapore ( $n = 43$ ), Malaysia ( $n = 25$ ) and the UK ( $n = 76$ ). Additional patients were recruited from a fourth site, in Milan, Italy ( $n = 22$ ), through the Bronchiectasis Program, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan. Relevant clinical characteristics, bronchiectasis etiology and patient demographics of the cross-sectional, longitudinal and metagenomic study cohorts can be found in Supplementary Tables 1, 5, 8 and 9.

**Ethics approval.** The study was approved by the IRBs of all of the participating institutes as follows: CIRB 2016/2073 mutually recognized by DSRB, NTU IRB-2016-01-031, NTU IRB-2017-07-023, NTU IRB-2017-12-010 (Singapore); UMMC 2018725-6524 (Malaysia); NHD 12/ES/0059, NHD 16/NW/0101 (Dundee, Scotland) and 255\_2020 (Comitato Etico Milano Area 2, Milan, Italy). All of the patients gave written informed consent to participate.

**Clinical data and specimen collection.** Sputum was obtained from each participant. Spontaneously expectorated representative sputum from a deep cough with the assistance of a chest physiotherapist or induction protocol (when appropriate) was collected into sterile containers and transported (on ice) for evaluation<sup>37</sup>. An equal volume of Sputasol (Thermo Fisher Scientific) was added to each sample and shaken for 15 min at 37 °C. Sputasol-homogenized samples were either stored (at  $-80$  °C) or mixed with two volumes of RNAlater (Sigma-Aldrich) for DNA extraction and microbiome analysis<sup>8,38</sup>. All of the specimens from clinical sites were transported promptly (within 4 h), appropriately (temperature-controlled) and processed centrally at a single site to ensure consistency and standardization of all experimental work. Samples from the three Singaporean hospitals were transported on ice by courier to the Nanyang Technological University ( $\leq 4$  h after collection). To ensure quality control of materials transported from sites outside Singapore, specimens were shipped on dry ice in temperature-controlled containers and their integrity checked on arrival before use. All of the DNA extraction experiments were performed at Nanyang Technological University, Singapore, using a single standardized protocol (described in further detail below).

**Sputum DNA and RNA extraction.** Sputum DNA was extracted from a 250 mg sample using methods previously described<sup>38</sup>. In brief, sputum samples in RNAlater were centrifuged at 8,000g for 10 min and resultant pellets resuspended in 500  $\mu$ l sterile PBS (GE Lifesciences) and transferred to sterile bead mill tubes (VWR) containing 1 mm sterile glass beads (Sigma-Aldrich). Next, homogenization was carried out using a bead mill homogenizer (VWR) with three rounds of bead beating at top speed (setting 6) for 30 s periods, with cooling on ice in between rounds, and DNA was purified using the Roche High-pure PCR Template Preparation Kit according to the manufacturer's instructions. For RNA extraction, sputum was subjected to mechanical lysis by bead beating in PowerBead garnet bead tubes (0.7 mm, Qiagen) using a bead mill homogenizer (VWR) as described above. Total RNA was then purified using the RNeasy Mini

Kit (Qiagen), with immediate conversion to complementary DNA with the iScript Reverse Transcription Supermix for RT-qPCR (Bio-Rad).

**Analysis of the bacterial and fungal communities.** For the cross-sectional arm of the study, targeted amplicon sequencing of the 16S rRNA gene (bacteriome) was carried out using a previously validated amplicon shotgun sequencing protocol with paired-end analysis ( $2 \times 101$  base pair reads) on an Illumina HiSeq 2500 platform<sup>8,39</sup>. Fungal community analysis (mycobiome analysis) was performed as detailed for the CAMEB cohort<sup>8</sup>. For analysis of longitudinal samples, an Illumina MiSeq platform was used in conjunction with validated wet-lab (Illumina) and cloud-based analytic workflows (Basespace Labs) for both bacterial 16S and fungal nuclear ribosomal ITS sequences<sup>40,40</sup>. Here specific primer sequences targeting the hypervariable regions V3 and V4 of the 16S rRNA gene and the ITS2 region were targeted using specific primer sets<sup>41,42</sup>. The 16S rRNA sequences were mapped to an Illumina-curated version of the greengenes 16S rRNA database (16S Metagenomics, v1.0.1), whereas ITS sequences were aligned to v7.2 of the UNITE database (ITS metagenomics v1.0.0) using a high performance implementation of the Ribosomal Database Project (RDP) classifier<sup>43–45</sup>. Blank DNA extractions were performed on sterile PBS and underwent both 16S rRNA and ITS analysis to assess the degree of background contamination associated with DNA extraction and sequencing reagents. DNA concentrations obtained from blank extractions were far lower than for test samples ( $0.31 \text{ ng } \mu\text{l}^{-1}$ , IQR  $0.01$ – $0.67 \text{ ng } \mu\text{l}^{-1}$  versus  $37.6 \text{ ng } \mu\text{l}^{-1}$ , IQR  $36.5$ – $104.3 \text{ ng } \mu\text{l}^{-1}$ ) and were pooled ( $n = 4$ ) to obtained sufficient material for amplification and sequencing, with the sample volume of purified amplicons added at approximately fourfold the volume of test samples in the sequencing library pool to achieve sufficient sequence data. Sequenced blanks contained read counts below that of test samples and therefore are unlikely to have had a substantial influence on observed microbiome profiles. Read counts in blanks as well as their taxonomic assignments are given in Extended Data Fig. 10. All of the sequence data from this study have been uploaded to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under project accession PRJNA590225.

**Quantitative PCR-based analysis of the viral community.** Real-time quantitative PCR (rt-qPCR) was performed using KAPA SYBR FAST Master Mix Universal (Kapa Biosystems) on an Applied Biosystems StepOnePlus Real-Time PCR System. A 10  $\mu$ l rt-qPCR reaction was performed for each well using a 96-well plate (Applied Biosystems) with 5  $\mu$ l 2X KAPA master mix, 0.2  $\mu$ l 50X ROX High, 0.2  $\mu$ l 10  $\mu$ M forward primer, 0.2  $\mu$ l 10  $\mu$ M reverse primer, 4.2  $\mu$ l water and 0.2  $\mu$ l DNA template. Cycling conditions were as follows: 95 °C activation (20 s) followed by 40 cycles of 95 °C denaturation (3 s) and 60 °C annealing and extension (30 s). A cycle threshold  $< 35$  was used for analysis, and viral load was quantified with reference to standard curves of synthetic target sequences of known concentration (gBlocks, Integrated DNA Technologies) for each of the 17 viral pathogens investigated (Supplementary Table 2). Primers used for each virus are detailed in Supplementary Table 11. In addition to a positive cycle threshold, fidelity of target amplification was assessed using melt curve analysis, and qualitative agarose gel electrophoresis of post-PCR amplicons was done to ensure consistent fragment size of positive samples.

**Integration of multi-biome data by SNF.** *Unweighted SNF.* Integration of bacterial and viral community data with the previously derived fungal mycobiome profiles of the respective participants from the CAMEB cohort was performed using SNF (that is, the R packages SNFtool and vegan), in an analysis involving all three datasets<sup>8,46,47</sup>. For each individual biome dataset, microbes prevalent in at least 5% of patients (that is,  $n \geq 10$ ) were considered for the integration, as follows: bacteria, 62 genera; fungi, 54 genera; and virus, four species. A Bray–Curtis similarity matrix was created for each dataset using the vegan package and subsequently integrated into a single similarity network using the SNFtool package, with implementation of spectral clustering to determine assignment of the integrated network into distinct patient groups<sup>47</sup>. The optimal number of clusters was determined using the eigengap method, and the value of  $K$  nearest neighbors was set based on the optimal silhouette width.

*Weighted SNF.* The WSNF algorithm is an extension of the SNF analytical pipeline described above and was developed for this study (Fig. 1b). In brief, if an individual biome demonstrates more taxa (for example, bacteria  $>$  viruses), there is a greater likelihood of influence from that biome on the overall multi-biome, an issue unaccounted for in standard (unweighted) SNF. Therefore, bacterial and viral profiles, for instance, cannot be considered equal (as is the case in conventional unweighted SNF) because the bacterial community has greater taxonomic richness (and therefore greater information). It is therefore appropriate to weight each biome during data integration based on taxonomic richness. For this study, the respective weights of each biome are assigned based on the richness of the data, as demonstrated by the number of genera (for bacteria and fungi) or species (for viruses) present in each individual biome at a minimum 5% prevalence. Similarly, a Bray–Curtis similarity matrix was created for each biome dataset using the vegan package, which was subsequently integrated using the WSNF analysis pipeline<sup>46</sup>. The optimal number of clusters was determined using the eigengap method and the value of  $K$  nearest neighbors was set based on the optimal silhouette width.

The codebase describing the modified WSNF algorithm can be found at [https://github.com/translational-respiratory-lab/The\\_Interactive](https://github.com/translational-respiratory-lab/The_Interactive).

The robustness of our identified clusters was assessed using a bootstrapping approach, with 70% of the integrated data being sampled in 100 bootstrap iterations, followed by spectral clustering with  $k$  (number of clusters) = 2 (the optimal  $k$  value obtained using all of the data) on this 70% bootstrap sample. The resulting bootstrap clusters (subsamples data, 70%) were compared with the original clusters (100%).

**Implementation of an online tool for WSNF analysis.** *Tool description.* To facilitate replication of our work and enable reproducible implementation of the WSNF analysis, an online webtool was developed (freely accessible at <https://integrative-microbiomics.ntu.edu.sg>). Given  $n$  microbiome datasets (views), the tool generates a corresponding patient (or sample) similarity network for each view based on the user-specified similarity measure before merging them using the user-specified algorithm. Furthermore, the tool then implements a spectral clustering algorithm to enable cluster analysis on the merged dataset, outputting the cluster assignments for each sample (or patient). The default optimum number of clusters is computed using ensemble-based voting for three different methodologies: best eigengap, rotation cost and average silhouette method. For a given value of  $k$  (the number of clusters), we calculate a score (or vote) using the following rules:

1. Average silhouette score > 0.7; Score = Score + 3
2.  $0.5 < \text{Average silhouette score} < 0.7$ ; Score = Score + 2
3.  $0.3 < \text{Average silhouette score} < 0.5$ ; Score = Score + 1
4.  $k$  equals the best value as derived from eigengap method; Score = Score + 3
5.  $k$  equals the 2nd-best value as derived from eigengap method; Score = Score + 2
6.  $k$  equals the best value as derived from rotation cost method; Score = Score + 3
7.  $k$  equals the 2nd-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 analyses such as label propagation and survival analysis<sup>47</sup>. The tool provides the options for choosing between four similarity measures (Bray–Curtis, Gower, Canberra and Jaccard), each appropriate for microbiome datasets used to construct a patient or sample similarity network, with either SNF or WSNF to integrate these networks.

For the implementation of WSNF, the following formula in SNF

$$P^{(\nu)} = S^{(\nu)} \times \frac{\sum_{k \neq \nu} P^k}{(m-1)} \times \left( S^{(\nu)} \right)^T, \nu = 1, 2, 3, \dots, m$$

was modified to

$$P^{(\nu)} = S^{(\nu)} \times \frac{\sum_{k \neq \nu} \omega_k \times P^{(k)}}{\sum_{k \neq \nu} \omega_k} \times \left( S^{(\nu)} \right)^T, \nu = 1, 2, 3, \dots, m$$

where  $\omega_k$  is the weight of the  $k$  dataset,  $m$  is the total number of views,  $P$  is the status matrix and  $S$  is the kernel matrix as defined by Wang et al.<sup>47</sup>. The tool assumes that each input microbiome dataset represents a view of an underlying biological system or disease. Reliable estimation of each view is assumed when using SNF<sup>48</sup>. However, it may not always be 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 present technologies and reference databases. In such cases, a WSNF approach is preferred, which still assumes that the input datasets share an underlying biological mechanism, but uses the user-specified weights to account for varying reliability in the microbiome data. The default weights are assigned based on the taxonomical richness of the datasets (that is, the number of microbes present). The user can also specify custom weights based on other considerations. The interface of the webtool was developed using Rshiny and is available through Shiny Server (open source) in conjunction with nginx-1.19.1, and is provided under a 3-clause BSD license. The tool is powered by custom scripts written in python 2.7 and R, and is containerized using Docker for ease of offline implementation ([https://hub.docker.com/r/jayanthkumar/integrative\\_microbiomics](https://hub.docker.com/r/jayanthkumar/integrative_microbiomics)).

**Testing and appraisal of the webtool using publicly available data.** In addition to the analysis of the study data, to assess the function and utility of our WSNF webtool, we applied it to two additional, independent, publicly available datasets that assessed corresponding bacterial and fungal profiles.

The first dataset was for oral lichen planus.<sup>49</sup> The data encompass salivary microbiome datasets for bacteria and fungi isolated from patients with oral lichen planus and control subjects. All of the paired-end fastq files containing ITS2 and 16S rRNA sequences and the accompanying metadata of the saliva samples under accession number SRP067603 were retrieved from the NCBI SRA. Pre-processing (filtering, trimming, dereplication, merging of paired reads, removal of primers

and chimeras) and taxonomy profiling (using the UNITE 02.02.2019 release for ITS2 and Silva v132 for 16S rRNA) were carried out using the DADA2 package<sup>50</sup>. The resulting datasets for the 52 samples were integrated using the integrative microbiomics webtool with  $K$  (nearest neighbors) = 6 and method = 'SNF'. Cluster consistency in each case was assessed using Average Silhouette score. Superior clustering was noted for our integrative microbiomics approach. The ability of the SNF tools to predict associated phenotypes was also investigated, which confirmed the ability of SNF to predict clinical status over and above the assessment of bacteria and fungi either alone or when concatenated without application of network fusion. Although inferior (albeit still significant) at predicting clinical class, the SNF approach highlighted differences in interleukin-17 and interleukin-23 not apparent on single biome clustering (Supplementary Table 12).

The second dataset was for soil ecosystems.<sup>51</sup> The data relate to taxonomic profiling of experimentally manipulated grassland ecosystems. Bacterial and fungal profiles were determined using 16S and ITS for 48 samples. Unsupervised clustering of the data was performed by inputting the described bacterial and fungal operational taxonomic unit (OTU) table into the webtool and evaluating clusters based on the available metadata downloaded from the respective sources. The dataset was integrated using the integrative microbiomics webtool with  $K=3$  and method = 'SNF'. Cluster consistency in each case was assessed using Average Silhouette score. The metadata assessed included the following soil variables: Decom (percentage decomposition of leaf litter), N2O (N<sub>2</sub>O emission from soil), grNt (total N in grasses), herbNt (total nitrogen in herbs), legNt (total nitrogen in legumes), grPt (total phosphorus in grasses), herbPt (total phosphorus in herbs), legPt (total phosphorus in legumes), Pleach (total P leached per 50 ml), Nleach (total N leached per 50 ml), aveMF (averaged multifunctionality, z-score average of all of the above response data) and pcaMF (multivariate multifunctionality, summed weighted PCA scores using all of the above response data) (Supplementary Table 13).

The integrative microbiomics analysis executed by the SNF webtool had similar or improved performance to that of individual or concatenated analysis of bacterial and fungal datasets with respect to clustering and association with soil variables, most noticeably for the generalized variables aveMF and pcaMF. The SNF webtool analysis significantly outperformed other analytical approaches in terms of defining microbe-derived clusters associated with environmental variation.

#### Co-occurrence analysis of microbial interaction within bronchiectasis patient clusters.

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<sup>52</sup>. To address this, Faust et al. developed a novel bootstrap and renormalization (Reboot) approach that mitigates these potential issues by calculating statistical significance thresholds that account for similarity due to pure compositionality<sup>16</sup>. A weighted ensemble-based co-occurrence analysis along with Reboot was implemented to identify the microbial association networks. Co-occurrence analysis was implemented with statistical significance testing using Reboot as described in Faust et al.<sup>16</sup>, with the following modifications. First, we included the implementation of Mutual Information as a similarity measure instead of Kullback–Leibler divergence. Second, we implemented a Mann–Whitney  $U$ -test, instead of a  $Z$ -test, for the pooled variance to compare between the null and bootstrap distributions, given that the distributions are not necessarily normal. Third, we merged the edge  $P$  values from the ensemble networks in a weighted fashion using the weighted Simes test. Last, the network edge scores were merged as a weighted aggregate of the normalized (calculated as a percentage of the maximum) absolute edge scores and the sign assigned based on GBLM (generalized boosted linear models), Spearman and Pearson correlation (more details described in the GitHub repository). This is important because an imbalance in the ensemble method can suppress actual signals and amplify errors. Finally, abundance and prevalence filters were applied, and only microbes at greater than 1% abundance in at least 5% of subjects were retained. These filters are applied to remove interactions resulting from random noise at the expense of sensitivity to weak signals. The stability of longitudinal multi-biomes was assessed by comparison of  $\alpha$ - and  $\beta$ -diversity for individual patients across the three assessed timepoints (baseline, exacerbation and post-exacerbation) using Kruskal–Wallis and ADONIS (permutational multivariate analysis of variance using distance matrices) testing, respectively. To study the stability of interactions longitudinally across the three timepoints (that is, baseline, exacerbation and post-exacerbation), the relative change in strength of an interaction (defined as maximal interaction strength minus minimal interaction strength) across timepoints was assessed (differential network analysis). Relative interaction change is plotted in Fig. 4g by comparing the change occurring between the baseline and exacerbation (B versus E) timepoints and between the exacerbation and post-exacerbation (E versus P) timepoints. Pairwise matrices indicate the comparative change in interaction observed between individual bacteria, fungi or viruses. A relative change in strength of interaction of  $<3$  was considered non-differential and reflective of a stable (core) interactome. The codebase to implement the co-occurrence analysis and differential network analysis can be found at [https://github.com/translational-respiratory-lab/The\\_Interactive](https://github.com/translational-respiratory-lab/The_Interactive).

**Predictive modeling of network response to antibiotic therapy.** To model the predicted impact of antibiotic therapy on the interactome, we simulated the effect of antibiotic treatment on the observed microbial networks. Antibiotic therapy exerts a substantial effect on the microbiome, and reduces the relative abundance of susceptible microbes<sup>53</sup>. In modeling network changes, antibiotic action was thus simulated by assuming a substantial reduction (that is, 75%; but not complete elimination) in the relative abundance of microbes targeted by  $\beta$ -lactam antibiotics (without documented intrinsic resistance) in the pre-antibiotic state for the following genera: *Streptococcus*, *Staphylococcus*, *Haemophilus*, *Moraxella*, *Actinomyces*, *Arachnia*, *Bacteroides*, *Bifidobacterium*, *Eubacterium*, *Fusobacterium*, *Lactobacillus*, *Leptotrichia*, *Peptococcus*, *Peptostreptococcus*, *Propionibacterium*, *Selenomonas*, *Treponema* and *Veillonella*. The interaction networks of pre-antibiotic state (baseline), simulated and observed antibiotic impact were determined by assessment of the interactome networks from patients in the longitudinal study arm who received  $\beta$ -lactam antibiotics ( $n = 12$ ; patient demographics are listed in Supplementary Table 6). Co-occurrence analysis to derive these three networks was performed to assess the utility of network analysis in the modeling and prediction of antibiotic impact on the interactome. To remove interactions resulting from random noise while maintaining sensitivity to weak signals and to enable comparison between the derived interactomes, abundance and prevalence filters were applied. Data were filtered for retention of microbes present at greater than 1% abundance in each of at least three subjects, in the pre-antibiotic or post-antibiotic or modeled antibiotic state. The modeled interactome was validated by comparing the network metrics and the rewiring of the pre-antibiotic interactome, between the modeled interactome and the post-antibiotic interactome. Network metrics such as node degree (busy), stress centrality (critical) and betweenness centrality (influential) were calculated and visualized in cytoscape<sup>54</sup>.

**Predictive modeling of time to next exacerbation.** Shannon diversity was assessed based on renormalized concatenated microbiome (bacteria, fungi and virus) datasets at the baseline, exacerbation and post-exacerbation timepoints. Distribution differences in Shannon diversity between two groups (time to exacerbation, <12 weeks and >12 weeks) was evaluated using Mann–Whitney *U*-test. Microbiome datasets were CLR (centered log ratio) transformed before the concatenation of microbes present in at least four patients at an abundance of 1%. Correlation between the abundance of each microbe and time to next exacerbation was assessed using Spearman's rank correlation with statistical testing. Multivariate adaptive regression spline (MARS), a non-linear regression model was implemented with microbes as the predictor variable to predict time to next exacerbation<sup>55</sup>. For interaction-based analysis, LIONESS, a single patient network inference framework, was implemented with GBLM as the network inference algorithm<sup>56</sup>. Correlation between the interaction strength of each pair of microbes (edge) and time to next exacerbation was assessed using Spearman's rank correlation with statistical testing. Model goodness of fit was evaluated by computing the R-squared (RSq) and the generalized R-squared (GRSq) metrics. A feature importance plot based on the generalized cross-validation (gcv) score was also constructed for the selected features 'microbes and interactions'. All of the above analyses were implemented in R using the packages Hmisc, earth, vegan, compositions and lionessR.

**Metagenomic analysis.** *Metagenomic sequencing.* The integrity of extracted sputum DNA was confi med using the Qubit dsDNA HS Assay Kit (Invitrogen) and sequenced on a HiSeq 2500 platform (Illumina) at the Nanyang Technological University core sequencing facility according to library preparation and DNA sequencing methods described previously<sup>57</sup>. Read counts in blanks as well as their taxonomic assignments are given in Extended Data Fig. 10. Patient demographics are listed in Supplementary Table 8 (preliminary functional analysis) and Supplementary Table 9 (validation).

*Quality control and manipulation of metagenomic sequencing data.* Trimmomatic (v0.39) (with parameters ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:1:TRUE, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, MINLEN:30) was used to clip Illumina adapters from the paired-end raw sequencing reads, remove low-quality bases at both ends, and drop reads less than 30 base pairs long after trimming<sup>58</sup>. FastQC (v0.11.8) was used to check the quality of the original sequencing reads as well as the trimmed ones, while bowtie2 (v2.3.5.1) was used subsequently to align the quality-filtered reads to the human reference genome (hg38)<sup>59</sup>. Unmapped, non-human reads were separated and sorted from the human reads using samtools (v1.9)<sup>60</sup>. FaSta files containing these reads, which were used for various downstream analyses, were obtained using bedtools (v2.28) and the R Bioconductor package ShortRead (v3.6.2)<sup>61,62</sup>.

*Functional analysis.* Read processing, quality assessment and functional annotation of derived metagenome sequences were performed using the MG-RAST annotation pipeline (v4.0.3) with reference to the Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>63,64</sup>.

*Taxonomic analysis.* Kaiju (v1.7.2) with default parameters was used to estimate the taxonomic composition in sputum. The NCBI BLAST nr+euk database (accessed

23 July 2019) for non-redundant (nr) proteins belonging to Archaea, Bacteria and Viruses, plus Fungi and microbial eukaryotes (euk) was used as a reference.

*Assessment of the virome.* The composition of the virome was assessed by performing metagenomic assembly using Megahit (v1.2.9), VirFinder (v1.1), bowtie2 (v2.3.3.1), samtools (v1.3.1), CONCOCT (v1.1.0) and Demovir (<https://github.com/feargalr/Demovir>)<sup>59,60,65–68</sup>. Megahit was initially used to cross-assemble all of the non-human reads into contigs, with contigs larger than 1 kilobase in length considered in the subsequent analysis. Contigs were classified as a viral or non-viral contig using VirFinder, trained on an in-built viral database. The non-human reads of each sample were then mapped back to the assembly viral contigs using bowtie2 and samtools to assess viral abundance, while CONCOCT was implemented with default parameters to generate the final coverage table containing the number of raw hits (*H*) to a specific viral contig. These numbers were normalized as described previously<sup>69</sup>, that is, by adjusting for read length (*R*), contig length (*L*) and non-human read depth (*N*) according to the following formula:

$$\frac{H}{\left(\frac{L-R}{10^3}\right) \left(\frac{N}{10^6}\right)} = \frac{H}{|L-R| \cdot N} \times 10^9$$

where the unit of normalized read counts (or the relative abundance) is given in reads per kilobase of reference sequence per million sample reads (RPKM). In parallel, Demovir script with its default parameters and database was used to assign the taxonomic rank of the viral contigs at the order and family levels, and blastn was used to search and align the viral contigs against data from the NCBI viral genomes resource (taxid 10239) to find the best matches.

**Statistical analysis and data visualization.** The Shapiro–Wilk normality test was used to examine data distributions. For continuous variables, statistical significance was determined using the Mann–Whitney or the Kruskal–Wallis test, with Dunn's test for post-hoc analysis and Benjamini–Hochberg correction when more than two groups were present. For categorical variables, Pearson's chi-squared test or Fisher's exact test (as appropriate) was implemented, with Bonferroni correction for multiple comparisons when contingency extended beyond 2 × 2 (ref. <sup>70</sup>). Relative risk was calculated using the R package, epitools<sup>71</sup>. In all of the cases, two-tailed analysis was considered, and differences were deemed significant at  $P < 0.05$ . Analysis was performed using R Statistical Software (v3.2.4 and v3.5.1), with the ggplot2 package for visualization<sup>72</sup>. Further information on study design and statistical information can be found in the online Nature Research Reporting Summary associated with this paper. For analysis of the microbiome, linear discriminant analysis effect size was implemented using the webtool available at <http://huttenhower.sph.harvard.edu/galaxy/> (ref. <sup>73</sup>). Co-occurrence networks were visualized and annotated in Cytoscape<sup>54</sup>. A Venn diagram summarizing differences in microbial interactions present at different longitudinal timepoints was plotted using the R package VennDiagram. The diffany application was used to assess for differential changes in interaction strengths across the measured longitudinal timepoints, and between the metagenomic (SC2) and targeted amplicon-derived (HEF) interactomes<sup>74</sup>.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All of the sequence data described in this study have been uploaded to the NCBI SRA under project accession PRJNA590225. Publicly available taxonomic and functional databases are referenced by short-read sequence classification tools used in this study as further described in the Nature Research Reporting Summary. Other associated data, including bacterial, fungal and viral profiles for all of the patients, as well as patient clinical attributes, are available at [https://github.com/translational-respiratory-lab/The\\_Interactome/tree/master/Data](https://github.com/translational-respiratory-lab/The_Interactome/tree/master/Data).

## Code availability

All code required for generation of the presented results, with accompanying documentation, are available at the study's online code repository ([https://github.com/translational-respiratory-lab/The\\_Interactome](https://github.com/translational-respiratory-lab/The_Interactome)).

## References

- Chalmers, J. D. et al. The bronchiectasis severity index. An international derivation and validation study. *Am. J. Respir. Crit. Care Med.* **189**, 576–585 (2014).
- Bateman, E. D. et al. Global strategy for asthma management and prevention: GINA executive summary. *Eur. Respir. J.* **31**, 143–178 (2008).
- Celli, B. R., MacNee, W. & Force, A. E. T. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur. Respir. J.* **23**, 932–946 (2004).
- Hill, A. T. et al. Pulmonary exacerbation in adults with bronchiectasis: a consensus definition for clinical research. *Eur. Respir. J.* **49**, 1700051 (2017).

37. Chotirmall, S. H. et al. Sputum *Candida albicans* presages FEV(1) decline and hospital-treated exacerbations in cystic fibrosis. *Chest* **138**, 1186–1195 (2010).
38. Coughlan, C. A. et al. The effect of *Aspergillus fumigatus* infection on vitamin D receptor expression in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **186**, 999–1007 (2012).
39. Ong, S. H. et al. Species identification and profiling of complex microbial communities using shotgun Illumina sequencing of 16S rRNA amplicon sequences. *PLoS One* **8**, e60811 (2013).
40. Hahn, A. et al. Different next generation sequencing platforms produce different microbial profiles and diversity in cystic fibrosis sputum. *J. Microbiol. Methods* **130**, 95–99 (2016).
41. Ward, T. L. et al. Development of the human mycobiome over the first month of life and across body sites. *mSystems* **3**, e00140-17 (2018).
42. Klindworth, A. et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, e1 (2013).
43. DeSantis, T. Z. et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
44. Koljalg, U. et al. Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* **22**, 5271–5277 (2013).
45. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
46. Oksanen, J. et al. Vegan: Community Ecology Package. R package version 2.0-7 (2013).
47. Wang, B. et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* **11**, 333–337 (2014).
48. Jiang, D. et al. Microbiome multi-omics network analysis: statistical considerations, limitations, and opportunities. *Front. Genet.* **10**, 995 (2019).
49. Li, Y. et al. Salivary mycobiome dysbiosis and its potential impact on bacteriome shifts and host immunity in oral lichen planus. *Int J. Oral Sci.* **11**, 13 (2019).
50. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
51. Wagg, C., Schlaeppi, K., Banerjee, S., Kuramae, E. E. & van der Heijden, M. G. A. Fungal–bacterial diversity and microbiome complexity predict ecosystem functioning. *Nat. Commun.* **10**, 4841 (2019).
52. Aitchison, J. The statistical analysis of compositional data. *J. R. Stat. Soc. B* **44**, 139–177 (1982).
53. Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl Acad. Sci. USA* **108**, 4554–4561 (2011).
54. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
55. Friedman, J. H. Multivariate adaptive regression splines. *Ann. Stat.* **19**, 1–67 (1991).
56. Kuijjer, M. L., Hsieh, P. H., Quackenbush, J. & Glass, K. lionessR: single sample network inference in R. *BMC Cancer* **19**, 1003 (2019).
57. Gusareva, E. S. et al. Microbial communities in the tropical air ecosystem follow a precise diel cycle. *Proc. Natl Acad. Sci. USA* **116**, 23299–23308 (2019).
58. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
59. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
60. Li, H. et al. The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
61. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
62. Morgan, M. et al. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* **25**, 2607–2608 (2009).
63. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
64. Meyer, F. et al. The metagenomics RAST server: a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**, 386 (2008).
65. Shkoporov, A. N. et al. The human gut virome is highly diverse, stable, and individual specific. *Cell Host Microbe* **26**, 527–541 (2019).
66. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
67. Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A. & Sun, F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* **5**, 69 (2017).
68. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
69. Tiew, P. Y. et al. Environmental fungal sensitisation associates with poorer clinical outcomes in COPD. *Eur. Respir. J.* **56**, 2000418 (2020).
70. MacDonald, P. L. & Gardner, R. C. Type I error rate comparisons of post hoc procedures for  $1 \times J$  chi-square tables. *Educ. Psychol. Meas.* **60**, 735–754 (2000).
71. Aragon, T.J., Fay, M.P. & Omidpanah, M.A. Package ‘epitools’ (2017).
72. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* 2nd edn (Springer, 2016).
73. Segata, N. et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
74. Van Landeghem, S., Van Parys, T., Dubois, M., Inze, D. & Van de Peer, Y. DiffNy: an ontology-driven framework to infer, visualise and analyse differential molecular networks. *BMC Bioinformatics* **17**, 18 (2016).

## Acknowledgements

This research is supported by the Singapore Ministry of Health’s National Medical Research Council under its Transition Award (NMRC/TA/0048/2016 to S.H.C.); the Clinician-Scientist Individual Research Grant (MOH-000141 to S.H.C.); NTU Integrated Medical, Biological and Environmental Life Sciences (NIMBELS) (NIM/03/2018 to S.H.C.); the British Lung Foundation through the GSK/British Lung Foundation Chair of Respiratory Research (to J.D.C.) and the Scottish Government Chief Scientist Office (SCAF/17/03). K.T.-A. gratefully acknowledges the financial support of the EPSRC via grant EP/N014391/1.

## Author contributions

M.M.A. and J.K.N. designed and performed the experiments, carried out data analysis and interpretation, statistical analysis, and wrote the manuscript. P.Y.T., A.Y.H.L., M.E.P., T.B.L., T.H.O., A. Tee, M.S.K. and J.A.A. carried out patient recruitment and collected clinical data and specimens for the Asian patients. P.Y.T., N.A.B.M.A., K.X.T., A. Tsang, Y.H.G., V.F.L.Y. and T.K.J. performed the experiments. F.X.I. provided bioinformatic support. H.R.K., A.J.D. and J.D.C. designed the overall study, interpreted the data, carried out patient recruitment and collected clinical data and specimens for the European patients and the longitudinal cohort (Scotland), and procured funding. M.O., F.B. and S.A. carried out patient recruitment and collected the clinical data and specimens for the European patients (Italy). B.O., K.T.-A., J.D.C. and S.H.C. made intellectual contributions. K.T.-A. carried out data analysis. S.H.C. conceived and designed the overall study and experiments, carried out data analysis and interpretation, statistical analysis, wrote the manuscript, and procured funding.

## Competing interests

J.D.C. has received research grants from GlaxoSmithKline, Boehringer-Ingelheim, AstraZeneca, Gilead Sciences, Grifols and Insmad, and has received personal fees from GlaxoSmithKline, Boehringer-Ingelheim, AstraZeneca, Chiesi, Grifols, Napp, Novartis, Insmad and Zambon, all outside of the submitted work. S.A. has received personal fees from Bayer Healthcare, Grifols, AstraZeneca and Zambon, and research grants and personal fees from Chiesi, Insmad, GlaxoSmithKline, Menarini, ZetaCube Srl and Fisher & Paykel, all outside of the submitted work. F.B. has received research grants and personal fees from AstraZeneca, Bayer, Chiesi, GlaxoSmithKline, Guidotti, Grifols, Insmad, Menarini, Novartis, Pfizer and Zambon, all outside of the submitted work. All other authors have no competing interests.

## Additional information

Extended data are available for this paper at <https://doi.org/10.1038/s41591-021-01289-7>.

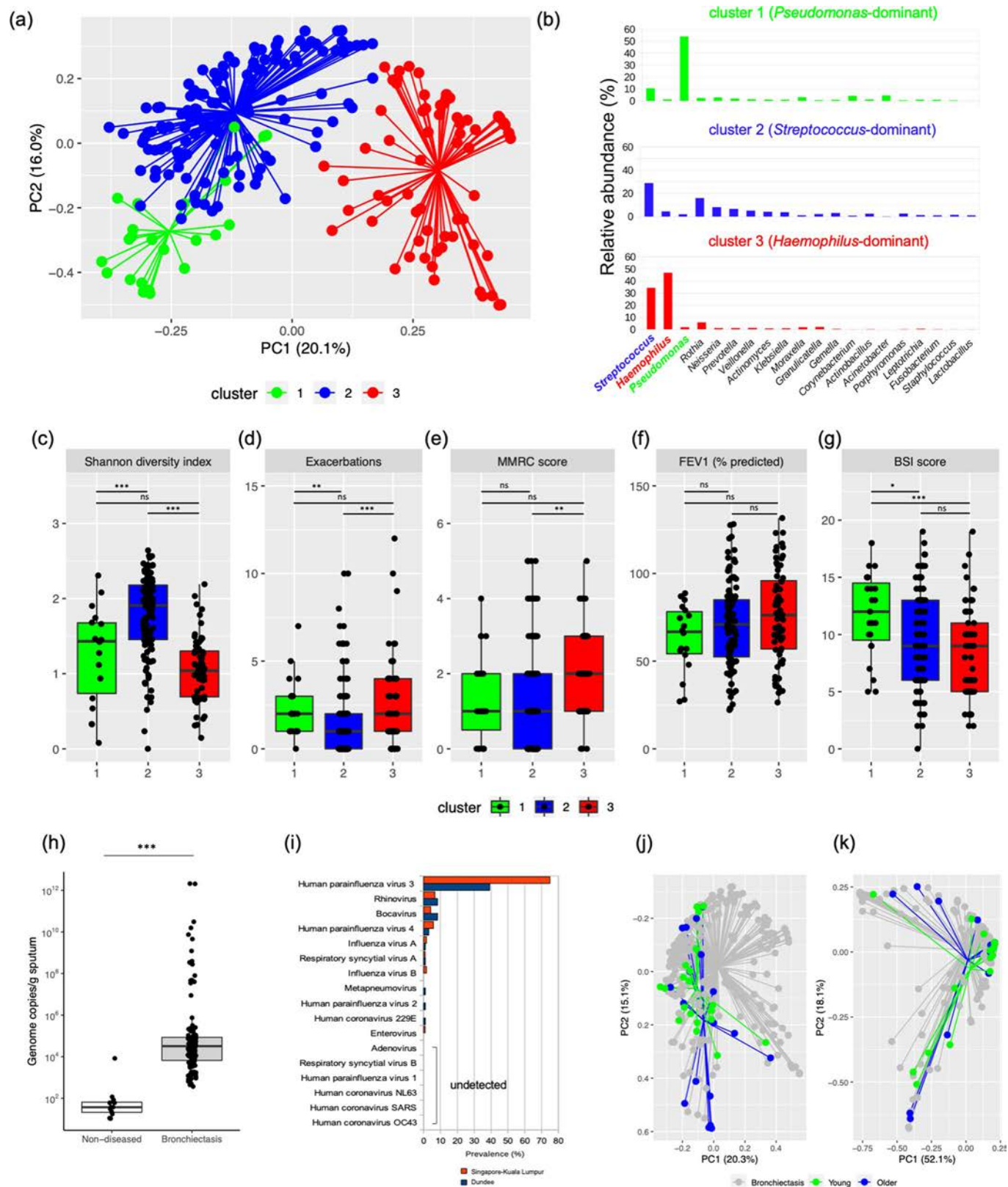
**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41591-021-01289-7>.

**Correspondence and requests for materials** should be addressed to S.H.C.

**Peer review information** *Nature Medicine* thanks Grant Waterer, Brian Keller and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Jennifer Sargent was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

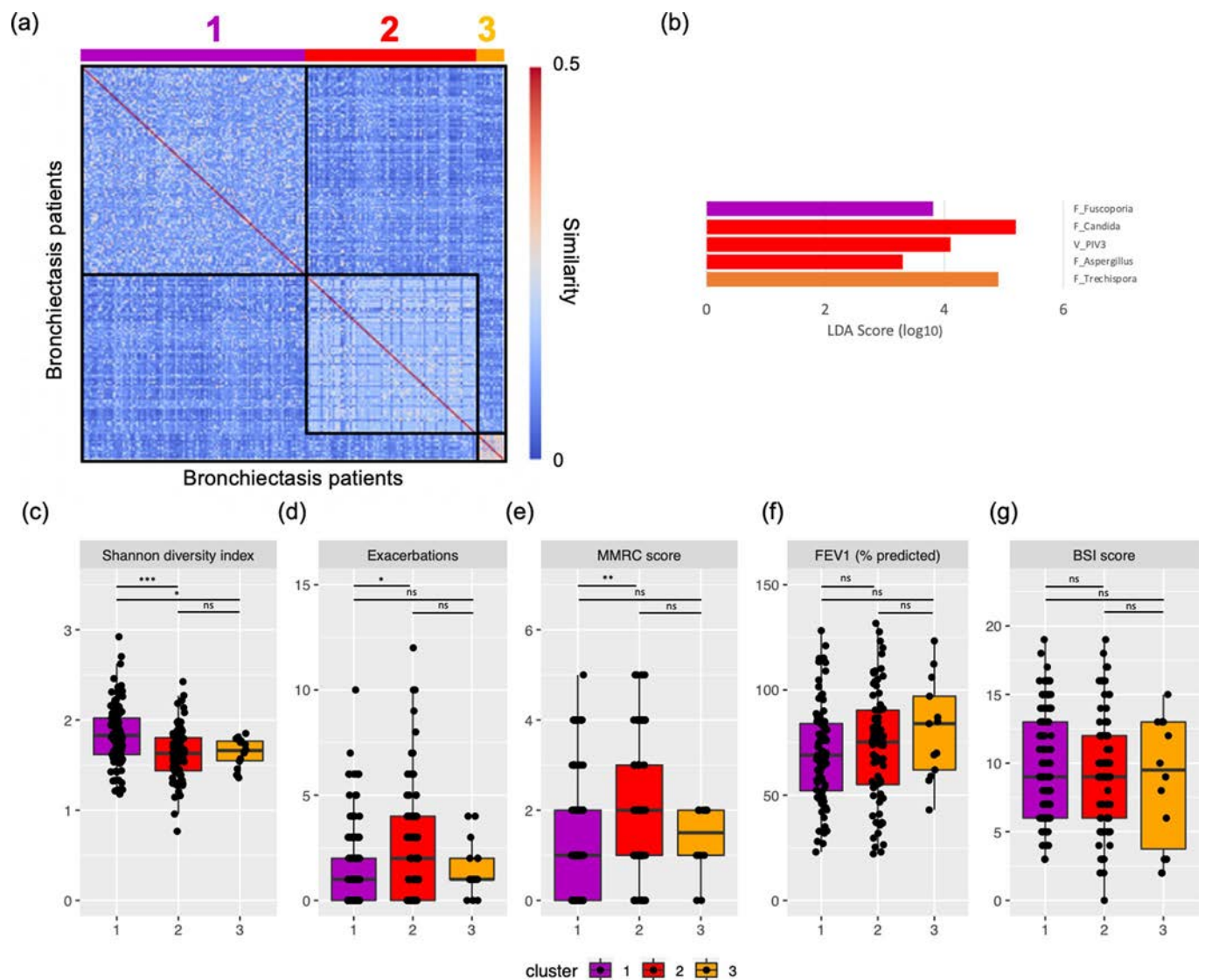
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).





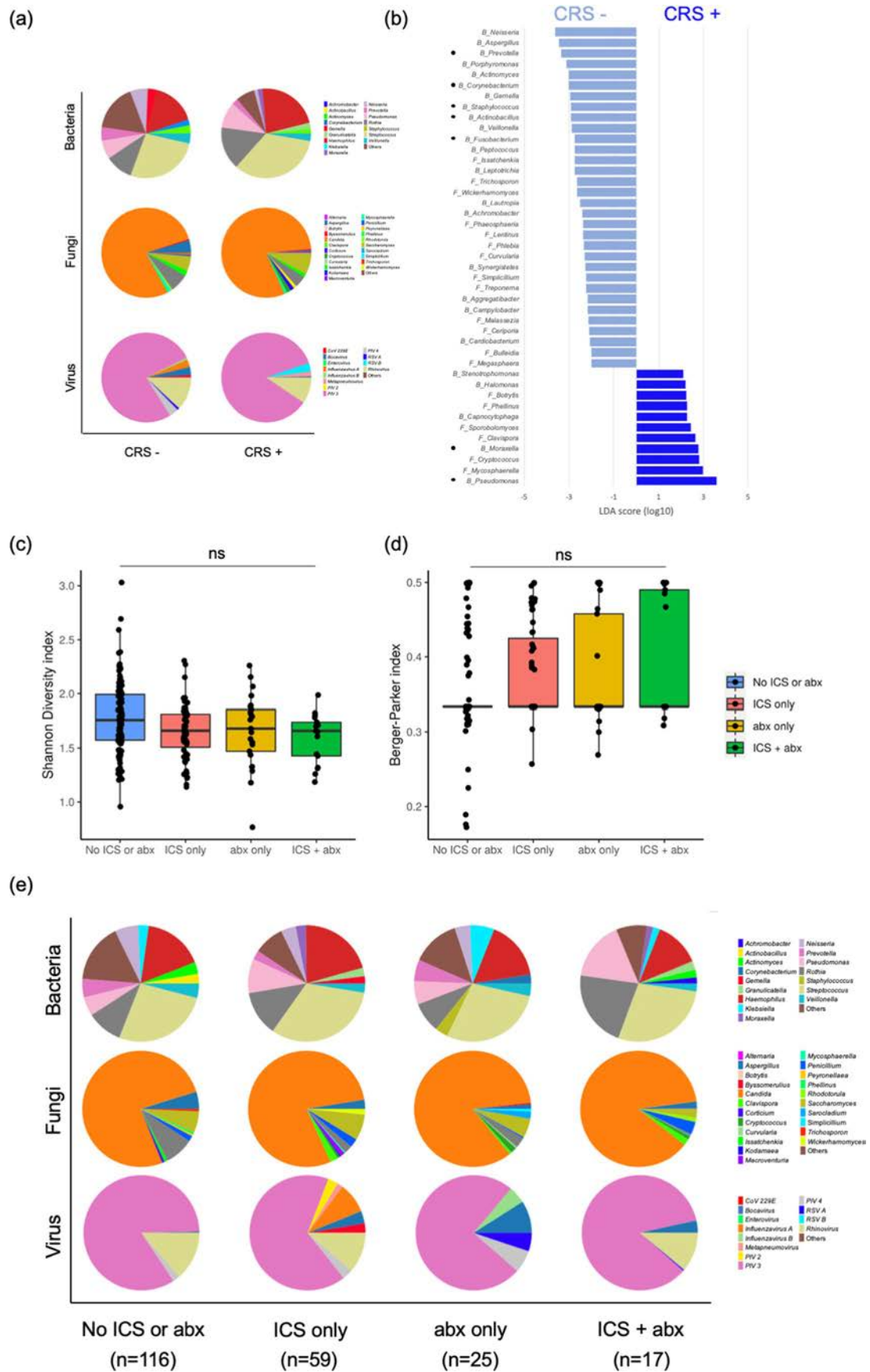
Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | Sub-analysis of discrete bacterial, viral and fungal profiles in stable-state bronchiectasis.** **a**, A Principal Coordinate Analysis (PCoA) of bacterial microbiome based on Bray-Curtis dissimilarity illustrates three patient clusters. Colour indicates groupings defined by spectral clustering analysis. **b**, The bacterial community profile of each identified cluster is illustrated as bar plots of average relative abundance for the top 20 identified genera; cluster 1: *Pseudomonas*-dominant (green), cluster 2: *Streptococcus*-dominant (blue), cluster 3: *Haemophilus*-dominant (red). Comparison of alpha diversity and clinical features among identified patient clusters (cluster 1; n = 23, cluster 2; n = 125, cluster 3, n = 69), derived from n = 217 biologically independent samples are illustrated (**c**) alpha diversity (Shannon diversity index,  $p = 2.9 \times 10^9$ ), (**d**) median number of exacerbations in the preceding year ( $p = 2.1 \times 10^4$ ), (**e**) breathlessness score (MMRC,  $p = 3.6 \times 10^3$ ), (**f**) lung function (as FEV1 % predicted,  $p = 1.8 \times 10^{-1}$ ) and (**g**) bronchiectasis severity index ( $p = 3.9 \times 10^{-3}$ ) for patients in each cluster. MMRC = modified medical research council, BSI = bronchiectasis severity index; FEV1% = forced expiratory volume in the 1<sup>st</sup> second (% predicted). Box plots reflect median and IQRs with whiskers bounding non-outlier values. Throughout, significance levels are indicated as follows: ns: non-significant; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$  (Kruskal-wallis with post-hoc Dunn test for multiple comparison). **h**, Viral community profiling in bronchiectasis, assessed by a 17-virus qPCR panel illustrates high viral load (in the stable disease state) compared to a non-diseased (healthy) state. Box and whisker plot illustrating overall viral load, expressed as viral genome copies/g sputum, quantified by qPCR in virus-positive sputum samples from non-diseased (healthy) individuals (n = 30) and patients with bronchiectasis (n = 217). **i**, A range of viruses and highest prevalence of human parainfluenza virus 3. Prevalence is expressed as percentage of patients demonstrating PCR-positivity originating from either Singapore-Kuala Lumpur (SG-KL, n = 120, red bars) or Dundee (DD, n = 97, blue bars). **j**, PCoA plot illustrating differences between bacteriome profiles from our stable state bronchiectasis cohort (grey, n = 217) compared to non-diseased (healthy) controls that include young (<40 years of age, green, n = 20) and older (>60 years of age, blue, n = 20) subjects. **k**, PCoA plot illustrating differences between mycobiome profiles from our stable state bronchiectasis cohort (grey, n = 217) compared to non-diseased (healthy) controls that include young (<40 years of age, green, n = 20) and older (>60 years of age, blue, n = 20) subjects.



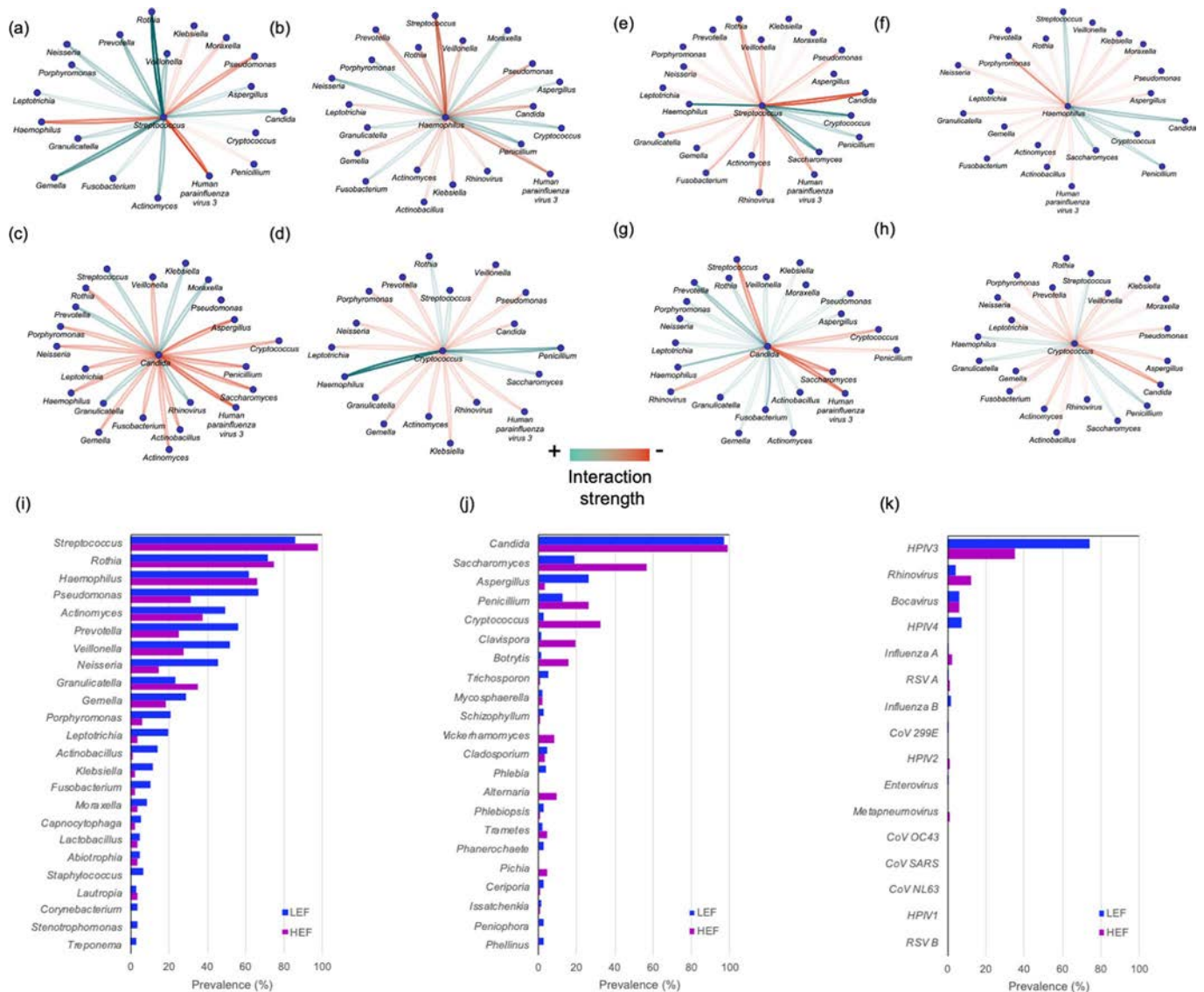
**Extended Data Fig. 2 | Un-weighted SNF analysis of multi-biome data provides inferior resolution of clinically relevant bronchiectasis patient clusters.**

**a**, A heatmap illustrating pairwise patient similarity scores (range; 0–0.5, blue–red) assessed by spectral clustering. Three distinct patient clusters are illustrated by purple, red and orange bars above the heatmap respectively. Clusters are colour-coded as (1) purple; (2) red and (3) orange. **b**, Linear discriminant effect size (LEfSe) analysis of the observed clusters illustrating taxa that discriminate between the multi-biome profiles of each group. A bar plot details each of the identified discriminant taxa ranked by effect size. Discriminant taxa with a log-transformed effect size of  $>3$  are presented ( $n=5$ ). Prefixes indicate whether taxa are bacterial (B), fungal (F) or viral (V). Comparison of alpha diversity and clinical features among identified patient clusters (cluster 1;  $n=115$ , cluster 2;  $n=88$ , cluster 3,  $n=14$ ) according to integrated multi-biome profiles, derived from  $n=217$  biologically independent samples are illustrated; (c) alpha diversity (Shannon diversity index,  $p=3.17 \times 10^{-6}$ ), (d) number of exacerbations in the preceding year,  $p=4.0 \times 10^{-2}$ , (e) breathlessness score (MMRC,  $p=3.0 \times 10^{-3}$ ), (f) lung function (as FEV1 % predicted,  $p=2.5 \times 10^{-1}$ ) and (g) bronchiectasis severity index (BSI,  $p=7.6 \times 10^{-1}$ ) for patients in each cluster. Box plots reflect median and IQRs with whiskers bounding non-outlier values. Significance levels for observed between-group differences are indicated as follows: ns: non-significant; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$  (Kruskal-wallis with post-hoc Dunn test for multiple comparison).

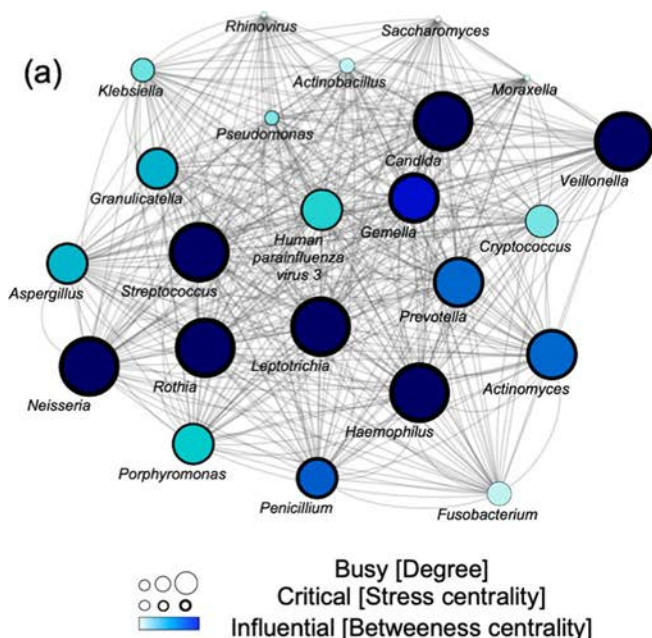


Extended Data Fig. 3 | See next page for caption.

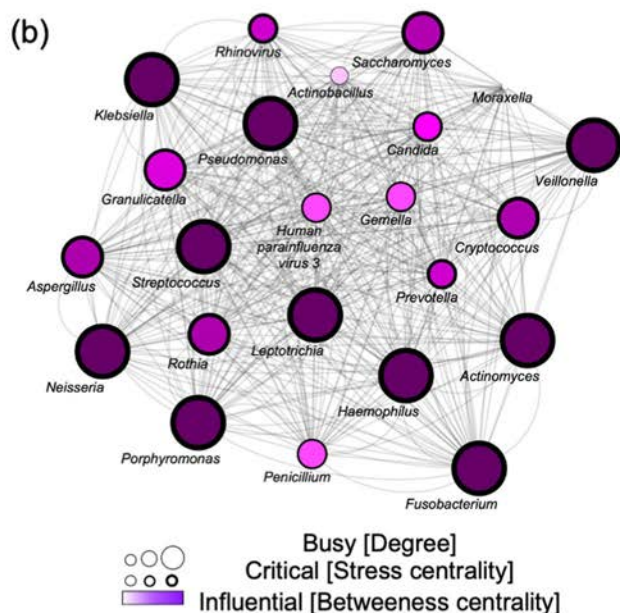
**Extended Data Fig. 3 | Sub-analysis of chronic rhinosinusitis (CRS) and antibiotic treatment effects on the microbiome.** **a**, Pie charts displaying microbiome profiles for bacteria, fungi and viruses for CRS + and CRS - patients with microbial identity indicated in the right-hand colour legend (n = 217 biologically independent samples). **b**, Linear discriminant effect size (LEfSe) analysis of patients with and without CRS illustrating taxa that discriminate between the multi-biome profiles of each group. A bar plot details each of the identified discriminant taxa ranked by effect size. Discriminant taxa with a log-transformed effect size of >2 are presented (n = 43). Prefixes indicate whether identified taxa are bacterial (B), fungal (F) or viral (V - none detected). Black circles indicated taxa previously established to be associated with CRS status. **c**, Box plots reflect median and IQRs with whiskers bounding non-outlier values and illustrate diversity metrics for each of the following groups; no inhaled corticosteroids (ICS) or antibiotic use (No ICS or abx), ICS only (ICS only), antibiotics only (abx only), or both (ICS + abx) indicating Shannon Diversity Index and **(d)** Berger-Parker Index. ns = not significant (Kruskal-wallis with post-hoc Dunn test for multiple comparison). **e**, Pie charts displaying microbiome profiles for bacteria, fungi and viruses for the same patient groups with microbial identity indicated in the right-hand colour legend.



**Extended Data Fig. 4 | Co-occurrence networks and prevalence of observed keystone taxa identified in the low (LEF) and high (HEF) exacerbation frequency clusters.** The interaction profiles of key taxa with other microbes within their network is illustrated for the (a–d) low-exacerbation frequency and (e–h) high-exacerbation frequency clusters respectively. Bacteria and fungi with established pathogenic potential including *Streptococcus*, *Haemophilus*, *Candida* and *Cryptococcus* are illustrated. Though common to both clusters, these ‘busy’, ‘critical’ and ‘influential’ microbes exhibited marked differences in their interaction networks when assessed individually thereby suggesting that microbial interactions, rather than mere presence or absence of a given microbe, may have a direct bearing on the clinical status of patients observed in the high frequency exacerbation cluster. Positive and negative interactions (expressed as edge weights) are colour-coded for illustration as indicated in the figure legend. The prevalence of microbial taxa (detected at >1% relative abundance) among each group is represented by horizontal bars for LEF (Blue) and HEF (purple) clusters. (i), Bacteria, (j) Fungi, (k) Viruses.

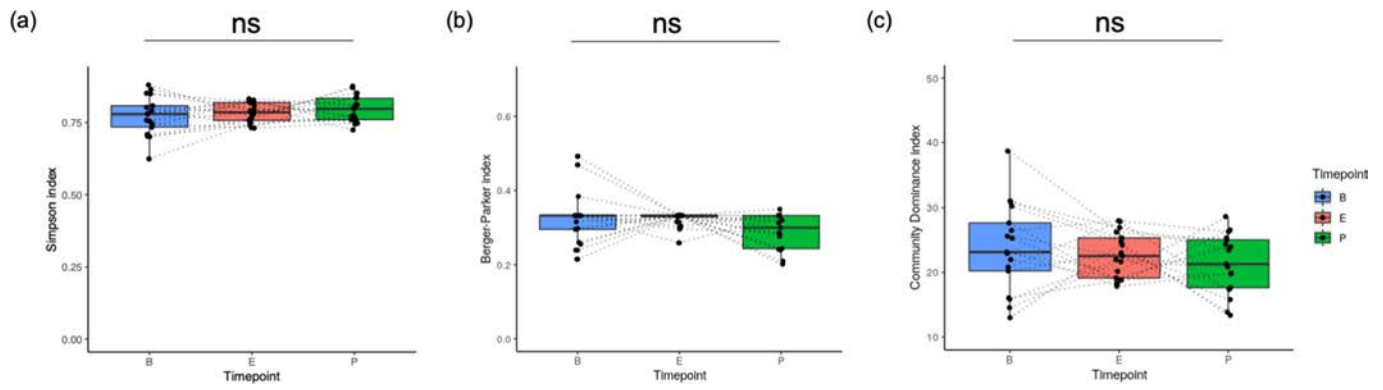


Non-Frequent Exacerbators			
Microbe	Busy (Degree)	Critical (Stress Centrality)	Influential (Betweenness Centrality)
<i>Candida</i>	44	52	0.0077
<b><i>Veillonella</i></b>	44	52	0.0077
<i>Streptococcus</i>	44	52	0.0077
<i>Rothia</i>	44	52	0.0077
<b><i>Neisseria</i></b>	44	52	0.0077
<b><i>Leptotrichia</i></b>	44	52	0.0077
<i>Haemophilus</i>	44	52	0.0077
<i>Gemella</i>	42	48	0.0072
<i>Penicillium</i>	40	40	0.0060
<i>Prevotella</i>	42	40	0.0058
<i>Actinomyces</i>	42	40	0.0058
<i>Granulicatella</i>	40	32	0.0047
<i>Aspergillus</i>	40	32	0.0046
<i>Porphyromonas</i>	40	30	0.0043
<i>PIV3</i>	40	28	0.0039
<i>Klebsiella</i>	36	20	0.0029
<i>Cryptococcus</i>	38	20	0.0028
<i>Pseudomonas</i>	34	18	0.0027
<i>Actinobacillus</i>	34	14	0.0019
<i>Fusobacterium</i>	36	14	0.0019
<i>Moraxella</i>	32	12	0.0017
<i>Rhinovirus</i>	32	12	0.0016
<i>Saccharomyces</i>	32	6	0.0008



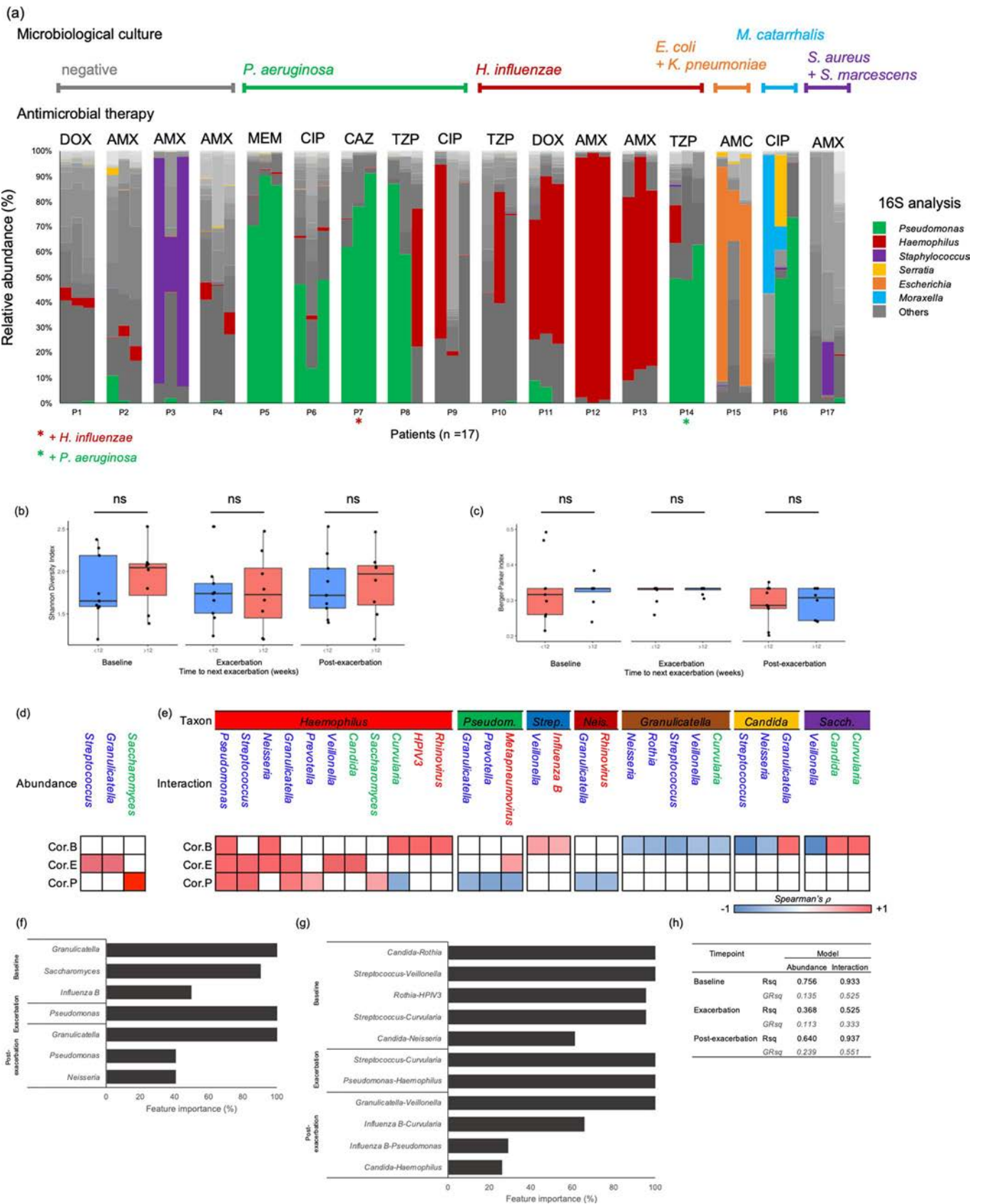
Frequent Exacerbators			
Microbe	Busy (Degree)	Critical (Stress Centrality)	Influential (Betweenness Centrality)
<b><i>Pseudomonas</i></b>	44	24	0.0028
<i>Klebsiella</i>	44	24	0.0028
<i>Veillonella</i>	44	24	0.0028
<i>Streptococcus</i>	44	24	0.0028
<i>Porphyromonas</i>	44	24	0.0028
<i>Neisseria</i>	44	24	0.0028
<i>Leptotrichia</i>	44	24	0.0028
<i>Haemophilus</i>	44	24	0.0028
<i>Fusobacterium</i>	44	24	0.0028
<b><i>Actinomyces</i></b>	44	24	0.0028
<i>Saccharomyces</i>	42	20	0.0023
<i>Cryptococcus</i>	42	20	0.0023
<i>Aspergillus</i>	42	20	0.0023
<i>Rothia</i>	42	20	0.0023
<i>Granulicatella</i>	42	18	0.0020
<i>Rhinovirus</i>	40	18	0.0021
<i>PIV3</i>	40	14	0.0016
<i>Penicillium</i>	40	14	0.0016
<i>Candida</i>	40	16	0.0018
<i>Prevotella</i>	40	18	0.0021
<i>Gemella</i>	40	14	0.0016
<i>Actinobacillus</i>	38	10	0.0011
<i>Moraxella</i>	36	8	0.0009

**Extended Data Fig. 5 | Supervised, non-SNF classification of the microbial interactome based on known exacerbation frequencies.** Network visualization of key microbial taxa in (a) non-frequent (<3 exacerbations/year) and (b) frequent (3 or more exacerbations/yr) are illustrated. Coloured circles represent microbes and grey lines their associated interactions within the network. Taxa present at >1% relative abundance in at least 5% of the patient cohort are included. Circle size (degree) reflects the number of direct interactions for a given microbe (termed 'busy'). Circle outline thickness represents the calculated stress centrality for each microbe termed 'critical' while circle colour depth reflects betweenness centrality or the 'influence' of the microbe within the network. A table to the right of each network illustrates details the composition of busy (degree), critical (stress centrality) and influential (betweenness centrality) microbes observed in non-frequent vs frequent exacerbators. Additional microbes present in the top taxa compared to SNF-based analysis (Fig. 3a,b) are highlighted by bold typeface.



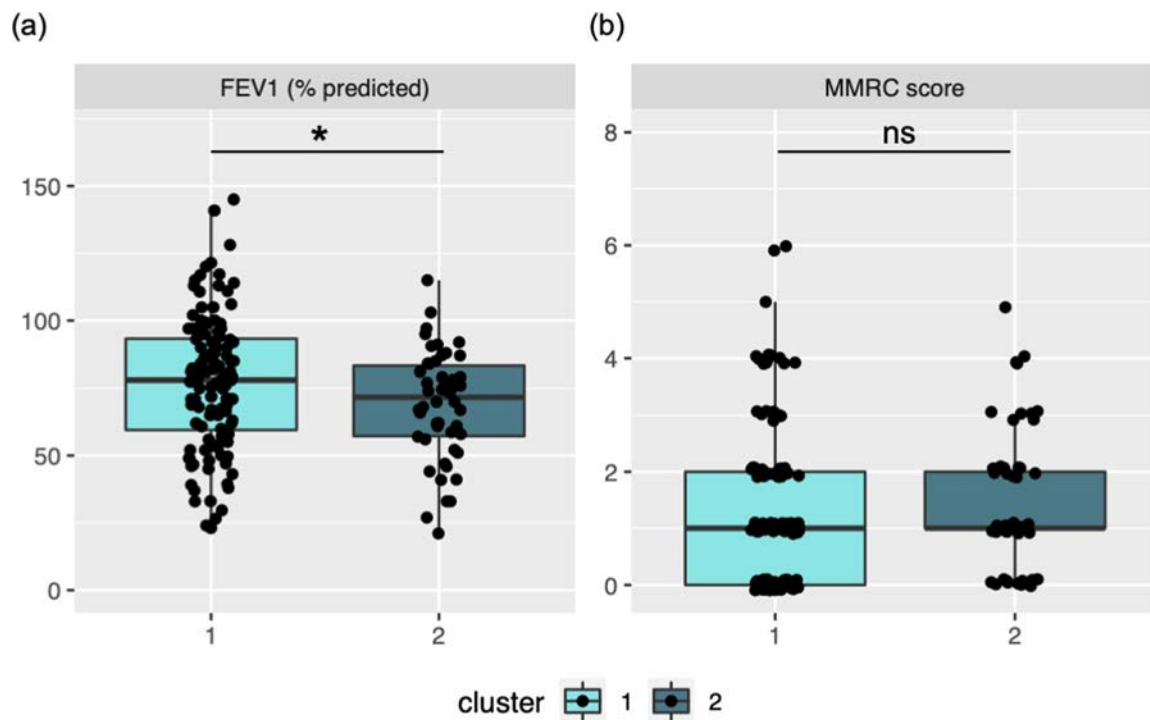
**Extended Data Fig. 6 | Additional diversity indices indicating comparable ecological diversity across longitudinal analysis of the integrated multi-biome during bronchiectasis exacerbations.** Box plots reflect median and IQRs with whiskers bounding non-outlier values and illustrate (a) Simpson (b) Berger-Parker and (c) Community dominance index<sup>1</sup> for combined bacterial, fungal and viral communities 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. Dotted lines indicate the longitudinal pattern of each individual patient ( $n=17$ ). Boxplots are coloured according to their respective longitudinal time-point and timepoints indicated by colour: blue = baseline, red = exacerbation, green = post-exacerbation. ns; not significant (Kruskal-Wallis).



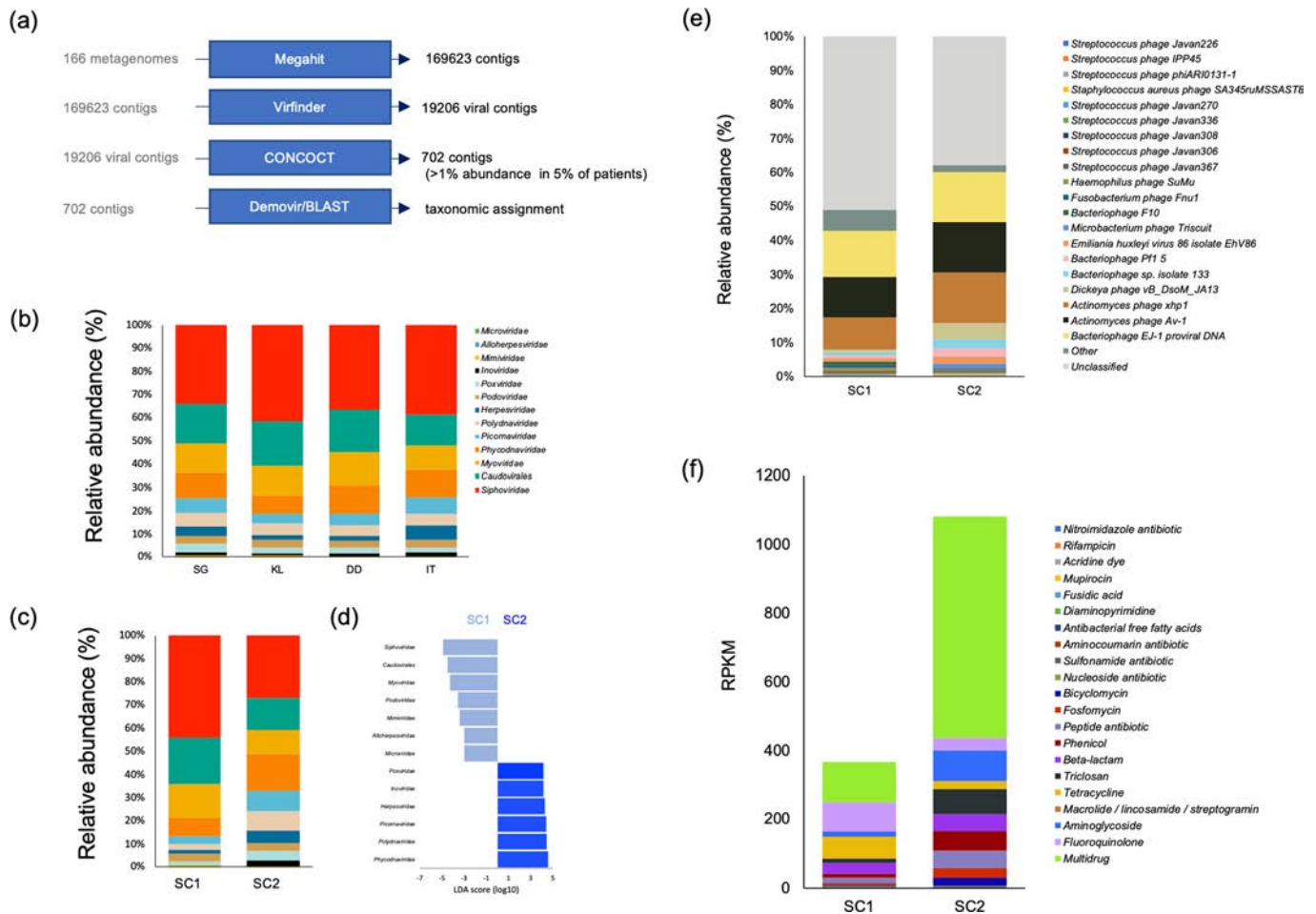


Extended Data Fig. 7 | See next page for caption.

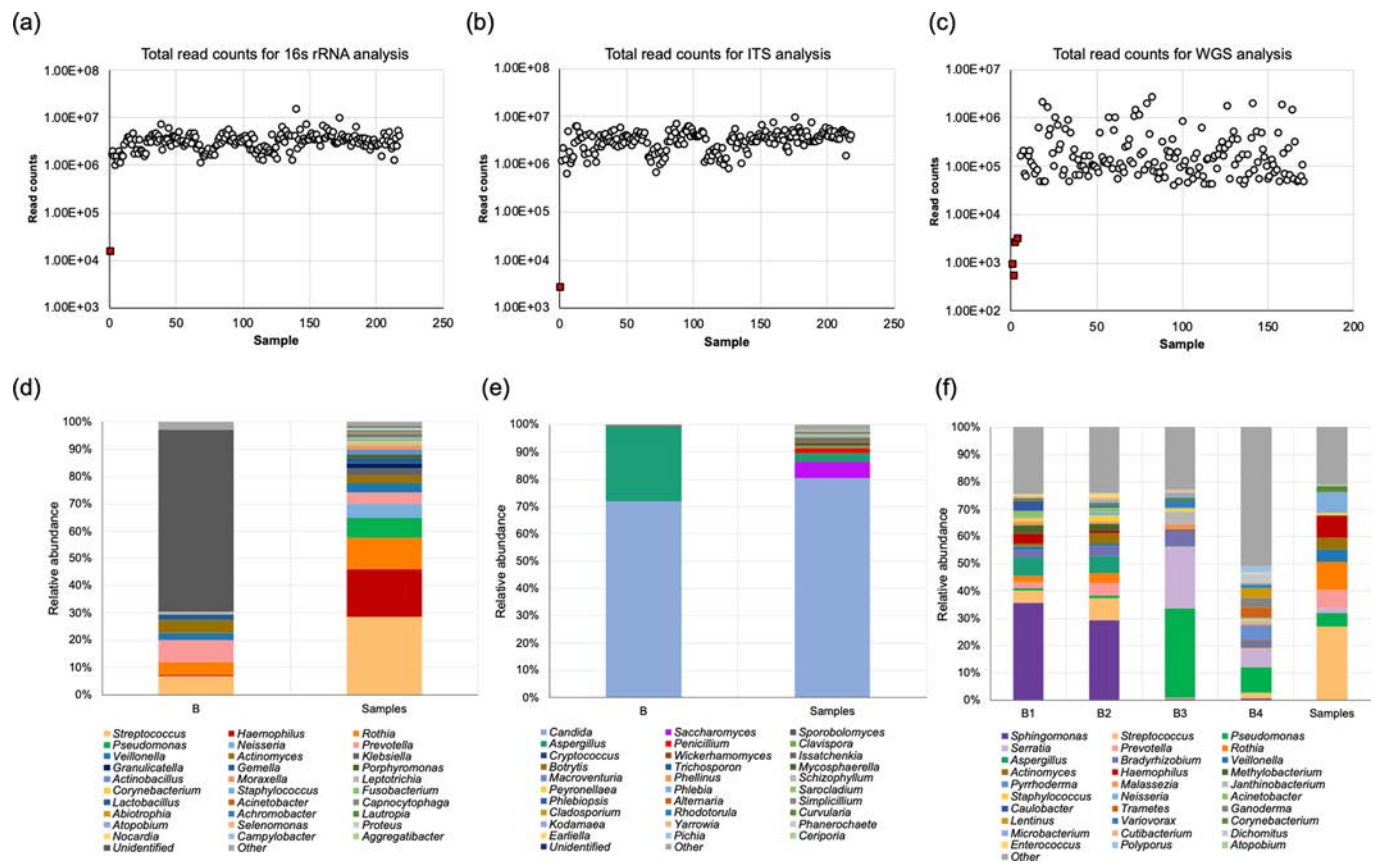
**Extended Data Fig. 7 | Comparison of culture data vs microbiome profiles and assessment of the interactome as a predictor of ‘time to next exacerbation’ in longitudinal samples. a**, For each patient (P1-17, n=17 biologically independent patient samples) three stacked barplots indicate the derived 16S rRNA profiles observed at each timepoint (n=3) from baseline to exacerbation and post-exacerbation (from left to right). Above each set of barplots are the indicated corresponding results from microbiological culture and the antibiotics used to treat the respective exacerbation (AMX - amoxicillin, AMC - amoxicillin-clavulanic acid, CAZ - ceftazidime, CIP - ciprofloxacin, DOX - doxycycline, MEM - meropenem, TZP - piperacillin-tazobactam). \*Asterisks indicate the co-isolation of *P. aeruginosa* and *H. influenzae* observed in two patients. A sub-analysis of microbiome diversity in patients suffering a subsequent exacerbation in the 12-week period following the initial occurrence versus those who remained exacerbation free for >12 weeks post-exacerbation is illustrated. **b**, Shannon Diversity Index at baseline (pre-exacerbation), during initial pulmonary exacerbation and post-exacerbation in those experiencing a second exacerbation in <12 weeks (red, n=9) and >12 weeks (blue, n=8). **c**, Berger-Parker index at baseline (pre-exacerbation), during initial pulmonary exacerbation and post-exacerbation in those experiencing a second exacerbation in <12 weeks (red) and >12 weeks (blue). Box plots reflect median and IQRs with whiskers bounding non-outlier values. ns; not significant (Kruskal-wallis). **d**, Correlation analysis of microbial abundance and **(e)** microbial interactions associated with time to next exacerbation. Heatmaps illustrate significant ( $p < 0.05$ ) associated correlations with colour indicating the strength of correlation (Spearman’s  $\rho$ ). Vertical text font colouration indicates kingdom membership; blue = bacteria, green = fungi, red = virus. *Psuedom.* = *Pseudomonas*, *Strep.* = *Streptococcus*, *Neiss.* = *Neisseria*, *Sacch.* = *Saccharomyces*. A multivariate adaptive regression spline (MARS) was implemented with both microbes and interactions (strength of interaction; edge weight) as predictor variables for ‘time to next exacerbation’ defined as; <12 weeks or >12weeks. Feature importance plots based on generalized cross-validation scores (gcv) illustrating the most important **(f)** microbes and **(g)** microbial interactions predicting the time to next exacerbation across baseline, exacerbation and post-exacerbation timepoints. Computed R-squared (RSq) and Generalized R-squared metric (GRsq) are indicated in table **(h)** reflecting goodness of the fit for each model.



**Extended Data Fig. 8 | Analysis of lung function and breathlessness in functionally-defined patient clusters.** Box and whisker plots showing (a) lung function (as FEV1 % predicted,  $p = 3.5 \times 10^{-2}$ ) and (b) breathlessness (symptom) score (as MMRC) in patients clustered according to functional analysis of sputum-derived metagenomes (FC1;  $n = 116$ , FC2;  $n = 50$ ). Box plots reflect median and IQRs with whiskers bounding non-outlier values. Significance levels for the observed between-group differences are indicated as follows: ns: non-significant;  $*p < 0.05$  (Mann-Whitney U). FEV1% = forced expiratory volume in the 1<sup>st</sup> second (% predicted). MMRC = modified medical research council.



**Extended Data Fig. 9 | Metagenomic analysis of the virome in an independent cohort of  $n = 166$  bronchiectasis patients recruited from Singapore, Malaysia, Scotland, and Italy.** To gain further insight into the respiratory virome we conducted an analysis of  $n = 166$  metagenomes from sputum. **a**, A schematic overview of the analysis pipeline used to construct virome profiles. Virome profiles were derived by first generating a metagenomic assembly using all available metagenomic sequences trimmed and free of contaminating human DNA followed by detection of viral contigs using Virfinder with subsequent refinement using CONCOCT yielding 702 contigs present in at least 1% relative abundance in 5% of patients. Taxonomic assignment was performed using Demovir and BLAST. **b**, Relative abundance of viral families identified across samples from Singapore (SG) Kuala Lumpur (KL) Scotland (DD) and Italy (IT). **c**, Relative abundance of viral families identified in wSNF clusters SC1 and SC2. **d**, LefSe discriminant taxa analysis comparing viral families observed in SC1 and SC2. **e**, BLAST analysis of identified viral contigs in SC1 and SC2. **f**, Analysis of antibiotic resistance gene family distribution in SC1 versus SC2 (RPKM - Reads Per Kilobase of transcript, per Million mapped reads).



**Extended Data Fig. 10 | Assessment of background DNA contamination in sequencing blank extraction controls.** Read counts of taxonomically assigned sequences for **(a)** 16s rRNA, **(b)** fungal ITS and **(c)** WGS metagenomic classification. Open circles represent samples while red squares indicate blank DNA extraction controls. Pooled blanks ( $n = 4$ ) were sequenced for 16s rRNA and ITS sequencing runs while four (un-pooled) extraction blanks were sequenced separately for WGS metagenomic analysis (B1-B4). Observed profiles of **(d)** bacterial, **(e)** fungal and **(f)** metagenomically assigned taxa identified in sequencing reads derived from DNA extraction blanks compared to aggregated sample profiles are indicated.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used for data collection.

Data analysis

All analysis performed in this study utilizes a suite of published packages and custom codes with modifications to previously published algorithms. The custom codes and all the analysis were implemented in Python(v2.7 & v3.6.1) and R (v3.2.4 & v3.5.1). Cytoscape(v3.7.1) was used to visualize networks and perform network analysis. Bioinformatic tools employed in in analysis of short-read sequencing data Trimmomatic (v0.39), FastQC (v0.11.8), bowtie2 (v2.3.3.1 & v2.3.5.1), samtools (v1.3.1 & 1.9), bedtools (v2.28), ShortRead (v3.6.2), MG-RAST annotation pipeline (v4.0.3), Kaiju version (v1.7.2), CONCOCT (v1.1.0) and Demovir (<https://github.com/feargalr/Demovir>). All codes central to our analytical pipeline are openly available at [https://github.com/translational-respiratory-lab/The\\_Interactome](https://github.com/translational-respiratory-lab/The_Interactome) and further codified as an online tool accessible at <https://integrative-microbiomics.ntu.edu.sg/>. Docker image; [https://hub.docker.com/r/jayanthkumar/integrative\\_microbiomics](https://hub.docker.com/r/jayanthkumar/integrative_microbiomics).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All sequence data described in this study has been uploaded to the National Centre for Biotechnology Information (NCBI) Sequence read archives (SRA) under project accession PRJNA590225. Other associated data including bacterial, fungal and viral profiles for all patients, as well as patients clinical attributes are available via the study's code repository [https://github.com/translational-respiratory-lab/The\\_Interactome](https://github.com/translational-respiratory-lab/The_Interactome).

Several publicly available reference datasets were employed in taxonomic and functional classification of the sequence reads; 16S targeted analysis; greengenes may 2013; [https://greengenes.secondgenome.com/?prefix=downloads/greengenes\\_database/gg\\_13\\_5/](https://greengenes.secondgenome.com/?prefix=downloads/greengenes_database/gg_13_5/) ITS targeted analysis; UNITE v7.2; <https://plutof.ut.ee/#/doi/10.15156/BIO/587475> KEGG functional classification; <https://www.genome.jp/kegg/> Kaiju source databases for metagenomic classification v1.7.2; <http://kaiju.binf.ku.dk/>

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Given the exploratory nature of the study, no formal power calculation was performed. Our observational study incorporating multi-kingdom microbiome data is the largest and only such study reported to date.
Data exclusions	Patients were recruited during routine visits to the outpatient clinic and were clinically stable at recruitment. Clinical stability was defined as the absence of new symptoms and where no change to bronchiectasis therapy had occurred in the preceding six week period. Patients were excluded if they had any other major respiratory diagnosis (asthma or COPD) out-ruled by clinical symptoms and established spirometry criteria, were pregnant or breastfeeding, had active mycobacterial disease or were on chemotherapy for malignancy. Patients with any active infection (necessitating acute use of antibiotics) or taking systemic corticosteroids in the four weeks preceding recruitment were excluded.
Replication	The study was not formally replicated but findings from the observational study were confirmed in our longitudinal analysis and independent metagenomic analysis in an independent patient cohort. Individual biologically independent sputum samples were taken from each patient in the observational study cohort. In the longitudinal arm, the microbiome was sampled were at three time-points.
Randomization	This is an observational study without treatment groups.
Blinding	This is an observational study with no blinding as there were no treatment or predefined patient groups requiring concealment from participants.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	All clinical data comprising the bronchiectasis severity index (BSI) including age, Body Mass Index (BMI), Medical Research Council (MRC) dyspnea score, FEV1 percentage predicted values, radiological severity, number of exacerbations in the preceding year, hospitalizations in the preceding year, microbial colonization with other organisms and colonization by <i>P. aeruginosa</i> was recorded for each patient, as was data on gender, disease etiology and smoking status. Patient demographics are described in Supplementary tables S1, S8 and S9.
Recruitment	Recruitment took place between March 2016 and Dec 2019. Patients with stable bronchiectasis were recruited as part of the CAMEB study; a cross-sectional Cohort of Asian and Matched European Bronchiectasis as previously described. Patients were recruited at three sites in Singapore, and a single site each in Malaysia, Scotland and Italy. Matching of Asian and European patients on age, sex and disease severity was employed to mitigate against geographic variability in these clinical confounders.

## Ethics oversight

The study was approved by the institutional review boards of all participating institutes as follows and all patients gave written informed consent to participate: CIRB 2016/2073 mutually recognized by DSRB; NTU IRB-2016-01-031; NTU IRB-2017-07-023; NTU IRB-2017-12-010 (Singapore); UMMC 2018725-6524 (Malaysia); NHD 12/ES/0059; NHD 16/NW/0101 (Dundee, Scotland) and 255\_2020 (Comitato Etico Milano Area 2, Milan, Italy).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

## Clinical trial registration

NA - Not a clinical trial.

## Study protocol

The study protocol was approved by the institutional review boards of all participating institutes as follows: CIRB 2016/2073 mutually recognized by DSRB; NTU IRB-2016-01-031; NTU IRB-2017-07-023; NTU IRB-2017-12-010 (Singapore); UMMC 2018725-6524 (Malaysia); NHD 12/ES/0059; NHD 16/NW/0101 (Dundee, Scotland) and 255\_2020 (Comitato Etico Milano Area 2, Milan, Italy) and is available on request.

## Data collection

Recruitment included three sites in Singapore (Singapore General Hospital, Changi General Hospital and Tan Tock Seng Hospital), one Malaysian site (UKM Medical Centre, Kuala Lumpur) and two European sites (Ninewells Hospital, Dundee, UK; Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy). Recruitment took place between March 2016 and Dec 2019.

## Outcomes

Our analysis was hypothesis generating. The clinical primary outcomes of interest were the clinical measures comprising the broncheictasis severity index (BSI) including age, Body Mass Index (BMI), Medical Research Council (MRC) dyspnea score, FEV1 percentage predicted values, radiological severity, number of exacerbations in the preceding year, hospitalizations in the preceding year, microbial colonization with other organisms and colonization by *Pseudomonas aeruginosa*.