Atopobium and *Fusobacterium* as novel candidates for sarcoidosis-associated microbiota

Alexandra Zimmermann[1], Henrik Knecht[1], Robert Häsler[1], Gernot Zissel[2], Karoline I. Gaede[3,4], Sylvia Hofmann[5], Almut Nebel[1], Joachim Müller-Quernheim[2], Stefan Schreiber[1,6] and Annegret Fischer[1]

Affiliations: [1]Institute of Clinical Molecular Biology, Kiel University and University Hospital Schleswig-Holstein, Kiel, Germany. [2]Dept of Pneumology, University of Freiburg, Freiburg, Germany. [3]BioMaterialBank Nord, Dept of Medicine, Research Center Borstel, Leibniz Center for Medicine and Biosciences, Borstel, Germany. [4]Airway Research Center North, Member of the German Center for Lung Research. [5]Dept of Conservation Biology, Helmholtz Centre for Environmental Research, Leipzig, Germany. [6]Clinic of Internal Medicine I, University Hospital Schleswig-Holstein, Kiel, Germany.

Correspondence: Stefan Schreiber, Institute of Clinical Molecular Biology, Kiel University, Rosalind-Franklin-Strasse 12, 24105 Kiel, Germany. E-mail: s.schreiber@mucosa.de

@ERSpublications
Sarcoidosis lung microbial profiles http://ow.ly/IfTC30gxm2U

Cite this article as: Zimmermann A, Knecht H, Häsler R, *et al. Atopobium* and *Fusobacterium* as novel candidates for sarcoidosis-associated microbiota. *Eur Respir J* 2017; 50: 1600746 [https://doi.org/10.1183/13993003.00746-2016].

ABSTRACT Sarcoidosis is a granulomatous disease that mainly affects the lung. A role of microbial factors in disease pathogenesis is assumed, but has not been investigated systematically in a large cohort.

This cross-sectional study compared the lung microbiota of 71 patients with sarcoidosis, 15 patients with idiopathic pulmonary fibrosis (non-infectious controls) and 10 healthy controls (HCs). Next-generation sequencing of 16S DNA was used on bronchoalveolar lavage samples to characterise the microbial composition, which was analysed for diversity and indicator species. Host genotypes for 13 known sarcoidosis risk variants were determined and correlated with microbial parameters.

The microbial composition differed significantly between sarcoidosis and HC samples (redundancy analysis ANOVA, p=0.025) and between radiographic Scadding types. *Atopobium* spp. was detected in 68% of sarcoidosis samples, but not in HC samples. *Fusobacterium* spp. was significantly more abundant in sarcoidosis samples compared with those from HCs. Mycobacteria were found in two of 71 sarcoidosis samples. Host-genotype analysis revealed an association of the rs2076530 (*BTNL2*) risk allele with a decrease in bacterial burden (p=0.002).

Our results indicate Scadding type-dependent microbiota in sarcoidosis BAL samples. *Atopobium* spp. and *Fusobacterium* spp. were identified as sarcoidosis-associated bacteria, which may enable new insights into the pathogenesis and treatment of the disease.

This article has supplementary material available from erj.ersjournals.com

Received: April 13 2016 | Accepted after revision: Aug 28 2017

Support statement: The study was supported by the Deutsche Forschungsgemeinschaft (DFG) through the Cluster of Excellence 'Inflammation at Interfaces'. The BioMaterialBank North is funded in part by the Airway Research Center North (ARCN), Member of the German Center for Lung Research (DZL), and is a member of the PopGen 2.0 network (P2N), which is supported by a grant from the German Ministry for Education and Research (01EY1103). Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: None declared.

Introduction

Sarcoidosis is an inflammatory disease with global prevalence, which is characterised by the presence of non-caseating granulomas [1]. Pulmonary involvement is common in about 90% of cases, but any other organ can be affected. Depending on the severity of disease manifestation, 1–5% of patients die from the consequences of respiratory failure and sarcoidosis-associated fibrosis [2, 3]. Five sarcoidosis types, known as Scadding stages, are distinguishable by chest radiography: Type 0 shows no abnormalities on chest radiography scans, while Type I is characterised by hilar or mediastinal lymphadenopathy, Type II by lymphadenopathy and parenchymal lesions, Type III by parenchymal disease only, and Type IV by pulmonary fibrosis. These radiographic types are helpful for clinical and scientific stratification of cohorts.

Like other complex diseases, sarcoidosis is thought to be caused by an interaction between genetic and environmental factors. While a number of genetic risk variants are known, the environmental triggers are mostly unidentified. Organic and inorganic factors such as occupational exposure to respirable dust have been investigated [2, 4]. Owing to the similarities with mycobacterial disease, an inhaled infectious agent has been hypothesised, and mycobacterial antigens as well as some other bacterial candidates such as *Chamydophilia pnemoniae* and *Propionibacterium acnes* have been examined as potential triggers [5–10]. However, results regarding probable microbial involvement in disease manifestation are inconsistent. There is compelling evidence for a role of the microbiota in other pulmonary disease conditions such as chronic obstructive pulmonary disease (COPD), asthma and idiopathic pulmonary fibrosis (IPF) [11–14]. For sarcoidosis, this has so far been investigated systemically in two studies, but both were severely limited by their small sample sizes [15, 16]. In the current study, we used next-generation sequencing (NGS) technologies enabling the detection of both unculturable and culturable bacteria in a large sample of German patients with sarcoidosis to gain further insight into the sarcoidosis lung microbiota compared with healthy controls (HCs) and the microbiota compared between sarcoidosis radiographic types. Patients with IPF were included as non-infectious controls with fibrotic aspects, and patients with pneumonia were included as proof of principle for the analysis of bacterial burden. In addition, we investigated a potential interaction of sarcoidosis lung microbiota with sarcoidosis risk-associated single nucleotide polymorphisms (SNPs).

Methods

Study participants

In total, 71 patients with sarcoidosis were included in the study (table 1), from University Hospital Freiburg (n=29) and the Research Center Borstel (n=42). All participants gave their written consent, and the study was approved by the local ethics committees. Of the patients with sarcoidosis, 39 were men and 32 were women. Age ranged between 24 and 67 years, with a median of 43 years. Stratified by radiographic type, three patients were assigned to radiographic Type 0, 17 to Type I, 39 to Type II, six to Type III and six to Type IV. None of the patients were receiving antibiotics at the time of bronchalveolar lavage (BAL) sampling or had any other lung diseases, except two of the patients with radiographic Type 0,

TABLE 1 Characteristics of study samples

	Sarcoidosis	IPF	Controls
Sample size	71	15	10
Age years range (median)	24–67 (43)	48–89 (72)	21–53 (23)
Males/females	38/33	9/6	6/4
Type 0/I/II/III/IV	3/17/39/6/6		
Never/former/current smokers	56/11/4	12/3/0	10/0/0
Medication#	0	0	0
Histologically confirmed yes/no/NA	50/12/9		
Time from initial diagnosis to BAL			
0 years	57		
1 year	4		
2 years	3		
4 years	3		
5 years	2		
30 years	1		
34 years	1		

IPF: idiopathic pulmonary fibrosis; NA: not available; BAL: bronchoalveolar lavage. #: steroids and immunosuppressants.

who also had bronchial asthma. Within the sarcoidosis samples, 56 were from nonsmokers, 11 from former smokers and 4 from occasional smokers. All subjects with Scadding Types I, II, III or IV fulfilled the requirement of pulmonary or lymphatic involvement.

10 healthy subjects from University Hospital Freiburg (n=5) and Research Center Borstel (n=5) were included as HCs, and 15 patients with IPF as the additional non-infectious patient group, for which low bacterial diversity was expected [14]. As a control for the measurement of a high bacterial burden, samples of 22 patients with different infectious pneumonia types of bacterial origin were included, for which bacterial burden was assessed, but no microbial profile was generated. All 22 pneumonia samples were characterised by an overgrowth of bacteria in the lower respiratory tract. The exact composition of these microbiomes was not determined for this study, as only the absolute amount of bacterial DNA was of relevance here.

BAL and sample preparation

BAL was performed as previously described at both recruiting sites for patients and controls [17]. In brief, 25 mL aliquots from a total volume of 200–300 mL of sterile saline (0.9% NaCl) were infused into a lingula or middle lobe segment of the lung and immediately aspirated. Aspirated lavage aliquots from one donor were pooled and stored until further preparation at −80°C. Microbial DNA and RNA were extracted and prepared by standard laboratory procedures (supplementary methods). The V1-V2 region of the 16S rRNA gene in the extracted DNA was amplified with primer 27F combined with 454 Life Sciences adapter B and primer 338R with 454 Life Sciences adapter A (Roche, Penzberg, Germany; primers from Metabion, Planegg, Germany). The reverse primer contained a multiplex barcode identifier sequence (10 bp), which allowed identification of individual samples. DNA was extracted with the Molysis Complete 5 Kit (Molzym, Bremen, Germany), which is also capable of extracting intracellular bacteria. The samples were prepared for pyrosequencing following the preparation procedure of Stratil *et al.* [18] (supplementary methods for details). To assess the differences between the total and the potentially active sarcoidosis microbiota, 16S rRNA was extracted with the MO BIO PowerMicrobiome™ RNA Isolation Kit, with additional DNA digestion, transcribed to cDNA by reverse transcriptase PCR (RT-PCR) and prepared for sequencing similarly to DNA. The 16S rRNA was compared with the 16S rDNA in terms of identifying physiologically active bacteria.

Pyrosequencing analysis

Amplicons were sequenced using a GS 454 FLX titanium technology and chemistry sequencer (Roche, Penzberg, Germany) in six runs with random sampling. Technical control samples were included to detect potential contamination of reagents with bacteria, and did not show systematic contamination. Sequence and quality files were barcode-sorted using PANGEA [19]. Sequences with a read length less than 200 bp and a quality score of less than 25 were rejected. Noise reduction was carried out using the software mothur [20]. Ambiguous sequences and sequences that differed in the primer or barcode sequence were eliminated, as were sequences with more than eight homopolymers and those with chimeric sequences. To correct for the variable number of the sequences, the output was normalised to 1000 sequences per sample.

Taxonomic analysis: operational taxonomic unit-based approach

Operational taxonomic units (OTUs) are groups of sequences that are clustered based on similarity, allowing the assignment of genera or species. The sequences were clustered with a threshold of 97% similarity, using the mothur pipeline. The microbial composition was analysed using established parameters and methods, such as the Shannon Index for alpha-diversity assignment [21] and redundancy analysis with Hellinger-transformed data for beta diversity [22]. Alpha diversity refers to the number and distribution of bacterial types within a sample, while beta diversity is the same comparison between different samples. To assess the differences in the species distribution between sarcoidosis, IPF and HC samples, the Kruskal–Wallis test was used. Differences in alpha diversity and bacterial burden were tested with Wilcox rank sum test, and beta diversity with ADONIS. More details are given in supplementary methods.

Genotyping and analysis for interaction

Human DNA was extracted from the BAL samples described above, using the DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany), and subsequent whole genome amplification was conducted using RPLI-g Single Cell Kit (QIAGEN). Genotypes were determined using Taqman® genotyping technology (Life Technologies, Foster City, USA) for the following 13 SNP variants, which are known to be associated with sarcoidosis, with the respective gene loci given in brackets: rs1049550 (*ANXA11*), rs2076530 and rs5007259 (both *BTNL2*), rs4143332 (*HLA-B*), rs9277542 (*HLA-DPB1*), rs479777 (chr11q13.1), rs1050045 (*OS9*), rs10484410 (*ZNF451*), rs1040461 (*RAB23*), rs12069782 (*IL23R*), rs4921492 (*IL12B*), rs653178

(*ATXN2*) and rs223498 (*NFKB1*/*MANBA*). A detailed description of the selected variants including references and genotype frequencies is given in supplementary table E1. Alpha diversity and bacterial burden were analysed for differences by host genotype using Wilcox rank sum test, and results corrected for multiple testing by false discovery rate (FDR) correction. Beta diversity was analysed using transformation-based redundancy analysis (tb-RDA) stratified by genotype.

Bacterial burden

To assess bacterial burden, the amount of bacterial DNA in the BAL samples was estimated by Taqman® [23], with primers Eub338F and Eub518R for the estimation of the bacterial burden and primers bActin_F and bActin_R to measure the amount of human DNA.

Results

In total, 374,341 sequences were obtained by NGS after quality checking and preprocessing, and these were then normalised to 1000 sequences per sample. Taxonomic analysis assigned these sequences to 3413 OTUs, including 121 OTUs above the abundance threshold (>0.1%). The total sum of sequences belonging to the OTUs >0.1% divided by the total sum of sequences in the whole dataset showed that 90% of the taxonomic information in the dataset was represented at this cutoff point.

Microbial diversity (alpha and beta diversity) and bacterial burden in sarcoidosis, IPF and controls

The average alpha diversity did not differ significantly between the sarcoidosis (Shannon index $(SI)_{Sarc}$=3.0, SD=0.52) and HC ($SI_{Control}$=2.8, SD=0.69) samples, whereas comparison of sarcoidosis and IPF samples revealed decreased diversity in IPF ($p<0.001$; SI_{IPF}=2.4±0.94; figure 1a). Discriminated by sarcoidosis radiographic types, alpha diversity did not differ significantly between the different types (diversity index: Type 0: SI_0=2.7, SD=1.2; Type I: SI_I=3.1, SD=0.34; Type II: SI_{II}=3.0, SD=0.46; Type III: SI_{III}=3.1, SD=0.41; and Type IV: SI_{IV}=3.0, SD=0.82; figure 2a).

The microbial composition differed significantly between sarcoidosis and HC samples ($ADONIS_{RDA}$, p=0.025; figure 1b) and between sarcoidosis and IPF samples ($ADONIS_{RDA,}$ p=0.001). Five OTUs were significantly imbalanced between sarcoidosis, IPF and HC samples: *Veillonella* (OTU91) was more abundant in HC samples, while *Streptococcus* (OTU1) was more abundant in HC and IPF samples, and *Atopobium* (OTU41) and *Fusobacterium* (OTU16) were more abundant in sarcoidosis samples. *Oribacterium* (OTU47) was less abundant in IPF samples compared with sarcoidosis and HC samples. Table 2 shows p-values and abundances. Regarding beta diversity of sarcoidosis radiographic types, a distinct clustering of sarcoidosis types and HC samples was observed ($ADONIS_{RDA}$, p=0.011; figure 2b). IPF samples clustered slightly apart from the sarcoidosis types and HC samples (supplementary figure E1).

Amount of human DNA did not correlate with bacterial burden (supplementary figure E2). Compared with HC samples, the bacterial burden was increased in IPF and pneumonia samples by a factor of 2.3 and 171, respectively, while the sarcoidosis samples showed a decrease in bacterial burden by a factor of 2.5 (figure 3), with a slight increase for Type I, and a decrease for Types II, III and IV (median burden: Controls=1.12×10^{-12}, I=3.03×10^{-12}, II=2.12×10^{-13}, III=2.58×10^{-14}, IV=5.14×10^{-14}).

Analysis of potentially confounding variables revealed no influence of sex, age, smoking history, recruitment site or sequencing batch on microbial profiles regarding alpha or beta diversity or bacterial burden.

Core microbiota and indicator species

Qualitative analysis of the 121 OTUs above the frequency threshold showed that 95 OTUS were shared between sarcoidosis, IPF and controls, while four OTUs (*Abiotrophica* spp., *Porphyromona*s spp., *Moraxella* spp., *Moryella* spp.) were found specifically in sarcoidosis samples. Core microbiota analysis demonstrated that *Streptococcus pseudopneumoniae* was found in 95% of the 71 sarcoidosis samples with the highest mean abundance of all OTUs (57%). Four *Streptococcus* (OTU1, OTU4, OTU6, OTU8), two *Prevotella* (OTU2, OTU9) and one *Veillonella* (OTU5) species were found in 90% of the sarcoidosis samples. A comparison of total bacterial composition showed that the representatives of these genera added up to more than 70% of the total bacterial load, and this was consistent with the results for the HC samples. Detailed bacterial species distribution in sarcoidosis samples is shown in figure 1c.

Further, the presence or absence of microbial species was analysed in order to identify microbial markers. In the analysis of sarcoidosis samples *versus* controls, *Atopobium* sp. (OTU41) occurred in 68% of the sarcoidosis samples with a mean relative abundance of 0.55%, but was not detected in HC samples (p=0.004). *Atopobium* was also present in IPF samples but less frequently and abundantly (mean±SD: IPF, 1.22±2.16; sarcoidosis, 5.51±7.31). A presence/absence map is given in supplementary figure E3. The

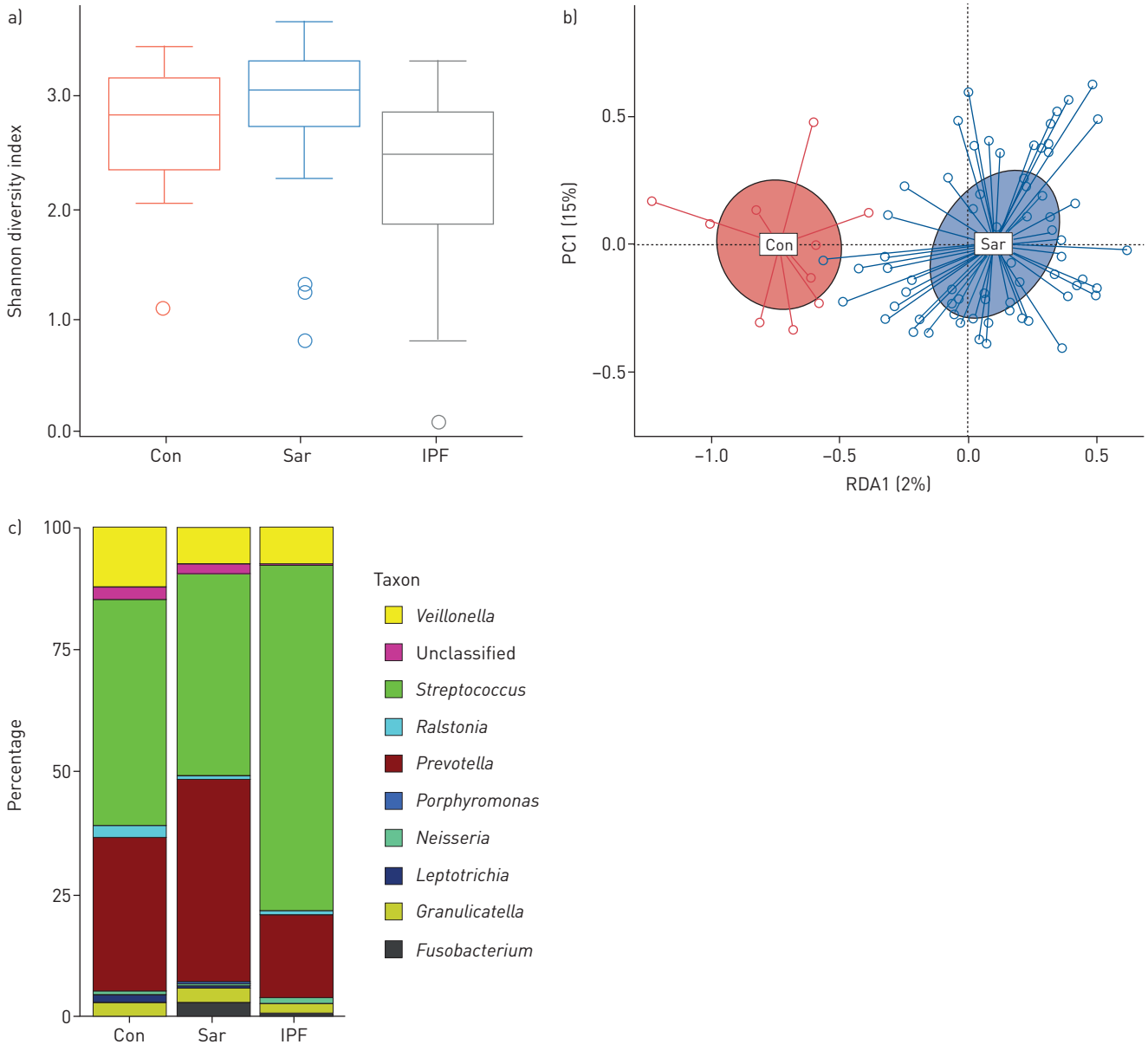

FIGURE 1 a) Alpha diversity of sarcoidosis (Sar) (n=71), idiopathic pulmonary fibrosis (IPF) (n=15) and healthy control (Con) (n=10) samples as measured by Shannon Index. The horizontal line inside the box represents the median. The whiskers represent the lowest and highest values within 1.5×interquartile range. Outliers and individual sample values are shown as circles. b) Beta diversity is shown as Hellinger-transformed redundancy analysis of sarcoidosis samples *versus* healthy controls. The axes show the constrained axis from redundancy analysis (RDA1) and the first principal component (PC1). c) Microbial composition of sarcoidosis, IPF and healthy control samples based on the nine most prominent genera.

amount of *Atopobium* was significantly increased for sarcoidosis types I and II compared with HC and IPF samples ($P_{I vs.C}$=0.036, $P_{II vs.C}$=0.036, $P_{I vs.IPF}$=0.045, FDR-corrected; figure 4). *Fusobacterium* spp. (OTU16) was significantly more abundant in sarcoidosis types II and III compared with HC samples ($P_{II vs.C}$=0.040, $P_{III vs.C}$=0.042; figure 4), and also occurred in IPF samples. One species (OTU91, *Veillonella* spp.) was specific for HCs.

Levels of Actinobacteria

As members of the class Actinobacteria, like Mycobacteria and Propionibacteria, are suggested to influence disease manifestation of sarcoidosis, *Propionibacterium*, *Mycobacterium* and the closely related *Corynebacterium* spp. were selected as candidates for detailed analysis [10, 24, 25]. *Mycobacterium* spp. were found at low levels in two out of 71 sarcoidosis samples (<5 observations) and in three IPF samples. *Propionibacterium* and *Corynebacterium* were detected at high frequency in 54% of the cases and 50% of

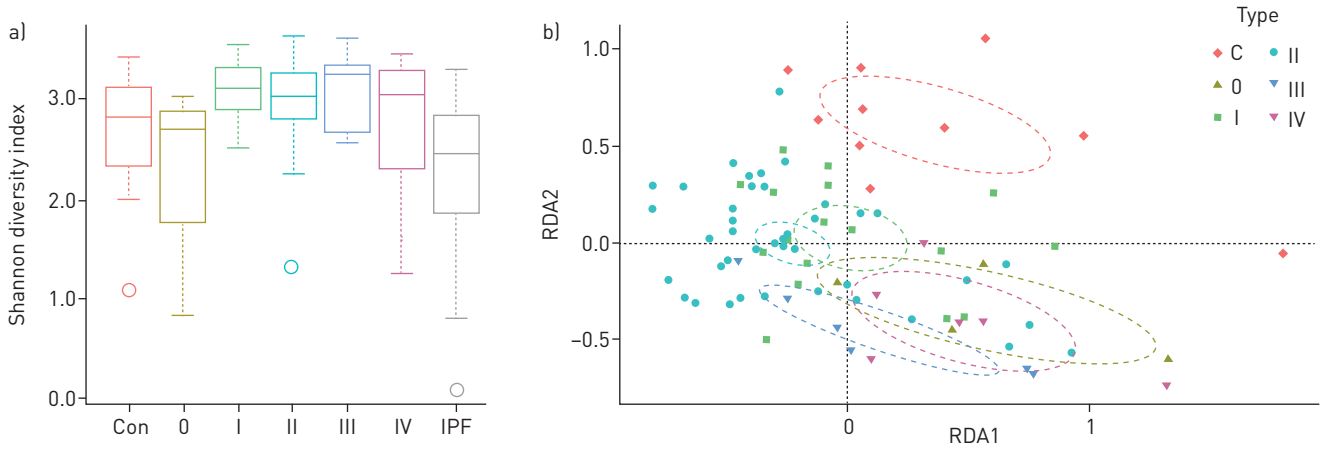

FIGURE 2 a) Alpha diversity in healthy control samples (Con) (n=10), sarcoidosis radiographic types (Type 0 (n=3), Type 1 (n=17), Type II (n=39, Type III (n=6), Type IV (n=6)) and idiopathic pulmonary fibrosis (IPF) samples (n=15). The horizontal line inside the box represents the median, and the whiskers represent the lowest and highest values within 1.5 interquartile range. Outliers and individual sample values are shown as circles. b) Cluster analysis of sarcoidosis types *versus* healthy controls by transformation-based redundancy analysis (tb-RDA), including 95% confidence intervals (dashed lines).

the HC samples, without displaying a significant difference in levels between IPF, HC and sarcoidosis samples or between sarcoidosis types.

Active microbiota

In the potentially alive and active microbiota (16S rRNA), the mean percentage of the genus *Streptococcus* was elevated (61.3% *versus* 39.1%), while the percentages of *Diaphorobacter* and *Ralstonia* were reduced (0% *versus* 5.4% and 0.2% *versus* 29.2%, respectively). The disease and Scadding type discriminating genera described above, including *Atopobium* and *Fusobacterium*, were found within the active microbiota.

Host–genotype microbiota interaction

13 variants were analysed for a potential association with alpha and beta diversity (supplementary table E2) and bacterial burden. After correction for multiple testing, the only significant result was retrieved for the carriership of the rs2076530 (*BTNL2*) risk allele A, which was significantly associated with a decrease in bacterial burden (FDR-corrected p=0.002; figure 5). An overview of statistical testing is given in supplementary table E2.

Discussion

In this study, we characterised the microbial profiles of the lower respiratory tract and demonstrated that in our sample, the microbiota differed in composition between sarcoidosis samples and controls, and within sarcoidosis radiographic types. We included IPF as an additional phenotype, as this fibrotic condition might resemble fibrosis of sarcoidosis to some degree. The significant differences between sarcoidosis samples compared with HC and IPF samples indicate the existence of sarcoidosis-specific microbial profiles.

TABLE 2 Results of Kruskal–Wallis test for unequal distribution of species between sarcoidosis and idiopathic pulmonary fibrosis (IPF) samples and controls

OTU	Genus	p-value	FDR-corrected p-value	Sarcoidosis mean	IPF mean	Control mean
Otu00091	*Veilonella*	1.73×10^{-5}	0.002	1.11	0.56	9.33
Otu00041	*Atopobium*	0.00012	0.004	5.51	1.22	0.00
Otu00047	*Oribacterium*	0.00013	0.004	4.19	0.39	2.89
Otu00001	*Streptococcus*	0.00013	0.004	81.24	293.17	138.33
Otu00016	*Fusobacterium*	0.00105	0.025	14.35	4.11	4.33

False discovery rate (FDR)-corrected p-values represent correction for all 121 operational taxonomic units (OTUs) tested. The mean abundance is given for clarity but was not used in the applied test statistics.

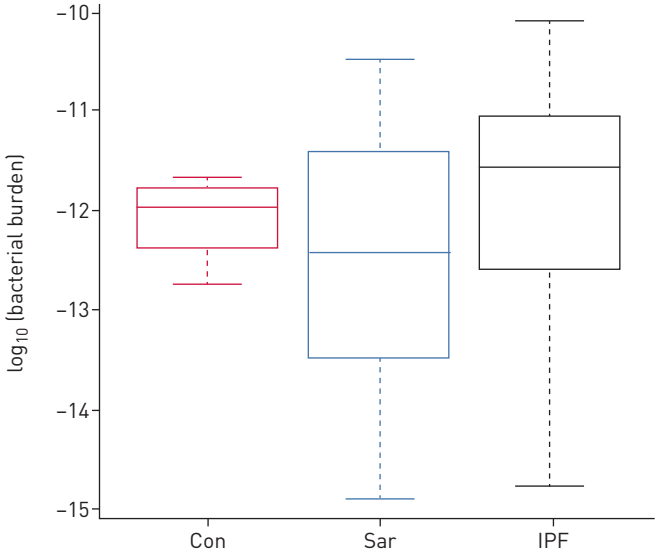

FIGURE 3 Bacterial burden of sarcoidosis (Sar), idiopathic pulmonary fibrosis (IPF) and healthy samples (Con). The horizontal line inside the box represents the median, and the whiskers represent the lowest and highest values within 1.5× interquartile range.

In microbial profiling of BAL, contamination of the samples can never be excluded completely, owing to the bronchoscopial sampling procedure. Nevertheless, a number of studies have described distinct microbial profiles in the various parts of the human respiratory tract [26, 27]. Technical controls were included in our study, which revealed sporadic presence of *Propionibacterium* spp. in the reagents, which is in line with previous reports [28]. This finding highlights *Propionibacterium* spp. as ubiquitous bacteria that are difficult to control, whereas none of the sarcoidosis-associated species were found in the technical controls. Although none of the probands were receiving corticosteroids at the time of BAL, we did not have information on previous usage of corticosteroids. Thus, we cannot exclude a potential long-term impact of inhaled corticosteroids on microbial diversity in the investigated samples [14, 29].

As expected, the bacterial burden was increased in IPF and pneumonia samples, while it was slightly decreased in the sarcoidosis samples. To exclude a technical bias due to different amounts of total DNA in the BAL samples, we tested for correlation of the amount of human DNA with the bacterial load and found no significant correlation. Further, there was no significant change in alpha diversity in sarcoidosis compared with HC samples. Previous studies showed contradicting results on alpha diversity of sarcoidosis samples. While Scher *et al.* [16] found a reduction in alpha diversity compared with HC samples, Garzoni *et al.* [15] did not find any differences. Both studies were limited by small sample size (n=10 and n=7, respectively). In this larger study (n=71), the sarcoidosis samples showed a weak elevation in alpha diversity, but this was not statistically significant. In line with the current model of sarcoidosis

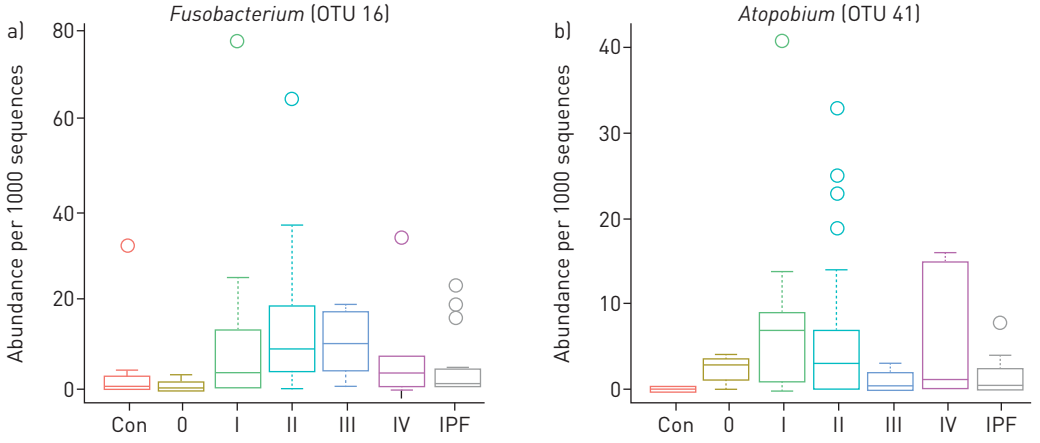

FIGURE 4 Relative abundance (%) of the candidate species *Atopobium* (OTU41) and *Fusobacterium* (OTU16) stratified by sarcoidosis (Sar) types (Type 0 (n=3), Type 1 (n=17), Type II (n=39), Type III (n=6), Type IV (n=6)) compared with control (Con) (n=10) and idiopathic pulmonary fibrosis (IPF) (n=15) samples. The horizontal line inside the box represents the median, and the whiskers represent the lowest and highest values within 1.5 interquartile range. Outliers and individual sample values are shown as circles.

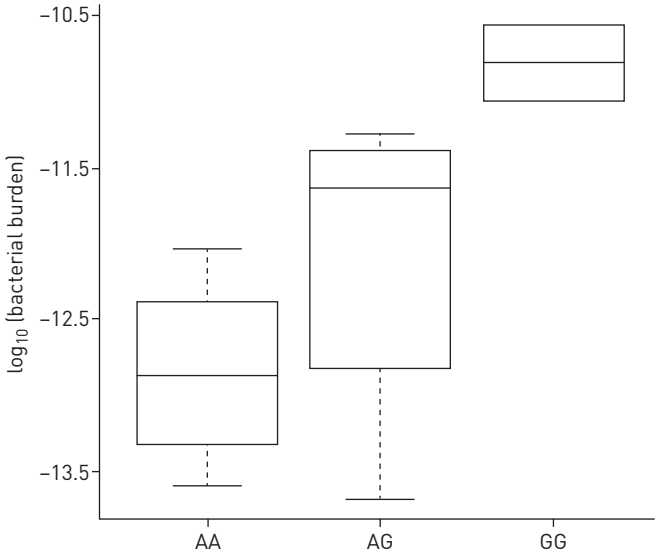

FIGURE 5 Bacterial burden of sarcoidosis samples stratified by BTNL2 (rs2076530) genotype (GG (n=6), AG (n=23), AA (n=16)). The horizontal line inside the box represents the median, and the whiskers represent the lowest and highest values within 1.5× interquartile range.

pathogenesis [8], this clearly argues against a classic infectious process in sarcoidosis, which would be characterised mainly by overgrowth of specific bacterial species and thereby reduced alpha diversity [30].

The most abundant genera in the investigated sarcoidosis BAL samples were *Streptococcus, Prevotella* and *Veillonella* species, which is in accordance with previous studies of lung microbiota [15, 31]. Five OTU abundances differed significantly between cases and controls. While small previous studies found a reduction in *Burkholderia* [16] or no differences at all [15] compared with healthy and diseased control samples, we found members of the genera *Veillonella, Oribacterium, Streptococcus, Atopobium* and *Fusobacterium* to be imbalanced between sarcoidosis, IPF and HC samples. The latter two genera were more abundant in sarcoidosis samples, and thus might represent novel candidates for sarcoidosis-associated bacteria. As a major limitation, our study setup did not allow conclusion of whether or not the identified species have a causal influence on sarcoidosis. However, the significant disease association of *Atopobium* and *Fusobacterium* spp. may indicate potential disease relevance. Moreover, the observation that these bacteria were alive and possibly reproducing in the lung of patients with sarcoidosis supports their possible role in the pathophysiology of sarcoidosis.

Atopobium is a vaginal commensal and member of the oral flora [32], and has been described previously in the context of COPD [33]. In the current study, *Atopobium* showed the highest abundances in sarcoidosis Types I and II, and might therefore be a candidate indicator for an early disease state represented by these types. It might also represent an initial trigger for sarcoidosis, which would have to be investigated in a longitudinal study design and an adequate granulomatous model. Of potential mechanistic relevance, *Atopobium* belongs to the same phylum of bacteria as *Mycobacterium tuberculosis*. Thus, it is likely that both species share highly conserved antigens, capable of inducing a similar immune response in patients with sarcoidosis [34]. A second possible mechanism is an auto-antigen-like reaction invoked by *Atopobium* antigens, as described for rheumatoid arthritis (RA) [35].

Elevated levels of *Fusobacterium* spp. were found in radiographic Type II and III samples, both of which are characterised by parenchymal disease. *Fusobacterium* is a commensal bacterium of the respiratory and intestinal flora, with pathogenic potential and tissue invading ability. Clinical relevant associations were found for disorders such as periodontitis, wound infections and colorectal carcinoma [36–38], with increased invasive properties in an inflammatory setting [39]. *Fusobacterium* was also found to be part of an inflammatory enhancing community in certain pneumotypes [40] and influences tetracycline response [41, 42]. Whether the specific presence of *Fusobacterium* in our sarcoidosis samples is a cause or a consequence of the parenchymal involvement and chronic disease in the radiographic types II and III remains to be elucidated.

To our knowledge, *Atopobium* has not been described previously in relation to sarcoidosis, while *Fusobacterium* was previously found to be highly abundant in two out of 10 sarcoidosis samples and in controls in a study on RA [16].

Mycobacteria and *Propionibacteria* have been widely discussed in the context of sarcoidosis, but partly with conflicting findings [8]. Our study does not add any fundamental evidence to this topic, as

Mycobacteria and *Propionibacteria* were not found to be imbalanced between sarcoidosis and HC BAL samples. However, as we did not investigate granuloma biopsies, we also cannot exclude the presence of mycobacterial or propionibacterial factors in the granuloma or a disease-initialising role of these bacteria. Instead, in our hypothesis-free approach, we suggest *Atopobium* and *Fusobacterium* as qualified novel disease-associated candidates that now require independent confirmation in different populations, in a longitudinal setup and with direct experimental approaches.

In addition, an association of the *BTNL2* rs2076530 risk allele with decreased bacterial burden was found in our study. Such a correlation of the microbiota with host genotype has been shown in well-studied areas such as complex diseases of the gut [43, 44], whereas in pulmonary diseases, only an interaction of *MUC5B* polymorphism with bacterial burden in IPF samples has so far been described [14]. The correlation observed in our study might be explained by the role of BTNL2 in the inhibition of T-cell proliferation [45]. As SNP rs2076530 leads to a loss of functionality of the BTNL2 protein, the presence of this variant might result in overactivation of T-cells [46] with a general overactivation of the immune system, resulting in granuloma formation. At the same time, this might lead to the observed reduction of the bacterial burden. We found that the bacterial burden was slightly increased in Type I, while it was decreased in Type II, III and IV sarcoidosis, suggesting loss of bacterial burden with progressed disease. To our knowledge, a potential Scadding type-specific association of rs2076530 has not yet been investigated.

Taken together, our findings from BAL samples support the hypothesis of a sarcoidosis-specific microbiota, with Scadding types showing distinct bacterial patterns. We could not confirm *M. tuberculosis* using NGS, but we found imbalances in the sarcoidosis microbiota, especially *Atopobium* spp. and *Fusobacterium* spp. associated with sarcoidosis-specific Scadding types. Thus, we describe for the first time an association of BTNL2 risk genotypes with bacterial burden in sarcoidosis. This highlights the need to further assess the role of lung microbiota for manifestation, progression and therapy of sarcoidosis, and its interplay with genetic factors.

Acknowledgements

We thank all study participants and physicians for their contribution. We further thank the laboratory team of the Institute of Clinical Molecular Biology (Kiel, Germany) for technical assistance.

Author contributions: Data analysis: A. Zimmermann, H. Knecht; samples or datasets: R. Häsler, G. Zissel, K.I. Gaede, J. Müller-Quernheim; study design: S. Hofmann, S. Schreiber, A. Fischer; manuscript draft: A. Zimmermann, H. Knecht, R. Häsler, A. Nebel, A. Fischer; manuscript revision and final manuscript: all authors.

References

1 Baughman RP, Culver DA, Judson MA. A concise review of pulmonary sarcoidosis. *Am J Respir Crit Care Med* 2011; 183: 573–581.
2 Valeyre D, Prasse A, Nunes H, *et al.* Sarcoidosis. *Lancet Lond Engl* 2014; 383: 1155–1167.
3 Statement on sarcoidosis. Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) adopted by the ATS Board of Directors and by the ERS Executive Committee, February 1999. *Am J Respir Crit Care Med* 1999; 160: 736–755.
4 Miller A. Sarcoidosis, firefighters sarcoidosis, and World Trade Center "sarcoid-like" granulomatous pulmonary disease. *Chest* 2007; 132: 2053–2053.
5 Robinson L, Smith P, SenGupta D, *et al.* Most symptomatic sarcoidosis patients have lymph nodes containing bacteria. *Chest* 2011; 140: 615A–615A.
6 Schupp JC, Tchaptchet S, Lützen N, *et al.* Immune response to *Propionibacterium acnes* in patients with sarcoidosis — in vivo and in vitro. *BMC Pulm Med* 2015; 15: 75.
7 Oswald-Richter KA, Beachboard DC, Seeley EH, *et al.* Dual analysis for Mycobacteria and Propionibacteria in sarcoidosis BAL. *J Clin Immunol* 2012; 32: 1129–1140.
8 Chen ES, Moller DR. Etiologic role of infectious agents. *Semin Respir Crit Care Med* 2014; 35: 285–295.
9 Gaede KI, Wilke G, Brade L, *et al.* Anti-*Chlamydophila* immunoglobulin prevalence in sarcoidosis and usual interstitial pneumoniae. *Eur Respir J* 2002; 19: 267–274.
10 Abe C, Iwai K, Mikami R, *et al.* Frequent isolation of *Propionibacterium acnes* from sarcoidosis lymph nodes. *Zentralblatt Für Bakteriol Mikrobiol Hyg 1 Abt Orig Med Mikrobiol Infekt Parasitol* 1984; 256: 541–547.
11 Larsen JM, Musavian HS, Butt TM, *et al.* Chronic obstructive pulmonary disease and asthma-associated Proteobacteria, but not commensal *Prevotella* spp., promote Toll-like receptor 2-independent lung inflammation and pathology. *Immunology* 2015; 144: 333–342.
12 Molyneaux PL, Mallia P, Cox MJ, *et al.* Outgrowth of the bacterial airway microbiome after rhinovirus exacerbation of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013; 188: 1224–1231.
13 Hilty M, Burke C, Pedro H, *et al.* Disordered microbial communities in asthmatic airways. *PloS One* 2010; 5: e8578.
14 Molyneaux PL, Cox MJ, Willis-Owen SA, *et al.* The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014; 190: 906–913.
15 Garzoni C, Brugger SD, Qi W, *et al.* Microbial communities in the respiratory tract of patients with interstitial lung disease. *Thorax* 2013; 68: 1150–1156.
16 Scher JU, Joshua V, Artacho A, *et al.* The lung microbiota in early rheumatoid arthritis and autoimmunity. *Microbiome* 2016; 4: 60.

17 Prasse A, Georges CG, Biller H, *et al.* Th1 cytokine pattern in sarcoidosis is expressed by bronchoalveolar CD4+ and CD8+ T cells. *Clin Exp Immunol* 2000; 122: 241–248.
18 Stratil SB, Neulinger SC, Knecht H, *et al.* Temperature-driven shifts in the epibiotic bacterial community composition of the brown macroalga *Fucus vesiculosus*. *Microbiologyopen* 2013; 2: 338–349.
19 Giongo A, Crabb DB, Davis-Richardson AG, *et al.* PANGEA: pipeline for analysis of next generation amplicons. *ISME J* 2010; 4: 852–861.
20 Schloss PD, Westcott SL, Ryabin T, *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009; 75: 7537–7541.
21 Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 194827: 379–423; 623–656.
22 Borcard D, Gillet F, Legendre P. Numerical ecology with R. Springer Science & Business Media, 2011.
23 Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2 (-delta delta C(T)) method. *Methods* 2001; 25: 402–408.
24 Eishi Y, Suga M, Ishige I, *et al.* Quantitative analysis of mycobacterial and propionibacterial DNA in lymph nodes of Japanese and European patients with sarcoidosis. *J Clin Microbiol* 2002; 40: 198–204.
25 Ishige I, Usui Y, Takemura T, *et al.* Quantitative PCR of mycobacterial and propionibacterial DNA in lymph nodes of Japanese patients with sarcoidosis. *Lancet* 1999; 354: 120–123.
26 Zimmermann A, Zissel G, Müller-Quernheim J, *et al.* Are bronchoalveolar lavages a good source for microbial profiling? Differences between throat and bronchoalveolar lavage microbiomes. *J Med Microbiol* 2015; 64: 948–951.
27 Nguyen LDN, Viscogliosi E, Delhaes L. The lung mycobiome: an emerging field of the human respiratory microbiome. *Front Microbiol* 2015; 6: 89.
28 Salter SJ, Cox MJ, Turek EM, *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014; 12: 87.
29 Esposito S, Marchese A, Tozzi AE, *et al.* DNA bacterial load in children with bacteremic pneumococcal community-acquired pneumonia. *Eur J Clin Microbiol Infect Dis* 2013; 32: 877–881.
30 Chang JY, Antonopoulos DA, Kalra A, *et al.* Decreased diversity of the fecal microbiome in recurrent *Clostridium difficile*-associated diarrhea. *J Infect Dis* 2008; 197: 435–438.
31 Charlson ES, Bittinger K, Haas AR, *et al.* Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* 2011; 184: 957–963.
32 Hsiao WW, Li KL, Liu Z, *et al.* Microbial transformation from normal oral microbiota to acute endodontic infections. *BMC Genomics* 2012; 13: 345.
33 Pragman AA, Kim HB, Reilly CS, *et al.* The lung microbiome in moderate and severe chronic obstructive pulmonary disease. *Ann Am Thorac Soc* 2014; 11: S77–S78.
34 Dubaniewicz A, Dubaniewicz-Wybieralska M, Sternau A, *et al. Mycobacterium tuberculosis* complex and mycobacterial heat shock proteins in lymph node tissue from patients with pulmonary sarcoidosis. *J Clin Microbiol* 2006; 44: 3448–3451.
35 Alam J, Kim YC, Choi Y. Potential role of bacterial infection in autoimmune diseases: a new aspect of molecular mimicry. *Immune Netw* 2014; 14: 7–13.
36 Han YW, Shi W, Huang GT-J, *et al.* Interactions between periodontal bacteria and human oral epithelial cells: Fusobacterium nucleatum adheres to and invades epithelial cells. *Infect Immun* 2000; 68: 3140–3146.
37 Shah C, Moolani Y, Campbell S. Epidemiological characteristics of infections associated with culture isolated Fusobacterium nucleatum: a retrospective analysis. *J Case Rep Stud* 2015; 2: 1.
38 Mima K, Nishihara R, Qian ZR, *et al. Fusobacterium nucleatum* in colorectal carcinoma tissue and patient prognosis. *Gut* 2016; 65: 1973–1980.
39 Strauss J, Kaplan GG, Beck PL, *et al.* Invasive potential of gut mucosa-derived *Fusobacterium nucleatum* positively correlates with IBD status of the host. *Inflamm Bowel Dis* 2011; 17: 1971–1978.
40 Segal LN, Clemente JC, Tsay J-CJ, *et al.* Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nat Microbiol* 2016; 1: 16031.
41 Marshall TG, Marshall FE. Sarcoidosis succumbs to antibiotics–implications for autoimmune disease. *Autoimmun Rev* 2004; 3: 295–300.
42 Bachelez H, Senet P, Cadranel J, *et al.* The use of tetracyclines for the treatment of sarcoidosis. *Arch Dermatol* 2001; 137: 69–73.
43 Tong M, McHardy I, Ruegger P, *et al.* Reprograming of gut microbiome energy metabolism by the FUT2 Crohn's disease risk polymorphism. *ISME J* 2014; 8: 2193–2206.
44 Knights D, Silverberg MS, Weersma RK, *et al.* Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome Med* 2014; 6: 107.
45 Nguyen T, Liu XK, Zhang Y, *et al.* BTNL2, a butyrophilin-like molecule that functions to inhibit T cell activation. *J Immunol* 2006; 176: 7354–7360.
46 Valentonyte R, Hampe J, Huse K, *et al.* Sarcoidosis is associated with a truncating splice site mutation in *BTNL2*. *Nat Genet* 2005; 37: 357–364.