bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Distinctive airway microbial ecology of smokers and asthmatics

Elena M. Turek[1], Michael J. Cox*[1], Michael Hunter[2,3], Jennie Hui[2,3,4], Phillip James[1], Saffron A.G. Willis-Owen[1], Leah Cuthbertson[1], Alan James[3,5], A. William Musk[2,3,6], Miriam F. Moffatt*[1], William O.C.M. Cookson*[1]

[1] Centre for Genomic Medicine, National Heart and Lung Institute, Imperial College London SW3 6LY

[2] School of Population and Global Health, University of Western Australia

[3] Busselton Population Medical Research Institute, Western Australia

[4] PathWest Laboratory Medicine, Queen Elizabeth II Medical Centre, Western Australia

[5] Department of Pulmonary Physiology, Sir Charles Gairdner Hospital, UWA Medical School, University of Western Australia

[6] Department of Respiratory Medicine Sir Charles Gairdner Hospital, UWA Medical School, University of Western Australia

*These authors contributed equally to the study

Correspondence should be addressed to:
Professor William Cookson
National Heart and Lung Institute
Imperial College London
London SW3 6LY, UK
Email: w.cookson@imperial.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The airways of the lung carry microbiota that contribute to respiratory health[1]. The ecology of normal airway microbial communities, their responses to environmental events, and the mechanisms through which they cause or modify disease are poorly understood. Cigarette smoking is the dominant malign environmental influence on lung function, causing 11·5% of deaths globally[2]. Asthma is the most prevalent chronic respiratory disease worldwide[3,4], but was uncommon 100 years ago[5]. The asthma pandemic is linked to urbanization, leading to considerations of protective microbiota loss (the "hygiene hypothesis")[6-8] and acquisition of strains that may damage the airway epithelia[9]. We therefore investigated oropharyngeal airway microbial community structures in a general population sample of Australian adults. We show here that airway bacterial communities were strongly organized into distinctive co-abundance networks ("guilds"), just seven of which contained 99% of all oropharyngeal operational taxonomic units (OTUs). Smoking was associated with diversity loss, negative effects on abundant taxa, profound alterations to network structure and marked expansion of *Streptococcus* spp.. These perturbations may influence chronic obstructive pulmonary disease[10] (COPD) and lung cancer[11]. In contrast to smokers, the loss of diversity in asthmatics selectively affected low abundance but prevalent OTUs from poorly understood genera such as *Selenomonas*, *Megasphaera* and *Capnocytophaga*, without coarse scale network disruption. The results open the possibility that replacement of specific organisms may mitigate asthma susceptibility.

We examined adults from the population of Busselton in Western Australia who were participating in a general health survey[12]. Direct sampling of the lung microbiota requires invasive procedures, such as bronchoscopy, that are not yet possible in epidemiological studies. In healthy individuals, however, the oropharynx and the intra-thoracic airways share similar microbiota[13,14]. We consequently used oropharyngeal swabs for population sampling,

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

accepting that the abundance of pathogens in the lower airways of diseased subjects is imperfectly reflected in the oropharynx[13,15],

We submitted oropharyngeal swabs from 578 subjects to 16S rRNA gene qPCR and sequencing, the latter yielding 44,290,100 high quality reads (Supplementary Figure 1.1 for analysis structure). After removal of 173 OTUs with high probability of being contaminants and 13,472 rare OTUs present in only one sample or with less than 20 reads, there remained 4,218 OTUs derived from 43,775,771 reads. To enable diversity analyses based on proportions, the samples were rarefied to a minimum of 6,543 reads, retaining 529 samples containing 4,005 OTUs and 3,461,247 reads. For consistency, unrarefied data from these same 529 samples were used to test differences between subject groups by DeSEQ, and network analyses. No systematic differences in results were seen if the larger sample was analyzed.

The average age of the 529 subjects was 56 years (Supplementary Table 1). Sixty subjects were current smokers and 216 were ex-smokers (with a mean 18 years since quitting). The mean levels of the forced expiratory volume in one second (FEV1) and the forced vital capacity (FVC) of the subjects were normal. There were 77 doctor-diagnosed asthmatics. There was only one case with a clinical diagnosis of COPD. The frequency of asthma and current smoking were not different to the whole Busselton cohort.

Subjects were not included if they were taking antibiotics within six weeks of the time of study. The annual rate of antibiotic prescription in the Australia population is 254 per 1000, and half of these will be for respiratory infections[16], so it is likely that many smokers will have intermittently taken antibiotics. Asthma is not considered an indication for antibiotics in the Australian healthcare system. Inhaled corticosteroids, used in the maintenance treatment of asthma, do not seem to affect microbial diversity[17].

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Structure of the normal airway microbiome

An estimate of Bray Curtis beta diversity (β) for the population gave the mean dissimilarity in microbial diversity (M) between subjects to be 0.51 ± SD=0.06 (on a scale of 0-1), indicating that on average individual airway microbiomes shared about half of their OTUs.

Five phyla contained 98.4% of all OTUs (Table 1, Supplementary Table 2.2). Firmicutes (predominately *Streptococcus* and *Veilonella* spp.) was the most common phylum, with 24 OTUs in the top 50, and 57.9% of all OTUs found in the complete dataset. Bacteroidetes (predominately *Prevotella* spp.) contained 14.1% of the OTUs, Proteobacteria (predominately *Neisseria* and *Haemophilus* spp.) contained 12.3%, Actinobacterium 9.1% and Fusobacterium 4.9%. Overall, the 50 most abundant OTUs accounted for 92% of the data (Supplementary Table 2.2).

Streptococcus spp. show high rates of clonal diversity and are poorly differentiated by standard culture and 16S sequences[18,19]. We therefore sequenced the methionine aminopeptidase gene (*map*) to further differentiate between *Streptococcus* taxa[19] in 483 subjects. After removal of *map*_OTUs only present in one sample or with fewer than 20 reads or negative correlations with qPCR abundance there remained 14,898 *map*_OTUs, suggesting substantial variation in Streptococcal strains in the population. β diversity estimates in rarefied data (to a level of 7,700 reads) found M = 0.84 ± SD=0.06, indicating low similarity of the streptococcal composition between subjects. The top 9 of the 10 most prevalent *map*_OTUs were identified as *S. salivarius*, with *S. parasanguinis* the next most prevalent. (Supplementary Table 2.6.1). The potential pathogen S. *mitis/pneumoniae* was detected in 58% of subjects, although at low abundance. Tools for further exploration of streptococcal clades are clearly wanted.

Microbial communities are formed through complex ecological interactions that can be exposed through network analyses[20]. On the assumption that correlations in the abundance of

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

different taxa would capture co-ordinated growth, we applied weighted correlation network analyses (WGCNA)[21] to the Busselton dataset.

We observed 13 discrete modules in which the abundance of members was strongly correlated. The WGCNA program labels modules with unique colour identifiers, but we have also named them according to their most abundant genera (Table 2). Just 13 OTUs remained unassigned to a network, referred to as the grey module. The 5 largest modules (in terms of abundance of members) contained contained 97.6% of all OTU sequence reads (Table 2). Individual hubs were stongly connected to their network vectors (range of P = 7.9E-266 to 1.9E-121) (Supplementary Table 2.4.1).

The strengths of association suggest that these co-abundance modules may represent "guilds" of co-operating bacteria that occupy ecological niches on the mucosa.

The largest guild (turquoise module: Prevotella.1) accounted for 42.7% of reads (Table 2, Supplementary Table 2.4.1). The most common organisms were within the genera *Prevotella*, *Veillonella*, *Actinomyces* and *Atopobium*. These organisms resemble common mucosal commensals at other body sites, and perhaps represent a base microbial carpet. The smaller guild (cyan) on the same division (B) of the network dendrogram (Supplementary Figure 1.2) was almost entirely made up of *Veillonella* spp. and may occupy a related ecological niche.

The blue module (Streptococcus.2) contained 21.6% of reads, predominately from the genera *Streptococcus*, *Haemophilus* and *Veillonella*. Network hubs included *Lactobacillales* and *Gemella*. The adjacent network (Neisseria: green) (Supplementary Figure 1.2) was dominated by *Neisseria*, with *Porphyromonas*, and *Capnocytophagia*. This may suggest a normal guild than can be occupied by *Proteobacteria* potential pathogens.

The magenta module (Streptococcus.1) (19.4% of reads) was completely dominated by *Streptococcus* taxa (40%) and an unidentified *Firmicutes* (60%) (Supplementary Table 2.4.1)

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

which is likely also to be streptococcal (based on phylogenetic clustering, not shown). Network hubs were also *Streptococcus*, identifying a streptococcal-specific guild in the mucosa.

Although our results were well powered to map microbial community composition, only limited functions could be surmised by genus assignments and the relationship of the networks to each other (Supplementary Figure 1.2). Metagenomic shotgun sequencing remains problematic for respiratory samples because of high proportions of human DNA, and we accept that a survey of the airway microbiota that includes systematic culture and genome sequencing is desirable.

The airway microbiome and clinical traits

A stepwise regression found that microbial diversity in individual airways was independently related to current cigarette smoking (R^2=6%, P<0.001), a current diagnosis of asthma (additional R^2=1.4%, P<0.005) and packyears of smoking (additional R^2=0.8%, P=0.04) (Supplementary Table 2.3), but not to age or sex. We partitioned the data into three subgroups: smoking + packyears>10 (n=159; asthmatic (n=77); and unaffected (n=233).

Smoking

A DeSEQ analysis to identify changes in specific taxa revealed marked effects of cigarette smoking. (Figure 1, Supplementary Table 2.5.1 and 2.5.2). The loss of diversity affected many abundant OTUs, including those in the genera *Fusobacterium*, *Neisseria*, *Haemophilus*, *Veillonella* and *Gemella*. By contrast, the OTUs increased in smokers were in general highly abundant *Streptococci*. Examination of *map* gene OTUs attributed increases in abundance to *S. parasanguinis* (Fold change 5.2, $P_{adjusted}$=1.75E-07), *S. mitis/pneumoniae* (3.62, 4.81E-09), *S. salivarius* (3.03, 5.59E-15) and *S. thermophilus* (2.53, 7.38E-05) (Supplementary Table 2.6.2).

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

To further explore the impact of smoking and asthma on the higher order structure of the airway microbiome, co-abundance networks were constructed separately in the asthmatic and current smoker portions of the cohort and compared with the full dataset (representing the whole population), limiting the analysis to the 4,207 OTU present in all three datasets.

The network structure of the communities was profoundly altered in current smokers. Whilst the largest guild (Prevotellla.1: commensal carpet) showed relative preservation, other modules showed markedly lower levels of conservation and were strongly positively or negatively associated with smoking status; either in terms of module eigenvectors or hubs (Figure 2, Table 2, Supplementary Table 2.4.1). In smokers, 276 OTUs became disconnected from any module. These most strongly featured *Streptococcus* (70 OTUs), unknown genera (41 OTUs) and *Veillonella* (35 OTUs).

Cigarette smoking has previously been shown to affect the airway microbiota[22], but the extent, magnitude and specificity of disruption here suggests an independent capacity to damage human health. The loss of diversity may predispose smokers to the recurrent infections that lead to COPD[10,22]. Smoking is accompanied by substantial changes in the bowel flora[23] that may mediate smoking influences on inflammatory bowel disease. Bacteria have known roles in the genesis of cancer in general[24] and in lung cancer specifically[11]. *Streptococcus* spp. produce an array of potent toxins that act against human cells or tissues[25], and the expansion of *Streptococcus* clades in smokers might be carcinogenic. Most patients with lung cancer have been heavy smokers and smoking often continues after diagnosis. Our results might also suggest that the local lung microbiota should be considered a factor in lung cancer responses to immuotherapy[26].

Asthma

Microbial diversity loss in asthmatics compared to non-smoking subjects was qualitatively different to the effects of smoking. Two genera (*Neisseria* and *Rothia* OTUs) were increased

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

in numbers ($P_{adjusted}$<0.05) (Figure 3, Supplementary Table 2.5.3). Of these, the *Neisseria* OTU was abundant (4.7% of reads in the population) and showing a 2-fold increase, consistent with increases in *Protebacteria* spp. previously observed in the thoracic airways of asthmatics[13,27].

Eighty-four OTUs were in relatively low abundance amongst asthmatic subjects (Figure 3, Supplementary Table 2.5.4). In marked contrast to smokers, the affected organisms were often in poorly characterized or potentially fastidious genera, including *Leptotrichia*, *Selenomonas*, *Megasphaera* and *Capnocytophaga*. Some representatives of the more common genera *Actinomyces*, *Prevotella* and *Veillonella* were also less abundant. This spectrum does not match organisms shown to be affected by inhaled corticosteroids[17].

These taxa were not concentrated in individual networks. The modules did not correlate with the presence of asthma either in terms of their vectors or their hubs, indicating that the general structure of oropharyngeal microbial communities in asthmatics was preserved (Figure 2).

Divergent (but potentially complementary) theories are offered on possible mechanisms by which microbial diversity might prevent asthma. The "immune deviation" hypothesis suggests that the adaptive immune system needs exposure to infections in order to avoid inappropriate reactions[28]. An extension of this model is that absence of commensal organisms leads to loss of local or systematic tonic signals that normally down-regulate immune responses at mucosal surfaces[29].

Our findings, of reduced numbers of distinctive low-abundance organisms, are consistent with immune modulation by such bacteria. Nevertheless, the lower thoracic airway microbiota in asthmatics consistently show significant excesses of potentially pathogenic *Proteobacteria* [13,27] (or *Streptococcus* spp. in severe disease [13,30,31]), and we detected a significant increase in the abundance of a common Proteobacterial OTU in the Busselton asthmatic subjects. It is therefore also possible that a diverse microbial community protects against asthma through

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

inhibition of colonization by potential pathogens through effects on growth, adherence or biofilm formation[32].

In each case, potential benefits appear likely from manipulating the specific airway microbiota found to be reduced in asthmatics. This provides a strong impetus to isolate and study the organisms that could provide protection.

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Methods

Five hundred and 78 Caucasian adults were recruited through the Busselton Health Study in Western Australia. Individuals with a diagnosis of cancer were excluded. Subjects completed a detailed questionnaire as previously described[12]. Subjects were classified as asthmatic if they answered yes to the question "Has your doctor ever told you that you have asthma". Other diagnoses potentially influencing the microbiome were diabetes (n=18 patients) and gastro-esophageal reflux (GERD, n=36). No associations were found for diabetes or GERD in any analyses, and we included subjects with these diagnoses in the unaffected group.

Samples for microbial analysis were taken under direct vision, using sterile rayon swabs that were rubbed gently with an even pressure around the oropharynx five times, strictly avoiding contact with tonsils, palate or nose. Swabs were immediately frozen and stored at -80°C prior to transportation on dry ice to Imperial College London, UK.

DNA was extracted from swab heads using the MP Bio FastDNA Spin Kit for Soil (http://www.mpbio.com). Blank controls with no sample added were taken from each DNA extraction kit to test for contamination[33]

PCR of the 16S rRNA V4 region was performed in quadruplicate using a custom indexed forward primer S-D-Bact-0564-a-S-15 (5' AYT GGG YDT AAA GNG 3'), reverse primer S-D-Bact-0785-b-A-18 (5' TAC NVG GGT ATC TAA TCC 3') and a high fidelity Taq polymerase master mix (Q5, New England Biolabs, Massachusetts, USA). Primer sequences were based on Klindworth et al.[34], with dual-barcoding as per Kozich et al.[35] with adaptors from Illumina (California, USA). A mock community (Table S1) was included to assess sequencing quality. PCR cycling conditions were: 95ºC for 2 minutes followed by 35 cycles of 95ºC for 20 seconds, 50ºC for 20 seconds and 72ºC for 5 minutes. Amplicons were purified, quantified and equi-molar pooled and the library paired-end sequenced (Illumina MiSeq V2

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

reagent kit) [35]as previously described[36]. Bacterial load was quantified by qPCR using KAPA BioSystems SYBR Fast qPCR Kit with the same 16S rRNA V4 primers used for sequencing.

Analysis of data was carried out in the R environment and details can be followed on github: https://tinyurl.com/y2onjblt. Sequence processing was performed in QIIME (version 1.9.0)[37]. Community level differences in alpha and beta diversity and Operational Taxonomic Unit (OTU) level differences, were analysed using Phyloseq in R (version 3.2.0). A phylogenetic tree was generated from the representative sequences using the default parameters of the make_phylogeny command[37]. Taxonomy of OTUs was assigned by matching representative sequences against release version 23 August 2013 of the Silva database[38] using the default parameters of the assign_taxonomy command[37]. OTUs occurring in only one sample or with less than 20 reads in the whole dataset were removed. Weighted and unweighted UniFrac beta diversity measures and subsequent principle co-ordinates analysis of them was carried out using the beta_diversity_through_plots script[37]. For the purposes of alpha diversity calculations, the raw counts tables were rarefied to a depth of 6,543 reads. Significant differences in alpha diversity between datasets were assessed using Mann–Whitney *U*-tests.

Potential kit contaminant OTUs were identified by the presence of negative Spearman's correlations between OTU abundance and bacterial burden (logged qPCR copy number), adjusted using Bonferroni corrected P-values < 0.05. OTUs subsequently of interest were cross-checked with a listing of potential contaminants[33].

We further differentiated *Streptococcus* spp. by sequencing the methionine aminopeptidase (*map*) gene[19] in 483 samples (constrained to 5 sequencing runs with controls). Of these subjects 234 were never-smoking and 53 were current smokers. We used barcoded primers map-up 5' GCWGACTCWTGTTGGGCWTATGC '3 and map-down 5' TTARTAAGTTCYTTCTTCDCCTTG '3. As positive controls, DNA from nine strains of

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Streptococcus with bacterial identity confirmed through Sanger sequencing was used for positive controls (S. agalactiae (DSMZ-2134); *S. constellatus* subsp. *Constellatus* (DSMZ-20575); *S. infantis* (DSMZ-12492); *S. parasanguinis* (DSMZ-6778); *S. pneumoniae* (DSMZ-20566); *S. pseudopneumoniae* (DSMZ-18670); *S. pyogenes* (DSMZ-20565); *S. sanguinis* (DSMZ-20567); and *S. mitis* (DSMZ-12643)). Analysis was performed in QIIME[37], using a clustering level of 95% to define OTUs. We attributed the most common OTU sequences to Streptococcal species by BLAST searches. Full details are online (http://hdl.handle.net/10044/1/63937).

Co-abundance networks between non-rarefied OTU abundances were analyzed using the WGCNA package[39]. Abundances were log transformed with 0.1 added to zeroes[40], and the topological adjacency matrix was constructed from Spearman's correlation coefficients with a β soft thresholding parameter of 3. Hierarchical clustering of the overlap matrix with dynamic tree cutting defined the co-abundance modules, with a minimum module size set at 20 OTUs. The significance of Spearman's correlation between module eigengenes and clinical variables was adjusted for multiple testing using the Benjamini and Hochberg method[41]. Module structure was contrasted between cohorts using the R package circlize (0.4.5).

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

1 Man, W.H., de Steenhuijsen Piters, W.A., & Bogaert, D., The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol* 15 (5), 259-270 (2017).

2 Collaborators, G.B.D.T., Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. *Lancet* 389 (10082), 1885-1906 (2017).

3 ISAAC, Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. The International Study of Asthma and Allergies in Childhood (ISAAC) Steering Committee. *Lancet* 351 (9111), 1225-1232 (1998).

4 Collaborators, G.B.D.C.R.D., Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet. Respiratory medicine* 5 (9), 691-706 (2017).

5 Burney, P.G., Chinn, S., & Rona, R.J., Has the prevalence of asthma increased in children? Evidence from the national study of health and growth 1973-86. *BMJ* 300 (6735), 1306-1310 (1990).

6 Strachan, D.P., Hay fever, hygiene, and household size. *Brit Med J* 299 (6710), 1259-1260 (1989).

7 Leibowitz, U. *et al.*, Epidemiological study of multiple sclerosis in Israel. II. Multiple sclerosis and level of sanitation. *J Neurol Neurosurg Psychiatry* 29 (1), 60-68 (1966).

8 Bach, J.F., The effect of infections on susceptibility to autoimmune and allergic diseases. *N Engl J Med* 347 (12), 911-920 (2002).

9 Cookson, W.O.C.M., Cox, M.J., & Moffatt, M.F., New opportunities for managing acute and chronic lung infections. *Nat Rev Microbiol* (2017).

10 Donaldson, G.C., Seemungal, T.A., Bhowmik, A., & Wedzicha, J.A., Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax* 57 (10), 847-852 (2002).

11 Jin, C. *et al.*, Commensal Microbiota Promote Lung Cancer Development via gammadelta T Cells. *Cell* (2019).

12 James, A. *et al.*, Rationale, design and methods for a community-based study of clustering and cumulative effects of chronic disease processes and their effects on ageing: the Busselton healthy ageing study. *BMC public health* 13, 936 (2013).

13 Hilty, M. *et al.*, Disordered microbial communities in asthmatic airways. *PLoS One* 5 (1), e8578 (2010).

14 Charlson, E.S. *et al.*, Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* 184 (8), 957-963 (2011).

15 Molyneaux, P.L. *et al.*, The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 190 (8), 906-913 (2014).

16 McCullough, A.R. *et al.*, Antibiotics for acute respiratory infections in general practice: comparison of prescribing rates with guideline recommendations. *Med J Aust* 207 (2), 65-69 (2017).

17 Durack, J. *et al.*, Features of the bronchial bacterial microbiome associated with atopy, asthma, and responsiveness to inhaled corticosteroid treatment. *J Allergy Clin Immunol* 140 (1), 63-75 (2017).

18 Bek-Thomsen, M., Tettelin, H., Hance, I., Nelson, K.E., & Kilian, M., Population diversity and dynamics of Streptococcus mitis, Streptococcus oralis, and

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Streptococcus infantis in the upper respiratory tracts of adults, determined by a nonculture strategy. *Infect Immun* 76 (5), 1889-1896 (2008).

19 Bishop, C.J. *et al.*, Assigning strains to bacterial species via the internet. *BMC biology* 7, 3 (2009).

20 Faust, K. & Raes, J., Microbial interactions: from networks to models. *Nat Rev Microbiol* 10 (8), 538-550 (2012).

21 Tong, M. *et al.*, A modular organization of the human intestinal mucosal microbiota and its association with inflammatory bowel disease. *PLoS One* 8 (11), e80702 (2013).

22 Charlson, E.S. *et al.*, Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS One* 5 (12), e15216 (2010).

23 Biedermann, L. *et al.*, Smoking cessation alters intestinal microbiota: insights from quantitative investigations on human fecal samples using FISH. *Inflamm Bowel Dis* 20 (9), 1496-1501 (2014).

24 Chang, A.H. & Parsonnet, J., Role of bacteria in oncogenesis. *Clin Microbiol Rev* 23 (4), 837-857 (2010).

25 Barnett, T.C. *et al.*, Streptococcal toxins: role in pathogenesis and disease. *Cell Microbiol* 17 (12), 1721-1741 (2015).

26 Routy, B. *et al.*, Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* 359 (6371), 91-97 (2018).

27 Huang, Y.J. *et al.*, Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. *J Allergy Clin Immunol* 127 (2), 372-381 e373 (2011).

28 Holt, P.G., Sly, P.D., & Bjorksten, B., Atopic versus infectious diseases in childhood: a question of balance? *Pediatr Allergy Immunol* 8 (2), 53-58 (1997).

29 Simpson, J.L. *et al.*, Airway dysbiosis: Haemophilus influenzae and Tropheryma in poorly controlled asthma. *Eur Respir J* (2015).

30 Huang, Y.J. *et al.*, The airway microbiome in patients with severe asthma: Associations with disease features and severity. *J Allergy Clin Immunol* 136 (4), 874-884 (2015).

31 Zhang, Q. *et al.*, Airway Microbiota in Severe Asthma and Relationship to Asthma Severity and Phenotypes. *PLoS One* 11 (4), e0152724 (2016).

32 Koo, H., Allan, R.N., Howlin, R.P., Stoodley, P., & Hall-Stoodley, L., Targeting microbial biofilms: current and prospective therapeutic strategies. *Nat Rev Microbiol* 15 (12), 740-755 (2017).

33 Salter, S.J. *et al.*, Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology* 12, 87 (2014).

34 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 (1), e1 (2013).

35 Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., & Schloss, P.D., Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79 (17), 5112-5120 (2013).

36 Ahmed, B. *et al.*, Comparison of the upper and lower airway microbiota in children with chronic lung diseases. *PLoS One* 13 (8), e0201156 (2018).

37 Navas-Molina, J.A. *et al.*, Advancing our understanding of the human microbiome using QIIME. *Methods Enzymol* 531, 371-444 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

38 Quast, C. *et al.*, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41 (Database issue), D590-596 (2013).

39 Langfelder, P. & Horvath, S., WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559 (2008).

40 Friedman, J. & Alm, E.J., Inferring correlation networks from genomic survey data. *PLoS Comput Biol* 8 (9), e1002687 (2012).

41 Benjamini, Y. & Hochberg, Y., Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B.* 57, 289-300 (1995).

Acknowledgements

The study was funded by a Wellcome Senior Investigator Award to WOCC and MFM (P46009), and the Asmarley Trust. The Busselton Healthy Ageing Study is supported by grants from the Government of Western Australia (Office of Science, Department of Health) and the City of Busselton, and from private donations to the Busselton Population Medical Research Institute. We thank the WA Country Health Service and the community of Busselton for their ongoing support and participation.

Author contributions

Overall study design: AWM, AJ, MH, JH, MJC, MFM, and WOCMC. Busselton Survey and sample collection: MH, JH, AJ, AWM. Microbial analysis strategy (laboratory and bioinformatics); MJC, MFM, PJ, EMT. Laboratory experiments: EMT, with assistance and advice by MJC, PJ, LC and MFM. Primary ecological analyses EMT with input from MJC and PJ. Network analysis EMT and SWO. Secondary analyses WOCC. EMT, MFM and WOCMC wrote the first draft of the paper. All authors have read and contributed to the final version of the paper.

Competing Interests

The authors have no competing interests to declare.

Materials & Correspondence

The raw data is available online at the European Nucleotide Archive at the European Bioinformatics Institute, with the accession number PRJEB29091

The R scripts for analysis are available at https://tinyurl.com/y2onjblt

Tables

Table 1 Principal phyla and genera of airway bacteria in a general population sample

Phylum	Genus	Abundance*
Firmicutes (53.4%)		
	Streptococcus	18.92%
	Veillonella	13.74%
	Unidentified_Firmicutes	11.79%
	Selenomonas	1.71%
	Gemella	1.64%
	Granulicatella	1.45%
	Johnsonella	0.70%
	Lachnoanaerobaculum	0.69%
	Megasphaera	0.66%
	Not known	0.46%
	Stomatobaculum	0.43%
	Oribacterium	0.43%
	Solobacterium	0.23%
	Peptostreptococcus	0.17%
	Peptococcus	0.16%
	Parvimonas	0.16%
	Butyrivibrio	0.10%
	Catonella	0.05%
	Filifactor	0.05%

Phylum	Genus	Abundance*
Bacteroidetes (17.7%)		
	Prevotella	15.36%
	Porphyromonas	1.45%
	Capnocytophaga	0.73%
	Tannerella	0.09%
	Bergeyella	0.08%
Fusobacteria (8.5%)		
	Fusobacterium	4.40%
	Leptotrichia	4.09%
Proteobacteria (8.5%)		
	Neisseria	4.59%
	Haemophilus	3.48%
	Not known	0.33%
	Campylobacter	0.07%
Actinobacteria (7.2%)		
	Actinomyces	4.62%
	Atopobium	2.11%
	Rothia	0.36%
	Bifidobacterium	0.08%
Other (0.25%)		

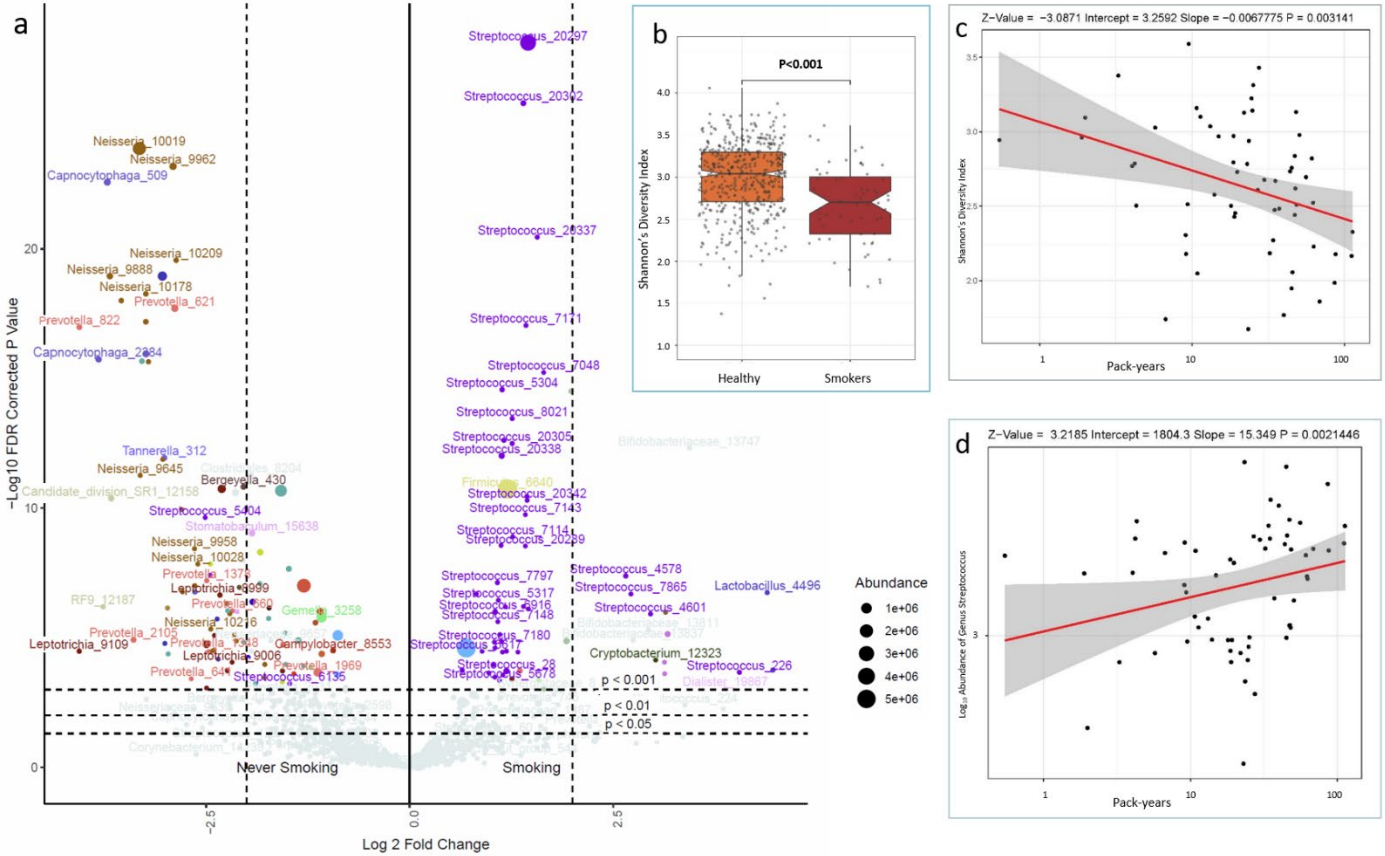
*Abundance based on total 43,652,299 high-quality sequence reads in 529 subjects

Table 2. WGCNA module summary and eigenvector associations with smoking

Module ID (Colour)	Number of OTUs	Total Abundance	Overall %	Cum %	Smoking R	Smoking *P*	Module description
Prevotella.1 (Turquoise)	2218	18,636,985	42.69	42.69			Commensal carpet: *Veilonella*, *Prevotella*, *Actinomyces*. *Veillonella* and *Atopobium* hubs
Streptococcus.2 (Blue)	472	9,433,313	21.61	64.30	-0.13	2.E-02	*Streptococcus* and *Haemophilus* prevalent. *Lactobacilliae* and *Gemella* hubs
Streptococcus.1 (Magenta)	126	8,480,289	19.43	83.73	0.18	8.E-04	*Streptococci* dominated
Fusobacteria (Brown)	583	3,099,110	7.10	90.83	-0.26	4.E-08	*Fusobacteria* and *Leptotrichia* hubs
Neisseria (Green)	204	2,969,651	6.80	97.63	-0.35	1.E-14	*Neisseria* dominated, prevalent *Capnocytophagia*
Prevotella.2 (Black)	136	387,098	0.89	98.52	0.15	4.E-03	*Prevotella*, *Parvimonas*, *Streptococci*, *Porphryomonas*
Veillonella (Cyan)	50	173,186	0.40	98.92	0.17	1.E-03	*Veillonella*
Prevotella.3 (Purple)	71	105,630	0.24	99.16			*Prevotella* dominated
Indeterminate (Tan)	55	101,284	0.23	99.39	0.15	4.E-03	*Prevotella* and *Treponema*
Porphymonas (Salmon)	50	89,951	0.21	99.60	0.15	6.E-03	*Porphyromonas* and *Prevotella*
Bifidobacteria (Pink)	134	86,562	0.20	99.80	0.32	8.E-13	*Bifidobacterium* hubs
Peptococcus (Midnightblue)	44	79,920	0.18	99.98	-0.16	2.E-03	*Peptococcus*
Contaminants (GreenYellow)	62	8,869	0.02	100.00	-0.12	3.E-02	*Herbaspirillum*: potential contaminants
Unconnected (Grey)	13	451	0.00				Unconnected OTUs: potential contaminants

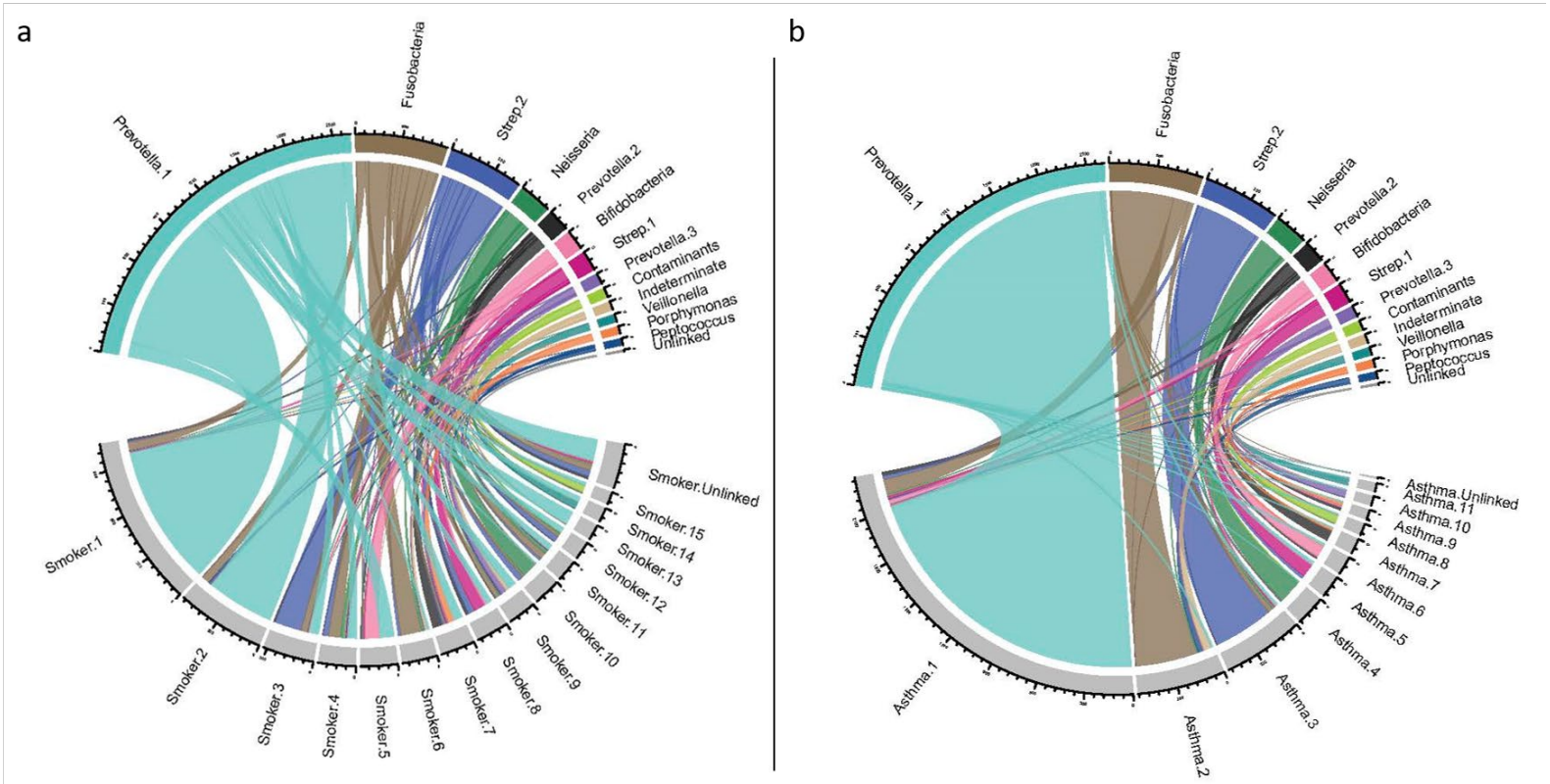
Figures

Figure 1. Smoking and the airway microbiome



a) The volcano plot shows significant differences in the abundance of OTUs between current smokers and the rest of the population. Fold change is shown on the x axis and -log10 *P* (FDR corrected) on the y axis. Relative abundances are reflected in the data point sizes; **b**) shows differences in alpha diversity between smokers and never smokers; **c**) shows how the extent and duration of smoking (packyears) progressively reduces diversity and **d**) shows progressive increase in the abundance of *Streptococcus* OTUs with increasing packyears.

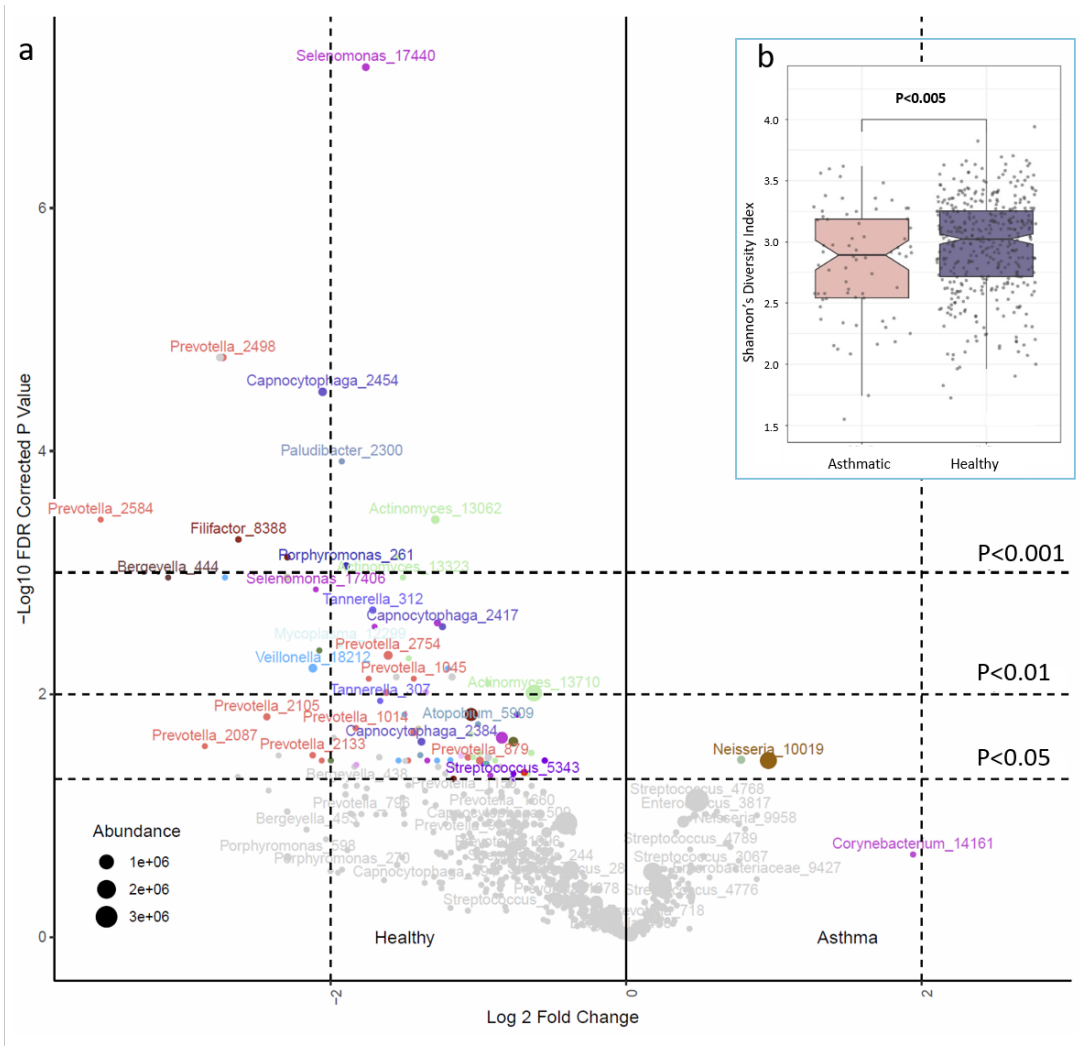
Figure 2. Network structure of the airway microbiome in normal subjects, compared to smokers and asthmatics



The Chord plots show sharing and discordance of 4,207 OTUs common to the three datasets for co-abundance networks. **a)** Network membership in the whole population (top half of plot) compared to current smokers (bottom half of plot); and **b)** compared to asthmatics. Module colours are arbitrarily assigned by WGCNA, and module bacterial names are derived from Table 2. Modules in smokers and asthmatics are simply named by size (Smoking.1, Asthma.2, etc.). There is a marked change of structure with fragmentation of major networks in the smokers, but high conservation of network membership between asthmatics and the whole cohort.

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 3. Asthma and the airway microbiome



a) The volcano plot shows significant differences in the abundance of OTUs between asthmatics and non-smoking subjects with less than 10 packyears of lifetime exposure. Fold change is shown on the x axis and -log10 *P* (FDR corrected) on the y axis. Relative abundances are reflected in the data point sizes; **b)** shows differences in alpha diversity between asthmatics and unaffected non-smoking subjects.

bioRxiv preprint doi: https://doi.org/10.1101/583559; this version posted March 21, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Author Affiliations

EMT, MJC, PJ, SAWO, LC, MFM, WOCMC. Centre for Genomic Medicine, National Heart and Lung Institute, Imperial College London SW3 6LY

MH, JH, AWM: School of Population and Global Health, University of Western Australia

MH, JH, AJ, AWM. Busselton Population Medical Research Institute, Western Australia

JH: PathWest Laboratory Medicine, Queen Elizabeth II Medical Centre, Western Australia

AJ: Department of Pulmonary Physiology, Sir Charles Gairdner Hospital, UWA Medical School, University of Western Australia

AWM: Department of Respiratory Medicine Sir Charles Gairdner Hospital, UWA Medical School, University of Western Australia

Supplementary Material

PDF 854KB