Online Supplement for: Genetic mannose binding lectin deficiency is associated with airway microbiota diversity and reduced exacerbation frequency in COPD

Alison J. Dicker[1]; Megan L. Crichton[1]; Andrew J. Cassidy[1]; Gill Brady[2]; Adrian Hapca[3]; Roger Tavendale[4]; Gisli G. Einarsson[5]; Elizabeth Furrie[6]; J. Stuart Elborn[5,7]; Stuart Schembri[1]; Sara E. Marshall[1]; Colin N.A. Palmer[4]; James D. Chalmers[1]*

1: Scottish Centre for Respiratory Research, University of Dundee, Ninewells Hospital & Medical School, Dundee, UK

2: Tayside Clinical Research Centre, James Arrott Drive, Dundee, UK

3: Dundee Epidemiology and Biostatistics Unit, University of Dundee, Ninewells Hospital & Medical School, Dundee, UK

4: Pat MacPherson Centre for Pharmacogenetics and Pharmacogenomics, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK

5: Centre for Experimental Medicine, School of Medicine, Dentistry and Biomedical Sciences, Queen's University, Belfast, Lisburn Road, Belfast, UK

6: Immunology, NHS Tayside, Ninewells Hospital and Medical School, Dundee, UK

7: National Heart and Lung Institute, Imperial College London, UK

***Corresponding Author**: James D Chalmers, Scottish Centre for Respiratory Research, University of Dundee, Ninewells Hospital & Medical School, Dundee DD1 9SY, UK, jchalmers@dundee.ac.uk, phone 01382383642

Contributions: All authors participated in study design, data analysis and interpretation of the data. All authors were involved in writing and revising the article prior to submission.

Funding: Chief Scientist Office, Scotland. Grant number ETM/262. James D Chalmers acknowledges fellowship support from the Wellcome Trust. Go-SHARE was supported by the Wellcome Trust (099177/Z/12/Z).

Conflicts of interest: SEM is now an employee of the Wellcome Trust

METHODS

SNP Genotyping

Go-TARDIS DNA samples were genotyped for four *MBL2* SNPs (promotor region SNP rs7096206, represented by X or Y, and exon 1 SNPs rs1800450, rs1800451, rs5030737 represented by A or O) which previous studies have used to define genetic MBL deficiency and are known to have the greatest impact on MBL serum levels. MBL SNP genotyping was carried out using custom Sequenom iPLEX gold assays on a Sequenom Massarray analyser for the rs7096206 and rs1800451 SNPs, with any individuals with a failed result for these SNPs being repeated using TaqMan realtime PCR assays (Life Technologies) on an Applied Biosystems 7900HT machine. All genotyping of rs5030737 and rs1800450 was carried out using TaqMan realtime assays. These SNPs were combined as YO/YO or XA/YO for MBL deficient, YA/YO or XA/XA for intermediate and YA/YA or XA/YA for high expressing haplotypes [1-3].

Microbiota Sub-cohort Study

Patients with known MBL genotype identified through Go-TARDIS and a second population based genetic registry (GoSHARE; http://www.goshare.org.uk/) were identified and invited to participate in a mechanistic sub-study. Patients were included if >40 years; with a FEV_1/FVC ratio <70% and a clinical diagnosis of COPD. Exclusion criteria included; the inability to give informed consent; previous adverse reaction to nebulised hypertonic saline; asthma; primary diagnosis of bronchiectasis; Cystic Fibrosis; active mycobacterial disease; and immunosuppression. Patients receiving long term antibiotic therapy or maintenance oral corticosteroid therapy were also excluded; additionally, patients needed to be clinically stable and free of antibiotic or corticosteroid therapy for 4 weeks prior to enrolment. The study was approved by the East of Scotland Research Ethics Committee (13/ES/0030) and all patients gave written informed consent.

Patients were enrolled during a period of disease stability and followed up for between 3 to 6 months during which patients completed EXACT-PRO diary cards and contacted the study team in the event of an exacerbation defined as 2 or more consecutive days where a patient reports 2 or 3 major symptoms (increased dyspnoea, sputum purulence or volume) or one major symptom plus and minor symptom (new or increased nasal discharge, wheeze, sore throat, fever or cough). At each visit, patients provided sputum samples after nebulisation with 5% hypertonic saline for 5 minutes according to standard procedures, sputum colour was assessed by utilisation of the charts described in [4]. Spirometry was performed according to ERS guidelines [5], blood samples were collected at each visit. Quality of life and symptoms were evaluated using the St. Georges Respiratory Questionnaire (SGRQ) and the COPD Assessment Test (CAT). At the onset of exacerbation, patients were assessed by the study team, samples collected, and given a standardised treatment of 40mg prednisolone for 7 days and doxycycline (200mg on day 1, followed by 100mg daily for 6 days, modified in the case of allergy or isolation of a pathogen not susceptible to doxycycline). Patients returned for a follow-up visit 10 days after start of exacerbation to confirm resolution of exacerbation. End of study visits were carried out during a period of disease stability when the patient had been free of antibiotic therapy for 14 days.

Systemic and Airway Inflammation

Sputum biomarkers (IL-1β, IL-8 and TNFα, EN-RAGE, Myeloperoxidase) were measured by ELISA or chromogenic assays as previously described.[6] Serum biomarkers (CRP, CD40L, P-selectin) were measured by commercial ELISAs (R+D systems, Abingdon, UK).

Sputum Microbiota

DNA was extracted from plugs selected from whole sputum, previously frozen at -80°C within 2 hours of collection, using the AllPrep DNA/RNA Mini kit on the QIAcube automation platform (QIAGEN) as follows: Whole sputum was incubated in an equal volume of 1 in 10

diluted Sputolysin (Calbiochem) in a shaking incubator for 30mins at 37°C, mixed with Buffer RLT as per the AllPrep kit protocol, then passed through QIAshredder columns (QIAGEN) with the resulting supernatant undergoing sequential DNA and RNA extraction on the QIAcube. Quality and quantity of the DNA was determined by Nanodrop and Qubit machine, using the Qubit dsDNA broad range kit (Thermo Scientific). Reagent controls were included to remove potential contamination. Illumina MiSeq sequencing of the bacterial 16S rRNA gene was performed following the recommended Illumina protocol using the primers S-D-Bact-0341-b-S-17 and S-D-Bact-0785-a-A-21 from [7], which target the V3 and V4 region of the 16S rRNA gene. Nextera XT Indices were added to each sample to allow multiplexing and the libraries sequenced using 2 x 300 paired end sequencing on the MiSeq platform using a MiSeq V3 kit (Illumina). Following sequencing on the Illumina MiSeq platform bioinformatic analysis of sequences were performed using the Quantitative Insights Into Microbial Ecology (QIIME) pipeline, version 1.9.0.[8] FastQ files were imported into QIIME and quality of reads checked; any reads with a Phred quality score less than Q20 were removed in QIIME when paired end reads were joined together for each sample. Un-joined reads were excluded from subsequent analysis. Sequences were clustered into operational taxonomic units (OTUs) based on 97% sequence similarity using the UCLUST algorithm [9], aligned against the Greengenes Core reference alignment (Version 13.8) [10] using PyNAST (Version 1.2.2) [11]. Taxonomy of the OTUs was assigned using the Ribosomal Database Project Classifier (Version 2.2) with the *de novo* OTU picking option [12]. OTUs were filtered to remove singletons and unassigned OTUs, or OTUs identified as Eukaryota, Human and Cyanobacteria. The dataset was normalised to the lowest number of OTUs and the Shannon-Wiener Diversity Index (S-WDI) of the samples determined as a measure of the richness and evenness of the microbial community. Identification of OTUs in the *Haemophilus* genera was carried out to species level by taking a representative sequence from each OTU and BLASTn

(https://blast.ncbi.nlm.nih.gov/Blast.cgi) searching it against the NCBI 16S Ribosomal RNA (Bacteria and Archaea) database. The top scoring hit (or hits if more than 1 hit had the same top score) was recorded and a consensus species identity determined by comparing the NCBI result with the QIIME result. Species assignments were validated by comparing against culture results in 20 individuals with positive sputum samples for *H. influenzae* and *H. parainfluenzae* resulting in 100% concordance between microbiota dominant species and culture identification. Statistical analysis of microbiota data was carried out using the non-parametric Spearman correlation and Mann Whitney test, where appropriate. All sequences generated were submitted to the NCBI Sequence Read Archive, accession number SRP073159.

Bacterial Binding Studies

The solid phase binding assay was performed as described in Lynch *et al*.[13] Bacterial strains were *P*. aeruginosa strain PA01, *S. aureus* strain RN6390, and clinical isolates of *H. influenzae* (designated B142), *S.* pneumoniae (designated B44), M. catarrhalis (B148), *E.coli* (BL21) and *B. cenocepacia (*kindly provided by Professor John Govan, University of Edinburgh). Bacteria were heat killed at 60^{o}C for 1 hour prior to use. Nunc maxisorb microtitre plates were coated with Mannan (positive control for MBL binding), acetylated bovine serum albumin (positive control for ficolin 2 and ficolin 3 binding) or bacteria as indicated in coating buffer (15mM Na_2CO_3, 35mM $NaHCO_3$, pH 9.6). After overnight incubation, wells were blocked with 300µl 1% human serum albumin (HSA) in TBS (10mM Tris-Cl, 140mM NaCl, pH 7.4) for 1 hour and subsequently washed with TBS/0.05% Tween 20 containing 5mM $CaCl_2$. Recombinant MBL, Ficolin-2 and Ficolin-3 (purchased from R+D systems) were diluted in 100µl of 10mM Tris-Cl, 140mM NaCl, 2mM $CaCl_2$, 0.1% HAS, pH 7.4. Diluted samples were added to the wells and incubated for 2 hours at 37^{o}C. After washing 4 times with wash buffer, bound lectins were detected with biotinylated secondary antibodies followed by streptavidin-HRP. After repeating the wash steps as above, 200µl of 3,3',5,5'-Tetramethylbenzidine was added to each

well for 30 minutes. The reaction was stopped with 100µl of HCL and absorbance measured at 450nm with subtraction of absorbance at 650nm using a microplate reader. Results are expressed as a percentage of the binding (optical density) achieved with the positive control.

Supplementary Results

Table E1: Patient metrics of the full study cohort compared to the sputum-producing producing sub-cohort. P values calculated through pairwise comparisons by T test for parametric data and Mann Whitney U test for non-parametric data. *indicates X^2 test across all 4 GOLD groups. CABG = coronary artery bypass graft, CCF = congestive cardiac failure, SGRQ = St. Georges Respiratory Questionnaire, LTOT = long term oxygen therapy, ICS = inhaled corticosteroid, LABA = long acting beta agonist, LAMA = long acting muscarinic antagonist, ARB = angiotensin receptor blocker, ACE = angiontensin converting enzyme.

	Full Cohort	Cohort giving at least 1 sputum sample	P value
Demographics and major comorbidities			
N	141	99	
MBL deficient (%)	34 (24.1)	24 (24.2)	
Age (SD)	71.3 (±8.0)	71.3 (±8.3)	1.0
Age at diagnosis (SD)	60 (±11.3)	59.77 (±11.5)	0.9
Male gender (%)	89 (63.1)	66 (66.7)	0.6
Active smokers (%)	29 (20.6)	24 (24.2)	0.5
Ex smokers (%)	99 (70.2)	71 (71.7)	0.8
Pack years (SD)	42 (±29.8)	42 (±29.0)	1.0
BMI (SD)	28.2 (±5.5)	28.3 (±5.7)	0.9
Myocardial Infarction (%)	16 (11.4)	15 (15.2)	0.4
CABG (%)	15 (10.6)	14 (14.1)	0.4
Angina (%)	30 (21.3)	24 (24.2)	0.6
Stroke (%)	13 (9.2)	9 (9.1)	1.0
Diabetes (%)	28 (19.9)	19 (19.2)	0.9
Cancer (%)	3 (2.1)	3 (3.0)	0.7
CCF (%)	7 (5.0)	5 (5.1)	1.0
Lung surgery (%)	6 (4.3)	4 (4.0)	0.9
COPD severity			
FEV_1 Percent Predicted (SD)	70.38 (±20.6)	70.34 (±21.7)	0.9
MRC Dyspnoea Score (SD)	2.68 (±1.4)	2.78 (±1.4)	0.6
Exacerbations per year (SD)	1.8 (±1.9)	2.1 (±2.0)	0.2
GOLD Score			0.6
A (%)	17 (12.1)	7 (7.1)	-
B (%)	52 (36.9)	36 (36.4)	-
C (%)	6 (4.3)	5 (5.1)	-
D (%)	66 (46.8)	51 (51.5)	-
SGRQ (SD)	41.63 (±23.0)	44.45 (±22.1)	0.3
On LTOT (%)	5 (3.6)	5 (5.1)	0.6
Medications			
Statins (%)	76 (53.9)	54 (54.6)	0.9
Inhaled Steroids (%)	84 (59.6)	61 (61.6)	0.8

LABA (%)	15 (10.6)	13 (13.1)	0.6
LAMA (%)	70 (49.7)	53 (53.5)	0.6
Theophylline (%)	9 (6.4)	8 (8.1)	0.6
Mucolytic (%)	17 (12.1)	13 (13.1)	0.8
Aspirin (%)	37 (26.2)	26 (26.3)	1.0
Beta Blocker (%)	18 (12.8)	13 (13.1)	0.9
ACE-inhibitor (%)	37 (26.2)	27 (27.3)	0.9
ARB (%)	12 (8.5)	6 (6.1)	0.5
Clopidogrel (%)	10 (7.1)	8 (8.1)	0.8
Anticoagulant (%)	11 (7.8)	5 (5.1)	0.4

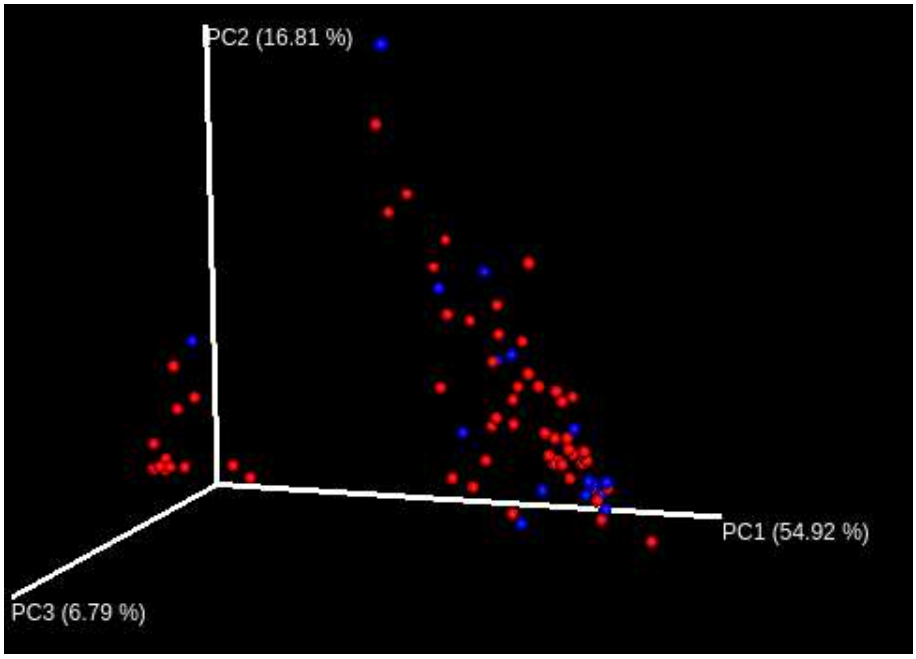

Figure E1: Beta diversity of all stable samples coded according to MBL genotype shown by weighted UNIFRAC PCoA plot, red representing non-deficient patients and blue MBL deficient patients.

Representativeness of the sample

The cohort providing DNA for analysis were compared to subjects included in the main TARDIS cohort and eligible for inclusion in the genetic study. We showed no clinical important differences in terms of age, gender, COPD severity, lung function or co-morbidity between the included patients and the overall cohort (Table E2).

Table E2. Comparison of selected clinical variables between the "parent" cohort TARDIS and the included patients according to MBL genotype.

	MBL deficient genotype	MBL intermediate genotype	MBL sufficient genotype	Overall eligible TARDIS cohort
N (% of study cohort)	240 (13.4)	586 (32.6)	970 (54.0)	4305 (100%)
Age at diagnosis (SD)	65.6 (± 9.0)	64.23 (± 9.6)	64.4 (± 9.6)	65.6 (± 9.5)
Male Gender (% of group)	131 (54.6)	308 (52.7)	506 (52.2)	2209 (51.3%)
Cigarette smoking by pack years (SD)	41.6 (± 20.6)	41.0 (± 20.3)	41.3 (± 22.7)	39.0 (± 21.8)
Body Mass Index (SD)	27.0 (± 5.6)	26.9 (± 5.4)	27.2 (± 5.6)	26.5 (± 5.6)
FEV_1 % predicted (SD)	79.6 (±20.8)	78.7 (±24.4)	78.0 (±22.1)	76.4 (± 22.0)
MRC dyspnoea score (SD)	2.44 (±1.0)	2.53 (±1.0)	2.48 (±1.0)	2.57 (±1.0)
FEV_1/FVC (SD)	0.59 (± 0.1)	0.58 (± 0.1)	0.58 (± 0.1)	0.58 (±0.1)
Cardiovascular disease (% of group)	24 (10.0)	53 (9.0)	78 (8.0)	532 (12.7)
Renal failure (% of group)	5 (2.1)	17 (2.9)	19 (2.0)	121 (2.9)

REFERENCES

1. Garred P, Honore C, Ma YJ, et al. MBL2, FCN1, FCN2 and FCN3-The genes behind the initiation of the lectin pathway of complement. *Mol Immunol* 2009;46(14):2737-44.
2. Chalmers JD, Fleming GB, Hill AT, et al. Impact of mannose-binding lectin insufficiency on the course of cystic fibrosis: A review and meta-analysis. *Glycobiology* 2011;21(3):271-82.
3. Chalmers JD, McHugh BJ, Doherty C, et al. Mannose-binding lectin deficiency and disease severity in non-cystic fibrosis bronchiectasis: a prospective study. *Lancet Respir Med* 2013;1(3):224-32.
4. Murray MP, Pentland JL, Turnbull K, et al. Sputum colour: a useful clinical tool in non-cystic fibrosis bronchiectasis. *Eur Respir J* 2009;34(2):361-64.
5. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J* 2005;26(2):319-38.
6. Chalmers JD, Smith MP, McHugh BJ, et al. Short- and Long-Term Antibiotic Treatment Reduces Airway and Systemic Inflammation in Non-Cystic Fibrosis Bronchiectasis. *Am J Respir Crit Care Med* 2012;186(7):657-65.
7. Klindworth A, Pruesse E, Schweer T, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 2013;41(1).
8. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7(5):335-6.
9. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26(19):2460-1.
10. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environl Microbiol* 2006;72(7):5069-72.
11. Caporaso JG, Bittinger K, Bushman FD, et al. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 2010;26(2):266-7.
12. Wang Q, Garrity GM, Tiedje JM, et al. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environl Microbiol* 2007;73(16):5261-7.
13. Lynch NJ, Roscher S, Hartung T, et al. L-ficolin specifically binds to lipoteichoic acid, a cell wall constituent of Gram-positive bacteria, and activates the lectin pathway of complement. *J Immunol* 2004;172(2):1198-202.