American Journal of Respiratory Cell and Molecular Biology/AJRCMB

Whole Blood Gene Expression in Pulmonary Non-tuberculous Mycobacterial Infection

Journal:	*American Journal of Respiratory Cell and Molecular Biology*
Manuscript ID	Red-2017-0230OC.R1
Manuscript Type:	OC - Original Contribution
Date Submitted by the Author:	12-Nov-2017
Complete List of Authors:	Cowman, Steven; National Heart and Lung Institute, Imperial College London; Royal Brompton Hospital, Department of Host Defence Jacob, Joseph; National Heart and Lung Institute, Imperial College London; Royal Brompton Hospital, Department of Radiology Hansell, David; National Heart and Lung Institute, Imperial College London; Royal Brompton Hospital, Department of Radiology Kelleher, Peter; Centre for Immunology and Vaccinology, Imperial College London; Royal Brompton Hospital, Department of Host Defence Wilson, Robert ; National Heart and Lung Institute, Imperial College London; Royal Brompton Hospital, Department of Host Defence Cookson, William; National Heart and Lung Institute, Imperial College London Moffatt, Miriam; National Heart and Lung Institute, Imperial College London Loebinger, Michael; National Heart and Lung Institute, Imperial College London; Royal Brompton Hospital, Department of Host Defence
Subject Category:	11.05 Non-Tuberculous Mycobacterial Disease < MYCOBACTERIAL DISEASES
Keywords:	

Title:

Whole blood gene expression in pulmonary non-tuberculous mycobacterial infection

Authors:

Steven A. Cowman[1,2], Joseph Jacob[1,3], David M. Hansell[1,3], Peter Kelleher[1,2], Robert Wilson[1,2], William O. C. Cookson[1], Miriam F. Moffatt*[1] and Michael R. Loebinger*[1,2].

1 National Heart and Lung Institute, Imperial College London, London, UK

2 Host Defence Unit, Royal Brompton Hospital, London, UK

3 Department of Radiology, Royal Brompton Hospital, London, UK

*These authors contributed equally.

Correspondence:

Correspondence should be addressed to:

Dr Michael Loebinger,

Department of Host Defence,

Royal Brompton Hospital,

Sydney Street,

London

SW3 6NP

UK

Tel: +442073518337

Fax: +442073518338

Email: M.Loebinger@rbht.nhs.uk

Author contributions:

S.A.C. collected samples and phenotypic information, performed the gene expression experiments and subsequent data analysis and composed the manuscript. R.W. treated the patients, contributed phenotypic information and critical input. D.H. and J.J. scored the CT scans. W.O.C.C and M.F.M. advised on experimental and study design and provided expertise in gene expression and data analysis. P.K. helped analyse immunological data. M.R.L. treated the patients, contributed phenotypic information and conceived and designed the study. All authors contributed to the final manuscript.

Funding:

S.A.C. and M.R.L. were funded by a grant from the Welton Foundation. Additional funding was provided by the Asmarley Trust and the Wellcome Trust. Support was also provided by the NIHR Respiratory Disease Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust and Imperial College London. D.M.H. is the recipient of a National Institute of Health Research Senior Investigator Award.

Running title:

Gene expression in pulmonary NTM infection

Descriptor:

11.5 Non-Tuberculous Mycobacterial Disease

Word count:

4040/5000 words

At a Glance Commentary

Scientific Knowledge on the Subject:

Pulmonary non-tuberculous mycobacterial infection (pNTM) is increasingly common and associated with a high mortality, although the clinical course is unpredictable. The factors predisposing towards the development of disease and its subsequent progression remain unclear. Impaired immune responses have been reported in individuals with pNTM but data are limited and inconsistent.

What This Study Adds to the Field:

The expression of many genes mediating cellular immunity was depressed in pNTM disease, including *IFNG*, which correlated negatively with markers of disease severity. Mortality was associated with the expression of transcripts related to the innate immune response and inflammation, whereas transcripts relating to T and B cell function were associated with improved survival. These findings suggest that pNTM is associated with an altered immune response which may both contribute to the development of disease as well as its severity. The expression of genes identified in this study warrant further investigation as potential markers of disease activity and suggest targets for potential therapeutic intervention.

This article has an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org

Abstract

Rationale: The factors predisposing towards the development of pulmonary non-tuberculous mycobacterial disease (pNTM) and influencing disease progression remain unclear. Impaired immune responses have been reported in individuals with pNTM but data are limited and inconsistent.

Objectives: To use gene expression profiling to examine the host response to pNTM.

Methods: Microarray analysis of whole blood gene expression was performed on 25 subjects with pNTM and 27 uninfected controls with respiratory disease. Gene expression results were compared to phenotypic variables and survival data.

Measurements and Main Results: Compared with uninfected controls, pNTM was associated with down-regulation of 213 transcripts enriched for terms related to T cell signalling including *IFNG*. Reduced *IFNG* expression was associated with more severe CT changes and impaired lung function. Mortality was associated with the expression of transcripts related to the innate immune response and inflammation, whereas transcripts related to T and B cell function were associated with improved survival.

Conclusions: These findings suggest that pNTM is associated with an aberrant immune response which may reflect an underlying propensity to infection, or result from NTM infection itself. There were important differences in the immune response associated with survival and mortality in pNTM.

Abstract word count: 193/250 words.

Key words: Nontuberculous mycobacteria, gene expression profiling

Introduction

Pulmonary non-tuberculous mycobacterial disease (pNTM) is an increasingly common and challenging infection. Whilst ubiquitous in the environment, disease usually occurs in the context of an existing pulmonary disease such as bronchiectasis or COPD, although in a subgroup of patients disease occurs in the absence of any known risk factors. The disease is associated with a high mortality ranging between 12.5% to 41.1% at five years (1-5). The clinical course of disease is variable and unpredictable, whilst some individuals develop progressive disease, some may remain stable without treatment (1, 6, 7). The factors governing the acquisition of disease and its subsequent clinical course are poorly understood.

Disseminated NTM disease occurs in the context of immunodeficiency such as AIDS or inherited and acquired defects of the IFN-γ/IL-12 pathway, and several studies have also demonstrated functional defects in the host immune response of subjects with pNTM (8-13). A recent study employing exome sequencing revealed an excess of variants affecting genes implicated in the immune response in subjects with pNTM compared with both the control population and unaffected family members (14). These findings suggest that defective immune responses may play a role in predisposing individuals to pNTM. The aims of this study were to investigate the host response to pNTM using global gene expression profiling of peripheral blood and to explore the relationship between gene expression and clinical outcomes.

Methods

Individuals with pNTM were recruited from the Royal Brompton Hospital and Chelsea and Westminster Hospital between September 2012 and November 2013. All participants met ATS 2007 criteria and were unanimously deemed to have active disease requiring treatment by two clinicians with expertise in pNTM (authors ML and RW). Control subjects with respiratory disease (bronchiectasis or COPD) but no radiological or microbiological evidence of mycobacterial disease were recruited from the same departments. Written consent was gained from all participants and the study was approved by the local Research Ethics Committee (reference 12/LO/1034).

Participants underwent clinical assessment and pulmonary function testing. Peripheral blood was taken for RNA extraction and clinical testing and spontaneously expectorated sputum was collected for phenol auramine microscopy and bacterial, mycobacterial and fungal cultures. If not performed within the previous six months, a HRCT was performed. For NTM cases CT scoring was performed by a specialist radiologist blinded to clinical details as described previously (15) and a composite score calculated by summing the scores for individual features. Further methods are provided in the online data supplement.

Gene expression analysis

Peripheral blood was collected into PAXgene RNA tubes (Becton, Dickinson and Company, NJ, USA) and RNA extracted using the PAXgene Blood RNA

Kit (PreAnalytiX, Hombrechtikon, Switzerland). Extracted RNA was amplified to complementary DNA using the Ovation Pico WTA System V2 (NuGEN Technologies, San Carlos, USA), then cDNA was fragmented and labelled using the Encore Biotin Module (NuGEN Technologies) and immediately hybridised to Affymetrix Human Gene 1.1 ST array plates and scanned using the GeneTitan instrument (Affymetrix, Santa Clara, USA). Samples were randomised prior to amplification and again prior to fragmentation, labelling and hybridisation.

Outliers were identified using the arrayQualityMetrics package (16) in the R environment version 3.1.2 and removed. Raw probe level data from .CEL files passing quality control were summarised, quantile normalised and $\log_2$ transformed using the Affymetrix Power Tools package. Expression data were re-imported into R where low-expressed and unannotated probes were removed and all downstream analyses performed.

Differentially expressed genes were identified using Significance Analysis of Microarrays (SAM) (17) applying a false discovery rate (FDR) of 5% for significance. Gene ontology analyses were performed using the WebGestalt tool (http://www.webgestalt.org) using the Gene Ontology Consortium (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. All expressed genes were used as the reference set, an adjusted P value of 0.05 was deemed significant.

Weighted Gene Co-expression Gene Network Analysis (WCGNA) was performed using the WGCNA package (18) with the default recommended parameters. All expressed genes were entered into the analysis, however only genes with a connectivity in the highest tertile were used to construct the topological overlap matrix.

Correlation between gene (or gene module) expression and clinical variables was performed using the Pearson or Spearman correlation for continuous variables according to their distribution, and polyserial correlation for categorical variables. The t-test and analysis of variance (ANOVA) were used to compare expression values between groups. The log-rank test was used to compare differences in survival curves.

Results

A total of 52 subjects (25 pNTM cases and 27 controls) matched for age, sex and ethnicity were studied. Clinical characteristics of the study cohort are detailed in Table 1 with further details provided in Tables E2 and E3 of the online data supplement. The proportion of subjects with a diagnosis of bronchiectasis was lower (although not significantly so) in the group of NTM cases due to the inclusion of subjects with “Lady Windermere’ syndrome. These individuals (N=7) had no underlying lung disease prior to their NTM infection, although at the time of enrolment bronchiectasis was present in all subjects on HRCT.

Gene expression and NTM disease

A total of 213 transcripts were identified as differentially expressed between cases and controls with a FDR of 5% (for full list see Table E4 in the online data supplement and Figure E1). All transcripts were down-regulated in pNTM disease. The top 25 genes with the highest fold change are detailed in Table 2.

One of the top differentially expressed genes in pNTM cases compared with controls was interferon gamma (*IFNG*) (Figure 1). Interferon gamma (IFN-γ) is the primary effector cytokine of the Th1 pathway which plays a vital role in mycobacterial immunity. In light of this it is notable that other differentially expressed genes included Nuclear factor of activated T cells 2 (*NFATC2*), Cytotoxic and regulatory T cell molecule (*CRTAM*) and X-C Motif Chemokine Ligand 1 and 2 (*XCL1*, *XCL2*). *NFATC2* encodes NFATc2, a member of the nuclear factor of activated T cells (NFAT) family of transcription factors. NFATc2 has been shown to be an *in vitro* regulator of both IFN- γ (19, 20) and TNF- α (21, 22) release. The product of *CRTAM* also promotes IFN-γ release from CD8+ T-cells (23-27) and NK cells (28, 29). *XCL2* encodes a chemokine which is chemotactic for CD8+ and CD4+ T-cells (30, 31) and its homologue, *XCL1*, encodes a chemokine associated with the Th1 response, the silencing of which has been shown to lead to reduced IFN-γ production and disordered granuloma formation in TB infected mice (32, 33).

Gene ontology (GO) analysis of the set of differentially expressed transcripts revealed enrichment for one biological process, “Negative regulation of alpha-

beta T cell proliferation" (enrichment ratio 36.01 adjusted P = 0.049) and a further ten GO cellular components. Additionally KEGG analysis revealed enrichment for the pathway "T cell receptor signaling" (enrichment ratio 6.12, adjusted P < 0.001), with nine genes in the pathway (*IFNG, AKT3, CD247, MAPK9, CBLB, NFATC2, ICOS, RASGRP1* and *ITK*) being significantly under-expressed in pNTM cases. Full details are given in Table E5 of the online data supplement.

To examine the influence of systemic steroid use and NTM treatment on the observed difference in gene expression, multiple linear modelling was performed using these as additional explanatory variables in addition to NTM disease. The expression of 168 of the 213 genes (79%) remained significantly associated with NTM disease (see Table E6 in the online data supplement).

The 213 transcripts were compared with the 380 gene 'meta-signature' of active tuberculosis identified by Blankley *et al*. (34). There was an overlap of only nine genes (*CDK5RAP2, EPHA4, FAIM3, FCRL3, GPR183, GZMK, KLF12, NELL2* and *SMA4*), most of which were also downregulated in tuberculosis except *CDK5RAP2* and *SMA4* which showed increased expression. Comparison with studies of sarcoidosis, another granulomatous disease, revealed that 11 of the 213 transcripts (*AKT3, CBLB, CCND2, CD247, ICOS, IFNG, IL2RB, ITK, MAPK9, NFATC2* and *RASGRP1*) are also present in the 31-gene sarcoid signature described by Zhou et al. (35) representing significant (P < 0.001) enrichment. However, none of the 20 genes in the experimentally derived 'unbiased' signature from this study were present and

there was little or no overlap with other reported gene signatures in sarcoidosis (36, 37).

In order to identify modules associated with pNTM, WGCNA was performed using all 13,360 expressed transcripts from the 52 subjects resulting in 10 modules being identified (detailed in Table E7 and E8 of the online data supplement). Three modules showed significantly ($P \leq 0.05$) different expression in pNTM cases compared to controls. These were enriched for multiple GO terms including those relating to T cell selection, differentiation, activation and signalling, and others associated with IFN-γ activity such as 'Natural killer (NK) cell mediated cytotoxicity', 'Antigen processing and presentation', and 'MHC class II protein complex' (Table 3). In all cases the expression of genes within each module was lower in subjects with NTM disease. Of the 213 genes identified to be differentially expressed in pNTM cases compared to controls, 114 (53.5%) were present in one of the three modules that showed association with pNTM.

IFNG expression and clinical outcomes

Given the essential role of IFN-γ in the immune response to mycobacteria, the relationship between *IFNG* expression in pNTM cases and clinical variables was explored further.

The expression of *IFNG* was significantly negatively correlated with several measures of lung function including FEV1 % predicted (Pearson's r = 0.43 P = 0.05), FVC % predicted (Pearson's r = 0.51 P = 0.019) and TLC % predicted

(Pearson's r = 0.48 P = 0.016). No significant correlations however were seen for transfer factor for carbon monoxide (TLCO) or the transfer factor corrected for alveolar volume (KCO).

Expression of *IFNG* was strongly negatively correlated with combined CT score (Figure 2a, Pearson's r = -0.53, P = 0.007). When individual CT features were examined (see Figure E2 in the online data supplement), *IFNG* expression was significantly lower in the presence of moderate or severely extensive bronchiectasis compared with a limited extent (P = 0.017 and P = 0.029 respectively).

In addition *IFNG* expression was associated with other markers of disease severity: it was positively correlated with serum albumin (Spearman's ρ = 0.50, P = 0.011), and negatively correlated with CRP (Spearman's ρ = -0.42 P = 0.034) and neutrophil count (Pearson's r = -0.52 P = 0.007). Interestingly, *IFNG* was also negatively correlated with time since diagnosis (Figure 2b, Spearman's ρ -0.60 P = 0.005) although there was no correlation between the combined CT score and disease duration (Spearman's ρ 0.24 P = 0.287).

No significant differences were seen in mean *IFNG* expression between survivors and non-survivors, although when subjects were dichotomised into low and high *IFNG* expression there was a trend towards reduced survival in the low expression group. This however did not reach significance (log-rank test P = 0.06).

For disease control subjects, no significant associations between any measures of lung function or markers of disease severity and *IFNG* expression were observed (see table E9 in the online data supplement), suggesting the changes were specific to NTM disease and not merely due to disease severity in general.

Gene expression and survival in NTM disease

Survival data was available for 24 of the pNTM cases who had a mortality rate of 33% (8 deaths) during the study. No deaths were observed in the control group. Survival analysis was performed on these 24 cases using SAM with an FDR of 5%, 215 genes were found to show association with decreased survival and 1,131 genes with increased survival. The top enriched GO terms in relation to survival are given in Table 4 (for full results of the analysis please see Tables E10-12 and Figure E3 in the online data supplement).

The top genes most strongly associated with reduced survival included genes forming part of the inflammasomes (Absent in melanoma 2 [*AIM2*], Caspase 1 [*CASP1*] and NLR family CARD domain containing 4 [*NLRC4*]) or related to them (Caspase 4 [*CASP4*] and caspase recruitment domain family member 16 and 17 [*CARD16*, *CARD17*]). The set of genes associated with reduced survival was enriched for 50 GO terms, many of which were related to the innate immune response and inflammation, including the IPAF and AIM2 inflammasome complexes. Additionally the genes were enriched for six KEGG pathways including “Toll-like receptor signaling pathway”.

The top genes most strongly associated with increased survival included *CD4*, the T-helper cell co-receptor; *CD28*, a co-receptor which mediates T cell activation; and *IL2RB*, the beta subunit of the receptor for interleukin 2 which plays a key role in the regulation of lymphocyte proliferation and activation, and is also of importance in NK cell function through its roles as a subunit of the interleukin 15 receptor. The set of genes associated with improved survival was enriched for 26 GO terms, the majority of which were related to T cell and B cell function, and four KEGG pathways including "Antigen processing and presentation".

WGCNA was repeated focussing only on the pNTM cases in order to investigate disease survival. Nine modules were identified, three of which showed significant differential expression in non-survivors compared with survivors (Table 5 and Tables E13 and E14 of the online data supplement). Of the 215 genes associated with reduced survival by SAM, 74 (34.6%) were present in either the 'white' or 'grey60' modules, whilst 666 (58.9%) of the 1,131 genes associated with increased survival were present in the 'darkorange' module.

Of note is the enrichment of the grey60 module for the GO term "RAGE receptor binding". RAGE (the receptor for advanced glycation endproducts) is a pattern recognition receptor which has been implicated in neutrophilic inflammation in tuberculosis (38, 39). Whilst grey60 module itself was not significantly correlated with neutrophil count (Pearson's r = 0.29, P = 0.152) many of the genes associated with RAGE were found to be significantly

correlated with neutrophilia (*HMBG2*: Pearson's r 0.5 P = 0.011, *S100A8*: Pearson's r 0.56 P = 0.003, *S100A9*: Pearson's r 0.56 P = 0.003, *S100A12*: Pearson's r 0.53 P = 0.006).

Discussion

In this study we have shown that NTM infection is associated with the differential expression of over two hundred genes. These were characterised by the reduced expression of many genes associated with cellular immunity, some of which have previously been identified to be important in the host response to mycobacterial infection. One of the top differentially expressed genes was *IFNG*, which plays an essential role in antimycobacterial immunity. Not only was *IFNG* expression lower in subjects with pNTM, but within this group reduced *IFNG* expression was significantly associated with more severe radiological disease and impaired lung function. There was also a trend towards an increased mortality in subjects with lower *IFNG* expression.

These data are consistent with experimental evidence from *in vitro* studies which have demonstrated diminished Type 1 cytokine responses in patients with pNTM, showing reduced IL-12 (8-10) and IFN-γ (8-13) release after stimulation with both mitogen and NTM. These studies have employed varying methodologies (different cell types, stimuli and read-outs) and not all such studies have detected a difference in the immune response in subjects with pNTM (40-42). The evidence from this present study therefore provides

valuable unbiased evidence for an impaired IFN-γ response in subjects with pNTM, and is the first study to link this impairment with adverse clinical features.

There is evidence that immunotherapy with IFN-γ may be beneficial in mycobacterial disease. Recombinant IFN-γ has been shown to improve sputum culture conversion in pulmonary TB infection (43) and has been used in refractory cerebral TB (21). Case reports have reported benefit from IFN-γ administration in NTM infection both in the presence (23, 25-27) and absence (28) of immunodeficiency. Two randomised controlled trials have been conducted in pNTM with one finding no benefit with inhaled IFN-γ (30) whilst the other observed a significant increase in treatment response with the addition of intramuscular IFN-γ (32).

These results were supported by the networks analysis, which identified gene modules significantly under-expressed in pNTM which were enriched for several terms linked to IFN-γ activity. In light of reports of immunodeficiency due to anti-cytokine autoantibodies (44) it is also interesting to note that one module was enriched for genes regulating thymic T cell selection.

The 215 genes associated with reduced survival were significantly enriched for multiple terms relating to innate immunity and pattern recognition receptors, including Toll-like receptor (TLR) signalling. Whilst TLR2 may have a protective role in murine models of NTM infection (45-49) and has been reported to display reduced expression in humans with pNTM (8) the pro-

inflammatory response triggered by TLR2 may also be deleterious in NTM infection. The 'rough' variant of *M. abscessus* has been shown to trigger a hyper-inflammatory response via TLR2 signalling (50) which may play a role in the virulence associated with this phenotype (51).

Interestingly, genes associated with inflammasome formation and pyroptosis were also overrepresented in the genes associated with reduced survival. Inflammasome activation has been reported in tuberculosis (52, 53) and has been proposed as a mechanism for the immune reconstitution inflammatory syndrome (IRIS) seen in HIV (54). To date only one report has demonstrated inflammasome activation by NTM (55) and little is known about their role in NTM infection.

Network analysis identified an association between mortality and the expression of genes related to RAGE, including the proteins S100A8 and S100A9 which together form calprotectin, an important pro-inflammatory mediator (56). In a murine model, expression of RAGE in lung tissue has been shown to increase in response to TB infection, and gene knock-out caused an increase in weight loss and mortality (57). The expression of both *S100A8* and *S100A9* have been associated with neutrophilic inflammation in tuberculosis (38, 39) with the expression of *S100A9* being suggested as a potential tuberculosis biomarker (58).

Improved survival was associated with 1131 genes, which were enriched for several GO terms predominantly relating to T-cell differentiation and

activation. Despite the trend towards higher expression in surviving NTM cases *IFNG* was not identified by SAM as being significantly associated with survival, although it was related to many of the enriched functional terms. Again the results of network analysis results were also supportive of these findings and identified a gene module associated with survival that was enriched for terms associated with T-cell development and signalling.

Taken together, these findings suggest that pulmonary NTM disease is associated with the down-regulation of genes involved in the T-cell response, including *IFNG*, which is consistent with the results of *in vitro* studies in subjects with NTM infection. In subjects in whom disease has developed, the expression of genes involved in the T-cell response is associated with a better prognosis, perhaps reflecting the preservation of an appropriate response to NTM infection. Conversely a poor prognosis is associated with an up-regulation of genes involved in innate immunity and inflammation, which may potentially mediate tissue damage and thus the progressive radiological changes observed.

Infection with *M. tuberculosis* has been reported to down-regulate *in vitro* IFN-γ release (59) thus it is not clear whether the differences in gene expression represent a predisposing defect in host response or are a result of immunomodulation following NTM infection. The reduction in *IFNG* seen as disease duration increases may point towards the latter. Similarly the question remains as to whether the up-regulation of the innate immune response associated with poor survival is in some way contributory to pathogenesis,

perhaps due to excessive inflammation and tissue damage, or is merely a response to an increased burden of mycobacterial infection in those with severe disease.

Pulmonary NTM disease is a heterogeneous condition containing a variety of clinical phenotypes, the most clinically relevant distinction being between fibro-cavitary and nodular-bronchiectatic disease. It is likely that the transcriptional response differs between these groups. Given the small sample size no comparison was made between different radiological patterns, however it is notable that *IFNG* expression strongly correlated with composite CT score, and was lower (although not significantly so) in individuals with severe cavitation. Another notable subgroup of individuals with NTM are those which demonstrate stable non-progressive disease in the absence of treatment (1, 6, 7) and it is also likely that their gene expression profiles differ to those seen in the individuals with active infection included in the present cohort.

Several studies have been conducted examining gene expression in tuberculosis, and recently Blankley *et al.* characterised a 380 gene 'meta-signature' of active tuberculosis showing consistent differential expression in at least nine datasets (34). There was little overlap between this signature and the genes associated with pNTM in the current study, and in contrast to pNTM gene expression in tuberculosis was characterised by an upregulation of IFN-γ regulated genes (although not *IFNG* itself). This difference may relate to the infecting organism, or reflect a more appropriate host response to tuberculosis

in a generally younger population with fewer comorbidities as opposed to the potentially aberrant response seen in individuals with pNTM. There was however a point of similarity in that genes related to T-cells were consistently downregulated in both tuberculosis and pNTM.

There was no overlap with experimentally derived expression signatures of sarcoidosis (35-37). However the 213 transcripts associated with pNTM were significantly enriched for genes in the 31-gene sarcoid signature described by Zhou *et al.* (35). These genes were chosen *a priori* as they are part of the T cell receptor, JAK STAT and cytokine-cytokine receptor signalling pathways implicated in sarcoidosis, and were able to differentiate sarcoidosis from controls with 82.2% accuracy. Eleven of these genes (35%) were also associated with pNTM in the current study, all of which were regulated in the same direction (down) in sarcoidosis. This is particularly interesting given the hypothesis that sarcoidosis may be driven by an aberrant immune response to mycobacterial antigens (60).

This study has some limitations. The selection of subjects with definite active disease was necessary to provide a cohort with a well-defined phenotype but has limited the sample size and therefore the statistical power of the study. This may therefore limit the ability to generalise these findings to the wider NTM population. Even with such selection, the NTM study population remains highly heterogeneous, further limiting the power to detect differences in expression. This may explain the relatively modest differences in expression seen between groups in contrast to studies of a more homogenous population

such as those seen in tuberculosis. The use of inhaled or systemic steroids may have an effect on gene expression, however the proportions were similar between NTM cases and controls, and the exclusion of these subjects would necessitate the exclusion of most subjects with COPD, an important group at risk of NTM disease, and would further limit the ability to generalise the findings.

Whilst being matched for basic demographics important phenotypic differences still remained between cases and controls, such as the high mortality seen in cases, systemic steroid use, and the treatment given for NTM. These differences may potentially contribute to the observed differences in gene expression. However, intervening to reduce the differences between groups by altering treatment would not have been feasible on ethical grounds, and the mortality in our cohort is in keeping with the known high mortality associated with NTM disease reported in other cohorts (1-5). Furthermore the expression of the majority of the differentially expressed transcripts (including *IFNG*) was significantly associated with NTM disease independent of systemic steroid use or NTM treatment. Similarly, several other factors were observed to be associated with survival (such as the presence of mycetoma) and the association seen between gene expression and survival may be therefore be confounded by these. Unfortunately the sample size precludes multivariate analyses to investigate these factors further.

Conclusions

This is the first study of gene expression associated with NTM infection. The expression of many genes mediating cellular immunity was depressed in NTM disease, including *IFNG*, which correlated negatively with markers of disease severity. Whilst requiring further validation in independent cohorts, these findings are mechanistically plausible and consistent with previous experimental data. Mortality in NTM disease was associated with reduced expression of genes related to cellular immunity and increased expression of genes related to innate immunity, suggesting that disease progression may be driven by immune dysregulation. To our knowledge this is the first time the immune response to NTM has been linked to differences in disease outcomes. The expression of genes identified in this study warrant further investigation as potential markers of disease activity and suggest targets for potential therapeutic intervention.

Acknowledgements

The authors would like to thank the staff of the Royal Brompton Hospital Respiratory Clinical Research Facility for their help in the collection of samples and phenotypic data.

References

1. Gochi M, Takayanagi N, Kanauchi T, Ishiguro T, Yanagisawa T, Sugita Y. Retrospective study of the predictors of mortality and radiographic deterioration in 782 patients with nodular/bronchiectatic Mycobacterium avium complex lung disease. *BMJ Open* 2015;5:e008058.
2. Hayashi M, Takayanagi N, Kanauchi T, Miyahara Y, Yanagisawa T, Sugita Y. Prognostic Factors of 634 HIV-Negative Patients with Mycobacterium avium Complex Lung Disease. *Am J Respir Crit Care Med* 2012;185:575–583.
3. Andréjak C, Thomsen VØ, Johansen IS, Riis A, Benfield TL, Duhaut P, Sørensen HT, Lescure F-X, Thomsen RW. Nontuberculous pulmonary mycobacteriosis in Denmark: incidence and prognostic factors. *Am J Respir Crit Care Med* 2010;181:514–521.
4. Fleshner M, Olivier KN, Shaw PA, Adjemian J, Strollo S, Claypool RJ, Folio L, Zelazny A, Holland SM, Prevots DR. Mortality among patients with pulmonary non-tuberculous mycobacteria disease. *Int J Tuberc Lung Dis* 2016;20:582–587.
5. Gommans EPAT, Even P, Linssen CFM, van Dessel H, van Haren E, de Vries GJ, Dingemans AMC, Kotz D, Rohde GGU. Risk factors for mortality in patients with pulmonary infections with non-tuberculous mycobacteria: A retrospective cohort study. *Respir Med* 2015;109:137–145.
6. Yamazaki Y, Kubo K, Takamizawa A, Yamamoto H, Honda T, Sone S. Markers indicating deterioration of pulmonary Mycobacterium avium-intracellulare infection. *Am J Respir Crit Care Med* 1999;160:1851–1855.

7. Lee G, Kim HS, Lee KS, Koh W-J, Jeon K, Jeong B-H, Ahn J. Serial CT findings of nodular bronchiectatic Mycobacterium avium complex pulmonary disease with antibiotic treatment. *American Journal of Roentgenology* 2013;201:764–772.

8. Ryu YJ, Kim EJ, Lee S-H, Kim SY, Suh GY, Chung MP, Kim H, Kwon OJ, Koh W-J. Impaired expression of Toll-like receptor 2 in nontuberculous mycobacterial lung disease. *Eur Respir J* 2007;30:736–742.

9. Vankayalapati R, Wizel B, Samten B, Griffith DE, Shams H, Galland MR, Reyn Von CF, Girard WM, Wallace RJ, Barnes PF. Cytokine profiles in immunocompetent persons infected with Mycobacterium avium complex. *J Infect Dis* 2001;183:478–484.

10. Kwon Y-S, Kim EJ, Lee S-H, Suh GY, Chung MP, Kim H, Kwon OJ, Koh W-J. Decreased cytokine production in patients with nontuberculous mycobacterial lung disease. *Lung* 2007;185:337–341.

11. Kartalija M, Ovrutsky AR, Bryan CL, Pott GB, Fantuzzi G, Thomas J, Strand MJ, Bai X, Ramamoorthy P, Rothman MS, Nagabhushanam V, McDermott M, Levin AR, Frazer-Abel A, Giclas PC, Korner J, Iseman MD, Shapiro L, Chan ED. Patients with nontuberculous mycobacterial lung disease exhibit unique body and immune phenotypes. *Am J Respir Crit Care Med* 2013;187:197–205.

12. Greinert U, Schlaak M, Rüsch-Gerdes S, Flad HD, Ernst M. Low in vitro production of interferon-gamma and tumor necrosis factor-alpha in HIV-seronegative patients with pulmonary disease caused by nontuberculous mycobacteria. *J Clin Immunol* 2000;20:445–452.

13. Safdar A, Armstrong D, Murray HW. A novel defect in interferon-gamma secretion in patients with refractory nontuberculous pulmonary mycobacteriosis. *Ann Intern Med* 2003;138:521.

14. Szymanski EP, Leung JM, Fowler CJ, Haney C, Hsu AP, Chen F, Duggal P, Oler AJ, McCormack R, Podack E, Drummond RA, Lionakis MS, Browne SK, Prevots DR, Knowles M, Cutting G, Liu X, Devine SE, Fraser CM, Tettelin H, Olivier KN, Holland SM. Pulmonary Nontuberculous Mycobacterial Infection: A Multisystem Multigenic Disease. *Am J Respir Crit Care Med* 2015;doi:10.1164/rccm.201502-0387OC.

15. Zoumot Z, Boutou AK, Gill SS, van Zeller M, Hansell DM, Wells AU, Wilson R, Loebinger MR. Mycobacterium avium complex infection in non-cystic fibrosis bronchiectasis. *Respirology* 2014;19:714–722.

16. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. *Bioinformatics* 2009;25:415–416.

17. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001;98:5116–5121.

18. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.

19. Kiani A, García-Cózar FJ, Habermann I, Laforsch S, Aebischer T, Ehninger G, Rao A. Regulation of interferon-gamma gene expression by nuclear factor of activated T cells. *Blood* 2001;98:1480–1488.

20. Erb KJ, Twardzik T, Palmetshofer A, Wohlleben G, Tatsch U, Serfling E.

Mice deficient in nuclear factor of activated T-cell transcription factor c2 mount increased Th2 responses after infection with Nippostrongylus brasiliensis and decreased Th1 responses after mycobacterial infection. *Infection and Immunity* 2003;71:6641–6647.

21. Lee J-Y, Yim J-J, Yoon B-W. Adjuvant interferon-γ treatment in two cases of refractory tuberculosis of the brain. *Clin Neurol Neurosurg* 2012;114:732–734.
22. Kaminuma O, Kitamura F, Kitamura N, Hiroi T, Miyoshi H, Miyawaki A, Miyatake S. Differential contribution of NFATc2 and NFATc1 to TNF-alpha gene expression in T cells. *J Immunol* 2008;180:319–326.
23. Squires KE, Murphy WF, Madoff LC, Murray HW. Interferon-gamma and Mycobacterium avium-intracellulare infection. *J Infect Dis* 1989;159:599–600.
24. Valle-Rios R, Patiño-Lopez G, Medina-Contreras O, Canche-Pool E, Recillas-Targa F, Lopez-Bayghen E, Zlotnik A, Ortiz-Navarrete V. Characterization of CRTAM gene promoter: AP-1 transcription factor control its expression in human T CD8 lymphocytes. *Mol Immunol* 2009;46:3379–3387.
25. Squires KE, Brown ST, Armstrong D, Murphy WF, Murray HW. Interferon-gamma treatment for Mycobacterium avium-intracellular complex bacillemia in patients with AIDS. *J Infect Dis* 1992;166:686–687.
26. Holland SM, Eisenstein EM, Kuhns DB, Turner ML, Fleisher TA, Strober W, Gallin JI. Treatment of refractory disseminated nontuberculous mycobacterial infection with interferon gamma. A preliminary report. *N*

Engl J Med 1994;330:1348–1355.

27. Hallstrand TS, Ochs HD, Zhu Q, Liles WC. Inhaled IFN-gamma for persistent nontuberculous mycobacterial pulmonary disease due to functional IFN-gamma deficiency. *Eur Respir J* 2004;24:367–370.

28. Chatte G, Panteix G, Perrin-Fayolle M, Pacheco Y. Aerosolized interferon gamma for Mycobacterium avium-complex lung disease. *Am J Respir Crit Care Med* 1995;152:1094–1096.

29. Mathew PA, Chuang SS, Vaidya SV, Kumaresan PR, Boles KS, Pham H-TK. The LLT1 receptor induces IFN-gamma production by human natural killer cells. *Mol Immunol* 2004;40:1157–1163.

30. Lam PK, Griffith DE, Aksamit TR, Ruoss SJ, Garay SM, Daley CL. Factors related to response to intermittent treatment of Mycobacterium avium complex lung disease. *Am J Respir Crit Care Med* 2006;173:1283–1289.

31. Cao W, Tang S, Yuan H, Wang H, Zhao X, Lu H. Mycobacterium tuberculosis antigen Wag31 induces expression of C-chemokine XCL2 in macrophages. *Curr Microbiol* 2008;57:189–194.

32. Milanes-Virelles MT, Garcia-Garcia I, Santos-Herrera Y, Valdes-Quintana M, Valenzuela-Silva CM, Jimenez-Madrigal G, Ramos-Gomez TI, Bello-Rivero I, Fernandez-Olivera N, Sanchez-De La Osa RB, Rodriguez-Acosta C, Gonzalez-Mendez L, Martinez-Sanchez G, Lopez-Saura PA, MACGAM Study Group. Adjuvant interferon gamma in patients with pulmonary atypical Mycobacteriosis: a randomized, double-blind, placebo-controlled study. *BMC Infect Dis* 2008;8:17.

33. Lei Y, Takahama Y. XCL1 and XCR1 in the immune system. *Microbes*

Infect 2012;14:262–267.

34. Blankley S, Graham CM, Levin J, Turner J, Berry MPR, Bloom CI, Xu Z, Pascual V, Banchereau J, Chaussabel D, Breen R, Santis G, Blankenship DM, Lipman M, O'Garra A. A 380-gene meta-signature of active tuberculosis compared with healthy controls. *European Respiratory Journal* 2016;doi:10.1183/13993003.02121-2015.

35. Zhou T, Zhang W, Sweiss NJ, Chen ES, Moller DR, Knox KS, Ma S-F, Wade MS, Noth I, Machado RF, Garcia JGN. Peripheral blood gene expression as a novel genomic biomarker in complicated sarcoidosis. In: Morty RE, editor. *PLoS ONE* 2012;7:e44818.

36. Koth LL, Solberg OD, Peng JC, Bhakta NR, Nguyen CP, Woodruff PG. Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. *Am J Respir Crit Care Med* 2011;184:1153–1163.

37. Monast CS, Li K, Judson MA, Baughman RP, Wadman E, Watt R, Silkoff PE, Barnathan ES, Brodmerkel C. Sarcoidosis extent relates to molecular variability. In: Hodkinson J, Chapel H, editors. *Clin Exp Immunol* 2017;188:444–454.

38. Marais S, Wilkinson KA, Lesosky M, Coussens AK, Deffur A, Pepper DJ, Schutz C, Ismail Z, Meintjes G, Wilkinson RJ. Neutrophil-associated central nervous system inflammation in tuberculous meningitis immune reconstitution inflammatory syndrome. *CLIN INFECT DIS* 2014;59:1638–1647.

39. Gopal R, Monin L, Torres D, Slight S, Mehra S, McKenna KC, Fallert Junecko BA, Reinhart TA, Kolls J, Báez-Saldaña R, Cruz-Lagunas A, Rodríguez-Reyna TS, Kumar NP, Tessier P, Roth J, Selman M,

Becerril-Villanueva E, Baquera-Heredia J, Cumming B, Kasprowicz VO, Steyn AJC, Babu S, Kaushal D, Zuniga J, Vogl T, Rangel-Moreno J, Khader SA. S100A8/A9 proteins mediate neutrophilic inflammation and lung pathology during tuberculosis. *Am J Respir Crit Care Med* 2013;188:1137–1146.

40. de Jong E, Lim A, Waterer G, Price P. Monocyte-derived macrophages do not explain susceptibility to pulmonary non-tuberculous mycobacterial disease. *Clin Transl Immunology* 2012;1:e2.

41. Lim A, Allison C, Price P, Waterer G. Susceptibility to pulmonary disease due to Mycobacterium avium–intracellulare complex may reflect low IL-17 and high IL-10 responses rather than Th1 deficiency. *Clinical Immunology* 2010;137:296–302.

42. Kim RD, Greenberg DE, Ehrmantraut ME, Guide SV, Ding L, Shea Y, Brown MR, Chernick M, Steagall WK, Glasgow CG, Lin J, Jolley C, Sorbara L, Raffeld M, Hill S, Avila N, Sachdev V, Barnhart LA, Anderson VL, Claypool R, Hilligoss DM, Garofalo M, Fitzgerald A, Anaya-O'Brien S, Darnell D, DeCastro R, Menning HM, Ricklefs SM, Porcella SF, *et al*. Pulmonary nontuberculous mycobacterial disease: prospective study of a distinct preexisting syndrome. *Am J Respir Crit Care Med* 2008;178:1066–1074.

43. Dawson R, Condos R, Tse D, Huie ML, Ress S, Tseng C-H, Brauns C, Weiden M, Hoshino Y, Bateman E, Rom WN. Immunomodulation with recombinant interferon-gamma1b in pulmonary tuberculosis. *PLoS ONE* 2009;4:e6984.

44. MD SKB, MD DSMH. Anticytokine autoantibodies in infectious

diseases: pathogenesis and mechanisms. *Lancet Infect Dis* 2010;10:875–885.

45. Sampaio EP, Elloumi HZ, Zelazny A, Ding L, Paulson ML, Sher A, Bafica AL, Shea YR, Holland SM. Mycobacterium abscessus and M. avium Trigger Toll-Like Receptor 2 and Distinct Cytokine Response in Human Cells. *American journal of respiratory cell and molecular biology* 2008;39:431–439.
46. Feng CG, Scanga CA, Collazo-Custodio CM, Cheever AW, Hieny S, Caspar P, Sher A. Mice lacking myeloid differentiation factor 88 display profound defects in host resistance and immune responses to Mycobacterium avium infection not exhibited by Toll-like receptor 2 (TLR2)- and TLR4-deficient animals. *J Immunol* 2003;171:4758–4764.
47. Gomes MS, Sousa Fernandes S, Cordeiro JV, Silva Gomes S, Vieira A, Appelberg R. Engagement of Toll-like receptor 2 in mouse macrophages infected with Mycobacterium avium induces non-oxidative and TNF-independent anti-mycobacterial activity. *Eur J Immunol* 2008;38:2180–2189.
48. Carvalho NB, Oliveira FS, Durães FV, de Almeida LA, Flórido M, Prata LO, Caliari MV, Appelberg R, Oliveira SC. Toll-like receptor 9 is required for full host resistance to Mycobacterium avium infection but plays no role in induction of Th1 responses. *Infection and Immunity* 2011;79:1638–1646.
49. Shin D-M, Yang C-S, Yuk J-M, Lee J-Y, Kim KH, Shin SJ, Takahara K, Lee SJ, Jo E-K. Mycobacterium abscessus activates the macrophage innate immune response via a physical and functional interaction

between TLR2 and dectin-1. *Cell Microbiol* 2008;10:1608–1621.

50. Roux A-L, Ray A, Pawlik A, Medjahed H, Etienne G, Rottman M, Catherinot E, Coppée J-Y, Chaoui K, Monsarrat B, Toubert A, Daffé M, Puzo G, Gaillard J-L, Brosch R, Dulphy N, Nigou J, Herrmann J-L. Overexpression of proinflammatory TLR-2-signalling lipoproteins in hypervirulent mycobacterial variants. *Cell Microbiol* 2011;13:692–704.
51. Catherinot E, Clarissou J, Etienne G, Ripoll F, Emile J-F, Daffé M, Perronne C, Soudais C, Gaillard JL, Rottman M. Hypervirulence of a rough variant of the Mycobacterium abscessus type strain. *Infection and Immunity* 2007;75:1055–1058.
52. Hagar JA, Miao EA. Detection of cytosolic bacteria by inflammatory caspases. *Current Opinion in Microbiology* 2014;17:61–66.
53. Man SM, Karki R, Kanneganti T-D. AIM2 inflammasome in infection, cancer, and autoimmunity: Role in DNA sensing, inflammation, and innate immunity. *Eur J Immunol* 2016;46:269–280.
54. Tan HY, Yong YK, Shankar EM, Paukovics G, Ellegård R, Larsson M, Kamarulzaman A, French MA, Crowe SM. Aberrant Inflammasome Activation Characterizes Tuberculosis-Associated Immune Reconstitution Inflammatory Syndrome. *The Journal of Immunology* 2016;doi:10.4049/jimmunol.1502203.
55. Lee H-M, Yuk J-M, Kim KH, Jang J, Kang G, Park JB, Son J-W, Jo E-K. Mycobacterium abscessus activates the NLRP3 inflammasome via Dectin-1-Syk and p62/SQSTM1. *Immunol Cell Biol* 2012;90:601–610.
56. Ehrchen JM, Sunderkötter C, Foell D, Vogl T, Roth J. The endogenous Toll-like receptor 4 agonist S100A8/S100A9 (calprotectin) as innate

amplifier of infection, autoimmunity, and cancer. *J Leukoc Biol* 2009;86:557–566.

57. van Zoelen MAD, Wieland CW, van der Windt GJW, Florquin S, Nawroth PP, Bierhaus A, van der Poll T. Receptor for advanced glycation end products is protective during murine tuberculosis. *Mol Immunol* 2012;52:183–189.

58. Xu D, Li Y, Li X, Wei L-L, Pan Z, Jiang T-T, Chen Z-L, Wang C, Cao W-M, Zhang X, Ping Z-P, Liu C-M, Liu J-Y, Li Z-J, Li J-C. Serum protein S100A9, SOD3, and MMP9 as new diagnostic biomarkers for pulmonary tuberculosis by iTRAQ-coupled two-dimensional LC-MS/MS. *Proteomics* 2015;15:58–67.

59. O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MPR. The immune response in tuberculosis. *Annu Rev Immunol* 2013;31:475–527.

60. Brownell I, Ramírez-Valle F, Sanchez M, Prystowsky S. Evidence for mycobacteria in sarcoidosis. *American journal of respiratory cell and molecular biology* 2011;45:899–905.

Figure Legends

Figure 1. Boxplots of *IFNG* expression in pNTM cases (grey) and controls (white). The central line, box boundaries and whiskers represent the median, first and third quartiles and the lowest and highest values lying within 1.5 times the interquartile range, respectively.

Figure 2 a) Expression of *IFNG* and combined CT score in NTM cases b) Expression of *IFNG* and duration of NTM disease.

Tables

	NTM cases N = 25	Controls N = 27
Age (years)	66.9 (±10.3)	65.4 (±7.5)
Male sex (%)	9 (36%)	10 (37%)
Ex- or current smoker (%)	14 (56%)	14 (52%)
Underlying diagnosis (%)		
COPD	7 (28%)	7 (26%)
Bronchiectasis	11 (44%)	20 (74%)
No underlying lung disease	7 (28%)	0
FEV1 % predicted	48.64 (±31.42)	71.27 (±25.98)
FVC % predicted	73.78 (±39.2)	102.25 (±20.87)
Prophylactic antibiotics (%)	6 (24%)	8 (29%)
Systemic corticosteroids (%)	3 (12%)	2 (7%)
Inhaled corticosteroids (%)	12 (48%)	17 (63%)
NTM treatment (%)	9 (36%)	-
NTM species		
MAC	14 (56%)	-
M. abscessus	4 (16%)	-
M. xenopi	3 (12%)	-
M. kansasii	2 (8%)	-
M. malmoense	1 (4%)	-
M. fortuitum	1 (4%)	-

Table 1. Clinical characteristics of the study population. Continuous variables are stated as means ± standard deviations. COPD = chronic obstructive pulmonary disease, FEV1 = forced expiratory volume in 1 second, FVC = forced vital capacity, MAC = *M. avium* complex.

Rank	Gene name	Entrez ID	Fold change	*d*-statistic	Name
1	*FLJ45825*	100505530	1.657	2.646	uncharacterized LOC100505530
2	*GZMK*	3003	1.61	2.857	granzyme K (granzyme 3; tryptase II)
3	*TARP*	445347	1.609	2.619	TCR gamma alternate reading frame protein
4	*TARP*	445347	1.609	2.619	TCR gamma alternate reading frame protein
5	*XCL2*	6846	1.553	3.269	chemokine (C motif) ligand 2
6	*A2M*	2	1.507	2.293	alpha-2-macroglobulin
7	*CRTAM*	56253	1.465	3.57	cytotoxic and regulatory T cell molecule
8	*PMS2P1*	5379	1.453	4.053	postmeiotic segregation increased 2 pseudogene 1
9	*FCRL3*	115352	1.446	3.26	Fc receptor-like 3
10	*PZP*	5858	1.435	2.275	pregnancy-zone protein
11	*TIGIT*	201633	1.4	3.333	T cell immunoreceptor with Ig and ITIM domains
12	*PPIH*	10465	1.393	2.897	peptidylprolyl isomerase H (cyclophilin H)
13	*AK5*	26289	1.39	2.515	adenylate kinase 5
14	*MUC12*	10071	1.39	2.388	mucin 12, cell surface associated
15	*IFNG*	3458	1.382	2.622	interferon, gamma
16	*FAHD2A*	51011	1.364	3.021	fumarylacetoacetate hydrolase domain containing 2A
17	*SMA4*	11039	1.361	3.602	glucuronidase, beta pseudogene
18	*VSIG1*	340547	1.356	2.784	V-set and immunoglobulin domain containing 1
19	*SAMD3*	154075	1.35	2.367	sterile alpha motif domain containing 3
20	*XCL1*	6375	1.348	2.247	chemokine (C motif) ligand 1
21	*IL2RB*	3560	1.347	2.242	interleukin 2 receptor, beta
22	*PSPH*	5723	1.339	2.232	phosphoserine phosphatase

23	*LDHB*	3945	1.336	2.312	lactate dehydrogenase B
24	*NELL2*	4753	1.323	2.276	NEL-like 2 (chicken)
25	*NFATC2*	4773	1.322	2.804	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2

Table 2. The top twenty five differentially expressed genes between pNTM cases and controls. Genes are ranked by fold-change, all displayed decreased expression in pNTM cases.

Module name	**Number of genes**	**Correlation with NTM disease**	***P* value**	**Relevant enriched GO terms**
Cyan	1,139	-0.31	0.02	Thymic T cell selection T cell receptor signalling pathway Antigen processing and presentation
Greenyellow	85	-0.32	0.02	Cellular defence response Positive regulation of immune response T cell activation T cell receptor complex Natural killer cell mediated cytotoxicity CCR1 chemokine receptor binding
Lightcyan	37	-0.27	0.05	B cell activation B cell receptor complex MHC class II protein complex

Table 3. Gene modules identified as significantly associated with NTM disease by WGCNA.

Top GO terms		Enrichment ratio	Adjusted *P* value
Associated with decreased survival			
response to other organism	GO:0051707	3.16	0.001
response to biotic stimulus	GO:0009607	3.00	0.001
activation of innate immune response	GO:0002218	5.56	0.001
detection of molecule of bacterial origin	GO:0032490	30.71	0.001
regulation of innate immune response	GO:0045088	4.23	0.001
Associated with increased survival			
T cell selection	GO:0045058	6.19	1×10^{4}
lymphocyte differentiation	GO:0030098	2.37	6×10^{4}
thymic T cell selection	GO:0045061	6.97	6×10^{4}
immune system development	GO:0002520	1.74	0.002
leukocyte differentiation	GO:0002521	2.01	0.002

Table 4. The top 5 most significantly enriched GO terms in genes associated with decreased and increased survival in NTM disease.

Module name	**Number of genes**	**Correlation with mortality**	***P* value**	**Relevant enriched GO terms**
darkorange	1,413	-0.5	0.01	T cell receptor signalling pathway Thymic T cell selection,
grey60	490	0.55	0.004	RAGE receptor binding
White	71	0.46	0.02	None

Table 5. Gene modules identified as significantly associated with decreased and increased survival in NTM disease.

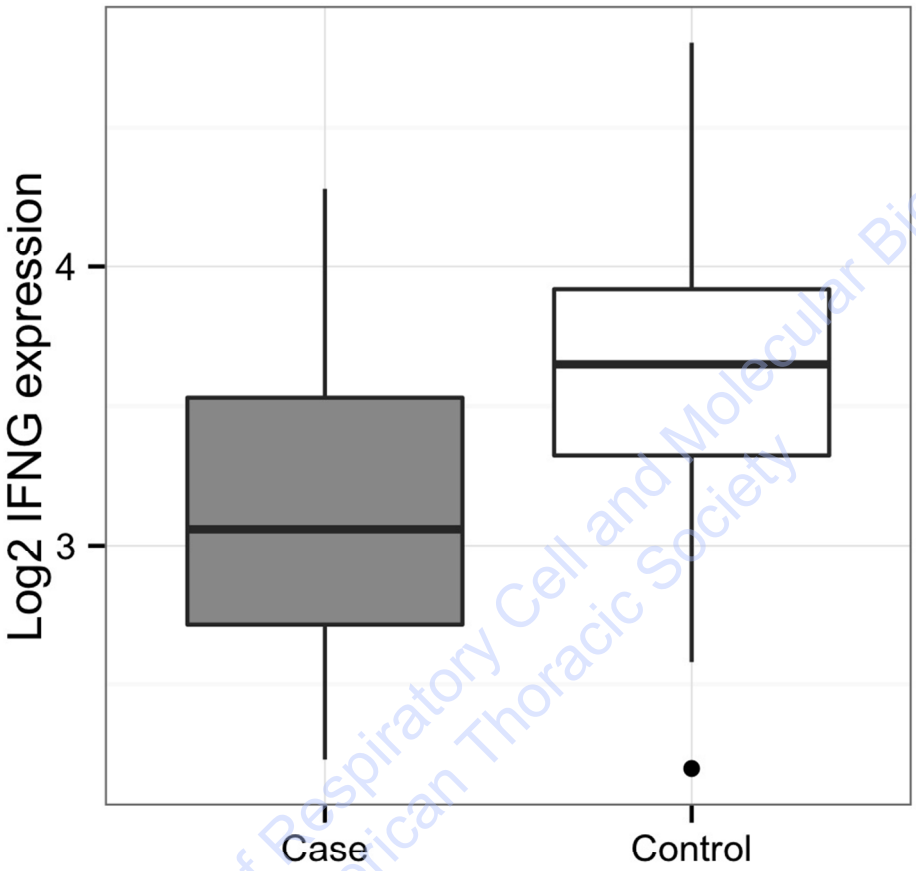

Figure 1.

374x374mm (72 x 72 DPI)

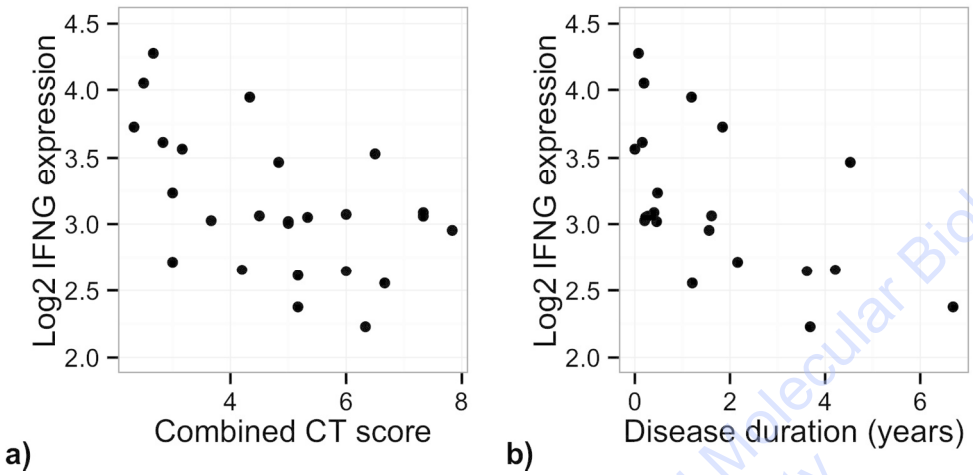

Figure 2.

564x282mm (72 x 72 DPI)

Whole Blood Gene Expression in Pulmonary Non-tuberculous Mycobacterial Infection

Steven A. Cowman, Joseph Jacob, David M. Hansell, Robert Wilson, William O. C. Cookson, Miriam F. Moffatt and Michael R. Loebinger

ONLINE DATA SUPPLEMENT

Supplementary methods

Inclusion and exclusion criteria:

The criteria for inclusion or exclusion from the study are detailed in table E1.

	Inclusion Criteria	Exclusion Criteria
NTM Cases	Clinical diagnosis of pulmonary NTM infection Age > 18 years	Pregnancy Immunosuppression (other than oral corticosteroids) Malignancy Cystic fibrosis Active tuberculosis Underlying lung disease other than bronchiectasis or COPD
Controls	Bronchiectasis or COPD Age > 18 years	Pregnancy Immunosuppression (other than oral corticosteroids) Malignancy Cystic fibrosis Fungal disease Active or past history of any mycobacterial disease Radiology suggestive of mycobacterial disease

Table E1. Study Inclusion and exclusion criteria

Sampling and assessment of studied subjects:

The following procedures were performed for all study participants: structured clinical interview, administration of the St George's Respiratory Questionnaire (SGRQ), measurement of height, weight and transcutaneous oxygen saturation, pulmonary function testing (spirometry, lung volumes by plethysmography and carbon monoxide transfer factor), peripheral venous blood sampling (for RNA extraction, full blood count and cell differential, urea

and creatine, electrolytes, liver function testing, bone profile, total serum vitamin D level and C-reactive protein (CRP) level), collection of spontaneously expectorated sputum (for phenol auramine microscopy, mycobacterial, bacterial and fungal culture), and high resolution CT thorax (unless done within previous 6 months).

CT scoring:

CT scoring was performed by a specialist radiologist with 5 years of experience in thoracic imaging, blinded to clinical details. For bronchiectasis extent, bronchiectasis severity, tree-in-bud opacification, nodules and consolidation, tertiles of the maximum possible score for each feature were used to categorise the features as mild, moderate or severe. For cavitating nodules, severe cavitation and aspergilloma scores for each feature were dichotomised as present or absent. As a measure of the overall severity of their radiological changes, composite CT scores were calculated for each participant by adding the individual scores for bronchiectasis extent, bronchiectasis severity, tree-in-bud opacification, nodules, consolidation, cavitating nodules, severe cavitation and aspergilloma.

Gene expression analysis:

RNA was isolated from peripheral blood using PAXgene RNA tubes (Becton, Dickinson and Company, NJ, USA) and extracted using the PAXgene Blood RNA Kit (PreAnalytiX, Hombrechtikon, Switzerland). Extracted RNA was

amplified to complementary DNA using the Ovation Pico WTA System V2 (NuGEN Technologies, San Carlos, USA), then cDNA was fragmented and labelled using the Encore Biotin Module (NuGEN Technologies) and immediately hybridised to Affymetrix Human Gene 1.1 ST array plates and scanned using the GeneTitan instrument (Affymetrix, Santa Clara, USA). Samples were randomised prior to amplification and again prior to fragmentation, labelling and hybridisation. RNA and cDNA were quantified using the Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, USA) and RNA integrity was assessed using the 2100 Bioanalyzer and RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, USA).

Outliers were identified using the arrayQualityMetrics package version 3.18.0 (1) in the R environment version 3.1.2. Any sample flagged as an outlier using any of the three criteria (distance between arrays, signal intensity boxplots or M-A plots) was removed. Raw probe level data from .CEL files passing quality control were summarised, quantile normalised and $\log_2$ transformed using the Affymetrix Power Tools package. Expression data were re-imported into R version 3.1.2 where all downstream analyses were performed. Low-expressed probes (those with expression level below the global median) and unannotated probes were removed.

Differentially expressed genes were identified using Significance Analysis of Microarrays (SAM) applying a false discovery rate (FDR) of 5% for significance. The Benjamini-Hochberg correction for multiple testing was used

to correct the FDR. For further details of SAM methodology please see the original description by Tusher *et al.* (2).

To investigate the possibility that differences in the use of systemic steroids or NTM treatment may be responsible for the difference in expression between NTM cases and controls, multiple linear modelling was performed in R version 3.1.2. The expression of each of the genes differentially expressed between NTM cases and controls was the dependent variable and NTM disease, NTM treatment and systemic steroid use plus interaction terms as the explanatory variables. A *P* value of 0.05 was deemed significant.

Gene ontology analyses were performed using the WebGestalt tool (http://www.webgestalt.org updated on 30/01/2013) using the Gene Ontology Consortium (GO) database version 1.2 (updated 11/11/2012) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (updated 21/03/2011). These databases were used to identify groups of functionally related genes which are present within a larger set of genes (for example a set of differentially expressed genes, or a gene module identified by WGCNA) more than would be expected by chance, referred to as 'enrichment'. All expressed genes were used as the reference set, the hypergeometric test was used to determine the significance of gene enrichment using the Benjamini-Hochberg correction for multiple testing. An adjusted *P* value of 0.05 was deemed significant. For the analysis of enrichment with the 31-gene sarcoid signature the hypergeometric test was performed using R version 3.1.2.

Weighted Gene Co-expression Gene Network Analysis (WCGNA) was performed using the WGCNA package (3) according to the authors' recommendations. All expressed genes were used as the initial input and genes with high (top 33%) connectivity were then selected to form the topological overlap matrix, using a soft thresholding power of 12. Hierarchical clustering was then used on this matrix to identify gene modules using a minimum module size of 30, and modules with greater than 75% similarity were merged. Module eigengenes were calculated as the first principal component of gene expression in each module and then used to calculate the correlation between module expression and phenotypic traits.

The expression data are available via the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97298).

Supplementary results

Recruitment:

Subjects were recruited using the criteria detailed in table E1. The control group was composed solely of Caucasian subjects. To prevent differences in ethnicity between groups from confounding the analysis of gene expression, two pNTM subjects of Chinese and Asian ethnicity were excluded prior to analysis as suitable controls matched for ethnicity could not be found.

Clinical data:

The results of peripheral blood testing for study participants are detailed in table E2.

	Cases	Controls
Haemoglobin	13.46 (±1.56)	12.65 (±3.83)
Platelets	264 (±112)	230 (±94)
Total leucocyte count	8.16 (±2.42)	7.39 (±1.86)
Neutrophils	5.57 (±2.14)	4.63 (±1.54)
Lymphocytes	1.67 (±0.64)	1.95 (±0.62)
Eosinophils	0.15 (±0.13)	0.17 (±0.1)
Monocytes	0.73 (±0.3)	0.61 (±0.19)
CRP	19.24 (±39)	6.74 (±8.61)
Albumin	37.92 (±5.3)	40.26 (±3.91)

Table E2. Blood results of study participants. Values are shown as mean ± standard deviation.

CT scoring:

The mean composite CT score was 4.34 (standard deviation ±1.41) ranging from 2.33 to 6.5. Details of individual CT features are shown in table E3.

CT feature	Mean score (±SD)	Score	Number of subjects (%)
Bronchiectasis extent	1.72 (±0.71)	Low	7 (28%)
		Intermediate	8 (32%)
		High	10 (40%)
Bronchiectasis severity	0.96 (±0.35)	Mild	10 (40%)
		Moderate	15 (60%)
		Severe	0
Tree-in-bud opacification	0.68 (±0.55)	Mild	19 (76%)
		Moderate	6 (24%)
		Severe	0
Nodules	0.74 (±0.55)	Mild	20 (80%)
		Moderate	5 (20%)
		Severe	0
Cavitating nodules	0.11 (±0.16)	Absent	15 (60%)
		Present	10 (40%)
Severe cavitation	0.13 (±0.23)	Absent	17 (68%)
		Present	8 (32%)
Aspergilloma	0.03 (±0.08)	Absent	22 (88%)
		Present	3 (12%)
Consolidation	0.46 (±0.46)	Mild	22 (88%)
		Moderate	3 (12%)
		Severe	0

Table E3. Scores for individual CT features in pNTM cases. SD = standard deviation.

Gene expression and NTM disease

RNA extraction:

The mean RNA yield per sample was 198 ng/μl (range 71.8 to 353 ng/μl). The mean RIN was 8.5 (range 7.1 to 9.3).

Quality control:

Five samples were identified as outliers using Array Quality Metrics (three cases, two controls). These samples were excluded, leaving 52 working samples for downstream analysis.

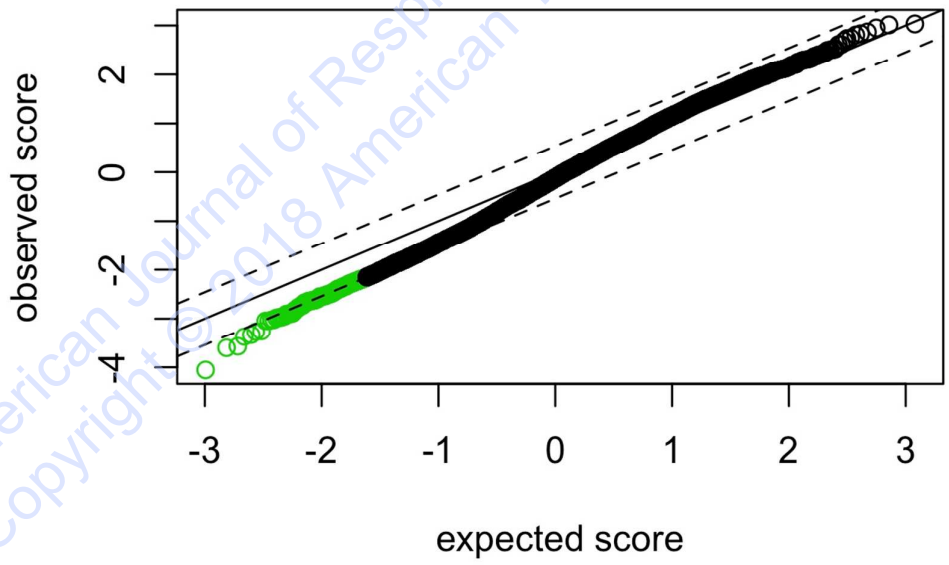

Figure E1. Quantile-quantile plot for differences in gene expression between pNTM and controls calculated using SAM. Dotted lines indicate the value of δ

corresponding to an FDR of 5%, transcripts showing significant differential expression are shown in green.

Table E4. Genes differentially expressed between pNTM cases and controls.

See attached document for table.

Table E5. GO analysis of genes differentially expressed between pNTM cases and controls

See attached document for table.

Table E6. Linear modelling of gene expression and NTM disease, NTM treatment and systemic steroid use.

See attached document for table.

Table E7. WGCNA gene modules for pNTM cases and controls

See attached document for table.

Table E8. GO analysis of WGCNA modules associated with pNTM

See attached document for table.

***IFNG* expression and clinical outcomes:**

The relationship between *IFNG* expression and individual CT features is shown in Figure E2.

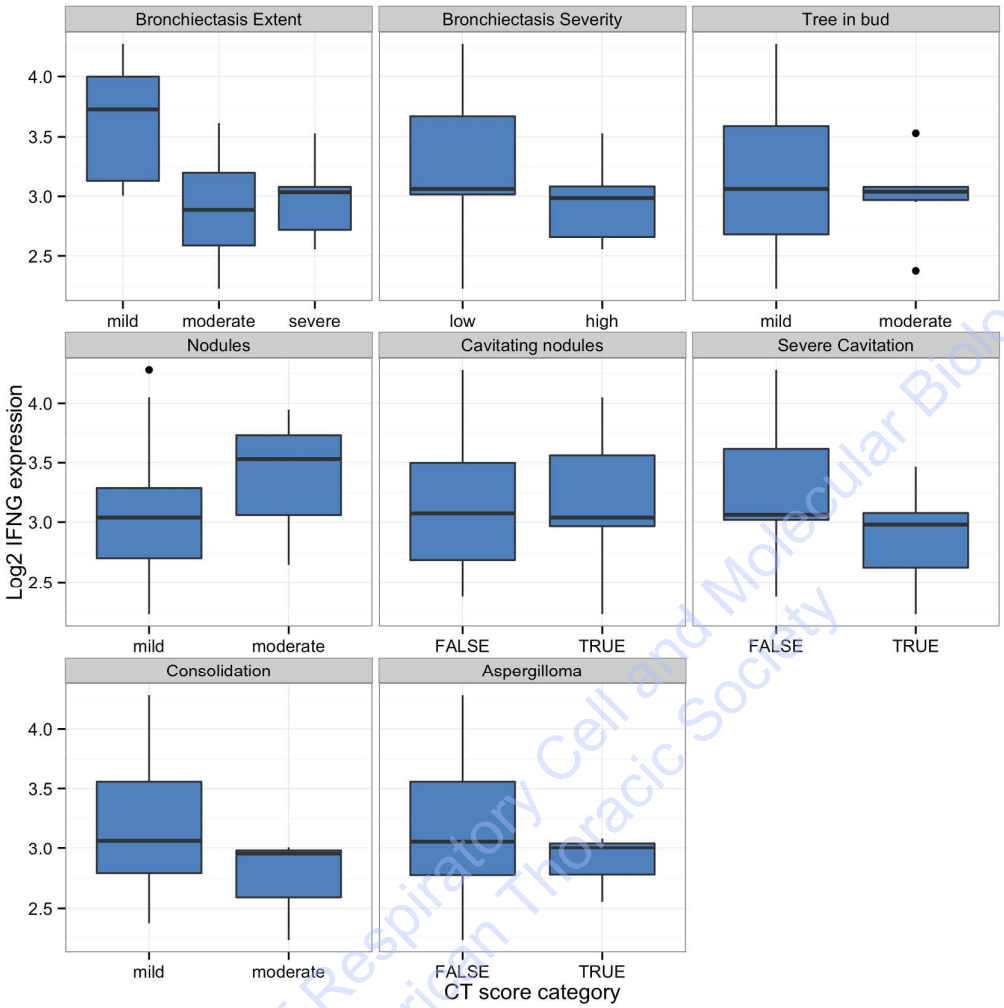

Figure E2. *IFNG* expression and individual CT features.

Clinical measure	Correlation coefficient	P - value
FEV1 (% predicted)	0.345	0.078
FVC (% predicted)	0.342	0.080
TLC (% predicted)	-0.009	0.965
TLCO (% predicted)	0.335	0.088
KCO (% predicted)	0.203	0.310
Serum albumin (g/L)	-0.091	0.652
Serum CRP (mg/L)	-0.015	0.939
Neutrophil count (cells x 10^9/L)	0.037	0.855

Table E9. Correlation between *IFNG* expression and clinical features in control subjects.

Gene expression and survival in NTM disease:

Table E10. Genes associated with survival in pNTM

See attached document for table.

Table E11. GO analysis of genes associated with decreased survival in pNTM

See attached document for table.

Table E12. GO analysis of genes associated with increased survival in pNTM

See attached document for table.

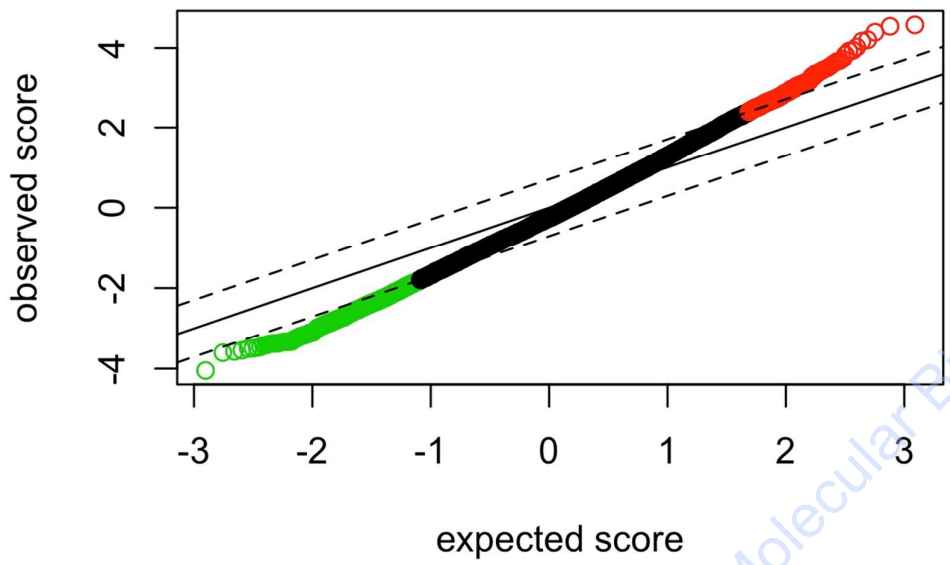

Figure E3. Quantile-quantile plot for gene expression and survival in pNTM cases calculated using SAM. Dotted lines indicate the value of δ corresponding to an FDR of 5%. Transcripts significantly associated with increased and decreased mortality are shown in red and green, respectively.

Table E13. WGCNA gene modules for pNTM cases

See attached document for table.

Table E14. GO analysis of WGCNA modules associated with survival in pNTM

See attached document for table.

Supplementary references

1. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. *Bioinformatics* 2009;25:415–416.
2. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001;98:5116–5121.
3. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.