Association between bacterial homoplastic variants and radiological pathology in tuberculosis

Louis Grandjean,1,2,3 Joha Monteserin,4 Robert Gilman,2,5 Julia Pauschardt,2 Sakib Rokadiya,6 Cesar Bonilla,7 Viviana Ritacco,6,4 Julia Rios Vidal,7 Julian Parkhill,8 Sharon Peacock,9 David AJ Moore,10 Francois Balloux11

ABSTRACT

Background Understanding how pathogen genetic factors contribute to pathology in TB could enable tailored treatments to the most pathogenic and infectious strains. New strategies are needed to control drug-resistant TB, which requires longer and costlier treatment. We hypothesised that the severity of radiological pathology on the chest radiograph in TB disease was associated with variants arising independently, multiple times (homoplasies) in the Mycobacterium tuberculosis genome.

Methods We performed whole genome sequencing (Illumina HiSeq2000 platform) on M. tuberculosis isolates from 103 patients with drug-resistant TB in Lima between 2010 and 2013. Variables including age, sex, HIV status, previous TB disease and the percentage of lung involvement on the pretreatment chest radiograph were collected from health posts of the national TB programme. Genomic variants were identified using standard pipelines.

Results Two mutations were significantly associated with more widespread radiological pathology in a multivariable regression model controlling for confounding variables (Rv2828c.141, RR 1.3, 95% CI 1.21 to 1.39, p<0.01; rpoC.1040 95% CI 1.77 to 2.16, RR 1.9, p<0.01). The rpoB.450 mutation was associated with less extensive radiological pathology (RR 0.81, 95% CI 0.69 to 0.94, p=0.03), suggestive of a bacterial fitness cost for this mutation in vivo. Patients with a previous episode of TB disease and those between 10 and 30 years of age also had significantly increased radiological pathology.

Conclusions This study is the first to compare the M. tuberculosis genome to radiological pathology on the chest radiograph. We identified two variants significantly positively associated with more widespread radiological pathology and one with reduced pathology. Prospective studies are warranted to determine whether mutations associated with increased pathology also predict the spread of drug-resistant TB.

INTRODUCTION

Pathology in TB disease depends on host, pathogen and environmental factors. Understanding to what extent pathogen genetic factors contribute to the variability in pulmonary pathology due to Mycobacterium tuberculosis could allow those patients that harbour the most pathogenic and thereby infectious organisms to be identified for enhanced treatments, transmission prevention measures and contact tracing. Multidrug-resistant tuberculosis (MDRTB) treatment outcomes are worse, medications are more toxic and the cost of treatment is 10 times greater than drug-susceptible TB.1 This makes prevention of new cases of MDRTB particularly important. Although there is now a greater understanding of which polymorphisms are associated with drug resistance in TB, relatively little is known about the consequences of the genetic makeup of the pathogen on disease severity and progression in vivo. So far, no studies have examined the association of the M. tuberculosis pathogen genome with radiologically quantified pathology on the chest radiograph.

Many studies have implicated the influence of M. tuberculosis lineage with TB pathogenesis in laboratory animals; however, these studies are limited by comparison to laboratory strains and typically lack information about whole genome variability.2,3 It is well established that there is a fitness cost arising from some rifampicin resistance conferring mutations in some drug-resistant strains in vitro,4,5 drug-resistant strains have been shown to cause less pathology in guinea pigs6 than controls. However, Comas et al demonstrated that this fitness cost in M. tuberculosis when measured in vitro by competitive fitness assays can be mitigated by compensatory mutations in the rpoA and rpoC genes.6 The

Key messages

What is the key question?

► Is there an association between bacterial genotype and pathology in drug-resistant TB?

What is the bottom line?

► This study demonstrates that polymorphisms in Rv2828C and rpoC are independently associated with more widespread pathology on the pretreatment chest radiograph of patients with drug-resistant TB.

Why read on?

► As the first study to compare the Mycobacterium tuberculosis genome to radiological pathology on the chest radiograph, it improves our understanding of how pathogen genetic factors contribute to pathology in TB, which could enable tailored treatments to the most pathogenic and infectious strains.
heterogeneity in the fitness cost of drug resistance mutations is also key to the future predictions of MDRTB spread.10 11

Convergent evolutionary (homoplastic) mutations emerging recurrently and independently at the same locus in multiple clades of the phylogeny form a subset of mutations likely to have been affected by strong directional natural selection. These homoplastic mutations represent excellent candidate mutations for polymorphisms involved in antimicrobial resistance, immune evasion and virulence12 13 and may influence radiological pathology and outcome.

Work by Ralph and colleagues14 has demonstrated that assessment of the degree of pulmonary parenchymal involvement on the chest radiograph can be reliably reproduced by independent observers with a high agreement between observers and that this is predictive of the sputum smear result at 2 months. We hypothesised that some convergent evolutionary polymorphisms could be significantly associated with lung pathology as quantified by the extent of pulmonary parenchymal involvement on the pretreatment diagnostic chest radiograph. In order to test this hypothesis, we performed whole genome sequencing on 103 TB patients with drug-resistant disease in Lima who had a chest radiograph performed before the start of treatment.

MATERIALS AND METHODS

Field methods, culture techniques and sample selection

Collection of clinical and demographic variables, sampling of sputa, culture and DNA extraction were undertaken as previously described. Briefly, samples with available pretreatment chest radiographs were selected from a prospective cohort study16 undertaken in the regions of Callao and Lima South between 2010 and 2013. Demographic data such as age and sex were collected from the clinical record at the health post. Previous TB disease was defined as any previous episode of clinically diagnosed and treated TB disease prior to the present episode. A pretreatment chest radiograph was defined as a radiograph taken before the onset of treatment for the current episode of TB disease at the same time as the initial diagnostic sampling. Testing for HIV was undertaken for all patients, and this result was confirmed from the clinical record. A bad treatment outcome was defined as those that died during treatment or had a bacteriological failure of treatment, while a good treatment outcome was defined as treatment completion or success as per the WHO guidelines.17 Those that transferred treatment outside the study area or those that abandoned treatment were regarded as unknown outcomes.

Sputum samples from all patients were transported to the regional reference laboratories and processed both on liquid (MODS) and solid Ogawa media. An aliquot of each positive culture was subcultured at Universidad Peruana Cayetano Heredia and DNA was extracted.15 All samples were retested and had drug susceptibility profiles confirmed at the national reference laboratory using the proportions method on agar.

Reading of chest radiographs and quantification of pathology

Chest radiographs included in the analysis were all correctly labelled, posteroanterior, baseline pretreatment films. Poor quality radiographs were excluded if they were over or under penetrated, over rotated or with inadequate field of view. Chest radiographs were read independently and prior to data analysis as described14 18 by two experienced clinicians blinded to the sequencing data and specialised in managing TB. The total percentage of the lung opacified with pathology was used as the pathology score outcome as previous studies have demonstrated that this is the only reliably reproducible indicator between observers. Briefly, the lungs were divided into six equally sized zones and the percentage of opacity in each zone and overall was estimated to the nearest 5%.18 Readings from the two observers were compared to quantify the degree of correlation between observers and an average of the two readings was used for the analysis. In addition to Spearman’s rank correlation, a Bland-Altman plot was used to compare the agreement between reviewers.

Genome sequence quality control

We prepared Illumina sequencing libraries with a 450bp insert size, using the manufacturer’s protocols, and then undertook sequencing on an Illumina HiSeq2000 with paired-end reads of length of 100bp. We multiplexed 96 samples per lane to attain an average depth of coverage of ~97-fold. We confirmed the species in the short reads using Kraken19 and then assembled the paired-end sequence reads with an improved assembly pipeline,20 based on Velvet.21 A list of isolates and their accession numbers in the European Nucleotide Archive is provided in online

Table 1  Clinical, demographic and pathogen data for the 103 isolates included in the study

| Variable                              | No. (±%*) |
|---------------------------------------|-----------|
| No. of patients                       | 103       |
| Male gender (%)                       | 71 (69)   |
| Age                                   |           |
| 10–19                                 | 21 (20)   |
| 20–29                                 | 46 (45)   |
| 30–39                                 | 13 (13)   |
| 40–49                                 | 11 (11)   |
| 50–59                                 | 7 (7)     |
| 60–69                                 | 2 (2)     |
| 70–79                                 | 2 (2)     |
| Missing                               | 1 (1)     |
| HIV positive                          |           |
| Radiological pathology % of lung involved (median/IQR) | 30 (15–50) |
| Drug susceptibility status            |           |
| Mono-resistant                        | 12 (12)   |
| MDR                                   | 91 (88)   |
| Previous tuberculosis disease         | 30 (29)   |
| Lineage                               |           |
| Lineage 2.2.1                         | 7 (7)     |
| Lineage 4                              | 7 (7)     |
| Lineage 4.1                           | 1 (1)     |
| Lineage 4.1.1                         | 5 (5)     |
| Lineage 4.1.1.3                       | 4 (4)     |
| Lineage 4.1.2.1                       | 10 (10)   |
| Lineage 4.3.2                         | 4 (4)     |
| Lineage 4.3.3                         | 43 (42)   |
| Lineage 4.3.4.1                       | 4 (4)     |
| Lineage 4.3.4.2                       | 17 (16)   |
| Lineage 4.8                            | 1 (1)     |

*% total 101 due to rounding.
†No missing data unless stated.
MDR, multidrug resistant.
supplementary file 1 (project number: ERP004677). Following this, short reads were mapped to the corrected H37Rv reference genome available from Casali et al.22 genome.cshlp.org/content/suppl/2012/02/01/gr.128678.1111.DC1/1_H37RvQM_embl.txt. In doing so, we employed SMALT V 0.7.4 (www.sanger.ac.uk/science/tools/smalt-0) using maximum and minimum insert sizes of 1000 and 50, respectively. To annotate single nucleotide polymorphisms (SNPs), we used SAMtools mpileup23 and BCFtools, as described by Harris et al.24 We included SNPs that were covered by at least two forward and two reverse short paired-end reads.25 A minimum base call quality of 50 and a minimum root mean squared mapping quality of 30 to call a SNP were used. Furthermore, the SNPs at sites with heterogeneous mapping where less than 75% of reads at that site covered the SNP were excluded from the analysis.24 We obtained the multiple alignment by generating pseudosequences, after ignoring small indels. Any site with >5% non-calls was excluded from the analysis, and an arbitrary minimum SNP prevalence threshold of 3% (3/103) and a homoplasy count of a minimum of three independent occurrences across the phylogenetic tree were chosen to avoid spurious comparison of low prevalence mutant strains.

Identification of convergent evolutionary variants
Variants were identified using phyC software12 to determine the number of independent occasions in which an ancestral base was different from the descendent base at any given site in the tree.

Sample size and model construction
As no previous study had been undertaken to provide an estimate of the expected effect size, formal power calculation was not performed. The study aim to recruit more than 100 patients was therefore based on pragmatic and systematic considerations. The extent of radiological pathology was Poisson distributed; therefore, a Poisson model was constructed using this variable as the dependent variable. Univariate analysis was first undertaken for each polymorphism. Polymorphisms with a p value of <0.05 on univariate analysis were selected for inclusion in a multivariable model together with confounding variables. The set of confounding variables that were included in the multivariable model was defined a priori. These included sex, age, previous TB disease, drug resistance status (mono-resistant or multidrug-resistant) and lineage. Multivariable Poisson regression was used to calculate the p values having corrected for confounding variables. The R package ‘stats’ was then used to plot the predicted pathology having accounted for other confounding variables in the multivariable model. All p values were Bonferroni corrected. Missing data were omitted and managed using the default `na.omit` setting under the glm function in R. Following reviewers’ comments, we performed additional population level correction

Phylogenetic analysis
A maximum likelihood tree was constructed with concatenated SNPs from the whole genome sequence data using R software (R Foundation for Statistical Computing, Vienna, Austria 2011, www.R-project.org) with the PHANGORN package.28 Clades were named according to the latest SNP-based bar code.27

Figure 1  Study phylogeny with branches coloured by sublineage as defined by Coll et al. Clinical outcome following treatment is displayed (bad outcome: black dots, good outcome: grey dots on the tips of the tree) in a ring surrounding the phylogeny with the pathology on chest radiograph (% of lung with disease) represented by a circle of varying size according to the extent of pathology.
using principal components to account for the effect of lineage on the association between the homoplastic polymorphisms and radiological pathology. We also performed a subanalysis of our principal findings with and without the inclusion of patients with previous TB disease as a variable.

**RESULTS**

A total of 103 different patients had radiological, clinical, demographic and pathogen data available for analysis (table 1). The proportion of patients with MDRTB was 88% (91/103) and 5% (5/103) were HIV positive. The predominant sublineage was lineage 4.3.3 (41%, 43/103) of the Euro-American lineage (Lineage 4). As a likely consequence of Chinese immigration to Peru in the mid-19th century to work on the railways and guano farms, 7% (7/103) of strains were from the Beijing family (table 1, figure 1, online supplementary figure 2). Evaluation of the extent of radiological pathology by two independent observers was significantly positively correlated using the Spearman rank correlation coefficient between observer pathology scores (Spearman $\rho = 0.66$, 95%CI 0.53 to 0.76, $p<0.001$) (online supplementary figure 1). The mean difference in pathology scores between observers in the Bland-Altman analysis was

| Reference position | Gene | Mutation | Homoplasies (n) | No. (Prevalence) | Reference base | Mutant base |
|--------------------|------|----------|-----------------|-----------------|----------------|-------------|
| 761160             | rpoB | Ser450Leu| 13              | 54 (52%)        | C              | T           |
| 2155176            | katG | Ser315Thr| 10              | 71 (69%)        | C              | G/T         |
| 4247436            | embB | Met306Val| 9               | 13 (13%)        | A              | G           |
| 1473254            | rs   | A1401G   | 5               | 5 (5%)          | A              | G           |
| 3135920            | Rv2828c | Thr141Arg   | 4               | 34 (33%)        | G              | C           |
| 4247438            | embB | Met306Ile| 4               | 16 (16%)        | G              | A/C/T       |
| 766493             | rpoC | Pro1040Arg| 3               | 3 (3%)          | C              | G/T         |
| 761114             | rpoB | Asp435Tyr| 3               | 3 (3%)          | G              | T           |
| 761144             | rpoB | His445Asp| 3               | 9 (9%)          | C              | G/T         |
| 2123153            | idD2 | Val311le| 3               | 13 (12%)        | C              | T           |
Tuberculosis

Figure 3  Distribution of radiograph pathology as a function of epidemiological and pathogen genetic variables; (A) previous episode of TB, rate ratio 1.2, p<0.01; (B) age range 10–30 versus other ages, rate ratio 1.2, p<0.01; (C) gender, rate ratio 1.13, p=0.29; (D) rpoB 450; Rr 0.8, p=0.03; (E) rpoC.1040, Rr 1.97, p<0.01; (F) Rv2828c.141, Rr 1.3, p<0.01.

4.7%, while the upper limit of agreement was 43.1% (95% CI 36.7 to 43.1) and the lower limit −32.9% (95% CI −26.5 to −39.4). Pathology appeared evenly distributed by sublineage before considering confounding variables and there were eight bad outcomes (failed treatment or died during treatment) (8/103) according to the WHO definitions (figure 1). The distribution of homoplastic variants across the phylogenetic tree is shown in figure 2; the most frequently observed homoplastic mutation was the rpoB Ser450Leu mutation, which occurred independently 13 times with a prevalence of 54% in the study population (figure 2, table 2).

Clinical and demographic variables associated with chest radiograph pathology
A previous episode of TB disease was associated with more extensive pathology on the chest radiograph (RR 1.2, p<0.01, 95% CI 1.10 to 1.28; figure 3A). Those aged between 30 and 39 had less severe disease than those between 10 and 19 years of age.
Table 3  Variables in the multivariable predictive regression model of radiological pathology in multidrug-resistant tuberculosis with rate ratios and their associated lower and upper 95% CI, SE and p values

| Variable | Rate ratio | Upper CI | Lower CI | SE  | P value* |
|----------|------------|----------|----------|-----|----------|
| Mutation |            |          |          |     |          |
| katG     | 1.14       | 1.28     | 1.01     | 0.07| 1.00     |
| rpoB.450 | 0.81       | 0.94     | 0.69     | 0.06| 0.03     |
| rpoC.1040| 1.97       | 2.16     | 1.77     | 0.10| <0.01   |
| rpoB.425 | 1.21       | 1.40     | 1.01     | 0.10| 1.00     |
| Rv2828c.141| 1.30    | 1.39     | 1.21     | 0.04| <0.01   |
| embB.306 | 0.86       | 0.97     | 0.74     | 0.06| 0.27     |
| rpoB.445 | 1.09       | 1.25     | 0.93     | 0.08| 1.00     |
| Male gender | 1.13   | 1.23     | 1.04     | 0.05| 0.29     |
| HIV positive | 0.84   | 1.02     | 0.66     | 0.09| 1.00     |
| Age 10–19 Comparison |  | | | | |
| Age 20–29 | 0.97     | 1.08     | 0.87     | 0.05| 1.00     |
| Age 30–39 | 0.66     | 0.82     | 0.50     | 0.08| <0.01   |
| Age 40–49 | 0.93     | 1.08     | 0.78     | 0.08| 1.00     |
| Age 50–59 | 1.20     | 1.37     | 1.02     | 0.09| 1.00     |
| Age 60–69 | 0.71     | 1.04     | 0.38     | 0.17| 1.00     |
| Age 70–79 | 0.43     | 0.79     | 0.07     | 0.18| <0.01   |
| MDR Comparison |  | | | | |
| Mono-resistant | 0.86   | 0.99     | 0.72     | 0.07| 0.75     |
| Previous TB | 1.19     | 1.28     | 1.10     | 0.05| <0.01   |
| Lineage 2 Comparison |  | | | | |
| Lineage 4    | 1.28     | 1.49     | 1.08     | 0.10| 0.48     |
| Lineage 4.1.1| 1.32   | 1.54     | 1.10     | 0.11| 0.33     |
| Lineage 4.1.3 | 1.06   | 1.34     | 0.78     | 0.14| 1.00     |
| Lineage 4.1.2.1| 1.34  | 1.52     | 1.16     | 0.09| 0.05     |
| Lineage 4.2.1  | 1.15    | 1.40     | 0.89     | 0.13| 1.00     |
| Lineage 4.3.3 | 1.02    | 1.18     | 0.87     | 0.08| 1.00     |
| Lineage 4.3.1.4 | 0.61   | 0.89     | 0.33     | 0.14| 0.02     |
| Lineage 4.3.2 | 1.26    | 1.43     | 1.08     | 0.09| 0.27     |
| Lineage 4.8 | 0.98 | 1.39     | 0.57     | 0.21| 1.00     |

*Bonferroni-corrected p value.

(RR 0.66, p<0.01, 95% CI 0.50 to 0.82). When age as a variable in the multivariable model was trichotomised into those less than 30, between 30 and 50 years old and those greater than 50 years of age, those less than 30 years of age were shown to have more severe disease than those between 30 and 50 years of age (RR 1.3, p<0.01, 95% CI 1.17 to 1.38, figure 3B). There was no significant difference in pathology observed by sex (figure 3C). HIV infection was not statistically associated with more severe disease.

Association between homoplastic polymorphisms and chest radiograph pathology

We detected 10 homoplastic polymorphisms in 103 strains. Other than the rrs A1401G mutation that occurred in an rRNA gene, all homoplastic polymorphisms detected in the dataset were non-synonymous (amino acid changing). Seven homoplastic polymorphisms were associated with the extent of lung pathology with a p value<0.05 on univariate analysis and these were included in the multivariable regression together with clinical and sociodemographic confounding variables. Three mutations were significantly associated with radiological pathology after controlling for confounding variables and Bonferroni correction (table 3).

The Pro1040Lys mutation in rpoC had the greatest effect size on radiological pathology with a rate ratio of 1.9 (95% CI 1.77 to 2.16, p<0.01; figure 3E). Patients that harboured a strain with this mutation had a mean pathology score almost twice that of those without the mutation. This was independent of other confounding variables, although, the prevalence of the rpoC.1040 mutation was only at a frequency of 3%, limiting its influence at the population level. In contrast, the Rv2828c Thr141Arg mutation occurred in 34% of the population at a rate ratio of 1.3 (95% CI 1.21 to 1.39, figure 3F), meaning that those with disease caused by a strain with the mutation had 30% more likely to have more widespread pathology than those without. The most prevalent homoplastic mutation rpoB S450L (known to be associated with rifampicin resistance) was associated with decreased pathology (rate ratio 0.8, p=0.03, 95%CI 0.69 to 0.94; figure 3D). However, the protective association of this rpoB.450 polymorphism on radiological pathology...
was not statistically significant when additional correction for population structure with principal component analysis (PCA) was performed, demonstrating that the association may be confounded by lineage. Neither did this polymorphism remain statistically significantly protective of pathology when a subanalysis was performed on only TB naïve patients, suggesting that this association was influenced by prior disease status.

DISCUSSION

This is the first published study to compare pathogen genome-wide data with radiological pathology in *M. tuberculosis*. We identified that MDRTB patients with disease caused by a pathogen harbouring the Rv2828c.141 mutation had 30% increased risk of more extensive radiological pathology. Those MDRTB patients with disease caused by strains carrying the rpoC.1040 polymorphism were twice as likely to have more extensive radiological disease. Conversely, the rpoB.450 mutation was 20% protective of radiological pathology, although this association was not observed to be independent of lineage nor to prior TB after additional PCA correction and subanalysis.

The finding that strains with the rifampicin resistance rpoB.450L mutation were associated with less widespread pathology and those with the rpoC.1040 polymorphism with more widespread pathology supports the notion that rifampicin resistance mediated by the rpoB.450 mutation is associated with a fitness cost as manifest by the delay to diagnosis is an unreliable variable to capture with often vague symptom start dates, this study is limited by the lack of data for this important confounding variable. Similarly, the socioeconomic status of the patient is likely to influence both the delay to presentation and adherence to prior treatment, making this a variable also likely to be associated with pathology on the chest radiograph. For those patients with previous TB disease, we did not capture data on the phenotype of the previous disease. However, a subanalysis restricted solely to TB naïve patients did not alter the principal findings.

The Bland-Altman comparison between observers demonstrated a 4.7% mean difference in pathology scores and the Spearman rank correlation between observers was significant. However, there was an increased difference between observers at the higher pathology scores. Future studies could improve the interobserver agreement by calibrating both observers on a training data set or employing AI pattern recognition algorithms instead from the outset.

In summary, this study provides a novel and interesting insight into the association of bacterial homoplasmic polymorphisms with the extent of pathology on the chest radiograph following infection with *M. tuberculosis*. With improved statistical power, it may be possible to identify a larger set of bacterial genetic markers that help predict pathology, disease outcome and transmission. Prospective cohort studies are also warranted to determine if polymorphisms in the bacterial genome of the index case such as these are independently predictive of secondary cases of disease and the development of severe radiological pathology among close contacts.

Author affiliations

1Department of Medicine, Imperial College London, London, UK
2Laboratorio de Investigacion y Enfermedades Infecciosas, Cayetano Heredia Pervinian University, Lima, Peru
3Institute of Child Health, UCL Division of Infection and Immunity, London, UK
4Instituto Nacional de Enfermedades Infecciosas INEI-ANLIS, Administración Nacional de Laboratorios e Institutos de Salud Dr Carlos G Malbrán, Buenos Aires, Argentina
5Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA
6Faculty of Medicine, Imperial College London, London, UK
7Unidad Tecnica de Tuberculosis MDR, Ministerio de Salud, Lima, Peru
8Pathogen Genomics Group, Wellcome Trust Sanger Institute, Cambridge, UK
9Faculty of Medicine, University of Cambridge, Cambridge, UK
10TB Centre, London School of Hygiene and Tropical Medicine, London, UK
11UC Genetics Institute, University College London, London, UK

Acknowledgements

Luz Caviedes, her drive and dedication to scientific research are sorely missed.

Contributors

Contributed to conception and design, acquisition, analysis and interpretation of data: LG, JM, RG, JP, SR, CB, VR, JRV, JP, SP, DAJM and FB. Contributed to drafting the work and critical revisions: LG, JM, RG, JP, SR, CB, VR, JRV, JP, SP, DAJM and FB. Contributed to final approval of the version submitted: LG, JM, RG, JP, SR, CB, VR, JRV, JP, SP, DAJM and FB. Contributed to the accountability, accuracy and integrity of the work: LG, JM, RG, JP, SR, CB, VR, JRV, JP, SP, DAJM and FB.

Funding

This work was supported by the Wellcome Trust (Grant Number 201470/Z/16/Z), JM acknowledges the EU FP7-PEOPLE-2013-IRSES-Marie Curie Action DEAN7 “International Research Staff Exchange Scheme” (19061), which covered the costs for a 3-month visit to London to work with LG. FB acknowledges support from the BBSC GCRF scheme and the National Institute for Health Research.

References

[1] Grandjean L, et al. Thorax 2020;75:584–591. doi:10.1136/thoraxjnl-2019-213281

This study was strengthened by the use of bacterial whole genome sequencing, the ability to make comparison to pretreatment chest radiographs and the availability of potentially confounding variables to adjust the findings in the context of a multivariable model. However, it is clear that the causes of radiological pathology at diagnosis are multifactorial. Although the delay to diagnosis is an unreliable variable to capture with often vague symptom start dates, this study is limited by the lack of data for this important confounding variable. Similarly, the socioeconomic status of the patient is likely to influence both the delay to presentation and adherence to prior treatment, making this a variable also likely to be associated with pathology on the chest radiograph. For those patients with previous TB disease, we did not capture data on the phenotype of the previous disease. However, a subanalysis restricted solely to TB naïve patients did not alter the principal findings.

The Bland-Altman comparison between observers demonstrated a 4.7% mean difference in pathology scores and the Spearman rank correlation between observers was significant. However, there was an increased difference between observers at the higher pathology scores. Future studies could improve the interobserver agreement by calibrating both observers on a training data set or employing AI pattern recognition algorithms instead from the outset.

In summary, this study provides a novel and interesting insight into the association of bacterial homoplasmic polymorphisms with the extent of pathology on the chest radiograph following infection with *M. tuberculosis*. With improved statistical power, it may be possible to identify a larger set of bacterial genetic markers that help predict pathology, disease outcome and transmission. Prospective cohort studies are also warranted to determine if polymorphisms in the bacterial genome of the index case such as these are independently predictive of secondary cases of disease and the development of severe radiological pathology among close contacts.

Author affiliations

1Department of Medicine, Imperial College London, London, UK
2Laboratorio de Investigacion y Enfermedades Infecciosas, Cayetano Heredia Pervinian University, Lima, Peru
3Institute of Child Health, UCL Division of Infection and Immunity, London, UK
4Instituto Nacional de Enfermedades Infecciosas INEI-ANLIS, Administración Nacional de Laboratorios e Institutos de Salud Dr Carlos G Malbrán, Buenos Aires, Argentina
5Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA
6Faculty of Medicine, Imperial College London, London, UK
7Unidad Tecnica de Tuberculosis MDR, Ministerio de Salud, Lima, Peru
8Pathogen Genomics Group, Wellcome Trust Sanger Institute, Cambridge, UK
9Faculty of Medicine, University of Cambridge, Cambridge, UK
10TB Centre, London School of Hygiene and Tropical Medicine, London, UK
11UC Genetics Institute, University College London, London, UK

Acknowledgements

Luz Caviedes, her drive and dedication to scientific research are sorely missed.

Contributors

Contributed to conception and design, acquisition, analysis and interpretation of data: LG, JM, RG, JP, SR, CB, VR, JRV, JP, SP, DAJM and FB. Contributed to drafting the work and critical revisions: LG, JM, RG, JP, SR, CB, VR, JRV, JP, SP, DAJM and FB. Contributed to final approval of the version submitted: LG, JM, RG, JP, SR, CB, VR, JRV, JP, SP, DAJM and FB. Contributed to the accountability, accuracy and integrity of the work: LG, JM, RG, JP, SR, CB, VR, JRV, JP, SP, DAJM and FB.

Funding

This work was supported by the Wellcome Trust (Grant Number 201470/Z/16/Z), JM acknowledges the EU FP7-PEOPLE-2013-IRSES-Marie Curie Action DEAN7 “International Research Staff Exchange Scheme” (19061), which covered the costs for a 3-month visit to London to work with LG. FB acknowledges support from the BBSC GCRF scheme and the National Institute for Health Research.
