Original research

Genome-wide association study of non-tuberculous mycobacterial pulmonary disease

Jaeyoung Cho,[1] Kyungtaek Park,[2] Sun Mi Choi,[1] Jinwoo Lee,[1,3] Chang-Hoon Lee,[1] Jung-Kyu Lee,[4] Eun Young Heo,[4] Deog Kyeom Kim,[3,4] Yeon Joo Lee,[5] Jong Sun Park,[3,5] Young-Jae Cho,[5] Ho Il Yoon,[3,5] Jae Ho Lee,[3,5] Choon-Taek Lee ,[3,5] Nayoung Kim,[3,6] Kyu Yeong Choi,[7] Kun Ho Lee,[7,8,9] Joohon Sung,[10,11] Sungho Won,[2,10,11] Jae-Joon Yim[1,3]

► Additional material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/thoraxjnl-2019-214430).

For numbered affiliations see end of article.

Correspondence to
Dr Jae-Joon Yim, Department of Internal Medicine, Seoul National University College of Medicine, Seoul 03080, Korea (the Republic of); yimjj@snu.ac.kr and Professor Sungho Won, Department of Public Health Sciences, Seoul National University Graduate School of Public Health, Seoul, Korea (the Republic of); won1@snu.ac.kr

JC and KP contributed equally.

Received 9 December 2019
Revised 25 August 2020
Accepted 23 September 2020
Published Online First 28 October 2020

To cite: Cho J, Park K, Choi SM, *et al*. *Thorax* 2021;**76**:169–177.

ABSTRACT

Background The prevalence of non-tuberculous mycobacterial pulmonary disease (NTM-PD) is increasing in South Korea and many parts of the world. However, the genetic factors underlying susceptibility to this disease remain elusive.

Methods To identify genetic variants in patients with NTM-PD, we performed a genome-wide association study with 403 Korean patients with NTM-PD and 306 healthy controls from the Healthy Twin Study, Korea cohort. Candidate variants from the discovery cohort were subsequently validated in an independent cohort. The Genotype-Tissue Expression (GTEx) database was used to identify expression quantitative trait loci (eQTL) and to conduct Mendelian randomisation (MR).

Results We identified a putatively significant locus on chromosome 7p13, rs849177 (OR, 2.34; 95% CI, 1.71 to 3.21; $p=1.36\times10^{-7}$), as the candidate genetic variant associated with NTM-PD susceptibility. Its association was subsequently replicated and the combined p value was 4.92×10^{-8}. The eQTL analysis showed that a risk allele at rs849177 was associated with lower expression levels of *STK17A*, a proapoptotic gene. In the MR analysis, a causal effect of *STK17A* on NTM-PD development was identified (β, −4.627; 95% CI, −8.768 to −0.486; p=0.029).

Conclusions The 7p13 genetic variant might be associated with susceptibility to NTM-PD in the Korean population by altering the expression level of *STK17A*.

Key messages

What is the key question?

► Is there genetic predisposition to non-tuberculous mycobacterial pulmonary disease (NTM-PD)?

What is the bottom line?

► Through a genome-wide association study (GWAS) we identified a putatively significant locus on chromosome 7p13, rs849177, which is possibly associated with susceptibility to NTM-PD in Korean populations by altering the expression levels of the proapoptotic *STK17A* gene.

Why read on?

► This first GWAS of NTM-PD with functional annotations helps offer insights into the pathogenesis of this disease.

INTRODUCTION

Non-tuberculous mycobacteria (NTM) are ubiquitous environmental organisms with a broad virulence spectrum. The incidence and prevalence of NTM pulmonary disease (NTM-PD) are increasing in many parts of the world,[1] and especially in South Korea where the annual prevalence of NTM infections has increased by more than fivefold over the last decade, reaching almost 40 cases/100 000 of the population in 2016.[2]

Disseminated NTM infections occur in patients with primary immunodeficiencies[3] as well as acquired immunocompromised states.[4] Previous studies have reported that Mendelian susceptibility to disseminated NTM infections is associated with gene mutations in the interleukin-12 dependent interferon-γ pathway.[3 5] However, NTM-PD develops in individuals without known immune defects, but typically in thin, postmenopausal women, sometimes with pectus excavatum, scoliosis or mitral valve prolapse.[6] As only a small subset of people get NTM-PD despite widespread exposure to NTM, the presence of the distinct phenotype,[6] familial clustering[7] and the ethnic disparity[8] of NTM-PD strongly suggest the presence of genetic predispositions to NTM-PD.

Unfortunately, genetic factors related to susceptibility to isolated NTM-PD remain elusive.[5] Although a whole-exome sequencing (WES) study of Caucasian patients with NTM-PD proposed that NTM-PD is a multigenic disease resulting from protein-affecting variants in immune, cystic fibrosis transmembrane conductance regulator, cilia and connective tissue genes in combination, it did not identify any specific underlying mutations.[9] Recently, two different WES studies identified macrophage stimulating 1 receptor (*MST1R*)[10] and TTK protein kinase (*TTK*)[11] as genetic risk factors for NTM-PD. However, only a small proportion of patients with NTM-PD could be explained by these variants, and these associations have not been replicated in other populations.[5]

Here, we carried out a genome-wide association study (GWAS) to elucidate genetic susceptibility to NTM-PD in the Korean population where NTM-PD is rapidly increasing.[2]

METHODS

Discovery cohort

Altogether, 412 patients with NTM-PD who participated in the ongoing prospective NTM cohort[12] from 1 July 2011 to 28 February 2017 at Seoul National University Hospital (SNUH) were included in this study. Diagnosis of NTM-PD was made based on the criteria of the American Thoracic Society/Infectious Diseases Society of America guidelines (online supplemental text).[13] Controls were selected from the Healthy Twin Study, Korea (HTK) cohort consisting of adult twin pairs and their relatives.[14] In the HTK cohort, 1056 participants from 364 families were genotyped using the same platform as that used for the discovery cohort cases. Among them, we randomly selected 494 participants who were one twin from each twin pair or their relatives.

Replication cohort

To validate genome-wide significant single-nucleotide polymorphisms (SNPs) in the discovery cohort, 187 patients with NTM-PD were enrolled at the SNUH, Seoul National University Bundang Hospital (SNUBH) and Seoul Metropolitan Government-Seoul National University Boramae Medical Center from 16 January 2017 to 20 November 2017. Controls were 500 healthy family members of patients with gastric cancer from SNUBH and 1195 healthy and cognitively normal participants from the Gwangju Alzheimer's disease and Related Dementia Cohort Research Center.[15]

All participants provided written informed consent and the study was conducted in accordance with the tenets of the Declaration of Helsinki.

GWAS with the discovery cohort

Both cases and controls in the discovery cohort were genotyped by the Illumina HumanCoreExome-12 V.1.0A on 298 930 variants. The first quality control (QC) stage was performed on cases and controls separately (online supplemental text and figure E1). After merging the cases and controls, the second QC stage was performed as suggested by Anderson *et al*.[16] The second QC stage involved two steps. First, participants were removed where (1) the call rates were less than 97%, (2) the differences in the heterozygosity rates from the means were larger than three times the SD and (3) the estimates of identity-by-descent with other participants were larger than 0.185, with the highest missing rates to remove participants with first-degree or second-degree relationships. Second, variants were removed when (1) the missing rates differed significantly between the cases and controls ($p<0.00001$), (2) the minor allele frequencies (MAFs) were smaller than 5%, (3) the call rates were less than 97% and (4) the p values for the Hardy-Weinberg equilibrium (HWE) exact test were smaller than 0.001.

After the second QC stage, we imputed the remaining 211 744 variants from the 709 participants using the Michigan Imputation Server (V.1.0.2) and its established pipeline. Haplotype Reference Consortium (HRC, r1.1 2016) was selected as the reference panel, Eagle (V.2.3) as the phasing programme, and Asian as the QC population.[17] The HRC panel consisted of 64 976 haplotypes and 32 611 samples from 20 different studies including the UK10K and 1000 Genomes Project.[18] Only variants with MAFs larger than 0.05, p values for the HWE exact test larger than 0.001, and R^2 values larger than 0.3 were analysed. Unknown body mass index (BMI) values for two patients with NTM-PD were imputed using the missForest method (R-package: missForest, V.1.4) for the following analyses.[19]

We conducted a logistic regression adjusting for age, age^2, sex and BMI. To adjust for population substructures, we calculated principal component (PC) scores with pruned variants. The 10 PC scores corresponding to the 10 largest eigenvalues were then included as covariates.[20] PLINK (V.1.90b3.44)[21] or R (V.3.4.4)

Table 1 Participant characteristics and the number of variants in the discovery and replication cohorts

	Discovery cohort		Replication cohort	
	Patients with NTM-PD (n=403)	Control (n=306)	Patients with NTM-PD (n=184)	Control (n=1680)
Age, mean (SD), years	63.5 (10.1)	54.5 (11.6)	65.8 (9.9)	67.3 (11.7)
Women, n (%)	265 (65.8)	175 (57.2)	123 (66.8)	1009 (60.1)
BMI, mean (SD), kg/m^2*	20.7 (2.6)	24.4 (3.1)	20.8 (2.6)	ND
Causative organism, n (%)†				
Mycobacterium avium complex	304 (75.8)	–	146 (79.8)	–
Mycobacterium abscessus complex	80 (20.0)	–	32 (17.5)	–
Others	17 (4.2)	–	5 (2.7)	–
Radiological types, n (%)				
Nodular bronchiectatic	333 (82.6)	–	145 (78.8)	–
Fibrocavitary	66 (16.4)	–	37 (20.1)	–
Unclassifiable	4 (1.0)	–	2 (1.1)	–
History of NTM treatment, n (%)‡	190 (47.1)	–	101 (54.9)	–
No. of variants after imputation	4 877 242		4 969 381	

*Including the imputed values for two patients with NTM-PD.
†The NTM species from two patients with NTM-PD in the discovery cohort and one in the replication cohort were not identified. Both *Mycobacterium avium* complex and *Mycobacterium abscessus* complex species were isolated in 64/401 patients (16.0%) in the discovery cohort and in 25/183 (13.7%) in the replication cohort. These patients were considered to have *M. avium* complex pulmonary disease when *M. avium* complex species were more often isolated than *M. abscessus* complex species, and vice versa. When these patients had the same number of *M. avium* complex and *M. abscessus* complex species isolated, we classified them based on the first isolated species.
‡Whether or not patients with NTM-PD had been treated was investigated on 18 May 2017 for the discovery cohort and on 29 November 2017 for the replication cohort.
BMI, body mass index; ND, not determined; NTM, non-tuberculous mycobacteria; NTM-PD, NTM pulmonary disease.

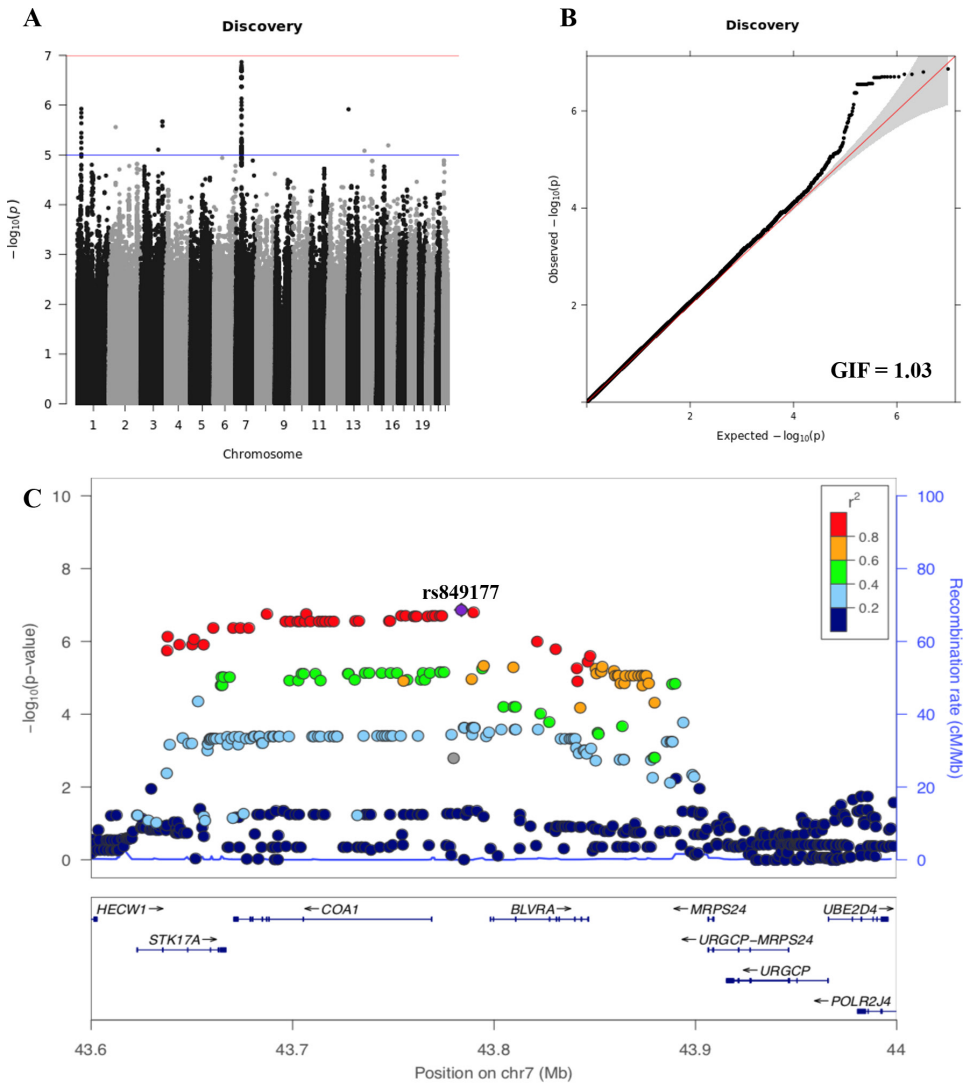

Figure 1 Plots showing putatively significant loci in the discovery cohort. (A) Manhattan plot. Red and blue lines indicate significance levels of 1.00×10^{-7} and 1.00×10^{-5}, respectively. (B) Quantile-quantile plot. CIs for the estimated p values are 0.95 and are coloured in grey. (C) Regional plot showing genes near to rs849177. Purple indicates rs849177 and other colours represent the degree of correlation between the variants and rs849177. Chr, chromosome; GIF, genomic inflation factor.

was used for all steps in the second QC stage and association tests.

Validation with the replication cohort

In the replication study, genotyping was done with the Affymetrix Axiom Korean Chip (K-CHIP) available through the K-CHIP consortium. K-CHIP was designed by the Center for Genome Science, Korea, National Institute of Health, Korea (4845–301, 3000–3031), and the K-CHIP consists of 827 783 variants. We conducted the first QC stage on cases and controls separately (online supplemental text and figure E1). After combining the cases and controls, the same steps used in the discovery cohort were performed in the second QC stage, except that participants were removed when the differences in the heterozygosity rates from the means were larger than four times the SD. After QC, 282 302 variants from 1864 participants were imputed, and some variants were removed using the same process as that used for the discovery cohort (online supplemental figure E1). Logistic regression was conducted by the same process with all covariates, except BMI, used in discovery analysis. BMI was not available for the controls in the replication cohort.

Genome-wide complex trait analysis

SNP-heritability for NTM-PD was estimated using Genome-wide complex trait analysis (GCTA) (V.1.26.0) with genome-based restricted maximum likelihood.[22] Only pruned variants with MAF larger than 0.1 were used for the estimation. We incorporated an estimate of the lifetime prevalence of NTM-PD in Korea as well as all covariates used for the GWAS except for the top 10 PC scores. The estimate of lifetime NTM-PD prevalence was 0.028, and was obtained by multiplying the point prevalence (39.6 cases/100 000 population in 2016)[2] by the average life span (70 years).[23]

Expression quantitative trait loci and PrediXcan analyses

Genes associated with the variants of interest and their expression levels were analysed using the Genotype-Tissue Expression (GTEx) Portal V.7 (https://www.gtexportal.org/). PrediXcan was also used with GTEx V.7 PredictDB to predict the gene expression levels in lung tissue using our imputed genotype data.[24] Linear regression was then conducted to compare the gene expression levels between cases and controls adjusting for the same covariates used in the GWAS.

Table 2 GWAS results for rs849177 in the discovery and replication cohorts

Variants	Chr	BP	Risk/protective alleles	Discovery cohort				Replication cohort				Combined P value
				P value	OR (95% CI)	Risk allele frequency*		P value†	OR (95% CI)	Risk allele frequency*		
						NTM-PD	Control			NTM-PD	Control	
rs849177‡	7	43 783 788	C/T	1.36×10^{-7}	2.34 (1.71 to 3.21)	0.70 (0.49(197)/0.42(168)/0.09(38))	0.57 (0.33(100)/0.49(149)/0.19(57))	0.017	1.29 (1.02 to 1.62)	0.68 (0.48(88)/0.40(73)/0.13(23))	0.63 (0.39(653)/0.47(794)/0.14(233))	4.92×10^{-8}

*Genotype frequencies and the exact number of genotypes are in parentheses and square brackets, respectively, in the following order: risk allele homozygotes (CC), risk allele heterozygotes (CT) and protective allele homozygotes (TT).
†One-tail p value.
‡rs849177 is the imputed variant. Its estimated squared-correlation values between the imputed and true genotype are 0.989 and 0.902 in the discovery and replication cohorts, respectively.
BP, base pair; Chr, chromosome; NTM-PD, non-tuberculous mycobacterial pulmonary disease.

Two-sample Mendelian randomisation analysis

We conducted a two-sample Mendelian randomisation (MR) analysis to identify the causal effect associated with the expression level of a candidate gene on the risk of NTM-PD development (online supplemental figure E2).[25] We selected genetic variants that were significantly associated with the expression level of a candidate gene in lung tissue from the GTEx V.7 data set. Then, we pruned out variants if their pairwise correlations with others were larger than 0.2 in our discovery cohort. Using the pruned-in variants as instrument variables (IVs) along with summary statistics from analyses of the GTEx data set and the discovery cohort, we applied the MR-Egger regression adjusting for correlations between the IVs.[25] The variant–NTM-PD and variant–candidate gene association estimates were obtained from our discovery cohort and the GTEx data set, respectively. To detect weak IVs explaining a relatively small proportion of variance in the expression level of a candidate gene,[26] we also applied the MR-pleiotropy residual sum and outlier (MR-PRESSO) test using the same IVs and summary statistics.[27] To adjust for correlations between the IVs, the CI and p value of the causal effect were estimated from 100 000 bootstrap replicates created by random sampling with replacement from the set of the selected variants before pruning. The same number of IVs were used for each MR-PRESSO test.

Transcriptome profiling analysis of NTM-infected macrophages

We conducted a transcriptome profiling analysis using GSE72821[28] and E-MTAB-1101,[29] two publicly available data sets. GSE72821 is an messenger RNA sequencing data set showing transcriptional events in THP-1-derived macrophages at 1, 4 or 24 hours post-infection with Smooth and Rough morphotypes of *Mycobacterium abscessus*.[28] We considered all 18 samples of macrophages infected with *M. abscessus* as the case group and 9 samples of uninfected macrophages as the control group. After genes whose expression level sums were smaller than one-tenth of the total sample size were removed, DESeq2 (V.1.14.1)[30] was used to identify genes with differential expression between the case and control groups.

E-MTAB-1101, an expression microarray data set, can decipher transcriptional responses in human macrophages after infection with *Mycobacterium avium*.[29] The CD14+ cells obtained from six donors were allowed to mature into macrophages by incubation for 5 days before they were infected with two *M. avium* subsp *hominissuis* isolates and one *M. avium* subsp *avium* isolate for 24 hours or left uninfected. We considered all nine samples of infected and three samples of uninfected macrophages as cases and controls, respectively. After normalisation, we analysed the expression data with limma (V.3.36.2).[31]

Gene Ontology enrichment analysis

Differentially expressed genes (DEGs) were defined based on a false discovery rate (FDR) of 0.05 from analyses of two data sets (GSE72821 and E-MTAB-1101). We imported those DEGs into the Gene Ontology (GO) consortium website (https://geneontology.org/) and applied Fisher's exact test.[32] The results were based on the database as of 21 February 2020.

Common variant and gene set analyses of previously identified genes

To examine whether genes identified in previous studies were associated with NTM-PD in our discovery cohort, we tested 43 genes previously reported to be associated with NTM-PD.[5 9–11 33 34]

Table 3 Association of rs849177 with *Mycobacterium avium* complex pulmonary disease and *Mycobacterium abscessus* complex pulmonary disease in the discovery cohort

	M. avium complex pulmonary disease (n=304)			*M. abscessus* complex pulmonary disease (n=80)		
Variants	OR (95% CI)	P value	Risk allele frequency*	OR (95% CI)	P value	Risk allele frequency*
rs849177	2.40 (1.71 to 3.37)	4.30×10^{-7}	0.70 (0.49/0.41/0.10)	2.14 (1.27 to 3.61)	4.21×10^{-3}	0.68 (0.46/0.44/0.10)

*The risk allele for rs849177 is C. Genotype frequencies are in parentheses in the following order: risk allele homozygotes (CC), risk allele heterozygotes (CT) and protective allele homozygotes (TT).

Their transcripts and 0.1 MB flanking regions were identified using the University of California Santa Cruz table browser (https://genome.ucsc.edu/cgi-bin/hgTables). We reported the common variant whose association with NTM-PD was strongest for each gene. In addition, we performed gene set analysis using the optimal sequence kernel association test.[35]

RESULTS

Putatively significant genome-wide association

After imputation, a total of 4 877 242 variants from 403 cases and 306 controls were considered for a GWAS to identify the disease susceptibility loci of NTM-PD. The clinical characteristics of these individuals are shown in table 1 and online supplemental table E1. The multidimensional scaling (MDS) analysis showed the presence of genetic heterogeneity between the discovery cohort and other populations in the 1000 Genomes Phase 3 data (online supplemental text and figure E3A), but no overall genetic heterogeneity between cases and controls was identified in the discovery cohort (online supplemental figure E3B). Only one peak around a significance level of 1×10^{-7} was observed, ranging from 43.6 to 43.8 MB of chromosome 7 (figure 1A). Among variants located in this region, rs849177 (chromosome 7p13; chr7:43783788) was most significantly associated with NTM-PD (OR, 2.34; 95% CI, 1.71 to 3.21; $p=1.36\times10^{-7}$; table 2). This association was consistently observed without adjustment for BMI ($p=2.85\times10^{-7}$; online supplemental figure E4). The frequencies of the rs849177 risk allele (C) were 0.70 for the cases and 0.57 for the controls. The genotype frequencies for rs849177 were 0.49 (risk allele homozygotes (CC); 197 participants), 0.42 (risk allele heterozygotes (CT); 168 participants) and 0.09 (protective allele homozygotes (TT); 38 participants) for the cases, and 0.33 (CC; 100 participants), 0.49 (CT; 149 participants) and 0.19 (TT; 57 participants) for the controls (table 2). The quantile-quantile plot shows that at 1.03 the value of the genomic inflation factor is very close to 1; therefore, type-1 analysis errors were well controlled (figure 1B). Figure 1C shows that the other variants near rs849177 are highly correlated with rs849177. However, the conditional analysis on rs849177 did not identify any other significant SNPs (online supplemental figure E5). Furthermore, the estimated heritability of NTM-PD explained by the pruned variants was 0.225 (p=0.266).

The association with rs849177 was consistent for patients with *M. avium* complex pulmonary disease (OR, 2.40; 95% CI, 1.71 to 3.37; $p=4.30\times10^{-7}$) and for those with *M. abscessus* complex pulmonary disease (OR, 2.14; 95% CI, 1.27 to 3.61; $p=4.21\times10^{-3}$; table 3). Although differences in the p values were primarily related to differences in the sample sizes, there were no significant differences in the associations of rs849177 with the two groups (p=0.769).

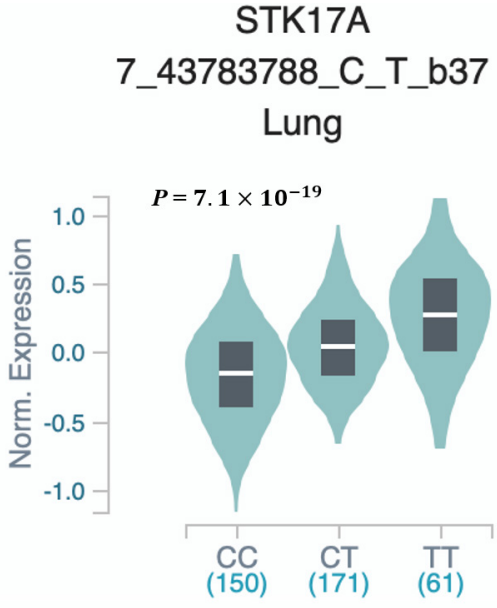

Figure 2 *STK17A* expression levels relating to the rs849177 genotypes in lung tissue from the Genotype-Tissue Expression database. The risk and protective alleles for rs849177 are C and T, respectively. The violin plot's width indicates the sample density. Overlaid boxes indicate IQR and centre-lines are the median. CC, risk allele homozygotes; CT, risk allele heterozygotes; TT, protective allele homozygotes; STK17A, serine/threonine kinase 17a.

Validation of the candidate SNP

We next validated the association study's findings using a replication cohort of 184 patients with NTM-PD and 1680 controls. The clinical characteristics of these individuals are shown in table 1 and online supplemental table E1. The MDS analysis showed that there was no overall genetic heterogeneity between the cases and controls in the replication cohort (online supplemental figure E3B). The nominal p value for rs849177 was 0.017 (OR, 1.29; 95% CI, 1.02 to 1.62; table 2). Moreover, as shown in table 2, the direction of effect of the variant was the same in both the discovery and replication cohorts. Using the Fisher's method, the combined p value for the discovery and replication cohorts was 4.92×10^{-8}. There were no significant variants other than rs849177 under the Fisher's method (online supplemental table E2).

Three-dimensional structural analysis

Serine/threonine kinase 17a (*STK17A*), cytochrome c oxidase assembly factor 1 homolog (*COA1*) and biliverdin reductase A (*BLVRA*) were found in the same topologically associating domain as rs849177 (online supplemental text and figure E6).

Candidate gene associated with NTM-PD rs849177 significantly affects *STK17A* expression in GTEx lung tissue ($p=7.1\times10^{-19}$; online supplemental table E3). In GTEx tissues other than lung tissue, the expression levels of *STK17A*, *COA1* and *BLVRA* were affected by rs849177 (online supplemental table E3). The expression quantitative trait loci analysis showed

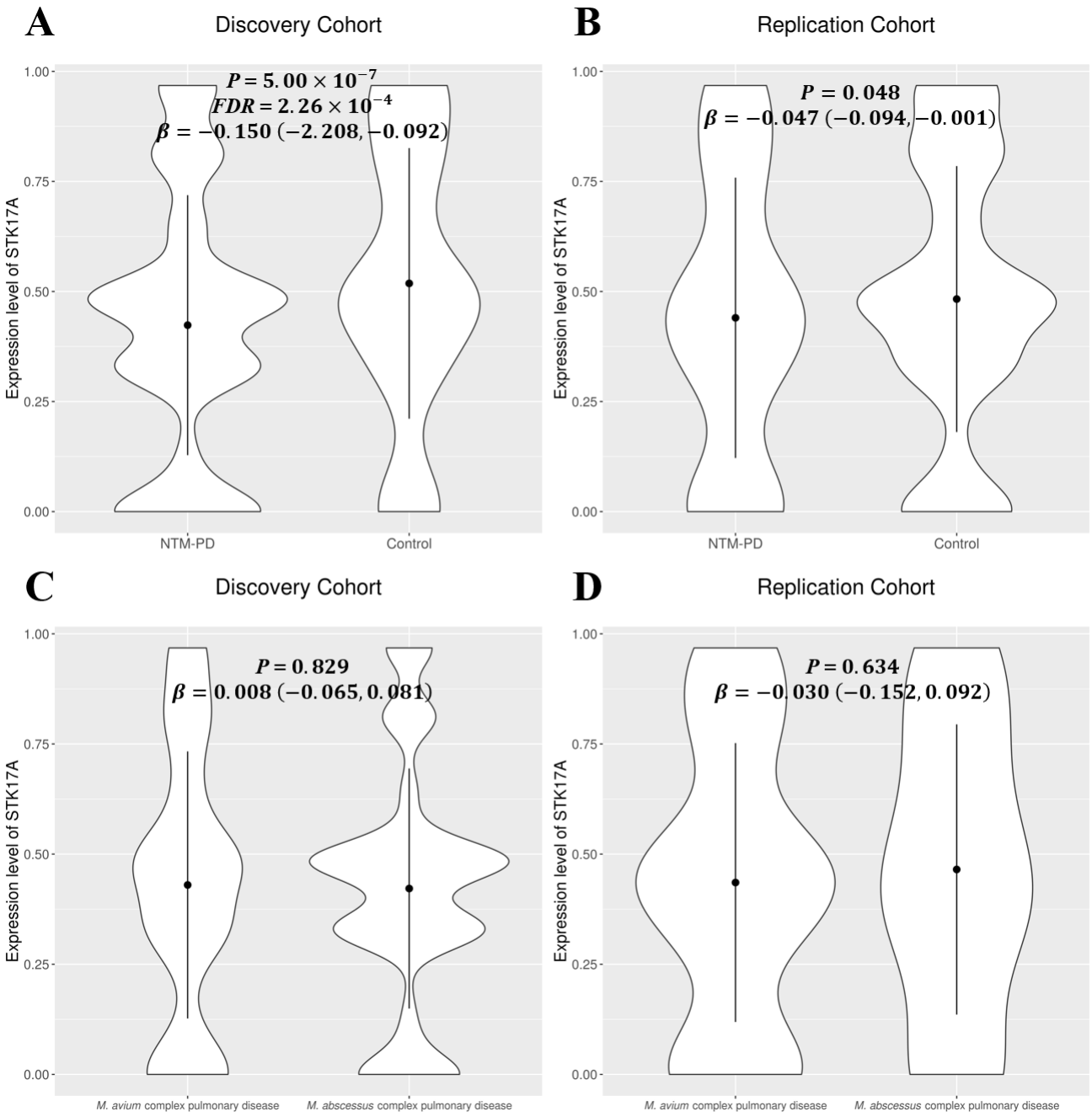

Figure 3 Expression levels of *STK17A* predicted by PrediXcan in patients with non-tuberculous mycobacterial pulmonary disease and controls in (A) the discovery and (B) replication cohorts, and in patients with *Mycobacterium avium* complex pulmonary disease and *Mycobacterium abscessus* complex pulmonary disease in (C) the discovery and (D) replication cohorts. The violin plot's width indicates the sample density. Overlaid black dots and vertical lines represent the means and one SD intervals, respectively. FDR, false discovery rate; NTM-PD, non-tuberculous mycobacterial pulmonary disease; STK17A, serine/threoninekinase 17a.

that individuals with C alleles at rs849177 tended to have lower *STK17A* expression levels in GTEx lung tissue (figure 2).

Next, lung tissue-specific gene expression levels were predicted using PrediXcan with our imputed genotype data. Consequently, *STK17A* was the most significantly differentially expressed gene between patients with NTM-PD and controls in the discovery cohort ($p=5.00\times10^{-7}$ and $FDR=2.26\times10^{-4}$; figure 3A), a result validated in the replication cohort (p=0.048; figure 3B). Additionally, the expression levels of *STK17A* did not differ between patients with *M. avium* complex pulmonary disease or *M. abscessus* complex pulmonary disease in the discovery cohort (p=0.829; figure 3C) and the replication cohort (p=0.634; figure 3D).

Causal effect of candidate genes on NTM-PD development

In the MR analysis, five variants associated with the expression levels of *STK17A* in lung tissue from the GTEx data set were used as IVs: rs7790685 (chr7:43597398), rs1100407 (chr7:43756898), rs757859 (chr7:43885849), rs929380 (chr7:43982139) and rs59505066 (chr7:44108863). The effects of these variants on the expression level of *STK17A* were significantly negatively associated with the effects of the variants on

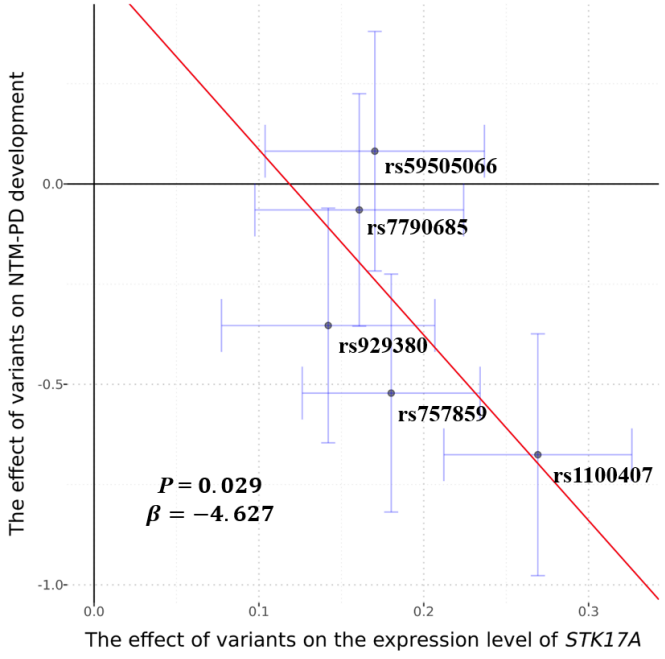

Figure 4 Plot of the variant–non-tuberculous mycobacterial pulmonary disease (NTM-PD) versus variant–*STK17A* expression regression coefficient for a Mendelian randomisation (MR) analysis with five variants. X and Y axes represent the effects of the variants on the expression level of *STK17A* and NTM-PD development, respectively. The MR-Egger regression estimate (β) is shown by a slope of red line. STK17A, serine/threonine kinase 17a.

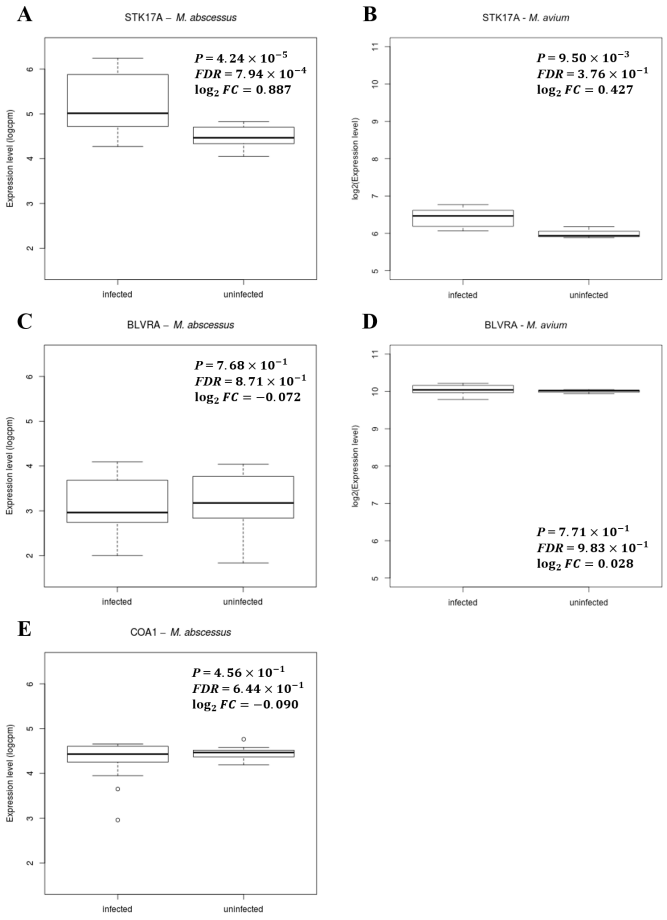

Figure 5 Expression levels of *STK17A* (A and B), *BLVRA* (C and D) and *COA1* (E) in *Mycobacterium abscessus*-infected or *Mycobacterium avium*-infected macrophages. A, C and E represent the results for GSE72821 (*M. abscessus*); B and D represent those for E-MTAB-1101 (*M. avium*). X and Y axes represent infected or uninfected macrophages and the expression levels of genes with log counts per million (logcpm) values or log2 expression levels, respectively. BLVRA, biliverdin reductase A; FC, fold change; FDR, falsediscovery rate; STK17A, serine/threonine kinase 17a.

NTM-PD development as shown using both the MR-Egger test (β, −4.627; 95% CI, −8.768 to −0.486; p=0.029; figure 4) and the MR-PRESSO test (β, −1.780; 95% CI, −2.974 to −1.618; p=0.049).

Differential expression of the candidate gene in NTM-infected macrophages

We performed transcriptome profiling using two publicly available data sets to examine the expression levels of *STK17A*, *BLVRA* and *COA1* in macrophages to determine whether their levels change on infection with NTM. The expression levels of *STK17A* increased significantly in both *M. abscessus*-infected macrophages ($p=4.24\times10^{-5}$ and $FDR=7.94\times10^{-4}$; figure 5A) and *M. avium*-infected macrophages ($p=9.50\times10^{-3}$ and $FDR=3.76\times10^{-1}$; figure 5B) compared with the uninfected macrophages at a nominal significance level of 0.05. However, neither *BLVRA* in the two data sets (figure 5C,D), nor *COA1* in the RNA-seq data set (figure 5E) was differentially expressed between the infected and uninfected macrophages. The differential expression level of *COA1* in the expression microarray data could not be determined because *COA1* was not included.

Enrichment of apoptotic processes in NTM-infected macrophages

The number of significant DEGs was 2939 and 143 in GSE72821 and E-MTAB-1101, respectively. GO analysis showed marked enrichment of apoptotic signalling pathways in NTM-infected macrophages (online supplemental table E4).

Association study of genes identified in previous studies

In the common variant analysis, rs3111229 (chr1:211363628) of repressor element-1-silencing transcription factor (REST) corepressor 3 (*RCOR3*) was significantly associated with NTM-PD as shown by the family-wise error rate of 0.05. In the gene set analysis, nuclear factor of activated T cells 2 (*NFATC2*) was significantly associated with NTM-PD with a FDR of 0.05 (online supplemental table E5).

DISCUSSION

Through GWAS, we identified rs849177 on chromosome 7p13 as a genetic variant associated with susceptibility to NTM-PD in a Korean population. The association was subsequently replicated in an independent cohort. In GTEx lung tissue, the rs849177 genotype was strongly correlated with the expression level of *STK17A*, a proapoptotic gene. Patients with NTM-PD were more likely to carry the risk allele at rs849177 associated with lower *STK17A* expression levels. Moreover, our findings using a MR approach supported the hypothesis that *STK17A* was causally related to risk of NTM-PD. Considering the transcriptome data showing upregulation of *STK17A* in macrophages on

infection with NTM, our findings suggest that the 7p13 variant associated with lower expression of *STK17A* could increase susceptibility to NTM-PD by impairing apoptosis of NTM-infected macrophages.

Activated macrophages can kill intracellular mycobacteria by enhancing phagosomal maturation,[36] induction of autophagy[37] and apoptosis.[38] Virulence of mycobacteria is largely attributed to their ability to survive inside the host's immune cells, particularly alveolar macrophages.[39] Apoptotic cell death, as part of the innate immune response, reduces mycobacterial viability. Host cell apoptosis is also important for the acquired immune response as apoptotic bodies containing mycobacterial antigens are phagocytosed by dendritic cells, which can lead to the presentation of such antigens and subsequent T cell activation.[40] It has been observed that virulent but not avirulent strains of *Mycobacterium tuberculosis* inhibit apoptosis of human alveolar macrophages.[41] This ability of virulent *M. tuberculosis* to inhibit apoptotic processes has been widely studied.[41 42]

STK17A encodes serine/threonine-protein kinase 17A, which is also known as the death-associated protein kinase-related apoptosis-inducing protein kinase (DRAK) 1.[43] DRAK1 is located in the nucleus and acts as a positive regulator of apoptosis.[44] Previous gene expression microarray data have shown that *STK17A* is upregulated in both *M. tuberculosis*-infected human macrophages[45] and *M. avium*-infected bovine macrophages.[46] The present study has also shown that *STK17A* is upregulated in human macrophages on infection with *M. avium* or *M. abscessus* by analysing the publicly available transcriptome data. Our data revealed that patients with NTM-PD had higher frequencies of the risk allele (C at rs849177) associated with lower *STK17A* expression. Given that apoptosis of alveolar macrophages infected with mycobacteria is crucial for human immunity, decreased *STK17A* gene expression might lead to reduced apoptosis of infected alveolar macrophages, allowing NTM to persist and replicate within the macrophages.

Cytochrome c oxidase assembly factor 1 homolog[47] and biliverdin reductase A[48] also play roles in apoptosis. However, unlike *STK17A*, the expression levels of *COA1* and *BLVRA* in lung tissue were not correlated with the rs849177 genotype. Furthermore, our analysis of recent transcriptome data showed that *COA1* and *BLVRA* were not differentially expressed in macrophages on infection with NTM, suggesting that they are not candidate NTM susceptibility genes associated with genetic variants on 7p13.

RCOR3, a gene associated with NTM-PD in gene-level analyses,[11] was also significantly associated with NTM-PD in the present study. RCOR3 is a member of the REST corepressors and its role in the pathogenesis of NTM-PD requires further exploration. In addition, *NFATC2* was significantly associated with NTM-PD in our gene set analysis. NFATc2 expression was downregulated in patients with NTM-PD.[33] As NFATc2 regulates the expression of interferon-γ[49] and tumour necrosis factor-α,[50] repression of NFATc2 might be associated with the development of NTM-PD.

A previous study reported that the SNP-heritability of pulmonary tuberculosis was 0.15.[23] As expected, the estimated heritability of NTM-PD calculated in our study was higher than that of pulmonary tuberculosis when the same method was used for estimating lifetime prevalence.

To appreciate the results of our study appropriately, its limitations have to be recognised. First, this study is moderately sized and could be underpowered, which might lead to failure in reaching the conventional genome-wide significance level of 5×10^{-8}. However, the combined p value from the Fisher method was 4.92×10^{-8}, which satisfies the conventional significance level. Second, the controls for the discovery and replication cohorts appear to have different allele frequencies at rs849177. However, as shown in online supplemental figure E3B, there was no overall genetic heterogeneity between the cases and controls within each cohort. Third, although we incorporated the data from the transcriptome analysis, the actual role for rs849177 and *STK17A* in human immunity against NTM infection awaits direct determination. The roles played by epithelial cells and immune cells other than macrophages should also be considered.

In conclusion, we identified a genetic variant on 7p13 that alters the expression of a proapoptotic gene and might be associated with susceptibility to NTM-PD in a Korean population. These findings offer insights into the pathogenesis of NTM-PD and may help with the search for novel therapeutics.

Author affiliations
[1]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Seoul National University Hospital, Jongno-gu, Seoul, Korea (the Republic of)
[2]Interdisciplinary Program of Bioinformatics, Seoul National University College of Natural Sciences, Seoul, Korea (the Republic of)
[3]Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea (the Republic of)
[4]Division of Pulmonary and Critical Care Medicine, Seoul Metropolitan Government–Seoul National University Boramae Medical Center, Dongjak-gu, Seoul, Korea (the Republic of)
[5]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Korea (the Republic of)
[6]Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Korea (the Republic of)
[7]Gwangju Alzheimer's disease and Related Dementia Cohort Research Center, Chosun University, Gwangju, Korea (the Republic of)
[8]Department of Biomedical Science, Chosun University, Gwangju, Korea (the Republic of)
[9]Aging Neuroscience Research Group, Korea Brain Research Institute, Daegu, Korea (the Republic of)
[10]Department of Public Health Sciences, Seoul National University Graduate School of Public Health, Seoul, Korea (the Republic of)
[11]Seoul National University Institute of Health and Environment, Seoul, Korea (the Republic of)

Contributors JJY and SW had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. JC, KP, JJY and SW contributed to the study concept and design, analysis and interpretation of data, and writing and revision of the manuscript. SMC, JL, CHL, JKL, EYH, DKK, YJL, JSP, YJC, HIY, JHL, CTL, NK, KYC, KHL and JS contributed to patient enrolment and data acquisition. All authors read and approved the final manuscript.

Funding This work was supported by the Seoul National University College of Medicine Research Fund (Grant number 0320160080).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study was approved by Institutional Review Board of Seoul National University Hospital (H-1605-111-763), Seoul National University Bundang Hospital (B-1610/366–303) and Chosun University (CHOSUN 2013-12-018-066).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Contact details: Jae-Joon Yim, MD, yimjj@snu.ac.kr.

ORCID iD
Choon-Taek Lee http://orcid.org/0000-0001-8030-3332

REFERENCES

1 Adjemian J, Olivier KN, Seitz AE, *et al*. Prevalence of nontuberculous mycobacterial lung disease in U.S. Medicare beneficiaries. *Am J Respir Crit Care Med* 2012;185:881–6.
2 Lee H, Myung W, Koh W-J, *et al*. Epidemiology of nontuberculous mycobacterial infection, South Korea, 2007-2016. *Emerg Infect Dis* 2019;25:569–72.
3 Filipe-Santos O, Bustamante J, Chapgier A, *et al*. Inborn errors of IL-12/23- and IFN-gamma-mediated immunity: molecular, cellular, and clinical features. *Semin Immunol* 2006;18:347–61.

4 Ristola MA, von Reyn CF, Arbeit RD, *et al*. High rates of disseminated infection due to non-tuberculous mycobacteria among AIDS patients in Finland. *J Infect* 1999;39:61–7.
5 Wu U-I, Holland SM. Host susceptibility to non-tuberculous mycobacterial infections. *Lancet Infect Dis* 2015;15:968–80.
6 Kim RD, Greenberg DE, Ehrmantraut ME, *et al*. Pulmonary nontuberculous mycobacterial disease: prospective study of a distinct preexisting syndrome. *Am J Respir Crit Care Med* 2008;178:1066–74.
7 Leung JM, Fowler C, Smith C, *et al*. A familial syndrome of pulmonary nontuberculous mycobacteria infections. *Am J Respir Crit Care Med* 2013;188:1373–6.
8 Adjemian J, Frankland TB, Daida YG, *et al*. Epidemiology of nontuberculous mycobacterial lung disease and tuberculosis, Hawaii, USA. *Emerg Infect Dis* 2017;23:439–47.
9 Szymanski EP, Leung JM, Fowler CJ, *et al*. Pulmonary nontuberculous mycobacterial infection. A multisystem, multigenic disease. *Am J Respir Crit Care Med* 2015;192:618–28.
10 Becker KL, Arts P, Jaeger M, *et al*. MST1R mutation as a genetic cause of lady Windermere syndrome. *Eur Respir J* 2017;49. doi:10.1183/13993003.01478-2016. [Epub ahead of print: 18 Jan 2017].
11 Chen F, Szymanski EP, Olivier KN, *et al*. Whole-Exome sequencing identifies the 6q12-q16 linkage region and a candidate gene, TTK, for pulmonary nontuberculous mycobacterial disease. *Am J Respir Crit Care Med* 2017;196:1599–604.
12 Kim SJ, Yoon SH, Choi SM, *et al*. Characteristics associated with progression in patients with of nontuberculous mycobacterial lung disease : a prospective cohort study. *BMC Pulm Med* 2017;17:5.
13 Griffith DE, Aksamit T, Brown-Elliott BA, *et al*. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am J Respir Crit Care Med* 2007;175:367–416.
14 Gombojav B, Song Y-M, Lee K, *et al*. The healthy twin study, Korea updates: resources for omics and genome epidemiology studies. *Twin Res Hum Genet* 2013;16:241–5.
15 Nijenhuis JT, Choi KY, Choi YY, *et al*. Differences between APOE carriers and Non-APOE carriers on neurocognitive tests: Jensen effects? *Am J Alzheimers Dis Other Demen* 2018;33:353–61.
16 Anderson CA, Pettersson FH, Clarke GM, *et al*. Data quality control in genetic case-control association studies. *Nat Protoc* 2010;5:1564–73.
17 Das S, Forer L, Schönherr S, *et al*. Next-Generation genotype imputation service and methods. *Nat Genet* 2016;48:1284–7.
18 McCarthy S, Das S, Kretzschmar W, *et al*. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016;48:1279–83.
19 Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012;28:112–8.
20 Zhao H, Mitra N, Kanetsky PA, *et al*. A practical approach to adjusting for population stratification in genome-wide association studies: principal components and propensity scores (PCAPS). *Stat Appl Genet Mol Biol* 2018;17.
21 Purcell S, Neale B, Todd-Brown K, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
22 Yang J, Lee SH, Goddard ME, *et al*. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88:76–82.
23 Speed D, Cai N, *et al*, UCLEB Consortium. Reevaluation of SNP heritability in complex human traits. *Nat Genet* 2017;49:986–92.
24 Gamazon ER, Wheeler HE, Shah KP, *et al*. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* 2015;47:1091–8.
25 Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44:512–25.
26 Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res* 2017;26:2333–55.
27 Verbanck M, Chen C-Y, Neale B, *et al*. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018;50:693–8.
28 Aulicino A, Dinan AM, Miranda-CasoLuengo AA, *et al*. High-Throughput transcriptomics reveals common and strain-specific responses of human macrophages to infection with Mycobacterium abscessus smooth and rough variants. *BMC Genomics* 2015;16:1046.
29 Agdestein A, Jones A, Flatberg A, *et al*. Intracellular growth of Mycobacterium avium subspecies and global transcriptional responses in human macrophages after infection. *BMC Genomics* 2014;15:58.
30 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol* 2014;15:550.
31 Smyth GK. Limma: linear models for microarray data. In: Smyth GK, Ritchie M, Thorne N, eds. *Bioinformatics and computational biology solutions using R and Bioconductor*. New York, NY: Springer, 2005: 397–420.
32 Mi H, Muruganujan A, Ebert D, *et al*. Panther version 14: more genomes, a new Panther GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* 2019;47:D419–26.
33 Cowman SA, Jacob J, Hansell DM, *et al*. Whole-Blood gene expression in pulmonary nontuberculous mycobacterial infection. *Am J Respir Cell Mol Biol* 2018;58:510–8.
34 Koh W-J, Kwon OJ, Kim EJ, *et al*. Nramp1 gene polymorphism and susceptibility to nontuberculous mycobacterial lung diseases. *Chest* 2005;128:94–101.
35 Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 2012;13:762–75.
36 Vergne I, Chua J, Singh SB, *et al*. Cell biology of mycobacterium tuberculosis phagosome. *Annu Rev Cell Dev Biol* 2004;20:367–94.
37 Gutierrez MG, Master SS, Singh SB, *et al*. Autophagy is a defense mechanism inhibiting BCG and Mycobacterium tuberculosis survival in infected macrophages. *Cell* 2004;119:753–66.
38 Molloy A, Laochumroonvorapong P, Kaplan G. Apoptosis, but not necrosis, of infected monocytes is coupled with killing of intracellular Bacillus Calmette-Guérin. *J Exp Med* 1994;180:1499–509.
39 Smith I. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev* 2003;16:463–96.
40 Briken V, Miller JL. Living on the edge: inhibition of host cell apoptosis by Mycobacterium tuberculosis. *Future Microbiol* 2008;3:415–22.
41 Keane J, Remold HG, Kornfeld H. Virulent Mycobacterium tuberculosis strains evade apoptosis of infected alveolar macrophages. *J Immunol* 2000;164:2016–20.
42 Sly LM, Hingley-Wilson SM, Reiner NE, *et al*. Survival of Mycobacterium tuberculosis in host macrophages involves resistance to apoptosis dependent upon induction of antiapoptotic Bcl-2 family member Mcl-1. *J Immunol* 2003;170:430–7.
43 Shiloh R, Bialik S, Kimchi A. The DAPK family: a structure-function analysis. *Apoptosis* 2014;19:286–97.
44 Sanjo H, Kawai T, Akira S. DRAKs, novel serine/threonine kinases related to death-associated protein kinase that trigger apoptosis. *J Biol Chem* 1998;273:29066–71.
45 Silver RF, Walrath J, Lee H, *et al*. Human alveolar macrophage gene responses to Mycobacterium tuberculosis strains H37Ra and H37Rv. *Am J Respir Cell Mol Biol* 2009;40:491–504.
46 Kabara E, Kloss CC, Wilson M, *et al*. A large-scale study of differential gene expression in monocyte-derived macrophages infected with several strains of Mycobacterium avium subspecies paratuberculosis. *Brief Funct Genomics* 2010;9:220–37.
47 Mick DU, Dennerlein S, Wiese H, *et al*. MITRAC links mitochondrial protein translocation to respiratory-chain assembly and translational regulation. *Cell* 2012;151:1528–41.
48 Baranano DE, Rao M, Ferris CD, *et al*. Biliverdin reductase: a major physiologic cytoprotectant. *Proc Natl Acad Sci U S A* 2002;99:16093–8.
49 Kiani A, García-Cózar FJ, Habermann I, *et al*. Regulation of interferon-gamma gene expression by nuclear factor of activated T cells. *Blood* 2001;98:1480–8.
50 Kaminuma O, Kitamura F, Kitamura N, *et al*. Differential contribution of NFATc2 and NFATc1 to TNF-alpha gene expression in T cells. *J Immunol* 2008;180:319–26.