Evaluation of the Host Genetic Effects of Tuberculosis-Associated Variants Among Patients With Type 1 and Type 2 Diabetes Mellitus

Huimin Zhong, *Emory University*
Matthew Magee, *Emory University*
Yunfeng Huang, *Emory University*
Qin Hui, *Emory University*
Marta Gwinn, *Emory University*
Neel Gandhi, *Emory University*
Yan Sun, *Emory University*

Journal Title: Open Forum Infectious Diseases

Volume: Volume 7, Number 4

Publisher: Oxford University Press Inc. | 2020-04-01, Pages ofaa106-ofaa106

Type of Work: Article | Final Publisher PDF

Publisher DOI: 10.1093/ofid/ofaa106

Permanent URL: https://pid.emory.edu/ark:/25593/vjk1v

Final published version: http://dx.doi.org/10.1093/ofid/ofaa106

Copyright information:

Open Forum Infectious Diseases

MAJOR ARTICLE

Evaluation of the Host Genetic Effects of Tuberculosis-Associated Variants Among Patients With Type 1 and Type 2 Diabetes Mellitus

Huimin Zhong,[1] Matthew J. Magee,[2] Yunfeng Huang,[1] Qin Hui,[1] Marta Gwinn,[1] Neel R. Gandhi,[1,2,3] and Yan V. Sun[1,4]

[1]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA, [2]Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA, [3]Division of Infectious Diseases, School of Medicine, Emory University, Atlanta, Georgia, USA, and [4]Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, Georgia, USA

Background. Understanding the link between tuberculosis (TB) and diabetes is increasingly important as public health responds to the growing global burden of noncommunicable diseases. Genetic association studies have identified numerous host genetic variants linked to TB; however, potential host genetic mechanisms linking TB and diabetes remain unexplored.

Methods. We used genetic and phenotypic data from the UK Biobank to evaluate the association of 6 previously reported TB-related host genetic variants (genome-wide significant associations from published studies) with diabetes. The study included 409 692 adults of European ancestry including 2177 with type 1 diabetes mellitus (T1DM) and 13 976 with type 2 diabetes mellitus (T2DM), defined by ICD-10 diagnosis codes.

Results. Of the 6 TB-associated single nucleotide polymorphisms (SNPs), 2 were associated with T1DM and 3 with T2DM, after adjusting for age, sex, body mass index, smoking, alcohol use, and population structure. After correction for multiple testing, SNPs rs2894257 and rs3135359 (*HLA-DRA-DQA1*) were associated with T1DM (rs2894257: odds ratio [OR], 1.32; 95% confidence interval [CI], 1.21–1.45; rs3135359: OR, 1.72; 95% CI, 1.57–1.88) and T2DM (rs2894257: OR, 1.11; 95% CI, 1.08–1.15; rs3135359: OR, 1.06; 95% CI, 1.025–1.096). The associations with T2DM weakened for rs2894257 and rs3135359 after further exclusion of probable T1DM cases defined by International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes. SNP rs4733781 on chromosome 8 (*ASAP1* gene) was associated with T2DM after exclusion of T1DM cases.

Conclusions. Our findings suggest that common host genetic effects may play a role in the molecular mechanism linking TB and diabetes. Future large genetic studies of TB and diabetes should focus on developing countries with high burdens of infectious and chronic diseases.

Keywords. tuberculosis; diabetes; host genetics; T1DM; T2DM.

More than 10 million incident cases of tuberculosis (TB) occur around the world each year [1]. In 2017, 1.6 million deaths from TB made it the most common cause of death from infectious disease and 1 of the top 10 causes of death worldwide. TB disease results from either rapid progression of a recently acquired *M. tuberculosis* (*Mtb*) infection or reactivation of a previous latent infection [2]. Rising noncommunicable disease (NCD) incidence is contributing to the profound burden of TB, as NCD comorbidities impact TB susceptibility and TB treatment outcomes. Diabetes mellitus (DM) increases susceptibility to TB, and an estimated 15% of global TB cases may be attributable to DM [3]. New biomedical approaches and technologies that provide insight into mechanisms of TB–diabetes synergy could help inform host-directed therapies for both diseases.

A number of genome-wide association studies (GWAS) have been performed to identify host genetic associations with either susceptibility to *Mtb* infection or the risk of progressing from *Mtb* infection to symptomatic, active TB disease [2]. GWAS studies have identified SNPs associated with susceptibility to active TB. These include rs4331426 (18q11.2; $P = 6.8\times10^{-9}$) [4] and rs2057178 ($P = 2.57\times10^{-11}$) [5], identified in African populations, and variants in the *ASAP1* gene on chromosome 8q24 (rs4733781: $P = 2.6\times10^{-11}$), identified in a Russian population [6]. These associations have been replicated in populations of different genetic ancestries such as Asian, European, and African (Supplementary Table 1) [5, 7–9]. Tian et al. conducted a GWAS of 23 common infections and infection-associated procedures and found several SNPs associated with *Mtb* infection in the 6p21.33 HLA region (eg, SNP rs2894257: $P = 8.16\times10^{-36}$; odds ratio [OR], 1.36; 95% confidence interval [CI], 1.33–1.39)

Received 16 December 2019; editorial decision 23 March 2020; accepted 24 March 2020.

Correspondence: Yan V. Sun, PhD, Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Rd. NE, Atlanta, GA 30322 (yan.v.sun@emory.edu).

Open Forum Infectious Diseases®

DOI: 10.1093/ofid/ofaa106

Table 1. Baseline Characteristics of the Cases and Controls of European Ancestry in the UK Biobank Cohort, 2006–2010

Variables	Controls Without DM (n = 310 267)		Cases With T1DM (n = 2177)		Cases With T2DM (n = 13 976)		P^a
	Mean ± SD or No. (%)		Mean ± SD or No. (%)		Mean ± SD or No. (%)		
Age, mean ± SD, y	56.7 ± 8.0		58.7 ± 7.7		60.8 ± 6.6		<.001
Sex							<.001
Female	169 482	54.62	880	40.42	5 220	37.35	
Male	140 785	45.38	1 297	59.58	8 756	62.65	
Body mass index, kg/m^2							<.001
Median (IQR)	26.54 (5.54)		29.23 (7.67)		31.15 (7.26)		
Mean ± SD	27.16 ± 4.57		30.07 ± 5.96		31.95 ± 5.82		
<25.0	107 518	34.65	479	22.00	1316	9.42	
≥25.0	202 749	65.35	1698	78.00	12 660	90.58	
Smoking status (current)							<.001
Yes	29 682	9.57	259	11.90	259	11.84	
No	280 585	90.43	1918	88.10	12 321	88.16	
Alcohol intake frequency							<.001
>3/wk	143 777	46.34	684	31.42	4511	32.28	
No or <3/wk	166 490	53.66	1493	68.58	9465	67.72	

Abbreviations: DM, diabetes mellitus; IQR, interquartile range; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

[a]For continuous and categorical variables, differences between cases (T1DM or T2DM) and controls were compared using the Kruskal-Wallis test and logistic regression test, respectively. Comparisons between the cases and the controls were statistically significant for both T1DM and T2DM. Therefore they were not listed separately.

[10]. Qi et al. also reported that *HLA-DQB1* 0201 (OR, 0.47; 95% CI, 0.26–0.69; Bonferroni-corrected $P = 2.80\times10^{-3}$), was associated with active TB disease [11].

Few studies have been conducted to evaluate shared host genetic factors between TB and diabetes. With diabetes prevalence increasing rapidly, particularly in countries with a high incidence of TB [12], further exploration of the genetic mechanisms underlying the comorbidity of TB and diabetes could be helpful for identifying vulnerable populations and appropriate interventions. In the present study, we investigated whether host genetic variants previously associated with *Mtb* infection susceptibility and TB progression were associated with prevalent diabetes mellitus. We used genotypic and phenotypic data from the UK Biobank (UKB) to examine the associations of 6 genetic variants with type 1 and type 2 diabetes [13].

METHODS

Study Design and Data Source

We conducted a cross-sectional genetic association analysis of baseline phenotype data from the UK Biobank. We focused on 452 264 participants of European ancestry with high-quality genetic data from 502 616 adults (aged 37 to 73 years) recruited between 2007 and 2010 [14]. After excluding closely related individuals, the present analysis included an unrelated sample of 342 574 participants of European ancestry with genetic data confirmed (Supplementary Figure 1). Social demographic factors (eg, education level, household income), lifestyle factors (eg, smoking history, alcohol consumption, physical activity), medical history (as documented by the 10th revision of the International Statistical Classification of Diseases and Related Health Problems [ICD-10] codes), and summary information relating to diagnoses were collected during the baseline visit and from hospital records [15]. Written informed consent was obtained from each participant before data collection.

Outcomes and Covariates

Study outcomes of interest included type 1 and type 2 diabetes mellitus (Supplementary Figure 1). Cases of type 1 diabetes mellitus (T1DM) were identified by the ICD-10 code "E10 insulin-dependent diabetes" (n = 2177). Cases of type 2 diabetes mellitus (T2DM) were defined by the primary and secondary ICD-10 diagnosis codes of "E11 non-insulin-dependent diabetes mellitus" (n = 13 976). Participants without T1DM or T2DM were considered the nondiabetic control group for this analysis (n = 310 267). Because we identified some participates who had both T1DM and T2DM codes, we created an additional T2DM-only case group (n = 12 502) by excluding those who had primary or secondary T1DM (n = 1474). Body mass index (BMI) was categorized as overweight/obese (BMI ≥ 25) and normal/underweight (BMI < 25). Self-reported cigarette smoking was dichotomized as current vs noncurrent. Self-reported alcohol consumption was dichotomized as high frequency (daily or almost daily, 3 or 4 times a week) and low frequency.

Genetic Data

Genotyping was performed using Affymetrix UK BiLEVE and UK Biobank Axiom arrays, and quality control and imputation were performed by a collaborative group headed by the Wellcome Trust Centre for Human Genetics, as previously described [15]. In addition to 3 genome-wide significant lead SNPs associated with TB (1 from each of 18q11.2, 11p13, and

Table 2. Genetic Associations of T1DM With TB-Related SNPs in the UK Biobank (2177 T1DM Cases, 310 267 Controls)

Chromosome	POS	SNP ID	Alleles[a]	Freq[b]	Model 1[c]		Model 2[d]	
					OR (95% CI)	*P* Value	OR (95% CI)	*P* Value
6	31628397	rs148844907	T/A	0.01	0.771 (0.555–1.073)	.123	0.768 (0.552–1.068)	.117
6	32390578	rs3135359	T/C	0.73	1.722 (1.573–1.886)	5.85×10^{-32}	1.721 (1.572–1.884)	7.15×10^{-32}
6	32433276	rs2894257	G/C	0.51	1.320 (1.205–1.445)	1.97×10^{-9}	1.315 (1.201–1.440)	3.27×10^{-9}
8	131296767	rs4733781	C/A	0.69	1.037 (0.972–1.105)	.272	1.037 (0.992–1.165)	.261
11	32364187	rs2057178	A/G	0.84	1.075 (0.992–1.165)	.076	1.073 (0.992–1.165)	.084
18	20190795	rs4331426	G/A	0.97	0.907 (0.773–1.065)	.234	0.906 (0.772–1.063)	.226

Abbreviations: CHROM, chromosome number; CI, confidence interval; Freq, frequency of effect allele; OR, odds ratio; POS, base pair position (GRCh37/hg19); SNP, single nucleotide polymorphism.

[a]The 2 alleles represent reference/effect alleles.

[b]Freq: frequency of effect allele.

[c]Model 1: basic adjusted model with age, sex, body mass index, principal components 1–10.

[d]Model 2: full adjusted model with age, sex, body mass index, smoking, alcohol, principal components 1–10.

ASAP1 loci), we selected 3 SNPs from the human leukocyte antigen (HLA) region for significant and independent associations with positive TB skin test [10]. Collectively, 6 independent TB-related genetic variants from 4 loci were identified through the published GWAS studies and included in this study (Tables 2–4). Details of the studies, including sample size, ethnic group, ORs, and *P* values, are given in Supplementary Table 1. The summarized 6 TB-related SNPs were the main exposure in this study of prevalent T2DM. Summary information of the candidate TB-related variants is given in Supplementary Table 2, including alleles, gene context, and imputation quality score.

Statistical Analyses

To identify common predisposing genetic factors for TB and diabetes, we performed cross-sectional analyses to assess the association of 6 TB-related genetic variants with prevalent T1DM and T2DM. Logistic regression models adjusted for social demographic factors (age, sex), lifestyle (smoke, alcohol consumption), and BMI at enrollment, as well as for population structure, using principal components (PCs) 1–10 from the genome-wide SNP data. PLINK2 (https://www.cog-genomics.org/plink/2.0/) was used to calculate genotype distributions, Hardy-Weinberg equilibrium (HWE), and summary statistics for genetic association (*P* value, odds ratio, with 95% CI). Test statistics comparing the phenotypes between subgroups were calculated using R, version 3.5.0 (https://www.r-project.org/). For each genetic variant, we performed 2 multivariate logistic regression analyses to calculate ORs and 95% CIs for the prevalence of T1DM/T2DM. The first model was adjusted for age, sex, BMI, and PC1-10. The second model included covariates from the first model, plus current smoking and alcohol consumption. Because multiple genetic associations were tested, we adjusted the statistical significance threshold using Bonferroni correction (.05/number of tested SNPs) to control for false positives.

RESULTS

The present study consisted of 310 267 individuals without T1DM or T2DM diagnosis, 2177 participants having T1DM

Table 3. Genetic Associations of T2DM With TB-Related SNPs in the UK Biobank (13 976 T2DM Cases, 310 267 Controls)

CHROM	POS	SNP ID	Alleles[a]	Freq[b]	Model 1[c]		Model 2[d]	
					OR (95% CI)	*P* Value	OR (95% CI)	*P* Value
6	31628397	rs148844907	T/A	0.01	0.974 (0.875–1.084)	.633	0.977 (0.877–1.088)	.67
6	32390578	rs3135359	T/C	0.73	1.061 (1.026–1.096)	.001	1.060 (1.025–1.096)	.001
6	32433276	rs2894257	G/C	0.51	1.111 (1.077–1.145)	1.42×10^{-11}	1.106 (1.073–1.140)	1.14×10^{-10}
8	131296767	rs4733781	C/A	0.69	1.030 (1.003–1.058)	.026	1.031 (1.004–1.058)	.024
11	32364187	rs2057178	A/G	0.84	0.988 (0.956–1.022)	.488	0.986 (0.954–1.020)	.427
18	20190795	rs4331426	G/A	0.97	0.973 (0.909–1.041)	.421	0.970 (0.906–1.037)	.371

Abbreviations: CHROM, chromosome number; CI, confidence interval; Freq, frequency of effect allele; OR, odds ratio; POS, base pair position (GRCh37/hg19); SNP, single nucleotide polymorphism.

[a]The 2 alleles represent reference/effect alleles.

[b]Freq: frequency of effect allele.

[c]Model 1: basic adjusted model with age, sex, body mass index, principal components 1–10.

[d]Model 2: full adjusted model with age, sex, body mass index, smoking, alcohol, principal components 1–10.

Table 4. Genetic Associations of T2DM (Excluding T1DM) and TB-Related SNPs in the UK Biobank (12 502 Non-T1DM T2DM Cases, 310 267 Controls)

Chromosome	POS	SNP ID	Alleles[a]	Freq[b]	Model 1[c]		Model 2[d]	
					OR (95% CI)	*P* Value	OR (95% CI)	*P* Value
6	31628397	rs148844907	T/A	0.01	0.942 (0.843–1.053)	.296	0.945 (0.845–1.056)	.319
6	32390578	rs3135359	T/C	0.73	1.024 (0.990–1.061)	.170	1.024 (0.989–1.060)	.187
6	32433276	rs2894257	G/C	0.51	1.084 (1.050–1.119)	8.65×10^{-7}	1.079 (1.045–1.114)	3.60×10^{-6}
8	131296767	rs4733781	C/A	0.69	1.030 (1.002–1.059)	.036	1.031 (1.003–1.060)	.033
11	32364187	rs2057178	A/G	0.84	0.972 (0.938–1.007)	.115	0.970 (0.936–1.005)	.096
18	20190795	rs4331426	G/A	0.97	0.981 (0.914–1.054)	.608	0.979 (0.911–1.051)	.552

Abbreviations: CHROM, chromosome number; CI, confidence interval; Freq, frequency of effect allele; OR, odds ratio; POS, base pair position (GRCh37/hg19); SNP, single nucleotide polymorphism.

[a]The 2 alleles represent reference/effect alleles.

[b]Freq: frequency of effect allele.

[c]Model 1: basic adjusted model with age, sex, body mass index, principal components 1–10.

[d]Model 2: full adjusted model with age, sex, body mass index, smoking, alcohol, principal components 1–10.

diagnosis codes, and 13 976 participants having T2DM diagnosis codes from the UK Biobank cohort from 2006 to 2010 (Supplementary Figure 1). Table 1 summarizes the distributions of age, sex, BMI, smoking, and alcohol consumption in the 2 case groups and controls at enrollment. Overall, members of the T1DM and T2DM case groups were older than the control group and included higher proportions of males, overweight or obese individuals, current smokers, and frequent consumers of alcohol.

Summarized in Table 2, the C alleles of rs2894257 and rs3135359 at the *HLA-DRA-DQA1* locus were associated with increased odds of T1DM after adjusting for age, sex, BMI, and top 10 PCs (model 1: rs2894257: OR, 1.32; 95% CI, 1.21–1.45; rs3135359: OR, 1.72; 95% CI, 1.57–1.89) and after adjusting for these variables in addition to smoking status and alcohol consumption (model 2: rs2894257: OR, 1.32; 95% CI, 1.20–1.44; rs3135359: OR, 1.72; 95% CI, 1.57–1.88). Associations with both SNPs remained statistically significant after Bonferroni correction for multiple testing.

Comparing 13 976 participants with T2DM diagnosis codes with non-DM controls, the TB risk alleles—2 SNPs in the HLA region (rs3135359 and rs2894257) and 1 in the *ASAP1* gene (rs4733781)—were associated with increased odds of T2DM (Table 3). After Bonferroni correction for multiple testing ($P = .05/6 = .0083$), the associations of both rs2894257 and rs3135359 remained significant.

To remove potential misclassification of T2DM as T1DM, we further excluded any T2DM patients who also had primary or secondary T1DM diagnosis from the T2DM case group (n = 12 502). As shown in Table 4, the OR for the association of rs2894257 with T2DM was reduced from 1.11 to 1.08 (model 1: 95% CI, 1.05–1.12; $P = 8.65\times10^{-7}$; 95% CI, 1.05–1.12; $P = 3.60\times10^{-6}$). None of the other 5 tested genetic associations were significant after Bonferroni correction for multiple testing. The SNP in the *ASAP1* gene, rs4733781, was marginally associated with T2DM at the nominal threshold (raw $P < .05$) with an OR of 1.03 (95% CI, 1.002–1.06).

Phenome-wide association studies (PheWAS) offer a complementary framework to investigate genetic associations across many disease traits simultaneously [16]. We examined genetic associations between target SNPs and clinical phenotypes using the GeneATLAS browser (http://geneatlas.roslin.ed.ac.uk/phewas/) [17]. Although PheWAS results for rs2894257 (C/G) were not available, we identified numerous diseases and traits associated with rs3135359. Phenome-wide associations for the T1DM-associated rs3135359 in the European ancestry subset of the UK Biobank data are summarized in Table 5. The TB-associated C allele of rs3135359 increased the risk of noncancer thyroid conditions by 81%. The top-ranked disease traits (ie, the lowest *P* values) associated with rs3135359 included thyroid conditions, neurological problems, blood cell traits, and other autoimmune diseases.

DISCUSSION

Based on >300 000 Caucasian participants from the UK Biobank, we found that a TB-related variant in the *HLA-DRA-DQA1* region (rs2894257) was associated with both T1DM (OR, 1.31; 95% CI, 1.21–1.45) and T2DM (OR, 1.08; 95% CI, 1.05–1.12). A second variant in this region, rs3135359, was also associated with T1DM but not T2DM after exclusion of T2DM cases who also had a T1DM diagnosis code. A third variant, rs4733781 (*ASAP1*), was moderately associated with T2DM (OR, 1.03; 95% CI, 1.00–1.06) but not T1DM.

The closest gene to SNP rs3135359 is *BTNL2*, which belongs to the butyrophilin-like B7 family of immunoregulators (proteins involved in immune surveillance) [18]. It mainly works as a negative T-cell regulator that decreases T-cell proliferation and cytokine release [19]. The naturally occurring mutations in *BTNL2* are associated with several diseases, including

Table 5. Summary of Phenome-Wide Search for Diseases and Traits Associated With rs3135359 C Allele in the UK Biobank

Trait	Beta	*P* Value	OR
Thyroid problem (not cancer)	0.0094	4.85×10^{-73}	1.18
Hypothyroidism/myxoedema	0.0083	4.29×10^{-68}	1.20
Insulin-dependent diabetes mellitus	0.0029	3.31×10^{-53}	1.55
Multiple sclerosis	–0.0019	1.48×10^{-47}	0.54
Demyelinating diseases of the central nervous system	–0.0020	3.68×10^{-46}	0.57
Chronic/degenerative neurological problem	–0.0026	2.25×10^{-45}	0.65
Hemoglobin concentration, g/dL	–0.0274	8.03×10^{-45}	–
Other rheumatoid arthritis	0.0032	1.03×10^{-41}	1.36
Other hypothyroidism	0.0055	2.81×10^{-40}	1.18
Rheumatoid arthritis	0.0032	1.80×10^{-37}	1.33
Disorders of thyroid gland (E00-E07)	0.0058	2.20×10^{-37}	1.16
Standing height, m	–0.1136	3.40×10^{-33}	–
Mean corpuscular hemoglobin, g/dL	–0.0376	4.58×10^{-29}	–
Mean platelet (thrombocyte) volume, fL	–0.0196	9.64×10^{-29}	–
Hayfever/allergic rhinitis	–0.0055	2.54×10^{-25}	0.90
Red blood cell (erythrocyte) distribution width, μm	0.0191	1.09×10^{-24}	–
Diabetes	0.0048	1.26×10^{-24}	1.11
Hematocrit, %	–0.0571	2.80×10^{-23}	–
Mean corpuscular hemoglobin concentration, g/dL	–0.0200	4.33×10^{-23}	–
Allergy/hypersensitivity/anaphylaxis	–0.0058	5.72×10^{-22}	0.92

rs3135359: minor allele frequency, 0.27; Hardy-Weinberg equilibrium *P* value = .09; imputation score, 1.00.

Abbreviation: OR, odds ratio.

sarcoidosis, ulcerative colitis, inflammatory bowel disease, T1DM, and prostate cancer [20].

The *ASAP1* gene (also known as *AMAP1* and *DDEF1*) encodes a multidomain ADP-ribosylation factor GTPase-activating protein, which is involved in the regulation of cytoskeletal dynamics, receptor recycling, and intracellular vesicle trafficking [6, 21]. Its expression is associated with poor prognosis for a variety of cancers and promotes cell migration, invasion, and metastasis. Although the functional role of *ASAP1* in diabetes is unclear, elevated expression of *ASAP1* mRNA was reported in adipose tissue isolated from obese mice and mice with diabetes compared with tissue isolated from wild-type mice [22]. Loss of *ASAP* led to delayed adipocyte development and reduced fat depot formation in mice [23]. Using the Genotype-Tissue Expression (GTEx) database to search for the functional links of genetic variants to gene expression levels, we found that rs4733781 is strongly associated with *ASAP1* expression in whole blood ($P = 6.5\times10^{-10}$), thyroid ($P = 1.5\times10^{-7}$), skin ($P = 1.7\times10^{-7}$), and lung ($P = 2.5\times10^{-6}$) tissues. The association we found between *ASAP1* and T2DM offers a clue to potential mechanisms underlying the comorbidity of TB and T2DM.

Several published studies have investigated the relationship between TB and diabetes, motivated by observed comorbidity between the 2 diseases [24–26], although very little is known about underlying molecular mechanisms [27]. Our study is another step toward understanding host genetic variants that could be related to both TB and diabetes. Although the relationship between TB and T2DM is suspected to be bidirectional, it is challenging to study the genetic risk of T2DM among people who have active TB or *Mtb* infection without large cohort studies that include host genetics. Most such studies, as well as PheWAS, have been conducted in US and European populations. Countries with high TB prevalence, which are best suited for studying factors underlying TB–diabetes comorbidity, are generally underrepresented in large genetics studies. Additional studies, particularly with a longitudinal design, will be needed to further explore relationships among host genetic risk factors for TB and diabetes.

In the present study, diabetes phenotypes were extracted from the baseline visit and from routine clinical health care follow-up records collected in the UK Biobank. Using ICD-10 codes from primary and secondary diagnoses, we were able to identify both T1DM and T2DM patients and examine specific genetic associations with each disease. The wide spectrum of phenotypes available in the UK Biobank study facilitates the PheWAS approach, which we used to search for additional disease phenotypes associated with *HLA-DRA-DQA1*.

This study had limitations. First, we identified T1DM and T2DM using diagnosis codes from the UK Biobank study. Although such disease definitions in large biobank studies permit efficient analysis of human genomics and diseases (including diabetes [28–30]), there is likely misclassification due to screening practices within the population. Potential underdiagnosis of diabetes may reduce the power of identifying genetic associations. Second, as some TB-associated genetic variants were identified in non-European populations, the underlying functional variants tagged by observed common variants may not exist or have low frequency in European ancestry (eg, UK Biobank Europeans). Such a difference in genetic ancestry may lead to negative associations using lead variants from previous GWAS. Ideally, we would repeat such association analyses in large genomic epidemiological studies of TB and DM with matching genetic ancestry (eg, Africans and South Asians). However, no large population studies with both phenotypic and genomic data in non-European ancestry are currently available to investigate the human genetics of TB and DM. Future human genetic and genomic studies need to establish large diverse cohorts representing global ancestries [31–33], which is particularly critical for TB—a global public health burden mostly in populations having non-European ancestries.

Our results support the hypothesis that TB-associated host genetic factors are also associated with T1DM and T2DM via immunologic functions. Moreover, common genetic factors and pathways may exist more broadly between other infectious and metabolic diseases, particularly via molecular mechanisms

associated with defense against infectious pathogens. A better understanding of the underlying relationships between infectious and chronic diseases may be particularly important for developing countries with a high prevalence of both.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

Financial support. This work was supported in part by National Institutes of Health grant R01AI139406. This research has been conducted using the UK Biobank Resource under Application Number '34031'.

Potential conflicts of interest. The authors have no conflicts of interest to report. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

Author contributions. Y.V.S. was responsible for the conception, organization, and execution of the research project. Y.V.S., H.Z., Y.H., and Q.H. were responsible for the design and execution of the statistical analysis. The first draft was written by Z.H. and Y.V.S. and developed further by M.J.M., M.G., and N.R.G. All co-authors contributed to critical review and editing of the manuscript.

References

1. World Health Organization. Global Tuberculosis Report 2018. Geneva: World Health Organization; **2018**.
2. Abel L, Fellay J, Haas DW, et al. Genetics of human susceptibility to active and latent tuberculosis: present knowledge and future perspectives. Lancet Infect Dis **2018**; 18:e64–75.
3. Lönnroth K, Roglic G, Harries AD. Improving tuberculosis prevention and care through addressing the global diabetes epidemic: from evidence to policy and practice. Lancet Diabetes Endocrinol **2014**; 2:730–9.
4. Thye T, Vannberg FO, Wong SH, et al; African TB Genetics Consortium; Wellcome Trust Case Control Consortium. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat Genet **2010**; 42:739–41.
5. Thye T, Owusu-Dabo E, Vannberg FO, et al. Common variants at 11p13 are associated with susceptibility to tuberculosis. Nat Genet **2012**; 44:257–9.
6. Curtis J, Luo Y, Zenner HL, et al. Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. Nat Genet **2015**; 47:523–7.
7. Chen C, Zhao Q, Hu Y, et al. A rare variant at 11p13 is associated with tuberculosis susceptibility in the Han Chinese population. Sci Rep **2016**; 6:24016.
8. Chimusa ER, Zaitlen N, Daya M, et al. Genome-wide association study of ancestry-specific TB risk in the South African coloured population. Hum Mol Genet **2014**; 23:796–809.
9. Lee SW, Lin CY, Chuang TY, Huang HH, Kao YH, Wu LSH. SNP rs4331426 in 18q11. 2 is associated with susceptibility to tuberculosis among female Han Taiwanese. J Microbiol Immunol Infect **2016**; 49:436–8.
10. Tian C, Hromatka BS, Kiefer AK, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nat Commun **2017**; 8:599.
11. Qi H, Zhang YB, Sun L, et al. Discovery of susceptibility loci associated with tuberculosis in Han Chinese. Hum Mol Genet **2017**; 26:4752–63.
12. Sola E, Rivera C, Mangual M, Martinez J, Rivera K, Fernandez R. Diabetes mellitus: an important risk factor for reactivation of tuberculosis. Endocrinol Diabetes Metab Case Rep **2016**; 2016:16-0035.
13. Day FR, Elks CE, Murray A, et al. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. Sci Rep **2015**; 5:11208.
14. Grant AV, El Baghdadi J, Sabri A, et al. Age-dependent association between pulmonary tuberculosis and common TOX variants in the 8q12-13 linkage region. Am J Hum Genet **2013**; 92:407–14.
15. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature **2018**; 562:203–9.
16. Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol **2013**; 31:1102–10.
17. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. Nat Genet **2018**; 50:1593–9.
18. Fukui M, Nakamura N, Nakano K, et al. HLA-associated cellular response to GAD in type 2 diabetes with antibodies to GAD. Endocr J **2000**; 47:753–61.
19. Fabregat M, Fernandez M, Javiel G, et al. The genetic profile from HLA and non-HLA loci allows identification of atypical type 2 diabetes patients. J Diabetes Res **2015**; 2015:485132.
20. Rich SS, French LR, Sprafka JM, et al. HLA-associated susceptibility to type 2 (non-insulin-dependent) diabetes mellitus: the Wadena City Health Study. Diabetologia **1993**; 36:234–8.
21. Hu X, Peng W, Chen X, et al. No significant effect of ASAP1 gene variants on the susceptibility to tuberculosis in Chinese population. Medicine (Baltimore) **2016**; 95:e3703.
22. King FJ, Hu E, Harris DF, et al. DEF-1, a novel Src SH3 binding protein that promotes adipogenesis in fibroblastic cell lines. Mol Cell Biol **1999**; 19:2330–7.
23. Schreiber C, Saraswati S, Harkins S, et al. Loss of ASAP1 in mice impairs adipogenic and osteogenic differentiation of mesenchymal progenitor cells through dysregulation of FAK/Src and AKT signaling. PLoS Genet **2019**; 15:e1008216.
24. Medellín-Garibay SE, Cortez-Espinosa N, Milán-Segovia RC, et al. Clinical pharmacokinetics of rifampin in patients with tuberculosis and type 2 diabetes mellitus: association with biochemical and immunological parameters. Antimicrob Agents Chemother **2015**; 59:7707–14.
25. Pearson F, Huangfu P, McNally R, et al. Tuberculosis and diabetes: bidirectional association in a UK primary care data set. J Epidemiol Community Health **2019**; 73:142–7.
26. Magee MJ, Salindri AD, Gujral UP, et al. Convergence of non-communicable diseases and tuberculosis: a two-way street? Int J Tuberc Lung Dis **2018**; 22:1258–68.
27. van Crevel R, Dockrell HM; TANDEM Consortium. TANDEM: understanding diabetes and tuberculosis. Lancet Diabetes Endocrinol **2014**; 2:270–2.
28. Said MA, Verweij N, van der Harst P. Associations of combined genetic and lifestyle risks with incident cardiovascular disease and diabetes in the UK Biobank study. JAMA Cardiol **2018**; 3:693–702.
29. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet **2018**; 50:1505–13.
30. Xue A, Wu Y, Zhu Z, et al; eQTLGen Consortium. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat Commun **2018**; 9:2941.
31. Gurdasani D, Barroso I, Zeggini E, Sandhu MS. Genomics of disease risk in globally diverse populations. Nat Rev Genet **2019**; 20:520–35.
32. Manolio TA. Using the data we have: improving diversity in genomic research. Am J Hum Genet **2019**; 105:233–6.
33. Peterson RE, Kuchenbaecker K, Walters RK, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. Cell **2019**; 179:589–603.