Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Huimin Zhong Date: April 22, 2019

Evaluate the host genetic effects of tuberculosis-associated variants on diabetes

By

Huimin Zhong
Master of Public Health

Global Epidemiology

_____

Yan Sun
Committee Chair

Evaluate the host genetic effects of tuberculosis-associated variants on diabetes

By

Huimin Zhong

B.Sc
University of Oregon
2017

Thesis Committee Chair: Yan Sun, Ph.D

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2019

Abstract

Evaluate the host genetic effects of tuberculosis-associated variants on diabetes
By Huimin Zhong

Background:

Numerous host genetic variants have been linked to tuberculosis (TB) from the genetic association studies. Comorbidities between TB and non-communicable diseases, such as diabetes, have been an increasing health burden. However, the molecular and genetic mechanism linking TB and diabetes hasn't been explored.

Methods:
We performed an association study to evaluate the effects of seven TB-related host genetic variants on type II diabetes (T2DM) using the genetic and phenotypic data from the UK Biobank. 409,692 participants of European ancestry including 13,976 T2DM cases and 2,177 T1DM cases defined by ICD-10 diagnosis codes.

Results:
One out of seven SNPs were significantly associated with T2DM adjusted for age, sex, BMI, smoke, alcohol use and population structure after correction for multiple testing. The C allele of SNP rs3135359 (*BTNL2* gene on HLA-A) is associated with increased risk for TB infection (OR 1.21 95% CI 1.18-1.24), increased risk for T2DM (OR 1.06 95% CI 1.03-1.10), and increased risk for T1DM (OR 1.72 95% CI 1.57-1.88). rs3135359 is strongly associated with thyroid disease and additional autoimmune diseases including multiple sclerosis and rheumatoid arthritis.

Conclusion:
Our findings support that the host genetic effects may partially explain observed disease comorbidities between TB susceptibility and T2DM. Our analyses demonstrated common genetic factors underlying chronic infection and non-communicable disease, particularly via immune functions.

Evaluate the host genetic effects of tuberculosis-associated variants on diabetes

By

Huimin Zhong

B.Sc
University of Oregon
2017

Thesis Committee Chair: Yan Sun, Ph.D

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2019

Acknowledgments

This thesis becomes a reality with the kind of support and help of many individuals. I would like to extend sincere thanks to all of them.

Foremost, I am grateful for the support and guidance of my thesis advisor Professor Yan Sun of the Rollins School of Public Health at Emory University. He was always interested and readily available to assist me in my work. The door to Prof. Sun office was always open whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work but steered me in the right direction whenever he thought I needed it.

I would also like to acknowledge Mr. Yunfeng Huang and Dr. Qin Hui of the Rollins School of Public Health at Emory University for sharing knowledge, technique, and all of the very valuable comments on this thesis.

Finally, I must express my very profound gratitude to my parents and to my colleges for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Huimin Zhong
April 22, 2019

Table of Contents

Chapter I: Background/Literature Review

Worldwide, Tuberculosis (TB) is one of the top 10 causes of death, that results from a rapid progression of a recently acquired *M. tuberculosis (Mtb)* infection or from reactivation.[1] The *Mtb* is spread through tiny droplets released into the air when coughs and sneezes.[2] After a TB infection, the bacteria could remain inside the body in an inactive state and cause no symptoms, which called latent TB infection.[1] It is able to progress into active TB over time. Thus, the treatment is necessary for the person with latent TB and to help control the spread of TB. Another consequence after TB infectious is active TB, which is the condition of sickness.[21] It could occur within the first few weeks after the *Mtb* infection with or it might occur years later.[1] The World Health Organization (WHO) estimated that 10.0 million people (9.0–11.1 million) developed TB disease in 2017 globally: 5.8 million men, 3.2 million women, and 1.0 million children.[36] There were cases in all countries and across all age groups. The severity of national epidemics varies widely among countries. In 2017, the three countries with the most incidences were India (27%), China (9%), and Indonesia (8%), together accounting for 44% of the total number of TB cases.[36] Tuberculosis remains one of the most deadly infectious diseases worldwide. Diabetes prevalence has been increasing rapidly, particularly in countries with high incidence of TB.[32] However, the comorbidity between TB and diabetes is mostly unexplained. Understanding the mechanism linking comorbid TB and diabetes

may shed lights on effective and efficient prevention and intervention for vulnerable populations.

To explore the potential host genetic factors related to TB and type II diabetes, peer-reviewed publications have been reviewed for the following questions.

1) Observed association between TB and T2DM

The link between diabetes mellitus and pulmonary TB has been known for a long time.[12] Many researchers believe there is a synergistic effect between those two diseases.[25] In 2007, a study showed that the diabetic women within age 20-45 and age 50-59 had significantly higher risk of TB.[13] Another study in 2018 found that incident rate of diabetes was substantially raised among individuals with a history of TB disease (incident rate ratio IRR=5.65 (95% CI 5.19 to 6.16)) compared with those without.[28] Another study reported that an early hyperglycemic environment caused by diabetes is likely being epigenetically imprinted on immune cells and alter the gene expression in patients with active TB.[33-34] In addition, the altered adipose tissue inflammation caused by type 2 diabetes could potentially contribute to the pathogenesis of TB disease.[27] These studies suggested that the relationship between TB and T2DM are likely bidirectional instead of one-way street.[24]

With the current global epidemic of obesity, there is now increasing public awareness in several countries of the association between TB and T2DM.[17]

However, there are few studies has been conducted to evaluate shared host factors between TB and T2DM. Therefore, we investigate host genetic variant to address the lack of mechanistic understanding of TB-T2DM comorbidity in the present study.

2) Possible host factors driving the association between TB and T2DM

Patients infected with *Mtb* will respond with active immunization system to defense the bacterium. A GWAS study of TB indicated *HLA* regions, *ASAP1*, *18q11.2* and downstream of *WT1* genomic regions that are work as components in cell wall biosynthesis, transcriptional regulation, and DNA repair pathways.[15] Since most of the SNPs found in GWAS studies were on identified immune genes, the mechanisms of immune defense exhibit biological plausibility and suggest other chronic diseases like diabetes were involved in the host containment of infection with TB.[29]

Numerous studies have explored the biological mechanism underlying the relationship between TB and T2DM. It's been found that both TB and T2DM diseases have endocrine alterations, which likely to play a role in certain immune-endocrine-metabolic associated disorders.[16] Then the T2DM was confirmed in patients with β-cell autoimmunity or HLA-DR genetic risk is more likely had developed diabetes.[3] A study in Mexico identified a significant association between glycemic control, increased TNF-α serum concentrations, and RIF pharmacokinetics in the patients had both TB and T2DM. These altered

metabolic and immune conditions are likely caused by anti-TB therapy and T2DM management,[25] but the directionality of the causal relationship hasn't been formally examined.–These results suggested a potential alter the mechanism of predisposition to TB.

3) host genetic factors associated with TB and T2DM

In the past decade, research for preventing and treating tuberculosis has focused on the genetic association between host and pathogen. Large numbers of genome-wide association studies (GWASs) based on microarray technologies have been performed using disease-specific definitions to identify genetic associations for tuberculosis.[1] A number of host genetic variants have been linked to tuberculosis from the past decade's studies.[5]

In a GWAS of TB susceptibility with 1,316 TB cases and 1,382 control, Thye et al. revealed SNP rs4331426 with P-value $< 5 \times 10^{-7}$.[33] Then there were two separate studies that repeated Thye's analysis in Asian with Han Taiwanese and Han Chinese population respectively. Chen repeated the analysis among Han Chinese group in 2016 and indicated that no gene expression levels were found to be associated with the genotype of rs4331426.[7] In 2016, Lee et al replicated genetic association of SNP rs4331426 with TB in the Han Taiwanese population, particularly in females ($p = 0.011$).[23] Unfortunately, both the repeated studies in Asian did not found any GWAS significant evidence for rs4331426 associated with TB susceptibility.

Another GWAS study published by Thye et al. in 2012 showed strong statistical evidence of associations in TB susceptibility with rs 2057178 and rs11031731 in Ghana population and replicated in Gambian, Russian and Indonesian population (N=22,680, $p=2.57 \times 10^{-11}$ for rs 2057178, $p=7.01 \times 10^{-9}$ for rs11031731). With HapMap3 imputed and 1000G imputed dataset, Chimusa et al. replicated the study conducted by Thye et al. 2012 in the South African population (N=733) and showed similar results for WT1's association with TB susceptibility but not GWAS significant ($p=2.71 \times 10^{-6}$).[8] In addition, Chen et al. replicated Thye's 2012 study in 2016 in Han, Chinese population (N=1,526) and found similar association between rs 2057178 and TB susceptibility but not GW significant.

In 2015, Curtis found an GW significantly association between TB and several variants located in the *ASAP1* gene on chromosome 8q24 ($P=2.60 \times 10^{-11}$ for rs4733781; $P=1.0 \times 10^{-10}$ for rs10956514, $P=4.1 \times 10^{-10}$ for rs12680942).[10,34] Later in 2016, Hu et al. repeated this study conducted by Curtis in western Chinese Han and Tibetan population (N=2,382) and found very similar association for two SNPs but not GW significant (p=0.86 for rs4733781 and p=0.45 for rs10956514). After that, Wang et al. replicated Thye's 2012 study in Chinese Xinjiang Muslim population (N=780) and found a non GW significant positive association between rs4733781 and TB susceptibility (Table S1).[35]

Tian et al. conducted GWAS studies of 23 common infections and infection-associated procedures, in which allele C of SNP rs3135359 was

significantly associated with a history of the positive TB skin test($P = 8.82 \times 10^{-21}$, OR= 1.21 95% CI 1.18-1.24). Qi et al. reported that gene HLA-DQB1 0201 (Bonferroni $P=2.80 \times 10^{-3}$), a candidate gene related to immune function, were associated with TB susceptibility.[30]

Besides the genetic factors, there are other risk factors could also play necessary roles in the mechanism of developing chronic diseases. It's been found the overweight and obese population (BMI >25) associated with higher risks for 48 adverse outcomes include cardio-metabolic categories.[11] Cassidy reported in 2016 that lifestyle factors include daily alcohol consumption and smoking habit associated with higher-risk characteristics of both CVD and type 2 diabetes.[6] In addition, those factors are likely to be clustered together, which means factors have a larger effect than just simply added.[6] Based on the summary from the past study, we decided to use social demographics factors (age, sex), lifestyle (smoke, alcohol consumption) and BMI, all collected at baseline, were chosen as the potential confounders and added to the analysis.

Chapter II: Manuscript

Evaluate the host genetic effects of tuberculosis-associated variants on diabetes

Abstract

Background:

Numerous host genetic variants have been linked to tuberculosis (TB) from the genetic association studies. Comorbidities between TB and non-communicable diseases, such as diabetes, have been an increasing health burden. However, the molecular and genetic mechanism linking TB and diabetes hasn't been explored.

Methods:

We performed an association study to evaluate the effects of seven TB-related host genetic variants on type II diabetes (T2DM) using the genetic and phenotypic data from the UK Biobank. 409,692 participants of European ancestry including 13,976 T2DM cases and 2,177 T1DM cases defined by ICD-10 diagnosis codes.

Results:

One out of seven SNPs were significantly associated with T2DM adjusted for age, sex, BMI, smoke, alcohol use and population structure after correction for multiple testing. The C allele of SNP rs3135359 (*BTNL2* gene on HLA-A) is associated with increased risk for TB infection (OR 1.21 95% CI 1.18-1.24), increased risk for T2DM (OR 1.06 95% CI 1.025-1.096), and increased risk for T1DM (OR 1.72 95% CI 1.57-1.88). The rs3135359 is strongly associated with thyroid disease and additional autoimmune diseases including multiple sclerosis and rheumatoid arthritis.

Conclusion:

Our findings support that the host genetic effects may partially explain observed disease comorbidities between TB susceptibility and T2DM. Our analyses demonstrated common genetic factors underlying chronic infection and non-communicable disease, particularly via immune functions.

Introduction

Worldwide, Tuberculosis (TB) is one of the top 10 causes of death, that results from a rapid progression of a recently acquired *M. tuberculosis (Mtb)* infection or from reactivation.[1] The *Mtb* is spread through tiny droplets released into the air when coughs and sneezes.[2] After a TB infection, the bacteria could remain inside the body in an inactive state and cause no symptoms, which called latent TB infection.[1] It is able to progress into active TB over time. Thus, the treatment is necessary for the person with latent TB and to help control the spread of TB. Another consequence after TB infectious is active TB, which is the condition of sickness.[21] It could occur within the first few weeks after the *Mtb* infection with or it might occur years later.[1] The World Health Organization (WHO) estimated that 10.0 million people (9.0–11.1 million) developed TB disease in 2017 globally: 5.8 million men, 3.2 million women, and 1.0 million children.[36] There were cases in all countries and across all age groups. The severity of national epidemics varies widely among countries. In 2017, the three countries with the most incidences were India (27%), China (9%), and Indonesia (8%), together accounting for 44% of the total number of TB cases.[36] Tuberculosis remains one of the most deadly infectious diseases worldwide. Diabetes prevalence has been increasing rapidly, particularly in countries with high incidence of TB.[32] However, the comorbidity between TB and diabetes is mostly unexplained. Understanding the mechanism linking comorbid TB and diabetes

may shed lights on effective and efficient prevention and intervention for vulnerable populations.

In the past decade, research for preventing and treating tuberculosis has focused on the genetic association between host and pathogen. Large numbers of genome-wide association studies (GWASs) based on microarray technologies have been performed using disease-specific definitions to identify genetic associations for tuberculosis.[1] A number of host genetic variants have been linked to tuberculosis from the past decade's studies.[5]

In a GWAS of TB susceptibility with 1,316 TB cases and 1,382 control, Thye et al. revealed SNP rs4331426 with P-value $< 5 \times 10^{-7}$.[33] Then there were two separate studies that repeated Thye's analysis in Asian with Han Taiwanese and Han Chinese population respectively. Chen repeated the analysis among Han Chinese group in 2016 and indicated that no gene expression levels were found to be associated with the genotype of rs4331426.[7] In 2016, Lee et al replicated genetic association of SNP rs4331426 with TB in the Han Taiwanese population, particularly in females ($p = 0.011$).[23] Unfortunately, both the repeated studies in Asian did not found any GWAS significant evidence for rs4331426 associated with TB susceptibility.

Another GWAS study published by Thye et al. in 2012 showed strong statistical evidence of associations in TB susceptibility with rs 2057178 and rs11031731 in Ghana population and replicated in Gambian, Russian and Indonesian population (N=22,680, $p=2.57 \times 10^{-11}$ for rs 2057178, $p=7.01 \times 10^{-9}$ for

rs11031731). With HapMap3 imputed and 1000G imputed dataset, Chimusa et al. replicated the study conducted by Thye et al. 2012 in the South African population (N=733) and showed similar results for WT1's association with TB susceptibility but not GWAS significant ($p=2.71 \times 10^{-6}$).[8] In addition, Chen et al. replicated Thye's 2012 study in 2016 in Han, Chinese population (N=1,526) and found similar association between rs 2057178 and TB susceptibility but not GW significant.

In 2015, Curtis found an GW significantly association between TB and several variants located in the *ASAP1* gene on chromosome 8q24 ($P=2.6 \times 10^{-11}$ for rs4733781; $P=1.0 \times 10^{-10}$ for rs10956514, $P=4.1 \times 10^{-10}$ for rs12680942).[10,34] Later in 2016, Hu et al. repeated this study conducted by Curtis in western Chinese Han and Tibetan population (N=2,382) and found very similar association for two SNPs but not GW significant (p=0.86 for rs4733781 and p=0.45 for rs10956514). After that, Wang et al. replicated Thye's 2012 study in Chinese Xinjiang Muslim population (N=780) and found a non GW significant positive association between rs4733781 and TB susceptibility (Table S1).[35]

Tian et al. conducted GWAS studies of 23 common infections and infection-associated procedures, in which allele C of SNP rs3135359 was significantly association with a history of the positive TB skin test($P = 8.82 \times 10^{-21}$, OR= 1.21 95% CI 1.18-1.24). Qi et al. reported that gene HLA-DQB1 0201 (Bonferroni $P=2.80 \times 10^{-3}$), a candidate gene related to immune function, were associated with TB susceptability.[30]

Since most of the SNPs found in GWAS studies were on identified immune genes, the mechanisms of immune defense exhibit biological plausibility and suggest other chronic diseases like diabetes were involved in the host containment of infection with TB.[5] There were numbers of studies published in the past century that have tried to explore the relationship between TB and T2DM. Many researchers believe there is a synergistic effect between those two diseases.[25] In 2007, a study showed that the diabetic women within age 20-45 and age 50-59 had significantly higher risk of TB.[13] Another study in 2018 found that incident rate of diabetes was substantially raised among individuals with a history of TB disease (incident rate ratio IRR=5.65 (95% CI 5.19 to 6.16)) compared with those without.[28] Another study reported that an early hyperglycemic environment caused by diabetes is likely being epigenetically imprinted on immune cells and alter the gene expression in patients with active TB.[33-34] In addition, the altered adipose tissue inflammation caused by type 2 diabetes could potentially contribute to the pathogenesis of TB disease.[27] These studies suggested that the relationship between TB and T2DM are likely bidirectional instead of one-way street.[24]

With the current global epidemic of obesity, there is now increasing public awareness in several countries of the association between TB and T2DM.[17] However, there are few studies has been conducted to evaluate shared host factors between TB and T2DM. Therefore, we investigate host genetic variant to

address the lack of mechanistic understanding of TB-T2DM comorbidity in the present study.

The UK Biobank (UKB) is a large, population-based cohort study examining the inter-relationships between environment, lifestyle, and genes.[11] Using available genotypic and phenotypic data from the UKB, we evaluated the host genetic effects of TB-associated variants on T2DM and T1DM. In addition, we completed a phenome-wide scan of the significantly associated SNP from our analysis to gain insights into other possible linked disease mechanisms.[12]

Methods

Study design and data source

To answer the research questions, a cross-sectional analysis was conducted on baseline phenotype and genotype data from the UK Biobank. Around 9.2 million invitations were mailed to recruit 502 616 adults aged between 37 and 73 years, between 2007 and 2010.[20] The self-reported ethnic background in the UK Biobank include 94.4% whites, 1.9% Asian or Asian British, 1.5% black or black British and other groups. After quality control of the UK Biobank dataset, there were 452,264 UK Biobank participants of European ancestry. After removed the related Caucasians group, this present analysis only included an unrelated white Caucasian group with genetic data confirmed (n=409,692) as shown in Figure 1. Genotyping was performed using the Affymetrix UK BiLEVE Axiom array on an initial approximately 50,000

participants. The Affymetrix UK Biobank Axiom array were used to genotype approximately 450,000 participants. Then the quality control and imputation were performed by a collaborative group headed by the Welcome Trust Centre for Human Genetics.[1] At the UKB phenotype dataset, social demographic factors (e.g., education level, household income), lifestyle factors (e.g., smoking history, alcohol consumption, physical activity), medical history (e.g. ICD-10 codes) and summary information relating to diagnoses made were collected in the baseline visit and during the hospital stay. Details of procedures have been previously published (https://www.ukbiobank.ac.uk/about-biobank-uk/). Participant's written informed consent was obtained prior to data collection. All data extracted were de-identified for analysis.

Definition of outcomes

Diagnosed T2DM was the outcome of interest and consisted of the main ICD-10 code and secondary ICD-10 code with T2DM (E11). Main ICD-10 codes show the patients reported and confirmed disease history and the secondary ICD-10 codes show the additional clinical detected disease. All distinct ICD-10-CM codes from each of the individuals' records were captured and translated into corresponding case groupings. ICD-10 code based T2DM was defined as participates recorded ICD-10 code "E11 non-insulin-dependent diabetes mellitus" in main or secondary ICD-10 code (n=13,976) as shown in Figure 1. The healthy reference group was defined as Caucasians participants with genetic

confirmation who do not have "E11 non-insulin-dependent diabetes mellitus" ICD-10 code, which was classified as the 'Without T2DM' control group in this analysis (n=310,970). With a brief check, there were some participates has both T1DM code and T2DM code. To avoid the possible misclassification of DM due to delay detection of T2DM, the secondary analysis was performed using T2DM cases excluding T1DM. Specifically, this confirmed T2DM group as defined as Caucasians with only T2DM (E11 non-insulin-dependent diabetes mellitus) in main or secondary ICD-10 code (n=12,502), which excluded those who also have T1DM (ICD-10 code as E10 insulin-dependent diabetes mellitus) in both main and secondary ICD-10 codes (n=1,474). With the significant SNPs found associated with T2DM, the effect's level was interested to compare the T1DM. Thus the T1DM group was build by ICD10-code "E10 Insulin-dependent diabetes" (n=2,177).

Genetic data

Collectively, there were seven TB-related genetic variants were identified through the published peer-reviewed GWAS studies in the past decade, details studies summary information included sample size, ethnic group, ORs and *p*-values were given in supplementary table S1. Those summarized seven tuberculosis-related SNPs were the main exposure in this study of prevalent T2DM. Summary information of the selected variants was given in supplementary table S2, includes alleles, gene context, and imputation score.

Covariates

Social demographics factors (age, sex), lifestyle (smoke, alcohol consumption) and BMI, all collected at baseline, were chosen as the potential confounders and added to the analysis. To improve accuracy, the principal components 1 to 10 were added as covariance in the analysis. Age was treated as a continuous variable. Sex was categorized as female and male. BMI was categorized as overweight/obese (≥25) and normal/underweight (<25). Smoke was dichotomous categorized as current, no now (previous, never, prefer not to answer). Alcohol consumption was also dichotomous categorized as high frequency (daily or almost daily, three or four times a week) and low frequency (never, once or twice a week, one to three times a month, special occasions only, prefer not to answer).

Statistical Analyses

All variables were tested for the association with the outcome of interest (p-value) and visually inspected whether they were distributed normally. The cross-section study analysis was conducted to access the host genetic effect of TB-related variances on the risk of T2DM. The analysis used PLINK2 version 2 APR 2019 (https://www.cog-genomics.org/plink/2.0/) to calculates case and control genotype distributions, the $\chi 2$ distribution, associated allelic P-value and allelic odds ratio (OR) with 95% CI. The algorithm was implemented as an R program version 3.5.0 (https://www.r-project.org/). We procedure that transforms

possible correlated variables into uncorrelated variables in smaller sample size by adding principle components 1-10 in the model. *Principal Component Analysis (*PCA) calculated through linear algebra and statistics. In detail, two multivariate analyses using binomial models were complete to capture the overall association between TB-related variances and T2DM with different confounders. The first model was the basic adjusted model includes the genotypes for the given SNPs and a list of PC 1-10, age, sex, and BMI. The second model was the complete model included the genotypes for the given SNPs and adjusted all interested exposure, PC 1-10, age, sex, BMI, smoke, and alcohol consumption. Both main and secondary ICD-10-CM code translation files were another input into the program. Ethic is proved to be an important confounder in genetic analysis, which was control at first places in these cases (only conducted on Caucasians group). Therefore, we have externally adjusted for ethnicity as a confounder of the bidirectional association between TB-associated variance and T2DM. The output tables list all SNP–disease associations and summarize the number of cases and controls for each diagnostic group, $\chi 2$ test statistic, P-value, allelic OR and Bonferroni level of significance. Because we tested genetic associations of multiple TB-related SNPs, we adjusted the statistical significance threshold using the Bonferroni correction (0.05/number of tested SNPs) to control for false positives.

Results

The present study consisted of 324, 946 individuals (46.16% of men, 53.86% of women) from the UK Biobank cohort from 2006 to 2010. Overall, 13,976 (4.3%) participants had T2DM recorded in ICD-10 codes. The mean age of entry (p<0.001) into the UK Biobank was 56.87 years old for the total study population, 60.77 years old for T2DM cases and 56.70 years old for the control group. There were 12,660 (90.58%) participants of cases group had BMI higher than 25, while 203,199 (65.34%) participants of the control group had a BMI higher than 25 (overweight or obese). The baseline characteristics of age, sex, BMI, smoking, alcohol consumption, and physical activity distribution for total, cases, and control group are given in table 1.

Table 2 shows the ORs with 95% CI of T2DM for TB-related genetic variances in basic adjusted model #1 and fully adjusted model #2. The rate of T2DM was significantly increased among those who had TB-related variances include rs3135359, rs12680942, rs10956514 and rs4733781. Although there have been small differences of p-values, those four SNPs remain significant across the two models. The C allele on SNP ID# rs3135359 at *BTNL2* gene located at HLA-A region was found associated with 6.1% (OR 1.061, 95% CI 1.026-1.096) and 6.0% (OR 1.060, 95% CI 1.025-1.096) greater likelihood of T2DM in the basic adjusted and full-adjusted model respectively. After multiple testing corrections, the Bonferroni level remains significant, 0.0056 in model 1, 0.0068 in model 2 (Table S3, S4). Then three out of three SNPs (rs12680942, rs10956514 and rs4733781) at

chromosome 8 that we focused on this analyze were tested significantly positively associated with T2DM across two models. The A allele on SNP ID rs12680942 at *ASAP1* gene was significantly positively associated with the risk of T2DM, OR=1.029, 95% CI 1.003-1.056 in adjusted model#1 and OR=1.030, 95% CI 1.003-1.057 in fully adjusted model #2. The ORs for the three SNPs on *ASAP1* gene remain very similar in numeric level with signature p-value <0.05 but shown non-statistically significant association after multiple testing corrections, Bonferroni significant level were 0.33 for rs12680942, 0.53 for rs10956514 and 0.28 for rs4733781 in full adjusted model (Table S4). Besides the SNPs tested with significant results, there were one more SNPs rs11031728 that have a positive association with T2DM but not statistically significant p-value. The detail analyzes results for rs11031728 in each model were included in supplemental table S4 and S5. The G allele on rs11031728 at the downstream of the WT1 gene was identified of negatively association with TB susceptibility in the previous studies, which showed a non-significantly positive association with T2DM across two models and similar results in (Non-T1) confirmed DM repeated analyze.

To avoid the misclassification of T1DM, the second analyses were done using confirmed T2DM (excluded any patients also had T1DM, Case/control=12,502/310,970). Results were included in table 4. In this analysis, SNPs rs3135359 was still positively associated with T2DM but no longer significant in both models, OR=1.024 with 95% CI (0.990-1.061) in model 1 and OR=1.024 with 95% CI (0.989-1.060) in model 2. Then the rs12680942 had a

significant p-value of positively associated with confirmed T2DM (OR=1.029 95% CI 1.001-1.058), but not significant after Bonferroni correction. Last, the rs4733781 was appeared with a significant p-value of positively associated with T2DM, OR=1.030 with 95% CI (1.002-1.059) in model 1 and OR=1.031 with 95% CI (1.003-1.060) in model 2, but not significant after Bonferroni correction. Other SNPs remain very similar results as the general T2DM analysis and details can be found as table 4.

Considering the SNPs rs3135359 had statistically significantly positive associated with T2DM, but no longer shows statistically significant in the repeated analysis with confirmed T2DM (after removed the misclassified T1DM). Additional analysis was done using T1DM as the case (N=2,177) and others as control (N=310,267). Results were included in table 5. The SNPs rs3135359 shows a very strong positive association with T1DM ($p= 5.85 \times 10^{-32}$ in model 1 and $p=7.15 \times 10^{-32}$). In detail, the C allele on SNP ID# rs3135359 at *BTNL2* gene located at HLA-A region found positively associated with 72.2% (OR 1.722, 95% CI 1.573-1.886) and 72.1% (OR 1.721, 95% CI 1.572-1.884) greater likelihood of T1DM in the basic adjusted and full-adjusted model respectively. After multiple testing corrections, the Bonferroni level remains strong significant, 7.02×10^{-31} in model 1, 8.58×10^{-31} in model 2. Besides rs3135359, there were four of the six rest SNPs had a non-significant positive association with T1DM (rs12680942, rs10956514, rs4733781 and rs2057178) and other two SNPs had a non-significant negative association with T1DM (rs11031728 and rs4331426).

Phenome-wide associations for the significant SNPs rs3135359 were concluded in Table 3. The top twenty 20 diseases with the lowest *p-value* were selected in table 3, which represents traits that had a significant association with the rs3135359. The diseases found to have a strongest association with allele C at rs3135359 is thyroid problem (not cancer) OR=1.81 ($p= 4.85 \times 10^{-73}$). Others include hypothyroidism/myxedema and E10 Insulin-dependent diabetes mellitus (i.e., T1DM)

Discussion

Based on >300,000 Caucasian participants in UK Biobank, the key finding of the present study confirmed that there is a host genetic effect of TB-related variants on T2DM and T1DM. In detail, this study suggested that the variances rs3135359 significantly associated with TB susceptibility found in formerly studies, which had a significant association with T1DM in the same direction. At the repeated analysis with confirmed T2DM as cases, the rs3135359 shows a positive association with T2DM but no longer significant in the statistic. This change of results brought up a question about the allele effect of rs3135359 on different subtypes of DM. After the exploration of additional analysis of T1DM, we found the C allele on rs3135359 at BTNL gene on HLA-A region has strong harmful effect on Type I diabetes (OR=1.72 95% CI 1.57-1.88) and moderate harmful effect on Type II diabetes (OR=1.02, 95% CI 0.99-1.06) according to the

full model. It's understandable that both T1DM and T2DM shared some physiological mechanisms.

Our results did not have statistically enough evidence to show that variation at *ASAP1* gene, downstream of the *WT1* gene and *18q11.2* are positively associated with DM. In detail, three variances (rs12680942, rs10956514 and rs4733781) on *ASAP1* gene, one variants rs11031728 at downstream of the WT1 gene and rs4331426 on 18q11.2 loci found significantly associated with TB susceptibility in previous studies, but showed none statistically significantly associations with T2DM and T1DM in reverse direction in this study.

To our knowledge, this is the study with the largest sample size to attempt to show the host genetic effect of TB-related genetic variants on T2DM prevalence in Caucasian. Form the statistical results of this study, there are significant host genetic effect plays an important role in the TB and T2DM association. In our analysis, we identified an association of several variants in *BTNL2* and *ASAP1* with T2DM. The risk of developing T2DM is higher in the host containment of infection with TB. It is plausible in the biological mechanism and also statistically proved in this study.

The physiological basis in literature review demonstrated the association of *BTNL2* gene *at HLA-A* region and *ASAP1* with several disorders respectively.[14] The *BTNL2* belongs to the butyrophilin-like B7 family of immunoregulators, which involved in immune surveillance.[18] It mainly works as a negative T-cell regulator, that decreasing T-cell proliferation and cytokine release.[14] The

naturally occurring mutations in *BTNL2* are associated with several diseases includes sarcoidosis, ulcerative colitis, inflammatory bowel disease, type 1 diabetes, prostate cancer, and etc.[31] Our results added one more significant association with T2DM. It is a reasonable physiological mechanism link to T2DM. Without proper function negative T-cell regulator, homeostasis between adipose tissue and the immune system cannot be maintained. Then it results in the altered insulin sensitivity. For *ASAP1* Gene, it encodes an ADP-ribosylation factor (ARF) GTPase-activating protein.[10,22] This gene possibly involved in the regulation of membrane trafficking and cytoskeleton remodeling.[10,22] No record indicates its relationship with any metabolic diseases. In this study, the key SNPs rs12680942, rs10956514 and rs4733781 in *ASAP1* found to be positively associated with T2DM (p-value <0.05, Bonferroni level >0.05), which indicated *ASAP1* plays an important role in the immune alteration. In the GeneATLAS browser (http://geneatlas.roslin.ed.ac.uk/phewas/), we confirmed the association of *ASAP1* with the disorder of mean platelet (thrombocyte) volume, mean corpuscular volume, and mean sphered cell volume at the variants level.

Additional studies will be needed to validate our cross-sectional association results. Our findings are an important step toward dissecting the host genetic effect contribution to variation in the chronic disease prevalence. Many experts believe the relationship between TB and T2DM is bidirectional. However, as far as we know, there is no adequate qualify cohort study to study the risk of T2DM among people who have had TB. Therefore, studies in high tuberculosis

prevalence countries are ideally needed to confirm the association. Unfortunately, those countries with high TB prevalence rate are mostly undeveloped or developing countries, which might have difficulties to conduct in practice due to the large sample size and research resources. Thus, a cohort study of diabetes among TB is also highly needed in developed countries to confirm this association.

Strengths and limitations

The key strength of our analysis is the size of the database. This large sample size is several orders of magnitude larger than most of the studies on this topic, which allow us to assess the host genetic effect of TB-related genetic variants in subgroups (TB susceptibility and early onset of PTB) on the risk of DM. The study also had several other strengths, including a good quality of data collected, availability of routine clinical healthcare record, and bio-lab detected diseases status. Our study's phenotype data was based on baseline visit and routine clinical healthcare follow up records collected in UK Biobank from 2006 to 2010. Disease status was self-recorded, tested and recorded in main ICD-10 codes by the nurse. With bio-lab test results, any additional founded disease was also recorded in secondary ICD-10 code by the nurse. Thus the main ICD-10 codes and secondary ICD-10 codes were highly likely being recorded accurately.

This study also had some limitations. First, this study was conducted in a cross-sectional study design, which cannot identify the causality of the

associations between TB-related variants and T2DM. Second, although we adjusted for key confounders of T2DM (age, sex, BMI, smoke, and alcohol), our estimate ORs could still be affected by residual confounding. Some variables like the physical activity and exercise level are hard to control in the analysis. Because the occupation type, transportation to work and exercise intensity and many other variables are linked to a person's activity level overall. Third, our examination of the data suggested that subtypes (E10, E11, E13, and E14) of diabetes may be less well recorded, may due to the T2DM test was not included in the routine healthcare test. Compare to the delay of T2DM diagnosed, T1DM status was better recorded because of its medical aid need and early diagnosed. Consider about the possibility for a small but unknown degree of misclassification between DM subtypes, we repeat our analysis with diabetes exclude T1DM (E10). Results are given in table 4. This secondary analysis confirms the significant positive association between certain SNPs and T2DM. Fourth, phenomic data was concluded from GeneAtlas of UK Biobanks dataset for the PheWAS summary. There could be other coding practices at a different institution or biobanks, which can affect the ability to replicate these results if the PheWAS is undertaken at another institution.

Summary

The results of the present study showed that the C allele of TB-related genetic variants rs3135359 at BTNL gene on HLA-A region has a strong harmful effect on T1DM (OR=1.721, 95% CI 1.572-1.884) and moderate harmful effect on T2DM (OR=1.024, 95% CI 0.989-1.060) according to the fully-adjusted model. Our analysis is cautiously added genomic epidemiological support to the hypothesis that the host genetic effects of TB-related variants on T2DM and postulated immunologic mechanisms. This may have implications for future healthcare for patients with TB history, who may not be routinely assessed for DM or informed that they might at higher risk of DM than others in the future. Moreover, our findings also yield insights into the metabolic diseases that are associated with infectious triggers. The relationship between infection, immunization and metabolic disorders are combined to a cycle. Indeed, our study identified and confirmed 20 more diseases have a significant association with the rs3135359 TB-related SNPs.

References

1. Abel L, Fellay J, Haas DW, et al. Genetics of human susceptibility to active and latent tuberculosis: present knowledge and future perspectives. *The Lancet Infectious Diseases.* 2018;18(3):e64-e75.
2. Ahmed M, Omer I, Osman SM, Ahmed-Abakur EH. Association between pulmonary tuberculosis and Type 2 diabetes in Sudanese patients. *International journal of mycobacteriology.* 2017;6(1):97-101.
3. Awa WL, Boehm BO, Rosinger S, et al. HLA-typing, clinical, and immunological characterization of youth with type 2 diabetes mellitus phenotype from the German/Austrian DPV database. *Pediatric diabetes.* 2013;14(8):562-574.
4. Basile KJ, Guy VC, Schwartz S, Grant SF. Overlap of genetic susceptibility to type 1 diabetes, type 2 diabetes, and latent autoimmune diabetes in adults. *Current diabetes reports.* 2014;14(11):550.
5. Blanco-Guillot F, Delgado-Sanchez G, Mongua-Rodriguez N, et al. Molecular clustering of patients with diabetes and pulmonary tuberculosis: a systematic review and meta-analysis. *PloS one.* 2017;12(9):e0184675.
6. Cassidy S, Chau JY, Catt M, Bauman A, Trenell MI. Cross-sectional study of diet, physical activity, television viewing and sleep duration in 233 110 adults from the UK Biobank; the behavioural phenotype of cardiovascular disease and type 2 diabetes. *BMJ open.* 2016;6(3):e010038.
7. Chen C, Zhao Q, Hu Y, et al. A rare variant at 11p13 is associated with tuberculosis susceptibility in the Han Chinese population. *Scientific reports.* 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. *Human molecular genetics.* 2013;23(3):796-809.
9. Cobat A, Poirier C, Hoal E, et al. Tuberculin skin test negativity is under tight genetic control of chromosomal region 11p14-15 in settings with different tuberculosis endemicities. *The Journal of infectious diseases.* 2014;211(2):317-321.
10. 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. *Nature genetics.* 2015;47(5):523.
11. Day FR, Elks CE, Murray A, Ong KK, Perry JR. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. *Scientific reports.* 2015;5:11208.
12. Dooley KE, Chaisson RE. Tuberculosis and diabetes mellitus: convergence of two epidemics. *The Lancet infectious diseases.* 2009;9(12):737-746.
13. Dyck RF, Klomp H, Marciniuk DD, et al. The relationship between diabetes and tuberculosis in Saskatchewan: comparison of registered Indians and other Saskatchewan people. *Canadian journal of public health = Revue canadienne de sante publique.* 2007;98(1):55-59.

14. Fabregat M, Fernandez M, Javiel G, Vitarella G, Mimbacas A. The Genetic Profile from HLA and Non-HLA Loci Allows Identification of Atypical Type 2 Diabetes Patients. *Journal of diabetes research.* 2015;2015:485132.
15. Farhat MR, Shapiro BJ, Kieser KJ, et al. Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. *Nature genetics.* 2013;45(10):1183.
16. Fernandez R, Diaz A, D'Attilio L, et al. An adverse immune-endocrine profile in patients with tuberculosis and type 2 diabetes. *Tuberculosis (Edinburgh, Scotland).* 2016;101:95-101.
17. Fisher-Hoch SP, Whitney E, McCormick JB, et al. Type 2 diabetes and multidrug-resistant tuberculosis. *Scandinavian journal of infectious diseases.* 2008;40(11-12):888-893.
18. Fukui M, Nakamura N, Nakano K, et al. HLA-associated cellular response to GAD in type 2 diabetes with antibodies to GAD. *Endocrine journal.* 2000;47(6):753-761.
19. Grant A, Sabri A, Abid A, et al. A genome-wide association study of pulmonary tuberculosis in Morocco. *Human genetics.* 2016;135(3):299-307.
20. 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. *The American Journal of Human Genetics.* 2013;92(3):407-414.
21. Hong EP, Go MJ, Kim H-L, Park JW. Risk prediction of pulmonary tuberculosis using genetic and conventional risk factors in adult Korean population. *PloS one.* 2017;12(3):e0174642.
22. Hu X, Peng W, Chen X, et al. No significant effect of ASAP1 gene variants on the susceptibility to tuberculosis in Chinese population. *Medicine.* 2016;95(21).
23. Lee S-W, Lin C-Y, Chuang T-Y, Huang H-H, Kao Y-H, Wu LS-H. SNP rs4331426 in 18q11. 2 is associated with susceptibility to tuberculosis among female Han Taiwanese. *Journal of Microbiology, Immunology and Infection.* 2016;49(3):436-438.
24. Magee M, Salindri A, Gujral U, et al. Convergence of non-communicable diseases and tuberculosis: a two-way street? *The International Journal of Tuberculosis and Lung Disease.* 2018;22(11):1258-1268.
25. Medellin-Garibay SE, Cortez-Espinosa N, Milan-Segovia RC, et al. Clinical Pharmacokinetics of Rifampin in Patients with Tuberculosis and Type 2 Diabetes Mellitus: Association with Biochemical and Immunological Parameters. *Antimicrobial agents and chemotherapy.* 2015;59(12):7707-7714.
26. Pan SW, Yen YF, Kou YR, et al. The Risk of TB in Patients With Type 2 Diabetes Initiating Metformin vs Sulfonylurea Treatment. *Chest.* 2018;153(6):1347-1357.
27. Pavan Kumar N, Nair D, Banurekha VV, et al. Type 2 diabetes mellitus coincident with pulmonary or latent tuberculosis results in modulation of adipocytokines. *Cytokine.* 2016;79:74-81.

28. Pearson F, Huangfu P, McNally R, Pearce M, Unwin N, Critchley JA. Tuberculosis and diabetes: bidirectional association in a UK primary care data set. *J Epidemiol Community Health.* 2019;73(2):142-147.
29. Png E, Alisjahbana B, Sahiratmadja E, et al. A genome wide association study of pulmonary tuberculosis susceptibility in Indonesians. *BMC medical genetics.* 2012;13(1):5.
30. Qi H, Zhang Y-B, Sun L, et al. Discovery of susceptibility loci associated with tuberculosis in Han Chinese. *Human molecular genetics.* 2017;26(23):4752-4763.
31. Rich SS, French LR, Sprafka JM, Clements JP, Goetz FC. HLA-associated susceptibility to type 2 (non-insulin-dependent) diabetes mellitus: the Wadena City Health Study. *Diabetologia.* 1993;36(3):234-238.
32. Sola E, Rivera C, Mangual M, Martinez J, Rivera K, Fernandez R. Diabetes mellitus: an important risk factor for reactivation of tuberculosis. *Endocrinology, diabetes & metabolism case reports.* 2016;2016.
33. Thye T, Owusu-Dabo E, Vannberg FO, et al. Common variants at 11p13 are associated with susceptibility to tuberculosis. *Nature genetics.* 2012;44(3):257.
34. Tian C, Hromatka BS, Kiefer AK, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nature communications.* 2017;8(1):599.
35. Wang X, Ma A, Han X, Litifu A, Xue F. ASAP1 gene polymorphisms are associated with susceptibility to tuberculosis in a Chinese Xinjiang Muslim population. *Experimental and therapeutic medicine.* 2018;15(4):3392-3398.
36. The World Health Organization. Global tuberculosis report 2018. 2018.

Tables:

Table 1: Baseline Characteristics of the Caucasian Participants (n=324,946) from the UK Biobank cohort in 2006-2010.

	Total study population (n=324,946)		Cases Patients with T2DM (n = 13,976)		Control - Patients without DM (n = 310,970)		*P
Variables	n (%)		n (%)		n (%)		
Age (Mean, SD)	56.87 ± 8.0		60.77 ± 6.59		56.70 ± 8.01		<0.0001
Sex							<0.0001
Female	175,020	53.86	5,220	37.35	169,800	54.60	
Male	149,926	46.14	8,756	62.65	141,170	45.40	
Body Mass Index (BMI)							<0.0001
< 25.0	109,087	33.57	1,316	9.42	107,771	34.66	
≥25.0	215,859	66.43	12,660	90.58	203,199	65.34	
Smoking status							<0.0001
Current	31,420	9.67	1,655	11.84	29,765	9.57	
Previous	114,697	35.30	6,605	47.26	108,092	34.76	
Never	177,665	54.68	5,621	40.22	172,044	55.32	
Prefer not to answer	1,164	0.36	95	0.68	1,069	0.34	
Alcohol intake frequency							<0.0001
Daily or almost daily	69,733	21.46	2,220	15.88	67,513	21.71	
Three or four times a week	78,843	24.26	2,291	16.39	79,708	24.50	
Once or twice a week	85,427	26.29	3,407	24.38	82,020	26.38	
One to three times a month	35,914	11.05	1,714	12.26	34,200	11.00	
Special occasions only	33,868	10.42	2,455	17.57	32,999	10.14	
Never	20,937	6.44	1,870	13.38	19,067	6.13	
Prefer not to answer	224	0.07	19	0.14	205	0.07	

*For continuous and categorical variables, differences between the two groups with or without TB-related SNPs were compared using Kruskal Wallis test and logistic regression test, respectively

Table 2 | Odds ratio of type II diabetes for TB-related SNPs effect in UK biobank (N=13,976)

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	32390578	**rs3135359**	T/C	**0.73**	**1.061 (1.026-1.096)**	0.0005	**1.060 (1.025-1.096)**	0.0006
8	131264033	**rs12680942**	G/A	**0.69**	**1.029 (1.003-1.056)**	0.0313	**1.030 (1.003-1.057)**	0.0278
8	131252758	**rs10956514**	G/A	**0.60**	**1.026 (1.001-1.052)**	0.0403	**1.026 (1.001-1.052)**	0.0438
8	131296767	**rs4733781**	C/A	**0.69**	**1.030 (1.003-1.058)**	0.0262	**1.031 (1.004-1.058)**	0.0236
11	32363616	rs11031728	C/G	0.16	1.011 (0.978-1.045)	0.5305	1.012 (0.979-1.047)	0.4668
11	32364187	rs2057178	A/G	0.84	0.988 (0.956-1.022)	0.4879	0.986 (0.954-1.020)	0.4271
18	20190795	rs4331426	G/A	0.97	0.973 (0.909-1.041)	0.4205	0.970 (0.906-1.037)	0.3705

CHROM chromosome, TB Tuberculosis, POS Base-pair coordinate, SNP Single Nucleotide Polymorphism, Freq frequency, OR odds ratio, CI confidence interval

a The alleles for each SNP are presented as low-risk/high-risk allele

b Freq Frequency of alternative allele

c Model 1 Basic adjusted model with age, sex, BMI, PC1-10

d Model 2 full adjusted model with age, sex, BMI, Smoke, Alcohol, PC1-10

Table 3 | Other phenotypes that are associated with the rs3135359 (C is the effect allele) in the UK Biobank

Trait	Beta	*p-value*	OR beta*
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
Haemoglobin concentration	-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
E00-E07 Disorders of thyroid gland	0.0058	2.20×10^{-37}	1.16
Standing height	-0.1136	3.40×10^{-33}	-
Mean corpuscular haemoglobin	-0.0376	4.58×10^{-29}	-
Mean platelet (thrombocyte) volume	-0.0196	9.64×10^{-29}	-
hayfever/allergic rhinitis	-0.0055	2.54×10^{-25}	0.90
Red blood cell (erythrocyte) distribution width	0.0191	1.09×10^{-24}	-
Diabetes	0.0048	1.26×10^{-24}	1.11
Haematocrit percentage	-0.0571	2.80×10^{-23}	-
Mean corpuscular haemoglobin concentration	-0.0200	4.33×10^{-23}	-
Allergy/hypersensitivity/anaphylaxis	-0.0058	5.72×10^{-22}	0.92

*MAF 0.2718; HWE 0.0924; Imputation score 0.9989

Table 4 | Odds ratio of Non-Type I, type II diabetes for TB-related SNPs effect in UK biobank (N=12,502)

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	32390578	rs3135359	T/C	0.73	1.024 (0.990-1.061)	0.1701	1.024 (0.989-1.060)	0.1872
8	131264033	**rs12680942**	G/A	0.69	**1.029 (1.001-1.058)**	0.0389	**1.030 (1.002-1.059)**	0.0352
8	131252758	rs10956514	G/A	0.60	1.024 (0.998-1.051)	0.0752	1.024 (0.997-1.051)	0.0808
8	131296767	**rs4733781**	C/A	0.69	**1.030 (1.002-1.059)**	0.0355	**1.031 (1.003-1.060)**	0.0325
11	32363616	rs11031728	C/G	0.16	1.027 (0.991-1.063)	0.1434	1.028 (0.993-1.065)	0.1201
11	32364187	rs2057178	A/G	0.84	0.972 (0.938-1.007)	0.1151	0.970 (0.936-1.005)	0.0957
18	20190795	rs4331426	G/A	0.97	0.981 (0.914-1.054)	0.6076	0.979 (0.911-1.051)	0.5518

TB Tuberculosis, POS Base-pair coordinate, SNP Single Nucleotide Polymorphism, OR odds ratio, CI confidence interval

a The alleles for each SNP are presented as low-risk/high risk allele

b Freq Frequency of alternative allele

c Model 1 Basic adjusted model with age, sex, BMI, PC1-10

d Model 2 full adjusted model with age, sex, BMI, Smoke, Alcohol, PC1-10

Table 5 | Odds ratio of type I diabetes for TB-related SNPs effect in UK biobank (N=2,177)

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	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}
8	131264033	rs12680942	G/A	0.69	1.036 (0.972-1.105)	0.2756	1.037 (0.973-1.106)	0.2638
8	131252758	rs10956514	G/A	0.60	1.042 (0.980-1.107)	0.1876	1.042 (0.980-1.107)	0.1899
8	131296767	rs4733781	C/A	0.69	1.037 (0.972-1.105)	0.2720	1.037 (0.992-1.165)	0.2608
11	32363616	rs11031728	C/G	0.16	0.937 (0.865-1.014)	0.1072	0.938 (0.866-1.016)	0.1176
11	32364187	rs2057178	A/G	0.84	1.075 (0.992-1.165)	0.0760	1.073 (0.992-1.165)	0.0844
18	20190795	rs4331426	G/A	0.97	0.907 (0.773-1.065)	0.2338	0.906 (0.772-1.063)	0.2263

TB Tuberculosis, POS Base-pair coordinate, SNP Single Nucleotide Polymorphism, Freq frequency, OR odds ratio, CI confidence interval

a The alleles for each SNP are presented as low-risk/high-risk allele

b Freq Frequency of alternative allele

c Model 1 Basic adjusted model with age, sex, BMI, PC1-10

d Model 2 full adjusted model with age, sex, BMI, Smoke, Alcohol, PC1-10

Figures/Figure Legends

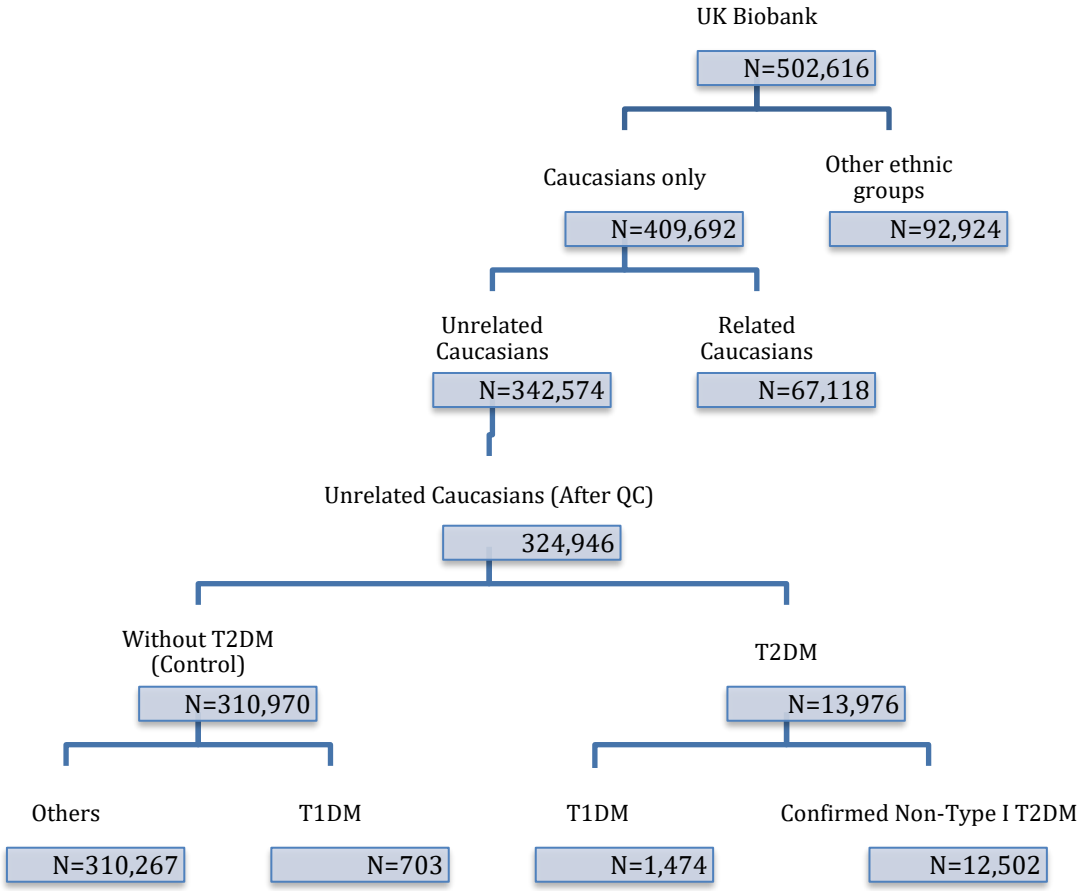

Figure 1 Flow chart of case and control group within the study population. T2DM, ICD-10 code based type II diabetes mellitus; T1DM, ICD10-code based Type I diabetes mellitus; T1DM under T2DM, misclassified T1DM inside the original T2DM group; QC, quality control.

Chapter III: Summary, Public Health Implications, Possible Future Directions

Summary:

The results of the present study showed one out of our seven SNPs were significantly associated with T2DM adjusted for age, sex, BMI, smoke, alcohol and PC 1-10, and corrected for multiple testing. The C allele of TB-related genetic variants rs3135359 at BTNL gene on HLA-A region previously found is associated with increased risk for TB infection (OR 1.21 95% CI 1.18-1.24), which also has a strong harmful effect on T1DM (OR 1.721, 95% CI 1.572-1.884) and moderate harmful effect on T2DM (OR 1.024, 95% CI 0.989-1.060) according to the fully adjusted model. The additional search identified the genetic association of rs3135359 with thyroid disease and several autoimmune diseases including multiple sclerosis and rheumatoid arthritis.

Public Health Implications:

Those findings support that the host genetic effects may partially explain disease comorbidities between TB infection and T2DM observed in epidemiologic studies. Our analysis is cautiously added genomic epidemiological support to the hypothesis that the host genetic effects of TB-related variants on T2DM and common genetic factors underlying chronic infection and non-communicable disease, particularly via immune functions. This may have implications for future healthcare for patients with TB history,

who may not be routinely assessed for DM or informed that they might at higher risk of DM than others in the future. Moreover, our findings also yield insights into the metabolic diseases that are associated with infectious triggers. The relationship between infection, immunization and metabolic disorders are combined to a cycle. Indeed, our study identified and confirmed 20 more diseases have a significant association with the rs3135359 TB-related SNPs.

Possible Future Directions:

Additional studies will be needed to validate our associations. Our findings are an important step toward dissecting the host genetic effect contribution to variation in the chronic disease prevalence. Many experts believe the relationship between TB and T2DM is bidirectional. However, as far as we know, there is no adequate qualify cohort study to study the risk of T2DM among people who have had TB. Therefore, studies in high tuberculosis prevalence countries are ideally needed to confirm the association. Unfortunately, those countries with high TB prevalence rate are mostly undeveloped or developing countries, which might have difficulties to conduct in practice due to the large sample size and research resources. Thus, a cohort study of diabetes among TB is also highly needed in developed countries to confirm this association.

Appendices

Supporting information

Table S1 | Selected of variants from published peer review study

CHROM	SNP	Sample size	Case/Control	Race/ethnicity	A1/A2[a]	Outcome	OR	(95% CI)	*P value* [b]
6	rs3135359	88,716	4,426 / 84,290	White (Caucasians)	T/C	TB susceptibility	1.21	(1.18-1.24)	8.82×10^{-21}
8	rs12680942	15,087	6,615 / 8,472	White (Russia)	G/A	TB susceptibility	0.85	(0.81-0.89)	4.10×10^{-10}
8	rs10956514	15,087	6,615 / 8,472	White (Russia)	G/A	TB susceptibility	0.85	(0.81-0.89)	1.00×10^{-10}
		2,382	1,031 / 1,351	Asian (Western Chinese Han and Tibetan population)	G/A	TB susceptibility	0.93	(0.77-1.12)	0.45
8	rs4733781	15,087	6,615 / 8,472	White (Russia)	C/A	TB susceptibility	0.84	(0.80-0.88)	2.6×10^{-11}
		2,382	1,031 / 1,351	Asian (Western Chinese Han and Tibetan population)	A/C	TB susceptibility	1.02	(0.86-1.22)	0.86
		780	400 / 380	Asian (Chinese Xinjiang Muslim population)	C/A	TB susceptibility	1.24	(1.00-1.54)	4.6×10^{-2}
11	rs11031728	22,680	8,821 / 13,859	Black + white + Asian (Ghana, replicated in Gambian, Russian and Indonesian.)	C/G	TB susceptibility	0.77	(0.71-0.84)	5.25×10^{-9}

		733	642 / 91	South African Coloured (SAC)	C/G	TB susceptibility	0.61	(0.50-0.75)	2.86×10^{-6}
11	rs2057178	22,680	8,821 / 13,859	Black + white + Asian (Ghana, replicated in Gambian, Russian and Indonesian.)	A/G	TB susceptibility	0.77	(0.71–0.84)	2.63×10^{-9}
		733	642 / 91	South African Coloured (SAC)	A/G	TB susceptibility	0.62	(0.50-0.75)	2.71×10^{-6}
		1,526	763 / 763	Asian (Han, Chinese population)	C/T	TB susceptibility	0.52	(0.35-0.78)	2.00×10^{-4}
18	rs4331426	11,425	1,316 / 1,382	Black (Akan, Ga-Adangbe, Ewe and several other ethnic groups from northern Ghana)	G/A	TB susceptibility	1.19	(1.13-1.27)	6.80×10^{-9}
		377	200 / 177	Asian (Han Taiwanese)	G/A	TB susceptibility	2.82	(1.11-7.18)	0.29
		1,526	763 / 763	Asian (Han, Chinese population)	A/G	TB susceptibility	1.06	(0.70-1.62)	0.84

CHROM chromosome #, SNP Single Nucleotide Polymorphism, OR odds ratio, CI confidence interval

a. A1/A2 are reference/derived alleles

b. P-value from publications

Table S2 | Summary of selected SNPs

CHROM	ID	POS	REF	ALT	INFO	Gene context	Frequency [a]
6	rs3135359	32390578	T	C	1.00	BTNL2 (HLA-A)	0.73
8	rs12680942	131264033	G	A	1.00	ASAP1	0.69
8	rs10956514	131252758	G	A	1.00	ASAP1	0.60
8	rs4733781	131296767	C	A	1.00	ASAP1	0.69
11	rs11031728	32363616	C	G	1.00	Downstream of the WT1 gene	0.16
11	rs2057178	32364187	A	G	1.00	Downstream of the WT1 gene	0.84
18	rs4331426	20190795	G	A	1.00	18q11.2	0.97

CHROM Chromosome, POS Position, INFO Imputation score, REF reference allele, ALT, alternate allele

a Frequency of alternative allele

Table S3 | Adjusted model with age, sex, BMI, PC1-10 (For T2DM cases=13,976)

Chromosome	POS	SNP ID	Alleles [a]	Freq [b]	BONF [c]	OR	(95% CI)	Z Statistics	*P value*
6	32390578	**rs3135359**	T/C	0.73	**0.01**	**1.061**	**(1.026-1.096)**	-3.50	0.0005
8	131264033	**rs12680942**	G/A	0.69	0.38	**1.029**	**(1.003-1.056)**	2.15	0.0313
8	131252758	**rs10956514**	G/A	0.60	0.48	**1.026**	**(1.001-1.052)**	2.05	0.0403
8	131296767	**rs4733781**	C/A	0.69	0.31	**1.030**	**(1.003-1.058)**	2.22	0.0262
11	32363616	rs11031728	C/G	0.16	1.00	1.011	(0.978-1.045)	-0.63	0.5305
11	32364187	rs2057178	A/G	0.84	1.00	0.988	(0.956-1.022)	-0.69	0.4879
18	20190795	rs4331426	G/A	0.97	1.00	0.973	(0.909-1.041)	0.81	0.4205

POS Base-pair coordinate, SNP Single Nucleotide Polymorphism, OR odds ratio, CI confidence interval

a The alleles for each SNP are presented as low-risk/high-risk allele

b Freq Frequency of alternative allele

c Bonferroni correction

Table S4 | Full adjusted model with age, sex, BMI, smoke, alcohol, PC1-10 (For T2DM cases=13,976)

CHROM	POS	SNP ID	Alleles [a]	Freq [b]	BONF [c]	OR	(95% CI)	Z Statistics	*P value*
6	32390578	**rs3135359**	T/C	0.73	**0.01**	**1.060**	**(1.025-1.096)**	-3.45	0.0006
8	131264033	**rs12680942**	G/A	0.69	0.33	**1.030**	**(1.003-1.057)**	2.20	0.0278
8	131252758	**rs10956514**	G/A	0.60	0.53	**1.026**	**(1.001-1.052)**	2.02	0.0438
8	131296767	**rs4733781**	C/A	0.69	0.28	**1.031**	**(1.004-1.058)**	2.26	0.0236
11	32363616	rs11031728	C/G	0.16	1.00	1.012	(0.979-1.047)	-0.73	0.4668
11	32364187	rs2057178	A/G	0.84	1.00	0.986	(0.954-1.020)	-0.79	0.4271
18	20190795	rs4331426	G/A	0.97	1.00	0.970	(0.906-1.037)	0.90	0.3705

CHROM chromosome, POS Base-pair coordinate, SNP Single Nucleotide Polymorphism, OR odds ratio, CI confidence interval

a The alleles for each SNP are presented as low-risk/high-risk allele

b Freq Frequency of alternative allele

b Bonferroni correction