UC Davis
UC Davis Previously Published Works

Title

A post-GWAS analysis of predicted regulatory variants and tuberculosis susceptibility

Permalink

https://escholarship.org/uc/item/6rp508xv

Journal

PLOS ONE, 12(4)

ISSN

1932-6203

Authors

Uren, Caitlin

Henn, Brenna M

Franke, Andre

et al.

Publication Date

2017

DOI

10.1371/journal.pone.0174738

Peer reviewed

A post-GWAS analysis of predicted regulatory variants and tuberculosis susceptibility

Caitlin Uren[1], Brenna M. Henn[2], Andre Franke[3], Michael Wittig[3], Paul D. van Helden[1], Eileen G. Hoal[1], Marlo Möller[1]*

1 SA MRC Centre for TB Research, DST/NRF Centre of Excellence for Biomedical TB Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, 2 Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York, United States of America, 3 Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Rosalind-Franklin-Strasse Kiel, Germany

* marlom@sun.ac.za

Abstract

Utilizing data from published tuberculosis (TB) genome-wide association studies (GWAS), we use a bioinformatics pipeline to detect all polymorphisms in linkage disequilibrium (LD) with variants previously implicated in TB disease susceptibility. The probability that these variants had a predicted regulatory function was estimated using RegulomeDB and Ensembl's Variant Effect Predictor. Subsequent genotyping of these 133 predicted regulatory polymorphisms was performed in 400 admixed South African TB cases and 366 healthy controls in a population-based case-control association study to fine-map the causal variant. We detected associations between tuberculosis susceptibility and six intronic polymorphisms located in *MARCO*, *IFNGR2*, *ASHAS2*, *ACACA*, *NISCH* and *TLR10*. Our post-GWAS approach demonstrates the feasibility of combining multiple TB GWAS datasets with linkage information to identify regulatory variants associated with this infectious disease.

Introduction

Genome-wide association studies (GWAS) have advanced the investigation of complex disease genetics and identified thousands of disease-associated variants. This study design compares allele frequencies of common genetic variants across the genome with phenotypic variation in large cohorts of cases and controls. GWAS is based on the premise that causal variants will be in linkage disequilibrium (LD) with the markers present on single nucleotide polymorphism (SNP) arrays. Since 2005, when the first GWAS was published [1], associations have been detected between numerous common genetic variants and several infectious diseases including TB [2–5].

More than 10 GWAS investigating TB susceptibility have been published to date (**Table 1**). These studies investigated the genetic factors associated with TB susceptibility in multiple populations. Thye et al. (2010) performed the first GWAS on TB susceptibility in a case-control cohort from Ghana and the Gambia and identified a region on chromosome 18q11.2 [6]. Within this region, there are numerous immune response genes such as cadherin 13 (*CDH13*), zinc finger protein 229 (*ZNF229*) and exportin 1 (*XPO1*). A meta-analysis which included data

from Ghana, the Gambia, Russia and Indonesia identified variants at 11p13 that were associated with TB susceptibility [7]. Chimusa et al. (2014) validated several of these loci and identified novel TB associations in a South African case-control cohort [8].

The majority of TB susceptibility variants previously identified are intronic (S1 Table) and may therefore have some regulatory functions. It has recently become feasible to predict regulatory effects of variants as computational tools, such as RegulomeDB [15] and Ensembl's Variant Effect Predictor, as well as information regarding the possible impact of regulatory regions have become available [15, 16]. We therefore applied a post-GWAS approach to TB susceptibility to identify possible variants contributing to disease development. A post-GWAS approach entails the use of previous GWAS associations and linkage disequilibrium (LD) data to identify further variants [and possibly the causative variant] that may be associated with the phenotype. This methodology was developed as pinpointing the exact targets of these associations is a challenge [17]. The post-GWAS analysis has previously been used to identify novel functional intronic variants associated with late-onset Alzheimer's disease [18], cardiovascular disease [19, 20] and human aging [21]. There has been no such analysis on susceptibility to TB.

Here we combine TB GWAS and candidate gene association studies and incorporate knowledge from RegulomeDB [15] and Ensembl's Variant Effect Predictor to fine-map putative regulatory variants that may predispose an individual to progress to active TB.

Methods

Study population

Sample collection was approved by the Health Research Ethics Committee of the Faculty of Health and Medical Sciences, Stellenbosch University (N95/072) and written informed consent was obtained from all study participants. Recruitment was done in two suburbs in the Western Cape, South Africa, where the incidence of TB is high (1340/100 000 population during 1996), although the HIV incidence at the time of sampling was low (~2% of population) [22]. Study participants self-identified as being from the South African Coloured (SAC) population. The admixed SAC population has genetic contributions from five ancestral populations. On average, Bantu-speaking populations contribute ~30%, the KhoeSan ~30%, Europeans 12–18% South Asian ~15%, and East Asian <10% [23–25]. Genotyping of ancestry informative markers (AIMS) was previously performed by Daya et al. (2013). These AIMS were used to infer admixture proportions using ADMIXTURE [25,26].

All study participants were HIV negative and unrelated. TB cases were bacteriologically confirmed (n = 400). The controls had no previous history of TB (n = 366) and were older than 18 years. Tuberculin skin tests (TST) were not performed, as the majority of adults in the communities are TST positive (> 80% of children older than 15 years) [27]. DNA was extracted from blood using the Nucleon BACC3 Kit (Amersham Biosciences, Buckinghamshire, UK).

Bioinformatics analysis

Data mining was done using the National Health Genome Research Institute's–European Bioinformatics Institute (NHGRI-EMBI) GWAS catalogue (http://www.ebi.ac.uk/gwas) and PubMed (www.ncbi.nlm.nih.gov/pubmed). Only single nucleotide polymorphisms (SNPs) that were reported to have $p < 0.05$ (after multiple testing correction during association tests) were recorded.

LD was calculated using SNAP [28]. These comparisons were made against Hapmap3 (release 2) data from 4 populations that best represented the ancestral populations of the SAC;

Table 1. Previous TB GWAS- results.

Population	Variant/Gene	Number of Cases	Number of Controls	Reference
Ghana	rs4331426 (gene desert)	921	1740	[6]
The Gambia		1316	1382	
Black, White, Asian from USA	rs4893980 (PDE11A)	48	57	[9]
	rs10488286 (KCND2)			
	rs2026414 (PCDH15)			
	rs10487416 (unknown gene)			
Thai and Japanese	Intergenic region between HSPEP1-MAFB	620	1524	[10]
Indonesia	rs1418267 (TXNDC4)	108	115	[11]
	rs2273061 (JAG1)			
	rs4461087 (DYNLRB2)			
	rs1051787 (EBF1)			
	rs10497744, rs1020941 (TMEFF2)			
	rs188872 (CCL17)			
	rs10245298 (HAUS6)			
	rs6985962 (PENK)			
Ghana	rs2057178 (WT1, intergenic)	2127	5636	[7]
The Gambia		1207	1349	
Russia		1025	983	
Indonesia		4441	5874	
South African Coloured	rs2057178, rs11031728 (WT1, intergenic)	642	91	[8]
	rs10916338,rs1925714 (RNF187)			
	rs6676375 (PLD5)			
	rs1075309 (SOX11)			
	rs958617 (CNOT6L)			
	rs1727757 (ZFPM2)			
	rs2505675 (LOC100508120)			
	rs1934954 (CYP2C8)			
	rs12283022,12294076 (DYNC2H1)			
	rs7105967,rs7947821 (DCUN1D5)			
	rs6538140 (E2F7)			
	rs1900442 (VWA8)			
	rs17175227 (SMOC1)			
	rs40363 (NAA60)			
	rs2837857 (DSCAM)			
	rs451390 (C2CD2)			
	rs3218255 (IL2RB)			
Russia	rs4733781,rs10956514,rs1017281,rs1469288, rs17285138,rs2033059, rs12680942 (ASAP1)	5530	5607	[3]
Morocco	rs358793 (Intergenic)	556	650	[12]
	rs17590261 (Intergenic)			
	rs6786408 (FOXP1)			
	rs916943 (AGMO)			
Uganda and Tanzania	rs4921437 (IL-12)	267	314	[13]
Iceland	rs557011, rs9271378 (located between HLADQA1 and HLA-DRB1)	8162	277643	[14]
	rs9272785 (HLA-DQA1)			

https://doi.org/10.1371/journal.pone.0174738.t001

Europeans (CEU), Han Chinese (CHB), Luhya (LWK) and Gujarati Indians (GIH). The Khoe-San ancestral component was represented by 2 ≠Khomani genomes. LD between the previously published SNPs and the ancestral genomes was calculated. An r^2 threshold of 0.8 and a window size of 500 kilobases were used as filters. A per population analysis was performed and SNPs were pooled across populations.

The potential functional impact of these variants was ascertained by RegulomeDB [15] and Ensembl's Variant Effect Predictor [16]. Only variants that had a RegulomeDB score of 1 (highest likelihood of a potential functional impact due to the variant) were used in further analysis due to the large number of variants. A summary of the filtering steps used are illustrated in **Fig 1**.

Genotyping

Genotyping was performed using the Agena MassARRAY® system (Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany). A total number of 133 SNPs were genotyped in 400 cases and 366 controls. Only SNPs that passed quality control (as determined by the confidence in allele call from the Typer Analyzer software package (version 4.0.20, Sequenom proprietary software)) were recorded.

Statistical analysis

The CaTS power calculator was used to perform power calculations [29]. The power to find a true deviation from the null hypothesis was calculated using a disease allele frequency of 0.06 and an alpha level of 0.05 to determine an odds ratio of 2. A TB disease prevalence of 1% was used in this analysis [30]. With the sample size available, the power to detect a deviation from null was calculated to be 98%.

All statistical analyses were done in R (www.r-project.org) using functions from the base R packages. The Fisher's exact test was used to calculate Hardy-Weinberg Equilibrium (HWE) *p*-values using functions from the *genetics* R package [31]. Logistic regression models were used to analyse the genotypic and allelic models. All models were adjusted for the confounding factors age, ancestry and sex by including these as covariates. The allelic models (recessive,

340 previously associated SNPs (S1 Table, *p* < 0.05)

1846 proxy SNPs from 5 ancestral populations in LD (r² > 0.8) with associated SNPs

133 SNPs with RegulomeDB results = 1 (S2 Table)

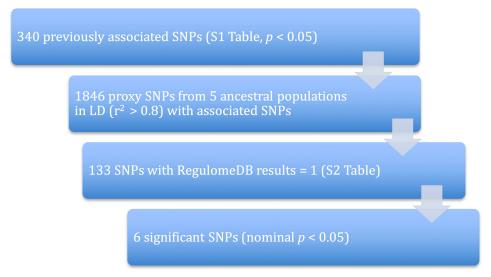6 significant SNPs (nominal *p* < 0.05)

Fig 1. Bioinformatics pipeline for the prioritization of variants.

https://doi.org/10.1371/journal.pone.0174738.g001

dominant and additive) were assessed and the model with the highest likelihood to correctly model the data was chosen [32]. Our SNP selection strategy was based on *a priori* evidence that the genes were associated with TB and Bonferroni corrections for multiple testing would be too stringent, risking the rejection of important findings [33–35]. For this reason, the Šidák step-up method was utilized [36]. Nominal *p*-values were corrected for multiple testing using the *multtest* package in R [37] and the cut-off for significance was *p* = 0.05.

Results

Descriptive statistics were generated for the cases and controls (**Table 2**). Age, sex and Khoe-San ancestry differed significantly between the cases and controls and were therefore all adjusted for in the logistic regression models. Prioritization of SNPs for genotyping is shown in **Fig 1**. Data mining identified 1800 SNPs that were found to be in LD ($r^2 > 0.8$) with the 230 SNPs previously associated with TB. After filtering for RegulomeDB scores of 1 (**S2 Table**), 133 SNPs remained and were genotyped in 400 TB cases and 366 healthy controls. All SNPs were in Hardy-Weinberg Equilibrium ($p > 0.01$). Of the 133 SNPs, 6 variants were found to be statistically significantly associated with TB susceptibility ($p < 0.05$) after adjusting for age, sex and ancestry (**Table 3**), but not after correcting for multiple testing. In this study, we view the methods of correction for multiple testing (including Bonferonni) to be too conservative for this analysis; there is *a priori* evidence that variants in LD with those reported here were associated with TB susceptibility (**Table 1**) [33–35]. For completeness however, nominal *p* values adjusted for multiple testing using the Šidák method are reported in **Table 3** [36].

The SNPs rs7599352, rs2600665, rs829161 and rs12233670 showed statistically significant associations with resistance to TB (nominal *p* = 0.021, 0.006, 0.037 and 0.022, respectively). The minor homozygote (T/T) of rs7599352 was found in fewer cases than controls and the genotype has an odds ratio of 0.416 (95% CI = 0.197–0.842). In contrast, rs2600665 followed an additive model of penetrance and thus for every copy of the G allele, an individual is ~30% less likely to progress to active disease. The ancestral homozygote (C/C) of rs829161 was found in double the amount of cases than controls and was associated with susceptibility to TB (nominal *p* = 0.037). This genotype has an odds ratio of 2.274 (95% CI = 1.13–4.73). The heterozygote genotype of rs12233670 was associated with TB susceptibility and with an odds ratio of 0.699 (95% CI = 0.491–0.994). Two SNPs, namely rs4687614 and rs2284555, were associated with increased susceptibility to TB. For every copy of the minor G allele of rs4687614, there was an increase of ~50% in the likelihood of progressing to active TB (*p* = 0.040, 95% CI = 1.09–1.98). In addition, rs2284555 was associated with TB susceptibility (*p* = 0.043); both the

Table 2. Case-control sample characteristics and TB susceptibility modelling results.

	TB Cases (n = 398)	Controls (n = 360)	*p* value*
Age (mean ± SD)	36.55 ± 11.26	30.69 ± 12.80	0.0001
Number of males (proportion)	211 (0.53)	111 (0.28)	< 0.0001
KhoeSan [IQR]	0.30 [0.20–0.39]	0.27 [0.18–0.36]	0.0224
West African [IQR]	0.27 [0.16–0.39]	0.25 [0.15–0.37]	0.3187
European [IQR]	0.18 [0.08–0.28]	0.19 [0.12–0.28]	0.7804
South Asian [IQR]	0.12 [0.03–0.19]	0.14 [0.06–0.22]	0.2767
East Asian [IQR]	0.09 [0.03–0.16]	0.10 [0.05–0.17]	NA[a]

[a] The East Asian component was not added to the model to avoid linear dependency.

SD standard deviation, IQR, interquartile range

*Statistic is an indication of the significance of the association between each factors with TB after adjusting for the other factors.

https://doi.org/10.1371/journal.pone.0174738.t002

Table 3. Association results of statistically significant regulatory SNPs.

rsID	Controls		HWE[c] p-val	TB Cases		HWE p-val	Association p-val Adjusted[d]	Model of penetrance	OR	2.5% CI	97.5% CI	Šidák p-val
	Count[a]	Prop[b]		Count	Prop							
rs2284555 (IFNGR2)												
	352		0.457	390		0.208						
G/G	37	0.11		25	0.06		0.043					0.997
A/G	145	0.41		168	0.43			Genotypic	2.146	1.174	3.988	
A/A	170	0.48		197	0.51			Genotypic	2.012	1.108	3.716	
G	219	0.31		218	0.28							
A	485	0.69		562	0.72		0.179					
rs829161 (ACACA)												
	348		0.539	387		0.015						
T/T	208	0.6		236	0.61		0.037					0.990
T/C	125	0.36		121	0.31							
C/C	15	0.04		30	0.08			Genotypic	2.274	1.133	4.727	
T	541	0.78		593	0.77		0.292					
C	155	0.22		181	0.23							
rs7599352 (MARCO)												
	327		0.034	353		0.438						
C/C	183	0.56		213	0.6		0.021					0.941
C/T	113	0.35		126	0.36							
T/T	31	0.09		14	0.04			Genotypic	0.416	0.197	0.842	
C	479	0.73		552	0.78		0.244					
T	175	0.27		154	0.22							
rs4687614 (NISCH)												
	355		1	391		0.733						
G/G	7	0.02		14	0.04		0.040					
A/G	87	0.25		114	0.29							
A/A	261	0.74		263	0.67							
G	101	0.14		142	0.18							
A	609	0.86		640	0.82		0.012	Additive	1.464	1.089	1.980	0.799
rs2600665 (AHSA2)												
	350		0.811	383		0.691						
T/T	156	0.45		207	0.54		0.019					
T/G	154	0.44		152	0.4							
G/G	40	0.11		24	0.06							
T	466	0.67		566	0.74		0.006	Additive	0.714	0.559	0.910	0.551
G	234	0.33		200	0.26							
rs12233670 (TLR10/1)												
	352		0.194	390		0.023						
T/T	105	0.3		136			0.022					
T/C	186	0.53		169				Genotypic	0.699	0.491	0.994	0.948
C/C	61	0.17		85								
T	396	0.56		441			0.815					

(*Continued*)

Table 3. (*Continued*)

rsID	Controls		HWE[c] *p*-val	TB Cases		HWE *p*-val	Association *p*-val Adjusted[d]	Model of penetrance	OR	2.5% CI	97.5% CI	Šidák *p*-val
	Count[a]	Prop[b]		Count	Prop							
C	308	0.44		339								

[a] Allelic and genotype counts

[b] Allelic and genotype proportions

[c] Statistic for the HWE exact test. This was stratified by TB susceptibility status.

[d] Statistic to indicate the association between genotype and TB susceptibility after adjusting for age, gender and ancestry. The allelic effect was modelled using the additive model.

https://doi.org/10.1371/journal.pone.0174738.t003

minor homozygote and heterozygote were associated with the phenotype. The minor homozygote has an odds ratio of 2.012 (95% CI = 1.108–3.716) whereas the heterozygote yielded an odds ratio of 2.146 (95% CI = 1.174–3.988).

Discussion

The genetic factors influencing TB susceptibility have been under investigation for many years, using various study designs, with limited success (S1 Table). This is due to a number of factors, including lack of power, the inability to identify causative variants and the complexity of admixed populations. We conducted a post-GWAS analysis of predicted functional variants and investigated their associations with TB susceptibility in the admixed SAC population. This resulted in the identification of 3 variants associated with an increased risk of progressing to active TB and another 3 variants associated with resistance to active TB.

MARCO (Macrophage Receptor With Collagenous Structure) is a member of the class A scavenger receptor family and has been implicated in innate antimicrobial activity, more specifically TB susceptibility, in the Gambian and Chinese Han populations [38, 39]. Statistically significant variants from Songane et al. (2012) and Bowdish et al. (2013) were used as query SNPs in the study presented here [39, 40]. The rs7599352 variant has not directly been implicated in TB susceptibility previously, but other intronic variants have been shown to have an effect on *MARCO* gene expression [38, 39]. This implies that intronic variants within this gene could have a greater functional effect than exonic variants.

Toll-like receptors (TLRs) are responsible for pathogen recognition and the resulting activation of an innate immune response. TLR genes, including *TLR2, TLR4* and *TLR8*, are known to contribute to TB susceptibility [41] and numerous variants within *TLR1* and *TLR10* have also been associated with TB susceptibility in previous studies [42, 43]. We provide evidence for a novel SNP (rs12233670) associated with a decrease in risk of TB susceptibility. This SNP was in LD with statistically significant SNPs reported in Ma et al. (2007) and Uciechowski et al. (2011) [42, 44].

Due to the long-range effects of LD with variants in Dual Specificity Phosphatase 14 (*DUSP14)* and *XPO1* (S1 Table), two novel genes were associated with increased TB susceptibility and increased resistance, namely, Acetyl-CoA Carboxylase Alpha (*ACACA)* and Activator Of Heat Shock 90kDa Protein ATPase Homolog 2 (*ASHA2*) respectively. An intronic variant (rs829161) in the *ACACA* gene was highly associated with the susceptibility phenotype with an odds ratio suggesting more than a 2.2 times increased chance of the disease progressing to an active state than remaining latent. The *ACACA* gene is involved in fatty acid carboxylation and in turn mediates the removal of cholesterol. Cholesterol is used as an energy source for *Mycobacterium tuberculosis (M.tb)* therefore any disruption of the level of the fatty acid may influence the dormancy of the bacteria [45]. A possible novel resistance pathway was

identified in this study, involving heat shock protein (HSP) activation and resulting ATPase activity by ASHA2. In this case the variant (rs260065) is located in the 5'-untranslated region and we therefore hypothesize that it has a profound regulatory function, potentially affecting gene expression. HSPs play a pivotal role in protein folding, stabilization and degradation, and are targets of chemotherapy in cancer patients as HSP modulates tumour cell apoptosis through protein kinase B, tumour necrosis factor receptor and NF-kB functioning. Due to the interaction with these known immune response genes, we hypothesize that ASHA2 could be involved in TB immune responses.

A variant (rs4687614) in the Nischarin (*NISCH)* gene was found to be associated with an increase in TB susceptibility. *NISCH* encodes the Nischarin protein. Nischarin has recently been identified as a dual effector that interacts with members of the Rac and Rab GTPase families [46]. The regulation of GTPases by Nischarin may regulate the maturation and acidification of vacuoles that are associated with phagocytosis of bacterial pathogens [46]. The odds ratio associated with the variant indicated an increased risk of progressing to active TB. An association was also found for a polymorphism in the Interferon-gamma Receptor 2 (*IFNGR2)* gene. An intronic variant (rs2284555) yielded a high odds ratio of 2.01 for the A/A genotype and 2.15 for the A/G genotype. *IFNGR2* has been previously implicated in various immunodeficiencies [47–50]. This study reiterates the role of IFNGR2 in antibacterial immune response and provides a novel causative SNP for TB.

Only 6 out of a total of 133 SNPs genotyped produced statistically significant associations with TB susceptibility. The predicted regulatory SNPs genotyped in our study were in strong LD with variants previously associated with TB in other populations. A comparison between the *p*-values of the original GWAS SNPs and the six variants identified here, shows that the *p*-values are lower in the original GWAS data. This could be an artefact due to the low sample size in our study, an increase in error variation or the moderate to low effect sizes associated with the six variants [51]. Susceptibility to TB is a complex disease and it is possible that numerous small/moderate effect variants at a frequency less than 0.05 will play a role in this phenotype [52]. It is also possible that some of the variants identified by previous studies are population-specific susceptibility variants and are therefore unlikely to be involved in disease predisposition in the SAC population. One might have expected many more associations due to the strong LD between SNPs genotyped in our study and variants previously identified as being associated with TB. The lack of associations suggest that these SNPs are not the causative SNPs that led to significant results in previous studies and that other SNPs in LD with the marker variants may play a role. Alternatively, the variants have smaller effect sizes than we could detect with our sample size. However, the power to detect a common variant with an odds ratio of 2 was 98%. Validation of these results in other case-control cohorts as well as the inclusion of recent GWAS results [5,12,14] is desirable, but complicated by the lack of available TB case-control cohorts with a similar genetic structure to that of the SAC population. In addition, since there was *a priori* evidence for an association, replication is arguably not necessary as this study attempted to fine-map the potential causal variants in loci identified by previous TB GWAS.

Age, KhoeSan ancestry and sex differed significantly between the TB cases and controls in this study, but were adjusted for in all analyses. We have previously shown that KhoeSan ancestry increases the risk of active TB [32] and old age is also a known TB risk factor [53]. Based on our experience in these communities, healthy males are less likely to attend clinics, while healthy females will accompany sick children and are therefore more likely to participate as controls. However, in most countries the TB notification rate is twice as high in males as in females [54] and evidence suggests that the X chromosome does play a role in TB susceptibility [55]. An investigation of sib-pair families from The Gambia indicated that chromosome Xq

might be involved in TB susceptibility [55]. Additionally, sex-specific TB associations for *TLR8*, an X-linked gene, has been identified in several populations [56–58], including the SAC population [59]. These loci could contribute to the observed male sex-bias in this population, but will require further investigation.

Subsequent to the analysis presented here, Chimusa et al. (2016) published a post-GWAS methodology utilizing LD information and the human protein–protein interaction network, which identified novel pathways associated with breast cancer [52]. The study by Chimusa et al. (2016) reiterates the need for alternative methodologies in the identification of regulatory variants associated with a phenotype.

In summary, the six predicted functional variants associated with TB susceptibility in the SAC population show that fine mapping of GWAS results can reveal candidate causal variants. Functional analyses are now required to elucidate the molecular mechanisms by which these polymorphisms may act, while well-powered TB GWAS and meta-analyses may continue to identify additional causal variants for TB susceptibility.

Supporting information

S1 Table. Variants associated with TB susceptibility. This table represents the variants that have been implicated in TB susceptibility by association studies. The closest gene to the variant is reported.
(ZIP)

S2 Table. The functional impact of proxy SNPs as determined by RegulomeDB. A score of 1 is the lowest score possible as determined by RegulomeDB. This indicates strong evidence of a functional impact. This has been determined by a variety of comprehensive datatypes.
(ZIP)

Acknowledgments

Author Contributions

Conceptualization: CU MM EH.

Data curation: CU.

Formal analysis: CU.

Funding acquisition: MM EH PvH.

Investigation: CU.

Methodology: CU MM.

Project administration: CU MM EH.

Resources: MM EH PvH MW AF.

Software: C.U.

Supervision: MM BH EH.

Validation: CU MM MW.

Visualization: CU.

Writing – original draft: CU.

Writing – review & editing: CU MM EH BH AF.

References

1. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005 Apr 15; 308(5720):385–9. https://doi.org/10.1126/science.1109557 PMID: 15761122

2. Wei Z, Liu Y, Xu H, Tang K, Wu H, Lu L, et al. Genome-Wide Association Studies of HIV-1 Host Control in Ethnically Diverse Chinese Populations. Sci Rep. 2015; 5:10879. https://doi.org/10.1038/srep10879 PMID: 26039976

3. Curtis J, Luo Y, Zenner HL, Cuchet-Lourenço D, Wu C, Lo K, et al. Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. Nat Genet. 2015 May; 47(5):523–7. https://doi.org/10.1038/ng.3248 PMID: 25774636

4. DeLorenze GN, Nelson CL, Scott WK, Allen AS, Ray GT, Tsai A- L, et al. Polymorphisms in HLA Class II Genes Are Associated With Susceptibility to Staphylococcus aureus Infection in a White Population. J Infect Dis. 2016 Mar 1; 213(5):816–23. https://doi.org/10.1093/infdis/jiv483 PMID: 26450422

5. Sobota RS, Stein CM, Kodaman N, Scheinfeldt LB, Maro I, Wieland-Alter W, et al. A Locus at 5q33.3 Confers Resistance to Tuberculosis in Highly Susceptible Individuals. Am J Hum Genet. 2016 Mar 3; 98(3):514–24. https://doi.org/10.1016/j.ajhg.2016.01.015 PMID: 26942285

6. Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat Genet. 2010 Sep; 42(9):739–41. https://doi.org/10.1038/ng.639 PMID: 20694014

7. Thye T, Owusu-Dabo E, Vannberg FO, van Crevel R, Curtis J, Sahiratmadja E, et al. Common variants at 11p13 are associated with susceptibility to tuberculosis. Nat Genet. 2012 Feb 5; 44(3):257–9. https://doi.org/10.1038/ng.1080 PMID: 22306650

8. Chimusa ER, Zaitlen N, Daya M, Möller M, van Helden PD, Mulder NJ, et al. Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. Hum Mol Genet. 2014 Feb 1; 23(3):796–809. https://doi.org/10.1093/hmg/ddt462 PMID: 24057671

9. Oki NO, Motsinger-Reif AA, Antas PR, Levy S, Holland SM, Sterling TR. Novel human genetic variants associated with extrapulmonary tuberculosis: a pilot genome wide association study. BMC Res Notes. 2011; 4:28. https://doi.org/10.1186/1756-0500-4-28 PMID: 21281516

10. Mahasirimongkol S, Yanai H, Mushiroda T, Promphittayarat W, Wattanapokayakit S, Phromjai J, et al. Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus for young tuberculosis. J Hum Genet. 2012 Jun; 57(6):363–7. https://doi.org/10.1038/jhg.2012.35 PMID: 22551897

11. Png E, Alisjahbana B, Sahiratmadja E, Marzuki S, Nelwan R, Balabanova Y, et al. A genome wide association study of pulmonary tuberculosis susceptibility in Indonesians. BMC Med Genet. 2012; 13:5. https://doi.org/10.1186/1471-2350-13-5 PMID: 22239941

12. Grant AV, Sabri A, Abid A, Abderrahmani Rhorfi I, Benkirane M, Souhi H, et al. A genome-wide association study of pulmonary tuberculosis in Morocco. Hum Genet. 2016 Mar; 135(3):299–307. https://doi.org/10.1007/s00439-016-1633-2 PMID: 26767831

13. Sobota RS, Stein CM, Kodaman N, Scheinfeldt LB, Maro I, Wieland-Alter W, et al. A Locus at 5q33.3 Confers Resistance to Tuberculosis in Highly Susceptible Individuals. Am J Hum Genet. 2016 Mar 3; 98(3):514–24. https://doi.org/10.1016/j.ajhg.2016.01.015 PMID: 26942285

14. Sveinbjornsson G, Gudbjartsson DF, Halldorsson BV, Kristinsson KG, Gottfredsson M, Barrett JC, et al. HLA class II sequence variants influence tuberculosis risk in populations of European ancestry. Nat Genet. 2016 Mar; 48(3):318–22. https://doi.org/10.1038/ng.3498 PMID: 26829749

15. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012 Sep; 22(9):1790–7. https://doi.org/10.1101/gr.137323.112 PMID: 22955989

16. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016; 17:122. https://doi.org/10.1186/s13059-016-0974-4 PMID: 27268795

17. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: Illuminating the Dark Road from Association to Function. Am J Hum Genet. 2013 Nov 7; 93(5):779–97. https://doi.org/10.1016/j.ajhg.2013.10.012 PMID: 24210251

18. Rosenthal SL, Barmada MM, Wang X, Demirci FY, Kamboh MI. Connecting the dots: potential of data integration to identify regulatory SNPs in late-onset Alzheimer's disease GWAS findings. PloS One. 2014; 9(4):e95152. https://doi.org/10.1371/journal.pone.0095152 PMID: 24743338

19. Cavalli M, Pan G, Nord H, Wadelius C. Looking beyond GWAS: allele-specific transcription factor binding drives the association of GALNT2 to HDL-C plasma levels. Lipids Health Dis. 2016; 15:18. https://doi.org/10.1186/s12944-016-0183-x PMID: 26817450

20. Bastami M, Nariman-Saleh-Fam Z, Saadatian Z, Nariman-Saleh-Fam L, Omrani MD, Ghaderian SMH, et al. The miRNA targetome of coronary artery disease is perturbed by functional polymorphisms identified and prioritized by in-depth bioinformatics analyses exploiting genome-wide association studies. Gene. 2016 Dec 5; 594(1):74–81. https://doi.org/10.1016/j.gene.2016.08.054 PMID: 27596011

21. Haider SA, Faisal M. Human aging in the post-GWAS era: further insights reveal potential regulatory variants. Biogerontology. 2015 Apr 17; 16(4):529–41. https://doi.org/10.1007/s10522-015-9575-y PMID: 25895066

22. Munch Z, Van Lill SWP, Booysen CN, Zietsman HL, Enarson DA, Beyers N. Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2003 Mar; 7(3):271–7.

23. de Wit E, Delport W, Rugamika CE, Meintjes A, Möller M, van Helden PD, et al. Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. Hum Genet. 2010 Aug; 128(2):145–53. https://doi.org/10.1007/s00439-010-0836-1 PMID: 20490549

24. Chimusa ER, Daya M, Möller M, Ramesar R, Henn BM, van Helden PD, et al. Determining ancestry proportions in complex admixture scenarios in South Africa using a novel proxy ancestry selection method. PLoS ONE. 2013; 8(9):e73971. https://doi.org/10.1371/journal.pone.0073971 PMID: 24066090

25. Daya M, van der Merwe L, Galal U, Möller M, Salie M, Chimusa ER, et al. A panel of ancestry informative markers for the complex five-way admixed South African coloured population. PloS One. 2013; 8 (12):e82224. https://doi.org/10.1371/journal.pone.0082224 PMID: 24376522

26. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009 Sep 1; 19(9):1655–64. https://doi.org/10.1101/gr.094052.109 PMID: 19648217

27. Gallant CJ, Cobat A, Simkin L, Black GF, Stanley K, Hughes J, et al. Impact of age and sex on mycobacterial immunity in an area of high tuberculosis incidence. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2010 Aug; 14(8):952–9.

28. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, Bakker PIW de. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008 Dec 15; 24 (24):2938–9. https://doi.org/10.1093/bioinformatics/btn564 PMID: 18974171

29. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006 Feb; 38(2):209–13. https://doi.org/10.1038/ng1706 PMID: 16415888

30. den Boon S., Van Lill SW, Borgdorff MW, Enarson DA, Verver S, Bateman ED, et al. High prevalence of tuberculosis in previously treated patients, Cape Town, South Africa. EmergInfectDis. 2007 Aug; 13 (8):1189–94.

31. Warnes G., Gorjanc G., Leisch F., Man M. genetics: Population Genetics R package. 2013 Mar

32. Daya M, van der Merwe L, van Helden PD, Möller M, Hoal EG. The role of ancestry in TB susceptibility of an admixed South African population. Tuberc Edinb Scotl. 2014 Jul; 94(4):413–20.

33. Perneger TV. What's wrong with Bonferroni adjustments. BMJ. 1998 Apr 18; 316(7139):1236–8. PMID: 9553006

34. Campbell H, Rudan I. Interpretation of genetic association studies in complex disease. PharmacogenomicsJ. 2002; 2(6):349–60.

35. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. AmJHumGenet. 2004 Apr; 74(4):765–9.

36. Šidák Z. Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. J Am Stat Assoc. 1967 Jun 1; 62(318):626–33.

37. Pollard K, Dudoit S, Laan M van der. Multiple Testing Procedures: R multtest Package and Applications to Genomics. UC Berkeley Division of Biostatistics Working Paper Series [Internet]. 2004 Dec 9; Available from: http://biostats.bepress.com/ucbbiostat/paper164

38. Ma M-J, Wang H-B, Li H, Yang J-H, Yan Y, Xie L-P, et al. Genetic variants in MARCO are associated with the susceptibility to pulmonary tuberculosis in Chinese Han population. PloS One. 2011; 6(8): e24069. https://doi.org/10.1371/journal.pone.0024069 PMID: 21886847

39. Bowdish DM, Sakamoto K, Lack NA, Hill PC, Sirugo G, Newport MJ, et al. Genetic variants of MARCO are associated with susceptibility to pulmonary tuberculosis in a Gambian population. BMC Med Genet. 2013; 14:47. https://doi.org/10.1186/1471-2350-14-47 PMID: 23617307

40. Schurz H, Daya M, Möller M, Hoal EG, Salie M. TLR1, 2, 4, 6 and 9 Variants Associated with Tuberculosis Susceptibility: A Systematic Review and Meta-Analysis. PLoS ONE. 2015; 10(10):e0139711. https://doi.org/10.1371/journal.pone.0139711 PMID: 26430737

41. Songane M, Kleinnijenhuis J, Alisjahbana B, Sahiratmadja E, Parwati I, Oosting M, et al. Polymorphisms in Autophagy Genes and Susceptibility to Tuberculosis. PLOS ONE. 2012 Aug 6; 7(8):e41618. https://doi.org/10.1371/journal.pone.0041618 PMID: 22879892

42. Ma X, Liu Y, Gowen BB. Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. PloS One. 2007; 2:1318.

43. Bulat-Kardum LJ, Etokebe GE, Lederer P, Balen S, Dembic Z. Genetic Polymorphisms in the Toll-like Receptor 10, Interleukin (IL)17A and IL17F Genes Differently Affect the Risk for Tuberculosis in Croatian Population. Scand J Immunol. 2015 Jul; 82(1):63–9. https://doi.org/10.1111/sji.12300 PMID: 25857634

44. Uciechowski P, Imhoff H, Lange C. Susceptibility to tuberculosis is associated with TLR1 polymorphisms resulting in a lack of TLR1 cell surface expression. J Leukoc Biol. 2011; 90:377–388. https://doi.org/10.1189/jlb.0409233 PMID: 21642391

45. Brzostek A, Pawelczyk J, Rumijowska-Galewicz A, Dziadek B, Dziadek J. Mycobacterium tuberculosis Is Able To Accumulate and Utilize Cholesterol. J Bacteriol. 2009 Nov 1; 191(21):6584–91. https://doi.org/10.1128/JB.00488-09 PMID: 19717592

46. Kuijl C, Pilli M, Alahari SK, Janssen H, Khoo P-S, Ervin KE, et al. Rac and Rab GTPases dual effector Nischarin regulates vesicle maturation to facilitate survival of intracellular bacteria. EMBO J. 2013 Mar 6; 32(5):713–27. https://doi.org/10.1038/emboj.2013.10 PMID: 23386062

47. Cooke GS, Campbell SJ, Sillah J, Gustafson P, Bah B, Sirugo G, et al. Polymorphism within the Interferon-γ/Receptor Complex Is Associated with Pulmonary Tuberculosis. Am J Respir Crit Care Med. 2006 Aug 1; 174(3):339–43. https://doi.org/10.1164/rccm.200601-088OC PMID: 16690980

48. Al-Muhsen S, Casanova J-L. The genetic heterogeneity of mendelian susceptibility to mycobacterial diseases. J Allergy Clin Immunol. 2008 Dec; 122(6):1043-1051-1053.

49. Hijikata M, Shojima J, Matsushita I, Tokunaga K, Ohashi J, Hang NTL, et al. Association of IFNGR2 gene polymorphisms with pulmonary tuberculosis among the Vietnamese. Hum Genet. 2012 May; 131 (5):675–82. https://doi.org/10.1007/s00439-011-1112-8 PMID: 22057826

50. Kong X- F, Vogt G, Itan Y, Macura-Biegun A, Szaflarska A, Kowalczyk D, et al. Haploinsufficiency at the human IFNGR2 locus contributes to mycobacterial disease. Hum Mol Genet. 2013 Feb 15; 22(4):769–81. https://doi.org/10.1093/hmg/dds484 PMID: 23161749

51. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol [Internet]. 2012 Dec [cited 2016 Sep 20]; 8(12). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531285/

52. Chimusa ER, Mbiyavanga M, Mazandu GK, Mulder NJ. ancGWAS: a post genome-wide association study method for interaction, pathway and ancestry analysis in homogeneous and admixed populations. Bioinforma Oxf Engl. 2016 Feb 15; 32(4):549–56.

53. Rieder HL, Disease IU against T and L. Epidemiologic basis of tuberculosis control. International Union Against Tuberculosis and Lung Disease; 1999. 180 p.

54. WHO | Global tuberculosis report 2016 [Internet]. WHO. [cited 2016 Dec 20]. Available from: http://www.who.int/tb/publications/global_report/en/

55. Bellamy R, Beyers N, McAdam KP, Ruwende C, Gie R, Samaai P, et al. Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. Proc Natl Acad Sci USA. 2000 Jul 5; 97(14):8005–9. https://doi.org/10.1073/pnas.140201897 PMID: 10859364

56. Davila S, Hibberd ML, Hari Dass R, Wong HEE, Sahiratmadja E, Bonnard C, et al. Genetic association and expression studies indicate a role of toll-like receptor 8 in pulmonary tuberculosis. PLoS Genet. 2008 Oct; 4(10):e1000218. https://doi.org/10.1371/journal.pgen.1000218 PMID: 18927625

57. Bukhari M, Aslam MA, Khan A, Iram Q, Akbar A, Naz AG, et al. TLR8 gene polymorphism and association in bacterial load in southern Punjab of Pakistan: an association study with pulmonary tuberculosis. Int J Immunogenet. 2015 Feb; 42(1):46–51. https://doi.org/10.1111/iji.12170 PMID: 25572425

58. Dalgic N, Tekin D, Kayaalti Z, Cakir E, Soylemezoglu T, Sancar M. Relationship between toll-like receptor 8 gene polymorphisms and pediatric pulmonary tuberculosis. Dis Markers. 2011; 31(1):33–8. https://doi.org/10.3233/DMA-2011-0800 PMID: 21846947

59. Salie M, van der Merwe L, Möller M, Daya M, van der Spuy GD, van Helden PD, et al. Associations between human leukocyte antigen class I variants and the Mycobacterium tuberculosis subtypes causing disease. J Infect Dis. 2014 Jan 15; 209(2):216–23. https://doi.org/10.1093/infdis/jit443 PMID: 23945374