Published in final edited form as:
Nat Genet. 2016 March ; 48(3): 318–322. doi:10.1038/ng.3498.

HLA class II sequence variants influence tuberculosis risk in populations of European ancestry

Gardar Sveinbjornsson[1,2], **Daniel F. Gudbjartsson**[1,2], **Bjarni V. Halldorsson**[1,3], **Karl G. Kristinsson**[4,5], **Magnus Gottfredsson**[4,6], **Jeffrey C. Barrett**[7], **Larus J. Gudmundsson**[1], **Kai Blondal**[8], **Arnaldur Gylfason**[1], **Sigurjon Axel Gudjonsson**[1], **Hafdis T. Helgadottir**[1], **Adalbjorg Jonasdottir**[1], **Aslaug Jonasdottir**[1], **Ari Karason**[1], **Ljiljana Bulat Kardum**[9], **Jelena Knežević**[10,11], **Helgi Kristjansson**[1,4], **Mar Kristjansson**[6], **Arthur Love**[4,12], **Yang Luo**[7], **Olafur T. Magnusson**[1], **Patrick Sulem**[1], **Augustine Kong**[1], **Gisli Masson**[1], **Unnur Thorsteinsdottir**[1,4], **Zlatko Dembic**[10], **Sergey Nejentsev**[13], **Thorsteinn Blondal**[8], **Ingileif Jonsdottir**[1,4,14], and **Kari Stefansson**[1,4]

[1]deCODE genetics / Amgen Inc., Sturlugata 8, Reykjavik, Iceland [2]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland [3]School of Science and Engineering, Reykjavik University, Reykjavík, Iceland [4]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland [5]Department of Clinical Microbiology, Landspitali, the National University Hospital of Iceland, Reykjavik, Iceland [6]Department of Infectious Diseases, Landspitali, the National University Hospital of Iceland, Reykjavik, Iceland [7]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK [8]Division of Communicable Disease Prevention and Control, Primary Health Care of the Capital Area, Reykjavik, Iceland [9]Department of Pulmology, Clinic of Internal Medicine, Clinical Hospital Center, University of Rijeka, Rijeka, Croatia [10]Laboratory of Molecular Genetics, Department of Oral Biology, Faculty of Dentistry, University of Oslo, Oslo, Norway [11]Division of Molecular Medicine, Ruđer Bošković Institute, Zagreb, Croatia [12]Department of Virology, Landspitali, the National University Hospital of Iceland, Reykjavik, Iceland [13]Department of Medicine, University of Cambridge, Cambridge,

Corresponding authors: Kari Stefansson, deCODE genetics / Amgen Inc., Sturlugata 8, 101 Reykjavík, Iceland. kstefans@decode.is, Phone:354-5701900, fax 354-5701901. Ingileif Jonsdottir, deCODE genetics / Amgen Inc., Sturlugata 8, 101 Reykjavik, Iceland. ingileif.jonsdottir@decode.is, Phone: 354-5701941, fax 354-5701901.

Author contributions

The study was designed and results were interpreted by G.S., D.F.G., B.V.H., A. Kong, U.T., T.B., I.J. and K.S.. Phenotype data ascertainment and Icelandic subject recruitment were coordinated and managed by T.B, K.G.K., M.G., L.J.G., A.L., M.K. and K.B.. S.N., J.C.B. and Y.L. coordinated, managed, genotyped and analyzed the Russian cohort sample set. L.B.K., J.K. and Z.D. coordinated and managed the Croatian cohort phenotypes and samples, which were genotyped and analysed by deCODE. G.S., H.T.H., G.M., S.A.G., O.T.M., U.T., and I.J. performed the sequencing, genotyping and expression analyses. G.S., D.F.G., B.V.H., S.A.G., A.G., A.J., A.J., A.K., H.K., and I.J. performed HLA typing and analysis of HLA data. G.S., D.F.G., B.V.H., A.G., S.A.G., P.S., A. Kong, G.M. and I.J. performed the statistical and bioinformatics analyses. The manuscript was drafted by G.S., D.F.G., S.N., I.J. and K.S. All authors contributed to the final version of the manuscript.

URLs. Allele frequencies, www.allelefrequencies.net; Broad's HLA reference, www.broadinstitute.org/gatk/media/docs/HLA_REFERENCE.zip; GATK/hla caller database, www.broadinstitute.org/gatk.

Competing financial interests

Authors affiliated with deCODE genetics/AMGEN Inc. declare competing financial interests as employees.

Europe PMC Funders Author Manuscripts

United Kingdom [14]Department of Immunology, Landspitali, the National University Hospital of Iceland, Reykjavik, Iceland

Abstract

Mycobacterium tuberculosis (*M. tuberculosis*) infections cause 9.0 million new tuberculosis (TB) cases and 1.5 million deaths annually1. To search for sequence variants that confer risk of TB we tested 28.3 million variants identified through whole-genome sequencing of 2,636 Icelanders for association with TB (8,162 cases and 277,643 controls), pulmonary TB (PTB), and *M. tuberculosis* infection. We found association of three sequence variants in the HLA class II region: rs557011[T] (MAF=40.2%) with *M. tuberculosis* infection (OR =1.14, $P=3.1\times10^{-13}$) and PTB (OR=1.25, $P=5.8\times10^{-12}$) and rs9271378[G] (MAF=32.5%) with PTB (OR=0.78, $P=2.5\times10^{-12}$), both located between *HLA-DQA1* and *HLA-DRB1*. Finally, a missense variant p.Ala210Thr in *HLA-DQA1*, (MAF=19.1%, rs9272785) shows association with *M. tuberculosis* infection ($P=9.3\times10^{-9}$, OR=1.14). The association of these variants with PTB was replicated in large samples of European ancestry from Russia and Croatia ($P< 5.9\times10^{-4}$). These findings demonstrate that the HLA class II region contributes to the complex genetic risk of tuberculosis, possibly through reduced presentation of protective *M. tuberculosis* antigens to T cells.

Mycobacterium tuberculosis (*M. tuberculosis*) causes 9.0 million new tuberculosis (TB) cases and 1.5 million deaths annually, and *M. tuberculosis* carriers who are not infected with HIV have a 10% lifetime risk of developing active TB disease1. The majority of infected individuals control the pathogen by mounting a successful, long-lived immune response, leading to clinically latent infection. Immunological impairment caused by malnutrition, diabetes, HIV/AIDS, aging and smoking, plays a major role in the epidemiology of TB and heritability studies have implicated genetic susceptibility2,3. It is believed that TB was rare in Iceland until the 19th century when it spread rapidly and reached its peak in the beginning of the nineteen thirties. In the year 1935 approximately 20.9% of 8 year olds and 34.2% of 13 years olds had positive tuberculin skin test (TST^+), indicating that a substantial part of the population had been exposed to *M. tuberculosis*4,5. Since then the incidence of TB in Iceland has decreased to the lowest in Europe6.

We imputed 28.3 million single nucleotide polymorphisms (SNPs) and insertions/deletions identified through whole-genome sequencing of 2,636 Icelanders into 104,220 chip-typed Icelanders and their first and second degree relatives7. The Icelandic Tuberculosis Data Registry contains TB diagnosis from 1900 to 2010, including confirmed TB diagnoses of 8,162 individuals with genotype information, whereof 3,686 had a confirmed pulmonary TB (PTB) (Table 1). 14,724 individuals with genotype information had contracted *M. tuberculosis* infection, since in addition to the TB cases, 6,562 individuals who did not develop TB, were recorded TST^+ and had not been Bacillus Calmette-Guerin (BCG) vaccinated (Supplementary Table 1). Since TST positivity may also be caused by vaccination with BCG, we excluded those who had been BCG vaccinated and considered the remaining TST^+ subjects as *M. tuberculosis*-infected. We tested the imputed sequence variants for association with PTB, TB, and *M. tuberculosis* infection (TB and/or TST^+)

using population controls (N>277,643). Sequence variants were weighed according to their prior probability of affecting gene function by applying thresholds for genome-wide significance that depend on the variant class. We allocated the type I error rate of 0.05 equally between three classes of variants. We tested 5,955 loss of function variants, 157,106 missense variants and 28,113,695 other variants yielding class specific Holm-Bonferroni genome-wide significance thresholds of 2.8×10^{-6} for loss of function, 1.1×10^{-7} for missense and 5.9×10^{-10} for other variants.

Using these criteria, several sequence variants in the human leukocyte antigen (HLA) region on chromosome 6p21 showed genome-wide significant association (Table 2, Supplementary Tables 2 and 3, Figure 1, Supplementary Figure 1). The strongest association was between rs557011[T] (MAF=40.2%), located between *HLA-DQA1* and *HLA-DRB1*, and *M. tuberculosis* infection ($P=3.1\times10^{-13}$, OR=1.14) and PTB ($P=5.8\times10^{-12}$, OR=1.25). rs9271378[G] (MAF=32.5%), also located between *HLA-DQA1* and *HLA-DRB1* (Figure 1), associates with reduced risk of PTB ($P=2.5\times10^{-12}$, OR=0.78). Finally, a missense variant p.Ala210Thr, (MAF=19.1%, rs9272785) in exon 4 of *HLA-DQA1* associates with *M. tuberculosis* infection ($P=9.3\times10^{-9}$, OR=1.14). p.Ala210Thr corresponds to the classical HLA-*DQA1*03* superallele. Three missense variantsp.Thr49Ser (rs1048023), p.Gly79Arg (rs12722072 and rs12722074) and p.Met99Val (rs1064944), that along with p.Ala210Thr define *HLA-DQA1*03,* did not align to the reference sequence. These variants were called by separately aligning each exon to one of the haplotypes found in the GATK/hla caller database8 and imputing for HLA typing (see methods). The variants correlate perfectly ($r^2>=0.99$) with rs9272785/p.Ala210Thr and show identical association (Table 2). The HLA-*DQA1*03* superallele (p.Ala210Thr/p.Thr49Ser/p.Gly79Arg/p.Met99Val) and the two non-coding variants are moderately correlated (pairwise r^2 between 0.11 and 0.35, Supplementary Table 4). The association of each of the two non-coding variants remains significant when conditioned on the other, while p.Ala210Thr remained significant when conditioned on rs9271378 but not rs557011 (Supplementary Table 5). No sequence variant associates significantly with TB disease as a whole, but rs557011 ($P=3.7\times10^{-9}$, OR=1.15) and rs9272785 ($P=2.3\times10^{-9}$, OR=0.86) still represent the top signals with consistent direction of effect.

We followed the three sequence variants identified in Iceland up in PTB samples from Russia (5,530 cases and 5,607 controls) and Croatia (438 cases and 1,009 controls). HLA-*DQA1*03* (represented by p.Ala210Thr) and the two non-coding variants associate with PTB in the Russian sample ($P<5.4\times10^{-4}$) and rs557011 associates with PTB in the smaller Croatian sample (P=0.0074). All three variants show a consistent direction of effects (Table 3). The ORs for rs557011 and rs9272785 are lower in the Russian sample than in Iceland (P<0.011).

To better understand the effect of the variants we compared PTB and TB cases to *M. tuberculosis*-infected individuals without TB, as well as non-pulmonary TB cases and *M. tuberculosis*-infected individuals to population controls (Table 4). HLA-*DQA1*03* and rs557011 both associate with *M. tuberculosis* infection in the absence of TB disease ($P=1.1\times10^{-7}$ and 2.5×10^{-6}, respectively), but not with TB disease among the *M. tuberculosis*-infected (P>0.13). rs557011 associates with PTB (P=0.0035) among the *M.*

tuberculosis-infected while HLA-*DQA1*03* does not (P=0.063). This pattern is inversed for rs9271378 which showed no association for *M. tuberculosis* infection in the absence of tuberculosis (P=0.20), but a significant association with TB (P=0.0035) and PTB ($P=3.7\times10^{-6}$) among *M. tuberculosis*-infected. For all of the variants, the association with non-pulmonary TB was weaker than for PTB, possibly because greater phenotypic heterogeneity among non-pulmonary TB cases. Microbiologically confirmed PTB cases and TB cases were additionally tested for association. All three variants associate more strongly with microbiologically confirmed cases (Supplementary Table 6). We use TST-positivity as a surrogate for *M. tuberculosis* infection, recognizing that it is not a perfect surrogate, since environmental mycobacteria may also cause TST positivity. It was reported that TST reactivity is controlled by the TST1 locus at 11p14 and the TST2 locus at 5p15 in a hyperendemic area in South-Africa 9 and TST negativity in a low endemic area in France was recently shown to be controlled by a chromosomal region at 11p14-15 close to TST1 and overlapping with the TNF1 locus10. We therefore tested the effects of the TB-associated variants reported here with TST negativity (Supplementary Table 1) and found no association (P>0.05).

We tested imputed HLA alleles of six of the classical HLA genes; *HLA-A, HLA-B, HLA-C, HLA-DQA1, HLA-DQB1* and *HLA-DRB1* for association with PTB, TB and *M. tuberculosis* infection (Supplementary Table 7). Of these, the *HLA-DQA1*03:01* allele showed the strongest association. However, the associations of rs9271378 and rs557011 cannot be fully explained by the available HLA alleles (Supplementary Table 8). This suggests either that the association is with an HLA allele that was not well imputed or that the association is not driven by the classical HLA alleles, with the possible exception of *HLA-DQA1*03:01*. Given that the strongest PTB association is with rs9271378, the possible involvement of other genes in that region cannot be excluded. The TB associated variants correlate with increased or decreased expression of *HLA-DQA1*, *HLA-DQB1* and/or *HLA-DRB1* in white blood cells, although other variants in the region are more strongly correlated with the level of expression of these genes, but had weaker association with TB in our study (Supplementary Table 9)11.

SNPs in the *ASAP1* gene on chromosome 8q24 have been associated with PTB in the Russian population 12. However, in Icelanders these SNPs do not associate with PTB (rs10956514[G] P=0.66, OR=0.99, Supplementary Table 10). Previously, a genome-wide association scan in African samples from Ghana and The Gambia yielded common variants associating with PTB: rs4331426[G] (OR=1.19, MAF=44%) at 18q11 and rs2057178[G] located downstream of the *WT1* gene on chromosome 11p13 (OR=0.77, MAF=32%) 13,14. We tested these variants for association in the Icelandic data. Both variants are rarer in Iceland than in the African populations and have smaller effects (Supplementary Table 10). These results resonate with findings of previous TB GWAS in the South African15 and Russian12 populations that neither showed TB association with rs4331426 on chromosome 18q11, but detected association of rs2057178 on chromosome 11p13. No SNPs in the HLA region were reported to have a genome-wide significant association with TB in the African samples, the most suggestive associations in the region were found with rs9469220 in *HLA-DQA1* (P=0.002, OR=1.15) and rs9272346 (P=0.022, OR=1.10). In the Icelandic data, rs9469220[A] (MAF=38.8%) shows association with PTB ($P=1.1\times10^{-5}$, OR=0.86,) and

correlates weakly with our top PTB signal rs9271378 (r^2=0.23). On the other hand, rs9272346[A] does not associate with TB in our data (P>0.21). Lack of replication of previously reported variants maybe due to factors such as phenotypic heterogeneity, BCG vaccination rates, exposure to environmental mycobacteria and the predominant circulating *M.tuberculosis* strains, that may cause effects of variants to vary between populations, affecting statistical power to detect their association, different LD to causal variants, in addition to the possibility of a spurious initial finding. We show genome-wide significant association of rs557011[T] and rs9271378[G] located between *HLA-DQA1* and *HLA-DRB1,* and rs9272785/p.Ala210Thr in *HLA-DQA1* with TB phenotypes and replication in follow up studies. However, it should be kept in mind when attempting to transfer results to other populations that effect sizes are likely to vary based on the case composition.

Our study suggests that allele *DQA1*03* may contribute to TB. *DQA1*03* associates with several autoimmune diseases. *DQA1*03* is a well known risk factor for the gluten sensitive enteropathy coeliac disease (CD) as part of the *DQA1*03-DQB1*03:02* haplotype, and *trans-DQA1*03:01/DQB1*02:01*, encoding the DQ8 and DQ2.3 molecules, respectively16. DQ8 and DQ2.3 bind gliadin peptide T cells epitopes, providing a functional evidence for the contribution of *DQA1*03* to the pathogenesis of the disease17,18. *DQA1*03* also increases the susceptibility to Type 1 Diabetes (T1D)19 and other autoimmune diseases 20,21. Many *M. tuberculosis* -derived epitopes are recognized by HLA-restricted CD4+ and CD8+ T cells in humans infected with *M. tuberculosis*22. In latently TB-infected individuals the CD4+ T cells recognizing *M. tuberculosis* epitopes are confined to the CXCR3+CCR6+ Th1 subset23 and high reactivity is associated with recognition of a few discrete dominant antigenic regions24. The ability of *DQA1*03:01*-containing HLA molecules to present dominant epitopes of critical protective *M. tuberculosis* antigens is unknown. Cell-surface expression of HLA-DQ alleles varies extensively, indicating an allelic hierarchy in the intrinsic stability of HLA-DQ molecules25, and DQ8, containing *DQA1*03*, is among the most unstable DQ molecules. Alterations in disease-associated amino acids located outside of the peptide-binding groove regulate the stability of DQ molecules25. The Ala210Thr mutation is in exon four, encoding the transmembrane region of the DQA1 chain, whereas the p.Thr49Ser, p.Gly79Arg and p.Met99Val missense mutations are in exon 2, in the vicinity of α−β chain dimerization interfaces at each end of the peptide binding groove26,27, and may affect peptide binding. Furthermore, the DQ8 molecule containing *DQA1*03* interacts poorly with HLA-DM (DM), that plays a critical role in peptide loading during antigen presentation, resulting in reduced loading and DM-editing of antigenic peptides26,28. Surface expression HLA class II peptide complexes on antigen presenting cells is also regulated by ubiquitination; their assembly, endocytosis, recycling and turnover29. The non-coding variants rs557011 and rs9271378 do not overlap with known biologically relevant regions (Supplementary Notes). However, a correlated marker rs1846190 (r^2=0.81 with rs557011) is located in an enhancer site in CD4+ CD25- IL17+ T cells30 and CTCF binding site which seems to regulate *HLA-DRB1* expression in lymphoblastoid cell lines31. It is conceivable that the contribution of *DQA1*03* and the missense mutations to the risk of *M. tuberculosis* infection and TB disease is through reduced stability of *DQA1*03* containing molecules, and poor presentation of critical *M. tuberculosis* antigens, resulting in poor activation of protective T cells. The effects of the

Europe PMC Funders Author Manuscripts

TB-associated variants on mRNA of *DQA1*03* and its surface expression in antigen presenting cells have not been studied.

The HLA region plays a key role in immune responses and associates with infections32 and autoimmune33,34 diseases and has been the focus of many tuberculosis candidate gene studies yielding conflicting results35. We found sequence variants, located in *HLA-DQA1* and between *HLA-DQA1* and *HLA-DRB1* that associate with TB. rs557011 associates with PTB and seems to confer susceptibility to both *M. tuberculosis* infection and risk of development of TB disease. On the contrary, rs9271378 does not associate with *M. tuberculosis* infection per se, but protects against development of TB disease in *M. tuberculosis*-infected individuals. For all variants the effects on PTB were the strongest.

Materials and Methods

The Icelandic discovery population

Individuals who had been diagnosed with TB or been infected with *M. tuberculosis* (TST positive) without developing TB during the 20th century, according to the Icelandic Tuberculosis Database (ITBDB) were invited to participate in the study. The ITBDB contains information on TB diagnosis, major and minor sites of infection based on *M. tuberculosis* culture and microscopic analysis, histology and roentgen result. It has also records on number and duration of TB episodes and hospitalizations, symptoms and signs, treatment and outcome and family history, as well as nationwide results of TST testing and BCG (see Supplementary Notes). In this study we used genotype data for 3,686 patients with confirmed pulmonary TB, 8,162 with any TB and 14,723 with *M. tuberculosis* infection with or without developing TB, using TST positivity as a surrogate for having been *M. tuberculosis* infected (excluding BCG vaccinated individuals) and ≥ 277,643 controls (remaining chip-typed Icelanders and their relatives).

The study was approved by the Icelandic Data Protection Authority (ref. 2004120649) and the National Bioethics Committee (ref. VSN 04-172 VSNb2004120008-03-1). All participating subjects who donated blood signed informed consent. Personal identities of the participants and biological samples were encrypted by a third party system approved and monitored by the Icelandic Data Protection Authority.

The Russian TB sample set

The study in the Russian cohort has been done as described previously12. Briefly, TB patients have been diagnosed using information about TB contact, medical history and clinical symptoms, presence of acid fast bacilli in sputum smear and symptoms characteristic of pulmonary TB on chest X rays. For all patients diagnosis has been confirmed by culture of *M. tuberculosis* from sputum. Patients with extra-pulmonary TB and all HIV-positive subjects were excluded. Controls were healthy adult blood bank donors with no history of TB. *M. tuberculosis* infection status of these controls was unknown.

The Croatian TB sample set

DNA was isolated from blood samples of PTB patients (N=244) and contact controls (n=85) treated at the Section of Pulmology, Department of Internal Medicine, Clinical Hospital Centre Rijeka, Rijeka and Hospital for Lung Diseases “Jordanovac”, University Hospital Center “Zagreb”, and healthy blood donors (n=924) collected at Departments of Transfusion Medicine in Rijeka and in Zagreb, Croatia, as previously described36,37. DNA samples from the blood of additional 194 PTB patients, treated in Hospital for Lung Disease “Jordanovac”, University Hospital Center “Zagreb”, Croatia, were isolated at deCODE, Reykjavik, Iceland. The age and gender frequency of the latter group of patients was not significantly different from the former cohort of cases. All study subjects provided oral and written informed consent. The Medical Research Council ethics committees at Zagreb and Rijeka approved the research.

Illumina chip genotyping

Icelandic chip-typed samples were assayed using the Illumina HumanHap300, HumanCNV370, HumanHap610, HumanHap1M, HumanHap660, Omni-1, Omni 2.5 or Omni Express bead chips at deCODE genetics. SNPs were excluded if they had (i) yield less than 95%, (ii) minor allele frequency less than 1% in the population or (iii) significant deviation from Hardy-Weinberg equilibrium in the controls ($P < 0.001$), (iv) if they produced an excessive inheritance error rate (over 0.001), or (v) if there was substantial difference in allele frequency between chip types (from just a single chip if that resolved all differences, but from all chips otherwise). All samples with a call rate below 97% were excluded from the analysis. For the HumanHap series of chips, 304,937 SNPs were used for long range phasing, whereas for the Omni series of chips 564,196 SNPs were included. The final set of SNPs used for long-range phasing was composed of 707,525 SNPs.

Genotyping of the Russian PTB cases and controls has been done using Affymetrix Genome-Wide Human SNP Array 6.0 and SNPs across the genome were imputed as described previously12. Association analysis of the HLA SNPs rs557011, rs9271378 and rs9272785 has been done using PLINK38.

Whole genome sequencing

Whole genome sequencing was performed for 2,636 Icelanders, selected for various conditions. All of the individuals were sequenced at a depth of at least 10X (average sequencing depth = 22X).

Template DNA fragments were hybridized to the surface of flow cells (GA PE cluster kit (v2) or HiSeq PE cluster kits (v2.5 or v3)) and amplified to form clusters using the Illumina cBot. In brief, DNA (2.512 pM) was denatured, followed by hybridization to grafted adaptors on the flow cell. Isothermal bridge amplification using Phusion polymerase was then followed by linearization of the bridged DNA, denaturation, blocking of 3’ ends and hybridization of the sequencing primer. Sequencing-by-synthesis (SBS) was performed on Illumina GAIIx and/or HiSeq 2000 instruments. Paired-end libraries were sequenced at 2 x 101 (HiSeq) or 2 x 120 (GAIIx) cycles of incorporation and imaging using the appropriate TruSeq™ SBS kits. Each library or sample was initially run on a single GAIIx lane for QC

validation followed by further sequencing on either GAIIx (³ 4 lanes) or HiSeq (³ 1 lane) with targeted raw cluster densities of 500800 k/mm2, depending on the version of the data imaging and analysis packages (SCS2.6-2-9/RTA1.6-1.9, HCS1.3.8-1.4.8/ RTA1.10.36-1.12.4.2). Real-time analysis involved conversion of image data to base-calling in real-time.

Generation of whole-genome genotype data

SNP and INDEL calling from the whole-genome sequence data of the 2,230 Icelanders and generation of imputed genotypes has been described previously39 (Supplementary Notes). Briefly, the genotypes identified were imputed into chip genotyped and long range phased Icelanders. Probabilities of genotypes were furthermore predicted for relatives of chip-typed individuals.

HLA typing

For each of the six MHC genes, the most common alleles present in the Icelandic population were selected from the allelefrequencies database, using the most ethnically related populations: Norway and Ireland. The exonic sequence of the alleles were downloaded from Broad's HLA reference.

We genotyped "in silico" a set of 2615 whole genome sequenced (WGS) individuals. For each sequenced individual we selected reads that either: 1) mapped to the public reference sequence using BWA or 2) were unmapped by BWA and could be aligned to one of the haplotypes found in the GATK/hla caller database.

For every gene genotyped we align each read in this set to each exon separately. We consider a read to belong to a given haplotype if: 1) The read can be aligned to the exon from GATK with no mismatches or indels, allowing for the possibility that the read only partially overlaps the exon as long as the overlap is at least 40 basepairs and the overlap does not introduce a mismatch or an indel. 2) The mate of this read can be aligned to some sequences occurring within +/- 1000 basepairs from the exon using BWA default parameters.

If a read r aligns to an exon of allele A we arbitrarily say that P(r|A) = 1-PE and if r does not align to A we set P(r|A) = PE. We arbitrarily choose PE = 0.001 and have P(r|A) = 0.9999 for reads r that align to the superallele A and P(r|A) = 0.001 for reads r that do not align to the superallele A. If we let R be the set of reads and assume independence of reads and the two alleles carried by an individual we can then compute

$$P(R \mid A1, A2) = \prod_{r \in R} \left(\frac{1}{2}P(r|A1) + \frac{1}{2}P(r \mid A2)\right).$$ The HLA alleles were imputed into the Icelandic sample set as previously described for genotypes39.

To check the accuracy of the imputation at least three individuals carrying each haplotype were HLA typed for the 6 genes using All-Set TM Gold SSP (Life Technologies, DQA1, DQB1, DRB1, HLA-A high resolution typing; HLA-B and HLA-C low resolution typing). Accuracy between imputation and wet-lab genotyping was 90%-99% and frequency weighted correlation was 95-99.6%. (Supplementary Table 12, Accuracy was 0.989 for DQA1)

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Genotyping of single variants

Single SNP genotyping of rs9271378 and rs9272785 in the Croatian sample set was carried out by deCODE Genetics in Reykjavik, Iceland, applying the Centaurus (Nanogen) platform40 . Sanger sequencing of rs557011 in the Croatian sample set was performed by deCODE Genetics.

Association testing

Logistic regression was used to test for association between sequence variants and disease (tuberculosis), treating disease status as the response and genotype counts as covariates. Other available individual characteristics that correlate with disease status were also included in the model as nuisance variables. These characteristics were: Sex, county of birth, current age or age at death (first and second order terms included), blood sample availability for the individual and an indicator function for the overlap of the lifetime of the individual with the timespan of phenotype collection (Supplementary Notes). Conditional analysis was performed by including the sequence variant being conditioned on as a covariate in the model under the null and the alternative in the generalized linear regression.

Principal component analysis

We calculated principal components using the EIGENSTRAT software41 (Supplementary Figure 2). In the Russian population, the first four principal components were included in the association analysis (Supplementary Table 11). The first 20 principal components only explain 0.6% of the variance of the Icelandic genotype data and the first two components correlate strongly with the county of birth, already included in the association test42. Including the first five principal components in the association testing of the chip typed Icelandic samples had a minimal effect (Supplementary Table 3) and principal components were not included in the final Icelandic association testing which also included relatives of the chip typed individuals.

Genotype imputation information

The informativeness of genotype imputation was estimated by the ratio of the variance of imputed expected allele counts and the variance of the actual allele counts:

$$\frac{Var(E(\theta|chip\,data))}{Var(\theta)},$$

where θ is the allele count. $Var(E(\theta|chip\ data))$ was estimated by the observed variance of the imputed expected counts and $Var(\theta)$ was estimated by $p(1-p)$, where p is the allele frequency. Sequence variants with imputation information below 0.8 were excluded from the analysis.

Gene and variant annotation

For the annotation of the exome data coordinates of variants were converted between hg18 and hg19 using the liftOver tool from UCSC43. Variants in hg19 coordinates were annotated

with information from Ensembl release 70 using Variant Effect Predictor (VEP) version 2.8 44. Only protein coding transcripts from RefSeq Release 5645 were considered.

Correction for relatedness of the Icelandic subjects and genomic control

Individuals in both the Icelandic case and control groups are related, causing the χ^2 test statistic to have a mean > 1 and median > 0.675. We estimated the inflation factor λ_g based on a subset of about 300,000 common variants and the *P* values adjusted by dividing the corresponding χ^2 values by this factor to adjust for both relatedness and potential population stratification46.

Thresholds for genome-wide significance

We weighed sequence variants according to their prior probability of affecting gene function by applying thresholds for genome-wide significance that depend on the variant class. The Bonferroni correction for multiple testing can be adjusted to account for prior importance of sequence variants. We performed a weighted Holm-Bonferroni correction based on giving equal weight to the classes of LoF, MSSNS, and other variants 47. For example sequence variants in the LoF class get weight $\frac{1}{3 \cdot 6,476}$. The sum of these weights over all the variants in the genome is 1 and the Bonferroni threshold for significance within a class containing m sequence variants will be $\frac{0.05}{3m}$.

Expression analysis

Samples of RNA from human peripheral blood were hybridized to Agilent Technologies Human 25K microarrays as described previously11. We quantified expression changes between two samples as the mean logarithm ($\log_{10}$) expression ratio (MLR) compared with a reference pool RNA sample. In comparing expression levels between groups of individuals with different genotypes, we denoted the expression level for each genotype as $10^{(\text{average MLR})}$, where the MLR is averaged over individuals with the particular genotype. We determined s.e.m. and significance by regressing the MLR values against the number of risk alleles carried. We took into account the effects of age, sex and differential cell type count in blood as explanatory variables in the regression. *P*-values were adjusted for familial relatedness of the individuals by simulation.

Assessment for potential overlap with regulatory regions

To identify TB associated variants that might have regulatory effects. We took the strongest non-coding signals (rs557011 and rs9271378) and identified all SNPs in LD at r2 > 0.8 (excluding SNPs with low imputation information values). This added rs1846190 and rs508318 (both r2=0.81 with rs557011). For each of the 4 variants identified, we searched for overlaps with known regulatory regions as follows: 1) First we used ENSEMBL to determine whether the variant had been assigned a regulatory region ENSR number. Then we examined the ENCODE data and looked for any evidence of ChIP-Seq transcription factor binding and DNaseI hypersensitivity sites48. 2) We also looked for enhancer and promoter chromatin segmentation states using the 25 state HMM from the Roadmap consortium30. 3) Then we looked for correlations between DNaseI hypersensitive sites and

local gene expression using results described by Sheffield et al31. 4) We examined SiPhy and GERP conservation scores49,50 and 5) we viewed the Factorbook and HaploReg_v3 data51,52 to search for potential changes in transcription factor binding motifs with ChIP-Seq evidence of the cognate transcription factor.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The study in Iceland was supported by the National Institute of Health/National Institute of Allergy and Infectious Diseases grant HHSN266200400064C. The authors thank all the participants in the study. We also thank the staff at the Patient Recruitment Center and the deCODE genetics core facilities.

The study of the Russian TB patients and controls was supported by the Wellcome Trust grants 088838/Z/09/Z and 095198/Z/10/Z, the EU Framework Programme 7 Collaborative grant 201483, the European Research Council Starting grant 260477, and the Royal Society grants UF0763346 and RG090638. S.N. is a Wellcome Trust Senior Research Fellow in Basic Biomedical Science and is also supported by the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre.

We thank Prof.dr.med. Sanja Balen (Department of Transfusion Medicine, Medical Faculty, University of Rijeka, Rijeka, Croatia) and Dr. Melita Balija (Department for Transfusion Medicine, Petrova 3, Zagreb, Croatia) who assisted in the collection of blood samples. We thank Assoc.prof.dr.med. Sanja Grle-Popovic for received assistance in collection of blood samples of tuberculosis patients treated at the Hospital for Lung Diseases “Jordanovac”, University Hospital Center “Zagreb”, Zagreb, Croatia. We thank Prof.dr.med. Jasminka Pavelic (Laboratory of Molecular Oncology, Division of Molecular Medicine, Ruđer Bošković Institute, Zagreb, Croatia) for providing resorces and helpful advice.

References

1. World Health Organization. WHO Global Tuberculosis Report 2014. World Health Organisation; 2014.
2. Fox GJ, Menzies D. Epidemiology of Tuberculosis Immunology. New Paradigm of Immunity to Tuberculosis. 2013; 783:1–32.
3. Comstock GW. Tuberculosis in twins: a re-analysis of the Prophit survey. Am Rev Respir Dis. 1978; 117:621–4. [PubMed: 565607]
4. Sigurdsson S. [Tuberculosis in Iceland. 1976]. Laeknabladid. 2005; 91:69–102. [PubMed: 16155306]
5. Sigurdsson S. Um berklaveiki á Íslandi. Læknablaðið. 1976; 62:3–50.
6. Bothamley GH, Ditiu L, Migliori GB, Lange C. Active case finding of tuberculosis in Europe: a Tuberculosis Network European Trials Group (TBNET) survey. Eur Respir J. 2008; 32:1023–30. [PubMed: 18550615]
7. Gudbjartsson DF, et al. Large-scale whole-genome sequencing of the Icelandic population. Nat Genet. 2015 **advance online publication**.
8. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–8. [PubMed: 21478889]
9. Cobat A, et al. Two loci control tuberculin skin test reactivity in an area hyperendemic for tuberculosis. J Exp Med. 2009; 206:2583–91. [PubMed: 19901083]
10. Cobat A, et al. Tuberculin Skin Test Negativity Is Under Tight Genetic Control of Chromosomal Region 11p14-15 in Settings With Different Tuberculosis Endemicities. Journal of Infectious Diseases. 2015; 211:317–321. [PubMed: 25143445]
11. Emilsson V, et al. Genetics of gene expression and its effect on disease. Nature. 2008; 452:423–8. [PubMed: 18344981]

12. Curtis J, et al. Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. Nat Genet. 2015 **advance online publication**.
13. Thye T, et al. Common variants at 11p13 are associated with susceptibility to tuberculosis. Nat Genet. 2012; 44:257–9. [PubMed: 22306650]
14. Thye T, et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat Genet. 2010; 42:739–41. [PubMed: 20694014]
15. Chimusa ER, et al. Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. Hum Mol Genet. 2014; 23:796–809. [PubMed: 24057671]
16. Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol. 2002; 2:647–55. [PubMed: 12209133]
17. Sollid LM, Qiao SW, Anderson RP, Gianfrani C, Koning F. Nomenclature and listing of celiac disease relevant gluten T-cell epitopes restricted by HLA-DQ molecules. Immunogenetics. 2012; 64:455–60. [PubMed: 22322673]
18. Lundin KEA, Scott H, Fausa O, Thorsby E, Sollid LM. T-Cells from the Small-Intestinal Mucosa of a Dr4,Dq7 Dr4,Dq8 Celiac-Disease Patient Preferentially Recognize Gliadin When Presented by Dq8. Hum Immunol. 1994; 41:285–291. [PubMed: 7883596]
19. Aly TA, et al. Extreme genetic risk for type 1A diabetes. Proc Natl Acad Sci U S A. 2006; 103:14074–9. [PubMed: 16966600]
20. Zanelli E, Breedveld FC, de Vries RR. HLA class II association with rheumatoid arthritis: facts and interpretations. Hum Immunol. 2000; 61:1254–61. [PubMed: 11163080]
21. Rider LG. The heterogeneity of juvenile myositis. Autoimmun Rev. 2007; 6:241–7. [PubMed: 17317616]
22. Lindestam Arlehamn CS, Lewinsohn D, Sette A. Antigens for CD4 and CD8 T cells in tuberculosis. Cold Spring Harb Perspect Med. 2014; 4:a018465. [PubMed: 24852051]
23. Lindestam Arlehamn CS, Sette A. Definition of CD4 Immunosignatures Associated with MTB. Front Immunol. 2014; 5:124. [PubMed: 24715893]
24. Arlehamn CS, et al. Dissecting mechanisms of immunodominance to the common tuberculosis antigens ESAT-6, CFP10, Rv2031c (hspX), Rv2654c (TB7.7), and Rv1038c (EsxJ). J Immunol. 2012; 188:5020–31. [PubMed: 22504645]
25. Miyadera H, Ohashi J, Lernmark A, Kitamura T, Tokunaga K. Cell-surface MHC density profiling reveals instability of autoimmunity-associated HLA. J Clin Invest. 2014
26. Busch R, et al. On the perils of poor editing: regulation of peptide loading by HLA-DQ and H2-A molecules associated with celiac disease and type 1 diabetes. Expert Rev Mol Med. 2012; 14:e15. [PubMed: 22805744]
27. Tollefsen S, et al. Structural and functional studies of trans-encoded HLA-DQ2.3 (DQA1*03:01/DQB1*02:01) protein molecule. J Biol Chem. 2012; 287:13611–9. [PubMed: 22362761]
28. Yin L, Maben ZJ, Becerra A, Stern LJ. Evaluating the Role of HLA-DM in MHC Class II-Peptide Association Reactions. J Immunol. 2015; 195:706–16. [PubMed: 26062997]
29. Cho KJ, Walseng E, Ishido S, Roche PA. Ubiquitination by March-I prevents MHC class II recycling and promotes MHC class II turnover in antigen-presenting cells. Proc Natl Acad Sci U S A. 2015
30. Chadwick LH. The NIH Roadmap Epigenomics Program data resource. Epigenomics. 2012; 4:317–24. [PubMed: 22690667]
31. Sheffield NC, et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. Genome Research. 2013; 23:777–88. [PubMed: 23482648]
32. Chapman SJ, Hill AV. Human genetic susceptibility to infectious disease. Nat Rev Genet. 2012; 13:175–88. [PubMed: 22310894]
33. Fernando MM, et al. Defining the role of the MHC in autoimmunity: a review and pooled analysis. PLoS Genet. 2008; 4:e1000024. [PubMed: 18437207]
34. de Bakker PI, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet. 2006; 38:1166–72. [PubMed: 16998491]

35. Meyer CG, Thye T. Host genetic studies in adult pulmonary tuberculosis. Semin Immunol. 2014; 26:445–453. [PubMed: 25307123]
36. Etokebe GE, et al. Toll-like receptor 2 (P631H) mutant impairs membrane internalization and is a dominant negative allele. Scand J Immunol. 2010; 71:369–81. [PubMed: 20500688]
37. Knezevic J, et al. Heterozygous carriage of a dysfunctional Toll-like receptor 9 allele affects CpG oligonucleotide responses in B cells. J Biol Chem. 2012; 287:24544–53. [PubMed: 22613717]
38. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. [PubMed: 17701901]
39. Styrkarsdottir U, et al. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. Nature. 2013; 497:517–20. [PubMed: 23644456]
40. Kutyavin IV, et al. A novel endonuclease IV post-PCR genotyping system. Nucleic Acids Research. 2006; 34:e128. [PubMed: 17012270]
41. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9. [PubMed: 16862161]
42. Price AL, et al. The impact of divergence time on the nature of population structure: an example from Iceland. PLoS Genet. 2009; 5:e1000505. [PubMed: 19503599]
43. Kent WJ, et al. The human genome browser at UCSC. Genome Res. 2002; 12:996–1006. [PubMed: 12045153]
44. McLaren W, et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics. 2010; 26:2069–70. [PubMed: 20562413]
45. Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. 2012; 40:D130–5. [PubMed: 22121212]
46. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999; 55:997–1004. [PubMed: 11315092]
47. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. 1979; 6:65–70.
48. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
49. Davydov EV, et al. Identifying a high fraction of the human genome to be under selective constraint using GERP++ PLoS Comput Biol. 2010; 6:e1001025. [PubMed: 21152010]
50. Garber M, et al. Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics. 2009; 25:i54–62. [PubMed: 19478016]
51. Wang J, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Research. 2012; 22:1798–812. [PubMed: 22955990]
52. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Research. 2012; 40:D930–4. [PubMed: 22064851]

Europe PMC Funders Author Manuscripts

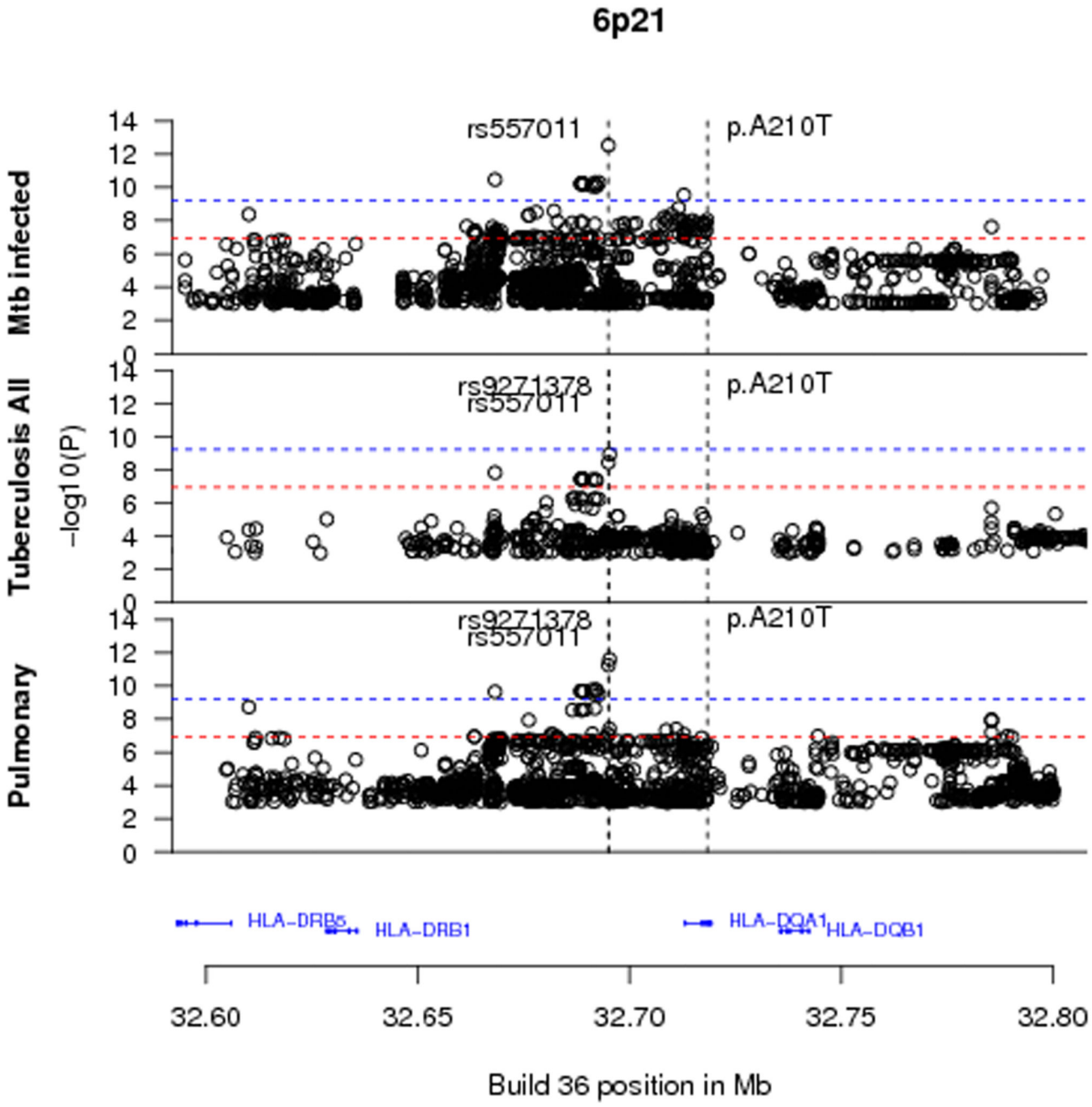

Figure 1.

Regional association plot of the HLA 6p21 loci for pulmonary tuberculosis (N=3,686), all tuberculosis (N=8,162) and *M. tuberculosis* infected w/wo TB (N=14,723). The -$\log_{10}$ P values (*y* axis) of each SNP are presented on the basis of their chromosomal positions (*x* axis). Only P values below 1×10^{-3} are plotted.The horizontal dotted lines represend significance thresholds for missense and non-coding variants.

Europe PMC Funders Author Manuscripts

Table 1

Characteristics of the Icelandic TB cases.

For both chip typed individuals and family imputed individuals the average year of birth and standard deviation is given as well as percent of individuals alive and percent of males.

	Pulmonary TB N=3,686		**Tuberculosis N=8,162**		***M. tuberculosis* infection (including TB) N=14,723**	
	Chip typed	**1st,2nd degree relatives**	**Chip typed**	**1st,2nd degree relatives**	**Chip typed**	**1st,2nd degree relatives**
N cases	1,188	2,498	2,765	5,397	6,105	8,618
Mean year of birth (SD)	1931(12)	1914(16)	1933(13)	1914(17)	1931(12)	1919(17)
Percent alive	53.0%	8.4%	57.4%	10.0%	54.7%	16.2%
Males %	40.7%	48.0%	41.1%	47.5%	42.3%	48.9%

Europe PMC Funders Author Manuscripts

Table 2

Sequence variants showing genome-wide significant association with Tuberculosis in Iceland.

For each sequence variant the reference SNP ID number (rs#), chromosome (Chr), hg18 position, its effect on a gene (Coding effect), minor allele frequency (MAF), gene name and modeled allele are provided in addition to the odds ratio (OR) for tuberculosis and the corresponding P values.

								Pulmonary TB			**Tuberculosis**			***M. tuberculosis* infection (including TB)**		
								N=3,686			**N=8,162**			**N=14,724**		
rs#	**chr**	**pos**	**MAF (%)**	**Info**	**coding**	**gene**	**Effect allele**	**P**	**OR**	**95% CI**	**P**	**OR**	**95% CI**	**P**	**OR**	**95% CI**
rs557011	chr6	32694991	40.2	0.997	-	-	T	**5.8×10^{-12}**	1.25	[1.17-1.33]	3.7×10^{-9}	1.15	[1.10-1.20]	**3.1×10^{-13}**	1.14	[1.10-1.18]
rs9271378	chr6	32695278	32.5	0.998	-	-	G	**2.5×10^{-12}**	0.78	[0.73-0.84]	1.2×10^{-9}	0.86	[0.82-0.90]	1.3×10^{-7}	0.90	[0.87-0.94]
rs9272785	chr6	32718379	19.1	0.996	p.Ala210Thr*	HLA-DQA1	A	3.5×10^{-7}	1.22	[1.13-1.32]	7.3×10^{-4}	1.10	[1.04-1.16]	**9.3×10^{-9}**	1.14	[1.09-1.19]

* p.Thr49Ser (rs1048023) , p.Gly79Ser (rs12722072) and p.Met99Val (rs1064944) correlate (r^2=0.99) with rs9272785/p.Ala210Thr and show identical association.

Europe PMC Funders Author Manuscripts

Table 3

Associations of the sequence variants from follow up in samples from Russia and Croatia.

Results from combined analysis of Iceland, Russia and Croatia are also provided. For each sequence variant the reference SNP ID number (rs#) and modeled allele are given in addition to the allele frequency in each cohort, the odds ratio (OR) for pulmonary tuberculosis, the corresponding P values and 95% confidence intervals. Significant heterogeneity was observed between the non-Icelandic and Icelandic samples for rs557011 (P=0.016) and rs9271378 (P=0.0024) but not for *DQA1*03* (P=0.21). The heterogeneity is driven by the difference between the Icelandic and the Russian samples.

			rs557011[T]				rs9271378[G]				p.Ala210Thr			
	#Cases	#Controls	P	OR	95% CI	AF%	P	OR	95% CI	AF%	P	OR	95% CI	AF%
PTB Iceland	**3,686**	**287,427**	5.8×10^{-12}	1.25	[1.17-1.33]	40.2	2.5×10^{-12}	0.78	[0.73-0.84]	32.5	3.5×10^{-7}	1.22	[1.13-1.32]	19.1
PTB Russia	**5,530**	**5,607**	8.5×10^{-5}	1.12	[1.06-1.19]	34.4	1.9×10^{-5}	0.89	[0.84-0.94]	46.6	5.4×10^{-4}	1.15	[1.06-1.24]	13.9
PTB Croatia	**438**	**1,009**	0.0074	1.26	[1.06-1.49]	36.1	0.052 *	0.85 *	[0.72-1.00]	57.5 *	0.66	1.06	[0.82-1.37]	10.9
PTB Russia+Croatia combined	**5,968**	**6,616**	4.7×10^{-6}	1.13	[1.07-1.20]	-	3.0×10^{-6}	0.89	[0.84-0.93]	-	5.9×10^{-4}	1.14	[1.06-1.23]	-
PTB all combined	**9,654**	**294,043**	2.0×10^{-15}	1.18	[1.13-1.23]	-	3.2×10^{-15}	0.85	[0.82-0.89]	-	1.9×10^{-9}	1.18	[1.12-1.25]	-

* Pvalue, OR and AF are given for rs113013369 which correlates with rs9271378 ($r^2 = 0.8$)

Europe PMC Funders Author Manuscripts

Table 4

Association of the sequence variants with TB sub-phenotypes.

For each sequence variant the reference SNP ID number (rs#), chromosome (Chr), hg18 position, its effect on a gene (Coding effect), minor allele frequency (MAF), gene name and modeled allele are provided in addition to the odds ratio (OR) for TB and the corresponding P values.

							PTB (N = 3,686) vs *M. tuberculosis*-infected wo TB (TST$^+$, N = 6,562)			**TB (N=8,162) vs *M. tuberculosis*-infected wo TB (TST$^+$, N = 6,562)**			**Non-pulmonary TB (N = 4,476)**			***M. tuberculosis*-infected wo TB (TST+, N = 6,562)**		
rs#	**chr**	**pos**	**MAF (%)**	**coding**	**gene**	**Effect allele**	**P**	**OR**	**95% CI**	**P**	**OR**	**95% CI**	**P**	**OR**	**95% CI**	**P**	**OR**	**95% CI**
rs5570011	chr6	32694991	40.2	-	-	T	0.0035	1.12	[1.04-1.21]	0.6	1.02	[0.95-1.10]	0.05	1.06	[1.00-1.12]	2.5×10^{-6}	1.11	[1.06-1.16]
rs9271378	chr6	32695278	32.5	-	-	G	3.7×10^{-6}	0.83	[0.77-0.90]	0.0035	0.91	[0.85-0.97]	0.028	0.94	[0.89-0.99]	0.2	0.97	[0.93-1.02]
rs9272285	chr6	32718379	19.1	p.Ala210Thr	HLA-DQA1	A	0.061	1.10	[1.00-1.22]	0.13	0.94	[0.87-1.02]	0.91	1.00	[1.00-1.00]	1.1×10^{-7}	1.16	[1.10-1.23]