Infection, Genetics and Evolution 71 (2019) 76–81

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

# An exome wide association study of pulmonary tuberculosis patients and their asymptomatic household contacts

Chandrika Bhattacharyya, Partha Pratim Majumder, Bhaswati Pandit*

*National Institute of Biomedical Genomics, PO: NSS, Kalyani 741251, West Bengal, India*

ARTICLE INFO

*Keywords:*
Tuberculosis
DNA polymorphisms
*SIGLEC15*
*HLA-DRA*
Haplotype

ABSTRACT

Tuberculosis is a leading cause of death in India. To identify genetic variants associated with susceptibility or resistance to *Mycobacterium tuberculosis* infection, we have performed an exome-wide association study with 0.2 million exonic variants among 119 pairs of tuberculosis patients and their clinically asymptomatic household contacts. The strongest association was identified for rs61104666[A], a synonymous variant (p.E292E) of exon 5 of the gene *SIGLEC15* (OR = 2.4, $p = 1.49 \times 10^{-5}$). We also found association of non-coding variants in the 3′UTR region of a gene encoding the class II human leukocyte antigens (HLAs), *HLA-DRA*. rs13209234[A] (minor allele frequency (MAF) = 13.8%) (OR = 0.35, $P = 2.5 \times 10^{-4}$) and rs3177928[A] (minor allele frequency (MAF) = 13.7%) (OR = 0.35, $P = 3.3 \times 10^{-4}$) were associated with protection from tuberculosis. These two SNPs, rs13209234 and rs3177928, are in complete linkage disequilibrium. These associations remained valid when additional data on freshly recruited individuals were jointly analyzed on 250 patient-control pairs. The identified gene, *HLA-DRA*, suggest involvement of immune regulation, indicating pathways associated with antigen presentation in tuberculosis infection.

## 1. Introduction

Tuberculosis (TB) is an aerosol transmitted infectious disease, caused by *Mycobacterium tuberculosis* (*M. tb.*). Tuberculosis remains one of the leading causes of death, with an annual incidence of 10 million and about 1.4 million deaths, despite significant advances made in diagnosis and treatment (WHO, 2018). In 2018, World Health Organization (WHO) reported that, 2.8 million new cases were diagnosed from India, with 0.4 million deaths.

Variability in the clinical response to *M. tb.* among infected individuals, resident in a restricted geographical area, often sharing the same household, is a strong indication of involvement of host genetics. Some individuals may not develop detectable infection (Ernst, 2012). Most individuals, resident in a high burden country such as India, are infected but do not develop clinical disease, while a small fraction of infected individuals develop clinical symptoms of disease.

Susceptibility to tuberculosis is a complex trait influenced by both genetic and environmental factors (Meyer and Thye, 2014). Several genome-wide linkage scans have been performed in human tuberculosis. Although genomic regions of putative linkage have been found, none of the linkage peaks identified in these studies attained a level of genome-wide significance (Baghdadi et al., 2006; Bellamy, 2000; Bellamy et al., 2000; Cooke et al., 2008; Ridruechai et al., 2010; Stein et al., 2008). Many candidate gene association studies have been performed in different global populations (Azad et al., 2012; Chen et al., 2013; Li et al., 2011; Moller and Hoal, 2010; Wu et al., 2014). One of the most convincing findings is the association of polymorphisms in the *SLC11A1* (*NRAMP1)* gene, (Li et al., 2011) replicated in multiple populations. Besides *SLC11A1*, *VDR*, HLA class II genes, *TLR*s and cytokine genes such as *IFNG*, *TNF*, and *IL10* are the most well studied candidate genes. Genome-wide association studies have also identified SNPs in multiple genes associated with tuberculosis, across African (Chimusa et al., 2014; Grant et al., 2016; Thye et al., 2010, 2012), European (Curtis et al., 2015; Sveinbjornsson et al., 2016) and Asian populations including Thais, Japanese and Indonesians (Hong et al., 2017; Mahasirimongkol et al., 2012; Png et al., 2012). Among African populations, SNPs located on chromosomes 18q11, 11p13 and 5q33 were discovered to be associated with tuberculosis susceptibility. Recently, two genome wide association studies, conducted in Icelandic (Sveinbjornsson et al., 2016) and Han Chinese (Qi et al., 2017) populations, revealed association with HLA class II region genes *HLA-DQA1* and *HLA-DQB1*. Association studies have indicated the possibility of ethnic differences contributing to differences in genetic susceptibility or protection to tuberculosis. No genome wide studies have been conducted among the Indian population, in spite of a high prevalence of tuberculosis in India. Candidate gene association studies conducted in

* Corresponding author.
*E-mail addresses:* cb1@nibmg.ac.in (C. Bhattacharyya), ppm1@nibmg.ac.in (P.P. Majumder), bp1@nibmg.ac.in (B. Pandit).

https://doi.org/10.1016/j.meegid.2019.03.006
Received 1 August 2018; Received in revised form 27 February 2019; Accepted 14 March 2019
Available online 18 March 2019
1567-1348/ 

India have revealed that variants in genes HLA class I and II (Mishra et al., 2014; Saikia et al., 2015), *TLR1* (Dittrich et al., 2015), *VDR* (Sharma et al., 2011), and cytokine genes (Abhimany, 2012; Joshi et al., 2015; Mishra et al., 2012; Sharma et al., 2010) are significantly associated with tuberculosis. Majumder et al. (2009) have reported a large number of novel variants from Indian populations groups in genes of the immune system (Bairagya et al., 2008; Majumder et al., 2009). They have also shown that haplotype frequency profiles of these genes in India are significantly different from those of European and African populations, possibly indicating local adaptation to pathogens (Mukherjee et al., 2009).

Here, we attempted to identify genes associated with susceptibility to tuberculosis infection in an Indian population using an exome wide search. We have hypothesized that an individual, living in close proximity (in same household) with a TB patient for an extended period of time, but not exhibiting clinical symptoms, may be genetically less susceptible. In order to identify those genetic factors, we have adopted a case-household contact association study where cases were active TB patients and controls were their biologically unrelated spouses without any clinical symptom of TB in spite of, being continuously exposed to *M. tb*. from their infected spouse.

## 2. Methods

### 2.1. Ethical statement

The study was approved by the Institutional Ethics Committees of the National Institute of Biomedical Genomics, India. Biological specimens were collected with written informed consent from each study participant.

### 2.2. Study design

In this case-control association study, cases were clinically symptomatic tuberculosis patients. Controls were clinically disease free, biologically unrelated spouses with (a) no history of past TB and (b) stayed under the same roof continuously for at least a six-month period of relevance and remained disease free. Tuberculosis patients were enrolled from a State General Hospital and DOTS (directly observed treatment, short-course) Centre under the Revised National Tuberculosis Control Program (RNTCP) in West Bengal, India. All recruited tuberculosis patients were positive for acid-fast smear (AFB) test of sputum, treatment naive and with no past history of tuberculosis. All the household contacts were tested AFB sputum culture at two time points; within an interval of three months. Each spouse who shared the physical dwelling unit with the patient was followed up for an additional three months to ascertain they had remained disease free, before being recruited as a control. They were also followed up for an additional period of two years. Household contacts were removed from further analysis, if they became infected during the course of this study. All study participants were HIV negative. We did not classify recruited household contacts based on IGRA test, as in a high-endemic setting such as India, IGRA is known to be an unreliable diagnostic test. All cases (Discovery Set; $n = 119$) and controls (Discovery Set; n = 119) were drawn from a homogeneous Bengali population of a single geographical location (Kalyani-Gayeshpur region) of West Bengal, India. For validation, we have recruited an additional 138 pairs of TB patients and household contacts, using the same inclusion and exclusion criteria.

### 2.3. Exome wide genotyping

Trained phlebotomists collected blood samples in BD Vacutainer® EDTA tubes, by venipuncture. Genomic DNA was isolated from study participants using QIAamp DNA Blood Midi Kit (Qiagen) using manufacturer's protocol. DNA concentration was determined using Qubit™ DNA Assay Kits (Invitrogen, USA) according to the manufacturer's instructions.

Genotyping was done using the Infinium HumanExome v1.2 beadchip (Illumina, cat # WG-353-1201) following manufacturer's protocol. Genotype base calling was done using Illumina Genome Studio following Genotyping Module, version 1.0. Only those loci at which, valid genotype base calls could be made for at least 90% of the study participants, were included for further analysis.

### 2.4. TaqMan SNP genotyping

In validation phase, we have genotyped study participants using TaqMan SNP genotyping assay (Assay ID = C_25747020_10 and C__2455631_10) and 7900HT system (Applied Biosystems). We visually checked all genotype clusters, assigned calls and extracted genotypes using TaqMan Genotyper software 1.4. Genotypes of four individuals were undetermined and hence, they are excluded from further analysis. We have also repeated 10% of samples, which were genotyped using Infinium HumanExome v1.2 beadchip, in TaqMan assay. All the samples were concordant in both.

### 2.5. Association analysis

Stringent quality control checks were used to exclude poor-quality data (i.e., missing genotype < 1% for both SNPs and individuals, minor allele frequency (MAF) > 1%, and SNPs not in Hardy-Weinberg equilibrium). Chi-square test for allelic and haplotype association were done using PLINK version 1.7 (Purcell et al., 2007). Manhattan plot and Haplotype blocks based on the exonic markers were created in Haploview 4.2 (Barrett et al., 2005). Regional association plot was prepared in LocusZoom v0.4.8 (Pruim et al., 2010).

### 2.6. In-silico analysis

Gene annotation was done using Variowatch version 3.0 and SNPnexus (Abu et al., 2018; Cheng et al., 2012). SNPs in transcription factor binding region, splicing regulator, or predicted miRNA binding sites were visualized using the HaploReg v4.1 (Ward and Kellis, 2016) and RegulomeDB version 1.1 (Boyle et al., 2012). The scoring system of RegulomeDB follows a score of 1a to 5, from highly significant regulatory variant to minimally important variant, based on available data types for a single coordinate (http://www.regulomedb.org/help#score). eQTLs were identified in Genotype-Tissue Expression (GTEx) database (2015).

Data curation, statistical analysis, and graphical representations were done using R, version 3.4.3 (R: Team, 2013).

## 3. Results

We generated genotype data on 2,44,700 markers which covered most of the known human genes. More than 80% ($N = 2,06,676$) of the genotyped loci turned out to be monomorphic (minor allele frequency (MAF) < 1%) in our study population, possibly a reflection of the European population-centric choice of markers placed on the DNA micro-array by the manufacturer. We noted that, 9285 variants had missing genotype data in > 1% of individuals, and 54 loci significantly ($p < .001$) deviated from Hardy–Weinberg equilibrium; these loci were removed. Seven (three TB patients and four controls) of the 238 study participants were removed because for each of them genotype information was missing at > 1% of loci. Three pairs of individuals showed cryptic relatedness (genetically assessed as having ~90% of markers identical-by-state); the member of each such pair with the higher genotyping call rate was retained and the other member was excluded.

Our final discovery data set, therefore, comprised data on 114 pairs of tuberculosis patients (Male = 92; Female = 22) and control spouses (Male = 20; Female = 94) with genotype calls at 28,755 SNP loci for

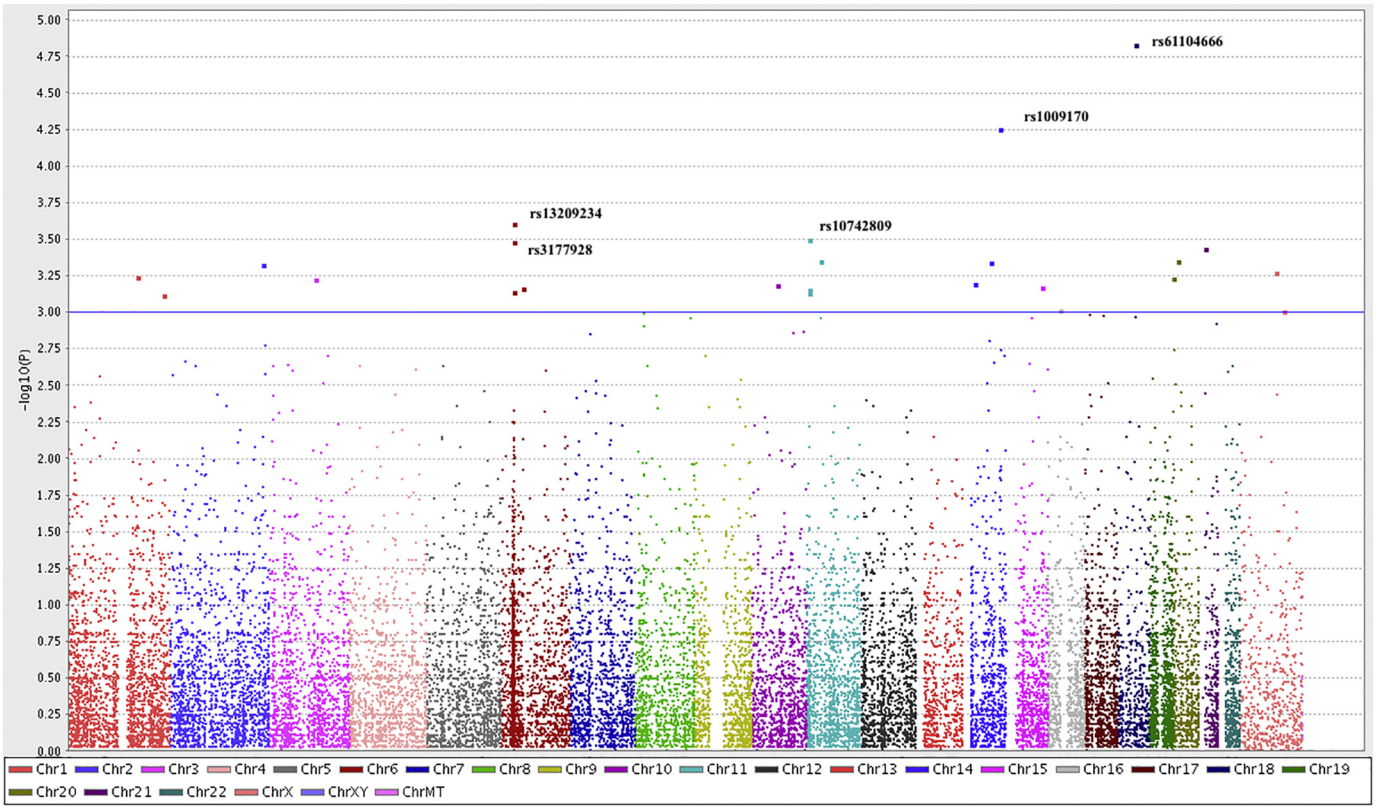


**Fig. 1. Manhattan plot depicting association results with 28,755 exonic markers.** Five most significantly associated SNPs are highlighted in the figure. rs13209234 and rs3177928 are on *HLA-DRA*.

each individual. Mean age (in years) of TB patients and household contacts were 45.58 (range 20–85) and 39.68 (range 19–70) respectively.

The Manhattan plot showing for the 28,755 exonic variants is provided in Fig. 1. None of the identified variants exceeds the genome wide significance level of $p$ value i.e. $-\log_{10}p \geq 6$, after multiple testing correction. We have used the threshold of $-\log_{10}p \geq 3$ for significance; this threshold corresponds to the value at which points start to deviate from the diagonal of the Q–Q plot (expected and observed $p$-values, without any multiple testing correction) (Fig. 1). Among the variants, which were unadjusted for multiple testing correction, the strongest association was identified for rs61104666, a synonymous variant (p.E292E) of exon 5 of the gene *SIGLEC15* (OR = 2.4 (C·I 99% 1.4–4.0), $p = 1.49 \times 10^{-5}$, Table 1). No other variant from the gene *SIGLEC15* was found to be associated. Next, rs1009170, an intergenic variant, ~7 kb downstream to *CPSF2* gene was found to be associated. Among the other most associated markers, two variants: rs13209234 (OR = 0.3 (C·I. 99% 0.2–0.7), $p = 2.5 \times 10^{-4}$) and rs3177928 (OR = 0.3 (C·I. 99% 0.2–0.7), $p = 3.3 \times 10^{-4}$) from the same gene *HLA-DRA* were associated with protection from tuberculosis infection (Table 1, Fig. 2). Another SNP, rs2239804, on 3′UTR of *HLA-DRA*, also showed association in the same direction (OR = 0.6 (C·I 99% 0.4–0.98), $p$ = .007) (Fig. 2).

Pairwise linkage disequilibrium (LD) measures ($r^2$) of this ~4.5 kb region (Chr6:32411523–32415975) of *HLA-DRA* 3′UTR are diagrammatically presented in Fig. 3. Two linkage disequilibrium (LD) blocks were estimated from this region, based on our genotype data. One larger LD block (3.54 kb) comprised of four SNPs; rs3177928, rs2227139, rs3129891 and rs13209234. Four haplotypes were phased from our study population, in this LD block. One of four haplotypes, “AAGA”, was found to be significantly more frequent in household contacts, than TB patients (Chi square test, $p$ = .00019 (Table 2). The other LD block (0.32 kb), of this region, consisted of three SNPs, rs2239804, rs7192, rs2239802. One of the four phased haplotypes, of this LD block, “ACG” was also observed at a significantly higher frequency in household contacts than TB patients (Chi square test, $p$ = .00020) (Table 2).

The $p$-values corrected for multiple testing were higher than 0.05, for the SNPs reported in earlier paragraphs. Possibly, in an exome-wide context, the sample size ($n$ = 114 “case-control” pairs) used by us for association-discovery is often considered to be small. We replenished our data set by an additional 136 tuberculosis patients (Male = 106; Female = 30) and their control spouses, using the same inclusion and exclusion criteria. In the additional recruits, the mean ages (in years) of

**Table 1**
Five most significant genetic variants, associated with tuberculosis in the study population.

| SNP | Chr | Position | Allele | MAF | Gene | Type of mutation | $p$ | OR | 99% CI |
|---|---|---|---|---|---|---|---|---|---|
| rs61104666 | 18 | 43,420,177 | A | 34.4 | *SIGLEC15* | p.E292E | $1.49 \times 10^{-5}$ | 2.39 | 1.42–4.02 |
| rap009170 | 14 | 92,636,713 | A | 39.6 | Intergenic | NA | $5.54 \times 10^{-5}$ | 0.46 | 0.28–0.76 |
| **rs13209234** | 6 | 32,415,975 | A | 13.6 | *HLA-DRA* | 3′UTR | $2.48 \times 10^{-4}$ | 0.35 | 0.16–0.75 |
| rs10742809 | 11 | 5,862,532 | C | 32.5 | *OR52E6* | p.Met199Arg | $3.17 \times 10^{-4}$ | 2.08 | 1.23–3.52 |
| **rs3177928** | 6 | 32,412,435 | A | 13.5 | *HLA-DRA* | 3′UTR | $3.3 \times 10^{-4}$ | 0.35 | 0.16–0.76 |

Significant SNPs from *HLA-DRA* gene are shown in bold. MAF: Minor Allele Frequency, OR: Odds Ratio, CI: Confidence Interval.

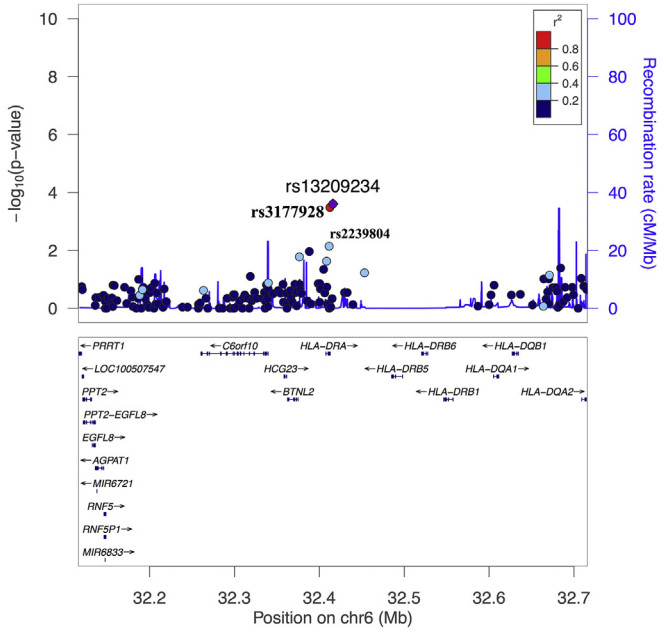


**Fig. 2. Regional association plot of *HLA-DRA* gene.** Among the five most associated markers, two variants: rs13209234 (OR = 0.3 (C·I 99% 0.2–0.7), $p = 2.5 \times 10^{-4}$) and rs3177928 (OR = 0.3 (C.I 99% 0.2–0.7), $p = 3.3 \times 10^{-4}$) are on the same gene *HLA-DRA*. Another SNP rs2239804, from 3'UTR of *HLA-DRA*, also showed association (OR = 0.6 (C.I 99% 0.4–0.98), $p = .007$).

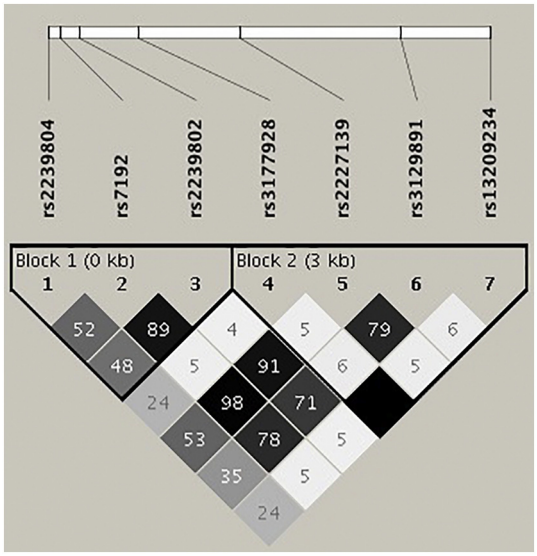


**Fig. 3. Plot of pairwise linkage disequilibrium (LD) among the significant SNPs in the 3'UTR region of *HLA-DRA*.** LD analysis showed that there was absolute LD ($r^2 = 1$) between the SNPs rs13209234-rs3177928.

**Table 3**
Significance of associations of two genetic variants remained valid on joint analysis ($n = 250$ patient-control pairs) with additional data: Results.

| SNP | MAF in TB patients | MAF in HC | *p* | OR | 99% CI |
|---|---|---|---|---|---|
| rs61104666 | 0.41 | 0.3 | 0.0005 | 1.58 | 1.12–2.22 |
| rs3177928 | 0.10 | 0.18 | 0.001 | 0.54 | 0.33–0.88 |

MAF: Minor Allele Frequency, OR: Odds Ratio, CI: Confidence Interval.

patients and household contacts were 44.75 (range 18–81) and 39.65 (range 15–66), respectively, similar to those observed among the original sets of recruits. On the additionally recruited patients and controls, we performed TaqMan assays to determine genotypes at SNPs (rs61104666 and rs3177928) discovered by us to be associated with the phenotype, with biological relevance. At these loci, the data with a combined sample size of 250 pairs of tuberculosis patients and household contacts were reanalyzed for validation of association discovered with a smaller (114 pairs) sample size. The results are presented in Table 3. All earlier results of significant association remained valid. rs13209234 and rs3177928, both from the gene *HLA-DRA*, are in complete LD ($r^2 = 1$). Therefore, only one of SNPs was genotyped in validation phase. In the joint analysis for validation, rs61104666 (OR = 1.58 (C·I 99% 1.1–2.2), $p = .0005$, Table 3) from *SIGLEC15* gene and rs3177928 (OR = 0.54 (C·I 99% 0.3–0.9), $p = .001$, Table 3) from *HLA-DRA* gene, remained significantly associated.

## 4. Discussion

In the present study, we have discovered novel associations of SNPs in and around *SIGLEC15* and *HLA-DRA* genes with tuberculosis in a homogeneous (Bengali) population of Indians. All TB patients, recruited into this study, were treatment naïve because we want to exclude the complex phenotypes of drug resistance cases, prevalent in India (WHO, 2018). Here, despite the fact that none of the observed association signals achieved stringent levels of genome wide significance after multiple testing corrections, likely due to small sample size of our study, the major findings from the study suggested a previously unexplored role of *HLA-DRA* gene in tuberculosis disease pathology.

The strongest association was with rs61104666, a synonymous exonic variant of the gene sialic acid binding immunoglobulin- like lectins 15 or *SIGLEC15*. Siglec-15 is highly conserved in vertebrates and is earlier reported to be involved in osteoclast differentiation (Macauley et al., 2014). Recent study shows that Siglec-15 is also expressed on tumor-associated macrophages and enhances TGF-β secretion in a DAP12–Syk signal transduction pathway (Takamiya et al., 2013). In our study, even though the SNP (rs61104666) identified is a synonymous change, but other SNPs located in the nearby region in LD with this SNP, may be of functional relevance. But, no other SNPs from this gene *SIGLEC15*, were present after quality control. The evidence of protein binding to the region in which this SNP is located is, however, weak (RegulomeDB score: 5). The association results with SNPs in the *HLA-DRA* region are more biologically relevant, as HLA class II antigens are

**Table 2**
Haplotypes from *HLA-DRA* gene were significantly associated with tuberculosis in the study population.

| SNPs | Haplotype | Frequency in TB | Frequency in HC | *p* |
|---|---|---|---|---|
| **rs3177928**\|rs2227139\|rs3129891\|**rs13209234** | **AAGA** | **0.07792** | **0.1978** | **0.00019** |
| | GGAG | 0.2554 | 0.2467 | 0.8297 |
| | GAAG | 0.04762 | 0.02203 | 0.1359 |
| | GAGG | 0.619 | 0.5335 | 0.0638 |
| **rs2239804**\|rs7192\|rs2239802 | AAC | 0.2371 | 0.2402 | 0.9377 |
| | AAG | 0.02155 | 0.0131 | 0.4872 |
| | **ACG** | **0.07759** | **0.1965** | **0.00020** |
| | GCG | 0.6638 | 0.5502 | 0.01253 |

(Tuberculosis patients: TB, Household contacts: HC) *HLA-DRA* SNPs, which are found significant in allelic association, are shown in bold.

involved in defense against bacterial infection. We have also discovered significant association with the "AAGA" haplotype, spanning two significantly associated variants (rs13209234 and rs3177928), with tuberculosis in the Indian population.

*HLA-DRA* is a member of HLA class II genes (Shiina and Inoko, 2005). It encodes the alpha chain of HLA-DR protein and heterodimerizes with beta chains (HLA-DRBs) anchoring in the cell membrane. Like other HLA molecules (major histocompatibility complex [MHC] class I and II), HLA-DR plays an important role in the immune system by presenting peptides on the cell surface of antigen-presenting cells (APCs, including B lymphocytes, dendritic cells, and macrophages) for recognition by T cells (Trowsdale, 2011). HLA Class II genes have been the focus of many association studies with bacterial infections, including tuberculosis (Meyer and Thye, 2014). Polymorphisms (rs557011, rs9271378, rs9272785 and rs41553512) in other HLA class II genes particularly *HLA-DRB1* and *HLA-DQB1* have been reported to be significantly associated with TB among Europeans and Asians (Qi et al., 2017; Sveinbjornsson et al., 2016). However, the reported SNPs from these studies were not present in our genotyping chip. Next, we checked whether these SNPs are in linkage disequilibrium, with our reported SNPs, as available from 1000 Genome Data on BEB population. We have found that rs557011, rs9272785 and rs41553512 are in linkage equilibrium with our significant SNPs rs13209234 and rs3177928, while only rs9271378 is in low LD (r2 = 0.10).

Regions near rs3177928, on *HLA-DRA* displays enhancer signature as indicated by presence of H3K4me1 and H3K27ac mark and DNase sensitive site in monocytes, natural killer cells and primary B cells from peripheral blood (RegulomeDB score: 1f). We have also observed presence of H3K4me3 mark, considered as promoter signature, around rs3177928, in primary B cells. POL2 and POL24H8 proteins also bind to this region in lymphoblastoid cell lines GM12878, GM12891 and GM18505 as determined by ChIP-Seq experiments (ENCODE Project Consortium, 2011). rs3177928 acts as a cis-eQTL for *HLA-DRB6* gene in whole blood tissue ($p = 6.8 \times 10^{-13}$). It also showed evidence for cis-eQTL for other HLA genes such as *HLA-DQ, HLA-DRB, MICB* in other tissue types, including lung ($p < .05$). rs13209234, the other associated SNP, also showed H3K27ac enhancer signature in Monocytes-CD14+ and primary B cells. This SNP is also considered a cis-eQTL for *HLA-DRB6* gene in whole blood tissue ($p = 7.3 \times 10^{-13}$) and for other HLA genes in several tissue types (RegulomeDB score: 3a). All these evidences indicate that SNPs present in the 3′UTR of *HLA-DRA* gene might have regulatory role for other class II HLA genes.

rs3177928, has long been associated with total cholesterol level among populations of European, East Asian, South Asian and African ancestry (Teslovich et al., 2010; Willer et al., 2013) and recently is also validated in an exome sequencing study of Mexican Americans (Gao et al., 2018). rs3177928-A significantly increases total cholesterol in blood than the other allele ($p < .05$). Moresoever, the ability of *M. tb.* to maintain a chronic infection was critically linked to its ability to acquire cholesterol from the host (Pandey and Sassetti, 2008).

Our study has identified novel genetic association with tuberculosis in the Indian (Bengali speaking) population. However, the major limitation of our study is lack of adequacy of sample size possibly resulting in lack of adequate statistical power for an exome wide study, even though we have replenished our sample set with additional samples for validation. A study with a larger sample size must be undertaken to validate the present results.

## Acknowledgements

We are grateful to all the study participants. Mr. Bijan Bhusan Bairagya is acknowledged for excellent technical assistance. We are thankful to Dr. Arindam Maitra and CoTeRI for conducting the genotyping and sequencing experiments. Ms. Anuradha Gautam is thanked for helping in TaqMan Genotyping.

## Funding

This work was supported by the grant BT/01/CEIB/11/VI/05 dated 23/11/2011, from Department of Biotechnology (DBT), Government of India, India. Partha P. Majumder was supported by the J.C.Bose Fellowship of the Government of India, Department of Science and Technology. Chandrika Bhattacharyya was supported by Junior and Senior Research Fellowship from University Grant Commission, India.

## Ethical approval

The study was approved by Institutional Ethics Committees of National Institute of Biomedical Genomics, Kalyani. Blood was drawn with written informed consent from volunteers and all the methods were carried out in accordance with the approved guidelines.

## Competing interest

None declared.

## References

Abhimany, M.B., 2012. Indian genome variation consortium. Footprints of genetic susceptibility to pulmonary tuberculosis: cytokine gene variants in north Indians. Indian J. Med. Res. 135, 763.
Abu, Z., Dayem Ullah, J.O., Wang, Jun, Nagano, Ai, Lemoine, Nicholas, Chelala, Claude, 2018. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. Nucleic Acids Res. 46, W109–W113.
Azad, A.K., Sadee, W., Schlesinger, L.S., 2012. Innate immune gene polymorphisms in tuberculosis. Infect. Immun. 80, 3343–3359.
Baghdadi, J.E., Orlova, M., Alter, A., Ranque, B., Chentoufi, M., Lazrak, F., Archane, M.I., Casanova, J.L., Benslimane, A., Schurr, E., Abel, L., 2006. An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. J. Exp. Med. 203, 1679–1684.
Bairagya, B.B., Bhattacharya, P., Bhattacharya, S.K., Dey, B., Dey, U., Ghosh, T., Maiti, S., Majumder, P.P., Mishra, K., Mukherjee, S., et al., 2008. Genetic variation and haplotype structures of innate immunity genes in eastern India. Infect. Genet. Evol. 8, 360–366.
Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265.
Bellamy, R., 2000. Identifying genetic susceptibility factors for tuberculosis in Africans: a combined approach using a candidate gene study and a genome-wide screen. Clin. Sci. (Lond.) 98, 245–250.
Bellamy, R., Beyers, N., McAdam, K.P., Ruwende, C., Gie, R., Samaai, P., Bester, D., Meyer, M., Corrah, T., Collin, M., et al., 2000. Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. Proc. Natl. Acad. Sci. U. S. A. 97, 8005–8009.
Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al., 2012. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 22, 1790–1797.
Chen, C., Liu, Q., Zhu, L., Yang, H., Lu, W., 2013. Vitamin D receptor gene polymorphisms on the risk of tuberculosis, a meta-analysis of 29 case-control studies. PLoS One 8, e83843.
Cheng, Y.C., Hsiao, F.C., Yeh, E.C., Lin, W.J., Tang, C.Y., Tseng, H.C., Wu, H.T., Liu, C.K., Chen, C.C., Chen, Y.T., Yao, A., 2012. VarioWatch: providing large-scale and comprehensive annotations on human genomic variants in the next generation sequencing era. Nucleic Acids Res. 40, W76–W81.
Chimusa, E.R., Zaitlen, N., Daya, M., Moller, M., van Helden, P.D., Mulder, N.J., Price, A.L., Hoal, E.G., 2014. Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. Hum. Mol. Genet. 23, 796–809.
Cooke, G.S., Campbell, S.J., Bennett, S., Lienhardt, C., McAdam, K.P., Sirugo, G., Sow, O., Gustafson, P., Mwangulu, F., van Helden, P., et al., 2008. Mapping of a novel susceptibility locus suggests a role for MC3R and CTSZ in human tuberculosis. Am. J. Respir. Crit. Care Med. 178, 203–207.
Curtis, J., Luo, Y., Zenner, H.L., Cuchet-Lourenco, D., Wu, C., Lo, K., Maes, M., Alisaac, A., Stebbings, E., Liu, J.Z., et al., 2015. Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. Nat. Genet. 47, 523–527.
Dittrich, N., Berrocal-Almanza, L.C., Thada, S., Goyal, S., Slevogt, H., Sumanlatha, G., Hussain, A., Sur, S., Burkert, S., Oh, D.Y., et al., 2015. Toll-like receptor 1 variations influence susceptibility and immune response to Mycobacterium tuberculosis. Tuberculosis (Edinb) 95, 328–335.
Ernst, J.D., 2012. The immunological life cycle of tuberculosis. Nat. Rev. Immunol. 12, 581–591.
Gao, C., Tabb, K.L., Dimitrov, L.M., Taylor, K.D., Wang, N., Guo, X., Long, J., Rotter, J.I., Watanabe, R.M., Curran, J.E., et al., 2018. Exome sequencing identifies genetic variants associated with circulating lipid levels in Mexican Americans: the insulin resistance atherosclerosis family study (IRASFS). Sci. Rep. 8, 5603.
Grant, A.V., Sabri, A., Abid, A., Abderrahmani Rhorfi, I., Benkirane, M., Souhi, H., Naji Amrani, H., Alaoui-Tahiri, K., Gharbaoui, Y., Lazrak, F., et al., 2016. A genome-wide

association study of pulmonary tuberculosis in Morocco. Hum. Genet. 135, 299–307.
Hong, E.P., Go, M.J., Kim, H.L., Park, J.W., 2017. Risk prediction of pulmonary tuberculosis using genetic and conventional risk factors in adult Korean population. PLoS One 12, e0174642.
http://www.tbfacts.org/tb-statistics-india/, 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660.
Joshi, L., Ponnana, M., Sivangala, R., Chelluri, L.K., Nallari, P., Penmetsa, S., Valluri, V., Gaddam, S., 2015. Evaluation of TNF-alpha, IL-10 and IL-6 cytokine production and their correlation with genotype variants amongst tuberculosis patients and their household contacts. PLoS One 10, e0137727.
Li, X., Yang, Y., Zhou, F., Zhang, Y., Lu, H., Jin, Q., Gao, L., 2011. SLC11A1 (NRAMP1) polymorphisms and tuberculosis susceptibility: updated systematic review and meta-analysis. PLoS One 6, e15831.
Macauley, M.S., Crocker, P.R., Paulson, J.C., 2014. Siglec-mediated regulation of immune cell function in disease. Nat. Rev. Immunol. 14, 653–666.
Mahasirimongkol, S., Yanai, H., Mushiroda, T., Promphittayarat, W., Wattanapokayakit, S., Phromjai, J., Yuliwulandari, R., Wichukchinda, N., Yowang, A., Yamada, N., et al., 2012. Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus for young tuberculosis. J. Hum. Genet. 57, 363–367.
Majumder, P.P., Staats, H.F., Sarkar-Roy, N., Varma, B., Ghosh, T., Maiti, S., Narayanasamy, K., Whisnant, C.C., Stephenson, J.L., Wagener, D.K., 2009. Genetic determinants of immune-response to a polysaccharide vaccine for typhoid. HUGO J. 3, 17–30.
Meyer, C.G., Thye, T., 2014. Host genetic studies in adult pulmonary tuberculosis. Semin. Immunol. 26, 445–453.
Mishra, G., Poojary, S.S., Raj, P., Tiwari, P.K., 2012. Genetic polymorphisms of CCL2, CCL5, CCR2 and CCR5 genes in Sahariya tribe of North Central India: an association study with pulmonary tuberculosis. Infect. Genet. Evol. *12*, 1120–1127.
Mishra, G., Kumar, N., Kaur, G., Jain, S., Tiwari, P.K., Mehra, N.K., 2014. Distribution of HLA-A, B and DRB1 alleles in Sahariya tribe of North Central India: an association with pulmonary tuberculosis. Infect. Genet. Evol. 22, 175–182.
Moller, M., Hoal, E.G., 2010. Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. Tuberculosis (Edinb) 90, 71–83.
Mukherjee, S., Sarkar-Roy, N., Wagener, D.K., Majumder, P.P., 2009. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. Proc. Natl. Acad. Sci. U. S. A. 106, 7073–7078.
Pandey, A.K., Sassetti, C.M., 2008. Mycobacterial persistence requires the utilization of host cholesterol. Proc. Natl. Acad. Sci. U. S. A. 105, 4376–4380.
Png, E., Alisjahbana, B., Sahiratmadja, E., Marzuki, S., Nelwan, R., Balabanova, Y., Nikolayevskyy, V., Drobniewski, F., Nejentsev, S., Adnan, I., et al., 2012. A genome wide association study of pulmonary tuberculosis susceptibility in Indonesians. BMC Med. Genet. 13, 5.
Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., Willer, C.J., 2010. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.
Qi, H., Zhang, Y.B., Sun, L., Chen, C., Xu, B., Xu, F., Liu, J.W., Liu, J.C., Jiao, W.W., Shen, C., et al., 2017. Discovery of susceptibility loci associated with tuberculosis in Han Chinese. Hum. Mol. Genet. 26, 4752–4763.
R: Team, R.C, 2013. R: a language and environment for statistical computing.
Ridruechai, C., Mahasirimongkol, S., Phromjai, J., Yanai, H., Nishida, N., Matsushita, I., Ohashi, J., Yamada, N., Moolphate, S., Summanapan, S., et al., 2010. Association analysis of susceptibility candidate region on chromosome 5q31 for tuberculosis. Genes Immun. 11, 416–422.
Saikia, B., Wanchu, A., Mahakur, S., Bind, M., Sarkar, K., Minz, R.W., 2015. Analysis of HLA association among North Indian HIV-positive individuals co-infected with Mycobacterium tuberculosis. Lung India 32, 449–452.
Sharma, S., Rathored, J., Ghosh, B., Sharma, S.K., 2010. Genetic polymorphisms in TNF genes and tuberculosis in North Indians. BMC Infect. Dis. 10, 165.
Sharma, P.R., Singh, S., Jena, M., Mishra, G., Prakash, R., Das, P.K., Bamezai, R.N., Tiwari, P.K., 2011. Coding and non-coding polymorphisms in VDR gene and susceptibility to pulmonary tuberculosis in tribes, castes and Muslims of Central India. Infect. Genet. Evol. 11, 1456–1461.
Shiina, T., Inoko, H., 2005. HLA gene region and its diversity. Nihon Rinsho 63 (Suppl. 4), 293–298.
Stein, C.M., Zalwango, S., Malone, L.L., Won, S., Mayanja-Kizza, H., Mugerwa, R.D., Leontiev, D.V., Thompson, C.L., Cartier, K.C., Elston, R.C., et al., 2008. Genome scan of M. tuberculosis infection and disease in Ugandans. PLoS One 3, e4094.
Sveinbjornsson, G., Gudbjartsson, D.F., Halldorsson, B.V., Kristinsson, K.G., Gottfredsson, M., Barrett, J.C., Gudmundsson, L.J., Blondal, K., Gylfason, A., Gudjonsson, S.A., et al., 2016. HLA class II sequence variants influence tuberculosis risk in populations of European ancestry. Nat. Genet. 48, 318–322.
Takamiya, R., Ohtsubo, K., Takamatsu, S., Taniguchi, N., Angata, T., 2013. The interaction between Siglec-15 and tumor-associated sialyl-Tn antigen enhances TGF-beta secretion from monocytes/macrophages through the DAP12-Syk pathway. Glycobiology 23, 178–187.
Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al., 2010. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466, 707–713.
Thye, T., Vannberg, F.O., Wong, S.H., Owusu-Dabo, E., Osei, I., Gyapong, J., Sirugo, G., Sisay-Joof, F., Enimil, A., Chinbuah, M.A., et al., 2010. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat. Genet. 42, 739–741.
Thye, T., Owusu-Dabo, E., Vannberg, F.O., van Crevel, R., Curtis, J., Sahiratmadja, E., Balabanova, Y., Ehmen, C., Muntau, B., Ruge, G., et al., 2012. Common variants at 11p13 are associated with susceptibility to tuberculosis. Nat. Genet. 44, 257–259.
Trowsdale, J., 2011. The MHC, disease and selection. Immunol. Lett. 137, 1–8.
Ward, L.D., Kellis, M., 2016. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. 44, D877–D881.
WHO, 2018. Global Tuberculosis Report.
Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al., 2013. Discovery and refinement of loci associated with lipid levels. Nat. Genet. 45, 1274–1283.
Wu, G., Zhao, M., Gu, X., Yao, Y., Liu, H., Song, Y., 2014. The effect of P2X7 receptor 1513 polymorphism on susceptibility to tuberculosis: a meta-analysis. Infect. Genet. Evol. 24, 82–91.