Published in final edited form as:

Infect Genet Evol. 2020 November ; 85: 104560. doi:10.1016/j.meegid.2020.104560.

CLEC4E (Mincle) genetic variation associates with pulmonary tuberculosis in Guinea-Bissau (West Africa)

Jasmine M. Olvany^{a,1}, Lindsay N. Sausville^{b,1}, Marquitta J. White^c, Alessandra Tacconelli^d, Gloria Tavera^e, Rafal S. Sobota^f, Cinzia Ciccacci^{g,h}, Anders S. Bohlbroⁱ, Christian Wejse^{j,k,l}, Scott M. Williams^{m,*}, Giorgio Sirugo^{n,o,}**

^aDepartment of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH 44106, USA

^bDepartment of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH 44106, USA

^cDepartment of Medicine, University of California San Francisco, San Francisco, CA 94158, USA

^dCentro di Ricerca, Ospedale San Pietro Fatebenefratelli, Rome, Italy

^eDepartment of Clinical Translational Science Collaborative, Case Western Reserve University, Cleveland, OH 44106, USA

^fNorthwestern Memorial Hospital, Northwestern University, Chicago, IL 60611, USA

^gUniCamillus, Saint Camillus International University of Health Sciences, 00131, Rome, Italy

^hDepartment of Biomedicine and Prevention, Genetics Section, University of Rome Tor Vergata, Rome, Italy

ⁱDepartment of Clinical Medicine, Aarhus University Hospital, Skejby, Denmark

^jBandim Health Project, Danish Epidemiology Science Centre and Statens Serum Institute, Bissau, Guinea-Bissau

^kDepartment of Infectious Diseases, Aarhus University Hospital, Skejby, Denmark

^lCenter for Global Health, School of Public Health, Aarhus University, Skejby, Denmark

^mDepartments of Population and Quantitative Health Sciences, and Genetics and Genome Sciences, Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, OH 44106, USA

ⁿDepartment of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

*Corresponding author smw54@case.edu (S.M. Williams). **Corresponding author at: Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

Giorgio.Sirugo@pennmedicine.upenn.edu (G. Sirugo).

¹Co-first authors

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.meegid.2020.104560>.

Declaration of competing interest
None declared.

^oDivision of Translational Medicine and Human Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract

Tuberculosis (TB) is the leading cause of death from a single infectious agent. According to the WHO, 85% of cases in 2018 were pulmonary tuberculosis (PTB), making it the most prevalent form of the disease. Although the bacillus responsible for disease, *Mycobacterium tuberculosis* (MTB), is estimated to infect 1.7 billion people worldwide, only a small portion of those infected (5–10%) will transition into active TB. Because such a small fraction of infected people develop active disease, we hypothesized that underlying host genetic variation associates with developing active pulmonary disease. Variation in *CLEC4E* has been of interest in previous association studies showing either no effect or protection from PTB. For our study we assessed 60 SNPs in 11 immune genes, including *CLEC4E*, using a case-control study from Guinea-Bissau. The 289 cases and 322 controls differed in age, sex, and ethnicity all of which were included in adjusted models. Initial association analysis with unadjusted logistic regression revealed putative association with seven SNPs ($p < 0.05$). All SNPs were then assessed in an adjusted model. Of the six SNPs that remained significant, three of them were assigned to the *CLEC4E* gene (rs12302046, rs0841847, and rs11046143). Of these, only rs10841847 passed FDR adjustment for multiple testing. Adjusted regression analyses showed that the minor allele at rs10841847 associated with higher risk of developing PTB (OR = 1.55, CI = 1.22–1.96, p -value = 0.00036). Based on these initial association tests, *CLEC4E* seemed to be the predictor of interest for PTB risk in this population. Haplotype analysis (2-SNP and 3-SNP windows) showed that minor alleles in segments including rs10841847 were the only ones to pass the threshold of global significance, compared to other haplotypes (p -value < 0.05). Linkage disequilibrium patterns showed that rs12302046 is in high LD with rs10841847 ($r^2 = 0.67$), and all other SNPs lost significance when adjusted for rs10841847 effects. These findings indicate that rs10841847 in *CLEC4E* is the single best predictor of pulmonary tuberculosis risk in our study population. These results provide evidence for the hypothesis that genetic variation of *CLEC4E* influences risk to TB in Guinea-Bissau.

Keywords

Mycobacterium tuberculosis; Tuberculosis; CLEC4E; Host genetics; Innate immunity; Mincle

1. Introduction

Tuberculosis (TB) is the leading cause of death from an infectious agent. The bacillus responsible for disease, *Mycobacterium tuberculosis* (*M. tuberculosis*), is estimated to latently infect 1.7 billion people worldwide (Programme, 2019). From this large infected population, only 5–10% of infected individuals will progress from latent to active disease (Programme, 2019). If the disease progresses into the active state within the first two years after infection, it is considered primary TB. Because of the discrepancy between the number of active TB patients and latently infected individuals, it has been hypothesized that underlying variability of host genetics is a factor in differentiating the two conditions. There are several different clinical presentations of TB that are covered under the blanket term of tuberculosis, but the pulmonary form (PTB) is by far the most prevalent (Programme,

T.W.D.oGT, 2019). There have been a number of previous reports, including twin, candidate gene, linkage analysis, and genome-wide association (GWA) studies, which altogether bolster the hypothesis that host genetics influence susceptibility to PTB (Bellamy et al., 2000; Fitness et al., 2004; Jepson et al., 2001; Thye et al., 2010).

The variability in host genetics can affect PTB risk in either of the two arms of the immune system, the innate or the adaptive. Innate immunity is a first line of defense against pathogens, based on non-fixed repertoire of receptors to rapidly sense and clear infectious agents; the more flexible, memory-based, adaptive immune system depends on innate immunity for activation (Sattler, 2017; Yatim and Lakkis, 2015). Immune cells can facilitate innate immunity by recognizing pathogens as non-self through pattern recognition receptors (PRRs) (Sattler, 2017). These receptors recognize pathogen-associated molecular patterns (PAMPs) of infectious agents and promote phagocytosis by macrophages or cause the release of chemokines and cytokines (Sattler, 2017). Classes of PRRs include Toll-like receptors (TLRs) and C-type lectin receptors (CLRs) (Liu et al., 2015; Sattler, 2017). Adaptive immunity utilizes specialized responses based on the recognition of invading pathogens. This type of immunity consists of cell-mediated and humoral immunity (Yatim and Lakkis, 2015). Cell-mediated immunity involves the activation of immune cells, such as macrophages and T-lymphocytes, to promote direct killing of pathogens, apoptosis of infected cells, and secretion of cytokines (Sattler, 2017; Yatim and Lakkis, 2015). In contrast, humoral immunity relies primarily on the production and secretion of macromolecules including antibodies (Yatim and Lakkis, 2015).

Candidate gene studies have already associated several innate immune-related genes with TB susceptibility (Bragina et al., 2019; Fitness et al., 2004; Silva-Ramirez et al., 2019; Stein et al., 2017; Wang et al., 2018). The heritability of tuberculosis, estimated to be up to 68%, has a wide range of estimates and methodologies to evaluate it, but overall there are many studies which support that there is a heritable component to host response to the bacterium (Abel et al., 2014; McHenry et al., 2020; Stein et al., 2003; Stein et al., 2017). As explained above there are many components in the pathways leading to cytokine expression and immune response, therefore we believe many of the genetic variants accounting for the heritability of PTB susceptibility could still be unknown. Here, we evaluated 60 SNPs in 11 candidate genes previously connected to the phenotype and/or relevant immunological pathways, to elucidate further the role that human genetics may have on transition into PTB in a well characterized West African cohort (Arce et al., 2004; Eskdale et al., 1997; Eyerich et al., 2017; Hashemi et al., 2013; Hijikata et al., 2016; Jaiswal and Srivastava, 2018; Lu et al., 2018; Nishida et al., 2018; White et al., 2014).

2. Materials and Methods

2.1. Participant recruitment

During a two-year collection period starting November 2003, participants were recruited at a demographic surveillance site, The Bandim Health Project (BHP) in Bissau, the capital of Guinea-Bissau. Cases were chosen from residents and long-term guests who were newly diagnosed with PTB and over the age of 15. Diagnosis was made in one of two ways: 1) three positive sputum samples for acid fast bacteria or 2) utilizing the World Health

Organizations (WHO) clinical criteria for active PTB (Programme, 2019). Initially 344 people with active PTB agreed to participate and all participants initiated tuberculosis treatment at local health clinics. Of the original volunteers, 289 patients produced viable DNA and were included in the study.

Recruitment of the controls occurred concurrently from a random sample of 200 households pulled from the same geographic region. Because the incidence of active PTB in the BHP community was estimated to be 284/100,000 people per year in 2004, it is likely the controls were exposed to the bacterium (Lemvik et al., 2014). For this reason, and to avoid previously identified study issues, households with any history of PTB in the last 2 years were excluded to prevent the inclusion of subclinical cases in the control group. Potential controls were also excluded if they had a cough for more than 2 weeks or personal PTB history. With all of these criteria, 322 controls were included in the study.

Prior to our study there were known distribution discrepancies between case and control populations in both sex and ethnicity due to sampling methodology. In this study, controls were derived from nuclear families, and thus many married couples were collected balancing out any discrepancy. However, willing PTB cases were more likely to be male, which caused the sex differences between the case and control populations. Additionally, there were differences in willingness to participate by ethnicity in the control sample, but PTB cases participated regardless. HIV tests for the controls were not performed as this would have negatively impacted participation in the study. HIV status was therefore not included in our analyses. During this time in Guinea-Bissau, the prevalence of HIV1 was 4.6% and HIV2 was 4.4% with HIV/TB coinfection rates of 66/100,000 cases in 2006, according to cross-sectional studies in Bissau (Jespersen et al., 2015; Olesen et al., 2018). A more detailed collection methodology has been outlined in a previously published paper on this data set (Olesen et al., 2007).

A breakdown of demographic information, including all of the aspects discussed is presented in Table 1.

2.2. Genotyping and data cleaning

Collected DNA samples were isolated via a standard salting out procedure, and then analyzed for purity using spectrophotometry. The 60 SNPs from the 11 genes were selected using r^2 values ($r^2 < 0.8$) based on the Yoruba reference population in HapMap (Frazer et al., 2007) (Table 2, Supplementary Table 1). We chose SNPs that were not in high LD with each other in order to achieve better coverage of the potential associated regions of the genes with the limited SNPs chosen. Most of the genes chosen are involved with the regulation of inflammation, or were previously reported to be important to the phenotype (Table 2). After selection, genotyping was carried out on the SNPs by a Sequenome MassARRAY™ system in the Center for Human Genetics Research Core Laboratory at Vanderbilt University or TaqMan at the MRC Unit, The Gambia, using standard protocols. DNA samples were archived in the Gambian DNA bank, and shared with collaborators according to the guidelines described by (Sirugo et al., 2004).

To ensure the validity of the dataset, quality control procedures were completed using PLINK! v1.07 (Purcell et al., 2007). On the individual level, a threshold of no less than 95% genotyping efficiency was applied, which did not exclude any of the individuals. Afterwards, the SNPs were evaluated using the following QC measures: SNP genotyping efficiency greater than 95%, minor allele frequency (MAF) greater than 10% (to exclude rare variants), and no evidence of deviation from Hardy-Weinberg equilibrium ($p > 0.01$). None of the SNPs were excluded on the basis of QC, leaving a total of 60 SNPs in 11 genes for genetic association in 611 individuals.

2.3. Statistical analysis

Summary statistics were performed in Rstudio 3.5.3, and all association analyses were performed in PLINK v.1.07 (Purcell et al., 2007). Demographic variables were compared between cases and controls. Binary variables were compared between the two groups using chi-square analysis. Continuous variables were tested for normality with the Shapiro-Wilkes test, and then compared either through the Student's t -test for normally distributed or the Wilcoxon rank-sum test for non-normally distributed variables. These analyses indicated that age and sex were both distributed differently between the case and control populations. Therefore, both age and sex were included in adjusted modeling of genetic association. Additionally, the logistic regression analysis was adjusted by ethnicity, because it has been previously associated with PTB in higher-ordered genetic interaction (White et al., 2014).

Unadjusted logistic regression was utilized as an initial test of SNP association with disease status. All markers were analyzed using adjusted logistic regression models. Covariates in the adjusted models were sex, age assessed in quartiles, and six self-reported ethnicities and another category for rare ethnicities. All results were corrected for multiple testing using a false discovery rate (FDR) of $q = 0.05$ based on the 60 SNPs tested. After additive modeling was explored, all markers were evaluated in both the dominant and recessive models.

Because the exploratory association identified an FDR significant SNP in the *CLEC4E* gene, follow-up haplotype association studies were conducted on this gene. Initially, to determine the haplotypes of interest, two and three-SNP sliding windows were used to generate omnibus p -values. To assess whether haplotype associations were driven by the single SNP that marginally associated with PTB, rs10841847, we tested all pairwise haplotypes that included this SNP. Any window that was originally under the $p < 0.05$ threshold was further evaluated for specific haplotypes using a generalized linear model. This allows the different alleles to be assigned a directionality of effect. We also ran logistic regression on all SNPs in *CLEC4E* conditioned on rs10841847. Finally, patterns of linkage disequilibrium, measured by r^2 , were visualized in case, control, and collectively using Haploview software (Barrett et al., 2005).

3. Results

3.1. Demographic description

After cleaning, the Guinea-Bissau cohort consisted of 289 cases and 322 controls (Fig. 1). The PTB cases were significantly younger than controls (36.86 years vs 39.17 years), and

there was an excess of males (60.9%). This sex differential is expected and has been previously reported (Jaillon et al., 2019; Schurz et al., 2018). Six common ethnic groups were identified in the study group by consisting of at least 5% of the overall collected population (Table 1).

3.2. Logistic regression association

Initial association analysis with unadjusted logistic regression revealed putative association with seven SNPs at a $p < 0.05$ (Table 3). For six of the seven SNPs the minor allele associated with increased odds of PTB. The minor allele for one SNPs (rs2227478 in *IL22*) was associated with decreased odds of PTB. Three of the seven associating SNPs were in *CLEC4E*.

The SNPs were then assessed in a model adjusted for age ($p = 0.0032$), sex ($p = 0.0026$), and ethnicity ($p = 0.014$). One of the seven originally associated SNPs was no longer significantly associated ($p < 0.05$) when these covariates were included in the model (Table 4). No additional SNPs became marginally associated in this model. The other six SNPs remained significant and had lower p -values in the adjusted models. After multiple testing adjustment (FDR $q = 0.05$), only one SNP was associated with PTB: rs10841847 in the *CLEC4E* gene in an additive genetic model. None of the other inheritance models showed any significant results after adjustment for multiple testing (Supplementary Tables 2 and 3).

3.3. CLEC4E focused analysis

Based on the initial association tests, *CLEC4E* seems to be single gene most associated with PTB in this population; for this reason, a deeper analysis of this gene was performed. Thus, the effects of different *CLEC4E* haplotypes on PTB risk was examined by evaluating all 8 SNPs in two- and three-SNP sliding windows. Only four haplotypes were globally associated with the disease phenotype ($p < 0.05$), and all contained the only individually significant SNP (rs10841847) (Table 5; Fig. 2).

To assess whether rs10841847 was driving these associations we created two-SNP haplotypes that included rs10841847 and each of the other SNPs in *CLEC4E*. All of these two-SNP haplotype windows either remained significantly associated or became significantly associated with PTB (Fig. 3, top).

The windows that were significant using the omnibus analyses were individually tested for haplotype directionality. This allowed us to see whether the alleles are associating with increased or decreased susceptibility to PTB. Overall the minor allele at rs10841847 (G > A) contributes to increased susceptibility to PTB across all haplotypes that include it, but not all haplotypes that include it are significant.

Most of the SNPs in the Guinea-Bissau population were not in high LD as predicted by the choice of SNPs from the Yoruba samples. Given that most SNPs were not in high LD, we tested whether the other SNPs in *CLEC4E* associated after adjustment for rs10841847. None of the other *CLEC4E* SNPs putatively associating with PTB in the marginal analyses remained significant after adjusting for rs10841847 (Table 6). These results and those of the

haplotype analyses indicate that rs10841847 or another variant in high LD with it, is responsible for the signal.

4. Discussion

Our results show that *CLEC4E* associates with PTB. The other genes in the study were theorized to be important to the phenotype or had previously been shown to be associated with it, but showed no statistical association in our Guinea-Bissau sample. It is possible that the other 10 genes studied here, may contribute to overall heritability of pulmonary tuberculosis in other populations. Within *CLEC4E*, one SNP, rs10841847, appears to be the sole driver of the association. Two separate tests we conducted bolster this assertion. Because no haplotype combination was more significant than the single variant and no other SNPs in the gene remain significant after the effects of rs10841847 are adjusted for, it is clear that this is the single best predictor in our sample population. As stated above, this indicates that rs10841847 is a variant of functional interest or it marks a non-genotyped variant in the region. We explored potential functional consequences of this SNP using the SNPinfo tool, FUNCPred, and found no role annotated or estimated to date (Xu and Taylor, 2009).

Our results indicate that the other candidate genes we assayed did not stand up to robust analyses and corrections for multiple testing. Although it is impossible to know why they did not show strong evidence of association it is possible that genetic variability between different ancestral populations and either subsequent backgrounds effects of environmental factors not measured could cause them to show no robust association (Stein et al., 2017). Also, differences in patterns of linkage disequilibrium among populations could cause effects to be lost if the genotyped variant was not the functional one. Thus is it not very surprising that genes of potential interest may not replicate between different samples (Sirugo et al., 2019). Nonetheless, we present all of the results, some of which have weaker evidence for association (e.g. *IL22* and *CTSZ*) so that others can compare to their findings.

CLEC4E is found in the natural killer region of chromosome 12 (Bowker et al., 2016). The protein encoded by *CLEC4E*, a receptor termed “macrophage-inducible C-type lectin” (Mincle), is an activating PRR located on the outer membrane of macrophages which recognizes the trehalose-6,6-dimycolate (TDM) component in the mycobacterial cell wall, and signals for the innate immune system to respond to infection (Lu et al., 2018; Schoenen et al., 2010). Two previously published papers focused on variants in the *CLEC4E* gene, but the findings were inconsistent (Bowker et al., 2016; Kabuye et al., 2019). One study on admixed South Africans presented evidence that *CLEC4E* variants may increase susceptibility to active TB, but the results were not statistically significant (Bowker et al., 2016). However, another study from Northern China reported that *CLEC4E* minor alleles associated with decreased predisposition to active TB (Kabuye et al., 2019). Given the gene’s role in innate immunity, it is plausible that *CLEC4E* could modulate vulnerability to active TB (Lu et al., 2018). Therefore, there is significant motivation to resolve the conflicting reports and gain etiological insights.

CLEC4E is likely involved with susceptibility in that *M. tuberculosis* is transmitted through inhalation of aerosolized mycobacteria into the lungs of a new host, where alveolar macrophages are the primary defense cells (Behler et al., 2012; Lu et al., 2018; Programme, 2019). Binding of Mincle to its ligand signals through Immunoreceptor Tyrosine-based Activation Motif (ITAM), paired with the Fragment crystallizable Receptor gamma (FcR γ) adapter protein, initiating a signaling pathway to activation of the Syk-CARD9-BCL10-Malt1 pathway which cascades leading to the canonical activation of NF- κ B (Shenderov et al., 2013). This activation eventually leads to a Th1/Th17 immune response including inflammasome recruitment, cytokine release, and eventually lung granuloma formation to defend against the invading pathogen (Shenderov et al., 2013). Of additional relevance, Mincle – / – mice lacked the ability to recruit and adhere neutrophils to the site of the invasion (Lee et al., 2012). A similar study also found that knockout mice had a higher bacterial load than the wild type, indicating that clearing the *Mtb* mycobacteria may require a Mincle-mediated immune response (Behler et al., 2012).

While some studies have found that the Mincle – / – mice can still mount a response to mycobacteria and form granulomas, these studies often do not capture the nuances involved in a sustained immune response or secondary TB (Heitmann et al., 2013). Mincle has been linked to recognizing cell death markers like SAP130 and mediating an inflammatory response to dying tissue (Lu et al., 2018; Nagata et al., 2017; Yamasaki et al., 2008). Overall, these findings support the interpretation that our results reflect the importance of Mincle function in resisting the progression to active TB. It has also been claimed that the *Mtb* strain-specificity can have an impact on the significance of Mincle, because the level of TDM PAMP on the cell wall varies from strain to strain (Lang, 2013). Either way, our results support the conclusion that altered Mincle function can influence the host's ability to defend against PTB.

All of our data taken together indicate that Mincle needs further evaluation in biological studies of TB. This is in contrast to the paper published in the South African population, which investigated the same allele for susceptibility to TB, but did not find statistical significance. We believe this might have been due to latently infected individuals being assigned to the control population, thus decreasing the power to find resolution between cases and controls. By adopting conservative parameters for recruiting controls (so as to exclude latent TB, theoretically), we may have circumvented the confounding factor that could have led to a negative association, and allowed us to detect a positive association between *CLEC4E* variation and PTB susceptibility. To the best of our knowledge, this is the first TB association study of *CLEC4E*/Mincle in a West African population. Our findings are also in contrast to the more recent study that reported the same allele in rs10841847 as a protective variant for PTB in a Northern Chinese population. We have no explanation for this discrepancy, and we believe it just points further to the need for a closer study of the role of *CLEC4E* in human susceptibility to active TB, as the preponderance of current data supports the high biological relevance of the receptor.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

Funding for this work was provided by the grants R56 A1130947 and T32 HL007567 from the National Institutes of Health of the United States.

References

- Abel L, El-Baghdadi J, Bousfiha AA, Casanova JL, Schurr E, 2014. Human genetics of tuberculosis: a long and winding road. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci* 369 20130428. [PubMed: 24821915]
- Arce I, Martinez-Munoz L, Roda-Navarro P, Fernandez-Ruiz E, 2004. The human C-type lectin CLECSF8 is a novel monocyte/macrophage endocytic receptor. *Eur. J. Immunol* 34, 210–220. [PubMed: 14971047]
- Barrett JC, Fry B, Mailer J, Daly MJ, 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. [PubMed: 15297300]
- Behler F, Steinwede K, Balboa L, Ueberberg B, Maus R, Kirchhof G, Yamasaki S, Welte T, Maus UA, 2012. Role of Mincle in alveolar macrophage-dependent innate immunity against mycobacterial infections in mice. *J. Immunol* 189, 3121–3129. [PubMed: 22869905]
- Bellamy R, Beyers N, McAdam KP, Ruwende C, Gie R, Samaai P, Bester D, Meyer M, Corrah T, Collin M, Camidge DR, Wilkinson D, Hoal-Van Helden E, Whittle HC, Amos W, van Helden P, Hill AV, 2000. Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. *Proc. Natl. Acad. Sci. U. S. A* 97, 8005–8009. [PubMed: 10859364]
- Bowker N, Salie M, Schurz H, van Heiden PD, Kinnear CJ, Hoal EG, Moller M, 2016. Polymorphisms in the pattern recognition receptor mincle gene (CLEC4E) and association with tuberculosis. *Lung* 194, 763–767. [PubMed: 27363694]
- Bragina EY, Babushkina NP, Garaeva AF, Rudko AA, Tsitrikov DY, Gomboeva DE, Freidin MB, 2019. Impact of the polymorphism of the PACRG and CD80 genes on the development of the different stages of tuberculosis infection. *Iran J. Med. Sci* 44, 236–244. [PubMed: 31182890]
- Eskdale J, Kube D, Tesch H, Gallagher G, 1997. Mapping of the human IL10 gene and further characterization of the 5' flanking sequence. *Immunogenetics* 46, 120–128. [PubMed: 9162098]
- Eyerich K, Dimartino V, Cavani A, 2017. IL-17 and IL-22 in immunity: driving protection and pathology. *Eur. J. Immunol* 47, 607–614. [PubMed: 28295238]
- Fitness J, Floyd S, Warndorff DK, Sichali L, Malema S, Crampin AC, Fine PE, Hill AV, 2004. Large-scale candidate gene study of tuberculosis susceptibility in the Karonga District of Nothern Malawi. *Am. J. Hum. Genet* 71, 341–349.
- Frazer KA, Ballinger DC, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Wayne MMY, Tsui SKW, Xue H, Wong JT-F, Galver LM, Fan J-B, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier J-F, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok P-Y, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui L-C, Mak W, Qiang Song Y, Tam PKH, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PIW, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Vernon Smith A, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Steve Qin Z, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Johnson T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y,

Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimma C, Royal CDM, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Wright Clayton E, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang IR, Holden AL, Brooks LD, McEwen JE, Guyer MS, Ota Wang V, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, The International HapMap, C., Genotyping centres: Perlegen, S., Baylor College of, M., ParAllele B., Beijing Genomics, I., Broad Institute of, H., Massachusetts Institute of, T., Chinese National Human Genome Center at B., Chinese National Human Genome Center at, S., Chinese University of Hong, K., Hong Kong University of, S., Technology, Illumine, McGill, U., Génome Québec Innovation, C., University of California at San, F., Washington, U., University of Hong, K., University of, T., Riken, Wellcome Trust Sanger, I., Analysis groups: Broad, I., Cold Spring Harbor, L., Johns Hopkins University School of, M., University of, M., University of, O., University of Oxford, W.T.C.f.H.G., Riken, Health, U.S.N.I.o., Information, U.S.N.I.o.H.N.C.f.B., Community engagement/public, c., sample collection groups: Beijing Normal, U., Beijing Genomics, I., Health Sciences University of Hokkaido, Shinshu, U., Howard, U., University of, I., University of, U., Ethical, I., social issues: Chinese Academy of Social, S., Genetic Interest, G., Kyoto U., Nagasaki U., University of Ibadan School of, M., University of, M., University of, O., Vanderbilt, U., Wellcome, T., Medicine, S.N.P.d.B.C.o., Washington, U., Scientific management: Chinese Academy of, S., Genome, C., Génome, Q., Japanese Ministry of Education, C.S.S., Technology, Ministry of, S., Technology of the People's Republic of, C., The Human Genetic Resource Administration of, C., The, S.N.P.C., 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861. [PubMed: 17943122]

Hashemi M, Eskandari-Nasab E, Moazeni-Roodi A, Naderi M, Sharifi-Mood B, Taheri M, 2013.

Association of CTSZ rs34069356 and MC3R rs6127698 gene polymorphisms with pulmonary tuberculosis. *Int. J. Tuberc. Lung Dis* 17, 1224–1228.

Heitmann L, Schoenen H, Ehlers S, Lang R, Holscher C, 2013. Mincle is not essential for controlling *Mycobacterium tuberculosis* infection. *immunobiology* 218, 506–516. [PubMed: 22784441]

Hijikata M, Matsushita I, Le Hang NT, Thuong PH, Tam DB, Maeda S, Sakurada S, Cuong VC, Lien LT, Keicho N, 2016. Influence of the polymorphism of the DUSP14 gene on the expression of immune-related genes and development of pulmonary tuberculosis. *Genes Immun* 17, 207–212. [PubMed: 26938665]

Jailion S, Berthenet K, Cauiaamla C, 2019. Sexual dimorphism in innate immunity. *Clin. Rev. Allergy Immunol* 56, 308–321.

Jaiswal S, Srivastava KK, 2018. Protein tyrosine kinase A modulates intracellular survival of mycobacteria through Galectin 3. *Biochem. Biophys. Res. Commun* 498, 884–890. [PubMed: 29545176]

Jepson A, Fowler A, Banya W, Singh M, Bennett S, Whittle H, Hill AV, 2001. Genetic regulation of acquired immune responses to antigens of *Mycobacterium tuberculosis*: a study of twins in West Africa. *Infect. Immun* 69, 3989–3994.

Jespersen S, Hønge BL, Oliveira I, Medina C, da Silva Te D, Correia FG, Erikstrup C, Laursen AL, Ostergaard L, Wejse C, Bissau H.I.V.Cs.g, 2015. Cohort profile: the Bissau HIV cohort—a cohort of HIV-1, HIV-2 and co-infected patients. *Int. J. Epidemiol* 44, 756–763. [PubMed: 25342251]

Kabuye D, Chu Y, Lao W, Jin G, Kang H, 2019. Association between CLEC4E gene polymorphism of mincle and pulmonary tuberculosis infection in a northern Chinese population. *Gene* 710, 24–29. [PubMed: 31075410]

Lang R, 2013. Recognition of the mycobacterial cord factor by Mincle: relevance for granuloma formation and resistance to tuberculosis. *Front. Immunol* 4, 5. [PubMed: 23355839]

Lee WB, Kang JS, Yan JJ, Lee MS, Jeon BY, Cho SN, Kim YJ, 2012. Neutrophils promote mycobacterial trehalose dimycolate-induced lung inflammation via the Mincle pathway. *PLoS Pathog* 8, e1002614. [PubMed: 22496642]

- Lemvik G, Rudolf F, Vieira F, Sodemann M, Ostergaard L, Rodrigues A, Gomes V, Aaby P, Wejse C, 2014. Decline in overall, smear-negative and HIV-positive TB incidence while smear-positive incidence stays stable in Guinea-Bissau 2004–2011. *Tropical Med. Int. Health* 19, 1367–1376.
- Liu Y, Liu J, Pang X, Liu T, Ning Z, Cheng G, 2015. The roles of direct recognition by animal lectins in antiviral immunity and viral pathogenesis. *Molecules* 20, 2272–2295. [PubMed: 25642837]
- Lu X, Nagata M, Yamasaki S, 2018. Mincle: 20 years of a versatile sensor of insults. *Int. Immunol* 30, 233–239. [PubMed: 29726997]
- McHenry ML, Bartlett J, Igo RP, Wampande E, Benchek P, Mayanja-Kizza H, Fluegge K, Hall NB, Gagneux S, Tishkoff SA, Wejse C, Sirugo G, Boom WH, Joloba M, Williams SM, Stein CM, 2020. Interaction between Host Genes and Mycobacterium Tuberculosis Lineage Can Affect Tuberculosis Severity: Evidence for Coevolution? *PLoS Genet* 16 (4), e1008728. 10.1371/journal.pgen.1008728. [PubMed: 32352966]
- Nagata M, Izumi Y, Ishikawa E., Kiyotake R, Doi R, Iwai S, Omande Z, Yamaji T, Miyamoto T, Samba T, Yamasaki S, 2017. Intracellular metabolite beta-glucosylceramide is an endogenous Mincle ligand possessing immunostimulatory activity. *Proc. Natl. Acad. Sci. U. S. A* 114, E3285–E3294. [PubMed: 28373578]
- Nishida N, Aiba Y, Hitomi Y, Kawashima M, Kojima K, Kawai Y, Ueno K, Nakamura H, Yamashiki N, Tanaka T, Tamura S, Mori A, Yagi S, Soejima Y, Yoshizumi T, Takatsuki M, Tanaka A, Harada K, Shimoda S, Komori A, Eguchi S, Maehara Y, Uemoto S, Kokudo N, Nagasaki M, Tokunaga K, Nakamura M, 2018. NELFCD and CTSZ loci are associated with jaundice-stage progression in primary biliary cholangitis in the Japanese population. *Sci. Rep* 8 8071. [PubMed: 29795304]
- Olesen R, Wejse C, Velez DR, Bisseye C, Sodemann M, Aaby P, Rabna P, Worwui A, Chapman H, Diatta M, Adegbola RA, Hill PC, Østergaard L, Williams SM, Sirugo G, 2007. DC-SIGN (CD209), pentraxin 3 and vitamin D receptor gene variants associate with pulmonary tuberculosis risk in West Africans. *Genes Immun* 8, 456–467.
- Olesen JS, Jespersen S, da Silva ZJ, Rodrigues A, Erikstrup C, Aaby P, Wejse C, Hønge BL, 2018. HIV-2 continues to decrease, whereas HIV-1 is stabilizing in Guinea-Bissau. *AIDS* 32, 1193–1198. [PubMed: 29683842]
- Programme, T.W.D.oGT, 2019. Global Tuberculosis Report 2019 Global Tuberculosis Report World Health Organization, pp. 1–283 Online.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC, 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet* 81, 559–575. [PubMed: 17701901]
- Sattler S, 2017. The role of the immune system beyond the fight against infection. *Adv. Exp. Med. Biol* 1003, 3–14. [PubMed: 28667551]
- Schoenen H, Bodendorfer B, Hitchens K, Manzanero S, Wemlinghaus K, Nimmerjahn F, Agger EM, Stenger S, Andersen P, Ruland J, Brown GD, Wells C, Lang IL, 2010. Cutting edge: Mincle is essential for recognition and adjuvanticity of the mycobacterial cord factor and its synthetic analog trehalose-dibehenate. *J. Immunol* 184, 2756–2760. [PubMed: 20164423]
- Schurz H, Kinnear CJ, Gignoux C, Wojcik G, van Helden PD, Tromp G, Henn B, Hoal EG, Moller M, 2018. A sex-stratified genome-wide association study of tuberculosis using a multi-ethnic genotyping array. *Front. Genet* 9, 678. [PubMed: 30713548]
- Shenderov K, Barber DL, Mayer-Barber KD, Guricha SS, Jankovic D, Feng CG, Oland S, Hieny S, Caspar P, Yamasaki S, Lin X, Ting JP, Trinchieri G, Besra GS, Cerundolo V, Sher A, 2013. Cord factor and peptidoglycan recapitulate the Th17-promoting adjuvant activity of mycobacteria through mincle/CARD9 signaling and the inflammasome. *J. Immunol* 190, 5722–5730. [PubMed: 23630357]
- Silva-Ramirez B, Saenz-Saenz CA, Bracho-Vela LA, Penuelas-Urquides K, Mata-Tijerina V, Escobedo-Guajardo BL, Gonzalez-Rios N,R, Vazquez-Monsivais O, Leon MB, 2019. Association between vitamin D receptor gene polymorphism and pulmonary tuberculosis in a Mexican population. *Indian J. Tuberc* 66, 70–75.
- Sirugo G, Schim van der Loeff M, Sam O, Nyan O, Pinder M, Hill AV, Kwiatkowski D, Prentice A, de Toma C, Cann HM, Diatta M, Jallow M, Morgan G, Clarke M, Corrah T, Whittle H, 2004. A National DNA Bank in the Gambia. West Africa. and Genomic Research in Developing Countries. *Nat Genet* 36 (8), 785–786. 10.1038/ng0804-785. [PubMed: 15284842]

- Sirugo G, Williams SM, Tishkoff SA, 2019. The Missing Diversity in Human Genetic Studies. *Cell* 177, 26–31. 10.1016/j.cell.2019.02.048. [PubMed: 30901543]
- Stem CM, Guwatudde D, Nakakeeto M, Peters P, Elston RC, Tiwari HK, Mugerwa R, Whalen CC, 2003. Heritability analysis of cytokines as intermediate phenotypes of tuberculosis-J. *Infect. Dis* 187, 1679–1685.
- Stein CM, Sausville L, Wejse C, Sobota RS, Zetola NM, Hill PC, Boom WH, Scott WK, Sirugo G, Williams SM, 2017. Genomics of human pulmonary tuberculosis: from genes to pathways. *Curr Genet Med Rep* 5, 149–166. [PubMed: 29805915]
- Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, Sirugo G, Sisay-Joof F, Enimil A, Chinbualt MA, Floyd S, Warndorff DK, Sichali L, Malema S, Crampin AC, Ngvira B, Teo YY, Small K, Rockett K, Kwiatkowski D, Fine PE, Hill PC, Newport M, Lienhardt C, Adegbola RA, Corrach T, Ziegler A, African TBGC, Wellcome Trust Case Control, C., Morris AP, Meyer CG, Horstmann RD, Hill AVS, 2010. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat. Genet* 42, 739–741. [PubMed: 20694014]
- Wang Y, Zhang MM, Huang WW, Wu SQ, Wang MG, Tang XY, Sandford AJ, He JQ, 2018. Polymorphisms in toll-like receptor 10 and tuberculosis susceptibility: evidence from three independent series. *Front. Immunol* 9, 309. [PubMed: 29527210]
- White MJ, Tacconelli A, Chen JS, Weise C, Hill PC, Gomes VF, Velez-Edwards DR, Ostergaard LJ, Hu T, Moore JH, Novelli G, Scott WK, Williams SM, Sirugo G, 2014. Epiregulin (EREG) and human V-ATPase (TCIRG1): genetic variation, ethnicity and pulmonary tuberculosis susceptibility in Guinea-Bissau and the Gambia. *Genes Immun* 15, 370–377. [PubMed: 24898387]
- Xu Z, Taylor JA, 2009. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res* 37, W600–W605. [PubMed: 19417063]
- Yamasaki S, Ishikawa E, Sakuma M, Hara H, Ogata K, Saito T, 2008. Mincle is an ITAM-coupled activating receptor that senses damaged cells. *Nat. Immunol* 9, 1179–1188. [PubMed: 18776906]
- Yatim KM, Lakkis FG, 2015. A brief journey through the immune system. *Chit. J. Am. Soc. Nephrol* 10, 1274–1281.

Fig. 1.
Flow chart of study exclusion/inclusion methodology

Fig. 2. 2–3 SNP sliding windows on LD plot.

Only markers in the *CLEC4E* gene are shown above. The red bars represent pairs or triple SNPs which passed a $p=0.05$ significance level during window/disease status inquiry. These windows include the following SNPs: *rs12302046*, *rs10841847*, *rs10841856*, and *rs4883167*. Superimposed on the data generated in PLINK is the Haploview Linkage Disequilibrium output, in r^2 measures.

Fig. 3. 2-SNP windows including rs10841847.

The red dots represent rs10841847-SNP pairs which passed a $p=0.05$ significance level during window/disease status inquiry. Superimposed on the data generated in PLINK is the Haploview Linkage Disequilibrium output, in r^2 measures.

Table 1:

Demographic Information for the Study Population

Variable	Case (n = 289)	Control (n = 322)	P-value
Mean Age (SD)	36.86 (13.82)	39.17 (12.22)	0.0032
Sex (% Male)	176 (60.90)	161 (50.00)	0.0026
Ethnicity:			0.014
Balanta (%)	41 (14.19)	60 (18.63)	
Fula (%)	43 (14.88)	39 (12.11)	
Mancanha (%)	23 (7.96)	39 (12.11)	
Mandinka (%)	20 (6.92)	23 (7.14)	
Manjago (%)	55 (19.03)	32 (9.94)	
Other (%)	47 (16.26)	35 (10.85)	
Pepel (%)	60 (20.76)	94 (29.19)	

For the Table 1, age was determined to be non-normally distributed, thus the p-value was generated via a Wilcoxon rank-sum test. The other two variables were considered normal and were generated via χ^2 analysis.

Table 2:

Investigated genes and their function

Chromosome	Gene	Function
1	<i>IL10</i>	regulatory cytokine
4	<i>EREG</i>	epidermal growth factor ligand
11	<i>TCIRG1</i>	vacuolar ATPase
12	<i>CLEC4D</i>	C-type lectin receptor
12	<i>CLEC4E</i>	C-type lectin receptor
12	<i>IL22</i>	stimulatory cytokine
14	<i>LGALS3</i>	regulator of apoptosis
17	<i>DUSP14</i>	negative regulator of T-cell proliferation and cytokine production
20	<i>MC3R</i>	G-protein coupled receptor
20	<i>CTSZ</i>	cysteine protease
20	<i>NELFCD</i>	member of NELF complex, inhibits RNA elongation by RNA polymerase II

References for the definitions are as follows: *IL10*¹, *EREG*², *TCIRG1*², *CLEC4D*³, *CLEC4E*⁴, *IL22*⁵, *LGALS3*⁶, *DUSP14*⁷, *MC3R*⁸, *CTSZ*⁸, *NELFCD*⁹

Table 3:

Unadjusted Logistic Regression

Unadjusted Logistic Regression				
SNP	Gene	OR	95% CI	p-value
rs10896289	<i>MIR6753</i>	1.29	1.01–1.66	0.044
rs12302046	<i>CLEC4E</i>	1.35	1.07–1.70	0.011
rs10841847	<i>CLEC4E</i>	1.48	1.17–1.87	0.00096
rs11046143	<i>CLEC4E</i>	1.32	1.05–1.66	0.017
rs2227478	<i>IL22</i>	0.769	0.598–0.989	0.041
rs2227472	<i>IL22</i>	1.37	1.08–1.73	0.0085
rs10369	<i>CTSZ</i>	1.32	1.03–1.69	0.030

Table 4:

Adjusted Logistic Regression

Adjusted Logistic Regression				
SNP	Gene	OR	95% CI	p-value
rs10896289	<i>MIR6753</i>	1.273	0.988–1.64	0.062*
rs12302046	<i>CLEC4E</i>	1.40	1.11–1.76	0.005
rs10841847	<i>CLEC4E</i>	1.55	1.22–1.96	0.00036 [±]
rs11046143	<i>CLEC4E</i>	1.35	1.07–1.71	0.013
rs2227478	<i>IL22</i>	0.772	0.598–0.997	0.047
rs2227472	<i>IL22</i>	1.414	1.11–1.80	0.0046
rs10369	<i>CTSZ</i>	1.40	1.08–1.81	0.011

[±]Effect remained significant after FDR correction (q=0.05)

* This SNP fell out of significance after adjusted for confounding variables detailed above.

Table 5

Haplotype Analysis

Haplotype	Haplotype Frequency		OR	P-value
rs12302046-rs10841847	Cases	Controls		
C-G	0.5	0.584	0.722	0.0046
A-G	0.057	0.067	0.843	0.48
C-A	0.029	0.02	1.51	0.29
A-A	0.413	0.329	1.44	0.0026

Haplotype	Haplotype Frequency		OR	p-value
rs12302046-rs10841847-rs10841856	Cases	Controls		
C-G-C	0.22	0.28	0.733	0.024
A-A-T	0.041	0.016	2.98	0.016
A-A-C	0.37	0.31	1.35	0.016
A-G-T	0.04	0.058	0.713	0.24
C-A-C	0.029	0.02	1.55	0.26
C-G-T	0.28	0.3	0.868	0.27
A-G-C *	0.017	0.0091	1.59	0.44
C-A-T *	2.8×10 ⁻⁹	2.0×10 ⁻⁹	NA	NA

Haplotype	Haplotype Frequency		OR	p-value
rs10841847-rs10841856	Cases	Controls		
G-C	0.242	0.288	0.765	0.044
G-T	0.315	0.362	0.821	0.11
A-T	0.041	0.016	2.97	0.016
A-C	0.401	0.333	1.39	0.0074

Haplotype	Haplotype Frequency		OR	p-value
rs10841847-rs10841856-rs4883167	Cases	Controls		
G-C-G	0.17	0.19	0.835	0.25
G-C-T	0.067	0.095	0.645	0.066
G-T-G	0.31	0.35	0.824	0.13
A-C-T	0.056	0.045	1.33	0.38
A-T-G	0.041	0.017	2.89	0.018
A-C-G	0.35	0.29	1.38	0.013
A-T-T *	8.7×10 ⁻⁹	2.0×10 ⁻⁷	NA	NA
G-T-T *	0.01	0.013	0.698	0.63

Global p-value = 0.012

Global p-value = 0.018

Global p-value = 0.0039

Global p-value = 0.033

Haplotypes are shown above as if walking down the gene from the 5' end, OR and both haplotype specific and global p-values were generated in PLINK v1.07.

* haplotypes were considered rare based on frequency, and any missing values occurred because there weren't enough occurrences for the software to calculate the measures reliably.

Table 6:Logistic Regression of the rest of *CLEC4E* Conditioned on rs10841847

SNP	OR	95% CI	P-value
rs12302046	0.952	0.630–1.44	0.81
rs10841856	1.13	0.857–1.50	0.38
rs4883167	0.878	0.622–1.24	0.46
rs11046143	1.08	0.811–1.45	0.59
rs10770847	1.08	0.818–1.42	0.60
rs10770850	1.12	0.866–1.45	0.39
rs10770854	1.16	0.909–1.48	0.23