A Sex-Stratified Genome-Wide Association Study of Tuberculosis Using a Multi-Ethnic Genotyping Array

Haiko Schurz[1,2]*, Craig J. Kinnear[1], Chris Gignoux[3], Genevieve Wojcik[4],
Paul D. van Helden[1], Gerard Tromp[1,2,5], Brenna Henn[6], Eileen G. Hoal[1] and Marlo Möller[1]

[1] DST-NRF Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, [2] South African Tuberculosis Bioinformatics Initiative, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, [3] Colorado Center for Personalized Medicine, Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, United States, [4] Department of Genetics, Stanford University, Stanford, CA, United States, [5] Centre for Bioinformatics and Computational Biology, Stellenbosch University, Cape Town, South Africa, [6] Department of Anthropology, UC Davis Genome Center, University of California, Davis, Davis, CA, United States

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, is a complex disease with a known human genetic component. Males seem to be more affected than females and in most countries the TB notification rate is twice as high in males than in females. While socio-economic status, behavior and sex hormones influence the male bias they do not fully account for it. Males have only one copy of the X chromosome, while diploid females are subject to X chromosome inactivation. In addition, the X chromosome codes for many immune-related genes, supporting the hypothesis that X-linked genes could contribute to TB susceptibility in a sex-biased manner. We report the first TB susceptibility genome-wide association study (GWAS) with a specific focus on sex-stratified autosomal analysis and the X chromosome. A total of 810 individuals (410 cases and 405 controls) from an admixed South African population were genotyped using the Illumina Multi Ethnic Genotyping Array, specifically designed as a suitable platform for diverse and admixed populations. Association testing was done on the autosome (8,27,386 variants) and X chromosome (20,939 variants) in a sex stratified and combined manner. SNP association testing was not statistically significant using a stringent cut-off for significance but revealed likely candidate genes that warrant further investigation. A genome wide interaction analysis detected 16 significant interactions. Finally, the results highlight the importance of sex-stratified analysis as strong sex-specific effects were identified on both the autosome and X chromosome.

Keywords: tuberculosis, GWAS, sex-bias, host genetics, X chromosome, sex-stratified, susceptibility

INTRODUCTION

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* (*M. tuberculosis*) is a global health epidemic and the leading cause of death due to a single infectious agent (World Health Organization [WHO], 2017). In 2016 1.3 million TB deaths were reported in HIV negative individuals and an additional 374000 deaths related to TB/HIV co-infection were recorded. The majority of these deaths occurred in southeast Asian and African countries (World Health Organization [WHO], 2017). TB is a complex disease, influenced by environmental and behavioral factors such as socio-economic status and smoking, as well as definite human genetic components. The contribution of the host genes to disease has been highlighted by numerous investigations, including animal (Pan et al., 2005), twin (Comstock, 1978; Sorensen et al., 1988; Flynn, 2006), linkage (Bellamy et al., 2000; Greenwood et al., 2000) and candidate gene association studies (Schurz et al., 2015). More recently genome-wide association studies (GWAS) in diverse populations have been done (Thye et al., 2010, 2012; Oki et al., 2011; Mahasirimongkol et al., 2012; Png et al., 2012; Chimusa et al., 2014; Curtis et al., 2015; Grant et al., 2016; Sobota et al., 2016; Qi et al., 2017).

Interestingly another influential factor in TB disease development is an individual's biological sex, which has been largely ignored in past TB studies and was usually only used as a covariate for adjusting association testing statistics. In 2016, males comprised 65% of the 10.4 million recorded TB cases, indicating that the TB notification rate is nearly twice as high in males as in females (World Health Organization [WHO], 2017). While socio-economic and behavioral factors do influence this ratio, it does not fully explain the observed sex-bias (Jaillon et al., 2017). Another factor that influences sex-bias is the effect

Abbreviations: 95CI, 95% confidence interval; *ARMCX1*, Armadillo repeat containing X-linked 1; *ARSF*, arylsulfatase F; *ASNS*, asparagine synthetase; *ATP2C1*, ATPase secretory pathway Ca^{2+} transporting 1; *C5orf64*, chromosome 5 open reading frame 64; *CFAP54*, cilia and flagella associated protein 54; *CIITA*, class II major histocompatibility complex transactivator; *CXorf51B*, chromosome X open reading frame 51B; *DIAPH2*, diaphanous related formin 2; DNA, deoxyribonucleic acid; *DPF3*, double PHD fingers 3; *DROSHA*, drosha ribonuclease III; *FRMPD4*, FERM and PDZ domain containing 4; *GRAMD2B*, GRAM domain containing 2B; GWAS, Genome-wide association study; HIV, human immunodeficiency virus; HWE, Hardy–Weinberg equilibrium; IL-10, interleukin 10; LD, linkage disequilibrium; *LINC00400*, long intergenic non-protein coding RNA 400; *LINC02153*, long intergenic non-protein coding RNA 2153; *LINC02246*, long intergenic non-protein coding RNA 2246; MAF, minor allele frequency; MEGA, multi-ethnic genotyping array; MHC, Major histocompatibility complex; *MIR514A1*, MicroRNA 514a-1; miRNA, micro RNA; *MTND6P12*, MT-ND6 pseudogene 12; *NCS1*, neuronal calcium sensor 1; *NF-kB*, nuclear factor kappa-light-chain-enhancer of activated B cells; *NTM*, neurotrimin; OR, odds ratio; P_comb, combined *p*-value using Stouffers method; P_Diff, *P*-value for sex-differentiation test; PBMC, peripheral blood mononuclear cell; *PCSK6*, proprotein convertase subtilisin/kexin type 6; pTB, pulmonary Tuberculosis; *RN7SKP120*, RNA, 7SK small nuclear pseudogene 120; RNA, ribonucleic acid; *RNF125*, ring finger protein 125; *RNF126*, ring finger protein 126; *RNU6-974P*, RNA, U6 small nuclear 974, pseudogene; *RTN4RL1*, reticulon 4 receptor like 1; SAC, South African colored; *SALL2*, spalt like transcription factor 2; SNP, single nucleotide polymorphism; *SRPX*, sushi repeat containing protein X-linked; TB, tuberculosis; *TBL1X*, transducin beta like 1 X-linked; *TENT4A*, terminal nucleotidyltransferase 4A; *TLR*, toll-like receptor; TNFα, tumor necrosis factor alpha; TST, tuberculin skin test; *URI1*, URI1, prefoldin like chaperone; *WT1*, Wilms tumor 1; XWAS, X chromosome wide association study.

that sex hormones (estrogen and testosterone) have on the immune system. Estrogen is an immune activator, upregulating pro-inflammatory cytokines (TNFα), while testosterone is an immune suppressor, upregulating anti-inflammatory cytokines (IL-10) (Cutolo et al., 2006). This could explain why men are more susceptible to infectious diseases compared to females (Jaillon et al., 2017). However, as sex-based differences in immune responses differ even between pre-pubertal boys and girls, as well as between post-menopausal women and elderly men, it shows that sex hormones do not fully explain the sex-bias (Klein et al., 2015). Thus, it has been proposed that the X chromosome and X-linked genes directly contribute to the observed sex-bias.

There are approximately 1,500 genes on the X chromosome, many of which are involved in the adaptive or innate immune system (Brooks, 2010). Since females have two X chromosomes, one requires silencing in order to equalize dosage of gene expression to that of men who only have one X chromosome. This silencing occurs randomly in each cell, making females functional mosaics for X linked genes and giving them a major immunological advantage over males (Jaillon et al., 2017). As males are haploid for X-linked genes any damaging polymorphisms or mutations on the X chromosome will have a more pronounced immunological effect in males than in mosaic females, thereby influencing the sex-bias (Abramowitz et al., 2014).

To date, eleven GWAS investigating susceptibility to clinical TB have been published (Thye et al., 2010, 2012; Oki et al., 2011; Mahasirimongkol et al., 2012; Png et al., 2012; Chimusa et al., 2014; Curtis et al., 2015; Grant et al., 2016; Sobota et al., 2016; Omae et al., 2017; Qi et al., 2017). There has not been significant overlap between the 11 published TB GWAS, but it seems that replication is more likely when populations with similar genetic backgrounds are compared: the *WT1* locus was associated with disease in populations from West and South Africa (Thye et al., 2012; Chimusa et al., 2014). Critically, genotyping microarrays that did not fully accommodate African genetic diversity were used in these studies (Thye et al., 2010, 2012; Chimusa et al., 2014; Curtis et al., 2015; Grant et al., 2016). It is therefore possible that unique African-specific susceptibility variants were not tagged by these initial arrays, since LD blocks are shorter in African populations (Campbell and Tishkoff, 2008). Moreover, none of the GWAS included or examined the X chromosome or sex-stratified analysis of the autosomes as was done in an asthma cohort (Mersha et al., 2015). Genetic differences between asthmatic males and females were identified on the autosome, with certain alleles having opposite effects between the sexes. Candidate gene association studies provide independent confirmation of the involvement of the X chromosome in TB susceptibility, through the association of X-linked *TLR8* susceptibility variants with active TB. Davila et al. (2008) investigated 4 *TLR8* variants (rs3761624, rs3788935, rs3674879, and rs3764880) in an Indonesian cohort and showed that all variants conferred susceptibility to TB in males but not females. The results for males were validated in male Russian individuals (Davila et al., 2008). These results were validated for rs3764880 in Turkish children, but no significant association

was found for rs3764879 (Dalgic et al., 2011). Hashemi-Shahri et al. (2014) found no significant *TLR8* associations in an Iranian population, while rs3764880 was significantly associated with TB susceptibility in both males and females in a Pakistani cohort (Bukhari et al., 2015). In admixed South African individuals rs3764879 and rs3764880 were significantly associated in both males and females, while rs3761624 was only significantly associated in females (Salie et al., 2015). Interestingly, in this cohort opposite effects were consistently found between the sexes for the same allele in all investigated *TLR8* variants (Salie et al., 2015), echoing the asthma findings of Mersha et al. (2015). Finally, in a Chinese cohort rs3764879 was significantly associated with TB disease in males but not females. While many of these variants did not reach genome wide significance they still provide evidence of the involvement of X-linked genes in TB susceptibility.

We report the first TB susceptibility GWAS with a specific focus on sex-stratified autosomal analysis and the X chromosome to elucidate the male sex-bias. Individuals from the unique five-way admixed SAC population, with ancestral contributions from Bantu-speaking African, KhoeSan, European, South and East Asian groups were genotyped in this study (Chimusa et al., 2013; Daya et al., 2013). These genetic contributions are due to both the complex colonization history of South Africa and the country's importance as a refreshment station on major trade routes during the fifteenth to nineteenth century (de Wit et al., 2010; Uren et al., 2016). This is therefore the first GWAS in the SAC that uses an array (Illumina Multi Ethnic Genotyping Array, see section "Genotyping") specifically designed to detect variants in the 4 most commonly studied populations, making it the most suitable platform for diverse and admixed populations at the time of genotyping.

MATERIALS AND METHODS

Study Population

Study participants were recruited from two suburbs in the Cape Town metropole of the Western Cape. These suburbs were chosen for its high TB incidence and low HIV prevalence (2%) at the time of sampling (1995–2005) (Kritzinger et al., 2009). Approximately 98% of the residents in these suburbs self-identify as SAC and have similar socio-economic status, which reduces confounding bias in the association testing (Chimusa et al., 2014). The cohort consists of 420 pulmonary TB (pTB) cases, bacteriologically confirmed to be culture and/or smear positive and 419 healthy controls from the same suburbs. Approximately 80% of individuals over the age of 15 years from these suburbs have a positive tuberculin skin test (TST), indicating exposure to *M. tuberculosis* (Gallant et al., 2010). All study participants were over 18 years of age and HIV negative.

Approval was obtained from the Health Research Ethics Committee of Stellenbosch University (project registration number S17/01/013 and 95/072) before participant recruitment. Written informed consent was obtained from all study participants prior to blood collection. DNA was extracted from the blood samples using the Nucleon BACC Genomic DNA extraction kit (Illumina, Buckinghamshire, United Kingdom). DNA concentration and purity was checked using the NanoDrop® ND-1000 Spectrophotometer and NanoDrop® v3.0.1 software (Inqaba Biotechnology, Pretoria, South Africa). The study adhered to the ethical guidelines as set out in the "Declaration of Helsinki, 2013 (World Medical Association [WMA], 2018).

Genotyping

Genotyping was done using the Illumina MEGA (Illumina, Miami, United States) which contains 1.7 million markers from various ethnicities making it highly suitable for diverse and admixed populations. The array is based on novel variants identified by the Consortium on Asthma among African ancestry populations in the Americas (CAAPA), the Illumina human core content for European and Asian populations as well as multi-ethnic exome content from African, Asian and European populations. The array also contains ancestry informative markers specific to the SAC population. While the KhoeSan population is not highly represented on the array, which could lead to a certain level of ascertainment bias, at the time of genotyping it was the most suitable platform for this diverse and admixed populations. Genome studio v2.04 (Illumina, Miami, United States) was used for SNP calling to calculate intensity scores and to call common variants (MAF $\geq$ 5%), followed by analysis with zCall to recall rare genotypes (MAF < 5%) (Goldstein et al., 2012).

Genotyping Quality Control

Quality control (QC) of the genotyping data was done using the XWAS version 2.0 software and QC pipeline to filter out low quality samples and SNPs (Chang et al., 2014; Gao et al., 2015). Data were screened for sex concordance, relatedness (up to third degree of relatedness) and population stratification (as determined by principal component analysis). Genotypes for males and females were filtered separately in order to maintain inherent differences between the sexes. SNPs were removed from the analysis if missingness correlated with phenotype (threshold = 0.01) as well as individual and SNP missingness (greater than 10%), MAF (less than 1%) and Hardy–Weinberg equilibrium (HWE) in controls (threshold = 0.01). Filtering continued iteratively until no additional variants or individuals were removed. Overlapping markers between the sexes were merged into a single dataset. X chromosome genotypes were extracted and variants were removed if the MAF or missingness was significantly different between the sexes (threshold = 0.01). A flow diagram explaining quality control steps and association testing of the data is shown in **Supplementary Figure S1**.

Admixture

The SAC population is a 5-way admixed population with ancestral contributions from Bantu-speaking African populations, KhoeSan, Europeans, and South and East Asians (Chimusa et al., 2013; Salie et al., 2015). To avoid confounding during association testing the ancestral components are included as covariates (Daya et al., 2014a). Admixture was estimated for the autosome (chromosomes 1–22) and the X chromosome

separately using the software ADMIXTURE (v1.3) (Alexander et al., 2009) and reference genotyping data for 5 ancestral populations. The reference populations used to infer ancestry were European (CEU) and South Asian (Gujarati Indians in Houston, Texas and Pathan of Punjab) extracted from the 1000 Genomes Phase 3 data (Sudmant et al., 2015), East Asian (Han Chinese in Beijing, China), African (Luhya in Webuye, Kenya, Bantu-speaking African, Yoruba from Nigeria) and San (Nama/Khomani) (Uren et al., 2016; Martin et al., 2017). Due to the limited number of individuals available for each reference population the SAC data had to be divided into 21 groups to equal the number of individuals per reference population. The number of individuals per reference population and admixed population has to be kept consistent in order to maximize the accuracy of the admixture results by not over-representing one particular population in the analysis. Therefore admixture inference was done separately for each of the 21 SAC groups, referred to as running groups. Each running group was analyzed five times at different random seed values. The results for each individual were averaged across the five runs in order to obtain the most accurate ancestry estimations [Shringarpure et al., 2016]. Four ancestral components (African, San, European, and South Asian (Salie et al., 2015)] were included as covariates in the logistic regression association testing with the smallest component (East Asian) excluded in order to avoid complete separation of the data.

Association Analysis

SNP Based Association Analysis

Autosomal TB association testing was done with sex-stratified and combined datasets using the additive model in PLINK (version 1.7[1]) (Purcell et al., 2007) in order to detect sex-based differences. TB association testing for the X chromosome were done separately in males and females using XWAS (version 2) and the results were combined using Stouffers method in order to obtain a combined association statistic (Chang et al., 2014; Gao et al., 2015). A sex-differentiated test was conducted for the X chromosome using the XWAS software to test for significant differences in genetic effects between males and females. SNP based association testing (sex-stratified or not) compares the frequency of alleles between cases and controls to determine if a specific allele co-occurs with a phenotype (TB) more often than would be expected by chance. The sex-differentiation test on the other hand compares the effect size (OR) of a variant between the sexes to determine if a variant has a different effect on risk between the sexes. The sex-differentiation test is explained in more detail by Chang et al. (2014). X chromosome inactivation states were also included in the association testing as covariates using a method developed by Wang et al. (2014). To include inactivation states in the association analysis the most likely state was determined for each SNP. A variant can either be inactivated, or it can be skewed toward the deleterious or normal allele or the variant can escape inactivation. To determine which of the four states is most probable the likelihood ratio for each one was calculated and the inactivation state that maximized

the likelihood ratio was applied to the SNP in question. This was done for each variant as inactivation states vary along the X chromosome [for a detailed description see Wang et al. (2014)]. Ancestry, sex and age were included in the analyses as covariates where applicable. Information on other risk factors known to influence TB susceptibility such as smoking and alcohol consumption was not available for this study cohort and could not be included as covariates. Multiple testing correction was done using the SimpleM method (Gao et al., 2010), which adjusts the significance threshold based on the number of SNPs that explains 95% of the variance in the study cohort. This method is less conservative than Bonferroni correction and is a close approximation of permutation results in a fraction of the time. For the autosome the genome-wide significance threshold was set to $5.0e^{-8}$ (Panagiotou and Ioannidis, 2012).

Gene Based Association Analysis

Gene-based association testing groups SNPs together and thus decreases the multiple testing burden and increase power to detect an association. Gene-based association testing was done using the XWAS v2 scripts, which were implemented using the Python[2] (version 2.7.10) and R programming environment [version 3.2.4, (R Development Core Team, 2013)] and R packages corpcor and mvtnorm. Reference files for the known canonical genes on the X chromosome for human genome build 37 were included in the XWAS v2 software package and used to group variants and p-values by gene (Chang et al., 2014; Gao et al., 2015). Bonferroni correction was used to adjust for multiple testing instead of SimpleM, as all genes, unlike SNPs, are independent of each other in the context of association testing and as such the multiple test correction cannot be less than the number of genes tested.

Interaction Analysis

Genome-wide SNP interaction analysis was done using CASSI[3] (v2.51). A joint effects model was implemented for a rapid overview of interactions of all variants across the genome (autosome and X chromosome). Variants from significant interactions were reanalyzed using a logistic regression approach with covariate correction, which would not be feasible for a genome-wide interaction analysis as it would be too computationally intensive. As there is no general consensus on the significance threshold for genome wide interaction analysis Bonferroni correction was used in order to avoid potential inflation of false positive results.

RESULTS

Cohort Summary

In total 410 TB cases and 405 healthy controls passed the sex-stratified QC procedure. General summary statistics for the cohort, including mean and standard deviation of age and global ancestry as well as the ratio of males to females in both cases and

---

[1]http://zzz.bwh.harvard.edu/plink/

[2]http://www.python.org
[3]https://www.staff.ncl.ac.uk/richard.howey/cassi/using.html

TABLE 1 | South African colored (SAC) sample characteristics showing case/control and sex distribution, mean and standard deviation of age and global ancestral components.

Group	Number	Female (%)	Age	San	African	European	South Asian	East Asian
TB cases	410	242 (59)	36.32 ± 11.04	33.89 ± 18.83	29.11 ± 19.80	17.10 ± 16.81	12.95 ± 10.82	7.08 ± 7.22
Controls	405	223 (55)	30.55 ± 12.91	33.75 ± 19.59	29.92 ± 20.46	16.12 ± 15.81	13.26 ± 12.38	7.04 ± 7.44

controls are shown in **Table 1**. Clear differences were observed between TB cases and controls for both age and ancestry, justifying the inclusion as covariates. Ancestral distributions were compared using the Wilcoxon signed-rank test and were shown to significantly differ (unpublished results) between the autosome and X chromosome (**Figure 1**). Y chromosome and mitochondrial haplogroup analysis also revealed strong sex biased admixture in the SAC population, with a strong female KhoeSan and male Bantu-speaking African and European bias (Quintana-Murci et al., 2010). As sex biased ancestry has been

shown to reflect in strong differences between the autosomal and X chromosome ancestral components they were included as covariates in the respective analyses (Wang et al., 2008; Bryc et al., 2010a,b).

SNP Based

The top results for the autosomal association testing are shown in **Table 2** and **Supplementary Figure S2**, with the QQ-plot indicating no constraints on the analysis or inflation of the results (**Supplementary Figure S2**). Following multiple test correction,

FIGURE 1 | Ancestral distribution on the X chromosome and autosome for males and females.
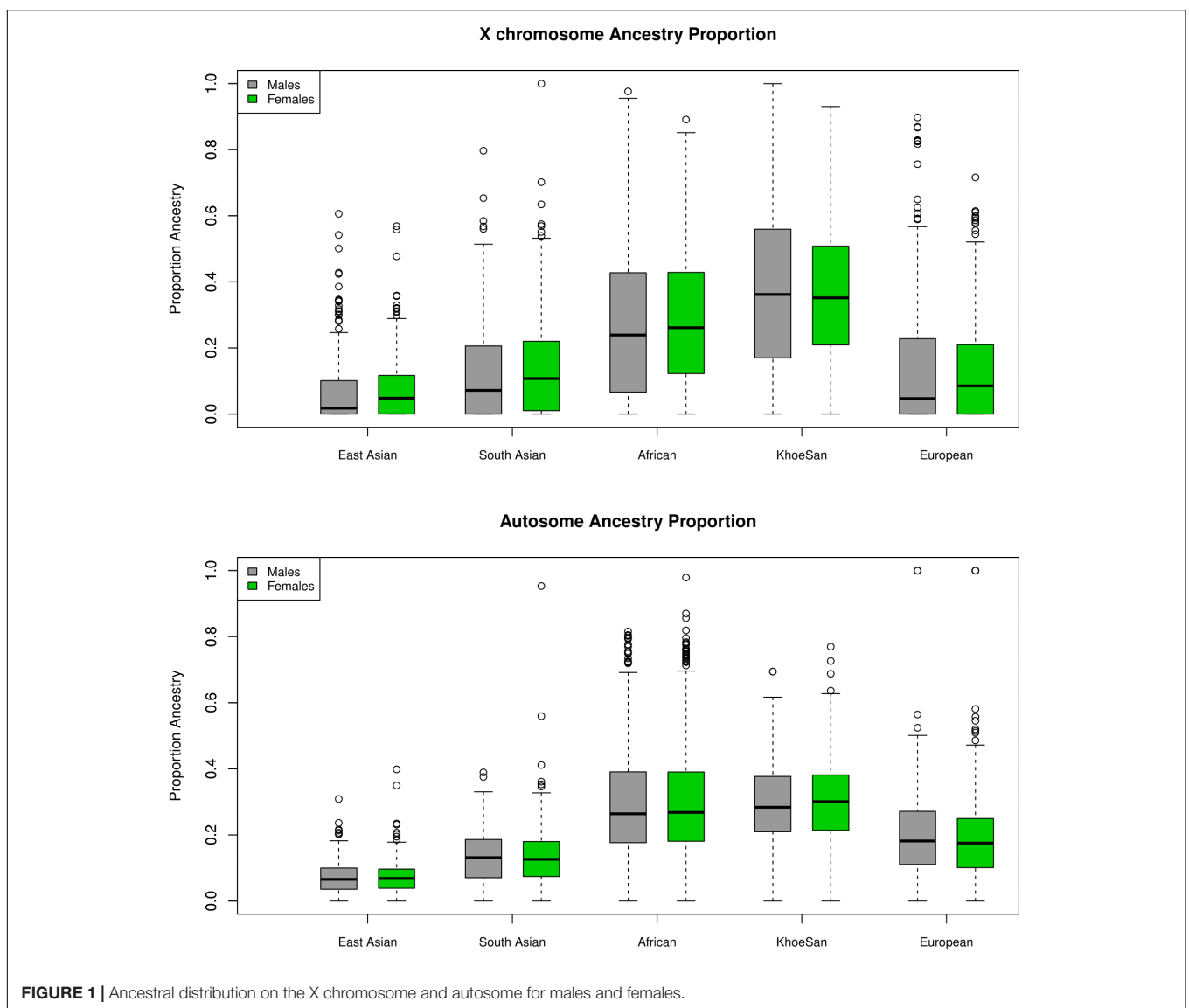
no significant associations were identified for the combined or sex-stratified analysis, but it is important to note that the top associations differed between the sex-stratified and combined analyses as well as for males and females (**Table 2**). The most significant variant for the combined autosomal association test was rs17410035 (OR = 0.4, p-value = $1.5e^{-6}$, **Table 2**), located in the $3'$-UTR of the *DROSHA* gene, which encodes a type 3 RNase. This RNase is involved in miRNA processing and miRNA biogenesis (Mullany et al., 2016). Although little evidence exists that rs17410035 has an impact on *DROSHA* gene expression or miRNA biogenesis (which could affect gene expression) it has been associated with increased colon cancer (OR = 1.22, p-value = 0.014) (Mullany et al., 2016) and cancer of the head and neck (OR = 2.28, p-value = 0.016) (Zhang et al., 2010). When the rs17410035 SNP interacts with other variants (rs3792830 and rs3732360) it can further increase the risk for cancer of the head and neck (Zhang et al., 2010), which illustrates the importance of doing interaction analysis. For the autosomal sex-stratified analysis the variant with the lowest p-value in males was rs11960504 (OR = 2.8, p-value = $7.21e^{-6}$, **Table 2**) located downstream of the *GRAMD2B* gene, a gene for which no information is available. The top hit in females was rs2894967 (OR = 2.17, p-value = $4.77e^{-6}$) a SNP located upstream of the *TENT4A* gene, a gene coding for a DNA polymerase shown to be involved in DNA repair (Ogami et al., 2013). Closer inspection of the data revealed that the effects between the sexes were in the same direction for all top hits in the combined analysis, whereas all variants identified in the sex-stratified analysis had effects in opposite directions between the sexes, or one sex had no effect, indicating that even on the autosome strong sex specific effects are prominent.

For the X chromosome specific association testing a sex-stratified test was conducted and the results were then combined using Stouffers method, which provided a good fit between expected and observed p-values (QQ-plot **Figure 2**) (Chang et al., 2014; Gao et al., 2015). The simpleM method indicated that of the 20,939 X-linked variants 17,600 explained 95% of the variance in the data resulting in a significance threshold of $2.8e^{-6}$ (0.05/17,600). No statistically significant associations with TB susceptibility were identified in either sex-stratified or the combined analysis (**Table 3** and **Figure 2**). The top hit for the X-linked combined (p-value = $2.62e^{-5}$) and females (OR = 1.83, p-value = $1.06e^{-4}$) only analysis was the same variant, rs768568, located in the *TBL1X* gene. For the males the lowest p-value was rs12011358 (OR = 0.37, p-value = $1.25e^{-4}$), a variant located in the *MTND6P12* gene. Both of these genes have not been previously associated with TB susceptibility and *MTND6P12* is a pseudogene with unknown expression patterns or function. Variants in *TBL1X* have been shown to influence prostate cancer (Park et al., 2016) and central hypothyroidism (Heinen et al., 2016) susceptibility. *TBL1X* is a regulator of nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kB) and is thus involved in the immune system which could impact TB susceptibility.

The method of modeling X chromosome inactivation states, developed by Wang et al. (2014), was also incorporated into the X-linked association testing, but no significant observations were observed. Although the p-values were generally lower than for the Stouffer method, the QQ-plot revealed that including estimations of X chromosome inactivation states inflated the p-values and increased the chance of type 1 errors and these results were therefore discounted (**Supplementary Table S1** and **Supplementary Figure S3**).

The sex differentiation test did not result in any significant associations (**Table 4**) and the variant with the lowest p-value was located in a pseudogene, *RNU6-974P* (p-value = $8.33e^{-5}$). The second lowest p-value was for a variant upstream of the *SRPX* (p-value = $2.18e^{-4}$) gene which has previously been shown to have a tumor suppressor function in prostate carcinomas (Kim et al., 2003). Whether these variants are associated with TB susceptibility or influence sex-bias is unclear, but the vastly opposite effects between the sexes are noteworthy. When comparing the OR for the sex differentiation test it is clear that variants can have major sex specific effects again highlighting the need for sex-stratified analysis (**Table 4**).
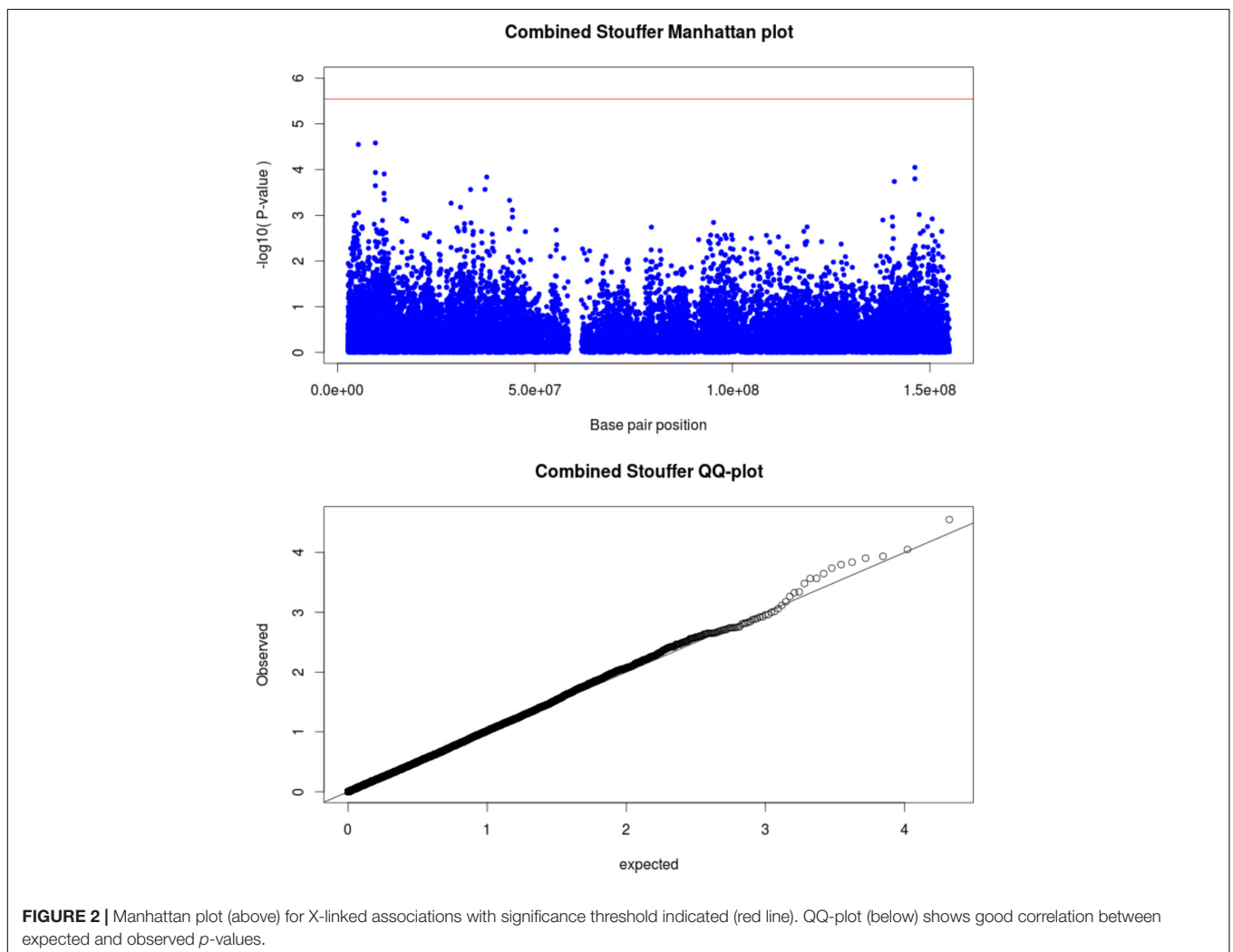
Gene Based

The X chromosome gene-based analysis, in which 1,105 X-linked genes were analyzed did not show any significant associations using a Bonferroni-adjusted significance threshold of $4.5e^{-5}$ (**Table 5**). The association with the lowest p-value for the combined analysis was in the chromosome X open reading frame 51B (*CXorf51B*) (p-value = $1.28e^{-4}$) coding for an uncharacterized protein (LOC100133053). The lowest p-value for males was in an RNA coding region that interacts with Piwi proteins (*DQ590189.1*, p-value = $1.7e^{-3}$), a subfamily of Argonaute proteins. While Piwi proteins are involved in germline stem cell maintenance and meiosis the function of the Piwi interacting RNA molecules are unknown (Girard et al., 2006). For females the top hit was *ARMCX1* (p-value = $6.07e^{-4}$), a tumor suppressor gene involved in cell proliferation and apoptosis of breast cancer cells. While this gene has not been previously implicated in TB susceptibility, *M. tuberculosis* has been shown to affect apoptosis pathways in order to evade the host immune response, suggesting that *ARMCX1* could affect TB susceptibility (Parandhaman and Narayanan, 2014). While not significant the analysis again reveals strong sex specific effects and the sex-stratified and combined analysis gave three different results (**Table 5**).

Interaction Analysis

A genome-wide interaction analysis was performed using the software Cassie. In total 1893973105 interactions were analyzed and following a Bonferroni correction for the number of interactions performed the significance threshold was set to $2.6e^{-11}$. For the joint effects model, 18 interactions passed the significance threshold (**Supplementary Table S2**). The top interaction was between rs1823897, upstream of the *ARSF* gene and rs7064174 in the *FRMPD4* gene (p-value = $7.23e^{-14}$), two genes for which not much information is available and it is unclear how they could be involved in TB susceptibility. The top 450 associations from the joint effects model were then

TABLE 2 | Top associations for the combined and sex-stratified autosomal association testing.

Chr	SNP	A1	Location	Gene	Group	OR	95CI	*P*-value
5	rs17410035	T	5'UTR	*DROSHA*	Combined	0.404	0.28–0.58	$1.50e^{-6}$
5	rs1501847	G	5'UTR	*C5orf64*	Combined	1.708	1.37–2.14	$2.64e^{-6}$
7	rs2665441	C	3'UTR	*ASNS*	Combined	1.681	1.34–2.10	$5.51e^{-6}$
9	rs1662230	G	5'UTR	*RN7SKP120*	Combined	2.278	1.58–3.27	$8.91e^{-6}$
12	rs199911028	G	Intronic	*CFAP54*	Combined	2.966	1.89–4.67	$2.58e^{-6}$
15	rs142644068	C	Intronic	*PCSK6*	Combined	0.132	0.06–0.30	$1.56e^{-6}$
5	rs11960504	T	3'UTR	*GRAMD2B*	Male	2.801	1.79–4.39	$7.21e^{-6}$
13	rs9315991	A	Intronic	*LINC00400*	Male	0.394	0.27–0.58	$2.03e^{-5}$
14	rs8016621	A	Intronic	*SALL2*	Male	0.252	0.14–0.46	$5.91e^{-6}$
5	rs2894967	C	5'UTR	*TENT4A*	Female	2.173	1.56–2.90	$4.77e^{-6}$
9	rs10819610	T	Intronic	*NCS1*	Female	0.514	0.39–0.67	$1.55e^{-6}$
14	rs7152005	T	Intronic	*DPF3*	Female	2.13	1.66–2.90	$1.52e^{-6}$
21	rs2150367	T	Intronic	*LINC02246*	Female	0.502	0.38–0.67	$1.56e^{-6}$

FIGURE 2 | Manhattan plot (above) for X-linked associations with significance threshold indicated (red line). QQ-plot (below) shows good correlation between expected and observed *p*-values.

retested using logistic regression and the same covariates as the SNP based association testing. No significant interactions (threshold of $2.6e^{-11}$) were observed in the logistic regression model (**Table 6**), but as Bonferroni correction is very conservative the top interactions should still be considered as they reach the significance level for SNP based GWAS.

Among the top hits in the logistic regression analysis (**Table 6**) some could impact TB susceptibility as they are involved in

TABLE 3 | Top hits for the X-linked associations, using Stouffers method to combine *p*-values.

Chr	SNP	A1	Location	Gene	Male			Female			P_comb
					OR	95CI	P	OR	95CI	P	
X	rs768568	C	Intron	*TBL1X*	1.69	1.0–2.86	$5.07 e^{-2}$	1.831	1.35–2.49	$1.06 e^{-4}$	$2.62e^{-5}$
X	rs12011358	T	5′UTR	*MTND6P12*	0.371	0.22–0.62	$1.25 e^{-4}$	0.715	0.53–0.96	$2.72 e^{-2}$	$2.82e^{-5}$
X	rs930631	T	3′UTR	*MIR514A1*	0.477	0.29–0.79	$3.66 e^{-3}$	0.664	0.49–0.90	$7.74 e^{-3}$	$8.94e^{-5}$

TABLE 4 | Sex-differentiation analysis.

Chr	SNP	A1	Location	Gene	Male			Female			P_Diff
					OR	95CI	P	OR	95CI	P	
X	rs145407087	C	3′UTR	*RNU6-974P*	0.427	0.12–1.54	0.193	7.147	1.95–26.19	0.003	$8.33e^{-5}$
X	rs5917743	C	5′UTR	*SRPX*	0.300	0.05–1.89	0.203	15.04	1.81–124.7	0.012	$2.18e^{-4}$
X	rs1337567	C	5′UTR	*DIAPH2*	1.45	0.88–2.39	0.146	0.571	0.42–0.78	$3.49e^{-4}$	$6.73e^{-4}$

immune functions. The interaction with the lowest *p*-value was between rs2631914, located upstream of *LINCO2153,* which is upregulated in people with major depressive disorder (Cui et al., 2016), and rs8067702, located downstream of *RTN4RL1*), previously associated with congenital heart disease, microcephaly and mild intellectual disability (Tang et al., 2015). While this interaction is not very informative in the context of TB three other interactions were identified that could impact TB susceptibility (**Table 6**).

The first interaction of interest is between *RNF125* gene (rs35996537) and *URI1* (rs1118924), involved in downregulation of CD^{4+}/CD^{38-} T-cells and PBMCs in HIV-1 positive individuals and NF-kB/CSN2/Snail pathway, activated by TNFα, respectively (Shoji-Kawata et al., 2007; Zhou et al., 2017). Second the interaction between rs386560079 (*ATP2C1),* which is involved in regulation of intracellular Ca^{2+}/Mn^{2+} concentrations through the Golgi apparatus (Deng and Xiao, 2017) and rs6498130 (*CIITA*). Variants in the *CIITA* gene reduce the expression of *MHC class II* proteins and receptors resulting in an immune privilege phenotype (Mottok et al., 2015). The final interaction of interest is between rs12286374 (*NTM*), which is mainly expressed in the brain and promotes neurite outgrowth and adhesion (Maruani et al., 2015) and rs2040739 (*RNF126*) a ring type E3 ligase involved in the Protein B kinase pathway which has been previously implicated in glucose metabolism, apoptosis, cell proliferation and transcription (Song et al., 2005). While none of these genes have previously been implicated in TB susceptibility the fact that some of them are involved in immune functions suggests a role in TB susceptibility.

DISCUSSION

In this GWAS we investigated TB susceptibility in the admixed SAC population, with a specific focus on sex-bias and the X chromosome. A sex-stratified QC protocol was applied to the data in order to conserve inherent differences between the sexes and all statistical analysis were conducted in a sex-stratified and

TABLE 5 | X chromosome gene-based association results.

Chr	Gene	Group	*P*-value
X	*CXorf51B*	Combined	$1.28e^{-4}$
X	*DQ580189.1*	Male	$1.7e^{-3}$
X	*ARMCX1*	Female	$6.07e^{-4}$

combined dataset in order to fully assess the impact of sex on TB susceptibility and the male sex-bias it presents with. We found no significant associations on the autosome or X chromosome for both the sex-stratified and combined SNP and gene-based association testing. A few significant interactions were identified, but the impact of these on TB susceptibility is unclear and will require further investigation to validate and functionally verify.

For the combined autosomal SNP based association testing the only potential variant of interest is rs17410035 located in the *DROSHA* gene (**Table 2**) which is potentially involved in miRNA biogenesis and could impact TB susceptibility if immune related regulatory miRNA is affected. For the X-linked association testing the top association in males was in an uninformative pseudogene, while the female and combined analysis revealed the same variant, rs768568 located in the *TBL1X* gene (**Table 3**). The TBL1X protein has been shown to be a co-activator of NF-kB mediated transcription of cytokine coding genes, but the mechanism of activation is unclear (Park et al., 2016). NF-kB is a vital component of the proinflammatory signaling pathway and is involved in multiple immune pathways including TLRs (Lawrence, 2009), which have previously been shown to influence TB susceptibility (Schurz et al., 2015). Based on this one could extrapolate that variants in the *TBL1X* gene could affect activation and proinflammatory signaling of NF-kB, which could have a direct effect on the immune system and thus TB susceptibility. The direction of effect for this variant was the same in males and females (**Table 3**), but was less significant in males probably due to loss of power when analyzing haploid genotypes. For the variants identified in the sex differentiated analysis it is unclear how they could influence TB susceptibility

TABLE 6 | Logistic regression interaction analysis with covariate adjustment.

Chr1	SNP1	Location	Gene1	Chr2	SNP2	Location	Gene2	P-value
8	rs2631914	5′UTR	LINC02153	17	rs8067702	3′UTR	RTN4RL1	$1.73e^{-9}$
2	rs6756958	Intronic	GALNT5	4	rs201376793	5′UTR	RNU6ATAC13P	$1.92e^{-9}$
4	rs882773	5′UTR	HMX1	18	rs9303903	3′UTR	METTL4	$5.76e^{-9}$
12	rs1798087	5′UTR	TSPAN1	13	rs2091337	3′UTR	LOC105370290	$5.8e^{-9}$
1	rs7517749	Intronic	RGS7	23	rs5907910	3′UTR	SPANXA2-OT1	$1.14e^{-8}$
7	rs757808	5′UTR	KIAA0087	8	rs12676973	3′UTR	FUT10	$1.77e^{-8}$
4	rs1919904	5′UTR	TMPRSS11E	11	rs10769029	Intronic	ALX4	$1.95e^{-8}$
5	rs10040477	5′UTR	LINC02148	12	rs1918193	Intronic	SYT1	$2.82e^{-8}$
1	rs6694239	5′UTR	TNR	2	rs985256	Intronic	SPATS2L	$3.22e^{-8}$
12	rs7975477	Intronic	MGAT4C	20	rs6123951	Intronic	PHACTR3	$3.29e^{-8}$
18	rs35996537	3′UTR	RNF125	19	rs1118924	Intronic	URI1	$3.82e^{-8}$
5	rs10040477	5′UTR	LINC02148	12	rs1918195	Intronic	SYT1	$3.87e^{-8}$
3	rs386560079	Intron	ATP2C1	16	rs6498130	Intronic	CIITA	$3.94e^{-8}$
11	rs4237591	3′UTR	CNTN5	14	rs11850085	Intronic	SLC8A3	$4.54e^{-8}$
1	rs1411276	Intronic	TGFBR3	4	rs1972127	Intronic	PRKG2	$4.86e^{-8}$
12	rs7962106	5′UTR	AVPR1A	18	rs200219001	Intronic	LDLRAD4	$4.94e^{-8}$
14	rs242402	Intronic	PELI2	19	rs2459744	5′UTR	SBK3	$5.22e^{-8}$
11	rs12286374	5′UTR	NTM	19	rs2040739	Intronic	RNF126	$5.49e^{-8}$

as the top hit is located in a pseudogene. However, the sex differentiated test did reveal just how big the difference in effects can be between the sexes for a specific variant (**Table 4**). If these variants with opposite effects are not analyzed in a sex-stratified way then the effects would cancel each other out and any information on sex specific effects would be lost. The X-linked gene-based association test revealed no significant associations despite having more power than the SNP based association testing. A possible reason for this could be that Bonferroni correction was used and as this is very conservative possible associations could have been missed. When looking at the most significant associations (**Table 5**) however, it is unclear how the identified genes could be implicated in TB susceptibility.

The joint effects interaction analysis revealed several significant interactions, but as association results have been previously shown to be severely influenced by admixture (Daya et al., 2014b) only the results for the logistic regression analysis will be discussed here. A few variants were identified in the logistic interaction analysis that could impact TB susceptibility (**Table 6**). URI1 (rs1118924) is activated by TNFα and is involved in the NF-kB/CSN2/Snail pathway, CIITA (rs6498130) impacts expression of MHC class II proteins and receptors and rs35996537 (RNF125) and rs2040739 (RNF126) are both E3 ubiquitin ligase proteins which affect a multitude of cellular functions, such as apoptosis (Song et al., 2005) and protein degradation (Shin et al., 2015). NF-kB, TNFα, MHC class II, E3 ligases, apoptosis and T-cells have all been implicated in TB susceptibility and could collectively contribute by influencing the immune response (Hirsch et al., 1999, 2005; Torres et al., 2006; Fallahi-Sichani et al., 2012; Bai et al., 2013; Parandhaman and Narayanan, 2014; Shin et al., 2015; Franco et al., 2017). As TB is a complex disease all potential influential factors need to be considered and as such the interaction

analysis cannot be ignored. Shortcomings of the interaction analysis are that they are very computationally intensive and suffer from a massive multiple test correction burden. Future research should thus focus on ways to prioritize variants for interaction analysis to decrease computation time as well as have sufficient sample size to minimize multiple test correction burden.

A previous GWAS in the SAC population found a significant association with TB susceptibility in the WT1 gene (rs2057178, OR = 0.62, p-value = $2.71e^{-6}$) (Chimusa et al., 2014). This association did not reach genome-wide significance in our study (OR = 0.75, p-value = 0.049). At the time of the GWAS by Chimusa et al., 2014) there were few African and KhoeSan (only 6 KhoeSan) individuals in the reference data used for imputation and the accuracy of imputation in this population was not known. As the identified variant (rs2057178) was imputed into the data it should have been validated in the SAC population using an appropriate genotyping approach. Secondly although the variant reached a significance threshold for the number of variants tested it did not reach genome wide significance threshold of $5.0e^{-8}$ (Panagiotou and Ioannidis, 2012). Finally, the GWAS performed by Chimusa et al. (2014) only contained 91 control individuals compared to 642 cases, which could affect the power of the study. Chimusa et al. (2014) were unable to replicate previous associations identified in the X-linked TLR8 gene (Davila et al., 2008). The two TLR8 variants in our data, rs3764880 (OR = 1.73, p-value = $3.1e^{-4}$) and rs3761624 (OR = 1.70, p-value = $3.94e^{-4}$) also did not show significant associations. While the haploid genotypes in males contributes to this, a second influential factor could be admixture. Chimusa et al. (2014) did not perform X chromosome specific admixture analysis, which could affect association testing of X-linked genes. Furthermore, only six KhoeSan reference individuals were available, which could affect the accuracy of admixture inference and severely

affect the results. For our study 307 KhoeSan individuals were available, improving the admixture inference and could explain why stronger effects (higher OR) were detected for the *TLR8* variants when compared to Chimusa et al. (2014). It is also important to note that using global ancestry components as covariates does not correct for ancestry at any specific locus and as a result each locus in this population could have up to five different ancestries. This could greatly reduce power and contribute to the lack of replication between studies. In order to address this future studies could incorporate local ancestry inference into the analysis in order to determine the number of ancestries at a locus of interest. Other candidate genes identified in previous GWAS studies were also separately analyzed here, but associations did not replicate (Online **Supplementary Data Sheet S2**).

We did not find any significant associations with TB susceptibility, but highlight the need for sex-stratified analysis. Closer inspection of the data revealed that a large number of SNPs with opposite direction of effects for not only the X chromosome, but the autosome too. Sex specific effects has previously been reported for autosomal variants associated with pulmonary function in asthma (Berhane et al., 2000). In the SAC population these opposite effects have previously been observed for X-linked variants in the *TLR8* gene (Daya et al., 2013) and the same is observed in this study. Sex-stratified analysis should therefore be included in association studies and incorporated in the study design. This can be done by keeping the male to female ratio balanced in the cases and controls. It would also be prudent to do the power calculation for the males and females separately. This will ensure sufficient power for sex-stratified analysis and could elucidate informative sex specific effects. This study was done in a 5-way admixed population. As was observed for the interaction analysis including admixture components significantly changes the association results. Furthermore it was observed (unpublished results) that the ancestral distribution between the X chromosome and autosome are different (**Figure 1**), which is an indication of sex-biased admixture (Goldberg and Rosenberg, 2015; Shringarpure et al., 2016) and highlights the importance of including X chromosome admixture components for X-linked and sex-bias analysis. It is important to note here that the ancestral components in the SAC present with a very wide range (**Figure 1**) and all this variability could affect the power of association studies. It is therefore desirable to increase the sample size when analyzing admixed individuals. Alternatively, a meta-analysis can be conducted, including data from all five ancestral populations, or local ancestry inference could be included in the analysis.

CONCLUSION

While no significant associations were identified this study shows the importance of conducting sex-stratified analysis. This analysis should be incorporated during the study design phase to ensure sufficient power and allow the inclusion of covariates with sex specific effects (in this case admixture components).

The sex-stratified analysis revealed that the effect of certain variants can differ between males and females, not only for the X chromosome but also for the autosome. TB is a complex disease with most genetic associations that do not replicate across different populations, which complicates the elucidation of the genetic impact on disease susceptibility. By including sex-stratified analysis and identifying sex specific effects and the cause for the male bias we can adjust treatment according to sex and potentially improve treatment outcome and survival.

DATA AVAILABILITY

The summary statistics from the case-control cohort will be made available to researchers on request, while access to the raw data will only be available to researchers who meet the criteria for access to confidential data after application to the Health Research Ethics Committee of Stellenbosch University. Requests can be sent to: Dr. Marlo Möller, E-mail: marlom@sun.ac.za.

AUTHOR CONTRIBUTIONS

HS, MM, CK, and GT conceived the idea for this study. CG, GW, and BH did the calling and QC of the raw genotyping data. HS did the analysis and wrote first draft. BH assisted with admixture analysis. All authors contributed to writing and proofreading for approval of the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00678/full#supplementary-material

REFERENCES

Abramowitz, L. K., Olivier-Van Stichelen, S., and Hanover, J. A. (2014). Chromosome imbalance as a driver of sex disparity in disease. *J. Genomics* 2, 77–88. doi: 10.7150/jgen.8123

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Bai, X., Feldman, N. E., Chmura, K., Ovrutsky, A. R., Su, W.-L., Griffin, L., et al. (2013). Inhibition of nuclear factor-kappa b activation decreases survival of mycobacterium tuberculosis in human macrophages. *PLoS One* 8:e61925. doi: 10.1371/journal.pone.0061925

Bellamy, R., Beyers, N., McAdam, K. P., Ruwende, C., Gie, R., Samaai, P., et al. (2000). Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. *Proc. Natl. Acad. Sci. U.S.A.* 5, 8005–8009. doi: 10.1073/pnas.140201897

Berhane, K., McCONNELL, R., Gilliland, F., Islam, T., James Gauderman, W., Avol, E., et al. (2000). Sex-specific effects of asthma on pulmonary function in children. *Am. J. Respir. Crit. Care Med.* 162, 1723–1730. doi: 10.1164/ajrccm.162.5.2001116

Brooks, W. H. (2010). X chromosome inactivation and autoimmunity. *Clin. Rev. Allergy Immunol.* 39, 20–29. doi: 10.1007/s12016-009-8167-5

Bryc, K., Auton, A., Nelson, M. R., Oksenberg, J. R., Hauser, S. L., Williams, S., et al. (2010a). Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc. Natl. Acad. Sci. U.S.A.* 12, 786–791. doi: 10.1073/pnas.0909559107

Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., et al. (2010b). Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc. Natl. Acad. Sci. U.S.A.* 107(Suppl. 2), 8954–8961. doi: 10.1073/pnas.0914618107

Bukhari, M., Aslam, M. A., Khan, A., Iram, Q., Akbar, A., Naz, A. G., et al. (2015). TLR8 gene polymorphism and association in bacterial load in southern Punjab of Pakistan: an association study with pulmonary tuberculosis. *Int. J. Immunogenet.* 42, 46–51. doi: 10.1111/iji.12170

Campbell, M. C., and Tishkoff, S. A. (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Hum. Genet.* 9, 403–433. doi: 10.1146/annurev.genom.9.081307.164258

Chang, D., Gao, F., Slavney, A., Ma, L., Waldman, Y. Y., Sams, A. J., et al. (2014). Accounting for eXentricities: analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. *PLoS One* 9:e113684. doi: 10.1371/journal.pone.0113684

Chimusa, E. R., Daya, M., Möller, M., Ramesar, R., Henn, B. M., van Helden, P. D., et al. (2013). Determining ancestry proportions in complex admixture scenarios in South Africa using a novel proxy ancestry selection method. *PLoS One* 8:e73971. doi: 10.1371/journal.pone.0073971

Chimusa, E. R., Zaitlen, N., Daya, M., Möller, M., Helden, P. D., Mulder, N. J., et al. (2014). Genome-wide association study of ancestry-specific TB risk in the South African coloured population. *Hum. Mol. Genet.* 1, 796–809. doi: 10.1093/hmg/ddt462

Comstock, G. W. (1978). Tuberculosis in twins: a re-analysis of the Prophit survey. *Am. Rev. Respir. Dis.* 117, 621–624.

Cui, X., Sun, X., Niu, W., Kong, L., He, M., Zhong, A., et al. (2016). Long non-coding RNA: potential diagnostic and therapeutic biomarker for major depressive disorder. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* 31, 5240–5248. doi: 10.12659/MSM.899372

Curtis, J., Luo, Y., Zenner, H. L., Cuchet-Lourenço, D., Wu, C., Lo, K., 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. doi: 10.1038/ng.3248

Cutolo, M., Capellino, S., Sulli, A., Serioli, B., Secchi, M. E., Villaggio, B., et al. (2006). Estrogens and autoimmune diseases. *Ann. N. Y. Acad. Sci.* 1089, 538–547. doi: 10.1196/annals.1386.043

Dalgic, N., Tekin, D., Kayaalti, Z., Cakir, E., Soylemezoglu, T., and Sancar, M. (2011). Relationship between toll-like receptor 8 gene polymorphisms and pediatric pulmonary tuberculosis. *Dis. Markers* 1, 33–38. doi: 10.3233/DMA-2011-0800

Davila, S., Hibberd, M. L., Hari, D. R., Wong, H. E., Sahiratmadja, E., Bonnard, C., et al. (2008). Genetic association and expression studies indicate a role of toll-like receptor 8 in pulmonary tuberculosis. *PLoS Genet.* 4:e1000218. doi: 10.1371/journal.pgen.1000218

Daya, M., Merwe, L., Galal, I., Möller, M., Salie, M., Chimusa, E. R., et al. (2013). A panel of ancestry informative markers for the complex five-way admixed south african coloured population. *PLoS One* 8:e82224. doi: 10.1371/journal.pone.0082224

Daya, M., van der Merwe, L., Gignoux, C. R., van Helden, P. D., Möller, M., and Hoal, E. G. (2014a). Using multi-way admixture mapping to elucidate TB susceptibility in the South African Coloured population. *BMC Genomics* 15:1021. doi: 10.1186/1471-2164-15-1021

Daya, M., van der Merwe, L., van Helden, P. D., Möller, M., and Hoal, E. G. (2014b). Investigating the role of gene-gene interactions in TB susceptibility. *PLoS One* 10:e0123970. doi: 10.1371/journal.pone.0123970

de Wit, E., Delport, W., Rugamika, C. E., Meintjes, A., Möller, M., van Helden, P. D., et al. (2010). Genome-wide analysis of the structure of the South African coloured population in the western cape. *Hum. Genet.* 128, 145–153. doi: 10.1007/s00439-010-0836-1

Deng, H., and Xiao, H. (2017). The role of the emphasis type ATP2C1 emphasis gene in hailey disease. *Cell Mol. Life Sci.* 1, 3687–3696. doi: 10.1007/s00018-017-2544-7

Fallahi-Sichani, M., Kirschner, D. E., and Linderman, J. J. (2012). NF-κB signaling dynamics play a key role in infection control in tuberculosis. *Front. Physiol.* 3:170. doi: 10.3389/fphys.2012.00170

Flynn, J. L. (2006). Lessons from experimental mycobacterium tuberculosis infections. *Microbes Infect.* 8, 1179–1188. doi: 10.1016/j.micinf.2005.10.033

Franco, L. H., Nair, V. R., Scharn, C. R., Xavier, R. J., Torrealba, J. R., Shiloh, M. U., et al. (2017). The ubiquitin ligase smurf1 functions in selective autophagy of mycobacterium tuberculosis and anti-tuberculous host defense. *Cell Host Microbe* 13, 421–423. doi: 10.1016/j.chom.2017.08.005

Gallant, C. J., Cobat, A., Simkin, L., Black, G. F., Stanley, K., Hughes, J., et al. (2010). Impact of age and sex on mycobacterial immunity in an area of high tuberculosis incidence. *Int. J. Tuberc. Lung Dis.* 1, 952–959.

Gao, F., Chang, D., Biddanda, A., Ma, L., Guo, Y., Zhou, Z., et al. (2015). XWAS: a software toolset for genetic data analysis and association studies of the X chromosome. *J. Hered.* 106, 666–671. doi: 10.1093/jhered/esv059

Gao, X., Becker, L. C., Becker, D. M., Starmer, J. D., and Province, M. A. (2010). Avoiding the high bonferroni penalty in genome-wide association studies. *Genet Epidemiol.* 34, 100–105. doi: 10.1002/gepi.20430

Girard, A., Sachidanandam, R., Hannon, G. J., and Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 13, 199–202. doi: 10.1038/nature04917

Goldberg, A., and Rosenberg, N. A. (2015). Beyond 2/3 and 1/3: the complex signatures of sex-biased admixture on the x chromosome. *Genetics* 201, 263–279. doi: 10.1534/genetics.115.178509

Goldstein, J. I., Crenshaw, A., Carey, J., Grant, G. B., Maguire, J., Fromer, M., et al. (2012). zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* 1, 2543–2545. doi: 10.1093/bioinformatics/bts479

Grant, A. V., Sabri, A., Abid, A., Abderrahmani Rhorfi, I., Benkirane, M., Souhi, H., et al. (2016). A genome-wide association study of pulmonary tuberculosis in Morocco. *Hum. Genet* 135, 299–307. doi: 10.1007/s00439-016-1633-2

Greenwood, C. M., Fujiwara, T. M., Boothroyd, L. J., Miller, M. A., Frappier, D., Fanning, E. A., et al. (2000). Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. *Am. J. Hum. Genet.* 67, 405–416. doi: 10.1086/303012

Hashemi-Shahri, S. M., Taheri, M., Gadari, A., Naderi, M., Bahari, G., and Hashemi, M. (2014). *Association Between TLR8 and TLR9 Gene Polymorphisms and Pulmonary Tuberculosis.* Available at: http://genecelltissue.portal.tools/?page=article&article_id=18316

Heinen, C. A., Losekoot, M., Sun, Y., Watson, P. J., Fairall, L., Joustra, S. D., et al. (2016). Mutations in TBL1X are associated with central hypothyroidism. *J. Clin. Endocrinol. Metab.* 101, 4564–4573. doi: 10.1210/jc.2016-2531

Hirsch, C. S., Johnson, J. L., Okwera, A., Kanost, R. A., Wu, M., Peters, P., et al. (2005). Mechanisms of apoptosis of T-cells in human tuberculosis. *J. Clin. Immunol.* 1, 353–364. doi: 10.1007/s10875-005-4841-4

Hirsch, C. S., Toossi, Z., Vanham, G., Johnson, J. L., Peters, P., Okwera, A., et al. (1999). Apoptosis and T Cell Hyporesponsiveness in Pulmonary Tuberculosis. *J Infect Dis.* 1, 945–953. doi: 10.1086/314667

Jaillon, S., Berthenet, K., and Garlanda, C. (2017). Sexual dimorphism in innate immunity. *Clin. Rev. Allergy Immunol.* doi: 10.1007/s12016-017-8648-x [Epub ahead of print].

Kim, C. J., Shimakage, M., Kushima, R., Mukaisho, K.-I., Shinka, T., Okada, Y., et al. (2003). Down-regulation of drs mRNA in human prostate carcinomas. *Hum. Pathol.* 34, 654–657. doi: 10.1016/S0046-8177(03)00240-5

Klein, S. L., Marriott, I., and Fish, E. N. (2015). Sex-based differences in immune function and responses to vaccination. *Trans. R. Soc. Trop. Med. Hyg.* 109, 9–15. doi: 10.1093/trstmh/tru167

Kritzinger, F. E., den, B. S., Verver, S., Enarson, D. A., Lombard, C. J., Borgdorff, M. W., et al. (2009). No decrease in annual risk of tuberculosis infection in endemic area in Cape Town. *South Africa. TropMedIntHealth.* 14, 136–142. doi: 10.1111/j.1365-3156.2008.02213.x

Lawrence, T. (2009). The nuclear factor NF-κB pathway in inflammation. *Cold Spring Harb. Perspect. Biol.* 1:a001651. doi: 10.1101/cshperspect.a001651

Mahasirimongkol, S., Yanai, H., Mushiroda, T., Promphittayarat, W., Wattanapokayakit, S., Phromjai, J., 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. doi: 10.1038/jhg.2012.35

Martin, A. R., Lin, M., Granka, J. M., Myrick, J. W., Liu, X., Sockell, A., et al. (2017). An unexpectedly complex architecture for skin pigmentation in africans. *Cell* 171, 1340.e14–1353.e14. doi: 10.1016/j.cell.2017.11.015

Maruani, A., Huguet, G., Beggiato, A., ElMaleh, M., Toro, R., Leblond, C. S., et al. (2015). 11q24.2-25 micro-rearrangements in autism spectrum disorders: relation to brain structures. *Am. J. Med. Genet A* 167A, 3019–3030. doi: 10.1002/ajmg.a.37345

Mersha, T. B., Martin, L. J., Biagini Myers, J. M., Kovacic, M. B., He, H., Lindsey, M., et al. (2015). Genomic architecture of asthma differs by sex. *Genomics* 1, 15–22. doi: 10.1016/j.ygeno.2015.03.003

Mottok, A., Woolcock, B., Chan, F. C., Tong, K. M., Chong, L., Farinha, P., et al. (2015). Genomic alterations in CIITA are frequent in primary mediastinal large b cell lymphoma and are associated with diminished MHC class II expression. *Cell Rep.* 17, 1418–1431. doi: 10.1016/j.celrep.2015.10.008

Mullany, L. E., Herrick, J. S., Wolff, R. K., Buas, M. F., and Slattery, M. L. (2016). Impact of polymorphisms in microRNA biogenesis genes on colon cancer risk and microRNA expression levels: a population-based, case-control study. *BMC Med. Genomics* 9:21. doi: 10.1186/s12920-016-0181-x

Ogami, K., Cho, R., and Hoshino, S. (2013). Molecular cloning and characterization of a novel isoform of the non-canonical poly(A) polymerase PAPD7. *Biochem. Biophys. Res. Commun.* 1, 135–140. doi: 10.1016/j.bbrc.2013.01.072

Oki, N. O., Motsinger-Reif, A. A., Antas, P. R., Levy, S., Holland, S. M., and Sterling, T. R. (2011). Novel human genetic variants associated with extrapulmonary tuberculosis: a pilot genome wide association study. *BMC Res. Notes* 31:28. doi: 10.1186/1756-0500-4-28

Omae, Y., Toyo-Oka, L., Yanai, H., Nedsuwan, S., Wattanapokayakit, S., Satproedprai, N., et al. (2017). Pathogen lineage-based genome-wide association study identified CD53 as susceptible locus in tuberculosis. *J. Hum. Genet.* 62, 1015–1022. doi: 10.1038/jhg.2017.82

Pan, H., Yan, B.-S., Rojas, M., Shebzukhov, Y. V., Zhou, H., Kobzik, L., et al. (2005). Ipr1 gene mediates innate immunity to tuberculosis. *Nature* 7, 767–772. doi: 10.1038/nature03419

Panagiotou, O. A., Ioannidis, J. P. A., and Genome-Wide Significance Project (2012). What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int. J. Epidemiol.* 41, 273–286. doi: 10.1093/ije/dyr178

Parandhaman, D. K., and Narayanan, S. (2014). Cell death paradigms in the pathogenesis of Mycobacterium tuberculosis infection. *Front. Cell Infect. Microbiol.* 4:31. doi: 10.3389/fcimb.2014.00031

Park, S.-Y., Na, Y., Lee, M.-H., Seo, J.-S., Lee, Y.-H., Choi, K.-C., et al. (2016). SUMOylation of TBL1 and TBLR1 promotes androgen-independent prostate cancer cell growth. *Oncotarget* 5, 41110–41122. doi: 10.18632/oncotarget.9002

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

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Qi, H., Zhang, Y.-B., Sun, L., Chen, C., Xu, B., Xu, F., et al. (2017). Discovery of susceptibility loci associated with tuberculosis in Han Chinese. *Hum. Mol. Genet.* 1, 4752–4763. doi: 10.1093/hmg/ddx365

Quintana-Murci, L., Harmant, C., Quach, H., Balanovsky, O., Zaporozhchenko, V., Bormans, C., et al. (2010). Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. *Am. J. Hum. Genet.* 9, 611–620. doi: 10.1016/j.ajhg.2010.02.014

R Development Core Team (2013). *R: A Language and Environment for Statistical Computing.* Vienna: R foundation for statistical computing.

Salie, M., Daya, M., Lucas, L. A., Warren, R. M., van der Spuy, G. D., van Helden, P. D., et al. (2015). Association of toll-like receptors with susceptibility to tuberculosis suggests sex-specific effects of TLR8 polymorphisms. *Infect. Genet Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* 34, 221–229. doi: 10.1016/j.meegid.2015.07.004

Schurz, H., Daya, M., Möller, M., Hoal, E. G., and Salie, M. (2015). TLR1, 2, 4, 6 and 9 variants associated with tuberculosis susceptibility: a systematic review and meta-analysis. *PLoS One* 10:e0139711. doi: 10.1371/journal.pone.0139711

Schurz, H., Kinnear, C. J., Gignoux, C. R., Wojcik, G. L., Helden, P. D., van Tromp, G. C., et al. (2018). A sex-stratified genome-wide association study of tuberculosis using a multi-ethnic genotyping array. *bioRxiv* [Preprint]. doi: 10.1101/405571

Shin, Y. J., Park, S. K., Jung, Y. J., Kim, Y. N., Kim, K. S., Park, O. K., et al. (2015). Nanobody-targeted E3-ubiquitin ligase complex degrades nuclear proteins. *Sci. Rep.* 16:14269. doi: 10.1038/srep14269

Shoji-Kawata, S., Zhong, Q., Kameoka, M., Iwabu, Y., Sapsutthipas, S., Luftig, R. B., et al. (2007). The RING finger ubiquitin ligase RNF125/TRAC-1 down-modulates HIV-1 replication in primary human peripheral blood mononuclear cells. *Virology* 10, 191–204. doi: 10.1016/j.virol.2007.06.028

Shringarpure, S. S., Bustamante, C. D., Lange, K., and Alexander, D. H. (2016). Efficient analysis of large datasets and sex bias with admixture. *BMC Bioinformatics* 23:218. doi: 10.1186/s12859-016-1082-x

Sobota, R. S., Stein, C. M., Kodaman, N., Scheinfeldt, L. B., Maro, I., Wieland-Alter, W., et al. (2016). A locus at 5q33.3 confers resistance to tuberculosis in highly susceptible individuals. *Am. J. Hum. Genet.* 3, 514–524. doi: 10.1016/j.ajhg.2016.01.015

Song, G., Ouyang, G., and Bao, S. (2005). The activation of Akt/PKB signaling pathway and cell survival. *J. Cell. Mol. Med.* 9, 59–71. doi: 10.1111/j.1582-4934.2005.tb00337.x

Sorensen, T. I., Nielsen, G. G., Andersen, P. K., and Teasdale, T. W. (1988). Genetic and environmental influences on premature death in adult adoptees. *N. Engl. J. Med.* 318, 727–732. doi: 10.1056/NEJM198803243181202

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., et al. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature* 1, 75–81. doi: 10.1038/nature15394

Tang, M., Yang, Y.-F., Xie, L., Chen, J.-L., Zhang, W.-Z., Wang, J., et al. (2015). Duplication of 10q22.3-q23.3 encompassing BMPR1A and NGR3 associated with congenital heart disease, microcephaly, and mild intellectual disability. *Am. J. Med. Genet. A* 167A, 3174–3179. doi: 10.1002/ajmg.a.37347

Thye, T., Owusu-Dabo, E., Vannberg, F. O., van Crevel, R., Curtis, J., Sahiratmadja, E., et al. (2012). Common variants at 11p13 are associated with susceptibility to tuberculosis. *Nat. Genet.* 44, 257–259. doi: 10.1038/ng.1080

Thye, T., Vannberg, F. O., Wong, S. H., Owusu-Dabo, E., Osei, I., Gyapong, J., et al. (2010). Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat. Genet.* 42, 739–741. doi: 10.1038/ng.639

Torres, M., Ramachandra, L., Rojas, R. E., Bobadilla, K., Thomas, J., Canaday, D. H., et al. (2006). Role of phagosomes and major histocompatibility complex class II (MHC-II) compartment in MHC-II antigen processing of mycobacterium tuberculosis in human macrophages. *Infect. Immun.* 74, 1621–1630. doi: 10.1128/IAI.74.3.1621-1630.2006

Uren, C., Kim, M., Martin, A. R., Bobo, D., Gignoux, C. R., Helden, P. D., et al. (2016). Fine-scale human population structure in southern africa reflects ecogeographic boundaries. *Genetics* 1, 303–314. doi: 10.1534/genetics.116. 187369

Wang, J., Yu, R., and Shete, S. (2014). X-chromosome genetic association test accounting for x-inactivation, skewed x-inactivation, and escape from x-inactivation. *Genet. Epidemiol.* 38, 483–493. doi: 10.1002/gepi.21814

Wang, S., Ray, N., Rojas, W., Parra, M. V., Bedoya, G., Gallo, C., et al. (2008). Geographic patterns of genome admixture in latin american mestizos. *PLoS Genet.* 4:e1000037. doi: 10.1371/journal.pgen.1000037

World Health Organization [WHO] (2017). *WHO | Global Tuberculosis Report 2017*. Available at: http://www.who.int/tb/publications/global_report/en/

World Medical Association [WMA] (2018). *The World Medical Association-WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects*. Available at: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

Zhang, X., Yang, H., Lee, J. J., Kim, E., Lippman, S. M., Khuri, F. R., et al. (2010). MicroRNA-related genetic variations as predictors for risk

of second primary tumor and/or recurrence in patients with early-stage head and neck cancer. *Carcinogenesis* 31, 2118–2123. doi: 10.1093/carcin/bgq177

Zhou, W., Wang, Q., Xu, Y., Jiang, J., Guo, J., Yu, H., et al. (2017). RMP promotes epithelial-mesenchymal transition through NF-κB/CSN2/Snail pathway in hepatocellular carcinoma. *Oncotarget* 20, 40373–40388. doi: 10.18632/oncotarget.16177

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.