



Local Ancestry Adjusted Allelic Association Analysis Robustly Captures Tuberculosis Susceptibility Loci

Yolandi Swart¹, Caitlin Uren^{1,2}, Paul D. van Helden¹, Eileen G. Hoal¹ and Marlo Möller^{1,2*}

¹DSI-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, ²Centre for Bioinformatics and Computational Biology, Stellenbosch University, Stellenbosch, South Africa

OPEN ACCESS

Edited by:

Ranjit Das,
Yenepoya University, India

Reviewed by:

Tesfaye B. Mersha,
Cincinnati Children's Hospital Medical
Center, United States
Mohammed S. Mustak,
Mangalore University, India

*Correspondence:

Marlo Möller
marlom@sun.ac.za

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 28 May 2021

Accepted: 01 October 2021

Published: 15 October 2021

Citation:

Swart Y, Uren C, van Helden PD,
Hoal EG and Möller M (2021) Local
Ancestry Adjusted Allelic Association
Analysis Robustly Captures
Tuberculosis Susceptibility Loci.
Front. Genet. 12:716558.
doi: 10.3389/fgene.2021.716558

Pulmonary tuberculosis (TB), caused by *Mycobacterium tuberculosis*, is a complex disease. The risk of developing active TB is in part determined by host genetic factors. Most genetic studies investigating TB susceptibility fail to replicate association signals particularly across diverse populations. South African populations arose because of multi-wave genetic admixture from the indigenous KhoeSan, Bantu-speaking Africans, Europeans, Southeast Asian and East Asian populations. This has led to complex genetic admixture with heterogeneous patterns of linkage disequilibrium and associated traits. As a result, precise estimation of both global and local ancestry is required to prevent both false positive and false-negative associations. Here, 820 individuals from South Africa were genotyped on the SNP-dense Illumina Multi-Ethnic Genotyping Array (~1.7M SNPs) followed by local and global ancestry inference using RFMix. Local ancestry adjusted allelic association (LAAA) models were utilized owing to the extensive genetic heterogeneity present in this population. Hence, an interaction term, comprising the identification of the minor allele that corresponds to the ancestry present at the specific locus under investigation, was included as a covariate. One SNP (rs28647531) located on chromosome 4q22 was significantly associated with TB susceptibility and displayed a SNP minor allelic effect (G allele, frequency = 0.204) whilst correcting for local ancestry for Bantu-speaking African ancestry (p -value = 5.518×10^{-7} ; OR = 3.065; SE = 0.224). Although no other variants passed the significant threshold, clear differences were observed between the lead variants identified for each ancestry. Furthermore, the LAAA model robustly captured the source of association signals in multi-way admixed individuals from South Africa and allowed the identification of ancestry-specific disease risk alleles associated with TB susceptibility that have previously been missed.

Keywords: South Africa, admixture mapping, TB susceptibility, ancestry-specific risk alleles, local ancestry adjustments, population genetics, host genetics

INTRODUCTION

Pulmonary tuberculosis (TB), caused by the *bacillus Mycobacterium tuberculosis (M.tb)*, is a complex disease which affects populations disproportionately and results from a multifactorial interaction between host and pathogen (Yim and Selvaraj, 2010). It is often said that approximately 5–10% of infected individuals (± 3 billion people worldwide) will go on to develop active TB whilst the majority will remain asymptomatic (Bañuls et al., 2015; El Kamel et al., 2015; Chaw et al., 2020). According to the World Health Organization (WHO), an estimated 10 million TB cases and 1.5 million deaths were reported in 2019 (WHO, 2019). TB therefore remains a global health burden and is of particular concern in low- to middle-income countries where a generally higher incidence rate (615 per 100 000 in South Africa) occurs, together with the limitations of currently available therapies and vaccines (Bao et al., 2016; WHO | Global tuberculosis report 2019, 2019). Numerous genetic and heritability studies have established the role of host genetic factors in susceptibility to TB (Rudko et al., 2016; Kinnear et al., 2017; Cai et al., 2019; Luo et al., 2019), but with minimal overlap between populations from various geographical regions (Thye et al., 2010; Oki et al., 2011; Mahasirimongkol et al., 2012; Png et al., 2012; Thye et al., 2012; Chimusa et al., 2014, 2014; Curtis et al., 2015; Schurz et al., 2015; Grant et al., 2016; Sobota et al., 2016; Uren et al., 2017a; Omae et al., 2017; Qi et al., 2017; Zheng et al., 2018). The variation observed between populations from diverse geographic regions indicates possible ancestry-specific differences that contribute to the host genetic variability observed in TB genome-wide association studies (GWAS) (van Helden et al., 2006; Chimusa et al., 2014; Schurz et al., 2019a; Cai et al., 2019).

Previous investigations into southern African history and population structure elucidated indigenous KhoeSan ancestry in the region, in addition to populations being multi-way admixed due to multiple inter- and intra-continental migrations (de Wit et al., 2010; Quintana-Murci et al., 2010; Uren et al., 2017b). This population history has resulted in admixture from indigenous KhoeSan, Bantu-speaking African, European, Southeast Asian and East Asian populations (de Wit et al., 2010; Quintana-Murci et al., 2010; Uren et al., 2016). Ancestral populations contributed linked alleles (haplotype blocks) resulting in a mosaic of phenotypic consequences. This admixture can be leveraged to identify associations between various TB phenotypes and genomic regions harbouring variants with highly differentiated allele frequencies among ancestral populations, known as admixture mapping (Wang et al., 2020). Hence, the unique and complex admixed individuals from southern Africa, harbouring genomic contributions from ancestral populations with differing historical disease burden, present an opportunity to investigate ancestry-specific disease risk alleles associated with TB susceptibility (Shriner, 2013; Wang et al., 2020).

Previous admixture mapping and association studies investigating TB susceptibility loci in South Africa were restricted by a low number of controls, small reference population sample size and low SNP density (de Wit et al.,

2010; Chimusa et al., 2014; Daya et al., 2014b, 2014a). With the recent adaption of computational algorithms to better suit multi-way admixed populations, a more suitable, high-density genotyping platform and the availability of large scale, population-specific datasets, we aimed to perform an updated scan for variants associated with TB using local ancestry adjusted allelic (LAAA) association models.

MATERIALS AND METHODS

Study Population and Ethics Approval

A total of 413 pulmonary TB cases and 407 healthy controls were recruited from the metropolitan area of Cape Town in the Western Cape Province, South Africa. The population from this area was elected due to the high incidence of TB as well as the equal socio-economic status and low prevalence of HIV at the time of sampling (Rossouw et al., 2003; Möller et al., 2009; Gallant et al., 2010). Furthermore, TB cases and controls were sampled from the same area, therefore socio-economic status is unlikely to be a confounding factor as previously determined by Chimusa et al. (2014). TB cases were distinguished through bacteriological confirmation (culture positive and/or smear positive). Healthy controls had no previous history of TB. However, 80% of individuals above 15 years of age in this area were estimated to have been exposed to *M.tb*, and could therefore be regarded as latently infected (Gallant et al., 2010). If study participants were under the age of 18 or were HIV-positive, they were excluded from the analysis.

Written informed consent was obtained from all study participants before recruitment and blood collection. Sample collection (protocol number 95/072) and this study (S20/02/041) were both approved by the Health Research Ethics Committee of the Faculty of Health Sciences (HREC), Stellenbosch University. The research was conducted according to the principles expressed in the Declaration of Helsinki (2013).

Genotyping, Data Merging and Quality Control

Genotype data on the case-control cohort was generated using the Illumina (Illumina, CA, United States) multi-ethnic genotyping array (MEGA) comprising ~ 1.7 million markers (Schurz et al., 2019b). The Sanger Imputation Server (SIS) (<https://imputation.sanger.ac.uk>) and the African Genome Resource (AGR) reference panel (Gurdasani et al., 2015) was utilised for the imputation of missing genotypes. The imputed data was subjected to iterative quality control as previously described by Schurz et al. (2019b). Thereafter, the data from the admixed individuals were merged with the respective appropriate source populations (summarised in **Table 1**) using PLINK v2.0 (<https://www.cog-genomics.org/plink/2.0/>) (Purcell et al., 2007) in order to generate input files required for global and local ancestry inference.

After merging of admixed and source ancestral populations, all individuals missing more than 10% genotypes were removed, SNPs with more than 3% missing data were excluded and a Hardy-Weinberg equilibrium (HWE) filter was used in controls

TABLE 1 | Ancestral populations included in analysis.

Population	<i>n</i>	Source
European (British)	60	1000G phase 3
African (Luhya)	50	1000G phase 3
East Asian (Chinese)	36	1000G phase 3
KhoeSan (Nama)	44	European Genome-Phenome archive- https://ega-archive.org/
South East Asian (Malay)	40	Wong et al. (2013)

(threshold < 0.01). The data was screened for relatedness using the software KING (Manichaikul et al., 2010) and individuals up to second degree relatedness were subsequently removed. Variants with a minor allele frequency (MAF) below 1% were removed. The final dataset after quality control and data filtering consisted of 392 TB cases and 346 controls in addition to 289 ancestral individuals. A total of 4,249,442 variants passed quality control and filtering parameters.

Global Ancestry Inference

ADMIXTURE was used to investigate the population substructure amongst our cohort, as well as to determine the correct number of contributing ancestries (Alexander and Lange, 2011; Zhou et al., 2011). This is a model-based approach to estimate individual ancestry coefficients of an individual's genome from *k* ancestral populations and corresponding ancestral genotype frequencies through cross validation. For the purpose of computational efficiency, redundant single-nucleotide polymorphisms (SNPs) were removed and only tagging SNPs representative of the genetic haplotype blocks remained. Therefore, each SNP that has a linkage disequilibrium (LD) r^2 of >0.1 within a 50-SNP sliding window (advanced by 10 SNPs at a time) was removed. A total of 261,694 autosomal markers after LD pruning and 820 individuals (413 cases and 407 controls) were used to infer ancestry in an unsupervised manner for $k = 3-10$ (5 iterations). All 820 individuals were grouped into running groups of equal size together with 289 ancestral populations whilst inferring global ancestry proportions. Related individuals were included in separate running groups. Running groups were created to ensure an equal number of reference populations and admixed populations whilst removing relatedness as a confounding factor during global ancestry assignment. After determining the correct *k* number of contributing ancestries through cross validation, the software RFMix was used to infer global ancestry proportions for downstream statistical analysis, since ADMIXTURE is not as accurate as haplotype-based analyses (Uren et al., 2020). The software PONG was used for visualisation of global ancestry proportions and amalgamation of multiple iterations into the major mode (Behr et al., 2016).

Local Ancestry Inference

Local ancestry inference requires phasing of haplotypes prior to inferring local ancestry. The software program SHAPEIT2 (Delaneau et al., 2013; Delaneau and Marchini, 2014) (utilizing the HapMap Genetic map – GRCh37) was used to

phase the merged dataset before inferring local ancestry for each position in the genome using RFMix (Maples et al., 2013). RFMix is 30X faster than other local ancestry inference software and is accurate in multi-way admixture scenarios (Maples et al., 2013; Uren et al., 2020). Default parameters were used, except for the number of generations since admixture, which was set to 15, consistent with previous studies (Uren et al., 2016). Both global and local ancestry was inferred for 1,027 individuals (392 TB cases, 346 controls and 289 ancestral individuals) and 4,249,442 autosomal SNPs.

Statistical Analysis

A Local Ancestry Adjusted Allelic (LAAA) model, first described by Duan et al. (2018), was used to investigate if there are allelic, ancestry-specific or ancestry-specific allelic associations with TB susceptibility in an admixed South African population (Duan et al., 2018). Dosage files were compiled at each locus as a biallelic state and were calculated as 0, 1 or 2 copies of a specific ancestry at any locus along the genome. Separate regression models for each ancestral group were fitted to investigate which ancestral population(s) drive the association between TB status and local ancestry at each locus. Genome-wide admixture proportions obtained from RFMix were included in all regression models to account for population structure. The smallest ancestry proportion (East Asian) was excluded as covariate to avoid complete separation of data. Therefore, four ancestral components (KhoeSan, African, European, and Southeast Asian) were included as covariates in association testing, together with age and gender. The number of alternate alleles (not the reference alleles) were counted, as these are more likely to be ancestry-specific. A total of 738 unrelated individuals (392 TB cases and 346 controls) and 4,249,442 autosomal markers were included in this analysis. The *glm()* function in R was used for logistic regression association testing.

The following four regression models were tested simultaneously to detect the source (allelic, ancestry or both ancestry-allelic effect) of the association signals observed:

1. Global ancestry proportions were included as covariates and thus represents the null model. This test is regularly used in GWAS to investigate whether an additive allelic dose affect exists on the phenotype, not considering local ancestry (Homozygous for the reference allele = 0; Heterozygous = 1; Heterozygous for the alternate allele = 2).
2. Local ancestry expressed in terms of the number of copies of a specific ancestry (Ancestry of interest = 1; Other ancestries = 0) at a locus were included as covariates. This model is often

utilised to conduct admixture mapping studies to elucidate ancestry effects of variants which showcases frequency disparities across ancestral populations (Homozygous for other ancestry = 0; Heterozygous = 1; Heterozygous for ancestry of interest = 2).

3. Minor allelic effects were used in an additive manner and were included as covariates whilst still adjusting for local ancestry. Therefore, jointly testing for model 1 + 2.
4. This model utilises the ancestry-specific minor alleles at a locus, thus the minor alleles together with the corresponding ancestry of the minor allele were included as covariates (Minor allele and ancestry not on the same haplotypes = 0; Minor allele and ancestry are on the same haplotype = 1). This model is an extension to the allelic (3) and local ancestry (2) model by modelling the combination of the minor allele present at a specific locus and the ancestry of the specific allele at that genomic locus. (Both minor allele and ancestry not on the same haplotype = 0; Heterozygote (only one haplotype has both minor allele and ancestry on the same haplotype = 1; Both minor allele and ancestry on the same haplotype = 2).

Since the true underlying causal variants as well as the LD between the marker under study are unknown, modelling all three terms simultaneously is the most effective approach to elucidate causal variants in an admixed cohort with minimal power loss (Duan et al., 2018). Therefore, we can determine if a specific minor allele, ancestry or both a minor allelic and ancestry co-occurs with TB status more often than would be expected by chance.

The development of power and sample size analysis tools for mapping ancestry-specific effects are lacking. The power to detect significant associations depends greatly on the proportion of admixture, differences in effect sizes between diverse ancestries and differences in the allele risk frequencies among ancestral populations. It is noteworthy to highlight that this information will vary for each admixture scenario. Nonetheless, it remains critical to conduct some sort of power calculation to ensure the reliability of elucidating ancestry-specific genomic regions amongst admixed individuals. Hence, we conducted a priori power analysis in order to ensure the reliability of results given our samples size using G*Power (Faul et al., 2007, 2009).

To account for the multiple testing burden, the R package *STEAM* (Significance Threshold Estimation for Admixture Mapping) (Grinde et al., 2019) was used to estimate the genome-wide significance threshold. *STEAM* is specifically designed to estimate genome-wide significance thresholds for admixture mapping studies given the admixture proportions and number of generations since admixture. We quantified the degree of inflation by generating a Quantile-Quantile plot of the residuals.

RESULTS

Global Ancestry Inference

After close inspection of global ancestry proportions generated using ADMIXTURE, the k number of

contributing ancestries was determined to be $k = 5$, since this was the lowest k -value through cross validation (**Supplementary Table S1**). Since haplotype-based admixture software is more accurate at global ancestry inference, ancestry proportions (genome-wide ancestral contributions) were inferred for all individuals using RFMix (Uren et al., 2020). **Figure 1** represents the global ancestry proportions plotted vertically for each admixed individual and contributing ancestral populations using RFMix ($k = 5$). It is evident from the global ancestry inference that the cohort is a complex five-way admixed group, with ancestral contributions from the indigenous KhoeSan (~35–40%), Bantu-speaking Africans (~27–30%), Europeans (~20%), Southeast Asians (~7–8%) and East Asians (~5%). Furthermore, extensive genetic heterogeneity can be observed, since genome-wide proportions differ vastly between individuals.

Local Ancestry Inference

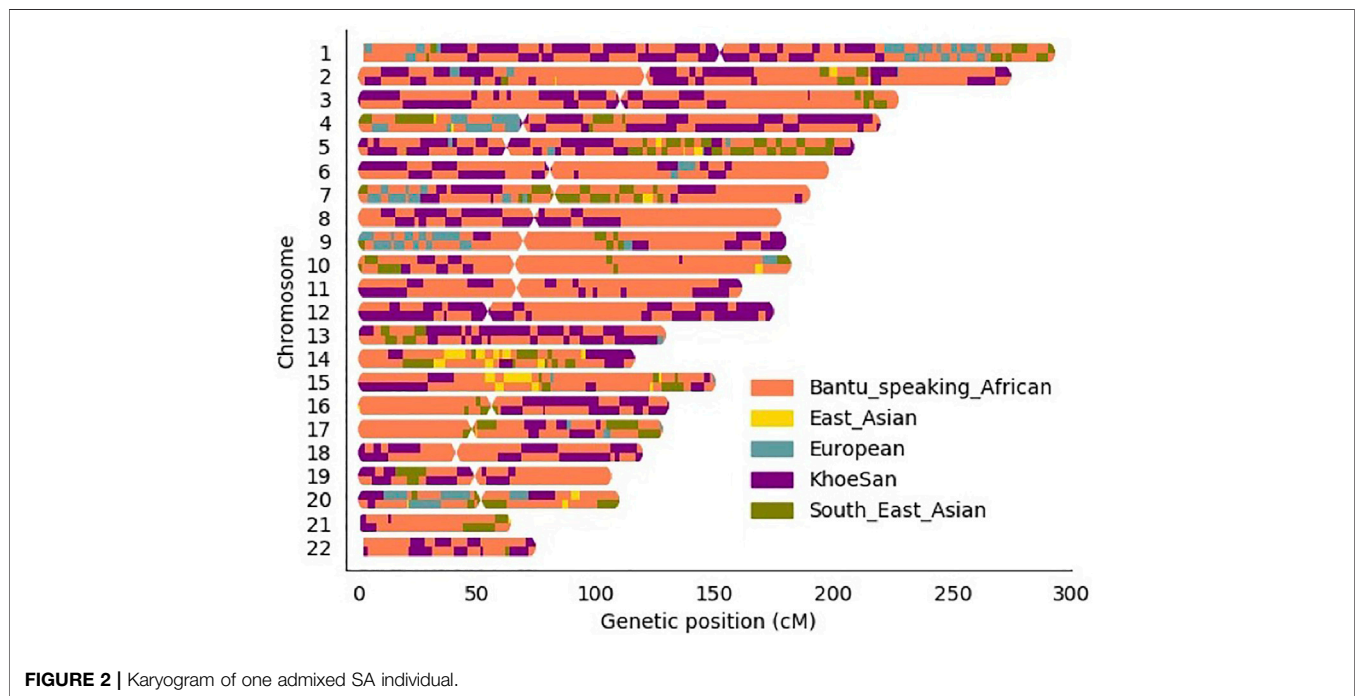
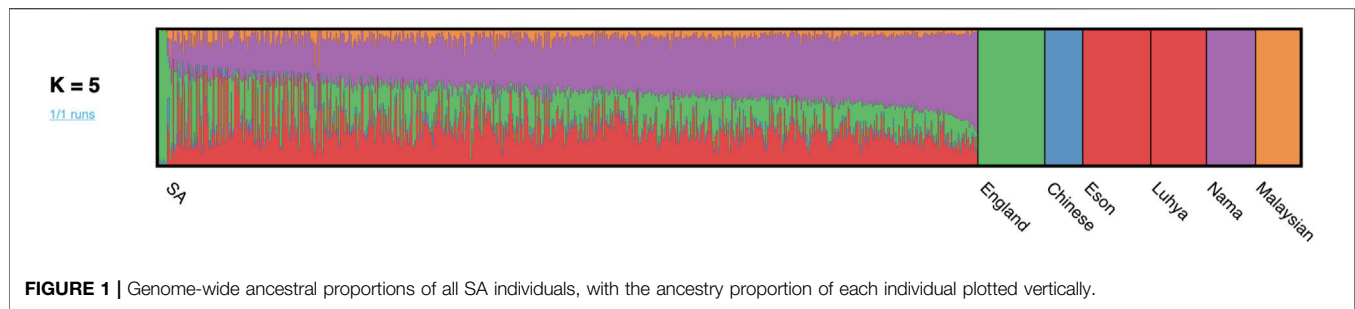
Local ancestry was estimated for all individuals and visually observed with karyograms. As shown in **Figure 2**, admixture between geographically distinct populations creates complicated ancestral-and admixture induced LD blocks. **Figure 2** represents a single five-way admixed individual. Since not all individuals will harbour the same number and length of ancestry segments, it is necessary to accurately infer local ancestry in every individual at each genomic locus.

Local Ancestry Allelic Adjusted Association Analysis

A total of 4,249,442 autosomal markers and 738 unrelated individuals (392 TB cases and 346 controls) were included in logistic regression models to assess whether any loci were significantly associated with TB status (adjusting for gender, age, and global ancestry proportions inferred by RFMix). More information regarding the distribution of age, gender and ancestry proportions of the cohort can be found in the **Supplementary Figures S1–S3** and **Supplementary Table S2**. LAAA models were successfully conducted for all five ancestries present in this highly complex admixed cohort.

One variant (rs28647531) was significantly associated with TB status (p -value $< 1.078 \times 10^{-6}$) due to an allelic SNP effect (G allele; 0.204 frequency) whilst adjusting for Bantu-speaking African local ancestry on chromosome 4 (OR = 3.065, p -value = 5.518×10^{-7}) (**Figure 3**). This variant is an intronic variant with a gene consequence on Follistatin-related protein (*FSTL5*), which is a protein coding gene involved in calcium ion binding. No restrictions on the analysis or inflation of results were observed as indicated by the Quantile-Quantile plot (**Supplementary Figure S4**). Although no other variants passed the significance threshold, multiple lead variants (p -value $< 1 \times 10^{-5}$) were identified. Furthermore, it is clear from our results that multiple distinct lead variants were identified for each ancestry.

The lead variants identified using only the global ancestry as covariates (model 1), are summarised in **Supplementary Table S3**. One lead variant (rs38672118) is near the protein



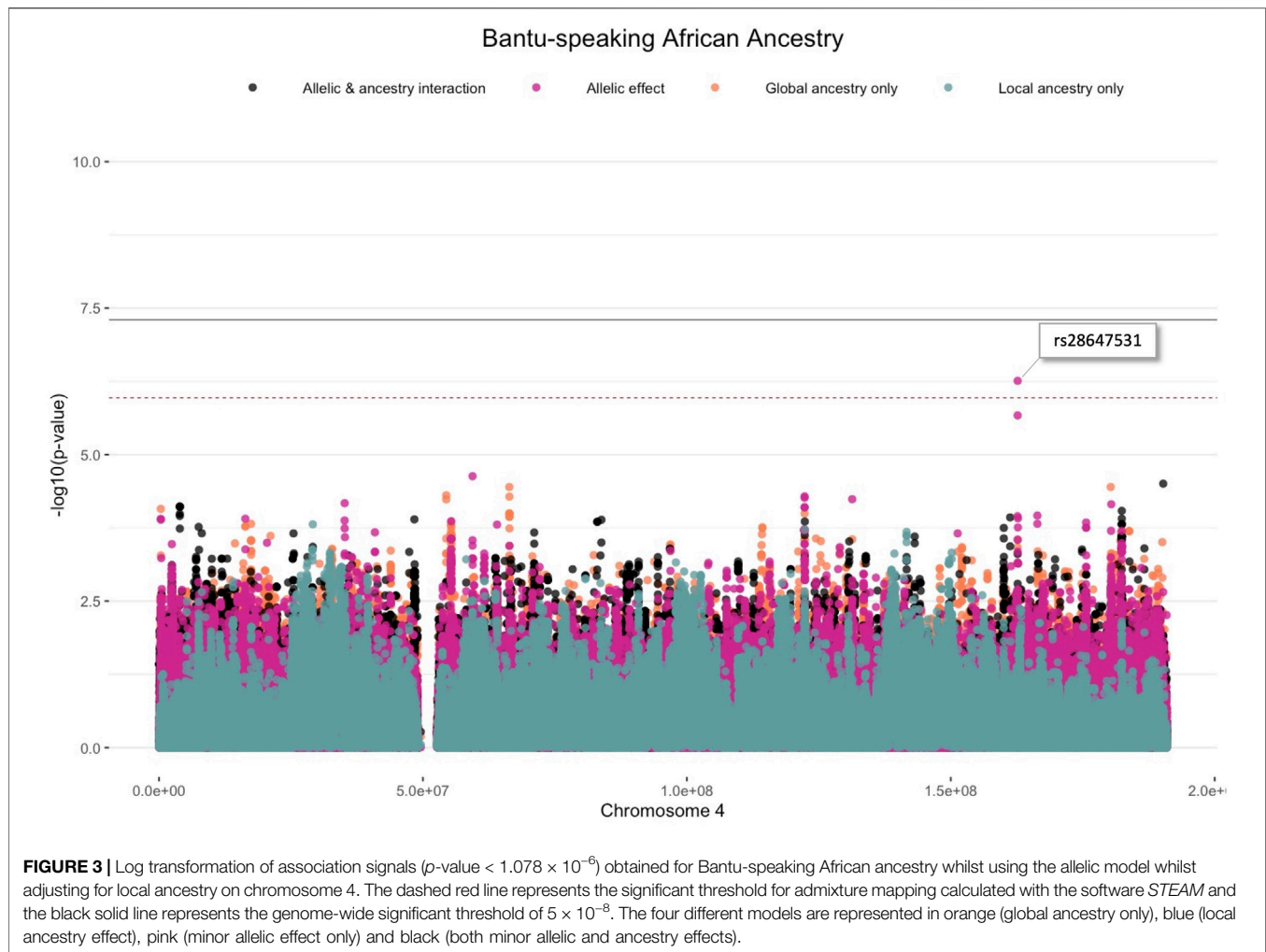
coding gene, *CUL2* (Cullin-2), located on chromosome 10. The lead variants identified by conducting admixture mapping (model 2), are summarised in **Supplementary Table S4**. Only one ancestry (European) identified a local ancestry peak on chromosome 15 (**Supplementary Figure S5**). The lead variants identified utilising the allelic model adjusting for local ancestry (model 3), are summarised in **Table 2**. The lead variants identified by the LAAA model (model 4) are summarised in **Table 3**. It is noteworthy that both the allelic model adjusting for local ancestry (model 3), and the LAAA model (model 4) captured association signals not previously observed for this cohort.

DISCUSSION

We conducted local ancestry allelic adjusted association analysis in a multi-way admixed South African (SA)

population to investigate whether ancestry-specific genetic regions are associated with TB susceptibility. Multi-way admixed populations allow the opportunity to simultaneously assess the association of TB status in multiple continental populations and elucidate possible ancestry-specific effects on TB susceptibility. Previous studies were confounded by the limited number of representative reference populations available to infer local ancestry and the use of the low-density Affymetrix gene chip array (~500k markers) in the analyses. New, more representative ancestral populations and an increase in accuracy of several software tools facilitated the novel findings presented here.

Global ancestry deconvolution suggested a five-way admixed scenario for the study cohort. This is in accordance with previous studies (de Wit et al., 2010; Chimusa et al., 2014; Uren et al., 2016). This diverse admixture and associated regional heterogeneity are reflected in the karyograms generated via



local ancestry inference (Figure 2). This scale of genetic heterogeneity suggests that no two individuals will harbour the same DNA segment from the same ancestral population, i.e., there is a high degree of locus-specific ancestry (Duan et al., 2018). The results presented here highlight that only including global ancestry proportions in the analysis is not sufficient to identify which ancestry is located on distinct chromosomal segments. The only lead variant (rs38672118) identified using the global ancestry-only model is near the protein coding gene, *CUL2*. Although the function of *CUL2* on *M.tb* clearance is still uncertain, *CUL2* forms an important part of the cullin-RING-based E3 ubiquitin-protein ligase complex and subsequently targets the ubiquitination of target proteins (Nguyen et al., 2017). The model used for admixture mapping (only utilising local ancestry) seems overconservative for complex multi-way admixed individuals, since only one admixture peak was close to the significance threshold for European ancestry (located on chromosome 15). This highlights the phenomenon of genetic heterogeneity where the presence of both admixture-induced LD blocks and haplotype LD blocks often results in missed association signals due to tagging SNPs being possibly located in different ancestral LD blocks (Duan et al., 2018).

One example of missing relevant associated variants in complex admixed populations, is the association signal obtained on chromosome 11q13 while adjusting for Bantu-speaking African- and European local ancestry. This lead variant indicated an association with the TIR Domain Containing Adaptor Protein (*TIRAP*) gene (Figure 4) and is involved in the toll-like receptor (TLR) 4 signalling pathway of the immune system via the TIR adaptor protein it codes for. *TIRAP* is a protein which identifies microbial pathogens through TLRs as part of the initial innate immune response (Selvaraj et al., 2010). This acts via *IRAK2* and *TRAF-6*, leading to the activation of NF-kappa-B, MAPK1, MAPK3 and JNK, which is essential for cytokine secretion in order to mount an inflammatory response (Capparelli et al., 2013). Polymorphisms in the *TIRAP* gene were previously identified to be associated with TB susceptibility in a South Indian population (Selvaraj et al., 2010), as well as a Chinese population (Zhang et al., 2011). This suggests a possible role of the *TIRAP* gene in TB susceptibility via activation of TLRs in order to recognize several components of *M.tb* during active TB disease. The T allele of TLR4 (rs4986791) was found to be associated with an increased risk for an Asian subgroup in a meta-analysis

TABLE 2 | Summary statistics of the top results (p -value $< 1 \times 10^{-5}$) whilst utilising the Additive allelic model whilst adjusting for local ancestry.

Chr	Position	rsID	Ref	Alt	Altfreq	OR	SE	p -value	Ancestry	Location	Gene
4	153960368	rs1024148	T	G	0.463	1.589	0.126	9.948e-06	European	None	None
13	24752695	rs7325698	T	C	0.310	1.363	0.150	7.721e-06	European	Intronic	SPATA13
13	24753449	rs2862243	A	G	0.263	1.300	0.174	3.811e-06	European	Intronic	SPATA13
13	50909771	rs67217502	T	C	0.103	1.108	0.225	6.474e-06	European	Intronic	DLEU1
13	50913874	rs12853498	A	T	0.103	1.108	0.225	6.474e-06	European	Intronic	DLEU1
13	50915280	rs17074141	C	T	0.103	1.108	0.225	6.474e-06	European	Intronic	DLEU1
13	50920890	rs17363026	T	C	0.104	1.110	0.225	4.129e-06	European	Intronic	DLEU1
13	50922773	rs17074143	T	A	0.099	1.104	0.230	5.059e-06	European	Intronic	DLEU1
13	50925317	rs34712361	T	C	0.098	1.103	0.230	7.398e-06	European	Intronic	DLEU1
13	50925565	rs67964536	C	T	0.098	1.103	0.230	7.398e-06	European	Intronic	DLEU1
13	50926076	rs79714483	A	G	0.098	1.103	0.230	7.398e-06	European	Intronic	DLEU1
14	48325261	rs447600	T	A	0.459	1.582	0.123	4.350e-06	European	None	None
2	52241352	rs2883609	C	G	0.318	1.374	0.122	9.361e-06	East Asian	nRNA_intronic	AC007682.1
2	180940603	rs13411512	T	C	0.274	1.315	0.130	7.428e-06	East Asian	Intronic	CWC22
12	9388842	ss1388098326	C	T	0.124	1.132	0.176	9.961e-06	East Asian	nRNA_intronic	A2MP1
14	48325261	rs447600	T	A	0.459	1.582	0.107	5.988e-06	East Asian	Intergenic	RP11-476J6.1
22	46046477	rs134850	A	G	0.223	1.250	0.131	7.042e-06	East Asian	Intergenic	ATXN10
1	151185502	rs4971014	A	G	0.187	1.206	0.150	3.279e-06	SouthEast Asian	Intronic	PIP5K1A
2	52241352	rs2883609	C	G	0.318	1.374	0.126	9.022e-06	SouthEast Asian	ncRNA_intronic	AC007682.1
2	180940603	rs13411512	T	C	0.274	1.315	0.132	5.645e-06	SouthEast Asian	Intergenic	CWC22
8	126754436	rs12547413	T	C	0.135	1.144	0.170	8.440e-06	SouthEast Asian	Intronic	CLU2
10	35527543	rs3867218	C	T	0.513	1.670	0.116	3.194e-06	SouthEast Asian	Intergenic	RNU6794P
14	90278083	rs10137384	T	C	0.121	1.129	0.192	8.522e-06	SouthEast Asian	ncRNA_intronic	RP1133N16.3
21	43759441	rs692544	C	T	0.508	1.662	0.115	4.414e-06	SouthEast Asian	Intergenic	TFF2
22	46036079	rs1894617	G	C	0.235	1.265	0.129	7.610e-06	SouthEast Asian	Intergenic	RNU6794P
22	46046477	rs134850	A	G	0.223	1.250	0.131	6.868e-06	SouthEast Asian	Intergenic	ATXN10
2	143729878	rs10928161	C	T	0.129	1.137	0.204	7.369e-06	KhoeSan	Intronic	KYNU
2	143730019	rs16855223	G	A	0.130	1.139	0.204	6.268e-06	KhoeSan	Intronic	KYNU
2	143731496	rs35991933	A	T	0.129	1.137	0.204	7.369e-06	KhoeSan	Intronic	KYNU
2	143731661	rs34891373	T	A	0.129	1.137	0.204	7.369e-06	KhoeSan	Intronic	KYNU
2	143737201	rs11904225	G	A	0.146	1.157	0.194	1.374e-06	KhoeSan	Intronic	KYNU
2	143742532	rs10496933	G	A	0.129	1.138	0.204	5.676e-06	KhoeSan	Intronic	KYNU
2	143743246	rs12463750	G	A	0.146	1.157	0.194	1.374e-06	KhoeSan	Intronic	KYNU
2	180940603	rs13411512	T	C	0.274	1.315	0.164	8.617e-06	KhoeSan	Intronic	KYNU
4	54413304	rs4864469	T	C	0.067	1.070	0.247	8.725e-06	KhoeSan	ncRNA_intronic	FIP1L1
4	114309839	rs6533681	C	T	0.356	1.427	0.157	4.856e-06	KhoeSan	Intergenic	ANK2
5	81172726	rs62368165	G	A	0.401	1.494	0.180	9.898e-06	KhoeSan	Intergenic	SHFM1P1
6	7328023	rs145663084	T	C	0.113	1.120	0.211	4.122e-06	KhoeSan	ncRNA_exonic	PRSS23
11	86632570	rs612410	T	C	0.337	1.400	0.158	8.791e-06	KhoeSan	Intronic	PRSS23
11	86641079	rs10792884	A	G	0.361	1.435	0.160	9.751e-06	KhoeSan	Intronic	PRSS23
11	86641484	rs7940935	C	T	0.363	1.437	0.160	9.775e-06	KhoeSan	Intronic	PRSS23
11	86642522	rs10792886	G	A	0.363	1.437	0.160	8.430e-06	KhoeSan	Intronic	PRSS23
11	86643022	rs7948323	C	A	0.364	1.439	0.160	8.311e-06	KhoeSan	Intronic	PRSS23
11	86643351	rs10751145	A	G	0.364	1.439	0.160	8.311e-06	KhoeSan	Intronic	PRSS23
11	86644159	rs10792887	G	A	0.364	1.439	0.160	8.311e-06	KhoeSan	Intronic	PRSS23
11	86644244	rs10792888	A	C	0.364	1.439	0.160	8.311e-06	KhoeSan	Intronic	PRSS23
11	86644300	rs7484279	C	T	0.364	1.439	0.160	8.311e-06	KhoeSan	Intronic	PRSS23
11	86644938	rs3740665	C	T	0.365	1.440	0.160	8.889e-06	KhoeSan	Intronic	PRSS23
11	86645157	rs10898560	A	G	0.366	1.442	0.160	9.506e-06	KhoeSan	Intronic	PRSS23
11	86645214	rs1902425	C	T	0.365	1.440	0.160	8.889e-06	KhoeSan	Intronic	PRSS23
11	124196782	rs676720	C	A	0.239	1.271	0.162	8.973e-06	KhoeSan	Intergenic	OR8B7P
11	124199570	rs7119360	A	G	0.239	1.271	0.162	8.973e-06	KhoeSan	Intergenic	OR8B7P
12	18844727	rs10841067	T	C	0.050	1.052	0.292	4.802e-06	KhoeSan	Intronic variant	PLCZ1
12	18845754	rs1973289	C	T	0.054	1.056	0.282	1.710e-06	KhoeSan	ncRNA_exonic	PLCZ1
12	18846108	rs2900416	A	G	0.054	1.056	0.282	1.710e-06	KhoeSan	ncRNA_exonic	PLCZ1
15	39513293	rs7176317	C	G	0.516	1.675	0.127	2.405e-06	KhoeSan	ncRNA_exonic	RP11-624L4.1
4	162663106	rs10517752	G	A	0.199	1.221	0.222	2.142e-06	African	Intronic	FSTL5
4	162663775	rs28647531	A	G	0.204	1.226	0.224	5.518e-07	African	Intronic	FSTL5
17	75030582	rs11077888	G	A	0.470	1.600	0.142	3.400e-06	African	Intergenic	AC015815.5
18	41342728	rs11659620	T	C	0.081	1.084	0.251	5.876e-06	African	Intergenic	RNU6443P
18	41351686	rs1822027	T	G	0.081	1.084	0.251	5.876e-06	African	Intergenic	RNU6443P
18	41352221	rs35810759	A	G	0.081	1.084	0.251	5.876e-06	African	Intergenic	RNU6443P
21	43759441	rs692544	C	T	0.508	1.662	0.136	1.094e-06	African	Intergenic	TFF2

TABLE 3 | Summary statistics of the top results (p -value $< 1 \times 10^{-5}$) whilst utilising the Local Ancestry Adjusted Allelic (LAAA) model.

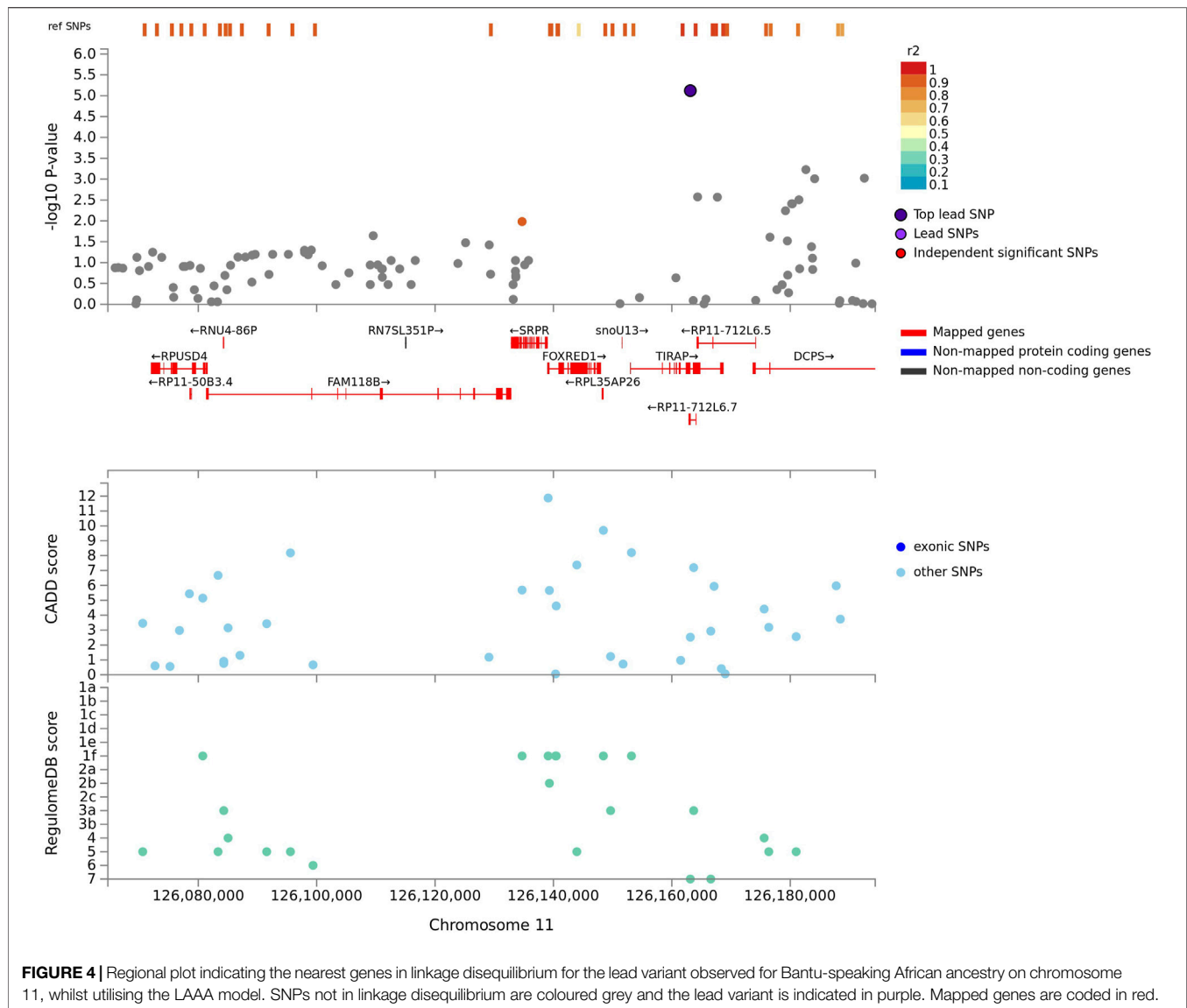
Chr	Position	rsID	Ref	Alt	Altfreq	OR	SE	p-value	Ancestry	Location	Gene
1	208027696	rs61821315	C	T	0.132	1.141	0.356	8.804e-06	African	None	None
1	208029947	rs7550821	C	T	0.132	1.141	0.358	4.771e-06	African	None	None
1	208030856	rs7551724	C	T	0.131	1.140	0.359	8.065e-06	African	None	None
2	59343477	rs17049931	C	T	0.148	1.160	0.318	5.919e-06	African	None	None
2	172987232	rs7583008	T	C	0.230	1.259	0.318	9.628e-06	African	None	None
2	172987357	rs7569224	G	C	0.230	1.259	0.318	9.628e-06	African	None	None
3	22648301	rs1449916	T	C	0.394	1.483	0.307	3.665e-06	African	None	None
3	104923287	rs13061116	G	A	0.216	1.241	0.357	5.134e-06	African	None	None
3	104923579	rs1525840	T	C	0.216	1.241	0.357	6.076e-06	African	None	None
3	104924774	rs11923672	A	T	0.216	1.241	0.357	6.076e-06	African	None	None
3	104924866	rs11926446	G	A	0.216	1.241	0.357	5.134e-06	African	None	None
3	104929569	rs9834777	T	C	0.217	1.242	0.357	5.161e-06	African	None	None
5	31584670	rs10940959	C	A	0.123	1.131	0.380	5.596e-06	African	None	None
9	73895875	rs7037178	T	G	0.459	1.582	0.235	8.903e-06	African	Intronic	<i>TRPM3</i>
9	73899145	rs1504387	T	C	0.461	1.586	0.234	3.558e-06	African	Intronic	<i>TRPM3</i>
11	126163124	rs609634	T	C	0.261	1.298	0.255	7.570e-06	African	Intronic	<i>TIRAP</i>
18	65322790	rs1444107	T	A	0.082	1.085	0.446	6.644e-06	African	Intronic	<i>DSEL-AS1</i>
18	65323846	rs2448767	A	G	0.082	1.085	0.446	6.644e-06	African	Intronic	<i>DSEL-AS1</i>
18	65324070	rs2448766	A	G	0.082	1.085	0.446	6.644e-06	African	Intronic	<i>DSEL-AS1</i>
2	183351225	rs1594304	T	C	0.421	1.523	0.253	3.557e-06	Khoesan	Intronic	<i>PDE1A</i>
5	26027283	rs12659706	C	T	0.337	1.401	0.340	9.503e-06	Khoesan	None	None
6	9576203	rs4715321	G	T	0.406	1.501	0.288	7.051e-06	Khoesan	None	None
5	142454386	rs13340374	C	T	0.061	1.063	0.532	3.421e-06	European	Intronic	<i>ARHGAP26</i>
9	119475712	rs72763937	C	T	0.184	1.202	0.320	8.548e-06	European	Intronic	<i>ASTN2</i>
11	126163124	rs609634	T	C	0.261	1.298	0.311	9.909e-06	European	Intronic	<i>TIRAP</i>
20	36998495	rs11698149	T	C	0.064	1.066	0.520	7.203e-06	European	Intronic	<i>LBP</i>
9	73099454	AS	T	C	0.324	1.383	0.531	9.295e-06	SouthEast Asian	Intronic	<i>KLF9-DT</i>

investigating TLR variants and susceptibility to TB (Schurz et al., 2015). Additionally, chromosome 11p13 was also previously associated with African ancestry in a previous GWAS (Thye et al., 2012; Chimusa et al., 2014). If the allelic model was not used while adjusting for local ancestry, this lead variant located near the *TIRAP* gene would have been missed due to the tagging SNP being located on a different ancestral haplotype LD block. This underlines the importance of including the LAAA models in association studies investigating complex multi-way admixed individuals.

One variant (rs28647531) passed the significance threshold and is located on chromosome 4q22 using the allelic model adjusting for Bantu-speaking African local ancestry (Figure 3). This variant is an intronic variant and located near the *FSTL5* gene, which has not been associated with TB susceptibility previously. This gene is a coding protein and was previously associated with colorectal cancer and acute myeloid leukaemia (Lv et al., 2017). Previous investigations of TB susceptibility in a southern African cohort identified African- and Khoesan ancestry to be associated with an increased risk for TB (Chimusa et al., 2014, 2014; Daya et al., 2014b). Likewise, previous association signals for TB susceptibility in Africans included the *WT1* gene located on chromosome 11p13 and locus 18q12 and polymorphisms in the *TLR8* genes (Thye et al., 2010, 2012; Chimusa et al., 2014). Although we did not validate these genes in our study, we did however elucidate a lead variant located on chromosome 18q12 for Bantu-speaking African ancestry whilst utilising the LAAA model, meaning both the

minor allele and ancestry co-occurs in this region. A previously unmapped protein coding gene (*DSEL-AS1*) was identified to be in LD with a leading SNP located on chromosome 18q12 for Bantu-speaking African ancestry (Supplementary Figure S6). *DSEL-AS1* is a lncRNA gene and was previously associated with unipolar depression, asparagine levels, bipolar disorder, body mass index and gut microbiome levels (Shi et al., 2011; Rhee et al., 2013; Winham et al., 2014; Ishida et al., 2020), but no biological pathways or interactions were reported for this lncRNA.

Moreover, another lead variant was identified for Bantu-speaking African ancestry. Transient receptor potential cation channel subfamily Melastatin member 3 (*TRPM3*), located on chromosome 9, is a protein coding gene which belongs to the family of transient receptor potential (TRP) channels. *TRPM3* is a permeable non-selective cation gene channel (Zhao et al., 2020, 3). Therefore, this gene is essential for cellular calcium signalling and homeostasis. Previous GWAS indicated the potential role of *TRPM3* in the measurement of mean platelet volume and were previously discovered in mostly European individuals (Aistle et al., 2016; Vuckovic et al., 2020). Another protein coding gene, Phosphodiesterase 1A (*PDE1A*), is involved in calcium signalling and was amongst the lead variants identified for Khoesan ancestry located on chromosome 2q14 by the LAAA model. This gene forms part of the cyclic nucleotide phosphodiesterases, which plays a role in signal transduction by regulating intracellular cyclic nucleotide concentrations through hydrolysis of cAMP and/or cGMP to their respective nucleoside 5-prime monophosphates. Therefore, this gene is



important for calmodulin binding and cGMP binding, as well as associated with urate measurement and glomerular filtration rate (Hellwege et al., 2019; Gill et al., 2021). Hence, there is evidence of the role of calcium ion channel activity in TB susceptibility, which includes the *FSTL5* gene and *TRPM3* gene for African ancestry, and the *PDE1A* gene for KhoeSan ancestry. *M.tb* modulates the levels and activity of key intracellular second messengers, such as calcium, to evade protective immune responses. Furthermore, calcium plays a crucial role in *M.tb* pathogenesis by activating differential transcription factors or mediating of the phagosome-lysosome fusion and cell survival (Sharma et al., 2016).

Our results demonstrate the benefit of simultaneously modelling allele, local ancestry, and ancestry-specific minor allelic effects when the admixed population under study exhibits extreme heterogeneity, since multiple distinct ancestry-specific genetic variants were identified for TB susceptibility that were previously missed by standard analyses. Thus, including an interaction term between the

minor allele present and the corresponding ancestry of that minor allele can robustly identify ancestry-specific effects on disease phenotypes in a complex admixed population. It is important to mention that only variants that met certain quality control criteria during the imputation procedure were included in our analysis. Furthermore, minor alleles might have become evident after populations diverged, or have occurred in recent human history, and they are more likely to be ancestry-specific (Qin et al., 2019). The LAAA model first described by Duan et al. (2018) counts the number of reference alleles, whereas we counted the number of copies of the alternate alleles. Minor alleles might have become evident after populations diverged, or have occurred in recent human history, and they are more likely to be ancestry-specific (Qin et al., 2019). Therefore, allowing the detection of minor ancestry-specific allelic effects.

Currently there is no clear best practise for deriving the significance cut-off threshold for admixture mapping studies.

Every admixture scenario is unique in terms of contributing ancestral source populations, density markers analysed and particularly generations since admixture occurred. Moreover, in the presence of correlated tests the Bonferroni correction for multiple testing burden is overconservative for admixture mapping studies and does not necessarily control for family-wise error rate control in association analysis (Grinde et al., 2019). For this reason, we used the method described by Grinde et al. (2019), which entails a test statistic simulation directly from the asymptotic distribution implemented in the R software package *STEAM*. It considers the number of contributing ancestral populations, number of generations since admixture occurred and the distribution of admixture proportions in the cohort of interest and permutes these factors 1,000 times to get a new cut-off for significance (Grinde et al., 2019).

A limitation of the current study is the small sample size and findings should be validated in additional larger cohorts from various ethnic groups. Given our sample size of 735 participants (392 TB cases and 346 controls), we have 95% chance to correctly rejecting the null hypothesis for large (>0.5) and medium effect sizes (>0.3). We do however lose power if the effect size is small (0.1–0.3) and any reported associations with a smaller effect size should therefore be interpreted with caution (**Supplementary Figure S7**). Furthermore, there is a possibility that the true effect could be smaller than 0.1 for ancestry-specific effects in five different continental populations, confounding the study power (Skotte et al., 2019). Since literature suggests that TB susceptibility is governed by numerous SNPs with small effect sizes, we may have missed true local ancestry effects (type 2 errors) due to our small sample size. To report on ancestry-specific susceptibility to TB in a multi-way admixed southern African population, we estimate that at least 5,568 participants are required to confidently identify markers with smaller effect sizes (0.1–0.3).

Future studies should also include *in silico* and *in vitro* validation. Moreover, progression to active TB might be explained by numerous variants having a small effect on disease outcome, or exceptionally rare variants (Schurz et al., 2015). Variants that are unique to different populations and at low frequency should also be interrogated in well-powered studies. In addition, the information on the infecting *M.tb* strain should also be included in association analysis, if possible, since it appears that *M.tb* co-evolved with humans (Brites and Gagneux, 2015) and that the interaction between host genes and *M.tb* lineage affects TB severity (Müller et al., 2021). The combination of the ancestral allele and older *M.tb* lineages, i.e., the genotype and lineage that co-existed historically, had the lowest average TB score (McHenry et al., 2020). According to the TB score system, individuals are ranked according to their relative risk of being infected with TB given certain diagnostic information. A TB score of more than 40 indicates that a TB diagnosis is highly likely, a score of 30–35 indicates a possible TB diagnosis and a score below 25 indicates an unlikely diagnosis (dos Santos et al., 2017). Thus, the host populations that were historically exposed to a specific lineage have a lower chance of disease. Similarly, the average TB score for the combinations of genotype and lineage that have not historically co-existed, were the highest (McHenry et al., 2020). Thus, the evolutionary history of both species should be considered together.

In conclusion, this is the first study to apply the LAAA model to a complex five-way admixed population from South

Africa which exhibits extensive genetic heterogeneity. This was enabled by newly developed algorithms for local ancestry inference, updated reference panels to represent contributing ancestral populations and a more suitable genotyping platform for diverse populations worldwide. We have demonstrated that the LAAA model robustly captured the source of association signals in highly complex admixed individuals. The true underlying architecture at each locus is unknown for most southern African populations, indicating that careful consideration of both global-and local ancestry is required for successful complex-trait mapping. Furthermore, local ancestry information across the genome is likely to become relevant to determine whether a genetic variant is expected to be useful in precision medicine, specifically in admixed populations.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: No new genetic data was generated for this study however, summary statistics for the quality and accuracy assessment of the genetic data will be made available to researchers who meet the criteria for access after application to the Health Research Ethics Committee of Stellenbosch University. Requests to access these datasets should be directed to MM, marlom@sun.ac.za.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Health Research Ethics Committee of the Faculty of Health Sciences, Stellenbosch University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YS performed all computational analyses. MM and CU conceived of the presented idea and supervised the work. EH and PvH contributed to the implementation of the study and established the sample bank. All authors contributed to the writing and proofreading of the final manuscript.

FUNDING

This research was partially funded by the South African government through the South African Medical Research Council and the National Research Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the South African Medical Research Council or the National Research Foundation. YS was supported by a Stellenbosch University Postgraduate Bursary.

ACKNOWLEDGMENTS

We would like to acknowledge and thank the study participants for their contribution and participation in projects from which genetic data was used.

REFERENCES

- Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation. *BMC Bioinformatics* 12, 246. doi:10.1186/1471-2105-12-246
- Astle, W. J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A. L., et al. (2016). The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* 167, 1415–1429. doi:10.1016/j.cell.2016.10.042
- Bañuls, A.-L., Sanou, A., Van Anh, N. T., and Godreuil, S. (2015). *Mycobacterium tuberculosis*: Ecology and Evolution of a Human Bacterium. *J. Med. Microbiol.* 64, 1261–1269. doi:10.1099/jmm.0.000171
- Bao, Z., Chen, R., Zhang, P., Lu, S., Chen, X., Yao, Y., et al. (2016). A Potential Target Gene for the Host-Directed Therapy of Mycobacterial Infection in Murine Macrophages. *Int. J. Mol. Med.* 38, 823–833. doi:10.3892/ijmm.2016.2675
- Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). Pong: Fast Analysis and Visualization of Latent Clusters in Population Genetic Data. *Bioinformatics* 32, 2817–2823. doi:10.1093/bioinformatics/btw327
- Brites, D., and Gagneux, S. (2015). Co-evolution of *Mycobacterium Tuberculosis* and *Homo Sapiens*. *Immunol. Rev.* 264, 6–24. doi:10.1111/imr.12264
- Cai, L., Li, Z., Guan, X., Cai, K., Wang, L., Liu, J., et al. (2019). The Research Progress of Host Genes and Tuberculosis Susceptibility. *Oxidative Med. Cell Longevity* 2019, 1–8. doi:10.1155/2019/9273056
- Capparelli, R., De Chiara, F., Di Matteo, A., Medaglia, C., and Iannelli, D. (2013). The MyD88 Rs6853 and TIRAP Rs8177374 Polymorphic Sites Are Associated with Resistance to Human Pulmonary Tuberculosis. *Genes Immun.* 14, 504–511. doi:10.1038/gene.2013.48
- Chaw, L., Chien, L.-C., Wong, J., Takahashi, K., Koh, D., and Lin, R.-T. (2020). Global Trends and Gaps in Research Related to Latent Tuberculosis Infection. *BMC Public Health* 20, 352. doi:10.1186/s12889-020-8419-0
- Chimusa, E. R., Zaitlen, N., Daya, M., Möller, M., van 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.* 23, 796–809. doi:10.1093/hmg/ddt462
- 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
- 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. doi:10.1186/1471-2164-15-1021
- Daya, M., van der Merwe, L., van Helden, P. D., Möller, M., and Hoal, E. G. (2014b). The Role of Ancestry in TB Susceptibility of an Admixed South African Population. *Tuberculosis* 94, 413–420. doi:10.1016/j.tube.2014.03.012
- 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
- Delaneau, O., Howie, B., Cox, A. J., Zagury, J.-F., and Marchini, J. (2013). Haplotype Estimation Using Sequencing Reads. *Am. J. Hum. Genet.* 93, 687–696. doi:10.1016/j.ajhg.2013.09.002
- Delaneau, O., Marchini, J., and Marchini, J. (2014). Integrating Sequence and Array Data to Create an Improved 1000 Genomes Project Haplotype Reference Panel. *Nat. Commun.* 5, 1–9. doi:10.1038/ncomms4934
- dos Santos, I. C. C., Genre, J., Marques, D., da Silva, A. M. G., dos Santos, J. C., de Araújo, J. N. G., et al. (2017). A New Panel of SNPs to Assess Thyroid Carcinoma Risk: a Pilot Study in a Brazilian Admixture Population. *BMC Med. Genet.* 18, 140. doi:10.1186/s12881-017-0502-8

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.716558/full#supplementary-material>

- Duan, Q., Xu, Z., Raffield, L. M., Chang, S., Wu, D., Lange, E. M., et al. (2018). A Robust and Powerful Two-step Testing Procedure for Local Ancestry Adjusted Allelic Association Analysis in Admixed Populations. *Genet. Epidemiol.* 42, 288–302. doi:10.1002/gepi.22104
- El Kamel, A., Joobeur, S., Skhiri, N., Cheikh Mhamed, S., Mribah, H., and Rouatbi, N. (2015). La lutte antituberculeuse dans le monde. *Revue de Pneumologie Clinique* 71, 181–187. doi:10.1016/j.pneumo.2014.03.004
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses. *Behav. Res. Methods* 41, 1149–1160. doi:10.3758/BRM.41.4.1149
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behav. Res. Methods* 39, 175–191. doi:10.3758/BF03193146
- 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.* 14, 952–959.
- Gill, D., Cameron, A. C., Burgess, S., Li, X., Doherty, D. J., Karhunen, V., et al. (2021). Urate, Blood Pressure, and Cardiovascular Disease. *Hypertension* 77, 383–392. doi:10.1161/HYPERTENSIONAHA.120.16547
- 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
- Grinde, K. E., Brown, L. A., Reiner, A. P., Thornton, T. A., and Browning, S. R. (2019). Genome-wide Significance Thresholds for Admixture Mapping Studies. *Am. J. Hum. Genet.* 104, 454–465. doi:10.1016/j.ajhg.2019.01.008
- Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., et al. (2015). The African Genome Variation Project Shapes Medical Genetics in Africa. *Nature* 517, 327–332. doi:10.1038/nature13997
- Hellwege, J. N., Velez Edwards, D. R., Giri, A., Qiu, C., Park, J., Torstenson, E. S., et al. (2019). Mapping eGFR Loci to the Renal Transcriptome and Phenome in the VA Million Veteran Program. *Nat. Commun.* 10, 3842. doi:10.1038/s41467-019-11704-w
- Ishida, S., Kato, K., Tanaka, M., Odamaki, T., Kubo, R., Mitsuyama, E., et al. (2020). Genome-wide Association Studies and Heritability Analysis Reveal the Involvement of Host Genetics in the Japanese Gut Microbiota. *Commun. Biol.* 3, 1–10. doi:10.1038/s42003-020-01416-z
- Kinney, C., Hoal, E. G., Schurz, H., van Helden, P. D., and Möller, M. (2017). The Role of Human Host Genetics in Tuberculosis Resistance. *Expert Rev. Respir. Med.* 11, 721–737. doi:10.1080/17476348.2017.1354700
- Luo, Y., Suliman, S., Asgari, S., Amariuta, T., Baglaenko, Y., Martínez-Bonet, M., et al. (2019). Early Progression to Active Tuberculosis Is a Highly Heritable Trait Driven by 3q23 in Peruvians. *Nat. Commun.* 10. doi:10.1038/s41467-019-11664-1
- Lv, H., Zhang, M., Shang, Z., Li, J., Zhang, S., Lian, D., et al. (2017). Genome-wide Haplotype Association Study Identify the FGFR2 Gene as a Risk Gene for Acute Myeloid Leukemia. *Oncotarget* 8, 7891–7899. doi:10.18632/oncotarget.13631
- 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
- Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., and Chen, W.-M. (2010). Robust Relationship Inference in Genome-wide Association Studies. *Bioinformatics* 26, 2867–2873. doi:10.1093/bioinformatics/btq559
- Maples, B. K., Gravel, S., Kenny, E. E., and Bustamante, C. D. (2013). RFMix: a Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *Am. J. Hum. Genet.* 93, 278–288. doi:10.1016/j.ajhg.2013.06.020
- McHenry, M. L., Williams, S. M., and Stein, C. M. (2020). Genetics and Evolution of Tuberculosis Pathogenesis: New Perspectives and Approaches. *Infect. Genet. Evol.* 81, 104204. doi:10.1016/j.meegid.2020.104204
- Möller, M., Nebel, A., Valentonyte, R., van Helden, P. D., Schreiber, S., and Hoal, E. G. (2009). Investigation of Chromosome 17 Candidate Genes in Susceptibility

- to TB in a South African Population. *Tuberculosis* 89, 189–194. doi:10.1016/j.tube.2008.10.001
- Müller, S. J., Schurz, H., Tromp, G., van der Spuy, G. D., Hoal, E. G., van Helden, P. D., et al. (2021). A Multi-Phenotype Genome-wide Association Study of Clades Causing Tuberculosis in a Ghanaian- and South African Cohort. *Genomics* 113, 1802–1815. doi:10.1016/j.ygeno.2021.04.024
- Nguyen, H. C., Wang, W., and Xiong, Y. (2017). Cullin-RING E3 Ubiquitin Ligases: Bridges to Destruction. *Subcell Biochem.* 83, 323–347. doi:10.1007/978-3-319-46503-6_12
- 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* 4, 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
- 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.* 26, 4752–4763. doi:10.1093/hmg/ddx365
- Qin, H., Zhao, J., and Zhu, X. (2019). Identifying Rare Variant Associations in Admixed Populations. *Sci. Rep.* 9, 5458. doi:10.1038/s41598-019-41845-3
- 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.* 86, 654. doi:10.1016/j.ajhg.2010.03.015
- Rhee, E. P., Ho, J. E., Chen, M.-H., Shen, D., Cheng, S., Larson, M. G., et al. (2013). A Genome-wide Association Study of the Human Metabolome in a Community-Based Cohort. *Cel. Metab.* 18, 130–143. doi:10.1016/j.cmet.2013.06.013
- Rossouw, M., Nél, H. J., Cooke, G. S., van Helden, P. D., and Hoal, E. G. (2003). Association between Tuberculosis and a Polymorphic NFκB Binding Site in the Interferon γ Gene. *The Lancet* 361, 1871–1872. doi:10.1016/S0140-6736(03)13491-5
- Rudko, A. A., Bragina, E. Y., Puzryev, V. P., and Freidlin, M. B. (2016). The Genetics of Susceptibility to Tuberculosis: Progress and Challenges. *Asian Pac. J. Trop. Dis.* 6, 680–684. doi:10.1016/S2222-1808(16)61109-X
- 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., Wojcik, G., van Helden, P. D., Tromp, G., et al. (2019a). A Sex-Stratified Genome-wide Association Study of Tuberculosis Using a Multi-Ethnic Genotyping Array. *Front. Genet.* 9. doi:10.3389/fgene.2018.00678
- Schurz, H., Müller, S. J., van Helden, P. D., Tromp, G., Hoal, E. G., Kinnear, C. J., et al. (2019b). Evaluating the Accuracy of Imputation Methods in a Five-Way Admixed Population. *Front. Genet.* 10. doi:10.3389/fgene.2019.00034
- Selvaraj, P., Parishankar, M., Singh, B., Jawahar, M. S., and Banurekha, V. V. (2010). Toll-like Receptor and TIRAP Gene Polymorphisms in Pulmonary Tuberculosis Patients of South India. *Tuberculosis* 90, 306–310. doi:10.1016/j.tube.2010.08.001
- Sharma, D., Tiwari, B. K., Mehto, S., Antony, C., Kak, G., Singh, Y., et al. (2016). Suppression of Protective Responses upon Activation of L-type Voltage Gated Calcium Channel in Macrophages during Mycobacterium Bovis BCG Infection. *PLOS ONE* 11, e0163845. doi:10.1371/journal.pone.0163845
- Shi, J., Potash, J. B., Knowles, J. A., Weissman, M. M., Coryell, W., Scheftner, W. A., et al. (2011). Genome-wide Association Study of Recurrent Early-Onset Major Depressive Disorder. *Mol. Psychiatry* 16, 193–201. doi:10.1038/mp.2009.124
- Shriner, D. (2013). Overview of Admixture Mapping. *Curr. Protoc. Hum. Genet.* 76, 2013 Unit 1.23. doi:10.1002/0471142905.hg0123s76
- Skotte, L., Jørsboe, E., Korneliusen, T. S., Moltke, I., and Albrechtsen, A. (2019). Ancestry-specific Association Mapping in Admixed Populations. *Genet. Epidemiol.* 43, 506–521. doi:10.1002/gepi.22200
- 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.* 98, 514–524. doi:10.1016/j.ajhg.2016.01.015
- 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., Vannberg, F. O., Wong, S. H., Owusu-Dabo, E., Osei, I., 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
- Uren, C., Henn, B. M., Franke, A., Wittig, M., van Helden, P. D., Hoal, E. G., et al. (2017a). A post-GWAS Analysis of Predicted Regulatory Variants and Tuberculosis Susceptibility. *PLoS ONE* 12, e0174738. doi:10.1371/journal.pone.0174738
- Uren, C., Hoal, E. G., and Möller, M. (2020). Putting RFMix and ADMIXTURE to the Test in a Complex Admixed Population. *BMC Genet.* 21, 40. doi:10.1186/s12863-020-00845-3
- Uren, C., Kim, M., Martin, A. R., Bobo, D., Gignoux, C. R., van Helden, P. D., et al. (2016). Fine-scale Human Population Structure in Southern Africa Reflects Ecogeographic Boundaries. *Genetics* 204, 303–314. doi:10.1534/genetics.116.187369
- Uren, C., Möller, M., van Helden, P. D., Henn, B. M., and Hoal, E. G. (2017b). Population Structure and Infectious Disease Risk in Southern Africa. *Mol. Genet. Genomics* 292, 499–509. doi:10.1007/s00438-017-1296-2
- van Helden, P. D., Möller, M., Babb, C., Warren, R., Walz, G., Uys, P., et al. (2006). TB Epidemiology and Human Genetics. *Novartis Found. Symp.* 279, 17–19. discussion 31–41, 216–219.
- Vuckovic, D., Bao, E. L., Akbari, P., Lareau, C. A., Mousas, A., Jiang, T., et al. (2020). The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* 182, 1214–e11. doi:10.1016/j.cell.2020.08.008
- Wang, K., Goldstein, S., Bleasdale, M., Clist, B., Bostoen, K., Bakwa-Lufu, P., et al. (2020). Ancient Genomes Reveal Complex Patterns of Population Movement, Interaction, and Replacement in Sub-saharan Africa. *Sci. Adv.* 6, eaaz0183. doi:10.1126/sciadv.aaz0183
- WHO (2019). Global Tuberculosis Report 2019. WHO. Available at: http://www.who.int/tb/publications/global_report/en/ (Accessed October 24, 2019).
- Winham, S. J., Cuellar-Barboza, A. B., Oliveros, A., McElroy, S. L., Crow, S., Colby, C., et al. (2014). Genome-wide Association Study of Bipolar Disorder Accounting for Effect of Body Mass index Identifies a New Risk Allele in TCF7L2. *Mol. Psychiatry* 19, 1010–1016. doi:10.1038/mp.2013.159
- Yim, J.-J., and Selvaraj, P. (2010). Genetic Susceptibility in Tuberculosis. *Respirology* 15, 241–256. doi:10.1111/j.1440-1843.2009.01690.x
- Zhang, Y. X., Xue, Y., Liu, J. Y., Zhao, M. Y., Li, F. J., Zhou, J. M., et al. (2011). Association of TIRAP (MAL) Gene Polymorphisms with Susceptibility to Tuberculosis in a Chinese Population. *Genet. Mol. Res.* 10, 7–15. doi:10.4238/vol10-1gmr980
- Zhao, S., Yudin, Y., and Rohacs, T. (2020). Disease-associated Mutations in the Human TRPM3 Render the Channel Overactive via Two Distinct Mechanisms. *Elife* 9, e5634. doi:10.7554/Elife.5634
- Zheng, R., Li, Z., He, F., Liu, H., Chen, J., Chen, J., et al. (2018). Genome-wide Association Study Identifies Two Risk Loci for Tuberculosis in Han Chinese. *Nat. Commun.* 9, 4072. doi:10.1038/s41467-018-06539-w
- Zhou, H., Alexander, D., and Lange, K. (2011). A Quasi-Newton Acceleration for High-Dimensional Optimization Algorithms. *Stat. Comput.* 21, 261–273. doi:10.1007/s11222-009-9166-3

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Swart, Uren, van Helden, Hoal and Möller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.