

# 1 Confirmation of HLA-II associations with TB susceptibility in admixed 2 African samples

3 Dayna Croock<sup>1</sup>, Yolandi Swart<sup>1</sup>, Haiko Schurz<sup>1</sup>, Desiree C. Petersen<sup>1</sup>, Marlo Möller<sup>1,2</sup>, Caitlin Uren<sup>1,2\*</sup>

4  
5 <sup>1</sup>DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre  
6 for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health  
7 Sciences, Stellenbosch University

8 <sup>2</sup>Centre for Bioinformatics and Computational Biology, Stellenbosch University

9 \*Corresponding author: [caitlinu@sun.ac.za](mailto:caitlinu@sun.ac.za)

10

## 11 Abstract

12 The International Tuberculosis Host Genetics Consortium (ITHGC) demonstrated the  
13 power of large-scale GWAS analysis across diverse ancestries in identifying tuberculosis  
14 (TB) susceptibility loci. Despite identifying a significant genetic correlate in the human  
15 leukocyte antigen (HLA)-II region, this association did not replicate in the African  
16 ancestry-specific analysis, due to small sample size and the inclusion of admixed samples.  
17 Our study aimed to build upon the findings from the ITHGC and identify TB susceptibility  
18 loci in an admixed South African cohort using the local ancestry allelic adjusted  
19 association (LAAA) model. We identified a near-genome-wide significant association  
20 (*rs3117230*,  $p$ -value =  $5.292 \times 10^{-6}$ , OR = 0.437, SE = 0.182) in the *HLA-DPB1* gene  
21 originating from KhoeSan ancestry. These findings extend the work of the ITHGC,  
22 underscore the need for innovative strategies in studying complex admixed populations,  
23 and confirm the role of the HLA-II region in TB susceptibility in admixed South African  
24 samples. [148/150 words]

25

## 26 Keywords

27 Human leukocyte antigen (HLA)-II, tuberculosis (TB), local ancestry, admixture, KhoeSan  
28 ancestry

29

## 30 Introduction

31 Tuberculosis (TB) is a communicable disease caused by *Mycobacterium tuberculosis* (*M.tb*)  
32 (World Health Organization, 2023). *M.tb* infection has a wide range of clinical manifestations

33 from asymptomatic, non-transmissible, or so-called “latent” infections to active TB (Möller

34 et al., 2018). Approximately 1/4 of the global population is infected with *M.tb*, but only 5-15%  
35 of infected individuals will develop active TB (Houben & Dodd, 2016). Several factors  
36 increase the risk of progressing to active TB, including co-infection with human  
37 immunodeficiency virus (HIV) and comorbidities, such as diabetes mellitus, asthma and  
38 other airway and lung diseases (Glaziou et al., 2018). Socio-economic factors including  
39 smoking, malnutrition, alcohol abuse, intravenous drug use, prolonged residence in a high  
40 burdened community, overcrowding, informal housing and poor sanitation also influence  
41 *M.tb* transmission and infection (Cudahy et al., 2020; Escombe et al., 2019; Laghari et al.,  
42 2019; Matose et al., 2019). Additionally, individual variability in infection and disease  
43 progression has been attributed to variation in the host genome (Uren et al., 2021; Verhein  
44 et al., 2018).

45  
46 Numerous genome-wide association studies (GWASs) investigating TB susceptibility have  
47 been conducted across different population groups. Compared to other well-studied  
48 diseases, GWASs investigating TB susceptibility are sparse and the results from these  
49 studies do not replicate well across populations (Möller & Kinnear, 2020; Möller et al., 2018;  
50 Uren et al., 2017). This lack of replication could be caused by small sample sizes, variation  
51 in phenotype definitions among studies, variation in linkage disequilibrium (LD) patterns  
52 across different population groups and the presence of population-specific effects (Möller &  
53 Kinnear, 2020). Additionally, complex LD patterns within population groups, produced by  
54 admixture, impede the detection of statistically significant loci when using traditional GWAS  
55 methods (Swart et al., 2020).

56  
57 The International Tuberculosis Host Genetics Consortium (ITHGC) performed a meta-  
58 analysis of TB GWAS results including 14 153 TB cases and 19 536 controls of African, Asian  
59 and European ancestries (Schurz et al., 2024). The multi-ancestry meta-analysis identified  
60 one genome-wide significant variant (*rs28383206*) in the human leukocyte antigen (HLA)-II  
61 region ( $p = 5.2 \times 10^{-9}$ , OR = 0.89, 95% CI = 0.84-0.95). The association peak at the *HLA-II* locus  
62 encompassed several genes encoding crucial antigen presentation proteins (including *HLA-*  
63 *DR* and *HLA-DQ*). While ancestry-specific association analyses in the European and Asian  
64 cohorts also produced suggestive peaks in the HLA-II region, the African ancestry-specific  
65 association test did not yield any associations or suggestive peaks. The authors described  
66 possible reasons for the lack of associations, including the smaller sample size compared to

67 the other ancestry-specific meta-analyses, increased genetic diversity within African  
68 individuals and population stratification produced by two admixed cohorts from the South  
69 African Coloured (SAC) population. The SAC population (as termed in the South African  
70 census (Lehohla, 2012)) form part of a multi-way (up to five-way) admixed population with  
71 ancestral contributions from Bantu-speaking African (~30%), KhoeSan (~30%), European  
72 (~20%), and East (~10%) and Southeast Asian (~10%) populations (Chimusa et al., 2013).  
73 The diverse genetic background of admixed individuals can lead to population stratification,  
74 potentially introducing confounding variables. However, the power to detect statistically  
75 significant loci in admixed populations can be improved by leveraging admixture-induced  
76 local ancestry (Swart et al., 2021; Swart, van Eeden, et al., 2022). Since previous  
77 computational algorithms were not able to include local ancestry as a covariate for GWASs,  
78 the local ancestry allelic adjusted association model (LAAA) was developed to overcome this  
79 limitation (Duan et al., 2018). The LAAA model identifies ancestry-specific alleles associated  
80 with the phenotype by including the minor alleles and the corresponding ancestry of the  
81 minor alleles (obtained by local ancestry inference) as covariates. The LAAA model has been  
82 successfully applied in a cohort of multi-way admixed SAC individuals to identify novel  
83 variants associated with TB susceptibility (Swart et al., 2021; Swart, van Eeden, et al.,  
84 2022).

85

86 Our study builds upon the findings from the ITHGC (Schurz et al., 2024) and aim to resolve  
87 the challenges faced in African ancestry-specific association analysis. Here, we explore  
88 host genetic correlates of TB in a complex admixed SAC population using the LAAA  
89 model.

90

## 91 **Methods**

### 92 *Data*

93 This study included the two SAC admixed datasets from the ITHGC analysis [RSA(A) and  
94 RSA(M)] as well as four additional TB case-control datasets obtained from admixed South  
95 African population groups (Table 1). Like the SAC population, the Xhosa population are  
96 admixed with rain-forest forager and KhoeSan ancestral contributions (Choudhury et al.,  
97 2021). All datasets were collected over the past 30 years under different research projects  
98 (Daya et al., 2013; Kroon et al., 2020; Schurz et al., 2018; Smith et al., 2023; Ugarte-Gil et  
99 al., 2020) and individuals that were included in the analyses consented to the use of their

100 data in future research regarding TB host genetics. Across all datasets, TB cases were  
101 bacteriologically confirmed (culture positive) or diagnosed by GeneXpert. Controls were  
102 healthy individuals with no previous or current history of TB disease or treatment. However,  
103 given the high prevalence of TB in South Africa [852 cases (95% CI 679-1026) per 100 00  
104 individuals 15 years and older (Cudahy et al., 2020)], most controls have likely been exposed  
105 to *M.tb* at some point (Gallant et al., 2010). For all datasets, cases and controls were  
106 obtained from the same community and thus share similar socio-economic status and  
107 health care access.

108

109 **Table 1.** Summary of the datasets included in analysis.

Dataset	Genotyping platform	Self-reported ethnicity	Cases/controls	Reference
<b>RSA(A)</b>	Affymetrix 500k	SAC	642/91	(Daya et al., 2013)
<b>RSA(M)</b>	MEGA array 1.1M	SAC	555/440	(Schurz et al., 2018; Swart et al., 2021)
<b>RSA(TANDEM)</b>	H3Africa array	SAC and Bantu-speaking African	161/133	(Swart, Uren, et al., 2022)
<b>RSA(NCTB)</b>	H3Africa array	SAC	49/111	(Oyageshio et al., 2023)
<b>RSA(Worcester)</b>	H3Africa array	SAC	61 cases	Unpublished
<b>RSA(Xhosa)</b>	Whole genome sequencing	IsiXhosa	44/120	Unpublished

110

111 A list of sites genotyped on the Infinium™ H3Africa array  
112 (<https://chipinfo.h3abionet.org/browse>) were extracted from the whole-genome sequenced  
113 [RSA(Xhosa)] dataset and treated as genotype data in subsequent analyses. Quality control  
114 (QC) of raw genotype data was performed using PLINK v1.9 (Purcell et al., 2007). In all  
115 datasets, individuals were screened for sex concordance and discordant sex information  
116 was corrected based on X chromosome homozygosity estimates ( $F_{\text{estimate}} < 0.2$  for females  
117 and  $F_{\text{estimate}} > 0.8$  for males). In the event that sex information could not be corrected based  
118 on homozygosity estimates, individuals with missing or discordant sex information were  
119 removed. Individuals with genotype call rates less than 90% and SNPs with more than 5%  
120 missingness were removed. Monomorphic sites were removed. Individuals were screened  
121 for deviations in HWE for each SNP and sites deviating from the HWE threshold of  $10^{-5}$  were  
122 removed. Sex chromosomes were excluded from the analysis. The genome coordinates  
123 across all datasets were checked for consistency and, if necessary, converted to GRCh37  
124 using the UCSC liftOver tool (Kuhn et al., 2013).

125

126 Genotype datasets were pre-phased using SHAPEIT v2 (Delaneau et al., 2013) and imputed  
127 using the Positional Burrows-Wheeler Transformation (PBWT) algorithm through the Sanger  
128 Imputation Server (SIS) (Durbin, 2014). The African Genome Resource (AGR) panel (n=4 956),  
129 accessed via the SIS, was used as the reference panel for imputation (Gurdasani et al., 2015)  
130 since it has been shown that the AGR is the best reference panel for imputation of missing  
131 genotypes for samples from the SAC population (Schurz et al., 2019). Imputed data were  
132 filtered to remove sites with imputation quality INFO scores less than 0.95. Individual  
133 datasets were screened for relatedness using KING software (Manichaikul et al., 2010) and  
134 individuals up to second degree relatedness were removed. A total of 7 544 769 markers  
135 overlapped across all six datasets. This list of intersecting markers was extracted from each  
136 dataset using PLINK --extract flag. The datasets were then merged using the PLINK v1.9. After  
137 merging, all individuals missing more than 10% genotypes were removed, markers with more  
138 than 5% missing data were excluded and a HWE filter was applied to controls (threshold <math>10^{-5}</math>).  
139 The merged dataset was screened for relatedness using KING and individuals up to second  
140 degree relatedness were subsequently removed. The final merged dataset after QC and data  
141 filtering (including the removal of related individuals) consisted of 1 544 individuals (952 TB  
142 cases and 592 healthy controls). A total of 7 510 057 variants passed QC and filtering  
143 parameters.

144

#### 145 *Global ancestry inference*

146 ADMIXTURE was used to determine the correct number of contributing ancestral proportions  
147 in our multi-way admixed population cohort (Alexander & Lange, 2011). ADMIXTURE  
148 estimates the number of contributing ancestral populations (denoted by K) and population  
149 allele frequencies through cross-validation (CV). All 1 544 individuals were grouped into  
150 running groups of equal size together with 191 reference populations (Table 2). Running  
151 groups were created to ensure approximately equal numbers of reference populations and  
152 admixed populations. Xhosa and SAC samples were divided into separate running groups.

153

154 **Table 2.** Ancestral populations included for global ancestry deconvolution.

Population	n	Source
European (British – GBR)	40	1000 Genomes (1000G) phase 3 (1000 Genomes Project Consortium et al., 2015)
East Asian (Chinese – CHB)	40	1000G phase 3

Bantu-speaking African (Yoruba – YRI)	40	1000G phase 3
Southeast Asian (Malaysian)	38	Singapore Sequencing Malay Project (SSMP) (Wong et al., 2013)
KhoeSan (Nama)	33	African Genome Variation Project (AGVP/ADRP) (Gurdasani et al., 2015)

155

156 Redundant SNPs were removed by PLINK through LD pruning by removing each SNP with LD  
157  $r^2 > 0.1$  within a 50-SNP sliding window (advanced by 10 SNPs at a time). Ancestral  
158 proportions were inferred in an unsupervised manner for  $K = 3-6$  (1 iteration). The best value  
159 of  $K$  for the data was selected by choosing the  $K$  value with the lowest CV error across all  
160 running groups. Ten iterations of  $K = 3$  and  $K = 5$  was run for the Xhosa and SAC individuals  
161 respectively. Since it has been shown that RFMix (Maples et al., 2013) outperforms  
162 ADMIXTURE in determining global ancestry proportions (C Uren et al., 2020), RFMix was also  
163 used to refine inferred global ancestry proportions. Global ancestral proportions were  
164 visualised using PONG (Behr et al., 2016).

165

#### 166 *Local ancestry inference*

167 The merged dataset and the reference file (containing reference populations from Table 2)  
168 were phased separately using SHAPEIT2. The local ancestry for each position in the genome  
169 was inferred using RFMix (Maples et al., 2013). Default parameters were used, but the  
170 number of generations since admixture was set to 15 for the SAC individuals and 20 for the  
171 Xhosa individuals (as determined by previous studies) (Uren et al., 2016). RFMix was run with  
172 three expectation maximisation iterations and the --reanalyse-reference flag.

173

#### 174 *Batch effect screening and correction*

175 Merging separate datasets generated at different timepoints and/or facilities, as we have  
176 done here, will undoubtedly introduce batch effects. Principal component analysis (PCA) is  
177 a common method used to visualise batch effects, where the first two principal components  
178 (PCs) are plotted with each sample coloured by batch, and a separation of colours is  
179 indicative of a batch effect (Nyamundanda et al., 2017). However, it is difficult to  
180 differentiate between separation caused by population structure and separation caused by  
181 batch effect using PCA alone. An alternative method to detect batch effects (Chen et al.,  
182 2022) involves coding case/control status by batch followed by running an association

183 analysis testing each batch against all other batches. If any single dataset has more positive  
184 signals compared to the other datasets, then batch effects may be responsible for producing  
185 spurious results. Batch effects can be resolved by removing those SNPs which pass the  
186 genome-wide significance threshold from the merged dataset. We have adapted this batch  
187 effect correction method for application in a highly admixed cohort with complex population  
188 structure (Croock et al., 2024). Our modified method was used to remove SNPs affected by  
189 batch effects from our merged dataset.

190

#### 191 *Local ancestry allelic adjusted association analysis*

192 The LAAA association model was used to investigate if there are allelic, ancestry-specific or  
193 ancestry-specific allelic associations with TB susceptibility in our merged dataset. Global  
194 ancestral components inferred by RFMix, age and sex were included as covariates in the  
195 association tests. Variants with minor allele frequency (MAF) < 1% were removed to improve  
196 the stability of the association tests. Dosage files, which code the number of alleles of a  
197 specific ancestry at each locus across the genome, were compiled. Separate regression  
198 models for each ancestral contribution were fitted to investigate which ancestral  
199 contribution is associated with TB susceptibility. Details regarding the models have been  
200 described elsewhere (Swart, van Eeden, et al., 2022); but in summary, four regression  
201 models were tested to detect the source of the association signals observed:

202

#### 203 *(1) Null model or global ancestry (GA) model:*

204 The null model only includes global ancestry, sex and age covariates. This test investigates  
205 whether an additive allelic dose exerts an effect on the phenotype (without including local  
206 ancestry of the allele).

207

#### 208 *(2) Local ancestry (LA) model:*

209 The LA model includes the number of alleles of a specific ancestry at a locus as covariates.  
210 This model is used in admixture mapping to identify ancestry-specific variants associated  
211 with a specific phenotype.

212

#### 213 *(3) Ancestry plus allelic (APA) model:*

214 The APA model simultaneously performs model (1) and (2). This model tests whether an  
215 additive allelic dose exerts an effect of the phenotype whilst adjusting for local ancestry.

216

217 (4) Local ancestry adjusted allelic (LAAA) model:

218 The LAAA model is an extension of the APA model, which models the combination of the  
219 minor allele and ancestry of the minor allele at a specific locus and the effect this interaction  
220 has on the phenotype.

221

222 The R package *STEAM* (Significance Threshold Estimation for Admixture Mapping) (Grinde et  
223 al., 2019) was used to determine the genome-wide significance threshold given the global  
224 ancestral proportions of each individual and the number of generations since admixture ( $g =$   
225 15). *STEAM* permuted these factors 10 000 times to derive a threshold for significance.  
226 Results were visualised in RStudio. A genome-wide significance threshold of  $p$ -value  $< 2.5 \times$   
227  $10^{-6}$  was deemed significant by *STEAM*.

228

## 229 Results

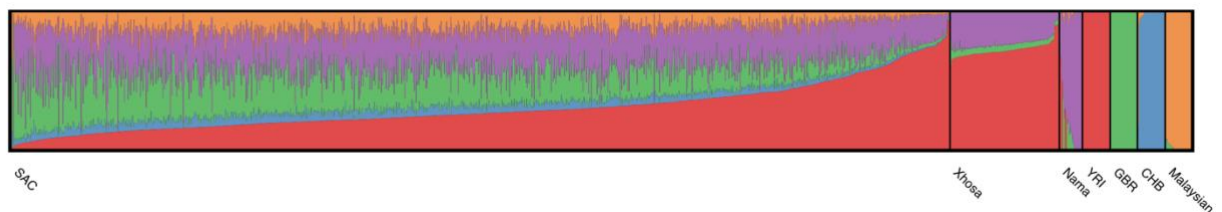
### 230 Global and local ancestry inference

231 After close inspection of global ancestry proportions generated using ADMIXTURE, the K  
232 number of contributing ancestries (the lowest k-value determined through cross-validation)  
233 was  $K = 3$  for the Xhosa individuals and  $K = 5$  for the SAC individuals (Figure 1). This is  
234 consistent with previous global ancestry deconvolution results (Chimusa et al., 2014;  
235 Choudhury et al., 2021). It is evident that our cohort is a complex, highly admixed group with  
236 ancestral contributions from the indigenous KhoeSan (~22 - 30%), Bantu-speaking African  
237 (~30 - 72%), European (~5 - 24%), Southeast Asian (~11%) and East Asian (~5%) population  
238 groups.

239

240 **K = 5**

241 [1/1 runs](#)  
242  



243

244 **Figure 1.** Genome-wide ancestral proportions of all individuals in the merged dataset. Ancestral proportions for each  
245 individual are plotted vertically with different colours representing different contributing ancestries.

246

247 Local ancestry was estimated for all individuals. Admixture between geographically distinct  
248 populations creates complex ancestral and admixture-induced LD blocks, which can be



249 visualised using local ancestry karyograms. Figure 2 shows karyograms for three individuals  
250 from the merged dataset. It is evident that, despite individuals being from the same  
251 population group, each possesses unique patterns of local ancestry arising from differing  
252 numbers and lengths of ancestral segments.

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

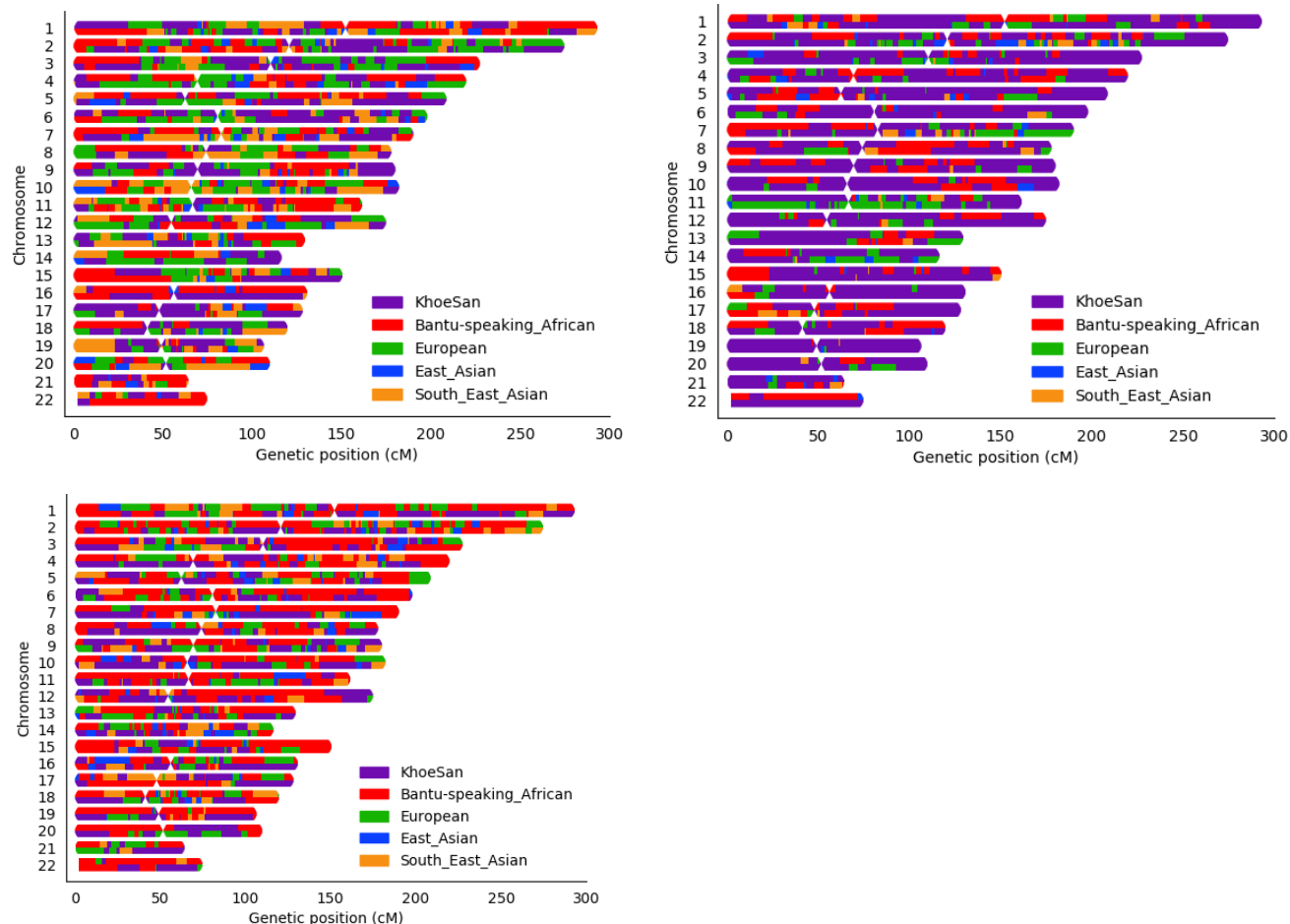
277

278

279

280

281



**Figure 2.** Local ancestry karyograms of three admixed individuals from the SAC population. Each admixed individual has unique local ancestry patterns generated by admixture among geographically distinct ancestral population groups.

### Local ancestry-allelic adjusted analysis

A total of 784 557 autosomal markers (with MAF > 1%) and 1 544 unrelated individuals (952 TB cases and 592 healthy controls) were included in logistic regression models to assess whether any loci and/or ancestries were significantly associated with TB status (whilst adjusting for sex, age, and global ancestry proportions). LAA models were successfully applied for all five contributing ancestries (Khoesan, Bantu-speaking African, European, East Asian and Southeast Asian). Only one variant (*rs74828248*) was significantly associated with

282 TB status ( $p$ -value  $< 2.5 \times 10^{-6}$ ) whilst utilising the LAAA model and whilst adjusting for Bantu-  
283 speaking African ancestry on chromosome 20 ( $p$ -value =  $2.272 \times 10^{-6}$ , OR = 0.316, SE = 0.244)  
284 (Supplementary Figure 1). No genomic inflation was detected in the QQ-plot for this region  
285 (Supplementary Figure 2). However, this variant is located in an intergenic region and the link  
286 to TB susceptibility is unclear.

287

288 Although no other variants passed the genome-wide significance threshold, multiple lead  
289 SNPs were identified. Notably, an appreciable peak was identified in the HLA-II region of  
290 chromosome 6 when using the LAAA model and adjusting for KhoeSan ancestry (Figure 3).  
291 The QQ-plot suggested minimal genomic inflation, which was verified by calculating the  
292 genomic inflation factor ( $\lambda = 1.05289$ ) (Supplementary Figure 3). The lead variants identified  
293 using the LAAA model whilst adjusting for KhoeSan ancestry in this region on chromosome 6  
294 are summarised in Table 3. The association peak encompasses the *HLA-DPA1/B1* (major  
295 histocompatibility complex, class II, DP alpha 1/beta 1) genes (Figure 4). It is noteworthy that  
296 without the LAAA model, this association peak would not have been observed for this cohort.  
297 This highlights the importance of utilising the LAAA model in future association studies when  
298 investigating disease susceptibility loci in admixed individuals, such as the SAC population.

299

300

301

302

303

304

305

306

307

308

309

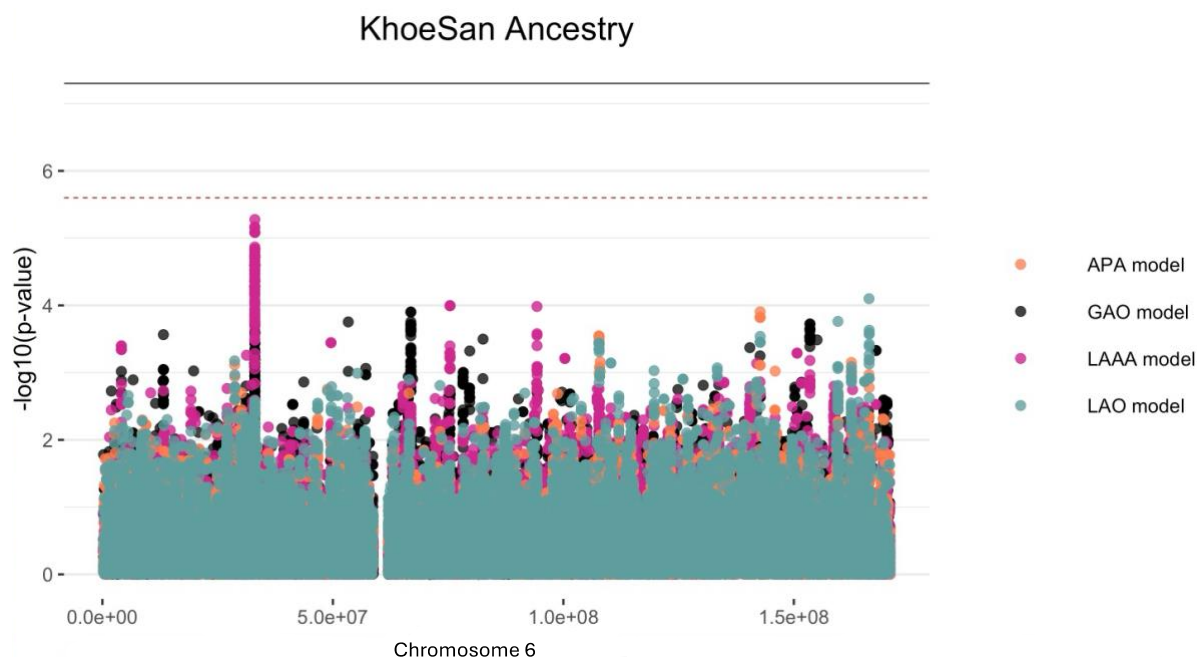
310

311

312

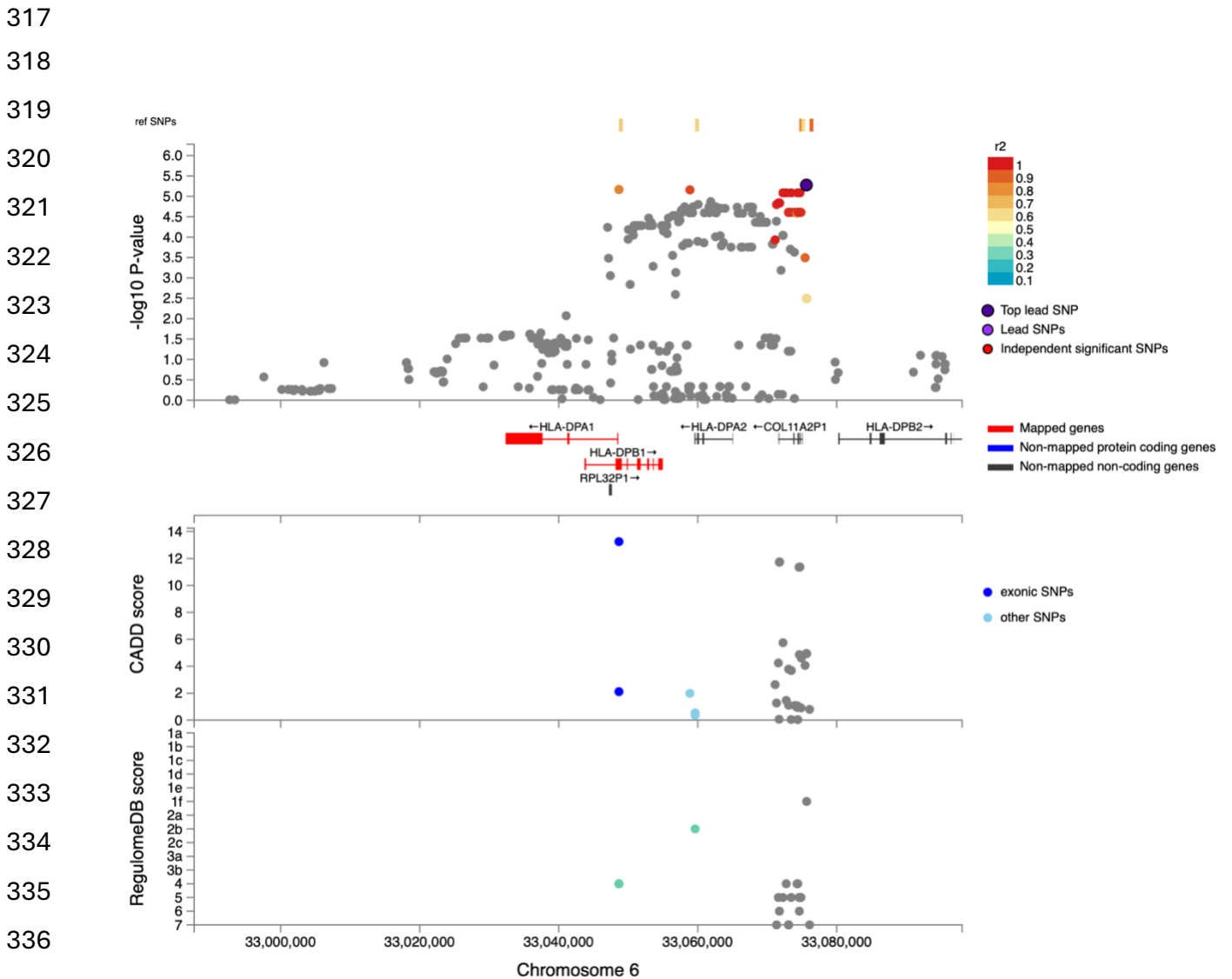
313

314



**Figure 3.** Log transformation of association signals obtained for KhoeSan ancestry whilst using the LAAA model on chromosome 6. The dashed red line represents the significant threshold for admixture mapping calculated with the software STEAM ( $p$ -value =  $2.5 \times 10^{-6}$ ) and the black solid line represents the genome wide significant threshold ( $p$ -

315 value =  $5 \times 10^{-8}$ ). The four different models are represented in black (global ancestry only - GAO), blue (local ancestry  
 316 effect - LAO), orange (ancestry plus allelic effect - APA) and pink (local ancestry adjusted allelic effect - LAAA).



338 **Figure 4.** Regional plot indicating the nearest genes in the region of the lead variant (*rs3117230*) observed on  
 339 chromosome 6. SNPs in linkage disequilibrium (LD) with the lead variant are coloured red/orange. The lead variant is  
 340 indicated in purple. Functional protein-coding genes are coded in red and non-functional (pseudo-genes) are indicated  
 341 in black.

342  
 343 The lead variant lies within *COL11A2P1* (collagen type X1 alpha 2 pseudogene 1).  
 344 *COL11A2P1* is an unprocessed pseudogene ([ENSG00000228688](https://www.ncbi.nlm.nih.gov/nuccore/ENSG00000228688)). Unprocessed  
 345 pseudogenes are seldomly transcribed and translated into functional proteins (Witek &  
 346 Mohiuddin, 2024). *HLA-DPB1* and *HLA-DPA1* are the closest functional protein-coding genes  
 347 to our lead variants.

348

349 **Table 3.** Suggestive associations ( $p$ -value <  $1e^{-5}$ ) for the LAAA analysis adjusting for KhoeSan local ancestry on  
350 chromosome 6

Position	Marker name	Ref	Alt	AltFreq	OR	SE	$p$ -value	Gene	Location
33075635	<i>rs3117230</i>	A	G	0.370	0.437	0.182	$5.292e^{-06}$	<i>HLA-DPB1</i>	Intergenic
33048661	<i>rs1042151</i>	A	G	0.325	0.437	0.184	$6.806e^{-06}$	<i>HLA-DPB1</i>	Exonic
33058874	<i>rs2179920</i>	C	T	0.369	0.445	0.180	$6.960e^{-06}$	<i>HLA-DPB1</i>	Intergenic
33072266	<i>rs2064478</i>	C	T	0.371	0.447	0.181	$8.222e^{-06}$	<i>HLA-DPB1</i>	Intergenic
33072729	<i>rs3130210</i>	G	T	0.371	0.447	0.181	$8.222e^{-06}$	<i>HLA-DPB1</i>	Intergenic
33073440	<i>rs2064475</i>	G	A	0.371	0.447	0.181	$8.222e^{-06}$	<i>HLA-DPB1</i>	Intergenic
33074348	<i>rs3117233</i>	T	C	0.371	0.447	0.181	$8.222e^{-06}$	<i>HLA-DPB1</i>	Intergenic
33074707	<i>rs3130213</i>	G	A	0.371	0.447	0.181	$8.222e^{-06}$	<i>HLA-DPB1</i>	Intergenic

351 Ref, reference allele; Alt, alternate allele; AltFreq, alternate allele frequency; OR, odds ratio; SE, standard error

352

### 353 Discussion

354 The LAAA analysis of host genetic susceptibility to TB, involving 942 TB cases and 592  
355 controls, identified one suggestive association peak adjusting for KhoeSan local ancestry.  
356 The association peak identified in this study encompasses the *HLA-DPB1* gene, a highly  
357 polymorphic locus, with over 2 000 documented allelic variants (Robinson et al., 2020). This  
358 association is noteworthy given that *HLA-DPB1* alleles have been associated with TB  
359 resistance (Dawkins et al., 2022; Ravikumar et al., 1999; Selvaraj et al., 2008). The  
360 direction of effect the lead variants in our study (Table 3) similarly suggest a protective effect  
361 against developing active TB. However, variants in *HLA-DPB1* were not identified in the ITHGC  
362 meta-analysis.

363

364 Population stratification arising from the highly heterogeneous admixed cohorts might have  
365 masked this association signal in the African ancestry-specific association analysis. The  
366 association peak in the HLA-II region was only captured using the LAAA model whilst  
367 adjusting for KhoeSan local ancestry. This underscores the importance of incorporating  
368 global and local ancestry in association studies investigating complex multi-way admixed

369 individuals, as the genetic heterogeneity present in admixed individuals (produced as a result  
370 of admixture-induced and ancestral LD patterns) may cause association signals to be missed  
371 when using traditional association models (Duan et al., 2018; Swart, van Eeden, et al.,  
372 2022).

373

374 We did not replicate the significant association signal in *HLA-DRB1* identified by the ITHGC.  
375 However, the ITHGC also did not replicate this association in their own African ancestry-  
376 specific analysis. The significant association, *rs28383206*, identified by the ITHGC appears  
377 to be tagging the *HLA-DQA1\*02:1* allele, which is associated with TB in Icelandic and Asian  
378 populations (Li et al., 2021; Sveinbjornsson et al., 2016; Zheng et al., 2018). It is possible  
379 that this association signal is specific to non-African populations, but additional research is  
380 required to verify this hypothesis. Both our study and the ITHGC independently pinpointed  
381 variants associated with TB susceptibility in different genes within the HLA-II locus (Figure 5).  
382 The HLA-II region spans ~0.8Mb on chromosome 6p21.32 and encompasses the *HLA-DP*, -  
383 *DR* and *-DQ* alpha and beta chain genes. The HLA-II complex is the human form of the major  
384 histocompatibility complex class II (MHC-II) proteins on the surface of antigen presenting  
385 cells, such as monocytes, dendritic cells and macrophages. The innate immune response  
386 against *M.tb* involves phagocytosis by alveolar macrophages. In the phagosome,  
387 mycobacterial antigens are processed for presentation on MHC-II on the surface of the  
388 antigen presenting cell. Previous studies have suggested that *M.tb* interferes with the MHC-II  
389 pathway to enhance intracellular persistence and delay activation of the adaptive immune  
390 response (Oliveira-Cortez et al., 2016). For example, *M.tb* can inhibit phagosome maturation  
391 and acidification, thereby limiting antigen processing and presentation on MHC-II molecules  
392 (Chang et al., 2005). Given that MHC-II plays an essential role in the adaptive immune  
393 response to TB and numerous studies have identified HLA-II variants associated with TB (Cai  
394 et al., 2019; Chihab et al., 2023; de Sá et al., 2020; Harishankar et al., 2018; Schurz et al.,  
395 2024; Selvaraj et al., 2008), additional research is required to elucidate the effects of HLA-II  
396 variation on TB risk status.

397

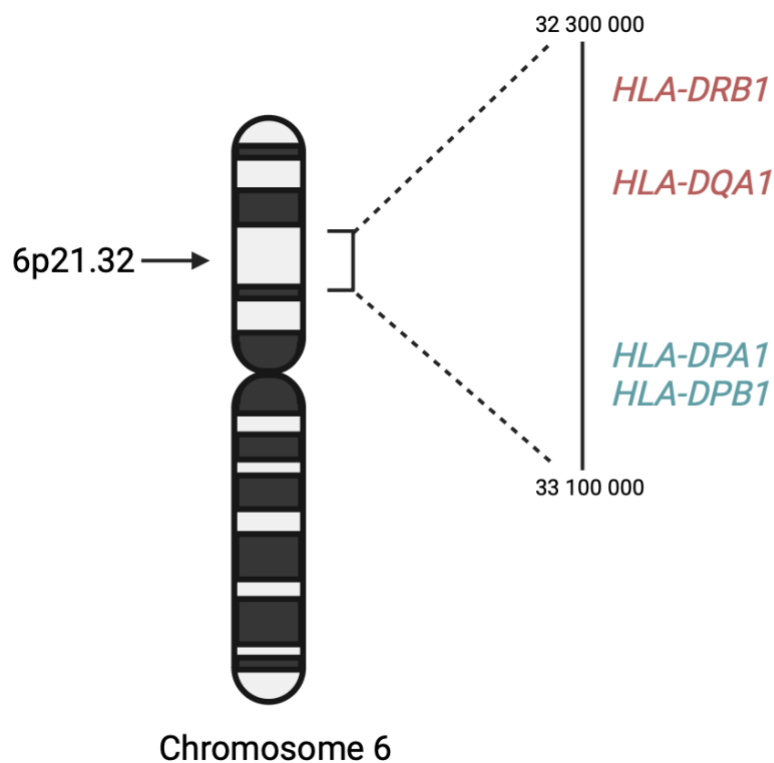
398

399

400

401

402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417



418 **Figure 5.** A schematic diagram the location of HLA-II genes associated with TB susceptibility. Genes in red were  
419 identified by the ITHGC. Genes in blue were identified by this study.  
420

421 In conclusion, application of the LAAA to a highly admixed SAC cohort revealed a suggestive  
422 association signal in the HLA-II region associated with protection against TB. Our study builds  
423 on the results of the ITHGC by demonstrating an alternative method to identify association  
424 signals in cohorts with complex genetic ancestry. This analysis shows the value of including  
425 individual global and local ancestry in genetic association analyses. Furthermore, we  
426 confirm HLA-II loci associations with TB susceptibility in an admixed South African  
427 population and hope that this publication will encourage greater appreciation for the role of  
428 the adaptive immune system in TB susceptibility and resistance.

429  
430 **Acknowledgements**

431 We acknowledge the support of the DSI-NRF Centre of Excellence for Biomedical  
432 Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis  
433 Research (SAMRC CTR), Division of Molecular Biology and Human Genetics, Faculty of  
434 Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. We also

435 acknowledge the Centre for High Performance Computing (CHPC), South Africa, for  
436 providing computational resources. This research was partially funded by the South African  
437 government through the SAMRC and the Harry Crossley Research Foundation.

438

#### 439 **Author ORCIDs**

440 Dayna Croock: 0000-0002-5107-8006

441 Yolandi Swart: 0000-0002-9840-3646

442 Haiko Schurz: 0000-0002-0009-3409

443 Desiree C. Petersen: 0000-0002-0817-2574

444 Marlo Möller: 0000-0002-0805-6741

445 Caitlin Uren: 0000-0003-2358-0135

446

#### 447 **Ethics**

448 Ethics approval was granted by the Health Research Ethics Committee (HREC) of  
449 Stellenbosch University, South Africa (project number S22/02/031).

450

#### 451 **Competing interests**

452 None declared.

453

#### 454 **References**

455 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E.  
456 P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis,  
457 G. R. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–  
458 74. <https://doi.org/10.1038/nature15393>

459 Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for  
460 individual ancestry estimation. *BMC Bioinformatics*, 12, 246.  
461 <https://doi.org/10.1186/1471-2105-12-246>

- 462 Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P., & Ramachandran, S. (2016). pong: fast  
463 analysis and visualization of latent clusters in population genetic data.  
464 *Bioinformatics*, 32(18), 2817–2823. <https://doi.org/10.1093/bioinformatics/btw327>
- 465 Cai, L., Li, Z., Guan, X., Cai, K., Wang, L., Liu, J., & Tong, Y. (2019). The research progress  
466 of host genes and tuberculosis susceptibility. *Oxidative Medicine and Cellular  
467 Longevity*, 2019, 9273056. <https://doi.org/10.1155/2019/9273056>
- 468 Chang, S. T., Linderman, J. J., & Kirschner, D. E. (2005). Multiple mechanisms allow  
469 Mycobacterium tuberculosis to continuously inhibit MHC class II-mediated antigen  
470 presentation by macrophages. *Proceedings of the National Academy of Sciences of  
471 the United States of America*, 102(12), 4530–4535.  
472 <https://doi.org/10.1073/pnas.0500362102>
- 473 Chen, D., Tashman, K., Palmer, D. S., Neale, B., Roeder, K., Bloemendal, A.,  
474 Churchhouse, C., & Ke, Z. T. (2022). A data harmonization pipeline to leverage  
475 external controls and boost power in GWAS. *Human Molecular Genetics*, 31(3), 481–  
476 489. <https://doi.org/10.1093/hmg/ddab261>
- 477 Chihab, L. Y., Kuan, R., Phillips, E. J., Mallal, S. A., Rozot, V., Davis, M. M., Scriba, T. J.,  
478 Sette, A., Peters, B., Lindestam Arlehamn, C. S., & SATVI Study Group. (2023).  
479 Expression of specific HLA class II alleles is associated with an increased risk for  
480 active tuberculosis and a distinct gene expression profile. *HLA : Immune Response  
481 Genetics*, 101(2), 124–137. <https://doi.org/10.1111/tan.14880>



- 482 Chimusa, E. R., Daya, M., Möller, M., Ramesar, R., Henn, B. M., van Helden, P. D.,  
483 Mulder, N. J., & Hoal, E. G. (2013). Determining ancestry proportions in complex  
484 admixture scenarios in South Africa using a novel proxy ancestry selection method.  
485 *Plos One*, 8(9), e73971. <https://doi.org/10.1371/journal.pone.0073971>
- 486 Chimusa, E. R., Zaitlen, N., Daya, M., Möller, M., van Helden, P. D., Mulder, N. J., Price,  
487 A. L., & Hoal, E. G. (2014). Genome-wide association study of ancestry-specific TB  
488 risk in the South African Coloured population. *Human Molecular Genetics*, 23(3),  
489 796–809. <https://doi.org/10.1093/hmg/ddt462>
- 490 Choudhury, A., Sengupta, D., Ramsay, M., & Schlebusch, C. (2021). Bantu-speaker  
491 migration and admixture in southern Africa. *Human Molecular Genetics*, 30(R1), R56–  
492 R63. <https://doi.org/10.1093/hmg/ddaa274>
- 493 Cudahy, P. G. T., Wilson, D., & Cohen, T. (2020). Risk factors for recurrent tuberculosis  
494 after successful treatment in a high burden setting: a cohort study. *BMC Infectious*  
495 *Diseases*, 20(1), 789. <https://doi.org/10.1186/s12879-020-05515-4>
- 496 Dawkins, B. A., Garman, L., Cejda, N., Pezant, N., Rasmussen, A., Rybicki, B. A., Levin,  
497 A. M., Benchek, P., Seshadri, C., Mayanja-Kizza, H., Iannuzzi, M. C., Stein, C. M., &  
498 Montgomery, C. G. (2022). Novel HLA associations with outcomes of Mycobacterium  
499 tuberculosis exposure and sarcoidosis in individuals of African ancestry using  
500 nearest-neighbor feature selection. *Genetic Epidemiology*, 46(7), 463–474.  
501 <https://doi.org/10.1002/gepi.22490>

- 502 Daya, M., van der Merwe, L., Galal, U., Möller, M., Salie, M., Chimusa, E. R., Galanter, J.  
503 M., van Helden, P. D., Henn, B. M., Gignoux, C. R., & Hoal, E. (2013). A panel of  
504 ancestry informative markers for the complex five-way admixed South African  
505 coloured population. *Plos One*, 8(12), e82224.  
506 <https://doi.org/10.1371/journal.pone.0082224>
- 507 de Sá, N. B. R., Ribeiro-Alves, M., da Silva, T. P., Pilotto, J. H., Rolla, V. C., Giacoia-  
508 Gripp, C. B. W., Scott-Algara, D., Morgado, M. G., & Teixeira, S. L. M. (2020). Clinical  
509 and genetic markers associated with tuberculosis, HIV-1 infection, and TB/HIV-  
510 immune reconstitution inflammatory syndrome outcomes. *BMC Infectious Diseases*,  
511 20(1), 59. <https://doi.org/10.1186/s12879-020-4786-5>
- 512 Delaneau, O., Howie, B., Cox, A. J., Zagury, J.-F., & Marchini, J. (2013). Haplotype  
513 estimation using sequencing reads. *American Journal of Human Genetics*, 93(4),  
514 687–696. <https://doi.org/10.1016/j.ajhg.2013.09.002>
- 515 Duan, Q., Xu, Z., Raffield, L. M., Chang, S., Wu, D., Lange, E. M., Reiner, A. P., & Li, Y.  
516 (2018). A robust and powerful two-step testing procedure for local ancestry adjusted  
517 allelic association analysis in admixed populations. *Genetic Epidemiology*, 42(3),  
518 288–302. <https://doi.org/10.1002/gepi.22104>
- 519 Durbin, R. (2014). Efficient haplotype matching and storage using the positional  
520 Burrows-Wheeler transform (PBWT). *Bioinformatics*, 30(9), 1266–1272.  
521 <https://doi.org/10.1093/bioinformatics/btu014>

- 522 Escombe, A. R., Ticona, E., Chávez-Pérez, V., Espinoza, M., & Moore, D. A. J. (2019).  
523 Improving natural ventilation in hospital waiting and consulting rooms to reduce  
524 nosocomial tuberculosis transmission risk in a low resource setting. *BMC Infectious*  
525 *Diseases*, 19(1), 88. <https://doi.org/10.1186/s12879-019-3717-9>
- 526 Gallant, C. J., Cobat, A., Simkin, L., Black, G. F., Stanley, K., Hughes, J., Doherty, T. M.,  
527 Hanekom, W. A., Eley, B., Beyers, N., Jaïs, J. P., van Helden, P., Abel, L., Alcaïs, A.,  
528 Hoal, E. G., & Schurr, E. (2010). Impact of age and sex on mycobacterial immunity in  
529 an area of high tuberculosis incidence. *The International Journal of Tuberculosis and*  
530 *Lung Disease*, 14(8), 952–959.
- 531 Glaziou, P., Floyd, K., & Raviglione, M. C. (2018). Global epidemiology of tuberculosis.  
532 *Seminars in Respiratory and Critical Care Medicine*, 39(3), 271–285.  
533 <https://doi.org/10.1055/s-0038-1651492>
- 534 Grinde, K. E., Brown, L. A., Reiner, A. P., Thornton, T. A., & Browning, S. R. (2019).  
535 Genome-wide Significance Thresholds for Admixture Mapping Studies. *American*  
536 *Journal of Human Genetics*, 104(3), 454–465.  
537 <https://doi.org/10.1016/j.ajhg.2019.01.008>
- 538 Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I.,  
539 Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M. O., Choudhury, A., Ritchie, G.  
540 R. S., Xue, Y., Asimit, J., Nsubuga, R. N., Young, E. H., Pomilla, C., Kivinen, K.,  
541 Rockett, K., Kamali, A., ... Sandhu, M. S. (2015). The African Genome Variation Project  
542 shapes medical genetics in Africa. *Nature*, 517(7534), 327–332.  
543 <https://doi.org/10.1038/nature13997>

- 544 Harishankar, M., Selvaraj, P., & Bethunaickan, R. (2018). Influence of genetic  
545 polymorphism towards pulmonary tuberculosis susceptibility. *Frontiers in Medicine*,  
546 5, 213. <https://doi.org/10.3389/fmed.2018.00213>
- 547 Houben, R. M. G. J., & Dodd, P. J. (2016). The Global Burden of Latent Tuberculosis  
548 Infection: A Re-estimation Using Mathematical Modelling. *PLoS Medicine*, 13(10),  
549 e1002152. <https://doi.org/10.1371/journal.pmed.1002152>
- 550 Kroon, E. E., Kinnear, C. J., Orlova, M., Fischinger, S., Shin, S., Boolay, S., Walzl, G.,  
551 Jacobs, A., Wilkinson, R. J., Alter, G., Schurr, E., Hoal, E. G., & Möller, M. (2020). An  
552 observational study identifying highly tuberculosis-exposed, HIV-1-positive but  
553 persistently TB, tuberculin and IGRA negative persons with M. tuberculosis specific  
554 antibodies in Cape Town, South Africa. *EBioMedicine*, 61, 103053.  
555 <https://doi.org/10.1016/j.ebiom.2020.103053>
- 556 Kuhn, R. M., Haussler, D., & Kent, W. J. (2013). The UCSC genome browser and  
557 associated tools. *Briefings in Bioinformatics*, 14(2), 144–161.  
558 <https://doi.org/10.1093/bib/bbs038>
- 559 Laghari, M., Sulaiman, S. A. S., Khan, A. H., Talpur, B. A., Bhatti, Z., & Memon, N. (2019).  
560 Contact screening and risk factors for TB among the household contact of children  
561 with active TB: a way to find source case and new TB cases. *BMC Public Health*,  
562 19(1), 1274. <https://doi.org/10.1186/s12889-019-7597-0>
- 563 Lehohla, P. (2012). *South African Census 2011 Meta-data* (Report No. 03-01-47; p. 130).  
564 South African Census.

- 565 Li, M., Hu, Y., Zhao, B., Chen, L., Huang, H., Huai, C., Zhang, X., Zhang, J., Zhou, W.,  
566 Shen, L., Zhen, Q., Li, B., Wang, W., He, L., & Qin, S. (2021). A next generation  
567 sequencing combined genome-wide association study identifies novel tuberculosis  
568 susceptibility loci in Chinese population. *Genomics*, 113(4), 2377–2384.  
569 <https://doi.org/10.1016/j.ygeno.2021.05.035>
- 570 Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010).  
571 Robust relationship inference in genome-wide association studies. *Bioinformatics*,  
572 26(22), 2867–2873. <https://doi.org/10.1093/bioinformatics/btq559>
- 573 Maples, B. K., Gravel, S., Kenny, E. E., & Bustamante, C. D. (2013). RFMix: a  
574 discriminative modeling approach for rapid and robust local-ancestry inference.  
575 *American Journal of Human Genetics*, 93(2), 278–288.  
576 <https://doi.org/10.1016/j.ajhg.2013.06.020>
- 577 Matose, M., Poluta, M., & Douglas, T. S. (2019). Natural ventilation as a means of  
578 airborne tuberculosis infection control in minibus taxis. *South African Journal of*  
579 *Science*, 115(9/10). <https://doi.org/10.17159/sajs.2019/5737>
- 580 Möller, M., Kinnear, C. J., Orlova, M., Kroon, E. E., van Helden, P. D., Schurr, E., & Hoal,  
581 E. G. (2018). Genetic Resistance to Mycobacterium tuberculosis Infection and  
582 Disease. *Frontiers in Immunology*, 9, 2219.  
583 <https://doi.org/10.3389/fimmu.2018.02219>
- 584 Möller, M., & Kinnear, C. J. (2020). Human global and population-specific genetic  
585 susceptibility to Mycobacterium tuberculosis infection and disease. *Current Opinion*

586 *in Pulmonary Medicine*, 26(3), 302–310.

587 <https://doi.org/10.1097/MCP.0000000000000672>

588 Nyamundanda, G., Poudel, P., Patil, Y., & Sadanandam, A. (2017). A novel statistical

589 method to diagnose, quantify and correct batch effects in genomic studies. *Scientific*

590 *Reports*, 7(1), 10849. <https://doi.org/10.1038/s41598-017-11110-6>

591 Oliveira-Cortez, A., Melo, A. C., Chaves, V. E., Condino-Neto, A., & Camargos, P. (2016).

592 Do HLA class II genes protect against pulmonary tuberculosis? A systematic review

593 and meta-analysis. *European Journal of Clinical Microbiology & Infectious Diseases*,

594 35(10), 1567–1580. <https://doi.org/10.1007/s10096-016-2713-x>

595 Oyageshio, O. P., Myrick, J. W., Saayman, J., van der Westhuizen, L., Al-Hindi, D.,

596 Reynolds, A. W., Zaitlen, N., Uren, C., Möller, M., & Henn, B. M. (2023). Strong effect

597 of demographic changes on tuberculosis susceptibility in south africa. *MedRxiv*.

598 <https://doi.org/10.1101/2023.11.02.23297990>

599 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller,

600 J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for

601 whole-genome association and population-based linkage analyses. *American*

602 *Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>

603 Ravikumar, M., Dheenadhayalan, V., Rajaram, K., Lakshmi, S. S., Kumaran, P. P.,

604 Paramasivan, C. N., Balakrishnan, K., & Pitchappan, R. M. (1999). Associations of

605 HLA-DRB1, DQB1 and DPB1 alleles with pulmonary tuberculosis in south India.

606 *Tubercle and Lung Disease : The Official Journal of the International Union against*

- 607     *Tuberculosis and Lung Disease*, 79(5), 309–317.
- 608     <https://doi.org/10.1054/tuld.1999.0213>
- 609     Robinson, J., Barker, D. J., Georgiou, X., Cooper, M. A., Flicek, P., & Marsh, S. G. E.
- 610     (2020). IPD-IMGT/HLA Database. *Nucleic Acids Research*, 48(D1), D948–D955.
- 611     <https://doi.org/10.1093/nar/gkz950>
- 612     Schurz, H., Kinnear, C. J., Gignoux, C., Wojcik, G., van Helden, P. D., Tromp, G., Henn,
- 613     B., Hoal, E. G., & Möller, M. (2018). A Sex-Stratified Genome-Wide Association Study
- 614     of Tuberculosis Using a Multi-Ethnic Genotyping Array. *Frontiers in Genetics*, 9, 678.
- 615     <https://doi.org/10.3389/fgene.2018.00678>
- 616     Schurz, H., Müller, S. J., van Helden, P. D., Tromp, G., Hoal, E. G., Kinnear, C. J., &
- 617     Möller, M. (2019). Evaluating the Accuracy of Imputation Methods in a Five-Way
- 618     Admixed Population. *Frontiers in Genetics*, 10, 34.
- 619     <https://doi.org/10.3389/fgene.2019.00034>
- 620     Schurz, H., Naranbhai, V., Yates, T. A., Gilchrist, J. J., Parks, T., Dodd, P. J., Möller, M.,
- 621     Hoal, E. G., Morris, A. P., Hill, A. V. S., & International Tuberculosis Host Genetics
- 622     Consortium. (2024). Multi-ancestry meta-analysis of host genetic susceptibility to
- 623     tuberculosis identifies shared genetic architecture. *ELife*, 13.
- 624     <https://doi.org/10.7554/eLife.84394>
- 625     Selvaraj, P., Raghavan, S., Swaminathan, S., Alagarasu, K., Narendran, G., &
- 626     Narayanan, P. R. (2008). HLA-DQB1 and -DPB1 allele profile in HIV infected patients

- 627 with and without pulmonary tuberculosis of south India. *Infection, Genetics and*  
628 *Evolution*, 8(5), 664–671. <https://doi.org/10.1016/j.meegid.2008.06.005>
- 629 Smith, M. H., Myrick, J. W., Oyageshio, O., Uren, C., Saayman, J., Boolay, S., van der  
630 Westhuizen, L., Werely, C., Möller, M., Henn, B. M., & Reynolds, A. W. (2023).  
631 Epidemiological correlates of overweight and obesity in the Northern Cape Province,  
632 South Africa. *PeerJ*, 11, e14723. <https://doi.org/10.7717/peerj.14723>
- 633 Sveinbjornsson, G., Gudbjartsson, D. F., Halldorsson, B. V., Kristinsson, K. G.,  
634 Gottfredsson, M., Barrett, J. C., Gudmundsson, L. J., Blondal, K., Gylfason, A.,  
635 Gudjonsson, S. A., Helgadóttir, H. T., Jonasdóttir, A., Jonasdóttir, A., Karason, A.,  
636 Kardum, L. B., Knežević, J., Kristjansson, H., Kristjansson, M., Love, A., ... Stefansson,  
637 K. (2016). HLA class II sequence variants influence tuberculosis risk in populations of  
638 European ancestry. *Nature Genetics*, 48(3), 318–322.  
639 <https://doi.org/10.1038/ng.3498>
- 640 Swart, Y., Uren, C., Eckold, C., Cliff, J. M., Malherbe, S. T., Ronacher, K., Kumar, V.,  
641 Wijmenga, C., Dockrell, H. M., van Crevel, R., Walzl, G., Kleynhans, L., & Möller, M.  
642 (2022). *cis* -eQTL mapping of TB-T2D comorbidity elucidates the involvement of  
643 African ancestry in TB susceptibility. *BioRxiv*.  
644 <https://doi.org/10.1101/2022.10.19.512814>
- 645 Swart, Y., Uren, C., van Helden, P. D., Hoal, E. G., & Möller, M. (2021). Local ancestry  
646 adjusted allelic association analysis robustly captures tuberculosis susceptibility  
647 loci. *Frontiers in Genetics*, 12, 716558. <https://doi.org/10.3389/fgene.2021.716558>

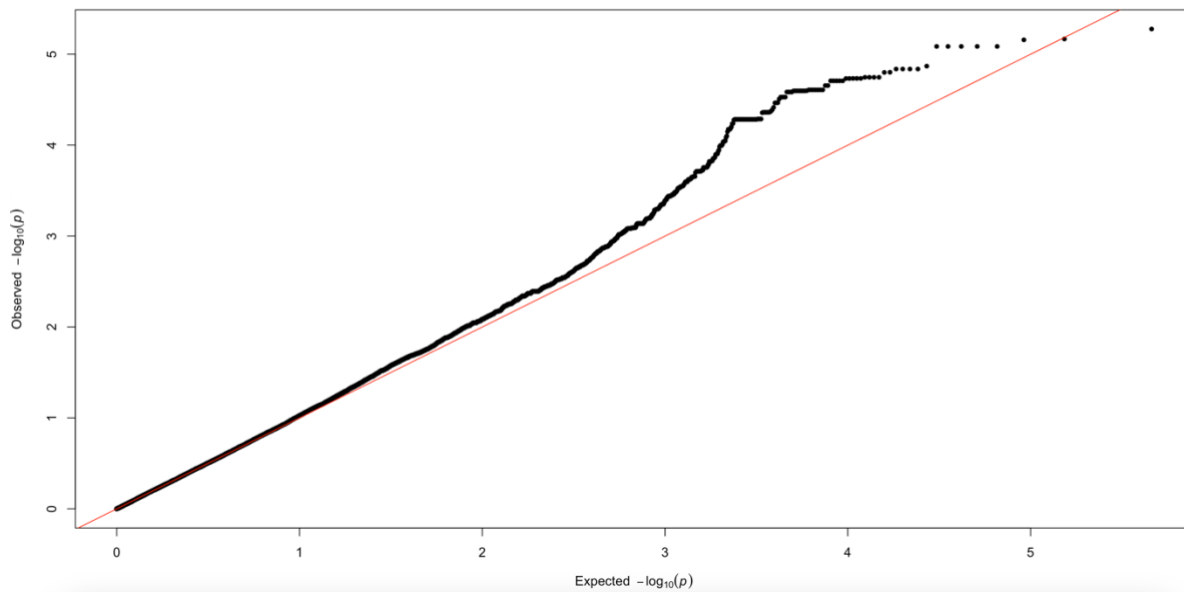


- 648 Swart, Y., van Eeden, G., Sparks, A., Uren, C., & Möller, M. (2020). Prospective avenues  
649 for human population genomics and disease mapping in southern Africa. *Molecular*  
650 *Genetics and Genomics*, 295(5), 1079–1089. [https://doi.org/10.1007/s00438-020-](https://doi.org/10.1007/s00438-020-01684-8)  
651 01684-8
- 652 Swart, Y., van Eeden, G., Uren, C., van der Spuy, G., Tromp, G., & Moller, M. (2022).  
653 GWAS in the southern African context. *Cold Spring Harbor Laboratory*.  
654 <https://doi.org/10.1101/2022.02.16.480704>
- 655 Ugarte-Gil, C., Alisjahbana, B., Ronacher, K., Riza, A. L., Koesoemadinata, R. C.,  
656 Malherbe, S. T., Cioboata, R., Llontop, J. C., Kleynhans, L., Lopez, S., Santoso, P.,  
657 Marius, C., Villaizan, K., Ruslami, R., Walzl, G., Panduru, N. M., Dockrell, H. M., Hill,  
658 P. C., Mc Allister, S., ... van Crevel, R. (2020). Diabetes Mellitus Among Pulmonary  
659 Tuberculosis Patients From 4 Tuberculosis-endemic Countries: The TANDEM Study.  
660 *Clinical Infectious Diseases*, 70(5), 780–788. <https://doi.org/10.1093/cid/ciz284>
- 661 Uren, C, Hoal, E. G., & Möller, M. (2020). Putting RFMix and ADMIXTURE to the test in a  
662 complex admixed population. *BMC Genetics*, 21(1), 40.  
663 <https://doi.org/10.1186/s12863-020-00845-3>
- 664 Uren, Caitlin, Henn, B. M., Franke, A., Wittig, M., van Helden, P. D., Hoal, E. G., & Möller,  
665 M. (2017). A post-GWAS analysis of predicted regulatory variants and tuberculosis  
666 susceptibility. *Plos One*, 12(4), e0174738.  
667 <https://doi.org/10.1371/journal.pone.0174738>

- 668 Uren, Caitlin, Hoal, E. G., & Möller, M. (2021). Mycobacterium tuberculosis complex  
669 and human coadaptation: a two-way street complicating host susceptibility to TB.  
670 *Human Molecular Genetics*, 30(R1), R146–R153.  
671 <https://doi.org/10.1093/hmg/ddaa254>
- 672 Verhein, K. C., Vellers, H. L., & Kleeberger, S. R. (2018). Inter-individual variation in  
673 health and disease associated with pulmonary infectious agents. *Mammalian*  
674 *Genome*, 29(1–2), 38–47. <https://doi.org/10.1007/s00335-018-9733-z>
- 675 Witek, J., & Mohiuddin, S. S. (2024). Biochemistry, Pseudogenes. In *StatPearls*.  
676 StatPearls Publishing.
- 677 Wong, L.-P., Ong, R. T.-H., Poh, W.-T., Liu, X., Chen, P., Li, R., Lam, K. K.-Y., Pillai, N. E.,  
678 Sim, K.-S., Xu, H., Sim, N.-L., Teo, S.-M., Foo, J.-N., Tan, L. W.-L., Lim, Y., Koo, S.-H.,  
679 Gan, L. S.-H., Cheng, C.-Y., Wee, S., ... Teo, Y.-Y. (2013). Deep whole-genome  
680 sequencing of 100 southeast Asian Malays. *American Journal of Human Genetics*,  
681 92(1), 52–66. <https://doi.org/10.1016/j.ajhg.2012.12.005>
- 682 World Health Organization. (2023). *Global Tuberculosis Report 2023* (World Health  
683 Organization, Ed.; p. 75). World Health Organization.
- 684 Zheng, R., Li, Z., He, F., Liu, H., Chen, J., Chen, J., Xie, X., Zhou, J., Chen, H., Wu, X., Wu,  
685 J., Chen, B., Liu, Y., Cui, H., Fan, L., Sha, W., Liu, Y., Wang, J., Huang, X., ... Ge, B. (2018).  
686 Genome-wide association study identifies two risk loci for tuberculosis in Han Chinese.  
687 *Nature Communications*, 9(1), 4072. <https://doi.org/10.1038/s41467-018-06539-w>  
688  
689  
690



742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769



**Supplementary Figure 3.** QQ-plot of expected  $p$ -values and observed  $p$ -values for the association signals obtained for Khoisan ancestry located on chromosome 6.