



OPEN

Replication of HLA class II locus association with susceptibility to podoconiosis in three Ethiopian ethnic groups

Tewodros Gebresilase^{1,2}, Chris Finan³, Daniel Suveges^{4,7}, Tesfaye Sisay Tessema², Abraham Aseffa¹, Gail Davey⁵, Konstantinos Hatzikotoulas^{4,8,9}, Eleftheria Zeggini^{4,8,9}, Melanie J. Newport^{5✉} & Fasil Tekola-Ayele⁶

Podoconiosis, a debilitating lymphoedema of the leg, results from barefoot exposure to volcanic clay soil in genetically susceptible individuals. A previous genome-wide association study (GWAS) conducted in the Wolaita ethnic group from Ethiopia showed association between single nucleotide polymorphisms (SNPs) in the HLA class II region and podoconiosis. We aimed to conduct a second GWAS in a new sample (N = 1892) collected from the Wolaita and two other Ethiopian populations, the Amhara and the Oromo, also affected by podoconiosis. Fourteen SNPs in the HLA class II region showed significant genome-wide association ($P < 5.0 \times 10^{-8}$) with podoconiosis. The lead SNP was rs9270911 ($P = 5.51 \times 10^{-10}$; OR 1.53; 95% CI 1.34–1.74), located near *HLA-DRB1*. Inclusion of data from the first GWAS (combined N = 2289) identified 47 SNPs in the class II HLA region that were significantly associated with podoconiosis (lead SNP also rs9270911 ($P = 2.25 \times 10^{-12}$)). No new loci outside of the HLA class II region were identified in this more highly-powered second GWAS. Our findings confirm the HLA class II association with podoconiosis suggesting HLA-mediated abnormal induction and regulation of immune responses may have a direct role in its pathogenesis.

The neglected tropical disease podoconiosis is a type of progressive tropical lymphoedema that mainly affects the lower leg. It results from long term exposure of bare feet to red clay soil derived from volcanic rock¹. The disease has been described in at least 32 countries, affecting susceptible individuals who do not wear shoes consistently^{2,3}. It is estimated that there are 4 million people living with podoconiosis globally, mainly living in highland regions of tropical countries in Africa, South and Central America and southeast Asia³. Ethiopia bears the largest burden of cases where nationwide disease mapping indicated a prevalence of 4%⁴ amounting to 1.5 million adults living with podoconiosis and a further 34.9 million at risk of the disease⁵. In addition to its physical consequences, which include disability and painful inflammatory episodes affecting the lower limb (known as acute dermatolymphangioadenitis)⁶, podoconiosis has considerable negative economic⁷ and psychosocial impact^{8–10} on affected individuals, families and communities.

Little is known about the pathogenesis of the disease. Early histopathology and electron microscopy studies suggest that soil particles cross the skin, are taken up by macrophages and transported to regional lymph nodes¹¹. It was proposed that the immune system is activated leading to inflammation and scarring particularly affecting the lymphatic system. The lymph vessels gradually become obstructed causing progressive lymphoedema with corresponding skin changes in the lower limb that are typical of podoconiosis. These include dermal nodules and a rough, velvet-like appearance to the skin known as mossy changes which are pathognomonic for the disease.

¹Armauer Hansen Research Institute (AHRI), Addis Ababa, Ethiopia. ²Unit of Health Biotechnology, Institute of Biotechnology, College of Natural and Computational Sciences, Addis Ababa University, Addis Ababa, Ethiopia. ³Institute of Cardiovascular Science, Faculty of Population Health, University College London, London, UK. ⁴Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. ⁵Brighton and Sussex Centre for Global Health Research, Brighton and Sussex Medical School, Brighton, UK. ⁶Epidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA. ⁷Present address: European Bioinformatics Institute, Hinxton, Cambridge, UK. ⁸Present address: Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany. ⁹Present address: TUM School of Medicine, Technical University of Munich and Klinikum Rechts Der Isar, Munich, Germany. ✉email: m.j.newport@bsms.ac.uk

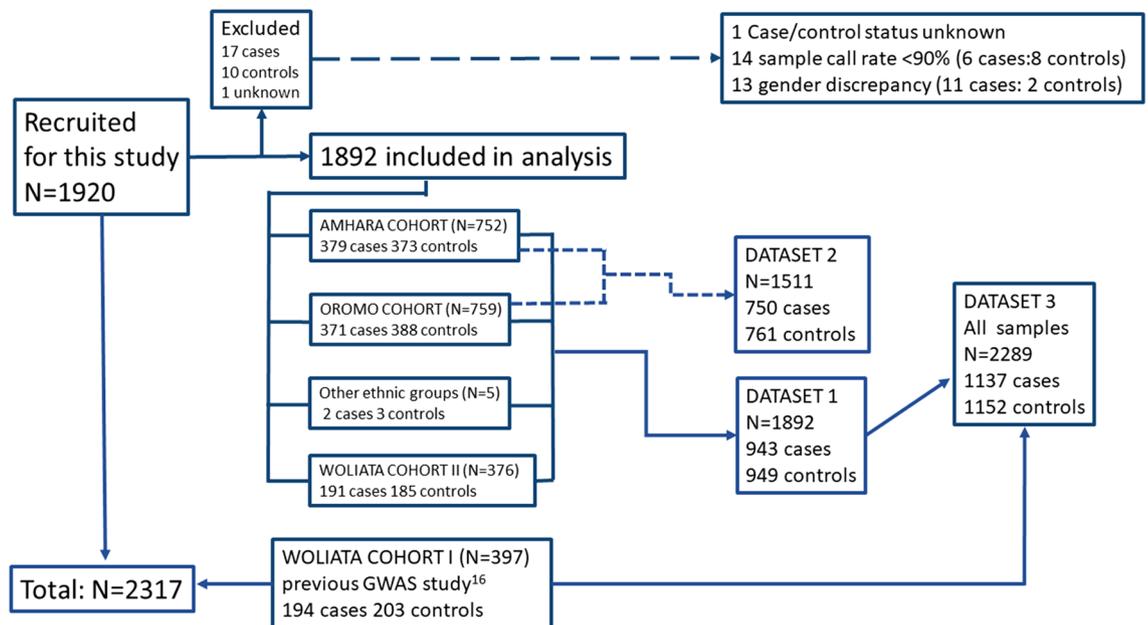


Figure 1. An overview of recruitment numbers and data analysis strategy. A total of 2317 DNA samples were available comprising 1920 new samples from three different Ethiopian ethnic groups (Amhara, Oromo and Wolaita) and 397 samples from a previous GWAS study on podoconiosis undertaken in a different cohort from the Wolaita group¹⁶. Twenty-eight new samples were excluded following QC checks. Data were analysed in three sets: dataset 1 comprised all newly recruited individuals from the three ethnic groups; dataset 2 comprised dataset 1 minus the Wolaita II cohort and ‘other ethnic groups’ samples; and dataset 3 comprised the total samples available for analysis from all cohorts.

The precise nature of the soil trigger is unknown but specific geological and climate conditions are required that lead to the formation of irritant soil, which explains the geographical distribution of podoconiosis in volcanic highland areas¹². Differences in particle and mineral composition between soil from endemic and non-endemic regions have been described, but correlation with the ability of the different soils to induce inflammation measured by haemolytic activity was not demonstrated¹³.

There is also evidence that genetic factors play a role in the pathogenesis of podoconiosis. This was first provided by a study conducted in Ethiopia in 1972, which demonstrated familial clustering of cases¹⁴. A segregation analysis of 59 multi-generational podoconiosis families from the Wolaita ethnic group in southwest Ethiopia indicated an autosomal co-dominant pattern of inheritance with an estimated sibling recurrence risk ratio (λ_s) and heritability of 5.07 and 0.63 respectively¹⁵. A genome-wide association study (GWAS) of 194 cases and 203 controls from the same population identified genome-wide significant association with one single-nucleotide polymorphism (SNP) [rs17612858, additive model: odds ratio, 2.19; 95% CI 1.66 to 2.90; $P = 3.44 \times 10^{-8}$], and genome-wide suggestive association with seven other SNPs ($P < 1.0 \times 10^{-5}$) in the HLA class II region of chromosome 6¹⁶. This finding was confirmed using a family-based association study involving 202 parent–child trios¹⁶.

The aims of this study were to undertake a second GWAS for podoconiosis in a much larger sample size to confirm this HLA association with higher confidence due to increased statistical power, and to determine whether other non-HLA loci were implicated. We included new samples from the Wolaita group and samples from two other ethnic groups, the Amhara and the Oromo, from other regions of Ethiopia where podoconiosis is endemic^{17,18}. We then included the data from the original Wolaita cohort described above in the final analysis to maximise the power of the study. Through these studies we aimed to advance our understanding of the molecular processes involved in the development of a condition that clearly results from gene–environment interactions that could be relevant to other non-communicable diseases, especially those that are also associated with class II HLA genes. Further studies of the genetic basis of podoconiosis are also warranted to develop a point-of-care test to identify those at risk of developing podoconiosis and correctly diagnosis those with the condition. This would allow preventive measures to be taken to avoid the development of podoconiosis, for example through rational allocation of limited resources such as protective footwear, and help differentiate podoconiosis from other forms of tropical lymphoedema such as lymphatic filariasis.

Results

Study subjects, sample sizes and quality control measures. DNA samples were available from a total of 2317 individuals. Of these, 1920 were newly-enrolled during this study from the Amhara, Oromo and Wolaita ethnic groups. The other 397 samples were collected from the Wolaita group during the first podoconiosis GWAS¹⁶ (194 cases and 203 controls). Figure 1 outlines the sources of the samples and the outcomes of the quality control (QC) process. The newly collected Wolaita cohort was designated Wolaita II and the previously collected cohort Wolaita I. Samples from 28 of the 1920 newly-enrolled individuals (17 cases, 10 controls and

Chromosome locus	SNP	Position (bp)*	A1	A2	Nearest gene	Ensembl annotation	MAF	P-value	OR (95% CI)	r ^{2**}
6p21.3	rs9270911 ^{††}	32,572,202	T	C	<i>HLA-DRB1</i>	Regulatory region variant	0.4413	5.512E-10	1.53 (1.34–1.74)	1
6p21.3	rs6906021 [†]	32,626,311	C	T	<i>HLA-DQB1</i>	Downstream gene variant	0.4406	3.478E-09	1.50 (1.31–1.72)	0.49
6p21.3	rs1129740	32,609,105	G	A	<i>HLA-DQA1</i>	Missense variant	0.4889	3.657E-09	0.67 (0.59–0.77)	0.76
6p21.3	rs482205 ^{†‡}	32,576,009	G	T	<i>HLA-DRB1</i>	Intergenic variant	0.3621	3.754E-09	1.51 (1.32–1.73)	0.54
6p21.3	rs1063355 [§]	32,627,714	T	G	<i>HLA-DQB1</i>	3'UTR variant	0.4894	3.831E-09	0.67 (0.59–0.77)	0.76
6p21.3	rs9273349 [§]	32,625,869	T	C	<i>HLA-DQB1</i>	Downstream gene variant	0.4894	3.831E-09	0.67 (0.59–0.77)	0.76
6p21.3	rs643889	32,575,918	T	A	<i>HLA-DRB1</i>	Intergenic variant	0.3614	3.996E-09	1.51 (1.32–1.74)	0.53
6p21.3	rs477515	32,569,691	A	G	<i>HLA-DRB1</i>	Intergenic variant	0.3338	5.033E-09	1.51 (1.32–1.74)	0.64
6p21.3	rs2516049	32,570,400	C	T	<i>HLA-DRB1</i>	Intergenic variant	0.3343	5.847E-09	1.51 (1.32–1.74)	0.64
6p21.3	rs17205647	32,637,418	A	G	<i>HLA-DQB1</i>	Upstream gene variant	0.3721	1.177E-08	1.48 (1.30–1.69)	0.35
6p21.3	rs1071630	32,609,126	T	C	<i>HLA-DQA1</i>	Missense variant	0.4862	1.207E-08	0.68 (0.60–0.78)	0.75
6p21.3	rs6928482 [†]	32,626,249	C	T	<i>HLA-DQB1</i>	Downstream gene variant	0.4447	1.482E-08	1.48 (1.29–1.69)	0.48
6p21.3	rs17843604 [§]	32,620,283	T	C	<i>HLADQA1</i>	Intergenic variant	0.4888	1.497E-08	1.47 (1.29–1.68)	0.8
6p21.3	rs4538748	32,657,505	C	T	<i>HLADQA1</i>	Intergenic variant	0.3743	2.377E-08	1.46 (1.28–1.67)	0.34

Table 1. Genome-wide significant results from a genome-wide association study (GWAS) on podoconiosis undertaken in three ethnic groups (Amhara, Oromo and Wolaita, dataset 1) from Ethiopia. SNP, single nucleotide polymorphism; A1, minor allele; A2, major allele; MAF, minor allele frequency; HWE, Hardy–Weinberg equilibrium. *The position is based on the University of California Santa Cruz (UCSC) human genome build 19 (hg19/GRCCh37). **r²:linkage disequilibrium (LD) with the lead SNP rs9270911. †Independent SNPs (podoconiosis-associated SNPs that were independent of each other (r² < 0.6) within 500 kb sliding windows, identified during annotation (see methods). ‡Lead SNPs: independent SNPs that are independent from each other (r² < 0.1). §SNPs that showed suggestive genome-wide linkage with podoconiosis in the first GWAS¹⁶.

one unknown) were excluded during the QC process (see below). The remaining 1892 subjects comprised: 379 cases and 373 controls from the Amhara ethnic group; 371 cases and 388 controls from the Oromo ethnic group; 191 cases and 185 controls from the Wolaita group (Wolaita II); and 2 cases and 3 controls from other ethnic groups. Demographic and clinical information is summarised in Supplementary Table 1.

Details of sample QC, SNP QC and allele frequency information can be found in Supplementary Tables 2–4. After QC, a total of 2,210,858 autosomal SNPs were genotyped with an average genotyping call rate of 99%. Multidimensional scaling (MDS) analyses by identity by descent (IBD) with African and other 1000 Genome populations¹⁹ revealed clustering in one group, indicating that our samples were of similar ancestry (Supplementary Fig. 1). Quantile–quantile (Q–Q) plotting indicated a genomic inflation factor (λ) of 1.0406 confirming the absence of overt population stratification effects on the analysis (Supplementary Fig. 2).

Following genotyping and QC, data were organised into three datasets for analysis (Fig. 1). Dataset 1 comprised the data derived from all newly-enrolled individuals from all three ethnic groups (n = 1982), to conduct a second GWAS on podoconiosis in Ethiopia; dataset 2 excluded the Wolaita data in case any association identified in dataset 1 was driven by this group (in which the first GWAS found association) so data from the 1511 Amhara and Oromo only was analysed; and dataset 3 comprised data derived from all the samples collected for this study plus the samples from the first podoconiosis GWAS (Wolaita cohort I) that originally identified the HLA association with podoconiosis, to maximise the study power (n = 2289).

Association results. Focusing firstly on confirming the previously reported association between class II HLA variants and podoconiosis, analysis of the newly collected data (dataset 1 comprising 943 cases and 949 controls from all three ethnic groups) identified 14 SNPs that achieved genome-wide significance ($P < 5.0 \times 10^{-8}$) using an additive allelic model (Table 1). All of them were located in the HLA class II locus at 6p21.3 as shown in the Manhattan plot (Fig. 2A). The lead SNP was rs9270911 ($P = 5.51 \times 10^{-10}$; OR = 1.53; 95% CI = 1.34–1.74), a regulatory region variant closest to *HLA-DRB1* and located approximately 48 kb upstream of the previously reported index SNP, rs17612858. Using LocusZoom, SNPs within 500 kb of the lead SNP rs9270911 were plotted based on their GWAS – log₁₀ P-values, NCBI build 37 genomic position, and recombination rates calculated from the 1000 Genomes Project reference data. (Fig. 3)¹⁹. To identify linkage-disequilibrium (LD)-independent SNPs, we utilised the clumping procedure in PLINK (using an r² threshold of 0.1 and a window size of 250 kb) and identified one independent signal, rs9270911. LD-analysis of the significantly associated variants indicated that they were moderately correlated (Supplementary Fig. 3).

To investigate whether the HLA class II association identified in this new GWAS was driven by the Wolaita group—the population in which the HLA class II association was previously reported—we excluded the Wolaita II data, analysing only the Amhara and Oromo data (dataset 2, 1511 samples, Fig. 1). We identified five HLA class II SNPs that reached genome-wide significance and the lead SNP was rs17205647 ($P = 6.469E-09$; OR = 1.56; 95% CI = 1.35–1.82; Fig. 2B and Supplementary Table 5). This variant is located upstream of *HLA-DQB1* and features in Table 1.

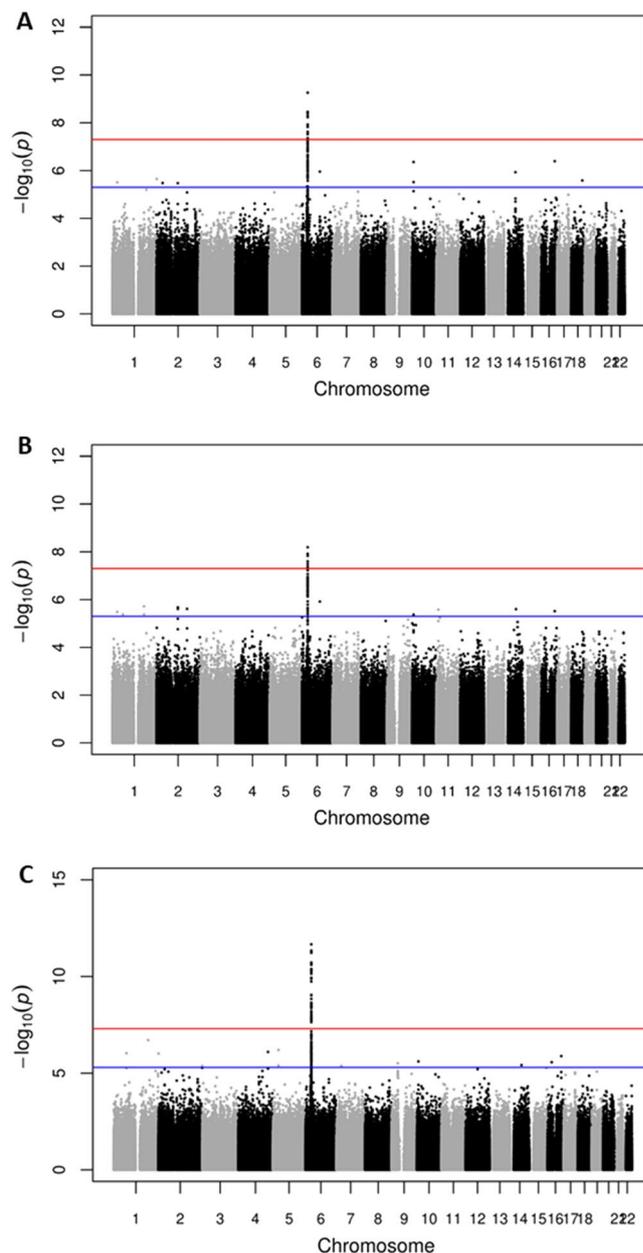


Figure 2. Manhattan plots showing genome-wide association with podoconiosis in Ethiopia. **(A)** dataset 1 comprising combined new samples from the Amhara, Oromo and Wolaita ethnic groups; (N = 1892, 943 cases and 949 controls); **(B)** dataset 2 from the Amhara and Oromo ethnic groups (N = 1511, 752 cases and 759 controls) and **(C)** dataset 3 comprising all available samples i.e. dataset 1 plus the samples from the previously-published podoconiosis GWAS undertaken in the Wolaita group (N = 2289, 1137 cases and 1152 controls). For each plot, the x-axis indicates chromosomal position and the y-axis indicates the $-\log_{10} P$ value. Points between the blue and red lines shows suggestive association ($P < 5.0 \times 10^{-7}$) whereas points above the red line shows genome-wide significance ($P < 5.0 \times 10^{-8}$) results. The R programme was used to generate the plots⁴⁸.

Finally, we undertook a genome-wide association analysis of all available data including those generated from the Wolaita I samples in which the HLA class II association was first reported (2289 samples, dataset 3, Fig. 1). A total of 47 SNPs reached genome-wide-significance, and the lead SNP was rs9270911 ($P = 2.25 \times 10^{-12}$; Fig. 2C and Supplementary Table 6). Two SNPs (rs17205647 and rs4538748) were significantly associated in all three dataset analyses in this study. Three SNPs (rs9273349, rs1063355, and rs17843604) which showed suggestive association with podoconiosis in the original GWAS study (i.e. genome-wide significance P values of $> 5.0 \times 10^{-8}$ but $< 1.0 \times 10^{-5}$)¹⁶, showed genome-wide significance in this more highly-powered analysis. Analysis of the larger combined dataset (N = 2289) did not identify any new loci outside the HLA class II region.

Since the lead SNP in the original study (rs17612858) was not included in the Illumina HumanOmni2.5 array used in this study, we determined the pairwise LD value between our top SNP (rs9270911) and rs9273349 which

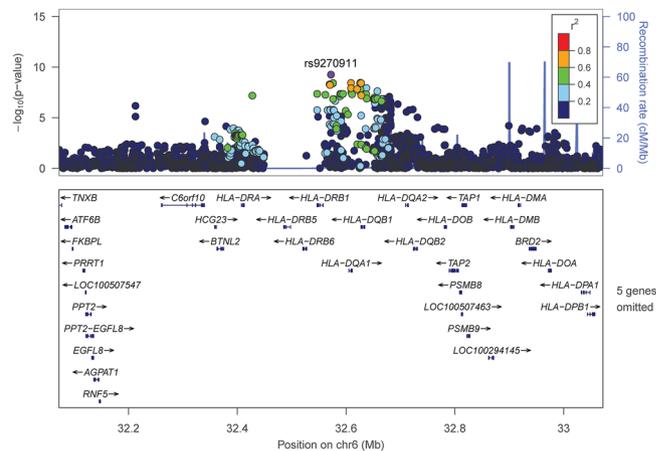


Figure 3. Association signals across a 1 Megabase region of the HLA class II region on chromosome 6 surrounding SNP rs9270911. Using LocusZoom, SNPs within 500 kb of the lead SNP rs9270911 were plotted based on their GWAS $-\log_{10} P$ -values, NCBI build 37 genomic position, and recombination rates calculated from the 1000 Genomes Project reference data. Each dot represents a SNP. The lead SNP rs9270911 is represented by a purple dot. The color scale of r^2 values is used to label SNPs based on their degree of linkage disequilibrium with rs9270911. Most SNPs are dark blue indicating low LD ($r^2 < 0.2$). The left y axis represented negative $\log_{10} P$ values, the right y-axis represents the recombination rate in centiMorgans (cM) per megabase (Mb). Genes found in the region are shown in relative position under the plot; arrows indicate the direction of transcription.

was the next best-scoring SNP in the original study ($P = 3.49 \times 10^{-07}$) that was also represented on the Illumina HumanOmni2.5 array. The LD value between these two variants was $r^2 = 0.75$ ($D' = 0.996046$), indicating that the SNPs are in high LD with each other.

Functional mapping and annotation results. Functional mapping and annotation of the 14 genome-wide significant SNPs using FUnctional Mapping (FUMA)²⁰ identified four independent ($r^2 < 0.6$) significant SNPs (rs9270911, rs482205, rs6928482, rs6906021), of which two (rs9270911, rs482205) were found to be lead SNPs ($r^2 < 0.1$). All were located within the genomic locus 6:32564784-6:32667548. *HLA-DRB1* co-localises to this region at position 6:32578769-6:589836.

A total of 173 unique SNPs in this locus were in LD ($r^2 \geq 0.6$) with the four independent significant SNPs, and were functionally annotated. These SNPs were mostly located in intergenic regions ($n = 140$; 80.9%) (Fig. 4A). Except for one SNP (rs28366343), all SNPs had a combined annotation dependent depletion (CADD) score of less than 12.37, indicating that they are not deleterious (Fig. 4B). Twelve SNPs (6.8%) had a RegulomeDB score of less than 4 suggesting the SNPs have a regulatory role (Fig. 4C). Except for one, all SNPs ($n = 172$; 99.4%) were located in an open chromatin state as indicated by a minimum chromatin state less than 7 (Fig. 4D). Review of the 173 SNPs in LD with the four independent associated SNPs in the NHGRI-EBI GWAS catalogue²¹ revealed associations with other immune-mediated diseases such as inflammatory bowel disease, type 1 diabetes, multiple sclerosis, asthma, hayfever and allergy, and responses to Epstein-Barr virus and hepatitis B vaccine (Supplementary Table 7).

Discussion

Using a genome-wide approach, we have confirmed the reported association between HLA class II variants and podoconiosis in an independent study that included 1892 new samples from three ethnic groups from Ethiopia, where podoconiosis is endemic. We found significant genome-wide association with 14 common variants in the HLA class II region on chromosome 6p21.3. The strongest association was with rs9270911, a regulatory variant located near *HLA-DRB1*. Other strongly associated variants were located near *HLA-DQA1* and *HLA-DQB1* which are contiguous and in LD with *HLA-DRB1*. The association was confirmed independently in the Oromia and Amhara groups, which have not previously been studied. The significance of the association was strengthened when the original Wolaita samples were included in the analysis (making a total of 2289 samples) with 47 common variants in the HLA class II region showing significant association with podoconiosis. This sample size was almost six times larger than that of the first podoconiosis GWAS (397 individuals) yet no variants outside the HLA class II region were found to be associated with podoconiosis suggesting a single major locus is predominantly responsible for genetic susceptibility to podoconiosis. This is consistent with a segregation analysis undertaken on multicaser Wolaita families¹⁵, but even larger association studies would be required to identify loci with minor effects.

Genes within the HLA class II region encode numerous molecules that have critical functions in the adaptive immune system. The polymorphic *HLA-DRB1* encodes the β chain of the HLA class II glycoprotein HLA-DR, which together with the α chain (encoded by *HLA-DRA* which is not polymorphic) forms a functional

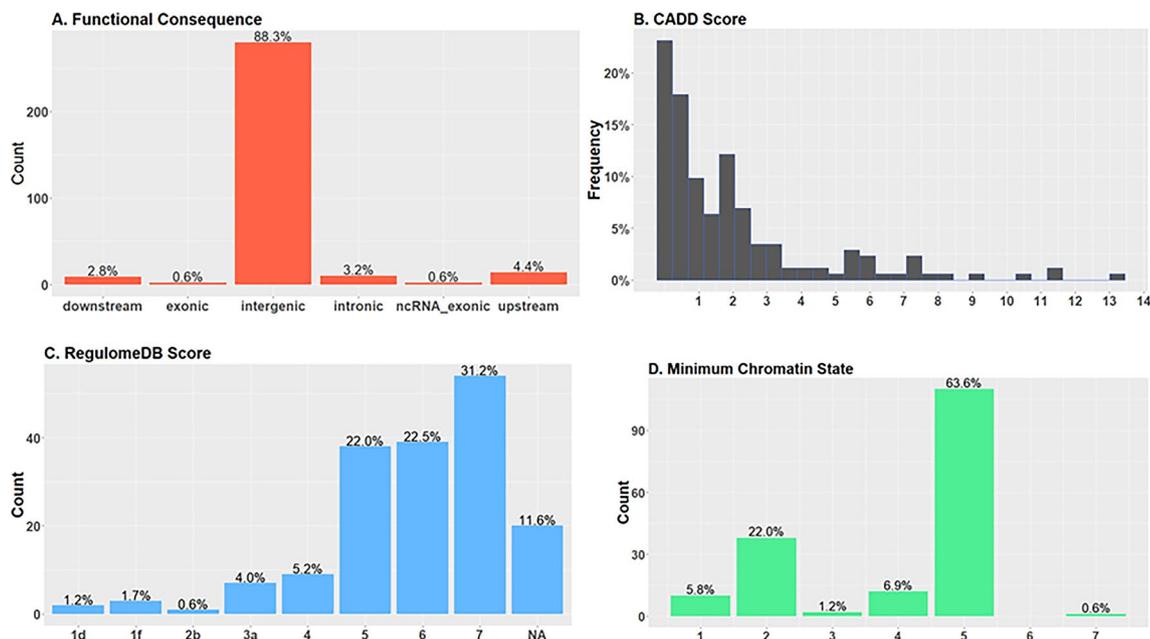


Figure 4. Functional annotation of 173 single nucleotide polymorphisms (SNPs) found to be in linkage disequilibrium ($r^2 > 0.6$) with the four independent SNPs showing genome-wide significant association with podoconiosis. **(A)** Distribution of functional consequences of SNPs in genomic risk loci **(B)** Distribution of combined annotation dependent depletion (CADD) score for SNPs in genomic risk loci. A score greater than 12.37 indicates deleterious SNPs. **(C)** Distribution of RegulomeDB score for SNPs in genomic risk loci. A score less than 4 indicates regulatory function. **(D)** Minimum chromatin state for 127 tissue/cell types for SNPs in genomic risk loci. A score less than 7 indicates open chromatin state, indicating higher accessibility.

antigen-binding heterodimer expressed on the surface of antigen presenting cells (APCs) such as macrophages, dendritic cells and B lymphocyte cells. Protein antigens processed by APCs into smaller peptides are loaded onto HLA molecules and presented to T-lymphocyte cells. Interaction between the HLA molecule, the peptide and the T-cell receptor (TCR) activates the T-cell to initiate an immune response specific to the antigen. Class II HLA molecules typically present peptide derived from exogenous (foreign) proteins to CD4⁺ T-cells whilst class I HLA molecules present endogenous ('self') peptides to CD8⁺ T cell. *HLA-DQA1* and *HLA-DQB1* encode α and β -chains for HLA-DQ molecules which have a similar role in antigen presentation to HLA-DR.

Association between HLA gene alleles and haplotypes and immune-mediated disorders is well established. These include autoimmune diseases such as type 1 diabetes mellitus (T1D), type 1 narcolepsy (T1N), coeliac disease and multiple sclerosis; infectious diseases such as leprosy and malaria; and aberrant reactions to molecules including beryllium and the anti-retroviral drug abacavir. In some cases there is a very strong predictable association with a single HLA allele—for example reaction to abacavir is associated with the HLA-B*5701 allele only and individuals can be genotyped for this variant before starting treatment²². In other disorders, such as multiple sclerosis, a number of haplotypes comprising different alleles are associated with susceptibility to disease. This may reflect epistatic interactions between the alleles, or selection of certain combinations of alleles that are more protective, through the ability to present a broader range of pathogen antigens than randomly inherited alleles would (bearing in mind the HLA has evolved to protect against pathogens)^{23,24}. It is, therefore, important to extend our studies in podoconiosis from SNP associations to more informative HLA allele and haplotype associations towards identifying functionally important variants that underlie the molecular mechanisms of disease. Our previous work using direct HLA typing found association between the HLA-DRB1*0701, DQA1*0201 and DQB1*0202 alleles, and the HLA-DRB1*0701-DQB1*0202 haplotype and podoconiosis¹⁶. We also performed HLA imputation using SNPs to demonstrate that HLA class II alleles can be predicted from SNP genotype data with a high level of accuracy at intermediate (two-digit) resolution in an African population²⁵. Further HLA typing and haplotype studies are being undertaken.

Despite the confounding effect of LD in the region, progress has been made towards understanding the mechanisms for some of these associations that will inform further studies as we seek to understand the molecular pathogenesis of podoconiosis. One of the fundamental postulates of autoimmunity is that the immune system mistakes self-antigens for foreign antigens to which the immune system has previously been exposed to and mounts a cross-reacting response to cause inflammation. Molecular mimicry is one mechanism by which this can happen, where the foreign antigen is derived from a microbe but closely resembles a self antigen²⁶. The autoimmune disease T1N is a well-characterised example, where hypothalamic cells that secrete hypocretin (a neuropeptide that stimulates wakefulness) are destroyed by auto-reactive CD4⁺ T-cells. The disorder is associated with the HLA-DQ heterodimer DQ0602 comprising α - and β -chains encoded by HLA-DQA1*0102 and HLA-DQB1*0602 alleles respectively, with risk further increased for HLA-DQB1*0602 homozygotes²⁷. The incidence of T1N cases increased after the 2009 H1N1 influenza pandemic and cases followed H1N1 influenza

vaccination²⁸. Screening of hypocretin and H1N1 influenza hemagglutinin identified a common peptide that could bind to HLA-DQ0602 and initiate T-cell mediated responses *in vitro*²⁹. Whilst molecular mimicry involves peptides usually derived from microorganisms and there is no epidemiological evidence linking podoconiosis with infection, soil harbours numerous microbes many of which exist in a viable but non-culturable state and have yet to be characterised³⁰, as well as other organic matter that could potentially be the source of a cross reactive peptide in podoconiosis.

Soil is also rich in metal ions and these can also trigger inflammatory disorders defined by their HLA associations. In chronic beryllium disease, granulomatous lung pathology develops in response to inhaled beryllium (an alkaline earth metal) in individuals carrying *HLA-DPBI* alleles that encode glutamic acid at position 69 of the HLA-DP β -chain. This allows positively-charged beryllium molecules to bind in the antigen-binding pocket to alter both the charge and conformation of the HLA-DP molecule allowing it to bind naturally-occurring peptides to create a ligand for pathogenic CD4⁺ T-cells³¹. Podoconiosis arises from prolonged contact with soil that is rich in metal ions. Alternatively, a soil mineral could induce changes in the structure or charge of a self-peptide or epitope, for example through post-translational modification, to create a neopeptide that is immunogenic, making this an attractive avenue to pursue in ongoing research on the soils that are involved in the aetiology of podoconiosis.

In many autoimmune diseases the self-peptide has yet to be identified, but progress has been made towards understanding the basis of the HLA association. Linkage between *HLA-DRB1* and T1D, in which an autoimmune process destroys the insulin-secreting β -islet cells of the pancreas, was well-established by family studies decades ago (and the role of HLA since confirmed by large-scale association studies³²). Further studies at the time revealed a stronger association with *HLA-DQB1*³³. Alleles that encoded the neutral amino acids serine, alanine or valine at position 57 of HLA-DQ1 were associated with susceptibility whilst aspartic acid at this position correlated with resistance to T1D. This finding was replicated in a Spanish population study, which also discovered that *HLA-DQA1* alleles encoding arginine at position 52 were associated with T1D³⁴. These amino acid positions are located within the antigen-binding pockets of the HLA molecules suggesting amino acid substitutions influence molecular interactions within the pocket and alter the repertoire of peptides they present. Variations at other amino acid positions within the peptide binding pockets of both HLA-DQB1 and HLA-DRB1 molecules have also been shown to influence susceptibility and certain heterozygote HLA combinations appear to increase the risk of T1D in a synergistic manner³⁵. Work is ongoing to identify the peptide that triggers immune response in diabetes—a number of candidates including autoantigens such as insulin are under investigation³⁶.

Further elucidation of the HLA alleles that are associated with susceptibility or resistance to podoconiosis¹⁶ and amino acid sequence analysis are required to further understand the role of HLA in the pathogenesis of podoconiosis. A better understanding of the HLA molecular structure could also allow prediction of the antigen(s) and its epitopes, which are unknown for podoconiosis, in parallel with ongoing studies in soil from endemic areas. In addition to the scientific advances made through the study of the molecular pathogenesis of podoconiosis and its contribution towards understanding gene environment interactions in complex traits, there are practical public health implications of this molecular work that have a wider impact for affected communities and disease control globally. Limited resources such as protective footwear can already be targeted to families with a history of podoconiosis. Confirmation of the class II HLA association could further refine this approach to identify those at risk who do not have a family history. Careful engagement with communities would be required to avoid increasing stigma associated with podoconiosis³⁷. Further effort may now be invested in developing a diagnostic test for podoconiosis, which is currently lacking. The condition affects up to 8% of people in affected communities³ and identification of a genetic marker, if not the causal variant, would refine identification of those at risk in whom podoconiosis could be prevented through consistent use of footwear. This would represent a major advance towards disease elimination.

There is a relative dearth of genetic data derived from African populations³⁸ and this study has contributed detailed genotype data from three different ethnic groups to existing datasets. Modern humans evolved out of Africa and African populations are genetically more diverse than any other populations. Being older, population structure and linkage disequilibrium patterns are more complex and many of the tools used to study genetics in health and disease were developed in European-ancestry populations which do not fully capture this diversity^{39,40}. This issue has been partly addressed through the availability of denser arrays that give better genome coverage, including the development of a 2.5 M African-specific GWAS array by the Human Health and Heredity in Africa (H3 Africa) Consortium⁴¹. However, there remains a need for population specific data to give better representation of the approximately 2000 ethnolinguistic groups that live in Africa³⁹.

Detailed HLA type data are lacking for African populations despite the role of HLA in immunity to common fatal infectious diseases such as malaria, HIV and tuberculosis. Better understanding of HLA in this context is also relevant to the development of new or better vaccines for such diseases. A future application of our data will be to develop a tool to predict HLA types from selected genotype data once we have HLA typed our study cohort, obviating the need for the time-consuming and expensive methods currently required for HLA typing prior to organ transplantation.

In conclusion, this project replicates the finding that HLA class II gene loci have significant associations with podoconiosis. The involvement of immune mechanisms in podoconiosis susceptibility seems plausible⁴² and further research is required to identify causal variants and further characterise the molecular basis of the association.

Methods

Study population and datasets. This was a population-based case–control study involving individuals from three ethnic groups (Amhara, Oromo, and Wolaita) from four different zones of Ethiopia. The Amhara samples were collected from 12 *woredas* (districts) in the East Gojjam and West Gojjam Zones. The Oromo

samples were collected from 6 *woredas* in the East Wellega Zone. The Wolaita samples comprised a newly collected cohort collected from 9 *woredas* within the Wolaita Zone (Wolaita II) and those from the original cohort (Wolaita I) which are described in more detail elsewhere¹⁶.

Three genome-wide analyses were conducted (Fig. 1) using (a) the Amhara, Oromo and Wolaita II samples collected during this study aiming to confirm the results from the first GWAS and determine whether loci outside the HLA class II region could be identified (dataset 1), (b) the Amhara and Oromo samples, but not the Wolaita II samples, to evaluate whether any positive findings from dataset 1 were driven by the Wolaita data since this was the group in which the HLA association was first identified (dataset 2) and c) samples from all individuals recruited during this study combined with the Wolaita I samples to examine the strength of the association and to identify novel signals (dataset 3). Sociodemographic characteristics and clinical information were collected for newly enrolled participants.

Case–control definition. Cases and controls were defined as described previously¹⁶. Briefly, cases were adults (aged 18 and above) with lymphoedema typical of podoconiosis and were resident in the study area for at least 5 years. Controls were healthy adults aged 50 and above who were resident in the podoconiosis endemic area for a minimum of 25 years and had no family history of podoconiosis and did not consistently use shoes. The age limit was chosen to allow sufficient contact with the volcanic soil over time for susceptible individuals to develop disease and not be mis-classified as controls¹⁶. Clinical assessment and disease staging were conducted by experienced nurses working at either the International Orthodox Christian Charities (IOCC) Debre Markos branch (Gojjam), the Catholic Church Clinic in Nekemte (Wellega), or Mossy Foot International in Wolaita using a validated clinical staging system for podoconiosis⁴³. There is no proven diagnostic test for the disease.

Power calculations. The online version of the GAS Power Calculator⁴⁴ was used to calculate the power of the study using the following assumptions in an additive model: significance level threshold of 5.0×10^{-8} and disease prevalence of 4%⁴. Different scenarios for the MAF and Odds Ratio (OR) were also used to estimate power for this study.

Genotyping and quality control (QC). DNA extracted from saliva samples using the Oragene saliva DNA kits (DNA Genotek, Ontario, Canada) was genotyped using the Illumina HumanOmni25-8 v.1–2 chip at the Wellcome Trust Sanger Institute (WTSI), Hinxton, UK. Standard quality control (QC) procedures⁴⁵ were performed using PLINK v1.9⁴⁶. Samples were excluded for cryptic relatedness/duplication ($\text{pi_hat} > 0.09$), gender discrepancy (based on the presence of heterozygous haploid genotypes), autosomal heterozygosity (> 3 SD from the mean), and call rate ($< 90\%$). Outlier samples were also checked visually from multidimensional scaling (MDS) and principal component analysis (PCA) plots and were removed from downstream analysis. SNPs that deviate Hardy–Weinberg equilibrium (HWE), call rate ($< 95\%$), and Minor Allele Frequency (MAF) ($< 5\%$) were excluded. In addition, insertions and deletions (indels), copy number variants (CNVs), non-autosomal and non-biallelic SNPs were removed.

Single-marker association analysis. Logistic regression under an additive model was conducted in PLINK adjusting for the first 10 principal components (PCs) as covariates. Age and shoe-wearing habit were not considered as covariates since they were controlled during participant recruitment. Genome-wide significance was set to $P < 5.0 \times 10^{-8}$ to account for multiple testing. Intensity cluster plots for the top associated SNPs were generated and manually examined to ensure only SNPs of high genotyping quality were considered for follow-up; poorly-called SNPs were removed from downstream analysis. Top associated SNPs showing good clustering were annotated using the Ensembl Variant Effect Predictor (VEP)⁴⁷ (build GRCh37). Quantile–Quantile (Q–Q) and Manhattan plots were generated using R statistical software⁴⁸. Regional association plots were made using LocusZoom⁴⁹. The linkage disequilibrium (LD) structure for the genome-wide significant SNPs ($P < 5.0 \times 10^{-8}$) was plotted in Haploview v4.2⁵⁰ using a haplotype definition described by Gabriel et al.⁵¹.

Haplotype association testing. In order to determine the haplotypes driving our top associated SNPs, we conducted haplotype-based analysis. Briefly, SNPs spanning chromosome 6 region 28,889,000–33,055,000 bps were extracted using PLINK. Haplotype and missing genotype inference as well as haplotype association tests were carried out using Beagle v3.2.2^{52,53}. The haplotype clusters which were found to be significantly associated with podoconiosis were further analyzed using the Beagle cluster2haps program to identify allele sequences that defined the haplotypes.

Identification of genomic risk loci. We employed FUNctional Mapping and annotation (FUMA), an online platform that combines various bioinformatics tools and data sources, on the summary statistics of the primary replication dataset to functionally annotate, map and prioritize the top associated variants²⁰. First, genome-wide significant SNPs ($P < 5.0 \times 10^{-8}$) that were independent of each other ($r^2 < 0.6$) within a 500 kb sliding window were identified (independent significant SNPs) using PLINK's clumping procedure⁴⁶. Lead SNPs were identified from the independent significant SNPs if they were independent from each other ($r^2 < 0.1$). Candidate SNPs to be used for functional annotations and gene mapping were identified as all SNPs (either from the GWAS summary statistics or phase 3 1000G AFR reference population) that had a MAF of greater than 1% and were in LD ($r^2 \geq 0.8$) with at least one of the independent significant SNPs¹⁹. If the LD blocks for the independent significant SNPs were 250 kb up or downstream from the most up- or downstream SNPs from each LD block, they were merged as one genomic locus. Thus, each genomic locus could contain multiple independent

significant SNPs and lead SNPs. SNPs that were previously reported to be associated with a phenotype in the NHGRI-EBI GWAS catalog²¹, and located in the same genomic loci were identified to provide further insight about the top associated variants.

Functional annotation. Functional consequences of candidate SNPs were determined using the combined annotation dependent depletion (CADD) score⁵⁴, RegulomeDB score⁵⁵, and the 15-core chromatin state^{56,57} as implemented in FUMA. The CADD score estimates the deleteriousness of SNPs, whereas the latter two estimate regulatory functions. A CADD score greater or equal to 12.37 indicates a more deleterious variant, whereas a lower RegulomeDB score (on a scale of 1 to 7) indicates variants with a regulatory function. A 15-core chromatin state value of less than or equal to 7 (of 15 categorical states) is an indication of an open chromatin state (accessible genomic region).

Ethical approval and informed consent. Ethical approval was obtained from the Armauer Hansen Research Institute (AHRI)/All Africa Leprosy and Tuberculosis Rehabilitation and Training Centre (ALERT) Ethics Review Committee (Ref:PO20/12) and the Ethiopian National Research Ethics Review Committee (Ref 310/577/06). Research Governance approval was given by Brighton and Sussex Medical School Research Governance and Ethics Committee (Ref 14/066/NEW). Written informed consent was obtained from each participant before enrolling them in the study. Rapid ethical appraisals was undertaken in all three study communities before sample collection began^{58,59}. This qualitative approach used focus groups and in-depth interviews to explore the views and concerns of the community and other stakeholders regarding the study and research more generally. The findings allowed the design of a contextualized consent process whilst meeting international ethics standards for biomedical research involving human subjects.

Received: 17 July 2020; Accepted: 29 December 2020

Published online: 08 February 2021

References

- Davey, G., Tekola, F. & Newport, M. J. Podoconiosis: non-infectious geochemical elephantiasis. *Trans. R. Soc. Trop. Med. Hyg.* **101**, 1175–1180. <https://doi.org/10.1016/j.trstmh.2007.08.013> (2007).
- Deribe, K. *et al.* The global atlas of podoconiosis. *Lancet Glob. Health* **5**, e477–e479. [https://doi.org/10.1016/S2214-109X\(17\)30140-7](https://doi.org/10.1016/S2214-109X(17)30140-7) (2017).
- Deribe, K., Cano, J., Trueba, M. L., Newport, M. J. & Davey, G. Global epidemiology of podoconiosis: a systematic review. *PLoS Neglect. Trop. Dis.* **12**, e0006324 (2018).
- Deribe, K. *et al.* Epidemiology and individual, household and geographical risk factors of podoconiosis in Ethiopia: results from the first nationwide mapping. *Am. J. Trop. Med. Hyg.* **92**, 148–158. <https://doi.org/10.4269/ajtmh.14-0446> (2015).
- Deribe, K. *et al.* Estimating the number of cases of podoconiosis in Ethiopia using geostatistical methods. *Wellcome Open Res.* **2**, 78. <https://doi.org/10.12688/wellcomeopenres.12483.2> (2017).
- Addiss, D. G. & Brady, M. A. Morbidity management in the global programme to eliminate lymphatic filariasis: a review of the scientific literature. *Filaria J.* **2**, 2. <https://doi.org/10.1186/1475-2883-6-2> (2007).
- Tekola, F., HaileMariam, D. & Davey, G. Economic costs of podoconiosis (endemic non-filarial elephantiasis) in Wolaita Zone, Ethiopia. *Trop. Med. Int. Health* **11**, 1136–1144 (2006).
- Tora, A., Franklin, H., Deribe, K., Reda, A. A. & Davey, G. Extent of podoconiosis-related stigma in Wolaita Zone, Southern Ethiopia: a cross-sectional study. *Springerplus* **3**, 647. <https://doi.org/10.1186/2193-1801-3-647> (2014).
- Yakob, B., Deribe, K. & Davey, G. High levels of misconceptions and stigma in a community highly endemic for podoconiosis in southern Ethiopia. *Trans. R. Soc. Trop. Med. Hyg.* **102**, 439–444 (2008).
- Bartlett, J. *et al.* Depression and disability in people with podoconiosis: a comparative cross-sectional study in rural Northern Ethiopia. *Int. Health* **8**, 124–131. <https://doi.org/10.1093/inthealth/ihv037> (2016).
- Price, E. W. *Podoconiosis: Non-filarial Elephantiasis* (Oxford Medical Publications, Oxford, 1990).
- Deribe, K. *et al.* Mapping and modelling the geographical distribution and environmental limits of podoconiosis in Ethiopia. *PLoS Negl. Trop. Dis.* **9**, e0003946. <https://doi.org/10.1371/journal.pntd.0003946> (2015).
- Le Blond, J. S. *et al.* Haemolytic activity of soil from areas of varying podoconiosis endemicity in Ethiopia. *PLoS ONE* **12**, e0177219. <https://doi.org/10.1371/journal.pone.0177219> (2017).
- Price, E. W. A possible genetic factor in non-filarial elephantiasis of the lower legs. *Eth. Med. J.* **10**, 87–93 (1972).
- Davey, G. *et al.* Podoconiosis: a tropical model for gene-environment interactions?. *Trans. R. Soc. Trop. Med. Hyg.* **101**, 91–96 (2007).
- Tekola Ayele, F. *et al.* HLA class II locus and susceptibility to podoconiosis. *N. Engl. J. Med.* **366**, 1200–1208 (2012).
- Molla, Y. B., Tomczyk, S., Amberbir, T., Tamiru, A. & Davey, G. Podoconiosis in East and West Gojam Zones, northern Ethiopia. *PLoS Negl. Trop. Dis.* **6**, e1744. <https://doi.org/10.1371/journal.pntd.0001744> (2012).
- Bekele, K. *et al.* Burden assessment of podoconiosis in Wayu Tuka woreda, east Wollega zone, western Ethiopia: a community-based cross-sectional study. *BMJ Open* **6**, e012308. <https://doi.org/10.1136/bmjopen-2016-012308> (2016).
- 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74. doi:<https://doi.org/10.1038/nature15393> (2015).
- Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826. <https://doi.org/10.1038/s41467-017-01261-5> (2017).
- MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901. <https://doi.org/10.1093/nar/gkw1133> (2017).
- Tangamornsuksan, W. *et al.* Association of HLA-B*5701 genotypes and abacavir-induced hypersensitivity reaction: a systematic review and meta-analysis. *J. Pharm. Pharm. Sci.* **18**, 68–76. <https://doi.org/10.18433/j39s3t> (2015).
- Meyer, D., Aguilar, V. R., Bitarello, B. D., Brandt, D. Y. & Nunes, K. A genomic perspective on HLA evolution. *Immunogenetics* **70**, 5–27. <https://doi.org/10.1007/s00251-017-1017-3> (2018).
- Trowsdale, J. & Knight, J. C. Major histocompatibility complex genomics and human disease. *Annu Rev. Genom. Hum. Genet.* **14**, 301–323. <https://doi.org/10.1146/annurev-genom-091212-153455> (2013).

25. Tekola Ayele, F. *et al.* Prediction of HLA class II alleles using SNPs in an African population. *PLoS ONE* **7**, e40206. <https://doi.org/10.1371/journal.pone.0040206> (2012).
26. Rojas, M. *et al.* Molecular mimicry and autoimmunity. *J. Autoimmun.* **95**, 100–123. <https://doi.org/10.1016/j.jaut.2018.10.012> (2018).
27. Mignot, E. *et al.* Complex HLA-DR and -DQ interactions confer risk of narcolepsy-cataplexy in three ethnic groups. *Am. J. Hum. Genet.* **68**, 686–699. <https://doi.org/10.1086/318799> (2001).
28. Partinen, M. *et al.* Narcolepsy as an autoimmune disease: the role of H1N1 infection and vaccination. *Lancet Neurol.* **13**, 600–613. [https://doi.org/10.1016/S1474-4422\(14\)70075-4](https://doi.org/10.1016/S1474-4422(14)70075-4) (2014).
29. Luo, G. *et al.* Autoimmunity to hypocretin and molecular mimicry to flu in type 1 narcolepsy. *Proc. Natl. Acad. Sci. USA* **115**, E12323–E12332. <https://doi.org/10.1073/pnas.1818150116> (2018).
30. Li, L., Mendis, N., Trigui, H., Oliver, J. D. & Faucher, S. P. The importance of the viable but non-culturable state in human bacterial pathogens. *Front. Microbiol.* **5**, 258. <https://doi.org/10.3389/fmicb.2014.00258> (2014).
31. Fontenot, A. P., Falta, M. T., Kappler, J. W., Dai, S. & McKee, A. S. Beryllium-induced hypersensitivity: genetic susceptibility and neoantigen generation. *J. Immunol.* **196**, 22–27. <https://doi.org/10.4049/jimmunol.1502011> (2016).
32. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
33. Todd, J. A., Bell, J. I. & McDevitt, H. O. HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* **329**, 599–604 (1987).
34. Gutierrez-Lopez, M. D. *et al.* Susceptibility to type 1 (insulin-dependent) diabetes mellitus in Spanish patients correlates quantitatively with expression of HLA-DQ alpha Arg 52 and HLA-DQ beta non-Asp 57 alleles. *Diabetologia* **35**, 583–588. <https://doi.org/10.1007/bf00400488> (1992).
35. Hu, X. *et al.* Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat. Genet.* **47**, 898–905. <https://doi.org/10.1038/ng.3353> (2015).
36. Purcell, A. W., Sechi, S. & DiLorenzo, T. P. The evolving landscape of autoantigen discovery and characterization in type 1 diabetes. *Diabetes* **68**, 879–886. <https://doi.org/10.2337/dbi18-0066> (2019).
37. Tora, A., Mengiste, A., Davey, G. & Semrau, M. Community involvement in the care of persons affected by podoconiosis—a lesson for other skin NTDs. *Trop. Med. Infect. Dis.* <https://doi.org/10.3390/tropicalmed3030087> (2018).
38. H3 Africa Consortium *et al.* Research capacity. Enabling the genomic revolution in Africa. *Science* **344**, 1346–1348, doi:<https://doi.org/10.1126/science.1251546> (2014).
39. Tishkoff, S. A. *et al.* The genetic structure and history of Africans and African Americans. *Science* **324**, 1035–1044 (2009).
40. Choudhury, A. *et al.* High-depth African genomes inform human migration and health. *Nature* **586**, 741–748. <https://doi.org/10.1038/s41586-020-2859-7> (2020).
41. Human Health and Heredity in Africa (H3 Africa) Consortium.
42. Alcantara, D. R. *et al.* Multiplexed gene expression analysis of HLA class II-associated podoconiosis implicates chronic immune activation in its pathogenesis. *Trans R Soc Trop Med Hyg* **in press** (2020).
43. Tekola, F., Ayele, Z., Mariam, D. H., Fuller, C. & Davey, G. Development and testing of a de novo clinical staging system for podoconiosis (endemic non-filarial elephantiasis). *Trop. Med. Int. Health* **13**, 1277–1283 (2008).
44. Johnson, J. L. & Abecasis, G. R. *GAS Power calculator: web-based power calculator for genetic association studies* (bioRxiv, 2017).
45. Weale, M. E. Quality control for genome-wide association studies. *Methods Mol. Biol.* **628**, 341–372. https://doi.org/10.1007/978-1-60327-367-1_19 (2010).
46. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
47. McLaren, W. *et al.* The ensembl variant effect predictor. *Genome Biol.* **17**, 122. <https://doi.org/10.1186/s13059-016-0974-4> (2016).
48. R Core Team. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, Vienna, Austria, 2013).
49. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337. <https://doi.org/10.1093/bioinformatics/btq419> (2010).
50. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
51. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229. <https://doi.org/10.1126/science.1069424> (2002).
52. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097. <https://doi.org/10.1086/521987> (2007).
53. Browning, B. L. & Browning, S. R. Efficient multilocus association testing for whole genome association studies using localized haplotype clustering. *Genet. Epidemiol.* **31**, 365–375. <https://doi.org/10.1002/gepi.20216> (2007).
54. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315. <https://doi.org/10.1038/ng.2892> (2014).
55. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797. <https://doi.org/10.1101/gr.137323.112> (2012).
56. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330. <https://doi.org/10.1038/nature14248> (2015).
57. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74, doi:<https://doi.org/10.1038/nature11247> (2012).
58. Tekola, F. *et al.* Tailoring consent to context: designing an appropriate consent process for a biomedical study in a low income setting. *PLoS Negl. Trop. Dis.* **3**, e482. <https://doi.org/10.1371/journal.pntd.0000482> (2009).
59. Gebresilase, T. T. *et al.* Rapid ethical appraisal: a tool to design a contextualized consent process for a genetic study of podoconiosis in Ethiopia [version 1; peer review: 3 approved with reservations]. *Wellcome Open Res.* <https://doi.org/10.12688/wellcomeopenres.12613.1> (2017).

Acknowledgments

We thank the International Orthodox Christian Charities (IOCC) in Debre Markos, the Catholic Church in Wellega and the International Mossy Foot International in Sodo, Wolaita for assistance with participant enrolment and data and sample collection in Ethiopia. We are also indebted for the regional health bureaus for their support and to all the study participants. This study was funded by the MRC UK (Grant Ref: MR/J008621/1). This work was also supported in part by the National Institute for Health Research (NIHR) Global Health Research Unit on NTDs at Brighton and Sussex Medical School using Official Development Assistance (ODA) funding. The views expressed here are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Author contributions

T.G.: sample collection, data analysis, figure production and wrote the first draft of the manuscript. C.F.: data analysis and interpretation. D.S.: data analysis and interpretation. T.S.T.: fieldwork supervision. A.A.: study conception and design, sample collection. G.D.: study conception and design, contribution of podoconiosis expertise, facilitation of sample collection. K.H.: data analysis and interpretation. E.Z.: study design, led data analysis and interpretation. M.N.: study conception and design, writing manuscript, PI for funding and overall study. F.T.A.: study design, data interpretation, contribution of samples. All authors edited manuscript and approved final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-81836-x>.

Correspondence and requests for materials should be addressed to M.J.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021