

## Genome-wide association study of Buruli ulcer in rural Benin

Jeremy Manry [1,2]\*#, Quentin B. Vincent [1,2] #, Maya Chrabieh [1,2], Lazaro Lorenzo [1,2], Ioannis Theodorou [3], Marie-Françoise Ardant [4,5], Christian Johnson [4,6], Estelle Marion [7], Annick Chauty [4,5], Laurent Marsollier [7], Laurent Abel [1,2,8], Alexandre Alcaïs [1,2]\* on behalf of the Franco-Beninese Buruli Research Group

[1] Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Recherche Médicale (INSERM) UMR 1163, Paris, France; [2] Imagine Institute, Paris Descartes University, Paris, France; [3] Center for Immunology and Infectious Diseases, INSERM UMR S 1135, Pierre and Marie Curie University, and AP-HP Laboratoire d'Immunologie et Histocompatibilité Hôpital Saint-Louis, Paris, France; [4] Fondation Raoul Follereau, Paris, France; [5] Centre de Dépistage et de Traitement de la Lèpre et de l'Ulcère de Buruli (CDTLUB), Pobè, Benin; [6] Centre Interfacultaire de Formation et de Recherche en Environnement pour le Développement Durable. Université d'Abomey Calavi, Bénin; [7] INSERM UMR-U892 and CNRS U6299, team 7, Angers University, Angers University Hospital, Angers, France; [8] St Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY, USA

# These authors contributed equally to this work

\* Corresponding authors:

Dr Jeremy Manry, Laboratory of Human Genetics of Infectious Diseases, INSERM U1163, Imagine Institute, 75015 Paris, France; [jeremy.manry@inserm.fr](mailto:jeremy.manry@inserm.fr)

Dr Alexandre Alcaïs, Laboratory of Human Genetics of Infectious Diseases, INSERM U1163, Imagine Institute, 75015 Paris, France; [alexandre.alcais@inserm.fr](mailto:alexandre.alcais@inserm.fr)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

## Abstract

Buruli ulcer, caused by *Mycobacterium ulcerans*, is the third mycobacterial disease worldwide characterized by devastating necrotizing skin lesions. The role of host genetics in susceptibility to Buruli ulcer has long been suggested. We conduct the first genome-wide association study of Buruli ulcer on a combined sample of 1,524 well characterized patients and controls from rural Benin. Two-stage analyses identify two novel associated loci located within lincRNA genes: rs9814705 in *ENSG00000240095.1* ( $P = 2.85 \times 10^{-7}$ ; odds ratio = 1.80 [1.43-2.27]), and rs76647377 in *LINC01622* ( $P = 9.85 \times 10^{-8}$ ; hazard ratio = 0.41 [0.28-0.60]). Furthermore, we replicate the protective effect of allele G of a missense variant located in *ATG16L1*, and previously shown to decrease bacterial autophagy (rs2241880,  $P = 0.003$ ; odds ratio = 0.31 [0.14-0.68]). Our results suggest lincRNAs and the autophagy pathway as critical factors in the development of Buruli ulcer.

## Introduction

Buruli ulcer is a chronic infectious disease caused by *Mycobacterium ulcerans*. It received little attention until about 15 years ago, despite being more prevalent than tuberculosis or leprosy in some areas <sup>1</sup>. Interest in this disease was stimulated by the identification, by the World Health Organization (WHO), of Buruli ulcer as a neglected emerging tropical disease and the launch of the Global Buruli Ulcer Initiative in 1998 <sup>2</sup>. Buruli ulcer is now recognized as the third most frequent human mycobacterial disease worldwide, after tuberculosis and leprosy <sup>3</sup>. It occurs mostly in rural areas of tropical countries, and West Africa is considered the principal endemic zone <sup>4</sup>. Prevalence estimates between 2007 and 2016 ranged from 0.32 per 1,000 [95% confidence interval (CI) 0.31-0.33] in Ivory Coast to 2.99 per 1,000 [2.35-3.07] in Benin <sup>5</sup>. In 2018, the WHO reported an increase in new cases of 39% relative to 2016 <sup>4</sup>.

Buruli ulcer is a devastating necrotizing skin infection characterized by pre-ulcerative lesions (nodules, plaques, edema) that may develop into deep ulcers with undermined edges potentially extending to bones and joints. *M. ulcerans* induces painless skin necrosis through the production of mycolactone, a toxin that has been shown to play a key role in bacterial virulence and analgesia <sup>6, 7, 8, 9</sup>. The painless nature of the disease often delays diagnosis, leading to late treatment and permanent disabilities, which affect more than 20% of patients, mostly children <sup>10</sup>. Considerable variability has been reported in the clinical presentation of Buruli ulcer, including spontaneous healing in both humans <sup>11, 12</sup> and specific mouse strains <sup>12</sup>. This observation, together with the familial clustering of cases <sup>13, 14</sup>, supports the view of a substantial contribution of host genetic factors to the response to infection with *M. ulcerans*.

This hypothesis is consistent with the discovery of rare inborn errors of immunity conferring a predisposition to severe infections with weakly virulent mycobacteria, such as the BCG vaccine, in the context of the syndrome of Mendelian susceptibility to mycobacterial diseases (MSMD) <sup>15, 16</sup>. Rare genetic defects have also been shown to underlie severe forms of tuberculosis <sup>17, 18</sup>, and homozygosity for a missense polymorphism of *TYK2* was recently identified as the first common monogenic cause of tuberculosis <sup>19, 20</sup>. A similar monogenic contribution to Buruli ulcer is supported by the recent

identification of a microdeletion on chromosome 8 in a familial form of severe Buruli ulcer <sup>21</sup>.

Genome-wide association studies (GWAS) have identified several common variants associated with leprosy <sup>22, 23, 24</sup>, while this kind of variants seems to play only a limited role in tuberculosis <sup>25</sup>. By contrast, the role of common host genetic variants in the development of Buruli ulcer has been investigated in only three candidate-gene studies. These studies explored a total of seven genes selected on the basis of their involvement in MSMD, tuberculosis or leprosy. Significant association with Buruli ulcer was reported for common variants of six of these genes: *SLC11A1* <sup>26</sup>, *PRKN*, *NOD2*, and *ATG16LI* <sup>27</sup> and *iNOS* and *IFNG* <sup>28</sup>. We performed a two-stage case-control GWAS, to investigate comprehensively the role of common variants in the development of Buruli ulcer, on a sample of 1,524 individuals recruited over the last 15 years at the Centre de Dépistage et de Traitement de la Lèpre et de l'Ulçère de Buruli (CDTLUB) of Pobè, Benin.

## Results

### Genome-wide analyses

GWAS was performed on the discovery sample of 402 Buruli ulcer cases (sex ratio: 0.79; median age: 11 years) and 401 exposed controls (sex ratio: 0.72; median age: 40 years) living in villages in the Ouémé and Plateau *départements* in which Buruli ulcer is endemic (Table 1; Supplementary Fig. 1). Principal component analysis on genotyped variants with a MAF > 0.05 revealed no evidence of population stratification in our sample (Supplementary Fig. 2a), as all our individuals clustered with those of the Yoruba population of Ibadan, Nigeria (YRI) and the Esan population of Nigeria (ESN) from the 1000 Genomes Project (Supplementary Fig. 2b). Following imputation, 10,014,109 high-quality autosomal variants were tested for association with two different phenotypic definitions of Buruli ulcer. We first used the binary case/control Buruli ulcer phenotype and considered the additive genetic model (see Methods). The results of the GWAS association tests are summarized in a Manhattan plot (Fig. 1a). No substantial deviation from expectations was observed, as shown by the genomic inflation factor ( $\lambda=1.027$ ) on the corresponding quantile-quantile plot (qq plot)

(Supplementary Fig. 3a). In total, 517 variants had  $P$  values  $< 5 \times 10^{-5}$ , including 10 with  $P$  values  $< 10^{-6}$ . The strongest signal for association was obtained for an imputed SNP (rs7246288) located on chromosome 19 with a  $P$  value =  $2.80 \times 10^{-7}$ , a MAF of 0.09 and the odds ratio (OR) of developing Buruli ulcer for TT vs. CT or CT vs. CC of 2.33 [1.62-3.34]. After LD pruning, focusing on variants fulfilling the replication criteria (see Methods), 51 variants were retained for replication tests (Supplementary Table 1).

Age at diagnosis can provide important information about the architecture of the genetic contribution to the disease. We performed another GWAS taking age at diagnosis of Buruli ulcer into account with a Cox model (see Methods). The corresponding qq plot ( $\lambda = 1.023$ ) and Manhattan plot are shown in Supplementary Fig. 3b and Fig. 1b, respectively. In total, 534 variants had an association  $P$  value  $< 5 \times 10^{-5}$ , including nine with a  $P$  value  $< 10^{-6}$ . The strongest association signal was obtained for a genotyped SNP (rs34060873) located on chromosome 3 ( $P$  value =  $1.99 \times 10^{-7}$ ). The minor allele A (MAF = 0.10) was protective, with a hazard ratio (HR) for the development of Buruli ulcer for AA vs. CA or CA vs. CC estimated at 0.51 [0.39-0.67]. The same LD pruning strategy led to the selection of 54 variants for replication analyses (Supplementary Table 1). Overall, considering the two phenotypic definitions described above as probably having independent signals, 105 variants, including 99 genotyped variants, were selected for replication analysis (Supplementary Table 1).

### **Replication study**

In the second phase, the 105 identified variants were genotyped in a replication sample composed of 455 cases (sex ratio: 0.77; median age: 15 years) and 238 controls (sex ratio: 0.70; median age: 22.5 years) (Table 1). After quality control, 100 variants were retained for association testing in the replication sample with the same phenotypic models (binary and age at onset) and the same allelic effect as observed for the discovery sample (Supplementary Table 1). Five variants could not be tested. Three of these variants were imputed in the discovery sample and could not be genotyped. Two were present in Alu sequences and the other in an LRT element, which are prone to genotyping and imputation errors. The two remaining variants were genotyped in the discovery sample but were found

to be in Hardy-Weinberg disequilibrium among the controls of the replication sample ( $P$  value  $< 10^{-4}$ ) (Supplementary Table 1). For the 100 high-quality variants retained, evidence of replication, with a  $P$  value less than 0.01 in one-tailed tests, was obtained for two SNPs (Supplementary Table 1 and Table 2).

The first evidence of replication was observed for the binary phenotype and rs9814705 on chromosome 3, for which the replication  $P$  value was  $6.5 \times 10^{-4}$ . As this variant had been imputed in the discovery sample, we decided to subject it to Sanger sequencing in the discovery sample. This resequencing effort slightly increased the discovery  $P$  value from  $3.71 \times 10^{-5}$  (Supplementary Table 1) to  $1.67 \times 10^{-4}$  (Table 2). The combined  $P$  value for this SNP in our two samples was close to the genome-wide significance level (combined  $P$  value =  $2.85 \times 10^{-7}$ ). The minor allele C was the risk allele (MAF = 0.14), and the OR for developing Buruli ulcer for CC vs. TC or TC vs. TT was estimated at 1.80 [95% CI, 1.43-2.27] (Table 2). The proportion of Buruli ulcer patients was 53% for TT carriers, and 76% for CC carriers (Fig. 2, Table 3). We then looked for additional variants in strong LD ( $r^2 > 0.5$ ) with rs9814705, using the 1000 Genomes Phase III data for the Yoruba population. Three SNPs were identified: rs1513419 ( $r^2 = 0.87$ , GWAS  $P$  value =  $4.37 \times 10^{-5}$ ), rs7637582 and rs7615284 ( $r^2 = 0.69$ , and GWAS  $P$  value =  $1.01 \times 10^{-5}$  for both) (Fig. 3a). This cluster of four variants is located on chromosome 3 in an intron of a lincRNA (ENSG00000240095.1) of unknown function, with the closest protein-coding gene (*PLOD2*) 107 kb away.

The second replicated SNP was rs76647377 on chromosome 6, which had an association  $P$  value of  $3.26 \times 10^{-3}$  in the second sample based on the age-of-onset phenotype. The combined  $P$  value for this SNP was close to the genome-wide significance level at  $9.85 \times 10^{-8}$ . The minor allele A (MAF = 0.03) was protective, with a HR for developing Buruli ulcer for AA vs. GA or GA vs. GG estimated at 0.41 [0.28-0.60] (Table 2). GG carriers were found to be prone to developing the disease at an earlier age than GA or AA carriers (Fig. 4). Given that there was only one AA individual, who was a control, we were unable to distinguish between an additive model and a dominant model for allele A, as the combined do not have the power to discriminate between an additive model, and a dominant model for

allele A that provided a very close combined  $P$  value at  $1.37 \times 10^{-7}$ . We identified three additional variants with an  $r^2 > 0.5$  with rs76647377: rs116809810 and rs11969790 ( $r^2 = 0.58$ , GWAS  $P$  value =  $4.55 \times 10^{-5}$  for both), and rs144839883 ( $r^2 = 0.54$ , GWAS  $P$  value =  $2.7 \times 10^{-3}$ ) (Fig. 3b). All these variants are located on chromosome 6, in the intron of a lincRNA *LINC01622* of unknown function (Table 2), and the closest protein-coding gene is *EXOC2*, 292 kb away.

### Signals overlapping with previously reported associations

As the microdeletion recently identified in a familial form of Buruli ulcer was close to a cluster of beta-defensin genes<sup>21</sup>, we also investigated the role of 5,447 variants located in 50 defensin genes and 91 eQTL for defensin genes (Supplementary Table 2). No enrichment in  $P$  values either  $< 0.01$  or  $< 0.001$  was observed among these variants in our discovery sample, the best hit being rs58925751, located in the *DEFB123* gene on chromosome 20 ( $P$  value =  $4.1 \times 10^{-4}$ ). We then checked the level of association in our GWAS for the six variants displaying significant association signals in the only three other candidate-gene association studies on Buruli ulcer performed to date<sup>26, 27, 28</sup> (Table 4). Only one of these variants — rs2241880, a missense T300A variant located in the *ATG16L1* gene — had a  $P$  value  $< 0.05$  in our GWAS under the additive model with the same allelic effect (one-tailed  $P$  value = 0.015). The association was even more significant ( $P$  value = 0.009) when considering a recessive model for allele G, the model of inheritance reported in the original work. Under this model, the OR for developing Buruli ulcer for GG vs. GA or AA subjects was estimated at 0.46 [0.25-0.86]. Furthermore, in the previous study<sup>27</sup>, GG homozygotes were found to be protected against the ulcerative clinical form of Buruli ulcer, with an OR of 0.35 [0.13–0.90]. Strikingly, when focusing only on ulcerative cases of our discovery sample (255 ulcerative cases vs. 401 controls), the association was more significant ( $P$  value = 0.003), with an OR of 0.31 [0.14-0.68] (Table 4). Finally, we also investigated whether some SNPs found to be associated with the other two most common mycobacterial diseases, tuberculosis and leprosy, overlapped with the SNPs identified in our GWAS on Buruli ulcer (see Methods for details). We first looked at the 105 SNPs selected for replication in our GWAS, and noted that rs9295218 — the strongest hit among the genotyped SNPs identified in our discovery sample (GWAS  $P$  value =  $5.17 \times 10^{-7}$ ; OR = 1.64 [1.35-2]) — was located in

an intron of the *PACRG* gene, variants of which have been associated with leprosy<sup>29</sup> (Supplementary Table 1). We observed the same trend towards association in the replication sample, although it was not significant (replication *P* value = 0.13), with an OR of 1.14 [0.91-1.42]). This variant is not in LD with variants of the *PRKN* (also called *PARK2*)/*PACRG* cluster associated with leprosy, including that reported to be associated with Buruli ulcer by Capela *et al.*<sup>27</sup> ( $r^2 = 0.007$ ) (Table 4). We then looked at the 35 SNPs shown to be of genome-wide significance in the GWAS performed on tuberculosis (eight independent variants; Supplementary Table 3a) or leprosy (27 variants in 19 different chromosomal regions; Supplementary Table 3b). None of these 35 SNPs was found to be associated with Buruli ulcer with a *P* value < 0.01. The lowest *P* value (0.017) was obtained for rs2269497, a missense variant of *RGS12* identified in a GWAS on tuberculosis in a Chinese population<sup>30</sup>, for which the minor G allele was found to be associated with a risk of tuberculosis, whereas we found it to have a protective effect in Buruli ulcer. Taken together, these results suggest that there is no strong overlap between common variants potentially involved in the development of Buruli ulcer and in that of tuberculosis or leprosy.

## Discussion

Less than 3% of GWAS to date have focused on sub-Saharan African populations (*vs.* ~50% on populations of European ancestry)<sup>31</sup>. The inaccessibility of many areas makes it difficult to set up GWAS focusing on a neglected tropical disease. In this context, we report the first GWAS investigating susceptibility to Buruli ulcer in a well-characterized sample, including more than 1,500 individuals, in Benin. This sample may be smaller than those used in GWAS on more common infectious diseases, such as leprosy and tuberculosis, but it is the largest sample ever used for a genetic association study on Buruli ulcer. The discovery sample had a power of 80% for detecting, with a type I error of  $5 \times 10^{-5}$  (the threshold to be tested for replication), variants with a MAF > 0.18 and an OR > 1.8. Individuals were recruited for this sample from areas of Benin in which Buruli ulcer is highly endemic, from the villages of Ouémé and Plateau (Supplementary Fig. 1). The high local prevalence in these areas, and our study design, in which the controls were older than the cases, maximize the



chances of controls having been exposed to *M. ulcerans*. The enrollment of unexposed controls would decrease the power of the study. In addition, the definition of Buruli ulcer for the discovery sample was based strictly on the laboratory confirmation of cases, increasing the quality of the criteria used for selecting cases and controls in this study.

The GWAS results highlighted two independent loci displaying a strongly suggestive association with Buruli ulcer rs9814705 for analyses with the binary affected status ( $P$  value =  $2.85 \times 10^{-7}$ ), and rs76647377 when age at Buruli ulcer onset was taken into account ( $P$  value =  $9.85 \times 10^{-8}$ ). These two variants are in regions of the genome devoid of protein-coding genes, the nearest such genes being *PLOD2* (107 kb away) for rs9814705, and *EXOC2* (292 kb away) for rs76647377. Both variants were located within the introns of lincRNAs: *ENSG00000240095.1* (rs9814705) and *LINC01622* (rs76647377). Unfortunately, few (for rs9814705) or no (for rs76647377) functional data are available for these lincRNAs, the functions of which remain unknown. SNPs rs9814705 and rs1513419 ( $r^2 = 0.87$  with rs9814705) are eQTL for lincRNA *ENSG00000240095.1* in esophagus-mucosa tissue ( $P$  value =  $3.9 \times 10^{-5}$  and  $P$  value =  $4.5 \times 10^{-5}$ , respectively)<sup>32</sup>. In the GTEX data, this lincRNA appeared to be expressed only in esophagus-mucosa tissue and in the minor salivary gland. In addition, rs9814705 and rs1513419 had  $P$  values of borderline significance (0.049) for being cis-eQTL of *PLSCR1* in lipopolysaccharide-stimulated monocytes from healthy individuals<sup>33</sup>. The *PLSCR1* gene, located 565 kb away from rs9814705, encodes the transmembrane protein phospholipid scramblase 1, which has been shown to be overexpressed after stimulation with type I interferons, and to play a crucial role in apoptosis<sup>34,35</sup>. This gene has also been identified as a core gene in the host response to tuberculosis in a meta-analysis of transcriptomic data<sup>36</sup>. Thus, *PLSCR1* may be an interesting candidate gene for involvement in the development of Buruli ulcer. Further studies are required to determine the precise role of the identified lincRNAs.

Using our GWAS data, we replicated the effect of the A/G SNP rs2241880, located in the *ATG16L1* gene, initially identified in another sample from Benin<sup>27</sup>. Individuals homozygous for the minor allele G (MAF of 0.33 in the African populations of gnomAD) are protected against Buruli ulcer, especially the ulcerative forms, with an OR of 0.31 [0.14-0.68] in our sample and 0.35 [0.13-0.90] in the original

study. The protein encoded by *ATG16L1* is required for autophagy<sup>37</sup>, and the G allele has been associated with an increase in the release of cytokines, such as IL-1 $\beta$  and IL-18, leading to a decrease in antibacterial autophagy<sup>38,39</sup>. This variant results in a T300A missense substitution in the principal isoform of *ATG16L1*. Remarkably, it has also been reported to be an eQTL in nine tissues, including skin, in which the G allele of *ATG16L1* is associated with higher levels of expression of this gene ( $P$  value =  $1.6 \times 10^{-25}$ )<sup>32</sup>. Allele G of rs2241880 was also found to increase the risk of Crohn's disease (CD) in several GWAS, with an OR of about 1.4 under an additive model<sup>40,41,42</sup>. In addition, *Atg16L1*-hypomorphic mice developed intestinal abnormalities resembling CD whilst displaying resistance to intestinal infection with the bacterium *Citrobacter rodentium*<sup>43,44</sup>. Published findings suggest that autophagy may serve as a rheostat for immune reactions<sup>45</sup>. These opposite effects in Buruli ulcer and CD provide another example of mirror genetic effects (*i.e.* an allele protective against infection but deleterious for inflammatory disorders), consistent with the view that the current high frequency of inflammatory/autoimmune diseases may reflect past selection for strong immune responses to combat infection<sup>46</sup>.

Despite significant practical and methodological challenges<sup>47,48,49</sup>, we successfully conducted a GWAS on a neglected tropical disease in rural sub-Saharan Africa, discovered two loci associated with Buruli ulcer and replicated a previously identified locus. Further studies are required to decipher the role of these variants and the function of the lincRNAs in which they are located. Detailed transcriptomic and epigenetic profiles for both whole blood (systemic response) and at the site of ulcers (local response) would complement our genomic approach and highlight genes and pathways involved in the disease. Mouse models of Buruli ulcer<sup>50</sup> will also be of major interest in this context, for further investigation of the role of the *ATG16L1* variant in the response to *M. ulcerans*, in particular.

## Methods

### Ethics statement

This study on genetic susceptibility to Buruli ulcer was approved by the institutional review board of the CDTLUB (Pobè, Benin), the National Ethical Review Board of the Ministry of Health in Benin (IRB00006860), and the ethics committee of the University Hospital of Angers, France (Comité d’Ethique du CHU d’Angers). Written informed consent was obtained from all adult participants. Parents or guardians provided informed consent on behalf of all minors participating in the study.

### Subjects

Between July 2003 and June 2012, 1,524 individuals from the CDTLUB in Pobè, Benin, were enrolled. All participants were living in villages distributed over the Ouémé and Plateau *départements* (an administrative area equivalent to a county) in an area in which Buruli ulcer was endemic<sup>51, 52</sup>. Based on the cases registered in these two areas between 2005 and 2012, the overall prevalence in the studied villages was 5.3 per 1,000; a detailed distribution by village of the cases recorded is shown in Supplementary Fig. 1. This discovery sample consisted of 408 HIV-free Buruli ulcer cases and 408 exposed controls (354 male and 462 female subjects; the mean ages of the cases and controls were 18.5 and 39.8 years, respectively). We chose to include controls that were older than the cases, as being older meant that they had been exposed to *M. ulcerans* for a longer time but had remained Buruli ulcer-free). All 408 cases were confirmed by IS2404 PCR, culture and/or Ziehl-Neelsen staining in the laboratory<sup>10, 53</sup>. The replication sample consisted of 708 individuals: 467 HIV-free Buruli ulcer patients and 241 exposed controls (303 male and 405 female subjects; the mean ages of cases and controls were 22.3 and 26.1 years, respectively). The cases included in the replication sample were either laboratory-confirmed (48.6%) or classified as highly probable (51.4%) on the basis of stringent clinical criteria and treatment efficacy.

### Genotyping, imputation and quality control

Genomic DNA was extracted from 5 to 10 mL blood samples with the Nucleon BACC2 Genomic DNA extraction kit (GE Healthcare), in accordance with the manufacturer's instructions. DNA

preparations from the discovery sample were randomized in nine 96-well plates (with an equal number of cases and controls per plate) and genotyped with the Illumina Omni2.5 chip array. Genotypes were assessed with GenTrain implemented in Illumina GenomeStudio software (v2011.1). Thirteen of the 816 samples were excluded from further analyses due to a call rate <95% (8 samples), lack of concordance between reported and genetically inferred gender (3 samples), or duplication (2 samples). Our ‘effective’ discovery sample therefore consisted of a total of 803 individuals: 402 cases and 401 controls (Table 1).

In total, we obtained 2,314,174 non-filtered autosomal variants, of which, 1,804,366, with a call rate > 95%, a minimum allelic frequency (MAF) > 0.01, and a Hardy-Weinberg  $P$  value >  $10^{-6}$  in controls were used for the two-step (phasing and genotype inference) imputation process. Haplotypes were phased with SHAPEIT2 v2.904<sup>54</sup>. For non-genotyped variants, genotype was inferred with IMPUTE2<sup>55</sup>, using the 1000 Genomes Phase III dataset as the reference panel and default values for all other parameters<sup>56,57</sup>. Only variants imputed with an *info* value<sup>55</sup> above 0.6 were retained for further analysis. In total, 10,014,109 autosomal variants with a MAF above 0.02 if genotyped (1,683,899 variants) or 0.05 if imputed (8,330,210 variants) were tested for association.

All variants for which evidence for association was obtained with a type I error of less than  $5 \times 10^{-5}$  (genotyped variants) or  $10^{-6}$  (imputed variants) were retained for genotyping and association testing in the replication sample. As the replication sample was derived from exactly the same population as the discovery sample, we were able to prune all the significant variants on the basis of their estimated linkage disequilibrium (LD), to limit the genotyping effort. Among the variants of a given bin, defined as a group of variants with an  $r^2 > 0.8$  with a core variant, we first selected the genotyped variant with the highest likelihood of successful genotyping in the replication sample, according to the manufacturer’s instructions (see below). For bins without a genotyped variant or with genotyped variant for which the chances of genotyping success were low, we selected the imputed variant with the highest *info* value. Following LD pruning, 105 variants (99 genotyped and 6 imputed) were genotyped and tested for association with Buruli ulcer in the replication sample. These variants were genotyped by Illumina GoldenGate genotyping with VeraCode technology. The replication sample

contained 708 individuals; 693 of these individuals (455 cases and 238 controls; Table 1) passed the quality control criteria (4 samples with a call rate < 95%, and 11 duplicated samples were excluded).

Variants with a call rate < 95% or a Hardy Weinberg  $P$  value below  $10^{-4}$  in controls were excluded.

Significant replication was defined as a variant having the same allelic effect as in the discovery sample, with a one-sided  $P$  value < 0.01 on the replication sample. As one of the replicated SNPs (rs9814705) was imputed in the discovery sample, we decided to perform Sanger sequencing on this SNP in the discovery sample to strengthen our conclusions. Sequences for the 803 samples were obtained with the Big Dye Terminator kit and a 3500xL automated sequencer from Applied Biosystems. Sequence files and chromatograms were inspected with GENALYS software<sup>58</sup>. Given the high concordance between imputation and Sanger sequencing data (~99%), genotypes inferred by the imputation process were used for the 143 missing Sanger-sequenced genotypes.

### **Statistical analyses**

Principal component (PC) analysis was performed on genotyped variants with a MAF > 0.05 to check for population structure, for the discovery sample and all the samples available from 1000 Genomes Project Phase III<sup>59</sup>, with the *ad hoc* functions implemented in PLINK v1.9<sup>60</sup>.

Two statistical designs were considered for tests of the association between genetic variants and the occurrence of Buruli ulcer, depending on the phenotypic definition use. We first considered Buruli ulcer as a binary phenotype (affected/unaffected). A classical case/control design was used and all analyses were performed within a logistic regression framework. Gender had no significant effect ( $P$  value = 0.41) on binary Buruli ulcer status. Given that the controls were older than the cases by design, we did not include age as a covariate, to avoid over-adjustment, and the effect of age was accounted for by specific survival analysis methods (see below). We also evaluated the potential impact of cryptic population stratification and/or hidden relatedness within our discovery sample, by comparing the results of the association tests obtained with a classical logistic regression model and those obtained with a mixed-effect logistic model<sup>61</sup>. We observed an extremely strong correlation ( $r = 0.90$ , Spearman  $P$  value <  $10^{-16}$ ) between the top 1000 variants obtained by classical logistic regression

and those obtained with a mixed model (Supplementary Fig. 4a). The largest difference for these top 1,000 variants, measured as the absolute difference between the  $-\log_{10}(P \text{ value})$  obtained in classical logistic regression that obtained with the mixed model, was less than 0.81. Consistent with these findings, we found no significant association between the three first principal components and the occurrence of Buruli ulcer.

We then considered age at onset for Buruli ulcer patients and age at examination for the exposed controls as the phenotype of interest. A survival analysis framework was used and all analyses were performed with Cox models<sup>62</sup>. The analysis strategy was identical to that used for the binary phenotype: i.e., we evaluated the potential impact of cryptic population stratification and/or hidden relatedness within the discovery sample by comparing the results of the association tests performed with a classical Cox model with those obtained with a mixed-effect Cox model<sup>63</sup>. Again, the correlation between the  $P$  values of the 1,000 top variants for the classical and mixed-effect Cox models was extremely strong ( $r = 0.97$ , Spearman  $P$  value  $< 10^{-16}$  (Supplementary Fig. 4b)). For these top 1,000 variants, the largest difference in terms of  $\log_{10}(P \text{ value})$  was at 0.27. Consistently, the use of the first three principal components had no impact on the results. The results reported here are therefore those for association tests performed with classical logistic or Cox models. All analyses were conducted assuming an additive genetic effect of the variants, and all the  $P$  values displayed were obtained in likelihood-ratio tests. All association analyses were performed with R (<https://www.r-project.org/>), PLINK 1.9<sup>60</sup>, GEMMA<sup>61</sup>, coxme<sup>63</sup>, SNPTEST v2.5.2<sup>64</sup>, and ProbABEL<sup>65</sup> software.

Finally, we calculated the power of our discovery sample (402 cases and 401 controls) for detecting a significant association with Buruli ulcer coded as a binary trait, assuming a type I error of  $5 \times 10^{-5}$ .

Power estimates are shown for an additive model, by MAF, and effect size, as estimated by calculation of the odds ratio (OR) of the tested variants in Supplementary Fig. 5. For example, our sample had 50% power for the detection of variants with an OR of 1.6 and a MAF  $> 0.2$ , and 80% power for detecting variants with an OR of 1.8 and a MAF  $> 0.18$ . All these power estimates were calculated with QUANTO v1.2.4 (<http://biostats.usc.edu/Quanto.html>).

### **Validation of variants previously found to be associated with common mycobacterial diseases**

We used our discovery sample to assess genetic findings already reported for Buruli ulcer and in the larger context of common mycobacterial diseases. We recently identified a microdeletion in a familial form of Buruli ulcer that overlaps with a cluster of genes encoding defensins<sup>21</sup>. We, therefore, first identified all the variants of the 50 defensin genes (located within 10 kb on either side of the gene) listed in the Ensembl GRCh37.p13 database and tested for association in our GWAS. We also considered all the variants reported to be eQTL (FDR < 0.05) for defensin genes in whole-blood and skin from the lower legs exposed to the sun in the GTEx database<sup>32</sup>. Results were available for 5,452 defensin-related SNPs (Supplementary Table 1). We then searched for variants reported to be associated with Buruli ulcer in the three published candidate gene studies<sup>26, 27, 28</sup>. A total of six variants on six genes were assessed. Finally, we investigated the role of host genetic factors potentially common to Buruli ulcer and the other two main mycobacterial diseases (i.e. tuberculosis and leprosy). We screened the GWAS catalog (<https://www.ebi.ac.uk/gwas/>) for variants associated with either tuberculosis or leprosy at the genome-wide level ( $P$  value <  $5 \times 10^{-8}$ ). We identified 36 independent SNPs, eight in TB, and 27 in leprosy, which we then assessed for association with Buruli ulcer in our GWAS discovery sample.

## References

1. Wansbrough-Jones M, Phillips R. Buruli ulcer: emerging from obscurity. *Lancet* **367**, 1849-1858 (2006).
2. World Health Organization. Neglected tropical diseases, hidden successes, emerging opportunities. (2009).
3. Sizaire V, Nackers F, Comte E, Portaels F. Mycobacterium ulcerans infection: control, diagnosis, and treatment. *Lancet Infect Dis* **6**, 288-296 (2006).
4. World Health Organization. Buruli ulcer (Mycobacterium ulcerans infection) Fact Sheet. (2018).
5. Simpson H, et al. Mapping the global distribution of Buruli ulcer: a systematic review with evidence consensus. *Lancet Glob Health* **7**, e912-e922 (2019).
6. Coutanceau E, et al. Modulation of the host immune response by a transient intracellular stage of Mycobacterium ulcerans: the contribution of endogenous mycolactone toxin. *Cell Microbiol* **7**, 1187-1196 (2005).
7. George KM, et al. Mycolactone: a polyketide toxin from Mycobacterium ulcerans required for virulence. *Science* **283**, 854-857 (1999).
8. Marion E, et al. Mycobacterial toxin induces analgesia in buruli ulcer by targeting the angiotensin pathways. *Cell* **157**, 1565-1576 (2014).
9. Marsollier L, et al. Colonization of the salivary glands of Naucoris cimicoides by Mycobacterium ulcerans requires host plasmacytes and a macrolide toxin, mycolactone. *Cell Microbiol* **7**, 935-943 (2005).
10. Vincent QB, et al. Clinical epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study. *Lancet Glob Health* **2**, e422-430 (2014).
11. O'Brien DP, Murrie A, Meggyesy P, Priestley J, Rajcoomar A, Athan E. Spontaneous healing of Mycobacterium ulcerans disease in Australian patients. *PLoS Negl Trop Dis* **13**, e0007178 (2019).
12. Marion E, et al. FVB/N Mice Spontaneously Heal Ulcerative Lesions Induced by Mycobacterium ulcerans and Switch M. ulcerans into a Low Mycolactone Producer. *J Immunol* **196**, 2690-2698 (2016).
13. O'Brien DP, et al. Exposure Risk for Infection and Lack of Human-to-Human Transmission of Mycobacterium ulcerans Disease, Australia. *Emerg Infect Dis* **23**, 837-840 (2017).
14. Sopoh GE, et al. Family relationship, water contact and occurrence of Buruli ulcer in Benin. *PLoS Negl Trop Dis* **4**, e746 (2010).
15. Bustamante J, Boisson-Dupuis S, Abel L, Casanova JL. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN-gamma immunity. *Semin Immunol* **26**, 454-470 (2014).
16. Casanova JL, Abel L. The genetic theory of infectious diseases: a brief history and selected illustrations. *Annu Rev Genomics Hum Genet* **14**, 215-243 (2013).
17. Alcais A, Fieschi C, Abel L, Casanova JL. Tuberculosis in children and adults: two distinct genetic diseases. *J Exp Med* **202**, 1617-1621 (2005).
18. Boisson-Dupuis S, et al. Inherited and acquired immunodeficiencies underlying tuberculosis in childhood. *Immunol Rev* **264**, 103-120 (2015).
19. Boisson-Dupuis S, et al. Tuberculosis and impaired IL-23-dependent IFN-gamma immunity in humans homozygous for a common TYK2 missense variant. *Sci Immunol* **3**, (2018).
20. Kerner G, et al. Homozygosity for TYK2 P1104A underlies tuberculosis in about 1% of patients in a cohort of European ancestry. *Proc Natl Acad Sci U S A* **116**, 10430-10434 (2019).
21. Vincent QB, et al. Microdeletion on chromosome 8p23.1 in a familial form of severe Buruli ulcer. *PLoS Negl Trop Dis* **12**, e0006429 (2018).
22. Zhang FR, et al. Genomewide association study of leprosy. *N Engl J Med* **361**, 2609-2618 (2009).
23. Zhang F, et al. Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. *Nat Genet* **43**, 1247-1251 (2011).



24. Liu H, *et al.* Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. *Nat Genet* **47**, 267-271 (2015).
25. Abel L, *et al.* Genetics of human susceptibility to active and latent tuberculosis: present knowledge and future perspectives. *Lancet Infect Dis* **18**, e64-e75 (2018).
26. Stienstra Y, *et al.* Susceptibility to Buruli ulcer is associated with the SLC11A1 (NRAMP1) D543N polymorphism. *Genes Immun* **7**, 185-189 (2006).
27. Capela C, *et al.* Genetic Variation in Autophagy-Related Genes Influences the Risk and Phenotype of Buruli Ulcer. *PLoS Negl Trop Dis* **10**, e0004671 (2016).
28. Bibert S, *et al.* Susceptibility to Mycobacterium ulcerans Disease (Buruli ulcer) Is Associated with IFNG and iNOS Gene Polymorphisms. *Front Microbiol* **8**, 1903 (2017).
29. Mira MT, *et al.* Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* **427**, 636-640 (2004).
30. Qi H, *et al.* Discovery of susceptibility loci associated with tuberculosis in Han Chinese. *Hum Mol Genet* **26**, 4752-4763 (2017).
31. Morales J, *et al.* A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol* **19**, 21 (2018).
32. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585 (2013).
33. Kim-Hellmuth S, *et al.* Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat Commun* **8**, 266 (2017).
34. Dong B, *et al.* Phospholipid scramblase 1 potentiates the antiviral activity of interferon. *J Virol* **78**, 8983-8993 (2004).
35. Sivagnanam U, Palanirajan SK, Gummadi SN. The role of human phospholipid scramblases in apoptosis: An overview. *Biochim Biophys Acta Mol Cell Res* **1864**, 2261-2271 (2017).
36. Sambarey A, *et al.* Meta-analysis of host response networks identifies a common core in tuberculosis. *NPJ Syst Biol Appl* **3**, 4 (2017).
37. Mizushima N, *et al.* Mouse Apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12-Apg5 conjugate. *J Cell Sci* **116**, 1679-1688 (2003).
38. Lassen KG, *et al.* Atg16L1 T300A variant decreases selective autophagy resulting in altered cytokine signaling and decreased antibacterial defense. *Proc Natl Acad Sci U S A* **111**, 7741-7746 (2014).
39. Lassen KG, Xavier RJ. Mechanisms and function of autophagy in intestinal disease. *Autophagy* **14**, 216-220 (2018).
40. Hampe J, *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* **39**, 207-211 (2007).
41. Kenny EE, *et al.* A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. *PLoS Genet* **8**, e1002559 (2012).
42. Rioux JD, *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* **39**, 596-604 (2007).
43. Cadwell K. Crosstalk between autophagy and inflammatory signalling pathways: balancing defence and homeostasis. *Nat Rev Immunol* **16**, 661-675 (2016).
44. Marchiando AM, *et al.* A deficiency in the autophagy gene Atg16L1 enhances resistance to enteric bacterial infection. *Cell Host Microbe* **14**, 216-224 (2013).
45. Martin PK, *et al.* Autophagy proteins suppress protective type I interferon signalling in response to the murine gut microbiota. *Nat Microbiol* **3**, 1131-1141 (2018).
46. Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet* **11**, 17-30 (2010).
47. Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. *Nat Rev Genet* **11**, 149-160 (2010).
48. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. *Nat Rev Genet* **11**, 356-366 (2010).
49. Peprah E, Xu H, Tekola-Ayele F, Royal CD. Genome-wide association studies in Africans and African Americans: expanding the framework of the genomics of human traits and disease. *Public Health Genomics* **18**, 40-51 (2015).

50. Bieri R, Bolz M, Ruf MT, Pluschke G. Interferon-gamma Is a Crucial Activator of Early Host Immune Defense against Mycobacterium ulcerans Infection in Mice. *PLoS Negl Trop Dis* **10**, e0004450 (2016).
51. Johnson RC, Sopoh GE, Barogui Y, Dossou A, Fourn L, Zohoun T. [Surveillance system for Buruli ulcer in Benin: results after four years]. *Sante* **18**, 9-13 (2008).
52. Wagner T, Benbow ME, Brenden TO, Qi J, Johnson RC. Buruli ulcer disease prevalence in Benin, West Africa: associations with land use/cover and the identification of disease clusters. *Int J Health Geogr* **7**, 25 (2008).
53. Stienstra Y, et al. Analysis of an IS2404-based nested PCR for diagnosis of Buruli ulcer disease in regions of Ghana where the disease is endemic. *J Clin Microbiol* **41**, 794-797 (2003).
54. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-181 (2011).
55. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-959 (2012).
56. Genomes Project C, et al. A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
57. van Leeuwen EM, et al. Population-specific genotype imputations using minimac or IMPUTE2. *Nat Protoc* **10**, 1285-1296 (2015).
58. Takahashi M, Matsuda F, Margetic N, Lathrop M. Automated identification of single nucleotide polymorphisms from sequencing data. *J Bioinform Comput Biol* **1**, 253-265 (2003).
59. Sudmant PH, et al. An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75-81 (2015).
60. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
61. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821-824 (2012).
62. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B (Methodological)* **34**, 187-220 (1972).
63. Therneau T. coxme: Mixed Effects Cox Models. R package version 2.2-3 [Internet]. 2012. Available from: <http://CRAN.R-project.org/package=coxme>. (2012).
64. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-913 (2007).
65. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).

## **Acknowledgments**

We thank Jean-Laurent Casanova, Emmanuelle Jouanguy and all the members of the HGID laboratory for useful discussions. We also thank Guillaume Vogt, Antoine Guérin, Jérémie Babonneau, Julien Guergnon, Grégoire Hure, Natasha Vladikine, and Myriam Berramdane for their contribution to genotyping and sequencing. We acknowledge support from the Fondation Raoul Follereau. LM and AA were supported by the Agence Nationale de la Recherche (ANR; grant no. ANR-17-BSV3-0013-01). JM, LA and AA were supported by the Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant no. ANR-10-LABX-62-IBEID) and the ANR under the "Investments for the Future" program (grant no. ANR-10-IAHU-01). QBV was supported by the Fondation Bettencourt Schueller through the MD/PhD program of the Imagine Institute. AA was supported by the Fondation pour la Recherche Médicale (grant no. DMI20091117308).

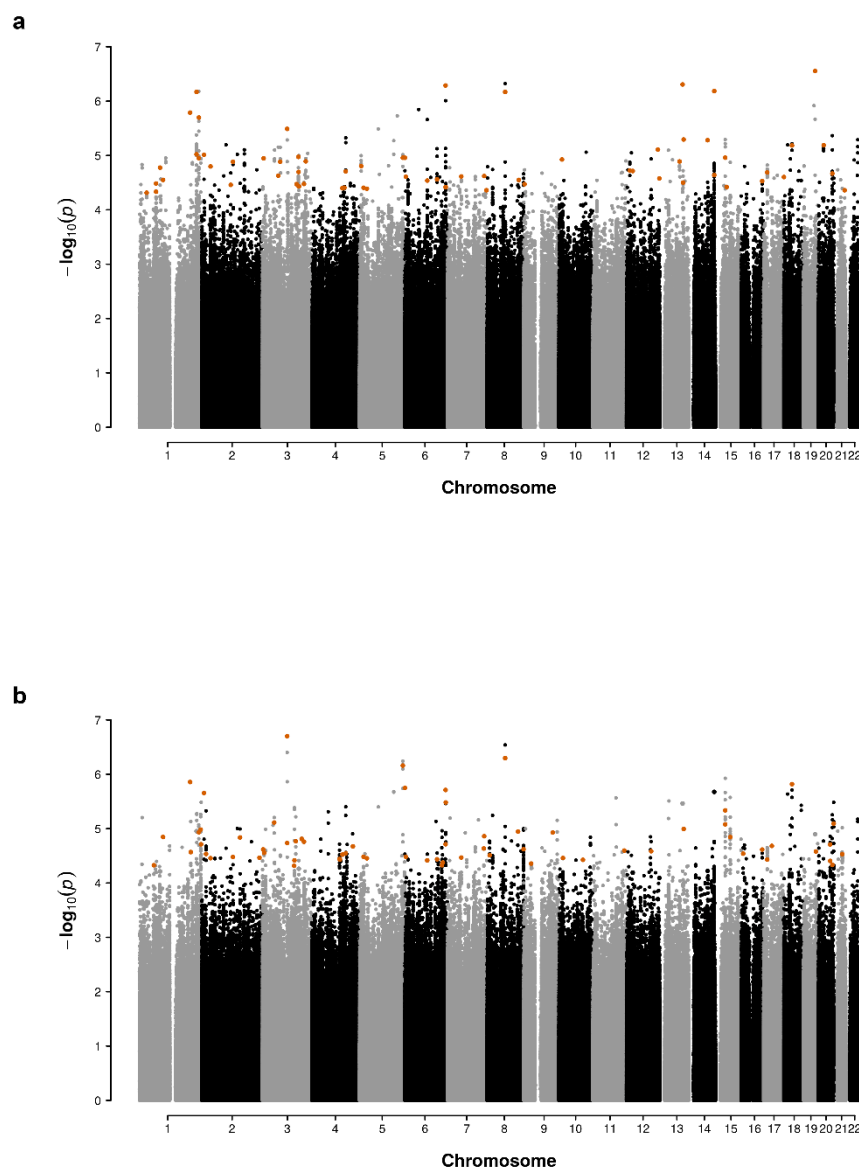
## **Author Contributions**

All authors were responsible for the study concept and design, and the analysis and interpretation of data. M-FA, CJ, LM, EM and AC collected the data. M-FA, LM, EM and AC did or confirmed diagnostics. QV, JM, MFA, EM, CJ, AC, MC, LL, IT and AA acquired the data. QV, JM, LA, and AA did the statistical analysis. JM, LA, and AA drafted the report. AA obtained the funding.

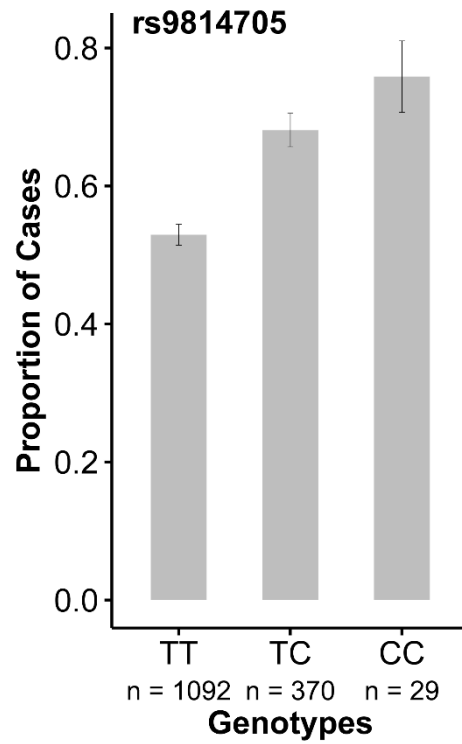
## **Competing Interests statement**

We declare no competing interests.

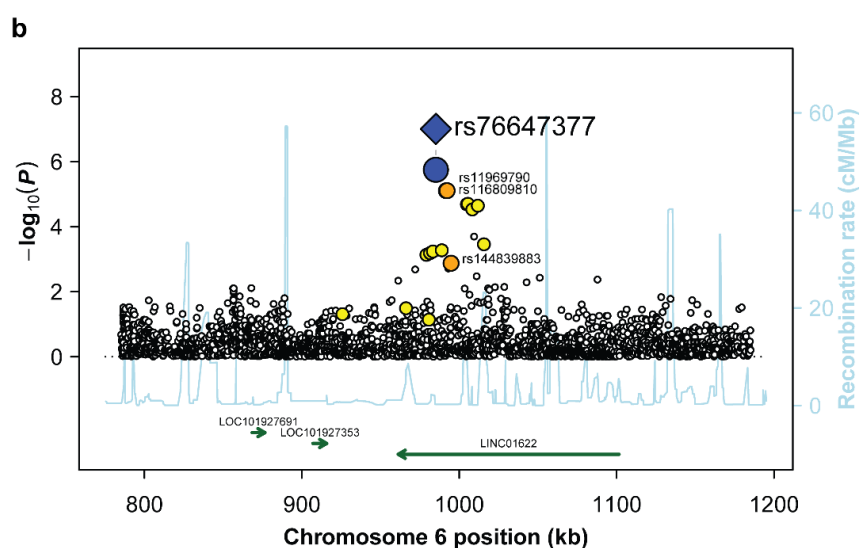
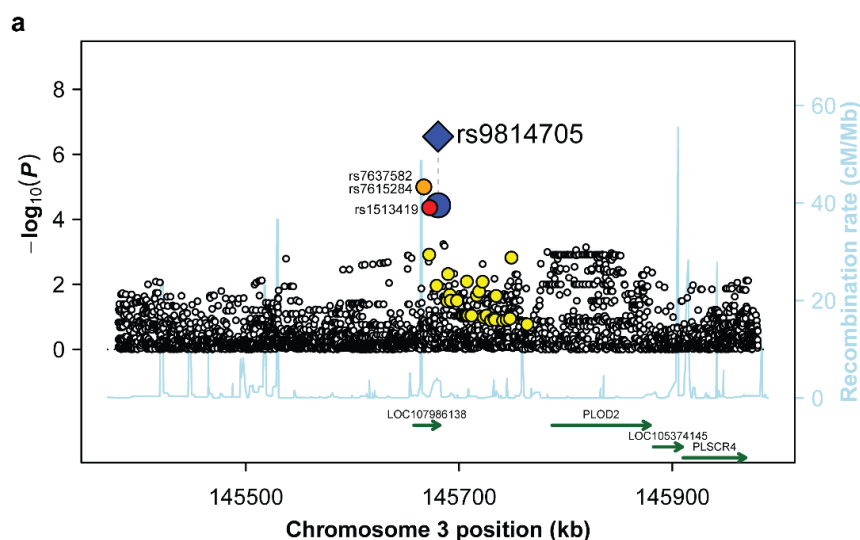
**Figure 1. Manhattan plots of a GWAS of susceptibility to Buruli ulcer in a population from Benin.** Manhattan plot of genome-wide  $P$  values for associations between variants and Buruli ulcer considering either a binary (affected/unaffected; panel A) or a censored (age at onset for Buruli ulcer patients and age at examination for the exposed controls; panel B) phenotype. The effects of variant genotypes on Buruli ulcer were investigated under an additive genetic model, using logistic regression and Cox models for the binary and censored phenotypes, respectively. The  $-\log_{10}P$ -values ( $y$ -axis) for SNPs are presented according to chromosomal positions ( $x$ -axis). Larger orange dots correspond to the 105 variants selected for replication.



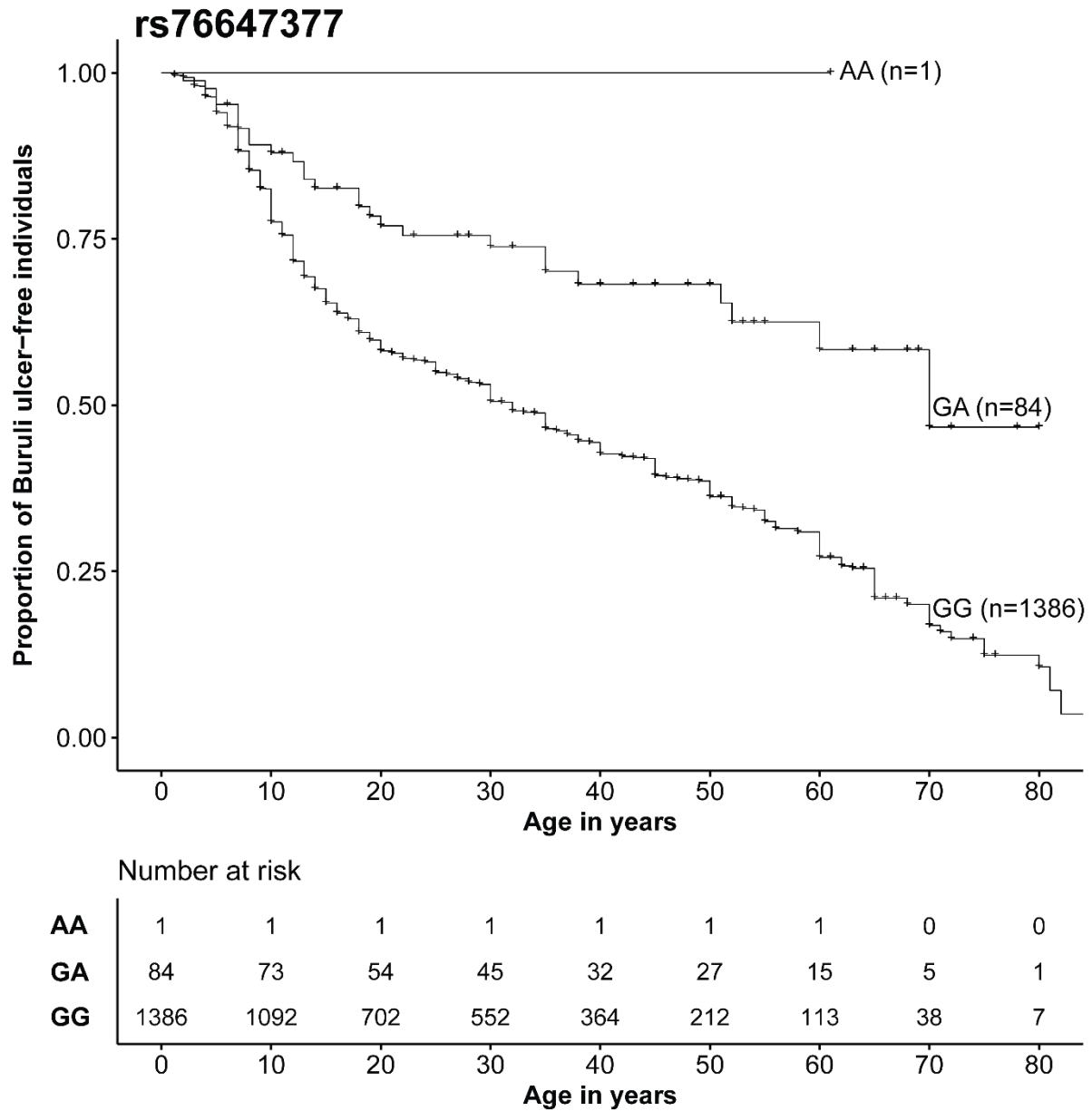
**Figure 2. Genotype at rs9814705 is associated with case-control status.** Distribution of the proportion of Buruli ulcer cases (y-axis) according to rs9814705 genotype (x-axis) in the combined sample ( $n$ : number of individuals).



**Figure 3. Association, linkage disequilibrium and gene maps for the two replicated association signals.** Plots, with GRCh37 coordinates, showing evidence of association ( $-\log_{10}(P)$ ; left y-axis) for the variants located in the vicinity of (A) rs9814705 and (B) rs76647377 (kb; x-axis). The distribution of recombination rates in these regions is also given (cM/Mb; Right y-axis; light blue line). For the two replicated SNPs (i.e. rs9814705 and rs76647377), blue diamonds correspond to the results observed in the combined analysis, whereas blue circles correspond to the results observed in the GWAS only. A color-coded scheme is used to display the degree of LD of the proximal variants with the replicated SNP (red:  $r^2 > 0.8$ , orange:  $0.5 < r^2 < 0.8$ , and yellow:  $0.2 < r^2 < 0.5$ ). Only variants with an  $r^2 > 0.5$  are labeled. Genes in the two chromosomal regions are plotted, with arrows indicating their orientation.



**Figure 4. Genotype at rs76647377 is associated with age at onset of Buruli ulcer.** Kaplan-Meier curves for the onset of Buruli ulcer (y-axis) according to rs76647377 genotype. Number of individuals at risk are provided by genotype and by ten-year age bins.



**Table 1. Clinical characteristics of the discovery and replication samples**

<b>Covariate</b>		<b>Discovery sample</b>		<b>Replication sample</b>		<b>Combined</b>	
		<b>Cases</b>	<b>Controls</b>	<b>Cases</b>	<b>Controls</b>	<b>Cases</b>	<b>Controls</b>
<b>Sex</b>	<b>Male</b>	177 (44.0)	168 (41.9)	198 (43.5)	98 (41.1)	375 (43.8)	266 (41.6)
	<b>Female</b>	225 (56.0)	233 (58.1)	257 (56.5)	140 (58.9)	482 (56.2)	373 (58.4)
	<b>Total</b>	402	401	455	238	857	639
<b>Age</b> <sup>a</sup>	<b>Mean age (SD)</b>	18.5 (16.7)	39.8 (17.3)	22.4 (17.3)	26.1 (18.7)	20.5 (17.1)	34.6 (19.0)
	<b>Median age</b>	11	40	15	22.5	13	35

NOTE. All values are expressed as *n* (%) unless otherwise indicated. SD: standard deviation.

<sup>a</sup> Age is given in years and corresponds to the age at the time of recruitment for controls and to age at diagnosis for cases.



**Table 2. Association results for the two replicated Buruli ulcer GWAS signals**

SNP	Chr	Position <sup>a</sup>	Gene	Consequence	M/m <sup>b</sup>	Discovery	Replication	Combined sample			
						<i>P</i> value	<i>P</i> value <sup>c</sup>	MAF Controls	MAF Cases	<i>P</i> value	Effect size <sup>d</sup>
rs9814705*	3	145680345	<i>ENSG00000240095.1</i>	intron	T/C	1.67×10 <sup>-4</sup>	6.51×10 <sup>-4</sup>	0.10	0.16	<b>2.85×10<sup>-7</sup></b>	1.80 (1.43 to 2.27)
rs76647377	6	985196	<i>LINC01622</i>	intron	G/A	1.78×10 <sup>-6</sup>	3.26×10 <sup>-3</sup>	0.05	0.02	<b>9.85×10<sup>-8</sup></b>	0.41 (0.28 to 0.60)

<sup>a</sup> GRCh37.p13

<sup>b</sup> The first allele *m* is the minor allele and *M* is the major allele in the combined cohort.

<sup>c</sup> One-tailed test.

<sup>d</sup> Effect size corresponds to odds ratios (95% confidence interval) when the GLM model is considered (rs9814705), and hazard ratios (95% confidence interval) when the Cox model is considered (rs76647377). Effect size is assessed under an additive model, using the minor allele as the reference.

\*rs9814705 was resequenced in the discovery cohort.

**Table 3. Genotypic distributions of the two replicated SNPs between cases and controls in the combined cohort**

SNP	Chr	Position <sup>a</sup>	Group	Genotype Frequencies (n) <sup>b</sup>		
				T/T	T/C	C/C
rs9814705	3	145680345	Cases	0.678 (578)	0.296 (252)	0.026 (22)
			Controls	0.804 (514)	0.185 (118)	0.011 (7)
rs76647377	6	985196		G/G	G/A	A/A
			Cases	0.968 (827)	0.032 (27)	0 (0)
			Controls	0.909 (582)	0.089 (57)	0.002 (1)

<sup>a</sup> GRCh37.p13

<sup>b</sup> Missing Sanger sequencing data for rs9814705 (143 samples) were obtained by best-guess imputation.

**Table 4. Validation study for six variants of three genes previously reported to be associated with Buruli ulcer**

SNP	Chr	M/m	Gene	Previous candidate gene studies						GWAS <sup>a</sup>		
				Study	Population	MAF	Model	OR (95% CI)	P value	MAF	OR (95% CI)	P value <sup>b</sup>
<b>rs1333955</b>	6	C/T	<i>PARK2</i>	Capela <i>et al.</i> <sup>27</sup>	Benin	0.24	Dom	1.43 (1.00-2.06)	0.05	0.27	1.05 (0.64-1.69)	0.357
<b>rs9302752</b>	16	C/T	<i>NOD2</i>	Capela <i>et al.</i> <sup>27</sup>	Benin	0.38	Dom	2.23 (1.14-4.37)	0.02	0.40	1.10 (0.72-1.67)	0.498
<b>rs2241880</b>	2	A/G	<i>ATG16L1</i>	Capela <i>et al.</i> <sup>27</sup>	Benin	0.30	Rec	0.35 (0.13-0.90)	0.02	0.26	0.46 (0.25-0.86)	<b>0.009</b>
<b>rs9282799</b>	17	G/A	<i>iNOS</i>	Bibert <i>et al.</i> <sup>28</sup>	Ghana	0.08	Add	1.99 (1.22-3.26)	0.006	0.04	1.33 (0.82-2.14)	0.110
<b>rs2069705</b>	12	A/G	<i>IFNG</i>	Bibert <i>et al.</i> <sup>28</sup>	Ghana	0.47	Add	1.56 (1.14-1.99)	0.007	0.43	1.00 (0.82-1.22)	0.482
<b>rs17235409</b>	2	G/A	<i>SCL11A1</i>	Stienstra <i>et al.</i> <sup>26</sup>	Ghana	0.07	Dom	2.50 (1.26-4.96)	0.007	0.09	1.02 (0.20-5.08)	0.257

Chr: chromosome; M: major allele; m: minor allele; MAF: minor allele frequency; Dom: dominant model; Rec: recessive model; Add: additive model.

<sup>a</sup> Results from the GWAS sample obtained with the binary case/control phenotype.

<sup>b</sup> P values for one-tailed tests are given for the marker alleles indicated in bold.