



**HAL**  
open science

## Genome-wide association study of Buruli ulcer in rural Benin highlights role of two LncRNAs and the autophagy pathway

Jeremy Manry, Quentin Vincent, Christian Johnson, Maya Chrabieh, Lazaro Lorenzo, Ioannis Theodorou, Marie-Françoise Ardant, Estelle Marion, Annick Chauty, Laurent Marsollier, et al.

### ► To cite this version:

Jeremy Manry, Quentin Vincent, Christian Johnson, Maya Chrabieh, Lazaro Lorenzo, et al.. Genome-wide association study of Buruli ulcer in rural Benin highlights role of two LncRNAs and the autophagy pathway. *Communications Biology*, 2020, 3, pp.177. 10.1038/s42003-020-0920-6 . hal-02573225

**HAL Id: hal-02573225**

**<https://hal.sorbonne-universite.fr/hal-02573225>**

Submitted on 14 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Genome-wide association study of Buruli ulcer in rural Benin highlights role of two LncRNAs and the autophagy pathway

Jeremy Manry <sup>1,2,9</sup>✉, Quentin B. Vincent<sup>1,2,9</sup>, Christian Johnson<sup>3,4</sup>, Maya Chrabieh<sup>1,2</sup>, Lazaro Lorenzo<sup>1,2</sup>, Ioannis Theodorou<sup>5</sup>, Marie-Françoise Ardant<sup>3,6</sup>, Estelle Marion <sup>7</sup>, Annick Chauty<sup>3,6</sup>, Laurent Marsollier<sup>7</sup>, Laurent Abel<sup>1,2,8</sup> & Alexandre Alcaïs <sup>1,2</sup>✉

Buruli ulcer, caused by *Mycobacterium ulcerans* and characterized by devastating necrotizing skin lesions, is the third mycobacterial disease worldwide. 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 sample of 1524 well characterized patients and controls from rural Benin. Two-stage analyses identify two variants located within LncRNA 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*, previously shown to decrease bacterial autophagy (rs2241880,  $P = 0.003$ ; odds ratio = 0.31 [0.14–0.68]). Our results suggest LncRNAs and the autophagy pathway as critical factors in the development of Buruli ulcer.

<sup>1</sup>Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale (INSERM) UMR 1163, Paris, France. <sup>2</sup>Université de Paris, Imagine Institute, Paris, France. <sup>3</sup>Fondation Raoul Follereau, Paris, France. <sup>4</sup>Centre Interfacultaire de Formation et de Recherche en Environnement pour le Développement Durable. Université d'Abomey, Calavi, Benin. <sup>5</sup>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. <sup>6</sup>Centre de Dépistage et de Traitement de la Lèpre et de l'Ulcère de Buruli (CDTLUB), Pobè, Benin. <sup>7</sup>INSERM UMR-U892 and CNRS U6299, team 7, Angers University, Angers University Hospital, Angers, France. <sup>8</sup>St Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY, USA. <sup>9</sup>These authors contributed equally: Jeremy Manry, Quentin B. Vincent. ✉email: [jeremy.manry@inserm.fr](mailto:jeremy.manry@inserm.fr); [alexandre.alcais@inserm.fr](mailto:alexandre.alcais@inserm.fr)

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 of Buruli ulcer as a neglected emerging tropical disease and the launch of the Global Buruli Ulcer Initiative by the World Health Organization (WHO) 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 1000 [95% confidence interval (CI): 0.31–0.33] in Ivory Coast to 2.99 per 1000 [95%CI: 2.35–3.07] in Benin<sup>5</sup>. In 2018, the WHO reported a global 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, and 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>6–8</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>9</sup>. Considerable variability has been reported in the clinical presentation of Buruli ulcer, including spontaneous healing in both humans<sup>10,11</sup> and specific mouse strains<sup>11</sup>. This observation, together with the familial clustering of cases<sup>12,13</sup>, support 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>14,15</sup>. Rare genetic defects have also been shown to underlie severe forms of tuberculosis<sup>16,17</sup>, and homozygosity for a missense polymorphism of *TYK2* was recently identified as the most common monogenic cause of tuberculosis<sup>18,19</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>20</sup>. Genome-wide association studies (GWAS) have identified several common variants associated with leprosy<sup>21–23</sup>, while this kind of variants seems to play only a limited role in tuberculosis<sup>24</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, leprosy or tuberculosis. Significant association with Buruli ulcer was reported for common variants of six of these genes: *SLC11A1*<sup>25</sup>, *PRKN*, *NOD2*, and *ATG16L1*<sup>26</sup> and *iNOS* and

*IFNG*<sup>27</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 1524 individuals recruited at the Centre de Dépistage et de Traitement de la Lèpre et de l'Ulcère de Buruli (CDTLUB) of Pobè, Benin. Our analysis identifies two novel variants independently associated with the onset of Buruli ulcer and confirms the protective effect of a missense variant in the bacterial autophagy gene *ATG16L1*.

## 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 (Benin) in which Buruli ulcer is endemic (Table 1; Fig. 1). Principal component analysis (PCA) on genotyped variants with a minor allele frequency (MAF) > 0.05 revealed no evidence of population stratification in our sample, either at the global (Supplementary Fig. 1a) or African-specific (Supplementary Fig. 1b) level 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. Refined PCA performed only on individuals from our study ruled out cryptic stratification as cases and controls were evenly distributed across the two first components (Supplementary Fig. 1c). Following imputation, 10,014,109 high-quality autosomal variants were tested for association with two different phenotypic definitions of Buruli ulcer under the assumption of an additive genetic model.

We first analyzed the binary ‘case/control’ Buruli ulcer phenotype. Consistent with the refined PCA analysis, the quantile–quantile (Q–Q) plot (Supplementary Fig. 2a) showed no substantial deviation between the observed and the expected, i.e. asymptotic, distribution of the test statistics at the genome-wide level (genomic inflation factor  $\lambda = 1.027$ ). The results of the GWAS tests are summarized in a Manhattan plot (Fig. 2a). In total, 517 variants had *P* values <  $5 \times 10^{-5}$ , including 10 with *P* values <  $10^{-6}$ . The most significant signal for association was observed for an imputed SNP (rs7246288) located on chromosome 19 with a *P* value =  $2.80 \times 10^{-7}$ , a MAF of 0.09 and an odds ratio (OR) estimated at 2.33 [95%CI: 1.62–3.34] for developing Buruli ulcer for TT vs. CT or CT vs. CC carriers. After linkage disequilibrium (LD) pruning (see the ‘‘Methods’’ section for details), 66 promising variants (i.e. displaying *P* value <  $5 \times 10^{-5}$  and  $10^{-6}$  for genotyped and imputed variants, respectively) were retained for genotyping and association testing in the replication sample (Supplementary Data 1).

Age at diagnosis has been shown to provide additional information about the architecture of the genetic contribution to complex diseases<sup>28</sup>. We, therefore, performed a GWAS taking

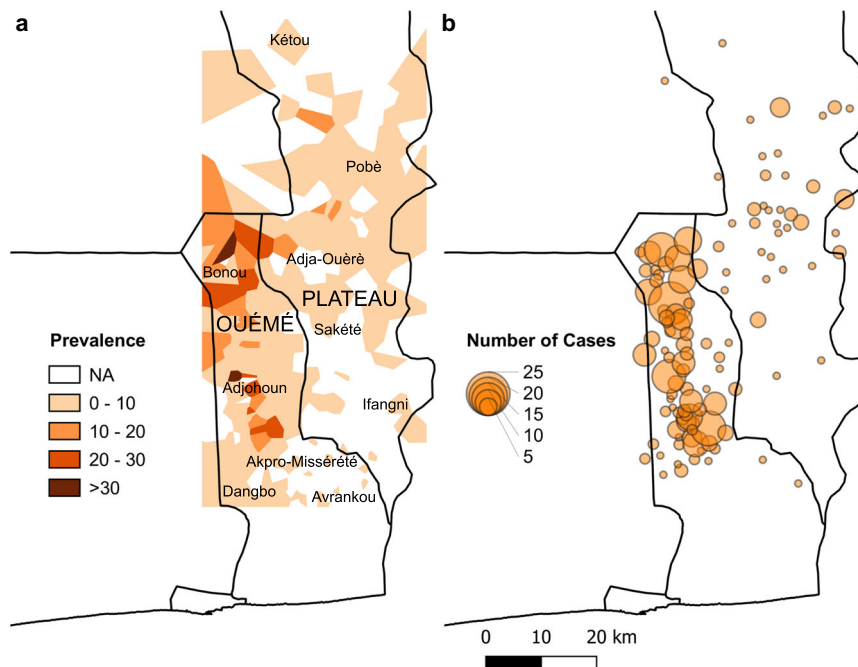
**Table 1 Distribution of sex and age according to the case/control status in the discovery and replication samples.**

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

SD standard deviation.

<sup>a</sup>Values are expressed as number (%).

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



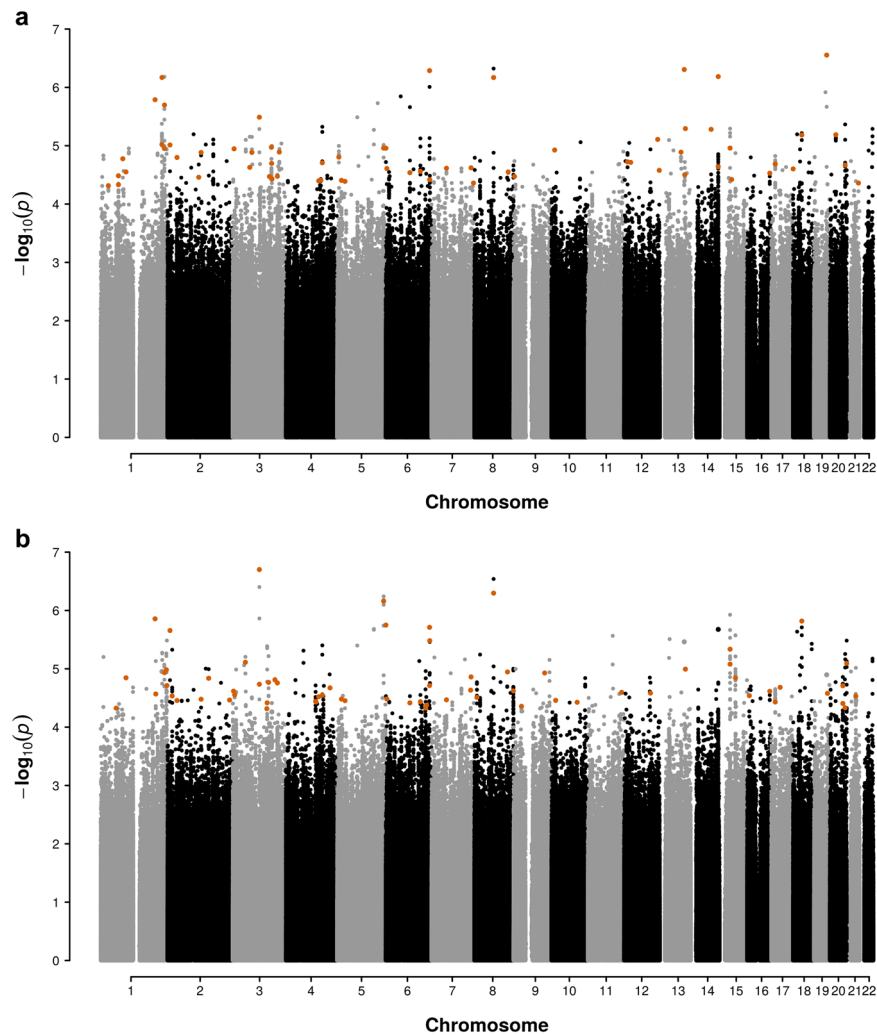
**Fig. 1** Geographic distribution of Buruli ulcer in Plateau and Ouémé in Benin. **a** Prevalence calculated from cases registered between 2005 and 2012 in the départements of OUÉMÉ and PLATEAU (upper case) and subsequently refined at municipality level (lower case); **b** Number of cases per village included in the discovery sample used for the GWAS.

age at diagnosis of Buruli ulcer into account by means of survival analytic tools (see the “Methods” section). As expected, Q–Q plot showed no substantial deviation ( $\lambda = 1.023$ ; Supplementary Fig. 2b). The results of the GWAS tests are summarized in a Manhattan plot (Fig. 2b). In total, 534 variants had a  $P$  value  $< 5 \times 10^{-5}$ , including nine with a  $P$  value  $< 10^{-6}$ . The most significant signal for association was observed for a genotyped SNP (rs34060873) located on chromosome 3 with a  $P$  value =  $1.99 \times 10^{-7}$ , a MAF = 0.10 and a hazard ratio (HR) estimated at 0.51 [95%CI: 0.39–0.67] for the development of Buruli ulcer for AA vs. CA or CA vs. CC carriers. Applying the same strategy as the one used for the binary phenotype, 68 variants were retained for genotyping and association testing in the replication sample (Supplementary Data 1). Among the 66 and 68 variants retained from the analysis of Buruli ulcer per se and the one taking age at onset into account, respectively, 29 displayed  $P$  values below our thresholds of selection in both analyses. For the replication analysis, the phenotypic definition providing the best  $P$  value for these variants was used. Altogether, analysis of the two phenotypic models led to the selection of 105 independent variants, 99 of which being genotyped and 6 being imputed (Supplementary Data 1).

**Replication study.** In the second phase of our study, these 105 variants were genotyped in an independent 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, five variants were excluded (see the “Methods” section) and 100 variants were retained for association testing in the replication sample with the same phenotypic model (binary or age at onset) and the same allelic effect as observed for the discovery sample (Supplementary Data 1). Out of the 100 high-quality variants tested for association with Buruli ulcer in this replication sample, rs9814705 and rs76647377 displayed significant evidence of replication, i.e. one-tailed  $P$  value of the association test  $< 0.01$  (Supplementary Data 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  $6.51 \times 10^{-4}$ . As this variant had been imputed in the discovery sample (imputation info value = 0.933), 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 Data 1) to  $1.67 \times 10^{-4}$  (Table 2). When merging our two samples, the combined  $P$  value for the association between rs9814705 and Buruli ulcer was  $2.85 \times 10^{-7}$ . The minor allele C (MAF = 0.14) was the risk allele, and the OR for developing Buruli ulcer for CC vs. TC or TC vs. TT carriers 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. 3, Supplementary Table 1). We then searched for additional variants in strong LD, i.e.  $r^2 > 0.5$ , with rs9814705, using the 1000 Genomes Phase III data for the YRI 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. 4). To confirm the hypothesis of a single signal driving the association with the onset of Buruli ulcer in this chromosomal region, we performed association tests with all local variants conditioning on rs9814705 (i.e., rs9814705 was considered as a covariate in the analysis of the other variants). No variant in LD ( $r^2 > 0.2$ , shown in Fig. 4) with rs9814705, including the three abovementioned SNPs, displayed  $P$  value  $< 0.05$  consistent with a single association signal accounted for by rs9814705 in this region (Supplementary Fig. 3a). It is important to note that this cluster of four SNPs in high LD ( $r^2 > 0.5$ ) spans 13.4 kb in a single intron of *ENSG00000240095.1*, a lncRNA of presently unknown function located on chromosome 3 with the closest identified protein-coding gene (*PLOD2*) located 107 kb away. While we cannot decide for the causal variant on the basis of  $P$  values, our results pinpoint this lncRNA as a solid candidate for further investigations.

The second evidence of replication was observed for the age of onset phenotype and rs76647377 on chromosome 6, for which



**Fig. 2** Manhattan plots of a GWAS of susceptibility to Buruli ulcer in a population from Benin. Distribution of observed  $P$  values for association between genetic variants and Buruli ulcer considering either a binary (affected/unaffected; **a**) or a censored (age at onset for Buruli ulcer patients and age at examination for the exposed controls; **b**) phenotype. The effects of variant genotypes on Buruli ulcer were tested 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 association are presented according to chromosomal positions (x-axis). Larger orange dots correspond to the 105 variants selected for replication.

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

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

<sup>a</sup>GRCh37.p13.

<sup>b</sup>Allele  $m$  is the minor allele and  $M$  is the major allele in the combined cohort.

<sup>c</sup>Discovery, replication and combined  $P$  values obtained when the logistic model (Buruli ulcer per se phenotype) is considered for rs9814705, and when the Cox model (taking age at onset into account) is considered for rs76647377. When considering mixed-models in the discovery sample, i.e. GEMMA for rs9814705 and coxme for rs76647377, discovery  $P$  values are  $2.13 \times 10^{-4}$  and  $2.06 \times 10^{-6}$ , respectively.

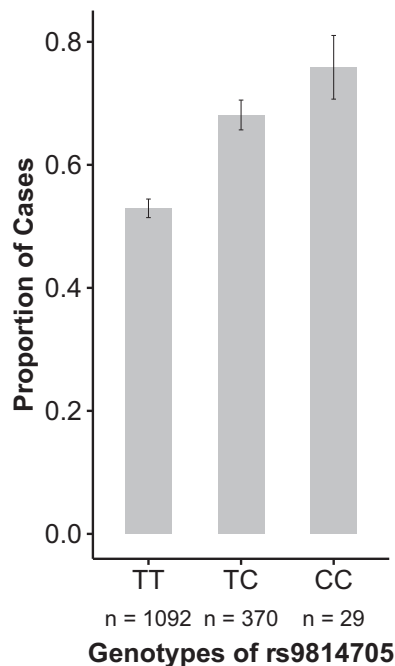
<sup>d</sup>One-tailed test.

<sup>e</sup>Effect size corresponds to odds ratios (95% confidence interval) when the logistic model (Buruli ulcer per se phenotype) is considered (rs9814705), and hazard ratios (95% confidence interval) when the Cox model (taking age at onset into account) is considered (rs76647377). Effect size is assessed under an additive model, using the minor allele as the reference.

\*rs9814705 was resequenced in the discovery sample.

the replication  $P$  value was  $3.26 \times 10^{-3}$ . When merging our two samples, the combined  $P$  value for the association between rs76647377 and Buruli ulcer was  $9.85 \times 10^{-8}$ . The minor allele A (MAF = 0.03) was protective and the HR for developing Buruli ulcer for AA vs. GA or GA vs. GG carriers was estimated at 0.41 [95%CI: 0.28–0.60] (Table 2). Stated differently, GG carriers are prone to develop Buruli ulcer at an earlier age than GA or AA

carriers (Fig. 5). Given that there was only one AA individual, who was a control, it is impossible to distinguish between an additive and a dominant effect for allele A. Following the same strategy as the one described above, screening of the region identified three SNPs displaying 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 =



**Fig. 3 Genotype at rs9814705 is associated with the onset of Buruli ulcer.**

Distribution of the proportion of Buruli ulcer cases (y-axis) according to rs9814705 genotype (x-axis) in the combined sample ( $n$ : number of individuals). Vertical bars represent standard errors.

$2.70 \times 10^{-3}$ ) (Fig. 6). To confirm the hypothesis of a single signal driving the association with the age of onset of Buruli ulcer in this chromosomal region, we performed association tests with all local variants conditioning on rs76647377. No variant in LD ( $r^2 > 0.2$ , shown in Fig. 6), including the aforementioned three SNPs, displayed  $P$  value  $< 0.05$  supporting a single association signal accounted for by rs76647377 in this region (Supplementary Fig. 3b). Although we cannot decide for the causal variant on the basis of  $P$  values, it is important to note that this cluster of four SNPs in high LD ( $r^2 > 0.5$ ) spans 9.7 kb in the intron of a lncRNA *LINC01622* of presently unknown function located on chromosome 6 (Table 2), with the closest reported protein-coding gene (*EXOC2*) located 292 kb away. Our results pinpoint this lncRNA as a good candidate for further investigations.

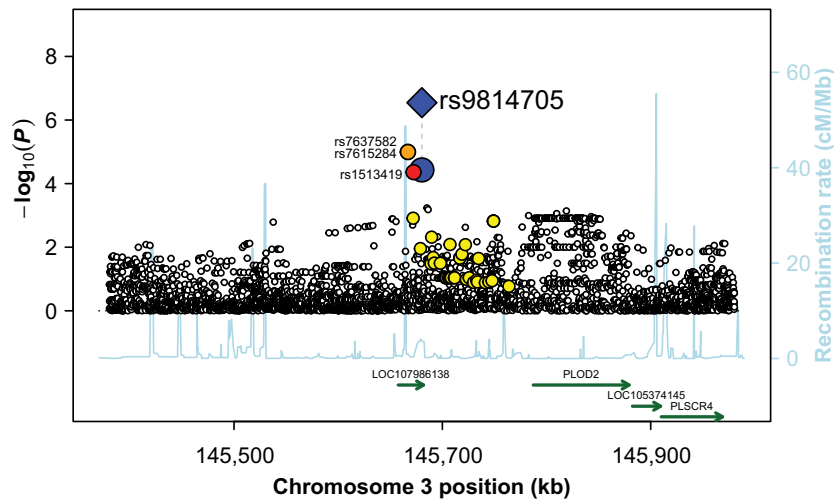
**Signals overlapping with previously reported associations.** As the microdeletion recently identified in a familial form of severe Buruli ulcer was close to a cluster of beta-defensin genes<sup>20</sup>, we investigated the role of 5455 variants located in 50 defensin genes and/or identified as eQTL for defensin genes (Supplementary Data 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 =  $4.13 \times 10^{-4}$ ). We then checked the evidence of association in our GWAS for the six variants displaying significant association with Buruli ulcer in the three candidate-gene association studies on Buruli ulcer performed to date<sup>25–27</sup> (Table 3). 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 (one-tailed  $P$  value = 0.009) when considering a recessive model for allele G (leading to the alanine amino acid change), as reported in the original work. Under this recessive model, the OR for developing Buruli ulcer for GG vs. GA or AA carriers was estimated at 0.46 [95%CI: 0.25–0.86]. Furthermore,

in the previous study<sup>26</sup>, GG homozygotes were found to be protected against the ulcerative clinical form of Buruli ulcer, with an OR of 0.35 [95%CI: 0.13–0.90]. Strikingly, when focusing on ulcerative cases of our discovery sample, the association was also more significant ( $P$  value = 0.003; OR = 0.31 [95%CI: 0.14–0.68]) despite the smaller sample size (255 ulcerative cases vs. 401 controls) (Table 3).

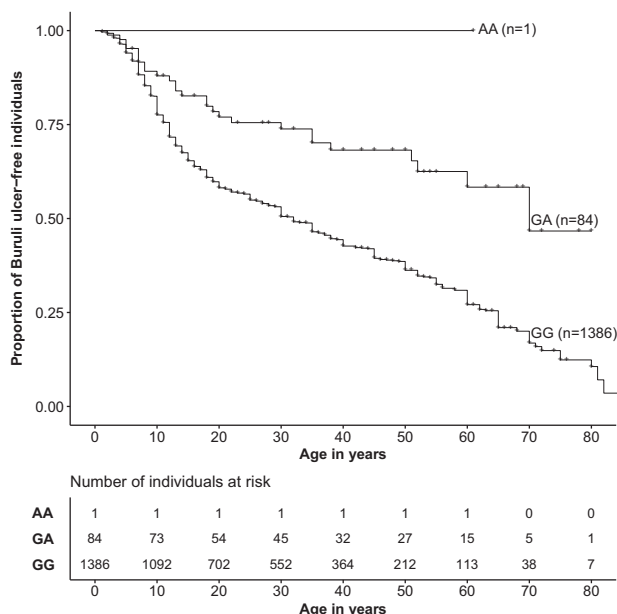
Finally, we also investigated whether some SNPs found to be associated with the other two most common mycobacterial diseases, tuberculosis and leprosy, overlapped with the SNPs identified in our GWAS on Buruli ulcer (see the “Methods” section for details). We first looked at the 105 SNPs selected for 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 [95%CI: 1.35–2.00])—was located in an intron of the *PACRG* gene, variants of which have been associated with leprosy<sup>29</sup> (Supplementary Data 1). We observed the same trend towards association in the replication sample (replication  $P$  value = 0.13, with an OR of 1.14 [95%CI: 0.91–1.42]). This variant is not in LD with variants of the *PRKN* (a.k.a. *PARK2*)/*PACRG* cluster associated with leprosy, including that reported to be associated with Buruli ulcer by Capela et al.<sup>26</sup> ( $r^2 = 0.007$ ) (Table 3). We then looked at the 35 SNPs reaching genome-wide significance in the GWAS performed on tuberculosis (eight independent variants in eight different chromosomal regions; Supplementary Data 3a) or leprosy (27 variants in 19 different chromosomal regions; Supplementary Data 3b). Using a type-I error of 0.01, none of these 35 SNPs was found to be significantly associated with Buruli ulcer. 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 straightforward 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<sup>31</sup>. The inaccessibility of many areas makes it difficult to set up GWAS focusing on a neglected tropical disease. In this context, to the best of our knowledge, we report the first GWAS investigating susceptibility to Buruli ulcer in a well-characterized sample, including more than 1500 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. Assuming a type I error of  $5 \times 10^{-5}$  (the threshold for a variant to be considered for replication), the discovery sample had a power of 80% for detecting variants with a MAF  $> 0.18$  and an OR  $> 1.80$  under the additive genetic model. Individuals were recruited for this study from the villages of Ouémé and Plateau, two areas of Benin in which Buruli ulcer is highly endemic (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 geographic clustering of *M. ulcerans* and its slow substitution rate suggest that patients from a given area are likely to be



**Fig. 4 Regional linkage disequilibrium plot for rs9814705.** Evidence of association with Buruli ulcer ( $-\log_{10}(P)$ ; left y-axis) for the variants located in the vicinity of rs9814705 (kb; x-axis). The distribution of recombination rates in this region is also given (cM/Mb; Right y-axis; light blue line). For the replicated SNP (i.e., rs9814705), the blue diamond corresponds to the result observed in the combined analysis, whereas the blue circle corresponds to the result 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:  $r^2 \geq 0.8$ , orange:  $0.5 \leq r^2 < 0.8$ , and yellow:  $0.2 \leq r^2 < 0.5$ ). Only variants with an  $r^2 > 0.5$  are labeled. Known genes in the chromosomal region are plotted, with arrows indicating their orientation.



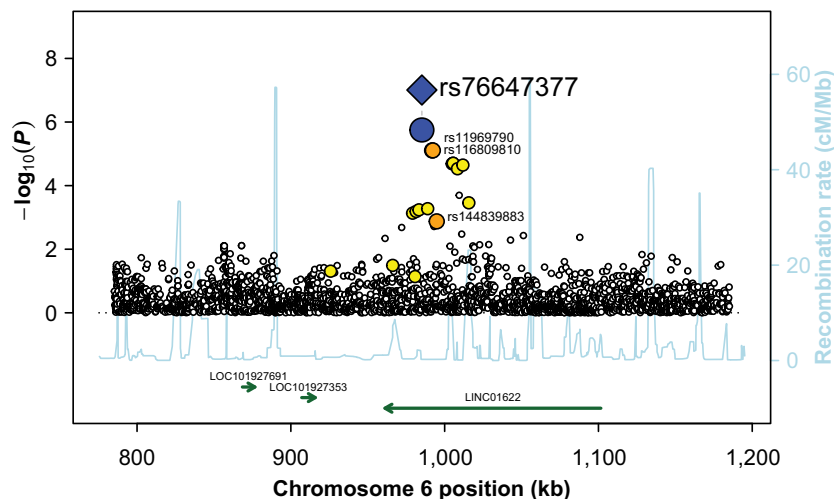
**Fig. 5 Genotype at rs76647377 is associated with age at onset of Buruli ulcer.** Kaplan-Meier curves for the onset of Buruli ulcer (y-axis) according to rs76647377 genotype. Number of individuals at risk are provided by genotype and by ten-year age bins.

infected by the same, or at least very similar *M. ulcerans* strains<sup>32,33</sup>. Strain variability, if any, would decrease the power of our GWAS, but not generate spurious signals of association.

Our study identified two independent loci associated with the onset (rs9814705,  $P$  value =  $2.85 \times 10^{-7}$ ) or the age at onset (rs76647377,  $P$  value =  $9.85 \times 10^{-8}$ ) of Buruli ulcer. These two variants are located in regions of the genome devoid of protein-coding genes: *PLOD2* is 107 kb away from rs9814705 and *EXOC2* is 292 kb away from rs76647377. Moreover, as expected in a sub-Saharan population, the LD pattern was of no help to resolve this issue. Both variants are located within the introns of LncRNAs: *ENSG00000240095.1* (rs9814705) and *LINC01622* (rs76647377).

Unfortunately, so far very limited functional data are available for *ENSG00000240095.1* and to the best of our knowledge none for *LINC01622*. Two correlated SNPs, rs9814705 and rs1513419 ( $r^2 = 0.87$  with rs9814705), are eQTL for LncRNA *ENSG00000240095.1* in esophagus-mucosa tissue ( $P$  value =  $3.9 \times 10^{-5}$  and  $P$  value =  $4.5 \times 10^{-5}$ , respectively)<sup>34</sup>. In the GTEx data, this LncRNA 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>35</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>36,37</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>38</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 LncRNAs.

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>26</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>39</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>40,41</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>34</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>42–44</sup>. In addition, *Atg16L1*-hypomorphic mice developed intestinal abnormalities resembling CD whilst displaying resistance to intestinal infection with the bacterium



**Fig. 6 Regional linkage disequilibrium plot for rs76647377.** Evidence of association with Buruli ulcer ( $-\log_{10}(P)$ ; left y-axis) for the variants located in the vicinity of rs76647377 (kb; x-axis). The distribution of recombination rates in this region is also given (cM/Mb; Right y-axis; light blue line). For the replicated SNP (i.e. rs76647377), the blue diamond corresponds to the result observed in the combined analysis, whereas the blue circle corresponds to the result 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:  $r^2 \geq 0.8$ , orange:  $0.5 \leq r^2 < 0.8$ , and yellow:  $0.2 \leq r^2 < 0.5$ ). Only variants with an  $r^2 > 0.5$  are labeled. Known genes in the chromosomal region are plotted, with arrows indicating their orientation.

**Table 3 Validation study for six variants located within six 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>
rs1333955	6	C/T	PARK2	Capela et al. <sup>26</sup>	Benin	0.24	Dom	1.43 (1.00-2.06)	0.05	0.27	1.05 (0.64-1.69)	0.357
rs9302752	16	C/T	NOD2	Capela et al. <sup>26</sup>	Benin	0.38	Dom	2.23 (1.14-4.37)	0.02	0.40	1.10 (0.72-1.67)	0.498
rs2241880	2	A/G	ATG16L1	Capela et al. <sup>26</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>
rs9282799	17	G/A	iNOS	Bibert et al. <sup>27</sup>	Ghana	0.08	Add	1.99 (1.22-3.26)	0.006	0.04	1.33 (0.82-2.14)	0.110
rs2069705	12	A/G	IFNG	Bibert et al. <sup>27</sup>	Ghana	0.47	Add	1.56 (1.14-1.99)	0.007	0.43	1.00 (0.82-1.22)	0.482
rs17235409	2	G/A	SCL11A1	Stienstra et al. <sup>25</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 under the same genetic model as the one used by the corresponding previous study.

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

*Citrobacter rodentium*<sup>45,46</sup>. Published findings suggest that autophagy may serve as a rheostat for immune reactions<sup>47</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/auto-immune diseases may reflect past selection for strong immune responses to combat infection<sup>48</sup>.

Despite important practical and methodological challenges<sup>49-51</sup>, we successfully conducted a GWAS on a neglected tropical disease in rural sub-Saharan Africa (Fig. 1), 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 lncRNAs 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>52</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 human genetic susceptibility to Buruli ulcer was approved by the Ethics Committee of the CDTLUB (Pobè, Benin), the International 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, 1524 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 is endemic<sup>53,54</sup>. Based on the cases registered in these two areas between 2005 and 2012, the prevalence in the studied villages was 5.3 per 1000; a detailed distribution by village of the cases recorded is shown in Fig. 1. The 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). 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). To limit the risk of misclassification of controls, we sampled controls living in the same endemic areas as cases, and who were older than the cases. As the population under study is sedentary, the older the individuals, the longer they have been exposed to *M. ulcerans* still remaining Buruli ulcer-free. The vast majority of cases were confirmed through the detection of *M. ulcerans* DNA by polymerase chain reaction (PCR) targeting the insertion element IS2404 which is the most sensitive and specific laboratory test currently available for the diagnostic of Buruli ulcer<sup>55,56</sup>. Two cases had positive culture examination only. Overall, 98.3% of cases in the discovery sample and 45% in the replication sample were laboratory-confirmed through PCR and/or culture examination. All the cases that were not laboratory-confirmed were classified as highly probable on the basis of stringent clinical criteria as defined by the WHO<sup>4</sup>. Therefore, the risk of misclassification of cases lacking laboratory confirmation was likely very low. This is best exemplified by a recent study estimating that in Buruli ulcer-endemic setting, the ability of trained clinicians to diagnose Buruli ulcer on the basis of clinical



criteria had a sensitivity of 92% [95%CI: 85–96%] and a specificity of 91% [95%CI: 81–96%]<sup>57</sup>. These estimates are particularly relevant in the context of our study as the CDTLUB was the main center of the abovementioned study providing approximately half of the total number of patients<sup>57</sup>.

**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% (eight samples), lack of concordance between reported and genetically inferred gender (three samples), or duplication (two samples). Our 'effective' discovery sample, therefore, consisted of a total of 803 individuals: 402 cases and 401 controls (Table 1).

In total, out of the 2,314,174 genotyped autosomal variants, a subset of 1,804,366, with a call rate >95%, a MAF >0.01, and a Hardy–Weinberg  $P$  value >  $10^{-6}$  in the group of controls were used for the two-step (phasing and genotype inference) imputation process. Haplotypes were phased with SHAPEIT2 v2.904<sup>58</sup>. For non-genotyped variants, genotype was inferred with IMPUTE2 v2.3.2<sup>59</sup>, using the 1000 Genomes Phase III dataset as the reference panel and default values for all other parameters<sup>60,61</sup>. Only variants imputed with an info value<sup>59</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 with Buruli ulcer in our discovery sample.

A type I error of less than  $5 \times 10^{-5}$  (genotyped variants) and  $10^{-6}$  (imputed variants) was used to select variants for further 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 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. Variants with a call rate <95% or a Hardy–Weinberg  $P$  value below  $10^{-4}$  in controls were excluded. Among the 105 selected variants, five were excluded. Three of these variants were imputed in the discovery sample, two of which were present in Alu sequences, including rs7246288 which was the best hit in the GWAS sample, and the other in an LRT element. 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 Data 1). The replication sample contained 708 individuals; 693 of these individuals (455 cases and 238 controls; Table 1) passed the quality control criteria as 4 samples with a call rate <95%, and 11 duplicated samples were excluded.

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>62</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.

**Statistics and reproducibility.** PCA 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>63</sup>, with the ad hoc functions implemented in PLINK v1.9<sup>64</sup>. To check for relatedness between individuals, we estimated the IBS between all individuals of the discovery sample using high quality genotyped variants with a MAF >5% by means of PLINK v1.9<sup>64</sup>. We identified 39 pairs of first degree relatives including 16 phenotypically concordant pairs (8 pairs of controls, 8 pairs of cases) and 23 phenotypically discordant pairs. We also identified 36 pairs of second-degree relatives including 17 concordant pairs (12 pairs of controls, 5 pairs of cases) and 19 discordant pairs. We investigated the possible influence of relatedness in our GWAS by the use of specific methods detailed in the next paragraph. By contrast, as only 105 variants were genotyped in the replication sample, we could not perform similar relatedness or PCA analysis in this sample. However, recruitment of individuals of the replication sample was performed in the same areas concomitantly with the discovery sample, and we expect both population structure and relatedness to be comparable between both samples.

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 Buruli ulcer status. Given that the controls were chosen to be older than the cases by design, we did not include age as a covariate, to avoid over-adjustment. Instead, effect of age was accounted for by specific survival analysis methods (see below). We also evaluated the potential impact of relatedness and cryptic population stratification 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>65</sup>. We observed a very significant 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 1000 variants, measured as the absolute difference between the  $-\log_{10}(P$  value) obtained in classical logistic regression and the one 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 of the PCA 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>66</sup>. In this analysis, we considered the Buruli ulcer diagnostic as the event of interest, i.e. the age at diagnosis being the failure time, and the age at which controls were recruited being the censored time. This age is indeed a good proxy of the duration of exposure since Buruli ulcer is highly endemic and the population is sedentary in the villages where the recruitment took place. The analysis strategy was identical to that used for the binary phenotype, i.e. we evaluated the potential impact of relatedness and cryptic population stratification 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>67</sup>. Again, the correlation between the  $P$  values of the 1000 top variants for the classical and mixed-effect Cox models was highly significant ( $r = 0.97$ , Spearman  $P$  value <  $10^{-16}$  (Supplementary Fig. 4b)). The largest difference within these top 1000 variants, measured as the absolute difference between the  $-\log_{10}(P$  value) obtained in classical Cox analysis and the one obtained with the mixed-effect Cox analysis was less than 0.28. Consistently, the use of the first three principal components of the PCA had no impact on the results.

Altogether, the analyses performed above confirmed the very limited impact of relatedness and/or population structure on the association results observed in our discovery sample using either the logistic (Buruli ulcer per se phenotype) or the Cox (taking age at onset into account) models. This is likely explained by the balanced distribution of phenotypically concordant and discordant relative pairs, and the ethnic homogeneity of our population. Therefore, all the results reported in the manuscript are 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 are the ones obtained by likelihood-ratio tests. All association analyses were performed with R (<https://www.r-project.org/>), PLINK 1.9<sup>64</sup>, GEMMA<sup>65</sup>, coxme<sup>67</sup>, SNPTEST v2.5.2<sup>68</sup>, and ProbABEL<sup>69</sup> software.

For the replication study, we defined as significant a variant having the same allelic effect under the same genetic (additive) model as in the discovery sample, with a one-tailed  $P$  value < 0.01 in the replication sample. Similar replication thresholds (0.01 or 0.05) have been used in previous GWAS after having selected a similar number of variants (~100) for replication<sup>21,70,71</sup>, and appears as a reasonable trade-off to limit the risk of false positive and false negative results. In our analysis the final interpretation of a replicated variant was based on the  $P$  value combining the results of the discovery and the replication samples that was compared to the classical genome-wide  $5 \times 10^{-8}$  threshold.

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 an additive genetic model and a type I error of  $5 \times 10^{-5}$ . This threshold is a pragmatic trade-off to limit the number of false positives and false negatives that was used before<sup>70</sup>. This calculation was performed under the assumption of absence of misclassification of controls. Although we chose controls older than cases and living in the same endemic areas, we cannot not completely rule out some degree of misclassification of controls that would lead to a reduced power. Power estimates are shown according to the MAF and effect size, as estimated by the OR of the tested variants (Supplementary Fig. 5). For example, our sample has 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<sup>72</sup>.

**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 common mycobacterial diseases, i.e. leprosy and tuberculosis. We recently identified a microdeletion in a familial form of Buruli ulcer that overlaps with a cluster of genes encoding defensins<sup>20</sup>. Therefore, we identified all the variants located within the 50 defensin genes (including those 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 significant (FDR < 0.05) eQTL for these defensin genes in whole-blood or skin from the lower legs exposed to the sun in the GTEx database<sup>34</sup>. Results were available for 5455 variants, of which 5450 were located in 50

defensin genes and 91 were identified as eQTL for defensin genes (86 of these eQTL were also located within defensin genes  $\pm 10$  kb) (Supplementary Data 2). We then searched for variants reported to be associated with Buruli ulcer in the three published candidate-gene studies<sup>25–27</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<sup>73</sup> for variants associated with either tuberculosis or leprosy at the genome-wide level ( $P$  value  $< 5 \times 10^{-8}$ ). We identified 35 SNPs, eight in TB, and 27 in leprosy, which we then assessed for association with Buruli ulcer in our discovery sample.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

The data that support the findings of this study are available within the paper and its Supplementary Information files. GWAS summary statistics for the two phenotypes studied, and source data for generating Fig. 5 are available at <https://doi.org/10.6084/m9.figshare.11941344>. All other relevant data are available from J.M. (jeremy.manry@inserm.fr) and L.A. (laurent.abel@inserm.fr) upon request.

Received: 6 November 2019; Accepted: 24 March 2020;

Published online: 20 April 2020

### References

- Wansbrough-Jones, M. & Phillips, R. Buruli ulcer: emerging from obscurity. *Lancet* **367**, 1849–1858 (2006).
- World Health Organization. Neglected tropical diseases, hidden successes, emerging opportunities (2009).
- Sizaire, V., Nackers, F., Comte, E. & Portaels, F. *Mycobacterium ulcerans* infection: control, diagnosis, and treatment. *Lancet Infect. Dis.* **6**, 288–296 (2006).
- World Health Organization. Buruli ulcer (*Mycobacterium ulcerans* infection) Fact Sheet (2018).
- Simpson, H. et al. Mapping the global distribution of Buruli ulcer: a systematic review with evidence consensus. *Lancet Glob. Health* **7**, e912–e922 (2019).
- 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).
- George, K. M. et al. Mycolactone: a polyketide toxin from *Mycobacterium ulcerans* required for virulence. *Science* **283**, 854–857 (1999).
- Marion, E. et al. Mycobacterial toxin induces analgesia in buruli ulcer by targeting the angiotensin pathways. *Cell* **157**, 1565–1576 (2014).
- Vincent, Q. B., Ardant, M. F., Marsollier, L., Chauty, A., Alcais, A. & Franco-Beninese Buruli Research, G. HIV infection and Buruli ulcer in Africa. *Lancet Infect. Dis.* **14**, 796–797 (2014).
- O'Brien, D. P., 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).
- 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).
- O'Brien, D. P. 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).
- Sopoh, G. E. et al. Family relationship, water contact and occurrence of Buruli ulcer in Benin. *PLoS Negl. Trop. Dis.* **4**, e746 (2010).
- Bustamante, J., Boisson-Dupuis, S., Abel, L. & Casanova, J. L. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN- $\gamma$  immunity. *Semin Immunol.* **26**, 454–470 (2014).
- Casanova, J. L. & Abel, L. The genetic theory of infectious diseases: a brief history and selected illustrations. *Annu Rev. Genomics Hum. Genet.* **14**, 215–243 (2013).
- Alcais, A., Fieschi, C., Abel, L. & Casanova, J. L. Tuberculosis in children and adults: two distinct genetic diseases. *J. Exp. Med.* **202**, 1617–1621 (2005).
- Boisson-Dupuis, S. et al. Inherited and acquired immunodeficiencies underlying tuberculosis in childhood. *Immunol. Rev.* **264**, 103–120 (2015).
- 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**, eaau8714 (2018).
- 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. USA* **116**, 10430–10434 (2019).
- Vincent, Q. B. et al. Microdeletion on chromosome 8p23.1 in a familial form of severe Buruli ulcer. *PLoS Negl. Trop. Dis.* **12**, e0006429 (2018).
- Zhang, F. R. et al. Genomewide association study of leprosy. *N. Engl. J. Med.* **361**, 2609–2618 (2009).
- Zhang, F. et al. Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. *Nat. Genet.* **43**, 1247–1251 (2011).
- Liu, H. et al. Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. *Nat. Genet.* **47**, 267–271 (2015).
- 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).
- Stienstra, Y. et al. Susceptibility to Buruli ulcer is associated with the SLC11A1 (NRAMP1) D543N polymorphism. *Genes Immun.* **7**, 185–189 (2006).
- 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).
- Bibert, S. et al. Susceptibility to *Mycobacterium ulcerans* disease (Buruli ulcer) is associated with IFNG and iNOS gene polymorphisms. *Front Microbiol.* **8**, 1903 (2017).
- Alcais, A., Quintana-Murci, L., Thaler, D. S., Schurr, E., Abel, L. & Casanova, J. L. Life-threatening infectious diseases of childhood: single-gene inborn errors of immunity? *Ann. N. Y. Acad. Sci.* **1214**, 18–33 (2010).
- Mira, M. T. et al. Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* **427**, 636–640 (2004).
- Qi, H. et al. Discovery of susceptibility loci associated with tuberculosis in Han Chinese. *Hum. Mol. Genet.* **26**, 4752–4763 (2017).
- 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).
- Coudereau, C. et al. Stable and local reservoirs of *Mycobacterium ulcerans* inferred from the nonrandom distribution of bacterial genotypes, Benin. *Emerg. Infect. Dis.* **26**, 491–503 (2020).
- Vandellannoote, K. et al. Multiple introductions and recent spread of the emerging human pathogen *Mycobacterium ulcerans* across Africa. *Genome Biol. Evol.* **9**, 414–426 (2017).
- G. TEx Consortium. The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
- Kim-Hellmuth, S. et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat. Commun.* **8**, 266 (2017).
- Dong, B. et al. Phospholipid scramblase 1 potentiates the antiviral activity of interferon. *J. Virol.* **78**, 8983–8993 (2004).
- Sivagnanam, U., Palanirajan, S. K. & Gummadi, S. N. The role of human phospholipid scramblases in apoptosis: An overview. *Biochim Biophys. Acta Mol. Cell Res.* **1864**, 2261–2271 (2017).
- Sambarey, A. et al. Meta-analysis of host response networks identifies a common core in tuberculosis. *NPJ Syst. Biol. Appl.* **3**, 4 (2017).
- 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).
- Lassen, K. G. et al. Atg16L1 T300A variant decreases selective autophagy resulting in altered cytokine signaling and decreased antibacterial defense. *Proc. Natl Acad. Sci. USA* **111**, 7741–7746 (2014).
- Lassen, K. G. & Xavier, R. J. Mechanisms and function of autophagy in intestinal disease. *Autophagy* **14**, 216–220 (2018).
- 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).
- Kenny, E. E. et al. A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. *PLoS Genet.* **8**, e1002559 (2012).
- Rioux, J. D. 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).
- Cadwell, K. Crosstalk between autophagy and inflammatory signalling pathways: balancing defence and homeostasis. *Nat. Rev. Immunol.* **16**, 661–675 (2016).
- Marchiando, A. M. et al. A deficiency in the autophagy gene Atg16L1 enhances resistance to enteric bacterial infection. *Cell Host Microbe.* **14**, 216–224 (2013).
- Martin, P. K. et al. Autophagy proteins suppress protective type I interferon signalling in response to the murine gut microbiota. *Nat. Microbiol.* **3**, 1131–1141 (2018).
- Barreiro, L. B. & Quintana-Murci, L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat. Rev. Genet.* **11**, 17–30 (2010).
- Teo, Y. Y., Small, K. S. & Kwiatkowski, D. P. Methodological challenges of genome-wide association analysis in Africa. *Nat. Rev. Genet.* **11**, 149–160 (2010).

50. Rosenberg, N. A., Huang, L., Jewett, E. M., Szpiech, Z. A., Jankovic, I. & Boehnke, M. Genome-wide association studies in diverse populations. *Nat. Rev. Genet.* **11**, 356–366 (2010).
51. Peprah, E., Xu, H., Tekola-Ayele, F. & Royal, C. D. 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).
52. Bieri, R., Bolz, M., Ruf, M. T. & 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).
53. Johnson, R. C., Sopoh, G. E., Barogui, Y., Dossou, A., Fourn, L. & Zohoun, T. Surveillance system for Buruli ulcer in Benin: results after four years. *Sante* **18**, 9–13 (2008).
54. Wagner, T., Benbow, M. E., Brenden, T. O., Qi, J. & Johnson, R. C. 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).
55. 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).
56. Zingue, D., Bouam, A., Tian R., B., D., & Drancourt, M. Buruli Ulcer, a prototype for ecosystem-related infection, caused by *Mycobacterium ulcerans*. *Clin. Microbiol. Rev.* **31**, e00045–17 (2018).
57. Eddyani, M. et al. Diagnostic accuracy of clinical and microbiological signs in patients with skin lesions resembling Buruli ulcer in an endemic region. *Clin. Infect. Dis.* **67**, 827–834 (2018).
58. Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).
59. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
60. Genomes Project Consortium. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
61. van Leeuwen, E. M. et al. Population-specific genotype imputations using minimac or IMPUTE2. *Nat. Protoc.* **10**, 1285–1296 (2015).
62. 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).
63. Sudmant, P. H. et al. An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).
64. Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M. & Lee, J. J. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
65. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
66. Cox, D. R. Regression models and life-tables. *J. R. Stat. Soc. Ser. B (Methodol.)* **34**, 187–220 (1972).
67. Therneau T. coxme: Mixed Effects Cox Models. R package version 2.2-3. <http://CRAN.R-project.org/package=coxme> (2012).
68. 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).
69. Aulchenko, Y. S., Struchalin, M. V. & van Duijn, C. M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinforma.* **11**, 134 (2010).
70. Patin, E. et al. Genome-wide association study identifies variants associated with progression of liver fibrosis from HCV infection. *Gastroenterology* **143**, 1244–1252 e1212 (2012).
71. Thye, T. et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat. Genet.* **42**, 739–741 (2010).
72. Gauderman, W. J. Sample size requirements for matched case-control studies of gene-environment interaction. *Stat. Med.* **21**, 35–50 (2002).
73. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005–D1012 (2019).

## Acknowledgements

We thank Marie Kempf and Jane Cottin (Laboratoire de Bactériologie, CHU d'Angers, Angers, France), Jean-Paul Saint-André (Laboratoire d'Anatomie Pathologique, CHU d'Angers, Angers, France), Ambroise Adeye and Aimé Goundote (CDTLUB, Pobè, Benin), and Jean Gabin Houezo and Didier Agossadou (Programme de Lutte Contre la Lèpre et l'Ulcère de Buruli, Ministère de la Santé, Cotonou, Benin). 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 Guernon, Grégoire Hure, Natasha Vladikine, and Myriam Berrandane 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; grants no. ANR-12-BSV3-0013-01 and ANR-17-BSV3-0013-01). J.M., L.A., and A.A. were supported by the Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant no. ANR-10-LABX-62-IBED) and the ANR under the "Investments for the Future" program (grant no. ANR-10-LAHU-01). QBV was supported by the Fondation Bettencourt Schueller through the MD/PhD program of the Imagine Institute. A.A. 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.-F.A., C.J., L.M., E.M., and A.C. collected the data. M.-F.A., L.M., E.M., and A.C. did or confirmed diagnostics. Q.V., J.M., M.F.A., E.M., C.J., A.C., M.C., L.L., I.T., and A.A. acquired the data. Q.V., J.M., L.A., and A.A. did the statistical analysis. J.M., L.A., and A.A. drafted the report. A.A. obtained the funding.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42003-020-0920-6>.

Correspondence and requests for materials should be addressed to J.M. or A.Ais.

Reprints and permission information is available at <http://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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020