A Genome-Wide Association Study Identifies \textit{SERPINB10}, \textit{CRLF3}, \textit{STX7}, \textit{LAMP3}, \textit{IFNG-AS1} and \textit{KRT80} As Risk Loci Contributing to Cutaneous Leishmaniasis In Brazil

Léa C. Castellucci, Lucas Almeida, Svetlana Cherlin, Michaela Fakiola, Richard W. Francis, Edgar M. Carvalho, Anadilton Santos da Hora, Tainã Souza do Lago, Amanda B. Figueiredo, Clara M. Cavalcanti, Natalia S. Alves, Katia LP Morais, Andréa Teixeira-Carvalho, Walderez O. Dutra, Kenneth J. Gollob, Heather J. Cordell, and Jenefer M. Blackwell

*Contributed equally

1National Institute of Science and Technology in Tropical Diseases, Brazil; 2Federal University of Bahia, Salvador, Brazil; 3Population Health Sciences Institute, Newcastle University, UK; 4INGM-National Institute of Molecular Genetics “Romeo ed Enrica Invernizzi” Milan, Milan, Italy; 5Telethon Kids Institute, The University of Western Australia, Western Australia; 6International Center for Research, AC Camargo Cancer Center, São Paulo, Brazil; 7Instituto Rene Rachou of Fundação Oswaldo Cruz (FIOCRUZ-Minas), Belo Horizonte, Brazil; 8Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; 9Núcleo de Ensino e Pesquisa, Instituto Mario Penna, Belo Horizonte, Brazil; 10Department of Pathology, University of Cambridge, UK; 11Present IFOM, the FIRC Institute of Molecular Oncology, Milan, Italy

© The Author(s) 2020. Published by Oxford University Press for the Infectious Diseases Society of America.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Corresponding author: Jenefer M. Blackwell (jenefer.blackwell@telethonkids.org.au)

PO Box 855, West Perth, Western Australia 6872; Hospital Avenue, Nedlands, Western Australia 6009; Phone: +61 8 63191000.
Summary: Genome-wide analysis of 2066 cases and 2046 controls, together with genotypic differences in antigen-specific interferon-γ made by CD3+ T cells, identifies IFNG-AS1 as a genetic risk factor for cutaneous leishmaniasis caused by *Leishmania braziliensis*.

A GWAS for cutaneous leishmaniasis in Brazil
Abstract

**Background.** Our goal was to identify genetic risk factors for cutaneous leishmaniasis (CL) caused by *Leishmania braziliensis*.

**Methods.** Genotyping 2066 CL cases and 2046 controls using Illumina HumanCoreExomeBeadChips provided data for 4,498,586 imputed single nucleotide variants (SNVs). Genome-wide association testing using linear mixed models took account of genetic diversity/ethnicity/admixture. Post-GWAS positional, expression quantitative trait locus (eQTL), and chromatin interaction mapping was performed in FUMA. Transcriptional data were compared between lesions and normal skin, and cytokines measured using flow cytometry and Bioplex assay.

**Results.** Positional mapping identified 32 genomic loci associated with CL, none achieving genome-wide significance (P<5x10^{-8}). Lead SNVs at 23 loci occurred at protein coding or non-coding RNA genes, 15 with eQTLs for functionally relevant cells/tissues and/or showed differential expression in lesions. Of these, the 6 most plausible genetic risk loci were: *SERPINB10* ($P_{\text{imputed}_{\text{1000G}}}=2.67x10^{-6}$), *CRLF3* ($P_{\text{imputed}_{\text{1000G}}}=5.12x10^{-6}$), *STX7* ($P_{\text{imputed}_{\text{1000G}}}=6.06x10^{-6}$), *KRT80* ($P_{\text{imputed}_{\text{1000G}}}=6.58x10^{-6}$), *LAMP3* ($P_{\text{imputed}_{\text{1000G}}}=6.54x10^{-6}$) and *IFNG-AS1* ($P_{\text{imputed}_{\text{1000G}}}=1.32x10^{-5}$). *LAMP3* ($P_{\text{adjusted}}=9.25x10^{-12}$; +6-fold), *STX7* ($P_{\text{adjusted}}=7.62x10^{-7}$; +1.3-fold) and *CRLF3* ($P_{\text{adjusted}}=9.19x10^{-5}$; +1.97-fold) were expressed more highly in CL biopsies compared to normal skin; *KRT80* ($P_{\text{adjusted}}=3.07x10^{-8}$; -3-fold) was lower. Multiple cis-eQTLs across *SERPINB10* mapped to chromatin interaction regions of transcriptional/enhancer activity in neutrophils, monocytes, B cells and haematopoietic stem cells. Those at *IFNG-AS1* mapped to transcriptional/enhancer regions in T, natural killer, and B cells. The percent peripheral blood CD3^+ T cells making antigen-specific interferon-γ differed significantly by *IFNG-AS1* genotype.
Conclusions. This first GWAS for CL identified multiple genetic risk loci including a novel lead to understanding CL pathogenesis through regulation of interferon-\(\gamma\) by IFNG antisense RNA 1.

Key words: Leishmania; GWAS; post-GWAS integrated analysis; interferon-gamma; IFNG-AS1
INTRODUCTION

American cutaneous leishmaniasis (ACL) caused by *Leishmania braziliensis* has multiple presentations including cutaneous (CL), mucosal (ML), and disseminated (DL) leishmaniasis. ML and DL are generally preceded by CL. The common CL form of disease is associated with localized skin lesions, mainly ulcers, on exposed body parts. Whilst normally self-limiting, the degree of pathology and rate of healing varies, with lesions leaving life-long scars. Not all infected individuals go on to develop disease. Subclinical infection is associated with *Leishmania*-specific cellular immune responses, measured as delayed type hypersensitivity (DTH) skin test responses [1]. *Leishmania* antigen-stimulated peripheral blood lymphocytes also produce interferon-γ (IFN-γ) and tumour necrosis factor (TNF) in subclinical infection, but at lower levels than CL [1]. In a longitudinal study, IFN-γ was associated with protection, but a positive skin-test response was not [2]. Indeed, a positive DTH response has high sensitivity for diagnosis of *L. braziliensis* CL [1, 3]. All forms of ACL are associated with exaggerated cellular immunity. In CL, there is a positive correlation between the frequency of CD4+ T cells expressing IFN-γ and TNF and lesion size [4], with higher levels in ML than CL [5]. The outcome of *L. braziliensis* infection is determined by a fine balance between pro-inflammatory IFN-γ and TNF and anti-inflammatory interleukin-10 (IL-10) [3, 5].

One question is whether host genetics influences these responses. Racial differences, familial clustering and murine studies support genetic control of leishmaniasis (reviewed [6]). Human family-based genetic epidemiology of CL caused by *L. peruviana*, a member of the *L. braziliensis* species complex, was consistent with a gene by environment multifactorial model, a two-locus model of inheritance providing best fit [7]. This suggested that major genetic risk factors might be found for CL. Candidate gene studies [8-16] of *L. braziliensis*
complex suggest that multiple genes associated with pro- and anti-inflammatory responses (TNFA, SLC11A1, CXCR1, IL6, IL10, CCL2/MCP1, IFNG) and/or wound healing (FLI1, CTGF, TGFBR2, SMAD2, SMAD3, SMAD7, COL1A1) influence CL or ML disease. Although frequently underpinned by functional data [12-14, 16] and/or supported by prior immunological studies [17, 18], these studies have generally lacked statistical power. 

Here we perform the first well-powered genome-wide association study for *L. braziliensis* CL, combining analysis across two cohorts comprising 2066 cases and 2046 controls. Integrative post-GWAS analysis [19] is used to positionally map genomic loci associated with CL, with functional annotation and experimental studies used to identify plausible genes that act as genetic risk factors for CL.

**METHODS**

**Ethical Considerations, Sampling and Clinical Data Collection**

The study, approved by the Hospital Universitário Professor Edgard Santos Ethical Committee (018/2008 and 22/2012) and the Brazilian National Ethical Committee (CONEP–305/2007; CONEP–1258513.1.000.5537), complied with principles of the Helsinki declaration. All participants or parents/guardians signed written consents. Post-quality control (QC) genotype data are lodged in the European Genome-phenome Archive (accession number EGAS00001004596). CL cases were ascertained at the Public Health Post, Corte de Pedra, Bahia, Brazil, where *L. braziliensis* is the confirmed species [9-12]. CL is defined as presence of chronic ulcerative lesions without mucosal involvement (ML) or dissemination to ≥10 sites (DL). ML and DL cases were excluded due to insufficient power. All CL cases had confirmed parasite detection and/or minimally met two of three criteria: positive leishmania-specific DTH, positive leishmania serology, leishmania histopathology. Endemic controls were attendants of cases with no current/previous history of CL, DL or ML,
including no scars. Samples were collected in two phases: 2008-2010 and 2016-2017. Blood bank controls were collected 2015-2017 at HEMOBA Foundation, Salvador. Demographic data (age, sex) were recorded. Blood (8 ml) was taken by venipuncture into dodecyl citrate acid-containing vacutainers (Becton Dickinson). Genomic DNA was prepared using proteinase K and salting-out and shipped to UK for genotyping at Cambridge Genomic Services, UK.

**Array Genotyping and Marker QC**

DNAs were genotyped on Illumina Infinium® HumanCoreExome Beadchips (Illumina Inc., San Diego, CA, USA) with probes for 551,004 single nucleotide variants (SNVs): 282,373 informative across ancestries; 268,631 exome-focused. Human genome build 37 (hg19) was used. Exclusions were individuals with missing data rate >5%, SNVs with genotype missingness >5%, minor allele frequency (MAF) <0.01, or deviation from Hardy-Weinberg equilibrium (threshold \( P < 1.0 \times 10^{-8} \)). Post-QC datasets comprised: 312,503 genotyped SNVs, 956 CL cases, 868 controls phase 1; 298,919 SNVs, 1110 CL cases, 1178 controls phase 2. Phase 1 and 2 had 52% and 81% power, respectively, the combined sample 99% power, to detect genome wide significance (\( P < 5 \times 10^{-8} \); [20]) assuming a disease allele frequency 0.25, effect size (genotype relative risk) 1.5, and disease prevalence 2%.

**SNV Imputation and GWAS**

Imputation was performed using the multi-ethnic 1000 Genomes Project phase 3 reference panel (1000G): 84.8 million variants; 2504 samples; 26 populations. The 293,563 post-QC genotyped SNVs common across phases 1 and 2 were imputed using the Michigan Imputation Server v1.0.4 [21]. Imputed SNVs with information metric <0.8 or genotype probability <0.9 were excluded. Remaining variants were converted to genotype calls and
filtered for <5% missingness and MAF>0.005. Imputation accuracy was assessed as the squared Pearson correlation between imputed SNV dosage and known allele dosage (r²>0.5).

Genome-wide association analysis was performed using a linear mixed model in FaST-LMM v2.07 under an additive model [22]. Population structure/relatedness were controlled using the genetic similarity matrix, computed from 32,696 phase 1 and 45,569 phase 2 linkage disequilibrium (LD)-pruned array variants. Systematic confounding was assessed using quantile-quantile (Q-Q) plots and an inflation factor (denoted λ; median observed/median theoretical chi-squared distributions). Manhattan plots were generated in R using mhtplot() in the genetic analysis package 'gap'. Regional association plots were created using LocusZoom [23]. The 32,696 phase 1 and 45,569 phase 2 LD-pruned variants were matched to HapMap populations and PCA plots prepared in R.

Post-GWAS annotation in FUMA

Functional Mapping and Annotation (FUMA) [19] was used to characterise regions of association based on positional, expression quantitative trait loci (eQTL) and chromatin interaction mapping. Summary statistics from the combined GWAS were loaded into FUMA. SNP2GENE was used to identify independent significant SNVs based on 1000G multi-ethnic LD data. SNP2GENE mapping used the default GWAS P<10⁻⁵ plus one manually entered seed hit at P=1.32x10⁻⁵. Independent significant SNVs and SNVs in LD with them were annotated for consequences on gene function using ANNOVAR, potential regulatory functions (Regulome DB score), and 15-core chromatin state predicted by ChromHMM for 127 tissue/cell types. Effects of SNVs on gene expression were determined using eQTLs from multiple tissue/cell types of healthy donors from databases: eQTLgen (44 different tissue types); BIOSQTL (BIO_eQTL_gene level, whole peripheral blood, 2,116...
healthy donors); DICE (B and T cells, monocytes, NK cells); and GTEx v8 (whole blood; cultured fibroblasts; skin exposed and not sun exposed).

Expression analysis in CL lesions

RNA expression for mapped genes was examined using published microarray data [24] comparing CL lesion biopsies (N=25) with normal skin (N=10) from non-endemic unexposed donors (GEO database: GSE55664). Between group comparisons were made on log transformed data using the GEO2R tool with Benjamini and Hochberg false discovery rate adjusted P-values.

Cytokine and Antigen-stimulated T cell responses

Plasma IFN$_\gamma$ was measured using BioPlex-220$^\circledR$ (Bio-Rad Laboratories Inc) with Cytokine Grp-I-panel 27-plex. Peripheral blood mononuclear cells from a subset (N=40) of untreated phase 2 CL patients were separated from heparinized blood over Ficoll and used to examine T cell responses by $IFNG-ASI$ genotype. Cells (1x10$^6$ cells/ml) were stimulated; 37°C/5%CO$_2$, 10 $\mu$g/mL $L$. braziliensis (strain MHOM / BR / 2001) log phase promastigote soluble Leishmania antigen, 1 $\mu$g/mL purified NA/LE anti-human CD28 (clone CD28.2, BD Biosciences, San Jose, CA, USA) 15h, then brefeldin A (BD Biosciences) 4h. Washed cells (PBS/0.2%BSA) were incubated (4°C;30min) with BUV661 anti-human CD3 monoclonal antibody (UCHT1 clone, BD Biosciences). Cells were fixed, washed, permeabilized using BD Cytofix/Cytoperm and incubated (4°C;30min) with BV605 anti-human IFN-$\gamma$ (B27 clone) and BUV395 anti-human TNF-alfa (MAB11 clone, BD Biosciences) in permeabilization buffer. Live/dead cells were distinguished using Fixable Viability Stain 575V (BD Biosciences). Data were acquired by BD FACSymphony A5 flow cytometry and
analysed using FlowJo 10.6.1 software (BD Biosciences), and differences between genotypes determined using non-parametric Kruskal-Wallis ANOVA with multiple comparisons.

RESULTS

Characteristics of the Study Population

Demographic and clinical details comparing cases and controls are provided in Supplementary Table 1. The younger age of endemic controls was counterbalanced by older age of blood bank controls. Phase 1 and 2 CL cases were matched for lesion number/size and DTH, with no correlation between lesion and DTH sizes. Blood bank controls fell within genetic heterogeneity of endemic controls (Supplementary Figures 1 and 2), with all controls matched to cases. A few outliers occurred in phase 1 endemic controls, which also showed greater heterogeneity in phase 2. Comparison against HapMap populations showed predominant admixture between Caucasian and African ethnicities. Linear mixed models used in association analyses take account of genetic heterogeneity.

Genome-wide Association Study

Manhattan and Q-Q plots for genotyped data for phases 1/2 (Supplementary Figures 3 and 4) showed no systematic bias ($\lambda$s 0.998/1). A Manhattan plot for the combined imputed genotype data (Figure 1) shows no hits at $P<5\times10^{-8}$. Four approaches were used to identify susceptibility genes: (i) integrative post-GWAS annotation in FUMA [19]; (ii) analysis of transcriptional data comparing lesions with normal skin [24]; (iii) review of gene function for relevance to parasite biology/immunopathology; and (iv) analysis of genotypic differences in T cell responses.
Integrative Post-GWAS Mapping and Annotation in FUMA

SNP2GENE identified 32 genomic loci associated with CL (Table 1; Supplementary Table 2). Positionally mapped SNVs localized to non-coding sequence, 58% intronic, 21% intergenic, 7% intronic in noncoding RNA genes, and 4% other. Most genomic loci (29/32) had a single lead SNV, with 5 additional independent significant SNVs at loci 9, 13, and 14. Top GWAS hits (=lead SNVs) at 23 loci were taken forward (Table 1): 18 at/near protein coding genes (21 genes: 15 intronic; 6 upstream/downstream), 5 intronic in noncoding RNA genes, and 1 intergenic <5kb. Nine lead SNVs intergenic at >5 kb from the nearest gene were excluded from further consideration.

GWAS SNVs are generally enriched for eQTLs [25]. Focussing on data from tissues (whole blood, skin) and cell types (immune cells) relevant to CL (Table 1), SNP2GENE mapped eQTLs associated with expression of MCCC1/LAMP3, PCMTD1, KRT80, IFNG-AS1, DYM, SERPINB10, S100B, and MAPK8IP2.

Transcriptional Analyses of Putative Susceptibility Genes

Additional evidence (Table 1) to support genes as candidates was sought by comparing expression in CL lesions versus normal skin [24]. Six genes were expressed at higher level in lesions (LAMP3, STX7, CALCR, CRLF3, PPP6R1, CHKB), 5 genes at lower levels (ZNF385D, MCCC1, PCMTD1, TSPAN9, KRT80). For 5 differentially expressed genes, ZNF385D, STX7, CALCR, TSPAN9, PPP6R1, SNP2GENE eQTL mapping provided no evidence for SNVs associated with expression in selected tissues (whole blood, skin) or cell types (fibroblasts, immune cells). A role for GWAS SNVs regulating expression of KRT80 was supported by eQTL and chromatin interaction mapping (Figure 2). The lead SNV and others in strong LD lie upstream of KRT80 in a region of strong transcriptional/enhancer activity and act as eQTLs in cultured fibroblasts, sun and non-sun exposed skin. Other genes
supported by both eQTL and lesion expression were in tandem with genes positionally mapped to the same lead SNV (Table 1), namely LAMP3/MCCC1, PXDNL/PCMTD1 and CHKB/MAPK8IP2. For PXDNL/PCMTD1 the lead SNV was intronic in PXDNL but neither it nor numerous mapped SNVs in LD with it were eQTL for PXDNL itself. Rather, they acted as eQTLs for PCMTD1 expression in fibroblasts (Supplementary Figure 5). For CHKB/MAPK8IP2, the lead SNV lies upstream of both genes transcribed in opposing directions. Mapped SNV act as eQTLs for MAPK8IP2 in sun and not-sun exposed skin but not for CHKB (Supplementary Figure 6). For LAMP3/MCCC1 the lead SNV and SNVs in LD with it map predominantly within LAMP3 (Supplementary Figure 7). While they act as cis-eQTLs for MCCC1 expression, data from CL lesions (Table 1) suggests stronger upregulation of LAMP3 compared to downregulation of MCCC1.

Ten genes showed no differential expression in lesions (Table 1). This included SERPINB10 across which multiple cis-eQTLs mapped to chromatin states of transcriptional/enhancer activity in neutrophils, monocytes, B cells and haematopoietic stem cells (Supplementary Figure 8). Four non-coding RNA genes (Table 1) not present on the chips used for CL lesion data [24] included IFNG-AS1 which had 10 eQTLs across a chromatin state region of transcriptional/enhancer activity in immune cells (whole blood: T cells, B cells, haematopoietic stem cells) that were associated with expression of IFNG-AS1, IFNG and IL26 (Figure 3).

Relevance to the Biology and Immunopathology of CL disease

In summary, of 32 positional mapped genomic loci, 23 occurred at protein coding or non-coding RNA genes of which 15 had eQTLs for expression in relevant cells/tissues and/or showed differential expression in CL lesions. To determine which genes in these 15 loci might act as CL susceptibility genes we reviewed gene function in relation to parasite biology
and CL immunopathology (Supplementary Table S3). A plausible functional role for 12 genes was not found, including $PCMTD1$ for which eQTL and lesion expression data were strong. A role for these genes cannot be discounted, but 6 genes had plausible links to CL pathogenesis (Table 2, Figure 4): $LAMP3$ and $STX7$ play a role in lysosome function; $KRT80$ and $CRLF3$ relate to skin perturbations; $SERPINB10$ and $IFNG-AS1$ play central roles in immune responses.

**Relating IFNG-AS1 genotypes to Antigen-Specific T cell Responses**

The GWAS (Supplementary Information) showed modest support ($P<0.01$) for some previous candidate genes, but not for $IFNG$ variants associated with $L. guyanensis$ CL in Brazil [16] including no association with plasma IFN$\gamma$ (Figure 5A/5B). IFNG-AS1 expression influences IFN$\gamma$ production [26]. While eQTL mapping supported SNVs at $IFNG-AS1$ acting as cis-eQTLs for $IFNG$ and $IL26$, they were stronger eQTLs for $IFNG-AS1$ (Figure 3). Plasma IFN$\gamma$ did not differ by $IFNG-AS1$ genotype (Figure 5C), but a significant difference in the percentage of antigen-specific IFN$\gamma$ producing T cells across genotypes was observed at rs4913269 (ANOVA $P_{\text{adjusted}}=0.044$) (Figure 5D). Individuals homozygous for the disease-associated G allele showed a significantly lower percentage of IFN$\gamma$ producing T cells compared to heterozygotes. Similar genotype associations occurred for TNF producing T cells (Figure 5E; ANOVA $P_{\text{adjusted}}=0.021$), which were strongly correlated with IFN$\gamma$ producing T cells (Figure 5F; $r^2=0.31$, $P=0.0003$). Parallel observations were made for 6 other SNVs in strong LD (see Figure 4D) with rs4913269.
DISCUSSION

Our GWAS provides the first hypothesis-free insights into genetic risk factors for *L. braziliensis* CL. Despite prior evidence for genetic regulation of leishmaniasis [7], and the robust well-powered study undertaken, no signals of association achieved genome-wide significance (P<5x10^{-8}) and only modest support was found for previous candidate gene studies (Supplementary Information). We therefore employed integrative approaches [19] to prioritise six genes as plausible genetic risk loci for CL.

Two genes relate to intracellular localization of *Leishmania* parasite in phagolysosomes [27]. *LAMP3* encodes lysosomal associated membrane protein 3, also known as dendritic cell LAMP or DCLAMP [28]. Expression of DCLAMP increases in activated dendritic cells, localizing to the MHC Class II compartment immediately before translocation of Class II to the cell surface [28]. LAMP3 was expressed at 5.9-fold higher levels in lesions compared to normal skin, supporting its role in CL. Since dendritic cells are the most potent antigen-presenting cells that induce primary T-cell responses, it is likely that variation at *LAMP3* will relate to presentation of *Leishmania* antigens to T cells. *STX7* encodes syntaxin 7 which influences vesicle trafficking to lysosomes, including phagosome-lysosome fusion [29]. Variants at *STX7* could influence delivery of *Leishmania* phagosomes to lysosomes of macrophages.

Two other susceptibility genes, *KRT80* and *CRLF3*, likely relate to skin perturbations and/or the wound healing response. Keratin 80 is a type II epithelial keratin with biased expression in skin keratinocytes [30]. Pathogens invading skin cause keratinocytes to produce chemokines which attract monocytes, natural killer cells, T cells, and dendritic cells [31]. *KRT80* and multiple other keratins (*KRT77/81/4/39/32/33B*) were expressed at lower levels in lesions compared to normal skin. Whether this reflects a paucity of keratinocytes in lesions, or specific down regulation of gene expression in keratinocytes within lesions,
requires further investigation. Keratinocytes play a role in wound healing, are potent producers of IL-10 and TGFβ [32], and can change to a sclerotic phenotype by gene silencing of Fli1 [33]. Although association of human CL and FLI1 [9, 10] was not replicated here, Fli1 is a confirmed murine CL susceptibility gene [34]. Our novel associations continue to focus on molecules/cells involved in wound healing. In contrast, the cytokine receptor-like factor 3 CRLF3 is expressed in normal skin, and shows pathologically enhanced expression in premalignant actinic keratosis and malignant squamous cell carcinoma [35]. CRLF3 appears to be similarly dysregulated in CL lesions.

SERPINB10 and IFNG-AS1 are highlighted as central regulators of immune responses. Serpin family B member 10 is a peptidase inhibitor expressed in bone marrow [36] in the monocytic lineage and can inhibit TNF-induced apoptosis [37]. Epithelial SERPINB10 contributes to allergic eosinophilic inflammation [38]. However, despite eQTL data showing SERPINB10 expressed in sun-exposed skin, we found no evidence for its expression in lesions, consistent with more central roles in immune regulation. IFNG antisense RNA 1 fine-tunes the magnitude of IFNγ responses [26]. It is expressed in mouse and human T helper1 cells and positively regulates Ifng expression [39, 40]. Transient over-expression of Ifng-as1 is associated with increased IFNγ and reduced susceptibility to Salmonella enterica [39]. Conversely, deletion of Ifng-as1 in mice compromises host defence against Toxoplasma gondii by reducing Ifng expression. Discordant expression of IFNG and IFNG-AS1 is seen in long-lasting memory T cells, where high IFNG-AS1 associated with low IFNG suggests feedback inhibition [26]. We observed that IFNG-AS1 genotype was associated with downstream effects on percentage of IFNγ-producing CD3+ T cells and highly correlated TNF-producing CD3+ T cells following antigen stimulation. Individuals homozygous for the disease-associated allele at 7 IFNG-AS1 associated SNVs had significantly lower percentages of IFNγ/TNF T cells, suggesting that lower IFNγ and TNF
regulated by *IFNG-AS1* causes increased disease risk. These two pro-inflammatory cytokines are important activators of macrophages for anti-leishmanial activity.

Our GWAS identified novel genetic risk factors for CL that provide interesting leads to further understanding CL pathogenesis, including through regulation of IFNγ responses.
Notes

Acknowledgements. We would like to thank the staff of the Health Post at Corte de Pedra for their assistance in the collection of samples and clinical and field data, as well as staff at the HEMOBA Foundation Blood Bank in Salvador.

Disclaimer. The study sponsor had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all study data and had final responsibility for the decision to submit for publication.

Financial support. This work was supported by: the British Medical Research Council (MRC) grant number MR/N017390/1; a Brazilian FAPEMIG grant in cooperation with MRC/CONFAP (CBB-APQ-00883-16), National Institute of Science and Technology in Tropical Diseases, Brazil (N° 573839/2008–5), CNPq (K.J.G. and W.O.D. are CNPq fellows), FAPESP (Fellowships for N.S.A. and A.B.F.); the National Institute of Science and Technology in Tropical Diseases, Brazil (N° 573839/2008–5); and the National Institute of Health NIH Grant AI 30639. HJC and SC were supported by the Wellcome Trust (grant number 102858/Z/13/Z).

Potential conflicts of interest. No authors reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.
References

1. Follador I, Araujo C, Bacellar O, et al. Epidemiologic and immunologic findings for the subclinical form of Leishmania braziliensis infection. Clin Infect Dis 2002; 34(11): E54-E8.
2. Muniz AC, Bacellar O, Lago EL, et al. Immunologic Markers of Protection in Leishmania (Viannia) braziliensis Infection: A 5-Year Cohort Study. J Infect Dis 2016; 214(4): 570-6.
3. Gomes-Silva A, de Cassia Bittar R, Dos Santos Nogueira R, et al. Can interferon-gamma and interleukin-10 balance be associated with severity of human Leishmania (Viannia) braziliensis infection? ClinExpImmunol 2007; 149(3): 440-4.
4. Antonelli LR, Dutra WO, Almeida RP, Bacellar O, Carvalho EM, Gollob KJ. Activated inflammatory T cells correlate with lesion size in human cutaneous leishmaniasis. Immunol Lett 2005; 101(2): 226-30.
5. Oliveira WN, Ribeiro LE, Schrieffer A, Machado P, Carvalho EM, Bacellar O. The role of inflammatory and anti-inflammatory cytokines in the pathogenesis of human tegumentary leishmaniasis. Cytokine 2014; 66(2): 127-32.
6. Blackwell JM, Fakiola M, Castellucci LC. Human genetics of leishmania infections. Hum Genet 2020.
7. Shaw MA, Davies CR, Llanos-Cuentas EA, Collins A. Human genetic susceptibility and infection with Leishmania peruviana. AmHumGenet 1995; 57: 1159-68.
8. Cabrera M, Shaw M-A, Sharples C, et al. Polymorphism in TNF genes associated with mucocutaneous leishmaniasis. JExpMed 1995; 182: 1259-64.
9. Castellucci L, Jamieson SE, Almeida L, et al. Wound healing genes and susceptibility to cutaneous leishmaniasis in Brazil. Infect Genet Evol 2012; 12(5): 1102-10.
10. Castellucci L, Jamieson SE, Miller EN, et al. FL1 polymorphism affects susceptibility to cutaneous leishmaniasis in Brazil. Genes Immun 2011; 12(7): 589-94.
11. Castellucci L, Jamieson SE, Miller EN, et al. CXCR1 and SLC11A1 polymorphisms affect susceptibility to cutaneous leishmaniasis in Brazil: a case-control and family-based study. BMC Med Genet 2010; 11: 10.
12. Castellucci L, Menezes E, Oliveira J, et al. IL6 -174 G/C promoter polymorphism influences susceptibility to mucosal but not localized cutaneous leishmaniasis in Brazil. J Infect Dis 2006; 194(4): 519-27.
13. Salhi A, Rodrigues V, Jr., Santoro F, et al. Immunological and genetic evidence for a crucial role of IL-10 in cutaneous lesions in humans infected with Leishmania braziliensis. J Immunol 2008; 180(9): 6139-48.
14. Ramasawmy R, Menezes E, Magalhaes A, et al. The -2518bp promoter polymorphism at CCL2/MCP1 influences susceptibility to mucosal but not localized cutaneous leishmaniasis in Brazil. Infect Genet Evol 2010; 10(5): 607-13.
15. Almeida L, Oliveira J, Guimaraes LH, Carvalho EM, Blackwell JM, Castellucci L. Wound healing genes and susceptibility to cutaneous leishmaniasis in Brazil: role of COL1A1. Infect Genet Evol 2015; 30: 225-9.
16. da Silva GAV, Mesquita TG, Souza VC, et al. A Single Haplotype of IFNG Correlating With Low Circulating Levels of Interferon-gamma Is Associated With Susceptibility to Cutaneous Leishmaniasis Caused by Leishmania guyanensis. Clin Infect Dis 2019.
17. D’Oliveira A, Jr., Machado P, Bacellar O, Cheng LH, Almeida RP, Carvalho EM. Evaluation of IFN-gamma and TNF-alpha as immunological markers of clinical outcome in cutaneous leishmaniasis. Rev Soc Bras Med Trop 2002; 35(1): 7-10.
18. Faria DR, Gollob KJ, Barbosa JJ, et al. Decreased in situ expression of interleukin-10 receptor is correlated with the exacerbated inflammatory and cytotoxic responses observed in mucosal leishmaniasis. Infect Immum 2005; 73: 7853-9.
19. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017; 8(1): 1826.
20. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. Eur J Hum Genet 2016; 24(8): 1202-5.
21. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. Nat Genet 2016; 48(10): 1284-7.
22. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. Nat Methods 2011; 8(10): 833-5.
23. Prum RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010; 26(18): 2336-7.
24. Novais FO, Carvalho LP, Passos S, et al. Genomic profiling of human Leishmania braziliensis lesions identifies transcriptional modules associated with cutaneous immunopathology. J Invest Dermatol 2015; 135(1): 94-101.
25. Croteau-Chonka DC, Rogers AJ, Raj T, et al. Expression Quantitative Trait Loci Information Improves Predictive Modeling of Disease Relevance of Non-Coding Genetic Variation. PLoS ONE 2015; 10(10): e0140758.
26. Petermann F, Pekowska A, Johnson CA, et al. The Magnitude of IFN-gamma Responses Is Fine-Tuned by DNA Architecture and the Non-coding Transcript of Ifng-as1. Molecular cell 2019; 75(6): 1229-42 e5.
27. Alexander J. Leishmania mexicana: inhibition and stimulation of phagosome-lysosome fusion in infected macrophages. Exp Parasitol 1981; 52(2): 261-70.
28. de Saint-Vis B, Vincent J, Vandenabeele S, et al. A novel lysosome-associated membrane glycoprotein, DC-LAMP, induced upon DC maturation, is transiently expressed in MHC class II compartment. Immunity 1998; 9(3): 325-36.
29. Wang H, Frelin L, Pevsner J. Human syntaxin 7: a Pep12p/Vps6p homologue implicated in vesicle trafficking to lysosomes. Gene 1997; 199(1-2): 39-48.
30. Fagerberg L, Hallstrom BM, Oksvold P, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics 2014; 13(2): 397-406.
31. Quaresma JAS. Organization of the Skin Immune System and Compartmentalized Immune Responses in Infectious Diseases. Clin Microbiol Rev 2019; 32(4).
32. Kim WH, An HJ, Kim JY, et al. Apamin inhibits TNF-alpha- and IFN-gamma-induced inflammatory cytokines and chemokines via suppressions of NF-kappaB signaling pathway and STAT in human keratinocytes. Pharmacol Rep 2017; 69(5): 1030-5.
33. Takahashi T, Asano Y, Sugawara K, et al. Epithelial Fli1 deficiency drives systemic autoimmunity and fibrosis: Possible roles in scleroderma. J Exp Med 2017; 214(4): 1129-51.
34. Sakthianandeswaren A, Curtis JM, Elso C, et al. Fine mapping of Leishmania major susceptibility Locus lmr2 and evidence of a role for Fli1 in disease and wound healing. Infect Immun 2010; 78(6): 2734-44.
35. Dang C, Gottschling M, Manning K, et al. Identification of dysregulated genes in cutaneous squamous cell carcinoma. Oncol Rep 2006; 16(3): 513-9.
36. Riewald M, Schleef RR. Molecular cloning of bomapin (protease inhibitor 10), a novel human serpin that is expressed specifically in the bone marrow. J Biol Chem 1995; 270(45): 26754-7.
37. Schleef RR, Chuang TL. Protease inhibitor 10 inhibits tumor necrosis factor alpha -induced cell death. Evidence for the formation of intracellular high M(r) protease inhibitor 10-containing complexes. J Biol Chem 2000; 275(34): 26385-9.
38. Mo Y, Zhang K, Feng Y, et al. Epithelial SERPINB10, a novel marker of airway eosinophilia in asthma, contributes to allergic airway inflammation. Am J Physiol Lung Cell Mol Physiol 2019; 316(1): L245-L54.
39. Gomez JA, Wapinski OL, Yang YW, et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. Cell 2013; 152(4): 743-54.

40. Collier SP, Collins PL, Williams CL, Boothby MR, Aune TM. Cutting edge: influence of Tmevpg1, a long intergenic noncoding RNA, on the expression of Ifng by Th1 cells. JImmunol 2012; 189(5): 2084-8.
Figure Legends

Figure 1. Manhattan plot of results from the combined analysis for the 4.46M high-quality 1000G imputed SNV variants common to Phase 1 and Phase 2 samples. Data are for analysis in FastLMM looking for association between SNVs and CL. The Y-axis indicates \(-\log_{10} P\) values for association, the X axis indicates the positions across each chromosome. The red dotted line indicates the \(P=5\times10^{-5}\) cut-off used to look for suggestive associations.

Figure 2. Results of positional, chromatin interaction, and eQTL activity mapping in FUMA for KRT80. (A) Maps the top lead SNV, and SNVs in LD with it according to the \(r^2\) colour-coded key, across the two genes. There were no additional independent significant SNV. (B) Chromatin-15 states colour coded for transcriptional/enhancer activity as shown in the key. Y-axis colour coding relates to cell/tissue types in which chromatin interaction was mapped. (C) eQTL activity for genes (Y-axis) in different cells/tissues from public domain databases as shown in the key. Full explanation of keys provided as preamble to supplementary figures.

Figure 3. Results of positional, chromatin interaction, and eQTL activity mapping in FUMA for IFNG-AS1. (A) Maps the top lead SNV, and SNVs in LD with it according to the \(r^2\) colour-coded key, across the two genes. There were no additional independent significant SNV. (B) Chromatin-15 states colour coded for transcriptional/enhancer activity as shown in the key. Y-axis colour coding relates to cell/tissue types in which chromatin interaction was mapped. (C) eQTL activity for genes (Y-axis) in different cells/tissues from public domain databases as shown in the key. Full explanation of keys provided as preamble to supplementary figures.
Figure 4. LocusZoom plots for GWAS associations identified as plausible genetic risk factors for CL following post-GWAS annotation: (A) LAMP3; (B) STX7; (C) KRT80; (D) CRLF3, (E) SERPINB10 and (F) IFNG-AS1. The $-\log_{10} P$ values (left y-axis) are shown in the top section of each plot. Dots representing individual SNVs are color coded (see key) based on their population-specific LD $r^2$ with the top SNV (annotated by rs ID) in the region. The right Y-axis is for recombination rate (blue line), based on HapMap data. The bottom section of each plot shows the positions of genes across the region.

Figure 5. Plots examining IFNG and IFNG-AS1 genotypes by IFN-γ and TNF responses. Plots (A), (B) and (C) show results for plasma levels of IFN-γ. (A) and (B) show that there is no association between plasma IFN-γ and genotypes for two SNVs at IFNG, rs1861494 that was associated with CL disease for L. guyanensis in a previous study [16] and rs2080414 that was in the strongest LD with rs2069705 that was associated with CL disease and plasma IFN-γ in that study (rs2069705 was not genotyped or imputed in the present study). (C) shows that there is no association between plasma IFN-γ and rs4913269 at IFNG-AS1. Plots (D) and (E) show differences in percentages of antigen-stimulated IFN-γ and TNF producing CD3+ T cells by IFNG-AS1 genotype for the top SNV rs4913269 at Chromosome 12 bp position 68407845 associated with CL disease in our study. (F) shows the correlation between percent IFN-γ+ and percent TNF+ CD3+ T cells for individuals genotyped.
Table 1. Summary of SNP2GENE Results for Lead GWAS SNVs and Associated Gene Information

| Genomic Location | Lead IndSig | rsID     | Nearest Gene | Type of Gene | Distance from Gene | Functional Location | N Pos | N eQT LSN | eQTL Database | eQTL Type | Lesion on vs Normal | Fold Change |
|------------------|-------------|----------|--------------|--------------|--------------------|---------------------|-------|-----------|---------------|-----------|----------------------|-------------|
| 1:175 28080 6    | rs127: 53656 | RP3-518E13.2: TNR | antisen: protein coding | 0 | ncRNA intronic intronic | 6 | 0 | (TNR) NS | ND | |
| 1:238 42720 3    | rs139: 14427 | RP11-136B18.1 | lincRNA | 459 | intergenic | 1 | 0 | NS | |
| 2:507 06764      | none        | NRXN1    | protein coding | 0 | intronic intronic | 40 | 0 | ND | |
| 3:221 17736      | rs139: 3086 | ZNF385 D | protein coding | 0 | intronic intronic | 3 | 0 | 6.28 E-04 | -1.8 | |
| 3:149 31426 7    | rs536: 03459 | WWTR1    | protein coding | 0 | intronic intronic | 11 | 0 | NS | |
| 3:182 85726 1    | s7428: 5558 | MCCC1    | protein coding | 231 | upstream | 4 | 12 | eQTLGen GTEx/v8 GTEx/v8 BIOSQTL | cis_e QTLs Skin SE Skin NSE Gene -level | 7.98 E-10 | -2.1 | |
| 3:182 85726 1    | s7428: 5558 | LAMP3    | protein coding | 0 | intronic | 14 | 1 | eQTLGen | cis_e QTLs | 9.25 E-12 | 5.9 | |
| 6:132 81556 4    | rs144: 48813 | STX7     | protein coding | 0 | intronic | 6 | 0 | ND | 0.00 E-08 | 1.3 | |
| 7:930 65079 8    | rs143: 58696 | CALCR    | protein coding | 0 | intronic | 8 | 0 | ND | 3.62 E-04 | 1.7 | |
| 8:402 45200      | rs125: 676 | CTA-392C11.2 | lincRNA | 0 | ncRNA intronic | 47 | 0 | ND | |
| 8:526 28820      | rs132: 61618 | PXDNL    | protein coding | 0 | intronic | 62 | 0 | NS | |
| 8:526 28820      | rs132: 61618 | PCMTD    | protein coding | 0 | downstream | 37 | 70 | eQTLGenGT Ex/v8 | cis_e QTLs Fibroblasts | 6.20 E-09 | -1.9 | |
| 11:80            | none        | RP11-     | lincRNA | 0 | ncRNA | 6 | 0 | ND | |

Downloaded from https://academic.oup.com/cid/advance-article/doi/10.1093/cid/ciaa1230/5896307 by guest on 25 August 2020
| rsID     | chr | Gene Symbol | Protein Coding | Intronic Location | Downstream/Upstream | GTEx/v8; GTEx/v8; GTEx/v8 | eQTLGenGT Ex/v8 BIOSQTL | cis-eQTLs Blood Gene-level |
|----------|-----|-------------|----------------|-------------------|---------------------|--------------------------|--------------------------|----------------------------|
| rs775    | 47010 | TSPAN9      | Protein coding | 167               | Downstream          | 3                        | 0                        | 2.54 E-05                  |
| rs107    | 59000 | KRT80       | Protein coding | 421               | Upstream            | 32                       | 31                       | 3.07 E-08                  |
| rs491    | 40784 | IFNG-AS1    | Antisense      | 0                 | Intronic            | 10                       | 10                       | ND                         |
| rs657    | 56088 | MDGA2:MDGA2 | Protein coding | 0                 | Intronic            | 4                        | 0                        | NS                         |
| rs125    | 68604 | AL1639      | lincRNA        | 0                 | Intronic            | 52                       | 0                        | ND                         |
| rs752    | 13612 | CRLF3       | Protein coding | 0                 | Intronic            | 8                        | 0                        | 9.19 E-09                  |
| rs493    | 76615 | DYM         | Protein coding | 0                 | Intronic            | 414                      | 390                      | NS                         |
| rs809    | 95567 | TCF4        | Protein coding | 0                 | Intronic            | 22                       | 6                        | NS                         |
| rs808    | 59876 | SERPINB10  | Protein coding | 0                 | Intronic            | 31                       | 28                       | NS                         |
| rs127    | 74688 | PPP6R1      | Protein coding | 0                 | Intronic            | 19                       | 0                        | 6.09 E-07                  |
| rs201    | 02364 | S100B       | Protein coding | 0                 | Intronic            | 14                       | 34                       | NS                         |
| rs112    | 03882 | CHK1        | Protein coding | 0                 | Downstream          | 28                       | 0                        | 2.60 E-05                  |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 32 | 22:51 | rs11297444 | MAPK8 | protein coding | 0 upstream | 32 | 29 | GTEx/v8 | GTEx/v8 | Skin SE | NS |
| 03882 | 4 | 9 | IP2 | | | | | | | | |

**Note:** Full details of genomic loci are provided in Supplementary Table 2.

**Abbreviations:** a Lead IndSigSNP = GWAS top SNV; b Bold indicates pairs of genes mapped with respect to the same Lead IndSigSNP; c Pos = positionally mapped SNVs from SNP2GENE analysis; eQTL type/Tissue-Cell Type: Skin SE = Skin Sun Exposed Lower Leg; Skin NSE = Skin Not Sun Exposed Suprapubic; Blood = Whole blood; Fibroblasts = cultured fibroblasts; f Analysed from data in GEO database GSE55664 using the GEO2R tool with Benjamini and Hochberg false discovery rate adjusted P-values, ND = not done, NS = not significant; g Fold-change for GEO2R lesion versus normal skin analysis.
Table 2. Top GWAS Hits in Genes of Plausible Functional Interest as Genetic Risk Factors for CL Caused by *L. braziliensis*

| Chr | Position (bp) | rsID | P-value | Odds Ratio (95% CI) | Beta (SE) | Allele | Variant Origin | Location | Gene | Function |
|-----|---------------|------|---------|---------------------|-----------|--------|----------------|----------|------|----------|
| 3   | 18285726      | rs74285558 | 6.54E-06 | 0.87 (0.82 - 0.92) | -0.034 (0.008) | T (C/T) | Global intron | LAMP3     |       | Lysosome associated membrane protein 3 |
| 6   | 13281556      | rs14448813 | 6.10E-06 | 0.82 (0.75 - 0.89) | -0.034 (0.007) | A (C/A) | African intron | STX7      |       | Syntaxin 7 |
| 12  | 52590004      | rs10783496 | 6.58E-06 | 1.06 (1.03 - 1.09) | 0.035 (0.008) | A (G/A) | Global intron | KRT80     |       | Keratin 80 |
| 12  | 68407845      | rs4913269  | 1.32E-05 | 1.06 (1.03 - 1.08) | 0.033 (0.008) | G (C/G) | Global intron | IFNG-AS1  |       | IFNG antisense RNA 1 |
| 17  | 29136126      | rs75270613 | 5.12E-06 | 0.83 (0.77 - 0.90) | -0.034 (0.008) | T (C/T) | African intron | CRLF3     |       | Cytokine receptor like factor 3 |
| 18  | 61598763      | rs8084306  | 1.56E-06 | 1.07 (1.04 - 1.10) | 0.038 (0.008) | C (T/C) | Global intron | SERPINB10 |       | Serpin family B member 10 |

**NOTE** Details of all post-GWAS candidate genes are provided in Supplementary Table S3. 
1Associated allele (ancestral/minor) for risk or protection as indicated by the odds ratio.
Figure 3
Figure 5

A. ANOVA NS

rs1861494 IFNG genotype

TT  TC  CC

B. ANOVA NS

rs2080416 IFNG genotypes

TT  TA  AA

C. ANOVA NS

rs4913269 IFNG-AS1 genotypes

CC  CG  GG

D. ANOVA p=0.044

rs4913269 IFNG-AS1 genotypes

CC  GC  GG

E. ANOVA p=0.021

rs4913269 IFNG-AS1 genotypes

CC  GC  GG

F. ANOVA p=0.003

rs4913269 IFNG-AS1 genotypes

 IFNγ T cells

HIV-1 T cells