



**HAL**  
open science

## **A Genome-Wide Association Study Identifies Novel Susceptibility loci in Chronic Chagas Cardiomyopathy**

Desiré Casares-Marfil, Mariana Strauss, Pau Bosch-Nicolau, María Silvina Lo Presti, Israel Molina, Christophe Chevillard, Edecio Cunha-Neto, Ester Sabino, Antonio Luiz P Ribeiro, Clara Isabel González, et al.

► **To cite this version:**

Desiré Casares-Marfil, Mariana Strauss, Pau Bosch-Nicolau, María Silvina Lo Presti, Israel Molina, et al.. A Genome-Wide Association Study Identifies Novel Susceptibility loci in Chronic Chagas Cardiomyopathy. *Clinical Infectious Diseases*, 2021, 73 (4), pp.672-679. 10.1093/cid/ciab090 . hal-03249865

**HAL Id: hal-03249865**

**<https://hal.science/hal-03249865>**

Submitted on 14 Feb 2022

**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.

# **A genome-wide association study identifies novel susceptibility *loci* in chronic Chagas cardiomyopathy.**

Desiré Casares-Marfil (1), Mariana Strauss (2), Pau Bosch-Nicolau (3), María Silvana Lo Presti (2), Israel Molina (3), Christophe Chevillard (4), Edecio Cunha-Neto (5), Ester Sabino (6), Antonio L. Ribeiro (7), Clara Isabel González (8), Chagas Genetics CYTED Network, Javier Martín\* (1), Marialbert Acosta-Herrera\* (1)

1 Instituto de Parasitología y Biomedicina López-Neyra, CSIC, Granada, Spain

2 Centro de Estudios e Investigación de la Enfermedad de Chagas y Leishmaniasis, FCM, INICSA-CONICET-UNC, Córdoba, Argentina,

3 Unidad de Medicina Tropical y Salud Internacional Hospital Universitari Vall d'Hebron, PROSICS, Barcelona, Spain

4 INSERM, Aix Marseille university, TAGC, UMR\_1090, Marseille, France.

5 Laboratory of Immunology, Heart Institute (InCor)/Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil,

6 Instituto de Medicina Tropical Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil,

7 Department of Internal Medicine, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil,

8 Grupo de inmunología y epidemiología molecular, Escuela de Microbiología, Universidad Industrial de Santander, Bucaramanga, Colombia.

\* These authors contributed equally as senior authors.

\*Corresponding authors

\*Javier Martín, Professor: Institute of Parasitology and Biomedicine López-Neyra, CSIC, Av. del Conocimiento 17. Armilla (Granada), Spain 18016. Phone: (+34) 958181669; Fax: (+34) 958181632; email: [javiermartin@ipb.csic.es](mailto:javiermartin@ipb.csic.es)

\*Marialbert Acosta-Herrera, PhD: Institute of Parasitology and Biomedicine López-Neyra, CSIC, Av. del Conocimiento 17. Armilla (Granada), Spain 18016. Phone: (+34) 958181621; Fax: (+34) 958181632; email: [m.acostaherrera@ipb.csic.es](mailto:m.acostaherrera@ipb.csic.es)

**Running title:** GWAS in chronic Chagas cardiomyopathy

**Summary:** A genome-wide association study of 3,413 samples from different Latin American countries identifies a significant genetic *loci* associated with the differential development of chronic Chagas cardiomyopathy.

**Keywords:** Chagas disease, cardiomyopathy, genetic association, GWAS, Latin American, meta-analysis, *Trypanosoma cruzi*

## **ABSTRACT**

**Background:** Chagas disease is an infectious disease caused by the parasite *Trypanosoma cruzi* and is endemic from Latin American countries. The goal of our study was to identify novel genetic *loci* associated with chronic Chagas cardiomyopathy development in Chagas disease patients from different Latin American populations.

**Methods:** We performed a cross-sectional, nested case-control study including three sample collections from Colombia, Argentina and Bolivia. Samples were genotyped to conduct a genome-wide association study (GWAS). These results were meta-analyzed with summary statistic data from Brazil, gathering a total of 3,413 Chagas disease patients. To identify the functional impact of the associated variant and its proxies we performed an *in silico* analysis of this region.

**Results:** The meta-analysis revealed a novel genome-wide statistically significant association with chronic Chagas cardiomyopathy development in rs2458298 (OR=0.90, 95%CI=0.87-0.94,  $p\text{-value}=3.27\times 10^{-08}$ ), nearby the *SAC3D1* gene. In addition, further *in silico* analyses displayed functional relationships between the associated variant and the *SNX15*, *BAFT2* and *FERMT3* genes, related to cardiovascular traits.

**Conclusions:** Our findings support the role of the host genetic factors in the susceptibility to the development of the chronic cardiac form of this neglected disease.

## INTRODUCTION

Chagas disease is an infectious disease caused by the parasitic protozoan *Trypanosoma cruzi* [1] that according to The World Health Organization (WHO) affects about 6-7 million people and its endemic areas are spread in 21 Latin American countries ([www.who.int/health-topics/chagas-disease#tab=tab\\_1](http://www.who.int/health-topics/chagas-disease#tab=tab_1)). Chagas disease is considered as an emerging infection as cases have been reported out of the endemic areas of transmission, mainly in United States and Europe, being this consequence of migration and globalization processes [1]. The disease comprises an acute phase, which occurs after the entry of the parasite into the host, and a chronic phase, which can occur even decades after the first contact with the infectious agent; however, approximately 30% of patients evolve to this chronic phase [1]. The hypothesis of the implication of host genetic component in the differential susceptibility to Chagas's disease arises from the high exposure to *T. cruzi* and the differential susceptibility to the infection in endemic areas, in addition to the remarkable variation in the disease progression during the chronic phase [1, 2]. In spite of the intensive investigation of the genetic basis of Chagas' disease using the candidate gene strategy, its understanding remains elusive. These studies have analyzed specific genetic polymorphisms mainly located in genes involved in the immune response such as chemokines, cytokines and the major histocompatibility complex; however there are recognized important limitations to identify solid genetic associations [3, 4]. Some of these limitations include reduced sample sizes and the assessment of a limited number of variants, usually based on plausible biological pathways involved in the trait or disease, precluding replication [4]. In this sense, genome-wide association studies (GWAS) have contributed to the assessment of thousands of genetic variants associated with the susceptibility to

complex traits, taking a step forward in the understanding of its genetic basis [5]. Regarding Chagas disease, only one GWAS has been performed in Brazilian population in chronic Chagas cardiomyopathy showing suggestive associations [6].

In order to better elucidate the genetic basis of Chagas disease and chronic Chagas cardiomyopathy, herein we perform a large GWAS and a meta-analysis of Latin American populations that, in combination with complementary *in-silico* functional evidence, would provide further insights into the pathogenesis of this neglected disease.

## **MATERIALS AND METHODS**

### **Study populations, sample collection and ethical considerations**

Samples from three different Latin American countries: Colombia, Bolivia and Argentina were included in this study and meta-analyzed with data from the previous GWAS in a Brazilian population, comprising a total of 3,699 genomic DNA recruited samples. The description of sample origin is available elsewhere [7]. Chagas disease cases were defined by their serological status using recombinant antigen and a commercial indirect hemagglutination test for ELISA assays. Participants were classified in seropositive and seronegative for *T. cruzi* antigens and comparisons were performed among them in order to assess susceptibility to the infection in the Colombian and Argentinian cohort. The Bolivian and Brazilian cohorts were only composed by seropositive individuals. In addition, all the seropositive patients recruited in this study were subjected to electrocardiogram, echocardiogram, and chest radiography to identify cardiac involvement. For the assessment of the chronic Chagas cardiomyopathy susceptibility, those patients with cardiac abnormalities were compared

to asymptomatic individuals. Data included from the Brazilian cohort was described elsewhere [6]. Sample size and demographic information are summarized in Table 1.

The study was approved by the Ethics Committees from the Industrial University of Santander and Cardiovascular Foundation, Colombia (Act No. 15/2005), the Vall D'Hebron University Hospital, Barcelona, Spain (PR (AMI) 297/2016), and the National Hospital of Clinics, National University of Cordoba, Argentina (CIEIS HNC 118/2012 and 2/16/2017). The Ethics Committees for the Brazilian cohort was described elsewhere [6]. Protocols used in the study followed the principles of the Declaration of Helsinki and all individuals included in the study signed written informed consents.

### **Genotyping, quality control and imputation**

Genomic DNA from blood samples from the Colombia, Bolivian and Argentinian cohorts were isolated using the *QIAamp MidiDNA* Kit following manufacturer's recommendations. The samples were genotyped using the Global Screening Array Platform (Illumina Inc., San Diego, CA, USA). Quality controls (QCs) of genotyped data and consistency of sex assignment were performed using Plink v.1.9 [8]. Individuals and variants with missing genotype rate <5% were excluded. Single nucleotide polymorphisms (SNPs) with different call rates between cases and controls ( $p$ -value < 0.05), or deviating from Hardy-Weinberg equilibrium (HWE) with a  $p$ -value  $\leq 1 \times 10^{-6}$  in the control group were removed.

A principal component analysis (PCA) was performed using Plink1.9 to compare our data with the data from “The 1000 Genomes Project” (1KGP). The

genomic inflation factor ( $\lambda$ ) was calculated to control the type I error rate with the package “gap” for R version 3.6.1.

After QCs, genotyped data were imputed with the Michigan Imputation Server using the admixed American population from 1KGP phase 3 as reference panel [9]. After imputation, variants were filtered out by their minor allele frequency (MAF<1%) and the imputation quality metric  $R_{sq}<0.3$ .

### **Statistical analysis**

For each case-control collection, association testing for imputed allele dosages was performed using a mixed model with EMMAX [10] implemented in EPACTS (<https://genome.sph.umich.edu/wiki/EPACTS>). Given the common environmental exposure to triatomine vectors of individuals living together in endemic areas, related individuals were maintained in the analysis. In this sense, genotyped data were used to calculate the pairwise identity by descent (IBD) proportion (PI\_HAT), duplicated individuals were removed (PI\_HAT>0.8) and related individuals were considered with a PI\_HAT $\geq$ 0.4. EMMAX calculates the kinship matrix to include it as a covariate in the statistical analysis, as well as sex and age. The summary-level statistics of each cohort, including the Brazilian cohort, were meta-analyzed using METASOFT [11] under a fixed and random-effect models based on variants' heterogeneity (Cochran's Q test *p-value*). As most of imputed variants (>95%) did not show heterogeneity (Cochran's Q test *p-value*>0.05), those analyzed under a random-effect model were omitted from further analysis. Conservative suggestive associations were established at a *p-value*<1x10<sup>-06</sup> and genome-wide significance at *p-value*<5x10<sup>-08</sup> as usually employed for GWAS. The SNPnexus software (<https://www.snp-nexus.org/v4/>) was used to annotate



the suggestive and significant signals, and LocusZoom (<http://locuszoom.org/>) for the regional association plots.

### **Functional annotation of associated variants**

An *in-silico* approach was used to investigate the potential biological consequences of the associated variant. The LDlink tool [12] was used to calculate proxies in high linkage disequilibrium (LD) with the top signal ( $r^2 > 0.8$ ) and different databases were queried. The Open Targets Genetic was used to assess their functional implication (<https://genetics.opentargets.org/>), and complemented with PheWeb version 1.1.17 (<http://pheweb.sph.umich.edu/SAIGE-UKB/>) to assess these variants in previous Phenome-Wide Association Studies (PheWAS). We used HaploReg v4.1 [13] to evaluate regulatory genomic regions and Capture HiC Plotter [14] to assess long-distance physical interactions. RegulomeDB (<https://regulomedb.org/regulome-search/>) and GTEx (<https://gtexportal.org/home/>) were used to evaluate expression-quantitative trait *loci* (eQTLs) in relevant tissues.

## **RESULTS**

A total of 3,413 postQCs samples from four independent Latin American cohorts from Colombia, Bolivia, Argentina and Brazil, were included in this study. The Supplementary Table 1 summarizes the pre and post QC information within each dataset. After QCs and imputation, 7,846,902 SNPs for the Colombian and 8,408,292 SNPs for the Argentinian cohorts were analyzed correcting by age, sex and kinship in the case of susceptibility to the infection by *T. cruzi* analysis, while in the case of chronic Chagas cardiomyopathy development, a total of 8,218,190 SNPs were tested in

the meta-analysis including the Colombia, Bolivian, Argentinian and Brazilian populations.

### **Population structure**

Given the ethnic admixture of Latin American populations, the first two principal components were used to evaluate the genetic heterogeneity existing in the samples included in this study and in comparison with the admixed American subpopulation from 1KGP (Supplementary Figure 1). The results of these analyses showed that the heterogeneity present in our samples is consistent with the observed in the admixed American subpopulation of 1KGP. Despite of this, the Argentinian and Bolivian sample collections seemed to be more homogenous, while a considerable genetic variability with the Brazilian cohort was detected, reinforcing the high degree of genetic admixture of this population. In the association analyses within each cohort, the effect of population structure was taken into account by using a mixed-model of association [10].

### **Differential susceptibility to *Trypanosoma cruzi* infection**

Individuals from the Colombian and Argentinian cohorts were classified and compared according to their serological status in seropositive and seronegative for *T. cruzi* infection. No statistically relevant results were identified in the association test in the Argentinian population, mainly due to the reduced sample size of this cohort. However, in the case of the Colombian cohort we detected seven suggestive associations ( $p\text{-value} < 1 \times 10^{-6}$ ) contained in four different *loci* due to the LD (Table 2, Supplementary figure 2). The strongest association is an intronic variant of the *EBF2* gene (rs147475322;  $p\text{-value} = 2.15 \times 10^{-07}$ , Odds Ratio [OR]=1.20, 95% Confidence

Interval [95%CI] 0.76-1.90). The other three SNPs are located in an intronic region of *CD247* gene (rs554994388) and map downstream and upstream of *SIK1* (rs229347) and *IL18* (rs4937075) genes, respectively (Table 2). Remarkably, a different SNP within the *IL18* gene was previously associated in a candidate gene assessment [7]; however, this variant was nominally associated in our study and did not reach the suggestive level of significance.

### **Differential susceptibility to the development of Chronic Chagas Cardiomyopathy**

In order to identify novel genetic variants associated with differential chronic Chagas cardiomyopathy development seropositive samples were stratified according to their cardiological status, and patients with cardiac abnormalities were compared to asymptomatic individuals in the association test.

After QCs, genotype imputation and filtering, a case-control association analysis from GWAS data was performed in the Colombian, Bolivian and Argentinian samples comprising a total of 2,309 individuals (981 chronic cardiac patients and 1,328 asymptomatic individuals). After correcting by age, sex and kinship, deviations in genomic inflation factors were not observed (Supplementary Table 1, supplementary figure 3). Next, summary statistics from every cohort, including the Brazilian data, was meta-analyzed by an inverse variance-weighted effect size and those shared by at least two data sets were considered. Additionally, we did not observe inflation in the meta-analysis ( $\lambda=1.03$ ). As result, one statistically significant associated signal was identified at GWAS-level ( $p\text{-value}<5\times 10^{-08}$ ) and 22 suggestive associations ( $p\text{-value}<1\times 10^{-06}$ ) in three different autosomal *loci* (Table 3; Figure 1). The strongest association (rs2458298;  $p\text{-value}=3.27\times 10^{-08}$ , OR=0.90, 95% CI 0.87-0.94), is located in chromosome 11 in an intronic region of the *NAALADLI* gene. This signal is followed by several proxy

variants in high or moderate LD ( $r^2 > 0.4$ ) located nearby *NAALADLI*, *SAC3DI* and *SNX15* genes (Figure 2). Regarding the rest of the suggestive signals, they are located in intergenic regions closed to *CDH8* and *KLF4* genes (Table 3).

### Functional analysis

To explore functional features of the genomic association obtained from the meta-analysis in chronic Chagas cardiomyopathy, several databases were queried. Regarding the top associated variant rs2458298, although is located in an intronic region of *NAALADLI*, it is closer to the transcription start site (TSS) of *SAC3DI*, being this considered as its nearest gene. PheWAS information from Open Targets Genetics revealed a risk association of this SNP with cardiovascular traits such as hypertension ( $p\text{-value}_{\text{PheWAS}} = 7.4 \times 10^{-06}$ ) and high blood pressure ( $p\text{-value}_{\text{PheWAS}} = 5.8 \times 10^{-06}$ ) in the UK Biobank database, as well as with abdominal aortic aneurysm ( $p\text{-value}_{\text{PheWAS}} = 1.9 \times 10^{-03}$ ) in a subset of patients. Significant eQTLs, (i.e. locus that explain part of the variance in gene expression in specific tissues) were described identifying rs2458298 as correlated with *SNX15* expression in heart atrial appendage ( $p\text{-value} = 9.9 \times 10^{-10}$ ) and heart left ventricle ( $p\text{-value} = 1.3 \times 10^{-11}$ ). The same SNP was also correlated with the expression of *FERMT3* in a lymphoblastoid cell line. Using the Capture Hi-C Plotter platform, 42 interactions of this SNP with different gene regions across chromosome 11 were identified in the experiments done by Mifsud and colleagues [15] in a lymphoblastoid cell line (GM12878). One of those relevant promoter interactions is with the *BATF2* gene, which has been previously implicated in *T. cruzi* infection [16]. All these results are summarized on Table 4.

### DISCUSSION

Several suggestive *loci* have been identified in association with differential susceptibility to *T. cruzi* infection in our study. Among these variants, highlight those located in intergenic regions near *IL18* and *CD247* given their implication in the immune response [17, 18]. *IL18* encodes the interleukin-18, a proinflammatory cytokine that is involved in the innate and adaptive response, and is critical for T cell differentiation into interferon- $\gamma$  (IFN- $\gamma$ ) producing Th1-type T cells and for IFN- $\gamma$  production by NK cells [18]. IFN- $\gamma$  is a key player in the acute *T. cruzi* infection and a pathogen resistance gene, aiding the control of *T. cruzi* parasitism by stimulating surrounding cells to produce TNF- $\alpha$  and other inflammatory mediators leading to the generation of peroxynitrite [19]. Additionally, IFN- $\gamma$  gene expression has been reported to increase in chronic cardiac patients in several studies conducted in Latin American populations, indicating an important role of this molecule in the pathogenesis of the chronic cardiac form through the induction of inflammatory damage [20]. Here, we confirm the association of this gene with the differential susceptibility to *T. cruzi* infection; although given the directed nature of candidate gene studies, different SNPs of the same locus were reported [7, 21].

Regarding *CD247*, this gene encodes a subunit of the T-cell receptor CD3 complex that plays an important role in antigen recognition [17]. In Chagas disease, previous works have confirmed lower expression of CD3 in isolated blood cells from individuals with high exposure to the parasite in Colombian populations [22]. Finally, genetic variants in the *CD247* were also previously associated with autoimmunity [23-25] pointing to the existence of shared genes in autoimmune and infectious diseases as described elsewhere [26]. In this way, our results reinforce the potential implication of *IL18* and *CD247* during immune response against *T. cruzi*.

Remarkably, a statistically significant association at genomic level was identified near the *SAC3D1* gene region (rs2458298) with the development of chronic Chagas cardiomyopathy. *SAC3D1*, also known as *SHD1*, has been identified as a transcriptional regulator of *STAT5* [27], also associated with cardioprotection in humans [28]. Additionally, the STAT-5 signaling by IL-2, IL-7 and IL-15 receptors has been shown to be perturbed in peripheral and heart-infiltrating T cells in chronic Chagas cardiomyopathy [29].

An *in silico* functional assessment of the most associated variant and its proxies was performed in order to acknowledge the interactions and processes where they are involved. Interestingly, physical contact of chromatin regions in this genetic location revealed interaction with *BATF2*. This gene encodes a transcription factor (TF) that has been related with Chagas disease, regulating the IL-23-Th17 pathway and suggesting an immunoregulatory function during *T. cruzi* infection in a mouse model [16]. On the other hand, the top signal has been associated with several cardiovascular traits in PheWas data from UK Biobank. Additionally, this variant is an eQTL of *SNX15* in heart atrial appendage and heart left ventricle. This gene encodes a member of the sorting nexin protein family, which has been related with cardiovascular diseases [30]. Additionally, the associated variant was also an eQTL for *FERMT3* (Fermitin Family Member 3) in lymphoblastoid cell lines. The *FERMT3* belongs to the Kindlins family and it encodes a protein that plays an important role in the regulation of thrombosis, as well as in maintaining the cytoskeleton of erythrocytes [31]. This gene was previously associated with serum triglyceride levels, considered as a risk factor for coronary heart disease [32]. Although, the diagnosis of chronic Chagas cardiomyopathy involves the

exclusion of other cardiomyopathies, the association with this gene may implicate common pathways among them.

Regarding other suggestive signals, stand out those located in chromosome 9 near to *KLF4*, which encodes the Kruppel-like factor 4, a TF related with cardiac mitochondrial homeostasis in a mouse model [33], as well as with a high expression of the inducible nitric oxide synthase and nitric oxide (NO) production in different cell types [34]. Taking into account that higher production of NO has been reported in chronic Chagas cardiomyopathy severity [35], it would be possible a relationship among *KLF4* and chronic Chagas cardiomyopathy through the NO production.

This study has certain limitations. First, is a cross-sectional assessment and the posterior seroconversion or development of cardiac events cannot be ruled out. In this case the inclusion of cases in the control group will biased the results towards the null hypothesis of no association, weakening the risk of a false positive result. Second, the associated genes have been nominated as related with the disease through an *in-silico* functional assessment in publicly available non-endemic populations. Further functional analyses will guarantee the acceptance or rebuttal of these hypotheses. Finally, the small sample size in the overall study but especially in the Chagas disease susceptibility assessment, bounds the statistical power for detecting lower frequency variants or with smaller effects. This is the reason why suggestive variants were highlighted, including their potential functional relevance.

The main strength of our study is that we reveal a consistent and reproducible association with chronic Chagas cardiomyopathy in four Latin American populations. Genetic studies in admixed populations are challenging given their differential ancestry proportions [36]; which is further complicated when taking also into account the genetic

diversity of the infectious agent [37]. This heterogeneity has been previously observed in a related parasitic disease such as Leishmania [38]. Further studies evaluating the human-parasite genetic interactions may improve the knowledge of disease pathogenesis.

Finally, the inclusion of underrepresented populations in genetic studies is a great opportunity to broaden the knowledge of complex diseases and deepen their molecular mechanisms [39], which is especially important in Chagas disease, due to its significant socio-economic burden in both endemic and non-endemic countries [40].

## **CONCLUSION**

In summary, in this work we describe the results of a genome-wide meta-analysis of Chagas disease in Latin American populations, identifying a novel locus near the gene region of *SAC3DI* in association with the development of chronic Chagas cardiomyopathy. Although larger studies are needed to validate our findings, this study provides important novel leads to understand the pathogenesis of this neglected disease.

## **ACKNOWLEDGEMENTS**

We thank all the patients who participated in this study and the Medical team from the different Latin American countries and Spain. This research is part of the doctoral degree awarded to DCM, within the Biomedicine program from the University of Granada entitled "Bases moleculares de la enfermedad de Chagas: Integrando Genómica, Transcriptómica y Epigenómica".

## **FUNDING**



This research was supported by grants from Red Iberoamericana de medicina genómica en enfermedad de Chagas (RIMGECH - 217RT0524) – CYTED and Ministerio de Ciencia y Tecnología de Córdoba (PID 2018) and Secretaría de Ciencia y Tecnología (Consolidar 2018-2021), Universidad Nacional de Córdoba, Argentina. MAH was supported by Juan de la Cierva fellowship (IJC2018-035131-I). ES and ALR was supported by the National Heart, Lung, and Blood Institute Retrovirus Epidemiology Donor Study-II (REDS-II), International Component Contract (HHSN268200417175C). ECN and CC were supported by an international program funded both by the Agence Nationale de Recherche (ANR, France) and the São Paulo State Research Funding Agency-(FAPESP, Brazil) agencies (Br-Fr-chagas). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

#### **DECLARATION OF COMPETING INTEREST**

The authors report no conflict of interests

## REFERENCES

1. Perez-Molina JA, Molina I. Chagas disease. *Lancet* **2018**; 391(10115): 82-94.
2. Chapman SJ, Hill AV. Human genetic susceptibility to infectious disease. *Nat Rev Genet* **2012**; 13(3): 175-88.
3. Cunha-Neto E, Chevillard C. Chagas disease cardiomyopathy: immunopathology and genetics. *Mediators Inflamm* **2014**; 2014: 683230.
4. Acosta-Herrera M, Strauss M, Casares-Marfil D, Martin J, Chagas Genetics CN. Genomic medicine in Chagas disease. *Acta Trop* **2019**; 197: 105062.
5. Tam V, Patel N, Turcotte M, Bosse Y, Pare G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet* **2019**; 20(8): 467-84.
6. Deng X, Sabino EC, Cunha-Neto E, et al. Genome wide association study (GWAS) of Chagas cardiomyopathy in Trypanosoma cruzi seropositive subjects. *PLoS One* **2013**; 8(11): e79629.
7. Strauss M, Acosta-Herrera M, Alcaraz A, et al. Association of IL18 genetic polymorphisms with Chagas disease in Latin American populations. *PLoS Negl Trop Dis* **2019**; 13(11): e0007859.
8. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **2015**; 4: 7.
9. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **2012**; 491(7422): 56-65.
10. Kang HM SJ, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. Variance component model to account for sample structure in genome-wide association studies. *Nature genetics* **2010**; 42(4): 348-54.

11. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet* **2011**; 88(5): 586-98.
12. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **2015**; 31(21): 3555-7.
13. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **2012**; 40(Database issue): D930-4.
14. Schofield EC, Carver T, Achuthan P, et al. CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets. *Bioinformatics* **2016**; 32(16): 2511-3.
15. Mifsud B, Tavares-Cadete F, Young AN, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet* **2015**; 47(6): 598-606.
16. Kitada S, Kayama H, Okuzaki D, et al. BATF2 inhibits immunopathological Th17 responses by suppressing IL23a expression during *Trypanosoma cruzi* infection. *J Exp Med* **2017**; 214(5): 1313-31.
17. Alcover A, Alarcon B, Di Bartolo V. Cell Biology of T Cell Receptor Expression and Regulation. *Annual review of immunology* **2018**; 36: 103-25.
18. Kaplanski G. Interleukin-18: Biological properties and role in disease pathogenesis. *Immunological reviews* **2018**; 281(1): 138-53.
19. Chevillard C, Nunes JPS, Frade AF, et al. Disease Tolerance and Pathogen Resistance Genes May Underlie *Trypanosoma cruzi* Persistence and Differential Progression to Chagas Disease Cardiomyopathy. *Front Immunol* **2018**; 9: 2791.

20. Cunha-Neto E, Dzau VJ, Allen PD, et al. Cardiac gene expression profiling provides evidence for cytokinopathy as a molecular mechanism in Chagas' disease cardiomyopathy. *The American journal of pathology* **2005**; 167(2): 305-13.
21. Leon Rodriguez DA, Carmona FD, Echeverria LE, Gonzalez CI, Martin J. IL18 Gene Variants Influence the Susceptibility to Chagas Disease. *PLoS Negl Trop Dis* **2016**; 10(3): e0004583.
22. Gomez-Olarte S, Bolanos NI, Cuellar A, Puerta CJ, Gonzalez JM. Diminished mitogen-induced T cell proliferation by *Trypanosoma cruzi* antigens associated with antigen-presenting cell modulation and CD3 signaling. *Cell Immunol* **2020**; 348: 103974.
23. Radstake TR, Gorlova O, Rueda B, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet* **2010**; 42(5): 426-9.
24. Takeuchi T, Suzuki K. CD247 variants and single-nucleotide polymorphisms observed in systemic lupus erythematosus patients. *Rheumatology* **2013**; 52(9): 1551-5.
25. Teruel M, McKinney C, Balsa A, et al. Association of CD247 polymorphisms with rheumatoid arthritis: a replication study and a meta-analysis. *PLoS One* **2013**; 8(7): e68295.
26. Matzaraki V, Kumar V, Wijmenga C, Zernakova A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome biology* **2017**; 18(1): 76.

27. Nakajima H, Tamura T, Ito M, et al. SHD1 is a novel cytokine-inducible, negative feedback regulator of STAT5-dependent transcription. *Blood* **2009**; 113(5): 1027-36.
28. Heusch G, Musiolik J, Kottenberg E, Peters J, Jakob H, Thielmann M. STAT5 activation and cardioprotection by remote ischemic preconditioning in humans: short communication. *Circ Res* **2012**; 110(1): 111-5.
29. Albareda MC, Perez-Mazliah D, Natale MA, et al. Perturbed T cell IL-7 receptor signaling in chronic Chagas disease. *Journal of immunology* **2015**; 194(8): 3883-9.
30. Yang J, Villar VAM, Rozyyev S, Jose PA, Zeng C. The emerging role of sorting nexins in cardiovascular diseases. *Clin Sci (Lond)* **2019**; 133(5): 723-37.
31. Kruger M, Moser M, Ussar S, et al. SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* **2008**; 134(2): 353-64.
32. Haas BE, Horvath S, Pietilainen KH, et al. Adipose co-expression networks across Finns and Mexicans identify novel triglyceride-associated genes. *BMC Med Genomics* **2012**; 5: 61.
33. Liao X, Zhang R, Lu Y, et al. Kruppel-like factor 4 is critical for transcriptional control of cardiac mitochondrial homeostasis. *J Clin Invest* **2015**; 125(9): 3461-76.
34. Mo X, Chen J, Wang X, et al. Kruppel-like factor 4 regulates the expression of inducible nitric oxide synthase induced by TNF-alpha in human fibroblast-like synoviocyte MH7A cells. *Mol Cell Biochem* **2018**; 438(1-2): 77-84.

35. Carvalho CM, Silverio JC, da Silva AA, et al. Inducible nitric oxide synthase in heart tissue and nitric oxide in serum of Trypanosoma cruzi-infected rhesus monkeys: association with heart injury. *PLoS Negl Trop Dis* **2012**; 6(5): e1644.
36. Conomos MP, Laurie CA, Stilp AM, et al. Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. *Am J Hum Genet* **2016**; 98(1): 165-84.
37. Zingales B. Trypanosoma cruzi genetic diversity: Something new for something known about Chagas disease manifestations, serodiagnosis and drug sensitivity. *Acta Trop* **2018**; 184: 38-52.
38. Castellucci LC, Almeida L, Cherlin S, et al. A Genome-Wide Association Study Identifies SERPINB10, CRLF3, STX7, LAMP3, IFNG-AS1 and KRT80 As Risk Loci Contributing to Cutaneous Leishmaniasis In Brazil. *Clin Infect Dis* **2020**.
39. Genetics for all. *Nature Genetics* **2019**; 51(4): 579-.
40. Lidani KCF, Andrade FA, Bavia L, et al. Chagas Disease: From Discovery to a Worldwide Health Problem. *Front Public Health* **2019**; 7: 166.

## Figure legends

**Figure 1.** Manhattan plot of the association test in chronic Chagas cardiomyopathy meta-analysis results. Y and X axes refer the  $-\log_{10}$  transformed *p-values* and positions in chromosomes, respectively. The red horizontal line refers the genome-wide association threshold, established at  $p < 5 \times 10^{-8}$ , and the blue one refers the suggestive threshold ( $p < 1 \times 10^{-6}$ ).

**Figure 2.** Regional association plot for the genome-wide significant signal. The  $-\log_{10}$  transformed *p-values* of associated variants are plotted by position. The genomic position indicated correspond the top variant associated in this region. The rest of the variants are colored in function of their degree of linkage disequilibrium (LD) with the associated signal. LD is based on pairwise  $r^2$  values from the American population from the 1000 Genomes Project. The blue line correspond the estimated recombination rates. Nearby genes to the associated and suggestive signals are showed in the bottom of the plot.

## Tables

**Table 1. Demographical characteristics and classification of data collections.**

	Seropositive								Seronegative
	Colombia		Bolivia		Argentina		Brazil*		Colombia
	CCC	ASY	CCC	ASY	CCC	ASY	CCC	ASY	
<b>Subjects, N</b>	577	442	97	518	100	62	207	306	1,104
<b>Sex (% females)</b>	280 (48 %)	282 (64%)	57 (59%)	364 (70%)	65 (65%)	40 (65%)	81 (39%)	154 (53%)	672 (61%)
<b>Age (Mean±SD)</b>	67.4±14.3	57.2±15.6	50.7±9.4	46.4±9.5	58.2±9.4	47.5±14.0	-	-	49.5±16.7

The sample sizes refer to data passing the genotyping quality controls.

Abbreviations: CCC, chronic chagas Cardiomyopathy; ASY, asymptomatic.

\* Data set from Deng et al, 2013



**Table 2. Suggestive\* signals associated with susceptibility to *Trypanosoma cruzi* infection in the Colombian cohort.**

Chr	Locus	Bp	SNP	SNP status	Rsq	Effect allele	Allele Freq	<i>p-value</i>	OR (95% CI)	Function
1	<i>CD247</i>	167431884	rs554994388	Imputed	0.50	G	0.03	5.35E-07	1.33 (1.19-1.49)	Intronic
8	<i>EBF2</i>	25979367	rs147475322	Imputed	0.86	A	0.03	2.15E-07	1.20 (0.76-1.90)	Intronic
11	<i>IL18</i>	112127040	rs4937075	Imputed	0.77	C	0.77	8.34E-07	1.02 (0.86-1.21)	Intergenic
21	<i>SIK1</i>	43428190	rs229347	Imputed	0.88	A	0.60	8.93E-07	1.07 (1.04-1.10)	Intergenic

\*Suggestive associations were considered with a  $p\text{-value} \leq 1E-06$

Associated *loci* were determined in order to the LD of contained variants and the most significant SNP of each *loci* were reported in the table.

The r-squared metric (Rsq) indicates the correlation between input genotypes and imputed dosages.

Abbreviations: Allele Freq allele frequency; bp base pair; chr chromosome; CI confidence interval; OR odds ratio; SNP info SNP information.

**Table 3. Genome-wide significant association and suggestive\* signals associated with chronic Chagas cardiomyopathy development in the meta-analysis.**

Chr	Locus	Bp	SNP	SNP status	Rsq	Effect allele	Allele Freq	<i>p-value</i> Q	<i>p-value</i>	OR (95% CI)	Function
5	<i>CDH8</i>	18718887	rs10472156	Imputed	0.81	A	0.73	0.72	8.89E-07	1.10 (1.06-1.14)	Intergenic
9	<i>KLF4</i>	107479876	rs10759240	Genotyped	-	C	0.54	0.98	4.68E-07	0.92 (0.89-0.95)	Intergenic
<b>11</b>	<b><i>SAC3DI</i></b>	<b>65047341</b>	<b>rs2458298</b>	Imputed	<b>0.95</b>	<b>G</b>	<b>0.26</b>	<b>0.98</b>	<b>3.27E-08</b>	<b>0.90</b> <b>(0.87-0.94)</b>	<b>Intronic</b>

\*Suggestive associations were considered with a  $p\text{-value} \leq 1E-06$

Associated *loci* were determined in order to the LD of contained variants and the most significant SNP of each *loci* were reported in the table.

In boldface the genome-wide significant signal.

The r-squared metric (Rsq) indicates the correlation between input genotypes and imputed dosages.

Abbreviations: Allele Freq allele frequency; bp base pair; chr chromosome; CI confidence interval; OR odds ratio; *p-value* Q Cochran's Q test; rsq r-square; SNP info SNP information.

**Table 4. *In silico* functional analysis of the associated variant with chronic Chagas cardiomyopathy development.**

Chr	SNP	Nearest gene	eQTLs	C-HiC Genes	PheWAS
11	rs2458298	<i>SAC3D1</i> <i>NAALADLI</i>	<i>SNX15</i> <sup>a</sup> <i>FERMT3</i> <sup>b</sup>	<i>BATF2</i> in GM12878 <sup>b</sup>	Hypertension, high blood pressure and abdominal aortic aneurysm

<sup>a</sup>eGENE in heart atrial appendage and heart left ventricle

<sup>b</sup>lymphoblastoid cell line

Queried data bases were Open Targets Genetics, GTEX, RegulomeDB, Capture HiC Plotter, and PheWEB

Abbreviations: Chr chromosome; C-HiC capture Hi-C; eQTL expression quantitative trait loci; PheWAS phenome-wide association study.