Common variants in the HLA-DRB1–HLA-DQA1 HLA class II region are associated with susceptibility to visceral leishmaniasis

LeishGEN Consortium1,2 & Wellcome Trust Case Control Consortium 21,2

To identify susceptibility loci for visceral leishmaniasis, we undertook genome-wide association studies in two populations: 989 cases and 1,089 controls from India and 357 cases in 308 Brazilian families (1,970 individuals). The HLA-DRB1–HLA-DQA1 locus was the only region to show strong evidence of association in both populations. Replication at this region was undertaken in a second Indian population comprising 941 cases and 990 controls, and combined analysis across the three cohorts for rs9271858 at this locus showed \( P_{\text{combined}} = 2.76 \times 10^{-17} \) and odds ratio (OR) = 1.41, 95% confidence interval (CI) = 1.30–1.52. A conditional analysis provided evidence for multiple associations within the HLA-DRB1–HLA-DQA1 region, and a model in which risk differed between three groups of haplotypes better explained the signal and was significant in the Indian discovery and replication cohorts. In conclusion, the HLA-DRB1–HLA-DQA1 HLA class II region contributes to visceral leishmaniasis susceptibility in India and Brazil, suggesting shared genetic risk factors for visceral leishmaniasis that cross the epidemiological divides of geography and parasite species.

Leishmania are protozoan parasites that live in macrophages. They are transmitted by sand flies and cause severe and debilitating cutaneous as well as fatal visceral disease in subtropical and tropical regions worldwide. Leishmaniasis is classified by the World Health Organization (WHO) as a neglected tropical disease. It affects 12 million people, and there are an estimated 1.5 million new cases annually1. Of these, 500,000 are cases of potentially fatal visceral leishmaniasis caused by the Leishmania donovani complex, 90% of which occur in three foci in India-Bangladesh-Nepal, Sudan and Brazil. Skin tests and lymphocyte response analyses indicate that only 1 in 5–10 infected individuals develop clinical disease2–4. The importance of host genetic factors is indicated by familial clustering of clinical disease5 and high sibling risk ratios6. However, human genetic studies undertaken to date (reviewed in refs. 7–9) provide inconsistent results. As part of the Wellcome Trust Case Control Consortium 2 (WTCCC2) study of 15 complex disorders and traits, we report the first genome-wide association study (GWAS) of visceral leishmaniasis across major foci of disease caused by L. donovani in India and Leishmania infantum chagasi in Brazil.

Subjects for the India discovery GWAS (Online Methods and Supplementary Table 1) were recruited from Bihar state in northeast India. Affected individuals and controls were matched by self-reported age, sex, religion, caste and geographic region of recruitment. The Brazilian family-based sample was collected as part of the Belém Family Study and from study sites near Natal6,10,11 (Online Methods and Supplementary Table 1).

All individuals were genotyped at the Wellcome Trust Sanger Institute on the custom Illumina Human660W-Quad chip (Online Methods). After performing stringent quality control procedures12 (Online Methods), the Indian discovery data set comprised 2,078 individuals (989 cases and 1,089 controls) genotyped at 526,731 SNPs, and the Brazilian discovery data set included a total of 1,970 individuals (357 cases from 308 families) genotyped at 553,323 SNPs.

As ancestry differences and close relationships between case and control individuals can confound association studies, we used a variance components method (sometimes referred to as a mixed model), which models the pairwise relatedness between individuals to account for population structure in the data13. This approach accounts for relatedness on different scales, from close relatives to distant ancestral structure, and it could thus be applied to both the case-control samples in India and the family data from Brazil. Similar models have recently been used in several other association studies14–18. An advantage of our implementation of the mixed model is that we are able to estimate effect sizes on the log-odds scale. Using this method, the genomic inflation factor (\( \lambda \)) was 1.03 for the Indian analysis and 1.02 for the Brazilian analysis, showing that it successfully handled any population structure or relatedness in the data. The mixed-model approach was used for all analyses presented here, for both discovery and replication data and in further dissection of association signals.

Previous studies attempting to identify the genetic components of visceral leishmaniasis susceptibility have typically been small, underpowered family studies or candidate gene studies19–23. We present the results from our discovery GWAS and replication data at these previously identified loci (Table 1). None of the loci previously found

---

1A full list of authors and affiliations appears at the end of the paper. 2A full list of members and affiliations is provided in the Supplementary Note.

Received 23 August 2012; accepted 6 December 2012; published online 6 January 2013; doi:10.1038/ng.2518
to be associated with visceral leishmaniasis consistently showed an association with visceral leishmaniasis in our three data sets.

Because different pathogen species are responsible for disease in India and Brazil, we first examined each GWAS separately. Our primary focus was on regions showing association in both GWAS analyses. This would be appropriate if similar host genetic factors affected susceptibility to the two different parasite species. Observing the same association in different populations with different study founders. Further replication in a Brazilian data set is required to assess this additional MHC signal: it could be a population- or parasite species–specific association or, alternatively, a false positive finding.

In the Indian discovery GWAS, the two SNPs most associated with visceral leishmaniasis in the MHC region were in strong linkage disequilibrium (LD) (rs9271858, \( r^2 = 1 \) in Indian replication = 0.58 at rs4428528; \( r^2 = 0.06 \) in the Brazilian controls). One of these SNPs was present in the replication data (rs9271858); therefore, for convenience, we focused on this SNP and refer to it as the top SNP. This SNP also showed significant association with visceral leishmaniasis in the MHC region; two highly correlated SNPs

A second Indian cohort was used in a replication analysis of the findings from the two discovery GWAS. These samples were genotyped on the Immunochip, a purpose-designed Illumina chip, some content of which was chosen by WTCCC2 for deep replication of its GWAS findings. After quality control (Online Methods), the Indian replication data set comprised 1,931 individuals (941 visceral leishmaniasis cases and 990 controls). Outside of the MHC region, none of the regions that were associated at \( P < 1 \times 10^{-5} \) in either the separate

In the Indian discovery cohort, when conditioning on the top SNP from the Indian discovery GWAS (rs9271858), there was still residual association signal in the MHC region; two highly correlated SNPs

### Table 1 Signals of association at previously reported non-HLA loci for visceral leishmaniasis

| SNP     | Chr. | Locus | Position (bp) | Risk allele | Populationa | Reference | This study | Odds ratio | 95% CI      | \( P \) value |
|---------|------|-------|---------------|-------------|--------------|-----------|-----------|------------|------------|-------------|
| rs4674259 | 2q35 | CXCR2 | 218699250     | A           | India        | 19        | Indian discovery | 0.92       | 0.81–1.04  | 0.195       |
| rs4672875c | 2q35 | CXCR1 | 218729630     | A           | India        | 19        | Indian replicationb | 0.87       | 0.77–1.00  | 0.040       |
| rs2276631 | 2q35 | SLC11A1 | 218957257   | A          | Sudan        | 23        | Indian discovery | 1.05       | 0.89–1.23  | 0.582       |
| rs3731865 | 2q35 | SLC11A1 | 218958247   | G          | Sudan        | 23        | Indian replication | 0.96       | 0.80–1.15  | 0.633       |
| rs17860704 | 6q27 | DLL1-FAM120B | 170444354 | G           | Sudan        | 22        | Indian discovery | 0.99       | 0.82–1.19  | 0.933       |
| rs159273 | 17q12 | CCL1 | 29714798     | C           | Brazil       | 21        | Indian discovery | 1.14       | 0.97–1.34  | 0.117       |
| rs2063979 | 17q12 | CCL16 | 31327679     | A           | Brazil       | 21        | Indian replication | 1.05       | 0.95–1.15  | 0.362       |
| rs854680 | 17q12 | CCL16 | 31333164     | C           | Brazil       | 21        | Indian replication | 1.15       | 0.86–1.56  | 0.334       |
| rs228953 | 22q12 | IL2RB | 35861382     | G           | Sudan        | 20        | Indian replication | 0.97       | 0.85–1.10  | 0.615       |

Chr., chromosome; NA, the SNP was not genotyped in this sample. Odd ratios are given for the previously reported risk allele. Positions are in NCBI human genome Build 36 coordinates.

aPopulation in which the initial SNP association was reported. bThe Indian replication sample reported in the table overlaps significantly with the case-control population described by Mehrotra et al.\(^{19}\) in which rs4674259 was genotyped as part of a separate Sequenom panel. c\( r^2 = 1 \) (HapMap Utah residents of Northern and Western Europe ancestry (CEU)) and 0.71 (HapMap Gujarati Indians in Houston, Texas (GIH)) with previously reported rs3138060.
both associated at $P < 1 \times 10^{-3}$ (rs9271252 and rs9271255; $r^2 = 1$ in Indian controls). For convenience, because one of these SNPs was typed in the discovery and replication samples, we focused on that SNP, rs9271255, which showed an association $P$ value of $8.79 \times 10^{-4}$ in the conditional analysis. This SNP was not highly correlated with rs9271858 ($r^2 = 0.21$ in Indian controls). When conditioning on both rs9271858 and rs9271255, no association signal remained at $P < 1 \times 10^{-3}$. In the Indian replication data, when conditioning on rs9271858, the association at rs9271255 was still significant ($P = 5.72 \times 10^{-3}$). This conditional analysis was repeated in the Brazilian discovery data: when conditioning on rs9271858, the association of rs9271255 was significant with a $P$ value of $1.91 \times 10^{-5}$.

We then turned to a more detailed examination of the replicated MHC signal (at rs9271858 and rs9271255). We focused on the Indian data, for which the Indian replication cohort was from a comparable population and could therefore be used to replicate cohort-specific association signals. To further explore the MHC signal in Indian cases, we estimated phase for the SNP genotypes for all individuals in the Indian discovery group. We subsequently fitted a model to the phased haplotypes, with different risk parameters for the four haplotypes defined by the phased genotypes at the two aforementioned SNPs, rs9271858 and rs9271255. After additional exploration of the model (Supplementary Note), we found that combining two of the four haplotypes to give a model with three risk classes fitted the data as well as the model with four risk classes (Supplementary Fig. 1) and also gave a better fit than the two-SNP model (Supplementary Note).

---

**Figure 1** Plot of genome-wide association results for the separate and combined discovery GWAS using a variance components method. SNPs in red show regions with replicated association to visceral leishmaniasis susceptibility. (a) Indian discovery data at 526,731 SNPs. (b) Brazilian discovery data at 553,323 SNPs. (c) Plot of the meta-analysis genome-wide association results.

---

**Figure 2** Regional association plots of the signal at the MHC region. (a-c) Results are shown for the Indian discovery (a), Brazilian discovery (b) and meta-analysis of the discovery cohorts (c). Top, $-\log_{10} P$ values. SNPs are colored on the basis of their LD with the labeled hit SNP, as calculated in the Indian controls (a) and Brazilian founders (b); no coloring is shown for the meta-analysis (c). The SNP colored in green is the hit SNP in the Brazilian cohort (a), Indian cohort (b) and Indian and Brazilian cohorts (c). Bottom, fine-scale recombination rates estimated from CEU HapMap population with genes marked as horizontal blue arrows. Genes flanking the hit SNP are colored red and are labeled.
Figure 3 Schematic of the HLA and SNP phased Indian discovery haplotypes in the MHC region. Each row represents an individual haplotype, and the three middle columns show different risk models, with the shading denoting in which class in the model that haplotype falls; column 2 shows the classification of haplotypes solely on the basis of the allele (risk or protective) present at the top SNP, rs9271858; column 3 shows the classification of haplotypes solely on the basis of the allele present at the top SNP in the conditional analysis, rs9271255; column 4 shows the three risk groups defined by the allele combinations (risk-risk, protective-risk, protective-protective) at SNPs rs9271858 and rs9271255, respectively, that most parsimoniously capture their phase-known haplotype associations with disease (Supplementary Fig. 1). The two outside columns show the HLA alleles at HLA-DRB1 and HLA-DQB1, with colors representing the presence of each allele. This analysis used the subset of individuals in the Indian discovery data for whom classical HLA alleles were available, and the plot shows only the 112 haplotypes for which the posterior probability of the estimated phase at both HLA-DRB1 and HLA-DQB1 was >0.9. Note that the haplotype carrying HLA-DRB1*0101 is rare and is not shown in the plot because it did not occur on a haplotype with a successfully phased HLA-DQB1 allele.

At rs9271858 and rs9271255, respectively, the haplotype with the highest risk was defined by the alleles GA, and the haplotype with intermediate risk was defined by AA alleles, with the haplotypes AG and GG being protective relative to the other two haplotypes and having statistically indistinguishable risks from them. This model with three risk classes had \( P = 1 \times 10^{-11} \) in the Indian discovery data compared to the null model where all haplotype classes had the same risk. When the protective haplotypes were set as the baseline with OR = 1, the intermediate haplotype OR (95% CI) was 1.28 (1.09–1.50), and the highest risk haplotype OR (95% CI) was 1.72 (1.48–2.00). This analysis replicated in the Immunochip data: the model with three risk classes had significant at \( P = 4.78 \times 10^{-6} \), with the intermediate haplotype OR (95% CI) equal to 1.19 (1.00–1.40) and the highest risk haplotype OR (95% CI) equal to 1.50 (1.27–1.76). In India (\( n = 3,732 \)), the frequencies of the 2-SNP haplotypes were AA, 0.301; AG, 0.310; GA, 0.372; and GG, 0.017. In Brazil (\( n = 1,026 \)), the frequencies of the 2-SNP haplotypes were AA, 0.406; AG, 0.097; GA, 0.390; and GG, 0.105.

Our results establish that common polymorphisms in the HLA-DRB1–HLA-DQA1 segment of the MHC region are genetic risk factors for visceral leishmaniasis, and, notably, their effects seem to cross the epidemiological divides of geography and parasite species. For GWAS associations, the top SNPs are not necessarily the causal variants but may be correlated with the functional variants. There are SNPs in the MHC region that are in moderate LD with rs9271858 (\( r^2 = 0.2 \)) and are located up to 500 kb away (for example, rs408359: \( r^2 = 0.19 \) calculated in the Indian controls). This extensive LD is common in the MHC region and makes it particularly difficult to assess possible functional candidates within the region of association.

One possibility is that the association signal could be driven by functional variation in the adjacent highly polymorphic classical HLA class II genes HLA-DRB1 and HLA-DQB1, each of which is a natural immunological candidate. The classical alleles at these loci were genotyped in a subset of individuals in both of the discovery cohorts and were phased (statistically) onto the haplotypes estimated from all the SNP data (Supplementary Note). In this phased data set, chromosomes carrying the protective haplotype at rs9271858 and rs9271255 always carried one of the three classical HLA-DRB1*15, HLA-DRB1*16 and HLA-DRB1*01 alleles, and the reverse also held true (Fig. 3 and Supplementary Table 5). We note that the HLA-DRB1*01 and HLA-DRB1*16 alleles are very rare in the Indian data (Supplementary Table 5), meaning that we had limited power to assess their possible roles. This correlation between protective SNP haplotype and classical HLA-DRB1*15, HLA-DRB1*16 and HLA-DRB1*01 alleles was also perfect in the 142 phased Brazilian discovery sample chromosomes (Online Methods), where all of the chromosomes carrying the protective haplotype defined by rs9271858 and rs9271255 also carried one of the HLA-DRB1*15, HLA-DRB1*16 or HLA-DRB1*01 alleles (Supplementary Fig. 2 and Supplementary Table 5). This lends support to the possibility that the association signal could be driven by classical HLA alleles and emphasizes the biological importance of endogenous processing and presentation of antigen from infected macrophages and dendritic cells to CD4+ T cells that drive the immune response in visceral leishmaniasis. Nevertheless, additional research will be needed to establish whether or not these HLA-DRB1 alleles are themselves functionally involved in protection against visceral leishmaniasis or are merely correlated with an as-yet-uncharacterized functional variant. Further details of the HLA analysis are presented in the Supplementary Note.

Previous studies of HLA associations in human leishmaniasis have provided confusing and often contradictory results, which have been generally compromised by small sample size and lack of power (reviewed in ref. 9). One small study did find protection associated with HLA-DRB1*15 and HLA-DRB1*16 for visceral leishmaniasis caused by Old World L. infantum, the same parasite species that causes visceral leishmaniasis in Brazil. Protection has also been associated with HLA-DR2 (an alternative name for the HLA-DRB1 serogroup that comprises the HLA-DRB1*15 and HLA-DRB1*16...
alleles) in small studies of cutaneous and mucocutaneous forms of leishmaniasis caused by *Leishmania mexicana* and *Leishmania braziliensis*, respectively.

There have been relatively few successful GWAS of infectious disease susceptibility—. Our GWAS has identified HLA class II region polymorphisms as significant genetic risk factors for visceral leishmaniasis, with the risk associated with the high-risk haplotype being high by GWAS standards: OR (95% CI) estimated from replication data of 1.50 (1.27–1.76). Notably, genetic variation in the MHC region has been associated with disease susceptibility, with the risk associated with the high-risk haplotype being even higher. Our findings focus attention on determining the precise nature of genetic variation in determining visceral leishmaniasis susceptibility and offer the potential to contribute to the development of strategies for disease control.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**ACKNOWLEDGMENTS**

Funding for this study was provided by the Wellcome Trust, as part of the WTCCC2 project (B98547/B/08/Z and B98547/Z/08/Z). We thank S. Bertrand, J. Bryant, S.L. Clark, J.S. Conquer, T. Dibbling, J.C. Eldred, S. Gamble, C. Hind, M.L. Perez, C.R. Stribling, S. Taylor and A. Wilk of the Wellcome Trust Sanger Institute's Sample and Genotyping Facilities for technical assistance. We thank D. Davison for making available his Shellfish program for calculating principal components in large genetic data sets. C.C.A.S. is supported by a Wellcome Trust Fellowship (097364/Z/11/Z). H.J.C. is supported by a Wellcome Senior Fellowship in Basic Biomedical Science (087436/Z/10/Z). P. Donnelly is supported in part by a Royal Society Wolfson Merit Award, and work was supported in part by Wellcome Trust Centre for Human Genetics core grants 090532/Z/09/Z and 075491/Z/04/B. Collection of samples and epidemiological data, sample preparation and sequencing-based HLA typing were supported by grants from the Wellcome Trust (074196/Z/04/Z and 085475/Z/08/Z to J.M.B., S.S., S.M.B.J. and M.E.W.) and the US National Institutes of Health (Tropical Medicine Research Center award P50 AI074321 to S.S. in India; Tropical Medicine Research Center award P50 AI 30639 to E.M. Carvalho in Brazil; and R01 AI076233 to M.E.W. and J.M.B.; R01 AI048822 to M.E.W., S.M.B.J. and J.M.B.). We give special thanks to all subjects who contributed samples and to clinicians and field staff in India and Brazil who helped with the recruitment of study subjects.

**AUTHOR CONTRIBUTIONS**

M.F., E.N.M., A.M., S.M., G.R.M., H.G.L., N.N.P., M.R., S.P., O.S., M.E.W., S.M.B.J., S.S. and J.M.B. oversaw cohort collections for LeishGEN. The WTCCC2 M.F., E.N.M., A.M., S.M., G.R.M., H.G.L., N.N.P., M.R., S.P.S., O.S., M.E.W., E. Gray, S.E., E. Kruger, H. Azevedo, E.S. & Kriger, H. Familial aggregation of *Leishmania chagasi* infection in northeastern Brazil. *Am. J. Trop. Med. Hyg.* 52, 364–369 (1999).

Petzl-Erler, M.L., Belich, M.P. & Queiroz-Telles, F. Association of mucosal helper 1 (Th1) cell– and Th2 cell–associated cytokines in Indian patients with visceral leishmaniasis. *J. Infect. Dis.* 200, 370–385 (2009).

Jeromino, S.M., et al. Genetic predisposition to self-curing infection with the protozoan *Leishmania chagasi*: a genomewide scan. *J. Infect. Dis.* 196, 1261–1269 (2007).

Ewing, N.A. et al. Genetic admixture in Brazilians exposed to infection with *Leishmania chagasi*. *Ann. Hum. Genet.* 73, 304–313 (2009).

Sawcer, S. Multiple sclerosis: the pathogenesis of a complex disease? *Nature* 476, 214–219 (2011).

Bellenguez, C., Straub, E., Freeman, C., Donnelly, P. & Spencer, C.C. A robust clustering algorithm for identifying problematic samples in genome-wide association studies. *Bioinformatics* 25, 2013–2019 (2009).

Bucheton, B. et al. Identification of a novel G245R polymorphism in the IL-2 receptor β membrane proximal domain associated with human visceral leishmaniasis. *Genes Immun.* 8, 79–83 (2007).

Carvalho, E.M. et al. Genome-wide and fine-resolution association analysis of malaria susceptibility genes in Brazil. *Genetics* 188, 84–90 (2007).

Fakiola, M. et al. Functional evidence implicating DLL1 as the gene that influences susceptibility to visceral leishmaniasis at chromosome 6q27. *J. Infect. Dis.* 204, 467–471 (2011).

Mohamed, H.S. et al. SLC11A1 (formerly NRAMP1) and susceptibility to visceral leishmaniasis in The Sudan. *Eur. J. Hum. Genet.* 12, 66–74 (2004).

Cortes, A. & Brown, M.A. Promise and pitfalls of the ImmunoChip. *Arthritis Res. Ther.* 13, 101 (2011).

Bakker, P.I. et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* 38, 1166–1172 (2006).

Sunder, S., Reed, S.G., Sharma, S., Mehrotra, A. & Murray, H.W. Circulating T helper (Th1) cell– and Th2 cell–associated cytokines in Indian patients with visceral leishmaniasis. *Am. J. Trop. Med. Hyg.* 55, 522–525 (1997).

Carvalho, E.M. et al. Immunologic markers of clinical evolution in children recently infected with *Leishmania donovani* chagasi. *J. Infect. Dis.* 165, 535–540 (1992).

Kubits, K. et al. Comparative microsatellite typing of new world leishmaniasis infantum reveals low heterogeneity among populations and its recent old world origin. *PLoS Negl. Trop. Dis.* 5, e1155 (2011).

Olive-Diaz, A. et al. Role of HLA class II alleles in susceptibility to and protection from localized cutaneous leishmaniasis. *Hum. Immunol.* 65, 255–261 (2004).

Cabrera, M. et al. Polymorphism in TNF genes associated with cutaneous leishmaniasis. *J. Exp. Med.* 182, 1259–1264 (1995).

Petz-Lewis, M.L., Belich, M.P. & Queiroz-Telles, F. Association of mucosal leishmaniasis with HLA. *Hum. Immunol.* 32, 254–260 (1991).

Jallow, M. et al. Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat. Genet.* 41, 657–665 (2009).

Thye, T. et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat. Genet.* 42, 739–741 (2010).

Khor, C.C. et al. Genome-wide association study identifies susceptibility loci for dengue shock syndrome at MHC and PLCe1. *Nat. Genet.* 43, 1139–1141 (2011).

Zhong, X. et al. Genome-wide association study of leishmaniasis. *Nat. Genet.* 45, 763–769 (2013).

31. M.E.W. et al. Genetic predisposition to self-curing infection with the protozoan *Leishmania chagasi*: a genomewide scan. *J. Infect. Dis.* 196, 1261–1269 (2007).

32. Astle, W. & Balding, D.J. Population structure and cryptic relatedness in genetic association studies. *Stat. Sci.* 24, 451–471 (2009).

33. Sawcer, S. Multiple sclerosis: the pathogenesis of a complex disease? *Nature* 476, 214–219 (2011).

34. Bellenguez, C., Straub, E., Freeman, C., Donnelly, P. & Spencer, C.C. A robust clustering algorithm for identifying problematic samples in genome-wide association studies. *Bioinformatics* 25, 2013–2019 (2009).

35. Bucheton, B. et al. Identification of a novel G245R polymorphism in the IL-2 receptor β membrane proximal domain associated with human visceral leishmaniasis. *Genes Immun.* 8, 79–83 (2007).

36. Petrozza, C.R. et al. Genetic predisposition to self-curing infection with the protozoan *Leishmania chagasi*: a genomewide scan. *J. Infect. Dis.* 196, 1261–1269 (2007).

1. Alvar, J. et al. Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE* 7, e35671 (2012).

2. Bucheton, B. et al. The interplay between environmental and host factors in the development of visceral leishmaniasis caused by *Leishmania mexicana* and *Leishmania braziliensis*, respectively. *PLoS ONE* 7, e35671 (2012).

3. Sacks, D.L., Lal, S.L., Shrivastava, S.N., Blackwell, J.M. & Neva, F.A. An analysis of T cell responsiveness in Indian Kala-azar. *J. Immunol.* 138, 908–913 (1987).
Michaela Fakiola\textsuperscript{1,29}, Amy Strange\textsuperscript{2,29}, Heather J Cordell\textsuperscript{3}, E Nancy Miller\textsuperscript{1,28}, Matti Pirinen\textsuperscript{2}, Zhan Su\textsuperscript{2}, Anshuman Mishra\textsuperscript{4}, Sanjana Mehrotra\textsuperscript{4}, Gloria R Monteiro\textsuperscript{5}, Gavin Band\textsuperscript{2}, Céline Bellenguez\textsuperscript{2}, Serge Dronov\textsuperscript{6}, Sarah Edkins\textsuperscript{6}, Colin Freeman\textsuperscript{2}, Eleni Giannoulatou\textsuperscript{2}, Emma Gray\textsuperscript{6}, Sarah E Hunt\textsuperscript{6}, Henio G Lacerda\textsuperscript{7}, Cordelia Langford\textsuperscript{6}, Richard Pearson\textsuperscript{2}, Núbia N Pontes\textsuperscript{5}, Madhukar Rai\textsuperscript{4}, Shri P Singh\textsuperscript{4}, Linda Smith\textsuperscript{8}, Olivia Sousa\textsuperscript{3}, Damjan Vukcevic\textsuperscript{2}, Elvira Bramon\textsuperscript{9}, Matthew A Brown\textsuperscript{10}, Juan P Casas\textsuperscript{11}, Aiden Corvin\textsuperscript{12}, Audrey Duncanson\textsuperscript{13}, Janusz Jankowski\textsuperscript{14}, Hugh S Markus\textsuperscript{15}, Christopher G Mathew\textsuperscript{16}, Colin N A Palmer\textsuperscript{17}, Robert Plomin\textsuperscript{18}, Anna Rautanen\textsuperscript{2}, Stephen J Sawcer\textsuperscript{19}, Richard C Trembath\textsuperscript{20}, Ananth C Viswanathan\textsuperscript{21,22}, Nicholas W Wood\textsuperscript{23}, Mary E Wilson\textsuperscript{24,25}, Panos Deloukas\textsuperscript{6}, Leena Peltonen\textsuperscript{6,28}, Frank Christiansen\textsuperscript{8}, Campbell Witt\textsuperscript{8}, Selma M B Jeronimo\textsuperscript{5}, Shyam Sundar\textsuperscript{4}, Chris C A Spencer\textsuperscript{2}, Jenefer M Blackwell\textsuperscript{1,26,30} & Peter Donnelly\textsuperscript{2,27,30}

\textsuperscript{1}Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Addenbrooke’s Hospital, Cambridge, UK. \textsuperscript{2}Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. \textsuperscript{3}Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK. \textsuperscript{4}Institute of Medical Sciences, Banaras Hindu University, Varanasi, India. \textsuperscript{5}Department of Biochemistry, Center for Biosciences, Universidade Federal do Rio Grande do Norte, Natal, Brazil. \textsuperscript{6}Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. \textsuperscript{7}Department of Infectious Diseases, Universidade Federal do Rio Grande do Norte, Natal, Brazil. \textsuperscript{8}Department of Clinical Immunology, Royal Perth Hospital, Perth, Western Australia, Australia. \textsuperscript{9}National Institute for Health Research (NIHR) Biomedical Research Centre for Mental Health at the South London and Maudsley National Health Service (NHS) Foundation Trust and Institute of Psychiatry King’s College London, London, UK. \textsuperscript{10}University of Queensland Diamantina Institute, Princess Alexandra Hospital, University of Queensland, Brisbane, Queensland, Australia. \textsuperscript{11}Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. \textsuperscript{12}Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland. \textsuperscript{13}Molecular and Physiological Sciences, The Welcome Trust, London, UK. \textsuperscript{14}Department of Oncology, University of Oxford, Oxford, UK. \textsuperscript{15}Stroke and Dementia Research Group, St George’s University of London, London, UK. \textsuperscript{16}Department of Medical and Molecular Genetics, King’s College London School of Medicine, Guy’s Hospital, London, UK. \textsuperscript{17}Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK. \textsuperscript{18}Social, Genetic and Developmental Psychiatry Centre, King’s College London Institute of Psychiatry, London, UK. \textsuperscript{19}Department of Clinical Neurosciences, University of Cambridge, Addenbrooke’s Hospital, Cambridge, UK. \textsuperscript{20}Oxford Centre for Diabetes, Endocrinology and Metabolism (ICDEM), Churchill Hospital, Oxford, UK. \textsuperscript{21}NIHR Biomedical Centre for Ophthalmology at Moorfields Eye Hospital NHS Foundation Trust, London, UK. \textsuperscript{22}University College London (UCL) Institute of Ophthalmology, London, UK. \textsuperscript{23}Department of Molecular Neuroscience, Institute of Neurology, London, UK. \textsuperscript{24}Department of Internal Medicine, University of Iowa, Iowa City, Iowa, USA. \textsuperscript{25}Veterans Affairs (VA) Medical Center, Iowa City, Iowa, USA. \textsuperscript{26}Telethon Institute for Child Health Research, Centre for Child Health Research, The University of Western Australia, Subiaco, Western Australia, Australia. \textsuperscript{27}Department of Statistics, University of Oxford, Oxford, UK. \textsuperscript{28}Deceased. \textsuperscript{29}These authors contributed equally to this work. \textsuperscript{30}These authors jointly directed this work. Correspondence should be addressed to P. Donnelly (donnelly@well.ox.ac.uk) or J.M.B. (jblackwell@ichr.uwa.edu.au).
For Brazilian samples, blood was collected by venipuncture from all available family members. Genomic DNA was prepared directly from blood in Natal. For Brazilian samples, blood was collected from study sites near Natal, where transmission of the parasite is focal and transient, allowing the identification of neighborhoods with ongoing or recent transmission. These families were ascertained from clinical, epidemiological, family structure and demographic information relating to the families is described elsewhere. All families were of equivalent socioeconomic status. The causative agent of visceral leishmaniasis was focal and transient, allowing the identification of neighborhoods with ongoing or recent transmission. These families were ascertained from medical records of the Fundação Nacional de Saúde in Rio Grande do Norte. All families were of equivalent socioeconomic status. The causative agent of visceral leishmaniasis was confirmed to be L. infantum chagasi in a subset (~10% over four stages) of individuals with visceral leishmaniasis. Further clinical, epidemiological, family structure and demographic information relating to the families is described elsewhere. Ethical approval for the Belém Family Study was obtained originally from the local ethics committee at the Instituto Evandro Chagas (Belém, Brazil). Approval for continued use of samples from the Belém Family Study and for collection and use of the samples from Natal was granted by the local institutional review board at the Universidade Federal do Rio Grande do Norte (CEP-UFRN 94–2004), nationally by the Comissão Nacional de Ética em Pesquisa (CONEP, 11019) and by the Ministerio Ciencia e Tecnologia for approval to sample blood samples out of Brazil (portaria 617; 28 September 2005). Informed written consent for sample collection was obtained from adults and from the parents of children under 18 years old. Approval for continued use of samples from the Belém Family Study and for collection and use of the samples from Natal was granted by the local institutional review board at the Universidade Federal do Rio Grande do Norte (CEP-UFRN 94–2004), nationally by the Comissão Nacional de Ética em Pesquisa (CONEP, 11019) and by the Ministerio Ciencia e Tecnologia for approval to sample blood samples out of Brazil (portaria 617; 28 September 2005). Informed written consent for sample collection was obtained from adults and from the parents of children under 18 years old. The Brazilian populations studied are long-term residents in northeast Brazil, where transmission of the parasite is focal and transient, allowing the identification of neighborhoods with ongoing or recent transmission. These families were ascertained from medical records of the Fundação Nacional de Saúde in Rio Grande do Norte. All families were of equivalent socioeconomic status. The causative agent of visceral leishmaniasis was confirmed to be L. infantum chagasi in a subset (~10% over four stages) of individuals with visceral leishmaniasis. Further clinical, epidemiological, family structure and demographic information relating to the families is described elsewhere.

Ethical approval for the Belém Family Study was obtained originally from the local ethics committee at the Instituto Evandro Chagas (Belém, Brazil). Approval for continued use of samples from the Belém Family Study and for collection and use of the samples from Natal was granted by the local institutional review board at the Universidade Federal do Rio Grande do Norte (CEP-UFRN 94–2004), nationally by the Comissão Nacional de Ética em Pesquisa (CONEP, 11019) and by the Ministerio Ciencia e Tecnologia for approval to sample blood samples out of Brazil (portaria 617; 28 September 2005). Informed written consent for sample collection was obtained from adults and from the parents of children under 18 years old. The Brazilian populations studied are long-term residents in northeast Brazil, where transmission of the parasite is focal and transient, allowing the identification of neighborhoods with ongoing or recent transmission. These families were ascertained from medical records of the Fundação Nacional de Saúde in Rio Grande do Norte. All families were of equivalent socioeconomic status. The causative agent of visceral leishmaniasis was confirmed to be L. infantum chagasi in a subset (~10% over four stages) of individuals with visceral leishmaniasis. Further clinical, epidemiological, family structure and demographic information relating to the families is described elsewhere.

Genotyping. DNA from the Indian discovery cases and controls and the Brazilian families was genotyped at the Wellcome Trust Sanger Institute on the Illumina Human660W-Quad chip, a custom chip designed by WTCCC2 that comprises Human550 SNPs supplemented with 60,000 additional probes that were intended to allow the genotyping of common copy-number variations (CNVs) from the Structural Variation Consortium. Replication genotyping for Indian samples was carried out at the Wellcome Trust Sanger Institute using the Illumina Illumicompochip, a custom chip designed by WTCCC2 and the Immunochip Consortium that comprises 196,524 SNPs. For both chips, bead intensity data were processed and normalized for each sample in BeadStudio software (Illumina); data for successfully genotyped samples were extracted, and genotypes were called using Illuminus software (Illumina).

Quality controls. SNPs. SNPs were excluded if the Fisher information for the allele frequency was not close to unity (information of <0.98), if the minor allele frequency (MAF) was very low (defined as <0.01%), if missingness was >0.02 or for extreme departures from Hardy-Weinberg equilibrium (HWE P-value of <1 × 10−20). After applying these filters in the Indian discovery, 526,731 autosomal SNPs remained for further analysis. For the Brazilian families, these filters left 553,323 autosomal SNPs. Of the 2 discovery cohorts, 521,134 overlapping SNPs passed quality control in both cohorts. Cluster plots of the hit SNPs described in this study are shown in Supplementary Figure 3.

Samples. For quality control of samples typed on the Indian and Brazilian Illumina Human660W-Quad data, a Bayesian clustering method was used to infer and exclude outlying individuals on the basis of call rate, heterozygosities and text. Input and outputs were shared between the Indian and Brazilian samples.

Statistical analyses. We carried out association analyses using a novel variance-components method (a linear mixed model, similar to that described by Kang et al.17). The linear mixed model explicitly accounts for correlations in individuals’ phenotypes due to their relatedness through specification of the phenotypic variance-covariance matrix in terms of parameters representing underlying genetic and environmental components of variance (estimated during the model fitting procedure) and pairwise relatedness measures (kinship coefficients; estimated before model fitting on the basis of the genome-wide genotype data). In this analysis, we used a standard linear mixed model γ = XB + Q + e, where y = (y1, y2, …, yN)T is a vector of responses (coded 1/0 for case/control) on N subjects, X = (X1, X2, …, Xm)T = a vector of genotype data and e = (e1, e2, …, eN)T is a vector of random errors with mean 0 and variance-covariance matrix σe2I. β is a vector of regression coefficients (to be estimated) representing the linear effects of the predictors on the response and Q and e are random effects assigned the distributions Q ~ N(0,σQ2I) and e ~ N(0,σe2I), where σQ2 and σe2 are variance parameters (to be estimated) representing the genetic and environmental components of variance, respectively, I is the n × n identity matrix and Φ is the n × n matrix of pairwise kinship coefficients.
Although use of this linear mixed model was originally proposed for pedigrees with known relationships, recently this approach has gained popularity for use with samples of unknown or uncertain relationships, including seemingly unrelated samples that may nevertheless show distant levels of common ancestry. For this purpose, the pairwise kinship coefficients modeling either close or distant relatedness are estimated (before fitting the linear mixed model) on the basis of genome-wide genotype data instead of being fixed at known theoretical values.

Computational considerations have led to the development of several faster approximations for constructing tests of the fixed effects of interest in the linear mixed model. These approximate tests have been implemented in various software packages, including EMMAX, Tassel, FaST-LMM and GenABEL. For the analyses presented here, we used our own implementation of the linear mixed model, developed for a previous study, which has the advantage, in common with the recently developed GEMMA package, of fitting the full (rather than an approximate) model, which in principle can lead to a small increase in power, depending on the underlying true model. Our implementation also allows a transformation between the parameters of the linear model described above and the logistic model (which is the usual model for case-control data), allowing the convenient estimation of effects on the log-odds scale and the generation of resulting OR estimates.

We confirmed that the Indian discovery results did not change significantly after the inclusion of the first ten principal components in the linear mixed model (data not shown).

Combined P values across Indian (discovery and replication) and Brazilian data sets were calculated in R using an inverse-variance fixed-effects meta-analysis.

In addition, for the Brazilian family study, we verified our results using two complementary alternative methods implemented in the software packages FBAT and ROADTRIPS. Results were comparable across the different methods (data not shown).

HLA typing. The HLA-DRB1 and HLA-DQB1 typing of 100 selected Indian discovery individuals was performed using a sequencing-based method. To select these individuals, SNPs in the class II region HLA-DRA–HLA-DQB2 locus (32,514,320–32,840,188 bp on chromosome 6 of NCBI Build 36) were first pruned in PLINK to generate a set of 78 SNPs in approximate linkage equilibrium (r^2 < 0.8) and with MAF > 0.1. LD blocks across these 78 SNPs were then visualized using Haploview, and the top 59 haplotype-tagging SNPs were selected and used to generate haplotypes for all 1,866 unrelated individuals in fastPHASE. The subset of 100 individuals selected for sequence-based HLA typing represented the 135 most common 59-SNP haplotypes in the region. Automated sequencing was carried out on ABI Prism 3730 or 3730xl Genetic Analyzers (Applied Biosystems), and HLA-DRB1 analysis was carried out using ASSENV4.0.1.36 (Conexon Genomics). Where possible, individuals were assigned HLA-DRB1 or HLA-DQB1 specificities to the four-digit level, that is, to specific HLA protein encoded on each chromosome (for example, HLA-DRB1*1501, where HLA-DRB1*1501 is a specific functional HLA-DRB1 protein). Where this was not possible, individuals were assigned to a two-digit allele group; for example, HLA-DRB1*15 indicates an allele group in which all specific functional HLA-DRB1 proteins share common ancestry.

HLA-DRB1 and HLA-DQB1 typing was further performed in a subset of 71 unrelated individuals from the Brazilian families using the same sequence-based method as was used for the Indian samples.

Phasing the SNP and HLA data. SNPs in the MHC region and HLA-DRB1 and HLA-DQB1 alleles were phased using IMPUTE2 (ref. 64) and PHASE. The Indian discovery and Brazilian replication cohorts were treated separately throughout the phasing pipeline.

For the Indian discovery data, IMPUTE2 (ref. 64) was used to phase all the SNPs on chromosome 6 for 1,866 individuals, using HapMap 3 populations as the haplotype reference panel. For the Brazilian data, 5,667 SNPs in the MHC region (20,733,613–41,984,313 bp) were phased. The phasing of the Brazilian data was performed in two stages, first with an unrelated set of 498 parents and second with an unrelated set of 268 offspring, totaling 766 individuals. The parent-child trio relationships were then used to examine the accuracy of both SNP and HLA allele phasing.

Treating the IMPUTE2 SNP phasing as fixed, PHASE was then used to phase the HLA alleles typed at HLA-DRB1 and HLA-DQB1 onto SNP haplotypes in the HLA region, selecting only haplotypes phased at a probability of >0.9. PHASE was used to phase the multiallelic HLA-DRB1 and HLA-DQB1 alleles onto the known SNP haplotypes, specifying known phase at all of the SNPs and the PHASE options —MR and —d1. The HLA alleles were given positions: HLA-DRB1 was set at 32,597,662 bp between rs28756238 and rs41546317, and HLA-DQB1 was set at 32,735,635 bp between rs28724231 and rs1063355. Both HLA-DRB1 and HLA-DQB1 alleles were phased in the Indian data, and HLA-DRB1 was phased in the Brazilian data.

After quality control of the PHASE data, in the Indian discovery set, there was a total of 3,732 SNP haplotypes, 154 with HLA-DRB1 alleles and 153 with HLA-DQB1 alleles. Of these, 112 were phased to the 4-digit level for both HLA-DRB1 and HLA-DQB1. In the Brazilian data set, there was a total of 1,481 SNP haplotypes, 142 of which were phased with 2-digit HLA-DRB1 alleles.

Haplotype analyses. Haplotype associations were investigated by applying two methods: a conditional analysis and a Bayesian approach. Both methods gave the same result, with each fitting a three-haplotype model of risk.

GENECLUSTER, which adopts a Bayesian approach, was used to look for primary and secondary association signals at known and putative SNPs. This method analyzes the genealogy of the case-control sample to find evidence for causal mutation(s). It uses the genealogy of a reference haplotype panel (in this case, HapMap 3 GIH haplotypes) to approximate the genealogy of the case-control sample at positions across the region by clustering the case-control haplotypes under the leaves of the reference genealogy. Placing a disease-associated mutation on a branch of the reference genealogy segregates the haplotypes at the leaves into two different risk groups: those case-control haplotypes that fall under the disease-associated mutation (and therefore carry the mutation) and those that do not. For a given mutation in the tree, GENECLUSTER carries out a Bayesian test of association. The model can also be extended to two disease mutations in the genealogy. For these data, the genealogical tree was estimated on chromosome 6 at 32,678,817 bp. The log10 (Bayes factor) for the 2-mutation (3-haplotype) model was 7.36 compared to a log10 (Bayes factor) of 5.42 for a single-mutation (2-haplotype) model.

After the linear mixed-model conditional analysis, we phased the SNP data for all individuals in the India discovery data and found that the two SNPs associated in the conditional analysis (rs9271858 and rs9271255) tagged the haplotypes from the GENECLUSTER analysis. The first mutation perfectly correlated with the protective haplotype, and the second mutation separated the risk haplotype into two groups, which are well tagged by the two SNP intermediate and risk haplotypes (r^2 protective = 1, intermediate = 0.49, risk = 0.60; overall correlation of 0.86).

37. Singh, S.P., Reddy, D.C., Mishra, R.N. & Sundar, S. Knowledge, attitude, and practices related to Kala-azar in a rural area of Bihar state, India. Am. J. Trop. Med. Hyg. 75, 505–508 (2006).
38. Manna, M., Majumder, H.K., Sundar, S. & Bhaduri, A.N. The molecular characterization of clinical isolates from Indian Kala-azar patients by MLEE and RAPD-PCR. Med. Sci. Monit. 11, BR220–BR227 (2005).
39. Chatterjee, M., Manna, M., Bhaduri, A.N. & Sarkar, D. Recent kala-azar cases in India: isozyme profiles of Leishmania parasites. Indian J. Med. Res. 102, 165–172 (1995).
40. Thakur, C.P., Dedet, J.P., Narain, S. & Prattlong, F. Leishmania species, drug unresponsiveness and visceral leishmaniasis in Bihar, India. Trans. R. Soc. Trop. Med. Hyg. 95, 187–189 (2001).
41. Sundar, S. et al. Resistance to treatment in Kala-azar: separation of isolates from northeast India. Am. J. Trop. Med. Hyg. 65, 193–196 (1991).
42. Blackwell, J.M. et al. Immunogenetics of leishmanial and mycobacterial infections: The Belem Family Study. Phil. Trans. R. Soc. Lond. B 352, 1331–1345 (1997).
43. Jeronimo, S.M. et al. An emerging peri-urban pattern of infection with Leishmania chagasi, the protozoan causing visceral leishmaniasis in northeast Brazil. Scand. J. Infect. Dis. 36, 443–449 (2004).
44. Khalil, E.A., Zijlstra, E.E., Kager, P.A. & El Hassan, A.M. Epidemiology and clinical manifestations of Leishmania donovani infection in two villages in an endemic area in eastern Sudan. Trop. Med. Int. Health 7, 35–44 (2002).
45. Fakiola, M. Classification and regression tree and spatial analyses reveal geographic heterogeneity in genome wide linkage study of Indian visceral leishmaniasis. PLoS ONE 5, e15807 (2010).
46. Zijlstra, E.E., el Hassan, A.M., Imsam, A. & G habil, H.W. Endemic kala-azar in eastern Sudan: a longitudinal study on the incidence of clinical and subclinical
infection and post-kala-azar dermal leishmaniasis. Am. J. Trop. Med. Hyg. 51, 826–836 (1994).

47. Conrad, D.F. et al. Origins and functional impact of copy number variation in the human genome. Nature 464, 704–712 (2010).

48. Barrett, J.C. et al. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. Nat. Genet. 41, 1330–1334 (2009).

49. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).

50. Fisher, R.A. The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. 52, 399–433 (1918).

51. Henderson, C.R. Estimation of variance and covariance components. Biometrics 9, 226–252 (1953).

52. Boerwinkle, E., Chakraborty, R. & Sing, C.F. The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. Am. Hum. Genet. 50, 181–194 (1986).

53. Chen, W.M. & Abecasis, G.R. Family-based association tests for genomewide association scans. Am. J. Hum. Genet. 81, 913–926 (2007).

54. Yu, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38, 203–208 (2006).

55. Aulchenko, Y.S., de Koning, D.J. & Haley, C. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics 177, 577–585 (2007).

56. Svishcheva, G.R., Axenovich, T.I., Belorogova, N.M., van Duijn, C.M. & Aulchenko, Y.S. Rapid variance components-based method for whole-genome association analysis. Nat. Genet. 44, 1166–1170 (2012).

57. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44, 821–824 (2012).

58. Horvath, S., Wei, E., Xu, X., Palmer, L.J. & Baur, M. Family-based association test method: age of onset traits and covariates. Genet. Epidemiol. 21 (suppl. 1), S403–S408 (2001).

59. Thornton, T. & McPeek, M.S. ROADTRIPS: case-control association testing with partially or completely unknown population and pedigree structure. Am. J. Hum. Genet. 86, 172–184 (2010).

60. Sayer, D. et al. HLA-DRB1 DNA sequencing based typing: an approach suitable for high throughput typing including unrelated bone marrow registry donors. Tissue Antigens 57, 46–54 (2001).

61. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265 (2005).

62. Scheet, P. & Stephens, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78, 629–644 (2006).

63. Marsh, S.G. et al. Nomenclature for factors of the HLA system, 2010. Tissue Antigens 75, 291–455 (2010).

64. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).

65. Stephens, M., Smith, N.J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68, 978–989 (2001).

66. Su, Z., Cardin, N., Donnelly, P. & Marchini, J. A Bayesian method for detecting and characterizing allelic heterogeneity and boosting signals in genome-wide association studies. Stat. Sci. 24, 430–450 (2009).