Published in final edited form as:

Nat Genet. 2014 December ; 46(12): 1333–1336. doi:10.1038/ng.3143.

Variation at *HLA-DRB1* is associated with resistance to enteric fever

Sarah J Dunstan^{1,2,3}, Nguyen Thi Hue^{1,4}, Buhm Han^{5,6,7,8}, Zheng Li⁹, Trinh Thi Bich Tram¹, Kar Seng Sim⁹, Christopher M Parry^{1,10}, Nguyen Tran Chinh¹¹, Ha Vinh¹¹, Nguyen Phu Huong Lan¹¹, Nga Tran Vu Thieu¹, Phat Voong Vinh¹, Samir Koirala¹², Sabina Dongol¹², Amit Arjyal¹², Abhilasha Karkey¹², Olita Shilpakar¹², Christiane Dolecek^{1,2}, Jia Nee Foo⁹, Le Thi Phuong¹³, Mai Ngoc Lanh¹³, Tan Do¹⁴, Tin Aung¹⁵, Do Nu Hon¹⁴, Yik Ying Teo^{16,17}, Martin L Hibberd^{9,18}, Katherine L Anders^{1,2}, Yukinori Okada^{19,20}, Soumya Raychaudhuri^{6,7,8,21,22}, Cameron P Simmons^{1,2,23}, Stephen Baker^{1,2,18}, Paul I W de Bakker^{24,25}, Buddha Basnyat¹², Tran Tinh Hien¹, Jeremy J Farrar^{1,2}, and Chiea Chuen Khor^{9,15,17,26,27}

¹Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam ²Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, Oxford, UK ³Nossal Institute of Global Health, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia ⁴Faculty of Biology, University of Science, Vietnam National University, Ho Chi Minh City, Vietnam ⁵Asan Institute for Life Sciences, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea ⁶Department of Medicine, Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA ⁷Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA ⁸Partners Center for Personalized Genetic Medicine, Boston, Massachusetts, USA ⁹Genome Institute of Singapore, Singapore ¹⁰Liverpool School of Tropical Medicine, Liverpool, UK ¹¹Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam ¹²Oxford University Clinical Research Unit–Nepal, Patan Academy of Health Sciences, Patan Hospital, Patan, Nepal ¹³Dong Thap Provincial Hospital, Cao Lanh, Vietnam ¹⁴Vietnam National Institute of Ophthalmology, Hanoi, Vietnam ¹⁵Singapore Eye Research Institute,

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

Correspondence should be addressed to, S.J.D. (sarah.dunstan@unimelb.edu.au) or C.C.K. (khorrcc@gis.a-star.edu.sg).

URLs. R statistical software, <http://www.r-project.org/>; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html; SNP2HLA, <https://www.broadinstitute.org/mpg/snp2hla/>.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

C.C.K. and S.J.D. are the principal investigators for the study who conceived and obtained funding for the project. C.C.K. organized and supervised the GWAS and replication genotyping pipeline, devised the overall analysis plan and wrote the first draft of the manuscript with input from S.J.D. S.J.D. established the enteric fever cohorts for the discovery and replication stages of this genetics study by working with J.J.F., T.T.H., N.P.H.L., N.T.H., T.T.B.T., C.M.P., N.T.C., H.V., L.T.P., M.N.L., N.T.V.T. and P.V.V. in Vietnam and with B.B., S.K., S.D., A.A., A.K., O.S., C.D. and S.B. in Nepal to coordinate the collection of clinical samples and phenotype data. K.L.A. and C.P.S. coordinated and collected the Vietnamese control cohorts. T.D., D.N.H. and T.A. contributed data from Vietnamese replication cohorts with other diseases. Z.L., K.S.S. and J.N.F. performed genotyping and DNA quality checks on all samples. C.C.K., Y.Y.T., P.I.W.d.B., B.H., Y.O., M.L.H. and S.R. analyzed the data and performed imputation. All authors critically reviewed manuscript revisions and contributed intellectual input to the final submission.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Singapore ¹⁶Department of Statistics and Applied Probability, National University of Singapore, Singapore ¹⁷Saw Swee Hock School of Public Health, National University of Singapore, Singapore ¹⁸London School of Tropical Medicine and Hygiene, London, UK ¹⁹Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan ²⁰Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan ²¹Department of Medicine, Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, USA ²²Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK ²³Department of Microbiology and Immunology, The University of Melbourne, Melbourne, Victoria, Australia ²⁴Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, the Netherlands ²⁵Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands ²⁶Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore ²⁷Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Abstract

Enteric fever affects more than 25 million people annually and results from systemic infection with *Salmonella enterica* serovar Typhi or Paratyphi pathovars A, B or C¹. We conducted a genome-wide association study of 432 individuals with blood culture–confirmed enteric fever and 2,011 controls from Vietnam. We observed strong association at rs7765379 (odds ratio (OR) for the minor allele = 0.18, $P = 4.5 \times 10^{-10}$), a marker mapping to the HLA class II region, in proximity to *HLA-DQB1* and *HLA-DRB1*. We replicated this association in 595 enteric fever cases and 386 controls from Nepal and also in a second independent collection of 151 cases and 668 controls from Vietnam. Imputation-based fine-mapping across the extended MHC region showed that the classical *HLA-DRB1**04:05 allele (OR = 0.14, $P = 2.60 \times 10^{-11}$) could entirely explain the association at rs7765379, thus implicating *HLA-DRB1* as a major contributor to resistance against enteric fever, presumably through antigen presentation.

Enteric (typhoid) fever remains a considerable public health problem worldwide. It has been estimated that there are 26.9 million new enteric fever infections per year globally, with 200,000 deaths^{1,2}. Individuals are normally infected with the causative agents (*S. enterica* Typhi (hereafter, *S. Typhi*) and *S. enterica* Paratyphi (hereafter, *S. Paratyphi*) pathovars) after the consumption of fecally contaminated food or water. Although industrialization, improvements in sanitation and the provision of clean water have effectively mitigated the disease burden in many countries^{3,4}, enteric fever continues to be endemic in many lower-income countries. Protective vaccines against *S. Typhi* exist, but they have limited efficacy and are not suitable for young children (who represent the most at-risk group). Consequently, these vaccines are not widely deployed in the populations with the greatest need. Notably, there is currently no licensed vaccine against enteric fever caused by *S. Paratyphi* pathovars, potentially constituting a huge problem as the incidence of *S. Paratyphi* A infection is increasing in many countries across Asia^{5,6}.

The interface between the human host and the pathogen likely has a critical role in determining outcome during an enteric fever infection. The genetic variability of the most common agent of enteric fever (*S. Typhi*) together with its virulence mechanisms and epidemiology have been investigated in detail^{7,8}, whereas little is known about the human host determinants influencing susceptibility to enteric fever. There are a number of very rare genetic diseases, for example, mediated by mutations in the interleukin (IL)-12 or interferon (IFN)- γ pathways, that result in hypersusceptibility to non-typhoidal *Salmonella* and other intracellular bacteria⁹, but the corresponding mutations have not, as yet, been associated with enteric fever susceptibility. Candidate gene studies^{10–13} have been compromised by limited sample size and consequent lack of power to identify robust associations¹⁴.

To further evaluate the role of host genetics in susceptibility to enteric fever, we performed a genome-wide association study (GWAS) on 432 individuals with blood culture-confirmed *S. Typhi* infection and 2,011 controls from Vietnam. These same cases and controls also underwent a second round of genotyping using the Illumina HumanExome BeadChip¹⁵ (Online Methods). An overall association analysis across all SNP markers tested showed little evidence of global test statistic inflation ($\lambda_{GC} = 1.05$; Supplementary Fig. 1), suggesting that cryptic population stratification between enteric fever cases and controls was minimal. This finding was indeed supported by principal-components analysis of all subjects (Supplementary Figs. 2 and 3, Supplementary Tables 1 and 2, and Supplementary Note). The strongest evidence for association mapped to SNP rs7765379 at the class II human leukocyte antigen (HLA) region near the *HLA-DQB1* and *HLA-DRB1* genes (*HLA-DRB1* mRNA, [NM_002124](#); *HLA-DQB1* mRNA, [NM_002123](#)). The minor allele of rs7765379 was under-represented in enteric fever cases (frequency = 1.04%) with respect to the control subjects (frequency = 5.5%; OR = 0.18, $P = 4.5 \times 10^{-10}$; Table 1). We also identified a second SNP marker exceeding genome-wide significance at rs6841458, which mapped near *GUCY1A3* on chromosome 4 (OR = 1.55, $P = 3.3 \times 10^{-8}$). No other genomic region exhibited significant evidence of association ($P > 5 \times 10^{-7}$; Supplementary Figs. 4 and 5).

We attempted to replicate the associations observed at rs7765379 in the major histocompatibility complex (MHC) region and the *GUCY1A3* variant rs6841458 in an independent collection of 595 enteric fever cases and 386 geographically matched controls from Nepal (Online Methods). We observed significant evidence of replication for rs7765379: the minor allele was again under-represented in the cases (frequency = 1.36%) in comparison to the controls (frequency = 3.9%; OR = 0.34, $P = 7.0 \times 10^{-4}$), a finding consistent with the discovery results (Table 1). However, the association with the *GUCY1A3* rs6841458 marker on chromosome 4 (OR = 0.97, $P = 0.71$) did not replicate (Supplementary Table 3). As the Nepal replication collection comprised groups with diverse ancestry for cases and controls, which could confound the association analysis, we reanalyzed the association between rs7765379 and enteric fever with stratification for self-reported ancestry using previously described methods¹⁶. The association observed remained robust and consistent across all ancestry groups where the marker was polymorphic (Supplementary Table 4). We also did not observe any difference in the magnitude of the association when the analysis was stratified for *S. Typhi* or *S. Paratyphi A* (Online Methods and Supplementary Table 5).

To provide additional confidence in the observed association at rs7765379, we performed further replication in samples originating from 151 individuals with enteric fever and 668 controls independently enrolled in Vietnam. We again observed significant association at rs7765379 (OR = 0.18, $P = 1.3 \times 10^{-4}$; Table 1). Meta-analysis of the data from the discovery and both replication collections showed a genome-wide significant association at rs7765379: the minor allele was associated with 4.55-fold increased resistance to enteric fever (OR_{meta} = 0.22, $P_{\text{meta}} = 2.29 \times 10^{-13}$; Table 1).

We next explored the possibility that the association at rs7765379 might be driven by functional variation within the classical HLA genes, which might not have been directly captured by the GWAS. To this end, we performed statistical imputation of classical HLA alleles and their corresponding amino acid polymorphisms using a combined reference panel of East Asian and European individuals with previously described methodologies^{17–19} (Online Methods). We observed the strongest association with the classical *HLA-DRB1*04:05* allele (OR = 0.14, $P = 2.60 \times 10^{-11}$; Fig. 1), which was unsurprisingly in tight linkage disequilibrium (LD) with rs7765379 ($r^2 = 0.83$). The association at this four-digit allele was more significant than any associations at SNPs or amino acid polymorphisms (Supplementary Table 6). We also performed omnibus tests for association at each amino acid position encoded by eight HLA genes to examine whether multiallelic polymorphism at any amino acid position could explain the association better²⁰. None of the omnibus test P values were more significant than the P value for *HLA-DRB1*04:05*. After we adjusted for the effect of *HLA-DRB1*04:05*, no other variant remained significant ($P > 1 \times 10^{-6}$)²¹, including rs7765379 (adjusted $P = 0.62$; Supplementary Table 6). In contrast, after we adjusted for the effect of rs7765379, there was still a nominally significant association for the *HLA-DRB1*04:05* allele (adjusted $P = 0.016$), making it more likely that the *HLA-DRB1*04:05* allele is a driving factor for this observed HLA association signal. To verify the accuracy of our imputation, we performed HLA typing for 140 individuals from the Vietnam GWAS discovery collection. The allelic concordance between the typed and imputed alleles for *HLA-DRB1*04:05* was 99.3%, thus reflecting the robustness of our imputation. Taking the directly typed HLA alleles as the ‘gold standard’, the sensitivity of the imputation was measured at 100% while the specificity of the imputation was 98.8% (refs. 17,22). Although our imputation accuracy was very high, we note that imputation errors normally result in loss of statistical power²³ and thus reduce the degree of statistical significance observed.

The *Salmonellae* are a genus of facultative intracellular pathogens with the ability to survive and replicate in an array of phagocytic and non-phagocytic cells by residing in a specialized cellular compartment, named the *Salmonella*-containing vacuole (SCV)^{24,25}. The intracellular survival and replication of *Salmonella* is critical for virulence and essential for systemic disease²⁶. *Salmonella* inject bacterial-derived effector proteins into host cells, such as macrophages, to influence host cell activation and promote bacterial survival. Additionally, *Salmonella* use dendritic cells for dissemination and restrict the ability of these cells to process antigens and present peptides²⁷. Enteric fever is the most severe manifestation of *Salmonella* infection in humans. It is fatal in 10–25% of patients without antimicrobial treatment, and surviving patients have longer fever clearance times and more severe clinical complications than those that received antimicrobial therapy²⁸. This

variability in disease severity and outcome is presumably dependent on external and environmental factors, such as exposure and the virulence of the pathogen, and host factors, such as variation in innate and adaptive immune responses during exposure and acute infection. The intimate and streamlined relationship between the human host and the invading *Salmonella*, in which the bacteria attempt to establish infection in the face of an immune response from the host²⁹, represents a fine balance in terms of pathogenicity.

HLA-DRB1 encodes the β chain of HLA-DR and belongs to the HLA class II molecules, which present antigen to CD4⁺ T lymphocytes. Variation in HLA-DR might affect the capacity of the HLA class II molecules to present antigen, which might in turn result in a differential immune response. Here we show that *HLA-DRB1*04:05* is strongly protective against enteric fever. Presence of the *HLA-DRB1*04:05* allele might tip the poise of pathogenicity between the invading pathogen and host immune response toward the human host, resulting in as much as fivefold greater protection from enteric fever. The minor allele of rs7765379 that we observed to confer resistance to enteric fever has previously been shown to strongly associate with increased susceptibility to Crohn's disease and rheumatoid arthritis^{30,31}. This observation is consistent with the possible counterbalancing effects between determinants of infectious disease susceptibility and potential autoimmunity in the host^{32,33} and perhaps more so with genetic determinants exerting as strong an effect as HLA molecules. The unequivocal detection of positive signals of evolutionary selection in the HLA region can be complicated by the presence of balancing selection occurring at this locus³⁴, so further efforts will be needed to explore the contribution of these opposing forces in depth.

Thus far, there have been few published GWAS of infectious disease susceptibility in comparison to other genetic traits and non-communicable diseases^{14,17,35–40}. In published studies of infectious disease susceptibility, the HLA region has been positively identified as encoding host factors for HIV control (but not for susceptibility to HIV infection itself)²⁹, leprosy, visceral leishmaniasis and chronic hepatitis B virus infection. Conversely, there was no signal of association implicating the HLA region for other infectious diseases such as meningococcal sepsis, dengue, tuberculosis and malaria. Before the introduction of technological advances permitting GWAS, candidate gene studies interrogating the involvement of the broad MHC region found variable evidence of associations in individuals with enteric fever at the class II and III regions without having the statistical power to be conclusive^{11–13}. Our previous study on 111 enteric fever cases and 77 controls in Vietnam showed evidence of association implicating *HLA-DRB1*04* in resistance against enteric fever ($P = 0.02$), but this finding was not robust to correction for multiple testing¹¹. We were able to confirm this association with *HLA-DRB1*04* in our current Vietnamese GWAS discovery collection ($OR = 0.33$, $P = 1.5 \times 10^{-9}$; Supplementary Table 6). This association is specifically driven by *HLA-DRB1*04:05* ($OR = 0.14$, $P = 2.6 \times 10^{-11}$; Supplementary Table 6), as performing the association analysis conditioning for the allele dosage of *HLA-DRB1*04:05* extinguished all evidence of association seen at *HLA-DRB1*04* ($P_{\text{conditioned}} = 0.056$).

To our knowledge, this study is the first large-scale, unbiased effort to investigate host genetic factors associated with enteric fever, providing strong evidence for a role for *HLA-*

*DRB1*04:05* as a protective factor against enteric fever. Our data document one of the more substantial effect sizes (averaging nearly ~5-fold greater resistance against disease for the minor allele) reported thus far for any genetic locus determining human susceptibility to an infectious disease. Further attention should now be focused on determining the precise nature of HLA class II variation in determining susceptibility to enteric fever and other invasive bacterial pathogens, thus offering the potential to contribute to improvements in the rational design of vaccines for enteric fever and other serious bacterial infections.

ONLINE METHODS

Enteric fever cases and population controls

DNA samples from individuals with enteric fever (Vietnam, $n = 432$ for the GWAS discovery stage and $n = 151$ for replication; Nepal, $n = 595$ for replication) were collected as part of larger epidemiological or clinical studies. Individuals with enteric fever were defined as children or adults with clinical signs and symptoms of enteric fever with culture-confirmed *S. Typhi* or *S. Paratyphi A* in their blood or bone marrow. In Vietnam, >99% of cases were infected with *S. Typhi*, whereas, in Nepal, 68% (402/595) of cases were colonized with *S. Typhi* and 32% (188/595) were colonized with *S. Paratyphi A*. Blood samples for research were collected at the time of study enrollment, and demographic and clinical information was recorded into case report forms for the duration of the hospital stay or study. The clinical studies in Vietnam were performed between 1992 and 2002 at the Hospital for Tropical Diseases in Ho Chi Minh City, the Dong Nai Pediatric Centre in Dong Nai Province and the Dong Thap Provincial Hospital in Dong Thap Province, Vietnam, and have been described previously^{44–48}. In Nepal, the clinical studies were performed between 2005 and 2014 at Patan Hospital, Kathmandu, and have been described previously^{49–51}.

The population control individuals from Vietnam and Kathmandu were cord blood controls. In Vietnam, cord blood control samples ($n = 2,011$ for the GWAS discovery stage and $n = 668$ for the replication stage) were collected from babies born between 2003 and 2012 at Hung Vuong Obstetric Hospital in Ho Chi Minh City and Dong Thap Hospital in Dong Thap province. In Nepal, cord blood samples ($n = 386$) were collected at Patan Hospital, Kathmandu, between 2008 and 2012. All cases and controls were unrelated, and the majority of Vietnamese individuals were of Vietnamese Kinh ancestry (>98%), as assessed by questionnaire. There was more ancestry diversity within the Nepalese cases and controls, where individuals self-reported as 1 of 15 different ancestry groups. Genomic DNA was extracted from blood samples using the Qiagen blood midi or maxi kits (Qiagen) or the Nucleon BACC genomic DNA extraction kits (GE Healthcare). The treating physician was responsible for obtaining informed consent from the patients or their parent or guardian for clinical study participation and from the baby's mother for population controls. Informed consent was obtained from all subjects involved in this research. Ethical approvals were granted by the scientific and ethical committees at the Hospital for Tropical Diseases, the Dong Thap Provincial Hospital and the Health Services of Dong Thap Province in Vietnam. The clinical trials in Nepal were approved by the Nepal Health Research Council, Kathmandu. Protocols were also approved by the Oxford Tropical Research Ethics Committee, Oxford University, UK.

Genotyping

Genome-wide genotyping for the Vietnamese enteric fever cases and cord blood controls was performed using the Illumina OmniExpress BeadChip and Illumina 660W BeadChip, which have been very well described⁴⁰. These samples also underwent exome-wide genotyping using the Illumina Human Exome BeadChip, which assays for 247,870 SNP markers enriched in the coding exome, following the manufacturer's specifications. Replication genotyping for the independently enrolled cases and controls from Nepal and Vietnam was performed with TaqMan allelic discrimination assays (Applied Biosystems) using routine laboratory techniques described elsewhere⁴⁰.

Statistical analysis

Stringent quality control filters were used to remove poorly performing samples and SNP markers in both the discovery (GWAS and exome chip) and replication (*de novo* genotyping) phases. SNPs with a call rate of <95%, non-polymorphic SNP markers and those showing significant deviation from Hardy-Weinberg equilibrium (P value for deviation of $<1 \times 10^{-6}$) were removed from further statistical analysis. For the exome chip that contained nonsynonymous variants, SNPs that were not monomorphic (corresponding to at least one heterozygous carrier individual) were included for downstream analysis. For the GWAS chip, SNPs with a MAF of >1% were included for downstream analysis. A total of 709,725 markers passed quality control filters in this manner (642,445 from the GWAS chip and 67,280 from the exome chip).

Similarly stringent quality control criteria were also applied on a per-sample basis, and samples with an overall genotyping success rate of <95% were removed from further analysis. The remaining samples were then subjected to biological relationship verification using the principle of variability in allele sharing according to the degree of relationship. Identity-by-state (IBS) information was derived using the PLINK software package⁵². For those pairs of individuals who showed evidence of relatedness reflecting pairs of first-degree relatives (for example, a parent and offspring, full siblings or monozygotic twins), we removed the sample with the lower call rate before performing principal-component analysis. Principal-component analysis was undertaken to account for spurious associations resulting from ancestral differences for individual SNPs, and principal-component plots were generated using the R statistical program. For the discovery stage, all enteric fever cases had perfectly matched controls, as visualized spatially on principal-component analysis (Supplementary Fig. 2), using previously reported criteria⁵³.

For both the discovery and replication stages, analysis of association contrasting SNP genotypes between enteric fever cases and controls was carried out using logistic regression, with adjustment for the first ten principal components of genetic stratification. This test models for a trend-per-copy effect of the minor allele on disease risk. The association P value was obtained by a log-likelihood ratio test comparing the likelihood of the null model against the likelihood of the fitted model. We verified non-departure from the additive genetic model by additionally performing the conventional trend test. Genotyping cluster plots for genetic markers surpassing genome-wide significance at the discovery stage were visually checked and verified to be of good quality (Supplementary Figs. 6 and 7).

Genome-wide imputation was performed using IMPUTE2 software (see URLs) with the 1000 Genomes Project reference panel⁵⁴.

HLA imputation and association analysis

We imputed classical HLA alleles and corresponding amino acid polymorphisms using a combined Asian and European reference panel with high-density SNP genotypes and 4-digit classical HLA allele genotypes (530 Asian individuals and 5,225 European individuals) using SNP2HLA software (see URLs), as described elsewhere^{18,19,22}. We extracted SNP genotypes located in the broad MHC region (chr. 6: 29–34 Mb) to impute 92 two-digit classical alleles, 152 four-digit classical alleles and 763 amino acid polymorphisms of the 8 class I and class II HLA genes (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1* and *HLA-DPB1*).

For the association analysis, we used a logistic regression model assuming additive effects of allele dosages on the log-odds scale. We defined HLA variants to include biallelic SNPs in the MHC region, two-digit and four-digit classical HLA alleles, biallelic amino acid polymorphisms and multiallelic amino acid positions. To account for potential population substructure, we included the top ten principal components as covariates. For HLA variants with m alleles ($m = 2$ for biallelic variants and $m > 2$ for multiallelic variants), we included $m - 1$ alleles as independent variables in the regression model, excluding the most frequent allele as a reference. This resulted in the following logistic regression model:

$$\log(\text{odds}) = \beta_0 + \sum_{j=1}^{m-1} \beta_{1,j} x_j + \sum_{l=1}^L \beta_{2,l} y_l + \varepsilon$$

where β_0 is the logistic regression intercept, $\beta_{1,j}$ is the additive effects of the dosage of allele j for the variant x_j , L is the number of principal components enrolled in the analysis, y_l is the l th principal component and $\beta_{2,l}$ is the effect of y_l . An omnibus P value was obtained by a log-likelihood ratio test comparing the likelihood of the null model against the likelihood of the fitted model. We further assessed the significance of the improvement in fit by calculating the deviance ($= -2$ times the log likelihood), which follows a χ^2 distribution with $m - 1$ degree(s) of freedom.

Statistical power calculation for the replication stage

Statistical power calculations were performed using previously described approaches for complex disease genetic studies⁵⁵. To observe a reasonably convincing replication (for example, at $P < 0.001$), our approach comprising a first-stage replication using 595 enteric fever cases and 386 controls from Nepal and a second-stage replication using 151 enteric fever cases and 668 controls from Vietnam would be able to reliably detect ($>90\%$ power) modest genetic effects ($OR > 1.5$ – 1.6) at MAFs in excess of 20%. For strong genetic effects with $OR > 3$, the replication stages were sufficiently powered to detect these even at MAFs as low as 5% (Supplementary Table 7). Thus, for potential true positive associations to achieve genome-wide significance in the combined discovery and replication data sets, we

brought forward sentinel SNPs showing $P < 5 \times 10^{-7}$ from the discovery stage for testing in the replication collections⁵⁶.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Wellcome Trust, UK as part of their Major Overseas Program in Viet Nam (089276/Z/09/Z) and by the Biomedical Research Council, Agency for Science, Technology and Research, Singapore. S.B. is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z). P.I.W.d.B. acknowledges support from the Netherlands Organization for Scientific Research (Vernieuwingsimpuls VIDI Award NWO project number 016.126.354).

References

1. Buckle GC, Walker CL, Black RE. Typhoid fever and paratyphoid fever: systematic review to estimate global morbidity and mortality for 2010. *J. Glob. Health.* 2012; 2:010401. [PubMed: 23198130]
2. Crump JA, Luby SP, Mintz ED. The global burden of typhoid fever. *Bull. World Health Organ.* 2004; 82:346–353. [PubMed: 15298225]
3. Cvjetanovi B, Grab B, Uemura K. Epidemiological model of typhoid fever and its use in the planning and evaluation of antityphoid immunization and sanitation programmes. *Bull. World Health Organ.* 1971; 45:53–75. [PubMed: 5316853]
4. Crump JA, Mintz ED. Global trends in typhoid and paratyphoid fever. *Clin. Infect. Dis.* 2010; 50:241–246. [PubMed: 20014951]
5. McGregor AC, Waddington CS, Pollard AJ. Prospects for prevention of *Salmonella* infection in children through vaccination. *Curr. Opin. Infect. Dis.* 2013; 26:254–262. [PubMed: 23591641]
6. Karki S, Shakya P, Cheng AC, Dumre SP, Leder K. Trends of etiology and drug resistance in enteric fever in the last two decades in Nepal: a systematic review and meta-analysis. *Clin. Infect. Dis.* 2013; 57:e167–e176. [PubMed: 23985342]
7. Kingsley RA, et al. Genome and transcriptome adaptation accompanying emergence of the definitive type 2 host-restricted *Salmonella enterica* serovar Typhimurium pathovar. *MBio.* 2013; 4:e00565–13. [PubMed: 23982073]
8. Wong VK, et al. Characterization of the yehUT two-component regulatory system of *Salmonella enterica* serovar Typhi and Typhimurium. *PLoS ONE.* 2013; 8:e84567. [PubMed: 24386394]
9. de Jong R, et al. Severe mycobacterial and *Salmonella* infections in interleukin-12 receptor-deficient patients. *Science.* 1998; 280:1435–1438. [PubMed: 9603733]
10. Dunstan SJ, et al. Typhoid fever and genetic polymorphisms at the natural resistance-associated macrophage protein 1. *J. Infect. Dis.* 2001; 183:1156–1160. [PubMed: 11237848]
11. Dunstan SJ, et al. Genes of the class II and class III major histocompatibility complex are associated with typhoid fever in Vietnam. *J. Infect. Dis.* 2001; 183:261–268. [PubMed: 11120931]
12. Dunstan SJ, et al. A *TNF* region haplotype offers protection from typhoid fever in Vietnamese patients. *Hum. Genet.* 2007; 122:51–61. [PubMed: 17503085]
13. Dharmana E, et al. *HLA-DRB1*12* is associated with protection against complicated typhoid fever, independent of tumour necrosis factor α . *Eur. J. Immunogenet.* 2002; 29:297–300. [PubMed: 12121274]
14. Fakiola M, et al. Common variants in the *HLA-DRB1–HLA-DQA1* HLA class II region are associated with susceptibility to visceral leishmaniasis. *Nat. Genet.* 2013; 45:208–213. [PubMed: 23291585]
15. Huyghe JR, et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat. Genet.* 2013; 45:197–201. [PubMed: 23263489]

16. Helgason A, et al. Refining the impact of *TCF7L2* gene variants on type 2 diabetes and adaptive evolution. *Nat. Genet.* 2007; 39:218–225. [PubMed: 17206141]
17. Pereyra F, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science.* 2010; 330:1551–1557. [PubMed: 21051598]
18. Jia X, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE.* 2013; 8:e64683. [PubMed: 23762245]
19. Pillai NE, et al. Predicting HLA alleles from high-resolution SNP data in three Southeast Asian populations. *Hum. Mol. Genet.* 2014; 23:4443–4451. [PubMed: 24698974]
20. Raychaudhuri S, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* 2012; 44:291–296. [PubMed: 22286218]
21. Gockel I, et al. Common variants in the *HLA-DQ* region confer susceptibility to idiopathic achalasia. *Nat. Genet.* 2014; 46:901–904. [PubMed: 24997987]
22. Okada Y, et al. Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. *Hum. Mol. Genet.* 2014 Jul 28.
23. Huang L, Wang C, Rosenberg NA. The relationship between imputation error and statistical power in genetic association studies in diverse populations. *Am. J. Hum. Genet.* 2009; 85:692–698. [PubMed: 19853241]
24. Jantsch J, Chikkaballi D, Hensel M. Cellular aspects of immunity to intracellular *Salmonella enterica*. *Immunol. Rev.* 2011; 240:185–195. [PubMed: 21349094]
25. Malik-Kale P, et al. *Salmonella*—at home in the host cell. *Front. Microbiol.* 2011; 2:125. [PubMed: 21687432]
26. Haraga A, Ohlson MB, Miller SI. *Salmonellae* interplay with host cells. *Nat. Rev. Microbiol.* 2008; 6:53–66. [PubMed: 18026123]
27. Cheminay C, Mohlenbrink A, Hensel M. Intracellular *Salmonella* inhibit antigen presentation by dendritic cells. *J. Immunol.* 2005; 174:2892–2899. [PubMed: 15728500]
28. van den Bergh ET, Gasem MH, Keuter M, Dolmans MV. Outcome in three groups of patients with typhoid fever in Indonesia between 1948 and 1990. *Trop. Med. Int. Health.* 1999; 4:211–215. [PubMed: 10223217]
29. Khor CC, Hibberd ML. Host-pathogen interactions revealed by human genome-wide surveys. *Trends Genet.* 2012; 28:233–243. [PubMed: 22445588]
30. Freudenberg J, et al. Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci. *Arthritis Rheum.* 2011; 63:884–893. [PubMed: 21452313]
31. Yamazaki K, et al. A genome-wide association study identifies 2 susceptibility loci for Crohn's disease in a Japanese population. *Gastroenterology.* 2013; 144:781–788. [PubMed: 23266558]
32. Sfriso P, et al. Infections and autoimmunity: the multifaceted relationship. *J. Leukoc. Biol.* 2010; 87:385–395. [PubMed: 20015961]
33. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. *Nat. Rev. Genet.* 2014; 15:379–393. [PubMed: 24776769]
34. de Bakker PI, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* 2006; 38:1166–1172. [PubMed: 16998491]
35. de Bakker PI, Telenti A. Infectious diseases not immune to genome-wide association. *Nat. Genet.* 2010; 42:731–732. [PubMed: 20802473]
36. Davila S, et al. Genome-wide association study identifies variants in the *CFH* region associated with host susceptibility to meningococcal disease. *Nat. Genet.* 2010; 42:772–776. [PubMed: 20694013]
37. Thye T, et al. Common variants at 11p13 are associated with susceptibility to tuberculosis. *Nat. Genet.* 2012; 44:257–259. [PubMed: 22306650]
38. Thye T, et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat. Genet.* 2010; 42:739–741. [PubMed: 20694014]
39. Fellay J, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science.* 2007; 317:944–947. [PubMed: 17641165]

40. Khor CC, et al. Genome-wide association study identifies susceptibility loci for dengue shock syndrome at *MICB* and *PLCE1*. *Nat. Genet.* 2011; 43:1139–1141. [PubMed: 22001756]
41. Dean M, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CCR5* structural gene. *Science.* 1996; 273:1856–1862. [PubMed: 8791590]
42. Huang Y, et al. The role of a mutant *CCR5* allele in HIV-1 transmission and disease progression. *Nat. Med.* 1996; 2:1240–1243. [PubMed: 8898752]
43. Modiano D, et al. Haemoglobin C protects against clinical *Plasmodium falciparum* malaria. *Nature.* 2001; 414:305–308. [PubMed: 11713529]
44. Cao XT, et al. A comparative study of ofloxacin and cefixime for treatment of typhoid fever in children. The Dong Nai Pediatric Center Typhoid Study Group. *Pediatr. Infect. Dis. J.* 1999; 18:245–248. [PubMed: 10093945]
45. Chinh NT, et al. A randomized controlled comparison of azithromycin and ofloxacin for treatment of multidrug-resistant or nalidixic acid-resistant enteric fever. *Antimicrob. Agents Chemother.* 2000; 44:1855–1859. [PubMed: 10858343]
46. Luxemburger C, et al. Risk factors for typhoid fever in the Mekong delta, southern Viet Nam: a case-control study. *Trans. R. Soc. Trop. Med. Hyg.* 2001; 95:19–23. [PubMed: 11280056]
47. Vinh H, et al. Double blind comparison of ibuprofen and paracetamol for adjunctive treatment of uncomplicated typhoid fever. *Pediatr. Infect. Dis. J.* 2004; 23:226–230. [PubMed: 15014297]
48. House D, et al. Cytokine release by lipopolysaccharide-stimulated whole blood from patients with typhoid fever. *J. Infect. Dis.* 2002; 186:240–245. [PubMed: 12134261]
49. Arjyal A, et al. Gatifloxacin versus chloramphenicol for uncomplicated enteric fever: an open-label, randomised, controlled trial. *Lancet Infect. Dis.* 2011; 11:445–454. [PubMed: 21531174]
50. Pandit A, et al. An open randomized comparison of gatifloxacin versus cefixime for the treatment of uncomplicated enteric fever. *PLoS ONE.* 2007; 2:e542. [PubMed: 17593957]
51. Koirala S, et al. Gatifloxacin versus ofloxacin for the treatment of uncomplicated enteric fever in Nepal: an open-label, randomized, controlled trial. *PLoS Negl. Trop. Dis.* 2013; 7:e2523. [PubMed: 24282626]
52. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007; 81:559–575. [PubMed: 17701901]
53. Tall AR. CETP inhibitors to increase HDL cholesterol levels. *N. Engl. J. Med.* 2007; 356:1364–1366. [PubMed: 17387130]
54. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature.* 2010; 467:1061–1073. [PubMed: 20981092]
55. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics.* 2003; 19:149–150. [PubMed: 12499305]
56. Purdue MP, et al. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat. Genet.* 2011; 43:60–65. [PubMed: 21131975]

Figure 1.

Association results for enteric fever within the broad HLA region. (a) We observed the most significant association at *HLA-DRB1*04:05* (OR = 0.14, $P = 2.6 \times 10^{-11}$). Squares denote directly genotyped SNPs, and diamonds denote imputed markers. *HLA-DRB1*04:05* was in tight LD with the most significant directly genotyped SNP, rs7765379 ($r^2 = 0.83$). (b) When conditioning the analysis on *HLA-DRB1*04:05*, we no longer observed any convincing evidence of association ($P > 4 \times 10^{-6}$).

Table 1
Association results for SNP rs7765379 in the class II HLA region with enteric fever in Vietnamese and Nepalese subjects

SNP (gene)	Reference allele	Effect allele	Genotype count ^d in cases (freq.)	Genotype count ^d in controls (freq.)	P value ^b	OR	Collection
rs7765379 (<i>HLA-DRB1-HLA-DQB1</i>)	A	C	0/9/423 (0.0104)	6/208/1,797 (0.055)	4.5×10^{-10} (2.53×10^{-8})	0.18	Discovery (Vietnam) ^c
	A	C	0/16/574 ^d (0.0136)	3/24/359 (0.039)	0.00070 (0.00062)	0.34	Replication (Nepal) ^e
	A	C	0/3/148 (0.010)	1/69/598 (0.053)	0.00013 (0.00098)	0.18	Replication (Vietnam) ^f
					2.29×10^{-13}	0.22 (0.15–0.34)	Meta-analysis

^a Genotype counts in cases and controls are given in the following format: homozygotes for the minor (effect) allele/heterozygous/wild type. The frequency of each allele is shown in parentheses.

^b P-values were derived from the likelihood ratio test using logistic regression (P-values in parentheses were derived from the trend test).

^c Discovery with 432 enteric fever cases and 2,011 controls.

^d A total of 590 of 595 cases were successfully genotyped.

^e Replication with 595 enteric fever cases and 386 controls.

^f Replication with 151 enteric fever cases and 668 controls.

The controls in the Nepalese replication show mild deviation from Hardy-Weinberg equilibrium ($0.015 < P < 0.001$), which is consistent with observations of SNPs strongly associated with resistance against infectious diseases^{41–43}.