Genome-wide association study identifies susceptibility loci for dengue shock syndrome at *MICB* and *PLCE1*

Chiea Chuen Khor[1–3], Tran Nguyen Bich Chau[4], Junxiong Pang[1,5], Sonia Davila[2,6], Hoang Truong Long[1,4], Rick T H Ong[2], Sarah J Dunstan[4,7], Bridget Wills[4,7], Jeremy Farrar[4,7], Ta Van Tram[8], Tran Thi Gan[8], Nguyen Thi Nguyet Binh[9], Le Trung Tri[9], Le Bich Lien[10], Nguyen Minh Tuan[10], Nguyen Thi Hong Tham[11], Mai Ngoc Lanh[11], Nguyen Minh Nguyet[4], Nguyen Trong Hieu[12], Nguyen VanVinh Chau[13], Tran Thi Thuy[14], Dennis E K Tan[6], Anavaj Sakuntabhai[15,16], Yik-Ying Teo[5,6,17], Martin L Hibberd[1,5,18] & Cameron P Simmons[4,7,18]

Hypovolemic shock (dengue shock syndrome (DSS)) is the most common life-threatening complication of dengue. We conducted a genome-wide association study of 2,008 pediatric cases treated for DSS and 2,018 controls from Vietnam. Replication of the most significantly associated markers was carried out in an independent Vietnamese sample of 1,737 cases and 2,934 controls. SNPs at two loci showed genome-wide significant association with DSS. We identified a susceptibility locus at *MICB* (major histocompatibility complex (MHC) class I polypeptide-related sequence B), which was within the broad MHC region on chromosome 6 but outside the class I and class II HLA loci (rs3132468, $P_{meta} = 4.41 \times 10^{-11}$, per-allele odds ratio (OR) = 1.34 (95% confidence interval: 1.23–1.46)). We identified associated variants within *PLCE1* (phospholipase C, epsilon 1) on chromosome 10 (rs3765524, $P_{meta} = 3.08 \times 10^{-10}$, per-allele OR = 0.80 (95% confidence interval: 0.75–0.86)). We identify two loci associated with susceptibility to DSS in people with dengue, suggesting possible mechanisms for this severe complication of dengue.

Dengue is an acute systemic viral infection caused by one of four serotypes of dengue virus, and globally it is the commonest mosquito-borne infection after malaria[1]. The burden of dengue is growing, with an estimated 100 million infections now occurring annually and with 2.5 billion people living in areas at risk of transmission. A wide variety of disease manifestations is seen, ranging from subclinical infection to severe and fatal disease. Severe dengue in children is characterized by an increase in vascular permeability that leads to life-threatening hypovolemic shock (DSS). This is often accompanied by thrombocytopenia and hemostatic dysfunction, which may result in severe bleeding. Children are at greatest risk of developing DSS, but with careful supportive care the case fatality rate is less than 1% (ref. 2). In southern Vietnam, serological studies have estimated that 85% of the population is exposed to dengue virus infection by the end of childhood (15 years old)[3], whereas DSS is estimated to occur in less than 1% of exposed individuals[2] (see population controls in Online Methods). Epidemiological studies have suggested a genetic basis in the host for susceptibility to severe dengue, and various candidate gene studies of modest sample sizes have investigated this possibility[4–8].

To estimate the genetic contribution underlying severe dengue, we genotyped 2,118 DNA samples from Vietnamese children with established or incipient DSS and 2,089 cord blood controls in a genome-wide association study (GWAS). After exclusion of samples for discrepancies between clinical and genetically inferred gender, for relatedness or for per-sample call rates of less than 95% (**Supplementary Fig. 1a**), there were 2,008 DSS cases and 2,018 controls available for analysis. The clinical and virological characteristics of the case population are described in **Supplementary Table 1**. A total of 657,366 SNPs were initially included within the Illumina 660W Beadchip used for genome-wide genotyping. After various stringent quality-control exclusions (**Supplementary Fig. 1b**), a total of 481,342 SNPs were retained for downstream association analysis.

We conducted the routine GWAS statistical tests (Online Methods) and identified strong evidence of disease association at two distinct loci (**Fig. 1**): *MICB* on chromosome 6 and *PLCE1* on chromosome 10,

[1]Infectious Diseases, Genome Institute of Singapore, Singapore. [2]Genome Institute of Singapore–National University of Singapore Centre for Molecular Epidemiology, National University of Singapore, Singapore. [3]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. [4]Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam. [5]Department of Epidemiology and Public Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. [6]Human Genetics, Genome Institute of Singapore, Singapore. [7]Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, Oxford, UK. [8]Tien Giang Hospital, My Tho City, Tien Giang Province, Vietnam. [9]Sa Dec Hospital, Sa Dec Town, Dong Thap Province, Vietnam. [10]Department of Dengue Hemorrhagic Fever, Children's Hospital No. 1, Ho Chi Minh City, Vietnam. [11]Dong Thap Hospital, Cao Lanh City, Dong Thap Province, Vietnam. [12]Hung Vuong Hospital, District 5, Ho Chi Minh City, Vietnam. [13]Hospital for Tropical Diseases, District 5, Ho Chi Minh City, Vietnam. [14]Department of Infectious Diseases, Children's Hospital No. 2, District 1, Ho Chi Minh City, Vietnam. [15]Institut Pasteur, Laboratoire de la Génétique de la réponse aux infections chez l'homme, Paris, France. [16]Department of Medicine, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand. [17]Department of Statistics and Applied Probability, Faculty of Science, National University of Singapore, Singapore. [18]These authors contributed equally to this work. Correspondence should be addressed to M.L.H. (hibberdml@gis.a-star.edu.sg) or C.P.S. (csimmons@oucru.org).

Received 24 May; accepted 9 September; published online 16 October 2011; doi:10.1038/ng.960

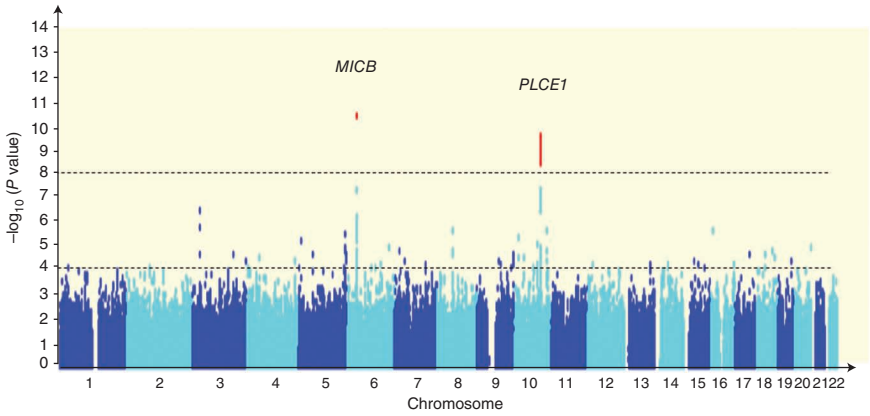

Figure 1 Manhattan plot showing directly genotyped SNPs plotted according to chromosomal location (*x* axis), with $-\log_{10}$ *P* values (*y* axis) derived from the 1-degree-of-freedom score test. The lower horizontal dashed line indicates the threshold for bringing SNPs forward to the replication stage ($P < 1 \times 10^{-4}$). SNPs surpassing $P < 1 \times 10^{-8}$ (upper horizontal dashed line) on combined analysis of both GWAS and replication data are shown in red, and gene names are given for these loci. SNPs in *MICB* and *PLCE1* have significant associations.

which were both represented by SNPs that were close to the formal threshold for genome-wide significance ($P = 5.38 \times 10^{-8}$ for rs3132468 at the *MICB* locus and $P = 5.84 \times 10^{-8}$ for rs3740360 at the *PLCE1* locus) (**Table 1**). Together with the SNPs at *MICB* and *PLCE1*, a total of 85 SNPs exceeded $P < 1 \times 10^{-4}$ on single SNP analysis (**Supplementary Table 2**). We were able to design assays for 72 out of these 85 SNPs using the Sequenom Mass-Array platform. The remaining 13 SNPs in the broad MHC region were refractory to assay design, thus necessitating the design of Applied Biosystems TaqMan assays for the sentinel SNP at *MICB* (rs3132468) and rs3134899 (also within *MICB*; GWAS $P = 1.03 \times 10^{-4}$, OR = 1.31). We then genotyped these 74 SNPs (72 non-MHC SNPs and two SNPs within *MICB*) in a replication sample of 1,824 DSS cases and 3,019 controls. We applied the same GWAS quality-control filters for the replication set: five SNPs had poor genotyping clusters and were excluded from analysis (**Supplementary Table 2**), and 132 samples (87 cases and 85 controls) had per-sample call rates of less than 95%; these were excluded from further analysis. This left 69 SNPs to be analyzed in 1,737 cases and 2,934 controls for the replication stage. In keeping with the GWAS observations, the strongest evidence of association was observed with SNPs at *MICB* (rs3132468, $P_{repl} = 9.32 \times 10^{-5}$ and rs3134899, $P_{repl} = 0.0082$) and *PLCE1* (three SNPs with P_{repl} ranging from 5.23×10^{-4} to 1.6×10^{-4}, **Table 1**). Using inverse-variance weights, data from both the GWAS and replication cohorts ($N = 3{,}745$ DSS cases and $N = 4{,}952$ controls) were combined in a formal meta-analysis, and this identified strong evidence of association with rs3132468 at *MICB* ($P = 4.41 \times 10^{-11}$; per-allele OR = 1.34 (95% confidence interval: 1.23–1.46)) and seven SNPs at *PLCE1* ($4.18 \times 10^{-9} \leq P \leq 3.08 \times 10^{-10}$; $0.75 \leq \text{OR} \leq 0.87$, **Table 1**). To aid in refining the original signal of association, we performed imputation analysis at regions flanking both loci (30–32 Mb on chromosome 6 and 95.5–96.5 Mb on chromosome 10). This did not identify signals of association beyond those from the directly genotyped SNPs. The associations observed at *MICB* and *PLCE1* were not specific to any dengue virus serotype on subgroup analysis of viral serotype, nor were they associated with the degree of thrombocytopenia or the degree of clinical shock (data not shown).

Found within the broad MHC locus, *MICB* lies just outside both the type I and type II human leukocyte antigen (HLA) regions, ~140,000 base pairs centromeric to the nearest class I gene (*HLA-B*) and slightly less than 1 million base pairs away from the nearest class II gene (*HLA-DRA*). Apart from the peak signal at rs3132468 that was observed directly within *MICB*, 12 other SNPs in this region also showed association signals exceeding $P < 1 \times 10^{-4}$ on single-SNP

Table 1 Association analysis between dengue shock syndrome and SNP genotypes at *MICB* and *PLCE1*

SNP (alleles)	Chromosome (position)	Candidate gene	Stage	MAF cases	MAF controls	OR	*P*	OR_{meta} (95% CI)	P_{meta}
rs3132468 (C/T)	6 (31583465)	*MICB*	GWAS	0.176	0.132	1.41	5.39×10^{-8}	1.34 (1.23–1.46)	4.41×10^{-11}
			Replication	0.163	0.134	1.27	9.32×10^{-5}		
rs3134899 (G/A)	6 (31581265)	*MICB*	GWAS	0.130	0.102	1.31	1.09×10^{-4}	1.26 (1.14–1.38)	4.08×10^{-6}
			Replication	0.114	0.096	1.20	0.0082		
rs3765524* (T/C)	10 (96048288)	*PLCE1*	GWAS	0.249	0.300	0.77	2.68×10^{-7}	0.80 (0.75–0.86)	3.08×10^{-10}
			Replication	0.265	0.302	0.83	1.60×10^{-4}		
rs2274223* (G/A)	10 (96056331)	*PLCE1*	GWAS	0.250	0.303	0.77	1.19×10^{-7}	0.81 (0.75–0.86)	6.89×10^{-10}
			Replication	0.267	0.300	0.85	5.23×10^{-4}		
rs3740360 (C/A)	10 (96015481)	*PLCE1*	GWAS	0.219	0.271	0.75	5.84×10^{-8}	0.80 (0.75–0.86)	1.15×10^{-9}
			Replication	0.242	0.273	0.85	0.0012		
rs12263737 (A/G)	10 (96034903)	*PLCE1*	GWAS	0.250	0.301	0.77	3.73×10^{-7}	0.81 (0.75–0.87)	1.22×10^{-9}
			Replication	0.266	0.300	0.84	3.95×10^{-4}		
rs11187842 (T/C)	10 (96042501)	*PLCE1*	GWAS	0.219	0.269	0.76	1.19×10^{-7}	0.80 (0.75–0.86)	1.78×10^{-9}
			Replication	0.240	0.271	0.85	0.0011		
rs753724 (T/G)	10 (96041407)	*PLCE1*	GWAS	0.219	0.269	0.76	1.28×10^{-7}	0.81 (0.75–0.86)	2.27×10^{-9}
			Replication	0.242	0.272	0.85	0.0012		
rs3781264 (G/A)	10 (96060365)	*PLCE1*	GWAS	0.229	0.278	0.77	3.43×10^{-7}	0.81 (0.76–0.87)	4.18×10^{-9}
			Replication	0.250	0.280	0.85	0.0011		

MAF cases, minor allele frequency in DSS cases; MAF controls, minor allele frequency in the controls; OR, odds of DSS per copy of the minor allele; *P*, *P* value using the one degree of freedom score test; OR_{meta}, odds ratio for the combined GWAS and replication cohorts; P_{meta}, *P* value for the combined GWAS and replication cohorts; 95% CI, 95% confidence interval for the OR; GWAS, sample size of 2,008 DSS cases and 2,018 cord blood controls; replication, sample size of 1,737 DSS cases and 2,934 cord blood controls. All SNPs are intronic to their respective genes except those marked with an asterisk (*), which are exonic.

analysis. We thus performed conditional analysis to assess the independence of the association observed at *MICB* rs3132468 from that of the nearby genes. Although the most significant SNP from the GWAS (rs3132468) could account for the majority of the association signal across the locus, we observed residual signals of association ($0.0003 < P < 0.05$) with SNPs near the vicinity of *HLA-B* and *HLA-C* as well as other neighboring genes (**Supplementary Fig. 2a**). These residual associations indicate that definitive identification of *MICB* as a gene associated with DSS could be complicated by its location within the broad MHC region, which is known for its extensive linkage disequilibrium spanning multiple genes (**Supplementary Fig. 2b**). This precludes definitive identification of the causative gene without extensive further fine-mapping and resequencing. With regard to *PLCE1* on chromosome 10, association analysis conditioning for the lead SNP (rs3743060, directly genotyped) did not identify any secondary signals of association (**Supplementary Fig. 3a**), which suggests that the lead SNP—or any of its close correlates in complete linkage disequilibrium with it and confined within their distinct genomic region (**Supplementary Fig. 3b**)—best explains the association signal at the locus. We did not observe any evidence of epistasis between SNPs at *MICB* and *PLCE1* ($P = 0.11$).

MICB seems to be a promising candidate on the basis of the present strength of the statistical associations observed in the chromosome 6 hit region. *MICB* encodes an inducible activating ligand for the NKG2D type II receptor on natural killer and $CD8^+$ T cells[9,10]. Ligation of NKG2D by MICB stimulates antiviral effector functions in natural killer cells, including cytokine expression and the cytolytic response[11]. We have previously reported that *MICB*, together with other genes associated with natural killer cell activation, are highly expressed in the leukocytes of people with acute dengue[12]. We therefore propose that the association between the *MICB* rs3132468 genotype and susceptibility to severe dengue might reflect altered or dysfunctional natural killer and/or $CD8^+$ T cell activation early in infection that results in a higher viral burden *in vivo*, which is a recognized factor in clinical outcome[13,14]. The recent finding that a SNP near the closely related *MICA* gene (rs2596542) is associated with hepatitis C virus–induced hepatocellular carcinoma is suggestive of a pivotal role for MIC proteins in the pathogenesis of these Flaviviridae infections[15].

Mutations within *PLCE1* are associated with nephrotic syndrome[16]. Nephrotic syndrome is a kidney disorder in which dysfunction of the glomerular basement membrane results in proteinuria and hypoproteinemia that, when severe, leads to reduced vascular oncotic pressure and edema. These elements of nephrotic syndrome have striking similarities with severe dengue and suggest a key role for *PLCE1* in maintaining normal vascular endothelial cell barrier function. In summary, our study identifies common variants in *MICB* and *PLCE1* that are associated with susceptibility to severe dengue.

URLs World Health Organization Guidelines for Diagnosis, Treatment, Prevention and Control of Dengue, http://whqlibdoc.who.int/publications/2009/9789241547871_eng.pdf; Sequenom, http://www.sequenom.com/; Applied Biosystems, http://www.appliedbiosystems.com/absite/us/en/home.html; R Project, http://www.r-project.org/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes *MICB* mRNA, NM_005931; *PLCE1* transcript variant 1 mRNA, NM_016341; *PLCE1* transcript variant 2 mRNA, NM_001165979.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This work was supported by the Wellcome Trust, United Kingdom (grant 088791/A/09/Z and 084368/Z/07/Z), and the Agency for Science, Technology and Research, Singapore.

AUTHOR CONTRIBUTIONS

M.L.H. and C.P.S. are the study principal investigators 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 M.L.H., C.P.S. and S.D. T.N.B.C. is the lead coordinator of clinical samples and phenotypes for both the discovery and replication stages. J.P. and D.E.K.T. performed genotyping and quality checks on all samples. C.C.K., S.D., R.T.H.O. and Y.-Y.T. analyzed the data. H.T.L., S.J.D., B.W., J.F., T.V.T., T.T.G., N.T.N.B., L.T.T., L.B.L., N.M.T., N.T.H.T., M.N.L., N.M.N., N.T.H., N.V.V.C., T.T.T. and A.S. coordinated and contributed patient and database phenotype collections as lead investigators for their respective sample collections. D.E.K.T. and J.P. performed genotyping and DNA quality checks. All authors critically reviewed manuscript revisions and contributed intellectual input to the final submission.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Gubler, D.J. Dengue and dengue hemorrhagic fever. *Clin. Microbiol. Rev.* **11**, 480–496 (1998).
2. Anders, K.L. *et al.* Epidemiological factors associated with dengue shock syndrome and mortality in hospitalized dengue patients in Ho Chi Minh City, Vietnam. *Am. J. Trop. Med. Hyg.* **84**, 127–134 (2011).
3. Thai, K.T. *et al.* Seroprevalence of dengue antibodies, annual incidence and risk factors among children in southern Vietnam. *Trop. Med. Int. Health* **10**, 379–386 (2005).
4. Loke, H. *et al.* Susceptibility to dengue hemorrhagic fever in vietnam: evidence of an association with variation in the vitamin D receptor and FC gamma receptor IIA genes. *Am. J. Trop. Med. Hyg.* **67**, 102–106 (2002).
5. Loke, H. *et al.* Strong HLA class I–restricted T cell responses in dengue hemorrhagic fever: a double-edged sword? *J. Infect. Dis.* **184**, 1369–1373 (2001).
6. Stephens, H.A. *et al.* HLA-A and -B allele associations with secondary dengue virus infections correlate with disease severity and the infecting viral serotype in ethnic Thais. *Tissue Antigens* **60**, 309–318 (2002).
7. Sakuntabhai, A. *et al.* A variant in the CD209 promoter is associated with severity of dengue disease. *Nat. Genet.* **37**, 507–513 (2005).
8. Vejbaesya, S. *et al. TNF* and *LTA* gene, allele and extended HLA haplotype associations with severe dengue virus infection in ethnic Thais. *J. Infect. Dis.* **199**, 1442–1448 (2009).
9. Steinle, A. *et al.* Interactions of human NKG2D with its ligands MICA, MICB, and homologs of the mouse RAE-1 protein family. *Immunogenetics* **53**, 279–287 (2001).
10. González, S., Lopez-Soto, A., Suarez-Alvarez, B., Lopez-Vazquez, A. & Lopez-Larrea, C. NKG2D ligands: key targets of the immune response. *Trends Immunol.* **29**, 397–403 (2008).
11. Garrity, D., Call, M.E., Feng, J. & Wucherpfennig, K.W. The activating NKG2D receptor assembles in the membrane with two signaling dimers into a hexameric structure. *Proc. Natl. Acad. Sci. USA* **102**, 7641–7646 (2005).
12. Hoang, L.T. *et al.* The early whole-blood transcriptional signature of dengue virus and features associated with progression to dengue shock syndrome in Vietnamese children and young adults. *J. Virol.* **84**, 12982–12994 (2010).
13. Libraty, D.H. *et al.* Differing influences of virus burden and immune activation on disease severity in secondary dengue-3 virus infections. *J. Infect. Dis.* **185**, 1213–1221 (2002).
14. Vaughn, D.W. *et al.* Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity. *J. Infect. Dis.* **181**, 2–9 (2000).
15. Kumar, V. *et al.* Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. *Nat. Genet.* **43**, 455–458 (2011).
16. Hinkes, B. *et al.* Positional cloning uncovers mutations in *PLCE1* responsible for a nephrotic syndrome variant that may be reversible. *Nat. Genet.* **38**, 1397–1405 (2006).

ONLINE METHODS

Patient enrollment and diagnosis. Blood samples for genotyping were collected from patients enrolled into one of two research studies of children with dengue. In both studies, children were eligible if they were ≤15 years of age and had clinical signs, symptoms and hematological findings that led to a clinical diagnosis of incipient or established DSS, as defined by World Health Organization criteria (see URLs). All patients were resuscitated with bolus intravenous fluid therapy (≥15 ml per kg body weight in the first hour). Summary laboratory and clinical findings were recorded into case record forms during the inpatient period until the patient was discharged from hospital, died or was transferred to another hospital. Blood samples for research and diagnostic tests were collected at the time of enrollment and again before patient discharge from hospital. The first study enrolled patients in the pediatric intensive care unit of the Hospital for Tropical Diseases (Ho Chi Minh City, Vietnam) between 2001 and 2009. The second study enrolled patients in high dependency rooms or the intensive care departments of Children's Hospital No. 1 and Children's Hospital No. 2 (Ho Chi Minh City, Vietnam), Tien Giang Provincial Hospital (My Tho City, Vietnam), Dong Thap Provincial Hospital (Cao Lanh City, Vietnam) and Sa Dec Hospital (Sa Dec Town, Vietnam) between 2008 and 2010. The parent or guardian of each participant gave written informed consent to participate. The Scientific and Ethical Committees of each study site approved the study protocols, as did the Oxford University Tropical Research Ethical Committee.

The GWAS was performed on DNA samples (n = 1,039) from patients enrolled between 2001 and 2009 at the Hospital for Tropical Diseases and from patients (n = 969) enrolled at the other five participating hospitals during 2008 only. The replication study was performed in patients (n = 1,737) enrolled between 2009 and 2010 at Children's Hospital No. 1, Children's Hospital No. 2, Tien Giang Provincial Hospital, Dong Thap Provincial Hospital and Sa Dec Hospital. All patients represented in the GWAS and replication phases had laboratory evidence of dengue, as shown by reverse transcription PCR detection of viral RNA in plasma collected at the time of enrollment and/or by serological detection of dengue-virus–reactive IgM or IgG in single or paired plasma specimens.

Cord blood DNA samples. Blood from the cord of newborn infants was collected in one of two prospective studies. The first study was conducted at Hung Vuong Hospital (Ho Chi Minh City, Vietnam) between 2004 and 2006. The second study was conducted at Hung Vuong Hospital and Dong Thap Hospital between 2009 and 2010. All participants gave written informed consent to participate. The Scientific and Ethical Committees of each study site approved the study protocols, as did the Oxford University Tropical Research Ethical Committee. DNA was extracted from cord blood using Nucleon BACC Genomic DNA Extraction Kits (GE Healthcare, USA).

The use of population controls. The number of potentially misclassified cord blood controls in the GWAS and replication stages was estimated to be 11 (out of 2,018) in the GWAS stage and 15 (out of 2,934) in the replication stage, based on the following three assumptions: that all individuals in a given birth cohort will experience two sequential infections by different serotypes during their lifetime[17]; that only up to 25% of these infections are clinically apparent[18–23]; and that 2% of clinically apparent secondary infections develop DSS. These assumptions estimate a lifetime population risk of DSS to be 0.5%. This is consistent with estimates of the prevalence of DSS cases expected over the first 15 years in a given birth cohort under the assumption that the incidence of DSS is constant (DSS incidence in southern Vietnam in 2009 was 26.59/100,000; based on statistics obtained from the Dengue Control program, Ministry of Health Vietnam, 2010). Under this assumption, we would expect 0.4% of a birth cohort to experience DSS before the age of 15 years.

Genotyping. Cases and controls were randomized on plates and were genotyped with Illumina Human 660W Quad BeadChips following manufacturer instructions. The successful use of this chip has been previously documented[24]. For the replication stage, 72 of the selected SNPs that were not on the broad MHC region were genotyped with the Sequenom (see URLs) MassArray primer extension iPLEX system. *MICB* rs3132468 and rs3134899 were genotyped using the Applied Biosystems (see URLs) TaqMan platform.

Statistical analysis. Stringent quality control filters were applied to remove poorly performing SNPs and samples using tools implemented in PLINK (version 1.7)[25]. The quality control criteria were as follows: SNPs that had genotypes with more than 5% missing, showed gross departure from Hardy–Weinberg equilibrium (a departure of $P < 10^{-7}$) or had a minor allele frequency below 1% were excluded from downstream analysis. For sample quality control, samples with an overall genotyping call rate of <95% were excluded from analysis. The remaining samples were then subjected to biological relationship verification by using the principle of variability in allele sharing according to the degree of relationship. Identity-by-state information was derived using PLINK[25]. For those pairs of individuals who showed evidence of cryptic relatedness (possibly owing to duplicated or biologically related samples), 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 of individual SNPs[26], and principal component plots were performed using the R statistical program package (R Project, see URLs).

For both the GWAS and replication stages, analysis of association with DSS was carried out using a 1 degree of freedom score-based test. This test models for a trend per copy of the minor allele on disease risk and has been extensively described[27,28]. It has the best statistical power to detect association for complex traits across a wide range of alternative hypotheses, with the exception of those involving rare recessive variants[29]. The threshold for significant independent replication was set at $P < 0.05$ in the combined replication data sets[30].

Meta-analysis was conducted using inverse variance weights for each cohort, which calculates an overall Z statistic, corresponding P value and accompanying ORs for each SNP analyzed[30,31]. Genotyping clusters were directly visualized for the 85 SNPs exceeding $P < 1 \times 10^{-4}$ and confirmed to be of good quality before inclusion for statistical analysis. **Supplementary Figure 4** shows the cluster plots for *MICB* rs3132468 and *PLCE1* rs3740360, the two SNPs that showed $P < 10^{-7}$ in the GWAS stage.

Analysis of linkage disequilibrium was performed using Haploview[32].

Statistical findings after routine GWAS analysis. Analysis of genetic ancestry using principal components revealed no significant population substructure between the DSS cases and controls (**Supplementary Fig. 5a**). As the individual principal components were nonsignificant (Bonferroni corrected $P > 0.05$, **Supplementary Table 3**) when tested as continuous covariates using logistic regression, we did not adjust for them in subsequent association analysis.

Single SNP analysis was performed using logistic regression assuming additive genetic effects relating genotype dosage (scoring for 0, 1 or 2 copies of the minor allele) to DSS. A quantile–quantile plot of the single SNP analysis showed a clear excess of extreme P values compared to the null distribution (**Supplementary Fig. 5b**). As this excess was observed against a background of minimal genome-wide inflation of test statistics ($\lambda gc = 1.024$), it excludes the possibility of substantial population substructure and differential genotyping call rate between cases and controls as a reason for the excess. Instead, this suggests that at least some of these extreme P values ($P < 1 \times 10^{-4}$) may represent true genetic associations with DSS.

17. Chau, T.N. *et al.* Dengue virus infections and maternal antibody decay in a prospective birth cohort study of Vietnamese infants. *J. Infect. Dis.* **200**, 1893–1900 (2009).
18. Tien, N.T. *et al.* A prospective cohort study of dengue infection in schoolchildren in Long Xuyen, Viet Nam. *Trans. R. Soc. Trop. Med. Hyg.* **104**, 592–600 (2010).
19. Balmaseda, A. *et al.* Trends in patterns of dengue transmission over 4 years in a pediatric cohort study in Nicaragua. *J. Infect. Dis.* **201**, 5–14 (2010).
20. Porter, K.R. *et al.* Epidemiology of dengue and dengue hemorrhagic fever in a cohort of adults living in Bandung, West Java, Indonesia. *Am. J. Trop. Med. Hyg.* **72**, 60–66 (2005).
21. Endy, T.P. *et al.* Epidemiology of inapparent and symptomatic acute dengue virus infection: a prospective study of primary school children in Kamphaeng Phet, Thailand. *Am. J. Epidemiol.* **156**, 40–51 (2002).
22. Mammen, M.P. *et al.* Spatial and temporal clustering of dengue virus transmission in Thai villages. *PLoS Med.* **5**, e205 (2008).
23. Endy, T.P. *et al.* Determinants of inapparent and symptomatic dengue infection in a prospective study of primary school children in Kamphaeng Phet, Thailand. *PLoS Negl. Trop. Dis.* **5**, e975 (2011).
24. Mells, G.F. *et al.* Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. *Nat. Genet.* **43**, 329–332 (2011).

doi:10.1038/ng.960

25. 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).
26. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
27. Purdue, M.P. *et al.* Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat. Genet.* **43**, 60–65 (2011).
28. Thomas, G. *et al.* A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (*RAD51L1*). *Nat. Genet.* **41**, 579–584 (2009).
29. Lettre, G., Lange, C. & Hirschhorn, J.N. Genetic model testing and statistical power in population-based association studies of quantitative traits. *Genet. Epidemiol.* **31**, 358–362 (2007).
30. McGovern, D.P. *et al.* Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat. Genet.* **42**, 332–337 (2010).
31. Asano, K. *et al.* A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. *Nat. Genet.* **41**, 1325–1329 (2009).
32. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).