



HAL
open science

Population genetics-informed meta-analysis in seven genes associated with risk to dengue fever disease.

Marisa Oliveira, Diana P Saraiva, Bruno Cavadas, Verónica Fernandes, Nicole Pedro, Isabelle Casademont, Fanny Koeth, Farida Alshamali, Nourdin Harich, Lotfi Cherni, et al.

► To cite this version:

Marisa Oliveira, Diana P Saraiva, Bruno Cavadas, Verónica Fernandes, Nicole Pedro, et al.. Population genetics-informed meta-analysis in seven genes associated with risk to dengue fever disease.. *Infection, Genetics and Evolution*, 2018, 62, pp.60-72. 10.1016/j.meegid.2018.04.018 . hal-01873838

HAL Id: hal-01873838

<https://hal.science/hal-01873838>

Submitted on 19 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Population genetics-informed meta-analysis in seven genes associated with risk to dengue fever disease

Marisa Oliveira^{1,2,3,4}, Diana P. Saraiva^{1,2,4}, Bruno Cavadas^{1,2,3}, Verónica Fernandes^{1,2}, Nicole Pedro^{1,2}, Isabelle Casademont^{4,5}, Fanny Koeth^{4,5}, Farida Alshamali⁶, Nourdin Harich⁷, Lotfi Cherni^{8,9}, Beatriz Sierra¹⁰, Maria G. Guzman¹⁰, Anavaj Sakuntabhai^{4,5,11} and Luisa Pereira^{1,2,12}

¹ i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, 4200-135 Porto, Portugal

² Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), 4200-135 Porto, Portugal

³ Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, 4050-313 Porto, Portugal

⁴ Institut Pasteur, Functional Genetics of Infectious Diseases Unit, 75724 Paris Cedex 15, France

⁵ Pasteur Kyoto International Joint Research Unit for Integrative Vaccinomics, Kyoto, Japan

⁶ General Department of Forensic Sciences and Criminology, Dubai Police General Headquarters, PO Box 1493, Dubai, United Arab Emirates

⁷ Laboratoire d'Anthropogenetique, Department de Biologie, Universite Chouaib Doukkali, El Jadida 24000, Morocco

⁸ Laboratory of Genetics, Immunology and Human Pathology, Faculté de Sciences de Tunis, Université de Tunis El Manar, Tunis 2092, Tunisia

⁹ Tunis and High Institute of Biotechnology, University of Monastir, 5000 Monastir, Tunisia

¹⁰ Virology Department, PAHO/WHO Collaborating Center for the Study of Dengue and its Vector, Pedro Kourí Institute of Tropical Medicine (IPK), 601 Havana, Cuba

¹¹ CNRS, Unité de Recherche Associée 3012, Paris, France

¹² Faculdade de Medicina da Universidade do Porto, Portugal

Short title: Meta-analysis in seven candidate genes for dengue disease

Corresponding author: Luisa Pereira (luisap@ipatimup.pt)

Abstract (maximum 300 words)

Population genetics theory predicted that rare frequent markers would be the main contributors for heritability of complex diseases, but meta-analyses of genome-wide association studies are revealing otherwise common markers, present in all population groups, as the identified candidate genes. In this work we applied a population-genetics informed meta-analysis to 10 markers located in seven genes said to be associated with dengue fever disease. Seven markers (in *PLCE1*, *CD32*, *CD209*, *OAS1* and *OAS3* genes) have high-frequency and the other three (in *MICB* and *TNFA* genes) have intermediate frequency. Most of these markers have high discriminatory power between population groups, but their frequencies follow the rules of genetic drift, and seem to not have been under strong selective pressure. There was a good agreement in directional consistency across trans-ethnic association signals, in East Asian and Latin American cohorts, with heterogeneity generated by randomness between studies and especially by low sample sizes. This led to confirm the following significant associations: with DF, odds ratio of 0.67 for *TNFA*-rs1800629-A; with DHF, 0.82 for *CD32*-rs1801274-G; with DSS, 0.55 for *OAS3*-rs2285933-G, 0.80 for *PLCE1*-rs2274223-G and 1.32 for *MICB*-rs3132468-C. The overall genetic risks confirmed sub-Saharan African populations and descendants as the best protected against the severer forms of the disease, while Southeast and Northeast Asians are the least protected ones. European and close neighbours are the best protected against dengue fever, while, again, Southeast and Northeast Asians are the least protected ones. These risk scores provide important predictive information for the largely naïve European and North American regions, as well as for Africa where misdiagnosis with other hemorrhagic diseases is of concern.

Keywords: Dengue disease; meta-analysis; East Asian and Latin American cohorts; worldwide diversity; global risk

Abbreviations:

DENV – dengue virus; DF – Dengue Fever; DHF – Dengue Hemorrhagic Fever; DSS – Dengue Shock Syndrome; EHH – Extended Haplotype Homozygosity; GWAS - genome-wide association studies; MAF – minimum allele frequency; SNP – Single Nucleotide Polymorphism; WHO - World Health Organization

Declarations of interest: none

Highlights (3 to 5 bullet points - maximum 85 characters, including spaces, per bullet point)

- The 10 dengue fever-associated markers present high- and intermediate-frequencies
- There was directional consistency across trans-ethnic association signals
- Significant associations were phenotype-specific, and mainly for DSS
- Sub-Saharan African populations are the best protected against DSS
- Southeast and Northeast Asians are the least protected against DSS and DF.

1. Introduction

Dengue fever (DF) is amongst the 17 neglected diseases prioritized by the World Health Organization (WHO) in the year 2012 [1]. The list will for sure enlarge as new epidemic agents emerge, such as chikungunya and Zika viruses [2]. Less than 10% of the global health stipends are destined to the treatment of these diseases, which infect around 90% of sick individuals in the world [3], contributing immensely to the global health burden.

Dengue is an acute systemic disease caused by four virus types (DENV1-4) which are transmitted to humans by an *Aedes* mosquito vector (mainly *Ae. aegypti* and *Ae. albopictus*). Most DENV infections are asymptomatic, but in some individuals, it can lead to a wide range of clinical symptoms, from mild fever to life threatening dengue hemorrhagic fever and dengue shock syndrome [4]. Usually, but not exclusively, the severe disease appears when the patient is infected by a heterologous DENV, in a non-completely understood cross-reaction with antibodies against the previous DENV type [5]. Control efforts have failed in counterbalance the effects of human international travel and global warming, which are introducing new vectors and pathogens into novel geographic areas [6]. *Ae. aegypti* is well adapted to urban environments, breeding in any container of water. This species was able to re-establish in countries of South and Central America, from where it had been eliminated in the mid-20th century, and it also dispersed throughout Southeast Asia during and after World War 2 [6]. The geographical range of the secondary dengue vector, *Ae. albopictus*, has also expanded substantially over the past 30 years, being the major potential vector of DENV in Europe [7]. This increased movement of people, vector and virus will also increase the co-circulation of serotypes, leading to an increase in the risk of sequential infections and severe disease.

In this high dispersion scenario, it is relevant to estimate the worldwide risk of dengue fever. Bhatt et al. [8] undertook an exhaustive assembly of known records of dengue occurrence worldwide (8,309 geo-located records), and used a formal modelling framework to map the global distribution of dengue risk based on evidence. Dengue was predicted to be ubiquitous throughout the tropics, with the highest risk zones in the Americas (14%) and Asia (70%; India alone contributed 34%), varying locally by the influence of rainfall, temperature and the degree of urbanization. Global values of 390 million (95% credible interval [284–528]) dengue infections per year, of which 96 million [67–136] are symptomatic, were estimated. These values

suppress more than thrice the dengue burden estimates by WHO, of 50–100 million infections per year [9]. Predicted risk in Africa, although unevenly distributed, was much more widespread than previously suggested, 16 [11–22] million infections, or 16% of the global total, nearly equivalent to that of the Americas. This disparity seems to support the fact that dengue disease can be masked by symptomatically similar illnesses in Africa [10]. Oceania contributed less than 0.2% of global apparent infections. The estimated 294 [217–392] million unapparent worldwide infections, although having no immediate implications for clinical management, is a huge potential viral reservoir, as asymptomatic individuals can also transmit the virus [11], having implications in designing vaccination campaigns.

A very important factor to consider in the worldwide risk evaluation to dengue disease is the host genetic component. A predictive map based on genetic risk information could provide useful insights into the still largely naïve regions of Europe and North America, and regions where registries/diagnosis mix-up predominates as in Africa. *Homo sapiens* is a young species dating around 200,000 years [12,13], with a single unique origin in Africa. The structure of the modern human population was initiated at 60,000 years ago, in the out-of-Africa event when a small group of Eastern Africans migrated towards the Arabian Peninsula/Near East, and thence, first to Asia (arriving in Australia at around 50,000 years ago) and then (from 45,000 years ago) to Europe [14]. The bottleneck in the origin of all Asians and Europeans explains their high homozygosity when compared with the high heterozygosity observed in Africans, so that the former have proportionally more deleterious genetic variation than the latter [15]. The out-of-Africa event was probably fuelled by the first noticeable population expansion [16], but continuous expansion of the human population begun in the Holocene, around 12,000 years ago, when climatic conditions became similar to present-day, also promoting the cultural Neolithic revolution of agriculture and domestication [17]. Population expansions lead to increased amounts of rare (low MAF, minimum allele frequency) and population-restricted variants. Thus, 73% of protein-coding variants and 86% of predicted-deleterious variants in humans arose only in the past 5,000-10,000 years, and are rare and population specific [18]. A considerable lower proportion of variants are shared between continents, presenting higher MAF (>1%), and being older [19]. Through genetic drift, variants will be randomly lost or fixed in the population along time, or, if they confer susceptibility/resistance to infectious diseases, they can be under the effect of selection [20]. Negative selection reduces frequencies of SNPs, which are maintained as

rare variants, while positive selection increases MAF of protective alleles, generating heterogeneity in selected polymorphisms among populations along time, as pathogens emerge and become important agents in specific geographical regions. The factors contributing to population divergence, in terms of (1) variants (especially the rare ones), (2) MAF of common variants, (3) linkage disequilibrium (LD) patterns (tag SNPs vary in LD with the causative SNP), (4) biological adaptations, (5) phenotypic prevalence, and (6) effective sizes, have important implications in the power and the reproducibility of association studies.

Association studies in dengue disease began by pinpointing potential candidate genes involved in the immune response to viral infection in Asian and Latin American populations [21]. One of the first identified candidates was DC-SIGN (dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin 1) protein, coded by *CD209* gene. DC-SIGN is an essential molecule for the interaction of virus particles with a cellular receptor responsible for its internalization in cells otherwise refractory to infection by the four DENV serotypes [22]. In accordance, cell surface DC-SIGN expression correlates with DENV infection rates in vitro, while antibodies against DC-SIGN can block infection [23,24]. At the genetic level, the G allele at position 336 (rs4804803) of DC-SIGN was associated with a dominant protection against DF (odds ratio, OR=4.90; $p=2 \times 10^{-6}$) but not against DHF, while among individuals with dengue, genotypes GG and GA strongly increased the risk of contracting DHF versus DF (OR=5.84; $p=1.4 \times 10^{-7}$) in three Thai dengue cohorts [25]. Further functional studies showed that dendritic cells from AG-heterozygous have a significantly higher cell surface DC-SIGN expression than AA-homozygous and a higher production of TNF- α , IL-12p40, and IP-10 in response to DENV infection [25]. However, against expectations, DENV replication was significantly lower in AG individuals, probably due to the enhanced production of IP-10 [25]. Replication was positive in Taiwanese cohorts [26], for GG/GA genotypes when comparing DHF with DF, other febrile diseases and controls ($p=0.003$; 3×10^{-5} and 0.001), with ORs in the range 2-5. Replication failed in India ([27]; although other markers around this SNP presented significant association values) and Mexico [28,29]. More difficult to explain was the closeness to significant association of GG genotype in Rio de Janeiro, Brazil, but in the other direction, being protective against severe dengue [30] (another Brazilian cohort from Salvador da Bahia showed no association [31]).

Another set of genes associated with dengue, and for which there is already functional evidence, is the oligoadenylate synthetase family (*OAS1*, *OAS3* and *OAS2* genes, in this order in chromosome 12), which consists in interferon-induced cytoplasmic double-stranded RNA (dsRNA) sensors. Upon activation by dsRNA, OAS converts ATP to oligoadenylate, which in turn activates the latent ribonuclease RNaseL that catalyses the degradation of viral and host RNA, thus reducing protein synthesis in the cell [32]. DENV infection leads to an upregulation of OAS expression in endothelium HUVEC cells [33], and the stable silencing of RNaseL in A549 lung cells increases DENV titers by 10- to 40-fold in agreement with its antiviral activity [34]. Some *OAS* polymorphisms (*OAS1* - rs1131454 and rs10774671, *OAS3* - rs2285932 and rs2072136, and *OAS2* - rs15895 and rs1732778) were correlated with clinical severity of dengue in India [35], when associated in haplotypes (*OAS2* G-G all patients vs. controls, p-corrected=0.012, OR=1.73 95% CI 1.16–2.59; *OAS3-OAS2* C-G-A-G all patients vs. controls, p-corrected =0.0486, OR=0.09, 95% CI 0.00–0.64). Even more interesting, a recent study in Thai dengue cohorts showed that the influence of the *OAS* polymorphism is probably dengue virus strain-dependent: *OAS3*-G allele (Ser381Arg; rs2285933) conferred protection against shock (OR=0.37; p<0.001) when the infection was caused by DENV2 (OR=0.13; p=0.007) but not when due to DENV1 [36]. Also, by conducting a multivariate analysis, there was a serotype-specific effect of the *OAS1*-G splicing variant (rs10774671) on the risk of plasma leakage (p=0.019) in DENV-2 infections (OR=3.17; 95% CI, 1.31–7.68) but not in DENV-1 infections (OR=0.81; 95% CI, .37–1.75), while no statistical significant associations were found for the *OAS3*-rs1859330 missense marker (Arg18Lys) in the Thai population.

CD32 or *FCGR2A* (Fc fragment of IgC receptor IIa) gene, together with other Fc γ receptors expressed in monocytes, macrophages and dendritic cells, mediate antibody dependent enhancement (ADE). This phenomenon occurs during secondary infection with a heterologous serotype, when antibodies formed against primary infecting serotype may enhance the entry of a different serotype and may contribute to disease severity [37]. A study from Cuba on DENV4 infected patients and *FCGR2A* p.R131H polymorphism reported that the R allele and R/R genotype are associated with protection to symptomatic dengue while the H allele and H/H genotype are associated with both DF and DHF [38]. Genotype frequencies of *FCGR2A* p.R131H were not statistically different between DF, DHF and controls in general in India, but

the RR genotype was observed significantly increased in DENV cases with thrombocytopenia [37].

TNFA (tumor necrosis factor alpha) gene codes a proinflammatory cytokine, mainly secreted by macrophages in response to bacterial or viral infections. Functional assays have shown that transient suppression of TNF- α production during the early period of ADE infection in THP-1 cells promoted the initiation of DENV replication [39]. These results were contradicted in another work, where TNF- α at high and medium concentration inhibited DENV replication in human dendritic cells, a primary DENV target cells (cited in [40]). The promoter -308A allele has been related to enhanced TNF- α gene transcriptional activity [41], and different studies have associated it with diverse dengue phenotypes or failed to do so: protective against development of DHF/DSS in Malaysia [40]; not associated in Mexico [42]; dominant model marginally protective against DF in Brazil [43]; risk factor for DHF in Venezuela [44] and Cuba [45].

As technological improvements enabled genotyping of thousands of SNPs at once in a chip, association surveys upgraded to unbiased genome-wide association studies (GWAS). The thousands of SNPs screened in GWAS elevate the statistical burden, pushing the significance threshold to the level of 10^{-8} , implying the genotyping of thousands of cases and controls. A common variant identified in the discovery population, if rare in a replication population, will imply a greater sample size to achieve comparable statistical power to confirm the significant association [46]. And, definitely, trans-ethnic studies are needed, as it is evident that no single population is sufficient to fully uncover the variants underlying a certain disease in all populations [46]. The only traditional genome-wide association study (GWAS) published so far in dengue context was performed in Vietnamese children (2,008 dengue shock cases and 2,018 controls; replicated in 1,737 cases and 2,934 controls) and showed significant association of polymorphisms within *MICB* (MHC class I polypeptide-related sequence B) and *PLCE1* (phospholipase C, epsilon 1) genes with dengue shock syndrome [47]. The authors hypothesise that the possible protective role of *PLCE1* gene is through maintenance of normal endothelial barrier function, as mutations in this gene are associated with nephrotic syndrome, a kidney disorder characterised by dysfunction of the glomerular basement membrane resulting in proteinuria and hypoproteinemia that, when severe, leads to reduced vascular oncotic pressure and edema. Still, functional assays are needed to fully ascertain the mechanism by which this

gene may be involved in dengue disease. Several linked *PLCE1* SNPs presented significant association values with DSS in Vietnamese children [47], and further works have confirmed association of some of these SNPs in other dengue cohorts (such as rs3740360 with dengue without shock in Vietnamese children and adults [48]; rs3765524 with DSS in Thai children [49]). An *in silico* assay predicted that rs2274223 (H1619R) is deleterious as it induces structural changes in the C2 domain of PLCE protein [50]. As occurs in most GWAS-discovered candidate markers, the ORs for these SNPs have a limited impact, varying around the value of 0.76-0.85. *MICB* gene is located on the highly gene dense HLA region in chromosome 6, and encodes a stress-inducible activating ligand for the NKG2D type II receptor on natural killer and CD8+ T cells. The ligation of NKG2D by MICB protein stimulates antiviral effector functions in natural killer cells, including cytokine expression and the cytolytic response [51]. The rs3132468-C risk allele is associated with lower mRNA expression [48] and may lead to dysfunctional natural killer and/or CD8+ T cell activation early in infection, resulting in a higher viral burden *in vivo* [47]. However, given the location of this gene in the rich HLA-A region, it is still possible that its association with dengue fever is due to hitchhiking.

Ethnicity is traditionally considered a confounding factor in association studies, leading to false-positives when cases and controls are not correctly paired for it, and a principal component-based correction is usually applied in GWAS [52]. Nevertheless, ethnicity is an important source of variability to be accounted for, and not just discarded as noise. In this work, we applied population genetics-informed meta-analyses to evaluate the genetic risk/protection conferred by 10 SNPs across seven genes mostly associated with dengue in the literature (*PLCE1*, *TNFA*, *DC-SIGN/CD209*, *OAS1*, *OAS3*, *FCGR2A/CD32* and *MICB*; Table 1). We surveyed the literature for case-control cohorts [25,26,27,29,30,35,36,38,40,42,43,44,48,49,53,54,55,56,57,58], took advantage of available datasets chip-genotyped ([47,59]and Oliveira et al. 2018), and added a few case-control cohorts of our own (mostly Cambodia, Vietnam and few Brazilian individuals). Then this information was used in genetic risk evaluation at the worldwide scale.

2. Material and Methods

Genotyping and datasets

The 10 SNPs were included in an array ordered from Life Technologies™. DNA with concentration higher than 10 ng/μl was automatically transferred from a 384-well plate to the array in AccuFill™ (Life Technologies™) by capillarity. The array was screened in ViiA™7 or QuantStudio™ 12K Flex (Life Technologies™). The results were analyzed in TaqMan® Genotyper software (Life Technologies™).

We screened the allele frequencies in our collection of samples from several worldwide populations. These population samples were collected from individuals at random, healthy in the moment of collection, in which the only selection criteria was the affiliation in that population for at least three generations back in time. Sampled individuals gave informed consent and the Ethics Committee from University of Porto guaranteed accordance to global ethical guidelines (17/CEUP/2012). Around 1,312 individuals from 15 countries were genotyped in this study. Other publically available population data were used for extracting the allele frequency information: 1000 Genomes database, phase 3 [60], composed of 2,504 complete genomes from 26 populations; 35 Koreans [61]; 100 Southeast Asian Malays [62]; 498 Dutch [63]. A few genome-wide SNP chip based population studies were also consulted [64,65,66] as well as Maasai data (n=90) from HapMap [67]. Thus, the number of individuals successfully genotyped for each SNP was between 4,055 and 4,830, and of populations between 43 and 61. The maximum number of populations per population group was: 11 European; 3 North African; 11 Sub-Saharan African; 11 Southwest Asian; 8 South Asian; 5 Northeast Asian; 4 Southeast Asian; 2 North American; 1 Caribe; and 5 Latin American. Data are presented in Supplemental Tables S1-S10.

A few control-case samples from Cambodia, Vietnam and Brazil were also genotyped for these markers, and data are presented in Table S11. Written informed consent was obtained from all subjects or, in case of individuals under 18 years of age, from their parents or tutors. Local ethics committees approved the protocol.

Population genetics and selection analyses

Hardy-Weinberg equilibrium at each locus for each population and F_{ST} genetic distances between populations were calculated in the software Arlequin vs 3.0 [68]. Heat maps, done in R

package fields, represent the average F_{ST} genetic distances between and within (except same population comparison) population groups and average MAFs for the 10 SNPs in each population group. To visualize the geographical distribution of allele frequencies and genetic risk, we constructed interpolation maps using the “Spatial Analyst Extension” of ArcView version 3.2 (www.esri.com/software/arcview/). We used the “Inverse Distance Weighted” (IDW) option with a power of two for the interpolation of the surface. IDW assumes that each input point has a local influence that decreases with distance. The geographic location used is the centre of the distribution area from which the individual samples of each population were collected.

Mantel tests, to evaluate the correlation between F_{ST} genetic distance matrices and great-circle geographic distance matrices were conducted in R using package Ape [69] and Mantel correlograms were obtained in mpmcorrelogram, by using 10,000 permutations. W Kendall’s coefficient of concordance and Friedman’ chi-square statistic tests were registered, while in the correlograms the rM referring to the vector with the computed Mantel correlations for each of 12 distance classes (and information on significant p-values after a progressive Bonferroni correction) were plot.

We used the package REHH [70] to evaluate natural or artificial selection in the candidate SNPs, through the extended haplotype homozygosity (EHH) measure described by Sabeti et al. [71]. This measure was calculated for each SNP in all 1000 Genomes [19] populations belonging to the three main human population groups: sub-Saharan Africa (Gambia, Sierra Leone, Nigerian Yoruba, Nigerian Esan and Kenya), Europe (Iberia, Italy, Great Britain, Finland) and East Asia (Vietnam, Chinese Dai, Chinese Han, South Chinese Han and Japan). 5Mb phased data around the focal SNP from 1000 Genomes database were inputted in REHH. For comparison purposes, we also obtained this measure for *MCM6* SNP rs4988235, which is known to be under strong positive selection in Northern European populations (reviewed in [72]).

By using the 1000 Genomes project populations of Great Britain, China-Han and Yoruba, as proxies for European, East Asian and African population groups, we checked the linkage disequilibrium (LD) for the two SNPs in *PLCE1*, two SNPs in *MICB* and three SNPs in *OAS1/3* in Haploview vs 4.2 [73].

Meta-analyses

Meta-analyses were conducted in the R package Metafor [74]. 2x2 contingency tables for MAF, and recessive and dominant genotypes were used as inputs for OR calculations in the following comparisons (whenever possible due to sample size/number of studies): DF vs controls; DHF vs controls; DSS vs controls; symptomatic vs controls. In a few studies, the control groups included asymptomatic individuals. Funnel and radial plots were obtained and checked to test for publication biases, and heterogeneity across studies was also assessed through Cochran's Q and I^2 statistics. Pooled ORs were obtained by using three models: an inverse-variance weighted fixed-effects model; Der Simonian-Laird estimator as a random-effects model; a restricted maximum-likelihood estimator as another random-effects model. A fixed-effect meta-analysis assumes that the true effect of intervention is equal in all studies (that is, fixed across studies), implying that the observed differences among study results are due solely to chance, i.e. that there is no statistical heterogeneity. A random-effects meta-analysis assumes that the effects being estimated in the different studies are not identical, but are random following some distribution.

Genetic risk

The dengue disease genetic risk scores for DF and DHF+DSS were calculated by multiplying each individual's risk allele count for each locus by the reported beta coefficient for that polymorphism and summing the product for all corresponding loci (as in [75]). The beta parameter for each SNP corresponds with the natural log of the ORs.

3. Results and Discussion

Worldwide diversity for the 10 SNPs

The 10 SNPs associated with dengue fever can be divided into two groups in terms of frequencies and distribution across the globe (Fig. 1). The SNPs *PLCE1*-rs2274223-G, *PLCE1*-rs3765524-T, *CD209*-rs4804803-G, *OAS1*-rs10774671-G, *OAS3*-rs1859330-G, *OAS3*-rs2285933-G and *CD32*-rs1801274-G are high frequent, rounding $0.20 > \text{MAF} > 0.50$ (Fig. 1A). In some cases, the minimum allele becomes even the maximum allele in Africa, and the amplitude in the frequency between the two alleles can be very high, being 0.41 in *OAS1* and 0.38 in *CD209*. For all these seven SNPs, the highest MAF frequencies are observed in sub-Saharan African populations, decreasing towards Europe and being lowest in Southeast/Northeast Asians (Fig. 1B and Fig S2-S6). This distribution of allele frequencies leads to the highest average F_{ST} genetic distances (Fig S1) being observed between sub-Saharan African populations or their descendants (Caribe, and less so Latin/North America that are mixes between populations having variable African input) and Southeast/Northeast Asians, reaching extreme values of 0.32 and 0.29 for *CD209* and *OAS1* SNPs. When applying the Mantel test to evaluate correlations between the genetic and the geographic distances for each SNP (Table 1; Supplemental Fig S10), it is possible to verify the existence of overall significant positive correlations, indicating that there is no deviation from the rule of lower exchange of genes as geographical distance increases. Also, checking for the possible effect of natural positive selection upon these SNPs (Fig 2 and Fig S11-17), the measure EHH indicates that no selective sweeps occurred in these makers. For comparison purposes, Fig. 2 (and Fig S18) includes the profiles for the highly positively selected marker rs4988235 in *MCM6* gene, which in European populations confers lactose tolerance in human adult life (an extended haplotype homozygosity is observed for the derived allele, especially in northern Europeans; the derived allele is absent in sub-Saharan Africans and East Asians) [72]. As can be observed in the figure, the extended haplotype in this case is considerable larger than the ones observed in the genes associated with dengue.

The remaining three SNPs, *TNFA*-rs1800629-A, *MICB*-rs3132468-C and *MICB*-rs3134899-C, are intermediately frequent ($\sim 0.05 > \text{MAF} > 0.20$) across the globe (Fig. 1A), presenting higher MAF in Europeans and North Africans/Southwest Asians, decreasing towards Africa and Asia (Fig. 1C; Fig S7-9). Distinctly, SNPs on *MICB* and *TNFA* genes are the ones that display lower

inter-population group divergence, with values of F_{ST} attaining at most 0.06-0.08 and especially between Asian and European ancestries (Fig. S1). These four markers also display a positive correlation in the Mantel test (Table 1; Supplemental Fig S10), except for *MICB*-rs3132468-C allele, where values are not significant. The three markers equally do not show signals of positive selection (Fig. S19-21).

Meta-analysis

The 2009 WHO classification for dengue disease phenotypes [9] introduced important changes in relation to the 1997 classification [76], aiming at a better clinical structuring of dengue patients. However, it led to addition of noise when joining cohorts from different studies, with considerable overlap between phenotypic classes. Despite these biases, we decided it was worth to compare the information gain from considering (a) stricter phenotypic classes and (b) a broader symptomatic class with higher sample size. Thus, we performed meta-analyses in the following phenotypic classes *versus* healthy controls: DSS (dengue shock syndrome); DHF (dengue hemorrhagic fever); DF (dengue fever); and Symptomatic (all the other classes together and a few cases that were identified as presenting symptoms but not allowing to classify patients in the previous classes).

The SNPs with larger cohorts are the ones discovered in the Vietnamese GWAS [47], rounding thousands of patients and controls (Table S11), although they have been mainly screened in East Asian populations. Beginning by the SNPs in *MICB* gene, *MICB*-rs3132468-C was genotyped in 9493 patients (4360 DSS, 224 DHF, 3971 DF and 938 with symptoms) versus 6785 controls, while *MICB*-rs3134899-C was screened in 4781 patients (3987 DSS, 225 DHF, 472 DF and 97 with symptoms) versus 5630 controls. The most significant results are for the comparison DSS vs controls (Fig. 3), although values are still significant for the symptomatic vs controls comparison (Fig. S24), while comparisons for the less severe phenotypes are not significant (Fig. S22-23). In the DSS vs controls, all the studies showed a high concordance between them, as can be confirmed in the funnel and radial plots (Fig 3B,C,E,F), and also through the non-significant Cochran's Q measure and nil I^2 values. The pooled OR values in DSS were of 1.32 [1.22-1.42] for *MICB*-rs3132468-C and 1.27 [1.16-1.39] for *MICB*-rs3134899-C, being highly statistically significant ($1.19e^{-12}$ and $3.13e^{-7}$, respectively). The concordance between these OR values reflect

the fact that the two markers are linked (Fig. S33) and may be tagging the not yet identified causative marker in this chromosome region.

For *PLCE1* markers, also thousands of patients and controls (Table S11), mainly (97%) Eastern Asians, have been genotyped. *PLCE1*-rs2274223-G was genotyped in 4763 patients (3987 DSS, 217 DHF, 192 DF and 90 with symptoms) versus 5614 controls, while *PLCE1*-rs3765524-T was screened in 5133 patients (3904 DSS, 157 DHF, 141 DF and 931 with symptoms) versus 5281 controls. The small Brazilian cohorts and the Thai cohort from Oliveira et al. (which has a higher frequency of heterozygous in DSS class than neighbour cohorts; Oliveira et al. 2018) introduced heterogeneity in the meta-analyses for *PLCE1*-rs2274223-G (Fig. S25A-C). If these samples are removed from the analyses, the other studies give a homogeneous signal, leading to the detection of a significant protective association for this allele and DSS (Fig. 4A), with an OR of 0.80 [0.76-0.86] ($p=1.93e-10$). The significant value is still detectable in the Symptomatic vs control comparison ($p=2.81e-10$), but not in the other phenotypes (Fig. S25D-L). Identical results are obtained for the linked (Fig. S33) *PLCE1*-rs3765524-T allele: OR=0.79 [0.75-0.85] ($p=1.79e-11$) in DSS (Fig. 4B; Fig S26).

CD32-rs1801274-G meta-analysis departed from 5343 patients (3988 DSS, 588 DHF, 673 DF and 94 with symptoms) versus 6195 controls, with 8% of the individuals being from Latin America. This SNP was contained in the chip screened in Khor et al. [47], and that is the reason for this large dataset. Notice that this SNP is not significantly associated with DSS in this Vietnamese cohort alone (see Fig. 5). A first test (Fig. S27) showed that our data from Cambodia and Cuban cohort from Garcia et al. [38] were introducing biases ($I^2 \approx 65\%$), in opposite directions. When removing these two studies, the other studies are homogeneous, as can be verified in the funnel and radial plots, as well as in Cochran's Q and I^2 measures. Association for this marker (Fig. 5) is close to association in DHF ($p=0.053$), with a protective OR (0.82 [0.67-1.00]), which becomes significant ($p=0.038$) in the larger symptomatic comparison with OR (0.94 [0.89-1.00]). Curiously, despite the large sample size in DSS comparison ($\sim 7x$ larger than DHF), the association is not significant in this phenotype class. It thus seem that CD32 is more closely associated with DHF.

The total sample sizes for the remaining SNPs are smaller given their absence from the chip screened in Vietnam [47]. *CD209*-rs4804803-G, one of the first dengue-associated genetic

marker, amounted to 3177 patients (123 DSS, 1262 DHF, 1682 DF and 110 with symptoms) versus 1616 controls. Around 10% of patients and 26% of controls are from Latin America, while the rest are of East/South Asian ancestry. This disparity in the sample sizes between ethnicities does not allow to test the hypothesis if the effect of this SNP could be opposite between East Asian and Latin American backgrounds, as advanced previously [30]. The strong significant causative effect of *CD209*-rs4804803-G in DF in the Thai cohort reported in Sakuntabhai et al. [25] was not obtained in any other work available so far. Even after removing that Thai cohort from the test, the remaining samples input a significant heterogeneity into the meta-analysis in DF vs control comparison (Fig. 6A), so that overall there is no significant effect in this phenotype in the three tested models. The same being true for the phenotype DHF (Fig. 6B), and when considering all Symptomatic (Fig. S28).

TNFA-rs1800629-A dataset sums up 1943 patients (165 DSS, 776 DHF, 906 DF and 96 with symptoms) versus 1934 controls, of which 51% patients and 74% controls are from Latin American ancestry. When samples are considered together, only the larger Symptomatic vs control comparison presents a significant protective effect (OR=0.81 [0.67-0.98]; p=0.022 to p=0.046 depending on the model), although heterogeneity is still around 20% for I^2 measure (Fig. 7A). Limiting the analysis to the Latin American samples, a significant protective effect (OR=0.67 [0.50-0.90]; p=0.0092) was observed for the DF vs control comparison (Fig. 7B), and not for the severer phenotypes (Fig. S29).

Lastly, the OAS family presented the lower sample sizes screened so far, and with the exceptions of a few individuals from India [35] and Cuba [59] typed for *OAS1*-rs10774671-G, the Thai cohorts from Simon-Loriere et al. [36] dominate these meta-analyses. Notice that Simon-Loriere et al. [36] did not provide controls, and in order to be consistent in our tests we included our Thai population screening as controls to compare with these Thai patient cohorts. We also decided to perform an overall evaluation, disregarding DENV strain, which was not accounted for in all other studies besides Simon-Loriere et al. [36]. Numbers amounted to 1125 patients (185 DSS, 505 DHF, 407 DF and 28 with symptoms) versus 402 controls for *OAS1*-rs10774671-G, 931 patients (187 DSS, 412 DHF, 304 DF and 28 with symptoms) versus 242 controls for *OAS3*-rs1859330-G, and 941 patients (187 DSS, 419 DHF, 307 DF and 28 with symptoms) versus 236 controls for *OAS3*-rs2285933-G. Values were non-significant for all *OAS1*-rs10774671-G and

linked (Fig. S33) *OAS3*-rs1859330-G comparisons (Fig. S30-31). While *OAS3*-rs2285933-G confers significant protection in DSS, with OR of 0.55 [0.34-0.90], $p=0.016$ considering the fixed-effect model (Fig. 8; Fig. S32).

Genetic risk scores

For markers confirmed here to be associated with the severer phenotype DSS, the protective are highly frequent in Africa (*PLCE1* SNPs and *OAS3*-rs2285933-G), while the causative are low frequent in Africa (*MICB* SNPs). The same is true for the *CD32*-rs1801274-G, which seems to protect from DHF. Thus, when summing up the risk scores conferred by *PLCE1*-rs2274223-G (linked *PLCE1*-rs3765524-T was left out), *MICB*-rs3132468-C (linked *MICB*-rs3134899-C was also left out), *OAS3*-rs2285933-G and *CD32*-rs1801274-G (Fig. 9A), sub-Saharan populations and their descendants are the best protected against the severer dengue phenotypes, in contrast with both Northeast Asian and Southeast populations, which are least protected. The difference in the means of genetic risk between those populations groups are highly significant (sub-Saharan Africa vs Southeast Asia, $p=2.96e^{-63}$; sub-Saharan Africa vs Northeast Asia, $p=7.66e^{-56}$). European populations and neighbor North African and Southwest Asian populations are in between those genetic risk scores, being significantly different from both (sub-Saharan Africa vs Europe, $p=2.02e^{-16}$; Europe vs Southeast Asia, $p=2.63e^{-8}$).

The protection conferred by *TNFA*-rs1800629-A to DF is higher in European, North African and Southwest Asian populations, and decreases towards Africa and especially East Asia (Fig 9B), being statistically significant different for the latter (sub-Saharan Africa vs Europe, $p=0.85$; Europe vs Southeast Asia, $p=2.22e^{-16}$).

4. Conclusions

Population genetic models predict that a large fraction of missing heritability for complex traits could be explained by loci that contain classes of rare (MAF <1%) susceptibility variants [77]. However, GWAS have been identifying mainly common variants, already present in the ancestral African population that migrated out-of-Africa [46]. In fact, by analysing 400 susceptibility-conferring SNPs across a spectrum of qualitative and quantitative traits, Park et al. [78] observed that both their number and their collective contribution to genetic variance were highest for more common MAF categories (30–40% or 40–50%), and dropped substantially for lower allele-frequency categories (5–10% or 10–20%). The authors confirmed that these results are not caused by lower statistical power for detection of association for low-frequency SNPs. Another simulation analysis [79] showed that variance due to rare alleles is only expected for diseases caused primarily by strong deleterious mutations (diseases tightly coupled to fitness). The 10 markers surveyed here for association with dengue disease follow these tendencies of SNPs uncovered by GWAS for non-strong uncoupling diseases. Seven markers (in *PLCE1*, *CD32*, *CD209*, *OAS1* and *OAS3* genes) have high-frequency MAF and the other three (in *MICB* and *TNFA* genes) have intermediate frequency MAF. Most of these markers have high discriminatory power between population groups. This is especially so for the SNPs in *CD209* and *OAS1* genes, which attain F_{ST} values between 0.20-0.32 in comparisons involving sub-Saharan African and Northeast/Southeast Asian ancestries. This level of values is attained by the so-called ancestry informative markers (our own calculations in SNPs referred in [80]), while reference values taking into account multiple SNPs based on whole genome sequencing are at the level of 0.106 for European-East Asian and 0.139 for European-West African comparisons [81]. Despite this high population discrimination, frequencies of these alleles follow the rules of genetic drift and seem not to have been under strong selective pressure.

If a large proportion of causal variants are common and shared across diverse populations, they will be replicated in trans-ethnic GWAS, as demonstrated in a large-scale review [82] for 28 diseases in European, East Asian and African ancestries. Even when power is insufficient to achieve statistically independent significance, there are high rates of directional consistency across trans-ethnic GWAS signals [83]. We confirmed this tendency to directional consistency between studies performed in East Asia and Latin America in dengue, with heterogeneity

generated by randomness between studies and especially by low sample sizes. Also the patterns of LD for the *PLCE1*, *MICB* and *OAS* genes are quite consistent in the three main population groups, and, so, the potential of replication in East Asia or Latin America for these markers is identical.

Another important result from Park et al. [78] study was that the genetic variance due to individual susceptibility SNPs, on average, remains fairly constant over different ranges of allele frequency, as for example in Crohn's disease, the average regression effects corresponded to OR of 1.08 for MAF=0.45, 1.13 for MAF=0.15, and 1.16 for MAF=0.05. Our results also corroborate this evidence, with close protective OR values, between 0.55 for *OAS3*-rs2285933-G, 0.67 for *TNFA*-rs1800629-A, 0.80 for *PLCE1*-rs2274223-G and 0.82 for *CD32*-rs1801274-G, despite differences in MAFs. Even the causative OR was of similar order of magnitude, of 1.32 for *MICB*-rs3132468-C.

Our meta-analysis was not informative enough to replicate overall the significant association between *CD209*-rs4804803-G and DF/DHF reported by Sakuntabhai et al. [25] in Thai, neither the opposite-direction association reported in Brazilians [30]. This result could be interpreted as surprising given the collected functional evidence that this polymorphism interferes with response to DENV infection [25], but follows many failed replication studies, even of associations that turned out to be genuine. The meta-analysis conducted by Lohmueller et al. [84] of 301 association studies of 25 disease loci, showed that only 11 loci were confirmed in the replications, and, even more striking, 24 of the 25 loci presented lower ORs than the initial study. This result is a consequence of the data set used to identify the variant of interest being a non-random population sample [85]. In fact, tests oversample affected individuals relative to their frequency in the population (an unreal similar number of cases and controls are tested), and there is a major ascertainment effect that occurs when a variant is of interest, specifically because it was significant for association. As most of the associations have modest effects, when there is a significant result, it may imply that the genotype counts of cases and controls are more different from each other than expected. Consequently, the estimates of effect size are biased upwards, an effect known as the "winner's curse". Gorroochurn et al. [86] developed a formula for the replication power of a second association study based on the p-value of an initial study, concluding that: (1) a p-value only slightly lower than the nominal α results in only

approximately 50% replication power; (2) very low p-values are required to achieve a replication power of at least 80% (e.g., at $\alpha=0.05$, a $p<0.005$ is required). The authors state that because many initially significant findings result in low replication power, replication failure should not be surprising or be interpreted as necessarily refuting the initial findings.

Our results also support the structuring of dengue patients into specific phenotype classes, despite the lowering in sample size for the statistical evaluation. It seems that particular polymorphisms can confer susceptibility or protection against particular molecular phenomena, which play a different role along the disease course, as we have recently (Oliveira et al. 2018) demonstrated in a GWAS conducted in Thai cohorts. In this study we were able to identify four genes involved in xenobiotic metabolism that are protecting against dengue fever phenotype, possibly through linking to the virus and conditioning of its localization in host cells. While a new gene (*PLCB4*) of the *PLCE1*-family protects against dengue shock syndrome and may be related with cytokine dynamics, inflammation and activation of vascular endothelium cells. In the current study, *TNFA* seems to protect for DF, *CD32* protects for DHF (the large sample size for DSS in the meta-analysis shows that the non-association with this phenotype class is not a sample size issue), while *PLCE1*, *MICB* and *OAS3* play a role at the level of DSS.

We were not able to evaluate the influence of dengue virus strain in the associations detected in these markers. Simon-Loriere et al. [36] showed statistically and functionally that this is an important factor when ascertaining the association for OAS family. Most studies do not provide information about the circulating strain when cohorts were collected, rendering it is impossible to conduct meta-analyses incorporating the strain information. Groups working in the field should henceforth make an effort to provide this information for future proper evaluation.

As failures in replications accumulate in GWAS, and sample sizes in the order of thousands are difficult to cope with (implying huge efforts in sample collection and financing), the net result was a general disinterest in performing GWAS, and specifically in neglected tropical diseases. Fortunately, some improvements in statistical methods are allowing to take advantage of ethnicity information in conducting more powerful association studies in lower sample size cohorts [87,88]. These methods analyse blocks of SNPs that can be affiliated to one of the ancestral populations. The blocks are randomly distributed by recombination across the chromosomes of the admixed descendant individuals, reflecting the frequency contributions of

the parental populations (expectation for the control group), except in genomic locations where a candidate gene confers susceptibility to a certain disease (significant frequency increase in the parental population with higher MAF for the susceptibility marker) in cases. As there are only hundreds of ancestral blocks, these tests have considerable lower statistical burdens than the traditional GWAS with thousands of SNPs. We took advantage of joined admixture-association studies and conducted work in Cuban [59] and Thai (Oliveira et al. submitted) cohorts, mapping new candidate genes: lipid metabolism-related *OSBPL10* and *RXRA*; and the already mentioned four genes involved in xenobiotic metabolism (*CHST10*, *AHRR*, *GRIP1* and *PPP2R5E*) and *PLCB4* of the *PLCE1*-family. The genetic risks conferred by the markers reported in these publications add precisely to the genetic risks displayed by the markers studied here. Sub-Saharan African populations and descendants are the best protected against the severer forms of the disease, while Southeast and Northeast Asians are the least protected ones. European and close neighbours are the best protected against dengue fever, while, again, Southeast and Northeast Asians are the least protected ones.

Acknowledgments

The research leading to these results has received funding from the European Commission Seventh Framework Programme [FP7/2007-2013] for the DENFREE project under Grant Agreement no. 282378. MO has a PhD grant from FCT (The Portuguese Foundation for Science and Technology - SFRH/BD/95626/2013). IPATIMUP integrates the i3S Research Unit, which is partially supported by FCT, FEDER and COMPETE (PEst-C/SAU/LA0003/2013).

References

1. WHO (2012) Accelerating work to overcome the global impact of neglected tropical diseases. A roadmap for implementation. Geneva: World Health Organization.
2. Gautret P, Simon F (2016) Dengue, chikungunya and Zika and mass gatherings: What happened in Brazil, 2014. *Travel medicine and infectious disease* 14: 7-8.
3. Morel CM (2003) Neglected diseases: under-funded research and inadequate health interventions. Can we change this reality? *EMBO reports* 4 Spec No: S35-38.
4. Guzman MG, Harris E (2015) Dengue. *Lancet* (London, England) 385: 453-465.
5. Vaughn DW, Green S, Kalayanarooj S, Innis BL, Nimmannitya S, et al. (2000) Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity. *J Infect Dis* 181: 2-9.
6. Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, et al. (2014) Global spread of dengue virus types: mapping the 70 year history. *Trends in microbiology* 22: 138-146.
7. Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, et al. (2015) The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *Elife* 4: e08347.
8. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, et al. (2013) The global distribution and burden of dengue. *Nature* 496: 504-507.
9. WHO (2009) Dengue: guidelines for diagnosis, treatment, prevention and control. Geneva: World Health Organization. pp. 160.
10. Stoler J, Al Dashti R, Anto F, Fobil JN, Awandare GA (2014) Deconstructing "malaria": West Africa as the next front for dengue fever surveillance and control. *Acta tropica* 134: 58-65.
11. Duong V, Lambrechts L, Paul RE, Ly S, Lay RS, et al. (2015) Asymptomatic humans transmit dengue virus to mosquitoes. *Proc Natl Acad Sci U S A* 112: 14688-14693.
12. McEvoy BP, Powell JE, Goddard ME, Visscher PM (2011) Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome Res* 21: 821-829.
13. Rito T, Richards MB, Fernandes V, Alshamali F, Cerny V, et al. (2013) The first modern human dispersals across Africa. *PLoS One* 8: e80031.
14. Soares P, Ermini L, Thomson N, Mormina M, Rito T, et al. (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84: 740-759.
15. Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, et al. (2008) Proportionally more deleterious genetic variation in European than in African populations. *Nature* 451: 994-997.
16. Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, et al. (2012) The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol Biol Evol* 29: 915-927.
17. Atkinson QD, Gray RD, Drummond AJ (2008) mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. *Mol Biol Evol* 25: 468-474.

18. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, et al. (2013) Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493: 216-220.
19. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56-65.
20. Karlsson EK, Kwiatkowski DP, Sabeti PC (2014) Natural selection and infectious disease in human populations. *Nat Rev Genet* 15: 379-393.
21. Coffey LL, Mertens E, Brehin AC, Fernandez-Garcia MD, Amara A, et al. (2009) Human genetic determinants of dengue virus susceptibility. *Microbes Infect* 11: 143-156.
22. Lozach PY, Burleigh L, Staropoli I, Navarro-Sanchez E, Harriague J, et al. (2005) Dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin (DC-SIGN)-mediated enhancement of dengue virus infection is independent of DC-SIGN internalization signals. *J Biol Chem* 280: 23698-23708.
23. Navarro-Sanchez E, Altmeyer R, Amara A, Schwartz O, Fieschi F, et al. (2003) Dendritic-cell-specific ICAM3-grabbing non-integrin is essential for the productive infection of human dendritic cells by mosquito-cell-derived dengue viruses. *EMBO reports* 4: 723-728.
24. Tassaneeritthep B, Burgess TH, Granelli-Piperno A, Trumpfheller C, Finke J, et al. (2003) DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. *The Journal of experimental medicine* 197: 823-829.
25. Sakuntabhai A, Turbpaiboon C, Casademont I, Chuansumrit A, Lowhnoo T, et al. (2005) A variant in the CD209 promoter is associated with severity of dengue disease. *Nat Genet* 37: 507-513.
26. Wang L, Chen RF, Liu JW, Lee IK, Lee CP, et al. (2011) DC-SIGN (CD209) Promoter -336 A/G polymorphism is associated with dengue hemorrhagic fever and correlated to DC-SIGN expression and immune augmentation. *PLoS Negl Trop Dis* 5: e934.
27. Alagarasu K, Damle IM, Bachal RV, Mulay AP, Shah PS, et al. (2013) Association of promoter region polymorphisms of CD209 gene with clinical outcomes of dengue virus infection in Western India. *Infect Genet Evol* 17: 239-242.
28. Noecker CA, Amaya-Larios IY, Galeana-Hernandez M, Ramos-Castaneda J, Martinez-Vega RA (2014) Contrasting associations of polymorphisms in FcγRIIIa and DC-SIGN with the clinical presentation of dengue infection in a Mexican population. *Acta tropica* 138: 15-22.
29. Vargas-Castillo AB, Ruiz-Tovar K, Vivanco-Cid H, Quiroz-Cruz S, Escobar-Gutierrez A, et al. (2017) Association of Single-Nucleotide Polymorphisms in Immune-Related Genes with Development of Dengue Hemorrhagic Fever in a Mexican Population. *Viral immunology*.
30. Xavier-Carvalho C, Gibson G, Brasil P, Ferreira RX, de Souza Santos R, et al. (2013) Single nucleotide polymorphisms in candidate genes and dengue severity in children: a case-control, functional and meta-analysis study. *Infect Genet Evol* 20: 197-205.
31. Silva LK, Blanton RE, Parrado AR, Melo PS, Morato VG, et al. (2010) Dengue hemorrhagic fever is associated with polymorphisms in JAK1. *Eur J Hum Genet* 18: 1221-1227.
32. Peisley A, Hur S (2013) Multi-level regulation of cellular recognition of viral dsRNA. *Cellular and molecular life sciences : CMLS* 70: 1949-1963.
33. Warke RV, Xhaja K, Martin KJ, Fournier MF, Shaw SK, et al. (2003) Dengue virus induces novel changes in gene expression of human umbilical vein endothelial cells. *Journal of virology* 77: 11822-11832.
34. Acosta EG, Kumar A, Bartenschlager R (2014) Revisiting dengue virus-host cell interaction: new insights into molecular and cellular virology. *Advances in virus research* 88: 1-109.
35. Alagarasu K, Honap T, Damle IM, Mulay AP, Shah PS, et al. (2013) Polymorphisms in the oligoadenylate synthetase gene cluster and its association with clinical outcomes of dengue virus infection. *Infect Genet Evol* 14: 390-395.

36. Simon-Lorriere E, Lin RJ, Kalayanarooj SM, Chuansumrit A, Casademont I, et al. (2015) High Anti-Dengue Virus Activity of the OAS Gene Family Is Associated With Increased Severity of Dengue. *J Infect Dis* 212: 2011-2020.
37. Alagarasu K, Bachal RV, Damle I, Shah PS, Cecilia D (2015) Association of FCGR2A p.R131H and CCL2 c.-2518 A>G gene variants with thrombocytopenia in patients with dengue virus infection. *Human immunology* 76: 819-822.
38. Garcia G, Sierra B, Perez AB, Aguirre E, Rosado I, et al. (2010) Asymptomatic dengue infection in a Cuban population confirms the protective role of the RR variant of the Fcγ3R polymorphism. *The American journal of tropical medicine and hygiene* 82: 1153-1156.
39. Chareonsirisuthigul T, Kalayanarooj S, Ubol S (2007) Dengue virus (DENV) antibody-dependent enhancement of infection upregulates the production of anti-inflammatory cytokines, but suppresses anti-DENV free radical and pro-inflammatory cytokine production, in THP-1 cells. *The Journal of general virology* 88: 365-375.
40. Sam SS, Teoh BT, Chinna K, AbuBakar S (2015) High producing tumor necrosis factor alpha gene alleles in protection against severe manifestations of dengue. *International journal of medical sciences* 12: 177-186.
41. Kroeger KM, Carville KS, Abraham LJ (1997) The -308 tumor necrosis factor-α promoter polymorphism effects transcription. *Molecular immunology* 34: 391-399.
42. Garcia-Trejo AR, Falcon-Lezama JA, Juarez-Palma L, Granados J, Zuniga-Ramos J, et al. (2011) Tumor necrosis factor alpha promoter polymorphisms in Mexican patients with dengue fever. *Acta tropica* 120: 67-71.
43. Santos A, de Moura EL, Ferreira JM, de Moura AWA, Ferreira ADS, et al. (2017) Association of TNFA (-308G/A), IFNG (+874 A/T) and IL-10 (-819 C/T) polymorphisms with protection and susceptibility to dengue in Brazilian population. *Asian Pacific journal of tropical medicine* 10: 1065-1071.
44. Fernandez-Mestre MT, Gendzekhadze K, Rivas-Vetencourt P, Layrisse Z (2004) TNF-α-308A allele, a possible severity risk factor of hemorrhagic manifestation in dengue fever patients. *Tissue antigens* 64: 469-472.
45. Perez AB, Sierra B, Garcia G, Aguirre E, Babel N, et al. (2010) Tumor necrosis factor-α, transforming growth factor-β1, and interleukin-10 gene polymorphisms: implication in protection or susceptibility to dengue hemorrhagic fever. *Human immunology* 71: 1135-1140.
46. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, et al. (2010) Genome-wide association studies in diverse populations. *Nat Rev Genet* 11: 356-366.
47. Khor CC, Chau TN, Pang J, Davila S, Long HT, et al. (2011) Genome-wide association study identifies susceptibility loci for dengue shock syndrome at MICB and PLCE1. *Nat Genet* 43: 1139-1141.
48. Whitehorn J, Chau TN, Nguyet NM, Kien DT, Quyen NT, et al. (2013) Genetic variants of MICB and PLCE1 and associations with non-severe dengue. *PLoS One* 8: e59067.
49. Dang TN, Naka I, Sa-Ngasang A, Anantapreecha S, Chanama S, et al. (2014) A replication study confirms the association of GWAS-identified SNPs at MICB and PLCE1 in Thai patients with dengue shock syndrome. *BMC Med Genet* 15: 58.
50. Taqi MM, Waseem D, Ismatullah H, Haider SA, Faisal M (2016) In silico transcriptional regulation and functional analysis of dengue shock syndrome associated SNPs in PLCE1 and MICB genes. 16: 335-345.
51. Muntasell A, Magri G, Pende D, Angulo A, Lopez-Botet M (2010) Inhibition of NKG2D expression in NK cells by cytokines secreted in response to human cytomegalovirus infection. *Blood* 115: 5170-5179.
52. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904-909.

53. Fernando AN, Malavige GN, Perera KL, Premawansa S, Ogg GS, et al. (2015) Polymorphisms of Transporter Associated with Antigen Presentation, Tumor Necrosis Factor-alpha and Interleukin-10 and their Implications for Protection and Susceptibility to Severe Forms of Dengue Fever in Patients in Sri Lanka. *Journal of global infectious diseases* 7: 157-164.
54. Alagarasu K, Mulay AP, Singh R, Gavade VB, Shah PS, et al. (2013) Association of HLA-DRB1 and TNF genotypes with dengue hemorrhagic fever. *Human immunology* 74: 610-617.
55. Sanchez-Leyva M, Sanchez-Zazueta JG, Osuna-Ramos JF, Rendon-Aguilar H, Felix-Espinoza R, et al. (2017) Genetic Polymorphisms of Tumor Necrosis Factor Alpha and Susceptibility to Dengue Virus Infection in a Mexican Population. *Viral immunology* 30: 615-621.
56. Dang TN, Naka I, Sa-Ngasang A, Anantapreecha S, Wichukchinda N, et al. (2016) Association of BAK1 single nucleotide polymorphism with a risk for dengue hemorrhagic fever. *BMC Med Genet* 17: 43.
57. Oliveira LF, Lima CP, Azevedo Rdo S, Mendonca DS, Rodrigues SG, et al. (2014) Polymorphism of DC-SIGN (CD209) promoter in association with clinical symptoms of dengue fever. *Viral immunology* 27: 245-249.
58. Loke H, Bethell D, Phuong CX, Day N, White N, et al. (2002) Susceptibility to dengue hemorrhagic fever in vietnam: evidence of an association with variation in the vitamin d receptor and Fc gamma receptor IIa genes. *The American journal of tropical medicine and hygiene* 67: 102-106.
59. Sierra B, Triska P, Soares P, Garcia G, Perez AB, et al. (2017) OSBPL10, RXRA and lipid metabolism confer African-ancestry protection against dengue haemorrhagic fever in admixed Cubans. *PLoS Pathog* 13: e1006220.
60. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. (2015) A global reference for human genetic variation. *Nature* 526: 68-74.
61. Zhang W, Meehan J, Su Z, Ng HW, Shu M, et al. (2014) Whole genome sequencing of 35 individuals provides insights into the genetic architecture of Korean population. *BMC Bioinformatics* 15 Suppl 11: S6.
62. Wong LP, Ong RT, Poh WT, Liu X, Chen P, et al. (2013) Deep whole-genome sequencing of 100 southeast Asian Malays. *Am J Hum Genet* 92: 52-66.
63. Consortium TGoN (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 46: 818-825.
64. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319: 1100-1104.
65. Hellenthal G, Busby GB, Band G, Wilson JF, Capelli C, et al. (2014) A genetic atlas of human admixture history. *Science* 343: 747-751.
66. Henn BM, Botigue LR, Gravel S, Wang W, Brisbin A, et al. (2012) Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet* 8: e1002397.
67. Consortium TIH (2005) A haplotype map of the human genome. *Nature* 437: 1299-1320.
68. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1: 47-50.
69. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20: 289-290.
70. Gautier M, Vitalis R (2012) rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28: 1176-1177.
71. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419: 832-837.
72. Gerbault P (2013) The onset of lactase persistence in Europe. *Human heredity* 76: 154-161.
73. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263-265.

74. Viechtbauer W (2010) Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw* 36: 1-48.
75. Marden JR, Walter S, Tchetgen Tchetgen EJ, Kawachi I, Glymour MM (2014) Validation of a polygenic risk score for dementia in black and white individuals. *Brain and behavior* 4: 687-697.
76. WHO (1997) Dengue haemorrhagic fever: diagnosis, treatment, prevention and control. Geneva: World Health Organization. 84 p.
77. Pritchard JK (2001) Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 69: 124-137.
78. Park JH, Gail MH, Weinberg CR, Carroll RJ, Chung CC, et al. (2011) Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc Natl Acad Sci U S A* 108: 18026-18031.
79. Simons YB, Turchin MC, Pritchard JK, Sella G (2014) The deleterious mutation load is insensitive to recent population history. *Nat Genet* 46: 220-224.
80. Marcheco-Teruel B, Parra EJ, Fuentes-Smith E, Salas A, Buttenschon HN, et al. (2014) Cuba: exploring the history of admixture and the genetic basis of pigmentation using autosomal and uniparental markers. *PLoS Genet* 10: e1004488.
81. Bhatia G, Patterson N, Sankararaman S, Price AL (2013) Estimating and interpreting FST: the impact of rare variants. *Genome Res* 23: 1514-1521.
82. Marigorta UM, Navarro A (2013) High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS Genet* 9: e1003566.
83. Ntzani EE, Liberopoulos G, Manolio TA, Ioannidis JP (2012) Consistency of genome-wide associations across major ancestral groups. *Human genetics* 131: 1057-1071.
84. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 33: 177-182.
85. Zollner S, Pritchard JK (2007) Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet* 80: 605-615.
86. Gorroochurn P, Hodge SE, Heiman GA, Durner M, Greenberg DA (2007) Non-replication of association studies: "pseudo-failures" to replicate? *Genetics in medicine : official journal of the American College of Medical Genetics* 9: 325-331.
87. Shriner D, Adeyemo A, Rotimi CN (2011) Joint ancestry and association testing in admixed individuals. *PLoS Comput Biol* 7: e1002325.
88. Maples BK, Gravel S, Kenny EE, Bustamante CD (2013) RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet* 93: 278-288.

Figure legends

Figure 1- Worldwide diversity for the 10 SNPs. (A) Heatmap for the minimum allele frequency (MAF) in the 10 SNPs. (B) Map of allele frequency distribution for *PLCE1*_rs2274223_G, which follows a pattern shared with *PLCE1*-rs3765524-T, *CD209*-rs4804803-G, *OAS1*-rs10774671-G, *OAS3*-rs1859330-G, *OAS3*-rs2285933-G and *CD32*-rs1801274-G (maps for these SNPs are displayed in Fig S2-7). (C) Map of allele frequency distribution for *TNFA*-rs1800629-A, which follows a pattern shared with *MICB*-rs3132468-C and *MICB*-rs3134899-C (maps for these are displayed in Fig S8-9).

Figure 2- Extended Haplotype Homozygosity (EHH) for *PLCE1*_rs2274223_G and the well-known positively selected European marker for lactase persistence *MCM6*-rs4988235. Data for one population representative from the three main human population groups (Nigerian Yorubans from sub-Saharan Africa, Great Britain from Europe and Chinese Han from East Asia; 1000 Genomes database).

Figure 3- Meta-analysis in *MICB*-rs3132468-C (A-C) and *MICB*-rs3134899-C (D-F) for DSS vs controls comparison. (A and D) Forest plots, pooled log OR ([95% CI] and p-values), Cochran's Q measure (and p-values) and I^2 values for the tested three models (weighted Mantel-Haenszel fixed-effect, Der Simonian-Laird, and restricted maximum-likelihood). (B and E) Funnel plots. (C and F) Radial plots.

Figure 4- Meta-analysis in *PLCE1*-rs2274223-G (A-C) and *PLCE1*-rs3765524-T (D-F) for DSS vs controls comparison. (A and D) Forest plots, pooled log OR ([95% CI] and p-values), Cochran's Q measure (and p-values) and I^2 values for the tested three models (weighted Mantel-Haenszel fixed-effect, Der Simonian-Laird, and restricted maximum-likelihood). (B and E) Funnel plots. (C and F) Radial plots.

Figure 5- Meta-analysis in *CD32*-rs1801274-G for DHF vs controls (A-C) and Symptomatic vs controls (D-F) comparisons. (A and D) Forest plots, pooled log OR ([95% CI] and p-values), Cochran's Q measure (and p-values) and I^2 values for the tested three models (weighted Mantel-Haenszel fixed-effect, Der Simonian-Laird, and restricted maximum-likelihood). (B and E) Funnel plots. (C and F) Radial plots.

Figure 6- Meta-analysis in *CD209*-rs4804803-G for DF vs controls (A-C) and DHF vs controls (D-F) comparisons. (A and D) Forest plots, pooled log OR ([95% CI] and p-values), Cochran's Q measure (and p-values) and I^2 values for the tested three models (weighted Mantel-Haenszel fixed-effect, Der Simonian-Laird, and restricted maximum-likelihood). (B and E) Funnel plots. (C and F) Radial plots.

Figure 7- Meta-analysis in *CD209*-rs4804803-G for DF vs controls (A-C) and DHF vs controls (D-F) comparisons. (A and D) Forest plots, pooled log OR ([95% CI] and p-values), Cochran's Q measure (and p-values) and I^2 values for the tested three models (weighted Mantel-Haenszel fixed-effect, Der Simonian-Laird, and restricted maximum-likelihood). (B and E) Funnel plots. (C and F) Radial plots.

Figure 8- Meta-analysis in *OAS3*-rs2285933-G for DSS vs controls comparison (A-C). (A) Forest plots, pooled log OR ([95% CI] and p-values), Cochran's Q measure (and p-values) and I^2 values for the tested three models (weighted Mantel-Haenszel fixed-effect, Der Simonian-Laird, and restricted maximum-likelihood). (B) Funnel plots. (C) Radial plots.

Figure 9- Genetic risk for DHF+DSS (A) and DF (B) in population samples reported in Tables S1-S10. (A) Values were calculated taking into account the significant association found here for *PLCE1*-rs2274223-G, *MICB*-rs3132468-C, *OAS3*-rs2285933-G and *CD32*-rs1801274-G. (B) Values were calculated for the significant association with *TNFA*-rs1800629-A.

Table 1- General information about the 10 SNPs screened in this work and results for the Mantel test (W= Kendall's coefficient of concordance; p-value for Friedman' chi-square statistic tests of W; significant values are indicated on bold). MAF refers to the minimum allele frequency in the global human population.

Gene	SNP	Position in build GRCh37	Function/Location	MAF	Mantel test	
					W	p-value
<i>PLCE1</i>	rs2274223	10:96066341	Missense (His1927Arg)	G	0.677	0.0001
<i>PLCE1</i>	rs3765524	10:96058298	Missense (Thr1777Ile)	T	0.659	0.0001
<i>TNFA</i>	rs1800629	6:31543031	Upstream gene variant	A	0.557	0.0219
DC-SIGN (<i>CD209</i>)	rs4804803	19:7812733	Upstream gene variant	G	0.616	0.0004
<i>OAS1</i> splicing	rs10774671	12:113357193	Splice acceptor variant	G	0.603	0.0010
<i>OAS3</i> K18R	rs1859330	12:113376388	Missense (Arg18Lys)	G	0.607	0.0010
<i>OAS3</i> S381R	rs2285933	12:113386779	Missense (Ser381Arg)	G	0.724	0.0001
FCgRIIa (<i>CD32</i>)	rs1801274	1:161479745	Missense (His166Arg)	G	0.646	0.0001
<i>MICB</i>	rs3132468	6:31475486	Intron variant	C	0.486	0.6919
<i>MICB</i>	rs3134899	6:31473286	Intron variant	C	0.585	0.0029

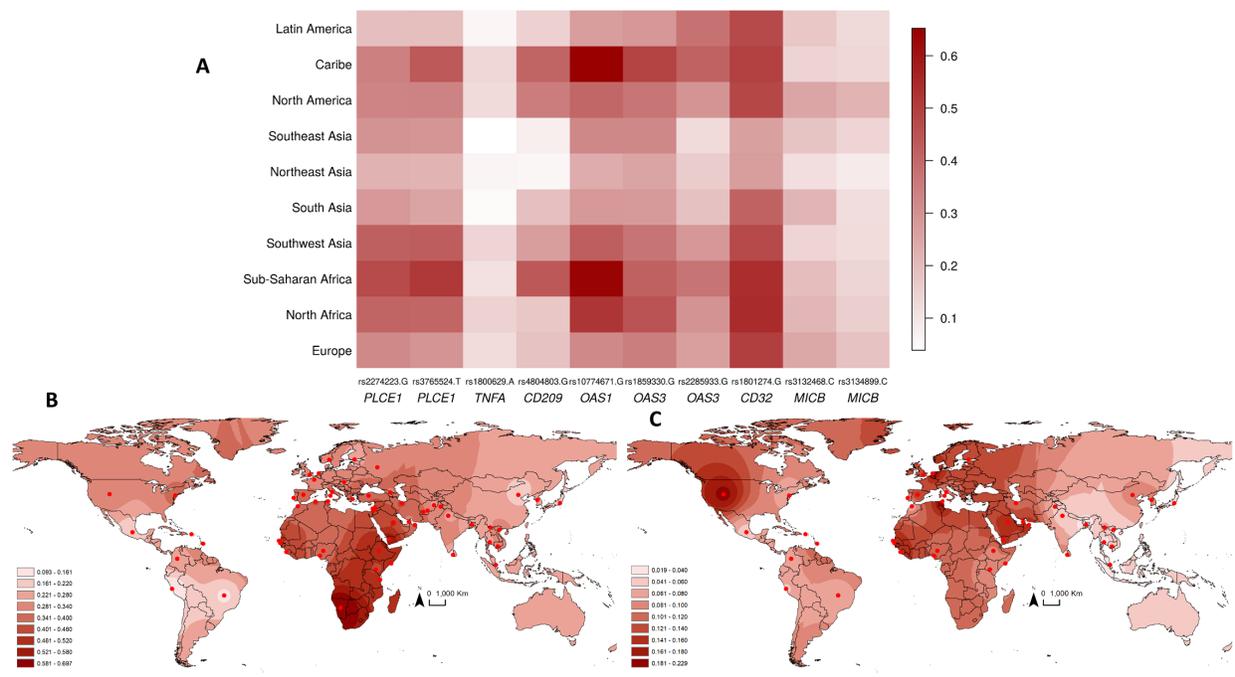
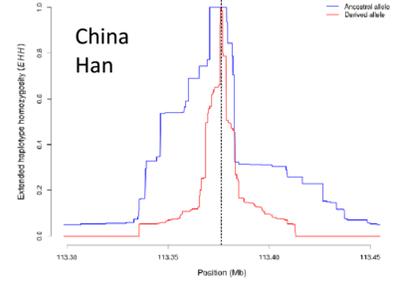
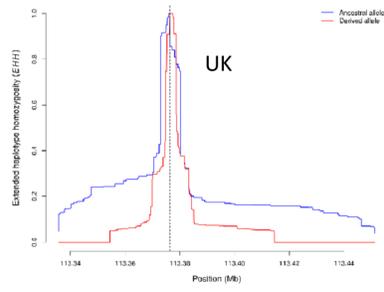
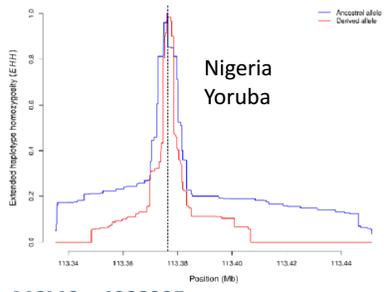


Figure 1

PLCE1-rs2274223



MCM6-rs4988235

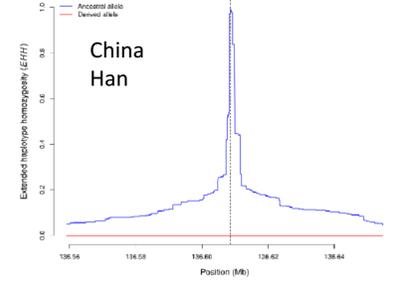
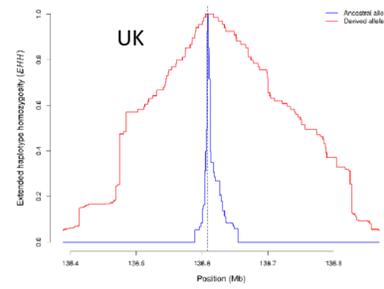
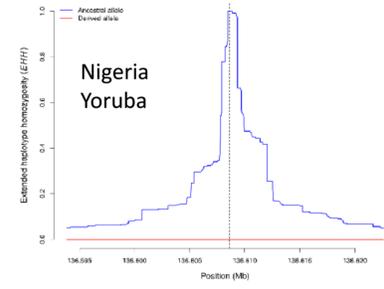


Figure 2

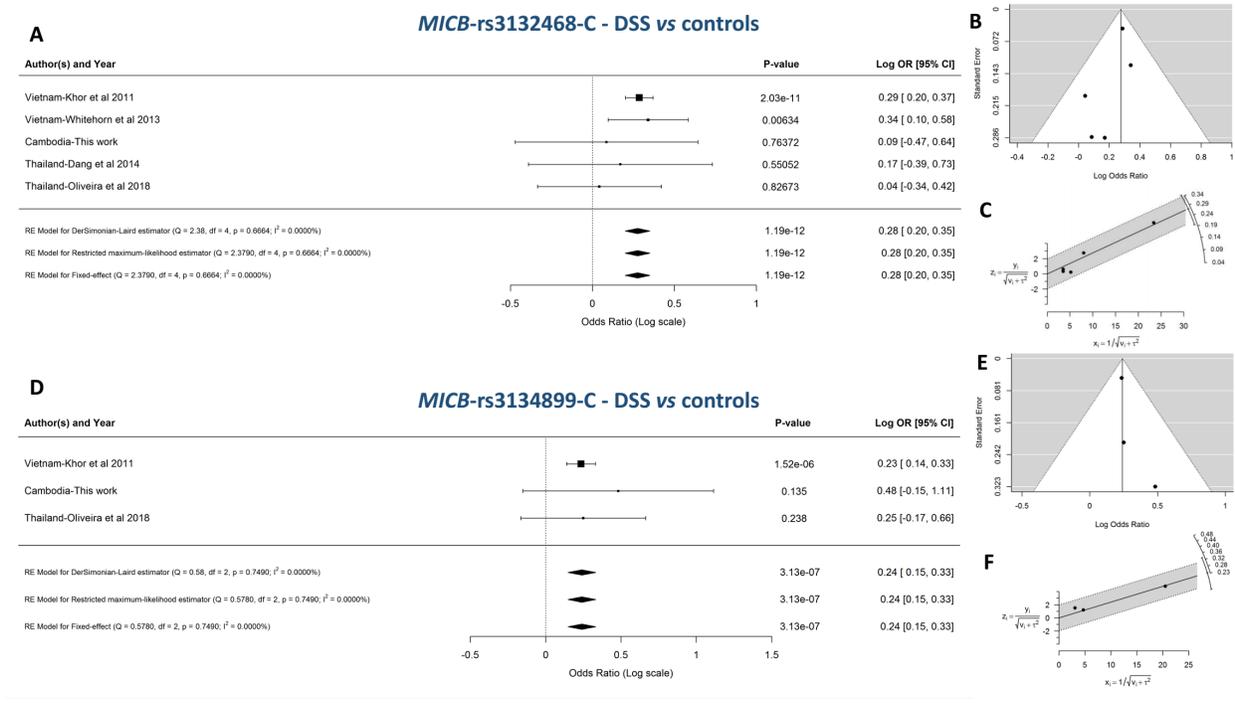


Figure 3

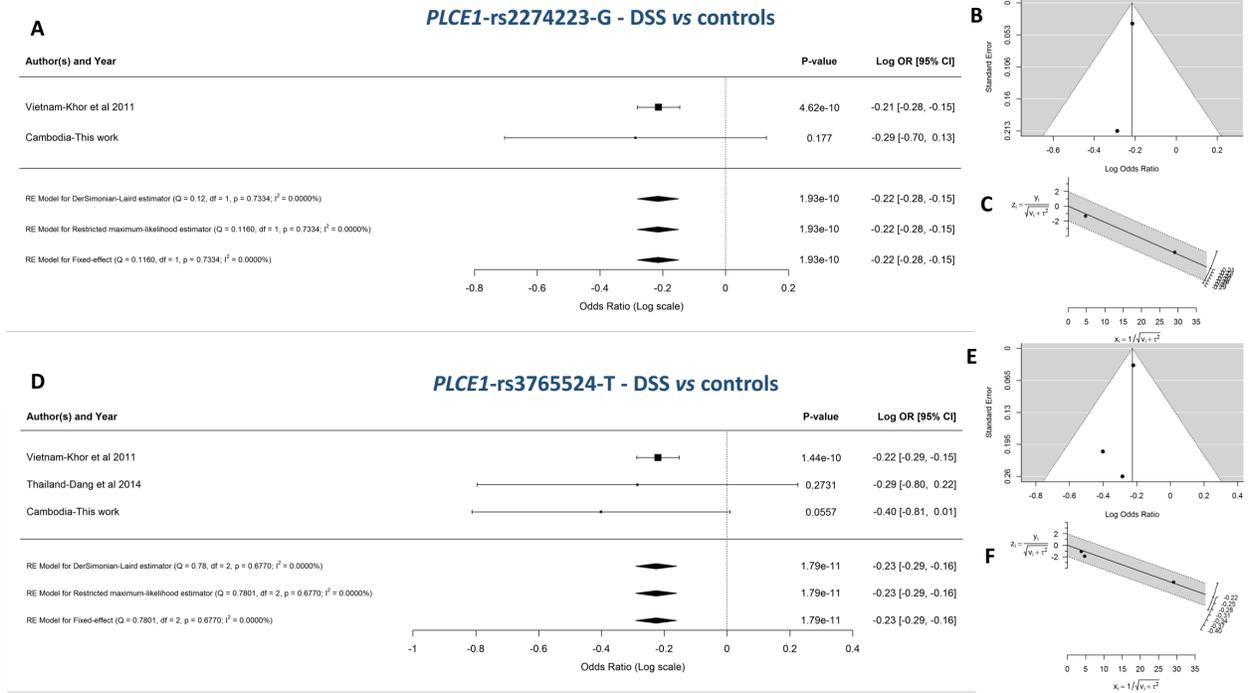
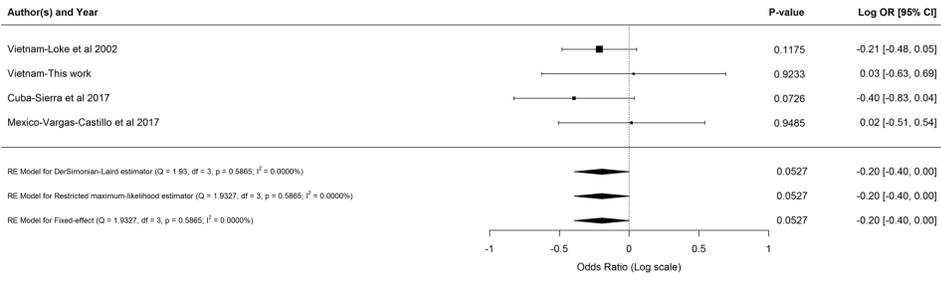


Figure 4

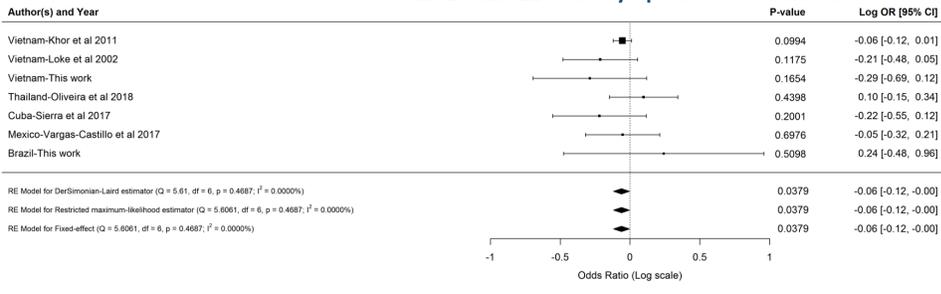
A

CD32-rs1801274-G - DHF vs controls

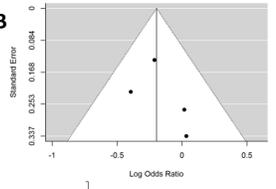


D

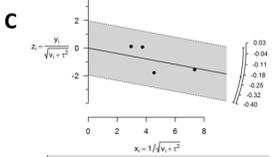
CD32-rs1801274-G - Symptomatic vs controls



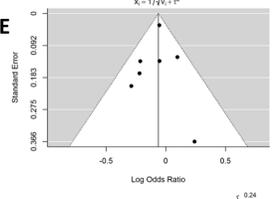
B



C



E



F

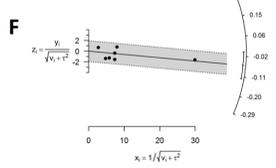


Figure 5

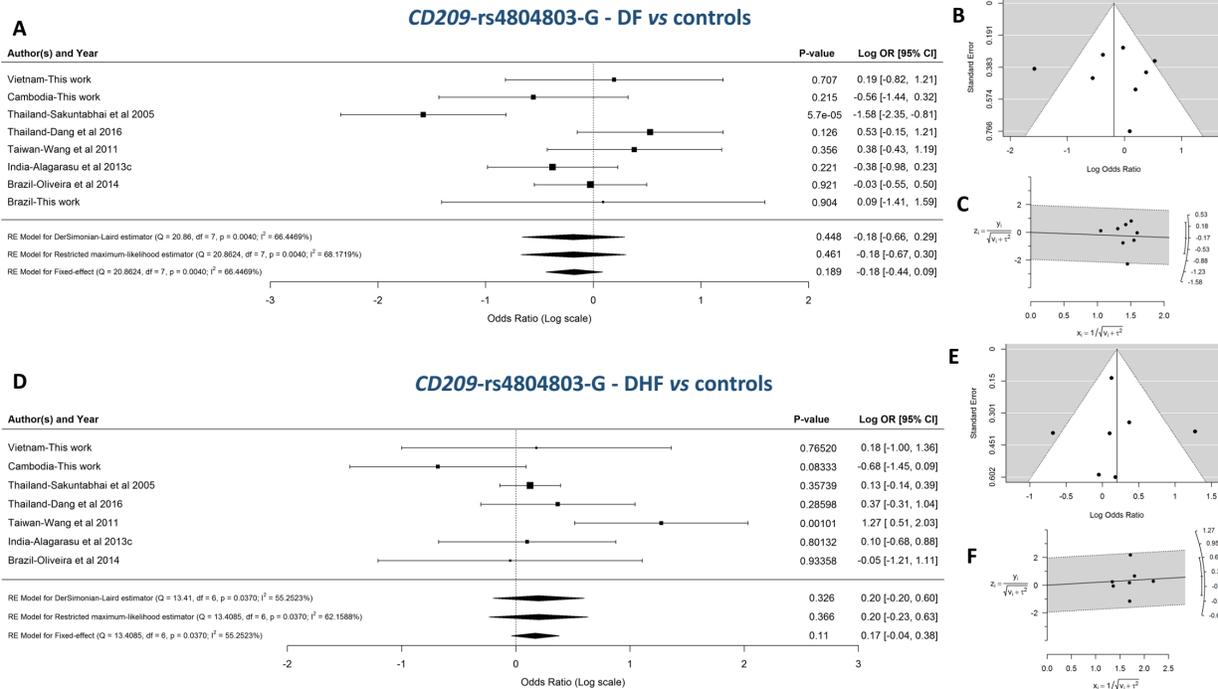


Figure 6

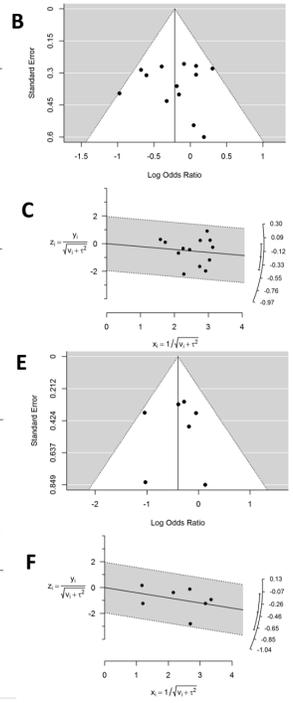
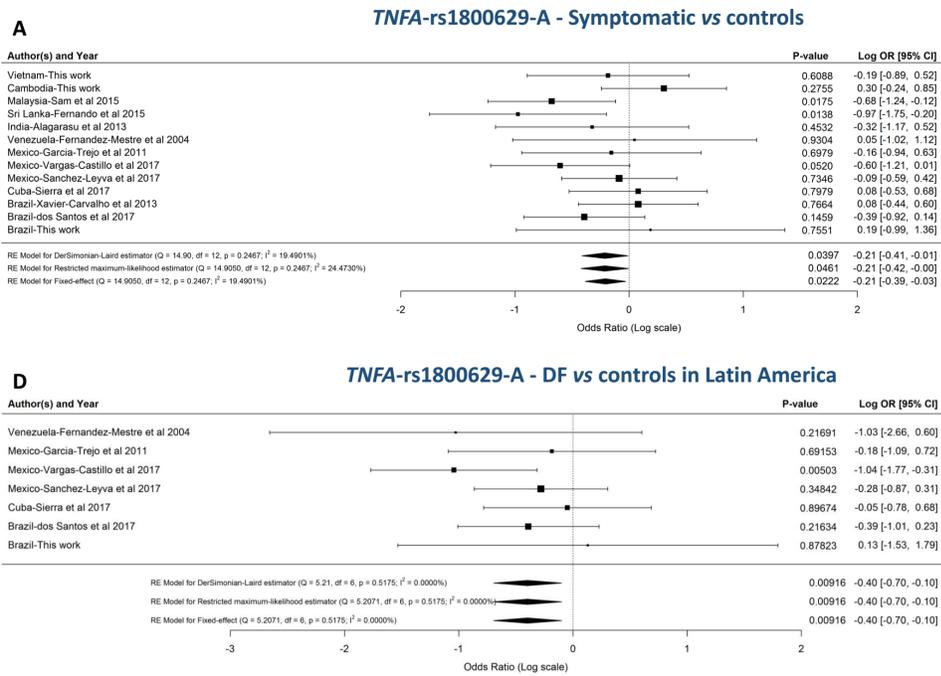


Figure 7

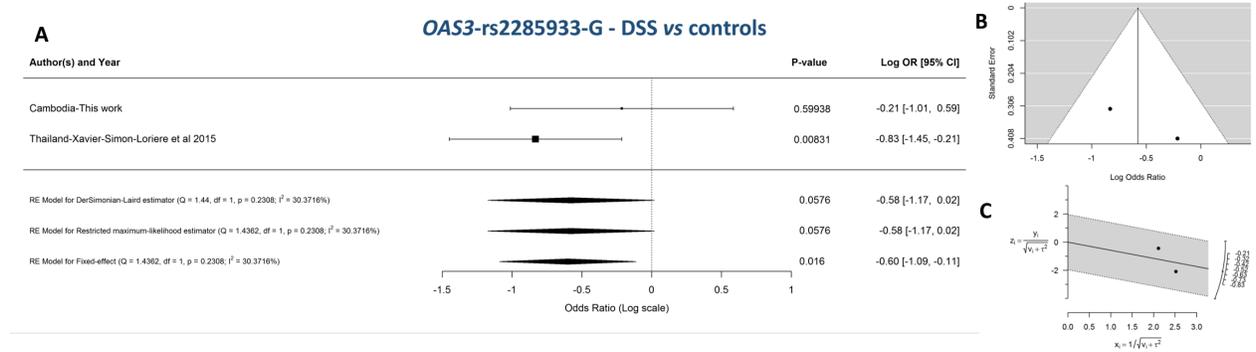


Figure 8

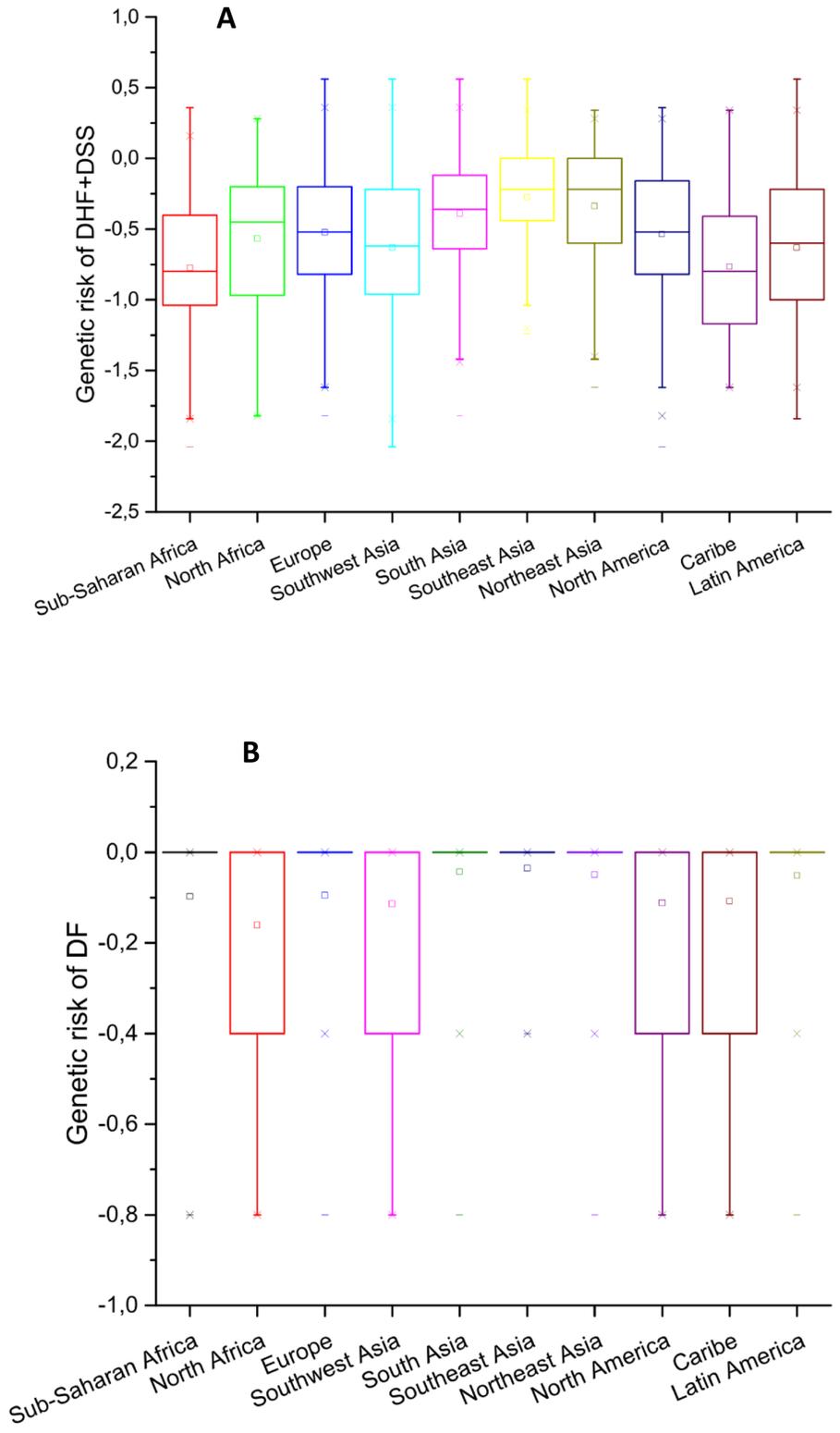


Figure 9