

Polymorphisms in host genes encoding NOSII, C-reactive protein, and adhesion molecules thrombospondin and E-selectin are risk factors for *Plasmodium falciparum* malaria in India

K. Kanchan¹ · S. S. Pati² · S. Mohanty² · S. K. Mishra² · S. K. Sharma³ · S. Awasthi⁴ · Indian Genome Variation Consortium⁵ · V. Venkatesh⁴ · S. Habib¹

Received: 11 February 2015 / Accepted: 1 July 2015 / Published online: 21 July 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Cytoadherence of *Plasmodium falciparum*-infected red blood cells (RBCs) in host microvasculature and complex regulation of the immune response are important contributors to the clinical outcome of disease. We tested the association of 23 single nucleotide polymorphisms (SNPs) and a microsatellite repeat in adhesion molecule genes *THBS1* and *ESEL*, and immune regulatory molecule genes *NOSII*, *CRP*, and *MBL2* with falciparum malaria in populations residing in a malaria-endemic and a non-endemic region of India. The *THBS1* haplotype CCCCA (rs1478604, rs7170682, rs2664141, rs12912082, rs3743125) was a risk factor in the endemic region (relative risk=3.78) and an *ESEL* SNP (rs5368, His468Tyr) associated with cerebral malaria (CM) [CM vs. non-cerebral malaria (NCM), odds ratio (OR)=2.23, $p=0.03$]. In the non-endemic region, an *ESEL* 3'UTR SNP (rs5359) associated with enhanced risk of disease (OR=3.62, $p=1 \times 10^{-4}$) and the CT genotype of the *CRP* promoter SNP (C/T/A) strongly associated with protection (severe vs. control, OR=0.29, $p=6 \times 10^{-5}$). Long repeat alleles of the *NOSII* promoter microsatellite (CCTTT)_n exhibited strong association with protection and the *NOSII* ATG haplotype (rs3729508, rs2297520, rs9282801) was strongly protective

against severe malaria in both regions (endemic, severe vs. control, OR=0.05, $p=0.0001$; non-endemic, severe vs. control, OR=0.3, $p=1 \times 10^{-5}$). Our results suggest differential contribution of variants of the investigated genes in determining the outcome of malaria in Indian populations.

Introduction

Plasmodium falciparum infection in humans is often characterized by severe clinical outcomes which have been associated with cytokine imbalance, high parasitemia, acidosis, and respiratory distress, as well as cerebral and non-cerebral cytoadherence of infected red blood cells (iRBCs) in the microvasculature. Cytoadherence is mediated by specific interactions between parasite ligands on the surface of iRBCs and host endothelial receptors, and helps the parasite evade immune clearance in the spleen [1]. iRBCs express cytoadherent ligands such as PfEMP1, clag, rifin, surfin, sequestrin, and modified band 3 protein, which play an important role in parasite–host interactions [2]. Human endothelial receptors implicated in cytoadherence are ICAM-1, PECAM-1, VCAM-1, CD36, selectins, thrombospondin (THBS), and glycosaminoglycans [2]. iRBCs also interact with a variety of immune cells [2], and immune responses to malaria are regulated by both the innate and adaptive immune system. These are critical for controlling *Plasmodium* infection by the stimulation of pathways which are mediated by inflammatory cytokines. Immune response pathways also involve downstream mediators known as immune regulatory molecules, such as cell surface receptors, soluble proteins, and signaling molecules. Genetic variations in genes encoding immune regulatory molecules influence disease outcome and severity of malaria [3]. Additionally, polymorphisms in genes encoding adhesion molecules ICAM-1, PECAM-1, CD36, and the “rosetting” factor

✉ S. Habib
saman.habib@gmail.com; saman_habib@cdri.res.in

¹ Division of Molecular and Structural Biology, CSIR—Central Drug Research Institute, Sector 10, Jankipuram Extension, Sitapur Road, Lucknow 226031, India

² Ispat General Hospital, Rourkela, India

³ National Institute of Malaria Research, New Delhi, India

⁴ King George's Medical University, Lucknow, India

⁵ CSIR—Institute of Genomics and Integrative Biology, Delhi, India

CR1 have shown differential patterns of susceptibility/resistance to malaria in world populations [3, 4].

THBS is a receptor of *P. falciparum* parasitized erythrocytes; it binds iRBCs in static assays and iRBCs interact with endothelial cells via THBS under flow conditions [5]. However, the iRBC ligand for THBS is controversial, with PfEMP1, erythrocyte-derived phosphatidylserine (PS), and modified band 3 protein as possible candidates [2]. THBS1 is a matricellular glycoprotein of the thrombospondin family. Polymorphisms of *THBS1* have not been extensively studied. However, a non-synonymous mutation (Asn700Ser) of THBS1 has been linked to increased risk of premature coronary artery disease [6]. E-selectin (CD62E) is a transmembrane glycoprotein expressed on activated endothelium and is also found as soluble protein in plasma or serum [7]. The results with a *P. falciparum* laboratory strain showed static adhesion to E-selectin, although adhesion of field isolates to E-selectin was either undetectable or very low [2], suggesting only a minor role of E-selectin in cytoadherence. However, markedly elevated levels of plasma E-selectin during the course of severe cerebral *P. falciparum* malaria have been reported [8]. Although a clear physiological role for E-selectin in malaria has not been defined, its elevated levels during infection suggest association with disease. The association of *ESEL* variants with malaria has not been investigated in detail. Only a single study on Nigerian children reported *ESEL* single nucleotide polymorphism (SNP) (rs1805193 G/T) to be strongly associated with both uncomplicated and severe malaria [9].

In humans, nitric oxide (NO) is produced by three NOS enzymes: an inducible type (iNOS or NOSII), and two constitutive types, neuronal cNOS (ncNOS or NOSI) and endothelial cNOS (ecNOS or NOSIII). NOSII is induced by the action of inflammatory cytokines, such as IFN γ , IL1 β , and TNF α , and endotoxins [10]. A larger quantity of NO is produced by the induction of NOSII and may play a role in host defense against bacterial and parasitic infections. The role of NO during human malaria is not clearly understood, as some studies have described its role as protective, while others have described it as a contributor to severe disease pathology. Experimental evidence from rodent models of malaria and in vitro studies have demonstrated that NO has a protective role against *P. falciparum* malaria [11]. Studies on human subjects have also demonstrated that NO exerts its protective effects during malarial infection by maintaining endothelial function and microvascular perfusion, preventing the activation of endothelial cells, reducing cytoadherence, and enhancing the production of inflammatory cytokines [12]. On the other hand, higher levels of NO are also associated with manifestation of severe malaria, particularly cerebral malaria. Several *NOSII* gene promoter polymorphisms, rs9282799 (C/T), rs1800482 (G/C), and microsatellite repeat (CCTTT)_n, have been studied in

the context of malaria severity in African and South-East Asian populations [13–15].

Mannose-binding lectin (MBL) is a serum protein of the C-type lectin superfamily and acts as a soluble pattern recognition receptor during the innate immune response. After binding to carbohydrates of infectious agents, it activates the lectin pathway of the complement system via interaction with MBL-associated serine proteases and kills the pathogen. MBL may also directly opsonize the microorganism for phagocytosis and modulate the release of proinflammatory cytokines [16]. During malaria infection, MBL is reported to bind the surface of infected erythrocytes and merozoites [17], thus contributing to consequent innate immune response and parasite clearance. Promoter and exonic SNPs of *MBL2* have a major effect on MBL protein structure and serum concentration [18, 19]. C-reactive protein (CRP) is an acute-phase protein found in blood plasma, and its levels are markedly increased in circulation during infection and/or inflammation. The severity of malarial infection and parasite density have been correlated with high circulating levels of CRP [20], and a low plasma level of CRP has been reported in individuals having asymptomatic *P. falciparum* infection [21]. CRP plays a role in complement-mediated hemolysis of infected erythrocytes and subsequent anemia, and has also been implicated in defense against pre-erythrocytic stages of malaria [22]. CRP induces the anti-inflammatory interleukin (IL)-10, which could affect the early immune response seen in malaria infections. An upstream triallelic *CRP* SNP (rs3091244, C/T/A) influences CRP levels and has been reported to be a contributing factor to increased susceptibility to uncomplicated *P. falciparum* malaria in African populations [23, 24].

We analyzed the association of selected *THBS*, *ESEL*, *NOSII*, *MBL2*, and *CRP* gene polymorphisms in populations drawn from a disease-endemic and a non-endemic region of India in a case-control study, and report the association of specific variants and haplotypes with severity/protection from *P. falciparum* malaria.

Methods

Populations and study subjects

Analysis of the frequency distribution of selected SNPs and haplotypes of the *THBS1* gene was carried out in the existing Indian Genome Variation Consortium (IGVC) panel II [25]. This panel consisted of 552 samples from 24 ethnically and linguistically diverse populations representing tribal, caste, and religious groups from different geographical regions of India. Panel II was derived from the initial IGVC sample set (panel I) of 1871 individuals from 55 diverse populations [26]. The population descriptors included linguistic affiliation (Indo-European, IE; Dravidian, DR; Tibeto-Burman, TB;

Austro-Asiatic, AA), followed by geographical zone (North, N; North-East, NE; South, S; East, E; West, W; Central, C) and ethnicity (caste, LP; tribe, IP; religious group, SP). A population of known African descent was included as an outgroup population (OG-W-IP) in panel II.

The case–control panel comprised existing patient and control samples collected from a *P. falciparum* hyper- to mesoendemic region (Antagarh, Chhattisgarh and Sundargarh, Orissa) and a non-endemic region (Lucknow and surrounding areas of Uttar Pradesh) of India, as described previously [4, 25, 27]. The control panel comprised individuals ethnically matched with the patient group; these belonged to the Bhumij, Munda, Oraon, and Gond tribal populations in the endemic region and the Aggarwal, Brahmin, Kayastha, Pasi, Thakur, Yadav, Shia, and Sunni caste and religious groups in the non-endemic region. Patient samples comprised 221 *P. falciparum* malaria patients (101 from endemic and 120 from non-endemic regions), compared with 192 control samples (102 from endemic and 90 from non-endemic regions) from ethnically matched and unrelated individuals [25]. World Health Organization (WHO) guidelines [28] were followed to categorize severe and non-severe malaria, as described previously. Severe patients were further categorized as those with cerebral malaria (CM) and severe non-cerebral malaria (NCM). CM was characterized by impaired consciousness (coma) with fever. Any one of the following symptoms indicated severe NCM: severe anemia, acidotic breathing, pulmonary edema, hypoglycemia, and increased serum creatinine levels [25].

Genotyping

Genetic marker information was extracted from the dbSNP database (<http://www.ncbi.nlm.nih.gov/SNP>). The IGVC sample set had been genotyped by Sequenom mass spectroscopy [4]. Genotyping of DNA samples from the blood of patients and controls was done by either the Illumina-based genotyping platform or SNaPshot analysis (Applied Biosystems) on an ABI 3130xl automated DNA analyzer.

Promoter and intronic *THBS* SNPs rs1478604, rs7170682, rs2664141, and rs12912082 were typed on the Illumina platform, while *THBS* rs2228262 and rs3743125 were genotyped by SNaPshot. One *ESEL* SNP rs5361 was typed on Illumina, while SNPs rs1805193, rs5355, rs5368, and rs5359 were genotyped by multiplex SNaPshot. *NOSII* SNPs rs3729508 and rs9282801 were genotyped by multiplex SNaPshot, rs9282799 by Illumina, and rs1800482 and rs2297520 by restriction fragment length polymorphism (RFLP) using *BsaI* and *NgoMIV* restriction enzymes, respectively. *MBL2* SNPs rs11003125, rs7096206, and rs1800450 were genotyped by multiplex SNaPshot and rs7095891, rs5030737, rs1800451, and rs4647964 were genotyped by Sanger DNA sequencing. The *CRP* triallelic SNP rs3091244 was genotyped by the SNaPshot method.

Genotyping of the microsatellite repeat lying in the promoter of *NOSII* was done by polymerase chain reaction (PCR) amplification followed by capillary electrophoresis on the ABI 3130xl analyzer. Primer pairs as described by Gómez et al. [29] (forward: 5'-ACCCCTGGAAGCCTACAACCTGCAT-3'; reverse: 5'-GCCACTGCACCCTAGCCTGTCTCA-3') were used with the forward primer tagged with the NED fluorophore. Genomic DNA (50 ng) was amplified in a reaction mixture containing 0.33 μ M of each primer, 25 μ M of each dNTP, 2.5 mM $MgCl_2$, 1 \times standard PCR buffer, and 0.6 U of GoTaq Flexi DNA polymerase (Promega). The initial denaturation (95 °C for 10 min) was followed by 40 cycles of 30 s at 95 °C, 30 s at 60 °C, and 30 s at 72 °C, followed by a final extension at 72 °C for 45 min. PCR fragments, together with the GeneScan LIZ-500 internal size standard (Applied Biosystems), were loaded onto the ABI 3130xl analyzer. GeneMapper analysis software v.3.7 (Applied Biosystems) was used to determine the size of individual fragments of the (CCTTT)_n microsatellite repeat.

Statistical analysis

Genotype and allele frequencies of SNPs were determined by direct counting. Hardy–Weinberg equilibrium (HWE) was assessed by Chi-square analysis (df=1) for each SNP. Average heterozygosity (H_o) and F_{st} values for SNPs were calculated by *Fst_win_vii* software [30]. The odds ratio (OR) and relative risk (RR) for risk assessment were estimated using Epi Info™ version 3.4 software that calculates *p*-values by Fisher's exact or the Mantel–Haenszel test. Haplotypes and linkage disequilibrium (LD) plots were generated using PHASE and HaploView software, respectively. Haplotype frequencies from genotype data were calculated by direct counting, and association of haplotypes with disease was also estimated by Epi Info™ version 3.4. A *p*-value <0.05 was considered statistically significant for all analyses. For multiple testing, Bonferroni correction was applied by multiplying the observed (uncorrected) *p*-value by the number of comparisons, and the resultant adjusted *p*-value was considered significant if it was <0.05. The prediction of transcription factor binding sites was done by using the online tool PROMO version 3.0.2.

Results and discussion

Selection of genetic variants for disease association

Genetic variants to be analyzed were selected depending upon their reported functional relevance in *falciparum* malaria, as well as other infectious diseases in other world populations. Minor allele frequencies (MAFs) of a total of 23 SNPs, six from *THBS1*, five from *ESEL*, five from *NOSII*, six from *MBL2*, and one from *CRP*, and a pentanucleotide repeat from

NOSII promoter were estimated in controls (C) as well as non-severe (NS) and severe (S) falciparum malaria patients from both a malaria-endemic region and a non-endemic region of India (Table 1).

Since there is no information available in the context of *THBS1* polymorphisms and their association with malaria, we selected SNPs from the entire gene depending upon their reported frequency in IGVC panel II [4, 25] and other Asian populations. Five *THBS* SNPs (rs1478604, rs7170682, rs2664141, rs12912082, rs3743125) with a mean MAF >20 % in IGVC panel II and Han Chinese and Japanese populations were genotyped. The five *THBS1* SNPs from IGVC panel II showed a wide range of MAF distributions in the 24 populations (Fig. 1). High heterozygosity (H_o) and low F_{st} values (rs1478604, $H_o=0.384$, $F_{st}=0.028$; rs7170682, $H_o=0.333$, $F_{st}=0.005$; rs266414, $H_o=0.386$, $F_{st}=0.006$; rs12912082, $H_o=0.323$, $F_{st}=0.015$; and rs3743125, $H_o=0.325$, $F_{st}=0.001$) of these SNPs may be indicative of balancing selection. A possible functional

non-synonymous polymorphism in *THBS1* exon 13 (rs2228262, A/G, Asn700Ser) in the most N-terminal calcium-binding repeat of THBS1 [6] was also included (Table 1). SNPs of the *ESEL* gene were selected on the basis of their reported functional relevance in other diseases. E-selectin consists of an N-terminal domain, epidermal growth factor (EGF)-like domain, lectin (ligand-binding) domain, variable short consensus repeats, and transmembrane domain, and recognizes its ligands through the carbohydrate recognition domain (CRD) (Fig. 2). *ESEL* exon 4 rs5361 (A/C, Ser/Arg) lies in the EGF-like domain and may cause altered ligand binding, exon 9 rs5368 (C/T, His/Tyr) is located in the fifth short consensus repeat domain, and exon 11 SNP rs5355 (C/T, Leu/Phe) lies in the transmembrane domain. Two regulatory SNPs from the 5' and 3'UTR of *ESEL* were also genotyped (Table 1). Exon 13 SNP of *THBS1* had a very low MAF (2 %) in control samples from both the endemic and non-endemic regions (Table 1), and was not considered further.

Table 1 Minor allele frequency (MAF) of *THBS1*, *ESEL*, *NOSII*, *MBL2*, and *CRP* polymorphisms in the malaria-endemic region (ER) and non-endemic region (NER) of India

rsID	DNA change	Position	MAF (%), ER			MAF (%), NER			
			C, n=101	NS, n=76	S, n=25	C, n=90	NS, n=14	S, n=106	
<i>THBS1</i>									
rs1478604	C/T	Promoter	39	48	40	50	33	61	
rs7170682	C/T	Intron 5	40	27	39	24	32	20	
rs2664141	C/T	Intron 7	49	46	48	46	39	39	
rs2228262	A/G	Exon 13	2	–	–	2	–	–	
rs12912082	C/T	Intron 13	41	29	36	22	27	22	
rs3743125	G/A	3'UTR	24	35	38	14	21	13	
<i>ESEL</i>									
rs1805193	G/T	Exon 2/5'UTR	9	15	19	7	21	11	
rs5361	A/C	Exon 4	8	9	7	7	11	4	
rs5368	C/T	Exon 9	13	7	19	15	8	11	
rs5355	C/T	Exon 11	14	21	12	11	8	14	
rs5359	A/G	Exon 14/3'UTR	28	35	24	15	36	39	
<i>NOSII</i>									
	(CCTTT) _n	Promoter	31	5	3	45	29	23	
rs3729508	G/A	Intron 7	32	41	35	35	46	35	
rs2297520	T/C	Intron 8	23	52	63	41	57	61	
rs9282801	G/T	Intron 16	29	28	22	33	25	28	
<i>MBL2</i>									
rs11003125	C/G	Promoter	26	25	28	28	38	28	
rs7096206	C/G	Promoter	20	17	14	21	17	18	
rs1800450	G/A	Exon 1	14	14	10	11	18	16	
<i>CRP</i>									
rs3091244	C/T/A	Promoter	T	19	27	36	35	13	32
			A	3	3	5	3	17	12

C controls; NS non-severe patients; S severe patients

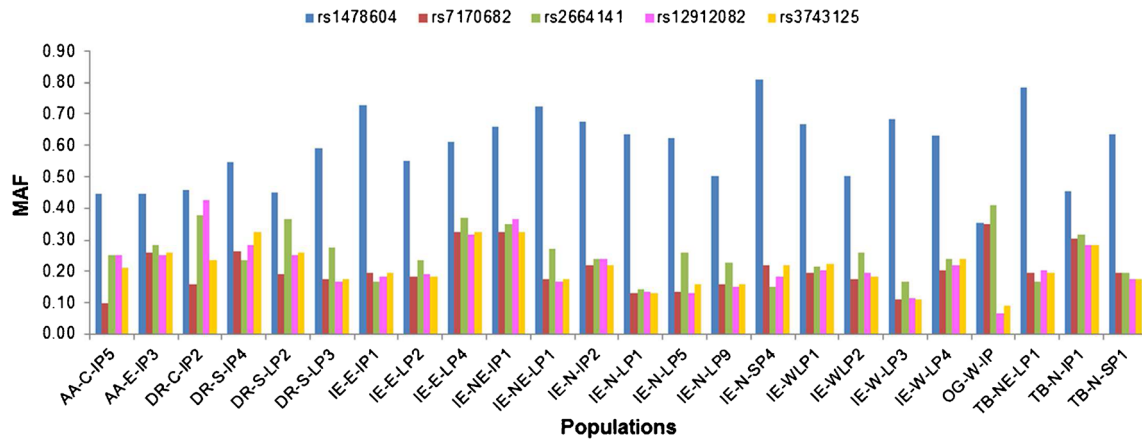


Fig. 1 Minor allele frequency (MAF) of *THBS1* single nucleotide polymorphisms (SNPs) across 24 populations of India [Indian Genome Variation Consortium (IGVC) panel II]

Of the *NOSII* SNPs, distal promoter SNPs rs9282799 and rs1800482 were found to be monomorphic in the control panel and were, thus, excluded from the study. For SNPs from the intergenic region of *NOSII*, we inspected IGVC panel I data from 53 endogamous populations of India [26] and identified three intronic SNPs with high mean frequency of the minor allele (mean MAFs of 0.29, 0.29, and 0.33 for rs3729508, rs2297520, and rs9282801, respectively) for inclusion in the case–control study.

All six SNPs selected for *MBL2* have been reported to have a major effect on MBL levels or protein structure [31]. Among these, rs7095891 was found to be monomorphic in our samples. Two non-synonymous SNPs, rs5030737 (Arg/Cys) and rs1800451 (Gly/Glu), also had a very low frequency (<5 %) among control groups and were not considered further. Of the three *MBL2* SNPs (Table 1) that were analyzed for association with malaria susceptibility/resistance, none showed association with disease in either the endemic or the non-endemic region. Studies examining the relationship between malaria and *MBL2* variants in African populations have produced conflicting results [18, 31–33]. Although the two missense mutations of exon 1, Gly54Asp (rs1800450) and Gly57Glu (rs1800451), are associated with risk of severe malaria in Gabonese [33] and Ghanaian populations [31], the former did not exhibit disease association in our study and the latter had very low frequency. The single triallelic SNP selected for

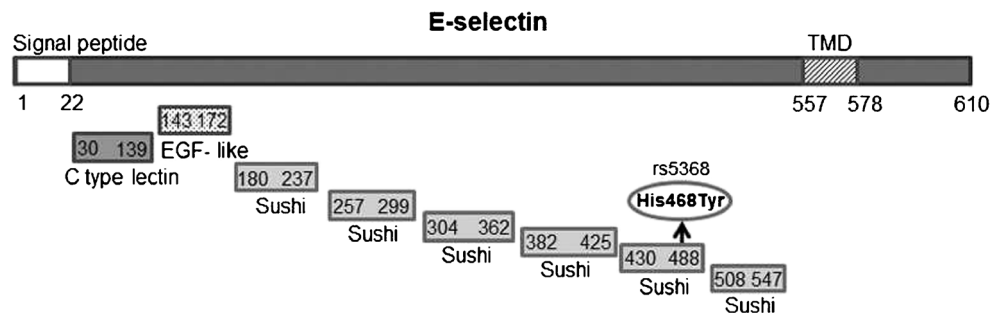
the *CRP* gene is known to effect CRP plasma levels and has been associated with malaria [23, 24].

Association of *THBS1* polymorphisms with susceptibility to falciparum malaria

Out of the five *THBS1* SNPs investigated for disease association, the 3’UTR SNP rs3743125 G/A exhibited association with malaria in the endemic region. Genotypes with the mutant A allele were significantly associated with manifestation of disease [AG+AA and GG, severe vs. control, OR=3.06, 95 % confidence interval (CI) = 1.13–8.65, $p=0.01$; non-severe vs. control, OR=3.37, 95 % CI = 1.66–6.87, $p=2 \times 10^{-4}$; severe vs. non-severe, OR=0.91, 95 % CI = 0.32–2.74, $p=0.84$], with no correlation with disease observed in the non-endemic region. The p -value in the endemic region was significant after Bonferroni correction. Interestingly, this SNP deviates from HWE in the endemic region (p -value=0.008), although it is in HWE in the non-endemic region, suggesting that the locus is under selection pressure in the malaria-endemic region. No significant association with disease was observed for the other individual *THBS1* SNPs.

The 3’UTR region often contains specific transcription factor binding sites contributing to the regulation of gene expression. Sequence analysis of the *THBS1* 3’UTR region around rs3743125 revealed the presence of DNA sequences

Fig. 2 Diagrammatic representation of domains of E-selectin and position of the rs5368 SNP

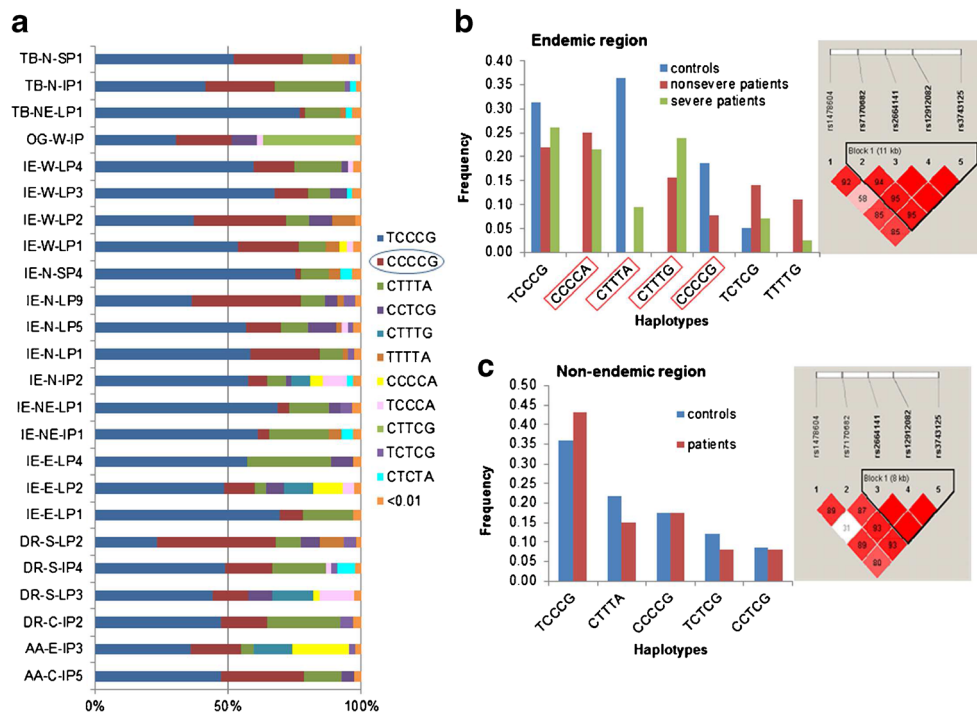


homologous to the binding sites of several regulatory proteins. Binding sites for three factors, c-ETS-1 (tG/Aggaag), ELK-1 (gttG/Aggaag), and c-MyB (ttcagtG/A), were common in both variants. Perfectly matched binding sites for five additional transcription factors- C/EBP-beta, NF-1, NFκB, FOXP3, and NF1/CTF were found in the G variant, and the A variant was predicted to have binding sites for four other transcription factors- GR-alpha, c-ETS-2, HNF-1C, and HNF-1B. Although small nucleotide changes in a regulatory region may not have dramatic effects on gene expression, they may fine tune regulatory functions, thus resulting in differences. The change in binding sites for transcription factors due to the *THBS1* 3'UTR SNP might have a broad range of phenotypic consequences.

Genotype data of five *THBS1* SNPs from IGVC panel II populations generated 32 haplotypes among which 11 haplotypes had a frequency >0.01 (Fig. 3a). The wild-type haplotype CCCCCG (-27 rs1478604 T/C, +3064 rs7170682 C/T, +6169 rs2664141 C/T, +9142 rs12912082 C/T, +14500 rs3743125 G/A) was not the predominant haplotype in all IGVC populations; TCCCG and CTTTA were the other major haplotypes. Haplotypes and pairwise LD plots for the five *THBS1* SNPs were generated from genotype data of controls and patient groups in the endemic and non-endemic regions (Fig. 3b, c). The three intronic SNPs (rs7170682 C/T, rs2664141 C/T, and rs12912082 C/T) and the 3'UTR SNP (rs3743125 G/A) were in strong LD and formed a haplotype block of ~11 kb (Fig. 3b) in the endemic region, whereas a haplotype block of ~8 kb was formed by rs266414, rs12912082, and rs3743125 in the non-endemic region (Fig. 3c). Haplotypes with a frequency >5 % in

either controls or patients were analyzed for association with malaria. The distribution of the CCCCCA haplotype which carries the risk allele A at the 3'UTR SNP was significantly different between control and patient categories in the endemic region (severe vs. control, $\chi^2=21.85$, $p=3 \times 10^{-6}$; non-severe vs. control, $\chi^2=26.5$, $p=3 \times 10^{-7}$) and had a relative risk value of 3.91 for severe and 3.78 for non-severe malaria. The p -values were significant after Bonferroni correction. The CCCCCG haplotype with the protective G allele at the 3'UTR SNP also exhibited differential distribution in the control and patient categories (severe vs. control, $\chi^2=8.99$, $p=0.002$; non-severe vs. control, $\chi^2=3.71$, $p=0.05$), although the non-severe vs. control comparison was not significant after Bonferroni correction; none of the severe patients in the endemic region carried this haplotype (Fig. 3b). These results are in agreement with the result of individual SNP genotype association of rs3743125 with disease manifestation. The CTTTG haplotype containing the three intronic SNP mutants was also differentially distributed in patients and controls (severe vs. control, $\chi^2=24.46$, $p=8 \times 10^{-7}$; non-severe vs. control, $\chi^2=18.92$, $p=1.3 \times 10^{-6}$) and conferred a relative risk of 4.0 for severe malaria and 3.18 for non-severe malaria. This suggests that, although intronic SNPs do not show any association individually, they have a combined effect in influencing disease outcome. Additionally, the CTTTA haplotype, which had mutant T allele at all three intronic SNPs together with the mutant A allele at the 3'UTR SNP, conferred protection from malaria (severe vs. control, $\chi^2=10.38$, $p=0.001$; non-severe vs. control, $\chi^2=29.68$, $p=1 \times 10^{-7}$); none of the non-severe patients in the endemic region carried this haplotype. Comparisons for the above haplotypes were significant

Fig. 3 **a** Distribution of *THBS1* haplotypes for the SNPs rs1478604, rs7170682, rs266414, rs12912082, and rs3743125 across 24 Indian populations. The wild-type haplotype is encircled. **b** Comparison of *THBS1* haplotype frequencies among control, non-severe, and severe patient groups of the endemic region, together with r^2 linkage disequilibrium (LD) plots for controls of the region. **c** *THBS1* haplotype frequencies among the control and severe patient groups of the non-endemic region. The insets show r^2 LD plots for controls of the region. The value in each cell is percent D' between SNP pairs



after Bonferroni correction. This suggests that even though the 3'UTR A allele is a risk factor for disease individually as well as in combination with some other SNPs, its effect may change dramatically in combination with T alleles of the intronic SNPs. No *THBS1* haplotype–disease correlation was found in the non-endemic region (Fig. 3c).

A regulatory region *ESEL* SNP is strongly associated with severe disease in the non-endemic region

Among the five SNPs investigated for the *ESEL* gene, rs5368 (exon 9, C/T) and rs5359 (3'UTR, A/G) showed association with disease in the endemic and non-endemic regions, respectively. Since patient samples in the non-endemic region were collected from a tertiary hospital setting, the number of non-severe patients was low. All comparisons were, thus, made between controls and severe patients for this region. The T allele of rs5368 was associated with susceptibility to cerebral malaria in the endemic region (CM vs. control, OR=2.23, 95 % CI=1.01–5.09, $p=0.03$; CM vs. non-severe, OR=4.43, 95 % CI=1.73–12.72, $p=5 \times 10^{-4}$; CM vs. NCM, OR=2.23, 95 % CI=1.01–5.09, $p=0.03$; NCM vs. control, OR=1, 95 % CI=0.4–2.49, $p=1$; non-severe vs. control, OR=0.5, 95 % CI=0.16–1.44, $p=0.15$; NCM vs. non-severe, OR=1.99, 95 % CI=0.7–6.14, $p=0.15$). However, the p -values for the CM vs. control and CM vs. NCM comparisons were not significant after Bonferroni correction. No association of this locus was found with cerebral malaria in the non-endemic region. The G allele of 3'UTR SNP rs5359 was associated with disease in the non-endemic region (severe vs. control, OR=3.62, 95 % CI=1.76–7.69, $p=1 \times 10^{-4}$). This association was even more pronounced when analysis was done with genotype count (severe vs. control, GG+AG and AA, OR=7.0, 95 % CI=3.53–13.96, $p=1 \times 10^{-7}$). p -Values for all comparisons for this SNP were significant after Bonferroni correction.

Association of the exon 9 rs5368 SNP of *ESEL* has been previously reported with hypertension in males in a Chinese population [34] and with immunoglobulin A nephropathy in Japan [35]. The SNP lies in the fifth short consensus repeat (sushi repeat) domain of the extracellular region of the protein (Fig. 2) and the His468Tyr variant may alter protein conformation, thus affecting its interactions with other proteins/ligands. However, this hypothesis remains to be validated. Association of the *ESEL* 3'UTR rs5359 SNP has not been reported previously for any disease. Analysis of the DNA sequence around rs5359 showed the presence of two overlapping GATA-1 transcription factor binding sites in the A variant. Interestingly, the core motif of the first GATA-1 site is altered by the A to G change (cagatA to cagatG), resulting in abolition of the first GATA-1 site in the G variant. The polymorphism is, thus, predicted to cause a change in the regulation of *ESEL* expression.

NOSII large (CCTTT)_n microsatellite repeats are associated with protection from disease manifestation

Genotyping of the microsatellite repeat in the promoter of *NOSII* revealed the presence of ten alleles, (CCTTT)_{7–17}, in both the endemic and non-endemic regions. The frequency of each microsatellite allele in the control and patient groups is shown in Fig. 4a, b. (CCTTT)₁₂ was the predominant allele in the controls of both regions. We, thus, categorized the repeats into small repeats of ≤ 12 (SR, CCTTT_{7–12}) and large repeats of ≥ 13 (LR, CCTTT_{13–17}) to analyze the association of the repeat polymorphism with disease susceptibility/resistance. Large repeats CCTTT_{13–17} exhibited association with protection from disease manifestation in both the endemic ($p < 1 \times 10^{-5}$) and non-endemic regions ($p < 1 \times 10^{-3}$) (Fig. 4c). The p -values were significant after Bonferroni correction.

Conflicting results have been reported in previous in vivo and in vitro studies of the (CCTTT)_n polymorphism, with shorter repeats (<11 repeats) being associated with fatal cerebral malaria in Gambian children [15] and longer forms (≥ 15 repeats) associated with severe malaria in Thai adults [36]. No association was observed between (CCTTT)_n repeat alleles and malaria disease severity in Kenyan children [14] and with malaria severity or NO production in Tanzanian or Gabonese children [37]. A case-only study on patients from India reported an association of longer forms of (CCTTT)_n (>13 repeats) with severe malaria [38]. Apart from population-specific factors, the apparent differences in the above results may partly be explained by the inconsistency in the classification of long/short repeats. We selected 12 as the cut-off length as the repeat copy number of 12 was maximally represented in the controls of both regions. Although African populations have a bimodal distribution of repeat length [15], (CCTTT)₁₂ has the highest frequency in Caucasian and Thai populations [15, 36]. Our results indicate that longer repeats (≥ 13) have a significant protective effect in populations residing in both a malaria-endemic and a non-endemic region of India.

The number of microsatellite repeats in the *NOSII* promoter region affects the level of iNOS transcripts and (CCTTT)_{14, 15, 17} microsatellite repeat alleles are strong enhancers of promoter activity, with enhancement being significantly greater than that with (CCTTT)_{12, 9} [39]. This would imply higher production of NO when the *NOSII* promoter carries long repeats. Since NO is toxic to malaria parasites in vitro and mediates host-protective effects in rodent models of malaria [11, 40], higher NO levels resulting from high copy number repeats offers an explanation for the association of large (CCTTT)_n repeats with protection against *P. falciparum* malaria observed by us.

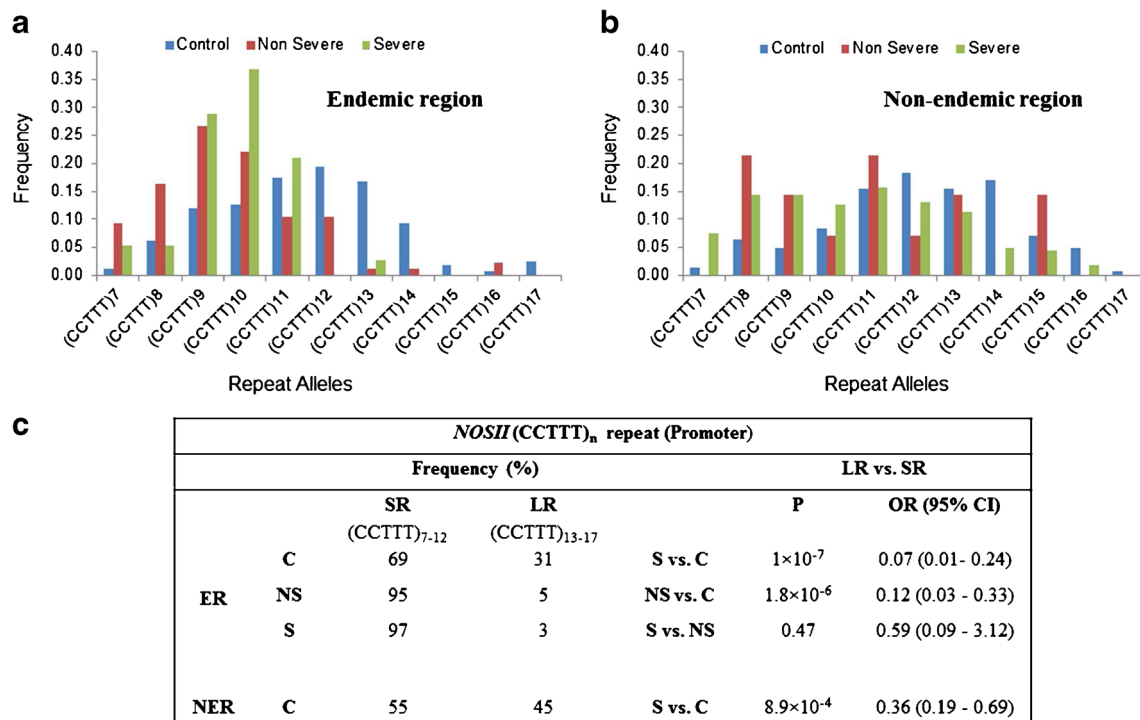


Fig. 4 Frequency distribution of *NOSII* (CCTTT)_n microsatellite repeat alleles in patient and control sets of the (a) endemic and (b) non-endemic regions. c Association analysis of short and long repeats with malaria susceptibility

Association of *NOSII* rs2297520 C allele and haplotypes with susceptibility to malaria

Of the three *NOSII* SNPs analyzed for disease association, the minor allele C of the intron 8 SNP (rs2297520, T/C) exhibited strong association with disease manifestation in both the endemic (severe vs. control, OR=5.7, 95 % CI=2.96–11.12, $p=1 \times 10^{-8}$; non-severe vs. control, OR=3.63, 95 % CI=1.89–7.01, $p=2 \times 10^{-5}$; severe vs. non-severe, OR=1.57, 95 % CI=0.86–2.88, $p=0.11$) and non-endemic region (severe vs. control, OR=2.25, 95 % CI=1.23–4.12, $p=0.004$). p -Values for both regions were significant after Bonferroni correction, although stronger disease association was observed in the endemic region. Sequence analysis of the *NOSII* intron 8 region flanking the rs2297520 SNP revealed the presence of DNA sequences homologous to the binding site of the transcription factor ENKTF-1. The ancestral T variant of rs2297520 is predicted to contain two overlapping binding sites for ENKTF-1. However, one binding site is abolished in the C variant, suggesting that the SNP would alter transcription levels of *NOSII*. However, this hypothesis needs to be validated.

We next investigated the association of haplotypes generated from the three intronic *NOSII* SNPs (+23526 rs3729508 G/A, +24516 rs2297520 T/C, and +36083 rs9282801 G/T) with *P. falciparum* susceptibility/resistance. Of the eight haplotypes found, those with a frequency >5 % in either control or patient groups were analyzed for disease association. The haplotype bearing ancestral alleles (GTG) was the predominant

haplotype in the controls of the endemic region (Fig. 5a), while ATG was the predominant haplotype in the controls of the non-endemic region (Fig. 5b). In the endemic region, the GCG haplotype bearing the risk allele C of rs2297520 was associated with disease manifestation (severe vs. control, OR=2.25, $p=0.05$; non-severe vs. control, OR=2.64, $p=0.003$) (Fig. 5a); this is in agreement with the results of individual SNP association analysis. Interestingly, the haplotype ACG with mutant allele A at rs3729508 and C at rs2297520 also showed very strong association with disease manifestation (severe vs. control, OR=8.11, $p=2 \times 10^{-7}$; non-severe vs. control, OR=11.41, $p=8 \times 10^{-5}$). This suggests that, even though the intron 7 SNP rs3729508 does not show association individually, it may have a combined effect with rs2297520 in influencing disease outcome. On the other hand, the ATG haplotype with ancestral T allele for rs2297520 was associated with protection from disease severity (severe vs. control, OR=0.05, $p=0.0001$; severe vs. non-severe, OR=0.05, $p=0.0001$) (Fig. 5a). p -Values were significant after Bonferroni correction.

In the non-endemic region as well, the GCG and ACG haplotypes were associated with severe disease (Fig. 5b), although the p -value for ACG was not significant after Bonferroni correction. Additionally, the ATT haplotype, which consists of mutant alleles at rs3729508 and rs9282801, also showed a strong association with disease manifestation in the non-endemic region (severe vs. control, OR=7.43, $p=0.0002$). As in the endemic region, the ATG

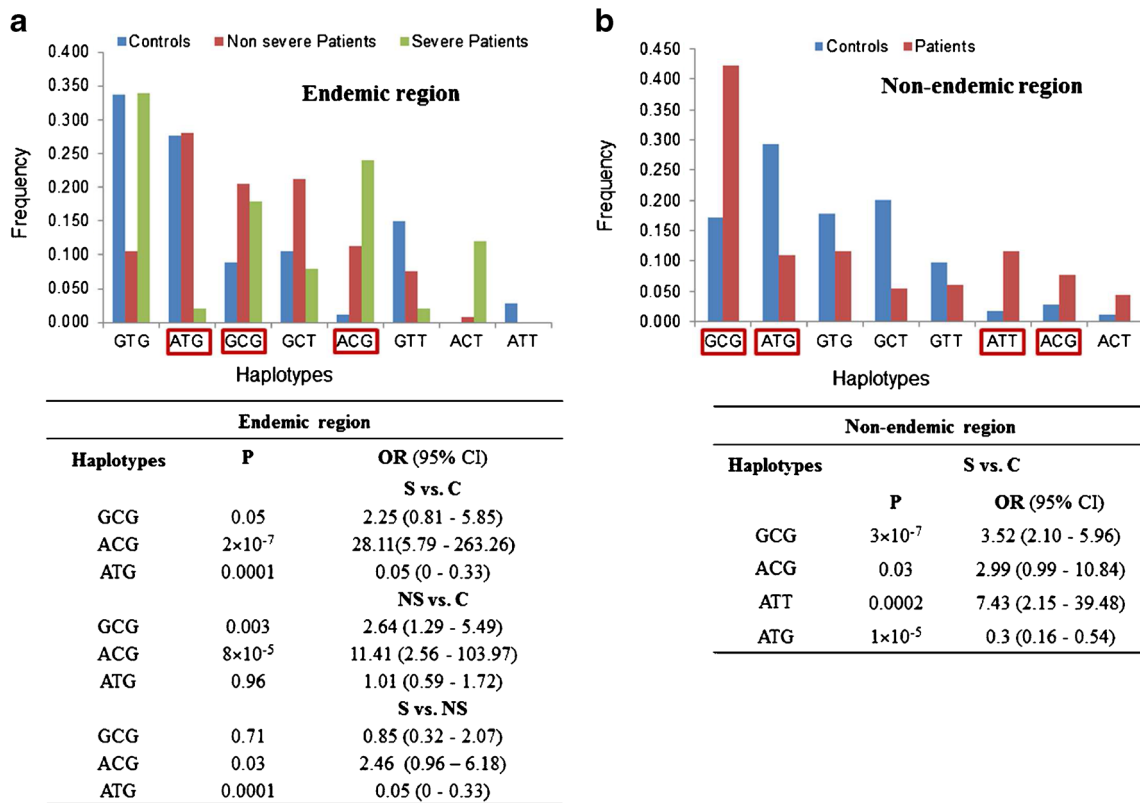


Fig. 5 Distribution of *NOSII* haplotypes in patient and control samples and their association analysis in the (a) endemic and (b) non-endemic region

haplotype was associated with protection from disease severity in the non-endemic region as well (severe vs. control, OR=0.3, $p=1 \times 10^{-5}$).

The T allele of the CRP SNP rs3091244 (C/T/A) is a risk factor for disease in the non-endemic region

The T allele of the triallelic promoter SNP rs3091244 of the *CRP* gene exhibited association with severe disease in the non-endemic region, with no association found in the endemic region. Although association was not inferred from allele frequency, the OR calculated on the basis of genotype count indicated association of the TT genotype with disease manifestation in the non-endemic region (severe vs. control, TT and CC, OR=4.17, 95 % CI=0.77–41.73, $p=0.05$; TT and CT, OR=9.22, 95 % CI=1.81–88.71, $p=0.001$). Interestingly, the CT genotype strongly associated with protection from disease in this region [severe vs. control, CT and all other genotypes (CC, TT, AA, AT, AC), OR=0.29, 95 % CI=0.15–0.56, $p=6 \times 10^{-5}$]. Although the endemic region population was in HWE for this SNP, the locus was not under HWE in the non-endemic region ($p=0.0001$), suggesting the possibility of the locus being under selection pressure in the latter.

The distribution of CRP rs3091244 (C/T/A) alleles and genotypes varies across world populations. The A allele is more common in African-American than in European

populations, and a marked difference in genotype frequencies of the SNP has been found between two sympatric ethnic groups, Fulani and non-Fulani, from Mali and Sudan [23]. The available frequency data from 55 endogamous populations in IGVC panel I revealed a low mean MAF (0.07) of the A allele in India [26]. The T allele was the predominant mutant allele in IGVC populations, with a wide range (0.10–0.47) of frequency distribution and a mean MAF of 0.23 [26]. The rs3091244 A allele has been reported to be associated with increased plasma levels of CRP and was also associated with increased susceptibility to uncomplicated *P. falciparum* malaria in a Sudanese population [24]. The frequency of the A allele was low in our sample set (Table 1) and no disease association was observed for genotypes (AA/AT/AC) carrying this allele. Instead, the mutant TT genotype was weakly associated with severe disease in the non-endemic region, whereas the CT genotype had a strong protective effect against malaria.

To our knowledge, this is the first report identifying a *THBS1* polymorphism and its specific haplotypes as risk factors for falciparum malaria and demonstrating association of an *ESEL* SNP with cerebral malaria in a disease-endemic region. Additionally, gene variants of two important immune regulatory molecules, *NOSII* and *CRP*, are associated with malaria disease outcome in Indian populations with differential association of a CRP promoter triallelic SNP observed in the endemic versus non-endemic regions. Although alteration

of transcription factor binding sites and protein conformational changes are predicted to result from some of these polymorphisms, their actual functional consequences remain to be validated.

Acknowledgments We are grateful to all donors and their families. KK received a Senior Research Fellowship from the Council of Scientific and Industrial Research, Government of India. This work was funded by a grant to SH and VV from the Department of Biotechnology, Government of India (BT/PR6065/MED/14/738/2005) and the CSIR network project BSC0104i to SH. This is CDRI communication number 9019.

Conflict of interest The authors declare no conflict of interest.

Compliance with ethical standards Ethical clearance for sample collection and approval of protocols was obtained from the Institutional Ethics Committee (Human Research) of participating institutes/hospitals. Informed written consent was obtained from each volunteer/guardian prior to the collection of blood samples.

References

- Sherman IW, Eda S, Winograd E (2003) Cytoadherence and sequestration in *Plasmodium falciparum*: defining the ties that bind. *Microbes Infect* 5:897–909
- Rowe JA, Claessens A, Corrigan RA, Arman M (2009) Adhesion of *Plasmodium falciparum*-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. *Expert Rev Mol Med* 11:e16
- de Mendonça VR, Goncalves MS, Barral-Netto M (2012) The host genetic diversity in malaria infection. *J Trop Med* 2012:940616
- Sinha S, Qidwai T, Kanchan K, Anand P, Jha GN, Pati SS, Mohanty S, Mishra SK, Tyagi PK, Sharma SK; Indian Genome Variation Consortium, Venkatesh V, Habib S (2008) Variations in host genes encoding adhesion molecules and susceptibility to falciparum malaria in India. *Malar J* 7:250
- Rock EP, Roth EF Jr, Rojas-Corona RR, Sherwood JA, Nagel RL, Howard RJ, Kaul DK (1988) Thrombospondin mediates the cytoadherence of *Plasmodium falciparum*-infected red cells to vascular endothelium in shear flow conditions. *Blood* 71:71–75
- Zwicker JI, Peyvandi F, Palla R, Lombardi R, Canciani MT, Cairo A, Ardissino D, Bernardinelli L, Bauer KA, Lawler J, Mannucci P (2006) The thrombospondin-1 N700S polymorphism is associated with early myocardial infarction without altering von Willebrand factor multimer size. *Blood* 108:1280–1283
- Gearing AJ, Hemingway I, Pigott R, Hughes J, Rees AJ, Cashman SJ (1992) Soluble forms of vascular adhesion molecules, E-selectin, ICAM-1, and VCAM-1: pathological significance. *Ann N Y Acad Sci* 667:324–331
- Hviid L, Kurtzhals JA, Adabayeri V, Loizon S, Kemp K, Goka BQ, Lim A, Mercereau-Puijalon O, Akanmori BD, Behr C (2001) Perturbation and proinflammatory type activation of $\gamma\delta 1^+ \gamma\delta$ T cells in African children with *Plasmodium falciparum* malaria. *Infect Immun* 69:3190–3196
- Amodu OK, Gbadegesin RA, Ralph SA, Adeyemo AA, Brenchley PE, Ayoola OO, Orimadegun AE, Akinsola AK, Olumese PE, Omotade OO (2005) *Plasmodium falciparum* malaria in south-west Nigerian children: is the polymorphism of ICAM-1 and E-selectin genes contributing to the clinical severity of malaria? *Acta Trop* 95:248–255
- Moncada S (1997) Nitric oxide in the vasculature: physiology and pathophysiology. *Ann N Y Acad Sci* 811:60–67, discussion 67–69
- Taylor-Robinson AW, Phillips RS, Severn A, Moncada S, Liew FY (1993) The role of TH1 and TH2 cells in a rodent malaria infection. *Science* 260:1931–1934
- Weinberg JB, Lopansri BK, Mwaikambo E, Granger DL (2008) Arginine, nitric oxide, carbon monoxide, and endothelial function in severe malaria. *Curr Opin Infect Dis* 21:468–475
- Levesque MC, Hobbs MR, O’Loughlin CW, Chancellor JA, Chen Y, Tkachuk AN, Booth J, Patch KB, Allgood S, Pole AR, Fernandez CA, Mwaikambo ED, Mutabingwa TK, Fried M, Sorensen B, Duffy PE, Granger DL, Anstey NM, Weinberg JB (2010) Malaria severity and human nitric oxide synthase type 2 (NOS2) promoter haplotypes. *Hum Genet* 127:163–182
- Hobbs MR, Udhayakumar V, Levesque MC, Booth J, Roberts JM, Tkachuk AN, Pole A, Coon H, Kariuki S, Nahlen BL, Mwaikambo ED, Lal AL, Granger DL, Anstey NM, Weinberg JB (2002) A new NOS2 promoter polymorphism associated with increased nitric oxide production and protection from severe malaria in Tanzanian and Kenyan children. *Lancet* 360:1468–1475
- Burgner D, Xu W, Rockett K, Gravenor M, Charles IG, Hill AV, Kwiatkowski D (1998) Inducible nitric oxide synthase polymorphism and fatal cerebral malaria. *Lancet* 352:1193–1194
- Jack DL, Turner MW (2003) Anti-microbial activities of mannose-binding lectin. *Biochem Soc Trans* 31:753–757
- Thévenon AD, Leke RG, Suguitan AL Jr, Zhou JA, Taylor DW (2009) Genetic polymorphisms of mannose-binding lectin do not influence placental malaria but are associated with preterm deliveries. *Infect Immun* 77:1483–1491
- Garred P, Nielsen MA, Kurtzhals JA, Malhotra R, Madsen HO, Goka BQ, Akanmori BD, Sim RB, Hviid L (2003) Mannose-binding lectin is a disease modifier in clinical malaria and may function as opsonin for *Plasmodium falciparum*-infected erythrocytes. *Infect Immun* 71:5245–5253
- Larsen F, Madsen HO, Sim RB, Koch C, Garred P (2004) Disease-associated mutations in human mannose-binding lectin compromise oligomerization and activity of the final protein. *J Biol Chem* 279:21302–21311
- Nahrevanian H, Gholizadeh J, Farahmand M, Assmar M (2008) Patterns of co-association of C-reactive protein and nitric oxide in malaria in endemic areas of Iran. *Mem Inst Oswaldo Cruz* 103:39–44
- Imrie H, Fowkes FJ, Michon P, Tavul L, Reeder JC, Day KP (2007) Low prevalence of an acute phase response in asymptomatic children from a malaria-endemic area of Papua New Guinea. *Am J Trop Med Hyg* 76:280–284
- Pied S, Nussler A, Pontent M, Miltgen F, Matile H, Lambert PH, Mazier D (1989) C-reactive protein protects against preerythrocytic stages of malaria. *Infect Immun* 57:278–282
- Israelsson E, Ekström M, Nasr A, Dolo A, Kearsley S, Arambepola G, Homann MV, Maiga B, Doumbo OK, Elghazali G, Giha HA, Troye-Blomberg M, Berzins K, Tornvall P (2009) Marked differences in CRP genotype frequencies between the Fulani and sympatric ethnic groups in Africa. *Malar J* 8:136
- Giha HA, Nasr A, Ekström M, Israelsson E, Arambepola G, Arnot D, Theander TG, Troye-Blomberg M, Berzins K, Tornvall P, ElGhazali G (2010) Association of a single nucleotide polymorphism in the C-reactive protein gene (–286) with susceptibility to *Plasmodium falciparum* malaria. *Mol Med* 16:27–33
- Jha P, Sinha S, Kanchan K, Qidwai T, Narang A, Singh PK, Pati SS, Mohanty S, Mishra SK, Sharma SK, Awasthi S, Venkatesh V, Jain S, Basu A, Xu S; Indian Genome Variation Consortium, Mukerji M, Habib S (2012) Deletion of the APOBEC3B gene strongly impacts susceptibility to falciparum malaria. *Infect Genet Evol* 12:142–148
- Indian Genome Variation Consortium (2008) Genetic landscape of the people of India: a canvas for disease gene exploration. *J Genet* 87:3–20

27. Kanchan K, Jha P, Pati SS, Mohanty S, Mishra SK, Sharma SK, Awasthi S, Venkatesh V, Habib S (2015) Interferon-gamma (IFNG) microsatellite repeat and single nucleotide polymorphism haplotypes of IFN-alpha receptor (IFNAR1) associated with enhanced malaria susceptibility in Indian populations. *Infect Genet Evol* 29: 6–14
28. World Health Organization (WHO) (2000) Severe falciparum malaria. World Health Organization, Communicable Diseases Cluster. *Trans R Soc Trop Med Hyg* 94(Suppl 1):S1–S90
29. Gómez LM, Anaya JM, Vilchez JR, Cadena J, Hinojosa R, Vélez L, Lopez-Nevot MA, Martín J (2007) A polymorphism in the inducible nitric oxide synthase gene is associated with tuberculosis. *Tuberculosis (Edinb)* 87:288–294
30. Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Ann Hum Genet* 47:253–259
31. Holmberg V, Schuster F, Dietz E, Sagarriga Visconti JC, Anemana SD, Bienzle U, Mockenhaupt FP (2008) Mannose-binding lectin variant associated with severe malaria in young African children. *Microbes Infect* 10:342–348
32. Boldt AB, Luty A, Grobusch MP, Dietz K, Dzeing A, Kombila M, Kreamsner PG, Kun JF (2006) Association of a new mannose-binding lectin variant with severe malaria in Gabonese children. *Genes Immun* 7:393–400
33. Luty AJ, Kun JF, Kreamsner PG (1998) Mannose-binding lectin plasma levels and gene polymorphisms in *Plasmodium falciparum* malaria. *J Infect Dis* 178:1221–1224
34. Wang Z, Liu Y, Liu J, Liu K, Lou Y, Wen J, Niu Q, Wen S, Wu Z (2010) E-selectin gene polymorphisms are associated with essential hypertension: a case–control pilot study in a Chinese population. *BMC Med Genet* 11:127
35. Takei T, Iida A, Nitta K, Tanaka T, Ohnishi Y, Yamada R, Maeda S, Tsunoda T, Takeoka S, Ito K, Honda K, Uchida K, Tsuchiya K, Suzuki Y, Fujioka T, Ujiie T, Nagane Y, Miyano S, Narita I, Gejyo F, Nihei H, Nakamura Y (2002) Association between single-nucleotide polymorphisms in selectin genes and immunoglobulin A nephropathy. *Am J Hum Genet* 70:781–786
36. Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Looareesuwan S, Tokunaga K (2002) Significant association of longer forms of CCTTT Microsatellite repeat in the inducible nitric oxide synthase promoter with severe malaria in Thailand. *J Infect Dis* 186:578–581
37. Kun JF, Mordmüller B, Perkins DJ, May J, Mercereau-Puijalon O, Alpers M, Weinberg JB, Kreamsner PG (2001) Nitric oxide synthase 2(Lambaréné) (G-954C), increased nitric oxide production, and protection against malaria. *J Infect Dis* 184:330–336
38. Dhangadamajhi G, Mohapatra BN, Kar SK, Ranjit MR (2009) The CCTTT pentanucleotide microsatellite in iNOS promoter influences the clinical outcome in *P. falciparum* infection. *Parasitol Res* 104:1315–1320
39. Warpeha KM, Xu W, Liu L, Charles IG, Patterson CC, Ah-Fat F, Harding S, Hart PM, Chakravarthy U, Hughes AE (1999) Genotyping and functional analysis of a polymorphic (CCTTT)(n) repeat of NOS2A in diabetic retinopathy. *FASEB J* 13:1825–1832
40. Seguin MC, Klotz FW, Schneider I, Weir JP, Goodbary M, Slayter M, Raney JJ, Aniagolu JU, Green SJ (1994) Induction of nitric oxide synthase protects against malaria in mice exposed to irradiated *Plasmodium berghei* infected mosquitoes: involvement of interferon gamma and CD8+ T cells. *J Exp Med* 180:353–358