Genome mapping of malaria resistance genes:

The host ligands of PfEMP1.

A thesis submitted for the degree of *Doctor of Philosophy*

Andrew E. Fry

Balliol College, University of Oxford

Trinity 2009

Genome mapping of malaria resistance genes: The host ligands of PfEMP1.

Andrew E. Fry

Balliol College, University of Oxford

A thesis submitted for the degree of *Doctor of Philosophy,* Trinity 2009

Abstract

Erythrocytes infected by mature forms of the *Plasmodium falciparum* parasite adhere to other components of the vascular space, a behavior considered critical to the pathogenesis of severe malaria. Adhesion is mediated by the *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), a highly variant antigen expressed by the parasite and subject to switching during the course of an infection. The host ligands of PfEMP1 include CD36, ICAM-1 and the ABO antigens. By employing a series of population- and family-based association studies from multiple African populations, we examined whether variation in the genes underlying these molecules affects susceptibility to severe malaria. Our results suggest that a common frameshift mutation in the ABO glycosyltransferase, responsible for blood group O, is associated with protection from severe malarial phenotypes (P=2x10^{-7}), particularly severe malarial anaemia. However, we found no significant disease associations with variation in either the *ICAM1* or *CD36* genes. We focused on two particular functional polymorphisms, the missense ICAM-1^{Kilifi} and the *CD36* nonsense mutation T1264G. We genotyped both markers in around 10,000 individuals, but neither demonstrated an association with severe malarial phenotypes. Malaria has been a profound selection pressure shaping human genetic diversity. The last decade has seen the development of several haplotype-based methods to detect signatures of recent positive evolutionary selection. These techniques are potentially invaluable tools in our hunt for genetic variants that protect from life threatening malaria. We used simulations and empirical data from the International HapMap Project to demonstrate the validity of searching for long regions of haplotype homozygosity, as an approach to finding alleles undergoing selective sweeps. We analysed genetic data from a range of populations, particularly those utilized by HapMap, to investigate whether our candidate genes were associated with signals of recent positive selection. We characterized the distribution of a selection event associated with the *CD36* 1264G allele, focused in Central-West Africa, and demonstrated a novel signal of low population differentiation at the *ABO* gene, suggestive of longstanding balancing selection. Our work confirms that variation in the host ligands of PfEMP1 modulates severe malaria susceptibility, and highlights the value of using signals of selection, along with functional experiments and genetic association studies, to dissect the biology of severe malaria.

Acknowledgements

I would like to thank my supervisor, Professor Dominic Kwiatkowski, for his support and advice, and for providing me with such a stimulating project. I thank the other members of the group for their diverse contributions to this work, particularly: Sarah Auburn, David Barnwell, Susana Campino, Man-Suen Chan, Taane Clark, Jantina de Vries, Mahamadou Diakite, Abier Elzein, Julian Forton, Anita Ghansah, Angie Green, Michael Griffiths, Jeremy Hull, Martin Kimber, Valentina Mangano, Marilyn McCreight, Kyle Minch, Bert Mohr, Catherine Moyes, Bonnie Pedersen, Claire Potter, Anna Richardson, Kirk Rockett, Elham Sadighi Akha, Miguel San Joaquin, Kerrin Small, Marryat Stevens, Yik Ying Teo, Clare Trafford and Aaron Vanderwal. I would like to express my appreciation to the patients and relatives who provided their DNA for our research, and to the scientific teams who labored to assemble these resources. I am also grateful to the Wellcome Trust for its support of my work through a Clinical Research Training Fellowship. From outside of the laboratory, I thank my good friends Geoffrey Ma, David Styles and Naomi Carter, for keeping me entertained. I would like to thank my father, Peter, for checking my spelling and grammar; and my mother, Janet, for her encouragement. Lastly, but critically, I thank my wife, Aureola, for her patience and sense of humor.

Tempus fugit

This thesis is comprised of the following set of scientific papers with the addition of an introduction, general discussion, and general conclusions.

1. **Fry AE**, Griffiths MJ, Auburn S, Diakite M, Forton JT, Green A, Richardson A, Wilson J, Jallow M, Sisay-Joof F, Pinder M, Peshu N, Williams TN, Marsh K, Molyneux ME, Taylor TE, Rockett KA, Kwiatkowski DP. Common variation in the ABO glycosyltransferase is associated with susceptibility to severe *Plasmodium falciparum* malaria. Human Molecular Genetics. 2008 Feb 15;17(4):567-76.

2. **Fry AE**, Auburn S, Diakite M, Green A, Richardson A, Wilson J, Jallow M, Sisay-Joof F, Pinder M, Griffiths MJ, Peshu N, Williams TN, Marsh K, Molyneux ME, Taylor TE, Rockett KA, Kwiatkowski DP. Variation in the *ICAM1* gene is not associated with severe malaria phenotypes. Genes and Immunity. 2008 Jul;9(5):462-9.

3. **Fry AE**, Ghansa A, Small KS, Palma A, Auburn S, Diakite M, Green A, Campino S, Teo YY, Clark TG, Jeffreys AE, Wilson J, Jallow M, Sisay-Joof F, Pinder M, Griffiths MJ, Peshu N, Williams TN, Newton CR, Marsh K, Molyneux ME, Taylor TE, Koram KA, Oduro AR, Rogers WO, Rockett KA, Sabeti PC, Kwiatkowski DP. Positive selection of a *CD36* nonsense variant in sub-Saharan Africa, but no association with severe malaria phenotypes. Human Molecular Genetics 2009 Jul 15;18(14):2683-92.

4. **Fry AE**, Trafford CJ, Kimber MA, Chan MS, Rockett KA, Kwiatkowski DP. Haplotype homozygosity and derived alleles in the human genome. American Journal of Human Genetics 2006 Jun;78(6):1053-9.

Contents

List of Figures

List of Tables

Abbreviations

ATS	Acidic terminal sequence
BCS	Blantyre Coma Score
CD	Cluster of differentiation
CEPH	Centre d'Etude du Polymorphisme Humain
CEU	CEPH individuals from Utah
CI	Confidence interval
CIDR	Cysteine rich interdomain region
CLR	Conditional logistic regression
CM	Cerebral malaria
cM	Centimorgan
CR1	Complement receptor 1
CSA	Chondroitin sulphate A
DARC	Duffy antigen/chemokine receptor gene
DBL	Duffy-binding-like
DNA	Deoxyribose nucleic acid
EHH	Extended haplotype homozygosity
FBAT	Family-based association test
F_{ST}	Fixation index
G6PD	Glucose-6-phosphate dehydrogenase
HA	Hylauronic acid
HapMap	International Haplotype Map Project
Hb	Haemoglobin

HBB	β-globin gene
HbS	Haemoglobin S
HCB	Chinese Han from Beijing
HGDP	Human Genome Diversity Project
HLA	Human leukocyte antigens
HS	Heparan sulphate
HWE	Hardy-Weinberg equilibrium
ICAM-1	Intercellular adhesion molecule-1
IgG	Immunoglobulin G
IgM	Immunoglobulin M
iHS	Integrated haplotype score
iRBC	Infected red blood cell
JPT	Japanese from Tokyo
KAHRP	Knob-associated histidine protein
Kb	Kilobase
kDa	Kilodalton
LD	Linkage disequilibrium
LFA-1	Lymphocyte function-associated antigen
LRH	Long-range haplotype test
MAF	Minor allele frequency
Mb	Megabase
MDA	Multiple displacement amplification
OR	Odds ratio

PECAM-1	Platelet/endothelial cell adhesion molecule-1
PEP	Primer extension pre-amplification
PfEMP1	*Plasmodium falciparum* erythrocyte membrane protein 1
PO-LRT	Parent-of-origin likelihood ratio test
RBC	Red blood cell
RFLP	Restriction fragment length polymorphism
RIF	Repetitive interspersed family
SA	Severe malarial anaemia
SELE	E-selectin
SELP	P-selectin
SICA	Schizont-infected cell agglutination
SNP	Single nucleotide polymorphism
STEVOR	Subtelomeric variant open reading frame
TDT	Transmission disequilibrium test
THBS1	Thrombospondin-1 gene
TNF	Tumor necrosis factor
TSP-1	Thrombospondin-1
VCAM-1	Vascular cell adhesion molecule
WHO	World Health Organization
XP-EHH	Cross population extended haplotype homozygosity
YRI	Yoruba from Ibadan, Nigeria

Chapter 1

Introduction

1.1 Malaria as a global health threat.

At the beginning of the 21st century malaria continues to impact human health on a global scale. Poverty, malnutrition, limited access to health services, insecticide resistance and the emergence of resistance to anti-malarial drugs, conspire to mean malaria remains a major public health problem in much of the developing world. Malaria is caused by a parasitic protozoa of the genus *Plasmodium*. Four main species of *Plasmodium* infect humans: *P. vivax, P. ovale, P. malariae* and *P. falciparum.* Occasionally other *Plasmodium* species also cause human disease (1). *P. falciparum* is generally considered to be the most dangerous (2). Today, over two billion people live at risk from *P. falciparum* infection, mainly populations living in sub-Saharan Africa, Southeast Asia and South America (Figure 1.1). Given the distribution of the disease in developing countries, exact statistics on mortality and morbidity are unavailable. However, estimates of disease burden suggest that approximately 500 million clinical episodes of *P. falciparum* occur every year, with over 1 million malaria-related deaths (3, 4). The majority of cases occur in sub-Saharan Africa, with children under the age of 5 years at particular risk from life-threatening disease. In pregnant women, malaria contributes to anaemia and adverse birth outcomes such as spontaneous abortion, stillbirth, premature delivery and low birth weight (4).

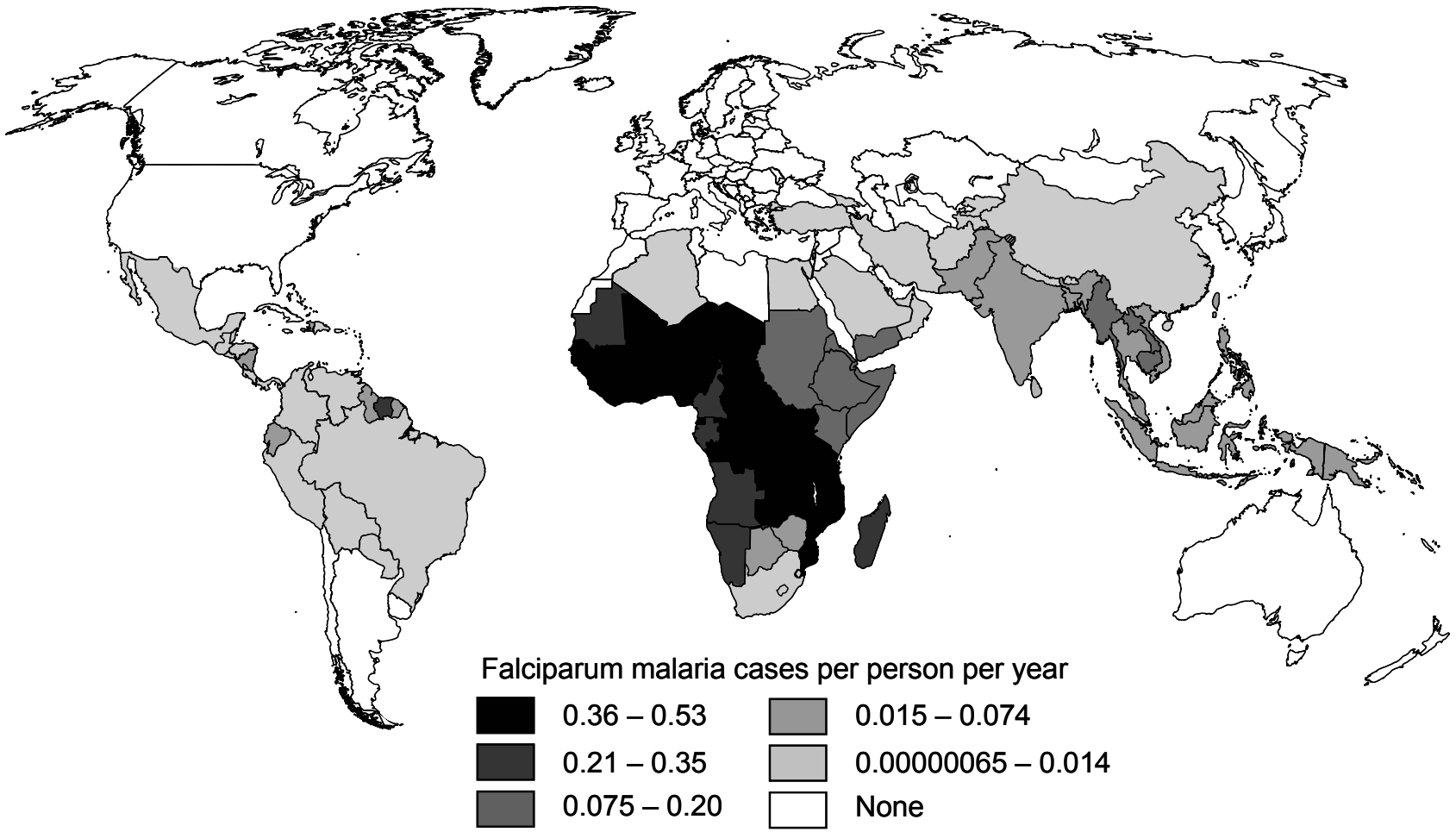


Figure 1.1. Estimated incidence of clinical *Plasmodium falciparum* episodes resulting from local transmission, country level averages, 2004 (4).

1.2 Natural history of malaria.

During a blood meal, the female Anopheles mosquito transmits malaria sporozoites from its salivary glands to the human host (Figure 1.2). The sporozoites promptly invade host hepatocytes. Following 1 to 2 weeks of parasite multiplication, the infected hepatocytes burst, releasing merozoites. The merozoites invade erythrocytes, where they replicate further, forming first trophozoites (which appear ring-shaped on blood films) and then schizonts. Haemolysis follows, with the release of many new merozoites into the circulation to infect further erythrocytes. A small proportion of merozoites differentiate into sexual forms, male and female gametocytes, which are taken up by mosquitoes during subsequent blood meals. In the mosquito foregut the gametocytes fertilize one another to form diploid ookinetes. These zygotes embed themselves in the outer layers of the insect's gut and produce multiple sporozoites which in turn migrate to the mosquito salivary glands, ready to initiate new infections. In human hepatocytes, some *P. vivax* and *P. ovale* parasites develop into hypnozoites. These can remain dormant in liver cells for months or years before reactivating and causing a disease relapse. Neither *P. falciparum* nor *P. malariae* produce hypnozoites, although the milder *P. malariae* can cause long-term infection if left untreated.

Clinical symptoms develop 1-3 weeks after the initial mosquito bite, depending on the *Plasmodium* species. The cyclical process of erythrocyte infection, parasite replication and haemolysis leads to the characteristically periodic symptoms of fever, chills, malaise, fatigue, headache, nausea and vomiting. *P. vivax* and *P. ovale* classically cause symptoms every 48 hours ('benign tertian malaria') and *P. malariae* every 72 hours ('benign quartan

malaria'). *P. falciparum* ('malignant tertian malaria') has a roughly 48 hour periodicity but tends to be more erratic, and is marked by prolonged fever of greater intensity. Malaria patients often develop hepatosplenomegaly and may become dehydrated, whilst extensive haemolysis can lead to jaundice and anaemia.

People living in endemic areas experience recurrent episodes of parasitaemia leading to the development of clinical immunity which is species and strain specific. The need for multiple exposures to the parasite goes some way to explaining why children are particularly susceptible to severe forms of disease, and why this risk diminishes with age. Unlike other *Plasmodium* species which rarely cause death, approximately 1-2% of untreated *P. falciparum* infections lead to severe life-threatening complications (5). 'Severe malaria' is the term commonly used to describe such infections. Cerebral malaria (CM) and severe malarial anaemia (SA) are two common complications of severe malaria found in African children (6). Other complications include hypoglycaemia, respiratory distress, acidosis, hypoglycaemia, or renal failure (7). The World Health Organisation (WHO) has published formal guidelines defining CM, SA and the other complications of severe malaria (7). In essence, CM is severely impaired consciousness not attributable to another cause, in the presence of *P. falciparum* parasitaemia, while SA is anaemia (packed cell volume less than 15% or haemoglobin less than 5g/dl) in the presence of significant parasitaemia. CM is associated with high rates of mortality (~20%) (8), but most patients recover rapidly from their coma. CM in children appears to have relatively low rates of long-term severe neurological sequelae (9), although some impairment persists in about 1 in 4 survivors (10, 11).

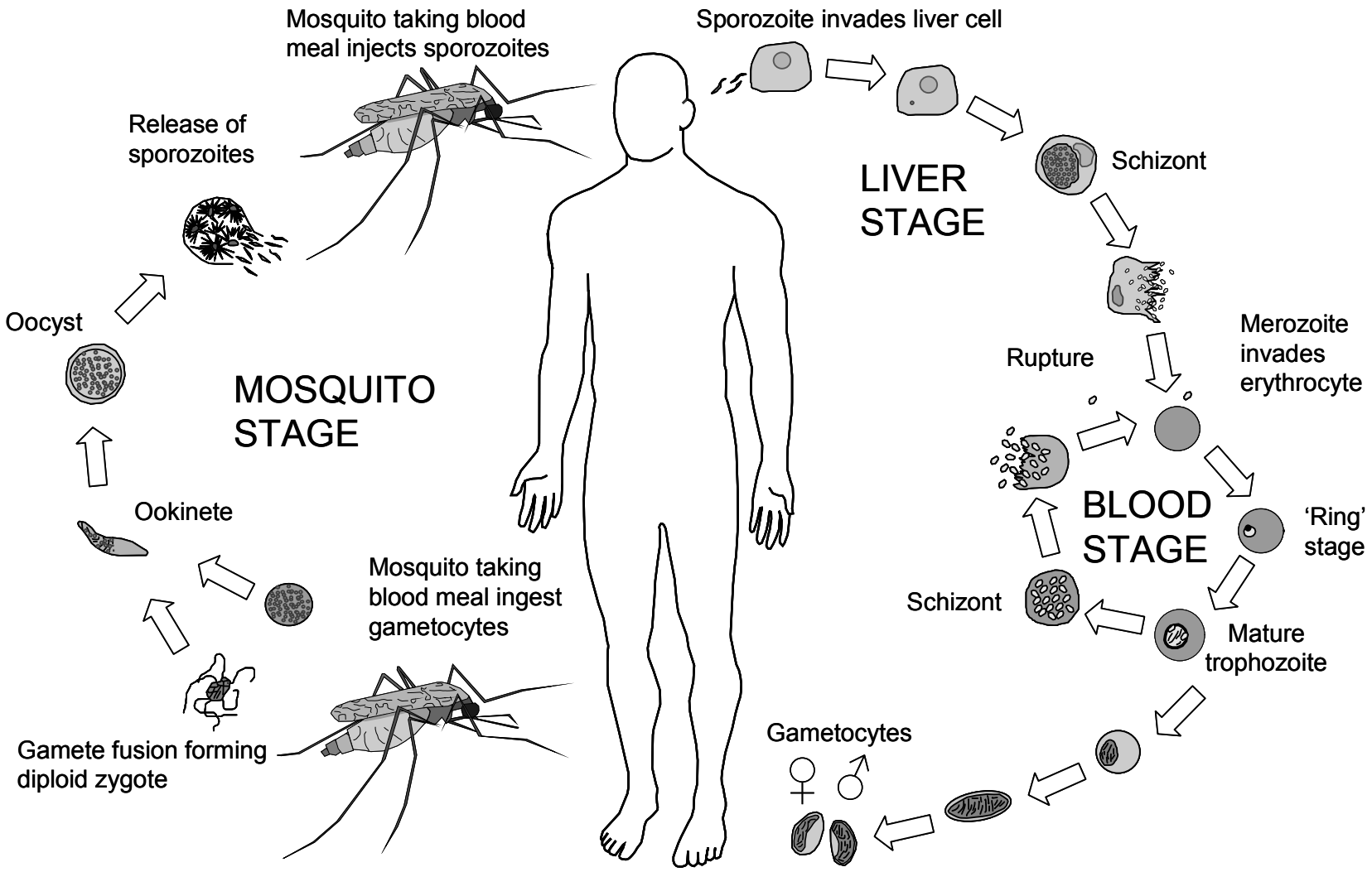


Figure 1.2. Life cycle of the *Plasmodium falciparum* parasite.

1.3 Pathogenesis of severe malaria.

The pathogenesis of severe malaria is complex and still poorly understood. There is growing appreciation, however, of the heterogeneous nature of the clinical syndromes of severe disease. A child with a life-threatening complication may have reached that state through many different disease-related mechanisms. For example, SA in an infected child can be due to a combination of haemolysis of parasitized erythrocytes; removal of erythrocytes (both parasitized and unparasitized) by the immune system; reduced erythropoiesis in bone marrow; interaction between malaria and other infections; and nutritional deficiencies (12). SA, hence lower oxygen carrying capacity of blood, contributes to the substantial metabolic consequences of severe disease. Acidosis and respiratory distress are both common in severe malaria and associated with increased mortality (6, 13). Hypovolaemia, hypoglycaemia, lactate production by parasites and decreased removal of lactate by the liver, all contribute to the metabolic burden of severe disease (12).

A key element of the pathogenesis of severe malaria is the ability of red blood cells (RBCs) infected with mature forms of the *P. falciparum* parasite (iRBCs) to adhere to components of the vascular compartment. Adhesive targets include endothelium (a behaviour called 'sequestration'), uninfected RBCs (described as 'rosetting' when seen *in vitro*, due to the appearance of the RBC clumps), platelets, and cells of the immune system (14). Stickiness has clear advantages for the parasite and disadvantages for the host. Adhesion of iRBCs to post-capillary venules helps the parasite avoid splenic passage and therefore reduces the chances of clearance by the immune system. An iRBC surrounding itself with uninfected RBCs cloaks the infected cell from the immune

system and clumps may get wedged in the microvasculature, once again avoiding splenic passage. Adhesion of iRBCs to leukocytes provides the opportunity to modulate components of the immune system (e.g. dendritic cells) (15).

Sequestration of iRBCs in the brain vasculature is clearly linked to the pathogenesis of CM. Post-mortem studies of people who have died of CM show greater accumulation of iRBCs on their cerebral microvasculature compared with the brains of non-CM patients, and in comparison with other organs in CM patients (16, 17). Widespread sequestration mechanically obstructs blood flow in vascular beds contributing to the failure of tissue perfusion and causing hypoxia. Sequestered iRBCs provoke local inflammatory responses such as nitric oxide release, induce oxidative stress and cause apoptosis of endothelial cells (18). Systemic release of proinflammatory cytokines, such as tumour necrosis factor (TNF)-α, up-regulates adhesion molecules, providing further targets for sequestration by iRBCs and host leukocytes. Parasite toxins and systemic metabolic disturbance are also thought to contribute to the insult to the cerebral microvasculature (19).The end result of these processes appears to be disruption of the blood-brain-barrier leading to vascular leak, cerebral oedema, accumulation of white blood cells, and haemorrhage into the brain (11, 20). The disparity in outcomes (some children dying with significant brain pathology while others recover quickly from coma with little or no neurological sequelae) underlines the diverse pathogenic mechanisms leading to impaired consciousness in severe malaria.

1.4 Human genetics as an approach to studying malaria.

In many villages in sub-Saharan Africa, everyone in the community is exposed to malaria several times a year. In such places, children will carry parasites in their blood almost constantly, yet only a fraction will go on to develop severe disease. This variation in outcome is explained in part by non-genetic factors such as age, nutrition, gender, the availability of bed nets and anti-malarials. However, there is strong evidence that a proportion of this interpersonal difference is due to host genetics. Studies of mortality among adoptees (21), ethnic differences in susceptibility (22), twin studies (23), linkage analysis in siblings (24), disease association studies (25), and longitudinal studies (26) have demonstrated a substantial genetic component to infectious disease susceptibility, and specifically to malaria susceptibility. Pedigree analysis in a longitudinal study of a rural Sri Lankan population found evidence that genetic variation modulated the frequency and intensity of *P. falciparum* infections. For clinical intensity of illness, around half of the repeatable differences between patients (about 10% of the total variation) was attributable to host genetics (26). Using similar methodology in a study of children living in a costal region of Kenya, investigators reported around a quarter of the total variation in mild and hospital-admitted malaria incidence was explained by additively acting host genes (27). Interestingly, the haemoglobin S (HbS) polymorphism explained only 2% of the total variation, suggesting that there are probably many genes affecting host susceptibility, each contributing small effects.

Over recent years genetic variants at hundreds of candidate genes have been studied and associations with diverse malaria phenotypes have been identified (see (25) for an extensive review). The candidate genes can be broadly categorized into a few groups: the

haemoglobins (e.g α-globin, β-globin); components of the erythrocyte surface [e.g. Duffy antigen (28), band 3 (29)]; red cell metabolism [e.g. glucose-6-phosphate dehydrogenase (G6PD) (30), pyruvate kinase (31)]; the immune system [e.g. human leukocyte antigens (HLA) (23, 32), TNF (33, 34), interferon receptor gamma (35), interlukin-4 (36)]; and host molecules used for cytoadherence by iRBC (see section 1.8). Many of the variants linked to protection from severe malaria were ascertained because of their association with deleterious haematological phenotypes and their geographical distribution in malarial regions. A common hypothesis is that the fitness consequences of the haematological phenotype are, in evolutionary terms, balanced against the benefit of protection from malaria. Globin gene variants were among the first genetic traits to be linked to malaria protection. J. B. S. Haldane, speaking at the 8th International Congress of Genetics in Stockholm in 1948, was the first to speculate that thalassaemia heterozygotes around the Mediterranean might receive protection from malaria (37). The hypothesis that sickle heterozygotes were protected from malaria was independently suggested and successfully tested by A. C. Allison in 1954 (38). Allison's results were arguably the first empirical evidence of natural selection in humans (39). The mutation responsible for sickle is a nonsynonymous SNP in codon 6 of the β-globin gene (*HBB*) which replaces a glutamate for a valine residue. The resulting HbS proteins tend to polymerise, becoming insoluble and leaving erythrocytes in a rigid 'sickle' shape which has difficulty fitting through tissue microvasculature. The distinctive geographical distribution of the sickle allele, present at a frequency of around 5-10% across sub-Saharan Africa and the Middle East, closely following the distribution of endemic malaria, led to the hypothesis that sickle operates as a balanced polymorphism. Sickle heterozygotes are protected from malaria while homozygotes suffer sickle cell anaemia, a syndrome which is rapidly fatal in childhood in developing countries. Further field and

laboratory studies have confirmed the link between HbS and protection from *P. falciparum* (40) as well as other point mutations in *HBB* such as haemoglobin C (41) and haemoglobin E (42). Another classical example of a genetic variant affecting malaria susceptibility is the Duffy antigen. Lack of expression of the Duffy antigen on erythrocytes is a fixed trait across much of Africa. Duffy antigen negativity is caused by a mutation at a GATA-1 erythroid transcription factor binding site in the promoter of the *DARC* gene, which abolishes antigen expression on erythrocytes but not other cell types (28). The Duffy antigen acts as the receptor for *P. vivax*, and removal of Duffy prevents invasion of host RBCs. Interestingly, Duffy negative blood group has also arisen, independently, in a region of Papua New Guinea where *P. vivax* is endemic (43). In contrast to sickle, Duffy antigen negativity appears to have reached fixation in African populations because it lacks deleterious consequences, with the exception of transfusion reactions (when a Duffy negative individual receives repeated Duffy positive blood transfusions) or haemolytic disease of the newborn (caused when a Duffy negative mother gives birth to a Duffy positive baby).

Dissection of how host genetic variation influences malaria susceptibility has a number of potential benefits. First, genetic epidemiology can highlight critical host genes and pathways involved in life threatening disease. Such studies may generate valuable insights into the biology of malaria, which can be exploited for vaccine and drug development. Second, it can be difficult when studying severe malaria to know whether a particular phenomenon (e.g. sequestration or rosette formation) is critical to pathogenesis, or merely a by-product (epiphenomenon) caused by severe disease. However, as an individual's genome is essentially fixed, any genetic variant associated with disease cannot be a 'side-effect', and is therefore likely to be causally linked to

severe disease. Thirdly, as one of the most significant infectious pathogens in recent human history, malaria has been a major factor in human evolution. By learning about how genetic variation contributes to malaria susceptibility, we may understand more about recent human evolution.

1.5 Detecting haplotypic signatures of selection.

Severe malaria kills many individuals who have not reached reproductive age. By doing so the parasite exerts a profound selection pressure on exposed populations. We expect that, over many generations, protective alleles will be selected for, while alleles that increase susceptibility will be selected against. Such evolutionary events often leave characteristic patterns of genetic variation, signatures of selection, in and around the genes involved (44). This presents an exciting opportunity. If we can detect these signals, they may guide us to key genetic variants influencing disease susceptibility. A wide range of techniques to detect signals of selection in population genetic data have been developed. These include methods looking for altered proportions of amino-acid changing mutations [e.g. the Ka/Ks test (45) or the McDonald-Kreitman test (46)], reduced genetic diversity [e.g. Tajima's D (47) or the Hudson-Kreitman-Aguade test (48)], excess high frequency derived alleles [e.g. Fay and Wu's H (49)], allele frequency differentiation between populations (e.g. F_{ST}), and analysis of haplotype structure. The first three approaches look for selection operating over long time periods, typically hundreds of thousands or millions of years, while population differentiation and haplotype-based methods tend to detect relatively recent selection (44). Haplotype-based methods are particularly suited to studying the effects of malaria, as they detect selection events from the last few thousand years; the period in which both humans and *P.*

falciparum have undergone rapid population growth (50). Haplotype-based methods include the long-range haplotype test (LRH) (51), integrated haplotype score (iHS) (52), haplosimilarity (53), and cross population extended haplotype homozygosity (XP-EHH) (54). The basic idea underlying these methods is the identification of alleles that have become common, while retaining an unusually similar haplotypic background (the combination of alleles at other markers along the same chromosome). An allele appearing by mutation in a population will exist, initially, on a single haplotype. In subsequent generations, if the allele persists, the haplotypic background of the allele will become mixed-up as recombination swaps material between different haplotypes in the population. However, if the new allele increases in frequency rapidly, because of positive selection, it will tend to retain its original haplotypic background (Figure 1.3). The key signal is therefore the presence of a long section of similar haplotypes (sometimes referred to as 'long' haplotypes) surrounding a core variant or, to consider it another way, unusual similarity of haplotypes at a given distance from the core. The significance of haplotype-based test statistics is difficult to determine analytically, as this requires detailed models of a population's demographic history and the effects of selection (55). Instead significance is generally determined empirically, by comparison with millions of other values calculated from across the genome. Empirical significance testing is convenient and can rank loci in terms of interest, but it leaves us uncertain as to what magnitude of signal constitutes a significant value. In principle, if few alleles in the human genome are under selection, the empirical significance testing could generate numerous false positives. Significance testing is also sensitive to differences in recombination rate across the genome, bias in SNPs ascertainment, and the precise demographic history of a population. In addition, we expect certain types of recent selection events to be undetectable, for example old alleles that become advantageous,

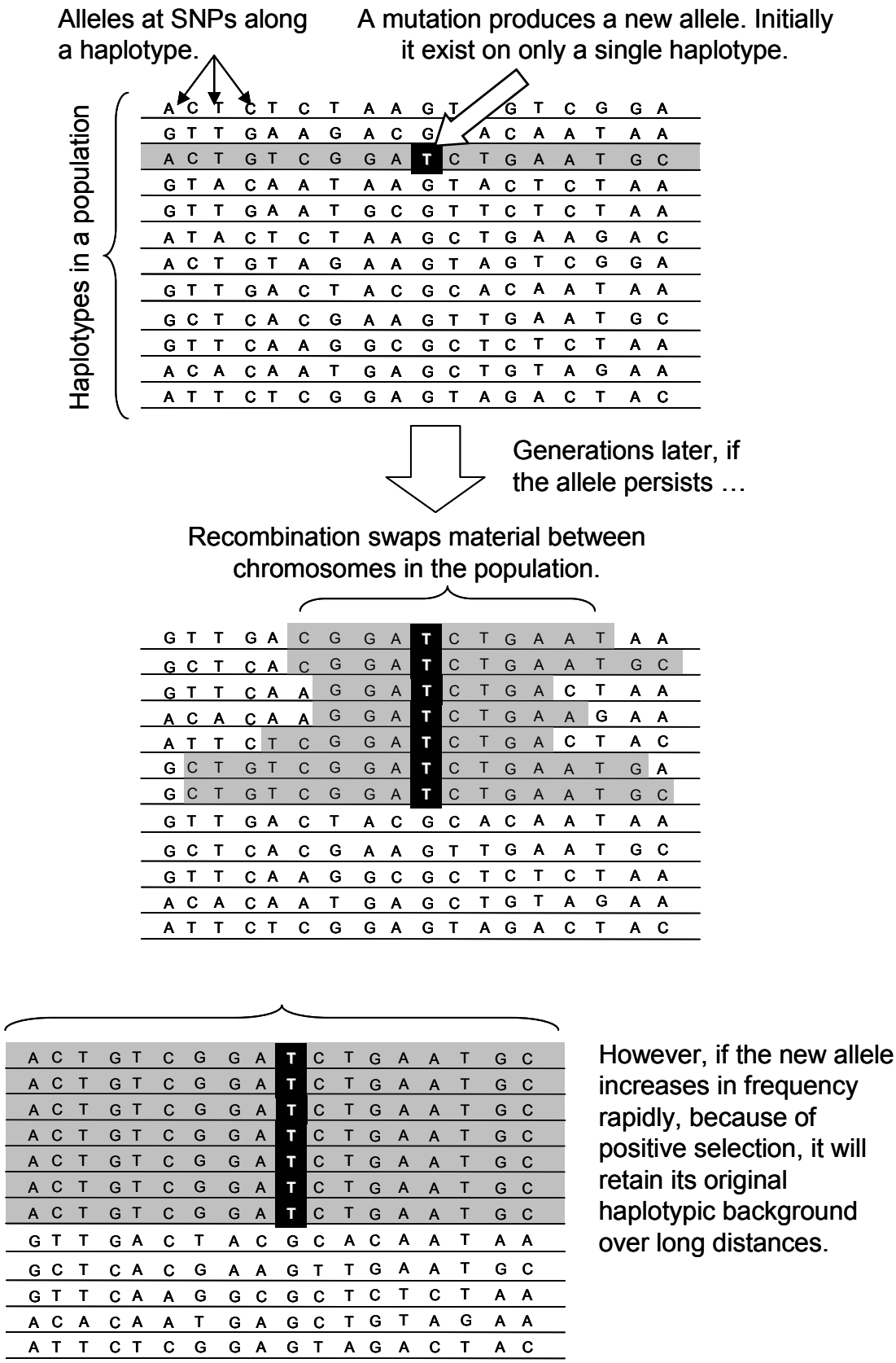


Figure 1.3. Basic concepts underlying haplotype-based methods for detecting recent positive selection

early selection sweeps, and sweeps near or at fixation. Haplotype-based methods for detecting recent positive selection have the potential to be extremely valuable in our efforts to map genetic variants that influence malaria susceptibility. Haplotype-based methods have already been used to show evidence for selection at a range of genes implicated in the genetic epidemiology of malaria, including *G6PD, CD40L* and *HBB* (51, 53, 56). *HBB* was also highlighted in a genome-wide screen (57). However, further work is required to determine the true efficacy of haplotype-based methods in practical situations, for example, when applied to genome-wide screening, or to study candidate genes with limited prior evidence for selection.

1.6 Sequestration: a history.

In 1890 the Italian scientists Amico Bignami and Giuseppe Bastianelli reported that RBCs infected by the mature stages of *P. falciparum* attached to the walls of vessels in deep tissues. This finding explained the absence of mature parasites from peripheral blood samples (58). Seventy years later transmission electron microscopy revealed protrusions, called 'knobs', on the surface of iRBCs. The appearance of these knob-structures coincided with the development of the mature forms of the parasite, and the sequestration phase (59). The knobs were soon shown to form focal junctions with the capillary endothelium, acting as attachment points for iRBCs (60). The number and size of knobs varies with the developmental stage of the intraerythrocytic parasite. Schizont-infected cells have smaller, more numerous knobs, whereas trophozoite-infected cells have larger, less numerous knobs. The number of knobs is also proportional to the number of parasites present in the cell (61). Knobs are closely related to the cytoskeleton underlying the red cell membrane (62). However, the relationship between knobs and

sequestration is more complex than it may initially seem. Considering other *Plasmodium* species, the mouse malaria *P. chabaudi* sequesters without having knob structures, while *P. malariae* and *P. brasiliarum* produce knobs yet do not sequester (62, 63). Immature gametocytes of *P. falciparum* are capable of a form of knob-independent adherence (64). Furthermore, clones of *P. falciparum* have been identified that do not form knobs but still have cytoadherent properties, although the binding observed is weaker (65). Knobs are not necessary for iRBC adherence, but they seem to increase the strength of binding and may expand the spectrum of adhesive targets.

The beginnings of our understanding of the molecular basis of sequestration can be traced to the experiments of Monroe D. Eaton in the 1930s. Eaton found that monkeys infected by *P. knowlsei* had serum that could agglutinate iRBCs, but not uninfected RBCs. This became the basis of an assay, called the schizont-infected cell agglutination (SICA) reaction (63). The SICA reaction demonstrated that the malaria parasite modified the repertoire of antigens present on the iRBC surface. Eaton went on to show two critical points: first, during an infection, there was sequential expression of different antigenic types and, second, that immunity was associated with the host developing type-specific antibodies. The SICA reaction was then used to investigate the sequential variation in antigens seen during an infection. It was eventually shown that parasites from a clonal line of *P. knowlesi*, essentially genetically identical haploid parasites, could vary in antigenic expression. This phenomenon, called clonal antigenic variation, has since been shown in a number of other *Plasmodium* species including *P. falciparum* (14).

The *P. knowlesi* SICA antigens were found to be high-molecular weight, polymorphic, trypsin sensitive, parasite-encoded proteins on the surface of the iRBC (66). Using a

similar immunoprecipitation strategy with strain-specific sera, a polymorphic, high-molecular-weight (~200–300 kDa) protein was isolated from human RBCs infected with *P. falciparum*. This molecule was named *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) (67). It was recognized that strain-specific sera could block the cytoadherence of iRBC *in vitro*, as could trypsin treatment of the erythrocytes (68). The shared features between the *P. knowlesi* SICA antigens and PfEMP1 suggested that the two were equivalent.

1.7 *Plasmodium falciparum* erythrocyte membrane protein 1.

In 1995 PfEMP1 was localized to the knob structures of iRBCs and reported to be encoded by the *var* gene family (69-71). There are around 60 *var* genes per haploid parasite genome; mainly located in the subtelomeric regions of the 14 *P. falciparum* chromosomes (14). The *var* gene family is highly diverse in both sequence and size (Figure 1.4). Despite their diversity, members of the *var* gene family share a conserved 3' acidic terminal sequence (ATS). The cytoplasmic ATS domain anchors PfEMP1 to elements of the knob structure by interacting with the binding domains of 'Knob-Associated Histidine Protein' (KAHRP), another parasite-derived protein. Disruption of KAHRP produces the knobless clones of *P. falciparum*. These clones still express PfEMP1 on the iRBC cell surface and undergo cytoadherence with reduced binding strength (72). The large variable 5' exon is exposed on the extracellular surface and is composed of two types of protein domains. Between two and five Duffy-binding-like (DBL) domains are present in a PfEMP1 molecule (73). These domains are categorized into five classes (α–ε) (74). DBL domains are interspersed by either cysteine rich interdomain regions (CIDR) or C2 domains. CIDR are also categorized into three classes

(α–γ). A typical pattern is an N-terminal DBLα domain followed by a CIDRα or CIDRγ domain, the rest of the molecule varies significantly (74). The *var* genes are subject to frequent recombination and rearrangement. These processes add novel variants to the *var* gene repertoire of parasite populations (14). The *var* genes are not the only variant multigene family found in the *P. falciparum* genome. Two other groups include the 'Repetitive Interspersed Family' (*rif*, ~200 genes per haploid genome), and the 'Subtelomeric Variant Open Reading Frame' family (*stevor*, 30-40 genes per haploid genome) (75). There is evidence that the products of both families (STEVOR and RIFIN proteins) are transported to the iRBC surface membrane, and undergo clonal variation. However, the extent to which these genes contribute to cytoadherence or immune evasion is not yet clear (63, 76). Only one *var* gene is expressed by an iRBC, but the active *var* gene is subject to switching during subsequent rounds of erythrocyte infection. Switching the PfEMP1 antigen displayed on iRBCs helps parasites evade the immune system and can lead to recrudescence in parasitaemia. Recent estimates have suggested the *var* gene switching rate may be as high as ~18% of parasites per generation *in vivo* (77). With up to 10^{10} parasites present in the infected host, this could rapidly lead to exhaustion of the full PfEMP1 repertoire. However, it is thought that v*ar* genes switch on and off at different rates, leading to a limited number of common antigenic types (78). The rates of switching are determined by *var* promoter sequences that control epigenetic silencing or activation of the genes (79). It appears that silenced *var* gene promoters compete for occupancy of a particular perinuclear compartment that permits transcription of only a single *var* gene. The diversity and switching of the *var* gene family explains why people living in malaria endemic regions must develop antibodies against different PfEMP1 variants in a piecemeal fashion, and why acquired immunity to *P. falciparum* often takes several years of persistent exposure to develop (80, 81).

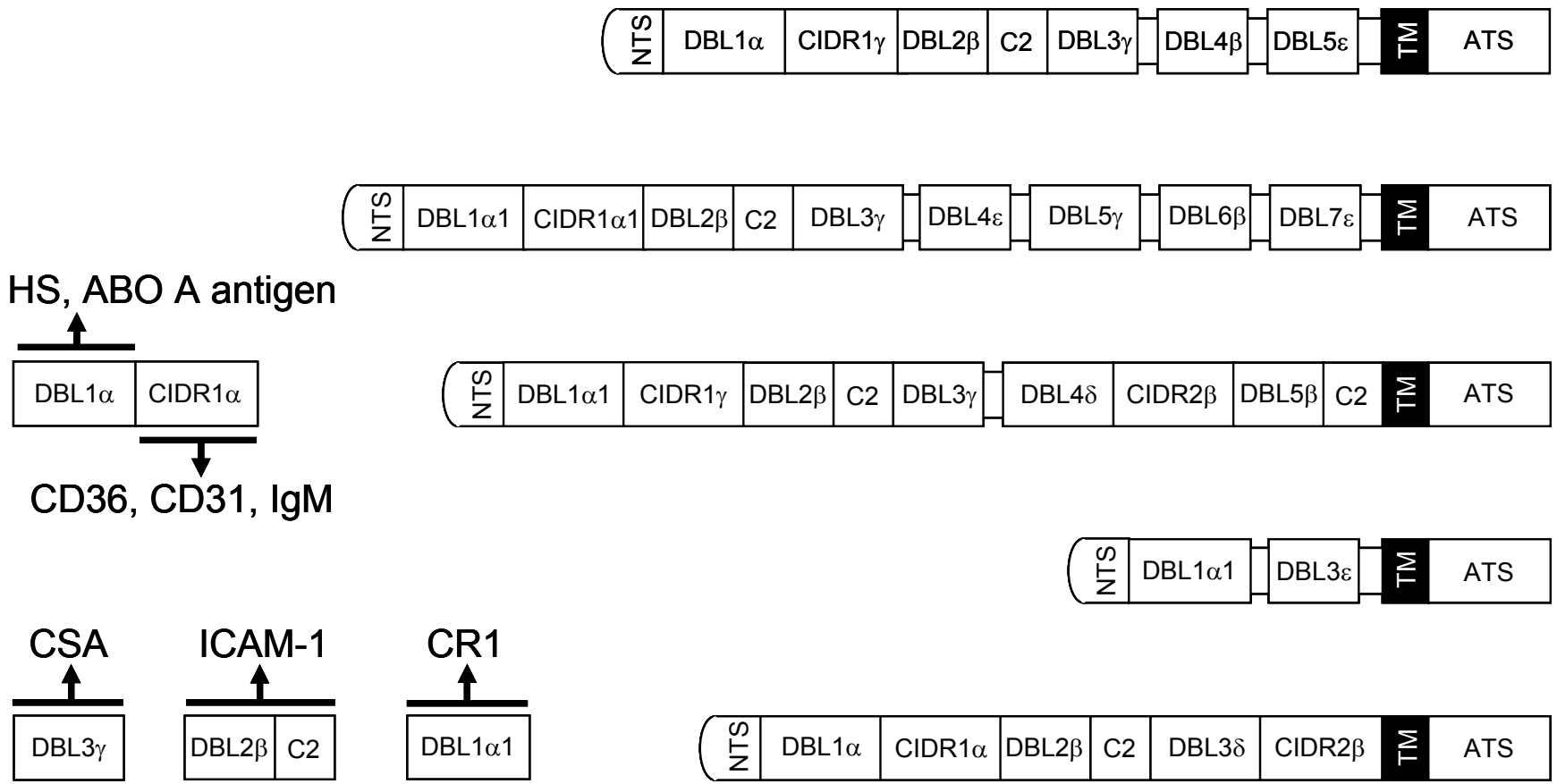


Figure 1.4. Examples of PfEMP1 domain structure and binding. Domain structures from 5 of the 59 PfEMP1 molecules encoded by the *P. falciparum* 3D7 parasite genome (right) and known host ligands for a range of PfEMP1 domains (left) (82). NTS, N-terminal segment; DBL, Duffy-binding-like domain; CIDR1, cysteine rich interdomain region; C2, C2 domain; TM, transmembrane; ATS, acidic terminal sequence; CR1, complement receptor 1; ICAM-1, Intercellular adhesion molecule; CSA, Chondroitin sulphate A, HS, heparan sulphate.

In addition to immune evasion through antigenic variation, the diverse cytoadherent properties of PfEMP1 variants provide the parasite with other strategies to modulate and avoid the host immune system. A wide range of host ligands of PfEMP1 have been reported.

1.8 The host ligands of PfEMP1.

Following characterization of the molecular basis of sequestration a number of host molecules have been identified as ligands for PfEMP1 (Figure 1.5). A series of genetic studies have focused on variation in the host genes underlying these molecules, and their impact on malaria. We shall now review this area in depth.

1.8.1 CD36

CD36 (Fatty Acid Translocase or Platelet Glycoprotein IV) is an 88 kDa glycoprotein and type B scavenger molecule found on diverse cell types including endothelial cells, adipocytes, platelets, macrophages and dendritic cells (83). The majority of *P. falciparum* clinical isolates and laboratory lines bind CD36 (73). CD36 is considered to provide a stable and long-lasting anchor for iRBCs. CD36 was first reported as a ligand of PfEMP1 in 1989 (84-86). A monoclonal antibody against an epitope of CD36 containing residues 139 to 184, an immunodominant region, blocks iRBC cytoadherence to CD36 (87), while antibodies against the CIDR1 domain of PfEMP1 block binding of iRBCs to CD36 (73). CD36 deficiency on platelets and other cell types (Nak^a negative blood group) is common in South East Asians (88) and populations of African origin (89). In 2000 Aitman *et al.* reported that a nonsense mutation, introduced by the T1264G

SNP (rs3211938) in exon 10 of the *CD36* gene, was common in samples of African Americans, Gambians and Kenyans (90) and distinct from the mutations found in Asia (91). CD36 polypeptides terminated by 1264G lack the C-terminal transmembrane domain, abolishing expression of CD36 on platelets and monocytes. These findings raised the possibility that CD36 deficiency had been selected for by malaria. To test this hypothesis Aitman *et al.* performed a case-control analysis using the two deficiency alleles found at >1% frequency in Gambians (T1264G and G1439C+1444delA) (90). Surprisingly, stratified analysis in Kenya (183 cases and 331 controls) and Gambia (415 cases and 430 controls) found an association between deficiency alleles and increased risk of severe disease (odds ratio (OR) 1.53, *P*=0.01) and CM (OR 1.49, *P*=0.04). The deficiency allele appeared to be predisposing children to malaria. Shortly after Pain *et al.* (92) analysed the T1264G allele in a Kenyan case-control study of 693 matched pairs. In contrast to Aitman *et al.* their data suggested that heterozygotes for the 1264G allele were protected against severe disease (OR 0·74, 95% CI 0·55–0·99; *P*=0.036). The protective effect included phenotypes such as SA and hypoglycemia, but not CM. When Omi *et al.* (93) screened *CD36* for variation in the Thai population they found 9 common variants (> 15% frequency) but did not encounter either T1264G or G1439C+1444delA. They performed a case-control study of 272 adults with severe *P. falciparum* infection and 203 with mild malaria. They found two associations between promoter polymorphisms and protection from malaria T-14C (*P*=0.016) and G-53T (*P*=0.05). Both SNPs were in high LD with each other and seemed to be embedded in one of two LD blocks spanning the gene. The authors went on to report that an allele of a nearby intronic TG dinucleotide repeat polymorphism was also present on haplotypes marked by the protective -14C and -53T alleles. The 12 TG repeat allele was associated with protection from malaria, CM versus mild malaria (*P*=0.0069) when compared against all

other alleles. In 2005 a smaller study of 223 Nigerian children was unable to find an association between T1264G and severe malaria (94). However, in 2007 a study of 471 severe malaria cases and 474 controls from the Luo people in Kenya once again reported that 1264G was marginally associated with severe malaria susceptibility (95).

1.8.2 Intercellular adhesion molecule -1

Intercellular adhesion molecule (ICAM)-1 (CD54) is a member of the immunoglobulin super-family typically expressed on endothelial cells, particularly brain microvascular endothelial cells and cells of the immune system (96). ICAM-1 is the ligand for lymphocyte function-associated antigen (LFA)-1 on lymphocytes and mediates homing and extravasation of leukocytes (97). This molecule is upregulated on cerebral endothelium by inflammatory cytokines (98). ICAM-1 was shown to be a receptor for iRBCs in 1989 (99) and a significant proportion of parasite strains bind to it (73). The minimal ICAM-1 binding domain of PfEMP1 is a complex motif consisting of the second DBL domain and an adjoining C2 domain (100). The binding site of iRBCs to ICAM-1 has been localized to the first of ICAM-1's five immunoglobulin-like domains (101). In 1997, Fernandez-Reyez *et al.* (102) sequenced the N-terminal domain of ICAM-1 in 24 asymptomatic Kenyan children and discovered a nonsynonymous polymorphism they designated as ICAM-1^{Kilifi}. ICAM-1^{Kilifi} was common in Gambians and Kenyans (~30%) but absent from a sample of Caucasians. These findings raised the possibility that the variant had been under selection by severe malaria. To explore this hypothesis the authors performed case-control analysis with 157 cases of cerebral malaria, 103 severe anaemia and 287 community controls. Surprisingly, ICAM-1^{Kilifi} was associated with susceptibility to CM, with a relative risk of 2.23 in homozygotes and

1.39 in heterozygotes (test for trend, *P*=0.0042). Laboratory work has found that isolates of *P. falciparum* that adhere to ICAM-1 have reduced binding to ICAM-1^{Kilifi} under both static and flow conditions (103). Further *in vitro* work has reported that ICAM-1^{Kilifi} has reduced binding avidity for LFA-1 and has reduced affinity for soluble fibrinogen (104). Taking into account the apparent association with malaria susceptibility and the allele frequency evidence of evolutionary selection in Africa, it has been proposed that ICAM-1^{Kilifi} may provide a selective advantage in chronic infection, by reducing leukocyte-mediated tissue damage (104). However, subsequent studies have found it difficult to replicate these initial association results. Bellamy *et al.* (105) reported a case-control study of ICAM-1^{Kilifi} in 1200 Gambian children. The experiment found no association. Kun *et al.* (106) found the derived allele was associated with protection from severe disease in Gabon in a small paired case-control study (100 cases and 100 controls). Three further studies in sub-Saharan Africa did not find an association between ICAM-1^{Kilifi} and severe malaria (94, 107, 108). In addition, Ohashi *et al.* (109) found no association between the ICAM-1^{Kilifi} variant and severe disease in Thailand, although the variant is at relatively low frequency in this population (~2%).

1.8.3 Complement receptor 1

Complement receptor 1 (CR1, CD35) encodes a membrane glycoprotein found on erythrocytes, neutrophils, glomerular podocytes, and follicular dendritic cells (110). CR1 is the receptor for complement components C3b and C4b. It acts as an inhibitory regulator of the complement cascade and also mediates the phagocytosis of opsonized material (111). Coding variation in the CR1 protein is responsible for the Knops blood group system. Differences in *CR1* allele frequencies between Africans and Europeans

have been highlighted as possible evidence for selection by severe malaria (112). CR1 involvement in malarial rosetting was first reported in 1997 (113). Rowe *et al.* later found the region of CR1 required for rosette formation was the C3b binding site. The presence of C3b was unnecessary for this activity (110). Specific monoclonal antibodies against CR1 or soluble CR1 could reverse rosetting.

Rosette formation is thought to be mediated by the first DBL domain of PfEMP1 (114). The iRBCs of Gambian children with CM have higher rates of rosetting behaviour compared with non-CM cases. The plasma from children with CM lacks anti-rosetting activity *in vitro*, whereas the plasma of children with mild disease can often disrupt rosettes (115, 116). Similar studies in Kenya have found clear trends towards higher rosette frequency with increasing severity of disease (117, 118). However, some studies have failed to find an association between rosetting and severe disease (119). The links between disease severity and rosetting, and between rosetting and CR1, along with evidence for recent selection at the locus, has generated significant interest in *CR1* polymorphisms. Nagayasu *et al.* (120) studied 185 Thai patients with malaria (55 severe, 130 uncomplicated) and observed that a *CR1* Hind III- Restriction Fragment Length Polymorphism (RFLP) allele, associated with low density CR1 expression, appeared to be associated with severe disease (P=0.005). In a case-control study of 1200 Gambian children Bellamy *et al.* found no disease association with the same *CR1* Hind III polymorphism (105). Zimmerman *et al* (121) studied the nonsynonymous *CR1* polymorphisms underlying the Knops blood group alleles. The Knops Sl:2 blood group had been associated with reduced *P. falciparum* rosette formation *in vitro* (117). However, the authors found no significant allele frequency differences between 463 severe malaria cases and 390 controls. Despite the lack of association with severe

malaria, the authors presented population genetic analysis suggestive of recent positive selection at the locus in Africans. Cockburn *et al.* (122) showed the that CR1 was absent from the erythrocytes of ~80% of health people from regions of Papua New Guinea with endemic malaria. Case-control analysis demonstrated that an exon 22 SNP associated with CR1 deficiency was also associated with protection from severe disease (180 severe malaria cases and 179 controls, OR 0.33; P=0.01).

1.8.4 Platelet/endothelial cell adhesion molecule-1

Platelet/endothelial cell adhesion molecule (PECAM)-1 (or CD31) is a 120-130 kDa integral membrane protein expressed by endothelial cells, platelets, neutrophils, monocytes and some T-cell subsets (123). Another member of the immunoglobulin super family, PECAM-1 is thought to mediate adhesion and migration of leukocytes through the vascular wall. PECAM-1 acts in a range of processes including thrombosis, inflammation, apoptosis and atherosclerosis (124). Binding of iRBCs to PECAM-1 on plastic, endothelial cells or transfected cells expressing the protein, was demonstrated in 1997. Soluble PECAM-1 and monoclonal antibodies against the antigen blocked adhesion (125). Case-control analysis in Kenya (99 controls, 297 cases) and Papua New Guinea (222 controls, 220 cases) of a single nonsynonymous SNP in codon 125 of *PECAM1* revealed no association with severe disease (126). A separate study in Thailand (132 cases, 78 controls) of both the codon 125 SNP and another nonsynonymous SNP at codon 563, reported that homozygotes for the 125V/563N haplotype were significantly associated with CM (P<0.01, OR 2.92) (127).

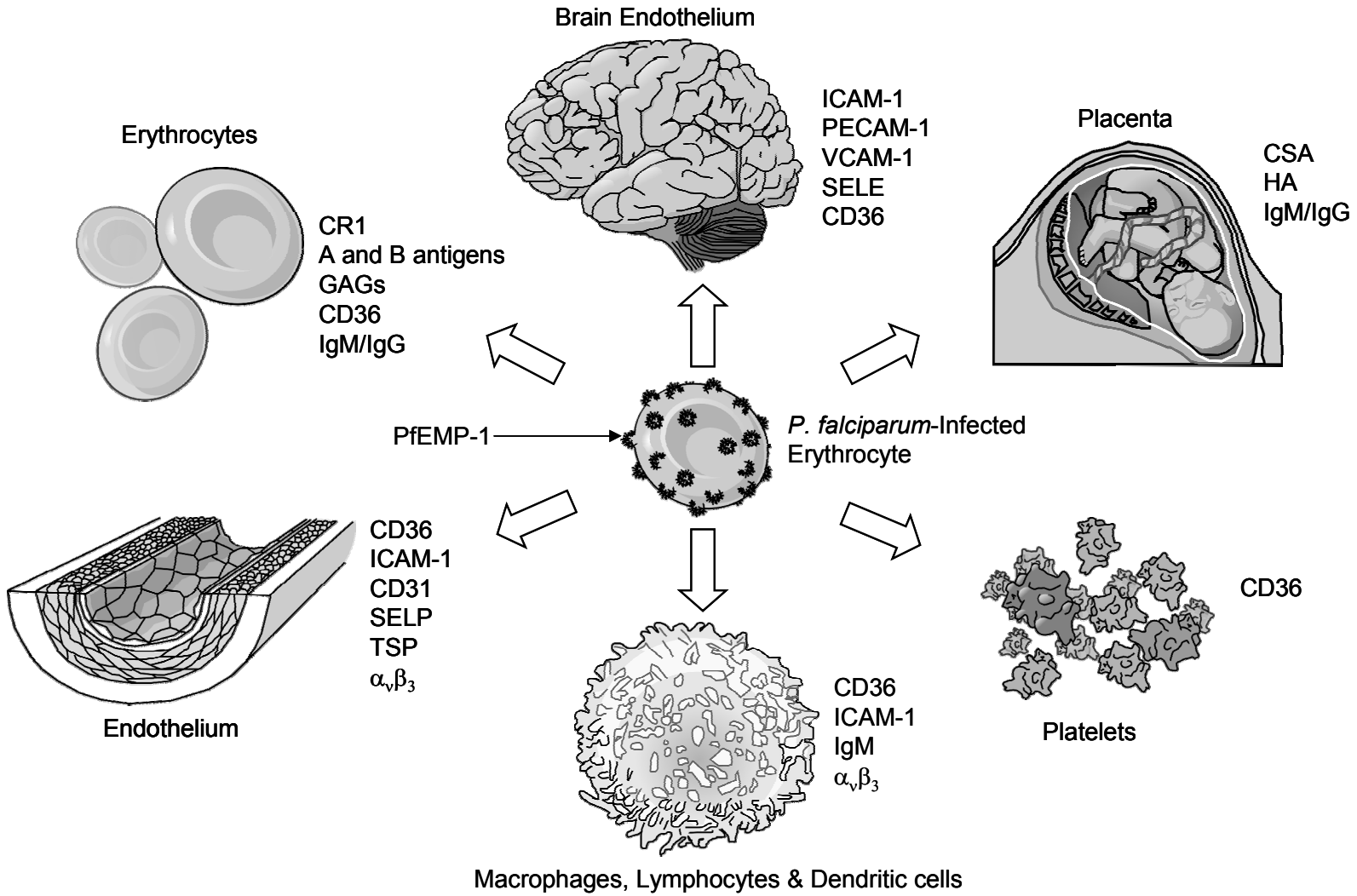


Figure 1.5. The host ligands of PfEMP1.

1.8.5 Thrombospondin-1

Thrombospondin (TSP)-1 is encoded by the *THBS1* gene. TSP-1 is a large trimeric molecule made of three identical 140kDa chains which is found on the surface of macrophages, melanoma cells, endothelium, the extracellular matrix and in platelet granules (63, 128). TSP-1 facilitates interactions between cells, and between cells and the extracellular matrix. Ligands for TSP-1 include fibrinogen, collagen, integrins, LDL receptor-related protein, syndecan-1, CD47 and CD36 (128, 129). TSP-1 functions in angiogenesis and platelet aggregation, and plays a role in tumorigenesis (130). TSP-1 binding to iRBCs was recognized in 1985, making it the first host ligand of PfEMP1 to be identified (131). Most clinical isolates and laboratory lines bind to TSP-1 (63). However, in contrast to ICAM-1 and CD36, PfEMP1 binding to TSP-1 is not variant-specific and is insensitive to immune sera (73). It is uncertain whether TSP-1 binding is stable under physiological conditions; it is not stable when bound to plastic slides (87), but it may be on vascular endothelium (132, 133). It is possible that the interaction between TSP-1 and iRBCs may be stabilized by other elements on the endothelium. It has also been suggested that another molecule on iRBCs, possibly modified band 3 or a parasite-derived molecule other than PfEMP1, is responsible for TSP-1 binding (73, 134).

1.8.6 P-selectin

P-selectin (SELP or CD62P) is a 140 kD adhesion molecule, which is expressed onto the surface of endothelium or platelets during inflammation or thrombosis. SELP mediates the interaction and rolling of leukocytes through its partner P-selectin glycoprotein

ligand-1 (CD162) (135). SELP was identified as ligand for iRBCs in 1998 (136) and for PfEMP1 specifically in 2001 (137). SELP binds iRBCs *in vitro* under flow conditions (87). To date, no human genetic association study between variation in the *SELP* gene and severe malaria has been published. However, evidence from rodent models suggests that SELP would be an interesting candidate (see Section 1.9).

1.8.7 ABO antigens

The ABO antigens are the basis of the ABO blood group system. The 3 different antigens A, B and O share a common complex carbohydrate precursor (the H antigen) modified as a final step by the ABO glycosyltransferase. Polymorphism of the ABO glycosyltransferase determines the final sugar residue added, which is either N-acetyl-galactosamine (A antigen), galactose (B-antigen) or nothing (O antigen). The ABO antigens are widely expressed on erythrocytes, endothelium and epithelium. Although discovered over a century ago (138), the function of the ABO system remains unclear (139). Humans without the A or B antigens, or those unable to express the entire H antigen (the 'Bombay phenotype') are healthy except for the risk of blood transfusion reactions. ABO blood group has been linked to susceptibility to a number of infectious diseases (140-143). Both blood group A and B antigens bind iRBCs leading to rosetting (144) . The binding site for the A antigen on PfEMP1 is within the N-terminal DBL1 and CIDR1 domains (145). Several field studies of ABO blood group and malaria susceptibility have suggested that blood group O individuals are relatively protected from severe malaria (146, 147). Similarly, a number of *in vitro* experiments conducted with parasite isolates from Kenya (117, 118) and Thailand (148), found that blood group

A, B and AB iRBCs were more likely to form rosettes, and that these blood groups were enriched amongst cases of severe disease.

1.8.8 Glycosaminoglycans

Chondroitin sulphate A (CSA), heparan sulphate (HS) and hylauronic acid (HA) are high molecular weight polysaccharide molecules added as post-translational modifications to proteins that are expressed on cell surfaces or, in the case of HA, secreted into the extracellular space. CSA is a heteropolymer of alternating glucuronic acid and sulphated N-acetylglucosamine residues, which is linked to the cell surface via a membrane-associated protein. There are clear links between malaria in pregnancy and the binding of iRBCs to placental CSA (149, 150). IRBC binding to CSA is mediated by PfEMP1 (151). CSA binds the DBL3 domains of PfEMP1 (152). PfEMP1 variants that bind to CSA tend not to bind CD36, although their binding domains are distinct (153). The placenta expresses abundant CSA which means pregnancy can open up a new niche for parasites in the circulation of a previously clinically immune individual. Accumulation of iRBCs in the placenta can lead to significant morbidity. In malaria endemic regions, approximately 19% of low birth weight in neonates is attributable to malaria in pregnancy. It is estimated that 6% of infant deaths are due to the low birth weight caused by malaria (154). Only a limited range of *var* gene products bind CSA. VAR2CSA seems to be the most likely PfEMP1 variant to mediate placental CSA binding (155). Women who acquire antibodies against CSA-binding PfEMP1 rapidly develop resistance to malaria in pregnancy. This explains the increased susceptibility to malaria during first and second pregnancies, and has led to suggestions that a vaccine against malaria in pregnancy could target VAR2CSA (156). HA has also been linked to placental

sequestration (157, 158). In contrast to CSA and HA, HS is a ligand for rosette formation (114). HS is thought to bind the first DBL domain of PfEMP1. HS on endothelial cells may also be involved in the sequestration of iRBCs (159). Binding of iRBCs to heparin (a molecule similar to HS) in Kenyan isolates of *P. falciparum* has also been associated with severe disease (118).

1.8.9 Other ligands

Non-immune IgM and IgG have both been shown to bind PfEMP1 (145, 160). Work with field isolates suggests that rosetting is influenced by IgM binding, and that IgM is the predominant class of natural antibodies binding to the surface of infected erythrocytes, rather than IgG (161). Non-immune IgM binds to the surface of iRBCs which have been selected for adhesion to CSA (162). Therefore it appears that PfEMP1 variants binding CSA also harbor binding domains for IgG and IgM (158, 163). Binding between iRBCs and vascular cell adhesion molecule (VCAM-1 or CD106) and E-Selectin (SELE or CD62E) has been reported (164). Both are interesting candidates as, along with ICAM-1, they are upregulated on cerebral endothelium during severe malaria (98). However, it appears that only a small minority of field isolates bind VCAM-1 or SELE (165). A small case-control study of malaria in Nigerian children (69 severe cases, 101 uncomplicated and 53 asymptomatic parasitaemia) has analysed a SNP in exon 4 of *SELE*, but no disease association was found (94). Finally, an *in vitro* flow assay suggested that $\alpha_V \beta_{III}$ integrin expressed on endothelial cells binds iRBCs (166). However the expression of this receptor on endothelium at the site of sequestration has not been confirmed.

1.9 Evidence from rodent models of sequestration.

Much of the functional evidence for the role of host ligands of PfEMP1 in severe malaria has come from *in vitro* studies of parasite isolates derived from peripheral blood samples. A possible disadvantage of this approach is that the parasites present in venous blood may not reflect those sequestered in end-organs; the parasites which we might presume are directly responsible for the complications of severe disease. Autopsy studies are viable but are difficult to perform on a large-scale. Rodent models of malaria are a potentially powerful alternative approach to studying the pathogenesis of severe disease. The rodent model provides investigators with opportunities to manipulate the host genome, specify parasite strain, control environmental variables and allows easy access to relevant tissues. Care must be taken in interpreting the results. One interesting illustration of the differences between human and murine malaria is that both *P. chabaudi* and *P. berghei* bind CD36 and undergo sequestration but neither possess PfEMP1 (167, 168). Human, rat and murine CD36 share 90% sequence homology at the amino acid level, and rat CD36 can avidly bind *P. falciparum* iRBCs *in vitro* (169). In mice CD36 is highly expressed on adipose tissue and some capillary endothelial cells (particularly those in skeletal or cardiac muscle tissue) but not cerebral endothelium (170). Real time observations of the infection process using transgenic luciferase producing *P. berghei* parasites found minimal evidence of CD36-mediated brain sequestration (168). CD36 wild type animals had significant sequestration of mature parasite forms in adipose tissue and lungs, which was abolished in CD36 deficient animals. However, cerebral complications still developed in CD36-null animals. Wild-type mice infected with *P. berghei* ANKA exhibited increased expression of ICAM-1, VCAM-1, and SELP (but not SELE) during infection (171). During *P. berghei* ANKA

infection mice susceptible to CM up regulate SELP in brain vessels while resistant mice do not. SELP-deficient mice infected with *P. berghei* ANKA have a reduced incidence of CM. Only mice deficient in endothelial SELP, and not platelet SELP, were protected from CM, and monoclonal antibodies against SELP did not prevent the development of cerebral disease (172). Mice deficient for ICAM-1 or SELP also exhibit increased survival after *P. berghei* infection (173, 174).

1.10 Aims

The ultimate goal of our work is to better understand the pathogenesis of severe malaria, and to aid in the development of strategies to treat and prevent life-threatening illness. This project focuses on the host ligands of PfEMP1. As we have seen, they are a diverse range of molecules linked by a common critical feature of parasite biology. There are numerous lines of evidence demonstrating the ways in which these host proteins and polysaccharides contribute to the pathogenesis of severe disease. However, relatively few of these experiments have addressed the impact of common genetic variation at the underlying loci, and many of the results to date have either been inconclusive or apparently contradictory. The core hypothesis of this project is that genetic variation in the host ligands of PfEMP1 modulates host susceptibility to severe malaria phenotypes. To explore this concept, we took two approaches: genetic association studies and population genetic analysis, and concentrated on three specific candidates: CD36, ICAM-1 and the ABO glycosyltransferase. Each candidate is of particular interest, CD36 is bound by the majority of *P. falciparum* isolates, ICAM-1 is highly expressed on the cerebral microvasculature, and the ABO antigens have been consistently reported to influence rosetting. Disease association studies provide us with an opportunity to

investigate the epidemiological impact of host genetic variation on key phenotypes such as CM or SA. Association studies come with a range of challenges, which include the practical (plus ethical, political and financial) implications of marshalling samples, phenotypes and genotypes; robust analysis of the results; and the interpretation of results in the context of existing data. Our second technique (analysis of population genetic data to detect recent positive selection) approaches our core hypothesis from a different direction. Population genetics provides us with a fascinating opportunity to unpick the evolutionary struggle between human and parasite. Malaria is clearly a powerful selective pressure operating on the human genome. If our core hypothesis is correct, signals of recent evolution are likely to be present at our candidate loci. We are entering an exciting time in human genetics; advanced genotyping technologies mean dense genome-wide datasets from multiple populations are becoming available. This provides opportunities to validate our theoretical tools for detecting recent positive selection, to screen candidate loci for signals of recent selection, and to determine the significance of putative selection events using empirical distributions of test statistics. The combination of both a significant disease association and signals of recent evolutionary selection, at one of our candidate genes, would be strong evidence that functional variation had been detected.

Chapter 2

Common variation in the ABO glycosyltransferase is associated with susceptibility to severe *Plasmodium falciparum* malaria.

Andrew E. Fry[1,*], Michael J. Griffiths[1,2], Sarah Auburn[1], Mahamadou Diakite[1], Julian T. Forton[1], Angela Green[1], Anna Richardson[1], Jonathan Wilson[1] , Muminatou Jallow[3], Fatou Sisay-Joof[3], Margaret Pinder[3], Norbert Peshu[2], Thomas N. Williams[2, 4], Kevin Marsh[2, 4], Malcolm E. Molyneux[5,7], Terrie E. Taylor[6,8], Kirk A. Rockett[1], Dominic P. Kwiatkowski[1, 9]

[1]The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, UK; [2]Kenya Medical Research Institute Centre for Geographical Medicine Research (Coast), P.O. Box 230, Kilifi, Kenya; [3]Medical Research Council, PO Box 273, Banjul, The Gambia; [4]Nuffield Department of Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK; The [5]Malawi–Liverpool–Wellcome Trust Programme of Clinical Tropical Research and [6]Blantyre Malaria Project, College of Medicine, PO Box 30096, Blantyre, Malawi; [7]Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK; [8]Department of Internal Medicine, College of Osteopathic Medicine, Michigan State University, East Lansing, Michigan 48824, USA. [9]Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK.

2.1 Abstract

There is growing epidemiological and molecular evidence that ABO blood group affects host susceptibility to severe *Plasmodium falciparum* infection. The high frequency of common ABO alleles means that even modest differences in susceptibility could have a significant impact on the health of people living in malaria endemic regions. We performed an association study, the first to utilize key molecular genetic variation underlying the ABO system, genotyping >9000 individuals across three African populations. Using population- and family-based tests, we demonstrated that alleles producing functional ABO enzymes are associated with greater risk of severe malaria phenotypes (particularly malarial anaemia) in comparison with the frameshift deletion underlying blood group O: Case-control allelic odds ratio (OR), 1.2; 95% confidence interval (CI), 1.09–1.32; P=0.0003; family-studies allelic OR, 1.19; CI 1.08 – 1.32; P=0.001; pooled across all studies allelic OR, 1.18; CI, 1.11–1.26; P=$2x10^{-7}$. We found suggestive evidence of a parent-of-origin effect at the *ABO* locus by analyzing the family trios. Non-O haplotypes inherited from mothers, but not fathers, are significantly associated with severe malaria (likelihood ratio test of Weinberg, P=0.046). Finally, we used HapMap data to demonstrate a region of low F_{ST} (-0.001) between the three main HapMap population groups across the *ABO* locus, an outlier in the empirical distribution of F_{ST} across chromosome 9 (~99.5 – 99.9th centile). This low F_{ST} region may be a signal of long-standing balancing selection at the *ABO* locus, caused by multiple infectious pathogens including *P. falciparum*.

2.2 Introduction

A link between the ABO blood group system and malaria susceptibility has long been suspected. Significant associations between blood group and *Plasmodium falciparum* malaria have been reported from cross-sectional and case–control studies in Brazil, Gabon, India, Sri Lanka and Zimbabwe (175-181) [see also recent reviews and references therein (146, 147)]. However, other studies in Colombia, India, Sudan and Nigeria could not find an association between malaria and blood group (182-187). The positive association studies have consistently suggested that blood group O individuals are relatively protected from severe malaria.

Human erythrocytes infected with mature forms of the *P. falciparum* parasite adhere to uninfected red blood cells, endothelia and other components of the vascular space (188). This adhesive behavior is mediated by *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), which is encoded by a family of highly variant parasite genes and subject to switching during the course of an infection (70). A range of human host molecules binding to PfEMP1 have been identified including CD36 (189), intercellular adhesion molecule (ICAM)-1 (190) and complement receptor-1 (CR1) (113). The ABO A and B antigens have been implicated in the formation of 'rosettes', the process by which infected red blood cells (iRBCs) surround themselves with uninfected erythrocytes (148, 191). PfEMP1 has been identified as the rosetting ligand of the parasite (113) and the blood group A antigen has been shown to bind to the semiconserved head structure of PfEMP1 (192). Erythrocyte rosetting is linked to the pathogenesis of severe malaria phenotypes such as cerebral malaria (CM) and severe malarial anaemia (SA) (116, 193). Indeed, fresh isolates of *P. falciparum* from Kenyan children with severe malaria bind A

antigen more frequently than strains from children with mild disease (194). Recent work on Kenyan and Malian isolates suggests a significant reduction in rosetting in blood group O individuals compared with non-O blood groups. Furthermore, parasites required rosetting activity for blood group O to offer protection from severe disease (195). Wild and laboratory parasite strains demonstrate blood group preferences, which appear to vary geographically. In general rosette formation has been shown to occur preferentially with blood group A, B or AB erythrocytes, and particularly groups A and AB in studies of African strains (148, 191, 196, 197).

Previous epidemiological studies have employed the phenotype of serology to determine the host ABO genotype. Although convenient, serological studies of the ABO system are unable to discriminate all genotypes, e.g. AO heterozygotes from AA homozygotes, and cannot assess subtler coding or non-coding variation. We took the novel approach of investigating the molecular genetic variation underlying this system. The ABO glycosyltransferase performs the terminal step in the biosynthesis of the ABO macromolecule, adding sugar residues to the precursor H antigen. The enzyme adds either *N*-acetylgalactosamine, to form the A antigen, galactose to form the B antigen, or is functionless, leaving the H (or O) antigen unmodified. A variety of polymorphisms have been reported in the ABO glycosyltransferase gene (198). The common key functional variants are: (i) a one nucleotide deletion in exon 6 (codon 87) leading to a reading frameshift and premature termination of the polypeptide before the N-terminal catalytic domain, producing the functionless O allele (199). Alternative ‘non-deletional’ O alleles exist but are rare in African populations (200). (ii) Four nonsynonymous single nucleotide polymorphisms (SNPs) (altering residues 176, 235, 266 and 268), which switch enzyme function from A transferase to B transferase activity. The third and fourth

SNPs (codons 266 and 268) have the greatest effect on determining the nucleotide-sugar donor used by the transferase. The second amino acid (codon 235) also affects nucleotide-sugar specificity but to a lesser degree. Finally the most 5' SNP (codon 176) has very little influence on donor specificity (201).

We designed genotyping assays for the O allele frameshift mutation (rs8176719) and the three key A/B non-synoymous SNPs in codons 235 (rs8176743), 266 (rs8176746) and 268 (rs8176747) in the catalytic domain. We conducted a multi-centred study employing samples from three African regions: The Gambia (West Africa), Malawi (South Central Africa) and Kenya (East Africa). We used two study types, first, population-based association studies in each of the regions employing children with severe malarial disease and local cord blood controls and, second, family-based association studies at each site looking for transmission distortion of ABO alleles between parents and affected offspring, an approach that has the advantage of being robust to population stratification. Our aims were: (i) given the presence of substantial linkage disequilibrium (LD) among the functional SNPs in the *ABO* locus, to define the most efficient marker set for large-scale genotyping; (ii) to perform SNP- and haplotype-based tests for association with severe malarial phenotypes; (iii) to employ our family-based genetic data to check for parent-of-origin effects, an approach not possible with previous case-control and cross-sectional study designs and (iv) to consider the population genetic implications of an association between *ABO* variation and severe falciparum malaria.

2.3 Results

2.3.1 Substantial LD between functional SNPs in the ABO glycosyltransferase gene.

The four key functional polymorphisms that we tested in the *ABO* locus are reported to be in substantial LD (198). To confirm this, we genotyped 1320 Gambian parent-offspring trios, and 30 Yoruba parent-offspring trios from Ibadan in Nigeria [cell line DNA used by International Haplotype Map (HapMap) Project] (57). This allowed us to look at the fine structure of the *ABO* locus, combining our functional SNPs with the high density HapMap marker dataset (Figure 2.1). Genotyping in both sample sets demonstrated near perfect LD (r^2, 0.9 – 1.0) among the three nonsynonymous SNPs that distinguish A and B haplotypes. The frameshift deletion underlying O alleles is in moderate LD with the three nonsynonymous SNPs (D', ~0.9; r^2, ~0.3). The majority of O alleles occur on an A haplotypic background, although the minority of recombinants with B haplotypes are still expected to produce truncated products. The implication of this haplotypic structure is that two markers, rs8176719 and one of the three nonsynonymous SNPs, can generally distinguish the three ABO alleles.

2.3.2 The common frameshift mutation underlying blood group O is associated with protection from severe malaria phenotypes, particularly SA.

Two markers, rs8176719 and rs8176746, were genotyped in 3906 cases of severe malaria plus population and family controls from The Gambia, Kenya and Malawi. Case-control analyses were performed on 2127 cases of severe malaria and 1931 population controls, and family-based association tests were performed on a different set of 1779 cases of severe malaria and their parents.

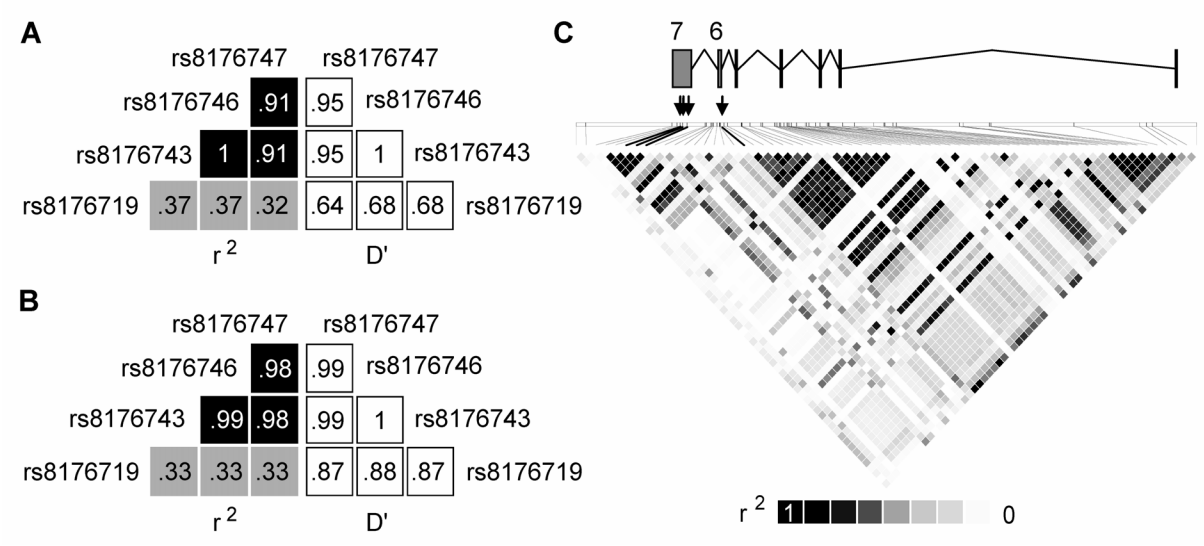


Figure 2.1. Linkage disequilibrium (LD) around the ABO glycosyltransferase gene. Strong LD exists between the four key functional SNPs (A) Yoruba HapMap parent-offspring trios from Ibadan in Nigeria. (B) 1320 Gambian parent-offspring trios. Both of these West-African populations have near perfect LD between the three nonsynonymous SNPs that differentiate A and B ABO alleles, and moderate but lower LD with the common deletion which generates the O allele. (C) r^2 values for the Yoruba population across the ABO gene. Phased genotypes (HapMap, July 2006) with additional genotyping for four functional polymorphisms in *ABO*. The *ABO* gene (total region shown chromosome 9, ~24kb, 133156822 – 133180999, NCBI Build 35) is illustrated 3' to 5' with exons 6 and 7 on the left. The SNPs differentiating A and B blood groups are in a high LD block, and indicated by three arrows under exon 7. The frameshift mutation responsible for the O allele (arrow under exon 6) is on the 3' edge of another high LD block.

Further details of severe malaria phenotypes and sample sizes from each population are given in Materials and Methods.

Our analysis took the statistical convention of using the commonest category (i.e. most frequent allele, haplotype or blood group) as reference, and comparing other alleles, haplotypes and inferred blood groups against this category. Single SNP analysis revealed that the minor allele of rs8176719, an insertion relative to the reference sequence (although almost certainly the ancestral allele), is associated with increased risk of severe malaria phenotypes. Consistent trends were found in both family and case-control studies, and across study sites (Table 2.1). However, some individual studies did not reach statistical significance; probably due to lack of power (see Materials and Methods for power calculations). Overall, the case-control allelic odds ratio (OR) was 1.2 and 95% confidence interval (CI) 1.09–1.32, P=0.0003; family-studies allelic OR 1.19, 95% CI 1.08–1.32, P=0.001; pooling data across all our studies, both family- and population-based, suggested an allelic OR of 1.18, 95% CI 1.11-1.26, $P= 2x10^{-7}$ for severe disease (Figure 2.2). The results suggest that the full-length allele of rs8176719 may be associated with a particular risk of anaemia during severe malaria infections. For the phenotype of SA, the allelic OR and 95% CI were 1.30 and 1.12–1.5 (P= 0.0004) and 1.34, 1.1 - 1.64 (P= 0.004) for the case–control and family-based studies respectively.

The situation with rs8176746 is more complex. The major allele, although defining the putatively ‘high risk’ A haplotype, occurs with the ‘low risk’ frameshift deletion upstream in about two-thirds of chromosomes. Thus, no simple genetic model is significantly associated with disease when this SNP is considered in isolation.

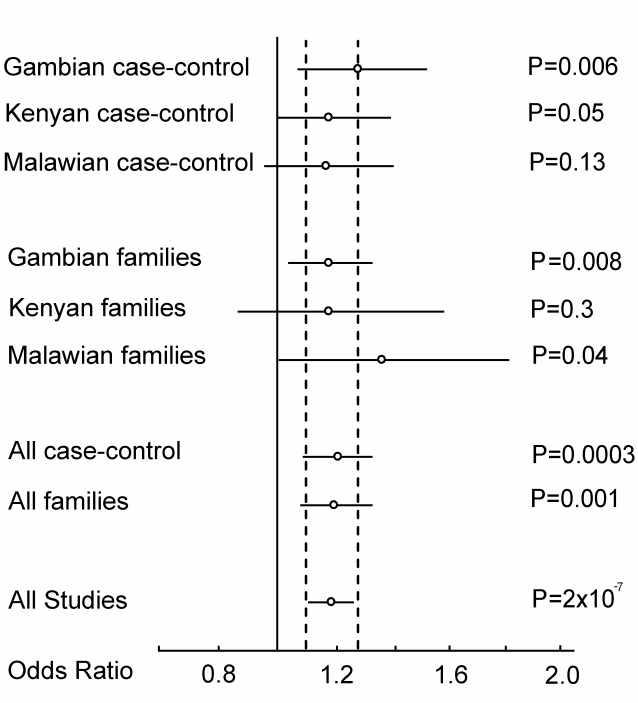


Figure 2.2. Estimated risk for *ABO* rs8176719 in severe malaria. For each study, the figure represents the estimated ORs (circular marks) and 95% CI (horizontal lines) for the full-length allele, in severe malaria. *P*-values were derived from regression analysis with covariates of ethnic group, gender and HbS genotype (for population-based studies), case-pseudo-control approach (for family studies) or UNPHASED analysis (pooled data across all studies). The interrupted lines represent the CI of the pooled data for all studies, both family and population based, using the UNPHASED application.

Table 2.1. Single SNP and haplotype association analysis for case-control and family studies.

			SNP						Haplotype				
			rs8176719			rs8176746			A			B	
Region	Phenotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
Population-based study													
Gambia	CM	1.21	0.98 - 1.49	0.08	0.97	0.75 - 1.25	0.81	1.37	1.04 - 1.79	0.025	1.09	0.83 - 1.42	0.54
Kenya	-	1.11	0.87 - 1.42	0.39	1.01	0.74 - 1.37	0.97	1.18	0.87 - 1.6	0.29	1.02	0.74 - 1.41	0.89
Malawi	-	1.18	0.97 - 1.43	0.09	1.02	0.81 - 1.29	0.84	1.31	1.02 - 1.68	0.034	1.07	0.83 - 1.37	0.60
All	-	1.17	1.04 - 1.32	0.011	1.00	0.86 - 1.16	0.99	1.29	1.1 - 1.51	0.001	1.06	0.91 - 1.24	0.46
Gambia	SA	1.29	1.02 - 1.65	0.037	1.01	0.77 - 1.34	0.94	1.41	1.04 - 1.91	0.028	1.17	0.86 - 1.58	0.31
Kenya	-	1.37	1.1 - 1.7	0.005	1.32	1.02 - 1.72	0.037	1.31	0.99 - 1.73	0.06	1.39	1.06 - 1.83	0.018
Malawi	-	1.16	0.83 - 1.6	0.38	0.96	0.63 - 1.46	0.86	1.29	0.85 - 1.97	0.23	1.04	0.67 - 1.6	0.87
All	-	1.30	1.12 - 1.5	0.0004	1.13	0.95 - 1.34	0.17	1.34	1.11 - 1.61	0.002	1.24	1.03 - 1.49	0.023
Gambia	Severe	1.27	1.07 - 1.5	0.006	1.06	0.87 - 1.29	0.55	1.40	1.12 - 1.74	0.003	1.16	0.94 - 1.44	0.17
Kenya	-	1.17	1 - 1.38	0.05	1.16	0.95 - 1.41	0.14	1.16	0.94 - 1.42	0.17	1.16	0.94 - 1.42	0.17
Malawi	-	1.16	0.96 - 1.39	0.13	1.01	0.8 - 1.27	0.95	1.28	1 - 1.63	0.047	1.05	0.82 - 1.34	0.71
All	-	1.20	1.09 - 1.32	0.0003	1.08	0.96 - 1.22	0.18	1.27	1.12 - 1.44	0.0003	1.13	0.99 - 1.28	0.06
Family-based study													
Gambia	CM	1.24	1.02 - 1.49	0.03	1.40	1.12 - 1.76	0.003	1.09	0.87 - 1.38	0.45	1.42	1.12 - 1.81	0.004
Kenya	-	1.26	0.84 - 1.87	0.27	1.10	0.67 - 1.82	0.70	1.22	0.76 - 1.97	0.41	1.21	0.69 - 2.12	0.50
Malawi	-	1.36	1.02 - 1.81	0.037	0.92	0.64 - 1.32	0.65	1.37	0.99 - 1.88	0.06	1.18	0.8 - 1.73	0.41
All	-	1.27	1.1 - 1.47	0.001	1.23	1.03 - 1.47	0.02	1.19	1 - 1.41	0.06	1.34	1.1 - 1.62	0.003
Gambia	SA	1.28	1.02 - 1.61	0.033	1.06	0.8 - 1.41	0.67	1.25	0.96 - 1.64	0.10	1.25	0.92 - 1.7	0.15
Kenya	-	1.27	0.76 - 2.12	0.36	0.95	0.51 - 1.78	0.87	1.29	0.73 - 2.28	0.37	1.04	0.54 - 1.99	0.91
Malawi	-	2.17	1.09 - 4.29	0.027	0.88	0.43 - 1.79	0.72	2.35	1.15 - 4.78	0.019	1.30	0.6 - 2.83	0.51
All	-	1.34	1.1 - 1.64	0.004	1.02	0.8 - 1.3	0.85	1.34	1.07 - 1.69	0.011	1.21	0.93 - 1.57	0.16
Gambia	Severe	1.17	1.04 - 1.32	0.008	1.15	0.99 - 1.33	0.06	1.13	0.98 - 1.3	0.10	1.23	1.05 - 1.44	0.01
Kenya	-	1.17	0.87 - 1.56	0.30	1.05	0.73 - 1.53	0.78	1.10	0.78 - 1.55	0.57	1.08	0.72 - 1.61	0.71
Malawi	-	1.35	1.01 - 1.78	0.04	0.91	0.64 - 1.29	0.60	1.36	1 - 1.87	0.05	1.15	0.79 - 1.69	0.46
All	-	1.19	1.08 - 1.32	0.001	1.10	0.97 - 1.25	0.13	1.16	1.03 - 1.31	0.018	1.20	1.05 - 1.37	0.009

P-values are from logistic regression with covariates of ethnic group, gender and HbS genotype (case-control) or case-pseudo-control approach and conditional logistic regression based on parental genotypes (family). Results for the multiplicative model. Commonest category (deletion allele and O haplotype) taken as reference.

Phenotypes: CM, cerebral malaria; SA, severe anaemia; Severe, all severe cases.

Table 2.2. Association analysis between severe malaria phenotypes and inferred blood group.

		Blood Group	A			B			AB	
Study/Region	Phenotype	OR	95% CI	*P*	OR	95% CI	*P*	OR	95% CI	*P*
Case-control										
Gambia	CM	1.13	0.78 - 1.61	0.52	0.96	0.67 - 1.37	0.83	2.14	1.12 - 4.07	0.021
Kenya	-	1.42	0.98 - 2.07	0.07	1.17	0.8 - 1.72	0.42	0.93	0.34 - 2.54	0.89
Malawi	-	1.53	1.11 - 2.12	0.01	1.09	0.8 - 1.5	0.58	1.32	0.75 - 2.34	0.34
All	-	1.36	1.11 - 1.66	0.003	1.07	0.87 - 1.31	0.53	1.51	1.02 - 2.22	0.04
Gambia	SA	1.52	1.02 - 2.26	0.041	1.22	0.82 - 1.83	0.33	1.76	0.79 - 3.91	0.17
Kenya	-	1.51	1.06 - 2.16	0.024	1.55	1.1 - 2.19	0.013	2.05	0.98 - 4.29	0.06
Malawi	-	1.66	0.96 - 2.86	0.07	1.15	0.65 - 2.02	0.63	1.33	0.5 - 3.5	0.57
All	-	1.54	1.22 - 1.96	0.00039	1.36	1.07 - 1.72	0.012	1.76	1.1 - 2.81	0.019
Gambia	Severe	1.27	0.95 - 1.7	0.10	1.07	0.8 - 1.42	0.65	2.04	1.16 - 3.58	0.013
Kenya	-	1.28	0.99 - 1.65	0.06	1.22	0.95 - 1.57	0.12	1.43	0.79 - 2.6	0.24
Malawi	-	1.49	1.08 - 2.04	0.014	1.06	0.78 - 1.45	0.70	1.31	0.75 - 2.3	0.34
All	-	1.33	1.13 - 1.56	0.00065	1.13	0.96 - 1.33	0.15	1.59	1.15 - 2.21	0.006
Family-based study										
Gambia	CM	1.14	0.83 - 1.57	0.42	1.56	1.11 - 2.21	0.011	2.66	1.4 - 5.05	0.003
Kenya	-	1.49	0.76 - 2.94	0.25	1.18	0.57 - 2.45	0.65	1.64	0.38 - 7.09	0.51
Malawi	-	1.60	1 - 2.57	0.05	1.09	0.63 - 1.86	0.77	1.85	0.79 - 4.35	0.16
All	-	1.30	1.01 - 1.66	0.038	1.37	1.05 - 1.8	0.02	2.23	1.38 - 3.61	0.001
Gambia	SA	1.56	1.06 - 2.29	0.025	1.43	0.94 - 2.17	0.10	1.46	0.69 - 3.09	0.32
Kenya	-	1.27	0.6 - 2.68	0.53	1.12	0.47 - 2.66	0.80	1.91	0.1 - 36.04	0.67
Malawi	-	1.60	0.53 - 4.8	0.41	0.98	0.31 - 3.12	0.97	2.13	0.49 - 9.16	0.31
All	-	1.51	1.09 - 2.09	0.014	1.32	0.92 - 1.88	0.13	1.58	0.83 - 3	0.16
Gambia	Severe	1.26	1.03 - 1.54	0.025	1.36	1.09 - 1.69	0.006	1.46	1 - 2.14	0.05
Kenya	-	1.20	0.75 - 1.9	0.45	1.02	0.6 - 1.72	0.94	1.23	0.39 - 3.84	0.72
Malawi	-	1.59	1 - 2.52	0.05	1.13	0.66 - 1.91	0.66	1.70	0.74 - 3.92	0.21
All	-	1.29	1.09 - 1.53	0.003	1.28	1.06 - 1.54	0.01	1.46	1.05 - 2.04	0.025

P-values are from logistic regression with covariates of ethnic group, gender and HbS genotype (case-control) or case-pseudo-control approach and conditional logistic regression based on parental genotypes (family). Commonest category (O blood group) taken as reference.

Phenotypes: CM, cerebral malaria; SA, severe anaemia; Severe, all severe cases.

2.3.3 ABO haplotypes producing A and B antigens are associated with susceptibility to severe malaria.

We inferred 2 SNP haplotypes using rs8176719 and rs8176746. In comparison with O haplotypes, both A (case-control OR, 1.27; 95% CI, 1.12–1.44, *P*= 0.0003; family OR, 1.16; 95% CI, 1.03–1.31, *P*= 0.018) and B (case-control OR, 1.13; 95% CI, 0.99–1.28, *P*= 0.06; family OR, 1.20; 95% CI, 1.05–1.37, *P*= 0.009) haplotypes are significantly associated with severe malaria (Table 2.1).

Using an individual's haplotypes, we inferred their ABO blood group (A and B haplotypes codominant, O haplotype recessive). Blood group A, B and AB individuals appear to be at significantly greater risk of severe malaria in comparison with blood group O (e.g. blood group A individuals: case-control OR, 1.33; 95% CI, 1.13–1.56, *P*= 0.00065; family OR, 1.29; 95% CI, 1.09–1.53, *P*= 0.003) (Table 2.2). Trends in the data suggest blood group B individuals may be at a subtly lower risk than blood group A, whereas blood group AB individuals are probably at the greatest risk of severe disease (case-control OR, 1.59; 95% CI, 1.15–2.21, *P*= 0.006; family OR, 1.46; 95% CI, 1.05–2.04, *P*= 0.025).

Non-O blood groups demonstrate particularly high risk of SA (e.g. blood group A: case-control OR, 1.54; 95% CI, 1.22–1.96, *P*= 0.00039; family OR, 1.51; 95% CI, 1.09–2.09, *P*= 0.014). The higher risk of SA experienced by individuals with non-O blood groups may reflect a pathophysiological effect, e.g. accelerated clearance of erythrocytes bound to iRBCs. Frequencies of genotypes, haplotypes and blood groups for all cases, controls and parents are documented in Appendix A, Table A.1.

2.3.4 ABO glycosyltransferase alleles and a possible parent-of-origin effect in disease susceptibility.

Previous association studies of the ABO system in severe malaria have examined serological data from unrelated individuals. In this study, having analysed genetic data from pedigrees, we were in the unique position of being able to look for parent-of-origin effects. We were surprised to find a substantial difference in transmission of high-risk alleles relating to their parent of origin. Full-length alleles of rs8176719 (marking A and B haplotypes) transmitted to offspring are associated with greater risk of severe disease if the allele was transmitted from a mother (OR 1.38, P= 0.0002) rather than a father (OR 1.05, P=0.6), conditional logistic regression fitting separate effects for maternal and paternal alleles supports the existence of a parent-of-origin effect (χ^2 = 3.96, P= 0.047) (Table 2.3).

Interactions between maternal genotype and child's genotype can masquerade as parent of origin effects. To investigate this scenario, we performed a parent-of-origin likelihood ratio test (PO-LRT) of Weinberg (202), which can allow for maternal genotype effects. The LRT demonstrates evidence of a parent-of-origin effect (P=0.046), but did not support a maternal genotype effect (P=0.21) (Table 2.4). Using 2-SNP haplotypes to identify the three ABO alleles, and repeating the conditional logistic regression, we found that both A and B haplotypes are associated with greater risk when transmitted by a mother although only the difference between maternal and paternal A haplotypes is significant (χ^2 = 5.19, P= 0.023).

Table 2.3. Evidence of a parent-of-origin effect at the ABO locus.

Variation/		CLR				TDT			
Region	Origin	OR	95% CI.	*P*	Wald	Observed	Expected	TDT χ^2	*P*
rs8176719									
Gambia	Maternal	1.35	1.1 - 1.65	0.004	-	233	202	9.51	0.002
-	Paternal	1.05	0.86 - 1.28	0.65	0.13	225	212	1.59	0.2
-	All	1.17	1.04 - 1.32	0.008		598	554	6.99	0.008
Malawi	Maternal	1.51	0.94 - 2.4	0.09	-	49	40	4.05	0.044
-	Paternal	1.20	0.74 - 1.93	0.46	0.55	43	37.5	1.61	0.2
-	All	1.35	1.01 - 1.78	0.04		113	98.5	4.27	0.039
Kenya	Maternal	1.48	0.92 - 2.35	0.1	-	46	39	2.51	0.11
-	Paternal	0.89	0.53 - 1.5	0.67	0.21	31	31	0.00	1
-	All	1.17	0.87 - 1.56	0.3		98	91	1.08	0.30
All	Maternal	1.38	1.16 - 1.65	0.0002	-	328	281	15.72	0.00007
-	Paternal	1.05	0.88 - 1.24	0.6	0.047	299	280.5	2.44	0.12
-	All	1.19	1.08 - 1.32	0.001		809	743.5	11.54	0.0007
ABO Haplotype									
A	Maternal	1.45	1.16 - 1.8	0.001	-	-	-	-	-
A	Paternal	0.99	0.8 - 1.22	0.894	0.023	-	-	-	-
B	Maternal	1.30	1.03 - 1.63	0.024	-	-	-	-	-
B	Paternal	1.17	0.92 - 1.49	0.203	0.58	-	-	-	-

CLR, conditional logistic regression; Wald, *P*-value for the Wald test comparing estimated parameters for maternal and paternal transmission; TDT, transmission disequilibrium test (203). Observed and expected rates of transmission of the minor allele along with classical TDT statistics are presented for comparison.

Table 2.4. Parent-of-origin likelihood ratio test for rs8176719.

Test	Model	-2LL	χ^2	df	*P*
-	Child and maternal genotypes, and PO	1365.582	-	-	-
Parent-of-origin effect	Child and maternal genotypes only	1369.58	3.998	1	0.046
Maternal genotype effect	Child genotype and PO only	1368.714	3.133	2	0.21

LL, log likelihood; χ^2, Log ratio test χ^2 value; df, degrees of freedom, *P*-value, *P*-value for the likelihood ratio test; PO, parent-of-origin effect included in model.

There was no significant difference in risk estimate for severe malaria between AO and AA genotypes in either the case-control study (AO: OR, 1.35; 95% CI, 1.13–1.60; AA: OR, 1.23; 95% CI, 0.84–1.79; Wald test *P*= 0.64) or family study (AO: OR, 1.31; 95% CI, 1.10–1.55; AA: OR, 1.13; 95% CI, 0.76–1.67; Wald *P*= 0.45). The same was true for BO versus BB genotypes, and grouping both non-O haplotypes. The phase of AO genotypes was distinguishable in the family study and, as noted above, suggested different risks of severe disease ($A^{mat}O^{pat}$: OR, 1.64; 95% CI, 1.27–2.12; $O^{mat}A^{pat}$: OR, 1.04; 95% CI, 0.81–1.33; Wald test *P*= 0.01). Differences between these genotypes and the AA genotype were not significant ($A^{mat}O^{pat}$ versus AA; Wald test *P*= 0.07; $O^{mat}A^{pat}$ versus AA; Wald test *P*= 0.92).

2.3.5 Low levels of population differentiation at the *ABO* locus represent a signal of longstanding balancing selection.

Given that one ABO allele is protective against a major selection pressure such as *P. falciparum* malaria, it is important to consider the reasons why the ABO system remains polymorphic in Africa. The answer may relate to the balance of protection offered by specific ABO alleles to other infectious disease. ABO antigens have been implicated not only in malaria, but also to the pathogenesis of *Escherichia coli* (142), *Helicobacter pylori* (143), Norwalk virus (141), Hepatitis C (204), and respiratory infections (140). There has also been speculation about whether historical agents such as smallpox and plague may have shaped the global distribution of ABO alleles (205, 206).

Blood group A, B and O alleles occur in other primate species. Sequence analysis has suggested that the common ancestral enzyme had A transferase activity and that the

presence of species-specific mutations underlying the non-human B and O alleles suggests functional polymorphism of the *ABO* gene has occurred through convergent evolution (207). High levels of nucleotide diversity at the *ABO* locus are considered to be significant evidence of non-neutral evolution in primate lineages.

Surveys for signatures of evolutionary selection in the human genome have demonstrated evidence of balancing selection at the *ABO* locus (44, 208). Using our merged haplotype data from the HapMap Yoruba population and the additional genotyping of four functional polymorphisms (rs8176719, rs8176746, rs8176747 and rs8176743), we investigated whether the long-range haplotype (LRH) test (51), which detects signals of recent positive selection, was associated with the functional *ABO* variants (Figure 2.3). Neither of our key SNPs appear to be associated with an extended haplotype homozygosity signal, which is consistent with a long-standing process of balancing selection. A recently reported high LRH score from the *ABO* locus (44) may reflect positive selection of nearby regulatory sequence controlling expression patterns rather than coding sequence itself. Interestingly, we observed very low levels of population differentiation across the *ABO* locus in HapMap populations (Figure 2.4A). The *ABO* locus including sequences 20-30 kb upstream, a region of 85 SNPs, has an F_{ST} (using a sample-size weighted metric of population differentiation between African, Asian and European populations) of -0.001, compared with a genome-wide average of ~0.1. Although more attention is generally given to high F_{ST} values, as signals of region-specific positive selection [e.g. the Duffy blood group locus under selection in Africa from *P. vivax* malaria (209)], here the low F_{ST} might reflect simultaneous balancing selection in all three populations, and would be consistent with a model of longstanding selection. The extension of the low F_{ST} region 20 – 30kb upstream of the *ABO* coding

sequence (while stopping just 3kb downstream) may indicate that the balancing selection is also shaping the *cis*-acting regulatory sequences immediately upstream of the gene. Although not a formal test of deviation from neutrality, we measured the empirical distribution of F_{ST} in windows across chromosome 9 and found the *ABO* region to be around the 99.5 – 99.9th centile in this distribution (Figure 2.4B).

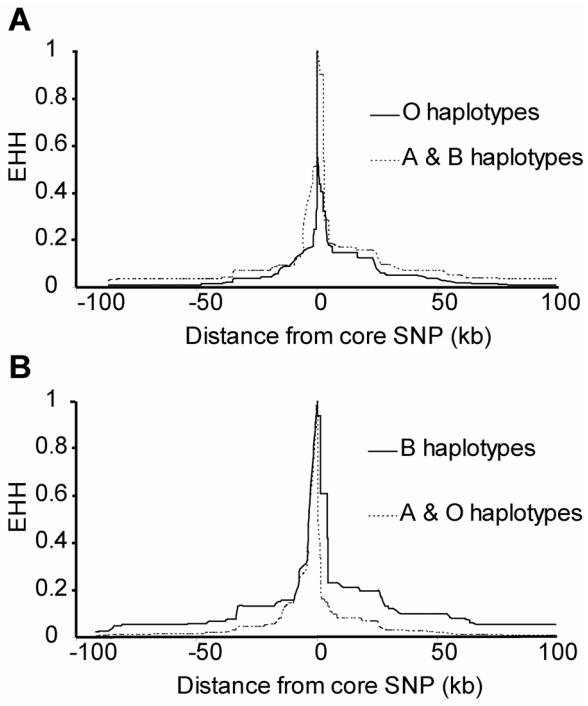


Figure 2.3. Extended haplotype homozygosity plots around functional SNPs in the *ABO* locus. Alleles that have risen rapidly in frequency because of recent positive or balancing selection (e.g. a partial selective sweep) can be surrounded by a region of similar haplotypes that can extend for hundreds of kilobases (44, 51). This occurs because recombination has had insufficient time to swap variation between the selected haplotype and others in the population. We plotted the decay of homozygosity (EHH) on phased haplotypes partitioned by the alleles of the functional *ABO* SNPs. (A) rs8176719, the frameshift mutation in exon 6 of the *ABO* gene, marking O haplotypes, which is associated with protection from severe malaria. (B) rs8176746, a nonsynonymous coding SNP in the N-terminal catalytic domain of ABO, one of the functional variants determining A/B glycosyltransferase activity. The lack of a pronounced EHH signal suggests that the balanced selection affecting variation at the *ABO* gene is longstanding.

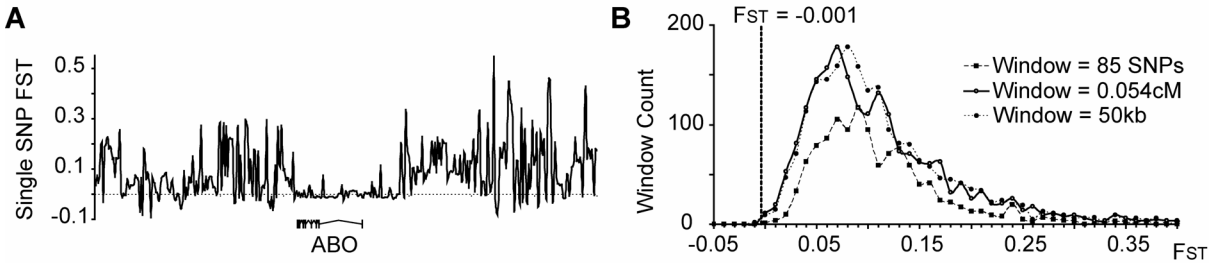


Figure 2.4. A region of Low F_{ST} region around the *ABO* gene. (A) 400 Single SNP F_{ST} values for three HapMap populations (CEU, YRI and combined Asian) surrounding the *ABO* gene. F_{ST} drops to a mean level of -0.001 in an 85 SNP window across the gene and the noncoding sequence ~20-30kb upstream. (B) Histogram representing an empirical distribution of F_{ST} determined by screening similarly sized windows across chromosome 9. Three different window sizes were used based on either marker numbers (85 SNPs), genetic distance (0.054 cM) or physical distance (50kb), only windows containing more than two markers were included. By all three distributions, the region around the *ABO* locus is a relative outlier ~99.5 – 99.9th centile for low F_{ST}.

2.4 Discussion

Our analysis strongly supports the hypothesis that blood group O individuals are relatively protected from severe malaria in comparison to other blood groups, particularly blood group A and AB. In addition to population-based studies, we conducted the first family-based association studies of *ABO* variation, with severe malaria. The analysis of parent-offspring trios is robust to artifacts of population stratification and allowed us to examine the possibility of parent-of-origin effects. We found that full-length alleles at rs8176719, particularly A haplotypes, inherited from a mother lead to greater risk for offspring than similar alleles inherited from a father. The analysis suggests a phenomenon such as genomic imprinting rather than an effect of maternal genotype. The tissue- and cell-type specific expression of *ABO* is controlled by epigenetic signals, particularly at a 5'CpG island in the promoter of the gene (210). The loss of *ABO* expression mediated by promoter hypermethylation is common in certain forms of malignancy (211). There has also been a small case series documenting preferential loss of the maternally derived ABO allele in adult leukaemia (212). Our results would be consistent with altered expression of the paternal ABO allele because of imprinting, leaving less target ligand for iRBCs to bind to, or shifting the pattern of sequestration to a 'safer' distribution. Paternal allele expression is not abolished, as demonstrated by the existence of blood group AB. However, it is interesting to speculate that during gestation it could be beneficial to modulate or suppress the paternal ABO allele, so as to minimize maternal-fetal ABO incompatibility and the risk of haemolytic disease of the newborn. To our knowledge, this is the first report suggestive of genomic imprinting in the genetics of malaria susceptibility. However, despite our sample size, we have limited power to detect the subtle difference in risk between maternal and

paternally derived alleles, it remains possible that this suggestive result is in fact type I error. Further experiments are required to confirm this finding, and could include family-based studies of *ABO* variation in malaria and other complex diseases, or functional experiments with family-derived samples to determine whether *ABO* gene expression is affected by parent of origin.

The sizes of our individual study datasets gives us limited power to analyze the differences between regions. The differences in OR between a family study and case-control study in a single population are of similar magnitude to the differences seen between regions, suggesting much of the variation observed is because of statistical fluctuation or methodological issues. Factors possibly contributing to the diversity of risk estimates reported between our populations, and in comparison with other reports include variation in phenotype definition and geographic differences in parasite strain, e.g. differing affinities for ABO antigens. The use of cord blood controls instead of age-matched controls could affect association results. Cord blood samples represent the distribution of genotypes in the populations at birth; since a minority will go on to develop severe disease, this will tend to underestimate the true odds ratio of any genetic susceptibility or resistance allele. In contrast, if blood group O individuals are depleted from the population between birth and mid-childhood (the age of the cases) because of non-malarial pathogens, this would lead to an overestimation of protection from blood group O. However, given the consistent results between the population- and family-based studies, the similar allele frequencies between cord blood samples and untransmitted parental alleles and the modest effect size found, such biases are probably limited.

Our study employed samples from children with severe malaria phenotypes and it has been noted that the ABO effect is harder to detect in samples incorporating adults with malaria (146). The gradual accumulation of antibodies against a range of iRBC surface antigens is thought to explain a component of the immunity to severe life-threatening malaria that develops in late childhood in endemic regions (213). It is possible that a limited repertoire of parasite epitopes binding A and B antigens could lead to a high prevalence of effective antibodies against these variants in adults, reducing the significance of host ABO blood group in older age groups. It is interesting to speculate whether vaccination, leading to effective immunity to A and B antigen binding PfEMP1 variants, could reduce morbidity among blood group A, B and AB children.

The modest effect size found is typical of validated associations with complex disease in humans. The effect size is likely, in part, to explain why some previous studies have been unable to demonstrate significant results, despite trending towards protection from blood group O, particularly when sample size was low. However, the global health impact of the frameshift deletion, underlying blood group O, needs to be put in context by considering its high frequency. Roughly half the peoples of sub-Saharan Africa, and many other human populations, at risk of life-threatening disease caused by *P. falciparum*, are homozygotes for this null mutation and are protected by being blood group O (214).

2.5 Materials and Methods

2.5.1 Subjects.

Patient samples were collected as part of ongoing epidemiological studies of severe malaria at the Royal Victoria Hospital, Banjul, The Gambia; the Queen Elizabeth Central Hospital, Blantyre, Malawi and Kilifi District Hospital, Kilifi, Kenya. Populations in these study sites are exposed to endemic malaria, with the burden of life-threatening disease being experienced by children (< 5 years old). Nuclear family trios comprised two parents and one affected child. All DNA samples were collected and genotyped following approval from the relevant research ethics committees and informed consent from participants. Controls were cord blood samples obtained from birth clinics in the same locality as the cases. See Appendix A, Table A.2 for further demographics of patients and controls.

2.5.2 Phenotype definition.

All cases were children admitted to hospital with evidence of *P. falciparum* on blood film and clinical features of severe malaria (215, 216). In our analyses of sub-phenotypes, we use Blantyre coma score of ≤2 as a criterion of CM, and haemoglobin <5g/dl or packed cell volume <15% as a criterion of SA. Some individuals had both CM and SA. Of the severe malaria cases that were not CM or SA by these criteria, most had lesser degrees of coma (Blantyre coma score 3) or anaemia (Hb 5-6g/dl), or other complications such as respiratory distress. Our samples comprised:

i) 701 Gambian cases (324 with CM and 217 with SA) and 624 controls.

ii) 718 Malawian cases (640 CM, 101 SA) and 405 controls.

iii) 708 Kenyan cases (216 CM, 270 SA) and 902 controls.

iv) 1320 Gambian trios (512 probands with CM and 343 with SA).

v) 225 Malawi trios (216 CM, 39 SA).

vi) 234 Kenyan trios (114 CM, 85 SA).

2.5.3 Power calculations.

Power calculations were performed using the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc) (217). A single regional case-control study had ~61% power, based on a sample size of 700 cases, 700 unselected controls, allelic OR of 1.2 (our risk estimate for non-O alleles compared with O alleles), high-risk allele frequency of 0.3 and a type I error rate of 0.05. Across all case-control studies, we would expect 97% power. The smaller family trio studies (n=230 trios) had ~25% power, while the larger Gambian study (n=1320 trios) had ~87% power.

2.5.4 Sample preparation and genotyping.

Genomic DNA samples underwent whole genome amplification through either primer extension pre-amplification (PEP) (218) or multiple displacement amplification (MDA) (219), before genotyping on a Sequenom MassArray genotyping platform (220). All assays achieved high rates of genotyping success and no deviations from Hardy-Weinberg equilibrium were encountered (Appendix A, Table A.2).

2.5.5 Statistical analysis.

Analysis was performed using STATA (v9.2 for windows) and the genassoc package (http://www-gene.cimr.cam.ac.uk/clayton/software/) written for STATA by David Clayton. In general, the results for the multiplicative model are presented. Although full-length alleles are dominant with regard to blood group phenotype, we found a trend towards increased susceptibility for blood group AB individuals (homozygotes for the full-length allele at rs8176719), which makes the multiplicative model valid. Case-control association analysis was undertaken by logistic regression and included the covariates of ethnic group, gender and Sickle status. DNA Sequenom genotyping for the Haemoglobin S (HbS) variant was performed for all samples. The HbS results for a proportion of the Gambian samples have previously been published (40). Gender was included in the regression analysis to control for sex-linked traits (e.g. G6PD). With regards to other variation thought to affect malaria susceptibility, Haemoglobin C was absent from our study populations and the range of deletions underlying the Thalassemias are not easily amenable to the high-throughput genotyping technology used.

The family-based association analysis was performed using a case-pseudo-control approach and conditional logistic regression (221). Trios were drawn from a larger pool of samples checked for relationship misspecification. All family samples were genotyped for 48 SNP markers and 15% of trios (indicated by their Mendelian errors rates) excluded from further analysis.

In the population-based studies two SNP haplotypes (and therefore the ABO alleles) were reconstructed using the snp2hap function, whereas in family studies phase can often be tracked from parent and child in the creation of the case-pseudo-control dataset. Data from the case-control studies was pooled in a single logistic regression analysis including covariates of ethnic group, gender and Sickle status, whereas family studies were combined in a single case-pseudo-control analysis. Pooling across all case-control and family-studies is less straight-forward. Here we used the UNPHASED application version 3.0 (http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/) (222, 223) which employs a retrospective likelihood framework for performing genetic association analysis, and can be used to combine data from nuclear families and unrelated subjects. Ethnic origin was found to be a significant confounder and was retained as a covariate in the UNPHASED analysis.

2.5.6 Integrating phased genotypes from the HapMap project.

Phased *ABO* locus haplotypes for the 30 Yoruba parent-offspring trios from Ibadan in Nigeria were downloaded from the HapMap website (http://www.hapmap.org) (57). Cell lines derived from the Yoruba individuals (obtained from the Coriell Cell Repositories, http://ccr.coriell.org/) were genotyped for the four key SNPs. The additional SNP data was combined with the known phased genotypes using PHASE version 2.1 (224, 225). LD patterns between our genotyped SNPs and between these genotypes and HapMap markers were visualized using HAPLOVIEW, version 3.32 (http://www.broad.mit.edu/mpg/haploview/) (226). Extended haplotype homozygosity values (51) were calculated using SWEEP, version 1.0 (http://www.broad.mit.edu/mpg/sweep/index.html).

2.5.7 F_{ST} calculation and window-based screen of chromosome 9.

Single nucleotide polymorphism genotyping data for chromosome 9 was downloaded from the HapMap website (Release 21a/Phase 2, Jan 2007). The dataset was derived from three populations: the 30 Yoruba trios; 30 US parent-offspring trios of northern and western European origin, collected by the Centre d'Etude du Polymorphisme Humain (CEPH); and a combined dataset of 45 unrelated individuals from the Tokyo, Japan and 45 unrelated individuals from Beijing, in China (57). A total of 94,678 SNPs across chromosome 9 were polymorphic in all three populations. The F_{ST} among the three populations was estimated, in windows across chromosome 9, using the formula for F_{ST} described in the supplementary information to the HapMap Consortial publication (57).

2.6 Acknowledgments

A.E.F. and M.J.G. were funded by Wellcome Trust Clinical Research Training Fellowships and T.W. by a Wellcome Trust Senior Fellowship. The sample collections and resources used in this study received funding from the UK Medical Research Council, the Wellcome Trust, the Bill and Melinda Gates Foundation, the National Institutes of Health and the European Union Network 6 BioMalPar Consortium. This manuscript is published with the permission of the director of the Kenya Medical Research Institute.

Chapter 3

Variation in the *ICAM1* gene is not associated with severe malaria phenotypes.

Andrew E. Fry[1], Sarah Auburn[1], Mahamadou Diakite[1], Angela Green[1], Anna Richardson[1], Jonathan Wilson[1,†], Muminatou Jallow[2], Fatou Sisay-Joof[2], Margaret Pinder[2], Michael J. Griffiths[1, 3], Norbert Peshu[3], Thomas N. Williams[3,4], Kevin Marsh[3, 4], Malcolm E. Molyneux[5,6], Terrie E. Taylor[7,8], Kirk A. Rockett[1], Dominic P. Kwiatkowski[1,9]

[1] The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, UK. [2] Medical Research Council, PO Box 273, Banjul, The Gambia. [3] Kenya Medical Research Institute Centre for Geographical Medicine Research (Coast), P.O. Box 230, Kilifi, Kenya. [4] Nuffield Department of Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK. [5] Malawi–Liverpool–Wellcome Trust Programme of Clinical Tropical Research, PO Box 30096, Blantyre, Malawi. [6] Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK. [7] Blantyre Malaria Project, College of Medicine, PO Box 30096, Blantyre, Malawi. [8] Department of Internal Medicine, College of Osteopathic Medicine, Michigan State University, East Lansing, Michigan 48824, USA. [9] Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK [†] Deceased

3.1 Abstract

Evidence from autopsy and *in vitro* binding studies suggests that adhesion of erythrocytes infected with *Plasmodium falciparum* to the human host intercellular adhesion molecule (ICAM)-1 receptor is important in the pathogenesis of severe malaria. Previous association studies between polymorphisms in the *ICAM1* gene and susceptibility to severe malarial phenotypes have been inconclusive and often contradictory. We performed genetic association studies with 15 single nucleotide polymorphisms (SNPs) around the *ICAM1* locus. All SNPs were screened in a family study of 1071 trios from The Gambia, Malawi and Kenya. Two key nonsynonymous SNPs with previously reported associations, rs5491 (K56M or 'ICAM-1^{Kilifi}') and rs5498 (K469E), were tested in an additional 708 Gambian trios and a case-control study of 4058 individuals. None of the polymorphisms were associated with severe malaria phenotypes. Pooled results across our studies for ICAM-1^{Kilifi} were, in severe malaria, odds ratio (OR), 1.02; 95% confidence interval (CI) 0.96 – 1.09, P=0.54, and cerebral malaria OR, 1.07; CI 0.97 – 1.17, P=0.17. We assess the available epidemiological, population genetic and functional evidence that links ICAM-1^{Kilifi} to severe malaria susceptibility.

3.2 Introduction

Several lines of evidence implicate intercellular adhesion molecule (ICAM)-1 in the pathogenesis of severe *Plasmodium falciparum* malaria. The causes of severe disease are complex, but a key feature is the adherence of parasite infected red blood cells (iRBC) to components of the vascular space and particularly endothelium (73). Adhesion of iRBCs to brain microvasculature is suspected to contribute to the development of cerebral malaria (CM) a manifestation of severe disease associated with high rates of mortality (20, 227). *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) is central to this adhesive behaviour. PfEMP1 is expressed by the parasite on the surface of iRBCs, where it is subject to antigenic variation during the course of an infection (228). PfEMP1 adheres to a range of molecules expressed by the human host including ICAM-1 (100). IRBCs bind the first N-terminal immunoglobulin-like domain of ICAM-1 (229). Autopsy studies of patients with fatal CM and severe malarial anaemia (SA) found iRBC sequestration on brain vascular endothelial cells, with enhanced expression of adhesion molecules, including ICAM-1, in the areas of iRBC binding (98, 230-232). Studies of parasite isolates have found high rates of *in vitro* ICAM-1 binding among wild strains, but reported correlations between ICAM-1 binding and disease severity have been inconsistent (113, 118, 233).

ICAM-1 (CD54), a member of the immunoglobulin superfamily, is typically expressed on endothelial cells, particularly in the brain where expression is strongly increased by proinflammatory cytokines (96). As a key component of the immune system ICAM-1's role in malaria susceptibility is not limited to direct interaction with the PfEMP1. ICAM-1 binds lymphocyte function-associated antigen (LFA)-1, a leukocyte cell surface

glycoprotein (234), allowing leukocytes passage through the blood brain barrier. Interaction between LFA-1 and ICAM-1 activates natural killer cells during *P. falciparum* infection (235). ICAM-1 is also a receptor for plasma protein fibrinogen (236), Mac-1 (237) and Rhinovirus (238).

In 1997 Fernandez-Reyes *et al.* (102) sequenced the N-terminal domain of ICAM-1 in asymptomatic Kenyan children and found a nonsynonymous coding polymorphism. The newfound allele, designated ICAM-1^{Kilifi} (rs5491), caused a lysine to methionine change at position 56 of the coding sequence (position 29 in the mature protein). This raised the possibility that the mutant allele had been selected for by endemic malaria. However, (to their surprise) a case-control study of Kenyan children (260 severe malaria cases and 287 community controls) suggested that the ICAM-1^{Kilifi} allele was associated with an increased susceptibility to CM. The authors noted ICAM-1^{Kilifi} was common (~30%) in Kenyan and Gambian populations, but not found in a sample of Caucasians, leading the authors to speculate that ICAM-1^{Kilifi} provided a compensatory selective advantage through an unknown mechanism. In contrast to the first study, a paired case-control analysis from Gabon (100 severe malaria cases and 100 mild cases), published in 1999, reported a protective effect from ICAM-1^{Kilifi} (106). No subsequent association study has replicated either of these initial findings. However, case-control studies in The Gambia (105), Thailand (109), Senegal (107), Nigeria (94) and further case-control (95) and longitudinal studies (108) in Kenya reported no significant association between malaria phenotypes and ICAM-1^{Kilifi}. Case-control analysis in Nigeria did suggest a marginal signal of association with susceptibility to another single-nucleotide-polymorphism (SNP) in exon 6 (rs5498).

Although there is a strong rationale why genetic variation around the *ICAM1* gene could affect susceptibility to severe malarial phenotypes, only a fraction of the common SNPs around the locus have been tested for disease associations, and the results for the closely studied ICAM-1^{Kilifi} variant are inconclusive. We designed and conducted an experiment to screen common variation across *ICAM1* while targeting the two nonsynonymous *ICAM1* SNPs (ICAM-1^{Kilifi} and rs5498, those with previously reported associations) with well-powered tests.

3.3 Results

3.3.1 Family-based association analysis of *ICAM1* SNPs with severe malaria phenotypes.

Fifteen SNP makers around the *ICAM1* gene were selected for genotyping in 1071 parent-offspring trios from The Gambia, Kenya and Malawi. Each trio comprised a child affected by severe malaria and their parents. The nonsynonymous SNPs ICAM-1^{Kilifi} and rs5498 were also genotyped in 708 additional Gambian families. Further details of subjects and severe malaria phenotypes are given in the Materials and methods (Section 3.5). Markers were tested for disease associations in additive, dominant and recessive models using the FBAT application. The results for the additive model in severe malaria, along with allele frequencies for parents and offspring in the family study are shown in Table 3.1 (the association results for sub-phenotypes and specific populations are documented in Appendix B, Table B.1).

No significant association was found between *ICAM1* SNPs and severe malaria phenotypes, or between the SNPs and severe malaria in a specific study region (all *P*-values > 0.01). Given that 15 markers are tested for association with three phenotypes, in three genetic models, all *P*-values are substantially above relevant thresholds following correction for multiple testing. With specific reference to ICAM-1^{Kilifi} this variant was not found to be associated with severe disease (N=1779 trios, P=0.93) or CM (N=842 trios, P=0.1) in our family study.

Table 3.1. *ICAM1* genotype frequencies and single marker family-based association analysis with severe malaria.

SNP	Major Allele	Minor Allele	Parents MAF	Cases MAF	*S*	*E(S)*	*Var(S)*	*P-value*
rs5490	A	C	26.0	26.0	429	430.0	190.50	0.94
rs5030340	C	T	5.5	5.5	100	105.0	52.00	0.49
rs5030344	G	A	4.2	4.3	87	86.5	43.25	0.94
rs5030351	C	T	24.6	24.8	414	417.5	183.75	0.80
rs5491	A	T	23.9	23.7	687	688.5	303.75	0.93
rs281432	G	C	24.8	24.7	383	398.5	180.25	0.25
rs5496	G	A	9.5	9.7	176	179.5	88.25	0.71
rs5498	A	G	11.4	11.0	346	368.0	182.00	0.10
rs3093030	C	T	5.8	5.6	100	112.0	55.50	0.11
rs281438	T	G	46.0	46.6	684	683.5	237.25	0.97
rs2569693	C	T	5.5	5.2	87	89.5	43.75	0.71
rs281439	C	G	38.8	38.6	583	589.5	236.25	0.67
rs281440	A	G	31.9	31.0	461	478.5	202.75	0.22
rs2075741	G	C	12.2	11.8	221	236.5	111.75	0.14
rs11575074	G	A	9.7	9.6	172	182.0	89.00	0.29

Abbreviations: MAF, minor allele frequencies; SNP, single nucleotide polymorphisms.

Minor allele frequencies (MAF, %) for parents and affected offspring. FBAT statistic ‘S’, expected FBAT statistic ‘E(S)’, variance ‘Var(S)’ and *P*-value are reported under the additive model for transmissions of the minor allele.

Table 3.2. FBAT haplotype association analysis for the 5' *ICAM1* haplotype block in severe malaria phenotypes

Phenotype	Haplotype	Freq.	S	E(S)	Var(S)	*P*
All severe	ACGCAGG	0.288	444.43	446.08	162.48	0.90
	- - - - - C -	0.241	372.16	379.23	135.87	0.54
	C - - T T - -	0.210	328.81	326.72	122.41	0.85
	- - - - - - A	0.087	149.73	150.40	67.36	0.94
	- T - - - - -	0.053	95.87	90.13	39.40	0.36
CM	ACGCAGG	0.279	248.30	247.77	92.88	0.96
	- - - - - C -	0.234	212.09	221.48	74.32	0.28
	C - - T T - -	0.222	209.42	206.29	73.68	0.72
	- - - - - - A	0.085	91.86	87.62	37.08	0.49
	- T - - - - -	0.062	62.86	60.23	25.49	0.60
SA	ACGCAGG	0.298	132.31	130.89	47.88	0.84
	- - - - - C -	0.227	100.32	102.52	36.75	0.72
	C - - T T - -	0.206	83.27	93.48	33.74	0.08
	- - - - - - A	0.089	43.80	43.79	17.86	1.00
	- T - - - - -	0.052	27.00	23.91	10.41	0.34

Abbreviations: CM, cerebral malaria; SA, severe malarial anaemia. Parental haplotype frequency derived from seven 50 *ICAM1* SNPs from rs5490 to rs5496, FBAT statistic 'S', expected FBAT statistic 'E(S)', variance 'Var(S)' and *P*-value are reported under the additive model for transmissions of the specific haplotype. Only haplotypes >5% frequency analysed.

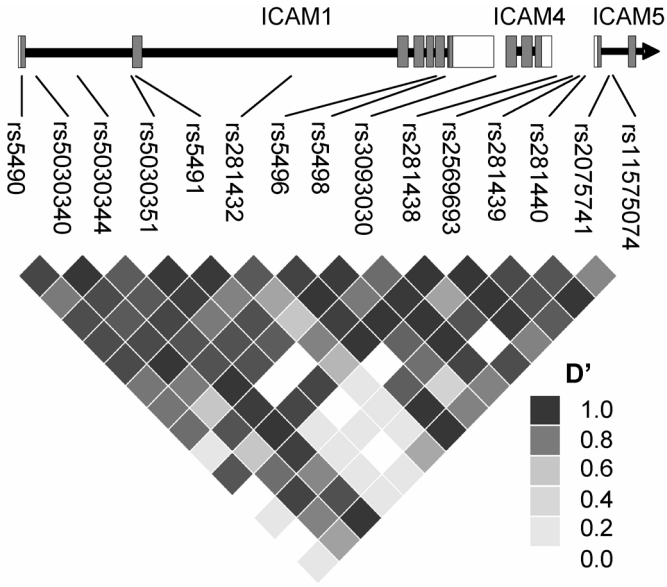


Figure 3.1. Linkage disequilibrium at the *ICAM1* locus in 612 Gambian family trios. An exonic model of the *ICAM1* gene (along with the nearby *ICAM4* and *ICAM5* genes), demonstrating the relative positions of our genotyped SNPs. All SNPs were within 5kb of the *ICAM1* gene. LD (D') calculated using the HAPLOVIEW application. A region of relatively high D' can be seen across the 5' of the gene, a section coding for the PfEMP1 binding N-terminal of ICAM-1. This haplotype block (7 SNPs from rs5490 to rs5496) was used in our haplotype association analysis. LD, linkage disequilibrium; SNP, single nucleotide polymorphisms.

3.3.2 Haplotype-specific association analysis

Using family trio data, we performed haplotype-based association analysis with severe malaria phenotypes in FBAT. Linkage disequilibrium between *ICAM1* SNPs was inspected using the HAPLOVIEW application. A haplotype block extending across seven SNPs in the 5' region of the gene was identified, common to all three study populations (Figure 3.1). This 5' haplotype block spans across the exons coding for the first N-terminal domain, which binds PfEMP1, LFA-1 and Rhinoviruses and is the location of ICAM-1^{Kilifi}. FBAT analysis of common haplotypes (>5%) revealed no significant disease associations (Table 3.2).

3.3.3 Population-based association analysis of ICAM-1^{Kilifi} and rs5498 with severe malarial phenotypes.

We genotyped the two nonsynonymous SNPs, with previously reported malaria associations, in three additional population-based studies. The case-control studies were independent of the family study and in total comprised 2127 cases of severe malaria and 1931 population controls. Further details of subjects and severe malaria phenotypes are given in the Materials and Methods (Section 3.5). Both markers were tested for disease associations in additive, dominant and recessive models using logistic regression. Analysis of all case-control studies together suggests neither marker is significantly associated with our severe malarial phenotypes (Table 3.3).

We specifically tested whether ICAM-1^{Kilifi} homozygotes were at altered risk of CM in individual regional populations. In the Gambian case-control study, ICAM-1^{Kilifi}

homozygotes demonstrated a possible association with CM susceptibility (taking wild-type homozygotes as reference; ICAM-1^{Kilifi} homozygotes odds ratio (OR), 2.5; 95% confidence interval (CI) 1.37 - 4.75, *P*=0.003), heterozygotes were not at significantly increased risk (OR, 1.03; 95% CI 0.75 - 1.42, *P*=0.85). This is reflected in the result for all severe cases in Gambia under the additive model (Table 3.3). A similar, but non-significant, trend was found in the family-study (ICAM-1^{Kilifi} homozygotes OR, 1.3; 95% CI 0.75 - 2.24, *P*=0.35 and heterozygotes OR, 1.0; 95% CI 0.784 - 1.3, *P*=0.94) [performed by case-pseudo-control logistic regression conditioning on parental genotypes (221)]. This finding merits cautious interpretation, as (a) ICAM-1^{Kilifi} homozygotes represent a relatively a small sub-group of the Gambians (6% of all Gambian cases, 50 individuals); (b) the trend was not followed in the other two populations, including Kenyans where a link with CM was originally reported (102) and (c) no association between CM and ICAM-1^{Kilifi} was reported in a previous Gambian study (105). No other significant region- and sub-phenotype-specific associations were observed. Pooled family- and population-based association analysis was performed using the UNPHASED application (Table 3.3). Using this large combined dataset we have narrowed the CI of the risk estimate for ICAM-1^{Kilifi}, which does not appear to be associated with severe disease (OR 1.02; 95% CI 0.96 – 1.09) or CM (OR, 1.07; 95% CI 0.97 – 1.17).

Table 3.3. Case-control and pooled case-control/family data association analysis between ICAM-1^{Kilifi} (rs5491), rs5498 and severe malaria phenotypes.

SNP	Phenotype	OR	95% CI.	P
Case-control studies [a]				
rs5491	CM	1.07	0.94 - 1.21	0.33
-	SA	1.10	0.95 - 1.28	0.21
-	All Severe	1.10	1 - 1.22	0.06
-	Gambia	1.23	1.01 - 1.5	0.04
-	Kenya	1.07	0.92 - 1.25	0.40
-	Malawi	1.05	0.87 - 1.27	0.62
rs5498	CM	1.12	0.95 - 1.31	0.19
-	SA	0.97	0.79 - 1.18	0.75
-	All Severe	1.06	0.93 - 1.21	0.41
-	Gambia	1.10	0.86 - 1.42	0.45
-	Kenya	1.07	0.87 - 1.32	0.52
-	Malawi	0.99	0.78 - 1.26	0.94
Pooled across case-control and family studies [b]				
rs5491	CM	1.065	0.97 - 1.17	0.17
-	SA	0.988	0.88 - 1.11	0.85
-	All Severe	1.021	0.96 - 1.09	0.54
rs5498	CM	1.087	0.97 - 1.22	0.16
-	SA	0.924	0.79 - 1.09	0.34
-	All Severe	1.029	0.94 - 1.12	0.53

Abbreviations: CM, cerebral malaria; SA, severe malarial anaemia.

[a] Results for severe malaria phenotypes across case-control studies and for severe malaria in regional case-control studies.

[b] Results for severe malaria phenotypes pooled across case-control and family studies (using UNPHASED).

3.4 Discussion

There is a strong rationale why genetic variation around the *ICAM1* gene, particularly the ICAM-1^{Kilifi} variant could affect susceptibility to severe malarial phenotypes. However, genetic epidemiological studies of the ICAM-1^{Kilifi} variant have demonstrated a pattern consistent with other candidate gene studies of complex human disease, in which early interesting results have not been replicated in subsequent studies (239) (Table 3.4). We conducted an association screen for common variation across the *ICAM1* gene, while targeting two nonsynonymous SNPs with well-powered tests, using family- and population-based methods. This has led to a substantial refinement of the risk estimates for ICAM-1^{Kilifi} and rs5498. Neither of these nonsynonymous SNPs, nor any other genotyped *ICAM1* SNP were significantly associated with severe disease or CM in our study. We cannot rule out the possibility that ICAM-1^{Kilifi} is associated with a small effect (e.g. an odds ratio of 1.1 or 1.05), particularly if it is limited to smaller subgroups for example, a sub-phenotype of severe disease, or only the rare ICAM-1^{Kilifi} homozygotes. However, our results do not support the existence of the large effect sizes initially reported. Differences in clinical phenotype definition between study sites could confound results. In our study, we have attempted to standardize phenotype definition using key clinical parameters (such as Blantyre Coma Score or haemoglobin concentration), but it is possible that disparity in phenotype definition between studies is responsible for the diversity of findings reported in the literature.

Table 3.4. Published risk estimates of severe malaria phenotypes associated with ICAM-1^{Kilifi}.

Study	Region	Design	n[a]	Phenotype[b]	Allele/ Genotype	OR	95% CI.	Test	*P*
Fernandez-Reyes *et al.*,1997 (102)	Kenya	Case-control	444	SA	T vs. A	1.15	0.82 - 1.6	Pearson's χ2	0.47
-	-	-	390	CM	T vs. A	1.55	1.17 - 2.06	Pearson's χ2	0.0028
-	-	-	547	Severe	T vs. A	1.38	1.08 – 1.77	Pearson's χ2	0.01
Kun *et al.*,1999 (106)	Gabon	Case-control	200	Severe vs. Mild	T vs. A	0.54	0.34 - 0.84	Pearson's χ2	0.0092
Bellamy *et al.*,1998 (105)	Gambia	Case-control	616	SA	T vs. A	0.71	0.51 - 0.99	Pearson's χ2	0.054
-	-	-	802	CM	T vs. A	0.89	0.69 - 1.14	Pearson's χ2	0.37
-	-	-	474	SA vs. Mild	T vs. A	0.67	0.47 - 0.94	Pearson's χ2	0.027
-	-	-	660	CM vs. Mild	T vs. A	0.83	0.63 - 1.09	Pearson's χ2	0.19
Ohashi *et al.*,2001 (109)	Thailand	Case-control	-	Mild, non-cerebral severe & cerebral malaria	-	-	-	-	NS[c]
Amodu *et al.*,2005 (94)	Nigeria	Case-control	200	Severe vs. Asymptomatic Parasitaemia	T vs. A	1.71	1.09 - 2.67	Pearson's χ2	0.025[d]
-	-	-	199	Severe vs. Mild	T vs. A	1.5	0.97 - 2.34	Pearson's χ2	0.086
Jenkins *et al.*,2005 (108)	Kenya	Cohort	1417	Malaria-specific clinic visits	AT vs AA	0.89[e]	0.74 - 1.08	Poisson regression	0.24
-	-	-	886	Malaria-specific clinic visits	TT vs AA	0.91	0.68 - 1.21	Poisson regression	0.52
Ndiaye *et al.*,2005 (107)	Senegal	Cohort	878	Traits related to infection and carriage	-	-	-	-	NS
Ayodo *et al.*,2007 (95)	Kenya	Case-control	915	Severe	T vs. A	0.71	0.42 - 1.21	Pearson's χ2	0.1
Present study	Gambia, Kenya & Malawi	Case-control & Family	3920	SA	T vs. A	0.99	0.88 - 1.11	UNPHASED	0.85
-			5637	CM	T vs. A	1.07	0.97 - 1.17	UNPHASED	0.17
-			9395	Severe	T vs. A	1.02	0.96 - 1.09	UNPHASED	0.54

Abbreviations: CI, confidence interval; CM, cerebral malaria; OR, odds ratio; SA, severe malarial anaemia.

Using original data, were possible, we performed the allelic test (2x2, Pearson's χ^2) deriving OR, 95% CI and *P*-value.

[a] Sample size (n) relates to the number of case and controls used in calculating association statistic.

[b] Cases with this disease phenotype are compared with population controls, unless indicated otherwise.

[c] Low allele frequency of ICAM-1^{Kilifi} in Thailand (~2%) means very low power to detect association. NS = unavailable /unreported non-significant p-value.

[d] Multivariate logistic regression adjusted for age and parasite density reported p=0.277.

[e] Incident rate ratio adjusted for sickle trait, age, sex, season, and ethnic group.

```
                                                  *
Homo sapiens        QTSVSPSKVI LPRGGSVLVT CSTSCDQPKL LGIETPLPKK ELLLPGNNRK VYELSNVQED
Pan troglodytes     ......P... .......Q.. ........D. .......... ....G...W. ..........
Pan paniscus        ......P... .......Q.. ........D. .......... ....G...W. ..........
Gorilla gorilla     ......P... .......... ........T. .......... ....L...Q. ..........
Pongo pygmaeus      H....SAN.F .........N ........T. .......... ...PG...W. M.........
Macaca mulatta      ....F.PE.. .......K.N ..A.....IS ..M....... .I.PG...W. M.........
Mus musculus        .V.IH.REAF ..Q....Q.N ..S..KEDLS ..L..QWL.D .-.ES.P.W. LF...EIG..
```

Figure 3.2. ICAM-1 amino acid alignments. The first 60 residues of the mature ICAM-1 protein in man, common chimpanzee, pygmy chimpanzee, gorilla, orangutan, rhesus monkey and mouse (Genbank identifiers AF340038, AF340033, AF340042, AF340036, AF340041, AF340040 and NM_010493). The location of the K to M nonsynonymous SNP rs5491 or ICAM-1^{Kilifi} variant is indicated with an asterisk (position 29 of the mature protein).

Variation in the prevalence of ICAM-1 binding strains between geographic regions (and in one region over time), could explain some of the differences between studies. Substantial research has been undertaken investigating the functional differences between wild type ICAM-1 and ICAM-1^{Kilifi}. Laboratory isolates of *P. falciparum* adherent to ICAM-1 show differences in adhesive phenotype between reference ICAM-1 and ICAM-1^{Kilifi} (103, 240). Binding of the A4 parasite line to ICAM-1^{Kilifi} is reduced under both static and flow conditions, whereas binding of the ItG line (derived from the same Brazilian isolate as A4 but with repeated selection on ICAM-1) is far less affected by the K29M residue change. ICAM-1 mutagenesis experiments have suggest that even closely related parasite strains can use different contact residues for adherence on ICAM-1 (103). This raises the possibility that some wild African *P. falciparum* strains express PfEMP1 molecules adapted to preferentially bind ICAM-1^{Kilifi} (or which shows little difference between ICAM-1 alleles). This could explain the absence of a significant disease association, despite the pronounced functional differences seen with laboratory strains.

The geographic distribution of the nonsynonymous ICAM-1^{Kilifi} has raised the suspicion of a selective event involving this variant. The ICAM-1^{Kilifi} allele SNP is at a derived allele frequency of around 20-30% in many African populations [Gambia 19.6%, Kenya 28.4%, Malawi 29.6% (present study), Nigeria 31.6%, Gabon 27% (106), Senegal 17%(107)]. However, ICAM-1^{Kilifi} is also present at frequencies around 5% in East Asia [Papua New Guinea 4.9% (106), Thailand 1.7% (109), Han Chinese 7.8%, Japanese 5.7% (57)] and is uncommon but not absent from European-derived populations [Sweden 1.1% (241), North Carolina Caucasians 0.4% (242)]. Although Caucasian allele frequencies might represent

recent African admixture, the presence of ICAM-1^{Kilifi} in several East Asian and Pacific Rim populations suggests that the allele has been common in humans for some time (potentially tens of thousands of years), and may have become rare in European-derived populations due to drift, demography or even selection.

The *ICAM1* locus has been screened in a number of genome-wide surveys for recent evolutionary selection (with metrics including heterozgosity, excess of rare alleles, high frequency derived alleles, population differentiation (F_{ST}) and long-range haplotype tests) the locus has not been considered an outlier (52, 57, 243). There is evidence of selection at *ICAM1* over longer timescales, as comparative sequence analysis has suggested the gene has been under selection in the human-chimp lineage (244). It should be noted that residue 29 is not conserved in primate evolution (Figure 3.2). If we include the ICAM-1^{Kilifi} protein, the modern range of functional primate ICAM-1 molecules includes four diverse types of amino-acid side chain: positively charged (Lysine), negatively charged (Aspartic acid), polar uncharged (Threonine) and non-polar (Methionine). This may indicate that the residue, located in a coiled loop which projects from the surface of the N-terminal domain (245) is subject to relatively few functional constraints, but would also be consistent with multiple changes to evade pathogen binding.

The incidence of intermediate frequency nonsynonymous mutations is probably greater than may often be considered. To date over 16 thousand polymorphic nonsynonymous SNPs have been genotyped in a sample of Yoruba ethnicity individuals from Nigeria (YRI) as part of the International Haplotype Map (HapMap) project (release 22) (57). For

nonsynonymous SNPs, polymorphic in YRI, 14.7% (2338/15883) have a Wright's F_{ST} between YRI and the CEPH samples (collected by the Centre d' Etude du Polymorphisme Humain, in Utah from individuals of northwestern European origin) greater than 0.143 (the level of population differentiation seen at rs5491). Using published estimates of ancestral status [see online supplementary data associated with (52) (http://hg-wen.uchicago.edu/selection/haplotter.htm)] we identified the derived allele in a subset of nonsynonymous SNPs. Among nonsynonymous SNPs, polymorphic in YRI, 48.1% (4016/8344) have a derived allele frequency greater than 0.25 (the derived allele frequency of rs5491 in YRI). However, in the specific situation, where no derived alleles are found in the CEPH samples (as with rs5491), only 7.9% (86/1089) of nonsynonymous SNPs, polymorphic in YRI, have a derived allele frequency greater that 0.25.

If we accept the geographical distribution of the ICAM-1^{Kilifi} allele as evidence of an African selection event, particularly in light of the strong functional evidence – how could we tally selection with lack of a significant association with severe malaria phenotypes? An interesting explanation would be a frequency-dependant model; here an equilibrium exists between the human polymorphism frequency, and parasite strain frequency, with competing strains preferring either ICAM-1^{Kilifi} or ICAM-1^{Ref} binding. Change in host allele frequency is opposed by expansion in the relevant binding parasite strains, selecting against the expanded allele and returning the system to equilibrium. At close to equilibrium all individuals would have similar risk of life-threatening malaria irrespective of their ICAM-1 genotype. A simpler explanation is that ICAM-1^{Kilifi} is under selection through processes unrelated to severe malaria. The ICAM-1 residue 29 is situated in the region binding LFA-

1(246) and human rhinoviruses (247). ICAM-1^{Kilifi} has reduced avidity for LFA-1, abolishes binding to soluble fibrinogen(104) and prevents binding of some rhinovirus serotypes (248). A longitudinal study of Kenyan children has reported a significantly reduced incidence of non-malarial febrile illness among ICAM-1^{Kilifi} homozygotes (108).

Future research is likely to include increasingly large-scale epidemiological studies, based in developing countries, encompassing a broad range of infectious diseases. It will be interesting to see whether the report of an association between ICAM-1^{Kilifi} and non-malarial infections is replicated in these experiments. The trend in genetic studies of complex disease has been a move from candidate-based to whole genome approaches. The possibility of host and parasite allele frequency-dependent models highlights the need to extend future work to capture information about the genome of the parasite as well. In conclusion, our analysis has demonstrated that variation in *ICAM1*, specifically ICAM-1^{Kilifi}, does not have the substantial impact on malaria susceptibility reported in early studies. However, it is apparent that genetic variation in the *ICAM1* gene can modulate immune function, and may be associated with susceptibility to non-malarial infections. Therefore, ICAM-1 remains an important candidate molecule for future studies of human health in developing countries.

3.5 Materials and Methods

3.5.1 Subjects

Patient samples were collected during ongoing epidemiological studies of severe malaria at the Royal Victoria Hospital, Banjul, The Gambia; the Queen Elizabeth Central Hospital, Blantyre, Malawi; and Kilifi District Hospital, Kilifi, Kenya. *P. falciparum* malaria is endemic in these study sites, with most life-threatening disease occurring in children under the age of 5 years. All DNA samples were collected and genotyped following approval from the relevant research ethics committees and informed consent from participants. Controls were cord blood samples obtained from birth clinics in the same regions as the cases. Further demographics of the cases and controls including ethnicity, sex and age distribution have previously been published (249) (see Appendix A, Table A.2).

3.5.2 Phenotype definition

All cases were children admitted to hospital with evidence of *P. falciparum* on blood film and clinical features of severe malaria.(6, 250) We used a Blantyre coma score of ≤ 2 as a criterion of CM, and haemoglobin <5g per 100ml or packed cell volume <15% as a criterion of SA. Some individuals had both CM and SA. Of the severe malaria cases that were not CM or SA by these criteria, most had lesser degrees of coma (Blantyre coma score 3) or anaemia (Hb 5-6g per 100ml), or other complications such as respiratory distress. Our family-based study comprised:

i) 612 Gambian trios genotyped for all markers and an additional 708 Gambian trios used to genotype ICAM-1^{Kilifi} and rs5498 (in total the Gambian trios included 512 CM cases and 343 cases of SA).

ii) 225 Malawi trios (216 CM, 39 SA).

iii) 234 Kenyan trios (114 CM, 85 SA).

The population-based study (used to genotype ICAM-1^{Kilifi} and rs5498) comprised:

vii) 701 Gambian cases and 624 controls (324 cerebral malaria, 217 severe malarial anaemia).

viii) 718 Malawian cases and 405 controls (640 cerebral malaria, 101 severe malarial anaemia).

ix) 708 Kenyan cases and 902 controls (216 cerebral malaria, 270 severe malarial anaemia).

3.5.3 Power calculations

Power calculations were performed using the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/)(217). Across all case-control studies we would expect 95% power (based on our sample size, allelic odd ratio of 1.2, high risk allele frequency of 0.25 and a type I error rate of 0.05), a single regional case-control study (for example 701 cases and 624 controls) would have 54% power. For a similar effect size across the total family trio study (1779 trios) we would have 92% power.

3.5.4 SNP selection

Data from HapMap release 19 (October 2005, www.hapmap.org) was used to identify 19 SNPs within 5 kb of the *ICAM1* gene, polymorphic in the Yoruba ethnic group from Ibadan in South West Nigeria (YRI). Assays were developed for these and two further nonsynonymous SNPs: rs5498 which was not typed in HapMap release 19, but has a reported association with severe malaria (94), and rs1799969, which was typed by HapMap but found to be monomorphic in the Yoruba. Following initial testing two assays (rs5030399, rs281437) were rejected on technical grounds, and four further assays were rejected due to low allele frequencies (less than 2 %) in Gambians (rs5030364, rs5030400, rs5030384, rs1799969).

3.5.5 Sample preparation and genotyping

Genomic DNA samples underwent whole genome amplification through either primer extension pre-amplification (PEP)(251) or multiple displacement amplification (MDA)(252), before genotyping on a Sequenom MassArray genotyping platform (253). There were low rates of missing data and all population control genotypes/ untransmitted parental alleles were in Hardy-Weinberg equilibrium ($P > 0.01$) (Appendix B, Table B.2).

3.5.6 Statistical analysis

Family-based association analysis was performed using FBAT version 1.7.2 (254, 255). Case-control association analysis was performed by logistic regression using covariates of ethnic group, gender and Sickle status. DNA Sequenom genotyping for the Haemoglobin S (HbS) variant was performed for all samples as previously described (249). For logistic regression we utilized STATA (v9.2 for windows) and the genassoc package (http://www-gene.cimr.cam.ac.uk/clayton/software/) written for STATA by David Clayton. In both family- and population based studies we tested each marker in additive, dominant and recessive models, with three phenotypes (CM, SA and all severe cases) and tested for SNP-based associations with severe disease in each study region (Gambia, Malawi and Kenya). Pooling across all case-control and family-studies was performed using the UNPHASED application version 3.0 (http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/) (256), which employs a retrospective likelihood framework for performing genetic association analysis, and can be used to combine data from nuclear families and unrelated subjects. Ethnic origin was found to be a significant confounder and was retained as a covariate in the UNPHASED analysis. Haplotype structure around the *ICAM1* region, in each of our family datasets was examined using HAPLOVIEW version 3.32 (226) (http://www.broad.mit.edu/mpg/haploview/), haplotype blocks were defined using a solid spine of LD (as defined by D'> 0.75). Haplotype disease association analysis was conducted within the FBAT framework using family data.

3.6 Acknowledgments:

This work was funded by a Wellcome Trust Clinical Research Training Fellowship (to A.E.F.) and the UK Medical Research Council (to D.P.K.). This manuscript is published with the permission of the director of the Kenya Medical Research Institute. T.W. was funded by the Wellcome Trust, MalariaGEN and the European Union Network 6 BioMalPar Consortium.

Chapter 4

Positive selection of a *CD36* nonsense variant in sub-Saharan Africa, but no association with severe malaria phenotypes.

Andrew E. Fry[1,*], Anita Ghansa[1,2], Kerrin S. Small[1], Alejandro Palma[3], Sarah Auburn[1,4], Mahamadou Diakite[1], Angela Green[1], Susana Campino[1,4], Yik Y. Teo[1], Taane G. Clark[1,4], Anna E. Jeffreys[1], Jonathan Wilson[1,†], Muminatou Jallow[5], Fatou Sisay-Joof[5], Margaret Pinder[5], Michael J. Griffiths[1,6], Norbert Peshu[6], Thomas N. Williams[6,7], Charles R. Newton[6,8], Kevin Marsh[6,7], Malcolm E. Molyneux[9,10], Terrie E. Taylor[11,12], Kwadwo A. Koram[2], Abraham R. Oduro[13], William O. Rogers[14], Kirk A. Rockett[1], Pardis C. Sabeti[3,15], Dominic P. Kwiatkowski[1,4]

[1] The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, UK. [2] Noguchi Memorial Institute for Medical Research, P. O. Box LG581, Legon-Accra, Ghana. [3] Broad Institute, Cambridge, Massachusetts, 02142, USA. [4] Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK [5] Medical Research Council, PO Box 273, Banjul, The Gambia. [6] Kenya Medical Research Institute Centre for Geographical Medicine Research (Coast), P.O. Box 230, Kilifi, Kenya.[7] Nuffield Department of Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK. [8] Institute of Child Health, University College of London, London WC1N 1EH [9] Malawi–Liverpool–Wellcome Trust Programme of Clinical Tropical Research, PO Box 30096, Blantyre, Malawi. [10] Liverpool School of

Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK. [11] Blantyre Malaria Project, College of Medicine, PO Box 30096, Blantyre, Malawi. [12] Department of Internal Medicine, College of Osteopathic Medicine, Michigan State University, East Lansing, Michigan, 48824, USA. [13] Navrongo Health Research Centre, P.O. Box 114, Navrongo, Ghana. [14] Naval Medical Research Unit #2, Jakarta, Indonesia. [15] Center for Systems Biology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, 02138, USA.

4.1 Abstract

The prevalence of CD36 deficiency in East Asian and African populations suggests that the causal variants are under selection by severe malaria. Previous analysis of data from the International HapMap Project indicated that a *CD36* haplotype bearing a nonsense mutation (T1264G; rs3211938) had undergone recent positive selection in the Yoruba of Nigeria. To investigate the global distribution of this putative selection event we genotyped T1264G in 3420 individuals from 66 populations. We confirmed the high frequency of 1264G in the Yoruba (26%). However, the 1264G allele is less common in other African populations and absent from all non-African populations without recent African admixture. Using long-range linkage disequilibrium we studied two West African groups in depth. Evidence for recent positive selection at the locus was demonstrable in the Yoruba, although not in Gambians. We screened 70 variants from across *CD36* for an association with severe malaria phenotypes, employing a case-control study of 1350 subjects, and a family study of 1288 parent-offspring trios. No marker was significantly associated with severe malaria. We focused on T1264G, genotyping 10,922 samples from 4 African populations. The nonsense allele was not associated with severe malaria (pooled allelic odds ratio, 1.0; 95% confidence interval, 0.89 – 1.12; P=0.98). These results suggest a range of possible explanations including the existence of alternative selection pressures on *CD36*, co-evolution between host and parasite, or confounding caused by allelic heterogeneity of CD36 deficiency.

4.2 Introduction

Erythrocytes infected with mature forms of the *Plasmodium falciparum* parasite adhere to endothelium, platelets, leucocytes and uninfected erythrocytes, a behavior considered key to the pathogenesis of severe falciparum malaria (73). The majority of *P. falciparum* clinical isolates bind CD36 (73, 118, 233, 257). A family of variant molecules called *P. falciparum* erythrocyte membrane protein (PfEMP)1 is responsible for binding CD36 and other host antigens (258). PfEMP1 is expressed by the parasite onto the surface of infected red blood cells (iRBC) and is subject to switching during the course of an infection. CD36 is found on a range of cell types including platelets, dendritic cells and endothelium (83). Adhesion of iRBCs to endothelial CD36 helps the parasite avoid splenic passage, contributes to microcirculatory occlusion and promotes local inflammatory responses (73). CD36-mediated binding of iRBC to dendritic cells inhibits their maturation and function (259), and CD36 on platelets is required for the formation of platelet-mediated clumps, which are associated with severe disease (260).

CD36 deficiency (or Nak^a negative blood group) has been reported in Japanese (261), Thais (262) and African-Americans (89). The prevalence of CD36 deficiency in East Asian and African populations raises the possibility that the responsible variants have been selected for by malaria. The molecular basis of CD36 deficiency in African-Americans is distinct from that found in East Asia (90). Common alleles reported in East Asia include a proline to serine missense mutation at codon 90 of the *CD36* gene (C478T) and a frameshift mutation at codon 317 (1159insA)(263). The commonest reported African CD36 deficiency allele is a

nonsense mutation in exon 10 of *CD36* (T1264G; rs3211938), which terminates the polypeptide before the second transmembrane domain (90).

Further evidence that evolutionary selection has shaped genetic variation at the *CD36* locus emerged from Phase 1 of the International Haplotype Map (HapMap) project (57). The HapMap project performed whole genome high resolution genotyping in samples from four populations including 30 parent-offspring trios from the Yoruba ethnic group from Ibadan in Southwestern Nigeria, 30 parent-offspring trios from Utah (of Northern and Western European ancestry), 45 unrelated Han Chinese from Beijing, and 45 unrelated Japanese from Tokyo. The *CD36* nonsense allele, 1264G, was common in the Yoruba (~25%) but not detected in the other populations. The 1264G allele was present on haplotypes found to be unusually similar over hundreds of kilobases (see Appendix C, Figure C.1). This signal, termed extended haplotype homozygosity (EHH), suggested that 1264G had been under recent positive evolutionary selection (44, 51).

Genetic epidemiological studies of severe falciparum malaria have examined whether CD36 deficiency alleles affect susceptibility, but have not produced consistent answers (Table 4.1). An initial case-control study of 1359 Gambian and Kenyan children found CD36 deficiency appeared to be associated with susceptibility to severe disease, leading the authors to propose that another infectious pathogen was responsible for the frequency of CD36 deficiency in East Asians and Africans (90). In contrast, a matched case-control study of 693 pairs of Kenyan children found that 1264G heterozygotes were protected from malaria, particularly from having multiple syndromes of malaria (92). A third study, of 223 children

from Ibadan in Nigeria was unable to detect a significant difference in T1264G frequency between children with severe disease and those with asymptomatic parasitaemia (94). However, a recent study of 913 Kenyan children has, once again, suggested that 1264G is associated with susceptibility to malaria (95).

We set out to analyze in detail the relationship between *CD36* variation and severe malaria. Our aims were to:

(i) Determine the distribution of the putative T1264G selection event, in a range of African and non-African populations.

(ii) Replicate the *CD36* EHH signal, in the Yoruba and, if possible, additional populations.

(iii) Test single nucleotide polymorphisms (SNPs) across *CD36*, for associations with severe malaria. Previous genetic association studies have focused on T1264G and only a fraction of variation across the gene has been analysed. Other SNPs in or near *CD36* may demonstrate novel disease associations, and could assist the fine-mapping of functional changes.

(iv) Employ a series of population- and family-based studies of severe malaria phenotypes to refine the risk estimate associated with T1264G.

Table 4.1. Published risk estimates of severe malaria phenotypes associated with CD36 deficiency alleles.

Publication	Population	Cases	Controls	Phenotype	Allele/Genotype	OR	95% CI	*P*
Aitman *et. al.* 2000	Gambia & Kenya	598	761	All cases	def vs. wt	1.53	1.09 - 2.15	0.01**
(90)		(388)*	" "	CM	def vs. wt	1.49	1.01 - 2.20	0.04
		(97)	(331)	CM	GG vs. rest	-	-	0.04
Pain *et. al.* 2001	Kenya	693	693	All cases	GG & GT vs. TT	0·74	0·55–0·99	0·036**
(92)		(104)	" "	3 or 4 syndromes	GG & GT vs. TT	0·47	0·22–0·95	0·024
		(589)	" "	1 or 2 syndromes	GG & GT vs. TT	0·79	0·58–1·07	0·11
		(413)	" "	CM	GG & GT vs. TT	0·84	0·60–1·18	0·27
		(250)	" "	Respiratory distress	GG & GT vs. TT	0·67	0·43–1·04	0·051
		(304)	" "	SA	GG & GT vs. TT	0·62	0·41–0·94	0·017
		(146)	" "	Hypoglycaemia	GG & GT vs. TT	0·54	0·30–0·96	0·026
Amodu *et. al.* 2005	Nigeria	101	53	Uncomplicated vs. asymptomatic	G vs. T	-	-	0.062†
(94)		69	" "	Severe vs. asymptomatic	G vs. T	-	-	0.326
Ayodo *et. al.* 2007	Luo (Kenya)	456	457	All cases	GT vs. rest	1.5	1.03–2.18	0.03
(95)		" "	" "	All cases	GG vs. GT vs. TT	-	-	0.061
		" "	" "	All cases	G vs. T	-	-	0.0023‡
		" "	" "	All cases	G vs. T	-	-	0.00043§

P, *P*-values; OR, odds ratio; CI, confidence interval; def, CD36 deficiency alleles 1264G and 1439C pooled; wt, *CD36* wild-type allele; the remaining alleles and genotypes refer to *CD36* T1264G. CM, cerebral malaria; SA, severe anaemia; Severe, all severe cases; Uncomplicated, non-severe malaria

* Sample numbers in brackets refer to subsets of the studies' cases or controls.

**Mantel–Haenszel stratified analysis

† Multivariate analysis

‡ Weighted for population differentiation (Yoruba vs. Masai)

§ Weighted for population differentiation (Yoruba vs. Kikuyu)

4.3 Results

4.3.1 A global survey of *CD36* haplotypes and T1264G.

We genotyped the T1264G polymorphism in 3420 individuals from 66 ethnic groups. These included 974 individuals from the Human Genome Diversity Project (HGDP)- Centre d'Etude du Polymorphisme Humain (CEPH) Diversity Panel (51 ethnic groups), and 2446 additional Gambian, Malawian, Kenyan and Ghanaian cord blood samples and population controls (15 ethnic groups). The 1264G allele was at a frequency of 26% in the Yoruba of Nigeria, a finding consistent with the HapMap Yoruba (24.2%) and another recent study of Nigerian children (94). The 1264G allele was present in populations across sub-Saharan Africa, but is generally less frequent than in the Yoruba (Figure 4.1). The 1264G allele was not seen in two African populations living in regions of lower malaria risk, the South African Bantu and the Namibian San. However, their limited sample sizes may mean the allele is present at low frequencies but undetected. The pygmies of central Africa, such as the Biaka (Central African Republic) and the Mbuti (Democratic Republic of Congo), are relatively genetically isolated from other sub-Saharan populations. There is, however, evidence that the Biaka have experienced admixture with neighbouring groups (264). This is consistent with the presence of 1264G alleles in the Biaka (5.8%) but not the Mbuti (0%). The 1264G allele was absent from almost all non-African ethnic groups. The Makrani of Pakistan where the only group outside of Africa (and the only Pakistani ethnic group) to have 1264G (2.1%). The presence of 1264G is consistent with past African admixture in the Makrani (265). For a full breakdown of samples genotyped and 1264G allele frequencies see Appendix C, Table C.1.

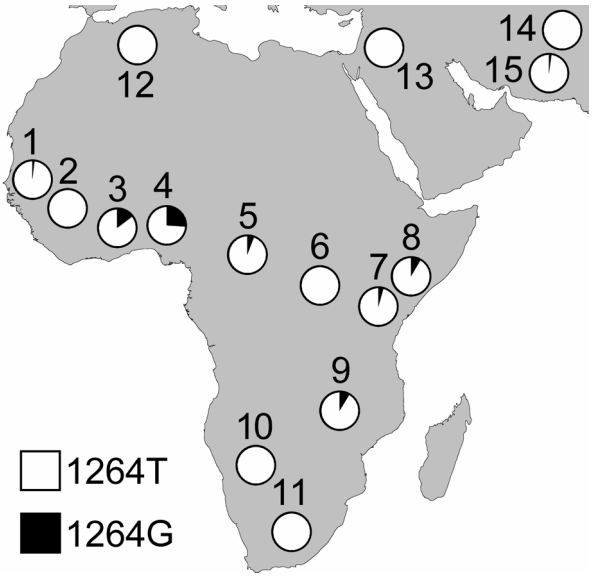


Figure 4.1. Frequencies of the *CD36* 1264G nonsense allele in a range of African and Middle Eastern populations. For each population, the country of origin, number of individuals sampled, and 1264G allele frequency are given (in parentheses) as follows: 1, Fula, Jola, Mandinka, Manjago, Serehuli, Serere and Wolloff (The Gambia, 582, 2%); 2, Mandenka (Senegal, 21, 0%); 3, Kasem, Nankan, and Buli (Ghana, 737, 14%); 4, Yoruba (Nigeria, 25, 26%); 5, Biaka Pygmy (Central African Republic, 26, 5.8%); 6, Mbuti Pygmy (Democratic Republic of Congo, 7, 0%); 7, Northeastern Bantu (Kenya, 11, 4.5%); 8, Chonyi, Duruma, Giriama and Kauma (Kenya, 722, 8.6%); 9, Malawians (Malawi, 405, 8.9%); 10, San (Namibia, 6, 0%); 11, Southeastern and Southwestern Bantu (South Africa, 7, 0%); 12, Mozabite (Mzab region of Algeria, 23, 0%); 13, Palestinian, Druze and Bedouin (Israel, 138, 0%); 14, Sindhi, Pathan, Kalash, Hazara, Burusho, Brahui and Balochi (Pakistan, 158, 0%); 15, Makrani (Pakistan, 24, 2.1%). The 1264G allele is absent from all other populations assayed but not shown in this figure.

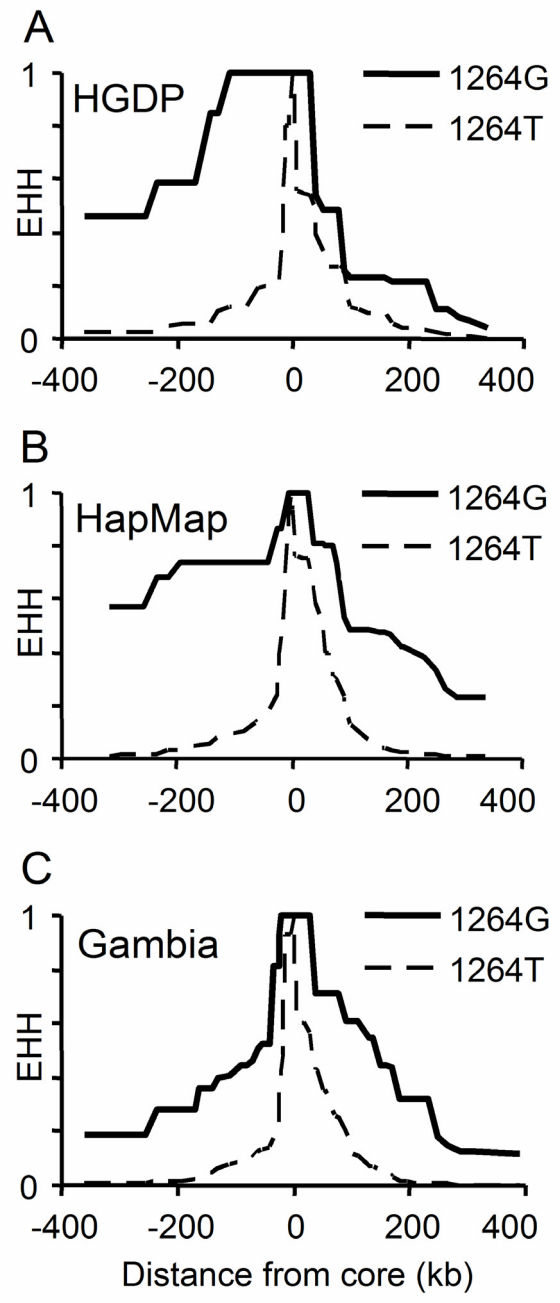


Figure 4.2. EHH decay surrounding T1264G in HGDP Yoruba, HapMap Yoruba and Gambian trios. Breakdown of extended haplotype homozygosity (EHH) with distance, on haplotypes partitioned by the alleles of T1264G (rs3211938; the 'core' SNP). We compared (A) 25 HGDP-CEPH Yoruba individuals, (B) HapMap genotypes from 60 Yoruba parents and (C) 202 Gambian parents.

We genotyped 53 additional SNPs, in and around *CD36*, in the HGDP-CEPH Diversity Panel. These markers comprised 12 intragenic SNPs (average density 1 SNP/6.3kb) and 41 SNPs extending 300kb up- and downstream of *CD36* (average density 1 SNP/15.3kb). We compared haplotypes carrying 1264G with those bearing the ancestral 1264T, calculating the EHH surrounding the core SNP (T1264G) (see Materials and Methods). EHH reflects the probability that two haplotypes, chosen at random from a population, will share identical markers, from the core SNP to a specified test position. Haplotypes carrying the 1264G allele were more similar over longer distances than those with the ancestral T-allele (Figure 4.2); a common trend when comparing ancestral and derived alleles (266). Measured ~100kb upstream from the core SNP (at rs1851937) the 1264G haplotype EHH was still 1.0 compared with 0.117 for 1264T, while at ~300kb upstream (rs304775) 1264G EHH was 0.462 vs. 0.026 for 1264T. Downstream, however, EHH values were lower, and there was less difference between the two haplotype subtypes, at ~100kb (rs12669309) EHH was 0.231 (G) vs. 0.12 (T), and 0.064 (G) vs. 0.006 (T) at ~300kb (rs2367090). To assess the significance of these differences we measured relative EHH (REHH, the ratio between EHH on haplotypes marked by 1264G and haplotypes bearing the T-allele) at a distance of around 0.25cM from the core SNP, and compared these measurements with the empirical genome-wide distribution of the long-range haplotype test (LRH) (57). The REHH was 17.6 (5') and 3.2 (3'). In other words, the EHH was 17-fold greater on haplotypes bearing 1264G at 0.25cM upstream. This 5' value is >99th centile compared with the empirical distribution of REHH and similar to (or greater than) previous reports of alleles under selection in Africans

[e.g. G6PD deficiency, haplotype frequency 18%, REHH 7 (51)]. However, the 3' REHH of 3.2 was not significant.

For comparison, we downloaded phased haplotypes from the HapMap Yoruba, paring down the dataset to the same 54 SNPs (including T1264G) (Figure 4.2). As expected, EHH measurements were greater on 1264G haplotypes. Upstream EHH was, at ~100kb, 0.741 (G) vs. 0.089 (T); at ~300kb, 0.571 (G) vs. 0.012 (T); downstream, at ~100kb, 0.483 (G) vs. 0.129 (T); and at ~300kb, 0.227 (G) vs. 0.005 (T). The HapMap Yoruba had REHH values of 29.7 (5') and 14.6 (3') [both >99th centile and consistent with the original observation (57)]. REHH values for our HGDP Yoruba are lower than for the HapMap Yoruba, particularly downstream of T1264G. There are a range of explanations for these differences: (i) the HapMap data had a slightly lower missing genotype rate (1% vs. 3% for HGDP); (ii) the HapMap haplotypes were inferred from higher density marker data; (iii) our HGDP sample size was smaller (50 haplotypes vs. 120 HapMap haplotypes); and (iv) the HapMap haplotypes were inferred from parent-offspring trios. Offspring genotypes can be used to unambiguously resolve phase at some parental markers. To test the impact of phasing accuracy on EHH we obtained Yoruba genotypes from the HapMap project and phased them without utilizing offspring data (treating the parents as 60 unrelated individuals). The resulting haplotypes had an REHH of 29.6 upstream, but only 5.1 downstream (see Appendix C, Figure C.2). This suggests that the asymmetry in HGDP Yoruba REHH, between 5' and 3' regions, may reflect technical artefacts affecting haplotypes estimation downstream (e.g. assay performance, phasing accuracy) rather than a biological signal (e.g. selection acting on sequences upstream of T1264G).

Too few 1264G haplotypes were sampled from the Makrani, the Biaka of the Central African Republic, or the Northeastern Bantu for a valid assessment of EHH to be made. To study 1264G haplotypes in detail, we selected 101 parent-offspring trios from our Gambian family study (see below). Twenty-six trios were selected because one or both parents had the 1264G allele. The remaining 75 were randomly chosen from the family study. Samples were genotyped and analysed in a manner similar to the HGDP-CEPH Diversity Panel. Offspring genotypes were used to assist phasing. In total, thirty 1264G haplotypes were compared with three hundred and seventy four 1264T haplotypes (Figure 4.2). EHH values on Gambian *CD36* 1264G haplotypes were raised. Upstream EHH was, at ~100kb, 0.405 (G) vs. 0.082 (T); at ~300kb, 0.184 (G) vs. 0.006 (T); downstream, at ~100kb, 0.607 (G) vs. 0.114 (T); and at ~300kb, 0.122 (G) vs. 0.003 (T). REHH was 19.1 upstream and 12.8 downstream. However, EHH is often raised around low frequency, derived alleles. The significance of these results was determined using phased genome-wide SNP data from 658 Gambian parent-offspring trios genotyped on an Illumina 650Y BeadArray platform (manuscript in preparation). The offspring in these trios were malaria cases; so to minimize bias we analysed only untransmitted parental chromosomes. We calculated the REHH for 15,807 derived alleles of frequency 1-3%, where REHH could be assessed at 0.25 ± 0.01cM from the core SNP. Comparison with this empirical, genome-wide distribution of the long-range haplotype test, suggest our Gambian REHH values are not significantly unusual, ~59th centile (12.8) and ~77th centile (19.1). This finding highlights that low frequency alleles are often surrounded by long haplotypes, limiting the power of EHH testing.

4.3.2 Genetic association studies with severe malaria phenotypes.

Seventy *CD36* SNPs, including T1264G and G1439C [a missense mutation in strong linkage disequilibrium in Gambians with the 1444delA frameshift mutation, another CD36 deficiency allele (90)] were tested for disease associations in Gambian case-control and family studies (see Material and Methods for further details of SNP selection). The family study comprised 1288 children affected by severe malaria and both parents [these trios included 508 cases of cerebral malaria (CM) and 333 cases of severe malarial anaemia (SA)]. The case-control study was separate from the family study and comprised 727 cases of severe malaria (339 CM, 226 SA) and 623 population controls. Both studies were powered to detect realistic effect sizes at a significance threshold of $P<0.05$. Only one marker, rs4728191, an intronic SNP, exceeded this threshold in both studies. The minor allele was associated with protection on both occasions (case-control, severe malaria, allelic odds ratio, (OR) 0.76; 95% CI, 0.61 – 0.96, $P=0.019$; family study, SA, allelic OR, 0.70; 95% CI, 0.5 – 0.99, $P=0.041$). This finding needs cautious interpretation as the associations were with different phenotypes and the Hardy-Weinberg *P*-value for rs4728191 in cord blood controls was marginal ($P=0.006$). All results from the case-control and family studies with $P<0.05$ are documented in Appendix C, Table C.2.

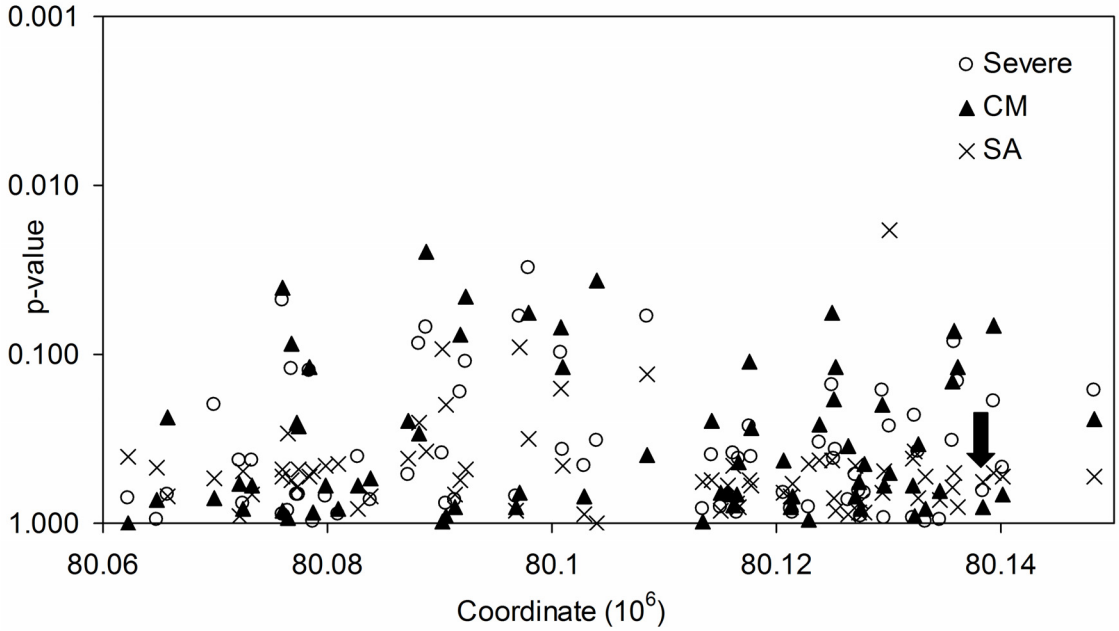


Figure 4.3. *P*-values for the pooled Gambian case-control and family studies. Allelic model *P*-values derived from UNPHASED analysis of 70 SNPs across *CD36*, for severe falciparum malaria (Severe), and the sub-phenotypes of cerebral malaria (CM) and severe malarial anaemia (SA). The arrow marks the position of T1264G (rs3211938). Chromosome 7 coordinates from NCBI build 36/dbSNP 126.

Using the UNPHASED application we combined the case-control and family data into a single analysis (Figure 4.3). Given that 70 SNPs have been tested, the formal significance threshold, with conservative Bonferroni correction, would be roughly 0.0007 (0.05/70). No marker in *CD36* had $P<0.002$ for any genetic model or phenotype. Rs4728191 was not significantly associated with severe malaria phenotypes in the combined analysis (severe malaria, allelic OR, 0.89; 95% CI, 0.79-1.01, $P=0.06$; SA allelic OR, 0.84; 95% CI, 0.67-1.06, $P=0.13$). Neither T1264G (severe malaria, allelic OR, 0.93; 95% CI 0.67-1.29, $P=0.65$), G1439C (severe malaria, allelic OR, 0.90; 95% CI, 0.68-1.19, $P=0.46$) or both deficiency alleles considered together (severe malaria, allelic OR, 0.91; 95% CI, 0.73-1.13, $P=0.39$) were significantly associated with severe malaria phenotypes. Full results from the pooled analysis are documented in Appendix C, Table C.3.

Given the previous report of protection associated with 1264G heterozygotes (92) we specifically checked for such an effect. Deficiency allele heterozygotes did tend to be underrepresented among cases (vs. wild type homozygotes, OR, 0.83; 95% CI, 0.66 - 1.05). In addition, no CD36 deficient individual (e.g. 1264G homozygote, 1439C homozygote or compound heterozygote) was found among the 727 cord blood controls, while several were seen among the cases (3 in the case-control study and 8 in the family study) (deficiency allele homozygotes vs. wild-type homozygotes OR, 1.59; 95% CI, 0.74 - 3.39). Together these results suggested a non-significant trend towards heterozygote advantage (overdominant model $P=0.09$). However, the low frequency of 1264G (1.9%) and 1439C (1.8%) in Gambians limited statistical power. To increase power we genotyped T1264G and

G1439C in 5 further sample sets derived from populations where 1264G is more common (8.2% in Kenyan controls, 9.0% in Malawian controls, and 14.2% in Ghanaian controls). These additional studies comprised:

i) 718 Malawian cases and 405 controls (640 CM, 101 SA).

ii) 708 Kenyan cases and 902 controls (216 CM, 270 SA).

iii) 225 Malawi trios (216 CM, 39 SA).

iv) 234 Kenyan trios (114 CM, 85 SA).

v) 792 Ghanaian cases and 806 controls (44 CM, 296 SA)

The 1439C allele was at a frequency of 0.1% in Ghanaian controls, but absent in Kenya and Malawi.

T1264G was not associated with severe malarial phenotypes in the additional 5 studies individually. Pooling data across all 7 studies (5681 cases and controls, and 1747 nuclear family trios) using the UNPHASED application we found no association between T1264G (severe malaria, allelic OR, 1.0; 95% CI, 0.89 – 1.12, P=0.98) or deficiency alleles (1264G and 1439C pooled, severe malaria, allelic OR, 0.98; 95% CI 0.89 - 1.09, P=0.76) and severe malaria (Table 4.2). The putative trend towards a heterozygous advantage, suggested by the Gambian data, was not supported by this larger data set, (deficiency allele heterozygotes vs. wild-type homozygotes, OR, 0.97; 95% CI, 0.86 - 1.09; deficiency allele homozygotes vs. wild-type homozygotes, OR, 1.09; 95% CI, 0.74 - 1.6; overdominant model P=0.53) (Table 4.3).

Table 4.2. Estimated risk for *CD36* T1264G in severe malaria.

Population	Cases	Controls	OR	95% CI	*P*
Population-based studies					
Gambia	727	623	0.74	0.4 - 1.37	0.34
Malawi	718	405	0.91	0.67 - 1.25	0.57
Kenya	708	902	1.20	0.93 - 1.55	0.15
Ghana	792	806	1.05	0.86 - 1.29	0.63
Family-based studies					
Gambia	1288	2576	0.92	0.59 - 1.45	0.73
Malawi	225	450	1.00	0.62 - 1.62	1.00
Kenya	234	468	0.86	0.55 - 1.34	0.50
Pooled	4692	6230	1.00	0.89 - 1.12	0.98

Allelic odds ratios (OR) and 95% confidence intervals (CI) for *CD36* 1264G compared with the T-allele, in severe malaria. *P*-values were derived from logistic regression analysis with covariates of ethnic group, gender and HbS genotype (population-based studies), case–pseudo-control analysis (family studies) or UNPHASED analysis (pooled data across all studies).

Table 4.3. Pooled analysis of CD36 deficiency alleles and severe malaria phenotypes.

Phenotype	Marker	Genotype	OR	95% CI	rec	over	dom	allelic
Severe	T1264G	GT vs. TT	1.00	0.88 - 1.14	-	-	-	-
	-	GG vs. TT	0.98	0.64 - 1.52	-	-	-	-
					0.94	0.99	0.99	0.98
	def	het vs. wt	0.97	0.86 - 1.09	-	-	-	-
	-	hom vs. wt	1.09	0.74 - 1.61	-	-	-	-
					0.62	0.53	0.64	0.76
SA	T1264G	GT vs. TT	1.05	0.86 - 1.28	-	-	-	-
	-	GG vs. TT	0.79	0.35 - 1.74				
					0.53	0.62	0.75	0.89
	def	het vs. wt	1.05	0.87 - 1.27	-	-	-	-
	-	hom vs. wt	0.89	0.43 - 1.84	-	-	-	-
					0.76	0.60	0.67	0.75
CM	T1264G	GT vs. TT	1.09	0.91 - 1.31	-	-	-	-
	-	GG vs. TT	0.78	0.37 - 1.64	-	-	-	-
					0.47	0.34	0.45	0.60
	def	het vs. wt	1.03	0.87 - 1.22	-	-	-	-
	-	hom vs. wt	1.03	0.55 - 1.93	-	-	-	-
					0.95	0.74	0.73	0.74

Pooled analysis performed from our 7 studies using the UNPHASED application; Phenotypes: Severe, severe falciparum malaria; SA, severe malarial anaemia ; CM, cerebral malaria; markers: T1264G or both CD36 deficiency (def) alleles pooled (G1439C+1444delA and T1264G); T1264G genotypes: GG, 1264G homozygotes; GT heterozygotes; TT, 1264T homozygotes; het, deficiency allele heterozygotes; hom, deficiency allele homozygotes; wt, wild-type homozygotes; OR, odds ratio; CI, confidence interval; *P*-values are reported for 4 genetic models rec(essive), over(dominant), dom(inant) and allelic.

4.4 Discussion

We set out to analyze the distribution of 1264G haplotypes in a range of global populations, to screen variation across the gene for association with severe malaria, and to target 1264G with well-powered tests of disease association. It is important to stress that the initial observation of raised EHH on 1264G haplotypes was only briefly noted in previous work, with little discussion of its biological significance (44, 57). This study represents the first detailed description of the *CD36* EHH signal found in the HapMap Yoruba. In addition, we presented novel data from the HGDP collection exploring the frequency of the nonsense allele across Africa, and replicating the EHH signal in the Yoruba. Our association study of severe malaria and T1264G employed a sample size more than twice that of all previous studies combined (see Table 4.1). Given our relative statistical power, the marginal significance of past results and their inconsistent outcomes, our data suggests that CD36 T1264G is not associated with severe malaria. Furthermore, no CD36 marker tested appeared to be significantly associated with disease susceptibility.

The reason for the expansion of the 1264G allele in West-Central Africa is unclear. Nigeria has a high prevalence of malaria infection, but so does The Gambia (1264G frequency 2%) and costal regions of Kenya (9%). The high frequency in Nigeria might indicate a recent origin in that region with subsequent migration. The lower allele frequency outside Nigeria could simply reflect selective sweeps in progress, but at earlier stages. There are a range of plausible explanations for the presence of selection in the absence of a malaria association. It is possible that the signal of selection detected around 1264G could be spurious; or that host-

parasite co-evolution has eliminated the advantage of this allele, subsequent to an initial selective sweep. Alternatively, the failure to detect an association might represent type II error. The 1264G allele could provide a modest degree of protection from severe malaria (e.g. an OR between 0.95 and 1); protect from an infrequent but life-threatening malaria phenotype; or only offer protection to the rare deficiency allele homozygotes. This association study would only have limited power to detect an effect in these situations. A small effect size might be sufficient to maintain the allele at low frequency in populations. Hypothetically, the expansion of 1264G alleles in the Yoruba could represent a local selection event on top of a low-level background of selection by malaria. The additional Yoruba event could also be malaria-related, for example the effects of a *P. falciparum* strain specific to Southwestern Nigeria. However, CD36 operates in a range of key biological processes including thrombostasis (267), glucose metabolism (268), lipid handling (269), immune function (270), angiogenesis (271) and possibly taste (272). Therefore other environmental factors need to be considered as selective pressures.

Combining evidence for natural selection with disease association statistics has been proposed as a technique to increase power to detect disease associations. Our data highlights a potential pitfall for this approach. In recently published work, an association between T1264G and severe malaria (P=0.03) was combined with the substantial allele frequency difference between the Yoruba (a population exposed to endemic malaria) and the Kikuyu and Masai (populations at low malaria risk). The combined P-values reported were highly significant (up to P=0.00043) (95). However, as we have seen, there is a substantial T1264G allele frequency difference between Yoruba and most African populations, including others

exposed to endemic malaria. This highlights the possibility that marginal evidence of a genetic association with disease A (e.g. severe malaria) could be conflated with strong evidence of natural selection caused by an unrelated process B (e.g. a novel dietary source). Combined analysis remains an exciting approach, but ensuring a signal of selection relates to the same disease as association data (or gauging the degree of confidence) will be a challenge.

It is possible that CD36 deficiency alleles cause deleterious consequences, particularly in the homozygous state, that balance any hypothetical advantages. These include neonatal immune thrombocytopenia (273), altered aerobic exercise capacity (274) and dysregulation of lipid metabolism (269). There is also, ironically, evidence that absence of CD36 is disadvantageous to the host during malaria infection. IRBCs are recognized and phagocytosed by monocytes and macrophages following CD36 binding (275, 276). One disease association study based in Thailand (475 adult patients with severe or mild malaria) reported an intronic dinucleotide repeat allele (in3$(TG)_{12}$) associated with protection from severe malaria (93). The authors showed that other alleles of the same microsatellite were associated with production of a short CD36 isoform altering the *P. falciparum*–binding epitope. In contrast, the protective in3$(TG)_{12}$ allele was associated with a full-length transcript, leading the authors to suggest that iRBC binding to intact CD36 could facilitate efficient clearance. The majority of *P. falciparum* isolates bind CD36 (73, 118, 233, 257) yet only a fraction of infections lead to life-threatening consequences. *In vitro* binding studies suggest that parasite strains from asymptomatic controls with parasitaemia have high levels of CD36 binding, while strains from individuals with non-severe, uncomplicated disease

have moderate levels of CD36 binding. The lowest levels of CD36 binding are found in isolates from individuals with severe phenotypes such as CM, and particularly SA (118, 233, 277, 278). CD36 (also known as Fatty Acid Translocase) is highly expressed in adipose tissue and skeletal muscle (83). Absence of CD36 as a target for sequestration on these large capillary beds may promote splenic passage and clearance of CD36 binding parasites. This could select for parasite clones which have switched to alternative host ligands. Post-mortem immunohistochemistry studies show relatively low expression of CD36 on brain endothelium without increased expression during malaria (73, 231). In contrast, other host ligands of PfEMP1, such as intercellular adhesion molecule-1 and E-selectin, are commonly expressed on cerebral endothelium, and induced by inflammation (231, 232).

CD36 deficiency alleles have been reported to modulate expression in a tissue specific way. Some variants prevent expression on platelets (CD36 deficiency type II) while other alleles abolish CD36 expression on a wider range of cells including platelets, macrophages, monocytes and probably other cell types (type I) (91, 279). The 1264G allele is considered to be a type I allele, but little is known about expression patterns (if any) on non-hematopoietic cells. In addition to T1264G and 1444delA a range of CD36 deficiency alleles have been reported in populations of African origin. These include 990delG and 1530ins14bp (90). Sequencing in Gambians has suggested that 1264G and 1444delA are the commonest deficiency alleles in this population, with others being less than 1% frequency (90). However, caution is required when extrapolating these findings across Africa. Allelic heterogeneity, particularly when unrecognized, has the potential to impair power in association analysis (280, 281) and could also be complicating our population genetic study.

The frequency distribution of CD36 deficiency across Africa could be quite different than suggested by typing just 1264G and 1439C+1444delA. Resequencing of *CD36* in a range of African and Asian populations is needed to describe the full repertoire and geographic distribution of CD36 deficiency alleles.

In conclusion, the relationship between CD36 and malaria is complex. The putative selection event associated with the *CD36* 1264G nonsense allele appears to be focused in West-Central Africa, although this and other CD36 deficiency alleles are present across sub-Saharan and Asian populations. A number of factors may explain the lack of an association between *CD36* variation and severe malaria phenotypes. However, the existence of alternative (or additional) selection pressures on CD36 deficiency alleles is an interesting possibility that needs further investigation. Additional field- and *in vitro* work is required to define the phenotypic consequences of CD36 deficiency alleles in health and disease.

4.5 Materials and Methods

4.5.1 Human Genome Diversity Panel

4.5.1.1 Sample preparation and genotyping.

DNA was derived from lymphoblastoid cell lines of 1064 individuals from the HGDP–CEPH Diversity Panel (http://www.cephb.fr/HGDP-CEPH-Panel/) (282). Samples from 15 known duplicate and atypical individuals were excluded (283). Genotyping was performed using Sequenom iPLEX assays. SNP identifiers, coordinates and genotyping success rates are documented in Appendix C, Table C.4. Following quality control for missing genotypes a set of 974 individuals were selected for further analysis (average genotyping success rate 96.7%; rs3211938 success rate 99.8%).

4.5.1.2 Statistical analysis.

Phased haplotypes and missing data were inferred using PHASE (version 2.1) (224, 225). Genotype data from each population was phased on its own. EHH calculations were performed using a web-based calculator (http://ihg2.helmholtz-muenchen.de/cgi-bin/mueller/webehh.pl). REHH was assessed as previously described (51, 284). REHH was assessed both sides of the core SNP (5' and 3'). The average REHH values for the two markers closest to and either side of 0.25cM were used for the LRH test [5', rs704871

(0.267cM) and rs2944398 (0.225cM); 3', rs10487878 (0.236cM) and rs4731861 (0.325cM)].

4.5.2 Genetic Association Studies

4.5.2.1 Phenotype definition.

All cases were children admitted to hospital with evidence of *P. falciparum* on blood film and clinical features of severe malaria (6, 250). We used a Blantyre coma score of ≤2 as a criterion of CM (due to the limitations of the available data we used ≤3 in Ghanaian cases), and <5g/dl or packed cell volume <15% as a criterion of SA. Some individuals had both CM and SA. Of the severe malaria cases that were not CM or SA by these criteria, most had lesser degrees of coma (Blantyre coma score 3) or anaemia (haemoglobin 5-6g/dl), or other complications such as respiratory distress. In The Gambia, Malawi and Kenya controls were cord blood samples obtained from birth clinics in the same locations as the cases. Ghanaian controls were community controls matched for age, ethnic group, and location of origin.

4.5.2.2 Subjects.

Patient samples were collected during ongoing epidemiological studies of severe malaria at the Royal Victoria Hospital, Banjul, The Gambia; the Queen Elizabeth Central Hospital, Blantyre, Malawi; Kilifi District Hospital, Kilifi, Kenya; and Navrongo War Memorial Hospital, Ghana. All DNA samples were collected and genotyped following approval from

the relevant research ethics committees and informed consent from participants. See Appendix C, Table C.5 for demographic details of patients and controls.

4.5.2.3 Power calculations.

Power calculations were performed using the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/) (217). For a type I error rate of 0.05 the Gambian case-control and family- studies had 85.7% and 98.7% power respectively. This is based on an allelic OR of 1.3 and a high risk allele frequency of 0.25 (the average minor allele frequency of the SNPs screened in the Gambians). With lower allele frequencies (0.1, 0.05 and 0.01) the case-control studies (each roughly 700 cases and 700 controls) had lower power, 59.1%, 36% and 11.4% respectively. The additional family studies (~250 trios) each had around 25.8%, 16.0% and 7.2% power. The power of the combined case-control study (2945 cases and 2736 controls) was 99.3%, 89.6% and 31.5%. The combined family-study (1747 trios) had 93.3%, 71.6% and 21.3% power.

4.5.2.4 SNP selection.

SNP data from HapMap release 21a (January 2007, http://www.hapmap.org) within 10kb of the *CD36* gene was download and assessed using HAPLOVIEW (version 3.32) (226). Polymorphic SNPs (minor allele frequency >5%) were identify in the HapMap Yoruba. Tagging SNPs were defined using the TAGGER algorithm (285). Sequenom iPLEX genotyping assays were designed for 69 tagging SNPs and three additional variants,

T1264G, G1439C and rs3173798 (located at a conserved splice site). Following quality control 70 SNPs were accepted for association analysis. This final set of SNPs tagged 116 of 133 (87%) polymorphic markers (>5% frequency) across *CD36* (including 10kb up- and downstream), using pairwise tagging and an r^2 greater than 0.8 (mean $r^2 = 0.97$), in the HapMap Yoruba.

4.5.2.5 Sample preparation and genotyping.

Genomic DNA samples used in the association analysis underwent whole genome amplification through either primer extension pre-amplification (251) or multiple displacement amplification (252), before Sequenom iPLEX genotyping. Genotype data underwent quality control for missing data (<10%), Hardy-Weinberg equilibrium (p>0.001) and Mendelian error rate. Genotype counts and quality control measures are documented in Appendix C, Table C.6.

4.5.2.6 Statistical analysis.

Case-control association analysis was performed by logistic regression using covariates of ethnic group, gender and Haemoglobin S (HbS) status. Sequenom genotyping for the HbS variant was performed for all samples as previously described (249). Family-based association analysis was performed using a case-pseudo-control approach and conditional logistic regression based on parental genotypes. Trios were drawn from a larger pool of samples assessed for relationship misspecification by Mendelian error rates. We utilized the

R statistical application (version 2.6.0) along with the dgc.genetics library (http://www-gene.cimr.cam.ac.uk/clayton/software/) and the SNPassoc package (286). In both family- and population-based studies we tested each marker in allelic (or 'multiplicative'), dominant, recessive and overdominant (or 'heterozygote advantage') genetic models, and with three phenotypes (CM, SA and all severe cases). Pooling across case-control and family studies was performed using the UNPHASED application (version 3.0.12) (http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/) (256). Ethnic origin was a significant confounder and was retained as a covariate in the UNPHASED analysis.

4.6 Acknowledgments:

This work was supported by the Wellcome Trust [GR074586AIA], through a Clinical Research Training Fellowship to A.E.F. and UK Medical Research Council funding (to D.P.K.). T.W. was funded by the Wellcome Trust, MalariaGEN and the European Union Network 6 BioMalPar Consortium. The work of the Ghanaian study was supported by funding from the National Institute of Allergy and Infectious Diseases/ National Institutes of Health [contract number AI95363] to the Noguchi Memorial Institute for Medical Research. We would like to thank the MalariaGEN consortium for providing access to the Illumina 650Y Gambian parent-offspring trio data. This manuscript is published with the permission of the director of the Kenya Medical Research Institute. The views expressed here are the personal ones of the authors and do not purport to represent those of the United States Department of the Navy.

Chapter 5

Haplotype homozygosity and derived alleles in the human genome.

Andrew E. Fry[1], Clare J. Trafford[1], Martin A. Kimber[1,2], Man-Suen Chan[1], Kirk A. Rockett[1], Dominic P. Kwiatkowski[1]

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom ; and [2] Tessella Support Services plc, Abingdon, United Kingdom

5.1 Abstract

Haplotype-based techniques are being used to estimate the relative age of alleles - particularly in screening loci for signals of recent positive selection - but does this approach capture even coarse age differences? Using simulations and empirical data from the International HapMap Project, we show that a simple pairwise metric of haplotype homozygosity gives significantly higher mean values for human single-nucleotide-polymorphism alleles that appear to be derived, than for those that appear to be ancestral, as determined by comparison with the chimpanzee genome. Our results support the use of haplotype-based techniques, such as extended haplotypic homozygosity, to assess the age of alleles.

5.2 Introduction

Gauging the age of alleles is a critical task in identifying SNPs that have been subject to recent positive selection and in understanding our evolutionary history. A new allele that arises by mutation will lie on a single haplotype, but, over time, the extensive linkage disequilibrium (LD) between the new allele and other markers on this ancestral haplotype breaks down as a result of recombination. Decay of LD around a target allele is considered to be a stopwatch by which its age can be estimated (287).

Diverse metrics are employed to measure the LD between markers (288); however, common ones, such as D' and r^2, do not differentiate between the alleles at a single locus. Haplotypic homozygosity (the probability of selecting two identical haplotypes at random from a population) is a measure of LD that has the ability to capture information about subgroups, or 'partitions', of haplotypes in the population marked by a specific allele (289). Furthermore, haplotypic homozygosity has recently become the basis of a strategy to detect loci that have undergone partial selective sweeps, by detecting 'young' core haplotypes or alleles (as judged by the decay of haplotypic homozygosity) that have reached high frequency (51). Extended haplotypic homozygosity (where the partitions of haplotypes are marked by a core set of markers) has been applied to detect selection at loci such as glucose-6-phosphate dehydrogenase (51), the CD40 ligand gene (51), the lactase gene (290), the spinocerebellar ataxia type 2 gene (291), the CCR5-Δ32 mutation, (284) and the haemoglobin E variant (56). A related approach, the haplosimilarity score, has been suggested for screening regional data sets and was applied to the haemoglobin S variant

(53). The recent publication by the International HapMap Consortium identified a number of outliers in the genomewide distribution of haplotype homozygosity, in the form of the long-range haplotype test statistic, as candidates for recent selective events (292).

Haplotype homozygosity metrics are being used in the literature to judge the relative age of alleles, but how sensitive is this approach? Can we test whether even coarse age differences are detected? We employed the chimpanzee genome as an extant out-group for humans, to split alleles into putative ancestral (very old) or derived (much younger) alleles. Then, we investigated whether a simple metric of haplotype homozygosity could be applied at the level of the whole genome to distinguish this broad dichotomy of alleles, which are likely to have very different ages.

First, we employ simulated SNP data to demonstrate the theoretical differences in haplotype homozygosity between ancestral and derived alleles. Second, we repeated this experiment, using empirical data from (i) the HapMap project (292), to examine genome-wide trends, and (ii) the ENCODE (ENCyclopedia Of DNA Elements) project, to compensate for ascertainment bias. Third, we examined the range of haplotype homozygosity values around ancestral and derived alleles at the same SNPs. Finally, we investigated "private" alleles (see below), as a subset of alleles we expect to contain some particularly young alleles. Our results support the use of haplotype-based techniques such as extended haplotypic homozygosity (51) or haplosimilarity score (53), to assess the age of alleles.

5.3 Methods

Given that the inference of extended haplotypes from genotype data is computationally intensive and intrinsically error prone, we opted for a simple pairwise metric of haplotypic homozygosity to rapidly investigate genome-wide trends. Consider two biallelic SNPs denoted A and B, where we are particularly interested in allele X of SNP A. We define the following metric:

$$H_X = h_{B:X} - h_B \ ,$$

where $h_{B:X}$ is the homozygosity observed at SNP B when we consider only haplotypes carrying allele X of SNP A, and h_B is the homozygosity observed at SNP B when we consider all haplotypes (Figure 5.1). We are particularly interested in situations where the homozygosity of the partitioned haplotypes is unusually high or low compared with the result expected from the general population. Therefore, we calculated a calibrated partition homozygosity (H_x) by subtracting the general population homozygosity at SNP B (h_B) from $h_{B:X}$.

A positive value of H_X implies that the homozygosity at locus B on haplotypes marked by allele X of SNP A is greater than would be expected, given the population allele frequency of SNP B. An important feature of this metric is that each allele of a single SNP is likely to receive a different H_x value, corresponding to the history of that allele.

Figure 1

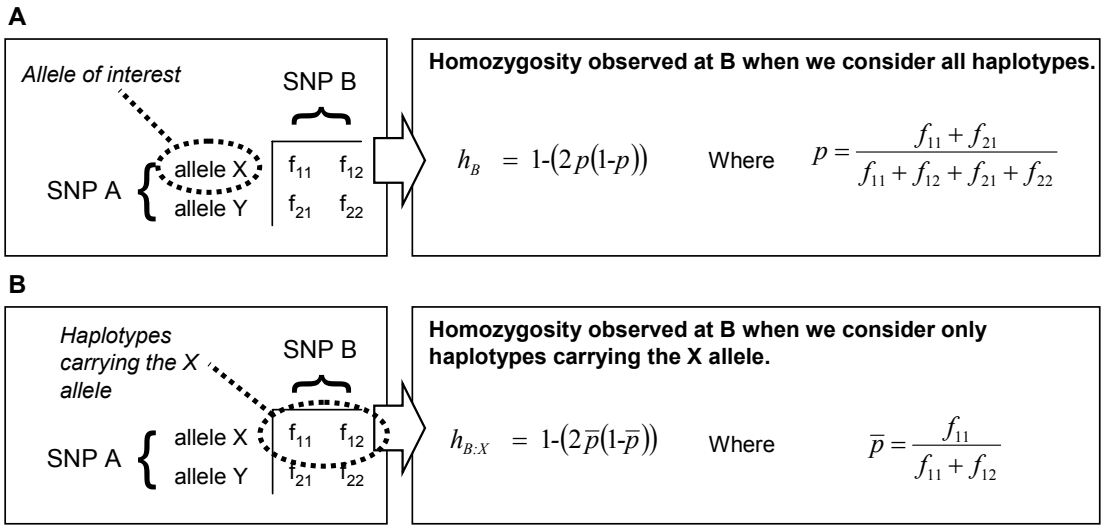


Figure 5.1. Calculation of the H_x metric. We define $H_X = h_{B:X} - h_B$, where $h_{B:X}$ and h_B are calculated as follows. A, Construct the classic 2x2 contingency table for marker combinations between SNPs A and B. Calculate h_B, the general population homozygosity at SNP B, as determined by Hardy-Weinberg equilibrium. B, Calculate $h_{B:X}$, the homozygosity at SNP B on the row, or partition, of haplotypes possessing allele X of SNP A.

To assess regional haplotypic homozygosity on haplotypes bearing the X allele of SNP A, we took the simple approach of averaging a target allele's H_x values for a window of nearby comparison SNPs. SNPs found to be monomorphic in a given population were ignored. The choice of window size was based on preliminary work that demonstrated a decay of H_x with increasing window size (because of greater average distance between target and comparison SNPs). A window size of 50kb centered on the target allele contains sufficient markers (~20 SNPs) to demonstrate our signal clearly; however alternative window sizes could be employed.

Conventional pairwise measures of LD are closely related to the H_x metric. Given the familiar 2x2 table (Figure 5.1) representing the relationship between two markers and containing the observed pairwise haplotype frequencies –and with the assumption that we regard the population allele frequencies as fixed – there is only 1 degree of freedom. We can express the relationships between H_x and the other metrics with the following equations:

$$r^2 = \frac{\left(f_A . H_A + f_a . H_a\right)}{\left(2 \times f_B \times f_b\right)}$$

and

$$D = \frac{\left(f_A^{\,2} \times H_A - f_a^{\,2} \times H_a\right)}{\left(2 \times \left(f_B - f_b\right)\right)} .$$

Here, f_A, f_a, f_B, f_b are the major- and minor-allele frequencies of SNPs A and B, whereas H_A and H_a are the partition haplotype homozygosities (H_x values) for the A and a alleles of SNP A, respectively. If $f_B = f_b$ (i.e., if SNP B has minor-allele frequency 0.5) D is simply $\left(f_{11} - f_{12}\right)/2$, which is of magnitude of $\sqrt{\left(r^2 \times f_a \times f_A\right)}/2$, but the sign is arbitrary

because, in this case, the assignment of one allele as major and the other as minor is meaningless. To consider this in biological terms, r^2 is the sum of the (scaled) homozygosities of two alleles – the histories of both alleles contribute to the overall correlation of the SNP – whereas D is related to the appropriately weighted difference between the homozygosities of two alleles.

5.4 Results

To explore the properties of the H_x metric in a system for which the properties of the data are clearly known, we performed coalescent simulations of a 50-kb of genomic sequence, using SelSim (293), version 2.1. The central SNP was set to attain specific allele frequencies (0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, and 0.95), and the target allele was modeled (i) as a neutral site and (ii) as an allele undergoing positive selection with different selection coefficients. The effective population size was N_e = 10,000, with one SNP every 300bp and a uniform population-scaled recombination rate of ρ = 0.4/kb (~1cM/Mb). At the end of each simulation, 120 chromosomes were sampled from the population. Mean H_x values and standard error of the mean were calculated for the region. One thousand simulations were run for each particular setting. Regional H_x and standard error of the mean values were averaged across the 1,000 runs for a given allele frequency (Figure 5.2).

Figure 2

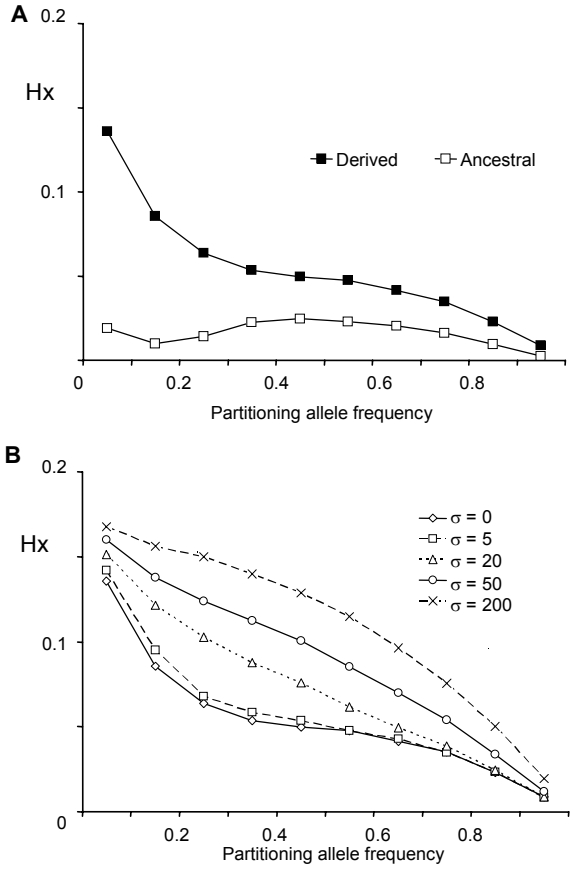


Figure 5.2. Comparison of mean regional H_x values for ancestral and derived alleles for simulated data sets, with the central target allele set to attain a range of frequencies. A, Ancestral and derived alleles modeled as neutral sites. B, Derived alleles under positive selection with different population-scaled selection coefficients (σ or 2Ns, where 2N is the population size and s is the selection coefficient). Each simulation was run 1,000 times for a given allele frequency. Standard errors of the mean were always <0.00045 – well within the symbols shown. As expected, H_x values decay with increasing allele frequency. Haplotype homozygosity was preserved on haplotypes marked by derived alleles and around alleles that have undergone rapid positive selection.

The allele frequency of a derived allele has a direct relationship with age (294), and, as predicted from theory, the simulations demonstrate haplotypic homozygosity decaying with allele frequency. The simulated data suggests that, throughout the frequency range, derived alleles tend to have higher mean regional H_x values than ancestral alleles. As expected, simulated alleles undergoing selective sweeps due to positive selection reach their target frequencies earlier and therefore retain greater regional haplotypic homozygosity.

We analysed empirical genome-wide trends in haplotypic homozygosity for ancestral and derived alleles. SNP genotypes were downloaded from HapMap release 16c.1 (phase 1, June 2005), along with ~18,000 SNPs from the ENCODE regions. Pairwise relationships between nearby SNPs were determined from genotypes of paired markers using an expectation-maximization algorithm. We determined the putative ancestral state for 964,842 HapMap SNPs by comparison with the *Pan troglodytes* genome (CHIMP1, November 2003). The implicit assumption with this approach is that, for the vast majority of loci, where the *Homo sapiens* genome is presently polymorphic, the *Pan troglodytes* genome remains in the state of our most recent common ancestor. This is a practical approximation with an estimated error rate for typical SNPs of ~0.5% (outside a CpG context) and an overall rate of ~1.6% (295).

Figure 3

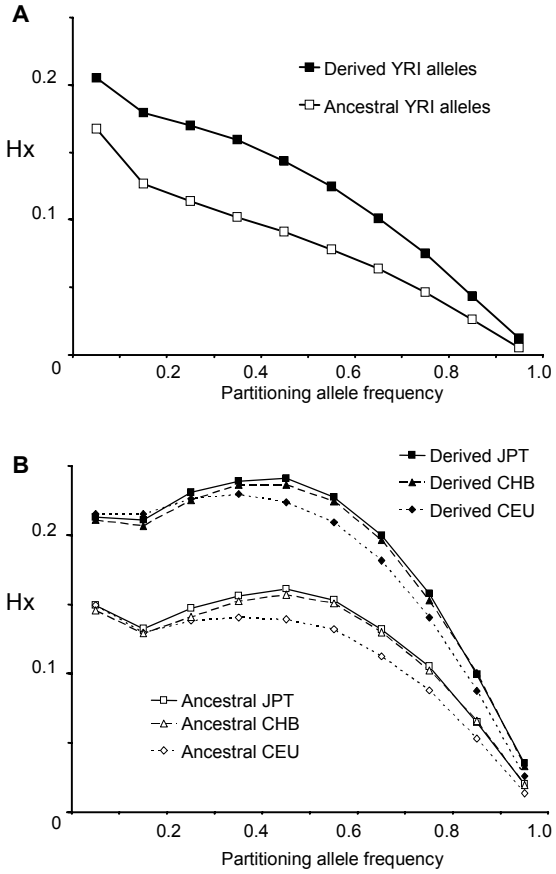


Figure 5.3. Comparison of mean regional H_x values for ancestral and derived alleles for HapMap SNPs, binned by partitioning allele frequency – that is, the frequency of target SNP A allele X. A, Data from 1,379,358 YRI alleles. B, Asian and European populations. Data from 1,194,358 JPT alleles, 1,307,740 CEU alleles, and 1,119,766 CHB alleles are plotted. Standard errors of the mean were between 0.00002 and 0.0004 – well within the symbols shown. Haplotypes marked by derived alleles had, on average, higher regional haplotypic homozygosity throughout the frequency spectrum and in all populations.

Regional H_x values for each HapMap allele were binned by target (i.e., SNP A) allele frequency. Then, the mean values for these bins plotted separately for ancestral and derived alleles. This was performed for all four HapMap populations (Figure 5.3): Yoruba from Ibadan, Nigeria (YRI); Japanese from Tokyo (JPT), Chinese Han from Beijing (CHB); and CEPH individuals from Utah (with northern and western European ancestry) (CEU). We found that haplotypes marked by derived alleles generally had higher regional haplotypic homozygosity throughout the frequency spectrum and in all populations. Alleles from the YRI had lower regional haplotypic homozygosity values than the Asian and European populations (CEU, CHB, and JPT), which all follow very similar patterns. The downward trend of haplotypic homozygosity with increasing allele frequency was confirmed – although the JPT, CHB, and CEU populations follow an apparently parabolic distribution. This analysis highlights the very different haplotypic structures between the YRI and the other three HapMap populations and reinforces the view that, because of their different demographic histories, the YRI have a relatively low genome-wide LD, compared with Asian and European populations (296).

It is well established that the ascertainment strategy employed in the HapMap project (resequencing of a small SNP discovery panel, followed by targeted genotyping in all samples) has led to a bias against the genotyping of rare alleles. Regional statistical attributes that depend on the spectrum of allele frequencies, such as nucleotide diversity, Tajima's D, F_{ST}, and LD (including H_x), will be affected by this ascertainment bias (297). The density of HapMap phase 1 data means that the sampling window around a partitioning SNP contains only a proportion of the alleles actually present; however, the sampled alleles

are skewed towards intermediate and high frequencies. To empirically determine whether the trends in H_x difference between ancestral and derived alleles would remain after correction for ascertainment bias, we used the ENCODE data. The ENCODE regions comprise ten 500-kb regions fully resequenced in 16 CEU, 16 YRI, 8 CHB, and 8 JPT samples. The polymorphisms found were genotyped in all 270 HapMap samples. Thus, the ENCODE data should be free of ascertainment bias. Regional H_x values for alleles in ENCODE regions were calculated and then binned by SNP A allele frequency. The mean values for these bins plotted separately for ancestral and derived alleles. This was performed for all four HapMap populations (Figure 5.4). Using the ENCODE data, we note a reduction in average H_x values; however, the general trends remained. Haplotypes marked by derived alleles generally have higher regional haplotypic homozygosity throughout the frequency spectrum and in all populations than did the ancestral alleles. Alleles from the YRI had lower regional haplotypic homozygosity values than the Asian and European populations, and there was still the decay of haplotypic homozygosity with increasing allele frequency. The lower H_x values calculated from ENCODE data, particularly YRI data, were closer to our simulations of H_x than those from Phase 1 HapMap data. The remaining differences between the empirical and simulated data are likely to be due to the simplicity of our simulated system with uniform recombination, evenly spaced SNPs, and lack of demographic events.

Figure 4

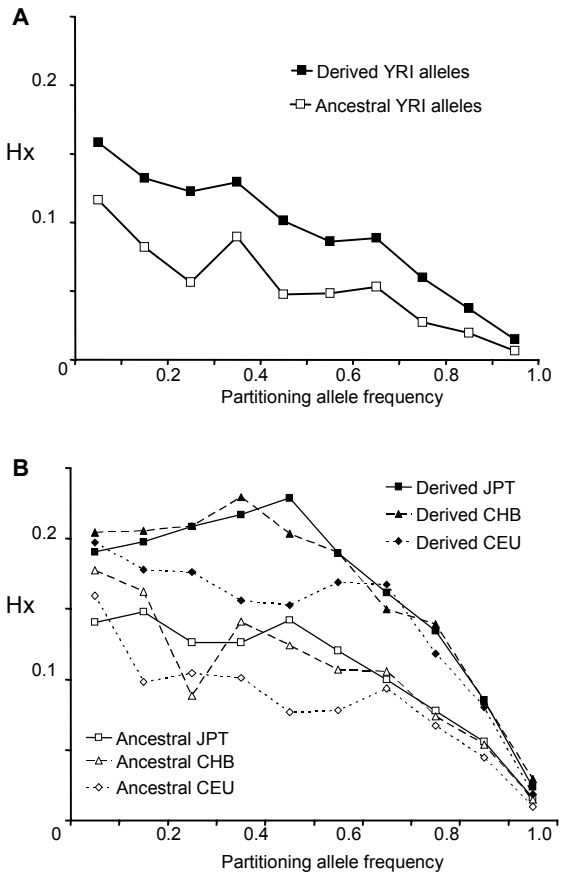


Figure 5.4. Comparison of mean regional H_x values for ancestral and derived alleles of SNPs in the ENCODE regions, binned by partitioning allele frequency – that is, the frequency of target SNP A allele X. A, Data from 15,614 YRI alleles. B, Asian and European populations. Data from 12,084 JPT alleles, 15,166 CEU alleles, and 11,042 CHB alleles are plotted. Standard errors of the mean were between 0.000046 and 0.0018 – well within the symbols shown. Even with complete SNP ascertainment, haplotypes marked by derived alleles had, on average, higher regional haplotypic homozygosity throughout the frequency spectrum and in all populations.

To compare regional H_x values for ancestral and derived alleles at the same loci, we additionally analysed SNPs with allele frequency of exactly 0.5, thus controlling for variation of H_x with allele frequency. Histograms of H_x distribution for ancestral and derived alleles from the CEU and YRI were plotted (Figure 5.5). Using a paired *t* test (two-tailed), we found a statistically significant difference between the regional H_x values for ancestral and derived alleles for both the YRI ($p = 1x10^{-192}$) and CEU ($p < 1x10^{-200}$). A nonancestral allele at frequency 0.5 has a higher regional haplotypic homozygosity value than its partner ancestral allele at the same SNP in ~65.8% of YRI SNPs and 71.5% of CEU SNPs.

The majority of HapMap phase 1 SNPs (~650,000) are polymorphic in all four populations, which suggests that, in general, the original mutations occurred some time before the groups separated. In contrast, 'private' alleles (alleles found in only one population) should include a subset of particularly recent mutations. We investigated YRI private alleles (alleles polymorphic in YRI but not in any other HapMap population) and CEU alleles not seen in the YRI. The CEU, CHB, and JPT share many of the alleles not seen in YRI, so rather than studying the relatively small number of completely private alleles in the CEU, we investigated only those CEU alleles that were private in relation to YRI alone.

Figure 5

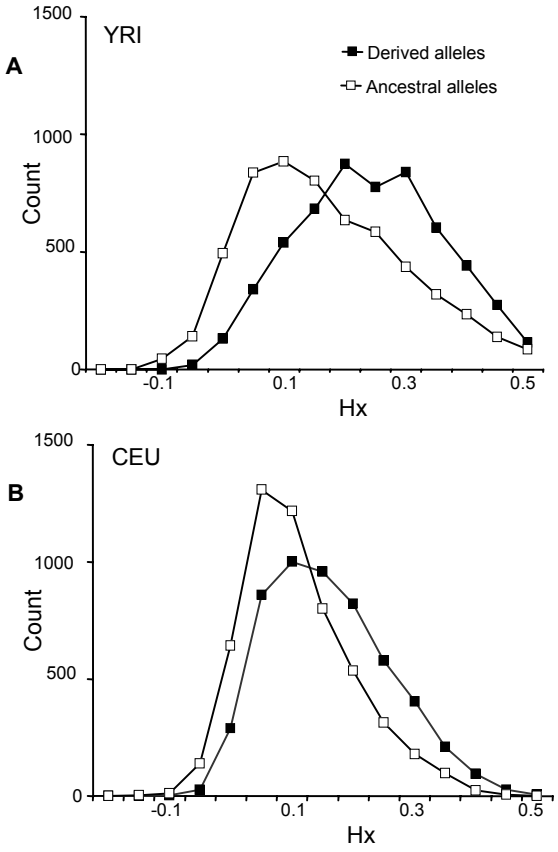


Figure 5.5. Histograms demonstrating the range of H_x values for ancestral and derived alleles with frequency of 0.5. A, Data from 5276 SNPs of the YRI population. B, Data from 5630 SNPs of the CEU population. Although both ancestral and nonancestral alleles can have a wide distribution of values, nonancestral alleles generally have higher regional H_x scores.

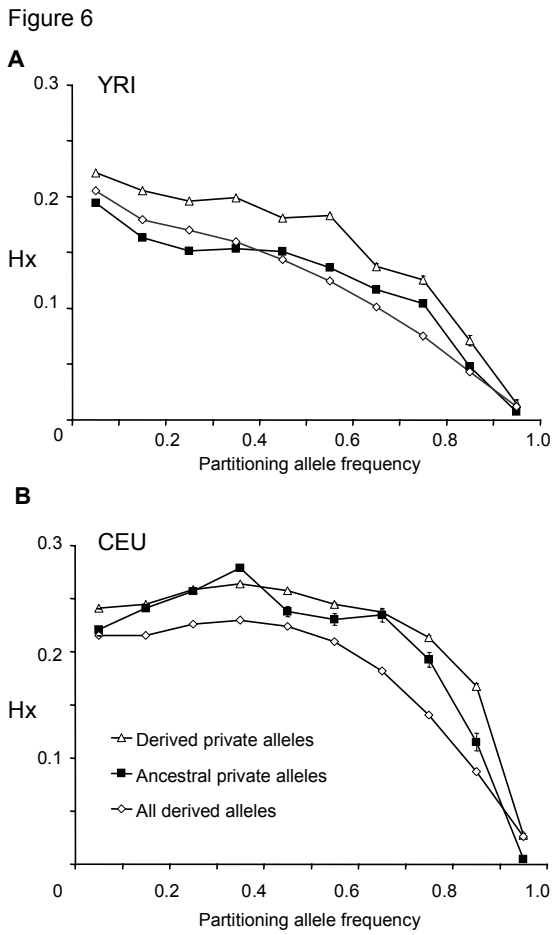


Figure 5.6. Comparison of mean regional H_x values for private derived alleles, private ancestral alleles, and all derived alleles, binned by partitioning allele frequency. A, Data from 62,207 YRI private derived alleles, 15,349 YRI private ancestral alleles, and 689,679 general derived alleles. B, Data from CEU alleles (as representative of the three Asian and European populations) that were private in relation to the YRI, including 41,312 CEU private derived alleles and 3,127 CEU private ancestral alleles, and from 653,870 general derived alleles. Standard errors of the mean are plotted. Partitions marked by private derived alleles had higher regional haplotypic homozygosity than either general derived alleles or private ancestral alleles, in the majority of partitioning allele frequency bins.

Many private alleles actually represent older mutations that occurred in the progenitor population before it split and, because of either lack of transmission or subsequent fixation, have become monomorphic in all but one of the populations. Therefore, we compared private derived alleles with private ancestral alleles. We expected private derived alleles to be enriched for young mutations specific to one population, whereas private ancestral alleles are likely to represent alleles that were present in the precursor population; however, the derived allele has become fixed in the comparison population(s). The plotted data (Figure 5.6) shows that both types of private alleles were associated with elevated regional haplotype homozygosity values in the CEU and YRI. However, in general, private derived alleles scored higher than nonprivate alleles and private ancestral alleles (particularly in the YRI). This is consistent with our expectation that the private derived alleles are enriched for a subset of particularly recent mutations.

5.5 Discussion

In these experiments, partitions of haplotypes marked by ancestral alleles gain haplotypic homozygosity as they fall in frequency. Demographic events such as founder effect, population bottlenecks, or admixture are likely to generate LD on haplotypes bearing ancestral and derived alleles. This is the most probable explanation of the high H_x values in the Asian and European populations. In our simulated datasets for which demographic events are absent, the ancestral allele can still pick up some haplotypic homozygosity as it drops in frequency; one possibility is that it picks up some LD by recombination with the homogenous haplotypes bearing the derived allele. Alternatively, as a partition of haplotypes becomes rarer, LD is generated relatively faster by genetic drift than it is removed by recombination (298). The existence of a minority of SNPs were the ancestral allele is misidentified could explain part of this trend, particularly the high H_x values seen among private ancestral alleles. In the future, a range of primate genomes could be used to predict ancestral alleles with greater accuracy. A final consideration is the effect of measuring haplotype homozygosity in a finite sample size. This leads to exaggeration of the H_x metric in rare partitions (e.g., a partition represented by 1 chromosome of 120 will automatically have maximum haplotype homozygosity). This is an issue in the first and second partitioning allele-frequency bins (a partition size of <10-20 chromosomes of 120), particularly when the population frequency of the comparison SNP (SNP B; see Figure 5.1) is close to 0.5, but rapidly diminishes thereafter.

To conclude, techniques such as haplotypic homozygosity are being used to estimate the date of origin of genetic variants thought to have undergone recent selection. The ability to reliably differentiate old and young alleles is critical to this strategy. Although other, more sophisticated haplotype-based methods may exist, we have shown that a simple metric of haplotype homozygosity that can rapidly be applied to the whole genome can distinguish between alleles of broadly different ages.

5.6 Acknowledgements

This work was funded by a Wellcome Trust Clinical Research Training Fellowship (to A.E.F) and the Medical Research Council.

Chapter 6

General Discussion

This project began with the hypothesis that genetic variation in the host ligands of PfEMP1 modulates susceptibility to severe malaria. During the course of our experiments we have encountered evidence that validates this hypothesis. Variation at the *ABO* locus demonstrated both disease associations and signals of longstanding balancing selection. We found that blood group O individuals are relatively protected from severe malarial phenotypes. A study published by Rowe *et al.* shortly before our own (a matched case-control study of 567 Malian children) confirmed the protective effect of blood group O, but suggested a greater effect size (OR 0.34, 95% CI 0.19-0.61) (195). Although the study of Malian children had a smaller sample size and was limited to one geographic region, it raises the possibility that our estimated effect size could be too low. A number of experimental factors can introduce bias into genetic association studies. These include genotyping error, sample handling error and choice of controls. We discussed in chapter 2 that the use of cord blood controls, rather than age-matched controls, could affect our results. Depletion of blood group O individuals during childhood (e.g. deaths due to an infectious pathogen) could exaggerate the scale of the protection. However, given the evidence that blood group O protects from at least one major source of childhood mortality (i.e. malaria) the opposite scenario is perhaps more relevant. If there were relatively more deaths among non-blood group O individuals during childhood, this would increase the

frequency of blood group O among our cases (who are generally aged between 6 months and 5 years). This would diminish the estimate of protection when compared with cord blood controls. A true OR of ~0.3 for blood group O would suggest that ABO blood group explains a significant proportion of interpersonal variation in severe malaria susceptibility. To put this in context, sickle heterozygotes have an OR of ~0.1, but generally represent only 10-15% of the population. In contrast, 45-50% of sub-Saharan populations are blood group O (40).

Failure to replicate the precise size of an effect is a relatively minor issue in comparison to the more common difficulty of being unable to replicate an association at all, a predicament we encountered both with CD36 and ICAM-1. Non-replication of results is a chronic problem among genetic association studies (239). The reasons for this phenomenon are multitudinous but tend to fall into two broad categories, either type I error in the initial study, or type II error in the replication experiment. Genotyping error, missing data and sample handling error have the capacity to generate both false positive and false negative results. Even large genotyping projects are not immune to such errors (for example, the key HapMap phase 1 data freeze, version 16ci, was published after releases, 16a, 16b and 16c sequentially needed correction). Studies based in developing countries face the additional challenges of collection, transport, storage and processing of blood samples in what are often difficult conditions. These factors reduce DNA quality and therefore the accuracy of the resulting genotype data. We made efforts to prevent (and if possible correct) errors by using automated genotype-calling algorithms, genotyping of replicate samples, genotyping samples with known genotypes, and employing quality-control filters such as Hardy-

Weinberg equilibrium, genotyping success rate and pedigree error rate. Careful visual inspection of raw genotyping data is still a sensitive (although subjective) way to detect serious problems. In future, as the amount of data grows, development of new automated genotype calling and quality-control methods will be vital.

Population stratification is often highlighted as a potential source of confounding in association studies (299). Differential sampling from genetically distinct sub-groups, between cases and controls, can cause neutral genetic markers to be correlated with disease. This is of particular concern in African populations, where there are high levels of genetic diversity and multiple ethnic groups sometimes living in close proximity. A range of analytical techniques have been employed to handle population stratification, these include Mantel-Haenszel testing and logistic regression. Logistic regression has the benefit of being able to accommodate multiple potentially confounding variables. However, such techniques rely on the investigator knowing what the relevant confounding variables are. It is possible that important elements of ethnic, social or cultural stratification may not be known. As genetic data sets have grown in size, new techniques have been developed to use large numbers of neutral markers to either calibrate the significance of association results (e.g. genomic control) (300), to assign individuals to discrete sub-populations (e.g. structured association analysis) (301) or more complex methods to dissect population structure (e.g. principal-component analysis) (302).

Another approach to the problem of population stratification is the use of family-based association studies. The nuclear family trio design we employed is robust to population

stratification and has roughly the same power per case as population-based studies (303). Furthermore, parents are relatively easy to collect in childhood-onset diseases like malaria. However, in comparison to case-control studies, the trio design incurs extra costs for the collection, processing, storage and genotyping of two parents for each case. There will also be some families where it is not feasible to collect parents, or where the wrong person is identified as a parent (e.g. a relative or another community member). Family trio studies can be biased by genotyping error, missing genotypes and pedigree misspecification (304). The design is also sensitive to loss of power from missing data (e.g. loss of a child's sample renders the two parental samples useless as controls, although it may be easier to impute some missing genotypes from parent-offspring data). Like case-control studies, the family trio design is susceptible to bias from changes in allele frequency between the ages of cases and controls. However, the 'time window' for such bias is substantially longer in trio studies and includes gamete production, fertilization, embryonic implantation, fetal life and parturition (as well as survival from birth to mid-childhood), a series of life-stages that are subject to strong selection pressures. A genetic variant affecting survival during any of these key events could distort the distribution of alleles in cases relative to their parents. Ironically, a false positive result caused by this type of bias might be replicated between experiments (assuming the cause of the transmission distortion is present in the different study populations).

Multiple testing is an important, although often ignored, source of type I error in genetic association studies. Even in a small study it is common for several polymorphisms to be tested in three or four genetic models, and with multiple phenotypes. The difficulty faced by

investigators is quantifying the amount of testing that has actually taken place. In a candidate gene study many tests will be strongly correlated (e.g. due to LD between SNPs, correlation between genetic models or phenotypes with common pathogenesis) and simple correction techniques, such as the Bonferroni method, are too conservative in this setting (299). It is notable that many of the unreplicated associations in the genetic epidemiology of malaria literature have *P*-values in the region of 0.05 to 0.01, a significance level unlikely to be robust to minor corrections for multiple testing. As genetic association studies get larger, and consider more markers, the challenge of multiple testing becomes greater.

It is possible that genetic differences between study populations could explain some instances of non-replication in genetic association studies of malaria. Haplotypic diversity in African populations is greater than in other human populations (305). This could lead to the situation were LD between a genotyped neutral marker 'A' and an ungenotyped functional variant 'B', is strong in one population, but weaker, absent or even reversed in another. As power to detect an association with SNP A is proportional to the r^2 between SNPs A and B, LD differences could lead to non-replication (306). Regional differences in the repertoire of functional alleles at candidate loci could also be a factor in some instances of non-replication. We encountered this issue in our study of CD36 deficiency and malaria. It has been clear since the initial publication by Aitman *et al.* that the genetic basis of CD36 deficiency in Africa is heterogeneous. Aitman identified several CD36 deficiency alleles but focused on 1264G and 1439C/1444delA because these were relatively common in Gambians. Most subsequent African studies have also focused on these two alleles although the frequency of other deficiency alleles outside of The Gambia is unknown. Allelic

heterogeneity causes loss of power in association studies, particularly when the other alleles are common and not genotyped (280, 307). The impact of unrecognized allelic heterogeneity on long-range haplotype testing is unknown, but loss of power is a likely consequence. In The Gambia, the REHH around *CD36* 1264G was raised, although not significantly different from alleles of a similarly low frequency. However, if there are other, unrecognized CD36 deficiency alleles under selection in The Gambia this would have increased EHH on 1264T haplotypes, reducing REHH on the 1264G haplotypes. Alternatively, competition between different deficiency alleles in The Gambia could be suppressing 1264G allele frequency in a manner similar to the interaction between HbS and HbC alleles (308), or there could be negative epistasis with another locus [in a manner similar to the relationship between HbS and α-thalassaemia (309)]. Despite the lack of a consistent disease association between CD36 deficiency alleles and malaria, the distribution of CD36 deficiency in southern Asia and Africa is striking and remains strongly suggestive of a link to malaria. The high endemicity African areas with low 1264G frequency may simply have other CD36 deficiency alleles present. A possible demographic event underlying the distribution of 1264G is the series of historical population migrations known as the 'Bantu expansions'. These migrations, which spread the Bantu language, culture and genes south and east across Africa, are thought to have begun in the region of Nigeria and Cameroon around three thousand years ago. This event would fit a model of 1264G originating in Central-West Africa and spreading mainly south and west. Recent migration of the allele westward to The Gambia might explain why a selection event there has only recently been initiated. Hypothetically, given time (i.e. several further human generations

affected by endemic *P. falciparum*), if the 1264G allele increases rapidly in frequency, its haplotype length may eventually become significantly different from that other alleles.

Many different phenotypes are available for use in genetic association studies of malaria. Fever and parasitaemia have been used in some clinical studies and are both related to the initial response to infection. They are relatively simple to quantify, however, both demonstrate substantial variation during the course of an infection and they are commonly raised in non-severe infections. Paradoxically, sequestration could reduce apparent parasitaemia while leading to life-threatening consequences. The clinical outcomes we wish to prevent are death and permanent disability caused by severe malaria. There are multiple levels at which host genetic variation could operate to prevent (or predispose to) severe disease: initial response to parasitaemia, risk of progression to severe disease, clearance of prolonged infection and ability to generate effective long-term immunity after infection. By choosing to study extreme, life-threatening phenotypes, we hoped to enrich our cases for those individuals most susceptible to them. This susceptibility may be due to genetic variation operating at any level of the host-parasite relationship. The problem for this approach is that severe malaria phenotypes only occur in a small minority of children (1-2%), making recruitment for large genetic studies harder. Differences in the phenotypes employed between studies may explain some episodes of non-replication. Even when a particular phenotype is nominally shared between two studies it may be measured with different thresholds, alternative parameters or by people with disparate clinical practices. These differences in phenotype definition could change the underlying phenomenon being studied. Our phenotypes were derived from primary clinical data with consistent criteria

applied to data from different study sites. Although this primary data is generally quantitative (e.g. BCS or haemoglobin) it still relies, to an extent, on the clinical practices of the recruiting teams. Features of malaria such as fever, prostration, respiratory distress, acidosis and anaemia are common symptoms in sick children presenting to clinics in sub-Saharan Africa. Many do have severe malaria, but a minority will have other diseases complicated by incidental parasitaemia or non-severe malaria. Illnesses which could mimic severe malaria include meningitis, other infections, toxicity to traditional medicines, febrile convulsions, over-sedation, or even long-standing epilepsy. Children with severe non-malarial illness who are inadvertently recruited to a genetic association study of malaria will contribute to loss of power and may bias the results.

Lack of power in association studies is an important factor leading to type II errors and may explain a significant proportion of non-replication in the wider literature of genetic association studies. There are several examples among candidate gene studies of the PfEMP1 host ligands where initial positive studies were followed-up by replication experiments using smaller sample sizes. The frequency of under-powered experiments reflects the accessibility of the 'small association study' to research groups with limited resources (for example those in developing countries). Although the collection of samples requires substantial effort, the basic design and analysis of case-control studies is relatively straightforward. Smaller groups are also more likely to use older genotyping technologies and analytical approaches. Publication is a key measure of success in research. However, prestigious scientific journals rarely agree to publish association studies with negative results. These factors reward authors who exaggerate the significance of their results, or

mine data for nominally significant *P*-values. When a false association report enters the literature it can be somewhat difficult to disprove. Our work used around 5-10 times the sample size of the initial studies, meaning lack of power is unlikely to explain our non-replication of the CD36 and ICAM-1 associations. A number of additional factors determine power in association studies. These include the ratio of cases to controls, genetic model (e.g. recessive or dominant) and allele frequency. For example, although the effect size of *ABO* rs8176719 was modest, the allele frequency was high and the genetic model was effectively multiplicative, both of which improved power. In contrast, the low allele frequency and possibly recessive mode of action for *CD36* T1264G will have reduced power.

Genetic variation underlying human blood group phenotypes has recurrently featured in our work. The ABO, Knops (CR1), and Nak^{a} (CD36) blood groups are all systems linked to malaria susceptibility. In the past, blood groups have been used as primitive genetic markers, studied in early association experiments in the hope that they were in LD with functional variation. However, in the case of malaria, blood group variation appears to be intrinsically functional, modulating the characteristics of the haemopoietic cell surface, a vital interface between host and parasite. There are around 300 human blood group systems of varying clinical significance (310), and the genetic variation underlying these systems is generally well characterized. It seems likely that human blood group systems will provide further interesting candidate genes for future studies of the genetic epidemiology of malaria.

Another theme in our work has been co-evolution between host and parasite. Given the functional differences between ICAM-1^{Kilifi} and the wild-type molecule (e.g. altered binding

for specific parasite strains), the lack of a disease association raises the possibility that some wild parasite strains have adapted to ICAM-1^{Kilifi}. Similarly, the prevalence of CD36 deficiency alleles in Africa without a disease association could also be explained by a historical advantage against malaria that has since been lost by parasite adaptation. Furthermore, the preference for ABO B antigen binding among malaria strains from South East Asia (146) may also represent co-evolution in response to the higher frequency of blood group B among local human populations. We reported low levels of F_{ST} around the *ABO* locus, and proposed that this lack of population differentiation represents a signal of balancing selection. Previous publications have considered the *ABO* gene using a number of selection metrics and there is near consensus that the locus has indeed experienced balancing selection. Our suggestion that low F_{ST} around *ABO* is evidence of balancing selection has subsequently been questioned in a paper by Calafell *et al.* (311). The authors found it implausible that similar selection constraints could have operated in different human populations. What mechanism might explain consistent selection constraints? Part of the difficulty in answering this question stems from our lack of understanding about the true function of the ABO system. One possibility is that during viral or bacterial replication host glycosyltransferases add ABO antigens to the surface of the microorganism. When the organism infects a new host the ABO antigens might present an opportunity to either prime the new host's immune system or to directly attack the parasite (312). Evidence for this model has been found with HIV (313). Such a mechanism would discourage fixation of the ABO alleles, although the optimum allele frequency would still be determined by the precise balance of selection constraints (314). The constraints on *ABO* may be similar in different populations because many viruses and bacteria are common across the globe (e.g. *H. pylori*,

rhinovirus, or *E. coli*). Alternatively, the same allele frequency may be optimum regardless of the sets of microbial pathogens involved.

The evidence for a parent-of-origin effect at *ABO* is a novel finding with potentially interesting biological implications. One possibility is that genomic imprinting at the *ABO* locus modulates tissue specific expression of ABO antigens. Imprinted genes generally control pre- and post-natal growth, including development of the placenta (315). The abundance of maternal-only expression among imprinted genes is thought to reflect the need to integrate mother and fetus (316). Field studies indicate that blood group O mothers are at increased risk of placental malaria in their first pregnancy (317). An explanation for this is that fetal and placental ABO antigens provide novel ligands for iRBC sequestration in blood group O mothers, in a manner similar to CSA. Blood group O primigravidae with a non-O blood group fetus will lack neutralizing antibodies against A or B antigen binding parasites. Suppression of paternal ABO alleles could be an evolutionary strategy to enhance fetal survival, maintain placental function, and improve pre- and post-natal growth. Another possibility is that our parent-of-origin effect actually reflects a loss of pregnancies in which the fetus and mother have discordant blood groups (e.g. through haemolytic disease, placental malaria or another mechanism). It is also important to remember that our evidence suggestive of a parent-of-origin effect requires further verification. Family trio studies have limited power to detect parent-of-origin effects as sample size is effectively half that of a conventional association test (paternal transmissions are compared against maternal transmissions). It would be interesting to extend field studies of ABO blood group and placental malaria to consider the relationship between infant and maternal blood group, and

disease susceptibility. Another possible avenue of research would be functional analysis of maternal and paternal *ABO* alleles in blood group AB individuals and their parents (e.g. through flow cytometry of erythrocytes) looking for differential expression between maternal and paternal alleles.

Chapter 7

Conclusions

This project has focused on the host ligands of PfEMP1, scrutinizing variation at three key genes for evidence of disease associations with severe malarial phenotypes. We have demonstrated compelling evidence for both a disease association and evolutionary selection at the *ABO* locus. In contrast, we found no evidence for disease associations at either the *ICAM1* or *CD36* genes. Population genetic analysis also suggested little evidence for recent selection at *ICAM1*. However, there was a signal of recent positive selection around the *CD36* nonsense allele 1264G, a selection event that appears to be focused in Central-West Africa. We employed a mixture of case-control and family-based association studies. This allowed us to consider the advantages and disadvantages of both study designs in the context malaria epidemiology. Along the way, we have demonstrated evidence for the validity of long-range haplotype methods to detect recent selection, and considered how genetic associations and selection data, in concert, can help define the biological significance of candidate genes. The host ligands of PfEMP1 are clearly linked to the pathogenesis of severe malaria, yet many of the underlying genes have never been analysed in genetic association studies. Further investigation of these candidates is likely to reveal valuable insights into the pathogenesis of life-threatening malaria and, perhaps, suggest novel strategies for therapeutic intervention.

Appendix A

Supplementary Tables for Chapter 2

Table A.1. Genotype, haplotype and ABO blood group frequencies for severe malaria cases, cord blood controls and parents.

SNP/Haplotype or Blood Group		Gambia				Malawi				Kenya			
		Control	%	Case	%	Control	%	Case	%	Control	%	Case	%
Population-based study													
rs8176719	1/1	275	(44.3)	269	(38.9)	193	(48.4)	303	(42.8)	467	(52.4)	330	(46.8)
	1/2	282	(45.4)	310	(44.8)	162	(40.6)	322	(45.5)	358	(40.2)	319	(45.2)
	2/2	64	(10.3)	113	(16.3)	44	(11.0)	83	(11.7)	66	(7.4)	56	(7.9)
rs8176746	1/1	413	(66.5)	449	(64.7)	270	(67.2)	485	(68.1)	642	(71.6)	483	(69.1)
	1/2	179	(28.8)	209	(30.1)	122	(30.3)	201	(28.2)	236	(26.3)	193	(27.6)
	2/2	29	(4.7)	36	(5.2)	10	(2.5)	26	(3.7)	19	(2.1)	23	(3.3)
ABO Haplotype	O	830	(67.0)	845	(61.2)	547	(68.9)	922	(65.5)	1286	(72.5)	971	(69.6)
	A	189	(15.3)	270	(19.6)	119	(15.0)	259	(18.4)	242	(13.6)	211	(15.1)
	B	219	(17.7)	265	(19.2)	128	(16.1)	227	(16.1)	246	(13.9)	214	(15.3)
Blood Group	O	274	(44.3)	268	(38.8)	193	(48.6)	301	(42.8)	465	(52.4)	327	(46.8)
	A	150	(24.2)	188	(27.2)	84	(21.2)	196	(27.8)	191	(21.5)	170	(24.4)
	B	172	(27.8)	185	(26.8)	99	(24.9)	164	(23.3)	206	(23.2)	177	(25.4)
	AB	23	(3.7)	49	(7.1)	21	(5.3)	43	(6.1)	25	(2.8)	24	(3.4)
Family-based study													
		Parents	%	Case	%	Parents	%	Case	%	Parents	%	Case	%
rs8176719	1/1	1166	(44.6)	538	(41.2)	193	(43.1)	87	(38.8)	241	(52.1)	110	(47.2)
	1/2	1140	(43.6)	606	(46.4)	201	(44.9)	103	(46.0)	186	(40.2)	103	(44.2)
	2/2	307	(11.7)	161	(12.3)	54	(12.1)	34	(15.2)	36	(7.8)	20	(8.6)
rs8176746	1/1	1778	(67.9)	859	(66.0)	301	(67.6)	158	(70.5)	333	(72.5)	164	(71.0)
	1/2	752	(28.7)	396	(30.4)	132	(29.7)	55	(24.6)	116	(25.3)	58	(25.1)
	2/2	88	(3.4)	47	(3.6)	12	(2.7)	11	(4.9)	10	(2.2)	9	(3.9)
ABO Haplotype	O	1597	(71.4)	1505	(67.2)	284	(70.6)	257	(63.9)	304	(74.9)	295	(72.7)
	A	358	(16.0)	392	(17.5)	65	(16.2)	88	(21.9)	58	(14.3)	64	(15.8)
	B	283	(12.6)	341	(15.2)	53	(13.2)	57	(14.2)	44	(10.8)	47	(11.6)
Blood Group	O	597	(53.4)	519	(46.4)	102	(50.7)	85	(42.3)	114	(56.2)	107	(52.7)
	A	272	(24.3)	298	(26.6)	53	(26.4)	67	(33.3)	47	(23.2)	54	(26.6)
	B	206	(18.4)	242	(21.6)	39	(19.4)	36	(17.9)	35	(17.2)	35	(17.2)
	AB	44	(3.9)	60	(5.4)	7	(3.5)	13	(6.5)	7	(3.4)	7	(3.4)

Table A.2. Sample sizes, genotyping assay performance and sample characteristics, by study design and region of origin.

Study Site	Samples	Size	Marker	Typing Success	HWE	Male (%)	Mean Age (Months)	Ethnic Groups
Population-based study								
Gambia	Cases	701	rs8176719	98.7%	0.16	51.0	51.9	103 Fula, 174 Jola, 264 Mandinka, 79 Wolloff, 81 other
			rs8176746	99.0%	0.09			
	Controls	624	rs8176719	99.5%	0.55	50.0	0	137 Fula, 89 Jola, 173 Mandinka, 94 Wolloff, 131 other
			rs8176746	99.5%	0.12			
Malawi	Cases	718	rs8176719	98.6%	0.90	49.3	42.1	*
			rs8176746	99.2%	0.43			
	Controls	405	rs8176719	98.5%	0.30	52.2	0	
			rs8176746	99.3%	0.48			
Kenya	Cases	708	rs8176719	99.6%	0.09	55.1	30.3	108 Chonyi, 320 Giriama, 41 Kauma, 239 other
			rs8176746	98.7%	0.56			
	Controls	902	rs8176719	98.8%	0.87	51.6	0	183 Chonyi, 405 Giriama, 55 Kauma, 259 other
			rs8176746	99.4%	0.70			
Family-based study								
Gambia	Trios	1320	rs8176719	98.9%	0.51	53.6	47.9	194 Fula, 226 Jola, 537 Mandinka, 177 Wolloff, 186 other
			rs8176743	98.8%	0.49			
			rs8176746	99.0%	0.60			
			rs8176747	97.3%	0.21			
Malawi	Trios	225	rs8176719	99.6%	0.63	54.7	38.5	*
			rs8176746	99.1%	0.08			
Kenya	Trios	234	rs8176719	99.1%	0.51	48.1	29.8	51 Chonyi, 134 Giriama, 15 Kauma, 34 other
			rs8176746	98.3%	0.68			

NOTE.- HWE=Chi-squared test for Hardy-Weinberg equilibrium in controls, or Z-test for Hardy-Weinberg equilibrium in untransmitted parental alleles treated as a diploid genotype. Controls are cord blood samples, representing population allele distribution at birth and are therefore age zero. *Malawians in the study region are considered relatively ethnically homogenous; therefore no ethnicity information was available.

Appendix B

Supplementary Tables for Chapter 3

Table B.1. FBAT *ICAM1* single SNP association results for malaria sub-phenotypes, and for severe malaria in specific populations.

SNP	S	E(S)	Var(S)	P-value	S	E(S)	Var(S)	P-value	S	E(S)	Var(S)	P-value
	Gambia Families				Malawi Families				Kenya Families			
rs5490	213	220.50	98.25	0.45	111	103.00	46.00	0.24	105	106.50	46.25	0.83
rs5030340	51	51.00	25.00	1.00	27	30.00	15.00	0.44	22	24.00	12.00	0.56
rs5030344	56	57.00	28.50	0.85	14	16.50	8.25	0.38	17	13.00	6.50	0.12
rs5030351	180	192.50	89.25	0.19	118	106.00	46.00	0.08	116	119.00	48.50	0.67
rs5491	453	470.00	212.00	0.24	122	108.00	47.00	**0.04**	112	110.50	44.75	0.82
rs281432	213	220.50	98.25	0.45	79	80.50	37.25	0.81	91	97.50	44.75	0.33
rs5496	123	125.50	61.25	0.75	41	35.00	17.50	0.15	12	19.00	9.50	**0.02**
rs5498	248	265.00	131.00	0.14	42	46.50	23.25	0.35	56	56.50	27.75	0.92
rs3093030	51	64.50	31.75	**0.02**	18	17.50	8.75	0.87	31	30.00	15.00	0.80
rs281438	365	369.00	126.50	0.72	157	154.00	53.50	0.68	162	160.50	57.25	0.84
rs2569693	47	53.00	26.00	0.24	12	10.50	5.25	0.51	28	26.00	12.50	0.57
rs281439	338	337.00	131.00	0.93	123	130.00	52.50	0.33	122	122.50	52.75	0.95
rs281440	276	281.50	119.25	0.61	74	84.50	38.25	0.09	111	112.50	45.25	0.82
rs2075741	120	136.00	65.50	0.05	43	41.50	19.75	0.74	58	59.00	26.50	0.85
rs11575074	91	90.50	44.25	0.94	46	48.50	23.75	0.61	35	43.00	21.00	0.08
	Severe Malaria				Cerebral Malaria				Severe Malarial Anaemia			
rs5490	429	430.00	190.50	0.94	255	255.00	112.50	1.00	125	129.50	56.25	0.55
rs5030340	100	105.00	52.00	0.49	63	67.50	33.75	0.44	26	27.50	13.75	0.69
rs5030344	87	86.50	43.25	0.94	42	45.00	22.50	0.53	24	21.00	10.50	0.35
rs5030351	414	417.50	183.75	0.80	249	245.50	107.25	0.74	98	115.50	51.75	**0.015**
rs5491	687	688.50	303.75	0.93	362	341.50	151.75	0.10	159	174.00	79.00	0.09
rs281432	383	398.50	180.25	0.25	216	231.00	105.00	0.14	103	108.00	48.50	0.47
rs5496	176	179.50	88.25	0.71	99	97.00	47.50	0.77	52	51.50	24.75	0.92
rs5498	346	368.00	182.00	0.10	172	183.00	91.50	0.25	87	94.50	45.25	0.26
rs3093030	100	112.00	55.50	0.11	54	62.00	31.00	0.15	33	33.00	16.00	1.00
rs281438	684	683.50	237.25	0.97	406	392.00	139.00	0.24	196	200.50	68.25	0.59
rs2569693	87	89.50	43.75	0.71	44	50.00	25.00	0.23	25	25.00	11.50	1.00
rs281439	583	589.50	236.25	0.67	327	336.50	134.75	0.41	163	168.50	68.75	0.51
rs281440	461	478.50	202.75	0.22	255	274.50	117.25	0.07	116	125.00	54.50	0.22
rs2075741	221	236.50	111.75	0.14	123	134.00	64.50	0.17	71	69.00	32.50	0.73
rs11575074	172	182.00	89.00	0.29	97	104.50	51.25	0.29	51	54.50	26.75	0.50

FBAT statistic 'S', expected FBAT statistic 'E(S)', variance 'Var(S)' and *P*-value are reported under the allelic model, for transmissions of the minor allele.

Table B.2. Genotyping assay performance by study design and region of origin.

Study Site	Marker	Success	MAF	HWE
Family-based association study				
Gambia	rs5491*	99.80%	0.221	0.62
	rs5498*	99.80%	0.115	0.72
	rs5490	96.40%	0.241	0.99
	rs5030340	97.40%	0.046	0.81
	rs5030344	99.80%	0.05	0.03
	rs5030351	97.10%	0.211	0.68
	rs281432	94.60%	0.261	0.36
	rs5496	98.90%	0.116	0.21
	rs3093030	94.10%	0.074	0.97
	rs281438	96.40%	0.465	0.56
	rs2569693	78.10%	0.069	0.92
	rs281439	96.00%	0.413	0.33
	rs281440	95.90%	0.321	0.26
	rs2075741	99.00%	0.14	0.65
	rs11575074	98.10%	0.082	0.59
Malawi	rs5490	97.90%	0.256	0.7
	rs5030340	99.00%	0.082	0.67
	rs5030344	99.40%	0.045	0.48
	rs5030351	98.50%	0.26	0.8
	rs5491	100.00%	0.258	0.77
	rs281432	98.70%	0.217	0.36
	rs5496	98.80%	0.076	0.23
	rs5498	100.00%	0.129	0.66
	rs3093030	99.10%	0.038	0.55
	rs281438	98.40%	0.47	0.04
	rs2569693	94.20%	0.023	0.74
	rs281439	98.70%	0.401	0.75
	rs281440	90.10%	0.334	0.88
	rs2075741	99.10%	0.1	0.1
	rs11575074	97.60%	0.129	0.74
Family-based association study (continued)				
Kenya	rs5490	97.00%	0.299	0.65
	rs5030340	98.30%	0.064	0.24
	rs5030344	99.70%	0.019	0.76
	rs5030351	98.90%	0.325	0.71
	rs5491	96.60%	0.323	0.65
	rs281432	98.40%	0.278	0.43
	rs5496	97.90%	0.058	0.77
	rs5498	98.40%	0.142	0.42
	rs3093030	99.40%	0.065	0.29
	rs281438	99.00%	0.444	0.42
	rs2569693	97.70%	0.054	0.4
	rs281439	98.90%	0.332	0.14
	rs281440	94.70%	0.339	0.63
	rs2075741	99.90%	0.132	0.26
	rs11575074	99.60%	0.123	0.36
Population-based association study				
Gambia	rs5491	99.30%	0.196	0.77
	rs5498	99.50%	0.11	0.82
Malawi	rs5491	99.60%	0.296	0.51
	rs5498	98.80%	0.149	0.37
Kenya	rs5491	99.10%	0.284	0.14
	rs5498	99.40%	0.134	0.76

NOTE.- HWE=Chi-squared test for Hardy-Weinberg equilibrium in controls, or Z-test for Hardy-Weinberg equilibrium in untransmitted parental alleles treated as a diploid genotype. Success = percentage of samples (cases, controls or parents) successfully genotyped for each marker. MAF = minor allele frequency. * Results for 1320 Gambian trios, rest of Gambian family results represent those from 612 trios.

Appendix C

Supplementary Tables and Figures for Chapter 4

Table C.1. Global survey of *CD36* T1264G allele frequencies.

Region	Population	Geographic origin	G allele	Total 2n	Freq	Climate	Prevalence
Sub-Saharan Africa	Yoruba	Nigeria	13	25	0.26	99.4%	48.0%
Sub-Saharan Africa	San	Namibia	0	6	0	7.6%	6.1%
Sub-Saharan Africa	Mbuti Pygmies	Democratic Republic of Congo	0	7	0	99.9%	55.3%
Sub-Saharan Africa	Mandenka	Senegal	0	21	0	96.9%	37.4%
Sub-Saharan Africa	Biaka Pygmies	Central African Republic	3	26	0.058	99.9%	41.7%
Sub-Saharan Africa	Bantu S.W. (Herero)	South Africa	0	2	0	14.8%	-
Sub-Saharan Africa	Bantu S.E. (Zulu)	South Africa	0	1	0	" "	-
Sub-Saharan Africa	Bantu S.E. (Tswana)	South Africa	0	2	0	" "	-
Sub-Saharan Africa	Bantu S.E. (S.Sotho)	South Africa	0	1	0	" "	-
Sub-Saharan Africa	Bantu S.E. (Pedi)	South Africa	0	1	0	" "	-
Sub-Saharan Africa	Kasem	Ghana	140	474	0.148	98.3%	48.6%
Sub-Saharan Africa	Nankan	Ghana	63	249	0.127	" "	" "
Sub-Saharan Africa	Buli	Ghana	3	14	0.107	" "	" "
Sub-Saharan Africa	Fula	Gambia	4	136	0.015	100.0%	36.0%
Sub-Saharan Africa	Jola	Gambia	2	89	0.011	" "	" "
Sub-Saharan Africa	Mandinka	Gambia	5	173	0.014	" "	" "
Sub-Saharan Africa	Manjago	Gambia	2	19	0.053	" "	" "
Sub-Saharan Africa	Serehuli	Gambia	5	52	0.048	" "	" "
Sub-Saharan Africa	Serere	Gambia	1	19	0.026	" "	" "
Sub-Saharan Africa	Wolloff	Gambia	5	94	0.027	" "	" "
Sub-Saharan Africa	Chonyi	Kenya	37	183	0.101	21.2%	36.3%
Sub-Saharan Africa	Duruma	Kenya	8	79	0.051	" "	" "
Sub-Saharan Africa	Giriama	Kenya	75	405	0.093	" "	" "
Sub-Saharan Africa	Kauma	Kenya	4	55	0.036	" "	" "
Sub-Saharan Africa	Bantu N.E.	Kenya	1	11	0.045	" "	" "
Sub-Saharan Africa	Malawian	Malawi	72	405	0.089	77.0%	58.2%
Oceania	Papuan	New Guinea	0	17	0	-	-
Oceania	NAN Melanesian	Bougainville	0	19	0	-	-
North Africa	Mozabite	Algeria (Mzab)	0	23	0	-	-
Middle East	Palestinian	Israel (Central)	0	50	0	-	-
Middle East	Druze	Israel (Carmel)	0	44	0	-	-
Middle East	Bedouin	Israel (Negev)	0	44	0	-	-
Europe	Tuscan	Italy	0	8	0	-	-
Europe	Sardinian	Italy	0	28	0	-	-
Europe	Russian	Russia	0	25	0	-	-
Europe	Orcadian	Orkney Islands	0	15	0	-	-
Europe	North Italian	Italy (Bergamo)	0	13	0	-	-
Europe	French Basque	France	0	22	0	-	-

Table C.1. Global survey of *CD36* T1264G allele frequencies (*continued*).

Region	Population	Geographic origin	G allele	Total 2n	Freq	Climate	Prevalence
Europe	French	France	0	29	0	-	-
Europe	Adygei	Russia Caucasus	0	17	0	-	-
Asia	Yizu	China	0	10	0	-	-
Asia	Yakut	Siberia	0	23	0	-	-
Asia	Xibo	China	0	5	0	-	-
Asia	Uygur	China	0	9	0	-	-
Asia	Tujia	China	0	10	0	-	-
Asia	Tu	China	0	6	0	-	-
Asia	Sindhi	Pakistan	0	23	0	-	-
Asia	She	China	0	9	0	-	-
Asia	Pathan	Pakistan	0	23	0	-	-
Asia	Oroqen	China	0	10	0	-	-
Asia	Naxi	China	0	10	0	-	-
Asia	Mongola	China	0	10	0	-	-
Asia	Miaozu	China	0	10	0	-	-
Asia	Makrani	Pakistan	1	24	0.021	-	-
Asia	Lahu	China	0	9	0	-	-
Asia	Kalash	Pakistan	0	24	0	-	-
Asia	Japanese	Japan	0	26	0	-	-
Asia	Hezhen	China	0	9	0	-	-
Asia	Hazara	Pakistan	0	21	0	-	-
Asia	Han	China	0	41	0	-	-
Asia	Daur	China	0	10	0	-	-
Asia	Dai	China	0	9	0	-	-
Asia	Cambodians	Cambodia	0	11	0	-	-
Asia	Burusho	Pakistan	0	21	0	-	-
Asia	Brahui	Pakistan	0	24	0	-	-
Asia	Balochi	Pakistan	0	22	0	-	-
America	Surui	Brazil	0	21	0	-	-
America	Pima	Mexico	0	25	0	-	-
America	Maya	Mexico	0	25	0	-	-
America	Karitiana	Brazil	0	24	0	-	-
America	Colombians	Colombia	0	13	0	-	-

NOTE.- G allele, number of G alleles found; Total 2n, total number of individuals genotyped; Freq, minor allele frequency; Climate, proportion of a country's population living in areas >75% suitable for malaria (i.e. endemic; perennial or seasonal areas); Prevalence, average national prevalence of malaria. Epidemiological statistics based on 1990 data (MARA-Lite Software 3.0.0, http://www.mara.org.za).

Table C.2. *CD36* SNP association results ($P < 0.05$), in Gambian family and case-control studies.

Study	Phenotype	Model	SNP	Lower	OR	Upper	*P*-value	Perm.
Family	CM	recessive	rs10216184	0.24	0.49	0.98	0.043	
CC	SA	overdominant	rs10480808	1.06	1.45	1.99	0.022	
CC	Severe	allelic	rs1049654	1.03	1.21	1.43	0.021	0.05
CC	Severe	recessive	rs1049654	1.04	1.43	1.96	0.028	
Family	CM	overdominant	rs10499859	0.55	0.71	0.92	0.0099	
CC	CM	overdominant	rs1537593	0.55	0.73	0.96	0.026	
Family	SA	allelic	rs1572230	0.45	0.65	0.95	0.026	0.038
Family	SA	dominant	rs1572230	0.43	0.64	0.94	0.025	
Family	SA	overdominant	rs1572230	0.44	0.66	0.97	0.034	
CC	SA	overdominant	rs1722501	0.42	0.59	0.83	0.0026	
CC	Severe	overdominant	rs1722501	0.61	0.78	0.98	0.036	
CC	SA	overdominant	rs1924	1.02	1.40	1.93	0.039	
CC	SA	recessive	rs3173798	0.07	0.23	0.77	0.017	
CC	Severe	recessive	rs3211817	0.06	0.24	0.92	0.037	
Family	CM	overdominant	rs3211822	1.01	1.25	1.54	0.038	
Family	CM	recessive	rs3211822	0.50	0.69	0.97	0.034	
Family	Severe	overdominant	rs3211822	1.03	1.18	1.34	0.014	
CC	SA	overdominant	rs3211828	1.02	1.41	1.96	0.037	
CC	SA	recessive	rs3211828	0.15	0.36	0.88	0.025	
CC	Severe	recessive	rs3211842	0.50	0.69	0.96	0.026	
CC	SA	overdominant	rs3211867	1.09	1.51	2.07	0.012	
CC	SA	recessive	rs3211869	0.11	0.33	0.95	0.040	
Family	CM	overdominant	rs3211883	0.60	0.74	0.92	0.0071	
CC	SA	allelic	rs3211899	0.35	0.56	0.89	0.015	0.0077
CC	SA	dominant	rs3211899	0.32	0.53	0.87	0.011	
CC	SA	overdominant	rs3211899	0.32	0.52	0.86	0.010	
Family	CM	recessive	rs3211909	0.11	0.32	0.93	0.036	
Family	Severe	recessive	rs3211928	0.36	0.53	0.77	0.0010	
Family	CM	allelic	rs3211930	0.37	0.58	0.92	0.020	0.02
Family	CM	dominant	rs3211930	0.37	0.59	0.95	0.029	
CC	Severe	allelic	rs3211931	1.01	1.22	1.47	0.036	0.467
CC	Severe	dominant	rs3211931	1.00	1.26	1.59	0.046	
Family	CM	recessive	rs3211942	1.74	15.03	129.80	0.013	
CC	Severe	allelic	rs4728191	0.61	0.76	0.96	0.019	0.034
CC	Severe	dominant	rs4728191	0.58	0.76	0.99	0.041	
Family	SA	allelic	rs4728191	0.50	0.70	0.99	0.042	0.072
Family	SA	dominant	rs4728191	0.48	0.69	0.99	0.043	
CC	CM	recessive	rs6968407	1.06	1.89	3.38	0.031	
CC	Severe	overdominant	rs6968407	0.62	0.79	0.99	0.045	

NOTE.- Study, family or case-control (CC), Phenotype, cerebral malaria (CM), severe malarial anaemia (SA) or all severe malaria cases (Severe); Model, genetic model; Genotype/allele, compared with genotype 1/1 (major allele homozygote) or allele 1 (major allele); OR, odds ratio; Lower/Upper, 95% confidence interval of odds ratio; Perm, permutation *P*-value calculated using HAPLOVIEW (100,000 permutations).

Table C.3A. UNPHASED analysis of pooled Gambian family and case-control studies (Severe malaria).

Severe	1/2 vs. 1/1		2/2 vs. 1/1					
Marker	OR	95% CI	OR	95% CI	Allelic	Over	Dom	Rec
def	0.83	0.66 - 1.05	1.59	0.74 - 3.39	0.390	0.091	0.194	0.152
G1439C	0.86	0.64 - 1.15	1.99	0.46 - 8.67	0.463	0.269	0.353	0.291
T1264G	0.93	0.67 - 1.31	0.72	0.08 - 6.48	0.649	0.705	0.672	0.796
rs10215288	0.91	0.81 - 1.03	1.01	0.82 - 1.22	0.457	0.100	0.185	0.536
rs10216184	1.11	0.99 - 1.26	1.11	0.84 - 1.47	0.098	0.108	0.077	0.714
rs10480808	1.12	0.99 - 1.27	1.14	0.97 - 1.34	0.086	0.288	0.057	0.462
rs1049654	1.02	0.9 - 1.15	1.02	0.86 - 1.2	0.814	0.821	0.768	0.959
rs10499857	1.00	0.89 - 1.12	0.91	0.71 - 1.17	0.673	0.824	0.890	0.443
rs1054516	0.99	0.89 - 1.12	0.97	0.77 - 1.21	0.806	0.991	0.876	0.787
rs1194178	0.99	0.87 - 1.11	0.97	0.74 - 1.27	0.762	0.842	0.782	0.864
rs1194180	0.93	0.83 - 1.04	0.92	0.74 - 1.14	0.199	0.260	0.171	0.698
rs11971740	1.00	0.89 - 1.12	0.95	0.78 - 1.16	0.722	0.828	0.909	0.576
rs12706912	0.99	0.88 - 1.11	0.91	0.76 - 1.09	0.381	0.728	0.685	0.277
rs13222488	1.06	0.94 - 1.19	1.20	0.95 - 1.52	0.125	0.602	0.234	0.183
rs13228738	1.12	0.99 - 1.28	1.17	0.82 - 1.67	0.069	0.105	0.068	0.553
rs1334511	0.89	0.79 - 1	0.98	0.81 - 1.19	0.322	0.049	0.092	0.555
rs1358337	1.01	0.9 - 1.14	1.03	0.87 - 1.22	0.731	0.987	0.801	0.767
rs1413659	0.95	0.85 - 1.07	0.95	0.76 - 1.18	0.428	0.454	0.384	0.848
rs1413661	1.03	0.92 - 1.16	0.98	0.8 - 1.19	0.877	0.493	0.652	0.664
rs1527463	1.06	0.93 - 1.22	1.06	0.73 - 1.56	0.397	0.413	0.380	0.859
rs1537593	1.04	0.91 - 1.18	1.12	0.95 - 1.31	0.167	0.694	0.401	0.177
rs1572230	0.89	0.77 - 1.03	0.74	0.45 - 1.2	0.060	0.155	0.084	0.313
rs17154155	1.01	0.89 - 1.14	1.07	0.91 - 1.26	0.429	0.700	0.713	0.362
rs17154246	1.01	0.9 - 1.14	0.98	0.77 - 1.25	0.966	0.817	0.887	0.851
rs1722501	0.93	0.83 - 1.04	0.99	0.8 - 1.21	0.403	0.175	0.222	0.757
rs1722502	1.03	0.91 - 1.15	1.03	0.84 - 1.26	0.688	0.722	0.662	0.906
rs1722505	1.02	0.91 - 1.14	0.98	0.8 - 1.2	0.968	0.651	0.796	0.722
rs17263407	1.13	1 - 1.29	1.26	0.9 - 1.78	0.031	0.088	0.039	0.311
rs17266080	1.14	0.99 - 1.3	0.96	0.65 - 1.43	0.163	0.054	0.081	0.600
rs1761661	1.01	0.9 - 1.13	0.97	0.79 - 1.18	0.892	0.780	0.954	0.697
rs1761662	1.02	0.9 - 1.14	0.92	0.71 - 1.19	0.849	0.668	0.918	0.485
rs1761663	1.02	0.9 - 1.16	1.03	0.88 - 1.21	0.676	0.909	0.696	0.789
rs1924	0.96	0.86 - 1.08	0.85	0.68 - 1.05	0.163	0.901	0.349	0.159
rs2366855	0.99	0.88 - 1.12	0.95	0.76 - 1.19	0.720	0.991	0.831	0.674
rs3173798	0.96	0.85 - 1.08	0.90	0.7 - 1.16	0.328	0.589	0.391	0.523
rs3173800	0.99	0.87 - 1.13	1.19	0.85 - 1.68	0.656	0.787	0.909	0.296

NOTE.- Heterozygotes (1/2) or minor allele homozygotes (2/2) compared with major allele homozygotes (1/1); OR, odds ratio; 95% C(onfidence) I(nterval); def, both CD36 deficiency alleles pooled; *P*-values for 4 genetic models: Rec(essive), Over(dominant), Dom(inant) and Allelic.

Table C.3A. UNPHASED analysis of pooled Gambian family- and case-control studies (Severe malaria) - *continued*.

Severe	1/2 vs. 1/1		2/2 vs. 1/1					
Marker	OR	95% CI	OR	95% CI	Allelic	Over	Dom	Rec
rs3211810	1.00	0.88 - 1.14	0.90	0.62 - 1.31	0.804	0.915	0.932	0.580
rs3211817	1.09	0.94 - 1.27	0.94	0.54 - 1.64	0.382	0.252	0.298	0.734
rs3211821	1.04	0.92 - 1.17	0.91	0.73 - 1.14	0.865	0.367	0.721	0.297
rs3211822	1.11	0.98 - 1.25	1.03	0.87 - 1.22	0.417	0.092	0.136	0.643
rs3211828	0.99	0.88 - 1.11	0.83	0.66 - 1.05	0.269	0.837	0.572	0.130
rs3211842	1.03	0.91 - 1.15	0.94	0.78 - 1.12	0.667	0.412	0.881	0.307
rs3211849	1.02	0.9 - 1.15	1.02	0.87 - 1.19	0.821	0.845	0.772	0.952
rs3211850	1.02	0.9 - 1.17	0.96	0.65 - 1.41	0.873	0.711	0.785	0.778
rs3211867	0.95	0.85 - 1.07	0.88	0.73 - 1.05	0.152	0.846	0.284	0.224
rs3211869	0.96	0.86 - 1.08	0.92	0.71 - 1.18	0.411	0.645	0.470	0.592
rs3211870	1.09	0.96 - 1.24	1.07	0.91 - 1.26	0.366	0.323	0.197	0.914
rs3211883	0.97	0.86 - 1.09	1.10	0.93 - 1.32	0.518	0.271	0.877	0.142
rs3211886	1.01	0.88 - 1.16	1.18	0.78 - 1.77	0.638	0.988	0.790	0.448
rs3211888	1.01	0.83 - 1.22	0.82	0.37 - 1.82	0.912	0.893	0.993	0.622
rs3211897	1.05	0.86 - 1.28	-	-	0.919	0.569	0.816	0.033
rs3211899	0.89	0.76 - 1.04	1.19	0.67 - 2.13	0.269	0.114	0.169	0.431
rs3211909	1.03	0.9 - 1.17	0.87	0.6 - 1.25	0.927	0.614	0.842	0.403
rs3211911	0.85	0.69 - 1.06	1.35	0.51 - 3.59	0.232	0.127	0.169	0.467
rs3211913	1.01	0.9 - 1.14	0.88	0.73 - 1.07	0.380	0.432	0.868	0.133
rs3211923	1.02	0.78 - 1.34	0.72	0.08 - 6.12	0.942	0.871	0.906	0.755
rs3211928	1.16	1.03 - 1.3	0.86	0.65 - 1.13	0.327	0.006	0.046	0.078
rs3211930	0.87	0.7 - 1.07	0.21	0.03 - 1.56	0.085	0.232	0.140	0.104
rs3211931	1.11	0.99 - 1.24	1.07	0.84 - 1.37	0.143	0.094	0.083	0.947
rs3211942	1.07	0.9 - 1.26	1.88	0.95 - 3.72	0.190	0.534	0.318	0.076
rs4545029	1.00	0.89 - 1.13	0.94	0.75 - 1.19	0.757	0.875	0.914	0.606
rs4728191	0.88	0.77 - 1.01	0.85	0.56 - 1.28	0.060	0.084	0.058	0.606
rs6960369	0.96	0.85 - 1.07	0.96	0.8 - 1.16	0.522	0.507	0.438	0.914
rs6968407	1.06	0.94 - 1.19	1.31	1.04 - 1.64	0.048	0.770	0.180	0.036
rs6969989	1.13	1 - 1.27	1.02	0.74 - 1.41	0.111	0.041	0.052	0.777
rs6972923	1.01	0.89 - 1.15	1.04	0.88 - 1.21	0.674	0.902	0.801	0.675
rs6973242	1.04	0.93 - 1.18	1.23	0.98 - 1.56	0.120	0.804	0.287	0.102
rs7807607	1.01	0.9 - 1.13	0.98	0.8 - 1.21	0.955	0.850	0.951	0.817
rs819436	1.00	0.89 - 1.12	1.08	0.86 - 1.36	0.707	0.815	0.914	0.490
rs9784998	1.00	0.89 - 1.12	0.84	0.67 - 1.06	0.364	0.649	0.750	0.128
rs997906	1.05	0.92 - 1.21	1.12	0.75 - 1.68	0.399	0.534	0.438	0.641

NOTE.- Heterozygotes (1/2) or minor allele homozygotes (2/2) compared with major allele homozygotes (1/1); OR, odds ratio; 95% C(onfidence) I(nterval); *P*-values for 4 genetic models: Rec(essive), Over(dominant), Dom(inant) and Allelic.

Table C.3B. UNPHASED analysis of pooled Gambian family- and case-control studies (CM).

CM	1/2 vs. 1/1		2/2 vs. 1/1					
Marker	OR	95% CI	OR	95% CI	Allelic	Over	Dom	Rec
def	0.85	0.6 - 1.21	1.52	0.5 - 4.63	0.634	0.351	0.472	0.462
G1439C	0.83	0.52 - 1.32	2.52	0.39 - 16.22	0.677	0.380	0.513	0.267
T1264G	1.00	0.62 - 1.62	-	-	0.811	0.901	0.951	0.317
rs10215288	0.93	0.78 - 1.11	1.00	0.74 - 1.35	0.696	0.392	0.487	0.755
rs10216184	1.26	1.05 - 1.51	1.01	0.65 - 1.57	0.069	0.011	0.019	0.612
rs10480808	1.12	0.93 - 1.37	1.13	0.89 - 1.44	0.296	0.447	0.215	0.710
rs1049654	1.00	0.84 - 1.2	1.00	0.78 - 1.29	0.973	0.969	0.964	0.998
rs10499857	0.98	0.82 - 1.17	0.70	0.47 - 1.05	0.235	0.824	0.552	0.087
rs1054516	1.00	0.84 - 1.2	1.02	0.72 - 1.44	0.941	0.994	0.967	0.926
rs1194178	0.91	0.76 - 1.09	1.19	0.8 - 1.79	0.825	0.221	0.457	0.266
rs1194180	0.97	0.82 - 1.16	0.95	0.7 - 1.3	0.707	0.817	0.723	0.842
rs11971740	0.95	0.8 - 1.13	0.94	0.69 - 1.27	0.543	0.630	0.523	0.819
rs12706912	1.00	0.84 - 1.19	1.00	0.77 - 1.3	0.984	0.988	0.983	0.994
rs13222488	1.08	0.9 - 1.3	1.35	0.95 - 1.91	0.119	0.682	0.248	0.139
rs13228738	1.21	1 - 1.48	1.48	0.87 - 2.5	0.025	0.085	0.034	0.246
rs1334511	0.81	0.68 - 0.98	0.80	0.6 - 1.08	0.037	0.080	0.023	0.519
rs1358337	0.99	0.83 - 1.19	1.15	0.9 - 1.46	0.348	0.458	0.797	0.183
rs1413659	0.96	0.81 - 1.15	0.93	0.67 - 1.29	0.600	0.785	0.642	0.747
rs1413661	0.98	0.82 - 1.16	1.09	0.8 - 1.5	0.819	0.648	0.927	0.496
rs1527463	1.13	0.92 - 1.38	1.15	0.68 - 1.93	0.249	0.288	0.239	0.755
rs1537593	0.96	0.79 - 1.16	1.24	0.98 - 1.58	0.076	0.060	0.803	0.009
rs1572230	0.97	0.78 - 1.2	0.87	0.44 - 1.74	0.665	0.799	0.718	0.732
rs17154155	1.03	0.85 - 1.24	1.07	0.84 - 1.36	0.586	0.995	0.682	0.652
rs17154246	1.02	0.86 - 1.22	1.02	0.7 - 1.5	0.830	0.840	0.819	0.957
rs1722501	0.91	0.76 - 1.08	1.01	0.75 - 1.36	0.603	0.243	0.350	0.641
rs1722502	0.97	0.81 - 1.16	1.18	0.86 - 1.62	0.602	0.517	0.995	0.236
rs1722505	0.96	0.81 - 1.14	1.02	0.74 - 1.41	0.856	0.584	0.694	0.760
rs17263407	1.19	0.98 - 1.44	1.32	0.78 - 2.21	0.057	0.103	0.059	0.451
rs17266080	1.14	0.93 - 1.4	1.12	0.58 - 2.15	0.243	0.233	0.215	0.874
rs1761661	0.95	0.8 - 1.13	1.03	0.75 - 1.41	0.833	0.512	0.639	0.713
rs1761662	0.92	0.77 - 1.1	1.19	0.81 - 1.75	0.928	0.262	0.537	0.258
rs1761663	1.07	0.88 - 1.31	1.14	0.9 - 1.45	0.269	0.964	0.371	0.390
rs1924	0.92	0.77 - 1.1	0.83	0.59 - 1.15	0.201	0.558	0.274	0.376
rs2366855	0.94	0.79 - 1.12	1.05	0.74 - 1.49	0.800	0.436	0.584	0.636
rs3173798	0.93	0.78 - 1.11	0.82	0.55 - 1.22	0.263	0.591	0.350	0.409
rs3173800	1.04	0.85 - 1.27	1.24	0.77 - 1.99	0.443	0.835	0.589	0.410

NOTE.- Heterozygotes (1/2) or minor allele homozygotes (2/2) compared with major allele homozygotes (1/1); OR, odds ratio; 95% C(onfidence) I(nterval); def, both CD36 deficiency alleles pooled; *P*-values for 4 genetic models: Rec(essive), Over(dominant), Dom(inant) and Allelic.

Table C.3B. UNPHASED analysis of pooled Gambian family- and case-control studies (CM) – *continued*.

CM	1/2 vs. 1/1		2/2 vs. 1/1					
Marker	OR	95% CI	OR	95% CI	Allelic	Over	Dom	Rec
rs3211810	0.99	0.82 - 1.21	0.80	0.45 - 1.42	0.660	0.964	0.823	0.453
rs3211817	1.13	0.9 - 1.43	0.38	0.11 - 1.25	0.791	0.217	0.452	0.082
rs3211821	1.02	0.85 - 1.22	1.09	0.77 - 1.53	0.682	0.958	0.782	0.667
rs3211822	1.12	0.93 - 1.35	1.07	0.83 - 1.38	0.435	0.267	0.246	0.940
rs3211828	0.93	0.78 - 1.1	0.74	0.52 - 1.06	0.111	0.679	0.239	0.148
rs3211842	1.03	0.86 - 1.22	0.85	0.65 - 1.12	0.420	0.392	0.929	0.163
rs3211849	1.01	0.83 - 1.22	1.03	0.81 - 1.31	0.801	0.924	0.891	0.795
rs3211850	1.02	0.84 - 1.25	0.74	0.42 - 1.3	0.700	0.675	0.994	0.268
rs3211867	0.88	0.74 - 1.06	0.79	0.6 - 1.03	0.057	0.485	0.099	0.206
rs3211869	0.93	0.77 - 1.11	0.77	0.52 - 1.16	0.185	0.596	0.294	0.276
rs3211870	1.12	0.92 - 1.36	1.21	0.95 - 1.53	0.118	0.802	0.173	0.277
rs3211883	0.91	0.76 - 1.08	1.16	0.89 - 1.51	0.693	0.098	0.537	0.095
rs3211886	1.04	0.84 - 1.29	1.19	0.67 - 2.1	0.572	0.804	0.662	0.594
rs3211888	1.10	0.83 - 1.47	0.50	0.11 - 2.22	0.813	0.459	0.624	0.331
rs3211897	1.18	0.87 - 1.6	-	-	0.593	0.198	0.356	0.071
rs3211899	0.91	0.72 - 1.16	0.99	0.43 - 2.28	0.506	0.453	0.465	0.949
rs3211909	1.04	0.86 - 1.27	0.60	0.33 - 1.11	0.602	0.485	0.957	0.085
rs3211911	0.89	0.65 - 1.21	3.87	0.95 - 15.73	0.913	0.400	0.636	0.037
rs3211913	1.00	0.84 - 1.2	0.83	0.63 - 1.1	0.339	0.539	0.756	0.157
rs3211923	1.07	0.7 - 1.64	2.01	0.2 - 20.78	0.642	0.762	0.697	0.560
rs3211928	1.17	0.98 - 1.41	1.12	0.75 - 1.67	0.144	0.102	0.088	0.915
rs3211930	0.79	0.57 - 1.09	-	-	0.072	0.188	0.116	0.142
rs3211931	1.11	0.93 - 1.32	1.29	0.9 - 1.85	0.119	0.411	0.174	0.270
rs3211942	1.14	0.89 - 1.47	3.10	1.21 - 7.9	0.067	0.372	0.157	0.018
rs4545029	0.91	0.76 - 1.09	1.15	0.82 - 1.63	0.916	0.213	0.487	0.261
rs4728191	0.94	0.77 - 1.15	0.82	0.46 - 1.45	0.394	0.612	0.469	0.543
rs6960369	0.90	0.76 - 1.07	0.88	0.66 - 1.18	0.249	0.359	0.216	0.679
rs6968407	1.11	0.92 - 1.32	1.48	1.05 - 2.09	0.041	0.622	0.135	0.047
rs6969989	1.18	0.99 - 1.42	1.36	0.86 - 2.13	0.045	0.111	0.049	0.347
rs6972923	1.09	0.89 - 1.32	1.15	0.9 - 1.47	0.252	0.873	0.321	0.410
rs6973242	1.08	0.9 - 1.29	1.41	0.99 - 1.99	0.087	0.759	0.231	0.077
rs7807607	0.94	0.79 - 1.12	1.02	0.74 - 1.4	0.735	0.430	0.538	0.730
rs819436	0.92	0.77 - 1.1	1.18	0.83 - 1.67	0.990	0.228	0.539	0.226
rs9784998	0.95	0.8 - 1.12	0.72	0.5 - 1.03	0.119	0.933	0.321	0.088
rs997906	1.07	0.87 - 1.32	1.38	0.8 - 2.37	0.270	0.648	0.398	0.281

NOTE.- Heterozygotes (1/2) or minor allele homozygotes (2/2) compared with major allele homozygotes (1/1); OR, odds ratio; 95% C(onfidence) I(nterval); *P*-values for 4 genetic models: Rec(essive), Over(dominant), Dom(inant) and Allelic.

Table C.3C. UNPHASED analysis of pooled Gambian family- and case-control studies (SA).

SA	1/2 vs. 1/1		2/2 vs. 1/1					
Marker	OR	95% CI	OR	95% CI	Allelic	Over	Dom	Rec
def	1.09	0.73 - 1.62	3.56	0.69 - 18.38	0.385	0.803	0.561	0.114
G1439C	1.19	0.73 - 1.95	-	-	0.529	0.483	0.483	-
T1264G	1.09	0.59 - 2.02	-	-	0.563	0.902	0.718	0.083
rs10215288	0.91	0.73 - 1.14	1.18	0.81 - 1.73	0.893	0.26	0.627	0.244
rs10216184	1.10	0.87 - 1.37	1.44	0.9 - 2.32	0.159	0.658	0.292	0.176
rs10480808	1.19	0.93 - 1.52	1.18	0.88 - 1.59	0.255	0.338	0.15	0.749
rs1049654	0.96	0.76 - 1.2	1.13	0.83 - 1.53	0.564	0.376	0.937	0.26
rs10499857	0.98	0.79 - 1.21	0.89	0.56 - 1.43	0.688	0.951	0.787	0.656
rs1054516	0.89	0.72 - 1.1	0.98	0.64 - 1.5	0.449	0.271	0.311	0.843
rs1194178	0.88	0.7 - 1.1	1.06	0.63 - 1.77	0.495	0.239	0.325	0.647
rs1194180	0.94	0.76 - 1.18	1.32	0.89 - 1.96	0.539	0.347	0.92	0.106
rs11971740	1.08	0.87 - 1.33	1.01	0.7 - 1.46	0.696	0.462	0.524	0.827
rs12706912	0.91	0.74 - 1.13	0.74	0.52 - 1.04	0.093	0.861	0.224	0.134
rs13222488	0.99	0.8 - 1.24	1.27	0.83 - 1.94	0.526	0.71	0.846	0.248
rs13228738	1.14	0.9 - 1.44	1.00	0.51 - 1.96	0.375	0.256	0.286	0.846
rs1334511	0.98	0.79 - 1.23	1.02	0.72 - 1.45	0.991	0.836	0.919	0.86
rs1358337	1.03	0.83 - 1.28	1.02	0.74 - 1.39	0.881	0.822	0.82	0.98
rs1413659	0.92	0.74 - 1.15	1.30	0.87 - 1.95	0.668	0.267	0.764	0.116
rs1413661	0.90	0.73 - 1.11	0.93	0.64 - 1.37	0.443	0.377	0.35	0.958
rs1527463	1.01	0.78 - 1.3	1.51	0.75 - 3.05	0.552	0.902	0.791	0.241
rs1537593	1.11	0.87 - 1.43	1.10	0.81 - 1.48	0.551	0.545	0.396	0.919
rs1572230	0.81	0.62 - 1.07	0.61	0.21 - 1.81	0.091	0.163	0.112	0.453
rs17154155	0.85	0.67 - 1.07	1.00	0.75 - 1.34	0.897	0.1	0.29	0.346
rs17154246	0.99	0.8 - 1.22	0.83	0.54 - 1.28	0.523	0.899	0.747	0.389
rs1722501	0.86	0.69 - 1.07	1.16	0.8 - 1.69	0.827	0.097	0.326	0.206
rs1722502	0.88	0.71 - 1.1	0.96	0.66 - 1.41	0.457	0.267	0.299	0.865
rs1722505	0.89	0.72 - 1.1	0.98	0.67 - 1.43	0.494	0.258	0.313	0.789
rs17263407	1.13	0.89 - 1.43	1.16	0.63 - 2.14	0.313	0.355	0.3	0.767
rs17266080	1.03	0.8 - 1.33	1.41	0.68 - 2.95	0.528	0.901	0.681	0.369
rs1761661	0.88	0.71 - 1.09	1.00	0.69 - 1.45	0.527	0.207	0.294	0.665
rs1761662	0.88	0.71 - 1.1	0.89	0.54 - 1.49	0.297	0.292	0.255	0.847
rs1761663	0.96	0.76 - 1.22	1.12	0.84 - 1.49	0.477	0.365	0.989	0.25
rs1924	1.02	0.83 - 1.26	0.82	0.53 - 1.26	0.657	0.604	0.995	0.304
rs2366855	0.93	0.75 - 1.16	0.99	0.66 - 1.48	0.671	0.518	0.555	0.902
rs3173798	1.02	0.82 - 1.26	0.67	0.4 - 1.12	0.424	0.562	0.854	0.101
rs3173800	1.02	0.8 - 1.3	1.01	0.54 - 1.89	0.871	0.852	0.854	0.997

NOTE.- Heterozygotes (1/2) or minor allele homozygotes (2/2) compared with major allele homozygotes (1/1); OR, odds ratio; 95% C(onfidence) I(nterval); def, both CD36 deficiency alleles pooled; *P*-values for 4 genetic models: Rec(essive), Over(dominant), Dom(inant) and Allelic.

Table C.3C. UNPHASED analysis of pooled Gambian family- and case-control studies (SA) – *continued*.

SA	1/2 vs. 1/1		2/2 vs. 1/1					
Marker	OR	95% CI	OR	95% CI	Allelic	Over	Dom	Rec
rs3211810	1.04	0.82 - 1.32	0.93	0.44 - 1.96	0.852	0.716	0.777	0.807
rs3211817	1.20	0.9 - 1.58	0.66	0.22 - 1.96	0.46	0.177	0.285	0.374
rs3211821	0.98	0.79 - 1.22	0.94	0.61 - 1.44	0.763	0.927	0.817	0.793
rs3211822	0.99	0.79 - 1.23	1.06	0.77 - 1.45	0.811	0.77	0.996	0.661
rs3211828	1.04	0.84 - 1.29	0.73	0.46 - 1.15	0.56	0.415	0.95	0.126
rs3211842	1.01	0.81 - 1.25	0.91	0.66 - 1.26	0.653	0.733	0.902	0.505
rs3211849	0.83	0.66 - 1.05	0.97	0.73 - 1.3	0.698	0.079	0.191	0.414
rs3211850	0.98	0.76 - 1.26	1.62	0.82 - 3.2	0.59	0.721	0.905	0.151
rs3211867	1.11	0.89 - 1.39	0.76	0.53 - 1.08	0.427	0.07	0.66	0.033
rs3211869	1.03	0.83 - 1.27	0.78	0.46 - 1.31	0.713	0.636	0.983	0.309
rs3211870	1.17	0.92 - 1.49	1.02	0.75 - 1.39	0.84	0.131	0.289	0.428
rs3211883	0.92	0.74 - 1.15	0.91	0.65 - 1.26	0.454	0.6	0.433	0.75
rs3211886	1.02	0.79 - 1.33	1.03	0.47 - 2.24	0.866	0.877	0.865	0.963
rs3211888	0.99	0.71 - 1.4	0.67	0.19 - 2.37	0.754	0.969	0.883	0.533
rs3211897	1.26	0.87 - 1.83	-	-	0.49	0.231	0.336	0.205
rs3211899	0.63	0.46 - 0.86	1.49	0.51 - 4.35	0.018	0.002	0.006	0.251
rs3211909	0.89	0.69 - 1.14	0.94	0.5 - 1.77	0.41	0.364	0.361	0.969
rs3211911	0.88	0.59 - 1.33	-	-	0.372	0.585	0.467	0.251
rs3211913	1.08	0.87 - 1.33	0.86	0.62 - 1.21	0.712	0.248	0.714	0.213
rs3211923	0.96	0.62 - 1.49	-	-	0.726	0.896	0.807	0.426
rs3211928	0.98	0.78 - 1.23	0.84	0.5 - 1.41	0.614	0.99	0.757	0.52
rs3211930	0.92	0.62 - 1.34	-	-	0.502	0.663	0.575	0.358
rs3211931	1.02	0.82 - 1.27	1.05	0.66 - 1.67	0.803	0.884	0.823	0.87
rs3211942	0.94	0.7 - 1.28	0.53	0.12 - 2.35	0.502	0.76	0.614	0.41
rs4545029	0.83	0.66 - 1.03	0.93	0.6 - 1.46	0.198	0.096	0.107	0.891
rs4728191	0.87	0.67 - 1.11	0.59	0.24 - 1.43	0.13	0.329	0.192	0.291
rs6960369	1.11	0.89 - 1.37	1.11	0.78 - 1.59	0.413	0.454	0.348	0.821
rs6968407	0.97	0.78 - 1.2	1.36	0.89 - 2.07	0.485	0.474	0.953	0.112
rs6969989	1.13	0.91 - 1.41	0.96	0.55 - 1.67	0.477	0.243	0.313	0.668
rs6972923	0.95	0.75 - 1.2	1.09	0.81 - 1.46	0.606	0.356	0.883	0.315
rs6973242	0.99	0.8 - 1.23	1.25	0.82 - 1.91	0.561	0.675	0.889	0.259
rs7807607	0.85	0.69 - 1.05	1.03	0.71 - 1.5	0.467	0.107	0.201	0.511
rs819436	0.81	0.66 - 1.01	1.19	0.77 - 1.84	0.404	0.033	0.119	0.208
rs9784998	1.01	0.82 - 1.25	0.74	0.47 - 1.16	0.459	0.615	0.848	0.164
rs997906	0.94	0.72 - 1.22	0.88	0.42 - 1.84	0.594	0.684	0.617	0.797

NOTE.- Heterozygotes (1/2) or minor allele homozygotes (2/2) compared with major allele homozygotes (1/1) ; OR, odds ratio; 95% C(onfidence) I(nterval); *P*-values for 4 genetic models: Rec(essive), Over(dominant),, Dom(inant) and Allelic.

Table C.4. 54 SNPs genotyped and analysed in the HGDP-CEPH Diversity Panel.

SNP	Coord	Success
rs4731338	79779138	96.9%
rs304753	79820301	97.6%
rs304775	79841610	99.4%
rs1878314	79861553	91.3%
rs568125	79882207	97.8%
rs704871	79903613	95.9%
rs2944398	79924797	97.6%
rs6467192	79945734	95.2%
rs10216027	79968467	96.4%
rs2103134	79997284	88.8%
rs1931694	80009513	97.7%
rs1851937	80028981	97.2%
rs2065668	80058544	99.1%
rs1722505	80078625	94.8%
rs10499859	80096746	96.6%
rs13306227	80114006	98.5%
rs1527463	80114267	99.8%
rs3211810	80114953	99.8%
rs3211828	80117539	99.7%
rs3211842	80120572	95.2%
rs3765187	80123939	99.9%
rs5957	80130272	97.4%
rs3211917	80133237	99.2%
rs3211938	80138385	99.8%
rs3211949	80140136	95.9%
rs1803256	80141286	96.7%
rs7755	80144207	98.9%
rs1405745	80146176	100.0%
rs1527480	80155968	97.6%
rs12534038	80167036	97.4%
rs6958863	80177838	99.8%
rs2366858	80178558	97.3%
rs954409	80189667	98.7%
rs1880957	80192165	98.6%
rs1527467	80197742	99.8%
rs1852538	80202937	90.0%
rs1527470	80208765	90.6%
rs1405743	80216084	95.6%
rs10246081	80228534	97.1%
rs12669309	80238106	94.6%
rs3807103	80249157	96.6%
rs12673477	80268683	94.3%
rs6467405	80274281	97.4%
rs17154508	80289376	95.8%
rs10487878	80298261	99.1%
rs4731861	80309471	96.5%
rs7795137	80328382	97.4%
rs7784762	80334149	99.0%
rs7779029	80370048	95.4%
rs2057880	80387869	87.6%
rs17154598	80404725	96.4%
rs740210	80425673	94.8%
rs2367090	80447883	97.2%
rs701327	80474845	98.8%

NOTE.- SNP, dbSNP reference SNP identifier; Coord, chromosome 7 coordinate values from NCBI build 36/dbSNP 126; Genotyping success rate in the 974 individuals used for our population genetic analysis.

Table C.5. Sample sizes and demographics by study design and region of origin.

Study Site	Samples	Size	Male (%)	Mean Age (Months)	Ethnic Groups
Population-based studies					
Gambia	Cases	727	50.8	52.1	106 Fula, 183 Jola, 274 Mandinka, 80 Wolloff, 84 Other
	Controls	623	50.1	0	136 Fula, 89 Jola, 173 Mandinka, 94 Wolloff, 131 Other
Malawi	Cases	718	49.3	42.1	*
	Controls	405	52.2	0	
Kenya	Cases	708	55.1	30.3	108 Chonyi, 320 Giriama, 41 Kauma, 239 other
	Controls	902	51.6	0	183 Chonyi, 405 Giriama, 55 Kauma, 259 other
Ghana	Cases	792	57.4	18.5	435 Kasem, 305 Nankan, 28 Buli, 28 other
	Controls	806	56.3	20.5	508 Kasem, 269 Nankan, 15 Buli, 14 other
Family-based studies					
Gambia	Trios	1288	54.0	47.7	189 Fula, 216 Jola, 526 Mandinka, 174 Wolloff, 183 Other
Malawi	Trios	225	54.7	38.5	*
Kenya	Trios	234	48.1	29.8	51 Chonyi, 134 Giriama, 15 Kauma, 34 other

NOTE.- Apart from the Ghanaians, all controls are cord blood samples are therefore from children age zero. *Malawians in the study region are considered relatively ethnically homogenous; therefore no ethnicity information was gathered. Gender (% male) and mean age of probands are reported.

Table C.6A. *CD36* genotyping assay performance (Gambian case-control study).

	Controls					HWE	Cases				
Marker	1/1	1/2	2/2	Freq	Success	*P*	1/1	1/2	2/2	Freq	Success
G1439C	594	22	0	0.018	98.9%	0.493	706	17	1	0.013	99.6%
T1264G	595	24	0	0.019	99.4%	0.579	706	20	0	0.014	99.9%
rs10215288	317	247	56	0.290	99.5%	0.478	371	267	79	0.296	98.6%
rs10216184	383	207	29	0.214	99.4%	0.952	446	243	34	0.215	99.4%
rs10480808	189	284	148	0.467	99.7%	0.049	191	371	161	0.479	99.4%
rs1049654	251	279	88	0.368	99.2%	0.503	255	338	130	0.414	99.4%
rs10499857	376	213	31	0.222	99.5%	0.972	448	243	31	0.211	99.3%
rs1054516	380	209	31	0.219	99.5%	0.818	427	251	45	0.236	99.4%
rs1194178	409	191	21	0.188	99.7%	0.906	482	211	30	0.187	99.4%
rs1194180	353	223	47	0.254	100.0%	0.185	404	260	61	0.263	99.7%
rs11971740	298	267	54	0.303	99.4%	0.647	347	302	73	0.310	99.3%
rs12706912	251	279	90	0.370	99.5%	0.422	306	327	92	0.352	99.7%
rs13222488	343	250	28	0.246	99.7%	0.045	420	256	46	0.241	99.3%
rs13228738	442	163	16	0.157	99.7%	0.932	523	183	19	0.152	99.7%
rs1334511	302	241	62	0.302	97.1%	0.206	353	268	89	0.314	97.7%
rs1358337	212	300	107	0.415	99.4%	0.980	253	344	128	0.414	99.7%
rs1413659	348	215	46	0.252	97.8%	0.138	396	255	59	0.263	97.7%
rs1413661	326	253	41	0.270	99.5%	0.433	378	280	63	0.282	99.2%
rs1527463	474	134	13	0.129	99.7%	0.421	546	158	20	0.137	99.6%
rs1537593	150	309	162	0.510	99.7%	0.956	181	350	191	0.507	99.3%
rs1572230	478	130	13	0.126	99.7%	0.313	585	129	8	0.100	99.3%
rs17154155	171	313	136	0.472	99.5%	0.794	219	354	150	0.452	99.4%
rs17154246	373	204	45	0.236	99.8%	0.029	422	265	38	0.235	99.7%
rs1722501	329	243	48	0.273	99.5%	0.798	396	250	77	0.279	99.4%
rs1722502	320	248	44	0.275	98.2%	0.727	371	274	67	0.287	97.9%
rs1722505	325	253	41	0.271	99.4%	0.424	380	282	62	0.280	99.6%
rs17263407	416	180	20	0.179	98.9%	0.971	494	205	24	0.175	99.4%
rs17266080	466	143	11	0.133	99.5%	0.880	538	172	10	0.133	99.0%
rs1761661	318	260	43	0.279	99.7%	0.338	375	287	64	0.286	99.9%
rs1761662	395	205	22	0.200	99.8%	0.531	454	240	31	0.208	99.7%
rs1761663	156	318	148	0.494	99.8%	0.613	194	345	186	0.494	99.7%
rs1924	337	239	45	0.265	99.7%	0.832	400	277	46	0.255	99.4%
rs2366855	362	214	40	0.239	98.9%	0.317	425	245	50	0.240	99.0%
rs3173798	374	212	35	0.227	99.7%	0.556	454	240	28	0.205	99.3%
rs3173800	456	147	18	0.147	99.7%	0.192	530	174	21	0.149	99.7%
rs3211810	476	129	16	0.130	99.7%	0.068	533	170	21	0.146	99.6%

NOTE.- Genotype counts for major allele homozygotes (1/1), heterozygotes (1/2) and minor allele homozygotes (2/2); Freq, minor allele frequency; Success, percentage genotyping success; HWE, exact test Hardy-Weinberg equilibrium *P*-value (controls only).

Table C.6A. *CD36* genotyping assay performance (Gambian case-control study) – *continued.*

	Controls					HWE	Cases				
Marker	1/1	1/2	2/2	Freq	Success	*P*	1/1	1/2	2/2	Freq	Success
rs3211817	512	96	11	0.095	99.4%	0.022	596	125	3	0.090	99.6%
rs3211821	359	226	32	0.235	99.0%	0.708	409	258	45	0.244	97.9%
rs3211822	219	285	108	0.409	98.2%	0.395	258	324	135	0.414	98.6%
rs3211828	358	217	41	0.243	98.9%	0.345	429	260	31	0.224	99.0%
rs3211842	258	262	101	0.374	99.7%	0.017	302	338	85	0.350	99.7%
rs3211849	179	305	135	0.464	99.4%	0.859	197	355	170	0.481	99.3%
rs3211850	492	115	14	0.115	99.7%	0.036	541	166	17	0.138	99.6%
rs3211867	278	261	81	0.341	99.5%	0.130	307	329	79	0.341	98.3%
rs3211869	379	208	34	0.222	99.7%	0.497	456	239	28	0.204	99.4%
rs3211870	159	303	156	0.498	99.2%	0.673	184	363	175	0.494	99.3%
rs3211883	276	270	76	0.339	99.8%	0.470	326	303	95	0.340	99.6%
rs3211886	477	134	12	0.127	100.0%	0.577	563	143	16	0.121	99.3%
rs3211888	557	56	6	0.055	99.4%	0.005	667	55	3	0.042	99.7%
rs3211897	569	50	2	0.043	99.7%	0.733	656	69	0	0.048	99.7%
rs3211899	508	110	3	0.093	99.7%	0.353	607	110	8	0.087	99.7%
rs3211909	454	145	22	0.152	99.7%	0.025	524	180	18	0.150	99.3%
rs3211911	552	62	3	0.055	99.0%	0.612	664	57	3	0.044	99.6%
rs3211913	268	262	91	0.357	99.7%	0.049	296	345	83	0.353	99.6%
rs3211923	582	35	2	0.032	99.4%	0.234	691	32	0	0.022	99.4%
rs3211928	344	236	37	0.251	99.0%	0.742	364	288	41	0.267	95.3%
rs3211930	559	61	1	0.051	99.7%	0.905	674	51	1	0.037	99.9%
rs3211931	348	236	39	0.252	100.0%	0.966	378	296	51	0.274	99.7%
rs3211942	528	82	8	0.079	99.2%	0.044	595	118	7	0.092	99.0%
rs4545029	367	222	31	0.229	99.5%	0.799	446	233	44	0.222	99.4%
rs4728191	453	143	24	0.154	99.5%	0.006	564	148	13	0.120	99.7%
rs6960369	277	286	60	0.326	100.0%	0.294	321	314	85	0.336	99.0%
rs6968407	343	249	30	0.248	99.8%	0.087	420	255	52	0.247	100.0%
rs6969989	361	230	17	0.217	97.6%	0.007	433	252	19	0.206	96.8%
rs6972923	154	313	147	0.494	98.6%	0.669	193	338	183	0.493	98.2%
rs6973242	343	248	30	0.248	99.7%	0.094	422	257	48	0.243	100.0%
rs7807607	322	248	42	0.271	98.2%	0.591	386	269	62	0.274	98.6%
rs819436	389	205	28	0.210	99.8%	0.955	451	229	41	0.216	99.2%
rs9784998	370	213	37	0.231	99.5%	0.446	429	264	30	0.224	99.4%
rs997906	454	151	13	0.143	99.2%	0.959	535	164	19	0.141	98.8%

NOTE.- Genotype counts for major allele homozygotes (1/1), heterozygotes (1/2) and minor allele homozygotes (2/2); Freq, minor allele frequency; Success, percentage genotyping success; HWE, exact test Hardy-Weinberg equilibrium *P*-value (controls only).

Table C.6B. *CD36* genotyping assay performance (Gambian family study).

	Parents					Cases				
Marker	1/1	1/2	2/2	Freq	Success	1/1	1/2	2/2	Freq	Success
G1439C	2454	110	7	0.024	99.8%	1220	65	2	0.027	99.9%
T1264G	2496	77	1	0.015	99.9%	1248	36	1	0.015	99.8%
rs10215288	1249	1031	213	0.292	96.8%	643	507	110	0.288	97.8%
rs10216184	1714	766	87	0.183	99.7%	858	382	41	0.181	99.5%
rs10480808	751	1262	539	0.458	99.1%	371	629	284	0.466	99.7%
rs1049654	869	1248	435	0.415	99.1%	442	620	218	0.413	99.4%
rs10499857	1520	907	143	0.232	99.8%	744	475	67	0.237	99.8%
rs1054516	1462	969	143	0.244	99.9%	748	454	85	0.242	99.9%
rs1194178	1638	835	86	0.197	99.3%	830	392	56	0.197	99.2%
rs1194180	1390	993	182	0.265	99.6%	717	483	87	0.255	99.9%
rs11971740	1216	1087	248	0.310	99.0%	615	547	121	0.307	99.6%
rs12706912	1066	1162	342	0.359	99.8%	540	580	167	0.355	99.9%
rs13222488	1560	881	126	0.221	99.7%	780	430	75	0.226	99.8%
rs13228738	1942	591	41	0.131	99.9%	967	293	27	0.135	99.9%
rs1334511	1210	969	363	0.333	98.7%	610	482	190	0.336	99.5%
rs1358337	901	1246	424	0.407	99.8%	450	623	212	0.407	99.8%
rs1413659	1386	960	181	0.262	98.1%	713	479	86	0.255	99.2%
rs1413661	1295	1105	173	0.282	99.9%	669	511	108	0.282	100.0%
rs1527463	1993	529	46	0.121	99.7%	992	271	20	0.121	99.6%
rs1537593	692	1267	611	0.484	99.8%	346	628	313	0.487	99.9%
rs1572230	2038	489	36	0.109	99.5%	1034	234	16	0.104	99.7%
rs17154155	778	1254	541	0.454	99.9%	388	606	293	0.463	99.9%
rs17154246	1543	897	123	0.223	99.5%	785	432	70	0.222	99.9%
rs1722501	1349	1008	217	0.280	99.9%	673	520	94	0.275	99.9%
rs1722502	1265	1061	174	0.282	97.0%	661	499	107	0.281	98.4%
rs1722505	1303	1097	173	0.280	99.9%	673	504	106	0.279	99.6%
rs17263407	1851	636	49	0.145	98.4%	925	312	28	0.145	98.2%
rs17266080	1934	552	66	0.134	99.1%	942	301	36	0.146	99.3%
rs1761661	1285	1101	183	0.286	99.7%	664	511	111	0.285	99.8%
rs1761662	1595	872	100	0.209	99.7%	813	415	59	0.207	99.9%
rs1761663	683	1327	559	0.476	99.7%	350	653	283	0.474	99.8%
rs1924	1364	997	197	0.272	99.3%	693	493	98	0.268	99.7%
rs2366855	1416	981	141	0.249	98.5%	725	464	83	0.248	98.8%
rs3173798	1543	879	131	0.223	99.1%	771	438	74	0.228	99.6%
rs3173800	1922	598	47	0.135	99.7%	977	280	30	0.132	99.9%
rs3211810	1854	669	50	0.149	99.9%	919	348	19	0.150	99.8%

NOTE.- Genotype counts for major allele homozygotes (1/1), heterozygotes (1/2) and minor allele homozygotes (2/2); Freq, minor allele frequency; Success, percentage genotyping success; HWE, exact test Hardy-Weinberg equilibrium *P*-value (controls only).

Table C.6B. *CD36* genotyping assay performance (Gambian family study) – *continued.*

	Parents					Cases				
Marker	1/1	1/2	2/2	Freq	Success	1/1	1/2	2/2	Freq	Success
rs3211817	2159	394	17	0.083	99.8%	1076	198	14	0.088	100.0%
rs3211821	1427	947	171	0.253	98.8%	719	478	84	0.252	99.5%
rs3211822	959	1217	381	0.387	99.3%	464	630	180	0.389	98.9%
rs3211828	1456	961	154	0.247	99.8%	728	474	83	0.249	99.8%
rs3211842	1054	1170	330	0.358	99.1%	535	576	176	0.361	99.9%
rs3211849	702	1286	584	0.477	99.8%	356	631	296	0.477	99.6%
rs3211850	1929	596	50	0.135	100.0%	955	307	22	0.137	99.7%
rs3211867	1042	1157	325	0.358	98.0%	542	559	176	0.357	99.1%
rs3211869	1547	873	128	0.222	98.9%	775	434	71	0.225	99.4%
rs3211870	683	1276	606	0.485	99.6%	334	640	310	0.491	99.7%
rs3211883	1136	1137	294	0.336	99.7%	588	519	171	0.337	99.2%
rs3211886	2033	497	29	0.108	99.3%	1027	237	18	0.106	99.5%
rs3211888	2315	255	5	0.051	100.0%	1149	133	5	0.056	99.9%
rs3211897	2358	204	13	0.045	100.0%	1175	110	0	0.043	99.8%
rs3211899	2173	382	19	0.082	99.9%	1102	175	10	0.076	99.9%
rs3211909	1869	617	51	0.142	98.5%	930	323	23	0.145	99.1%
rs3211911	2338	189	9	0.041	98.4%	1181	92	3	0.038	99.1%
rs3211913	1104	1136	290	0.339	98.2%	570	573	138	0.331	99.5%
rs3211923	2440	112	2	0.023	99.1%	1225	61	1	0.024	99.9%
rs3211928	1384	833	119	0.229	90.7%	686	450	42	0.227	91.5%
rs3211930	2370	201	2	0.040	99.9%	1189	97	0	0.038	99.8%
rs3211931	1502	901	151	0.236	99.1%	744	480	56	0.231	99.4%
rs3211942	2240	309	7	0.063	99.2%	1137	141	6	0.060	99.7%
rs4545029	1528	913	131	0.228	99.8%	773	443	71	0.227	99.9%
rs4728191	1946	571	51	0.131	99.7%	982	284	20	0.126	99.8%
rs6960369	1102	1119	307	0.343	98.1%	560	563	153	0.341	99.1%
rs6968407	1562	874	126	0.220	99.5%	781	427	77	0.226	99.8%
rs6969989	1638	814	72	0.190	98.0%	819	405	43	0.194	98.4%
rs6972923	681	1309	554	0.475	98.8%	347	643	283	0.475	98.8%
rs6973242	1555	880	128	0.222	99.5%	779	431	76	0.227	99.8%
rs7807607	1288	1057	168	0.277	97.6%	661	497	106	0.280	98.1%
rs819436	1524	935	104	0.223	99.5%	774	436	76	0.229	99.8%
rs9784998	1448	974	150	0.248	99.8%	724	480	82	0.250	99.8%
rs997906	1953	518	36	0.118	97.3%	1001	258	17	0.114	99.1%

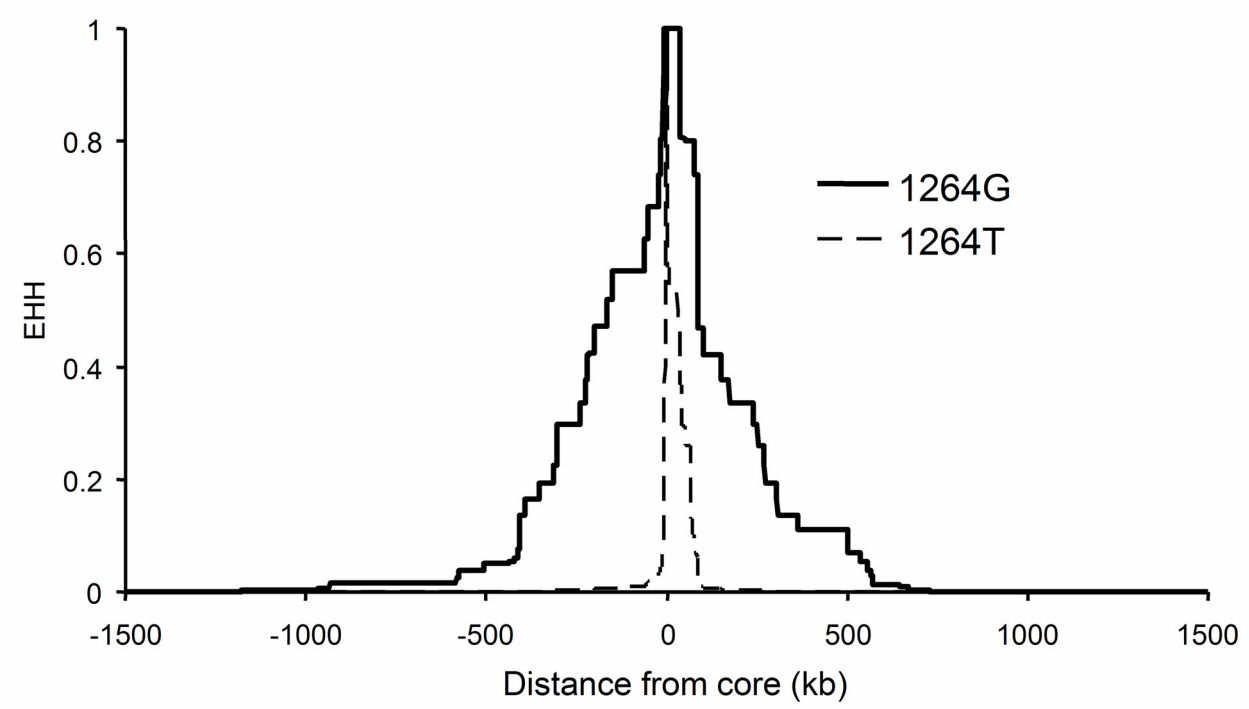
NOTE.- Genotype counts for major allele homozygotes (1/1), heterozygotes (1/2) and minor allele homozygotes (2/2); Freq, minor allele frequency; Success, percentage genotyping success; HWE, exact test Hardy-Weinberg equilibrium *P*-value (controls only).

Table C.6C. *CD36* genotyping assay performance (Malawian, Kenyan & Ghanaian studies).

Marker		1/1	1/2	2/2	Freq	Success	HWE
T1264G							
Family-based studies							
Malawi	Parents	375	84	3	0.097	98.7%	-
	Cases	188	31	8	0.104	97.0%	-
Kenya	Parents	371	67	8	0.093	99.1%	-
	Cases	182	42	0	0.094	99.6%	-
Population-based studies							
Malawi	Controls	334	64	4	0.090	99.3%	0.544
	Cases	603	111	4	0.083	100.0%	-
Kenya	Controls	754	140	4	0.082	99.6%	0.506
	Cases	568	126	4	0.096	98.6%	-
Ghana	Controls	553	183	15	0.142	93.2%	1.0
	Cases	561	174	22	0.144	95.6%	-
G1439C							
Population-based studies							
Ghana	Controls	689	2	0	0.001	85.7%	1.0
	Cases	727	2	0	0.001	92.1%	-

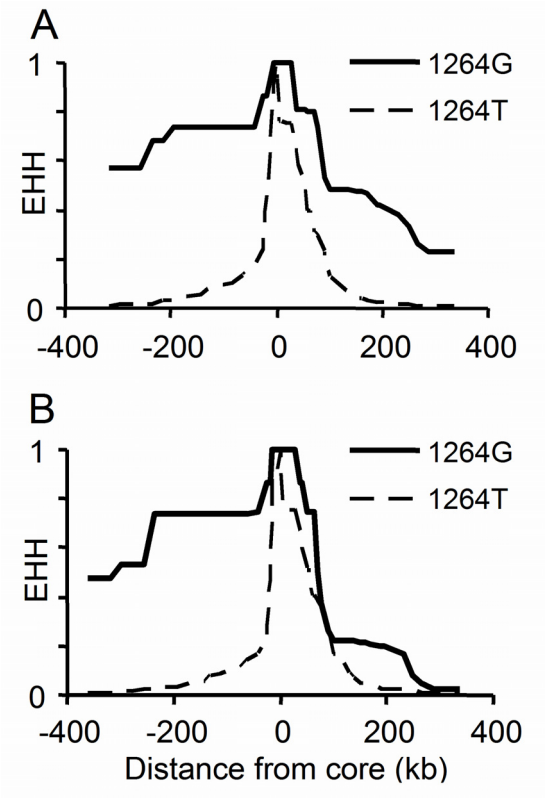
NOTE.- Genotype counts for major allele homozygotes (1/1), heterozygotes (1/2) and minor allele homozygotes (2/2); Freq, minor allele frequency; Success, percentage genotyping success; HWE, exact test Hardy-Weinberg equilibrium *P*-value (controls only).

Figure C.1. Distant decay in EHH surrounding T1264G in the HapMap Yoruba.



Breakdown of extended haplotype homozygosity (EHH) with distance on haplotypes partitioned by the alleles of T1264G (rs3211938; the 'core' SNP). Phased haplotype data (comprising 3627 polymorphic markers) from 30 Yoruba parent-offspring trios (120 parental haplotypes) downloaded from the HapMap project (HapMap release 21a, January 2007). Subtle differences between 1264G- and T-haplotypes are detectable over 1.1 Mb upstream and 0.7 Mb downstream of the core SNP.

Figure C.2. The effects of using offspring genotypes to resolve phase of parental haplotypes.



EHH decay surrounding T1264G in 30 HapMap Yoruba parent-offspring trios. We compared (A) parental haplotypes phased using offspring genotypes to help resolve phase, and (B) parental haplotypes phased without using offspring genotypes to help resolve phase. Note that the lack of offspring genotypes particularly impacts EHH values 3' to T1264G.

Hello Blake

FRY 2009

Bibliography

1. Singh, B., Kim Sung, L., Matusop, A., Radhakrishnan, A., Shamsul, S.S., Cox-Singh, J., Thomas, A. and Conway, D.J. (2004) A large focus of naturally acquired Plasmodium knowlesi infections in human beings. *Lancet,* 363, 1017-1024.
2. World Health Organization. (1999) The World Health Report 1999: Making a Difference. World Health Organization, Geneva.
3. Snow, R.W., Guerra, C.A., Noor, A.M., Myint, H.Y. and Hay, S.I. (2005) The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature,* 434, 214-217.
4. World Health Organization (2005) World Malaria Report 2005. World Health Organization, Geneva.
5. Greenwood, B. and Mutabingwa, T. (2002) Malaria in 2002. *Nature,* 415, 670-672.
6. Marsh, K., Forster, D., Waruiru, C., Mwangi, I., Winstanley, M., Marsh, V., Newton, C., Winstanley, P., Warn, P., Peshu, N. *et al.* (1995) Indicators of life-threatening malaria in African children. *N Engl J Med,* 332, 1399-1404.
7. World Health Organization (2000) Severe falciparum malaria. World Health Organization, Communicable Diseases Cluster. *Trans R Soc Trop Med Hyg,* 94 Suppl 1, S1-90.
8. Newton, C.R. and Krishna, S. (1998) Severe falciparum malaria in children: current understanding of pathophysiology and supportive treatment. *Pharmacol Ther,* 79, 1-53.
9. van Hensbroek, M.B., Palmer, A., Jaffar, S., Schneider, G. and Kwiatkowski, D. (1997) Residual neurologic sequelae after childhood cerebral malaria. *J Pediatr,* 131, 125-129.
10. John, C.C., Bangirana, P., Byarugaba, J., Opoka, R.O., Idro, R., Jurek, A.M., Wu, B. and Boivin, M.J. (2008) Cerebral malaria in children is associated with long-term cognitive impairment. *Pediatrics,* 122, e92-99.
11. Idro, R., Jenkins, N.E. and Newton, C.R. (2005) Pathogenesis, clinical features, and neurological outcome of cerebral malaria. *Lancet Neurol,* 4, 827-840.

12. Miller, L.H., Baruch, D.I., Marsh, K. and Doumbo, O.K. (2002) The pathogenic basis of malaria. *Nature,* 415, 673-679.
13. Taylor, T.E., Borgstein, A. and Molyneux, M.E. (1993) Acid-base status in paediatric Plasmodium falciparum malaria. *Q J Med,* 86, 99-109.
14. Flick, K. and Chen, Q. (2004) var genes, PfEMP1 and the human host. *Mol Biochem Parasitol,* 134, 3-9.
15. Urban, B.C., Ferguson, D.J., Pain, A., Willcox, N., Plebanski, M., Austyn, J.M. and Roberts, D.J. (1999) Plasmodium falciparum-infected erythrocytes modulate the maturation of dendritic cells. *Nature,* 400, 73-77.
16. MacPherson, G.G., Warrell, M.J., White, N.J., Looareesuwan, S. and Warrell, D.A. (1985) Human cerebral malaria. A quantitative ultrastructural analysis of parasitized erythrocyte sequestration. *Am J Pathol,* 119, 385-401.
17. Pongponratn, E., Turner, G.D., Day, N.P., Phu, N.H., Simpson, J.A., Stepniewska, K., Mai, N.T., Viriyavejakul, P., Looareesuwan, S., Hien, T.T. *et al.* (2003) An ultrastructural study of the brain in fatal Plasmodium falciparum malaria. *Am J Trop Med Hyg,* 69, 345-359.
18. Pino, P., Taoufiq, Z., Nitcheu, J., Vouldoukis, I. and Mazier, D. (2005) Blood-brain barrier breakdown during cerebral malaria: suicide or murder? *Thromb Haemost,* 94, 336-340.
19. Medana, I.M. and Turner, G.D. (2006) Human cerebral malaria and the blood-brain barrier. *Int J Parasitol,* 36, 555-568.
20. van der Heyde, H.C., Nolan, J., Combes, V., Gramaglia, I. and Grau, G.E. (2006) A unified hypothesis for the genesis of cerebral malaria: sequestration, inflammation and hemostasis leading to microcirculatory dysfunction. *Trends Parasitol,* 22, 503-508.
21. Petersen, L., Andersen, P.K. and Sorensen, T.I. (2008) Genetic and environmental effects on mortality before age 70 years. *Epidemiology,* 19, 472-476.
22. Dolo, A., Modiano, D., Maiga, B., Daou, M., Dolo, G., Guindo, H., Ba, M., Maiga, H., Coulibaly, D., Perlman, H. *et al.* (2005) Difference in susceptibility to malaria between two sympatric ethnic groups in Mali. *Am J Trop Med Hyg,* 72, 243-248.

23. Jepson, A., Sisay-Joof, F., Banya, W., Hassan-King, M., Frodsham, A., Bennett, S., Hill, A.V. and Whittle, H. (1997) Genetic linkage of mild malaria to the major histocompatibility complex in Gambian children: study of affected sibling pairs. *Bmj,* 315, 96-97.

24. Timmann, C., Evans, J.A., Konig, I.R., Kleensang, A., Ruschendorf, F., Lenzen, J., Sievertsen, J., Becker, C., Enuameh, Y., Kwakye, K.O. *et al.* (2007) Genome-wide linkage analysis of malaria infection intensity and mild disease. *PLoS Genet,* 3, e48.

25. Kwiatkowski, D.P. (2005) How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet,* 77, 171-192.

26. Mackinnon, M.J., Gunawardena, D.M., Rajakaruna, J., Weerasingha, S., Mendis, K.N. and Carter, R. (2000) Quantifying genetic and nongenetic contributions to malarial infection in a Sri Lankan population. *Proc Natl Acad Sci U S A,* 97, 12661-12666.

27. Mackinnon, M.J., Mwangi, T.W., Snow, R.W., Marsh, K. and Williams, T.N. (2005) Heritability of malaria in Africa. *PLoS Med,* 2, e340.

28. Tournamille, C., Colin, Y., Cartron, J.P. and Le Van Kim, C. (1995) Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet,* 10, 224-228.

29. Cattani, J.A., Gibson, F.D., Alpers, M.P. and Crane, G.G. (1987) Hereditary ovalocytosis and reduced susceptibility to malaria in Papua New Guinea. *Trans R Soc Trop Med Hyg,* 81, 705-709.

30. Guindo, A., Fairhurst, R.M., Doumbo, O.K., Wellems, T.E. and Diallo, D.A. (2007) X-linked G6PD deficiency protects hemizygous males but not heterozygous females against severe malaria. *PLoS Med,* 4, e66.

31. Ayi, K., Min-Oo, G., Serghides, L., Crockett, M., Kirby-Allen, M., Quirt, I., Gros, P. and Kain, K.C. (2008) Pyruvate kinase deficiency and malaria. *N Engl J Med,* 358, 1805-1810.

32. Hill, A.V., Allsopp, C.E., Kwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., Bennett, S., Brewster, D., McMichael, A.J. and Greenwood, B.M. (1991) Common west African HLA antigens are associated with protection from severe malaria. *Nature,* 352, 595-600.

33. McGuire, W., Knight, J.C., Hill, A.V., Allsopp, C.E., Greenwood, B.M. and Kwiatkowski, D. (1999) Severe malarial anemia and cerebral malaria are associated with different tumor necrosis factor promoter alleles. *J Infect Dis,* 179, 287-290.

34. Knight, J.C., Udalova, I., Hill, A.V., Greenwood, B.M., Peshu, N., Marsh, K. and Kwiatkowski, D. (1999) A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nat Genet,* 22, 145-150.

35. Koch, O., Awomoyi, A., Usen, S., Jallow, M., Richardson, A., Hull, J., Pinder, M., Newport, M. and Kwiatkowski, D. (2002) IFNGR1 gene promoter polymorphisms and susceptibility to cerebral malaria. *J Infect Dis,* 185, 1684-1687.

36. Luoni, G., Verra, F., Arca, B., Sirima, B.S., Troye-Blomberg, M., Coluzzi, M., Kwiatkowski, D. and Modiano, D. (2001) Antimalarial antibody levels and IL4 polymorphism in the Fulani of West Africa. *Genes Immun,* 2, 411-414.

37. Haldane, J.B. (1949) The rate of mutation of human genes. *Hereditas,* Suppl 35, 267-273.

38. Allison, A.C. (1954) Protection afforded by sickle-cell trait against subtertian malareal infection. *Br Med J,* 1, 290-294.

39. Lederberg, J. (1999) J. B. S. Haldane (1949) on infectious disease and evolution. *Genetics,* 153, 1-3.

40. Ackerman, H., Usen, S., Jallow, M., Sisay-Joof, F., Pinder, M. and Kwiatkowski, D.P. (2005) A comparison of case-control and family-based association methods: the example of sickle-cell and malaria. *Ann. Hum. Genet.,* 69, 559-565.

41. Modiano, D., Luoni, G., Sirima, B.S., Simpore, J., Verra, F., Konate, A., Rastrelli, E., Olivieri, A., Calissano, C., Paganotti, G.M. *et al.* (2001) Haemoglobin C protects against clinical Plasmodium falciparum malaria. *Nature,* 414, 305-308.

42. Chotivanich, K., Udomsangpetch, R., Pattanapanyasat, K., Chierakul, W., Simpson, J., Looareesuwan, S. and White, N. (2002) Hemoglobin E: a balanced polymorphism protective against high parasitemias and thus severe P falciparum malaria. *Blood,* 100, 1172-1176.

43. Zimmerman, P.A., Woolley, I., Masinde, G.L., Miller, S.M., McNamara, D.T., Hazlett, F., Mgone, C.S., Alpers, M.P., Genton, B., Boatin, B.A. *et al.* (1999)

Emergence of FY*A(null) in a Plasmodium vivax-endemic region of Papua New Guinea. *Proc Natl Acad Sci U S A,* 96, 13973-13977.

44. Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D. and Lander, E.S. (2006) Positive natural selection in the human lineage. *Science,* 312, 1614-1620.
45. Hughes, A.L. and Nei, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature,* 335, 167-170.
46. McDonald, J.H. and Kreitman, M. (1991) Adaptive protein evolution at the Adh locus in Drosophila. *Nature,* 351, 652-654.
47. Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics,* 105, 437-460.
48. Hudson, R.R., Kreitman, M. and Aguade, M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics,* 116, 153-159.
49. Fay, J.C. and Wu, C.I. (2000) Hitchhiking under positive Darwinian selection. *Genetics,* 155, 1405-1413.
50. Joy, D.A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Krettli, A.U., Ho, M., Wang, A., White, N.J., Suh, E. *et al.* (2003) Early origin and recent expansion of Plasmodium falciparum. *Science,* 300, 318-321.
51. Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J. *et al.* (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature,* 419, 832-837.
52. Voight, B.F., Kudaravalli, S., Wen, X. and Pritchard, J.K. (2006) A map of recent positive selection in the human genome. *PLoS Biol,* 4, e72.
53. Hanchard, N., Rockett, K., Spencer, C., Coop, G., Pinder, M., Jallow, M., Kimber, M., McVean, G., Mott, R. and Kwiatkowski, D.P. (2006) Screening for Recently Selected Alleles by Analysis of Human Haplotype Similarity. *Am J Hum Genet*, 153-159.
54. Sabeti, P.C. and Varilly, P. and Fry, B. and Lohmueller, J. and Hostetter, E. and Cotsapas, C. and Xie, X. and Byrne, E.H. and McCarroll, S.A. and Gaudet, R. *et al.*

(2007) Genome-wide detection and characterization of positive selection in human populations. *Nature,* 449, 913-918.

55. Williamson, S.H., Hubisz, M.J., Clark, A.G., Payseur, B.A., Bustamante, C.D. and Nielsen, R. (2007) Localizing recent adaptive evolution in the human genome. *PLoS Genet,* 3, e90.
56. Ohashi, J., Naka, I., Patarapotikul, J., Hananantachai, H., Brittenham, G., Looareesuwan, S., Clark, A.G. and Tokunaga, K. (2004) Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. *Am J Hum Genet,* 74, 1198-1208.
57. The International HapMap Consortium (2005) A haplotype map of the human genome. *Nature,* 437, 1299-1320.
58. Bignami, A. and Bastianelli, G. (1890) Osservazioni suile febbri malariche estive-autunnall. *La Riforma Medica,* 232, 1334–1335.
59. Trager, W., Rudzinska, M.A. and Bradbury, P.C. (1966) The fine structure of Plasmodium falciparum and its host erythrocytes in natural malarial infections in man. *Bull World Health Organ,* 35, 883-885.
60. Luse, S.A. and Miller, L.H. (1971) Plasmodium falciparum malaria. Ultrastructure of parasitized erythrocytes in cardiac vessels. *Am J Trop Med Hyg,* 20, 655-660.
61. Sherman, I.W., Crandall, I.E., Guthrie, N. and Land, K.M. (1995) The sticky secrets of sequestration. *Parasitol Today,* 11, 378-384.
62. Oh, S.S., Chishti, A.H., Palek, J. and Liu, S.C. (1997) Erythrocyte membrane alterations in Plasmodium falciparum malaria sequestration. *Curr Opin Hematol,* 4, 148-154.
63. Sherman, I.W., Eda, S. and Winograd, E. (2003) Cytoadherence and sequestration in Plasmodium falciparum: defining the ties that bind. *Microbes Infect,* 5, 897-909.
64. Day, K.P., Hayward, R.E., Smith, D. and Culvenor, J.G. (1998) CD36-dependent adhesion and knob expression of the transmission stages of Plasmodium falciparum is stage specific. *Mol Biochem Parasitol,* 93, 167-177.
65. Ruangjirachuporn, W., Afzelius, B.A., Paulie, S., Wahlgren, M., Berzins, K. and Perlmann, P. (1991) Cytoadherence of knobby and knobless Plasmodium falciparum-infected erythrocytes. *Parasitology,* 102 Pt 3, 325-334.

66. Howard, R.J., Barnwell, J.W. and Kao, V. (1983) Antigenic variation of Plasmodium knowlesi malaria: identification of the variant antigen on infected erythrocytes. *Proc Natl Acad Sci U S A,* 80, 4129-4133.

67. Leech, J.H., Barnwell, J.W., Miller, L.H. and Howard, R.J. (1984) Identification of a strain-specific malarial antigen exposed on the surface of Plasmodium falciparum-infected erythrocytes. *J Exp Med,* 159, 1567-1575.

68. Leech, J. (1991) Plasmodium falciparum cytoadherence. *Res Immunol,* 142, 681-686.

69. Baruch, D.I., Pasloske, B.L., Singh, H.B., Bi, X., Ma, X.C., Feldman, M., Taraschi, T.F. and Howard, R.J. (1995) Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell,* 82, 77-87.

70. Smith, J.D., Chitnis, C.E., Craig, A.G., Roberts, D.J., Hudson-Taylor, D.E., Peterson, D.S., Pinches, R., Newbold, C.I. and Miller, L.H. (1995) Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell,* 82, 101-110.

71. Su, X.Z., Heatwole, V.M., Wertheimer, S.P., Guinet, F., Herrfeldt, J.A., Peterson, D.S., Ravetch, J.A. and Wellems, T.E. (1995) The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. *Cell,* 82, 89-100.

72. Crabb, B.S., Cooke, B.M., Reeder, J.C., Waller, R.F., Caruana, S.R., Davern, K.M., Wickham, M.E., Brown, G.V., Coppel, R.L. and Cowman, A.F. (1997) Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. *Cell,* 89, 287-296.

73. Newbold, C., Craig, A., Kyes, S., Rowe, A., Fernandez-Reyes, D. and Fagan, T. (1999) Cytoadherence, pathogenesis and the infected red cell surface in Plasmodium falciparum. *Int J Parasitol,* 29, 927-937.

74. Smith, J.D., Subramanian, G., Gamain, B., Baruch, D.I. and Miller, L.H. (2000) Classification of adhesive domains in the Plasmodium falciparum erythrocyte membrane protein 1 family. *Mol Biochem Parasitol,* 110, 293-310.

75. Cheng, Q., Cloonan, N., Fischer, K., Thompson, J., Waine, G., Lanzer, M. and Saul, A. (1998) stevor and rif are Plasmodium falciparum multicopy gene families which potentially encode variant antigens. *Mol Biochem Parasitol,* 97, 161-176.

76. Blythe, J.E., Yam, X.Y., Kuss, C., Bozdech, Z., Holder, A.A., Marsh, K., Langhorne, J. and Preiser, P.R. (2008) Plasmodium falciparum STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. *Infect Immun,* 76, 3329-3336.

77. Gatton, M.L., Peters, J.M., Fowler, E.V. and Cheng, Q. (2003) Switching rates of Plasmodium falciparum var genes: faster than we thought? *Trends Parasitol,* 19, 202-208.

78. Horrocks, P., Pinches, R., Christodoulou, Z., Kyes, S.A. and Newbold, C.I. (2004) Variable var transition rates underlie antigenic variation in malaria. *Proc Natl Acad Sci U S A,* 101, 11129-11134.

79. Voss, T.S., Healer, J., Marty, A.J., Duffy, M.F., Thompson, J.K., Beeson, J.G., Reeder, J.C., Crabb, B.S. and Cowman, A.F. (2006) A var gene promoter controls allelic exclusion of virulence genes in Plasmodium falciparum malaria. *Nature,* 439, 1004-1008.

80. Bull, P.C., Lowe, B.S., Kortok, M., Molyneux, C.S., Newbold, C.I. and Marsh, K. (1998) Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nat Med,* 4, 358-360.

81. Vestergaard, L.S., Lusingu, J.P., Nielsen, M.A., Mmbando, B.P., Dodoo, D., Akanmori, B.D., Alifrangis, M., Bygbjerg, I.C., Lemnge, M.M., Staalsoe, T. *et al.* (2008) Differences in human antibody reactivity to Plasmodium falciparum variant surface antigens are dependent on age and malaria transmission intensity in northeastern Tanzania. *Infect Immun,* 76, 2706-2714.

82. Kraemer, S.M. and Smith, J.D. (2006) A family affair: var genes, PfEMP1 binding, and malaria disease. *Curr Opin Microbiol,* 9, 374-380.

83. Febbraio, M., Hajjar, D.P. and Silverstein, R.L. (2001) CD36: a class B scavenger receptor involved in angiogenesis, atherosclerosis, inflammation, and lipid metabolism. *J Clin Invest,* 108, 785-791.

84. Barnwell, J.W., Asch, A.S., Nachman, R.L., Yamaya, M., Aikawa, M. and Ingravallo, P. (1989) A human 88-kD membrane glycoprotein (CD36) functions in vitro as a receptor for a cytoadherence ligand on Plasmodium falciparum-infected erythrocytes. *J Clin Invest,* 84, 765-772.

85. Ockenhouse, C.F., Tandon, N.N., Magowan, C., Jamieson, G.A. and Chulay, J.D. (1989) Identification of a platelet membrane glycoprotein as a falciparum malaria sequestration receptor. *Science,* 243, 1469-1471.

86. Oquendo, P., Hundt, E., Lawler, J. and Seed, B. (1989) CD36 directly mediates cytoadherence of Plasmodium falciparum parasitized erythrocytes. *Cell,* 58, 95-101.

87. Baruch, D.I., Rogerson, S.J. and Cooke, B.M. (2002) Asexual blood stages of malaria antigens: cytoadherence. *Chem Immunol,* 80, 144-162.

88. Yamamoto, N., Ikeda, H., Tandon, N.N., Herman, J., Tomiyama, Y., Mitani, T., Sekiguchi, S., Lipsky, R., Kralisz, U. and Jamieson, G.A. (1990) A platelet membrane glycoprotein (GP) deficiency in healthy blood donors: Naka- platelets lack detectable GPIV (CD36). *Blood,* 76, 1698-1703.

89. Curtis, B.R. and Aster, R.H. (1996) Incidence of the Nak(a)-negative platelet phenotype in African Americans is similar to that of Asians. *Transfusion,* 36, 331-334.

90. Aitman, T.J., Cooper, L.D., Norsworthy, P.J., Wahid, F.N., Gray, J.K., Curtis, B.R., McKeigue, P.M., Kwiatkowski, D., Greenwood, B.M., Snow, R.W. *et al.* (2000) Malaria susceptibility and CD36 mutation. *Nature,* 405, 1015-1016.

91. Kashiwagi, H., Tomiyama, Y., Honda, S., Kosugi, S., Shiraga, M., Nagao, N., Sekiguchi, S., Kanayama, Y., Kurata, Y. and Matsuzawa, Y. (1995) Molecular basis of CD36 deficiency. Evidence that a 478C-->T substitution (proline90-->serine) in CD36 cDNA accounts for CD36 deficiency. *J Clin Invest,* 95, 1040-1046.

92. Pain, A., Urban, B.C., Kai, O., Casals-Pascual, C., Shafi, J., Marsh, K. and Roberts, D.J. (2001) A non-sense mutation in Cd36 gene is associated with protection from severe malaria. *Lancet,* 357, 1502-1503.

93. Omi, K., Ohashi, J., Patarapotikul, J., Hananantachai, H., Naka, I., Looareesuwan, S. and Tokunaga, K. (2003) CD36 polymorphism is associated with protection from cerebral malaria. *Am J Hum Genet,* 72, 364-374.

94. Amodu, O.K., Gbadegesin, R.A., Ralph, S.A., Adeyemo, A.A., Brenchley, P.E., Ayoola, O.O., Orimadegun, A.E., Akinsola, A.K., Olumese, P.E. and Omotade, O.O. (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.

95. Ayodo, G., Price, A.L., Keinan, A., Ajwang, A., Otieno, M.F., Orago, A.S., Patterson, N. and Reich, D. (2007) Combining evidence of natural selection with association analysis increases power to detect malaria-resistance variants. *Am J Hum Genet,* 81, 234-242.

96. Dietrich, J.B. (2002) The adhesion molecule ICAM-1 and its regulation in relation with the blood-brain barrier. *J Neuroimmunol,* 128, 58-68.

97. Hubbard, A.K. and Rothlein, R. (2000) Intercellular adhesion molecule-1 (ICAM-1) expression and cell signaling cascades. *Free Radic Biol Med,* 28, 1379-1386.

98. Armah, H., Dodoo, A.K., Wiredu, E.K., Stiles, J.K., Adjei, A.A., Gyasi, R.K. and Tettey, Y. (2005) High-level cerebellar expression of cytokines and adhesion molecules in fatal, paediatric, cerebral malaria. *Ann Trop Med Parasitol,* 99, 629-647.

99. Berendt, A.R., Simmons, D.L., Tansey, J., Newbold, C.I. and Marsh, K. (1989) Intercellular adhesion molecule-1 is an endothelial cell adhesion receptor for Plasmodium falciparum. *Nature,* 341, 57-59.

100. Smith, J.D., Craig, A.G., Kriek, N., Hudson-Taylor, D., Kyes, S., Fagen, T., Pinches, R., Baruch, D.I., Newbold, C.I. and Miller, L.H. (2000) Identification of a Plasmodium falciparum intercellular adhesion molecule-1 binding domain: a parasite adhesion trait implicated in cerebral malaria. *Proc Natl Acad Sci U S A,* 97, 1766-1771.

101. Smith, J.D., Kyes, S., Craig, A.G., Fagan, T., Hudson-Taylor, D., Miller, L.H., Baruch, D.I. and Newbold, C.I. (1998) Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte membrane protein-1 identifies a CD36 binding domain. *Mol Biochem Parasitol,* 97, 133-148.

102. Fernandez-Reyes, D., Craig, A.G., Kyes, S.A., Peshu, N., Snow, R.W., Berendt, A.R., Marsh, K. and Newbold, C.I. (1997) A high frequency African coding

polymorphism in the N-terminal domain of ICAM-1 predisposing to cerebral malaria in Kenya. *Hum Mol Genet,* 6, 1357-1360.

103. Tse, M.T., Chakrabarti, K., Gray, C., Chitnis, C.E. and Craig, A. (2004) Divergent binding sites on intercellular adhesion molecule-1 (ICAM-1) for variant Plasmodium falciparum isolates. *Mol Microbiol,* 51, 1039-1049.
104. Craig, A., Fernandez-Reyes, D., Mesri, M., McDowall, A., Altieri, D.C., Hogg, N. and Newbold, C. (2000) A functional analysis of a natural variant of intercellular adhesion molecule-1 (ICAM-1Kilifi). *Hum Mol Genet,* 9, 525-530.
105. Bellamy, R., Kwiatkowski, D. and Hill, A.V. (1998) Absence of an association between intercellular adhesion molecule 1, complement receptor 1 and interleukin 1 receptor antagonist gene polymorphisms and severe malaria in a West African population. *Trans R Soc Trop Med Hyg,* 92, 312-316.
106. Kun, J.F., Klabunde, J., Lell, B., Luckner, D., Alpers, M., May, J., Meyer, C. and Kremsner, P.G. (1999) Association of the ICAM-1Kilifi mutation with protection against severe malaria in Lambarene, Gabon. *Am J Trop Med Hyg,* 61, 776-779.
107. Ndiaye, R., Sakuntabhai, A., Casademont, I., Rogier, C., Tall, A., Trape, J.F., Spiegel, A., Dieye, A. and Julier, C. (2005) Genetic study of ICAM1 in clinical malaria in Senegal. *Tissue Antigens,* 65, 474-480.
108. Jenkins, N.E., Mwangi, T.W., Kortok, M., Marsh, K., Craig, A.G. and Williams, T.N. (2005) A polymorphism of intercellular adhesion molecule-1 is associated with a reduced incidence of nonmalarial febrile illness in kenyan children. *Clin Infect Dis,* 41, 1817-1819.
109. Ohashi, J., Naka, I., Patarapotikul, J., Hananantachai, H., Looareesuwan, S. and Tokunaga, K. (2001) Absence of association between the allele coding methionine at position 29 in the N-terminal domain of ICAM-1 (ICAM-1(Kilifi)) and severe malaria in the northwest of Thailand. *Jpn J Infect Dis,* 54, 114-116.
110. Rowe, J.A., Rogerson, S.J., Raza, A., Moulds, J.M., Kazatchkine, M.D., Marsh, K., Newbold, C.I., Atkinson, J.P. and Miller, L.H. (2000) Mapping of the region of complement receptor (CR) 1 required for Plasmodium falciparum rosetting and demonstration of the importance of CR1 in rosetting in field isolates. *J Immunol,* 165, 6341-6346.

111. Craig, M.L., Waitumbi, J.N. and Taylor, R.P. (2005) Processing of C3b-opsonized immune complexes bound to non-complement receptor 1 (CR1) sites on red cells: phagocytosis, transfer, and associations with CR1. *J Immunol,* 174, 3059-3066.

112. Krych-Goldberg, M., Moulds, J.M. and Atkinson, J.P. (2002) Human complement receptor type 1 (CR1) binds to a major malarial adhesin. *Trends Mol Med,* 8, 531-537.

113. Rowe, J.A., Moulds, J.M., Newbold, C.I. and Miller, L.H. (1997) P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature,* 388, 292-295.

114. Chen, Q., Barragan, A., Fernandez, V., Sundstrom, A., Schlichtherle, M., Sahlen, A., Carlson, J., Datta, S. and Wahlgren, M. (1998) Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) as the rosetting ligand of the malaria parasite P. falciparum. *J Exp Med,* 187, 15-23.

115. Treutiger, C.J., Hedlund, I., Helmby, H., Carlson, J., Jepson, A., Twumasi, P., Kwiatkowski, D., Greenwood, B.M. and Wahlgren, M. (1992) Rosette formation in Plasmodium falciparum isolates and anti-rosette activity of sera from Gambians with cerebral or uncomplicated malaria. *Am J Trop Med Hyg,* 46, 503-510.

116. Carlson, J., Helmby, H., Hill, A.V., Brewster, D., Greenwood, B.M. and Wahlgren, M. (1990) Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. *Lancet,* 336, 1457-1460.

117. Rowe, A., Obeiro, J., Newbold, C.I. and Marsh, K. (1995) Plasmodium falciparum rosetting is associated with malaria severity in Kenya. *Infect Immun,* 63, 2323-2326.

118. Heddini, A., Pettersson, F., Kai, O., Shafi, J., Obiero, J., Chen, Q., Barragan, A., Wahlgren, M. and Marsh, K. (2001) Fresh isolates from children with severe Plasmodium falciparum malaria bind to multiple receptors. *Infect Immun,* 69, 5849-5856.

119. al-Yaman, F., Genton, B., Mokela, D., Raiko, A., Kati, S., Rogerson, S., Reeder, J. and Alpers, M. (1995) Human cerebral malaria: lack of significant association between erythrocyte rosetting and disease severity. *Trans R Soc Trop Med Hyg,* 89, 55-58.

120. Nagayasu, E., Ito, M., Akaki, M., Nakano, Y., Kimura, M., Looareesuwan, S. and Aikawa, M. (2001) CR1 density polymorphism on erythrocytes of falciparum malaria patients in Thailand. *Am J Trop Med Hyg,* 64, 1-5.

121. Zimmerman, P.A., Fitness, J., Moulds, J.M., McNamara, D.T., Kasehagen, L.J., Rowe, J.A. and Hill, A.V. (2003) CR1 Knops blood group alleles are not associated with severe malaria in the Gambia. *Genes Immun,* 4, 368-373.

122. Cockburn, I.A., Mackinnon, M.J., O'Donnell, A., Allen, S.J., Moulds, J.M., Baisor, M., Bockarie, M., Reeder, J.C. and Rowe, J.A. (2004) A human complement receptor 1 polymorphism that reduces Plasmodium falciparum rosetting confers protection against severe malaria. *Proc Natl Acad Sci U S A,* 101, 272-277.

123. Newman, P.J. (1997) The biology of PECAM-1. *J Clin Invest,* 100, S25-29.

124. Gong, N. and Chatterjee, S. (2003) Platelet endothelial cell adhesion molecule in cell signaling and thrombosis. *Mol Cell Biochem,* 253, 151-158.

125. Treutiger, C.J., Heddini, A., Fernandez, V., Muller, W.A. and Wahlgren, M. (1997) PECAM-1/CD31, an endothelial receptor for binding Plasmodium falciparum-infected erythrocytes. *Nat Med,* 3, 1405-1408.

126. Casals-Pascual, C., Allen, S., Allen, A., Kai, O., Lowe, B., Pain, A. and Roberts, D.J. (2001) Short report: codon 125 polymorphism of CD31 and susceptibility to malaria. *Am J Trop Med Hyg,* 65, 736-737.

127. Kikuchi, M., Looareesuwan, S., Ubalee, R., Tasanor, O., Suzuki, F., Wattanagoon, Y., Na-Bangchang, K., Kimura, A., Aikawa, M. and Hirayama, K. (2001) Association of adhesion molecule PECAM-1/CD31 polymorphism with susceptibility to cerebral malaria in Thais. *Parasitol Int,* 50, 235-239.

128. Chen, H., Herndon, M.E. and Lawler, J. (2000) The cell biology of thrombospondin-1. *Matrix Biol,* 19, 597-614.

129. Ge, Y. and Elghetany, M.T. (2005) CD36: a multiligand molecule. *Lab Hematol,* 11, 31-37.

130. Streit, M., Velasco, P., Brown, L.F., Skobe, M., Richard, L., Riccardi, L., Lawler, J. and Detmar, M. (1999) Overexpression of thrombospondin-1 decreases angiogenesis and inhibits the growth of human cutaneous squamous cell carcinomas. *Am J Pathol,* 155, 441-452.

131. Roberts, D.D., Sherwood, J.A., Spitalnik, S.L., Panton, L.J., Howard, R.J., Dixit, V.M., Frazier, W.A., Miller, L.H. and Ginsburg, V. (1985) Thrombospondin binds falciparum malaria parasitized erythrocytes and may mediate cytoadherence. *Nature,* 318, 64-66.

132. Siano, J.P., Grady, K.K., Millet, P., Swerlick, R.A. and Wick, T.M. (1997) Plasmodium falciparum: soluble thrombospondin increases cytoadherence of parasitized erythrocytes to human microvascular endothelium under shear flow conditions. *Exp Parasitol,* 87, 69-72.

133. Rock, E.P., Roth, E.F., Jr., Rojas-Corona, R.R., Sherwood, J.A., Nagel, R.L., Howard, R.J. and Kaul, D.K. (1988) Thrombospondin mediates the cytoadherence of Plasmodium falciparum-infected red cells to vascular endothelium in shear flow conditions. *Blood,* 71, 71-75.

134. Eda, S., Lawler, J. and Sherman, I.W. (1999) Plasmodium falciparum-infected erythrocyte adhesion to the type 3 repeat domain of thrombospondin-1 is mediated by a modified band 3 protein. *Mol Biochem Parasitol,* 100, 195-205.

135. Geng, J.G., Chen, M. and Chou, K.C. (2004) P-selectin cell adhesion molecule in inflammation, thrombosis, cancer growth and metastasis. *Curr Med Chem,* 11, 2153-2160.

136. Ho, M., Schollaardt, T., Niu, X., Looareesuwan, S., Patel, K.D. and Kubes, P. (1998) Characterization of Plasmodium falciparum-infected erythrocyte and P-selectin interaction under flow conditions. *Blood,* 91, 4803-4809.

137. Senczuk, A.M., Reeder, J.C., Kosmala, M.M. and Ho, M. (2001) Plasmodium falciparum erythrocyte membrane protein 1 functions as a ligand for P-selectin. *Blood,* 98, 3132-3135.

138. Landsteiner, K. (1901) Über Agglutinationserscheinungen normalen menschlichen Blutes. *Wien Klin Wochenschr,* 14, 1132–1134.

139. Dean, L. (2005) Chapter 5: The ABO blood group. In *Blood Groups and Red Cell Antigen*. National Center for Biotechnology Information.

140. Raza, M.W., Blackwell, C.C., Molyneaux, P., James, V.S., Ogilvie, M.M., Inglis, J.M. and Weir, D.M. (1991) Association between secretor status and respiratory viral illness. *BMJ,* 303, 815-818.

141. Lindesmith, L., Moe, C., Marionneau, S., Ruvoen, N., Jiang, X., Lindblad, L., Stewart, P., LePendu, J. and Baric, R. (2003) Human susceptibility and resistance to Norwalk virus infection. *Nat. Med.,* 9, 548-553.

142. Blackwell, C.C., Dundas, S., James, V.S., Mackenzie, D.A., Braun, J.M., Alkout, A.M., Todd, W.T., Elton, R.A. and Weir, D.M. (2002) Blood group and susceptibility to disease caused by Escherichia coli O157. *J. Infect. Dis.,* 185, 393-396.

143. Boren, T., Falk, P., Roth, K.A., Larson, G. and Normark, S. (1993) Attachment of Helicobacter pylori to human gastric epithelium mediated by blood group antigens. *Science,* 262, 1892-1895.

144. Barragan, A., Kremsner, P.G., Wahlgren, M. and Carlson, J. (2000) Blood group A antigen is a coreceptor in Plasmodium falciparum rosetting. *Infect Immun,* 68, 2971-2975.

145. Chen, Q., Heddini, A., Barragan, A., Fernandez, V., Pearce, S.F. and Wahlgren, M. (2000) The semiconserved head structure of Plasmodium falciparum erythrocyte membrane protein 1 mediates binding to multiple independent host receptors. *J Exp Med,* 192, 1-10.

146. Uneke, C.J. (2007) Plasmodium falciparum malaria and ABO blood group: is there any relationship? *Parasitol. Res.,* 100, 759-765.

147. Cserti, C.M. and Dzik, W.H. (2007) The ABO blood group system and Plasmodium falciparum malaria. *Blood,* 110, 2250-2258.

148. Carlson, J. and Wahlgren, M. (1992) Plasmodium falciparum erythrocyte rosetting is mediated by promiscuous lectin-like interactions. *J. Exp. Med.,* 176, 1311-1317.

149. Rogerson, S.J., Chaiyaroj, S.C., Ng, K., Reeder, J.C. and Brown, G.V. (1995) Chondroitin sulfate A is a cell surface receptor for Plasmodium falciparum-infected erythrocytes. *J Exp Med,* 182, 15-20.

150. Fried, M. and Duffy, P.E. (1996) Adherence of Plasmodium falciparum to chondroitin sulfate A in the human placenta. *Science,* 272, 1502-1504.

151. Reeder, J.C., Cowman, A.F., Davern, K.M., Beeson, J.G., Thompson, J.K., Rogerson, S.J. and Brown, G.V. (1999) The adhesion of Plasmodium falciparum-

infected erythrocytes to chondroitin sulfate A is mediated by P. falciparum erythrocyte membrane protein 1. *Proc Natl Acad Sci U S A,* 96, 5198-5202.

152. Buffet, P.A., Gamain, B., Scheidig, C., Baruch, D., Smith, J.D., Hernandez-Rivas, R., Pouvelle, B., Oishi, S., Fujii, N., Fusai, T. *et al.* (1999) Plasmodium falciparum domain mediating adhesion to chondroitin sulfate A: a receptor for human placental infection. *Proc Natl Acad Sci U S A,* 96, 12743-12748.
153. Gamain, B., Gratepanche, S., Miller, L.H. and Baruch, D.I. (2002) Molecular basis for the dichotomy in Plasmodium falciparum adhesion to CD36 and chondroitin sulfate A. *Proc Natl Acad Sci U S A,* 99, 10020-10024.
154. Guyatt, H.L. and Snow, R.W. (2004) Impact of malaria during pregnancy on low birth weight in sub-Saharan Africa. *Clin Microbiol Rev,* 17, 760-769, table of contents.
155. Resende, M., Nielsen, M.A., Dahlback, M., Ditlev, S.B., Andersen, P., Sander, A.F., Ndam, N.T., Theander, T.G. and Salanti, A. (2008) Identification of glycosaminoglycan binding regions in the Plasmodium falciparum encoded placental sequestration ligand, VAR2CSA. *Malar J,* 7, 104.
156. Viebig, N.K., Gamain, B., Scheidig, C., Lepolard, C., Przyborski, J., Lanzer, M., Gysin, J. and Scherf, A. (2005) A single member of the Plasmodium falciparum var multigene family determines cytoadhesion to the placental receptor chondroitin sulphate A. *EMBO Rep,* 6, 775-781.
157. Beeson, J.G., Rogerson, S.J., Cooke, B.M., Reeder, J.C., Chai, W., Lawson, A.M., Molyneux, M.E. and Brown, G.V. (2000) Adhesion of Plasmodium falciparum-infected erythrocytes to hyaluronic acid in placental malaria. *Nat Med,* 6, 86-90.
158. Rasti, N., Namusoke, F., Chene, A., Chen, Q., Staalsoe, T., Klinkert, M.Q., Mirembe, F., Kironde, F. and Wahlgren, M. (2006) Nonimmune immunoglobulin binding and multiple adhesion characterize Plasmodium falciparum-infected erythrocytes of placental origin. *Proc Natl Acad Sci U S A,* 103, 13795-13800.
159. Barragan, A., Fernandez, V., Chen, Q., von Euler, A., Wahlgren, M. and Spillmann, D. (2000) The duffy-binding-like domain 1 of Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) is a heparan sulfate ligand that requires 12 mers for binding. *Blood,* 95, 3594-3599.

160. Flick, K., Scholander, C., Chen, Q., Fernandez, V., Pouvelle, B., Gysin, J. and Wahlgren, M. (2001) Role of nonimmune IgG bound to PfEMP1 in placental malaria. *Science,* 293, 2098-2100.

161. Rowe, J.A., Shafi, J., Kai, O.K., Marsh, K. and Raza, A. (2002) Nonimmune IgM, but not IgG binds to the surface of Plasmodium falciparum-infected erythrocytes and correlates with rosetting and severe malaria. *Am J Trop Med Hyg,* 66, 692-699.

162. Creasey, A.M., Staalsoe, T., Raza, A., Arnot, D.E. and Rowe, J.A. (2003) Nonspecific immunoglobulin M binding and chondroitin sulfate A binding are linked phenotypes of Plasmodium falciparum isolates implicated in malaria during pregnancy. *Infect Immun,* 71, 4767-4771.

163. Semblat, J.P., Raza, A., Kyes, S.A. and Rowe, J.A. (2006) Identification of Plasmodium falciparum var1CSA and var2CSA domains that bind IgM natural antibodies. *Mol Biochem Parasitol,* 146, 192-197.

164. Ockenhouse, C.F., Tegoshi, T., Maeno, Y., Benjamin, C., Ho, M., Kan, K.E., Thway, Y., Win, K., Aikawa, M. and Lobb, R.R. (1992) Human vascular endothelial cell adhesion receptors for Plasmodium falciparum-infected erythrocytes: roles for endothelial leukocyte adhesion molecule 1 and vascular cell adhesion molecule 1. *J Exp Med,* 176, 1183-1189.

165. Udomsangpetch, R., Taylor, B.J., Looareesuwan, S., White, N.J., Elliott, J.F. and Ho, M. (1996) Receptor specificity of clinical Plasmodium falciparum isolates: nonadherence to cell-bound E-selectin and vascular cell adhesion molecule-1. *Blood,* 88, 2754-2760.

166. Siano, J.P., Grady, K.K., Millet, P. and Wick, T.M. (1998) Short report: Plasmodium falciparum: cytoadherence to alpha(v)beta3 on human microvascular endothelial cells. *Am J Trop Med Hyg,* 59, 77-79.

167. Mota, M.M., Jarra, W., Hirst, E., Patnaik, P.K. and Holder, A.A. (2000) Plasmodium chabaudi-infected erythrocytes adhere to CD36 and bind to microvascular endothelial cells in an organ-specific way. *Infect Immun,* 68, 4135-4144.

168. Franke-Fayard, B., Janse, C.J., Cunha-Rodrigues, M., Ramesar, J., Buscher, P., Que, I., Lowik, C., Voshol, P.J., den Boer, M.A., van Duinen, S.G. *et al.* (2005) Murine

malaria parasite sequestration: CD36 is the major receptor, but cerebral pathology is unlinked to sequestration. *Proc Natl Acad Sci U S A,* 102, 11468-11473.

169. Serghides, L., Crandall, I., Hull, E. and Kain, K.C. (1998) The Plasmodium falciparum-CD36 interaction is modified by a single amino acid substitution in CD36. *Blood,* 92, 1814-1819.

170. Greenwalt, D.E., Scheck, S.H. and Rhinehart-Jones, T. (1995) Heart CD36 expression is increased in murine models of diabetes and in mice fed a high fat diet. *J Clin Invest,* 96, 1382-1388.

171. Bauer, P.R., Van Der Heyde, H.C., Sun, G., Specian, R.D. and Granger, D.N. (2002) Regulation of endothelial cell adhesion molecule expression in an experimental model of cerebral malaria. *Microcirculation,* 9, 463-470.

172. Combes, V., Rosenkranz, A.R., Redard, M., Pizzolato, G., Lepidi, H., Vestweber, D., Mayadas, T.N. and Grau, G.E. (2004) Pathogenic role of P-selectin in experimental cerebral malaria: importance of the endothelial compartment. *Am J Pathol,* 164, 781-786.

173. Sun, G., Chang, W.L., Li, J., Berney, S.M., Kimpel, D. and van der Heyde, H.C. (2003) Inhibition of platelet adherence to brain microvasculature protects against severe Plasmodium berghei malaria. *Infect Immun,* 71, 6553-6561.

174. Li, J., Chang, W.L., Sun, G., Chen, H.L., Specian, R.D., Berney, S.M., Kimpel, D., Granger, D.N. and van der Heyde, H.C. (2003) Intercellular adhesion molecule 1 is important for the development of severe experimental malaria but is not required for leukocyte adhesion in the brain. *J Investig Med,* 51, 128-140.

175. Beiguelman, B., Alves, F.P., Moura, M.M., Engracia, V., Nunes, A.C., Heckmann, M.I., Ferreira, R.G., da Silva, L.H., Camargo, E.P. and Krieger, H. (2003) The association of genetic markers and malaria infection in the Brazilian Western Amazonian region. *Mem. Inst. Oswaldo. Cruz.,* 98, 455-460.

176. Fischer, P.R. and Boone, P. (1998) Short report: severe malaria associated with blood group. *Am. J. Trop. Med. Hyg.,* 58, 122-123.

177. Lell, B., May, J., Schmidt-Ott, R.J., Lehman, L.G., Luckner, D., Greve, B., Matousek, P., Schmid, D., Herbich, K., Mockenhaupt, F.P. *et al.* (1999) The role of

red blood cell polymorphisms in resistance and susceptibility to malaria. *Clin. Infect. Dis.,* 28, 794-799.

178. Migot-Nabias, F., Mombo, L.E., Luty, A.J., Dubois, B., Nabias, R., Bisseye, C., Millet, P., Lu, C.Y. and Deloron, P. (2000) Human genetic factors related to susceptibility to mild malaria in Gabon. *Genes Immun.,* 1, 435-441.
179. Pant, C.S., Gupta, D.K., Sharma, R.C., Gautam, A.S. and Bhatt, R.M. (1992) Frequency of ABO blood groups, sickle-cell haemoglobin, G-6-PD deficiency and their relation with malaria in scheduled castes and scheduled tribes of Kheda District, Gujarat. *Indian J. Malariol.,* 29, 235-239.
180. Pathirana, S.L., Alles, H.K., Bandara, S., Phone-Kyaw, M., Perera, M.K., Wickremasinghe, A.R., Mendis, K.N. and Handunnetti, S.M. (2005) ABO-blood-group types and protection against severe, Plasmodium falciparum malaria. *Ann. Trop. Med. Parasitol.,* 99, 119-124.
181. Singh, N., Shukla, M.M., Uniyal, V.P. and Sharma, V.P. (1995) ABO blood groups among malaria cases from district Mandla, Madhya Pradesh. *Indian J. Malariol.,* 32, 59-63.
182. Martin, S.K., Miller, L.H., Hicks, C.U., David-West, A., Ugbode, C. and Deane, M. (1979) Frequency of blood group antigens in Nigerian children with falciparum malaria. *Trans. R. Soc. Trop. Med. Hyg.,* 73, 216-218.
183. Kassim, O.O. and Ejezie, G.C. (1982) ABO blood groups in malaria and schistosomiasis haematobium. *Acta Trop.,* 39, 179-184.
184. Bayoumi, R.A., Bashir, A.H. and Abdulhadi, N.H. (1986) Resistance to falciparum malaria among adults in central Sudan. *Am. J. Trop. Med. Hyg.,* 35, 45-55.
185. Akinboye, D.O. and Ogunrinade, A.F. (1987) Malaria and loaisis among blood donors at Ibadan, Nigeria. *Trans. R. Soc. Trop. Med. Hyg.,* 81, 398-399.
186. Thakur, A. and Verma, I.C. (1992) Malaria and ABO blood groups. *Indian J. Malariol.,* 29, 241-244.
187. Montoya, F., Restrepo, M., Montoya, A.E. and Rojas, W. (1994) Blood groups and malaria. *Rev. Inst. Med. Trop. Sao. Paulo.,* 36, 33-38.
188. Flick, K. and Chen, Q. (2004) var genes, PfEMP1 and the human host. *Mol. Biochem. Parasitol.,* 134, 3-9.

189. Baruch, D.I., Ma, X.C., Singh, H.B., Bi, X., Pasloske, B.L. and Howard, R.J. (1997) Identification of a region of PfEMP1 that mediates adherence of Plasmodium falciparum infected erythrocytes to CD36: conserved function with variant sequence. *Blood,* 90, 3766-3775.

190. Smith, J.D., Craig, A.G., Kriek, N., Hudson-Taylor, D., Kyes, S., Fagen, T., Pinches, R., Baruch, D.I., Newbold, C.I. and Miller, L.H. (2000) Identification of a Plasmodium falciparum intercellular adhesion molecule-1 binding domain: a parasite adhesion trait implicated in cerebral malaria. *Proc. Natl. Acad. Sci. USA,* 97, 1766-1771.

191. Barragan, A., Kremsner, P.G., Wahlgren, M. and Carlson, J. (2000) Blood group A antigen is a coreceptor in Plasmodium falciparum rosetting. *Infect. Immun.,* 68, 2971-2975.

192. Chen, Q., Heddini, A., Barragan, A., Fernandez, V., Pearce, S.F. and Wahlgren, M. (2000) The semiconserved head structure of Plasmodium falciparum erythrocyte membrane protein 1 mediates binding to multiple independent host receptors. *J. Exp. Med.,* 192, 1-10.

193. Newbold, C., Warn, P., Black, G., Berendt, A., Craig, A., Snow, B., Msobo, M., Peshu, N. and Marsh, K. (1997) Receptor-specific adhesion and clinical disease in Plasmodium falciparum. *Am. J. Trop. Med. Hyg.,* 57, 389-398.

194. Heddini, A., Pettersson, F., Kai, O., Shafi, J., Obiero, J., Chen, Q., Barragan, A., Wahlgren, M. and Marsh, K. (2001) Fresh isolates from children with severe Plasmodium falciparum malaria bind to multiple receptors. *Infect. Immun.,* 69, 5849-5856.

195. Rowe, J.A., Handel, I.G., Thera, M.A., Deans, A.-M., Lyke, K.E., Koné, A., Diallo, D.A., Raza, A., Kai, O., Marsh, K. *et al.* (2007) Blood group O protects against severe Plasmodium falciparum malaria through the mechanism of reduced rosetting. *Proc. Natl. Acad. Sci. USA,* 104, 17471-17476.

196. Udomsangpetch, R., Todd, J., Carlson, J. and Greenwood, B.M. (1993) The effects of hemoglobin genotype and ABO blood group on the formation of rosettes by Plasmodium falciparum-infected red blood cells. *Am. J. Trop. Med. Hyg.,* 48, 149-153.

197. Rowe, A., Obeiro, J., Newbold, C.I. and Marsh, K. (1995) Plasmodium falciparum rosetting is associated with malaria severity in Kenya. *Infect. Immun.,* 63, 2323-2326.

198. Yip, S.P. (2002) Sequence variation at the human ABO locus. *Ann. Hum. Genet.,* 66, 1-27.

199. Yamamoto, F., Clausen, H., White, T., Marken, J. and Hakomori, S. (1990) Molecular genetic basis of the histo-blood group ABO system. *Nature,* 345, 229-233.

200. Franco, R.F., Simoes, B.P. and Zago, M.A. (1995) Relative frequencies of the two O alleles of the histo-blood ABH system in different racial groups. *Vox. Sang.,* 69, 50-52.

201. Yamamoto, F. and Hakomori, S. (1990) Sugar-nucleotide donor specificity of histo-blood group A and B transferases is based on amino acid substitutions. *J. Biol. Chem.,* 265, 19257-19262.

202. Weinberg, C.R., Wilcox, A.J. and Lie, R.T. (1998) A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *Am. J. Hum. Genet.,* 62, 969-978.

203. Spielman, R.S. and Ewens, W.J. (1996) The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.,* 59, 983-989.

204. Poujol-Robert, A., Boelle, P.Y., Wendum, D., Poupon, R. and Robert, A. (2006) Association between ABO blood group and fibrosis severity in chronic hepatitis C infection. *Dig. Dis. Sci.,* 51, 1633-1636.

205. Adalsteinsson, S. (1985) Possible changes in the frequency of the human ABO blood groups in Iceland due to smallpox epidemics selection. *Ann. Hum. Genet.,* 49, 275-281.

206. Doughty, B.R. (1977) The changes in ABO blood group frequency within a mediaeval English population. *Med. Lab. Sci.,* 34, 351-354.

207. Kermarrec, N., Roubinet, F., Apoil, P.A. and Blancher, A. (1999) Comparison of allele O sequences of the human and non-human primate ABO system. *Immunogenetics,* 49, 517-526.

208. Akey, J.M., Eberle, M.A., Rieder, M.J., Carlson, C.S., Shriver, M.D., Nickerson, D.A. and Kruglyak, L. (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol.,* 2, e286.

209. Hamblin, M.T. and Di Rienzo, A. (2000) Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet.,* 66, 1669-1679.

210. Kominato, Y., Hata, Y., Takizawa, H., Tsuchiya, T., Tsukada, J. and Yamamoto, F. (1999) Expression of human histo-blood group ABO genes is dependent upon DNA methylation of the promoter region. *J. Biol. Chem.,* 274, 37240-37250.

211. Gao, S., Worm, J., Guldberg, P., Eiberg, H., Krogdahl, A., Liu, C.J., Reibel, J. and Dabelsteen, E. (2004) Genetic and epigenetic alterations of the blood group ABO gene in oral squamous cell carcinoma. *Int. J. Cancer,* 109, 230-237.

212. Dobrovic, A., O'Keefe, D., Sage, R.E. and Batchelder, E. (1993) Imprinting and loss of ABO antigens in leukemia. *Blood,* 82, 1684-1685.

213. Bull, P.C. and Marsh, K. (2002) The role of antibodies to Plasmodium falciparum-infected-erythrocyte surface antigens in naturally acquired immunity to malaria. *Trends Microbiol.,* 10, 55-58.

214. Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A. (1996) *The History and Geography of Human Genes.* Abridged edition ed. Princeton University Press.

215. Marsh, K., Forster, D., Waruiru, C., Mwangi, I., Winstanley, M., Marsh, V., Newton, C., Winstanley, P., Warn, P., Peshu, N. *et al.* (1995) Indicators of life-threatening malaria in African children. *N. Engl. J. Med.,* 332, 1399-1404.

216. (1990) Severe and complicated malaria. World Health Organization, Division of Control of Tropical Diseases. *Trans. R. Soc. Trop. Med. Hyg.,* 84 Suppl 2, 1-65.

217. Purcell, S., Cherny, S.S. and Sham, P.C. (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics,* 19, 149-150.

218. Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W. and Arnheim, N. (1992) Whole genome amplification from a single cell: implications for genetic analysis. *Proc. Natl. Acad. Sci. USA,* 89, 5847-5851.

219. Gonzalez, J.M., Portillo, M.C. and Saiz-Jimenez, C. (2005) Multiple displacement amplification as a pre-polymerase chain reaction (pre-PCR) to process difficult to amplify samples and low copy number sequences from natural environments. *Environ. Microbiol.,* 7, 1024-1028.

220. Ross, P., Hall, L., Smirnov, I. and Haff, L. (1998) High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat. Biotechnol.,* 16, 1347-1351.

221. Cordell, H.J., Barratt, B.J. and Clayton, D.G. (2004) Case/pseudocontrol analysis in genetic association studies: A unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. *Genet. Epidemiol.,* 26, 167-185.

222. Dudbridge, F. (2003) Pedigree disequilibrium tests for multilocus haplotypes. *Genet. Epidemiol.,* 25, 115-121.

223. Dudbridge, F. (2006) UNPHASED user guide. *Technical Report 2006/5.* MRC Biostatistics Unit, Cambridge, UK.

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

225. Stephens, M. and Donnelly, P. (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.,* 73, 1162-1169.

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

227. Berendt, A.R., Tumer, G.D. and Newbold, C.I. (1994) Cerebral malaria: the sequestration hypothesis. *Parasitol Today,* 10, 412-414.

228. Chen, Q., Schlichtherle, M. and Wahlgren, M. (2000) Molecular aspects of severe malaria. *Clin Microbiol Rev,* 13, 439-450.

229. Ockenhouse, C.F., Betageri, R., Springer, T.A. and Staunton, D.E. (1992) Plasmodium falciparum-infected erythrocytes bind ICAM-1 at a site distinct from LFA-1, Mac-1, and human rhinovirus. *Cell,* 68, 63-69.

230. Porta, J., Carota, A., Pizzolato, G.P., Wildi, E., Widmer, M.C., Margairaz, C. and Grau, G.E. (1993) Immunopathological changes in human cerebral malaria. *Clin Neuropathol,* 12, 142-146.

231. Turner, G.D., Morrison, H., Jones, M., Davis, T.M., Looareesuwan, S., Buley, I.D., Gatter, K.C., Newbold, C.I., Pukritayakamee, S., Nagachinta, B. *et al.* (1994) An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. *Am J Pathol,* 145, 1057-1069.

232. Silamut, K., Phu, N.H., Whitty, C., Turner, G.D., Louwrier, K., Mai, N.T., Simpson, J.A., Hien, T.T. and White, N.J. (1999) A quantitative analysis of the microvascular sequestration of malaria parasites in the human brain. *Am J Pathol,* 155, 395-410.

233. Rogerson, S.J., Tembenu, R., Dobano, C., Plitt, S., Taylor, T.E. and Molyneux, M.E. (1999) Cytoadherence characteristics of Plasmodium falciparum-infected erythrocytes from Malawian children with severe and uncomplicated malaria. *Am J Trop Med Hyg,* 61, 467-472.

234. Marlin, S.D. and Springer, T.A. (1987) Purified intercellular adhesion molecule-1 (ICAM-1) is a ligand for lymphocyte function-associated antigen 1 (LFA-1). *Cell,* 51, 813-819.

235. Baratin, M., Roetynck, S., Pouvelle, B., Lemmers, C., Viebig, N.K., Johansson, S., Bierling, P., Scherf, A., Gysin, J., Vivier, E. *et al.* (2007) Dissection of the Role of PfEMP1 and ICAM-1 in the Sensing of Plasmodium falciparum-Infected Erythrocytes by Natural Killer Cells. *PLoS ONE,* 2, e228.

236. Languino, L.R., Plescia, J., Duperray, A., Brian, A.A., Plow, E.F., Geltosky, J.E. and Altieri, D.C. (1993) Fibrinogen mediates leukocyte adhesion to vascular endothelium through an ICAM-1-dependent pathway. *Cell,* 73, 1423-1434.

237. Diamond, M.S., Staunton, D.E., de Fougerolles, A.R., Stacker, S.A., Garcia-Aguilar, J., Hibbs, M.L. and Springer, T.A. (1990) ICAM-1 (CD54): a counter-receptor for Mac-1 (CD11b/CD18). *J Cell Biol,* 111, 3129-3139.

238. Staunton, D.E., Merluzzi, V.J., Rothlein, R., Barton, R., Marlin, S.D. and Springer, T.A. (1989) A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. *Cell,* 56, 849-853.

239. Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D., Bailey-Wilson, J.E. *et al.* (2007) Replicating genotype-phenotype associations. *Nature,* 447, 655-660.

240. Adams, S., Turner, G.D., Nash, G.B., Micklem, K., Newbold, C.I. and Craig, A.G. (2000) Differential binding of clonal variants of Plasmodium falciparum to allelic forms of intracellular adhesion molecule 1 determined by flow adhesion assay. *Infect Immun,* 68, 264-269.

241. Ma, J., Mollsten, A., Falhammar, H., Brismar, K., Dahlquist, G., Efendic, S. and Gu, H.F. (2007) Genetic association analysis of the adiponectin polymorphisms in type 1 diabetes with and without diabetic nephropathy. *J Diabetes Complications,* 21, 28-33.

242. Register, T.C., Burdon, K.P., Lenchik, L., Bowden, D.W., Hawkins, G.A., Nicklas, B.J., Lohman, K., Hsu, F.C., Langefeld, C.D. and Carr, J.J. (2004) Variability of serum soluble intercellular adhesion molecule-1 measurements attributable to a common polymorphism. *Clin Chem,* 50, 2185-2187.

243. Walsh, E.C., Sabeti, P., Hutcheson, H.B., Fry, B., Schaffner, S.F., de Bakker, P.I., Varilly, P., Palma, A.A., Roy, J., Cooper, R. *et al.* (2006) Searching for signals of evolutionary selection in 168 genes related to immune function. *Hum Genet,* 119, 92-102.

244. Wildman, D.E., Uddin, M., Liu, G., Grossman, L.I. and Goodman, M. (2003) Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: enlarging genus Homo. *Proc Natl Acad Sci U S A,* 100, 7181-7188.

245. Casasnovas, J.M., Stehle, T., Liu, J.H., Wang, J.H. and Springer, T.A. (1998) A dimeric crystal structure for the N-terminal two domains of intercellular adhesion molecule-1. *Proc Natl Acad Sci U S A,* 95, 4134-4139.

246. Staunton, D.E., Dustin, M.L., Erickson, H.P. and Springer, T.A. (1990) The arrangement of the immunoglobulin-like domains of ICAM-1 and the binding sites for LFA-1 and rhinovirus. *Cell,* 61, 243-254.

247. Register, R.B., Uncapher, C.R., Naylor, A.M., Lineberger, D.W. and Colonno, R.J. (1991) Human-murine chimeras of ICAM-1 identify amino acid residues critical for rhinovirus and antibody binding. *J Virol,* 65, 6589-6596.

248. Xiao, C., Tuthill, T.J., Bator Kelly, C.M., Challinor, L.J., Chipman, P.R., Killington, R.A., Rowlands, D.J., Craig, A. and Rossmann, M.G. (2004) Discrimination among

rhinovirus serotypes for a variant ICAM-1 receptor molecule. *J Virol,* 78, 10034-10044.

249. Fry, A.E., Griffiths, M.J., Auburn, S., Diakite, M., Forton, J.T., Green, A., Richardson, A., Wilson, J., Jallow, M., Sisay-Joof, F. *et al.* (2008) Common variation in the ABO glycosyltransferase is associated with susceptibility to severe Plasmodium falciparum malaria. *Hum Mol Genet,* 17, 567-576.
250. (1990) Severe and complicated malaria. World Health Organization, Division of Control of Tropical Diseases. *Trans R Soc Trop Med Hyg,* 84 Suppl 2, 1-65.
251. Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W. and Arnheim, N. (1992) Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci U S A,* 89, 5847-5851.
252. Gonzalez, J.M., Portillo, M.C. and Saiz-Jimenez, C. (2005) Multiple displacement amplification as a pre-polymerase chain reaction (pre-PCR) to process difficult to amplify samples and low copy number sequences from natural environments. *Environ Microbiol,* 7, 1024-1028.
253. Ross, P., Hall, L., Smirnov, I. and Haff, L. (1998) High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat Biotechnol,* 16, 1347-1351.
254. Horvath, S., Xu, X., Lake, S.L., Silverman, E.K., Weiss, S.T. and Laird, N.M. (2004) Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. *Genet Epidemiol,* 26, 61-69.
255. Laird, N.M., Horvath, S. and Xu, X. (2000) Implementing a unified approach to family-based tests of association. *Genet Epidemiol,* 19 Suppl 1, S36-42.
256. Dudbridge, F. (2008) Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum Hered,* 66, 87-98.
257. Udomsangpetch, R., Reinhardt, P.H., Schollaardt, T., Elliott, J.F., Kubes, P. and Ho, M. (1997) Promiscuity of clinical Plasmodium falciparum isolates for multiple adhesion molecules under flow conditions. *J Immunol,* 158, 4358-4364.
258. Baruch, D.I., Gormely, J.A., Ma, C., Howard, R.J. and Pasloske, B.L. (1996) Plasmodium falciparum erythrocyte membrane protein 1 is a parasitized erythrocyte receptor for adherence to CD36, thrombospondin, and intercellular adhesion molecule 1. *Proc Natl Acad Sci U S A,* 93, 3497-3502.

259. Urban, B.C., Willcox, N. and Roberts, D.J. (2001) A role for CD36 in the regulation of dendritic cell function. *Proc Natl Acad Sci U S A,* 98, 8750-8755.

260. Pain, A., Ferguson, D.J., Kai, O., Urban, B.C., Lowe, B., Marsh, K. and Roberts, D.J. (2001) Platelet-mediated clumping of Plasmodium falciparum-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. *Proc Natl Acad Sci U S A,* 98, 1805-1810.

261. Ikeda, H., Mitani, T., Ohnuma, M., Haga, H., Ohtzuka, S., Kato, T., Nakase, T. and Sekiguchi, S. (1989) A new platelet-specific antigen, Naka, involved in the refractoriness of HLA-matched platelet transfusion. *Vox Sang,* 57, 213-217.

262. Urwijitaroon, Y., Barusrux, S., Romphruk, A. and Puapairoj, C. (1995) Frequency of human platelet antigens among blood donors in northeastern Thailand. *Transfusion,* 35, 868-870.

263. Yanai, H., Chiba, H., Fujiwara, H., Morimoto, M., Abe, K., Yoshida, S., Takahashi, Y., Fuda, H., Hui, S.P., Akita, H. *et al.* (2000) Phenotype-genotype correlation in CD36 deficiency types I and II. *Thromb Haemost,* 84, 436-441.

264. Bamshad, M.J., Wooding, S., Watkins, W.S., Ostler, C.T., Batzer, M.A. and Jorde, L.B. (2003) Human population genetic structure and inference of group membership. *Am J Hum Genet,* 72, 578-589.

265. Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A., Zerjal, T., Tyler-Smith, C. and Mehdi, S.Q. (2002) Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet,* 70, 1107-1124.

266. Fry, A.E., Trafford, C.J., Kimber, M.A., Chan, M.S., Rockett, K.A. and Kwiatkowski, D.P. (2006) Haplotype homozygosity and derived alleles in the human genome. *Am J Hum Genet,* 78, 1053-1059.

267. Asch, A.S., Barnwell, J., Silverstein, R.L. and Nachman, R.L. (1987) Isolation of the thrombospondin membrane receptor. *J Clin Invest,* 79, 1054-1061.

268. Lepretre, F., Linton, K.J., Lacquemant, C., Vatin, V., Samson, C., Dina, C., Chikri, M., Ali, S., Scherer, P., Seron, K. *et al.* (2004) Genetic study of the CD36 gene in a French diabetic population. *Diabetes Metab,* 30, 459-463.

269. Love-Gregory, L., Sherva, R., Sun, L., Wasson, J., Schappe, T., Doria, A., Rao, D.C., Hunt, S.C., Klein, S., Neuman, R.J. *et al.* (2008) Variants in the CD36 gene

associate with the metabolic syndrome and high-density lipoprotein cholesterol. *Hum Mol Genet,* 17, 1695-1704.

270. Savill, J., Hogg, N. and Haslett, C. (1991) Macrophage vitronectin receptor, CD36, and thrombospondin cooperate in recognition of neutrophils undergoing programmed cell death. *Chest,* 99, 6S-7S.

271. Dawson, D.W., Pearce, S.F., Zhong, R., Silverstein, R.L., Frazier, W.A. and Bouck, N.P. (1997) CD36 mediates the In vitro inhibitory effects of thrombospondin-1 on endothelial cells. *J Cell Biol,* 138, 707-717.

272. Laugerette, F., Passilly-Degrace, P., Patris, B., Niot, I., Febbraio, M., Montmayeur, J.P. and Besnard, P. (2005) CD36 involvement in orosensory detection of dietary lipids, spontaneous fat preference, and digestive secretions. *J Clin Invest,* 115, 3177-3184.

273. Curtis, B.R., Ali, S., Glazier, A.M., Ebert, D.D., Aitman, T.J. and Aster, R.H. (2002) Isoimmunization against CD36 (glycoprotein IV): description of four cases of neonatal isoimmune thrombocytopenia and brief review of the literature. *Transfusion,* 42, 1173-1179.

274. Yanai, H., Watanabe, I., Ishii, K., Morimoto, M., Fujiwara, H., Yoshida, S., Hui, S.P., Matsuno, K. and Chiba, H. (2007) Attenuated aerobic exercise capacity in CD36 deficiency. *J Med Genet,* 44, 445-447.

275. McGilvray, I.D., Serghides, L., Kapus, A., Rotstein, O.D. and Kain, K.C. (2000) Nonopsonic monocyte/macrophage phagocytosis of Plasmodium falciparum-parasitized erythrocytes: a role for CD36 in malarial clearance. *Blood,* 96, 3231-3240.

276. Smith, T.G., Serghides, L., Patel, S.N., Febbraio, M., Silverstein, R.L. and Kain, K.C. (2003) CD36-mediated nonopsonic phagocytosis of erythrocytes infected with stage I and IIA gametocytes of Plasmodium falciparum. *Infect Immun,* 71, 393-400.

277. Newbold, C., Warn, P., Black, G., Berendt, A., Craig, A., Snow, B., Msobo, M., Peshu, N. and Marsh, K. (1997) Receptor-specific adhesion and clinical disease in Plasmodium falciparum. *Am J Trop Med Hyg,* 57, 389-398.

278. Traore, B., Muanza, K., Looareesuwan, S., Supavej, S., Khusmith, S., Danis, M., Viriyavejakul, P. and Gay, F. (2000) Cytoadherence characteristics of Plasmodium

falciparum isolates in Thailand using an in vitro human lung endothelial cells model. *Am J Trop Med Hyg,* 62, 38-44.

279. Kashiwagi, H., Tomiyama, Y., Nozaki, S., Kiyoi, T., Tadokoro, S., Matsumoto, K., Honda, S., Kosugi, S., Kurata, Y. and Matsuzawa, Y. (2001) Analyses of genetic abnormalities in type I CD36 deficiency in Japan: identification and cell biological characterization of two novel mutations that cause CD36 deficiency in man. *Hum Genet,* 108, 459-466.
280. Longmate, J.A. (2001) Complexity and power in case-control association studies. *Am J Hum Genet,* 68, 1229-1237.
281. Clark, T.G., Fry, A.E., Auburn, S., Campino, S., Diakite, M., Green, A., Richardson, A., Teo, Y.Y., Small, K., Wilson, J. *et al.* (2009) Allelic heterogeneity of G6PD deficiency in West Africa and severe malaria susceptibility. *Eur J Hum Genet*.
282. Cann, H.M., de Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W.F., Bonne-Tamir, B., Cambon-Thomsen, A. *et al.* (2002) A human genome diversity cell line panel. *Science,* 296, 261-262.
283. Rosenberg, N.A. (2006) Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet,* 70, 841-847.
284. Sabeti, P.C., Walsh, E., Schaffner, S.F., Varilly, P., Fry, B., Hutcheson, H.B., Cullen, M., Mikkelsen, T.S., Roy, J., Patterson, N. *et al.* (2005) The Case for Selection at CCR5-Delta32. *PLoS Biol,* 3, e378.
285. de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J. and Altshuler, D. (2005) Efficiency and power in genetic association studies. *Nat Genet,* 37, 1217-1223.
286. Gonzalez, J.R., Armengol, L., Sole, X., Guino, E., Mercader, J.M., Estivill, X. and Moreno, V. (2007) SNPassoc: an R package to perform whole genome association studies. *Bioinformatics,* 23, 644-645.
287. Stephens, J.C., Reich, D.E., Goldstein, D.B., Shin, H.D., Smith, M.W., Carrington, M., Winkler, C., Huttley, G.A., Allikmets, R., Schriml, L. *et al.* (1998) Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. *Am J Hum Genet,* 62, 1507-1515.

288. Guo, S.W. (1997) Linkage disequilibrium measures for fine-scale mapping: a comparison. *Hum Hered,* 47, 301-314.

289. Sabatti, C. and Risch, N. (2002) Homozygosity and linkage disequilibrium. *Genetics,* 160, 1707-1719.

290. Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J.A., Rhodes, M., Reich, D.E. and Hirschhorn, J.N. (2004) Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet,* 74, 1111-1120.

291. Yu, F., Sabeti, P.C., Hardenbol, P., Fu, Q., Fry, B., Lu, X., Ghose, S., Vega, R., Perez, A., Pasternak, S. *et al.* (2005) Positive Selection of a Pre-Expansion CAG Repeat of the Human SCA2 Gene. *PLoS Genet,* 1, e41.

292. Altshuler, D., Brooks, L.D., Chakravarti, A., Collins, F.S., Daly, M.J. and Donnelly, P. (2005) A haplotype map of the human genome. *Nature,* 437, 1299-1320.

293. Spencer, C.C. and Coop, G. (2004) SelSim: a program to simulate population genetic data with natural selection and recombination. *Bioinformatics,* 20, 3673-3675.

294. Kimura, M. and Ota, T. (1973) The age of a neutral mutant persisting in a finite population. *Genetics,* 75, 199-212.

295. (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature,* 437, 69-87.

296. Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R. *et al.* (2001) Linkage disequilibrium in the human genome. *Nature,* 411, 199-204.

297. Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H. and Nielsen, R. (2005) Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res,* 15, 1496-1502.

298. Terwilliger, J.D., Zollner, S., Laan, M. and Paabo, S. (1998) Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'drift mapping' in small populations with no demographic expansion. *Hum Hered,* 48, 138-154.

299. Balding, D.J. (2006) A tutorial on statistical methods for population association studies. *Nat Rev Genet,* 7, 781-791.

300. Devlin, B., Roeder, K. and Wasserman, L. (2001) Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol,* 60, 155-166.

301. Pritchard, J.K., Stephens, M., Rosenberg, N.A. and Donnelly, P. (2000) Association mapping in structured populations. *Am J Hum Genet,* 67, 170-181.

302. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet,* 38, 904-909.

303. McGinnis, R., Shifman, S. and Darvasi, A. (2002) Power and efficiency of the TDT and case-control design for association scans. *Behav Genet,* 32, 135-144.

304. Teo, Y.Y., Fry, A.E., Sanjoaquin, M.A., Pederson, B., Small, K.S., Rockett, K.A., Kwiatkowski, D.P. and Clark, T.G. (2008) Assessing Genuine Parents-Offspring Trios for Genetic Association Studies. *Hum Hered,* 67, 26-37.

305. Conrad, D.F., Jakobsson, M., Coop, G., Wen, X., Wall, J.D., Rosenberg, N.A. and Pritchard, J.K. (2006) A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet,* 38, 1251-1260.

306. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science,* 296, 2225-2229.

307. Slager, S.L., Huang, J. and Vieland, V.J. (2000) Effect of allelic heterogeneity on the power of the transmission disequilibrium test. *Genet Epidemiol,* 18, 143-156.

308. Allison, A.C. (2004) Two lessons from the interface of genetics and medicine. *Genetics,* 166, 1591-1599.

309. Williams, T.N., Mwangi, T.W., Wambua, S., Peto, T.E., Weatherall, D.J., Gupta, S., Recker, M., Penman, B.S., Uyoga, S., Macharia, A. *et al.* (2005) Negative epistasis between the malaria-protective effects of alpha(+)-thalassemia and the sickle cell trait. *Nat Genet*.

310. Daniels, G. (2005) The molecular genetics of blood group polymorphism. *Transpl Immunol,* 14, 143-153.

311. Calafell, F., Roubinet, F., Ramirez-Soriano, A., Saitou, N., Bertranpetit, J. and Blancher, A. (2008) Evolutionary dynamics of the human ABO gene. *Hum Genet,* 124, 123-135.

312. Preece, A.F., Strahan, K.M., Devitt, J., Yamamoto, F. and Gustafsson, K. (2002) Expression of ABO or related antigenic carbohydrates on viral envelopes leads to

neutralization in the presence of serum containing specific natural antibodies and complement. *Blood,* 99, 2477-2482.

313. Neil, S.J., McKnight, A., Gustafsson, K. and Weiss, R.A. (2005) HIV-1 incorporates ABO histo-blood group antigens that sensitize virions to complement-mediated inactivation. *Blood,* 105, 4693-4699.

314. Seymour, R.M., Allan, M.J., Pomiankowski, A. and Gustafsson, K. (2004) Evolution of the human ABO polymorphism by two complementary selective pressures. *Proc Biol Sci,* 271, 1065-1072.

315. Isles, A.R. and Holland, A.J. (2005) Imprinted genes and mother-offspring interactions. *Early Hum Dev,* 81, 73-77.

316. Wolf, J.B. and Hager, R. (2006) A maternal-offspring coadaptation theory for the evolution of genomic imprinting. *PLoS Biol,* 4, e380.

317. Senga, E., Loscertales, M.P., Makwakwa, K.E., Liomba, G.N., Dzamalala, C., Kazembe, P.N. and Brabin, B.J. (2007) ABO blood group phenotypes influence parity specific immunity to Plasmodium falciparum malaria in Malawian women. *Malar J,* 6, 102.