

# Reappraisal of known malaria resistance loci in a large multicenter study

Malaria Genomic Epidemiology Network\*

Many human genetic associations with resistance to malaria have been reported, but few have been reliably replicated. We collected data on 11,890 cases of severe malaria due to *Plasmodium falciparum* and 17,441 controls from 12 locations in Africa, Asia and Oceania. We tested 55 SNPs in 27 loci previously reported to associate with severe malaria. There was evidence of association at  $P < 1 \times 10^{-4}$  with the *HBB*, *ABO*, *ATP2B4*, *G6PD* and *CD40LG* loci, but previously reported associations at 22 other loci did not replicate in the multicenter analysis. The large sample size made it possible to identify authentic genetic effects that are heterogeneous across populations or phenotypes, with a striking example being the main African form of *G6PD* deficiency, which reduced the risk of cerebral malaria but increased the risk of severe malarial anemia. The finding that *G6PD* deficiency has opposing effects on different fatal complications of *P. falciparum* infection indicates that the evolutionary origins of this common human genetic disorder are more complex than previously supposed.

It was recognized over half a century ago that malaria has been a major force of evolutionary selection on the human genome and that certain hematological disorders have risen to high frequency in malaria-endemic areas because they reduce the risk of death due to malaria<sup>1–3</sup>. Sickle hemoglobin (HbS) and glucose-6-phosphate dehydrogenase (*G6PD*) deficiency are often-quoted examples of natural selection due to malaria, and many other genetic associations with resistance or susceptibility to malaria have been reported<sup>2–9</sup>. However, the current literature contains many conflicting lines of evidence based on relatively small studies whose results have not been independently replicated.

To address this problem, we conducted a large multicenter case-control study of severe malaria across 12 locations in Burkina Faso, Cameroon, The Gambia, Ghana, Kenya, Malawi, Mali, Nigeria, Tanzania, Vietnam and Papua New Guinea (**Supplementary Fig. 1** and **Supplementary Table 1**). The structure of this consortial project has been described elsewhere<sup>10</sup>, and information about each of the partner studies can be found on the Malaria Genomic Epidemiology Network (MalariaGEN) website (see URLs). We used the World Health Organization (WHO) definition of severe malaria, which comprises a broad spectrum of life-threatening clinical complications of *P. falciparum* infection<sup>11–15</sup>. In this report, we examine genetic associations with severe malaria in general and with two distinct clinical forms of severe malaria: cerebral malaria with a Blantyre coma score of less than 3 and severe malarial anemia with a hemoglobin level of less than 5 g/dl or a hematocrit level of less than 15%.

## RESULTS

### Samples and clinical data

The first stage of work was to collect standardized clinical data on severe malaria from multiple locations (**Supplementary Table 2**).

This effort presented many practical challenges, as severe malaria is an acute illness that mainly occurs in resource-poor settings where laboratory facilities are limited and medical records can be unreliable. It was necessary to allow for variations in the design and implementation of the study in different settings, with study characteristics depending on a range of local circumstances. Investigators at different sites agreed at the outset on principles for sharing data and on standardized clinical definitions, and they also worked together to define best ethical practices across different local settings, including the development of guidelines for informed consent<sup>10,16,17</sup>. A set of web tools was developed to enable investigators to curate data in their locally used format before transforming them to the standardized format necessary for data from different sites to be merged.

After data curation and quality control (Online Methods), 11,890 cases of severe malaria and 17,441 controls were included for analysis (**Table 1** and **Supplementary Table 3**). Controls were intended to be representative of the populations to which the cases belonged; that is, a minority of controls may have subsequently gone on to develop severe malaria. The ancestry composition of the cases and controls at each location is shown in **Supplementary Table 4**. A total of 6,283 cases had cerebral malaria or severe malarial anemia, of which 3,345 had cerebral malaria only, 2,196 had severe malarial anemia only and 742 had both cerebral malaria and severe malarial anemia (**Table 1**). A further 5,607 cases did not have cerebral malaria or severe malarial anemia according to the criteria used here but satisfied the WHO definition of severe malaria, which includes a range of other clinical complications such as acidosis, respiratory distress and hypoglycemia that are not explored in detail in the present analysis<sup>11</sup>.

Most of the cases of severe malaria were young children, with median ages ranging from 1.3 to 3.8 years at different study sites, except in Vietnam where most cases were young adults with a median

\*A full list of authors and affiliations appears at the end of the paper.

**Table 1 Clinical phenotype case counts and percentage of case fatalities**

Study site	Cases					Controls
	All severe malaria	As phenotypes			Other severe malaria	
		Cerebral malaria	Severe malarial anemia	Cerebral malaria and severe malarial anemia		
The Gambia	2,425 (14)	785 (26)	458 (4)	126 (26)	1,056 (6)	3,342
Mali	453 (14)	86 (26)	185 (8)	74 (20)	108 (25)	344
Burkina Faso	865 (5)	107 (20)	399 (10)	20 (20)	339 (4)	729
Ghana (Navrongo)	682 (4)	22 (27)	248 (2)	14 (36)	398 (3)	489
Ghana (Kumasi)	1,496 (4)	230 (10)	551 (2)	75 (13)	640 (2)	2,042
Nigeria	77 (4)	6 (17)	8 (12)	0 (0)	63 (2)	40
Cameroon	621 (5)	39 (19)	82 (7)	8 (50)	492 (4)	578
Kenya	2,268 (11)	909 (14)	160 (8)	214 (16)	985 (7)	3,949
Tanzania	429 (11)	34 (41)	182 (4)	28 (29)	185 (11)	453
Malawi	1,388 (17)	873 (15)	132 (7)	166 (19)	217 (2)	2,697
Vietnam	794 (11)	211 (18)	31 (6)	8 (12)	544 (9)	2,538
Papua New Guinea	392 (2)	43 (5)	120 (2)	9 (11)	220 (0)	240
Total	11,890 (10)	3,345 (18)	2,196 (4)	742 (20)	5,607 (5)	17,441

Numbers are given for cases of all severe malaria by study site, with the percentage of case fatalities shown in parentheses. Cases are further divided into those with cerebral malaria only, severe malarial anemia only, both cerebral malaria and severe malarial anemia and other severe malaria.

age of 29 years. The median age for severe malarial anemia was 32 months, and the median age for cerebral malaria was 62 months (Supplementary Table 3). Cerebral malaria was more common than severe malarial anemia in The Gambia, Kenya, Malawi and Vietnam, whereas severe malarial anemia was more common than cerebral malaria in Burkina Faso, Cameroon, Ghana, Mali, Papua New Guinea and Tanzania. It has previously been observed that severe malarial anemia particularly affects young African children exposed to extremely high levels of malaria transmission, whereas cerebral malaria particularly affects older African children exposed to lower levels of malaria transmission<sup>18,19</sup>. The mean case fatality rate after treatment was 18% for cerebral malaria, 4% for severe malarial anemia and 10% for severe malaria overall (Table 1).

### Genetic loci analyzed

We tested previously reported associations with severe malaria for 55 SNPs in 27 gene regions: *ABO*, *ADORA2B*, *ATP2B4*, *C6*, *CD36*, *CD40LG*, *CR1*, *ACKR1* (*DARC*), *G6PD*, *GNAS*, *HBB*, *ICAM1*, *IL1A*, *IL1B*, *IL4*, *IL10*, *IL13*, *IL22*, *IRF1*, *LTA*, *NOS2*, *SPTB*, *TLR1*, *TLR4*, *TLR6*, *TLR9* and *TNF* (for corresponding references, see Supplementary Tables 5–7). Note that this report focuses on SNPs that we were able to reliably genotype using the Sequenom platform; that is, we do not here consider structural variants such as those responsible for  $\alpha$ -thalassaemia. Other SNPs were included for the purposes of quality control (Supplementary Fig. 2 and Supplementary Tables 5–7). All SNPs were initially tested for association with severe malaria using a standard logistic regression method assuming fixed effects across different populations and under a variety of models of inheritance (Online Methods). We repeated these tests of association separately for the severe malaria subphenotypes of cerebral malaria and severe malarial anemia. Results for each SNP are shown in Supplementary Figure 3 and Supplementary Tables 8–10, and below we discuss those SNPs that showed strong evidence of association in the multicenter analysis.

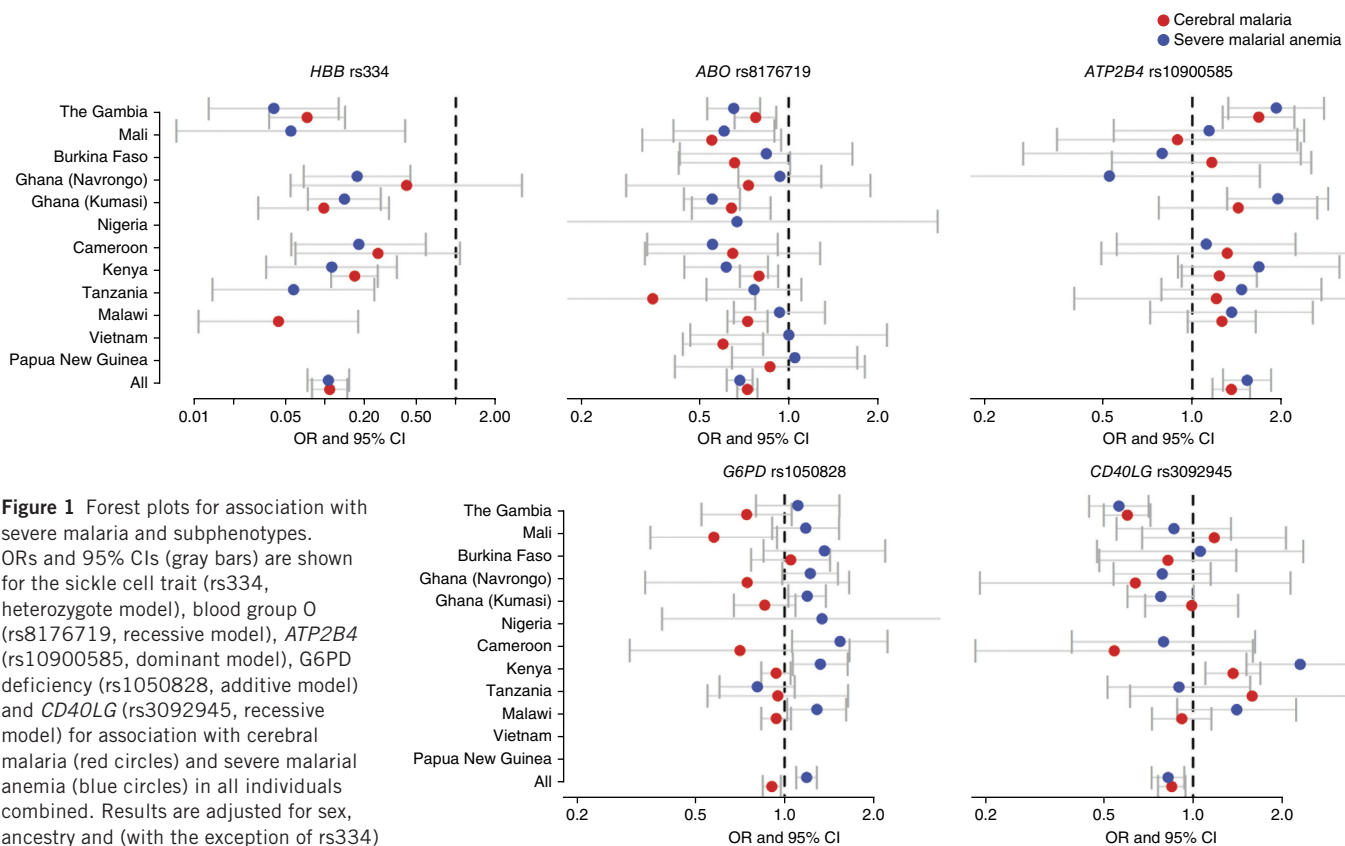
*HBB*. The *HBB* gene encodes  $\beta$ -globin, which has three well-known structural variants that have been associated with resistance to malaria: hemoglobin S (HbS), hemoglobin C (HbC) and hemoglobin E (HbE)<sup>2,7–9</sup>. The SNP responsible for HbS, rs334, was present at all the African sites, with heterozygote frequencies in

controls ranging from 0.05 (Malawi) to 0.22 (Nigeria). Heterozygotes had reduced risk of severe malaria (odds ratio (OR) = 0.14;  $P = 1.6 \times 10^{-225}$ ), cerebral malaria (OR = 0.11;  $P = 4.7 \times 10^{-88}$ ) and severe malarial anemia (OR = 0.11;  $P = 9.3 \times 10^{-65}$ ) (Fig. 1, Table 2 and Supplementary Table 11). The SNP responsible for HbC, rs33930165, was present only in individuals from West Africa (Burkina Faso, The Gambia, Ghana, Mali, Cameroon and Nigeria), with frequencies of the derived (non-ancestral/non-reference) alleles ranging from 0.01 to 0.15 in controls. The strongest signal of association with severe malaria was seen under an additive genetic model; that is, the greatest protective effect was seen in homozygotes. Each copy of the derived allele reduced the risk of severe malaria by 29% (OR = 0.71;  $P = 6.9 \times 10^{-9}$ ), reduced the risk of cerebral malaria by 28% (OR = 0.72;  $P = 0.01$ ) and reduced the risk of severe malarial anemia by 26% (OR = 0.74;  $P = 2.1 \times 10^{-3}$ )

(Table 2). The SNP responsible for HbE, rs33950507, was found at a derived allele frequency of 0.4 in the S'tieng ancestry group in Vietnam but was rare or absent in other ancestry groups, such that the sample was too small to estimate association with severe malaria.

*ABO*. The *ABO* gene encodes the glycosyltransferase enzyme that determines ABO blood group. Individuals who are homozygous for a single-nucleotide deletion (rs8176719) in *ABO* have an inactive form of the glycosyltransferase and can be classified as blood group O (ref. 20). Estimated by this criterion, the frequency of blood group O in control samples ranged from 0.32 in Papua New Guinea to 0.62 in Nigeria (Supplementary Table 12). Aggregated analysis across all sites showed that blood group O was associated with decreased risk of severe malaria (OR = 0.74;  $P = 5.0 \times 10^{-32}$ ), with reduced risk of cerebral malaria (OR = 0.73;  $P = 8.9 \times 10^{-16}$ ) and with reduced risk of severe malarial anemia (OR = 0.68;  $P = 7.9 \times 10^{-14}$ ) (Fig. 1 and Table 2). We also analyzed rs8176746, a nonsynonymous coding SNP in *ABO* that is in linkage disequilibrium (LD) with rs8176719 and determines the production of B antigens such that the majority of individuals carrying the derived allele express blood group B (refs. 20–22). The derived allele was associated in a dominant gene model with increased risk of severe malaria (OR = 1.25;  $P = 2.0 \times 10^{-17}$ ) (Table 2).

*G6PD*. The *G6PD* gene is an X-linked gene encoding glucose-6-phosphate dehydrogenase with many allelic variants<sup>23</sup>. The major form of *G6PD* enzyme deficiency in Africa is encoded by the derived allele of rs1050828 (*G6PD* c.202C>T), commonly known as *G6PD*+202T<sup>24</sup>. Two other SNPs that cause *G6PD* deficiency were found in individuals from The Gambia but were rare in the other populations studied here<sup>25</sup>. In this study, the *G6PD*+202T allele was present at frequencies ranging from 0.03 in The Gambia to 0.28 in Nigeria (Supplementary Table 13). Aggregated across all African sites, we found an increased risk of severe malarial anemia in male hemizygotes (OR = 1.49;  $P = 3.6 \times 10^{-5}$ ) and in female homozygotes under a recessive model of association (OR = 1.94;  $P = 1.9 \times 10^{-3}$ ) (Table 3 and Supplementary Tables 14 and 15). In contrast, there was a trend toward decreased risk of cerebral malaria in female heterozygotes (OR = 0.87;  $P = 0.06$ ) and male hemizygotes (OR = 0.81;  $P = 0.01$ ) (Supplementary Tables 14, 16 and 17). Below, we discuss this heterogeneity of effect in more detail. Similar but weaker trends were observed for rs1050829, which marks the ancestral lineage on which *G6PD*+202 originated<sup>24</sup>.



**Figure 1** Forest plots for association with severe malaria and subphenotypes.

ORs and 95% CIs (gray bars) are shown for the sickle cell trait (rs334, heterozygote model), blood group O (rs8176719, recessive model), *ATP2B4* (rs10900585, dominant model), *G6PD* deficiency (rs1050828, additive model) and *CD40LG* (rs3092945, recessive model) for association with cerebral malaria (red circles) and severe malarial anemia (blue circles) in all individuals combined. Results are adjusted for sex, ancestry and (with the exception of rs334) the sickle cell trait. Results are not presented

when the sample size was too small (fewer than five cases or controls with the relevant genotype) or for locations where the derived allele was absent. Further details are available in **Supplementary Tables 11–19**. OR = 1, representing no effect, is highlighted by the vertical dashed lines.

***ATP2B4*.** The *ATP2B4* gene, encoding a calcium transporter found in the plasma membrane of erythrocytes, has been identified by genome-wide association study (GWAS) as a malaria resistance locus<sup>26</sup>. We typed four SNPs in this gene that were found to be in LD; the derived alleles of rs10900585 and rs55868763 were associated with increased risk of severe malaria, whereas the derived alleles of rs4951074 and rs1541255 were associated with decreased risk (**Table 2** and **Supplementary Tables 8–10** and **18**). When aggregated across all African sites, individuals carrying at least one copy of the derived allele at rs10900585 had an OR of 1.32 for severe malaria ( $P = 1.7 \times 10^{-9}$ ), whereas individuals homozygous for the derived allele at rs4951074 had an OR of 0.77 ( $P = 7.6 \times 10^{-7}$ ). In both cases, the magnitude of the genetic effect was similar for cerebral malaria and severe malarial anemia (**Fig. 1**).

***CD40LG*.** The *CD40LG* gene is a gene on the X chromosome encoding CD40 ligand that has previously been associated with severe malaria<sup>27</sup>. Homozygotes for the derived allele of a SNP in the 5' UTR (rs3092945) showed reduced risk of severe malaria (OR = 0.85;  $P = 1.1 \times 10^{-6}$ ), with a similar trend of protection in both males (OR = 0.90;  $P = 0.01$ ) and females (OR = 0.78;  $P = 8.9 \times 10^{-5}$ ) when the data were aggregated across sites (**Table 3**). However, when sites were analyzed individually, the results were strikingly different between sites: homozygotes for the derived allele showed significantly reduced risk of severe malaria in The Gambia (OR = 0.54;  $P = 2.3 \times 10^{-22}$ ) but significantly increased risk in Kenya (OR = 1.42;  $P = 7.8 \times 10^{-6}$ ) (**Supplementary Table 19**).

**Other loci.** None of the other loci tested here showed consistent evidence of association with severe malaria in the multicenter analysis with a significance of  $P < 1 \times 10^{-4}$ . All variants tested, some of which

had weak associations that merit further investigation, are shown in **Supplementary Figure 3** and **Supplementary Tables 8–10**. At the *CD36* locus, heterozygotes for the codon variant rs201346212 tended to have reduced risk of severe malaria (OR = 0.67;  $P = 4.2 \times 10^{-4}$ ). Other weak signals of association ( $P$  values in the range of 0.05 to 0.001) were observed for *CD36*, *IL1A* and *IRF1* with severe malaria overall, for *CRI* and *IL4* with cerebral malaria and for *IL20RA* with severe malarial anemia. Although it is clear from these data that many genetic associations reported in the literature might have been false positives, as has been observed for other common diseases<sup>28</sup>, it is undoubtedly also the case that authentic genetic associations might be missed by multicenter studies if the effect is weak and there is heterogeneity of effect across different study sites.

### Epistasis between significantly associated loci

Epistasis between malaria resistance loci has been reported in previous studies<sup>29,30</sup>. We therefore tested for pairwise interaction between all SNPs that showed significant association at the *HBB*, *ABO*, *G6PD*, *ATP2B4* and *CD40LG* loci (**Supplementary Fig. 4** and **Supplementary Table 20**). This analysis did not identify any strong evidence of interaction, but a marginally significant effect was observed between the *ATP2B4* locus (rs10900585) and the allele for HbC (rs33930165;  $P = 1.3 \times 10^{-3}$ ), such that the ancestral allele of rs10900585, which was the minor allele in Africa, tended to reverse the protective effect of the HbC allele. This association warrants further investigation, as *ATP2B4* encodes the major erythrocyte calcium channel and intracellular calcium levels have been noted to affect the clinical phenotype of sickling disorders<sup>31</sup>.

**Table 2 Autosomal SNPs with strong association signals**

Gene	SNP	Alleles (ancestral/derived)	Frequency of the derived allele in controls (ancestral hom/het/derived hom)	Case phenotype	Frequency of the derived allele in cases (ancestral hom/het/derived hom)	Model <sup>a</sup>	Model OR (95% CI)	Model <i>P</i>
<i>ABO</i>	rs8176719	I/D	0.69 (1,700/7,215/8,238)	Severe malaria	0.64 (1,506/5,533/4,750)	R	0.74 (0.7–0.78)	$4.99 \times 10^{-33}$
	rs8176719	I/D	0.69 (1,700/7,215/8,238)	Cerebral malaria only	0.64 (415/1,537/1,373)	R	0.73 (0.67–0.79)	$8.85 \times 10^{-16}$
	rs8176719	I/D	0.69 (1,700/7,215/8,238)	Severe malarial anemia only	0.62 (302/1,054/816)	R	0.68 (0.62–0.76)	$7.97 \times 10^{-14}$
	rs8176746	C/A	0.17 (11,943/4,852/565)	Severe malaria	0.2 (7,512/3,817/507)	D	1.25 (1.19–1.32)	$2.01 \times 10^{-17}$
	rs8176746	C/A	0.17 (11,943/4,852/565)	Cerebral malaria only	0.2 (2,148/1,049/138)	A	1.27 (1.18–1.36)	$2.00 \times 10^{-11}$
	rs8176746	C/A	0.17 (11,943/4,852/565)	Severe malarial anemia only	0.22 (1,342/746/97)	D	1.28 (1.16–1.42)	$1.71 \times 10^{-6}$
<i>ATP2B4</i>	rs10900585	G/T	0.66 (1,644/5,722/5,737)	Severe malaria	0.68 (868/4,056/4,203)	D	1.32 (1.21–1.45)	$1.69 \times 10^{-9}$
	rs10900585	G/T	0.66 (1,644/5,722/5,737)	Cerebral malaria only	0.69 (245/1,168/1,236)	D	1.35 (1.17–1.57)	$3.06 \times 10^{-5}$
	rs10900585	G/T	0.66 (1,644/5,722/5,737)	Severe malarial anemia only	0.69 (159/816/851)	D	1.53 (1.27–1.84)	$3.68 \times 10^{-6}$
	rs1541255	A/G	0.32 (6,241/5,667/1,439)	Severe malaria	0.29 (4,558/3,922/743)	R	0.75 (0.68–0.83)	$4.87 \times 10^{-9}$
	rs1541255	A/G	0.32 (6,241/5,667/1,439)	Cerebral malaria only	0.29 (1,334/1,122/200)	R	0.7 (0.59–0.82)	$4.03 \times 10^{-6}$
	rs1541255	A/G	0.32 (6,241/5,667/1,439)	Severe malarial anemia only	0.29 (932/788/137)	R	0.67 (0.55–0.82)	$3.96 \times 10^{-5}$
<i>HBB</i>	rs334	A/T	0.07 (12,773/1,791/77)	Severe malaria	0.02 (10,388/213/84)	H	0.14 (0.12–0.16)	$1.62 \times 10^{-225}$
	rs334	A/T	0.07 (12,773/1,791/77)	Cerebral malaria only	0.01 (3,041/42/5)	H	0.11 (0.08–0.15)	$4.67 \times 10^{-88}$
	rs334	A/T	0.07 (12,773/1,791/77)	Severe malarial anemia only	0.03 (1,965/31/45)	H	0.11 (0.07–0.15)	$9.25 \times 10^{-65}$
	rs33930165	G/A	0.03 (9,341/515/74)	Severe malaria	0.04 (6,866/445/46)	A	0.71 (0.63–0.8)	$6.87 \times 10^{-9}$
	rs33930165	G/A	0.03 (9,341/515/74)	Cerebral malaria only	0.02 (1,412/56/5)	A	0.72 (0.56–0.94)	0.01
	rs33930165	G/A	0.03 (9,341/515/74)	Severe malarial anemia only	0.04 (1,465/116/10)	A	0.74 (0.6–0.9)	$2.11 \times 10^{-3}$

Odds ratios (ORs), 95% confidence intervals (95% CIs) and *P* values for the optimal genotypic model are presented for males and females combined. Results are adjusted for sex, ancestry and (with the exception of rs334) the sickle cell trait. Study sites at which a SNP was monomorphic were excluded from analysis. Further models are shown in **Supplementary Tables 8–13**. Association *P* values for all SNPs can be found in **Supplementary Figure 3**. Hom, homozygote; het, heterozygote; I/D, insertion/deletion.

<sup>a</sup>Models: A, additive; D, dominant; H, heterozygote advantage; R, recessive.

### Heterogeneity of effect

The large sample size of this study allowed us to investigate the heterogeneity in effect of malaria resistance loci in greater detail than has hitherto been possible. When associations with *HBB*, *ABO*, *G6PD*, *ATP2B4* and *CD40LG* were analyzed by population and when cerebral malaria and severe malarial anemia were treated as separate phenotypic entities, various different patterns were observed (**Fig. 1**). Standard approaches to genetic association analysis, such as the logistic regression methods used above, apply a fixed-effects model that assumes that true associations should be constant across different studies. However, patterns of disease association could potentially vary owing to a range of genetic, environmental and biological factors, and by understanding this variation we might gain important scientific insights. We therefore used a Bayesian statistical framework to evaluate different models of genetic association, allowing for heterogeneity of effect (see the Online Methods and ref. 32). In essence, this approach weighs up the evidence that a genetic effect is fixed or heterogeneous when it is compared across different populations and clinical phenotypes, for example, cerebral malaria and severe malarial anemia. Here we allowed for two kinds of heterogeneous effect: correlated effects, which are not fixed but tend to behave in a similar way, and independent effects, which have no tendency to behave in a similar way. We estimated the posterior probability of each model under the a priori assumption that all models were equally likely. The results for specific SNPs in *HBB*, *ABO*, *G6PD*, *ATP2B4* and *CD40LG* are shown in **Figure 2**, and the results for all SNPs are shown in **Supplementary Figure 5**.

For HbS and blood group O, there was strong evidence for fixed or correlated effects across different populations and clinical phenotypes (**Figs. 1 and 2**). HbS confers a remarkably fixed level of protection against cerebral malaria and severe malarial anemia, with heterozygotes

for the derived allele showing 89% reduced risk for both conditions when data were averaged across sites (**Table 2**). HbC and blood group O were less strongly protective than HbS, but they too had very similar effects on cerebral malaria and severe malarial anemia. There are many different theories about the molecular and cellular mechanisms by which HbS, HbC and blood group O act to protect against malaria<sup>22,33–36</sup>. The fact that these variants protect equally against clinical complications as disparate as coma and severe anemia implies that they act through some general mechanism, for example, by suppressing parasite density, rather than through a specific effect on particular pathological processes such as cerebral malaria or severe malarial anemia.

At the *CD40LG* and *ATP2B4* loci, there was evidence for heterogeneity of effect across populations (**Figs. 1 and 2**). For rs3092945 in *CD40LG*, the posterior probability of independent effects in different populations was greater than 90%, whereas the effects on clinical phenotype were relatively constant. In the case of rs10900585 in *ATP2B4*, the posterior probability was more evenly balanced between the various models, with some evidence for both fixed and independent effects in different populations. Heterogeneity in effect across populations might indicate some source of biological variation such as epistasis or gene-environment interactions, or it might be due to the associated SNPs simply being genetic markers that have variable patterns of LD with the true malaria resistance alleles. Variable patterns of LD are particularly common in Africa, such that authentic genetic associations may fail to replicate in different locations unless the causal variant is directly genotyped, but in general this form of population heterogeneity might prove useful in the genetic fine mapping of causal variants<sup>32,37,38</sup>.

The genetic effects observed for *G6PD* were strikingly different in character from those seen at the *HBB*, *ABO*, *ATP2B4* and *CD40LG*



**Table 3 X-chromosome SNPs with strong association signals**

Gene	SNP	Alleles (ancestral/derived)	Sample	Frequency of the derived allele in controls (ancestral hom/het/derived hom)	Case phenotype	Frequency of the derived allele in cases (ancestral hom/het/derived hom)	Model <sup>a</sup>	Model OR (95% CI)	Model <i>P</i>
CD40LG	rs3092945	T/C	All	0.27 (11,030/2,621/3,197)	Severe malaria	0.29 (7,123/2,035/2,250)	R	0.85 (0.79–0.91)	$1.11 \times 10^{-6}$
	rs3092945	T/C	All	0.27 (11,030/2,621/3,197)	Cerebral malaria only	0.27 (2,102/576/596)	R	0.85 (0.76–0.94)	$2.45 \times 10^{-3}$
	rs3092945	T/C	All	0.27 (11,030/2,621/3,197)	Severe malarial anemia only	0.32 (1,197/411/452)	R	0.82 (0.73–0.94)	$2.97 \times 10^{-3}$
	rs3092945	T/C	F	0.27 (4,581/2,621/849)	Severe malaria	0.3 (2,636/2,035/513)	R	0.78 (0.69–0.88)	$8.93 \times 10^{-5}$
	rs3092945	T/C	F	0.27 (4,581/2,621/849)	Cerebral malaria only	0.29 (813/576/157)	R	0.9 (0.74–1.09)	0.27
	rs3092945	T/C	F	0.27 (4,581/2,621/849)	Severe malarial anemia only	0.33 (424/411/103)	R	0.71 (0.56–0.9)	$3.49 \times 10^{-3}$
	rs3092945	T/C	M	0.27 (6,449/0/2,348)	Severe malaria	0.28 (4,487/0/1,737)	M	0.9 (0.83–0.98)	0.01
	rs3092945	T/C	M	0.27 (6,449/0/2,348)	Cerebral malaria only	0.25 (1,289/0/439)	M	0.85 (0.75–0.97)	0.01
	rs3092945	T/C	M	0.27 (6,449/0/2,348)	Severe malarial anemia only	0.31 (773/0/349)	M	0.87 (0.75–1.01)	0.07
	G6PD	rs1050828	C/T	All	0.15 (11,552/1,770/1,279)	Severe malaria	0.15 (8,516/1,152/1,000)	A	1.02 (0.97–1.06)
rs1050828		C/T	All	0.15 (11,552/1,770/1,279)	Cerebral malaria only	0.13 (2,513/338/230)	A	0.91 (0.85–0.97)	$6.08 \times 10^{-3}$
rs1050828		C/T	All	0.15 (11,552/1,770/1,279)	Severe malarial anemia only	0.18 (1,563/215/263)	A	1.19 (1.1–1.28)	$2.62 \times 10^{-5}$
rs1050828		C/T	F	0.15 (5,069/1,770/174)	Severe malaria	0.14 (3,705/1,152/134)	A	0.95 (0.88–1.03)	0.23
			F	0.15 (5,069/1,770/174)	All severe malaria	0.14 (3,705/1,152/134)	R	1.15 (0.9–1.46)	0.27
			F	0.15 (5,069/1,770/174)	All severe malaria	0.14 (3,705/1,152/134)	H	0.9 (0.82–0.99)	0.02
rs1050828		C/T	F	0.15 (5,069/1,770/174)	Cerebral malaria only	0.14 (1,129/338/39)	A	0.92 (0.82–1.04)	0.19
			F	0.15 (5,069/1,770/174)	Cerebral malaria only	0.14 (1,129/338/39)	R	1.09 (0.76–1.57)	0.65
			F	0.15 (5,069/1,770/174)	Cerebral malaria only	0.14 (1,129/338/39)	H	0.87 (0.76–1.01)	0.06
rs1050828		C/T	F	0.15 (5,069/1,770/174)	Severe malarial anemia only	0.16 (669/215/41)	A	1.12 (0.96–1.3)	0.15
			F	0.15 (5,069/1,770/174)	Severe malarial anemia only	0.16 (669/215/41)	R	1.94 (1.3–2.89)	$1.92 \times 10^{-3}$
			F	0.15 (5,069/1,770/174)	Severe malarial anemia only	0.16 (669/215/41)	H	0.93 (0.77–1.11)	0.42
rs1050828		C/T	M	0.15 (6,483/0/1,105)	Severe malaria	0.15 (4,811/0/866)	M	1.1 (0.99–1.22)	0.07
rs1050828		C/T	M	0.15 (6,483/0/1,105)	Cerebral malaria only	0.12 (1,384/0/191)	M	0.81 (0.68–0.96)	0.01
rs1050828	C/T	M	0.15 (6,483/0/1,105)	Severe malarial anemia only	0.2 (894/0/222)	M	1.49 (1.24–1.79)	$3.55 \times 10^{-5}$	

Odds ratios (ORs), 95% confidence intervals (95% CIs) and *P* values for the optimal genetic model. Results are adjusted for sex (all individuals combined), ancestry and the sickle cell trait. Study sites at which a SNP was monomorphic were excluded from analysis. Further models are shown in **Supplementary Tables 8–10** and **14–17**. Association *P* values for all SNPs can be found in **Supplementary Figure 3**. Hom, homozygote; het, heterozygote; M, male; F, female.

<sup>a</sup>Models: A, additive; D, dominant; H, heterozygote advantage; M, male hemizygote; R, recessive.

loci (**Figs. 1** and **2**). There was strong heterogeneity of effect across the phenotypes, with more than 80% posterior probability that the G6PD+202T allele had independent effects on cerebral malaria and severe malarial anemia. This difference in effect was observed consistently across the different populations. Previous studies have concluded that G6PD deficiency is protective against *P. falciparum*, although there has been debate about whether the protective effect is confined to female heterozygotes or is also present in male hemizygotes<sup>4,5,25,39</sup>. With a much larger sample size than any previous study, our data show that male hemizygotes and female homozygotes have increased risk of severe malarial anemia, whereas male hemizygotes and female heterozygotes have reduced risk of cerebral malaria. If males and females are combined in an additive model, the G6PD+202T allele confers reduced risk of cerebral malaria (OR = 0.91; *P* =  $6.1 \times 10^{-3}$ ) but increased risk of severe malarial anemia (OR = 1.19; *P* =  $2.6 \times 10^{-5}$ ), and the overall effect on severe malaria is close to neutral (OR = 1.02; *P* = 0.15) (**Table 3** and **Supplementary Tables 13–17**).

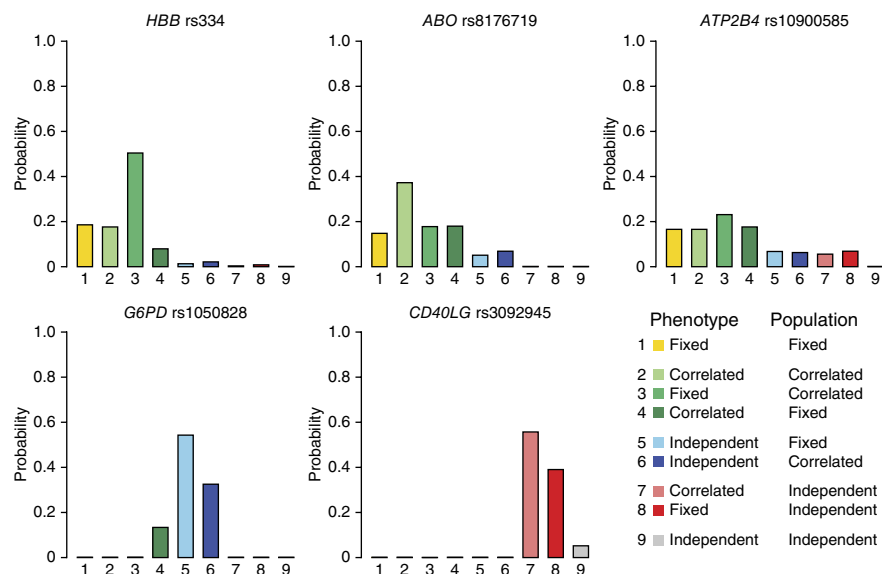
## DISCUSSION

Severe malarial anemia is a complex pathological entity<sup>40,41</sup>. Parasites invade and destroy erythrocytes as they replicate, and the host response to infection also leads to erythrocyte destruction and bone marrow suppression. In the developing world, the resulting anemia

is often aggravated by chronic nutritional deficiency and helminthic infection<sup>41,42</sup>. G6PD deficiency and the hemoglobinopathies are interesting examples of a biological tradeoff between an inherent tendency to cause anemia and the potential to protect against anemia by protecting against malaria. The present study allows us to observe the outcome of this tradeoff with greater resolution than has hitherto been possible, showing that the risk of severe malarial anemia is significantly reduced for HbS heterozygotes, HbC homozygotes and HbC heterozygotes, whereas for G6PD-deficient male hemizygotes and female homozygotes the risk is significantly increased.

G6PD deficiency has excited much interest among evolutionary biologists because it is so common in the human population and displays such a remarkable diversity of allelic forms throughout the tropics and subtropics<sup>3,6,9,23,39,43</sup>. The main force for evolutionary selection is widely assumed to be severe malaria due to *P. falciparum*, but this focus is called into question by the present findings showing that G6PD deficiency has little effect on the overall risk of severe malaria. These data do not exclude the possibility that *P. falciparum* has had a role, as the present-day risk of cerebral malaria and severe malarial anemia might not accurately reflect patterns of disease and fatality caused by *P. falciparum* in the past, particularly before antimalarial drugs became widely used. However, it is necessary to consider the possibility of some other evolutionary driving force, and

**Figure 2** Genetic heterogeneity for the severe malaria subtypes cerebral malaria only and severe malarial anemia only within and across African sites for significant loci. Bar plots show the distribution of probability between each of nine models of association where the effects on each phenotype are fixed, independent or correlated within a site in combination with being fixed, independent or correlated across all sites. Models are a priori assumed to be equally likely (see the Online Methods and supplementary material for details). Results are shown for SNPs rs334 (HbS, heterozygote model) in *HBB*, rs8176719 homozygotes (blood group O) in *ABO*, rs10900585 in *ATP2B4*, G6PD+202 (rs1050828) in *G6PD* and rs3092945 in *CD40LG*.



an obvious candidate is *Plasmodium vivax*, once regarded as benign but increasingly recognized as causing a substantial burden of fatality and severe disease<sup>44</sup>. In a region of Thailand where both *Plasmodium* species are endemic, G6PD deficiency has been observed to suppress *P. vivax* infection more effectively than *P. falciparum* infection<sup>45</sup>. An important biological difference of *P. vivax* relative to *P. falciparum* is that it preferentially infects reticulocytes (young erythrocytes), which have higher levels of G6PD enzyme activity than older erythrocytes, and it is conceivable that in these circumstances G6PD deficiency might exert a stronger protective effect. The geographical distribution of G6PD deficiency broadly coincides with the transmission range of *P. vivax*, with the notable exception of large parts of sub-Saharan Africa, where G6PD deficiency is common but *P. vivax* is absent<sup>43,46</sup>. It has recently been discovered that chimpanzees and gorillas in Africa carry parasites that are closely related to *P. vivax*<sup>47</sup>. Evolutionary analysis of these ape parasite lineages indicates that human *P. vivax* is of African origin, suggesting that *P. vivax* might have been common in Africa before the selective sweep of the FYB<sup>E5</sup> (FY\*O) allele (Duffy negative blood group) led to its elimination from most of the continent<sup>48,49</sup>.

Large multicenter studies have transformed the field of human genetics over the past decade, but major practical challenges remain in conducting such studies in the developing world. Obtaining reliable phenotypic data can be difficult in resource-poor settings, particularly for acute conditions whose diagnosis depends on accurate clinical records at the time of illness, such as severe malaria. We have endeavored to overcome these obstacles by establishing systems for standardizing and sharing data from research groups in different countries. The extremely strong phenotypic associations observed for the sickle cell trait and blood group O and their remarkably consistent effects on different clinical forms of severe malaria illustrate the effectiveness of this approach and provide a benchmark for evaluating other loci. A key finding is that most previously reported candidate gene associations do not replicate, and an important use of the sample collections established by this project will be to discover authentic susceptibility loci by conducting GWAS on a larger scale than was previously possible. A critical consideration when conducting multicenter GWAS is that true genetic effects can exhibit marked heterogeneity across different populations, and here we propose a statistical framework for dealing with this problem. There are many potential sources of heterogeneity, including interactions with other infections and environmental variables. A major source of heterogeneity could be parasite genetic variation, whose impact on the clinical outcome of malaria remains very poorly understood, and this is an area of future investigation that could

yield important biological insights. Previously, this has been difficult to address in a systematic manner, but with growing knowledge about the natural landscape of genome variation in both the host and the parasite there is the potential to discover new genetic interactions, for example, between parasite ligands and host receptors for erythrocyte invasion, that could be of great practical importance for vaccine development. This work will require a framework for large-scale genetic epidemiology studies, and here we demonstrate the feasibility of integrating data across multiple locations to achieve scientific insights that could not be achieved by individual studies in isolation.

**URLs.** MalariaGEN Partner Sites, <http://www.malariagen.net/projects/cp1>; MalariaGEN case report form, <http://www.malariagen.net/files/downloads/23.pdf>; MalariaGEN ethics and governance, <http://www.malariagen.net/community/ethics-governance>; R project, <http://www.r-project.org/>.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

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

## ACKNOWLEDGMENTS

The MalariaGEN Project is supported by the Wellcome Trust (WT077383/Z/05/Z) and the Bill and Melinda Gates Foundation through The Foundation for the National Institutes of Health (FNIH, USA) (566) as part of the Grand Challenges in Global Health Initiative. The Resource Centre for Genomic Epidemiology of Malaria is supported by the Wellcome Trust (090770/Z/09/Z). This research was supported by the UK Medical Research Council (G0600718 and G0600230) and by the Wellcome Trust Biomedical Ethics Enhancement Award (087285) and Strategic Award (096527). D.P.K. receives support from the UK Medical Research Council (G19/9). C.C.A.S. was supported by a Wellcome Trust Career Development Fellowship (097364/Z/11/Z). The Wellcome Trust also provides core awards to the Wellcome Trust Centre for Human Genetics (090532/Z/09/Z) and the Wellcome Trust Sanger Institute (098051/Z/05/Z). The Malaria Research and Training Center—Bandiagara Malaria Project (MRTC-BMP) in Mali group is supported by an Interagency Committee on Disability Research (ICDR) grant from the National Institute of Allergy and Infectious Diseases/US National Institutes of Health (NIAID/NIH) to the University of Maryland and the University of Bamako (USTTB) and by the Mali-NIAID/NIH International Centers for Excellence in Research (ICER) at USTTB. Contributions from Nigeria to CP1 were supported financially by a grant within the BioMalPar European Network of Excellence (LSHP-CT-2004-503578). E.A. received partial funding from the European

Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement 242095-EVIMaLaR and the Central African Network for Tuberculosis, HIV/AIDS and Malaria (CANTAM) funded by the European and Developing Countries Clinical Trials Partnership (EDCTP). T.N.W. is funded by Senior Fellowship awards from the Wellcome Trust (076934/Z/05/Z and 091758/Z/10/Z) and through the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement 242095-EVIMaLaR. The Kenya Medical Research Institute (KEMRI)–Wellcome Trust Programme is funded through core support from the Wellcome Trust. This paper is published with the permission of the director of KEMRI. C.M.N. is supported through a strategic award to the KEMRI–Wellcome Trust Programme from the Wellcome Trust (084538). The Joint Malaria Programme, Kilimanjaro Christian Medical Centre in Tanzania received funding from a UK MRC grant (G9901439). We would like to thank all the Vietnamese individuals who agreed to provide samples for this study. We acknowledge the work of the clinical staff from the Hospital of Tropical Diseases, HCMC and Phuoc Long and Dong Xoai District Hospitals in Binh Phuoc province, Vietnam, who initially diagnosed and studied the individuals with severe malaria. We would like to thank N.T. Hieu and his staff from Hung Vuong Obstetric Hospital for the collection of the cord blood controls. The clinical component of this study was funded through the Wellcome Trust Major Overseas Program in Vietnam (089276/Z/09/Z). L.M. was supported through Basser (Royal Australasian College of Physicians) and National Health and Medical Research Council (NHMRC) scholarships. M.L. was supported through a Fogarty Foundation Scholarship. T.M.E.D. was supported through an NHMRC practitioner fellowship.

#### AUTHOR CONTRIBUTIONS

All senior authors from each partner site have approved the final manuscript and their site authorship, acknowledgments and full contributors list. Writing group: G.M.C., D.P.K., S.M., K.A.R. and C.C.A.S. Project management: E.A., T. Agbenyega, S.A., A.A., O.A., K.A.B., D.J.C., V.C., T.M.E.D., O.D., C.D., S.J.D., J.F., T.T.H., K.J.J., H.K., A.K., S.K., K.A.K., D.P.K., K.M., P.M., D.M., M.M., I.M., A. Niangaly, N.P., M. Pinder, B.P., H.R., E.R., K.A.R., P.S., S.B.S., G.S., S.S., T.T., M.A.T., T.N.W. and M.D.W. Sample clinical data collection and management: S.A., A.A., L.N.A.-E., O.A., T. Apinjuh, K.A.B., E.C.B., G.M.C., D.J.C., S.J.D., A.E., J.E., K.F., A. Ghansah, L.H., M.J., D.K., H.K., A.K., S.K., M.L., A. Macharia, V.D.M., A. Manjurano, L.M., P.M., S.M., R.M., A. Niangaly, C.M.N., A. Ndi, V.N., S.O., N.P., N.H.P., M. Pinder, B.P., N.T.N.Q., H.R., K.A.R., M.S., G.S., F.S.-J., S.S., T.T., C.Q.T., M.A.T., O.T., S. Usen, S. Uyoga and A.V. Sample processing, genotyping and management: A.A., L.N.A.-E., O.A., T. Apinjuh, E.C.B., R.C., A.E., A. Ghansah, A. Green, L.H., C.H., M.J., A.E.J., D.K., H.K., M.L., A. Macharia, V.D.M., A. Manjurano, L.M., S.M., R.M., A. Niangaly, C.M.N., A. Ndi, V.N., S.O., N.H.P., N.T.N.Q., K.A.R., K.R., F.S.-J., C.Q.T., O.T., S. Uyoga and A.V. Analysis: G.B., T.G.C., G.M.C., D.P.K., S.Q.L., S.M., M. Pirinen, K.A.R., N.S. and C.C.A.S. Full details on contribution are given in the supplementary material.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

- Haldane, J.B.S. Disease and evolution. *Ric. Sci. Suppl.* **A 19**, 68–76 (1949).
- Allison, A.C. Protection afforded by sickle-cell trait against subtertian malarial infection. *BMJ* **1**, 290–294 (1954).
- Allison, A.C. Genetic factors in resistance to malaria. *Ann. NY Acad. Sci.* **91**, 710–729 (1961).
- Bienzie, U., Ayeni, O., Lucas, A.O. & Luzzatto, L. Glucose-6-phosphate dehydrogenase and malaria. Greater resistance of females heterozygous for enzyme deficiency and of males with non-deficient variant. *Lancet* **1**, 107–110 (1972).
- Ruwende, C. *et al.* Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. *Nature* **376**, 246–249 (1995).
- Tishkoff, S.A. *et al.* Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* **293**, 455–462 (2001).
- Kwiatkowski, D.P. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet.* **77**, 171–192 (2005).
- Verra, F., Mangano, V.D. & Modiano, D. Genetics of susceptibility to *Plasmodium falciparum*: from classical malaria resistance genes towards genome-wide association studies. *Parasite Immunol.* **31**, 234–253 (2009).
- Hedrick, P.W. Population genetics of malaria resistance in humans. *Heredity (Edinb.)* **107**, 283–304 (2011).
- Malaria Genomic Epidemiology Network. A global network for investigating the genomic epidemiology of malaria. *Nature* **456**, 732–737 (2008).
- World Health Organization. Severe falciparum malaria. *Trans. R. Soc. Trop. Med. Hyg.* **94** (suppl. 1), S1–S90 (2000).
- World Health Organization. *Guidelines for the Treatment of Malaria* 2nd edn. (World Health Organization, Geneva, 2010).
- Marsh, K. *et al.* Indicators of life-threatening malaria in African children. *N. Engl. J. Med.* **332**, 1399–1404 (1995).
- Taylor, T. *et al.* Standardized data collection for multi-center clinical studies of severe malaria in African children: establishing the SMAC network. *Trans. R. Soc. Trop. Med. Hyg.* **100**, 615–622 (2006).
- Jallow, M. *et al.* Clinical features of severe malaria associated with death: a 13-year observational study in The Gambia. *PLoS ONE* **7**, e45645 (2012).
- Chokshi, D.A. *et al.* Valid consent for genomic epidemiology in developing countries. *PLoS Med.* **4**, e95 (2007).
- Parker, M. *et al.* Ethical data release in genome-wide association studies in developing countries. *PLoS Med.* **6**, e1000143 (2009).
- Snow, R.W. *et al.* Relation between severe malaria morbidity in children and level of *Plasmodium falciparum* transmission in Africa. *Lancet* **349**, 1650–1654 (1997).
- Reyburn, H. *et al.* Association of transmission intensity and age with clinical manifestations and case fatality of severe *Plasmodium falciparum* malaria. *J. Am. Med. Assoc.* **293**, 1461–1470 (2005).
- Yamamoto, F., Clausen, H., White, T., Marken, J. & Hakomori, S. Molecular genetic basis of the histo-blood group ABO system. *Nature* **345**, 229–233 (1990).
- Ogasawara, K. *et al.* Molecular genetic analysis of variant phenotypes of the ABO blood group system. *Blood* **88**, 2732–2737 (1996).
- Fry, A.E. *et al.* Common variation in the ABO glycosyltransferase is associated with susceptibility to severe *Plasmodium falciparum* malaria. *Hum. Mol. Genet.* **17**, 567–576 (2008).
- Cappellini, M.D. & Fiorelli, G. Glucose-6-phosphate dehydrogenase deficiency. *Lancet* **371**, 64–74 (2008).
- Hirono, A. & Beutler, E. Molecular cloning and nucleotide sequence of cDNA for human glucose-6-phosphate dehydrogenase variant A. *Proc. Natl. Acad. Sci. USA* **85**, 3951–3954 (1988).
- Clark, T.G. *et al.* Allelic heterogeneity of G6PD deficiency in West Africa and severe malaria susceptibility. *Eur. J. Hum. Genet.* **17**, 1080–1085 (2009).
- Timmann, C. *et al.* Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature* **489**, 443–446 (2012).
- Sabeti, P. *et al.* CD40L association with protection from severe malaria. *Genes Immun.* **3**, 286–291 (2002).
- Hirschhorn, J.N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genet. Med.* **4**, 45–61 (2002).
- Williams, T.N. *et al.* Negative epistasis between the malaria-protective effects of  $\alpha^+$ -thalassaemia and the sickle cell trait. *Nat. Genet.* **37**, 1253–1257 (2005).
- Atkinson, S.H. *et al.* Epistasis between the haptoglobin common variant and  $\alpha^+$ -thalassaemia influences risk of severe malaria in Kenyan children. *Blood* **123**, 2008–2016 (2014).
- Steinberg, M.H., Eaton, J.W., Berger, E., Coleman, M.B. & Oelshlegel, F.J. Erythrocyte calcium abnormalities and the clinical severity of sickling disorders. *Br. J. Haematol.* **40**, 533–539 (1978).
- Band, G. *et al.* Imputation-based meta-analysis of severe malaria in three African populations. *PLoS Genet.* **9**, e1003509 (2013).
- Fairhurst, R.M. *et al.* Abnormal display of PfEMP-1 on erythrocytes carrying haemoglobin C may protect against malaria. *Nature* **435**, 1117–1121 (2005).
- Bunn, H.F. The triumph of good over evil: protection by the sickle gene against malaria. *Blood* **121**, 20–25 (2013).
- Cserti, C.M. & Dzik, W.H. The ABO blood group system and *Plasmodium falciparum* malaria. *Blood* **110**, 2250–2258 (2007).
- Rowe, J.A. *et al.* Blood group O protects against severe *Plasmodium falciparum* malaria through the mechanism of reduced rosetting. *Proc. Natl. Acad. Sci. USA* **104**, 17471–17476 (2007).
- Teo, Y.Y., Small, K.S. & Kwiatkowski, D.P. Methodological challenges of genome-wide association analysis in Africa. *Nat. Rev. Genet.* **11**, 149–160 (2010).
- Teo, Y.Y. *et al.* Genome-wide comparisons of variation in linkage disequilibrium. *Genome Res.* **19**, 1849–1860 (2009).
- Luzzatto, L. G6PD deficiency and malaria selection. *Heredity* **108**, 456 (2012).
- Menendez, C., Fleming, A.F. & Alonso, P.L. Malaria-related anaemia. *Parasitol. Today* **16**, 469–476 (2000).
- Weatherall, D.J. & Kwiatkowski, D. Hematologic disorders of children in developing countries. *Pediatr. Clin. North Am.* **49**, 1149–1164 (2002).
- Calis, J.C. *et al.* Severe anemia in Malawian children. *N. Engl. J. Med.* **358**, 888–899 (2008).
- Howes, R.E., Battle, K.E., Satyagraha, A.W., Baird, J.K. & Hay, S.I. G6PD deficiency: global distribution, genetic variants and primaquine therapy. *Adv. Parasitol.* **81**, 133–201 (2013).
- Price, R.N. *et al.* Vivax malaria: neglected and not benign. *Am. J. Trop. Med. Hyg.* **77**, 79–87 (2007).
- Louicharoen, C. *et al.* Positively selected G6PD-Mahidol mutation reduces *Plasmodium vivax* density in Southeast Asians. *Science* **326**, 1546–1549 (2009).
- Guerra, C.A. *et al.* The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Negl. Trop. Dis.* **4**, e774 (2010).
- Liu, W. *et al.* African origin of the malaria parasite *Plasmodium vivax*. *Nat. Commun.* **5**, 3346 (2014).
- Miller, L.H., Mason, S.J., Clyde, D.F. & McGinniss, M.H. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, *FyFy*. *N. Engl. J. Med.* **295**, 302–304 (1976).
- Hamblin, M.T., Thompson, E.E. & Di Rienzo, A. Complex signatures of natural selection at the Duffy blood group locus. *Am. J. Hum. Genet.* **70**, 369–383 (2002).

The complete list of authors is as follows:

Kirk A Rockett<sup>1-3</sup>, Geraldine M Clarke<sup>1,3</sup>, Kathryn Fitzpatrick<sup>1</sup>, Christina Hubbart<sup>1</sup>, Anna E Jeffreys<sup>1</sup>, Kate Rowlands<sup>1</sup>, Rachel Craik<sup>1</sup>, Muminatou Jallow<sup>4,5</sup>, David J Conway<sup>4,6</sup>, Kalifa A Bojang<sup>4</sup>, Margaret Pinder<sup>4</sup>, Stanley Usen<sup>4</sup>, Fatoumatta Sisay-Joof<sup>4</sup>, Giorgio Sirugo<sup>4</sup>, Ousmane Toure<sup>7</sup>, Mahamadou A Thera<sup>7</sup>, Salimata Konate<sup>7</sup>, Sibiry Sissoko<sup>7</sup>, Amadou Niangaly<sup>7</sup>, Belco Poudiougou<sup>7</sup>, Valentina D Mangano<sup>8</sup>, Edith C Bougouma<sup>9</sup>, Sodiomon B Sirima<sup>9</sup>, David Modiano<sup>8</sup>, Lucas N Amenga-Etego<sup>10</sup>, Anita Ghansah<sup>11</sup>, Kwadwo A Koram<sup>11</sup>, Michael D Wilson<sup>11</sup>, Anthony Enimil<sup>12</sup>, Jennifer Evans<sup>13,14</sup>, Olukemi Amodu<sup>15</sup>, Subulade Olaniyan<sup>15</sup>, Tobias Apinjah<sup>16</sup>, Regina Mugri<sup>17</sup>, Andre Ndi<sup>17</sup>, Carolyne M Ndila<sup>18</sup>, Sophie Uyoga<sup>18</sup>, Alexander Macharia<sup>18</sup>, Norbert Peshu<sup>18</sup>, Thomas N Williams<sup>18,19</sup>, Alphaxard Manjurano<sup>20,21</sup>, Eleanor Riley<sup>21</sup>, Chris Drakeley<sup>20,21</sup>, Hugh Reyburn<sup>20,21</sup>, Vysaul Nyirongo<sup>22</sup>, David Kachala<sup>23</sup>, Malcolm Molyneux<sup>24</sup>, Sarah J Dunstan<sup>23-25</sup>, Nguyen Hoan Phu<sup>23,26</sup>, Nguyen Thi Ngoc Quyen<sup>23</sup>, Cao Quang Thai<sup>23,26</sup>, Tran Tinh Hien<sup>23,26</sup>, Laurens Manning<sup>27,28</sup>, Moses Laman<sup>27</sup>, Peter Siba<sup>27</sup>, Harin Karunajeewa<sup>28</sup>, Steve Allen<sup>29</sup>, Angela Allen<sup>30</sup>, Timothy M E Davis<sup>28</sup>, Pascal Michon<sup>27,31</sup>, Ivo Mueller<sup>27,32,33</sup>, Angie Green<sup>1</sup>, Sile Molloy<sup>1</sup>, Kimberly J Johnson<sup>1</sup>, Angeliki Kerasidou<sup>1,34</sup>, Victoria Cornelius<sup>1</sup>, Lee Hart<sup>1</sup>, Aaron Vanderwal<sup>1</sup>, Miguel SanJoaquin<sup>1</sup>, Gavin Band<sup>1</sup>, Si Quang Le<sup>1</sup>, Matti Pirinen<sup>1</sup>, Nuno Sepúlveda<sup>35</sup>, Chris C A Spencer<sup>1,3</sup>, Taane G Clark<sup>35</sup>, Tsiri Agbenyega<sup>12,36</sup>, Eric Achidi<sup>17</sup>, Ogobara Doumbo<sup>7</sup>, Jeremy Farrar<sup>23,24</sup>, Kevin Marsh<sup>18</sup>, Terrie Taylor<sup>37</sup> & Dominic P Kwiatkowski<sup>1-3</sup>

<sup>1</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>2</sup>Wellcome Trust Sanger Institute, Hinxton, UK. <sup>3</sup>Medical Research Council Centre for Genomics and Global Health, University of Oxford, Oxford, UK. <sup>4</sup>Medical Research Council Unit, Serrekunda, The Gambia. <sup>5</sup>Royal Victoria Teaching Hospital, Banjul, The Gambia. <sup>6</sup>Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, London, UK. <sup>7</sup>Malaria Research and Training Centre, Faculty of Medicine, University of Bamako, Bamako, Mali. <sup>8</sup>Dipartimento di Sanita Pubblica e Malattie Infettive, University of Rome La Sapienza, Rome, Italy. <sup>9</sup>Centre National de Recherche et de Formation sur le Paludisme (CNRFP), Ouagadougou, Burkina Faso. <sup>10</sup>Navrongo Health Research Centre, Navrongo, Ghana. <sup>11</sup>Noguchi Memorial Institute for Medical Research, Accra, Ghana. <sup>12</sup>Komfo Anokye Teaching Hospital, Kumasi, Ghana. <sup>13</sup>Department of Molecular Medicine, Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany. <sup>14</sup>Kumasi Centre for Collaborative Research, Kumasi, Ghana. <sup>15</sup>Institute of Child Health, College of Medicine, University of Ibadan, Ibadan, Nigeria. <sup>16</sup>Department of Biochemistry and Molecular Biology, University of Buea, Buea, Cameroon. <sup>17</sup>Department of Medical Laboratory Sciences, University of Buea, Buea, Cameroon. <sup>18</sup>KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya. <sup>19</sup>Faculty of Medicine, Department of Medicine, Imperial College, London, UK. <sup>20</sup>Joint Malaria Programme, Kilimanjaro Christian Medical Centre, Moshi, Tanzania. <sup>21</sup>Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. <sup>22</sup>Malawi-Liverpool Wellcome Trust Clinical Research Programme, College of Medicine, University of Malawi, Blantyre, Malawi. <sup>23</sup>Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam. <sup>24</sup>Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, Oxford, UK. <sup>25</sup>Nossal Institute of Global Health, University of Melbourne, Melbourne, Victoria, Australia. <sup>26</sup>Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam. <sup>27</sup>Papua New Guinea Institute of Medical Research, Goroka, Papua New Guinea. <sup>28</sup>School of Medicine and Pharmacology, Fremantle Hospital Unit, University of Western Australia, Fremantle, Western Australia, Australia. <sup>29</sup>College of Medicine, Swansea University, Swansea, UK. <sup>30</sup>Weatherall Institute of Molecular Medicine, Oxford University, Oxford, UK. <sup>31</sup>Faculty of Health Sciences, Divine Word University, Madang, Papua New Guinea. <sup>32</sup>Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria, Australia. <sup>33</sup>Barcelona Centre for International Health Research (CRESIB), Barcelona, Spain. <sup>34</sup>Ethox Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK. <sup>35</sup>London School of Hygiene and Tropical Medicine, London, UK. <sup>36</sup>School of Medical Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. <sup>37</sup>Blantyre Malaria Project, College of Medicine, University of Malawi, Blantyre, Malawi. Correspondence should be addressed to K.A.R. (kirk.rockett@well.ox.ac.uk) or D.P.K. (dominic.kwiatkowski@well.ox.ac.uk).



## ONLINE METHODS

Cases of severe malaria were recruited on admission to hospital, usually as part of a larger program of clinical research on malaria, designed and led by local investigators. A control group was recruited at each of the study sites to match the ancestry composition of the cases (Supplementary Tables 1, 2 and 4). The control group was intended to be representative of the general population, and cord blood samples were used as controls at several study sites. We describe elsewhere the details of study design at individual sites and local epidemiological conditions, including malaria endemicity (see refs. 21,50–53 and the MalariaGEN website (see URLs)). Following consultation with the Severe Malaria in African Children network and other clinical experts, a standardized case report form was developed to record the clinical features of severe malaria<sup>14</sup> (the case report form can be found on the MalariaGEN website; see URLs). This form was not intended to replace local practice but to encourage uniformity in the core data collected across the different sites. A secure web application was developed to enable investigators to upload and curate their data and to transform it into standardized units and format before releasing it to the consortial database. A data fellow was appointed at each site with responsibility for the process of integrating local clinical data with the central database. As part of a capacity-building program, data fellows received training in data management and analysis. Moreover, ethics advice, support and training/capacity building in ethics were provided to data fellows and partners throughout the life of the study.

The normalized clinical data from each study site were combined to ascertain phenotypes in a standardized manner across the entire data set (Supplementary Table 21). A case of severe malaria was defined as an individual admitted to a hospital or clinic with *P. falciparum* parasites in the blood film and with clinical features of severe malaria as defined by WHO criteria<sup>11,12</sup>. Severe malaria comprises a number of overlapping syndromes, with the most commonly reported being cerebral malaria and severe malarial anemia. In keeping with standard criteria, cerebral malaria was defined here as a case of severe malaria with a Blantyre coma score of <3 for a child or a Glasgow coma score of <9 for an adult. Severe malarial anemia was defined here as a case of severe malaria with a hemoglobin level of <5 g/dl or a hematocrit level of <15%. In this report, we did not attempt to classify other severe malaria syndromes, such as respiratory distress, that are more complicated to standardize among study sites, although they would be present in our data set. Control samples were either collected from cord blood or, if sampled from the local population, were microscopically negative for malaria.

**Ethics.** We All studies were collected under the approval of the appropriate ethics committees, and all participants gave informed consent. Please refer to Supplementary Table 1 and the MalariaGEN website (see URLs) for further details.

**Genotyping.** We selected the Sequenom iPLEX MassARRAY platform for genotyping because of its high-throughput capacity for samples, adaptability for assay design and ability to genotype up to 40 SNPs in a single reaction. We chose to design two iPLEX multiplexes as a compromise between maximizing the number of SNPs we could type on any given sample and minimizing the time and cost to genotype all the samples submitted to the MalariaGEN Resource Centre for the various projects.

Altogether, assays for 89 SNPs were finally designed and tested in 2 rounds of multiplex design on 33,138 samples (Supplementary Fig. 2 and Supplementary Tables 5–7). Of these, 16 SNPs were excluded during the multiplex design and testing phase for a number of reasons: 9 were missing data in more than 20% of the samples; 1 showed a mismatch between the published sequence and the human reference genome; 1 was monomorphic; and 5 could not be redesigned into the final multiplexes (Supplementary Table 7). Of the remaining 73 SNPs genotyped, 55 were included on the basis of a known genetic association with severe malaria (Supplementary Table 5), 3 were used to confirm or type sex and 15 were selected to aid in sample quality control (Supplementary Tables 6 and 22–24, and Supplementary Note). In the quality control phase, 1 of the 55 SNPs with a known association with severe malaria, rs1800750 (*TNF* c.–376G>A), showed a large deviation in Hardy-Weinberg equilibrium and was removed from further analysis (see the ‘Statistical analysis’ section below).

**Samples.** A total of 38,926 individual records comprising 16,433 cases of severe malaria and 22,492 controls were obtained from across the 12 study sites (Supplementary Table 2). Clinical data were missing for sex in 4% of records (we confirmed or typed sample sex by genotyping) and for ancestry in 2% of records. A total of 33,138 samples were genotyped. Each sample was assessed for inclusion in the analysis if it was successfully genotyped at more than 90% of the 65 analysis SNPs (excluding *ATP2B4* SNPs): we excluded 789 samples on this basis (Supplementary Fig. 6). The majority of sample failures were found to be due to blood storage and DNA extraction issues. After quality control of both phenotypic and genotypic data, 11,890 severe malaria cases and 17,441 controls were included for analysis (Table 1).

There were 213 different ancestry groups, of which 41 comprised at least 5% of the individuals at a study site; these included Mandinka, Jola, Wolof and Fula (The Gambia); Bambara, Malinke, Peulh and Sarakole (Mali); Mossi (Burkina Faso); Akan, Frarra, Nankana and Kasem (Ghana); Yoruba (Nigeria); Bantu and Semi-Bantu (Cameroon); Chonyi, Giriama and Kauma (Kenya); Mzigua, Wasambaa and Wabondei (Tanzania); Chewa (Malawi); Madang and Sepik (Papua New Guinea); and Kinh (Vietnam). For purposes of analysis, we classified ancestry groups with a very small sample size (less than 5% of individuals at any study site) as ‘other’ (Supplementary Table 4).

**Statistical analysis.** All statistical analyses were performed using the R statistical software environment (see URLs). As part of the genotyping quality control process, we identified SNPs with large deviations from Hardy-Weinberg equilibrium that might signify assay failure. Overall Hardy-Weinberg equilibrium was assessed from the distribution of Hardy-Weinberg equilibrium *P* values calculated for each SNP by country and ancestry group, after discarding groups where the calculated allele frequency was less than  $5/2N$  (where *N* was the number of individuals in the group)<sup>54</sup>. An assay was marked for potential exclusion if the results deviated from Hardy-Weinberg equilibrium ( $P < 1 \times 10^{-4}$ ) in more than four ancestry groups (Supplementary Fig. 7).

Single-SNP tests, adjusted for HbS genotype, sex and ancestry, for association with severe malaria and the severe malaria subtypes cerebral malaria only and severe malarial anemia only were performed for the 55 SNPs with a known association with severe malaria. Standard logistic regression models were used for tests of association at each autosomal SNP (Supplementary Table 25). Primary analyses comprised tests of association between each SNP and severe malaria phenotypes across all individuals combined as well as separately by sex (X-chromosome SNPs only) and study site: genotypic, additive, dominant, recessive and heterozygote advantage genetic models of inheritance were considered. For X-chromosome SNPs, males were treated as homozygous females. Therefore, when analyzing the males only for X-chromosome SNPs, the genotypic, dominant, recessive and additive models were equivalent and the heterozygous model was redundant: in this case, we present the results from the dominant model (referred to as the male hemizygote model for males at X-chromosome SNPs) and note that the ORs correspond to the change in the odds of disease for males hemizygous for the derived allele in comparison to males hemizygous for the ancestral allele. For combined analyses of males and females at X-chromosome SNPs, robust estimates of variance were used to account for the unequal variance<sup>55</sup>; all models were then appropriate. In secondary analyses, we considered additional genetic models comparing effects between homozygotes and heterozygotes at selected SNPs. ORs and 95% CIs were derived from Wald tests applied to regression coefficients. Significance was assessed using likelihood ratio tests of association, except for in combined analyses of males and females at X-chromosome SNPs where Wald tests were applied using the robust variance estimates. Results are presented with respect to the association between the derived (non-ancestral) allele and the severe malaria phenotype in question.

Standard logistic regression analyses assume that effects are fixed across all sites. To investigate evidence for genetic heterogeneity across severe malaria subtypes both within and across African sites, we compared different models of association in a Bayesian statistical framework. The models we considered comprised fixed, independent or correlated effects between subtypes within a site crossed with fixed, independent or correlated effects of each subtype across all sites. See the supplementary material for further details. For each SNP, we assumed a normally distributed prior on the log OR of association with a mean of 0 and standard deviation  $\sigma$ , where  $\sigma = 1$  for rs334 to reflect

the prior belief that the effect size is large, consistent with the observed ORs of approximately 0.1;  $\sigma = 0.4$  for SNPs found to be significant in fixed-effects analysis; and  $\sigma = 0.2$  otherwise. To model fixed, independent or correlated effects either within or across sites, we set the correlation parameters between subtypes to 1, 0.1 and 0.96, respectively. Multinomial regression was used to make independent maximum-likelihood estimates of the effect of each SNP on these mutually exclusive subtypes for all individuals combined at each African site. Estimates were adjusted for sex, ancestry and, with the exception of rs334, the sickle cell trait. Approximate Bayes factors (ABFs) were then calculated for each SNP and model and used to estimate a posterior probability of each of the models for each SNP.

We also tested for interaction between all pairs of SNPs that were significant in the single-SNP analysis; 25 pairs of markers were tested (**Supplementary Table 20**). We considered two different statistical models of interaction: (i) a 1-degree-of-freedom 'best model' test for the optimal genetic model for each SNP (as defined by association with severe malaria for all individuals across all sites in a fixed-effects model adjusted for ancestry and sex) at each of the interacting loci and (ii) a more general 'genotype' test using a model that allowed for separate effects for heterozygous and homozygous genotypes at each of the interacting loci. At X-chromosome SNPs, male individuals were treated as homozygous females and only additive effects were considered. For a

pair of autosomal SNPs, the genotype test was then a 4-degrees-of-freedom test of interaction; for a pair comprising an autosomal SNP and an X-chromosome SNP, it was a 3-degrees-of-freedom test; and, for a pair of X-chromosome SNPs, it was a 1-degree-of-freedom test. Tests of interaction were performed by testing whether the regression coefficients that represent interaction terms in the corresponding logistic regression model were equal to zero or not. These tests are described in more detail in Cordell<sup>56</sup>.

50. Apinjoh, T.O. *et al.* Association of cytokine and Toll-like receptor gene polymorphisms with severe malaria in three regions of Cameroon. *PLoS ONE* **8**, e81071 (2013).
51. Manjurano, A. *et al.* Candidate human genetic polymorphisms and severe malaria in a Tanzanian population. *PLoS ONE* **7**, e47463 (2012).
52. Dunstan, S.J. *et al.* Variation in human genes encoding adhesion and proinflammatory molecules are associated with severe malaria in the Vietnamese. *Genes Immun.* **13**, 503–508 (2012).
53. Toure, O. *et al.* Candidate polymorphisms and severe malaria in a Malian population. *PLoS ONE* **7**, e43987 (2012).
54. Vallone, P.M., Decker, A.E. & Butler, J.M. Allele frequencies for 70 autosomal SNP loci with U.S. Caucasian, African-American, and Hispanic samples. *Forensic Sci. Int.* **149**, 279–286 (2005).
55. Clayton, D. Testing for association on the X chromosome. *Biostatistics* **9**, 593–600 (2008).
56. Cordell, H.J. Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet.* **10**, 392–404 (2009).