

Functional analysis of genome-wide dataset from 17000 individuals identifies multiple candidate malaria resistance genes enriched in malaria pathogenic pathways

Delesa Damena¹, Francis Agamah¹, Peter O. Kimathi¹, Ntumba E. Kabongo¹, Hundaol Girma¹, Wonderful T. Choga¹, Lemu Golassa³ and Emile R. Chimusa^{1,2*}

¹Division of Human Genetics, Department of Pathology, University of Cape Town, Private Bag, Rondebosch, 7700, Cape Town, South Africa.

²Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Private Bag, Rondebosch, 7700, Cape Town, South Africa.

³Addis Ababa University, Aklilu Lema Institute of Pathobiology, P.O.Box 1176, Addis Ababa, Ethiopia

Abstract

Recent genome-wide association studies (GWASs) of severe malaria have identified several association variants. However, much about the underlying biological functions are yet to be discovered. Here, we systematically predicted plausible candidate genes and pathways from functional analysis of severe malaria resistance GWAS summary statistics (N=17,000) meta-analyzed across eleven populations in malaria endemic regions. We applied positional mapping, expression quantitative trait locus (eQTL), chromatin interaction mapping and gene-based association analyses to identify candidate severe malaria resistance genes. We performed network and pathway analyses to investigate their shared biological functions. We further applied rare variant analysis to raw GWAS datasets (N=11,000) of three malaria endemic populations including Kenya, Malawi and Gambia and performed various population genetic structures of the identified genes in the three populations and global populations.

Our functional mapping analysis identified 57 genes located in the known malaria genomic loci while our gene-based GWAS analysis identified additional 125 genes across the genome. The identified genes were significantly enriched in malaria pathogenic pathways including multiple overlapping pathways in erythrocyte-related functions, blood coagulations, ion channels, adhesion molecules, membrane signaling elements and neuronal systems. Our population genetic analysis revealed that the minor allele frequencies (MAF) of the single nucleotide polymorphisms (SNPs)

residing in the identified genes are generally higher in the three malaria endemic populations compared to global populations. Overall, our results suggest that severe malaria resistance trait is attributed to multiple genes; highlighting the possibility of harnessing new malaria therapeutics that can simultaneously target multiple malaria protective host molecular pathways.

Introduction

Malaria is still one of the global health problems with approximately 228 million cases and 405,000 deaths in 2018 (1). African countries disproportionately carry the global burden of malaria accounting for 93% and 94% of cases and deaths, respectively (1). *P. falciparum* malaria is still one of the leading causes of child mortality in endemic regions, particularly in sub-Saharan Africa. According to the World Health Organization (WHO), malaria killed about 285,000 under five children in 2016 (2). About 10-20% of children who recover from severe malaria develop neurological sequelae and sub-optimal neuronal development (3). Severe malaria (SM) is defined as demonstration of asexual forms of the malaria parasites in the blood of a patient with a potentially fatal manifestation or complication of malaria in whom other diagnosis have been excluded (4). The SM complications include rapid progression to severe malarial anaemia (SMA), hypoglycaemia, cerebral malaria (CM), acidosis and death (4).

The global malaria eradication program, accelerated by the WHO led Roll Back Malaria (RBM) partnership, is largely focusing on scaling up the coverage of the available intervention strategies such as the distribution of long-lasting insecticide treated nets (LLINs), indoor residual insecticide spraying, intermittent treatment for pregnant women in high transmission settings and rapid diagnosis and effective treatments using artemisinin-based combination therapies (ACTs) (5). This led to the significant decline of malaria burden in many parts of the endemic regions. Despite the successes gained in reducing the global burden of malaria, the progress towards global malaria elimination is challenged by wide arrays of problems including decreased funding, lack of political commitments in some countries, emergence of drug resistant parasites and insecticide resistant mosquitoes and lack of effective vaccine among others (6).

P. falciparum has a complex life cycle alternate between vertebrate and female *Anopheles mosquito*. During its blood meal, the infected mosquitoes inoculates the transmissible form the parasite, the sporozoites, into human skin. From the skin the sporozoites enter into the blood

circulation or up-taken by the lymphatic system and invade liver (7). After maturation, the parasite buds off the hepatocytes and released in to the circulation in the form of merozoites containing hundreds of thousands of merozoites that infect erythrocytes (7). The erythrocyte stage also called blood stage lifecycle is complex multi-step process that involve repeated invasion, growth, replication and egress events (8). To create favourable environment for its survival and growth, the parasite remodels the infected erythrocyte (IE). The remodelling events include creation of a parasitophorous vacuolar membrane (PVM) that surrounds the parasite and modification of antigenic and structural properties of the IE by transporting a protein called *P. falciparum erythrocyte membrane protein 1 (PfEMP1)* to the surface (8). The clinical symptoms of the SM have been linked to the blood stage life cycle (8).

Even though SM is one of the commonest reasons for admission to hospital and is a major cause of hospital death in children aged 1-5 years in endemic areas, it constitutes only a small subset (1-2%) of the infected children as the majority of malaria infections is mild (9). It has been shown that such clinical variations is partly attributable to human genetic factors (10, 11). Thus, a comprehensive understanding of the human genetic causes of variation in malaria clinical outcome may potentially provide clues to design new intervention strategies such as therapeutics and vaccines which can facilitate the global malaria eradication program (12, 13).

Aiming at shedding more light to the genetic basis of severe *P. falciparum* malaria, several genome-wide association studies (GWASs) have been conducted in diverse malaria endemic populations over the last decade (14–18). The GWASs have replicated some of the well-known malaria susceptibility genomic risk loci including sickle cell (*HBB*) and *ABO* blood group loci and identified new variants in *ATP2B4* and Glycophorin regions (14–18). Due to the single-marker testing approach commonly used, single SNP-based GWASs may miss candidate variants with weak genetic effects and therefore, combining all effects from multiple variants within a gene and deconvoluting the interactions between genes underlying the malaria resistance trait may provide important insights in to underlying genetics (19). Today, a number of gene-based and pathway-level statistical analytic methods have recently been developed and successfully implemented in complex disease studies (20–23). These methods integrate functional information from advanced biological databases including the Genotype-Tissue Expression (GTEx)(24), Encyclopedia of DNA Elements (ENCODE) (25), Roadmap Epigenomics Project (26) and chromatin interaction

information (27). Owing to the fact that direct functional follow-up of several candidate causal variants and genes is expensive, application of such computational method to prioritize genes and their respective biological pathways are proven to be useful in complex diseases studies (20). Furthermore, gene and gene-set analysis can improve the study power by aggregating the joint effects of weakly associated markers, a common challenge in a polygenic trait studies (28).

Here, we implement several gene-set, pathway and network analytic methods on summary statistics of severe malaria GWAS from 17,000 individuals meta-analysed across eleven populations and systematically predicted plausible genes and pathways. We further performed rare variant analysis on raw GWAS dataset ($N = \sim 11,000$) of Kenya, Gambia and Malawi populations. Finally, we performed population genetic structure analysis of the identified genes in the three malaria endemic countries and across global populations. Established over the course of long co-evolution time, blood stage life cycle of the parasite constitutes the most extensive interplay between host and parasite genomes which leads to the clinical symptoms of SM. Therefore, our results suggest that severe malaria resistance is polygenic and attributed to multiple genes aggregated in pathogenic pathways linked to the erythrocyte stage lifecycle of *P. falciparum*.

Results

Functional mapping and annotations

We applied three functional mapping strategies to the severe malaria GWAS summary statistics ($N=17,000$ samples, ~ 17 million SNPs) meta-analysed across 11 malaria endemic populations in Africa, Asia and Oceania (see **Material and Methods**). We identified 19 lead SNPs out of 69 significant SNPs across 6 genomic loci (**Supplementary Data 1-3**). The genomic locus was defined as the region that contain independent lead SNPs and nominally significant SNPs ($p < 0.05$) in linkage disequilibrium (LD) block with lead SNPs. An independent significant SNP was defined as a genome-wide significant SNP ($P\text{-value} < 5e-8$) within the genomic boundary of LD threshold of $r^2 > 0.6$. Lead SNPs were selected from independent significant SNPs at LD threshold of $r^2 > 0.1$. SNPs in close proximity (< 250 kb) were considered as a single locus and thus, each genomic locus can contain multiple independent significant SNPs and lead SNPs. These SNPs

were significantly enriched in ncRNA-intronic, intronic, intergenic and ncRNA-exonic regions (**Fig1.A**).

Our functional mapping strategies yielded a total of 57 protein-coding genes (**Table 1, Supplementary Data 4**). These include 29, 23 and 14 genes identified by eQTL mapping, chromatin interaction mapping and positional mapping, respectively (**Fig1.B**). Two genes including *ATP2B4* and *HBD* were identified by all the three gene mapping strategies while five genes including *GYPB*, *HBG2*, *TRIM6-TRIM34*, *OR51F2* and *TRIM68* were predicted by two of the three mapping strategies (**Supplementary Data 4**).

The identified genes were enriched in five cytogenic positions including 11p15 ($p=2.65e-18$), chr9q34 ($p=4.63e-9$), chr4p31 ($p=4.6e-8$), chr1q32 ($p=2.6e-7$), chr3q26 ($p=7.97e-4$) (S1 Table). We noted that the majority (33%) of the identified genes were clustered on 11p15 (**Fig 2**). These include beta globin gene cluster: *HBB*, *HBD*, *HBG1*, *HBG2* and *HBE1*; Tripartite motif-containing (*TRIM*) family genes including *TRIM68*, *TRIM21*; and genes involved in olfactory receptors and G protein-coupled signalling (GPCR) such as: *CCKR*, *OR51F2* and *OR51L* (**Supplementary Table 1**). About two third (13/19) of the genes in this locus are in eQTL and chromatin interactions (**Supplementary Data 4 and Supplementary Fig 1**).

All the implicated genes in chr9q34 locus are located outside the genomic risk locus and were identified by eQTL mapping (**Supplementary Data 4, Supplementary Fig 2**). These include surfait gene cluster such as: *SURF2*, *SURF4*, *MED22* and *SURF6*; a metalloprotease gene, *ADAMTS13*; and a gene encoding *ORS* blood group system coding gene (*GBGT1*). In the remaining enriched cytogenic positions, the known genes including *ATP2B4* (chr1q32); *FREM3*, *GYPE* and *GYPB* (chr4p31) were replicated. Other notable genes include *BTG2* a tumour suppressor gene on chr1q32 and *B3GALNT1* on chr3q26 (**Supplementary Data 4**).

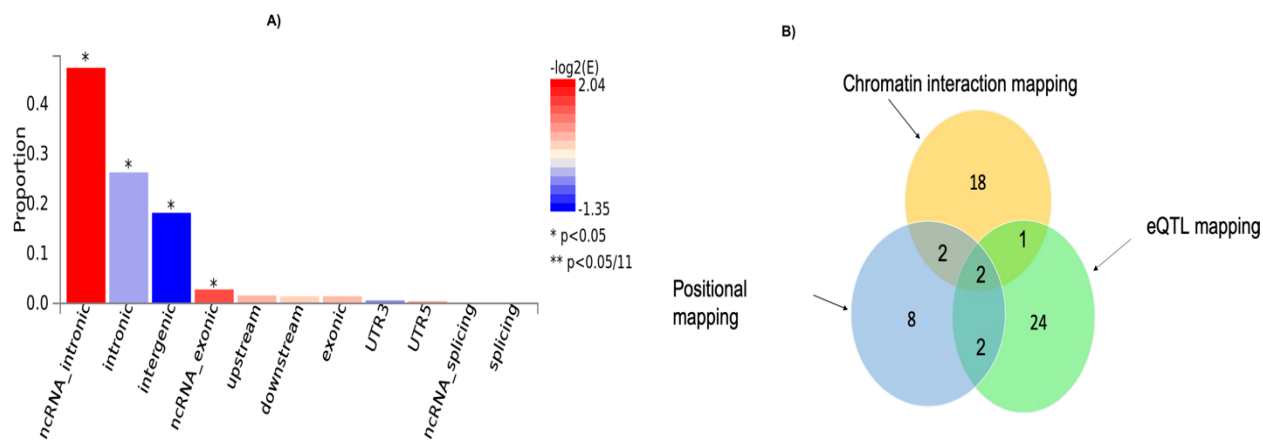


Fig.1. A) Proportion of malaria resistance GWAS SNPs in different genomic annotation categories. **B)** The number of genes identified by each of the three functional mapping strategies including positional mapping, eQTL and chromatin interactions. The intersection sections of the circles depict the number overlapped genes between the respective mapping strategies.

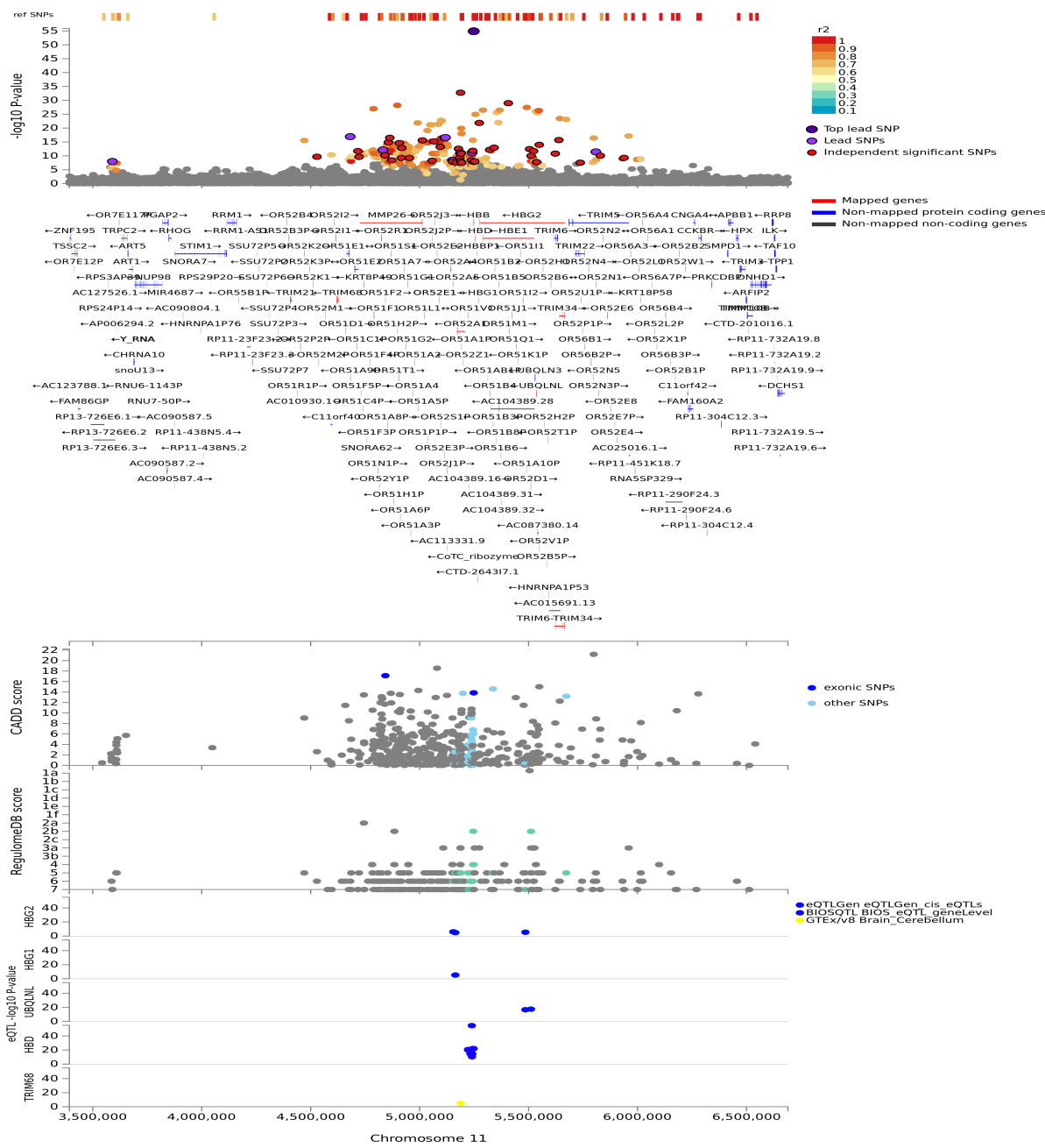


Fig. 2. Regional plot of severe malaria susceptibility GWAS locus on chromosome 11. Non-GWAS-tagged SNPs are shown at the top of the plot as rectangles since they do not have a P-value from the GWAS. Prioritized genes are highlighted in red. eQTLs are plotted per gene and coloured based on tissue types. CADD score, RegulomeDB score and eQTLs, SNPs which are not mapped to any gene are coloured grey.

Table 1. Fifty-seven severe malaria resistance candidate genes identified by eQTL mapping, chromatin interaction mapping and positional mapping strategies implemented in FUMA

Genes -Ensg	Symbols	Chr	Cytoband	Start	End	Biotype	Independent significant SNPs
ENSG00000160323	<i>ADAMTS13</i>	9	q34.2	136279478	136324508	protein coding	rs8176751; rs687621
ENSG00000197859	<i>ADAMTS2</i>	9	q34.2	136397286	136440641	protein coding	rs8176751
ENSG00000179674	<i>ARL14</i>	3	q25.33	160394948	160396233	protein coding	rs116423146
ENSG00000058668	<i>ATP2B4</i>	1	q32.1	203595689	203713209	protein coding	rs4951370
ENSG00000169255	<i>B3GALNT1</i>	3	q26.1	160801671	160823172	protein coding	rs116423146
ENSG00000159388	<i>BTG2</i>	1	q32.1	203274619	203278730	protein coding	rs4951370
ENSG00000110148	<i>CCKBR</i>	11	p15.4	6280966	6293357	protein coding	rs113892119;rs28576676
ENSG00000133063	<i>CHIT1</i>	1	q32.1	203181955	203242769	protein coding	rs4951370
ENSG00000113758	<i>DBN1</i>	5	q35.3	176883609	176901402	protein coding	rs687621
ENSG00000122176	<i>FMOD</i>	1	q32.1	203309756	203320617	protein coding	rs4951370
ENSG00000183090	<i>FREM3</i>	4	q31.21	144498455	144621828	protein coding	rs201510180
ENSG00000109458	<i>GAB1</i>	4	q31.21	144257915	144395721	protein coding	rs111374053
ENSG00000148288	<i>GBGT1</i>	9	q34.2	136028340	136039332	protein coding	rs687621
ENSG00000250361	<i>GYPB</i>	4	q31.21	144917257	145061844	protein coding	rs201510180
ENSG00000197465	<i>GYPE</i>	4	q31.21	144792020	144826716	protein coding	rs34330779
ENSG00000244734	<i>HBB</i>	11	p15.4	5246694	5250625	protein coding	rs334
ENSG00000223609	<i>HBD</i>	11	p15.4	5253908	5256600	protein coding	rs334; rs4290259; rs79681613; rs113892119;rs28576676
ENSG00000213931	<i>HBE1</i>	11	p15.4	5289582	5526847	protein coding	rs145843585
ENSG00000213934	<i>HBG1</i>	11	p15.4	5269313	5271122	protein coding	rs7927066
ENSG00000196565	<i>HBG2</i>	11	p15.4	5274420	5667019	protein coding	rs145843585; rs183322782; rs148179286;rs7927066;rs11037724
ENSG00000203813	<i>HIST1H3H</i>	6	p22	27777842	27778314	protein coding	rs8176751
ENSG00000122188	<i>LAX1</i>	1	q32.1	203734304	203745361	protein coding	rs4951370
ENSG00000148297	<i>MED22</i>	9	q34.2	136205160	136214986	protein coding	rs8176751; rs687621
ENSG00000108960	<i>MMD</i>	17	q22	53469974	53499353	protein coding	rs8176751
ENSG00000167346	<i>MMP26</i>	11	p15.4	4726157	5013659	protein coding	rs141862673; rs145429724
ENSG00000169251	<i>NMD3</i>	3	q26.1	160822484	160971320	protein coding	rs116423146
ENSG00000184881	<i>OR51B2</i>	11	p15.4	5344541	5345582	protein coding	rs145843585
ENSG00000176925	<i>OR51F2</i>	11	p15.4	4842551	4843686	protein coding	rs141862673; rs145429724; rs113892119;rs28576676
ENSG00000176798	<i>OR51L1</i>	11	p15.4	5020213	5021160	protein coding	rs113892119;rs28576676
ENSG00000182070	<i>OR52A1</i>	11	p15.4	5172239	5207612	protein coding	rs116780407
ENSG00000228474	<i>OST4</i>	2	p23.3	27293340	27294641	protein coding	rs8176751
ENSG00000142657	<i>PGD</i>	1	p36.22	10458649	10480201	protein coding	rs687621
ENSG00000143850	<i>PLEKHA6</i>	1	q32.1	204187979	204346793	protein coding	rs4951370
ENSG00000163590	<i>PPM1L</i>	3	q25.33	160473390	160796695	protein coding	rs116423146
ENSG00000188783	<i>PRELP</i>	1	q32.1	203444956	203460480	protein coding	rs4951370
ENSG00000170955	<i>PRKCDBP</i>	11	p15.4	6340176	6341877	protein coding	rs28576676
ENSG00000160271	<i>RALGDS</i>	9	q34.2	135973107	136039301	protein coding	rs8176751; rs687621

ENSG00000148300	<i>REXO4</i>	9	q34.2	136271186	136283164	protein coding	rs8176751
ENSG00000080345	<i>RIF1</i>	2	q23.3	152266397	152364527	protein coding	rs8176751
ENSG00000170153	<i>RNF150</i>	4	q31.21	141780961	142134031	protein coding	rs111374053
ENSG00000136193	<i>SCRN1</i>	7	p14.3	29959719	30029905	protein coding	rs687621
ENSG00000160326	<i>SLC2A6</i>	9	q34.2	136336217	136344259	protein coding	rs8176719; rs687621
ENSG00000196542	<i>SPTSSB</i>	3	q26.1	161062580	161090668	protein coding	rs116423146
ENSG00000148290	<i>SURF1</i>	9	q34.2	136218610	136223552	protein coding	rs8176751; rs687621
ENSG00000148291	<i>SURF2</i>	9	q34.2	136223428	136228045	protein coding	rs8176751
ENSG00000148248	<i>SURF4</i>	9	q34.2	136228325	136242970	protein coding	rs8176751
ENSG00000148296	<i>SURF6</i>	9	q34.2	136197552	136203235	protein coding	rs8176751; rs687621
ENSG00000196628	<i>TCF4</i>	18	q21.2	52889562	53332018	protein coding	rs687621
ENSG00000132109	<i>TRIM21</i>	11	p15.4	4406127	4414926	protein coding	rs28576676
ENSG00000132274	<i>TRIM22</i>	11	p15.4	5710919	5758319	protein coding	rs28576676
ENSG00000258659	<i>TRIM34</i>	11	p15.4	5640994	5665628	protein coding	rs183322782; rs148179286
ENSG00000213186	<i>TRIM59</i>	3	q25.33	160150233	160203561	protein coding	rs116423146
ENSG00000121236	<i>TRIM6</i>	11	p15.4	5617339	5634188	protein coding	rs28576676
ENSG00000258588	<i>TRIM6- TRIM34</i>	11	p15.4	5617955	5665628	protein coding	rs183322782; rs148179286; rs28576676
ENSG00000167333	<i>TRIM68</i>	11	p15.4	4619902	4629489	protein coding	rs10837488; rs4290259;rs113892119;rs28576676
ENSG00000175518	<i>UBQLNL</i>	11	p15.4	5535623	5537935	protein coding	rs11037724
ENSG00000109445	<i>ZNF330</i>	4	q31.21	142142041	142155851	protein coding	rs111374053

Candidate genes identified by gene-based GWAS analysis

Taking the polygenic nature of severe malaria resistance trait in to consideration (11), we applied a pathway scoring algorithm(Pascal) (22) method that aggregates all SNPs within a gene and capture polygenic effects at the gene level (see **Material and Methods**). The Pascal analysis replicated 13 genes that were identified by our functional annotation methods in malaria genomic risk loci (**Supplementary Table 2**) and identified 125 additional genes across the genome (**Supplementary Data 5**). The genes with top scores outside genomic risk loci include *CSMD1* ($p=1.58e-12$) on chr8p23.2 and *RBFOX1* ($p=9.76e-11$) on chr16p13.3. *CSMD1* is an important regulator of complement activation and inflammation (29, 30) while *RBFOX1* encodes for an mRNA-splicing factor linked to autism spectrum disorders (31). A previous study in Tanzanian population reported association of variants in *RBFOX* gene with SM (17).

Other important genes identified by gene-based GWASs include neural adhesion molecules including *CNTN4* ($p=3.88e-9$) on chr3p26.3-p26.2; *PCSK5* ($p=2.88e-11$) on chr9q21.13; *CDH13* ($p=4.19e-8$) on chr16q23.3) and *TMEM132* ($p=2.18e-8$) on chr17q12 (**Supplementary Data 5**). These genes were reported to be linked to autism spectrum disorders and other neurodevelopmental conditions (32). Furthermore, protein kinases including *FLT4* ($p=9.96e-8$) on 5q35.3; and *PTPRT* ($p=4.92e-7$) on chr20q12-q13 and *PRKG1* ($p=1.2e-6$) on 10q11.2-q21.1 were among the genes with top scores (**Supplementary Data 5**). *PTPRT* is a tyrosine phosphatase receptor involved in STAT3 pathway and was recently reported to be associated with mild malaria susceptibility in Benin populations (33). *PRKG1* is a cyclic guanosine monophosphate (GMP) dependent protein kinase which plays important roles in relaxation of vascular smooth muscle and inhibition of platelet aggregation (34). *FLT4* acts as a cell-surface receptor for vascular endothelial growth factor C (VEGFC) and vascular endothelial growth factor D (VEGFD), and plays an essential role in the development of the vascular network (35). It has been shown that VEGF and its receptor-related molecules are expressed in the brain tissues and reported to play protective during CM (36).

Gene-based rare variant association

Because rare variants are known to play role in the variation of most complex traits, we applied optimal unified sequence kernel association test (SKAT-O), which combines burden and variance-component analyses (37) to the raw genotype GWAS dataset of Gambia, Kenya and Malawi populations (**see Material and Methods**). The SKAT-O analysis identified a total of six and nine nominally significant genes in Gambia and Malawi populations, respectively. These include nine long intergenic non-protein coding RNAs (LincRNAs), *MIR4282*, *GLYR1*, *NDNF*, *EPB41L2*, *ATP8A1* and *WASF3* (**Supplementary Table 3**). However, none of these genes were significant after correction for multiple testing.

Functional networks and subnetworks of severe malaria resistance candidate genes

To investigate the functional interaction between all the candidate malaria resistance candidate genes identified in this study, we implemented network analysis (**see Materials and Methods**). Our global network generated 351 functional interactions between 268 genes. Topology analysis identified *ABO*, *HBB*, *HBD*, *HBE1*, *ATP2B4* as highly influential connector hub genes influencing at least two subnetworks/communities while *TRIM21* and *OR5F2* constituted independent communities. *MED22* and *OR551B6* constituted provincial hub genes (**Fig 3**).

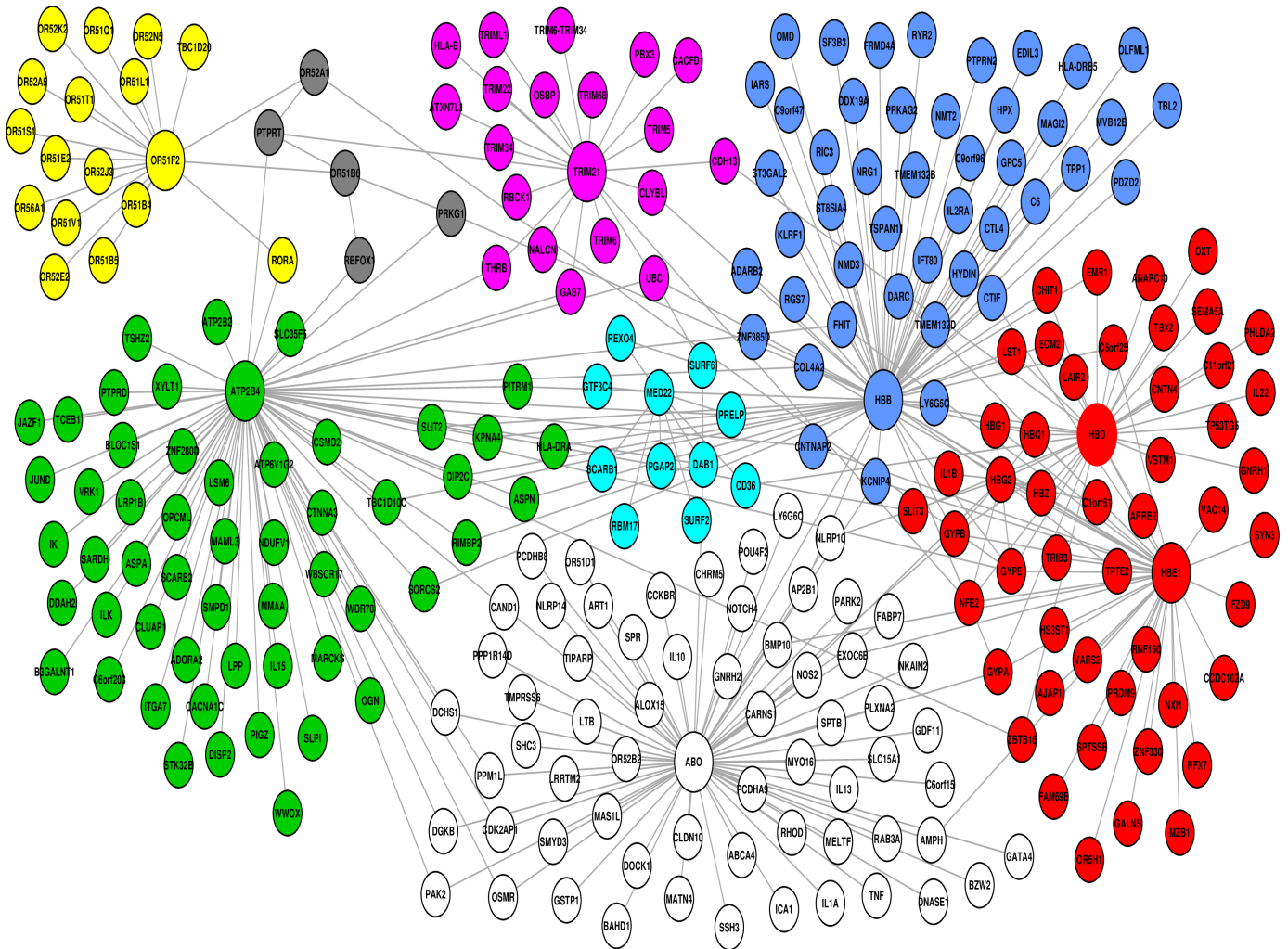


Fig. 3. Network generated from predominant severe malaria protective candidate genes, comprising of 351 interactions between 268 nodes. Topology analysis identified *ABO*, *HBB*, *HBD*, *HBE1*, *ATP2B4* as highly influential connector hub genes influencing at least two subnetworks/communities while *TRIM21* and *OR5F2* constituted independent communities. *MED22* and *OR551B6* constituted provincial hub genes.

Molecular functions and pathways of candidate severe malaria resistance genes

To test whether the genes predicted by the three functional mapping strategies overlapped in functional gene sets and pathways, we conducted gene enrichment analysis implemented in FUMA (20) using MsigDBc5 (38) gensets as background (see **Material and Methods**). The gene enrichment analysis identified several shared biological functions linked to erythrocyte-related pathways including three gene ontology (GO) cellular components, eight GO molecular functions and fourteen GO biological processes (**Table 2**). The implicated cellular components include haptoglobin-haemoglobin complex ($p=7.6e-8$), haemoglobin-complex ($p=7.63e-8$) and cytosolic-part ($p=6.73e-3$). The enriched molecular functions include haptoglobin binding ($p=4.87e-8$), oxygen carrier activity ($3.8e-7$), oxygen binding ($p=4.87e-5$) and other activities related to haemoglobin functions. The shared biological activities include oxygen transport ($p=4.22e-6$), gas transport ($p=4.87e-6$), hydrogen peroxide catabolism ($p=9.08e-5$), protein hetero-oligomerization ($p=2.86e-3$), protein complex-oligomerization ($p=4.23e-3$), interferon gamma mediated signalling pathways ($p=3.09e-2$) and blood coagulation ($p=3.09e-2$).

In addition to genes within severe malaria genomic risk loci, we performed functional analysis and pathway analysis for the genes identified by the gene-based GWAS using Database for Annotation, Visualization and Integrated Discovery (DAVID) method (39) and Pascal (22), respectively (see **Material and Methods**). The DAVID analysis yielded eight functional categories, the majority of which are linked to malaria pathogenesis (**Table 3**) including GPCR signalling, membrane/transmembrane proteins, Na⁺/K⁺ transporting ATPases, cell adhesions, haemoglobin related functions, calcium signalling and actin binding activities.

Similarly, Pascal analysis implicated eleven significant pathways the majority of which are linked to malaria pathogenesis in RBCs, vasculatures and brain (**Fig 4A**). These include G protein-coupled receptor signalling ($p=7.88e-15$), haemostasis ($p=4.52e-10$), neuronal system ($p=1.25e-9$), axon guidance ($p=5.93e-8$), calcium signalling ($p=1.10e-7$), chemical transmission across synapses ($p=1.75e-7$), immune system ($p=2.61e-6$), signalling by Rho GTPase ($p=1.45e-5$) and

tight junction ($p=3.57 \times 10^{-5}$). Furthermore, differential expression of genes implicated by Pascal method showed significant enrichment in blood vessels (**Fig 4B**).

Table 2. Gene enrichment results of functionally annotated genes in malaria genomic risk loci identified by FUMA method

GO terms	GeneSet	N. genes	N. enriched genes	P-value	adjusted P-value	Genes
Cellular components	<u>Haptoglobin-haemoglobin complex</u>	11	5	8.91e-11	7.63e-8	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Haemoglobin complex</u>	12	5	1.52e-10	7.63e-8	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Cytosolic part</u>	239	7	6.73e-6	2.25e-3	<i>HBB, HBD, HBG1, HBG2, HBE1, DBN1, SURF6</i>
Biological functions	<u>Haptoglobin binding</u>	10	5	4.87e-11	8.02e-8	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Oxygen carrier activity</u>	14	5	3.84e-10	3.16e-7	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Oxygen binding</u>	36	5	6.87e-8	3.77e-5	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Molecular carrier activity</u>	41	5	1.35e-7	5.56e-5	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Oxidoreductase activity acting on peroxide as acceptor</u>	56	5	6.66e-7	2.19e-4	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Haemoglobin binding</u>	7	3	8.69e-7	2.38e-4	<i>HBB, HBD, HBE1</i>
	<u>Antioxidant activity</u>	85	5	5.35e-6	1.26e-3	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Tetrapyrrole binding</u>	136	5	5.23e-5	1.08e-2	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
Biological Processes	<u>Oxygen transport</u>	15	5	5.74e-10	4.22e-6	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Gas transport</u>	19	5	2.20e-9	8.10e-6	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Hydrogen peroxide catabolic process</u>	32	5	3.71e-8	9.08e-5	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Antibiotic catabolic process</u>	50	5	3.74e-7	6.88e-4	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Drug catabolic process</u>	108	6	8.05e-7	1.18e-3	<i>CHIT1, HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Cofactor catabolic process</u>	66	5	1.52e-6	1.87e-3	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Protein hetero-oligomerization</u>	133	6	2.73e-6	2.86e-3	<i>HBB, HBD, HBG1, HBG2, HBE1, HIST1H3H</i>
	<u>Protein complex oligomerization</u>	551	10	4.60e-6	4.23e-3	<i>TRIM21, HBB, HBD, HBG1, HBG2, HBE1, TRIM6, TRIM34, TRIM22, HIST1H3H</i>
	<u>Antibiotic metabolic process</u>	91	5	7.49e-6	6.11e-3	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Cellular detoxification</u>	107	5	1.65e-5	1.21e-2	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Protein trimerization</u>	54	4	2.00e-5	1.34e-2	<i>TRIM21, TRIM6, TRIM34, TRIM22</i>
	<u>Detoxification</u>	122	5	3.11e-5	1.91e-2	<i>HBB, HBD, HBG1, HBG2, HBE1</i>
	<u>Interferon gamma mediated signalling pathway</u>	70	4	5.60e-5	3.09e-2	<i>TRIM21, TRIM68, TRIM34, TRIM22</i>
	<u>Coagulation</u>	335	7	5.89e-5	3.09e-2	<i>HBB, HBD, HBG1, HBG2, HBE1, HIST1H3H, ADAMTS13</i>
	<u>Oxygen transport</u>	15	5	5.74e-10	4.22e-6	<i>HBB, HBD, HBG1, HBG2, HBE1</i>

Table 3: Functional categories of genes identified by gene-based GWAS analysis grouped using DAVID method

Functional group	Genes	Enrichment Score
GPCR signalling pathways and olfactory receptors	<i>OR51B6, FZD10, TMEM132C, OR51V1, OR51B2, OR52K2, SORCS2, OR56A1, OR52B4, TMEM132D, OR51E2, OR51T1, OR51B4, OR51B5, VSTM1, OR51F2, THSD7B</i>	3.15
Transmembrane protein and, Na ⁺ /K ⁺ transporting ATPase	<i>TSPAN11, VSTM1, TMEM132D, SURF4, NKAIN2, TMEM132C, EVC, SLC35F3</i>	3.07
Tyrosine phosphatase, tyrosine kinase, cell adhesion molecule-like	<i>OPCML, PTPRD, PTPRT, CNTN5, FLT4, NTM, CNTN4, PTPRN2, VSTM1, PTPRS</i>	2.51
Haemoglobin related activities	<i>HBG2, HBE1, HBB, HBD</i>	2.42
Sodium leak and Potassium channel interacting protein	<i>KCNIPI, NALCN, KCNIP4, KCTD1</i>	1.14
zinc finger protein	<i>SMYD3, TSHZ2, ZNF385B, ZNF385D</i>	0.61
Actin binding LIM protein family and RAR-related orphan receptor A	<i>RORA, THRB, GLIS3, ABLIM2</i>	0.45
Calcium/calmodulin-dependent protein kinase, cGMP-dependent kinase and fms-related tyrosine kinase	<i>PRKG1, CAMK1D, FLT4, VRK1</i>	0.44

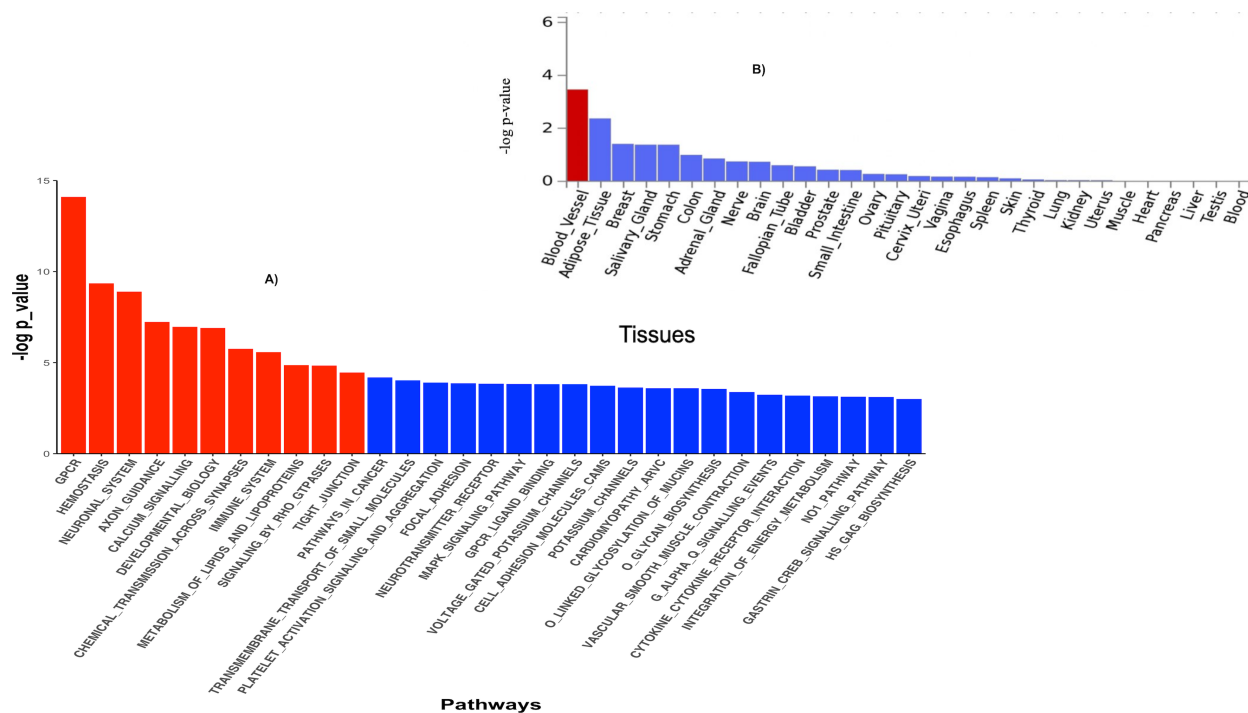


Fig. 4. A) Pathway scores obtained from pascal analysis. Significant Pathways at Bonferroni corrected $P\text{-value} \leq 0.05$ are coloured in red. **B)** Tissue specific gene expression and shared biological functions of genes identified by pascal in GTEx v8 30 general tissue types. Input genes were tested against each of the precomputed differentially expressed sets using the hypergeometric test. Significant enrichment at Bonferroni corrected $P\text{-value} \leq 0.05$ are coloured in red.

Population genetic structure of malaria resistance candidate genes

We noted that the minor allele frequencies (MAF) of the SNPs residing in the identified genes are generally higher in the three malaria populations compared to 20 ethnic groups (**Supplementary Fig 1 and 2**). We further observed that the proportion of pathogenic SNPs in a total of eighteen genes is much higher in the three malaria endemic populations compared to other populations (**Supplementary Data 6 and 7**). These include *TRIM* family genes such as *TRIM21*, *TRIM22*, *TRIM68*, *TRIM6-TRIM34* and *TRIM34* in which the pathogenic SNP proportion ranges from 13.3-25%; olfactory receptors genes such as: *OR51B4*, *OR51B6*, *OR51B2*, *OR56A1*, *OR51L1*, *OR52K2* and *OR51E2* in which the pathogenic SNP proportion ranges from (27.3 -100%).

The principal component analysis based on the SNPs residing in the identified genes effectively clustered all the populations according their ancestry (**Fig 5**).

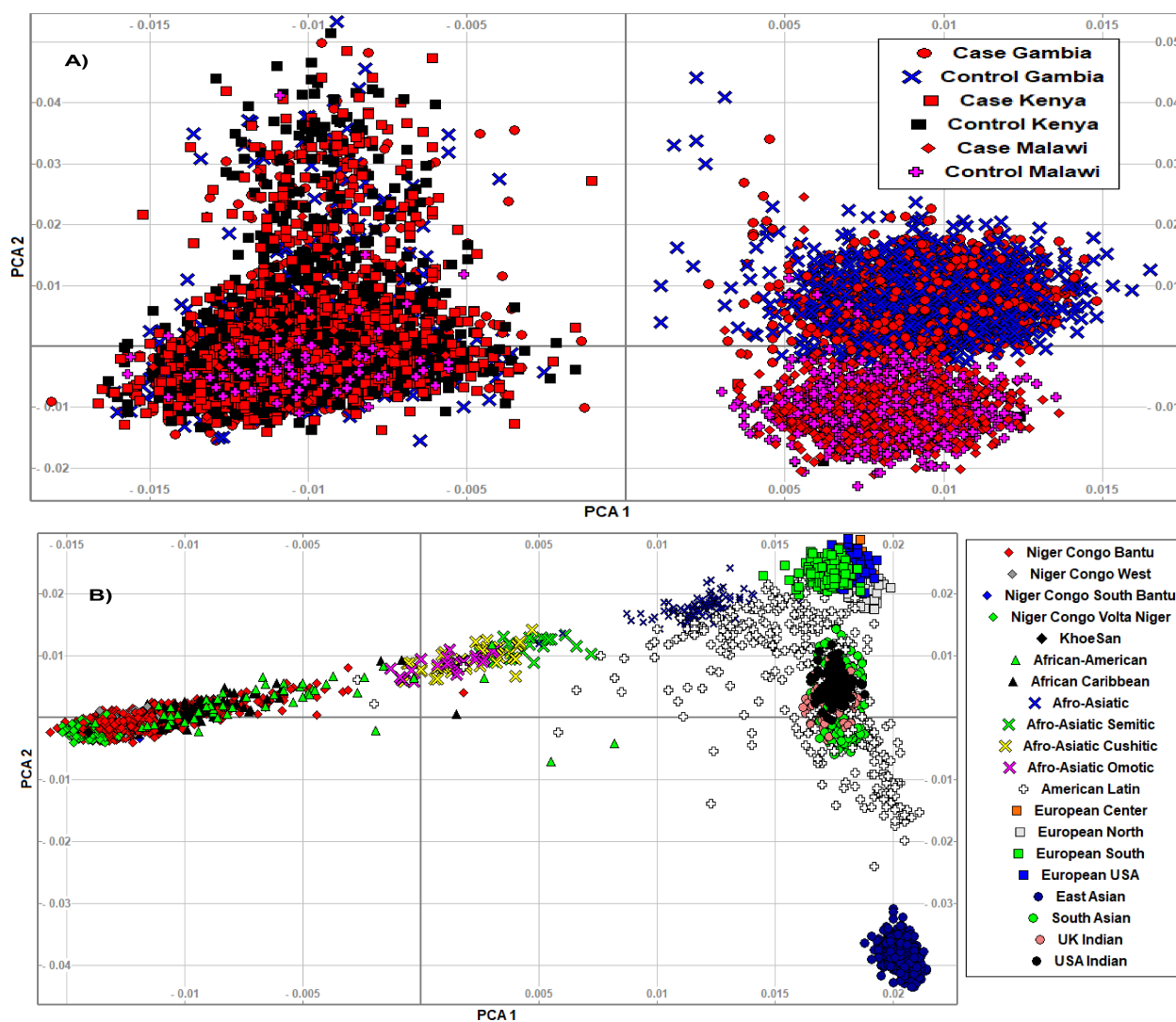


Fig .5. A) We clustered the merged malaria GWAS dataset containing only SNPs residing in the identified malaria resistance candidate genes (N= 10578 samples, 15,675 SNPs) using smartpca software. The three populations and their case/control status were indicated by different colours and symbols. **B).** We clustered the AGVP dataset containing only SNPs residing in the identified malaria resistance candidate genes (N=4932 samples, 93,5549 SNPs) in to sub-regions/populations using smartpca software. The populations were indicated by different colours and symbols.

Discussions

In this study, we applied statistical functional analytic method to the largest ever severe malaria susceptibility GWAS dataset and identified the well-known malaria resistance loci and a number of novel genes that can guide future functional experiments. We noted that severe malaria resistance is attributed to multiple genes and pathways linked to malaria pathogenesis during blood stage life cycle of the parasite including merozoite invasion, parasite growth, cytoadherence, and signal transduction. The genes that were identified by our three mapping strategies might have equal importance; genes that were identified by positional mapping may act at protein level through structural changes while the genes identified by eQTL and chromatin interactions exert their influences through quantitative changes at gene expression levels (20).

The fact that the functionally mapped genes are clustered on chromosome 11p15 is consistent with our recent work in which we reported the disproportionate concentration of SNP-heritability on chromosome eleven. This might reinforce the need for targeting this chromosome in the future severe malaria susceptibility studies (11). We noted that in addition to the sickle trait gene (*HBB*), our mapping strategies identified other members of beta globin gene cluster that cause various forms of beta-thalassemia (*HBE1*, *HBD*, *HBG* and *HBG2*).

Our network analysis showed that all these genes constituted hub with which several other genes are connected; which reaffirm the importance of hemoglobinopathies in resistance against severe malaria. Beta-thalassemia and other hemoglobinopathies are thought to confer protection against severe malaria by suppressing the parasite growth and by mitigating associated pathogenic effects (40). The proposed protective mechanisms of hemoglobinopathies against severe malaria has been extensively reviewed elsewhere (40, 41). In addition to the beta globin gene cluster, the Tripartite motif (*TRM*) containing gene family (*TRIM68* and *TRIM21*) identified in this locus are known to play critical role in down regulating Toll-like receptors(TLR)- and Rig-like receptors (RLR)-induced responses and protect from autoimmune and inflammations (42, 43). Mal-adapted inflammatory reactions is one of the hall mark pathogenic pathways in severe malaria (44, 45). The olfactory receptors super-family genes (*CCKR*, *OR51F2* and *OR51L*) identified in this locus might involve in G protein-coupled receptors(GPCR) signalling activities which is important in blood stage life cycle of *P. falciparum* (46). However, it is also possible that these genes were

detected because of their abundance and close proximity to with globin gene cluster(47). We also be noted that the genes in the *TRM* family and olfactory receptors super-family contain higher proportion of pathogenic SNPs in the three-malaria endemic population compared to the global populations. The majority of the well-known malaria protective genes have deleterious variants and were evolved under balancing selections [42].

Our eQTL mapping identified *ADAMTS13* gene on chr9q34 outside the malaria genomic risk locus which would not have been identified by the conventional SNP mapping approach (20). *ADAMTS13* is a zinc-containing metalloprotease enzyme that cleaves, von Willebrand factor vWF, a large protein derived from endothelial surface and megakaryocytes which plays a crucial role in basic haemostasis (48, 49). Following activation of endothelial cells, vWF is directly released in to plasma and basement membrane or is stored in Weibel–Palade bodies (WPBs) from where it is released by regulated secretion to promote adhesion of platelets at the sites of vascular injury and facilitate vascular healing (49). However, abnormal accumulations of vWF caused by deficiency of plasma *ADAMTS13* trigger intravascular platelet aggregation and micro thrombosis leading to a vascular disease, Thrombotic Thrombocytopenic Purpura (TTP) (48). Indeed, recent works have linked the platelet-mediated clumping of infected erythrocytes in microvasculature during cerebral malaria with increased level of VWF in plasma caused by mutations in *ADAMTS13* genes (50–52).

In the same genomic locus, our eQTL functional mapping identified surfeit gene cluster, metalloprotease genes linked to epithelial adhesions and blood coagulations. *SURF4* gene has been implicated in epithelial cell adhesion trait (53), *MED22* gene is linked to VWF factor/factor VIII level measurement (54) and *SURF6* has been implicated in epithelial ovarian cancer (55). Our network analysis showed that *MED22* forms a central hub to which the rest of surfeit gene cluster are connected. This may suggest that the greater importance of *MED22* gene compared with the other members of the cluster. However, further comprehensive studies are needed to better understand the association of these genes with SM resistance. In the remaining genomic risk loci, the well-known genes including *ATP2B4* ([chr1q32](#)), *FREM3*, *GYPE* and *GYPB* (chr4p31) were replicated and few additional genes were identified. The glycoporphin gene cluster, *GYPB* and *GYPB* encode the MNS blood group system and are host-erythrocyte receptors for *P. falciparum*; suggesting that polymorphisms in these genes play protective role by interfering with the invasion

processes [55]. *ATP2B4* variants may impair the parasite lifecycle in erythrocyte by affecting intracellular Calcium homeostasis (18, 56).

Novel genes identified in the loci include *BTG2* on chr1q32 and *B3GALNT1* on chr3q26. *BTG2* is a tumour suppressor gene known to be linked to RBC related traits including MCHC level, RBC distribution and reticulocyte count (57). *B3GALNT1* encodes globoside blood group system which is determined by P antigen (58). Globoside/P antigen is the most abundant neutral glycolipid in the erythrocyte membrane and has been recognized as a cellular receptor for parvo-B19 virus (59). Individuals lacking this receptor are resistant to parvo-B19 virus and uro-pathogenic *E.coli* infections (59, 60). Further investigations are needed to establish the link between these genes and SM resistance.

We noted that the mapped genes share cellular components including haptoglobin binding, haemoglobin complex and cytosolic part and several overlapped molecular functions and biological processes linked with the blood stage life cycle of the parasite. *P. falciparum* spends most of its lifecycle within RBCs, where it undergoes multiple rounds of invasion, growth, replication and egress; causing the signs and symptoms of malaria (44, 61). The majority of the classical haemoglobin variants confer protection against severe malaria by restricting invasion process and intraerythrocytic growth of the parasite (40). Haptoglobin is an acute phase glycoprotein present in human plasma. It forms stable complexes with extracellular haemoglobin that is released from lysed RBCs and thereby curtail the haemoglobin-induced oxidative tissue damage (62).

P. falciparum ingest the host cell cytosol to obtain nutrients and space for growth in the RBCs (63). A recent study showed that the host cytosol uptake process is mediated by parasite's protein called *VPS45* (64). The fact that the identified candidate genes in this study were enriched in the cytosol part of the cellular component might suggest that these genes might arrest the nutrient uptake of the parasites and thereby confer protection against their pathogenic effects. In addition to haemoglobin related functions, some of the candidate genes were enriched in other pathways linked to malaria pathogenesis including blood coagulation related processes malaria (65, 66) and interferon gamma mediated signalling pathways (67, 68). This may suggest that the host genetic factors might interfere with parasite development and its pathogenic outcomes at multiple levels to confer protection the life-threatening form of malaria.

Our gene-based GWAS analysis replicated the well-known malaria resistance candidate genes in the genomic risk loci and identified several genes across the genome. Genes with top scores encode for different malaria relevant functions such as: regulation of inflammation (*CSMD1*), neural adhesion (*CNTN4*, *PCSK5*, *CDH13*, *TMEM132*), vascular epithelial development (*FLT4*) and protein kinases (*PTPRT*, *PRKG1*). Furthermore, functional analysis of the candidate genes yielded functional categories linked with the blood stage lifecycle of the parasite and associated pathologies. The top enriched functions include GPCR signalling, membrane/transmembrane proteins, [Na⁺/K⁺] transporting ATPases, Sodium leak and Potassium channel interacting proteins, cell adhesions molecules and Calcium/Calmodulin-dependent protein kinases (CDPKs). CDPKs have crucial functions in calcium signalling at various stages of the parasite's life cycle and is proposed to be one of the potential drug targets against malaria (69). It has been shown that *P. falciparum* infection activates host signalling pathway involving protein kinase C (PKC) (70). Similarly, host GPCR signalling pathways have been shown to play vital roles in invasion, intra-erythrocyte parasite development and egress processes (71, 72); suggesting the existence of substantial interactions between host membrane/transmembrane signalling and parasite signalling elements which might mediate the disease severity. Further studies are needed to decouple the host-parasite interface of signal transduction and explore the potential target for new therapeutics. Sodium leak and Potassium channel interacting proteins, [Na⁺/K⁺] transporting ATPases play critical role in maintaining electrochemical equilibrium in normal erythrocytes. However, upon invasion by trophoblast stage of the parasite, the ion pump-leak balance is perturbed; with increased leak rate and decreased pump rate resulting in a remarkable increase in [Na⁺] and decrease in [K⁺] in the erythrocyte cytosol (73, 74). This results in formation of a new permeability pathway (NPP) in the erythrocyte membrane which allow the transport of nutrients and waste products necessary for the parasite. The composition and physiological role of NPP has been reviewed elsewhere (75, 76). Studies have shown that both parasites driven proteins encoded by *clag3.1* and *clag3.2* (77, 78) and host benzodiazepine receptor mediate the formation of NPP (79, 80). Our result may suggest the existence of multiple host genes that are involved in mediating this process. Given that TPP can be a potential target for new therapeutics, further studies are needed to investigate the role of host and parasites genetics in mediating the channel and its pathophysiology.

In addition to the well-known pathways such as: haemostasis and immune system that play role in malaria resistance (13), our analysis identified novel pathways including cell adhesion molecules, Rho GTPase activities, tight junction, neuronal system and axon guidance. One of the key virulence factors of *P. falciparum* is its capacity to modify iRBCs to adhere to the endothelium of the vasculature and, thereby, sequester in capillaries and postcapillary venules in vital organs leading to severe disease manifestations (61). Adhesion phenotype is mediated by expression of *P. falciparum* erythrocyte membrane protein 1 (*PfEMP1*) on the iRBCs (81, 82). The binding of iRBCs with endothelium involve various adhesion molecules including CD36, ICAM-1, E-selectin and chondroitin sulfate A (CSA) that are variably expressed in different organs (83–85). The neural adhesion molecules identified in the current study might involve in receptor activities and their polymorphisms might play protective roles against SM. Furthermore, adhesion events have been shown to activate Rho kinase signalling pathway which is strongly implicated in various vascular diseases (86). The genes that are enriched in these pathways might provide protection against severe malaria by weakening the cytoadherence interactions and associated pathologies. Other genes that are enriched in neuronal system, axon guidance and tight junction might be linked with intra cerebral pathogenesis of SM (87). Furthermore, the candidate malaria resistance genes identified by gene-based GWAS were differential expressed in blood vessels; suggesting that the majority of the identified genes are likely counteract *P. falciparum* induced endothelial disfunctions in microvasculature and capillaries (44).

In conclusion, our functional mapping analysis identified 57 genes located in the known malaria genomic loci while our gene-based GWAS analysis identified additional 125 genes across the genome which can potentially guide future experimental studies. The identified genes were significantly enriched in malaria pathogenic pathways including multiple overlapping pathways in erythrocyte-related functions, blood coagulations, ion channels, adhesion molecules, membrane signaling elements and neuronal systems. Overall, our results suggest that severe malaria resistance trait is attributed to multiple genes that are enriched in overlapping pathways linked to severe malaria pathogenesis; highlighting, the possibility of harnessing new malaria therapeutics that can simultaneously target multiple malaria protective host molecular pathways. Further experimental studies are needed to validate the findings in the current study.

Materials and Methods

Description of the study datasets

We accessed a previous severe malaria GWAS datasets (16) ($N = \sim 11,000$) of three African populations including Kenya, Gambia and Malawi from European Phenome Genome Archive (EGA) following the standard data access protocols outlined in (88, 89). Children with severe malaria cases were recruited on admission to hospital using definitions as per WHO guidelines for cerebral malaria (Blantyre coma score < 3 in children or Glasgow coma score < 11 in adults), severe malarial anaemia (haemoglobin < 5 g/100 mL or haematocrit $< 15\%$) and other malaria-related symptoms (90). Control samples were obtained from representative of the ethnic groups of the cases or in some study sites from the local population (89). The samples were genotyped on Illumina Omni 2.5M array and QC filtered as described in (14). In addition to the genotype dataset, we obtained a set of severe malaria susceptibility GWAS summary statistics ($N=17,000$) meta-analysed across eleven population in Africa, Oceania and Asia from (18). The dataset contained information on GWASs of individual study populations and their meta-analysis. We additionally accessed a merged data set from 1000 Genomes Project and African Genome Variation Project (AGVP)(91) of which 20 world-wide ethnic groups (**Supplementary Table 4**) have been grouped following ethno-linguistic information (92).

Functional mapping and annotations

We used the meta-analyzed malaria GWAS summary statistics ($N=17,000$ samples, 17million SNPs) across eleven populations(18) for functional mapping and annotations. We implemented FUMA(20), a pipeline that determines genomic risk loci and prioritize potential causal genes by incorporating information from multiple sources including GTEx (93), ENCODE (25), Roadmap Epigenomics Project (26) and chromatin interaction information (27). Briefly, a pre-calculated LD structure based on 1000 Genome version 3 of the African population was used to determine the risk loci and independent significant SNPs from the GWAS summary statistics data. Independent significant SNP was defined as a genome-wide significant SNP (P -value $< 5e-8$) within LD threshold of $r^2 > 0.6$. Other nominally significant SNPs ($p < 0.05$) that are in LD ($r^2 \leq 0.6$) with

independent significant SNPs were designated as candidate SNPs and included for annotation analysis. From identified independent significant SNPs, lead SNPs were selected at LD threshold of $r^2 > 0.1$. SNPs in close proximity (< 250 k) were considered as a single locus and thus, each genomic locus can contain multiple independent significant SNPs and lead SNPs.

We then implemented three gene mapping strategies including positional mapping, expression Quantitative Trait Locus (eQTL) mapping, and chromatin interaction implemented in FUMA(20). Positional mapping was performed by ANNOVAR tool (94) using Ensembl (build 85; <http://www.ensembl.org/>) dataset. A maximum distance of 10kb window size upstream and downstream was used to map SNPs to genes. SNPs filtering was carried out based on CADD score(95), RegulomeDB score (96) and 15-core chromatin state (97).

eQTL mapping was performed for genes within 1 Mb of the most significant variant using datasets that contain eQTL information related to severe malaria such as brain and blood. These include:PsychENCODE(98), GTExv8 (93), BRAINEAC (99), DICE (100), eQTLGen (101) (and Blood eQTLbrowser (102) and scRNA_eQTLs(103). Chromatin interaction mapping was performed using dataset including Hi-C data of 21 tissues/cell types obtained from GSE87112. (104), Hi-C loops from Giusti-Rodriguez et al. 2019 (105) which contains pre-processed enhancer-promoter and promoter-promoter interactions based on Hi-C data for adult and foetal human brain samples(105), Hi-C based data from PsychENCODE (98) which is composed of Enhancer-Promoter links based on Hi-C and Promoter anchored Hi-C loops and Enhancer-Promoter correlations from FANTOM5 (106). We restricted our analysis to tissues related to severe malaria pathogenesis. Genes prioritized by any of these three strategies were considered as potential causal genes. To gain insights in to the biological functions of prioritized genes, we performed gene enrichment analysis using an hypergeometric test in which gene-sets obtained from MsigDB (38) and WikiPathways (107) were used as background genes. We further tested differential gene expression values on 54 tissues obtained from the GTEx (93) as indicted FUMA (20).

Gene-based genome-wide association analysis

Considering the polygenic nature of severe malaria susceptibility trait (11), we applied Pascal (22), a gene-based GWAS analysis that can aggregate all SNPs within a gene and thus, capture modest effects. Briefly, Pascal sum of chi-squared statistics (SOCS) analysis was applied to nominally significant GWAS SNPs (p -value < 0.05) and to compute the corresponding gene scores (p -values) in 50kb upstream and downstream window. LD information for estimation of correlation structure was obtained from African dataset in 1000G phase 3 (108). We further categorized the prioritized genes into different functional groups using DAVID tools (39). Significant genes were subjected to differential gene expression analysis implemented in FUMA software using 54 tissues obtained from the GTEx (93). Pathway scores were computed by combining the scores of genes that belong to the same gene-set. The analysis doesn't require a gene score threshold and incorporates weakly associated genes for pathway enrichment. Gene-fusion parameter was set to 1Mb so that all pathway-member genes within 1Mb apart were fused together. Genes in HLA-region and those containing more than 3000 SNPs were removed as outlined in Pascal documentations (22).

Gene burden and rare-variants association tests

Given that the GWAS assumption is based on common variant common disease hypothesis, GWAS approach always miss potential association signal from rare variants. To examine the contribution of rare variants, we applied optimal unified sequence kernel association test (SKAT-O), which combines burden and variance-component analyses (37) to the GWAS dataset of Gambia, Kenya and Malawi populations. Briefly, we aligned the VCF files including Gambia (N=4920 samples, 1.6 million SNPs), Malawi (N=2560 samples, 1.6 million SNPs) and Kenya (N=3143 samples, 1.6 million SNPs) to GWAS dataset to 1000 Genome v-3 reference haplotypes using Genotype Harmonizer (109) and removed SNPs with position and strand mismatches and phased using SHAPEITv2 (110). We performed imputation using impute 2 (111) and obtained ~20 million from each population. After removal SNPs with low genotype rate and imputation accuracy, we retained ~15,000 SNPs in each population. We then applied SKAT-O test to the quality filtered data following the procedure outlined in SKAT package (37).

Network analysis

We performed a network analysis to investigate the functional interaction between malaria resistance candidate genes. Briefly, we obtained functional interaction network from of the identified candidate malaria resistance genes using Multiple Association Network Integration Algorithm (geneMANIA) tool (112). Using this information, we computed network parameters including degree, betweenness and closeness centrality metrics to evaluate the topology of nodes(genes) and edges(interactions) in the network using networkX (113) and R igraph packages (114).

Closeness measures the average distance from the node to all other nodes in the network, indicating which nodes represent a greater “risk” (maximally close with lowest sum of edge weights) for eliciting other nodes. Betweenness measures the number of times that a node lies on the shortest path between two other nodes, indicating which nodes serve as a “hub” between other nodes (114). The degree of a node is described as the number of direct connections it has with other nodes within the network. Low degree nodes usually connect to nodes within their local community whereas high degree nodes usually extend to the neighbouring community (114). Using the centrality scores, we quantified node centrality to identify hub genes by investigating the contribution of the edges and the weight of the edges towards node centrality. The hubs genes make strong contributions to the subnetwork and/or global network integrity. Connector hubs and provincial hubs refers to nodes that link other nodes across different communities and local communities, respectively.

Population genetic structure of malaria-specific resistance candidate genes

We performed the population genetic structure of the identified genes in malaria endemic populations (Kenya, Gambia and Malawi) and global populations of 20 ethnic groups obtained from African Genome Variation Project (AGVP) (91). We mapped a total of 14,106,476 SNPs to the identified candidate genes using dbSNP database. We merged the quality filtered GWAS datasets of the three malaria endemic populations using PLINK software (115) and retained dataset that contains the SNPs that are mapped to the identified genes (N= 10578 samples, 15,675 SNPs). We performed basic quality control, removed structural variants and ambiguous SNPs

using PLINK software (115) as described elsewhere (116). We retained AGVP dataset containing only the SNPs that are mapped to the identified candidate genes (N=4932 samples, 93,5549 SNPs). We then partitioned these datasets in to a total 23 different Ethnic groups (20 from AGVP and 3 from malaria endemic populations) based on population or country labels information. We clustered the merged malaria GWAS dataset and the AGVP dataset containing only SNPs residing in the identified genes in to sub-regions/populations using smartpca software (117). For each ethnic group, we computed various population genetics analysis including minor allele frequency (MAF) at various bins and gene-specific in SNPs MAF by aggregating MAF of all associated SNPs within gene from dbSNPs. We finally computed proportion of pathogenic SNPs within the candidate genes using ANNOVAR software (94) by computing the number of SNPs reported to be pathogenic over total numbers associated SNPs within gene from dbSNPs.

Acknowledgements

This work was supported through the DELTAS Africa Initiative [grant 107740/Z/15/Z]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107740/Z/15/Z] and the UK government. The authors would also like to thank the National Research Foundation of South Africa for funding (NRF) [grant # RA171111285157/119056] and The Centre for High-Performance Computing (CHPC, www.chpc.ac.za) for computing resources This study makes use of data generated by MalariaGEN. A full list of the investigators who contributed to the generation of the data is available from www.malariagen.net. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

Funding

DD, FA, PK, NK, HG and LG are funded by The Developing Excellence in Leadership and Genetics Training for Malaria Elimination in sub-Saharan Africa (DELGEME) program (grant #PD00217ML). EC is funded by NIH projects

Declaration of interest

The authors declare that they have no competing interests.

Authors contributions

DD designed, performed the data analysis and drafted the manuscript, FA, PK, NK, WT, HG and LG contributed in data analysis and revision of the manuscript, EC supervised the work

References

1. World Health Organization(WHO). World malaria report. 2019
2. World Health Organization(WHO) .World malaria report. 2018
3. Egan A, Keusch GT, Breman JG. The intolerable burden of malaria: a new look at the numbers.200. *Am J Trop Med Hyg*;64
4. World Health Organization(WHO). Severe Malaria. 2014;19: 7–131. doi:10.1111/tmi.12313
5. Dhiman S. Are malaria elimination efforts on right track ? An analysis of gains achieved and challenges ahead. *IDP*; 2019; 1–19.
6. Newby G, Bennett A, Larson E, Cotter C, Shretta R, Phillips AA, et al. The path to eradication : a progress report on the malaria-eliminating countries. *The Lancet*.2016; 1775–1784.
7. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. *Nature*. 2002;415: 673–679. doi:10.1038/415673a
8. Cowman AF, Crabb BS. Invasion of Red Blood Cells by Malaria Parasites. *Cell*.2006; 755–766. doi:10.1016/j.cell.2006.02.006
9. Kevin M, Dayo F, Catherine W, Isiah M, Maria W, Victoria M, Et al. Indicators of life-threatening malaria in african children *N Engl J Med*.1995;332: 1399–1404.
10. Damena D, Denis A, Golassa L, Chimusa ER. Genome-wide association studies of severe *P . falciparum* malaria susceptibility : progress , pitfalls and prospects. *BMC Med.Genom.*; 2019; 1–14.
11. Damena D, Chimusa ER. Genome-wide heritability analysis of severe malaria resistance reveals evidence of polygenic inheritance.*HMG*. 2020;29: 168–176. doi:10.1093/hmg/ddz258
12. Teo Y, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association

- analysis in Africa. *Nat Rev Genet.* 2013;11: 149–160. doi:10.1038/nrg2731.
13. Kwiatkowski DP. How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria. *Am J Hum Genet.* 2005;77: 171–192. doi:10.1086/432519
 14. Jallow M, Teo YY, Small KS, Rockett KA, Clark TG, Kivinen K, et al. Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat Genet.* 2010;41: 657–665. doi:10.1038/ng.388. Genome-wide
 15. Timmann C, Thye T, Vens M, Evans J, May J, Ehmen C, et al. Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature.* 2012;489. doi:10.1038/nature11334
 16. Band G, Rockett KA, Spencer CCA, Kwiatkowski DP, Si Le Q, Clarke GM, et al. A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature.* 2015;526: 253–257. doi:10.1038/nature15390
 17. Ravenhall M, Campino S, Sepu N, Nadjm B, Mtove G, Wangai H, et al. Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania. *PloS Genet.* 2018; 1–15.
 18. Gavin B, Quang S Le, Geraldine MC, Katja K, Christina H, Anna EJ, et al. Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania. *Nat Commun.* 2019;10: 5732. doi:doi.org/10.1038/s41467-019-13480-z |
 19. Chimusa ER, Mbiyavanga M, Mazandu GK, Mulder NJ. AncGWAS: A post genome-wide association study method for interaction, pathway and ancestry analysis in homogeneous and admixed populations. *Bioinformatics.* 2015;32: 549–556. doi:10.1093/bioinformatics/btv619
 20. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun;* 2019: 1–10. doi:10.1038/s41467-017-01261-5
 21. Leeuw CA De, Mooij JM, Heskes T, Posthuma D. MAGMA : Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol;* 2015; 1–19. doi:10.1371/journal.pcbi.1004219
 22. Lamparter D, Marbach D, Rueedi R, Kutalik Z. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PloS Comput Biol.* 2016; 1–20. doi:10.1371/journal.pcbi.1004714
 23. Yoon S, Nguyen HCT, Yoo YJ, Kim J, Baik B, Kim S, et al. Efficient pathway enrichment and network analysis of GWAS summary data using GSA-SNP2. *Nucliec acids Res.* 2018;46. doi:10.1093/nar/gky175
 24. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science.* 2015;348: 648–661.
 25. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489:57–74 .
 26. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;48:317–330
 27. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* 2016; 17, 2042–2059
 28. Dudbridge F. Polygenic Epidemiology. *Genet Epidemiol.* 2016;40: 268–272. doi:10.1002/gepi.21966
 29. Sun PC, Uppaluri R, Schmidt AP, Pashia ME, Quant EC, Sunwoo JB, et al. Transcript Map of the 8p23 Putative Tumor Suppressor. *Genomics;* 2001;75: 17–25. doi:10.1006/geno.2001.6587
 30. Lee AS, Rusch J, Lima AC, Usmani A, Huang N, Lepamets M, et al. Rare mutations in the complement regulatory gene CSMD1 are associated with male and female infertility. *Nat*

- Commun. 2019; 10:1–16. doi:10.1038/s41467-019-12522-w
31. Hamada N, Ito H, Nishijo T, Iwamoto I, Morishita R. Essential role of the nuclear isoform of RBFOX1 , a candidate gene for autism spectrum disorders , in the brain development. *Sci Rep*. 2016; 1–19. doi:10.1038/srep30805
 32. Fernandez T, Morgan T, Davis N, Klin A, Morris A, Farhi A, et al. Disruption of Contactin 4 (CNTN4) Results in Developmental Delay and Other Features of 3p Deletion Syndrome. *Am. J. Hum. Genet*. 2004;74: 1286–1293.
 33. Milet J, Boland A, Luisi P, Sabbagh A, Sadissou I, Sonon P, et al. First genome-wide association study of non-severe malaria in two birth cohorts in Benin. *Hum Genet*.2019; 138:1341-1357
 34. Mro'zek K, Tanner M, Heinonen K, Bloomfield C. Molecular Cytogenetic Characterization of the KG-1 and KG-1a Acute Myeloid Leukemia Cell Lines by Use of Spectral Karyotyping and Fluorescence In Situ Hybridization. *Gen, Chrom & Canc* 2003;38:252: 249–252. doi:10.1002/gcc.10274
 35. Aprelikova O, Pajusola K, Partanen J, Armstrong E, Alitalo R, Bailey SK, et al. FLT4) a Novel Class III Receptor Tyrosine Kinase in Chromosome. *Canc Res*.1992; 746–748.
 36. Yeo TW, Lampah DA, Gitawati R, Tjitra E, Kenangalem E, Piera K, et al. Angiopoietin-2 is associated with decreased endothelial nitric oxide and poor clinical outcome in severe falciparum malaria.PNAS. 2008;105: 17097–17102.
 37. Ionita-laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence Kernel Association Tests for the Combined Effect of Rare and Common Variants. *Am J Hum Genet*. The American Society of Human Genetics; 2013;92: 841–853. doi:10.1016/j.ajhg.2013.04.015
 38. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3 . 0.Bioinformatics. 2011;27: 1739–1740. doi:10.1093/bioinformatics/btr260
 39. Jiao X, Sherman BT, Huang DW, Stephens R, Baseler MW, Lane HC, et al. DAVID-WS : a stateful web service to facilitate gene / protein list analysis. *Bioinformatics*.2012;28: 1805–1806. doi:10.1093/bioinformatics/bts251
 40. Taylor SM, Cerami C, Fairhurst RM. Hemoglobinopathies: Slicing the Gordian Knot of Plasmodium falciparum Malaria Pathogenesis. *PLoS Pathog*. 2013;9. doi:10.1371/journal.ppat.1003327
 41. Hedrick P. Population genetics of malaria resistance in humans. *Heredity*. 2011;10716: 283–304. doi:10.1038/hdy.2011.16
 42. Wynne C, Lazzari E, Smith S, McCarthy EM, Ni' Gabhann J, et al. TRIM68 Negatively Regulates IFN- b Production by Degrading TRK Fused Gene , a Novel Driver of IFN- b Downstream of Anti-Viral Detection Systems. *PloS ONE* 2014;9. doi:10.1371/journal.pone.0101503
 43. Foss S, Bottermann M, Jonsson A, Sandlie I, James LC, Andersen JT. TRIM21 — From Intracellular Immunity to Therapy. *Frontiers Immun*.2019;10: 1–14. doi:10.3389/fimmu.2019.02049
 44. Moxon A.Christopher, Matthew PG, McGuinness D, Danny AM, Matthias M. New Insights into Malaria Pathogenesis.*Ann Rev*. 2020;15:313-343
 45. Milner DA. Malaria Pathogenesis. *Cold Spring Harb Perspect Med*. 2018; 1–11.
 46. Mbengue A, Yam XY, Braun-breton C. Human erythrocyte remodelling during Plasmodium falciparum malaria parasite growth and egress. *BJH*.2012; 171–179. doi:10.1111/j.1365-2141.2012.09044.x

47. Alonso S, Lo S, Izagirre N. Overdominance in the Human Genome and Olfactory Receptor Activity. *Mol. Biol. Evol.* 2008;25:997–1001. doi:10.1093/molbev/msn049
48. Zheng X, Long. ADAMTS13 and von Willebrand Factor in Thrombotic Thrombocytopenic Purpura. *Annu Rev Med.* 2015; 66:211–225. doi:10.1146/annurev-med-061813-013241.
49. Sporn LA, Marder VJ, Wagner DD, Stel H V, Mourik JA Van. Differing Polarity of the Constitutive and Regulated Secretory Pathways for von Willebrand Factor in Endothelial Cells. *J. Cell. Biol.* 1989;108: 1283–1289.
50. Adams Y, Kuhnrae P, Higgins MK, Ghumra A, Rowe JA. Rosetting Plasmodium falciparum-infected erythrocytes bind to human brain microvascular endothelial cells in vitro, demonstrating a dual adhesion phenotype mediated by distinct P. falciparum erythrocyte membrane protein 1 domains. *Infect Immun.* 2014;82: 949–959. doi:10.1128/IAI.01233-13
51. Hollestelle MJ, Mantey EA, Chakravorty SJ, Craig A, Akoto AO, Donnell JO, et al. von Willebrand factor propeptide in malaria : evidence of acute endothelial cell activation. *BJH.* 2006; 562–569. doi:10.1111/j.1365-2141.2006.06067.x
52. Kraisin S, Naka I, Patarapotikul J, Nantakomol D, Nuchnoi P. Association of ADAMTS13 polymorphism with cerebral malaria. *Mal.J.* 2011; 1–8.
53. Ahsan M, Ek WE, Rask-andersen M, Karlsson T, Lind-thomsen A, Enroth S, et al. The relative contribution of DNA methylation and genetic variants on protein biomarkers for human diseases. *PloS Genet.* 2017; 13:1–24.
54. Sabater-Lleal M, Jennifer EH, Paul S de V, Jonathan M. Genome-Wide Association Transethnic Meta-Analyses Identifies Novel Associations Regulating Coagulation Factor VIII and von Willebrand Factor Plasma Levels. *Circulations.* 2019; 139:620–635. doi:10.1161/CIRCULATIONAHA.118.034532
55. Chen K, Ma H, Li L, Zang R, Wang C, Song F, et al. Corrigendum : Genome-wide association study identifies new susceptibility loci for epithelial ovarian cancer in Han Chinese women. *Nat Commun.* 2014;5682: 6828. doi:10.1038/ncomms6828
56. Tiffert T, Lew VL, Ginsburg H, Krugliak M, Croisille L, Mohandas N. The hydration state of human red blood cells and their susceptibility to invasion by Plasmodium falciparum. *Blood.* 2005;105: 4853–4860. doi:10.1182/blood-2004-12-4948
57. Astle WJ, Elding H, Jiang T, Ouwehand WH, Butterworth AS, Soranzo N, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease Resource. *Cell.* 2017; 1415–1429. doi:10.1016/j.cell.2016.10.042
58. Hellberg Å, Poole J, Olsson ML. Molecular Basis of the Globoside-deficient P k Blood Group. *J. Biol. Chem.* 2002;277: 29455–29459. doi:10.1074/jbc.M203047200
59. Kevin EB, Jonathan RH, Giorgio G, Stancie MA, Elton DL, Peggy M, et al. Resistance to Parvovirus B19 infection due to Lack of virus receptor (Erythrocyte Antigen). *N Engl J Med.* 1994;330: 1192–6.
60. Lund B, Lindberg FP, Baga M, Normark S. Globoside-Specific Adhesins of Uropathogenic Escherichia coli Are Encoded by Similar Trans-Complementable Gene Clusters. *J. Bacteriol.* 1985;162: 1293–1301.
61. Rowe JA, Claessens A, Corrigan RA, Arman M. Adhesion of Plasmodium falciparum-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. *Expert Rev Mol Med.* 2009;11: e16. doi:10.1017/S1462399409001082
62. Kristiansen M, Graversen JH, Jacobsen C, Hoffman È, Law SKA, Sonne O, et al. Identification of the haemoglobin scavenger receptor. *Nature.* 2001;409: 1999–2002.
63. Francis SE, Sullivan DJ, Goldberg DE. hemoglobin metabolism in the malaria parasite

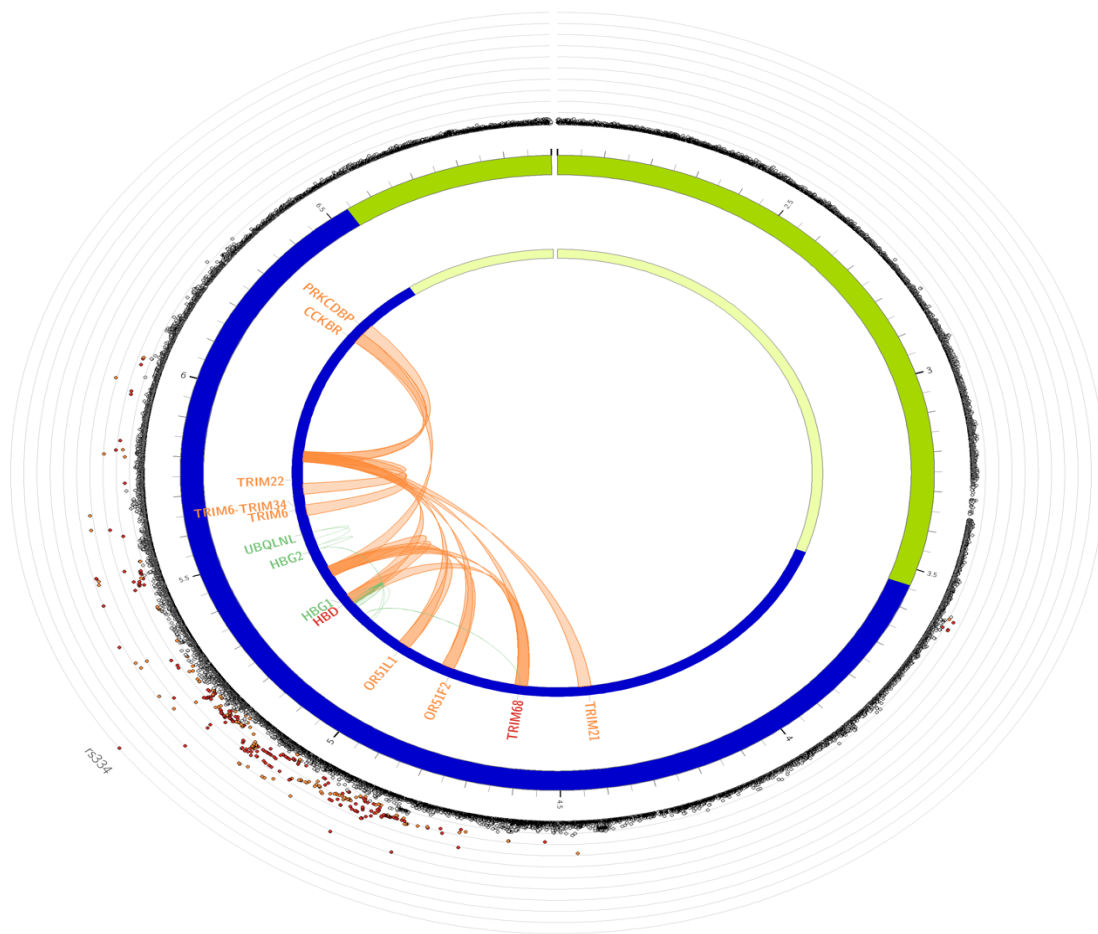
- Plasmodium falciparum. *Annu. Rev. Microbiol.* 1997; 97–123.
64. Jonscher E, Flemming S, Schmitt M, Spielmann T, Jonscher E, Flemming S, et al. PfVPS45 Is Required for Host Cell Cytosol Uptake by Malaria Blood Stage Parasites. *Cell Host Microb.* 2019; 166–173. doi:10.1016/j.chom.2018.11.010
 65. Francischetti IVOMB, Seydel KB, Monteiro RQ. Blood Coagulation , Inflammation , and Malaria. *Microcirculation.*2008; 81–107. doi:10.1080/10739680701451516
 66. Sullivan JMO, Preston RJS, Regan NO, Donnell JSO. Emerging roles for hemostatic dysfunction in malaria pathogenesis. *Blood.* 2016;127: 2281–2289. doi:10.1182/blood-2015-11-636464.
 67. Hunt NH, Ball HJ, Hansen AM, Khaw LT, Guo J, Bakmiwewa S, et al. Cerebral malaria : gamma-interferon redux. *Frontiers Cell Infec.Microbiol.* 2014;4: 1–12. doi:10.3389/fcimb.2014.00113
 68. Tosevski V, Forman R, Muller W, Couper N. crossm Gamma Interferon Mediates Experimental Cerebral Malaria by Signaling within Both the Hematopoietic and Nonhematopoietic Compartments. *Infect Immun.* 2017;85: 1–13.
 69. Ghartey-kwansah G, Yin Q, Li Z, Gumpfer K, Sun Y, Yang R, et al. Calcium-dependent Protein Kinases in Malaria Parasite Development and Infection. *Cell Transplant.*2020; 1–12. doi:10.1177/0963689719884888
 70. Sicard A, Semblat J, Doerig C, Hamelin R, Moniatte M, Dorin-semblat D, et al. Activation of a PAK-MEK signalling pathway in malaria parasite-infected erythrocytes. *Cell Microbiol.*2011;13: 836–845. doi:10.1111/j.1462-5822.2011.01582.x
 71. Brochet M, Billker O. Calcium signalling in malaria parasites. *Cell.Microbiol.*2016; 12. doi:10.1111/mmi.13324
 72. Millholland MG, Mishra S, Dupont CD, Love MS, Patel B, Shilling D, et al. Article A Host GPCR Signaling Network Required for the Cytolysis of Infected Cells Facilitates Release of Apicomplexan Parasites. *Cell Host Microb.*2013; 13:15–28.
 73. Pillai AD, Addo R, Sharma P, Nguitragool W, Srinivasan P, Desai SA. Malaria parasites tolerate a broad range of ionic environments and do not require host cation remodelling. *Mol Microbiol.*2013;88: 20–34. doi:10.1111/mmi.12159
 74. Desai SA. Why do malaria parasites increase host erythrocyte permeability ? *Cell.* 2014;30: 151–159.
 75. Kiaran K. Ion Regulation in the Malaria Parasite. *Biochem. J.* 2014;457, 1–18
 76. Kirk K, Lehane AM. Membrane transport in the malaria parasite and its host erythrocyte. *Biochem. J.* 2014;18: 1–18. doi:10.1042/BJ20131007
 77. Nguitragool W, Bokhari AAB, Pillai AD, Rayavara K, Sharma P, Turpin B, et al. Malaria Parasite clag3 Genes Determine Channel-Mediated Nutrient Uptake by Infected Red Blood Cells. *Cell.*2011;145: 665–677. doi:10.1016/j.cell.2011.05.002
 78. Eklund EH, Akabas MH, Fidock DA. Previews Taking Charge : Feeding Malaria via Anion Channels. *Cell.* 2011;145: 645–647. doi:10.1016/j.cell.2011.05.012
 79. Bouyer G, Egée S, Thomas SL. Three types of spontaneously active anionic channels in malaria-infected. *BCMD.*2006;36: 248–254. doi:10.1016/j.bcmd.2006.01.005
 80. Bouyer G, Cueff A, Gallagher PG, Thomas SLY. Erythrocyte peripheral type benzodiazepine receptor / voltage-dependent anion channels are upregulated by Plasmodium falciparum. *Blood.*2011;118: 2305–2312. doi:10.1182/blood-2011-01-329300.
 81. Heatwoie VM, Wertheimer SP, Peterson DS, Ravetch JA, Weilems TE. The Large Diverse Gene Family w Encodes Proteins Involved in Cytoadherence and Antigenic Variation of

- Plasmodium falciparum-Infected Erythrocytes. *Cell*. 1995;82: 89–100.
82. Smith JD, Craig AG, Roberts DJ, Hudson-taylor DE, Peterson DS, Pinches R, et al. Switches in Expression of Plasmodium falciparum VW Genes Correlate with Changes in Antigenic and Cytoadherent Phenotypes of Infected Erythrocytes. *Cell*. 1995;82: 101–110.
 83. Raventos-suarez C, Kaul DK, Macaluso F, Nagel RL. Membrane knobs are required for the microcirculatory obstruction induced by Plasmodium falciparum-infected erythrocytes. *Proc. Natl. Acad. Sci.* 1985;82: 3829–3833.
 84. Rowe JA, Moulds JM, Newbold CI, Miller LH. mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Proc. Natl. Acad. Sci.* 1997;388: 292–295.
 85. Fried M, Duffy PE. Adherence of Plasmodium falciparum to Chondroitin sulfate A in the Human Placenta. *Science*. 1996;272: 4–6.
 86. Taoufiq Z, Gay F, Balvanyos J, Ciceron L, Tefit M, Lechat P, et al. Rho Kinase Inhibition in Severe Malaria : Thwarting Parasite-Induced Collateral Damage to Endothelia. *JID*. 2008;197. doi:10.1086/528988
 87. Nishanth G, Schlüter D. Blood – Brain Barrier in Cerebral Malaria : Pathogenesis and Therapeutic Intervention. *Trends Parasitol.* 2019;35: 516–528. doi:10.1016/j.pt.2019.04.010
 88. Parker M, Bull SJ, Vries J De, Agbenyega T, Doumbo OK, Dominic P. Ethical Data Release in Genome-Wide Association Studies in Developing Countries. *PloS Med*. 2009;6: 4–7. doi:10.1371/journal.pmed.1000143
 89. Achidi EA, Agbenyega T, Allen S, Amodu O, Bojang K, Conway D, et al. A global network for investigating the genomic epidemiology of malaria. *Nature*. 2008;456: 732–737. doi:10.1038/nature07632
 90. Trampuz A, Jereb M, Muzlovic I, Prabhu RM. Clinical review: Severe malaria. *Crit Care*. 2003;7: 315–323. doi:10.1186/cc2183
 91. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature*. 2015;517: 327–332. doi:10.1038/nature139942
 92. Michalopoulos S. The Origins of Ethnolinguistic Diversity *Am. Econom Rev.* 2012,. 2012;102: 1508–1539.
 93. Guerini D, Pan B, Carafoli E. Expression, purification, and characterization of isoform 1 of the plasma membrane Ca²⁺ pump. Focus on calpain sensitivity. *J Biol Chem*. 2003;278: 38141–38148. doi:10.1074/jbc.M302400200
 94. Wang K, Li M, Hakonarson H. ANNOVAR : Functional annotation of genetic variants from high-throughput sequencing data. 2010;38: 1–7. doi:10.1093/nar/gkq603
 95. Kircher M, Witten DM, Jain P, Roak BJO, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nucl Ac Res*.. 2014; doi:10.1038/ng.2892
 96. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genom Res*. 2012; 1790–1797. doi:10.1101/gr.137323.112.
 97. Ernst J, Kellis M. ChromHMM : automating chromatin- state discovery and characterization. *Nat Method*. 2012;10:1-2
 98. Li M, Santpere G, Kawasawa YI, Evgrafov OV, Gulden FO, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science*. 2018; doi:10.1126/science.aat7615

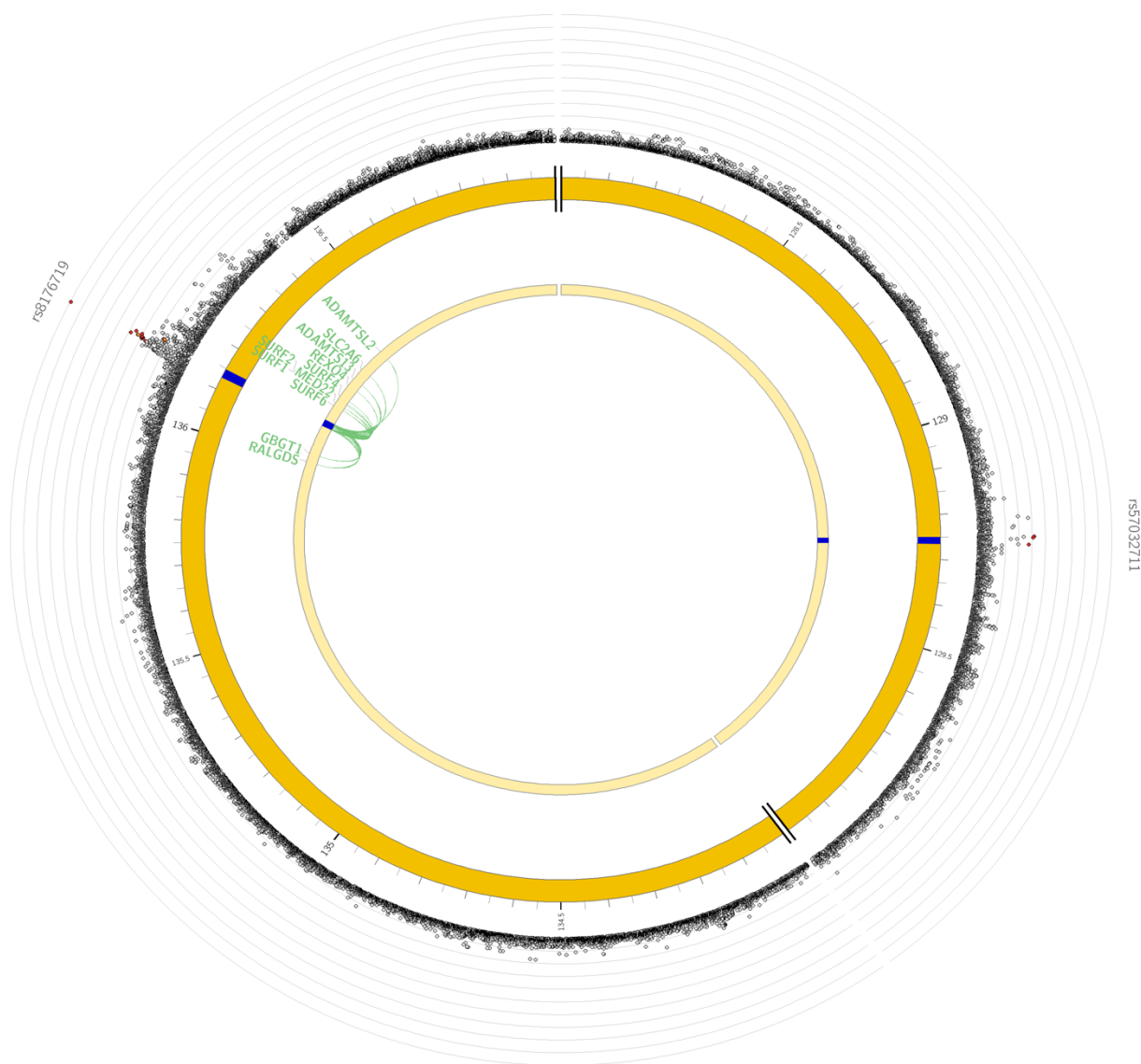
99. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat.Neuroscience*. 2014; doi:10.1038/nn.3801
100. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AL, White BM, Zapardiel-Gonzalo J, et al. Impact of Genetic Polymorphisms on Human Immune Resource Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*.2018; 1701–1715.
101. Vösa U, Claringbould A, Westra H, Bonder MJ, Zeng B, Kirsten H, et al. Unraveling the polygenic architecture of complex traits using blood eQTL meta- analysis. *bioRxiv* 2018; 1–57. doi: <https://doi.org/10.1101/447367> . this
102. Zhernakova D V, Deelen P, Vermaat M, Itersson M Van, Galen M Van, Arindrarto W, et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Blood*.2016; doi:10.1038/ng.3737
103. Wijst MGP Van Der, Brugge H, Vries DH De, Deelen P, Swertz MA. Single-cell RNA sequencing identifies celltype- specific cis-eQTLs and co-expression QTLs.*Nat.Genet*. 2018. 50:493–497; <https://doi.org/10.1038/s41588-018-0089-9>
104. Schmitt AD, Hu M, Jung I, Lin Y, Barr CL. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell*.2016; 2042–2059.
105. Huckins LM, Dobbyn A, Ruderfer DM, Hoffman G, Wang W, Pardiñas AF, et al. Gene expression imputation across multiple brain Regions provides insights into schizophrenia risk. *Nat.Genet*.2019; 51: 659–674 <https://doi.org/10.1038/s41588-019-0364-4>
106. Andersson R, Gebhard C, Miguel-escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014.504:455–461 doi:10.1038/nature12787
107. Kutmon M, Riutta A, Nunes N, Hanspers K, Willighagen EL, Bohler A, et al. WikiPathways : capturing the full diversity of pathway knowledge. 2016; *Nucl.Ac.Res*. 44: 488–494. doi:10.1093/nar/gkv1024
108. The 1000 Genomes Project Consortium. Europe A map of human genome variation from population scale sequencing. *Nature*. 2011;467: 1061–1073. doi:10.1038/nature09534.A
109. Mostafavi S, Ray D, Warde-farley D, Grouios C, Morris Q. GeneMANIA : a real-time multiple association network integration algorithm for predicting gene function. *Nucl.Ac.Res*.2008;9: 1–15.
110. Hagberg AA, National LA, Alamos L. Exploring Network Structure , Dynamics , and Function using NetworkX. *Proc. SciPy*.2008; 11–15.
111. Csardi G, Tam’as N. The igraph software package for complex network research. *IJCS*
112. Deelen P, Bonder MJ, Velde KJ Van Der, Westra H, Winder E, Hendriksen D, et al. Genotype harmonizer : automatic strand alignment and format conversion for genotype data integration.*BMC.Res.Not*. 2014; 1–4.
113. Delaneau O, Marchini J, Zagury J. High-resolution whole-genome haplotyping using limited seed data.*Nat.Method*. 2013;9.
114. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5. doi:10.1371/journal.pgen.1000529
115. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 2007;81: 559–575. doi:10.1086/519795
116. Marees AT, Stringer S, Claire M, Derks EM. A tutorial on conducting genome - wide

- association studies : Quality control and statistical analysis. *Int J Methods Psychiatr Res.*2018; 1–10. doi:10.1002/mpr.1608
117. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. *PloS Genet.*2006.2. doi:10.1371/journal.pgen.0020190

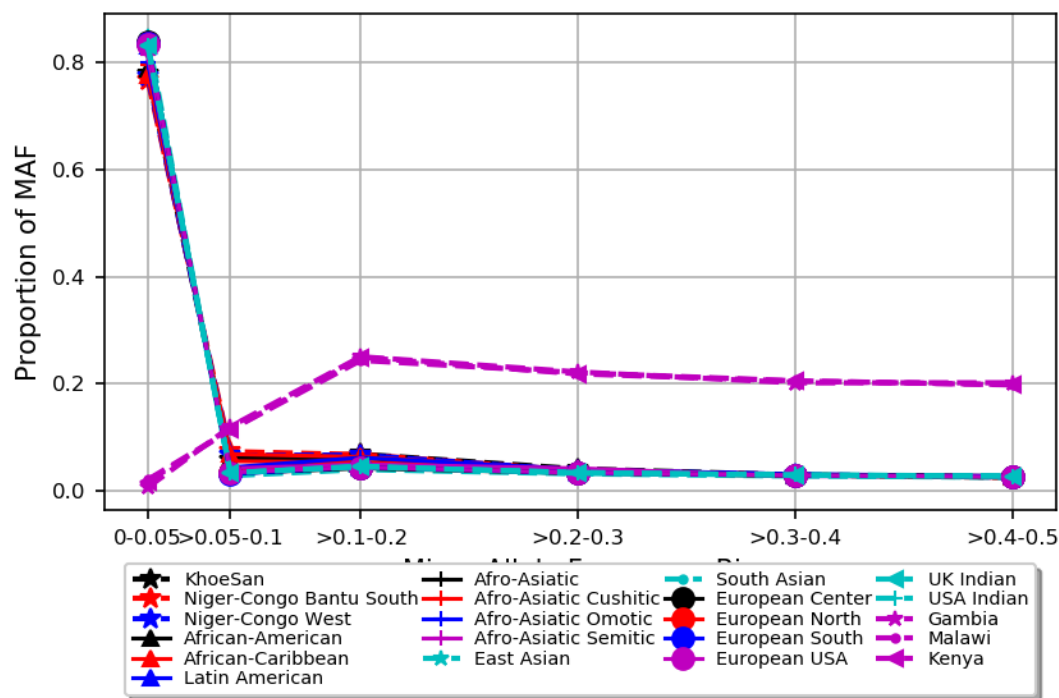
Supplementary Figures and Tables



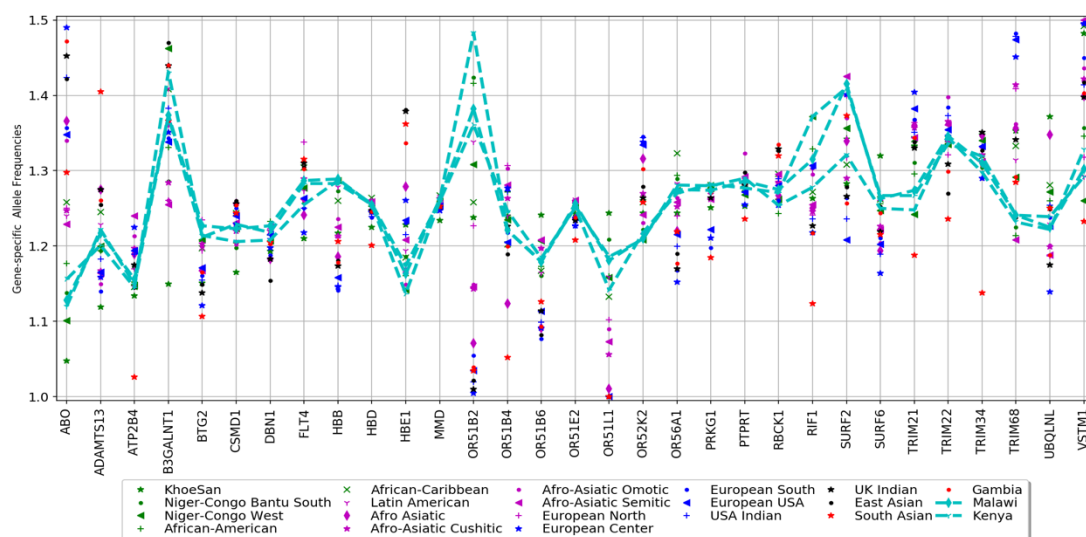
Supplementary Fig 1 . Chromatin interactions and eQTLs of severe malaria resistance candidate genes on chr 11 risk locus. The most outer layer is the Manhattan plot displaying SNPs with P-value < 0.05. Candidate SNPs are coloured based on the highest r^2 to one of the independent significant loci (red: $r^2 > 0.8$, orange: $r^2 > 0.6$). Other SNPs are coloured in grey. The outer circle is the chromosome coordinate and genomic risk loci are highlighted in blue. Genes mapped by either Hi-C or eQTLs are shown on the inner circle. Genes identified by chromatin interaction and eQTLs are coloured orange and green respectively while genes mapped by both are coloured red.



Supplementary Fig2. Chromatin interactions and eQTLs of severe malaria resistance candidate genes on chr 9 risk locus. The most outer layer is the Manhattan plot displaying SNPs with P-value < 0.05. Candidate SNPs are coloured based on the highest r^2 to one of the independent significant loci (red: $r^2 > 0.8$, orange: $r^2 > 0.6$). Other SNPs are coloured in grey. The outer circle is the chromosome coordinate and genomic risk loci are highlighted in blue. Genes mapped by either Hi-C or eQTLs are shown on the inner circle. Genes identified by chromatin interaction and eQTLs are coloured orange and green respectively while genes mapped by both are coloured red



Supplementary Fig 3. Minor allele frequency bins of SNPs mapped to severe malaria resistance candidate genes in three malaria endemic populations (Gambia, Malawi, Kenya) and global populations of 20 ethnic groups. Y-axis represent allele frequency, X-axis represent different MAF bins. Populations were represented by different colors and symbols



Supplementary Fig 4. Gene specific MAF of SNPs mapped to severe malaria resistance candidate genes in three malaria endemic populations (Gambia, Malawi, Kenya) and global populations of 20 ethnic groups. Y-axis represent gene specific allele frequency, X-axis represent genes. Populations were represented by different colors and symbols

Supplementary Table 1. Positional enrichment of genes identified by FUMA method using MsigDB genes as background

Ensg	entrezID	HUGO	symbol	chromosome	Cytoband	Start	end	Biotype
ENSG00000058668	493	ATP2B4	ATP2B4	1	q32.1	203595689	203713209	Protein coding
ENSG00000148296	6838	SURF6	SURF6	9	q34.2	136197552	136203235	Protein coding
ENSG00000148297	6837	MED22	MED22	9	q34.2	136205160	136214986	Protein coding
ENSG00000148290	6834	SURF1	SURF1	9	q34.2	136218610	136223552	Protein coding
ENSG00000148291	6835	SURF2	SURF2	9	q34.2	136223428	136228045	Protein coding
ENSG00000148248	6836	SURF4	SURF4	9	q34.2	136228325	136242970	Protein coding
ENSG00000132109	6737	TRIM21	TRIM21	11	p15.4	4406127	4414926	Protein coding
ENSG00000176925	119694	OR51F2	OR51F2	11	p15.4	4842551	4843686	Protein coding
ENSG00000244734	3043	HBB	HBB	11	p15.4	5246694	5250625	Protein coding
ENSG00000223609	3045	HBD	HBD	11	p15.4	5253908	5256600	Protein coding
ENSG00000196565	3048	HBG2	HBG2	11	p15.4	5274420	5667019	Protein coding
ENSG00000213931	3046	HBE1	HBE1	11	p15.4	5289582	5526847	Protein coding
ENSG00000184881	79345	OR51B2	OR51B2	11	p15.4	5344541	5345582	Protein coding

Supplementary Table 2. Candidate malaria resistance genes identified by both functional mapping and gene-based GWAS analysis

Position	N	N	P-value	adjusted P	Genes
chr11p15	297	19	8.86e-21	2.65e-18	<i>TRIM21, TRIM68, MMP26, OR51F2, OR51L1, OR52A1, HBB, HBD, HBG1, HBG2, HBE1, OR51B2, UBQLNL, TRIM6, TRIM6-TRIM34, TRIM34, TRIM22, CCKBR, PRKCDBP</i>
chr9q34	214	11	3.10e-11	4.63e-9	<i>RALGDS, GBGT1, SURF6, MED22, SURF1, SURF2, SURF4, REXO4, ADAMTS13, SLC2A6, ADAMTSL2</i>
chr4q31	67	6	4.61e-8	4.60e-6	<i>RNF150, ZNF330, GAB1, FREM3, GYPE, GYPB</i>
chr1q32	147	7	2.62e-7	1.96e-5	<i>CHIT1, BTG2, FMOD, PRELP, ATP2B4, LAX1, PLEKHA6</i>
chr3q26	61	3	7.97e-4	4.77e-2	<i>B3GALNT1, NMD3, SPTSSB</i>

Supplementary Table 3. Nominally significant genes obtained from burden and rare variant analysis using SKAT package

Population	Gene	chromosome	Cytoband	start	end	Biotype	P.value	N.Marker.Rare
Gambia	<i>LOC105377246</i>	4	q13.1	58,976,071	58,984,194	ncRNA	1.72e-4	29
	<i>LOC100507464</i>	8	q11.21	49,496,763	49,512,234	ncRNA	9.68e-05	36
	<i>MIR4282</i>	6	q13	72967687	72967753	lncRNA	4.83e-4	7
	<i>GLYR1</i>	16	P13.3	4803203	4847288	Protein coding	3.72e-4	139
	<i>LINC00520</i>	14	q22.3	55781132	55796731	lncRNA	9.70e-05	19
	<i>LOC105371381</i>	-	-	-	-	-	1.84e-4	15
Malawi	<i>ATP8A1</i>	4	P13	42408373	42657105	Protein coding	4.91e-4	244
	<i>EPB41L2</i>	6	q23.2	130839347	131063322	Protein coding	3.91e-4	207
	<i>LOC105376161</i>	9	q22.32	95,764,922	95,776,282	ncRNA	1.51e-4	9
	<i>LOC105377731</i>	5	q35.1	173,269,431	173,280,929	ncRNA	3.13e-4	29
	<i>NDNF</i>	4	q27	121035613	121073021	Protein coding	9.53e-05	28
	<i>LINC00676</i>	13	q13.34	109,728,274	109,730,034	lncRNA	3.40e-4	10
	<i>LOC105370198</i>	13	q14.2	48,108,123	48,166,377	ncRNA	4.87e-4	122
	<i>LOC653786</i>	-	-	-	-		4.53e-4	29
	<i>WASF3</i>	13	q12.13	26557683	26688948	Protein coding	2.54e-4	193

