

**Genetic fine mapping and functional analysis
of a chromosomal region within *ATP2B4*
associated with resistance against severe malaria in humans**

Dissertation

With the aim of achieving a doctoral degree
at the Faculty of Mathematics, Informatics and Natural Sciences
Department of Biology

submitted by
Christina Strauß

2017 in Hamburg

Evaluators of the Dissertation

1. Prof. Dr. Julia Kehr
2. Dr. Christian Timmann

Date of the Disputation

21.07.2017

The laboratory work for this study was conducted
in the Department of Molecular Medicine
of the Bernhard Nocht Institute for Tropical Medicine, Hamburg
from Mai 2014 until April 2017.

Table of content

I	ABBREVIATIONS	IV
II	ABSTRACT	V
III	ZUSAMMENFASSUNG	VI
IV	LIST OF FIGURES	VII
V	LIST OF TABLES	VIII
1	INTRODUCTION	1
1.1	Severe malaria	1
1.1.1	The parasite life cycle	2
1.1.2	Pathology, signs and symptoms	3
1.1.3	Genetic resistance factors	4
1.2	A recent GWAS on severe malaria (Timmann et al. 2012)	5
1.3	PMCA4	7
1.4	The calcium balance in <i>Plasmodium</i>-infected erythrocytes	8
1.5	The challenge of interpreting GWAS results	9
1.5.1	Circular chromosome conformation capture	11
2	AIM AND OBJECTIVES	13
3	MATERIAL	14
3.1	Supplies for cell culture	14
3.2	Enzymes and enzyme inhibitors	14
3.3	General chemicals, water and lab consumables	14
3.4	PCR primer and fluorescence-labeled probes	14
3.5	Lab equipment	15
3.6	Buffers	15
3.7	Kits and ready-to-use experimental systems	16

3.8	Cell lines	16
3.9	Samples for functional experiments	16
3.10	Software and online tools	17
4	METHODS	18
4.1	Imputation	18
4.1.1	Association testing	20
4.1.2	Visualization of association data	20
4.1.3	LD analysis	20
4.2	Quantification of nucleic acid	21
4.3	4C-Seq	21
4.3.1	Cell culture	21
4.3.2	Fixation of nuclei	22
4.3.3	Digestion of nuclei	23
4.3.4	Preparation of sequencing library	24
4.3.5	Data analysis	26
4.4	Genotyping of donor samples for functional experiments	27
4.4.1	DNA extraction from dried blood spots	27
4.4.2	FRET-MCA assay	28
4.5	Gene-expression assays	29
4.5.1	Samples	29
4.5.2	RNA isolation and cDNA synthesis	29
4.5.3	Quantitative PCR	30
4.5.4	Statistics	31
4.6	Parasite-proliferation assay	32
4.6.1	Samples	32
4.6.2	Parasite culture	33
4.6.3	Determination of Parasitemia	34
4.6.4	Statistics	34
5	RESULTS	36
5.1	Computational fine mapping	36
5.1.1	Quality control of basic data	36
5.1.2	Imputation and QC of imputed data	36
5.1.3	Association testing	37
5.1.4	LD analysis	38
5.2	4C-Seq	39
5.2.1	Chromatin-chromatin interactions of the resistance locus	39
5.2.2	Reciprocal experiments	41

5.3	Genotyping of study group for functional experiments	43
5.4	Expression of <i>ATP2B4</i> in Ghanaian whole blood	44
5.5	Parasite proliferation referring to the rs10900585 genotype	46
6	DISCUSSION	49
6.1	Computational fine mapping	49
6.2	Chromatin-chromatin interactions of the resistance locus	53
6.2.1	Methods of 4C-Seq analysis	58
6.3	Expression of <i>ATP2B4</i>	60
6.4	Parasite-proliferation assay	61
7	CONCLUSIONS AND PROSPECTS	63
8	LIST OF LITERATURE	64
9	APPENDIX	75
9.1	Primers	75
9.1.1	Genotyping	75
9.1.2	4C-Seq	75
9.2	Tables	77
9.2.1	Computational fine mapping	77
9.2.2	Expression of <i>ATP2B4</i> in Ghanaian whole blood	79
9.2.3	Parasite proliferation referring to the rs10900585 genotype	80
9.3	Figures	81
9.3.1	Computational fine mapping	81
9.3.2	4C-Seq	84
9.3.3	Expression of <i>ATP2B4</i> in Ghanaian whole blood	86
	Acknowledgement	88
	Declaration on oath	89
	Language Certificate	90

I Abbreviations

CaM-BD	calmodulin-binding domain
DHE	dihydroethidium
DNA	desoxyribonucleic acid
FDR	false discovery rate
FRET	fluorescence resonance energy transfer
FSC	forward scatter
HRM	high resolution melting
HWE	Hardy-Weinberg equilibrium
IE	infected erythrocyte
IQR	interquartile range
MAF	minor allele frequency
MCA	melting curve analysis
NGS	next-generation sequencing
OR	odds ratio
PV	parasitophorous vacuole
PVM	parasitophorous vacuole membrane
QC	quality control
RBC	red blood cell (erythrocyte)
RE	restriction enzyme
RNA	ribonucleic acid
RPM	reads per million (normalized fragment read count)
RT	room temperature
SLRP	small leucine-rich repeat proteoglycans
SNP	single nucleotide polymorphism
SSC	side scatter
TAD	topologically associating domain
UTR	untranslated region
WHO	World Health Organization

II Abstract

In a recent genome-wide association study (GWAS) including more than 2.500 affected children in Ghana, West Africa, two new genetic resistance loci for severe malaria were identified (Timmann et al. 2012). One of them was indicated by the single-nucleotide polymorphism (SNP) rs10900585, which is located in the second intron of the gene *ATP2B4* encoding PMCA4, the major calcium pump of erythrocytes. In the current study, the mechanism of malaria resistance associated with the protective rs10900585 allele was examined.

Fine-mapping studies yielded no additional variants that were stronger associated than the index SNP rs10900585, identified in the GWAS. Applying circular chromosome conformation capture (4C-Seq) in a number of cell lines related to malaria pathology, we obtained evidence for a physical interaction between the rs10900585 region in the second intron of *ATP2B4* and the promoter region of the same gene in various cell lines. This suggested an influence of the rs10900585 region on *ATP2B4* transcriptional activity. Despite that, *ATP2B4* expression studies in whole blood samples of 200 Ghanaian donors did not show any differences corresponding to the rs10900585 genotype in the expression levels of the gene or its known splice variants. Studying the growth of *Plasmodium falciparum*, 3D7 strain, in erythrocytes of Ghanaian individuals with the three rs10900585 genotypes, however, showed significantly decreased proliferation rates in erythrocytes from donors carrying the resistance-associated genotype.

Including a recent finding that the minor allele of a SNP in high linkage disequilibrium with rs10900585 was associated with a decreased PMCA4 expression in erythrocyte membranes of Hungarian donors (Zambo et al. 2017), the data generated for this study suggests a genetic mechanism that has an impact on a downstream step of the translation of the *ATP2B4* mRNA. If a decreased translation can be verified in respective African samples, it could explain the impaired parasite proliferation, as the parasite employs the host's PMCA4 to generate the Ca²⁺-concentration required for its development inside the erythrocyte.

III Zusammenfassung

In einer kürzlich veröffentlichten genomweiten Assoziationsstudie (GWAS) die mehr als 2,500 Kinder aus Ghana, West Afrika, einschloss, konnten zwei neue Resistenz-Loci für schwere Malaria identifiziert werden (Timmann et al. 2012). Einer dieser beiden Loci wird durch den Einzelnukleotid-Polymorphismus (SNP) rs10900585 gekennzeichnet, der im zweiten Intron des Gens *ATP2B4* lokalisiert ist. Dieses codiert das Protein PMCA4, welches für den maßgeblichen Teil des aktiven Transports von Calcium-Ionen durch die erythrozytäre Membran verantwortlich ist. In der vorliegenden Studie wurden die Mechanismen untersucht, welche die Malaria-Resistenz vermitteln, die mit dem seltenen Allel von rs10900585 assoziiert ist.

Eine computergestützte Feinkartierung ergab keine weiteren Varianten, die eine deutlich stärkere Assoziation zeigen als der zuvor identifizierte Index SNP rs10900585. Unter Anwendung der Methode „circular chromosome conformation capture (4C-Seq)“ konnte in mehreren Zelllinien eine Interaktion zwischen der rs10900585-Region und der Promotor-Region von *ATP2B4* auf Chromatin-Ebene gezeigt werden. Dies deutet einen möglichen Einfluss des Resistenz-Locus auf die Transkription von *ATP2B4* an. Expressions-Analysen in Vollblut von 200 ghanaischen Individuen zeigten aber keinen Unterschied der mRNA-Spiegel des Gens und der bekannten Spleißvarianten in Bezug auf den rs10900585-Genotyp der Spender. Bei der Untersuchung des Wachstums des *Plasmodium falciparum*-Laborstamms 3D7 in Erythrozyten ghanaischer Spender, konnte jedoch eine signifikant verringerte Proliferation in roten Blutzellen von Spendern mit dem resistenz-assoziierten rs10900585-Genotyp beobachtet werden.

Eine neue Studie zeigt eine Assoziation von verringerter PMCA4-Expression in der Erythrozyten-Membran ungarischer Spender mit der seltenen Variante eines SNPs (Zambo et al. 2017), der in hohem Kopplungsungleichgewicht (LD) mit dem hier beschriebenen SNP rs10900585 liegt. Schließt man diese Entdeckung mit ein, so deuten die Daten der vorliegenden Studie auf Mechanismen hin, die sich auf Vorgänge auswirken, die auf die Translation der *ATP2B4* mRNA folgen. Sollte eine verringerte Translation in entsprechenden afrikanischen Proben verifiziert werden, so könnte dies die verringerte Proliferation der Parasiten erklären, da diese die PMCA4-Moleküle aus der erythrozytären Membran des Wirtes nutzen, um die für ihre Entwicklung im Erythrozyten essentielle Ca^{2+} -Konzentration stabil zu halten.

IV List of figures

Figure 1: Global Distribution of Malaria in 2014 (www.rollbackmalaria.org).....	1
Figure 2: Life cycle of <i>P. falciparum</i> (Cowman et al. 2012).....	2
Figure 3: Frequencies of the two alleles of the HbS SNP in sub populations from Phase 3 of the 1000 Genomes Project.....	4
Figure 4: Schematic illustration of the resistance locus.....	6
Figure 5: Schematic image of the secondary structure of PMCAs.....	7
Figure 6: Schematic illustration of different compartments of an erythrocyte, infected with plasmodia.....	9
Figure 7: Workflow for functional analyses of non-coding GWAS loci (Edwards et al. 2013).....	10
Figure 8: Schematic image of the experimental steps of 4C-Seq.....	12
Figure 9: Flow chart of the steps of <i>in silico</i> fine mapping from raw data to statistical analysis.....	18
Figure 10: Schematic figure of the viewpoint region (VP_Locus) and restriction sites.....	23
Figure 11: Regional association plot of the resistance locus generated from the dataset keepFRET-MCA_AFR.....	37
Figure 12: FourCSeq analysis of sequenced chromosomal fragments interacting with VP_Locus in fixed nuclei of six selected cell lines.....	39
Figure 13: Basic4Cseq analysis of sequenced chromosomal fragments interacting with VP_Locus in fixed nuclei of six selected cell lines.....	40
Figure 14: Basic4Cseq analysis of sequenced chromosomal fragments interacting with VP_Hit4 in fixed nuclei of six selected cell lines.....	42
Figure 15: Relative expression of (A) total ATP2B4, (B) its common splice variant and (C) the rare splice variant in samples of the Ghanaian study group.....	44
Figure 16: Relative expression ratios of (A) the common and (B) the rare splice variant in samples of the Ghanaian study group with respect to total <i>ATP2B4</i> expression.....	45
Figure 17: Parasitemia in infected erythrocytes of donors with different rs10900585 genotypes.....	46
Figure 18: Parasite proliferation in erythrocytes of donors with different rs10900585 genotypes.....	47
Figure 19: Regulatory elements within the locus region (3' region of intron 1-2, exon 2 intron 2-3 of <i>ATP2B4</i>).....	51
Figure 20: Hi-C data from an analysis of the resistance locus and its respective four 4C-Seq hits in K562 (Rao et al. 2014).....	54
Figure 21: Regulatory elements within the fragments of the four 4C-Seq hits of the locus region.....	55
Figure 22: Z-Scores calculated for the fragment counts in 4C-Seq libraries for VP_Locus from six selected cell lines.....	59
Figure 23: QQ-plots of the quantiles of fragmet counts in 4C-Seq libraries for VP_Locus from six selected cell lines.....	59

V List of tables

Table 1: Laboratory equipment that was applied for this study.	15
Table 2: Composition of PBS stock solution (10x).	15
Table 3: Composition of TE(PCR) stock solution (10x).	15
Table 4: Kits and ready-to-use experimental systems that were applied for this study.	16
Table 5: Cell lines that were applied in this study.	16
Table 6: Software that was applied in this study.	17
Table 7: Cell lines for 4C-Seq experiments in the context of severe malaria.	21
Table 8: Composition of lysis buffer for nuclei fixation during 4C-Seq.	22
Table 9: Composition of T4 DNA ligase reaction buffer (as provided by NEB).	23
Table 10: Composition of the PCR mix for the preparation of 4C-Seq libraries.	25
Table 11: Cycling conditions for the PCR generating 4C-Seq libraries.	25
Table 12: Sequences of primers for amplification of 4C-Seq libraries of different viewpoints.	25
Table 13: Composition of the PCR mix for genotyping by FRET-MCA.	28
Table 14: Specific cycling conditions of the PCR for assessing the HbS genotype by FRET-MCA.	28
Table 15: Cycling conditions of the PCR for genotyping by FRET-MCA.	29
Table 16: Specific annealing temperatures and melting gradients for genotyping by FRET-MCA.	29
Table 17: Composition of the reaction mix for cDNA synthesis from RNA.	30
Table 18: Cycling conditions for cDNA synthesis from RNA.	30
Table 19: TaqMan® assays used for mRNA quantification of selected genes.	31
Table 20: Composition of the PCR mix for TaqMan® Gene-Expression Assays.	31
Table 21: Cycling conditions for TaqMan® Gene-Expression assays.	31
Table 22: Reaction mix for determination of parasitemia with flow cytometry.	34
Table 23: Chromosomal fragments interacting with the resistance locus (VP_Locus).	41
Table 24: Genotype frequencies of the resistance SNPs and rs334 among the 200 Ghanaian donors of the study group.	43

1 Introduction

1.1 Severe malaria

Despite all efforts, initiatives and investments over the past years, malaria still remains one of the biggest global burdens for humans. The yearly World Malaria Report of the World Health Organization (WHO) stated in the World Malaria Report 2015 that transmission of parasites occurs in 95 countries (WHO 2016). This results in approximately 3.2 billion people at risk of being infected and developing the disease, mostly in low-income countries of the southern hemisphere (Figure 1). In these countries, malaria is the seventh most frequent cause of death (WHO 2017).

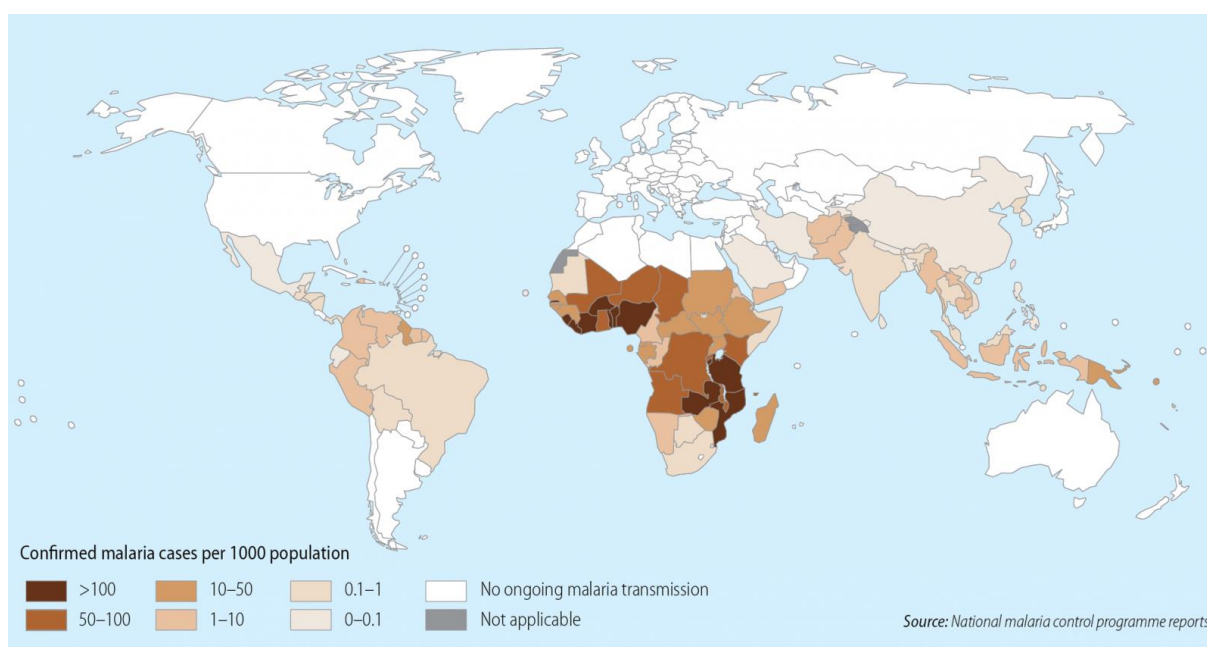


Figure 1: Global Distribution of Malaria in 2014 (www.rollbackmalaria.org).

Almost half a million people died in 2015 from the consequences of an infection with one of the five species of the genus *Plasmodium* that are known to cause the human malaras. Of these, *P. falciparum* is responsible for most of the deaths, since it predominantly causes severe malaria, the most aggressive form of the disease. It is also the most prevalent *Plasmodium* species in Africa. Therefore, 90 % of all fatal outcomes of malaria are recorded in this area. Especially children are susceptible to a fatal course of the infection, making malaria the fourth highest cause of death amongst African children under the age of five according to WHO reports for 2015.

1.1.1 The parasite life cycle

Alphonse Laveran (1881) was the first to discover parasites in the blood of a malaria patient in Algeria. Almost 20 years later, Ronald Ross (1898) described the *Anopheles* mosquito as the probable vector of malaria transmission. Roughly another 50 years later, the discovery of a part of the parasite development in the human liver led to the final piece for unraveling the complex life cycle of *P. falciparum* (Shortt and Garnham 1948). This life cycle is separated into the asexual reproduction in liver and blood of the human host and the sexual stage within the mosquito (Figure 2).

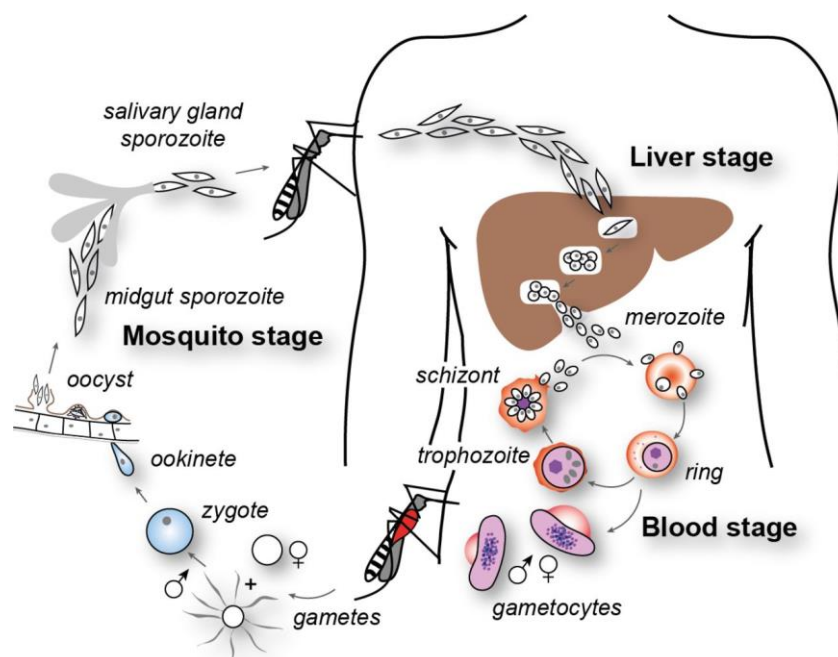


Figure 2: Life cycle of *P. falciparum* (Cowman et al. 2012).

The latter starts when a female *Anopheles* mosquito bites an infected human and takes up mature male and female gametocytes. These sexual forms further develop within the gut and finally fuse. Up to 10,000 sporozoites are produced (Aly et al. 2009) and migrate to the hemolymph through which they reach the mosquito's salivary gland and are transmitted to a new host with the next blood meal.

In the vertebrate host, sporozoites migrate from the site of the insect bite through the blood stream and invade hepatocytes. There they proliferate and form up to 30,000 merozoites, which are released into the blood stream by a sophisticated process of budding of parasite-filled vesicles (Sturm et al. 2006). There, they invade erythrocytes and pass the stages of intraerythrocytic development: a ring stage, trophozoite and schizont. The mature schizont finally gives rise to a new generation of invasive merozoites. This cycling progress of the blood stages enables a successive massive

proliferation of the parasite within the human host. The occasional development of sexual gametocytes from ring stages closes the parasite's life cycle, when taken up by a female *Anopheles* mosquito.

1.1.2 Pathology, signs and symptoms

In some cases humans do not develop any symptoms after being infected with *P. falciparum*, in others the infection results in a multi-symptomatic manifestation called severe malaria, often with a fatal outcome. All signs and symptoms of clinical malaria result from pathological processes during the blood stage of the parasite's development (Miller et al. 2002). Nonspecific flu-like symptoms are characteristic for mild clinical malaria, but may develop into the more specific clinical features and laboratory findings of severe malaria. Those include impaired consciousness or coma, cerebral convulsions, prostration, metabolic acidosis with acidotic breathing, circulatory collapse or shock, jaundice, pulmonary edema, hypoglycemia, severe anemia and renal impairment (WHO 2010).

The signs and symptoms of severe malaria are mostly explained by the property of the late blood stages of *P. falciparum* to present a set of parasite proteins on the surface of infected red blood cells (RBCs) that mediate so-called cytoadherence. These surface proteins cause infected RBCs to adhere to human endothelial cells (Udeinya et al. 1981), to form rosettes with uninfected RBCs (Udomsangpetch 1989) and to enable platelet-mediated agglutination with other RBCs (Pain et al. 2001). By adherence to the vascular endothelium, infected RBCs sequester from the circulation, a property possibly evolved to escape removal (so-called clearance) of bulky infected RBCs by the spleen. Therefore, only the slim ring stages are found in the peripheral blood of individuals infected with *P. falciparum* (Pouvelle et al. 2000). Most importantly, endothelial adherence of infected RBCs causes an impaired microvascular blood flow, which appears to be responsible for the various forms of organ damage seen in severe malaria. These include seizures and coma caused by cytoadherence in small blood vessels of the brain. Much research over the past years focused on tissue-specific receptor molecules on human endothelial cells and the parasite-derived surface proteins of infected RBCs mediating tissue-specific binding (Kraemer and Smith 2006; Sherman et al. 2003). Another clinical manifestation of severe malaria is severe anemia, which occurs due to (a) the rupture of erythrocytes when merozoites are released, (b) the accelerated

splenic removal of infected RBCs (Buffet et al. 2011) and (c) the impaired erythropoiesis that was observed in severe malaria patients (Skorokhod et al. 2010).

Adult individuals in endemic regions for *falciparum* malaria with a stable transmission of the parasite are often able to control infections (Trampuz et al. 2003). There is evidence that humans develop an immunological resistance when permanently exposed to the parasite over a couple of years. In addition, genetic factors can protect from severe malaria.

1.1.3 Genetic resistance factors

Through its high fatality, severe malaria has exerted a high selection pressure in endemic regions. Therefore, it may have had the greatest influence on the evolution of the human genome in recent history (Kwiatkowski 2005). It began approximately 10,000 years ago, when domestic agriculture was established in Africa. Humans formed centralized communities and the wells and watered fields started to serve as extensive breeding grounds for the *Anopheles* vectors. Accordingly, almost all known resistance variants in the human genome emerged in the past 10.000 years (Hedrick 2011).

Since the parasite's blood stage is responsible for the development of fatal malaria, it is plausible that most of the known genetic resistance affects RBCs. There is a group of variants that provide protection within the genes encoding the two globin chains. They either lead to a reduced expression of one of the respective molecules (β -/ α -thalassemia) or to the expression of a malformed β -globin (hemoglobinopathies). The most prominent hemoglobinopathy is the sickle-cell disease, caused by one variant of a single nucleotide polymorphism (SNP) referred to as the HbS allele. Homozygous carriers of the HbS allele suffer from sickle-cell disease that commonly has been fatal at a relatively young age. In contrast, the heterozygous genotype conveys increased resistance to severe malaria (May et al. 2007), resulting in a substantial allele frequency of the HbS allele malaria-endemic areas, in particular in the African population (Figure 3). In contrast, the allele is virtually absent in populations that are not under the evolutionary pressure of *falciparum* malaria.



Figure 3: Frequencies of the two alleles of the HbS SNP in sub populations from Phase 3 of the 1000 Genomes Project.

The abbreviations mean the following: African (AFR), American (AMR) East Asian (EAS), European (EUR) and South Asian (SAS), healthy allele (T) and disease allele (A) (www.ensembl.org)

Apart from the mutations affecting the hemoglobin proteins, mutated proteins involved in essential metabolic processes can have a protective effect when a malaria infection occurs. A deficiency of the glucose-6-phosphat dehydrogenase, for example, leads to a generally shortened lifespan of the RBC and is associated with a resistance to cerebral malaria (Luzzatto et al. 1969; Clarke et al. 2017). Also genetic alterations in erythrocyte surface molecules have an effect on the susceptibility towards malaria, as some of them are essential for the parasite invasion into RBCs or the later occurring cytoadherence (Lelliott et al. 2015). A lot of the functional mechanisms underlying the multiple genetic variants with associations to malaria resistance have not been discovered yet.

Many of these gene variants were only detected because the regional occurrence of the diseases they cause corresponds to the current or historic geographical distribution of malaria (Allison 1954). It was for example the high rate of β -thalassemia, a usually rare inherited blood disorder, in the Mediterranean population that led to the first assumption of genetic factors influencing the susceptibility towards malaria. It was the first time someone proposed that a genetic resistance to an infectious disease could act as an evolutionary force on the human genome (Haldane 1949). Although a lot of genetic factors that influence the reaction to an infection with *P. falciparum* have been detected, these cannot explain the level of susceptibility of every individual. To further unravel genetic foundations of malaria resistance, genome-wide association studies (GWAS) constitute an effective tool.

1.2 A recent GWAS on severe malaria (Timmann et al. 2012)

A large GWAS conducted among children in Ghana, associated two previously unknown resistance loci with severe malaria (Timmann et al. 2012). For this study, samples from 1,325 cases and 828 unaffected controls were included. Patients were defined as cases, when the WHO criteria for cerebral malaria (Blantyre coma score < 3) or severe anemia (hemoglobin concentration < 5 g/dl) were fulfilled in the presence of *P. falciparum* in blood smears.

After quality control, genotype data of 804,895 SNPs derived from the Affimetrix Genome-Wide Human SNP array 6.0 was used to perform a genome-wide imputation with the genotype data of 174 African individuals from the 2010-08 release of the 1000 Genomes Project. After an additional quality control of the enriched dataset, a total of 4,205,739 SNPs remained that were subsequently subjected to genome-wide statistical analyses.

SNPs from four different chromosomal loci reached significant levels of association to the trait of severe malaria in a joint analysis after replication. Two of these four loci were the previously known ABO and HbS loci. Their appearance confirms the high quality of the study group. The two novel resistance loci are located in the upstream region of the gene *MARVELD3* and in an intronic region of *ATP2B4*. Although the latter locus was replicated in several subsequent studies (Malaria Genomic Epidemiology Network 2014; Bedu-Addo et al. 2013; Band et al. 2015), the causal variant(s) and the underlying functional mechanisms remain unknown. This makes it an interesting locus for follow-up investigations.

ATP2B4 is an approximately 117kb large gene, located on chromosome 1 of the human genome (Figure 4A). The genomic sequence codes for a total of 22 potential exons that undergo differential splicing. The associated resistance locus covers approximately 7kb within the second intron of the gene (Figure 4B). It is defined by five SNPs (rs10900585, rs2365860, rs10900589, rs2365858, rs4951074), which were significantly associated in the mentioned GWAS. These SNPs are further referred to as resistance SNPs.

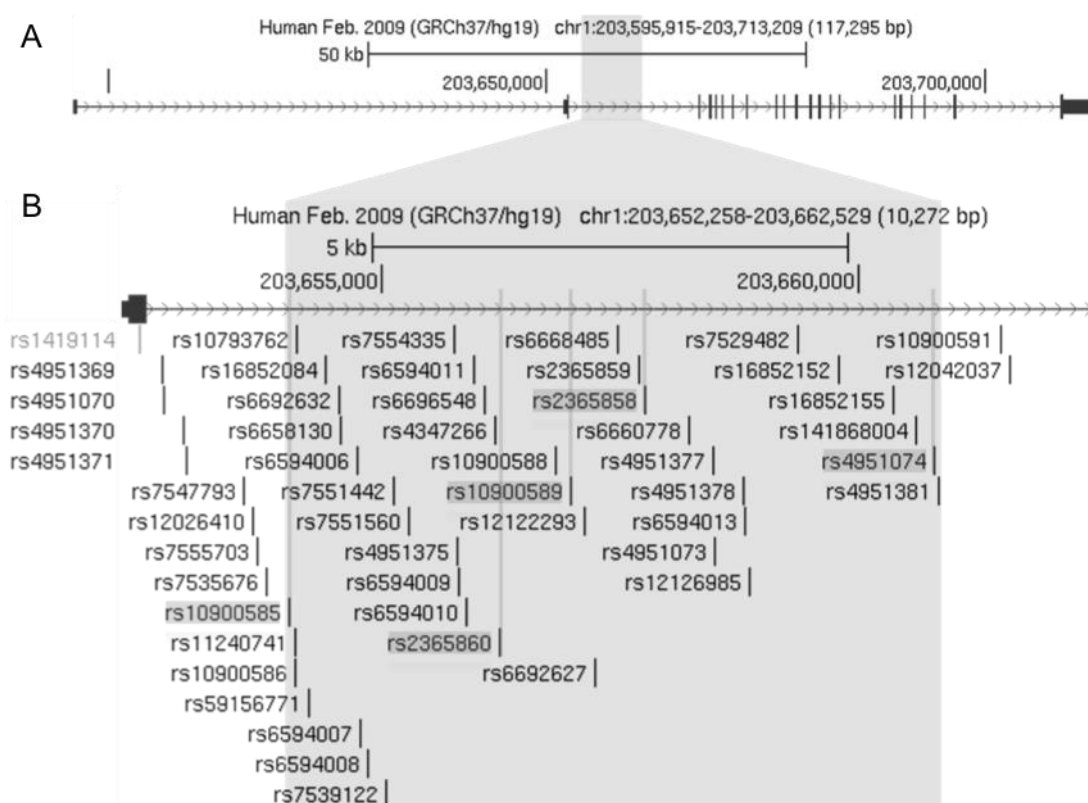


Figure 4: Schematic illustration of the resistance locus.

The image shows (A) the complete *ATP2B4* transcript NM_001684 and (B) a zoom of the chromosomal region within intron 2 (gray marked area) where the resistance SNPs map. The zoomed region shows common SNPs with a minor allele frequency $\geq 1\%$. The resistance-associated SNPs from the GWAS are highlighted with darker gray. Gray lines indicate their chromosomal position (image created on <http://genome.ucsc.edu>).

The analysis revealed a recessive mode of inheritance to be most likely, meaning that only homozygous carriers of the respective minor alleles have an increased resistance to severe malaria. Similar to the HbS allele, all associated SNPs show an allele frequency that is dependent on the geographical origin of the population. Consistent with the theory, the protective variants are most frequent within the African population that is endemically exposed to *P. falciparum*.

1.3 PMCA4

The *ATP2B4* gene encodes the plasma-membrane calcium ATPase 4 (PMCA4). PMCA4 is one of at least four known members (PMCA1,-2,-3,-4) of a family of active and thereby ATP-dependent transporter molecules that extrude calcium ions from the cell against large electrochemical gradients. It is a rather large protein (~1200aa) that is located in the plasma membrane and contains ten transmembrane domains (Figure 5). All four PMCAs show similar sequences and secondary structures. Only a minor part of their sequences is presented extracellularly. However, two large loops as well as the N- and C-terminus extend into the cytosol. The catalytic domain of the proteins, where ATP is bound and hydrolyzed is found in the second of the two cytosolic loops (marked with a P in Figure 5). In its inactive state PMCA4 is auto-inhibited, as this domain is conformationally blocked by the C-terminus of the molecule (Figure 5A).

The regulatory domain that binds the ubiquitous messenger protein calmodulin (CaM-BD) is located at the C-terminus of PMCA4 and also immerges into the cytosol. When free calcium ions (Ca^{2+}) are present in the cytosol, they form a complex with calmodulin, which then binds to the respective domain of PMCA4 (Figure 5B). This leads to a change in conformation that makes the catalytic domain accessible, reverses the auto inhibition and starts the extrusion of calcium.

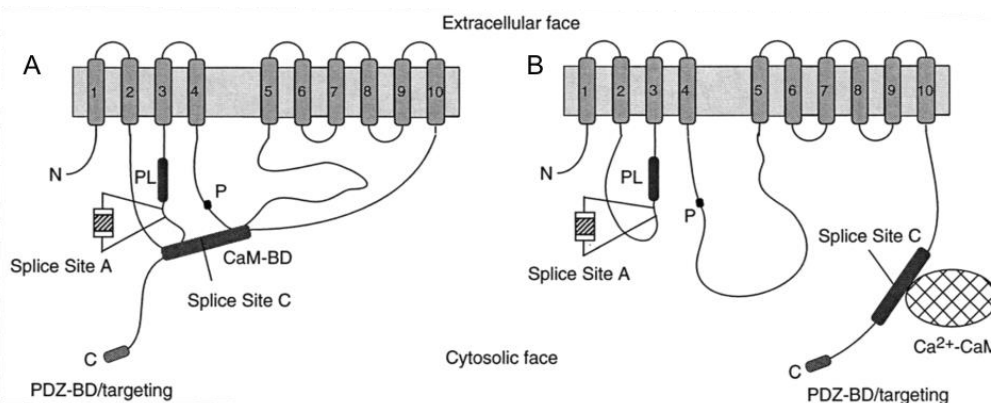


Figure 5: Schematic image of the secondary structure of PMCAs. The illustration shows the protein in its (A) autoinhibited conformation and (B) after activation by binding of the Ca^{2+} -CaM complex (Strehler 2001).

Tissue-specific differential splicing further increases the diversity of PMCA (Stauffer et al. 1995). The transcription products of all four genes show two distinct splice sites, which were verified *in vivo*. These are located close to the catalytic and regulatory domains of the encoded proteins (Strehler 2001). The functional consequences of differential splicing at splice site A are not fully examined. In contrast, it has been shown that an exon that is differentially spliced at splice site C (within the CaM-BD) has a strong influence on its affinity to the Ca²⁺-CaM complex (Caride et al. 2007). In general, the regulation of differential splicing in humans is still poorly understood, but the diversity of PMCA splice variants and their tissue specific distribution suggest that it plays a crucial role in maintaining intracellular calcium level (Di Leva et al. 2008).

Of the four known PMCA, only PMCA1 and PMCA4 are expressed ubiquitously (Krebs 2015). In the case of erythrocytes, these two molecules are the only active Ca²⁺ transporters within the membrane (Virgilio et al. 2003). PMCA4 appears to play the major role in controlling the intracellular Ca²⁺ level as it is expressed more strongly than PMCA1 (Stauffer et al. 1995). It is known that the concentration of cytosolic Ca²⁺ ions is extremely low within RBCs in contrast to the concentration in the surrounding human blood plasma (Bogdanova et al. 2013). This large electrochemical gradient needs to be maintained along the plasma membrane of red blood cells.

1.4 The calcium balance in *Plasmodium*-infected erythrocytes

In multiple studies it was shown, that the malaria parasite's development depends on Ca²⁺-mediated cellular processes (Enomoto et al. 2012), while the concentration of free cytosolic calcium is very low. This implies a demand for free calcium ions in the extracellular environment that could not be covered by the low cytosolic content of RBCs. This contradiction is resolved by the way the parasite invades the erythrocyte via membrane invagination and the parasitophorous vacuole (PV) and the corresponding membrane (PVM) are formed.

A recent study showed that the PV surrounding the parasite has a Ca²⁺ concentration of approximately 40µM (Gazarini et al. 2003) and is thereby much higher than within the parasite where the concentration is at the nanomolar level (Figure 6). In the same study multiple processes were examined that stress the importance of erythroid PMCA for the malaria parasite: (a) human PMCA were transferred to the PVM, (b) these active transporters appear to be responsible for the maintenance of the relatively high

Ca^{2+} concentration within the PV and (c) the surrounding high Ca^{2+} content is essential for a proper development of the parasite within the RBC.

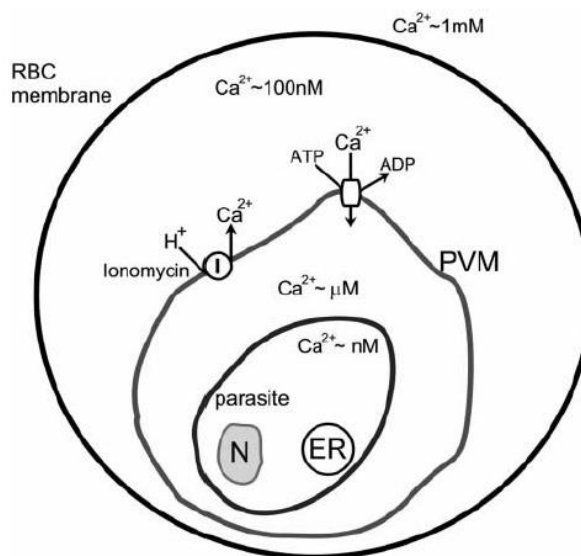


Figure 6: Schematic illustration of different compartments of an erythrocyte, infected with plasmodia. The concentration levels of free calcium ions in the different compartments are indicated (N: nucleus, ER: endoplasmatic reticulum) (Gazarini et al. 2003).

The complex expression mechanisms of *ATP2B4* and the importance of a functionally intact gene product for the development of the malaria parasite increase the probability that the associated resistance SNPs within the intronic region might have a functional impact on the gene itself. But it has to be kept in mind that intronic SNPs, even if they are causal, do not necessarily have an influence on the corresponding gene.

1.5 The challenge of interpreting GWAS results

The first successful genome-wide association study was published a bit more than a decade ago and revealed the association of two SNPs with a trait of age-related macular degeneration (Klein et al. 2005). Since then, the progress in the development of NGS techniques and genotyping arrays made the conduction of GWAS affordable for larger study populations, thereby improving the power of subsequent statistics. The possibility to elucidate the influence of human genetics on the susceptibility towards complex diseases, and the course and outcome of any possible trait, led to an increase of published GWAS. Until today more than 2,500 of these studies have been published containing over 24,000 unique SNP-trait associations for all kinds of complex diseases from psychiatric disorders, over different kinds of cancer to a variety of infectious diseases (MacArthur et al. 2017). Unfortunately, 88% of these SNPs map in non-coding regions (Edwards et al. 2013) leading to the assumption that genetic variants altering

gene regulation might play a bigger role in the manifestation of diseases than those directly changing amino acid sequences.

Examining functional consequences of associated genetic variants in non-coding regions and thereby explaining cellular mechanisms that have an influence on the susceptibility towards a certain disease represents the challenge of the so-called post-GWAS era. The regulatory mechanisms of gene expression, like transcription regulation and differential splicing, are still not fully elucidated. The discovery of more and more functions that regulatory RNAs seem to assume in these processes (reviewed by Wery et al. 2011) has even increased the complexity of the issue.

Numerous strategies to examine GWAS loci in non-coding regions have been published over the past years (Hou and Zhao 2013; Huang 2015; van der Sijde et al. 2014). In general it is advisable to start with a fine mapping approach for the associated locus to further narrow down the possible causal variants (Figure 7). The remaining variants of interest can then be further annotated by utilizing public data bases containing for example information about regulatory elements or the linkage disequilibrium (LD).

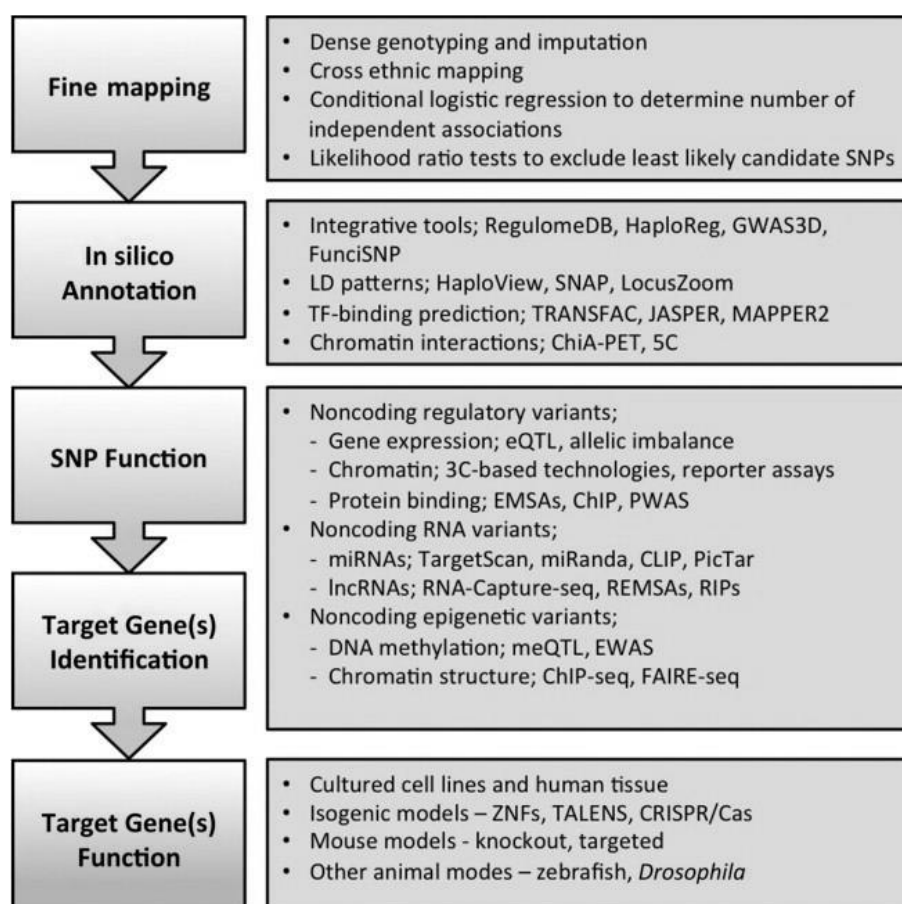


Figure 7: Workflow for functional analyses of non-coding GWAS loci (Edwards et al. 2013).

Subsequently, the SNP function is examined to possibly detect the affected gene(s). This can be done by multiple experimental approaches like the 4C method for investigation of chromatin interactions of the locus (1.5.1) or quantifying the influence different variants have on gene expression. The final step involves working with animal models or cell culture to elucidate final functional outcome of the variant in the physiological context of the analyzed trait.

1.5.1 Circular chromosome conformation capture

In numerous cases, the very vague interpretation of intronic or intergenic SNPs associated to certain traits, concentrates on the gene closest to the variation. This assumption is more likely to hold true if the regulatory region of that gene interacts with the region containing the associated SNP. These chromatin-chromatin interactions are mediated through proteins that bind to certain motifs of the DNA. To examine the physical proximity of two chromosomal regions, the so-called 3C-based technologies have been introduced (de Laat and Dekker 2012).

Among these, circular chromosome conformation capture (4C) constitutes a possibility to find chromosomal regions interacting with a certain known viewpoint region in a hypothesis-free manner. The adapted method 4C-Seq (Figure 8), which contains two enzymatic digestion and a final analysis step using next generation sequencing (NGS) was established at the University Medical Center Utrecht (Splinter et al. 2012; van de Werken et al. 2012). When using 4C-Seq to examine GWAS results, the viewpoint corresponds to the trait-associated locus.

The first step when applying this method is fixing the three-dimensional status of the chromatin within the nucleus with formaldehyde. The formaldehyde cross-links proteins to proteins and DNA. Then a restriction enzyme (RE) is applied followed by ligating the generated sticky ends. A subsequent incubation with proteinase separates the fixed chromosomal interactions and results in large circles of DNA containing multiple fragments that were in physical proximity within the nucleus. These are further digested and then incubated again with ligase to form smaller DNA circles.

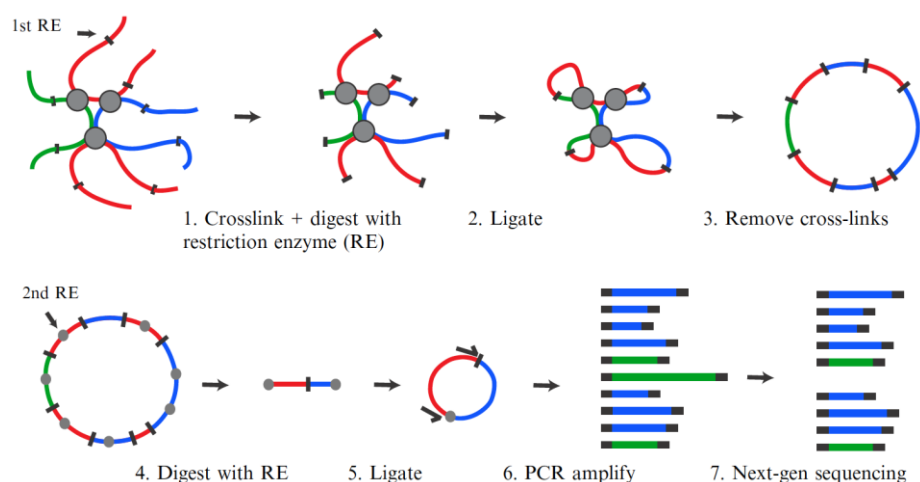


Figure 8: Schematic image of the experimental steps of 4C-Seq.

Big gray circles represent fixed protein-DNA links; black dashes show location of restriction sites of 1st RE and the resulting sticky ends; small gray circles are restriction sites of the 2nd RE and the resulting sticky ends; red is the viewpoint area; green and blue are unknown interacting chromosomal regions (van de Werken et al. 2012).

These contain a small part of known sequences of the viewpoint region, bordered by the restriction sites of the first and the second RE and ligated with interacting chromosomal fragments of unknown sequences. By designing primers that are directed from within the known sequence into the unknown, this unknown sequence is amplified. The PCR product is analyzed using NGS. Chromosomal fragments that are detected more often than random proximity could explain are assumed to specifically interact with the viewpoint region. If these fragments contain regulatory elements like promoters or enhancers, the expression of the respective gene might be influenced by the variants within the viewpoint region. The application of 4C-Seq has previously contributed to the successful investigation of functional mechanisms underlying trait association of SNPs in non-coding chromosomal regions (Smemo et al. 2014).

2 Aim and objectives

The aim of this project was to advance our understanding of the functional mechanisms underlying the association of the previously described *ATP2B4* locus with an increased resistance to severe malaria (Timmann et al. 2012). Two aspects were considered in order to achieve that.

First, computational fine mapping was conducted to identify associated SNPs that might have been undetected by previous attempts. Second, the cellular processes that might cause the statistically identified resistance of the carriers of the respective variants were examined:

(a) The chromatin-chromatin interactions of the associated locus region of *ATP2B4* were identified in order to narrow down the number of possible genes on which the resistance variants might exert a regulatory effect. For this, 4C-Seq was established and applied.

(b) The expression levels of *ATP2B4* were determined in whole blood samples from donors analyzed with respect to their variants of previously published resistance SNPs.

(c) Furthermore, the possibility of a variant-dependent effect on the parasite development within the erythrocyte was examined. The latter was achieved by establishing a cell culture experiment with infected erythrocytes from donors with different genotypes of the respective resistance SNPs.

For all functional experiments it was essential to acquire sample material from donors with the same ethnical background as those from the original study. Consequently, a blood sampling campaign with 200 donors was conducted in the Blood Donor Clinic of the Komfo Anokye Teaching Hospital in Kumasi, Ghana.

3 Material

3.1 Supplies for cell culture

All culture media, media additives and accutase to detach adherent cells were purchased from PAN (PAN-Biotech GmbH, Aidenbach, Germany). All consumables for cell culture and blood drawing were ordered from Sarstedt (Sarstedt AG & Co., Nürnberg, Germany), except for the 30- μ m cell strainers (Sigma-Aldrich Chemie GmbH, Munich, Germany) and the Nunc™ MicroWell™ 96-well plates (Thermo Fisher Scientific Inc., Waltham, USA). The Plasmodipur filters which were used to purify erythrocytes for parasite cultures were manufactured by Europroxima (Europroxima BV, Anrhem, The Netherlands).

3.2 Enzymes and enzyme inhibitors

Restriction enzymes, ligase and the commercially available respective buffer solutions were ordered from NEB (New England Biolabs, Ipswich, USA). RNase, Proteinase K as well as proteinase inhibitors were purchased from Roche (F. Hoffmann-La Roche AG, Basel, Switzerland).

3.3 General chemicals, water and lab consumables

If not stated otherwise, all chemicals and solvents were purchased from Merck (Merck KGaA, Darmstadt, Germany), Sigma-Aldrich (Sigma-Aldrich Chemie GmbH, Munich, Germany), Biomol (Biomol GmbH, Hamburg, Germany) and Roth (Carl Roth GmbH + Co. KG, Karlsruhe, Germany). For preparation of PCR mastermixes and dilution of template nucleic acid, nuclease-free water from Qiagen (QIAGEN GmbH, Hilden, Germany) was used. If solutions were subsequently autoclaved, they were prepared with deionized water, otherwise they were prepared with HPLC-grade water (Merck KGaA, Darmstadt, Germany). All plastic consumables were purchased from Sarstedt (Sarstedt AG & Co., Nürnberg, Germany).

3.4 PCR primer and fluorescence-labeled probes

All PCR and sequencing primers were ordered from Eurofins (Eurofins MWG GmbH, Ebersberg, Germany). The fluorescence-labeled probes for genotyping were synthesized by Biomers (biomers.net GmbH, Ulm, Germany).

3.5 Lab equipment

Table 1: Laboratory equipment that was applied for this study.

device	type designation	distributor
bioanalyzer	2100	Agilent Technologies, Inc., Santa Clara, USA
biological safety cabinet	51424/5 Microflow	Nunc GmbH & Co. KG, Wiesbaden, Germany
centrifuges	4515 R, 4515 C, 4515 D, 5804R	Eppendorf AG, Hamburg, Germany
counting chamber	Improved Neubauer, bright line double ruling	BRAND GMBH + CO KG, Wertheim, Germany
flow cytometer	Accuri C6, LSRII	Becton, Dickinson and Company, Franklin Lakes, USA
fluorometer	Qubit™ 3.0	Thermo Fisher Scientific Inc., Waltham, USA
heating block	5436, comfort	Eppendorf AG, Hamburg, Germany
incubator	New Brunswick Innova Co-170	Eppendorf AG, Hamburg, Germany
PCR cyler	T3 Thermocycler	Biometra GmbH, Göttingen, Germany
pipetting robot	freedom evo	Tecan Group Ltd., Männedorf, Switzerland
RT-PCR cyler	LightCycler 480 Instrument	F. Hoffmann-La Roche AG, Basel, Switzerland
tube roller mixer	SRT2	Sigma-Aldrich Chemie GmbH, Munich, Germany
water bath	1083	Gesellschaft für Labortechnik mbH, Burgwedel, Hamburg

3.6 Buffers

PBS (10x)

Table 2: Composition of PBS stock solution (10x).

component	final concentration
NaCl	1,37 M
Na ₂ HPO ₄	160 mM
NaH ₂ PO ₄ ; pH 7,2	40 mM

TE(PCR) (10x)

Table 3: Composition of TE(PCR) stock solution (10x).

component	final concentration
EDTA; pH 8	2 mM
TrisHCl; pH 8	100 mM

3.7 Kits and ready-to-use experimental systems

Table 4: Kits and ready-to-use experimental systems that were applied for this study.

	distributor
Agilent 7500 DNA Kit	Agilent Technologies, Inc., Santa Clara, USA
Expand™ Long Template PCR System	F. Hoffmann-La Roche AG, Basel, Switzerland
High capacity cDNA Reverse Transcription Kit with RNase Inhibitor	Thermo Fisher Scientific Inc., Waltham, USA
High Pure PCR Product Purification Kit	F. Hoffmann-La Roche AG, Basel, Switzerland
mag™ maxi kit	LGC Genomics GmbH, Berlin, Germany
QIAquick PCR Purification Kit	QIAGEN GmbH, Hilden, Germany
Qubit™ quantitation assays	Thermo Fisher Scientific Inc., Waltham, USA
TaqMan®	Thermo Fisher Scientific Inc., Waltham, USA

3.8 Cell lines

Table 5: Cell lines that were applied in this study.

cell line	source
KMOE-2	ordered from DSMZ; ACC 37
HEL	stock (in 2005 ordered from DSMZ; ACC 11)
TF1	ordered from DSMZ; ACC 334
K562	stock (in 2005 ordered from DSMZ; ACC 10)
HepG2	stock (in 2008 obtained in-house from Dr. Volker Heussler)
HPMEC	stock (in 2010 obtained from Dr. James Kirkpatrick, Johannes Gutenberg University Mainz)

3.9 Samples for functional experiments

Blood samples were taken from regular blood donors of the Blood Donor Clinic of the Komfo Anokye Teaching Hospital (KATH) in Kumasi, Ghana. Ethical clearance for the study was obtained from the committee on Human Research Publication and Ethics from the Kwame Nkrumah University of Science and Technology (KNUST) and the committee of ethics of the regional Medical Association Hamburg, Germany. Informed consent was obtained from each blood donor recruited for the study. Information about ethical background, age and sex were obtained from the study participants. The ethics represented in the study group are Akan (79.2 %), Ewe (4.0 %), Grusi (2.5 %), Mole-Dagbani (1.9 %) and Ga Adangbe (1.5 %). The average age is 32 years (17 y – 59 y). Among the participants, 86.1 % were male and 13.9 % were female. The blood groups

were determined by the serological laboratory of the KATH. The distribution of blood groups in the study population shows 53.0 % participants with blood group O⁺, 21.3 % with blood group B⁺, 14.4 % with blood group A⁺, 5.4 % with blood group O⁻, 3.0 % with blood group AB⁺, 2.0 % with blood group B⁻, 1.0 % with blood group A⁻ and no participant with blood group AB⁻.

3.10 Software and online tools

Table 6: Software that was applied in this study.

	applied version
PLINK	1.07
Gtool	0.7.5
IMPUTE2	2.3.0
SNPtest	2.4.1
LocusZoom	http://locuszoom.org/genform.php?type=yourdata ; 18.11.2016, 11:37
Haploview	4.2
R	3.2.0
FastX-toolkit	0.0.14
Bowtie	0.12.8
SAMtools	0.1.18
Haploreg	4.1
Basic4CSeq	1.6.0
FourCSeq	1.4.0
LightCycler 480 Software	release 1.5.0 SP3

4 Methods

4.1 Imputation

Genotype data from (a) the Affymetrix Genome-wide Human SNP Array 6.0 was merged with (b) a set of genotype data generated by melting curve analysis (MCA) with assays based on fluorescence resonance energy transfer (FRET) before imputation of reference data from the 1000 Genomes Project (Phase 3 v5; Auton et al. 2015). The computational fine mapping consisted of multiple steps (Figure 9)

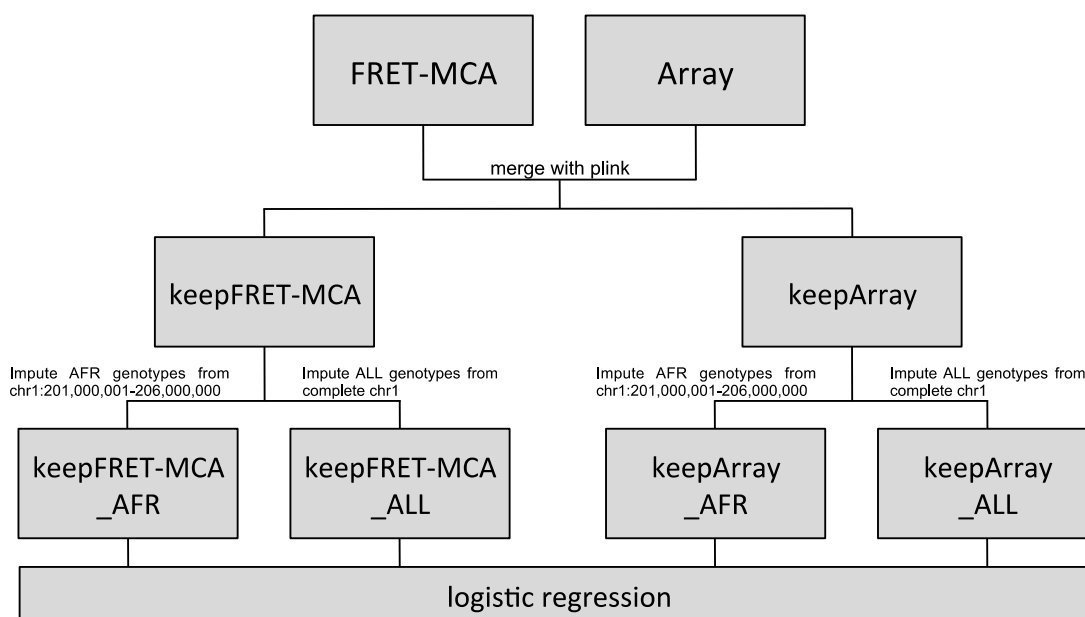


Figure 9: Flow chart of the steps of *in silico* fine mapping from raw data to statistical analysis.

The first dataset consisted of raw genotype data generated by FRET-based genotyping of the 2,153 individuals from the previous GWAS and in part from the donors of the mentioned replication groups. The data covered 102 SNPs from chromosome 1. It was subjected to quality control (QC) to reach the same requirements as the data from the first set. The QC was performed using PLINK. Genotypes were recoded from binary to base format and adjusted to the sequence of the forward strand. For some SNPs, the recoding had to be performed manually with respect to the published minor alleles (Ensembl genome browser 88, Yates et al. 2016). If the minor allele frequency (MAF) exceeded 45 % the SNP was excluded from the dataset to avoid false allele assignment. All chromosomal positions were updated to GRCh37/hg19 (February 2009) and genotype data of all individuals that do not belong to the original GWAS cohort was excluded. Finally, only SNPs reaching a genotyping rate of at least 0.95, a MAF of 0.01 and a Hardy-Weinberg-Equilibrium (HWE) of 10^{-6} were considered further.

The other set of data from the SNP array contained the filtered genotypes after QC as previously applied and analyzed in the mentioned GWAS (Timmann et al. 2012). In contrast to that previous study, a MAF threshold of 0.01 was applied for this study. In total, 63,371 genotyped SNPs on chromosome 1 were included.

Merging the datasets was performed using PLINK. Duplicates were treated in two different ways. For one merged dataset, the genotypes as determined by FRET-MCA were kept in case of duplicate SNPs (further referred to as “keepFRET-MCA”). For a second set the genotypes as originally identified by SNP array were considered (further referred to as “keepArray”). Subsequent analyses were performed with both merged datasets, respectively.

To convert the data from PLINK format (PED/MAP) to the format required for the imputation, the software Gtool was used. The PED-files were converted to GEN-files and SAMPLE-files were generated from the original MAP-files. The haplotype estimation (“pre-phasing”) was performed using Impute2 (Howie et al. 2009), the software that was also applied for subsequent imputation. The chromosomes were separated into segments of 5 Mb and the phasing was performed with a 250 kb buffer to each side of the segments. GRCh37/hg19 was used as the reference genome. For all statistical parameters influencing the accuracy of the pre-phasing (flags `-k`, `-iter`, `-burnin`, `-Ne`), the default values as recommended by the authors of Impute2 were applied (Howie et al. 2012).

In the following step, genotypes determined and published by the 1000 Genomes Project (Phase 3 v5; Auton et al. 2015) and not contained in the pre-phased dataset were imputed (default for `-Ne` was used). Two different imputations were performed using (a) only the genotype data of the African individuals (further referred to as AFR) and (b) using genotype data from all individuals of the 1000 Genomes Project (further referred to as ALL). For the AFR-imputation only genotypes from the area surrounding the *ATP2B4* locus (chr1: 201,000,001-206,000,000) were imputed.

The merged and imputed genotype files were converted back into PLINK format using the Gtool software. They were then subjected to QC, by excluding all SNPs with a MAF smaller than 0.01 and an info score smaller than 0.5. The info score is a value that Impute2 assigns to every imputed SNP, reflecting the certainty with which the genotype of this SNP was imputed. It ranges from 0 (very uncertain) to 1 (very certain). After QC, four different datasets (keepArray_AFR, keepArray_ALL, keepFRET-MCA_AFR, keepFRET-MCA_ALL) were prepared for statistical analyses.

4.1.1 Association testing

For all SNPs from the datasets, the association with the trait of severe malaria was calculated by performing logistic regression. The calculation was accomplished under the assumption of the previously determined recessive mode of inheritance of the *ATP2B4* locus. Furthermore, it included corrections for age, gender and the population structure represented by a principal component analysis (PCA). This resembles the proceeding for the previous GWAS. The calculations for the *ATP2B4* locus were performed using the software PLINK. A p-value $< 10^{-4}$ was defined as a threshold for an association with the trait.

4.1.2 Visualization of association data

The results of the logistic regression were visualized using the online tool of the LocusZoom software (Pruim et al. 2010). The designated index SNP was rs10900585, the best associated SNP from the published GWAS. As a complementary feature, the tool offers the possibility to plot the recombination rates. These are calculated from the data of African individuals of the 2012 release of the 1000 Genomes Project (Abecasis et al. 2012). All chromosomal positions refer to GRCh37/hg19. Data from chr1:203,595,000-203,750,000 was included for an image, allowing an overview. The area containing all SNPs in an LD of $r^2 > 0.2$ with the index SNP rs10900585 was defined as region of interest.

4.1.3 LD analysis

Analyses of the linkage disequilibrium were performed using the software Haploview. For this, genotype data of all SNPs that showed an association with the trait of severe malaria (p-value $< 10^{-4}$ [4.1.1]) were entered. The r^2 -values were calculated as a measure of linkage between the different SNPs. The threshold for tagging SNPs was determined to be $r^2 > 0.8$.

The association analyses of haplotypes were performed using PLINK. Different haplotypes of the five published resistance SNPs were tested by logistic regression.

4.2 Quantification of nucleic acid

For this project, quantification of DNA and RNA was performed using the Qubit™ 3.0 fluorometer and the respective Qubit™ quantitation assays following the manufacturer's instructions. Only the quantification of 4C-Seq libraries for multiplexing equal molarities before sequencing was performed with a bioanalyzer, using the Agilent 7500 DNA Kit.

4.3 4C-Seq

The execution of the experimental steps for the 4C-Seq experiments followed the published standards (Splinter et al. 2012). Incubation times and temperatures were optimized. The viewpoint sequences determined the choice of restriction enzymes and the design of the applied primers.

4.3.1 Cell culture

For 4C-Seq experiments, cell lines representing the different tissues and cells involved during pathophysiological processes of severe malaria were chosen (Table 7). The erythroid cell lines were selected due to their potential of induced hemoglobin expression.

Table 7: Cell lines for 4C-Seq experiments in the context of severe malaria.

cell line	represented malaria stage	origin	culture medium	medium additives
KMOE-2	blood stage	erythroid leukemia	RPMI 1640	20 % FCS, 100 U/ml penicillin, 10 µg/ml streptomycin
HEL	blood stage	erythroid leukemia	RPMI 1640	10 % FCS, 100 U/ml penicillin, 10 µg/ml streptomycin
TF1	blood stage	erythroid leukemia	RPMI 1640	20 % FCS, 100 U/ml penicillin, 10 µg/ml streptomycin, 5 ng/ml GM-CSF
K562	blood stage	erythroid leukemia	RPMI 1640	10 % FCS, 100 U/ml penicillin, 10 µg/ml streptomycin
HepG2	liver stage	hepatocellular carcinoma	MEM	10 % FCS, 100 U/ml penicillin, 10 µg/ml streptomycin
HPMEC	endothelial sequestration of IEs	pulmonary microvascular endothelium	Endopan MV	

Various cell lines were cultivated in different media according to their respective requirements (Table 7). All cells were cultivated in T25 or T75 cell culture flasks and were permanently kept under sterile conditions.

4.3.2 Fixation of nuclei

Adherently growing cells were detached from the bottom of the cell culture flask by removal of culture medium and incubation in 300 μ l/25 cm² accutase solution (>500 U/ml) for 10 min at 36 °C. The cells were suspended in 10 ml of the respective culture medium and were subsequently handled like the suspension cells.

Suspension cells were counted by using a Neubauer chamber. They were washed once in PBS and then resuspended in 10 ml PBS with 10 % FCS per 1 x 10⁷ total cells in the batch. The cell suspension was applied to a cell strainer with 30- μ m pores. Per 10 ml cell suspension, 540 μ l of 37 % formaldehyde were added after that to obtain a 2 % concentration of formaldehyde in the solution. After 10 min of rolling incubation at RT, 712 μ l of 2 M glycine were added per 10 ml cell suspension. The suspension was immediately placed on ice. After centrifugation (400 x g, 8 min, 4 °C), the supernatant was discarded and cells were resuspended in 1 mL ice cold lysis buffer (Table 8) per 1 x 10⁷ total cells in the pellet.

Table 8: Composition of lysis buffer for nuclei fixation during 4C-Seq.

component	final concentration
TrisHCl	50 mM
NaCl	150 mM
EDTA	5 mM
Igepal	0,5%
TX-100	1%
cComplete™ Protease Inhibitor Cocktail	1 x

The suspension was incubated on ice until complete lysis of cells was achieved (10-15 min). The progress of lysis was checked microscopically by incubating small aliquots of the suspension in methyl-green pyronin solution. The lysed cells were centrifuged (900 x g, 5 min, 4 °C) and the supernatant was discarded. The pellet was then thoroughly resuspended in ice cold PBS to a concentration of 1 x 10⁷ cells/ml. The suspension was distributed by pipetting 1 ml each into 2 ml reaction tubes. The reaction tubes were centrifuged and after discarding the supernatant, snap-frozen in liquid nitrogen. The fixed nuclei pellets were stored at -80°C until further use. They were suitable for application up to a maximum of six months.

4.3.3 Digestion of nuclei

For the digestion steps of the 4C-Seq protocol, appropriate restriction enzymes had to be selected. For the first restriction BsrGI was selected, an enzyme with a 6-bp restriction site. The localization of these sites defines the borders of the experiment's viewpoint region. Digesting the human DNA with BsrGI produces a viewpoint fragment containing all resistance-associated SNPs from the locus that was identified in the GWAS (VP_Locus, Figure 10).

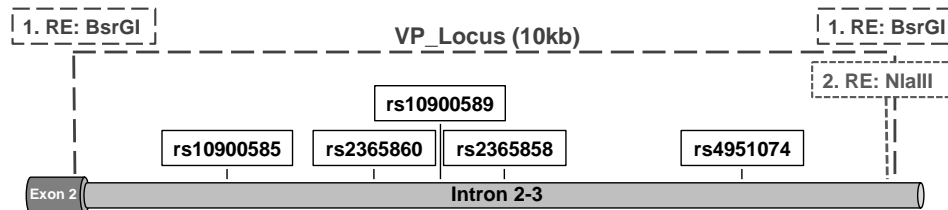


Figure 10: Schematic figure of the viewpoint region (VP_Locus) and restriction sites.

NlaIII, a restriction enzyme with a 4-bp restriction site was chosen for the second digestion. It cuts next to the downstream restriction site of BsrGI.

First restriction

The frozen nuclei pellets were resuspended in 450 μ l water (HPLC-grade). Then, 60 μ l 10 x NEBuffer 2.1 and 10 μ l 10% SDS were added and the sample was incubated for 10 min in a block heater at 65 $^{\circ}$ C and 900 rpm shaking. Triton-X 100 was added (30 μ l of a 20% solution) and the reaction tube was transferred for 10 min to a 37 $^{\circ}$ C heat block, shaking at 900 rpm.

This was followed by a step-wise addition of the restriction enzyme BsrGI, by first adding 2 x 200 U BsrGI with 4 h incubation in between. Then, the sample was incubated overnight and 1 x 200 U BsrGI were added in the morning followed by 4 h of incubation. During the complete time, the suspension was kept at 37 $^{\circ}$ C with 900 rpm shaking. BsrGI was then inactivated by 20 min incubation at 80 $^{\circ}$ C. The complete sample was transferred to a 50 ml reaction tube and diluted in 5,7 ml water (HPLC-grade) and 700 μ l ligase buffer (Table 9). After that, 1000 U T4 DNA-Ligase were added and the reaction tube was incubated in a 16 $^{\circ}$ C water bath overnight.

Table 9: Composition of T4 DNA ligase reaction buffer (as provided by NEB).

component	final concentration
TrisHCl	50 mM
MgCl ₂	10 mM
ATP	1 mM
DTT	10 mM

The ligated DNA was purified by adding 30 μ l Proteinase K (10 mg/ml) to the suspension and incubating it in a water bath while shaking at 60 $^{\circ}$ C overnight. Subsequently, 30 μ l RNase A (10 mg/ μ l) were added and the solution was incubated at 37 $^{\circ}$ C while 900 rpm shaking for 45 min. Recovery of DNA was accomplished by a phenol-chloroform extraction. For this, 7 ml of a premixed phenol-chloroform solution for DNA extraction was added and after thorough vortexing the solution was centrifuged for 15 min at 3.500 x g. The upper phase was transferred to a 50 ml reaction tube and diluted with 7 ml water (HPLC-grade), 1 ml of a 3 M sodium-acetate solution (pH=5,6), 5 μ l glycogen (20 mg/ml) and 35 ml absolute ethanol. After mixing, the solution was frozen at -80 $^{\circ}$ C. Subsequently it was thawed and centrifuged at 10.000 x g and 4 $^{\circ}$ C for 20 min. The supernatant was discarded and the pellet was washed in 10 ml 70 % ethanol (10.000 x g, 15 min, 4 $^{\circ}$ C). The supernatant was again discarded and the pellet was dried at RT overnight. It was then resuspended in 200 μ l 1 x TE.

Second restriction

The 200 μ l DNA suspension from the previous step was transferred to a 2 ml reaction tube and diluted in 500 μ l water (HPLC-grade), 50 μ l CutSmart Buffer and 50 U NlaIII. The sample was incubated in a block heater at 37 $^{\circ}$ C and 900 rpm overnight. The RE was heat-inactivated for 20 min at 65 $^{\circ}$ C. The complete suspension was transferred to a 50 ml reaction tube and diluted in 12.1 ml water (HPLC-grade) and 1.4 μ l ligase reaction buffer (Table 9). Subsequently, 1000 U T4 DNA ligase were added and the reaction tube was incubated in a 16 $^{\circ}$ C water bath overnight.

Again, 500 μ l of a 3 M sodium-acetate solution (pH=5,6), 5 μ l glycogen (20 mg/ml, ABI) and 35 ml absolute ethanol were added and the solution was frozen at -80 $^{\circ}$ C. After thawing a centrifugation step followed (10.000 x g, 40 min, 4 $^{\circ}$ C) and after discarding the supernatant the remaining pellet was diluted in 10 ml 70 % ethanol. After centrifugation (10.000 x g, 30 min, 4 $^{\circ}$ C), the supernatant was discarded and the pellet was dried at RT overnight. The pellet was dissolved in 150 μ l TrisHCl and then purified with the QIAquick PCR Purification Kit (Qiagen), following the manufacturer's instructions.

4.3.4 Preparation of sequencing library

The primers for the PCR generating the sequencing libraries were designed according to the sequence of the fragment between the sites of the first and the second restriction enzymes (Figure 10). The previously purified DNA was diluted and distributed into 100 μ l aliquots containing 3.2 μ g DNA. These were then used for the amplification of

viewpoint-bound DNA fragments. The composition of the PCR mix (Table 10) and the settings of the cycling conditions (Table 10) were chosen as suggested by the manual of the Expand™ Long Template PCR System.

Table 10: Composition of the PCR mix for the preparation of 4C-Seq libraries.

component	volume per reaction [μ l]	final concentration
Buffer 1 (10x)	80	1x
dNTP mix (10mM)	16	0,2 mM
reading Primer (100 μ M)	11,2	1,4 μ M
index Primer (100 μ M)	11,2	1,4 μ M
Polymerase (5U/ μ l)	11,2	0,07 U
template-DNA (32 ng/ μ l)	100	25ng/ μ l
water (nuclease-free)	570,4	
sum	800	

Table 11: Cycling conditions for the PCR generating 4C-Seq libraries.

	temperature[$^{\circ}$ C]	duration
initialization	94	5 min
denaturation	94	15 sec
annealing	55	1 min
elongation	68	3 min
final elongation	68	5 min
storage	8	∞

30x {

Sequencing libraries were prepared for different viewpoints. DNA fragments bound to the *ATP2B4* locus (VP_Locus) were initially sequenced in all selected cell lines. The identified interacting fragments were then used as viewpoint regions for the preparation of further sequencing libraries. These reciprocal experiments were performed in all six cell lines with four viewpoints that were identified to interact with the locus (VP_Hit1, VP_Hit2, VP_Hit3, VP_Hit4). Depending on the viewpoint, different primer sets were designed for the preparation of the sequencing library (Table 12).

Table 12: Sequences of primers for amplification of 4C-Seq libraries of different viewpoints.

viewpoint	reading primer	index primer
VP_Locus	TGACTTCTTAAGGTTGCTATTGTACA	GGGCAGGTCAAGGCTCTTAG
VP_Hit1	CCAGAGAATTAGTTGTTTTGATTGTACA	ACAGCAAACCCCGTCTCTA
VP_Hit2	CCCTAATACTTTTGCCTGAATGTACA	ATTTCCCTTCTCTCCCTGC
VP_Hit3	TCTGAAAGGGAAACACAAATGTACA	TCCTCACTTGCCTCAGACA
VP_Hit4	CTTTGTCTCTATGTGTGTGTGTACA	CAGAGGCAGATAGACAGTCAGA

To be able to multiplex libraries during sequencing, 2-bp barcodes were integrated prior to the reading primers. Furthermore, the Illumina Adapter sequences P5 and P7 were

added in front of the barcoded reading primer and the index primer, respectively. Lists of complete primer sequences can be found in the Appendix (Table A - 2, Table A - 3). The sequencing reaction was performed at BGI (Szhenzhe, China) on the Illumina HighSeq platform, generating 49-bp single end reads.

4.3.5 Data analysis

Sequence datasets were de-multiplexed according to their cell line-specific barcodes and their viewpoint-specific reading primer, using the FastX-Toolkit. For selecting sequences according to the 2-bp barcode, the FastX-barcode-splitter was applied. The first two 5' nucleotides were scanned and the sequences separated into groups, allowing for no mismatches. After trimming off the barcodes with the FastX-trimmer, the groups were again divided according to their 5' nucleotides using the FastX-barcode-splitter. This time, one mismatch in the sequences of the viewpoint-specific primers was allowed. The primer sequences were then trimmed off, so that only the sequences of viewpoint-annealing fragments remained.

Those were aligned to the human genome (GRCh37/hg19) using the software Bowtie. The -n alignment mode was chosen allowing for a maximum of two mismatches in the sequences (-n 2, -l 100). The selection of reported alignments was defined and reads that were aligned more than once were excluded (-m 1, --best, --strata). The resulting files in the SAM-format were reformatted into BAM-files using the software SAMtools.

To analyze and visualize the data of the sequenced libraries of different cell lines and viewpoints, two published analysis pipelines for 4C data were used. Basic4C-Seq (Walter et al. 2014) and FourC-Seq (Klein et al. 2015b) are both R-packages. They were applied as described in the respective manuals (Walter et al. 2015; Klein et al. 2015a). Basic4Cseq is a tool that can be used to normalize the number of counted fragments and visualize sequencing data from 4C-Seq experiments. FourCSeq has the additional feature of statistical analyses that compare fragment counts of different libraries from experiments with identical primers to identify significantly increased values representing chromosomal regions that interact with the viewpoint region.

4.4 Genotyping of donor samples for functional experiments

4.4.1 DNA extraction from dried blood spots

For genotyping of the 200 Ghanaian blood donors, DNA was extracted from dried blood spots. For that, components of the magTMmaxi kit, a ready-to-use magnetic bead-based system, were applied. Each spot contained 75 μ l whole blood. For DNA extraction, one spot from the Spot Saver Card was cut into 4-6 pieces and transferred into a 15 ml reaction tube. They were covered with 100 μ l water (HPLC-grade), 250 μ l buffer BL, 25 μ l proteinase K (20 mg/ml) and incubated shaking in a 55 °C warm water bath overnight. The pieces of filter paper were then removed and the remaining volume of the mixture was estimated. One estimated volume of buffer BLM and absolute ethanol were added to the mixture, respectively. After adding 20 μ l of the resuspended magnetic-beads suspension the sample was incubated rolling at RT for at least two hours.

Subsequently, the DNA bound to the magnetic beads was washed by first spinning the suspension (100 x g, 1 min, RT) and then applying the tubes to a magnetic rack. The supernatant was removed and 0.9 ml wash buffer 1 was added. After removing the tubes from the rack, the magnetic beads were resuspended in the buffer and transferred to a 2 ml reaction tube. To ensure the complete transfer of the beads, the ones remaining in the 15 ml tube were resuspended in another 0.9 ml wash buffer 1 and transferred to the smaller tube. The latter was then incubated for 10 min shaking at RT. It was again centrifuged (100 x g, 1 min, RT) before applying it to a magnetic rack and discarding the supernatant. This washing step was repeated another three times always using 1.8 ml wash buffer 1. After discarding the wash buffer for the third time, the beads were dried with open lid at 55 °C for 10 min.

The elution of DNA from the beads was performed in two steps. First 70 μ l TE(PCR) were added to the dried beads and incubated for 30 minutes shaking at 55 °C. The suspension was centrifuged (100 x g, 1 min, RT) and the tube subsequently applied to a magnetic rack. The supernatant containing the DNA was transferred to a new reaction tube. A second elution step was performed using another 100 μ l TE(PCR). After incubation, centrifugation and application to the magnetic rack as previously, the supernatant was pooled with the other DNA suspension. Samples were kept at 4-8 °C until application in FRET-MCA assays for genotyping.

4.4.2 FRET-MCA assay

For genotyping of different polymorphisms in the 200 blood samples from Ghanaian donors, FRET-MCA assays were performed using the DNA solution isolated from dried blood spots (4.4.1). The genotypes of the HbS polymorphism and the five published resistance SNPs from the *ATP2B4* locus were determined. For each sample the amplification was performed by preparing the different reaction mixes (Table 13). The appropriate amounts of water, MgCl₂ and primers varied between the different assays, following established laboratory routines. A list of complete primer sequences can be found in the Appendix (Table A - 1).

Table 13: Composition of the PCR mix for genotyping by FRET-MCA.

component	volume per reaction [μl]	final concentration
Buffer BD, no MgCl ₂ (10x)	1	1x
Solution S	0 - 1	0 - 1x
MgCl ₂ (25mM)	0.8 - 1.6	2 - 4 mM
forward Primer (100μM)	0.01 - 0.05	0.1 - 0.5 μM
reverse Primer (100μM)	0.01 - 0.05	0.1 - 0.5 μM
Anchor (100μM)	0.02	0.2 μM
Sensor (100μM)	0.02	0.2 μM
dNTPs (10 mM)	0.2	200 μM
Polymerase (5U/μl)	0.2	1 U
template-DNA	2	
water (nuclease-free)	4.3 - 5.7	
sum	10	

The FRET-MCA assay for assessing the HbS genotype had a special PCR program (Table 14). The general PCR program for the other five assays (Table 15) was run with assay-specific annealing temperatures

Table 14: Specific cycling conditions of the PCR for assessing the HbS genotype by FRET-MCA.

	temperature [°C]	duration
45x {	initialization	94 2 min
	denaturation	94 15 sec
	annealing	63 30 sec
	elongation ramp	up to 75 0.5 °C / sec
	elongation	75 2 sec
	final elongation	72 2 min
	35 2 min	
storage	8 ∞	

Table 15: Cycling conditions of the PCR for genotyping by FRET-MCA.

	temperature [°C]	duration
initialization	94	4 min
denaturation	94	20 sec
annealing	see Table 16	30 sec
elongation	72	30 sec
final elongation	72	5 min
storage	8	∞

45x

Melting curve analyses were performed with the Light Cycler 480. The corresponding software was applied to determine the genotypes. Sensor-specific melting gradients were chosen while recording the fluorescence signal during melting (Table 16).

Table 16: Specific annealing temperatures and melting gradients for genotyping by FRET-MCA.

	annealing temperature [°C]	melting gradient [°C]
HbSC	63	35 - 80
rs10900585	58	40 - 75
rs2365860	55	40 - 70
rs10900589	55	40 - 70
rs2365858	55	40 - 70
rs4951074	55	40 - 70

4.5 Gene-expression assays

The expression of *ATP2B4* and its two known splice variants was determined by qPCR.

4.5.1 Samples

Gene-expression assays were performed in whole blood samples of all 200 donors from the Ghanaian cohort. For this, 200 µl of freshly drawn, anti-coagulated blood (CPDA-1) were pipetted into 1.5 ml cryotubes. Under sterile conditions, 750 µl Tri Reagent BD and 20 µl of 5 N acidic acid were added. The samples were mixed by vortexing and stored at -80°C until isolation of RNA.

4.5.2 RNA isolation and cDNA synthesis

The RNA isolation was performed following the supplier's instructions for RNA isolation from whole-blood samples with Tri Reagent BD. Samples were thawed and brought to room temperature. After thoroughly vortexing them, they were incubated at RT for 5 min and 200 µl chloroform were added. The samples were shaken for 15 sec and then kept at RT for 3 minutes. For phase separation, the samples were centrifuged

(12.000 x g, 15 min, 4 °C) and the upper phase was transferred to a fresh 1.5 ml reaction tube. It was mixed with 500 µl Isopropanol and left at RT for 10 min. After another centrifugation step (12.000 x g, 8 min, 4 °C) the supernatant was removed completely. The RNA was washed with 75 % ethanol (12.000 x g, 5 min, 4 °C) and the supernatant was again completely removed. The pellet was air-dried for 10 min with an open lid and solubilized in 100 µl nuclease-free water by slowly shaking at 56 °C for 10 min. The solution was either directly used for cDNA synthesis or stored at -80°C.

Synthesis of cDNA from isolated RNA was performed using the High Capacity cDNA Reverse Transcription Kit with RNase Inhibitor. For each sample the reaction mix (Table 17) and settings of the thermocycler (Table 18) were chosen according to the manual.

Table 17: Composition of the reaction mix for cDNA synthesis from RNA.

component	volume per reaction [µl]	final concentration
RT-Buffer (10x)	2	1x
dNTP mix (100mM)	0,8	4 mM
RT-random Primer (10x)	2	1x
MultiScribe Reverse Transcriptase (50 U/µl)	1	2,5 U/µl
RNase Inhibitor (20 U/µl)	1	1 U/µl
template-RNA	10	
water (nuclease-free)	3,2	
sum	20	

Table 18: Cycling conditions for cDNA synthesis from RNA.

	temperature [°C]	duration [min]
step 1	25	10
step 2	37	120
step 3	85	5
step 4	4	∞

4.5.3 Quantitative PCR

The synthesized cDNA was diluted 1:2 with nuclease-free water and 2 µl of each sample were pipetted in triplicates on a 384-well reaction plate. The relative mRNA expression of total *ATP2B4* (further referred to as *ATP2B4_compl*) and the two known transcript variants (further referred to as *AP2B4_common* (NM_001684) and *ATP2B4_rare* (NM_001001396)) from the Reference Sequence collection (RefSeq, NCBI) were determined. As a housekeeping gene for expression analysis in whole blood, the peptidylprolyl isomerase B (PPIB) was chosen according to literature

(Pachot et al. 2004). The quantitative PCR was performed using TaqMan® Gene-Expression Assays for the respective genes (Table 19).

Table 19: TaqMan® assays used for mRNA quantification of selected genes.

	RefSeq mRNA	#
PPIB	NM_000942	Hs00168719_m1
ATP2B4	NM_001001396	Hs00608066_m1
ATP2B4	NM_001684	Hs01067999_m1
ATP2B4	NM_001684	Hs01067999_m1
ATP2B4	NM_001001396	Hs01061213_m1

All reactions were performed in triplicates following the manufacturer's suggestions regarding the composition of the reaction mix (Table 20) and the settings of the thermocycler (Table 21).

Table 20: Composition of the PCR mix for TaqMan® Gene-Expression Assays.

component	volume per reaction [μl]	final concentration
TaqMan® Universal MasterMix II, no UNG (2x)	5	1x
TaqMan® Gene-Expression Assay (20x)	0,5	1x
template cDNA	2	
water (nuclease-free)	2,5	
sum	10	

PCR reactions were performed using the Light Cycler 480. During the reaction, the fluorescence signal was recorded using a wavelength of 465 nm for excitation and 510 nm for detection. The C_t -values were determined by the analysis software of the RT-PCR cycler.

Table 21: Cycling conditions for TaqMan® Gene-Expression assays.

	temperature [°C]	duration
initialization	95	10 min
denaturation	95	15 sec
annealing & elongation	60	1 min
storage	37	∞

45x {

4.5.4 Statistics

Statistical analyses of the expression assays were performed using the software R. For each of the 200 donor samples the mean and the standard deviation of the C_t -values of the triplicates were calculated. Samples with a standard deviation over 0.3 were

excluded from further analyses. Only in case of the assay that assesses the expression of the rare splice variant of *ATP2B4* a different threshold was defined. Due to the low concentration, the fluorescence signal was detected at relatively high C_t -values. This causes smaller differences in the amount of template transcript to result in bigger differences in recorded C_t -values. After reviewing the distribution of the standard deviations of this assay, a cut off of 1.5 was chosen. Subtracting the corresponding values of the expression of the housekeeping gene PPIB was performed to normalize the expression levels of the different transcripts.

For statistical testing, samples were divided into groups referring to the three different rs10900585 genotypes of the donors (GG, GT, TT). The Shapiro-Wilk test was applied to check whether the values of relative gene expression in all three groups are normally distributed. Following the results of this test, a Kruskal-Wallis test was applied to compare the median values of relative expression of the different transcripts in the designated groups. The null hypothesis that medians of the groups do not differ was rejected when p -value < 0.05 .

4.6 Parasite-proliferation assay

4.6.1 Samples

From the cohort of 200 Ghanaian donors, 30 were chosen for the parasite-proliferation assay. Donors with blood groups other than O^+ and carriers of a protective HbS allele were excluded from this experiment. For the choice of the final group of 30 donors, the genotype of rs10900585 was defined as the index marker. For each of the three genotypes (GG, GT, TT), 10 donors that fit all respective criteria were randomly selected from the cohort. One extra donor with a heterozygous genotype was chosen as a negative control. From each donor, 6,5 ml venous blood were drawn and anticoagulated with CPDA-1. The respective genotypes were again determined and samples were only included when the original genotypes were confirmed. The blood and all separated parts of it that were used in the assay, were stored for a maximum of three weeks at 4-8 °C until application.

The complete process of purification and washing of erythrocytes was performed under sterile conditions. Of each sample, 3 ml blood were diluted in 7 ml PBS and applied to an individual Plasmodipur filter to separate the erythrocytes from the leucocytes. The resulting pellet of red blood cells was washed two times with PBS. Of the packed RBCs, 600 μ l were diluted in 15 ml of the medium that was subsequently used for parasite

culture (RPMI 1640 containing 0.5% Albumax). Thereby, an approximate hematocrit of 4% was obtained. The exact concentration of erythrocytes was determined, using the flow cytometer Accuri C6 (BD). Thresholds on forward scatter (FSC) and side scatter (SSC) were set to discriminate erythrocytes from debris. The number of erythrocytes in 10 μ l of cell suspension was recorded. The lowest detected count was chosen as the target-erythrocyte concentration for dilution of the other suspensions with culture medium. The adjusted erythrocyte suspensions were prepared on day -1 and kept at 4-8 °C overnight.

4.6.2 Parasite culture

An inoculation culture of the parasite cell line 3D7 was cultivated in 0⁺-erythrocytes (University Clinic Eppendorf, Hamburg) and maintained until a preferably high parasitemia was reached (ideal: 5-10%).

On day 0, the parasitemia of the inoculation culture was determined by flow cytometry (4.6.3). The culture was diluted with culture medium until a 1:1 mixture with the diluted donor erythrocytes (4.6.1) resulted in a culture with an approximate hematocrit of 2% and a parasitemia of 0.1%. The volume *X* of original inoculation culture that was needed for 200 mL of this diluted suspension was calculated as followed:

$$X [mL] = \frac{200 \text{ mL} * 0.2\%}{\text{parasitemia of inoculation culture } [\%]}$$

With each donor-erythrocyte suspension, 10 mL of starting culture with a hematocrit of 2% and a parasitemia of 0.1% were prepared. A parasite culture with the same hematocrit and the same starting parasitemia like the donor cultures but with random 0⁺-erythrocytes (University clinic Eppendorf, Hamburg) served as positive control. A negative control was a purified, but not infected erythrocyte suspension from a heterozygous donor with a hematocrit of 2%.

The parasitemia of the starting cultures was verified by flow cytometry. Then, they were distributed into identical 96-well plates (Nunc-Microwell-plates, ThermoFisher Scientific). Therefore, 200 μ l of each of the 30 starting cultures and the two controls were pipetted in triplicates into the wells. For each determination of parasitemia, one plate was used. It was recorded every 24 h until the parasitemia in the positive control exceeded 10 %. This was defined as the end point measurement.

All plates were kept together at 37 °C under hypoxic conditions. Every 24 h, 100 µl of the medium from all wells in all plates were discarded and substituted with 100 µl of fresh medium. The plates remained at RT for the same amount of time during change of medium. Order of stacking was changed in a regular fashion over time of cultivation to avoid a potential bias.

4.6.3 Determination of Parasitemia

To determine the parasitemia of the 3D7-infected erythrocyte culture, the cells were stained with the two DNA-intercalating fluorescence dyes dihydroethidium (DHE) and HO33342. To measure the parasitemia of a culture, cells were incubated with dye in a 5-mL FACS tube. The reaction mix (Table 22) was incubated in the dark at RT for 20 minutes.

Table 22: Reaction mix for determination of parasitemia with flow cytometry.

component	volume per reaction [µl]	final concentration
culture medium (RPMI 164, 0.5 % Albumax)	68	
parasite culture (hematocrit 2 %)	30	
DHE in DMSO (0.5 mg/mL)	1	5 µg/mL
HO33342 in DMSO (0.45 mg/mL)	1	4.5 µg/mL
sum	102	

To inactivate the parasites, 400 µl of a 0.0003 % solution of glutaraldehyde in culture medium were added. All subsequent steps were not performed under sterile conditions. The amount of fluorescing cells was detected using the flow cytometer LSRII (BD). The blue laser (488 nm) with a 575-nm band pass filter was applied to stimulate and detect DHE. Cells containing HO33342 were detected using the UV laser (355 nm) combined with a 450-nm band pass filter. Thresholds for these two as well as FSC (390 V) and SSC (330 V) were adjusted to determine the amount of DNA-containing and ergo infected erythrocytes (parasitemia) in the culture. For each of the triplicates of the different cultures and controls 100,000 events were recorded.

4.6.4 Statistics

The software R was applied for the statistical analyses of the parasite-proliferation assay. For each sample and each day that the parasitemia was determined by flow cytometry, the mean of the triplicates was calculated.

The samples were separated into groups depending on the erythrocyte donor's genotype of the index SNP rs10900585 (GG, GT, TT). For each day, the median and the interquartile range (IQR) of the parasitemia in the three groups were calculated.

A potential genetic effect on the parasite proliferation can only be measured after the parasites have invaded the donor's erythrocytes. One proliferation cycle of *P. falciparum* during the blood stage takes approximately 48 hours. At that time point, all parasites should be within donor erythrocytes. Hence, the determined parasitemia at the end point of the cultivation was divided by the parasitemia after 48 hours to generate the normalized variable of proliferation rate, which was subjected to statistical analysis.

Due to the small groups, the median values of the proliferation rates were compared. Therefore, a Kruskal-Wallis test was applied. If the null hypothesis was rejected, Dunn's test was applied for pairwise comparison. The null hypotheses that the median values of the growth ratios do not differ was rejected, when $p\text{-value} < 0.05$.

5 Results

5.1 Computational fine mapping

A computational approach was chosen to fine map the resistance locus within *ATP2B4*. Genotype data from the previously published GWAS and from additional FRET-MCA assays were combined and formed the basis for an imputation of public data from the 1000 Genomes Project.

5.1.1 Quality control of basic data

The raw data of 102 SNPs, generated by FRET-MCA, was subjected to QC as previously described (4.1) and 48 SNPs were excluded due to the applied thresholds. Of these 48 SNPs, a total of 22 variants failed due to their low genotyping rates, eight because their MAF was too low and eleven because they did not reach either of these two thresholds. Two further SNPs were excluded because of their low HWE. Five SNPs were excluded because they were binary coded and their MAF was over 45 %. For these, a clear assignment of alleles during recoding was not possible.

During merging of the data, 35 SNPs on chromosome 1 were identified that were contained in both datasets. After merging the filtered FRET-MCA data with the original array data with the two different modes PLINK offers, the dataset contained FRET-MCA genotyped data for either 54 (keepFRET-MCA) or 19 (keepArray) SNPs. The merged datasets consisted of a total of 63,390 genotyped SNPs on chromosome 1.

5.1.2 Imputation and QC of imputed data

Genotype data of additional variants was imputed into the two basis sets (keepFRET-MCA and keepArray) using data from the 1000 Genomes Project (Phase 3 v5; Auton et al. 2015) from either all individuals (ALL) or only African individuals (AFR). The SNPs were afterwards filtered as previously described (4.1).

Imputing genotype data from 2,504 individuals of diverse ancestry (ALL) into the two basis datasets of the complete chromosomes led to an approximate total number of 6.5 Million genotyped SNPs on chromosome 1. Roughly 80 % of these SNPs did not pass QC. In each of the two datasets, approximately 1.3 Million SNPs on chromosome 1 remained.

As previously described, the imputation of genotype data from 174 individuals of African ancestry (AFR) was only performed for a 5-Mb segment containing the

ATP2B4 locus on chromosome 1. After imputation, the datasets contained around 50,000 genotyped SNPs of which slightly more than 40% did not pass QC. The filtered datasets for analysis contained around 29,000 SNPs.

5.1.3 Association testing

The filtered data was subjected to statistical analysis by logistic regression. The calculations were corrected for age, sex and population structure. Figure 11 shows the negative decadic logarithm of the p-values from logistic regression of the genotyped SNPs from the keepFRET-MCA dataset. The region of interest that was defined by a pairwise LD of > 0.2 with the index SNP rs10900585 (4.1.2) is highlighted in blue.

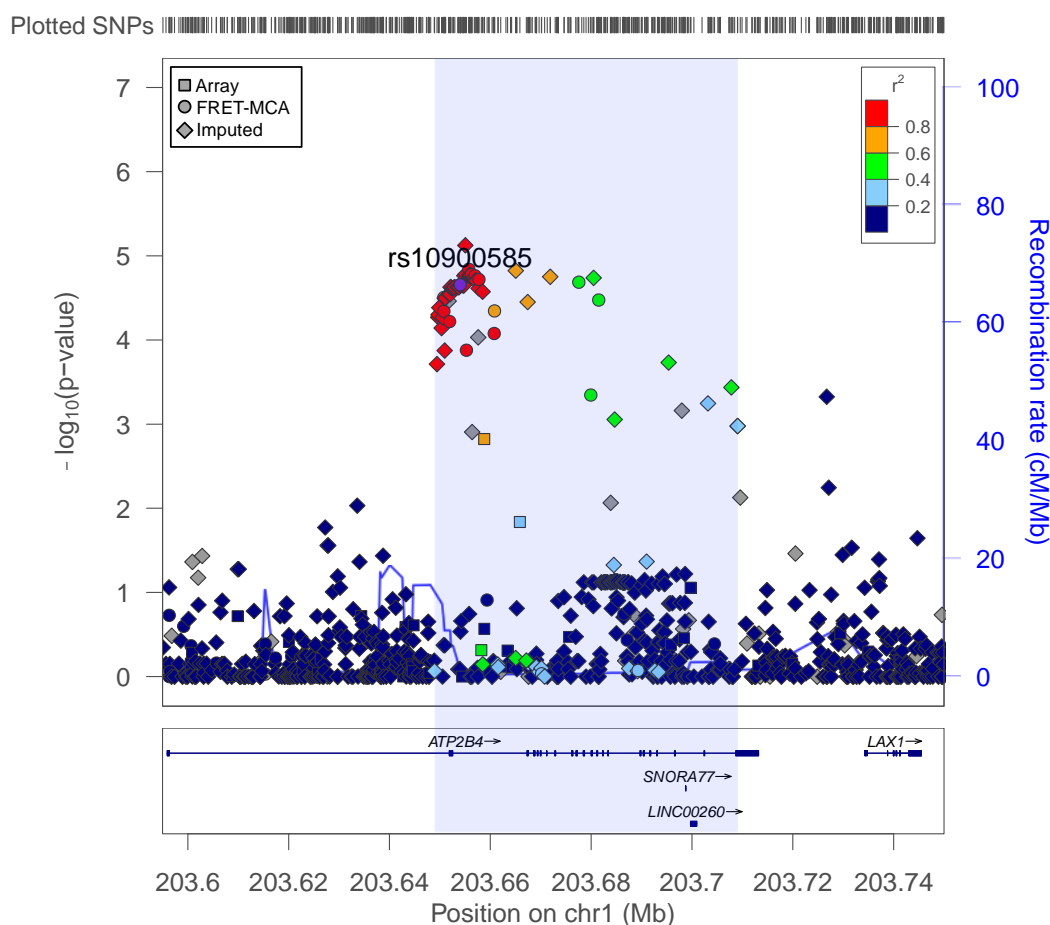


Figure 11: Regional association plot of the resistance locus generated from the dataset keepFRET-MCA_AFR. The origin of genotype data is encoded in the shape of the data point and the LD with the index SNP rs10900585 in its color.

An accumulation of SNPs that were associated with the trait of severe malaria with p-values ranging around 10^{-5} can be observed in the genetic area around the second exon of *ATP2B4*. The SNPs belonging to this accumulation are all within the region of interest defined by high LD with the index SNP (indicated by color-coded r^2). Two peaks of recombination rates border the chromosomal area containing the associated SNPs

(p-value < 10^{-4}). The corresponding data from the other datasets (keepFRET-MCA_ALL, keepArray_AFR, keepArray_ALL) shows no noticeable differences (Figure A - 1, Figure A - 2, Figure A - 3).

In total, 68 SNPs from the chromosomal region of *ATP2B4* were associated with the trait of severe malaria with a p-value < 10^{-4} in at least one of the four datasets (Table A - 5, Table A - 6). Of these SNPs, 56 were associated with a p-value smaller than 10^{-4} in all four datasets. In total, only three SNPs show an association with a p-value < 10^{-5} . All 68 associated SNPs show odds ratios (ORs) ranging from 0.54 to 0.60 and 59 of them are located within intron 1-2, exon 2 and intron 2-3 of *ATP2B4*. The remaining nine SNPs map in the rear part of the gene. All exonic SNPs are either located in untranslated regions or are synonymous SNPs.

The determined ORs of the five resistance SNPs as well as the corresponding p-values roughly match the values that were previously published (Table A - 4).

5.1.4 LD analysis

The software HapMap was applied to calculate the pairwise linkage disequilibrium (r^2) of the five published, clustered and most significantly associated resistance SNPs as well as the LD of all SNPs that were associated with resistance to severe malaria showing a p-value < 10^{-4} in at least one of the four datasets.

For the five resistance SNPs, calculations on the basis of genotype data from FRET-based genotyping showed roughly the same strength of LD as calculations using the GWAS data (Figure A - 4). The three SNPs (rs2365860, ts10900589, rs2365858) in the middle of these five SNPs are in almost complete linkage with each other. With the flanking two SNPs (rs10900585, rs4951074) occasional recombination takes place. Statistical analyses by logistic regression of all present haplotypes (MAF > 0.01) of the five SNPs did not result in any stronger associations or effect sizes than previous regressions of the single SNPs (Table A - 7).

The linkage disequilibrium of all SNPs with a p-value < 10^{-4} (Table A - 5, Table A - 6) was analyzed (Figure A - 5). For this, the dataset keepArray_AFR was used and calculations with HapMap were performed with a threshold of $r^2 > 0.8$ for tagging SNPs. The variants are distributed into four blocks of which all have been further investigated by experimental genotyping of at least one tagging SNP. Therefore, no experimental genotyping of further variants was performed in this study.

5.2 4C-Seq

The 4C-Seq method was applied to examine whether the resistance-associated locus within *ATP2B4* interacts with other genomic regions on a chromatin level. As previously described, the analyzed cell lines were chosen to reflect the most prominent cells included in the pathophysiology of severe malaria. The data was analyzed by applying the two analysis pipelines FourCSeq (Klein et al. 2015b) and Basic4Cseq (Walter et al. 2014).

5.2.1 Chromatin-chromatin interactions of the resistance locus

The associated resistance locus was chosen to be the viewpoint for the design of the first batch of experiments. Data was first analyzed using FourCSeq (Figure 12).

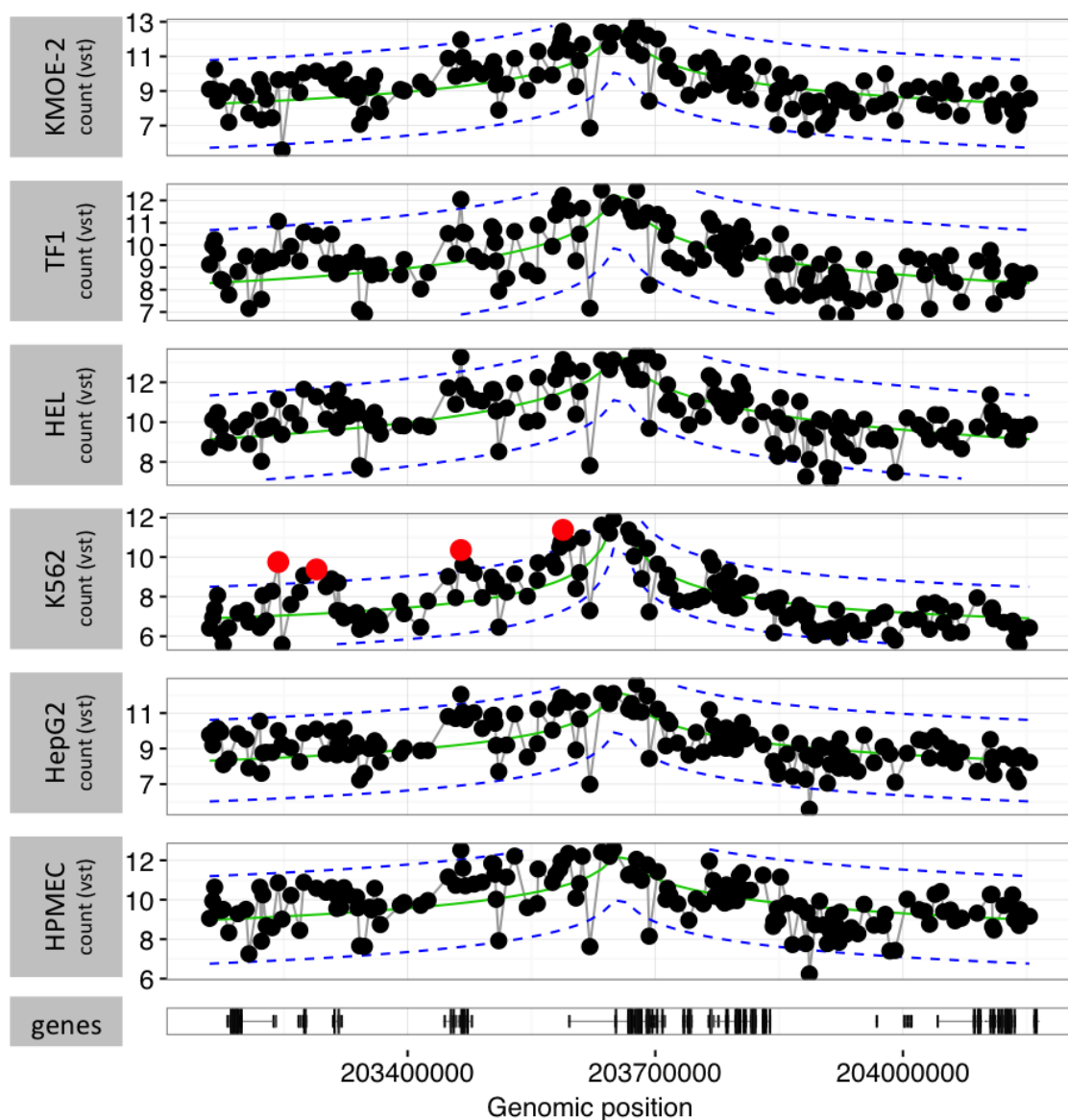


Figure 12: FourCSeq analysis of sequenced chromosomal fragments interacting with VP_Locus in fixed nuclei of six selected cell lines.

Data analysis with FourCSeq identified four fragments on chromosome 1 that were detected in a significantly increased number in the K562 library. In the data-visualizing output of FourCSeq these elevated values of the normalized count are marked in red (Figure 12, thresholds: FDR (false discovery rate) <math><0.01</math>, ATP2B4. In the remaining five libraries no significant hits were identified.

In Figure 13 the normalized fragment read count (reads per million [RPM]) of aligned fragments from all six libraries is visualized using Basic4Cseq.

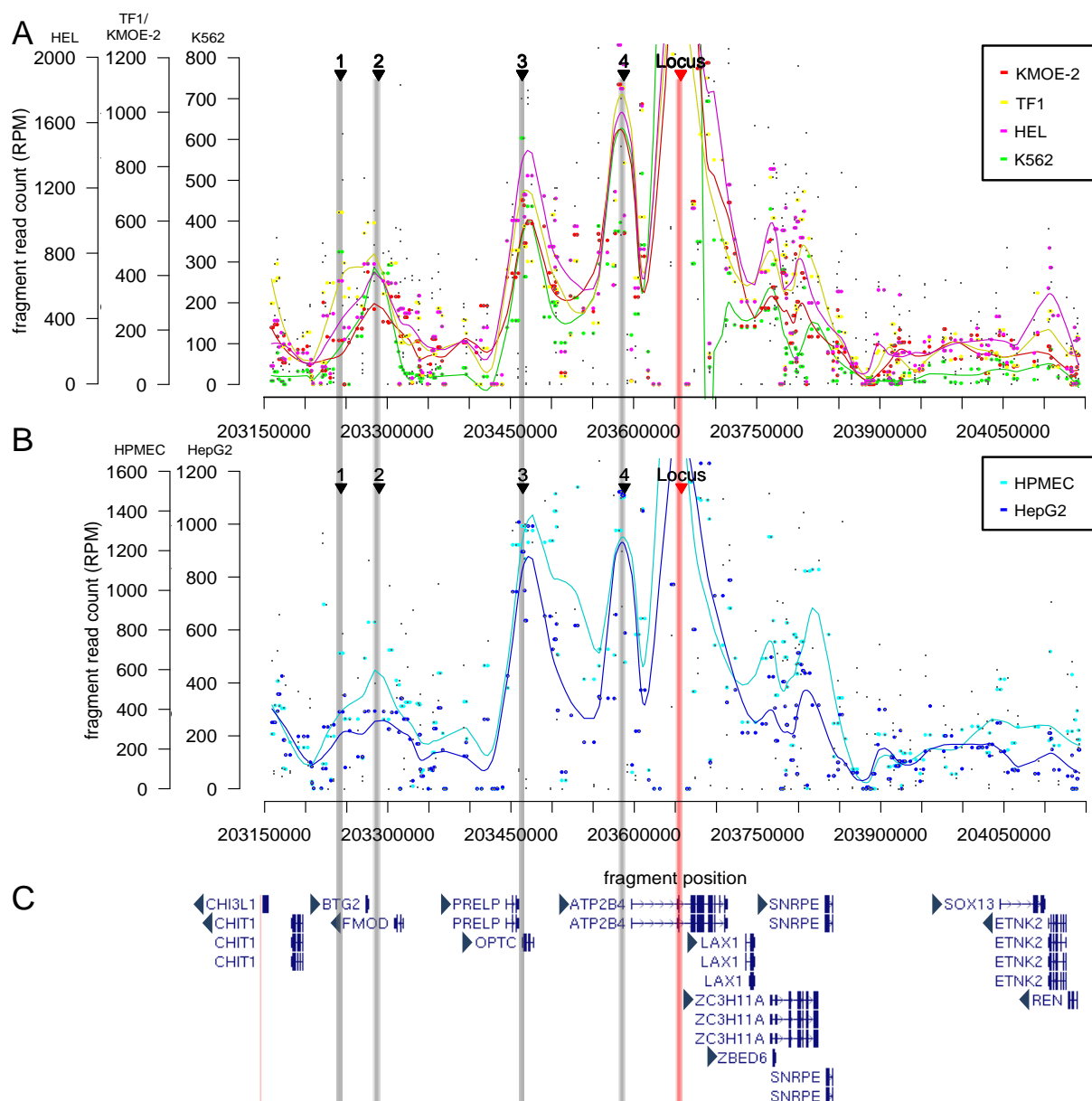


Figure 13: Basic4Cseq analysis of sequenced chromosomal fragments interacting with VP_Locus in fixed nuclei of six selected cell lines.

The figure shows (A) results from four erythroid cell lines, (B) results from HPMEC and HepG2 and (C) genes that are located in the affected region of chromosome 1 (image created on <http://genome.ucsc.edu>). The red line indicates the chromosomal position of VP_Locus. The four interactions that were detected in K562 cells by FourCSeq analysis are indicated by gray lines and numbered 1 to 4 according to their chromosomal position.

The RPM level of a peak depends on the 3-dimensional chromatin structure, but also on methodological factors like the accessibility of the nuclear DNA during fixation. These cell-specific characteristics generate the different scales of the y-axis. However, they do not affect the interpretation of the data.

Figure 13 shows that, although the four spots of elevated read counts were only detected as being statistically significant in the K562 library, the number of the respective fragments was also elevated in all other examined libraries. Apart from that, it can be observed that the general pattern of chromosomal interactions of the *ATP2B4* locus is similar in the various cell lines. Data exploration by Basic4CSeq did not show any further prominent read counts on chromosome 1 (data not shown).

The exact borders of the interacting fragments are defined by the localization of BsrGI-restriction sites (Table 23). The first interacting fragment maps to an intergenic region and lies upstream of two genes: *CHIT1* in 5' direction and *BTG2* in 3' direction. Hit 2 is located downstream of *BTG2* as well as downstream of the coding region of the gene *FMOD*.

Table 23: Chromosomal fragments interacting with the resistance locus (VP_Locus).

interacting fragment	start of fragment	end of fragment	size [bp]
Hit 1	chr1:203,239,154	chr1:203,247,278	8,125
Hit 2	chr1:203,278,771	chr1:203,300,487	21,717
Hit 3	chr1:203,462,509	chr1:203,466,461	3,953
Hit 4	chr1:203,587,124	chr1:203,589,312	2,189

The third fragment with significantly elevated read counts covers the immediate upstream region of the *OPTC* gene and parts of its coding region including the first three exons. This genomic area is also located downstream of the *PRELP* gene. The fourth fragment, interacting with the associated intronic region of *ATP2B4* maps to the genomic region directly upstream of the gene itself.

5.2.2 Reciprocal experiments

A reciprocal experimental design of all four hits was applied to replicate and validate the chromatin interactions of the resistance locus, which were identified by analyzing the first batch of libraries. For this, new primers were designed to hybridize to the detected fragments, making these hits the viewpoints of the respective reciprocal experiments. Then, template digested with the same restriction enzymes as before was used for library preparation.

When analyzing the four resulting sample sets with FourCSeq, all interactions of the locus fragment with the four hits were confirmed in K562 (Figure A - 6). Among other fragments with significantly increased fragment counts, the ones referring to the resistance locus showed significantly elevated values in all four reciprocal K562 experiments (indicated in Figure A - 6 by red arrows). In the sequencing libraries from the other five cell lines the fragment containing the resistance locus was not detected in a significantly increased number.

Like previously, the data from all reciprocal experiments was explored using Basic4CSeq. The results for the reciprocal 4C-Seq experiment of Hit 4 are shown in Figure 14.

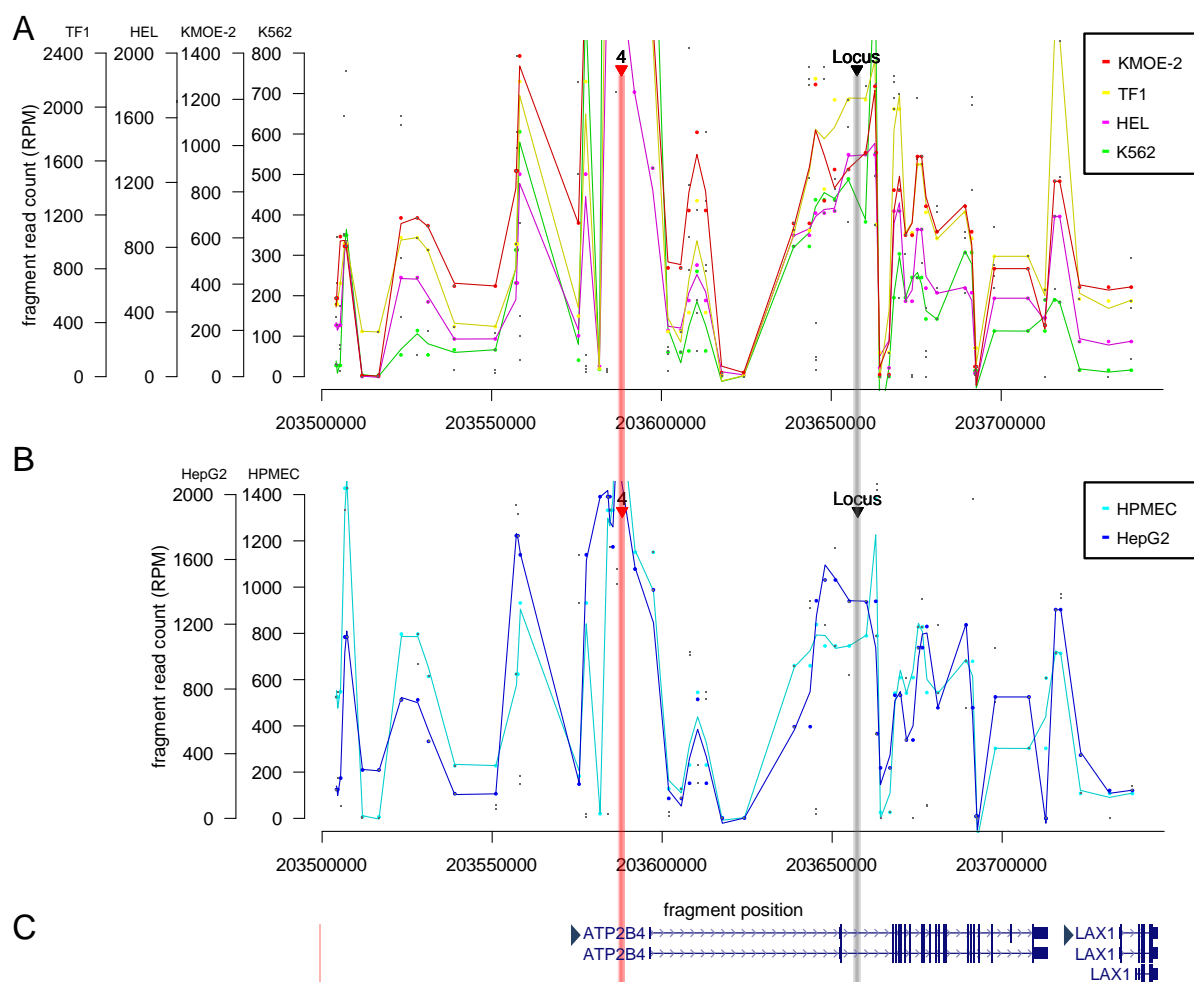


Figure 14: Basic4Cseq analysis of sequenced chromosomal fragments interacting with VP_Hit4 in fixed nuclei of six selected cell lines.

The figure shows (A) results from four erythroid cell lines, (B) results from HPMEC and HepG2 and (C) genes that are located in the affected region of chromosome 1 (image created on <http://genome.ucsc.edu>). The red line indicates the chromosomal position of VP_Hit4. The chromosomal position of the fragment that contains the resistance locus is indicated by the gray line.

The peak of fragment read counts was aligned around the respective viewpoint region upstream of *ATP2B4*. In the six cell lines a local maximum of fragment counts can be observed in the locus region. Therefore, the interaction between the resistance locus and the upstream region of *ATP2B4* was confirmed when applying an experimental design that is reciprocal to the initial experiment.

The reciprocal experiments for Hit 1, Hit 2 and Hit 3 showed similar results (Figure A - 7, Figure A - 8, Figure A - 9). The chromosomal region surrounding the fragment of the resistance locus showed elevated read counts in all examined cell lines. The peak of the read counts was not always exactly at the locus fragment, but occasionally shifted to the fragment upstream or downstream of it.

5.3 Genotyping of study group for functional experiments

For the selection of donors for the parasite-proliferation assay and statistical analysis of the functional experiments, the genotypes of the five published resistance SNPs and the SNP rs334, known to determine the sickle-cell allele, were analyzed in samples from 200 Ghanaian donors (Table 24).

Table 24: Genotype frequencies of the resistance SNPs and rs334 among the 200 Ghanaian donors of the study group.
* (ensembl.org; 23.03.17; 11:30)

resistance SNP	genotype		frequency in study group	frequency in African individuals from phase 3 of the 1000 Genomes Project *
	(hom. major allele)	(het)		
rs10900585	(hom. minor allele)	TT	0.340	0.336
		TG	0.425	0.487
		GG	0.230	0.176
rs2365860		AA	0.365	0.397
		AC	0.450	0.466
		CC	0.185	0.137
rs10900589		TT	0.365	0.397
		TA	0.450	0.466
		AA	0.185	0.137
rs2365858		CC	0.365	0.397
		CG	0.450	0.466
		GG	0.185	0.137
rs4951074		GG	0.430	0.397
		GA	0.410	0.466
		AA	0.160	0.137
rs334 HbS/C		AA	0.736	0.810
		AS	0.159	0.180
		SS	0.005	0.010
		CC	0.005	-
		AC	0.095	-
	CS	0.000	-	

The frequencies of the respective genotypes did not differ significantly from the frequencies calculated from data of African individuals included in the 1000 Genomes Project. In case of the C allele of rs334, no information was available from the Ensembl genome browser 88 (Yates et al. 2016) as it only provides biallelic information for each variable.

5.4 Expression of *ATP2B4* in Ghanaian whole blood

In whole blood samples from 200 Ghanaian donors, the expression levels of mRNA of total *ATP2B4* (*ATP2B4_compl*) and the two known splice variants (RefSeq, *AP2B4_common* [NM_001684], *ATP2B4_rare* [NM_001001396]) were determined by applying TaqMan® Gene-Expression Assays (4.5). This was done to assess whether the associated resistance SNPs might affect *ATP2B4* expression in a quantitative manner but also to evaluate whether they have any impact on the ratio of the two known splice variants. For statistical analysis the 200 donors were separated into groups corresponding to their genotypes of the index SNP rs10900585.

The relative expression of the various *ATP2B4* mRNAs in these groups is shown in Figure 15. Testing for normal distribution by applying a Shapiro-Wilk test showed that the null hypothesis of normal distribution is not commonly accepted for all three genotype groups in any of the applied assays (Table A - 8). Therefore, median values were compared.

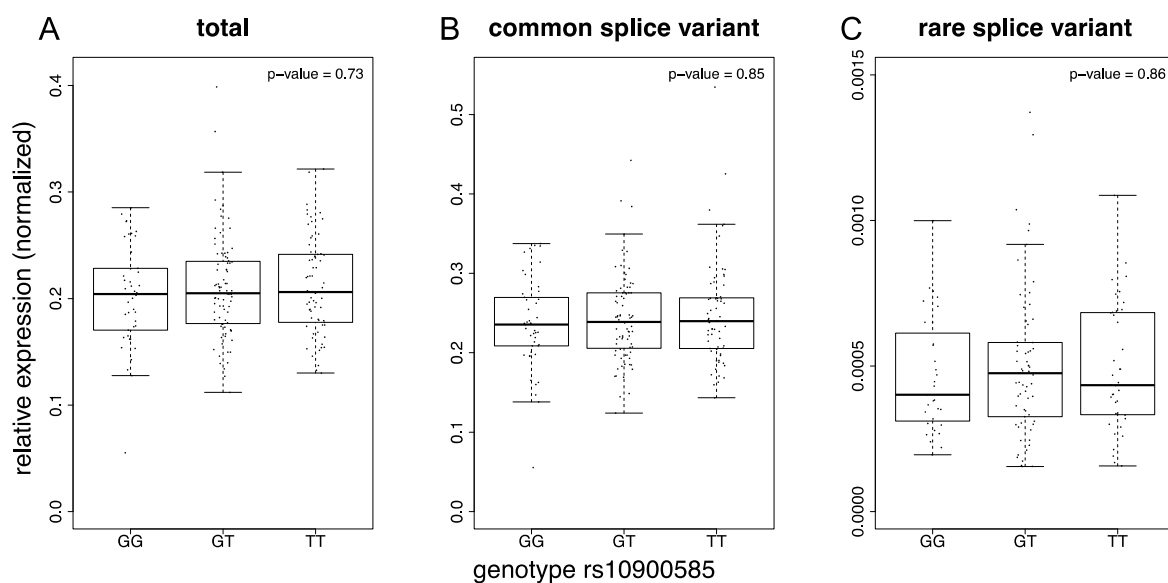


Figure 15: Relative expression of (A) total *ATP2B4*, (B) its common splice variant and (C) the rare splice variant in samples of the Ghanaian study group.

Data points show values of single samples. The respective boxplots visualize the 25th percentile (bottom of the box), the median (band within the box) and the 75th percentile (top of the box). The length of the whiskers is determined by the last value within the respective 1.5-fold interquartile range. The p-values refer to the performed Kruskal-Wallis test.

A statistical comparison by applying the Kruskal-Wallis test showed that the differences in total mRNA expression of *ATP2B4* were not significant (p-value = 0.73). The examination of the expression of the two known splice variants of *ATP2B4* showed the same tendencies as the total expression of the gene. In both cases the statistical comparison of the medians of relative expression did not indicate a significant difference (p-value_{ATP2B4_common} = 0.85, p-value_{ATP2B4_rare} = 0.86). The data for each expression assay is further summarized in Table A - 9.

Furthermore, the expression ratio of the respective splice variants and the total expression of *ATP2B4* were compared with respect to the genotype groups. The Shapiro-Wilk Test showed that the values in at least one of the genotype groups of each assay were not normally distributed (Table A - 10). Accordingly, the Kruskal-Wallis test was applied for subsequent statistical analysis.

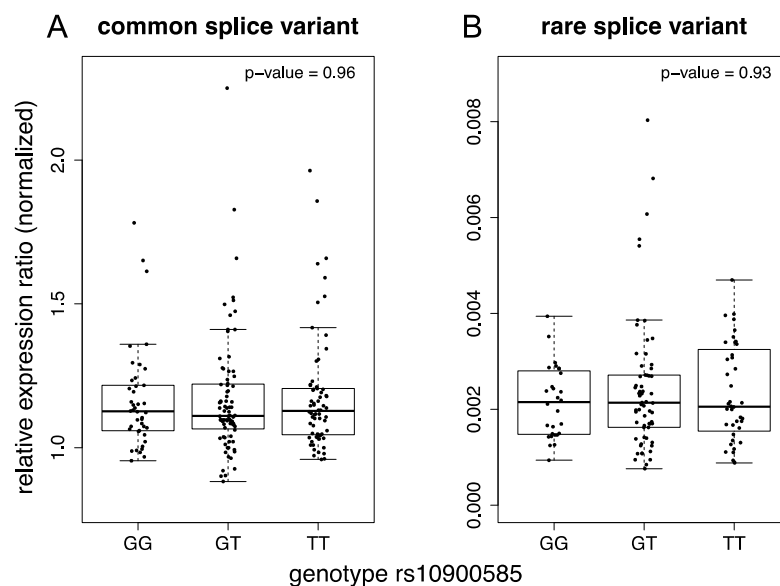


Figure 16: Relative expression ratios of (A) the common and (B) the rare splice variant in samples of the Ghanaian study group with respect to total *ATP2B4* expression. Data points show values of single samples. The respective boxplots visualize the 25th percentile (bottom of the box), the median (band within the box) and the 75th percentile (top of the box). The length of the whiskers is determined by the last value within the respective 1.5-fold interquartile range. The p-values refer to the performed Kruskal-Wallis test.

The median values of the ratios of relative expression of both splice variants and total expression of *ATP2B4* did not differ significantly between the three groups (Figure 16). The summary of the data is shown in Table A - 11.

To examine the relationship between the relative expression of the splice variants and the total expression of the gene, a linear regression analysis was performed (Figure A - 10). Applying a linear model resulted in an adjusted r^2 -value of 0.97 for the common

variant. The r^2 -value for applying the linear model on the expression levels of the rare variant was 0.83.

5.5 Parasite proliferation referring to the rs10900585 genotype

A parasite-proliferation assay (4.6) was performed to examine whether the statistical association of the resistance locus within *ATP2B4* relates to a cellular process that influences the development of the parasite during the blood stage of its life cycle.

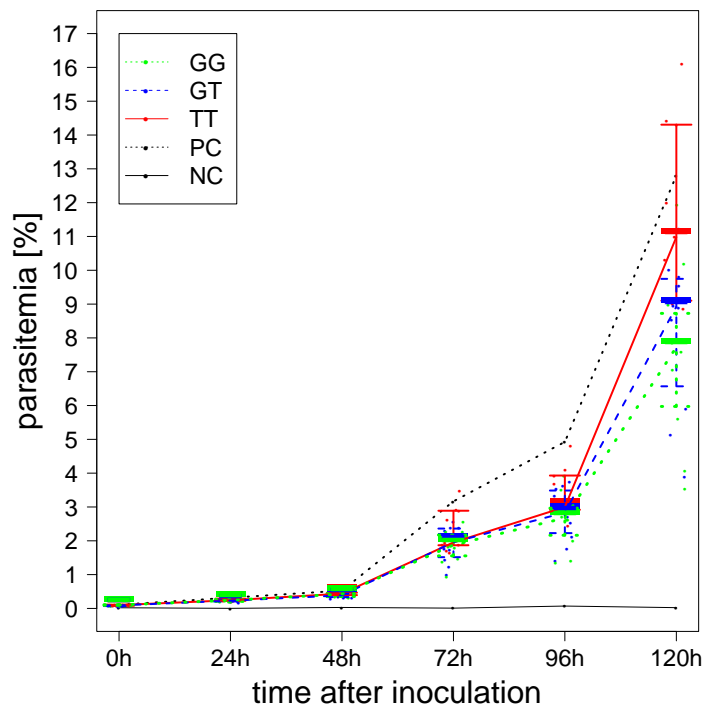


Figure 17: Parasitemia in infected erythrocytes of donors with different rs10900585 genotypes.

The graphs show the development of the median values of the detected parasitemia in the three genotype groups. The values of different groups are color-coded. Error bars correspond to the respective interquartile ranges. Values of samples are shown as color-coded dots. The graphs of the positive control (PC) and the negative control (NC) refer to the mean value of the triplicates at respective time points.

After 120 h of incubation, the median values of parasitemia in cultures of RBCs from the three genotype groups showed substantial differences (Figure 17). The highest parasitemia was detected in erythrocytes from carriers of the homozygous major allele (TT). The corresponding median value was 10.98 %. The lowest parasitemia after 120 h of cultivation was reached in erythrocytes from the group of donors with the GG genotype (median: 7.73 %). With a median value of 8.93 % the cultures with erythrocytes from heterozygous donors showed an intermediate parasitemia.

The parasite culture with random 0⁺-erythrocytes, functioning as positive control, showed a parasitemia of 12.80 % after 120 h of cultivation. Thereby the pre-defined end point was reached. The negative control showed no parasites throughout the complete time of cultivation. The difference between the parasitemia in erythrocytes from donors with different rs10900585 genotypes increased with an increasing parasitemia in the culture over time.

One of the ten samples that were selected for the TT group was later genotyped as GT and therefore excluded from all analyses. For the 29 remaining samples the rs10900585 genotype was confirmed. As previously described (4.6.4), the values of the detected parasitemia of the end point measurement were divided by the respective values of the parasitemia after 48 hours, since all parasites should have invaded donor RBCs after this time. The resulting proliferation rates are displayed in Figure 18.

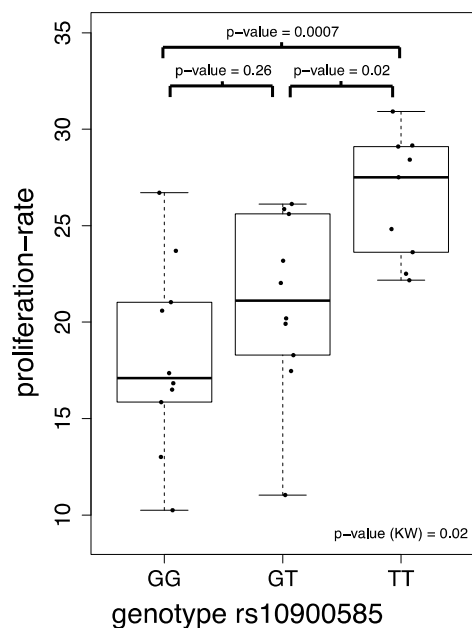


Figure 18: Parasite proliferation in erythrocytes of donors with different rs10900585 genotypes.

Data points show means of the triplicates of each culture from erythrocytes of the different donors. The boxplots visualize the 25th percentile (bottom of the box), the median (band within the box) and the 75th percentile (top of the box) of the values of each genotype group. The length of the whiskers is determined by the last value within the respective 1.5-fold interquartile range. The p-value (KW) refers to the performed Kruskal-Wallis test. All other p-values result from Dunn's test for pairwise comparison of the indicated groups.

During the time between 48 h and 120 h of cultivation, the parasites proliferated 27.5-fold within erythrocytes from donors with the TT genotype of rs10900585. This rate was the highest among the three groups. The lowest (17.1-fold) was detected

within erythrocytes from carriers of the homozygous GG allele. The cultures from heterozygous donors showed an intermediate value of 21.2-fold.

Statistical analysis by applying a Kruskal-Wallis test revealed that the median values of the proliferation rates differ significantly between the three genotype groups (p-value = $2,67 \times 10^{-3}$). The subsequent analysis with Dunn's test for pairwise comparisons showed a significant difference between the two groups of donors with the homozygous genotypes (GG vs. TT: p-value = $6,96 \times 10^{-4}$). Furthermore, the difference between the TT group and the heterozygous group was significant (GT vs. TT: p-value = 0.02) while the third pairwise comparison (GG vs. GT) did not reach a level of significance (p-value = 0.26). All descriptive details of the experiment are summarized in Table A - 12.

The differences of the parasites' proliferation in erythrocytes from donors with homozygous genotypes of the selected index SNP were confirmed by the results of a replication experiment using cells from the same blood donors (Figure A - 11, Figure A - 12, Table A - 13). In contrast to the original experiment, the proliferation rate of parasites within erythrocytes from donors with a heterozygous genotype was similar to the one in RBCs from carriers of the homozygous major allele (TT). The end point measurement (parasitemia in positive control over 10%) was performed after 144 hours.

6 Discussion

6.1 Computational fine mapping

The previously published resistance locus for severe malaria within the gene *ATP2B4* was confirmed by the computational fine mapping that was performed for this study. All SNPs that were associated with severe malaria show odds ratios ranging from 0.54 to 0.60 and are therefore resistance-associated.

For the current study, a more recent and more extensive set of data from the 1000 Genomes Project was available for imputation into the datasets of the 2,153 individuals from the original study cohort (Timmann et al. 2012). Data from 661 humans of African origin was imputed, in contrast to 174 African individuals for the analysis performed in 2012. As previously shown, this increase in number also increases the quality of the imputed missing variant genotypes (Zheng et al. 2015). Altogether, a much higher number of variants passed QC during the current study and was applied for association testing.

In the first study, imputation of African genotypes from the 1000 Genomes Project and subsequent QC resulted in a total of 6,844 variants within the region surrounding the resistance locus (chr1:201,000,001-206,000,000). The respective imputation and QC for the current study generated 28,672 variants tested within the same chromosomal area. This increase was not only due to the improved quality of imputation but also to an increased number of variants that is covered by the phase 3 release of the 1000 Genomes Project. Furthermore, the MAF threshold for the variant QC before regression analysis was lowered in comparison to the previous study (from 0.10 to 0.01).

Although almost four times more SNPs were subjected to statistical analysis, it was not possible to identify a variant with a substantially stronger association (p-value) or effect size (OR) than the ones that were published. It is not known whether genotyping by FRET-MCA or by an array produces the more reliable results. Therefore, the two mentioned strategies for handling duplicates were chosen (4.1) and the resulting two respective data sets were subjected to imputation. In addition, there is an ongoing discussion, whether it is necessary to exclude data of all individuals from the 1000 Genomes Project that do not match the ethnicities of the study group, before imputation (Howie et al. 2011). It was shown, that sample size improves the quality of

imputation rather than ethnic homogeneity does, but whether this holds true has not been shown for the latest release of the 1000 Genomes Project containing data from a large number of individuals of African ancestry. However, the two potential sources of error were avoided by generating and analyzing the four different datasets by logistic regression without observing substantial discrepancies (5.1.3).

The imputation of the latest data from the 1000 Genomes Project did not reveal any new hints for identifying a potential causal variant that might have been undetected until now. Previous studies showed that this is technically possible, for example when the imputation of a dataset derived from an Africa population succeeded in the re-detection of the sickle cell allele with much higher level of significance compared to the locus-tagging SNPs of that GWAS on severe malaria (Jallow et al. 2009).

One possible explanation would be that the causal variant, or a causal combination of variants is already among the resistance-associated SNPs, but does not stand out with a substantially higher p-value or OR. This theory is supported by the massive number of variants, including rare variants, analyzed in this study. Previous studies showed, that calculating a statistical association of a rare variant is possible by imputing its genotypes (Magi et al. 2012). The argument that potentially associated variants might specifically occur in the Ghanaian population and have not been imputed due to the absence of Ghanaian genotypes from the imputed dataset can be weakened by the fact that the locus within *ATP2B4* has been replicated in numerous African populations, including ones from the phase 5 release of the 1000 Genomes Project (Malaria Genomic Epidemiology Network 2014). Nevertheless, genetic heterogeneity between African populations cannot be ruled out to affect the quality of imputation of variants surrounding the resistance-associated locus. In addition, there are other analyses that show that the imputation of the majority of SNPs with a MAF < 0.03 is of poor quality (Zheng et al. 2015). Thus, even if a rare variant causing resistance to severe malaria was contained within the 1000 Genomes data, it might not have been properly imputed.

If this were the case, and also if there were a causal variant not covered by the 1000 Genomes Project, an additional approach would be to re-sequence the locus region in DNA samples from some or all of the 2,153 individuals of the GWAS study population. A suitable method would be probe-based target enrichment with subsequent NGS. This would enable the discovery of potentially trait-associated rare variants that have not been detected by the methods previously applied. Attempts have been made by analyzing the genomic region of *ATP2B4* by high resolution melting (HRM), but no new

rare variants were discovered (unpublished data). Nevertheless, NGS is a much more sensitive method that could detect rare variants that were not found by HRM.

Taken together, there is no hint from the association signals within the *ATP2B4* locus whether the underlying mechanism inducing resistance to severe malaria is caused by (a) one or more of the already described common resistance SNPs with a weak effect, (b) by common, but unknown genetic variants in strong LD with these resistance SNPs or (c) by rare unknown genetic variants in LD having a strong effect on the phenotype.

In the published GWAS, it was shown that variants for which the logistic regression of genotype data from the study cohort (2,153 donors) generates a p-value $< 10^{-4}$, can reach genome wide significance (p-value $< 10^{-8}$) if the number of samples is extended with a replication group (Timmann et al. 2012). That is why, a p-value lower than 10^{-4} was defined to indicate an association with the trait in this study. Analyses revealed a list of over 60 SNPs that pass this threshold and are considered associated with resistance ($OR < 1$). The majority of these SNPs are located within the 3' region of intron 1-2, within exon 2 and within intron 2-3 (Table A - 5, Table A - 6). Public genome browsers show regulatory elements within this region in various cell lines (Figure 19). Sequence analyses in K562 cells predicted areas with potential promoter or enhancer functions. The same can be seen in other cell lines like HeLa-S3 and HUVEC. Nevertheless, these predictions remain to be very unspecific.

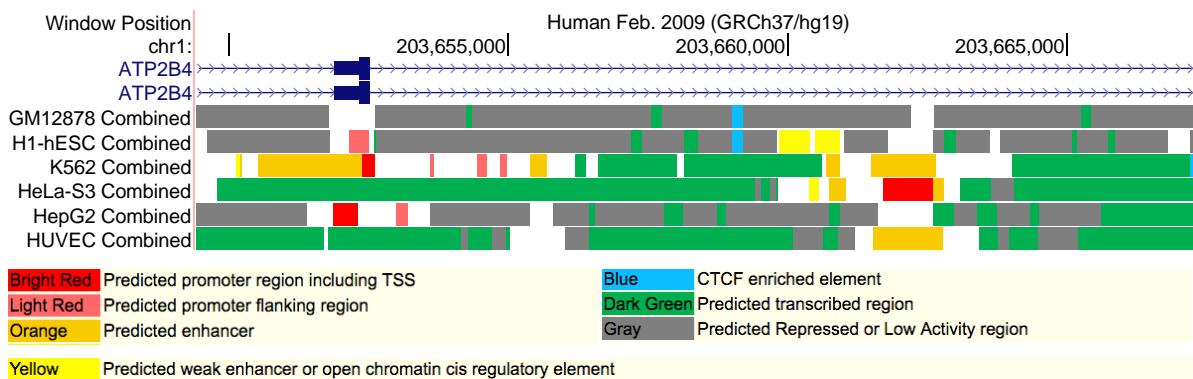


Figure 19: Regulatory elements within the locus region (3' region of intron 1-2, exon 2 intron 2-3 of *ATP2B4*). The respective types of elements are indicated by color coding. They were derived by application of a consensus merge of the segmentations produced by the ChromHMM and Segway software (image created on <http://genome.ucsc.edu>).

When examining gene regulation in the context of non-coding SNPs that are associated to a trait by GWAS, research concentrates on three cellular processes: transcription, splicing and translation (Fung et al. 2015). Regarding transcription, the quantity of generated mRNA of the affected gene is of interest. Over the last years expression quantitative trait loci (eQTL) have been a pivotal topic (Westra and Franke 2014;

Nica and Dermitzakis 2013). These eQTLs are genomic loci that contribute to variations of transcription-levels.

Westra et al. (2013) published an eQTL-meta study in which 26 out of the 60 SNPs here associated with resistance to malaria in the present study were associated with a quantitative change in expression of *ATP2B4* and *LAX1*. The latter is a gene that maps directly downstream of the *ATP2B4*-coding region. The minor alleles of the 26 associated variants that map to intron 1-2, exon 2 and intron 2-3 of *ATP2B4* were found to mediate an increase in *ATP2B4* expression and a decrease of *LAX1* expression. For the associated SNPs from the 3' region of *ATP2B4*, the data about the direction of the effect is inconsistent. The majority of all SNPs that influence the expression levels of either *ATP2B4* or *LAX1*, maps to the genomic region of *ATP2B4* and is located within intron 1-2, exon 2 and intron 2-3. This comprises exactly the resistance locus identified in 2012 and verified by this study.

In contrast to eQTLs, there is only very little genome-wide data published on variant specific differential splicing. A genome-wide study on gene expression in whole-blood samples of the Framingham Heart Study identified the strongest association of a SNP in the 5' region of *ATP2B4* (rs970) to be associated with variant-specific splicing of *ATP2B4* RNA (Zhang et al. 2015). Also, the index variant (rs10900585) was identified as an sQTL (splice quantitative trait locus). Both variants appear to cause differential expression on exon-specific, but not whole-transcript levels of *ATP2B4*. Therefore, results from this study are partly contradictory to the previously described eQTL study. The results from both previous studies (Westra et al. 2013; Zhang et al. 2015) have not been further analyzed. When consulting results from large transcriptome studies, it has to be kept in mind that these studies deal with huge sets of data. This requires intensive quality control and statistical analyses need to be corrected for multiple testing. Nevertheless, the results might always contain false positives. Therefore, they should never be used to directly explain cellular processes that were not specifically addressed by the study.

Another difficulty when using these studies for interpreting the resistance association of the *ATP2B4* locus results from the fact that the respective experiments were performed with samples from individuals of European ancestry. The regulation of genes differs substantially between populations from different ethnical backgrounds (Spielman et al. 2007). Therefore, results cannot be directly transferred, but have to be confirmed in the specific population.

In a very recent study, an association between a haplotype within *ATP2B4* and the amount of PMCA4 in erythrocyte membranes was found in blood samples from Hungarian donors (Zambo et al. 2017). This haplotype covers parts of introns 1-2, exon 2 and intron 2-3, just like the majority of the resistance-associated SNPs from this study with African donors. The probable tagging SNP rs1541252 that is used by Zambo et al. for association-calculation lies within exon 2. This exon mostly encodes the 5' UTR of the *ATP2B4*-mRNA. The study shows a gradual increase (additive MOI) of protein expression in erythrocyte membranes depending on the rs1541252 genotype of the donor. The least PMCA4 was detected in samples from homozygous carriers of the minor allele and the highest in samples from donors with the homozygous major allele. Furthermore, it was shown that the increased PMCA4 expression is accompanied by a more efficient extrusion of calcium ions from the cell. These findings are of great interest as the SNP rs1541252 was associated with resistance to malaria in this study (Table A - 5) and is in strong LD with the five resistance SNPs from the GWAS (Figure A - 5).

Considering the fact that the parasite employs PMCA4 from the RBC membrane to stabilize the Ca^{2+} -concentration within the parasitophorous vacuole (1.4), a decrease of PMCA4 expression might directly affect its development within the erythrocyte.

In summary, there is substantial evidence that the causal SNPs underlying the statistical association of the resistance locus within *ATP2B4* may cause variant dependent shifts in the regulation of the gene. However, all findings must be further examined in African samples as partly addressed by the functional experiments of this project.

6.2 Chromatin-chromatin interactions of the resistance locus

To examine the chromatin-chromatin interactions of the resistance locus within *ATP2B4*, 4C-Seq experiments were conducted with various cell lines. These analyses revealed four chromosomal regions interacting with the published resistance locus on a chromatin level. The interactions were found in all six examined cell lines and were verified by reciprocal 4C-Seq analyses. Recently, chromatin-chromatin interactions and their influence on gene regulation became a focus of research in human genetics (Bouwman and Laat 2015). Interactions between fragments within *ATP2B4* and the chromosomal regions around the fragments that were detected in this study (Table 23) were previously published for various cell lines in multiple studies of chromatin interactions (e.g. Javierre et al. 2016; Mifsud et al. 2015). Figure 20 shows chromatin interactions of the upstream region of *ATP2B4* and the resistance locus within

K562 cells generated from Hi-C experiment, an advancement of 4C (Wang et al. 2017). The squares that directly or approximately connect the resistance locus and the four hits (indicated by gray triangles), show high intensities compared to other squares connecting chromosomal regions of similar distance. These interactions have the potential to mediate a regulatory effect that could be affected by sequence variation within one of the interacting fragments (Smemo et al. 2014).

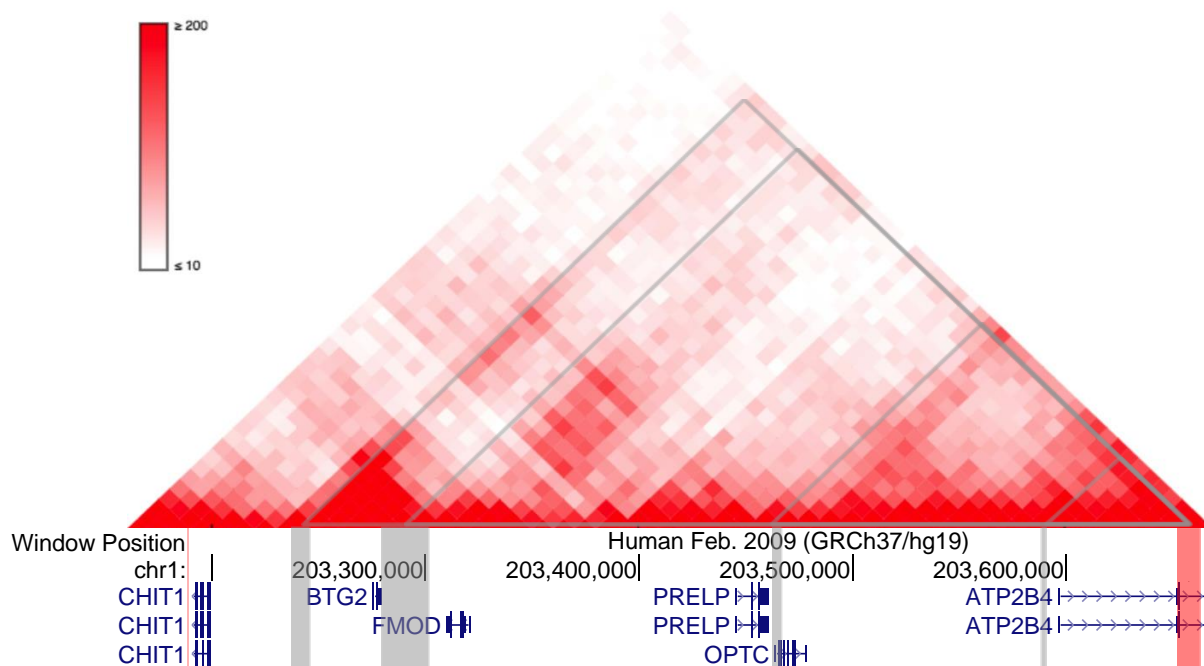


Figure 20: Hi-C data from an analysis of the resistance locus and its respective four 4C-Seq hits in K562 (Rao et al. 2014). Each of the red-shaded squares connects two 10-kb fragments of the selected region of chromosome 1. The darker the red square, the more often the two fragments interact with each other (image created on <http://promoter.bx.psu.edu/hi-c/>).

Assuming that the target fragment contains regulatory domains, an alteration of the chromatin-chromatin interaction could affect the expression of genes, which are located on or adjacent to these fragments. Sequence analyses of various cell lines of the Encode Project (ENCODE Project Consortium 2012) generated predictions about their chromosomal regulatory domains (Figure 21). The target fragments Hit 1 and Hit 2 of the chromatin interactions of the resistance locus within *ATP2B4* contain multiple regulatory segments. Not only predicted promoter (red/light red) and enhancer (orange/yellow) regions can be found within the respective sequences, but also sites of enriched binding of the transcription factor CTCF (blue). These are known to border and stabilize DNA loops into so-called topologically associating domains (TADs), which enable chromatin-chromatin interaction and thereby potential gene regulation (Bouwman and Laat 2015).

From the sequences of Hit 3 and Hit 4, no regulatory elements can be predicted except for a single weak enhancer at the beginning of the coding region of *OPTC* in human embryonic stem cells.

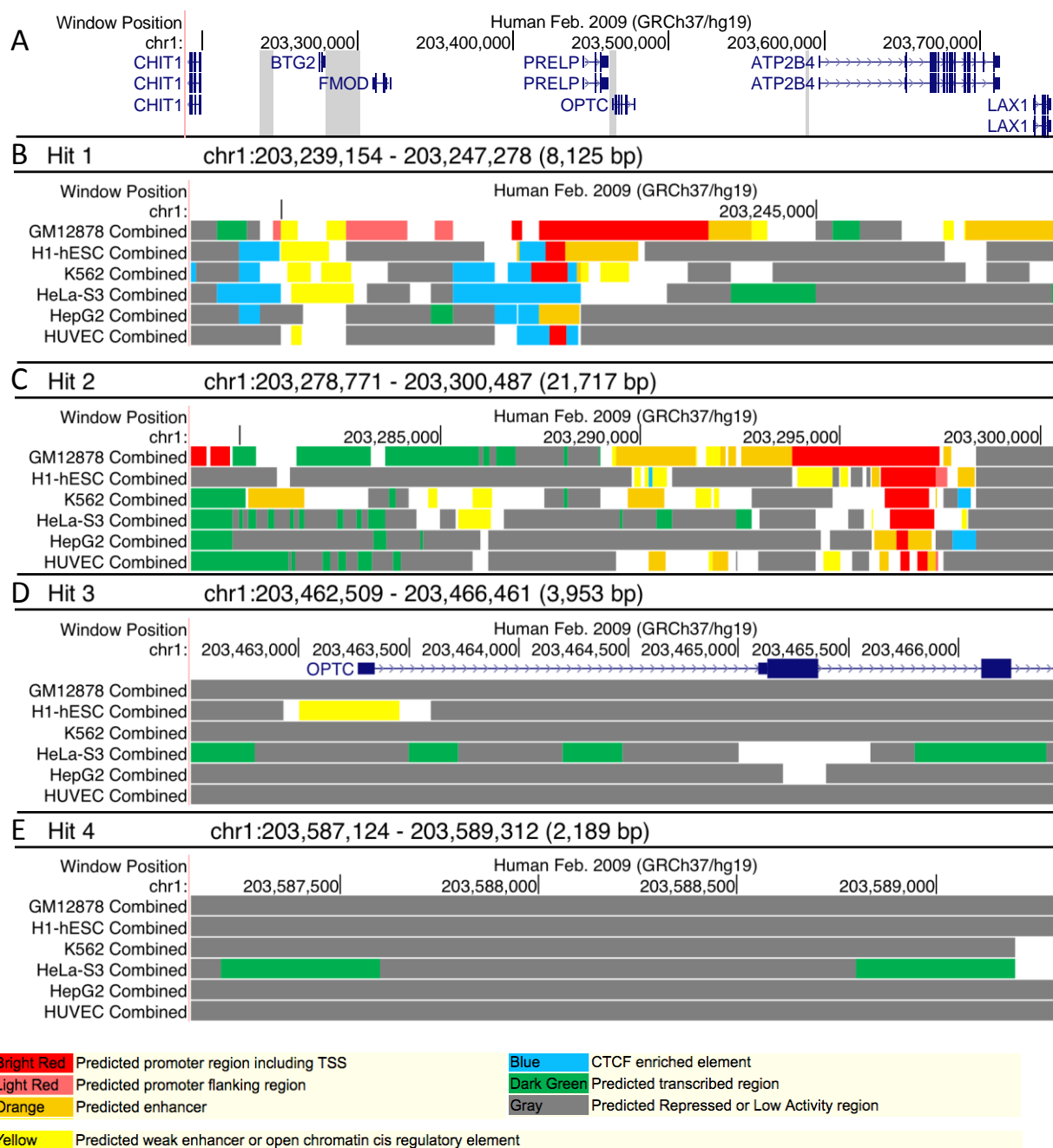


Figure 21: Regulatory elements within the fragments of the four 4C-Seq hits of the locus region.

The localization of the fragments is indicated by gray boxes (A). The regulatory elements within the fragments of (B) Hit1, (C) Hit2, (D) Hit3 and (E) Hit4 are shown in the figure. The respective types of elements are indicated by color-coding. They were derived by application of a consensus merge of the segmentations produced by the ChromHMM and Segway software (image created on <http://genome.ucsc.edu>).

Still, a potential influence of sequence variation cannot be excluded in the resistance-locus region that is interacting with the respective fragments. Since the predicted regulatory elements that are shown in Figure 21 are generated from sequence analysis and not from functional experiments it is still possible that regulatory elements

with unknown sequence patterns are contained. The fragments referring to Hit 3 and Hit 4 both map to chromosomal regions directly upstream of *OPTC* and *ATP2B4*, respectively. This is, where in most cases the promoter regions of genes are located.

The genes adjacent to the chromosomal regions that have been detected to interact with the published resistance locus are diverse in their functions and have been associated with different traits.

CHIT1

The genomic area upstream of Hit 1 encodes the protein chitinase 1, an enzyme that is able to break down glycosidic bonds of chitin. It is secreted by activated macrophages of vertebrates and seems to play a role in the degradation of chitin-containing pathogens like fungi (van Eijk et al. 2005). In the context of malaria resistance, it is noteworthy that serum levels of beta-thalassemia patients were found to be elevated (Barone et al. 1999). Hence, the resistance of these patients towards malaria may have other reasons than the abnormal structure of their hemoglobin. Also patients with atherosclerosis, a disease influencing the vascular endothelium that is important for the pathophysiology of malaria, have elevated levels of chitinase 1 (Artieda et al. 2003).

BTG2

The gene *BTG2* that is located between the fragments identified as Hit 1 and Hit 2 by 4C-Seq encodes the protein “B-cell translocation gene anti-proliferation factor 2”. It is a well studied tumor suppressor that is involved in all steps of the cell cycle (Mao et al. 2015). *BTG2* appears not to be involved in cellular processes relevant for the pathophysiology of malaria.

Small leucine-rich repeat proteoglycans (SLRPs)

The genes *FMOD*, *PRELP* and *OPTC* that map adjacent to Hit 2 and Hit 3 do all encode for proteins that belong to the same family. Their respective products fibromodulin, proline and arginine rich end leucine rich repeat protein and opticin are assigned to the group of “small leucine-rich repeat proteoglycans” (SLRPs). These are biologically active components of the extracellular matrix, which are able to interact with multiple molecules and induce inflammation (Merline et al. 2009).

In the past, sequence variants in SLRP genes have been associated to different traits. Interestingly, all of them induce ocular diseases (Schaefer and Iozzo 2008). Different loss-of-function variants within *FMOD*, *PRELP* and *OPTC* are, for example, associated

with high myopia (Majava et al. 2007). Fibromodulin expression has also been reported in carotid atherosclerotic plaques with subsequent cerebrovascular events (Shami et al. 2015).

In the context of resistance to severe malaria, these findings do not yield any obvious functional clues. Since SLRPs are involved in multiple signaling pathways it is not possible to exclude a potential protective impact on severe malaria that a shift in regulation of the three concerned genes could have. Nevertheless, it has been reported that certain SLRPs compensate each other if one of them is altered in its expression (Svensson et al. 1999). Therefore, it remains questionable whether quantitative expression changes due to impaired chromatin interactions would have a significant cellular consequence at all.

ATP2B4

As previously described, *ATP2B4* encodes the plasma membrane calcium pump 4 (1.3). It is a protein that actively extrudes calcium ions from the cell and plays a pivotal role in erythrocytes. The chromosomal interactions between the resistance locus and the upstream region of *ATP2B4* (Hit 4) could be affected by sequence variations within the locus. This might cause an expression shift of the gene. With respect to the functional characteristics of the protein and its dominant expression in RBC membranes, a varied expression might affect the parasites development during the blood stage. Therefore, this gene is the most likely of all genes adjacent to the detected 4C-Seq hits to contribute to the mechanisms underlying the published resistance locus. The fact that different resistance-associated SNPs are also associated to the RNA and protein levels in human samples (6.1), support this hypothesis.

A general difficulty of interpreting the results from the present 4C-Seq analyses is the fact that all experiments have been conducted in continuous cell lines. Although they were chosen to represent different host cells and tissues involved in the life cycle of *P. falciparum*, they can only be seen as a model system. Since all cell lines are derived from tumors it is also unknown whether the haplotype structure of the viewpoint region corresponds to the haplotype structure of healthy tissues. A thorough examination of the potential influence that sequence variants have on the chromosomal interactions, would require a comparison of data from 4C-Seq experiments of individuals with a haplotype characterized as being protective and a non-protective. Here, the lack of

knowing the target cell line in which the variants mediate the protective effect becomes crucial.

4C-Seq experiments are a valuable method for the research question of this study. Yet, it is hard to acquire an appropriate sample material. Nevertheless, the results from the experiments conducted for this project show a pattern of the chromatin interactions of the published resistance locus that appears to be independent from the selected cell type.

6.2.1 Methods of 4C-Seq analysis

One major difficulty in the application of 4C-Seq is the lack of a commonly accepted statistical analysis method. This might be largely due to the fact that it is a relatively new method. Multiple pipelines, partially containing algorithms categorizing the interactions of the viewpoint, have been published (Walter et al. 2014; Klein et al. 2015a). Yet, they have not been recited in application. In many of the studies in which 4C-Seq was applied, a simple observation of an increase of fragment counts was interpreted as indicating chromosomal interaction without further statistical analysis (Hughes et al. 2013).

For the current study, an approach of combining two methods of evaluation was chosen. The analysis with FourCSeq yielded four hits that were only identified as being significant in one cell line (K562). In contrast, data visualization with Basic4CSeq also showed increased fragment counts for all hits in the other cell lines (Figure 13).

A closer look at the data revealed a shortcoming of the FourCSeq analysis pipeline when applied to the data generated for this study. The manual of Felix Klein offers multiple steps of data quality control (Klein et al. 2015b), but only very vague instructions on how data should be discarded. In Figure 22 and Figure 23, histograms of the data's z-scores and the QQ-plots of the quantiles are shown.

In both cases it can be observed that the plots generated from the data of K562 cells differs from those obtained with the other cell lines' data. Nevertheless, the examples shown in the FourCSeq manual generate the impression that the differences seen in the present data might not be so evident that the data should be excluded from the analysis. However, the reference count from which z-scores are calculated is generated from the complete mass of data that is included into the evaluation method. Therefore, it is plausible that cell line-specific fragment counts that differ in their distribution from the other input data generate significant signals.

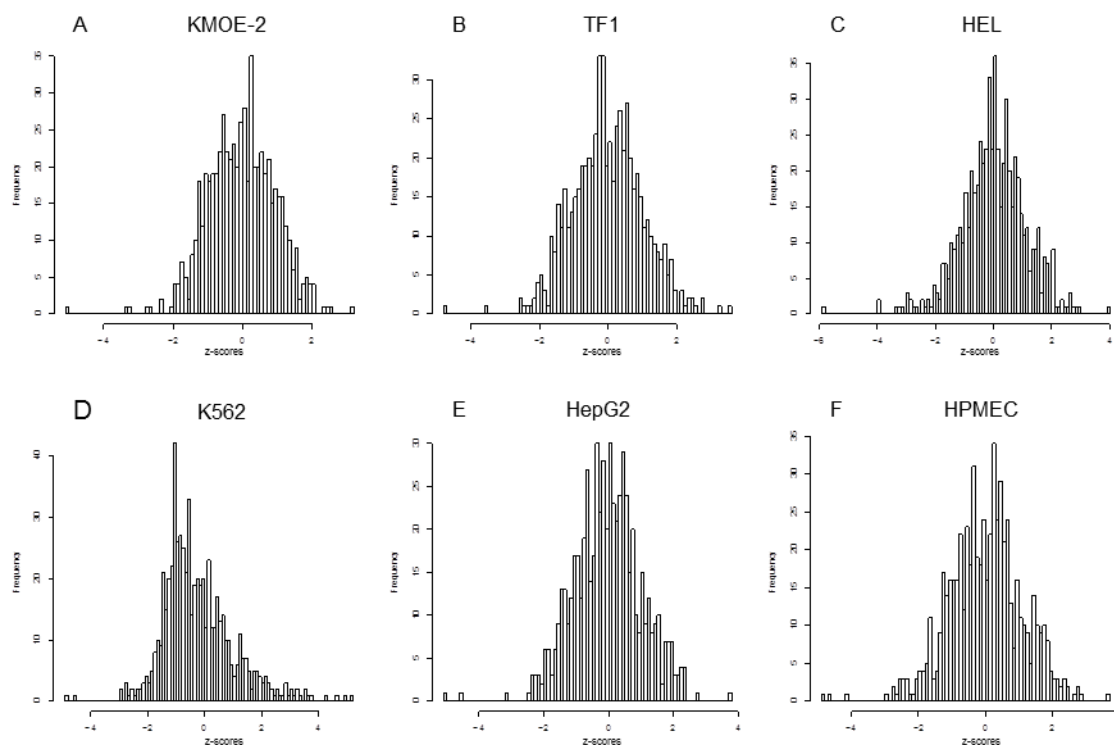


Figure 22: Z-Scores calculated for the fragment counts in 4C-Seq libraries for VP_Locus from six selected cell lines. The cell lines are (A) KMOE-2, (B) TF1, (C) HEL, (D) K562, (E) HepG2, (F) HPMEC.

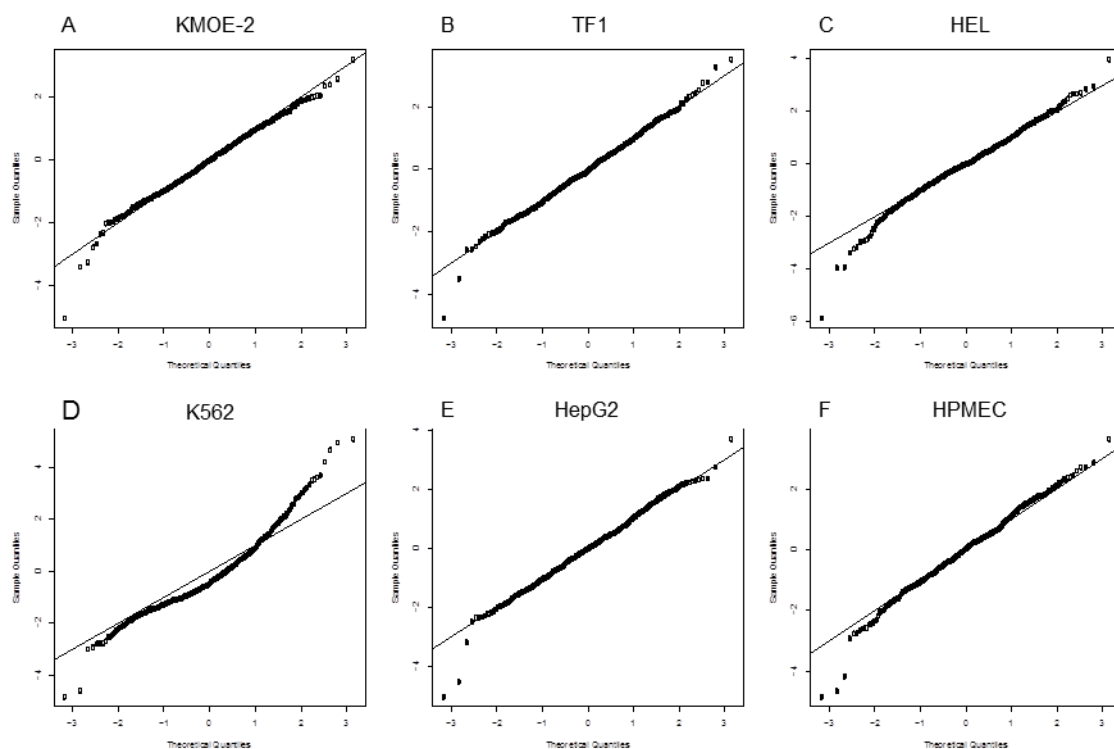


Figure 23: QQ-plots of the quantiles of fragment counts in 4C-Seq libraries for VP_Locus from six selected cell lines. The cell lines are (A) KMOE-2, (B) TF1, (C) HEL, (D) K562, (E) HepG2, (F) HPMEC.

However, the visualization of the complete data by using Basic4CSeq and the occurrence of the respective chromatin-chromatin interactions in published data confirms the presence of these interactions in numerous cell lines. Also, the reciprocal 4C-Seq

analyses of the same samples underline the results. Still, the described difficulties in the application of published analysis pipelines emphasize the need for a generally accepted and reliable pipeline to analyze 4C-Seq data.

6.3 Expression of *ATP2B4*

Statistical analyses in this study did not show differences of the mRNA expression of *ATP2B4* in whole blood samples from donors with different genotypes of the variant rs10900585, which is associated with resistance to severe malaria. Neither the level of complete *ATP2B4* mRNA nor the individual levels of the two known splice variants varied in samples from the three donor groups of African ancestry. Moreover, the expression ratio of the respective splice variants and the total expression of *ATP2B4* do not differ with regard to the donor's genotype. The application of the linear model rather implies that the level of one isoform is simply determined by the level of total *ATP2B4* expression.

These findings suggest that the five published SNPs with the strongest association (Timmann et al. 2012), or the causal variant(s) they might tag, do not mediate a genotype-specific quantitative regulation of overall *ATP2B4* transcription or splicing of *ATP2B4* pre-mRNAs into the two known isoforms in humans of African ancestry. This is contradictory to previous results from large eQTL and sQTL studies conducted with whole blood samples from donors of European ancestry (6.1).

However, the results from the conducted expression analyses do not completely exclude the possibility that a quantitative change in *ATP2B4* expression or differential splicing is responsible for the published resistance association. First, a study group of 200 samples might not generate the statistical power to display the difference in expression levels that is associated to the five resistance SNPs published, especially when assuming them to be not causal but tagging the causal variant(s). In comparison, both previous studies that contain genome-wide screening for eQTLs and sQTLs included samples from approximately 5,300 individuals. A second shortcoming of the qPCR approach from the present study is the choice of sample material. There were two reasons for choosing to determine RNA expression from whole blood samples. The first was the possibility to use the previously published results from European donors as a reference. The second reason was the accessibility of donor material. With regard to the general pathophysiology of a *P. falciparum* infection there are three cell types from the host, which are most likely to affect the course of infection when changes in their cellular

processes appear: hepatocytes, vascular endothelial cells and erythrocytes. Of these, the latter is the only sample material which is easy to access. Unfortunately, RBCs are unfavorable for genetic studies due to the lack of a nucleus. Therefore, erythroid precursors that still contain a nucleus (e.g. erythroblast) or recently ejected it and therefore most likely still contain all mRNA (reticulocytes) would be the appropriate material. No protocol for erythroblast isolation was found that would have been feasible to apply under lab conditions in Ghana. Attempts to isolate sufficient reticulocytes from a reasonable volume of blood from untreated donors failed during establishing methods for the sample collection. If a potential difference in gene regulation is only present in that specific cell type it is most likely that it cannot be detected in whole blood due to effect dilution by leucocytic mRNA. Hence, the chromatin-chromatin interaction between the locus and the upstream region of *ATP2B4* that is present in multiple cell lines is not affected by sequence variation within the locus in a way that a quantitatively influence on *ATP2B4* transcription and splicing of *ATP2B4* pre-mRNA can be detected in RNA isolates from whole blood.

6.4 Parasite-proliferation assay

The parasite culture experiment conducted in this study shows that the proliferation rate of *P. falciparum* lab strain 3D7 is significantly impaired in erythrocytes from donors that carry the homozygous minor allele (G) of rs10900585 in comparison to the ones carrying the homozygous major allele (T). This complies with the result from the logistic regression, that the G allele is associated to a resistance to severe malaria (Timmann et al. 2012).

The assay established for this study, enables the verification of relatively small differences of malaria parasite growth in erythrocytes from donors with different index genotypes. In general, cultivation of *P. falciparum* is an established system (Trager and Jensen 1976). The parasite proliferation in culture is very sensitive to environmental changes (Schuster 2002) and therefore it is essential to keep all cultures at the same time and under the exact same conditions when examining quantitative differences. This limited the number of included erythrocyte samples to 30. An automated cultivation system would enable to examine more samples, which would yield more power for statistical analyses and prove even smaller differences in parasite growth. The relatively large interquartile range (IQR) of the median values of the three groups shows how many factors apparently influence the parasite proliferation even if

known factors like different blood groups of the ABO system, the rhesus factor and the protective HbS alleles are excluded. The fact that the conducted assay revealed a reproducible and significant difference in proliferation of parasites within RBCs from the examined donor groups emphasizes the validity of the result.

It was observed, that the end point defined by a parasitemia of over 10 % in the positive control was reached 24 h later in the replication experiment. When one compares the graphs of the parasitemia of the two experiments (Figure 17, Figure A - 11) it is apparent that the parasites proliferate slower in the replicate. This is most likely due to the fact that the RBCs imported from Africa had to be stored a week longer than for the initial experiment. The development of *P. falciparum* within erythrocytes from donors with the heterozygous rs10900585 genotype showed divergent results in the initial experiment and the replication. It might be possible to determine the tendency of parasite proliferation within these cultures if the sample size and thereby the statistical power of the results is increased. The assay does not unravel the cellular changes mediated by the causal genetic variant(s). It focuses on the asexual proliferation of the parasite in the human RBC and excludes all other aspects of the complex pathophysiology of a *P. falciparum* infection. Conceivably, other stages of the infection could further influenced by the *ATP2B4* variants.

In summary, it was shown that a resistance-associated SNP is associated with an impaired parasite proliferation within the erythrocyte. The observed reduction in parasite-proliferation rate could be a biological explanation for the resistance association, especially when considering the exponential increase of IEs in the infected human. The fact that a significant difference of parasite proliferation with respect to the index SNP rs10900585 was demonstrated in a relatively small number of samples indicates that the underlying mechanism is caused by a common genetic variant with low OR rather than a rare variant with a strong effect. Most likely, rare variants would not influence the median values for proliferation of the genotypic groups compared in this assay.

To identify the parasite's metabolic processes that are affected and cause the impaired proliferation, a range of sophisticated parasitological methods would have to be applied. However, a first step to narrow down the possibilities would be the identification of the part of the blood stage that is inhibited. Whether it is the parasite's invasion, its maturation within the RBC or the stage when infectious merozoites are released can be determined by life-cell imaging.

7 Conclusions and prospects

Combining all findings from this study and the discussed results from recently published studies, one can hypothesize how the resistance to severe malaria which is associated with the minor allele of numerous variants in LD with the index SNP rs10900585 is mediated. It appears as if transcription and splicing, the early stages of gene expression are not influenced by the genotypes of these variants. However, they seem to affect processes downstream of the translation of *ATP2B4* mRNA. Exon 2 that is contained within the resistance-locus region partly encodes the 5' region of the mRNA. The latter is majorly responsible for the translation of genes (Mignone et al. 2002). This implies that sequence variation within this region might cause insufficient translation by influencing the efficiency of ribosome binding leading to the observed decrease of PMCA4 expression in RBC membranes of Hungarian donors (Zambo et al. 2017). The resulting impairment of calcium extrusion from erythrocytes appears not to have a significant effect on the essential functionality of uninfected RBCs as there is no such phenotype known in humans carrying the respective genotype. However, it is likely that it causes a respective imbalance of ions within the parasitophorus vacuole, as the parasite with its active metabolism demands a highly efficient maintenance of calcium balance. This might impair the parasite proliferation as shown in the proliferation assay in this study and thereby leaves more time for an infected human to control the infection and prevent severe malaria.

To reveal the missing steps of this mechanism, first the reduced PMCA4 expression and the thereby impaired calcium extrusion in RBCs from donors carrying the resistance-associated genotype need to be confirmed in samples from African donors. To prove that an insufficient translation of *ATP2B4* mRNA is responsible for that, *in vitro* translation should be performed using mRNA extracts of samples from donors carrying the respective genotype.

8 List of literature

Abecasis, Goncalo R.; Auton, Adam; Brooks, Lisa D.; DePristo, Mark A.; Durbin, Richard M.; Handsaker, Robert E. et al. (2012): An integrated map of genetic variation from 1,092 human genomes. In: *Nature* 491 (7422), S. 56–65. DOI: 10.1038/nature11632.

Allison, A. C. (1954): Protection Afforded by Sickle-cell Trait Against Subtertian Malarial Infection. In: *Br Med J.* 1954 (1(4857)), 290–294.

Aly, Ahmed S. I.; Vaughan, Ashley M.; Kappe, Stefan H. I. (2009): Malaria parasite development in the mosquito and infection of the mammalian host. In: *Annual review of microbiology* 63, S. 195–221. DOI: 10.1146/annurev.micro.091208.073403.

Artieda, Marta; Cenarro, Ana; Ganan, Alberto; Jerico, Ivonne; Gonzalvo, Carmen; Casado, Juan M. et al. (2003): Serum chitotriosidase activity is increased in subjects with atherosclerosis disease. In: *Arteriosclerosis, thrombosis, and vascular biology* 23 (9), S. 1645–1652. DOI: 10.1161/01.ATV.0000089329.09061.07.

Auton, Adam; Brooks, Lisa D.; Durbin, Richard M.; Garrison, Erik P.; Kang, Hyun Min; Korbelt, Jan O. et al. (2015): A global reference for human genetic variation. In: *Nature* 526 (7571), S. 68–74. DOI: 10.1038/nature15393.

Band, Gavin; Rockett, Kirk A.; Spencer, Chris C. A.; Kwiatkowski, Dominic P. (2015): A novel locus of resistance to severe malaria in a region of ancient balancing selection. In: *Nature* 526 (7572), S. 253–257. DOI: 10.1038/nature15390.

Barone, R.; Di Gregorio, F.; Romeo, M. A.; Schiliro, G.; Pavone, L. (1999): Plasma chitotriosidase activity in patients with beta-thalassemia. In: *Blood cells, molecules & diseases* 25 (1), S. 1–8. DOI: 10.1006/bcmd.1999.0221.

Bedu-Addo, George; Meese, Stefanie; Mockenhaupt, Frank P. (2013): An ATP2B4 polymorphism protects against malaria in pregnancy. In: *The Journal of infectious diseases* 207 (10), S. 1600–1603. DOI: 10.1093/infdis/jit070.

Bogdanova, Anna; Makhro, Asya; Wang, Jue; Lipp, Peter; Kaestner, Lars (2013): Calcium in red blood cells-a perilous balance. In: *International journal of molecular sciences* 14 (5), S. 9848–9872. DOI: 10.3390/ijms14059848.

- Bouwman, Britta A. M.; Laat, Wouter de (2015): Getting the genome in shape: the formation of loops, domains and compartments. In: *Genome biology* 16, S. 154. DOI: 10.1186/s13059-015-0730-1.
- Buffet, Pierre A.; Safeukui, Innocent; Deplaine, Guillaume; Brousse, Valentine; Prendki, Virginie; Thellier, Marc et al. (2011): The pathogenesis of Plasmodium falciparum malaria in humans: insights from splenic physiology. In: *Blood* 117 (2), S. 381–392. DOI: 10.1182/blood-2010-04-202911.
- Caride, Ariel J.; Filoteo, Adelaida G.; Penniston, John T.; Strehler, Emanuel E. (2007): The plasma membrane Ca²⁺ pump isoform 4a differs from isoform 4b in the mechanism of calmodulin binding and activation kinetics: implications for Ca²⁺ signaling. In: *The Journal of biological chemistry* 282 (35), S. 25640–25648. DOI: 10.1074/jbc.M701129200.
- Clarke, Geraldine M.; Rockett, Kirk; Kivinen, Katja; Hubbart, Christina; Jeffreys, Anna E.; Rowlands, Kate et al. (2017): Characterisation of the opposing effects of G6PD deficiency on cerebral malaria and severe malarial anaemia. In: *eLife* 6. DOI: 10.7554/eLife.15085.
- Cowman, Alan F.; Berry, Drew; Baum, Jake (2012): The cellular and molecular basis for malaria parasite invasion of the human red blood cell. In: *The Journal of cell biology* 198 (6), S. 961–971. DOI: 10.1083/jcb.201206112.
- Di Leva, Francesca; Domi, Teuta; Fedrizzi, Laura; Lim, Dmitry; Carafoli, Ernesto (2008): The plasma membrane Ca²⁺ ATPase of animal cells: structure, function and regulation. In: *Archives of biochemistry and biophysics* 476 (1), S. 65–74. DOI: 10.1016/j.abb.2008.02.026.
- Edwards, Stacey L.; Beesley, Jonathan; French, Juliet D.; Dunning, Alison M. (2013): Beyond GWASs: illuminating the dark road from association to function. In: *American journal of human genetics* 93 (5), S. 779–797. DOI: 10.1016/j.ajhg.2013.10.012.
- ENCODE Project Consortium (2012): An integrated encyclopedia of DNA elements in the human genome. In: *Nature* 489 (7414), S. 57–74. DOI: 10.1038/nature11247.

- Enomoto, Masahiro; Kawazu, Shin-ichiro; Kawai, Satoru; Furuyama, Wakako; Ikegami, Tohru; Watanabe, Jun-ichi; Mikoshiba, Katsuhiko (2012): Blockage of spontaneous Ca²⁺ oscillation causes cell death in intraerythrocytic *Plasmodium falciparum*. In: *PloS one* 7 (7), e39499. DOI: 10.1371/journal.pone.0039499.
- Fung, Jenny N.; Rogers, Peter A. W.; Montgomery, Grant W. (2015): Identifying the biological basis of GWAS hits for endometriosis. In: *Biology of reproduction* 92 (4), S. 87. DOI: 10.1095/biolreprod.114.126458.
- Gazarini, Marcos L.; Thomas, Andrew P.; Pozzan, Tullio; Garcia, Celia R. S. (2003): Calcium signaling in a low calcium environment: how the intracellular malaria parasite solves the problem. In: *The Journal of cell biology* 161 (1), S. 103–110. DOI: 10.1083/jcb.200212130.
- Haldane, J. B.S. (1949): The rate of mutation of human genes. In: *Int. Congr. Genet.* 1949 (8th).
- Hedrick, P. W. (2011): Population genetics of malaria resistance in humans. In: *Heredity* 107 (4), S. 283–304. DOI: 10.1038/hdy.2011.16.
- Hou, Lin; Zhao, Hongyu (2013): A review of post-GWAS prioritization approaches. In: *Frontiers in genetics* 4, S. 280. DOI: 10.3389/fgene.2013.00280.
- Howie, Bryan; Fuchsberger, Christian; Stephens, Matthew; Marchini, Jonathan; Abecasis, Goncalo R. (2012): Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. In: *Nature genetics* 44 (8), S. 955–959. DOI: 10.1038/ng.2354.
- Howie, Bryan; Marchini, Jonathan; Stephens, Matthew (2011): Genotype imputation with thousands of genomes. In: *G3 (Bethesda, Md.)* 1 (6), S. 457–470. DOI: 10.1534/g3.111.001198.
- Howie, Bryan N.; Donnelly, Peter; Marchini, Jonathan (2009): A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. In: *PLoS genetics* 5 (6), e1000529. DOI: 10.1371/journal.pgen.1000529.
- Huang, Qingyang (2015): Genetic study of complex diseases in the post-GWAS era. In: *Journal of genetics and genomics = Yi chuan xue bao* 42 (3), S. 87–98. DOI: 10.1016/j.jgg.2015.02.001.

- Hughes, Jim R.; Lower, Karen M.; Dunham, Ian; Taylor, Stephen; Gobbi, Marco de; Sloane-Stanley, Jacqueline A. et al. (2013): High-resolution analysis of cis-acting regulatory networks at the alpha-globin locus. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 368 (1620), S. 20120361. DOI: 10.1098/rstb.2012.0361.
- Jallow, Muminatou; Teo, Yik Ying; Small, Kerrin S.; Rockett, Kirk A.; Deloukas, Panos; Clark, Taane G. et al. (2009): Genome-wide and fine-resolution association analysis of malaria in West Africa. In: *Nature genetics* 41 (6), S. 657–665. DOI: 10.1038/ng.388.
- Javierre, Biola M.; Burren, Oliver S.; Wilder, Steven P.; Kreuzhuber, Roman; Hill, Steven M.; Sewitz, Sven et al. (2016): Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. In: *Cell* 167 (5), 1369-1384.e19. DOI: 10.1016/j.cell.2016.09.037.
- Klein, Felix A.; Pakozdi, Tibor; Anders, Simon; Ghavi-Helm, Yad; Furlong, Eileen E. M.; Huber, Wolfgang (2015a): FourCSeq analysis work flow. Online: <http://bioconductor.org/packages/release/bioc/html/FourCSeq.html>.
- Klein, Felix A.; Pakozdi, Tibor; Anders, Simon; Ghavi-Helm, Yad; Furlong, Eileen E. M.; Huber, Wolfgang (2015b): FourCSeq: analysis of 4C sequencing data. In: *Bioinformatics (Oxford, England)* 31 (19), S. 3085–3091. DOI: 10.1093/bioinformatics/btv335.
- Klein, Robert J.; Zeiss, Caroline; Chew, Emily Y.; Tsai, Jen-Yue; Sackler, Richard S.; Haynes, Chad et al. (2005): Complement factor H polymorphism in age-related macular degeneration. In: *Science (New York, N.Y.)* 308 (5720), S. 385–389. DOI: 10.1126/science.1109557.
- Kraemer, Susan M.; Smith, Joseph D. (2006): A family affair: var genes, PfEMP1 binding, and malaria disease. In: *Current opinion in microbiology* 9 (4), S. 374–380. DOI: 10.1016/j.mib.2006.06.006.
- Krebs, Joachim (2015): The plethora of PMCA isoforms: Alternative splicing and differential expression. In: *Biochimica et biophysica acta* 1853 (9), S. 2018–2024. DOI: 10.1016/j.bbamcr.2014.12.020.

Kwiatkowski, Dominic P. (2005): How malaria has affected the human genome and what human genetics can teach us about malaria. In: *American journal of human genetics* 77 (2), S. 171–192. DOI: 10.1086/432519.

Laat, Wouter de; Dekker, Job (2012): 3C-based technologies to study the shape of the genome. In: *Methods (San Diego, Calif.)* 58 (3), S. 189–191. DOI: 10.1016/j.jymeth.2012.11.005.

Laveran, Alphonse (1881): Un nouveau parasite trouvé dans le sang de malades atteints de fièvre palustre. Origine parasitaire des accidents de l'impaludisme. In: *Bull Mém Soc Méd Hôpitaux Paris* 1881 (17), S. 158–164.

Lelliott, Patrick M.; McMorran, Brendan J.; Foote, Simon J.; Burgio, Gaetan (2015): The influence of host genetics on erythrocytes and malaria infection: is there therapeutic potential? In: *Malaria journal* 14, S. 289. DOI: 10.1186/s12936-015-0809-x.

Luzzatto, L.; Usanga, F. A.; Reddy, S. (1969): Glucose-6-phosphate dehydrogenase deficient red cells: resistance to infection by malarial parasites. In: *Science (New York, N.Y.)* 164 (3881), S. 839–842.

MacArthur, Jacqueline; Bowler, Emily; Cerezo, Maria; Gil, Laurent; Hall, Peggy; Hastings, Emma et al. (2017): The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). In: *Nucleic acids research* 45 (D1), D896–D901. DOI: 10.1093/nar/gkw1133.

Magi, Reedik; Asimit, Jennifer L.; Day-Williams, Aaron G.; Zeggini, Eleftheria; Morris, Andrew P. (2012): Genome-wide association analysis of imputed rare variants: application to seven common complex diseases. In: *Genetic epidemiology* 36 (8), S. 785–796. DOI: 10.1002/gepi.21675.

Majava, Marja; Bishop, Paul N.; Hagg, Pasi; Scott, Paul G.; Rice, Aine; Inglehearn, Chris et al. (2007): Novel mutations in the small leucine-rich repeat protein/proteoglycan (SLRP) genes in high myopia. In: *Human mutation* 28 (4), S. 336–344. DOI: 10.1002/humu.20444.

- Malaria Genomic Epidemiology Network (2014): Reappraisal of known malaria resistance loci in a large multicenter study. In: *Nature genetics* 46 (11), S. 1197–1204. DOI: 10.1038/ng.3107.
- Mao, Bijing; Zhang, Zhimin; Wang, Ge (2015): BTG2: a rising star of tumor suppressors (review). In: *International journal of oncology* 46 (2), S. 459–464. DOI: 10.3892/ijo.2014.2765.
- May, Jurgen; Evans, Jennifer A.; Timmann, Christian; Ehmen, Christa; Busch, Wibke; Thye, Thorsten et al. (2007): Hemoglobin variants and disease manifestations in severe falciparum malaria. In: *JAMA* 297 (20), S. 2220–2226. DOI: 10.1001/jama.297.20.2220.
- Merline, Rosetta; Schaefer, Roland M.; Schaefer, Liliana (2009): The matricellular functions of small leucine-rich proteoglycans (SLRPs). In: *Journal of cell communication and signaling* 3 (3-4), S. 323–335. DOI: 10.1007/s12079-009-0066-2.
- Mifsud, Borbala; Tavares-Cadete, Filipe; Young, Alice N.; Sugar, Robert; Schoenfelder, Stefan; Ferreira, Lauren et al. (2015): Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. In: *Nature genetics* 47 (6), S. 598–606. DOI: 10.1038/ng.3286.
- Mignone, Flavio; Gissi, Carmela; Liuni, Sabino; Pesole, Graziano (2002): Untranslated regions of mRNAs. In: *Genome biology* 3 (3).
- Miller, Louis H.; Baruch, Dror I.; Marsh, Kevin; Dumbo, Ogobara K. (2002): The pathogenic basis of malaria. In: *Nature* 415 (6872), S. 673–679. DOI: 10.1038/415673a.
- Nica, Alexandra C.; Dermitzakis, Emmanouil T. (2013): Expression quantitative trait loci: present and future. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 368 (1620), S. 20120362. DOI: 10.1098/rstb.2012.0362.
- Pachot, Alexandre; Blond, Jean-Luc; Mougin, Bruno; Miossec, Pierre (2004): Peptidylpropyl isomerase B (PPIB): a suitable reference gene for mRNA quantification in peripheral whole blood. In: *Journal of biotechnology* 114 (1-2), S. 121–124. DOI: 10.1016/j.jbiotec.2004.07.001.

- Pain, A.; Ferguson, D. J.; Kai, O.; Urban, B. C.; Lowe, B.; Marsh, K.; Roberts, D. J. (2001): Platelet-mediated clumping of Plasmodium falciparum-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. In: *Proceedings of the National Academy of Sciences of the United States of America* 98 (4), S. 1805–1810. DOI: 10.1073/pnas.98.4.1805.
- Pouvelle, B.; Buffet, P. A.; Lepolard, C.; Scherf, A.; Gysin, J. (2000): Cytoadhesion of Plasmodium falciparum ring-stage-infected erythrocytes. In: *Nature medicine* 6 (11), S. 1264–1268. DOI: 10.1038/81374.
- Pruim, Randall J.; Welch, Ryan P.; Sanna, Serena; Teslovich, Tanya M.; Chines, Peter S.; Gliedt, Terry P. et al. (2010): LocusZoom: regional visualization of genome-wide association scan results. In: *Bioinformatics (Oxford, England)* 26 (18), S. 2336–2337. DOI: 10.1093/bioinformatics/btq419.
- Rao, Suhas S. P.; Huntley, Miriam H.; Durand, Neva C.; Stamenova, Elena K.; Bochkov, Ivan D.; Robinson, James T. et al. (2014): A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. In: *Cell* 159 (7), S. 1665–1680. DOI: 10.1016/j.cell.2014.11.021.
- Ross, Ronald (1898): The role of the mosquito in the evolution of the malaria parasite. In: *Lancet* 1898 (2), S. 489.
- Schaefer, Liliana; Iozzo, Renato V. (2008): Biological functions of the small leucine-rich proteoglycans: from genetics to signal transduction. In: *The Journal of biological chemistry* 283 (31), S. 21305–21309. DOI: 10.1074/jbc.R800020200.
- Schuster, F. L. (2002): Cultivation of Plasmodium spp. In: *Clinical Microbiology Reviews* 15 (3), S. 355–364. DOI: 10.1128/CMR.15.3.355-364.2002.
- Shami, Annelie; Tengryd, Christoffer; Ascitto, Giuseppe; Bengtsson, Eva; Nilsson, Jan; Hultgardh-Nilsson, Anna; Goncalves, Isabel (2015): Expression of fibromodulin in carotid atherosclerotic plaques is associated with diabetes and cerebrovascular events. In: *Atherosclerosis* 241 (2), S. 701–708. DOI: 10.1016/j.atherosclerosis.2015.06.023.

- Sherman, Irwin W.; Eda, Shigetoshi; Winograd, Enrique (2003): Cytoadherence and sequestration in *Plasmodium falciparum*: defining the ties that bind. In: *Microbes and infection* 5 (10), S. 897–909.
- SHORTT, H. E.; GARNHAM, P. C. (1948): Pre-erythrocytic stage in mammalian malaria parasites. In: *Nature* 161 (4082), S. 126.
- Skorokhod, Oleksii A.; Caione, Luisa; Marrocco, Tiziana; Migliardi, Giorgia; Barrera, Valentina; Arese, Paolo et al. (2010): Inhibition of erythropoiesis in malaria anemia: role of hemozoin and hemozoin-generated 4-hydroxynonenal. In: *Blood* 116 (20), S. 4328–4337. DOI: 10.1182/blood-2010-03-272781.
- Smemo, Scott; Tena, Juan J.; Kim, Kyoung-Han; Gamazon, Eric R.; Sakabe, Noboru J.; Gomez-Marin, Carlos et al. (2014): Obesity-associated variants within FTO form long-range functional connections with IRX3. In: *Nature* 507 (7492), S. 371–375. DOI: 10.1038/nature13138.
- Spielman, Richard S.; Bastone, Laurel A.; Burdick, Joshua T.; Morley, Michael; Ewens, Warren J.; Cheung, Vivian G. (2007): Common genetic variants account for differences in gene expression among ethnic groups. In: *Nature genetics* 39 (2), S. 226–231. DOI: 10.1038/ng1955.
- Splinter, Erik; Wit, Elzo de; van de Werken, Harmen J. G.; Klous, Petra; Laat, Wouter de (2012): Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. In: *Methods (San Diego, Calif.)* 58 (3), S. 221–230. DOI: 10.1016/j.ymeth.2012.04.009.
- Stauffer, T. P.; Guerini, D.; Carafoli, E. (1995): Tissue distribution of the four gene products of the plasma membrane Ca²⁺ pump. A study using specific antibodies. In: *The Journal of biological chemistry* 270 (20), S. 12184–12190.
- Strehler, Emanuel E. (2001): Role of Alternative Splicing in Generating Isoform Diversity Among Plasma Membrane Calcium Pumps. In: *Physiological Reviews* 81 (1).
- Sturm, Angelika; Amino, Rogerio; van de Sand, Claudia; Regen, Tommy; Retzlaff, Silke; Rennenberg, Annika et al. (2006): Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. In: *Science (New York, N.Y.)* 313 (5791), S. 1287–1290. DOI: 10.1126/science.1129720.

- Svensson, L.; Aszodi, A.; Reinholt, F. P.; Fassler, R.; Heinegard, D.; Oldberg, A. (1999): Fibromodulin-null Mice Have Abnormal Collagen Fibrils, Tissue Organization, and Altered Lumican Deposition in Tendon. In: *Journal of Biological Chemistry* 274 (14), S. 9636–9647. DOI: 10.1074/jbc.274.14.9636.
- Timmann, Christian; Thye, Thorsten; Vens, Maren; Evans, Jennifer; May, Jurgen; Ehmen, Christa et al. (2012): Genome-wide association study indicates two novel resistance loci for severe malaria. In: *Nature* 489 (7416), S. 443–446. DOI: 10.1038/nature11334.
- Trager, W.; Jensen, J. B. (1976): Human malaria parasites in continuous culture. In: *Science (New York, N.Y.)* 193 (4254), S. 673–675.
- Trampuz, Andrej; Jereb, Matjaz; Muzlovic, Igor; Prabhu, Rajesh M. (2003): Clinical review: Severe malaria. In: *Critical care (London, England)* 7 (4), S. 315–323. DOI: 10.1186/cc2183.
- Udeinya, I.; Schmidt, J.; Aikawa, M.; Miller, L.; Green, I. (1981): Falciparum malaria-infected erythrocytes specifically bind to cultured human endothelial cells. In: *Science* 213 (4507), S. 555–557. DOI: 10.1126/science.7017935.
- Udomsangpetch, R. (1989): Plasmodium falciparum-infected erythrocytes form spontaneous erythrocyte rosettes. In: *Journal of Experimental Medicine* 169 (5), S. 1835–1840. DOI: 10.1084/jem.169.5.1835.
- van de Werken, Harmen J. G.; Vree, Paula J. P. de; Splinter, Erik; Holwerda, Sjoerd J. B.; Klous, Petra; Wit, Elzo de; Laat, Wouter de (2012): 4C technology: protocols and data analysis. In: *Methods in enzymology* 513, S. 89–112. DOI: 10.1016/B978-0-12-391938-0.00004-5.
- van der Sijde, Marijke R.; Ng, Aylwin; Fu, Jingyuan (2014): Systems genetics: From GWAS to disease pathways. In: *Biochimica et biophysica acta* 1842 (10), S. 1903–1909. DOI: 10.1016/j.bbadis.2014.04.025.
- van Eijk, Marco; van Roomen, Cindy P. A. A.; Renkema, G. Herma; Bussink, Anton P.; Andrews, Laura; Blommaart, Edward F. C. et al. (2005): Characterization of human phagocyte-derived chitotriosidase, a component of innate immunity. In: *International immunology* 17 (11), S. 1505–1512. DOI: 10.1093/intimm/dxh328.

- Virgilio, L. Lew; Daw, Nuala; Perdomo, Deisy; Etzion, Zipora; Bookchin, Robert M.; Tiffert, Teresa (2003): Distribution of plasma membrane Ca²⁺ pump activity in normal human red blood cells. In: *Blood* 102 (12), S. 4206–4213. DOI: 10.1182/blood-2003-06-1787.
- Walter, Carolin; Schuetzmann, Daniel; Rosenbauer, Frank; Dugas, Martin (2014): Basic4Cseq: an R/Bioconductor package for the analysis of 4C-seq data. In: *Bioinformatics (Oxford, England)* 30 (22), S. 3268–3269. DOI: 10.1093/bioinformatics/btu497.
- Walter, Carolin; Schuetzmann, Daniel; Rosenbauer, Frank; Dugas, Martin (2015): Basic4Cseq: an R/Bioconductor package for analyzing 4C-seq data (Vignette). Online: <https://www.bioconductor.org/packages/release/bioc/html/Basic4Cseq.html>.
- Wang, Yanli; Zhang, Bo; Zhang, Lijun; An, Lin; Xu, Jie; Li, Daofeng et al. (2017): The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions.
- Wery, Maxime; Kwapisz, Marta; Morillon, Antonin (2011): Noncoding RNAs in gene regulation. In: *Wiley interdisciplinary reviews. Systems biology and medicine* 3 (6), S. 728–738. DOI: 10.1002/wsbm.148.
- Westra, Harm-Jan; Franke, Lude (2014): From genome to function by studying eQTLs. In: *Biochimica et biophysica acta* 1842 (10), S. 1896–1902. DOI: 10.1016/j.bbadis.2014.04.024.
- Westra, Harm-Jan; Peters, Marjolein J.; Esko, Tonu; Yaghootkar, Hanieh; Schurmann, Claudia; Kettunen, Johannes et al. (2013): Systematic identification of trans eQTLs as putative drivers of known disease associations. In: *Nature genetics* 45 (10), S. 1238–1243. DOI: 10.1038/ng.2756.
- WHO (2010): Guidelines for the treatment of Malaria. 3rd. Geneva: World Health Organization.
- WHO (2016): World Malaria Report 2015. Geneva: World Health Organization. Online: <http://ebookcentral.proquest.com/lib/gbv/detail.action?docID=4778804>.

WHO (2017): The top 10 causes of death. World Health Organization. Online: <http://www.who.int/mediacentre/factsheets/fs310/en/index1.html>, zuletzt aktualisiert am 13.03.2017, zuletzt geprüft am 13.03.2017.

Yates, Andrew; Akanni, Wasiu; Amode, M. Ridwan; Barrell, Daniel; Billis, Konstantinos; Carvalho-Silva, Denise et al. (2016): Ensembl 2016. In: *Nucleic acids research* 44 (D1), D710-6. DOI: 10.1093/nar/gkv1157.

Zambo, Boglarka; Varady, Gyorgy; Padanyi, Rita; Szabo, Edit; Nemeth, Adrienn; Lango, Tamas et al. (2017): Decreased calcium pump expression in human erythrocytes is connected to a minor haplotype in the ATP2B4 gene. In: *Cell calcium*. DOI: 10.1016/j.ceca.2017.02.001.

Zhang, Xiaoling; Joehanes, Roby; Chen, Brian H.; Huan, Tianxiao; Ying, Saixia; Munson, Peter J. et al. (2015): Identification of common genetic variants controlling transcript isoform variation in human whole blood. In: *Nature genetics* 47 (4), S. 345–352. DOI: 10.1038/ng.3220.

Zheng, Hou-Feng; Rong, Jing-Jing; Liu, Ming; Han, Fang; Zhang, Xing-Wei; Richards, J. Brent; Wang, Li (2015): Performance of genotype imputation for low frequency and rare variants from the 1000 genomes. In: *PloS one* 10 (1), e0116487. DOI: 10.1371/journal.pone.0116487.

9 Appendix

9.1 Primers

9.1.1 Genotyping

Table A - 1: Primer sequences for applied genotyping assays (FRET-MCA).

Hbs	
forward primer	ACATTTGCTTCTGACACAAC
reverse primer	GCCCAGTTTCTATTGGTCTCC
sensor	CTCCTGTGGAGAAGTCTGC-fluorescein
anchor	other-GTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGA-phosphate
rs10900585	
forward primer	TAAATCACAAGGTATATAACTGTACCCATC
reverse primer	ATTAGCCACATGCCTGTAAC
sensor	other-GGCAACAAGAGTGAGACTCC-phosphate
anchor	CCATTGCACTCCAGCCTGCCT-fluorescein
rs2365860	
forward primer	CTTTATTCTTCTTACACACTAGGACG
reverse primer	TCTTTATGAGAGATGGGGTCTTG
sensor	TGGTAACTCTTACATTACTACAACT-fluorescein
anchor	other-TGAGGCAGAAGAAGTCTTGAGTCCAG-phosphate
rs10900589	
forward primer	AGTTTATTACATTTCTGATACTAACATGGT
reverse primer	TTCATGCCTTCTGATTATACTCC
sensor	other-TAGCTAAGATTCTCTGTGAGACT-phosphate
anchor	TGCATGTAGTCCAGCACTATCTGTGC-fluorescein
rs2365858	
forward primer	GCCATGTAGCTAAACATCTC
reverse primer	CTTAGCCTCCTGGTACTG
sensor	AAGGAACACAAGCCACTATCAAA-fluorescein
anchor	other-CATTGGGCTGAGGCTGCGCAT-phosphate
rs4951074	
forward primer	ATGAAGTACAGAAGGTGGCAGAGT
reverse primer	GCACCTCCAGTTTAATACACGCAT
sensor	CTAGAGATAAAGCAGTTCGATT-fluorescein
anchor	other-AAGGTCACAGCCAGAGGATGATGAAC-phosphate

9.1.2 4C-Seq

Table A - 2: Primer sequences for 4C-Seq experiments of VP_Locus.

VP_Locus	
ALL_P7_rv	CAAGCAGAAGACGGCATAACGAGAT GGGCAGGTCAAGGCTCTTAG
KMOE-2_P5_fw	AATGATACGGCGACCACCGAGATCTT ACACTCTTCCCTACACGACGCTCTTCCGATCT CT TGACTTCTTAAGGTTGCTATTGTACA
TF1_P5_fw	AATGATACGGCGACCACCGAGATCTT ACACTCTTCCCTACACGACGCTCTTCCGATCT AC TGACTTCTTAAGGTTGCTATTGTACA
HEL_P5_fw	AATGATACGGCGACCACCGAGATCTT ACACTCTTCCCTACACGACGCTCTTCCGATCT CG TGACTTCTTAAGGTTGCTATTGTACA
K562_P5_fw	AATGATACGGCGACCACCGAGATCTT ACACTCTTCCCTACACGACGCTCTTCCGATCT TG TGACTTCTTAAGGTTGCTATTGTACA
HepG2_P5_fw	AATGATACGGCGACCACCGAGATCTT ACACTCTTCCCTACACGACGCTCTTCCGATCT TA TGACTTCTTAAGGTTGCTATTGTACA
HPMEC_P5_fw	AATGATACGGCGACCACCGAGATCTT ACACTCTTCCCTACACGACGCTCTTCCGATCT GA TGACTTCTTAAGGTTGCTATTGTACA

Table A - 3: Primer sequences for 4C-Seq experiments of reciprocal experiments.

VP_Hit1	
ALL_P7_rv	CAAGCAGAAGACGGCATAACGAGAT CAGAGGCAGATAGACAGTCAGA
KMOE-2_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT AC CTTTGTCTCTATGTGTGTGTGTACA
TF1_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATC TG CTTTGTCTCTATGTGTGTGTGTACA
HEL_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT GA CTTTGTCTCTATGTGTGTGTGTACA
K562_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT AC CTTTGTCTCTATGTGTGTGTGTACA
HepG2_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT CT CTTTGTCTCTATGTGTGTGTGTACA
HPMEC_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT TC CTTTGTCTCTATGTGTGTGTGTACA
VP_Hit2	
ALL_P7_fw	CAAGCAGAAGACGGCATAACGAGAT TCCTCACTTGCACTCAGACA
KMOE-2_P5_rv	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT GA TCTGAAAGGGAAACACAAATGTACA
TF1_P5_rv	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT CT TCTGAAAGGGAAACACAAATGTACA
HEL_P5_rv	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT TC TCTGAAAGGGAAACACAAATGTACA
K562_P5_rv	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT GA TCTGAAAGGGAAACACAAATGTACA
HepG2_P5_rv	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT AG TCTGAAAGGGAAACACAAATGTACA
HPMEC_P5_rv	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT AC TCTGAAAGGGAAACACAAATGTACA
VP_Hit3	
ALL_P7_rv	CAAGCAGAAGACGGCATAACGAGAT ATTTCCCTTCTCTCCCTGC
KMOE-2_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT TG CCCTAATACTTTTGCTGAATGTACA
TF1_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT GA CCCTAATACTTTTGCTGAATGTACA
HEL_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT CT CCCTAATACTTTTGCTGAATGTACA
K562_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT TG CCCTAATACTTTTGCTGAATGTACA
HepG2_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT TC CCCTAATACTTTTGCTGAATGTACA
HPMEC_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT AG CCCTAATACTTTTGCTGAATGTACA
VP_Hit4	
ALL_P7_rv	CAAGCAGAAGACGGCATAACGAGAT ACAGCAAACCCCTCTCTA
KMOE-2_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT CT CCAGAGAATTAGTTGTTTTGATTGTACA
TF1_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT TC CCAGAGAATTAGTTGTTTTGATTGTACA
HEL_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT AG CCAGAGAATTAGTTGTTTTGATTGTACA
K562_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT CT CCAGAGAATTAGTTGTTTTGATTGTACA
HepG2_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT AC CCAGAGAATTAGTTGTTTTGATTGTACA
HPMEC_P5_fw	AATGATACGGCGACCACCGAGATCT ACATCTTTCCCTACACGACGCTCTTCCGATCT TG CCAGAGAATTAGTTGTTTTGATTGTACA

9.2 Tables

9.2.1 Computational fine mapping

Table A - 4: Results from logistic regressions of the resistance SNPs from this study (array, FRET-MCA) and the published GWAS.

	array		FRET-MCA		published	
	OR	P-value	OR	P-value	OR	P-value
rs10900585			0.6	2.21×10^{-5}	0.6	2.0×10^{-5}
rs2365860	0.58	2.45×10^{-5}	0.57	1.67×10^{-5}	0.58	2.5×10^{-5}
rs10900589	0.57	1.90×10^{-5}	0.57	1.73×10^{-5}	0.57	2.4×10^{-5}
rs2365858	0.57	1.40×10^{-5}	0.57	1.91×10^{-5}	0.56	1.1×10^{-5}
rs4951074	0.55	5.79×10^{-6}	0.58	8.37×10^{-5}	0.55	6.5×10^{-6}

Table A - 5: SNPs within Intron 1-2 and exon 2 of *ATP2B4* that are associated with the trait of severe malaria (p -value $< 10^{-4}$) after logistic regression in this study (*SNParray, #SNParray and FRET-MCA, ° RefSeq: NM_001684).

SNP	BP	ORIGIN	Position ATP2B4	keepHRM_ALL (64)		keepHRM_AFR (60)		keepGWAS_ALL (63)		keepGWAS_AFR (60)	
				OR	P	OR	P	OR	P	OR	P
rs7546390	203649423	Imputed	Intron1-2					0,57	7,06E-05		
rs7546599	203649668	Imputed	Intron1-2			0,57	5,38E-05	0,56	2,34E-05		
rs7514742	203649716	Imputed	Intron1-2			0,57	5,02E-05	0,58	6,1E-05	0,59	9,1E-05
rs12035565	203649878	Imputed	Intron1-2	0,60	9,91E-05	0,58	4,14E-05	0,57	3,05E-05	0,58	3,44E-05
rs11240731	203650336	Imputed	Intron1-2	0,60	8,89E-05	0,59	7,22E-05	0,58	4E-05	0,58	4,22E-05
rs11240733	203650592	Imputed	Intron1-2	0,59	4,7E-05	0,59	5,4E-05	0,59	5,06E-05	0,59	5,18E-05
rs10751449	203650784	Genotyped#	Intron1-2	0,59	4,55E-05	0,59	4,55E-05	0,59	4,55E-05	0,59	4,55E-05
rs10736845	203650786	Genotyped#	Intron1-2	0,54	3,13E-05	0,54	3,13E-05	0,54	3,13E-05	0,54	3,13E-05
rs10751450	203650945	Imputed	Intron1-2					0,63	9,04E-05	0,63	7,5E-05
rs10751451	203650978	Imputed	Intron1-2	0,58	3,18E-05	0,58	3,18E-05	0,58	3,12E-05	0,58	3,18E-05
rs10751452	203651030	Imputed	Intron1-2	0,58	3,18E-05	0,58	3,18E-05	0,58	3,12E-05	0,58	3,18E-05
rs35014299	203651063	Imputed	Intron1-2	0,58	3,18E-05	0,58	3,16E-05	0,58	3,12E-05	0,58	3,18E-05
rs546240990	203651063	Imputed	Intron1-2	0,59	4,89E-05	0,59	4,93E-05	0,58	4,28E-05	0,59	4,67E-05
rs34690356	203651160	Imputed	Intron1-2	0,58	3,18E-05	0,58	3,16E-05	0,58	3,12E-05	0,58	3,18E-05
rs35068094	203651731	Imputed	Intron1-2	0,54	3,32E-05	0,54	3,46E-05	0,54	3,21E-05	0,54	3,21E-05
rs11240734	203651824	Imputed	Intron1-2	0,58	2,86E-05	0,58	2,86E-05	0,58	2,86E-05	0,58	2,86E-05
rs1541252	203651927	Genotyped#	Exon2	0,59	6,03E-05	0,59	6,03E-05	0,59	6,03E-05	0,59	6,03E-05
rs1541253	203652040	Imputed	Exon2	0,58	2,86E-05	0,58	2,86E-05	0,58	2,86E-05	0,58	2,86E-05
rs1541254	203652140	Imputed	Exon2	0,57	2,36E-05	0,57	2,36E-05	0,58	2,93E-05	0,58	2,93E-05
rs1541255	203652141	Imputed	Exon2	0,57	2,36E-05	0,57	2,36E-05	0,58	2,93E-05	0,58	2,93E-05
rs1419114	203652444	Imputed	Exon2	0,57	2,45E-05	0,57	2,45E-05	0,58	2,93E-05	0,58	2,93E-05

Table A - 6: SNPs within intron2-3, intron 7-8, intron 10-11, intron 12-13, intron 13-14, intron 19-20, intron 20-21 and Exon 21 of *ATP2B4* that are associated with the trait of severe malaria (p -value $< 10^{-4}$) after logistic regression in this study (*SNParray, #SNParray and FRET-MCA, ° RefSeq: NM_001684).

SNP	BP	ORIGIN	Position in ATP2B4	keepHRM_ALL [°] (64)		keepHRM_AFR [°] (60)		keepGWAS_ALL [°] (63)		keepGWAS_AFR [°] (60)	
				OR	P	OR	P	OR	P	OR	P
rs4951369	203652683	Imputed	Intron2-3	0,57	2,41E-05	0,57	2,41E-05	0,59	5,04E-05	0,58	4,61E-05
rs4951070	203652698	Imputed	Intron2-3	0,57	2,41E-05	0,57	2,41E-05	0,59	5,04E-05	0,58	4,61E-05
rs4951370	203652905	Imputed	Intron2-3	0,57	1,46E-05			0,58	2,3E-05		
rs4951371	203652946	Imputed	Intron2-3	0,57	2,36E-05	0,57	2,36E-05	0,58	2,93E-05	0,58	2,93E-05
rs7547793	203653544	Imputed	Intron2-3	0,57	2,36E-05	0,57	2,36E-05	0,58	2,93E-05	0,58	2,93E-05
rs7555703	203653690	Imputed	Intron2-3	0,57	2,36E-05	0,57	2,36E-05	0,58	3,04E-05	0,58	2,93E-05
rs7535676	203653786	Imputed	Intron2-3	0,57	2,36E-05	0,57	2,36E-05	0,58	3,04E-05	0,58	2,93E-05
rs10900585	203654024	Genotyped#	Intron2-3	0,60	2,21E-05	0,60	2,21E-05	0,60	2,21E-05	0,60	2,21E-05
rs10900586	203654085	Imputed	Intron2-3	0,57	2,23E-05	0,57	2,23E-05	0,58	2,97E-05	0,58	2,86E-05
rs10793762	203654098	Imputed	Intron2-3	0,57	2,23E-05	0,57	2,23E-05	0,58	2,97E-05	0,58	2,86E-05
rs6692632	203654546	Imputed	Intron2-3	0,57	2,23E-05	0,57	2,23E-05	0,58	2,97E-05	0,58	2,86E-05
rs6658130	203654568	Imputed	Intron2-3	0,57	2,23E-05	0,57	2,23E-05	0,58	2,97E-05	0,58	2,86E-05
rs10594838	203654668	Imputed	Intron2-3	0,57	2,23E-05	0,57	2,23E-05	0,58	2,97E-05	0,58	2,97E-05
rs6594006	203654738	Imputed	Intron2-3	0,57	2,23E-05	0,57	2,23E-05	0,58	2,97E-05	0,58	2,97E-05
rs6594007	203654772	Imputed	Intron2-3	0,57	2,23E-05	0,57	2,23E-05	0,58	2,97E-05	0,58	2,97E-05
rs6594008	203654851	Imputed	Intron2-3	0,57	1,73E-05	0,57	1,73E-05	0,57	1,97E-05	0,57	1,97E-05
rs7539122	203655043	Imputed	Intron2-3	0,56	7,97E-06	0,56	7,51E-06	0,56	1,02E-05	0,56	9,01E-06
rs7551442	203655121	Imputed	Intron2-3	0,57	1,65E-05	0,57	1,65E-05	0,57	1,97E-05	0,57	1,96E-05
rs7551560	203655270	Genotyped#	Intron2-3					0,58	2,45E-05	0,58	2,45E-05
rs7554335	203655743	Genotyped#	Intron2-3	0,57	1,45E-05	0,57	1,45E-05	0,58	2,91E-05	0,58	2,91E-05
rs4951375	203655785	Imputed	Intron2-3	0,57	1,73E-05	0,57	1,73E-05	0,58	2,45E-05	0,58	2,45E-05
rs6594010	203655882	Imputed	Intron2-3	0,57	1,73E-05	0,57	1,73E-05	0,58	2,45E-05	0,58	2,45E-05
rs6594011	203655955	Imputed	Intron2-3	0,57	1,71E-05	0,57	1,68E-05	0,57	2,15E-05	0,57	2,2E-05
rs6696548	203656067	Imputed	Intron2-3	0,57	1,73E-05	0,57	1,73E-05	0,58	2,45E-05	0,58	2,45E-05
rs4347266	203656178	Imputed	Intron2-3	0,57	1,73E-05	0,57	1,73E-05	0,58	2,45E-05	0,58	2,45E-05
rs2365860	203656230	Genotyped*	Intron2-3	0,57	1,67E-05	0,57	1,67E-05	0,58	2,45E-05	0,58	2,45E-05
rs10900588	203656814	Imputed	Intron2-3	0,57	1,73E-05	0,57	1,78E-05	0,58	2,45E-05	0,58	2,45E-05
rs10900589	203656974	Genotyped*	Intron2-3	0,57	1,73E-05	0,57	1,73E-05	0,57	1,9E-05	0,57	1,9E-05
rs12122293	203657117	Imputed	Intron2-3	0,57	2,55E-05	0,57	1,96E-05	0,58	2,45E-05	0,58	2,45E-05
rs6692627	203657220	Imputed	Intron2-3	0,57	2,59E-05	0,57	1,99E-05	0,58	2,45E-05	0,58	2,45E-05
rs5780192	203657601	Imputed	Intron2-3	0,59	6,9E-05	0,60	9,31E-05				
rs2365859	203657688	Imputed	Intron2-3	0,58	3,76E-05	0,57	2,44E-05	0,58	2,45E-05	0,58	2,45E-05
rs2365858	203657749	Genotyped*	Intron2-3	0,57	1,91E-05	0,57	1,91E-05	0,57	1,4E-05	0,57	1,4E-05
rs4951377	203658471	Imputed	Intron2-3	0,58	3,34E-05	0,58	2,67E-05	0,58	2,45E-05	0,58	2,53E-05
rs4951074	203660781	Genotyped*	Intron2-3	0,58	8,37E-05	0,58	8,37E-05	0,55	5,79E-06	0,55	5,79E-06
rs4951381	203660838	Genotyped*	Intron2-3	0,54	4,52E-05	0,54	4,52E-05	0,53	1,83E-05	0,53	1,83E-05
rs3851298	203665047	Imputed	Intron2-3	0,52	1,42E-05	0,52	1,5E-05	0,51	8,72E-06	0,52	1,28E-05
rs2228445	203667409	Imputed	Intron2-3	0,53	2,41E-05	0,54	3,55E-05	0,54	2,92E-05	0,54	4,05E-05
rs4951081	203671877	Imputed	Intron7-8	0,54	1,5E-05	0,54	1,78E-05	0,55	2,89E-05	0,56	5,29E-05
rs6697384	203677565	Genotyped*	Intron10-11	0,54	2,06E-05	0,54	2,06E-05	0,53	1,31E-05	0,53	1,31E-05
rs6655965	203680492	Imputed	Intron12-13	0,53	1,67E-05	0,53	1,83E-05	0,54	2,75E-05	0,55	3,76E-05
rs12407860	203681503	Genotyped#	Intron13-14	0,54	3,35E-05	0,54	3,35E-05	0,54	3,35E-05	0,54	3,35E-05
rs9326596	203695360	Imputed	Intron19-20	0,58	9,45E-05						
rs12060019	203703164	Imputed	Intron20-21	0,55	7,97E-05						
rs7547344	203707804	Imputed	Intron20-21	0,55	3,63E-05						
rs955865	203709034	Imputed	Exon21	0,55	9,05E-05						
rs955866	203709068	Imputed	Exon21	0,55	9,04E-05						

Table A - 7: Results from logistic regressions of different haplotypes of the five published resistance SNPs (clear: minor allele, gray-shaded with dots: major allele).

rs10900585	rs2365860	rs10900589	rs2365858	rs4951074	keep GWAS			keep FRET		
					Freq	OR	p	Freq	OR	p
T	A	T	C	G	59,3	1,25	0.000527	59,8	1,23	0.00148
G	C	A	G	A	29,9	0,76	9e-05	29,9	0,76	6.8x10 ⁻⁵
G	C	A	G	G	5,3	1,18	0.258	5,3	1,19	0.239
G	A	T	C	A	3,4	0,93	0.674	3,4	0,96	0.805
G	A	T	C	G	1,3	1,00	0.999	1,2	1,05	0.869

9.2.2 Expression of *ATP2B4* in Ghanaian whole blood

Table A - 8: Results from Shapiro-Wilk test with the values of the relative expression of total *ATP2B4* and its two known splice variants in the samples of different genotype groups (rs10900585).

	genotype	ATP2B4_compl	ATP2B4_common	ATP2B4_rare
Shapiro-Wilk test	GG	0.24	0.14	0.03
(p-value)	GT	2.06 x 10 ⁻³	0.1	1.18 x 10 ⁻⁴
	TT	0.26	2.32 x 10 ⁻⁵	0.07

Table A - 9: Summarized information about the determined data of the relative expression of total *ATP2B4* and its two known splice variants in the samples of different genotype groups (rs10900585).

genotype	ATP2B4_compl			ATP2B4_common			ATP2B4_rare		
	GG	GT	TT	GG	GT	TT	GG	GT	TT
n	42	84	66	45	80	66	31	65	40
mean	0.202	0.210	0.212	0.236	0.243	0.249	0.466 x 10 ⁻³	0.502 x 10 ⁻³	0.489 x 10 ⁻³
sd	0.048	0.048	0.046	0.058	0.056	0.066	0.200 x 10 ⁻³	0.257 x 10 ⁻³	0.217 x 10 ⁻³
min	0.055	0.112	0.130	0.055	0.124	0.143	0.195 x 10 ⁻³	0.155 x 10 ⁻³	0.157 x 10 ⁻³
Q1	0.167	0.177	0.178	0.209	0.206	0.206	0.311 x 10 ⁻³	0.326 x 10 ⁻³	0.333 x 10 ⁻³
median	0.204	0.205	0.206	0.235	0.239	0.240	0.401 x 10⁻³	0.475 x 10⁻³	0.434 x 10⁻³
Q3	0.227	0.235	0.241	0.270	0.276	0.269	0.613 x 10 ⁻³	0.581 x 10 ⁻³	0.683 x 10 ⁻³
max	0.285	0.399	0.322	0.338	0.442	0.535	0.999 x 10 ⁻³	1.372 x 10 ⁻³	1.09 x 10 ⁻³
Kruskal-Wallis test (p-value)	0.73			0.85			0.86		

Table A - 10: Results from Shapiro-Wilk test with the ratios of the relative expression of the two known splice variants and the total expression of *ATP2B4* in the samples of different genotype groups (rs10900585).

	genotype	ATP2B4_common/ ATP2B4_compl	ATP2B4_rare/ ATP2B4_compl
Shapiro-Wilk test	GG	1.59 x 10 ⁻⁵	0.19
(p-value)	GT	5.38 x 10 ⁻⁹	1.69 x 10 ⁻⁷
	TT	2.78 x 10 ⁻⁸	0.04

Table A - 11: Summarized information about the determined data of the relative expression of the two known splice variants and the total expression of *ATP2B4* in the samples of different genotype groups (rs10900585).

genotype	ATP2B4_common/ ATP2B4_compl			ATP2B4_rare/ ATP2B4_compl		
	GG	GT	TT	GG	GT	TT
n	42	80	66	28	64	40
mean	1.165	1.170	1.180	2.146×10^{-3}	2.430×10^{-3}	2.319×10^{-3}
sd	0.178	0.211	0.205	0.749×10^{-3}	1.405×10^{-3}	1.020×10^{-3}
min	0.955	0.883	0.959	0.939×10^{-3}	0.763×10^{-3}	0.884×10^{-3}
Q1	1.061	1.067	1.045	1.485×10^{-3}	1.624×10^{-3}	1.578×10^{-3}
median	1.126	1.111	1.128	2.150×10^{-3}	2.137×10^{-3}	2.057×10^{-3}
Q3	1.216	1.221	1.205	2.780×10^{-3}	2.713×10^{-3}	3.193×10^{-3}
max	1.782	2.250	1.963	3.943×10^{-3}	8.032×10^{-3}	4.699×10^{-3}
Kruskal-Wallis test (p-value)	0.96			0.93		

9.2.3 Parasite proliferation referring to the rs10900585 genotype

Table A - 12: Summarized information about the development of parasite proliferation in erythrocytes from donors with different rs10900585 genotypes.

genotype	proliferation rate (parasitemia 120 h/parasitemia 48 h)		
	GG	GT	TT
n	10	10	9
mean	18.18	20.97	26.47
sd	4.92	4.69	3.24
min	10.25	11.04	22.17
Q1	16.01	18.69	23.63
median	17.10	21.11	27.51
Q3	20.93	25.00	29.10
max	26.71	26.12	30.92
Kruskal-Wallis test (p-value)	$2,67 \times 10^{-3}$		
Dunn's test for pairwise comparison (p-value)	GG vs. GT $6,96 \times 10^{-4}$	GG vs. TT 0.26	GT vs. TT 0.02

Table A - 13: Summarized information about the development of parasite proliferation in erythrocytes from donors with different rs10900585 genotypes (replication).

genotype	proliferation rate (parasitemia 144 h/parasitemia 48 h)		
	GG	GT	TT
n	10	10	9
mean	26.41	34.87	34.93
sd	7.86	6.89	7.43
min	15.63	22.87	26.21
Q1	20.90	32.85	29.07
median	23.61	36.69	33.78
Q3	31.80	38.46	41.32
max	40.35	46.24	46.06
Kruskal-Wallis test (p-value)	0.04		
Dunn's test for pairwise comparison (p-value)	GG vs. GT 0.02	GG vs. TT 0.03	GT vs. TT 0.98

9.3 Figures

9.3.1 Computational fine mapping

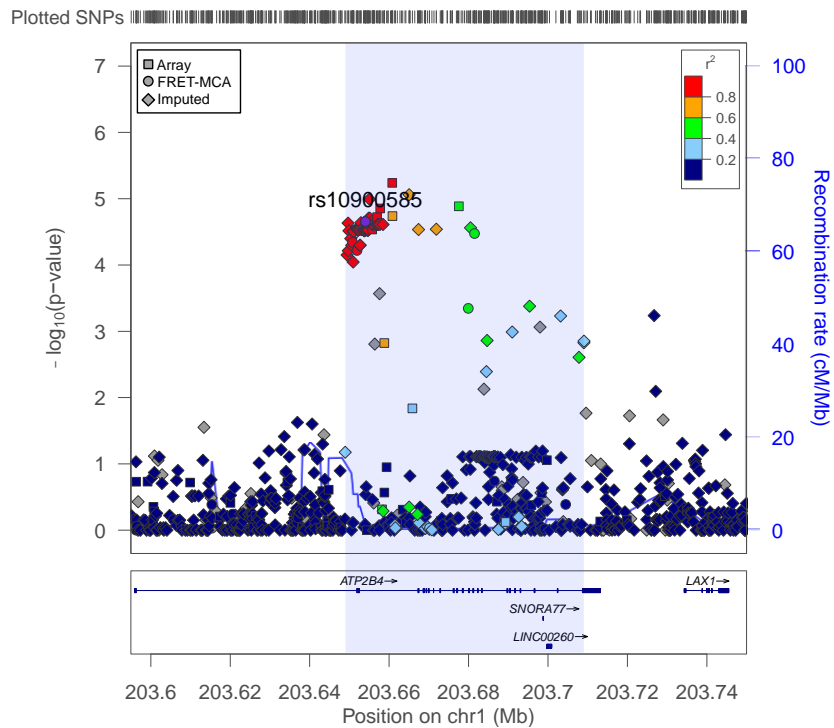


Figure A - 1: Regional association plot of the resistance locus generated from the dataset keepArray_ALL. The origin of genotype data is coded in the shape of the data point and the LD with the index SNP rs10900585 in its color.

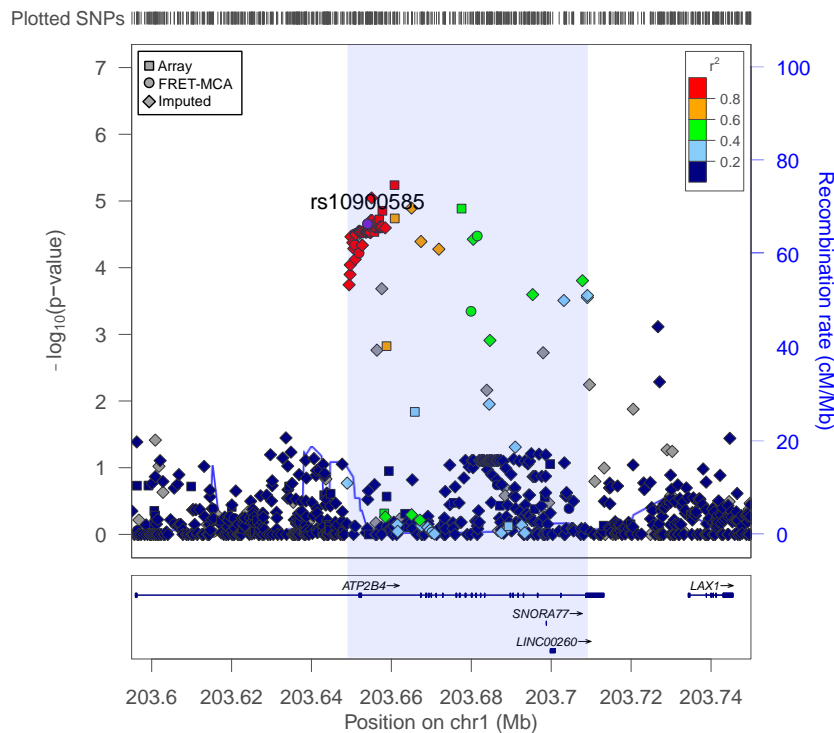


Figure A - 2: Regional association plot of the resistance locus generated from the dataset keepArray_AFR. The origin of genotype data is coded in the shape of the data point and the LD with the index SNP rs10900585 in its color.

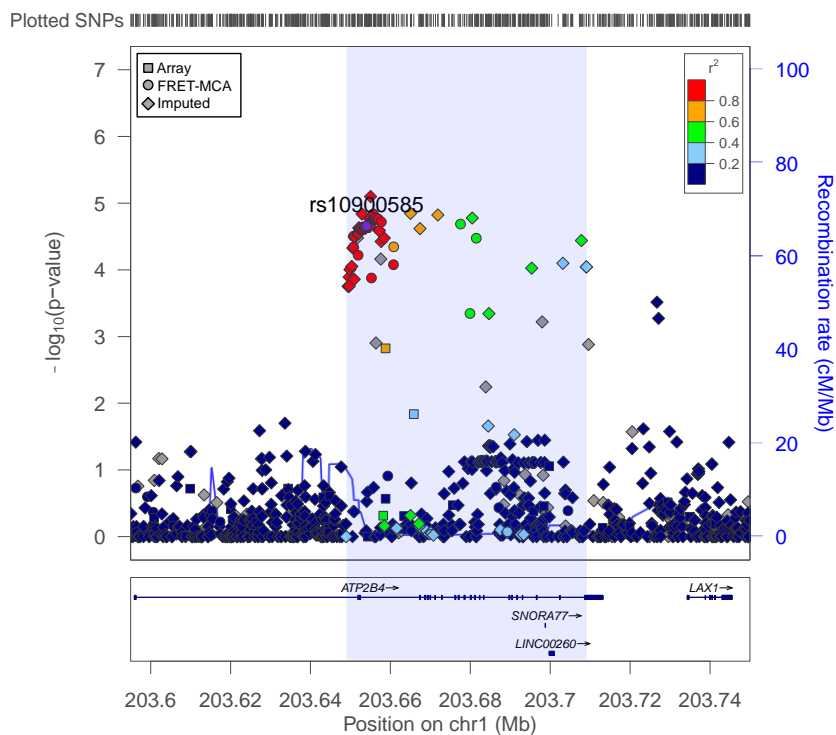


Figure A - 3: Regional association plot of the resistance locus generated from the dataset keepFRET-MCA_ALL. The origin of genotype data is coded in the shape of the data point and the LD with the index SNP rs10900585 in its color.

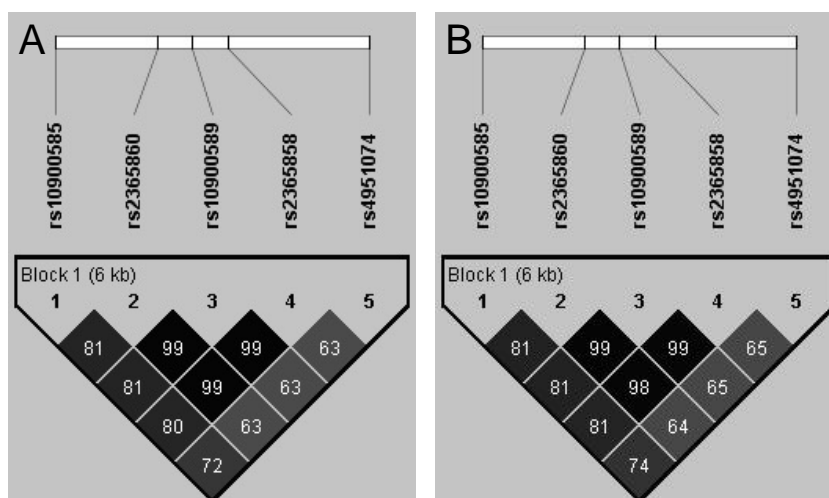


Figure A - 4: LD of the resistance SNPs. r^2 -values were calculated from genotype information from (A) the array data or (B) FRET-MCA.

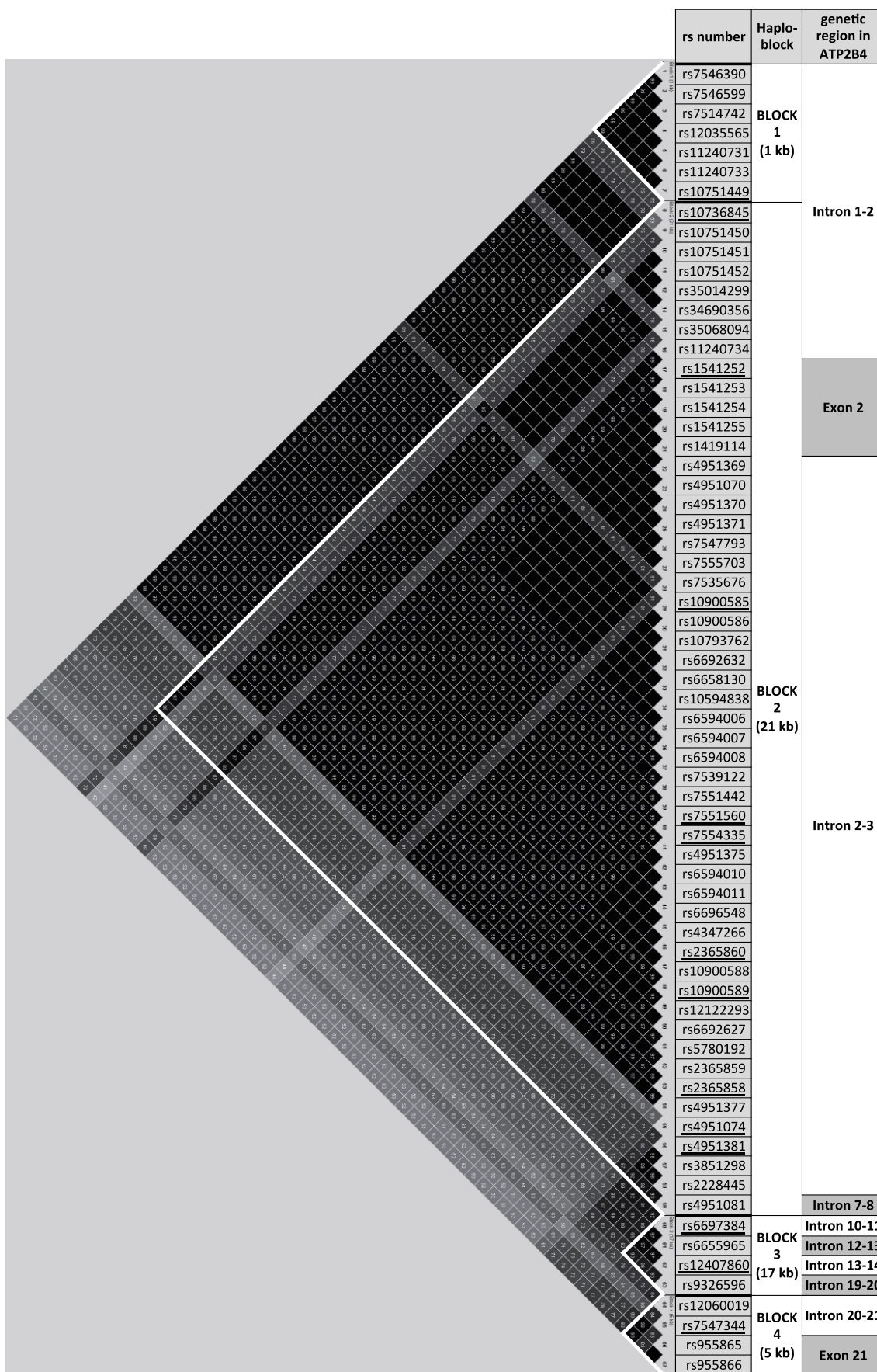


Figure A - 5: LD structure and localization of all associated SNPs (p -value $< 10^{-4}$) within *ATP2B4*. Haploblocks are indicated by white triangles and SNPs which have already been genotyped via array or FRET-MCA are underlined.

9.3.2 4C-Seq

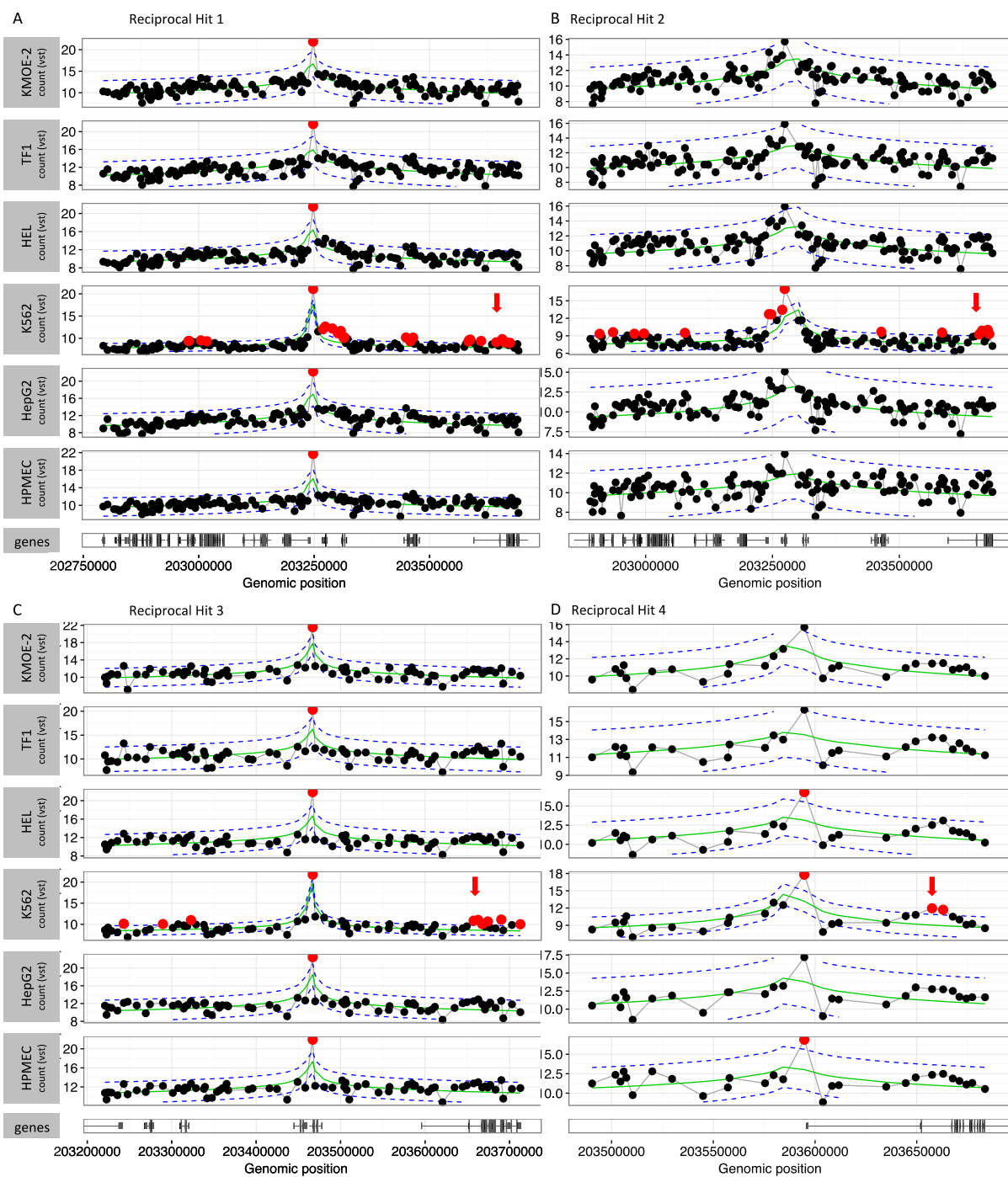


Figure A - 6: FourCSeq analysis of sequenced chromosomal fragments interacting with (A) Hit 1, (B) Hit 2, (C) Hit3 and (D) Hit 4 in fixed nuclei of six selected cell lines (reciprocal experiments).

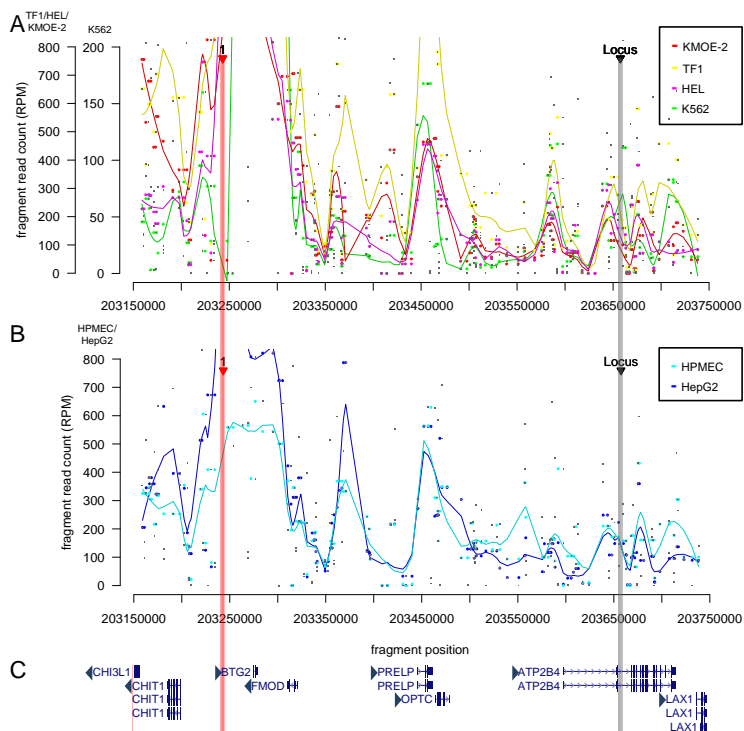


Figure A - 7: Basic4Cseq analysis of sequenced chromosomal fragments interacting with VP_Hit4 in fixed nuclei of six selected cell lines.

The figure shows (A) results from four erythroid cell lines, (B) results from HPMEC and HepG2 and (C) genes that are located in the affected region of chromosome 1 (image created on <http://genome.ucsc.edu>). The red line indicates the chromosomal position of VP_Hit1. The chromosomal position of the fragment that contains the resistance locus is indicated by the gray line.

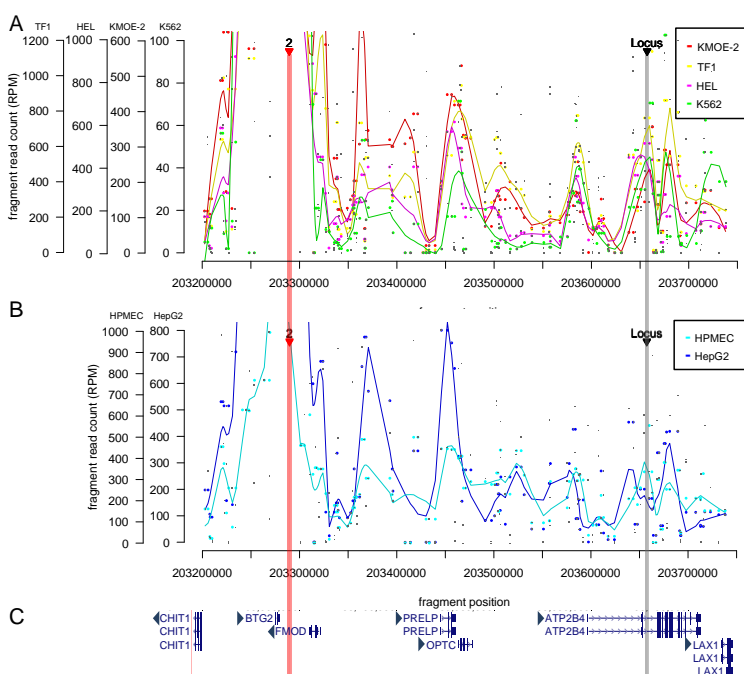


Figure A - 8: Basic4Cseq analysis of sequenced chromosomal fragments interacting with VP_Hit4 in fixed nuclei of six selected cell lines.

The figure shows (A) results from four erythroid cell lines, (B) results from HPMEC and HepG2 and (C) genes that are located in the affected region of chromosome 1 (image created on <http://genome.ucsc.edu>). The red line indicates the chromosomal position of VP_Hit2. The chromosomal position of the fragment that contains the resistance locus is indicated by the gray line.

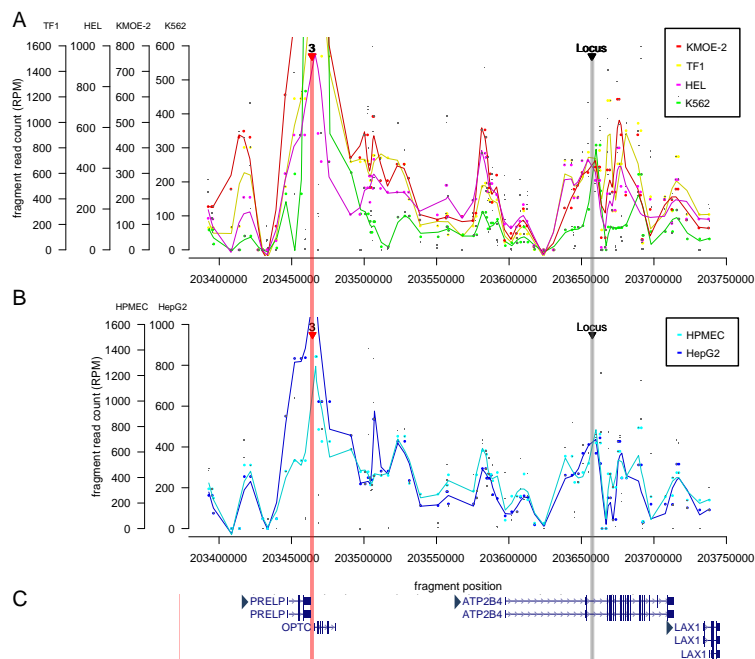


Figure A - 9: Basic4Cseq analysis of sequenced chromosomal fragments interacting with VP_Hit4 in fixed nuclei of six selected cell lines.

The figure shows (A) results from four erythroid cell lines, (B) results from HPMEC and HepG2 and (C) genes that are located in the affected region of chromosome 1 (image created on <http://genome.ucsc.edu>). The red line indicates the chromosomal position of VP_Hit3. The chromosomal position of the fragment that contains the resistance locus is indicated by the gray line.

9.3.3 Expression of *ATP2B4* in Ghanaian whole blood

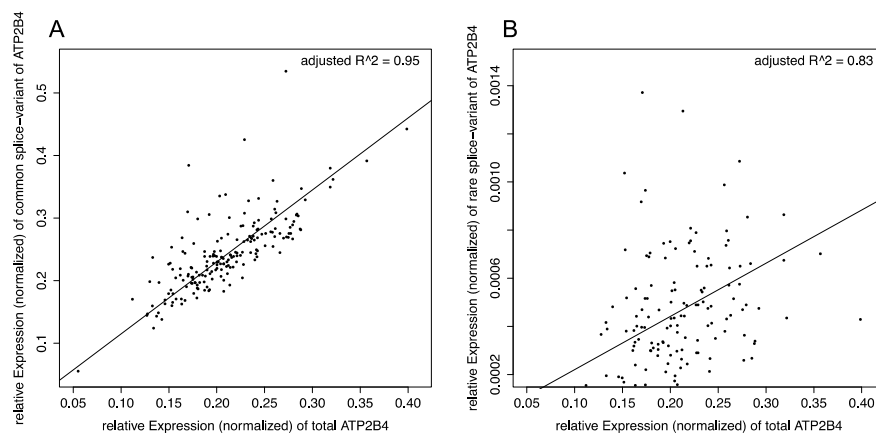


Figure A - 10: Linear regression of (A) the relative expression of the common splice variant of *ATP2B4* with relative expression of total *ATP2B4* and (B) the relative expression of the rare splice variant of *ATP2B4* with relative expression of total *ATP2B4* proliferation.

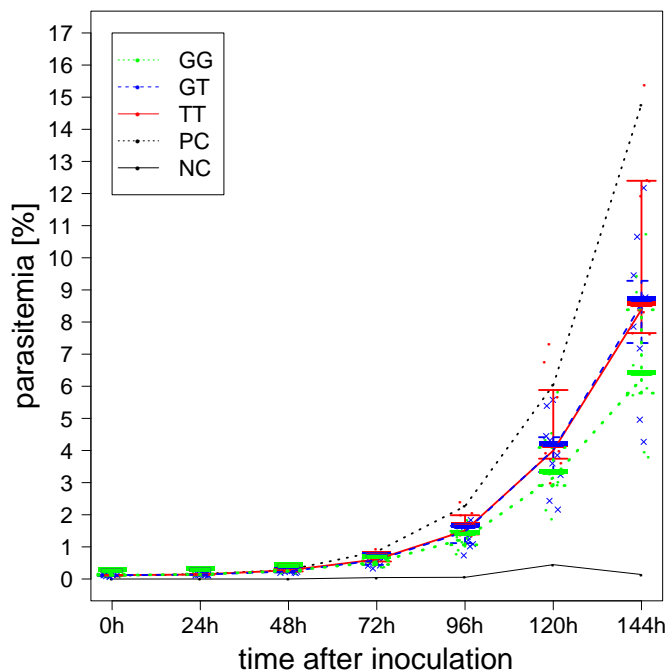


Figure A - 11: Parasitemia in infected erythrocytes of donors with different rs10900585 genotypes (replication). The graphs show the development of the median values of the detected parasitemia in the three genotype groups. The values of different groups are color-coded. Error bars correspond to the respective interquartile ranges. Values of samples are shown as color-coded dots. The graphs of the positive control (PC) and the negative control (NC) refer to the mean value of the triplicates at respective time points.

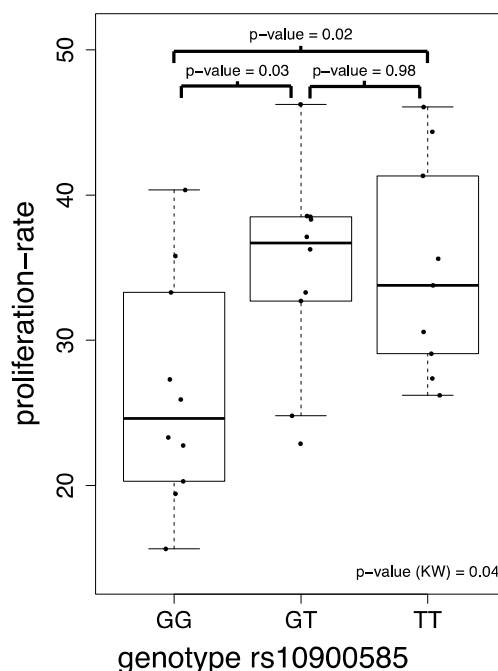


Figure A - 12: Parasite proliferation in erythrocytes of donors with different rs10900585 genotypes (replication). Data points show means of the triplicates of each culture from erythrocytes of the different donors. The boxplots visualize the 25th percentile (bottom of the box), the median (band within the box) and the 75th percentile (top of the box) of the values of each genotype group. The length of the whiskers is determined by the last value within the respective 1.5-fold interquartile range. The p-value (KW) refers to the performed Kruskal-Wallis test. All other p-values result from Dunn's test for pairwise comparison of the indicated groups.

Acknowledgement

I would like to thank all the people who were officially involved in the realization of this dissertation: Rolf Horstmann for giving me the possibility to conduct my experimental work in his lab group and for being part of my examination committee, Julia Kehr and Christian Timmann for evaluating this dissertation and in case of Julia Kehr also for being a member of my examination committee, as well as Tim-Wolf Gilberger, and Tobias Spielmann and Anna Bachmann for being my co-supervisors and for always having an open ear and valuable advices, as well as Thomas Jacobs.

I am also grateful to the numerous lab groups at the BNI who supported me during my experiments by sharing their expertise, their ideas and their laboratories with me: AG Spielmann, AG Gilberger, AG Tannich, AG May, AG Jacobs and especially “my” group, the Department of Molecular Medicine. Of all, I would like to thank two people in particular: Gerd for his efforts to help me prepare and successfully accomplish my sampling campaign in Ghana, not least his racing-car driver like performance to save my samples, and Kathrin for her valuable guidance not only through scientific challenges.

For the help and support I experienced during my time in Ghana I would like to thank Shirley Owuso-Ofori and the amazing staff of the Blood Donor Centre of the Komfo Anokye Teaching Hospital in Kumasi, as well as Albert Dompseh for being so generous to provide me the lab space. Furthermore, I would like to thank everyone at the KCCR for their support and for making miracles happen by acquiring liquid nitrogen when there was technically none available. In addition to their professional help, I would like to thank Ingrid Sobel and Micha Nagel for numerous enjoyable occasions with BBQ, beer, cake and brunch. I had an amazing time.

Last and foremost, I would like to thank my family and friends:

You make me who I am.

Eidesstattliche Erklärung

Declaration on oath

Hiermit erkläre ich an Eides statt, dass die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg,

Datum/date

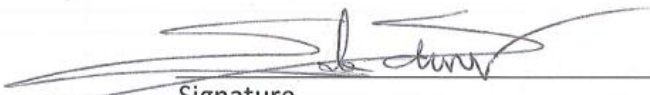
Christina Strauß

Unterschrift/signature

Language Certificate

I am a native English speaker, have read the present Dissertation of
Christina Strauß, and hereby confirm that it complies with the rules of the English language.

Münster, 24.04.2017
City, Date


Signature

Prof. E. Scherer
Institut für Mineralogie
Universität Münster
Corrensstraße 24
48149 Münster