# Dissertation

Submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
Andris Schulz
Diploma-Biologist
born in Braunschweig

Oral examination:……

# Studies on variations in host genes coding for *Plasmodium* liver stage binding partners and the susceptibility to malaria

Referees: Prof. Dr. Michael Lanzer

Prof. Dr. Kai Matuschewski

A gold bug -
I hurl into the darkness
and feel the depth of night.

*Takahama Kyoshi*

**Acknowledgements**

Zuerst möchte ich Prof. Michael Lanzer für die Bereitschaft das Erstgutachten zu übernehmen und Prof. Kai Matuschewski für die Rolle als Zweitgutachter danken.

Steffen Borrmann danke ich besonders für die Möglichkeit dieser Arbeit. Vor allem durch Deine großzügige Unterstützung konnte ich vieles Erleben und Erlernen, das nicht selbstverständlich ist.

I also want to thank Tom Williams, Alex Macharia and all the people in the team who helped me a lot in Kilifi. I really enjoyed working in your nice team. Moreover, my special thanks go to all the patients, who participated in the studies and to the busy field workers who gathered all the information.

Ich bedanke mich auch bei Ann-Kristin Müller, Kai Matuschewski und Freddy Frischknecht für die Ideen und Hilfe bei meinem Projekt. Dank gilt auch den Kollegen in der AG Müller, besonders Sabrina, die mir viel geholfen haben und Geduld bewiesen, wenn ich wieder die Sterilbank blockierte.

Besonderer Dank gilt auch Lars Beckmann für die Hilfe bei dem theoretischen und mathematischen Hintergrund dieser Arbeit.

Ich danke Georgina Montagna für die Unterstützung bei der Zellkultur.

Bedanken möchte ich mich auch herzlich bei allen meinen Kollegen, die mir oft das Leben und Arbeiten erleichtert haben:

Anja, vielen Dank für die Zeit und Mühe, die Du investiert hast um noch schnell etwas für mich zu bestellen und nach Kilifi zu schicken. Es hat Spaß gemacht mit Dir zu arbeiten und zu reden.

Judith, vielen Dank, dass man in Deiner Schublade häufig alles schnell wiedergefunden hat. Ich fand unsere Gespräche über die Welt auch sehr inspirierend, besonders unsere gemeinsamen Reisen.

Ifey, Dir danke ich für die vielen interessanten Gespräche. Du warst ein Sonnenschein im Labor- auch bei trüben Tagen.

Elise, bei Dir möchte ich mich für Dein immer offenes Ohr bedanken und für Deine Hilfe und Anteilnahme.

Außerdem möchte ich mich bei allen bedanken, die zeitweise meine Wege gekreuzt haben: Henrike, Barbara, Yvonne, Katharina, Johannes und Dirk- vielen Dank für die schöne Zeit mit Euch.

Meinen Eltern gilt der größte Dank, da ich ohne Ihre Mühen wohl nicht so weit gekommen wäre. Vielen Dank für Eure durchgehende Unterstützung! Meinem Bruder danke ich für die immer schnelle Hilfe, wenn die Computer nicht so wollten wie ich.

Meiner Freundin Petra danke ich für die Geduld und Beistand während der letzten Jahre.

Bedanken möchte ich mich bei allen Freunden, die trotz meiner wenigen Lebenszeichen und der geringen Zeit, die ich mit ihnen verbracht habe, meine Freunde geblieben sind.

## Table of content


**Acknowledgements** **I**

**Abbreviations** **IX**

**Summary** **1**

**Zusammenfassung** **2**

**1. Introduction** **3**

1.1 Malaria: past and present 3

1.2 Characteristics of *Plasmodium* 5

1.3 Clinical features of malaria 7

1.4 Immunology of malaria 7

1.5 Hepatocyte stage proteins and host-parasite interactions 11

1.6 Genetic host factors influencing protection against malaria 13

1.7 Effects of evolutionary selection on the human genome by malaria 16

1.8 Association studies and signatures of selection 17

1.9 Association and cohort studies 19

**2. Aim of study** **21**

**3. Materials and methods** **22**

**3.1 Materials** **22**

3.1.1 Equipment 22

3.1.2 Chemicals 23

3.1.3 Consumables 24

3.1.4 Buffers, media, solutions 25

3.1.5 Enzymes, ladder, antibodies, vector 26

3.1.6 Kits 26

3.1.7 Oligonucleotides 27

3.1.8 The Mild and Severe Malaria Cohorts 31

3.1.8.1 Mild Malaria Cohort dynamic follow-up study 31

3.1.8.2 Severe Malaria Cohort study 32

3.1.9 Databases 32

3.1.10 Software 33

**3.2 Methods** **34**

**3.2.1 Molecular biology** **34**

3.2.1.1 Isolation and purification of nucleic acids 34

3.2.1.1.1 Isolation of genomic DNA 34

3.2.1.1.2 Isolation of RNA 34
3.2.1.2 Purification of nucleic acids 34
3.2.1.2.1 Ethanol precipitation 34
3.2.1.2.2 Purification of GenomiPhi samples 34
3.2.1.2.3 Purification of sequencing products 35
3.2.1.2.4 Purification of PCR products 35
3.2.1.3 Determination of DNA/ RNA concentration 36
3.2.1.4 Amplification of nucleic acids 36
3.2.1.4.1 Cleaving of gDNA 36
3.2.1.4.2 Transcription of RNA to cDNA 36
3.2.1.4.3 Whole Genome Amplification 36
3.2.1.4.4 Polymerase Chain Reaction 37
3.2.1.4.4.1 PCR for *APOA1* re-sequencing 37
3.2.1.4.4.2 Long Range PCR for SNaPshot assay 38
3.2.1.4.4.3 PCR for rs670 RFLP assay 39
3.2.1.5 Gel electrophoresis 40
3.2.1.6 MspI Restriction Fragment Length Polymorphism 40
3.2.1.7 Sequencing of PCR products 40
3.2.1.8 Primer extension assay with SNaPshot Multiplex Kit 42
**3.2.2 Cell culture 44**
3.2.2.1 Hepatoma cell culture 44
3.2.2.2 Determination of cell concentration 44
3.2.2.3 Freezing and thawing of hepatocytes 45
3.2.2.4 Transfection of hepatocytes 45
3.2.2.5 APOA1 overexpression assay 46
3.2.2.5.1 Isolation of *P. berghei* sporozoites 46
3.2.2.5.2 Infection of hepatocytes with *P. berghei* sporozoites 46
3.2.2.6 APOA1 overexpression assay analysis 47
3.2.2.6.1 Isolation of RNA for real-time PCR 47
3.2.2.6.2 Quantification of cDNA using Real-Time PCR 47
3.2.2.6.3 IFA staining and microscopy 48

**3.2.3 Bioinformatical methods** **49**
3.2.3.1 Database screening 49
3.2.3.2 DNA sequence analysis and manipulation 49
3.2.3.3 Calculation of tag SNPs 50
3.2.3.4 Primer Design 50
3.2.3.4.1 Primer Design for PCR and real-time PCR 50
3.2.3.4.2 Primer Design for SNaPshot Multiplex assay 50
3.2.3.5 Genotype data validation 51
3.2.3.6 Analysis of Mild Malaria Cohort genotype data 52
3.2.3.6.1 Poisson regression analysis 52
3.2.3.6.2 Multiple Failure-Time Analysis 54
3.2.3.7 Power Calculation for the Severe Malaria Cohort genotyping 55
3.2.3.9 Association analysis of rs670 in the Severe Malaria Cohort 55
3.2.3.9.1 Association analysis of rs670 using $\chi^2$ –test 56
3.2.3.9.2 Association analysis of rs670 using logistic regression 56
3.2.3.10 Calculation of linkage disequilibrium and haplotype blocks 57
3.2.3.11 Reconstruction of haplotypes 59
3.2.3.12 Screening for signatures of recent selection 59
3.2.3.12.1 Detection of extended haplotypes using Sweep software 59
3.2.3.12.2 Detection of signatures of selection using Haplotter software 60
**4. Results**
**4.1 Identification of SNPs in candidate genes *APOA1, FABP1* and *SNAPAP*** **62**
4.1.1 Re-sequencing of *APOA1* confirmed 11 SNPs 62
4.1.2 Selection of 11 SNPs and 2 tagSNPs in *FABP1* by database screening 63
4.1.3 Selection of 12 SNPs in *SNAPAP* by database screening 64
**4.2 Genotyping of the Mild Malaria Cohort** **68**
4.2.1 Genotyping Results of the Mild Malaria Cohort Study 68
4.2.1.1 Alleles of 10 SNPs in *APOA1* 70
4.2.1.2 Alleles of 11 SNPs in *FABP1* 70
4.2.1.3 Alleles of 12 SNPs in *SNAPAP* 71

**4.3 Genotype-phenotype association analyses using single SNP approach** **72**
4.3.1 Reduced risk in rs670 G carriers in Poisson regression analyses 74
4.3.2 Multiple failure-time analysis revealed a prolonged time between infections by rs670 75
4.3.3 No association of rs670 with parasite density or epistasis with HbAS 77
**4.4 Association of SNP rs670 with risk of severe malaria** **79**
4.4.1 Power calculation for design of genotyping Severe Malaria Cohort 79
4.4.2 rs670 genotyping of the Severe Malaria Cohort 79
4.4.3 Lowered risk for severe malaria in rs670 G carriers 81
**4.5 Screening for signatures of selection in *APOA1, FABP1* and *SNAPAP*** **82**
4.5.1 rs670 A allele is present at low to intermediate frequencies around the world 83
4.5.2 Haplotype blocks in *APOA1, FABP1* and *SNAPAP* 84
4.5.3 Haplotype analysis of rs670 revealed G as ancestral allele 88
4.5.4 No signatures of selection in the genomic region of *APOA1, FABP1* and *SNAPAP* 90
**4.6 Influence of APOA1 on *P. berghei* liver cell stage development** **94**
4.6.1 IFA staining revealed increased size of *P. berghei* liver stages in APOA1 overexpressing hepatocytes 94
4.6.2 Quantification using real-time PCR indicates elevated Pb18s in APOA1 overexpressing hepatocytes. 97
**5. Discussion** **100**
**References** **109**

**Figures**


**Figure 1: The spatial distribution of *P. falciparum* malaria endemicity. 4**

**Figure 2: Life cycle of human *Plasmodium.* 6**

**Figure 3: Population indices of immunity to *P. falciparum.* 8**

**Figure 4: *P. falciparum* life cycle with host immune responses. 10**

**Figure 5: Time scales for signatures of selection. 18**

**Figure 6: Primer extension with SNaPshot Multiplex Kit. 42**

**Figure 7: Gene structure of *APOA1* and identified SNPs. 65**

**Figure 8: Gene structure of *FABP1* and selected SNPs. 66**

**Figure 9: Gene structure of *SNAPAP* and selected SNPs. 67**

**Figure 10: Representative examples of SNaPshot Multiplex assay electropherograms for *APOA1, FABP1* and *SNAPAP.* 69**

**Figure 11: Frequency histogram of observations in the *APOA1* dataset by months in children below 10 years. 73**

**Figure 12: Incidence rate ratios for malaria of rs670 AA vs. GG and HbAA vs. HbAS by age. 75**

**Figure 13: Kaplan-Meier failure estimates in children 1-5 years by genotypic group. 76**

**Figure 14: Dot plot of asexual *P. falciparum* parasites in children from 1-5 years by rs670 (a) and HbAS (b) genotypes. 78**

**Figure 15: Rs670 MspI RFLP assay. 80**

**Figure 16: Odds ratio of rs670 and *HBB* alleles in children <5 years. 82**

**Figure 17: World map of rs670 A allele frequencies. 84**

**Figure 18: Distribution of rs670 genotypes in children 1-5 years from the Mild Malaria Cohort. 85**

**Figure 19: LD and haplotype blocks in *APOA1, FABP1* and *SNAPAP* 88**

**Figure 20: Haplotype bifurcation diagrams of *APOA1*. 89**

**Figure 21: Screen for signatures of selection in *HBB* and *APOA1* 92**

**Figure 22: Screen for signatures of selection in *FABP1* and *SNAPAP* 93**

**Figure 23: Number of *P. berghei* liver cell stages in WT and ApoTrans. 95**

**Figure 24: Size of *P. berghei* liver cell stages in WT and ApoTrans. 96**

**Figure 25: Pb18s in Huh7 (a,c) and Hepa 1.6 (b,d) cell lines at 24, 48 and 60 hours post invasion 98**

**Figure 26: Pb18s ratio of ApoTrans/ WT in Huh7 and Hepa 1.6 99**

**Figure 27: Influence of rs670 on *Plasmodium* liver cell stages** **108**

**Tables**

**Table1: Host factors associated with malaria.** **15**
**Table 2: Primer for *APOA1* re-sequencing.** **27**
**Table 3: Primer for Long Range PCR.** **27**
**Table 4 a: Primer for *APOA1* SNaPshot Multiplex assay.** **28**
**Table 4 b: Primer for *FABP1* SNaPshot Multiplex assay.** **29**
**Table 4 c: Primer for *SNAPAP* SNaPshot Multiplex assay.** **29**
**Table 5: Primer for PCR product used in rs670 RFLP assay.** **30**
**Table 6: Primer for Real-Time PCR.** **30**
**Table 7: Results of *APOA1* genotyping in Mild Malaria Cohort with SNaPshot Multiplex assay.** **70**
**Table 8: Results of *FABP1* genotyping in Mild Malaria Cohort with SNaPshot Multiplex assay.** **71**
**Table 9: Results of *SNAPAP* genotyping in Mild Malaria Cohort with SNaPshot Multiplex assay.** **72**
**Table 10: Number of subjects and observations from the genotyped Mild Malaria Cohort.** **73**
**Table 11: Covariates influencing variability in *P. falciparum* incidence in children 1-5 years.** **77**
**Table 12: Distribution of rs670 in the genotyped Severe Malaria Cohort.** **80**
**Table 13: $\chi^2$ –test of rs670 genotypes in children <5 years.** **81**

**Abbreviations:**

| | |
|---|---|
| APOA1 | apolipoprotein A1 |
| bp | base pairs |
| °C | degrees Celsius |
| CAD | coronary artery disease |
| cDNA | complementary DNA |
| CI | confidence interval |
| CSP | *P. falciparum* circumsporozoite protein |
| DNA | Deoxyribonucleic acid |
| EDTA | Ethylendiaminotetraacetic acid |
| FABP1 | fatty acid binding protein 1 |
| gDNA | Genomic DNA |
| h | hour |
| HWE | Hardy-Weinberg equilibrium |
| Kb | kilobase pairs |
| MAF | minor allele frequency |
| Mb | megabase pairs |
| ORF | open reading frame |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| pi | post invasion |
| PVM | parasitophorous vacuole membrane |
| PV | parasitophorous vacuole |
| RAS | radiation-attenuated sporozoites |
| RNA | Ribonucleic acid |

| | |
|---|---|
| rRNA | Ribosomal RNA |
| RT | Room temperature |
| RT-PCR | real time polymerase chain reaction |
| SB-RI | scavenger-receptor-BI |
| SD | Standard deviation |
| SNAPAP | SNARE-associated protein-associated protein |
| SNP | single nucleotide polymorphism |
| TAE | Tris/Acetic acid/EDTA |
| TE | Tris/EDTA |
| UV | ultraviolet |
| V | Volt |
| WGA | whole genome amplification |
| WT | wild type |
| µ | micro (1 x 10-6) |
| µg | microgram |
| µl | microliter |
| µM | micromolar |

## Summary

Liver stage development is the first obligate replication phase of the malarial parasite *Plasmodium* in the mammalian host. Rapid intracellular growth and differentiation of a single parasite in the hepatocyte result in release of thousands of pathogenic blood-stage merozoites. Parasite-host interactions during this clinically silent pre-erythrocytic phase and their roles in disease outcome of *Plasmodium* infections remain largely unknown.

It was previously shown that essential parasite-encoded proteins at the pathogen/host interface cooperate with host proteins, such as apolipoprotein A1 (APOA1), Fatty Acid Binding Protein (FABP1) and SNAP-associated Protein (SNAPAP), in the liver.

In my study I used high-density screening of genetic variants in patient cohorts in a malaria-endemic area to examine the influence of genetic variations in the genes, coding for these liver stage ligands, on *P. falciparum* incidence and disease outcome.

My results revealed that a single nucleotide polymorphism (SNP) rs670 (G/A) in the 5' untranslated region of human *APOA1* affects both incidence and outcome of *P. falciparum* malaria. Non-immune children below 5 years carrying the G allele had a 50% lower risk of re-infections and significantly reduced odds of severe malaria. It is well documented that the sickle cell trait (HbAS) and other established polymorphisms protect against disease caused by pathogenic *Plasmodium* blood stages. Further, I performed *in vitro* experiments that suggest an enhanced development of *P. berghei* late liver cell stages due to high level of APOA1.

In this study, I establish that additional selection for alleles affecting pre-erythrocytic parasite development may play previously unrecognised roles in malaria disease outcome. My results verify the relevance of clinically silent pathogen expansion phases, such as maturation of *Plasmodium* liver stages, for the clinical epidemiology of infectious diseases.

## Zusammenfassung

Das Leberzellstadium ist die erste notwendige Replikationsphase des *Plasmodium*-Parasiten im Säugetier. Schnelles intrazelluläres Wachstum und Ausdifferenzierung eines einzelnen Parasiten in der Leberzelle führt zur Freisetzung von tausenden Merozoiten, die anschließend Erythrozyten infizieren. Interaktionen zwischen Parasit und Wirt während dieser klinisch unauffälligen Phase sind größtenteils unbekannt.
Es wurde gezeigt, dass lebenswichtige Proteine des Parasiten mit Proteinen des Wirts, wie Apolipoprotein A1 (APOA1), Fatty Acid Binding Protein (FABP1) und SNAP-assozierten Protein (SNAPAP), kooperieren.
In meiner Studie untersuchte ich in Patienten aus malaria-endemischen Gebieten genetische Varianten in den drei Genen, die diese Leberstadien-spezifischen Liganden kodieren, und analysierte ihren Einfluss auf die Häufigkeit von *P. falciparum* Infektionen und den Krankheitsverlauf.
Meine Ergebnisse zeigen, dass der Einzelnukleotidpolymorphismus rs670 (G/A), im 5' untranslatierten Bereich von *APOA1,* die Häufigkeit und den Ausgang von *P. falciparum* Malaria beeinflusst. Nicht immune Kinder unter 5 Jahren, die das G Allel haben, hatten ein 50% geringeres Risiko einer Reinfektion und eine signifikant reduziertes Risiko für schwere Malaria. Es ist beschrieben, dass genetische Eigenschaften wie Sichelzellanämie (HbAS) und andere Varianten vor Komplikationen schützen, die durch *P. falciparum* Blutstadien ausgelöst werden. Zusätzliche *in vitro* Experimente lassen auf eine verbesserte Entwicklung von späten *P. berghei* Leberstadien durch APOA1 schließen.
In dieser Studie zeige ich, dass zusätzliche Selektion von Allelen, die einen Einfluss auf die prä-erythrozytäre Parasitenentwicklung haben, eine Rolle im Krankheitsverlauf von Malaria spielen können, die bisher unbekannt war. Meine Ergebnisse bestätigen die Relevanz von klinisch unauffälligen Wachstumsphasen, wie die Reifung der *Plasmodium* Leberstadien, für die klinische Epidemiologie von Infektionskrankheiten.

# 1. Introduction

## 1.1 Malaria: past and present

Historical records of seasonal fevers associated with marshes and biting insects exist in different cultures like the ancient Chinese, Indians and Assyrians. In the 5th century BC Hippocrates (460-370 BC) observed a connection between appearance of disease and seasons of the year and described the clinical picture of malaria. *Celsus* (25 BC-54 AD) portrayed a detailed picture of various types of malaria. In the 18 century the people related these typical fevers to the foul air near swamps and therefore it was widely named as malaria from the Italian "mala aria".

In 1880 the army surgeon *A. Laveran* (1845-1922) first described malaria parasites in human blood. Around ten years later *D.L. Romanowski* (1861-1921) developed a polychrome staining method for demonstrating plasmodia in blood smears. The transmission was still unresolved until *Sir Patrick Manson* (1844-1922) proposed a transmission from person to person by mosquitoes in 1894. His theory was supported by Sir Ronald Ross' (1857-1932) findings of pigmented cysts in *Anopheles* mosquitoes in 1897. A year later *Ross* worked out the complete cycle of bird malaria. In 1898 *G. Grassi* (1854-1925), *A. Bignami* (1861-1919) and *G. Bastianelly* (1862-1959) in Italy described the cycle of human malaria parasites. *Ross*' work earned him the 1902 Nobel Prize in medicine.

Until now more than 120 Plasmodium species were identified of which *P. falciparum*, *P. vivax, P. knowlesi, P. ovale* and *P. malariae* are able to infect human hosts. In the past malaria was widely distributed around the world. Records of indigenous malaria reached from Russia down to Argentina (*Hay et al., 2004*). In these endemic areas malaria free regions were found because of environmental factors limiting the transmission. From the Middle Ages until last century parts of Europe and North America were endemic areas. In the 1950s it was eradicated from North America and in the 1960s from Southern Europe.

Today malaria is one of the most important health burdens and reason of child mortality worldwide. *P. falciparum* causes about half a billion episodes of disease and kills 1 to 2 million children in Africa annually (*Snow et al., 2005*). Other less lethal *species P. vivax, P. ovale* and *P. malariae* are responsible for additional hundreds of million cases.

Worldwide 1.38 billion people are at risk of *P. falciparum* malaria: 0.69 billion in Central and South East Asia, 0.66 billion in Africa, Yemen, and Saudi Arabia, and 0.04 billion in Middle and South America (*Hay et al., 2009*). These areas differ by the level of *P. falciparum* endemicity and transmission intensities (Figure 1). Uniformly low endemic levels are found from Middle to South America. Low endemicity of *P. falciparum* malaria is widespread in Asia with small areas of intermediate and very rarely high transmission. High levels of endemicity are common in Africa whereas *P. vivax* occurs mainly outside of Africa in tropical and sub-tropical regions. In areas where both species co-exist mixed infections are common (*Snounou and White, 2004*).

The health- and socio-economic impact of *P. falciparum* malaria is highest in Sub-Saharan Africa. 70% of all cases and 75% of all mortality cases are in children younger than five years were concentrated in this region (*Snow et al., 2005*; *Breman, 2001*). The high death toll and continuous health burden affects the economic growth in endemic areas. The impact on local economy includes costs of health care, working days lost due to sickness, days lost in education, decreased productivity due to brain damage from cerebral malaria, and loss of investment and tourism (*Greenwood et al., 2005*). *Gallup and Sachs (2001)* described that the level of gross national product per capita in a malarious country is reduced to more than half of that of a non-malarious country. Therefore, the fight against malaria was set as a main goal to reduce extreme poverty in these regions at the Millennium Summit in September 2000.

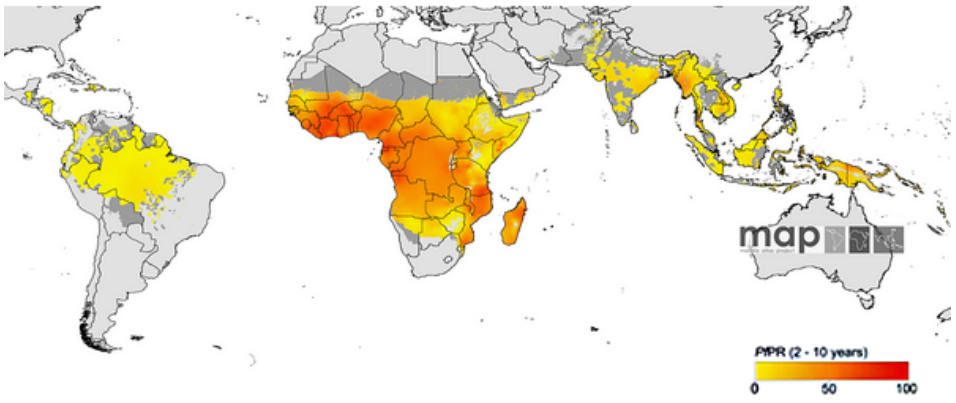



**Figure 1: The spatial distribution of *P. falciparum* malaria endemicity**
Annual *P. falciparum* parasite rates (PfPR) in children from 2-10 years within the stable spatial limits of transmission (yellow to red, 0%–100%). The rest of the land area was defined as unstable risk (dark grey). (*Hay et al., 2009*).

### 1.2 Characteristics of *Plasmodium*

*Plasmodium* parasites are protozoa of the genus *Plasmodium.* Genealogically they are organized in the phylum of *Apicomplexa*, class of *Sporozoa*, order of *Coccidia*, suborder of *Haemosporidiae* and family of *Plasmodiiae*.

*Plasmodium* parasites are found in a spectrum of hosts like mammals, reptiles and birds. The great majority of the parasites are transmitted by various mosquito species whereas human malaria parasites are exclusively transmitted by anophelines. All *Plasmodium* species show strong host specificity. The believed last common ancestor with a two-host life cycle in Dipterans and vertebrates that evolved around 130 million years ago indicates a long history of co-adaption between host and parasite (*Carter and Mendis 2002*).

The life cycle of *Plasmodium* is complex and depends on 2 hosts. Therefore, the parasite must adapt to different environmental and host specific conditions and factors. All mammalian *Plasmodium* species have comparable life cycles. The sexual stage is linked to the invertebrate host whereas the asexual stage only occurs in the mammal.

The parasite in its sporozoite form is injected into the host by the female mosquito during the blood meal. After short inoculation times the sporozoites enter the bloodstream and finally invade hepatocytes. Liver resident macrophages, so called Kupffer cells, serve as gateways for sporozoites to enter the hepatocytes (*Frevert, 2004*). Immediately after invasion the parasitophorous vacuole membrane (PVM) forms around the sporozoite. The liver stage undergoes nuclear replication and matures to hepatic schizonts (*Amino et al., 2006, Vaughan et al., 2008*). These liver cell stages release thousands of merozoites in the blood stream surrounded by host cell membrane, so called merosomes (*Sturm et al., 2006*). In a rodent model the majority of merosomes exit the liver intact and accumulate in the lungs, where the merozoites are released (*Baer et al., 2007*). *P. vivax* and *P. ovale* liver cell stages are exceptional because some can arrest their development and form hypnozoites.

Released merozoites are capable to infect erythrocytes. After invasion a PVM forms around the merozoite. At first, it differentiates into a ring trophozoite and then into a mature throphozoite that enlarges to form a schizont and finally divides into intraerythrocytic merozoites until the multiplying parasite causes the infected red blood cell to rupture and release of a new generation of merozoites (*Cowman and Crabb,*

*2006*). The duration of this phase differs between different *Plasmodium* species and strains, typically 48 hours for *P. falciparum, vivax* and *ovale* or 72 hours for *P. malariae*. Upon erythrocyte infection, some merozoites differentiate into sexual forms, so-called gametocytes. The male gametocytes undergo a rapid nuclear division if a female mosquito takes them up via blood meal. Their nucleus divides into four to eight nuclei that exflagellate. These so called microgametes fertilize the female macrogametes. The resulting ookinete penetrates the midgut wall and transforms under the midgut's basal membrane into an oocyst. After rupture of the mature oocyst more than 1,000 sporozoites are released and some invade the mosquito's salivary glands via the haemolymph. They can be transmitted into a susceptible host via the saliva during the mosquito's next blood meal and complete the life cycle.

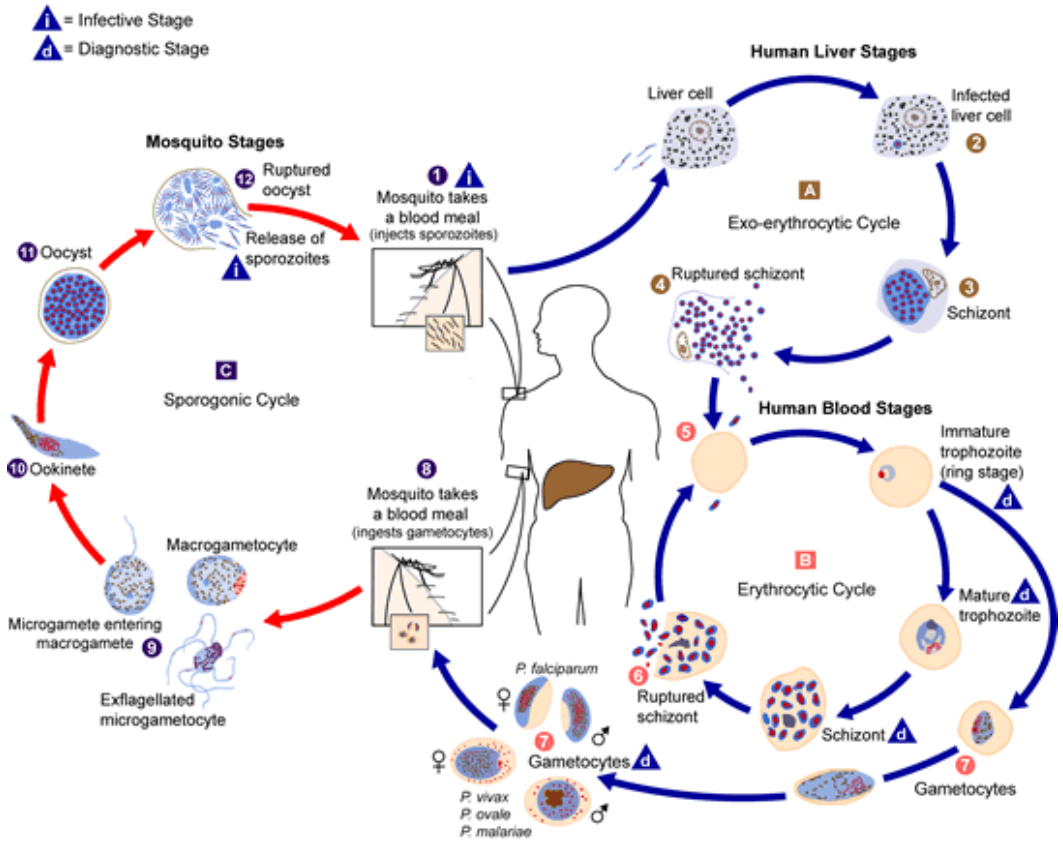



**Figure 2: Life cycle of human *Plasmodium***
Details described in the text (*www.dpd.cdc.gov*).

### 1.3 Clinical features of malaria

Malaria is an acute febrile illness. Disease outcome depends on species of parasite, age, genetic factors, nutritional status, state of health and immunity of the patient. After infection the liver cell stage is clinical silent.

The first clinical sign is fever followed by febrile paroxysms. 2 to 3 days before the first paroxysm patients often complain of malaise, headache, dizziness, fatigue, pain, anorexia and nausea. The paroxysms are divided into three stages: the cold stage (shivering), the hot stage (fever) and the sweating stage with subsequent declining fever. The paroxysms periods repeat every 48 or 72 hours in dependence of the *Plasmodium* species. Due to these patterns malaria was given its other names: Benign tertian malaria (*P. vivax*), malignant tertian (*P. falciparum*), tertian (*P. vivax*, *P. ovale*) and quartan malaria (*P. malariae*). It only occurs with a synchronized parasite population in the patient. In *P. falciaprum* malaria the periodicity is often not clearly defined resulting in continuous or remittent fever in patients.

In non-immune patients, the disease can rapidly progress to life-threatening severe malaria, which is caused by sequestration in different organs and host immune responses (*Mackintosh et al., 2004*). Severe malaria can develop into a complex multisystem disorder, which is characterized by cerebral malaria, hepatic dysfunction, anaemia, renal dysfunction, pulmonary oedema, metabolic acidosis, hypoglycaemia, circulatory collapse and/ or coma (*Rasti et al., 2004*).

Malaria during pregnancy can result in anaemia in the pregnant woman, low birth weight, premature birth, abortion and increased perinatal death. The risk for the woman is highest at the first pregnancy and diminishes in subsequent pregnancies (*Andrews and Lanzer, 2002*). The severity of disease is determined by a combination of parasite virulence factors and host inflammatory responses (*Mackintosh et al., 2004*).

### 1.4 Immunology of malaria

The clinical outcome of a malaria infection depends on many factors. Besides key influences from host and parasite, the transmission dynamics of the disease play an important role (*Baird et al., 1998*). In regions with stable year-round *Plasmodium* transmission by infected mosquito bites exist a persistent level of low parasitaemia in the population. Only a small percentage of the infected patients develop severe

malaria. All clinical outcomes can be observed but the majority shows no or only mild clinical symptoms. Populations in low transmission areas have a higher risk of severe complications per infection. Stable transmission seems to play a major role in development of immunity against malaria *(Mbogo et al. 1995, Snow et al. 1997)*. Acquisition of immunity is a slow process. Even in high transmission areas it takes years and only leads to clinical protection but not sterile immunity against the parasite (*Gilles and Warrell, 2002*). In high-transmission areas malaria is a severe threat for children from 0 to 5 years but a mild condition in adults *(Baird et al., 1998).* Severity of malaria declines from the 4$^{th}$ year and one-year later death is a scarce disease outcome. A substantial immunity can be already acquired with 1 or 2 episodes of malaria *(Gutpa et al., 1999).* In adults mild symptoms like fever and headache are common disease outcomes. However, all age groups are still susceptible to *Plasmodium* infections and parasite prevalence is high. In regions with unstable malaria transmission the acquisition of immunity is delayed. Moreover, immunity appears not to be stable.

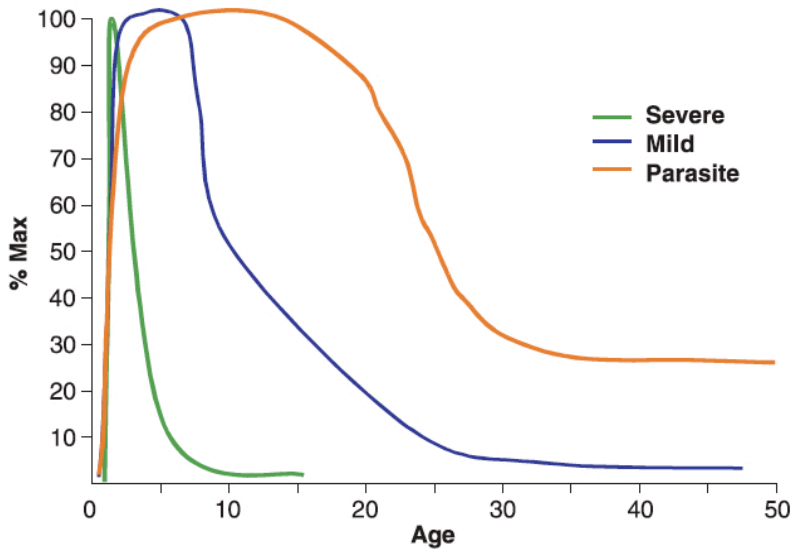



**Figure 3**: **Population indices of immunity to *P. falciparum*.**
The age pattern of asymptomatic parasite prevalence and the period prevalence of both mild and severe clinical malaria are shown in relation to maximum prevalences recorded from studies in Kilifi at the Kenyan coast (*Marsh and Kinyanjui, 2006*).

*Plasmodium* is a complex organism with different morphological forms in the human host. Accordingly, the host immune response is confronted by intra- and extracellular parasites with thousands of recognizable epitopes. Immunity is first acquired against life-threatening disease, then later against symptomatic infection and finally partial immunity to parasitaemia (*Schofield and Grau, 2005*). Immunity to malaria consists of cellular and humoral responses against all stages of infection. Antibodies play an important role against extracellular *Plasmodium* forms. It was shown that they could immobilize sporozoites inoculated in the skin (*Vanderberg and Frevert, 2004*) or block host cell invasion of merozoites (*Cohen and Butcher 1970*). Against intra-hepatocytic *Plasmodium* stages interferon-γ (IFN- γ), $CD4_+$ T cells and CD8+ T-cell responses are key factors (*Stevenson and Riley, 2004; Doolan and Martinez-Alier 2006*). Details to immune responses at different *P. falciparum* life stages are described in Figure 4.

Until today a major hope to fight and finally eradicate malaria is the development of a malaria vaccine. However, an affordable and effective vaccine has not been found yet. Different vaccine approaches targeting pre-erythrocytic, erythrocytic and transmission phase of *Plasmodium* are in development.

*Nussenzweig et al.* (*1967*) have demonstrated that a short-lived protection against malaria can be achieved by vaccination with irradiated sporozoites. Immunization with radiation-attenuated sporozoites (RAS) could be induced in rodents, non-human primates and humans nevertheless their deployment in endemic areas was not possible yet (*Doolan and Hoffman, 2000; Ballou, 2007*).

As a metabolic center, the liver plays an important role in carbohydrate and lipid metabolism, glycogen storage, plasma protein synthesis, hormone production, bile formation, detoxification and breakdown of compounds. Compared with other organs, the liver is particularly enriched with cells of the innate immune system and antigens from the gastrointestinal tract first encounter the innate immune system defences there (*Mehal et al., 2001*). Accordingly, strategies against liver cell stages are suggested as a key vaccine target. Blocking sporozoite invasion into the liver by antibodies or elimination of liver cell stages by Cytotoxic T lymphocytes (CTLs) would lead to a sterile protection.

A vaccination with living *Plasmodium* sporozoites that are able to invade hepatocytes, but do not replicate due to radiation-, genetic or chemical attenuation are promising approaches against an extreme complex parasite like *Plasmodium* (*Luke*

*and Hoffman, 2003; Mueller et al., 2005 (1,2); Jobe et al., 2007; Purcell et al., 2008*). Another liver-stage–based strategy is a recombinant vaccine, called RTS,S/AS01 consisting of the *P. falciparum* circumsporozoite protein (CSP) found on the surface of the sporozoite. Due to its low immunogenic capacity it is fused with a hepatitis B surface antigen formulated in an adjuvant system for a higher immune response (*Ballou, 2009*). Until now, it seems to be the only out of 47 candidates that reach phase III (*Langhorne et al., 2008*).

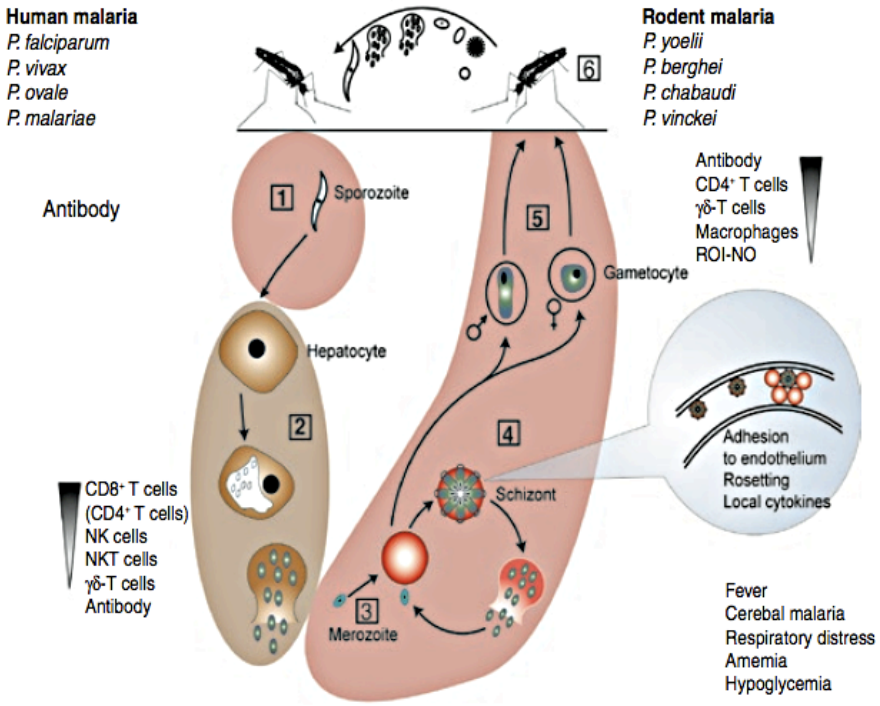



**Figure 4**: ***P. falciparum*** **life cycle with host immune responses.**
Top left and right: plasmodia that infect humans and rodents. Numbers indicate effector mechanisms thought to be effective against *Plasmodium* in the mammalian host and in blocking transmission to the mosquito: **1**: antibodies to sporozoites neutralize sporozoites and/or block invasion of hepatocytes; **2**: IFN- and CD8+ T cells (CD4+ dependent), natural killer (NK) cells, natural killer T (NKT) and T cells kill intrahepatic parasites; **3**: antibodies to merozoites opsonize merozoites for uptake and/or inhibit invasion of RBCs; **4**: antibodies to infected RBCs surface opsonize infected RBCs for phagocytosis and/or block the adhesion of infected RBCs to endothelium; TNF and IFN-γ activate macrophages to phagocytose and/or kill infected RBCs and merozoites; antibodies to glycosylphosphatidylinositol neutralize parasite toxins and prevent the induction of excessive inflammation; **5**: antibodies to infected RBCs prevent the sequestration of gametocytes, which prevents the sequestration and maturation of gametocytes; **6:** antibody and complement taken up in the blood meal mediate the lysis of gametocytes and prevent fertilization and further development of the parasite in the mosquito. ROI, reactive oxygen intermediate; NO, nitrous oxide**.** Fever and severe malaria are associated with the parasite cycle in the blood, as well as rosetting of infected RBCs to blood vessel endothelium (blue insert) *(Langhorne et al., 2008).*

### 1.5 Hepatocyte stage proteins and host-parasite interactions

Until now, very little is known about the molecular mechanisms of invasion and development of liver cell stages. The PVM acts as barrier between the parasite and the host cell but also as a gateway for interactions. A functional PVM is required for liver cell stage development and final maturation to merozoites. This was supported by a double knock-out mutant of *P. yoelii* that was unable to form a stable PVM (*Labaied et al., 2007*).

So far, several essential interactions between the host hepatocyte and the parasite could be observed in rodent models. *Silvie et al. (2008)* identified an asparagine rich protein named "sporozoite and liver stage asparagin-rich protein" (SLARP) that is only expressed in sporozoites and liver cell stages. *P. yoelii* SLARP knock-out experiments showed a normal invasion but an impaired development in the liver. *Matuschewski et al. (2002)* identified 2 proteins, referred as "Upregulated in infective sporozoites gene" (UIS) 3 and 4, which are exclusively expressed in the liver cell stage of *P. berghei*. These proteins, members of the "early transcribed membrane proteins" (ETRAMP) family are located in the PV tubovesicular network with their C-terminal ends in the hepatocyte's cytoplasma (*Spielmann et al., 2006 1*).

*Uis3* and *uis4* knock-out mutants showed an impaired development and failed finally in establishing a blood infection in mice. Located in the PVM, both proteins interact with host ligands like Apolipoprotein A1 (APOA1) and Fatty Acid Binding Protein 1 (FABP1), which were identified in a yeast-two hybrid screen (*Mueller, personal communication*). A third ligand, SNAP Associated Protein (SNAPAP), was found interacting with UIS7, which contrary to UIS3 or UIS4 has no homolog in *P. falciparum* (*Mueller Ak, unpublished*). A reverse immunization with *uis3*- and *uis4*-deficient sporozoites conferred complete protection against infectious sporozoite challenge in a rodent malaria model.

Both, APOA1 and FABP1, are involved in the lipid metabolism of the host. Mainly expressed in the liver FABP1 is the key player in the lipid composition of the liver (*Fisher et al., 2007*). This small, highly conserved, cytoplasmic protein is able to bind long-chain fatty acids, bile acids and other hydrophobic ligands. An essential interaction between FABP1 and *P. yoelii* liver cell stages could be demonstrated by *Mikolajczak et al.* (*2007*). In humans a coding SNP in *FABP1* was associated with fasting triglycerides and LDL-cholesterol concentrations in women (*Fisher et al.,*

*2007*).

The conserved APOA1 is the major protein component of high-density lipoprotein (HDL) in plasma, which transports cholesterol from tissues to the liver for excretion. It is a cofactor for lecithin cholesterolacyltransferase (LCAT), which is responsible for the formation of most plasma cholesteryl esters. Several diseases or beneficial effects are linked to variation in APOA1. Amyloidosis (*Soutar et al. 1992*), hypertriglyceridemia *(Stocks et al., 1987*), dyslipoproteinemia (*Strobl et al., 1988*) and many more diseases are influenced by APOA1. The best-known protective effect of APOA1 is against cardiovascular diseases (*Ajees et al., 2006*). Recently, a complex of APOA1, apolipoprotein L-1 (apoL1) and haptoglobin-related protein (Hpr) was observed in killing trypanosomes through anionic pore formation in the lysosomal membrane of the parasite (*Paysa and Vanhollebeke, 2009; Harrington et al., 2009).*

SNAPAP, a small protein of 139 residues, is component of the SNARE complex of proteins that is required for synaptic vesicle docking and fusion *(Ilardi et al., 1999*). Its function is not clear but it has been shown that it interacts with SNAP-25, BLOC-1, SNAP23, Dysbindin, TRPV1, PLDN and RGS7 (*Starcevic and Dell'Angelica, 2003; Hunt et al., 2003; Morenilla-Palao et al., 2006*).

The role of lipids for *Plasmodium* remains unclear. It is hypothesized that the fast development of the parasite in the liver requires a high amount of lipids and thus interaction with lipid carriers of the host is essential (*Mikolajczak et al., 2007*). A change in serum lipid and lipoprotein compositions in the patients due to *Plasmodium* infections was observed (*Blair et al. 2002; Faucher et al. 2002*). In malaria-exposed primigravidae APOA1 plasma level in are associated with severe malaria (*Simpson et al., 2010*). Because of high demand for lipids *Mitamura and Palacpac* (*2003*) proposed the lipid metabolism as an attractive target for anti-malarial drugs.

Initially, it was thought that *Plasmodium* was incapable of fatty acid *de novo* synthesis but components of type II fatty acid biosynthesis (FAS-II) pathway were discovered (*Ralph et al., 2004*). Localized in the apicoplast it was shown that FAS-II is essential for malaria parasite late liver stage development (*Vaughan et al., 2009*). In contrast to the related *Toxoplasma gondii* that depends on LDL (*Sehgal et al., 2005*) lipid trafficking between HDL and *Plasmodium falciparum*-infected red blood cells was observed *(Grellier et al., 1991).*

In liver cell stages cholesterol seems to play a major role (*Rodrigues et al. 2008*). A disruption of the host cell receptor "scavenger-receptor-BI*"* (SR-BI) showed a negative impact on liver stage development (*Yalaoui et al., 2008*). SR-BI is a high-affinity HDL receptor that plays a role in lipid concentration in the plasma (*Trigatti et al., 2003, Masson et al., 2009*). Moreover, HDL/SR-BI interaction with CD81 seems to influence infectivity of the host cell to hepatitis C or *Plasmodium* (*Rodrigues et al., 2008; Yalaoui et al., 2008; Dreux et al., 2009).*

### 1.6 Genetic host factors influencing protection against malaria

*Plasmodium* is highly adapted to its hosts. In the whole life cycle the parasite depends on specific interactions with the host for access to nutrients, overcome physical barriers or avoidance of immune responses. On all stages the parasite is susceptible to interruptions leading to a reduced survival of the parasites and an increased fitness for the exposed host. The genetic basis of resistance to malaria is complex and probably many unknown genes are involved. Until now, several human genetic traits have been identified that effect susceptibility to and/ or pathogenesis of *P. falciparum* malaria. These are shown to affect resistance to malaria, disease oucome and the immune response of the host (*Kwiatkowski, 2005*). An overview of host factors associated to malaria and their characteristics are listed in table 1 (a,b and c).

The best-described example for genetically induced protection is the *HBB* gene. Three different single nucleotide polymorphisms (SNPs) confer protection against malaria by amino acid substitution: Glu6Val (HbS), Glu6Lys (HbC) and Glu26Lys (HbE). Structurally each haemoglobin molecule consists of a haeme molecule surrounded by four polypeptide chains referred to as α- and β-globins. All substitutions result in alterations of the β-chain of the haemoglobin molecule. β-thalassaemia mutations in the *HBB* gene lead to either complete loss of function or production of low levels of the β-globin protein.

The red blood cells of *HBB* S heterozygotes, so-called sickle cell trait (HbAS), carry a normal and an abnormal β-chain with an unobtrusive phenotype under normal conditions. Homozygotes' (HbSS) erythrocytes show under hypoxia an abnormal sickle shape as a result of 2 abnormal β-chains. Sickle cell disease has a poor

prognosis and, if untreated, can be fatal in early childhood. In spite of homozygote mortality, the *HBB* S allele is found in increased frequencies over 20% in malaria-endemic areas compared to malaria free regions (*Kwiatkowski, 2005*). Due to a 90% reduced risk of severe malaria the polymorphism is an example for balanced selection where the deleterious effects of the HbSS genotype prevent the S allele's fixation (*Williams et al., 2005; Kwiatkowski, 2005*). The other variants of the *HBB* gene—namely, HbC and HbE or beta thalassaemia show only mild clinical symptoms (*Stuart and Nagel, 2004*).

Variations in the *HBA* gene, causing α-thalassaemia, are affecting risk of malaria as well. Deletion of one or both α-globin genes due to a mutation in *HBA* results in an autosomal recessive disease. In Africa the most common is the heterozygous $\alpha^{-}$- and homozygous $\alpha^{+}$–thalassaemia trait. Because of the malaria-protective character and the absence of strong negative effects on health, the allele frequencies have risen to high frequencies (up to 80%) in many populations (*Williams, 2006*). Individuals with $\alpha^{+}$–thalassaemia are usually asymptomatic except for occasional mild anaemia and low red blood cell indices.

Mechanisms of protection afforded by Hb variants remain largely unknown. For HbS two mechanisms are hypothesized: The suppression of parasite growth in red cells and enhanced splenic clearance of parasitized erythrocytes (*Kwiatkowski, 2005*). Both, difficult to examine *in vivo,* is supported by a reduced incidence of clinical malaria and lower parasite densities observed in children with HbAS (*Williams et al. 2005 (2)*). Furthermore, HbS seems to have a synergistic effect on acquired immunity*: Williams et al.* (*2005, 1*) observed an increased malaria-protective effect in 2-10 years old Kenyan children.

Like for sickle cell, the protective mechanisms of α-thalassemia are largely unknown. *Pattanapanyasat et al. (1999)* showed a reduced parasite growth in erythrocytes expressing the $a^{+}$-thalassaemia phenotype. Other *in vitro* studies support different mechanisms like an impaired ability to form rosettes, better splenic clearance, reduced expression of complement factor 1 (CR1) and higher expression of antigens in infected erythrocytes (*Williams, 2006*). However, both $\alpha^{-}$- and $\alpha^{+}$–thalassaemia protect against severe and fatal malaria but has no effect either on mild or rather symptomless malaria or parasite densities in the blood (*Williams et al., 2005, 3 and 2006*). All of these genetic polymorphisms could influence each other and thus susceptibility to malaria and disease outcome. Indeed, *Wiiliams et al.*

*(2005, 3)* observed a negative epistasis between the malaria-protective effects of a+-thalassemia and the sickle cell trait. The incidence of mild and severe *falciparum* malaria was close to baseline if both phenotypes were inherited together.

**Table 1: Host factors associated with malaria.**
Different human factors are listed affecting resistance (a), polymorphisms of host receptors for cytoadherence by *P. falciparum* infected erythrocytes influencing susceptibility or resistance (b) and immune gene associations with resistance and susceptibility to malaria (c) (adapted from *Kwiatkowski, 2005*)

a) **Common Erythrocyte Variants That Affect Resistance to Malaria**

| Gene | Protein | Function | Reported Genetic Associations with Malaria |
|---|---|---|---|
| *FY* | Duffy antigen | Chemokine receptor | FY*O allele completely protects against *P. vivax* infection. |
| *G6PD* | Glucose-6-phosphatase dehydrogenase | Enzyme that protects against oxidative stress | G6PD deficiency protects against severe malaria. |
| *GYPA* | Glycophorin A | Sialoglycoprotein | GYPA-deficient erythrocytes are resistant to invasion by *P. falciparum*. |
| *GYPB* | Glycophorin B | Sialoglycoprotein | GYPB-deficient erythrocytes are resistant to invasion by *P. falciparum*. |
| *GYPC* | Glycophorin C | Sialoglycoprotein | GYPC-deficient erythrocytes are resistant to invasion by *P. falciparum*. |
| *HBA* | α-Globin | Component of hemoglobin | $\alpha^+$ Thalassemia protects against severe malaria but appears to enhance mild malaria episodes in some environments. |
| *HBB* | β-Globin | Component of hemoglobin | HbS and HbC alleles protect against severe malaria. HbE allele reduces parasite invasion. |
| *HP* | Haptoglobin | Hemoglobin-binding protein present in plasma (not erythrocyte) | Haptoglobin 1-1 genotype is associated with susceptibility to severe malaria in Sudan and Ghana. |
| *SCL4A1* | CD233, erythrocyte band 3 protein | Chloride/bicarbonate exchanger | Deletion causes ovalocytosis but protects against cerebral malaria. |

b) **Host Molecules That Mediate Cytoadherence by *P. falciparum*–Infected Erythrocytes and That Have Been Reported to Show Association with Resistance or Susceptibility to Malaria**

| Gene | Protein | Interaction with Parasitized Erythrocyte[a] | Reported Genetic Associations with Malaria |
|---|---|---|---|
| *CD36* | CD36 antigen, thrombospondin receptor | PE-binding receptor on endothelium and dendritic cells | *CD36* polymorphisms show variable associations with severe malaria in the Gambia, Kenya, and Thailand. |
| *CR1* | CR1, complement receptor 1 | PE-binding receptor on erythrocytes | *CR1* polymorphisms show variable associations with severe malaria in the Gambia, Thailand, and Papua New Guinea. |
| *ICAM1* | CD54, intercellular adhesion molecule-1 | PE-binding receptor on endothelium | *ICAM1* polymorphisms show variable associations with severe malaria in Kenya, Gabon, and the Gambia. |
| *PECAM1* | CD31, platelet-endothelial cell–adhesion molecule | PE-binding receptor on endothelium | *PECAM1* polymorphisms show variable associations with severe malaria in Thailand, Kenya, and Papua New Guinea. |

[a] PE = parasitized erythrocyte.

c) **Immune Genes Reported to Be Associated with Different Malaria Phenotypes**

| Gene | Protein | Function | Reported Genetic Associations with Malaria |
|---|---|---|---|
| *FCGR2A* | CD32, low affinity receptor for Fc fragment of IgG | Clearance of antigen-antibody complexes | Association with severe malaria in the Gambia |
| *HLA-B* | HLA-B, a component of MHC class I | Antigen presentation that leads to cytotoxic T cells | HLA-B53 association with severe malaria in the Gambia |
| *HLA-DR* | HLA-DR, a component of MHC class II | Antigen presentation that leads to antibody production | HLA-DRB1 association with severe malaria in the Gambia |
| *IFNAR1* | Interferon α receptor component | Cytokine receptor | Association with severe malaria in the Gambia |
| *IFNG* | Interferon γ | Cytokine with antiparasitic and proinflammatory properties | Weak associations with severe malaria in the Gambia |
| *IFNGR1* | Interferon γ receptor component | Cytokine receptor | Association with severe malaria in Mandinka people of the Gambia |
| *IL1A/IL1B* | Interleukin-1α and -1β | Proinflammatory cytokines | Marginal associations with severe malaria in the Gambia |
| *IL10* | Interleukin-10 | Anti-inflammatory cytokine | Haplotypic association with severe malaria in the Gambia |
| *IL12B* | Interleukin-12 β subunit | Promotes development of Th1 cells | Association with severe malaria in Tanzania |
| *IL4* | Interleukin-4 | Promotes antibody-producing B cells | Association with antimalarial antibody levels in Fulani people of Burkina Faso |
| *MBL2* | Mannose-binding protein | Activates classic complement | Association with severe malaria in Gabon |
| *NOS2A* | Inducible NO synthase | Generates NO, a free radical | Various associations with severe malaria in Gabon, the Gambia, and Tanzania |
| *TNF* | Tumor necrosis factor | Cytokine with antiparasitic and proinflammatory properties | Various associations with severe malaria and reinfection risk in the Gambia, Kenya, Gabon, and Sri Lanka |
| *TNFSF5* | CD40 ligand | T cell–B cell interactions leading to immunoglobulin class switching | Association with severe malaria in the Gambia |

### 1.7 Effects of evolutionary selection on the human genome by malaria

Malaria was one of the most driving forces in history of mankind. With the introduction of agriculture, settlement and thus growing population size, malaria gained in importance at beginning of the Neolithic period around 10,000 years ago. The high mortality and morbidity of malaria especially of *P. falciparum* applied a strong selective pressure on the human genome (*Klein and Takahata, 2002; Kwiatkowski, 2005*). In recent years an increasing number of identified human genes affecting malaria bear witness of the strong influence of *Plasmodium* in history of humankind. The heritability of malaria susceptibility was shown in Gambia by *Jepson et al. (1997*). They observed a higher chance of malaria fever attacks in monozygotic than in dizygotic twins. In two cohort studies in Kenyan children it was found that 25% of the total variation in incidence of mild clinical malaria and hospital admissions to malaria was explained by additively acting host genes. HbAS, the strongest known resistance genetic factor explained only 2% suggesting the existence of many unidentified malaria-protective genes (*MacKinnon et al., 2005*).

Adaptations to malaria can have disadvantageous effects like the high mortality of HbSS or the thalassemias. They are the most common Mendelian diseases of humans and they constitute a major global health problem today. About 75% of the global haemoglobinopathies occur in sub-Saharan Africa with more than 200,000 affected newborns per year (*Weatherall and Clegg, 2001*).

Inter-ethnic comparisons of susceptibility to malaria show that different malaria-resistance alleles have arisen in different places. Human population diversity in Africa is exemplified by over 2,000 distinct ethnic groups. Recent studies observed extensive genetic variation among even geographically close African populations (*Tishkoff and Williams, 2002; Tishkoff et al., 2009*). For example the *HBB* S allele is common in Africa but rare in Southeast Asia, whereas the opposite is true for the *HBB* E. In West Africa *HBB* C is more frequent than the *HBB* S allele (*Kwiatkowski, 2005*). Other different geographic distributions of genetic traits like of the thalassaemias, glucose-6-phosphate dehydrogenase (G6PD) deficiency or the Duffy-negative blood group support the theory that they evolved in different populations. Moreover, it seems that the same genotype has arisen several times independently in different populations like HbS that is found in 4 different haplotypes (*Lapoumëroulie et al., 1992*) or the geographic allele diversity of G6PD (*Tishkoff et al., 2006*).

### 1.8 Association studies and signatures of selection

The idea that malaria has been acting as a major evolutionary force in recent human history has driven the development of tools to explore the human genome for signatures of positive selection (*Sabeti et al, 2006*). Identifications of the proposed candidates for selection, like *HBB S* in Africa or lactase (*LCT)* in Europe, have strong functional support because due to the direct influence of the mutation on the phenotype. Most proposed candidates lack compelling biological evidence and are only relied on comparative and population genetic evidences (*Sabeti et al., 2006*). Well-defined phenotypes, one or two genetic disease loci with a high penetrance, a small phenocopy rate and usually small allele frequencies characterize Mendelian traits that are rare in population. Most common diseases in humans are complex traits, like lipid metabolism disorders or hypertension, with no direct correspondence between a single genotype and the disease phenotype. Often they show diffuse phenotypes, high phenocopy rate, genetic heterogeneity, low penetrance, epistasis, mitochondrial inheritance and are often influenced by environmental factors (*Lander and Schork, 1994*).

In Darwin's theory the fixation of adaptive mutations in a population is an important theme on the origin of species by natural selection. In the late 1960s *Motoo Kimura et al.* proposed in his "neutral theory of molecular evolution" that most polymorphisms are selectively neutral and are maintained or fixed in population by chance, so-called genetic drift. In comparison with this background genetic variation signatures of positive or negative selection can be identified (*Bamshad and Wooding, 2003*). Other effects due to population demographic history, such as periods of reduced population size (bottlenecks), expansions, and subdivided populations are problematic to determine whether a signature is caused by selection. Different approaches have shown to be suitable for detecting selection depending on the window of evolutionary time. Figure 5 names briefly the different signatures of selection in an estimated evolutionary time frame, in which it can be used to detect moderately to strong selection in humans.

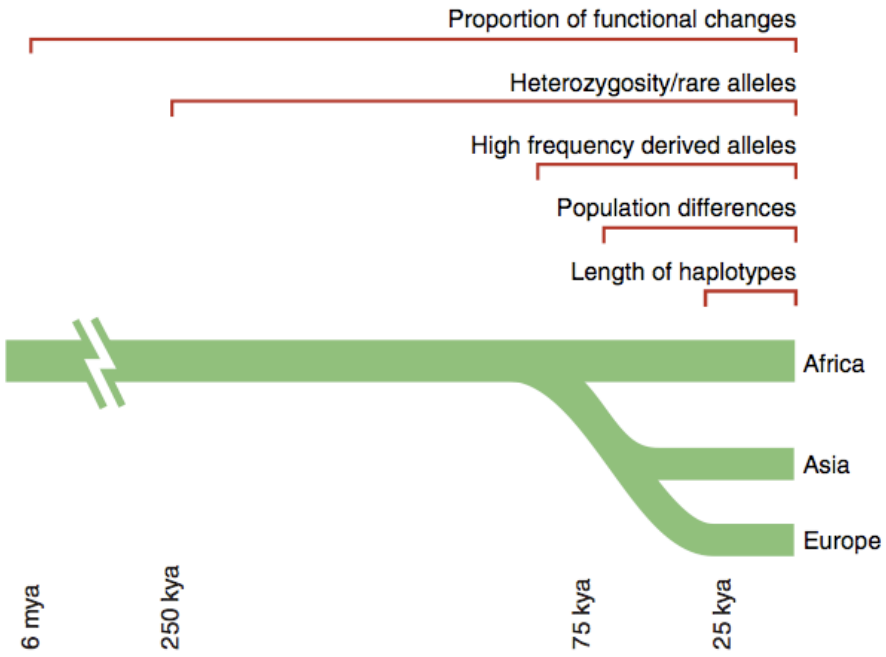



**Figure 5**: **Time scales for signatures of selection.**
Different signatures of selection are listed top down. Brief explanation and the common statistical methods for their detection are put in brackets.
*Proportion of functional changes* (mutations that alter protein function are usually deleterious and are thus less likely to become common or reach fixation; McDonald-Kreitman test); *Heterozygosity/rare alleles* (Reduction in genetic diversity, increase of allele also increase variants at nearby genomic locations leads to low diversity; Tajima's D, Fu and Li's D); *High-frequency derived alleles* (hitchhiking with favoured new allele creates region containing many high-frequency derived alleles; Fay and Wu's H); *Populations* differences (selection in subpopulations due to different environments; $F_{st}$ test); *Length haplotypes* (allele rises rapidly enough that recombination does not break down the association with nearby alleles creating long haplotypes; EHH and iHS) (mya, kya: million years, thousand years; *adapted from Sabeti et al., 2006*).

### 1.9 Association and cohort studies

Different genetic variants like single nucleotide polymorphisms (SNPs) and microsatellites are markers that are widely used to detect association of a gene to a phenotype. The usually bi-allelic SNPs are defined to have at least an allele frequency of 1% or more in the population. They occur every 100-300bp and show a low heterozygosity. SNPs have the advantage of low mutation rates and little genotyping costs for large-scale and high-troughput genotyping (*Brookes, 1999; Gray et al., 2000; Emahazion et al., 2001; Burgner et al., 2006*). The frequency of occurrence and genomic location is proportional to the SNP's influence on the gene: From non-coding regions, regulatory hotspots to coding regions with synonymous, non-synonymous, missense and nonsense effects on translation. More than 11.5 million SNPs are found contributing to 90% of the genetic variation in the human genome and because of their low mutation rate they serve as reliable markers of molecular evolution (*Lander et al., 2001, Burgner et al., 2006; Madsen et al., 2007*).

Mainly two kinds of study designs are used to localize the genes underlying human disease: genome-wide and candidate gene association studies. An advantage of genome-wide association, mostly linkage mapping, studies is that no assumptions about potential involved genes have to be made whereas candidate gene studies that concentrate on polymorphisms in specific genes have a higher density of tested polymorphisms. The problem of low resolution is exemplified in a genome-wide association study in West Africa, in which HbS was not associated with malaria until the genomic region was re-sequenced (*Jallow et al., 2009*).

Candidate gene association studies are better suited for detecting genes underlying common and more complex diseases where the risk associated with any given candidate gene is relatively small (*Risch and Merikangas, 1996; Kwon and Goate, 2000*). Furthermore, in contrast to linkage mapping studies, that require large families with disease affected and unaffected individuals, candidate gene approaches can be performed with unrelated subjects or small families. A major difficulty lies in the identification of potential candidate genes by experimental data.

Both study types can be realized within cohort or case-control studies. Additional types of epidemiological studies include cross-sectional and ecological, or correlation studies providing a weaker informative value compared to cohort and case-control studies (*Schoenbach and Rosamond, 2000*).

A prospective cohort study data is the longitudinal observation of healthy subjects

with the characteristic being studied, e.g. disease. Cohort studies are considered as gold standard in observational epidemiology (*Schoenbach and Rosamond, 2000*).

Case-control studies start with the identification of persons with the disease of interest and a suitable control group of subjects without the disease. The purpose of the control group is to estimate the baseline frequency of the genetic variant in the population. The main advantage of cohort studies is that it allows complete information on the subject's disease, including quality control of data, provides a clear temporal sequence of disease, measurement of potential confounders and give an opportunity to study multiple outcomes related to a specific disease. Disadvantages include a possible change during the course of the study, high-cost for follow-up monitoring and thus a small number of subjects (*Szklo, 1998*). In contrast, case-control studies allow studying a large number of subjects at relative low cost. Case-control studies however are prone to selection bias and are more suitable to study rare diseases but lack of validation of information on exposure or other covariates and their retrospective character limits the power of conclusions.

## 2 Aim of study

Despite all international efforts *P. falciparum* malaria is one of the major health burdens in developing countries. The recent detection of tolerance against artemisinins, the key component of all current first-line antimalarial drug combinations (*Dondorp et al., 2009*) underscores the vital need for new affordable drugs and innovative malaria intervention strategies. Critical to present malaria control efforts is the reduction of *Plasmodium* sporozoite transmission from mosquito to man. With the success of immunization with RAS and RTS,S/AS01 as the most advanced vaccine candidate so far, the sporozoites and liver cell stages are brought into perspective as key players. Three genes, *apolipoprotein A1 (APOA1), fatty acid binding protein (FABP1)* and *SNAP-associated protein (SNPAPAP),* coding for host cell ligands, have been identified to interact with parasite proteins (UIS3, UIS4 and UIS7) in a yeast two-hybrid screen. However, the roles, if any, of *Plasmodium* liver stage/host interactions on malaria disease outcome remain unknown. Accordingly, identification of lipid-binding proteins as host cell ligands for essential proteins at the *Plasmodium*/host cell interface permitted a natural genetics approach to validate important roles for pre-erythrocytic parasite development.

The aim of this study was to test the hypothesis whether genetic variants encoding for proteins that participate in host-parasite liver stage interactions, as exemplified by UIS4/APOA1, UIS3/L-FABP and SNAPAP/UIS7, correlate with risk of natural infections in a malaria-endemic area. To this end, I genotyped single nucleotide polymorphisms (SNPs) in two independent well-characterised groups of children from the Kenyan coast (*Williams et al., 2005, 2*) and analyzed their relationship with susceptibility to *P. falciparum* infections and malaria outcome. Additionally, different statistical approaches were used to identify signatures of recent selection within the genomic regions of the target genes to explain the evolutionary history of associated polymorphisms or genes in the context of host/parasite co-evolution. A further aim was to determine the influence of APOA1 on liver cell stage development of in a *P. berghei* model.

# 3 Materials and methods

## 3.1 Materials

### 3.1.1 Equipment

| | |
|---|---|
| Analytic Scales: | |
| Adventurer Pro | Ohaus, Mannheim |
| Adventurer Pro (fine scale) | Ohaus, Mannheim |
| Autoklave | Systec GmbH, Wettenberg |
| Centrifuges: | |
| J2-MC | Beckman Coulter, |
| Megafuge 2.0R | Heraeus Instruments, Hanau |
| RC5Bplus | Sorvall |
| Fridge | Liebherr, Biberach |
| Freezer | Liebherr, Biberach |
| Gel electrophoresis system MiniSub-Cell | Biorad, Munich |
| Genetic Analyzer: | |
| 3130 Genetic Analyzer | Applied Biosystems, Darmstadt |
| 3130xl Genetic Analyzer | Applied Biosystems, Darmstadt |
| Incubator Hera Cell | Heraeus Instruments, Hanau |
| Microscopes: | |
| Confocal Laserscanning Microsc. LSM510 | Zeiss, Jena |
| Microscope (fluorescence) Axiovert 100M | Zeiss, Jena |
| Microscope (light) Axiolab | Zeiss, Jena |
| Microwave | Intellowave, LG; Willich |
| Multiscreen Vacuum Manifold | Millipore, Schwalbach |
| Neubauer chamber | Labotec, Göttingen |
| Nitrogen tank | CBS, USA |
| pH-Meter | Schott |
| PCR thermal cycler: | |
| GeneAmp PCR System 9700 | Applied Biosystems, USA |
| Mastercycler epgradient S | Eppendorf, Hamburg |

Pipettes:

| | |
|---|---|
| Research 10µl, 20µl 200µl, 1000µl | Eppendorf, Hamburg |
| Pipetman Gilson P2, P20, P200, P1000 | Abimed, Langenfeld |
| Power supply Power Pac 300 | Biorad, München |
| Spectrophotometer NanoDrop 1000 | Peqlab, Erlangen |
| Sterile bench Herasafe | Heraeus Instruments, Hanau |
| ThermoMixer Compact | Eppendorf, Hamburg |
| Vortex Mixer 7-2020 | neoLab, Heidelberg |
| Water bath Julabo 7A | Julabo |

### 3.1.2 Chemicals

| | |
|---|---|
| Agarose | Invitrogen, Karlsruhe |
| β-mercaptoethanol | Merck, Darmstadt |
| BSA (Bovine Serum Albumin) | Roth, Karlsruhe |
| DMEM-Medium (high Glucose) | Pan Biotech, Aidenbach |
| Dimethyl sulfoxide (DMSO) | Merck, Darmstadt |
| EDTA (Na2EDTA) | Clonetech, Heidelberg |
| Ethidium bromide | Sigma Aldrich, Taufkirchen |
| Ethanol 100% | Roth, Karlsruhe |
| Fetal calf serum (FCS) | Gibco Invitrogen, Karlsruhe |
| Formaldehyde | Merck, Darmstadt |
| Geneticin (G418) | Gibco Invitrogen, Karlsruhe |
| Glycerol | Roth, Karlsruhe |
| HBSS (Hanks Balanced Salt Sol.) | Pan Biotech, Aidenbach |
| Hi-Di formamide | Applied Biosystems, Darmstadt |
| HPLC $H_2O$ | Roth, Karlsruhe |
| Methanol | Roth, Karlsruhe |
| Sodium acetate | Roth, Karlsruhe |
| Sodium chloride | Merck, Darmstadt |
| Trypan blue | Roth, Karlsruhe |
| PBS (tablets) | Gibco Invitrogen, Karlsruhe |
| POP7 polymer | Applied Biosystems, Darmstadt |
| RPMI 1640-Medium (+25mM HEPES, + L-Glutamine) | Gibco Invitrogen Karlsruhe |

| | |
|---|---|
| Panserin PX10 (serum-free medium) | Pan Biotech, Aidenbach |
| Trypsin/EDTA solution | Gibco Invitrogen, Karlsruhe |

### 3.1.3 Consumables

| | |
|---|---|
| ABgene PCR plates | Thermo Scientific, Waltham,USA |
| chamberslides | LabTek |
| cell culture flasks (25cm$^2$, 75cm$^2$) | Renner, Dannstadt-Schauernheim |
| cover slides | Marienfeld, Lauda-Königshofen |
| cryo-tubes | NALGENE |
| culture well plates (24, 6 well) | Greiner Bio-one, Frickenhausen |
| Eppendorf tubes (1,5ml) | Eppendorf, Hamburg |
| Eppendorf tubes (0,5ml) | Eppendorf, Hamburg |
| filter, sterile (Ø 0,2µm) | Schleicher und Schuell, Einbeck |
| Filter Tip 10E, 20, 100, 200, 1000 | Greiner bio-one, Frickenhausen |
| glass slides | Marienfeld, Lauda-Königshofen |
| gloves | VZM, Heidelberg |
| Parafilm | Pechiney Plastic Packaging |
| PCR reaction tubes (0,2ml) | Greiner Bio-one, Frickenhausen |
| PCR softstrips 0,2ml, colour | Biozym Scientific, Oldendorf |
| pipette tips | Greiner Bio-one, Frickenhausen |
| pipette tips w. filter | Greiner Labortechnik, Nütringen |
| RNase free Biopure 1,5 ml | Eppendorf, Hamburg |
| single use pipettes (1, 2, 10 ,25ml) | Corning Incorporation, Bodenheim |
| syringes | BD Microlance, Becton Dickinson |
| syringe needles | BD Microlance, Becton Dickinson |

### 3.1.4 Buffers, Media and Solutions

*TE-Buffer (Tris-EDTA):*

10 mM Tris/HCl pH 7.6

1 mM EDTA

*50xTAE Buffer:*

242 g Tris

57.1 ml Acetic acid

50 mM EDTA pH 8

*PBS (pH 7,4):*

137mM NaCl

27,mM KCl

9,2 mM $Na_2HPO_4$

1,5 mM $KH_2PO_4$

*DMEM complete:*

445ml DMEM-Medium (high-glucose)

50ml FCS

5ml Penicillin/Streptomycin (sterile filtered)

*Freezing solution:*

20% DMSO,

80% FCS

*Round-up solution* (100x; sterile filtered):

60mg Penicillin

100mg Kanamycin

50mg Fluocytosin

10mg Chloramphenicol

### 3.1.5 Enzymes, ladder, antibodies, vector

| | |
|---|---|
| *Taq* Polymerase | Fermentas, St. Leon-Rot |
| Endonuclease MspI | New England Biolabs, Frankfurt M. |
| Shrimp Alkaline Phosphotase | USB, Staufen |
| Exosap-it | USB, Staufen |
| Exonuclease I | Fermentas, St. Leon-Rot |
| Generuler 1kb, 100bb, 50bp | Fermentas, St. Leon-Rot |
| α-PbHSP70 | AK. Müller, Heidelberg |
| α-Mouse Alexa 488 | Invitrogen, Karlsruhe |
| pcDNA™3.1, 5-His | Invitrogen, Karlsruhe |

### 3.1.6 Kits

| | |
|---|---|
| 5 Prime MasterMix | 5 Prime/ VWR, Darmstadt |
| BigDye v1.1 Cycle Sequencing Kit | Apllied Biosystems, Darmstadt |
| GenomiPhi DNA Amplification Kit | GE Healthcare, Freiburg |
| HiFect Kit | Lonza, Cologne |
| LongRange PCR Kit | Qiagen, Hilden |
| Power SYBR Green Master Mix | Applied Biosystems, Darmstadt |
| QIAamp DNA Blood Mini Kit | Qiagen, Hilden |
| RETROscript Kit | Ambion, Darmstadt |
| RNeasy Mini Kit | Qiagen, Hilden |
| SNaPshot Multiplex System Kit | Applied Biosystems, Darmstadt |

### 3.1.7 Oligonucleotides

All primers were purchased from MWG-Biotech (Ebersberg).

**Table 2: Primer for *APOA1* re-sequencing**

| ID | sequence | length (bp) |
|---|---|---|
| Apo_Promo_for | TTTTGGAGGCGGACAATATC | 20 |
| Apo_Promo_rev | CCTGTTGCTGCTCACTGGT | 19 |
| Apo_A1_for | GGGACAGAGCTGATCCTTGA | 20 |
| Apo_A2_rev | CTGAGATCTGAGCCGAAAGG | 20 |
| Apo_B1_for | AGAGGCAGCAGGTTTCTCAC | 20 |
| Apo_B2_rev | TTCTCCTCTCCGAAGACAGC | 20 |
| Apo_C1_for | CTTCTTCACCACCTCCTCTGC | 22 |
| Apo_C2_rev | ATCTCCTCCTGCCACTTCTTCT | 22 |
| Apo_D1_for | AGAGATGAGCAAGGATCTGGAG | 22 |
| Apo_D2_rev | AAGCTGCTTCCCACTTTG | 19 |
| Apo_E1_for | GAGCGCTCTCGAGGAGTACA | 22 |
| Apo_E2_rev | TCCGCTGTGACTTCCTTTCTAC | 22 |

**Table 3: Primer for Long Range PCR**

| ID | sequence | length (bp) | product (kb) | ET |
|---|---|---|---|---|
| ApoA1long_for | AAGTTCCACATTGCCAGG | 19 | 2.4 | 2.5’ |
| ApoA1long_rev | TGAGCCTTGCTAAGGCAG | 19 | | |
| FABP1long_for | GCAGGTCAGTCGTGAAGAGG | 20 | 3.4 | 3.5’ |
| FABP1long_rev | TGTCACCCAATGTCATGGTC | 20 | | |
| Snapaplong_for | AGGGCTCTTGGTCTTTCTGG | 20 | 5 | 5’ |
| Snapaplong_rev | AGGGCTCTTGGTCTTTCTGG | 20 | | |

**Table 4 a,b,c Primer for SNaPshot Multiplex assays**

All primer for the SNaPshot Multiplex assays are listed in order to the run length in the electrophoresis. The ID column contains the internally used name and the official dbSNP identifier in brackets. The characters +/- indicate the direction of the primer in relation to the 5' transcription start point. In the 4th column concentrations of each primer in the used primer mixes are listed. The last column names the nucleotide substitution.

**Table 4 a: Primer for *APOA1* SNaPshot Multiplex assay**

| ID | sequence | length (bp) | µM in Mix | SNP |
|---|---|---|---|---|
| 3b- (rs5076) | TTTTTTTTTCCAGCC TATCAGGGGTGA | 27 | 0,8 | G/A |
| 4b- (rs7116797) | TTTTTATTTTTTAAAG-GAGACA-GAGCTGGGACT | 33 | 0,8 | G/A |
| 6d (rs5072) | TTTTTTTTTATTTTTCTG CCTGGA-GATCCCATTCC | 35 | 0,8 | G/A |
| 5- (rs5073) | TTTTTTTTATTTTTTCA CAATGTGGCTCTGTGA TCA | 36 | 1 | C/T |
| 7b- (rs2070665) | TTTTTTTATTTTTTTTT TATTTTTTA-CAGGCCTCTGCCCCCT | 42 | 0,8 | G/A |
| 1b- (rs12718463) | TTTTTTTTTTTTTTTTTTT TTTTTTTCGATCTTGGC CCTAAGACG | 45 | 0,8 | C/T |
| 8- (rs5070) | TTTTTATTTTT TATTTTTTTTTATTTTTT CACGGGGATTTAGGGA GAA | 48 | 0,8 | G/A |
| 9b- (rs5069) | TTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTC AGGTACCCAGAGGCCC | 53 | 0,6 | G/A |
| 10b+ (rs670) | TTTTTTATTTTTTTTTTTT TATTTTTTTTT TATTTTTTGTGAGCAG CAACAGGGCC | 56 | 1,5 | G/A |
| 2c+ (rs5081) | TTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTT TTAACAACTCCGTGCC CA GAC | 59 | 0,8 | A/T |

**Table 4 b: Primer for *FABP1* SNaPshot Multiplex assay**

| ID | sequence | length (bp) | µM in Mix | SNP |
|---|---|---|---|---|
| rs1d+ (rs13400963) | TTTTTTTTTTTTTCCAAAAAAC-TAAGCTGGGTACC | 35 | 1 | A/G |
| rs2b+ (rs7577401) | TTTTTTTTTTCATGTA-CAGGAGGATGTGGC | 30 | 0,8 | C/T |
| rs3+ (rs2919871) | TTTTTTTTTTTACCCA-CAAAATTTTTGTGAACCTCA | 36 | 0,8 | T/A |
| rs4b+ (rs13399861) | TTTTTTTTTTTTTTTTTGGTGGTGATAGGGAGA-GACTT | 38 | 0,8 | C/T |
| rs5+ (rs2241883) | TTTTTTTTTTTTTTTTTTTGAAGGTGA-CAATAAACTGGTGACA | 43 | 1 | A/G |
| rs6- (rs1545224) | TTTTTTTTTTTTTTTTTTTTTTTTTTGTTGCGCTGAGCAGAAAGG | 46 | 0,8 | A/G |
| rs7+ (rs2197076) | TTTTTTTTTTTTTTTTTTTTTTTTTTGGTCCTACAGGAAACCAGA | 49 | 1 | T/C |
| rs8+ (rs6711957) | TTTTTTTTTTTTTTTTTTTTTTTTTGAA-GAATAATA-CATACCTGGTTGGG | 52 | 1 | T/C |
| rs9+ (rs28452712) | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGCTTGTCAGCAG-CATC | 55 | 0,6 | A/G |
| rs10- (rs11678114) | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGACACAACTAGGCA-CACAGC | 58 | 1 | C/T |
| rs11- (rs1123605) | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAGGAAA-GAAGTTTGGGGGTTG | 60 | 0,6 | A/G |

**Table 4 c: Primer for *SNAPAP* SNaPshot Multiplex assay**

| ID | sequence | length (bp) | µM in Mix | SNP |
|---|---|---|---|---|
| rs1b- (rs3806234) | TTTTTTTTTTGAAAT-GAGTCTAGCGCGAAC | 31 | 0,8 | C/T |
| rs3c+ (rs35903030) | TTTTTTTTTTTTTTTTTTTTTTCGACCTTTTCGCGAAGG | 40 | 0,8 | G/C |
| rs4+ (rs16835479) | TTTTTTTTTTTTTTTTTTTTTCTGGAGTTCCTGCGACC | 39 | 1 | A/C |

| | | | | |
|---|---|---|---|---|
| rs5b+ (rs16835481) | TTTTTTTTTTTTTTTTTTTGCTTTCCCTCA-GAACAAGAGAT | 42 | 0,8 | G/A |
| rs12+ (rs12742195) | TTTTTTTTTTTTTTTTTTTTTTTCTCTTGCCTGTAATCCCAGC | 45 | 1,5 | A/G |
| rs6- (rs8532) | TTTTTTTTTTTTTTTTTTTTTTTTCAGGGCCACCTTCTGATC | 45 | 0,8 | C/T |
| rs7b+ (rs6692298) | TTTTTTTTTTTTTTTTTTTTTTAATTTTAAAATTCTTTGGGGAAAG | 52 | 0,8 | G/C |
| rs8b- (rs1802461) | TTTTTTTTTTTTTTTTTTTTTTTTTTTGCTGTTTCCTTGGCAACAC | 51 | 0,8 | A/T |
| rs9+ (rs16835489) | TTTTTTTTTTTTTTTTTTTTTTTTTTGAT-GAGCCTATGGACTCAGTAGC | 54 | 0,8 | A/C |
| rs13+ (rs12755522) | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACCAGCCTGAC-CAACATA | 57 | 1,5 | G/T |
| rs10b+ (rs41305092) | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCTATGGACTCAGTAGCA-CAAGTA | 57 | 1 | C/T |
| rs11- (rs7345) | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTCCTGAAGAAGG-TAAGCCTCA | 60 | 0,8 | A/C |

**Table 5: Primer for PCR product used in rs670 RFLP assay**

| ID | sequence | length (bp) | product (bp) |
|---|---|---|---|
| ApoA1_MspI_for | CCCACTGAACCCTTGACC | 18 | 235 |
| ApoA1_MspI_rev | CACCTCCTTCTCGCAGTCTC | 20 | |

**Table 6: Primer for Real-Time PCR**

| ID | sequence | length (bp) |
|---|---|---|
| PbMSPI_for | AAATAAATCTGGTTTGGTAGGAGAAGG | 27 |
| PbMSPI_rev | CCGCAGTTTGACAACCAGCAGTTGG | 25 |
| Pb18s_for | AAGCATTAAATAAAGCGAATACATCCTTAC | 30 |
| Pb18_rev | GGAGATTGGTTTTGACGTTTATGTG | 25 |
| GAPDH_for | CTTCTTGCACCACCAACTGC | 20 |
| GAPDH_rev | CGTTCAGTTCAGGGATGACC | 20 |

| | | |
|---|---|---|
| ApoMouse_for | GGAAAACTGGGACACTCTGG | 20 |
| ApoMouse_rev | CGGTAGAGCTCCACATCCTC | 20 |
| ApoHomo_for | ACCTGGAAAAGGAGACAGAGG | 21 |
| ApoHomo_rev | AGTGGGCTCAGCTTCTCTTG | 20 |
| ApoRatus_for | CACTGTGTACGTGGATGTGC | 20 |
| ApoRatus_rev | CGCATCTGATCGCTGTAGAG | 20 |

### 3.1.8 The Mild and Severe Malaria cohorts

#### 3.1.8.1 Mild Malaria Cohort dynamic follow-up study

The Mild Malaria cohort was designed as a dynamic cohort with entries and exits of subjects during the time of follow-up. Details in design of the mild-disease cohort study have been described in detail previously (*Williams et al., 2005, 2).* In brief, participants living within the area of Ngerenya and Chonyi of Kilifi District on the Kenyan Coast were monitored for clinical events from September 1998 to August 2001 by weekly active surveillance. Additionally, intercurrent clinical events were tracked through a dedicated outpatient clinic. Children born into study households were recruited at birth. Participants exited the study if informed consent was withdrawn, if they died or if they moved away for more than 2 months. In addition, 4 cross-sectional surveys to evaluate the prevalence of *P. falciparum* parasites were conducted to assess the prevalence of *falciparum* parasites during the study interval. The data include anthropometric data like age, sex, ethnic, genotypes of sickle cell and α-thalassaemia. Malariometric consists of species and parasite density calculated by number of *P. falciparum* parasites per microlitre of blood. It was calculated by reference to the number of parasites / 200 white blood cells or 500 red blood cells. If a blood count was available the parasitaemia was calculated by reference to the actual blood count. If no blood count was available the parasitaemia was calculated assuming a white blood cells of $8x10^9$/L or a red blood cells count of $5x10^{12}$/ L. Accessory data about seasons and if the patient was absent at the regular monitoring were included.

#### 3.1.8.2 Severe Malaria cohort study

The Severe Malaria cohort was designed to record children with severe malaria (cases) and children without known severe events (controls). The design and accomplishment has been described in detail in *Snow et al. (2000) and Williams et al. (2005, 2).* Briefly, between May 1992 and April 1995 children born within a defined rural study area to the north of the Kilifi research unit were recruited into a fixed birth cohort. The study area was the same but larger part of the same geographic area in which the Mild Malaria Cohort study was conducted. Admissions to the paediatric wards at Kilifi District Hospital (the closest hospital facility to the study area) from participants of this cohort until December 1997 were recorded. Between May and October 2000 2,695 resident surviving members of this cohort were invited to provide a blood sample. In 2,104 children typing for both HbS and α-thalassaemia was completed.

### 3.1.9 Databases

*Case Western Reserve University (CWRU) Genomics*
http://cmbi.bjmu.edu.cn/genome/candidates/SNPs.html

dbSNP
http://www.ncbi.nlm.nih.gov/projects/SNP/

*Ensembl Genome Browser*
http://www.ensembl.org/index.html

*Haplotter*
http://hg-wen.uchicago.edu/selection/haplotter.htm

*International HapMap Project*
http://www.hapmap.org/

*PharmGKB (The Pharmacogenomics Knowledge Base)*
http://www.pharmgkb.org/

*UCSC Genome Bioinformatics*
http://genome.ucsc.edu/

### 3.1.10 Software

| | |
|---|---|
| AutoDimer v. 1.0 | http://www.cstl.nist.gov/strbase/AutoDimerHomepage-/AutoDimerProgramHomepage.htm |
| BioEdit v 7.0.9 | http://www.mbio.ncsu.edu/BioEdit/bioedit.html |
| DNA Sequencing-Analysis v. 5.2. | Applied Biosystems, Darmstadt |
| Genemapper v. 4.0 | Applied Biosystems, Darmstadt |
| gPLINK v. 1.04 | http://pngu.mgh.harvard.edu/~purcell/plink/gplink.shtml |
| Haplotter | http://hg-wen.uchicago.edu/selection/haplotter.htm |
| Haploview v. 3.32 | http://www.broadinstitute.org/mpg/haploview |
| ImageJ v. 1.42q | http://rsbweb.nih.gov/ij/ |
| Microsoft Office 2008 | Microsoft Corporation, Santa Rosa, USA |
| PGA package | http://dceg.cancer.gov/bb/tools/pga |
| PHASE v. 2.1 | http://www.stat.washington.edu/stephens/phase.html |
| Primer3 | http://frodo.wi.mit.edu/primer3/ |
| Prism v. 5.0b | GraphPad Software Inc., La Jolla, USA |
| SBEprimer v.1.1β | http://www.zaik.uni-koeln.de/AFS/Projects/Bioinformatics/sbeprimer.html |
| STATA MP 10.1 | Stata Corporation, College Station, USA |
| Sweep v. 1.0 | http://www.broadinstitute.org/mpg/sweep/ |

## 3.2 Methods

### 3.2.1 Molecular biology

#### 3.2.1.1 Isolation and purification of nucleic acids

##### 3.2.1.1.1 Isolation of genomic DNA

Genomic DNA (gDNA) isolation from blood and hepatocyte culture was conducted with a silica-membrane-based DNA purification method from QIAamp DNA Blood Mini Kit (Qiagen). Cells containing DNA were lysed and the DNA was stabilized by a combination of buffers. After adding ethanol the solution was loaded on a column where the DNA was bound to the silica membrane. The bound DNA was eluted in $H_2O$ after washing with different buffers to remove debris and salts. The manufacture's protocol was followed for all steps.

##### 3.2.1.1.2 Isolation of RNA

Isolation of RNA was performed with a silica-membrane based method using Qiagen's RNeasy Mini Kit. The protocol provided by the manufacturer was followed.

#### 3.2.1.2 Purification of nucleic acids

##### 3.2.1.2.1 Ethanol precipitation

DNA is solved in water due to its polarity. Adding a salt to this solution the charge of the DNA is neutralized and it precipitates. This reaction is facilitated by ethanol and temperature. Centrifugation allows removing the supernatant. A subsequent washing step with 70% ethanol removes the salt from the DNA. Re-suspending in TE-Buffer allows long-time storage of solved DNA.

##### 3.2.1.2.2 Purification of GenomiPhi samples

For better a performance of downstream applications all GenomiPhi samples (3.2.1.4.3) were purified using ethanol precipitation. To each sample 20µl $dH_2O$ and 4µl of a pre-mixed sodium acetate (1.5M)/ EDTA (250mM) buffer was added. After

mixing, 100µl chilled ethanol (100%) was added and this mix was stored overnight at -20°C. Next day it was centrifuged at 12,000rpm for 20 minutes at 4°C. The supernatant was discarded and 150µl of chilled 70% ethanol was added. Again, it was centrifuged at 12,000rpm for 10 minutes and the supernatant was discarded. The pellet was dried at RT for 10 minutes, re-suspended in 40µl of TE buffer and stored at 4°C.

#### 3.2.1.2.3 Purification of sequencing products

Best purification of sequencing products (3.2.1.7) was achieved with a different protocol for ethanol precipitation:

80µl of HPLC $H_20$ and 10µl sodium acetate (3M) were added to the complete sequencing reaction. After adding 250µl of ethanol (100%) the mix incubated 20 minutes at RT followed by a centrifugation step of 20 minutes at 13,000rpm at RT.

The supernatant was discarded and the pellet was washed with 250µl of 70% ethanol. Again the mix was centrifuged at 13,000rpm for 10 minutes and the supernatant removed. The pellet was dried at 37°C for 5 minutes and re-suspended in 20µl Hi-Di formamide to prevent formation of secondary structures.

#### 3.2.1.2.4 Purification of PCR products

All PCR products used for sequencing were purified with the vacuum filtration application MultiScreen$^{HTS}$ Vacuum Manifold (Millipore). In the single sealed MultiScreen PCR 96 Filter Plates primer and other low molecular weight material were washed away whereas nucleic acids were hold back if vacuum was applied. Nucleic acids were resolved in $H_2O$ and retrieved by aspiration. PCR products were diluted in 100µl $dH_20$ respectively and loaded into the plate. Empty wells were sealed with plate sealing tape. Filtration time was around 13 minutes until no liquid in the well was left. 50µl of H2O were added to each well and PCR products were re-suspended. All taken steps were done in accordance to the manufacturer's protocol.

#### 3.2.1.3 Determination of DNA/ RNA concentration

All DNA and RNA concentrations were determined by photometry. Nucletotides absorb light at a maximum of 260nm whereas proteins have maximum absorption at 280nm.

1 µl of sample containing DNA or RNA was loaded on the spectrophotometer NanoDrop 1000 (Thermo Scientific). The manufacturer's recommended default settings were used for absorption measurement and calculation of the nucleic acid concentration.

#### 3.2.1.4 Amplification of nucleic acids

##### 3.2.1.4.1 Cleaving of gDNA

To clear DNA contamination in RNA prior to cDNA synthesis the samples were incubated with *DNase I*. This enzyme cleaves double stranded DNA in 5' phosphorylated oligodeoxynucleotides. For a better performance the *TURBO DNase* (Ambion) was used with the provided protocol of the manufacturer but with an increased incubation time of 45 minutes at 37°C.

##### 3.2.1.4.2 Transcription of RNA to cDNA

The enzyme Reverse Transcriptase is used to transcribe RNA to its complementary cDNA as template for downstream real-time PCR. Here, the RETROscript First Strand Synthesis Kit (Ambion) was used. First the RNA was heat denaturated to clear secondary structures, then cDNA was amplified by the reverse transcriptase using the RNA as matrices and at last the enzyme was heat inactivated.

For transcription of total RNA 1µl oligo dT and 1µl random decamer primer were used in the reaction. In all steps manufacturer's protocol was followed.

##### 3.2.1.4.3 Whole Genome Amplification

Whole genome amplification (WGA) amplifies the whole genome of a sample and thus increases the amount of available DNA for further experiments. The proofread-

ing *Phi29* DNA polymerase amplifies nucleic acids strands via a strand displacement reaction.

The whole gDNA of all samples from the Mild Malaria Cohort were amplified with the GenomiPhi DNA Amplification Kit (GE Healthcare).
Therefore, for each sample 1µl of gDNA was mixed with 9µl of provided sample buffer containing random-sequenced hexamer primer. These primer align randomly thus allowing random amplification. The mix was denatured at 95°C for 3 minutes and immediately cooled down to 4°C on ice allowing the primer to bind. Pre-mixed solution of 9µl reaction buffer and 1µl of *Phi29* polymerase was added to the cold sample and incubated for 2 hours at 30° C. In this time dNTPs in the reaction were converted into high-molecular-weight fragment copies of the template DNA by the polymerase. The latter was inactivated by a post-amplification heat inactivation step at 65°C for 10 minutes. All GenomiPhi products were stored at 4°C until purification (3.2.1.2.2).

#### 3.2.1.4.4 Polymerase Chain Reaction

Polymerase chain reaction (PCR) was used to amplify DNA fragments *in vitro*. For selection of the desired segments two primer have been used complementary to the template DNA. The reaction takes place in 3 steps:
At the first step the hydrogen bonds between the DNA strands are released at 94°C. Then the reaction is cooled down to the temperature between 50-61°C at which the primer anneal. At the last step, so called elongation step, the 3' end of the primer were elongated with dNTPs by the polymerase. All 3 steps were repeated in cycles that allow copying the fragment million times.

##### 3.2.1.4.4.1 PCR for *APOA1* re-sequencing

For the re-sequencing of *APOA1* 6 sets of overlapping DNA fragments were amplified with PCR from 50 random individuals from Kilifi at the East Coast of Kenya. The total product size was ~2,4kb. Polymerase and ingredients from Fermentas were used according to the manufacturer's protocol. For all PCR a single reaction without

DNA was used as a negative control. Additionally, a positive control of known performance was used in all established PCR.
All sequences were re-assembled in the BioEdit v. 7.0.9 software (3.2.3.2).

The Master Mix for each reaction of 50µl was prepared as follows:

| | |
|---|---|
| $H_2O$ | 36,5µl |
| 10x PCR-Buffer | 5µl |
| $MgCl_2$ | 3µl |
| dNTPs | 2µl |
| 5' primer | 1µl |
| 3' primer | 1µl |
| *Taq* | 0.5µl |
| DNA | 1µl |

The reaction was run on a Mastercycler epgradient S (Eppendorf) cycler with following programme:

| | | |
|---|---|---|
| 95°C | 3min | |
| 95°C | 30sec | ⎫ |
| 57°C | 50sec | ⎬ 33 cycles |
| 72°C | 1min | ⎭ |
| 72°C | 10min | |
| 4°C | hold | |

#### 3.2.1.4.4.2 Long Range PCR for SNaPshot assay

The Long Range PCR Kit (Qiagen) was used for amplifying the whole gene of *APOA1*, *FABP1* and *SNAPAP* in the subset of the Mild Malaria Cohort. The amplified products included all target SNPs of each gene. They were used as a template for the SNaPshot assays (3.2.1.8).

The protocol provided by the manufacturer was followed. A final volume of 25µl per reaction was composed of:

| | |
|---|---|
| $H_2O$ | 19.85µl |
| LongRange-Buffer | 2.5µl |
| dNTP-Mix | 1.25µl |
| 5' primer | 0.2µl |
| 3' primer | 0.2µl |
| LongRange-polymerase | 0.2µl |
| DNA | 0.8µl |

The PCR was performed on ABI 9700 Thermal Cycler:

| | | |
|---|---|---|
| 93°C | 3min | |
| 93°C | 15sec | 38 cycles |
| AT | 30sec | |
| 68°C | ET (see table 3) | |
| 4°C | hold | |

#### 3.2.1.4.4.3 PCR for rs670 RFLP assay

PCR with the 5-Prime-MasterMix (5-Prime) was used to amplify the template for the SNP rs670 RFLP assay (3.2.1.6).

A single reaction with a Volume of 20µl consisted of:

| | |
|---|---|
| $H_2O$ | 10µl |
| 5 Prime MasterMix | 8µl |
| 5' primer | 0.6µl |
| 3' primer | 0.6µl |
| DNA | 0.8µl |

The reaction was performed on an ABI 9700 Thermal Cycler with the cycle protocol from the PCR for *APOA1* re-sequencing (3.2.1.4.4.1).

#### 3.2.1.5 Gel electrophoresis

Gel electrophoresis was used to visualize DNA fragments from PCR or RFLP assays. In an electric field the negatively charged DNA is forced to the anode. The DNA fragments were embedded in a matrix of agarose polymers. Separation is achieved due to movement of DNA fragments in this matrix proportional to their size. The agarose powder was dissolved in 1xTAE-Buffer by heating and 1µg/ml ethidium bromide was added. For fragments smaller than 500bp a 2% and for bigger sizes a 1% agarose concentration was used. The DNA samples were mixed with 0.3 volumes of Loading Dye (Fermentas) and loaded into the gel pockets. 8µl of 1kb or a 50bp Marker (Fermentas) was used for sizing and the gel was run at 5V/cm in a gel chamber (Biorad).

The intercalated ethidium bromide in the DNA was excited using UV-light and photographed with the Gel Doc XR or Chemi Doc XRS (Biorad).

#### 3.2.1.6 MspI Restriction Fragment Length Polymorphism Assay

A Restriction Fragment Length Polymorphism assay (RFLP) was used to genotype samples from the Severe Malaria Cohort for the SNP rs670. Endonucleases are widely used to cleave phosphodiester bonds in DNA strands. Here, the endonuclease MspI (New England Biolabs) was used that recognize and cleaves (^) a sequence consisting of 5'-CCG^G-3'.

A PCR product of 235bp was amplified in a PCR with 5-Prime-Mastermix (3.2.1.4.4.3). 5µl of this product was digested overnight with 3U of MspI in 4µl $H_2O$ and 1µl NEBuffer 4 (10x) at 37°C.

Next day the fragments were separated in 2.5% agarose gels by electrophoresis and visualized with the GelDoc XR or ChemiDoc XRS. The PCR product was only cleaved when the G genotype (reverse strand: C) was present.

For all master mixes a negative control and a positive control for all the three genotypes (AA, GA or GG) were performed.

#### 3.2.1.7 Sequencing of PCR products

Here, the chain terminator sequencing method developed by *Sanger et al.* (*1977*) was used. Comparable to PCR the amplification is performed with primer, a poly-

merase and dNTPs. Latter contain a low concentration of labelled chain terminating dideoxyribonucleotide triphosphate (ddNTPs). Integration of these leads to an amplification termination that result finally in a mixture of DNA fragments with varying length that are terminated only at positions where ddNTP are integrated. The fragments are then size-separated by electrophoresis and the sequence is determined by the run length and different labelled ddNTPs, which fluoresces at a different wavelength. Contrary to PCR the sequencing reaction is not exponential.

All sequencing was performed by GATC (Konstanz) or with the 3130 Genetic Analyzer (ABI) in combination with the Big Dye Terminator Mix v1.1 Kit (ABI). For latter the purified PCR product (3.2.1.2.4) was added to a master mix with a total volume of 10µl: The final master mix was composed of:

| | |
|---|---|
| HPLC $H_2O$ | 4.5µl |
| Big Dye Sequencing Buffer (5x) | 1.5µl |
| Primer (5nM) | 1µl |
| Big Dye Terminator Mix v1.1 | 1µl |
| PCR product | 2µl |

The PCR was set up in a Mastercycler epgradient S (Eppendorf):

| | | |
|---|---|---|
| 96°C | 1min | |
| 96°C | 10sec | ⎫ |
| 58 | 5sec | ⎬ 25 cycles |
| 60°C | 4min | ⎭ |
| 4°C | hold | |

After completion the product was purified using ethanol precipitation (3.2.1.2.3) and the pellet was re-suspended in 20µl Hi-Di formamide. Then all reactions are loaded into the 3130 Genetic Analyzer. The manufacturer's default settings for Big Dye v1.1 in combination with a 36cm capillary and POP7 polymer were used for electrophoresis. The results were analyzed with ABI DNA Sequencing Analysis Software v. 5.2.

### 3.2.1.8 Primer extension assay with SNaPshot Multiplex Kit

In a primer extension assay with the SNaPshot Multiplex Kit (ABI) a primer binds to the template and only a single nucleotide complementary to the matrix is incorporated at 3' end of the primer (Figure 6). Contrary to PCR or sequencing the single incorporation is caused by exclusive presence of fluorescent labelled ddNTPs in the SNaPshot Multiplex reaction. This method was approved to be accurate, cost-effective and suitable for large amounts of clinical data (*Pati et al., 2003; Ben-Avi et al., 2004*).

The primer with labelled nucleotide is separated by electrophoresis. Multiplexing is achieved by varying primer length.

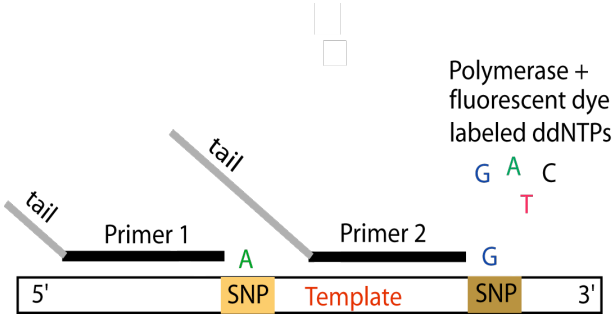



**Figure 6: Primer extension with SNaPshot Multiplex Kit**
SNP primer anneals 5' of SNP and is extended by a single base unit. Multiplexing is achieved by using ascending lengths of poly-T tails, which results in different electrophoretic mobilities thus all primer in the reaction can be separated by their size.

In the SNapShot Multiplex assay the product from the Long Range PCR (3.2.1.4.4.2) was used as the template for each gene. The single reaction contained 10 primer for *APOA1*, 12 for *FABP1* and 12 for *SNAPAP*. Differentiation between the primer was possible due to a varying length of a poly-T tail at 5' of the primer. Separation of the SNaPshot products was achieved by electrophoresis on a 3130 or 3130xl Genetic Analyzer. Final Analysis was conducted with the Genemapper software v. 4.0 (ABI).

For the SNaPshot Multiplex reaction 5µl of the template were digested with 2µl of a pre-mixed solution containing 3U Exonuclease 1 and 2U Shrimp Alkaline Phospotase (SAP) to degrade primer and dNTPs from prior PCR. The Exonuclease

1 digests single stranded DNA whereas SAP catalyses the release of 5'- and 3'-phosphate groups resulting in un-incorporatable dNTPs/ ddNTPs.
The mix was incubated at 37°C for 65 minutes and enzyme was heat inactivated at 75°C for 15 minutes. After cooling down to 4°C 3µl were transferred to a pre-cooled mix that consisted of 2.5µl SNaPshot Multiplex Ready Reaction Mix, 1µl of the specific primer-mix and 3.5µl HPLC $H_2O$.
Subsequent thermal cycling was perfomed on ABI Thermal Cycler 9700 or Eppendorf Fast Cycler:

| | | |
|---|---|---|
| 96°C | 1min | |
| 96°C | 10sec | ⎫ |
| 50 | 5sec | ⎬ 25 cycles |
| 60°C | 30sec | ⎭ |
| 4°C | hold | |

To avoid co-migration of unincorporated ddNTPs along with the primer fragments the whole reaction mix was digested with 2U SAP for 60 minutes at 37°C with adjacent enzyme heat inactivation at 65°C for 15 minutes.

0.8µl of the digested SNaPshot product was mixed with 8.9µl of Hi-Di formamide and 0.3µl LIZ 120 size standard (ABI) and loaded into the 3130 or 3130xl Genetic Analyzer. In all master mixes a negative control without template DNA was performed to test for contamination.
The electrophoresis was run with a 36cm capillary and POP7 polymer on the Genetic Analyzer. The manufacturer's recommended program set up was used. Only the run time was reduced to 600 seconds to speed up the assay.

### 3.2.2 Cell culture

#### 3.2.2.1 Hepatoma cell culture

The human hepatoma cell line Huh7 (provided by *AK. Müller*) and the murine hepatoma cell line Hepa1.6 (provided by *G. Montagna*) were cultured *in vitro*. They were kept in cell culture flasks in 15ml DMEM complete medium in an incubator with 37ºC and 5% $CO_2$.

For a stable growth the adherent cells were splitted when the flask bottom was confluently overgrown. Therefore, medium was removed and the cells were washed with HBSS medium to remove serum remnants. For detaching the hepatocytes were incubated with 3ml Trypsin/EDTA solution for 3 minutes at 37ºC. Then the re-suspended cells were centrifuged at 2000rpm for 3 minutes and the supernatant was removed. The cell pellet was washed in 6ml HBSS, centrifuged and the supernatant was again discarded. The pellet was re-suspended in 3ml DMEM complete. A 1ml aliquot was transferred into a new flask. All media were pre-warmed to 37ºC using a water bath.

#### 3.2.2.2 Determination of cell concentration

Counting the hepatocytes was necessary for plating a specific cell number in wells for the experiments. Washed and re-suspended hepatocytes (3.2.2.1) were stained with Trypan blue in a 1:10 dilution. This vital staining allows counting of living cells because cells with a functional membrane remain uncoloured in the blue background. The living cells were counted in a chamber with Neubauer rulings. The concentration of cells were calculated by:

$$\frac{cells}{ml} = \left( \frac{counted\ cells\ in\ 4\ squares}{4} \right) * dilution\ factor\ * 10^4$$

#### 3.2.2.3 Freezing and thawing of hepatocytes

An aliquot of hepatocytes (3.2.2.1) were centrifuged for 3 minutes at 2000rpm and the pellet was re-suspended in the *freezing solution*. Immediately the suspension was frozen at -80ºC. Later the frozen cells were stored in liquid nitrogen.

For thawing the frozen aliquot was thawed quickly in a water bath at 37ºC and added to pre-warmed 6ml of DMEM complete. After centrifugation at 2000rpm for 3 minutes and supernatant removal the pellet was re-suspended in 3ml DMEM complete and then added to a cell culture flask with 10ml DMEM complete. The flask was stored in the incubator at 37ºC.

#### 3.2.2.4 Transfection of hepatocytes

For an APOA1 overexpression model the hepatoma cell line Huh7 was transfected. The Huh7 cell line was provided by *AK. Müller* (Heidelberg) and the pcDNA3.1/V5-His A vector was provided by *G. Montagna* (Max PIanck Institute, Berlin). The used vector expressed *ApoA1* cDNA of *R. norvegicus*. *ApoA1* expression was controlled by the Cytomegalovirus-Promotor (CMV). The HiFect Kit (Lonza) was used for transfection of the vector into the cell lines. 24 hours before transfection 100,000 hepatocytes per well were plated in 500µl DMEM complete in a 24well plate. The transfection was done according to the manufacturer's protocol for 24well plates. Five hours after transfection the medium was replaced with DMEM complete and another 24 hours later the medium was changed again including 800µg/ml Geneticin G-418 (Gibco Invitrogen) as a selection marker. Geneticin G-418 blocks the polypeptide synthesis in eukaryotic and prokaryotic cells by inhibition of the elongation step. The vector contains the neo gene from Tn5 encoding an aminoglycoside 3'-phosphotransferase that results in resistance to G-418.

A mock-transfected Huh7 was created the same way but with a different vector that was not expressing ApoA1. The transfected Hepa1.6 cell line was provided by *G. Montagna*. Successful translation of ApoA1 was verified by Western blot (*Montagna G, personal communication*).

#### 3.2.2.5 APOA1 overexpression assay

##### 3.2.2.5.1 Isolation of *P. berghei* sporozoites

*Anopheles stephensi* infected with *P. berghei* were provided by AK Müller. For isolation of *P. berghei* sporozoites the salivary glands from mosquitos were isolated 17 days after the infection with *P.berghei*. Therefore, the salivary glands were dissected out of dazed mosquitos by tweezers, stored in 200µl RPMI 1640 medium with 3% BSA on ice and homogenized with a pestle. The homogenate was centrifuged for 3 minutes at 1000rpm and the supernatant was transferred to a new Eppendorf tube. The pellet was re-suspended in 200µl RPMI 1640/ 3% BSA medium and centrifuged again at 1000rpm for 3 minutes. Both supernatants were merged and the sporozoite concentration in a 1:10 dilution was determined in count chamber with Neubauer rulings (3.2.2.2).

##### 3.2.2.5.2 Infection of hepatocytes with *P. berghei* sporozoites

Hepatocytes were plated 24h prior to infection with *P. berghei*. For later RNA isolation 100,000 cells per well were plated into 24well plates whereas 30,000 cells per well were plated into Labtek slides. Only Panserin PX10 (Pan-Biotech) serum-free medium was used.

The medium was removed before the sporozoites infection. In 24well plates 50,000 sporozoites per well and in the Labtek slides 10,000 sporozoites per well were added in 100µl and 50µl Panserin PX10 medium, respectively. An incubation time of 90 minutes at 37ºC allowed the sporozoites to invade the hepatocytes. Then the medium was replaced with Panserin PX10 containing *round up solution* (1:1000) to prevent contaminations. Every 24 hours the medium was renewed.

#### 3.2.2.6 APOA1 overexpression assay analysis

##### 3.2.2.6.1 Isolation of RNA for real-time PCR

At 24, 48 and 65 hours post infection (pi) the hepatocytes were harvested out of the 24well plates for RNA isolation (3.2.1.1.2). The cells were washed with HBSS medium and then incubated in Trypsin/ EDTA solution for 3 minutes at 37ºC. The cells were added to 1ml DMEM complete, centrifuged at 2000rpm for 3 minutes. DMEM was removed and every pellet was re-suspended in 350µl RLT Buffer (Qiagen) containing 10µl/ml β-mercaptoethanol. They were stored at -80ºC.

##### 3.2.2.6.2 Quantification of cDNA using Real-Time PCR

Real-Time PCR enables detection and quantitative measurement of the amount of specific DNA molecules in a sample. Based on the principle of PCR, additionally the amount of dsDNA is detected after every amplification cycle. For detection of the amplified product SYBR Green was used. It forms a complex with dsDNA that emits green light if excited with light at 494 nm. The cycle was employed for quantification when the fluorescence exceeded the threshold ct. *Gapdh (H.apiens* or *M.musculus)* was measured for normalization and relative quantification to *APOA1 (R.norvegicus*), *APOA1 (H.sapiens), 18s (P.berghei*) and *msp1 (P.berghei)*. The Master mix with $H_2O$ and the original RNA was used as negative controls. All reactions and controls were tested in triplicates. The primer were tested for non-specific amplifications and primer dimer. Therefore, a single dissociation cycle was run subsequently to a completed real-time PCR. The dissociation curve showed the melting point of the amplified product. Primer with a dissociation curve showing more than one peak were excluded from the assay.

Quantification of targeted cDNA copies from the APOA1 overexpression experiment was performed on the 7500 Real-Time System (ABI) with Power SYBR Green PCR Master Mix (ABI). Fluorescence detection was set at the 60°C step.

A single reaction mix of 25µl was composed of:

| | |
|---|---|
| $H_2O$ | 10µl |
| Power SYBR Mix | 12µl |
| 5'primer (5nM) | 1µl |
| 3'primer (5nM) | 1µl |
| cDNA | 1µl |

Programme for real-time PCR:

| | | |
|---|---|---|
| 95°C | 3min | |
| 95°C | 10sec | ⎫ |
| 55°C | 15sec | ⎬ 40 cycles |
| 60°C | 45sec | ⎭ |
| *Dissociation step:* | | |
| 95°C | 15sec | |
| 60°C | 1min | |
| 95°C | 15sec | |

Each reaction was performed in triplicates and the mean ct value was calculated. The delta ct (Δct) and the delta delta ct (ΔΔct) were calculated by:

Δct=18s-GAPDH and ΔΔct=Δ ApoTrans – Δ WT

**3.2.2.6.3 IFA staining and microscopy**

The Indirect Fluorescent Antibody (IFA) staining was used to detect liver cells stages. In this method an antibody (AB) interacts with a specific antigen and is labelled with a second fluorescent AB.

The medium in Labtek slides was removed at 24, 48 and 65 hours pi. The cells were fixed with -20ºC cold methanol for 10 minutes at RT and carefully washed twice with FCS/PBS (1:100). Then they were blocked with FCS/PBS (1:10) and stored at 4ºC overnight. For staining of liver cell stages the monoclonal anti-*P. berghei* HSP70 antibody (α-PbHSP70) from *M. musculus* was added in a dilution of 1:300 in FCS/PBS (1:10).

After an incubation time of 45 minutes at 37ºC the slides were washed 3 times with FCS/PBS (1:100). Then the anti-mouse AB (α-mouse-IgG-Alexa 488) in a 1:300 dilution in FCS/PBS (1:10) was added. Again the slides were incubated at 37ºC for 45 minutes and subsequently washed 3 times with FCS/PBS (1:100). The nuclei of the hepatocytes and parasites were stained with Hoechst (1:1000) for 15 minutes at RT in FCS/PBS (1:10). Again the slides were washed 3 times with FCS/PBS (1:100). The slides were mounted with glycerol/PBS (1:10) and cover slips. The latter were sealed with nail polish.

Liver cell stages were counted under the microscope Axiovert 100M with a 40x magnification. For measurement of liver cell stage sizes they were photographed in the Confocal Laserscanning Mikroscope LSM 510 using the 63x objective. 25 liver cell stages per well and time point were photographed and their area was determined in ImageJ v. 1.42q (National Institute of Health). The means of liver stage area size were compared for significance with a non-parametric unpaired two-tailed *Student's t-test*. All graphs and calculations were done in Prism v. 5.0b (GraphPad Software Inc.).

### 3.2.3 Bioinformatical methods

#### 3.2.3.1 Database screening

Different databases were used to examine and download information about DNA sequences, genomic regions, genes and SNPs. DNA sequences were obtained from the ENSEMBL Genome Browser and NCBI databases.

SNP information was gathered at the NCBI dbSNP database and completed with information from the CWRU Genomics, PharmGKB, UCSC Genome Browser and the HapMap *(The International HapMap Consortium, 2007)* databases.

#### 3.2.3.2 DNA sequence analysis and manipulation

DNA sequences downloaded from databases or from sequencing were analysed with the BioEdit v. 7.0.9 software (*Hall et al., 1999*). All re-sequenced fragments from *APOA1* (3.2.1.4.4.1) were reassembled and aligned in BioEdit using *APOA1*

DNA sequences from NCBI (NM_000039) and Ensembl (ENSG00000118137) as matrices.

#### 3.2.3.3 Calculation of tag SNPs

For minimizing the number of SNPs in the SNaPshot Multiplex assays and thus reducing genotyping costs, SNPs in the HapMap database were tagged. Tag SNPs are a subset of all SNPs in a genomic region. Based on LD they are able to capture most information of all SNPs in the genomic region where one SNP allele is unique to a particular haplotype (*Johnson et al., 2001*). *De Bakker et al. (2006)* verified the usage of HapMap DNA samples for selecting tag SNPs in many different ethnics around the world.

Tag SNPs in the HapMap Yoruba population (YRI) and in U.S. residents with northern and western European ancestry collected (CEU) by the Centre d'Etude du Polymorphisme Humain (CEPH) data were calculated using the implemented Tagger tool (*de Bakker et al., 2005*). Within each gene they were calculated with an $r^2 \geq 0.8$ and a minor allele frequency (MAF) cut-off of 0.05. Only tag SNPs were chosen that showed comparable tagging in the CEU and YRI dataset.

#### 3.2.3.4 Primer Design

##### 3.2.3.4.1 Primer Design for PCR and real-time PCR

Except for the SNaPshot Multiplex assay, all primer were designed using the Primer3 web-tool (*Rozen et al., 2000*). For designing the DNA sequences of targeted genome regions were copied from ENSEMBL or NCBI databases into the online accessible Primer3 interface.

##### 3.2.3.4.2 Primer Design for SNaPshot Multiplex assay

Primer for the SNapShot Multiplex assays were designed with the SBEprimer v.1.1β software (*Kaderali et al., 2003*). For every target gene all SNPs flanked by 20bp in 5' and 3' directions were imported into the programme. The default settings were used

and with the “slow check” option possible forward and reverse primer were calculated. Out of all these primer appropriate combinations were chosen. Hence SBEprimer is used for multiplex assay with attached primer all chosen combinations were revised for possible formations of primer-dimer and hairpins with the Autodimer v.1 software (*Vallone et al., 2004*) in its default set up.
A passed primer was first tested solely in a SNaPshot Multiplex reaction for correct priming in a sequenced control template. Then all primer for each target gene combined in one mix were tested for primer interactions and artefacts.

The Genemapper v. 4.0 software (ABI) was used to analyze data from the SNaPshot Multiplex assays. The assay data were imported directly into the software and manufacturer’s settings for analyses were applied. For the automatic allele calling I created a panel for each primer combination of the target genes. Therefore, every primer was tested solely and a bin for every allele was generated. All these bins (two bins per primer) were implemented in the panel for each of the target genes. A bin identifies the run length and the fluorescence of the single incorporated labelled ddNTP in the electropherogram from each run.
Genemapper v. 4.0 with applied panels detected and listed the specific alleles for every primer automatically. All lists, containing the detected alleles, were exported to the software Excel 2008 (Microsoft Corporation).
All electropherograms were inspected two times manually to account for possible electrophoretic shifts due to age and condition of the POP7 polymer and capillaries.

#### 3.2.3.5 Genotype data validation

Genotype data were transformed into the PED-file format. Hence, I used data from unrelated individuals thus the mother’s and father’s identifiers were set to “0” and the individual ID to “1”. The individual sample name was defined in the pedigree name column. Details are available at the gPlink v 1.04 manual.
For summary statistics and data file validation the genotype PED files were imported into the gPLINK v. 1.04 software (*Purcell et al., 2007*). The genotype files were searched for input errors, duplicates and missing data.

#### 3.2.3.6 Analysis of Mild Malaria cohort genotype data

The Mild Malaria Cohort database was merged with the validated genotype data (3.2.3.5**)** from the SNaPshot Multiplex assay for every gene respectively. Criteria for merging were the individual sample names (serial) that excluded the not genotyped samples of the original cohort.

##### 3.2.3.6.1 Poisson regression analysis

The incidence rate (IR) is defined as the number of events (numerator) per person-time units (denominator). It describes the rate that the event occurs in the population at risk. In the analysis of the Mild Malaria Cohort the presence of *P. falciparum* asexual parasites on thick blood smears from the patient with fever was defined as event.

An established method to summarize incidence rates for multiple events with different individual follow-up time is the Poisson regression (*Frome et al., 1985*). It assumes the response variable has a Poisson distribution meaning that events occur in a fixed period of time and independently of the time since the last event. An advantage is its independence from normally distributed data. Additionally, it is capable to test for complex interactions with covariates.

The Poisson regression model expresses the logarithm of the event rate as a linear function of predictor variables.

In this model the incidence rate is defined for the *j*-th observation by:

$$r_j = e^{\beta_0 + \beta_1 x_1 + ... + \beta_k x_{k,j}}$$

The expected number of events ( $C_i$ ) and exposure ( $E_i$ ) is calculated by:

$$C_i = E_i e^{\beta_0 + \beta_1 x_1 + ... + \beta_k x_{k,j}} = e^{\ln(E_i)\beta_0 + \beta_1 x_{1,i} + ... + \beta_k x_{k,j}}$$

The incidence rate ratio (IRR) is obtained by comparing the IR between the genotype groups:

$$e^{\beta_i} = \frac{e^{\ln(E)i\beta_0 + \beta_1(x_1+1)+...+\beta_k x_{k,i}}}{e^{\ln(E)i\beta_0 + \beta_1 x_1 +...+\beta_k x_{k,i}}}$$

All following calculations were done in STATA v. 10.1 (STATA Corporation):
Analogous to *Wambua et al*. *(2006)* the IRR and the effects of explanatory variables on incidence rate was estimated by Poisson regression for all SNPs in each gene. The follow-up cohort attrition bias was compensated using person-times for every individual. Additionally, several covariates were included to the analysis to test for interactions. Therefore, all genotypes and covariates were converted to indicator variables by the *xi*-option in STATA. Influence between covariates and genotypes were tested using the in STATA implemented *xi-interaction expansion.* Furthermore possible interactions of the genotypes with sickle cell trait (HbAS) were determined with the *Wald test s*tatistics.

The unadjusted IRR was calculated. Additionally, in a second calculation the IRR was adjusted for covariates like age, sex, ethnic group, season of year, sickle-cell disease (HbAS) and $\alpha$+–thalassaemia with the STATA *xi-interaction expansions.* Observations where the subjects were not available at the weekly monitoring were not included in the statistics.

A goodness-of-fit-test was performed using the *estat gof*-option in STATA to validate the Poisson regression model for each result. Based on deviance statistics the Poisson regression model is inappropriate if the p-value of this test is significant.

The relationships between the SNP genotypes and the log-transformed parasite densities in the Mild Malaria Cohort dataset were determined using linear regression analyses. When the dependent is a quantitative trait, linear regression is a adequate model for a association analysis in which a line is fit to the response in terms of the predictor's count value and a p-value is computed for goodness of fit.

For both analyses HbAS and $\alpha^+$-thalassaemia positive children were excluded due to association of both phenotypes on asexual *falciparum* parasite densities in the blood (*Williams et al., 2005 2; Wambua et al., 2006)*.

In all calculations data of sickle cell genotypes from the Mild Malaria Cohort was used as a reference and positive control. The *robust variance estimation* option was used in un- and adjusted IRR calculations to account for potential clustering of multiple events within the same individuals (*Lin and Wei, 1989*).

#### 3.2.3.6.2 Multiple Failure-Time Analysis

Failure-time analysis examines the relationship of the survival distribution to covariates. Multiple failure-time analyses are widely used for time-to-failure data when multiple failures/ events were measured in the same subject. Like in the Poisson regression the presence *of P. falciparum* asexual parasites with fever was defined as failure. Hence the parasite infection could only have occurred one by one a method for ordered failure data was used. The Andersen-Gill model (*Andersen and Gill, 1982*) for multiple ordered failures analyses was performed. This approach measures time to first failure, time from previous to next failure and finally to the time point of observation dropout. Test of association was done with a Cox regression (*Cox, 1972*). It evaluates the effects of explanatory variables on the hazard rate using a proportional hazard regression model. The hazard rate reflects the probability of the event of interest happening in time and is defined for the *j*th subject in the data:

$$h_i(t \mid x_j) = h_0(t)\exp(x_j\beta_x)$$

The regression coefficient $\beta_x$ is estimated from the data and the baseline hazard $h_0(t)$ is left unestimated. An advantage in this model is that no assumptions about the hazard over time have to be made.

The calculated hazard rates are frequencies or probabilities that a failure occurs at a given time point and the hazard ratios were an estimation of the ratios of the hazard rate between the genotype groups.

The validated data from the Mild Malaria Cohort was transformed into the required format of the Andersen-Gill model. For analysis all genotypes were converted in indicator variables by the *xi*-option in STATA. Again the *robust variance estimator* was used in regard of multiple measurements in the same individuals.

The in STATA v. 10.1 implemented Kaplan-Meier estimator was used to plot the failure function of the different genotypes from the multiple time-to-failure data. With every failure the subject is counted as lost but the subject re-enters immediately in the analysis until it finally drops out with its observation end. In all calculations sickle cell disease data was used as a reference.

#### 3.2.3.7 Power Calculation for the Severe Malaria Cohort genotyping

Statistical power calculations inform and help to design case-control studies. They state the actual sample size needed to find a true genotype-phenotype correlation under the study specifications *(Gordon and Finch, 2005)*. Additionally, they can help explaining possible causes for false or ambiguous results.

To measure an effect of rs670 severe malaria power calculations were done with the *Power for Genetic Association Analyses* (PGA) software *(Menashe et al., 2008)*. The program assumes that SNPs are biallelic and in Hardy-Weinberg equilibrium (HWE). It was used to determine the required number of genotyped cases and controls to appoint the influence of rs670 genotypes in the Severe Malaria Cohort Study.

The threshold for the minimum sample size was set where the statistical power exceeded 0.8 and α 0.05 for association analysis. The power reflects the Type II error (β) that describes a false negative result. β is defined as 1-power. A false positive result is defined as type I error (α).

The hazard ration calculated in the multiple failure-time analyses was used as relative risk (RR).

#### 3.2.3.9 Association analysis of rs670 in the Severe Malaria Cohort

All validated rs670 genotype data from the RFLP assay (3.2.1.6) were validated (3.2.2.5) and were analyzed for association using $\chi^2$ –test, Fisher's exact test and logistic regression. For all analyses the data was transformed into a binary dataset with children with past severe malaria events as 1 and children without as 0.

#### 3.2.3.9.1 Association analysis of rs670 using $\chi^2$ – and Fisher's exact test

Both, the $\chi^2$ – and Fisher's exact test use contingency tables in the analysis. The $\chi^2$ –test calculates deviations from a $\chi^2$ –distribution. If a variable equals the sum of the squares of a set of statistically independent standard Gaussian random variables it is in $\chi^2$ –distribution and the null hypothesis is true, meaning the observed equals the expected distribution. Contrary to $\chi^2$ –test the advantage of Fisher's exact test is the independence of large sample sizes without relying on assumptions. For test of association using allele frequncies the "gencc" command in STATA was used for calculating odds ratios (OR) with Cornfield correction (*Cornfield, 1956*). Because of the large sample size when using alleles the Cornfield approximation is a better approach compared to the $\chi^2$ –test (*Kung-Jong and Chii-Dean, 2003*).

#### 3.2.3.9.2 Association analysis of rs670 using logistic regression

The logistic regression calculates the prediction of the probability of occurrence of an event by fitting data to a logistic curve. It is a generalized linear model of a binomial regression model that predicts the probability of a binary event (e.g. sick or healthy) in relation between to one or more risk factors (e.g. age, ethnics) and an outcome. The logistic function is given by:

$$P = \frac{1}{1 + e^{-z}}$$

where *P* is the probability of the outcome and *z* is the exposure to risk factors that is defined by

$$z = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

with β describing the coefficients of the variants of $x_1 \ldots x_k$.

The odds are calculated by: $odds = \frac{P}{1 - P}$ .

The odds ratio is calculated to compare the odds across groups.

Logistic regression was used to test and adjust the association of rs670 alleles with severe malaria. The adjusted logistic model accounted for age, ethnic group, season of year, sickle-cell disease (HbAS) and $\alpha$+–thalassaemia. Therefore, all covariates

were used with the *x-expansion* in STATA v. 10.1. For a higher statistical power the allele frequencies were used.

#### 3.2.3.10 Calculation of linkage disequilibrium and haplotype blocks

Linkage Disequilibrium (LD) is the non-random association between bi-allelic alleles at two loci (*Goldstein, 2001*). All loci in LD are inherited together. The 4 possible gametes of two bi-allelic markers (A, B) in linkage equilibrium have a frequency (x) of:

$A_1B_1$ $x_{11}$
$A_1B_2$ $x_{12}$
$A_2B_1$ $x_{21}$
$A_2B_2$ $x_{22}$

These frequencies can be calculated by their allele frequencies:

$A_1$ $p_1 = x_{11} + x_{12}$ $(x_{11}= p_1 q_1)$
$A_2$ $p_2 = x_{21} + x_{22}$ $(x_{21}= p_2 q_1)$
$B_1$ $q_1 = x_{11} + x_{21}$ $(x_{12}= p_1 q_2)$
$B_2$ $q_2 = x_{12} + x_{22}$ $(x_{22}= p_2 q_2)$

If the 2 bi-allelic marker are in linkage disequilibrium a deviation in the expected and observed gamete frequencies can be calculated using**:** $D= x_{11}x_{22} – x_{12}x_{21}$

*D* depends on allele frequencies so that changes in *D* reflect both changes in the intensity of the linkage correlation but as well changes in gene frequency. Therefore, LD is commonly measured by the ratio of D: $D' = D/D_{max}$
where $D_{max}$ is the maximal possible value of *D* for the given allele frequencies (*Lewontin, 1964*). *D*' ranging from 0 to 1, which 1 describes complete LD.

Haplotypes are the combinations of alleles in a population. Loci in these blocks are in strong LD. The haplotypes were disrupted by mutation, genetic drift and/or recombination. These events causing a fragmentation of haplotypes that leads to haplotype blocks (*Gabriel et al., 2002*).
Conclusions about evolutionary history and/or selection in the genomic region can be made out of the block sizes and numbers. In a neutral model of molecular evolution (1.8) mutations and recombination events reduce the size of a haplotype thus older and relatively common alleles were found in small haplotype blocks. Young or

alleles influenced by selection will be associated in long-range blocks with a high frequency in the population (*Bamshad and Woodding, 2003*). In general, LD between two loci decreases more rapidly with increasing distance. The block size is influenced by varying recombination rates across the human genome with an average block length of 1-173kb *(Gabriel et al., 2002).*

Genotype data in the PED file format were imported into the Haploview v.3.32 software (*Barret et al., 2005*). Haplotype blocks were defined with the implemented variant of the algorithm developed by *Wang et al. (2002)* with a gamete frequency cut-off of 0.01. In this method the population frequencies of the 4 possible two-marker haplotypes are computed for each marker pair. If all 4 possible haplotypes were observed, a recombination event occurred. A haplotype frequency cut-off of 0.01 was used. Sequent markers formed haplotype blocks where only 3 gametes were observed. The LD is presented in the Haploview's GOLD heatmap color scheme.

Additionally, allele frequencies and the Hardy-Weinberg Equilibrium (HWE) were calculated. The frequency of a given genotype will reach equilibrium in a randomly mating population and will stay constant over many generations in the absence of selection, migration or mutation.

HWE is defined by: $p^2 + 2pq + q^2 = 1$ where p and q are the frequencies of the 2 alleles respectively.

In several publications deviations from HWE were used as quality control for any DNA genotyping method and recommended as an essential part of any genome scan. Markers that are not in HWE should be excluded (*Gomes et al.1999; Weiss et al., 2001; Xu et al., 2002*). On the contrary *Zou and Donner (2006)* suggested the inclusion of these markers in further analyses.
For Haploview's implemented HWE calculation the observed was compared with the predicted heterozygosity in the cohort data. Latter was calculated with the minor allele frequency (MAF) using the equation: $2 \times MAF \times (1 - MAF)$.

#### 3.2.3.11 Reconstruction of haplotypes

Haplotypes of the genotyped unrelated individuals in the Mild Malaria Cohort were reconstructed with the PHASE v. 2.1 software (*Stephens et al., 2001, 2003*). This software implements a model-based Bayesian method for estimating haplotypes from unrelated individual genotype data. This approach calculates the "best" reconstruction haplotypes for each individual by an "approximate coalescent" model that reflects the fact that haplotypes in a population tend to group together in clusters of similar haplotypes. It assumes that the population has been evolving with a constant size for a long time and is randomly mating and that the locus under study is evolving neutrally (*Stephens and Scheet, 2005*). Besides the high accuracy in haplotype reconstruction its advantages lie in allowance for missing data strong and performance even when recombination occurred (*Xu et al., 2004; Marchini et al., 2006; Li et al., 2007).*

For detection of signatures of recent selective sweeps in the target genes in the Sweep software (3.2.2.10.1) the genotyped data from the Mild Malaria Cohort was transformed into the format specified by the PHASE programme's manual. SNPs that violated the HWE were excluded.

PHASE v. 2.1 was run with the standard settings. For each individual only the haplotype combination with the highest probability was used for further analyses from the PHASE output file.

#### 3.2.3.12 Screening for signatures of recent selection

For detecting signatures of recent selection in the genomic region of *APOA1, FABP1* and *SNAPAP* two different apporaches were used:

##### 3.2.3.12.1 Detection of extended haplotypes using Sweep software

The Sweep software v. 1.0 (*Sabeti et al., 2002)* analyses haplotype structures in the genome to detect evidences of natural selection. It examines alleles in high frequency in a haplotype with a long range LD across the genomic region. The decay of LD is a scale of the age the haplotype (1.8 and 3.2.2.8). It is measured by the extended haplotype homozygosity (EHH). The EHH is defined by the probability that two randomly chosen chromosomes carrying the core haplotype of interest are iden-

tical by descent from the core over a genomic distance (*Sabeti et al., 2002)*. For each allele, the haplotype homozygosity decays from 1 to 0 with increasing distance from the core.

To visualize the LD decay from a core SNP haplotype bifurcation diagrams were calculated displaying the LD breakdown and splitting of the haplotypes from a defined core SNP. This approach moves from the core SNP to the next marker and splits if two alleles are present.

For determination of the LD decay in *APOA1, FABP1* and *SNAPAP* only the calculated haplotypes of each individual with the highest probability were used (3.2.3.11). All haplotypes were arranged in the required format for the Sweep software like described in the Sweep manual.

#### 3.2.3.12.2 Detection of signatures of selection using Haplotter software

Haplotter, a web-based application developed by *Voight et al. (2006),* scans genomic regions for signs of selection in the phase I and II HapMap data (3.2.2.1). It calculates the integral of the decay of the EHH, a so called "integrated haplotype score" (iHS), away from the core allele until EHH reaches 0.05. iHS summed for the ancestral ($iHH_A$) or derived allele ($iHH_D$) over both directions away from the core SNP.

The final iHS is obtained by: $iHS = \ln\left(\frac{iHH_A}{iHH_D}\right)$

An extreme positive iHS score (iHS > 2) means that haplotypes on the ancestral allele background are longer compared to derived allele background whereas an extreme negative iHS score (iHS < -2) means that the haplotypes on the derived allele background are longer compared to the haplotypes associated with the ancestral allele (*Voight et al., 2006*).

This approach has the highest power to detect recent selection for alleles with medium frequencies that have not reached fixation.

The implemented gene-centered approach was used for *APOA1, FABP1* and *SNAPAP* to calculate and visualize the iHS. Additionally, Haplotter's implemented *Fay and Wu's H* (*Fay and Wu, 2000*) and *Tajima's D (Tajima, 1989)* were calculated to identify deviations in polymorphism frequencies due to selection. Both methods are based on the frequencies of the segregating polymorphisms in the genomic region

of interest. A positive or negative selection will shift the frequencies in different directions compared to the neutral model.

*Fay and Wu's H* is sensitive to high allele frequencies that exist due to a selective sweep. It compares the number of derived nucleotide variants at low and high frequencies with the number of variants at intermediate frequencies (*Bamshad and Wooding, 2003*). This test requires knowledge of the ancestral allele (*Sabeti et al., 2006*).

An excess of low frequency alleles can be observed if these alleles are hitchhiking with the region under positive selection. *Tajima's D* compares the number of observed polymorphisms with the mean pairwise difference between sequences. *Fay and Wu's H* and *Tajima's D* have the greatest power to detect a selective sweep when the selected allele is fixed or close to fixation. In summary *Fay and Wu's H* calculates departures from neutrality that are measured in the difference between high-frequency and intermediate-frequency alleles whereas *Tajima's D* reflect the difference between low-frequency and intermediate frequency alleles thus predict departures from neutrality. Hence *Tajima's D* is sensitive to population effects both tests combined can help to differ selection from population effects. *Fay and Wu's H* is the most commonly used test for derived alleles (*Sabeti et al., 2006*). It is based on derived alleles thus knowledge of the ancestral alleles is required. In practice closely related species like chimpanzees are usually used as human ancestors.

Because populations in different environments facing different selection pressures the allele frequencies in these populations should differ in the selected genomic regions. The *Wright Fixation Index* $F_{ST}$ can be used to measure the level differentiation between populations at a locus (*Lewontin and Krakauer, 1973*). It correlates the randomly chosen alleles within the same sub-population relative to its frequency in the entire population (*Holsinger and Weir, 2009*). It is an indicator to the degree of resemblance among individuals within populations and directly related to the variance in allele frequencies among populations (*Holsinger and Weir, 2009*). A small $F_{ST}$ value indicates a similar allele frequencies within each population whereas large, it means different allele frequencies. In Haplotter the implemented $F_{ST}$ statistic was used to compare the heterozygosity around the genomic region of *APOA1*, *FABP1* and *SNAPAP* between the YRI and CEU population from the HapMap dataset. All these approaches can provide information about the evolutionary history of a genetic marker (Figure 5).

# 4. Results

## 4.1 Identification of SNPs in candidate genes *APOA1, FABP1* and *SNAPAP*

*P. falciparum* malaria has exerted strong selection pressure on the human genome for the last 10,000 years. In the context of high variability of individual disease outcome the genetics of host response are likely to play an important role (*Mackinnon et al, 2005*). The protein UIS3 and UIS4 play an essential role in *P. berghei* liver cell stages. *Mueller et al*. could identify three host genes coding for liver ligands: *APOA1, FABP1* and *SNAPAP*. I reasoned that if genetic polymorphisms in these three human candidates could alter host/parasite interactions children in endemic areas carrying protective alleles should be less susceptible to *falciparum* infections and suffer fewer malaria incidences. To address this question all three genes were screened for polymorphisms.

### 4.1.1 Re-sequencing of *APOA1* confirmed 11 SNPs

I initiated my study by re-sequencing the genomic region around *APOA1* (ENSG00000118137) in 50 DNA samples from the area of Kilifi at the Kenyan Coast. They were randomly chosen and no anthropometric information was available. I sequenced every sample in 6 overlapping fragments spanning a genomic region of 2.5kb around *APOA1* and re-aligned them using the BioEdit v. 7.0.9.0 software. The *APOA*1 gene consists of 1.86kb and is located on Chromosome 11 at position 116,706,467 to 116,708,666. The open reading frame (ORF) is oriented to the opposite to the chromosomal direction that is annotated in the Ensembl database. The gene consists of 4 exons and 3 introns. The Apolipoprotein A1 Precursor protein consists of around 243 amino acids (AS) with a molecular weight of 30kDa. Five different transcripts (mRNA) are annotated.

In re-sequenced data I found 11 SNPs at the *APOA1* locus. Two were identified in the 5' UTR, another two in the 3'UTR whereas the remaining SNPs were scattered within the introns of the gene. In all 100 re-sequenced chromosomes no insertions or deletions were localized in this genomic region.

All localized SNPs were indentified in the dbSNP database and named by the official identifier (Figure 7). The allele frequencies between the re-sequenced Kenyan sam-

ples and data from the PharmGKB and CWRU Genomics databases were compared. Latter consisted of sample data from Harare in Zimbabwe with 80 chromosomes and samples from 68 individuals from Michigan in the United States. The PharmGKB dataset contained a mix of samples from African-Americans and US-Whites. No information about SNPs in the genomic region of *APOA1* was available in HapMap Phase I or II data thus I could not compare frequencies between the populations from this database.
For all 11 SNPs no information about function or effect was obtainable except for rs670 and rs5069. These two SNPs were associated with different phenotypes (*Kamboh et al, 1999*). SNP rs12718467 was excluded because of its low minor allele frequency (MAF) of 1%.

### 4.1.2 Selection of 11 SNPs and 2 tagSNPs in *FABP1* by database screening

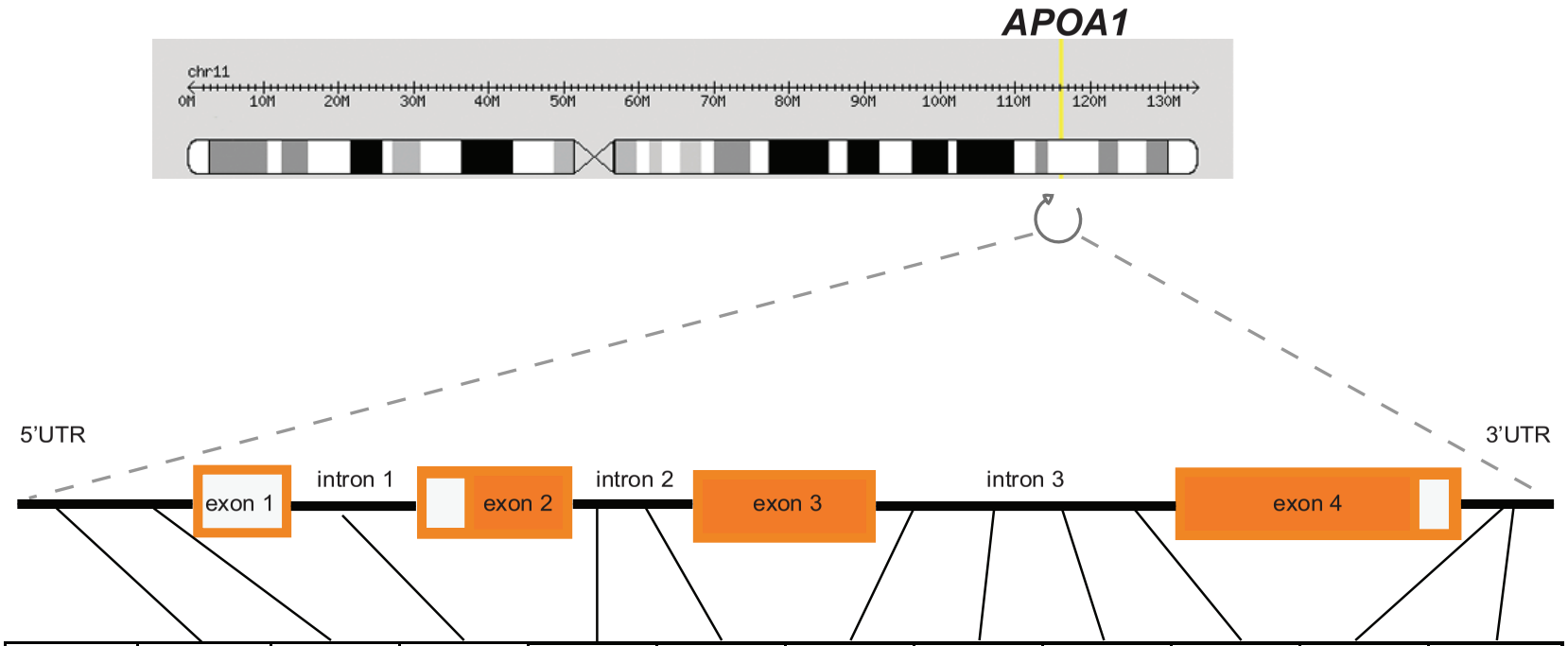
For identification of SNPs in *FABP1* and *SNAPAP* I screened the dbSNP, CWRU Genomics, PharmGKB, UCSC and the HapMap databases.
The gene of *FABP1* (ENSG00000163586) reaches from position 88,422,510 to 88,427,635 on Chromosome 2. It consists of 4 exons and 3 large introns. Two transcripts are coding for a small protein of 127 AS with a molecular weight of approximately 14kDa. SNPs were identified in the dbSNP and HapMap databases. Like *APOA1* its ORF points to the opposite direction of the chromosome. I chose 10 SNPs in the databases for the SNaPshot Multiplex assay (3.2.3.1)**.** Their allele frequencies from the HapMap datasbase in the Yoruba (YRI) population of Ibadan in Nigeria containing 30 sets of samples from two parents and an adult child were compared with the dataset of U.S. residents with northern and western European ancestry (CEU) by the Centre d'Etude du Polymorphisme Humain (CEPH) (Figure 8). Latter consisted of thirty U.S. trios provided samples, which were collected in 1980s.

Here, allele frequencies from the YRI were compared with the CEU to identify SNPs that differ in frequency due to possible selection. All SNPs and their frequencies are listed in figure 8. No insertion or deletions were found in *FABP1* using HapMap dataset according to *McCarroll et al. (2006).*

Tag SNPs obtain information about other SNPs without testing them directly because they are in high LD thus they minimize the number of SNPs in the assays. Usage of the HapMap dataset for tag SNPs in different populations was validated (*De Bakker 2006)*. Therefore, I calculated two tag SNPs using the in HapMap implemented Tagger (3.2.2.3). rs1545223 and rs1530273 were tagged by rs13399861 and rs6711957 respectively. In both cases the tagging with a MAF cut off of 0.5 and $r^2$ cut off of 0.8 was verified in the YRI and CEU HapMap dataset. All SNPs in *FABP1* were distributed within the introns with unknown phenotype except for rs2241883. For the latter a shift from G to A results in a substitution from alanine to threonine that is associated with fasting triglycerides and LDL-cholesterol levels in females (*Fisher et al., 2007*).

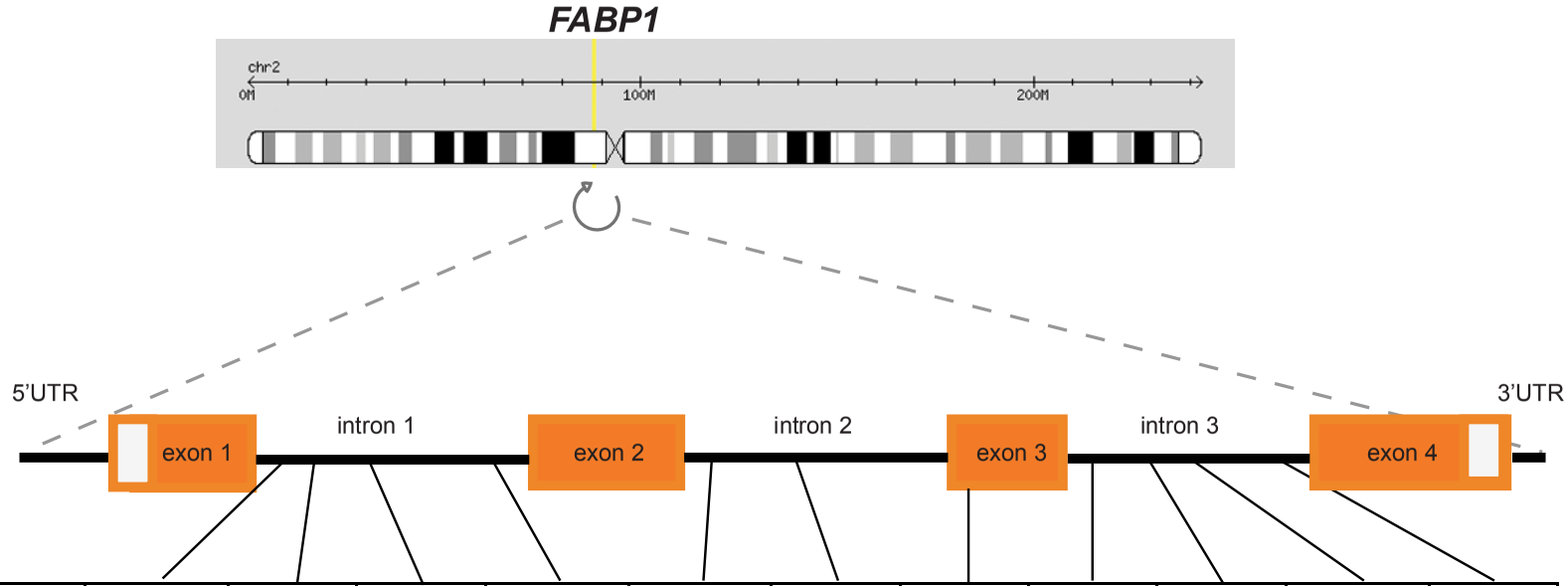
### 4.1.3 Selection of 12 SNPs in *SNAPAP* by database screening

*SNAPAP* is located on chromosome 1 between position 153,631,145 and 153,634,325. It is organized in 4 exons and 3 introns. The ORF is oriented in the same direction like the chromosome. Four transcripts are annotated in the Ensembl database. The protein consists of 136 AS with a relative molecular weight of 15kDa.
Like in *FABP1* I identified SNPs using database screening and compared their allele frequencies in the YRI and CEU datasets from HapMap. Again the SNPs were selected according to their potential influence on gene function, known differences in allele frequencies between populations and the reliability of the primer in the SNaPshot Multiplex assay (3.2.3.4.2). 12 SNPs were selected for further genotyping (Figure 9). One SNP was located in the 5'UTR, three in the 3'UTR, four in the exons and another 4 in the introns. The allele change (A→T) of rs1802461 leads to an AS exchange from serine to cysteine without known phenotype. Nucleotide substitutions in SNP rs16835479 and rs8532 result in a synonymous AS change. Like in FABP1 I found no insertions or deletion in SNAPAP using the HapMap data (*McCarroll et al. 2006)*. Furthermore, no tag SNP could be verified in the HapMap YRI or CEU dataset.



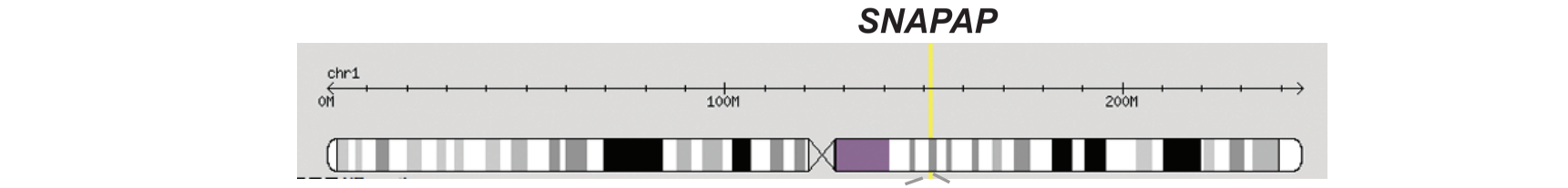

| dbSNP ID | rs12718467 | rs670 | rs5069 | rs5070 | rs2070665 | rs5072 | rs5073 | rs7116797 | rs5076 | rs5081 | rs12718463 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| alleles | G/T | G/A | C/T | T/C | C/T | C/T | G/A | C/T | C/T | T/A | A/G |
| position | 116213912 | 116213623 | 116213464 | 116213230 | 116212894 | 116212793 | 116212723 | 116212548 | 116212400 | 116211556 | 116211553 |
| Kenya | 99/ 1 | 85/15 | 69 / 30 | 69/ 30 | 86/ 14 | 86/ 14 | 94/ 6 | 37/ 63 | 78/ 22 | 93/ 7 | 84/ 16 |
| PharmGKB | 99/ 1 | 81/ 19 | 88/ 12 | 44 /56 | 93 / 7 | 92/ 8 | 96/ 4 | 80/ 20 | 88/ 12 | 92/ 8 | 84/ 16 |
| CWRU | - | - | 75/ 25 Zimb. | 70 /30 Zimb. | - | 95/ 5 US | 95/ 5 Zimb. | - | 95/ 5 Zimb. | 65/ 35 Zimb. | - |

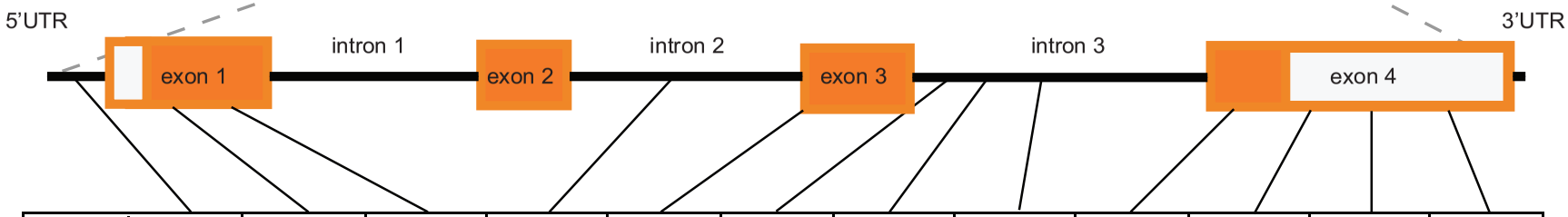
**Figure 7: Gene structure of *APOA1* and identified SNPs**
Position of *APOA1* is marked by the yellow line on chromosome 11 (upper figure). The circle indicates the inverse orientation of the ORF to the chromosome. Exons are presented in orange boxes (filled: translated, white: untranslated regions (UTR)). Introns are graphed as black lines.
In the first line SNPs are listed by the official dbSNP identifier.The alleles, their frequencies (percent) and positions on the chromosome are listed below. Last three lines show the allele frequencies i) in the re-sequenced samples from Kilifi, ii) in the PharmGKB and iii) in the CWRU database. Latter includes the samples origin: Harare, Zimbabwe (Zimb.) and from Michigan with european ancestry (US). Data from the PharmGKB contain children with North- and African-American background that were sampled in Chicago.




| dbSNP ID | rs13400963 | rs7577401 | rs2919871 | rs28452712 | rs13399861 | rs6711957 | rs2241883 | rs1123605 | rs1545224 | rs11678114 | rs2197076 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **alleles** | G/A | G/A | A/T | A/G | G/A | C/T | G/A (miss.) | C/T | G/A | G/A | C/T |
| **position** | 88208060 | 88207888 | 88207561 | 88207140 | 88206371 | 88205735 | 88205181 | 88205023 | 88204931 | 88204131 | 88203873 |
| **YRI** | 22/ 78 | 7/ 93 | 35/ 65 | - | 72/ 28 | - | 15/ 85 | - | 4/ 96 | - | 96/ 4 |
| **CEU** | 0/ 100 | 0/ 100 | 18/ 82 | - | 100/ 0 | - | 29/ 71 | - | 26/ 74 | - | 79/ 21 |
| | | | | | ***tagging:*** | ***tagging:*** | | | | | |
| | | | | | *rs1545223* | *rs1530273* | | | | | |
| | | | | | *88204625* | *88205822* | | | | | |

**Figure 8: Gene structure of *FABP1* and selected SNPs**
Position of *FABP1* is marked by the yellow line on chromosome 2 (upper figure). The circle indicates the inverse orientation of the ORF to the chromosome. Exons are presented in orange boxes (filled: translated, white: untranslatedregions (UTR)). Introns are graphed as black lines.
In the first line SNPs are listed by the official dbSNP identifier. The alleles, their function (miss.: missense mutation), their frequencies (percent) and positions on the chromosome are listed below. Last two lines show the allele frequencies in the HapMap data Phase I from the Yoruban (YRI) and U.S. residents with northern and western European ancestry (CEU). Below the tag SNPs rs13399861 and rs6711957 the identifier of the tagged SNPs and their chromosomal positions are listed.




| dbSNP ID | rs3806234 | rs35903030 | rs16835479 | rs16835481 | rs8532 | rs6692298 | rs12742195 | rs12755522 | rs1802461 | rs16835489 | rs41305092 | rs7345 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| alleles | C/T | G/C (frame.) | A/C (syn) | G/A | C/T (syn) | G/C | A/G | G/T | A/T (miss.) | A/C | C/T | A/C |
| position | 151897695 | 151897927 | 151897948 | 151898508 | 151898567 | 151898763 | 151899502 | 151899568 | 151900324 | 151900427 | 151900434 | 151900682 |
| YRI | 75/ 25 | - | - | 13/ 87 | - | 70/ 30 | - | - | - | 96/ 4 | - | 78/ 22 |
| CEU | 67/ 33 | - | - | 0/ 100 | - | 92/ 8 | - | - | - | 100/ 0 | - | 47/ 53 |

**Figure 9: Gene structure of *SNAPAP* and selected SNPs**
Position of *SNAPAP* is marked by the yellow line on chromosome 1 (upper figure). Exons are presented in orange boxes (filled: translated, white: untranslated regions (UTR)). Introns are graphed as black lines. In the first line SNPs are listed by the official dbSNP identifier. The alleles, their function (frame.: frameshift, syn.: synonymous and miss.: missense mutation), their frequencies (percent) and positions on the chromosome are listed below. Last two lines show the allele frequencies in the HapMap data Phase I from the Yoruban (YRI) and U.S. residents with northern and western European ancestry (CEU).

## 4.2 Genotyping of the Mild Malaria Cohort

### 4.2.1 Genotyping Results of the Mild Malaria Cohort Study

Due to the low amount of DNA in samples from the Mild Malaria Cohort I amplified 728 sample genomes using the GenomiPhi Kit. This subset of the cohort contained duplicates or triplicates that were excluded from further analysis. All remaining samples were used for genotyping of the targeted SNPs in *APOA1, FABP1* and *SNAPAP* with the SNaPshot Multiplex assay (3.2.1.8).

The efficiency of the Long Range PCR (3.2.1.4.4.2) was proportional to the template product sizes of *APOA1* (2.4kb), *FABP1* (3.4kb) and *SNAPAP* (5kb). For this reason the obtained genotyped sample number for the *APOA1* was highest followed by *SNAPAP* and finally *FABP1*. This was probably caused by progressing disintegration of the sample DNA due to the long-term storage and repeated freezing and thawing (*Williams TN, personal communication*). Using the Genemapper v. 4.0 software all alleles were called automatically in the SNaPshot Multiplex assay. An example of SNaPshot Multiplex electropherogram for each gene is displayed with its specific panel in Figure 10.

Genotyping results of the SNaPshot Multiplex assays are presented in table 7,8 and 9 for each tested gene, respectively. The SNPs are labelled with the dbSNP identifier. In the 2$^{nd}$ column the alleles are shown with first the major and then the minor allele. The 3$^{rd}$ column gives the percentage of non-missing genotypes for this marker and the total number (n) of genotyped subjects. The minor allele frequency (MAF) is given in the 4$^{th}$ column. The major allele frequency can be calculated by 1-MAF. For calculation of the Hardy-Weinberg equilibrium (HWE) the observed was compared with the predicted heterozygosity. Significant deviations were given by the p-value of HWE in the last column. I used HWE as quality control (*Gomes et al.1999; Weiss et al., 2001; Xu et al., 2002*). Contrary to these, markers that violated HWE were included tentatively in further analyses like proposed by *Zou and Donner (2006)*.

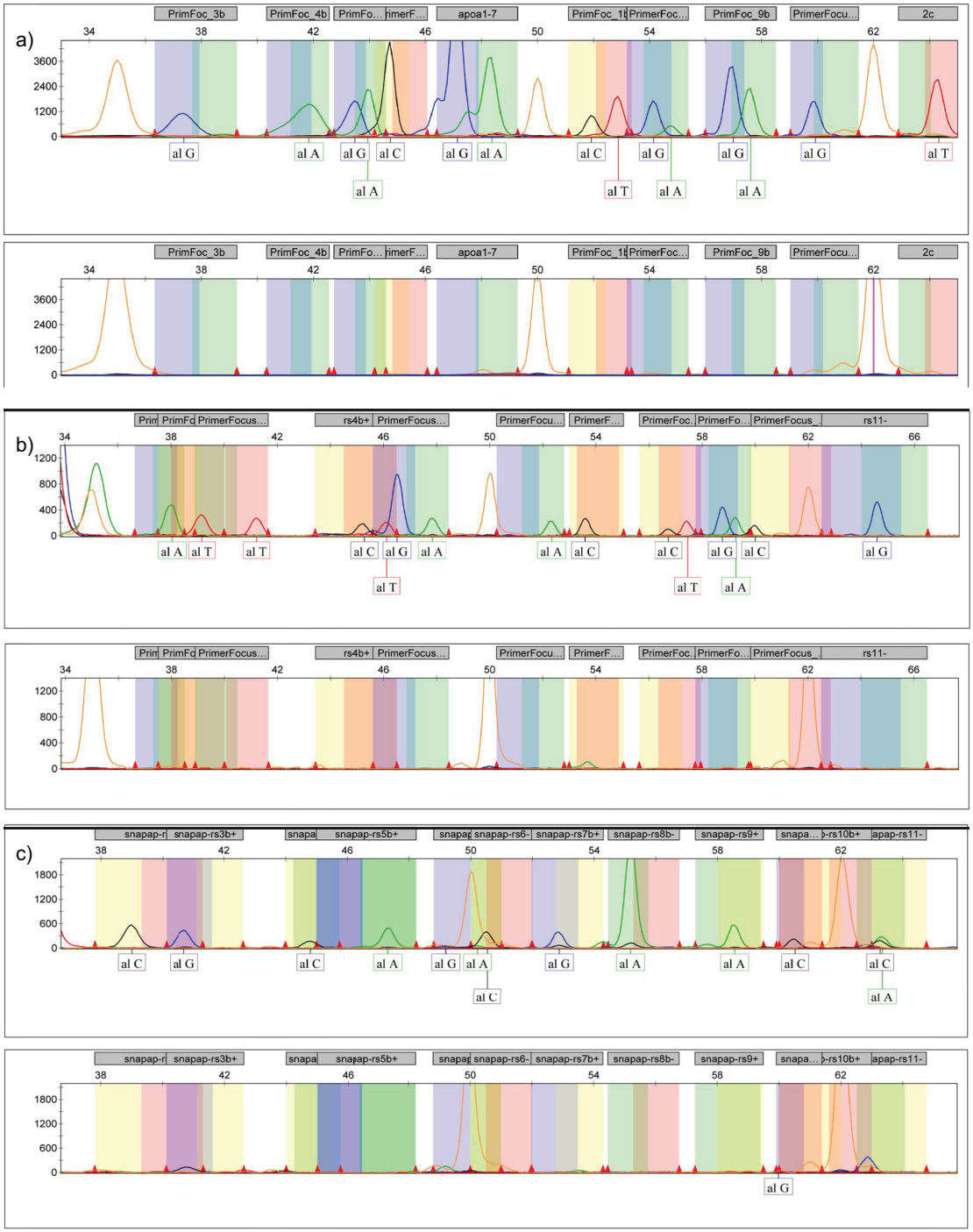



**Figure 10: Representative examples of SNaPshot Multiplex assay electropherograms for *APOA1* (a), *FABP1* (b) and *SNAPAP* (c)**
In all three figures a representative example of the specific SNaPshot Multiplex assay is given with DNA (top) and without DNA (below). X-axes represent the run length in bp and the y-axes the fluorescence intensity. Grey boxes show the primer with coloured bins for the expected alleles in the panel (blue: G, green: A, yellow: C and red: T). Primers appear in order of table 4. Called alleles are listed below each electropherogram. An allele is only called if the peak emerges in the correct bin. Peaks appearing in the negative controls are smaller in presence of DNA due to exhaustion of the causal primers in the reaction. Heterozygous alleles are called in relation to their peak heights. Peak heights are influenced by amount of DNA and primer in the reaction and SNaPshot product loaded into the Genetic Analyzer.

#### 4.2.1.1 Alleles of 10 SNPs in *APOA1*

For *APOA1* 568 samples from the Mild Malaria Cohort were genotyped (Table 7). All 10 targeted SNPs were genotyped with an efficiency above 99%. No marker violated the HWE p-value threshold of 0.001.

**Table 7: Results of *APOA1* genotyping in Mild Malaria Cohort with SNaPshot Multiplex assay**

| dbSNP ID | alleles | non-missing alleles (%) / n | MAF | obs. HET | pred. HET | HWE p-value |
|---|---|---|---|---|---|---|
| **rs12718463** | T:C | 100 / 568) | 0.34 | 0.42 | 0.44 | 0.31 |
| **rs5081** | T:A | 100 / 568 | 0.27 | 0.37 | 0.39 | 0.39 |
| **rs5076** | G:A | 99.8 / 567 | 0.15 | 0.18 | 0.25 | 0.002 |
| **rs7116797** | G:A | 100 / 568 | 0.45 | 0.45 | 0.49 | 0.04 |
| **rs5073** | C:T | 100 / 568 | 0.03 | 0.05 | 0.05 | 1.0 |
| **rs5072** | G:A | 100 / 568 | 0.16 | 0.27 | 0.26 | 1.0 |
| **rs2070665** | G:A | 100 / 568 | 0.17 | 0.30 | 0.29 | 0.25 |
| **rs5070** | A:G | 99.8 / 567 | 0.42 | 0.44 | 0.48 | 0.06 |
| **rs5069** | G:A | 100 / 568 | 0.20 | 0.30 | 0.32 | 0.20 |
| **rs670** | G:A | 100 / 568 | 0.24 | 0.36 | 0.36 | 0.76 |

#### 4.2.1.2 Alleles of 11 SNPs in *FABP1*

From the Mild Malaria Cohort 497 samples were genotyped for the 11 chosen SNPs in *FABP1* (Table 8). Only 3 markers were below 100% of non-missing genotypes. The observed heterozygosity differed significantly from the predicted heterozygosity for marker rs2241883 and rs13399861 thus violated the HWE.

**Table 8: Results of *FABP1* genotyping in Mild Malaria Cohort with SNaPshot Multiplex assay**

| dbSNP ID | alleles | non-missing alleles (%) / n | MAF | obs. HET | pred. HET | HWE p-value |
|---|---|---|---|---|---|---|
| **rs2197076** | C:T | 99.8 / 496 | 0.07 | 0.12 | 0.13 | 0.07 |
| **rs11678114** | C:T | 100 / 497 | 0.03 | 0.06 | 0.06 | 1.0 |
| **rs1545224** | A:G | 100 / 497 | 0.12 | 0.19 | 0.22 | 0.01 |
| **rs1123605** | G:A | 99.6 / 495 | 0.003 | 0.006 | 0.006 | 1.0 |
| **rs2241883** | A:G | 100 / 497 | 0.13 | 0.17 | 0.22 | 0.001 |
| **rs6711957** | C:T | 100 / 497 | 0.39 | 0.43 | 0.47 | 0.0484 |
| **rs13399861** | C:T | 100 / 497 | 0.39 | 0.56 | 0.47 | 0.00001 |
| **rs28452712** | G:A | 100 / 497 | 0.34 | 0.44 | 0.45 | 0.76 |
| **rs2919871** | A:T | 100 / 497 | 0.36 | 0.45 | 0.46 | 0.75 |
| **rs7577401** | T:C | 99.8 / 496 | 0.03 | 0.07 | 0.07 | 0.95 |
| **rs13400963** | A:G | 100 / 497 | 0.15 | 0.24 | 0.27 | 0.05 |

#### 4.2.1.3 Alleles of 12 SNPs in *SNAPAP*

All 12 SNPs in *SNAPAP* were genotyped in 548 samples from the Mild Malaria Cohort (Table 9). Missing alleles were found in marker rs35903030, rs12742195 and rs7345. Marker rs16835479, rs16835479 and rs12742195 fell below the HWE p-value threshold. SNP rs12742195 and rs12755522 were excluded due to low sized peaks and ambiguous allele calling in the electropherograms of the SNaPshot Multiplex assay.

**Table 9: Results of *SNAPAP* genotyping in Mild Malaria Cohort with SNaPshot Multiplex assay**

| dbSNP ID | alleles | non-missing alleles (%) / n | MAF | obs. HET | pred. HET | HWE p-value |
|---|---|---|---|---|---|---|
| **rs3806234** | C:T | 100 / 548 | 0.16 | 0.28 | 0.27 | 0.36 |
| **rs35903030** | G:C | 99.8 / 547 | 0.001 | 0.002 | 0.002 | 1.0 |
| **rs16835479** | C:A | 100 / 548 | 0.11 | 0.16 | 0.20 | 0.0001 |
| **rs16835481** | A:G | 100 / 548 | 0.01 | 0.16 | 0.18 | 0.13 |
| **rs8532** | C:T | 100 / 548 | 0.13 | 0.21 | 0.22 | 0.45 |
| **rs6692298** | G:C | 100 / 548 | 0.35 | 0.21 | 0.46 | $1*10^{-32}$ |
| **rs12742195** | A:G | 100 / 548 | 0.48 | 0.91 | 0.50 | $1*10^{-91}$ |
| **rs12755522** | G:T | 100 / 548 | 0.002 | 0.004 | 0.004 | 1.0 |
| **rs1802461** | T:A | 100 / 548 | 0.001 | 0.002 | 0.002 | 1.0 |
| **rs16835489** | A:C | 100 / 548 | 0.01 | 0.03 | 0.03 | 1.0 |
| **rs41305092** | C:T | 100 / 548 | 0.03 | 0.07 | 0.06 | 1.0 |
| **rs7345** | A:C | 99.1 / 543 | 0.22 | 0.37 | 0.35 | 0.06 |

### 4.3 Genotype-phenotype association analyses using single SNP approach

Association studies based on SNPs are a widely used approach for identifying genes associated with diseases, pathomechanisms or phenotypes (e.g. body height, melanization) (*Brooks, 1999*). Here, I applied this approach to associate the phenotype, number of *P. falciparum* malaria events (presence of asexual blood stages with fever) in relation to the individual follow-up time, with the genotype in the Mild Malaria Cohort subset. Therefore, the validated genotype datasets from each gene were merged with the sample data from the Mild Malaria Cohort Study, respectively (3.2.3.6).

*Williams et al. (2005, 1)* observed a decreasing protective effect of the sickle cell trait (HbAS) in children older than 10 years supposedly caused by an accelerated

immune acquisition of immunity. Accordingly, observations from subjects above 10 years were excluded from analyses. For each gene the numbers of subjects and observations in the total and the reduced datasets containing only children younger than 10 years are listed in table 10. The distributions of observations by age are comparable between the three genes. Figure 11 exemplifies the distribution of observations by age groups for *APOA1*.

**Table 10: Number of subjects and observations from the genotyped Mild Malaria Cohort.**

The number of tested SNPs, subjects (n), malaria events and observations are listed for each gene in the total dataset and for children below 10 years. Additionally, the observations were translated into children years of follow-up (cyfu).

| Gene | no. SNP | total | | | | <10 years | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | n | no. obs. | no. events | cyfu | n | no. events | no. obs. | cyfu |
| ***APOA1*** | 10 | 568 | 86594 | 3047 | 1665 | 465 | 2767 | 70308 | 1352 |
| ***FABP1*** | 11+2 | 497 | 74299 | 2633 | 1429 | 401 | 2345 | 59789 | 1150 |
| ***SNAPAP*** | 10 | 548 | 82870 | 2862 | 1594 | 445 | 2574 | 67107 | 1291 |

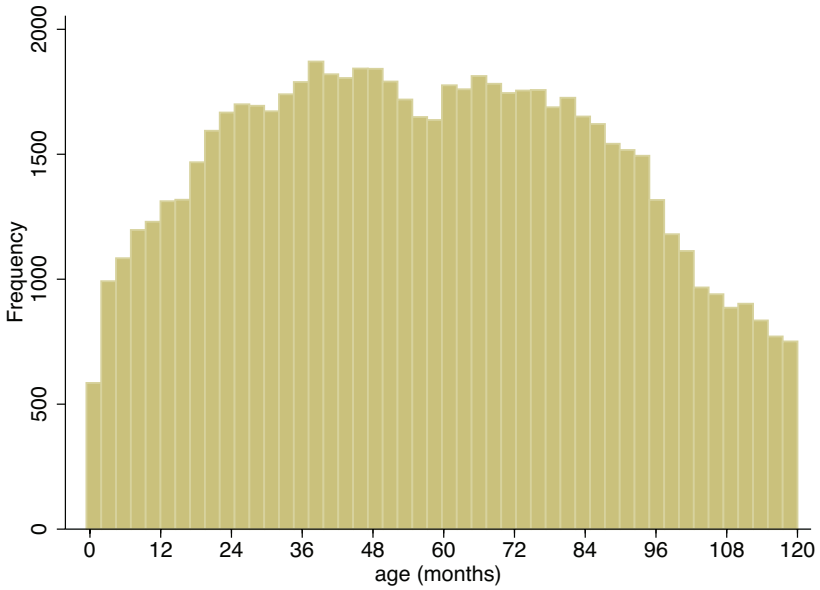



**Figure 11: Frequency histogram of observations in the *APOA1* dataset by months in children below 10 years.**

All observations from the genotyped Mild Malaria Cohort for *APOA1* are presented by age in months. The height of the bars is equal to the number of observations in the age group thus the sum of the heights is equal to the total number of observations/ weekly visits.

### 4.3.1 Reduced risk in rs670 G carriers in Poisson regression analyses

I used Poisson regression to determine the influence of the SNP genotypes and possible covariates on the incidence of *P. falciparum* malaria (3.2.3.6.1). Incidence rate ratios (IRR) were compared between the three possible genotypes of each SNP. Additionally, association was analysed in different age groups in each dataset to identify age-related associations.

All SNP genotypes were tested for association with the occurrence of asexual parasite blood stages in the cohort participants. Among all 33 finally used SNPs only rs670 (G/A) located in 5' UTR of *APOA1* was consistently associated with incidence and time to infection with *P. falciparum.* In analogy to HbAS (*Williams et al., 2005, 1)*, which I used as a positive control, the association depended on age of the subjects in the cohort (Figure 12). Contrary to HbAS, association of rs670 was already lost in individuals older than 8 years. In both traits no association was observed in children younger than 12 months.

In 306 non-immune children between 1 and 5 years, the most susceptible age group, rs670 was associated with *P. falciparum* incidence (p=0.003). In this age group I found an almost 50% lower risk of *P. falciparum* incidence in GG compared to AA carriers, reflected by the unadjusted median IRR of 0.52 (95% CI 0.34 - 0.80). Accounting for possible confounding variates I have found a similar IRR of 0.51 (95% CI 0.32 - 0.84; p=0.007) in a separate Poisson regression model, which was adjusted for sex, age, season, location, ethnic group, HbAS and $\alpha$+–thalassaemia. The effect of rs670 GG on *P. falciparum* incidence is comparable to HbAS. An un- and adjusted Poisson regression for HbAS showed an IRR of 0.59 (95% CI, 0.45-0.78; p<0.001) and 0.60 (95% CI, 0.43-0.82; p=0.002), respectively.

I accounted for a possibly inflated Type I error caused by multiple testing in a Poisson regression model in which all SNPs were included. The SNP rs670 remained significantly associated (unadjusted p=0.006; adjusted p=0.01) in 1-5 year old children. The Poisson regression model for the association of rs670 was validated by a goodness-of-fit-test with a non-significant p-value of 1.

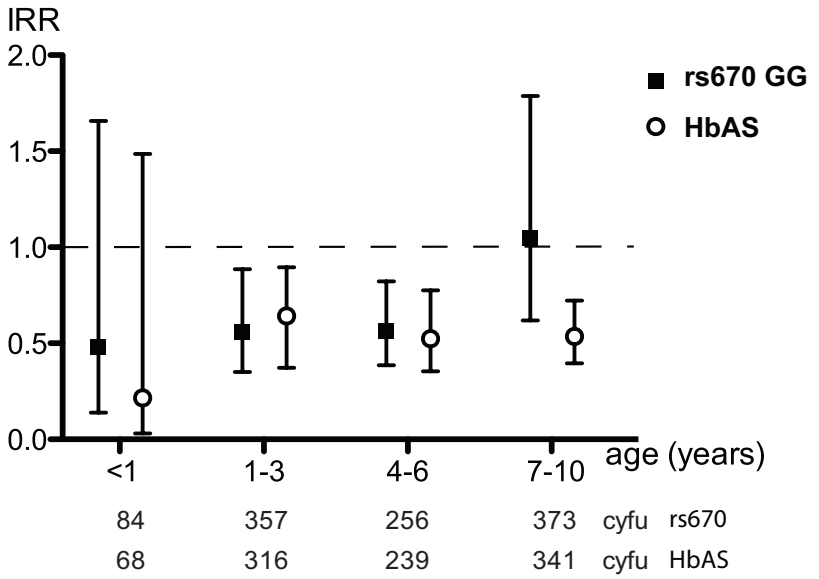



**Figure 12: Incidence rate ratios for malaria of rs670 AA vs. GG and HbAA vs. HbAS by age**

The unadjusted IRR are presented for each age group. Below the children years of follow-up (cyfu) for each age group are listed.

### 4.3.2 Multiple failure-time analysis revealed a prolonged time between infections by rs670

An influence of rs670 on incidence of *P. falciparum* is probably caused by differences in time between infections. I used a multiple failure-time analysis based on the Anderson-Gill model for ordered multiple-failures to determine the time differences between *P. falciparum* infections and re-infections in relation to the SNP genotypes (3.2.3.6.2).

Similar to the Poisson regression analyses an age-related effect of the rs670 genotypes and time to *P. falciparum* infections was found. Again rs670 was significantly associated in the most susceptible age group from 1-5 years. In a Cox regression model a significant association from AA to GG was observed ($p=0.005$) in this age group. Analogous, this association was lost in children older than 8 years ($p>0.05$), contrary to HbAS, which remained significant ($p<0.001$). For both genetic traits I

found no association in children below 12 months. The association strength of rs670 GG and HbAS with time to infection was comparable with hazard ratios of 0.49 (95% CI, 0.29 – 0.81) and 0.55 (95% CI, 0.42 – 0.73), respectively.

The time to detectable blood stage re-infections was extended by 63 days in GG carriers (median failure time of 126 days, 95% CI 98 - 154) compared to AA homozygotes (median of 63 days; 95% CI 7 – 98). GA carriers had similar median failure times of 119 (95% CI 91 - 140) to GG homozygotes. This compares to a prolonged median time of 266 days in children with HbAS (95% CI 105 – 420) vs. children with normal HbAA (median 112 days, 95% CI 91 – 126). These results indicate a delay of detectable blood stages induced by both genetic traits. Times to-re-infection by genetic trait were graphed by Kaplan-Meier failure estimations for multiple failures in figure 13. One child carrying HbAS and the rs670 AA genotype showed a median survival time of 561 days suggesting dominance of HbAS thus it was excluded from the analyses.

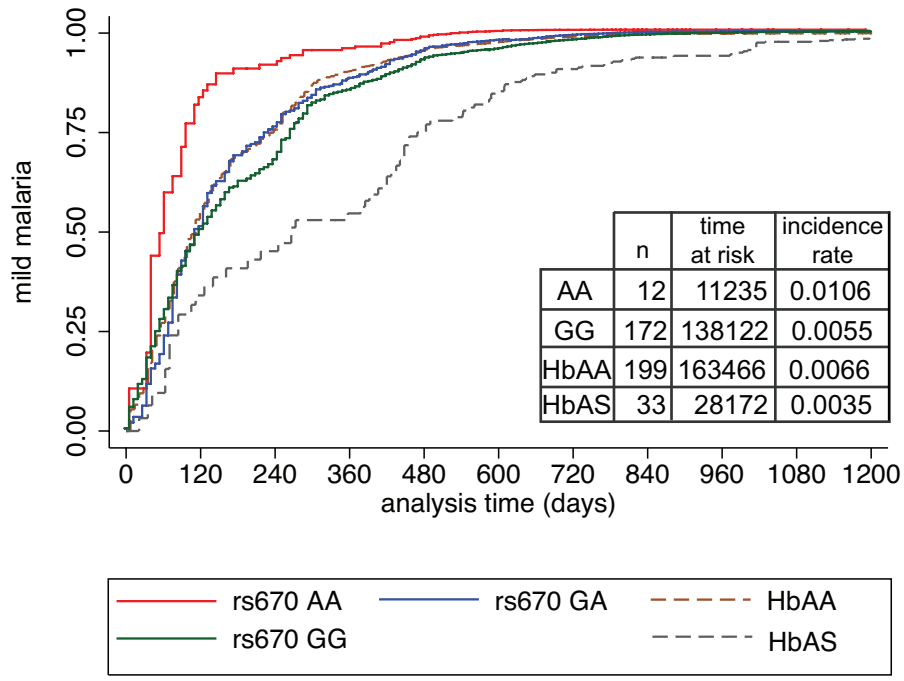



**Figure 13: Kaplan-Meier failure estimates in children 1-5 years by genotypic group**

Failure-time estimations are graphed for multiple failures in children by genotypes of rs670 and HbAS. Number of subjects (n), their total time at risk (days) and the incidence rate for the main genotypes are given in the box below graphs.

Using the cox regression analysis I could determine the influence of different factors on the variability in *P. falciparum* malaria incidence in children from 1 to 5 years. In total, all tested covariates (season, age, ethnic group, sex, location, α+–thalassaemia, HbAS and rs670) explained 51% (95% CI, 46-60) of the variation. Broken down into the single variates (table 11) I found only three major factors influencing incidence of *P. falciparum*. By far most of the variability is explained by the seasons underscoring exposure as a key factor of *P. falciparum* incidence. On the contrary host genetic factors seem to play a subordinate role on variability. rs670 or HbAS explained only 1.6% and 2%, respectively.

**Table 11: Covariates influencing variability in *P. falciparum* incidence in children 1-5 years**

All variates are given with their explaining fraction and 95% CI on *P. falciparum* incidence.

| factor | explained variability | 95% CI |
|---|---|---|
| season | 0.46 | 0.40 - 0.56 |
| age | 0.001 | <0.001 - 0.007 |
| ethnic group | <0.001 | <0.001 - 0.004 |
| sex | <0.001 | <0.001 - 0.007 |
| location | 0.002 | <0.001 - 0.009 |
| α+–thalassaemia | <0.001 | <0.001 - 0.007 |
| HbAS | 0.02 | 0.006 - 0.038 |
| rs670 | 0.016 | 0.003 - 0.036 |
| total | 0.51 | 0.40 - 0.60 |

### 4.3.3 No association of rs670 with parasite density or epistasis with HbAS

HbAS is affecting the development of intraerythrocytic *P. falciparum* stages. *Williams et al.* (2005, 2) described an association of HbAS on asexual parasites densities in the patient's blood. Accordingly, determination of the influence of rs670 on *P. falciparum* blood stages could provide evidence to localize its effect. Therefore, I performed linear regression analyses with the rs670 genotypes and the *P. falciparum* parasite densities that were measured in the subjects from the Mild Malaria Cohort (3.2.3.6.1). Based on full blood count data the $\log_{10}$ transformed numbers of

*P. falciparum* asexual stages were used. In figure 12 the log numbers of parasites are plotted by the genotypes of rs670 and HbAS: In comparison the rs670 genotypes show similar median parasite densities that do not differ significantly whereas the median number of parasites is considerably reduced in children with HbAS. As expected,HbAS was significantly associated with parasite densities in children between 1-5 years (p=0.001) whereas I found no association with rs670 (AA vs. GA: p=0.16 and AA vs. GG: p=0.28).

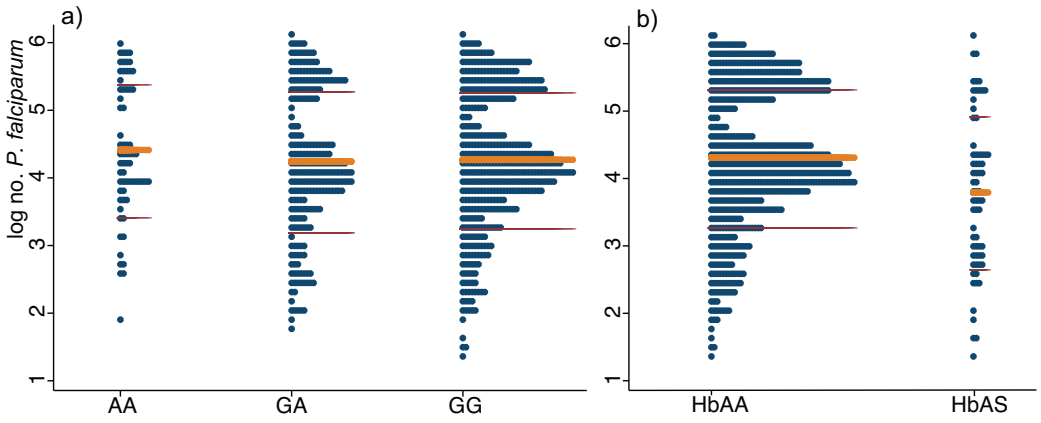



**Figure 14: Dot plot of asexual *P. falciparum* parasites in children from 1-5 years by rs670 (a) and HbAS (b) genotypes.**

The log parasite numbers of each subject were plotted by the genotype. Each dot stand for one observation thus the length represents the number. The median is marked in orange and the standard deviation in red. Subjects with HbAS or $\alpha^{+}$-thalassaemia were excluded from the analysis of rs670 whereas $\alpha^{+}$-thalassaemia carriers were left out from the HbAS analysis.

*Williams et al.* (2005, 3) observed a negative epistasis between $\alpha$+–thalassaemia and HbAS, both affecting *P. falciparum* erythrocytic stages. To detect possible interaction of rs670 and HbAS I examined the genotypes of both genetic traits using Wald statistical tests. A significant interaction between rs670 and HbAS was not observed (p>0.4). In summary both results indicate a spatial and/ or functional separation between the effects of both genetic traits. In contrast to HbAS I found no evidence that rs670 affects intra-erythrocytic development of *P. falciparum* or interacts

with other protective traits against this parasite stage. On this account, an effect of rs670 on pre-erythrocytic *P. falciparum* parasites is the most likely explanation.

### 4.4 Association of SNP rs670 with risk of severe malaria

In the Mild Malaria Cohort an association of rs670 with *P. falciparum* infections and time to re-infection was observed. A possible influence on malaria pathology, mainly caused by blood stages, could not be tested. To this end I genoyped rs670 in a second independent study group using a RFLP assay. This case-control study consisted of subjects from the same but larger geographic area like the previous cohort (3.1.8.2). The study was designed to test for possible enrichment of rs670 alleles in children, who were admitted to Kilifi District Hospital with severe malaria, compared to the baseline prevalence in children without known severe malaria events.

#### 4.4.1 Power calculation for design of genotyping Severe Malaria Cohort

In cohorts based on case-control designs it is helpful to calculate a benchmark of number of samples that is required to achieve sufficient statistical power.

I used the *Power for Genetic Association Analyses* (PGA) software to obtain the required sample size from the Severe Malaria Cohort (3.2.3.7). The determined hazard ratio of 0.49 in children between 1-5 years (4.3.2) carrying the G was used as relative risk to calculate the minimum sample size for an association analysis. The control to case ratio was set to equal. For a statistical power of 0.8 and α of 0.5 and using severe malaria disease prevalence of 0.02 and a frequency of the disease predisposing rs670 G allele of 0.76 I calculated a minimum sample size of 438 cases and controls respectively.

#### 4.4.2 rs670 genotyping of the Severe Malaria Cohort

For genotyping rs670 in a subset of the Severe Malaria Cohort I used an RFLP assay with the restriction endonuclease MspI (3.2.1.6). The digested PCR products were photographed and rs670 genotypes were determined manually for every individual as exemplified in figure 15. All AA homozygous samples were confirmed in a

second repeated digestion to exclude a potential enzymatic failure in the first digestion.

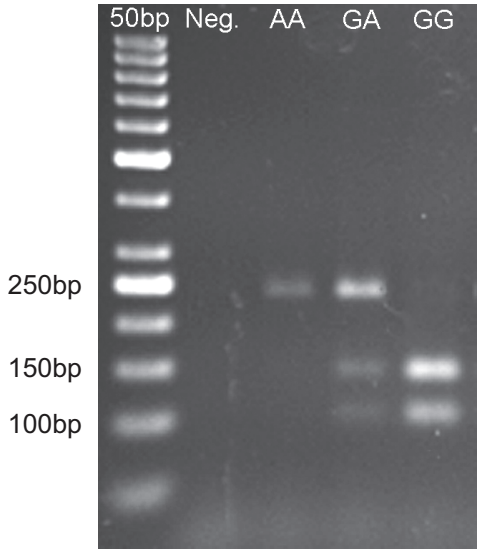



**Figure 15: Rs670 MspI RFLP assay**
PCR products were digested using MspI. The enzyme only cleaved if the G allele was present. At the top, the marker, the negative control and the genotypes of rs670 are listed. The AA genotype resulted in the 235bp PCR product. Heterozygotes showed the original and the two cleaved products with a length of 135 and 100bp. In GG carriers only the cleaved products were left.

In total I genotyped rs670 in 797 children with severe malaria and in 850 controls drawn from the population. Because of the strong influence of rs670 on *P. falciparum* infections in children below 5 all older subjects were excluded from further analyses. The rs670 genotype numbers of children with severe malaria events and controls drawn from the population for the total Severe Malaria Cohort subset and for children below 5 years are listed in table 11. Again HWE was used for quality control. Testing for deviance I did not observe any violation of HWE in the total ($p>0.5$) as well in the two single datasets ($p>0.5$).

**Table 12: Distribution of rs670 in the genotyped Severe Malaria Cohort.**
The number of subjects (n) and genotypes of rs670 are listed for the total dataset and for children below 5 years. The sums of the subjects, genotypes or the average allele frequencies are given in the last line.

| | total | | | | | | <5 years | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **n** | **GG** | **GA** | **AA** | **G %** | **A %** | **n** | **GG** | **GA** | **AA** | **G %** | **A %** |
| **sev. mal.** | 797 | 496 | 270 | 31 | 79 | 21 | 704 | 441 | 238 | 25 | 80 | 20 |
| **controls** | 850 | 585 | 244 | 21 | 83 | 17 | 768 | 538 | 211 | 19 | 84 | 16 |
| **sum/ Ø** | 1647 | 1081 | 514 | 52 | 81 | 19 | 1472 | 983 | 449 | 44 | 82 | 18 |

### 4.4.3 Lowered risk for severe malaria in rs670 G carriers

In table 11 the differences in the rs670 allele frequencies between the children suggest an enrichment of the A allele in children that suffered from severe malaria. A $\chi^2$ –test using allele frequencies showed a significant association of rs670 with severe malaria (p=0.003). Calculation of an odds ratio (OR) of 0.75 (95% CI, 0.62-0.91) suggests a 33% lower risk to develop severe malaria for G allele carriers. Comparing the three genotypes in a $\chi^2$ –test the association remained significant with a risk of 22% (p=0.01; Pearson $\chi^2$=9.29). This compares to children with HbAS that have a 426% lower risk for severe malaria events (p<0.001). In table 12 the OR for the alleles and genotypes of rs670 and HbAS are listed. No differences were observed using Fisher's exact test compared to the $\chi^2$ –test.

**Table 13: $\chi^2$ –test of rs670 genotypes in children <5 years.**

Number of cases and controls, their odds ratios with 95% CI obtained by $\chi^2$ –test are listed for rs670 and HbAS (without HbSS) as reference. The OR and CI of the alleles are obtained by Cornfields' correction.

| allele/ genotype | cases | controls | odds ratio | 95% CI |
|---|---|---|---|---|
| **A** | 288 | 249 | 1.32 | 1.10 – 1.60 |
| **G** | 1120 | 1287 | 0.75 | 0.62 – 0.90 |
| **AA** | 25 | 19 | 1.32 | 0.72 - 2.38 |
| **GA** | 238 | 211 | 1.13 | 0.93 - 1.36 |
| **GG** | 441 | 538 | 0.82 | 0.72 - 0.93 |
| **AA+GA** | 263 | 230 | 1.14 | 0.96 – 1.36 |
| **HbAA** | 674 | 564 | 1.19 | 1.06 - 1.33 |
| **HbAS** | 19 | 98 | 0.19 | 0.12 - 0.32 |

Additionally, I used logistic regression analyses to account for possible confounding factors. Because of the low number of AA homozygotes I used allele frequencies to increase statistical power. In an unadjusted logistic regression model an OR of 0.75 (95% CI, 0.60-0.90; p=0.003) for rs670 G carriers was observed, reconfirming the previous $\chi^2$ –test. Similar to previous comparison between genotypes, this effect was weaker to the well-established *HBB S* allele (OR 0.22, 95% CI 0.14-0.34; p<0.001). An adjustment for age, sampling location, ethnic group, HbAS and $\alpha$+–thalassaemia

lowered the OR to 0.63 (95%, 0.47-0.90; p=0.009) for rs670 G whereas *HBB* S remained at almost the same level (OR 0.20; 95% CI, 0.1-0.38; p<0.001). In Figure 16 the unadjusted odds ratios of rs670 and *HBB* alleles are shown. It illustrates the weak effect of rs670 G on the odds to develop severe malaria compared to *HBB* S. The difference in strength of the effect suggests an influence of rs670 on pre-erythrocytic *P. falciparum* stages in contrast to HbAS that directly affects development of intra-erythrocytic stages. Again testing for epistasis between rs670 alleles with *HBB S* no interaction was found (p=0.8) using Wald statistics. HbAS and $\alpha$+– thalassaemia showed a borderline interaction (OR 3.7, 95% CI 0.95-14.9 p=0.06).

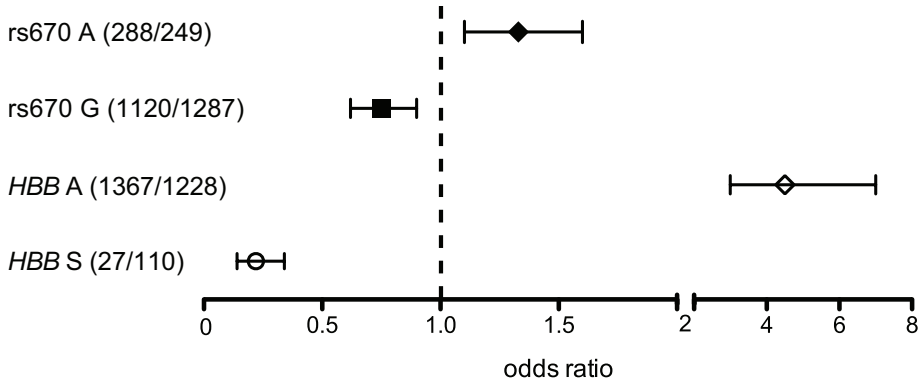



**Figure 16: Odds ratio of rs670 and *HBB* alleles in children <5 years.**
Unadjusted OR are graphed with 95% CI by rs670 and *HBB* alleles. Numbers in brackets show the allele number in children with/ without severe malaria.

### 4.5 Screening for signatures of selection in *APOA1, FABP1* and *SNAPAP*

Malaria is the strongest known force for evolutionary selection in the recent history of the human genome (*Kwiatkowski, 2005*). Different signatures reveal old or recent selection events occurring in the human genome. They include skewed allele frequencies between populations or conserved haplotypes. Here, I used different statistical approaches to discover such signatures in the three targeted genes *APOA1, FABP1* and *SNAPAP.* HbAS was used as reference. It is the best-described protective genetic trait against malaria. Despite the high and life-threatening morbidity in SS homozygotes its allelic frequency is elevated in malaria endemic regions.

### 4.5.1 rs670 A allele is present at low to intermediate frequencies around the world

Allele frequencies differ between populations as a result of random genetic drift or if they are under selection of local environmental factors. In the two study sites of the Mild Malaria Cohort, Chonyi and Ngerenya, rs670 genotypes are similar distributed with frequencies of 0.24, respectively (Figure 18). However, in the subset of the Severe Malaria Cohort representing a larger geographic area, I found a lower A allele frequency of 0.17. This variation in frequencies is probably caused by the different ethnic composition between both sampled populations in the cohorts.

In the combined data set of the total Mild Malaria Cohort and the children without severe malaria I calculated an A allele frequency of 0.2. I excluded the children with a past severe malaria event because of enrichment of the A allele due to its higher risk of severe malaria.

Because of its association with different diseases rs670 was genotyped in many epidemiological studies. In Figure 14 I mapped these frequencies for different countries around the world. Highest allele frequencies are found in central Asia with dropping frequencies towards South-East Asia. American and European populations show a low to intermediate frequencies (0.1-0.19) whereas lowest frequency was observed in Nigeria. These results indicate a decrease in A allele frequency from east to west. In regard to the negative selection of the A allele by malaria Kenya (East Africa) could be the gateway between the high abundance of the A allele in Asia and low in Africa.

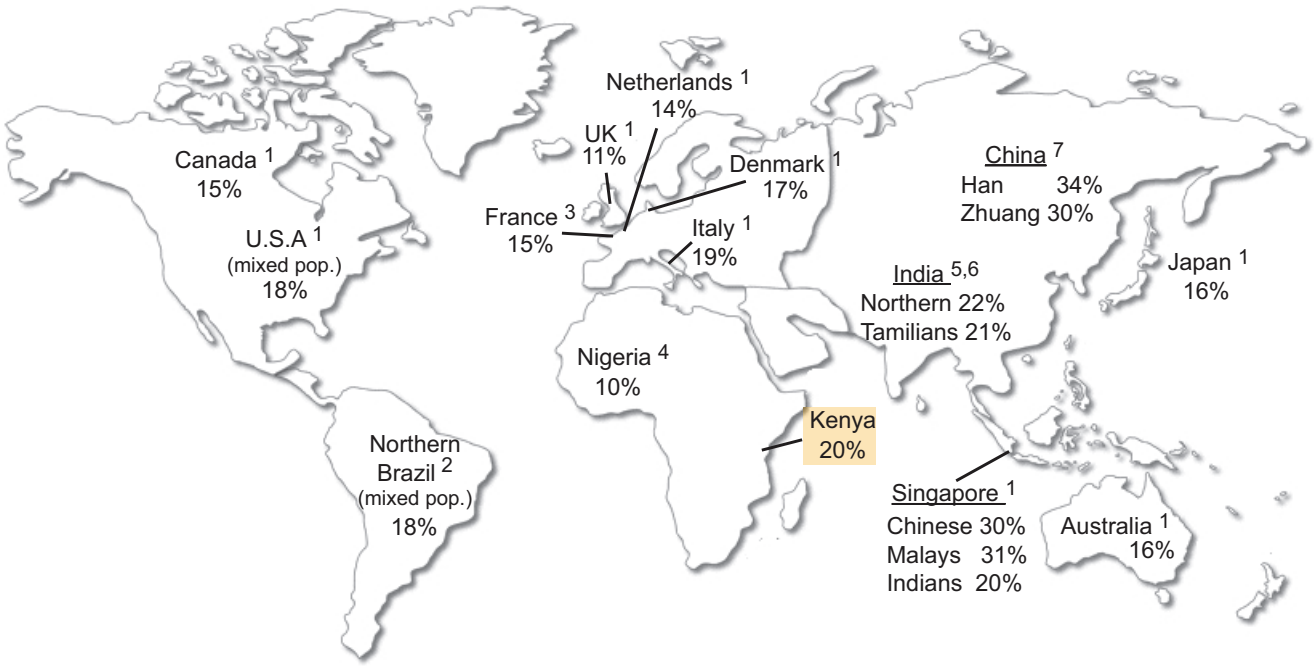



**Figure 17: World map of rs670 A allele frequencies**

Allele frequencies of the minor rs670 A allele are given in percent for different populations (1: reviewed in *Juo et al., 1999*; 2: *de Franca et al., 2005*; 3: *Helbecque et al., 2008*; 4: *Kamboh et al., 1999*; 5: *Chhabra et al., 2005*; 6: *Padmaja et al., 2009*; 7: *Li et al., 2008).*

### 4.5.2 Haplotype blocks in *APOA1, FABP1* and *SNAPAP*

The linkage disequilibrium (LD) composition in *APOA1, FABP1* and *SNAPAP* was examined to find signatures of selective events in the targeted genes. Haplotypes are the combinations of alleles in a population. Haplotypes can be disrupted by mutation, genetic drift and/or recombination. Events causing a fragmentation of haplotypes lead to haplotype blocks (*Gabriel et al., 2002*). The markers localized in these blocks are in linkage disequilibrium (LD), the non-random association of alleles at different sites. The size and number of these blocks show the evolutionary history of the genomic region.

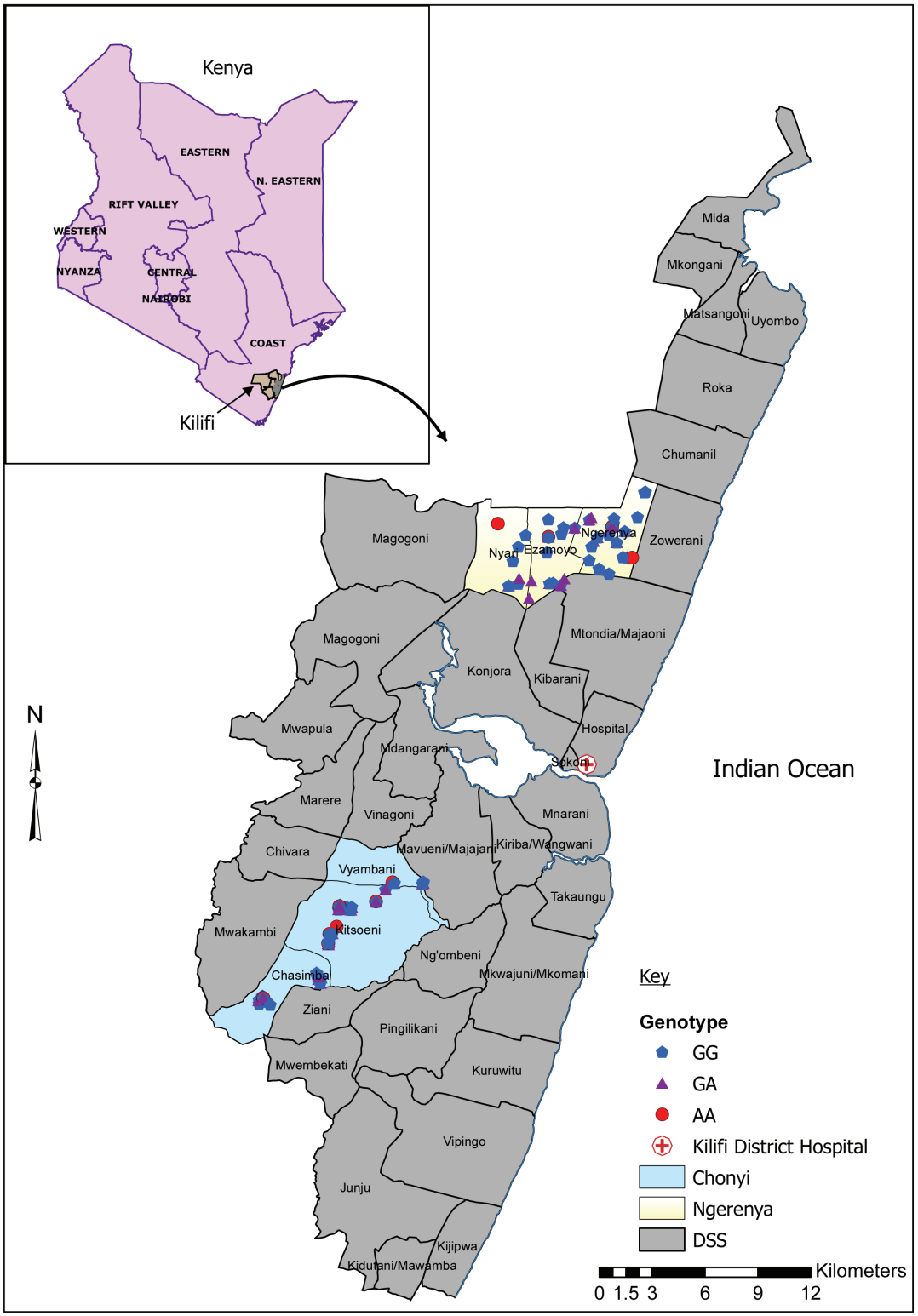



**Figure 18: Distribution of rs670 genotypes in children 1-5 years from the Mild Malaria Cohort.**
Genotpyes of rs670 are shown at the two study sites Chonyi and Ngerenya. Subjects from the Severe Malaria Cohort were sampled in the whole area of the DSS surveillance.

Old and common mutations will be found in a smaller haplotype block than younger with lower frequencies. Patterns of LD and haplotype blocks can indicate signatures of recent selection (*Nordborg et al., 2002*). Between two haplotype blocks recombination events probably have occurred in the evolutionary history of the population (*Wang et al., 2002*).

I calculated haplotype blocks from the Mild Malaria Cohort genotype datasets for the three genes with Haploview v. 4.1 software (3.2.3.10). The LD between the marker and calculated haplotype blocks are visualized in Figure 19:

Around the genomic region of *APOA1* all genotyped markers are in higher LD compared to *FABP1* and *SNAPAP* indicating a higher conservation in the evolutionary history of the population. However, only two haplotype blocks could be determined. One reaches out from the associated marker rs670 in the 5'UTR into the 1$^{st}$ intron with rs5069. The 2$^{nd}$ block ranges from marker rs5073 in the 3$^{rd}$ intron to rs5081 in the 3'UTR framing the 4$^{th}$ exon.

Compared to *APOA1* the LD between the tested markers of *FABP1* is low. Two haplotype blocks were calculated at both ends of the gene. In the 1$^{st}$ intron two markers (rs13400963 and rs7577401) forming a block. Despite the low LD (D'=0.01) only three of the 4 possible two-marker haplotypes are found above the frequency of 0.01 thus it is unlikely that a recombination is deemed to have taken place. Another spans of 1kb spanned across the 3$^{rd}$ intron from marker rs1123605 to rs2197076. Again the LD within this block is low but in all combinations only 3 gametes were present.

In general the genotyped markers in *SNAPAP* are in high LD. Especially marker rs35903030 in the 1$^{st}$ exon and marker rs1802461 in the 4$^{th}$ exon are in maximum LD (D'=1) with the other genotyped SNPs. Again two haplotype blocks could be identified: The first ranges from the 1$^{st}$ exon with marker rs35903030 across the 2$^{nd}$ exon to marker rs16835481 in the 2$^{nd}$ intron. The 2$^{nd}$ haplotype block is found from rs12755522 in the 3$^{rd}$ intron to the last tested marker rs7345 in the 3'UTR.

These LD patterns indicate a higher evolutionary conservation of *APOA1* and *SNAPAP* than *FABP1*. Intriguingly in *APOA1* and *SNAPAP* the 4$^{th}$ exon seems to be conserved. In FABP1 all markers close to the 4$^{th}$ exon are in LD indicating a conservation of this region as well. The high LD in region of the 4$^{th}$ exon is possibly caused by essential coding sequences for protein functions.

A reason for the observed LD patterns between these genes can probably be found in the different recombination rates between the three genes, calculated in the Hapmap Phase II data (*Winckler et al., 2005; The International HapMap Consortium et al., 2007*). The decay of LD, and thus the relative size of haplotype blocks, depends on local recombination rates. In *FABP1* and particular before the 3$^{th}$ exon recombination rates between rs13400963 and rs6711957 achieved 9 cM/ Mb marking this region as a recombination hot spot. Between rs2241883 and rs2197076 the recombination dropped to 1cM/ Mb, which was still almost ninefold higher compared to *SNAPAP* (0.12 cM/ Mb). Around *APOA1* no recombination rates could be calculated due to the low number of genotyped SNPs in HapMap.

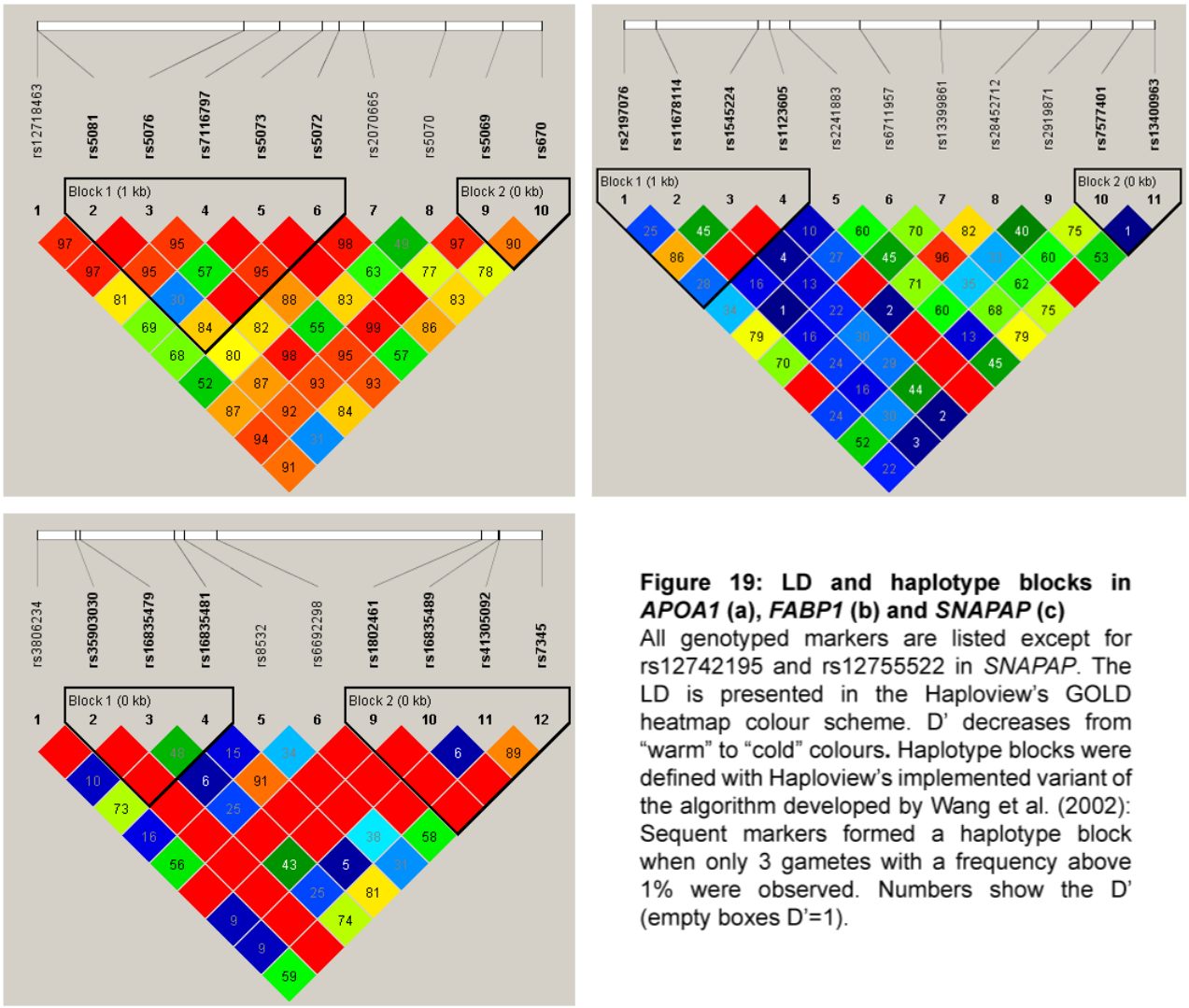



**Figure 19: LD and haplotype blocks in *APOA1* (a), *FABP1* (b) and *SNAPAP* (c)**
All genotyped markers are listed except for rs12742195 and rs12755522 in *SNAPAP*. The LD is presented in the Haploview's GOLD heatmap colour scheme. D' decreases from "warm" to "cold" colours. Haplotype blocks were defined with Haploview's implemented variant of the algorithm developed by Wang et al. (2002): Sequent markers formed a haplotype block when only 3 gametes with a frequency above 1% were observed. Numbers show the D' (empty boxes D'=1).

### 4.5.3 Haplotype analysis of rs670 revealed G as ancestral allele

Under neutral evolution, new alleles require a long time to achieve high frequency in the population and recombination events will break down the LD in the genomic region (*Sabeti et al., 2002*). Accordingly, old and common alleles have only LD over a short distance on the chromosome whereas younger ones can have either long- or short-range LD. Haplotypes with high levels of haplotype homozygosity in combination with a high frequency in the population indicate evidence of recent positive selection.

To find signatures of selection, I examined the haplotype structures and distribution of rs670 by calculating individual haplotypes from all subjects of the Mild Malaria

Cohort. The bifurcation diagrams of haplotypes with the rs670 alleles as core SNP are visualized in figure 20 using SWEEP (3.2.3.12.1).

The haplotypes containing the rs670 G allele are highly branched at each of the 10 genotyped SNPs thus the haplotype frequencies are dropping to low frequencies with ongoing distance from the core. This fast decay of LD is not seen with the A allele as core SNP. Latter was highly abundant in a single major haplotype. From the core to the 7$^{th}$ marker (rs7116797) bifurcations with only low frequencies arise from every node. At marker rs5076 the major haplotype was splitted into two with comparable frequencies. In comparison, the low bifurcation rate of the A allele indicate a conserved evolutionary history or recent appearance in population history. The G seems to be ancestral allele. The latter conclusion is supported by a G allele in that position in Chimpanzees, which are used as genetic ancestors to humans in the UCSC database.

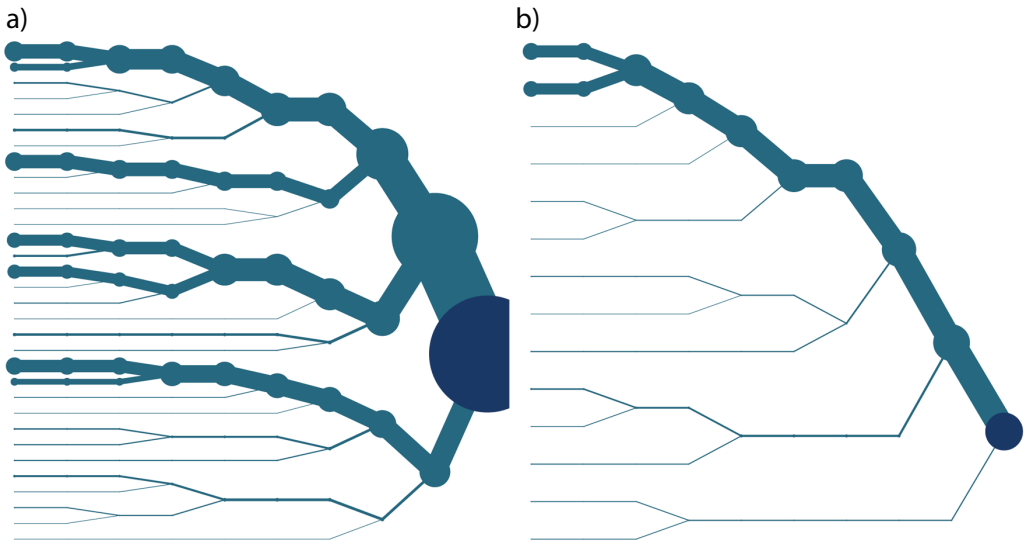



**Figure 20: Haplotype bifurcation diagrams of *APOA1*.**
The alleles G (a) or A (b) of rs670 are set as core SNP, indicated by a dark blue circle. Markers are ordered as seen in table 7. Moving from rs670 to the end of *APOA1* each node represents a genotyped marker. The diagram divides if both alleles are present. The thickness of the lines corresponds to the number of samples carrying the haplotype.

### 4.5.4 No signatures of selection in the genomic region of *APOA1, FABP1* and *SNAPAP*

The genomic region around *APOA1, FABP1* and *SNAPAP* were examined for signatures of selection with the web-based Haplotter software. Applying the gene-centered approach the iHS, *Fay and Wu's* H, *Tajima's* D and the *Wright fixation index* ($F_{ST}$) were calculated in 500kb of the proximal and distal genomic region from each gene, respectively. Each of theses statistical approaches provide information on different time scales in human history (Figure 5; Details: 3.2.3.12.2). Briefly, alleles under selection are embedded in extended haplotypes showing a high homozygosity that can be measured using iHS statistics. Due to selection allele frequencies can be skewed compared to the neutral model. This can be determined with *Fay and Wu's* H and *Tajima's* D. These tests differ in statistical power for detecting selection events and in their dependence on fixation of the marker in the population.

For all calculations the HapMap Phase II dataset was used. Results of the statistical tests are graphed for *HBB* and *APOA1* in Figure 21 and for *FABP1* and *SNAPAP* in Figure 22. Here, only the CEU and YRI populations were compared in regard of different selection pressure from *P. falciparum* malaria. The East Asian (ASN) dataset was ignored due to fusion of Japanese individuals from Tokyo and Han Chinese from Beijing (*Voight et al., 2006*).

Around *APOA1* no significant evidence of selection was observed in the two CEU and YRI datasets. The highest iHS of 0.6 was seen in the YRI population. In the vicinity of *APOA1* the highest iHS peak was observed in the genomic region of the serine/threonine-protein kinase KIAA0999 at 200kb downstream of *APOA1*. A comparable result was observed with *Fay and Wu's* H. The highest H value with 0.4 was measured in the CEU population suggesting low high frequencies of derived alleles thus objecting a selection event. *Tajima's* D showed almost the same characteristics. The highest value was observed in the CEU population of 1.75 whereas 1.25 was measured in the YRI population. This indicates that *APOA1* might not have evolved neutrally in the CEU population. However, a comparison between the CEU and YRI populations with a $F_{ST}$ of 1 showed a standard degree of population differentiation.

In *FABP1* no signatures of selection were observed in the CEU or YRI datasets. Both iHS scores of the CEU and YRI population reached the value of 0.3 indicating presence of only short haplotypes. *Fay and Wu's* H and *Tajima's* D showed no different frequency shifts from the neutral model in both populations. Both approaches had a maximum of 0.3 in the two populations. In almost the same manner the $F_{ST}$ value reached only 0.3 indicating a shared evolutionary background of this locus in both populations.

Like in *APOA1* and *FABP1* no high iHS score was observed in the genomic region of *SNAPAP* in CEU or YRI. The maximum value reached 0.6 in the CEU population. Also, *Fay and Wu's* H and *Tajima's* D showed no indication for a selective event in evolutionary history. The H score of 0.8 in the CEU was higher than in the YRI population with 0.3. *Tajima's* D score showed with 0.25 in CEU a slightly higher shift in allele frequencies than in YRI with 0.1. In both populations no signature of selection was observed compared with the $F_{ST}$ statistics. The value of the CEU vs. YRI score showed a maximum of 0.25.

For the HBB, which I used as reference, I also failed to find indications of long-range haplotypes in *HBB* due to a low iHS score of 0.5. The only significant iHS score in vicinity is the *Olfactory receptor 51B2* (*OR51B2*) around 200kb upstream of *HBB* in YRI. Furthermore, the score of *Fay and Wu's H* achieved only a low value of 0.1. Moreover, the low *Tajima's D* score of 0.1 indicating normal heterozygosity in this genomic region. In comparison of both populations I found no significant difference between the YRI and CEU reflected by the low $F_{st}$ score of 0.5.

In summary I found no evidence of signatures of selection either in APOA1, FABP1 and SNAPAP, or in *HBB*. In contrast to *FABP1* and *SNAPAP* in which no selection is known, the absence of any signature of selection in *APOA1* or at least *HBB* is surprising.

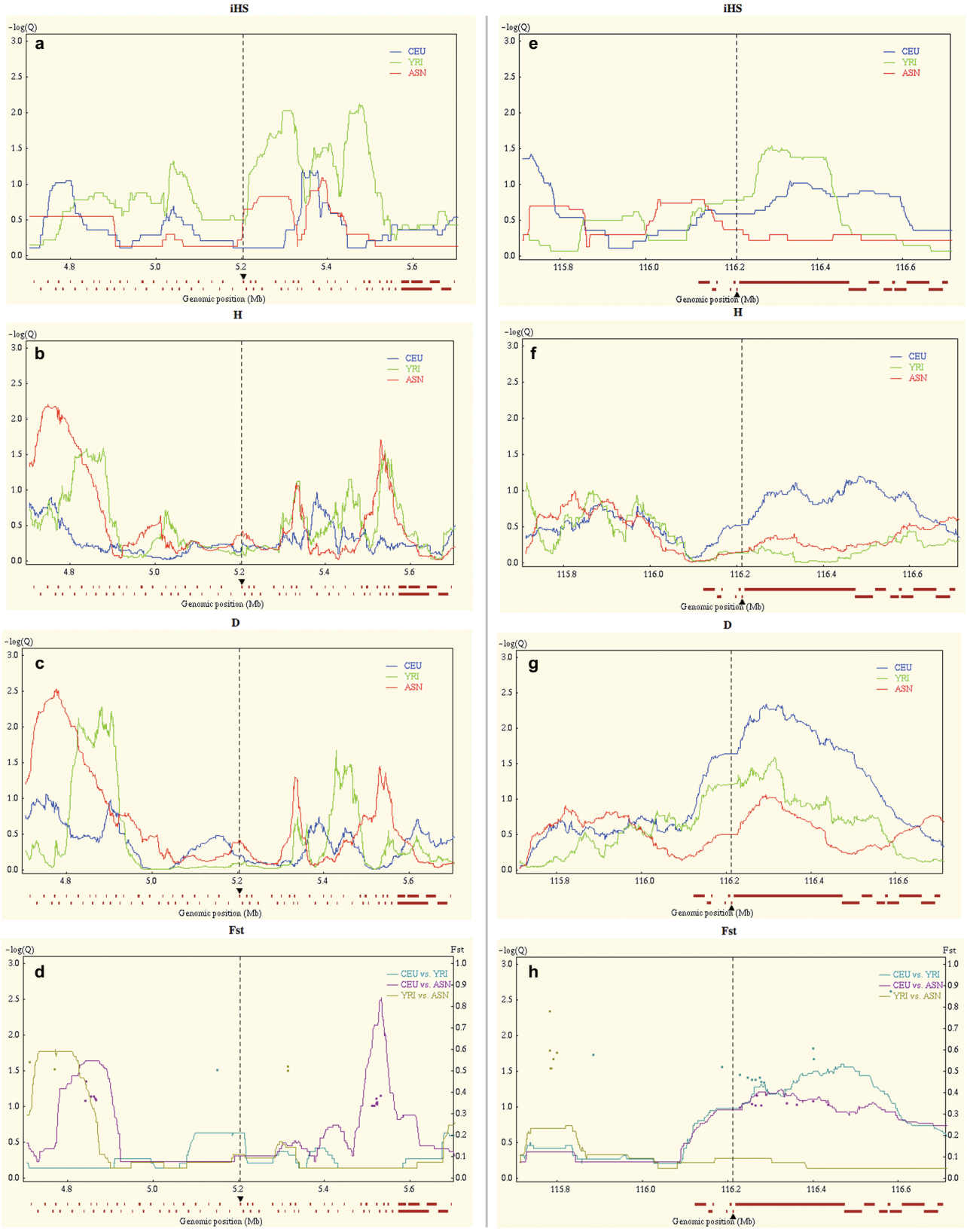



**Figure 21: Screen for signatures of selection in *HBB* and *APOA1***
Graphs show score of iHS, Fay and Wu's H, Tajima's D and Fst in the genomic region of *HBB* (a,b,c,d) and *APOA1* (e,f,g,h) in the CEU, YRI and ASN populations. Position of the gene (dotted line) is framed by 0.5Mb up- and downstream of the chromosome. Red blocks below x-axes represent genes in the genomic region.

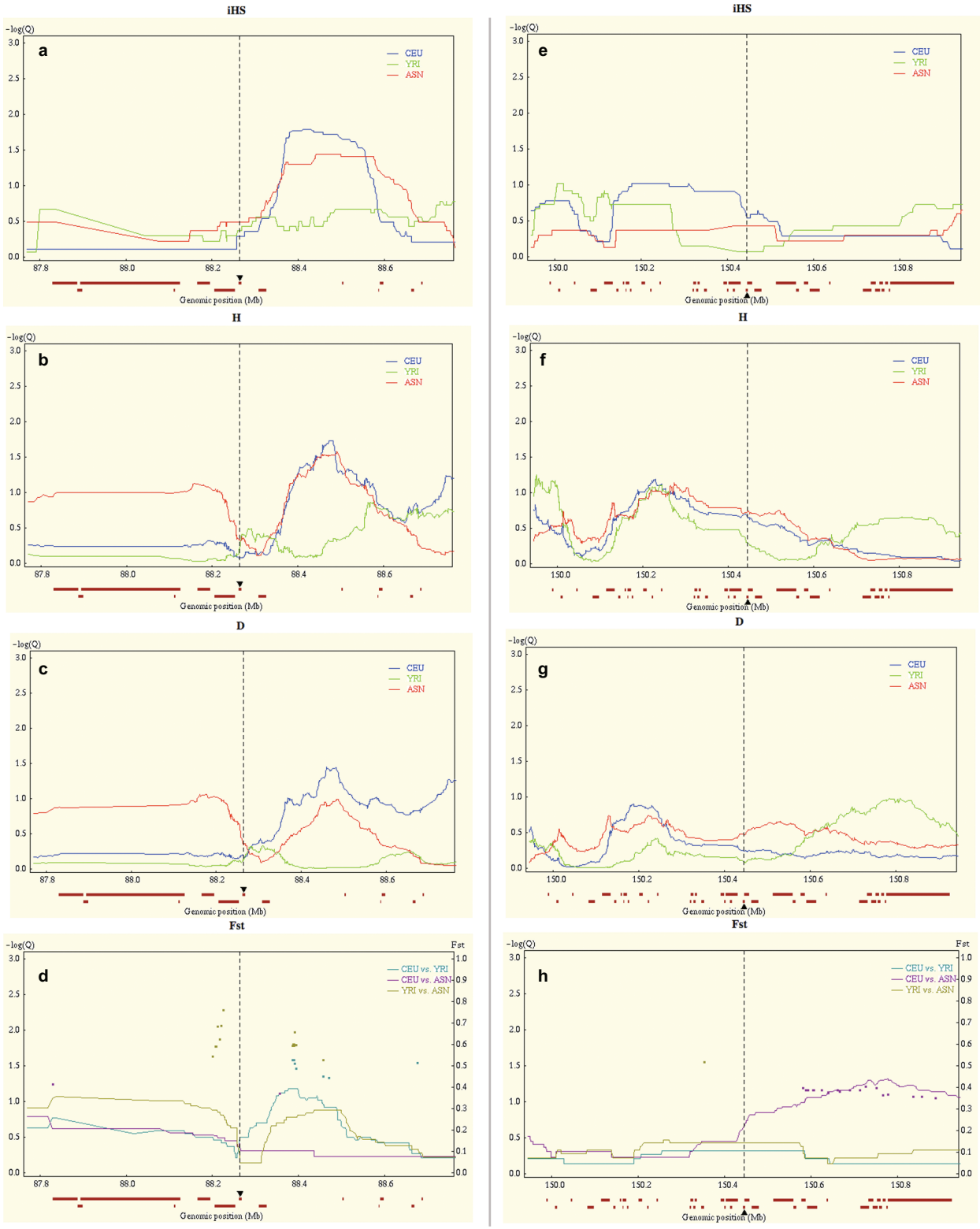



**Figure 22: Screen for signatures of selection in *FABP1* and *SNAPAP***
Graphs show score of iHS, Fay and Wu's H, Tajima's D and Fst in the genomic region of *FABP1* (a,b,c,d) and *SNAPAP* (e,f,g,h) in the CEU, YRI and ASN populations. Position of the gene (dotted line) is framed by 0.5Mb up- and downstream of the chromosome. Red blocks below x-axes represent genes in the genomic region.

### 4.6 Influence of APOA1 on *P. berghei* liver cell stage development

The development of *P. falciparum* liver cell stages seems to be dependent on cholesterol (*Mikolajczak et al., 2007; Rodrigues et al. 2008).* HDL and its major component APOA1 is the key provider for cholesterol into the liver *(Yalaoui et al., 2008).* rs670 is associated with APOA1 expression but data from previous *in vitro* studies are contradictory (*Tuteja et al. 1992; Agnotti et al. 1994)*. Therefore, I aimed to determine the influence of APOA1 on the Plasmodium liver stage development directly. A previous knock-down of APOA1 using siRNA showed an impaired development of *P. berghei* (*Müller AK, personal communication*). For testing the effect of an oversupply of APOA1 I performed experiments in transfected Huh7 hepatocytes, overexpressing rat APOA1, which were infected with *P. berghei* (3.2.2.5).

First I aimed to measure the difference in development of *P. berghei* liver cell stages in WT and the transfected Huh7 hepatocytes. In parallel, the same experiment was performed in murine Hepa 1.6 hepatocytes. A second experiment intended to quantify the difference in growth in two transfected Huh7 cell lines: One overexpressing rat APOA1 and the other was transfected with a mock vector to account for the transfection. To measure differences in number and morphology of liver cell stages, they were stained by IFA (3.2.2.6.3). For quantification of *P. berghei* liver cells stages I used a real time PCR assay (3.2.2.6.2).

#### 4.6.1 IFA staining revealed increased size of *P. berghei* liver stages in APOA1 overexpressing hepatocytes

The liver cell stages were stained with anti-hsp70 antibody from mice and green-fluorescent labelled anti-mouse antibody. The DNA from hepatocytes and parasites were stained with Hoechst. In the 1$^{st}$ experiment the wild type was compared to the transfected Huh7 cell line. This experiment was done once with duplicates for each cell line at three time points post invasion of *P berghei*. Because Hepa1.6 hepatocytes showed an impaired growth they were excluded from analyses.

In Figure 23 the liver cell stage numbers were plotted for Huh7 hepatocytes at different time points post invasion (pi). At time point 24h pi the number of infected hepatocytes was around 30% higher in the WT compared to the transfected, APOA1

overexpressing Huh7 (ApoTrans). 48 hours pi the liver stages in the WT were reduced to the level of the ApoTrans cell line. The latter showed constant liver cells stage numbers suggesting appropriate growth conditions in contrast to WT. Last count was done on time point 60h pi. Here, I observed a drastic reduced number of liver cell stages in both cell line likely due to evasion of matured merosomes after 50 hours post invasion. The remaining appeared to have a lagged development indicated by their small size that was similar to 24h pi.

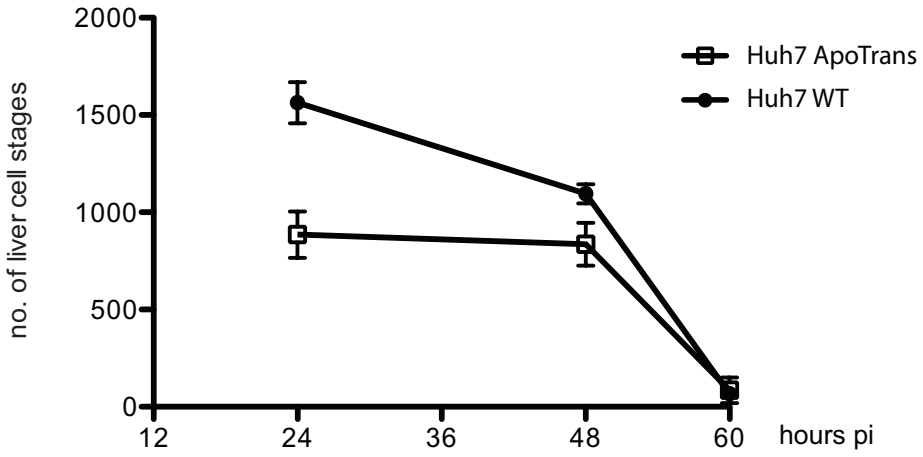



**Figure 23: Number of *P. berghei* liver cell stages in WT and ApoTrans.**
*P. berghei* liver cell stage numbers were counted at 24, 48 and 60 hours post invasion.

I determined differences in liver cell stage sizes of 50 liver cell stages (25 per duplicate) at every time point and for each cell line. *P. berghei* liver cells stages were photographed and the green-fluorescent area of the parasite was measured (3.2.2.6.3). In Figure 23 all size measurements are displayed except from last time point of 60h pi.

After 24 hours pi the liver stages of both cell lines showed a comparable mean circumferential area size whereas at time point 48h the means of the area size between the two cell lines differed. Although the sizes were highly variable I found significantly larger liver cell stages in the overexpressing APOA1 Huh7 compared to the wild type in an unpaired two-tailed *Student t-test* analysis ($p<0.001$). In compari-

son the mean circumferential area was 25% bigger and their oval shape suggested a higher level of maturity in presence of high level of APOA1 (Figure 23 a,b). A *Welch*-corrected student *t-test* was used to account for significant variances of the standard deviations between the two cell lines ($p<0.01$). In contrast to 48h no significant difference was observed between the both cell lines at 24h pi ($p>0.5$).

In the second experiment using a Huh7, transfected with a different mock-vector, I could not confirm the results. Measuring the circumreferential area size no significant difference between the mean sizes was observed (data not shown). The number of *P. berghei* liver cell stages between the two transfected cell lines did not differ.

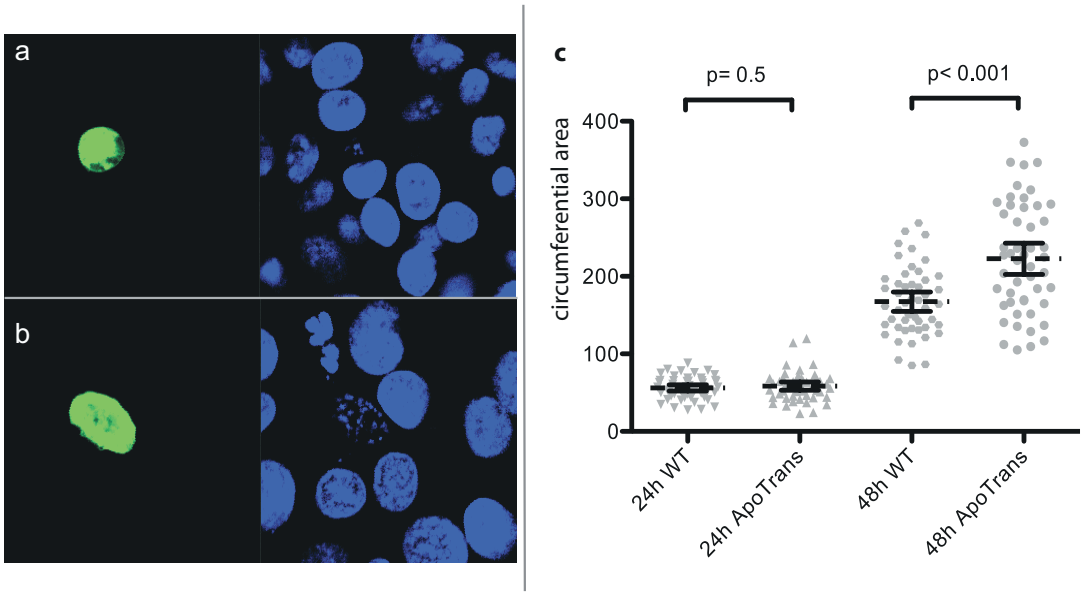



**Figure 24: Size of *P. berghei* liver cell stages in WT and ApoTrans.**
**a,b**: *P. berghei* liver cell stage sizes of WT (a) and ApoTrans (b) at 48h pi labelled with α-hsp70 (green) and with Hoechst stained DNA (blue). **c**: Size measurements are mapped at the Y-axis with a dimensionless circumferential areas size for WT and ApoTrans at 24 and 48h pi. Mean sizes are given by the dotted line. Error bars represent 95% CI.

### 4.6.2 Quantification using real-time PCR indicates elevated Pb18s in APOA1 overexpressing hepatocytes.

For a quantification of *P. berghei* liver cells stages I used a real-time PCR assay. I aimed to measure the amount of Pb18s between infected cell lines at different time points post invasion using cDNA. Pb18s was normalised using host *GAPDH* as internal control to account for differences in isolation of RNA and amplification to cDNA between the samples (3.2.2.6.2). Expression of rat and host APOA1 was controlled.

For Hepa 1.6 and Huh7 the relative expression of Pb18s is graphed for each time point in Figure 25 (a,b). Each cell line was set into relation to the first measurement at time point 24h.

In Huh7 and Hepa 1.6 the WT strain shows a decreasing expression of Pb18s at the subsequent time points. In the APOA1 overexpressing Hepa1.6 cell strain a maximum amount of Pb18s was observed at 48h pi, whereas in the transfected Huh7 only a slight decrease of Pb18s was observed. At 60h the amount of Pb18s was reduced severely in all cell lines. However, the high variance within the ApoTrans and WT samples prevents any conclusions at this last time point.

For comparing differences in growth between the WT and transfected cell lines each time point was set in relation to the WT at 24h pi. The results are graphed for Hepa1.6 and Huh7 in Figure 25 (c,d).

Like in Figure 25 a,b the amount of Pb18s decreases from 24h to subsequent time points in the WT strains. In the transfected Hepa 1.6 a maximum of Pb18s was seen at 48h pi whereas it remained almost stable in the transfected Huh7. At the last time point of 60h pi the expression level was comparable between the transfected and the wild type strains.

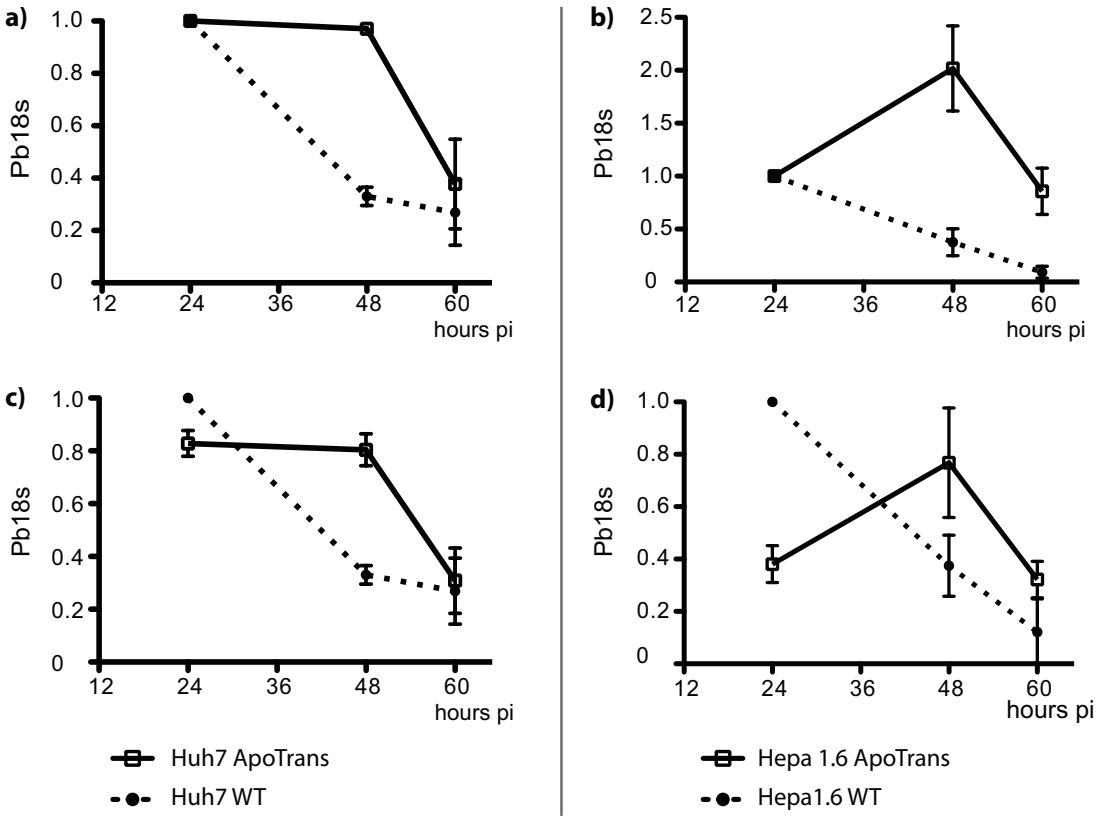



**Figure 25: Pb18s in Huh7 (a,c) and Hepa 1.6 (b,d) cell lines at 24, 48 and 60 hours post invasion**
Amount of Pb18s in WT and transfected hepatocytes at 48h and 60h were related to 24h for Huh7 (a) and Hepa 1.6 (b). Below, Pb18s in Huh7 (c) and Hepa 1.6 (d) are set in relation to the specific WT at 24h. Error bars represent standard deviations.

Calculating the ratio of Pb18s between ApoTrans and WT strains at each time point I found a significantly elevated level in Huh7 ($p=0.001$) and Hepa 1.6 ($p<0.001$) at 48h compared to 24h using a unpaired two-tailed *Student t-test* (Figure 26). In the transfected strains Pb18s was around 2.5 times higher compared to the WT at 48h pi. The quantity of Pb18s was similar between transfected and WT hepatocytes at 24h or 60h pi.

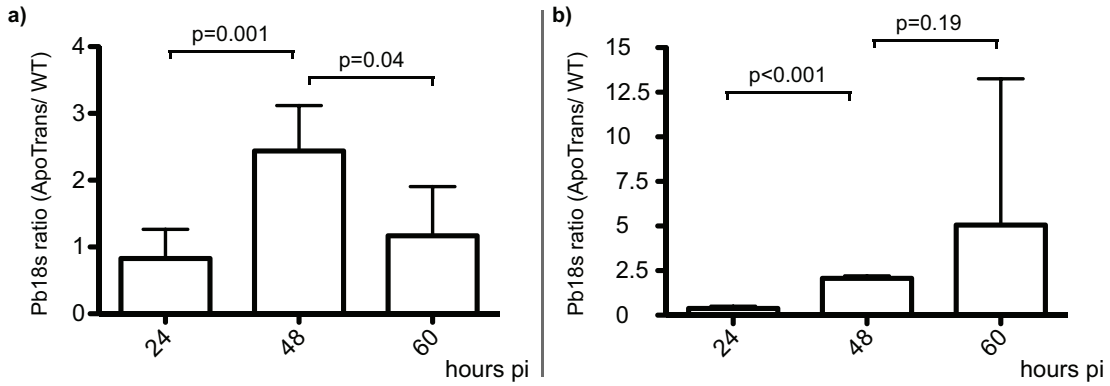



**Figure 26: Pb18s ratio of ApoTrans/ WT in Huh7 and Hepa 1.6**
The mean ratio of Pb18s in Huh7 (a) and Hepa 1.6 (b) at 24, 48 and 60 hours post invasion is given by the dotted line. Error bars represent standard deviations.

In the second experiment designed to control for the transfection event I could not confirm these results. Quantification of Pb18s in the ApoTrans and the mock-transfected strain showed no difference at any time point could be observed (data not shown). Both transefected cell lines differed in their duration of their cultivation. The mock-transfected cell linie, transfected shortly before the experiment, showed a decelerated growth compared to the ApoTrans cell line.

In summary, I found a higher amount of Pb18s in the APOA1 overexpressing strains than in the WT indicating stable growth conditions in transfected Huh7 or even an increased growth of *P. berghei* in transfected Hepa 1.6 hepatocytes. The increase of Pb18s at 48h suggests an essential role of APOA1 for the development of liver cell stages at this time point. With regard to the finding of comparable numbers of liver cell stages but enlarged size the increase of Pb18s suggests a correlation of Pb18s to size or higher metabolic rate in *P. berghei* liver cell stages. However, a second experiment using a transfected cell line, which is not expressing rat APOA1 did not reinforce these results. The observed constrained growth of the mock-transfected Huh7 compared the ApoTrans could have been an explanation.

# 5. Discussion

The recent detection of tolerance against artemisinine, the central components in current antimalarial combination drugs, underscores the vital need for new affordable drugs and innovative malaria intervention strategies (*Dondorp et al., 2009*). Critical to present malaria control efforts is the reduction of *Plasmodium* sporozoite transmission from mosquito to man. To date, the combination of validated intervention tools, including vector control, exposure prophylaxis and an anti-sporozoite subunit vaccine, termed RTS,S/AS01, achieve significant reductions of sporozoite inoculations to the population at risk (*Bejon et al., 2008*). Ideally, these anti-transmission strategies may be complemented by interventions that target the intra-hepatic expansion phase of the malaria parasite before the onset of clinical symptoms, which are exclusively caused by asexual blood stage parasites (*Hoffman and Doolan, 2000*).

Until to date, the roles of *Plasmodium* liver stage/ host interactions in malaria disease outcome remain unknown. After invasion of a suitable hepatocyte, the parasite resides within the parasitophorous vacuole (PV), which separates the parasite from the cytoplasm of the host cell (*Labaied et al., 2007)*. Two parasite-encoded PV transmembrane proteins, termed “Up regulated in Sporozoites proteins 3 and 4” (UIS3 and UIS4) have been assigned an essential role in the development of the *Plasmodium berghei* liver stages (*Müller et al., 2005 1,2*). Yeast-Two-Hybrid screens of whole hepatocyte transcript libraries for factors discovered interactions of host ligands with the *Plasmodium* UIS4 and UIS3 proteins (*Mueller et al., personal communication*). Apolipoprotein A1 (APOA1) and liver fatty acid binding protein 1 (FABP1) were isolated as binding partners for UIS4 and UIS3, respectively. A third ligand called SNAP-associated protein (SNAPAP) was identified in *P. berghei* to interact with the UIS7 protein. In contrast to UIS3 and UIS4 no homolog of UIS7 is found in *P. falciparum*.

In this thesis, I wanted to test whether human polymorphisms that alter the host-parasite liver stage interactions, as exemplified by UIS4/APOA1, UIS3/FABP1 and UIS7/SNAPAP correlate with risk of natural infections in a malaria-endemic area. Accordingly, I genotyped and analysed single nucleotide polymorphisms (SNPs) in two independent well-characterised groups of children from the Kenyan coast (*Williams et al., 2005 1,2,3; Wambua et al., 2006*) using the cost-effective SNaPshot

primer extension assay with its high accuracy *(Pati et al., 2003)* and then determined their relationship with susceptibility to and outcome of *P. falciparum* infections. SNPs can be directly responsible for the observed phenotypes or be in linkage with the causal mutations. Genome wide approaches have been commonly used in European or Asian populations but they have been problematic in Africa due to the high ethnic diversity (*Jallow et al., 2009*). Candidate gene approaches have a high SNP resolution of the targeted genomic region but they are susceptible to false positive discoveries due to population structure (*Altshuler et al., 2008*). Here, I attempted to use information about ethnic groups in the cohort data for controlling for potential bias.

Out of the 33 analyzed SNPs only rs670 (G/A) located in the 5' untranslated region (UTR) of *APOA1* was significantly associated with incidence and time to re-infection with *P. falciparum.* Interestingly, I did not find any association in *FABP1*, which is described to affect development of *P. yoelii* liver cell stages (*Mikolajczak et al., 2007*). In contrast the lack of association with *SNAPAP* is unsurprising because *P. falciparum* does not appear to encode for a UIS7 homologue.

Analogous to the fading protective effect of HbAS with increasing age (*Williams et al., 2005 (1)*), the association of rs670 appeared to vary with age of the patient, possibly due to maternally transmitted antibodies below one year of age and acquisition of malaria-specific immunity in individuals older than 8 years; although the effect of HbAS did not fade in the studied subset of the Mild Malaria Cohort. All further analyses concentrated on children from 1 to 5 years as the most susceptible age group for *P. falciparum* malaria. In this group protection against *P. falciparum* infections of almost 50% in children carrying the rs670 G allele was comparable to the effect of HbAS. This protective effect is most likely caused by the time differences between subsequent infections. In support of this notion, I observed a two-fold increase in time to infection in GG homozygotes when compared to AA allele carriers and half the time in children with HbAS in this age group. This observed effect of HbAS on time to re-infections was consistent with the findings in a longitudinal study of 225 individuals in Mali *(Crompton et al., 2008)*. The prolonged time until the parasites are detectable in the blood suggests a difference in development of parasites due to both genetic traits.

Because it is hardly possible to measure directly the *P. falciparum* infections in the liver of patients I used presence of asexual parasites in the blood as marker. For this reason I could not determine the direct influence of rs670 on liver cell stages. A delayed appearance of asexual blood stages could be explained by a slowed development in the liver or an inhibited growth of intra-erythrocytic stages like described for HbAS (*Pasvol et al., 1978, Kwiatkowski, 2005; Williams, 2006*). In support for a liver stage-specific effect, the rs670 SNP does neither affect parasite blood stage parasite densities nor display epistasis with other blood-stage specific protective traits, e.g. $\alpha$+-thalassaemia with HbAS, as it was observed for HbAS *(Williams et al., 2005, 1; Williams et al., 2005, 3)*. Multivariate analyses adjusted for potential confounding demonstrated independent protective effects for rs670 and HbAS, but not $\alpha$+-thalassaemia. The latter showed no association on uncomplicated *P. falciparum* incidence or on parasitaemia as expected (*Wambua et al., 2006*).

To minimize the risk of a false-positive association the effect of rs670 was confirmed in a second independent study group. Moreover, the risk to develop symptoms of severe malaria was evaluated. I observed 33% lower odds for developing severe malaria in children below 5 years carrying the G allele suggesting only a weak effect on disease outcome and thus likely no effect *P. falciparum* erythrocytic stages. This compares to the strong effect of the well-established *HBB S* allele influencing blood stage development.

Because APOA1 is one of the best-known risk markers to coronary heart disease (*Chhabra et al., 2005; McQueen et al., 2008*) numerous studies of rs670 in different countries have been performed in the past. In many epidemiological studies the A allele was associated with a higher level of APOA1 and HDL but others could not confirm these observations. *(Jenah et al., 1990; Xu et al, 1993; Needham et al., 1994; Talmud et al., 1994; Wang et al, 1996; Kamboh et al; 1999*). A systematic meta-analysis suggests an average increase of APOA1 plasma level by 5mg/dl in the A allele carriers (*Juo et al., 1999*).

Published reports of the effect of rs670 on a shifted expression of *APOA1 in vitro* are also contradictory. Some studies observed a link between *APOA1* transcription and the A allele (*Tuteja et al., 1992; Angotti et al., 1994*) whereas other studies did not confirm these observations (*Smith et al., 1992; Barre et al., 1994*). *Angotti et al.* (1994) stated that elevated APOA1 serum levels are caused by a decreased binding affinity of an unidentified nuclear factor due to the A allele. Three transcriptional factors are described to bind in the region from -89 to -64 of *APOA1*. The hepatic nuclear factor 4 (HNF4) promotes whereas apolipoprotein regulatory protein 1 (ARP1) and Ear3/COUP-TF repress transcription of *APOA1* (*Lai et al., 2005*). These findings point to a complex regulation of the tandemly organized *APOA1-APOC3-APOA4-APOA5* gene cluster that is regulated in a multifactorial mode including the interaction of multiple genes and multiple environmental factors (*Papazafiri et al., 1991; Danek et al., 1998*).

Collectively, these studies support a change in APOA1 expression due to rs670. Accordingly, the most likely molecular interpretation of my epidemiological findings is a direct influence of reduced *APOA1* expression levels due to the rs670 G allele on parasite development in the liver. Impairment of *Plasmodium* liver-stage development is akin to the phenotype seen in wild-type mice infected with *uis4(-)*-sporozoites (*Mueller et al., 2005, 2*). It was suggested that the fast development in the liver requires external lipids from the host *(Mikolajczak et al., 2007). Rodrigues et al.* (*2008*) showed that the depletion of all sources of cholesterol reduces infection, including liver cell stage development. An impaired development could result in a delay of release and/ or a reduced number of merosomes/ merozoites out of the liver into the blood.

Both scenarios would explain the observed extended time to detectable blood stages in patients in the Mild Malaria Cohort. This is consistent with the weak effect of rs670 on severe malaria compared to HbAS. After the release of the parasites from the liver rs670 would have no influence on their growth in the blood in contrast to HbAS that act primarily on this stage (*Williams et al., 2005*). The higher risk of severe malaria due to the A allele could be caused by a accelerated development or high numbers of merosomes/ merozoites released from the liver resulting in a fast establishment of a detectable parasitaemia in the child.

Due to the complex regulation of the *APOA1-APOC3-APOA4-APOA5* gene cluster and previous contradictory studies a direct measurement of the SNP on liver stage development *in vitro* was not reasonable. Supposing an impact of rs670 on APOA1 expression I tested the influence of APOA1 on *P. berghei* hepatocyte stages in an overexpression model. In the human Huh7 and murine Hepa 1.6 hepatoma cell lines, overexpressing rat APOA1, the *P. berghei* liver cell stages were significantly larger and contained more 18s copies compared to the WT at 48h post invasion. The total number of hepatocyte stages was similar between both cell lines at this time point. A decrease of established liver cell stages from 24h to 48h in Huh7 WT indicates an increased invasion in these cells but suboptimal growing conditions for *P. berghei*. This pattern was observed before (*Frischknecht F, personal communication*). A higher 18s copy number suggests an increased number of merozoites per liver cell stage or an elevated metabolism which could result in healthier merozoites. The latter might be more successful in invading erythrocytes.
However, in a second experiment using mock-transfected hepatoma cell lines, which were not expressing additional APOA1, instead of WT these results could not be reproduced. One reason could be the observed impaired growth of the mock-transfected hepatocytes or a change in the composition of the growth media used in both experiments containing different quantities of external APOA1 concentrations (*Pan-Biotech, personal communication*).

On a more general note, the use of promiscuous and fast growing *P. berghei* might be not representative with regard to human infections with *P. falciparum*. Reassuringly, a knock-down screen using APOA1 siRNA showed an impaired development of *P. berghei in vitro (Mueller AK, unpublished)*. Furthermore, *P. berghei* seems to have a delayed time to detectable blood stages in APOA1 knock-out mice *(Mueller AK, personal communication*). Furthermore, in knock-out and overexpression experiments targeting the scavenger receptor BI (SR-BI) show comparable results to my first APOA1 overexpression experiment: In WT the liver cell stages were significantly larger compared to SB-RI $^{-/-}$ mutants at 48h post invasion possibly due to lack of cholesterol. *P. yoelli* invasion in hepatocytes requires SR-BI as co-receptor to CD81 (*Yalaoui et al., 2008*). Besides the influence of SR-BI on hepatocyte permissiveness it is implicated in HDL uptake in hepatocytes (*Rhainds and Brissette, 2004*).

The frequency of the beneficial *HBB* S allele in malaria-endemic areas is raised by positive selection pressure (*Kwiatkowski, 2005; Williams et al., 2006*), but the deleterious effects of HbSS on the health preclude its fixation in any given population. Despite a higher chance of severe malaria and thus a negative selective pressure the A allele frequency of 0.2 in my sample was more abundant than in the Nigerian population (0.1) *(Kamboh et al, 1999)* or in different European populations (0.1-0.19) (*Siggurdson et al, 1992; Xu et al, 1994; Juo et al., 1999; Helbecque et al., 2008)*. However, its frequency is similar to South-East Asians (~0.2) (*Heng et al, 2001; Chhabra et al., 2005*; *Padmaja et al., 2009)* and lower compared to the Chinese population (~0.3) (*Li et al., 2008*). This pattern probably reflects the historical and ongoing immigration from East Asia to the East African Coast and mixing with the local population (*Herzig and Pascale, 2006*). The corresponding reduction of allele A frequencies further westwards could be caused by negative *P. falciparum* selection pressure.
Interestingly, the A allele seems to be abundant in areas that are historically endemic for *P. vivax*. This non-lethal *Plasmodium* species causes morbidity and fewer severe complications (*Breman, 2001, Mendis et al., 2001*). It is hypothesized that in *P. falciparum* and *P. vivax* endemic areas infections with *P. vivax* that occur earlier in life could have a protective effect against later *P. falciparum* malaria (*Maitland et al., 1996; Williams et al., 1996, Smith et al., 2001*). Accordingly a permissive effect of rs670 A on *P. vivax* would result in a protective benefit in *P. falciparum* infection.

As an alternative hypothesis I propose to also consider a similar beneficial effect in *P. falciparum* high-transmission areas. *Gupta et al.* (1999) observed low incidences of severe malaria in areas where transmission is highest. It is hypothesized that infections in early infancy, when the children are protected by maternal antibodies, will quickly develop adaptive immune responses, which protect against severe malaria later in life. A considerable degree of immunity against the parasite is already achieved after 1 or 2 episodes of malaria (*Gupta et al., 1999*). An accelerated acquisition of infections, as seen in rs670 AA homozygotes, during infancy would promote a fast acquisition of immunity during a time when the children are still passively protected by maternal antibodies. Contrary to high-transmission areas this effect would diminish in low-transmission areas.

APOA1 could also influence directly the immune response and acquisition of immunity. It may have a role in limiting inflammation (*Bresnihan et al., 2004)* and shows a negative correlation with TNF-α (*Hyka et al., 2001*). Furthermore, T-cells that are stimulated by APOA1-pretreated dendritic cells (DC) produced low levels of IFN-γ and APOA1 seems to inhibit cross talk between DC and natural killer cells (NK) *(Kim et al., 2005).* The latter, component of the innate immune system, induce apoptosis in infected cells whereas IFN-γ plays an important role in killing intrahepatocytic parasites by stimulating the production of nitric oxides by the infected hepatocytes (*Oliveira et al., 2008*). TNF-α is involved in regulation of immune cells and apoptotic cell death. Therefore, a reduction of IFN-γ and TNF-α caused by high level of APOA1 could increase the survival of *Plasmodium* liver cell stages.

The genomic region around *APOA1* is more conserved as indicated by higher LD between the SNPs compared to *FABP1* and *SNAPAP*. As described before, rs670 was in LD with rs5069 that is thought to be relevant to APOA1 and HDL levels (*Kamboh et al., 1996; Wang et al; 1996)* but rs5069 was not associated to malaria incidence in my analyses. Because rs670 was not in LD with other associated SNPs close to *APOA1* future studies should analyse the whole *APOA1-APOC3-APOA4-APOA5* gene cluster to identify all markers in LD with rs670 that might explain the inconsistent effect of rs670 in different ethnic groups.

Interestingly, the association of rs670 with other diseases, especially coronary artery disease (CAD), indicate that the allele frequencies may be influenced by multiple factors (*Chhabra et al., 2005; Vollbach et al., 2005; Dixit et al., 2007*).

While infectious pathogens, such as the malaria parasite, are the most likely driving forces for allelic frequencies in different populations, no signatures of selection on this locus or *HBB* in a population-genetic-based analysis of publicly accessible HapMap Phase II data was detected using Haplotter (*Voight et al., 2006*). Using different statistical approaches (iHs, Fay and Wu's H, Tajima's D and $F_{ST}$) I hoped to gain information about the evolutionary time point of a possible selectional event (*Figure 5; Sabeti et al., 2006*). None showed a significant signature of selection in *APOA1, FABP1, SNAPAP* or *HBB*. Only *Tajima's D* in the CEU dataset indicated conserved region around *APOA1* probably caused by its sensitivity to population

structures. These findings may be explained by mild selection pressure, reflected by low-odds ratio of rs670 against severe malaria. Population history could also be a factor: Mildly deleterious alleles can rise to moderate frequency in populations that have undergone recent expansion or neutral alleles might confer susceptibility to disease due to changes in living conditions like the burden of malaria accompanying human settlement (*Altshuler et al., 2008*). Additionally, rs670 could be prevented from fixation because of a balanced selection similar to HbAS (*Williams et al. 2005, 1*).

Alternative reasons, which would also explain the failure to pick up signatures of positive selection at the *HBB* locus, include the small sample sizes, the low SNP density, or absence of the relevant SNPs in the dataset. Especially in 40kb around *APOA1* no SNP was genotyped in the YRI. Marker rs670 was missing in all dataset. This draws attention to the present limitations of the HapMap project and supporting the continued importance of candidate-gene association approaches. However, more studies will be required to confirm the described association of rs670 on *P. falciparum* malaria. Recent advances in sequencing technology and approaches, such as the 1000 Genomes Project, are promising new tools (*Jallow et al., 2009*).

Collectively, my data suggests that polymorphisms, which protect against natural malaria transmission, exist. In my thesis I focussed on host cell proteins that upon infection are specifically recruited by the developing liver stage parasite. I indeed detected a robust protective effect of the SNP rs670 in the 5' UTR of *APOA1* against both uncomplicated and severe malaria using two independent screening methods in malaria-endemic populations. Association of rs670 with the risk of *P. falciparum* re-infections and severe malaria emphasises the epidemiologically important role for protein-protein interactions at the pathogen/host interface. My candidate gene association gene approach can now be extended to host cell factors that play important roles in sporozoite invasion (*Silvie et al., 2003, Rodrigues et al., 2008, Yalaoui et al., 2008*). Most importantly, my study is the first description of host cell polymorphisms that act specifically on the clinically silent malaria parasite expansion phase, a concept that may be extended to other human pathogens. My results also highlight the importance of lipid metabolism in host/pathogen interactions. Furthermore, it might provide a plausible explanation for the observed variations in

apolipoprotein, lipid levels and cardiovascular risk between ethnic groups (*Kamboh et al., 1999; Mbalilaki et al., 2008; McQueen et al., 2008*). It was shown that Asian Indians have considerably higher prevalence of premature coronary artery disease (CAD) and standardized mortality rates for CAD compared with Europeans, Chinese and Malays (*Mohan et al., 2001*). African blacks or Black Americans have a high prevalence of hypertension in association with coronary heart disease (*Yanci, 2001; Marijon et al., 2007*). It is implicated that genetic traits, such as rs670, have a strong influence on lipid composition besides cultural, social and environmental factors (*Talmud et al., 1994, Shanker et al., 2008*). In contrast to *P. falciparum,* the selectional forces of late-onset diseases such as CAD are negligible. Therefore, genetic traits leading to cardiovascular diseases could be enriched in populations living in malaria-endemic regions due to a benefit from protection against *P. falciparum*.

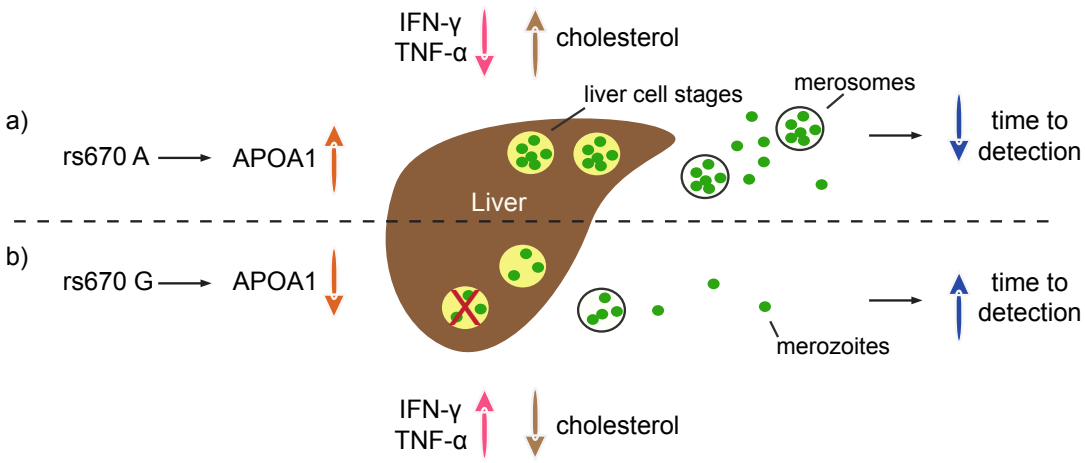



**Figure 27: Influence of rs670 on *Plasmodium* liver cell stages**
This hypothetical model illustrates the effect of rs670 on liver cell stages.
**a:** The rs670 A allele increases the amount of APOA1 decreasing IFN-γ and TNF-α that results in a weak immune response against liver cell stages. The high level of available cholesterol stimulates the development of the intra-hepatocytic parasites. Finally, a high number of merosomes and/or merozoites are released that rapidly establish a detectable blood infection. The optimal growth conditions could result in an accelerated maturation of liver cell stages and early release from the liver. **b:** The rs670 G allele causes low level of APOA1 that reduces the amount of cholesterol leading to an impaired development of intra-hepatocytic parasites. The level of IFN-γ and TNF-α are increased resulting in apoptosis of infected hepatocytes (X). As a result, low numbers of released merosomes and/or merozoites lead to a prolonged time to detectable asexual blood stages. The lack of cholesterol could result in delayed growth of liver cell stages and thus maturation to infective merozoites.

## References:


Ajees AA, Anantharamaiah GM, Mishra VK, Hussain MM, Murthy HM.
Crystal structure of human apolipoprotein A-I: insights into its protective effect against cardiovascular diseases.
Proc Natl Acad Sci U S A. 2006 Feb 14;103(7):2126-31.

Altshuler D, Daly MJ, Lander ES.
Genetic mapping in human disease.
Science. 2008 Nov 7;322(5903):881-8.

Andersen PK and Gill RD.
Cox's regression model for counting processes: A large sample study.
Annals of Statistics 1982, 10: 1100–1120.

Andrews KT, Lanzer M.
Maternal malaria: Plasmodium falciparum sequestration in the placenta.
Parasitol Res. 2002 Aug;88(8):715-23.

Angotti E, Mele E, Costanzo F, Avvedimento EV.
A polymorphism (G/ A transition) in the k78 position of the apolipoprotein A-I promoter increases transcription efficiency.
J Bio Chem 1994 ; 269 : 17 371–4.

Amino R, Thiberge S, Martin B, Celli S, Shorte S, Frischknecht F, Ménard R.
Quantitative imaging of Plasmodium transmission from mosquito to mammal.
Nat Med. 2006 Feb;12(2):220-4.

Baer K, Klotz C, Kappe SH, Schnieder T, Frevert U.
Release of hepatic Plasmodium yoelii merozoites into the pulmonary microvasculature.
PLoS Pathog. 2007 Nov;3(11):e171.

Baird JK, Masbar S, Basri H, Tirtokusumo S, Subianto B, Hoffman SL.
Age-dependent susceptibility to severe disease with primary exposure to Plasmodium falciparum.
J Infect Dis. 1998 Aug;178(2):592-5.

Ballou WR.
Obstacles to the development of a safe and effective attenuated pre-erythrocytic stage malaria vaccine.
Microbes Infect 2007; 9: 761–766.

Ballou WR.
The development of the RTS,S malaria vaccine candidate: challenges and lessons.
Parasite Immunol. 2009 Sep;31(9):492-500.

Bamshad M, Wooding SP
Signatures of natural selection in the human genome.
Nat Rev Genet. 2003 Feb;4(2):99-111.

Barre DE, Guerra R, Verstraete R, Wang Z, Grundy SM, Cohen JC.
Genetic analysis of a polymorphism in the human apolipoprotein A-I gene promoter: effect on plasma HDL-cholesterol levels,
J. Lipid Res. 35 1994 1292 – 1296.

Barrett JC, Fry B, Maller J, Daly MJ.
Haploview: analysis and visualization of LD and haplotype maps.
Bioinformatics. 2005 Jan 15;21(2):263-5.

Bejon P, Lusingu J, Olotu A, Leach A, Lievens M, Vekemans J, Mshamu S, Lang T, Gould J, Dubois MC, Demoitié MA, Stallaert JF, Vansadia P, Carter T, Njuguna P, Awuondo KO, Malabeja A, Abdul O, Gesase S, Mturi N, *et al.*
Efficacy of RTS,S/AS01E vaccine against malaria in children 5 to 17 months of age.
N Engl J Med. 2008 Dec 11;359(24):2521-32.

Ben-Avi L, Durst R, Shpitzen S, Leitersdorf E, Meiner V.
Apolipoprotein E genotyping: accurate, simple, high throughput method using ABI Prism SNaPshot Multiplex System.
J Alzheimers Dis. 2004 Oct;6(5):497-501

Blair S, Carmona J, Correa A.
Malaria in children: links between nutrition and immunity
Rev Panam Salud Publica. 2002 Jan;11(1):5-14.

Breman, J. G
"The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden."
Am J Trop Med Hyg. 2001, 64 1-11.

Bresnihan B, Gogarty M, FitzGerald O, Dayer JM, Burger D.
Apolipoprotein A-I infiltration in rheumatoid arthritis synovial tissue: a control mechanism of cytokine production?
Arthritis Res Ther. 2004;6(6):R563-6.

Brookes, AJ.
The essence of SNPs
Gene. 1999, 234, 177–186

Burgner D, Jamieson SE, Blackwell JM.
Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better?
Lancet Infect Dis. 2006 Oct;6(10):653-63.

Carter, R. and K. N. Mendis
Evolutionary and historical aspects of the burden of malaria.
Clin Microbiol Rev. 2002 15(4): 564-594.

Chhabra S, Narang R, Lakshmy R, Das N.
APOA1-75 G to A substitution associated with severe forms of CAD, lower levels of HDL and apoA-I among northern Indians.
Dis Markers. 2005;21(4):169-74.

Cohen S, Butcher GA.
Properties of protective malarial antibody.
Nature. 1970 Feb 21;225(5234):732-4.

Cornfield, J.
A statistical problem arising from retrospective studies.
Proc. Third Berkeley Symp. on Math. Statist. and Prob., 1956, Vol. 4 (Univ. of Calif. Press, 1956), 135-148.

Cowman AF, Crabb BS.
Invasion of red blood cells by malaria parasites.
Cell. 2006 Feb 24;124(4):755-66.

Cox DR.
Regression models and life-tables.
J. Roy. Statist. Soc. 1972; B. 34:187-220.

de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D.
Efficiency and power in genetic association studies.
Nat Genet. 2005 Nov;37(11):1217-23.

de Bakker PI, Burtt NP, Graham RR, Guiducci C, Yelensky R, Drake JA, Bersaglieri T, Penney KL, Butler J, Young S, Onofrio RC, Lyon HN, Stram DO, et al.
Transferability of tag SNPs in genetic association studies in multiple populations.
Nat Genet. 2006 Nov;38(11):1298-303.

de França E, Alves JG, Hutz MH.
APOA1/C3/A4 gene cluster variability and lipid levels in Brazilian children.
Braz J Med Biol Res. 2005 Apr;38(4):535-41.

Dixit M, Choudhuri G, Saxena R, Mittal B.
Association of apolipoprotein A1-C3 gene cluster polymorphisms with gallstone disease
Can J Gastroenterol. 2007 Sep;21(9):569-75.

Dondorp AM, Nosten F, Yi P, Das D, Phyo AP, Tarning J, Lwin KM, Ariey F, Hanpithakpong W, Lee SJ, Ringwald P, Silamut K, Imwong M, et al.
Artemisinin resistance in Plasmodium falciparum malaria.
N Engl J Med. 2009 Jul 30;361(5):455-67.

Doolan DL, Hoffman SL.
The complexity of protective immunity against liver-stage malaria.
J Immunol. 2000 Aug 1;165(3):1453-62.

Doolan DL and Martinez-Alier N.
Immune response to pre-erythrocytic stages of malaria parasites.
Curr Mol Med. 2006 6(2): 169-185.

Dreux M, Dao Thi VL, Fresquet J, Guérin M, Julia Z, Verney G, Durantel D, Zoulim F, Lavillette D, Cosset FL, Bartosch B.
Receptor complementation and mutagenesis reveal SR-BI as an essential HCV entry factor and functionally imply its intra- and extra-cellular domains.
PLoS Pathog. 2009 Feb;5(2):e1000310.

Emahazion T, Feuk L, Jobs M, Sawyer SL, Fredman D, St Clair D, Prince JA, Brookes AJ.
SNP association studies in Alzheimer's disease highlight problems for complex disease analysis.
Trends Genet. 2001 Jul;17(7):407-13.

Faucher JF, Ngou-Milama E, Missinou MA, Ngomo R, Kombila M, Kremsner P.
The impact of malaria on common lipid parameters.
*Parasitol Res* 2002, 88:1040-1043.30.

Fay J.C., Wu C.I.
Hitchhiking under positive Darwinian selection.
Genetics. 2006; 155(3):1405-13

Fisher E, Weikert C, Klapper M, Lindner I, Möhlig M, Spranger J, Boeing H, Schrezenmeir J, Döring F.
L-FABP T94A is associated with fasting triglycerides and LDL-cholesterol in women.
Mol Genet Metab. 2007 Jul;91(3):278-84.

Frevert U.
Sneaking in through the back entrance: The biology of malaria liver stages.
Trends Parasitol. 2004, 20: 417-424.

Fried M, Nosten F, Brockman A, Brabin BJ, Duffy PE.
Maternal antibodies block malaria.
Nature. 1998 Oct 29;395(6705):851-2.

Frome EL, Checkoway H.
Epidemiologic programs for computers and calculators. Use of Poisson regression models in estimating incidence rates and ratios.
Am J Epidemiol. 1985 Feb;121(2):309-23.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D.
The structure of haplotype blocks in the human genome.
Science. 2002 Jun 21;296(5576):2225-9.

Gallup, J. L. and J. D. Sachs.
The economic burden of malaria.
Am J Trop Med Hyg. 2001, 64(1-2 Suppl): 85-96.

Gilles, H. and D. Warrel.
Essential Malariology 4th Edition
Arnold, Hodder Headline Group. 2002, London

Goldstein DB.
Islands of linkage disequilibrium.
Nat Genet. 2001 Oct;29(2):109-11.

Gomes I, Collins A, Lonjou C, Thomas NS, Wilkinson J, Watson M, Morton N.
Hardy-Weinberg quality control.
Ann Hum Genet. 1999 Nov;63(Pt 6):535-8.

Gordon D, Finch SJ.
Factors affecting statistical power in the detection of genetic association.
J Clin Invest. 2005 Jun;115(6):1408-18.

Gray IC, Campbell DA, Spurr NK.
Single nucleotide polymorphisms as tools in human genetics.
Hum Mol Genet. 2000 Oct;9(16):2403-8.

Greenwood BM, Bojang K, Whitty CJ, Targett GA.
Malaria.
Lancet 2005, 365 (9469): 1487–1498.

Grellier P, Rigomier D, Clavey V, Fruchart JC, Schrevel J.
Lipid traffic between high density lipoproteins and Plasmodium falciparum-infected red blood cells.
J Cell Biol. 1991 January 2; 112(2): 267–277.

Gupta S, Snow RW, Donnelly CA, Marsh K, Newbold C.
Immunity to non-cerebral severe malaria is acquired after one or two infections.
Nat Med. 1999 Mar;5(3):340-3.

Hall, T.A.
BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.
Nucl. Acids: Symp. Ser. 41:95-98.

Harrington JM, Howell S, Hajduk SL.
Membrane permeabilization by trypanosome lytic factor, a cytolytic human high density lipoprotein.
J Biol Chem. 2009 May 15;284(20):13505-12.

Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW.
The global distribution and population at risk of malaria: past, present, and future. Lancet Infect Dis. 2004 Jun;4(6):327-36.

Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, Kabaria CW, Manh BH, Elyazar IR, Brooker S, Smith DL, Moyeed RA, Snow RW.
A world malaria map: Plasmodium falciparum endemicity in 2007.
PLoS Med. 2009 Mar 24;6(3)

Helbecque N, Codron V, Cottel D, Amouyel P.
An apolipoprotein A-I gene promoter polymorphism associated with cognitive decline, but not with Alzheimer's disease.
Dement Geriatr Cogn Disord. 2008;25(2):97-102.

Heng, Ck; Low, PS; Saha, N. Variations in the promoter region of the apolipoprotein A-1 gene influence plasma lipoprotein(a) levels in Asian Indian neonates from Singapore.
Pediatr Res. 2001;49:514–8.

Herzig, Pascale
South Asians in Kenya: Gender, Generation and Changing Identities in Diaspora.
LIT Verlag, Münster 2006

Hoffman SL, Doolan DL.
Malaria vaccines-targeting infected hepatocytes.
Nat Med. 2000 Nov;6(11):1218-9.

Holsinger KE, Weir BS.
Genetics in geographically structured populations: defining, estimating and interpreting F(ST).
Nat Rev Genet. 2009 Sep;10(9):639-50.

Hunt RA, Edris W, Chanda PK, Nieuwenhuijsen B, Young KH
Snapin interacts with the N-terminus of regulator of G protein signaling 7.
Biochem Biophys Res Commun. 2003, 303 (2): 594-9.

Hyka N, Dayer JM, Modoux C, Kohno T, Edwards CK, Roux-Lombard P, Burger D,
Apolipoprotein A-I inhibits the production of interleukin-1beta and tumor necrosis factor-alpha by blocking contact-mediated activation of monocytes by T lymphocytes.
Blood. 2001, 97 2381–2389.

Iardi JM, Mochida S, Sheng ZH.
Snapin: a SNARE-associated protein implicated in synaptic transmission".
Nat Neurosci. 1999, 2 (2): 119-24

Jallow M, Teo YY, Small KS, Rockett KA, Deloukas P, Clark TG, Kivinen K, Bojang KA, Conway DJ, Pinder M, Sirugo G, Sisay-Joof F, Usen S, Auburn S, Bumpstead SJ, et al.
Genome-wide and fine-resolution association analysis of malaria in West Africa.
Nat Genet. 2009. 41, 657–665

Jeenah M, Kessling A, Miller N, Humphries S.
G to A substitution in the promoter region of the apolipoprotein AI gene is associated with elevated serum apolipoprotein AI and high density lipoprotein cholesterol concentrations.
Mol Biol Med. 1990 Jun;7(3):233-41

Jepson A, Sisay-Joof F, Banya W, Hassan-King M, Frodsham A, Bennett S, Hill AV, Whittle H
Genetic linkage of mild malaria to the major histocompatibility complex in Gambian children: study of affected sibling pairs.
BMJ. 1997 Jul 12;315(7100):96-7.

Jobe O, Lumsden J, Mueller A., Williams J., Silva-Rivera H, Kapp, SH, Schwenk, RJ, Matuschewski, K, Krzych U.
Genetically Attenuated Plasmodium berghei Liver Stages Induce Sterile Protracted Protection that Is Mediated by Major Histocompatibility Complex Class I-Dependent Interferon- gamma -Producing CD8+T Cells.
J Infect Dis. 2007, 196: 599-607.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA.
Haplotype tagging for the identification of common disease genes.
Nat Genet. 2001 Oct;29(2):233-7

Juo SH, Wyszynski DF, Beaty TH, Huang HY, Bailey-Wilson JE.
Mild association between the A/G polymorphism in the promoter of the apolipoprotein A-I gene and apolipoprotein A-I levels: a meta-analysis.
Am J Med Genet. 1999 Jan 29;82(3):235-41

Kaderali L, Deshpande A, Nolan JP, White PS.
Primer-design for multiplexed genotyping.
Nucleic Acids Res. 2003 Mar 15;31(6):1796-802

Kamboh MI, Aston CE, Nestlerode CM, McAllister AE, Hamman RF:
Haplotype analysis of two APOA1/ *Msp* I polymorphisms in relation to plasma levels of apoA-I and HDL cholesterol.
Atherosclerosis 1996; 127: 255–262.

Kamboh MI, Bunker CH, Aston CE, Nestlerode CS, McAllister AE, Ukoli FA
Genetic association of five apolipoprotein polymorphisms with serum lipoprotein-lipid levels in African blacks.
Genet Epidemiol. 1999;16(2):205-22.

Kim KD, Lim HY, Lee HG, Yoon DY, Choe YK, Choi I, Kim YS, Yang Y, Lim JS.
Apolipoprotein A-I induces IL-10 and PGE2 production in human monocytes and inhibits dendritic cell differentiation and maturation.
Biochem Biophys Res Commun. 2005 Dec 16;338(2):1126-36.

Kimura M.
The neutral theory of molecular evolution.
Cambridge 1983.

Klein J, Takahata N.
Where do we come from? The molecular evidence for human descent.
Springer, New York,2002

Kung-Jong L; Chii-Dean L.
A revisit on comparing the asymptotic interval estimators of odds ratio in a single 2x2 table.
Biometrical Journ. 2003, 45, p226-237

Kwiatkowski DP.
How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria.
Am J Hum Genet. 2005 August; 77(2): 171–192

Kwon JM and Goate AM.
The Candidate Gene Approach.
Alcohol Research & Health, 2000 Vol. 24,

Labaied M, Harupa A, Dumpit RF, Coppens I, Mikolajczak SA, Kappe SH.
Plasmodium yoelii sporozoites with simultaneous deletion of P52 and P36 are completely attenuated and confer sterile immunity against infection.
Infect Immun. 2007 Aug;75(8):3758-68.

Lai CQ, Parnell LD, Ordovas JM.
The APOA1/C3/A4/A5 gene cluster, lipid metabolism and cardiovascular disease risk.
Curr Opin Lipidol. 2005 Apr;16(2):153-66.

Lander ES, Schork NJ.
Genetic dissection of complex traits.
Science. 1994 Sep 30;265(5181):2037-48.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al.
Initial sequencing and analysis of the human genome.
Nature. 2001 Feb 15;409(6822):860-921.

Langhorne J, Ndungu FM, Sponaas AM, Marsh K.
Immunity to malaria: more questions than answers.
Nat Immunol. 2008 Jul;9(7):725-32

Lapouméroulie C, Dunda O, Ducrocq R, Trabuchet G, Mony-Lobé M, Bodo JM, Carnevale P, Labie D, Elion J, Krishnamoorthy R.
A novel sickle cell mutation of yet another origin in Africa: the Cameroon type.
Hum Genet. 1992 May;89(3):333-7.

Li SS, Cheng JJ, Zhao LP.
Empirical vs Bayesian approach for estimating haplotypes from genotypes of unrelated individuals.
BMC Genet. 2007 Jan 29;8:2

Li Y, Yin R, Zhou Y, Deng Y, Yang D, Pan S, Lin W.
Associations of the apolipoprotein A-I gene polymorphism and serum lipid levels in the Guangxi Hei Yi Zhuang and Han populations.
Int J Mol Med. 2008 Jun;21(6):753-64.

Lin D, Wei L (1989)
The robust inference for the Cox Proportional Hazards Model.
J Am Stat Assoc 84: 1074–1079.

Lewontin RC.
The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models.
Genetics. 1964 Jan;49(1):49-67.

Lewontin RC, Krakauer J.
Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms.
Genetics. 1973 May;74(1):175-95.

Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN.
Heritability of malaria in Africa.
PLoS Med. 2005 Dec;2(12):e340.

Mackintosh CL, Beeson JG, Marsh K.
Clinical features and pathogenesis of severe malaria.
Trends Parasitol. 2004 Dec;20(12):597-603.

Madsen BE, Villesen P, Wiuf C.
A periodic pattern of SNPs in the human genome.
Genome Res. 2007 Oct;17(10):1414-9.

Maitland K, Williams TN, Bennett S, Newbold CI, Peto TE, Viji J, Timothy R, Clegg JB, Weatherall DJ, Bowden DK.
The interaction between Plasmodium falciparum and P. vivax in children on Espiritu Santo island, Vanuatu.
Trans R Soc Trop Med Hyg. 1996 Nov-Dec;90(6):614-20.

Marchini J, Cutler D, Patterson N, Stephens M, Eskin E, Halperin E, Lin S, Qin ZS, Munro HM, Abecasis GR, Donnelly P; International HapMap Consortium.
A comparison of phasing algorithms for trios and unrelated individuals.
Am J Hum Genet. 2006 Mar;78(3):437-50.

Marijon E, Trinquart L, Jani D, Jourdier H, Garbarz E, Ferreira B.
Coronary heart disease and associated risk factors in sub-Saharan Africans
J of Hum Hypertens (2007) 21, 411–414.

Marsh K, Kinyanjui S.
Immune effector mechanisms in malaria.
Parasite Immunol. 2006 Jan-Feb;28(1-2):51-60.

Masson D, Koseki M, Ishibashi M, Larson CJ, Miller SG, D King B, Tall AR.
Increased HDL Cholesterol and ApoA-I in Humans and Mice Treated With a Novel SR-BI Inhibitor
Arterioscler Thromb Vasc Biol, December 1, 2009; 29(12): 2054 - 2060.

Matuschewski K, Ross J, Brown SM, Kaiser K, Nussenzweig V, Kappe SH.
Infectivity-associated changes in the transcriptional repertoire of the malaria parasite sporozoite stage.
J Biol Chem. 2002 Nov 1;277(44):41948-53.

Mbalilaki JA, Masesa Z, Strømme SB, Høstmark AT, Sundquist J, Wändell P, Rosengren A, Hellenius ML.
Daily energy expenditure and cardiovascular risk in Masai, rural and urban Bantu Tanzanians.
Br J Sports Med. 2008 Jun 3.

Mbogo CN, Snow RW, Khamala CP, Kabiru EW, Ouma JH, Githure JI, Marsh K, Beier JC.
Relationships between Plasmodium falciparum transmission by vector populations and the incidence of severe disease at nine sites on the Kenyan coast.
Am J Trop Med Hyg. 1995 Mar;52(3):201-6.

McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett J, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM.
Common deletion polymorphisms in the human genome.
Nature Genetics 38: 86-92, 2006.

McQueen MJ, Hawken S, Wang X, Ounpuu S, Sniderman A, Probstfield J, Steyn K, Sanderson JE, Hasani M, Volkova E, Kazmi K, Yusuf S
Lipids, lipoproteins, and apolipoproteins as risk markers of myocardial infarction in 52 countries (the INTERHEART study): a case-control study.
Lancet. 2008 Jul 19;372(9634):224-33.

Mehal WZ, Azzaroli F, Crispe IN.
Immunology of the healthy liver: Old questions and new insights.
Gastroenterology 2001, 120:250-260.

Menashe I, Rosenberg PS, Chen BE.
PGA: power calculator for case-control genetic association analyses.
BMC Genet. 2008 May 13;9:36.

Mendis K, Sina BJ, Marchesini P, Carter R.
The neglected burden of Plasmodium vivax malaria.
Am J Trop Med Hyg. 2001 Jan-Feb;64

Mikolajczak SA, Jacobs-Lorena V, MacKellar DC, Camargo N, Kappe SH.
L-FABP is a critical host factor for successful malaria liver stage development.
Int J Parasitol. 2007 Apr;37(5):483-9.

Mitamura T, Palacpac NM.
Lipid metabolism in Plasmodium falciparum-infected erythrocytes: possible new targets for malaria chemotherapy.
Microbes Infect. 2003 May;5(6):545-52.

Mohan V, Deepa R, Rani SS, Premalatha G; Chennai Urban Population Study.
Prevalence of coronary artery disease and its relationship to lipids in a selected population in South India: The Chennai Urban Population Study (CUPS No. 5).
J Am Coll Cardiol. 2001 Sep;38(3):682-7.

Morenilla-Palao C, Planells-Cases R, García-Sanz N, Ferrer-Montiel A.
Regulated exocytosis contributes to protein kinase C potentiation of vanilloid receptor activity.
J Biol Chem. 2004 Jun 11;279(24):25665-72.

Mueller AK, Camargo N, Kaiser K, Andorfer C, Frevert U, Matuschewski K, Kappe SH.
Plasmodium liver stage developmental arrest by depletion of a protein at the parasite-host interface.
Proc Natl Acad Sci U S A. 2005 Feb 22;102(8):3022-7. Epub 2005 Feb 7.

Mueller AK, Labaied M, Kappe SH, Matuschewski K.
Genetically modified Plasmodium parasites as a protective experimental malaria vaccine.
Nature. 2005 Jan 13;433(7022):164-7.

Needham EW, Mattu RK, Rees A, Stocks J, Galton DJ.
A polymorphism in the human apolipoprotein AI promoter region: a study in hypertriglyceridaemic patients.
Hum Hered. 1994 Mar-Apr;44(2):94-9.

Nordborg M, Tavaré S.
Linkage disequilibrium: what history has to tell us.
Trends Genet. 2002 Feb;18(2):83-90.

Nussenzweig RS, Vanderberg, J, Most H. Orton, C.
Protective immunity produced by the injection of x-irradiated sporozoites of plasmodium berghei.
Nature. 1967,216: 160-162.

Oliveira GA, Kumar KA, Calvo-Calle JM, Othoro C, Altszuler, Nussenzweig, V, Nardin EH.
Class II-Restricted Protective Immunity Induced by Malaria Sporozoites
Infect Immun. 2008 March; 76(3): 1200–1206.

Padmaja N, Ravindra Kumar M, Adithan C
Association of polymorphisms in apolipoprotein A1 and apolipoprotein B genes with lipid profile in Tamilian population.
Indian Heart J. 2009 Jan-Feb;61(1):51-4.

Papazafiri P, Ogami K, Ramji D, Nicosia A, Monaci P, Cladaras C, Zannis V.
Promoter elements and factors involved in hepatic transcription of the human Apo A-I gene positive and negative regulators bind to overlapping sites.
J. Biol Chem. 1991. 2661 5790 – 5797

Pasvol G, Weatherall DJ, Wilson RJ.
Cellular mechanism for the protective effect of haemoglobin S against P. falciparum malaria.
Nature. 1978. 274, 701-703.

Pati N, Schowinsky V, Kokanovic O, Magnuson V, Ghosh S.
A comparison between SNaPshot, pyrosequencing, and biplex invader SNP genotyping methods: accuracy, cost, and throughput.
J Biochem Biophys Methods. 2004 Jul 30;60(1):1-12.

Pattanapanyasat K, Yongvanitchit K, Tongtawe P, Tachavanich K, Wanachiwanawin W, Fucharoen S, Walsh DS
Impairment of Plasmodium falciparum growth in thalassemic red blood cells: further evidence by using biotin labeling and flow cytometry.
Blood 1999, 93:3116–3119

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC.
PLINK: a tool set for whole-genome association and population-based linkage analyses.
Am J Hum Genet. 2007 Sep 81(3):559-75.

Purcell LA, Wong KA, Yanow SK, Lee M, Spithill TW, Rodriguez A.
Chemically attenuated Plasmodium sporozoites induce specific immune responses, sterile immunity and cross-protection against heterologous challenge.
Vaccine. 2008 Sep 8;26(38):4880-4

Ralph SA, van Dooren GG, Waller RF, Crawford MJ, Fraunholz MJ, Foth BJ, Tonkin CJ, Roos DS, McFadden GI.
Tropical infectious diseases: metabolic maps and functions of the Plasmodium falciparum apicoplast.
Nat Rev Microbiol. 2004 Mar;2(3):203-16.

Rasti, N., M. Wahlgren and Q. Chen.
Molecular aspects of malaria pathogenesis.
FEMS Immunol Med Microbiol. 2004, 41(1): 9-26.

Rhainds, D., and Brissette, L.
The role of scavenger receptor class B type I (SR-BI) in lipid trafficking. defining the rules for lipid traders.
Int. J. Biochem. Cell Biol. 2004; 36, 39–77.

Risch N, Merikangas K.
The future of genetic studies of complex human diseases.
Science. 1996 Sep 13;273 (5281):1516-7.

Roberts DJ, Williams TN.
Haemoglobinopathies and resistance to malaria.
Redox Rep. 2003;8(5):304-10.

Rodrigues CD, Hannus M, Prudêncio M, Martin C, Gonçalves LA, Portugal S, Epiphanio S, Akinc A, Hadwiger P, Jahn-Hofmann K, Röhl I, *et al.*
Host scavenger receptor SR-BI plays a dual role in the establishment of malaria parasite liver infection.
Cell Host Microbe. 2008 Sep 11;4(3):271-82.

Rozen, S. and Skaletsky, H.J.
Primer3 on the WWW for general users and for biologist programmers.
Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, 2000 NJ, pp 365-386

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES.
Detecting recent positive selection in the human genome from haplotype structure.
Nature. 2002 Oct 24;419(6909):832-7.

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES.
Positive natural selection in the human lineage.
Science. 2006 Jun 16;312 (5780):1614-20.

Sanger F, Nicklen S, Coulson AR.,
DNA sequencing with chain-terminating inhibitors.
Proc Natl Acad Sci. 1977 Dec;74(12):5463-7

Schofield L, Grau GE.
Immunological processes in malaria pathogenesis.
Nat Rev Immunol. 2005 Sep;5(9):722-35.

Schoenbach J, Rosamond W.
Understanding the Fundamentals of Epidemiology
Edition 2000, www.epidemiolog.net

Sehgal A, Bettiol S, Pypaert M, Wenk MR, Kaasch A, Blader IJ, Joiner KA, Coppens I: Peculiarities of host cholesterol transport to the unique intracellular vacuole containing *Toxoplasma*.
*Traffic* 2005, 6:1125-1141.

Shanker J, Perumal G, Rao VS, Khadrinarasimhiah NB, John S, Hebbagodi S, Mukherjee M, Kakkar VV.
Genetic studies on the APOA1-C3-A5 gene cluster in Asian Indians with premature coronary artery disease.
Lipids Health Dis. 2008 Sep 19;7:33.

Silvie O, Rubinstein E, Franetich JF, Prenant M, Belnoue E, Rénia L, Hannoun L, Eling W, Levy S, Boucheix C, Mazier D.
Hepatocyte CD81 is required for Plasmodium falciparum and Plasmodium yoelii sporozoite infectivity.
Nat Med. 2003 Jan;9(1):93-6.

Simpson DC, Kabyemela E, Muehlenbachs A, Ogata Y, Mutabingwa TK, Duffy PE, Fried M.
Plasma levels of apolipoprotein A1 in malaria-exposed primigravidae are associated with severe anemia.
PLoS One. 2010 Jan 21;5(1):e8822.

Smith T, Genton B, Baea K, Gibson N, Narara A, Alpers MP.
Prospective risk of morbidity in relation to malaria infection in an area of high endemicity of multiple species of Plasmodium.
Am J Trop Med Hyg. 2001 May-Jun;64(5-6):262-7.

Snounou G, White NJ.
The co-existence of Plasmodium: sidelights from falciparum and vivax malaria in Thailand.
Trends Parasitol. 2004 Jul;20(7):333-9.

Snow RW, Omumbo JA, Lowe B, Molyneux CS, Obiero JO, Palmer A, Weber MW, Pinder M, Nahlen B, Obonyo C, Newbold C, Gupta S, Marsh K.
Relation between severe malaria morbidity in children and level of Plasmodium falciparum transmission in Africa.
Lancet. 1997 Jun 7;349(9066):1650-4.

Snow RW, Howard SC, Mung'Ala-Odera V, English M, Molyneux CS, et al.
Paediatric survival and re-admission risks following hospitalization on the Kenyan coast.
Trop Med Int Health. 2000 5: 377–383.

Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI.
The global distribution of clinical episodes of Plasmodium falciparum malaria.
Nature. 2005. 434(7030): 214-217.

Soutar AK, Hawkins PN, Vigushin DM, Tennent GA, Booth SE, Hutton T, Nguyen O, Totty NF, Feest TG, Hsuan JJ, et al.
Apolipoprotein AI mutation Arg-60 causes autosomal dominant amyloidosis.
Proc Natl Acad Sci U S A. 1992 Aug 15;89(16):7389-93.

Spielmann T, Gardiner DL, Beck HP, Trenholme KR, Kemp DJ.
Organization of ETRAMPs and EXP-1 at the parasite-host cell interface of malaria parasites.
Mol Microbiol. 2006 Feb;59(3):779-94.

Starcevic M, Dell'Angelica EC.
Identification of snapin and three novel proteins (BLOS1, BLOS2, and BLOS3/reduced pigmentation) as subunits of biogenesis of lysosome-related organelles complex-1 (BLOC-1).
J Biol Chem. 2004 Jul 2;279(27):28393-401.

Stephens, M., Smith N, and Donnelly, P.
A new statistical method for haplotype reconstruction from population data.
American Journal of Human Genetics. 2001; 68, 978--989.

Stephens, M., and Donnelly, P.
A comparison of Bayesian methods for haplotype reconstruction from population genotype data.
American Journal of Human Genetics. 2003; 73:1162-1169.

Stephens M, Scheet P.
Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation.
Am J Hum Genet. 2005 Mar;76(3):449-62.

Stevenson MM and Riley EM.
Innate immunity to malaria.
Nat Rev Immunol. 2004, 4: 169-180.

Stocks J, Paul H, Galton D.
Haplotypes identified by DNA restriction-fragment-length polymorphisms in the A-1 C-III A-IV gene region and hypertriglyceridemia.
Am J Hum Genet. 1987 August; 41(2): 106–118.

Strobl W, Jabs HU, Hayde M, Holzinger T, Assmann G, Widhalm K.
Apolipoprotein A-I (Glu 198----Lys): a mutant of the major apolipoprotein of high-density lipoproteins occurring in a family with dyslipoproteinemia.
Pediatr Res. 1988 Aug;24(2):222-8.

Stuart MJ, Nagel RL
Sickle-cell disease.
Lancet 2004, 364(9442):1343-1360.

Sturm A, Amino R, van de Sand C, Regen T, Retzlaff S, Rennenberg A, KruegerA, Pollok JM, Menard R, Heussler VT.
Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids.
Science. 2006 Sep 1;313(5791):1287-90.

Szklo M.
Population-based cohort studies.
Epidemiol Rev. 1998;20(1):81-90.

Tajima, F.
Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.
Genetics 1989; 123:585-595

Talmud PJ, Ye S, Humphries SE.
Polymorphism in the promoter region of the apolipoprotein AI gene associated with differences in apolipoprotein AI levels: the European Atherosclerosis Research Study.
Genet Epidemiol. 1994;11(3):265-80.

The International HapMap Consortium *et al.*
A second generation human haplotype map of over 3.1 million SNPs.
Nature. 2007 Oct 18;449(7164):851-61.

Tishkoff SA, Williams SM.
Genetic analysis of African populations: human evolution and complex disease.
Nat Rev Genet. 2002 Aug;3(8):611-21.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, et al.
The genetic structure and history of Africans and African Americans.
Science. 2009 May 22;324(5930):1035-44.

Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J.
Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance.
Science 2001, 293:455-462.

Trigatti BL, Krieger M, Rigotti A.
Influence of the HDL receptor SR-BI on lipoprotein metabolism and atherosclerosis.
Arterioscler Thromb Vasc Biol. 2003 Oct 1;23(10):1732-8

Tuteja R, Tuteja N, Melo C, Casari G, Baralle FE.
Transcription efficiency of human apolipoprotein A-I promoter varies with naturally occurring A to G transition .
FEBS Lett 1992 ; 304 : 98–101.

Vaughan AM, Aly AS, Kappe SH.
Malaria parasite pre-erythrocytic stage infection: gliding and hiding.
Cell Host Microbe. 2008 Sep 11;4(3):209-18.

Vaughan AM, O'Neill MT, Tarun AS, Camargo N, Phuong TM, Aly AS, Cowman AF, Kappe SH.
Type II fatty acid synthesis is essential only for malaria parasite late liver stage development.
Cell Microbiol. 2009 Mar;11(3):506-20

Vallone PM, Butler JM.
AutoDimer: a screening tool for primer-dimer and hairpin structures.
Biotechniques. 2004 Aug;37(2):226-31.

Vanderberg JP, Frevert U.
Intravital microscopy demonstrating antibodymediated immobilization of Plasmodium berghei sporozoites injected into skin by mosquitoes.
Int J Parasitol. 2004, 34: 991-996.

Voight BF, Kudaravalli S, Wen X, Pritchard JK.
A map of recent positive selection in the human genome.
PLoS Biol. 2006 Mar;4(3):e72.

Voight BF, Kudaravalli S, Wen X, Pritchard JK.
A map of recent positive selection in the human genome.
PLoS Biol. 2006 Mar;4(3):e72.

Vollbach H, Heun R, Morris CM, Edwardson JA, McKeith IG, Jessen F, Schulz A, Maier W, Kölsch H.
APOA1 polymorphism influences risk for early-onset nonfamiliar AD.
Ann Neurol. 2005 Sep;58(3):436-41.

Wambua S, Mwangi TW, Kortok M, Uyoga SM, Macharia AW, Mwacharo JK, Weatherall DJ, Snow RW, Marsh K, Williams TN.
The effect of alpha+-thalassaemia on the incidence of malaria and other diseases in children living on the coast of Kenya.
PLoS Med. 2006 May;3(5)

Wang XL, Badenhop R, Humphrey KE, Wilcken DEL.
New *MspI* polymorphism at +83 of the human apolipoprotein AI gene: association with increased circulating high density lipoprotein cholesterol levels.
Genet Epidemiol., 1996, 13:1–10.

Wang N, Akey JM, Zhang K, Chakraborty R, Jin L.
Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation.
Am J Hum Genet. 2002 Nov;71(5):1227-34.

Weatherall DJ, Clegg JB.
Inherited haemoglobin disorders: an increasing global health problem.
Bull World Health Organ. 2001;79(8):704-12.

Weiss ST, Silverman EK, Palmer LJ.
Case-control association studies in pharmacogenetics.
Pharmacogenomics J. 2001;1(3):157-8.

Williams TN, Maitland K, Bennett S, Ganczakowski M, Peto TE, Newbold CI, Bowden DK, Weatherall DJ, Clegg JB.
High incidence of malaria in alpha-thalassaemic children.
Nature. 1996 Oct 10;383(6600):522-5

Williams TN, Weatherall DJ, Newbold CI.
The membrane characteristics of Plasmodium falciparum-infected and uninfected heterozygous alpha(0)thalassaemic erythrocytes.
Br J Haematol. 2002, 118:663–670

Williams TN (1), Mwangi TW, Roberts DJ, Alexander ND, Weatherall DJ, Wambua S, Kortok M, ,SnowRW, Marsh K.
An immune basis for malaria protection by the sickle cell trait.
PLoS Med. 2005 May;2(5)

Williams TN (2), Mwangi TW, Wambua S, Alexander ND, Kortok M, Snow RW, Marsh K.
Sickle cell trait and the risk of Plasmodium falciparum malaria and other childhood diseases.
J Infect Dis. 2005 Jul 1;192(1):178-86

Williams TN (3), Mwangi TW, Wambua S, Peto TE, Weatherall DJ, Gupta S, Recker M, Penman BS, Uyoga S, MachariaA, Mwacharo JK, Snow RW,MarshK.
Negative epistasis between the malaria-protective effects of alpha+-thalassemia and the sickle cell trait.
Nat Genet. 2005 Nov;37(11):1253-7

Williams TN.
Human red blood cell polymorphisms and malaria.
Curr Opin Microbiol. 2006 Aug;9(4):388-94.

Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P, Altshuler D.
Comparison of fine-scale recombination rates in humans and chimpanzees.
Science. 2005 Apr 1;308(5718):107-11.

Xu C-F, Angelico F, DelBen M, Humphries SE. Role of genetic variation at the apoAI-CIII-AIV gene cluster in determining plasma apoAI levels in boys and girls. Genet Epidemiol 1993 ; 10 : 113–22.

Xu, L., Turner, A., Little, J., Bleeker, E. R. & Meyers, D. A.
Positive results in association studies are associated
with departure from Hardy-Weinberg equilibrium: hint for genotyping error ?
Human Genetics 2002 11, 573–574.

Xu H, Wu X, Spitz MR, Shete S.
Comparison of haplotype inference methods using genotypic data from unrelated individuals.
Hum Hered. 2004;58(2):63-8.

Yalaoui S, Huby T, Franetich JF, Gego A, Rametti A, Moreau M, Collet X, Siau A, van Gemert GJ, Sauerwein RW, Luty AJ, Vaillant JC, Hannoun L, Chapman J, Mazier D, Froissard P.
Scavenger receptor BI boosts hepatocyte permissiveness to Plasmodium infection.
Cell Host Microbe. 2008 Sep 11;4(3):283-92.

Yanci CW.
Heart failure in blacks: etiologic and epidemiologic differences.
Curr Cardiol Rep. 2001 May;3(3):191-7.

Yui, T Aoyama, H Morishita, M Takahashi, Y Takatsu, and C Kawai
Serum prostacyclin stabilizing factor is identical to apolipoprotein A-I (Apo A-I). A novel function of Apo A-I.
J Clin Invest. 1988 September; 82(3): 803–807.

Ich erkläre hiermit, dass ich die vorgelegte Dissertation selbst verfasst und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe.

Ich erkläre hiermit, dass ich an keiner anderen Stelle ein Prüfungsverfahren beantragt bzw. die Dissertation in dieser oder anderer Form bereits anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe.

Heidelberg, 02.02.2010 _____

Andris Schulz