



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

**Genomweite Assoziationsstudie zu zerebraler Malaria und
Charakterisierung eines assoziierten Locus
auf Chromosom 2p22.3**

Dissertation

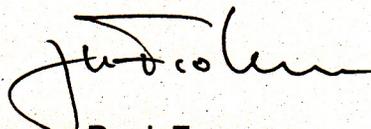
Zur Erlangung der Würde des Doktors der Naturwissenschaften
des Fachbereichs Biologie, der Fakultät für Mathematik, Informatik und
Naturwissenschaften, der Universität Hamburg

vorgelegt von
Ulrike Herzog
aus Wilhelmshaven

Hamburg, im August 2012

Genehmigt vom Fachbereich Biologie
der Fakultät für Mathematik, Informatik und Naturwissenschaften
an der Universität Hamburg
auf Antrag von Professor Dr. R. HORSTMANN
Weitere Gutachterin der Dissertation:
Professor Dr. I. BRUCHHAUS
Tag der Disputation: 25. Mai 2012

Hamburg, den 10. Mai 2012



Professor Dr. J. Fromm
Vorsitzender des Promotionsausschusses
Biologie

ABBILDUNGSVERZEICHNIS	V
TABELLENVERZEICHNIS	VI
ZUSAMMENFASSUNG	VII
SUMMARY	VIII
ABKÜRZUNGSVERZEICHNIS	IX
1 EINLEITUNG	1
1.1 Malaria	1
1.1.1 Malaria tropica.....	2
1.1.2 Zerebrale Malaria	2
1.1.2.1 Klinik und Diagnose	2
1.1.2.2 Pathogenese.....	3
1.2 Humangenetik der Malaria	5
1.3 Auswahl des Studiendesigns.....	7
1.4 Zielsetzung	10
2 PATIENTEN, MATERIAL UND METHODEN	12
2.1 Studiendesign und Studienpopulation	12
2.1.1 Studiendesign	12
2.1.2 Studienpopulation	12
2.2 Material	15
2.2.1 Geräte	15
2.2.2 Chemikalien und Verbrauchsmaterial	15
2.2.3 Puffer und Lösungen.....	16
2.2.4 Oligonukleotide	16
2.2.5 RNA aus humanem Gewebe.....	16
2.2.6 cDNA aus humanem Gewebe.....	16
2.2.7 cDNA-Klone	17
2.2.8 Vektoren.....	17
2.2.9 Reagenziensysteme.....	17
2.2.10 Kultivierung von <i>E. coli</i>	18
2.2.10.1 Bakterienstämme	18
2.2.10.2 Medien	18

2.2.10.3	Antibiotika	18
2.3	Methoden	19
2.3.1	Statistische Analysen	19
2.3.1.1	Qualitätskontrolle der genomweiten Assoziationsanalyse.....	19
2.3.1.1.1	Ausschluss von Studienteilnehmern	19
2.3.1.1.2	Ausschluss von SNPs.....	19
2.3.1.2	Assoziationsanalyse.....	20
2.3.1.3	Populationsstratifikation	22
2.3.1.4	Powerberechnung.....	22
2.3.1.5	Transmissions-Disequilibrium-Test	23
2.3.2	Molekularbiologische Methoden.....	23
2.3.2.1	Polymerase-Kettenreaktion (PCR).....	23
2.3.2.2	Genotypisierung von genetischen Polymorphismen	25
2.3.2.2.1	<i>Genome-Wide Human SNP Array 6.0</i> von Affymetrix®	25
2.3.2.2.2	SNPlex® Technologie.....	25
2.3.2.2.3	<i>Dynamic allele-specific hybridization</i> (DASH).....	25
2.3.2.3	RNA-Isolierung aus humanen Venen.....	26
2.3.2.4	Reverse Transkription.....	27
2.3.2.5	Quantitative <i>Real-Time</i> PCR.....	27
2.3.2.6	DNA-Sequenzierung	30
2.3.2.7	<i>Fine-Mapping</i>	31
2.3.2.8	<i>Rapid Amplification of cDNA-Ends</i> (RACE)	31
2.3.2.8.1	<i>5'-Rapid Amplification of cDNA Ends</i> (5'-RACE)	31
2.3.2.8.2	<i>3'-Rapid Amplification of cDNA Ends</i> (3'-RACE)	33
2.3.2.9	Aufreinigung von Amplifikationsprodukten	34
2.3.2.10	Klonierung von Amplifikationsprodukten	35
2.3.2.11	Gelelektrophoretische Auftrennung von DNA	35
2.3.3	Zellbiologische Methoden.....	36
2.3.3.1	Kultivierung von <i>E. coli</i>	36
2.3.3.2	Transformation von <i>E. coli</i>	36
2.3.3.3	Isolierung von Plasmid-DNA aus <i>E. coli</i>	37
2.3.4	<i>In silico</i> -Methoden	37
2.3.4.1	Charakterisierung von genomischen Bereichen.....	37

2.3.4.2	Sequenzanalyse	37
2.3.4.3	Untersuchungen zur putativen Proteinkodierung	37
2.3.4.4	SNP-Datenbanken	37
2.3.4.5	<i>Expressed Sequence Tag</i> -Datenbanken	38
2.3.4.6	Oligonukleotid- <i>Design</i>	38
2.3.5	Auswahl des Assoziationsbereichs	38
3	ERGEBNISSE	39
3.1	Genomweite Assoziationsstudie zu zerebraler Malaria	39
3.1.1	Qualitätskontrolle der Einzelnukleotidaustausche	39
3.1.2	Populationsstratifikation der GWA-Studiengruppe	39
3.1.3	GWA-Studie zur zerebralen Malaria	42
3.2	Kontrolle der GWA-Ergebnisse in einer unabhängigen Fall-Kontroll-Gruppe.	43
3.3	Untersuchungen der kombinierten GWA- und Replikationsgruppen	45
3.4	Transmissions-Disequilibrium-Test (TDT) zur Bestätigung der bisherigen Assoziation	45
3.5	Fine Mapping der mit zerebraler Malaria assoziierten Region auf Chromosom 2p22.3	46
3.6	Nachweis von Transkripten in der mit zerebraler Malaria assoziierten chromosomalen Region 2p22.3	50
3.6.1	Recherche in Internet-Datenbanken	51
3.6.2	Suche nach offenen Leserahmen in den Transkripten	54
3.6.3	Sequenzanalyse von cDNA-Klonen der Transkripte	55
3.6.4	Gewebespezifische Genexpression der <i>Expressed Sequence Tags</i> ..	56
3.6.5	Quantifizierung der Transkriptmengen	58
3.6.6	Analysen zur Länge der cDNA-Sequenz der Transkripte	61
4	DISKUSSION	66
4.1	Populationsidentifikation	66
4.2	Assoziationsanalyse	67
4.3	Analysen zur Sensitivität der genomweiten Assoziationsstudie	68
4.4	Suche nach der kausalen Variante	70
4.5	<i>Long noncoding RNAs</i> im Assoziationsbereich?	71

4.6	Expression der Transkripte.....	72
4.7	Versuche zur Amplifikation vollständiger Transkripte	73
4.8	Mögliche Funktionen der RNAs.....	75
4.9	Ausblick.....	77
LITERATURVERZEICHNIS		79
ANHANG.....		87
DANKSAGUNG		108

ABBILDUNGSVERZEICHNIS

Abbildung 1:	Räumliche Verteilung der Malaria-Risikogebiete von 2009	1
Abbildung 2:	Zytoadhärenz parasitierter Erythrozyten im Lumen eines zerebralen Blutgefäßes	4
Abbildung 3:	Schematische Übersicht über Haplotypen	10
Abbildung 4:	Quantile-Quantile-Plot der p-Werte des additiven Modells der GWA-Studiengruppe ...	20
Abbildung 5:	Schematische Abbildung der 5'-RACE mit der <i>CapFinder</i> -Methode	32
Abbildung 6:	Schematische Abbildung der 3'-RACE PCR	34
Abbildung 7:	1. und 2. Dimension der multidimensionalen Skalierung (MDS), stratifiziert nach ethnischer Herkunft	40
Abbildung 8:	1. und 2. Dimension der multidimensionalen Skalierung (MDS), stratifiziert nach Fall-Kontroll-Status	41
Abbildung 9:	1. und 2. Dimension der multidimensionalen Skalierung (MDS), stratifiziert nach Geschlecht	41
Abbildung 10:	<i>Manhattan-Plot</i> der GWA-Studiengruppe	42
Abbildung 11:	Kopplungsungleichgewicht [LD] von genetischen Varianten in einem Bereich von 20 kb um rs12479448	49
Abbildung 12:	Locus auf Chromosom 2p22.3 mit Assoziation zu zerebraler Malaria	50
Abbildung 13:	Annotierte Transkripte auf Chromosom 2p22.3	53
Abbildung 14:	Putative offene Leserahmen (ORF) abgeleitet aus cDNA-Sequenzen	54
Abbildung 15:	Expressionsstärke der Gene der Transkripte	60
Abbildung 16:	Schematische Abbildung der durch <i>Rapid Amplification of cDNA Ends</i> (RACE)-PCR ermittelten 3'- und 5'-Enden des AI/AW/BX-Transkripts	62
Abbildung 17:	Schematische Abbildung der durch <i>Rapid Amplification of cDNA Ends</i> (RACE)-PCR ermittelten 3'- und 5'-Enden des AA-Transkripts	64
Abbildung 18:	Schematische Abbildung der durch <i>Rapid Amplification of cDNA Ends</i> (RACE)-PCR ermittelten 3'- und 5'-Enden des DA-Gens	65

TABELLENVERZEICHNIS

Tabelle 1:	Übersicht über relevante Gene, für deren Varianten eine Assoziation mit Malaria tropica nachgewiesen werden konnte.....	6
Tabelle 2:	Demographische Charakteristika der Studiengruppen	13
Tabelle 3:	Angabe der klinischen Symptome der zerebralen Malaria-Fälle.....	14
Tabelle 4:	Standard-PCR-Ansatz und -Programm mit der Firepol® Polymerase (Solis)	24
Tabelle 5:	Standard-PCR-Ansatz und -Programm mit der AmpliTaq Gold® 360 DNA Polymerase (Invitrogen).....	24
Tabelle 6:	PCR-Ansatz und -Programm und der FRET-Analyse.....	26
Tabelle 7:	Ansatz und Programm der quantitativen <i>Real-Time</i> PCR für den LightCycler 480® Instrument	29
Tabelle 8:	Effizienzen der Oligonukleotidpaare der quantitativen Real-Time PCR	29
Tabelle 9:	Reaktionsansatz und PCR-Programm der Sequenzierung nach Sanger.....	30
Tabelle 10:	Ergebnisse der Assoziationsstudien in der GWA-Studiengruppe und der Replikationsgruppe.....	44
Tabelle 11:	Kopplungsungleichgewicht und Entfernung der für eine Genotypisierung im gesamten Studienkollektiv ausgewählten genetischen Varianten zum SNP rs12479448	47
Tabelle 12:	Ergebnisse von Assoziationstests nach Kombination der GWA- und der Replikationsgruppe und von ausgewählten Varianten mit einem Korrelationskoeffizient von $r^2 \geq 0,6$ zu rs12479448 nach dem <i>Fine Mapping</i> einer 100 kb-Region auf Chromosom 2p22.3	48
Tabelle 13:	Charakteristika der Transkripte des Assoziationsbereichs auf Chromosom 2p22.3 aus den Internet-basierten Datenbanken von UCSC und NCBI	52
Tabelle 14:	Proteine mit einer Sequenzidentität von > 50 % zu den offenen Leserahmen der Transkripte aus <i>blastp</i>	55
Tabelle 15:	Länge der Exons und Introns der Transkripte im Assoziationsbereich.....	57
Tabelle 16:	Gewebespezifische Expression der Transkripte	58

ZUSAMMENFASSUNG

Zerebrale Malaria ist eine Form der schweren Malaria, deren Pathogenese bis heute Gegenstand der Forschung ist. Für zerebrale Malaria, andere Formen der schweren Malaria und auch für milde Malaria wurden in der Vergangenheit menschliche genetische Varianten beschrieben, die einen Schutz vor der Erkrankung vermitteln. Diese Arbeit nutzt einen hypothesenfreien Ansatz, um mit einer genomweiten Assoziationsstudie in einer Studiengruppe von 568 Kindern mit zerebraler Malaria und 2123 gesunden Kontroll-Kindern weitere genetische Varianten zu suchen, die Schutz oder Empfänglichkeit hinsichtlich der Erkrankung beeinflussen. Ziel ist es, durch Identifikation von funktionellen genetischen Varianten Stoffwechselwege zu erkennen, die eine entscheidende Rolle in der Pathogenese spielen.

Mit einer Assoziationsstudie, welche eine logistische Regression für Korrekturen auf ethnische Stratifikation, Geschlecht und Alter beinhaltete, wurde in der Fall-Kontroll-Studie ein Locus auf Chromosom 2p22.3 identifiziert. Der Einzelnukleotidaustausch rs12479448 zeigte die stärkste Assoziation mit zerebraler Malaria (OR 1,6; 95 % KI 1,2 – 1,9; $p = 1,2 \times 10^{-8}$). Zusätzlich wurde mit einem familienbasierten genetischen Testverfahren, dem Transmissions-Disequilibriumstest, das Assoziationssignal bestätigt. Der assoziierte Locus befindet sich in einer intergenischen Region. In einem Assoziationsbereich, der durch Einzelnukleotidaustausche mit einem Signifikanzniveau von $p < 0,05$ und mit einem Kopplungsungleichgewicht von $r^2 > 0,05$ zur Variante rs12479448 definiert wurde, fanden sich Sequenzinformationen von fünf Transkripten. Mit einer *In silico*-Analyse der cDNA-Sequenzen dieser Transkripte konnte keine Protein-kodierende Funktion aufgezeigt werden. Expressionsanalysen in 21 untersuchten cDNAs aus unterschiedlichen Geweben belegten eine sehr geringe und gewebespezifische Expression der Transkripte. Weitere Untersuchungen an den Transkripten machten das Vorkommen von alternativ gespleißten Isoformen für vier der Transkripte wahrscheinlich.

Die geringe Konzentration und die fehlende Protein-kodierende Funktion der Transkripte deuten auf eine mögliche regulatorische Funktion dieser nicht kodierenden RNAs hin. Sie können über diesen Mechanismus eine Rolle in der Pathogenese der zerebralen Malaria spielen.

SUMMARY

Cerebral malaria is a form of severe malaria whose pathogenesis has been subject of research up to now. In the past, human genetic variants have been described that mediate protection against cerebral malaria and other forms of severe malaria as well as for mild malaria. In its search for further genetic variants which impact upon protection against the disease or susceptibility to it, this study employs an hypothesis free approach via a genome wide association study of a group consisting of 568 children with cerebral malaria and 2123 healthy children. The objective is to detect metabolic pathways which play a decisive role in the pathogenesis through functional genetic variants.

With the aid of an association study comprising a logistic regression corrected for population stratification, sex and age, a locus on chromosome 2p22.3 was identified in the case-control study. The single nucleotide polymorphism rs12479448 showed the strongest association with cerebral malaria (OR 1,6; 95 % KI 1,2 – 1,9; $p = 1,2 \times 10^{-8}$). Additionally, the association signal was confirmed by a family based genetic test procedure, the transmission disequilibrium test. The associated locus is in an intergenic region. In an associated area, which was defined by SNPs with a significance level of $p < 0,05$ and a linkage disequilibrium of $r^2 > 0,05$ to the variant rs12479448, sequence information of five transcripts was found. Using *in silico* analysis, no protein coding function could be shown in the cDNA sequences of the transcripts. Moreover, expression analysis of 21 examined cDNAs of different tissues proved a very low as well as a tissue specific expression of the transcripts. Further research on the transcripts made the occurrence of alternative spliced isoforms in four of the transcripts likely.

The low concentration of the transcripts and the missing protein coding function give a hint of a possible regulatory function of these noncoding RNAs. Via this mechanism, the transcripts could play a role in the pathogenesis of cerebral malaria.

ABKÜRZUNGSVERZEICHNIS

%	Prozent
A	Adenin
AA	AA861168
ABO	ABO-Blutgruppensystem
ACTB	Beta Aktin
AI	AI218168
AIDS	<i>Acquired immunodeficiency syndrome</i>
Amp	Ampicillin
AW	AW665463
AS	Aminosäure
bp	Basenpaare
BLAST	<i>Basic Local Alignment Tool</i>
bzw.	beziehungsweise
BX	BX091678
°C	Grad Celsius
C	Cytosin
CD36	<i>Cluster of Differentiation 36</i>
cDNA	komplementäre DNA
CR	Komplementrezeptor
Ct	Cycle of Threshold
D'	Korrelationskoeffizient D als Maß für Kopplungsungleichgewicht
DA	DA227815
DASH	<i>Dynamic allele-specific hybridization</i>
ddH ₂ O	Doppeltdestilliertes Wasser
dl	Deziliter
DMSO	Dimethylsulfoxid
dNTP	Desoxynukleosidtriphosphat
DNA	Desoxyribonukleinsäure
DTT	Dithiothreitol
E	Effizienz
<i>E. coli</i>	Escherichia coli
EDTA	Ethylendiamintetraessigsäure
<i>et al.</i>	und andere
EtOH	Ethanol
EST	<i>Expressed Sequence Tag</i>
FRET	<i>Fluorescence Resonance Energy Transfer</i>
g	Gramm
G	Guanin
G6PD	Glucose-6-Phosphat-Dehydrogenase
gDNA	genomische DNA
GSP	<i>Gene specific primer</i>
GWA-Studie	genomweite Assoziationsstudie

GYP	Glycophorin
H	Stunde/n
H ₂ O	Wasser
HbB	Hämoglobin
HbC	Hämoglobin C
HbE	Hämoglobin E
HbS	Hämoglobin S
ICAM-1	<i>Intracellular Adhesion Molecule 1</i>
IFNGR1	Gen des Interferon γ Rezeptors 1
IL	Interleukin
iNOS	induzierte Stickstoffsynthetase
<i>IRF1</i>	Gen des Interferon regulatorischen Faktors 1
<i>in vitro</i>	außerhalb eines Organismus
<i>in vivo</i>	innerhalb eines Organismus
<i>in silico</i>	Computer generiert
kb	Kilobasenpaare
KI	Konfidenzintervall
l	Liter
LB	Luria Bertani
LD	Kopplungsungleichgewicht (<i>Linkage Disequilibrium</i>)
Mb	Megabasen
MDS	Multidimensionale Skalierung
MgCl ₂	Magnesiumchlorid
ml	Milliliter
mM	millimolar
min	Minute/n
μ g	Mikrogramm
μ l	Mikroliter
n	Anzahl
NaCl	Natriumchlorid
NE	normalisierte Expression
ng	Nanogramm
nm	Nanometer
NOS2A	Gen der induzierbaren Stickstoffoxid Synthetase
NCBI	<i>National Center for Biotechnology Information</i>
OR	<i>Odds Ratio</i>
ORF	<i>Open reading frame</i>
p	p-Wert
PCR	Polymerase-Kettenreaktion
pg	Pikogramm
pH	pH-Wert
PPP1CA	katalytische α -Untereinheit der Protein Phosphatase 1

<i>P. sp.</i>	<i>Plasmodium species</i>
q	kurzer Arm eines Chromosoms
qRT-PCR	quantitative Real-Time PCR
r^2	Quadrat des Korrelationskoeffizienten r als Maß für Kopplungsungleichgewicht
RACE	<i>Rapid Amplification of cDNA ends</i>
RNA	Ribonukleinsäure
rpm	Umdrehungen pro Minute
RR	Relatives Risiko
RT	Raumtemperatur
RT-PCR	Reverse Transkription mit folgender cDNA-Synthese
<i>SCL4A1</i>	Gen des Bande 3 Proteins
sec	Sekunde
SNP	Einzelnukleotidaustausch (<i>single nucleotide polymorphism</i>)
T	Thymin
TBE	TRIS-Borat-EDTA
TDT	Transmissions-Disequilibriumstest
TE	TRIS-EDTA
TNF	Tumor Nekrose Faktor
U	Reaktionseinheiten
UCSC	<i>University of Santa Cruz</i>
UV	Ultraviolette Strahlung
V	Volt
WHO	<i>World Health Organization</i>
w/v	<i>weight/volume</i> (Masse/Volumen)

1 Einleitung

1.1 Malaria

Malaria ist neben AIDS und Tuberkulose eine der bedeutendsten Infektionskrankheiten. Etwa 40 % der Weltbevölkerung sind dem Risiko einer Malaria ausgesetzt (Abbildung 1). Für das Jahr 2010 wurden von der Weltgesundheitsorganisation (WHO) weltweit 216 Millionen Malariafälle geschätzt (WHO, 2011), von denen etwa eine Million Erkrankte starben (Murray *et al.*, 2012). In den Endemiegebieten zählen meist Kinder unter 5 Jahren und Schwangere zu den Opfern. Aufgrund der hohen Anzahl an Krankheits- und Todesfällen verursacht Malaria große sozioökonomische Belastungen für die betroffenen Länder (Sachs & Malaney, 2002).

Verursacht wird Malaria durch Parasiten der Gattung *Plasmodium*, von denen bis heute fünf humanpathogene Spezies bekannt sind: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae* und *Plasmodium knowlesi*.

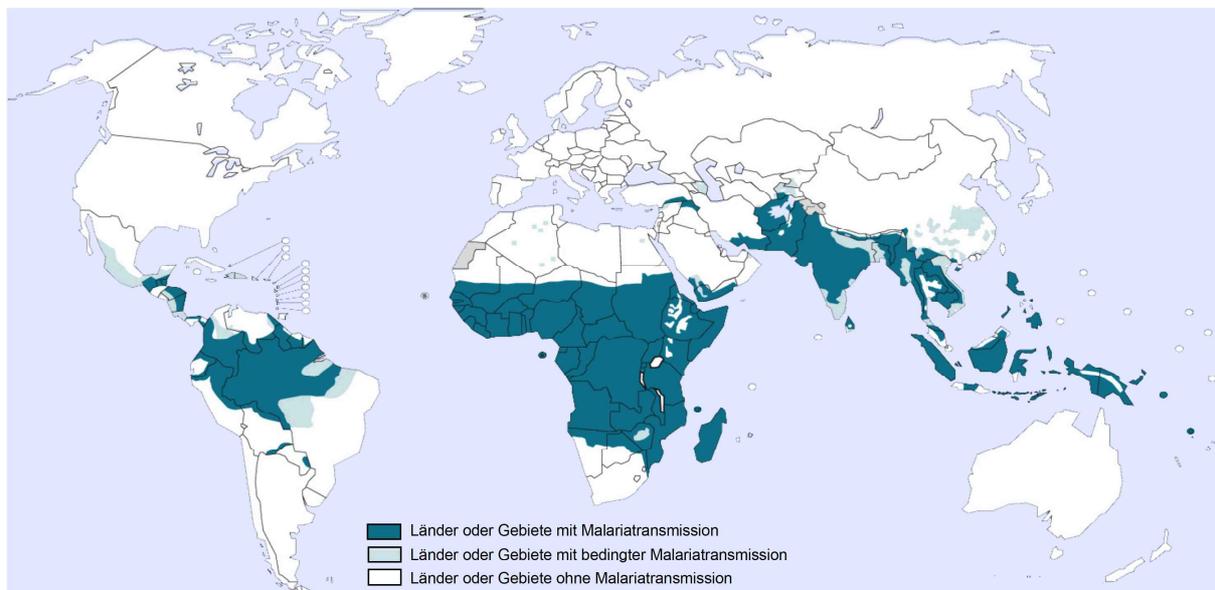


Abbildung 1: Malaria-Risikogebiete von 2009
(verändert nach WHO, 2010)

1.1.1 Malaria tropica

Der Parasit *P. falciparum* ist der Erreger der Malaria tropica. Das klinische Bild der Infektion reicht von asymptomatischer Parasitämie bis zu schwerer Malaria mit tödlichem Verlauf. Meist sind die Symptome vielseitig und wenig spezifisch, üblich sind Fieberschübe mit Schüttelfrost, Kopf- und Gliederschmerzen sowie Durchfall (WHO, 2000). Es können aber lebensbedrohliche Komplikationen wie schwere Anämie, zerebrale Malaria, Hypoglykämie, metabolische Azidose und ein Atemnotsyndrom auftreten (Idro *et al.*, 2010). In Endemiegebieten entwickeln von den erkrankten Kindern etwa 0,5 % ein solches Krankheitsbild, das von der WHO „schwere Malaria“ genannt wurde (Roca-Feltrer *et al.*, 2008). Schwere Verlaufsformen werden ganz überwiegend durch *P. falciparum* verursacht.

1.1.2 Zerebrale Malaria

Zerebrale Malaria zählt nach der schweren Anämie zu der zweithäufigsten Komplikation einer Malaria tropica (Newton *et al.*, 1998). Die Inzidenz liegt in endemischen Regionen Afrikas bei 1120 Kindern/100000/Jahr (Idro *et al.*, 2010). Dort erkranken Kinder unter 5 Jahren am häufigsten an zerebraler Malaria (Idro *et al.*, 2010). Allein in Afrika entwickeln jährlich mehr als 575000 Kinder diese neurologische Komplikation der Erkrankung (Idro *et al.*, 2010).

1.1.2.1 Klinik und Diagnose

Die WHO definiert zerebrale Malaria als lebensbedrohliche Krankheit, die durch einen komatösen Zustand charakterisiert ist und durch asexuelle Formen von *P. falciparum* im Blut verursacht wird (WHO, 2000). Andere Ursachen für eine Bewusstlosigkeit wie Fieberkrämpfe, Hypoglykämie oder Meningitis müssen ausgeschlossen werden. Durch die fehlende Spezifität der Diagnose könnten sonst Patienten mit einem Koma anderer Genese und simultan auftretender Parasitämie unrichtig als zerebrale Malaria diagnostiziert werden (Berkley *et al.*, 1999; Idro *et al.*, 2007; Schubart *et al.*, 2006; Taylor *et al.*, 2004). Dies beruht auf der Tatsache, dass Individuen in Endemiegebieten eine Teilimmunität gegenüber *P. falciparum* entwickeln und auch bei Parasitämie asymptomatisch bleiben können (Snow *et al.*, 1997). Da hauptsächlich Kinder unter die Erkrankten fallen, wurde zur Klassifizierung

der Schwere des Komas eigens für sie der *Blantyre Coma Scale* entworfen (Molyneux *et al.*, 1989).

Bei Kindern setzt das Koma meist plötzlich ein und kann von Krämpfen begleitet werden (Idro *et al.*, 2010). Einige Kinder fallen aus einem Zustand fortschreitender Schwäche und Erschöpfung ins Koma. Zu weiteren häufig beobachteten Zeichen zählen Hirnödeme, intrakranielle Hypertension und Anzeichen auf Dysfunktionen im Stammhirn. Als zusätzliche Begleiterscheinungen können Anämie, metabolische Azidose, Ungleichgewicht des Elektrolythaushalts, Hypoglykämie sowie ein Schockzustand beobachtet werden (Idro *et al.*, 2010). In vielen Fällen konnten Schwellungen der Sehnerven, aber auch weiße Netzhautveränderungen (*retinal whitening*) oder Blutungen in der Netzhaut festgestellt werden (Beare *et al.*, 2006). In über 10 % der Fälle einer zerebralen Malaria kommt es zu neurologischen Residuen (Brewster *et al.*, 1990). Zu den Risikofaktoren der neurologischen und kognitiven Folgeerscheinungen werden eine Vielzahl an Krämpfen, tiefes und lang anhaltendes Koma, Hypoglykämie sowie intrakranielle Hypertension gezählt (Idro *et al.*, 2006). Bereits 11 % der erkrankten Kinder weisen bei ihrer Entlassung neurologische Defizite auf (Brewster *et al.*, 1990). Komplikationen wie Erblindung, Störungen in der Bewegungskoordination und zentrale Hypotonie bilden sich mit der Zeit zurück (van Hensbroek *et al.*, 1997). Bei 25 % der erkrankten Kinder kommt es jedoch zu langfristigen Beeinträchtigungen, insbesondere der kognitiven Fähigkeiten (John *et al.*, 2008). In 10 – 15 % der Fälle sterben die Kinder an zerebraler Malaria (Idro *et al.*, 2010).

1.1.2.2 Pathogenese

In der Literatur finden sich unterschiedliche Hypothesen zur Pathogenese der zerebralen Malaria, so dass diese trotz der bekannten involvierten Prozesse nach wie vor als unverstanden gilt (Idro *et al.*, 2010; van der Heyde *et al.*, 2006). Es wird vermutet, dass die Zytoadhärenz von infizierten Erythrozyten an Rezeptoren des humanen Wirtsendothels am Pathomechanismus vieler schwerer Verläufe einer Malaria tropica beteiligt ist. Diese so genannte Sequestrierung erfolgt über variable Oberflächenproteine des Parasiten in der Blutphase von *P. falciparum* (Baruch *et al.*, 1996; Smith *et al.*, 1995). Patienten mit zerebraler Malaria zeigen eine erhöhte

Sequestrierung am zerebralen mikrovaskulären Endothel (Abbildung 2; Pongponratn *et al.*, 1991; Pongponratn *et al.*, 2003; Ponsford *et al.*, 2011).

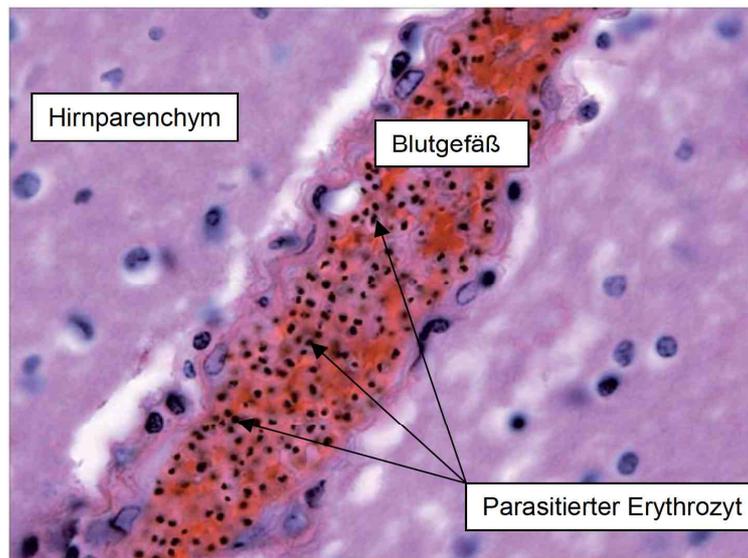


Abbildung 2: Zytoadhärenz parasitierter Erythrozyten im Lumen eines zerebralen Blutgefäßes
Schnitt durch Hirngewebe nach Hämatoxylin-Eosin-Färbung eines an zerebraler Malaria erkrankten Kindes, 1000-fache Vergrößerung (O'Connell, 2007).

Mit der Bindung an das humane Endothel umgeht der Parasit die Passage durch die Milz, in der infizierte Erythrozyten ausgesondert werden (Saul, 1999). Infizierte Erythrozyten binden außerdem an umliegende infizierte Erythrozyten (Autoagglutination) und nicht-infizierte Erythrozyten (*Rosetting*; Miller *et al.*, 2002). Im peripheren Gefäßsystem kann die Zytoadhärenz der infizierten Erythrozyten zum Verschluss von kleinen Blutgefäßen führen. Dieser wird als Auslöser von vermindertem Blutfluss, Sauerstoffmangel und Blutungen diskutiert (Idro *et al.*, 2005; Schofield & Grau, 2005). In der Pathogenese der zerebralen Malaria wird daher der Sequestrierung eine entscheidende Rolle in der Funktionsstörung des Endothels zugesprochen. Gestützt wird diese Hypothese von Beobachtungen, die eine Korrelation zwischen dem Grad der intrazerebralen Sequestrierung mit dem Schweregrad des Komas zeigten (Pongponratn *et al.*, 2003). Als wichtigster Rezeptor zur Bindung der infizierten Erythrozyten an das zerebrale Endothel wird ICAM-1 (*Intracellular Adhesion Molecule 1*) angesehen (Ochola *et al.*, 2011; Turner *et*

al., 1994). Zusätzlich zur Sequestrierung von infizierten Erythrozyten konnte eine Sequestrierung und Akkumulation von Thrombozyten und Leukozyten im Gehirn beobachtet werden (Coltel *et al.*, 2004; Grau *et al.*, 2003). Darüber hinaus könnten die Zytoadhärenz und dadurch ausgelöste Prozesse die Barrierefunktion des Endothels beeinträchtigen (Francischetti, 2008). So initiiert zum Beispiel die Aktivierung des Hirnendothels eine Reihe von Prozessen, die möglicherweise in einer Fehlfunktion der Bluthirn-Schranke resultieren (Adams *et al.*, 2002; Medana & Turner, 2006). Beteiligt sind inflammatorische Prozesse, Fehlfunktionen von Nervenzellen, Freisetzung vaskulärer und thrombozytärer Mikropartikel und letztendlich die Apoptose von Endothel- und Nervenzellen in der betroffenen Region (Combes *et al.*, 2004; Medana *et al.*, 2007; Moxon *et al.*, 2011; Schofield & Grau, 2005; Turner *et al.*, 1994).

1.2 Humangenetik der Malaria

Haldane beschrieb eine geografische Übereinstimmung in der Verbreitung des vererbaren Hämoglobindefekts β -Thalassämie und Malaria im Mittelmeerraum (Haldane, 1949). Er postulierte, dass der Nachteil einer homozygoten Ausprägung des β -Thalassämie-Allels durch den Vorteil einer heterozygoten Ausprägung über einen partiellen Schutz vor Malaria ausgeglichen werden würde. Infolge dessen stellt sich ein Gleichgewicht der Allele im Genpool der Population ein (balancierter Genpolymorphismus). Dieses Prinzip der balancierten Selektion soll in Malaria-endemischen Gebieten zu einer relativ hohen Frequenz von den Hämoglobinallelen Hämoglobin S (HbS) und Hämoglobin C (HbC) geführt haben (Flint *et al.*, 1998).

Schätzungen zufolge beeinflussen humane genetische Faktoren zu 25 % das Risiko, an einer Malaria zu erkranken (Mackinnon *et al.*, 2005). Bis heute sind eine Reihe von Genvarianten publiziert, die Resistenz bzw. Empfänglichkeit gegenüber Malaria beeinflussen (Tabelle 1). Die schützenden Mechanismen der aufgeführten genetischen Varianten sind bislang Gegenstand der Forschung.

In einigen der in Tabelle 1 aufgelisteten Gene wurden Assoziationen von Varianten mit zerebraler Malaria nachgewiesen. So konnte ein Schutz-vermittelnder Mechanismus durch eine Deletion im *CD36*-Gen gezeigt werden, dessen Protein bei

der Sequestrierung infizierter Erythrozyten eine entscheidene Rolle spielt (Omi *et al.*, 2003).

Tabelle 1: Übersicht über relevante Gene, für deren Varianten eine Assoziation mit Malaria tropica nachgewiesen werden konnte

Gen	Protein	Publikation
<i>HBA1, HBA2</i>	α -Globin	Flint <i>et al.</i> , 1986
<i>HBB</i>	β -Globin	Willcox <i>et al.</i> , 1983
<i>SCL4A1</i>	Bande 3 Protein	Allen <i>et al.</i> , 1999
<i>CD36</i>	Thrombospondin Rezeptor	Aitman <i>et al.</i> , 2000
<i>ICAM-1</i>	Interzelluläres Adhäsionsmolekül-1	Fernandez-Reys <i>et al.</i> , 1997
<i>ABO</i>	Glycosyl-Transferase	Fischer <i>et al.</i> , 1998
<i>GYPC</i>	Glycophorin C	Patel <i>et al.</i> , 2001
<i>TNFA</i>	Tumornekrosefaktor α	McGuire <i>et al.</i> , 1994
<i>IL-1A</i> <i>IL-1B</i>	Interleukin 1 α Interleukin 1 β	Walley <i>et al.</i> , 2004
<i>IL-10</i>	Interleukin 10	Wilson <i>et al.</i> , 2005
<i>IL-12b</i>	Interleukin 12 β	Morahan <i>et al.</i> , 2002
<i>IFNGR1</i>	Interferon γ Rezeptor 1	Koch <i>et al.</i> , 2002
<i>IRF1</i>	Interferon regulatorischer Faktor 1	Mangano <i>et al.</i> , 2008
<i>CR1</i>	Komplementrezeptor 1	Teeranaipong <i>et al.</i> , 2008
<i>NOS2A</i>	Induzierte Stickstoffsynthetase	Kun <i>et al.</i> , 1999
<i>G6PD</i>	Glucose-6-Phosphat-Dehydrogenase	Ruwende <i>et al.</i> , 1995

Darüber hinaus wurden mehrere Promotorvarianten in immunrelevanten Genen mit zerebraler Malaria assoziiert. Darunter fallen die Gene von IL-12b (Marquet *et al.*, 2008; Morahan *et al.*, 2002; Naka *et al.*, 2009), TNF- α (Knight *et al.*, 1999; McGuire *et al.*, 1994) und Komplementrezeptor 1 (Teeranaipong *et al.*, 2008). Kontroverse Ergebnisse finden sich in der Literatur zu Assoziationen mit zerebraler Malaria von Polymorphismen im Gen der induzierten Stickstoffsynthetase, *NOS2A* (Burgner *et al.*, 2003; Burgner *et al.*, 1998; Hobbs *et al.*, 2002; Kun *et al.*, 1998; Levesque *et al.*,

1999) und einer Variante im *ICAM-1*-Gen (Fernandez-Reyes *et al.*, 1997; Fry *et al.*, 2008; Kun *et al.*, 1999). In Bezug auf eine für zerebrale Malaria spezifische Assoziation wurde eine zum Krankheitsbild der Ovalozytose führende Deletion von 27 bp im Gen des Bande 3 Proteins, eine Promotorvariante des Interferon- γ -Rezeptors und die Hämoglobin C Variante mit einem Schutz beschrieben (Allen *et al.*, 1999; Koch *et al.*, 2002; May *et al.*, 2007).

1.3 Auswahl des Studiendesigns

Wie oben beschrieben liegen für Malaria hauptsächlich Untersuchungen an Kandidatengen vor. Kandidatengenansätze sind dann möglich, wenn biologische Prozesse bei der Entstehung einer Erkrankung bereits bekannt sind. Gene, deren Produkte in diesen Prozessen eine Rolle spielen, können so als mögliche krankheitsassoziierte Gene identifiziert werden. In der vorliegenden Studie sollten hingegen in einem hypothesenfreien, genomweiten Ansatz unerwartete Wirtsfaktoren gefunden werden, die die Entwicklung einer zerebralen Malaria beeinflussen.

Man unterscheidet in der genetischen Epidemiologie zwischen Fall-Kontroll-, Querschnitts- und Kohortenstudien. Eine häufige Anwendung finden Fall-Kontroll-Studien, da sie sich besonders für seltene Erkrankungen eignen und keine lange Beobachtungszeit voraussetzen (Bickemöller & Fischer, 2007). Da schwere Malaria nur von etwa 0,5 % der an Malaria erkrankten Personen entwickelt wird, wurde für die Untersuchung ein Fall-Kontroll-Studiendesign gewählt. Man unterscheidet zwischen Kopplungs- und Assoziationsstudien.

Von Kopplung spricht man, wenn in einer Familie Allele überzufällig häufig an erkrankte Nachkommen vererbt werden. Man unterscheidet Untersuchungen an Eltern/Kind- und Geschwisterpaaren. Kopplungsanalysen können genomweit oder anhand von Kandidatengen bzw. mutmaßlich assoziierten Loci durchgeführt werden. In genomweiten Analysen wird oftmals Kopplung mit Erkrankungen in genomweiten Regionen gefunden, die mehrere Megabasen große chromosomale Bereiche umfassen. In diesen Regionen können jedoch einige 100 bis 1000 Gene lokalisiert sein (Cardon & Bell, 2001). Genomweite Kopplungsanalysen können daher nur als Grobkartierung von assoziierten Loci dienen.

In einer genomweiten Kopplungsanalyse zu schwerer Malaria konnte ein schwaches Signal in einer Region auf Chromosom 5q31-33 gezeigt werden (Garcia *et al.*, 1998b). Dieser Locus zeigte auch in genomweiten Kopplungsstudien zur milden bzw. asymptomatischen Malaria eine Assoziation (Milet *et al.*, 2010; Sakuntabhai *et al.*, 2008; Timmann *et al.*, 2007). Die Kopplung dieses Locus konnte inzwischen durch eine populations- und familienbasierte Analyse auf einen Polymorphismus im *IL3*-Gen zurückgeführt werden (Meyer *et al.*, 2011). Kopplungsanalysen eignen sich besonders für Analysen von monogenen Erkrankungen mit seltenen Mutationen (Lander & Schork, 1994). Da Malaria jedoch zu den komplexen Erkrankungen gezählt wird (Garcia *et al.*, 1998a; Rihet *et al.*, 1998), wurde die vorliegende Arbeit als Assoziationsanalyse konzipiert. Um dabei eine hypothesenfreie und systematische Suche nach Assoziationen von genetischen Markern mit zerebraler Malaria zu gewährleisten, wurde eine genomweite Assoziationsstudie (GWA-Studie) durchgeführt.

GWA-Studien ermöglichen die Suche von Assoziationen zwischen Erkrankungen und Varianten in Genen, die vorher nicht mit der Erkrankung in Verbindung gebracht wurden. Assoziationsstudien vergleichen die Frequenzen von genetischen Varianten in einer Fall- und einer Kontrollgruppe. Im Gegensatz zur Kopplungsanalyse werden in Assoziationsstudien keine familienbasierten Ähnlichkeiten von Vererbungsmustern, sondern populationsbasierte Frequenzen betrachtet. Ein Nachteil, der sich daraus für Assoziationsstudien ergibt, ist seine erhöhte Anfälligkeit gegenüber Populationsstratifikationen (Lander & Schork, 1994).

Um Assoziationen, die mithilfe von GWA-Studien nachgewiesen werden, zu bestätigen, ist es möglich, einen Transmissions-Disequilibriumstest (TDT) durchzuführen. Beim TDT handelt es sich um einen familienbasierten Test zur Prüfung von Assoziation bei gleichzeitig vorliegender Kopplung zwischen einem genetischen Locus und einer Erkrankung (Spielman *et al.*, 1993). In so genannten Trios (Kernfamilien) wird die Vererbungsfrequenz von Allelen untersucht. Dem Test liegt das Prinzip zugrunde, dass ein elterliches Allel, wenn es mit einer Erkrankung gekoppelt und assoziiert ist, häufiger an ein erkranktes Kind vererbt wird als ein nicht-assoziiertes/nicht-gekoppeltes Allel. Ausschließlich heterozygote Eltern sind hierbei informativ. Der Vorteil des TDTs ist die automatische interne Kontrolle durch die nicht weitervererbten Allele, welche ein Auftreten von falsch-positiven

Ergebnissen durch Populationsstratifikation (ethnische, soziale und geographische Stratifizierungen) ausschließt. Da lediglich innerhalb von Familien getestet wird, leidet der Test jedoch unter einer relativ geringen statistischen *Power*.

Als genetische Marker werden in GWA-Studien Einzelnukleotidaustausche (*single nucleotide polymorphisms*, SNPs) verwendet. Sie stellen mit schätzungsweise 15 Mio. SNPs bei einer Allelfrequenz größer 1 % die häufigste Variante des humanen Genoms dar (1000 Genomes Project Consortium, 2010). Im Vergleich zu anderen genetischen Varianten weisen sie eine geringe Mutationsrate auf und gelten daher als stabile Marker, auch in der humanen Evolution (Sachidanandam *et al.*, 2001). Ihre meist biallelische Eigenschaft erlaubt es, SNPs in einem weitgehend automatisierten Hochdurchsatzverfahren zu genotypisieren und mathematisch auszuwerten (Sachidanandam *et al.*, 2001). Ob SNPs funktionelle Auswirkungen besitzen, hängt von ihrer genomischen Position ab. In kodierenden Regionen führen so genannte nicht-synonyme SNPs zum Austausch einer Aminosäure oder zum Kettenabbruch eines Proteins. Auch können an Spleißstellen gelegene SNPs zu Veränderungen in der RNA-Prozessierung führen. Synonyme SNPs hingegen bleiben ohne Folgen für das Protein, da durch sie kein Aminosäureaustausch hervorgerufen wird. Die meisten SNPs liegen in nicht-kodierenden Regionen des Genoms (Risch, 2000). Sind sie jedoch in regulatorischen Elementen wie Promotoren, *Enhancern* oder *Silencern* lokalisiert, können sie ebenfalls funktionelle Auswirkungen besitzen.

Mit den verfügbaren Genotypisierungsplattformen kann jeweils nur eine definierte Anzahl an SNPs typisiert werden. Die Vererbung benachbarter SNP-Allele erfolgt jedoch nicht zufallsgemäß sondern gekoppelt in definierten Haploblöcken. Man spricht dabei von einem so genannten Kopplungsungleichgewicht (*linkage disequilibrium*, LD). Besteht zwischen zwei SNPs ein hohes LD, kann aufgrund des Genotyps eines SNPs mit hoher Wahrscheinlichkeit der Genotyp des anderen SNPs vorausgesagt werden (Abbildung 3). SNPs, die ein hohes LD zu weiteren SNPs aufweisen, werden als *Tagging* SNPs bezeichnet. Durch die Genotypisierung von *Tagging* SNPs kann so in GWA-Studien eine möglichst hohe genomische Abdeckung erreicht werden. Diese Herangehensweise wird als *LD-Mapping* bezeichnet.

Im Gegensatz zu anderen verbreiteten Erkrankungen ist die Suche nach Resistenz- bzw. Empfänglichkeitsallelen bei Infektionskrankheiten langsamer vorangeschritten (de Bakker & Telenti, 2010). Bis heute wurde nur eine GWA-Studie zu *Malaria tropica* in einer Studienpopulation aus Gambia veröffentlicht (Jallow *et al.*, 2009). In dieser Studie wurde jedoch keine neue genetische Variante mit einem Effekt auf schwere Malaria gefunden.

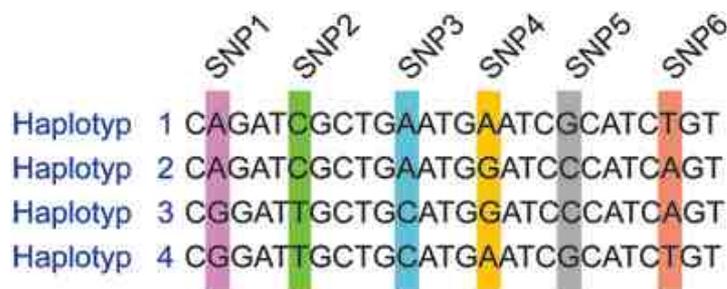


Abbildung 3: Schematische Übersicht über Haploblöcke

Einzelnukleotidaustausche (SNPs) sind mit farblichen Balken hervorgehoben. SNP 1, 2 und 3 sowie SNP 4, 5 und 6 weisen eine hohe Kopplung auf. Jeder SNP dieser gekoppelten SNPs kann als *Tagging* SNP der anderen zwei SNPs dienen (nach Manolio *et al.*, 2008).

1.4 Zielsetzung

In der vorliegenden Studie sollten in einem hypothesenfreien Ansatz neue Gene oder Stoffwechselwege identifiziert werden, die Hinweise auf die Pathogenese der zerebralen Malaria geben. Für dieses Vorgehen wurde eine GWA-Studie nach dem Fall-Kontroll-Studiendesign gewählt.

In der GWA-Studie sollten mithilfe des *Affymetrix® Genome Wide SNP Array 6.0* die Genotypen von mehr als 906600 SNPs in einer ghanaischen Studienpopulation, bestehend aus 409 Fällen mit zerebraler Malaria und 828 gesunden Kontrollen, bestimmt werden. Auf Basis der Genotypen der Fälle und Kontrollen sollten Assoziationsberechnungen durchgeführt werden. Assoziierte Loci sollten anhand einer weiteren Studiengruppe aus derselben Studienregion, bestehend aus 159 Fällen zerebraler Malaria und 2123 gesunden Kontrollen, bestätigt oder verworfen werden. Bei bestätigter Assoziation sollte ein *Fine Mapping* in der chromosomalen Region durchgeführt werden, durch das die kausale Variante eingegrenzt oder

gefunden werden könnte. Darüber hinaus sollten Transkripte, die in der Nähe des assoziierten Locus liegen, charakterisiert und hinsichtlich ihres Expressionsverhaltens untersucht werden.

2 Patienten, Material und Methoden

2.1 Studiendesign und Studienpopulation

2.1.1 Studiendesign

Das Ziel dieser genomweiten Assoziationsstudie (GWA-Studie) war es, genetische Varianten zu finden, die eine Resistenz gegenüber zerebraler Malaria vermitteln. Da es sich bei zerebraler Malaria um eine seltene Erkrankung handelt, wurde ein Fall-Kontroll-Studiendesign gewählt. Die berücksichtigten Fälle wurden aus dem *Komfo Anokye Teaching Hospital* in Kumasi (Ghana) rekrutiert. Bei den Kontrollen handelt es sich um Stichproben aus der Bevölkerung im Einzugsgebiet des Krankenhauses.

2.1.2 Studienpopulation

Die Rekrutierung der Studienteilnehmer erfolgte im Zeitraum von 2001 bis 2005 in der Ashanti Region, einem meso- bis hyperendemischem Gebiet der Malaria mit ganzjährigen hohen Infektionsraten von *P. falciparum* sowie mäßigen Infektionsraten von *P. ovale* und *P. malariae* (Binka *et al.*, 1994; Kobbe *et al.*, 2006).

In die GWA-Studie zur zerebralen Malaria wurden im Zeitraum von 2001 bis 2005 568 Kinder im Alter von 2 Monaten bis 12 Jahren mit einer zerebralen Malaria nach den Kriterien der Weltgesundheitsorganisation (WHO) aus dem *Komfo Anokye Teaching Hospital* aufgenommen (WHO, 2000). Die Aufnahmekriterien der Studie waren ein Nachweis asexueller *P. falciparum*-Parasiten im Blut sowie ein *Blantyre Coma Score* von < 3 bei zwei Untersuchungen in 30 minütigem Abstand. Zusätzlich zu den Zeichen der zerebralen Malaria traten überwiegend folgende weitere Symptome von schwerer Malaria auf: Hyperparasitämie (> 200.000 Parasiten/ μ l), metabolische Azidose (Laktatwert > 5 mmol/L und/oder Blutbasenwerte < -5 mEq/L), Hypoglykämie (Blutzuckerwert < 2.2 mmol/L), respiratorische Insuffizienz sowie Erschöpfung (diagnostiziert nach altersbedingtem Unvermögen von körperlichen Fähigkeiten wie Sitzen, Essen, Stehen). Die Wahl der Kontroll-Individuen erfolgte unter Berücksichtigung einer Anpassung an den Altersaufbau, die Verteilung der Ethnien und das Geschlechterverhältnis der Individuen in der Fallgruppe. Die

insgesamt 3519 Studienteilnehmer, die nach Qualitätskontrollen (siehe 2.3.1.1) in der Studie verblieben, wurden in zwei Gruppen aufgeteilt. Die GWA-Studiengruppe zur initialen Suche nach Assoziation von SNPs mit zerebraler Malaria bestand aus 409 Fällen mit zerebraler Malaria und 828 Kontrollen (Tabelle 2). Die Fälle wurden so ausgewählt, dass sie neben den Symptomen einer zerebralen Malaria zusätzlich Zeichen einer Azidose wie Hyperlaktatämie, einer Atemnot oder einem Basendefizit aufwiesen. Die Replikationsgruppe zur Bestätigung der Assoziation der SNPs nach Analyse der GWA-Studiengruppe setzte sich aus 159 Fällen mit zerebraler Malaria und 2123 Kontrollen zusammen. Unter den Kontrollen der Replikationsgruppe befanden sich 1042 Individuen, die in die GWA-Studie aufgenommen wurden und 1081 Individuen aus weiteren Studien zur Malaria tropica aus der umliegenden Region (Kobbe *et al.*, 2007; Timmann *et al.*, 2007).

Tabelle 2: Demographische Charakteristika der Studiengruppen

	Alter (Monate)	Geschlecht (männl:weibl)	Ethnizität (%)				Anzahl
			Akan	Ga- Adangbe	Northerners	Ewe	
GWA-Studiengruppe							
Fälle	24 (2-120)	56:44	62	1	34	1	409
Kontrollen	20 (2-120)	54:45	68	1	28	2	828
Replikationsgruppe							
Fälle	28 (4-147)	54:46	66	-	30	3	159
Kontrollen	40 (3-161)	51:49	66	< 1	33	1	2123

Abgebildet ist der Median des Alters mit der Spannweite in Klammern.

In beiden Studiengruppen war die Ethnie Akan mit 62 – 68 % am stärksten vertreten (Tabelle 2). Die unter *Northerners* zusammengefasste Gruppe setzte sich aus Individuen verschiedener Ethnien zusammen, die nicht näher eingegrenzt werden konnten. Diese Gruppe von Ethnien war mit 28 – 34 % in den beiden Studiengruppen vertreten. Zwischen 1 und 3 % der Individuen gehörten der Ethnie Ewe an. Die Ethnie Ga-Adangbe war mit bis zu 1 % in den Studiengruppen vertreten.

Als Untersuchungsmaterial dienten mit EDTA versetzte Blutproben aus Routineuntersuchungen der Patienten. Den Kindern der Kontrollgruppen wurde entsprechend Blut entnommen.

Die Charakteristika der klinischen Untergruppen finden sich in Tabelle 3. Einverständniserklärungen der Eltern oder deren Betreuer lagen für alle teilnehmenden Kinder vor. Eine Genehmigung zur Durchführung der Studie wurde erteilt durch die Ethik-Kommission der *School of Medical Sciences, Kwame Nkrumah University of Sciences and Technology, Kumasi, Ghana*.

Tabelle 3: Angabe der klinischen Symptome der zerebralen Malaria-Fälle

	Symptom						
	Zerebrale Malaria	Schwere Anämie*	Hyperlaktämie*	Basendefizit*	Atemnot*	Hyperparasitämie*	Hypoglykämie*
GWA-Studiengruppe							
Fälle	409	129	262	203	183	179	66
Replikationsgruppe							
Fälle	159	22	53	23	38	45	13

*Zusätzlich zum klinischen Phänotyp einer zerebralen Malaria konnten sechs weitere klinische Phänotypen der schweren Malaria beobachtet werden. Die weiteren Phänotypen überschneiden sich mit zerebraler Malaria und teilweise auch untereinander.

2.2 Material

2.2.1 Geräte

ABI PRISM® 3100 Genetic Analyser	Applied Biosystems, Darmstadt
Agagel Mini Elektrophoresekammer	Biometra, Göttingen
T3 Thermocycler	Biometra, Göttingen
Transilluminator	Biorad, Hercules, USA
CERTOMAT®MO SchüttlerBraun	BioTech International, Göttingen
5415R Zentrifuge	Eppendorf AG, Hamburg
Biophotometer 6131	Eppendorf AG, Hamburg
Thermomixer 5436	Eppendorf AG, Hamburg
Sicherheitswerkbank HERAsafe HS15	Heraeus Instruments GmbH, Hanau
JK Ultra-Turrax T8-Homogenisierer	IKA®-Werke GmbH & Co. KG, Stau- fen
Wasserbad Julabo 5B	Julabu, Seelbach
Inkubator	Memmert, Schwalbach
BIO PAD Elektrophoresekammer	MWG-Biotech, Ebersberg
Innova™ 4400 Incubator Shaker	New Brunswick Scientific, USA
Thermocycler Primus 96 advanced	PEQLAB Biotechnologie GmbH, Er- langen
Pipetten-Set	Rainin, Giessen
Light Cycler® 480	Roche, Mannheim
3E Platten-Zentrifuge	Sigma, Osterode am Harz
4K10 Platten-Zentrifuge	Sigma, Osterode am Harz
Pipettierroboter Genesis Workstation 200	Tecan, Crailsheim

2.2.2 Chemikalien und Verbrauchsmaterial

Allgemein verwendete Chemikalien stammen von Merck (E. Merck, Darmstadt), Roth (Carl Roth GmbH, Karlsruhe) sowie Sigma (Sigma-Aldrich GmbH, Steinheim).

Verbrauchsmaterialien wurden von BD (Becton, Dickinson and Company, USA), Biozym (Biozym Scientific GmbH, Hessisch Oldendorf), Eppendorf (Eppendorf AG, Hamburg), Nunc (Nunc GmbH & Co KG, Wiesbaden) und Sarstedt (Sarstedt AG & Co, Nümbrecht), Applied Biosystems™ (Darmstadt) bezogen. PCR Reagenzien, sofern sie nicht als Reagenziensystem aufgeführt werden, stammen von Solis BioDyne (Estland). Aqua ad injectabilia wurde von Delta Select (Dreieich) bezogen.

Sofern erforderlich erfolgte eine Sterilfiltration von Lösungen mithilfe von Filtern einer Porengröße von 0,22 µm von Schleicher&Schuell (Dassel).

2.2.3 Puffer und Lösungen

Agarose-Gel-Auftragspuffer (10x)	0,25% (w/v) Bromphenolblau 0,25% (w/v) Xylencyanol 50% (v/v) Glycerin
Ethidiumbromid-Stammlösung	5 mg/ml Ethidiumbromid in ddH ₂ O
TBE (10x)	0,45 M Tris-Base 0,45 M Borsäure 10 mM EDTA, pH 8,0 autoklaviert
RNase away	Molecular BioProducts, USA

2.2.4 Oligonukleotide

Die verwendeten Oligonukleotide stammten von MWG-BIOTECH AG (Ebersberg) oder Biomers (Ulm/Donau). Ein Oligonukleotid zur Synthese von cDNA wurde von der Firma USB (USA, Cleveland) bezogen. Im Anhang findet sich eine Liste der verwendeten Oligonukleotide (siehe Anhang).

2.2.5 RNA aus humanem Gewebe

Humane Gesamt-RNA wurde von BioCat (Heidelberg) erworben oder aus Venengewebe isoliert, die von der herzchirurgischen Abteilung des Asklepiosklinikums Hamburg bereitgestellt wurde. Die pulmonale Endothelzelllinie HPMEC wurde von Prof. Dr. James Kirkpatrick des pathologischen Instituts der Universität Mainz zur Verfügung gestellt (Krump-Konvalinkova *et al.*, 2001). Im Anhang findet sich eine Liste der erworbenen RNAs.

2.2.6 cDNA aus humanem Gewebe

Erststrang-cDNA aus unterschiedlichen humanen Geweben wurden von BioCat (Heidelberg) sowie Primerdesign (Southampton, UK) bezogen. Eine cDNA Bibliothek aus mikrovaskulären primären Endothelzellen wurde von Aligent Technologies (Waldbronn, ehemals Stratagene) bezogen. Eine Übersicht über die verwendeten cDNAs findet sich im Anhang. Zudem wurde aus der isolierten RNA der Beinvene sowie aus der erworbenen RNA selbst cDNA synthetisiert.

2.2.7 cDNA-Klone

AI218168

Source Bioscience (ehemals
ImaGenes GmbH), Berlin

AA861168

Source Bioscience (ehemals
ImaGenes GmbH), Berlin

2.2.8 Vektoren

pCR2.1 Cloning Vector

Invitrogen, Darmstadt

2.2.9 Reagenziensysteme

BigDye® Terminator v3.1 Cycle

Applied Biosystems, Darmstadt

HiDi-Formamid Sequencing Kit

Applied Biosystems, Darmstadt

GeneRuler™ low

Fermentas, St. Leon-Roth

MSB® Spin PCRapace

Invitex, Berlin

Superscript® III First Strand Synthesis System
For RT-PCR

Invitrogen, Darmstadt

SuperScript® II/III Reverse Polymerase

Invitrogen, Darmstadt

Second Strand Buffer

T4 DNA Ligase

Invitrogen, Darmstadt

TA Cloning® Kit

Invitrogen, Darmstadt

Thermoscript™ Reverse Transcriptase

Invitrogen, Darmstadt

TOPO TA Cloning® Kit

Invitrogen, Darmstadt

TRIzol® Reagent

Invitrogen, Darmstadt

Sephadex G-50

Millipore GmbH, Schwal-
bach/Ts)

NucleoSpin® Gel and PCR Clean-up

Macherey und Nagel, Düren

peqGOLD DNA-Sizer XII

PeqLab, Erlangen

peqGOLD 100 bp DNA-Leiter

PeqLab, Erlangen

GoTaq® Flexi DNA Polymerase

Promega, USA

HotStarTaq® PCR

Qiagen GmbH, Hilden

RNeasy Mini Kit

Qiagen GmbH, Hilden

QIAprep® Spin Miniprep Kit

Qiagen GmbH, Hilden

QIAquick Gel Extraction Kit

Qiagen GmbH, Hilden

QIAquick PCR Purification Kit

Qiagen GmbH, Hilden

5'/3' RACE Kit, 2nd Generation

Roche, Mannheim

LightCycler® 480 SYBR Green I Master

Roche, Mannheim

2.2.10 Kultivierung von *E. coli*

2.2.10.1 Bakterienstämme

Es wurden chemisch kompetente Bakterienstämme mit dem TOPO TA Cloning® Kit von Invitrogen (Darmstadt) geliefert, die für die Transformation von *E. coli* verwendet wurden.

2.2.10.2 Medien

LB (*Luria Broth*) wurde in ddH₂O angesetzt. Das Medium wurde bei 120°C für 25 min unter einem Druck von 1 bar autoklaviert. Antibiotika wurden ab einer Temperatur von 55°C zum Medium gegeben.

LB-Medium 10 g/l Bacto-Trypton, 5 g/l Hefe-Extrakt, 10 g/l NaCl

Medien für die Agarplatte enthielten zusätzlich 15 g/l Bacto-Agar.

2.2.10.3 Antibiotika

Die verwendeten Antibiotika wurden von der Firma Roth (Kanamycinsulphat) und Sigma (Ampicillin) bezogen. Ampicillin wurde in einer Endkonzentration von 30 µg/ml angewendet, Kanamycin mit 50 µg/ml.

2.3 Methoden

2.3.1 Statistische Analysen

2.3.1.1 Qualitätskontrolle der genomweiten Assoziationsanalyse

2.3.1.1.1 Ausschluss von Studienteilnehmern

Es wurden diejenigen Individuen ausgeschlossen, deren Genotypisierungsrate unter 96 % und Heterozygotenrate außerhalb 26 bis 31 % lag. Ebenso wurden Individuen ausgeschlossen, denen eine biologische Verwandtschaft mit einem identischen Anteil der ererbten Anlagen von über 12,5 % zu anderen Teilnehmern der GWA-Studiengruppe nachgewiesen werden konnte. Die Einzelnukleotidaustausche (SNPs) rs334 (HbS) und rs33930165 (HbC) wurden in allen verbleibenden Studienteilnehmern genotypisiert. Die Sichelzellerkrankung verursacht neben Anämie auch eine erhöhte Anfälligkeit für Sepsis. Da Anämie und Sepsis auch zu Bewusstlosigkeit führen können und dies als Phänokopie einer zerebralen Malaria angesehen werden kann, wurden Individuen mit einem HbSS- und HbSC-Genotyp nicht in die Assoziationsanalyse (siehe 2.3.1.2) eingeschlossen.

2.3.1.1.2 Ausschluss von SNPs

Von der Assoziationsanalyse wurden SNPs der X und Y Chromosomen sowie mitochondriale und pseudoautosomale SNPs ausgeschlossen. Autosomale SNPs mit einer Minor-Allelfrequenz von < 1 % und einer Genotypisierungsrate von < 96 % wurden ebenfalls nicht für eine Analyse verwendet. Die Plots von Signalintensitäten der Genotypen der genomweiten SNP-Typisierung, die so genannten *cluster plots*, wurden in den Fällen visuell auf ihre Qualität geprüft, Genotypen zweifelsfrei zu identifizieren, wenn SNPs ein Signifikanzniveau von $p = 10^{-4}$ unterschritten. SNPs mit mehrdeutigen Ergebnissen wurden ebenfalls ausgeschlossen. Darüber hinaus wurden SNPs, die eine Abweichung vom Hardy-Weinberg-Gleichgewicht von $p < 10^{-7}$ aufwiesen, nicht in die Assoziationsberechnungen eingeschlossen. Nach diesen Kriterien der Qualitätskontrolle blieben 774800 SNPs für die Assoziationsanalyse.

Abschließend wurden die berechneten 774800 p-Werte der Assoziationsanalyse in einem so genannten Quantile-Quantile-Plot aufgetragen um einen Hinweis darauf zu erhalten, ob in der Studie mehr signifikante Ergebnisse erhalten wurden als zu erwarten wären. Dabei wurden die beobachteten p-Werte der Reihe nach vom kleinsten zum höchsten aufgelistet und der theoretischen Verteilung der p-Werte gegenübergestellt (Abbildung 4). Unter einer vollständigen Null-Verteilung beträgt der erwartete Inflationsfaktor $\lambda = 1$. Der Inflationsfaktor λ betrug nach Adjustierung mit den Variablen der Multidimensionalen Skalierung (MDS) 1,012. Ein erhöhtes Auftreten von falsch-positiven Assoziationen durch Populationsstratifikation (siehe 2.3.1.3) oder Genotypisierungsfehler gilt daher als unwahrscheinlich.

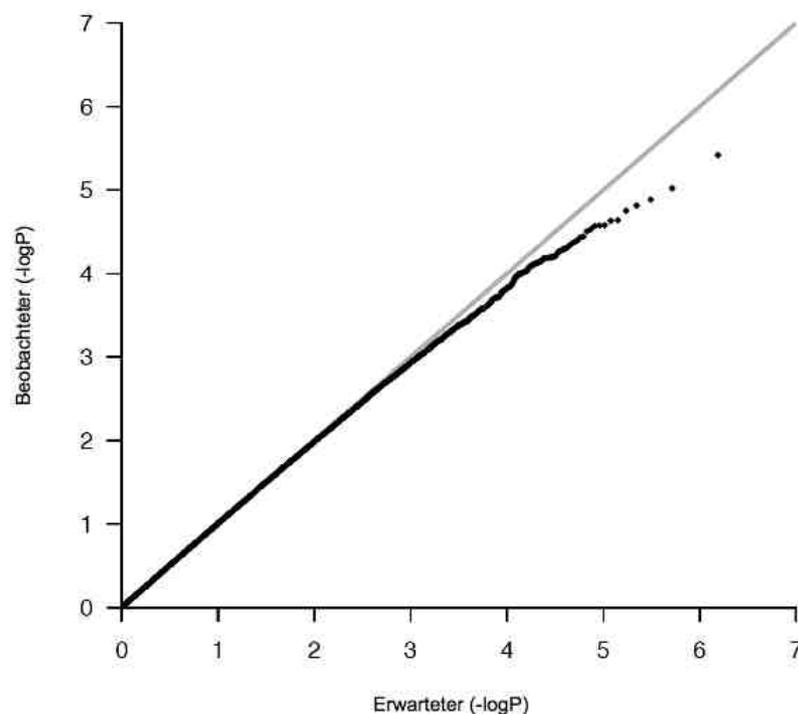


Abbildung 4: Quantile-Quantile-Plot der p-Werte des additiven Modells der GWA-Studiengruppe

2.3.1.2 Assoziationsanalyse

Für die statistische Analyse wurde die Software PLINK v1.07 verwendet (Purcell *et al.*, 2007). Die Assoziationsberechnungen erfolgten mit einem multivariaten

logistischen Regressionsmodell. Dadurch konnten mehrere Faktoren (Kovariaten) berücksichtigt werden, die das Krankheitsrisiko einer zerebralen Malaria beeinflussen und zu Verzerrungen führen. SNPs, die nach der Genotypisierung in der GWA-Studiengruppe eine Assoziation mit einem Trend in Richtung genomweiter Signifikanz mit p-Werten $\leq 5 \times 10^{-5}$ im additiven Vererbungsmodell aufwiesen, wurden anschließend in der Replikationsgruppe überprüft. Um auch andere Vererbungsmodi als das additive Modell zu testen, wurden Assoziationssanalysen unter Annahme eines dominanten oder rezessiven Vererbungsmodells für den neu gefundenen Locus durchgeführt.

Zunächst wurden eine Nullhypothese H_0 („*Es besteht keine Assoziation zwischen dem Krankheitsbild und den genetischen Varianten*“) und eine Alternativhypothese H_1 („*Es besteht eine Assoziation zwischen dem Krankheitsbild und genetischen Varianten*“) festgelegt. In der Medizin hat sich ein Signifikanzniveau von $p = 0,05$ etabliert. Liegt der berechnete p-Wert unterhalb dieses Werts, wird die Nullhypothese abgelehnt, und es gilt die Alternativhypothese H_1 . Da in dieser Studie nahezu 800000 SNPs getestet wurden (multiples Testen), steigt die Wahrscheinlichkeit für das Auftreten von falsch-positiven Ergebnissen. Mit der Bonferroni-Korrektur ist es jedoch möglich, dieser Fehlerwahrscheinlichkeit entgegenzuwirken, indem man das Signifikanzniveau α durch die Anzahl der durchgeführten Tests dividiert. Für genomweite Assoziationsstudien (GWA-Studien) hat sich ein Signifikanzlevel von 5×10^{-8} durchgesetzt, für den zufällige Assoziationen als hinreichend unwahrscheinlich gelten (Pe'er *et al.*, 2008). Dieses berechnet sich aus dem Signifikanzniveau α , dividiert durch die in der Assoziationsstudie verwendeten unabhängigen SNPs (Hardy & Singleton, 2009; Smith & Newton-Cheh, 2009). Unterschreiten p-Werte einzelner SNPs diesen Wert, wird Signifikanz für eine genomweite Studie angenommen.

Die *Odds Ratio* (OR) ist in der Epidemiologie ein Assoziationsmaß zwischen Risikofaktoren und einer Erkrankung. Liegt der Wert bei 1, gibt es keine Assoziation zwischen beiden Komponenten. Ein Wert unter 1, weist auf einen Schutz, ein Wert über 1 auf eine Empfänglichkeit gegenüber der Erkrankung hin. Die OR erfolgt unter Angabe des 95 % Konfidenzintervalls und wird meist für das Minor-Allel angegeben.

Das relative Risiko beschreibt die Wahrscheinlichkeit, um welche sich ein Merkmal (von zwei Alternativmerkmalen) in zwei Gruppen unterscheidet. Dazu wird der Quotient der Wahrscheinlichkeiten eines Merkmals errechnet.

Das Kopplungsungleichgewicht (*linkage disequilibrium*, LD) zwischen SNPs kann mit dem Korrelations-Koeffizienten r^2 und D' beschrieben werden. Zwei Marker stehen in einem vollständigen LD, wenn $D' = 1$ ergibt. r^2 gibt die Korrelation der Allele der untersuchten Varianten an und ermöglicht bei geringen Allelfrequenzen eine verlässlichere Aussage über das LD zwischen SNPs als D' . Ein Korrelations-Koeffizient von $r^2 = 1$ entspricht einem vollständigen LD. Der Korrelations-Koeffizient r^2 kann dafür verwendet werden, *Tagging* SNPs zu definieren.

Die Analysen zur Kopplung der Varianten wurden mit dem Programm Haploview 12.9.0 berechnet (Barrett *et al.*, 2005).

2.3.1.3 Populationsstratifikation

Treten in Untergruppen einer ethnisch gemischten Population unterschiedliche Allel- und Krankheitshäufigkeiten auf, können falsch-positive Assoziationen bestimmt werden. Sie treten dann auf, wenn sich die Anteile zwischen Fällen und Kontrollen unterscheiden (Lander & Schork, 1994; McCarthy *et al.*, 2008; Pearson & Manolio, 2008). Das Verfahren der multidimensionalen Skalierung (MDS) wurde als biomathematisch-genetisches Verfahren zur Bestimmung unterschiedlicher ethnischer Gruppen in der vorliegenden Studienpopulation verwendet. Die mittels der MDS-Transformation abgeleiteten Variablen der 1. bis 3. Ordnung dienten in dieser Studie zugleich zur Korrektur der Assoziationsberechnungen. Grafisch dargestellt werden die Daten mit der Software R (<http://www.R-project.org>).

2.3.1.4 Powerberechnung

Die statistische *Power* drückt die Wahrscheinlichkeit aus, mit der ein statistischer Test die Alternativhypothese H_1 erkennt. Als Variable spielen unter anderem die Fallzahlen, das Signifikanzniveau sowie die Inzidenz der Erkrankung eine Rolle für die Berechnung. Die Inzidenz der zerebralen Malaria in subsaharischen Gebieten beträgt 1 - 12 Fälle/1000 Kinder/Jahr (Idro *et al.*, 2010). Die Bestimmung der statistischen *Power* dieser GWA-Studie erfolgte mit der Software CaTS (Skol *et al.*,

2006). Diese ermöglichte mit einer Wahrscheinlichkeit von 80 % den Nachweis von Assoziationen von SNPs mit einer Frequenz des risikotragenden Allels von 5 % und einem relativen Risiko von 1,67 mit der verwendeten Studienteilnehmerzahl unter Annahme des additiven Vererbungsmodells.

2.3.1.5 Transmissions-Disequilibrium-Test

Für die Berechnungen des Transmissions-Disequilibriumstests wurde das Programm UNPHASED v3.1.3 verwendet. Das Programm ermöglicht, neben den Assoziationsberechnungen mit vollständigen Trios (Eltern und erkranktes Kind) auch Berechnungen mit Duos (ein Elternteil und erkranktes Kind), indem die Genotypen der fehlenden Eltern statistisch aus der Population der vorhandenen Eltern geschätzt werden.

2.3.2 Molekularbiologische Methoden

2.3.2.1 Polymerase-Kettenreaktion (PCR)

Die Amplifikation von DNA-Sequenzen erfolgte mittels Polymerase-Kettenreaktionen (*polymerase chain reaction*, PCR). Die verwendeten Oligonukleotide wurden zu Stocklösungen mit einer 100 pmol/µl Konzentration angesetzt und bei -20°C gelagert. Die Gebrauchslösung bestand aus 1:10-Verdünnungen der Stocklösung. Es wurde mit der FirePol® Polymerase sowie mit der HotFirePol® (SOLIS) und der AmpliTaq Gold® 360 DNA Polymerase (Invitrogen) gearbeitet und Variationen anhand der Magnesiumkonzentration (1,5 – 3 mM) im Reaktionsansatz durchgeführt. Je nach verwendeter Polymerase und Matrize wurden dem Reaktionskit beigefügte PCR-Zusätze wie Q-Solution oder Solution S hinzugegeben. Diese Zusätze ermöglichen eine Veränderung der Schmelztemperatur der Matrize und erleichtern die Amplifikation schwieriger Ausgangsmaterialien. Außerdem wurden verschiedene Temperaturen zur Oligonukleotid-Hybridisierung (50 - 62°C) getestet sowie die Temperatur zur Polymerisierung zwischen 68°C und 72°C variiert.

Der Standardansatz einer PCR mit der FirePol® Polymerase war im Allgemeinen wie in Tabelle 4 dargestellt zusammengesetzt:

Tabelle 4: Standard-PCR-Ansatz und -Programm mit der Firepol® Polymerase (Solis)

	<u>finale Konzentration</u>			
Polymerase-Puffer BD [10x]	1x			
Solution S [10x]	1x	95°C	5 min	
MgCl ₂ [25mM]	1,5 – 3,0 mM	95°C	30 sec	
Oligonukleotid vorwärts [10 pmol/μl]	0,5 μM	X°C	30 sec	45 Zyklen
Oligonukleotid rückwärts [10 pmol/μl]	0,5 μM	72°C	1 min	
dNTPs [je 10 mM]	200 μM	72°C	10 min	
FIREPol® Polymerase [5 U/μl]	1 U	4°C		
DNA-Matrize	x μl			
ddH ₂ O	x μl			
Gesamtvolumen	20 μl			

Der Standardansatz einer PCR mit der AmpliTaq Gold® 360 DNA Polymerase war im Allgemeinen wie in Tabelle 5 dargestellt zusammengesetzt:

Tabelle 5: Standard-PCR-Ansatz und -Programm mit der AmpliTaq Gold® 360 DNA Polymerase (Invitrogen)

	<u>finale Konzentration</u>			
Polymerase-Puffer [10x]	1x			
Q Solution [5x]	1x	95°C	5 min	
MgCl ₂ [25mM]	1,5 – 3,0 mM	95°C	30 sec	
Oligonukleotid vorwärts [10 pmol/μl]	0,5 μM	X°C	30 sec	45 Zyklen
Oligonukleotid rückwärts [10 pmol/μl]	0,5 μM	72°C	1 min	
dNTPs [je 10 mM]	200 μM	72°C	10 min	
AmpliTaq Gold® Polymerase [5 U/μl]	1 U	4°C		
DNA-Matrize	x μl			
ddH ₂ O	x μl			
Gesamtvolumen	20 μl			

2.3.2.2 Genotypisierung von genetischen Polymorphismen

2.3.2.2.1 *Genome-Wide Human SNP Array 6.0* von Affymetrix®

Die in dieser Studie untersuchten Fälle und Kontrollen wurden im Rahmen einer genomweiten Assoziationsstudie (GWA-Studie) zu *P. falciparum*-Malaria mit dem *Genome-Wide Human SNP Array 6.0* von der Firma Affymetrix® in den USA genotypisiert. Diese Genotypisierungsplattform von Affymetrix® ermöglicht die Genotypisierung von mehr als 906600 autosomalen, gonosomalen und mitochondrialen SNPs. Die Genotypen der untersuchten SNPs der verwendeten Fälle und Kontrollen lagen für das GWA-Studienkollektiv vor.

2.3.2.2.2 SNPLex® Technologie

Das SNPLex® *Genotyping System* erlaubte die simultane Genotypisierung von 48 SNPs innerhalb einer biologischen Probe (Tobler *et al.*, 2005). Es erfolgte eine Genotypisierung ausgewählter SNPs in der Replikationsgruppe mithilfe dieser Methode. Die Genotypisierungen wurden vom Institut für Klinische Molekularbiologie der Universität Kiel durchgeführt.

2.3.2.2.3 *Dynamic allele-specific hybridization (DASH)*

Eine weitere Methode zur Genotypisierung von SNPs ermöglicht DASH (Howell *et al.*, 1999). Dabei wird sich die durch ein Auftreten von SNPs hervorgerufene unterschiedliche Hybridisierungstemperatur von DNA-Strängen zu Nutzen gemacht. Die Zielsequenz wurde unter Anwesenheit von zwei Hybridisierungssonden (Anker und Sensor) amplifiziert. Die Sonden hybridisierten innerhalb des Amplikons in der Nähe des zu typisierenden SNPs. Der Sensor hybridisierte an die Sequenz um die zu charakterisierende genetische Variante. Zwischen Anker und Sensor, die mit unterschiedlichen Fluophoren markiert waren, erfolgte ein Energietransfer, dessen Fluoreszenz messbar war, wenn beide Sonden simultan hybridisieren. Befand sich an der Position der Variante ein *mismatch* durch einen Nukleotidaustausch, war die Bindungsaffinität von Sensor zu Zielsequenz geringer als bei vollständiger Komplementärsequenz. Die genetischen Polymorphismen ließen sich durch ihre unterschiedlichen Hybridisierungstemperaturen unterscheiden; der abweichende

SNP besaß die geringere Hybridisierungstemperatur. Dieses Ereignis ließ sich über den Abfall der Fluoreszenzintensität durch die Dissoziation des Sensors während einer stetigen Temperaturerhöhung im *LightCycler* (Schmelzkurvenanalyse) messen. Mit der Software des *LightTypers* (Roche Diagnostics, Mannheim) wurden die Genotypen der SNPs bestimmt. Ausgangsprodukt dieser Methode waren 2 µl einer 1:200 Verdünnung von zuvor über das *GenomiPhi HY DNA-Amplification Kit* (GE Healthcare, Braunschweig) genomweit amplifizierte DNA des gesamten Studienkollektivs. Der obligate PCR-Ansatz sowie die obligaten PCR-Bedingungen der Genotypisierung finden sich in Tabelle 6.

Tabelle 6: PCR-Ansatz und -Programm und der FRET-Analyse

	<u>Volumen</u>			
Polymerase-Puffer BD/B [10x]	1 µl			
Solution S [10x]	0-1 µl			
MgCl ₂ [25mM]	0,4 – 1,2 µl	95°C	3 min	
Oligonukleotid vorwärts [10 mol/µl]	0,2 – 1,0 µl	95°C	1 min	
Oligonukleotid rückwärts [10 mol/µl]	0,2 – 1,0 µl	X°C	1 min	45 Zyklen
Anker [5 pmol/µl]	0,4 µl	72°C	1 min	
Sensor [5 pmol/µl]	0,4 µl	72°C	10 min	
dNTPs [je 10 mM]	200 µl	4°C		
DMSO	0-0,5 µl			
FIREPol® Polymerase [5 U/µl]	0,2 µl			
Matrizen-DNA	2 µl			
ddH ₂ O	ad 10,0 µl			

2.3.2.3 RNA-Isolierung aus humanen Venen

Die Durchführung der RNA-Isolierung aus vaskulärem Gewebe erfolgte nach einem modifizierten Protokoll von Townsend *et al.* (Townsend *et al.*, 1999). Die Pulverisierung der humanen Beinvenen wurde mit Mörser und Stößel unter RNase-freien Konditionen und Zugabe von flüssigem Stickstoff erzielt. Eine Zugabe von TRIzol® Reagenz diente einem zusätzlichen Schutz vor RNA-Degradierung. Die Menge des Reagenzes richtete sich nach dem Gewicht des aufzuarbeitenden Gewebes, welches im Vorfeld bestimmt wurde. Eine Hälfte des erforderlichen TRIzol® Reagenzes wurde während des Pulverisierungsprozesses beigefügt. Die zweite Hälfte wurde dem Gemisch nach der Pulverisierung zugefügt. Die gefrorene

Flüssigkeit wurde anschließend bei Raumtemperatur aufgetaut. Ein Homogenisierer ermöglichte die vollständige Lösung des Gemischs. Das Homogenat wurde anschließend mit 1/5 seines Volumens an Chloroform versetzt und 15 - 30 sec stark gemischt. Anschließend erfolgte eine Inkubation auf Eis für 30 min mit einer nachfolgenden 30-minütigen Zentrifugation bei 4°C und 10000 x g. Die wässrige RNA-haltige Phase wurde separiert und ein äquivalentes Volumen Isopropanol zur Fällung hinzugegeben. Das Gemisch wurde bei -20°C für 30 min inkubiert und im Anschluss für 30 min bei 4°C und 10000 x g zentrifugiert. Der Niederschlag wurde mit 70 % Ethanol gewaschen, und erneut bei 4°C für 5 min und 10000 x g zentrifugiert. Die RNA wurde für 10 min bei 56°C luftgetrocknet und in 100 µl Wasser gelöst. Anschließend erfolgte eine Aufreinigung nach Herstellerangaben mit dem *RNeasy Mini Kit* (Qiagen). Einbezogen in diesen Arbeitsschritt war ein 15-minütiger DNase-Verdau zur Eliminierung einer etwaigen Kontamination mit genomischer DNA (gDNA).

Die RNA-Konzentration der Ausbeute wurde in 50-fach RNase-freiem Wasser mit einem Bio-Photometer (Eppendorf) bei 260 nm gemessen.

2.3.2.4 Reverse Transkription

Die cDNA wurde nach Angaben des Herstellers mit dem *SuperScript II/III First-Strand Synthesis System for RT-PCR* (Invitrogen) generiert. Hierfür dienten Oligo(dT)₁₈₋₂₂ (Invitrogen) als Oligonukleotide. Es wurden optimalerweise 1 µg RNA für die Reverse Transkription verwendet. Stets diente ein Ansatz ohne Matrizen-RNA als Kontrolle. Inkubiert wurde für 1 Stunde bei 45°C.

Für schwierige Sekundärstruktur-bildende oder GC-reiche RNA-Sequenzen wurde die *Thermoscript™ Reverse Transcriptase* (Invitrogen) verwendet. Mit dieser Reversen Transkriptase ist es möglich, die cDNA bei bis zu 65°C zu synthetisieren.

2.3.2.5 Quantitative *Real-Time* PCR

Die quantitative *Real-Time* PCR (qRT-PCR) erlaubt es, die Amplifikation während einer PCR in Echtzeit zu quantifizieren. Der interkalierende Fluoreszenz-Farbstoff *SYBR® Green I* bindet an doppelsträngige DNA und ermöglicht über die proportional steigende Emission eine Detektion der entstehenden Amplifikationsprodukte.

Ein Schwellenwert (*cycle of threshold*, Ct-Wert) sowie die PCR-Effizienz wurden hierfür als Auswertungsmaß verwendet. Der Ct-Wert determiniert den Zyklus, in dem die Fluoreszenz des PCR-Produktes signifikant über dem des Hintergrunds liegt. In diesem Fall liegt in allen Proben eine äquivalente Menge neu synthetisierter DNA vor (Pfaffl, 2004). Die PCR-Effizienz gibt die Amplifikationseffizienz in einem PCR-Zyklus an und beträgt optimalerweise 2 (Verdopplung des Ausgangsstranges). Die Effizienzen sowie die Ct-Werte des Referenz- und des Zielgens wurden durch die Software *LightCycler® 480 Instrument – Software Version 1.5* (Roche) automatisch bestimmt und anschließend manuell miteinander korreliert (Muller *et al.*, 2002):

$$NE = \frac{E_{\text{Referenz}}^{Ct_{\text{Referenz}}}}{E_{\text{Zielgen}}^{Ct_{\text{Zielgen}}}}$$

NE = normalisierte Expression

E = Effizienz

Ct = *Crossing Point*

Als Referenzgene dienten zwei konstitutiv exprimierte Gene: *Beta Aktin (ACTB)* und das Gen der katalytischen α -Untereinheit der *Protein Phosphatase 1 (PPP1CA)* (Bernard *et al.*, 1996; Dolter & Braman, 2001; Hsiao *et al.*, 2001). *ACTB* und *PPP1CA* werden in Zellen unterschiedlich stark exprimiert: *ACTB* gehört zu den stark transkribierten Genen, *PPP1CA* dagegen zu den wenig transkribierten Genen. Das Oligonukleotidpaar zur Amplifikation des Referenzgens *ACTB* stammte aus einer Publikation (Galiveti *et al.*, 2009). Es hybridisiert mit dem Vorwärts-Oligonukleotid an einer Spleißstelle der mRNA und kann somit nur Amplifikate des Gens anhand von *complementary DNA* (cDNA), nicht aber anhand von gDNA generieren.

Angesetzt wurde die PCR nach Angaben des Herstellers mit dem *LightCycler® 480 SYBR Green I Master-Kit* (Roche) und durchgeführt auf einem *LightCycler 480® Instrument* (Roche) unter folgenden PCR-Bedingungen:

Tabelle 7: Ansatz und Programm der quantitativen *Real-Time* PCR für den LightCycler 480® Instrument

SYBRgreen I Master, 2x	5 µl	96°C	5 min	
Oligonukleotid vorwärts [10 pmol]	1 µl	96°C	20 sec	
Oligonukleotid rückwärts [10 pmol]	1 µl	58°C	20 sec	45 Zyklen
H ₂ O	x µl	72°C	30 sec	
cDNA-Matrize	x µl			
Gesamtvolumen	20 µl	4°C		

Die Auswertung erfolgte mit der oben angeführten *Software*. Um Pipettierfehler zu minimieren, wurden mindestens Duplex-Ansätze verwendet.

Zur Optimierung der quantitativen Auswertung wurde zudem eine Schmelzkurvenanalyse der Amplifikate durchgeführt. Hierbei wird nachfolgend an die PCR eine Temperaturerhöhung vollzogen. Bei der Trennung der DNA-Stränge fällt das Fluoreszenzsignal ab, da der Fluoreszenzfarbstoff *SYBRgreen* von Doppelstrang dissoziiert. Die Produkte derselben Fragmente weisen bei Dissoziation einen identischen Schmelzpunkt auf. So kann sicher gegangen werden, dass es sich bei dem zu analysierten Produkt um das richtige handelt.

Für die Etablierung der quantitativen *Real-Time* PCR wurden zunächst die Effizienzen der verwendeten Oligonukleotidpaare bestimmt (Tabelle 8).

Tabelle 8: Effizienzen der Oligonukleotidpaare der quantitativen *Real-Time* PCR

	Exon	Effizienz	Matrize
ACTB	4/5	1,88 (± 0,06)	cDNA
PPP1CA	6/7	1,97 (± 0,03)	cDNA
AI/AW	1	1,84 (± 0,06)	gDNA
BX	2	1,92 (± 0,01)	gDNA
AA	ungespleißt	1,77 (± 0,03)	gDNA
DA	2	2,00 (± 0,03)	gDNA
DA	6	1,97 (± 0,01)	gDNA

ACTB, Beta Aktin; *PPP1CA*, Protein Phosphatase 1; AI/AW, AI218168/AW66536; BX, BX091678; DA, DA227815; cDNA, *complementary* DNA; gDNA, *genomic* DNA

Dafür wurde für jedes untersuchte Exon der Transkripte je eine Verdünnungsreihe in den Stufen 100 ng/µl, 10 ng/µl, 1 ng/µl und 100 pg/µl im Triplex-Ansatz angesetzt. Die Effizienzbestimmung der Referenzgene erfolgte aufgrund der Oligonukleotid-Bindungsstellen an Sequenzen unterschiedlicher Exons anhand von cDNA. Für die Transkripte wurden die Effizienzbestimmungen an gDNA vorgenommen. Die Berechnungen zu den Oligonukleotid-Effizienzen erfolgten mit der *LightCycler® 480 Instrument – Software*.

2.3.2.6 DNA-Sequenzierung

Sequenzen der PCR-Produkte wurden mit der Kettenabbruchmethode nach Sanger gewonnen (Sanger *et al.*, 1977). Dazu wurde 1 µl des PCR-Produktes mit dem *BigDye™ Terminator Cycle Sequencing Ready Reaction Kit* amplifiziert. Anschließend an die *BigDye*-Reaktion erfolgte eine Aufreinigung des Amplifikats mittels Sephadex G-50 (Millipore GmbH, Schalbach/Ts), um unerwünschte Oligonukleotide aus dem Reaktionsansatz zu entfernen. Das Produkt wurde in HiDi-Formamid (Applied Biosystems) gegeben und im 3100 *Genetic Analyser* (Applied Biosystems) kapillarelektrophoretisch aufgetrennt.

Tabelle 9: Reaktionsansatz und PCR-Programm der Sequenzierung nach Sanger

	Volumen			
<i>BigDye</i> Reaktionsmix	2 µl	96°C	5 min	
PCR-Puffer	2 µl	96°C	30 sec	
Oligonukleotid vorwärts/rückwärts [10 pmol]	1 µl	50°C	15 sec	25 Zyklen
DNA-Matrize	1 µl	60°C	4 min	
ddH ₂ O	4 µl	4°C		

Die DNA-Sequenzen wurden mit Hilfe der *SeqScape® Software Version 2.5* (Applied Biosystems) gegen genomische DNA (gDNA) oder gegen in Online-Datenbanken hinterlegten Sequenzen des Humangenoms des NCBI oder UCSC (siehe 2.3.4.2) abgeglichen.

2.3.2.7 *Fine-Mapping*

Das *Fine Mapping* begrenzte sich auf einen Bereich von 100 kb um eine Variante mit der stärksten Assoziation. Da es sich bei diesem Abschnitt um eine hochrepetitive Region handelt, war es notwendig, den Bereich in einer ersten PCR (*outer PCR*) zunächst in etwa 3000 bp lange DNA-Fragmente aufzuteilen. Dessen Amplifikationsprodukt diente als Matrize für die zweite *BigDye*-Reaktion (*inner PCR*) die anschließend wie unter 2.3.2.6 beschrieben weiter behandelt wurde.

2.3.2.8 *Rapid Amplification of cDNA-Ends (RACE)*

Die *Rapid Amplification of cDNA Ends (RACE)* ist ein PCR-basiertes Verfahren zur Amplifikation von cDNA-Enden. Da bei dieser Art PCR eine Vielzahl von unspezifischen PCR-Produkten entsteht und der zu untersuchende Bereich einen hohen Anteil an repetitiven Sequenzen aufweist, wurde eine *HotStarTaq* DNA Polymerase (Qiagen, Hilden) verwendet. Diese vermindert durch eine initiale Aktivierung bei 94°C für 15 min eine Amplifikation von unspezifischen PCR-Produkten und Oligonukleotid-Dimeren.

Für die RACE war es nötig, zwischen verschiedenen DNA-Polymerasen und Zusammensetzungen der Reaktionsbedingungen zu variieren, da die verschiedenen Matrizen mit ihren unterschiedlichen Oligonukleotiden jeweils eigene Amplifikationsbedingungen benötigten (siehe Anhang). Die Zusammensetzung der verwendeten Reaktionsansätze finden sich in Tabelle 4 und 5.

2.3.2.8.1 *5'-Rapid Amplification of cDNA Ends (5'-RACE)*

Die 5'-RACE erfolgte entsprechend einem modifizierten Protokoll nach Schramm *et al.* (Schramm *et al.*, 2000). Das Prinzip der 5'-RACE ist in Abbildung 5 dargestellt. Voraussetzung für diese Methode ist eine bekannte Teilsequenz des Transkriptes. Diese dient als Matrize, indem ein komplementäres genspezifisches Oligonukleotid 1 (GSP1) an die Matrize hybridisiert und mithilfe der Reversen Transkriptase ein cDNA-Erststrang in Richtung des 5'-Endes synthetisiert wird. Ist das Ende des Transkriptes erreicht, fügt die Reverse Transkriptase drei bis fünf Cytosine an das 3'-Ende des cDNA-Erststranges.

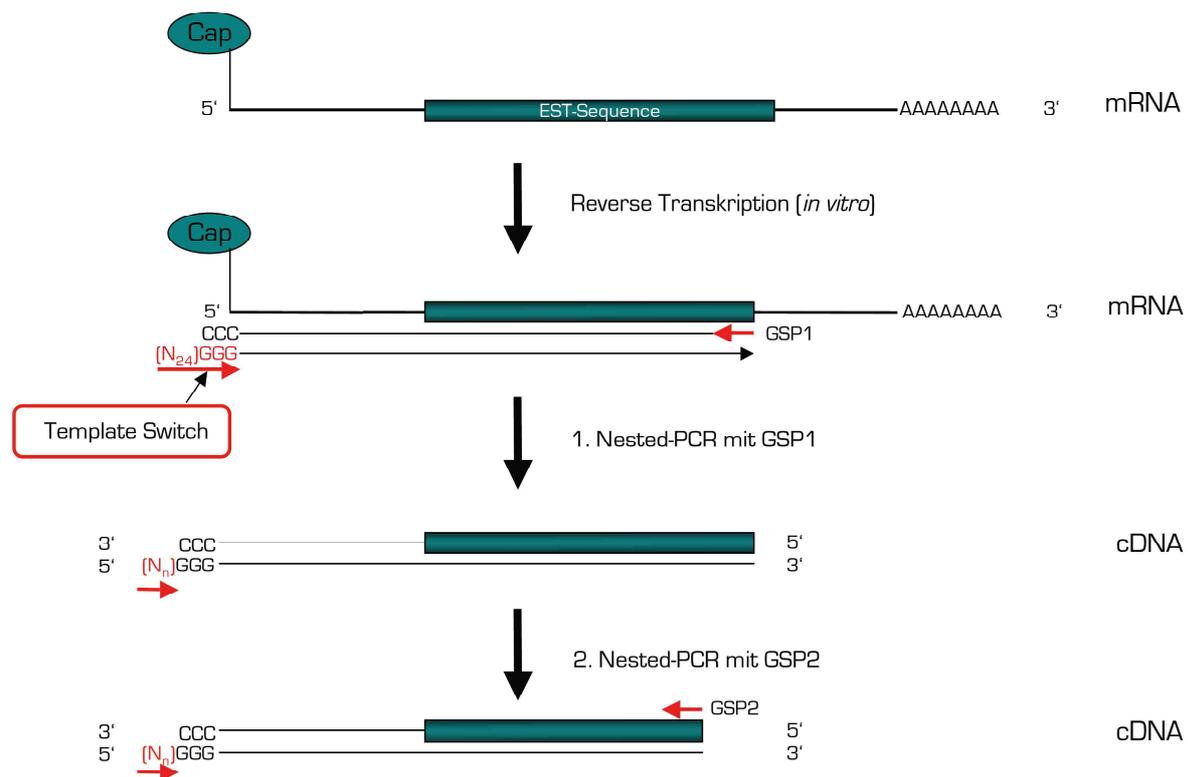


Abbildung 5: Schematische Abbildung der 5'-RACE mit der CapFinder-Methode

Anhand einer bekannten Sequenz wird mit einem GSP1 ein cDNA-Erststrang der RNA-Matrize synthetisiert. Die Reverse Transkriptase fügt 3-5 Cytosine an das 5'-Ende des neu synthetisierten Strangs an, an deren Sequenz das 3'-Ende des CapFinder-Oligonukleotids hybridisiert. Entlang dieses neuen Oligonukleotids schreibt die Reverse Transkriptase nach einem *template switch* den cDNA-Zweitstrang, so dass am Ende ein cDNA-Doppelstrang vorliegt. Anhand dieser CapFinder-cDNA kann mit weiteren GSP und dem komplementären CapFinder-Oligonukleotid eine 5'-RACE vollzogen werden bis das gewünschte Produkt gewonnen wird. Cap, 5'-Capstruktur des Transkripts; EST, *Expressed Sequence Tag*; mRNA, *messenger RNA*; GSP, genspezifisches Oligonukleotid; N, Anzahl angefügter Basen; C, Cytosin; G, Guanin; A, Adenin; cDNA, *complementary DNA*; RACE, *Rapid Amplification of cDNA Ends*, modifiziert nach Schramm *et al.*, 2000.

Diese Cytosine dienen als Bindungsstelle für ein generiertes Oligonukleotid (*CapFinder*-Oligonukleotid) mit einer Komplementärsequenz zu diesen Tri- bis Pentanukleotiden, die dann um die Sequenz des Oligonukleotids bis zu einer Länge von 30 bp verlängert werden. Diese angefügte Sequenz dient bei nachfolgenden PCRs als Hybridisierungsstelle für ein weiteres Oligonukleotid (*CapFinder B*). Sobald das *CapFinder*-Oligonukleotid an den cDNA-Erststrang hybridisiert hat, vollzieht die Reverse Transkriptase einen sogenannten *Template-Switch* und schreibt entlang des Erststrangs den Zweitstrang. Am Ende der cDNA-Synthese liegt ein cDNA-Doppelstrang von der bekannten Sequenz des Transkripts bis zum 5'-Ende vor. Dieses Produkt wurde für eine *semi-nested* PCR mit einem genspezifischen

Oligonukleotid 2 (GSP2) und dem *CapFinder* B-Oligonukleotid eingesetzt. Diese PCR erfolgte für 20 Zyklen. Mit einem genspezifischen Oligonukleotid 3 (GSP3) und dem *CapFinder* B-Oligonukleotid wurde aus 1 µl diesem PCR-Ansatz eine weitere *semi-nested* PCR für 30 Zyklen angesetzt. Konnte nach dieser PCR kein Produkt über die Gelelektrophorese detektiert werden, wurde die PCR um weitere 10 Zyklen erweitert. Sollten mehr als eine Bande entstanden sein, wurde 0,5 - 1 µl des Amplifikats in eine PCR mit einem weiteren GSP eingesetzt. Waren so genannte Schmierbanden auf dem Gel zu erkennen, wurde die PCR-Lösung verdünnt und mit 1 µl der verdünnten Lösung eine weitere *semi-nested* PCR vollzogen. Je nach Intensität der Schmierbanden wurden Verdünnungen von 1:50 bis 1:500 angesetzt. Für Transkripte, bei denen mit der *CapFinder*-Methode keine 5'-Enden detektiert werden konnten, wurde ein 5'/3'-RACE-Kit (Roche, Mannheim) herangezogen. Hierbei wurde zunächst ein Einzelstrang mit einem GSP synthetisiert und im Anschluss mit einer Terminalen Transferase eine oligo(dT)-Sequenz am 5'-Ende angefügt. In der ersten *semi-nested* PCR mit dieser Matrize wurde ein Oligonukleotid verwendet, das über eine komplementäre Sequenz zum oligo(dT)-Nukleotid verfügt, am 5'-Ende jedoch eine angefügte Oligonukleotidsequenz besitzt. Diese Sequenz wurde als Bindungsstelle für weitere PCRs verwendet, um eine gewisse Spezifität zu erhalten. Man behilft sich am 5'-Ende mit der oligo(dT)-Sequenz, da in kodierenden und 5'-Regionen von Säugern eine Neigung zu GC-reichen Sequenzen besteht. Die Wahrscheinlichkeit einer Hybridisierung des oligo(dT)-Nukleotids an adeninreiche Sequenzen im 3'-Bereich eines Transkripts wird dadurch verringert. Die PCR-Bedingungen sind unter Tabelle 4 und 5 zusammengefasst.

2.3.2.8.2 3'-Rapid Amplification of cDNA Ends (3'-RACE)

Die 3'-RACE basiert auf einem ähnlichen Prinzip wie die 5'-RACE (Abbildung 6). Als Matrize dient cDNA, die mit einem oligo(dT)-Oligonukleotid synthetisiert wurde. Für die erste *semi-nested* PCR wurde ein Oligonukleotid verwendet, das komplementär zum poly(A)-Schwanz der generierten cDNA ist und darüber hinaus weitere 30 Nukleotide an seinem 3'-Ende besitzt. Das zweite Oligonukleotid wurde so gewählt, dass es in der genspezifischen Sequenz der cDNA hybridisierte (GSP1).

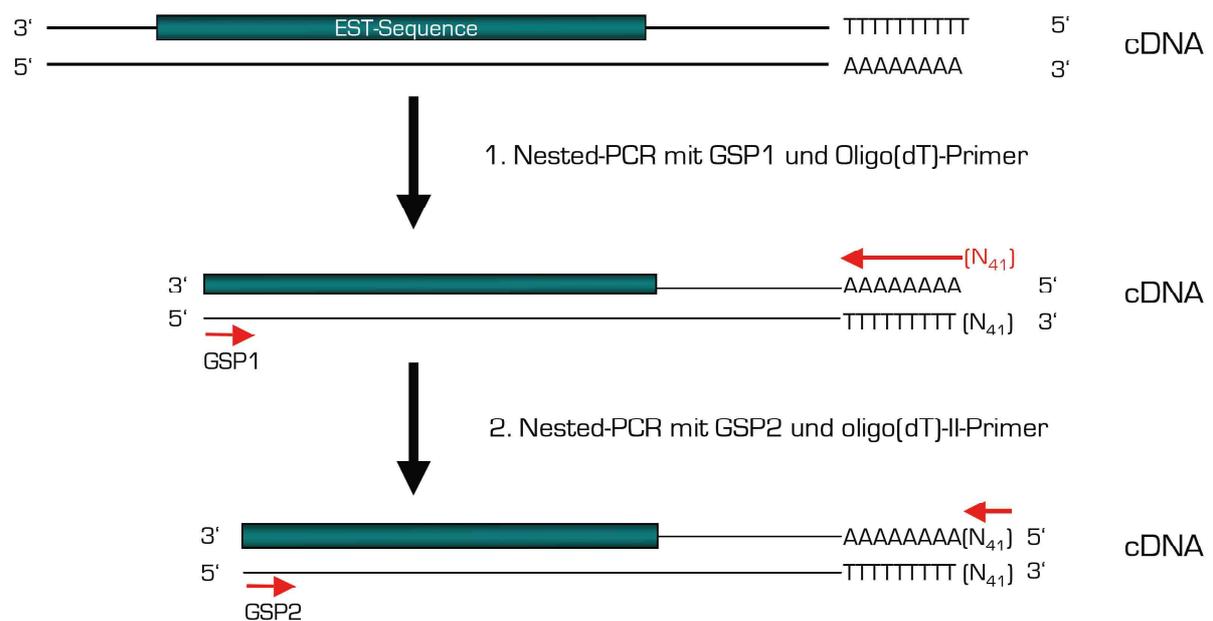


Abbildung 6: Schematische Abbildung der 3'-RACE PCR

Die 3'-RACE verlangt cDNA, die mittels oligo(dT)-Nucleotiden generiert wurde als Matrize. An deren poly(A)-Schwanz bindet das oligo(dT)-Oligonucleotid, das an seinem 5'-Ende einen Überhang trägt, der nicht an die cDNA-Sequenz hybridisieren kann. Die Sequenz des Überhangs diente nachfolgend als Bindungsstelle für weitere PCRs mit genspezifischen Oligonucleotiden (GSP). Dieses Prinzip konnte angewendet werden, bis das gewünschte PCR-Produkt amplifiziert wurde. EST, *Expressed Sequence Tag*; PCR, Polymerase Kettenreaktion; A, Adenin; T, Thymin; GSP, genspezifisches Oligonucleotid; N, Anzahl angefügter Basen; cDNA, *complementary DNA*; RACE, *Rapid Amplification of cDNA Ends*; modifiziert nach Schramm *et al.*, 2000.

Mit diesen Oligonucleotiden wurde eine PCR mit 20 Zyklen gestartet. Aus diesem Amplifikat wurde nach dem Prinzip der 5'-RACE eine zweite *semi-nested* PCR angesetzt, in der ein GSP2 sowie ein Oligonucleotid komplementär zum angefügten Überhang hybridisierten. Dieses Verfahren erfolgte, bis das 3'-Ende des Transkripts detektiert wurde.

2.3.2.9 Aufreinigung von Amplifikationsprodukten

Eine Trennung von PCR-Produkten von restlichen Oligonucleotiden und PCR-Reaktionsbeigaben erfolgte mit dem *NucleoSpin® Gel and PCR Clean-up* Aufreinigungssystem (Macherey & Nagel, Düren) sowie dem *MSB® Spin PCRapace* Aufreinigungssystem (Invitek, Berlin) nach Angaben des Herstellers.

2.3.2.10 Klonierung von Amplifikationsprodukten

Die Klonierung von Amplifikationsprodukten erfolgte mit dem *Topo® TA Cloning® Kit* nach Angaben des Herstellers (Invitrogen, Darmstadt). Dafür wurden zunächst die Produkte mithilfe des *NucleoSpin® Miniprep Kit* (Macherey & Nagel, Düren) oder des *QIAquick® PCR Purification Kits* (Qiagen, Hilden) von restlichen Oligonukleotiden und Oligonukleotid-Dimeren befreit.

Plasmide wurden unter Kanamycin oder Ampicillin selektioniert und die Bakterienkolonien über eine PCR auf ein *Insert* untersucht (*Colony-PCR*). Die *Colony-PCR* erfolgte nach den unter Tabelle 4 aufgeführten Bedingungen über 30 Amplifikationszyklen. Das Amplifikationsprodukt wurde sequenziert und über NCBI Blastn (<http://www.blast.ncbi.nlm.nih.gov/Blast.cgi>) seine genomische Position bestimmt. Handelte es sich dabei um die gesuchte Sequenz, wurde eine Flüssigkultur dieses Klones angesetzt und dieser über Nacht bei 125 rpm bei 37°C in einem Certomat® MO Kreisschüttler inkubiert (BioTech International, Göttingen). Die Plasmide der Bakterienkultur wurden über das *QIAprep® Spin Miniprep Kit* isoliert (Qiagen, Hilden).

2.3.2.11 Gelelektrophoretische Auftrennung von DNA

Die Analyse der PCR-Produkte erfolgte über Auftrennung der Amplifikate mittels Gelelektrophorese. Agarose der Firma ROTH wurde in TBE-Puffer in Konzentrationen von 1,5 – 2,5 % (w/v) angesetzt und mit 200 µg/ml Ethidiumbromid versetzt. Als Laufpuffer diente 1x TBE-Puffer. Das Gel wurde für 30 min bei 120 V laufen gelassen. Anschließend erfolgte eine Visualisierung der Produkte über UV-Licht (254 nm) und eines Geldokumentationssystems.

Als Größenstandard dienten, je nach erwarteter Größe der Amplifikate, *peqGOLD DNA-Sizer XII*, *peqGOLD DNA Leiter Plus* (PEQLAB Biotechnologie GmbH, Erlangen) sowie *GeneRuler™ Low* (Fermentas, St. Leon-Roth).

2.3.3 Zellbiologische Methoden

2.3.3.1 Kultivierung von *E. coli*

Kultur auf Mediumplatten

Die Kultur von *E. coli* auf Mediumplatten erfolgte mit 50 und 200 µl einer Bakteriensuspension unter Zugabe der erforderlichen Menge eines Antibiotikums zum LB-Medium. Nach einer Transformation eines Plasmids mit Resistenzgen wurde zudem eine Blau-Weiß-Färbung durchgeführt indem der Bakteriensuspension 40 µl einer 2%igen X-Gal-Lösung zugeführt wurde. Die Inkubation erfolgte über Nacht bei 37°C bis Einzelkolonien erkennbar waren.

Flüssigkultur

Flüssigkulturen wurden aus Einzelkolonien in 3 ml Gefäßen mit LB-Medium angesetzt. Die Kultivierung erfolgte über Nacht bei 225 rpm und einer Temperatur von 37°C in einem *Innova™ 4400 Incubator Shaker* (New Brunswick Scientific, USA). Dem Medium wurden gegebenenfalls erforderliche Mengen eines Antibiotikums zugeführt.

2.3.3.2 Transformation von *E. coli*

Die Transformation von *E. coli* erfolgte mit chemisch kompetenten Zellen. Ein Aliquot von 200 µl der Zellen wurde auf Eis aufgetaut, mit 1 µl des zu transformierenden Plasmids vermischt und für 15 min auf Eis inkubiert. Nach einem Hitzeschock bei 42°C in einem Wasserbad für 35 sec und einer erneuten Inkubation auf Eis für 2 min wurden die Zellen mit 250 µl SOC-Medium vermengt. Nach einer Inkubation von 30 min bei einer Temperatur von 37°C und 225 rpm in einem Certomat® MO Kreisschüttler (BioTech International, Göttingen) wurden 50 und 200 µl der Zellsuspension auf zwei Kulturplatten ausgestrichen.

2.3.3.3 Isolierung von Plasmid-DNA aus *E. coli*

Die Isolierung bakterieller Plasmid-DNA erfolgte mit dem *QIAprep® Spin Miniprep Kit* (Qiagen) nach Anleitung des Herstellers. Die DNA wurde in 30 µl TE-Puffer eluiert und anschließend mit einem Bio-Fotometer (Eppendorf) die DNA-Konzentration gemessen.

2.3.4 *In silico*-Methoden

2.3.4.1 Charakterisierung von genomischen Bereichen

Sequenzierte DNA-Bereiche wurden mithilfe des MapViewers von NCBI (<http://www.ncbi.nlm.nih.gov/mapview>) und des Genome Browsers von UCSC (<http://www.genome.ucsc.edu/cgi-bin/hgGateway>) ermittelt.

2.3.4.2 Sequenzanalyse

Die Auswertung von sequenzierten DNA-Fragmenten wurden mit dem Programm *Sequencing Analysis 5.1* und *SeqScape v2.1.1* durchgeführt. Zusätzlich wurde NCBI Blastn für die Sequenzanalyse verwendet.

2.3.4.3 Untersuchungen zur putativen Proteinkodierung

Für Untersuchungen zu möglichen Protein-kodierenden Sequenzen in den Transkripten wurde die Sequenz mit der Internetsoftware ExPASy (Gasteiger *et al.*, 2003) geprüft und die möglichen Aminosäurenabfolgen mit NCBI Blastp (<http://www.blast.ncbi.nlm.nih.gov/Blast.cgi>) auf Sequenzähnlichkeiten zu anderen Proteinen durchsucht.

2.3.4.4 SNP-Datenbanken

Für die Analyse von SNPs wurden folgende Datenbanken herangezogen:

- SNPper (<http://www.snpper.chip.org>)
- dbSNP (<http://www.ncbi.nlm.nih.gov>)

- Ensemble BioMart (<http://www.ensembl.org>)
- 1000 Genomes (<http://www.1000genomes.org>)

2.3.4.5 *Expressed Sequence Tag*-Datenbanken

Die Informationen zu den *Expressed Sequence Tags* (ESTs) wurden den Datenbanken von NCBI (<http://www.ncbi.nlm.nih.gov/unigene>) und UCSC (<http://www.genome.ucsc.edu/cgi-bin/hgGateway>) entnommen.

2.3.4.6 Oligonukleotid-*Design*

Der Entwurf der Oligonukleotidpaare für PCR-basierte Methoden erfolgte über *Primer3 v0.4.0* (Rozen & Skaletsky, 2000). Als genomische Grundlage dienten die bei NCBI Blast hinterlegten Sequenzen.

Der Entwurf der Oligonukleotide der DASH-Methode erfolgte über die *LightCycler® Probe Design Software 2.0*. Die Grundlage dieser Sequenzen waren die in dbSNP hinterlegten umliegenden Sequenzen der jeweils zu untersuchenden Variante.

2.3.5 Auswahl des Assoziationsbereichs

Für weiterführende molekularbiologische Untersuchungen wurde ein Assoziationsbereich um den SNP rs12479448 festgelegt, der die höchste Assoziation mit zerebraler Malaria zeigte. Der Assoziationsbereich wurde begrenzt durch ein LD von $r^2 > 0,05$ zu rs12479448 und einem Signifikanzniveau von $p < 0,05$ (Abbildung 12). Die SNPs rs7557477 und rs1823800 mit den genomischen Position 35046574 und 35123318 bp (*Genome Build* 36.3) stellten jeweils die äußeren Grenzen des insgesamt 76,7 kb betragenden Assoziationsbereiches dar.

3 Ergebnisse

3.1 Genomweite Assoziationsstudie zu zerebraler Malaria

3.1.1 Qualitätskontrolle der Einzelnukleotidaustausche

Von den 906622 genomweiten und mitochondrialen Einzelnukleotidaustauschen (SNPs), die mit dem *Genome-Wide Human SNP Array 6.0* von Affymetrix® typisiert wurden, erfüllten 774800 autosomale SNPs die Standards der Qualitätskontrolle (siehe Material und Methoden) und konnten in die nachfolgenden statistischen Analysen der genomweiten Assoziationsstudie (GWA-Studie) aufgenommen werden.

3.1.2 Populationsstratifikation der GWA-Studiengruppe

Für Untersuchungen auf Populationsstratifikation wurde das Verfahren der multidimensionalen Skalierung (MDS) auf die Genotypen angewendet, die von 774800 SNPs in der GWA-Studiengruppe von 1237 Individuen ermittelt wurden. Von den drei mit der MDS-Transformation errechneten Dimensionen, die als Variable für eine molekulargenetisch bestimmte Klassifikation der genetischen Heterogenität der Individuen verwendet wurden, sind zur Illustration die Dimensionen 1 und 2 dargestellt (Abbildung 7).

In der Abbildung sind zwei benachbarte *Cluster* von Individuen sowie ein Areal mit Individuen zu erkennen, die außerhalb dieser *Cluster* liegen. Eine Stratifikation nach ethnischer Selbstauskunft fasst die Individuen aus der mit ca. 65% am häufigsten angegebenen Ethnie Akan im Wesentlichen im linken *Cluster* zusammen. Während dessen bilden Individuen, die nach eigenen Angaben der zweitgrößten Gruppe *Northerners* zuzuordnen waren, mit ca. 31% den Großteil des rechten *Clusters*. In diesen beiden *Clustern* sind auch die meisten Individuen der angegebenen Ethnien Ga-Adangbe und Ewe vertreten, die nur einen sehr geringen Anteil von jeweils weniger als 2% an der Studienpopulation haben (siehe Tabelle 2). Die Streuung außerhalb der beschriebenen Haupt-*Cluster* wird überwiegend von Individuen erzeugt, die nach eigenen Angaben der Gruppe *Northerners* zuzuordnen waren. Es wurde somit eine Populationsstratifikation in der Studiengruppe gefunden, die mit der

alleinigen Zuordnung der Individuen zu ghanaischen Ethnien bzw. Gruppen von Ethnien nur unzureichend beschrieben werden kann. Um die beobachtete Populationsstratifikation in den folgenden Assoziationsberechnungen zu berücksichtigen, wurden die ersten drei MDS-Dimensionen als Kovariaten in die nachfolgende statistische Analyse einbezogen.

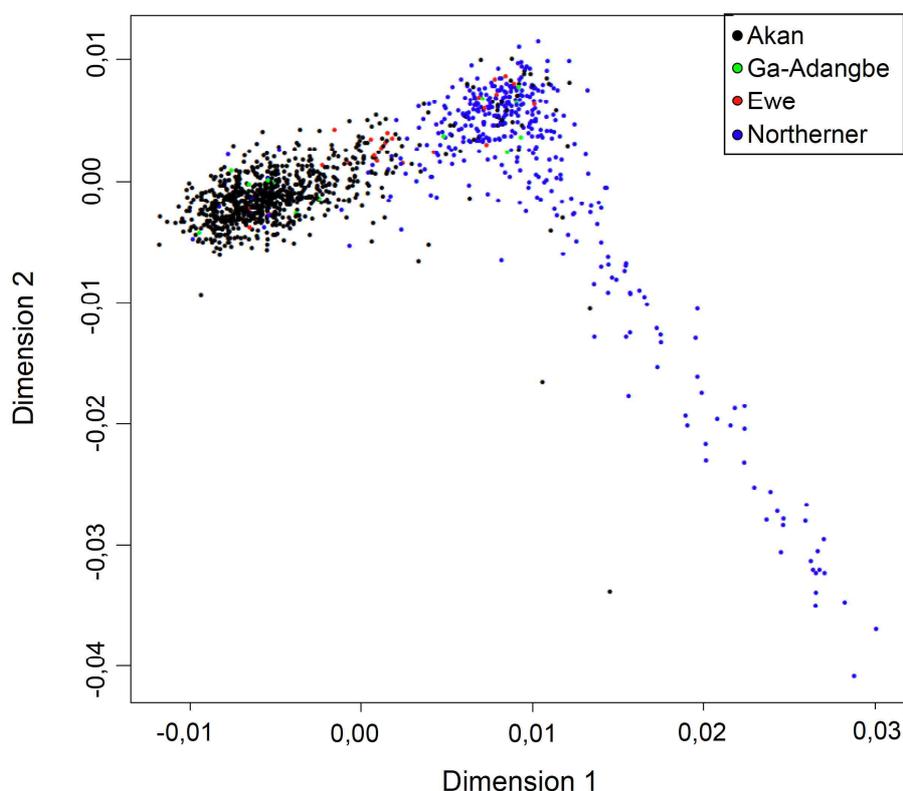


Abbildung 7: 1. und 2. Dimension der multidimensionalen Skalierung (MDS), stratifiziert nach ethnischer Herkunft

Die MDS-Dimensionen wurden aus 774800 SNP-Genotypen von 409 Fällen mit zerebraler Malaria und 828 gesunden Kontrollen ermittelt. Jeder Punkt stellt ein Individuum dar.

Nach Stratifikation des *Plots* der 1. und 2. Dimension der MDS konnten keine systematischen Abweichungen der Verteilungen zwischen den Fällen und Kontrollen und den Geschlechtern erkannt werden (Abbildung 8 und Abbildung 9).

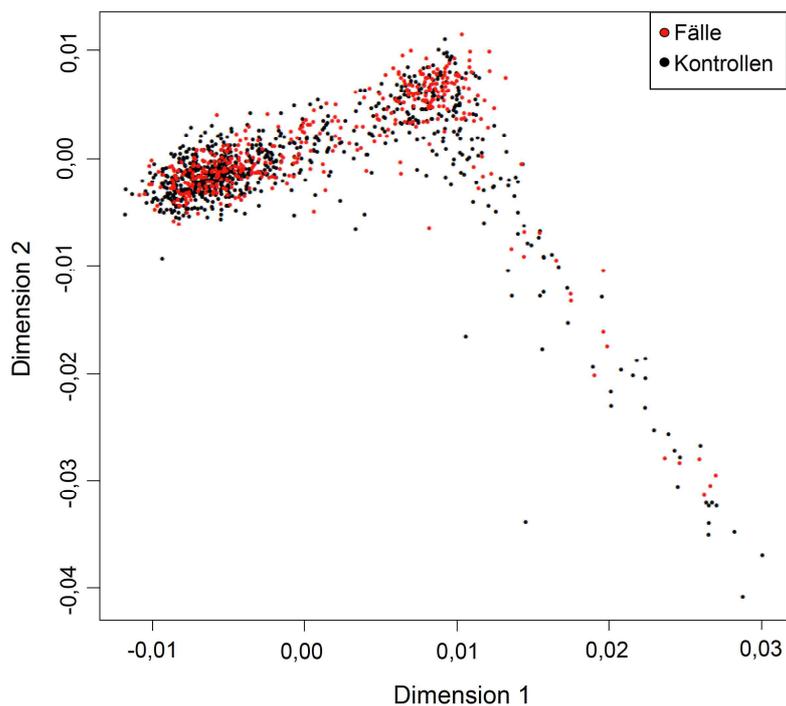


Abbildung 8: 1. und 2. Dimension der multidimensionalen Skalierung (MDS), stratifiziert nach Fall-Kontroll-Status

Die MDS-Dimensionen wurden aus 774800 SNP-Genotypen von 409 Fällen mit zerebraler Malaria und 828 gesunden Kontrollen ermittelt. Jeder Punkt stellt ein Individuum dar.

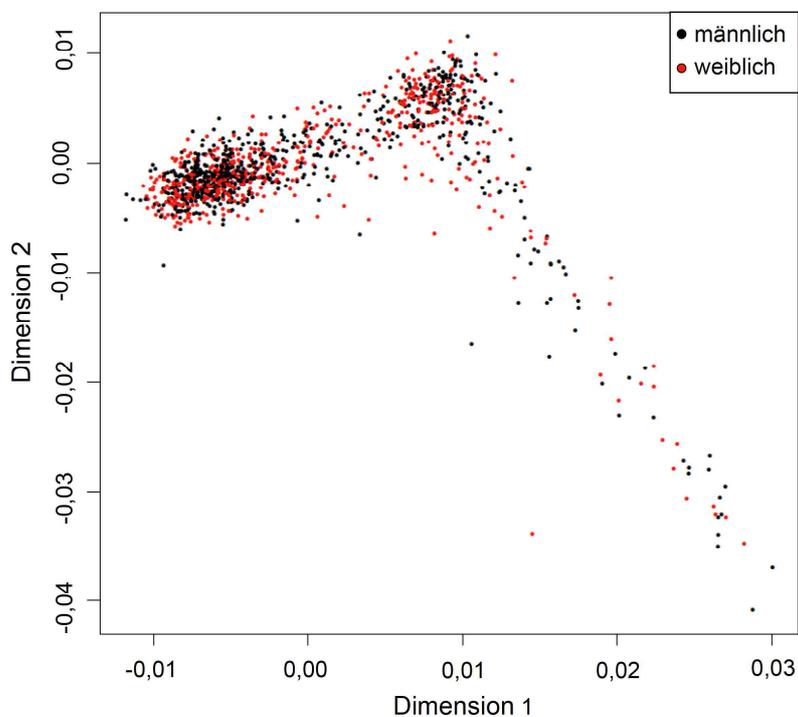


Abbildung 9: 1. und 2. Dimension der multidimensionalen Skalierung (MDS), stratifiziert nach Geschlecht

Die MDS-Dimensionen wurden aus 774800 SNP-Genotypen von 409 Fällen mit zerebraler Malaria und 828 gesunden Kontrollen ermittelt. Jeder Punkt stellt ein Individuum dar.

3.1.3 GWA-Studie zur zerebralen Malaria

Für die Untersuchungen zur Assoziation von zerebraler Malaria mit 774800 SNPs wurden die Genotypen der 409 Fälle der GWA-Studiengruppe mit denen der 828 gesunden Kontrollen verglichen. Die statistische Auswertung erfolgte mit einer multivariaten logistischen Regression unter Einbeziehung der ersten drei MDS-Dimensionen und unter Annahme eines additiven Vererbungsmodells.

Mit dem Ziel, eine Auswahl an SNPs zu treffen, die nach einer gemeinsamen Analyse mit weiteren Studienteilnehmern den empfohlenen Schwellenwert für genomweite Signifikanz von $p < 5 \times 10^{-8}$ (Manolio *et al.*, 2008; Pe'er *et al.*, 2008) erreichen könnten, wurde ein willkürlicher Schwellenwert von $p < 5 \times 10^{-5}$ gesetzt. SNPs, die diesen Schwellenwert unterschritten, wurden für eine weitere Analyse in Betracht gezogen. Eine genomweite Übersicht der errechneten Signifikanzniveaus ist in einem so genannten *Manhattan-Plot* dargestellt (Abbildung 10).

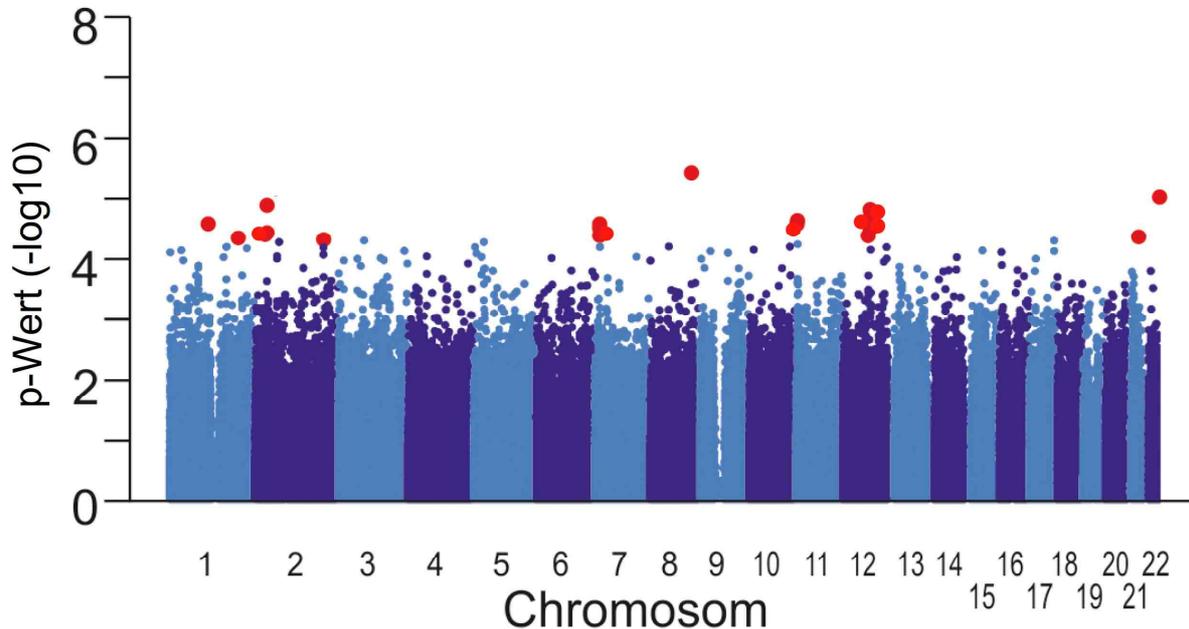


Abbildung 10: Manhattan-Plot der GWA-Studiengruppe

Signifikanzniveaus von 774800 autosomalen SNPs der GWA-Studiengruppe, aufgetragen gegen ihre genomische Position. Statistische Analysen erfolgten mit multivariater logistischer Regression unter Annahme eines additiven Vererbungsmodells. Rote Punkte repräsentieren die für die Replikationsstudie ausgewählten SNPs mit einem p-Wert von $< 5 \times 10^{-5}$, blaue Punkte stehen für SNPs mit einem p-Wert von $\geq 5 \times 10^{-5}$.

Unter Berücksichtigung des Schwellenwertes wurden 19 SNPs auf acht unterschiedlichen Chromosomen identifiziert (Tabelle 10). Fielen mehrere dieser SNPs in einen chromosomalen Bereich von bis zu 20 kb, wurden diese aufgrund der Annahme, dass unter ihnen ein Kopplungsungleichgewicht (*linkage disequilibrium*, LD) besteht, einem Locus zugeordnet. Dies führte zu einer Verteilung der 19 SNPs auf 13 unabhängige Loci.

Von den 13 identifizierten Loci wurden 11 Loci weiter bearbeitet. Zwei SNPs, die in dieser Studie nicht weiter analysiert wurden (rs10836540 und rs12805180), liegen auf Chromosom 11 in der Nähe des *HBB*-Gens (Tabelle 10). *HBB* kodiert für die Kette des β -Hämoglobins und stellt somit einen bekannten Resistenzlocus der schweren Malaria dar (May *et al.*, 2007).

Die Genotypen der 11 am jeweiligen Locus am stärksten assoziierten SNPs wurden mit unabhängigen Genotypisierungsmethoden (SNPlex® oder DASH, *dynamic allele-specific hybridization*; je nach Verfügbarkeit) in derselben Studiengruppe erneut typisiert. Die Ergebnisse beider Genotypisierungsmethoden wurden miteinander verglichen und eine Konkordanz von über 98% wurde festgestellt.

3.2 Kontrolle der GWA-Ergebnisse in einer unabhängigen Fall-Kontroll-Gruppe

Die 11 SNPs mit der stärksten Assoziation am jeweiligen Locus wurden in einer unabhängigen Studiengruppe, der Replikationsgruppe bestehend aus zusätzlichen 159 Fällen mit zerebraler Malaria und 1042 gesunden Kontrollen, genotypisiert.

Um das 11-fache Testen der untersuchten SNPs zu berücksichtigen, wurde das für eine Replikation zu unterschreitende Signifikanzniveau auf eine Schwelle von $p < 4,5 \times 10^{-3}$ ($p = 0,05$ bei 11 SNPs) angepasst. Die statistische Analyse mit multivariater logistischer Regression zeigte unter Annahme eines additiven Vererbungsmodells und bei Korrektur auf die Variablen Alter, Geschlecht und Ethnie für den SNP rs12479448 eine signifikante Assoziation ($p = 2,9 \times 10^{-4}$, Tabelle 10).

Tabelle 10: Ergebnisse der Assoziationsstudien in der GWA-Studiengruppe und der Replikationsgruppe

SNPs* ausgewählter Loci der GWA-Studie					GWA-Gruppe					Replikationsgruppe				
Chr.	Position# [bp]	rs- Nummer	Anzahl assoziierter SNPs* am Locus	Verer- bungs- modell	Minor-/ Major- Allel	MAF Fall- gruppe [%]	MAF Kontroll- gruppe [%]	p-Wert	OR (95% KI)	Minor-/ Major- Allel	MAF Fall- gruppe [%]	MAF Kontroll- gruppe [%]	p-Wert	OR (95% KI)
1	112146150	rs7517611	1		A/C	15,4	10,2	$2,7 \times 10^{-5}$	1,7 (1,3 – 2,2)	A/C	11,5	12,0	$5,9 \times 10^{-1}$	0,9 (0,6 - 1,3)
1	199439727	rs832160	1		A/G	22,8	30,6	$4,6 \times 10^{-5}$	0,7 (0,5 – 0,8)	A/G	25,8	28,6	$2,6 \times 10^{-1}$	0,9 (0,6 - 1,1)
2	3015354	rs20289833	1		C/T	4,7	1,6	$4,0 \times 10^{-5}$	2,9 (1,8 – 4,9)	C/T	1,5	1,9	$7,0 \times 10^{-1}$	0,8 (0,3 - 2,3)
2	35069984	rs12479448	2		A/T	20,9	13,6	$1,3 \times 10^{-5}$	1,7 (1,3 – 2,1)	A/T	24,5	16,2	$2,9 \times 10^{-4}$	1,7 (1,3 - 2,3)
2	198939676	rs2467041	1		T/C	46,7	44,4	$4,9 \times 10^{-5}$	1,4 (1,2 – 1,7)	T/C	46,5	46,3	$8,8 \times 10^{-1}$	1,0 (0,8 - 1,3)
7	13667180	rs12538929	3		T/A	38,5	46,1	$2,6 \times 10^{-5}$	0,7 (0,6 - 0,8)	T/A	46,9	42,8	$2,6 \times 10^{-1}$	1,2 (0,9 - 1,5)
8	120629877	rs7008680	1	additiv	A/G	42,1	33,1	$3,8 \times 10^{-6}$	1,5 (1,3 – 1,8)	A/G	34,9	34,6	$8,6 \times 10^{-1}$	1,0 (0,8 - 1,3)
11	4707349	rs10836540	1		T/C	31,0	34,0	$2,7 \times 10^{-5}$	0,7 (0,5 – 0,8)	nicht weiter bearbeitet, siehe 3.1.3				
11	5180937	rs12805180	1		A/C	29,0	27,0	$2,3 \times 10^{-5}$	1,5 (1,3 – 1,8)	nicht weiter bearbeitet, siehe 3.1.3				
12	75614451	rs1806597	1		T/C	29,0	22,1	$4,2 \times 10^{-5}$	1,5 (1,2 – 1,8)	T/C	24,2	27,8	$1,4 \times 10^{-1}$	0,8 (0,6 - 1,1)
12	81924270	rs10082865	4		C/G	48,5	39,4	$1,5 \times 10^{-5}$	1,5 (1,2 – 1,7)	C/G	40,4	42,3	$6,2 \times 10^{-1}$	0,9 (0,7 - 1,2)
21	3490573	rs2835314	1		A/G	17,9	24,5	$4,4 \times 10^{-5}$	0,6 (0,5 – 0,8)	A/G	21,0	21,9	$7,7 \times 10^{-1}$	1,0 (0,7 - 1,3)
22	47123181	rs1883129	1		G/C	43,9	46,6	$9,5 \times 10^{-6}$	1,5 (1,2 – 1,8)	C/G	48,7	47,4	$1,3 \times 10^{-1}$	1,2 (0,9 - 1,5)

GWA-Gruppe (409 Fälle, 828 Kontrollen), genotypisiert mit dem *Genome-wide Human SNP Array 6.0* (Affymetrix). Replikationsgruppe (159 Fälle, 1042 Kontrollen), genotypisiert mittels SNPlex® oder DASH (*dynamic allele-specific hybridization*). Chr., Chromosom; bp, Basenpaar; rs-Nummer, NCBI-Nummer annotierter SNPs; SNPs, Einzelnukleotidaustausche; MAF, Minor-Allelfrequenz; A, Adenin; C, Cytosin; G, Guanin; T, Thymin; GWA, genomweite Assoziation; OR, *Odds Ratio*; KI, Konfidenzintervall; *, SNPs, die nach statistischer Analyse einen p-Wert von $p < 5 \times 10^{-5}$ aufwiesen, Assoziationsberechnungen ausgehend vom Minor-Allel; #, NCBI *Genome Build* 36.3.

3.3 Untersuchungen der kombinierten GWA- und Replikationsgruppen

Mit dem Ziel, Assoziationsberechnungen in den kombinierten Studiengruppen vorzunehmen, wurden für den SNP rs12479448 die Genotypen der GWA- und der Replikationsgruppe in einen gemeinsamen Datensatz zusammengeführt. Die statistische Auswertung erfolgte wie unter 3.2 beschrieben. Anhand der Genotypen der insgesamt 568 Fälle und 1870 Kontrollen ergab sich eine Assoziation der Variante rs12479448 mit zerebraler Malaria (*Odds Ratio* [OR] = 1,6; 95% Konfidenzintervall [KI] = 1,4 – 1,9; $p = 7,8 \times 10^{-8}$).

Um die statistische *Power* der Assoziationsstudie weiter zu erhöhen, wurden die Genotypen einer zusätzlichen Kontrollgruppe aus derselben ghanaischen Population, bestehend aus 1081 Individuen, eingeschlossen. In der Analyse von nunmehr insgesamt 3519 Individuen zeigte sich für den SNP rs12479448 eine genomweit signifikante Assoziation mit einem p -Wert von $2,4 \times 10^{-8}$ (Tabelle 12).

Zusätzlich zum additiven Modell wurden Berechnungen zum dominanten, rezessiven und überdominanten Vererbungsmodell durchgeführt. Eine stärkere Assoziation als die von rs12479448 mit zerebraler Malaria wurde nicht gefunden (Tabelle 12).

3.4 Transmissions-Disequilibrium-Test (TDT) zur Bestätigung der bisherigen Assoziation

Um die Assoziation der Variante rs12479448 mit den Allelen A und T mit einem weiteren, nicht populationsbasierten Testverfahren zu bestätigen, wurde als ein familienbasiertes Testverfahren ein TDT durchgeführt, für den 169 Trios (Eltern und ein an zerebraler Malaria erkranktes Kind) sowie 216 Duos (ein Elternteil und ein erkranktes Kind) zur Verfügung standen. Für die Berechnungen wurde das Programm *UNPHASED* verwendet, das es ermöglicht, neben der Analyse von Trios auch Duos zu berücksichtigen.

Der TDT zeigte für das Allel A von rs12479448 ein für einfaches Testen signifikantes Transmissionsungleichgewicht (OR = 1,7; 95 % KI = 1,3 – 1,9; $p = 4,4 \times 10^{-4}$) für eine bevorzugte Vererbung dieses Allels von den Eltern auf ihre erkrankten Kinder.

3.5 Fine Mapping der mit zerebraler Malaria assoziierten Region auf Chromosom 2p22.3

Mit dem Ziel, genetische Varianten mit stärkerer Assoziation und letztendlich die kausale Variante zu finden, wurde eine DNA-Sequenzierung von 24 Fällen und 24 Kontrollen über einen Bereich von 100 kb durchgeführt. Der sequenzierte Bereich erstreckte sich 50 kb 5' und 3' der chromosomalen Position 35069984 bp der am stärksten assoziierten Variante rs12479448. Es wurden 1013 genetische Varianten nachgewiesen, von denen 380 nicht in SNP-Datenbanken (Stand: März 2012) annotiert waren. Die 380 neuen genetischen Varianten setzen sich zusammen aus 323 SNPs, 32 Insertionen und 25 Deletionen (siehe Anhang).

Mithilfe des Programms Haploview (www.broad.mit.edu/mpg/haploview) wurden Analysen zum Kopplungsungleichgewicht zwischen den Varianten durchgeführt. Insgesamt zeigt der Bereich ein relativ geringes LD ohne eindeutig identifizierbare Haplotypen. In Abbildung 11 ist das LD als Quadrat des paarweisen Korrelationskoeffizienten r (r^2) zwischen den Varianten dargestellt, die in einem Ausschnitt von 20 kb um den SNP rs12479448 liegen und ein minimales LD von 0,01 zu rs12479448 aufweisen.

Drei der 1013 untersuchten Varianten stehen in einem vergleichsweise hohen LD von $r^2 \geq 0,6$ zu rs12479448 (rs12476596, rs34996117 und rs6543892; Tabelle 11 und Abbildung 11). Diese wurden ausgewählt, um sie in der gesamten Studienpopulation zu genotypisieren.

Die statistische Analyse wurde wie unter 3.2 beschrieben durchgeführt und zeigte im gesamten Studienkollektiv eine genomweit signifikante Assoziation der Variante rs34996117 (OR = 1,6; 95 % KI = 1,3 – 1,9; $p = 3,1 \times 10^{-8}$; Tabelle 12). Eine stärkere Assoziation als die für rs12479447 gefundene mit zerebraler Malaria, konnte für keine der drei genetischen Varianten nachgewiesen werden (Tabelle 12).

Tabelle 11: Kopplungsungleichgewicht und Entfernung der für eine Genotypisierung im gesamten Studienkollektiv ausgewählten genetischen Varianten zum SNP rs12479448

SNP	Chromosomale Position [#] [bp]	Entfernung zu rs12479448 [bp]	Austausch	r ²
rs6543892	35061277	- 8707	C/T	0,615
rs12476596	35065680	- 4304	C/A	0,636
[rs12479448]	35069984	0	A/T	1
rs34996117	35073041	+ 3057	delT	0,906

Der Wert $r^2 = 1$ entspricht einer vollständigen Kopplung der genetischen Varianten, der Wert 0 entspricht keiner Kopplung. Für die Analyse wurden die Genotypen der Varianten herangezogen, die in 24 Fällen mit zerebraler Malaria und 24 Kontrollen durch DNA-Sequenzierung ermittelt wurden. In Klammern steht der SNP rs12479448, auf den Bezug genommen wurde. SNP, Einzelnukleotidaustausch; bp, Basenpaar; r², Korrelationskoeffizient; rs-Nummer, NCBI-Nummer annotierter SNPs; C, Cytosin; T, Thymin; A, Adenin; del, Deletion; [#], NCBI *Genome Build* 36.3.

Tabelle 12: Ergebnisse von Assoziationstests nach Kombination der GWA- und der Replikationsgruppe und von ausgewählten Varianten mit einem Korrelationskoeffizient von $r^2 \geq 0,6$ zu rs12479448 nach dem *Fine Mapping* einer 100 kb-Region auf Chromosom 2p22.3

Chr.	SNP der GWA-Studie			Kombinierte Analyse					
	Position [#] [bp]	rs-Nummer	Annotation am Locus	Minor-/ Major- Allel	MAF Fallgruppe [%]	MAF Kontrollgruppe [%]	Vererbungs- modell	p-Wert	OR (KI 95%)
2	35069984	rs12479448	intergenisch	A/T	21,9	15,0	additiv	$2,4 \times 10^{-8}$	1,6 (1,4 - 1,9)
							dominant	$8,7 \times 10^{-8}$	1,7 (1,4 - 2,0)
							rezessiv	$4,2 \times 10^{-3}$	2,0 (1,3 - 3,3)
SNPs des <i>Fine Mappings</i>				Kombinierte Analyse					
2	35061277	rs6543892	intergenisch	C/G	22,71	25,02	additiv	$2,1 \times 10^{-5}$	1,4 (1,2 - 1,6)
2	35065680	rs12476596	intergenisch	A/C	13,78	15,64	additiv	$5,6 \times 10^{-7}$	1,5 (1,3 - 1,8)
2	35073041	rs34996117	intergenisch	delT/T	15,17	17,44	additiv	$3,1 \times 10^{-8}$	1,6 (1,3 - 1,9)

Kombinierte Studiengruppe mit 568 Fällen mit zerebraler Malaria und 2951 Kontrollen (GWA-Gruppe, 409 Fälle und 828 Kontrollen; Replikationsgruppe, 159 Fälle und 1042 Kontrollen; 1081 zusätzliche Kontrollen), Genotypisierung mittels DASH (*dynamic allele-specific hybridization*). Assoziationsberechnungen ausgehend vom Minor-Allel. GWA-Studie, genomweite Assoziationsstudie; Chr., Chromosom; bp, Basenpaar; rs-Nummer, NCBI-Nummer annotierter SNPs; MAF, Minor-Allelfrequenz; OR, *Odds Ratio*; KI, Konfidenzintervall; A, Adenin; T, Thymin; C, Cytosin; del, Deletion; [#], NCBI *Genome Build* 36.3.



Abbildung 11: Kopplungsungleichgewicht [LD] von genetischen Varianten in einem Bereich von 20 kb um rs12479448

Die Analyse des LD mithilfe des Programms Haploview basiert auf den Genotypen von 97 genetischen Varianten, die in der DNA-Sequenzierung von 24 Fällen mit zerebraler Malaria und 24 Kontrollen ermittelt wurden und in einem minimalen LD von $r^2 \geq 0,01$ zu rs12479448 stehen. Die Quadrate symbolisieren mit ihren Farben die Stärke des paarweisen LDs der entsprechenden Varianten. Die Positionen des SNPs rs12479448 sowie die der aufgrund ihres hohen LDs mit rs12479448 für eine Genotypisierung im gesamten Studienkollektiv ausgewählten genetischen Varianten im NCBI *Genome Build* 36.3 sind mit der jeweiligen rs-Nummer gekennzeichnet. Weitere genetische Varianten sind entsprechend ihrer Position auf dem Chromosom markiert. Die roten Rechtecke heben das LD von rs12479448 zu weiteren Varianten des 20 kb Bereichs hervor, grüne Quadrate das LD der im gesamten Studienkollektiv genotypisierten Varianten zu rs12479448. kb, Kilobasen; chr., chromosomale; rs-Nummer, NCBI-Nummer annotierter SNPs; r^2 , Korrelationskoeffizient.

Abbildung 12 gibt einen Überblick über die gefundenen Signifikanzniveaus am Locus auf Chromosom 2p22.3.

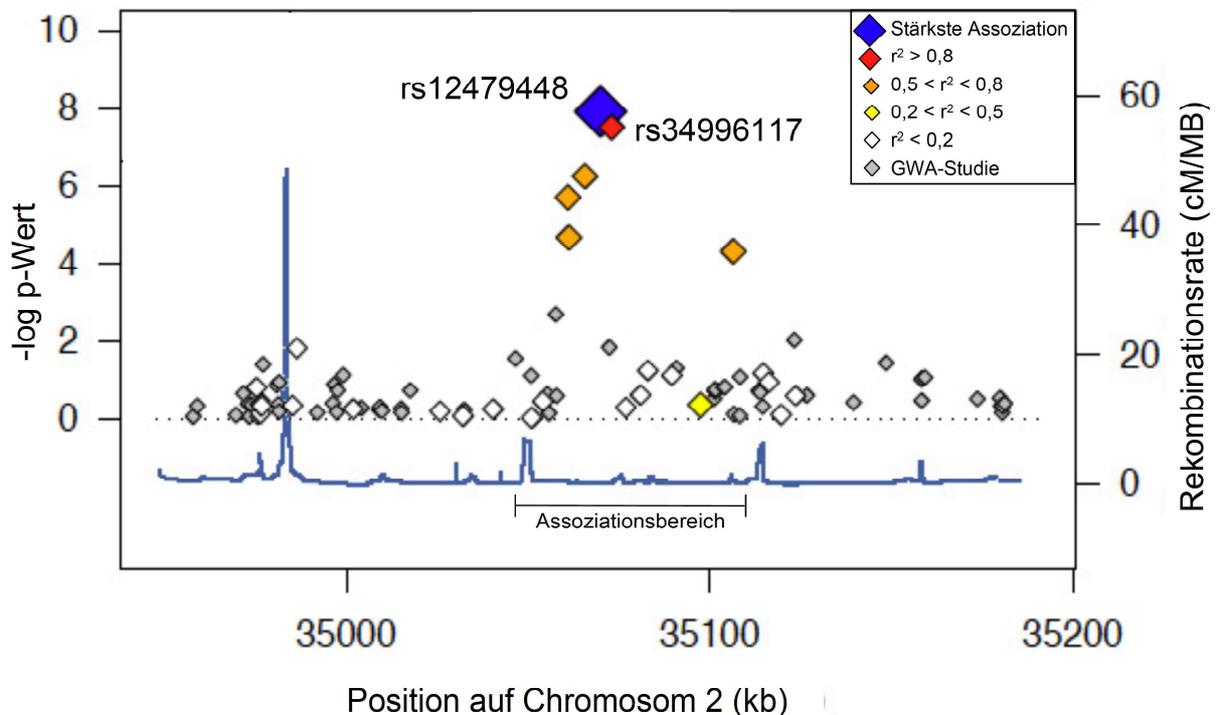


Abbildung 12: Locus auf Chromosom 2p22.3 mit Assoziation zu zerebraler Malaria

Darstellung der Signifikanzniveaus von SNPs aufgetragen gegen ihre chromosomale Position (NCBI *Genome Build* 36.3). Graue Rauten stehen für SNPs, die mit dem *Genome-Wide SNP Array* 6.0 von Affymetrix® in der GWA-Studiengruppe, bestehend aus 409 Fällen und 828 Kontrollen, genotypisiert wurden. Farbige Rauten stehen für die genetischen Varianten, die mittels DASH (*dynamic allele-specific hybridization*) oder SNPlex® in 568 Fällen und 2951 Kontrollen (GWA-Studiengruppe eingeschlossen) genotypisiert wurden. Die Farbabstufungen von rot nach weiß geben das Kopplungsungleichgewicht als Quadrat des Korrelationskoeffizienten r^2 zu dem am stärksten assoziierten SNP rs12479448 (blau) wieder. Die dunkelblaue Linie kennzeichnet die Rekombinationsrate dieses chromosomalen Bereichs in der HapMap CEU-Population (Population von 180 Individuen aus Utah nord- und westeuropäischer Abstammung). Darunter ist der für spätere Untersuchungen verwendete Assoziationsbereich abgebildet. SNPs mit genomweiter Signifikanz ($p < 5 \times 10^{-8}$) sind mit ihrer rs-Nummer gekennzeichnet. cM/MB, Centimorgan/Megabasen; kb, Kilobasenpaare; GWA-Studie, genomweite Assoziationsstudie.

3.6 Nachweis von Transkripten in der mit zerebraler Malaria assoziierten chromosomalen Region 2p22.3

Mit dem Ziel, in der Region 2p22.3 kodierte Transkripte zu untersuchen, wurde zunächst ein Untersuchungsbereich festgelegt, der im Folgenden als „Assoziationsbereich“ bezeichnet wird. Dieser wurde durch SNPs begrenzt, die einerseits ein LD $r^2 > 0,05$ zu rs12479448 und ein Signifikanzniveau von $p = 0,05$

erreichten (siehe 2.3.5). Der Assoziationsbereich umfasste 76,7 kb und reichte von den chromosomalen Positionen 35046574 – 35123318 bp.

3.6.1 Recherche in Internet-Datenbanken

In den Datenbanken von NCBI (www.ncbi.nlm.nih.gov) und UCSC (www.genome.ucsc.edu) sind mit dem Stand von März 2012 im Assoziationsbereich keine Protein-kodierenden Gene annotiert. Es sind jedoch Angaben zu fünf Transkripten unbekannter Funktion in der GenBank-Datenbank hinterlegt, die vollständig oder unvollständig im Assoziationsbereich liegen oder diesen umspannen. Die Sequenzinformation der Transkripte basiert auf der *complementary* DNA (cDNA)-Sequenz von RNA aus verschiedenen Geweben. Die cDNA wurde mit unterschiedlichen Methoden synthetisiert (Tabelle 13). Abbildung 13 gibt einen schematischen Überblick über die Transkripte im Assoziationsbereich und der umliegenden chromosomalen Region. Zur Vereinfachung wurden die Transkripte wie folgt abgekürzt: DA227815 in DA, AA861169 in AA, AI218168 in AI, AW665463 in AW und BX091678 in BX.

Die Transkripte AA, AI und AW liegen vollständig im Assoziationsbereich. Das Transkript BX beginnt außerhalb des Assoziationsbereichs, endet jedoch mit dem 3. Exon innerhalb des Assoziationsbereichs. Das Transkript DA umspannt mit einem 327590 bp langen Intron den Assoziationsbereich.

In Tabelle 13 sind Charakteristika der Transkripte aus den genannten Datenbanken zusammengefasst. Da ein Großteil der Sequenzinformation von AI und AW identisch ist und sie sich ebenfalls hinsichtlich ihrer Spleißstellen gleichen, wurden sie als Varianten eines Transkripts dargestellt, welches im Folgenden als AI/AW bezeichnet wird. AI/AW besitzt in der nach UCSC beschriebenen Orientierung auf dem Plusstrang keine konventionellen Spleißstellen entsprechend der GT-AG-Regel (Breathnach *et al.*, 1978; Mount, 1982).

Tabelle 13: Charakteristika der Transkripte des Assoziationsbereichs auf Chromosom 2p22.3 aus den Internet-basierten Datenbanken von UCSC und NCBI

Transkript	Annotierte cDNA-Länge in UCSC-/NCBI-Datenbank [bp]	Strang-orientierung*	Anzahl der Exons	konventielle Spleißstellen
AI/AW ¹	503/503	+	3	nein [#]
BX ¹	383/385	-	3	ja
AA ²	356/356	-	1	kein Spleißprozess
DA ³	580/580	+	6	ja

AI/AW, AI218168/AW665463; BX, BX091678; AA, AA861169; DA, DA227815; cDNA, *complementary DNA*; bp, Basenpaare; ¹ keine Angabe zur Methode der cDNA-Synthese; ² cDNA-Synthese mit einem Oligo(dT)-Nukleotid mit NotI-Restriktionsstelle; ³ cDNA-Synthese mit der *Oligo-Capping-Methode*; * Strangorientierung nach UCSC zur genomischen DNA-Sequenz, gelesen in Plusstrangrichtung vom Telomer des kurzen Arms des Chromosoms zum Telomer des langen Arms des Chromosoms in 5'-3'-Richtung; [#] unter Annahme einer Kodierung auf dem Minusstrang ergeben sich für dieses Transkript die konventionellen Spleißstellen.

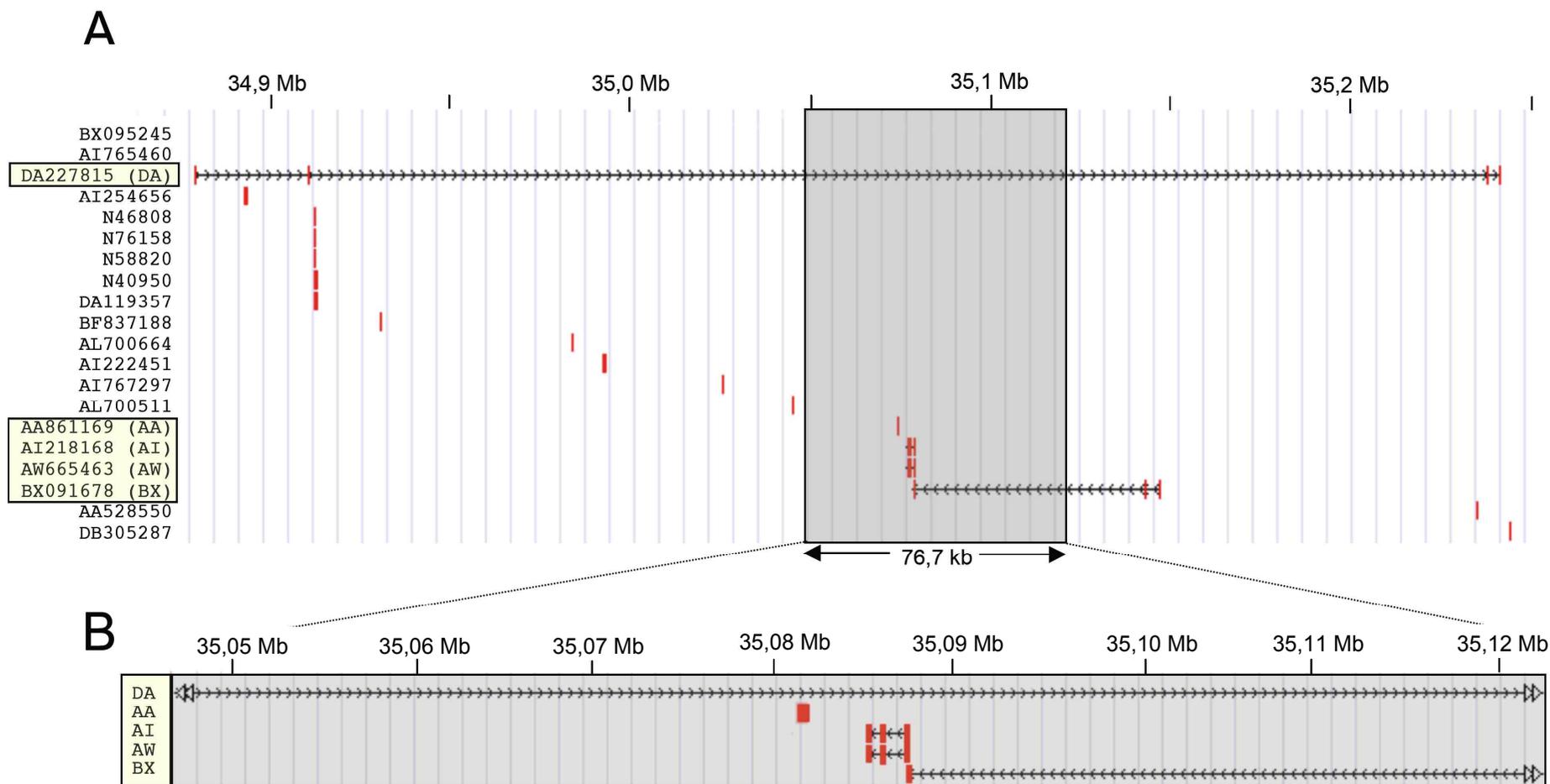


Abbildung 13: Annotierte Transkripte auf Chromosom 2p22.3

A: Darstellung der primären Transkripte aus dem 76,7 kb umfassenden Assoziationsbereich (grau) in einer Region von 400 kb auf Chromosom 2p22.3 (aus dem UCSC *Genome Browser*; NCBI *Genome Build* 36.3). B: Vergrößerte Darstellung des Assoziationsbereichs. Die Annotationen der Transkripte, die im Assoziationsbereich kodiert sind oder diesen Bereich mit einem Intron umspannen, sind gelb hervorgehoben. Pfeile repräsentieren Introns in annotierter Strangorientierung, vertikale rote Striche repräsentieren Exons; Mb, Megabasenpaare; kb, Kilobasenpaare.

3.6.2 Suche nach offenen Leserahmen in den Transkripten

Mit dem Ziel, offene Leserahmen (ORF) in den Transkripten zu ermitteln, wurden aus den in den Datenbanken hinterlegten cDNA-Sequenzinformationen mithilfe des ExPASy *Translate Tools* (web.expasy.org/translate) *in silico* die Aminosäuresequenzen (AS-Sequenzen) abgeleitet (siehe Anhang). Abbildung 14 gibt einen Überblick über die möglichen ORFs der Transkripte, die eine Mindestlänge von 90 bp aufwiesen. Unter diesen Voraussetzungen konnte für jedes Transkript mindestens ein ORF gefunden werden.

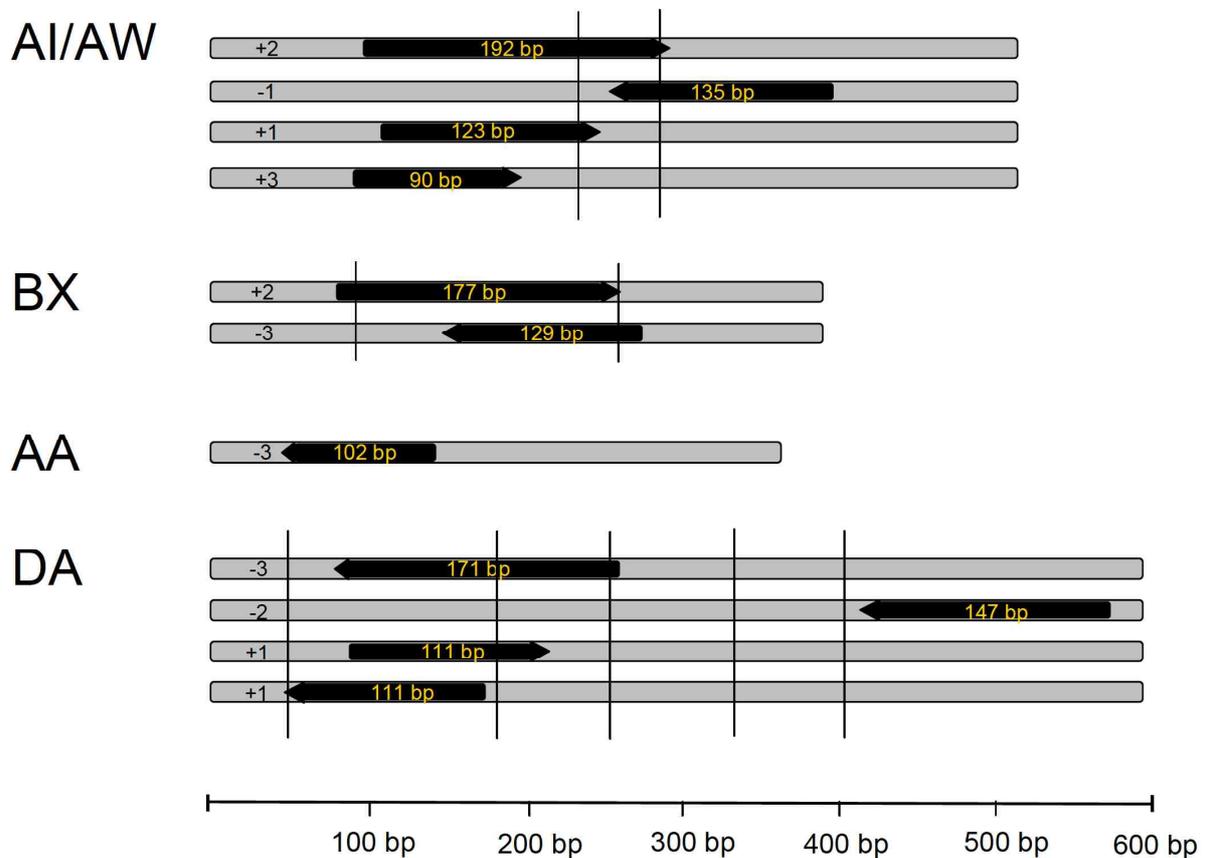


Abbildung 14: Putative offene Leserahmen (ORF) abgeleitet aus cDNA-Sequenzen

Dargestellt sind mögliche ORFs ab einer cDNA-Sequenzlänge von 90 bp. Die Sequenzlänge der jeweiligen Transkripte ist an der unten angegebenen Skala abzulesen. Graue Balken stehen für die gesamte Transkriptlänge, schwarze Balken markieren die Position eines ORFs innerhalb des Transkripts. Die Sequenzlänge des ORFs ist innerhalb der schwarzen Balken notiert. Die Zahlen in den grauen Balken geben die jeweiligen Leserahmen an. Senkrechte Striche durch die Transkripte geben die Positionen von Spleißstellen wieder. AI/AW, AI218168/AW665463; BX, BX091678; AA, 861169; DA, DA227815; cDNA, *complementary DNA*; bp, Basenpaar.

Die Sequenzen der putativen Peptide aus den in Abbildung 14 abgebildeten ORFs wurden verwendet, um nach Sequenzidentitäten zu weiteren Proteinen in der Proteindatenbank *blastp* des NCBI zu suchen. Eine Übersicht der Peptide, die unter den vorgegebenen Sucheinstellungen von *blastp* eine Sequenzidentität von mehr als 50% ihrer AS-Sequenz zu annotierten Proteinen aufwies, findet sich in Tabelle 14. Mit dem Stand von März 2012 konnte für den ORF -1 von AI/AW ein Protein mit 98% Sequenzidentität gefunden werden. Die Sequenzübereinstimmung umfasst 43 von insgesamt 45 AS des humanen hypothetischen Proteins hCG2041770. Für den ORF +3 von DA konnten drei Proteine mit jeweils 66, 72 und 58% Sequenzidentität gefunden werden. Die Suche zeigte unter den gewählten Bedingungen keine Sequenzidentität zu annotierten Proteinen für die ORFs aus BX und AA.

Bei AI/AW und DA handelt es sich um prozessierte Transkripte entsprechend der GT-AG-Regel für Exon-Intron-Grenzen. Die gefundenen ORFs beider Transkripte erstrecken sich jedoch nicht über die gesamte prozessierte Sequenz, sondern nur über Teilsequenzen der gespleißten cDNA.

Tabelle 14: Proteine mit einer Sequenzidentität von > 50 % zu den offenen Leserahmen der Transkripte aus *blastp*

Transkript	ORF	Peptidlänge [AS]	Protein	Organis- mus	Protein- länge [n AS]	Sequenz- identität [%]
AI/AW	-1	64	<i>hCG2041770*</i>	<i>Homo sapiens</i>	45	98
DA	+3	37	<i>hCG1999602*</i>	<i>Homo sapiens</i>	148	66
DA	+3	37	<i>Transcription cofactor vestigial-like protein 1</i>	<i>Pan troglodytes</i>	266	72
DA	+3	37	<i>LOC100605180*</i>	<i>Normascus leucogenys</i>	148	58

ORF, Offener Leserahmen; AS, Aminosäuren; n, Anzahl; AI/AW, AI218168/AW665463; DA, DA227815; * hypothetisches Protein.

3.6.3 Sequenzanalyse von cDNA-Klonen der Transkripte

Über die Firma *Source Bioscience* wurden cDNA-Inserts der Transkripte AI/AW und AA jeweils in einem pT7T3D-Pacl-Vektor in *E. coli* DH10B erworben. Die cDNA-

Inserts von BX und DA waren nicht erhältlich. Für eine Sequenzanalyse wurden die erworbenen cDNA-*Inserts* sequenziert und mit der in den Datenbanken UCSC und NCBI hinterlegten Information abgeglichen.

Die Sequenz und Länge des cDNA-*Inserts* von AA stimmten mit den in den Datenbanken hinterlegten Sequenzen überein. Die Sequenz des cDNA-*Inserts* von AI jedoch zeigte eine abweichende Länge von 794 nt und entsprach einer zusammengeführten Sequenz aus AI/AW und BX. Eine Orientierung der Sequenzinformation auf dem Minusstrang zeigte zudem Spleißstellen entsprechend der GT-AG-Regel. Dies gibt erneut einen Hinweis darauf, dass das Transkript AI/AW, entgegen seiner Annotation von UCSC, möglicherweise auf dem Minusstrang kodiert ist (siehe Tabelle 13) und zusätzlich mit BX ein zusammenhängendes Transkript bilden könnte. Die Transkripte AI/AW und BX werden im Folgenden als zusammengehöriges Transkript (AI/AW/BX) dargestellt, wobei nicht entschieden werden kann, ob Spleißvarianten oder Artefakte der cDNA vorliegen.

3.6.4 Gewebespezifische Genexpression der *Expressed Sequence Tags*

Es wurden 20 verschiedene Gewebe sowie die Endothelzelllinie HPMEC (*human pulmonary microvascular endothelial cells*) ausgewählt, um eine gewebespezifische Genexpression der Transkripte in kommerziell erworbener sowie selbst synthetisierter cDNA zu untersuchen. Zunächst wurden mithilfe des Referenzgens *Protein Phosphatase 1 (PPP1CA)* über eine Exon-übergreifende Polymerase Kettenreaktion (*Polymerase chain reaction*, PCR) die cDNAs aller zu untersuchenden Gewebe auf eine Kontamination mit genomischer DNA (gDNA) geprüft. Es wurden lediglich cDNAs für weitere Analysen verwendet, die keine gDNA-Kontamination aufwiesen.

Für die Expressionsanalysen wurden jene Exons der Transkripte ausgewählt, die mindestens eine Länge von 100 bp besaßen, um optimale Bedingungen für den Entwurf von Oligonukleotiden für eine PCR zu schaffen und eine Spezifität der Amplifikationsprodukte nach erfolgter Sequenzierung gewährleisten zu können. Für das Transkript DA wurden dementsprechend Fragmente aus Exon 2 und 6, für AI/AW/BX solche aus Exon 2, 3 und 5 gewebespezifisch untersucht. Für das Transkript AA ist ein solches aus dem einzigen Exon analysiert worden (Tabelle 15).

Tabelle 15: Länge der Exons und Introns der Transkripte im Assoziationsbereich

	DA [bp]	AA [bp]	AI/AW/BX [bp]
	Expressionsgewebe [NCBI-Datenbank]		
	Gehirn	Hoden	fötale Lunge, Hoden, B-Zellen
Exon 1	49	356	92
Intron 1	31463		4211
Exon 2	130		176
Intron 2	327590		63727
Exon 3	72		237
Intron 3	391		938
Exon 4	82		58
Intron 4	3147		455
Exon 5	64		231
Intron 5	91		
Exon 6	182		

Die Länge der Exons und Introns wurden anhand der annotierten cDNA-Sequenz der Transkripte von UCSC ermittelt. Exons, für die eine Nachweis-PCR etabliert worden ist, sind fett dargestellt. bp, Basenpaare; DA, DA227815; AA, AA861168; AI/AW/BX, AI21868/AW665463/BX091678.

Eine Zusammenfassung der Ergebnisse der qualitativen Analyse ist in Tabelle 16 dargestellt. Es fällt auf, dass für das Gen des mutmaßlich zusammenhängenden Transkripts AI/AW/BX die geprüften Exons weder einheitlich noch zusammenhängend exprimiert werden. Die in den Datenbanken für dieses Transkript beschriebene Expression in Hodengewebe wurde lediglich für Exon 5 bestätigt.

Eine Expression des Gens von AA wurde in neun unterschiedlichen Geweben und der Endothelzelllinie HPMEC nachgewiesen. Die in den Datenbanken beschriebene Expression in Hodengewebe wurde bestätigt.

Für das Gen des Transkripts DA wurde für Exon 2 und 6 in sieben unterschiedlichen Geweben ein identisches Expressionsprofil beobachtet. Die beschriebene Expression in Gehirngewebe wurde zumindest in der cDNA eines der zwei cDNA-Isolate bestätigt.

Tabelle 16: Gewebespezifische Expression der Transkripte

	Arterie [#]	Vene [#]	Leukozyten [#]	Haut [#]	Gehirn [#]	Gehirn ⁺	Hoden ⁺	Schilddrüse ⁺	Herz ⁺	Dünndarm ⁺	Lunge [#]	Lunge ⁺	Milz ⁺	Bauchspeicheldrüse ⁺	Skelettmuskel ⁺	Niere ⁺	Leber [#]	Leber ⁺	Fettgewebe ⁺	Plazenta ⁺	Ovarien ⁺	Dickdarm ⁺	Beinvene [*]	HPMEC [*] , Zelllinie
AI/AW/BX, Exon 5	+	+	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-
AI/AW/BX, Exon 3	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AI/AW/BX, Exon 2	+	-	+	+	+	+	-	-	+	-	-	-	-	+	-	+	+	-	-	+	+	+	+	-
AA	+	+	+	+	-	-	+	-	+	-	+	+	-	-	-	+	+	+	-	-	-	-	-	+
DA, Exon 2	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	+	-	+
DA, Exon 6	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	+	-	+
<i>PPP1CA</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Aqua bidest</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Als Positivkontrolle diente das Referenzgen Protein Phosphatase 1 (*PPP1CA*), für die Negativkontrolle wurde *Aqua bidest* verwendet. [#] cDNA der Firma BioChain; ⁺ cDNA der Firma PrimerDesign; * cDNA, synthetisiert aus isolierter RNA; HPMEC, *human pulmonary microvascular endothelial cell*.

Die Untersuchungen zeigten somit eine gewebespezifische Expression für alle Transkripte. Während Doppeluntersuchungen in cDNA aus Lungengewebe identische Ergebnisse ergaben, wurden bei Doppeluntersuchungen in cDNAs der Gewebsformen Arterie, Gehirn und Leber nicht immer konforme Ergebnisse erzielt.

3.6.5 Quantifizierung der Transkriptmengen

Zur Bestimmung der relativen Expression der Gene von AA/AW/BX, AA und BX wurden quantitative *Real-Time* Polymerase-Kettenreaktionen (qRT-PCRs) durchgeführt und deren Ergebnisse mit denen von zwei Referenzgenen verglichen. Die qRT-PCRs wurden für jedes der zu untersuchenden Transkripte und die Referenzgene durchgeführt. Um einen Ausschluss von gDNA-Kontamination zu gewährleisten, wurden die PCR-Oligonukleotide der Referenzgene so gewählt, dass sie in unterschiedlichen Exons binden. Die Berechnung der relativen Expression erfolgte anhand der unter 2.3.2.5 angegebenen Formel, wobei die Effizienz der

entsprechenden PCR-Oligonukleotidpaare zuvor bestimmt und zur Korrektur verwendet wurden. Die quantitativen Expressionsstudien der Gene wurden in den Geweben durchgeführt, in denen zuvor qualitativ eine Expression nachgewiesen wurde (siehe Tabelle 16). Hierfür dienten die gleichen PCR-Oligonukleotidpaare, die bereits zuvor Anwendung fanden. Mit dem PCR-Oligonukleotidpaar, das im 3. Exon des Gens von AI/AW/BX bindet, konnten jedoch mithilfe des Reagenziensystems *LightCycler® 480 SYBR Green I Master* von ROCHE keine Amplifikate erzeugt werden. Ein quantitativer Nachweis der Expression dieses Gens war jedoch anhand von PCR-Oligonukleotidpaaren, die innerhalb des 2. sowie des 5. Exons dieses Transkripts binden, möglich.

Als Referenzgene für die Expressionsstärke wurden *Beta Aktin (ACTB)* und *PPP1CA* gewählt, da sie unterschiedliche Expressionslevels besitzen. *ACTB* repräsentiert ein Gen mit starkem, *PPP1CA* ein Gen mit geringem Expressionsniveau. Die Expression beider Referenzgene konnte in allen untersuchten Geweben nachgewiesen werden. Die Expression der Gene, die für die Transkripte kodieren, konnte nicht in allen Geweben nachgewiesen werden (Abbildung 15). Dieses Problem ließ sich auch durch die Verwendung verschiedener cDNA-Konzentrationen nicht beheben.

Abbildung 15 zeigt die Expression der Gene aller Transkripte. Von AI/AW/BX wurden Expressionsanalysen der Exons 2 und 5 durchgeführt, von DA wurden Exon 2 und 6 untersucht. Vom ungespleißten Transkript AA wurden Expressionsanalysen von einem Fragment des einzigen Exons der cDNA durchgeführt.

Die Analysen zeigten in den untersuchten Geweben sehr schwache Expressionsniveaus aller Gene, die für die Transkripte kodieren. Für alle Transkripte konnte in Relation zu *ACTB* eine zwischen zwei bis acht Potenzen schwächere Expression beobachtet werden. Zu *PPP1CA* wurden Expressionsstärken nachgewiesen, die von einem äquivalenten Niveau bis zu einem um fünf Potenzen schwächeren Niveau reichten. Zusammenfassend wurden die höchsten Expressionsniveaus in vaskulärem und/oder zerebralem Gewebe nachgewiesen. Das höchste Expressionsniveau fand sich beim Transkript AA aus cDNA der endothelialen Zelllinie HPMEC. Die niedrigsten Expressionsniveaus wurden für das zweite Exon von DA aus Plazenta- und Beinvenengewebe gefunden (Abbildung 15).

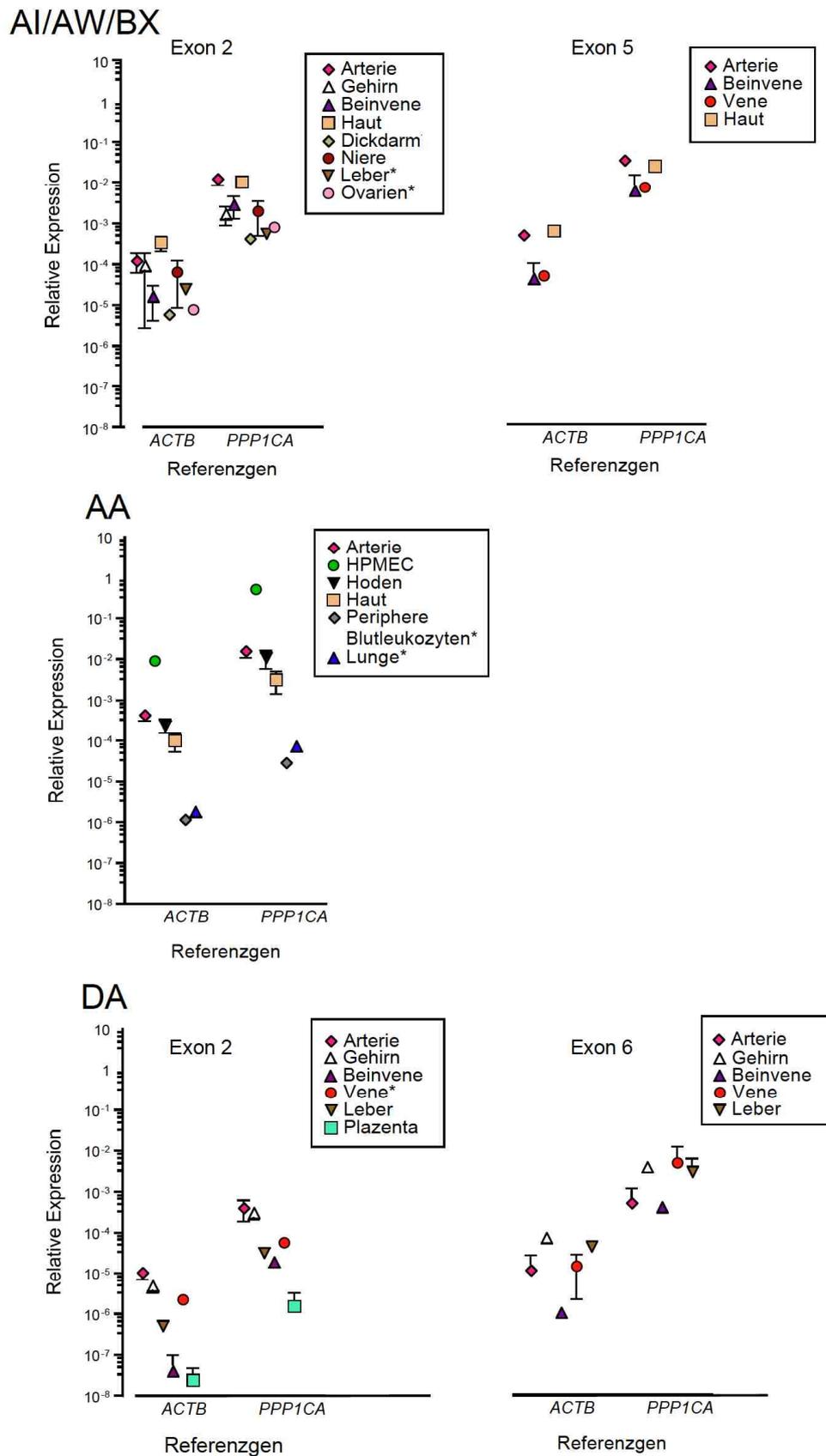


Abbildung 15: Expressionsstärke der Gene der Transkripte

Dargestellt sind jeweils die relativen Expressionen der Gene der Transkripte zu den Referenzgenen Beta Actin (*ACTB*) und Protein Phosphatase 1 (*PPP1CA*). Von AI/AW/BX und DA wurden die relativen

Expressionen in zwei Exons untersucht. Angegeben ist jeweils der Mittelwert der relativen Expression des Gens zu *ACTB* und *PPP1CA* aus mindestens einem Duplex-Ansatz einer quantitativen *Real-Time* PCR. Ausnahmen, die auf Einzelbestimmungen basieren, wurden mit * versehen. Für AI/AW/BX konnte aus dem 2. Exon aus peripheren Bluteukozyten sowie aus Gewebe aus Herz, Bauchspeicheldrüse und Plazenta kein Expressionsprofil des Gens erstellt werden. Aus dem 5. Exon konnte aus peripheren Bluteukozyten, Leber- und Hodengewebe ebenfalls kein Expressionsprofil des Gens erstellt werden. Weiterhin konnte für AA aus Venen-, Schilddrüsen-, Nieren- und Lebergewebe kein Expressionsprofil des Gens erstellt werden. Vom Transkript DA konnte anhand von PCR-Oligonukleotiden, die im 2. Exon binden, kein Expressionsprofil des Gens in Dickdarmgewebe erstellt werden. Aus dem 6. Exon konnte kein Expressionsprofil aus Dickdarm- sowie Plazentagewebe gewonnen werden. HPMEC, *human pulmonary microvascular endothelial cell*.

3.6.6 Analysen zur Länge der cDNA-Sequenz der Transkripte

Da es sich bei den annotierten cDNA-Sequenzen der Transkripte um unvollständige Teilsequenzen handeln konnte, wurden Untersuchungen zur vollständigen cDNA-Sequenz der jeweiligen Transkripte über 3'- und 5'-*Rapid Amplification of cDNA Ends* (RACE)-PCRs durchgeführt. Die Amplifikation der korrekten 3'- und 5'-Enden eines Transkriptes erwies sich als schwierig, vermutlich durch die geringe Transkriptanzahl (Abbildung 15) und einen hohen Anteil an repetitiven Sequenzen im zu amplifizierenden Bereich. Um spezifische Amplifikate zu erhalten, wurden Versuche zur Optimierung der Bedingungen der RACE-PCR vorgenommen, die folgende Parameter betrafen: *Annealing*- und *Extension*-Temperatur, Magnesiumkonzentration, eingesetzte cDNA-Menge und Art der DNA-Polymerase. Zusätzlich wurden je nach verwendeter DNA-Polymerase die PCR-Zusätze *Solution S* und *Q-Solution* verwendet, die eine Amplifikation von GC-reichen DNA-Matrizen erleichtern. Trotz der gewählten Veränderungen der PCR-Reaktionsansätze konnte nur in wenigen Geweben eine Bestimmung des 3'- und/oder 5'-Endes erfolgen.

AI/AW/BX

Die vollständige cDNA-Sequenz des AI/AW/BX-Transkripts konnte nicht ermittelt werden. In 5'-Richtung jedoch konnte in cDNA der Gewebe Arterie, Vene und Haut jeweils ein alternativer Spleißprozess zu einem bisher nicht beschriebenen, 239 bp langen Exon gefunden (Abbildung 16). Während in vaskulärem Gewebe ein Spleißprozess vom annotierten 3. Exon zum unbekanntem Exon erfolgte, war in Hautgewebe ein Spleißprozess vom annotierten 4. Exon zum gleichen unbekanntem Exon festzustellen (Abbildung 16). In Arterien- und Beinvenengewebe wurden

zusätzlich zwei unterschiedliche Genprodukte gefunden, die *antisense* zum annotierten Transkript liegen (Abbildung 16). In arteriellem Gewebe umspannte das Gen das annotierte 2. Exon, im Gewebe der Beinvene umspannte es das annotierte 5. Exon. In diesen Transkripten zeigte sich kein Spleißvorgang. Das in cDNA aus arteriellem Gewebe gefundene Transkript wies ein konventionelles Polyadenylierungssignal (5'-AAUAAA-3') auf (Mandel *et al.*, 2006). Eine vollständige Amplifikation der cDNA-Sequenzen wurde nicht erreicht, jedoch konnten überlappende Sequenzen der 3'- und 5'-RACE-PCR amplifiziert werden.

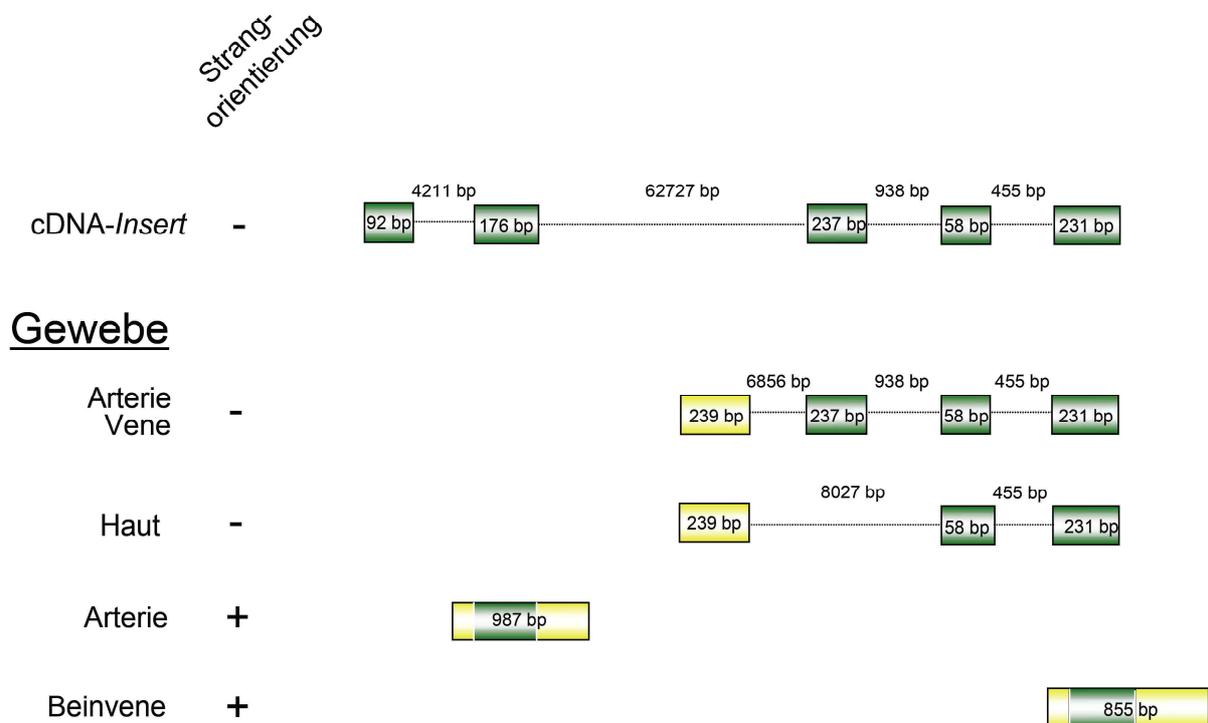


Abbildung 16: Schematische Abbildung der durch *Rapid Amplification of cDNA Ends (RACE)-PCR* in cDNA aus Arterie, Vene, Haut und Beinvene ermittelten 3'- und 5'-Enden des Transkripts AI/AW/BX im Vergleich zum cDNA-Insert von AI

Das Transkript entsprechend der cDNA-Sequenz des käuflich erworbenen cDNA-Inserts im Vergleich zur cDNA-Sequenz der AI/AW/BX-Transkripte, die in Arterien-, Haut-, und Beinvenengewebe nachgewiesen wurden. Die Exons aus der cDNA-Sequenz des *Inserts* sind in grünen Blöcken dargestellt, verbindende Striche stehen für Introns. Übereinstimmungen in der cDNA-Sequenz des cDNA-Inserts und der neuen Transkripte sind ebenfalls in Grün angegeben. Gelbe Blöcke zeigen vom cDNA-Insert abweichende Isoformen der nachgewiesenen Transkripte. Zusätzlich ist die vermutliche Strangorientierung der cDNA-Sequenz angegeben. AI/AW/BX, AI218168/AW665463/BX091678; bp, Basenpaar.

Aufgrund des Auftretens der konventionellen Spleißstellen entsprechend der GT-AG-Regel ist anzunehmen, dass die Gene der alternativen RNA-Isoformen aus Haut-

und Arterien- und Venengewebe auf dem Minusstrang kodiert sind. Die Ergebnisse der *CapFinder*-Methode und der 3'-RACE der ungespleißten Transkripte aus dem Gewebe der Beinvene und der Arterien weisen im Gegensatz dazu auf eine Orientierung auf dem Plusstrang hin.

AA

Die Bestimmung der 3'- und 5'-Enden des AA-Transkripts erfolgte in Arterien- und Hodengewebe.

In Arterien- und Hodengewebe wurden zwei 5'-Enden und ein 3'-Ende identifiziert (Abbildung 17). Die 5'-Enden wurden mit dem 5'-3'-RACE-Reagenziensystem von ROCHE ermittelt, da mit der *CapFinder*-Methode eine Amplifikation erfolglos blieb. Die Sequenz des annotierten 3'-Endes konnte um 89 nt verlängert werden. Es befindet sich kein Polyadenylierungssignal am Ende der Sequenz, jedoch eine Folge von sieben Adeninen. Aufgrund der komplementären Basenpaarungen des für die RACE-PCR verwendeten Oligo(dT)-Nukleotids zum poly(A)-Ende des Transkripts besteht die Möglichkeit, dass das Oligonukleotid während der cDNA-Synthese an diese Adeninreiche Sequenz gebunden hat und somit zu einem verkürzten 3'-Ende führte. Um dies auszuschließen wurde eine PCR durchgeführt, mit der ein Transkript nachgewiesen werden kann, das sich über das möglicherweise verkürzte 3'-Ende hinaus erstreckt. Mithilfe dieser PCR konnte jedoch kein längeres 3'-Ende für das Transkript ermittelt werden.

In Hodengewebe wurde ein 3'-Ende gefunden, für welches ein konventionelles Polyadenylierungssignal (5'-AAUAAA-3') in der cDNA-Sequenz identifiziert werden konnte (Mandel *et al.*, 2006). Dieses 3'-Ende ist 44 bp kürzer als das in arteriellem Gewebe nachgewiesene 3'-Ende des Transkripts. Das 5'-Ende des Transkripts in Hodengewebe konnte weder mit der *CapFinder*-Methode, noch mit dem Reagenziensystem von ROCHE ermittelt werden. Es konnte jedoch ein Amplifikat nachgewiesen werden, welches sich über das in Hodengewebe gefundene 3'-Ende und über das in arteriellem Gewebe gefundene 5'-Ende der Isoform 1 hinaus erstreckt (Abbildung 17). Die Amplifikation einer vollständigen cDNA des AA-Transkripts war in keinem Gewebe möglich.

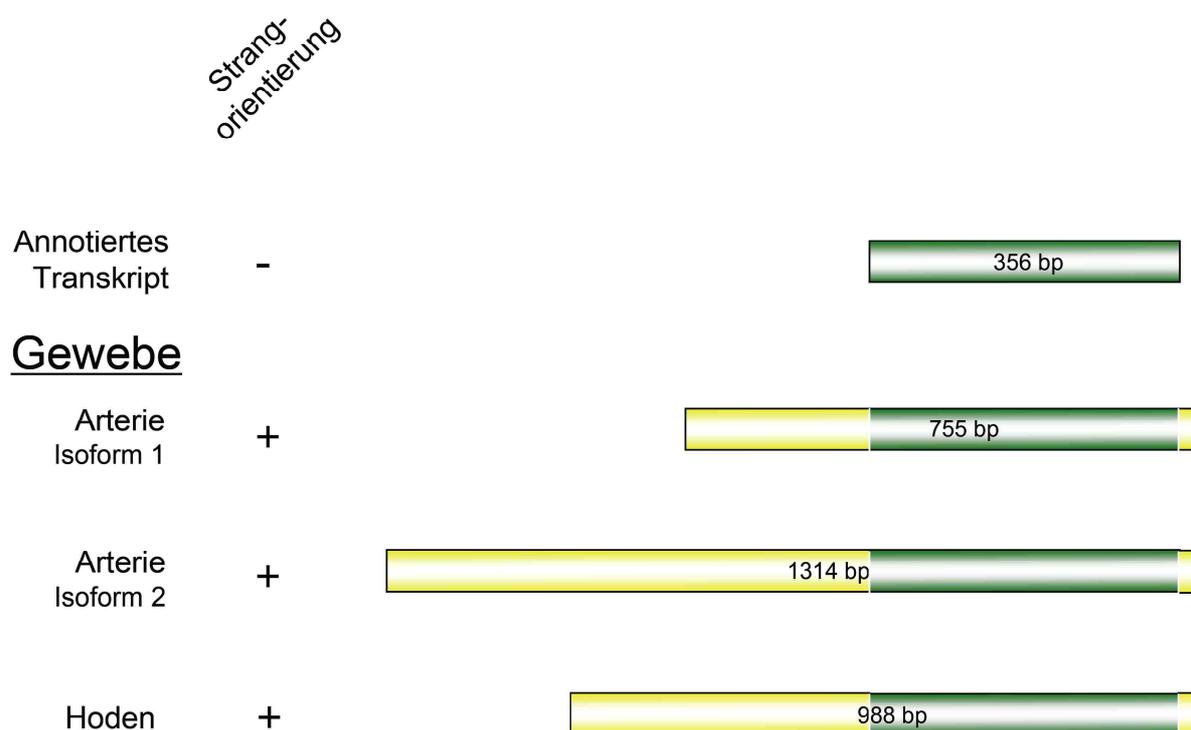


Abbildung 17: Schematische Abbildung der durch *Rapid Amplification of cDNA Ends (RACE)-PCR* in cDNA aus Arterie und Hoden ermittelten 3'- und 5'-Enden des Transkripts AA im Vergleich zum annotierten Transkript

Das annotierte Transkript AA entsprechend der in der UCSC-Datenbank hinterlegten cDNA-Sequenz im Vergleich zur cDNA-Sequenz der Transkripte von AA, die in Arterien- und Hodengewebe nachgewiesen wurden. Die annotierte cDNA-Sequenz ist Grün dargestellt. Übereinstimmungen in der cDNA-Sequenz des annotierten Transkripts und der neuen Transkripte sind ebenfalls Grün. Gelbe Blöcke zeigen vom annotierten Transkript abweichende Sequenzen. Zusätzlich ist die Strangorientierung der cDNA-Sequenz angegeben. bp, Basenpaar; AA, AA861169.

DA

Für das Gen des Transkripts DA konnte in zerebralem und arteriellem Gewebe jeweils ein 5'-Ende ermittelt werden (Abbildung 18). Beide 5'-Enden liegen im annotierten 2. Exon, sie unterscheiden sich jedoch in ihrer Länge. Während das in Gehirngewebe nachgewiesene 5'-Ende 70 nt des annotierten 2. Exons umfasst, beträgt das 5'-Ende aus arterieller cDNA 113 nt. In beiden Geweben konnten ebenfalls 3'-Enden ermittelt werden. Das 3'-Ende aus arteriellem Gewebe besitzt anschließend an das annotierte 6. Exon ein weiteres, 32 nt langes, 7. Exon mit konventionellem Polyadenylierungssignal (Abbildung 18).

In Gehirngewebe konnte durch die 5'-RACE eine alternative Spleißvariante detektiert werden, die eine Exzision des 4. Exons aufweist (Abbildung 18). Ein 3'-Ende konnte

mit der Exzision des annotierten 6. Exons nachgewiesen werden. Diese cDNA besitzt das 4. Exon und zeigt somit zwei mögliche Isoformen des Gens in diesem Gewebe. Die cDNA-Sequenz des neuen 3'-Endes dieser alternativen Spleißvariante endet an einer poly(A)-reichen Sequenz.

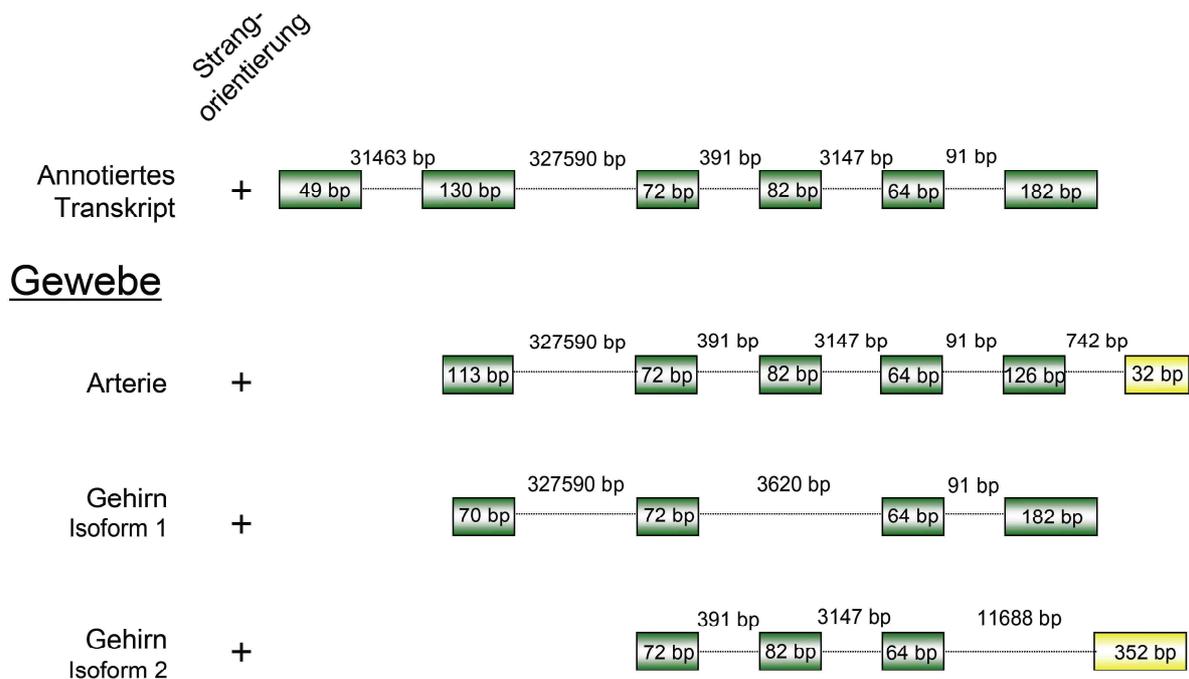


Abbildung 18: Schematische Abbildung der durch *Rapid Amplification of cDNA Ends (RACE)-PCR* in cDNA aus Arterie und Gehirn ermittelten 3'- und 5'-Enden des Transkripts DA im Vergleich zum annotierten Transkript

Das annotierte Transkript DA entsprechend der in der UCSC-Datenbank hinterlegten cDNA-Sequenz im Vergleich zum Aufbau der Transkripte, die in Arterien- und Gehirngewebe nachgewiesen wurden. Die Exons aus der annotierten cDNA-Sequenz des Transkripts sind in grünen Blöcken dargestellt, verbindende Striche stehen für Intronsequenzen. Übereinstimmungen in der cDNA-Sequenz des annotierten Transkripts und der neuen Transkripte sind ebenfalls in Grün angegeben. Gelbe Blöcke zeigen von der annotierten cDNA-Sequenz abweichende Isoformen der nachgewiesenen Transkripte. Zusätzlich ist die Strangorientierung der cDNA-Sequenz angegeben. DA, DA227815; bp, Basenpaar.

Es wurde für keines der Transkripte eine vollständige cDNA amplifiziert jedoch überlappen sich die Sequenzen der RACE-PCRs aus Arterien- und Gehirngewebe, so dass davon auszugehen ist, dass diese ein zusammenhängendes Transkript bilden. Die mittels 5'-RACE-PCR gewonnene cDNA aus Gehirn stimmt aufgrund der Exzision des 4. Exons nicht mit der Sequenz der 3'-RACE überein. Es muss sich daher um zwei Isoformen handeln, deren jeweiliges 3'- bzw. 5'-Ende es noch zu identifizieren gilt.

4 Diskussion

4.1 Populationsidentifikation

Populationen in afrikanischen Regionen können sehr heterogene Ethnien aufweisen (Campbell & Tishkoff, 2008; Sikora *et al.*, 2011; Teo *et al.*, 2010). Die Population dieser genomweiten Assoziationsstudie (GWA-Studie) konnte nach Selbstauskunft und Namensanalyse in die drei verschiedenen Ethnien Akan, Ewe und Ga-Adangbe sowie einer Gruppe von nicht weiter spezifizierten Ethnien, die als *Northerner* bezeichnet wurde, unterteilt werden. Da ethnische Stratifikation in Assoziationsstudien zu unrichtigen Ergebnissen führen kann (Lander & Schork, 1994), wurde die GWA-Studiengruppe zusätzlich auf genetischer Basis ausgewertet. Dazu wurden die 774800 Genotypen der Individuen mithilfe der multidimensionalen Skalierung (MDS) analysiert, einer Methode, die auch genetische Heterogenität darstellen kann (Kruskal, 1978). Diese Untersuchung lässt eine Populationsstratifikation in der GWA-Studiengruppe erkennen. Es konnte dabei gezeigt werden, dass bei der Mehrzahl der Individuen der Ethnien Akan und der Gruppe *Northerners* die Zuordnung durch ethnische Selbstauskunft mit der Zuordnung der *Cluster* der MDS-Analyse weitgehend übereinstimmte. Andere Individuen der Ethnien Ewe, Ga-Adangbe und viele Individuen der Gruppe *Northerners* zeigten jedoch eine abweichende und inhomogenere Verteilung, so dass sich durch eine ausschließliche Berücksichtigung der ethnischen Selbstauskunft die Populationsstratifikation nicht ausreichend beschreiben ließ.

Um eine Korrektur auf Populationsstratifikation vorzunehmen, wurden die Variablen der MDS-Analyse der GWA-Studiengruppe in die Assoziationsberechnungen der GWA-Studie einbezogen. Ein Auftreten von falsch-positiven Assoziationen aufgrund von Populationsstratifikation erscheint in der GWA-Studiengruppe daher eher unwahrscheinlich. Bedingt durch die fehlende genomweite Typisierung lagen für die Replikationsgruppe keine MDS-Variablen vor. Es ist deshalb nicht auszuschließen, dass in der Replikationsgruppe sowie in der kombinierten Gruppe die Verwendung der Ethnien zur Korrektur auf Populationsstratifikation zu Ungenauigkeiten in der Assoziationsanalyse geführt hat. In der GWA-Studiengruppe können falsch-positive Assoziationen allerdings zufällig durch multiples Testen, Genotypisierungsfehler oder weitere unentdeckte *Confounder* generiert worden sein (Thomas *et al.*, 2009).

4.2 Assoziationsanalyse

Die Analyse der Genotypen von 774800 Einzelnukleotidaustauschen (SNPs) in der GWA-Studiengruppe zeigte zunächst eine Assoziation von 13 unabhängigen Loci mit zerebraler Malaria mit einem p-Wert von $p < 5 \times 10^{-5}$. Zwei dieser Loci lagen in der Nähe des bekannten Resistenzgens Hämoglobin B (*HBB*) und wurden in einer anderen GWA-Studie zu schwerer Malaria weiter bearbeitet. Von den verbleibenden 11 Loci unterschritt lediglich der SNP rs12479448 (A/T) nach der Analyse der Genotypen der gesamten Studiengruppe den Schwellenwert für eine genomweite Signifikanz ($p = 5 \times 10^{-8}$; Pe'er *et al.*, 2008; Pearson & Manolio, 2008). Die stärkste Assoziation wurde im additiven Vererbungsmodell mit einem p-Wert von $2,4 \times 10^{-8}$ (*Odds Ratio* [OR] = 1,6; 95 % Konfidenzintervall [KI] = 1,4 – 1,9) gefunden. Träger des Allels A tragen ein 1,6-fach erhöhtes Risiko, an zerebraler Malaria zu erkranken, als Träger des Allels T. Das krankheitsassoziierte Allel A findet sich in der Fallgruppe mit einer Häufigkeit von 21,9 % und in der Kontrollgruppe mit einer Häufigkeit von 15 %.

Zusätzlich zur populationsbasierten Assoziationsanalyse konnte ein familienbasierter Test, der Transmissions-Disequilibriumstest (TDT), der Populationsstratifikationen gegenüber robust ist (Spielman *et al.*, 1993), die Assoziation des SNPs mit zerebraler Malaria bestätigen. Um die Teststärke des TDTs zu erhöhen, wurden neben 169 Trios auch 166 Duos in die Analyse eingeschlossen. Der fehlende Genotyp eines Elternteils wird in Duos jedoch statistisch geschätzt und kann daher zu geringen Verzerrungen in den Aussagen zur Alleltransmission führen (Sun *et al.*, 1999).

In der Kontrollgruppe der ghanaischen Studienpopulation wurde das schützende Allel T mit einer Frequenz von 85 % gefunden. In weiteren afrikanischen und asiatischen Populationen ist das schützende Allel T mit einer Häufigkeit von 72 - 91 % vertreten (Yoruba, Nigeria = 85 %; Luhya, Kenia = 85 %; Maasai, Kenia = 72 %; Beijing, China = 86 %; Tokio, Japan = 91 %; <http://www.ncbi.nlm.nih.gov/projects/SNP>). In zwei kaukasischen Populationen liegt die Frequenz des T-Allels bei 44 bzw. 47 % (Individuen aus Utah mit nord- und westeuropäischer Abstammung = 47 %; Toskana, Italien = 44 %; <http://www.ncbi.nlm.nih.gov/projects/SNP>). Die hohe Frequenz der schützenden Variante in afrikanischen sowie asiatischen Populationen könnte auf

eine positive Selektion des Allels in Malaria-endemischen Gebieten hinweisen, wie sie auch für Allele bekannter Resistenzgene beschrieben wurde (Miller, 1996).

Der mit zerebraler Malaria assoziierte SNP rs12479448 liegt auf Chromosom 2p22.3 in einer intergenischen Region ohne annotierte Protein-kodierende oder regulatorische Einheiten. Für rs12479448 wurde bisher keine Assoziation mit einer anderen Erkrankung beschrieben. Ebenfalls findet sich in der Literatur keine weitere krankheitsassoziierte Variante innerhalb der umliegenden 100 kb. Die Distanz von rs12479448 zum nächstgelegenen annotierten Protein-kodierenden Gen beträgt 1,3 Mb.

Dass in dieser GWA-Studie eine intergenische Variante gefunden wurde, entspricht den Ergebnissen vieler anderer GWA-Studien, die ebenfalls krankheitsassoziierte genetische Varianten in intergenischen oder intronischen Bereichen beschreiben (Hindorff *et al.*, 2009; Slavin *et al.*, 2011; Thye *et al.*, 2010). Es wird vermutet, dass in diesen Regionen regulatorische Einheiten kodiert sind, die einen Einfluss auf die Expression von umliegenden Genen besitzen könnten (Cooper & Shendure, 2011).

4.3 Analysen zur Sensitivität der genomweiten Assoziationsstudie

Im Hinblick auf eine Teststärke von 80 % können in dieser GWA-Studie Assoziationen von SNPs mit einer Minor-Allelfrequenz von 5 % und einem relativen Risiko (RR) von 1,67 mit zerebraler Malaria nachgewiesen werden. Da es sich bei zerebraler Malaria um eine seltene Erkrankung handelt, ist das RR in diesem Fall mit der OR vergleichbar (Davies *et al.*, 1998). Die Wahrscheinlichkeit, eine krankheitsassoziierte Variante in einer GWA-Studie zu finden, wird außerdem durch die *genome coverage* (dtsh. genomische Abdeckung) der verwendeten Genotypisierungsplattform beeinflusst. *Genome coverage* wird durch die Stärke des Kopplungsungleichgewichts (*linkage disequilibrium*, LD) der genotypisierten Varianten zu den verbleibenden genetischen Varianten, die in einer Population vertreten sind, bestimmt. Je stärker das LD zwischen den genotypisierten SNPs zu den verbleibenden genetischen Varianten ist, desto höher ist die *genome coverage* einer Genotypisierungsplattform. Aus diesem Grund kann die *genome coverage* einer Genotypisierungsplattform zwischen Populationen unterschiedlich hoch ausfallen. Da afrikanische Populationen im Vergleich zu kaukasischen und asiatischen Populationen über ein geringes LD verfügen (Li *et al.*, 2008a), könnte der

Affymetrix® Genome-wide SNP Array 6.0 in einer Population wie der GWA-Studiengruppe lediglich eine schwache *genome coverage* erzielen (Frazer *et al.*, 2007). Die verwendete Genotypisierungsplattform wurde nach den Erkenntnissen des HapMap-Projektes entworfen, dessen Priorität es war, SNPs zu genotypisieren, die sowohl in kaukasischen als auch in asiatischen und afrikanischen Populationen vertreten sind (Clark *et al.*, 2005). Neue Untersuchungen zur genetischen Diversität zeigen jedoch neben bisher unbekanntem SNPs, die ausschließlich in afrikanischen Populationen vertreten sein können, auch solche, die kein LD zu umliegenden Varianten aufweisen (Altshuler *et al.*, 2010; Conrad *et al.*, 2006; Ke *et al.*, 2008). Dadurch sinkt die Wahrscheinlichkeit, in GWA-Studien in afrikanischen Populationen genomweite Signifikanzlevel zu unterschreiten (Teo *et al.*, 2010). Zudem wird gerade bei komplexen Erkrankungen wie Malaria (Garcia *et al.*, 1998a) vermutet, dass ihre Pathologie durch die Interaktion mehrerer Gene hervorgerufen wird, die ihrerseits schwache oder moderate Effektstärke besitzen (Manolio *et al.*, 2009; Risch, 2000). Es ist daher wahrscheinlich, dass die Assoziation von weiteren unbekanntem genetischen Varianten mit zerebraler Malaria in dieser GWA-Studie nicht nachzuweisen war.

So wurde zum Beispiel die Deletion rs8176719 im *ABO*-Locus, die zum Phänotyp der Blutgruppe 0 führt und einen Schutz vor schwerer Malaria vermittelt (Rowe *et al.*, 2007), nicht mit dem *Affymetrix® Genome-wide SNP Array 6.0* genotypisiert. Auch benachbarte SNPs dieser Deletion zeigten unterhalb des gesetzten Schwellenwertes von $p = 5 \times 10^{-5}$ keine Assoziation mit der Erkrankung. Eine Assoziation mit schwerer Malaria konnte jedoch auch in dieser Studiengruppe in einer GWA-Studie durch direkte DNA-Genotypisierung von rs8176719 gezeigt werden (p -Wert = $1,1 \times 10^{-6}$; OR = 1,4; KI = 1,2 – 1,6; Timmann, *Biology and Parasitology of Malaria Parasites*, Heidelberg, 20.05.2009). Der Mangel an Sensitivität dieser Studie, die Assoziation am *ABO*-Locus nachzuweisen, war vermutlich durch ein zu geringes LD der Marker, die mit dem *Affymetrix® Genome-wide SNP Array 6.0* genotypisiert werden, zur kausalen Variante bedingt.

Da die Teststärke neben der Effektstärke unter anderem von der Allelfrequenz beeinflusst wird, konnte die Assoziation dieser Variante mit schwerer Malaria trotz einer geringen Effektstärke aufgrund der hohen Frequenz des schützenden Allels nachgewiesen werden.

Auf Chromosom 11 zeigten zwei SNPs in der Nähe des *HBB*-Gens eine Assoziation der Region mit zerebraler Malaria unterhalb des gewählten Schwellenwertes von $p = 5 \times 10^{-5}$. In der kodierenden Sequenz des *HBB*-Gens liegen nebeneinander die SNPs rs334 und rs33930165, die für Hämoglobin S (*HbS*) bzw. Hämoglobin C (*HbC*) kodieren. Die Schutz-vermittelnden Allele der beiden SNPs wurden nicht mit dem *Affymetrix® Genome-wide SNP Array 6.0* genotypisiert. Eine differenzierte Betrachtung bezüglich ihrer Assoziation mit zerebraler Malaria war daher nur durch eine direkte Genotypisierung mit anschließender statistischer Analyse möglich. In einer Kandidatengenanalyse wurde für den SNP rs33930165, der zu *HbC* führt, eine geringe Effektstärke und Frequenz in der gleichen Studienpopulation nachgewiesen (OR = 0,64; KI = 0,45 – 0,91; $p = 0,03$; Minorallel-Frequenz Fälle = 5,3 %, Minor-Allelfrequenz Kontrollen = 20,35 %), die auch für die GWA-Studie herangezogen wurde (May *et al.*, 2007). Es ist daher anzunehmen, dass die Assoziation dieses Locus in der GWA-Studiengruppe durch *HbS* hervorgerufen wird, da die Schutz-vermittelnde Variante einen starken Effekt besitzt (OR = 0,065; KI = 0,037 – 0,11; $p = 1,5 \times 10^{-21}$; Minor-Allelfrequenz Fälle = 0,53 %; Minor-Allelfrequenz Kontrollen = 7,2 %; Timmann, *Biology and Parasitology of Malaria Parasites*, Heidelberg, 20.05.2009; Williams, 2006). Die Teststärke der GWA-Studie ist vermutlich nicht ausreichend, um in der vorliegenden Studienpopulation eine Assoziation des SNPs, der für *HbC* kodiert, mit zerebraler Malaria nachzuweisen (May *et al.*, 2007).

4.4 Suche nach der kausalen Variante

Es ist unwahrscheinlich, dass es sich bei rs12479448 direkt um die kausale und somit funktionelle Variante handelt, die einen Einfluss auf die Manifestation der zerebralen Malaria hat (Smith & Newton-Cheh, 2009), da in nur wenigen GWA-Studien initial die kausale Variante gefunden wird (McCarthy *et al.*, 2008). Zudem lag das Auswahlkriterium der SNPs, die mit dem *Affymetrix® Genome-wide SNP Array 6.0* genotypisiert werden, nicht in ihrer Funktion (Li *et al.*, 2008b).

Um eine Variante zu finden, die eine stärkere Assoziation als rs12479448 mit zerebraler Malaria zeigt, wurde durch Re-Sequenzierung eines 100 kb Bereiches um rs12479448 in 24 Fällen und 24 Kontrollen ein so genanntes *Fine Mapping* durchgeführt. Von den gefundenen Varianten wurden über die ermittelten Genotypen das LD zu rs12479448 berechnet. Drei Varianten, die ein LD von $r^2 > 0,6$ zu

rs12479448 aufweisen, wurden in der gesamten Studienpopulation auf eine genomweite Assoziation untersucht. Eine stärkere Assoziation als die von rs12479448 konnte jedoch nicht gefunden werden. Die Analyse des LDs von rs12479448 zu den Polymorphismen des sequenzierten Bereichs zeigte zudem wenige Polymorphismen, die in einem starken LD zu rs12479448 stehen. Somit könnte eine Variante mit stärkerer Assoziation als $p = 5 \times 10^{-8}$ durch das auf einem hohen LD basierenden Auswahlverfahren für die Genotypisierung in der gesamten Studienpopulation herausgefiltert worden sein. Darüber hinaus ist es möglich, dass sich die kausale Variante außerhalb des gewählten 100 kb Bereiches befindet. Bereits beim *HbS*-Locus zeigte sich in einer GWA-Studie zu schwerer Malaria, dass Varianten, die ein starkes LD zur kausalen Variante aufweisen, auch außerhalb dieses Bereiches mit der Erkrankung assoziiert sein können (Jallow *et al.*, 2009). In der vorliegenden GWA-Studie zeigten zwei SNPs, die mehr als 400 kb voneinander entfernt liegen, eine Assoziation des *HBB*-Locus mit zerebraler Malaria.

4.5 *Long noncoding RNAs* im Assoziationsbereich?

Um Transkripte aus der umliegenden Region des SNPs rs12479448 mit der stärksten Assoziation mit zerebraler Malaria zu untersuchen, wurde ein 76,7 kb langer Assoziationsbereich ausgewählt, der durch SNPs mit einer Signifikanz von $p = 0,05$ und einem LD von r^2 von 0,05 zu rs12479448 begrenzt wurde. In den Online-Datenbanken des NCBI und UCSC (<http://www.ncbi.nlm.nih.gov>; <http://www.genome.ucsc.edu>) waren in diesem Bereich fünf Transkripte beschrieben. *In silico*-Analysen zeigten, dass die Transkripte keinen offenen Leserahmen (ORF) besitzen, der eine Länge von 64 Aminosäuren (AS) übersteigt. Humane, Protein-kodierende RNAs verfügen hingegen zu 95 % über einen ORF mit einer Länge von mehr als 100 AS (Frith *et al.*, 2006). Darüber hinaus weist keiner der ermittelten ORFs eine Sequenz auf, die alle beschriebenen Exons einschließt. Vergleiche der Sequenz von annotierten Proteinen zu den Peptidsequenzen der ORFs, zeigten lediglich zu einem hypothetischen Protein aus dem Genom von Hominiden substanzielle Ähnlichkeit. Aufgrund der *in silico* ermittelten Charakteristika lässt sich mit hoher Wahrscheinlichkeit eine Protein-kodierende Funktion der Transkripte ausschließen. Vermutlich handelt es sich bei den Transkripten um *noncoding RNAs*

(ncRNAs), die durch ihre Länge von über 200 nt der Klasse der *long* ncRNAs (lncRNAs) zuzuordnen wären.

4.6 Expression der Transkripte

Im Assoziationsbereich konnte die Expression von Segmenten der fünf in den Online-Datenbanken des NCBI und UCSC annotierten Transkripte über Polymerase Kettenreaktion (PCR)-basierte Methoden nachgewiesen werden. Dabei konnte für alle Transkripte eine gewebespezifische Expression wahrscheinlich gemacht werden. Eine gewebespezifische Expression wurde zuvor bereits für andere lncRNAs beschrieben (Ponjavic *et al.*, 2007) und auf eine regulierte Transkriptionsmaschinerie zurückgeführt (Gustincich *et al.*, 2006). Die Transkripte konnten vorrangig in zerebralem, vaskulärem und hepatischem Gewebe nachgewiesen werden. Im zerebralen und vaskulären Gewebe fanden sich die höchsten Expressionslevel der Transkripte. Möglicherweise deutet dies darauf hin, dass die Transkripte in der Pathogenese der zerebralen Malaria eine Rolle spielen könnten.

Am Beispiel des Transkripts DA und AI/AW zeigt sich, dass in *complementary* DNA (cDNA) eines gleichen Gewebes verschiedener Chargen voneinander abweichende Expressionsmuster auftreten können. Dies könnte auf eine RNA-Degradierung während der RNA-Isolierung zurückzuführen sein, so dass die Transkripte bei der cDNA-Synthese nicht als Matrize zur Verfügung standen. Ebenfalls ist denkbar, dass eine zu geringe Konzentration der Matrize im PCR-Ansatz die Amplifikation des Transkripts nicht erlaubte. Darüber hinaus könnte die RNA auch aus verschiedenen Zelltypen eines Gewebes isoliert worden sein, die jeweils über ein individuelles Expressionsprofil verfügen. Individuelle Expressionsmuster von Transkripten wurden selbst in homogenen Zellen beschrieben (Fiering *et al.*, 2000; McAdams & Arkin, 1999). Die unterschiedliche Expression von verschiedenen Exons eines Transkripts könnte zudem durch alternative Spleißprozesse bedingt sein (Gilbert, 1978; Modrek & Lee, 2002). In Untersuchungen zur vollen Transkriptlänge konnten alternative Spleißprozesse bereits für AI/AW/BX und DA bestätigt werden.

Die quantitativen Expressionsanalysen in verschiedenen Geweben zeigen zwischen der relativen Expression der Gene der Transkripte zum Referenzgen Beta Aktin (*ACTB*) und zum Referenzgen Protein Phosphatase 1 (*PPP1CA*) größtenteils eine Differenz von zwei Potenzen. Das in der Literatur beschriebene Auftreten von

starken Expressionsschwankungen von Referenzgenen in unterschiedlichen Geweben (Radonic *et al.*, 2004; Vandesompele *et al.*, 2002) ist daher für diesen Fall eher unwahrscheinlich. Analysen des verwendeten Referenzgens *ACTB*, das ein hohes Expressionsniveau besitzt, zeigten in humanen Endothelzellen etwa 1000 Transkriptkopien pro Zelle (Wada *et al.*, 2011). Aus den quantitativen Expressionsanalysen lässt sich daher schließen, dass die untersuchten Transkripte mit einer Kopienzahl von etwa $10^{-2} - 10^{-7}$ pro Zelle vertreten sind. Daraus ist abzuleiten, dass vermutlich nicht in allen Zellen der untersuchten Gewebe Kopien eines Transkripts vorhanden waren. Möglicherweise kommt es in den Zellen zu temporären Unterschieden in der Genexpression. Da sich ein Gewebe aus verschiedenen Zelltypen zusammensetzt, ist ebenfalls denkbar, dass nur in bestimmten Zelltypen die Expression eines entsprechenden Transkripts erfolgte. Lediglich in der Endothelzelllinie wurde für das Gen des Transkripts AA ein mit dem Referenzgen *PPP1CA* vergleichbares Expressionslevel nachgewiesen. Dies könnte darauf zurückzuführen sein, dass es sich dabei um einen homogenen Zelltyp handelt, der eine einheitliche Genexpression in den Zellen aufweist. Die Ergebnisse der Expressionsanalysen stimmen am ehesten mit Untersuchungsergebnissen überein, die zeigen, dass Loci mit nicht-kodierender Funktion in vielen Fällen niedrige Expressionslevel aufweisen (Ponting *et al.*, 2009). Es wird angenommen, dass ncRNAs hauptsächlich regulatorische Funktionen übernehmen (Mattick & Makunin, 2006).

4.7 Versuche zur Amplifikation vollständiger Transkripte

Mithilfe der *Rapid Amplification of cDNA-Ends* (RACE)-PCR konnten einige der 5'- und 3'-Enden der Transkripte aus dem Assoziationsbereich bestimmt werden. Die Methode benötigt in vielen Fällen eine umfangreiche Optimierung der PCR-Bedingungen, um das exakte Ende eines Transkripts zu amplifizieren (McGrath, 2011). Die mangelnde Spezifität wird durch die nur einseitig spezifische Amplifikation verursacht, die bei der RACE eines unbekanntes cDNA-Endes mit universellen Oligonukleotiden und genspezifischen Oligonukleotiden erfolgt. Bei der 3'-RACE bindet das universelle Oligonukleotid an den 200-250 nt langen Poly(A)-Schwanz am 3'-Ende von RNA Polymerase II synthetisierten Transkripten. Eine spezifische Amplifikation wird dabei nur über die genspezifischen Oligonukleotide im 5'-Bereich

erreicht. Erschwert wurde die spezifische Amplifikation der Transkripte, da ihre Gene in einer genomisch hoch-repetitiven Region lokalisiert sind (UCSC Genome Browser). Auch die Suche nach Oligonukleotiden, die eine ausschließlich genspezifische Bindung aufweisen, wurde dadurch erheblich erschwert. Zusätzlich zeigten sich in sequenzierten Amplifikaten der 3'-RACE jedes Transkripts eine oder mehrere aufeinander folgende Adenin-reiche Basenabfolgen. Diese ermöglichten eine Bindung des 5'-Oligo(dT)-Endes des universellen Oligonukleotids und begünstigten die Amplifikation von verkürzten 3'-Enden (Frohman *et al.*, 1988). Da über die Adenin-reichen Abschnitte in der Sequenz der Transkripte eine cDNA-Synthese mit Oligo(dT)-Nukleotiden erfolgen kann, könnten zusätzlich fehlerhafte Rückschlüsse auf eine Transkription mit RNA-Polymerase II geschlossen werden. Trotz der Beobachtung, dass RNA-Polymerase II eine Vielzahl an ncRNAs mit Poly(A)-Schwanz synthetisiert (Cai *et al.*, 2004; Lee *et al.*, 2004; Ponting *et al.*, 2009), wurden darüber hinaus RNAs beschrieben, die keinen Poly(A)-Schwanz besitzen (Wilusz & Spector, 2010). Eine 3'-RACE unter Verwendung eines Oligo(dT)-Nukleotids kann in diesem Fall nicht zu einer Amplifikation des (vollständigen) 3'-Endes des Transkripts führen. Außerdem können ncRNAs auch durch RNA Polymerase III synthetisiert werden (White, 2011), die ihrerseits keine Polyadenylierung durchführt. Es bietet sich daher an, zur Bestimmung der cDNA-Enden unterstützend weitere Methoden anzuwenden, die unabhängig von einer Polyadenylierung sind (White, 2011). Das Polyadenylierungssignal 5'-AAUAAA-3' der RNA-Polymerase II (Mandel *et al.*, 2006) ließ sich in der Sequenz von AA und DA in bestimmten Isoformen der Transkripte nachweisen. Ein Großteil aller Gene verfügt darüber hinaus über zwei oder mehr Polyadenylierungssignale (Yan & Marr, 2005), so dass durch 3'-RACE mit einem genspezifischen Oligonukleotid mehrere Amplifikationsprodukte entstehen können.

Die Amplifikation eines unbekanntes 5'-Endes anhand einer bekannten Teilsequenz erfolgte in dieser Studie mithilfe von zwei verschiedenen Strategien. Die *CapFinder*-Strategie ermöglicht hierbei durch die 5'-Cap-Struktur die Amplifikation von RNA Polymerase II synthetisierten Transkripten (Schramm *et al.*, 2000). Die vollständige Amplifikation des 5'-Endes wird gewährleistet, da die RNA durch die 5'-Cap-Struktur vor einer Degradation geschützt ist. Im Vergleich zum annotierten 5'-Ende von DA konnten zwei unterschiedliche Isoformen mit verkürzten 5'-Enden amplifiziert werden.

Da die Bindung des universellen Oligonukleotids ausschließlich an der 5'-Cap-Struktur erfolgt, ist davon auszugehen, dass es sich bei den Isoformen nicht um Artefakte handeln kann. Denkbar wäre, dass das Gen unterschiedliche Transkriptionsstartpunkte besitzt (Sandelin *et al.*, 2007). Eine weitere Strategie der 5'-RACE fügt an das 5'-Ende synthetisierter cDNA durch eine terminale Transferase einen Poly(A)-Schwanz an. Das Enzym ist in der Lage, auch möglicherweise degradierte 5'-Enden mit einem oligo(dT)-Schwanz zu versehen. Da mit der *CapFinder*-Strategie beim Transkript AA keine Ergebnisse erzielt werden konnten, wurde trotz des Nachteils diese Strategie zur Amplifikation des 5'-Endes gewählt. Dadurch konnten zwei 5'-Enden für das Transkript nachgewiesen werden. Ob das verkürzte 5'-Ende der Isoform 1 auf eine Degradation der RNA oder auf unterschiedliche Transkriptionsstartpunkte im Gen zurückzuführen ist, konnte nicht geklärt werden. Da für das Transkript AA über 3'-RACE ein Polyadenylierungssignal gefunden wurde, ist davon auszugehen, dass es durch RNA Polymerase II synthetisiert wird und dadurch cotranskriptionell eine 5'-Cap-Struktur erhält. Eine fehlende 5'-Cap-Struktur am Transkript AA scheidet daher als mögliche Ursache für das Scheitern der *CapFinder*-Methode aus. Vielmehr liegt nahe, dass durch die nur einseitig spezifische Amplifikation der RACE sowie eines im 5'-Bereich gelegenen *DNA repeat elements* (UCSC Genome Browser, *Repeatmasker*) die Amplifikation des 5'-Endes erschwert wurde und die optimalen Bedingungen der RACE-PCR bislang nicht erzielt werden konnten.

4.8 Mögliche Funktionen der RNAs

Von den etwa 90 % des humanen Genoms, die transkribiert werden, kodieren lediglich etwa 2 % für Protein-kodierende Gene (Birney *et al.*, 2007; The ENCODE Project Consortium, 2004). Durch die geringen Sequenzähnlichkeiten der verbleibenden Transkripte zu Protein-kodierenden Transkripten stellt sich die Frage, ob erstere durch ein so genanntes transkriptionelles „Hintergrundrauschen“ (*transcriptional noise*) erzeugt werden (Johnson *et al.*, 2005; Struhl, 2007), oder zur Gruppe von funktionellen ncRNAs zählen (Ponting *et al.*, 2009). *Transcriptional noise* beruht auf einer ungenauen Bindung der RNA-Polymerase II an schwache Promotorsequenzen im Genom, die zu einer permanenten Transkription entsprechender Bereiche führt (Struhl, 2007). Die hohe Konservierung der Regionen

im Assoziationsbereich (UCSC *Genome Browser, Build 36.3*), in denen die Transkripte kodiert sind, sowie die vorhandenen Spleißstellen von AI/AW/BX und DA könnten möglicherweise einen Hinweis auf eine positive Selektion geben (Mattick, 2009; Ponjavic *et al.*, 2007). Dies könnte auf eine eventuelle Funktion der beiden Transkripte verweisen und einer Synthese durch *transcriptional noise* entgegenstehen.

Mittlerweile konnten für einige der zahlreichen ncRNAs des menschlichen Genoms Funktionen beschrieben werden, die auch auf die untersuchten Transkripte zutreffen könnten. Diese Funktionen betreffen eine Reihe zellulärer Prozesse wie Transkription, Translation, Gen-Stillegung, Replikation, RNA-Prozessierung, -Modifikation und -Stabilität sowie Proteinstabilität und -translokation (Mercer *et al.*, 2009; Orom *et al.*, 2011; Storz, 2002). Zudem könnten die Transkripte als Vorläufer für kurze RNAs dienen (Jia *et al.*, 2010; Wilusz *et al.*, 2009).

Es wurde gezeigt, dass ncRNAs eine Rolle in komplexen Erkrankungen spielen können (Perez *et al.*, 2008; Taft *et al.*, 2010; Wapinski & Chang, 2011). Es wäre daher denkbar, dass genetische Varianten die Expression oder Funktion der untersuchten Transkripte beeinflussen und so eine mögliche Rolle in der Pathogenese der zerebralen Malaria spielen können. In einem *Plasmodium berghei* ANKA-Mausmodell für zerebrale Malaria konnte bereits eine veränderte Expression von drei ncRNAs in Gehirnzellen gezeigt werden (El-Assaad *et al.*, 2011). Es ist jedoch anzumerken, dass das Mausmodell einer zerebralen Malaria nicht auf die humane zerebrale Malaria übertragbar ist. Im murinen zerebralen Malaria-Modell binden Leukozyten und Thrombozyten an das Hirnendothel, mit *Plasmodium* infizierte Erythrozyten jedoch nicht (White *et al.*, 2010). Trotz dessen wäre denkbar, dass die Infektion mit *P. falciparum* einen Einfluss auf das Expressionsverhalten der Gene der Transkripte haben könnte. So löste zum Beispiel eine Infektion mit dem Epstein-Barr-Virus *in vitro* eine erhöhte Expression von lncRNAs in humanen B-Zellen aus (Mrazek *et al.*, 2007). Auch konnte im Mausmodell mit dem SARS-assoziierten Coronavirus ein verändertes Expressionsverhalten bei etwa 500 annotierten lncRNAs und etwa 1000 bisher nicht annotierten Loci in Lungenzellen beobachtet werden (Peng *et al.*, 2010).

Ebenso könnte eine mögliche Interaktion der Transkripte mit umliegenden Genen, Genprodukten oder regulatorischen Einheiten Auswirkungen auf die Manifestation

einer zerebralen Malaria haben. So wurde beispielsweise in GWA-Studien zu Typ-II-Diabetes, intrakraniellen Aneurysmen, verschiedener Krebsarten und zum Herzinfarkt ein Empfänglichkeitslocus gefunden, in dem die lncRNA ANRIL (*antisense lncRNA of the INK4 locus*) kodiert ist (Pasmant *et al.*, 2010). SNPs in und um diesen Genlocus besitzen einen direkten Einfluss auf das Expressionsverhalten und die Prozessierung von ANRIL (Burd *et al.*, 2010; Yap *et al.*, 2010). Darüber hinaus beeinflusst ANRIL in *cis* die Expression eines umliegenden Genlocus (Yap *et al.*, 2010). Sollte die kausale Variante des krankheitsassoziierten Locus tatsächlich in einem veränderten Expressionsverhalten oder einer alternativen Prozessierung in einem der Transkripte resultieren, könnte dies möglicherweise eine entscheidende Rolle in der Pathogenese der zerebralen Malaria spielen.

4.9 Ausblick

Um die kausale Variante zu finden, die eine Resistenz gegenüber zerebraler Malaria verleiht, sollten weitere Ansätze im *Fine Mapping* verfolgt werden. Da das LD zwischen der kausalen Variante und der die Assoziation mit der Erkrankung anzeigenden Variante in afrikanischen Populationen sehr schwach sein kann (Jallow *et al.*, 2009), sollten ebenfalls Varianten aus dem Assoziationsbereich mit einem LD von $r^2 < 0,6$ zu rs12479448 untersucht werden. Diesbezüglich könnte bei der Auswahl der genetischen Varianten ein Fokus auf jene Varianten gelegt werden, die in der genomischen DNA-Sequenz der Transkripte lokalisiert sind.

Zusätzlich zum *Fine Mapping* kann durch Imputation die Suche nach der kausalen Variante unterstützt werden (Browning & Browning, 2007). Imputation bezeichnet die statistische Methode, auf Basis der bekannten Haplotypen von SNPs eines Individuums fehlende Allele benachbarter SNPs bei einem zu untersuchenden Individuum abzuleiten. Für diese Analysen stehen durch das 1000-Genom-Projekt Datensätze mit Haplotypen von Individuen afrikanischen Ursprungs zur Verfügung (1000 Genomes Project Consortium, 2010). Mit den vorhergesagten Genotypen lassen sich *in silico* Assoziationsberechnungen durchführen, die in GWA-Studien die Teststärke erhöhen können (Spencer *et al.*, 2009).

Untersuchungen zur möglichen Funktion der Transkripte als *Enhancer* oder als Repressoren von Genen könnten über RNA-Interferenz erfolgen (Fire *et al.*, 1998). Dabei würden die Transkripte mithilfe von doppelsträngiger microRNA degradiert

werden. Expressionsanalysen der umliegenden Gene könnten durch *Real-Time* PCR durchgeführt werden. Genomweite oder regionale Expressionsanalysen könnten mithilfe von *Microarrays* oder cDNA-Sequenzierung erfolgen.

LITERATURVERZEICHNIS

- ADAMS, S., BROWN, H. & TURNER, G. (2002) Breaking down the blood-brain barrier: signaling a path to cerebral malaria? *Trends Parasitol*, 18, 360-6.
- AITMAN, T. J., COOPER, L. D., NORSWORTHY, P. J. *et al.*, (2000) Malaria susceptibility and CD36 mutation. *Nature*, 405, 1015-6.
- ALLEN, S. J., O'DONNELL, A., ALEXANDER, N. D. *et al.*, (1999) Prevention of cerebral malaria in children in Papua New Guinea by southeast Asian ovalocytosis band 3. *Am J Trop Med Hyg*, 60, 1056-60.
- ALTSHULER, D. M., GIBBS, R. A., PELTONEN, L. *et al.*, (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, 467, 52-8.
- BARRETT, J. C., FRY, B., MALLER, J. *et al.*, (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21, 263-5.
- BARUCH, D. I., GORMELY, J. A., MA, C. *et al.*, (1996) Plasmodium falciparum erythrocyte membrane protein 1 is a parasitized erythrocyte receptor for adherence to CD36, thrombospondin, and intercellular adhesion molecule 1. *Proc Natl Acad Sci U S A*, 93, 3497-502.
- BEARE, N. A., TAYLOR, T. E., HARDING, S. P. *et al.*, (2006) Malarial retinopathy: a newly established diagnostic sign in severe malaria. *Am J Trop Med Hyg*, 75, 790-7.
- BERKLEY, J. A., MWANGI, I., MELLINGTON, F. *et al.*, (1999) Cerebral malaria versus bacterial meningitis in children with impaired consciousness. *Qjm*, 92, 151-7.
- BERNARD, K., AUPHAN, N., GRANJEAUD, S. *et al.*, (1996) Multiplex messenger assay: simultaneous, quantitative measurement of expression of many genes in the context of T cell activation. *Nucleic Acids Res*, 24, 1435-42.
- BICKEMÖLLER, H. & FISCHER, C. (2007) Einführung in die genetische Epidemiologie.
- BINKA, F. N., MORRIS, S. S., ROSS, D. A. *et al.*, (1994) Patterns of malaria morbidity and mortality in children in northern Ghana. *Trans R Soc Trop Med Hyg*, 88, 381-5.
- BIRNEY, E., STAMATOYANNOPOULOS, J. A., DUTTA, A. *et al.*, (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447, 799-816.
- BREATHNACH, R., BENOIST, C., O'HARE, K. *et al.*, (1978) Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc Natl Acad Sci U S A*, 75, 4853-7.
- BREWSTER, D. R., KWIATKOWSKI, D. & WHITE, N. J. (1990) Neurological sequelae of cerebral malaria in children. *Lancet*, 336, 1039-43.
- BROWNING, S. R. & BROWNING, B. L. (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*, 81, 1084-97.
- BURD, C. E., JECK, W. R., LIU, Y. *et al.*, (2010) Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet*, 6, e1001233.
- BURGNER, D., USEN, S., ROCKETT, K. *et al.*, (2003) Nucleotide and haplotypic diversity of the NOS2A promoter region and its relationship to cerebral malaria. *Hum Genet*, 112, 379-86.
- BURGNER, D., XU, W., ROCKETT, K. *et al.*, (1998) Inducible nitric oxide synthase polymorphism and fatal cerebral malaria. *Lancet*, 352, 1193-4.
- CAI, X., HAGEDORN, C. H. & CULLEN, B. R. (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna*, 10, 1957-66.
- CAMPBELL, M. C. & TISHKOFF, S. A. (2008) African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet*, 9, 403-33.
- CARDON, L. R. & BELL, J. I. (2001) Association study designs for complex diseases. *Nat Rev Genet*, 2, 91-9.
- CLARK, A. G., HUBISZ, M. J., BUSTAMANTE, C. D. *et al.*, (2005) Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res*, 15, 1496-502.
- COLTEL, N., COMBES, V., HUNT, N. H. *et al.*, (2004) Cerebral malaria -- a neurovascular pathology with many riddles still to be solved. *Curr Neurovasc Res*, 1, 91-110.

- COMBES, V., TAYLOR, T. E., JUHAN-VAGUE, I. *et al.*, (2004) Circulating endothelial microparticles in malawian children with severe falciparum malaria complicated with coma. *Jama*, 291, 2542-4.
- CONRAD, D. F., JAKOBSSON, M., COOP, G. *et al.*, (2006) A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet*, 38, 1251-60.
- CONSORTIUM, T. E. P. (2004) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Science*, 306, 636-40.
- CONSORTIUM, T. G. P. (2010) A map of human genome variation from population-scale sequencing. *Nature*, 467, 1061-73.
- COOPER, G. M. & SHENDURE, J. (2011) Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet*, 12, 628-40.
- DAVIES, H. T., CROMBIE, I. K. & TAVAKOLI, M. (1998) When can odds ratios mislead? *Bmj*, 316, 989-91.
- DE BAKKER, P. I. & TELENTI, A. (2010) Infectious diseases not immune to genome-wide association. *Nat Genet*, 42, 731-2.
- DOLTER, K. E. & BRAMAN, J. C. (2001) Small-sample total RNA purification: laser capture microdissection and cultured cell applications. *Biotechniques*, 30, 1358-61.
- EL-ASSAAD, F., HEMPEL, C., COMBES, V. *et al.*, (2011) Differential microRNA expression in experimental cerebral and noncerebral malaria. *Infect Immun*, 79, 2379-84.
- FERNANDEZ-REYES, D., CRAIG, A. G., KYES, S. A. *et al.*, (1997) A high frequency African coding polymorphism in the N-terminal domain of ICAM-1 predisposing to cerebral malaria in Kenya. *Hum Mol Genet*, 6, 1357-60.
- FIERING, S., WHITELAW, E. & MARTIN, D. I. (2000) To be or not to be active: the stochastic nature of enhancer action. *Bioessays*, 22, 381-7.
- FIRE, A., XU, S., MONTGOMERY, M. K. *et al.*, (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391, 806-11.
- FISCHER, P. R. & BOONE, P. (1998) Short report: severe malaria associated with blood group. *Am J Trop Med Hyg*, 58, 122-3.
- FLINT, J., HARDING, R. M., BOYCE, A. J. *et al.*, (1998) The population genetics of the haemoglobinopathies. *Baillieres Clin Haematol*, 11, 1-51.
- FLINT, J., HILL, A. V., BOWDEN, D. K. *et al.*, (1986) High frequencies of alpha-thalassaemia are the result of natural selection by malaria. *Nature*, 321, 744-50.
- FRANCISCHETTI, I. M. (2008) Does activation of the blood coagulation cascade have a role in malaria pathogenesis? *Trends Parasitol*, 24, 258-63.
- FRAZER, K. A., BALLINGER, D. G., COX, D. R. *et al.*, (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449, 851-61.
- FRITH, M. C., FORREST, A. R., NOURBAKHS, E. *et al.*, (2006) The abundance of short proteins in the mammalian proteome. *PLoS Genet*, 2, e52.
- FROHMAN, M. A., DUSH, M. K. & MARTIN, G. R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A*, 85, 8998-9002.
- FRY, A. E., AUBURN, S., DIAKITE, M. *et al.*, (2008) Variation in the ICAM1 gene is not associated with severe malaria phenotypes. *Genes Immun*, 9, 462-9.
- GALIVETI, C. R., ROZHDESTVENSKY, T. S., BROSIUS, J. *et al.*, (2009) Application of housekeeping npcRNAs for quantitative expression analysis of human transcriptome by real-time PCR. *Rna*, 16, 450-61.
- GARCIA, A., COT, M., CHIPPAUX, J. P. *et al.*, (1998a) Genetic control of blood infection levels in human malaria: evidence for a complex genetic model. *Am J Trop Med Hyg*, 58, 480-8.
- GARCIA, A., MARQUET, S., BUCHETON, B. *et al.*, (1998b) Linkage analysis of blood *Plasmodium falciparum* levels: interest of the 5q31-q33 chromosome region. *Am J Trop Med Hyg*, 58, 705-9.
- GASTEIGER, E., GATTIKER, A., HOOGLAND, C. *et al.*, (2003) ExpASY: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, 31, 3784-8.
- GILBERT, W. (1978) Why genes in pieces? *Nature*, 271, 501.

- GRAU, G. E., MACKENZIE, C. D., CARR, R. A. *et al.*, (2003) Platelet accumulation in brain microvessels in fatal pediatric cerebral malaria. *J Infect Dis*, 187, 461-6.
- GUSTINCICH, S., SANDELIN, A., PLESSY, C. *et al.*, (2006) The complexity of the mammalian transcriptome. *J Physiol*, 575, 321-32.
- HALDANE, J. B. S. (1949) The rate of mutations of human genes. Proceedings of the VIII Congress of Genetics. *Hereditas*, 35, 267-273.
- HARDY, J. & SINGLETON, A. (2009) Genomewide association studies and human disease. *N Engl J Med*, 360, 1759-68.
- HINDORFF, L. A., SETHUPATHY, P., JUNKINS, H. A. *et al.*, (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*, 106, 9362-7.
- HOBBS, M. R., UDHAYAKUMAR, V., LEVESQUE, M. C. *et al.*, (2002) A new NOS2 promoter polymorphism associated with increased nitric oxide production and protection from severe malaria in Tanzanian and Kenyan children. *Lancet*, 360, 1468-75.
- HOWELL, W. M., JOBS, M., GYLLENSTEN, U. *et al.*, (1999) Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. *Nat Biotechnol*, 17, 87-8.
- HSIAO, L. L., DANGOND, F., YOSHIDA, T. *et al.*, (2001) A compendium of gene expression in normal human tissues. *Physiol Genomics*, 7, 97-104.
- IDRO, R., CARTER, J. A., FEGAN, G. *et al.*, (2006) Risk factors for persisting neurological and cognitive impairments following cerebral malaria. *Arch Dis Child*, 91, 142-8.
- IDRO, R., JENKINS, N. E. & NEWTON, C. R. (2005) Pathogenesis, clinical features, and neurological outcome of cerebral malaria. *Lancet Neurol*, 4, 827-40.
- IDRO, R., MARSH, K., JOHN, C. C. *et al.*, (2010) Cerebral malaria: mechanisms of brain injury and strategies for improved neurocognitive outcome. *Pediatr Res*, 68, 267-74.
- IDRO, R., NDIRITU, M., OGUTU, B. *et al.*, (2007) Burden, features, and outcome of neurological involvement in acute falciparum malaria in Kenyan children. *Jama*, 297, 2232-40.
- J. B. KRUSKAL, M. W. (1978) Multidimensional Scaling. 7.
- JALLOW, M., TEO, Y. Y., SMALL, K. S. *et al.*, (2009) Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat Genet*, 41, 657-65.
- JIA, H., OSAK, M., BOGU, G. K. *et al.*, (2010) Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *Rna*, 16, 1478-87.
- JOHN, C. C., BANGIRANA, P., BYARUGABA, J. *et al.*, (2008) Cerebral malaria in children is associated with long-term cognitive impairment. *Pediatrics*, 122, e92-9.
- JOHNSON, J. M., EDWARDS, S., SHOEMAKER, D. *et al.*, (2005) Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet*, 21, 93-102.
- KE, X., TAYLOR, M. S. & CARDON, L. R. (2008) Singleton SNPs in the human genome and implications for genome-wide association studies. *Eur J Hum Genet*, 16, 506-15.
- KNIGHT, J. C., UDALOVA, I., HILL, A. V. *et al.*, (1999) A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nat Genet*, 22, 145-50.
- KOBBE, R., ADJEI, S., KREUZBERG, C. *et al.*, (2007) Malaria incidence and efficacy of intermittent preventive treatment in infants (IPTi). *Malar J*, 6, 163.
- KOBBE, R., NEUHOFF, R., MARKS, F. *et al.*, (2006) Seasonal variation and high multiplicity of first Plasmodium falciparum infections in children from a holoendemic area in Ghana, West Africa. *Trop Med Int Health*, 11, 613-9.
- KOCH, O., AWOMOYI, A., USEN, S. *et al.*, (2002) IFNGR1 gene promoter polymorphisms and susceptibility to cerebral malaria. *J Infect Dis*, 185, 1684-7.
- KRUMP-KONVALINKOVA, V., BITTINGER, F., UNGER, R. E. *et al.*, (2001) Generation of human pulmonary microvascular endothelial cell lines. *Lab Invest*, 81, 1717-27.
- KUN, J. F., KLABUNDE, J., LELL, B. *et al.*, (1999) Association of the ICAM-1 Kilifi mutation with protection against severe malaria in Lambarene, Gabon. *Am J Trop Med Hyg*, 61, 776-9.
- KUN, J. F., MORDMULLER, B., LELL, B. *et al.*, (1998) Polymorphism in promoter region of inducible nitric oxide synthase gene and protection against malaria. *Lancet*, 351, 265-6.
- LANDER, E. S. & SCHORK, N. J. (1994) Genetic dissection of complex traits. *Science*, 265, 2037-48.

- LEE, Y., KIM, M., HAN, J. *et al.*, (2004) MicroRNA genes are transcribed by RNA polymerase II. *Embo J*, 23, 4051-60.
- LEVESQUE, M. C., HOBBS, M. R., ANSTEY, N. M. *et al.*, (1999) Nitric oxide synthase type 2 promoter polymorphisms, nitric oxide production, and disease severity in Tanzanian children with malaria. *J Infect Dis*, 180, 1994-2002.
- LI, J. Z., ABSHER, D. M., TANG, H. *et al.*, (2008a) Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, 319, 1100-4.
- LI, M., LI, C. & GUAN, W. (2008b) Evaluation of coverage variation of SNP chips for genome-wide association studies. *Eur J Hum Genet*, 16, 635-43.
- MACKINNON, M. J., MWANGI, T. W., SNOW, R. W. *et al.*, (2005) Heritability of malaria in Africa. *PLoS Med*, 2, e340.
- MANDEL, C. R., KANEKO, S., ZHANG, H. *et al.*, (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature*, 444, 953-6.
- MANGANO, V. D., LUONI, G., ROCKETT, K. A. *et al.*, (2008) Interferon regulatory factor-1 polymorphisms are associated with the control of Plasmodium falciparum infection. *Genes Immun*, 9, 122-9.
- MANOLIO, T. A., BROOKS, L. D. & COLLINS, F. S. (2008) A HapMap harvest of insights into the genetics of common disease. *J Clin Invest*, 118, 1590-605.
- MANOLIO, T. A., COLLINS, F. S., COX, N. J. *et al.*, (2009) Finding the missing heritability of complex diseases. *Nature*, 461, 747-53.
- MARQUET, S., DOUMBO, O., CABANTOUS, S. *et al.*, (2008) A functional promoter variant in IL12B predisposes to cerebral malaria. *Hum Mol Genet*, 17, 2190-5.
- MATTICK, J. S. (2009) The genetic signatures of noncoding RNAs. *PLoS Genet*, 5, e1000459.
- MATTICK, J. S. & MAKUNIN, I. V. (2006) Non-coding RNA. *Hum Mol Genet*, 15 Spec No 1, R17-29.
- MAY, J., EVANS, J. A., TIMMANN, C. *et al.*, (2007) Hemoglobin variants and disease manifestations in severe falciparum malaria. *Jama*, 297, 2220-6.
- MCADAMS, H. H. & ARKIN, A. (1999) It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet*, 15, 65-9.
- MCCARTHY, M. I., ABECASIS, G. R., CARDON, L. R. *et al.*, (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*, 9, 356-69.
- MCGRATH, P. T. (2011) Characterizing cDNA Ends by Circular RACE. *Methods Mol Biol*, 772, 257-65.
- MCGUIRE, W., HILL, A. V., ALLSOPP, C. E. *et al.*, (1994) Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. *Nature*, 371, 508-10.
- MEDANA, I. M., IDRO, R. & NEWTON, C. R. (2007) Axonal and astrocyte injury markers in the cerebrospinal fluid of Kenyan children with severe malaria. *J Neurol Sci*, 258, 93-8.
- MEDANA, I. M. & TURNER, G. D. (2006) Human cerebral malaria and the blood-brain barrier. *Int J Parasitol*, 36, 555-68.
- MERCER, T. R., DINGER, M. E. & MATTICK, J. S. (2009) Long non-coding RNAs: insights into functions. *Nat Rev Genet*, 10, 155-9.
- MEYER, C. G., CALIXTO FERNANDES, M. H., INTEMANN, C. D. *et al.*, (2011) IL3 variant on chromosomal region 5q31-33 and protection from recurrent malaria attacks. *Hum Mol Genet*, 20, 1173-81.
- MILET, J., NUEL, G., WATIER, L. *et al.*, (2010) Genome wide linkage study, using a 250K SNP map, of Plasmodium falciparum infection and mild malaria attack in a Senegalese population. *PLoS One*, 5, e11616.
- MILLER, L. H. (1996) Malaria. Protective selective pressure. *Nature*, 383, 480-1.
- MILLER, L. H., BARUCH, D. I., MARSH, K. *et al.*, (2002) The pathogenic basis of malaria. *Nature*, 415, 673-9.
- MODREK, B. & LEE, C. (2002) A genomic view of alternative splicing. *Nat Genet*, 30, 13-9.
- MOLYNEUX, M. E., TAYLOR, T. E., WIRIMA, J. J. *et al.*, (1989) Clinical features and prognostic indicators in paediatric cerebral malaria: a study of 131 comatose Malawian children. *Q J Med*, 71, 441-59.

- MORAHAN, G., BOUTLIS, C. S., HUANG, D. *et al.*, (2002) A promoter polymorphism in the gene encoding interleukin-12 p40 (IL12B) is associated with mortality from cerebral malaria and with reduced nitric oxide production. *Genes Immun*, 3, 414-8.
- MOUNT, S. M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res*, 10, 459-72.
- MOXON, C. A., GRAU, G. E. & CRAIG, A. G. (2011) Malaria: modification of the red blood cell and consequences in the human host. *Br J Haematol*.
- MRAZEK, J., KREUTMAYER, S. B., GRASSER, F. A. *et al.*, (2007) Subtractive hybridization identifies novel differentially expressed ncRNA species in EBV-infected human B cells. *Nucleic Acids Res*, 35, e73.
- MULLER, P. Y., JANOVJAK, H., MISEREZ, A. R. *et al.*, (2002) Processing of gene expression data generated by quantitative real-time RT-PCR. *Biotechniques*, 32, 1372-4, 1376, 1378-9.
- MURRAY, C. J., ROSENFELD, L. C., LIM, S. S. *et al.*, (2012) Global malaria mortality between 1980 and 2010: a systematic analysis. *Lancet*, 379, 413-31.
- NAKA, I., PATARAPOTIKUL, J., TOKUNAGA, K. *et al.*, (2009) A replication study of the association between the IL12B promoter allele CTCTAA and susceptibility to cerebral malaria in Thai population. *Malar J*, 8, 290.
- NEWTON, C. R., TAYLOR, T. E. & WHITTEN, R. O. (1998) Pathophysiology of fatal falciparum malaria in African children. *Am J Trop Med Hyg*, 58, 673-83.
- O'CONNELL, D. (2007) Malaria: Fatal attraction. *Nature Reviews Microbiology*, 5.
- OCHOLA, L. B., SIDDONDO, B. R., OCHOLLA, H. *et al.*, (2011) Specific receptor usage in Plasmodium falciparum cytoadherence is associated with disease outcome. *PLoS One*, 6, e14741.
- OMI, K., OHASHI, J., PATARAPOTIKUL, J. *et al.*, (2003) CD36 polymorphism is associated with protection from cerebral malaria. *Am J Hum Genet*, 72, 364-74.
- OROM, U. A., DERRIEN, T., GUIGO, R. *et al.*, (2011) Long noncoding RNAs as enhancers of gene expression. *Cold Spring Harb Symp Quant Biol*, 75, 325-31.
- PASMANT, E., SABBAGH, A., VIDAUD, M. *et al.*, (2010) ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *Faseb J*, 25, 444-8.
- PATEL, S. S., MEHLOTRA, R. K., KASTENS, W. *et al.*, (2001) The association of the glycophorin C exon 3 deletion with ovalocytosis and malaria susceptibility in the Wosera, Papua New Guinea. *Blood*, 98, 3489-91.
- PE'ER, I., YELENSKY, R., ALTSHULER, D. *et al.*, (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*, 32, 381-5.
- PEARSON, T. A. & MANOLIO, T. A. (2008) How to interpret a genome-wide association study. *Jama*, 299, 1335-44.
- PENG, X., GRALINSKI, L., ARMOUR, C. D. *et al.*, (2010) Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. *MBio*, 1.
- PEREZ, D. S., HOAGE, T. R., PRITCHETT, J. R. *et al.*, (2008) Long, abundantly expressed non-coding transcripts are altered in cancer. *Hum Mol Genet*, 17, 642-55.
- PFAFFL (2004) Real-Time RT-PCR: Neue Ansätze zur exakten mRNA Quantifizierung. *BIOspektrum*, 4.
- PONGPONRATN, E., RIGANTI, M., PUNPOOWONG, B. *et al.*, (1991) Microvascular sequestration of parasitized erythrocytes in human falciparum malaria: a pathological study. *Am J Trop Med Hyg*, 44, 168-75.
- PONGPONRATN, E., TURNER, G. D., DAY, N. P. *et al.*, (2003) An ultrastructural study of the brain in fatal Plasmodium falciparum malaria. *Am J Trop Med Hyg*, 69, 345-59.
- PONJAVIC, J., PONTING, C. P. & LUNTER, G. (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res*, 17, 556-65.
- PONSFORD, M. J., MEDANA, I. M., PRAPANSILP, P. *et al.*, (2011) Sequestration and Microvascular Congestion Are Associated With Coma in Human Cerebral Malaria. *J Infect Dis*.
- PONTING, C. P., OLIVER, P. L. & REIK, W. (2009) Evolution and functions of long noncoding RNAs. *Cell*, 136, 629-41.
- PURCELL, S., NEALE, B., TODD-BROWN, K. *et al.*, (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81, 559-75.

- RADONIC, A., THULKE, S., MACKAY, I. M. *et al.*, (2004) Guideline to reference gene selection for quantitative real-time PCR. *Biochem Biophys Res Commun*, 313, 856-62.
- RIHET, P., ABEL, L., TRAORE, Y. *et al.*, (1998) Human malaria: segregation analysis of blood infection levels in a suburban area and a rural area in Burkina Faso. *Genet Epidemiol*, 15, 435-50.
- RISCH, N. J. (2000) Searching for genetic determinants in the new millennium. *Nature*, 405, 847-56.
- ROCA-FELTRER, A., CARNEIRO, I. & ARMSTRONG SCHELLENBERG, J. R. (2008) Estimates of the burden of malaria morbidity in Africa in children under the age of 5 years. *Trop Med Int Health*, 13, 771-83.
- ROWE, J. A., HANDEL, I. G., THERA, M. A. *et al.*, (2007) Blood group O protects against severe *Plasmodium falciparum* malaria through the mechanism of reduced rosetting. *Proc Natl Acad Sci U S A*, 104, 17471-6.
- ROZEN, S. & SKALETSKY, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 132, 365-86.
- RUWENDE, C., KHOO, S. C., SNOW, R. W. *et al.*, (1995) Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. *Nature*, 376, 246-9.
- SACHIDANANDAM, R., WEISSMAN, D., SCHMIDT, S. C. *et al.*, (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409, 928-33.
- SACHS, J. & MALANEY, P. (2002) The economic and social burden of malaria. *Nature*, 415, 680-5.
- SAKUNTABHAI, A., NDIAYE, R., CASADEMONT, I. *et al.*, (2008) Genetic determination and linkage mapping of *Plasmodium falciparum* malaria related traits in Senegal. *PLoS One*, 3, e2000.
- SANDELIN, A., CARNINCI, P., LENHARD, B. *et al.*, (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet*, 8, 424-36.
- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 74, 5463-7.
- SAUL, A. (1999) The role of variant surface antigens on malaria-infected red blood cells. *Parasitol Today*, 15, 455-7.
- SCHOFIELD, L. & GRAU, G. E. (2005) Immunological processes in malaria pathogenesis. *Nat Rev Immunol*, 5, 722-35.
- SCHRAMM, G., BRUCHHAUS, I. & ROEDER, T. (2000) A simple and reliable 5'-RACE approach. *Nucleic Acids Res*, 28, E96.
- SCHUBART, C. D., MTURI, N., BELD, M. G. *et al.*, (2006) Role of viruses in Kenyan children presenting with acute encephalopathy in a malaria-endemic area. *Am J Trop Med Hyg*, 75, 1148-50.
- SIKORA, M., LAAYOUNI, H., CALAFELL, F. *et al.*, (2011) A genomic analysis identifies a novel component in the genetic structure of sub-Saharan African populations. *Eur J Hum Genet*, 19, 84-8.
- SKOL, A. D., SCOTT, L. J., ABECASIS, G. R. *et al.*, (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*, 38, 209-13.
- SLAVIN, T. P., FENG, T., SCHNELL, A. *et al.*, (2011) Two-marker association tests yield new disease associations for coronary artery disease and hypertension. *Hum Genet*.
- SMITH, J. D., CHITNIS, C. E., CRAIG, A. G. *et al.*, (1995) Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell*, 82, 101-10.
- SMITH, J. G. & NEWTON-CHEH, C. (2009) Genome-wide association study in humans. *Methods Mol Biol*, 573, 231-58.
- SNOW, R. W., OMUMBO, J. A., LOWE, B. *et al.*, (1997) Relation between severe malaria morbidity in children and level of *Plasmodium falciparum* transmission in Africa. *Lancet*, 349, 1650-4.
- SPENCER, C. C., SU, Z., DONNELLY, P. *et al.*, (2009) Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet*, 5, e1000477.

- SPIELMAN, R. S., MCGINNIS, R. E. & EWENS, W. J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet*, 52, 506-16.
- STORZ, G. (2002) An expanding universe of noncoding RNAs. *Science*, 296, 1260-3.
- STRUHL, K. (2007) Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol*, 14, 103-5.
- SUN, F., FLANDERS, W. D., YANG, Q. *et al.*, (1999) Transmission disequilibrium test (TDT) when only one parent is available: the 1-TDT. *Am J Epidemiol*, 150, 97-104.
- TAFT, R. J., PANG, K. C., MERCER, T. R. *et al.*, (2010) Non-coding RNAs: regulators of disease. *J Pathol*, 220, 126-39.
- TAYLOR, T. E., FU, W. J., CARR, R. A. *et al.*, (2004) Differentiating the pathologies of cerebral malaria by postmortem parasite counts. *Nat Med*, 10, 143-5.
- TEERANAIPONG, P., OHASHI, J., PATARAPOTIKUL, J. *et al.*, (2008) A functional single-nucleotide polymorphism in the CR1 promoter region contributes to protection against cerebral malaria. *J Infect Dis*, 198, 1880-91.
- TEO, Y. Y., SMALL, K. S. & KWIATKOWSKI, D. P. (2010) Methodological challenges of genome-wide association analysis in Africa. *Nat Rev Genet*, 11, 149-60.
- THOMAS, D. C., CASEY, G., CONTI, D. V. *et al.*, (2009) Methodological Issues in Multistage Genome-wide Association Studies. *Stat Sci*, 24, 414-429.
- THYE, T., VANNBERG, F. O., WONG, S. H. *et al.*, (2010) Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat Genet*, 42, 739-41.
- TIMMANN, C., EVANS, J. A., KONIG, I. R. *et al.*, (2007) Genome-wide linkage analysis of malaria infection intensity and mild disease. *PLoS Genet*, 3, e48.
- TOBLER, A. R., SHORT, S., ANDERSEN, M. R. *et al.*, (2005) The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. *J Biomol Tech*, 16, 398-406.
- TOWNSEND, C. F., NEWMAN, C. M. & FRANCIS, S. E. (1999) Efficient extraction of RNA from vascular tissue. *Methods Mol Med*, 30, 39-46.
- TURNER, G. D., MORRISON, H., JONES, M. *et al.*, (1994) An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. *Am J Pathol*, 145, 1057-69.
- VAN DER HEYDE, H. C., NOLAN, J., COMBES, V. *et al.*, (2006) A unified hypothesis for the genesis of cerebral malaria: sequestration, inflammation and hemostasis leading to microcirculatory dysfunction. *Trends Parasitol*, 22, 503-8.
- VAN HENSBROEK, M. B., PALMER, A., JAFFAR, S. *et al.*, (1997) Residual neurologic sequelae after childhood cerebral malaria. *J Pediatr*, 131, 125-9.
- VANDESOMPELE, J., DE PRETER, K., PATTYN, F. *et al.*, (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*, 3, RESEARCH0034.
- WADA, Y., LI, D., MERLEY, A. *et al.*, (2011) A multi-gene transcriptional profiling approach to the discovery of cell signature markers. *Cytotechnology*, 63, 25-33.
- WALLEY, A. J., AUCAN, C., KWIATKOWSKI, D. *et al.*, (2004) Interleukin-1 gene cluster polymorphisms and susceptibility to clinical malaria in a Gambian case-control study. *Eur J Hum Genet*, 12, 132-8.
- WAPINSKI, O. & CHANG, H. Y. (2011) Long noncoding RNAs and human disease. *Trends Cell Biol*, 21, 354-61.
- WHITE, N. J., TURNER, G. D., MEDANA, I. M. *et al.*, (2010) The murine cerebral malaria phenomenon. *Trends Parasitol*, 26, 11-5.
- WHITE, R. J. (2011) Transcription by RNA polymerase III: more complex than we thought. *Nat Rev Genet*, 12, 459-63.
- WHO (2000) Severe falciparum malaria. World Health Organization, Communicable Diseases Cluster. *Trans R Soc Trop Med Hyg*, 94 Suppl 1, S1-90.
- WHO (2011) World Malaria Report 2011.

-
- WILLCOX, M., BJORKMAN, A. & BROHULT, J. (1983) Falciparum malaria and beta-thalassaemia trait in northern Liberia. *Ann Trop Med Parasitol*, 77, 335-47.
- WILLIAMS, T. N. (2006) Red blood cell defects and malaria. *Mol Biochem Parasitol*, 149, 121-7.
- WILSON, J. N., ROCKETT, K., JALLOW, M. *et al.*, (2005) Analysis of IL10 haplotypic associations with severe malaria. *Genes Immun*, 6, 462-6.
- WILUSZ, J. E. & SPECTOR, D. L. (2010) An unexpected ending: noncanonical 3' end processing mechanisms. *Rna*, 16, 259-66.
- WILUSZ, J. E., SUNWOO, H. & SPECTOR, D. L. (2009) Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev*, 23, 1494-504.
- YAN, J. & MARR, T. G. (2005) Computational analysis of 3'-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. *Genome Res*, 15, 369-75.
- YAP, K. L., LI, S., MUNOZ-CABELLO, A. M. *et al.*, (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell*, 38, 662-74.

ANHANG

Oligonukleotide für die Detektion der Transkripte, RACE-PCR, qRT-PCR und DASH

Oligonukleotid	Oligonukleotid-Sequenz	Oligonukleotid	Oligonukleotid-Sequenz
AI218168.1I_for	tgattacatgggtttcttctatgctctca	AI2181.2_E1_for	gttaacttggcaggcact
OdT-T7 II	gagagaggatccaagtactaatacagctcactatagg	AI2181.2_E1o_rev	agcatcacaatgaaggaaaaca
OdT-T7 I	gagagaggatccaagtactaatacagctcactatagg gagatTTTTTTTTTTTTTTTT(agc)	DA227815E6_for	gatatgcaaggaagctcagtctc
cDNA_AI_neu_for	ttcagtatggattgaacagatagtg	DA227815E6_rev	taccctgacacagtgatg
AI218168-Cap.1F	gtgaaaagttggcttacaggga	PPP1CA.for	ctggtgacacttttctcagc
AI218168.2I_for	cctctcaaaaggaaatcagcctcctgt	PPP1CA.rev	cttcccctgttctgtcgg
AI21E3n.3'testR	tcagtctctttattatgtctgtca	qRT-ACTB-for	tcaagatcattgctcctctgag
Hs.542102.PW1F	atcatgggaagattgggtcc	qRT-ACTB-rev	acatctgctggaagtgagaca
Hs.542102.PW3F	tctcaagcatgggtgatg	CFB_for	gagagaacgctgacgagagactga cag
Hs.542102_for-2	ttatgaaacccttttaagtatcctgt	CapFinderB	gagagaacgctgacgagagactga ca(rgrgrg)
Hs.542102_for	tcacaatgccgtaagaatgc	BX091678.2_E1_for	tgttctgacagaggaaaaaac
Hs.542102_rev	cactgaaatcttctggaacataa	BX091678.2_E1_rev	tttctctcagcagggtgtg
Hs.542102.PW5F	gctgctaagaaataggtgcaaa	fydaw_E2+3_for	tctctgcctcaaacgttctt
DA227815E2_RNAr	cctttcagtgccctttcag	fydaw_E2+3_rev	tctctcgttctcactgttg
DA227815E3f	tgaccacgggtttctggagt	AI-BX_for	ttgtggcaggcactgtactt
DA227815E2-3.brF	cgaagctaaaaccagatactgctttc	oligo(dT)12-18	TTTTTTTTTTTT
Da227815-Ex5.1R	tacatccatcaaaatctgcaaaaaata	AI_4_for	ttcatgattacatgggttcttct
pT7T3D_for	ggaaatcagcctcctgttg	AI_5_for	gctctcaattctcagggtttca
pT7T3D_rev	atgtgccatccagttcaacc	Oligo(dT)-2_short	gagagaggatccaagtactaatacga ctc
DA227815E1_for	cctcctgaaataccacctgt	DA227815-Ex1.1F	tcctgaaataccacctgtaaataga
M13_F	gta aaa cga cgg cca g	DA227815-Ex3.1R	actccatacgaaaaaactccaga
M13_R	cag gaa aca gct atg ac	DA227815-Ex6.1R	cagttatgcttaccgtgc
Hs.542102.2I_rev	gcaatgaccacagtagaacgcatgg	DA227815-Ex2.1F	gctgctatggggggatg
Hs.542102.1I_rev	tgatggacagcaactttggaactgtg	DA227815-Ex5.1F	ttggcagattttgatgatgtac
AI218168_in2.2.rev	gagggtgaaagttcgtatctg	CapFc	acgcgtgacgagagactgacag
AI218168_in2.1_rev	tctgctagggatgtctgct	DA227815E2-3.brF	cgaagctaaaaccagatactgctttc
DA227815E6_seqF	ctatTTTggccaatagtagccc	AI218168-Cap.1F	gtgaaaagttggttcttacaggga
DA227815artcDNAf	gtgtcatgcagctgctatgg	AI218168-Cap.2F	tggagatgagagagaaagctgtga
DA227815artcDNAr	ctcccaataatacaatattct(ac)	DA227815brCap1	gaattcaagactgccttctacc
rs12476596F	agttaatagattactaattactctcagaag	DA227815brCap2	gcctcttctcagtagacgaagct
rs12476596R	agtaattatgggcaaaacatatctagtg	DA227815br3polA	ggtttccagataggcaacaag
rs12476596S	tatcattactacatatagaaaacaagtgaaat- Fluorescein	DA227815artpolA2	aagacagagtctcgtctgtcacc
rs12476596A	LCRED-aatcatctgtattgaaactatctgtgctttatg- Phosphate	Hs.542102.3-end.rev	gaactactgccattgatattattc
rs34996117F	agacaattcatctcaataagattttacct	AI218168E4.2r	ccccacttacttccatacag
rs34996117R	gggaaatgctcattaaaagcctca	AI218168E4.2f	gtatggaagtaaagtgggggttc
rs34996117A	LCRed-agatggcatgacattctccaaaagagattttaa- Phosphate	AI218168E4.2.1f	ttcattgaactctgtagtctc
rs34996117S	atacttattgctttgtcattaaagtaagttaa- Fluorescein	AI218168E4.3f	gtaaagtggggttctgccc
rs6543892F	taggtgtctgtttgtgatttagttttc	Hs.542102_CapFf.1	gaccactaacgggttctga
rs6543892R	tcctatctccattgtatcagaaaataact	Hs.542102_CapFf.2	cataccatccacaaccaga

rs6543892S	ccattgtctcttttgtttgttcattgctttgagatct- Fluorescein	Hs.542102.CapF3r	tcagaaccogttagtggtgc
rs6543892A	LCRed-atcataaattctttgtctaggtccaggctctaaagag- Phosphate	AI218168E4.3r	cactttactgtctgattggttc
35111669F	cagagaatgtaccactgtaaagtca	AI218168E4.3f	actgaggaaccaatcagcaagt
35111669R	tccaggcattcggacactg	AI218168E4.4f	aggacactctgtaatggaggaca
35111669A	LCRedtccctccctctcctcctcctccctcttct- Phosphate	AI218168E4.4r	ttgtaactttaaaaatattgctgta
35111669S	gctcgccctccctctcc-Fluorescein	AI218168E4.CapFf	ctggagtacagtggtccgatta
AI4E4_CapTestF	ttgagacaaggcttactctgttg	Hs.542102.Cap2outF	tgtgaaaattagacaattcatctcaaa
Hs.542102+3prR	ttgccattgatatttaccaca	Hs.542102Cap2F	catattaatgagttaactggtgcat
AI4E4_CapTestF	ttgagacaaggcttactctgttg	AI218168.5'2R	tttagaacctgccagacattca
Hs.542102+3prR	ttgccattgatatttaccaca	AI218168.5'1R	gggtttctccatgttggaatc
DA227815E6_UTRr1	caatggaagcaaaaactccaa	AI218168E4.5f	tccacctcgaaggctcagt
DA227815E6_UTRr2	aagcagaaggcagcttcaat	AI218168E4.5R	taatcggaccactgtactccag
AI218168E4f	gagcctgggcaagaatacat	AI218168E3nF	tgaatgtctggcaggttctaaa
AI218168E4r	ggcaagaacccccactttac	AI218168E3nR	gagacggagtttcgctctg

Oligonukleotide der Sequenzierung des *Fine Mappings*

Oligonukleotid	Oligonukletid-Sequenz	Oligonukleotid	Oligonukletid-Sequenz
gap1_f1	gattgtaaaggaataagtaagccta	seq_f87	ctgcataaaattatagcctaatttattc
seq_f1	cctatctatgctctggccatctct	pr16_f	tgcagcagtagatgggaaa
seq_f2	aggctggctctgaactcctga	seq_f88	gatattgcatctctcattacactctgg
seq_f3	ctcataatggatcatgaatctaaatgaagagct	seq_f89	ctgatcccaaacatgaacatctgc
seq_f4	gatggcatagtcatctttagacatcttgac	seq_f90	caaaaactggcctttctgtaatgtg
seq_f5	ctttgaattgaacatacgcaactttgtgg	seq_f91	ggtggtctctcatgtgtaattggtgc
pr1_fvs2	ggaggaggagggaagcagtttctag	seq_f92	catttttaattatgctggctttccg
seq_f6	gctatactgcaagagtcagcaaatatg	gap7_f1	tttcagtaggtgactgtagcctgta
seq_f7	ccagcctgagtaatatagcaagacctg	seq_f93	aatggcttagcaaatggcacacc
seq_f8	agcaagccgtgaggtatctgc	seq_f94	tggacctccagcaaacccaacagac
seq_f9	cccttcaaatagttaacatagctcg	seq_f95	gggactatgtgaaaagcaaatctg
pr2_f	tcaacttgaaagctcagttca	gap7_f2	ggaaaggacaaccagtagcagc
seq_f10	aacacgaatgaatccccagtg	seq_f96	ttaacacccccactgtcaacattagac
seq_f11	ttgtctgtattgaaacaaggagc	seq_f97	actaaaatcagagcagaacggaag
seq_f12	aaaaactatagctgcatcttctagctc	seq_f98	agaccaatatccctgatgaactg
seq_f13	gcaagaagaacaaggaaatgatagtgct	seq_f99	gataagcaactcagcaagctctcag
gap2_f1vs3	tcatgacagccttagtgactca	gap7_f3	gaaagaatcagtagcaaggtggc
seq_f14	tacacattattgaatggtggattagc	seq_f100	gtctaaaacacaaaagcaatggc
seq_f15	atgatggtgcggcctgtagtc	seq_f101	cggcactattcacaatagcaaacagctg
seq_f16	gattcaaagtattctgcactttgacag	pr18_f	aactgatgaatggcaagcct
seq_f17	ctggtggagggtgattgtagcatag	seq_f102	cacaagattagggaagcaacatg
seq_f18	agtcttacagtaaattgatgaaatgtag	seq_f103	ttatctattccaccagttgatgggtg
seq_f19	ccaggatcaatgaaacactgaagaag	seq_f104	caaaaagataagttcagccagatgg
pr3_f	ggcacaattgccaatgaac	gap8_f1	attgtggctacagtagtggtt
seq_f20	gcttagaaaactctataaaagcataaaaggtc	seq_f105	ggctacatagattccattgtgatg
seq_f21	ctcagtagcttttaaaaacaatgttcagtagac	seq_f106	gagttcattgtagattctgtagattagcc
seq_f22	gaaccaaaaggcccaagaaagc	gap8_f2	catccctagtaacagagcgtaaaac
gap3_f1	agtcacgttttagaggttaaattac	seq_f107	gagagtacaactattaaattctggtcaattg
seq_f23	tccagctctcctatttcaaaagttc	pr19_f	gaattgctacagctccaagcc
seq_f24	accctttcccaggagaagcctac	seq_f108	cagttttccaactggtcatgctcttc
seq_f25	gttaggagaatggaagcaaatcatagactg	seq_f109	gctatcaatattaaagctgatttgaaggtg

seq_f26	cccaaattccaccaacaggaagtg	gap9_f1	aggtaaaattcaaggcattacatgtct
gap3_f2	attgtgatggggaagctgga	seq_f110	cctgaaaatacaaaaactgcacccaagac
seq_f27	cctcctccatggattatgagtggtg	seq_f111	cctgtattctgattgacctagaagg
seq_f28	cttagaaggcatcttccaacag	seq_f112	tactgggcttaggatgggagacttag
seq_f29	ttcacttgaatccttagaggcaggc	gap9_f2	ggagataatatgttatagacttgga
seq_f30	aagaagacaggggcaataggc	seq_f113	gatgagattttcagattggcccacttag
pr4_f	actgttgacctctgaggac	seq_f114	gcatacactagatgttgaaggcattgg
seq_f31	agactgtgcaacagttttgttg	seq_f115	gtttgttagcttctgattggatc
pr5_f	ttcctaagcaggaaacat	gap9_f3	gctatttgaccacgattaaaatgccag
seq_f32	tgccacaacctggaagcaaaatgag	seq_f116	ctttgatttaacctgagggtgactc
seq_f33	tgggcagcagaacgagacactg	seq_f117	cagaaaatttggtacacagaattgaaacc
seq_f34	gtgacttagctatcaagacaataaaggc	seq_f118	cactgtagcaccattaaattgattatctg
seq_f35	gcatacctactcattgtctgcag	seq_f119	cactgcagtgtaataataaagaaggac
pr6_f	acaaaggctctcacagtattagaaa	seq_f120	gtggacactgaaatcacagatagactg
seq_f36	tgccccataaatgtacaattagtatg	seq_f121	cagaaagctgtatagttccatgagctg
seq_f37	cttaagtgcagtcaaacctggctac	seq_f122	tgggcttcatcccattggatgggtg
seq_f38	aggcaggagaaatggcgtgaac	gap9_f5	tcatgcagtaatgtcacagaatcct
seq_f39	ttaaaactgcagacattgcgagaagc	seq_f123	gtcccatgttctgtccacatccag
gap4_f1	ctagtaggctgtcgagaaaacg	seq_f124	agctcatggtgccaagaatttcag
seq_f40	atgaaaagaagtcgagaagttgag	seq_f125	ggtttaattgactaacagttaccatg
seq_f41	gaatgaatatgtataaacatcaagtgc	gap9_f6	acctaataatgactcactcgtga
seq_f42	caagttcaatactcataccatacactagg	seq_f126	ctccaatagtagcagaatctgcaagtc
gap4_f2	gtgcatgtctgaggctcagg	seq_f127	gaaaactgttcatcatgctactgtctacc
seq_f43	tcgaactcaataaaatccatccagaag	seq_f128	ccagggtgattttcatattgactg
seq_f44	tcaataaacacggtgcacatgcag	gap9_f7	gtgggtcaactgtagctcac
seq_f45	ggagagggctgggcttttaagt	seq_f129	ccttcatgcagattatcactgtcag
pr7_f	gggacaatacattctgggtaaga	seq_f130	ctgatgtaaatagatggaaccactcag
seq_f46	caatggggccacagaaaatggg	seq_f131	ctcaggattatctttatagcagtggtg
gap5_f1	aagtgtggaagagagaacagattcc	gap9_f8	tcaagctggtgcataaatctg
seq_f47	tcatccacgtttctcaagtgac	seq_f132	ctcaacgccagccatgaaagcag
seq_f48	tgagtttgactcctggtgattctg	seq_f133	gctctctctcctgcatgtaagtcatg
seq_f49	aagtaaggtaatgtgctgcctctag	seq_f134	atggtagtctaagttcccaagg
gap5_f2vs2	tcctaggtataagatcgttatcagcagc	gap9_f9	ttaccaatgcagcctacatattctg
seq_f50	tttcaaatccactgctgttattag	seq_f135	ccctaggccatagatgtttgtaagtg
seq_f51	tgggcaggaaccactatctcact	seq_f136	cagtagaactccatgcctcgttg
seq_f52	aggaaaactagtgctgatgtctc	seq_f137	ttgagttctacctaccaccaacggtg
pr8_f	tgtgggagagctgtgtgtt	seq_f138	actcatttagtaattaactctagaatcagc
seq_f53	ttgtcccactggaaggtctttagg	gap9_f10	ctaattatggatgcttaggtgtttg
seq_f54	cacagagaacatcacagactttattgg	seq_f139	cagactcagaattaagtaactctctg
seq_f55	ctaaaaatgctcaggggcagaaggttac	seq_f140	ctaattcttaagtgctgctgagatcc
seq_f56	actagctcagggctcaaaaggc	seq_f141	ccaacaactcctattttcaaatcatcattg
pr9_f	ggcatgggataatgagcaga	gap9_f11	agggactgtggtactgctga
seq_f57	agcaggatttctgacactggac	seq_f142	tgttaccatgaaaccagaaatcctc
seq_f58	cactaatgttagatttgaggtttcaagac	seq_f143	aatagatgaagtaagagaccctcc
seq_f59	tgctagatctagactgcatgctg	seq_f144	ggtatgaaaagatattgtatgtaaggag
gap6_f1	agaaaaaattgaaactgtggagagtataagt	gap9_f12vs2	atcattgtcagtgctcacatgggtag
seq_f60	ttgctgtaatcccagctacttg	seq_f145	cattccaaccagctagtagtatataagtg
seq_f61	gagagagcatagtatgtggaccttg	seq_f146	ggtaacatcttttgaagatccag
gap6_f2vs2	aaacaatgactgaatcctcacactctgg	seq_f147	gccttcaatccatgaatgtgtatg
seq_f62	cttctgggaagggtgagaacttag	seq_f148	atagcagtagtgagtgtagacatgctg
seq_f63	agaaggaacaaaactctggacacacgatg	seq_r2	ccatatttctgtcagtaagaacaaatag
pr10_f	aatgctgctattgacagccc	seq_r3	agatacatagctggagtggaattac
seq_f64	gtgccgctgttaactgtcatac	seq_r7	gtcctgctgggaggtgttg
seq_f65	gttttacatatttaagtagttaactggtgc	seq_r8	gagattaataatgagatcatgacttttc

seq_f66	gggtgatgtaaaactatgaatatgtgctgg	gap5_r1	agacaaagatgccacattcatca
seq_f67	tgacacagttccaaaagtgtcg	gap5_r2	gttaaatccttgaagtactagataaagaatc
pr11_fvs2	tgagtttaggtaagttagaggatga	seq_r81	ctttctatccatgagcatggaatgttttc
seq_f68	caaacctgaaagatgttggtaggag	seq_r82	ctcctgtgtccatgtgttcattg
seq_f69	gtcccaggatcagcgctatctg	seq-pr16_fr	ttcaccaacgacatttcaagt
seq_f70	agtatctcattcatagtacctgccg	seq_r88	ttggatgatgcttcagacatga
seq_f71	atggggttacaatctcccagacg	seq_r97	taagcttctgatgtgctgctgga
pr_12fvs2	tgtaactttgtggcaggcactg	seq_r98	ctgttattggtgataagaatgctgtgga
seq_f72	agaaaaggcctgggaagattaggag	seq_r108	gaggcctttttccaacaca
seq_f73	cgggcaagtagtttagtaagaatatgg	pr19-r	gcagcatgatcacagcctta
seq_f74	cagttacatgaaggtattctggcatagc	gap2_f1vs3rev	ctggtgttttgttaacgtgtcga
seq_f75	tagggatgctcagctggagatctac	seq_r14	gaacctcctgttcaagtgtacc
pr13_f	caatgtggacagatgatttgct	seq_r8vs2	gaaatcgacatattgtaactattgaaaggg
seq_f76	gattaatagtctccaaacccattcagg	gap5_r2vs2	gcaatggatcctaacaaaagttaaactcctg
seq_f77	ctgagaaatctatttcaagtacaatgc	seq_r109	caagtgtaaatattgagtctgatgaaagc
seq_f78	cctcataaacttctaaggttagcacaacc	seq_r14vs2	ctgagacggagctcgtctctg
seq_f79	ctctttaaagatggactatctcaagg	seq_r37	gctctgtcgcccaggctggag
pr14_fvs2	aataccaatgcttattatctgctctccag	seq_r38	gatataaatcaatctgcattgagaaaaacgc
seq_f80	atgccaaaatattttacagatgagtg	seq_r43	catcgcttaattattgatcgcaatgg
seq_f81	catgatcatcataatagacacagaaaaggc	seq_r44	gctgtttttccccataaaacttaacg
seq_f82	cagctaacaaggggaagtgaaggac	gap6_f1rev	tgcaacctccacctcctgttct
seq_f83	ctcaagatggattaaggcttaaatgtg	seq_r60	tccatccggaatcattgtgtctc
pr15_f	ggttcagaccctctccttga	seq_r67	caccctaagtgaccagctc
seq_f84	gaggcacaatctaacctgtgaataaccag	pr11_fvs2rev	tcctaattcccccaagactag
seq_f85	caacaacagtgcatctacttaacgag	gap7_frev	caatgagcgggttccgtggacgtag
seq_f86	gatgagtaaatatcattactaaagtggacagtg	seq_r93	cgtaacagatggggttgggtggaatg

Unbeschriebene genetische Varianten aus dem Bereich 50 kb strangauf- und abwärts der Position des SNPs rs12479448

No.	Genomposition [Build 36.3]	Austausch	No.	Genomposition [Build 36.3]	Austausch
1	35030205	T>A	190	35079904	A>C
2	35030249	T>A	191	35079958	G>A
3	35030256	T>C	192	35080027	T>G
4	35030263	T>C	193	35080231	C>insT
5	35030280	T>C	194	35080246	G>A
6	35030379	T>A	195	35081290	G>A
7	35030808	A>G	196	35081667	C>T
8	35031665	A>T	197	35082430	G>A
9	35031993	A>G	198	35082446	T>G
10	35032008	A>G	199	35082459	C>A
11	35032090	C>T	200	35082484	T>A
12	35032999	T>C	201	35082544	T>A
13	35033191	T>A	202	35082554	T>A
14	35033314	G>A	203	35082936	T>delT
15	35033345	T>A	204	35083106	G>A
16	35033381	T>A	205	35083257	T>C
17	35033611	A>G	206	35083631	A>delA
18	35033945	T>C	207	35083763	A>delA
19	35033964	C>T	208	35084227	A>C
20	35034222	A>C	209	35084752	A>delA

21	35034804	T>delTAT	210	35084838	G>T
22	35035929	T>G	211	35085128	A>G
23	35035983	A>G	212	35085475	T>C
24	35036059	G>A	213	35085565	T>C
25	35036067	T>G	214	35085710	C>T
26	35036172	T>G	215	35085842	A>delA
27	35036192	A<delA	216	35086254	A>G
28	35036458	A>T	217	35086505	C>T
29	35036465	C>T	218	35086742	A>C
30	35036486	G>A	219	35087251	T>A
31	35036488	G>A	220	35087384	G>A
32	35036490	G>T	221	35087430	A>C
33	35036497	T>C	222	35087523	A>G
34	35036518	T>C	223	35087555	A>G
35	35036732	T>insT	224	35087557	A>G
36	35037290	T>C	225	35087573	T>A
37	35037488	C>T	226	35087787	A>T
38	35037655	G>A	227	35088341	T>insT
39	35037831	G>A	228	35089235	G>C
40	35038123	A>insG	229	35089242	T>A
41	35038372	C>A	230	35089244	G>T
42	35038506	G>C	231	35090406	A>insA
43	35038545	T>G	232	35090502	T>C
44	35038726	C>T	233	35090854	C>T
45	35038868	A>C	234	35091019	C>T
46	35038982	G>A	235	35091226	A>G
47	35039266	T>ins15xA	236	35091276	A>G
48	35039487	C>A	237	35091291	G>A
49	35039692	T>C	238	35091369	G>C
50	35039758	C>T	239	35091439	T>C
51	35040199	G>A	240	35091474	C>T
52	35040323	A>T	241	35091599	G>A
53	35040324	A>T	242	35091877	T>A
54	35040616	C>G	243	35092033	T>C
55	35040617	C>G	244	35092071	A>insA
56	35040777	C>T	245	35092258	A>T
57	35040809	C>T	246	35092262	A>T
58	35041588	C>T	247	35092297	A>G
59	35041717	T>insT	248	35092323	C>T
60	35041926	A>insT	249	35092561	T>C
61	35042241	G>A	250	35092586	T>C
62	35042269	C>insAG	251	35092629	C>T
63	35042271	G>C	252	35092655	C>T
64	35042497	G>C	253	35092679	A>G
65	35042559	A>GATA	254	35092882	G>T
66	35042602	T>insAGAT	255	35092943	C>A
67	35043135	A>T	256	35092945	A>G
68	35043597	C>T	257	35092963	T>C
69	35043801	C>A	258	35092973	C>A
70	35043843	G>A	259	35093010	A>C
71	35043844	A>G	260	35093049	A>insA
72	35043848	C>A	261	35093086	A>T
73	35043877	T>A	262	35093235	G>T
74	35044247	C>T	263	35093236	C>insC

75	35044865	C>T	264	35093331	C>G
76	35045405	T>C	265	35093721	C>G
77	35045795	A>insA	266	35093886	A>C
78	35046285	T>C	267	35093949	A>T
79	35046492	T>C	268	35094247	T>C
80	35046735	C>G	269	35094339	C>G
81	35046979	C>T	270	35094553	A>C
82	35047087	C>G	271	35095835	T>G
83	35047165	G>insTAGAG	272	35095870	T>C
84	35047719	T>G	273	35095994	T>A
85	35047993	C>delAG	274	35095997	T>C
86	35048022	G>C	275	35096056	A>C
87	35048534	G>A	276	35096058	A>C
88	35048584	T>C	277	35096076	C>T
89	35048774	C>A	278	35096137	C>A
90	35048897	T>C	279	35096241	G>A
91	35049920	C>T	280	35096258	C>G
92	35050049	C>T	281	35096289	C>T
93	35050060	A>insA	282	35096572	A>G
94	35051252	A>G	283	35096573	T>C
95	35051871	T>G	284	35096601	G>A
96	35052012	A>G	285	35096753	T>A
97	35052307	A>G	286	35096805	G>insG
98	35052508	C>T	287	35096820	T>A
99	35052713	T>G	288	35097005	G>T
100	35053305	T>C	289	35097097	T<A
101	35053520	G>A	290	35097103	T>A
102	35053542	C>T	291	35097118	A>G
103	35053623	G>C	292	35097266	T>C
104	35053658	G>T	293	35097337	A>G
105	35053732	G>A	294	35097423	G>T
106	35054771	G>A	295	35097504	T>C
107	35054930	C>G	296	35097761	G>A
108	35055965	C>A	297	35098680	G>T
109	35056817	C>T	298	35098784	A>C
110	35056907	G>A	299	35098837	A>G
111	35057648	T>C	300	35098992	C>T
112	35058111	T>C	301	35099679	T>A
113	35058554	T>C	302	35099808	C>T
114	35058740	A>delATCA	303	35099840	T>G
115	35059811	G>A	304	35099907	T>C
116	35060071	A>G	305	35099909	C>G
117	35060223	C>A	306	35099934	G>A
118	35060380	A>T	307	35099948	T>C
119	35060462	C>T	308	35100001	C>A
120	35060736	G>C	309	35100539	G>A
121	35061410	A>G	310	35100637	G>delG
122	35061437	A>G	311	35100951	C>G
123	35061467	T>A	312	35101330	A>T
124	35061706	T<delT	313	35101925	C>A
125	35061903	T>A	314	35102127	G>insG
126	35061930	T>C	315	35102833	A>T
127	35062369	T>delT	316	35104290	A>C
128	35062851	T>C	317	35104825	C>T

129	35062904	T>A	318	35105599	A>G
130	35062955	G>delG	319	35106111	A>G
131	35063732	T>delT	320	35106579	G>A
132	35063773	T>delT	321	35106818	C>T
133	35064115	G>A	322	35107082	A>insA
134	35064134	A>G	323	35107347	A>C
135	35066470	C>T	324	35108657	C>delCT
136	35066582	A>delA	325	35108950	T>insA
137	35066709	A>delA	326	35109558	A>ins/delA
138	35067414	C>T	327	35110859	C>A
139	35067565	A>T	328	35111496	T>C
140	35068188	C>T	329	35111665	A>G
141	35068547	A>G	330	35111669	G>A
142	35069005	C>G	331	35111676	C>A
143	35069222	A>C	332	35111824	T>G
144	35069292	G>A	333	35112187	T>A
145	35069344	C>G	334	35112444	G>A
146	35069399	A>C	335	35112451	T>G
147	35069434	G>A	336	35112509	G>T
148	35070363	A>delA	337	35112601	A>T
149	35070545	G>A	338	35113790	C>G
150	35070557	G>C	339	35114654	G>T
151	35070941	A>G	340	35115127	A>G
152	35071109	C<A	341	35115590	C>T
153	35071136	T>G	342	35115908	A>G
154	35071148	A>G	343	35116010	A>insA
155	35071248	T>C	344	35116194	G>C
156	35071326	G>C	345	35116255	G>C
157	35071331	C>T	346	35116282	A>insA
158	35071470	G>A	347	35116464	A>G
159	35071665	G>A(del?)	348	35117736	A>delA
160	35071734	G>T	349	35117878	G>C
161	35072471	C>T	350	35118019	G>T
162	35072709	C>A	351	35119164	A>G
163	35073040	T>insT	352	35119366	T>G
164	35074312	TT>delTT	353	35119395	G>T
165	35074356	A>C	354	35120244	T>insTTAT
166	35074801	A>insA	355	35120336	A>G
167	35075213	T>A	356	35120617	A>T
168	35075340	C>T	357	35120891	G>delG
169	35075411	T>insATTTT	358	35121071	A>G
170	35075923	T>delTCCC	359	35121606	T>A
171	35075939	C>insT/C/CTCCC/A	360	35121710	G<A
172	35075972	T>C	361	35121786	A>G
173	35075973	T>C	362	35121823	G>A
174	35075974	C>T	363	35121849	A>insA
175	35075977	C>T	364	35122628	C>G
176	35075986	C>T	365	35123190	T>C
177	35075991	T>C	366	35123669	C>T
178	35076003	C>T	367	35123769	A>C
179	35076007	T>C	368	35125066	G>C
180	35076008	C>T	369	35126527	T>C
181	35077143	C>insC	370	35126739	A>C
182	35077699	G>A	371	35126885	A>G

183	35077896	T>C	372	35126892	G>delG
184	35077964	A>G	373	35127468	A>G
185	35078787	C>T	374	35127806	C>T
186	35078890	A>G	375	35128055	G>C
187	35079271	A>G	376	35128443	T>delTTG
188	35079698	A>G	377	35128964	A>C
189	35079879	C>T	378	35129285	G>A
190	35079904	A>C	379	35129352	A>G
191	35079958	G>A	380	35129814	G>A

Ins: Insertion; del, Deletion

Genomische Positionen der Transkripte

DA227815

	Arterie	Gehirn Isoform 1	Gehirn Isoform 2	annotiertes Transkript [NCBI]
Exon 1				34878421-34878470
Intron 1				34878471-34909933
Exon 2	34909950-34910063	34909996-34910063		34909934-34910063
Intron 2	34910064-35237653	34910064-35237653		34910064-35237653
Exon 3	35237654-35237725	35237654-35237725	?-35237725	35237654-35237725
Intron 3	35237726-35238116	35237726-35241345	35237726-35238116	35237726-35238116
Exon 4	35238117-35238198	35241346-35241409	35238117-35238198	35238117-35238198
Intron 4	35238199-35241345	35241410-35241500	35238199-35241345	35238199-35241345
Exon 5	35241346-35241409	35241501-?	35241346-35241409	35241346-35241409
Intron 6	35241410-35241500		35241410-35253098	35241410-35241500
Exon 6	35241501-35241627		35253099-35253450	35241501-35241682
Intron 6	35241628-35242368			
Exon 7	35242369-35242400-			

NCBI Genome Build 36.3

AA861168

	Arterie Isoform 1	Arterie Isoform 2	Hoden	annotiertes Transkript [NCBI]
Exon 1	35073478-35074232	35072919-35074232	?-35074188	35073840-35074195

NCBI Genome Build 36.3

AI218168/AW665463/BX091678

	Arterie/ Vene	Haut	Arterie	Beinvene	annotiertes Transkript [NCBI]
Exon 1				35076774-35077597	35146971 -35146879
Intron 1					35146878-35142666
Exon 2	35085840-35085620				35142667-35142492
Intron 2	35085619-35078763				35142491-35078582
Exon 3	35078764-35078527	35085840-35085620			35078582-35078527
Intron 3	35078526-35077592	35085619-35077592			35078526-35077592
Exon 4	35077593-35077532	35077593-35077532			35077593-35077532
Intron 4	35077531-35077074	35077531-35077074	35142387-35143395		35077531-35077074
Exon 5	35077075-?	35077075-?			35077075-35076841

NCBI Genome Build 36.3

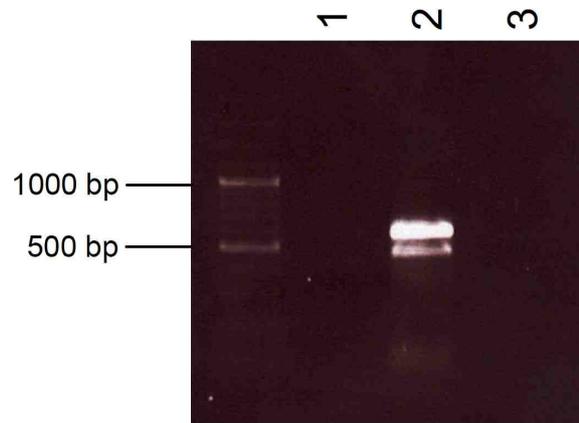
PCR-Bedingungen zur Amplifikation der 5'-RACE-Produkte der Transkripte

	Oligonukleotide	Annealing- Temperatur	Extentions- Temperatur	Extentions- zeit	Zyklen	Verdünnung
AI218168, Exon 1, Beinvene						
1. PCR	AI2181.2_E1o_rev CFB_for	54°C	72°C	2 min	25	unverdünnt
2. PCR	AI218168_revII CapFinderB	54°C	68°C	2 min	30	1:500
3. PCR	AI218168.1l_rev CapFinderB	58°C	68°C	1,5 min	30	1:500
AA861168, Arterie, Isoform 1						
1. PCR	oligo(dT)anchor Hs.542102_CapR1	60°C	72°C	2 min	20	unverdünnt
2. PCR	anchor-Primer Hs.542102_CapR2	60°C	72°C	2 min	30	1:30
AA861168, Arterie, Isoform 2						
1. PCR	oligo(dT)anchor Hs.542102_rev	55°C	72°C	2 min	20	unverdünnt
2. PCR	Hs.542102ri3 anchor-Primer	54°C	72°C	2 min	30	1:30
DA227815, Exon 6, Arterie und Gehirn						
1. PCR	DA227815E6o3_rev CFB_for	58°C	72 °C	2 min	20	unverdünnt
2. PCR	DA227815E6_RNAf CapFinderB	55°C	72°C	2 min	30	1:200

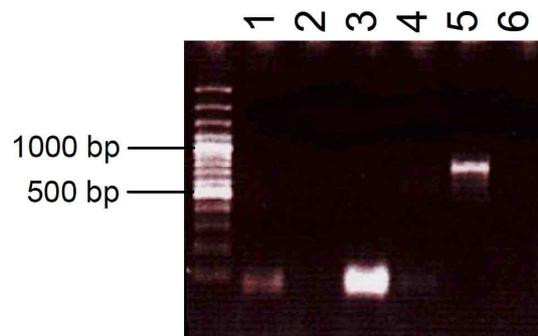
 PCR-Bedingungen zur Amplifikation der 3'-RACE-Produkte der Transkripte

	Oligonukleotide	Annealing- Temperatur	Extentions- Temperatur	Extentions- zeit	Zyklen	Verdünnung
AI218168, Exon 1, Beinvene						
1. PCR	AI218168.1I_for Oligo(dT)	60°C	72°C	2 min	20	unverdünnt
2. PCR	AI218168.2I_for OdT-T7 II	62°C	72°C	2 min	30	unverdünnt
AA861168, Hoden						
1. PCR	Hs.542102_for Oligo(dT)	55°C	72°C	1,5 min	40	unverdünnt
2. PCR	Hs.542102_for-2 OdT-T7 II	58°C	72°C	2min	40	unverdünnt
DA227815, Exon 6, Beinvene						
1. PCR	DA227815-Ex2.1F Oligo(dt)	58°C	72°C	2 min	20	unverdünnt
2. PCR	DA227815E3f Hs.542102_for-2	60°C	72°C	2 min	30	1:200
DA227815, Exon 6, Arterie						
1. PCR	DA227815E6_for Oligo(dT)	58°C	72°C	2 min	20	unverdünnt
2. PCR	DA227815E6_RNAr OdT-T7 II	58°C	72°C	2 min	30	1:100
DA227815, Exon 2, Arterie						
1. PCR	DA227815E3f Oligo(dT)	50°C	72°C	2 min	20	Unverdünnt
2. PCR	DA227815E4f OdT-T7 II	58°C	72°C	2 min	30	1:500

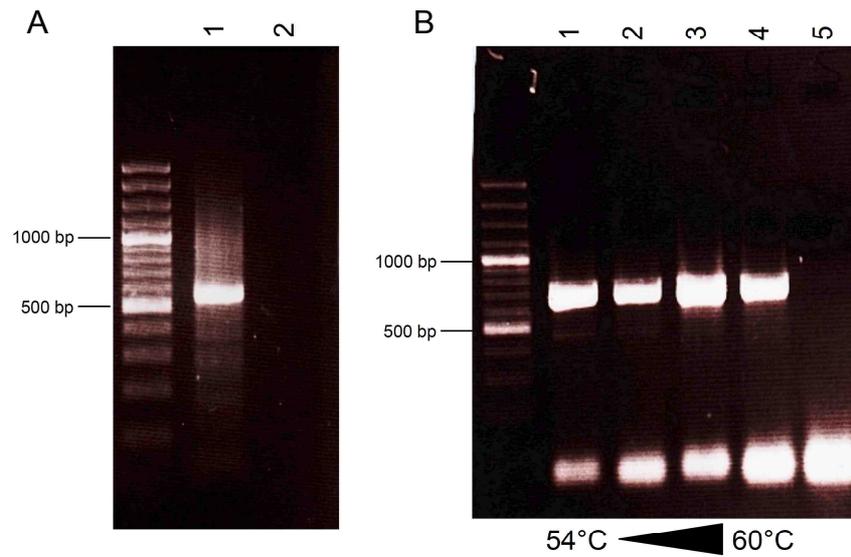
Amplifikationsprodukte der 5'- und 3'-RACE

**3'-RACE von AI218168/AW665463/BX091678**

Amplifikationsprodukte von AI218168 und AW6653 aus dem 1. Exon aus arterieller cDNA aus einer PCR mit (*lane 1*) und ohne Q-Solution (*lane 2*) aus einer 1. RACE-PCR mit Q-Solution. Ohne Q-Solution-Zusatz konnten zwei, etwa 500 und 600 bp lange, Fragmente amplifiziert werden. Über Klonierung und DNA-Sequenzierung wurde das größere Fragment als gespleißte Isoform des Transkripts identifiziert. *Lane 3*, Leerwert

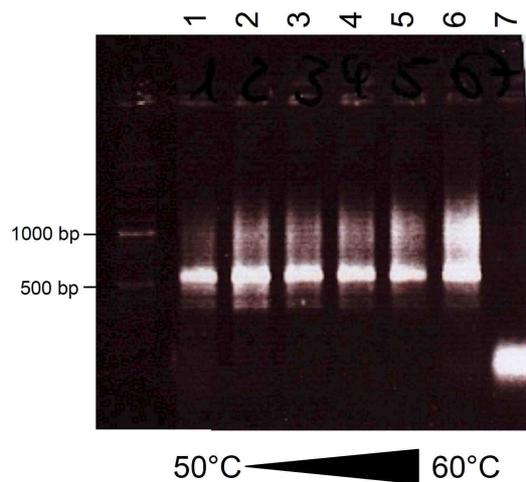
**3'-RACE von AI218168/AW665463/BX091678**

Amplifikationsprodukte einer 3'-RACE von AI218168/AW665463/BX091678 bei einer Annealingtemperatur von 58°C aus dem ersten Exon aus venöser (*lane 2 und 5*) und arterieller cDNA (*lane 1 und 4*). Die PCR wurde mit Q-Solution angesetzt (*lane 1-6*). Ausgangsprodukt war eine 1. RACE-PCR mit (*lane 1-3*) und ohne (*4-6*) Solution Q. *Lane 3 und 6*, Leerwert



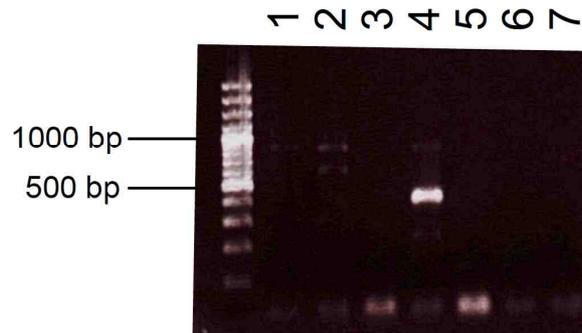
5'- RACE von AA861169

A) Cap 274 bp vor dem 5'-Ende des annotierten Transkripts. *Lane 1*, 5'RACE aus arterieller cDNA; *lane 2*, Leerwert. B) Cap 833 bp vor dem 5'-Ende des annotierten Transkripts. *Lane 1 – 4*, 5'-RACE aus arterieller cDNA bei einem Temperaturgradienten von 54 – 60°C (je 2°C Temperaturdifferenz pro *lane*); *lane 5*, Leerwert.



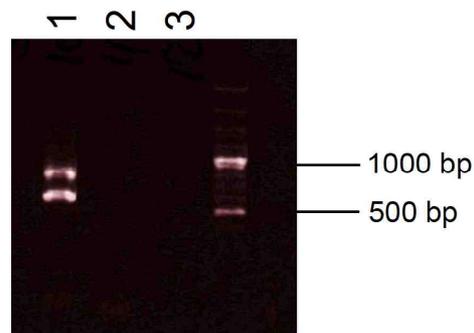
3'RACE von AA861169

3'-Ende aus arterieller cDNA. *Lane 1-6*, 3'-RACE bei einem Temperaturgradienten von 54 – 60°C (je 2°C Temperaturdifferenz pro *lane*); *lane 7*, Leerwert.



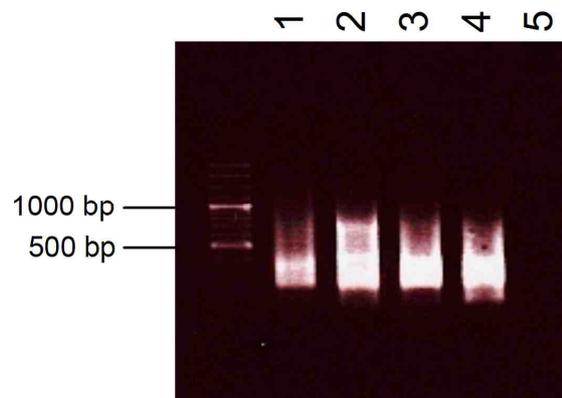
3'-RACE-Produkt von DA227815

3'-RACE-PCR aus arterieller und venöser cDNA. Nur für Arterie konnte unter diesen PCR-Bedingungen (ohne Solution Q) unter vorheriger Amplifikation ohne Solution Q eine starke Bande detektiert werden, die über Sequenzierung bestätigt worden ist. Bei den schwächeren Banden der anderen Reihen handelt es sich um unspezifische Nebenprodukte. (7) Leerwert.



3'-RACE von DA227815 aus zerebraler cDNA

3'-Ende des Gehirn-Transkripts von Brain mit zwei GSP generiert. Bei der unteren Bande handelte es sich um das gesuchte Transkript, welches durch Sequenzierung bestätigt wurde. (1) Hirn-cDNA aus pre-PCR mit Solution Q, (2) Hirn-cDNA aus pre-nested-PCR ohne Solution Q, (3) Leerwert.



5'- und 3'-RACE von DA227815 aus zerebraler cDNA

B) 5'-RACE aus Arterie (1 und 3) und Gehirn (2 und 4) mit und ohne Solution S. (Primer 193-CapFinder).

PCR-Bedingungen für die Genotypisierung genetischer Varianten

	Annealing- Temperatur [°C]	MgCl ₂ - Konzentration [mM]	Puffer / Solution S	Verhältnis for:rev
rs6543892	55	3	Puffer BD Solution S	1:5
rs12476596	58	3	Puffer B Solution S	1:5
rs34996117	58	3	Puffer B Solution S	5:1

RNA

Gewebe	Firma	Geschlecht des Donors	Alter des Donors [Jahre]	Gewebetyp	Art
Arterie	Biochain	♂	44	gesund	total RNA
Gehirn	BioChain	♂	23	gesund	total RNA
Leber	Biochain	♂	64	gesund	total RNA
Lunge	BioChain	♂	26	gesund	total RNA
Vene	BioChain	♂	24	gesund	total RNA
Beinvene	selbst isoliert, Asklepios Klinik St. Georg	unbekannt, mehrere Donoren	unbekannt	unbekannt	total RNA
Pulmonares mikrovaskuläres Endothel	selbst isoliert aus HPMEC-Zelllinie	♂	63	gesund	total RNA

cDNA

Gewebe	Firma	Geschlecht des Donors	Alter des Donors [Jahre]	Gewebetyp	Synthese
Arterie I	Biochain	♂	68	gesund	oligo(dT) Primer
Arterie II	selbst synthetisiert, RNA von Biochain	♂	44	gesund	oligo(dT) Primer
Dickdarm	PrimerDesign	k. A.	k. A.	gesund	oligo(dT) + random Primer
Fettgewebe	PrimerDesign	k. A.	k. A.	gesund	oligo(dT) + random Primer
Gehirn I	PrimerDesign	k. A.	k. A.	gesund	oligo(dT) + random Primer
Gehirn II	selbst synthetisiert, RNA von Biochain	♂	23	gesund	oligo(dT) Primer
Haut	BioChain	♂ ♀	24 60	gesund	oligo(dT) Primer
Herz	PrimerDesign	k. A.	k. A.	gesund	oligo(dT) + random Primer
Hoden	PrimerDesign	k. A.	k. A.	gesund	oligo(dT) + random Primer
Leber	Biochain	♂	64	gesund	oligo(dT) Primer
Leber II	selbst synthetisiert, RNA von Biochain	♂	64	gesund	oligo(dT) Primer
Lunge I	PrimerDesign	k. A.	k. A.	gesund	oligo(dT) + random Primer
Lunge II	selbst synthetisiert, RNA von Biochain	♂	26	gesund	oligo(dT) Primer
Milz	PrimerDesign	k. A.	k. A.	gesund	oligo(dT) + random Primer
Niere	PrimerDesign	k. A.	k. A.	gesund	oligo(dT) + random Primer
Schilddrüse	PrimerDesign	k. A.	k. A.	gesund	oligo(dT) + random Primer
Skelettmuskel	PrimerDesign	k. A.	k. A.	gesund	oligo(dT) + random Primer
Periphere Blutleukozyten	Biochain	♀ ♂	63 53	gesund	oligo(dT) Primer
Plazenta	PrimerDesign	k. A.	k. A.	gesund	oligo(dT) + random Primer
Vene I	Biochain	♀ ♂	78 24	gesund	oligo(dT) Primer

EST-Sequenzen

Ein putatives Startcodon (ATG) ist fett markiert. Quelle: UCSC Genome Browser

AI218168

TTTAAAGATACTGTCTCAACATTTATTTTTCTTTTCAGT**ATGG**ATTGAACAGATAGTGTT
AACTTTGTGGCAGGCACTGTACTTAC**ATG**ATTAC**ATGGG**TTTCTTTCT**ATG**CTCTCAATT
CTTCAGGTTTTTAAAATTTTTCTCTCAAAGGAAATCAGCCTCCTTGTTGTTAATTTT
TT**ATGTAGTCATG**TTTTCTTCATTGTGATGCTTACATTGATATTGACTAGGCCTCAAGA
ACAATCTTCTCGTTCTTCACTGTTGGACCCAGGCCTGGCCTGGTAACGTTTTTCTCTCTT
GCTTAGTCCAAGTTATTTTCAGTTACCTTGAAGTATTTCTGCATAGCAATAAAA

AW665463

TTTTTTTTTACTGTCTCAACATTTATTTTTCTTTTCAGT**ATGG**ATTGAACAGATAGTGTT
TAACTTTGTGGCAGGCACTGTACTTAC**ATG**ATTAC**ATGGG**TTTCTTTCT**ATG**CTCTCAAT
TCTTCAGGTTTTTCAAATTTTTCTCTCAAAGGAAATCAGCCTCCTTGTTGTTAATTT
TTT**ATGTAGTCATG**TTTTCTTCATTGTG**ATG**CTTACATTGATATTGACTAGGCCTCAAG
AACAATCTTCTCGTTCTTCACTGTTGGACCCAGGCCTGGCCTGTTAACGTTTTTCTCTCT
TGCTTAGTCCAAGTTACTTACAGTTAC**ATGA**AGGTATTCTGGCATAGCAATAAAACAACCG
GAAAGTCTTTGAGTGTGTTGAAGTAGACATTTTTCAATACAGTTTCTAATAAATCT
ACTGTTTTCTTCTCACGGAGGTGTGAAGGTGTGCAGAGTACCAATAAAGACTTCATAAAG
TCGTAAGCACTGTGTGAAGAGTG

BX901678

TTTTTTTGTGCCATTTTTATTTTTTTTTGTGCCATCCAGTTTTTTATTGTGCCATCCAGT
TCAACCTACAAGCCC**ATGGATGG**TACCTACAGCTTTGGCAAGAACCCCACTTTACTTCC
ATACAGATTTGAAACCAGACATAGCTCTTTCACACTA**ATG**TTGGATCC**ATGT**CAAATACA
TAGAGGACTACGAAGNTCA**ATGAAATG**TATTCTTGCCAGGCTCATCTCGAAATCCTGGG
CTCAAACAAAATTAGCTGG**ATG**TGCTGGTTGAAACATTCTGTTCTTGAACAGAGGAAAA
ACACTCTTACACAGTGCTTACGAGTTTCTGNAGTCTTTATTGGTACTCTGCACACCTTC
ACACCTCCGTGAGAAGAAAACAGTA

AA861169

TTTTGTTTGAACACTTGGCATTGATATTTATTCACCACAGAGGAAAATAACAACACTACTT
TTAGAAATATACATACAGAATTAATTCCTGAATTTTTCTATTGGAACATAATTCTAAAT
GCAGTTCCATGTTGCTTATCTAGCAAACCATAGGCTCAA**ATG**TTTTAAAATACTTT**GAT**
GGACAGCAACTTTTTGGAAGTGTGTCAGGGACAGAAATTTAGACAGAGATTGGCTTTTACAG
AATGCTTAAACAGTTATTTTCAGACTTTTTGGCA**ATG**ACCCACAGTAGAACGCATTTGA
TTTTACAACCTCAGTATAGGCAAACTTTAAACACATATACAAATATATAATATAGAC

DA227815

ACATAACTGCTTTCTCTGAAATACCACCTTGTAAGATAGAGAAA**ATG**TTGCACAGAT
TTCTCTGTGTC**ATG**CAGCTGCT**ATGGGGGATGG**AGGAGGAGTGGCATTGGCAATTCAAG
ACTGCCTTTCTACCCTTTTCAGTGCCTCTTTTCAGTGATACGAAGCTAAAACCAGATACT
GCTTTCATAGTACTTTTCTGACCACGGTTTTCTGGAGTTTTTTTCGT**ATGG**AGTTTGTAAAG
CTTCTGAC**ATG**GATTATAACTGTACTT**GATG**GAGCACAGTTATAATAGCTGCTCTAGAG
TCCTTGATAATTCCCACATCTGAGTCATCTCAAGTTATTTTTGGCAGATTTT**GATGGATG**
TACATTGATGGATAGCCACGTGGATCCAGAAAATAAAGACTTCTACTCTGCATTTGACTA
AAGCAGAAGATATGCAAGGAAGCTCAGTCTCTCAATTGAGTGGATCCTGAAATTTATCCT

GTTGTCTTCAGCTACTAAGGCTAATTTGCAAATCACACTGTGCACGGGTAAGCATAAACT
GTAAGCATATATATTAATCTTCATTTTTTTAAATTGTAAT

Proteinsequenzen der Transkripte

AI218168

5'3' Offener Leserahmen 1

FKD TVSTFIFSF S Met D Stop TDSVNFVAGTVLT Stop LHGFLS Met LSILQVF
KIFSSQKEISLLVNFVFLCSHVFLHCDAYIDID Stop ASRTIFSFFTVGP RP
GLVTF FSLA Stop SKLFQLP Stop SISA Stop Q Stop

5'3' Offener Leserahmen 2

LKILSQHLFFLSVWIEQIVLTLWQALYLHDY Met GFFLCSQFFRFLKFFP
LKRKSASLLLIFYVVMet F SFIV Met LTLILTRPQE QSSRSSLLDPGLAW
Stop RFSLLLSPSYFSYLEVFLHSNK

5'3' Offener Leserahmen 3

Stop RYCLNIYFFFQYGLNR Stop C Stop LCGRHCTY Met ITWVSFYALNSSGF
Stop NFFLSKGNQPPCC Stop FF Met Stop SCFPSL Stop CLH Stop Y Stop LGLKNN
LLVLHCWTQAWPGNVFLSCLVQVISVTLKYFCIAIK

3'5' Offener Leserahmen 1

FYCYAEILQGN Stop NNLD Stop AREKNVTRPGLGPTVKNEKIVLEA Stop SI
S Met Stop ASQ Stop RKT Stop LHKKLTTRRLISF Stop EEKILKT Stop RIESIERN
PCNHVSTVPATKLTLSVQSILKEKINVETVSL

3'5' Offener Leserahmen 2

FIAMet QKYFKVTEITWTKQERKTLPGQAWVQQ Stop RTRRLFLRPSQYQ
CKHHNEGKHDYIKN Stop QQG Stop FPFERKKF Stop KPEELRA Stop KETHV
IMet Stop VQCLPQS Stop HYLFPY Stop KKK Stop Met LRQYL Stop

3'5' Offener Leserahmen 3

LLLCRNTSR Stop LK Stop LGLSKREKRYQARPGSNSEEREDCS Stop GLVNI
NVSIT Met KEN Met TT Stop KINNKEADFLLRGKNFKNLKN Stop EHRKKP Met
Stop SCKYSACHKVNTICSIHTERKNKC Stop DSIFK

AW665463

5'3' Offener Leserahmen 1

FFFYCLNIYFFFQYGLNR Stop C Stop LCGRHCTY Met ITWVSFYALNSSGF
QNFFLSKGNQPPCC Stop FF Met Stop SCFPSL Stop CLH Stop Y Stop LGLKNNLL
VLHCWTQAWPVNVFLSCLVQVTSVT Stop RYSGIAIKQPE S PLSVFEVDI
FQYSF Stop YKSTVFFSRRCEGVQSTNKDFINS Stop ALCEE

5'3' Offener Leserahmen 2

FFFTVSTFIFSF S Met D Stop TDSVNFVAGTVLT Stop LHGFLS Met LSILQVF
KIFSSQKEISLLVFNFLCSHVFLHCDAYIDID Stop ASRTIFSFFT V GPRP
GLLTF FSLA Stop SKLLQLHEGILA Stop Q Stop NNRKVL Stop VCLK Stop TFFNT
VSNINLLFSSHGGVVKVCRVPIKTS Stop TRKHCVKS

5'3' Offener Leserahmen 3

FFLLSQHLFFLSVWIEQIVLTLWQALYLHDY Met GFFLCSQFFRFSKFFP
LKRKSASLLLIFYVV Met F SFIV Met LTLILTRPQE QSSRSLLDPGLAC
Stop RFSLLLSPSYFSY Met KVFWSNKT TGKSFECV Stop SRHFSIQFLI
Stop IYCFLLTEV Stop R CAEYQ Stop RLHKL VSTV Stop RV

3'5' Offener Leserahmen 1

HSSHAYEF Met K SLLVLC T PSHLREKKTVDLY Stop KLY Stop K Met STSNT
LKGLSGCFIAMet PEYLHVTEVTWTKQERKTLTGQAWVQQ Stop RTRRLF
LRPSQYQCKHHNEGKHDYIKN Stop Q QGG Stop FPFERKKF Stop KPEELRA
Stop KETHVI Met Stop VQCLPQS Stop HYLFNPY Stop KKK Stop Met LRQ Stop KK

3'5' Offener Leserahmen 2

TLHTVLTSL Stop SLYWYSAHLHTSVRRKQ Stop IYIRNCIEKCLLQTHSKD
FPVVL L LCQNTF Met Stop LK Stop LGLSKREKR Stop QARPGSNSEEREDCS
Stop GLVNINVSIT Met KEN Met TT Stop KINNKEADFLLRGKNFENLKN Stop E
HRKKP Met Stop SCKYSACHKVNTICSIHTERKNKC Stop DSKKK

3'5' Offener Leserahmen 3

LFTQCLRVEVFIGTLHTFTPP Stop EENSRIELETVLKNVYFKHTQRTFR
LFYCYARIPSCN Stop SNLD Stop AREKNVNR PGLGPTVKNEKIVLEA Stop S
IS Met Stop ASQ Stop RKT Stop LHKKLTTRRLISF Stop EEKILKT Stop RIESIERN
PCNHVSTVPATKLTLSVQSILKEKINVETVKKK

BX091678

5'3' Offener Leserahmen 1

FFCAIFIFFCAIQFFIVPSSSTYKP Met DGTYSFGKNPHFTSIQI Stop NQT
Stop LFHTNVGS Met SNT Stop RTTKXNE Met YSCPGSSRN PGLKQKLAGCA
G Stop NILFLNRGKTLFTQCLRVSXVFIGTLHTFTPP Stop EENS

5'3' Offener Leserahmen 2

FFVPFLFFFVPSSFLLCHPVQPTSPW Met VPTALARTPTLLPYRFETR H
SSFTL Met LDPCQIHRGLRXS Met KCILAQAHL EILGSNKN Stop LDVLVETF
CS Stop TEEKHSSHAYEFLX SLLVLC T PSHLREKKT V

5'3' Offener Leserahmen 3

FLCHFYFFLCHPVFYCAIQFNLQAAGWYLQLWQEPPLYFHTDLKPDIA
LSH Stop CWIHVKYIEDYEXQ Stop NVFLPRLISKSWAQT KISW Met CWLKH
SVLEQRKNTLHTVLT SFXS LYWYSAHLHTSVRRKQ

3'5' Offener Leserahmen 1

Y C F L L T E V Stop R C A E Y Q Stop R X Q K L V S T V Stop R V F F L C S R T E C F N Q H I Q L
I F V Stop A Q D F E Met S L G K N T F H Stop X S Stop S S Met Y L T W I Q H Stop C E R A Met S
G F K S V W K Stop S G G S C Q S C R Y H P W A C R L N W Met A Q Stop K T G W H K K K Stop
K W H K K

3'5' Offener Leserahmen 2

T V F F S R R C E G V Q S T N K D X R N S Stop A L C E E C F S S V Q E Q N V S T S T S S Stop
F L F E P R I S R Stop A W A R I H F I X L R S P L C I Stop H G S N I S V K E L C L V S N L Y G S K
V G V L A K A V G T I H G L V G Stop T G W H N K K L D G T K K N K N G T K K

3'5' Offener Leserahmen 3

L F S S H G G V K V C R V P I K T X E T R K H C V K S V F P L F K N R Met F Q P A H P A N F C L
S P G F R D E P G Q E Y I S L X F V V L Y V F D Met D P T L V Stop K S Y V W F Q I C Met E V K
W G F L P K L Stop V P S Met G L Stop V E L D G T I K N W Met A Q K K I K Met A Q K

AI/AW/BX

5'3' Offener Leserahmen 1

K D T V S T F I F S F S Met D Stop T D S V T L W Q A L Y L H D Y Met G F F L C S Q F F R F L K F
F P L K R K S A S L L L I F Y V V Met F S F I V Met L T L I L T R P Q E Q S S R S S L L D P G L A C
Stop R F S L L L S P S Y F S Y Met K V F W H S N K T T G K S F E C V Stop S R H F S I Q F L I
Stop I Y C F L L T E V Stop R C A E Y Q Stop R L Q K L V S T V Stop R V F F L C S R T E C F N

5'3' Offener Leserahmen 2

K I L S Q H L F F L S V W I E Q I V Stop L C G R H C T Y Met I T W V S F Y A L N S S G F Stop N F
F L S K G N Q P P C C Stop F F Met Stop S C F P S L Stop C L H Stop Y Stop L G L K N N L L V L H
C W T Q A W P V N V F L S C L V Q V I S V T Stop R Y S G I A I K Q P E S P L S V F E V D I F Q Y
S F Stop Y K S T V F F S R R C E G V Q S T N K D F R N S Stop A L C E E C F S S V Q E Q N V S

5'3' Offener Leserahmen 3

R Y C L N I Y F F F Q Y G L N R Stop C N F V A G T V L T Stop L H G F L S Met L S I L Q V F K I F
S S Q K E I S L L V N F L C S H V F L H C D A Y I D I D Stop A S R T I F S F F T V G P R P G L L
T F F S L A Stop S K L F Q L H E G I L A Stop Q Stop N N R K V L Stop V C L K Stop T F F N T V S
N I N L L F S S H G G V K V C R V P I K T S E T R K H C V K S V F P L F K N R Met F Q

3'5' Offener Leserahmen 1

V E T F C S Stop T E E K H S S H S A Y E F L K S L L V L C T P S H L R E K K T V D L Y Stop K L Y
Stop K Met S T S N T L K G L S G C F I A Met P E Y L H V T E I T W T K Q E R K T L T G Q A W V
Q Q Stop R T R R L F L R P S Q Y Q C K H H N E G K H D Y I K N Stop Q Q G G Stop F P F E R K K
F Stop K P E E L R A Stop K E T H V I Met Stop V Q C L P Q S Y T I C S I H T E R K N K C Stop D S
I F

3'5' Offener Leserahmen 2

L K H S V L E Q R K N T L H T V L T S F Stop S L Y W Y S A H L H T S V R R K Q Stop I Y I R N C I
E K C L L Q T H S K D F P V V L L L C Q N T F Met Stop L K Stop L G L S K R E K R Stop Q A R P
G S N S E E R E D C S Stop G L V N I N V S I T Met K E N Met T T Stop K I N N K E A D F L L R G K
N F K N L K N Stop E H R K K P Met Stop S C K Y S A C H K V T L S V Q S I L K E K I N V E T V S

3'5' Offener Leserahmen 3

Stop N I L F L N R G K T L F T Q C L R V S E V F I G T L H T F T P P Stop E E N S R F I L E T V L K
N V Y F K H T Q R T F R L F Y C Y A R I P S C N Stop N N L D Stop A R E K N V N R P G L G P T V
K N E K I V L E A Stop S I S Met Stop A S Q Stop R K T Stop L H K K L T T R R L I S F Stop E E K I
L K T Stop R I E S I E R N P C N H V S T V P A T K L H Y L F N P Y Stop K K K Stop Met L R Q Y L

AA861169

5'3' Offener Leserahmen 1

F C L N Y L P L I F I H H R G K Stop Q L L L E I Y I Q N Stop I H Stop I F H W N I I L N A V P C C L
S S K T I G S N V Stop K Y F D G Q Q L L E L C Q G Q N L D R D W L L Q N A Stop T S Y F Q T F
W Q Stop P T V E R I Stop F Y N S V Stop A K L Stop T H I Q I Y N I D

5'3' Offener Leserahmen 2

F V Stop T T C H Stop Y L F T T E E N N N Y F Stop K Y T Y R I K F T E F F I G T Stop F Stop Met Q
F H V A Y L A K P Stop A Q Met F K N T L Met D S N F W N C V R D R I Stop T E I G F Y R Met L K
P V I F R L F G N D P Q Stop N A F D F T T Q Y R Q N F K H I Y K Y I I Stop

5'3' Offener Leserahmen 3

L F E L L A I D I Y S P Q R K I T T T F R N I H T E L N S L N F S L E H N S K C S S Met L L I Stop Q
N H R L K C L K I L Stop W T A T F G T V S G T E F R Q R L A F T E C L N Q L F S D F L A Met T H
S R T H L I L Q L S I G K T L N T Y T N I Stop Y R

3'5' Offener Leserahmen 1

V Y I I Y L Y Met C L K F C L Y Stop V V K S N A F Y C G S L P K S L K I T G L S I L Stop K P I S V
Stop I L S L T Q F Q K L L S I K V F L N I Stop A Y G F A R Stop A T W N C I Stop N Y V P Met K N
S V N L I L Y V Y F Stop K Stop L L F S S V V N K Y Q W Q V V Q T K

3'5' Offener Leserahmen 2

S I L Y I C I C V Stop S F A Y T E L Stop N Q Met R S T V G H C Q K V Stop K Stop L V Stop A F C
K S Q S L S K F C P Stop H S S K S C C P S K Y F Stop T F E P Met V L L D K Q H G T A F R I Met
F Q Stop K I Q Stop I Stop F C Met Y I S K S S C Y F P L W Stop I N I N G K Stop F K Q

3'5' Offener Leserahmen 3

L Y Y I F V Y V F K V L P I L S C K I K C V L L W V I A K K S E N N W F K H S V K A N L C L N S V
P D T V P K V A V H Q S I F K H L S L W F C Stop I S N Met E L H L E L C S N E K F S E F N S V C
I F L K V V V I F L C G E Stop I S Met A S S S N K

DA227815

5'3' Offener Leserahmen 1

T Stop L L S S Stop N T T L Stop R Stop R K C C T D F S V S C S C Y G G Met E E E W H W Q F K
T A F P T L F S A S F S D T K L K P D T A F I V L F Stop P R F S G V F S Y G V C K L P D Met D Y
N C T Stop W S T V I I A A L E S L I I P T S E S S Q V I F G R F Stop W Met Y I D G Stop P R G S R
K Stop R L L L C I Stop L K Q K I C K E A Q S L N Stop V D P E I Y P V V F S Y Stop G Stop F A N H
T V H G Stop A Stop T V S I Y I N L H F F K L Stop

5'3' Offener Leserahmen 2

H N C F P P E I P P C K D R E N V A Q I S L C H A A A Met G G W R R S G I G N S R L P F L P F S
V P L S V I R S Stop N Q I L L S Stop Y F S D H G F L E F F R Met E F V S F L T W I I T V L D G A Q
L Stop Stop L L Stop S P Stop Stop F P H L S H L K L F L A D F D G C T L Met D S H V D P E N K D
F Y S A F D Stop S R R Y A R K L S L S I E W I L K F I L L S S A T K A N L Q I T L C T G K H K L
Stop A Y I L I F I F L N C N

5'3' Offener Leserahmen 3

I T A F L L K Y H L V K I E K Met L H R F L C V Met Q L L W G D G G G V A L A I Q D C L S Y P F Q
C L F Q Stop Y E A K T R Y C F H S T F L T T V F W S F F V W S L Stop A S Stop H G L Stop L Y L
Met E H S Y N S C S R V L D N S H I Stop V I S S Y F W Q I L Met D V H Stop W I A T W I Q K I K T
S T L H L T K A E D Met Q G S S V S Q L S G S Stop N L S C C L Q L L R L I C K S H C A R V S I N
C K H I Y Stop S S F F Stop I V

3'5' Offener Leserahmen 1

I T I Stop K N E D Stop Y I C L Q F Met L T R A Q C D L Q I S L S S Stop R Q Q D K F Q D P L N
Stop E T E L P C I S S A L V K C R V E V F I F W I H V A I H Q C T S I K I C Q K Stop L E Met T Q
Met W E L S R T L E Q L L Stop L C S I K Y S Y N P C Q E A Y K L H T K K L Q K T V V R K V L
Stop K Q Y L V L A S Y H Stop K R H Stop K G Stop E R Q S Stop I A N A T P P P S P H S S C Met
T Q R N L C N I F S I F T R W Y F R R K A V Met

3'5' Offener Leserahmen 2

L Q F K K Met K I N I Y A Y S L C L P V H S V I C K L A L V A E D N R I N F R I H S I E R L S F L A
Y L L L Stop S N A E Stop K S L F S G S T W L S I N V H P S K S A K N N L R Stop L R C G N Y Q
G L Stop S S Y Y N C A P S S T V I I H V R K L T N S I R K N S R K P W S E K Y Y E S S I W F Stop
L R I T E R G T E K G R K G S L E L P Met P L L L H P P I A A A Stop H R E I C A T F S L S L Q G G
I S G G K Q L C

3'5' Offener Leserahmen 3

Y N L K K Stop R L I Y Met L T V Y A Y P C T V Stop F A N Stop P Stop Stop L K T T G Stop I S G S
T Q L R D Stop A S L H I F C F S Q Met Q S R S L Y F L D P R G Y P S Met Y I H Q N L P K I T Stop
D D S D V G I I K D S R A A I I T V L H Q V Q L Stop S Met S G S L Q T P Y E K T P E N R G Q K S
T Met K A V S G F S F V S L K E A L K R V G K A V L N C Q C H S S S I P P Stop Q L H D T E K S V
Q H F L Y L Y K V V F Q E E S S Y

DANKSAGUNG

Prof. Dr. Rolf Horstmann danke ich für die Bereitstellung des interessanten Themas und der Räumlichkeiten sowie der freien Hand, die mir während der Erstellung meiner Arbeit gelassen wurde. Prof. Dr. Iris Bruchhaus gilt besonderer Dank für ihre engagierte Funktion als zweite Dissertationsgutachterin.

Prof. Dr. Egbert Tannich und Prof. Dr. Jürgen May danke ich für die Übernahme der Ko-Betreuerfunktion. Dr. med. Christian Timmann rechne ich seine stetige, kritische und intensive Betreuung hoch an. Besonderer Dank gilt auch Prof. Dr. Christian Meyer für seine Unterstützung bei der Ausarbeitung der Dissertation.

Der gesamten AG Horstmann danke ich für die angenehme und lustige Arbeitsatmosphäre sowie die beständige Hilfe in allen Lebenslagen. Herzlicher Dank gilt dabei Dr. Kathrin Schuldt und Dr. Thorsten Thye für das Einführen in die genetische Epidemiologie. Gerd Ruge danke ich für mein täglich Brot! Für schöne Mittagstische in der Sonne und verregnete aber spaßige Tage an der Laborbank danke ich Dr. Claudia Esser, Hanna Matthews und Elena Terhalle.

Anna Bachmann und Jenny Matthiesen danke ich für ihre Ideen und Hilfestellungen zu RNA-abhängigen Problematiken.

Besonderer Dank für den geteilten PhD-Vorsitz und die gemeinsamen Stunden am BNI und darüber hinaus gilt Arlett Heiber.

Für eine ausgeglichene Work-Life-Balance sorgten außerdem Ina Roesner und Christian Amon – danke! Maike Meinen gilt besonderer Dank für distanzlose Unterstützung. ♥

Herzlicher Dank gilt auch Felix Maliers und Christopher O'Donnell für ihr scharfes Korrekturauge. Mein Dank richtet sich auch an Gesine Müller und Dr. Katja Barth für ihr offenes Ohr.

Prof. Dr. James Kirkpatrick (Johannes Gutenberg Universität Mainz, Institut für Allgemeine Pathologie) danke ich für die Bereitstellung der HPMEC-Zelllinie, ebenso Prof. Dr. Jörg Ostermeyer und Dr. Matthias Betzold (Asklepiosklinik St. Georg, Herzchirurgie) für die humanen Venen.

Meiner Mutter danke ich für ihren bedingungslosen Rückhalt.