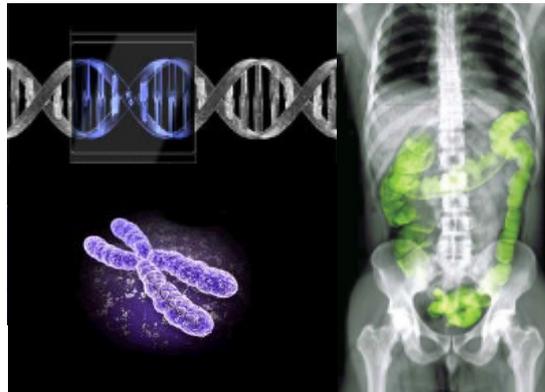

**Positionsklonierung eines Kandidatengens für
chronisch entzündliche Darmerkrankungen
auf dem humanen Chromosom 16**



Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität

zu Kiel

vorgelegt von

Dipl.-Biol. Andreas Rüther

**Kiel
2005**

Referent:

Korreferent:

Tag der mündlichen Prüfung:

Zum Druck genehmigt: Kiel, den

.....

Der Dekan



**Menschen mit einer neuen Idee gelten so lange als Spinner,
bis sich die Sache durchgesetzt hat.**

Mark Twain (30.11.1835 – 21.04.1910)

INHALTSVERZEICHNIS

1	EINLEITUNG	8
1.1	Chronisch entzündliche Darmerkrankungen	8
1.1.1	Charakterisierung des Krankheitsphänotyps und der Ätiopathogenese	8
1.1.2	Genetische Grundlagen der chronisch entzündlichen Darmerkrankungen	9
1.2	Prinzip der Positionsklonierung von Kandidatengen	11
1.3	Zielsetzung der Studie	12
2	METHODEN	13
2.1	Rekrutierung und Aufbau der Studienpopulation	13
2.2	DNA Proben	14
2.2.1	DNA Isolation aus EDTA-Vollblut	14
2.2.2	Fertigung der PCR-Platten für die TaqMan®-Genotypisierung	15
2.3	Genotypisierung	16
2.3.1	Prinzip der biallelischen Genotypisierung	16
2.3.2	Methode der biallelischen Genotypisierung	18
2.4	Mutationsdetektion und SNP-Verifikation	24
2.4.1	Prinzip der Sequenzierung	24
2.4.2	Methode der Sequenzierung nach Sanger	26
2.5	Chromosomale Lokalisierung von Kandidatenregionen	32
2.5.1	Prinzip der Fluoreszenz-in-situ-Hybridisierung	32
2.5.2	Methode der Fluoreszenz-in-situ-Hybridisierung	33
2.6	Interne Datenbank	34
2.6.1	Verwaltung der Patienten-, Phänotyp- und Genotypdaten	34
2.6.2	Qualitätskontrolle	34
2.7	Statistische Methoden	36
2.7.1	Kopplungsungleichgewicht	36
2.7.2	Transmission Disequilibrium Test	38
2.7.3	χ^2 - und Odds Ratio-Statistik der Fall-Kontroll-Studie	40

3	ERGEBNISSE	44
3.1	Feinkartierungsphase I	44
3.1.1	Definition der Kandidatenmarker	44
3.2	Feinkartierungsphase II	49
3.2.1	Erhöhung der Markerichte und Aktualisierung der physikalischen Karte	49
3.2.2	Physikalische Kartierung der Kandidatenmarker durch die Fluoreszenz-in-situ-Hybridisierung (FISH)	52
3.2.3	Abgrenzung und Definition der Kandidatenregionen KR 1, KR 2 und KR 3	53
3.3	Feinkartierungsphase III und Auflösung der Kandidatenregionen in Kandidatengenen	59
3.3.1	ZFP90 (Zinkfingerprotein 90)	60
3.3.2	ZNF19 und ZNF23 (Zinkfingerproteine 19 und 23)	63
3.3.3	GPR56 (G-Protein gekoppelter Rezeptor 56)	67
3.3.4	BRD7 (Bromodomänen Protein 7)	70
4	DISKUSSION	77
4.1	Interpretation der Kandidatengenstatistik	77
4.1.1	Zinkfingerproteine ZFP90, ZNF19 und ZNF23	77
4.1.2	G-Protein gekoppelter Rezeptor 56 (GPR56)	79
4.1.3	Bromodomänen Protein 7 (BRD7)	80
4.2	Potentielle Einflussgrößen in der Assoziationsstudie und ihre Relevanz für die vorliegende Studie	81
4.2.1	Populationsstratifikation	81
4.2.2	Probenumfang, statistische Aussagekraft und Risikoabschätzung	83
4.2.3	LD- und Haplotypstruktur	85
4.2.4	Replikation von Assoziationssignalen	86
4.2.5	Vergleich der familienbasierten und populationsbasierten Studie	88
4.3	Schlussfolgerung und Ausblick der Studie	90
5	ZUSAMMENFASSUNG	92
6	SUMMARY	93
7	LITERATUR	94

8 ABKÜRZUNGSVERZEICHNIS 108

9 ANHANG 111

9.1 Chemikalien und Geräte 111

9.2 Patientenfragebogen 113

1 Einleitung

1.1 Chronisch entzündliche Darmerkrankungen

1.1.1 Charakterisierung des Krankheitsphänotyps und der Ätiopathogenese

Die chronisch entzündlichen Darmerkrankungen (CED, MIM 601458), zu denen Morbus Crohn (MC, MIM 266600) und Colitis ulcerosa (CU, MIM 191390) als die beiden Hauptformen gezählt werden, sind der Gruppe der komplexen Erkrankungen zuzuordnen. Hierzu werden Krankheiten gezählt, denen eine polygenetische und umweltbedingte Ätiologie zu Grunde liegt. Kennzeichnend für den Phänotyp der CED ist eine massive und destruktive Veränderung des humanen Gastrointestinaltraktes, die das Resultat chronischer Entzündungsprozesse darstellt. Der Krankheitsverlauf ist durch eine chronische, in Schüben verlaufende Symptomatik charakterisiert, welche sich in dem Auftreten von Bauchschmerzen, Durchfällen, Fieber, Anämie, Arthritis und Körpergewichtsverlusten ausdrückt. Die phänotypische Charakterisierung der CED und die Unterscheidung der Subphänotypen Morbus Crohn und Colitis ulcerosa erfolgt über endoskopische, histologische, radiologische und klinische Kriterien. Hierbei ist der Phänotyp MC durch eine diskontinuierliche, segmentale Ausbreitung, die den gesamten Gastrointestinaltrakt betreffen kann und oftmals mit der Ausbildung von Stenosen, Fisteln und Abszessen einhergeht, gekennzeichnet. Demgegenüber erfolgt die dauerhafte Schädigung des Verdauungstraktes beim Phänotyp CU in einer kontinuierlichen Weise vom Rektum nach proximal und ist durch die Bildung von Pseudopolypen und Ulcera charakterisiert. Histologisch ist Morbus Crohn durch eine inflammatorische Infiltration der gesamten Darmwand gekennzeichnet. Colitis ulcerosa wird hingegen durch eine auf die Mukosa begrenzte Entzündung charakterisiert. Die modellhafte Ätiopathogenese der chronisch entzündlichen Darmerkrankungen ist wie folgt zu umschreiben: Umweltfaktoren in Form von infektiösen Agenzien, die aus einzelnen Bakterienspezies bishin zur gesamten physiologischen Darmflora bestehen, besonderen Nahrungsmittelbestandteilen (z.B. Konservierungsstoffe) und/oder Umwelttoxinen (z.B. Nikotingenuss) wirken als Antigene im Darmbereich. Auf der Basis genetischer, prädisponierender Faktoren rufen diese Antigene eine pathologische Immunreaktion in Form einer chronischen Aktivierung des mukosalen Immunsystems hervor. Somit sind die chronisch entzündlichen Darmerkrankungen als Folge einer gestörten Immuntoleranz vor dem Hintergrund genetischer und umweltbedingter Faktoren zu betrachten [Macdonald und Monteleone, 2005; Ott *et al.*, 2004; Rogler, 2004; Schreiber *et al.*, 2005]. Studien, die eine ansteigende Krankheitsinzidenz von 6-10 Neuerkrankungen pro 100.000 Einwohner/Jahr innerhalb der Bevölkerung Nordeuropas beschreiben, unterstreichen hierbei die Bedeutung

der Umweltfaktoren und weisen dem Lebensstil westlicher Industriegesellschaften eine deutliche Krankheitsrelevanz zu [Shivananda *et al.*, 1996]. Es wird geschätzt, dass die Lebenszeitprävalenz für die Entwicklung einer chronisch entzündlichen Darmerkrankung etwa 0,2% für Einwohner der westlichen Industrienationen beträgt [Binder, 2004; Probert *et al.*, 1996]; das Hauptmanifestationsalter liegt dabei in der frühen Adoleszenz, wobei ein nennenswerter Anteil der Patienten schon in der Kindheit erkrankt [Russell und Satsangi, 2004].

Der Phänotyp der chronischen Colitis kann mit einer bedeutenden Prävalenz auch in domestizierten Tierformen wie dem Hund (*Canis lupus f. familiaris*) oder der Katze (*Felis silvestris f. catus*) beobachtet werden [Cave, 2003; Feinstein und Olsson, 1992; Hostutler *et al.*, 2004], und wird dort insbesondere vor dem Hintergrund diätetischer Faktoren diskutiert [Guilford und Matz, 2003; Simpson, 1998]. Ein Beispiel für die Entwicklung einer CED in einem Wildtier stellt der in den Regenwäldern Nordkolumbiens beheimatete Lisztaffe (*Saguinus oedipus*) dar. Er entwickelt mit einer hohen Prävalenz eine idiopathische Colitis, die der Colitis ulcerosa ähnelt. Den Bakterien *Helicobacter spec.* und *Escherichia coli* wurde hierbei eine bedeutende Rolle in der Pathogenese zugesprochen [Mansfield *et al.*, 2001; Saunders *et al.*, 1999], was als weiterer Hinweis auf umweltbasierte Triggerfaktoren für die Ausbildung einer CED gewertet werden kann.

1.1.2 Genetische Grundlagen der chronisch entzündlichen Darmerkrankungen

Die maßgebliche Beteiligung genetischer Faktoren an der Pathogenese der CED konnte durch zahlreiche epidemiologische Studien bewiesen werden. Eine familiäre Häufung, die durch ein relatives Geschwisterrisiko (λ_s) im Bereich von 15-35 für den Phänotyp MC und 6-9 für den Phänotyp CU gekennzeichnet ist [Kuster *et al.*, 1989; Meucci *et al.*, 1992; Orholm *et al.*, 1991; Probert *et al.*, 1993], und die hohe Konkordanz von 50-60% zwischen monozygoten Zwillingen gegenüber 4% in dizygoten Zwillingen dokumentieren dieses eindrucksvoll [Jess *et al.*, 2005; Orholm *et al.*, 2000; Tysk *et al.*, 1988]. Auch ist der signifikante Unterschied der Krankheitsprävalenz zwischen verschiedenen ethnischen Gruppen innerhalb einer Bevölkerung - und somit ähnlichen Umweltbedingungen - ein weiterer Beleg dafür [Jayanthi *et al.*, 1992; Probert *et al.*, 1996]. Erste Studien ließen vermuten, dass dem Phänotyp MC ein rezessives und dem Phänotyp CU ein dominantes Krankheitsmodell unter Beteiligung eines jeweiligen Hauptrisikogens zu Grunde liegen [Orholm *et al.*, 1993]. Neuere Untersuchungen konnten jedoch zeigen, dass die chronisch entzündlichen Darmerkrankungen einem polygenetischen Vererbungsmodell folgen, und sowohl gemeinsame als auch exklusive Risikogene für die beiden Subphänotypen MC und CU

verantwortlich sind [Binder, 1998]. Eine monogene Vererbungsweise konnte durch umfangreiche Familienstudien ausgeschlossen werden, da sowohl eine unvollständige Penetranz als auch eindeutige Abweichung von klassischen - den Mendelschen Regeln folgenden - Vererbungsmodellen beobachtet wurden [Laharie *et al.*, 2001; Lander und Schork, 1994; Orholm *et al.*, 1993]. Gestützt wird diese Annahme durch den Nachweis multipler Kopplungsregionen, die in mikrosatellitenbasierten Kopplungsanalysen von erkrankten Geschwisterpaaren generiert wurden [Duerr *et al.*, 2000; Hampe *et al.*, 1999; Ma *et al.*, 1999; Satsangi *et al.*, 1996]. Insgesamt existieren derzeit neun etablierte CED-Kopplungsregionen (IBD1-IBD9), in denen potentielle Krankheitsgene unterschiedlicher Effektstärke vermutet werden [Hugot, 2004; Mathew und Lewis, 2004; Newman und Siminovitch, 2005]. Die Gesamtzahl aller krankheitsrelevanten Gene lässt sich jedoch nicht ausschließlich über diese abschätzen, da insbesondere Risikogene moderater Effektgröße durch Kopplungsanalysen nur bedingt erfasst werden. Die Entdeckung des CED-Krankheitsgen *DLG5* konnte dieses eindrucksvoll bestätigen [Stoll *et al.*, 2004].

Bisherige Studien konnten insgesamt zwei Krankheitsgene in Form von *NOD2* (*CARD15*) und *DLG5* sowie einen Risikohaplotyp, der die Gene *OCTN1* (*SLC22A4*) und *OCTN2* (*SLC22A5*) beinhaltet, identifizieren [Hampe, Cuthbert *et al.*, 2001; Hugot *et al.*, 2001; Ogura *et al.*, 2001; Peltekova *et al.*, 2004; Stoll *et al.*, 2004]. Gemeinsam gliedern sich diese Risikogene in ein funktionelles Krankheitsmodell der CED, welches als Schwächung der Integrität der Magendarmbarriere definiert werden kann [Schreiber *et al.*, 2005]. Während die funktionelle Relevanz der Gene *DLG5*, *OCTN1* und *OCTN2* hinsichtlich der strukturellen Stabilität des Darmepithels bzw. des Ionentransportes innerhalb des Epithels noch hypothetischen Charakter besitzt [McGovern und Ahmad, 2005], ist *NOD2* als bisher stärkstes Risikogen für den Phänotyp MC in seiner mechanistischen Funktionsweise besser erforscht [Chamaillard *et al.*, 2003; Eckmann und Karin, 2005; Inohara *et al.*, 2003]. So konnte gezeigt werden, dass drei Mutationen des *NOD2*-Gens (3020insC, G908R und R702W) zur strukturellen Veränderungen des C-terminalen Leucin-Restes des *NOD2*-Proteins führen. Hierdurch wird eine Reduzierung bzw. Störung der *NOD2*-basierten Bakterienerkennung innerhalb der Darmepithelzellen hervorgerufen, die eine chronische bakterielle Infektion der Colon-Epithelzellen zur Folge haben kann [Eckmann und Karin, 2005; Girardin *et al.*, 2003]. Trotz des starken Risikobeitrages der *NOD2*-Varianten, der sich in einem relativen Risiko (RR) von ca. 20 für die homozygote und einem RR von 2-4 für die heterozygote Trägerschaft des Risikoallels äußert [Mathew und Lewis, 2004], besitzen nur 30-40% der MC-Patienten Nordeuropas gegenüber 14% der gesunden Bevölkerung mindestens eine der drei Risikovarianten [Cuthbert *et al.*, 2002; Lesage *et al.*, 2002]. Hieraus lässt sich ableiten, dass für die Ausbildung des Krankheitsphänotyps der Besitz einer *NOD2*-

Mutation weder zwingend noch alleinig notwendig ist, und damit der Ansatz eines polygenetischen Krankheitsmodells für die CED untermauert wird.

1.2 Prinzip der Positionsklonierung von Kandidatengen

Das Ziel einer Kandidatengenstudie für eine komplexe Erkrankung ist die Entdeckung von krankheitsrelevanten Risikogenen. Im Mittelpunkt steht dabei die Erforschung der genetischen Grundlagen der Krankheitsprädisposition und der Ätiopathogenese, um auf diese Weise eine Optimierung der Krankheitsdiagnostik, -prävention und -therapie zu erlangen [Palmer und Cardon, 2005]. Gekennzeichnet ist der Aufbau einer solchen Studie durch das methodische Grundkonzept der Positionsklonierung [Botstein und Risch, 2003; Wild und Rioux, 2004]. Dieses lässt sich in sechs Etappen unterteilen, von denen die erste aus der Etablierung der Krankheitsheritabilität besteht. Konnte der Nachweis auf einen genetischen Krankheits hintergrund durch epidemiologische Untersuchungen (z.B. familiäre Krankheitshäufung) erbracht werden, erfolgt in der zweiten Phase die Identifizierung chromosomaler Kopplungsregionen durch familienbasierte Kopplungsstudien. Hierbei werden in der Regel hochpolymorphe Mikrosatelliten verwendet, die auf der Basis eines familiären Vererbungsmusters als Marker für eine mit dem Krankheitsphänotyp gekoppelte Genregion dienen. In der nächsten Phase der Positionsklonierung gilt es mit Hilfe familien- oder populationsbasierter Assoziationsstudien die Krankheitsgenregion über eine Feinkartierung einzugrenzen. Dabei kommt es zum Einsatz biallelischer Assoziationsmarker in Form von Einzelbasenmutationen (SNPs). Die Grundlage der indirekten Assoziationsstudie bildet hierbei eine allelische Kopplung des Assoziationsmarkers mit der Risikovariante des Krankheitsgens. In der populationsbasierten Studie (Fall-Kontroll-Studie) wird ein SNP (*single nucleotid polymorphism*) als mit der Krankheit assoziiert betrachtet, wenn die Häufigkeit oder Verteilung der SNP-Allele in einer Patientengruppe signifikant von der in einer gesunden Kontrollgruppe abweicht. Dagegen wird in der familienbasierten Assoziationsstudie die Vererbung der Markerallele von den Eltern auf die erkrankten Kinder betrachtet, und eine Krankheitsassoziation als eine gerichtete Über- oder Untertransmission eines Markerallels definiert. Konnten auf diese Weise einzelne assoziierte Genloci isoliert werden, gilt es in der nächsten Etappe die funktionellen Sequenzvariationen des assoziierten Gens durch eine Sequenzierung der entsprechenden Genregionen zu identifizieren. Wurden auf diese Weise krankheitsrelevante Genvariationen entdeckt, wird in der fünften Phase der Kandidatengenstudie versucht, die Validität der ermittelten Risikomutationen durch eine Replikation der statistischen Krankheitsassoziation in unabhängigen Populationen zu bestätigen. Verläuft auch diese Phase erfolgreich, steht am Ende der Positionsklonierung die

Identifikation eines Krankheitsgens und seiner assoziierten Risikoallele, der sich funktionelle Studien anschließen.

1.3 Zielsetzung der Studie

Das der *NOD2*-Entdeckung zu Grunde liegende Kopplungssignal auf Chromosom 16 wird durch die Varianten im *NOD2*-Gen nicht vollständig erklärt. Unabhängige Kopplungsstudien konnten durch eine Stratifizierung der Studienfamilien gegen die *NOD2*-Riskovarianten den Hinweis erbringen, dass auf dem q-Arm des Chromosom 16 weitere Risikogene lokalisiert sein könnten [Annese *et al.*, 2003; Hampe *et al.*, 2002; van Heel *et al.*, 2003]. Das Ziel dieser Studie ist es, über den methodischen Ansatz der Positionsklonierung ein neues Krankheitsgen für die chronisch entzündlichen Darmerkrankungen auf dem Chromosom 16q zu identifizieren. Im einzelnen ergaben sich hieraus folgende Aufgaben- und Fragestellungen:

- I. Kann durch eine SNP-Feinkartierung der *NOD2*-unabhängigen Kopplungsregion auf Chromosom 16q ein neues Krankheitsgen auf der Grundlage eines populations- und familienbasierten Studiendesigns identifiziert werden?
- II. Können die krankheitsassoziierten Marker in einer unabhängigen Studienpopulation repliziert werden?
- III. Ist es möglich, den identifizierten Risikoallelen eine funktionelle Bedeutung für die CED zuzuordnen?

2 Methoden

Eine Auflistung der verwendeten Gerätschaften und Chemikalien befindet sich im Anhang (siehe 9.1).

2.1 Rekrutierung und Aufbau der Studienpopulation

Die Studienpopulation in Form von DNA (*desoxiribonucleic acid*)-Proben gliederte sich in vier Gruppen. Diese bestanden zum einen aus Familien, die mindestens zwei erkrankte Kinder und die gesunden Eltern umfasste (Mutlplex-Familien). Die zweite Gruppe bildeten die sogenannten Eltern-Kind-Trios nach Falk und Rubinstein [Falk und Rubinstein, 1987], zu denen jeweils ein erkranktes Kind und die beiden gesunden Elternteile zählten (Monoplex-Familien). Einzelne, unverwandte Patienten bildeten die dritte Probengruppe, und die vierte Gruppe stellten die gesunden, unverwandten Kontrollindividuen dar. In dem zweigeteilten Studiendesign, welches sowohl den Ansatz einer familien- als auch populationsbasierten Assoziationsstudie beinhaltet, wurden folgende Populationen voneinander abgegrenzt:

Die initiale Feinkartierungspopulation setzte sich zum einen aus 181 Familien und 564 Trios und zum anderen aus 538 unverwandten, gesunden Blutspender zusammen. Das Verhältnis von Morbus Crohn- zu Colitis ulcerosa-Patienten belief sich dabei auf etwa 2:1. Insgesamt umfasste die erste Studienpopulation 2936 Individuen deutscher Abstammung. Die zweite Studiengruppe bestand als unabhängige Replikationspopulation aus 726 Trios und 534 Einzelpatienten mit einem Phänotypverhältnis von 1:1 für Morbus Crohn zu Colitis ulcerosa. Desweiteren beinhaltet diese Gruppe 1040 populationsbasierte Kontrollindividuen, die durch eine zufällige Auswahl innerhalb der norddeutschen Bevölkerung rekrutiert wurden und Bestandteil einer umfangreichen Bevölkerungsstudie innerhalb des ‚popgen‘-Projektes der Uniklinik Kiel sind [Lamina *et al.*, 2005]. In der Gesamtheit umfasste die Replikationspopulation damit 3752 deutschstämmige Individuen.

Tab. 2.1: Zusammensetzung der Studienpopulation

	CED-Patienten	MC-Patienten	CU-Patienten	nichterkrankte Familienmitglieder	Anzahl der Familien	unverwandte, gesunde Kontrollen
Einzelpatienten	534	176	358	-	-	
Trios	1290	825	465	2580	1290	1578
Familien	432	287	145	274	181	
total	2256	1288	968	2854	1471	1578

Die phänotypische Unterscheidung zwischen Patienten und gesunden Kontrollen erfolgte auf der Basis einheitlicher Ein- bzw. Ausschlusskriterien. Der Patientenstatus einer Studienperson basierte dabei auf einer eindeutigen Krankheitsdiagnose, die auf einem endoskopischen, radiologischen, histologischen sowie klinischen Befund beruhte. Zusätzlich wurden umfangreiche, fragebogenbasierte Daten zum Gesamtkrankheitsbild erhoben (siehe 9.2). Die Patienten und ihre Familienmitglieder wurden an der I. Medizinischen Klinik des Universitätsklinikum Schleswig-Holstein (Campus Kiel), der Charité Universitätsklinik (Berlin) und über eine Patientenvereinigung (Deutsche Morbus Crohn/Colitis ulcerosa Vereinigung) rekrutiert. Die als Kontrollen verwendeten Blutspender wurden über das Institut für Transfusionsmedizin im Universitätsklinikum Schleswig-Holstein (Campus Kiel) gesammelt und sowohl über diagnostische Blutparameter als auch durch eine ausführliche Befragung als gesund eingestuft. Eine Gesamtübersicht der Studienpopulation findet sich in der Tabelle 2.1 wieder.

2.2 DNA Proben

2.2.1 DNA Isolation aus EDTA-Vollblut

Zur Extraktion genomischer DNA aus Blutproben wurde das Präparationskit „Invisorb® Blood Giga Kit“ (Invitek, Berlin, Deutschland) unter veränderten, optimierten Protokollbedingungen verwendet. Im ersten Reaktionsschritt, der Lyse der Erythrozyten, wurden 10 ml Ethylendiamintetraessigsäure (EDTA)-Vollblut mit 30 ml Puffer I versetzt und bei Raumtemperatur 10 Minuten inkubiert. Anschließend wurde die Probe 3 Minuten bei 3000 U/min zentrifugiert, und der Überstand bestehend aus den Erythrozytenfragmenten wurde verworfen. Dieser Schritt wurde mit 20 ml Puffer I solange wiederholt bis das verbleibende Zellpellet, welches die DNA-haltigen Leukozyten enthielt, farblos und damit frei vom Häm der Erythrozyten war. Eine Probenkontamination mit Häm hätte u.a. eine Hemmung der PCR (*polymerase chain reaction*)-Reaktion zur Folge gehabt [Heath *et al.*, 1999]. Das gereinigte Zellpellet wurde im Anschluss mit 3 ml Puffer II und 50 µl Proteinase K resuspendiert, und es erfolgte eine zweistündige Inkubation bei 60 °C in einem Wasserbad unter kontinuierlichem Schütteln. Dieser Abschnitt diente der Lyse der Leukozyten und der Trennung der DNA vom Zellkern. Um die gelösten Proteine zu Fällern und damit von der DNA zu trennen, wurden 1,8 ml Puffer III nach der Inkubation hinzugefügt, und der Probenmix wurde nach kurzem, energischem Schütteln für 5 Minuten auf Eis gelagert. Danach folgte ein 15 minütiger Zentrifugationsschritt bei 5000 U/min, und der klare Überstand wurde in ein 15 ml Zentrifugationsröhrchen überführt. Die Präzipitation der DNA erfolgte nach dem Prinzip der

Alkohol-Fällung [Bearden, 1974; Shapiro, 1981; Wilcockson, 1975], indem 10 ml Ethanol (96%) dem Überstand zugefügt und das DNA-Alkohol-Gemisch vorsichtig geschwenkt wurde. Im Anschluss wurde die präzipitierte DNA mit einer Pipettenspitze gesammelt und zur Entfernung der Salzreste in ein 2 ml Eppendorf-Röhrchen, das 1 ml Ethanol (70%) enthielt, überführt. Das DNA-Pellet wurde hierauf mit Ethanol gewaschen und nach einem Vortex-Zwischen-Schritt bei 13.000 U/min für 2 Minuten abzentrifugiert. Es folgte das Abpipettieren des Ethanol-Überstandes, und das DNA-Pellet wurde bei Raumtemperatur für 10 Minuten getrocknet, um die flüchtigen Ethanolreste zu entfernen. Das gereinigte DNA-Pellet wurde zum Abschluss der DNA-Isolation in 500 µl Tris-EDTA-Puffer (TE-Puffer, einfache Arbeitsverdünnung) resuspendiert und bei -20 °C gelagert. Eine kurzfristige Lagerung bei +4°C erfolgte bei der direkten Weiterverarbeitung der genomischen DNA ohne dass es zu Qualitätseinbußen kam.

2.2.2 Fertigung der PCR-Platten für die TaqMan[®]-Genotypisierung

Die isolierten und aufgereinigten DNA-Proben wurden für die Genotypisierungsexperimente nach einem definierten Verteilungsschema, dem sogenannten Platten-Layout, auf 96-Kammer-Platten (*96-well plates*) (Costar Corning Inc., Cambridge, MA, USA) angeordnet. Dabei wurden jeweils 3 Plattenkammern auf den Positionen D3, D10 und H12 nicht mit DNA befüllt, um im Experiment als Leerwertkontrollen (Negativ-Kontrollen) gegenüber Kontamination durch Fremd-DNA zu fungieren. Die Kammer C3 beinhaltete auf jeder Platte die DNA des selben Individuums einer CEPH (Centre d'Etude du Polymorphisme Humain)-Zelllinie, die als Kontroll-DNA (Positiv-Kontrolle) für die Qualitätskontrolle des TaqMan[®]-Assays und anderer Einflussgrößen des Typisierungsexperimentes diente. Jeweils vier 96-Kammer-Platten wurden zu einer 384-Kammer-Platte zusammengestellt, wobei das 384er-Plattenlayout eine Mischform der vier 96er-Plattenlayouts darstellte. Proben von Kontroll-Individuen und Patienten befanden sich auf separaten PCR-Platten, wobei DNA-Proben von Familienmitgliedern auf der selben Platte verteilt wurden. Jede 96- und 384-Kammer-Platte erhielt eine individuelle Barcode-Nummer, sodass jedes Experiment durch die Plattennummer, dem Platten-Layout, der TaqMan[®]-Assay-Nummer und den Probennummern eindeutig charakterisiert war.

Nach der Konzentrationsbestimmung der isolierten DNA-Proben durch ein PicoGreen[®]-Messprotokoll [Ahn *et al.*, 1996; Rengarajan *et al.*, 2002] wurde ein Aliquot der Proben in 96- oder 384-Kammer-Lagerplatten (*96-/384-deep-well plates*) (ABgene, Epsom, UK) überführt und durch zwei Verdünnungsschritte auf eine Arbeitskonzentration von 0,5

ng/µl eingestellt. Das Pipettieren der Platten erfolgte dabei mit Hilfe eines automatisierten Pipettierroboters (Genesis RSP 150 Multipipettierroboter, Firma TECAN Deutschland GmbH, Crailsheim, Deutschland). Jeweils 5 µl der Arbeitsverdünnung wurden mit einem 96- oder 384-Kanal-Hydra-Mikrodispenser (Robbins Scientific, Sunnyvale, CA, USA) in die 96- oder 384-Kammer-PCR-Platten pipettiert, sodass in dem Genotypisierungsexperiment jeweils 2,5 ng DNA eingesetzt wurden. Die Arbeitsplatten wurden für ca. 2 Stunden bei 60 °C getrocknet, mit Folie durch einen ‚Plattenversiegler‘ (ABgene, Hamburg, Deutschland) verschweißt und bei Raumtemperatur gelagert.

2.3 Genotypisierung

2.3.1 Prinzip der biallelischen Genotypisierung

Für die Bestimmung der Genotypen von Einzelbasenpolymorphismen (SNPs) wurde die Methode der biallelischen Genotypisierung auf der Grundlage des 5'-Nuklease-Assays [De la Vega *et al.*, 2005; Holland *et al.*, 1991; Livak *et al.*, 1995] verwendet. Hierbei handelt es sich um eine Messung von allelspezifischer Fluoreszenz, welche eine eindeutige Zuordnung des jeweiligen Genotyps ermöglicht. Die Grundlage der Methode ist die PCR nach K. B. Mullis [Mullis *et al.*, 1986; Mullis und Faloona, 1987]. Hierbei handelt es sich um eine Methode der sequenzspezifischen DNA-Amplifikation, bei der das Enzym Taq-Polymerase eine zentrale Rolle spielt und dessen 5'-Nuklease-Aktivität die Basis des 5'-Nuklease-Assays bildet.

Im Allgemeinen besteht eine PCR-Reaktion aus drei Reaktionsabschnitten, die sich in Zyklen wiederholen bis eine ausreichende Menge an DNA synthetisiert ist (Abb. 2.1). Der erste Abschnitt ist die Denaturierung der vorgelegten, doppelsträngigen DNA (*template* DNA) in einzelsträngige DNA. Danach erfolgt im zweiten Abschnitt eine Anlagerung (*annealing*) von sequenzspezifischen Oligonukleotiden (Primer), welche die Zielsequenz flankieren, an die Einzelstrang-DNA. Im dritten Abschnitt werden die Primer mit Hilfe der Polymerase und Einzelnukleotiden (dNTPs) am 3'-Ende verlängert (*elongation*), wobei die Einzelstrang-DNA als Matrize dient, sodass am Ende des ersten Zyklus die Anzahl der DNA-Abschnitte der Zielsequenz verdoppelt wurde. Durch eine Abfolge mehrerer Zyklen wird auf diese Weise eine exponentielle Anreicherung von spezifischen DNA-Fragmenten erreicht.

Der 5'-Nuklease-Assay kombiniert die PCR-Reaktion, deren Zielsequenz den zu genotypisierenden SNP umschließt, mit markerspezifischen Oligonukleotiden (SNP-Sonden, *probes*), die für das jeweilige Allel des biallelischen SNPs charakteristisch sind. Die doppelt markierten Fluoreszenz-Sonden besitzen am 5'-Ende einen fluoreszierenden Reporterfarbstoff und am 3'-Ende einen *Quencher*-Farbstoff, der die Fluoreszenzfähigkeit

des Reporterfarbstoffes im sondengebundenen Zustand unterdrückt. Ein SNP-spezifisches Sondenpaar unterscheidet sich dabei zum einen in der Nukleotidsequenz an der Stelle der Einzelpunktmutation und zum anderen in dem Reporter-Farbstoff, der jeweils eine individuelle Fluoreszenz bei unterschiedlicher Wellenlänge besitzt. Dadurch erlangen die SNP-Sonden ihre Allelspezifität und ermöglichen eine Diskriminierung zwischen den beiden Markerallelen.

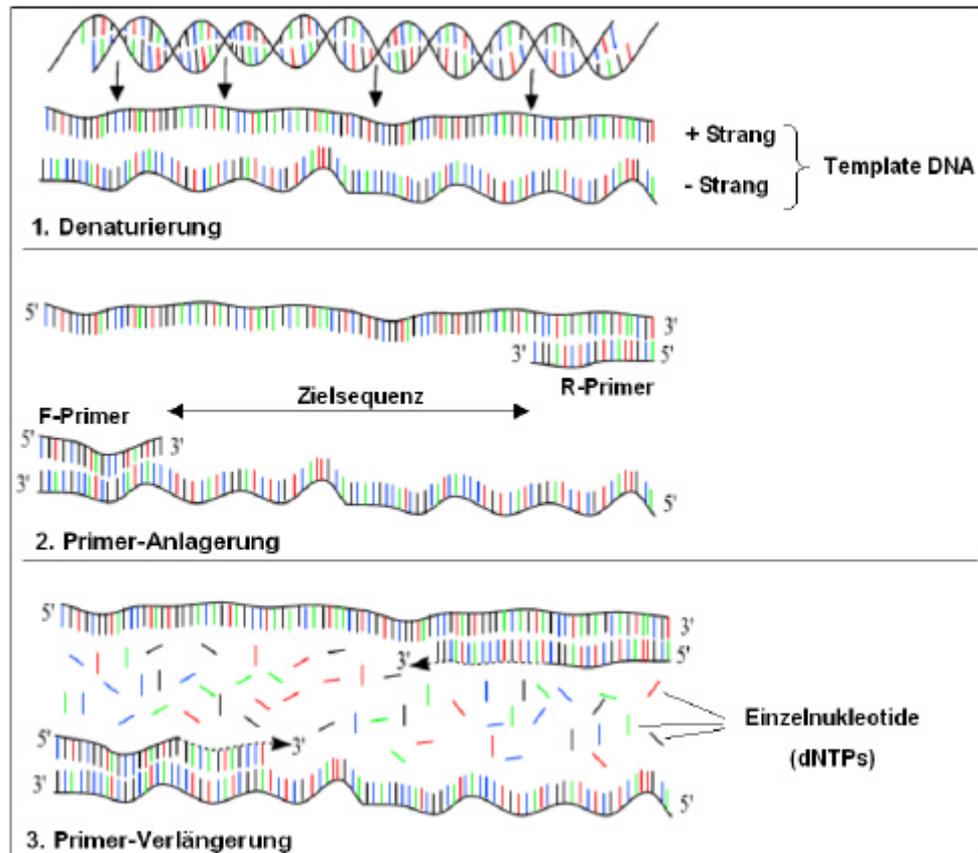


Abb. 2.1: Prinzip der Polymerase-Ketten-Reaktion (PCR).

Während der PCR hybridisiert die SNP-komplementäre Sonde an den reversen DNA Einzelstrang und wird durch die 5'-Nukleaseaktivität der Taq-Polymerase wieder abgespalten. Dabei kommt es zur Ablösung des Reporterfarbstoffes von der Sonde und die Reporterfluoreszenz wird aktiviert (Abb. 2.2). Analog zur exponentiellen Zunahme der DNA-Fragmente nimmt auch die Stärke der Fluoreszenz während der PCR-Reaktion zu. Dabei kommt es entsprechend des Genotyps der *template*-DNA zu einer Zunahme eines einzelnen Fluoreszenzsignals im Falle der Homozygotie oder zweier Fluoreszenzsignale bei vorliegender Heterozygotie für den zu untersuchenden SNP. Die Messung der generierten Fluoreszenz erfolgt nach Beendigung der PCR-Reaktion in einer sogenannten

Endpunktmessung mittels der Laser-Scan-Technologie. Hierbei wird mit Hilfe von Laserlicht der ungebundene Reporterfarbstoff energetisch angeregt und die resultierende Fluoreszenzemission wird spektrografisch erfasst und über einen CCD (*charge coupled device*)-Chip elektronisch gespeichert. Da sich die Fluoreszenz der Reporterfarbstoffe in der Wellenlänge unterscheidet, ist somit eine eindeutige, indirekte Zuordnung des jeweiligen Genotyps möglich.

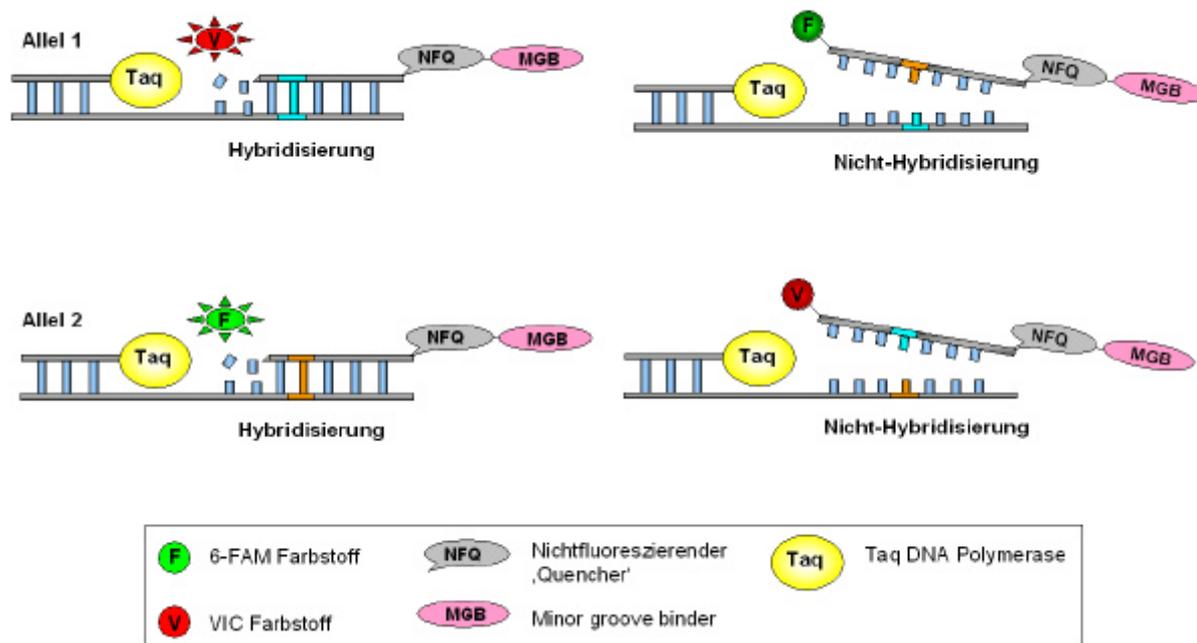


Abb. 2.2: Prinzip des 5'-Nuklease-Assays.

2.3.2 Methode der biallelischen Genotypisierung

Die Genotypisierung der untersuchten Gruppen erfolgte sowohl auf 96-Kammer-PCR-Platten als auch auf 384-Kammer-PCR-Platten. Hierzu wurde zunächst der Reaktionsmix – im weiteren als Assay bezeichnet – entsprechend des Probenumfangs zusammenpipettiert. Ein Assay ist SNP-spezifisch und besteht zum einen aus dem Mastermix, der sich aus der Taq-Polymerase, den dNTPs (Desoxynukleotide), dem Reaktionspuffer und dem Magnesiumchlorid ($MgCl_2$) als Cofaktor der Polymerase zusammensetzt, und zum anderen aus dem eigentlichen SNP-Assay, der die SNP-spezifischen Sonden (TaqMan[®] probes) und die SNP-Sequenz flankierenden Oligonukleotiden umfasst.

Bei der Herstellung der SNP-Assays kamen drei verschiedene Arten des Designs, der Synthese und der funktionellen Überprüfung zur Anwendung. Die Mehrheit der Assays wurde als gebrauchsfertiger TaqMan[®]MGB-Assay bezogen und verwendet. Gegenüber der konventionellen SNP-Sonde (siehe 2.3.1) eines klassischen TaqMan[®]-Assays (Eurogentec S.A., Seraing, Belgien) besitzt die MGB-Assay-Sonde am 3'-Ende des Quencher-Farbstoffes noch ein sogenanntes *minor groove binder* Molekül (MGB, *N*-Methylpyrrolecarboxamide). Dieses lagert sich mit seiner sichelförmigen Struktur bei der Hybridisierung der Sonde an die kleine Furche (*minor groove*) der Hybriddoppelhelix an und erhöht somit die Bindungsstärke zwischen Sonde und DNA [Afonina *et al.*, 1997; Kutyavin *et al.*, 1997]. Das hat zur Folge, dass Sonden mit kürzerer Nukleotidsequenz bei gleicher Schmelztemperatur gegenüber konventionellen Sonden verwendet werden können, was sich in einer erhöhten Spezifität der Sonde niederschlägt. Desweiteren ersetzt ein nicht fluoreszierender Quencher den Quencher-Farbstoff „TAMRA[™]“ (6-Carboxytetramethylrhodamin) der konventionellen Sonde, was eine Reduzierung der Hintergrundfluoreszenz zur Folge hat und die spektrale Diskriminierung erhöht. TaqMan[®]MGB-Assays wurden in drei Varianten verwendet, die sich in der Art der Assay-Herstellung unterscheiden. Es kam zum einen der Assay-on-Demand (AoD, Applied Biosystems Inc., Foster City, CA, USA) und zum anderen der Assay-by-Design (AbD, Applied Biosystems Inc., Foster City, CA, USA) zum Einsatz. Wichtigstes Unterscheidungsmerkmal zwischen diesen ist die Art des Assay-Designs, d.h. auf welche Weise die Primer- und Sondensequenzen generiert werden. Ein AoD ist beim Hersteller auf der Basis öffentlicher Datenbanksequenzen berechnet, synthetisiert und auf Funktionsfähigkeit in jeweils 45 Individuen verschiedener ethnischer Populationen (kaukasisch, chinesisch, japanisch, afrikanisch) überprüft worden, und ist kurzfristig, vorabproduziert verfügbar. Da während des Verlaufs der Studie nur 160.000 AoD versus ca. 3 Millionen bekannter SNPs verfügbar waren, wurde vorwiegend der AbD verwendet. Bei diesem erfolgte die Berechnung und Zusammenstellung des Assays basierend auf der vom Kunden gelieferten DNA-Sequenz, welche den zu untersuchenden SNP als auch benachbarte SNPs - falls bekannt - enthält. Die verwendeten Sequenzen stammten dabei zum einen aus öffentlichen Datenbanken wie dem *National Center for Biotechnology Information* (NCBI) oder dem *The Ensembl Project* (European Bioinformatics Institute, Wellcome Trust Sanger Institute) und zum anderen aus sequenzieller Mutationsdetektion (siehe 2.4). Im Anschluss erfolgte die Assay-Synthese und Funktionsüberprüfung auf Seiten des Herstellers, sodass beide TaqMan[®]MGB-Assay Formen gebrauchsfertig, ohne weitere Optimierungsschritte eingesetzt werden konnten. Als Reporter-Farbstoff wurden beim TaqMan[®]MGB-Assay „FAM[™]“ (6-Carboxyfluoreszein) und „VIC[®]“ (Markenname der Firma Applied Biosystems Inc.) eingesetzt.

Für einen kleinen Teil der untersuchten SNPs wurde die dritte Assay-Form, der sogenannte selbstdesignte (manuelle) TaqMan[®]MGB-Assay (Applied Biosystems Inc., Foster City, CA, USA) verwendet. Da dieser Assay-Typ vor der Genotypisierungsreaktion einer Optimierung bedurfte und nur im 96-Kammer-Maßstab verwendet werden konnte, kam er ausschließlich bei Markern zum Einsatz, die als AoD oder AbD nicht verfügbar bzw. synthetisierbar waren, für die Studie eine hohe Relevanz besaßen und nicht durch benachbarte, gekoppelte SNPs ersetzt werden konnten. Für die Konzeption des Assays, die im Wesentlichen das Design der Oligonukleotide umfasst, wurde die Software Primer Express 2.0 (Applied Biosystems Inc., Foster City, CA, USA) verwendet. Die in der Programmbeschreibung empfohlenen Richtlinien für das Assay-Design wurden übernommen [Dieffenbach *et al.*, 1993; Lowe *et al.*, 1990] und durch nachfolgend beschriebene Designvorgaben seitens des Assay-Herstellers (Applied Biosystems Inc.) ergänzt. Die Schmelztemperatur (T_m) der Primer wurde zwischen 58°C und 60°C gewählt, wobei der Temperaturunterschied zwischen Vorwärts (*forward*)- und Rückwärts (*reverse*)-Primer (F- und R-Primer) weniger als 2°C betrug. Der Gehalt an Guanin- und Cytosin-Nukleotiden (GC-Gehalt) wurde auf 30-80% festgelegt und die Primerlänge variierte zwischen 15 und 35 Nukleotiden. Desweiteren bestanden die letzten 5 Nukleotide des 3'-Endes aus maximal 2 GC-Basen und die letzte Base war kein Thymin. Die MGB-Sonden besaßen eine um 7°C höhere Schmelztemperatur als das Primerpaar und die Temperaturdifferenz zwischen den Sonden war niedriger als 1°C. Der GC-Gehalt lag in dem Intervall 20% bis 80% und die Sondenlänge betrug 13 bis 20 Basen. Die Nukleotidsequenz wurde aus Gründen der Bindungsspezifität sowohl bei den Primern als auch bei den Sonden möglichst variabel gewählt; insbesondere bei den Sonden wurde eine direkte Abfolge von mehr als drei Guanin-Basen vermieden. Auch durfte das 5'-Ende der Sonde nicht mit einer Guanin-Base beginnen und der Guaningehalt sollte den Cytosingehalt nicht überschreiten. Als wichtiger Punkt galt es die Lage des SNP in der Sondensequenz zu berücksichtigen; wobei eine Position als optimal zwischen Sondenmitte und dem 3'-Ende definiert wurde; der Mindestabstand zum 3'-Ende musste allerdings 2 Nukleotide betragen. Das Amplifikat sollte sich in einer Größenordnung von 50 bis 150 Basenpaaren (Bp) befinden, wobei das 3'-Ende der Primer so nah als möglich, jedoch ohne Überlappung, neben die Sonde positioniert wurde. Während des Primerdesigns wurde eine mögliche Primerdimerisierung hinsichtlich Bindungsstärke, Überlappungslänge und Fehlamplifikation überprüft. Weiterhin wurden sowohl die Primer als auch Sonden auf die potentielle Ausbildung von Sekundärstrukturen in Form von Schleifen- oder *hairpin*-Formationen, welche eine herabgesetzte Verfügbarkeit während der PCR-Reaktion zur Folge hätten, überprüft. Als Berechnungsgrundlage der jeweiligen Schmelztemperaturen der Oligonukleotide wurde eine Konzentration von 50 nM und eine Salzkonzentration von 50 mM verwendet.

Die Assay-Optimierung erfolgte hauptsächlich durch Veränderung des PCR-Protokolls (Tab. 2.2). Vor der Verwendung eines selbstdesignten TaqMan[®]MGB-Assays wurde jeder Assay auf je 47 DNA-Proben bei verschiedenen Annealing-/Elongation-Temperaturen (AE-Temperatur) getestet. In der ersten Optimierungsrunde wurde dabei eine AE-Temperatur von 60°C bzw. 62°C gewählt. Falls ein Assay nicht mindestens bei einer der beiden Temperaturen funktionell überzogen konnte, d.h. in der Endpunktmessung eine gute Diskriminierung der Genotypen möglich war, so wurde in einer zweiten Runde die AE-Temperatur auf 58°C bzw. 64°C erniedrigt bzw. erhöht. Desweiteren kam es bei suboptimalen Ergebnissen der zweiten Optimierungsläufe zu einer Veränderung der eingesetzten Primer- und/oder MgCl₂-Konzentrationen, und falls es sich um Amplifikate großer Länge handelte, wurde auch das AE-Zeitfenster vergrößert.

Grundsätzlich wurde bei allen drei verwendeten Assay-Typen ein einheitliches PCR-Protokoll verwendet (Tab. 2.2), das nur im Falle der selbstdesignten Assays ggf. abgeändert worden ist. In einem klassischen PCR-Protokoll wird zwischen der Anlagerungstemperatur und der Strangverlängerungstemperatur unterschieden, wobei erstgenannte im Bereich der Primerschmelztemperatur und letztere im Temperaturoptimum der Taq-Polymerase bei 72°C liegt. Hierbei wird dann von einem 3-Schritt-Protokoll gesprochen, da zwischen Denaturierung, Primeranlagerung und Strangverlängerung differenziert wird. In der vorliegenden Studie wurde hingegen ein 2-Schritt-Protokoll verwendet, das eine gemeinsame AE-Temperatur von 60°C beinhaltet; bei selbstdesignten Assays lag diese auf Grund der Optimierung zwischen 58°C und 64°C. Die Vorteile des 2-Schritt-Protokolles liegen dabei in einer verkürzten Gesamtzeit der PCR-Reaktion bei unveränderter Genotypisierungseffizienz des Assays.

Tab. 2.2: PCR-Bedingungen für den TaqMan^aMGB-Assay. (* Die AE-Temperatur wich zum Teil bei manuellen Assays auf Grund der Optimierung von der Standardtemperatur ab)

Temperatur	Zeit	Prozess	Funktion
95 °C	10 min	konstant	Polymerase Aktivierung
95 °C	15 s	45 Zyklen	DNA Denaturierung
60 °C*	1 min		Anlagerung und Verlängerung
4 °C	∞	konstant	

Tab. 2.3: Zusammensetzung der Reaktionslösung für die biallelische Genotypisierung.

Ansatz für 5ml Einzelreaktion	Assay-by-Design	Assay-on-Demand	Manueller Assay
	Volumen [μ l]	Volumen [μ l]	Volumen [μ l]
TaqMan[®] PCR MasterMix	2,50	2,50	2,50
FAM[®] Sonde [10μM]	-	-	0,10
VIC[®] Sonde [10μM]	-	-	0,10
F-Primer [20μM]	-	-	0,225
R-Primer [20μM]	-	-	0,225
gebrauchsfert. SNP-Assay-Mix	0,0625	0,25	-
Wasser, PCR-Qualität	2,4375	2,25	1,85
Gesamtvolumen	5,00	5,00	5,00

Im Anschluss an die PCR-TaqMan[®]-Reaktion wurde die Fluoreszenz jeder Probe mittels einer Fluoreszenzmesseinheit ermittelt. Abhängig von dem PCR-Plattentyp kamen bei der Verwendung des 96-Kammer-Formates das Lesegerät „ABI Prism[®] 7700 Sequence Detection System“ oder im 384er-Maßstab der Fluoreszenzleser „ABI Prism[®] 7900HT Sequence Detection System“ (beide Geräte: Applied Biosystems Inc., Foster City, CA, USA) zum Einsatz. Die Meßmethode ist hierbei bei beiden Systemen die gleiche (siehe 2.3.1) und die Messung erfolgte nach einmaliger Kalibrierung und protokoll- bzw. farbstoffspezifischer Konfiguration vollautomatisch für alle Genotypisierungsexperimente. Die ermittelten Fluoreszenzwerte der DNA-Proben einer PCR-Platte wurden nach der Messung im zweidimensionalen Koordinatensystem aufgetragen (Abb. 2.3), wobei die Fluoreszenz des Farbstoffes „VIC[®]“ dem Allel1 (X-Achse) und „FAM[™]“ dem Allel2 (Y-Achse) zugeordnet worden ist. Bei einer erfolgreich durchgeführten TaqMan[®]-Reaktion fielen die Messwerte gemäß der drei potentiellen Genotypen in drei Wolken (Cluster) zusammen, da die Signalstärke der Farbstoffe proportional zu der Anzahl gebundener Sonden und damit ein indirektes Maß für die Allele der jeweiligen DNA-Probe ist. DNA-freie Kammern der PCR-Platte (Abb. 2.4: Kammer O12, O24, P12 und P24) dienten der Qualitätskontrolle und deren Fluoreszenzwerte bildeten den vierten Cluster in der Nähe des Ursprunges des Koordinatensystems. Auf diese Weise wurde gegen Kontamination durch Fremd-DNA oder eine zu hohe Hintergrundfluoreszenz des Assays getestet. Die eigentliche Zuordnung der Genotypen zu den Clustern erfolgte manuell in der Software des Messgerätes, wobei Messwerte, die einem Cluster nicht eindeutig zugeordnet werden konnten, als Fehltypisierung aus der weiterführenden Analyse herausgenommen wurden. Der Abschluss der Genotypisierung bildete der Datenimport der generierten Genotypen in eine Datenbank, welche die Basis für die späteren Auswertungen und statistischen Analysen bildete.

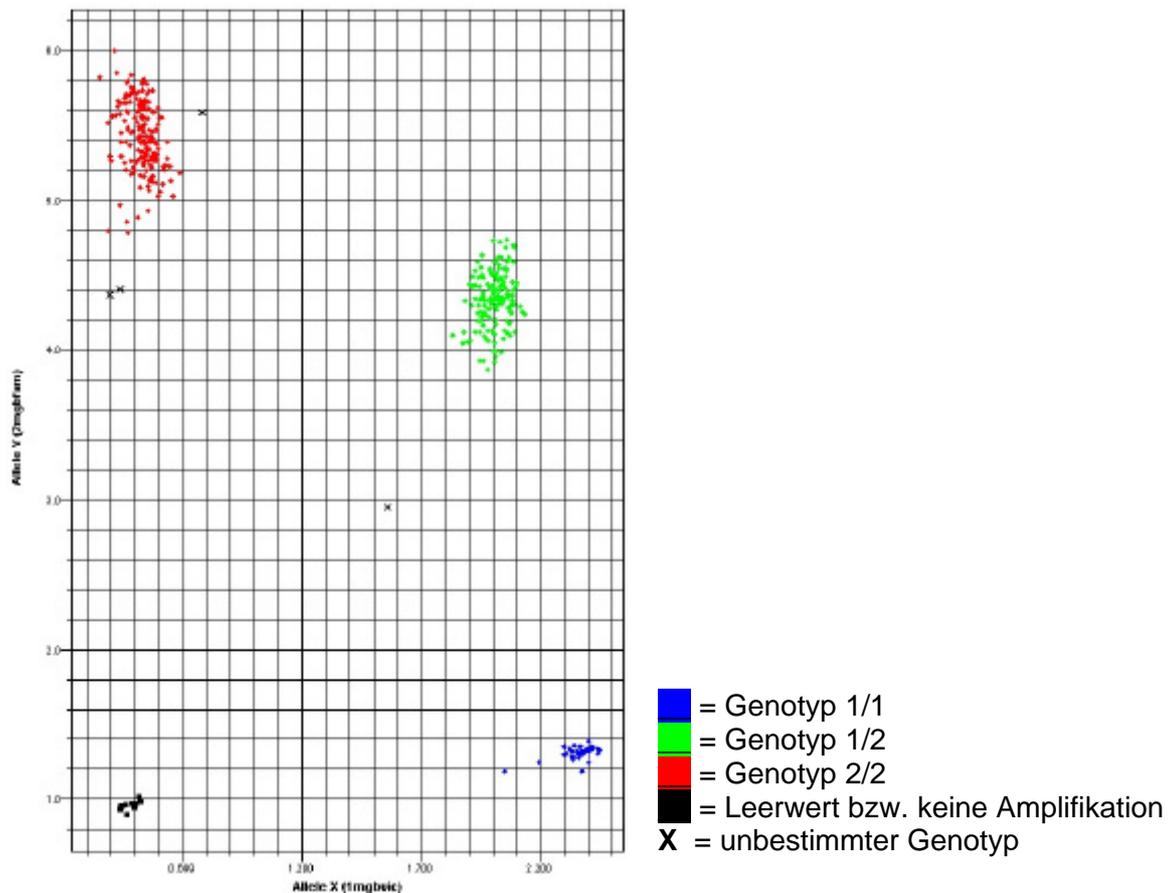


Abb. 2.3: TaqMan[®]-Fluoreszenzcluster. Die Fluoreszenzwerte der beiden allelspezifischen Farbstoffe werden je Einzelprobe gegeneinander aufgetragen. Die Stärke der Fluoreszenz ist proportional zu der Anzahl gebundener Sonden und damit zu der Anzahl der allelischen Amplifikate. Der homozygote Genotyp entspricht der Fluoreszenz eines einzelnen Farbstoffes in der Probe; werden beide Fluoreszenzen in einer Probe gemessen, so entspricht dieses dem heterozygoten Genotyp. Eine Hintergrundfluoreszenz muss bei der Clusterdefinition - der so genannten Allelzuordnung (*allele calling*) - berücksichtigt werden.

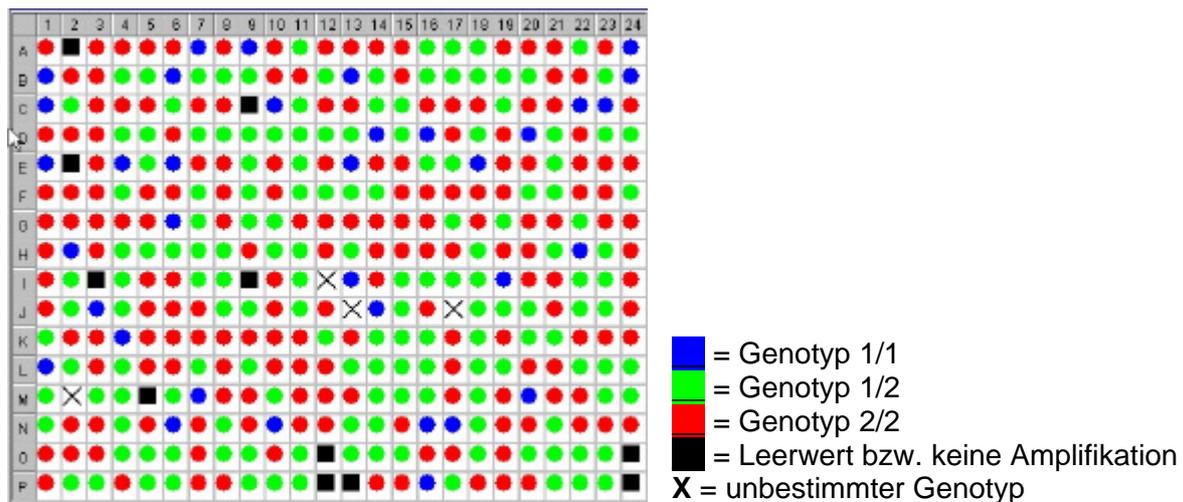


Abb. 2.4: Zuordnung der Genotypen. Basierend auf der manuellen Abgrenzung der Genotypencluster (Abb. 2.3) erfolgt die Zuweisung des Genotyps zu den einzelnen Kammern der PCR-Platte. Die Genotypenverteilung auf der PCR-Platte wird zum Abschluss des Genotypisierungsexperimentes in eine Datenbank überführt, in der die Genotypen den DNA-Proben und damit den Individuen zugeordnet werden.

2.4 Mutationsdetektion und SNP-Verifikation

2.4.1 Prinzip der Sequenzierung

Das Ermitteln („Lesen“) der Basenabfolge eines DNA-Abschnittes wird als Sequenzierung bezeichnet. In der vorliegenden Studie wurde hierzu eine konventionelle PCR-Reaktion (siehe 2.3.1) mit einer Kettenabbruch-Synthese nach Sanger [Sanger *et al.*, 1977] kombiniert. Im ersten Schritt wurde mittels einer PCR-Reaktion der Ziel-DNA-Abschnitt amplifiziert. Anschließend erfolgte eine Kettenabbruch-Reaktion, indem ein Teil des aufgereinigten PCR-Produktes zusammen mit dem F- oder R-Primer, einer DNA-Polymerase, dNTPs und sogenannten Didesoxynukleotiden (ddNTPs) unter Durchlaufen mehrerer Temperaturzyklen reagierte. Die ddNTPs besitzen keine 3'-Hydroxygruppe, sodass nach Einbau eines solchen Nukleotids eine Strangverlängerung nicht mehr möglich ist und es zu einem Syntheseabbruch kommt. Desweiteren ist jedes der vier verschiedenen ddNTPs mit einem unterschiedlichen Fluoreszenzfarbstoff markiert, sodass eine Diskriminierung der resultierenden DNA-Fragmente an Hand der Fluoreszenz und Fragmentlänge möglich ist. Die dNTPs lagen in einer höheren Konzentration als die ddNTPs vor, sodass hierdurch der Einbau eines ddNTPs seltener erfolgte und auch längere fluoreszenzmarkierte Fragmente synthetisiert werden konnten. Durch das Durchlaufen mehrerer Reaktionszyklen sollte

gewährleistet werden, dass am Reaktionsende eine ausreichende Anzahl aller potentiell möglichen, fluoreszenzmarkierten DNA-Einzelstränge unterschiedlicher Länge vorlag. Im Anschluss an die Kettenabbruchreaktion wurde das aufgereinigte Gemisch der DNA-Fragmente mittels einer Kapillar-Gel-Elektrophorese in die einzelnen Längenfraktionen aufgetrennt. Gleichzeitig erfolgte mit Hilfe eines Lasers und eines spektrografischen Detektors die Zuordnung der basenspezifischen Fluoreszenz gegenüber den DNA-Fragmenten. Da die Wandergeschwindigkeit der Fragmente umgekehrt proportional zu der Fragmentlänge ist, und gleichzeitig Fragmente gleicher Länge mit dem gleichen ddNTP enden (Ausnahme: Fragment endet mit einem SNP), ist die Abfolge der Fluoreszenzmessungen direkt der Basenabfolge des amplifizierten DNA-Abschnittes zuzuordnen (Abb. 2.5).

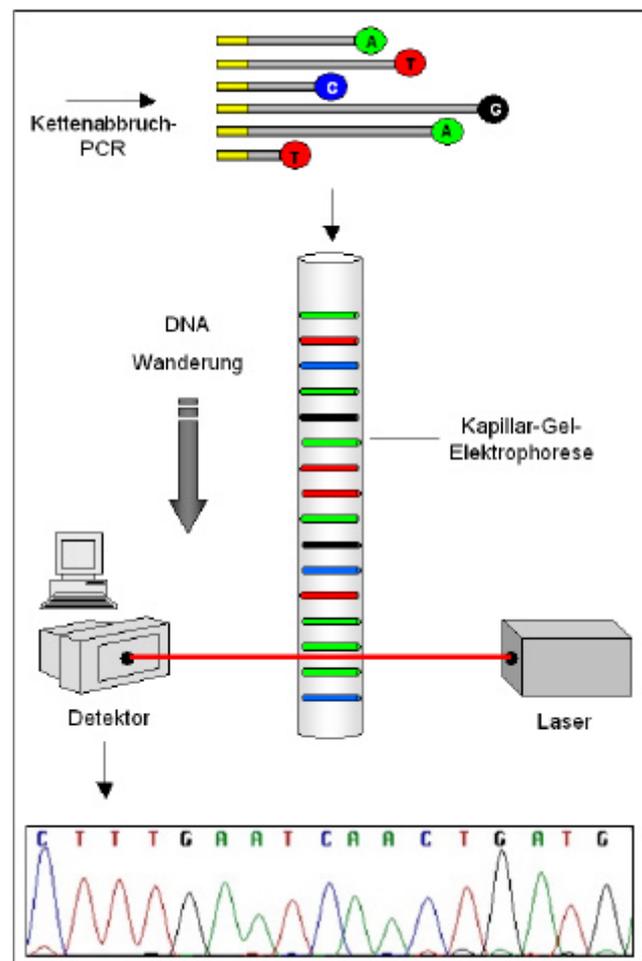


Abb. 2.5: Prinzip der Sequenzierung nach Sanger. Die fluoreszenzmarkierten DNA-Fragmente der Kettenabbruchreaktion werden elektrophoretisch in einer Gelkapillare aufgetrennt. Eine fixierte Laser-Detektor-Einheit misst die Fluoreszenz der vorbeiziehenden Fragmente und speichert diese ab. Die Abfolge der basenspezifischen Fluoreszenzmessungen wird abschließend direkt in die Basenabfolge des amplifizierten Sequenzabschnittes übersetzt.

2.4.2 Methode der Sequenzierung nach Sanger

Bei der Auswahl von SNPs aus öffentlichen Datenbanken (NCBI, Ensembl) musste ein gewisser Anteil der Marker mit Hilfe der Sequenzierung vor Beginn der Genotypisierung verifiziert werden. Dieses war für solche SNPs erforderlich, die keine Frequenzangaben für eine kaukasische Population - basierend auf spezifischen Sequenzier- oder Genotypisierungsexperimenten – hatten. Gleiches galt für SNPs, dessen Datenbankeinträge ausschließlich auf Sequenzvergleichen [Ning *et al.*, 2001] genomweiter Klon-Sequenzierungen beruhten. Desweiteren wurde die Sequenzierung zur Detektion unbekannter SNPs in Kandidatengenomen verwendet. Im Wesentlichen wurden hierbei die jeweiligen Exone, Intron-Exon-Übergänge (*splice sites*) und Promoterregionen sequenziert.

Der erste Schritt eines jeden Sequenzier-Experimentes bestand aus dem Design der Primerpaare für den zu sequenzierenden DNA-Abschnitt. Hierbei wurden die selben Kriterien des Designs verwendet wie sie im Kapitel 2.3.2 beschrieben sind. Die Amplikonlänge wurde dabei auf maximal 500 Bp begrenzt, wodurch akkurater sequenziert werden konnte. Unter Verwendung des F- und R-Primers konnte zudem das PCR-Produkt in beide Strangrichtungen abgelesen werden. Somit lag jedes sequenzierte DNA-Fragment am Ende des Experimentes als *antisense* (codogener) Strang und *sense* Sequenz vor, was die Sicherheit der Sequenzierung durch eine Vergleichbarkeit beider Stränge erhöhte. Hierauf erfolgte eine Temperaturoptimierung der Reaktion, in dem die optimale Anlagerungstemperatur der Oligonukleotide in einer Gradienten-PCR ermittelt wurde. Dabei wurde die selbe DNA in allen Reaktionsansätzen (Tab. 2.4) verwendet, um eine Vergleichbarkeit der unterschiedlichen Reaktionsbedingungen zu gewährleisten. Die Gradienten-PCR-Reaktion ist dadurch gekennzeichnet, dass in der PCR-Maschine ein Temperaturgradient definierter Größe während der PCR-Zyklen aufgebaut wurde. Der Gradient verlief in horizontaler Richtung auf der 96-Kammer-Platte, sodass jeweils die DNA-Proben einer Plattenspalte der gleichen Reaktionstemperatur ausgesetzt waren und der Gradient sich in 12 Abstufungen mit einer Temperaturzunahme von Plattenspalte 1 nach 12 unterteilte. Zur Erhöhung der spezifischen Primerbindung in der PCR-Reaktion wurde das *touchdown*-PCR-Prinzip [Don *et al.*, 1991] verwendet. Hierbei wurde in der ersten Hälfte der PCR-Reaktion die Anlagerungstemperatur zyklweise um 0,5°C erniedrigt (Tab. 2.5). Die mittlere Anlagerungstemperatur des ersten Zyklus lag dabei über der rechnerisch ermittelten. Somit konnte die Bildung von Primerdimeren und Artefakten reduziert und die Ausbeute an spezifischem PCR-Produkt erhöht werden.

Tab. 2.4: Reaktionsansatz für die Primeroptimierung der Sequenzierung.

Ansatz für 25ml Einzelreaktion	Volumen [ml]
Puffer (10fach)	2,50
MgCl ₂	2,00
dNTPs [10mM]	0,50
F-Primer [20µM]	0,10
R-Primer [20µM]	0,10
Taq Polymerase	0,15
DNA [5ng/µl]	1,00
Wasser, PCR-Qualität	18,65
Gesamtvolumen	25,00

Tab. 2.5: PCR-Bedingungen der *touchdown*-Gradienten-PCR. Innerhalb der ersten 15 Zyklen wurde die Anlagerungstemperatur pro Zyklus um 0,5°C abgesenkt. Die angegebenen Anlagerungstemperaturen beziehen sich auf die mittlere Gradiententemperatur.

Temperatur	Zeit	Prozess	Anmerkung
95 °C	5 min	konstant	
95 °C	30 s	15 Zyklen	* mittlere <i>touchdown</i> Starttemperatur, td=-0,5°C/Zyklus
66 °C*	30 s		
72 °C	30 s		
95 °C	30 s	18 Zyklen	* mittlere <i>touchdown</i> Endtemperatur
58 °C*	30 s		
72 °C	30 s		
72 °C	10 min	konstant	
4 °C	∞	konstant	

Im Anschluss an die PCR-Reaktion wurde das PCR-Produkt zur qualitativen Beurteilung auf einem Agarose-Gel elektrophoretisch aufgetrennt [Takahashi *et al.*, 1969] (Abb. 2.6) und die optimale Anlagerungstemperatur der Primerpaare für die nachfolgende PCR- und Sequenzier-Reaktion definiert. Angestrebt wurde hierbei eine klare, distinkte Gelbande, welche auf eine spezifische PCR-Reaktion rückschließen lässt. Zusätzliche Dimerprodukte der beiden Oligonukleotide geringer Fragmentgröße oder Artefakte bildeten ein Ausschlusskriterium für die Temperatúrauswahl.

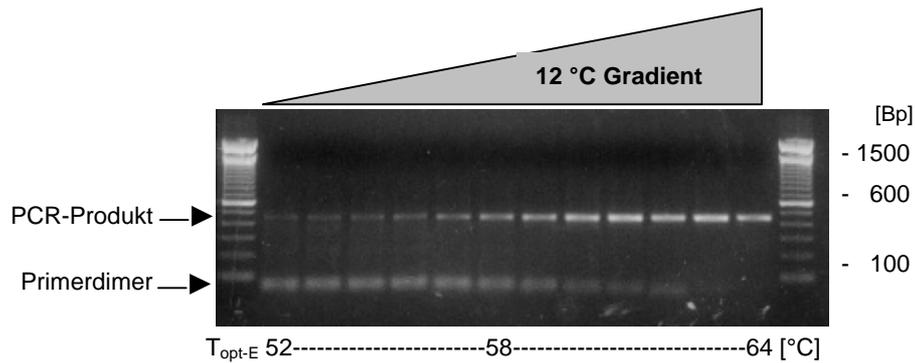


Abb. 2.6: Qualitative Bestimmung von PCR-Produkten einer *touchdown*-Gradienten-PCR mittels Agarose-Gel-Elektrophorese. Abgebildet ist das PCR-Produkt der Primerpaaroptimierung für das Exon 5 des Kandidatengens *ZNF23* (Zinkfingerprotein 23). Die Amplikonlänge beträgt 449 Bp bei einer errechneten Primeranlagerungstemperatur von 58,6°C (F-Primer) bzw. 58,9°C (R-Primer).

Der Primeroptimierung schloss sich die eigentliche PCR-Reaktion unter angepassten PCR-Bedingungen an. Für die Mutationsdetektion bzw. SNP-Verifikation wurde eine DNA-Auswahl von 47 Proben verwendet, die im gleichen Verhältnis aus Morbus Crohn und Colitis ulcerosa Patienten bestand. Auf einer 96-Kammer-Platte wurden die 47 DNA-Proben in jeweils 2 Plattenkammern gefüllt, sodass in der eigentlichen Sequenzierreaktion jede DNA-Probe sowohl mit dem F- als auch R-Primer sequenziert werden konnte. Insbesondere bei niederfrequenten SNPs konnte hierdurch eine hohe Sequenziertrate erzielt werden. Gleichzeitig stellte dieser Ansatz eine interne Qualitätskontrolle für den jeweilig sequenzierten Gegenstrang dar. Zusätzlich durchliefen zwei DNA-freie Kammern den gesamten Sequenzier-Prozess, um als Leerkontrollen u.a. hinsichtlich Kontamination durch Fremd-DNA zu dienen. Die PCR-Reaktion als solches entsprach der *touchdown*-PCR-Reaktion der Primeroptimierung mit dem Unterschied, dass anstelle des Gradienten nun für alle Proben die optimierte *touchdown*-Start- bzw. Endtemperatur verwendet wurde. Die spezifische Anlagerungstemperatur variierte demzufolge im jeweiligen Experiment (Tab. 2.7) und der Reaktionsansatz erfolgte gemäß des Primeroptimierungsprotokolls (Tab. 2.4).

Tab. 2.6: PCR-Bedingungen der Sequenzier-*touchdown*-PCR-Reaktion. Die Anlagerungstemperatur T_{opt} ist spezifisch für die benutzten Primer und wurde in einer jeweilig vorhergehenden *touchdown*-Gradienten-PCR ermittelt.

Temperatur	Zeit	Prozess	Anmerkung
95 °C	5 min	konstant	
95 °C	30 s	16 Zyklen	* optimierte <i>touchdown</i> Starttemperatur, $td=-0,5^{\circ}\text{C}/\text{Zyklus}$
$T_{\text{opt-S}} [^{\circ}\text{C}]^*$	30 s		
72 °C	30 s		
95 °C	30 s	19 Zyklen	* optimierte <i>touchdown</i> Endtemperatur
$T_{\text{opt-E}} [^{\circ}\text{C}]^*$	30 s		
72 °C	30 s		
72 °C	10 min	konstant	
4 °C	∞	konstant	

Anschließend wurde vor der eigentlichen Kettenabbruch-Sequenzierreaktion ein Teil des PCR-Produktes für die weitere Verwendung mit den Enzymen Exonuklease I (Exo I) und *shrimp alkaline phosphatase* (SAP) aufgereinigt. Der enzymatische Verdau zielte zum einen auf die Beseitigung der restlichen Primer ab, da diese in der Nachfolgereaktion zur Sequenzierung von unleserlichen Mischsequenzen geführt hätten. Dieses konnte durch die Eigenschaft der Exo I einzelsträngige Oligonukleotide abzubauen [Berthold und Geider, 1976] erreicht werden. Zum anderen hätten die verbleibenden dNTPs der PCR-Reaktion das dNTP/ddNTP-Verhältnis der Sequenzierreaktion unkalkulierbar und störend verändert. Daher kam hier die dephosphorylierende SAP zum Einsatz [Sauer *et al.*, 2000]. Zum Abschluss der enzymatischen Aufreinigung war es zwingend notwendig die Verdauenzyme zu inaktivieren (Tab. 2.8), da besonders die SAP sehr stabil ist und die Sequenzierreaktion ansonsten nachhaltig gestört hätte.

Tab. 2.7: Reaktionsansatz für die enzymatische Aufreinigung des PCR-Produktes.

Ansatz für 10 _{ml} Einzelreaktion	Volumen [ml]
SAP (Shrimp Alkaline Phosphatase)	0,30
Exo I (Exonuklease I)	0,15
Wasser, PCR-Qualität	1,55
PCR Produkt	8,00
Gesamtvolumen	10,00

Tab. 2.8: Temperaturprotokoll der enzymatischen Aufreinigung des PCR-Produktes. Verbliebende Primer- und dNTP-Reste werden für die nachfolgende Sequenzierreaktion abgebaut.

Temperatur	Zeit	Prozess	Anmerkung
37 °C	15 min	konstant	Aufreinigung (Enzymverdau) des PCR-Produktes
72 °C	15 min	konstant	Enzym-Inaktivierung
10 °C	∞	konstant	

In der nachfolgenden Kettenabbruchreaktion reagierte ein Teil des aufgereinigten PCR-Produktes zusammen mit dem „BigDye[®]“-Reaktionsmix (Applied Biosystems Inc., Foster City, CA, USA) [Salas-Solano *et al.*, 1998], welcher im Wesentlichen aus DNA-Polymerase, dNTPs und fluoreszenzmarkierten ddNTPs bestand, und einem der beiden PCR-Primer, die nun als Sequenzier-Primer eingesetzt wurden (siehe 2.4.1). Die Sequenzierreaktion verlief in 25 Zyklen analog zu einer PCR-Reaktion, wobei die Primeranlagerungstemperatur entsprechend der Optimierung gewählt wurde (Tab 2.10). Da der dritte Zyklusabschnitt bedingt durch den „BigDye[®]“-Reaktionsmix sein Temperaturoptimum bei 60°C besaß, wurde die Obergrenze der Anlagerungstemperatur ebenfalls bei 60°C definiert, und es konnte zum Teil zu geringen Abweichungen gegenüber der optimierten Anlagerungstemperatur kommen [Wen, 2001].

Tab. 2.9: Reaktionsansatz für die Sequenzierreaktion.

Ansatz für 10ml Einzelreaktion	Volumen [ml]
Sequenzier-Reaktions-Mix (BigDye [®]), 1:5 verdünnt	2.00
F- oder R-Primer [1,6 µM]	1.00
Wasser, PCR-Qualität	5.00
PCR Produkt, aufgereinigt	2.00
Gesamtvolumen	10.00

Tab. 2.10: Temperaturprotokoll der Sequenzierreaktion.

Temperatur	Zeit	Prozess	Anmerkung
96 °C	30 s	25 Zyklen	* optimierte Anlagerungstemperatur Kettenabbruch-Reaktion
T _{opt-E} [°C]*	15 s		
60 °C	4 min		
4 °C	∞	konstant	

Bevor das Produkt der Kettenabbruchreaktion mit Hilfe der Kapillar-Gel-Elektrophorese in die verschiedenen Längenfraktionen aufgetrennt werden konnte, erfolgte ein mechanischer Aufreinigungsschritt. Nicht eingebaute, markierte ddNTPs bildeten eine Störgröße in Form von Fluoreszenz-Hintergrundrauschen für die Bestimmung der unterschiedlichen, fluoreszenzmarkierten DNA-Fragmente. Salze aus dem Reaktionspuffer, Reste des PCR-Produktes und nicht eingebaute dNTPs besaßen ebenfalls die Fähigkeit die Elektrophorese negativ zu beeinflussen. Daher wurde das Sequenzierprodukt durch eine Gel-Filtration auf Basis des modifizierten Polysaccharids Dextran („Sephadex™ G-50“ - superfein, Amersham Biosciences, Freiburg, Deutschland) aufgearbeitet [Mort *et al.*, 1998]. „Sephadex™“-Pulver ging in Kombination mit Wasser und einem mehrstündigen Quellprozess in einen gelartigen Zustand über, wobei die Dextranmoleküle sich in einem dreidimensionalen Netzwerk anordneten und eine porenartige Struktur im Gel ausgebildet wurde. Die Porengröße lag dabei im Rahmen von 20µm bis 80 µm, sodass die oben genannten Störelemente bei der Filtration in diesen Poren zurückgehalten wurden. Nachdem das verdünnte Sequenzierprodukt auf die „Sephadex™“-Gelsäule aufgetragen worden war, folgte ein Zentrifugationsschritt, der das Gemisch aus Sequenzierprodukt und Reaktionsreagenzien voneinander trennte. Das Zentrifugat wurde in einer 96-Kammer-Platte, Modell „MicroAmp® optical 96-well reaction plate“ (Applied Biosystems Inc., Foster City, CA, USA), aufgefangen und mit einer selbstklebenden Aluminiumfolie versiegelt.

Tab. 2.11: Protokoll zur „Sephadexä“-Aufreinigung des Sequenzierproduktes.

Ansatz für 96-Proben (Sephadex™-platte MAHVN 4550)
1) Sephadex™ G-50 superfein in die Säulen-Ladevorrichtung füllen. 2) Sephadexplatte™ mittels Ladevorrichtung befüllen. 3) Je Kammer 300 µl PCR-Wasser zugeben. 4) Mind. 2 Stunden Quellzeit für das Sephadex™-Gel. 5) 5 min Abzentrifugieren des Restwassers bei 2100 U / min. 6) Waschen des Gels mit 150 µl pro Kammer. 7) 5 min Abzentrifugieren des Waschwassers bei 2100 U / min. 8) Sequenzierprodukt mit 10 µl PCR-Wasser pro Kammer verdünnen. 9) Auftragen des Sequenzierproduktes auf die Gelsäule. 10) 5 min Abzentrifugieren des gereinigten Produktes bei 2100 U / min in die Sequenzierplatte.

Die abschließende Kapillar-Gel-Elektrophorese wurde auf einem automatisierten Hochdurchsatz-96-Kapillar-Sequenzierer (Modell „96-capillary fluorescence detection system ABI PRISM® 3700 DNA Analyzer“, Firma Applied Biosystems Inc., Foster City, CA, USA) durchgeführt und die Sequenzchromatogramme wurden in der institutsinternen Datenbank gespeichert. Die Auswertung der Sequenzdaten in Form von Chromatogramm-*alignments*

(Sequenzabgleiche) und Sequenzvergleichen mit den Consensussequenzen öffentlicher Datenbanken (NCBI, Ensembl) erfolgte mit der Auswertesoftware Sequencher, Version 4.2 und 4.5 (Gene Codes Corporation, Ann Arbor, MI, USA).

2.5 Chromosomale Lokalisierung von Kandidatenregionen

2.5.1 Prinzip der Fluoreszenz-in-situ-Hybridisierung

Im Allgemeinen ist die in-situ-Hybridisierung (ISH) als Methode charakterisiert, Nukleinsäuresequenzen in Form von Ribonukleinsäure oder DNA in Geweben, Zellen, Zellkernen oder Chromosomen optisch hervorzuheben und damit zu lokalisieren [John *et al.*, 1969; Pardue und Gall, 1969]. Die Fluoreszenz-in-situ-Hybridisierung (FISH), bei der es zum Einsatz von Fluoreszenzfarbstoffen als optische Marker kommt, stellt eine Form der ISH-Methodik dar [Lawrie und Gosden, 1980; Lichter *et al.*, 1990]. Hierbei wird die zu hybridisierende Sonde in Form einer spezifischen DNA-Sequenz mit Markermolekülen wie z.B. Biotin oder Digoxigenin versehen. Diese Markermoleküle wiederum können durch den Einsatz von markerspezifischen Antikörpern, die ihrerseits mit einem Fluoreszenzfarbstoff gekoppelt sind, detektiert werden.

Der schematische Ablauf einer FISH von Chromosomen beginnt mit der Präparation von Interphasekernen oder Metaphasechromosomen, indem z.B. kultivierte Blutzellen durch eine hypotone Lösung zum Zerplatzen gebracht und auf einem Objektträger fixiert werden. Der Grad der resultierenden Chromosomenspreitung ist dabei maßgeblich für die Auswertbarkeit des FISH-Experimentes. Desweiteren können Cytoplasmaresten oder zelluläre Strukturen den Zugang der DNA-Sonde zu den Chromosomen behindern und so einen Anstieg der Hintergrundstrahlung verursachen. Die chromosomale DNA wird im nächsten Schritt mit Hitze denaturiert und im anschließenden Renaturierungsschritt mit der markierten Sonde hybridisiert. Die Sondensequenz ist dabei mehrere Kilobasen lang und gleichzeitig spezifisch für die zu detektierende chromosomale Region. Es erfolgt ein Waschschrift, um die nicht hybridisierten Sonden zu entfernen. Im Anschluss wird der fluoreszenzmarkierte Antikörper zugegeben. Im Beispiel des Markers Biotin erfolgt die Zugabe des Biotinhemmstoffes Avidin, das eine hohe Bindungsspezifität und -affinität gegenüber Biotin vergleichbar einer Antigen-Antikörper-Bindung besitzt; Avidin ist jedoch per Definition kein Antikörper. Das Avidin wiederum ist z.B. als FITC-Avidin mit dem Fluoreszenzfarbstoff Fluoresceinisothiocyanat gekoppelt. Die Lokalisation des Markermoleküls erfolgt mit Hilfe eines Fluoreszenzmikroskops, das mit einem

Anregungsfilter, welcher der Emissionswellenlänge des Fluoreszenzfarbstoffes entspricht, ausgestattet ist. Zusätzlich werden die Chromosomen mit einer Gegenfärbung durch z.B. 4',6-Diamidino-2-phenylindol-dihydrochlorid sichtbar gemacht. Hierdurch wird eine Identifizierung der einzelnen Chromosomen und somit eine absolute und relative Zuordnung des Markermoleküls zu dem hybridisierten DNA-Abschnitt ermöglicht (Abb. 2.7 B).

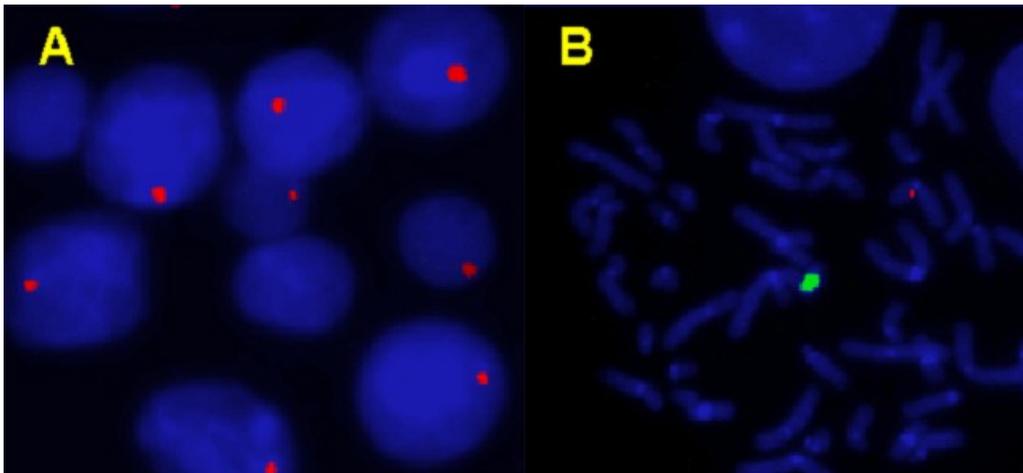


Abb. 2.7: Fluoreszenz-in-situ-Hybridisierung. A) Hybridisierung einer x-chromosomalen Sonde in Interphasezellen. B) Hybridisierung von x- und y-chromosomalen Sonden gegen gespreitete Metaphasechromosomen.

2.5.2 Methode der Fluoreszenz-in-situ-Hybridisierung

In der vorliegenden Arbeit wurden Sonden in Form von markierten BAC (*bacterial artificial chromosome*)-Klonen verwendet. Dabei beinhaltete das klonierte DNA-Fragment die flankierende Sequenz der SNPs, die per FISH auf dem Chromosom 16 nachgewiesen und relativ zueinander lokalisiert werden sollten. Die BAC-Klone wurden in der Datenbank des deutschen Ressourcenzentrums für Genomforschung (RZPD) ausgewählt und anschließend über die BAC-Klon-Sammlung des RZPDs bezogen. Die Kultivierung der BAC-Klone erfolgte institutsintern, und das FISH-Experiment wurde in Kooperation mit dem Institut für Humangenetik im Universitätsklinikum Schleswig-Holstein (Campus Kiel) durchgeführt.

2.6 Interne Datenbank

2.6.1 Verwaltung der Patienten-, Phänotyp- und Genotypdaten

Die Speicherung und Verwaltung jeglicher Information, die im Zusammenhang mit dem verwendeten Probenmaterial erfasst oder generiert wurde, erfolgte mittels eines internen Datenbanksystems [Hampe, Wollstein *et al.*, 2001]. Dazu gehörten auf Seiten der Patienteninformationen Angaben wie Alter, Geschlecht, regionale Herkunft, phänotypische Charakterisierung und Einteilung in die Subphänotypen der CED, sowie die Verwandtschaftsverhältnisse (Stammbaumstruktur) zwischen Individuen einer Familie. Jede Probe und jedes Individuum erhielt eine einmalig zugeordnete, eindeutige Identifikationsnummer, sodass sowohl eine Probenverwechslung als auch eine Rückverfolgung zur Identität eines Patienten durch diese Form der anonymisierten Identifizierung ausgeschlossen wurde. Angaben zur Person wie Name oder Adresse wurden nicht in dieser Datenbank gespeichert. Eine Verknüpfung der Angaben zur Konzentration und Plattenposition der einzelnen DNA-Proben mit den Patientennummern erfolgte ebenfalls in der Datenbank, wobei jede PCR-Platte als solche durch einen individuellen Barcode gekennzeichnet war. Desweiteren wurden die SNP-Assay-Informationen (Primer-/Probensequenzen und -konzentrationen, Chromosomenposition, Locusinformation) und individuellen Genotypdaten in der Datenbank gespeichert und für die einzelnen Analysen beim Datenexport mit den Patienteninformationen verknüpft. Das Format des Datenexports entsprach den Kriterien der für die Datenanalyse verwendeten Programme (siehe 2.7). Über eine Auswahl verschiedener Software-Applikationen konnte u.a. der Genotypisierungsstatus der einzelnen SNPs für ein Einzelindividuum, eine Familie oder eine Zusammenstellung verschiedener PCR-Platten erfolgen. Ebenso war es möglich SNPs in unterschiedlichen Kombinationen für verschiedene Patientensätze zu analysieren und miteinander zu vergleichen. Auf Seiten der Laborverwaltung kam die Datenbank z.B. bei der Gestaltung neuer PCR-Platten-Layouts, der Steuerung der automatisierten Plattenproduktion oder DNA-Quantifizierung zum Einsatz.

2.6.2 Qualitätskontrolle

Bevor die Analyse der Genotypisierungsdaten erfolgte, durchliefen die Typisierungsergebnisse mehrere Stationen der Qualitätskontrolle. Als erstes erfolgte die Überprüfung der Fluoreszenzwerte der Leerwertpositionen (Negativ-Kontrollen) jeder einzelnen Platte, nachdem die Zuordnung der Genotypen durch das Abgrenzen der

Fluoreszenzclustern abgeschlossen war. Befanden sich Leerwerte zwischen oder in den Clustern, so war dieses ein Hinweis auf Kontamination mit Fremd-DNA und die Ergebnisse mussten verworfen werden. Ausschließlich Platten deren Leerwerte in der Umgebung des Ursprungs des Fluoreszenzkoordinatensystems lagen gelangten in die nächste Phase, den Datenimport. Ähnliche Kriterien galten bei der Beurteilung der Genotypisierung der Positiv-Kontrollen in Form einer Zelllinien-DNA, die sich auf jeder Typisierungsplatte befand. Verschiedene Platten, die mit dem selben TaqMan[®]-Assay genotypisiert worden waren, mussten den gleichen Genotyp für die Positiv-Kontrolle zeigen, um in die Analyse zu gelangen. Der nächste Schritt vor dem Import der Genotypen in die Datenbank bestand aus der Kontrolle der Genotypen eines Stammbaumes (Familien, Trios) hinsichtlich Mendelscher Vererbungsfehler [Douglas *et al.*, 2002; Ewen *et al.*, 2000]. Zeigten sich vereinzelt Familien mit Genotypzuordnungen, die basierend auf den Mendelschen Vererbungsregeln nicht möglich waren, so wurde als erstes das entsprechende Fluoreszenzdiagramm auf eine eindeutige Zuordnung der Familienmitglieder überprüft. Handelte es sich um grenzwertige Genotypzuordnungen, so wurde der Stammbaum für den entsprechenden SNP auf den Genotyp 0 gesetzt. Waren die Genotypen eindeutig zugeordnet, so handelte es sich entweder um vertauschte DNA-Proben oder um nicht leibliche Väter. In beiden Fällen wurden die entsprechenden Elternteile mit dem Genotyp 0 versehen. TaqMan[®]-Assays, die eine hohe Anzahl von vermeintlichen Vererbungsfehlern aufzeigten, wurden als nicht glaubwürdig eingestuft und aus der Analyse ausgeschlossen. Es folgte die Kontrolle der Genotypisierungsrate der einzelnen Assays, wobei das Ausschlusskriterium für einen Assay bei einer Typisierungseffizienz kleiner 95% lag. Der Abschluss der Qualitätskontrolle für die Genotypisierungsexperimente bestand in der Abschätzung, inwieweit die Genotypen einer Population von nichterkrankten, unverwandten Personen vom Hardy-Weinberg-Gleichgewicht abwichen [Hosking *et al.*, 2004; Xu *et al.*, 2002]. Hierzu wurde ein χ^2 -Test mit einem Signifikanzschwellenwert von einem Prozent verwendet.

Bei der Sequenzierung bestand die Qualitätskontrolle aus der Verwendung des F- und R-Primers in der Sequenzierreaktion. Somit konnte zum einen die Genauigkeit der generierten Sequenzen durch Vergleiche der korrespondierenden plus- und minus-Sequenzen überprüft werden. Zum anderen musste ein Basenaustausch sowohl auf dem codogenen als auch komplementären Strang vorhanden sein, um als SNP charakterisiert zu werden. Analog zur Genotypisierungsroutine gehörte auch bei der Sequenzierung die Kontaminationskontrolle durch die Verwendung von Negativ-Kontrollen in Form von DNA-freien PCR-Platten-Kammern zum Standardprotokoll.

2.7 Statistische Methoden

Für die Genkartierung der Kandidatengene wurden Assoziationsanalysen in Form des familienbasierten Ansatzes durch den Transmissions-Ungleichgewicht-Test (*transmission disequilibrium test, TDT*) [Spielman *et al.*, 1993] und der auf Einzelpatienten und Einzelkontrollen beruhenden Fall-Kontroll-Analyse (*case control, CC*) in Form einer χ^2 - und *odds ratio* (OR)-Statistik durchgeführt. Dabei erfolgten die Analysen für die drei diagnostischen Krankheitskategorien CED, MC und CU. Allelische und genotypische Frequenzangaben für die einzelnen Assoziationsmarker wurden jeweils getrennt für die Kontroll- und Patientengruppe berechnet, wobei in letztere nur jeweils ein Patient pro Familie miteinbezogen wurde. Somit sollte ein systematischer Fehler durch einen unterschiedlich starken, familiären Beitrag vermieden werden. Zur Berechnung von Haplotypblockstrukturen und der Abschätzung von allelischer Kopplung zwischen Markern wurde das Kopplungsungleichgewicht als Hauptkriterium herangezogen.

Die einzelnen Analyseformen werden nachfolgend hinsichtlich ihrer Berechnungsgrundlage und Aussagefähigkeit beschrieben:

2.7.1 Kopplungsungleichgewicht

Das Kopplungsungleichgewicht (LD, *linkage disequilibrium*), auch als gametisches Ungleichgewicht bekannt, wird auf der Grundlage der spezifischen Allelfrequenz zweier Loci A und B beschrieben. LD liegt vor, wenn die Wahrscheinlichkeit für das Vorhandensein eines spezifischen Allels am Locus A und Locus B in einer einzelnen Keimzelle nicht dem Produkt der Einzelhäufigkeiten der Allele entspricht. Handelt es sich bei den beiden Allelen um ein Markerallel und ein Allel eines Krankheitsgens, so resultiert das Kopplungsungleichgewicht in einer Assoziation zwischen diesen. Entsprechend detektierbar wird diese Assoziation durch den Umstand, dass bestimmte Markerallele häufiger in erkrankten als in zufällig ausgewählten Personen einer Population auftreten. Somit bildet ein Kopplungsungleichgewicht die Grundlage der Assoziationsstudie [Maniatis *et al.*, 2004; Morton, 2005; Zhang *et al.*, 2002]. Es gilt zu beachten, dass eine Kopplung (*linkage*) von Markerlocus und Kandidatengenlocus im Gegensatz zum LD eine gemeinsame - gekoppelte - Vererbung in Familien beschreibt, die unabhängig von der Allelspezifität beider Loci ist. Dieses ist ein wesentliches Unterscheidungsmerkmal für Kopplungs- und Assoziationsstudien.

Die Berechnung des LD erfolgte paarweise für die biallelischen SNP-Marker. Die Häufigkeiten P der Allelkombinationen für den Locus A (Allele A, a) und den Locus B (Allele B, b) wird in Form einer Kontingenztabelle dargestellt.

Tab. 2.12: Kontingenztabelle der LD-Berechnung.

		Locus B		Σ
		B	b	
Locus A	A	P_{AB}	P_{Ab}	P_A
	a	P_{aB}	P_{ab}	P_a
Σ		P_B	P_b	1

Im Kopplungsgleichgewicht sind die erwarteten Haplotypfrequenzen gleich dem Produkt der Allelfrequenzen: $P_{AB} = P_A \cdot P_B$.

Die Abweichung D vom Erwartungswert des entsprechenden Haplotyps wird berechnet durch:

$$D_{AB} = P_{AB} - P_A \cdot P_B = P_{AB} \cdot P_{ab} - P_{Ab} \cdot P_{aB}$$

Für biallelische Loci gilt, dass der absolute Wert der Abweichung für alle vier möglichen Haplotypen der gleiche ist und somit die Abweichung für einen Haplotyp die, der anderen beschreibt. Um das LD von verschiedenen, sich in den Allelfrequenzverteilungen unterscheidenden Locus-Paaren vergleichen zu können, wird die Abweichung D normalisiert. Die am häufigsten verwendeten Standardisierungen sind D' (*Dprime*) [Lewontin, 1964, 1988] und r^2 (auch Δ^2 genannt) [Hill und Robertson, 1968]:

$$D' = \begin{cases} \frac{D}{\min(P_A \cdot P_b, P_a \cdot P_B)} & : D > 0 \\ \frac{D}{\max(-P_A \cdot P_B, -P_a \cdot P_b)} & : D < 0 \end{cases}, \quad r^2 = \frac{D^2}{(P_A \cdot P_B \cdot P_a \cdot P_b)}$$

Als vollständiges LD wird ein D' von 1 und als perfektes LD ein r^2 von 1 bezeichnet. Letzteres tritt ausschließlich in Populationen auf, in denen ein ‚Flaschenhals-Ereignis‘ (*bottleneck effect*) bzw. genetische Drift maßgeblich stattgefunden hat. Generell gilt es zu beachten, dass die wesentlichen Einflussfaktoren für ein Kopplungsungleichgewicht das Alter der Mutationen, die Rekombinationsfraktion, die Entwicklungsgeschichte der Population und die genetische Drift sind, und somit das LD selbst bei eng benachbarten Loci stark variiert [Suarez und Hampe, 1994]. Für Assoziationsstudien bedeutet das vor allem, dass ein LD auf der Kopplung von Marker- und Krankheitsallel beruhen kann, aber nicht muss. Eine Stratifizierung innerhalb der beprobten Population z.B. durch die unvollständige Durchmischung von Subpopulation (*admixture*) kann sich ebenfalls in einem erhöhten LD niederschlagen und somit Fehler erster Art (falsch positiv) in der Studie generieren.

D' ist als relatives Maß für ein LD nur bedingt brauchbar, da schon das Fehlen eines Haplotyps den D' -Wert auf 1 setzt [Ardlie *et al.*, 2002]. Insbesondere bei seltenen SNPs mit einer geringen Frequenz des weniger häufigen Allels (*minor allele frequency*) tritt diese Situation häufig auf. Ebenso haben ein kleiner bis moderater Probenumfang einen Einfluss auf den D' -Wert [Terwilliger *et al.*, 2002]. Somit ist D' eher als Indikator für fehlende Haplotypen bzw. Regionen häufiger Rekombination (*recombination hotspots*) als für die Abschätzung des Kopplungsungleichgewichts zu verwenden. Hingegen ist der r^2 -Wert diesen Störgrößen nicht erlegen und damit ein direkt proportionales Maß für das LD und eine indirekt proportionale Größe für das Auffinden von Assoziationen [Teare *et al.*, 2002].

In der vorliegenden Studie diente die Berechnung des Kopplungsungleichgewichtes der Abschätzung von LD-Blockstrukturen, um die Abhängigkeit von Assoziationssignalen auf Basis der physikalischen Lage der Marker zu beurteilen. Die initiale Kopplungsregion auf Chromosom 16q sollte damit schrittweise durch eine Feinkartierung in Kandidatengenregionen aufgelöst werden. Zusätzlich wurden LD-basierte Haplotypblöcke definiert und die Haplotyp-kennzeichnenden (*haplotype tagging*, ht) Marker bestimmt [Carlson *et al.*, 2004; Johnson *et al.*, 2001; Stram, 2004]. Neben den kodierenden SNPs stellten diese einen Grossteil der Marker für die Feinkartierungsexperimente dar. Die Berechnungen erfolgten unter Verwendung des Programms Haploview (Vers. 3.2) [Barrett *et al.*, 2005].

2.7.2 Transmission Disequilibrium Test

Familienbasierte Assoziationsanalysen wie der TDT bieten gegenüber der Fall-Kontroll-Statistik den Vorteil, falsch positive Assoziationssignale (*spurious association*), die durch Populationsstratifizierung allelische Kopplung vortäuschen (siehe 2.7.1), zu vermeiden. Der

TDT [Spielman *et al.*, 1993] überprüft gleichzeitig für einen biallelischen Marker mit den Allelen M1 und M2 die Kopplung und das LD des Markers. Als Assoziationstest identifiziert der TDT ein Kopplungsungleichgewicht nur, wenn auch gleichzeitig eine Kopplung vorliegt, was ihn inert gegenüber Populationsstratifizierung macht. Gleichzeitig ist der TDT auch ein Kopplungstest, der Kopplung nur in Anwesenheit von LD detektiert und deshalb ein geeignete Methode für den Kandidatengenansatz darstellt. Die Fähigkeit Kopplung mittels TDT zu identifizieren steigt mit zunehmendem Kopplungsungleichgewicht.

Grundlage für den TDT sind Kernfamilien - sogenannte Trios [Falk und Rubinstein, 1987] - in denen die Übertragung (Transmission) bzw. Nicht-Übertragung von Allelen von den Eltern auf das erkrankte Kind berechnet wird. Dabei wird jedes Elternteil gemäß des jeweilig übertragenden und nicht-übertragenden Allels klassifiziert (Tab. 2.13), wobei ausschließlich heterozygote Eltern als informativ erachtet werden. Die Nullhypothese ist $H_0: E(n_{12}) = E(n_{21})$, wobei n_{ij} , $i, j = 1, 2$, der Anzahl der $M_i M_j$ Eltern entspricht, die das Allel i an das Kind vererbt haben. Als Teststatistik wurde die McNemar-Statistik verwendet [Agresti, 1996], welche unter der Nullhypothese asymptotisch einer χ^2 -Verteilung mit einem Freiheitsgrad entspricht (Abb. 2.10):

$$\chi^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

Tab. 2.13: 2x2-Felder-Kontingenztabelle der übertragenen gegen die nicht-übertragenen Allele.

		nicht übertragenes Allel	
		M1	M2
übertragenes Allel	M1	n_{11}	n_{12}
	M2	n_{21}	n_{22}

Der TDT stellte neben der OR- und CC-basierten χ^2 -Statistik die wesentliche Analyseverfahren zur Auswertung der Genotypisierungsdaten dar. Die Berechnung erfolgte zum einen mit dem Programm TRANSMIT (Vers. 2.5) [Clayton, 1999], das den

ursprünglichen TDT-Ansatz in einer erweiterten Form nutzt. Hierbei handelt es sich um einen Vergleich der Anzahl der tatsächlich übertragenen Allele mit der zu erwartenden Anzahl. Letztere basiert auf der Allelfrequenz der Elternpopulation und wird unter Annahme des Hardy-Weinberg-Gleichgewichtes berechnet. Zur Überprüfung der statistischen Validität der TRANSMIT-Ergebnisse wurde ein *bootstrap*-Signifikanztest für 1000 *bootstrap*-Datensätze durchgeführt. Der TDT in seiner ursprünglichen Form wurde mit Hilfe des Programms Genehunter (Vers. 2.1_r2 beta) [Kruglyak *et al.*, 1996] berechnet. Hierbei wurde der TDT sowohl für die einzelnen Marker als auch für Multi-Locus-Gruppen ermittelt. Die Signifikanz der Genehunter-Ergebnisse wurde mittels eines Permutationstestes überprüft, indem neue Datensätze durch zufällige Vertauschung der jeweilig übertragenen Allele generiert wurden. Für die Einzelmarkeranalyse erfolgte dieses 10.000x, für die Multi-Marker-Analyse 1.000x. Jedes Einzelergebnis der simulierten Datensätze wurde mit dem Ergebnis der Originaldaten verglichen. In einer abschließenden Statistik wurde die Anzahl der Simulationen zusammengefasst, die die Schwellenwerte eines signifikanten ($p \leq 0,05$), hochsignifikanten ($p \leq 0,01$) und höchstsignifikanten ($p \leq 0,001$) Ergebnisses überschritten. In beiden Programmen erfolgte die Assoziationsanalyse für die einzelnen Marker und für 2- bzw. 3-Marker-Haplotypen, wobei die Haplotypen systematisch aus den jeweilig benachbarten SNPs (gleitendes 2- bzw. 3-Marker-Fenster) gebildet wurden.

2.7.3 χ^2 - und Odds Ratio-Statistik der Fall-Kontroll-Studie

Neben der familienbasierten TDT-Statistik kam eine weitere Methode der assoziationsbasierten Genkartierung in Form der CC-Studie zum Einsatz. Diese basiert auf dem Vergleich von allelischen oder genotypischen SNP-Frequenzen zwischen erkrankten Personen (*cases*) und gesunden Individuen (*controls*), die für die zu untersuchende Krankheit als nicht betroffen eingestuft werden. Da die CC-Statistik im Vergleich zum TDT anfälliger gegenüber einer Populationstratifizierung ist, kommt der Zusammenstellung der Patienten und Kontrollen eine besondere Bedeutung zu. Dabei gilt es zu beachten, dass Faktoren wie Alters- und Geschlechtsverteilung, ethnische und geographische Herkunft sowie von der Studie unabhängige Erkrankungen eine wichtige Größe bei der Angleichung der beiden Studiengruppen darstellen. Wesentliche Vorteile des CC-Ansatzes sind die schneller zu rekrutierenden Fall-Kontroll-Gruppen gegenüber den Familien/Trios der TDT-Statistik und die höhere statistische Verwertbarkeit aller Individuen gegenüber dem Ausschluss der nicht informativen, homozygoten Eltern in der TDT-Analyse.

Die statistische Auswertung erfolgte zum einen über den populationsbasierten OR, der den Quotienten aus der Expositionsquote der Erkrankten gegen die Expositionsquote der Gesunden darstellt. Dabei ist die Expositionsquote das Verhältnis der Anzahl Exponierter zur Anzahl Nichtexponierter unter den Erkrankten bzw. unter den Gesunden. Übertragen auf die SNP-basierte CC-Studie ist das Verhältnis der beiden Allele eines Loci der Expositionsquote gleichzusetzen (Tab. 2.14). In einem solchen Fall spricht man von einem allelischen OR.

Tab. 2.14: Kreuztabelle des Chancenverhältnis (Odds Ratio) in einer Fall-Kontroll-Studie. (H: hohes Risiko, G: geringes Risiko, n: Anzahl der Patienten, m: Anzahl der Kontrollen)

		Erkrankung	
		Ja (Fälle)	Nein (Kontrollen)
Risiko	H (Allel 1)	n_H	m_H
	G (Allel 2)	n_G	m_G
Σ		n	m

Der OR stellt in der CC-Analyse eine retrospektive Kenngröße dar. Er dient der Abschätzung, aber nicht der Bestimmung des relativen Risikos (RR). Das RR ist eine Maßzahl der prospektiven Studie und wird durch den OR nur dann annähernd abgebildet, wenn der Probenumfang der CC-Studie repräsentativ für die Bevölkerung ist und es sich um eine seltene Krankheit handelt:

$$OR = \frac{n_H \cdot m_G}{n_G \cdot m_H} \quad RR = \frac{n_H \cdot (n_G + m_G)}{n_G \cdot (n_H + m_H)}$$

Desweiteren können OR-Berechnungen auf Basis von Genotypen für verschiedene Risikomodelle der Erkrankung erfolgen (Tab. 2.15).

Tab. 2.15: Verschiedene Formen des allelischen und genotypischen Odds Ratio.

OR-Typ	hohes Risiko	geringes Risiko
OR Allel 1	häufiges Allel 1	seltene Allel 2
OR Träger Allel 1	Genotyp 1/1, 1/2	Genotyp 2/2
OR Heterozygot versus Allel 1	Genotyp 1/2	Genotyp 1/1
OR Homozygote Allel 1	Genotyp 1/1	Genotyp 2/2
OR Allel 2	seltene Allel 2	häufiges Allel 1
OR Träger Allel 2	Genotyp 1/2, 2/2	Genotyp 1/1
OR Heterozygot versus Allel 2	Genotyp 1/2	Genotyp 2/2
OR Homozygot Allel 2	Genotyp 2/2	Genotyp 1/1

Der Odds Ratio sagt somit in erster Linie etwas über Chancenerhöhung oder -erniedrigung aus, ein bestimmtes Allel oder einen bestimmten Genotyp bei einem Patienten vorzufinden. Weicht der OR signifikant von 1 ab, in dem das 95%-Konfidenzintervall (KI) den OR von 1 nicht einschließt, kann man von einer Assoziation des untersuchten Markers mit der Krankheit sprechen.

Eine weitere Auswertemethode der CC-Daten stellte die χ^2 -Statistik nach Pearson dar. Hierbei erfolgte die Transformation der Allel- bzw. Genotypenanzahl für die Fall- bzw. Kontrollgruppe in eine Kontingenztabelle (Tab. 2.16, Tab. 2.17).

Tab. 2.16: 2x3-Felder-Kontingenztabelle der Genotypen gegen Fälle bzw. Kontrollen.

	Genotyp			Σ
	1 / 1	1 / 2	2 / 2	
Fälle	n_{11}	n_{12}	n_{22}	n
Kontrollen	m_{11}	m_{12}	m_{22}	m

Tab. 2.17: 2x2-Felder-Kontingenztabelle der Allele gegen Fälle bzw. Kontrollen.

	Allel		Σ
	1	2	
Fälle	$2n_{11} + n_{12}$	$n_{12} + 2n_{22}$	$2n$
Kontrollen	$2m_{11} + m_{12}$	$m_{12} + 2m_{22}$	$2m$

Anschließend wurde der χ^2 -Test nach Pearson angewendet [Ott, 1985], welcher die Unabhängigkeit zweier Merkmale überprüft, indem der Unterschied zwischen dem Beobachtungswert und dem Erwartungswert bei Merkmalsunabhängigkeit bestimmt wird. In der vorliegenden CC-Studie sind diese beiden Merkmale der Erkrankungsstatus (Patient vs. Kontrolle) und der biallelische Marker (Allel 1 vs. Allel 2) unter der Nullhypothese, dass die Allelfrequenzen in der Patientengruppe denen der Kontrollgruppe entsprechen. Für den allelischen χ^2 -Wert ergibt sich folgende Berechnung basierend auf einer 2x2-Felder-Kontingenztabelle (Tab. 2.17) mit einem Freiheitsgrad:

$$\chi^2 = \frac{((2n_{11} + n_{12}) \cdot (m_{12} + 2m_{22}) - (n_{12} + 2n_{22}) \cdot (2m_{11} + m_{12}))^2 \cdot 2 \cdot (n + m)}{(2n_{11} + n_{12} + 2m_{11} + m_{12}) \cdot (n_{12} + 2n_{22} + m_{12} + 2m_{22}) \cdot 2n \cdot 2m}$$

Die Signifikanz der Merkmalsabhängigkeit bzw. der Assoziation des Markers mit der Krankheit wurde basierend auf der χ^2 -Verteilung in Form des abgeleiteten p-Wertes für das 95%-Konfidenzintervall dargestellt. Mit Hilfe von Permutationstests im Rahmen von 1.000 bis 10.000 Permutationen pro Marker wurde die Validität der errechneten p-Werte überprüft. Analog zum allelischen erfolgte die Berechnung des genotypischen χ^2 -Wertes unter Verwendung der 2x3-Felder-Kontingenztafel (Tab. 2.16) bei einem Freiheitsgrad von 2. Es galt zu berücksichtigen, dass der genotypische χ^2 -Wert bei Berechnungen unter dominanten oder rezessiven Krankheitsmodellen (Freiheitsgrad=1) von der modellfreien, ungewichteten Analyse abweichen konnte. Desweiteren konnte die χ^2 -Statistik als valider Test nur für Marker angewendet, deren Allele statistisch unabhängig im Individuum vorlagen, was einer Genotypverteilung innerhalb des Hardy-Weinberg-Gleichgewichtes (HWG) entsprach. SNPs mit signifikanten Abweichungen vom HWG hätten eine nicht tolerierbare Erhöhung der Typ I Fehlerrate zur Folge gehabt [Sasieni, 1997; Schaid und Jacobsen, 1999].

Die verwendete CC-Analysesoftware setzte sich aus institutsinterner Software und dem Programm haplo_ccmap (Autor: Jochen Hampe, Institut für klinische Molekularbiologie, Universitätsklinikum Schleswig-Holstein, Campus Kiel), welches auf dem Programm HAPMAX (Autor: Michael Krawczak, Institut für Medizinische Informatik und Statistik, Universität Kiel) basiert, zusammen. Letzteres verwendet einen haplotypbasierten CC-Algorithmus, sodass bei der Berechnung von Einzelpunktanalysen Abweichungen zum χ^2 -Test nach Pearson auftreten konnten. Die Visualisierung der Analyseergebnisse erfolgte über die Programme Haploview (Vers. 3.2) [Barrett *et al.*, 2005] und Genedigger (Vers. 2.0) (Autoren: Jochen Hampe, Tim Lu, Stefan Wächter & Andreas Wollstein, Institut für klinische Molekularbiologie, Universitätsklinikum Schleswig-Holstein, Campus Kiel).

3 Ergebnisse

3.1 Feinkartierungsphase I

3.1.1 Definition der Kandidatenmarker

Auf dem Hintergrund des *NOD2*-unabhängigen Kopplungssignals für Chromosom 16q [Annese *et al.*, 2003; Hampe *et al.*, 2002; van Heel *et al.*, 2003] erfolgte in der ersten Phase der Assoziationsstudie die Genotypisierung von 298 SNPs, welche über eine Region von 37,5 Megabasen (Mb) mit einer durchschnittlichen Markerdichte von einem SNP pro 126 Kilobasen (Kb) auf dem langen Arm des Chromosom 16 verteilt waren. Auswahlkriterien für die Marker waren validierte, in-vitro Frequenzangaben öffentlicher Datenbanken (NCBI, Ensembl, The SNP Consortium Ltd.) für mindestens eine kaukasische Population. Zudem mussten sich die Marker auf Chromosom 16q befinden; die chromosomale Zuordnung erfolgte hierbei auf Basis der Datenbankversion hg9 des *Human Genome Project* (HGP), welche der NCBI Version Build 27 entsprach. Physikalisch erstreckte sich der Bereich der Feinkartierung von 48.755 Kb bis 86.283 Kb, sodass sowohl die zu untersuchende Kopplungsregion um den Mikrosatellitenmarker D16S496 als auch die Assoziationsregion um das Gen *NOD2* miteingeschlossen wurden. Die hier verwendete Genotypisierungsgruppe setzte sich aus 181 Familien, 472 Trios und 359 Kontrollen zusammen (siehe 2.1), sodass das erste Genotypisierungsexperiment 2481 Individuen umfasste.

Zur regionalen Aufschlüsselung des Kopplungssignals und Charakterisierung sowie Abgrenzung definierter Kandidatenregionen, welche in der Phase II der Feinkartierung verfolgt werden sollten, wurden Assoziationsanalysen in Form der TDT- und CC-Statistik durchgeführt (siehe 2.7). Eine erste Übersicht hinsichtlich der Anzahl und lokalen Verteilung der statistisch signifikanten Marker lieferte eine graphische Gegenüberstellung der TDT- und CC-Resultate für die Einzelpunktanalyse des Phänotyps CED (Abb. 3.1). Eindeutig gekennzeichnet durch höchstsignifikante Marker trat der *NOD2*-Locus im Bereich von 51,7 Mb hervor. Darüber hinaus erwiesen sich 26 SNPs in mindestens einer der beiden Analyseformen als statistisch auffällig mit einer Signifikanz von $p \leq 0,05$. SNPs in mittelbarer Nähe zu *NOD2* wurden auf Grund eines hohen LD zu den Markern der *NOD2*-Region als *NOD2*-abhängig betrachtet und von der Kandidatenmarkerauslese ausgeschlossen.

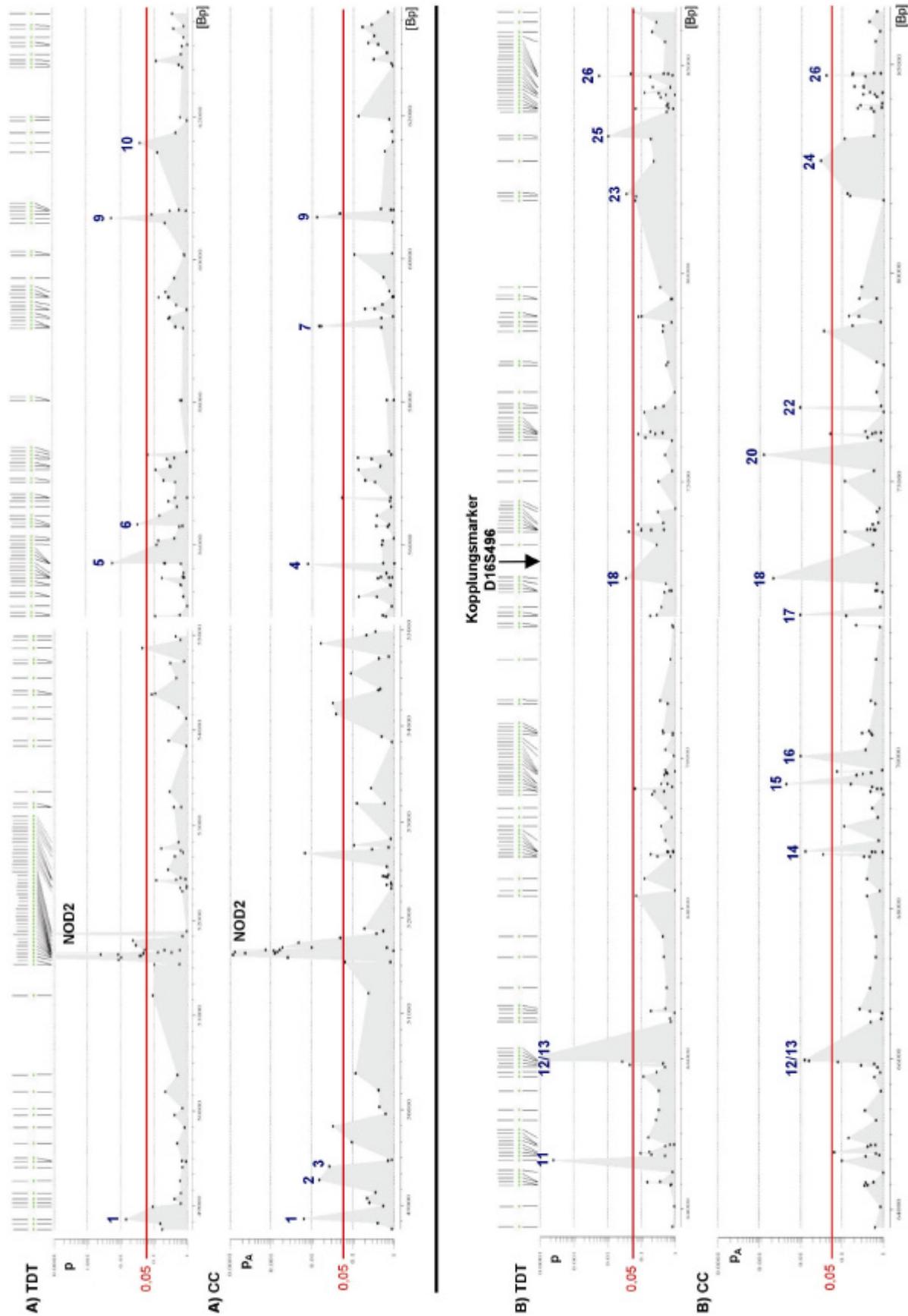


Abb. 3.1: Phase I der SNP Feinkartierung. TDT- und CC-Einzelmarker-Statistik des Phänotyps CED für die Region A (48,8–63,7 Mb) und B (63,7–86,3 Mb) auf Chr16q. [Bp: Basenpaar, p_A : allelbasiertes Signifikanzniveau]

Von den 26 signifikanten Markern (M1–M26) wurden M19 und M21 auf Grund signifikanter Abweichungen vom HWG - bedingt durch eine überdurchschnittliche Mendelfehlerrate - als nicht vertrauenswürdig verworfen. Die restlichen 24 Marker wurden anschließend hinsichtlich ihrer Signalstärke in Kandidatenmarker und Marker geringerer Priorität unterteilt. Hierzu wurde ein SNP als Kandidat für die Feinkartierungsphase II bestimmt, wenn er hochsignifikant in mindestens einer der beiden Statistiken oder signifikant in beiden war. In der Tabelle 3.1 wird die TDT-Statistik nach McNemar [Agresti, 1996] und die CC-Analyse basierend auf Pearson's χ^2 -Test [Ott, 1985] dargestellt. Geringe Abweichungen gegenüber der Darstellung in der Abbildung 3.1, die die TRANSMIT-TDT- und haplo_ccmap-Statistik abbildet, sind auf Unterschiede in den Berechnungsalgorithmen der Auswertemethoden zurückzuführen (siehe 2.7.2 und 2.7.3).

Tab. 3.1: TDT- und CC-Einzelmarker-Analyse für die signifikanten Marker der Feinkartierungsphase I. [MA=selteneres Allel, trans: Anzahl transmittierter MA-Allele, untrans: Anzahl nicht-transmittierter MA-Allele, f_{Kont} : MA-Allelfrequenz in der Kontrollgruppe, f_{Pat} : MA-Allelfrequenz in der Patientengruppe, Chi^2 : Assoziationskoeffizient, p: statistisches Signifikanzniveau, p_A : allelbasiertes Signifikanzniveau, p_G : genotypenbasiertes Signifikanzniveau, OR_A : allelbasierter Odds Ratio für MA, OR_{Het} : heterozygotenbasierter Odds Ratio für MA, OR_{Hom} : homozygotenbasierter Odds Ratio für MA].

Nr.	Marker	Position [Bp]	TDT				CC						
			trans	untrans	Chi^2	p	f_{Kont}	f_{Pat}	p_A	p_G	OR_A	OR_{Het}	OR_{Hom}
M1	rs1345436	48,860,471	93	128	5.54	0.0186	0.13	0.09	0.0043	0.0013	0.65	0.55	1.77
M2	rs1510986	49,263,904	311	315	0.03	0.8730	0.47	0.42	0.0277	0.0754	0.81	0.87	0.65
M3	rs41206	49,408,468	138	154	0.88	0.3491	0.18	0.14	0.0160	0.0435	0.74	0.70	0.63
M4	rs1894941	55,735,869	301	268	1.91	0.1665	0.27	0.34	0.0034	0.0096	1.35	1.31	2.09
M5	rs1420549	55,745,951	47	82	9.50	0.0021	0.05	0.04	0.4704	0.4596	0.85	0.84	-
M6	rs2042447	56,284,382	322	269	4.75	0.0292	0.33	0.34	0.7317	0.5639	1.03	0.94	1.22
M7	rs733463	59,058,663	276	291	0.40	0.5287	0.44	0.38	0.0096	0.0352	0.78	0.75	0.63
M8	rs1672866	59,364,415	204	196	0.16	0.6892	0.18	0.19	0.5034	0.0102	1.08	1.37	0.50
M9	rs729334	60,583,269	268	311	3.19	0.0739	0.38	0.32	0.0076	0.0207	0.77	0.71	0.63
M10	rs1423980	61,639,563	299	247	4.95	0.0261	0.38	0.39	0.8553	0.2029	1.02	1.23	0.92
M11	rs160527	64,647,785	217	149	12.63	0.0004	0.17	0.21	0.0472	0.1020	1.27	1.36	1.29
M12	rs1370235	65,972,680	70	111	9.29	0.0023	0.08	0.06	0.0657	0.0326	0.72	0.79	-
M13	rs1370249	65,986,459	65	106	9.83	0.0017	0.08	0.06	0.0458	0.0270	0.69	0.76	-
M14	rs893198	68,760,003	242	236	0.08	0.7838	0.22	0.26	0.0477	0.0064	1.25	1.56	0.97
M15	rs722175	69,654,746	57	92	8.22	0.0041	0.10	0.06	0.0035	0.0021	0.61	0.72	0.07
M16	rs1080941	70,027,612	304	297	0.08	0.7752	0.31	0.37	0.0053	0.0157	1.32	1.25	1.92
M17	rs1424139	71,802,093	159	148	0.39	0.5301	0.09	0.13	0.0271	0.0884	1.41	1.44	1.59
M18	rs765171	72,683,392	295	250	3.72	0.0539	0.30	0.36	0.0038	0.0007	1.34	1.00	2.37
M20	rs1979717	75,644,292	60	59	0.01	0.9270	0.02	0.05	0.0127	0.0423	1.98	1.96	-
M22	rs1112183	76,777,445	77	83	0.23	0.6353	0.10	0.07	0.0131	0.0410	0.67	0.67	0.35
M23	rs1392487	81,910,425	243	300	5.98	0.0144	0.32	0.29	0.1625	0.3524	0.87	0.83	0.79
M24	rs1297137	82,698,998	156	128	2.76	0.0966	0.09	0.13	0.0114	0.0427	1.49	1.48	2.23
M25	rs2040508	83,299,354	224	180	4.79	0.0286	0.22	0.23	0.5173	0.4228	1.08	1.19	0.91
M26	rs714852	84,746,192	57	92	8.22	0.0041	0.07	0.05	0.0246	0.0468	0.65	0.66	-

0.05 > p > 0.01	0.01 > p > 0.001	0.001 > p
1.3 < OR < 1.5	1.5 < OR < 2	2 < OR
0.77 > OR > 0.67	0.67 > OR > 0.5	0.5 > OR
OR [95% Konfidenzintervall inkl. 1]		

7 Kandidaten-SNPs (K1–K7) (siehe Abb. 3.2) wurden für die Nachverfolgung durch Genotypisierung zusätzlicher, benachbarter SNPs und der damit verbundenen besseren LD- und Haplotyp-Auflösung der einzelnen Markerregionen ausgewählt. Im einzelnen galt die Aufmerksamkeit folgenden, vorläufigen Kandidaten-SNPs:

K1 (M11) – rs160527

In der TDT-Analyse zeigte sich eine höchstsignifikante ($p=0,0004$) Transmission des selteneren Alleles (*minor allele*, MA) ergänzt durch eine moderate Signifikanz ($p=0,0472$) für den allelischen CC-Test. Der genotypische OR für die heterozygote Form untermauerte die Assoziation mit dem Wert von 1,36 [KI 1,02-1,82].

K2 (M12) – rs1370235

Der TDT zeigte eine hochsignifikante Übertragung ($p=0,0023$) des häufigeren Allels (Wildtyp-Allel, Abk.: WA), und die genotypische CC-Analyse erwies sich mit $p=0,0326$ als signifikant. Der allelische OR von 0,72 musste auf Grund des KI von 0,50-1,02 als nicht vertrauenswürdig eingestuft werden.

K3 (M13) – rs1370249

Der Marker K3 befand sich durch eine geringe Distanz von 13,8 Kb in einem hohen LD zu dem SNP K2. Die gleichen Allelfrequenzen des MA ($f_{\text{kont}}(K2, K3)=0,08$) unterstrichen zudem den gemeinsamen Haplotyp von K2 und K3. Dieses kam in der Statistik durch ein ähnlich starkes Assoziationssignal für K3 in Form von $p=0,0017$ für den TDT und $p=0,0270$ für den genotypischen CC-Test zum Ausdruck. K3 zeigte zusätzlich für die allelische CC-Analyse eine moderate Signifikanz von $p=0,0458$ und einen beachtenswerten allelischen OR von 0,69, der jedoch durch ein weites KI von 0,48-0,99 relativiert wurde.

K4 (M15) – rs722175

K4 zeigte in beiden Analyseformen ein hochsignifikantes Assoziationssignal mit einem TDT-basierten p von 0,0041 und einem allelischen wie auch genotypischen CC-basierten p in der Größenordnung von 10^{-3} ($p_A=0,0035$, $p_G=0,0021$). Der allelische OR von 0,61 [KI 0,43-0,85] betonte den Kandidatenmarkerstatus.

K5 (M18) – rs765171

In der TDT-Statistik konnte der SNP K5 mit einem p -Wert von 0,0539 nur als hinweislich betrachtet werden, jedoch zeigte er sich in der CC-Analyse mit einem allelischen p -Wert von 0,0038 und einem p_G von 0,0007 als höchstsignifikanter Assoziationsmarker. Bestätigt wurde dieses mit einem OR_A von 1,34 [KI 1,10-1,64] und einem OR_{Hom} von 2,37 [KI 1,48-3,81].

K6 (M23) – rs1392487

Marker K6 zeigte im TDT eine moderate Assoziation ($p=0,0144$) verbunden mit einer nicht-signifikanten CC-Analyse. Auf Grund seiner deutlichen Signifikanz im 2-Marker-Haplotyp-TDT sowohl mit dem 5'-flankierendem SNP rs2032918 ($p=0,0083$) als auch dem 3'-flankierenden Marker M24 ($p=0,0034$) wurde ihm trotz der mäßigen Einzelmarkerassoziation der Kandidatenstatus zugesprochen.

K7 (M24) – rs1297137

Für den Marker K7 galten ähnliche Bedingungen wie für den SNP K6. Die Einzelmarkeranalyse war schwach signifikant mit einer negativen TDT-Statistik und einem moderaten CC-Ergebnis von $p_A=0,0114$ und $p_G=0,0427$. Der OR_A mit 1,49 [KI 1,09-2,03] und OR_{Het} mit 1,48 [KI 1,05-2,09] unterstrichen das überschaubare CC-Signal. Entscheidend jedoch für die Weiterverfolgung der K6- und K7-Markerregion war das hochsignifikante TDT-Ergebnis für den 2-Marker-Haplotyp mit $p=0,0034$ für den häufigen ($f_{Hap}=0,51$) und $p=0,0006$ für den seltenen Haplotyp ($f_{Hap}=0,08$). Es galt zu beachten, dass die Marker K6 und K7 in einer Entfernung von 788 Kb positioniert waren, sodass eine Lokalisation der beiden SNPs auf verschiedenen Haplotypen durchaus wahrscheinlich war und das Assoziationssignal gegebenenfalls auf einen Kandidatenlocus hinwies, der zwischen den Markern lag.

Insgesamt zeigte sich der in der ersten Genotypisierungsphase generierte Datensatz als sehr robust mit einer Typ I Fehlerrate von ca. 1% für die TDT-Analyse. Eine Permutationsanalyse mit 10.000 *Bootstrap*-Wiederholungen lieferte in 136 von 10.000 Fällen einen höheren χ^2 -Wert als den Maximalwert der Einzelpunktanalyse. In 225 von 10.000 Permutationen wurden genauso viele Ergebnisse mit einem p-Wert $\leq 0,01$ und in 30 Fällen $\leq 0,001$ wie in der zu Grunde liegenden Analyse ermittelt.

Um einen ersten Eindruck der LD-Blockstruktur basierend auf den 298 SNPs zu erlangen, wurde der paarweise D' -Wert zwischen den Markern berechnet und graphisch über eine Farbcodierung der LD-Stärke wiedergegeben (Abb. 3.2). Ein hohes LD wurde liberal mit einem $D'>0,5$ definiert, D' -Werte $\leq 0,5$ wurden graduell von mittlerer über geringer Stärke bis hin zu fehlendem LD dargestellt. Entsprechend der großen chromosomalen Region von 38 Mb und einer durchschnittlichen Markerdichte von 1 SNP/125 Kb war eine Unterteilung der gesamten Region in wenige, wohldefinierte LD-Blöcke noch nicht möglich. Hinzu kam das einige Areale durch eine heterogene Markerverteilung Lücken von bis zu 780 Kb aufwiesen. Somit konnte die haplotypbasierte Analyse zu diesem Zeitpunkt nur bedingt und unter Vorbehalt angewendet werden. Als Konsequenz galt es, in der zweiten Genotypisierungsrunde die Regionen der Kandidatenmarker mit einer höheren SNP-Dichte zu versehen und wesentliche Markerlücken zu schließen.

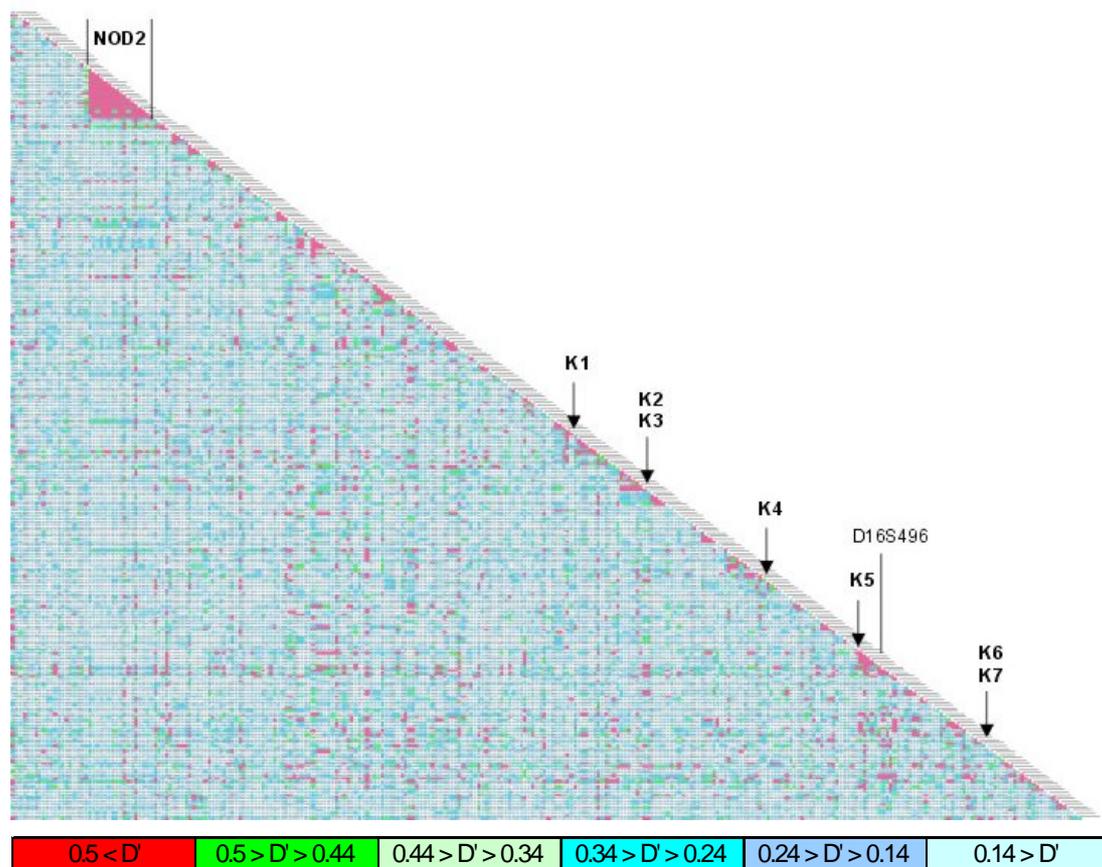


Abb. 3.2: LD-Struktur der 16q-Region basierend auf 298 Feinkartierungsmarkern und 28 SNPs des *NOD2*-Locus. [D': liberaler Kopplungskoeffizient, K: Kandidatenmarker, D16S496: *NOD2*-unabhängiger Kopplungsmarker auf Chr16q, *NOD2*: CED-Risikogen].

3.2 Feinkartierungsphase II

3.2.1 Erhöhung der Markerdichte und Aktualisierung der physikalischen Karte

In der zweiten Phase der Assoziationsstudie wurden 29 SNPs, die vornehmlich in der Nachbarschaft der Kandidatenmarker und in großen Lücken der Markerverteilung platziert waren, genotypisiert. Dabei wurde die Typisierungsgruppe um weitere 179 weibliche Kontrollen zur verbesserten Angleichung des Geschlechterverhältnisses zwischen der Kontroll- und Patientengruppe und zur Erhöhung der statistischen Aussagekraft vergrößert. Zudem wurde die Patientengruppe um 92 CED Trios erweitert, wobei das Verhältnis der Phänotypen MC zu CU von 2:1 beibehalten wurde. Sowohl die 29 zusätzlichen SNPs als auch 298 Basismarker wurden in der gesamten, erweiterten Gruppe von nun 2936

Individuen genotypisiert. Während der statistischen Auswertung des erweiterten Datensatzes wurde die physikalische Karte der chr16q-Region in den öffentlichen Datenbanken durch die Veröffentlichung der Version hg12 maßgeblich verändert. Als Folge der Positionsaktualisierungen veränderte sich die Kandidatenmarkerabfolge und -distanz so entscheidend, dass - ergänzt durch die weitere Analyse der 327 SNPs - die Marker K4, K5 und K7 den Kandidatenstatus verloren. Die gesamte Kartierungsregion verschob sich centromerisch um 20 Mb und vergrößerte sich auf 44 Mb. Die Kandidatenmarkerabfolge lautete demzufolge K1, K2, K3, K6, K7, K5, K4. Im einzelnen stellten sich die Kandidatenmarker einschließlich ihrer benachbarten SNPs in der Studie nun wie folgt dar:

K1 - rs160527

4 zusätzliche benachbarte Marker erhöhten die SNP-Dichte um K1 auf 1 SNP/5 Kb und die TDT-Statistik erwies sich sowohl in der Einzelpunkt- als auch 2- bzw. 3-Marker-Haplotyp-Analyse mit $p=0,0002$, $p=0,0001$ bzw. $p=0,0001$ als äußerst robust und höchstsignifikant. Das moderate CC-Signal der ersten Typisierungsrunde wurde trotz Erhöhung der Individuenzahl beibehalten ($p=0,0419$), was als Signalstabilisierung gewertet wurde.

K2/K3 - rs1370235 / rs1370249

Die Marker K2 und K3 wurden durch ihre enge Kopplung ($D'=0,99$) gemeinsam betrachtet. 2 neue SNPs im 3'-Bereich verringerten die Distanz um 501 Kb auf 5 Kb und das starke TDT-Signal der ersten Analyse wurde für die Einzelmarker mit $p=0,0021$, dem häufigeren 2-Marker-Haplotyp mit $p=0,0005$ und dem häufigeren 3-Marker-Haplotyp mit $p=0,0112$ bestätigt. Zusätzlich konnte nun auch für den 2- und 3-Marker-Haplotyp in der CC-Analyse eine Signifikanz in der Größenordnung von 10^{-2} neben einem Einzelmarkersignal von $p_A=0,0094$ und $0,0062$ beobachtet werden.

K4 - rs722175

Durch die Typisierung von 3 zusätzlichen SNPs in der K4-Umgebung konnte die mittlere Distanz auf 10 Kb reduziert werden. Das hochsignifikante Ergebnis der initialen Genotypisierungsphase konnte in der zweiten mit einer reduzierten CC- und TDT-Signifikanz von $p_A=0,0462$ respektive $p=0,0116$ nicht bestätigt werden. Die 2- und 3-Marker-Haplotyp-Analyse zeigte im TDT lediglich für den selteneren Haplotyp ($f_{\text{Hap}}=0,08$) eine Relevanz mit $p=0,0016$ bzw. $p=0,0010$.

K5 - rs765171

Die Markerlücken von 131 und 803 Kb in der Umgebung von K5 wurden durch 4 zusätzliche SNPs verkleinert. Der TDT erwies sich analog zur ersten Runde als Trend zur Signifikanz ($p=0,0839$ für den Einzelmarker) und die CC-Analyse erbrachte einen allelischen p-Wert von 0,0134. Die Haplotypanalyse erbrachte ausschließlich moderate Signale mit Werten von 10^{-2} , sodass K5 insgesamt an Signalstärke einbüßte.

K6 - rs1392487

5 neue Marker wurden in der 788 Kb großen 3'-Lücke zwischen K6 und K7 genotypisiert. Der Einzelpunkt-TDT erlangte durch die Gruppenvergrößerung eine stärkere Signifikanz mit $p=0,0054$. Auch die allelische CC-Analyse konnte jetzt ein moderates Assoziationssignal mit $p=0,011$ aufweisen. Die TDT-Haplotypanalyse zeigte sich gleichbleibend hochsignifikant mit Werten von 10^{-3} für die 2- und 3-Marker-Haplotypen ($f_{\text{Hap}}=0,30$), und auch in der CC-Statistik erwies sich der 2-Marker-Haplotyp mit $p=0,0092$ nun als signifikant.

K7 - rs1297137

Sowohl im TDT als auch in der CC-Analyse zeigte K7 keine Assoziation mehr für die Einzelmarkeranalyse. Moderate bis mittlere Signifikanzen in der Haplotypanalyse basierten sowohl im TDT als auch in der CC-Statistik entweder auf seltenen oder langen Haplotypen (> 400 Kb) und wurden deshalb vernachlässigt.

Basierend auf diesen Ergebnissen wurde der Fokus nun auf eine etwa 8 Mb große Region, welche in 5'-Richtung durch K1-3 und in 3'-Richtung durch K6 begrenzt war, gelegt. Der Schwerpunkt in dem nachfolgenden Studienabschnitt wurde dabei auf die Validierung der physikalischen Karte für die 8Mb-Region gerichtet. Ebenso kam der Markersaturierung nun eine noch gewichtigere Bedeutung zu, da eine ausgeprägte LD-Blockstruktur die Basis für das Eingrenzen von Kandidatenregionen und -genen war.

3.2.2 Physikalische Kartierung der Kandidatenmarker durch die Fluoreszenz-in-situ-Hybridisierung (FISH)

Zur chromosomalen in-situ Kartierung der Kandidatenmarker K1, K2/K3 und K6 wurde die Methode der Fluoreszenz-in-situ-Hybridisierung (FISH) verwendet (siehe 2.5). Als Markersonden wurden BAC-Klone benutzt, deren DNA-Fragmente die entsprechende Kandidatenmarkersequenz enthielten. Über die BAC-Klon-Datenbank des Deutschen Ressourcenzentrums für Genomforschung (RZPD) wurden die zuvor in öffentlichen Datenbanken (NCBI, Ensembl) ausgewählten BAC-Klone bezogen und vor Ort kultiviert. Vor der Verwendung eines BAC-Klons für die FISH erfolgte die Verifizierung des integrierten DNA-Fragmentes mittels markerspezifischer PCR.

Das Ergebnis der FISH-Experimente bestätigte die in den öffentlichen Datenbanken annotierten relativen Positionen der Marker K1, K2/K3 und K6 zueinander. Die Lage der K1-Sonde (RP11-206F18) relativ zur K6-Sonde (RP11-432I5) stimmte mit der annotierten Distanz von 8,4 Mb überein (Abb. 3.3).

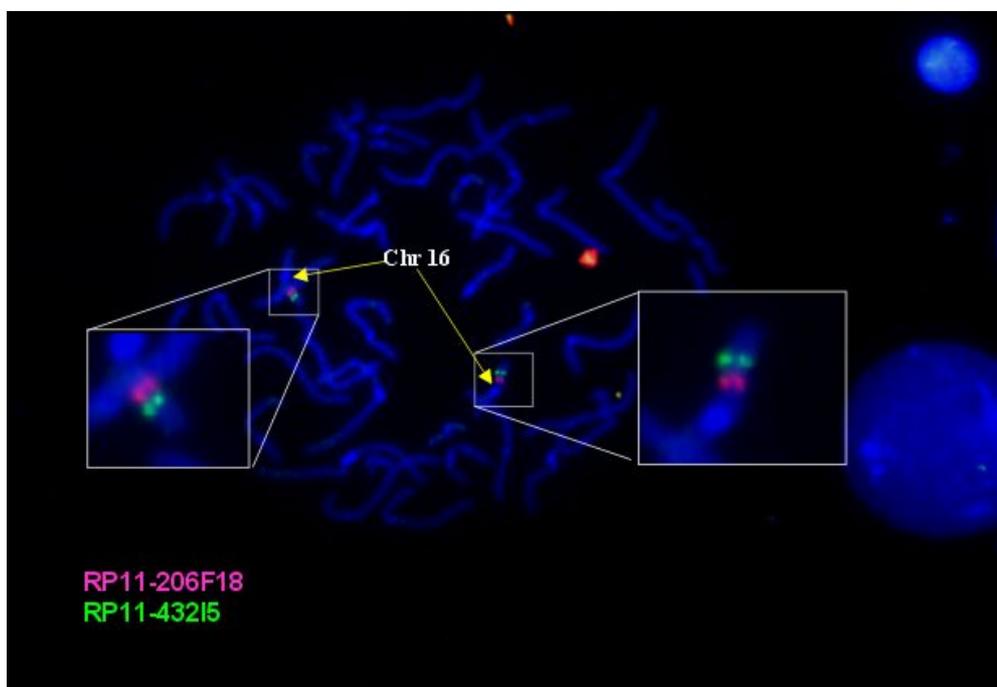


Abb. 3.3: FISH-Kartierung der Kandidatenmarker K1 und K6. [K1-Sonde: RP11-206F18, K6-Sonde: RP11-432I5, Chr16: Chromosom16].

Bei der Hybridisierung der K1-Sonde gegen die K2/K3-Sonde (RP11-562N11) konnte auf Grund der geringen Distanz der Marker von 1,1 Mb die 5'-3'-Ausrichtung und damit die größere Nähe des K1-Markers zum Centromer nur erahnt werden (Abb. 3.4).

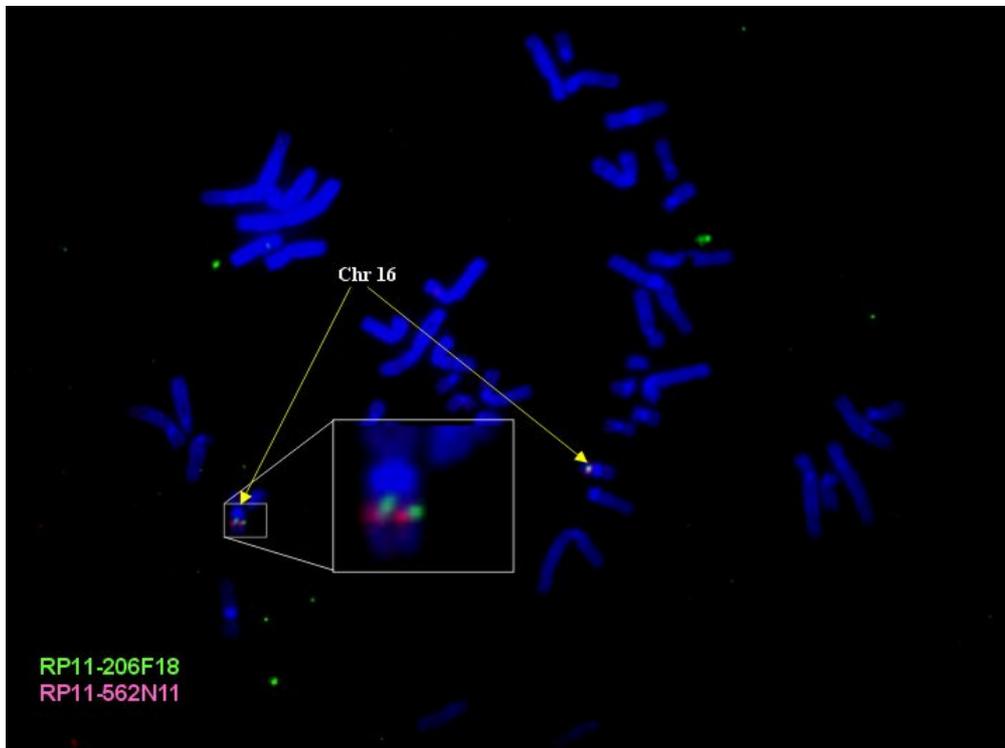


Abb. 3.4: FISH-Kartierung der Kandidatenmarker K1 und K2/K3. [K1-Sonde: RP11-206F18, K2/K3-Sonde: RP11-562N11, Chr16: Chromosom16].

Für jeden der vier Marker wurde die absolute Lage auf dem q-Arm des Chromosom 16 durch die FISH bestätigt. Als weitere Überprüfung der verwendeten Chromosomenkarte für die chr16q-Region wurde ein Datenbankvergleich der NCBI-basierten Positionsangaben mit der *Radiation Hybrid* (RH)-Karte des Stanford Human Genome Center (SHGC) durchgeführt. Es wurden keine Unterschiede festgestellt, sodass die Markerpositionen als vertrauenswürdig eingestuft wurden.

3.2.3 Abgrenzung und Definition der Kandidatenregionen KR 1, KR 2 und KR 3

Mit einer erneuten Aktualisierung der Sequenzdatenbank im Frühjahr 2003 von der Version hg13 (NCBI Build 31) auf die Version hg14 (NCBI Build 32) verschob sich die 8Mb-Kandidatenmarkerregion um 10 Mb in Richtung Telomer des 16q-Armes. Eine

anschließende TDT- und CC-Analyse der 75 SNPs der 8Mb-Region - basierend auf den aktualisierten Positionsangaben - bestätigte die Resultate der vorangegangenen Auswertestatistik. Die Konstruktion von LD-basierten Haplotypen, die die Kandidatenmarker überspannten, war zu diesem Zeitpunkt nicht möglich, da die aktuelle Markerdichte distinkte LD-Blockstrukturen nicht beschreiben konnte (Abb. 3.5).

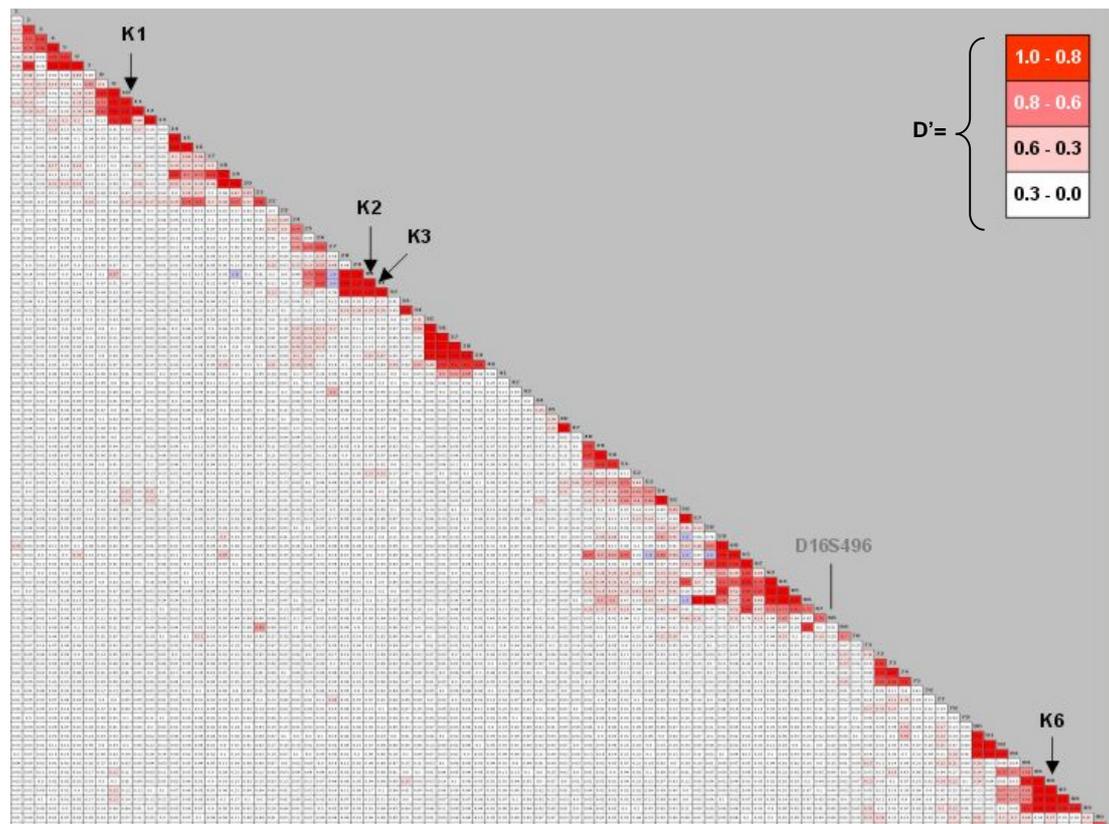


Abb. 3.5: D'-basierter LD-Plott der 8Mb-Kandidatenmarkerregion. Die Grundlage bildeten 75 SNPs aus der Feinkartierungsphase I und II, deren Positionsangaben auf der Datenbankversion hg14 basierten. [D': liberaler Kopplungskoeffizient, K: Kandidatenmarker, D16S496: *NOD2*-unabhängiger Kopplungsmarker auf Chr16q].

Um unter den etwa 80 in der Region befindlichen Genen Kandidaten auf Locusebene zu definieren, wurden 69 zusätzliche SNPs zur höheren Auflösung der LD-Struktur und Charakterisierung von Kandidatenregionen in der bereits verwendeten Genotypisierungsgruppe typisiert. Die Markerdichte erhöhte sich dabei auf 1 SNP/60 Kb und schlug sich in der LD-Karte partiell durch eine moderate LD-Blockbildung nieder. Bedingt durch eine heterogene, eingeschränkte Verfügbarkeit von annotierten und validierten SNPs in den öffentlichen Datenbanken konnten 2 Regionen in der Größenordnung von 1,2 und 2,0 Mb in ihrer LD-Struktur nicht aufgelöst werden (Abb. 3.6).

Die Auswertung der Genotypdaten einschließlich der nach MC und UC differenzierten Analyse für alle 144 Marker resultierte in der Abgrenzung der drei Kandidatenregionen KR1, KR2 und KR3:

KR1

Die Grundlage der Kandidatenregion 1 bildeten die Marker K2 und K3 mit ihrer guten Einzelpunktsignifikanz im TDT für den Phänotyp CED von $p=0,002$ und der starken Assoziation von $p=0,0005$ für den häufigeren 2-Marker-Haplotyp. Für die Subphänotypen MC und CU konnte ebenfalls ein signifikantes Ergebnis mit p -Werten in der Größenordnung von 10^{-2} für beide SNPs ermittelt werden. Die Suche nach geeigneten Kandidatengen in dieser Region zeigte auf, dass sich in einem Umkreis von 3 Mb um K2 und K3 kein validiertes oder hypothetisches Gen in den Datenbankannotationen der Version hg16 (NCBI Build 34) befand. Diese ‚Genwüste‘ wurde im 5'-Bereich durch das Gen *CDH8* (Cadherin 8) und im 3'-Bereich durch *CDH11* (Cadherin 11) begrenzt. Die darauffolgende Feinkartierung und Analyse der beiden Gene mit jeweils 8 Markern wies durch eine negative TDT- und CC-Statistik keinerlei Bedeutung für KR1 auf. Die Datenbankrecherche nach homologen Mausgenen, regulatorischen Elementen oder ESTs (*expressed sequence tag*) als Hinweis auf Gentranskripte verlief ebenfalls erfolglos. Als dritter und letzter Abschnitt der Signalverfolgung für KR1 wurde die LD-Struktur um K2 und K3 näher betrachtet. Es konnte beobachtet werden, dass diese beiden Marker zusätzlich zu ihrer regionalen Kopplung auch ein überregionales LD-Muster aufwiesen (Abb. 3.6). Insgesamt handelte es sich dabei um 9 SNPs, welche sich in einem hohen LD ($D' > 0,5$) mit K2 und K3 befanden. Die Einzelmarker- und Haplotypanalyse für alle 9 SNPs inklusive der benachbarten Marker zeigte sich statistisch unauffällig, sodass KR1 als Kandidatenregion für die Studie keine Relevanz mehr besaß.

KR2

4 Marker in einem Bereich von 29 Kb zeigten in der Einzelmarker-TDT-Analyse für den Phänotyp CED signifikante Werte von $p=0,0370$ (rs1728786) bis $p=0,0007$ (rs1728773). Für den Phänotyp CU verstärkte sich das Signal mit Werten von $p=0,0090$ (rs1103286) bis $p=0,0003$ (rs1728773). Die CC-Analyse verlief für die einzelnen Marker negativ, zeigte jedoch für den 3-Marker-Haplotyp (rs1728786-rs1103286-rs1728787) mit $p=0,0046$ basierend auf einer guten LD-Auflösung für KR2 ein beachtenswertes Signal. Da die LD-Struktur sich über KR2 hinaus bis in die benachbarten Regionen ohne nennenswerte Lücken präsentierte und das Assoziationssignal auf KR2 beschränkt war, wurde gezielt mittels Datenbankrecherche nach Kandidatengen innerhalb der Kandidatenregion 2 gesucht. Als einziges Gen wurde *ZFP90* (Zinkfingerprotein 90) identifiziert. Die Assoziationsmarker

rs1728773 und rs1728786 konnten im 5'-Bereich des Genlocus und rs1103286 und rs1728787 im eigentlichen Gen lokalisiert werden. Somit war *ZFP90* als Kandidatengen für die Feinkartierungsphase III und Mutationsdetektion qualifiziert.

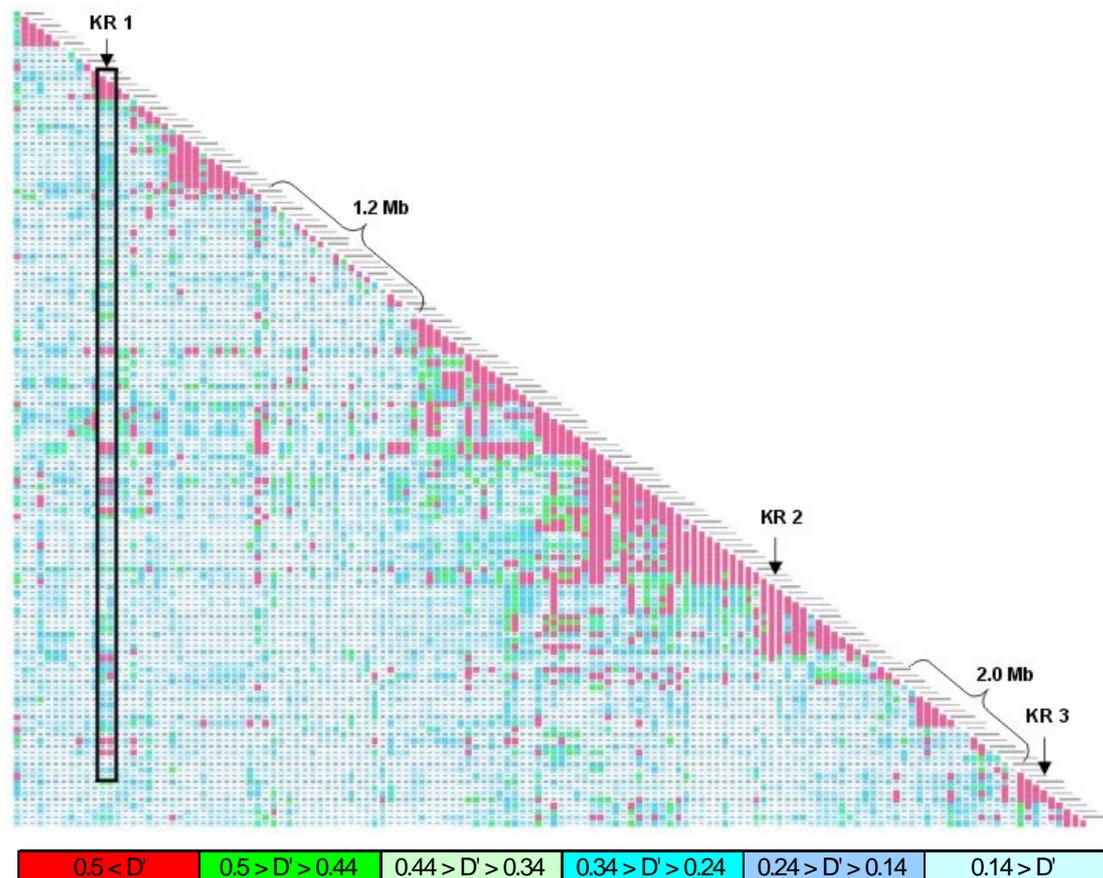


Abb. 3.6: LD-Struktur der 8Mb-Region basierend auf 144 Markern der Feinkartierungsphase I und II. Neben der deutlichen Ausbildung von Blockstrukturen weisen 2 Regionen mit 1,2 und 2,0 Mb Lücken im LD-Muster auf Grund einer z.T. heterogenen Markerverteilung auf. Die Kandidatenregion 1 zeigt neben einem hohen lokalen LD auch eine überregionale Kopplung mit 9 SNPs unterschiedlichster Entfernung. [D' : liberaler Kopplungskoeffizient, KR: Kandidatenregion, Mb: Megabase].

KR3

Eine 180 Kb große Region, welche sich in einer Entfernung von 281 Kb im 5'-Bereich des Kandidatenmarkers K6 befand, zeigte für 6 SNPs konsistente Assoziationssignale sowohl in der TDT- als auch CC-Analyse. Für den Phänotyp CED wurden Signifikanzen in der CC-Statistik von $p=0,021$ (rs1424139) bis $p=0,0064$ (rs732318) ermittelt. Der TDT zeigte die stärksten Signale für den Phänotyp CU mit Werten zwischen $p=0,016$ (rs109556) und

$p=0,0061$ (rs1345891). Die gesamte Region (474 Kb) einschließlich des den 3'-Bereich begrenzenden Markers K6 erwies sich als genreiches Areal, das u.a. 2 Zinkfingerogene (*ZNF19*, *ZNF23*) und ein Gen, das das kalzium-bindende Protein 2 (*CALB2*) kodierte, beinhaltete. Um die Eingrenzung der potentiellen Kandidatengene zu ermöglichen, wurden 13 zusätzliche, genbasierte SNPs typisiert. Hierbei zeigte sich in der Analyse, dass sowohl die Marker des Gens *CALB2* als auch *CHST4* (Carbohydrat-Sulfotransferase 4) moderate bis nicht-signifikante Resultate in der Auswertung erbrachten. Lediglich die SNPs der benachbarten Zinkfingerogene *ZNF19* und *ZNF23* zeigten sowohl in der TDT- als auch CC-Statistik eindeutige Assoziationen. Im TDT waren 4 von 7 SNPs für den Phänotyp MC mit Werten von $p=0,0438$ (rs2288489) bis $p=0,0063$ (rs2070832) signifikant. Auch in der CC-Analyse konnten die höchsten Signifikanzen mit $p=0,0017$ (rs732318) und $p=0,0046$ (rs2288489) für den Phänotyp MC ermittelt werden. Die LD-Struktur der Kandidatenregion 3 zeigte sich bei liberaler Abschätzung über die D' -Berechnung als undifferenzierte, stark gekoppelte Einheit (Abb. 3.7 A). Eine genbasierte LD-Blockstruktur als Grundlage der Kandidatengenabgrenzung war somit über die D' -Kalkulation nicht auflösbar. Erst die konservativere LD-Berechnung über die Bestimmung des r^2 ermöglichte es, Substrukturen zwischen den eng benachbarten Genloci aufzudecken (Abb. 3.7 B). *CALB2* konnte hierdurch aus dem LD-Block der beiden Zinkfingerogene ausgeschlossen werden. Im 3'-Bereich bildete der mit einem Marker schwach assoziierte *CHST4*-Locus ($p=10^{-2}$) die LD-Blockbegrenzung. Lediglich K6 zeigte außerhalb des Blocks ein hohes LD mit den Markern des *ZNF23*. Somit kristallisierten sich die beiden Zinkfingerogene *ZNF19* und *ZNF23* - sowohl über die Assoziationssignale als auch über die LD-Struktur der KR3 - als Kandidatengene heraus.

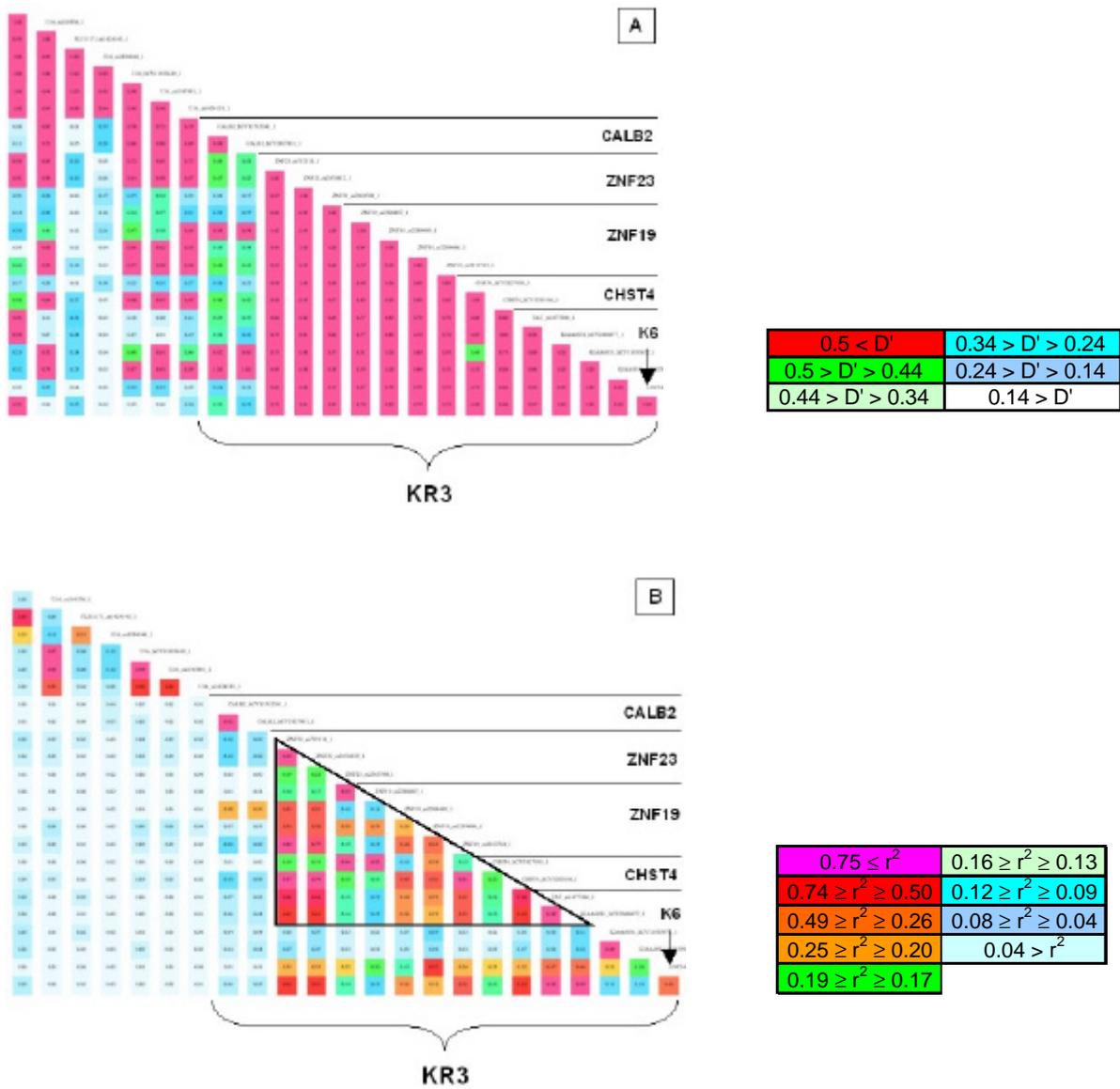


Abb. 3.7: Vergleich der LD-Struktur der Kandidatenregion 3 durch eine liberale und konservative LD-Abschätzung. A) D' -Statistik, B) r^2 -Statistik. [KR: Kandidatenregion, D' : liberaler Kopplungskoeffizient, r^2 : konservativer Kopplungskoeffizient].

3.3 Feinkartierungsphase III und Auflösung der Kandidatenregionen in Kandidatengene

In der dritten und letzten Phase der Studie wurde der Focus auf die in der Feinkartierungsphase II isolierten Kandidatengene *ZFP90*, *ZNF19* und *ZNF23* gelegt. Die Typisierung zusätzlicher, locusbasierter Marker für die einzelnen Gene zielte auf die Auflösung LD-basierter Haplotypblöcke zur Eingrenzung und Identifizierung von krankheitsrelevanten Mutationen. Desweiteren wurden die proteinkodierenden und transkriptionsregulierenden Genregionen (Exone, Promoter, *splice-sites*) in 47 Patienten-DNA-Proben der Genotypisierungsgruppe sequenziert. Das Phänotyp-Verhältnis war hierbei für MC zu CU mit 24 zu 23 Patienten ausgewogen. Über einen Sequenzvergleich der Probensequenzen mit der jeweiligen Referenzsequenz aus den öffentlichen Datenbanken wurden die relevanten Genbereiche auf Mutationen untersucht. Die aus der Mutationsdetektion resultierenden, validierten bzw. detektierten Polymorphismen wurden im Anschluss in der kompletten Genotypisierungspopulation (Gruppe A) der ersten beiden Feinkartierungsphasen typisiert. Zur statistischen Validierung der Assoziationssignale wurden alle Marker der Feinkartierungsphase III zusätzlich in einer unabhängigen Replikationspopulation (Gruppe B) genotypisiert. Diese bestand aus 726 Trios und 534 Einzelpatienten mit einem Phänotypverhältnis von 1:1 für MC zu CU. 1040 populationsbasierte Kontrollen gehörten ebenfalls zur Gruppe B, sodass insgesamt 3752 Individuen für eine Signalreplikation zur Verfügung standen (siehe 2.1).

Zwei weitere Gene, die außerhalb der Kandidatenregionen KR1-3 lokalisiert waren, wurden ebenfalls als Kandidatengene feinkartiert und sequenziert. Dieses war zum einen *GPR56* (G-Protein gekoppelter Rezeptor 56), das mit dem signifikanten Marker M7 (Tab. 3.1) in der CC-Statistik der ersten Phase auffällig war, jedoch außerhalb der Kandidatenregionen der zweiten Phase lag. Mit der Publikation eines *NOD2*-unabhängigen Kopplungssignals auf Chromosome 16q für den Phänotyp Colitis ulcerosa [Annese *et al.*, 2003] bekam die M7-Region jedoch eine höhere Priorität, da der signalbestimmende Kopplungsmarker D16S408 im 1,3 Mb entfernten 5'-Bereich von *GPR56* lokalisiert ist. Das fünfte Kandidatengen stellte *BRD7* (Bromodomänen Protein 7) dar. Bedingt durch die Aktualisierung der Chromosomenkarten des humanen Genoms durch die Veröffentlichung der Datenbankversion hg17 (NCBI Build 35) im Mai 2004 wurde der komplette Markerbestand der Studie hinsichtlich der Positionsangaben und Genannotation aktualisiert und re-analysiert. Die beiden Marker hCV183346 und hCV148486 wiesen nun mit signifikanten Ergebnissen sowohl in der TDT-Statistik ($p=0,0187$, $p=0,0033$) als auch CC-

Analyse ($p=0,0199$, $p=0,0044$) und ihrer intragenischen Lokalisation auf *BRD7* als vielversprechenden Kandidaten hin.

Die zusammenfassende Auswertung der Mutationsdetektion, Feinkartierung und Replikationsgenotypisierung für die einzelnen Kandidatengene wird nachfolgend für jede Genregion dargestellt.

3.3.1 *ZFP90 (Zinkfingerprotein 90)*

Der Genlocus *ZFP90* erstreckte sich über 28 Kb, beinhaltete 5 Exone und codierte über ein 4,7 Kb großes Transkript ein Protein von 468 Aminosäuren. Über die Mutationsdetektion wurden 4 SNPs (ZF4, ZF8, ZF13, ZF16), welche in der 5'-Umgebung bzw. im intronischen 5'-UTR-Bereich lokalisiert waren, detektiert. Desweiteren konnten die öffentlich annotierten SNPs ZF3, ZF5, ZF6, ZF9 und ZF10 validiert werden. Diese befanden sich ebenfalls in der nicht-translatierten Genregion. Aus den Feinkartierungsphasen I bis III befanden sich 11 Marker aus den öffentlichen Datenbanken in der *ZFP90*-Region. Somit wurden insgesamt 20 SNPs für den Locus *ZFP90* in der Genotypisierungsgruppe A und Replikationsgruppe B typisiert und über die TDT- und CC-Statistik analysiert.

Die TDT-Analyse der Gruppe A (Tab. 3.2) zeigte für 13 Marker signifikante Ergebnisse im Rahmen von $p=0,0196$ bis $p=0,0011$ für den Phänotyp CED. 10 dieser SNPs waren auch für den Phänotyp CU mit p -Werten zwischen 0,0480 und 0,0056 assoziiert. Für den Phänotyp MC konnte dagegen nur für die Marker rs1728773 und rs1103286 ein moderates Signal von $p=0,0291$ und $p=0,0256$ ermittelt werden. Der für einen aminosäureaustauschkodierende SNP rs1177648 (ZF18) erzielte mit $p=0,0523$ für CED und $p=0,0519$ für CU ein hinweisendes, aber nicht signifikantes Resultat. In der Gruppe B zeigte keiner der 20 Marker eine statistische Assoziation mit einem der drei Phänotypen. Die CC-Statistik erwies sich mit Ausnahme der Marker ZF3 und ZF4 über beide Typisierungsgruppen als nicht signifikant. Die geringe Allelfrequenz von $f_{\text{Kont}}=0,006$ und $f_{\text{Kont}}=0,029$ für die beiden schwach signifikanten SNPs verwies dabei die positive CC-Statistik in den Bereich des statistischen Typ I Fehlers.

Die Analyse der LD- und Haplotypstruktur (Abb. 3.8) zeigte über die gesamte Genregion ein konsistentes Kopplungsungleichgewicht der Marker ZF2 bis ZF20. Unterstrichen wurde dieses durch einen gen-überspannenden Haplotypblock von 19 Markern, der sich in 5 Haplotypen mit Häufigkeiten von 6%–45% aufteilte. Der zweithäufigste Haplotyp mit 21% konnte dabei als der in der TDT-Statistik assoziierte identifiziert werden. Das seltenere Allel der 11 Marker ZF2, ZF5, ZF8-12, ZF14, ZF17, ZF19 und ZF20 zeigte durchgehend eine Häufigkeit von etwa 21% und war für alle 11 SNPs ausschließlich auf dem

zweithäufigsten Haplotyp zu finden (Abb. 3.8). Der 5-Marker-Haplotyp der SNPs ZF8-ZF12 unterstrich diesen Befund mit einer Signifikanz von $p=0,00035$ in der TDT-Analyse für den Phänotyp CED.

Tab. 3.2: TDT-Statistik ZFP90. [MA=selteneres Allel, Dist: Markerdistanz, f_{kont} : MA-Allelfrequenz in der Kontrollgruppe, trans: Anzahl transmittierter MA-Allele, untrans: Anzahl nicht-transmittierter MA-Allele, Chi^2 : Assoziationskoeffizient, p: statistisches Signifikanzniveau].

Nr.	Marker	Dist [Bp]	f_{kont}	Gruppe A											
				CED				MC				CU			
				trans	untrans	Chi^2	p	trans	untrans	Chi^2	p	trans	untrans	Chi^2	p
ZF1	rs696587		0.286	320	323	0.01	0.9058	198	180	0.86	0.3545	122	143	1.66	0.1970
ZF2	rs1728773	17,172	0.212	239	315	10.43	0.0012	148	188	4.76	0.0291	91	127	5.94	0.0148
ZF3	rs7193495	363	0.006	22	9	5.45	0.0196	11	5	2.25	0.1336	11	4	3.27	0.0707
ZF4	1pe1zfp90	88	0.029	27	50	6.87	0.0088	17	27	2.27	0.1317	10	23	5.12	0.0236
ZF5	rs1728770	29	0.213	245	293	4.28	0.0385	154	171	0.89	0.3457	91	122	4.51	0.0337
ZF6	rs1728779	3,376	0.448	389	366	0.70	0.4026	226	210	0.59	0.4435	163	156	0.15	0.6951
ZF7	rs1728786	2,206	0.363	350	409	4.59	0.0322	215	243	1.71	0.1908	135	166	3.19	0.0740
ZF8	16ZFP90_01	3,358	0.218	246	304	6.12	0.0134	153	182	2.51	0.1131	93	122	3.91	0.0480
ZF9	rs1170444	223	0.216	245	304	6.34	0.0118	154	175	1.34	0.2470	91	129	6.56	0.0104
ZF10	rs1170443	8	0.215	243	311	8.35	0.0039	148	183	3.70	0.0544	95	128	4.88	0.0271
ZF11	rs1182968	650	0.214	245	303	6.14	0.0132	152	181	2.53	0.1120	93	122	3.91	0.0480
ZF12	rs1170442	119	0.217	247	292	3.76	0.0526	156	170	0.60	0.4381	91	122	4.51	0.0337
ZF13	16ZFP90_23	32	0.144	197	194	0.02	0.8794	120	116	0.07	0.7946	77	78	0.01	0.9360
ZF14	rs1103286	4,529	0.216	255	321	7.56	0.0060	156	198	4.98	0.0256	99	123	2.59	0.1072
ZF15	hCV11674972	4,236	0.363	354	399	2.69	0.1010	214	237	1.17	0.2788	140	162	1.60	0.2055
ZF16	16ZFP90_31	9,471	0.057	79	64	1.57	0.2097	39	33	0.50	0.4795	40	31	1.14	0.2855
ZF17	rs1728787	710	0.208	243	312	8.58	0.0034	156	191	3.53	0.0603	87	121	5.56	0.0184
ZF18	rs1177648	4,965	0.365	333	385	3.77	0.0523	197	215	0.79	0.3752	136	170	3.78	0.0519
ZF19	hCV1964457	1,661	0.209	240	317	10.64	0.0011	151	187	3.83	0.0502	89	130	7.68	0.0056
ZF20	rs3203684	772	0.209	235	293	6.37	0.0116	151	174	1.63	0.2020	84	119	6.03	0.0140

0.05 > p > 0.01 **0.01 > p > 0.001** **0.001 > p**

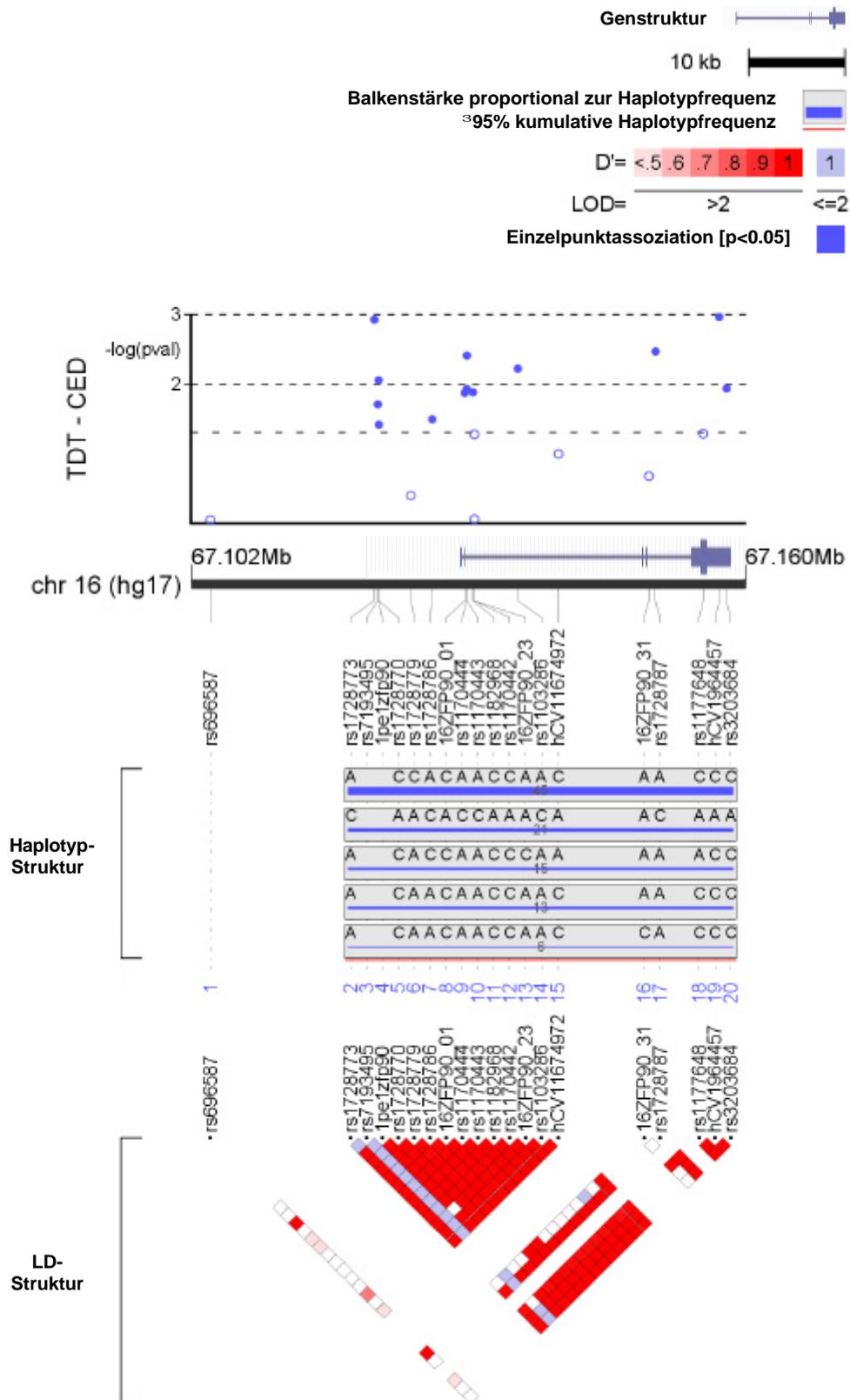


Abb. 3.8: Assoziationskartierung, Markerverteilung, Haplotyp- und LD-Struktur für ZFP90. [pval=p-Wert, D': liberaler Kopplungskoeffizient, LD: Kopplungsungleichgewicht, LOD: Kopplungswahrscheinlichkeit, Mb: Megabase, p: statistisches Signifikanzniveau]

3.3.2 ZNF19 und ZNF23 (Zinkfingerproteine 19 und 23)

Die Genloci *ZNF19* und *ZNF23* lagen mit einem Abstand von 12 Kb in unmittelbarer Nähe zueinander und wurden deshalb gemeinsam betrachtet. Beide Genregionen umfassten jeweils ca. 15 Kb und pro Gen waren jeweils 6 Exone in der Genstruktur beschrieben. Das mRNA-Transkript des *ZNF19* Gens belief sich auf 2,6 Kb und codierte ein 458 Aminosäuren großes Protein. Für *ZNF23* waren es analog 3,2 Kb und 643 Aminosäuren. Die Sequenzierung der 12 Exone und der Promoterregionen lieferte 6 bis dahin unbeschriebene und 5 bereits annotierte SNPs. Für *ZNF23* konnten 3 intronische SNPs (ZN5, ZN7, ZN8), ein codierender SNP (ZN2) im Exon 6 und ein nicht-codierender SNP (ZN4) im Exon 4 ermittelt werden. Die Mutationsdetektion für *ZNF19* erbrachte 5 exonische SNPs, von denen einer (ZN16) einen Aminosäureaustausch codierte, 3 (ZN10-12) in dem 3'-UTR-Bereich lagen und einer (ZN14) als ‚stille‘, nicht kodierende Mutation beschrieben war. Ein weiterer Marker (ZN19) konnte in der 5'-Region von *ZNF19* identifiziert werden. Insgesamt wurden in der Region um *ZNF19* und *ZNF23* 19 SNPs (ZN1-19) in der Genotypisierungsgruppe A und B typisiert.

Die CC-Analyse (Tab. 3.4) für CED zeigte für die 5 Marker ZN1, ZN3, ZN7, ZN15 und ZN18 signifikante Ergebnisse für den allelischen χ^2 in einem Bereich von $p=0,0387$ (ZN3) bis $p=0,0042$ (ZN15). Der signifikanteste, allelische OR wurde für rs2288489 (ZN15) mit 0,74 [KI 0,61-0,91] ermittelt. 4 dieser 5 SNPs zeigten eine allelische Assoziation mit dem Phänotyp MC im Signifikanzbereich $p=0,0398$ bis $p=0,0094$. Für den Phänotyp CU erzielte lediglich der für CED und MC signifikante Marker rs2288489 (ZN15) ein schwaches, nicht signifikantes Assoziationssignal von $p_A=0,0864$. Die CC-Analyse der Genotypdaten der Replikationsgruppe B zeigte für keinen der 19 SNPs eine Signifikanz. In der TDT-Statistik (Tab. 3.5) wurden signifikante Ergebnisse ausschließlich für den Phänotyp MC in der Gruppe A ermittelt. Die 5 TDT-assoziierten Marker rs732318, rs2070832, rs8051212, rs2288489 und rs3813743 waren die selben Marker, die sich in der CC-Analyse für CED als signifikant ausgewiesen hatten. Die TDT-Signalstärke lag in einem Signifikanzbereich zwischen $p=0,0395$ (Zn15) und $p=0,0075$ (ZN3).

Die LD-Struktur für die Gene *ZNF19* und *ZNF23* (Abb. 3.9) basierend auf den 19 Markern stellte sich als homogen mit einem hohen LD zwischen den einzelnen SNPs dar. Abgrenzende Blockstrukturen für die einzelnen Gene oder Teilbereiche dieser konnten nicht detektiert werden. Ebenso war es nicht möglich, die assoziierten SNPs einem gemeinsamen Haplotyp zuzuordnen. Somit verblieben die Assoziationssignale für den Phänotyp MC auf regionalem, 2 Gene umschließendem Niveau. Der codierende SNP rs2070832 (ZN3) konnte als einziger sowohl eine funktionelle Relevanz als auch eine phänotyp-assoziierte Signifikanz in der TDT- und CC-Statistik aufweisen.

Tab. 3.4: CC-Statistik ZNF19 und ZNF23. [MA=selteneres Allel, Dist: Markerdistanz, f_{Kont} : MA-Allelfrequenz in der Kontrollgruppe, f_{Pat} : MA-Allelfrequenz in der Patientengruppe, p_A : allelbasiertes Signifikanzniveau, p_G : genotypenbasiertes Signifikanzniveau, OR_A : allelbasierter Odds Ratio für MA, OR_{Het} : heterozygotenbasierter Odds Ratio für MA, OR_{Hom} : homozygotenbasierter Odds Ratio für MA].

Locus	Nr.	Marker	Dist [Bp]	Gruppe A							
				f_{Kont}	f_{Pat}	p_A	p_G	OR_A	OR_{Het}	OR_{Hom}	
CED	U16	ZN1	rs732318	0.321	0.280	0.0229	0.0603	0.82	0.78	0.73	
	ZNF23	ZN2	1e6zfn23	3,605	0.155	0.162	0.6405	0.2472	1.05	1.18	0.73
	ZNF23	ZN3	rs2070832	3,709	0.312	0.274	0.0387	0.1029	0.84	0.80	0.73
	ZNF23	ZN4	rs11641967	746	0.151	0.164	0.3817	0.5134	1.10	1.16	0.97
	ZNF23	ZN5	1e4zfn23	223	0.142	0.139	0.8558	0.5887	0.98	1.03	0.58
	ZNF23	ZN6	rs1609869	2,041	0.289	0.295	0.7180	0.7630	1.03	1.09	0.99
	ZNF23	ZN7	rs8051212	5,308	0.326	0.279	0.0119	0.0376	0.80	0.77	0.68
	ZNF23	ZN8	rs8051439	269	0.158	0.165	0.6406	0.3347	1.05	1.16	0.77
	ZNF23	ZN9	rs2303788	2,125	0.290	0.296	0.7467	0.7965	1.03	1.08	0.99
	ZNF19	ZN10	3e6zfn19	10,311	0.141	0.137	0.7577	0.7235	0.97	1.00	0.67
	ZNF19	ZN11	2e6zfn19	272	0.158	0.163	0.7592	0.6527	1.03	1.10	0.85
	ZNF19	ZN12	1e6zfn19	142	0.295	0.305	0.5842	0.8509	1.05	1.06	1.08
	ZNF19	ZN13	rs2288487	945	0.296	0.306	0.5689	0.6676	1.05	1.11	1.03
	ZNF19	ZN14	rs2288488	46	0.300	0.307	0.6877	0.8606	1.04	1.07	1.03
	ZNF19	ZN15	rs2288489	51	0.196	0.153	0.0042	0.0131	0.74	0.71	0.65
	ZNF19	ZN16	rs8050871	111	0.348	0.317	0.0986	0.2393	0.87	0.84	0.79
	ZNF19	ZN17	rs2288486	3,238	0.483	0.447	0.0709	0.1905	0.87	0.84	0.76
	ZNF19	ZN18	rs3813743	10,017	0.311	0.272	0.0301	0.0923	0.83	0.82	0.68
	ZNF19	ZN19	1e1zfn19	210	0.141	0.139	0.8590	0.8692	0.98	1.00	0.76
MC	U16	ZN1	rs732318	0.321	0.275	0.0249	0.0487	0.80	0.73	0.73	
	ZNF23	ZN2	1e6zfn23	3,605	0.155	0.177	0.1844	0.1099	1.17	1.34	0.86
	ZNF23	ZN3	rs2070832	3,709	0.312	0.270	0.0398	0.0814	0.82	0.76	0.75
	ZNF23	ZN4	rs11641967	746	0.151	0.154	0.8670	0.9563	1.02	1.04	0.96
	ZNF23	ZN5	1e4zfn23	223	0.142	0.143	0.9439	0.7106	1.01	1.06	0.64
	ZNF23	ZN6	rs1609869	2,041	0.289	0.283	0.7692	0.9317	0.97	1.00	0.92
	ZNF23	ZN7	rs8051212	5,308	0.326	0.272	0.0094	0.0292	0.77	0.74	0.63
	ZNF23	ZN8	rs8051439	269	0.158	0.179	0.2120	0.0921	1.16	1.33	0.80
	ZNF23	ZN9	rs2303788	2,125	0.290	0.287	0.8555	0.9730	0.98	0.97	0.98
	ZNF19	ZN10	3e6zfn19	10,311	0.141	0.142	0.9470	0.9184	1.01	1.03	0.84
	ZNF19	ZN11	2e6zfn19	272	0.158	0.177	0.2777	0.2407	1.14	1.26	0.88
	ZNF19	ZN12	1e6zfn19	142	0.295	0.289	0.7527	0.9195	0.97	0.95	0.97
	ZNF19	ZN13	rs2288487	945	0.296	0.293	0.8682	0.9615	0.98	1.01	0.95
	ZNF19	ZN14	rs2288488	46	0.300	0.292	0.7102	0.9207	0.96	0.95	0.95
	ZNF19	ZN15	rs2288489	51	0.196	0.152	0.0096	0.0315	0.74	0.72	0.56
	ZNF19	ZN16	rs8050871	111	0.348	0.327	0.3292	0.6082	0.91	0.93	0.81
	ZNF19	ZN17	rs2288486	3,238	0.483	0.453	0.1767	0.3914	0.89	0.86	0.79
	ZNF19	ZN18	rs3813743	10,017	0.311	0.274	0.0705	0.1811	0.84	0.81	0.73
	ZNF19	ZN19	1e1zfn19	210	0.141	0.144	0.8351	0.9571	1.03	1.04	0.96

0.05 > p > 0.01	0.01 > p > 0.001	0.001 > p
1.3 < OR < 1.5	1.5 < OR < 2	2 < OR
0.77 > OR > 0.67	0.67 > OR > 0.5	0.5 > OR
OR [95% Konfidenzintervall inkl. 1]		

Tab. 3.5: TDT-Statistik ZNF19 und ZNF23. [MA=selteneres Allel, Dist: Markerdistanz, trans: Anzahl transmittierter MA-Allele, untrans: Anzahl nicht-transmittierter MA-Allele, Chi²: Assoziationskoeffizient, p: statistisches Signifikanzniveau].

Locus	Nr.	Marker	Dist [Bp]	Gruppe A				
				trans	untrans	Chi ²	p	
U16	ZN1	rs732318		175	227	6.73	0.0095	
ZNF23	ZN2	1e6znf23	3,605	126	121	0.10	0.7504	
ZNF23	ZN3	rs2070832	3,709	163	215	7.15	0.0075	
ZNF23	ZN4	rs11641967	746	118	101	1.32	0.2507	
ZNF23	ZN5	1e4znf23	223	116	115	0.00	0.9475	
ZNF23	ZN6	rs1609869	2,041	177	165	0.42	0.5164	
ZNF23	ZN7	rs8051212	5,308	159	207	6.30	0.0121	
ZNF23	ZN8	rs8051439	269	135	121	0.77	0.3816	
ZNF23	ZN9	rs2303788	2,125	187	168	1.02	0.3133	
MC	ZNF19	ZN10	3e6znf19	10,311	110	113	0.04	0.8408
	ZNF19	ZN11	2e6znf19	272	132	122	0.39	0.5304
	ZNF19	ZN12	1e6znf19	142	180	175	0.07	0.7907
	ZNF19	ZN13	rs2288487	945	187	176	0.33	0.5637
	ZNF19	ZN14	rs2288488	46	181	173	0.18	0.6707
	ZNF19	ZN15	rs2288489	51	112	145	4.24	0.0395
	ZNF19	ZN16	rs8050871	111	187	208	1.12	0.2907
	ZNF19	ZN17	rs2288486	3,238	197	233	3.01	0.0826
	ZNF19	ZN18	rs3813743	10,017	168	211	4.88	0.0272
	ZNF19	ZN19	1e1znf19	210	116	116	0.00	1

0.05 > p > 0.01 **0.01 > p > 0.001** **0.001 > p**

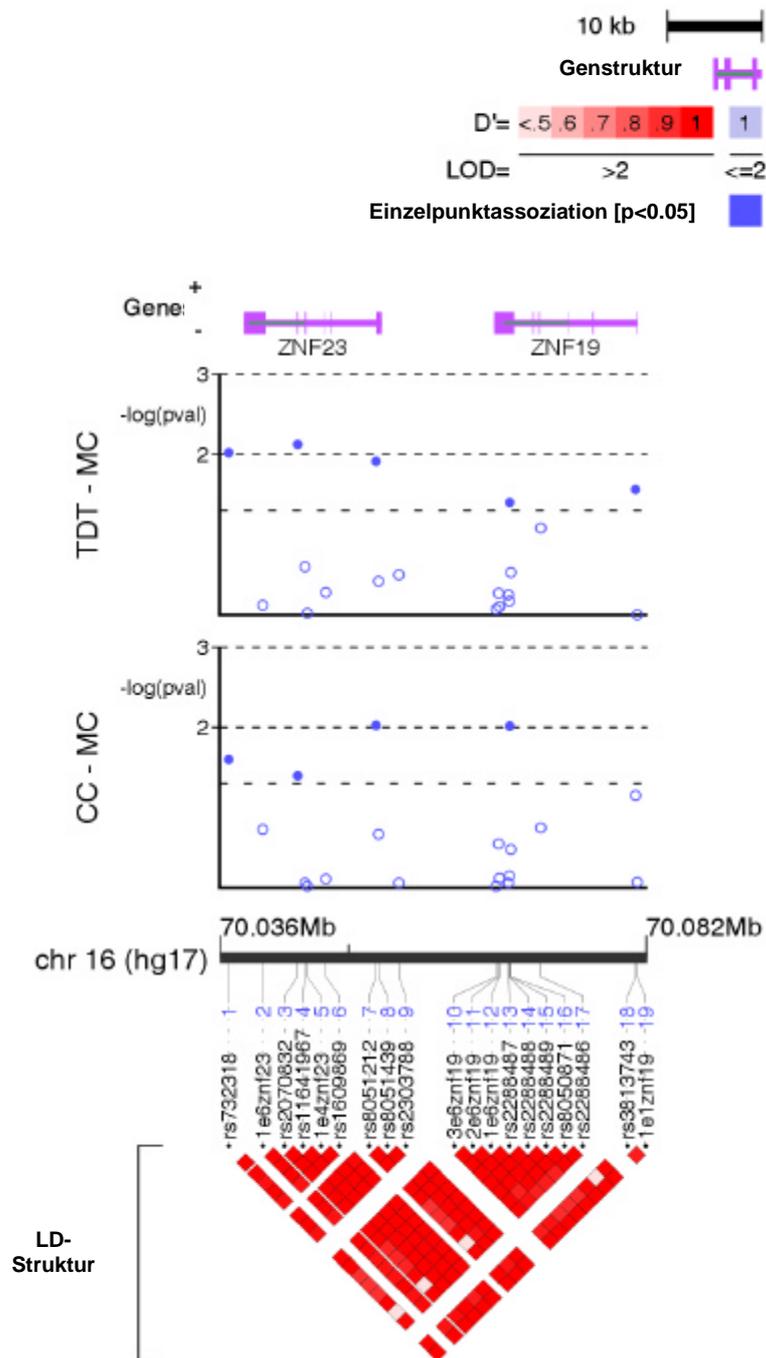


Abb. 3.9: Assoziationskartierung, Markerverteilung und LD-Struktur für *ZNF19* und *ZNF23*. [pval=p-Wert, D': liberaler Kopplungskoeffizient, LD: Kopplungsungleichgewicht, LOD: Kopplungswahrscheinlichkeit, Mb: Megabase, p: statistisches Signifikanzniveau]

3.3.3 GPR56 (G-Protein gekoppelter Rezeptor 56)

Das Gen *GPR56* erstreckte sich über eine Länge von 45 Kb, beinhaltete 15 Exone und wurde durch ein 4,1 Kb großes Transkript charakterisiert, welches ein Protein von 687 Aminosäuren codierte. 12 der 15 genotypisierten Marker in der *GPR56*-Region wurden mittels Sequenzanalyse validiert. Dazu gehörten 5 SNPs im 5'-UTR-Bereich, 3 intronische und 4 exonische SNPs, von denen 3 als codierend im Sinne eines Aminosäureaustausches definiert waren. Zusätzlich konnte ein neuer SNP im 5'-UTR-Bereich detektiert werden.

Die statistische Auswertung zeigte ausschließlich in der CC-Analyse (Tab. 3.6) signifikante Ergebnisse für den Locus *GPR56*. Für den Phänotyp CED konnten 6 Marker (G2-G6, G10) mit einer allelischen Assoziation zwischen $p=0,0442$ (G5) und $p=0,0012$ (G3 und G4) identifiziert werden. Der OR für den selteneren homozygoten Genotyp (OR_{Hom}) erwies sich für die Marker rs935746 (G2), hCV15977020 (G5) und rs8054093 (G6) mit Werten von 1,53 [KI 1,11-2,11], 1,53 [KI 1,09-2,13] und 1,61 [KI 1,15-2,25] als beachtenswert. Da jedoch alle 3 SNPs im Test auf Abweichung vom HWG moderate bis signifikante Ergebnisse ($p=0,093$ bis $p=0,031$) bedingt durch einen Heterozygotenüberschuss erzielten, musste ihre statistische Aussagekraft relativiert werden. Hingegen unterstrich der OR_{Hom} von 0,63 [KI 0,46-0,87] für die Marker G3 und G4 ihre allelische ($p=0,0012$) und genotypische ($p=0,003$) CED-Assoziation. Für den Phänotyp MC zeigten noch 5 der 6 SNPs ein statistisches Signal im Rahmen von $p_A=0,0472$ (G5) bis $p_A=0,0008$ (G3 und G4). Der OR_{Hom} für die Marker rs733464 (G3) und rs733463 (G4) steigerte sich gegenüber der CED-Analyse von 0,63 [KI 0,46-0,87] auf 0,58 [KI 0,40-0,83]. Der allelische und heterozygotenbasierte OR erwies sich ebenfalls als signifikanter gegenüber der CED-Statistik mit $OR_A=0,74$ [KI 0,62-0,88] und $OR_{Het}=0,67$ [KI 0,51-0,88]. Für den Phänotyp CU erbrachten noch 4 Marker (G2-G4, G10) ein signifikantes Resultat für den allelischen χ^2 -Test mit Werten von $p_A=0,0481$ (G10) bis $p_A=0,0215$ (G3). Die CC-Analyse der Gruppe B erbrachte keine Replikation der in der Gruppe A nachgewiesenen Assoziationen. In der TDT-Statistik der Gruppe A zeigte lediglich der SNP hCV1318779 (G11) ein moderates Signal für den Phänotyp MC mit $p=0,0129$. Für die Gruppe B konnte keine TDT-Assoziation nachgewiesen werden.

Das D'-basierte LD-Muster für den Locus *GPR56* (Abb. 3.10) wies eine eindeutig heterogene Struktur auf, in der eine klare LD-Blockbildung nur in der Markerregion G2 bis G6 beobachtet werden konnte. Die für diesen Block definierten Haplotypen erlaubten keine Zuordnung der assoziierten Marker auf dem Hintergrund eines einzelnen, assoziierten Risiko-Haplotyps. Ebenso lagen die selteneren, stark gekoppelten Allele der hochsignifikanten Marker G3 und G4 nicht auf einem einzelnen Haplotyp. Für die

codierenden SNPs rs1801257, rs1801255 und rs17379472 konnte in keiner der hier durchgeführten Auswertestatistiken eine Krankheitsassoziation ermittelt werden.

Tab. 3.6: CC-Statistik GPR56. [MA=selteneres Allel, Dist: Markerdistanz, f_{kont} : MA-Allelfrequenz in der Kontrollgruppe, f_{pat} : MA-Allelfrequenz in der Patientengruppe, p_A : allelbasiertes Signifikanzniveau, p_G : genotypenbasiertes Signifikanzniveau, OR_A : allelbasierter Odds Ratio für MA, OR_{Het} : heterozygotenbasierter Odds Ratio für MA, OR_{Hom} : homozygotenbasierter Odds Ratio für MA].

Nr.	Marker	Dist [Bp]	Gruppe A						
			f_{kont}	f_{pat}	p_A	p_G	OR_A	OR_{Het}	OR_{Hom}
G1	hCV1318758		0.239	0.254	0.3926	0.3092	1.08	1.19	0.97
G2	rs935746	4,281	0.445	0.493	0.0171	0.0052	1.21	0.96	1.53
G3	rs733464	421	0.445	0.382	0.0012	0.0030	0.77	0.69	0.63
G4	rs733463	69	0.445	0.382	0.0012	0.0028	0.77	0.68	0.63
G5	hCV15977020	4,970	0.408	0.447	0.0442	0.0046	1.18	0.91	1.53
G6	rs8054093	400	0.405	0.449	0.0259	0.0015	1.20	0.90	1.61
G7	rs1801257	1,671	0.465	0.474	0.6515	0.8187	1.04	0.99	1.08
G8	rs1801255	420	0.263	0.251	0.4881	0.1581	0.94	0.82	1.14
G9	rs1376041	78	0.293	0.307	0.4454	0.1290	1.07	0.92	1.41
G10	hCV15977030	1,382	0.410	0.450	0.0434	0.0783	1.18	1.05	1.43
G11	hCV1318779	7,748	0.399	0.385	0.4829	0.5913	0.94	1.01	0.86
G1	hCV1318758		0.239	0.259	0.3026	0.1263	1.11	1.29	0.92
G2	rs935746	4,281	0.445	0.492	0.0387	0.0306	1.20	1.00	1.51
G3	rs733464	421	0.445	0.372	0.0009	0.0027	0.74	0.67	0.58
G4	rs733463	69	0.445	0.372	0.0008	0.0024	0.74	0.67	0.58
G5	hCV15977020	4,970	0.408	0.452	0.0472	0.0139	1.20	0.95	1.57
G6	rs8054093	400	0.405	0.452	0.0327	0.0095	1.21	0.96	1.63
G7	rs1801257	1,671	0.465	0.480	0.4990	0.6396	1.06	0.98	1.13
G8	rs1801255	420	0.263	0.257	0.7389	0.1873	0.97	0.83	1.23
G9	rs1376041	78	0.293	0.304	0.5864	0.5009	1.06	0.96	1.28
G10	hCV15977030	1,382	0.410	0.438	0.1961	0.1560	1.12	0.95	1.32
G11	hCV1318779	7,748	0.399	0.379	0.3679	0.6141	0.92	0.96	0.83
G1	hCV1318758		0.239	0.254	0.4980	0.8011	1.08	1.08	1.14
G2	rs935746	4,281	0.445	0.497	0.0330	0.0363	1.23	1.04	1.58
G3	rs733464	421	0.445	0.390	0.0215	0.0418	0.80	0.71	0.67
G4	rs733463	69	0.445	0.393	0.0277	0.0503	0.81	0.71	0.69
G5	hCV15977020	4,970	0.408	0.442	0.1503	0.0638	1.15	0.93	1.45
G6	rs8054093	400	0.405	0.446	0.0900	0.0288	1.18	0.94	1.55
G7	rs1801257	1,671	0.465	0.466	0.9622	0.9368	1.00	1.05	1.00
G8	rs1801255	420	0.263	0.228	0.0936	0.0843	0.83	0.72	0.90
G9	rs1376041	78	0.293	0.307	0.5214	0.4785	1.07	0.97	1.33
G10	hCV15977030	1,382	0.410	0.458	0.0481	0.1372	1.22	1.24	1.48
G11	hCV1318779	7,748	0.399	0.400	0.9389	0.3944	1.01	1.18	0.93

0.05 > p > 0.01	0.01 > p > 0.001	0.001 > p
1.3 < OR < 1.5	1.5 < OR < 2	2 < OR
0.77 > OR > 0.67	0.67 > OR > 0.5	0.5 > OR
OR [95% Konfidenzintervall inkl. 1]		

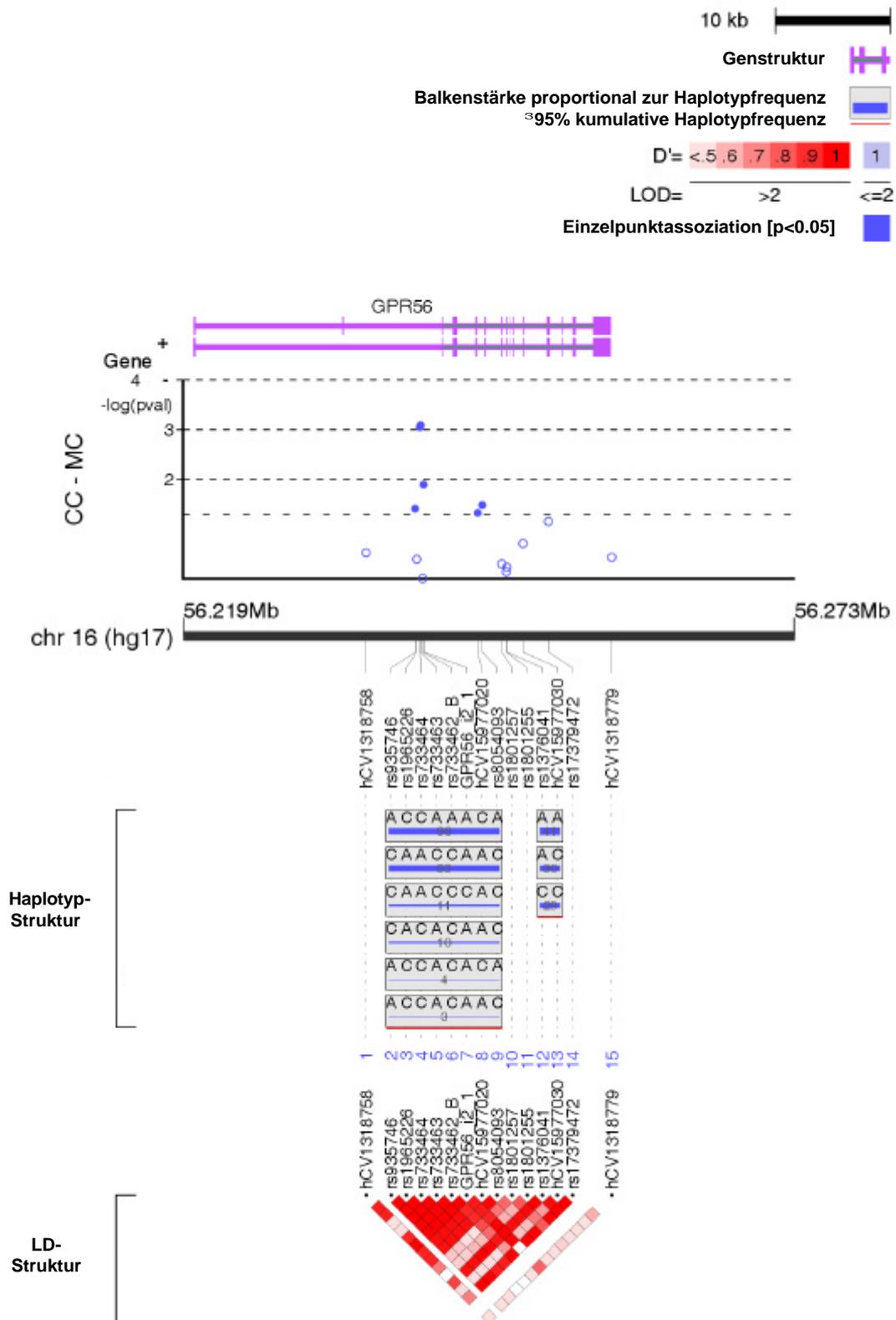


Abb. 3.10: Assoziationskartierung, Markerverteilung, Haplotyp- und LD-Struktur für *GPR56*. [pval=p-Wert, D': liberaler Kopplungskoeffizient, LD: Kopplungsungleichgewicht, LOD: Kopplungswahrscheinlichkeit, Mb: Megabase, p: statistisches Signifikanzniveau]

3.3.4 *BRD7 (Bromodomänen Protein 7)*

Der Locus *BRD7* überspannte eine Region von 49,9 Kb und beinhaltete 18 Exone. Über ein 2,2 Kb großes Transkript wurde das *BRD7*-Protein, welches aus 652 Aminosäuren bestand, codiert. Die Sequenzierung der Exone und des Promoterbereiches konnte lediglich einen intronischen SNP (B1) und einen nicht codierenden, exonischen SNP (rs4477702) validieren. Neue, bis dahin unbeschriebene Marker konnten durch die Mutationsdetektion nicht ermittelt werden.

Die CC-Analyse (Tab. 3.7) der Genotypisierungsgruppe A wies für alle drei Phänotypen assoziierte Marker in der *BRD7*-Region auf. Für den Phänotyp CED konnten signifikante Ergebnisse im allelischen χ^2 -Test für 5 SNPs (B1, B4-B7) mit Werten von $p_A=0,0187$ (B4) bis $p_A=0,0033$ (B6) ermittelt werden. Der homozygotenbasierte OR zeigte für die SNPs rs8059133 (B5) und hCV148486 (B6) starke Assoziationen mit $OR_{Hom}=0,45$ [KI 0,24-0,85] und $OR_{Hom}=0,41$ [KI 0,22-0,76]. Die selben Marker und der SNP rs8046783 (B2) zeigten für den Phänotyp MC eine allelische Assoziation im Bereich von $p_A=0,0444$ (B4) bis $p_A=0,0040$ (B1). Die deutlichsten OR-Signale zeigten dabei erneut die Marker B5 und B6 mit $OR_{Hom}=0,46$ [KI 0,22-0,96] und $OR_{Hom}=0,42$ [KI 0,21-0,86]. Die Marker B5, B6 und B7 erbrachten auch für den Phänotyp CU eine statistische Relevanz mit $p_A=0,0150$ (B5, B7), $p_A=0,0100$ (B6) und einem OR_{Hom} für den SNP B6 von 0,41 [KI 0,18-0,92]. Die Auswertung der Genotypendaten der Typisierungsgruppe B zeigte eine Replikation der Assoziationssignale für den Phänotyp CED für die Marker B1 und B4 mit $p_A=0,0201$ und $p_A=0,0298$. Für MC konnte in der Gruppe B eine Assoziationsreplikation für die Marker B1, B4, B5 und B6 mit einer allelischen Signifikanz zwischen $p_A=0,0367$ (B4) und $p_A=0,0253$ (B6) verzeichnet werden. In der TDT-Statistik (Tab. 3.8) erbrachten für den Phänotyp CED die selben Marker, die in der CC-Analyse auffällig waren, signifikante Resultate. Diese erstreckten sich über einen Signifikanzbereich von $p=0,0199$ (B4) bis $p=0,0030$ (B7). Für die Marker rs8059133 (B5) und hCV148486 (B6) konnte das TDT-Signal für CED in der Gruppe B mit $p=0,0305$ und $p=0,0309$ repliziert werden. Für den Phänotyp MC wurden signifikante TDT-Ergebnisse für die SNPs B1 und B4 bis B7 im Rahmen von $p=0,0420$ (B4) bis $p=0,0034$ (B1) ermittelt. Für die Marker B1, B5 und B6 konnte die TDT-Assoziation für den Phänotyp MC mit $p=0,0486$ (B1), $p=0,0201$ (B5) und $p=0,0172$ (B6) in der Gruppe B repliziert werden.

Tab. 3.7: CC-Statistik BRD7. [MA=selteneres Allel, Dist: Markerdistanz, f_{Kont} : MA-Allelfrequenz in der Kontrollgruppe, f_{Pat} : MA-Allelfrequenz in der Patientengruppe, p_A : allelbasiertes Signifikanzniveau, p_G : genotypenbasiertes Signifikanzniveau, OR_A : allelbasierter Odds Ratio für MA, OR_{Het} : heterozygotenbasierter Odds Ratio für MA, OR_{Hom} : homozygotenbasierter Odds Ratio für MA].

Nr.	Marker	Dist [Bp]	Gruppe A							
			f_{Kont}	f_{Pat}	p_A	p_G	OR_A	OR_{Het}	OR_{Hom}	
CED	B1	rs7184802	0.212	0.167	0.0038	0.0177	0.74	0.75	0.57	
	B2	rs8046783	3,905	0.409	0.374	0.0737	0.1849	0.86	0.92	0.74
	B3	rs4785408	1,838	0.190	0.196	0.6619	0.7872	1.05	1.01	1.25
	B4	hCV183346	1,325	0.218	0.181	0.0187	0.0520	0.79	0.83	0.56
	B5	rs8059133	7,639	0.192	0.150	0.0048	0.0137	0.74	0.80	0.45
	B6	hCV148486	1,035	0.197	0.154	0.0033	0.0062	0.74	0.81	0.41
	B7	rs4785412	13,730	0.183	0.143	0.0056	0.0214	0.74	0.77	0.50
MC	B1	rs7184802	0.212	0.161	0.0040	0.0179	0.71	0.73	0.49	
	B2	rs8046783	3,905	0.409	0.365	0.0466	0.1388	0.83	0.88	0.68
	B3	rs4785408	1,838	0.190	0.196	0.7292	0.9311	1.04	1.03	1.13
	B4	hCV183346	1,325	0.218	0.182	0.0444	0.1404	0.80	0.81	0.64
	B5	rs8059133	7,639	0.192	0.151	0.0169	0.0490	0.75	0.81	0.46
	B6	hCV148486	1,035	0.197	0.154	0.0106	0.0255	0.74	0.82	0.42
	B7	rs4785412	13,730	0.183	0.147	0.0287	0.0813	0.77	0.81	0.49
CU	B1	rs7184802	0.212	0.181	0.1169	0.2985	0.82	0.83	0.68	
	B2	rs8046783	3,905	0.409	0.385	0.3307	0.3925	0.90	1.01	0.78
	B3	rs4785408	1,838	0.190	0.193	0.8731	0.9660	1.02	1.00	1.11
	B4	hCV183346	1,325	0.218	0.186	0.1209	0.2025	0.82	0.88	0.55
	B5	rs8059133	7,639	0.192	0.148	0.0150	0.0668	0.73	0.76	0.51
	B6	hCV148486	1,035	0.197	0.151	0.0100	0.0375	0.72	0.79	0.41
	B7	rs4785412	13,730	0.183	0.140	0.0150	0.0638	0.72	0.74	0.52
			Gruppe B							
			f_{Kont}	f_{Pat}	p_A	p_G	OR_A	OR_{Het}	OR_{Hom}	
CED	B1	rs7184802	0.187	0.161	0.0201	0.0595	0.83	0.84	0.64	
	B2	rs8046783	3,905	0.390	0.378	0.3910	0.5142	0.95	0.90	0.93
	B3	rs4785408	1,838	0.197	0.207	0.4296	0.5936	1.06	1.03	1.26
	B4	hCV183346	1,325	0.191	0.166	0.0298	0.0775	0.84	0.87	0.63
	B5	rs8059133	7,639	0.163	0.143	0.0655	0.1792	0.86	0.86	0.72
	B6	hCV148486	1,035	0.164	0.143	0.0542	0.1534	0.85	0.85	0.72
	B7	rs4785412	13,730	0.158	0.139	0.0675	0.1661	0.86	0.87	0.65
MC	B1	rs7184802	0.187	0.158	0.0306	0.0865	0.82	0.79	0.73	
	B2	rs8046783	3,905	0.390	0.369	0.2143	0.4291	0.91	0.89	0.85
	B3	rs4785408	1,838	0.197	0.204	0.6251	0.7309	1.04	1.08	0.96
	B4	hCV183346	1,325	0.191	0.162	0.0367	0.1122	0.82	0.82	0.68
	B5	rs8059133	7,639	0.163	0.136	0.0346	0.0966	0.81	0.83	0.57
	B6	hCV148486	1,035	0.164	0.135	0.0253	0.0755	0.80	0.81	0.57
	B7	rs4785412	13,730	0.158	0.135	0.0595	0.1553	0.83	0.84	0.60
CU	B1	rs7184802	0.187	0.164	0.1040	0.1825	0.85	0.89	0.56	
	B2	rs8046783	3,905	0.390	0.388	0.8938	0.5022	0.99	0.90	1.04
	B3	rs4785408	1,838	0.197	0.210	0.4037	0.1787	1.08	0.97	1.59
	B4	hCV183346	1,325	0.191	0.170	0.1453	0.2572	0.87	0.91	0.61
	B5	rs8059133	7,639	0.163	0.150	0.3323	0.6020	0.91	0.89	0.88
	B6	hCV148486	1,035	0.164	0.151	0.3274	0.5969	0.90	0.89	0.88
	B7	rs4785412	13,730	0.158	0.143	0.2352	0.4703	0.88	0.90	0.71

0.05 > p > 0.01	0.01 > p > 0.001	0.001 > p
1.3 < OR < 1.5	1.5 < OR < 2	2 < OR
0.77 > OR > 0.67	0.67 > OR > 0.5	0.5 > OR
OR [95% Konfidenzintervall inkl. 1]		

Tab. 3.8: TDT-Statistik *BRD7*. [MA=selteneres Allel, Dist: Markerdistanz, trans: Anzahl transmittierter MA-Allele, untrans: Anzahl nicht-transmittierter MA-Allele, Chi²: Assoziationskoeffizient, p: statistisches Signifikanzniveau].

Nr.	Marker	Dist [Bp]	Gruppe A				Gruppe B			
			trans	untrans	Chi ²	p	trans	untrans	Chi ²	p
B1	rs7184802		175	233	8.25	0.0041	172	210	3.78	0.0519
B2	rs8046783	3,905	325	375	3.57	0.0588	304	347	2.84	0.0919
B3	rs4785408	1,838	247	250	0.02	0.8930	217	236	0.80	0.3720
CED B4	hCV183346	1,325	242	296	5.42	0.0199	175	211	3.36	0.0669
B5	rs8059133	7,639	194	252	7.54	0.0060	159	200	4.68	0.0305
B6	hCV148486	1,035	206	268	8.11	0.0044	160	201	4.66	0.0309
B7	rs4785412	13,730	168	227	8.81	0.0030	152	179	2.20	0.1378
B1	rs7184802		100	146	8.60	0.0034	108	139	3.89	0.0486
B2	rs8046783	3,905	191	219	1.91	0.1667	199	238	3.48	0.0621
B3	rs4785408	1,838	149	138	0.42	0.5161	151	165	0.62	0.4310
MC B4	hCV183346	1,325	147	184	4.14	0.0420	111	140	3.35	0.0672
B5	rs8059133	7,639	120	162	6.26	0.0124	102	138	5.40	0.0201
B6	hCV148486	1,035	131	172	5.55	0.0185	102	139	5.68	0.0172
B7	rs4785412	13,730	103	149	8.40	0.0038	99	126	3.24	0.0719
B1	rs7184802		75	87	0.89	0.3458	64	71	0.36	0.5469
B2	rs8046783	3,905	134	156	1.67	0.1964	105	109	0.07	0.7845
B3	rs4785408	1,838	98	112	0.93	0.3340	66	71	0.18	0.6692
CU B4	hCV183346	1,325	95	112	1.40	0.2374	64	71	0.36	0.5469
B5	rs8059133	7,639	74	90	1.56	0.2115	57	62	0.21	0.6467
B6	hCV148486	1,035	75	96	2.58	0.1083	58	62	0.13	0.7150
B7	rs4785412	13,730	65	78	1.18	0.2770	53	53	0.00	1

0.05 > p > 0.01 **0.01 > p > 0.001** **0.001 > p**

Die Berechnung des D'-basierten Kopplungsungleichgewichts über die komplette Markerzahl der *BRD7*-Region (Abb. 3.11) zeigte eine homogene LD-Struktur im gesamten Genbereich auf und ließ eine nähere Einordnung der detektierten Assoziationssignale in Substrukturen des *BRD7*-Locus nicht zu. Lediglich die intronischen SNPs B5, B6 und B7 konnten einem einzelnen Haplotyp mit einer Frequenz von 13,7% (CED) bzw. 13,5% (MC) in der Patientengruppe A zugeordnet werden. Dieses äußerte sich in der TDT-Statistik für den 3-Marker-Haplotyp B5-B6-B7 mit deutlichen Assoziationen von $p=0,00038$ für CED und $p=0,00046$ für MC.

Für das Kandidatengen *BRD7* galt es, die physikalische Nähe zu dem MC assoziierten Krankheitsgen *NOD2* zu berücksichtigen. Die Distanz von 328 Kb legte den Verdacht nahe, dass die für *BRD7* ermittelten Assoziationssignale über eine regionale Kopplung die *NOD2*-Assoziation widerspiegeln. Die CC-Analyse der in der *BRD7-NOD2*-Region typisierten Marker (Abb. 3.12) zeigte die lokale Anhäufung von CED- und MC-assoziierten SNPs für die beiden Genregionen einschließlich des *NOD2*-benachbarten Gens *CYLD* (Zylindromatose). Für den Phänotyp CU wurde hingegen ausschließlich für den *BRD7*-Locus eine statistische Assoziation mehrerer, genbasierter SNPs beobachtet. Der statistischen Auswertung lagen dabei die Genotypdaten für die Typisierungsgruppe der Feinkartierungsphase I zu Grunde (siehe 3.1.1).

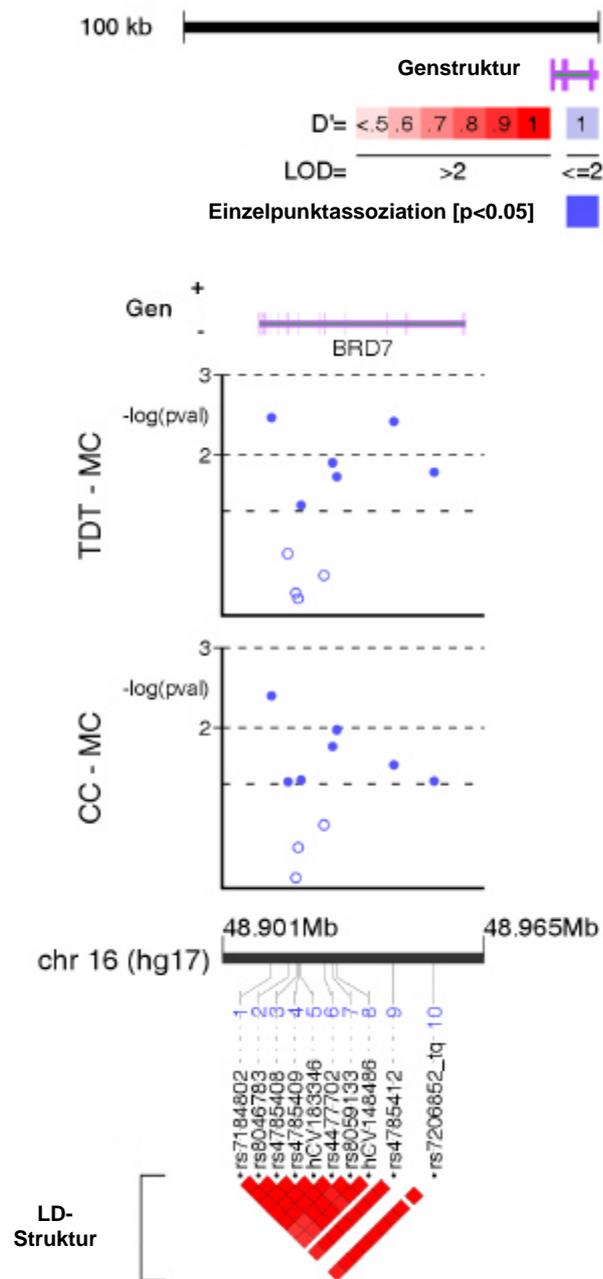


Abb. 3.11: Assoziationskartierung, Markerverteilung und LD-Struktur für *BRD7*.
 [pval=p-Wert, D': liberaler Kopplungskoeffizient, LD: Kopplungsungleichgewicht, LOD: Kopplungswahrscheinlichkeit, Mb: Megabase, p: statistisches Signifikanzniveau]

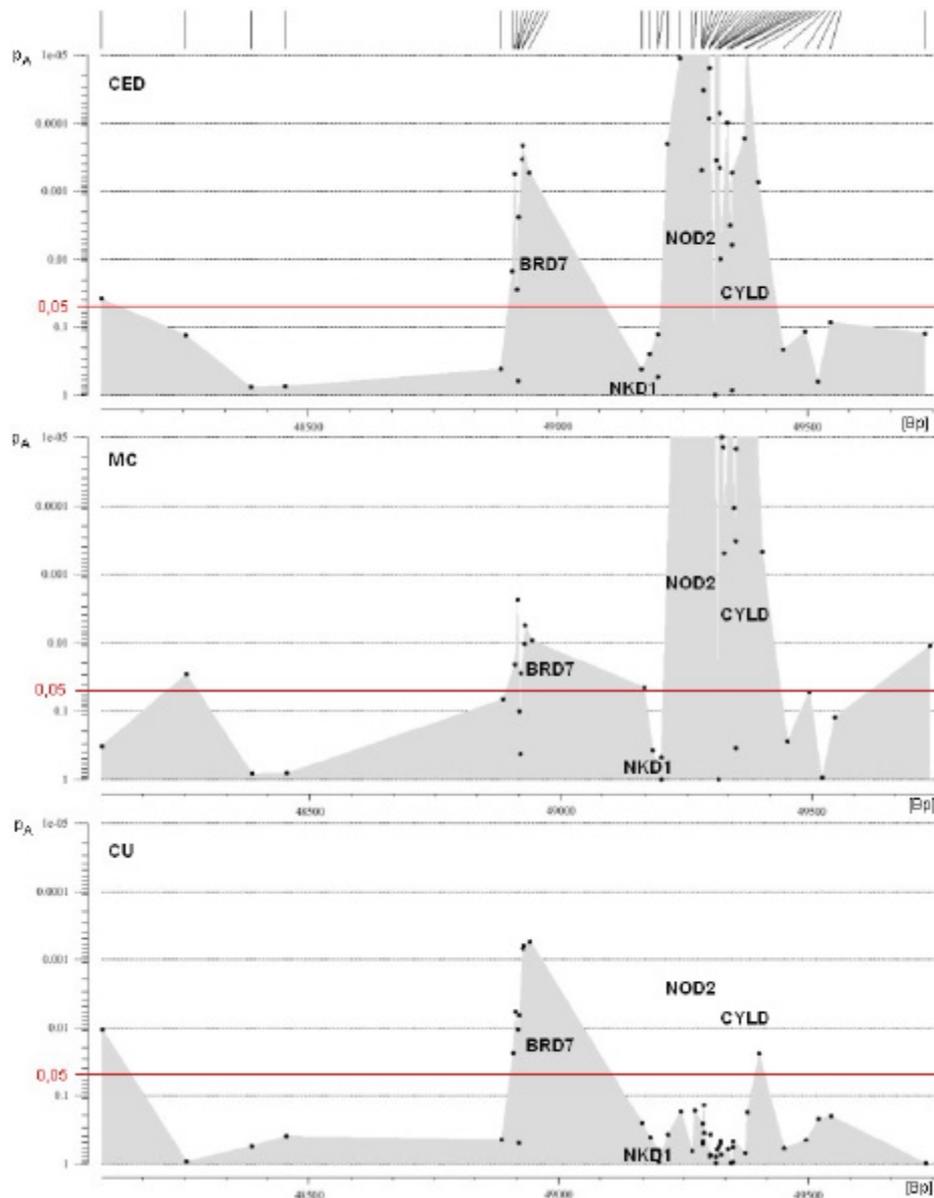


Abb. 3.12: CC-Analyse der *BRD7-NOD2*-Region für die Phänotypen CED, MC und CU. [Bp: Basenpaar, p_A : allelbasiertes Signifikanzniveau].

Die LD-Struktur über die gesamte Region (Abb. 3.13) zeigte eine deutliche, dreigeteilte Blockbildung. In der darauffolgenden Analyse wurde der LD-Block1 dem Locus *BRD7*, Block2 dem Gen *NKD1* und Block3 den Genen *NOD2* und *CYLD* zugeordnet. Zur Abschätzung der Signalkopplung bzw. *NOD2*-Unabhängigkeit der *BRD7*-Assoziation wurde die blockbasierte Genotypenpaarung für die Patienten- und Kontrollgruppe durchgeführt; die resultierenden, blockspezifischen Datensätze wurden im Anschluss über die CC-Statistik analysiert. Das Grundkonzept der Genotypenpaarung lag dabei in der Selektion von Patienten- und Kontrollindividuen nach dem Kriterium der maximalen Genotypübereinstimmung für einen bestimmten LD-Block.

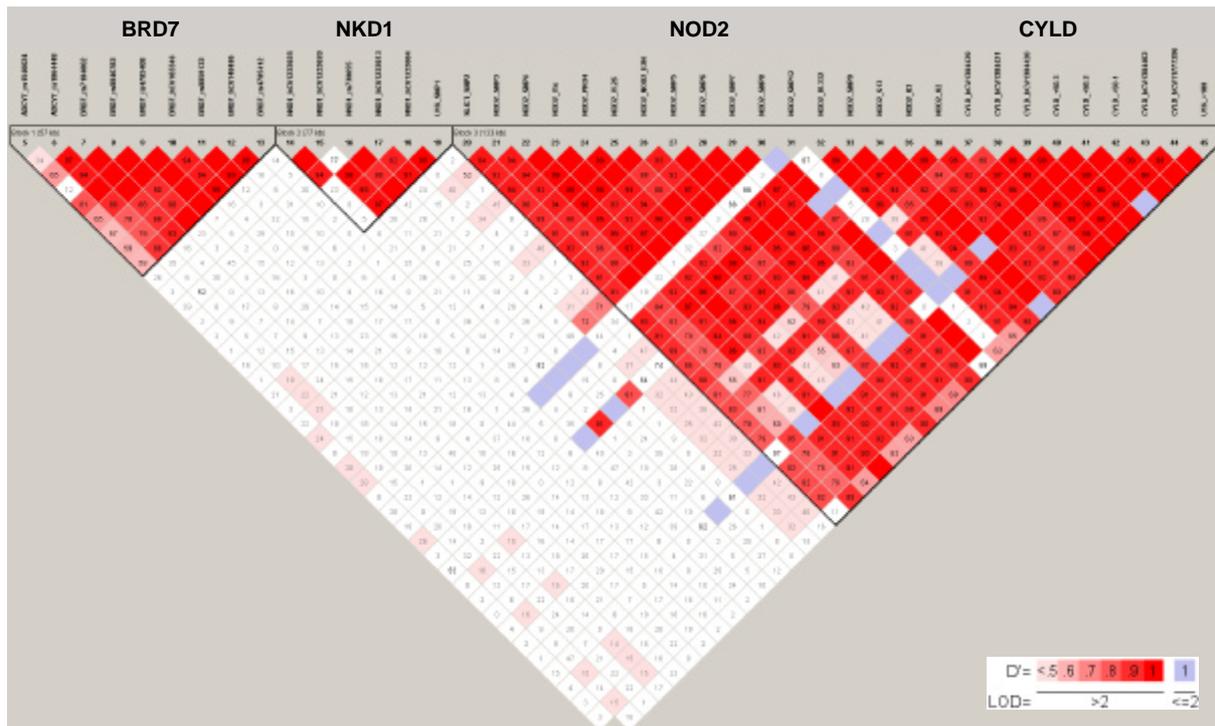


Abb. 3.13: D'-basierte LD-Struktur der *BRD7-NOD2*-Region. [D': liberaler Kopplungskoeffizient, LOD: Kopplungswahrscheinlichkeit]

So wurde für jeden der drei LD-Blöcke ein Datensatz mit Patienten und Kontrollen generiert, indem nur die Patienten in die CC-Analyse gelangten, die für die Gesamtheit der Marker des zugrundeliegenden LD-Blocks in ihrer Genotypkombination eine übereinstimmende Kontrollperson besaßen. Patienten und Kontrollen, die nach Abschluss der paarweisen Zuordnung noch verblieben, wurden aus dem Datensatz entfernt. Somit wurde erzielt, dass die CC-Assoziation für die der Genotypenangleichung zu Grunde liegenden Region künstlich aufgehoben wurde, um die kopplungsbedingte Abhängigkeit der Nachbarregionen zu überprüfen. Konnte eine statistische Assoziation für die *BRD7*-Region nachgewiesen werden, nachdem die unterschiedlichen Genotypenverteilung in den Patienten und Kontrollen für die Marker des *NOD2*-beinhaltenden LD-Block3 angeglichen wurde, wäre dieses ein Hinweis auf eine *NOD2*-unabhängige Krankheitsassoziation.

Die CC-Analyse der blockbezogenen Datensätze zeigte für den Phänotyp CED bei der Genotypenpaarung für den Block1 (Abb. 3.14 A) eine starke *NOD2*-Assoziation, welche sich in einer Größenordnung von $p_A < 10^{-5}$ befand und damit der Signalstärke der ungepaarten Analyse (Abb. 3.12) entsprach. Für einige *NOD2*-Marker konnte eine tendenzielle Signalschwächung beobachtet werden, die gegebenenfalls auch auf die um bis zu 35% reduzierte Individuenzahl in den genotypgepaarten Datensätzen zurückzuführen war. Die Auswertung der nach Block3 (*NOD2-CYLD*) angepassten Genotypdaten zeigte für den Phänotyp CED (Abb. 3.14 B) ein reduziertes, aber signifikantes Ergebnis für die *BRD7*

Marker B1, B5, B6 und B7 mit allelischen Assoziationen von $p_A=0,0285$ (B1) bis $p_A=0,0128$ (B6). Eine moderate Signifikanz von $p_A=0,049$ und $p_A=0,045$ erzielten die *BRD7* SNPs B5 und B6 für den Phänotyp CU unter der selbigen Block3-Anpassung der Datenanalyse (Abb. 3.14 C).

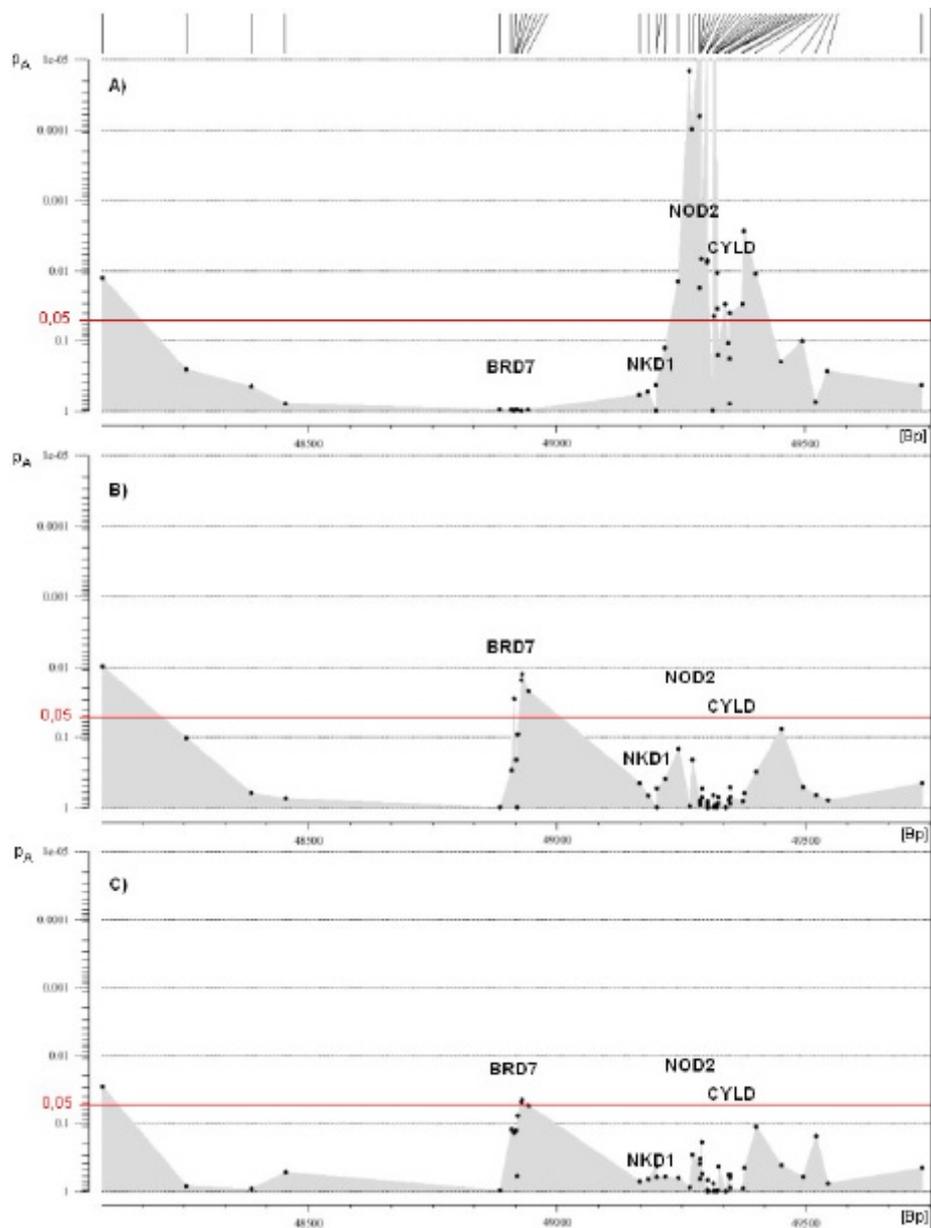


Abb. 3.14: CC-Analyse der LD-blockbasierten Genotypenpaarung für die *BRD7-NOD2*-Region. A) Phänotyp CED, Genotypenpaarung Block 1. B) Phänotyp CED, Genotypenpaarung Block 3. C) Phänotyp CU, Genotypenpaarung Block 3. [Bp: Basenpaar, p_A : allelbasiertes Signifikanzniveau].

4 Diskussion

4.1 Interpretation der Kandidatengenstatistik

Mit Hilfe der Positionsklonierung wurde in einer drei Kartierungsphasen umfassenden Assoziationsstudie eine 35 Mb große Region des Chromosom 16q hinsichtlich eines Risikogens für den Krankheitsphänotyp CED untersucht. Hierbei wurden insgesamt 436 Assoziationsmarker in Form biallelischer SNPs in einer umfangreichen und repräsentativen Probensammlung von insgesamt 6.688 Individuen genotypisiert. Durch die Auswertung der experimentell ermittelten Genotypdaten konnten die Gene *BRD7*, *GPR56*, *ZFP90*, *ZNF19* und *ZNF23* mittels der CC- und TDT-Statistik als assoziierte Kandidatengene identifiziert werden. Die zu beobachtenden Effektstärken variieren deutlich, sowohl innerhalb als auch zwischen den Kandidatenloci und weisen ein breites Signalspektrum auf, das sowohl moderate als auch höchstsignifikante Assoziationen beinhaltet. Einen funktionellen Risikostatus auf der Basis von Mutationen, die einem Aminosäureaustausch zu Grunde liegen oder transkriptionsregulatorische bzw. transkriptverändernde Eigenschaften besitzen, konnte bisher keinem der fünf Gene auf Basis der vorliegenden Untersuchung zugesprochen werden.

Aus funktioneller Sicht lassen sich die Gene *ZFP90*, *ZNF19* und *ZNF23* als Mitglieder der Zinkfinger-Protein-Genfamilie zusammenfassen, sodass in der nachfolgenden Betrachtung drei Kandidatengruppen hinsichtlich ihrer Krankheitsrelevanz erörtert werden.

4.1.1 Zinkfingergene *ZFP90*, *ZNF19* und *ZNF23*

Zinkfingerproteine besitzen eine DNA-bindende Domäne und spielen insbesondere als Transkriptionsfaktoren eine wesentliche, regulatorische Rolle bei der Genexpression [Bray *et al.*, 1991]. Sie repräsentieren die größte Proteinklasse innerhalb der transkriptionsregulierenden Elemente und besitzen einen Anteil von etwa 3% der bis dato identifizierten humanen Gene [Klug, 2005]. Die Fähigkeit der Zinkfingerdomäne, sich an spezifische DNA-Regionen zu heften, kommt u.a. auch in der Erforschung von therapeutischen Fusionsproteinen, welche eine spezifischen Kontrolle der Genexpression gewährleisten sollen, zum Einsatz [Klug, 2005]. Neben ihrer DNA-bindenden Eigenschaft besitzen zahlreiche Vertreter der Zinkfingerproteine auch die Fähigkeit sich an RNA anzulagern [Brown, 2005]. Ein Beispiel hierfür, welches gleichzeitig auch die Relevanz von Zinkfingerproteinen innerhalb akuter und chronischer Entzündungsprozesse darlegt, ist das

Protein Tristetraprolin (TTP), welches über das Zinkfingergen *ZFP36* (Zinkfingerprotein 36) kodiert ist. TTP ist ein zytosolisches Protein, welches regulatorisch in die Synthese des systemischen Entzündungsmediators TNF- α (Tumor Nekrose Faktor alpha) eingreift [Carballo *et al.*, 2001; Carballo *et al.*, 1998; Cooper *et al.*, 2001]. Zwei weitere Beispiele für entzündungsassoziierte Zinkfingerproteine sind die Proteine *A20* und *ZAS3*, die durch die Gene *TNFAIP3* (Tumor Nekrose Faktor alpha induziertes Protein 3) und *HIVEP3* (Humanes Immundefekt Virus Typ I Enhancer-Bindeprotein 3) codiert werden. Beiden Proteinen wird eine inhibitorische Rolle in der Aktivierung des proinflammatorischen Transkriptionsfaktors NF- κ B (nukleärer Faktor κ B) zugewiesen [Evans *et al.*, 2004; Hong *et al.*, 2003; Hong und Wu, 2005].

Die in der vorliegenden Studie untersuchten Zinkfingergene *ZFP90*, *ZNF19* und *ZNF23* sind derzeit hinsichtlich ihrer Proteine weder in einer funktionellen Spezifität noch krankheitsrelevanten Kausalität beschrieben; lediglich *ZFP90* ist als Transkriptionsrepressor charakterisiert [Thiel *et al.*, 2001]. Ihre Existenz und chromosomale Lokalisation ist dagegen eindeutig nachgewiesen [Cannizzaro *et al.*, 1993; Nagase *et al.*, 2001; Rousseau-Merck *et al.*, 1993]. Auffällig in der statistischen Auswertung der Genotypdaten der 20 *ZFP90*-Marker ist die komplett negative CC-Statistik. Die deutliche CED- und moderate CU-Assoziation im TDT verliert durch diesen Umstand ihre funktionelle Krankheitsrelevanz für die Durchschnittsbevölkerung. Trotz eines deutlichen Transmissionsungleichgewichts für die CED, welches auf 6 Einzelmarkersignalen in der Größenordnung von $p=10^{-3}$, einem hochsignifikanten 5-Marker-Haplotyp ($p=0,00035$) und zahlreichen moderat assoziierten SNPs beruht, liegt der Verdacht nahe, dass der *ZFP90*-Assoziation ein statistischer Fehler erster Ordnung anhaftet. Ursächlich hierfür könnte eine gerichtete Vererbung (siehe 4.2.5) des *ZFP90*-Locus oder einer mit ihm gekoppelten Region sein [Zollner *et al.*, 2004]. Die fehlende Assoziation funktioneller SNPs und eine negative Assoziationsstatistik für die Replikationsgruppe B unterstreichen diese Vermutung. Vertrauenswürdiger präsentieren sich dagegen die Morbus Crohn assoziierten Kandidatengene *ZNF19* und *ZNF23*. Die vier SNPs rs732318, rs8051212, rs2288489 und rs2070832, von denen der letztgenannte einen Aminosäureaustausch codiert, zeigen sowohl für die CC- als auch TDT-Analyse eine statistische Signifikanz im schwachen bis mittleren Spektrum und eine moderate Erhöhung des Krankheitsrisikos um das 1,3 bis 1,4fache für den heterozygoten Genotyp. Ob es sich hierbei jedoch um eine valide und für die CED bedeutsame Assoziation handelt, muss in Hinblick auf die fehlende Replikation und die moderate statistische Effektgröße in Frage gestellt werden. Hinzu kommt eine homogene LD-Struktur, welche beide Genloci überspannt und somit eine differenzierte Zuordnung des mit *ZNF19* und *ZNF23* assoziierten Risikos nicht ermöglicht.

4.1.2 G-Protein gekoppelter Rezeptor 56 (GPR56)

G-Protein gekoppelte Rezeptoren (GPCR) bilden mit mehr als 800 bekannten Repräsentanten die wahrscheinlich größte und mannigfaltigste Proteinfamilie innerhalb des humanen Proteoms [Fredriksson *et al.*, 2003]. Als Transmembranrezeptoren ist ihre primäre Funktion durch die Transduktion eines extrazellulären Stimulus (Bindung eines Liganden) in ein intrazelluläres Signal gekennzeichnet [Kroeze *et al.*, 2003]. Das intrazelluläre G-Protein (Guaninnukleotid-bindendes Protein) fungiert hierbei in seiner aktivierten Form als Signaltransduktor, der in die unterschiedlichsten, intrazellulären Signalwege wie z.B. der Regulation metabolischer Enzyme und Ionenkanäle oder zellulären Prozessen wie der Transkription und Sekretion eingebunden ist [Hamm, 1998; Neves *et al.*, 2002]. Etwa die Hälfte aller GPCRs werden der olfaktorischen und gustatorischen Wahrnehmung zugeordnet [Takeda *et al.*, 2002]. Dem gegenüber steht u.a. die maßgebliche Beteiligung von G-Protein gekoppelten Rezeptoren an Entzündungsreaktionen, intestinaler Immunantwort und T-Zell-Entwicklung [Qin *et al.*, 1998; Zabel *et al.*, 1999]. Ein prominentes Beispiel hierfür ist die Rolle des Histamins als Mediator akuter und chronischer Entzündungen und die Histamin-Rezeptoren vom Typ GPCR [Blaser *et al.*, 2006; de Esch *et al.*, 2005; Jutel *et al.*, 2005]. G-Protein gekoppelte Rezeptoren besitzen eine herausragende Bedeutung für die pharmazeutische Industrie, da sie die Grundlage bilden, den Zellstoffwechsel zu beeinflussen ohne das ein Medikament in die Zelle eindringen muss; es wird vermutet, dass 30%-50% der modernen Medikamente auf GPCRs basieren [Flower, 1999; Robas *et al.*, 2003].

Als viertes potentiell Kandidatengen wurde das Gen *GPR56* in der vorliegenden Assoziationsstudie untersucht. Hinweise aus der Literatur deuten auf eine Assoziation von *GPR56*-Mutationen und einer Fehlbildung des cerebralen Kortex in Form der bilateralen frontoparietalen Polymikrogyrie (BFPP) hin [Jansen und Andermann, 2005; Piao *et al.*, 2005; Piao *et al.*, 2004]. Gleichzeitig wird *GPR56* auf der Basis gewebsspezifischer Expressionsmuster auch als Risikofaktor für Krebserkrankungen, insbesondere für die Entwicklung des Glioblastoms (Glioblastoma multiforme), diskutiert [Shashidhar *et al.*, 2005; Zendman *et al.*, 1999]. Die starke CC-Assoziation der Marker rs733464 ($p_A=0,0009$) und rs733463 ($p_A=0,0008$) mit dem Phänotyp MC scheint auf den ersten Blick, das Risikospektrum um die chronisch-entzündlichen Darmerkrankungen zu erweitern. Bei genauerer Betrachtung der Auswertestatistiken fällt jedoch auf, dass die *GPR56*-Assoziation alleinig auf der CC-Studie begründet ist. Hinzu kommt der Umstand, dass das statistische Signal nur durch zwei gekoppelte SNPs, die in der 5'-UTR-Region lokalisiert sind, getragen wird. Sowohl die funktionellen (hier codierenden) als auch mit den Signalmarkern gekoppelten SNPs zeigen keine oder eine nur mäßige Assoziation. Hinzu kommt eine

fehlende Signalreplikation in der zweiten Genotypisierungsgruppe, sodass das Gesamtbild für den *GPR56*-Locus die beobachtete Assoziation in die Kategorie der falsch positiven Signale einordnet. Als Erklärungsansatz könnte das Phänomen der weitreichenden Kopplung (siehe 4.2.3) von genetisch nicht zu unterscheidenden SNPs dienen [Lawrence, Evans, Morris *et al.*, 2005]. Hierdurch würde man allerdings die Marker-Assoziation als statistisch valide einordnen, jedoch die Lokalisation des assoziierten Risikoallels in einer mit den *GPR56*-SNPs gekoppelten Region vermuten.

4.1.3 Bromodomänen Protein 7 (*BRD7*)

Bromodomänen sind 110 Aminosäuren lange Sequenzen, die evolutiv konserviert als Bestandteil zahlreicher, eukaryotischer Transkriptionsfaktoren vorkommen und in Proteinen der signalabhängigen Transkriptionsregulation nachgewiesen wurden [Winston und Allis, 1999; Zhou *et al.*, 2004]. Die Transkriptionsaktivierung über die Beeinflussung der Chromatin-Remodellierung scheint hierbei der wesentliche Wirkmechanismus der Bromodomänen-Proteine zu sein und wurde im speziellen für *BRD7* eindeutig nachgewiesen [Cong *et al.*, 2005; Filetici *et al.*, 2001]. Desweiteren wurde *BRD7* in Genexpressionsstudien als potentielles Tumorsuppressorgen für das Nasopharynxcarcinom in der asiatischen Bevölkerung ausgewiesen [Zhang *et al.*, 2003; Zhou *et al.*, 2004].

Als einziges der fünf untersuchten Kandidatengene dieser Studie zeigt *BRD7* eine statistische Krankheitsrelevanz gleichermaßen in der initial genotypisierten Gruppe A und in der Replikationsgruppe B. Gleichzeitig kann für vier der sieben Marker sowohl in dem familien- als auch populationsbasierten Studiendesign eine hochsignifikante Assoziation ($p=0,006-0,003$) mit dem Phänotyp CED und eine moderate Signifikanz ($p=0,03-0,003$) für die Subphänotypen MC und CU beobachtet werden. Dem gegenüber steht zum einen eine fehlende funktionelle Assoziation, da alle assoziierten SNPs in der intronischen Genregion lokalisiert sind. Zum anderen besteht die Möglichkeit, dass die relative Nähe zu dem MC-Risikogen *NOD2* die Grundlage der *BRD7*-Assoziation bildet, indem eine genetische Kopplung der beiden Genregionen das vermeintliche Krankheitsrisiko der *BRD7*-Variationen generiert. Die Analyse der LD-blockbasierten Genotypenpaarung wiederum weist auf eine *NOD2*-Unabhängigkeit des *BRD7*-Risikos für die chronisch entzündlichen Darmerkrankungen und die Colitis ulcerosa hin. Die D' -basierte LD-Analyse für die *BRD7*-*NOD2*-Region untermauert diese Annahme durch eine größtenteils schwache bzw. vernachlässigbare Kopplung der beiden Genregionen mit Durchschnittswerten von $D' \leq 0,2$. Zusammengefasst qualifizieren diese Ergebnisse *BRD7* und die umgebenden Gene für weiterführende Experimente hinsichtlich einer Krankheitsrelevanz für die Colitis ulcerosa,

insbesondere vor dem Hintergrund der für diesen Phänotyp fehlenden NOD2-Assoziation [Hampe, Cuthbert *et al.*, 2001; Hugot *et al.*, 2001; Ogura *et al.*, 2001].

4.2 Potentielle Einflussgrößen in der Assoziationsstudie und ihre Relevanz für die vorliegende Studie

4.2.1 Populationsstratifikation

Die am häufigsten zitierte Fehlerquelle in populationsbasierten Assoziationsstudien ist die unentdeckte Populationsstratifikation [Cardon und Palmer, 2003]. Sie gilt sowohl für den statistischen Fehler 1. Ordnung in Form eines unechten Krankheitsrisikos, als auch den Fehler 2. Ordnung, der dem Übersehen echter Suszeptibilitätsfaktoren entspricht, als hauptverantwortlich. Ebenfalls wird ihr der Großteil der nicht replizierenden Assoziationssignale zugesprochen [Tabor *et al.*, 2002; Weiss und Terwilliger, 2000]. Per Definition gilt eine Studiengruppe als populationsstratifiziert, wenn sie aus Subpopulation besteht, die unabhängig vom Krankheitsstatus für die betrachteten Marker unterschiedliche Allelfrequenzen besitzen und dadurch eine falsche LD-Struktur für die Gesamtpopulation begründen. Die Basis hierfür bildet zum einen die Ausbildung und Etablierung sozialer, religiöser und ethnischer Substrukturen innerhalb einer Bevölkerung [Cardon und Palmer, 2003]. Zum anderen kann die Vermischung genetisch unterschiedlicher Populationen z.B. durch Völkerwanderungen eine Substrukturierung des genetischen Hintergrundes der Gesamtbevölkerung zur Folge haben [Cavalli-Sforza und Piazza, 1993]. Geht mit diesen Faktoren eine unterschiedliche Krankheitsprävalenz in der Patienten- und Kontrollpopulation einher, so sind die Grundvoraussetzungen für ein falsches bzw. verzerrtes Studienergebnis geschaffen [Wacholder *et al.*, 2000]. Für die CC-Studie würde dieses bedeuten, dass sich der genetische Hintergrund sowohl zwischen als auch innerhalb der Patienten- und Kontrollgruppen unterscheiden könnte und somit eine Differenzierung zwischen krankheits- und stratifikationsbedingten Assoziationssignalen nicht möglich wäre.

Ein prominentes Beispiele für eine falsch positive Krankheitsassoziation, die auf Populationsstratifikation basiert, ist die Assoziation zwischen einem HLA-Haplotyp und der Diabetes bei den Pima Indianern Nordamerikas [Knowler *et al.*, 1988]. Als Grundlage des starken Assoziationssignales ($OR=0,27$) konnte die Vermischung europäischer und indianischer Vorfahren der Studienpopulation identifiziert werden. Ein weiteres Beispiel ist die unechte Assoziation des Dopamin D2-Rezeptors mit dem Alkoholismus, die auf der Variation von Krankheitsprävalenz und SNP-Frequenz in Populationen unterschiedlicher Ethnien basiert [Gelernter *et al.*, 1993; Pato *et al.*, 1993]. Diese eindeutigen Beispiele sind

jedoch selten in der Literatur zu finden, sodass der empirische Beleg für die unentdeckte Populationstratifikation als Hauptverursacher des Typ I Fehlers in der CC-Studie noch erbracht werden muss [Cardon und Palmer, 2003].

Als Kontrolle bzw. Test für Substrukturen des genetischen Hintergrundes einer Studienpopulation kann die Genotypisierung von genomweiten, unabhängigen SNPs, die in genarmen oder stark konservierten Regionen des Genoms lokalisiert sind, verwendet werden [Pritchard und Rosenberg, 1999; Pritchard, Stephens und Donnelly, 2000]. Wird auf diese Weise eine Populationsstratifikation entdeckt, so ergeben sich zwei Möglichkeiten, die Teststatistik entsprechend anzupassen. Die erste ist die des *genomic control* (GC) Ansatzes, der die Populationstratifikation als multiplikativen Faktor für die χ^2 -Statistik ansieht, welcher fälschlicherweise die Assoziationssignifikanz erhöht bzw. erniedrigt. Dieser Faktor kann über einen GC-Markersatz bestimmt werden und als Korrekturfaktor die Teststatistik reskalieren [Bacanu *et al.*, 2000; Devlin und Roeder, 1999; Devlin *et al.*, 2001; Pritchard und Donnelly, 2001; Pritchard, Stephens, Rosenberg *et al.*, 2000]. Eine hohe Verlässlichkeit erzielt diese Methode vor allem in großen Studien mit einer hohen Probenanzahl, da die Fähigkeit Populationstratifikation aufzudecken mit der Zunahme des Probenumfangs einhergeht [Pritchard und Donnelly, 2001]. Die Anzahl der GC-Marker hat ebenfalls einen entscheidenden Einfluss auf die statistische Leistungsfähigkeit des GC-Experimentes, sodass bei starken Assoziationen wenige Dutzend, bei moderaten Effekten mehrere Hundert verwendet werden müssen [Freedman *et al.*, 2004; Turakulov und Easteal, 2003]. Als alternative Methode wird die ‚strukturierte‘ Assoziationsstatistik (SA) angesehen [Pritchard und Donnelly, 2001; Pritchard, Stephens, Rosenberg *et al.*, 2000]. Die Aufdeckung von Substrukturen wird hierbei analog zum GC-Ansatz durchgeführt, jedoch erfolgt im Anschluss eine Zuteilung der Einzelindividuen in genetisch angeglichenen, homogene Subpopulation, die frei von Populationstratifikation sind. Die sich anschließende Teststatistik beruht auf einer Kombination der jeweiligen Einzelergebnisse für die Subpopulationen.

Die in der vorliegenden Studie untersuchten Individuen stammen im Falle des familienbasierten Studiendesigns aus allen Teilen Deutschlands. Für die CC-Studie wurden ebenfalls Patienten aus ganz Deutschland verwendet, wohingegen bedingt durch eine Rekrutierung der in Kiel ansässigen Blutspender bzw. der Verwendung von Kontrollen des popgen-Projektes die Kontrollgruppe zum größten Teil aus dem norddeutschen Raum stammt. Somit stellt sich die Frage, ob eine Populationsstratifikation auf dem Hintergrund regionaler genetischer Unterschiede in der deutschen Bevölkerung eine Relevanz für die Studienergebnisse hätte. Eine aktuelle Studie beschäftigt sich auf der Basis von 210 GC-Markern und 3 populationsbasierten Kontrollgruppen aus dem nord-, ost und süddeutschen Raum mit dieser Fragestellung [Lamina *et al.*, 2005]. Die verwendeten SNPs setzen sich dabei zu einem Teil aus 140 Markern, die abseits funktioneller Genregionen in sogenannten

,Genwüsten' lokalisiert und damit geringer Selektion ausgesetzt sind, zusammen. Der andere Teil besteht aus 70 SNPs, die eine moderate Allelfrequenz besitzen, gleichmäßig über das Genom verteilt und zwischen den Genen positioniert sind. Somit soll gewährleistet werden, dass ausschließlich neutrale Drift- und Migrationsfaktoren der demographischen Geschichte in Form von Populationsstratifikation detektiert werden können. Die statistische Auswertung erfolgt sowohl über die GC- als auch SA-Statistik und ist zum gegenwärtigen Zeitpunkt noch nicht abgeschlossen [Lamina *et al.*, 2005].

Somit lässt sich vor dem Hintergrund derzeitiger Erkenntnisse eine populationsbasierte Verzerrung der Resultate in der CC-Statistik der 16q-Studie mit absoluter Sicherheit nicht ausschließen.

4.2.2 Probenumfang, statistische Aussagekraft und Risikoabschätzung

Dem Probenumfang wird in den Assoziationsstudien eine besondere Bedeutung zugesprochen, da die ermittelten Krankheitsrisiken sich signifikant und gerichtet zwischen kleinen und großen Studien unterscheiden. Kleine Studien postulieren stärkere Effekte, wohingegen große Studien die Effektgrößen konservativer definieren [Ioannidis *et al.*, 2003]. Hinzu kommt, dass in Metaanalysen nur ca. 15% der publizierten Risikoassoziationen konsistent repliziert werden [Ioannidis *et al.*, 2003] und dass in Erstpublikationen assoziationsbasierte Risikoallele in ihrer Effektstärke tendenziell überschätzt werden [Lohmueller *et al.*, 2003]. Letzteres ist zum Teil auch auf einen Publikationsbias zurückzuführen [Dickersin, 1997; Dickersin und Min, 1993], der beispielsweise in der Tatsache zum Ausdruck kommt, dass sich die Publikationsrate einer Krankheitsassoziation bei einer positiven Erstbeschreibung annähernd verdoppelt [Trikalinos *et al.*, 2004].

Ein weiterer Faktor, der die Untergrenze des Probenumfangs für eine solide Assoziationsstudie mitbestimmt, ist die Risikoabschätzung für die zu erwartenden Kandidatengene. So zeigen die meisten der bisher validierten Risikogene für polygene Erkrankungen eine moderate Effektstärke in der Größenordnung von $OR=1,1-1,5$ [Hirschhorn, 2005]. Um solche Effekte verlässlich von statistischen Typ I und Typ II Fehlern zu unterscheiden, sind Studiengruppen im Bereich von mehreren Tausend Individuen notwendig [Risch und Merikangas, 1996]. Dem gegenüber steht die Mehrzahl aktueller Studien mit Probensammlungen in der Größenordnung von einigen Hundert [Hirschhorn, 2005]. Die zusammenfassende Auswertung mehrerer Studien (Metaanalyse) und die damit verbundene Vergrößerung des Stichprobenumfangs und Erhöhung der statistischen Aussagekraft scheint hierbei ein geeignetes Mittel zu sein, genetische Risikofaktoren und den Studienbias verlässlich gegeneinander abzuwägen [Ioannidis *et al.*, 2001].

Neben der zu erwartenden Effektstärke geht auch die Häufigkeit des zu entdeckenden Risikoallels in die Berechnung des Probenumfangs ein. Hinweise auf verbreitete Mutationen (>5%) in häufigen (Volks-)Krankheiten sind die Grundlage für das *common variant/common disease* (CV/CD) Modell, dessen Grundannahme die Existenz einzelner, hochfrequenter Risikomutationen in der betroffenen Population postuliert [Chakravarti, 1999; Lander und Schork, 1994; Reich und Lander, 2001]. Als Erklärungsansatz dient zum einen der Umstand, dass polygene Volkskrankheiten evolutiv weniger von Nachteil sind wie monogene Krankheiten, die häufig lethal oder reproduktionseinschränkend sind. Desweiteren zeigen einzelne Risikoallele bei polygenetischen Erkrankungen gemäß dem multiplikativen Krankheitsmodell eine geringe Penetranz, sodass der Selektionsdruck auf die einzelne Mutation gering ist. Zusätzlich lässt die in Bezug auf monogene Erkrankungen relative Häufigkeit von Volkskrankheiten vermuten, dass einige Krankheitsmutationen basierend auf der demographischen Geschichte der menschlichen Bevölkerung eine hohe Verbreitung erfahren haben [Reich und Lander, 2001]. Nicht zuletzt unterstreichen empirische Untersuchungen wie z.B. die Entdeckung häufiger, krankheitsrelevanter Splice-Mutationen das CV/CD Modell [Lohmueller *et al.*, 2003; Ueda *et al.*, 2003; Valentonyte *et al.*, 2005]. Fasst man alle relevanten Faktoren zusammen und berücksichtigt auch die geringe statistische Aussagekraft früherer Assoziationsstudien [Trikalinos *et al.*, 2004], so kommt man zu dem Schluss, dass der Probenumfang bei den zu erwartenden moderaten Einzeleffekten mindestens in der Größenordnung von 10^3 liegen sollte. Dieses gilt gleichermaßen für das CC-Design [Ioannidis *et al.*, 2003] als auch für die familienbasierte Assoziationsstudie [Risch und Merikangas, 1996].

Betrachtet man nun vor diesem Hintergrund die für die chr16q-Studie verwendete Probengruppe A, von der 2398 Individuen in die TDT-Statistik und 1283 in die CC-Analyse eingegangen sind, so kann man der vorliegenden Studie seitens des Probenumfangs eine adäquate statistische Aussagekraft attestieren. Die Replikationsgruppe B mit 2178 Proben im familienbasierten und 2300 im populationsbasierten Studiendesign unterstreicht diesen Anspruch. Daneben lässt sich die Effektstärke der noch zu entdeckenden Risikogene für die CED und im Speziellen für MC basierend auf den bisherigen Erkenntnissen einer moderaten Größenordnung zuweisen. Ausgangspunkt für diese Schätzung ist das relative Geschwisterrisiko (λ_s) für die CED, welches für MC im Bereich von 15-35 [Kuster *et al.*, 1989; Meucci *et al.*, 1992; Probert *et al.*, 1993] und für CU bei 6-9 liegt [Meucci *et al.*, 1992; Orholm *et al.*, 1991] und damit der Summe aller genetischen Faktoren gleichgesetzt werden kann. Die bisher identifizierten Krankheitsgene für CED liefern für den Phänotyp MC und die heterozygote Trägerschaft des Risikoallels in der kaukasischen Bevölkerung im Falle von *NOD2* eine Erhöhung des Krankheitsrisikos um den Faktor 3 bis 4 [Cuthbert *et al.*, 2002;

Hampe, Cuthbert *et al.*, 2001; Hugot *et al.*, 2001] und für den *IBD5*-Haplotyp der Gene *OCTN1* und *OCTN2* eine Risikoerhöhung um das 2,5fache [Peltekova *et al.*, 2004]. Gleichzeitig wurde der statistische Nachweis einer Interaktion zwischen *NOD2* und *IBD5* erbracht [Mirza *et al.*, 2003], und die Vermutung eines gegebenenfalls epistatischen Zusammenwirkens durch eine gemeinsame Erhöhung des Erkrankungsrisikos um den Faktor 7 bis 10 unterstrichen [Peltekova *et al.*, 2004]. Als drittes Risikogen geht *DLG5* mit einer mäßigen Erhöhung des Krankheitsrisikos um das 1,5fache in das genetische Gesamtrisiko für den MC ein [Stoll *et al.*, 2004]. Basierend auf der polygenen Ätiologie und der Zugrundelegung eines multiplikativen Krankheitsmodells ergibt sich so für weitere Krankheitsgene der CED bzw. MC ein eher moderates Risikospektrum mit zu erwartenden Assoziationen in der Größenordnung von $OR=1,5$. Betrachtet man aus diesem Blickwinkel die assoziierten Marker der fünf postulierten Kandidatengene der chr16q-Studie, so entsprechen sie hinsichtlich ihrer Effektstärke grundsätzlich den Erwartungen.

4.2.3 LD- und Haplotypstruktur

Die Grundvoraussetzung für die Entdeckung von krankheitsassoziiierter Risikovarianten mittels indirekt assoziierter Marker, ist die Kopplung des Assoziationsmarkers mit der risikotragenden Mutation. Dabei üben sowohl die Marker- als auch Risikoallelcharakteristika einen vielfältigen Einfluss auf die statistische Möglichkeit aus, eine Kopplung zwischen beiden zu entdecken. Relevante Faktoren sind hierbei das Alter der krankheitsverursachenden Mutation, das Vererbungsmodell der Krankheit, die Entfernung zwischen Marker und Krankheitslocus sowie die Markerdichte [Jorde, 2000]. Letztere wird in konservativen Simulationen mit mindestens einem SNP pro 3-5Kb empfohlen [Kruglyak, 1999; Pritchard und Przeworski, 2001], jedoch haben empirische Untersuchungen gezeigt, dass für einen Großteil der Regionen im humanen Genom eine hohe Kopplung für Regionen von 30-60Kb vorliegt. Diese Größenordnung von Haplotypblöcken ist jedoch auf kaukasische Populationen beschränkt und ist für afrikanische Populationen auf Grund einer größeren genetischen Variabilität wesentlich kleiner [Goddard *et al.*, 2000; Jorde *et al.*, 1994; Peterson *et al.*, 1995; Reich *et al.*, 2001]. Zwischen diesen Blöcken hoher genetischer Kopplung liegen kleinere Regionen hoher Rekombination (*recombination hot spots*) [Gabriel *et al.*, 2002; Goldstein, 2001; Maniatis *et al.*, 2002]. Dieses Erkenntnis ist jedoch nicht als eine grundsätzliche Pauschalisierung zu verstehen, da das Kopplungsungleichgewicht neben seiner Populationsspezifität auch immer eine hohe Variation in Genregionen aufweist. Dieses kann als Hinweis auf einen fluktuierenden Selektionsdruck und eine unterschiedliche Funktionalität bzw. Überlebensrelevanz der einzelnen Gene gedeutet werden [Goddard *et*

al., 2000; Long *et al.*, 2004; Stephens *et al.*, 2001]. Eine weitere Besonderheit im Hinblick auf die Kopplung von SNPs stellen die genetisch ununterscheidbaren giSNPs (*genetically indistinguishable SNPs*) dar. Etwa 50% aller SNPs besitzen einen genetisch nicht zu unterscheidenden (nichtrekombinierten) ‚Partner-SNP‘, mit dem sie in perfektem LD ($r^2=1$) sind [Lawrence, Evans, Morris *et al.*, 2005]. Ein Drittel dieser giSNPs wiederum liegen nicht in direkter (rekombinationsfreier) Nachbarschaft, sondern sind als sogenannte nicht-konsequente Marker in unterschiedlichen, bis zu mehreren hundert Kilobasen entfernten Haplotypblöcken lokalisiert [Lawrence, Evans, Morris *et al.*, 2005; Peterson *et al.*, 1995; Zavattari *et al.*, 2000]. Diese weitreichende Kopplung von SNPs hat einen wesentlichen Einfluss auf die Assoziationsstudie, da die Feinkartierung und Lokalisierung des vermeintlichen Krankheitsgenes durch assoziierte giSNPs fehlgeleitet werden kann [Abecasis *et al.*, 2001; Pritchard und Przeworski, 2001].

Die Markerdichte in der vorliegenden Studie liegt mit einem SNP pro 80Kb in einem Bereich moderater statistischer Verlässlichkeit und muss als eine mögliche Ursache für die fehlende Identifizierung eines *NOD2*-unabhängigen Risikogens in der chr16q-Kopplungsregion betrachtet werden. Hinzu kommt eine z.T. stark heterogene Markerverteilung, die sich in der LD-Struktur der Studie widerspiegelt. Für die starke CC-Assoziation des *GPR56*-Locus, welche im wesentlichen auf zwei eng gekoppelte Marker beruht, muss die Möglichkeit in Erwägung gezogen werden, dass die weitreichende Kopplung von giSNPs *GPR56* fälschlicherweise als Kandidatenlocus charakterisiert haben und das eigentliche Risikogen an einer anderen chromosomalen Stelle lokalisiert ist.

4.2.4 Replikation von Assoziationssignalen

Eine Replikation von Assoziationssignalen ist unverzichtbar, da nur etwa 5%-15% der neu veröffentlichten Assoziationen auch konsistent mit einer Replikationsquote größer 75% repliziert werden, und die Mehrheit der veröffentlichten Risikomutationen für polygenetische Erkrankungen nicht in einer unabhängigen Studie bestätigt werden [Cardon und Bell, 2001; Hirschhorn *et al.*, 2002; Ioannidis *et al.*, 2001; Ioannidis *et al.*, 2003; Terwilliger und Goring, 2000]. Die Gefahr, einem Typ I Fehler bei der Untersuchung einzelner Marker (*single hit associations*) aufzusitzen, ist entsprechend hoch und muss bei der Verfolgung von statistischen Signalen jedweder Größenordnung berücksichtigt werden. Es wurde gezeigt, dass erste Assoziationsbeschreibungen häufig die höchste Signalstärke beinhalteten und bei anschließender Replikation eine signifikante Schwächung des Risikobeitrages verzeichnet werden musste. Als Ursache werden sowohl ein Studienbias in Form von verzerrten bzw. überschätzten Ergebnissen [Hirschhorn, 2005; Lohmueller *et al.*, 2003] als auch genetische

Unterschiede zwischen der Erstbeschreibungs- und der Replikationspopulation beschrieben [Ioannidis *et al.*, 2001] (siehe auch 4.2.2). Im Allgemeinen muss das Aufspüren und Replizieren von genetischen Risikofaktoren für multifaktorielle Erkrankungen als ein komplexes Unterfangen charakterisiert werden. Zurückzuführen ist dies auf die Umstände, dass die krankheitsprädisponierenden Gene oftmals durch einen moderaten Risikobeitrag gekennzeichnet sind, Gen-Gen- und Gen-Umwelt-Interaktionen in der Krankheitsentstehung und dem Krankheitsverlauf eine Rolle spielen, das Studiendesign häufig eine geringe statistische Kraft besitzt und eine Heterogenität der genetischen und umweltbedingten Krankheitsdeterminanten zwischen den Studienpopulationen existiert [Cardon und Bell, 2001; Risch und Merikangas, 1996; Terwilliger und Goring, 2000]. Umso stärker wird dadurch die Bedeutung des Studiendesigns in Hinblick auf einen uniformen Manifestationsgrad der Patienten und der Verwendung gleicher Phänotypisierungskriterien für die zu untersuchende Krankheit. Zusätzlich gilt es einer Geschlechter- und Altersanpassung innerhalb und zwischen den Studienpopulationen eine besondere Aufmerksamkeit zukommen zu lassen, da z.B. bei unterschiedlichen Maxima des statistischen Ausbruchsalter einer Krankheit die altersabhängige Ätiologie auf unterschiedliche Risikogene hinweisen kann und in der Studienplanung berücksichtigt werden muss. Desweiteren muss eine Replikation in einer Population gleicher Ethnie erfolgen, da die Allelfrequenz von krankheitsrelevanten Risikoallelen zwischen Ethnien bzw. „Rassen“ stark variiert [Goldstein und Hirschhorn, 2004; Ioannidis *et al.*, 2004] bis hin zur kompletten Abwesenheit von krankheitsassoziierten Haplotypen oder Einzelmutationen in Populationen unterschiedlicher ethnischer Herkunft [Croucher *et al.*, 2003]. Ist die Risikovariante allerdings in der Population präsent, bleibt der biologische Eintrag und damit die Krankheitsprädisposition in ihrer Effektstärke jenseits der „Rassenschranken“ konstant [Goldstein und Hirschhorn, 2004; Ioannidis *et al.*, 2004].

Die in der 16q-Studie verwendeten DNA-Proben wurden unter der Verwendung einheitlicher Phänotypisierungs- und Ausschlusskriterien in Form klinischer Standarddiagnoseverfahren bzw. einer umfangreichen, fragebogenbasierten Datenerhebung rekrutiert (siehe 2.1). Gleichzeitig wurde insbesondere für die CC-Studie eine adäquate Geschlechterverteilung in den Patienten- und Kontrollgruppen angestrebt. Somit muss die fehlende Replikation der Assoziationssignale für die Gene *ZNF19*, *ZNF23*, *ZFP90* und *GPR56* in erster Linie mit einer erhöhten Wahrscheinlichkeit eines Typ I Fehlers betrachtet werden. Da die verwendete Replikationspopulation in ihrem Umfang in etwa der ersten Studiengruppe entspricht, würde eine deutliche Vergrößerung der Replikationspopulation um mindestens den Faktor 2-3 nicht nur auf eine höhere statistische Gewichtung abzielen, sondern auch die Validität des ersten Assoziationsbefundes klarer von einer unechten Erhöhung des Krankheitsrisikos abgrenzen können. Demgegenüber stände jedoch die

erschwerte praktische Umsetzung in Hinblick auf die Kostenerhöhung und Verfügbarkeit des Probenmaterials.

4.2.5 Vergleich der familienbasierten und populationsbasierten Studie

Betrachtet man in einer zusammenfassenden Gegenüberstellung die beiden unterschiedlichen Studienansätze und Auswertestatistiken, insbesondere vor dem Hintergrund der oben aufgeführten Einflussgrößen, so besitzen grundsätzlich beide Methoden ein gleichstarkes statistisches Potential, um Krankheitsassoziation aufzudecken [Ackerman *et al.*, 2005; Risch, 2000]. Die Vorteile der TDT-Analyse liegen dabei in erster Linie in der Resistenz gegenüber einer Populationsstratifikation, was in der starken Zunahme von familienbasierten Studien in den letzten Jahren zum Ausdruck kommt [Cardon und Palmer, 2003]. Auch kann der TDT unter Verwendung der nicht-transmittierten Allele als Kontrollgruppe zur direkten Risikoabschätzung genutzt werden [Ackerman *et al.*, 2005; Falk und Rubinstein, 1987], wengleich nur unter der Prämisse nach Spielman *et al.* (1993), dass die nichtvererbten Allele die Grundpopulation repräsentieren. Auf Seiten der Nachteile finden sich dagegen die schwierigere Rekrutierung der Eltern-Kind-Trios und die Herabsetzung der statistischen Aussagekraft bei gleichem Probenumfang, da nur ein Teil der Genotypen in die Statistik eingeht [Morton und Collins, 1998]. Eine weitere Einschränkung ist der Umstand, dass ausschließlich heterozygote Elternteile auswertbar sind, was im Falle von seltenen, niederfrequenten Markern nach einem großen Probenumfang verlangt. Der für die Assoziationsstudie sicherlich schwerwiegendste Störfaktor in der TDT-Statistik ist allerdings die gerichtete Transmission (*transmission distortion*) [Zollner *et al.*, 2004], die auch für Kopplungsanalysen von Relevanz ist [Edwards, 2003; Greenwood und Morgan, 2000]. Hierunter ist eine signifikante Abweichung von der unter Abwesenheit selektiver Faktoren zu erwartenden zufälligen Vererbung der maternalen und paternalen Chromosomen zu verstehen. Es ist erwiesen, dass in Teilen des menschlichen Genoms der genetische Anteil, den sich Geschwister teilen, signifikant über dem nach Mendel zu erwartenden liegt (50,43% vs. 50%) [Zollner *et al.*, 2004]. Ursächlich hierfür kann zum einen die in-utero (postzygotische) Selektion gegen schädliche Mutationen und damit für oder gegen bestimmte Genotypen sein. Die meiotische Drift in der weiblichen Keimbahn, die durch eine asymmetrische Zellteilung der Oocyten, einer funktionellen Asymmetrie der meiotischen Pol-Spindeln und einer Heterozygotie der Spindelanheftungsloci gekennzeichnet ist, stellt eine weitere Quelle der gerichteten Transmission auf der Basis nichtzufälliger Chromosomenverteilungen in der meiotischen Zellteilung dar [LeMaire-Adkins und Hunt, 2000; Lyttle, 1993; Pardo-Manuel de Villena und Sapienza, 2001]. Ebenso wirkt sich eine

maternal-fötale Inkompatibilität oder der unterschiedliche Befruchtungserfolg der Spermatozoen (gametische Selektion) gerichtet auf den Vererbungsmechanismus aus [Piccinni *et al.*, 1998; Westendorp *et al.*, 2001]. Besteht der Verdacht, dass die TDT-Statistik durch die Anwesenheit einer gerichteten Transmission beeinflusst ist, kann sich zum einen über die Bestimmung des Transmissionsverhältnis in Kontroll-Trios der gesunden Bevölkerung Klarheit verschafft werden [Eaves *et al.*, 1999; Naumova *et al.*, 2001; Zollner *et al.*, 2004]. Als Alternative kann eine krankheitsunabhängige Grundtransmission durch die Bestimmung und den Vergleich der Allelfrequenzen der nicht-transmittierten Elternallele mit den Allelfrequenzen einer ethnisch und altersangepassten Kontrollpopulation ermittelt werden [Ackerman *et al.*, 2005]. Dabei gilt die Grundannahme, dass die Allelfrequenz der nicht-transmittierten Allele bei Abwesenheit von gerichteter Vererbung der Allelfrequenz der Kontrollpopulation entspricht [Spielman *et al.*, 1993; Thomson, 1995].

Die Möglichkeiten und Einschränkungen der populationsbasierten Studie zeigen sich im Vergleich zum familienbasierten Ansatz in ähnlicher Mannigfaltigkeit. In Bezug auf die relative statistische Aussagekraft des CC-Designs gegenüber der TDT-Statistik, zeichnet sich der populationsbasierte Ansatz durch eine höhere Effizienz aus. Insbesondere in pharmakogenetischen Studien und bei dem Auffinden moderater Risikogene wird der populationsbasierte Ansatz deshalb bevorzugt [Risch und Teng, 1998; Risch, 2000]. Weitere Vorteile stellen die gute methodische Etablierung durch den Gebrauch in der Epidemiologie und eine effizientere Probenrekrutierung dar. Ebenso besteht im Gegensatz zur familienbasierten Studie die Möglichkeit, Krankheiten zu studieren, deren Ausbruch vornehmlich im fortgeschrittenen Alter angesiedelt ist. Demgegenüber kommen die limitierenden Faktoren der CC-Analyse in einem hohen Maß an inkonsistenten, nicht replizierenden Assoziationen zum Ausdruck [Terwilliger und Goring, 2000; Weiss und Terwilliger, 2000]. Die Ursachen hierfür liegen u.a. in der biologischen und phänotypischen Komplexität multifaktorieller Erkrankungen, der Populationsspezifität von LD-Strukturen, den oftmals zu gering dimensionierten Probensammlungen und der Anfälligkeit gegenüber Populationstratifikation [Ioannidis *et al.*, 2001; Ioannidis *et al.*, 2003; Risch, 2000; Risch und Zhang, 1996; Terwilliger und Goring, 2000; Weiss und Terwilliger, 2000]. Zumindest letzteres könnte durch eine familienbasierte CC-Studie, in der die Kontrollpopulation aus Familienangehörigen besteht, umgangen werden [Gauderman *et al.*, 1999; Witte *et al.*, 1999]. Allerdings würde durch diese Form der Überanpassung die Effizienz der Risikoabschätzung reduziert, da eine höhere genetische Ähnlichkeit zwischen Fällen und Kontrollen existieren würde; bei der Verwendung gesunder Geschwister als Kontrollen würde dieses eine Einbusse von bis zu 50% bedeuten [Risch und Teng, 1998].

Wägt man die Vor- und Nachteile der beiden Studienformen gegeneinander ab, so kann man zu dem Entschluss gelangen, dass eine ideale Assoziationsstudie beide

Methoden beinhalten sollte. Somit lässt sich das Risiko, ein falsch positives Assoziationssignal zu verfolgen und entsprechende Ressourcen zu vergeuden, deutlich reduzieren. In der vorliegenden Studie zeigt sich deutlich, dass wie im Falle der assoziierten Gene *ZFP90* und *GPR56* beide Teststatistiken unabhängig voneinander ein statistisches Signal erzeugen können. Betrachtet man nun die CED-Krankheitsgene *NOD2* und *DLG5* [Hampe, Cuthbert *et al.*, 2001; Hugot *et al.*, 2001; Ogura *et al.*, 2001; Stoll *et al.*, 2004] als Referenzen für starke und moderate Risikogene, so muss man in einer Assoziationsstudie, die sowohl den familien- als auch populationsbasierten Ansatz beinhaltet, ein valides Assoziationssignal in beiden Teststatistiken wiederfinden.

4.3 Schlussfolgerung und Ausblick der Studie

Die in dieser Assoziationsstudie verwendete SNP-Auswahl und Probensammlung führte trotz eines umfangreichen und robusten Studiendesigns nicht zur eindeutigen Identifizierung eines neuen Krankheitsgens für die CED in der von Hampe *et al.* (2002), van Heel *et al.* (2003) und Annese *et al.* (2003) postulierten *NOD2*-unabhängigen Kopplungsregion des Chromosom 16. Das Gen *BRD7* konnte als einziges von fünf in der Studie assoziierten und postulierten Kandidatengenen in einer unabhängigen Replikationspopulation bestätigt werden. Jedoch konnte keine Assoziation von funktionellen Risikoallelen auf der Basis codierender Mutationen oder Promotervariationen und den Krankheitsphänotypen CED, MC oder CU entdeckt werden. Hinweise aus aktuellen Studien zeigen jedoch, dass dieses kein absoluter Ausschlussfaktor für die Krankheitsrelevanz eines Kandidatengens ist, was beispielsweise durch die zunehmende Entdeckung krankheitsrelevanter Splice-Mutationen verdeutlicht wird [Valentonyte *et al.*, 2005; Wessagowit *et al.*, 2005].

Die relative physikalische Nähe des *BRD7*-Locus zu dem MC-Risikogen *NOD2* [Annese *et al.*, 2003; Hampe *et al.*, 2002; Hugot *et al.*, 2001; Ogura *et al.*, 2001] muss als potentieller Einflussfaktor auf die *BRD7*-Assoziation betrachtet werden. Eine statistische Überprüfung zur Abschätzung dieser *NOD2*-Relevanz weist auf eine unabhängige Assoziation des *BRD7*-Locus mit dem Phänotyp CED und CU hin. Letzterer ist nicht mit *NOD2* assoziiert, sodass Folgeexperimente zur Untersuchung eines möglichen Risikobeitrages des *BRD7*-Locus und der umgebenden Gene für den Phänotyp CU den vorliegenden Assoziationsbefund untermauern könnten. Insbesondere die Replikation der *BRD7*-Assoziation in weiteren, unabhängigen Studienpopulationen wäre hierbei richtungsweisend.

Ein limitierender Faktor für das Studiendesign und damit auch für die Erfolgsaussichten der durchgeführten Studie stellte die begrenzte Verfügbarkeit von

validierten Markern in den öffentlichen Datenbanken dar. Dieses kommt in der zum Teil heterogenen Markerverteilung zum Ausdruck und wird besonders bei der Berechnung der LD-Strukturen durch größere Regionen geringer Markerkopplung deutlich (Abb. 3.2). Im Oktober 2002 belief sich die Anzahl annotierter SNPs für das humane Genom in der NCBI Datenbank auf 2,8 Mio., von denen 500.000 durch Sequenzier- und Genotypisierungsdaten validiert waren. Drei Jahre später - im Oktober 2005 - konnten Assoziationsstudien dagegen auf 10,4 Mio. annotierte und 4,9 Mio. validierte Marker zurückgreifen. Die hierdurch bedingte bessere Auflösungsöglichkeit von LD-Strukturen wird in zukünftigen Assoziationsstudien die Chancen auf die Entdeckung neuer genetischer Risikofaktoren maßgeblich erhöhen [Lawrence, Evans und Cardon, 2005]. Eine weitere und entscheidende Neuerung ist die Entdeckung der humanen Haplotypblockstrukturen, die zur Entstehung des HapMap-Projektes geführt haben [HapMap-Consortium, 2003]. Ziel des Projektes ist es, die Haplotypblöcke im humanen Genom durch die Genotypisierung mehrerer Millionen SNPs in Populationen europäischen, afrikanischen und asiatischen Ursprungs zu charakterisieren. Über die öffentliche Verfügbarkeit der Genotypdaten kann eine effizientere Studienplanung für die Kartierung von Krankheitsgenen erfolgen, indem die Auswahl der Marker ausschließlich auf die vollständige Erfassung der blockbildenden Haplotypen reduziert wird, und redundante SNPs ohne zusätzlichen statistischen Erkenntnisgewinn ausgeschlossen werden [Johnson *et al.*, 2001; Kruglyak, 2005]. Die Verwendung solcher haplotyp-charakterisierenden htSNPs (ht = *haplotype tagging*) ermöglicht eine hohe Ersparnis bei den Genotypisierungskosten und ein statistisch aussagekräftigeres Studiendesign. Neue Studienansätze wie z.B. die fokussierte Genotypisierung codierender SNPs (cSNPs), die auf die Aufdeckung direkter Krankheitsassoziationen ausgerichtet ist [Hughes, Packer, Welch, Bergen *et al.*, 2005; Hughes, Packer, Welch, Chanock *et al.*, 2005; Wang *et al.*, 2005], oder die gezielte Genotypisierung von Markern, die in Genen krankheitsrelevanter und/oder pharmakologisch auswertbarer Stoffwechselfade lokalisiert sind [Iida *et al.*, 2006; Roses, 2004] können ebenfalls einen entscheidenden Beitrag bei der zukünftigen, assoziationsbasierten Entdeckung von krankheitsrelevanten Genmutationen leisten. Desweiteren gehören auch Neuerungen wie die Durchführung genomweiter Genexpressionsstudien [Costello *et al.*, 2005; Stranger *et al.*, 2005] und die Entwicklung zeit- und kostensparender Genotypisierungsmethoden in Form der *chiparray*- oder multiplexbasierten Techniken [Craig *et al.*, 2005; De la Vega *et al.*, 2005; Gunderson *et al.*, 2005] zu den neuen ‚Werkzeugen‘ der Assoziationsstudie. Auf der Grundlage dieser neuen Möglichkeiten wäre es denkbar, in einer Folgestudie dem Hinweis auf ein *NOD2*-unabhängiges Risikogen in der chr16q-Region durch ein ergänzendes Studiendesign nachzugehen.

5 Zusammenfassung

Die chronisch entzündlichen Darmerkrankungen (CED) zählen mit ihren beiden Subphänotypen Morbus Crohn (MC) und Colitis ulcerosa (CU) zu den komplexen Erkrankungen, denen eine polygenetische und umweltbedingte Ätiologie zu Grunde liegt. Das bisher stärkste Risikogen für die CED konnte in Form von *NOD2* auf der Basis einer Kopplungsregion (IBD1) auf dem Chromosom 16 identifiziert werden. Eine Stratifizierung gegen die *NOD2*-Riskovarianten in unabhängigen Kopplungsanalysen erbrachte den Hinweis auf ein weiteres, *NOD2*-unabhängiges Krankheitsgen auf dem Chromosom 16q; dieses galt es in einer Kandidatengenstudie zu identifizieren. In einem umfassenden Feinkartierungsexperiment wurden 436 SNP-Marker einer 35Mb großen Region in insgesamt 6.688 Individuen genotypisiert. Das Studiendesign war als familien- und populationsbasierte Assoziationsstudie angelegt und die Auswertung über die TDT- und Fall-Kontroll-Statistik konnte die Gene *BRD7*, *GPR56*, *ZFP90*, *ZNF19* und *ZNF23* als Kandidatengene ausweisen. Ein 5-Marker-Haplotyp des *ZFP90*-Gens zeigte eine deutliche CED-Assoziation ($p=0,00035$) basierend auf einer gerichteten Vererbung in den Patienten. Desweiteren konnte eine statistische Erhöhung des Krankheitsrisikos für den Phänotyp MC um den Faktor 1,4 ($OR=0,74$, 95%-KI 0,61-0,91) für einzelne Varianten der Gene *ZNF19* und *ZNF23* identifiziert werden. Zwei Mutationen der 5'-UTR-Region des *GPR56*-Locus zählten durch eine starke allelische Assoziation in der Fall-Kontroll-Statistik ($p=0,0009$ und $p=0,0008$) ebenfalls zu den ermittelten Riskovarianten für diesen Krankheitsphänotyp. Das Gen *BRD7* erhielt den bedeutesten Kandidatengenstatus in dieser Studie, da eine Replikation der assoziierten Genvarianten in einer unabhängigen Studienpopulation erbracht werden konnte. Fünf Einzelbasenpolymorphismen (SNPs) des *BRD7*-Locus zeigten sowohl in der familien- als auch populationsbasierten Teststatistik eine starke Assoziation mit dem Phänotyp CED ($p=0,006-0,003$) und eine moderate Bedeutung für die Subphänotypen MC und CU ($p=0,03-0,003$). Die physikalische Nähe (328 Kb) des *BRD7*-Locus zu dem MC-Krankheitsgen *NOD2* wurde als potentieller Einflussfaktor auf die *BRD7*-Assoziation betrachtet und mittels einer LD-Block-basierten Genotypenpaarung der Patienten und Kontrollindividuen als nicht relevant für den Phänotyp CED und CU beurteilt. Die D' -basierte LD-Analyse der *BRD7*-*NOD2*-Region unterstrich diese Annahme durch eine schwache Kopplung der beiden Genloci ($D' \leq 0,2$). Einen funktionellen Risikostatus konnte auf der Basis der vorliegenden Untersuchung bisher keinem der fünf Gene zugesprochen werden. Folgestudien zur Untersuchung der Krankheitsrelevanz des *BRD7*-Locus für die Colitis ulcerosa könnten jedoch insbesondere vor dem Hintergrund der für diesen Phänotyp nicht vorhandenen Assoziation des *NOD2*-Krankheitsgens von größter Bedeutung sein.

6 Summary

Inflammatory bowel disease (IBD) includes two clinical subtypes: Crohn's disease (CD) and ulcerative colitis (UC). IBD is a multifactorial disorder characterized by chronic inflammation of the gastrointestinal tract. The etiology of IBD is supposed to be based on a complex interaction of polygenetic and environmental factors. Based on a linkage region on chromosome 16 (IBD1) an innate immunity gene, namely *NOD2*, has been identified as a strong risk gene for CD. Stratification against the disease causing *NOD2* variants in further linkage analyses led to the assumption of an additional, *NOD2*-independent susceptibility locus for IBD on chromosome 16q. Therefore, the *NOD2*-independent linkage region has been investigated in the present candidate gene study. Based on a positional cloning approach a 35Mb region on chr16q was fine-mapped by genotyping 436 SNP markers in a total of 6,688 individuals. The study design implemented a family- and a population-based approach, and TDT and case-control statistics identified 5 candidate genes: *BRD7*, *GPR56*, *ZFP90*, *ZNF19* and *ZNF23*. A 5-marker-haplotype of *ZFP90* was identified in IBD patients by significant transmission distortion ($p=0,00035$). Furthermore, variants of *ZNF19* and *ZNF23* were associated with CD by a significant increase of the disease risk of 1.4 (OR=0,74, 95%-CI 0,61-0,91). Two polymorphisms in the 5'-UTR region of the *GPR56* locus also showed strong allelic association with CD in the case-control statistics ($p=0.0009$ and $p=0.0008$, respectively). In this study, the most promising candidate gene was *BRD7* due to association replication in an independent study group of patients and controls. Briefly, 5 SNPs of the *BRD7* locus revealed a strong association with IBD in family- and population-based statistics ($p=0.006-0.003$). Additionally, moderate significance was shown for CD and UC ($p=0.03-0.003$). *BRD7* is located within a close distance to *NOD2* (328 kb), but LD calculations indicated negligible LD between the two loci ($D' \leq 0,2$). An additional in-depth analysis based on LD-based genotype matching demonstrated a *NOD2*-independency of the *BRD7*-association for IBD and UC. Functional evidence for the 5 candidate genes could not be demonstrated by the present investigations, but further studies regarding the relevance of *BRD7* as a diseases gene for UC are promising, particularly with regard to the lack of association of *NOD2* and UC. Therefore, main emphasis should be placed on signal replication of the *BRD7* association.

7 Literatur

- Abecasis, G.R. et al. Extent and distribution of linkage disequilibrium in three genomic regions. *Am J Hum Genet* **68**, 191-197 (2001).
- Ackerman, H. et al. A comparison of case-control and family-based association methods: the example of sickle-cell and malaria. *Ann Hum Genet* **69**, 559-65 (2005).
- Afonina, I. et al. Efficient priming of PCR with short oligonucleotides conjugated to a minor groove binder. *Nucleic Acids Res* **25**, 2657-60 (1997).
- Agresti, A. An introduction to categorical data analysis. 227-228 (1996).
- Ahn, S.J., Costa, J. & Emanuel, J.R. PicoGreen quantitation of DNA: effective evaluation of samples pre- or post-PCR. *Nucleic Acids Res* **24**, 2623-5 (1996).
- Annese, V. et al. Linkage of ulcerative colitis to the pericentromeric region of chromosome 16 in Italian inflammatory bowel disease families is independent of the presence of common CARD15 mutations. *J Med Genet* **40**, 837-41 (2003).
- Ardlie, K.G., Kruglyak, L. & Seielstad, M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* **3**, 299-309 (2002).
- Bacanu, S.A., Devlin, B. & Roeder, K. The power of genomic control. *Am J Hum Genet* **66**, 1933-44 (2000).
- Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-5 (2005).
- Bearden, J., Jr. Isolation of nucleolar proteins by ethanol precipitation of nucleic acids. *Biochim Biophys Acta* **361**, 109-13 (1974).
- Berthold, V. & Geider, K. Interaction of DNA with DNA-binding proteins. The characterization of protein HD from *Escherichia coli* and its nucleic acid complexes. *Eur J Biochem* **71**, 443-9 (1976).
- Binder, V. Genetic epidemiology in inflammatory bowel disease. *Dig Dis* **16**, 351-5 (1998).
- Binder, V. Epidemiology of IBD during the twentieth century: an integrated view. *Best Pract Res Clin Gastroenterol* **18**, 463-79 (2004).
- Blaser, K., Jutel, M. & Akdis, C.A. The role of histamine in regulation of immune responses. *Chem Immunol Allergy* **91**, 174-87 (2006).
- Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* **33**, 228-37. (2003).
- Bray, P., Lichter, P., Thiesen, H.J., Ward, D.C. & Dawid, I.B. Characterization and mapping of human genes encoding zinc finger proteins. *Proc Natl Acad Sci U S A* **88**, 9563-7 (1991).

- Brown, R.S. Zinc finger proteins: getting a grip on RNA. *Curr Opin Struct Biol* **15**, 94-8 (2005).
- Cannizzaro, L.A., Aronson, M.M. & Thiesen, H.J. Human zinc finger gene ZNF23 (Kox16) maps to a zinc finger gene cluster on chromosome 16q22, and ZNF32 (Kox30) to chromosome region 10q23-q24. *Hum Genet* **91**, 383-5 (1993).
- Carballo, E. et al. Decreased sensitivity of tristetraprolin-deficient cells to p38 inhibitors suggests the involvement of tristetraprolin in the p38 signaling pathway. *J Biol Chem* **276**, 42580-7 (2001).
- Carballo, E., Lai, W.S. & Blakeshear, P.J. Feedback inhibition of macrophage tumor necrosis factor- α production by tristetraprolin. *Science* **281**, 1001-5 (1998).
- Cardon, L.R. & Bell, J.I. Association study designs for complex diseases. *Nat Rev Genet* **2**, 91-9 (2001).
- Cardon, L.R. & Palmer, L.J. Population stratification and spurious allelic association. *Lancet* **361**, 598-604 (2003).
- Carlson, C.S. et al. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* **74**, 106-20 (2004).
- Cavalli-Sforza, L.L. & Piazza, A. Human genomic diversity in Europe: a summary of recent research and prospects for the future. *Eur J Hum Genet* **1**, 3-18 (1993).
- Cave, N.J. Chronic inflammatory disorders of the gastrointestinal tract of companion animals. *N Z Vet J* **51**, 262-74 (2003).
- Chakravarti, A. Population genetics--making sense out of sequence. *Nat Genet* **21**, 56-60 (1999).
- Chamaillard, M., Girardin, S.E., Viala, J. & Philpott, D.J. Nods, Nalps and Naip: intracellular regulators of bacterial-induced inflammation. *Cell Microbiol* **5**, 581-92 (2003).
- Clayton, D. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am J Hum Genet* **65**, 1170-7. (1999).
- Cong, P. et al. The transcriptional regulation role of BRD7 by binding to acetylated histone through bromodomain. *J Cell Biochem* (2005).
- Cooper, P., Potter, S., Mueck, B., Yousefi, S. & Jarai, G. Identification of genes induced by inflammatory cytokines in airway epithelium. *Am J Physiol Lung Cell Mol Physiol* **280**, L841-52 (2001).
- Costello, C.M. et al. Dissection of the Inflammatory Bowel Disease Transcriptome Using Genome-Wide cDNA Microarrays. *PLoS Medicine* **2**(2005).
- Craig, D.W. et al. Identification of disease causing loci using an array-based genotyping approach on pooled DNA. *BMC Genomics* **6**, 138 (2005).

- Croucher, P.J. et al. Haplotype structure and association to Crohn's disease of CARD15 mutations in two ethnically divergent populations. *Eur J Hum Genet* **11**, 6-16 (2003).
- Cuthbert, A.P. et al. The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. *Gastroenterology* **122**, 867-74 (2002).
- de Esch, I.J., Thurmond, R.L., Jongejan, A. & Leurs, R. The histamine H4 receptor as a new therapeutic target for inflammation. *Trends Pharmacol Sci* **26**, 462-9 (2005).
- De la Vega, F.M., Lazaruk, K.D., Rhodes, M.D. & Wenz, M.H. Assessment of two flexible and compatible SNP genotyping platforms: TaqMan SNP Genotyping Assays and the SNPLex Genotyping System. *Mutat Res* **573**, 111-35 (2005).
- Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
- Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* **60**, 155-66 (2001).
- Dickersin, K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* **9**, 15-21 (1997).
- Dickersin, K. & Min, Y.I. Publication bias: the problem that won't go away. *Ann N Y Acad Sci* **703**, 135-46; discussion 146-8 (1993).
- Dieffenbach, C.W., Lowe, T.M. & Dveksler, G.S. General concepts for PCR primer design. *PCR Methods Appl* **3**, S30-7. (1993).
- Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. & Mattick, J.S. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* **19**, 4008 (1991).
- Douglas, J.A., Skol, A.D. & Boehnke, M. Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. *Am J Hum Genet* **70**, 487-95 (2002).
- Duerr, R.H., Barmada, M.M., Zhang, L., Pfutzer, R. & Weeks, D.E. High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. *Am J Hum Genet* **66**, 1857-62 (2000).
- Eaves, I.A. et al. Transmission ratio distortion at the INS-IGF2 VNTR. *Nat Genet* **22**, 324-5 (1999).
- Eckmann, L. & Karin, M. NOD2 and Crohn's disease: loss or gain of function? *Immunity* **22**, 661-7 (2005).
- Edwards, J.H. Sib-pairs in multifactorial disorders: the sib-similarity problem. *Clin Genet* **63**, 1-9 (2003).
- Evans, P.C. et al. Zinc-finger protein A20, a regulator of inflammation and cell survival, has de-ubiquitinating activity. *Biochem J* **378**, 727-34 (2004).

- Ewen, K.R. et al. Identification and analysis of error types in high-throughput genotyping. *Am J Hum Genet* **67**, 727-36 (2000).
- Falk, C.T. & Rubinstein, P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet* **51**, 227-33. (1987).
- Feinstein, R.E. & Olsson, E. Chronic gastroenterocolitis in nine cats. *J Vet Diagn Invest* **4**, 293-8 (1992).
- Filetici, P., P, O. & Ballario, P. The bromodomain: a chromatin browser? *Front Biosci* **6**, D866-76 (2001).
- Flower, D.R. Modelling G-protein-coupled receptors for drug design. *Biochim Biophys Acta* **1422**, 207-34 (1999).
- Fredriksson, R., Lagerstrom, M.C., Lundin, L.G. & Schioth, H.B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **63**, 1256-72 (2003).
- Freedman, M.L. et al. Assessing the impact of population stratification on genetic association studies. *Nat Genet* **36**, 388-93 (2004).
- Gabriel, S.B. et al. The structure of haplotype blocks in the human genome. *Science* **296**, 2225-9 (2002).
- Gauderman, W.J., Witte, J.S. & Thomas, D.C. Family-based association studies. *J Natl Cancer Inst Monogr*, 31-7 (1999).
- Gelernter, J., Goldman, D. & Risch, N. The A1 allele at the D2 dopamine receptor gene and alcoholism. A reappraisal. *Jama* **269**, 1673-7 (1993).
- Girardin, S.E., Hugot, J.P. & Sansonetti, P.J. Lessons from Nod2 studies: towards a link between Crohn's disease and bacterial sensing. *Trends Immunol* **24**, 652-8 (2003).
- Goddard, K.A., Hopkins, P.J., Hall, J.M. & Witte, J.S. Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. *Am J Hum Genet* **66**, 216-34 (2000).
- Goldstein, D.B. Islands of linkage disequilibrium. *Nat Genet* **29**, 109-11 (2001).
- Goldstein, D.B. & Hirschhorn, J.N. In genetic control of disease, does 'race' matter? *Nat Genet* **36**, 1243-4 (2004).
- Greenwood, C.M. & Morgan, K. The impact of transmission-ratio distortion on allele sharing in affected sibling pairs. *Am J Hum Genet* **66**, 2001-4 (2000).
- Guilford, W.G. & Matz, M.E. The nutritional management of gastrointestinal tract disorders in companion animals. *N Z Vet J* **51**, 284-91 (2003).
- Gunderson, K.L., Steemers, F.J., Lee, G., Mendoza, L.G. & Chee, M.S. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* **37**, 549-54 (2005).
- Hamm, H.E. The many faces of G protein signaling. *J Biol Chem* **273**, 669-72 (1998).

- Hampe, J. et al. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* **357**, 1925-8 (2001).
- Hampe, J. et al. Evidence for a NOD2-independent susceptibility locus for inflammatory bowel disease on chromosome 16p. *Proc Natl Acad Sci U S A* **99**, 321-6. Epub 2001 Dec 18. (2002).
- Hampe, J. et al. A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am J Hum Genet* **64**, 808-16 (1999).
- Hampe, J. et al. An integrated system for high throughput TaqMan based SNP genotyping. *Bioinformatics* **17**, 654-5 (2001).
- HapMap-Consortium, T.I. The International HapMap Project. *Nature* **426**, 789-96 (2003).
- Heath, E.M., O'Brien, D.P., Banas, R., Naylor, E.W. & Dobrowolski, S. Optimization of an automated DNA purification protocol for neonatal screening. *Arch Pathol Lab Med* **123**, 1154-60 (1999).
- Hill, W.G. & Robertson, A. Linkage disequilibrium in finite populations. *Theor Appl Genet* **38**, 226-231 (1968).
- Hirschhorn, J.N. Genetic approaches to studying common diseases and complex traits. *Pediatr Res* **57**, 74R-77R (2005).
- Hirschhorn, J.N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genet Med* **4**, 45-61 (2002).
- Holland, P.M., Abramson, R.D., Watson, R. & Gelfand, D.H. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc Natl Acad Sci U S A* **88**, 7276-80. (1991).
- Hong, J.W., Allen, C.E. & Wu, L.C. Inhibition of NF-kappaB by ZAS3, a zinc-finger protein that also binds to the kappaB motif. *Proc Natl Acad Sci U S A* **100**, 12301-6 (2003).
- Hong, J.W. & Wu, L.C. Structural characterization of the gene encoding the large zinc finger protein ZAS3: implication to the origin of multiple promoters in eukaryotic genes. *Biochim Biophys Acta* **1681**, 74-87 (2005).
- Hosking, L. et al. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *Eur J Hum Genet* **12**, 395-9 (2004).
- Hostutler, R.A. et al. Antibiotic-responsive histiocytic ulcerative colitis in 9 dogs. *J Vet Intern Med* **18**, 499-504 (2004).
- Hughes, A.L. et al. Effects of natural selection on interpopulation divergence at polymorphic sites in human protein-coding Loci. *Genetics* **170**, 1181-7 (2005).
- Hughes, A.L., Packer, B., Welch, R., Chanock, S.J. & Yeager, M. High level of functional polymorphism indicates a unique role of natural selection at human immune system loci. *Immunogenetics* **57**, 821-7 (2005).

- Hugot, J.P. Inflammatory bowel disease: a complex group of genetic disorders. *Best Pract Res Clin Gastroenterol* **18**, 451-62 (2004).
- Hugot, J.P. et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599-603 (2001).
- Iida, A. et al. Japanese single nucleotide polymorphism database for 267 possible drug-related genes. *Cancer Sci* **97**, 16-24 (2006).
- Inohara, N. et al. Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J Biol Chem* **278**, 5509-12 (2003).
- Ioannidis, J.P., Ntzani, E.E. & Trikalinos, T.A. 'Racial' differences in genetic effects for complex diseases. *Nat Genet* **36**, 1312-8 (2004).
- Ioannidis, J.P., Ntzani, E.E., Trikalinos, T.A. & Contopoulos-Ioannidis, D.G. Replication validity of genetic association studies. *Nat Genet* **29**, 306-9 (2001).
- Ioannidis, J.P., Trikalinos, T.A., Ntzani, E.E. & Contopoulos-Ioannidis, D.G. Genetic associations in large versus small studies: an empirical assessment. *Lancet* **361**, 567-71 (2003).
- Jansen, A. & Andermann, E. Genetics of the polymicrogyria syndromes. *J Med Genet* **42**, 369-78 (2005).
- Jayanthi, V., Probert, C.S., Pinder, D., Wicks, A.C. & Mayberry, J.F. Epidemiology of Crohn's disease in Indian migrants and the indigenous population in Leicestershire. *Q J Med* **82**, 125-38 (1992).
- Jess, T. et al. Disease concordance, zygosity, and NOD2/CARD15 status: follow-up of a population-based cohort of Danish twins with inflammatory bowel disease. *Am J Gastroenterol* **100**, 2486-92 (2005).
- John, H.A., Birnstiel, M.L. & Jones, K.W. RNA-DNA hybrids at the cytological level. *Nature* **223**, 582-7 (1969).
- Johnson, G.C. et al. Haplotype tagging for the identification of common disease genes. *Nat Genet* **29**, 233-7 (2001).
- Jorde, L.B. Linkage disequilibrium and the search for complex disease genes. *Genome Res* **10**, 1435-44 (2000).
- Jorde, L.B. et al. Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. *Am J Hum Genet* **54**, 884-98 (1994).
- Jutel, M., Blaser, K. & Akdis, C.A. Histamine in allergic inflammation and immune modulation. *Int Arch Allergy Immunol* **137**, 82-92 (2005).
- Klug, A. Towards therapeutic applications of engineered zinc finger proteins. *FEBS Lett* **579**, 892-4 (2005).

- Knowler, W.C., Williams, R.C., Pettitt, D.J. & Steinberg, A.G. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am J Hum Genet* **43**, 520-6 (1988).
- Kroeze, W.K., Sheffler, D.J. & Roth, B.L. G-protein-coupled receptors at a glance. *J Cell Sci* **116**, 4867-9 (2003).
- Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* **22**, 139-44 (1999).
- Kruglyak, L. Power tools for human genetics. *Nat Genet* **37**, 1299-300 (2005).
- Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E.S. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* **58**, 1347-63 (1996).
- Kuster, W., Pascoe, L., Purmann, J., Funk, S. & Majewski, F. The genetics of Crohn disease: complex segregation analysis of a family study with 265 patients with Crohn disease and 5,387 relatives. *Am J Med Genet* **32**, 105-8 (1989).
- Kutyavin, I.V., Lukhtanov, E.A., Gamper, H.B. & Meyer, R.B. Oligonucleotides with conjugated dihydropyrroloindole tripeptides: base composition and backbone effects on hybridization. *Nucleic Acids Res* **25**, 3718-23 (1997).
- Laharie, D. et al. Inflammatory bowel disease in spouses and their offspring. *Gastroenterology* **120**, 816-9 (2001).
- Lamina, C. et al. Genetic diversity in German and European populations: looking for substructures and genetic patterns. *Gesundheitswesen* **67 Suppl 1**, S127-31 (2005).
- Lander, E.S. & Schork, N.J. Genetic dissection of complex traits. *Science* **265**, 2037-48 (1994).
- Lawrence, R. et al. Genetically indistinguishable SNPs and their influence on inferring the location of disease-associated variants. *Genome Res* **15**, 1503-10 (2005).
- Lawrence, R.W., Evans, D.M. & Cardon, L.R. Prospects and pitfalls in whole genome association studies. *Philos Trans R Soc Lond B Biol Sci* **360**, 1589-95 (2005).
- Lawrie, S.S. & Gosden, J.R. The identification of human chromosomes by quinacrine fluorescence after hybridisation in situ. *Hum Genet* **53**, 371-3 (1980).
- LeMaire-Adkins, R. & Hunt, P.A. Nonrandom segregation of the mouse univalent X chromosome: evidence of spindle-mediated meiotic drive. *Genetics* **156**, 775-83 (2000).
- Lesage, S. et al. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* **70**, 845-57 (2002).
- Lewontin, R.C. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49-67 (1964).
- Lewontin, R.C. On measures of gametic disequilibrium. *Genetics* **120**, 849-52 (1988).

- Lichter, P. et al. High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* **247**, 64-9 (1990).
- Livak, K.J., Flood, S.J., Marmaro, J., Giusti, W. & Deetz, K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl* **4**, 357-62. (1995).
- Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. & Hirschhorn, J.N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* **33**, 177-82 (2003).
- Long, J.R. et al. Patterns of linkage disequilibrium and haplotype distribution in disease candidate genes. *BMC Genet* **5**, 11 (2004).
- Lowe, T., Sharefkin, J., Yang, S.Q. & Dieffenbach, C.W. A computer program for selection of oligonucleotide primers for polymerase chain reactions. *Nucleic Acids Res* **18**, 1757-61 (1990).
- Lyttle, T.W. Cheaters sometimes prosper: distortion of mendelian segregation by meiotic drive. *Trends Genet* **9**, 205-10 (1993).
- Ma, Y. et al. A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm Bowel Dis* **5**, 271-8 (1999).
- Macdonald, T.T. & Monteleone, G. Immunity, inflammation, and allergy in the gut. *Science* **307**, 1920-5 (2005).
- Maniatis, N. et al. Positional cloning by linkage disequilibrium. *Am J Hum Genet* **74**, 846-55 (2004).
- Maniatis, N. et al. The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. *Proc Natl Acad Sci U S A* **99**, 2228-33 (2002).
- Mansfield, K.G. et al. Enteropathogenic *Escherichia coli* and ulcerative colitis in cotton-top tamarins (*Saguinus oedipus*). *J Infect Dis* **184**, 803-7 (2001).
- Mathew, C.G. & Lewis, C.M. Genetics of inflammatory bowel disease: progress and prospects. *Hum. Mol. Genet.*, ddh079 (2004).
- McGovern, D. & Ahmad, T. New IBD genes? *Gut* **54**, 1060-1 (2005).
- Meucci, G. et al. Familial aggregation of inflammatory bowel disease in northern Italy: a multicenter study. The Gruppo di Studio per le Malattie Infiammatorie Intestinali (IBD Study Group). *Gastroenterology* **103**, 514-9 (1992).
- Mirza, M.M. et al. Genetic evidence for interaction of the 5q31 cytokine locus and the CARD15 gene in Crohn disease. *Am J Hum Genet* **72**, 1018-22 (2003).
- Mort, A.J., Zhan, D. & Rodriguez, V. Use of scavenger beads to remove excess labeling reagents from capillary zone electrophoresis samples. *Electrophoresis* **19**, 2129-32 (1998).

- Morton, N.E. Linkage disequilibrium maps and association mapping. *J Clin Invest* **115**, 1425-30 (2005).
- Morton, N.E. & Collins, A. Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci U S A* **95**, 11389-93 (1998).
- Mullis, K. et al. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* **51 Pt 1**, 263-73 (1986).
- Mullis, K.B. & Faloona, F.A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* **155**, 335-50 (1987).
- Nagase, T., Kikuno, R. & Ohara, O. Prediction of the coding sequences of unidentified human genes. XXII. The complete sequences of 50 new cDNA clones which code for large proteins. *DNA Res* **8**, 319-27 (2001).
- Naumova, A.K., Greenwood, C.M. & Morgan, K. Imprinting and deviation from Mendelian transmission ratios. *Genome* **44**, 311-20 (2001).
- Neves, S.R., Ram, P.T. & Iyengar, R. G protein pathways. *Science* **296**, 1636-9 (2002).
- Newman, B. & Siminovitch, K.A. Recent advances in the genetics of inflammatory bowel disease. *Curr Opin Gastroenterol* **21**, 401-7 (2005).
- Ning, Z., Cox, A.J. & Mullikin, J.C. SSAHA: A Fast Search Method for Large DNA Databases. *Genome Res.* **11**, 1725-1729 (2001).
- Ogura, Y. et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603-6 (2001).
- Orholm, M., Binder, V., Sorensen, T.I., Rasmussen, L.P. & Kyvik, K.O. Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. *Scand J Gastroenterol* **35**, 1075-81 (2000).
- Orholm, M. et al. Investigation of inheritance of chronic inflammatory bowel diseases by complex segregation analysis. *Bmj* **306**, 20-4 (1993).
- Orholm, M. et al. Familial occurrence of inflammatory bowel disease. *N Engl J Med* **324**, 84-8. (1991).
- Ott, J. A chi-square test to distinguish allelic association from other causes of phenotypic association between two loci. *Genet Epidemiol* **2**, 79-84 (1985).
- Ott, S.J. et al. Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* **53**, 685-93 (2004).
- Palmer, L.J. & Cardon, L.R. Shaking the tree: mapping complex disease genes with linkage disequilibrium. *Lancet* **366**, 1223-34 (2005).
- Pardo-Manuel de Villena, F. & Sapienza, C. Nonrandom segregation during meiosis: the unfairness of females. *Mamm Genome* **12**, 331-9 (2001).
- Pardue, M.L. & Gall, J.G. Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc Natl Acad Sci U S A* **64**, 600-4 (1969).

- Pato, C.N., Macciardi, F., Pato, M.T., Verga, M. & Kennedy, J.L. Review of the putative association of dopamine D2 receptor and alcoholism: a meta-analysis. *Am J Med Genet* **48**, 78-82 (1993).
- Peltekova, V.D. et al. Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet* **36**, 471-5 (2004).
- Peterson, A.C. et al. The distribution of linkage disequilibrium over anonymous genome regions. *Hum Mol Genet* **4**, 887-94 (1995).
- Piao, X. et al. Genotype-phenotype analysis of human frontoparietal polymicrogyria syndromes. *Ann Neurol* **58**, 680-7 (2005).
- Piao, X. et al. G protein-coupled receptor-dependent development of human frontal cortex. *Science* **303**, 2033-6 (2004).
- Piccinni, M.P. et al. Defective production of both leukemia inhibitory factor and type 2 T-helper cytokines by decidual T cells in unexplained recurrent abortions. *Nat Med* **4**, 1020-4 (1998).
- Pritchard, J.K. & Donnelly, P. Case-control studies of association in structured or admixed populations. *Theor Popul Biol* **60**, 227-37 (2001).
- Pritchard, J.K. & Przeworski, M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet* **69**, 1-14 (2001).
- Pritchard, J.K. & Rosenberg, N.A. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* **65**, 220-8 (1999).
- Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-59 (2000).
- Pritchard, J.K., Stephens, M., Rosenberg, N.A. & Donnelly, P. Association mapping in structured populations. *Am J Hum Genet* **67**, 170-81 (2000).
- Probert, C.S. et al. Prevalence and family risk of ulcerative colitis and Crohn's disease: an epidemiological study among Europeans and south Asians in Leicestershire. *Gut* **34**, 1547-51 (1993).
- Probert, C.S., Jayanthi, V., Rampton, D.S. & Mayberry, J.F. Epidemiology of inflammatory bowel disease in different ethnic and religious groups: limitations and aetiological clues. *Int J Colorectal Dis* **11**, 25-8 (1996).
- Qin, S. et al. The chemokine receptors CXCR3 and CCR5 mark subsets of T cells associated with certain inflammatory reactions. *J Clin Invest* **101**, 746-54 (1998).
- Reich, D.E. et al. Linkage disequilibrium in the human genome. *Nature* **411**, 199-204 (2001).
- Reich, D.E. & Lander, E.S. On the allelic spectrum of human disease. *Trends Genet* **17**, 502-10 (2001).
- Rengarajan, K., Cristol, S.M., Mehta, M. & Nickerson, J.M. Quantifying DNA concentrations using fluorometry: a comparison of fluorophores. *Mol Vis* **8**, 416-21 (2002).

- Risch, N. & Merikangas, K. The Future of Genetic Studies of Complex Human Diseases. *Science* **273**, 1516-1517 (1996).
- Risch, N. & Teng, J. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. *Genome Res* **8**, 1273-88 (1998).
- Risch, N.J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847-56 (2000).
- Risch, N.J. & Zhang, H. Mapping quantitative trait loci with extreme discordant sib pairs: sampling considerations. *Am J Hum Genet* **58**, 836-43. (1996).
- Robas, N., O'Reilly, M., Katugampola, S. & Fidock, M. Maximizing serendipity: strategies for identifying ligands for orphan G-protein-coupled receptors. *Curr Opin Pharmacol* **3**, 121-6 (2003).
- Rogler, G. Update in inflammatory bowel disease pathogenesis. *Curr Opin Gastroenterol* **20**, 311-7 (2004).
- Roses, A.D. Pharmacogenetics and drug development: the path to safer and more effective drugs. *Nat Rev Genet* **5**, 645-56 (2004).
- Rousseau-Merck, M.F. et al. Chromosomal localization of 9 KOX zinc finger genes: physical linkages suggest clustering of KOX genes on chromosomes 12, 16, and 19. *Hum Genet* **92**, 583-7 (1993).
- Russell, R.K. & Satsangi, J. IBD: a family affair. *Best Pract Res Clin Gastroenterol* **18**, 525-39 (2004).
- Salas-Solano, O. et al. Routine DNA sequencing of 1000 bases in less than one hour by capillary electrophoresis with replaceable linear polyacrylamide solutions. *Anal Chem* **70**, 3996-4003 (1998).
- Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-7 (1977).
- Sasieni, P.D. From genotypes to genes: doubling the sample size. *Biometrics* **53**, 1253-61 (1997).
- Satsangi, J. et al. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* **14**, 199-202 (1996).
- Sauer, S. et al. A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nucleic Acids Res* **28**, E13 (2000).
- Saunders, K.E. et al. Novel intestinal *Helicobacter* species isolated from cotton-top tamarins (*Saguinus oedipus*) with chronic colitis. *J Clin Microbiol* **37**, 146-51 (1999).

- Schaid, D.J. & Jacobsen, S.J. Biased tests of association: comparisons of allele frequencies when departing from Hardy-Weinberg proportions. *Am J Epidemiol* **149**, 706-11 (1999).
- Schreiber, S., Rosenstiel, P., Albrecht, M., Hampe, J. & Krawczak, M. Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nat Rev Genet* **6**, 376-88. (2005).
- Shapiro, D.J. Quantitative ethanol precipitation of nanogram quantities of DNA and RNA. *Anal Biochem* **110**, 229-31 (1981).
- Shashidhar, S. et al. GPR56 is a GPCR that is overexpressed in gliomas and functions in tumor cell adhesion. *Oncogene* **24**, 1673-82 (2005).
- Shivananda, S. et al. Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). *Gut* **39**, 690-7. (1996).
- Simpson, J.W. Diet and large intestinal disease in dogs and cats. *J Nutr* **128**, 2717S-2722S (1998).
- Spielman, R.S., McGinnis, R.E. & Ewens, W.J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* **52**, 506-16. (1993).
- Stephens, J.C. et al. Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293**, 489-93 (2001).
- Stoll, M. et al. Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nat Genet* **36**, 476-80. Epub 2004 Apr 11. (2004).
- Stram, D.O. Tag SNP selection for association studies. *Genet Epidemiol* **27**, 365-74 (2004).
- Stranger, B.E. et al. Genome-Wide Associations of Gene Expression Variation in Humans. *PLoS Genet* **1**, e78 (2005).
- Suarez, B.K. & Hampe, C.L. Linkage and association. *Am J Hum Genet* **54**, 554-9; author reply 560-3. (1994).
- Tabor, H.K., Risch, N.J. & Myers, R.M. Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* **3**, 391-7. (2002).
- Takahashi, M., Ogino, T. & Baba, K. Estimation of relative molecular length of DNA by electrophoresis in agarose gel. *Biochim Biophys Acta* **174**, 183-7 (1969).
- Takeda, S., Kadowaki, S., Haga, T., Takaesu, H. & Mitaku, S. Identification of G protein-coupled receptor genes from the human genome sequence. *FEBS Lett* **520**, 97-101 (2002).
- Teare, M.D., Dunning, A.M., Durocher, F., Rennart, G. & Easton, D.F. Sampling distribution of summary linkage disequilibrium measures. *Ann Hum Genet* **66**, 223-33 (2002).

- Terwilliger, J.D. & Goring, H.H. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum Biol* **72**, 63-132 (2000).
- Terwilliger, J.D., Haghghi, F., Hiekkalinna, T.S. & Goring, H.H. A bias-ed assessment of the use of SNPs in human complex traits. *Curr Opin Genet Dev* **12**, 726-34 (2002).
- Thiel, G., Lietz, M., Bach, K., Guethlein, L. & Cibelli, G. Biological activity of mammalian transcriptional repressors. *Biol Chem* **382**, 891-902 (2001).
- Thomson, G. Mapping disease genes: family-based association studies. *Am J Hum Genet* **57**, 487-98 (1995).
- Trikalinos, T.A., Ntzani, E.E., Contopoulos-Ioannidis, D.G. & Ioannidis, J.P. Establishment of genetic associations for complex diseases is independent of early study findings. *Eur J Hum Genet* **12**, 762-9 (2004).
- Turakulov, R. & Eastal, S. Number of SNPS loci needed to detect population structure. *Hum Hered* **55**, 37-45 (2003).
- Tysk, C., Lindberg, E., Järnerot, G. & Floderus-Myrhed, B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* **29**, 990-6. (1988).
- Ueda, H. et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* **423**, 506-11 (2003).
- Valentonyte, R. et al. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat Genet* **37**, 357-64 (2005).
- van Heel, D.A. et al. The IBD6 Crohn's disease locus demonstrates complex interactions with CARD15 and IBD5 disease-associated variants. *Hum. Mol. Genet.* **12**, 2569-2575 (2003).
- Wacholder, S., Rothman, N. & Caporaso, N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst* **92**, 1151-8 (2000).
- Wang, J. et al. Preparation and analysis of cSNP chip on hepatocellular carcinoma-related genes. *Hepatobiliary Pancreat Dis Int* **4**, 398-402 (2005).
- Weiss, K.M. & Terwilliger, J.D. How many diseases does it take to map a gene with SNPs? *Nat Genet* **26**, 151-7. (2000).
- Wen, L. Two-step cycle sequencing improves base ambiguities and signal dropouts in DNA sequencing reactions using energy-transfer-based fluorescent dye terminators. *Mol Biotechnol* **17**, 135-42 (2001).
- Wessagowit, V., Nalla, V.K., Rogan, P.K. & McGrath, J.A. Normal and abnormal mechanisms of gene splicing and relevance to inherited skin diseases. *J Dermatol Sci* **40**, 73-84 (2005).

- Westendorp, R.G., van Dunne, F.M., Kirkwood, T.B., Helmerhorst, F.M. & Huizinga, T.W. Optimizing human fertility and survival. *Nat Med* **7**, 873 (2001).
- Wilcockson, J. The differential precipitation of nucleic acids and proteins from aqueous solutions by ethanol. *Anal Biochem* **66**, 64-8 (1975).
- Wild, G.E. & Rioux, J.D. Genome scan analyses and positional cloning strategy in IBD: successes and limitations. *Best Pract Res Clin Gastroenterol* **18**, 541-53 (2004).
- Winston, F. & Allis, C.D. The bromodomain: a chromatin-targeting module? *Nat Struct Biol* **6**, 601-4 (1999).
- Witte, J.S., Gauderman, W.J. & Thomas, D.C. Asymptotic bias and efficiency in case-control studies of candidate genes and gene-environment interactions: basic family designs. *Am J Epidemiol* **149**, 693-705 (1999).
- Xu, J., Turner, A., Little, J., Bleecker, E.R. & Meyers, D.A. Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: hint for genotyping error? *Hum Genet* **111**, 573-4 (2002).
- Zabel, B.A. et al. Human G protein-coupled receptor GPR-9-6/CC chemokine receptor 9 is selectively expressed on intestinal homing T lymphocytes, mucosal lymphocytes, and thymocytes and is required for thymus-expressed chemokine-mediated chemotaxis. *J Exp Med* **190**, 1241-56 (1999).
- Zavattari, P. et al. Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Hum Mol Genet* **9**, 2947-57 (2000).
- Zendman, A.J., Cornelissen, I.M., Weidle, U.H., Ruiter, D.J. & van Muijen, G.N. TM7XN1, a novel human EGF-TM7-like cDNA, detected with mRNA differential display using human melanoma cell lines with different metastatic potential. *FEBS Lett* **446**, 292-8 (1999).
- Zhang, W., Collins, A., Maniatis, N., Tapper, W. & Morton, N.E. Properties of linkage disequilibrium (LD) maps. *Proc Natl Acad Sci U S A* **99**, 17004-7 (2002).
- Zhang, X.M., Wang, X.Y., Sheng, S.R., Wang, J.R. & Li, J. Expression of tumor related genes NGX6, NAG-7, BRD7 in gastric and colorectal cancer. *World J Gastroenterol* **9**, 1729-33 (2003).
- Zhou, J. et al. BRD7, a novel bromodomain gene, inhibits G1-S progression by transcriptionally regulating some important molecules involved in ras/MEK/ERK and Rb/E2F pathways. *J Cell Physiol* **200**, 89-98 (2004).
- Zollner, S. et al. Evidence for extensive transmission distortion in the human genome. *Am J Hum Genet* **74**, 62-72. Epub 2003 Dec 15. (2004).

8 Abkürzungsverzeichnis

I_s	relatives Geschwisterrisiko
A20	Zinkfingerprotein 20
AbD	<i>Assay-by-Design</i>
AE	<i>annealing/elongation</i>
AoD	<i>Assay-on-Demand</i>
Bp	Basenpaar
BRD7	Bromodomänen Protein 7
CALB2	kalzium-bindendes Protein 2
CC	<i>case control</i>
CCD	<i>charge coupled device</i>
CDH	Cadherin
CED	chronisch entzündliche Darmerkrankungen
CEPH	Centre d'Etude du Polymorphisme Humain
Chi-p	χ^2 -basierter p-Wert
Chr16q	Chromosom 16q
CHST4	Carbohydrat-Sulfotransferase 4
CU	Colitis ulcerosa
CYLD	cylindromatosis
D'	liberaler Kopplungskoeffizient
DCCV	Deutsche Morbus Crohn/Colitis ulcerosa Vereinigung
DNA	<i>desoxribonucleic acid</i>
dNTP	Desoxynukleotid
EDTA	Ethylendiamintetraessigsäure
EST	<i>expressed sequence tag</i>
Exo I	Exonuklease I
f_{Hap}	Haplotypfrequenz
FISH	Fluoreszenz-in-situ-Hybridisierung
FITC	Fluoresceinisothiocyanat
f_{Kont}	MA-Allelfrequenz in der Kontrollgruppe
f_{Pat}	MA-Allelfrequenz in der Patientengruppe
F-Primer	<i>forward primer</i>
GC	<i>genomic control</i>
GC-Gehalt	Guanin/Cytosin-Gehalt
giSNP	<i>genetically indistinguishable SNP</i>

GPCR	G-Protein gekoppelter Rezeptor
GPR56	G-Protein gekoppelter Rezeptor 56
G-Protein	Guaninnukleotid-bindendes Protein
H₀	Nullhypothese
HGP	Human Genome Project
HLA	<i>human leukocyte antigen</i>
ht	<i>haplotype tagging</i>
HWG	Hardy-Weinberg-Gleichgewicht
ISH	in-situ-Hybridisierung
Kb	Kilobase
KI	95%-Konfidenzintervall
KR	Kandidatenregion
LD	<i>linkage disequilibrium</i>
LOD	<i>logarithm of the odds</i>
MA	selteneres Allel
Mb	Megabase
MC	Morbus Crohn
m_G	Anzahl der Kontrollen mit geringem Erkrankungsrisiko
MGB	<i>minor groove binder</i>
MgCl₂	Magnesiumchlorid
m_H	Anzahl der Kontrollen mit hohem Erkrankungsrisiko
MIM	Mendelian Inheritance in Man
NCBI	National Center for Biotechnology Information
NF-κB	nuklearer Faktor κB
n_G	Anzahl der Patienten mit geringem Erkrankungsrisiko
n_H	Anzahl der Patienten mit hohem Erkrankungsrisiko
NKD1	<i>naked cuticle homolog 1 (Drosophila)</i>
NOD2	<i>nucleotid oligodimerisation domain 2</i>
OCTN1	<i>organic cation transporter 1</i>
OCTN2	<i>organic cation transporter 2</i>
OR	<i>odds ratio</i>
OR_A	allelbasiertes OR
OR_{Het}	heterozygotenbasiertes OR
OR_{Hom}	homozygotenbasiertes OR
p_A	allelbasierte Häufigkeit
PCR	<i>polymerase chain reaction</i>
p_G	genotypenbasierte Häufigkeit

P_{het}	heterozygotenbasierte Häufigkeit
P_{hom}	homozygotenbasierte Häufigkeit
r²	konservativer Kopplungskoeffizient
RH	<i>radiation hybrid</i>
RNA	<i>ribonucleic acid</i>
R-Primer	<i>reverse</i> Primer
RR	relatives Risiko
RZPD	Deutsches Ressourcenzentrum für Genomforschung GmbH
SAP	<i>shrimp alkaline phosphatase</i>
SHGC	Stanford Human Genome Center
SNP	<i>single nucleotid polymorphism</i>
TDT	<i>transmission disequilibrium test</i>
T_m	Schmelztemperatur
TNF-α	Tumor Nekrose Faktor alpha
T_{opt-E}	optimierte PCR Endtemperatur
T_{opt-S}	optimierte PCR Starttemperatur
Trans	Anzahl transmittierter Allele
TTP	Tristetraprolin
Untrans	Anzahl nicht-transmittierter Allele
WA	häufigeres (Widtyp-)Allel
WTSI	Welcome Trust Sanger Institute
ZFP90	Zinkfingerprotein 90
ZNF19	Zinkfingerprotein 19
ZNF23	Zinkfingerprotein 23

9 Anhang

9.1 Chemikalien und Geräte

Material	Hersteller / Anbieter
100 bp DNA ladder	Invitrogen; Karlsruhe, Deutschland
2.2 ML storage plate (96 well)	ABgene, Epsom, UK
384 deep well storage plate	ABgene, Epsom, UK
Agarose	Eurogentec; Köln, Deutschland
AmpliTaq [®] DNA Polymerase	Applied Biosystems; Weiterstadt, Deutschland
AmpliTaq [®] Gold DNA Polymerase	Applied Biosystems; Weiterstadt, Deutschland
AmpliTaq [®] DNA Polymerase	Applied Biosystems; Weiterstadt, Deutschland
Bacillo [®]	Bode Chemie; Hamburg, Deutschland
BigDye Terminator Ready reaction kit	Applied Biosystems; Weiterstadt, Deutschland
Bromophenol blue	Sigma; München, Germany
Cell culture flasks (250 ml; canted neck)	BD Biosciences; Heidelberg, Deutschland
Cryotubes (2ml)	Greiner Bio-One GmbH; Frickenhausen, Germany
DNAzol [®]	Molecular Research Center, inc. Cincinnati, Ohio, USA
dNTP set (100mM solutions 100µM each)	Amersham Biosciences; Freiburg, Deutschland
Easy peel heat seal foil	ABgene, Epsom, UK
EDTA	Sigma; München, Deutschland
EDTA blood vial 9 ml	Sarstedt; Nümbrecht, Deutschland
Ethanol <i>p. a.</i>	Merck; Darmstadt, Deutschland
Ethanol, technisch	Bundesmonopol für Branntwein (BfB); Offenbach, Deutschland
Ethidium bromide solution (10mg/ml)	Invitrogen; Karlsruhe, Deutschland
ExoI (Exonuclease I)	Amersham Biosciences; Freiburg, Deutschland
GeneAmp PCR buffer system (10x buffer w/o MgCl ₂ ; 25mM MgCl ₂ solution)	Applied Biosystems; Weiterstadt, Deutschland
Glycerol	Sigma; München, Deutschland
Invisorb Blood Giga Kit	Invitex, Berlin, Deutschland
isopropanol	Merck; Darmstadt, Deutschland
MgCl ₂	Merck; Darmstadt, Deutschland
MicroAmp [®] optical 96 well reaction plate	Applied Biosystems; Weiterstadt, Deutschland
MicroAmp [®] single strips	Applied Biosystems; Weiterstadt, Deutschland
MicroAmp [®] single tubes	Applied Biosystems; Weiterstadt, Deutschland
Microtiter 384 well plates	Sarstedt; Nürnberg, Deutschland

Material	Hersteller / Anbieter
Microtiter 384 well plates	Greiner Bio-One GmbH; Frickenhausen, Deutschland
Microtiter 96 well plates	Sarstedt; Nürnberg, Deutschland
Microtiter 96 well plates	Costar Corning Incorporated; Cambridge, MA, USA
Microtiter plates, 96-well, round bottom, with lid	Sarstedt; Nümbrecht, Deutschland
Microtiter strips (Nunc-ImmunoMaxisorp™, 8-well, flat bottom) with 96-well frame	Nunc; Wiesbaden, Deutschland
Multiscreen column loader	Amersham Biosciences; Freiburg, Deutschland
PEQLAB DNA isolation system	PEQLAB Biotechnology GmbH; Erlangen, Deutschland
PicoGreen	Molecular Probes Europe BV, Leiden, The Netherlands
Pipette (serological, sterile with filter 5 / 10 / 25 ml)	Sarstedt; Nümbrecht, Deutschland
Pipette tips with filter (10 / 200 / 1,000 µl)	Sarstedt; Nürnberg, Deutschland
protein-kinase K	Invitek, Berlin, Deutschland
proteinase K	Molecular Research Center, inc. Cincinnati, Ohio, USA
saccharose	Merck; Darmstadt, Deutschland
SAP shrimp alkaline phosphatase	Amersham Biosciences; Freiburg, Deutschland
Sephadex powder (G50 superfine)	Amersham Biosciences; Freiburg, Deutschland
Sephadex spin column plates MAHVN 4550	Amersham Biosciences; Freiburg, Deutschland
SmartLadder DNA marker	Eurogentec; Köln, Deutschland
TAE Buffer 25x ready pack	Amresco; Solon, OH, USA
TaqMan® Universal PCR Master Mix	Applied Biosystems; Weiterstadt, Deutschland
TBE buffer 10x ready pack	Amresco; Solon, OH, USA
TEMED	Sigma; München, Deutschland
Tris	Merck; Darmstadt, Deutschland
Triton-X	Sigma; München, Deutschland
Trypsin / EDTA (0.25% / 1 mM)	Invitrogen/Gibco; Karlsruhe, Deutschland
Tubes (0.5 / 1.5 / 2.0 mL)	Eppendorf; Köln, Deutschland
Tubes (0.5, 1.5, 2 mL)	Eppendorf; Köln, Deutschland
Tubes, flat bottom (60 mL)	Sarstedt; Nümbrecht, Deutschland
Tubes, sterile (15 mL)	Sarstedt; Nümbrecht, Deutschland
Tubes, sterile (50 mL)	BD Biosciences; Heidelberg, Deutschland
Xylene Cyanol FF	Sigma; München, Deutschland

9.2 Patientenfragebogen

Patient

Fragebogen zur Familienstudie bei chronisch entzündlichen Darmerkrankungen

Persönliche Daten

männlich <input type="checkbox"/>	weiblich <input type="checkbox"/>	Geburtsjahr: _____
-----------------------------------	-----------------------------------	--------------------

In verschiedenen Volksgruppen treten die chronisch entzündlichen Darmerkrankungen in verschiedener Häufigkeit und Ausprägung auf

Ethnische Herkunft (z.B. deutsch, türkisch, italienisch, polnisch): _____

<u>Woran sind Sie erkrankt?</u>	nicht erkrankt	<input type="checkbox"/>
Morbus Crohn <input type="checkbox"/>	Crohn/Colitis nicht entschieden	<input type="checkbox"/>
Colitis ulcerosa <input type="checkbox"/>	Collagene colitis	<input type="checkbox"/>

Welche Ihrer Verwandten hat ebenfalls Morbus Crohn oder Colitis Ulcerosa?

Welcher Ihrer Verwandten hat Symptome wie Sie selbst auch, ist aber bisher nicht untersucht worden?

Wie viele Geschwister haben Sie? _____
(Bitte mit Alter und Geschlecht angeben, wenn verstorben bitte Jahr und Alter angeben) _____

Wie viele Kinder haben Sie selbst _____
(Bitte mit Alter und Geschlecht angeben) _____

Rauchen/Rauchten Sie Ja Nein Wenn ja,
wie viele Jahre: von.....bis.....
Wie viele Zigaretten pro Tag: _____

In einer englischen Studie wurde ein Zusammenhang zwischen dem Stadtleben und dem Risiko an Morbus Crohn zu erkranken hergestellt:

<u>Wo sind Sie aufgewachsen ?</u>	<u>Wo lebe Sie heute?</u>
Dorf <input type="checkbox"/>	Dorf <input type="checkbox"/>
Kleinstadt <input type="checkbox"/>	Kleinstadt <input type="checkbox"/>
Großstadt <input type="checkbox"/>	Großstadt <input type="checkbox"/>

In einem Haushalt mit wie vielen Personen wuchsen Sie auf? _____

In einem Haushalt mit wie vielen Personen leben Sie heute? _____

<u>Welcher Sanitärkomfort herrschte in Ihrer Kindheit?</u>			
Fließendes Wasser	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>	
Fließendes Warmwasser	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>	
Wasserklosett	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>	
Zentralheizung	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>	

Wurden Sie gegen Tuberkulose (Tbc) geimpft?	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>
Wurden Sie gegen Masern geimpft?	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>
Sind Sie als Kind an Masern erkrankt	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>
Besteht bei Ihnen Asthma?	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>
wenn ja, seit wann:		
Besteht bei Ihnen eine Neurodermitis/Ekzem?	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>
Besteht bei Ihnen eine rheumatische Organerkrankung?	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>
Wenn ja, welche:		
Seit wann:		

Die folgenden Fragen beziehen sich nur auf erkrankte Familienmitglieder: Ihre individuelle genetische Veranlagung kann möglicherweise den Krankheitsverlauf beeinflussen:

<u>Welche Darmteile sind bei Ihnen betroffen?</u>		
Dünndarm	<input type="checkbox"/>	Übergang vom Dünndarm zum Dickdarm <input type="checkbox"/>
Teile des Dickdarms	<input type="checkbox"/>	welche:
ganzer Dickdarm:	<input type="checkbox"/>	
Enddarm	<input type="checkbox"/>	
genauere Angaben:		
<u>Haben Sie während Ihrer Erkrankung Fisteln entwickelt?</u> Ja <input type="checkbox"/> Nein <input type="checkbox"/>		
Wenn ja, welche:		
Enddarm (Analfisteln)	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	weiß nicht <input type="checkbox"/>
Fisteln zwischen Darmschlingen	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	weiß nicht <input type="checkbox"/>
Fisteln vom Darm zur Haut	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	weiß nicht <input type="checkbox"/>
Fisteln zu anderen Organen	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	weiß nicht <input type="checkbox"/>
<u>Haben sich bei Ihnen im Verlauf der Erkrankung Stenosen (Darmverengungen) ausgebildet?</u> Ja <input type="checkbox"/> Nein <input type="checkbox"/>		
<u>Haben/Hatten Sie extraintestinale Beschwerden?</u> Ja <input type="checkbox"/> Nein <input type="checkbox"/>		
Gelenkschmerzen während der Schübe	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	
Hautentzündungen	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	
Augenentzündungen	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	
Gallengangsentzündungen	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	
Seit welchem Lebensjahr ist die Diagnose klar?		
Seit welchem Lebensjahr hatten Sie erste Beschwerden?		
<u>Wie wurde Ihre Erkrankung bestätigt?</u>		
Feingewebeschnitt	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	
Darmspiegelung	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	
Darmröntgen	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	
Wie viele - und welche - Operationen sind aufgrund Ihrer Erkrankung notwendig gewesen?		

Haben/Hatten Sie einen künstlichen Darmausgang? Ja <input type="checkbox"/> Nein <input type="checkbox"/>	
<u>Wie viele stationäre Krankenhausaufenthalte sind bei Ihnen notwendig gewesen?</u>	
Nie <input type="checkbox"/>	<u>Wie lange insgesamt im Krankenhaus?</u>
Weniger als alle 2 Jahre <input type="checkbox"/>	Weniger als 1 Monat <input type="checkbox"/>
ca. 1x pro Jahr <input type="checkbox"/>	1-3 Monate <input type="checkbox"/>
1-3 x pro Jahr <input type="checkbox"/>	3-6 Monate <input type="checkbox"/>
Mehr als 3 x pro Jahr <input type="checkbox"/>	Mehr als 6 Monate <input type="checkbox"/>
<u>Wie lange haben Sie insgesamt Kortison > 10mg pro Tag nehmen müssen?</u>	
Mehr als ein Jahr <input type="checkbox"/>	
bis zu einem Jahr <input type="checkbox"/>	
3-6 Monate <input type="checkbox"/>	
1-3 Monate <input type="checkbox"/>	
Weniger als 1 Monat <input type="checkbox"/>	
Noch nie <input type="checkbox"/>	

<u>Wie viele Schübe haben Sie in etwa pro Jahr?</u>	
Weniger als alle 2 Jahre <input type="checkbox"/>	
ca. 1x pro Jahr <input type="checkbox"/>	
1-3 x pro Jahr <input type="checkbox"/>	
Mehr als 3 x pro Jahr <input type="checkbox"/>	
Wann war der letzte Schub:	
<u>Sind Sie wegen Ihrer chronisch-entzündlichen Darmerkrankung vorzeitig berentet worden?</u> Ja <input type="checkbox"/> Nein <input type="checkbox"/>	
Wenn ja, in welchem Lebensalter?	

Was würden Sie in der Betreuung von Patienten mit chronisch-entzündlichen Darmerkrankungen verändern?

Nach Eingang der Proben wird Ihre Adresse und Ihr Name aus unseren Unterlagen gelöscht. Aufgrund vieler Rückfragen von Patienten möchten wir Sie hier aber noch einmal fragen.

Bitte Löschen Sie meine Adresse und Namen nach Eingang sofort:

Ja Nein

Ich möchte weiter über die allgemeinen Ergebnisse der Studie (die in ca. 3 Jahren zu erwarten sind) auf dem Laufenden gehalten werden:

Ja Nein

Vielen Dank für Ihre Mitwirkung!

Bei Rückfragen melden Sie sich bitte gerne unter Telefon 0431 597 1373

Danksagung

Mein grundsätzlicher Dank gilt vorab all denjenigen, die mich bei der Entstehung und der Fertigstellung dieser Arbeit sowohl beruflich als auch privat begleitet, unterstützt und motiviert haben. Ich habe das Glück, von besonderen Menschen umgeben zu sein, denen stillschweigend mein innigster Dank gilt. Nichtsdestotrotz möchte ich an dieser Stelle ganz besonders danken:

Herrn Prof. Dr. Stefan Schreiber für die Vergabe eines spannenden Themas, das mich aus der Welt der domestizierten Meerschweinchen in die entlegenen Winkel des menschlichen Körpers geführt hat.

Herrn Prof. Dr. Günther Hartl für seine Bereitschaft, diese neue Erfahrung in Form von anregenden Diskussionen mit mir zu teilen und meiner Arbeit begleitend beizustehen.

Der großen ‚Mukosafamilie‘, denn ohne sie wäre diese Arbeit nicht zu Stande gekommen. Großer Dank geht insbesondere ins Labor, wo ich eine enorme Unterstützung erhalten habe.

Meinen Freunden und meiner Familie für die moralische Unterstützung in der ‚heißen‘ Phase.

Danke Katja!

Erklärung

Hiermit erkläre ich, dass die vorliegende Arbeit - abgesehen von der Beratung durch meine akademischen Lehrer - nach Inhalt und Form meine eigene Arbeit ist. Die Arbeit wurde bis jetzt weder vollständig noch in Teilen einer anderen Stelle im Rahmen eines Prüfungsverfahrens vorgelegt. Ferner erkläre ich, dass ich noch keine früheren Promotionsversuche unternommen habe.

Für die mündliche Prüfung wird die Form der Disputation gewählt.

Der Zulassung von Zuhörern / Zuhörerinnen bei der mündlichen Prüfung wird nicht widersprochen.

Kiel, den

.....

(Andreas Rüther)

Lebenslauf

Persönliche Daten

Name	Andreas Rüter
Geburtsdatum	17.12.1971
Geburtsort	Duisburg / Nordrhein-Westfalen
Staatsangehörigkeit	Deutsch
Familienstand	Ledig

Berufliche Entwicklung

ab 2001	Wissenschaftlicher Angestellter am Institut für Klinische Molekularbiologie, Universitätsklinikum Schleswig-Holstein, Campus Kiel
1994 - 2001	Diplomstudium der Biologie an der Christian-Albrechts-Universität zu Kiel (CAU) mit Hauptprüfung im Fach Zoologie, Nebenfächer Biologische Meereskunde und Meereschemie, Diplomarbeit am Institut für Haustierkunde: „Biochemisch-genetische Variabilität und Differenzierung bei wilden und domestizierten Meerschweinchen (Gattung <i>Cavia</i>)“, Abschlussnote: 1,0
1992 - 1994	Ausbildung zum Chemisch - Technischen Assistenten an den beruflichen Schulen des Hochsauerlandkreises in Olsberg, Abschlussnote: 2,3

Wehrdienst

1991 - 1992	Standort Göttingen und Schwalmstadt (Heer)
-------------	--

Schulbildung

1984 - 1991	Gymnasium Petrinum in Brilon, 1. Leistungsfach: Biologie, 2. Leistungsfach: Deutsch, Abschlussnote: 1,9
1982 - 1984	Abtei-Gymnasium Hamborn in Duisburg
1978 - 1982	Städtische Grundschulen in Duisburg