

Aus der Kinderklinik und Kinderpoliklinik im Dr. von Haunerschen Kinderspital

Abteilung Pneumologie/Allergologie

Direktor: Prof. Dr. Dr. h.c. Dietrich Reinhardt

GENETISCHE DETERMINANTEN VON ASTHMA

METHODIK UND ANWENDUNG



Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Martin Depner

aus
Pforzheim

2009

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter: Prof. Dr. Erika von Mutius

Mitberichterstatter: Prof. Dr. Katja Radon, MSc
Priv. Doz. Dr. Annette Pohl-Koppe

Mitbetreuung durch den
promovierten Mitarbeiter: Dr. Sabina Illi

Dekan: Prof. Dr. med. Dr. h.c. M. Reiser, FACR, FRCR

Tag der mündlichen Prüfung: 08.10.2009

Inhaltsverzeichnis

Abkürzungsverzeichnis	6
Abbildungsverzeichnis	9
Tabellenverzeichnis.....	10
1 Einleitung	11
1.1 Relevanz der Identifizierung von Asthmagenen	11
1.2 Asthma	13
1.2.1 Definition	13
1.2.2 Zellmechanismen (T-Zellen).....	14
1.2.3 Umbauprozesse der Atemwege („airway remodeling“).....	15
1.3 Atopie.....	16
1.4 Atopisches und nicht-atopisches Asthma.....	17
1.5 Das genetische Material	18
1.5.1 DNA	18
1.5.1.1 Struktur der DNA.....	18
1.5.1.2 DNA-Abschnitte Gen, Chromosomen, Locus, Genom	21
1.5.1.2.1 Chromosomen	22
1.5.1.2.2 Locus	24
1.5.1.2.3 Genom	24
1.5.1.3 Variationen der DNA-Sequenz	24
1.5.2 Weitergabe genetischen Materials	28
1.5.2.1 Rekombination.....	28
1.5.2.2 Kopplungsungleichgewicht.....	30
1.5.2.3 Erbgänge	31
1.6 Auf der Suche nach Asthma-Genen	35
1.6.1 Auswahl bestimmter Gene: Überblick über Vorgehensweisen.....	35
1.6.2 Genomweiter Ansatz	36
1.6.2.1 Genomweite Assoziationsstudien	36
1.6.2.2 „Positional Cloning“	37
1.6.2.2.1 ADAM33.....	38
1.6.3 Kandidatengenansatz.....	39
1.6.3.1 Kandidatengene <i>Toll-like Rezeptoren (TLRs)</i>	41
1.7 Ableitung der Fragestellung	44
2 Methoden.....	45
2.1 Grundbegriffe Assoziationsstudien	45
2.1.1 SNPs als genetische Variablen	46
2.1.2 Genotypkombinationen als genetische Variablen	46
2.1.3 Haplotypen als genetische Variablen	47
2.1.4 Hardy-Weinberg-Equilibrium	48
2.1.5 Power und Studiendesign	50
2.1.6 Stratifikation.....	51
2.1.7 Kopplungsungleichgewicht oder „linkage disequilibrium“	52
2.1.7.1 LD-Koeffizienten.....	52
2.1.7.2 LD-Blöcke und „tagging SNPs“	55
2.1.7.3 LD bei der Kartierung	57
2.1.8 Multiples Testen und Replikation	58
2.1.9 Interaktion	58
2.2 Statistische Methoden zur Auswertung von Assoziationsstudien.....	60

2.2.1	Zusammenhänge zwischen einem SNP und qualitativen Phänotypen.....	60
2.2.1.1	Allel zählendes Modell	61
2.2.1.2	Dominantes und rezessives Modell	61
2.2.1.3	Genotypentest ohne Modell	62
2.2.1.3.1	Kategorielles Modell	62
2.2.1.3.2	Dummy-Modelle	62
2.2.1.4	Allelisches Modell	63
2.2.2	Zusammenhänge zwischen einem SNP und quantitativen Phänotypen.....	64
2.2.3	Zusammenhänge zwischen einem Haplotyp und qualitativen Phänotypen	65
2.2.3.1	Auswahl der SNPs für einen Haplotyp	65
2.2.3.2	Auswahl der Stichprobe für die Haplotypanalyse	66
2.2.3.3	Schätzen der Haplotypen	67
2.2.3.4	Assoziationstests für Haplotypen.....	68
2.2.3.5	Effektgrößen bei Haplotypanalysen, Haplotyp-trendregression	70
2.2.4	Besonderheiten bei GWA-Studien	71
2.3	Bestimmung der Genotypen.....	74
2.3.1	CHIP.....	74
2.3.2	MALDITOF	76
2.3.3	Generelle Qualitätskontrollen	77
2.4	Für die Auswertungen verwendete Stichproben	78
2.4.1	Querschnittsstudien (ISAAC II und Leipzig).....	80
2.4.2	Fall-Kontroll-Studie MDL1872	82
2.4.3	Stichprobe für Genomweite Assoziationsstudie	83
2.4.4	MAS	85
2.5	Definition der relevanten Zielvariablen	86
2.6	Verwendete Methoden	88
2.6.1	Population und Typisierung	88
2.6.2	Statistische Methoden	88
3	Ergebnisse	90
3.1	ORMDL3 als Vertreter der GWA.....	90
3.1.1	Ergebnisse der GWA („Hits“).....	90
3.1.2	Betrachtung des Locus auf Chromosom 17	93
3.1.2.1	Gene und Kopplungsungleichgewicht	93
3.1.2.2	ORMDL	94
3.1.2.3	Replikation des ursprünglichen Befundes	95
3.1.2.4	Spezifizierung des ursprünglichen Befundes.....	100
3.1.2.5	Spezifizierung des ORMDL-Effektes.....	102
3.2	ADAM33 als Vertreter des „positional cloning“	105
3.3	TLR als Vertreter der Kandidatengene	108
4	Diskussion	113
4.1.	Zusammenfassung wesentlicher Ergebnisse	113
4.2	Einordnung der Ergebnisse in Kontext der Asthmaforschung.....	115
4.2.1	Gene, deren Assoziation mit Asthma oft repliziert wurde.....	115
4.2.2	Varianten der Definition von Replikation.....	116
4.2.3	<i>ORMDLs</i> im Kontext der Replikationsstudien.....	116
4.2.4	<i>ADAM33</i> im Kontext der Replikationsstudien.....	118
4.2.5	<i>TLRs</i> im Kontext der Replikationsstudien	121
4.2.6	Methodologische Überlegungen zu GWAs.....	125
4.2.6.1	Publikationsbias	125
4.2.6.2	„Coverage“.....	126
4.2.6.3	„Common disease – common variant“-Hypothese	128

4.2.6.4	Allelfrequenz, Power und Effektgrößen	129
4.2.6.5	Kombination der Methoden zur Identifizierung von Asthmagenen	133
4.3	Funktionen der Gene	136
4.3.1	Klassifikation von Kandidatengenem	136
4.3.2	Bedeutung der Gene auf Chromosom 17	138
4.4	Rückschlüsse auf weiteres Vorgehen bei GWAs	142
5	Zusammenfassung	144
6	Literaturverzeichnis	147
	Danksagung	162
	Lebenslauf	163

Abkürzungsverzeichnis

AA	amino acid
ADAM33	a disintegrin and metalloproteinase
ADRB2	adrenergic beta-2 receptor
AF	allele frequency
AHR	airway hyperreactivity
AS	Aminosäure
BHR	bronchial hyperreactivity
CC16	clara cell protein 16
CCL	chemokine (C-C motif) ligand
CD	cluster of differentiation
CDCV-Hypothese	common disease - common variant-Hypothese
CHI3L1	Chitinase 3-like 1
CNG	Centre National de Génotypage
CNV	copy number variation
COL29A1	collagen, type XXIX, alpha 1
CYFIP2	cytoplasmic FMR1 interacting protein 2
D	Disequilibriumskoeffizient
D	Dresden
D'	D prime (Maß für LD)
D1940	Population von 1940 Kindern aus Dresden
DEFB1	defensin, beta 1
DFNA5	deafness, autosomal dominant 5
DLM4264	Population von 4264 Kindern (aus Dresden, Leipzig und München)
D_{\max}	maximal möglicher Disequilibriumskoeffizient
D_{\min}	minimal möglicher Disequilibriumskoeffizient
DNA	desoxyribonucleic acid
DPP10	dipeptidyl-peptidase 10
dsRNA	double stranded RNA
EBV	Eppstein-Barr Virus
EBVL	EBV-transformierte B-Zell-Linien
EGF	epidermal growth factor
ELB-Algorithmus	Excoffier-Laval-Balding Algorithmus
EM-Algorithmus	expectation-maximization Algorithmus
EMSA	electrophoretic mobility shift assay
eQTL	expression quantitative trait locus
eQTNs	expression quantitative trait nucleotides
ERO1LB	ERO1-like beta (<i>S. cerevisiae</i>)
FCER1B	Fc epsilon receptor I beta-chain
FDA	food and drug administration
FDR	false discovery rate
FEV ₁	forced expiratory volume in 1 second
FLG	Filaggrin
FVC	forced vital capacity
GATA3	GATA binding protein 3
Gfi1	growth factor independent 1
GPR154	G protein receptor 154
GPRA	G-protein coupled receptor for asthma susceptibility

GSDM1	Gasdermin 1
GSDML	Gasdermin-like
GWA	Genome-wide association study
HDL	high density lipoprotein
HLA	major histocompatibility complex, class II
HLADRB1	Major histocompatibility complex, class II, DR beta 1
htSNPs	haplotyp tagging SNPs
HWE	Hardy-Weinberg equilibrium
IE	intronic enhancer
IFN- γ	interferon gamma
IgE	Immunglobulin E
IL	Interleukin
IRAKM	IL-1 receptor-associated-kinase-M
ISAAC	International Study of Asthma and Allergies in Childhood
IU/ml	international units/milliliter
K.I.	Konfidenz-Intervall
L	Leipzig
L1165	Population von 1165 Kindern aus Leipzig
LD	linkage disequilibrium
LP	Lipoproteine
LPS	Lipopolysaccharid
M	München
M1159	Population von 1159 Kindern aus München
MAF	minor allele frequency
MAGICS	Multicentre Asthma Genetics in Childhood Study
MALDITOF	Matrix Assisted Laser Desorption/Ionisation Time Of Flight
MAS	Multicenter Allergy Study
MD3099	Population von 3099 Kindern (aus Dresden, Leipzig und München)
MDL1872	Population von 1872 Kindern (aus München, Dresden und Leipzig)
MLZE	melanoma-derived leucine zipper
MRC	medical research council
mRNA	messenger RNA
MS	Massenspektrometrie
NOD1/2	nucleotide-binding oligomerization domain containing 1/2
NPSR1	neuropeptide S receptor 1
NRG1	Neuregulin 1
OR	Odds Ratio
ORF	open reading frame
ORMDL	ORM1 (<i>S. cerevisiae</i>)-like genes
ORMDL3	ORM1-like 3 (<i>S. cerevisiae</i>)
PAMP(s)	pathogen associated molecular pattern(s)
PBMC	peripheral blood mononuclear cell
PC ₂₀ FEV ₁	provocative concentration inducing a 20% fall in FEV ₁
PCR	polymerase chain reaction
PG	Peptidoglykan
PHD	plant homeodomain
PHF11	PHD finger protein 11
PRR	pattern recognition receptors
PTGER2	prostaglandin E receptor 2
R ²	R square (Maß für LD)
RNA	ribonucleic acid

RS-Virus	respiratory syncytial virus
SNP(s)	single nucleotide polymorphism(s)
SPINK5	serine peptidase inhibitor, Kazal type 5
ssRNA	single stranded RNA
STAT6	signal transducer and activator of transcription 6
TBX21	T-box 21
TF	transcription factor
T _h -Zellen	T-Helfer-Zellen
TLR	Toll-like Rezeptor
T _{regs}	regulatorische T-Zellen
USF	upstream transcription factor
UTR	untranslatierte Region
YKL	human cartilage glycoprotein
ZBP2	zipcode binding protein 2
ZCCHC11	zinc finger, CCHC domain containing 11
ZNFN1A3	zinc finger , Protein, Subfamily 1 A, 3(Aiolos)

Abbildungsverzeichnis

Abbildung 1: Phänotypische Heterogenität der Atopie	16
Abbildung 2: DNA und Gene	18
Abbildung 3: DNA Struktur.....	19
Abbildung 4: DNA Replikation (A) und Transkription (B)	19
Abbildung 5: Änderung des Proteins durch Änderung im Gen.....	20
Abbildung 6: Struktur eines Gens (oben) und Änderungen nach der Transkription	21
Abbildung 7: Beispiel für ein Teilstück der DNA in haploider und diploider Form ohne Variation.	22
Abbildung 8: Darstellung eines Chromosoms	23
Abbildung 9: Beispiel für ein Teilstück der DNA in haploider und diploider Form mit Möglichkeit der Variation.....	24
Abbildung 10: Beispiel für ein Allel.	25
Abbildung 11: Beispiel für einen SNP und die 3 Varianten, die eine Person tragen kann.	26
Abbildung 12: Unterscheidung von Genotyp und Haplotyp.....	27
Abbildung 13: SNP als Marker (nur ein Ausschnitt der umliegenden Sequenz ist gezeigt) ...	28
Abbildung 14: Beispiel für „crossing over“.....	29
Abbildung 15: Beziehung zwischen LD (r , y-Achse) und der Anzahl von Generationen, seit die Mutation in die Population eingeführt wurde (x-Achse).	31
Abbildung 16: Beispiel für Mendelsche Vererbung: rezessiver, dominanter und kodominanter Erbgang.....	32
Abbildung 17: Modell für eine komplexe Erkrankung.	34
Abbildung 18: Schema der Domänen-Struktur der ADAMs.	38
Abbildung 19: Unterschied zwischen Haplotyp und kombiniertem Genotyp	47
Abbildung 20: Hardy–Weinberg-Gleichgewicht für zwei Allele.	49
Abbildung 21: Varianten der SNP- und Stichprobenselektion für Haplotypanalysen.....	65
Abbildung 22: Variationsmöglichkeiten der Haplotypen bei Trägern mehrerer heterozygoter SNPs	67
Abbildung 23: Darstellung eines „cluster plots“.....	76
Abbildung 24: Populationen, auf denen die Auswertungen dieser Arbeit basieren.....	78
Abbildung 25: Studiendesign der GWA	90
Abbildung 26: Hits aus der GWA-Studie (nach Moffatt (Moffatt et al. 2007))	91
Abbildung 27: A) Lage der zehn Hits auf Chromosom 17 und B) deren „linkage disequilibrium“ (berechnet als R^2) in der Population MD3099 aus ISAAC ...	93
Abbildung 28: Expressionsmuster von <i>ORMDL1</i> , <i>ORMDL2</i> , und <i>ORMDL3</i> in verschiedenen Geweben aus Analysen einer „RealTime-Polymerase chain reaction“ („RealTime-PCR“).	95
Abbildung 29: Lage der untersuchten SNPs in <i>ADAM33</i>	105
Abbildung 30: TLR1, TLR6 und TLR10 SNPs erhöhen proinflammatorische und Th1- assoziierte Zytokinlevels und reduzieren Th2-assoziierte <i>IL4</i> -Expression nach TLR2/1 und TLR2/6 Stimulation.....	111
Abbildung 31: Überblick über die am häufigsten replizierten Gene	115
Abbildung 32: Kombination von Kandidatengenansatz mit genomweiten Ansätzen.....	134
Abbildung 33: Klassifikation von Kandidatengen nach funktionellen Mechanismen.	137
Abbildung 34: Integrierter Ansatz zur Untersuchung von Genotyp-Phänotyp- Zusammenhängen.	142

Tabellenverzeichnis

Tabelle 1: Überblick über die humanen <i>Toll-like</i> Rezeptoren und ihre Liganden.	42
Tabelle 2: Beispiel zur Berechnung von „linkage disequilibrium“	52
Tabelle 3: Beispiel für LD auf Grund von Populationsstruktur	55
Tabelle 4: Beispiel zur Bestimmung von „haplotype tagging SNPs“	56
Tabelle 5: Kodierungen eines SNPs bei unterschiedlichen genetischen Modellen.....	60
Tabelle 6: Häufigkeiten in Fällen und Kontrollen nach Genotyp	61
Tabelle 7: Häufigkeiten in Fällen und Kontrollen im dominanten Modell	62
Tabelle 8: Häufigkeiten in Fällen und Kontrollen im allelischen Modell.....	63
Tabelle 9: Programme zur Schätzung von Haplotypen	68
Tabelle 10: Kodierungen von Haplotypen	68
Tabelle 11: Formalismus zum Testen auf Assoziationen eines individuellen Haplotyps mit dem Phänotyp	69
Tabelle 12: Typischer Ablauf einer genomweiten Assoziationsstudie	74
Tabelle 13: Deskriptive Charakteristik wichtiger Phänotypen für alle deutschen Kinder in der gesamten (N=7940) und genotypisierten (N=4264) Studienpopulation	80
Tabelle 14: Deskriptive Charakteristik wichtiger Phänotypen für alle deutschen Kinder in der Studienpopulation nach Städten geschichtet	81
Tabelle 15: Verteilung von Fällen und Kontrollen in der Fall-Kontroll-Studie MDL1872.....	82
Tabelle 16: Anzahl der (deutschen) Kinder, die an der GWA teilnahmen	84
Tabelle 17: Marker der GWA Analyse, welche unter der 5%-FDR-Schranke bleiben.	92
Tabelle 18: Assoziation von Asthma und ORMDL3 mit SNPs von Chromosom17q21 in kombinierten und Replikationsstichproben.....	97
Tabelle 19: Replikationen von Assoziationen des CHR 17 mit Asthma, Locus Teil 1.	98
Tabelle 20: Replikationen von Assoziationen des CHR 17 mit Asthma, Locus Teil 2	99
Tabelle 21: Atopieeffekte in der Querschnittpopulation aus Kindern von Dresden, Leipzig und München (DLM 4264)	101
Tabelle 22: SNP-Beschreibung der Polymorphismen auf Chromosom 17 in DLM4264.....	102
Tabelle 23: Wechselseitig adjustierte Analyse für Effekte von drei SNPs, die stellvertretend für die <i>ORMDL</i> -GENE stehen, auf atopisches Asthma	103
Tabelle 24: Assoziation von SNPs aus <i>ADAM33</i> mit Asthma (für MDL1872 auch Asthma+BHR).....	106
Tabelle 25: Assoziation von Haplotypen in <i>ADAM33</i> mit Asthma	107
Tabelle 26: Ergebnisse der Assoziationsstudien für 21 SNPs aus der Familie der <i>Toll-like Rezeptoren</i> in MDL1872	109
Tabelle 27: Assoziation von <i>TLR1</i> , <i>TLR6</i> und <i>TLR10</i> Polymorphismen mit Asthma, Atopie und Subformen von Asthma und Atopie in einer Querschnittpopulation aus Dresdner und Münchner Kindern (MD3099).	110
Tabelle 28: Replikationsmatrix, um den Grad der Evidenz einer Replikation einzuordnen..	116
Tabelle 29: Widersprüchliche Befunde von Replikationsstudien für SNPs von <i>ADAM33</i> ...	119
Tabelle 30: Überblick über Odds Ratios signifikanter Effekte auf Einzel-SNP-Ebene innerhalb eigener Auswertungen für Publikationen der Arbeitsgruppe Kabesch.	131

1 Einleitung

1.1 *Relevanz der Identifizierung von Asthmagenen*

Die Prävalenz allergischer Erkrankungen ist in den letzten 30 Jahren deutlich angestiegen. Dies gilt vor allem für die westlichen Länder (Jones 2008). Besonders Kinder erkranken immer häufiger an Allergien und Asthma, und im pädiatrischen Bereich ist Asthma mittlerweile die häufigste chronische Krankheit der westlichen Welt.

Eine neuere Kostenanalyse, die vom Forschungszentrum der europäischen Kommission angefordert wurde, hat die Gesamtkosten für die Behandlung asthmatischer Kinder unter 15 Jahren in 25 Mitgliedsstaaten der Europäischen Union auf etwa 3 Billionen Euro im Jahr geschätzt (European Lung White Book 2003). Die Kosten wären mindestens doppelt so hoch, wenn man das Asthma von Erwachsenen oder berufsbezogenes Asthma zusätzlich in Betracht ziehen würde. Allergische Erkrankungen werden zukünftig die Volkswirtschaften sowie deren Gesundheitssysteme in immer stärkerem Maße belasten und immer mehr Menschen in ihrer Lebensqualität einschränken. Eine bedeutende Reduktion der Lebensqualität zeigt sich vor allem auch bei asthmatischen Kindern. Die gegenwärtigen Therapien für Asthma sind zwar effektiv für milde Formen von Asthma; ernstere Formen von Asthma bleiben hingegen schwer zu behandeln. Etwa 80% der Gesamtkosten machen aber gerade jene 20% der Patienten mit schwerem Asthma aus. So ist es heute zwar möglich, bei exakter Diagnose und Identifikation von Auslösern, allergische Erkrankungen und Asthma zu kontrollieren, doch sind trotz intensiver Bemühungen auf dem Gebiet der biomedizinischen Forschung Allergien selten und Asthma bis heute nicht heilbar.

Untersuchungen, die sich mit der Identifizierung von Ätiologie und Therapiemöglichkeiten von Asthma beschäftigen, haben daher eine hohe gesundheitspolitische Relevanz. Sie können zudem für die Primärprävention, also die Verhinderung von Krankheitsfällen durch frühzeitiges Erkennen von Risikofaktoren und deren Eliminierung bedeutend sein, die epidemische Ausbreitung von Allergien und Asthma weltweit zu bekämpfen.

Es ist unwahrscheinlich, dass die Prävalenz allergischer Erkrankungen nur aus genetischen Veränderungen oder nur aus Umweltfaktoren resultiert. Vielmehr werden Ursachen für Asthma derzeit als sowohl umweltbezogen als auch genetisch sowie als Folge deren Inter-

aktion vermutet. Da jedoch in beiden Forschungsfeldern noch keine prädiktiven Marker mit ausreichender Sensitivität und Spezifität identifiziert werden konnten, die sich für die Vorhersage des Krankheitseintritts und die Prognose bereits bestehender allergischer Erkrankungen eignen, gilt es in beiden Feldern die Asthmaforschung voran zu treiben. Wie auch bei anderen chronischen Erkrankungen konzentriert sich eine Form der Suche auf genetische Faktoren, die sich zum Beispiel für eine Diagnostik einsetzen lassen.

1.2 Asthma

1.2.1 Definition

Asthma ist keine klar einzugrenzende Krankheit, sondern eine Ansammlung verschiedener Krankheiten. Ein Konsens der Asthmadefinition besteht darin, Asthma als eine chronisch entzündliche Krankheit der Atemwege zu definieren, bei der viele Zelltypen und zelluläre Elemente eine Rolle spielen, insbesondere Mastzellen, Eosinophile und T-Zellen (Bousquet et al. 2000; Lemanske et al. 2003).

Bei betroffenen Individuen verursacht die Entzündung sich wiederholende Episoden von Giemen, Atemnot, Engegefühl in der Brust und Husten, der besonders in der Nacht und/oder am frühen Morgen auftritt. Diese Symptome sind gewöhnlich mit weit ausgebreiteter, aber variabler Atemwegsobstruktion verknüpft, welche entweder spontan oder durch Behandlung zumindest teilweise reversibel ist. Die Entzündung ist variabel mit Veränderungen von Hyperreaktivität der Atemwege gekoppelt und wirkt sich akut oder chronisch auf eine Einengung der Atemwege aus (Lemanske et al. 2003). Die Hyperreaktivität der Atemwege wird mit „airway hyperreactivity“ (AHR), die bronchiale Hyperreaktivität mit „bronchial hyperreactivity“ (BHR) abgekürzt.

Die Entstehung und Fortschreibung von Asthma ist sehr komplex und bezieht das Zusammenspiel zahlreicher Entzündungszellen ein (Bousquet et al. 2000). Die Teilnahme vieler Zelltypen und ihrer ausgeschütteten Moleküle konstituiert ein dynamisches Netzwerk, welches sich sowie die Krankheit konstant über die Zeit ausgestaltet und zwar sowohl auf zellulärer Ebene als auch auf Ebene des Organismus. Dies betrifft auch die Symptome von Asthma, die sich über das Leben verändern und so die Suche nach genetischen oder nichtgenetischen Risikofaktoren kompliziert gestalten (Carroll 2005). In der Kindheit ist Asthma meist mit der Reaktion auf inhalative Allergene gekoppelt und eine bedeutende Anzahl von Kindern hat lebenslang persistierende Symptome. Die Krankheit kann aber auch später im Leben auftreten und wird dann mit anderen Faktoren wie zum Beispiel dem Rauchen in Verbindung gebracht.

Asthma schließt neben akuten Symptomen, welche durch Bronchodilatoren schnell aufgehoben werden können, auch durch chronische Entzündungen verursachte Exazerbationen ein, die zwar unterdrückt oder durch antientzündliche Medikamente gemildert werden. Daneben gibt es aber einen Prozess der Umbildung der Atemwegswand, das sogenannte „airway remodeling“, für das es bislang keine validierte Behandlung gibt. So kann Asthma auch als ein Kontinuum von Symptomen bis hin zum „airway remodeling“ betrachtet werden, wobei die

Folgen und der Schweregrad dieser Ereignisse von Patient zu Patient hochgradig variabel sind.

1.2.2 Zellmechanismen (T-Zellen)

Insbesondere Mastzellen, Eosinophile und T-Zellen spielen eine wichtige Rolle bei der Asthmapathogenese. Das Triggern einer Immunantwort auf der Oberfläche von Mastzellen oder Basophilen durch allergenspezifisches Immunglobulin E (IgE) kann zu sofortigen Symptomen durch die Ausschüttung von Histamin oder anderen Mediatoren führen, während Eosinophile zu AHR bei Asthmatikern beiträgt (Kay 2001). Die Aktivierung von B-Zellen zur Produktion von IgE und die Anhäufung von Eosinophilen sind unter der Kontrolle von sogenannten T_H2 -Zellen. T_H2 -Zellen lösen die Rekrutierung und Aktivierung anderer Zelltypen aus, die in einer komplexen Kaskade von Entzündungsmechanismen mündet. Biologische Mechanismen, die mit T_H2 -Zellen in Verbindung gebracht werden, bieten daher gute Erklärungsansätze für allergische Reaktionen.

T-Zellen repräsentieren eine heterogene Familie von Zellen, die sowohl eine protektive als auch eine schädliche Funktion auf das Immunsystem ausüben kann. Während Effektorzellen, die auch T-Helferzellen (T_H) genannt werden, gegen exogene schädigende Erreger schützen, wird von regulatorischen T-Zellen (T_{regs}) ein Schutz gegen autoimmune Reaktionen und ein Stoppen der Reaktionen von Effektorzellen angenommen, sobald deren Wirkung gefährlich wird (Romagnani et al. 2009). Zwei Haupttypen von Effektor-T-Zellen kommen beim Menschen vor: T-Helfer-Zellen Typ 1 (T_H1 -Zellen) und Typ 2 (T_H2) (Romagnani 1991).

Typischerweise produzieren nach Antigenkontakt aktivierte T_H1 -Zellen unter dem Einfluss von Interleukin 12 (IL-12) als Hauptzytokin Interferon Gamma ($IFN-\gamma$), einen wichtigen Vermittler der zytotoxischen Immunreaktion, Tumorzellabwehr, Zytolyse von virusinfizierten Zellen, Transplantatabstoßung und Autoimmunprozessen (Ngoc et al. 2005).

Dagegen produzieren T_H2 -Zellen unter dem Einfluss von IL-4 und bei Abwesenheit von IL-12 wiederum IL-4 sowie IL-5, IL-9 und IL-13, aber kein $IFN-\gamma$. Diese Interleukine aktivieren dann B-Zellen (IL-4 und IL-13), Mastzellen (IL-4 und IL-10) und Eosinophile (IL-5). IL-9 und IL-13 sind zudem Zytokine, die Einfluss auf AHR haben (Robinson 2000). Die Verbindung von T_H2 -typischen Reaktionen mit allergischen Erkrankungen wurde auch durch Beobachtungen in Mausmodellen unterstützt (Mosmann et al. 1996).

Der durch verbreitete Antigene auslösende Mechanismus für die allergische Reaktion wird in der Entstehung eines Ungleichgewichts bei der T-Zell-Differenzierung zugunsten einer T_H2 -

Antwort vermutet. Stark vereinfacht lässt sich sagen, dass die T_h1 -Antwort herunterreguliert ist, wenn die T_h2 -Antwort überwiegt und umgekehrt.

T_{regs} üben hingegen zusätzlich einen hemmenden Einfluss auf die Entwicklung von sowohl T_h1 - als auch T_h2 -Zellen aus (Ngoc et al. 2005). Werden T_{regs} aufgrund von mangelnder Auseinandersetzung mit Mikroorganismen weniger stimuliert, sinkt ihr hemmender Einfluss auf die T_h1 - und T_h2 -Zellen. Dies liefert ein Erklärungsmodell, warum die Inzidenz sowohl von T_h1 -vermittelten Autoimmunerkrankungen als auch von T_h2 -vermittelten allergischen Erkrankungen wie dem Asthma bronchiale stark zugenommen hat (Smit, J. J. et al. 2004).

Inzwischen wurden noch weitere T-Zellen beschrieben, die entweder IL-4 und IFN- γ (T_h0 -Zellen) oder IL-17 (T_h17 -Zellen) produzieren, und die Mechanismen weiter ausdifferenzieren (Romagnani et al. 2009). Dennoch hat die Hypothese Bestand, dass ein Ungleichgewicht der T_h -Zellen vom Typ 1 und 2 auf die Entwicklung von Allergien Einfluss nimmt.

1.2.3 Umbauprozesse der Atemwege („airway remodeling“)

Die Atemwege von Asthmatikern zeigen eine charakteristische Infiltration mit Entzündungszellen. Daneben gibt es strukturelle Veränderungen der Atemwege, die auch schon in früher Kindheit gefunden werden können (Barbato et al. 2006; Warner et al. 2008). So gesehen können Umbauprozesse selbst ohne Entzündung stattfinden. Insofern ist ein Verständnis, inwieweit solche Umbauprozesse zur Progression von Asthma beitragen, insbesondere im Hinblick auf Therapie und Krankheitsmanagement, von großer Bedeutung. Pathomechanismen, welche mit solchen Umbauprozessen assoziiert sind, können Erklärungsansätze für das Aufrechterhalten allergischer Symptomatik bieten.

Das „airway remodeling“ kann als die Gemeinsamkeit von Veränderungen in der Komposition, Organisation, Zellquantität und den extrazellulären Bestandteilen der Atemwegwand definiert werden (Warner et al. 2008). Wie bei der Asthmasymptomatik sind hiervon einige Komponenten durch Behandlung oder spontan reversibel, jedoch werden die meisten Abnormalitäten, so wie sie im Epithel, in der glatten Muskulatur, den Blutgefäßen und der extrazellulären Matrix gefunden werden, als permanent und therapieresistent betrachtet. Die unterschiedlichen Charakteristika lassen darauf schließen, wie vielfältig sich das „airway remodeling“ äußern kann. Dazu gehören Veränderungen der glatten Muskulatur und des Epithels, eine Anhäufung von Fibroblasten und Becherzellen sowie Verdickungen der *lamina reticularis*. Einige jüngere Studien, die sehr gut zusammengefasst wurden (Warner et al. 2008), untersuchen diese Komponenten des „airway remodeling“.

Exemplarisch sei die allgemein erhöhte Proliferationsrate im Epithel erwähnt, die besonders bei Patienten mit schwerem Asthma Gültigkeit hat (Ricciardolo et al. 2003; Cohen, L. et al.

2007). Wenig Zweifel gibt es daher über die Abnormität des Epithels bei Asthmatikern, obwohl dessen substanzielle Schädigung kontrovers diskutiert wird. Es wird angenommen, dass die verminderte innerzelluläre Adhäsion die „epidermal growth factor“ (EGF)-Rezeptor abhängige T_H2 -Zellen begünstigte Aktivität fördert und dadurch das Epithel weiter durch Entzündungsmechanismen schädigt (Ricciardolo et al. 2003). Eine EGF-ähnliche Domäne besitzt beispielsweise auch das Strukturprotein ADAM33.

1.3 Atopie

Eine bedeutende Rolle im Zusammenhang mit Asthma spielt die atopische Sensibilisierung oder Atopie. Als Allergie wird eine überschießende Abwehrreaktion des Immunsystems auf bestimmte und normalerweise harmlose Umweltstoffe, die Allergene, bezeichnet. Eine solche Reaktion äußert sich in typischen, durch entzündliche Prozesse vermittelten Symptomen.

Während genau genommen der Begriff der Allergie Reaktionen des Immunsystems einschließt, die auf verschiedene Mechanismen zurückgehen, ist Atopie stets auf Mechanismen zurückzuführen, die das Immunglobulin E einschließen. Atopie ist IgE-vermittelt und damit ein Teilaspekt der Allergie. Als Atopie wird die Neigung, als Reaktion auf geringe Dosen von Allergenen IgE-Antikörper zu bilden, definiert und als atopische Erkrankungen diejenigen, die mit Atopie, d.h. der Produktion allergenspezifischer IgE-Antikörper einhergehen. Dazu gehören die atopische Dermatitis, der Heuschnupfen und das Asthma bronchiale (von Mutius, E. 1999), wobei sich diese Krankheitsbilder überschneiden (Abbildung 1).

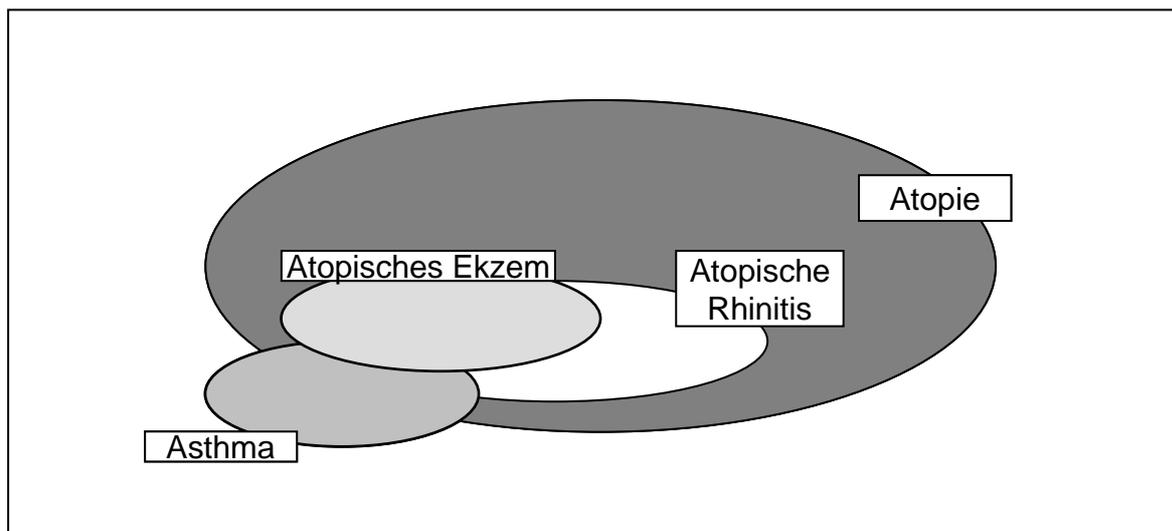


Abbildung 1: Phänotypische Heterogenität der Atopie (in Anlehnung an Feijen et al. (Feijen et al. 2000))

Die spezifischen Antikörper gegen bestimmte Allergene werden als spezifisches IgE bezeichnet und können entweder im Serum mit messtechnischen Methoden nachgewiesen oder indirekt über die Haut durch einen Prick-Test aufgezeigt werden.

1.4 Atopisches und nicht-atopisches Asthma

Atopie spielt insofern eine Rolle bei Asthma, da sich dieses auf Basis des Atopiestatus in verschiedene Subformen einteilen lässt. Je nachdem, ob eine atopische Sensibilisierung vorliegt oder nicht, wird zwischen atopischem (extrinsischem) und nicht-atopischem (intrinsischem) Asthma unterschieden, für deren Entstehung verschiedene Mechanismen hypothetisiert werden. Das extrinsische Asthma bronchiale wird durch Allergene ausgelöst und führt zur sogenannten allergischen Frühreaktion mit einer vermehrten Bildung von IgE und damit erhöhten IgE-Spiegeln im Blut-Serum. Die allergische Reaktion stellt hier die vermutliche Ursache der Symptome dar. Nach intrakutaner Injektion von Extrakten aero gener Antigene zeigt sich die allergische Reaktion in Form von Quaddelbildung und Rötung der Haut.

Eine IgE-vermittelte Reaktion auf verbreitete Allergene ist die häufigste Variante der Krankheit in der Kindheit und im frühen Erwachsenenalter (Platts-Mills et al. 1996). Jedoch dürfte auch bei der nicht-allergischen Form des Asthma eine immunologische Komponente eine Rolle spielen, da die pathologischen Merkmale der Entzündung in beiden Varianten größtenteils ähnlich sind (Bentley et al. 1994; Humbert et al. 1996). Beim intrinsischen Asthma laufen zwar ebenfalls Abwehrreaktionen im Körper ab, diese werden aber nachgewiesenermaßen nicht durch ein Allergen hervorgerufen, denn es kommt seltener zu einem Anstieg von IgE. Oft lässt sich für die Hyperreagibilität der Bronchien keine erklärbare Ursache finden. Als Auslöser des asthmatischen Geschehens können hier meist Virusinfektionen wie der „respiratory syncytial“(RS)-Virus, Rhinovirus, Parainfluenzavirus oder Influenzavirus oder andere Triggermechanismen, wie Kaltluft, Staub, Tabakrauch, körperliche Anstrengung oder emotionaler Stress angenommen werden (Romanet-Manent et al. 2002). Da Kinder mit atopischem Wheeze eine schlechtere Prognose und einen schwereren Verlauf der Krankheit zeigen als solche mit nicht-atopischem Wheeze (Illi et al. 2006), gilt es die phänotypische Heterogenität in die Analysen einzubeziehen. Daher schließt in der vorliegenden Arbeit der Fokus auf Asthma immer auch einen Blick auf die Subformen (atopisches und nicht-atopisches Asthma) ein.

1.5 Das genetische Material

Da ein Grundverständnis einiger Fachbegriffe aus der Genetik bei der Interpretation der vorliegenden Arbeit unabdingbar ist, soll nachfolgend die menschliche Erbsubstanz, die **Desoxyribonukleinsäure** („desoxyribonucleic acid“ bzw. **DNA**), deren Struktur, Funktion und insbesondere deren Variationsmöglichkeit eingeführt werden. Die Hauptvariable der Genetik wird später der „single nucleotide polymorphism“ (SNP) sein (s. 1.5.1.3), womit eine spezielle Variation in der DNA gemeint ist.

Bei den Zellen von sogenannten Eukaryoten, also Pflanzen, Tieren und Pilzen, ist der Großteil der DNA im Zellkern als Chromosomen organisiert. Gene sind Bestandteile der Chromosomen (Abbildung 2) und ein bestimmtes Set von Chromosomen bildet das gesamte Genom.

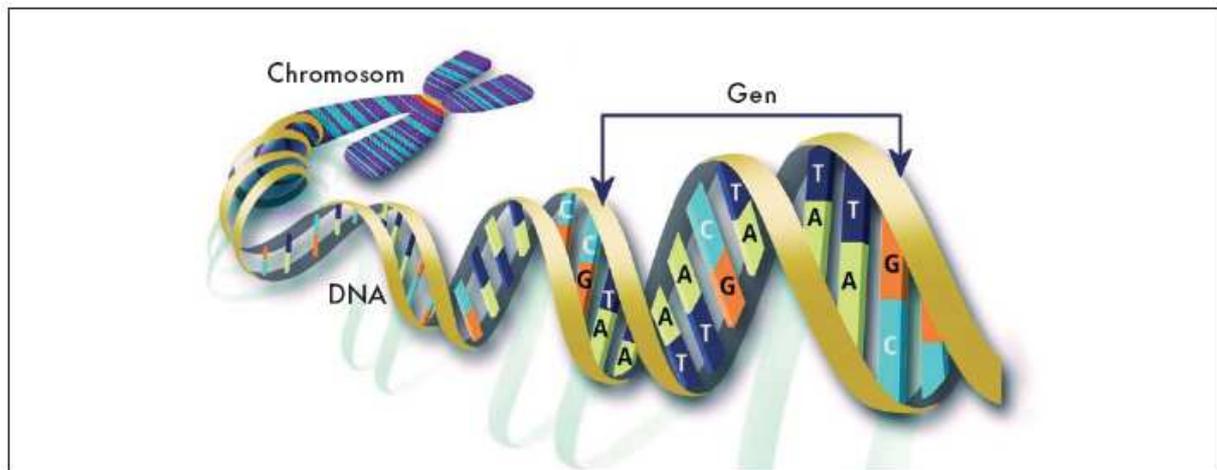


Abbildung 2: DNA und Gene

Chromosomen beinhalten Gene, die fundamentale Bausteine unseres Erbmaterials sind. Gene stellen Baupläne für Proteine dar und bestehen aus dem biologischen Makromolekül DNA (nach Schneeberger et al. 2005)

1.5.1 DNA

1.5.1.1 Struktur der DNA

Das menschliche Genom besteht aus einer langen Sequenz sogenannter Nukleotide, die je nachdem, durch welche Base das Nukleotid gekennzeichnet ist, in vier Formen vorliegen: Adenin (A), Cytosin (C), Guanin (G) und Thymin (T). Die Nukleotide verbinden sich zu einem Molekül (Moleküle sind Teilchen, die aus zwei oder mehreren Atomen bestehen, welche durch kovalente Bindungen verbunden sind) aus zwei Nukleotidsträngen, welches die Form einer Doppelhelix hat. In Abbildung 3 ist ein Beispiel für eine solche DNA-Sequenz gezeigt. Starke Bindungen halten die Basen an einem Strang und schwächere Bindungen verbinden die beiden Stränge, wobei sich immer A mit T und C mit G paart.

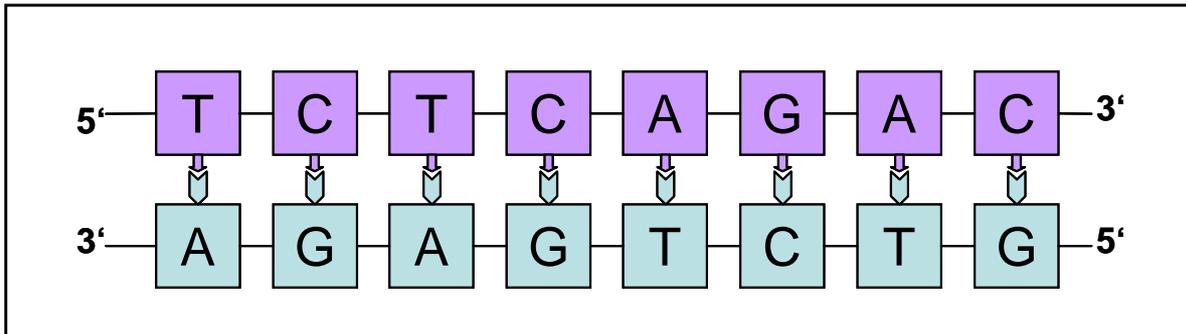


Abbildung 3: DNA Struktur

Jeder Einzelstrang hat zwei verschiedene Enden, welche 5' und 3' genannt werden und die in die entgegengesetzte Richtung orientiert sind. Im Normalfall liegt die DNA im Kern einer Zelle als Doppelstrang vor. Aufgrund der komplementären Struktur kann die DNA sich bei der Zellteilung selbst verdoppeln. Sie teilt sich ähnlich einem Reißverschluss in zwei Einzelstränge, die dann jeweils wieder zu einem Doppelstrang ergänzt werden (Abbildung 4A).

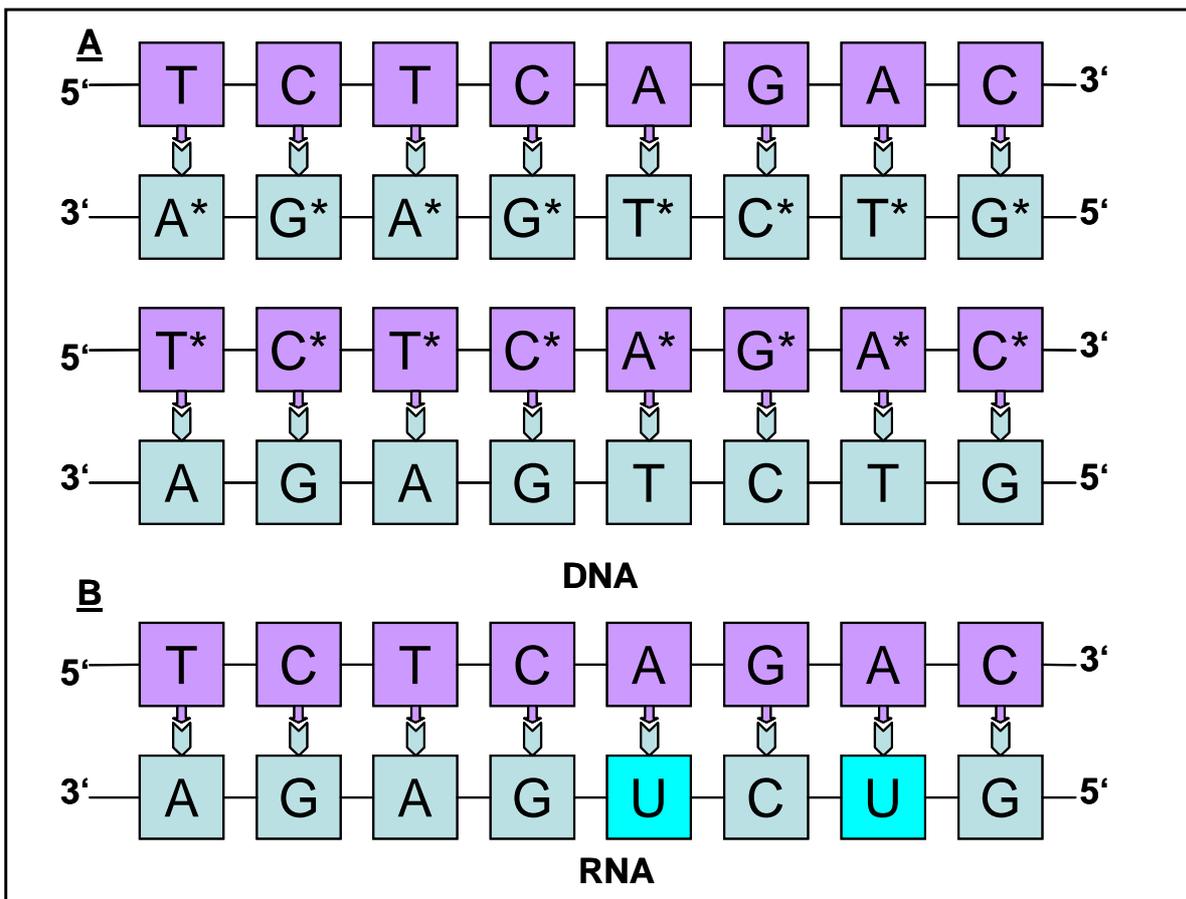


Abbildung 4: DNA Replikation (A) und Transkription (B)
(in Anlehnung an Burton (Burton et al. 2005))

Ein einzelner DNA-Strang kann auch als Vorlage („template“) zur Bildung eines komplementären Strangs von **Ribonukleinsäure** („**ribonucleic acid**“, bzw. **RNA**) dienen. Dieser Vor-

gang wird als **Transkription** bezeichnet, wobei die transkribierte RNA der DNA sehr ähnlich ist. Der einzige Unterschied besteht darin, dass Thymin (T) durch Uracil (U) ersetzt wird (Abbildung 4B).

Durch posttranskriptionale Prozesse wird aus prä-RNA mRNA („messenger RNA“) erstellt. Im Anschluss folgt die **Translation**, bei der mRNA in eine definierte Abfolge von Aminosäuren, welche die Bausteine für ein Protein darstellen, übersetzt wird.

Dieses Protein übernimmt im Körper eine spezifische Funktion. Durch eine Veränderung der DNA-Sequenz kann die Erstellung oder Funktion eines Proteins beeinflusst werden (Abbildung 5). Änderungen des Proteins können sich wiederum auf komplexe physiologische Prozesse und damit den Gesundheitszustand des gesamten Organismus auswirken.

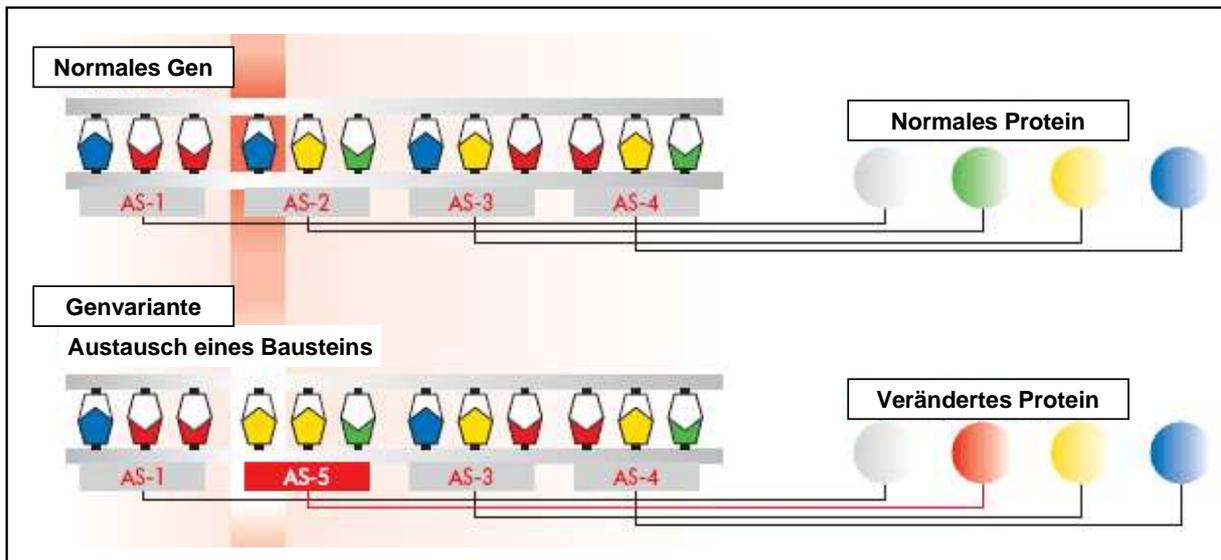


Abbildung 5: Änderung des Proteins durch Änderung im Gen
Eine Veränderung im Gen (unterschiedliche Farben markieren hier unterschiedliche Nukleotide) führt zu einer veränderten Aminosäure (AS), wodurch ein verändertes Protein gebildet wird (Schneeberger et al. 2005)

1.5.1.2 DNA-Abschnitte Gen, Chromosomen, Locus, Genom

Ein **Gen** ist ein Abschnitt auf der DNA, welcher die Grundinformationen zur Herstellung einer biologisch aktiven RNA enthält. Da ein Gen ein Teilstück einer DNA-Sequenz ist, kann man sich auch das Gen als Abfolge von Nukleotiden vorstellen. Dabei werden **Introns**, **Exons** und **untranslatierte Regionen** unterschieden (Abbildung 6). Nach der Transkription findet das Spleißen statt, bei dem bestimmte Elemente aus der mRNA herausgeschnitten werden. Als Exon werden die Teile eines Gens bezeichnet, die nach dem Spleißen erhalten bleiben. Introns werden hingegen beim Spleißen herausgeschnitten und abgebaut (Abbildung 6).

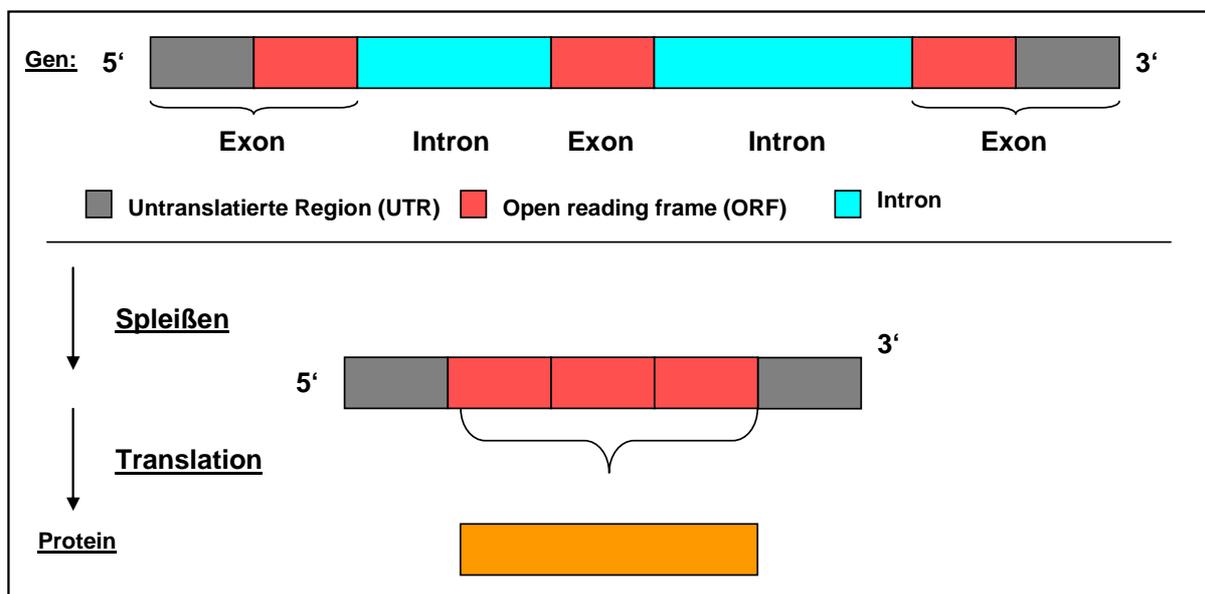


Abbildung 6: Struktur eines Gens (oben) und Änderungen nach der Transkription

Innerhalb der Exons unterscheidet man die translatierten Regionen, also Sequenzen, die in ein Protein übersetzt werden (kodierende Sequenz), und untranslatierte Regionen. Als untranslatierte Regionen werden die Randbereiche der mRNA bezeichnet, welche nicht für Proteine kodieren. Somit bilden nur die in Abbildung 6 rot markierten Bestandteile die kodierende Sequenz, bzw. das **„open reading frame“ (ORF)**, was mit „offener Leserahmen“ übersetzt werden kann.

Während man früher davon ausging, dass ein Gen jeweils die „Bauanleitung“ für ein bestimmtes Protein liefert, geht man heute davon aus, dass ein bestimmtes Gen mit Hilfe komplexer Regulierungsmechanismen verschiedene Proteine kodieren kann. Protein kodierende Gene machen aber nach heutigen Erkenntnissen nur einen kleinen Anteil der DNA aus. Etwa 95 Prozent der DNA sind u. a. regulatorische Einheiten, größtenteils aber Sequenzabschnitte mit bislang unbekannter Funktion (sogenannte „junk DNA“).

1.5.1.2.1 Chromosomen

Die DNA liegt in der Zelle immer in Verbindung mit bestimmten Proteinmolekülen vor. Chromosomen sind Strukturen des Zellkerns, die aus DNA und Proteinen bestehen, und die Erbsubstanz der höheren Organismen repräsentieren. Die DNA ist dabei in den Chromosomen zu kompakten Strukturen zusammengefaltet und mit bestimmten Proteinen verbunden. Die einzelnen DNA-Protein-Pakete werden als Nucleosomen bezeichnet.

Bei jeder Art von Organismen liegen die Chromosomen in charakteristischer Anzahl und Form vor. In humanen Körperzellen befindet sich ein doppelter Chromosomensatz, das heißt, es gibt in der Regel zwei Chromosomen, die in Form und Größe gleich und in der Sequenz ähnlich sind. Man nennt diese Chromosomen **homologe** Chromosomen. Der Homologiebegriff wird später aber auch in einem weiteren Sinne verwendet und bezeichnet dann Ähnlichkeiten in der DNA-Sequenz zwischen verschiedenen Spezies (s. 3.1.2.2).

In Organismen, die sich geschlechtlich fortpflanzen, wird bei zwei homologen Chromosomen je eine Kopie vom Vater und eine von der Mutter vererbt. Zellen, die einen doppelten Chromosomensatz beinhalten, werden als **diploide** Zellen bezeichnet. Zellen, die einen einfachen Chromosomensatz tragen, werden als **haploid** bezeichnet. Abbildung 7 verdeutlicht den Unterschied zwischen haploid und diploid für eine bestimmte DNA-Sequenz.

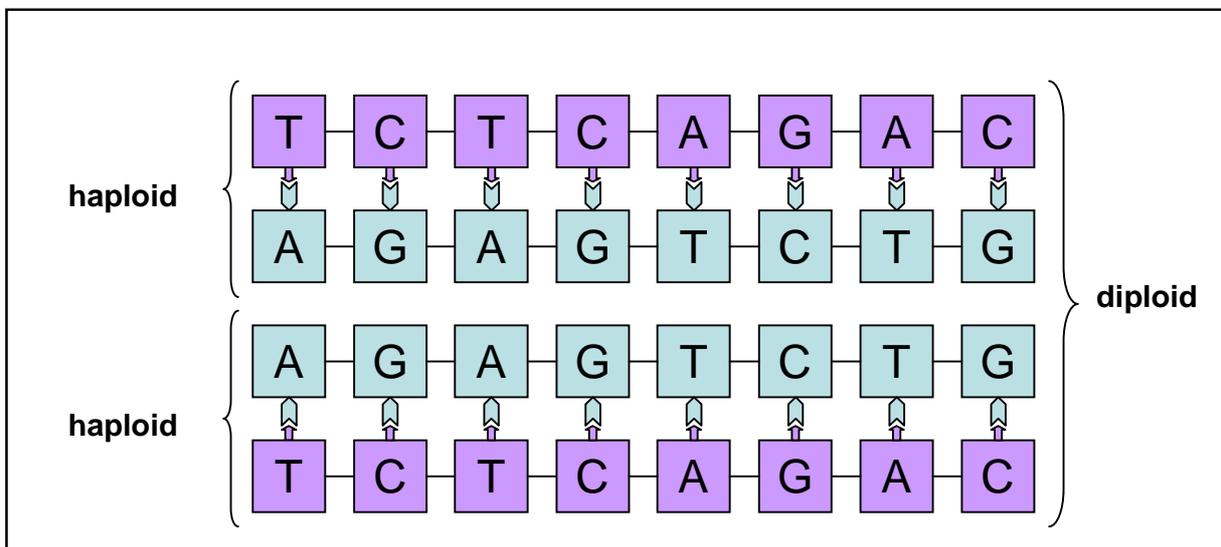


Abbildung 7: Beispiel für ein Teilstück der DNA in haploider und diploider Form ohne Variation. Ein Nachkomme erbt die haploiden Teilstücke des gleichen DNA-Abschnitts von Mutter und Vater, die zusammen wieder diploide Erbinformation tragen.

Das menschliche Genom besteht aus einem diploiden Chromosomensatz von 23 Chromosomenpaaren, 22 autosomalen und einem Geschlechtschromosomenpaar, XX (weiblich) oder XY (männlich). Befruchtete Eizellen und die meisten Zellen im Körper, die daraus hervorge-

hen, tragen zwei Kopien des Genoms und sind demzufolge diploid. Insofern liegt dort auch jedes Chromosom, jedes Gen und jedes Teilstück eines Gens in einer diploiden Form vor. Ein Individuum trägt somit in jeder Körperzelle die gesamte Erbinformation.

In der Meiose (Geschlechtsteilung) wird der diploide Chromosomensatz halbiert, so dass Keimzellen sowie unbefruchtete Eizellen nur einen haploiden Chromosomensatz tragen. Nach der Befruchtung trägt die Eizelle wieder einen diploiden Chromosomensatz, der sich aus den haploiden Chromosomensätzen zusammensetzt, die je eine Variante von Vater und Mutter repräsentieren.

Ein menschliches Chromosom ist im Durchschnitt nur etwa zehn Mikrometer groß und ist in Abbildung 8 dargestellt.

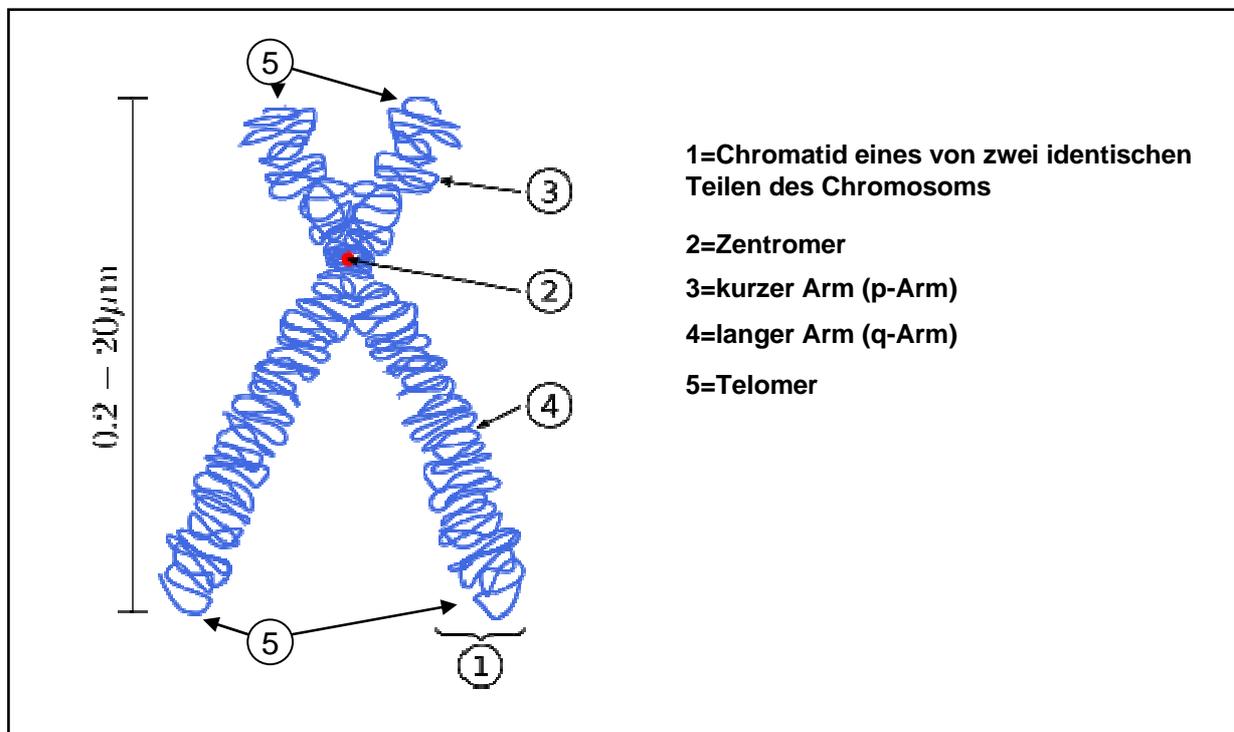


Abbildung 8: Darstellung eines Chromosoms

Der Bereich, an dem die väterlicher- und mütterlicherseits vererbten DNA-Doppelstränge, die sogenannten Chromatiden, zusammenhängen, wird als Zentromer bezeichnet. Das Zentromer teilt das Chromosom in zwei Arme, die als p (frz. „petite“) für den kurzen oder q (frz. „queue“) für den langen Arm stehen.

Innerhalb jeden Armes gibt es Banden, die beim Anfärben der Chromosomen entstehen. Entsprechend dieser Banden kann ein Chromosom in Regionen unterteilt werden, z.B. 11q23: die dritte Bande der zweiten Region auf dem langen Arm von Chromosom 11.

1.5.1.2.2 Locus

Als **Locus** wird eine eindeutige chromosomale Lage bezeichnet, welche die physikalische Position eines individuellen Gens oder einer DNA-Sequenz im Genom definiert. Oftmals wird unter einem Locus ein bestimmter Genort verstanden. Genau genommen kann der Term Locus sich aber auch auf Regionen beziehen, die mehrere Gene einschließen und dabei auch nicht kodierende Sequenzen der DNA umfassen.

1.5.1.2.3 Genom

Die komplette DNA-Sequenz entspricht dem menschlichen Genom und bildet das Repertoire der Proteine für das Proteom, die Gesamtheit aller Proteine in einem Lebewesen. Durch Sequenzierung wurde das menschliche Genom komplett bestimmt. Dabei enthalten die Chromosomen jeweils eine unterschiedliche Anzahl von Genen. Gene, die für ein Protein kodieren können, werden innerhalb des menschlichen Genoms auf etwa 20000 bis 25000 geschätzt (*Finishing the euchromatic sequence of the human genome* (2004)).

1.5.1.3 Variationen der DNA-Sequenz

Abbildung 9 zeigt die Möglichkeit der Variation in der DNA-Sequenz, denn an der markierten Stelle sind verschiedene Nukleotide zu erkennen. Falls solch eine Variationsmöglichkeit besteht, kann das Nukleotid zwischen zwei homologen Chromosomen innerhalb eines Individuums variieren.

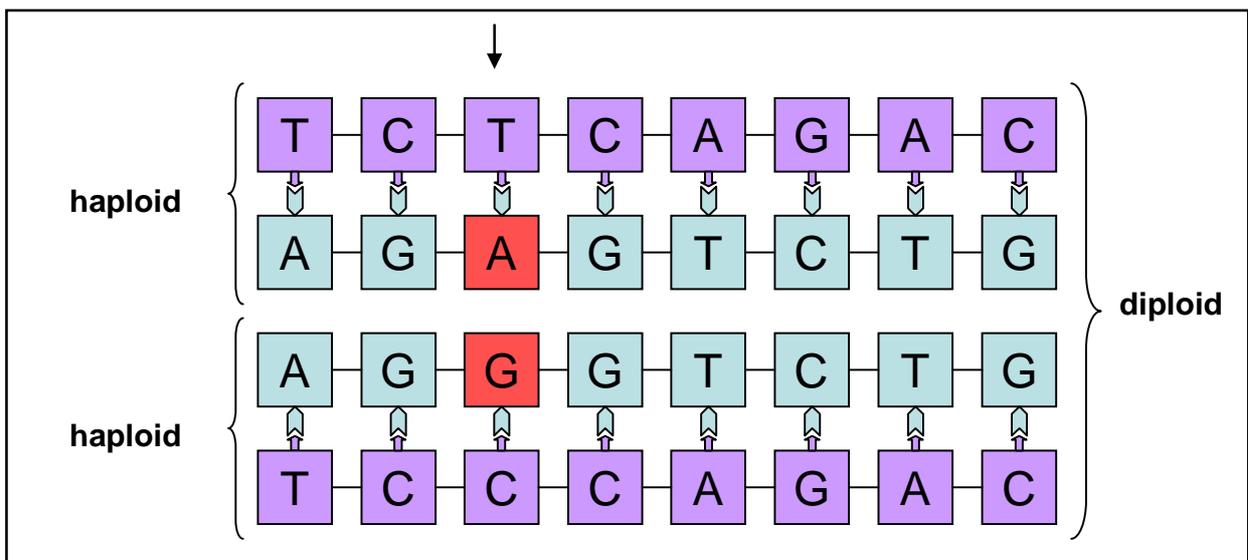


Abbildung 9: Beispiel für ein Teilstück der DNA in haploider und diploider Form mit Möglichkeit der Variation.

Eigentlich möchte man Variation aber zwischen Individuen betrachten, wenn sie als Einflussfaktor auf Krankheiten gelten soll.

Wenn die DNA-Sequenz an einer bestimmten Stelle, einem bestimmten Locus, zwischen verschiedenen Individuen in der Population variiert, werden die unterschiedlichen Versionen der DNA-Sequenz als **Allele** bezeichnet. Der Begriff des Allels bezieht sich somit auf mögliche Variation innerhalb einer Population und bedeutet nicht, dass es die Variation innerhalb eines Individuums geben muss. Die einfachste Form eines Allels besteht darin, dass es eine Variation innerhalb eines einzigen Nukleotids gibt. So kann in Abbildung 10 das Allel sowohl als G als auch als A auftreten.

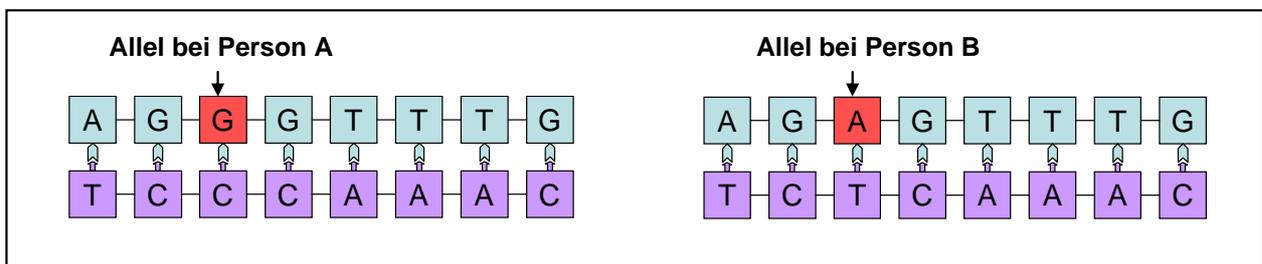


Abbildung 10: Beispiel für ein Allel.
Dargestellt sind zwei DNA-Sequenzen, die an einer Stelle variieren.

Wenn es an einem Genort zwei Allele gibt, so wird das in einer Population weniger häufig auftretende Allel als „minor allele“, das häufiger auftretende als „wild type allele“ bezeichnet. Entsprechend spricht man von der „minor allele frequency“ (**MAF**), wenn man die Häufigkeit des seltenen Allels in einer Population beschreibt.

Kehrt man von der Populationsebene auf die individuelle Ebene zurück, so gilt es sich noch einmal bewusst zu machen, dass die Zellen in der Regel diploid sind und damit die entsprechende Sequenz innerhalb eines Individuums zweimal getragen wird. Dies bedeutet, dass innerhalb des doppelten Chromosomensatzes das Allel am entsprechenden Locus zweimal in der gleichen Ausprägung (A/A oder G/G) oder in unterschiedlichen Ausprägungen (A/G) vorliegen kann. Abbildung 11 zeigt die drei Varianten und ist die graphische Veranschaulichung eines **SNPs**.

Mit **„single nucleotide polymorphisms“ (SNPs)** werden Loci in einer DNA-Sequenz bezeichnet, die in einer einzigen Base variieren. Da Menschen einen doppelten Chromosomensatz haben, hat jeder Mensch in seinen diploiden Zellen auf den beiden homologen Chromosomen am betreffenden Genort entweder zwei unterschiedliche Allele eines Gens oder aber zwei gleiche Allele des betreffenden Gens. Im ersten Fall spricht man von Heterozygotie, im zweiten Fall von Homozygotie.

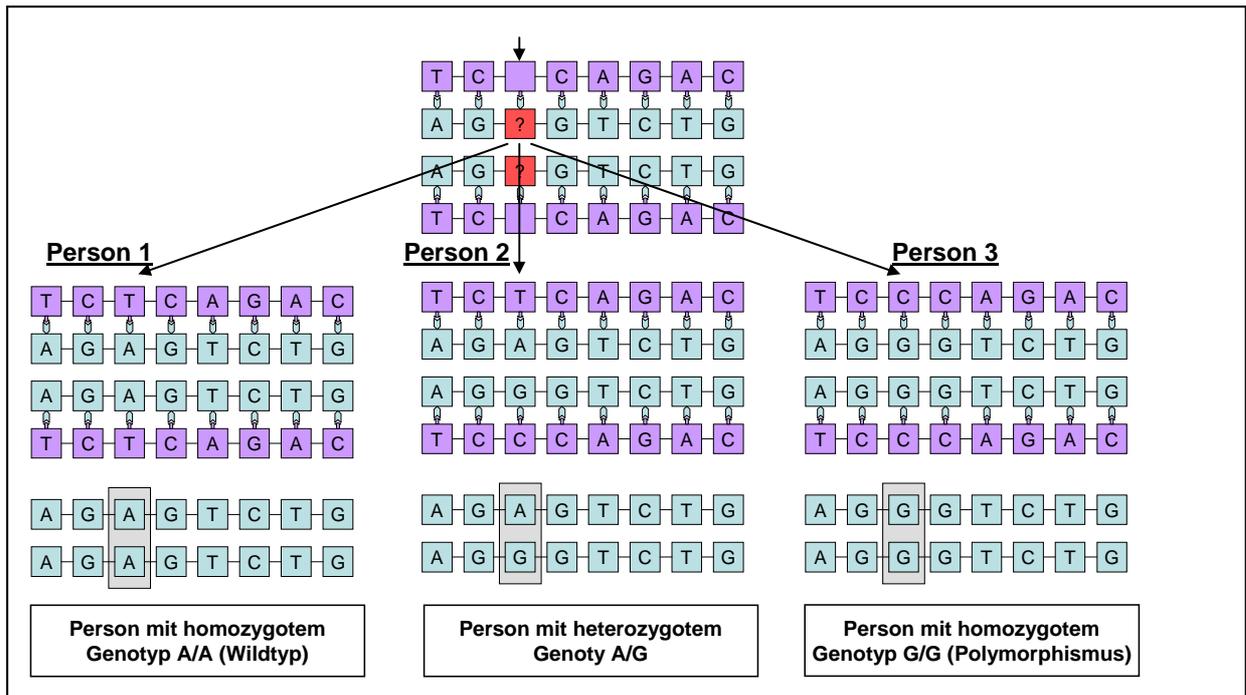


Abbildung 11: Beispiel für einen SNP und die 3 Varianten, die eine Person tragen kann. Je nach Kombination der beiden Allele spricht man von homozygotem oder heterozygotem Genotyp. A sei das häufige Allel, dann ist AA der Wildtyp.

Ein SNP kann somit drei Ausprägungen haben:

1. zwei der (in der Population) häufig auftretenden Allele: Wildtyp (homozygot)
2. zwei unterschiedliche Allele: heterozygot(er) Genotyp
3. zwei der (in der Population) selten auftretenden Allele: Polymorphismus (homozygot)

Bei einer Person kann immer nur eine dieser drei Varianten eines Genotyps am gleichen Locus auftreten und Individuen können danach unterschieden werden, welchen Typ sie tragen. In der Regel gibt es in jedem Gen mehrere Allele, die Unterschiede in der DNA-Sequenz aufweisen und somit auch eine Vielzahl von SNPs.

Phylogenetisch kann man sich die Entstehung einer Variation vorstellen, indem man annimmt, dass irgendwann in einer Population bei einer Person ein Nukleotid auftritt, was zuvor nicht an dieser Position der DNA zu finden war, z.B. ein G, wo zuvor nur A war. Dies wird dann als Mutation bezeichnet.

Setzt sich ein neues Allel in einer Population durch, so bedeutet dies, dass an dieser Stelle in der DNA ein Gen bezüglich seines Nukleotids zu einem gewissen Grad (z.B. zu mehr als 1 %) im Genpool einer Population variieren kann. Durch diese höhere Frequenz werden SNPs von Mutationen abgegrenzt (Crawford et al. 2005), deren Frequenz kleiner ist. Manchmal werden SNPs daher auch als „erfolgreiche“ Punktmutationen bezeichnet, das heißt als genetische Veränderungen an einem Punkt der DNA-Sequenz, die sich durchgesetzt haben.

Es wurde geschätzt, dass im menschlichen Genom etwa 10 Millionen solcher SNPs vorkommen (2003) und dass diese etwa 90% der Variation in der Population erklären (Kruglyak et al. 2001). Neben den SNPs gibt es andere Variationsmöglichkeiten, die anderweitig erklärt sind (Burton et al. 2005) und im Rahmen der vorliegenden Arbeit keine Rolle spielen.

Unterschiedliche Allele finden sich auch in Regionen der DNA, die keine Protein kodierende Funktion haben. SNPs in Regionen, die für ein Protein kodieren, werden als nicht-synonym oder synonym bezeichnet, je nachdem, ob eine Änderung in der Sequenz der DNA die Sequenz der Aminosäuren im Genprodukt verändert. Synonyme SNPs ändern die Aminosäure nicht. Nicht-synonyme SNPs bewirken eine Änderung in der Aminosäure und werden auch als kodierende SNPs bezeichnet. Es sind jedoch nicht ausschließlich SNPs, die für eine Aminosäureänderung kodieren, welche Einfluss auf die Ausprägung der Phänotypen nehmen.

Während der Genotyp auf genetische Information bezogen ist, handelt es sich beim **Phänotyp** um nicht-genetische, messbare Charakteristika eines Individuums, also hier zum Beispiel die atopischen Erkrankungen oder deren Symptome.

Auch andere Regulationsmechanismen, wie zum Beispiel die Regulation der Transkription eines Proteins, können Einfluss auf eine Krankheit oder gewisse Phänotypen nehmen (Burton et al. 2005).

Gelegentlich wird in der genetischen Epidemiologie die genetische Ausprägung eines Individuums an einem Locus als Genotyp bezeichnet, wobei hier die Allele des diploiden Chromosomensatzes berücksichtigt sind. Dies ist abzugrenzen von Haplotypen, welche nur Informationen eines haploiden Chromosomensatzes berücksichtigen (Abbildung 12).

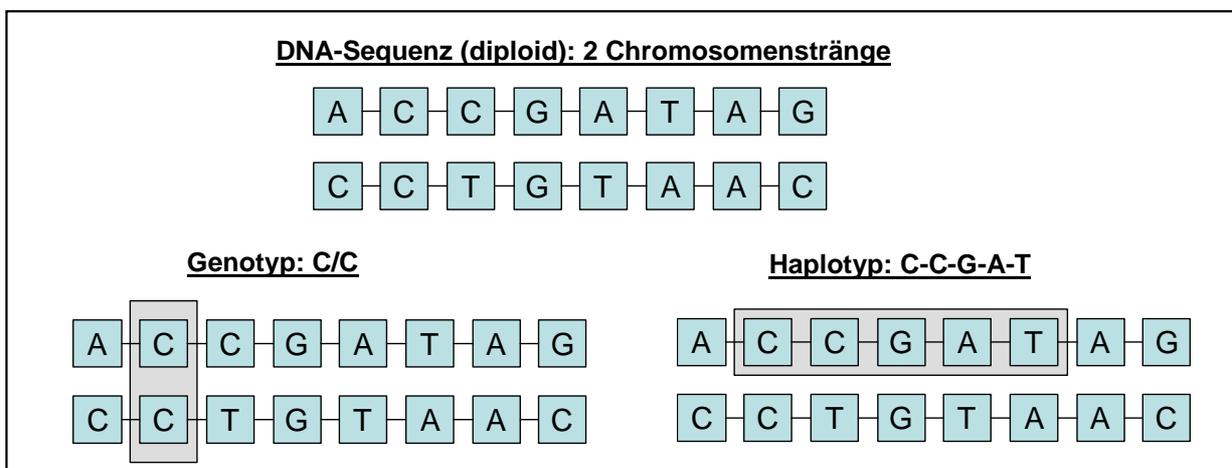


Abbildung 12: Unterscheidung von Genotyp und Haplotyp

Die allelische Konfiguration entlang eines Chromosoms wird **Haplotyp** genannt und die Information über den Haplotyp wird als **Phase** bezeichnet. Haplotypen müssen nicht zusammenhängende DNA-Sequenzen sein, sind aber als allelische Kombination innerhalb eines Chromosoms definiert.

Ein **genetischer Marker** ist eine variable DNA-Sequenz, welche sowohl eine hinreichend spezifische nichtvariable DNA-Sequenz hat, um einen Genort zu lokalisieren, als auch eine hinreichend heterogene Sequenz besitzt, um Differenzen zwischen Individuen zu identifizieren. Mittels genetischer Marker kann man somit Variation innerhalb des Genoms auf bestimmte Regionen einengen. SNPs können durch die umliegende DNA-Sequenz eindeutig lokalisiert werden (Abbildung 13).

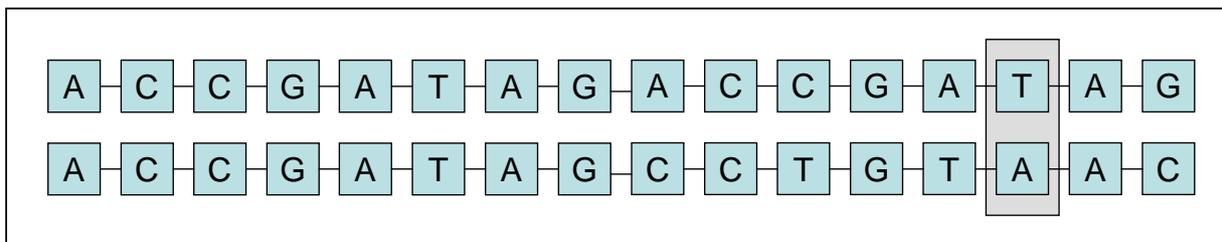


Abbildung 13: SNP als Marker (nur ein Ausschnitt der umliegenden Sequenz ist gezeigt)

1.5.2 Weitergabe genetischen Materials

1.5.2.1 Rekombination

In der Meiose wird über **Rekombination** das Erbmateriale von jedem Elternteil neu kombiniert und über die geschlechtliche Fortpflanzung in der Zygote (der befruchteten Eizelle) zu einem neuen Individuum aggregiert (Busselmaier et al. 2007). Unter **Rekombination** versteht man im engeren Sinne den Austausch von Allelen und im weiteren Sinne die Verteilung und Neuordnung von genetischem Material (DNA, RNA). Durch Rekombination kommt es zu neuen Gensequenzen und Merkmalskombinationen.

Da diploide Lebewesen über einen doppelten Chromosomensatz verfügen, muss bereits bei der Entstehung der Keimzellen, die haploid sind, entschieden werden, in welcher Form die Auswahl des weiterzugebenden Erbmaterials aus dem schon Vorhandenen geschehen soll. Homologe Chromosomen werden in der Meiose bei jedem Elternteil getrennt, so dass eine 50%-ige Wahrscheinlichkeit besteht, eine Kopie eines homologen Chromosomenpaares von einem Elternteil zu erhalten. Da ein Elternteil 23 Chromosomenpaare hat, von denen jeweils nur der haploide Satz weiter vererbt wird, gibt es bereits bei der Auswahl des haploiden

Chromosomensatzes für die Weitervererbung 2^{23} Möglichkeiten der Kombination von Chromosomen bei jedem Elternteil.

Obwohl bereits interchromosomale Rekombination es nahezu unmöglich macht, das gleiche genetische Material eines Elternteils zu erhalten, gibt es weitere Mechanismen, welche die Variation potenzieren. Zwei homologe Chromosomen können kurze Abschnitte untereinander austauschen, indem sie jeweils an gleicher Stelle brechen und sich dann über Kreuz wieder zusammenfügen. Dies wird als **„crossing over“** bezeichnet. Auf diese Weise kommt es zu einer Rekombination väterlichen Erbmaterials mit dem mütterlichen (Abbildung 14).

Dabei erhält eine Keimzelle nach „crossing over“ eine Mischung von Teilen zwei homologer Chromosomen. Das „crossing over“ wird durch sehr komplexe Mechanismen reguliert (Cromie et al. 2007).

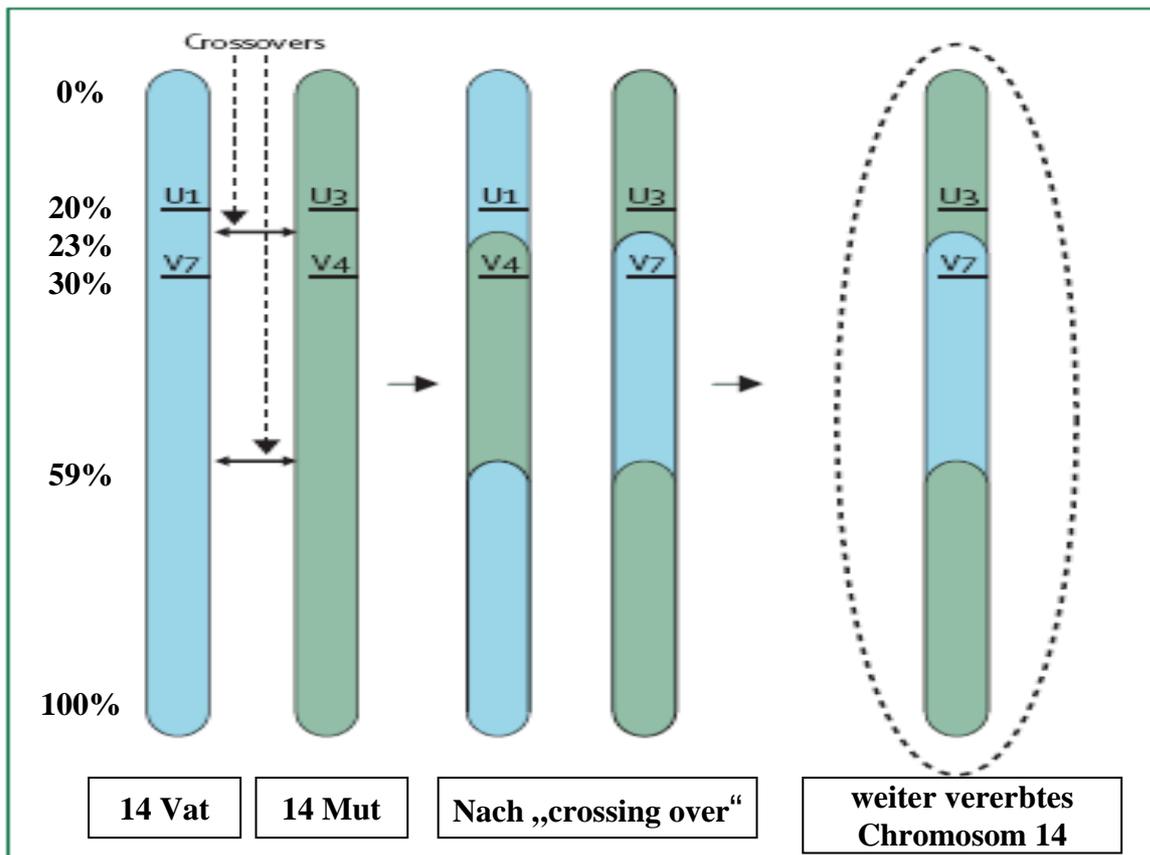


Abbildung 14: Beispiel für „crossing over“.

Zwei hypothetische Loci, U und V, nehmen 20 oder 30% der Größe entlang der Länge des Chromosoms 14 ein. Sie existieren als Allel U1 und V7 auf Chromosom 14 Vat (das Chromosom, das ursprünglich vom Vater kam) und als Allel U3 und V4 auf dem Chromosom 14 Mut (das Chromosom, das ursprünglich von der Mutter kam).

Crossovers bei 23% und 59% entlang des Chromosoms produzieren zwei gemischte Chromosomen. In diesem Beispiel wird das rechte Chromosom, welches die Allele U3 und V7 trägt, weitervererbt. Diese zwei Allele wurden unabhängig von Vater und Mutter entwickelt (nach Burton (Burton et al. 2005)).

1.5.2.2 Kopplungsungleichgewicht

Bei der genetischen Kartierung, der Bestimmung der Lage eines Gens mit bekannter Funktion auf einem DNA-Molekül, macht man sich das Auftreten der Rekombination zunutze. Je weiter zwei Loci voneinander entfernt sind, desto größer ist die Wahrscheinlichkeit von „crossing over“. Der Rekombinationsanteil, also der Anteil von Meiosen, welche in Rekombination resultieren, wird als Indikator für die Entfernung zweier Gene herangezogen. Dieser Indikator kann in eine erwartete Anzahl von „crossing overs“ umgerechnet werden. Regionen innerhalb des Genoms, in welchen viel oder wenig Rekombinationsereignisse stattfinden, werden als „hot spots“ und „cold spots“ bezeichnet (Petes 2001).

Ohne Rekombination werden nebeneinander liegende Loci gemeinsam oder gekoppelt weiter vererbt. Kopplung kann zu Kopplungsungleichgewicht führen. Eine gebräuchliche Definition von **Kopplungsungleichgewicht** („**linkage disequilibrium**“ oder **LD**) lautet: Sind zwei Allele in einer Population an zwei oder mehreren Genorten häufiger oder weniger häufig verknüpft als nach ihren individuellen (Allel-)Frequenzen zu erwarten, dann befinden sie sich im Kopplungsungleichgewicht; andernfalls sind sie im Kopplungsgleichgewicht.

Man beachte, dass LD ein Maß der Populationsgenetik ist und ähnlich einer Korrelation den Zusammenhang zwischen zwei Genotypen beschreibt.

Sobald eine Variation in die Population eingeführt wird, ist sie perfekt mit einer benachbarten Variante korreliert. Jedoch kommt es im Laufe der nachfolgenden Generationen zu Rekombinationsereignissen, weshalb die Korrelation aufgebrochen wird. Der Einfluss der Rekombinationsrate c ist in Abbildung 15 dargestellt.

Beim Auftreten einer Neumutation in einer Population, gibt es komplettes LD, das heißt die Krankheitsmutation tritt ausschließlich mit einem Set bestimmter Marker auf. Über die nachfolgenden Generationen nimmt dann das LD ab (Palmer et al. 2005).

Da LD auch durch phylogenetische Einflüsse (wie Mutation, die zufällige Veränderung von Allelfrequenzen („random drift“) oder natürliche Selektion) oder Vererbungsgeschichte einer Population (wie historische Populationsgrößen, Intensität und Arten künstlicher Selektion, Populationsmischungen oder Vererbungsmuster) beeinflusst sein kann, ist die Annahme, dass Änderungen des LD nur auf Rekombination zurückzuführen sind, eine zu starke Vereinfachung.

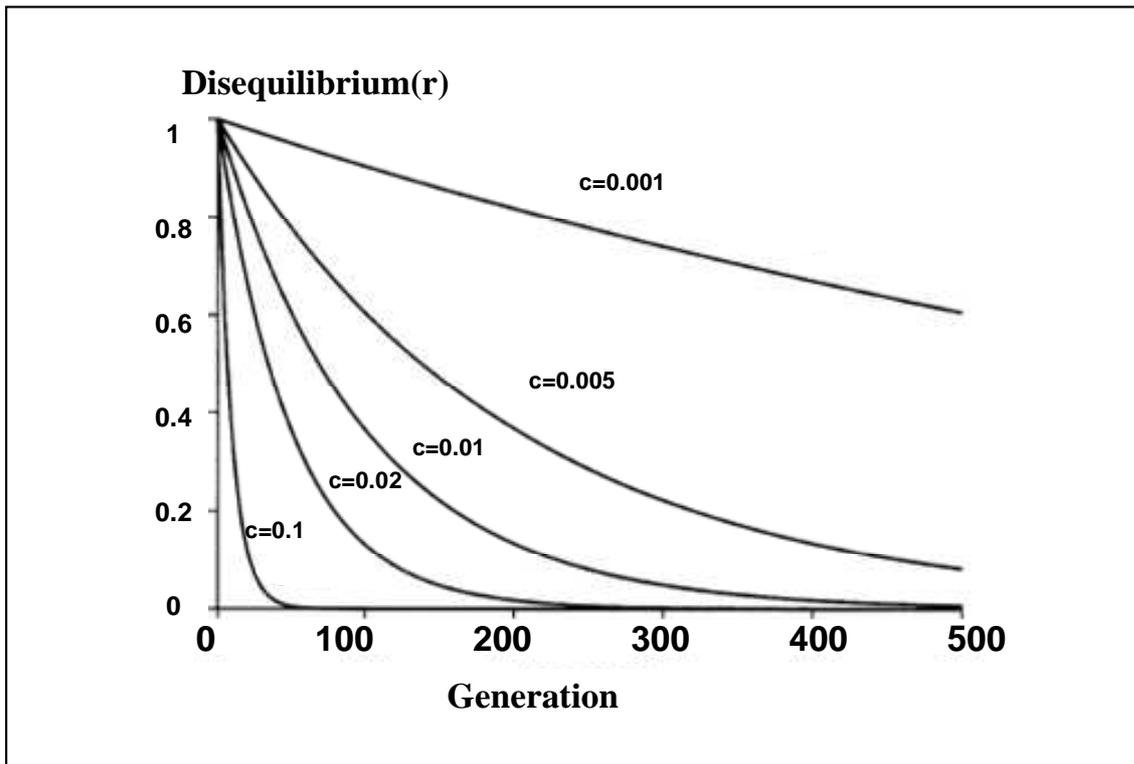


Abbildung 15: Beziehung zwischen LD (r , y-Achse) und der Anzahl von Generationen, seit die Mutation in die Population eingeführt wurde (x-Achse). Werte von r werden für verschiedene Rekombinationsraten c geplottet. Dabei legt der Graph die Standardgleichung $r_t = r_0(1-c)^t$ zugrunde, wobei t die aktuelle Generation, r_t Disequilibrium in Generation t und r_0 das anfängliche Disequilibrium in Generation 0, c die Rekombinationsrate (Jorde 1995)

Für die Kartierung einer genetischen Region hofft man dennoch, dass Rekombination die anderen Einflussquellen dominiert. Unter dieser Annahme lassen sich dann Loci ausmachen, die mehr oder weniger stark mit einem potenziellen Krankheitslocus in Zusammenhang stehen, bzw. mit diesem gekoppelt sind (Ennis et al. 2001). Dies liegt daran, dass nun hohes Kopplungsungleichgewicht auf gekoppelte Vererbung zurückzuführen ist und von Regionen in hohem LD angenommen werden kann, dass sie etwas miteinander zu tun haben.

1.5.2.3 Erbgänge

Im Rahmen der später durchgeführten Analysen für Zusammenhänge zwischen Genotypen und Phänotypen wird auf Annahmen zurückgegriffen, denen verschiedene Erbgänge zugrunde liegen.

Mendelsche Erbgänge sind rezessiv und dominant (Chudley 1998). Der dominante Erbgang ist eine Form der Vererbung, bei der bereits ein defektes Allel auf einem der beiden homologen Chromosomen zur Merkmalssausprägung reicht. Dies bedeutet, dass Allele sich bereits in heterozygotem Zustand in ihrer Wirkung auf den Phänotyp beobachtbar auswirken. Der rezessive Erbgang ist hingegen eine Form der Vererbung, bei dem defekte Allele auf bei-

den homologen Chromosomen vorliegen müssen, damit die Krankheit bzw. das Merkmal zum Ausbruch kommt. Es erkranken nur homozygote Träger des betroffenen Allels. In Abbildung 16 ist jeweils gezeigt, wie sich der Phänotyp in Abhängigkeit von Genotyp und Erbgang ergibt. Dabei sei das farbig markierte Puzzleteil der vererbte Phänotyp, z.B. eine vererbte Krankheit. Für den dominanten Erbgang gilt zum Beispiel, dass Genotyp 2 und Genotyp 3 den kranken Phänotyp bedingen.

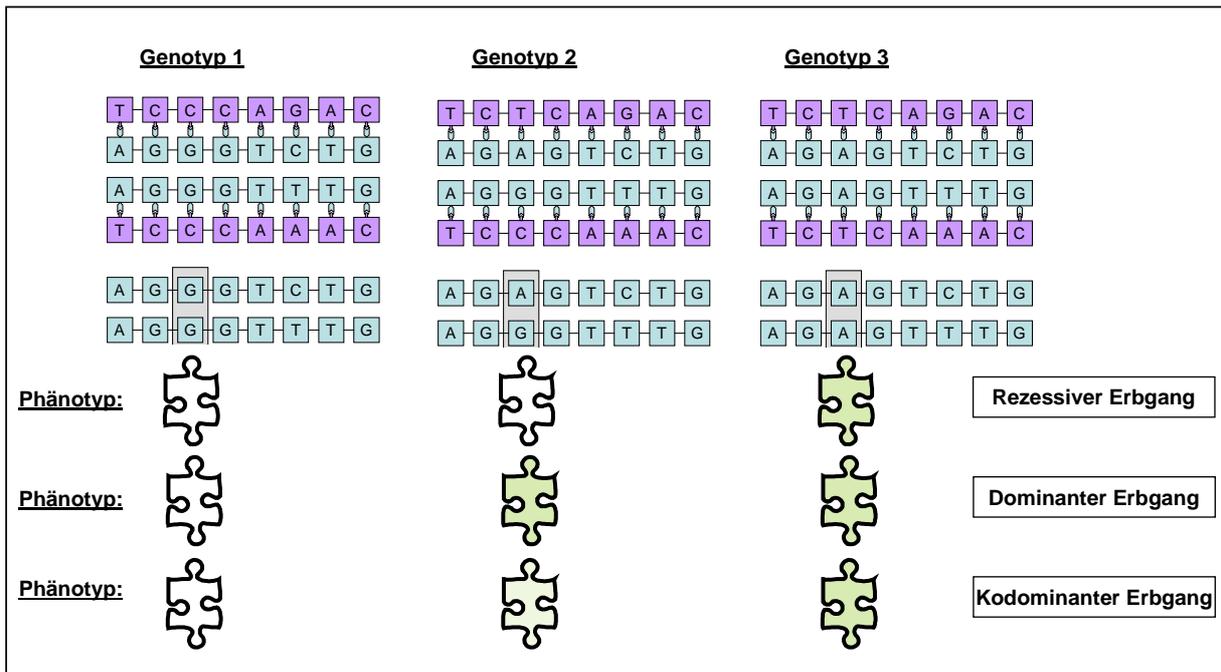


Abbildung 16: Beispiel für Mendelsche Vererbung: rezessiver, dominanter und kodominanter Erbgang. Weiß bedeutet gesunder Phänotyp. Farbig bedeutet pathologischer Phänotyp. Risikoallel ist A. Dargestellt sind die Phänotypen in Abhängigkeit von Genotyp und Erbgang.

Der Begriff der kodominanten Vererbung umschreibt in allgemeiner Form die Möglichkeit, dass die beiden Formen, die für ein Allelpaar homozygot sind, vom heterozygoten Zustand im Phänotyp unterschiedlich sind (dritte Variante in Abbildung 16).

Eine Übertragung der Allelkonfiguration an einem SNP auf ein kodominantes Modell sind die additiven Modelle: Bei diesen wird die Anzahl von 0 (Wildtyp), 1 (heterozygoter Typ) und 2 (Polymorphismus) Risikoallelen unterschieden. Dies wird häufig einem kodominanten Vererbungsgang zugeordnet, bei welchem das Vorhandensein eines Risikoallels zu einem intermediären Phänotyp führt. Manchmal wird auch von Gen-Dosis-Effekt gesprochen, da man mit einer Häufung von Risikoallelen ein größeres Risiko für die Ausprägung eines pathologischen Phänotyps verbindet.

Da sich die Mendelschen Erbgänge jeweils auf Loci in bestimmten Chromosomen beziehen, spricht man auch von autosomal dominant oder autosomal rezessiver Vererbung, wenn das zu vererbende Merkmal auf einem Autosom liegt. Ein X-chromosomaler Erbgang ist dagegen ein Erbgang, bei dem das merkmalstragende Gen auf dem X-Chromosom liegt. X-chromosomale Erbgänge sind von autosomalen Erbgängen abzugrenzen, da die Vererbungsmuster hier bei Frauen und Männern unterschiedlich ablaufen. Männliche Individuen haben in der Regel neben dem Y-Chromosom ein X Chromosom und werden daher als hemizygot bezeichnet. Darum tritt bei ihnen ein phänotypisches Merkmal, das auf dem X-Chromosom kodiert ist, auch bei rezessiver Vererbung immer auf (zu den genauen Vererbungsmodalitäten siehe z.B. Busselmaier (Busselmaier et al. 2007)).

Bei **monogenen Erkrankungen** sind Mutationen in einem Gen hinreichend, um den klinischen Phänotyp zu produzieren. Bei **komplexen Störungen** mit multiplen Ursachen resultiert erst die Kombination von Variationen in mehreren Genen, die für verschiedene Proteine kodieren, in einer genetischen Prädisposition für einen klinischen Phänotyp (Peltonen et al. 2001). Dadurch haben die Vererbungsmuster, die auf Mendel zurückgehen, dort keine Gültigkeit mehr. Mutationen in einem Gen sind hier oftmals weder hinreichend noch notwendig, den Phänotyp zum Ausbruch zu bringen. Die Hauptunterscheidung zwischen monogenen und komplexen genetischen Erkrankungen besteht somit darin, dass Letztere nicht den Mendelschen Regeln der Vererbung folgen (Barnette et al. 2005). Abbildung 17 zeigt, wie man sich eine komplexe Erkrankung stattdessen vorstellen kann.

Genetische Variationen gibt es an verschiedenen Stellen im Genom. Dabei ist bei den komplexen Erkrankungen die Variation an einer Stelle nicht mehr hinreichend, den Phänotyp zu erklären. Stattdessen wird angenommen, dass bestimmte Ausprägungen des Genotyps an unterschiedlichen Loci zusammentreffen müssen, um die Entwicklung der Krankheit zu begünstigen. Auch weiß man nicht, welche Art der Variation in der DNA (SNP, Haplotyp, anderes) und welcher Erbgang die Krankheit am besten erklärt. Es ist nicht einmal auszuschließen, dass an unterschiedlichen Loci andere Erbgänge stattfinden können.

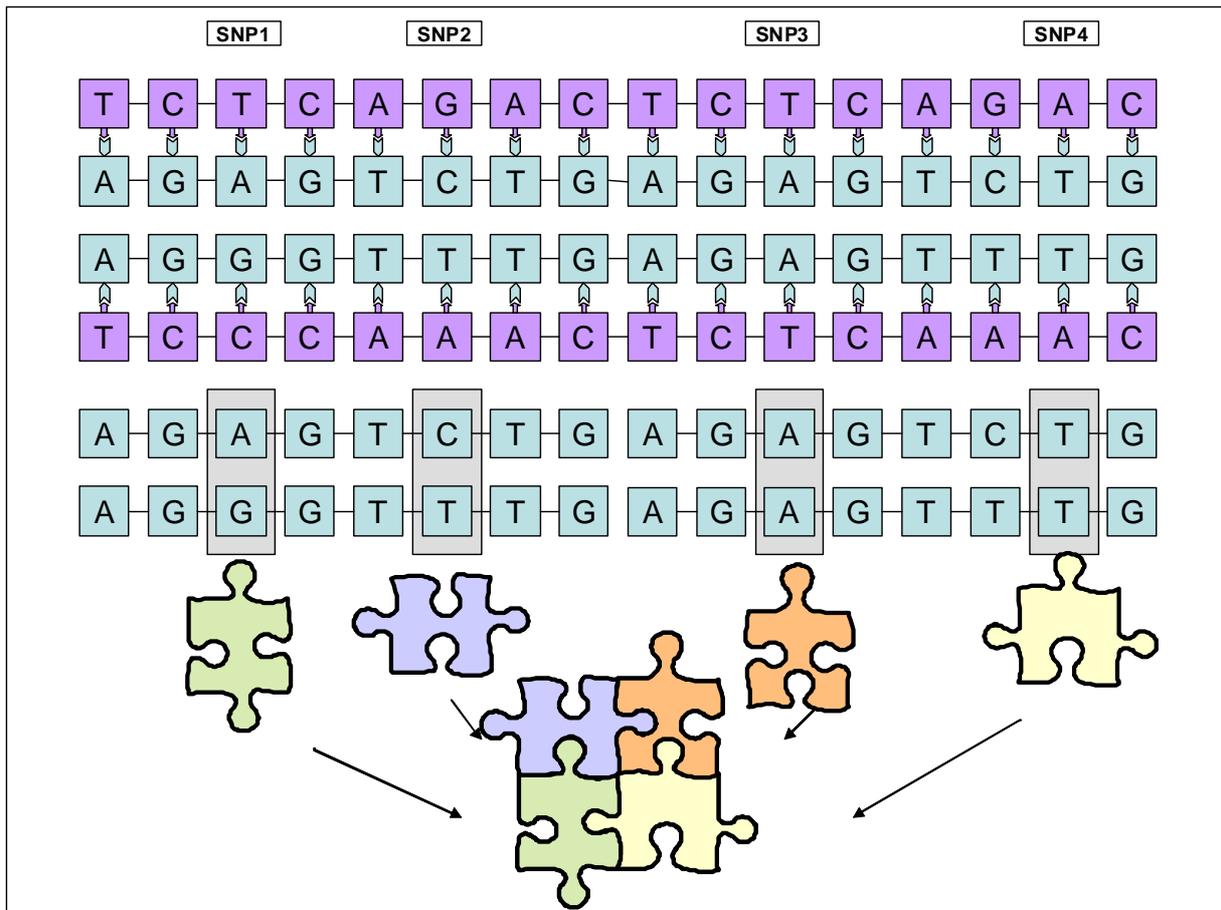


Abbildung 17: Modell für eine komplexe Erkrankung.
 Verschiedene Variationen an verschiedenen Stellen im Genom sind verantwortlich, die komplexe Erkrankung zu erklären

Diese enorme Komplexität macht es so schwierig, genetische Determinanten komplexer Erkrankungen zu bestimmen. Bei Asthma geht man davon aus, dass es sich um eine komplexe Erkrankung handelt. Da auch hier die genetischen Determinanten lange nicht erschlossen sind, werden Effekte von vielen einzelnen Einflussgrößen und somit auch von einzelnen Genen untersucht. Man erwartet dabei jedoch auf Grund des komplexen Zusammenspiels vieler Faktoren eher geringe Effekte.

Neben den genetischen Einflussgrößen sind auch Umweltfaktoren zu berücksichtigen. Auch bei Umweltvariablen besteht die Möglichkeit eines komplexen Zusammenspiels vieler Faktoren. Die Komplexität wird also noch erweitert, wenn man bedenkt, dass neben genetischen Faktoren auch die Umweltfaktoren und mögliche Interaktionen von Gen und Umwelt die Entwicklung von Erkrankungen beeinflussen können.

1.6 Auf der Suche nach Asthma-Genen

1.6.1 Auswahl bestimmter Gene: Überblick über Vorgehensweisen

Asthma ist eine komplexe Erkrankung, die durch verschiedene genetische Faktoren und Umwelteinflüsse determiniert wird. Seit langem ist anerkannt, dass es eine familiäre Häufung von Asthma gibt und dass eine positive Familiengeschichte von Asthma bei Eltern oder Geschwistern einen starken Risikofaktor für die Entwicklung von Asthma darstellt (Kabesch 2004). Diese klinischen Beobachtungen wurden durch Zwillingsstudien bestätigt, welche dabei eine wichtige Rolle spielten, den genetischen Anteil im Vergleich zum Umweltanteil auszumachen (Edfors-Lubs 1971). Bei Asthma wurde der genetische Anteil auf eine Größenordnung von 30-75% geschätzt (Feijen et al. 2000). Einen weiteren Eindruck in die genetische Komponente von Asthma lieferten Segregationsstudien. Demnach ist die Annahme eines Einflusses vieler Gene mit kleinen Effekten für komplexe Erkrankungen wahrscheinlich.

Die Suche nach krankheitsrelevanten Genen wird dadurch kompliziert, dass man für intermediäre Phänotypen auch unterschiedliche Vererbungsmuster annimmt. Zusammengefasst bleibt auch heute der Eindruck, dass Asthma und Atopie komplexe Phänotypen polygenetischen Ursprungs sind, die durch Umweltfaktoren beeinflusst werden (z.B. Feijen et al. 2000; Kabesch 2004). Entsprechend schwierig ist die Frage zu beantworten, mit welchem Ansatz die Suche nach krankheitsrelevanten Genen am besten fortgeführt werden sollte.

Heutzutage sind es vor allem Assoziationsstudien, von denen man sich die Aufdeckung von relevanten Effekten bei Genen erhofft (Risch, N. et al. 1996). Assoziationsstudien vergleichen erkrankte und gesunde Individuen einer Population. Sie untersuchen dabei die Frequenz von Polymorphismen bestimmter (sorgfältig ausgewählter) Gene und sind umso aussagekräftiger, je relevanter das untersuchte Gen für das betroffene Krankheitsbild ist.

Assoziationsstudien treten in unterschiedlichen Kontexten auf.

Werden Gene über das gesamte Genom auf Assoziationen hin untersucht, spricht man von einem **genomweiten Ansatz**. Genomweite Assoziationsstudien (im Folgenden auch als GWAs bezeichnet) sind abzugrenzen von genomweiten Kopplungsstudien.

Genomweite Kopplungsstudien zielen ebenfalls darauf ab, Krankheitsgene durch ein systematisches Absuchen des menschlichen Genoms zu identifizieren. Chromosomale Regionen, die für eine Krankheit von Bedeutung sein könnten, werden bei dieser Methode dadurch identifiziert, dass Marker, die mit der Krankheit assoziiert sind, häufiger in betroffenen Familienmitgliedern übertragen werden. Eine genauere Untersuchung dieser identifizierten Regionen be-

nutzt im Anschluss auch Assoziationsstudien und führt im besten Falle zur Identifizierung des Krankheitsgens.

Diese Methode für die Identifizierung krankheitsrelevanter Gene, bei welcher die näherungsweise Lokation auf einem Chromosom bei unbekannter Funktion ermittelt und dann eingengt wird, bezeichnet man als „**positional cloning**“.

Werden nur bestimmte Gene ausgewählt, spricht man vom **Kandidatengenansatz**.

Im Rahmen dieser Arbeiten sollen die drei verschiedenen Ansätze skizziert werden.

- Genomweite Assoziationsstudie (hypothesenfrei)
- „positional cloning“=genomweite Kopplungsstudie (hypothesenfrei) und Feinkartierung
- Kandidatengenansatz: Assoziationsstudie innerhalb eines Gens (hypothesengeleitet)

Die Identifizierung von Genen, die in die Genese einer Erkrankung involviert sind, ist auf Grund der verschiedenen Ansätze recht vielfältig und kompliziert geworden. Inzwischen sind etliche Gene identifiziert, die in unterschiedlichem Ausmaß an der Entstehung von Asthma beteiligt sein können. Zum Teil gehen sie aus genomweiten Ansätzen, zum Teil aus dem Kandidatengenansatz hervor.

1.6.2 Genomweiter Ansatz

1.6.2.1 Genomweite Assoziationsstudien

Den genomweiten Ansätzen gemein ist ein hypothesenfreies Vorgehen. Genomweite Assoziationsstudien untersuchen systematisch, ob sie einen möglichen Krankheitslocus finden, indem sie alle Variationen über das gesamte Genom auf eine Assoziation mit dem Phänotyp testen. Damit stehen GWAs in Kontrast zu den Kandidatengenstudien, bei welchen nur Variationen in Kandidatengen oder Kandidatenregionen analysiert werden, für die es *a priori* Hypothesen gibt. GWAs sind ein großer Fortschritt gegenüber genomweiten Kopplungsstudien, da sie auch kleinere Effekte aufzudecken vermögen und Loci in einer zuvor nicht erreichbaren Dichte über das gesamte Genom bestimmen können. GWAs nutzen in der Regel SNPs als genetische Marker, da sie leicht zu typisieren sind und umfangreich im menschlichen Genom vorkommen. Während die ursprünglichen genomweiten Kopplungsstudien 300-500 Marker über das Genom setzten, können nun mit modernen Technologien bis zu einer Million SNPs pro Individuum gleichzeitig gesetzt werden. Bereits 12 Millionen SNPs sind katalogisiert (2. National Center for Biotechnology Information et al. 2007), und es ist anzunehmen, dass sich in Zukunft Technologien entwickeln lassen, die auch mehrere Millionen SNPs gleichzeitig berücksichtigen können.

Obwohl es einen wachsenden Fokus auch für populationsbasierte GWAs oder GWAs mit Trios gibt, wurden die meisten GWAs bisher als Fall-Kontroll-Studien durchgeführt. Deren großer Vorteil ist die relativ leichte Durchführbarkeit und die Möglichkeit, sehr kleine Effekte aufzudecken. Abgesehen von der Auswahl der zu genotypisierenden Polymorphismen entspricht der Ablauf der GWA den Standards einer üblichen Assoziationsstudie. Eine typische GWA-Studie besteht aus vier Teilen:

1. Selektion einer hohen Anzahl von Individuen mit der Krankheit oder dem interessierenden Phänotyp und einer geeigneten Vergleichsgruppe
2. Auswahl eines Chips, der genomweite Marker setzt, Isolierung der DNA, Genotypisierung und Datenbetrachtung, um eine hohe Genotypisierungsqualität zu garantieren
3. Statistische Tests auf Assoziation der SNPs, welche die Qualitätskontrolle passiert haben, mit dem untersuchten Phänotyp
4. Replikation der gefundenen Ergebnisse in einer unabhängigen Population und experimentelles Prüfen funktioneller Implikationen der Ergebnisse

GWAs weisen im Kontext der Assoziationsstudien einige Besonderheiten auf, die vor allem durch die Vielzahl der Variablen begründet sind. Diese Besonderheiten sollen später verdeutlicht werden. Dazu werden die wesentlichen Merkmale einer Assoziationsstudie im Methodenteil ausführlich skizziert, bevor explizit auf Besonderheiten von GWAs eingegangen wird.

1.6.2.2 „Positional Cloning“

Beim „positional cloning“, das ebenfalls ohne eine *a priori* spezifizierte Hypothese startet, besteht der erste Schritt in der Identifizierung einer chromosomalen Region, die innerhalb von Familien zusammen mit einer Krankheit vererbt wird. Kopplungsanalysen machen sich dabei die gemeinsame Vererbung vorher ausgewählter Marker mit einer Krankheit innerhalb von Stammbäumen zunutze. Die Rekrutierung von Familien ist allerdings sehr aufwändig und die eher kleineren Stichproben bergen den Nachteil, kleine Effekte schwer aufzudecken. Wird eine Variante besonders häufig mit einer Krankheit weiter gegeben, geht man daher von einem bedeutenden Effekt aus und der vermutete Bezug der Region für die entsprechende Krankheit gewinnt an Bedeutung. Mittels der Entdeckung von Signalstärken, welche durch verschiedene Marker angezeigt werden, kann man Regionen nur bis zu einem gewissen Grad eingrenzen. Anschließend müssen solche Regionen weiter untersucht werden. Das „positional cloning“ lässt der Identifizierung einer relevanten Region eine Vielzahl von Techniken folgen, um das krankheitsrelevante Gen identifizieren und klonieren zu können (Ennis et al. 2001). Die Feinkartierung von Regionen, die durch Kopplungsstudien entdeckt wurden, kann

sehr mühsam sein, da diese Region zum Teil mehr als zehn Megabasen aufspannen und dadurch zahlreiche mögliche Gene enthalten. In der identifizierten Region werden in der Regel Assoziationsstudien durchgeführt. Die Assoziationsstudien können weiter eingegrenzt werden, wenn ein Gen innerhalb einer solchen Region für besonders relevant gehalten wird. Solche Relevanz kann ein Gen durch eine plausible biologische Hypothese liefern, warum es auf den untersuchten Phänotyp Einfluss nehmen kann.

1.6.2.2.1 ADAM33

Nicht nur ein Beispiel für „positional cloning“, sondern auch das erste Gen, welches durch „positional cloning“ identifiziert wurde, ist *ADAM33* („A Disintegrin And Metalloproteinase“) (Van Eerdewegh et al. 2002). *ADAM33* gehört zu einer Familie von 40 ADAM Proteinen (Holgate et al. 2006). Die ADAM-Familie ist eine Subgruppe einer Superfamilie von zinkabhängigen Metalloproteinasen (Stone et al. 1999; Primakoff et al. 2000). ADAMs bestehen in der Regel aus 750-800 Aminosäuren und sind gekennzeichnet durch mehrere Domänen (Abbildung 18).

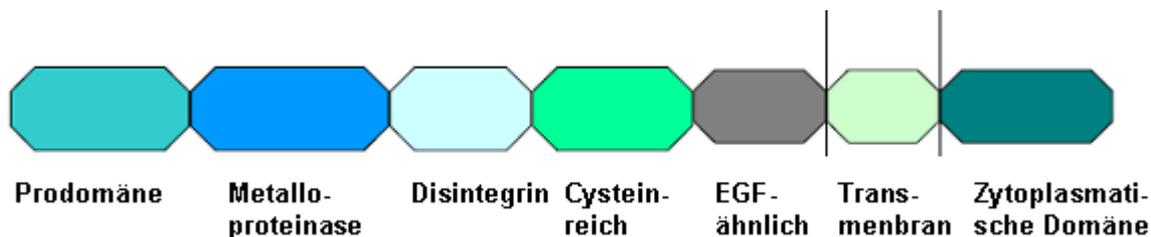


Abbildung 18: Schema der Domänen-Struktur der ADAMs.
Die senkrechten Linien deuten die Zellmembran an.

Das N-terminale extrazelluläre Ende wird von einer Pro-Domäne gebildet, auf welche die Metalloproteinase-Domäne folgt. Daran schließt sich eine Disintegrin-Domäne an, gefolgt von einer cysteinreichen Domäne, einer EGF-ähnlichen Domäne, sowie einer Transmembran-Domäne und einem zytoplasmatischen carboxyterminalen Anteil.

Das Gen *ADAM33* besteht aus 22 Exons, die für eine Signalsequenz sowie für die verschiedenen typischen ADAM-Domänen kodieren (Van Eerdewegh et al. 2002). Die verschiedenen Domänen vermitteln mehrere Funktionen von *ADAM33*: Aktivierung, Proteolyse, Adhäsion, Fusion und intrazelluläre Signalweiterleitung (Holgate et al. 2006). *ADAM33* wird auf Grund von Sequenzhomologien zu *ADAM12*, *13*, *15* und *19* zu einer Untergruppe der ADAM-Familie gezählt, deren Mitglieder vor allem proteolytische Aktivität aufweisen (Van Eerdewegh et al. 2002). Die proteolytische Eigenschaft der Metalloproteinase-Domäne von

ADAM33 wurde zudem in Versuchen mit *Drosophila* nachgewiesen (Garlisi et al. 2003; Zou et al. 2004).

Die von den Zellen sezernierten proteolytischen Enzyme bauen Bestandteile der extrazellulären Matrix ab. Dies ist zum Beispiel erforderlich, wenn Zellen die Basalmembran durchdringen müssen. Von essentieller Bedeutung für die Entwicklung und Homöostase eines Organismus sind auch die Wechselwirkungen zwischen Zellen und von Zellen mit der Matrix. Zellen müssen während der Embryogenese und Morphogenese wandern, adhären, miteinander interagieren und auch fusionieren (Wolfsberg et al. 1995).

Expriert wird *ADAM33* vor allem in der glatten Muskulatur des Bronchialsystems und in Lungenfibroblasten, nicht aber in Epithelzellen, T-Zellen oder Leukozyten (Powell et al. 2004). Der Nachweis von *ADAM33* in glatten Muskelzellen des embryonalen Bronchialsystems deutet darauf hin, dass *ADAM33* bei der Entwicklung von bronchialen, vaskulären und neuronalen Strukturen, deren Funktion sowie der Differenzierung von mesenchymalen Zellen wie Fibroblasten und glatten Muskelzellen in diesen Strukturen beteiligt ist. Analysen von Variationen in *ADAM33* legen nahe, dass es durch seinen Effekt auf die mesenchymale Zelldifferenzierung das „airway remodeling“ beeinflussen könnte (Warner et al. 2008). Außerdem spielt *ADAM33* möglicherweise eine Rolle bei der Entstehung der bronchialen Hyperreaktivität durch Proliferation der Fibroblasten und glatten Bronchialmuskulatur, Verdickung der Bronchialwände und somit Obstruktion der Atemwege, vielleicht ausgelöst durch eine veränderte Proteinaktivität (Schoen 2008). Inzwischen kann *ADAM33* als ein positionelles Kandidatengene aufgefasst werden, das durch „positional cloning“ identifiziert wurde.

1.6.3 Kandidatengenansatz

Ein Kandidatengene ist ein Gen, für welches es eine begründete Hypothese gibt, am Krankheitsgeschehen beteiligt zu sein. Zur Identifizierung solcher Gene arbeitet man mit dem sogenannten Kandidatengenansatz. Die Suche nach krankheitsrelevanten Loci mittels Kandidatengenansatz beinhaltet die systematische Suche nach genetischen Varianten innerhalb der Kandidatengene sowie Untersuchungen der einzelnen Varianten auf Assoziation mit relevanten Phänotypen. Der Kandidatengenansatz kann somit als hypothesengesteuerter Ansatz verstanden werden. Basis für den Kandidatengenansatz ist das Vorhandensein von detaillierten Informationen zu den entsprechenden Genen. Je nach vorliegender Information kann man Kandidatengene in verschiedene Kategorien einteilen, die jedoch ineinander übergehen können:

Bei **funktionellen** Kandidaten handelt es sich um Gene, deren Produkte in Beziehung zum untersuchten Phänotyp stehen. Diese Beziehungen sind meist experimentell gesichert (z.B. ein Enzym beeinflusst nachgewiesenermaßen einen Stoffwechselweg, der zur Ausprägung eines Merkmals beiträgt). Ein fundiertes Wissen dieser Funktionsweisen ist wichtig für eine erfolgreiche Selektion von Kandidatengen. Die Annahme einer spezifischen biologischen Funktion kann ausgeweitet werden (siehe Zhu et al. 2007). So können z.B. auch Expressionsmuster als Hinweise genutzt werden, Kandidatengene auszuwählen (Cheung et al. 2002; Duan et al. 2008).

Positionelle Kandidaten sind Gene, die auf Grund ihrer Lage im Genom mit dem Krankheitsgeschehen in Verbindung gebracht werden können. Solche Gene werden häufig im Rahmen von Studien ermittelt, welche die Region des Krankheitslocus zwar auf einen bestimmten Chromosomenabschnitt einengen, den ursächlichen Locus jedoch nicht ausmachen. Hierbei können Informationen aus Kopplungsstudien verwendet werden. Zur Funktion des Genproduktes müssen dabei keine Erkenntnisse vorliegen.

Bei **vergleichenden** Strategien werden Gene als Kandidaten gewählt, bei denen ein Effekt Arten übergreifend festgestellt wird. Solche Kandidatengene sind entweder funktionell konserviert oder strukturell homolog, wenn man sie in verschiedenen Spezies betrachtet.

Die Überprüfung, inwieweit ein Kandidatengen tatsächlich krankheitsrelevant ist, geschieht im Rahmen von Assoziationsstudien. Da man in der Genetik komplexer Erkrankungen von kleinen Effekten ausgeht, werden große Stichproben benötigt, um die vermuteten Effekte aufzudecken. Obwohl der Kandidatengenansatz nützlich ist, schnell über die spezifische Assoziation einer genetischen Variante mit dem Phänotyp zu entscheiden und in genetischen Assoziationsstudien vielfach zum Einsatz kommt (Tabor et al. 2002), bleibt das Problem, dass der Anteil ursächlicher Gene, welche Merkmale von biomedizinischer Bedeutung beeinflussen, noch klein ist. Damit ist auch die Auswahl an Kandidatengen begrenzt (Zhu et al. 2007).

Da die Pathomechanismen für Asthma vielfältig und nicht gänzlich verstanden sind, ist es nicht eindeutig, auf welche Mechanismen man sich beziehen sollte. Einige Ansatzpunkte, die gut etabliert sind, bieten Modelle veränderter T-Zell-Funktion, Expression bestimmter Zytokine oder Veränderungen in der Barrierefunktion der Lunge (Kabesch 2004).

1.6.3.1 Kandidatengene *Toll-like Rezeptoren (TLRs)*

Von den Toll-like Rezeptoren (TLRs), Proteinen, die bestimmte Pathogenmuster auf Mikroorganismen erkennen, wird angenommen, dass sie Einfluss auf das T_H1 - T_H2 Gleichgewicht, insbesondere durch eine Förderung der T_H1 -Antwort, ausüben. Da damit die Hypothese einer Verringerung des Risikos für die Entwicklung allergischer Erkrankungen einhergeht, geben die Gene, welche für die TLRs kodieren, ein Beispiel für eine Gruppe von Kandidatengenen und sollen genauer beschrieben werden.

Den Ausgangspunkt für die Untersuchung der Toll-like Rezeptoren liefert die immunologische Hygienehypothese, die 1989 das erste Mal in einem Artikel des britischen Wissenschaftlers David P. Strachan beschrieben wurde (Strachan 1989). Entsprechend der Hygienehypothese ist die Entwicklung von allergischen Erkrankungen durch bakterielle und virale Infektionen sowie durch die Exposition mit nichtinfektiösen Mikroben beeinflusst (von Mutius, E. 2007).

Pathogene wie sie die Mikroben (Bakterien, Pilze oder Viren) darstellen, weisen einige immer wiederkehrende Muster, die es dem Immunsystem erlauben, Gefahr zu erkennen. Solche wiederkehrenden Muster werden in der Immunologie als „pathogen associated molecular patterns“ (PAMPs) bezeichnet. Für die Erkennung solcher fremden Pathogene hat das Immunsystem Proteine bereits auf der Stufe des angeborenen Immunsystems entwickelt, die als Antwort auf Erkennung fremder Pathogene eine Entzündungsreaktion auslösen. Moleküle, welche PAMPs identifizieren und Immunantworten auslösen können, werden „pattern recognition receptors“ (PRRs) genannt.

Die am besten charakterisierte Gruppe der PRR sind die genannten Toll-like Rezeptoren (TLRs) (Janeway et al. 2002). TLRs nehmen eine zentrale Rolle in der Abwehr von Infektionen ein, erkennen PAMPs und lösen nach Bindung derselben die Aktivierung des angeborenen und schließlich des adaptiven Immunsystems aus (Medzhitov et al. 1998; Akira et al. 2001). Die Aktivierung von TLRs führt zur Ausschüttung von proinflammatorischen Zytokinen, welche zu einer stärkeren Bildung von T_H1 -typischen Zytokinen beitragen kann. T_H1 -Antworten sind wiederum mit einem Schutz vor Allergien assoziiert (Secrist et al. 1993). Man geht davon aus, dass Schlüsselzytokine des T_H2 -Systems wie IL-4, IL-5 und IL-13 gehemmt werden, während die Ausschüttung von IL-12 und IL-23, also typischer T_H1 -Zytokine, gefördert wird. Dadurch werden *TLRs* zu vielversprechenden Kandidatengenen für atopische Erkrankungen wie Asthma. Im Menschen sind bisher zehn TLR-Mitglieder identifiziert worden (Tabelle 1).

TLR	Ligand	Lokalisation	Referenz
TLR1 + TLR2	Triacyl-Lipopeptide	Zelloberfläche	(Takeuchi et al. 2001)
TLR2	Lipoteichonsäure	Zelloberfläche	(Ozinsky et al. 2000)
TLR2 + TLR6	Diacyllipopeptide	Zelloberfläche	(Takeuchi et al. 2001)
TLR3	Doppelstrangige RNA	Endosomale Kompartimente	(Alexopoulou et al. 2001)
TLR4	LPS	Zelloberfläche	(Poltorak et al. 1998)
TLR5	Flagellin	Zelloberfläche	(Hayashi et al. 2001)
TLR7 TLR8	Einzelstrangige RNA, Imidazoquinoline	Endosomale Kompartimente	(Heil et al. 2004)
TLR9	unmethylierte CpG-DNA	Endosomale Kompartimente	(Hemmi et al. 2000)
TLR10		unbekannt	

Tabelle 1: Überblick über die humanen Toll-like Rezeptoren und ihre Liganden.

TLR1, *TLR2*, *TLR4*, *TLR5*, and *TLR6* sind auf der Zelloberfläche exprimiert. Im Gegensatz dazu werden *TLR3*, *TLR7*, *TLR8* und *TLR9* endosomal exprimiert. *TLR3* erkennt doppelstrangige RNA, während *TLR7* und *TLR8* einzelstrangige RNA erkennen. *TLR9* erkennt nicht methylierte CpG-Dinucleotide in bakterieller DNA (Hughes et al. 2008).

Eine direkte Rolle der TLRs im Zusammenhang mit Erkrankungen zeigt sich an deren Erkennung von bakteriellen Substanzen. Bakterielle Toxine sind als Induktoren inflammatorischer Prozesse bekannt. Man unterscheidet dabei zwei Klassen: a) Toxine, die entweder durch einen aktiven Prozess von pathogenen Keimen nach außen abgegeben werden, und b) Toxine, die Bestandteil der bakteriellen Zellwand sind. Diese toxischen Zellwandkomponenten werden durch passive Prozesse, z.B. beim Absterben der Bakterien oder bei der Behandlung mit Antibiotika, in die Umgebung freigesetzt. In dieser zweiten Gruppe bakterieller Toxine finden sich Zellwandkomponenten wie das Lipopolysaccharid (LPS), auch Endotoxin genannt, Lipopeptide/-proteine (LP) und Peptidoglycan (PG), die wesentliche Bestandteile der Zellwand von Gram-negativen und/oder Gram-positiven Bakterien sind.

Lipoproteine der Zellmembran von Gram+/- Bakterien, Mykobakterien, Mykoplasmen und anderen Organismen induzieren intrazelluläre Signale in den Zellen des Immunsystems über Toll-like Rezeptor 2 (TLR2). Nach bisher vorliegenden Befunden ist TLR2 vor allem als Heterodimer zusammen mit TLR1 oder TLR6 wirksam. Die Rolle von TLR6 als Heterodimer mit TLR2 scheint wichtig für die Bindung von Lipoproteinen und Peptidoglykanen zu sein

(Nakao et al. 2005). TLR6 erkennt Diacyl-Lipopetide, während TLR1 Triacyl-Lipopetide erkennt (Krishnan et al. 2007).

Auch direkte Bindung von TLR2 wurde für Lipoproteine demonstriert (Vasselon et al. 2004). TLR4 reagiert auf Lipopolysaccharide (Guillot et al. 2004). TLR4 formt zusammen mit CD14 und anderen Kofaktoren auf der Oberfläche von Immunzellen einen Komplex, um Endotoxine zu erkennen und die intrazelluläre Signalübertragung zu aktivieren. Daneben erkennt TLR5 bakterielle Flagelline und Toll-like Rezeptor 9 (TLR9) ist für die Erkennung von viraler und intrazellulärer bakterieller DNA verantwortlich (Hemmi et al. 2000). Auf Grund dieser Zusammenhänge von Toll-like-Rezeptoren mit bakteriellen Toxinen und einer daraus resultierenden Einflussnahme auf atopische Erkrankungen sind sämtliche *TLRs* plausible Kandidatengene für Asthma.

1.7 Ableitung der Fragestellung

Drei methodische Annäherungen zur Identifizierung krankheitsrelevanter Gene wurden eingeführt: der Kandidatengenansatz, das „positional cloning“ und genomweite Assoziationsstudien.

Im Rahmen der vorliegenden Arbeit werden drei Studien behandelt, an denen ich maßgeblich beteiligt war (Schedel et al. 2006; Moffatt et al. 2007; Kormann et al. 2008).

Bezüglich der GWA war ich verantwortlich für die Integration der deutschen Teilstichproben, das heißt speziell deren Ziehung sowie die Auswertungen innerhalb der Replikationskohorten. Sämtliche statistischen Analysen der hier dargestellten Assoziationsstudien zu *ADAM33* sowie zu den Toll-like Rezeptoren wurden von mir durchgeführt, deren Ergebnisse interpretiert und die daraus abgeleitete Schlüsse in die Manuskripte eingearbeitet.

Jede Studie kann mit je einer der methodischen Annäherungen in Verbindung gebracht werden. Dazu gehören die erste durchgeführte GWA für Asthma als Vertreter des genomweiten Ansatzes, eine Assoziationsstudie für *ADAM33* als „positional cloned“ Gen sowie eine Kandidatengenstudie, welche die Rolle von *TLRs* bei der Entwicklung von Asthma untersucht. Die drei Studien sollen in dieser Reihenfolge vorgestellt und diskutiert werden.

Die Ausgangsfrage der GWA lautet:

- Gibt es Gene (Loci), die mit dem Phänotyp (Asthma) assoziiert sind?

Sollten sich spezifische Loci finden lassen, lässt sich als weitere Frage ableiten:

- Inwieweit erfahren die Ergebnisse der GWA zu Asthma Bestätigung? (insbesondere für das Gen oder die Loci, welche sich aus der GWA als neue Kandidaten für Asthma ergeben)

Daneben interessieren auch die Studien der anderen Ansätze:

- Welche Rolle spielt *ADAM33* bei der Entwicklung von Asthma im Kindesalter?
- Welche Rolle spielen die *TLRs* bei der Entwicklung von Asthma im Kindesalter?

Neben dem Bericht der Ergebnisse dieser drei Studien soll der Frage nachgegangen werden, welche Bedeutung die untersuchten Gene im Kontext des jeweiligen Ansatzes haben.

Abschließend sollen die methodischen Annäherungen noch einmal im Kontext der Ergebnisse diskutiert werden.

2 Methoden

2.1 Grundbegriffe Assoziationsstudien

Mittels Assoziationsstudien lassen sich genetische Varianten identifizieren, welche in die Entwicklung komplexer Krankheiten einbezogen sind (Editorial 1999).

Das Ziel jeder Assoziationsstudie ist die Identifizierung von genetischen Variationen (**SNPs**, **Haplotypen** oder **Kombinationen von Genotypen**), die zwischen Gesunden und Kranken variieren (Balding 2006).

Die Aufgabe der Identifizierung genetischer Determinanten einer Krankheit kann durch einiges kompliziert werden. Gelegentlich macht die Bestimmung der Genotypen im Labor Probleme, weshalb man Qualitätskontrollen sowie einen Test auf Abweichung vom **Hardy-Weinberg-Gleichgewicht** vornimmt. Im Rahmen verschiedener **Studiendesigns** geht es nach der Genotypisierung um die Aufdeckung einer Assoziation. Um solche Assoziationen zu finden, versucht man Tests mit hoher **Power** zu benutzen.

Selbst wenn die Qualität der Genotypisierung nachgewiesen gut ist und man im Anschluss eine Assoziation eines SNPs mit der Krankheit gefunden hat, so ist die Wertigkeit der gefundenen Assoziation nicht sofort klar. Man kann a) einen ursächlichen Locus gefunden haben oder b) einen Locus entdeckt haben, der in starker Beziehung (hohem **„linkage disequilibrium“** oder **LD**) mit dem wirklich ursächlichen Locus steht. Leider ist es auch möglich, dass c) die Assoziation durch Mischung von Populationen (**Populationstruktur oder Stratifikation**) begründet ist oder d) es sich um einen falsch positiven Befund handelt.

Falsch positive Befunde werden oft durch zuvor mangelnde Korrektur für **multiple Testen** begründet. Daher gehört eine Analyse, die für multiple Tests korrigiert, heute ebenso zum Standard wie **Replikationen** an unabhängigen Stichproben. Alle Begriffe sollen kurz eingeführt werden, wobei der Darstellung des „linkage disequilibrium“ auf Grund der Bedeutung in verschiedenen Kontexten etwas mehr Raum gegeben wird.

2.1.1 SNPs als genetische Variablen

Ihre Vielzahl und die Einfachheit, sie in automatisierter Form zu bestimmen, machen SNPs derzeit zum Gegenstand der Wahl für genetische Assoziationsstudien.

Wie in 1.5.1 erläutert, wird ein SNP aus zwei Allelen gebildet und jeder Mensch kann in seinen diploiden Zellen nur eine von drei Ausprägungen für einen SNP haben, den Wildtyp (zwei häufige Allele), den heterozygoten Typ (ein häufiges, ein seltenes Allel) oder den Polymorphismus (zwei seltene Allele).

Es gibt verschiedene Möglichkeiten der Analysen, die sich entsprechend ändern, je nachdem, wie diese drei Ausprägungen eines SNPs behandelt werden. Statistische Auswertungen, die den Zusammenhang eines einzelnen SNPs mit einem Phänotyp zum Gegenstand haben, werden als Einzel-SNP-Analysen bezeichnet und werden in 2.2.1 noch genauer erläutert.

Ein Ansatz der Einzel-SNP-Analyse geht auf die Annahme eines Mendelschen Erbgangs zurück. Ein anderer Ansatz ist es, eine Erkrankung zu untersuchen, je nachdem ob die untersuchten Personen das interessierende Allel entweder gar nicht, einmal oder zweimal haben. Und ein dritter Ansatz besteht darin, die Allelhäufigkeiten bei Fällen und Kontrollen zu vergleichen (Sasieni 1997).

Die wissenschaftliche Bedeutung von SNPs liegt in ihrem häufigen Auftreten und der hohen Variabilität. Meist ist allerdings recht wenig über die Funktionsweise eines bestimmten SNP bekannt. Die dichte Verbreitung über das Genom macht SNPs vor allem für die genomweiten Studien interessant (Palmer et al. 2005). SNPs treten durchschnittlich einmal etwa alle 200 Basenpaare auf und mehr als neun Millionen SNPs sind in öffentlichen Datenbanken katalogisiert. Jedoch haben viele SNPs eine sehr geringe Allelfrequenz. So bleiben noch etwa sechs Millionen SNPs über das gesamte Genom, welche eine Frequenz von mindestens 5% aufweisen. Diese werden als häufige Varianten bezeichnet.

2.1.2 Genotypkombinationen als genetische Variablen

Auch Kombinationen von SNPs lassen sich als Analyseeinheit verwenden. Dabei können die Informationen des kompletten Genotyps (also beider Allele) ebenso verwendet werden wie Kombinationen, welche nur ein Allel an einem Locus berücksichtigen.

Wird die Information des Genotyps herangezogen, so wird wieder die diploide Information weiter verarbeitet. In Abbildung 19 werden zwei Genotypen (A/T und G/C) kombiniert, wobei hier die Genotypkombination nur für ein Individuum dargestellt ist. Da in einer Population drei mögliche Genotypen existieren gibt es 3^n Kombinationen an Genotyppartitionierungen.

Zum Beispiel können bei einer Analyse zweier SNPs $3 \times 3 = 9$ verschiedene Kombinationen an Genotypen entstehen, bei drei SNPs bereits 27. Diese Kombinationen können auch wieder in unterschiedlicher Weise interpretiert werden. Häufig betrachtet man das gleichzeitige Auftreten zweier Polymorphismen als besonders risikoreich.

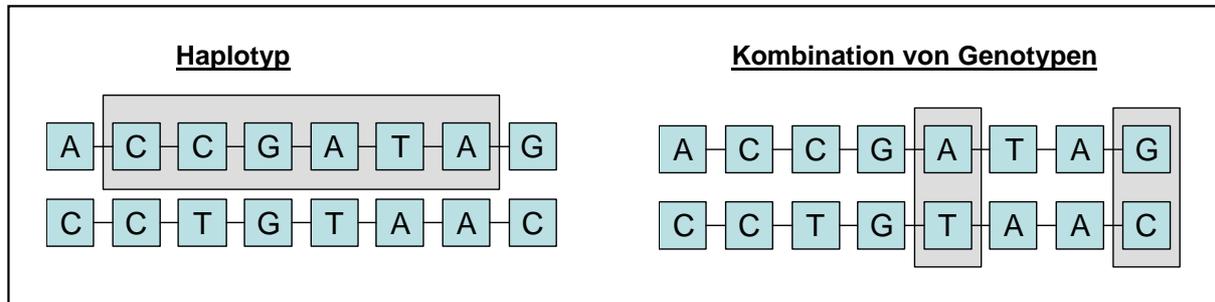


Abbildung 19: Unterschied zwischen Haplotyp und kombiniertem Genotyp

Die Anzahl der einzubeziehenden SNPs ist nicht begrenzt. Da die Anzahl der Kombinationen möglicher Genotypen jedoch schnell wächst, sind hier sinnvolle Partitionierungen zu überlegen (Nelson et al. 2001). Werden viele verschiedene SNPs gleichzeitig untersucht, wird häufig auch von Multilocus-Ansätzen gesprochen, da SNPs an unterschiedlichen Loci im Gen oder Genom liegen.

Spezielle Verfahren der Multilocus-Ansätze findet man auch unter dem Begriff der Interaktion (Vermeulen et al. 2007). Dabei handelt es sich um Analysemethoden, die einen gemeinsamen Effekt für mehrere SNPs berechnen. Zu unterscheiden sind additive Effekte, bei denen sich Effekte aufsummieren, von epistatischen Effekten, bei welchem der Effekt eines SNPs durch Variationen eines anderen SNPs modifiziert wird (Cordell 2002).

2.1.3 Haplotypen als genetische Variablen

Den Ausgangspunkt der Haplotypanalyse bildet ebenfalls die Idee, dass Polymorphismen eines Gens bevorzugt in spezifischen Kombinationen auf einem Chromosom vorkommen. Gegenstand sind hierbei jedoch Kombinationen von Allelen von einzelnen Nukleotiden auf dem gleichen Chromosomenstrang (Abbildung 19), nicht Kombinationen des gesamten Genotyps. Auf Grund der unabhängigen Vererbung verschiedener Chromosomen macht die Analyse von Haplotypen nur innerhalb eines Chromosomenstrangs Sinn. Während Multilocusansätze somit Informationen beobachtbarer Genotypen über das ganze Genom kombinieren können, beschränken sich Haplotypen auf Ketten von Allelen auf einem Chromosomenstrang. Sämtliche Verfahren setzen selbstverständlich eine korrekte Genotypisierung mit entsprechenden Labormethoden voraus.

Die folgenden Argumente, die im Kontext der Haplotypanalysen herangezogen werden, gelten ebenso für Multilocusansätze: Da die Nukleotide verschiedenen SNPs zugehören, wird es für möglich gehalten, dass Haplotypen mehr Information über die komplexe Beziehung zwischen DNA-Variation und Phänotyp liefern als jeder einzelne SNP (Stephens, J. C. et al. 2001). Dies unterstreicht die Relevanz, Kandidatengene auch auf Zusammenhänge mit Haplotypen zu untersuchen (Schaid et al. 2002). Folgende weitere Gründe sprechen für die Verwendung von Haplotypen (Clark 2004):

- Es gibt eine biologische Erklärung für den Haplotypeffekt: Die biologische Erklärung für Haplotypeffekte besteht darin, dass einige Mutationen in einem Gen mehrere Aminosäureänderungen im Proteinprodukt verursachen und der gemeinsame Effekt dieser Aminosäureänderungen einen viel größeren Einfluss auf die Funktion des Proteinprodukts haben kann als eine einzelne Aminosäureänderung (Schaid et al. 2002).
- Die statistische Power von Assoziationstests wird erhöht, wenn die Dimension der zu untersuchenden Einheiten reduziert wird. Die Verwendung von Haplotypen kann somit die Power bei Assoziationstests verbessern (Judson et al. 2000; Akey et al. 2001).

Haplotypen werden jedoch im Unterschied zu Multilocusgenotypen als Einheit vererbt und bilden daher eine genetisch noch besser interpretierbare Variante.

2.1.4 Hardy-Weinberg-Equilibrium

Das Hardy-Weinberg-Gesetz, 1908 unabhängig von dem britischen Mathematiker G.H. Hardy und dem deutschen Biologen W. Weinberg aufgestellt, beschreibt eine Regel über die Verteilung von Erbmerkmalen in großen Populationen, nach der die ursprünglichen Proportionen von Allelen in allen folgenden Generationen erhalten bleiben. Damit einher geht eine Konstanz der Genotypfrequenzen. Man spricht von Hardy-Weinberg-Gleichgewicht oder „Hardy-Weinberg Equilibrium“ (HWE).

Die Gültigkeit von HWE beruht auf zahlreichen Vorannahmen. Dazu gehören ein diploider Organismus, nicht überlappende Generationen, gleiche Allelfrequenz bei Frauen und Männern, was wichtig bei der Analyse X-chromosomaler Gene ist, zufällige Paarung sowie eine große, eigentlich unbegrenzte Population. Zudem sollte Migration vernachlässigbar sein, Mutationen ignoriert werden können und natürliche Selektion nicht die unter Betrachtung stehenden Allele betreffen. Außerdem muss der unter Betrachtung stehende Locus zwei Allele haben (Hartl 1997). Trotz des theoretischen Konstrukts der idealen Population lassen sich die Formeln durchaus mit Erfolg in der Praxis einsetzen. Die beiden Formeln für das Hardy-Weinberg Gleichgewicht lauten entsprechend Formel 1:

- $p^2 + 2pq + q^2 = 1$
- $p + q = 1$, wobei p die relative Häufigkeit des Auftretens des Allels A , q die Allelfrequenz des (zu A komplementären) Allels a

Formel 1: Hardy-Weinberg-Gleichgewicht

Mit Hilfe der Formeln des Hardy-Weinberg-Gesetzes lässt sich bei zufälliger Vererbung die Häufigkeit eines Allels in einer Population berechnen, wenn die Häufigkeiten der Genotypen bekannt sind bzw. die Häufigkeit eines Genotyps, wenn die Allelfrequenz bekannt ist. Für die Genotypfrequenzen ergibt sich somit: $p^2 = \text{Häufigkeit}(AA)$, $2pq = \text{Häufigkeit}(Aa)$, $q^2 = \text{Häufigkeit}(aa)$. Hieran ist auch zu erkennen, dass sich die Allelfrequenz auf den Anteil der Heterozygoten auswirkt (Abbildung 20).

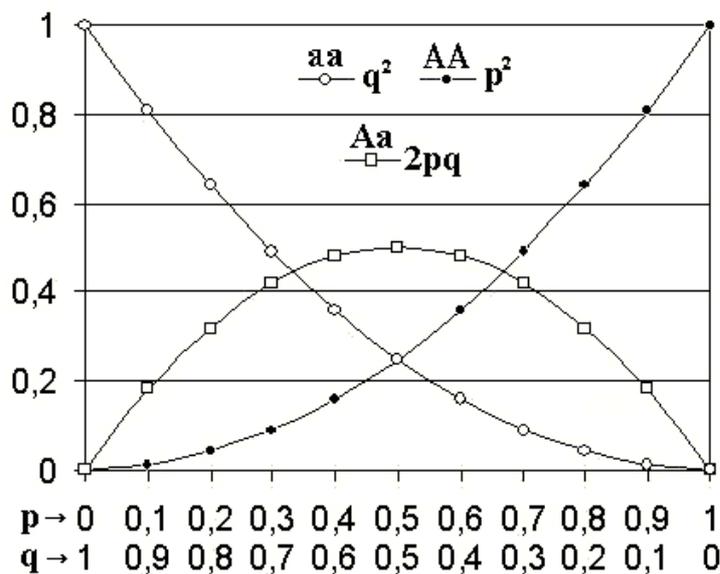


Abbildung 20: Hardy-Weinberg-Gleichgewicht für zwei Allele.

Die horizontale Achse zeigt die beiden Allelfrequenzen p und q , die vertikale Achse zeigt die Genotypfrequenzen. Die drei möglichen Genotypen sind durch unterschiedliche Zeichen dargestellt.

Häufig wird ein Test auf Abweichung vom HWE für die Überprüfung der Genotypisierungsqualität angewendet (Li, M. et al. 2008; Ryckman et al. 2008). In der Praxis wird bei sehr starken Abweichungen (z.B. $\chi^2 > 25$ bei Durchführung eines χ^2 -Tests mit der Nullhypothese keiner Abweichung vom HWE) von einem Problem bei der Genotypisierung ausgegangen. Umgekehrt beweisen die üblichen Chi-Quadrat-Tests, mit denen auf Abweichungen von Hardy-Weinberg getestet wird, die Gültigkeit des Hardy-Weinberg-Gleichgewichts auch dann nicht, wenn die Nullhypothese keiner Abweichung vom HWE beibehalten wird. Einen Beleg für die Gültigkeit im statistischen Sinne können hier nur Äquivalenztests liefern (Wellek 2004). Abweichungen vom HWE sind nicht auf Genotypisierungsfehler begrenzt, sondern können auch echt sein. Weichen die Genotypfrequenzen von denen ab, die man unter Gültig-

keit von HWE erwarten würde, kann dies auch auf Faktoren der Evolution zurückzuführen sein.

2.1.5 Power und Studiendesign

Das Ziel jeder Assoziationsstudie besteht darin, Effekte zwischen einer Krankheit und einer genetischen Variation zu finden. Die **Power** oder Teststärke gibt an, mit welcher Wahrscheinlichkeit ein statistisches Verfahren zugunsten einer spezifischen Alternativhypothese (z.B. „Es gibt eine Assoziation“) entscheidet, falls diese richtig ist. Gemeint ist die Wahrscheinlichkeit einen vorhandenen Effekt auch wirklich aufdecken zu können. Oftmals wird gerade bei nicht detektierten Effekten vermutet, dass ein statistischer Test ungenügende Power hatte, was zum Beispiel durch zu kleine Stichproben bedingt sein kann. Varianten im **Studiendesign** können sich somit auf die Power auswirken.

Die einfachste Form eines Studiendesigns für eine Assoziationsstudie ist die Fall-Kontroll-Studie, die auch in der klassischen Epidemiologie weit verbreitete Anwendung findet. Durch den Vergleich der Häufigkeit bestimmter Allele bei Patienten und gesunden Kontrollpersonen lassen sich die relativen Erkrankungsrisiken abschätzen, die von genetischen Determinanten ausgehen können. Dabei werden Fälle je nachdem rekrutiert, welchen Phänotyp die Studie untersucht. So würde man bei einer Studie, die Asthma zum Gegenstand hat, erst Fälle von Asthmatikern sammeln und diesen eine geeignete Kontrollgruppe von Personen ohne Asthma gegenüberstellen. Von einer eingebetteten Fall-Kontroll-Studie oder einem „nested-case-control“-Studiendesign spricht man, wenn Schätzungen aus einer Substichprobe einer etablierten Kohorte generiert werden, die den Schätzungen ähnlich sind, welche man bei Verwendung der gesamten Kohorte bekommen würde (Xue et al. 2009). Dabei werden zu jedem Fall Kontrollen ergänzt, die aus der gleichen Kohorte kommen, aber den entsprechenden Phänotyp der Fälle nicht tragen.

Ergebnisse aus Studien dieses Typs werden als Verhältnisse angegeben. Das Verhältnis, das in Odds Ratio angegeben wird, ist ein Maß dafür, um wie viel größer in der Gruppe mit Risikofaktor die Chance zu erkranken ist, wenn man sie mit der Gruppe ohne Risikofaktor vergleicht. Dabei können genetische Varianten die Risikofaktoren darstellen.

Ebenso wie Querschnittsstudien werden Fall-Kontroll-Studien einmalig durchgeführt. Querschnittsstudien untersuchen das Verhältnis zwischen einer Krankheit und anderen Expositionen, so wie sie in einer definierten Population zu einem bestimmten Zeitpunkt auftreten. Sie bieten einen besseren Ansatzpunkt zur gleichzeitigen Untersuchung verschiedener Phänotypen, wenn sie populationsbasiert eine geeignete repräsentative Bevölkerung zugrunde legen.

Im Gegensatz zu Fall-Kontrollstudien sind dort nämlich keine Verzerrungen der Prävalenzen anzunehmen.

Da nahezu sämtliche Auswertungen der vorliegenden Arbeit auf Fall-Kontroll- oder Querschnittsstudien basieren, sei an dieser Stelle nur auf weitere Studiendesigns wie Längsschnittsuntersuchungen oder Kohortenstudien hingewiesen. Studiendesigns der klinischen Forschung (Hartung et al. 2009) oder der genetischen Epidemiologie (Boddeker et al. 2000) sind überblicksartig an anderer Stelle zusammengefasst.

2.1.6 Stratifikation

Ungeachtet des Designs oder der Power werden immer wieder Bedenken geäußert, dass genetische Unterschiede zwischen ethnischen oder geographischen Schichten in der zu untersuchenden Population zu einer Verfälschung der Risikoschätzungen führen könnten. Wenn Patienten und Kontrollpersonen hierbei systematisch aus unterschiedlichen Schichten rekrutiert würden, könnte dies in der Tat zu genetischen Unterschieden zwischen den Stichproben führen, die nicht mit der Krankheit im Zusammenhang stünden, sondern durch deren Heterogenität begründet ist. Die Möglichkeit von Populationssubstrukturen wird auch als Stratifikation bezeichnet. Sobald die gemischten Subpopulationen sich sehr stark in ihrer Genetik unterscheiden, kann dies die Ergebnisse einer Assoziationsstudie verzerren. Daher wird man z.B. kaukasische nicht mit nicht-kaukasischen Populationen mischen.

Die Höhe des Einflusses von Populationsstratifikation ist zwar nicht wirklich geklärt (Thomas et al. 2002; Wacholder et al. 2002), es gilt aber vorsichtig zu sein, da Stratifikation im ungünstigen Fall zu falsch positiven Assoziationen führen kann. Geeignete Designs für familienbasierte Assoziationsstudien umgehen das Problem der Stratifikation. Dies geschieht jedoch auf Kosten einer reduzierten Power. Die Stratifikation kann aber auch in Fall-Kontrollstudien durch verschiedene Strategien ausgeglichen werden (Cordell et al. 2005):

- Matchen für Variablen, welche auf die Zielgröße Einfluss nehmen könnten
- Suche nach informativen Markern und deren Einbeziehung als Kovariate
- genomische Kontrolle

Der Einfluss von Stratifikation auf den Alpha-Fehler kann somit gering bleiben, sobald die Fälle und Kontrollen gut gematcht sind (Zang et al. 2007). Ansonsten wird Stratifikation durch spezielle Methoden ausgeglichen. Dabei setzt die genomische Kontrolle Verfahren ein, die für genetische Heterogenität korrigieren (Devlin et al. 1999; Devlin et al. 2004).

2.1.7 Kopplungsungleichgewicht oder „linkage disequilibrium“

Das Kopplungsungleichgewicht oder „linkage disequilibrium“ (LD) ist definiert als die statistische Assoziation zwischen Sequenzvarianten verschiedener Positionen entlang eines Chromosoms (Weiss et al. 2002). LD wird auch als allelische Assoziation bezeichnet, da es einen Zusammenhang von Allelen beschreibt. Gemessen wird also ein Zusammenhang zwischen genetischen Varianten, zum Beispiel zwischen SNPs. Zuerst muss diese Größe quantifizierbar gemacht werden (siehe 2.1.7.1), um sie später in Assoziationsstudien nutzen zu können.

Einerseits kann man sich im Vorfeld einer Studie innerhalb eines Sets von SNPs, die hoch korreliert sind, zugunsten der Genotypisierung nur eines einzigen SNPs entscheiden. Dieser SNP, der mit allen anderen SNPs dieses Sets korreliert ist, kann dann als „tagging SNP“ stellvertretend für das ganze Set typisiert werden (siehe 2.1.7.2). So können enorme Kosten eingespart werden.

Im Anschluss an Berechnungen von Assoziationen kann andererseits darauf geschlossen werden, welche Varianten mit einem assoziierten SNP hoch korreliert sind (Kruglyak 1999; Risch, N. J. 2000). Dadurch lassen sich SNPs finden, welche die gleiche oder eine ähnliche Assoziation aufweisen wie ein untersuchter SNP, möglicherweise aber eine stärkere, direkte funktionelle Bedeutung haben.

2.1.7.1 LD-Koeffizienten

Um Rückschlüsse auf Zusammenhänge zwischen SNPs ziehen zu können, muss das LD messbar gemacht werden. Wünschenswerte Eigenschaften für ein solches LD-Maß sind dessen gute Interpretierbarkeit und praktischer Nutzen. Um diesen Ansprüchen zu genügen, wurden verschiedene Maße entwickelt. Ein Maß ist Lewontins Disequilibriumskoeffizient D_{ij} . Dieses Maß misst im Falle paarweiser biallelischer Loci die Differenz des beobachteten gemeinsamen Auftretens zweier Allele und der erwarteten Häufigkeit des gemeinsamen Auftretens dieser Allele unter der Annahme von Equilibrium (berechnet als Produkt der Allelfrequenzen). Diese Formel sei an Tabelle 2 erklärt.

		Locus B		
		B ₁	B ₂	Gesamt
Locus A	A ₁	P ₁₁	P ₁₂	p ₁
	A ₂	P ₂₁	P ₂₂	p ₂
Gesamt		q ₁	q ₂	1

Tabelle 2: Beispiel zur Berechnung von „linkage disequilibrium“

Seien A und B zwei Loci mit je zwei Allelen (jeweils A_1 und A_2 an Locus A oder B_1 und B_2 an Locus B), die Frequenzen der Allele an den jeweiligen Loci p_1 , p_2 , q_1 und q_2 , und die Frequenzen der Haplotypen A_1B_1 , A_1B_2 , A_2B_1 und A_2B_2 jeweils P_{11} , P_{12} , P_{21} und P_{22} , so gilt Equilibrium, wenn $P_{11} = p_1 \times q_1$ (oder $P_{12} = p_1 \times q_2$ oder $P_{21} = p_2 \times q_1$ oder $P_{22} = p_2 \times q_2$). Disequilibrium gilt, wenn die Produkte der Allelfrequenzen von den Häufigkeiten der Haplotypen abweichen. Entsprechend ergibt sich die Höhe des LD-Koeffizienten nach Lewontin durch Formel 2.

$$D = P_{11} - p_1 \times q_1 = P_{12} - p_1 \times q_2 = P_{21} - p_2 \times q_1 = P_{22} - p_2 \times q_2.$$

Formel 2: Lewontins Disequilibriumskoeffizient

Ob D positiv oder negativ ausfällt, ist durch das beliebige Benennen der Allele bedingt. Für jede beschriebene Allelfrequenz ist der kleinste mögliche Wert D_{\min} und der größte mögliche Wert D_{\max} von D wie folgt definiert.

$$D_{\min} \text{ ist der größere von } -p_1 \times q_1 \text{ und } -p_2 \times q_2.$$

$$D_{\max} \text{ ist der kleinere von } p_1 \times q_2 \text{ und } p_2 \times q_1.$$

Dieses Maß ist jedoch abhängig von der Allelfrequenz und schwierig zu interpretieren. Daher wurden zahlreiche Anstrengungen unternommen, ein LD-Maß weniger allelabhängig und besser interpretierbar zu gestalten. Von diesen haben sich D' ¹ (Formel 3) und R^2 (Formel 4) als die in der Literatur am häufigsten verwendeten durchgesetzt. Ein Wert von 1 steht für komplettes LD, während 0 die Abwesenheit von LD bedeutet.

$$D' = \frac{D}{D_{\max}}$$

Formel 3: LD-Koeffizient D'

$$R^2 = \frac{(P_{12} - p_1 p_2)^2}{p_1 p_2 q_1 q_2}$$

Formel 4: LD-Koeffizient R^2

¹ Wegen der Zweifelhaftigkeit des Vorzeichens wird häufig auch der Betrag von D' angegeben.

Einige andere Maße, die auf Lewontins Disequilibriumskoeffizienten D basieren, sind entwickelt worden (Du et al. 2007). Diese Maße, die in Übersichtsarbeiten zusammengestellt sind (Devlin et al. 1995; Jorde 2000; Pritchard et al. 2001), spielen in der Praxis aber eine weniger starke Rolle. Im Rahmen unserer Arbeiten wird später ausschließlich auf den Koeffizienten R^2 zurückgegriffen. An dieser Stelle soll jedoch auch D' betrachtet werden, da sich dieses Maß sehr häufig in Studien finden lässt. Beide Maße (D' und R^2) liegen zwischen 0 und 1 und sind damit gut interpretierbar: je näher der Koeffizient an 1 liegt, desto höher ist das LD. Wird der Koeffizient 0, besteht keinerlei Zusammenhang zwischen zwei Loci.

D' und R^2 haben noch weitere Vorteile: D' wird genau dann 1, wenn mindestens ein Allel des einen mit einem Allel des anderen Locus verknüpft ist. Somit bleibt D' nach einer Neumutation auf dem maximalen Wert 1, bis ein Rekombinationsereignis den originären Haplotyp durchbricht. Jedoch sind Werte $D' < 1$ schwieriger zu interpretieren und in der praktischen Anwendbarkeit bei der Vergleichbarkeit von Markerpaaren ist D' begrenzt, da es zu häufig den Wert 1 annimmt (Ardlie et al. 2002; Du et al. 2007).

R^2 repräsentiert die statistische Korrelation und ist das gebräuchlichste Maß. Es lässt sich damit wie ein Korrelationskoeffizient interpretieren. Neben der Nähe zum Korrelationskoeffizienten bietet dieses Maß noch weitere Vorteile (Jorde 2000; Pritchard et al. 2001; Weiss et al. 2002). So lässt sich eine Beziehung zwischen prädisponierenden Krankheitsloci, SNPs in LD und einer geforderten Stichprobengröße herstellen, die vor allem bei Powerberechnungen eine Rolle spielen kann. Geht man davon aus, nicht den kausativen Locus (SNP 1), sondern einen SNP 2 in LD mit dem kausativen Locus zu untersuchen, muss die Stichprobe um $1/R^2$ erhöht werden, um die Assoziation, die man für SNP 1 mit einem Phänotyp vermutet, auch für SNP 2 zu entdecken (Pritchard et al. 2001; Carlson et al. 2003).

Wie bei Effektmaßen üblich sind die statistische Signifikanz, die stets von der Stichprobengröße abhängt, und die Höhe des LD streng zu trennen (Weiss et al. 2002). Die statistische Signifikanz des LD kann mittels Tests geprüft werden, die für Kontingenztafeln üblich sind. Dies sind typischerweise Fishers exakter Test oder ein χ^2 -Test.

Häufig wird ein hohes Kopplungsungleichgewicht zwischen SNPs auf deren gekoppelte Vererbung zurückgeführt, obgleich dies nicht den einzigen Grund für ein hohes LD liefert. Beispielsweise kann auch Stratifikation hohes LD bedingen. In folgendem Lehrbuchbeispiel (Hartl 1997) hat Subpopulation 1 eine Allelfrequenz von 0,05 für A_1 und B_1 , während Subpopulation 2 eine Allelfrequenz von 0,95 für A_1 und B_1 hat. In Tabelle 3 ist zu erkennen, dass in der gemischten Population LD besteht, obwohl dies in den Subgruppen nicht vorliegt.

	Häufigkeiten	Subpopulation 1	Subpopulation 2	Gemischte Population
A_1B_1	P_{11}	0,0025	0,9025	0,4525
A_1B_2	P_{12}	0,0475	0,0475	0,0475
A_2B_1	P_{21}	0,0475	0,0475	0,0475
A_2B_2	P_{22}	0,9025	0,0025	0,4525
	D	0	0	0,2025
	Dmin	-0,0025	-0,0025	-0,2500
	Dmax	0,0475	0,0475	0,2500
	D'	0	0	0,81

Tabelle 3: Beispiel für LD auf Grund von Populationsstruktur

Die Folgerung, dass hohes LD zwischen SNPs gleichzusetzen ist mit hoher Kopplung, ist demzufolge ungenau. Kopplung ist zwar meist hinreichend, aber nicht notwendig für Kopplungsungleichgewicht.

2.1.7.2 LD-Blöcke und „tagging SNPs“

Durch die Bestimmung von SNPs, die in einem Set von SNPs die meiste genetische Information beinhalten, sogenannter „**tagging SNPs**“, können die Anzahl zu genotypisierender SNPs und damit die Kosten einer Genotypisierung stark reduziert werden.

Wie ein hohes Maß an genetischer Information aufzufassen ist, wird unten genauer erläutert. Die Idee dabei ist, dass wenige SNPs den Großteil der Information widerspiegeln, die man bei Betrachtung aller SNPs erhalten würde. Dies gelingt für ein bestimmtes Set von SNPs am besten.

Eine Beschränkung auf „tagging SNPs“ kann für Assoziationsstudien auch eine Erhöhung der Power beinhalten, da man später weniger Tests durchführt. Bei der Identifizierung von „tagging SNPs“ haben sich verschiedene Vorgehensweisen etabliert. Man unterscheidet zwei Vorgehensweisen, die LD basierte und Haplotyp basierte Selektion.

Bei LD basierten Ansätzen werden sogenannte „bins“ dadurch definiert, dass alle paarweisen LD-Koeffizienten innerhalb eines „bins“ ein gewisses LD-Maß nicht unterschreiten (Cousin et al. 2003; Carlson et al. 2004). Ausgewählt wird dann ein SNP, welcher den „bin“ im Hinblick darauf am besten repräsentiert, dass er mit allen SNPs des „bins“ hohes LD aufweist. Dieser häufig verwendete Selektionsmechanismus verwendet den Begriff „bin“ für eine Gruppe von SNPs in hohem LD (Carlson et al. 2004) und ist einer der gebräuchlichsten Algorithmen, deren Anwendung z.B. mittels Haploview (Barrett et al. 2005) frei zugänglich ist. Die Schwelle für hohes LD kann beliebig variiert werden, wobei ein R^2 von 0,8 ein gängiges

Maß ist. Auf diesen Ansatz zur Bestimmung von „tagging SNPs“ wurde im Rahmen unserer Arbeiten zurück gegriffen. Der Begriff „bin“ wird in manchen Publikationen durch den Begriff „Block“ ersetzt, was durch die Mehrfachbelegung des Begriffs „Block“ missverständlich sein kann.

Der genaue Algorithmus zur Bestimmung von „tagging SNPs“ für „bins“ gestaltet sich wie folgt: Es wird eine Schwelle für R^2 vorgegeben. Der SNP, welcher diesen Schwellenwert mit den meisten anderen SNPs überschreitet, wird identifiziert und als maximal informativer Locus bezeichnet. Dieser Locus wird mit allen SNPs, mit denen er assoziiert war, zu einem „bin“ zusammengefasst. Weil ein „bin“ zunächst nur auf Basis eines einzigen SNPs bestimmt wurde, werden dann alle LD-Maße innerhalb des „bins“ neu berechnet. Jeder SNP, der mit allen anderen SNPs innerhalb eines „bins“ die vorgegebene Schwelle überschreitet, kann dann als „tagging SNP“ gewählt werden. Welchen der SNPs man innerhalb eines „bins“ als „tagging SNP“ verwendet, kann z.B. vom genomischen Kontext (kodierend oder nicht kodierender SNP) oder der Durchführbarkeit der Genotypisierung abhängen. Dieser Prozess der Bestimmung von „bins“ wird für die verbleibenden SNPs solange iterativ fortgesetzt, bis alle Loci einem „bin“ zugeordnet sind. Am Ende erhält man ein Set von „tagging SNPs“, das stellvertretend für die Gesamtheit der betrachteten SNPs steht.

Bei haplotypbasierten Ansätzen zur Bestimmung von „tagging SNPs“ folgt man einem Ansatz, der eigentlich unabhängig von LD-Maßen ist. Dieser Ansatz spielt jedoch eine Rolle, da aus diesem Vorgehen der Begriff des „tagging SNPs“ oder „haplotype tagging SNPs“ (htSNPs) hervorging. Haplotypbasierte Ansätze wählen das kleinste Set von SNPs, mit dem es gelingt, alle gängigen Haplotypen innerhalb eines chromosomalen Abschnitts zu unterscheiden. Das nicht mehr redundante Set von „haplotype tagging SNPs“ kann eine Vielzahl von Haplotypen unterscheiden (Thompson et al. 2003) und deckt in der Regel den Hauptanteil der Variation möglicher Haplotypen ab. Ein Beispiel soll dies verdeutlichen.

	SNP 1 (A/C)	SNP2 (A/C)	SNP 3 (G/T)	SNP 4 (G/T)
Haplotyp 1	A	A	T	G
Haplotyp 2	A	C	G	T
Haplotyp 3	C	A	G	G

Tabelle 4: Beispiel zur Bestimmung von „haplotype tagging SNPs“

In Tabelle 4 genügt die Genotypisierung der ersten beiden SNPs, um auf die Ausprägungen von SNP 3 und SNP 4 schließen zu können. Da nur diese Haplotypen vorliegen, weiß man, dass der Haplotyp A-A-T-G sein muss, wenn die ersten beiden Allele A und A sind. Die Ty-

pisierung dieser zwei „haplotype tagging SNPs“ (SNP1 und SNP2) genügt somit, auf den gesamten Haplotyp rückzuschließen zu können. Verschiedene Annäherungen wurden vorgeschlagen, um „htSNPs“ zu bestimmen, die zwischen weniger und stark formalisierten Methoden variieren.

2.1.7.3 LD bei der Kartierung

LD-Maße wurden auch verwendet, um spezifische LD-Muster durch das ganze Genom zu finden (Daly et al. 2001; Gabriel et al. 2002).

Als Blöcke bezeichnet man hier chromosomale Abschnitte mit hohem Kopplungsungleichgewicht und geringer Haplotyp-Diversität. Diese Blöcke sind nicht gleichzusetzen mit den oben eingeführten „bins“, sondern bezeichnen hier zusammenhängende Segmente innerhalb des Genoms. Entsprechend werden auch andere Algorithmen verwendet, um diese Segmente zu definieren. Beispielsweise definiert Gabriel solch einen Block als eine Region, in der die große Mehrheit der informativen Marker (95%) ein hohes paarweises LD aufweist (Gabriel et al. 2002). Entsprechend liefert nur ein geringer Anteil von Markern ein kleines LD-Maß. In diesen Blocksegmenten gibt es somit wenig Evidenz für historische Rekombination. Zudem werden innerhalb dieser Segmente nur wenige häufige Haplotypen beobachtet.

Bei der genomweiten Kartierung wird das Genom dann als eine Serie von Blöcken beschrieben, welche durch hohes LD charakterisiert sind und die durch kleinere Segmente mit sehr geringem LD getrennt sind. Die kleineren Segmente werden typischerweise auf Rekombinationshotspots zurückgeführt (Daly et al. 2001; Patil et al. 2001; Gabriel et al. 2002; Cardon et al. 2003). „Hot spots“ werden Loci genannt, an denen eine besonders hohe Rekombinationsrate vorliegt. Auch hier werden die Blöcke nicht ausschließlich durch Methoden definiert, die paarweise LD-Maße benutzen (Gabriel et al. 2002; Phillips et al. 2003; Stumpf et al. 2003), sondern können auch durch Methoden, die sich auf häufige Haplotypen beziehen, charakterisiert werden (Daly et al. 2001; Patil et al. 2001; Zhang, K. et al. 2002).

LD ist von fundamentaler Bedeutung bei der Genkartierung. Es kann nicht nur bei der genomweiten Kartierung dienlich sein. Auch im Rahmen des „positional cloning“ kann es verwendet werden, durch Feinkartierung die exakte Region aufzuspüren, welche das Kopplungssignal verursacht hat (z.B. (Hastbacka et al. 1992)). Darüber hinaus können LD-Muster dazu dienen, die Verteilung von „crossing over“-Ereignissen zu schätzen und evolutionäre Ereignisse zu entschlüsseln (Wall et al. 2003).

2.1.8 Multiples Testen und Replikation

Das Problem des multiplen Testens besteht darin, bei Durchführen zahlreicher Tests eine erhöhte Wahrscheinlichkeit zu haben, eine Assoziation nur als Zufallsbefund ausfindig zu machen. Umso mehr Tests durchgeführt werden, desto stärker rückt dieses Problem in den Vordergrund, da mit steigender Zahl von Tests auch die Wahrscheinlichkeit steigt, zufällig ein signifikantes Ergebnis zu bekommen. Es wurden daher verschiedene Korrekturverfahren entwickelt, die für die Anzahl durchgeführter Tests korrigieren. Traditionelle Methoden kontrollieren die Wahrscheinlichkeit, eine falsche Entscheidung zu treffen, was unter Statistikern auch als die Kontrolle der familienweisen Fehlerrate bezeichnet wird. Die Kontrolle dieses Fehlers erfordert, dass zu jedem von n durchgeführten Tests das Fehlerniveau korrigiert werden muss. Am prominentesten ist die Bonferroni-Korrektur, welche das Signifikanzniveau einfach auf α/n herabsetzt, wobei α für das gewählte Signifikanzniveau und n für die Anzahl der durchgeführten Tests steht. Die Bonferroni-Korrektur gilt als recht konservativ in dem Sinne, dass sich Assoziationen nur schwer aufdecken lassen. Obgleich Prozeduren mit mehr Power entwickelt wurden, ist den Verfahren, welche die familienweise Fehlerrate kontrollieren, gemein, dass sie bei großen Familien von untersuchten Endpunkten einen hohen Verlust an Power aufweisen.

Neben der Bonferroni-Korrektur hat sich vor allem die **„false discovery rate“ (FDR)** etabliert (Hochberg et al. 1990; Benjamini et al. 1995; Sabatti et al. 2003). Die FDR betrachtet die Kontrolle des Fehlers bei multiplen Tests von einem anderen Standpunkt. Die FDR ist der erwartete Anteil von falschen Entdeckungen unter allen Entdeckungen. Zwei mögliche Prozeduren, die FDR zu kontrollieren, wurden von Benjamini eingeführt (Benjamini et al. 1995). Die FDR hat den großen Vorteil, nicht so konservativ zu sein und wird als geeignete Balance zwischen dem Vernachlässigen der Korrektur für multiples Testen und einer zu strengen Korrektur betrachtet.

Im Unterschied zur Berechnung eines korrigierten Signifikanzniveaus wird auch die Wiederholung von Ergebnissen als Bestätigung eines Resultats aufgefasst. Daher gelten Replikationsstudien als wünschenswert. Einen Beleg ähnlich der Replikation liefern auch Homogenitätsanalysen, die einen Zusammenhang für verschiedene Subgruppen einer großen Population nachweisen können. Dies können zum Beispiel unterschiedliche Studienzentren sein.

2.1.9 Interaktion

Es ist allgemein akzeptiert, dass genetische und Umweltfaktoren zur Entwicklung komplexer Erkrankungen beitragen. Daher ist die Gen-Umwelt-Interaktion ein heiß diskutiertes Thema

in der Humangenetik und es bestehen große Erwartungen für potentielle Anwendungen (Dempfle et al. 2008).

Interaktion ist in der genetischen Epidemiologie jedoch ein schwer greifbarer und nicht eindeutig definierter Begriff. Ganz allgemein gesprochen wird Interaktion oft so verstanden, dass sich zwei Gegenstände in irgendeiner Weise beeinflussen. SNP-SNP-Interaktion wird in der genetischen Epidemiologie gelegentlich auch als „epistasis“ bezeichnet. Dies bedeutet im ursprünglichen Sinne, dass SNPs nicht unabhängig voneinander mit einem Merkmal assoziiert sind. Bisläng gibt es keinen Konsens, wie Interaktion im Rahmen der genetischen Epidemiologie zu modellieren ist.

Insbesondere gibt es kein Standardverfahren, den Einfluss von zwei oder mehreren SNPs auf einen Phänotyp zu analysieren. Bisherige Vorschläge der genetischen Epidemiologie folgen vielmehr ganz unterschiedlichen statistischen Methoden, die darauf ausgerichtet sind, Sets von SNP zu identifizieren oder Sets von SNPs darauf hin zu bewerten, ob deren Kombination einen signifikanten Beitrag zur Vorhersage einer Krankheit liefert. In einem Übersichtsartikel von Musani (Musani et al. 2007) zur Gen-Gen-Interaktion werden verschiedene Verfahren zur Berechnung von Interaktion vorgestellt. Ein Versuch, diese Verfahren zu klassifizieren unterscheidet dabei parametrische und nicht-parametrische Varianten zur Interaktionsberechnung. Ein klassisches Verfahren zur Berechnung von Interaktionen liefert die logistische Regression, bei welcher Interaktion als Produktterm zweier Variablen in ein Modell, das eine Krankheit prognostizieren möchte, aufgenommen wird.

2.2 Statistische Methoden zur Auswertung von Assoziationsstudien

2.2.1 Zusammenhänge zwischen einem SNP und qualitativen Phänotypen

Der SNP besteht aus drei möglichen Varianten von Allelpaaaren. Dadurch kann die Kodierung des SNPs in Form einer Variablen mit drei Kategorien vorgenommen werden. Die drei Kategorien stellen die drei Ausprägungen des Genotyps dar (Wildtyp, heterozygoter Typ, Polymorphismus). Je nachdem, wie diese dreigestufte Variable weiter interpretiert wird, ergeben sich unterschiedliche Auswertemethoden. Tabelle 5 fasst die gängigen Kodierungen unter verschiedenen genetischen Vorannahmen zusammen. Nicht dargestellt sind hier Tests zwischen lediglich zwei der drei Genotypen (z.B. Polymorphismus vs. Wildtyp), die gelegentlich auch berechnet werden. Das allelische Modell ist hier ebenfalls nicht aufgeführt, da es einen Sonderfall darstellt und nicht der Genotyp, sondern das Allel Gegenstand der Untersuchung ist. Es wird unten beschrieben (2.2.1.4.).

		Wildtyp	Heterozygoter Typ	Polymorphismus
Dominantes Modell		0	1	1
Rezessives Modell		0	0	1
Allel zählendes Modell		0	1	2
Dummy-Modell	Dummy 1	0	1	0
	Dummy 2	0	0	1
Kategorielles Modell		x	y	z

Tabelle 5: Kodierungen eines SNPs bei unterschiedlichen genetischen Modellen

Offensichtlich resultieren verschiedene Nullhypothesen, je nachdem ob man ein dominantes, rezessives oder Allel zählendes Modell annimmt oder gar kein genetisches Modell zugrunde legt. In der Regel gibt es bei Kandidatengenansätzen wenig Wissen über das zugrunde liegende genetische Modell. Da aber häufig fraglich bleibt, ob überhaupt eine Annahme gerechtfertigt ist, wird gelegentlich auf neutralere Modelle ausgewichen. So werden z.B. SNPs auch in der Dummy-Kodierung bei Modellschätzungen verwendet oder es werden sämtliche genetischen Modelle getestet. Letzteres verschärft allerdings wieder das Problem des multiplen Testens. Es wurde vorgeschlagen, bei nicht bekanntem Vererbungsmodell ein Allel zählendes Modell zu verwenden (Schaid 1996). Dieses Modell besitzt die höchste Power, auch bei Gültigkeit anderer genetischer Modelle Assoziationen aufzudecken. Dieses Modell wird daher derzeit den meisten genomweiten Studien zugrunde gelegt.

2.2.1.1 Allel zählendes Modell

Betrachten wir eine Aufteilung der Genotypen innerhalb einer Gruppe von Kranken (Fälle) und nicht Erkrankten (Kontrollen), so ergibt sich folgendes Schema (Tabelle 6).

	Anzahl Allele			Gesamt
	Wildtyp	heterozygot	polymorph	
	0	1	2	
Fälle	r_0	r_1	r_2	R
Kontrollen	s_0	s_1	s_2	S
Gesamt	n_0	n_1	n_2	N

Tabelle 6: Häufigkeiten in Fällen und Kontrollen nach Genotyp

Im Allel zählenden Modell wird getestet, ob eine lineare Änderung (also eine Zunahme oder Abnahme) des Risikos bei ansteigender Allelzahl vorliegt. Dieses Modell wird gelegentlich auch als kodominant oder additiv bezeichnet. Dies geht auf die biologisch interpretierbare Variante zurück, den Genotyp entsprechend der Anzahl der Allele zu kodieren, die im Genotyp vorkommen (also 0-1-2). Eine andere lineare Kodierung ändert allerdings nichts an Effektgrößen und Signifikanztests innerhalb der Analysen auf univariater Ebene. Der gängige Test für dieses Modell ist der Armitage-Trend Test (Armitage 1955; Sasieni 1997). Die χ^2 -verteilte Teststatistik wird wie folgt berechnet (Formel 5).

$$Y^2 = \frac{N \{ N(r_1 + 2r_2) - R(n_1 + 2n_2) \}^2}{R(N - R) \{ N(n_1 + 4n_2) - (n_1 + 2n_2)^2 \}}, \text{ wobei die Notation entsprechend Tabelle 6}$$

Formel 5: Armitage-Trend Test

2.2.1.2 Dominantes und rezessives Modell

Zusammenfassungen von Kategorien innerhalb des Genotyps führen zum dominanten oder rezessiven Modell. Das dominante Modell fasst beide Kategorien, die zumindest einmal das seltene Allel haben, zusammen und testet gegen den Wildtyp; das rezessive Modell testet das gleichzeitige Auftreten zwei seltener Allele, was als Polymorphismus bezeichnet wird, gegen die beiden anderen Genotypen. Zur Ermittlung der relevanten Teststatistik kann die gängige Formel eines Chi-Quadrat-Tests herangezogen werden.

$$\chi^2 = \sum_{k=1}^n \frac{(f_{\text{observed}} - f_{\text{expected}})^2}{f_{\text{expected}}}$$

Formel 6: Chi-Quadrat-Test

Auch die Effektmaße basieren in diesem Fall auf einer Vierfeldertafel (Tabelle 7). So berechnet sich die Odds Ratio im Falle eines dominanten Modells entsprechend Formel 7.

	Anzahl Allele		Total
	0	1 oder 2	
Fälle	r_0	r_1+r_2	R
Kontrollen	s_0	s_1+s_2	S
Total	n_0	n_1+n_2	N

Tabelle 7: Häufigkeiten in Fällen und Kontrollen im dominanten Modell

$$\frac{r_1 + r_2}{s_1 + s_2} \cdot \frac{r_0}{s_0}$$

Formel 7: Beispiel zur Berechnung einer Odds Ratio

2.2.1.3 Genotyptest ohne Modell

2.2.1.3.1 Kategorielles Modell

Ein weiterer möglicher Test, der keinerlei genetisches Modell zugrunde legt, ist der χ^2 -Test mit 2 Freiheitsgraden, bei dem getestet wird, ob die Effekte innerhalb der genetischen Variation homogen sind. Wird die globale Nullhypothese keines Effekts angenommen, erübrigen sich weitere Tests. Bei Ablehnung der Homogenitätshypothese, ist jedoch noch nicht spezifiziert, welche Unterschiede die bedeutsamen sind. Nicht auszuschließen ist dabei auch ein „Heterozygoten-Effekt“, also eine überzufällig häufige oder seltene Anzahl von Krankheitsträgern in der Gruppe der heterozygoten im Vergleich zu beiden homozygoten. Dieser Fall entspricht keinem der bekannten genetischen Modelle und bleibt daher schwer interpretierbar. Zu beachten ist auch, dass bei geringen Zellbesetzungen die Annahme einer χ^2 -verteilten Testgröße in Frage steht und in solchen Fällen Fishers exakter Test eine bessere Alternative darstellt. Dennoch kommt dieser Test in der Praxis gelegentlich zur Anwendung und ist auch in Standardpaketen (wie z.B. SAS Genetics) implementiert.

2.2.1.3.2 Dummy-Modelle

Eine weitere Kodierung eines SNPs, welches sich auf kein genetisches Modell festlegt, ist die Verwendung von Dummy-Variablen. Dabei wird eine Variable mit n Kategorien durch n-1 Dummies präsentiert. Zur Darstellung eines SNPs benötigen wir also zwei Dummy-Variablen (siehe Tabelle 5). Die einzelne Dummy-Variable gibt den Effekt der jeweiligen Kategorie zur Referenzkategorie an. In der Regel wird der Wildtyp als Referenzkategorie verwendet. Dann lassen sich durch die Dummies die Effekte des heterozygoten Genotyps, bzw. des Poly-

morphismus abbilden. Dies sind klar interpretierbare Effekte², die man häufig in der Epidemiologie zur Beschreibung einer kategoriellen Variablen verwendet. Es werden also zwei Effekte einem Modell zugrunde gelegt, der Effekt vom Wildtyp zum heterozygoten Typ sowie der Effekt vom Wildtyp zum Polymorphismus. Daneben lassen sich weitere Varianten der Darstellung eines SNPs über spezielle Design-Variablen einführen, womit man den genetischen Vorstellungen noch näher kommen will. Dabei ist vor allem die Zerlegung des Effekts des Genotyps in eine additive und dominante Komponente zu erwähnen (z.B. (Zeng et al. 2005)).

2.2.1.4 Allelisches Modell

Beim allelischen Modell ist nicht mehr der Genotyp Untersuchungseinheit, sondern das Allel. Da jedes Individuum an einem autosomalen Locus zwei Allele hat, gibt es zweimal so viel Allele wie Personen. Verglichen werden Häufigkeiten des seltenen und häufigen Allels zwischen den Fällen und Kontrollen. Somit kann man hier von einer Haplotypanalyse im einfachsten Sinne sprechen.

Dabei tragen alle Personen mit Wildtyp das häufige Allel zweimal, alle Personen mit Polymorphismus tragen das seltene Allel zweimal. Personen mit heterozygotem Genotyp tragen jeweils einmal das häufige und einmal das seltene Allel. Entsprechend ergibt sich die Darstellung in Tabelle 8. Die χ^2 -verteilte Teststatistik des allelischen Tests wird entsprechend Formel 8 berechnet

	Anzahl Allele		Total
	Allel 1	Allel 2	
Fälle	r_1+2r_2	r_1+2r_0	R
Kontrollen	s_1+2s_2	s_1+2s_0	S
Total	n_1+2N_2	n_1+2N_0	N

Tabelle 8: Häufigkeiten in Fällen und Kontrollen im allelischen Modell

$$Y_A^2 = \frac{2N \{2N(r_1 + 2r_2) - 2R(n_1 + 2n_2)\}^2}{(2R)2(N - R) \{2N(n_1 + 2n_2) - (n_1 + 2n_2)^2\}}$$

Formel 8: Allelischer Test

Obwohl der Armitage-Trend-Test den Vorteil hat, gegenüber Abweichungen vom Hardy-Weinberg-Gleichgewicht robust zu sein, wird auch der Allel-basierte Test häufig angewandt (Zheng et al. 2007).

² Ein SNP kann hier als signifikant betrachtet werden, wenn das Modell, welches beide Dummy-Variablen einbezieht, eine signifikante Assoziation mit dem Phänotyp zeigt. Dies kann man mit einem Waldtest prüfen.

2.2.2 Zusammenhänge zwischen einem SNP und quantitativen Phänotypen

Auch beim Nachweis eines Zusammenhangs von SNPs mit quantitativen Zielvariablen lassen sich für die SNPs wieder ein dominantes, rezessives und allelisches Modell anwenden. Ebenso ist die Verwendung des SNPs in einer Allel zählenden Kodierung oder Dummy-Kodierung möglich. Da die wesentlichen Zielgrößen dieser Arbeit dichotome Variablen sind, sei hier nur auf die quantitativen Zielgrößen verwiesen, die in Zusammenhang mit Asthma eine Rolle spielen können. Dies sind vor allem IgE-Werte (auf Grund fehlender Gültigkeit der Normalverteilungsannahme meist in einer transformierten Form) oder Parameter der Lungenfunktion (meist als Prozente vom Soll). Für Auswertungen dieser Variablen ist nicht die logistische, sondern die multiple Regression das Mittel der Wahl. Bei einfachen Modellen genügen je nach Gültigkeit der Normalverteilungsannahme auch t-Tests und univariate Varianzanalysen oder Wilcoxon und Kruskal-Wallis-Tests.

2.2.3 Zusammenhänge zwischen einem Haplotyp und qualitativen Phänotypen

Bei einer Haplotypanalyse wird der Auslöser einer Krankheit nicht auf einem speziellen Locus, sondern in einer Kombination verschiedener Loci vermutet. Haplotypanalysen testen, ob eine solche Kombination (ein Haplotyp) mit einer Krankheit assoziiert ist.

Vor einem solchen Test muss bedacht werden, in welcher Stichprobe die Haplotypen geschätzt werden und welche SNPs in die Analyse einbezogen werden. Erst nach diesen Entscheidungen kann dann mittels diverser Tests entschieden werden, ob es Unterschiede zwischen Kranken und Gesunden hinsichtlich eines Haplotyps gibt (Schaid 2004).

Daraus ergeben sich folgende Vorüberlegungen

1. Welche SNPs sollen in die Haplotypanalyse einbezogen werden?
2. Welcher Teil der Stichprobe soll in die Haplotypanalyse einbezogen werden?
3. Welcher Algorithmus soll herangezogen werden, um die Haplotypen zu schätzen?
4. Welcher Test soll verwendet werden, um Unterschiede zwischen Fällen und Kontrollen aufzudecken

2.2.3.1 Auswahl der SNPs für einen Haplotyp

Die Auswahl der SNPs, mit welchen die Haplotypanalyse durchgeführt werden soll, ist abhängig von der Vorauswahl der SNPs innerhalb eines Projektes. Abbildung 21 zeigt mögliche Varianten.

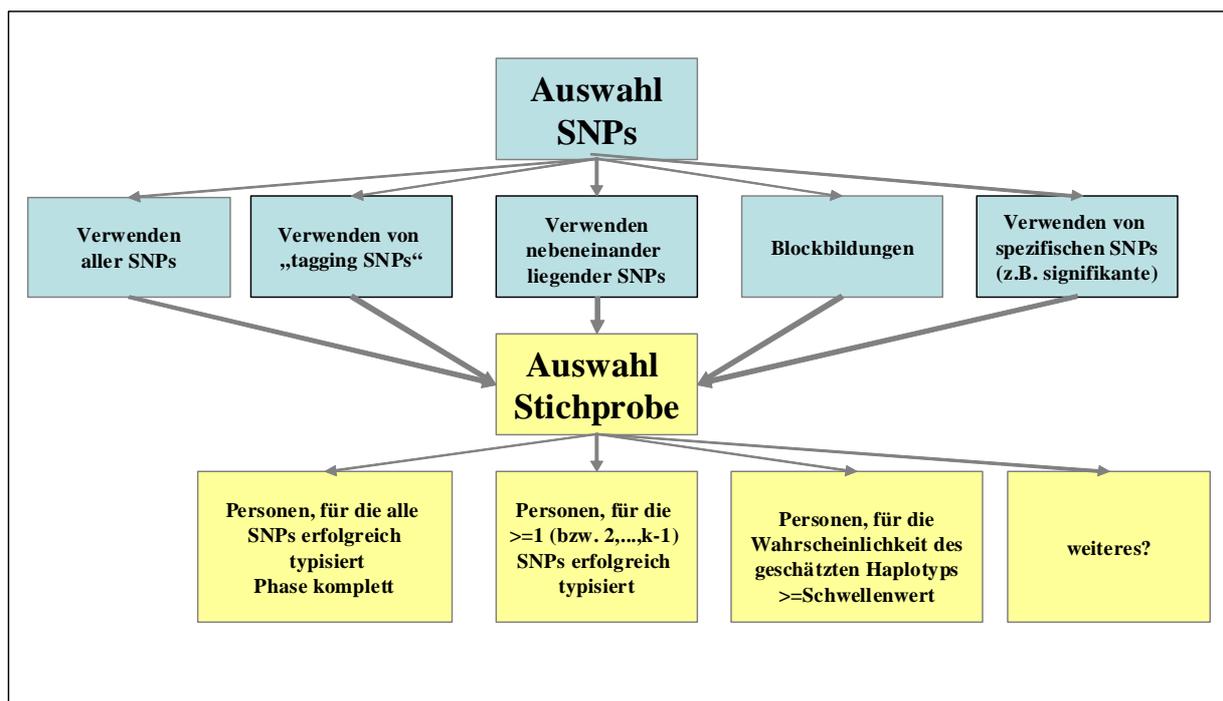


Abbildung 21: Varianten der SNP- und Stichprobenselektion für Haplotypanalysen

In der Regel wird im Rahmen des Kandidatengenansatzes ein Gen oder eine bestimmte Region untersucht. Soll das ganze Gen durch den Haplotyp abgebildet werden, so ist eine nahe liegende Option, sämtliche SNPs, welche in dieser Region gefunden und typisiert wurden, in die Haplotypanalyse einzubeziehen.

Auch das ausschließliche Benutzen von „tagging SNPs“ kann eine ganze Region abbilden, wenngleich, dadurch dass einzelne „tagging SNPs“ stellvertretend für einige andere SNPs stehen können, dies weniger genau ist. Mit der sogenannten „Fenster-technik“ werden Schritt für Schritt kleine Regionen innerhalb des gesamten Bereichs untersucht, indem stets n =Anzahl der SNPs nebeneinander liegende SNPs zusammengefasst werden. Fasst man nebeneinander liegende SNPs mit hohem LD zusammen, so ergibt sich die „spine-Methode“, bei welcher Marker innerhalb eines Blocks in hohem LD sind, nicht aber die dazwischen liegenden Segmente. Elaboriertere Methoden zur Segmentbildung sind die Konfidenzmethode von Gabriel (Gabriel et al. 2002) oder die Vier-Gameten-Regel (Wang, N. et al. 2002).

2.2.3.2 Auswahl der Stichprobe für die Haplotypanalyse

Die Auswahl der Stichprobe ist im Wesentlichen von der Qualität der Genotypisierung abhängig. Im optimalen eher theoretischen Fall gelingt für jede Person die Typisierung auf jedem SNP. Dann wird man auch Haplotypanalysen auf der gleichen Stichprobe rechnen, für welche die Studie geplant war. In der Regel gibt es aber Personen, für welche die Typisierungen weniger erfolgreich laufen. Es empfiehlt sich daher, Personen von den Haplotypanalysen auszuschließen. Misslingt die Typisierung auf sämtlichen Markern, so ist von einem Problem der DNA auszugehen, weshalb solche Personen generell ausgeschlossen werden. Ansonsten lassen sich Schwellen bilden, die Personen ausschließen, für welche die Typisierung auf einer bestimmten Anzahl von SNPs misslungen ist. Wenn Information auf mindestens einem Marker vorliegt, lassen sich auch Imputationsalgorithmen benutzen, welche die fehlenden Werte ersetzen (Zhao, Z. et al. 2008).

Eine etwas andere Vorgehensweise ist Folgende: Für jede Person werden individuelle Wahrscheinlichkeiten einer Kombination von Haplotypen getestet. Die Haplotypen mit der höchsten Wahrscheinlichkeit werden dann einer Person zugeordnet. Sobald diese maximale Wahrscheinlichkeit über einer gewissen Schwelle liegt, wird die Person berücksichtigt (z.B. werden alle Personen berücksichtigt, deren maximale Wahrscheinlichkeit über 90% liegt). Um die Richtigkeit der Ergebnisse sicher zu stellen, empfehlen sich insbesondere beim Einschluss von Personen, für die viele Werte geschätzt werden müssen, Sensitivitätsanalysen. Diese können z.B. der Frage nachgehen, wie sich Häufigkeiten der Haplotypen für komplett erfolgreich

typisierte Personen von Subgruppen unterscheiden, bei denen viele Genotypen geschätzt werden mussten.

2.2.3.3 Schätzen der Haplotypen

Haplotypen sind in der Regel keine direkt beobachtbaren Größen. Ein Haplotyp wird gebildet aus einer Kette von Allelen auf einem Chromosomenstrang. Ebenso wie es zwei Allele pro SNP in einer Person gibt, ergeben sich auch zwei Haplotypen für ein Set von SNPs innerhalb einer Person. In der Regel ist nicht bekannt, welches Allel auf welchem Chromosomenstrang liegt. Die Haplotypen sind daher zunächst nur bei einer Kombination homozygoter SNPs festgelegt, da dann das gleiche Allel auf beiden Chromosomensträngen zu finden ist. Gibt es nur einen heterozygoten SNP, so liegen die beiden möglichen Haplotypen immer noch fest, da die zwei möglichen Haplotypen nur durch einen SNP variiert werden können. Sobald es aber mehrere heterozygote Stellen gibt, gibt es mehrere nicht mehr eindeutig nachweisbare Varianten. Abbildung 22 zeigt, dass für einen Träger von zwei heterozygoten Genotypen mehrere Möglichkeiten bestehen, Haplotypen zu bilden.

	SNP 1 A/G	SNP 2 C/T	SNP 3 C/C
Haplotypen Variante 1	A	C	C
	G	T	C
Haplotypen Variante 2	G	C	C
	A	T	C

Abbildung 22: Variationsmöglichkeiten der Haplotypen bei Trägern mehrerer heterozygoter SNPs

Die wirklich vorliegenden Varianten können im Rahmen einfacher Genotypisierungen nicht bestimmt werden, bzw. deren Bestimmung im Labor wäre derzeit noch zu aufwändig. Daher werden die Häufigkeiten von Haplotypen in einer Population in der Regel geschätzt. Hierzu wurden verschiedene Algorithmen vorgeschlagen. Dazu gehören unter anderem ein Sparsamkeitsmodell (Clark 1990), ein Bayes-basiertes populationsgenetisches Modell (Stephens, M. et al. 2003) und Algorithmen, die auf der Maximum-Likelihood-Theorie begründet sind

(Excoffier et al. 1995; Excoffier et al. 2003). Von den verschiedenen vorgeschlagenen Schätzverfahren basieren die meisten Programme auf dem „expectation-maximization“ (EM)-Algorithmus. Die Vor- und Nachteile verschiedener Software wurden 2005 für 46 Programme beschrieben (Salem et al. 2005). Einige der wichtigsten Programme sind in Tabelle 9 zusammengefasst.

Programm (Algorithmus)	Web-Seite	Referenz
Arlequin (EM ¹ und ELB ²)	http://lgb.unige.ch/arlequin/	(Excoffier et al. 1995; Excoffier et al. 2003)
Clark or HAPINFEX	(keine)	(Clark 1990)
HAPLO (EM ¹)	http://krunch.med.yale.edu/haplo/	(Hawley et al. 1995)
Haplotyper	http://www.people.fas.harvard.edu/~junliu/Haplo/docMain.htm	(Niu et al. 2002)
PHASE	http://www.stat.washington.edu/stephens/software.html	(Stephens, M. et al. 2001) (Stephens, M. et al. 2003)

Tabelle 9: Programme zur Schätzung von Haplotypen (Crawford et al. 2005)

¹EM= „expectation-maximization“

²ELB= „Excoffier-Laval-Balding algorithm“

2.2.3.4 Assoziationstests für Haplotypen

Erst nach Abschluss dieser Vorüberlegungen werden Assoziationstests durchgeführt. Ein globaler Test bezieht sich dabei auf die gesamte Verteilung der Haplotypen, während individuelle Tests einen speziellen Haplotyp herausgreifen.

	Haplotyp 1	Haplotyp 2	(...)	Haplotyp i	(...)	Haplotyp k	Total
Fälle	r ₁	r ₂		r _i		r _k	R
Kontrollen	s ₁	s ₂		s _i		s _k	S
Total	n ₁	n ₂		n _i		n _k	N

Tabelle 10: Kodierungen von Haplotypen

Die Möglichkeit, mittels diverser Schätzalgorithmen die Häufigkeiten von Haplotypen zu schätzen, erlaubt eine Schätzung der Haplotypfrequenzen für Fälle und Kontrollen. Mittels Likelihood Ratio Tests kann getestet werden, ob sich Fälle signifikant von Kontrollen unterscheiden, wobei die log Likelihoods (bestimmte statistische Funktionen) getrennt für Kontrollen, Fälle und die aus Fällen und Kontrollen gepoolte Stichprobe maximiert werden. Unter der Nullhypothese keiner Unterschiede ist die Teststatistik aus Formel 9 χ^2 -verteilt mit Anzahl der Freiheitsgrade $df = df_{cases} + df_{controls} - df_{pooled}$ (Zhao, J. H. et al. 2000).

$$\chi^2 = 2(L_{cases} + L_{controls} - L_{pooled})$$

Formel 9: Likelihood Ratio Test zum Testen auf Assoziationen des globalen Haplotyps mit dem Phänotyp

Ein Problem dieser Testgröße ist, dass die Annahme einer χ^2 -Verteilung im Falle sehr vieler Haplotypen auf Grund zu geringer Zellbesetzungen nicht haltbar ist. Das könnten z.B. alle Haplotypen mit einer Frequenz <3% in der Stichprobe sein. Ein weiterer Nachteil der Testgröße des globalen Haplotypentests besteht darin, dass keine Adjustierung für Kovariaten möglich und Hardy-Weinberg-Equilibrium in allen SNPs vorausgesetzt ist.

Eine Alternative ist daher die Beschränkung des globalen Tests auf Haplotypen, die einen gewissen Mindestprozentsatz ausmachen (z.B. $\geq 3\%$). Zur Berechnung von Assoziationen zwischen individuellen Haplotypen und einem Phänotyp betrachte man Tabelle 11.

	Haplotyp 1	Haplotyp 2	Total
Fälle	c_{11}	c_{12}	t_1
Kontrollen	c_{21}	c_{22}	t_2
Total	h_1	h_2	T

Tabelle 11: Formalismus zum Testen auf Assoziationen eines individuellen Haplotyps mit dem Phänotyp

Hier steht $T = 2n = t_1 + t_2 = h_1 + h_2$ für die Gesamtzahl von Haplotypen in der Stichprobe (man beachte, dass es immer doppelt so viele Haplotypen wie Personen gibt).

Haplotyp 1 bezieht sich hierbei auf den gerade zu testenden Haplotyp, während Haplotyp 2 für alle anderen Haplotypen steht. Der Wert c_{ij} steht für die „pseudobeobachtete“ Anzahl von Individuen mit dem Phänotyp i und Haplotyp j. Man beachte, dass diese Zahlen nicht notwendigerweise ganzzahlig sein müssen, da die Haplotypen nicht beobachtet, sondern geschätzt werden.

Aus der Vierfeldertafel lässt sich dann wieder ein Chi-Quadrat-Test mit einem Freiheitsgrad ableiten, wobei gilt:

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(c_{ij} - t_i h_j / T)^2}{t_i h_j / T}$$

Formel 10: χ^2 -Test zum Testen auf Assoziationen eines individuellen Haplotyps mit dem Phänotyp

2.2.3.5 Effektgrößen bei Haplotypanalysen, Haplotyptrendregression

Wenn es gelingt, einen Haplotyp zu identifizieren, der mit einem Phänotyp assoziiert ist, können die darin vorkommenden SNPs auch wieder als ein Subset wichtiger SNPs zur Erklärung eines Phänotyps aufgefasst werden. In diesem Sinne kann auch hier von einer Interaktion gesprochen werden. Im Unterschied zu den Multilocusmodellen, welche Genotypen beobachten können, beruhen Haplotypanalysen aber in der Regel auf geschätzten Entitäten. Daher sind die Analysen etwas vorsichtiger zu interpretieren.

Wie auf Einzel-SNP-Ebene werden Effektmaße berechnet, die eine Erhöhung (oder Verminderung) des Risikos für eine Erkrankung auf eine genetische Variante (Haplotypen) beziehen. Odds Ratios für Haplotypen können zwar berechnet werden, sind allerdings in der Interpretation problematisch. Die Ausgangslage ist ja, dass pro Individuum einem Phänotyp (Fall oder Kontrolle) zwei Haplotypen zugeordnet werden. Ergebnisse der Haplotypanalyse sind aber Odds Ratios für einzelne Haplotypen. Zudem beruhen die Vergleiche auf Populationsschätzungen und das Tragen eines Haplotyps ist noch nicht auf Individuen übertragbar. Es gibt jedoch Vorgehensweisen, die einem Individuum basierend auf den Haplotypschätzungen die wahrscheinlichsten Haplotypenpaare zuordnen.

Regressionsmodelle im Kontext von Haplotypen sind zusammen mit weiteren Assoziations-test, die Haplotypen einbeziehen, anderweitig exzellent zusammengefasst (Schaid 2004). Im Rahmen der vorliegenden Arbeit wurde nur auf eine Variante von Regressionsmodellen, die Haplotyptrendregression, zurück gegriffen (Zaykin et al. 2002). Bei diesem Verfahren werden geschätzte Wahrscheinlichkeiten individueller Haplotypen als unabhängige Variablen logistischer Regressionsmodelle verwendet. Regressionsmodelle liefern methodische Vorteile, die sich in einer möglichen Einbeziehung von Kovariaten, der Durchführung von schrittweisen Selektionen oder dem Berechnen von Haplotyp-Umwelt-Interaktionen zeigen. Dennoch sind die bestehenden Methoden nicht perfekt, genetische Modelle zu repräsentieren (Schaid 2004).

2.2.4 Besonderheiten bei GWA-Studien

Wie eingangs erwähnt, weisen GWAs spezielle Probleme auf, die es in den Griff zu bekommen gilt (Ziegler et al. 2008). Dazu gehören die Frage nach der gewählten Stichprobe und möglicher Stratifikation sowie Fragen nach Power und Korrektur für multiples Testen.

Auch bei GWAs geht man davon aus, dass kleine Effekte aufgedeckt werden. Dies verlangt große Stichproben. Hierbei ist zu Gunsten einer höheren Power eine Anreicherung der Fälle gewünscht, was allerdings zu phänotypischer Heterogenität führen kann. Insgesamt lässt sich zwar sagen, dass sich mit einem breiten Spektrum von Rekrutierungsvarianten Studiendesigns erstellen lassen, die mit GWAS gut kompatibel sind. Dennoch gilt es sich folgender Punkte bewusst zu sein:

- Da die Fälle meist aus Kliniken stammen, stellt sich die Frage der Repräsentativität, da man in Kliniken besonders milde Formen bestimmter Krankheiten nicht so leicht findet.
- Umstritten ist auch die Wahl einer geeigneten Kontrollgruppe (Consortium, Wellcome Trust Case Control 2007). Fraglich werden Ergebnisse vor allem dann, wenn es große Unterschiede zwischen Fällen und Kontrollen hinsichtlich der interessierenden Phänotypen oder genetischer Risikofaktoren gibt.
- Gelegentlich werden auch historische Kontrollen genommen, Dies kann jedoch zu Ungleichheiten bei Fällen und Kontrollen führen. Schon in der Fragestellung kann es Unterschiede geben, was die Analysen bestimmter Variablen etwas problematisch gestalten kann. Allerdings kann informative Heterogenität durch Erhöhung der Stichprobe ausgeglichen werden (McCarthy et al. 2008).
- Ansonsten übliche Vergleichstabellen (Chanock et al. 2007) hinsichtlich wichtiger Charakteristika von Fällen, bzw. von Kontrollen findet man selten in GWAs (Pearson et al. 2008), wären aber wünschenswert

Die Mischung von Populationen, um große Stichproben zu erhalten, rückt das Problem der Stratifikation wieder in den Vordergrund. Daher sollte in jedem Fall geprüft werden, ob sich Substrukturen finden lassen, wenn Populationen verschiedener Länder gepoolt werden. Inzwischen gibt es allerdings gute Werkzeuge, Stratifikation zu korrigieren (Price et al. 2006; Setakis et al. 2006; Zheng et al. 2006). Tools um Substruktur zu entdecken (Paschou et al. 2007; Price et al. 2008; Tian et al. 2008) sind ein willkommenes Nebenprodukt der GWAs.

Besondere Aufmerksamkeit muss aufgrund der Vielzahl von Tests im Rahmen einer GWA dem Problem des multiplen Testens gewidmet werden. Würde man das übliche Signifikanz-

niveau von 0,05 heranziehen, könnte man bei 1 Million SNPs mit etwa 50000 assoziierten SNPs rechnen. Daher muss das Signifikanzniveau deutlich herabgesetzt werden (Risch, N. et al. 1996; Consortium, The International HapMap 2005; Hoggart et al. 2008). Gewöhnlich wird eine Bonferroni-Korrektur in Abhängigkeit von der Anzahl der Marker vorgenommen (Lin 2006). Diese Schwellen wurden kritisiert, da die SNPs aufgrund von LD nicht unabhängig sein dürften. Zudem ist die Power bei Verwendung dieser traditionellen Vorgehensweise eher herabgesetzt, weshalb sich auch hier die FDR (Benjamini et al. 1995; Benjamini et al. 2001) etabliert hat (Sun, L. et al. 2006). Auch Bayessche Korrekturverfahren wurden vorgeschlagen. Sämtlichen Verfahren ist es gemeinsam, dass sie eine sehr niedrige Schwelle für den α -Fehler in der Größenordnung von 5×10^{-8} zugrunde legen.

Um eine weitere Inflation von Tests zu vermeiden, beschränkt man sich in der Analyse meist auf ein genetisches Modell. Auch im Rahmen der GWAs ist der übliche Test ein Test mit einem Freiheitsgrad, in der Regel der Cochran-Armitage-Trend-Test, auf jedem Marker. Solche Tests können mit und ohne Adjustierungen vorgenommen werden. Dabei kann u.a. auch für Populationsstruktur adjustiert werden. Dieser Test hat die meiste Power und wird daher in den meisten GWAs eingesetzt. Daneben kommen vor allem dominante oder rezessive Modelle zumindest explorativ zum Einsatz. Die Verwendung dieser genetischen Modelle geht aber ein höheres Risiko ein, mögliche Krankheitsvarianten zu übersehen.

Als Darstellung der wichtigsten Ergebnisse einer GWA wird in der Regel ein Plot der negativen logarithmieren p-Werte gezeigt, der leichter zu lesen ist als ein Plot, der die p-Werte abbildet.

Die Geschichte vieler fehlender Replikationen in Kandidatengenstudien, die hohe Dimensionalität von GWA-Studien, die damit verbundene hohe Vulnerabilität an einer Vielzahl von Stellen und die potentiellen Möglichkeiten von Verzerrungen und Fehlern erklären bereits die eminent große Bedeutung von Replikationsstudien. Diese sind damit bereits Standard der meisten genomweiten Studien und meist in mehrstufige Designs integriert.

Die Verwendung mehrstufiger Designs dient der Minimierung falsch positiver Befunde ebenso wie der Kostenreduzierung unter Beibehaltung gleicher Power (Hirschhorn et al. 2005). Meistens wird ein Zweiphasen-Design verwendet (Lin 2006), aber auch andere Designs sind möglich (Thomas et al. 2005). Bei einem Zweiphasen-Design werden in der ersten Phase alle ausgewählten SNPs in einem Teil der Stichprobe und nachfolgend die besten Ergebnisse in der Gesamtstichprobe getestet (Satagopan et al. 2002; Satagopan et al. 2004). Es werden aber auch Verfahren vorgeschlagen, die vom Zweistufendesign wieder abrücken und mit der gleichen oder höherer Power einstufige Designs verwenden (Skol et al. 2006).

Replikationsstudien sollten zunächst möglichst den identischen Phänotyp wie in der ursprünglichen Studie im Hinblick auf eine Assoziation mit genetischen Varianten untersuchen. Später werden GWAs dann meist auf verwandte Phänotypen ausgedehnt (Chanock et al. 2007). In jedem Fall ist es klug, erst einmal eine definitiv validierte Assoziation gefunden und bestätigt zu haben, bevor man weitere Studien durchführt, um ein mögliches Scheitern der Replikation dann auch interpretieren zu können. Assoziationsstudien liefern selbst nach positiven Replikationsstudien noch wenig Information über die Genfunktion (außer vielleicht, dass SNPs bereits auf Basis erhöhter Genexpression ausgewählt wurden). Funktionelle Studien, die genetische Manipulationen in Zelllinien oder Tiermodellen beinhalten, sollten daher der Identifizierung neuer Kandidatengenregionen folgen (Frayling et al. 2007).

Ein letzter GWA-spezifischer Punkt betrifft die Auswahl von SNPs. Diese ist derzeit auf die limitierte Auswahl von Chips begrenzt, welche eine bequem durchführbare GWA ermöglichen (Barrett et al. 2006; Pe'er et al. 2006). Dies kann Lücken in interessanten Regionen zur Folge haben. Man nimmt an, dass 67-89% der häufigen SNPs in europäischen Populationen durch die Chips abgedeckt ist (Consortium, The International HapMap 2005; Frazer et al. 2007). Mit nachfolgendem „fine-mapping“ können solche Regionen genauer untersucht werden.

Da das Set von SNPs auf einem Chip mit einer gewissen Wahrscheinlichkeit die kausale Variante nicht einschließt, kann man den genomweiten Ansatz so verstehen, als ob Marker auf dem Chip als Prädiktoren von möglichen untypisierten Varianten bestehen (Marchini et al. 2007). Dabei ist weithin akzeptiert, dass dieser Ansatz für Studien dieser Sorte nicht derjenige mit der meisten Power ist. Eine der gegenwärtigen Herausforderungen für die Analyse besteht daher darin, zusätzliche Information über die nicht typisierten Daten zu bekommen. Dies kann durch sogenannte Imputationsverfahren geschehen, die Information über die genotypisierten Marker mit Information hinsichtlich der Variationen aus existierenden Datenbanken wie z.B. HapMap (<http://www.hapmap.org/>) kombinieren.

Kurz gefasst werden dabei a) Daten aus einer Studie und b) Daten aus einer Referenzpopulation *in silico*, meist HapMap-Daten genutzt und dann c) ein genetisches Modell verwendet, um Genotypen zu simulieren oder imputieren, die in einer Studie nicht typisiert worden sind. Solche *in silico* geschätzten Genotypen können im Anschluss so behandelt werden, als wären die SNPs real typisiert worden. Verschiedene Verfahren sind entwickelt worden (Halperin et al. 2009), aber es besteht kein Konsens darüber, welche Methode am besten geeignet ist (Marchini et al. 2007).

2.3 Bestimmung der Genotypen

2.3.1 CHIP

Die Fortschritte in der Technologie der Genotypisierungen haben GWAs erst möglich gemacht (Teo 2008). Die DNA-Chip-Technologie nutzt Techniken aus der Halbleiterfertigung, um bekannte Gene auf einem winzigen Plastik- oder Glasplättchen, dem Microarray, zu identifizieren und deren Aktivität zu messen. Microarrays können sowohl zur Bestimmung relativer Änderungen der Genexpression, als auch zur Genotypisierung eingesetzt werden. Bis zu einer Millionen genetischer Varianten können heutzutage gleichzeitig mittels vorab präparierten Microarrays untersucht werden. Solche Microarrays werden z.B. von Affymetrix (Santa Clara, California, USA) oder Illumina (San Diego, California, USA) erstellt.

ABLAUF	
Wahl des Studiendesigns	z.B. Fall-Kontroll-Studien (Betroffene Fälle, gesunde Kontrollen)
Auswahl der Variablen	Genetische Marker (SNPs) typischerweise mit 2 Formen A und B, einer der 3 Genotypen: AA, AB, BB
Auswahl des Array (nebenstehend einige Beispiele)	<ul style="list-style-type: none"> • Affymetrix Gene Chip Human Mapping 500K Array Set (ca. 500000 SNPs) • Affymetrix Genome-Wide Human SNP Array 5.0 (die gleichen SNPs wie beim Array-Set, aber auf einem einzigen Chip; zudem sind 420000 Mikrosatelliten, bzw. CNVs¹ hinzugefügt) (ca. 500000 SNPs) • Affymetrix Genome-Wide Human SNP Array 6.0 (ca. 1000000 SNPs) • Illumina HumanHap550 Genotyping BeadChip (ca. 550000 SNPs) • Illumina HumanHap650Y Genotyping BeadChip (Y zeigt, dass der Chip etwa 100000 SNPs enthält, die spezifisch für die Yoruba-Population sind) (ca. 650000 SNPs) • Illumina Human1M BeadChip (ca. 1000000 SNPs)
Vorbereitung der genomischen DNA	
Hybridisierung	Auftragen der markierten Proben auf den Array
Scannen	Detektieren der Farbintensitäten, in welcher Größenordnung DNA auf dem CHIP verbleibt
Normalisierung	Standardisierung der gemessenen Werte und anschließende Bestimmung des Genotypen
Qualitätskontrolle	<ul style="list-style-type: none"> • Analyse der „cluster plots“ • Ausschluss von SNPs mit geringer Qualität • Ausschluss von Probanden mit Genotypen von geringer Qualität • Ausschluss von Probanden mit kontaminierter DNA
Wahl des genetischen Modells und statistische Analyse	

Tabelle 12: Typischer Ablauf einer genomweiten Assoziationsstudie (nach Ziegler (Ziegler et al. 2008))

¹ CNV=„Copy number variation“

Tabelle 12, welche den Genotypisierungsprozess in den Ablauf einer GWA einbindet, gibt einen Überblick über bekannte Chips.

Meist wird ein bekannter Chip verwendet, so dass die SNPs im Vorfeld einer GWA festliegen. Kundenangepasste Chips sind zwar möglich, aber derzeit noch recht teuer. Wie bei jeder Genotypisierungsmethode ist das Ziel der Chip-Technologie, Variationen in der DNA-Sequenz in spezifischen Genen oder Regionen des menschlichen Genoms zu identifizieren.

Die einzelnen Felder des Arrays auf einem Chip sind dazu mit einzelsträngigen DNA-Stücken beschichtet. Durch Zugabe der mit einem Fluoreszenzfarbstoff markierten Untersuchungsproben, der sogenannten Hybridisierung, binden diese dann bei komplementärer Basenabfolge an die DNA auf dem Chip. Die Position, Intensität und Wellenlänge der entstehenden Mischfarbe werden mit einer hochauflösenden Laserkamera detektiert. Dabei entsteht eine Serie von Hybridisierungsintensitäten, welche in den aktuellen Genotyp übersetzt werden müssen. Dieser Prozess des Übersetzens ist auch als Callen der Genotypen bekannt.

Um die Variabilität innerhalb einer Probe oder Bündeln von Proben im Ausmaß der Bindungsstärke oder den Betrag der zugegebenen DNA zu berücksichtigen, müssen die Rohwerte der Hybridisierungsintensitäten normalisiert werden. Erst dann können sinnvolle Vergleiche zwischen den Proben vorgenommen werden. Die Normalisierung ist nötig, da kürzlich entwickelte Algorithmen beim Callen der Genotypen typischerweise Strategien der Clusterbildung einsetzen, welche Informationen über verschiedene Stichproben für jeden SNP einsetzen und Genotypzuweisungen dann für das gesamte Cluster vornehmen (Teo 2008). Dies steht im Unterschied zu früheren Strategien, welche die Intensitäten auf jede individuelle Probe bei jedem SNP detektierten, und von denen daher angenommen wurde, bevorzugt den heterozygoten Genotyp zuzuweisen. Das Poolen der Information über alle Individuen für die Genotypzuweisung erhöht die Glaubwürdigkeit des Calls, da von Proben mit ähnlichen Intensitätsprofilen auch erwartet wird, denselben Genotyp zu tragen. Allerdings kann die Genauigkeit bei dieser Strategie von der Anzahl der Proben innerhalb eines Clusters abhängig sein und das Callen für kleine Stichproben damit problematisch werden. Ein Überblick über einige mögliche Algorithmen wird anderweitig gegeben (Teo 2008).

Abhängig vom Algorithmus entstehen dann die Genotypinformationen, die in Form sogenannter „cluster plots“ visualisiert werden können. Abbildung 23 zeigt eine solche visuelle Präsentation der Signale für den Genotyp innerhalb einer Gruppe von Individuen bezüglich eines SNPs. Je nach Stärke der für die beiden Allele stehenden Farbintensitäten können nun Genotypen zugewiesen werden.

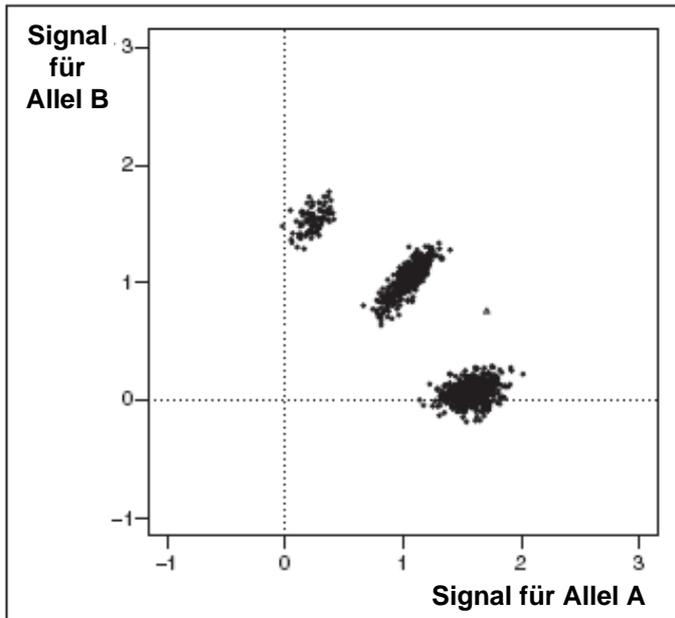


Abbildung 23: Darstellung eines „cluster plots“

Die horizontale und vertikale Achse repräsentieren die normalisierten Signale der Hybridisierung für die beiden möglichen Allele eines SNPs, benannt als Allel A und B. Jeder Punkt stellt das Signalprofil einer Probe dar. Dabei sind entsprechend einiger Normalisierungsmethoden auch negative Signale möglich. Das Cluster von Punkten, die ein hohes Signal für Allel A und kein Signal für Allel B haben, wird als das Cluster mit dem Genotyp AA angenommen. Umgekehrt sollte das Cluster, welches den Genotyp BB repräsentiert, ein hohes Signal für B und keines für A aufweisen. Das Cluster von Punkten mit nahezu ähnlichen Signalprofilen für A und B entspricht dem heterozygoten Genotyp.

Die Probe, welche hier zwischen zwei Clustern liegt und durch ein kleines Dreieck abgebildet ist, zeigt ein Profil, welches sich nicht unzweifelhaft einem der Cluster zuordnen lässt. Gute Algorithmen sollten solchen Proben einen fehlenden Wert zuweisen (nach Teo (Teo 2008)).

Die Analyse solcher „cluster plots“ gehört zur Qualitätskontrolle. Drei verschiedene Punktwolken sollten möglichst eindeutig die drei verschiedenen Genotypen widerspiegeln. Häufig ist eine klare Trennung aber nur ein Idealfall und es bedarf sehr genauer Betrachtungen der Ursachen, wenn sich eine solche Trennung nicht vornehmen lässt. Bei der Beurteilung der Genotypisierungsqualität ist daher neben den üblichen Qualitätskontrollen die Betrachtung der „cluster plots“ sehr aufschlussreich (Chanock et al. 2007). Verschiedene Konstellationen zeigen dabei gewisse Ungenauigkeiten der Genotypisierung.

2.3.2 MALDITOF

Eine andere Genotypisierungsmethode, bei der eine spezielle Einzel-SNP-Auswahl möglich ist, ist die „Matrix Assisted Laser Desorption/Ionisation Time of Flight“ (MALDITOF)-Massenspektrometrie (MS). Hierbei können auch mehrere SNPs gleichzeitig, maximal derzeit aber etwa nur im zweistelligen Bereich analysiert werden.

Im Rahmen dieser Methode ionisiert man die DNA-Probe mittels Laserbeschuss über eine Matrix, in welche die eigentliche Probe eingebettet ist. Dazu wird die zu untersuchende Probe mit dem hundert- bis tausendfachen Überschuss an Matrix versetzt und auf einem metalli-

schen Träger kokristallisiert. Anschließend wird dieser Komplex im Hochvakuum eines Massenspektrometers einem intensiven Laserimpuls von wenigen Nanosekunden Dauer ausgesetzt (Karas et al. 1988). Durch den Laserbeschuss verdampft die Matrix und es findet gleichzeitig eine Ionisierung der zu analysierenden Einheiten statt. Die entstandenen Ionen werden in einem elektrischen Feld beschleunigt und ein Ionendetektor wandelt die ankommenden Ionen in ein elektrisches Signal um. Die Flugzeit („time of flight“) zeigt sich in Abhängigkeit von Masse und Ladungszahl. Definierte Flugzeiten charakterisieren dann unterschiedliche Genotypen. Die am weitesten für die Genotypisierung von Einzelbasenaustausch optimierte, auf MALDI-TOF MS basierende Methode ist die Primer-Extensions-Reaktion. Hier entstehen durch eine Primer-Verlängerung SNP-spezifische Primer-Extensionsprodukte, die sich aufgrund ihres unterschiedlichen Gewichtes massenspektrometrisch wie beschrieben unterscheiden lassen (Griffin et al. 2000).

2.3.3 Generelle Qualitätskontrollen

Eine genaue Qualitätskontrolle aller SNPs ist in jeder Studie unabdingbar. Dazu gehören die Bestimmung der Anzahl der Personen, für die ein SNP gemessen werden kann („**Callrate**“) sowie die Bestimmung von Allelfrequenz und Konkordanzraten der bestimmten Genotypen bei Mehrfachtypisierungen, die möglichst über 99,5% liegen sollte. Mehrfachtypisierungen können für jeden SNP vorgenommen werden, was jedoch zu einer enormen Kostensteigerung führt. Im Rahmen einer GWA sollte stets der am stärksten assoziierte SNP mit einer anderen Methode typisiert werden. Eine weitere sinnvolle Qualitätskontrolle für die Richtigkeit der Genotypisierung eines SNPs liefert ein Test auf Abweichung vom HWE. Wird dieser signifikant, liegt häufig ein Problem vor. Neben dem Ausschluss von SNPs mit geringer Qualität sollte auch ein Ausschluss von Probanden mit vielen Genotypen von geringer Qualität oder kontaminierter DNA vorgenommen werden.

2.4 Für die Auswertungen verwendete Stichproben

Basis der Auswertungen für die Studien dieser Arbeit waren vor allem die deutschen Kohorten der „International Study of Asthma and Allergies in Childhood“ II (ISAAC II)-Studie (Weiland et al. 1999) aus München und Dresden. Daneben gingen Kinder aus einer Studie mit Leipziger Kindern in die Analysen ein (von Mutius, E. et al. 1998; Weiland et al. 1999).

Von sämtlichen 7940 Kindern aus München (M), Dresden (D) oder Leipzig (L), die rekrutiert worden waren, sollten bei den genetischen Analysen in den Querschnittpopulationen nur diejenigen einbezogen werden, für welche DNA vorlag und die Variable für Gesamt-IgE-Werte angegeben war. Zudem wurden sämtliche genetischen Analysen auf Kinder deutscher Herkunft eingeschränkt. Die Querschnittpopulationen der genetischen Analyse sind daher gegenüber der Ursprungspopulation auf 4264 Kinder reduziert (Abbildung 24). Diese Population wurde als DLM4264 bezeichnet.

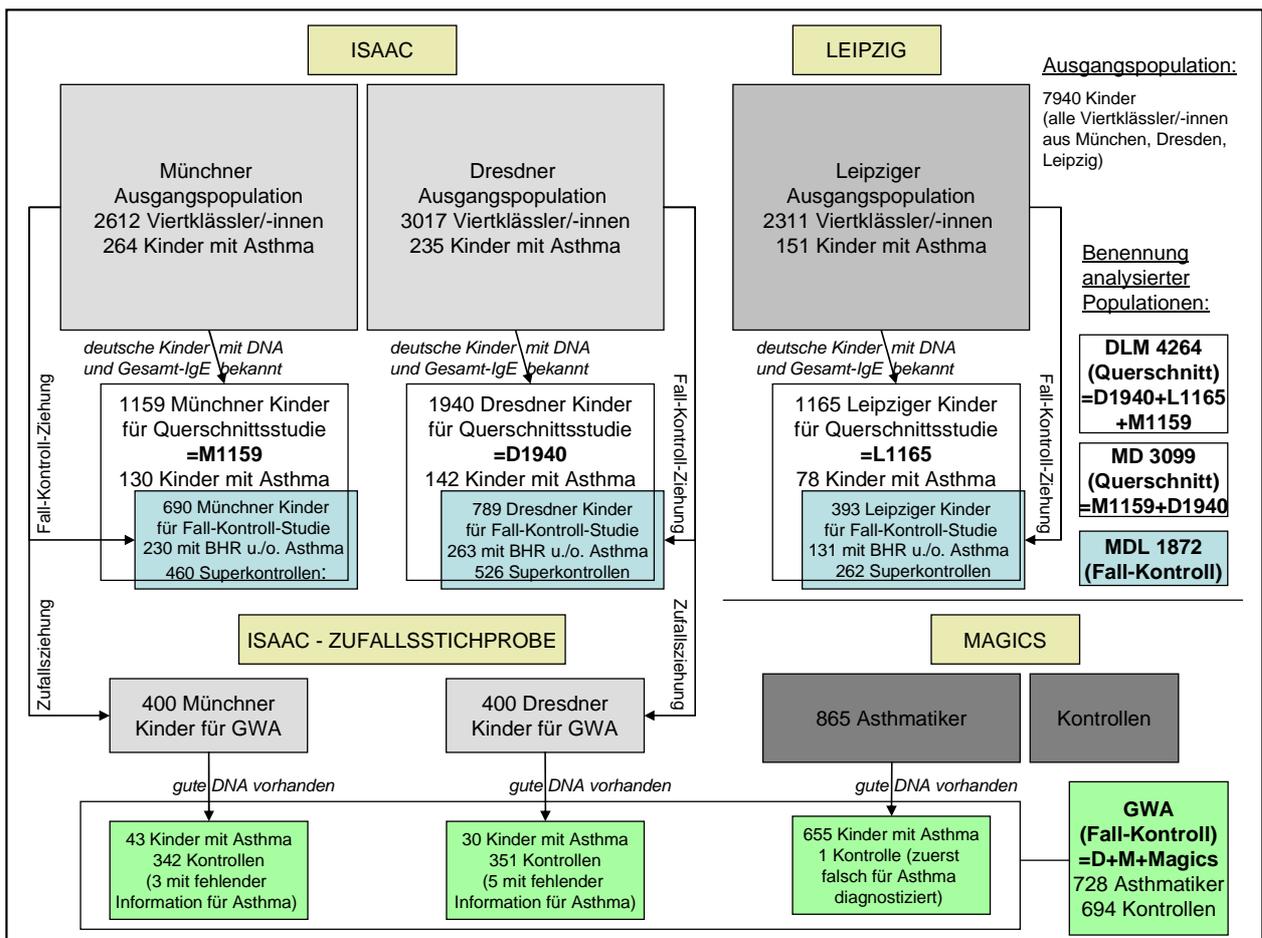


Abbildung 24: Populationen, auf denen die Auswertungen dieser Arbeit basieren

Auch die Teilpopulationen der einzelnen Zentren (M1159, D1940 oder L1165) wurden in Studien der eigenen Arbeitsgruppe herangezogen. Daneben wurden zahlreiche Studien in der Querschnittspopulation durchgeführt, welche nur Münchner und Dresdner Kinder einbeziehen (MD3099).

Neben den Populationen für die Querschnittsuntersuchungen gab es zwei weitere eingebettete Studiendesigns, die aus der gleichen Grundpopulation abgeleitet wurden.

Das erste Studiendesign ist ein eingebettetes Fall-Kontroll-Design, das sich aus Kindern aus München, Dresden und Leipzig zusammen setzt (siehe z.B. (Kormann et al. 2008)). Dabei wurden alle Kinder mit einer Diagnose für Asthma oder BHR (oder beides) als Fälle herangezogen. Diese 624 Fälle wurden in einem Verhältnis 1:2 pro Stadt mit sogenannten Superkontrollen ergänzt. Superkontrollen sind hier Kinder, für welche gleich mehrere Krankheiten oder Risikofaktoren zur Entwicklung atopischer Erkrankungen auszuschließen sind, und werden im entsprechenden Abschnitt genau definiert (2.4.2). Diese Population von Fällen und Superkontrollen wird als MDL1872 bezeichnet. Sie wurde in Zusammenhang mit Auswertungen für *TLRs* und *ADAM33* verwendet.

In einem weiteren Studiendesign wurde eine geschichtete Zufallsstichprobe von je 400 Dresdner und Münchner Kindern gezogen, die zusammen mit Asthmatikern einer anderen Studie, der „Multicentre Asthma Genetics in Childhood Study“ (MAGICS), die relevante Stichprobe deutscher Kinder für die GWA bildete (2.4.3).

Die drei Teilstichproben (Querschnittspopulationen DLM4264, Fall-Kontroll-Population MDL1872 und GWA-Teilstichprobe) wurden zu unterschiedlichen Zeitpunkten unabhängig voneinander aus der Ausgangspopulation gezogen. Daher sind die Ziehungen nicht komplett in den Querschnittspopulationen enthalten. Dennoch gibt es natürlich eine große Überschneidung. Chronologisch wurde erst MDL1872, dann DLM4264 und dann die GWA-Teilstichprobe erstellt.

Dass nicht alle Münchner und Dresdner Kinder aus der Fall-Kontroll-Studie Teil der Querschnittspopulation sind, liegt daran, dass die Fall-Kontroll-Studie zuerst durchgeführt wurde. Dadurch war es möglich, dass Kinder, für die keine ausreichende Menge an DNA vorlag, zwar in MDL1872 analysiert werden konnten, aber aus den Querschnittspopulationen rausfallen. Die Teilstichprobe von ISAAC-Kindern für die GWA ist ebenfalls nicht komplett Teil der Querschnittspopulation DLM4264, da für diese Ziehung der Zufallsstichprobe die Variable IgE keine Rolle spielte und zudem Kinder mit österreichischer Nationalität zugelassen waren.

Daneben gab es bei der Studie, die Assoziationen von Asthma mit *ADAM33* untersuchte, Auswertungen mit Kindern aus der Kohortenstudie „Multicenter Allergy Study“ (MAS), wobei hier vorwiegend auf die Daten der Kinder im Alter von zehn Jahren zurückgegriffen wurde (Schedel et al. 2006), so dass die Auswertungsmethoden denjenigen einer Querschnittsstudie entsprachen. Nachfolgend sollen die Populationen der Kinder, die in die genetischen Analysen eingingen, kurz dargestellt werden.

2.4.1 Querschnittsstudien (ISAAC II und Leipzig)

Im Rahmen der „International Study of Asthma and Allergies in Childhood“ Phase II (ISAAC II), einer internationalen Vergleichsstudie zur Erforschung von Asthma und Allergie, wurden 1995/96 Viertklässler im Alter von 9-11 Jahren in zwei deutschen Städten, München und Dresden, rekrutiert (Weiland et al. 1999). Entsprechend einem der ISAAC II-Studie sehr ähnlichen Protokoll wurden in denselben Jahren auch in Leipzig Schulkinder der vierten Klasse für eine weitere Studie im Bereich Asthmaforschung rekrutiert und untersucht (von Mutius, E. et al. 1998). Die Prävalenzen wichtiger Phänotypen unterscheiden sich aber kaum zwischen der genotypisierten Teilstichprobe und der gesamten Studienpopulation. Dies gilt sowohl für die aus drei Zentren gepoolte Population (Tabelle 13) als auch für die einzelnen Studienzentren (Tabelle 14).

	Gesamte Studienpopulation		Genotypisierte Studienpopulation	
	N=7940		N=4264	
	n ¹	[%]	n ¹	[%]
Buben	3879/7655	50,67%	2151/4257	50,53%
Alter ²	9,7 (0,60)		9,7 (0,60)	
Asthma	650/7750	8,39%	350/4185	8,36%
BHR	462/3821	12,09%	300/2468	12,16%
Heuschnupfen	639/7743	8,05%	336/4174	8,05%
Atopische Dermatitis ³	923/5417	17,04%	540/2994	18,04%
Ekzem ⁴	1200/7349	16,33%	679/3960	17,15%
Atopie (Pricktest)	1474/5866	25,13%	1056/4110	25,69%

Tabelle 13: Deskriptive Charakteristik wichtiger Phänotypen für alle deutschen Kinder in der gesamten (N=7940) und genotypisierten (N=4264) Studienpopulation

¹ n betroffene / n Personen, für die Daten verfügbar

² präsentiert als Mittelwert und Standardabweichung (\pm SD)

³ nur Dresdner und Münchner Kinder

⁴ Ekzem wurde als eine Kombination zweier Varianten zusammengefasst: Bericht von „itchy dermal changes“ und „ever eczema“ in Leipzig und Dr. Diagnose für atopische Dermatitis in München oder Dresden.

<u>DRESDEN</u>	Gesamte Studienpopulation		Genotypisierte Studienpopulation	
	N=3017		N=1940	
	n¹	[%]	n¹	[%]
Buben	1565/3016	51,89%	994/1940	51,24%
Alter²	9,72 (0,60)		9,72 (0,60)	
Asthma	235/2961	7,94%	142/1906	7,45%
BHR	164/1042	15,74%	128/823	15,55%
Heuschnupfen	289/2958	9,77%	183/1903	9,62%
Ekzem³	482/2902	16,61%	327/1871	17,48%
Atopie (Pricktest)	576/2242	25,69%	493/1859	26,52%
<u>LEIPZIG</u>	Gesamte Studienpopulation		Genotypisierte Studienpopulation	
	N=2017		N=1165	
	n¹	[%]	n¹	[%]
Buben	1026/2028	50,59%	590/1158	50,95%
Alter²	9,87 (0,61)		9,89 (0,60)	
Asthma	151/2219	6,80%	78/1131	6,90%
BHR	118/1875	6,29%	68/1129	6,02%
Heuschnupfen	115/2252	5,11%	56/1142	4,90%
Ekzem³	277/1932	14,34%	139/966	14,39%
Atopie (Pricktest)	473/1812	26,10%	279/1108	25,18%
<u>MÜNCHEN</u>	Gesamte Studienpopulation		Genotypisierte Studienpopulation	
	N=2612		N=1159	
	n¹	[%]	n¹	[%]
Buben	1288/2611	49,33%	567/1159	48,92%
Alter²	9,49 (0,56)		9,50 (0,55)	
Asthma	264/2570	10,27%	130/1148	11,32%
BHR	180/904	19,91%	104/516	20,16%
Heuschnupfen	235/2533	9,28%	97/1129	8,59%
Ekzem³	441/2515	17,53%	213/1123	18,97%
Atopie (Pricktest)	425/1812	23,45%	284/1143	24,85%

Tabelle 14: Deskriptive Charakteristik wichtiger Phänotypen für alle deutschen Kinder in der Studienpopulation nach Städten geschichtet

¹ n betroffene / n Personen, für die Daten verfügbar

² präsentiert als Mittelwert und Standardabweichung (\pm SD)

³ Ekzem wurde als eine Kombination zweier Varianten zusammengefasst: Bericht von „itchy dermal changes“ und „ever eczema“ in Leipzig und Dr. Diagnose für atopische Dermatitis in München oder Dresden.

2.4.2 Fall-Kontroll-Studie MDL1872

Die Fall-Kontroll-Population, bestehend aus insgesamt 1872 deutschen Kindern (München n=690, Dresden n=789, Leipzig n=393) wurde wie folgt erstellt: Alle Kinder mit der Arzt Diagnose Asthma und/oder bronchialer Hyperreaktivität wurden als Fälle definiert und in einem Verhältnis von 1:2 mit Superkontrollen untersucht. Es handelt sich somit um eine eingebettete Fall-Kontroll-Studie.

Es ist wichtig zu sehen, dass die reine Asthmadefinition hier nicht gleich der Falldefinition ist. Für Analysen, welche sich auf Asthma beziehen, wurden nur die Kinder als Fälle eingeschlossen, welche die Arzt Diagnose Asthma erhalten hatten und somit Kinder, die BHR, aber kein Asthma hatten, ausgeschlossen. Als Kontrollgruppe diente jeweils die Gruppe der Superkontrollen.

Superkontrollen sollten als nicht-asthmatische, nicht-atopische Kinder ohne erhöhte IgE-Werte (Gesamt-IgE < 50 IU/ml) und ohne BHR definiert werden (Tabelle 15).

	Exakte Definition der Fälle		Exakte Definition der Kontrollen („Superkontrollen“)	gezogene aus möglichen	
München (N=690)	Münchener Fälle	230	Münchener Superkontrollen	460	
	Asthma+BHR	21	Asthma negativ + Hauptpricktest negativ + <IGE <= 50 + BHR negativ	174	174
	nur Asthma	117	Asthma negativ + Hauptpricktest negativ + <IGE <= 50 + BHR fehlend	286	296
	nur BHR	92			
Dresden (N=789)	Dresdner Fälle	263	Dresdner Superkontrollen	526	
	Asthma+BHR	25	Asthma negativ + Hauptpricktest negativ + <IGE <= 50 + BHR negativ	276	276
	nur Asthma	128	Asthma negativ + Hauptpricktest negativ + <IGE <= 50 + BHR fehlend	250	383
	nur BHR	110			
Leipzig (N=363)	Leipziger Fälle	131	Leipziger Superkontrollen	262	
	Asthma+BHR	15	Asthma negativ + Hauptpricktest negativ + <IGE <= 50 + BHR negativ	262	377
	nur Asthma	63			
	nur BHR	53			
GESAMT	alle Fälle	624	alle Superkontrollen	1248	
	alle Fälle mit BHR	316			
	alle Fälle mit Asthma	369			

Tabelle 15: Verteilung von Fällen und Kontrollen in der Fall-Kontroll-Studie MDL1872

Da nicht für alle Kinder die Variable BHR vorlag, wurde ein zweistufiges Vorgehen gewählt.

Zuerst wurde aus allen Kindern, welche die vier genannten Kriterien erfüllten, städteweise eine Zufallsstichprobe herangezogen, um das Verhältnis 1:2 zu bekommen. In München und Dresden gab es nicht genügend Kinder, welche alle vier Kriterien erfüllten, um genügend Superkontrollen zu erhalten. Daher wurden in einem zweiten Schritt Kinder zufällig aus allen möglichen Kindern ausgewählt, bei denen zwar die BHR-Diagnose unbekannt war, die aber die drei verbleibenden Kriterien erfüllten (nicht-asthmatische, nicht-atopische Kinder ohne erhöhte IgE-Werte). Aus diesen wurde wieder eine Zufallsstichprobe in München und Dresden gezogen. Auf diese Weise konnte das Verhältnis 1:2 zwischen Fällen und Kontrollen städteweise erhalten werden.

Für die Auswertungen auf der Variablen Asthma wurden später die Fälle auf diejenigen mit Asthma und BHR oder nur Asthma reduziert (N=369, München n=138, Dresden n=153, Leipzig n=78).

2.4.3 Stichprobe für Genomweite Assoziationsstudie

Den Großteil der Asthmatiker in der GWA bildeten Kinder der MAGIC-Studie. MAGIC ist eine multizentrische Studie, welche im Zeitraum von 2001 und 2007 mit dem Ziel der Identifikation von für die Manifestation atopischer Erkrankungen relevanten Polymorphismen in Kandidatengenomen durchgeführt wurde. Hierzu wurden in sieben verschiedenen Zentren aus Österreich oder Deutschland Kinder mit Asthma rekrutiert (in fünf Zentren wurden zusätzlich auch nicht-asthmatische Patienten der Klinik als Kontrollen rekrutiert).

Im finalen MAGIC-Datensatz waren 1255 Kinder eingeschlossen, von welchen Information vorlag, 865 Asthmatiker und 390 Kontrollen. Die Information setzte sich zusammen aus Elternfragebogen, Information aus ärztlichen Untersuchungen der Asthmatiker oder Kontrollen sowie Informationen aus Bluttests. Ein ausgefüllter Elternfragebogen lag zum Zeitpunkt der vorliegenden Arbeit nur für 1205 dieser Kinder (845 Asthmatiker) vor. Tabelle 16 zeigt, wie sich diese über die Studienzentren verteilen.

Für die genomweite Assoziationsstudie wurden die Proben, die zum Zeitpunkt der Durchführung der ersten GWA zu Asthma vorlagen, somit von 753³ Kindern, zur Genotypisierung ans „Centre National de Génotypage“ (CNG) in Paris geschickt. Von diesen passierten 656 die Qualitätskontrolle und gingen in die Auswertung ein⁴.

³ 8 davon sind aus verschiedenen Gründen nicht mehr im MAGIC-Datensatz (meist fehlende Informationen).

⁴ Ein Kind von diesen 656 wurde nachträglich als Kontrolle identifiziert und entsprechend ausgewertet.

		zur Genotypisierung ans Centre National de Génotypage nach Paris	ausgewertet
MAGICS	mit Fragebogen	(in MAGICS nur Asthmatiker)	in GWA
Bochum	222	135	120
Feldkirch	98	16	8
Freiburg	154	128	101
Köln	65	26	26
München	327	183	174
Wesel	312	243	205
Wien	27	22	22
ISAAC II			
München		400 (45 Asthma)	388 (43 Asthma)
Dresden		400 (30 Asthma)	386 (30 Asthma)

Tabelle 16: Anzahl der (deutschen) Kinder, die an der GWA teilnahmen
Dargestellt sind Zahlen der Kinder, deren DNA ans Centre National de Génotypage zur Genotypisierung geschickt und derer, die später ausgewertet wurden

Neben den Kindern aus MAGICS wurde eine stratifizierte Zufallsstichprobe von 800 Kindern aus ISAAC (jeweils 400 Kinder aus München und Dresden) gezogen, um die GWA-Stichprobe zu komplettieren. Davon gingen später 774 in die Auswertungen ein.

Um auszuschließen, dass die gezogenen Zufallsstichprobe für die GWA gegenüber der Grundgesamtheit aller Kinder, für die DNA vorlag, verzerrt sein könnte, wurde für wichtige Variablen auf eine solche Abweichung getestet. Alter, Geschlecht sowie asthmarelevante Phänotypen (Asthma, Heuschnupfen, Neurodermitis und Atopie) zeigten dabei keine signifikanten Unterschiede, so dass Verzerrungen durch die Stichprobenziehung in diesen Variablen ausgeschlossen werden können.

Der Anteil von deutschen Kindern in der GWA bestand aus 728 Asthmatikern und 694 Kontrollen⁵⁶. Dieser wurde um 266 asthmatische und 549 nicht-asthmatische kaukasische Kinder der UK ergänzt, um die Power der Fall-Kontroll-Untersuchungen zu erhöhen.

Generell sind in der ersten GWA zu Asthma auch Personen aus UK eingeschlossen. Diese Personen umfassen eine Stichprobe von Familien (MRC-A) und eine Stichprobe von Kontrollkindern (UK-C). Beide Gruppen entstammen der MRC UK „National family collection“. Sämtliche asthmatischen Kinder der UK stammen aus dem MRC-A-Panel. Die Kinder der Assoziationsstudie, die als Kontrollen verwendet wurden, setzen sich hingegen zusammen aus

⁵ Da 8 von den Kindern aus ISAAC 2 die Asthmavariablen fehlte, addiert sich das nur zu 1422 statt 1430 wie in der Tabelle.

⁶ In späteren Untersuchungen können sich diese Zahlen auf Grund von ergänzenden Qualitätskontrollen, unzureichender DNA leicht ändern.

437 Kindern aus dem UK-C-Panel sowie 112 Kindern aus dem MRC-A-Panel (siehe auch Abbildung 25).

Wie eingangs erwähnt besteht zwischen einem Teil der deutschen Stichprobe, die in der GWA verwendet wurde, und der Querschnittspopulation aus ISAAC eine Überschneidung. Daher ist darauf hinzuweisen, dass für Replikationsuntersuchungen, die Ergebnisse der GWA-Studie in ISAAC wieder zu finden versuchen, stets nur die Kinder analysiert wurden, die nicht Bestandteil der GWA-Stichprobe sind. Dies sind in München und Dresden 2359 Kinder (86 asthmatische Kinder aus München und 114 asthmatische Kinder aus Dresden).

2.4.4 MAS

Die „Multicenter Allergy Study“ (MAS) ist eine prospektive Beobachtungsstudie, die 1990 begann, um den prädiktiven Wert verschiedener klinischer und immunologischer Parameter sowie die Bedeutung früher Allergenexposition und des Einflusses anderer umweltbedingter Einflussfaktoren für die Entwicklung atopischer Erkrankungen bei Kindern zu untersuchen (Bergmann et al. 1994).

Hierbei wurden unter der Federführung von Professor Wahn, Charité Berlin, anfänglich 1314 Kinder aus fünf verschiedenen Zentren von der Geburt an beobachtet (Illi et al. 2004). Die Zentren waren Berlin, Düsseldorf, Freiburg, Mainz und München. Jährliche „follow ups“ wurden regelmäßig bis zum Alter von 13 Jahren durchgeführt und schlossen dabei standardisierte Interviews und körperliche Untersuchungen ein. Die Studie ist anderweitig genau beschrieben (z.B. (Kulig et al. 1999; Liu et al. 2003)).

Um genetische Analysen durchführen zu können, wurden die Eltern aller Kinder später explizit um Zustimmung gebeten, von den Kindern auch DNA entnehmen zu dürfen. Da nicht alle damit einverstanden waren, reduzierte sich die Anzahl der Kinder, für die genetische Analysen durchgeführt werden konnten, auf 888 Kinder. Diese Kinder, für die Genotypisierungsdaten vorliegen, wurden in der Studie zu *ADAM33* auf die 824 Kinder beschränkt, deren Nationalität Deutsch war (Schedel et al. 2006).

2.5 Definition der relevanten Zielvariablen

Die wesentliche Zielgröße jeder Studie war Asthma. Dies war in der Population der MAGICS-Kinder nicht weiter zu bestimmen, da nur Asthmatiker aus verschiedenen Kliniken in die Studie der GWA einbezogen wurden. Kinder aus der ISAAC-Population sowie aus Leipzig wurden über einen Elternfragebogen als Asthmatiker klassifiziert. Gaben die Eltern an, dass bei ihrem Kind durch einen Arzt Asthma oder wiederholt eine spastische, asthmatische oder obstruktive Bronchitis diagnostiziert wurde, so wurde dies als Asthma gewertet. In der MAS-Studie wurde Asthma entsprechend den ISAAC-Kern-Fragen so abgeleitet, dass sie den Fragen der ISAAC-Kinder entsprach.

Analysen, welche die Atopie betreffen, wurden bislang nicht innerhalb der GWA durchgeführt. Bei den Kindern aus Dresden, Leipzig und München wurde Atopie mittels Pricktest gemessen. Beim Prick-Test wird ein definierter Allergenextrakt auf die Haut aufgetropft. Anschließend wird die Haut mit einer Lanzette leicht angestochen, so dass die jeweiligen Substanzen in die Oberhaut eindringen. Die Testreaktion kann nach 20 Minuten im Vergleich zu zwei immer mitgeführten Kontrollen (Positivkontrolle mit Histamin und wirkstofffreier Negativkontrolle) abgelesen werden. Beurteilt werden die Hautrötung und die Quaddelgröße. Zeigte sich in einem Pricktest mit einem Allergenextrakt nach Abzug von der Negativkontrolle eine Quaddelgröße von mindestens 3mm, galt der Prick-Test als positiv. Ein Kind wurde als atopisch bezeichnet, wenn eine positive Reaktion auf mindestens eines von sechs inhalativen Allergenen (nämlich Hausstaubmilben *Dermatophagoides pteronyssinus* oder *farinae*, Schimmelpilz *Alternaria tenuis*, Katzenhaare, Gräser- und Baumpollenmischungen) im Hautallergietest auftrat. Die Kinder aus Leipzig wurden ebenfalls auf eine Reaktion auf sechs Allergene getestet (*Dermatophagoides pteronyssinus*, Gräser, Birke und Haselnusspollen sowie Hund und Katze) (von Mutius, E. et al. 1998). Eine atopische Sensibilisierung innerhalb der MAS-Kinder wurde als eine Serumkonzentration von mindestens 0,35 kU/l des jeweiligen spezifischen IgE bezeichnet. Ein Kind galt als atopisch, wenn es zum entsprechenden Zeitpunkt auf mindestens eines der getesteten Allergene sensibilisiert war.

Als Kinder mit atopischem Asthma wurden diejenigen definiert, welche sowohl Asthma als auch atopische Sensibilisierung aufwiesen. Kinder mit nicht-atopischem Asthma sind hingegen diejenigen, für welche zwar eine Asthmadignose, aber keine atopische Sensibilisierung vorlag. Als Referenzgruppe für atopisches und nicht-atopisches Asthma wurden in der Regel Kinder ohne Asthma herangezogen - und zwar unabhängig vom Status der Atopie. Bei den Analysen zu *TLRs* wurde auch die Kontrollgruppe von Kindern ohne Atopie und ohne Asth-

ma herangezogen, da sich nur diese Variante im Falle der Superkontrollen definieren ließ. Analysen in Querschnittsstudien belegten aber, dass die Definition der Kontrollgruppe nur zu minimalen Unterschieden bei den Auswertungen führt.

Die Lungenfunktion wurde in der ISAAC-Population an einer 50%igen randomisierten Teilstichprobe mittels Spirometrie mit Bestimmung der Fluss-Volumen-Kurve durch MasterScope Version 4.1 (Jäger, Würzburg, Deutschland) entsprechend den Kriterien der American Thoracic Society für die Durchführung von reproduzierbaren und zufriedenstellenden Spirogrammen (1987) gemessen (von Ehrenstein et al. 2002). Ein Minimum von zwei Spirogrammen wurde erhoben und dabei das höchste zweier reproduzierbarer Messungen des „forced expiratory volume in 1 second“ (FEV_1) als Ausgangswert benutzt.

BHR wurde als Veränderung im FEV_1 nach Inhalation einer nebulisierten, hyperosmolaren Kochsalzlösung (4,5%) bestimmt. Kinder mit einem Lungenfunktionsabfall von 15% in FEV_1 nach Provokation wurden als BHR-positiv diagnostiziert (Weiland et al. 1999).

In Leipzig wurde Lungenfunktion nach Kaltluftprovokation unter Verwendung entsprechender Protokolle gemessen (von Mutius, E. et al. 1998). Das 95. Perzentil der Verteilung einer Referenzpopulation wurde bestimmt und korrespondierte mit einem 9%igen Abfall der FEV_1 . Entsprechend wurde ein FEV_1 -Abfall von 9% in Leipzig als BHR-positiv definiert (Nicolai et al. 1993).

BHR in MAS wurde nur zum Zeitpunkt bestimmt, als die Kinder sieben Jahre alt waren. Dabei wurden Provokationen ansteigender Histaminkonzentrationen (gewöhnlich von 0,5mg/ml bis 0,8 mg/ml) entsprechend Standardprozeduren vorgenommen. Das 90. Perzentil der Verteilung von der Provokationskonzentration, die einen 20%igen Abfall der FEV_1 ($PC_{20}FEV_1$) in einer gesunden Teilstichprobe bewirkte, betrug 0,85mg/ml (Niggemann et al. 2001). Bronchiale Hyperreaktivität wurde als eine $PC_{20}FEV_1$ definiert, die über diesem Wert lag.

2.6 *Verwendete Methoden*

2.6.1 *Population und Typisierung*

Die GWA-Population, die Kombination von Teilstichproben deutscher Kinder aus zwei Studien (ISAAC und MAGICS), wurde für die erste genomweite GWA zu Asthma mit einer britischen Kohorte kombiniert. Eine Population von Familien aus UK wurde in der GWA zusätzlich untersucht.

Die Population für die GWA-Replikation entstammt Kindern von München und Dresden aus ISAAC. Dies ist somit die MD3099-Populationen ohne Kinder, die in der GWA analysiert wurden (siehe Abbildung 24, Seite 78). Analysen zum Kopplungsungleichgewicht wurden auch in der gesamten MD3099-Population durchgeführt. Vertiefende Analysen zu Befunden, die aus der GWA stammen, wurden auch in der gesamten Querschnittspopulation DLM4264 durchgeführt.

Für die Auswertungen zu *ADAM33* wurden die Kinder der eingebetteten Fall-Kontroll-Studie MDL1872 sowie die Kinder der MAS-Studie herangezogen. Auch bei der Kandidatengenstudie zu den *TLRs* wurden zuerst die Kinder von MDL1872 analysiert, wonach einige vertiefende Analysen in der Querschnittspopulation MD3099 folgten.

Um die unabhängigen Variablen, die bei genetischen Assoziationsstudien der vorliegenden Arbeit immer die SNPs darstellen, zu erhalten, wurden die entsprechenden Technologien der Genotypisierung genutzt. Während bei der GWA die Chip-Technologie zum Einsatz kam, um die Genotypen zu bestimmen, wurde bei den anderen Studien MALDI-TOF Massenspektrometrie verwendet. Auch die Genotypisierungen für die Replikationen der GWA-Ergebnisse wurden mittels MALDI-TOF Massenspektrometrie bestimmt.

2.6.2 *Statistische Methoden*

Mit einem Chi-Quadrat-Test wurde für jeden SNP getestet, ob er Abweichungen vom HWE zeigte. Bei stärkeren Abweichungen wurde die Richtigkeit der Genotypisierung in Teilstichproben mit einer anderen Genotypisierungsmethode überprüft. Zur Bestimmung der LD-Koeffizienten wurde R^2 als LD-Maß zugrunde gelegt und das Programm Haploview zur Berechnung paarweiser LD-Maße verwendet (Barrett et al. 2005). Sowohl bei der GWA als auch bei der Studie, welche die *ORMDL*-Gene untersuchte, wurde für die Assoziationstests ein lineares Modell zugrunde gelegt. Für die Berechnung der Assoziationen der SNPs von *ADAM33* und den *TLRs* mit den entsprechenden Phänotypen wurde ein dominantes Modell zugrunde gelegt. Die Effekte wurden mit Methoden der logistischen Regression bestimmt.

Haplotypen wurden mit einem EM-Algorithmus geschätzt (Excoffier et al. 1995) und Haplotypenanalysen wurden als globale Tests mit Cocaphase (Dudbridge 2003) oder als Tests für individuelle Haplotypen durchgeführt. Zudem wurden Haplotypeneffekte bei den *TLRs* und *ORMDLs* mit Haplotyp-trendregression berechnet (Zaykin et al. 2002). Zur Berechnung von Unterschieden in Parametern der Lungenfunktion zwischen den Genotypen wurden Varianzanalysen oder t-Tests verwendet. Lungenfunktionsparameter gingen als Prozent von Sollwerten in die Analysen ein.

Wenn nicht explizit vermerkt, sind die Effekte (meist Odds Ratios) und p-Werte für nicht adjustierte Analysen berichtet. Eine Adjustierung für Kovariaten wurde in der Regel nicht durchgeführt. Eine Adjustierung für multiples Testen wurde an verschiedenen Stellen ergänzt. Dabei wurde als Standardprozedur eine Bonferronikorrektur über die Anzahl der analysierten SNPs oder analysierten Haplotypen herangezogen. Als Korrekturverfahren für multiples Testen diente bei der GWA die Berechnung einer FDR. Dieses Verfahren kam auch bei dem *TLR*-Projekt zum Einsatz (Benjamini et al. 1995). Zudem wurde im Rahmen der GWA mittels einer speziellen Methode der logistischen Regression (Setakis et al. 2006) auf Stratifikation geprüft. Bei diesem Korrekturverfahren wurden 102 zufällig ausgewählte SNPs als Kovariaten in eine schrittweise logistische Regression einbezogen. Soweit nicht anders vermerkt wurden sämtliche Analysen mit dem Programm SAS oder dessen Zusatzmodul SAS/Genetics durchgeführt.

3 Ergebnisse

3.1 ORMDL3 als Vertreter der GWA

3.1.1 Ergebnisse der GWA („Hits“)

Im Rahmen der ersten genomweiten Assoziationsstudie für Asthma wurde ein neuer möglicher genetischer Krankheitslocus für Asthma bei Kindern auf Chromosom 17q21 gefunden (Moffatt et al. 2007).

Das Studiendesign zielte darauf ab, bereits intern positive Replikationen potentiell positiver Ergebnisse durch Familien- und Fall-Referenz-Panels der Versuchspersonen zu bekommen. Abbildung 25 zeigt das genaue Studiendesign.

	aus UK		Deutsch		
	MRC-A		MAGICS/ ISAAC	Gesamt	
	Eltern	Kinder	Kinder		
ohne Asthma		112	437	694	1243
mit Asthma		266	0	728	994
Gesamt	405	378	437	1422	

primäre Assoziationsstudie
 Familienstudie
 Genexpressionsmessungen

Abbildung 25: Studiendesign der GWA

Durchgeführt wurde eine Assoziationsstudie mit deutschen und britischen Kindern (primäre Assoziationsstudie), eine Familienstudie in britischen Familien sowie Genexpressionsmessungen in britischen Kindern. Darüber hinaus gab es Replikationsuntersuchungen in weiteren hier nicht dargestellten Populationen (in Anlehnung an Moffat (Moffatt et al. 2007)).

Zuerst wurden alle Kinder der primären Assoziationsstudie mit dem Illumina Sentrix Human Hap 300 Bead Chip genotypisiert. SNPs mit einer „minor allele frequency“ unter 0,5%, extrem starker Abweichung vom Hardy-Weinberg-Gleichgewicht ($\chi^2 > 25$) oder einer zu geringen Callrate (<95%) wurden von den Analysen ausgeschlossen. So kamen 307328 SNPs zur Auswertung. Außerdem wurden im Panel der Kernfamilien (MRC-A-Panel) weitere 91923

SNPs mit 36% durchschnittlicher Heterozygotität mit dem Illumina Sentrix Human-1 Genotyping Bead Chip typisiert (Familienstudie in Abbildung 25).

Um die Effekte der Polymorphismen auf die Transkription positioneller Kandidatengene zu untersuchen, wurden zudem Genexpressionsanalysen bei Probanden und Geschwistern des Familienpanels untersucht (Moffatt et al. 2007).

Sämtliche auf dem verwendeten Chip befindlichen SNPs wurden in der Population deutscher und britischer Kinder auf Assoziation mit Asthma getestet (dabei entspricht die deutsche Teilstichprobe der GWA-Stichprobe entsprechend Abbildung 24). Auf Grund der Poolung von Kindern aus verschiedenen Ländern wurde geprüft, ob Populationsstratifikation vorliegt. Da diese nicht nachgewiesen wurde, stammen die Hauptergebnisse aus dem über die Länder (Großbritannien und Deutschland) kombinierten Datensatz von 994 Asthmatikern und 1243 Kontrollen.

SNPs, welche bei einem Test auf Assoziation mit Asthma unter einer bestimmten Schranke des p-Wertes blieben, wurden als sogenannte „Hits“ gefiltert. Dazu wurden die 1%-FDR-Schranke als $p \leq 6,8 \times 10^{-7}$ und die 5%-FDR-Schranke als $p \leq 5,0 \times 10^{-6}$ berechnet. Die Punkte über den roten Linien in Abbildung 26 zeigen Regionen an, für welche dieses Signifikanzniveau nach Korrektur für multiples Testen nicht überschritten wurde.

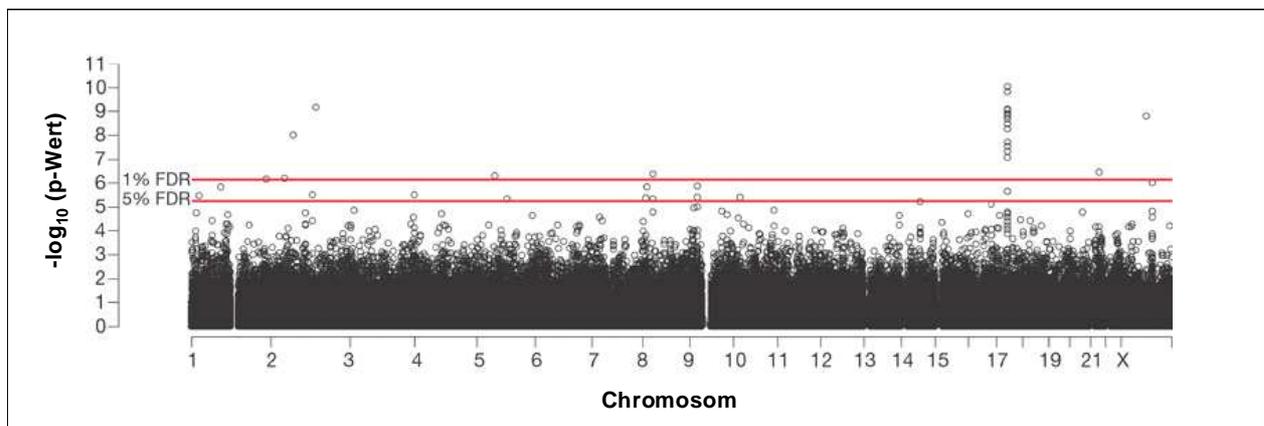


Abbildung 26: Hits aus der GWA-Studie (nach Moffatt (Moffatt et al. 2007))

Daneben wurden die SNPs nach p-Werten aufsteigend sortiert und die SNPs, welche die kleinsten p-Werte bezüglich einer Assoziation mit Asthma zeigten, wurden als „top hits“ bezeichnet. Aus Abbildung 26 geht hervor, dass die meisten der „top hits“ auf Chromosom 17 liegen. Neben diesen blieben weitere Hits auf Chromosom 2, 8, 21 und X unter der 1%-FDR-Schranke. Von 34 „top hits“ gab es sechzehn, welche unter der 5%-FDR-Schranke blieben, und zwar auch dann, wenn ein Korrekturverfahren für Stratifikation vorgenommen wurde. 12

SNPs blieben nach Anwendung dieses Korrekturverfahrens sogar unter der 1%-FDR-Schranke (Tabelle 17).

Marker	Chromosom	Lokation (bp) ¹	Unkorrigierter P-Wert ²	Korrigierter P-Wert ³	
rs10924993	1	234.526.408	7E-07	3E-06	
rs6716266	2	69.098.732	1E-08	9E-08	<1% FDR
rs8179521	2	127.867.394	3E-06	8E-07	
rs3791244	2	138.448.235	7E-10	5E-08	<1% FDR
rs4512342	8	32.727.416	4E-07	2E-09	<1% FDR
rs2666781	10	22.245.682	4E-06	4E-06	
rs907092	17	35.175.785	1E-09	8E-07	
rs9303277	17	35.229.995	2E-09	5E-08	<1% FDR
rs11557467	17	35.282.160	8E-10	3E-08	<1% FDR
rs8067378	17	35.304.874	9E-10	4E-08	<1% FDR
rs2305480	17	35.315.722	3E-09	5E-07	<1% FDR
rs2290400	17	35.319.766	2E-10	4E-09	<1% FDR
rs7216389	17	35.323.475	9E-11	3E-09	<1% FDR
rs4795405	17	35.341.943	2E-09	5E-07	<1% FDR
rs2037986	21	28.398.348	4E-07	6E-07	<1% FDR
rs2311978	X	75.705.247	2E-09	2E-07	<1% FDR

Tabelle 17: Marker der GWA Analyse, welche unter der 5%-FDR-Schranke bleiben.

12 SNPs bleiben auch unter der 1%-FDR-Schranke und stellen die „top hits“ der Studie dar.

¹ Die Angaben beziehen sich auf HG18, einer entsprechenden Datenbank zur Angabe der Lokation der SNPs, im März 2006, bp=„base pairs“

² Unkorrigierte p-Werte stammen aus logistischen Regressionsmodellen der einzelnen Marker.

³ Korrigierte p-Werte stammen aus logistischen Regressionsmodellen, welche 102 zufällige Marker einbeziehen, die für Populationsstratifikation kontrollieren.

An dieser Stelle bieten sich folgende Vorgehensweisen an, die Analysen fortzusetzen, von denen die ersten beiden im Rahmen der vorliegenden Arbeit verfolgt wurden:

1.) Das Naheliegende ist die genauere Betrachtung der sogenannten „top hits“ und der Region, in der sie liegen. Damit einher geht eine genauere Analyse der dort liegenden Gene, Replikation der gefundenen Ergebnisse sowie funktionelle Studien für SNPs, die als kausale Varianten von Asthma hypothetisiert werden. Der Locus auf Chromosom 17q21 als vermuteter Krankheitslocus bildet dafür den Ausgangspunkt.

2) Zudem lassen sich neue Kandidatengene ableiten. Dazu gehören vor allem die Gene, welche in einer Region liegen, in der ein Hit gefunden wurde. Von Interesse können aber ebenso Gene sein, die in einer Beziehung zu einem der assoziierten Loci oder einem Gen innerhalb der hoch assoziierten Loci stehen.

3.) Eine weitere Forschungsrichtung dehnt die interessanten Regionen auf weitere Hits aus, die sich nicht auf die Region in Chromosom 17 begrenzen. Die genomweite Betrachtung aller Hits soll im Rahmen der vorliegenden Arbeit jedoch nicht weiter verfolgt werden.

3.1.2 Betrachtung des Locus auf Chromosom 17

3.1.2.1 Gene und Kopplungsungleichgewicht

Nach Entdeckung des neuen Locus, der hoch assoziiert mit Asthma war, stellte sich die Frage, ob sich damit bestimmte Gene verbinden lassen und inwieweit die „top hits“ miteinander korreliert sind. Abbildung 27 zeigt die Lage der „top hits“, dazugehörige Gene und das zwischen den zehn „top hits“ aus der GWA bestehende LD, das hier für die Querschnittspopulation aus Münchner und Dresdner Kindern (MD3099) berechnet ist.

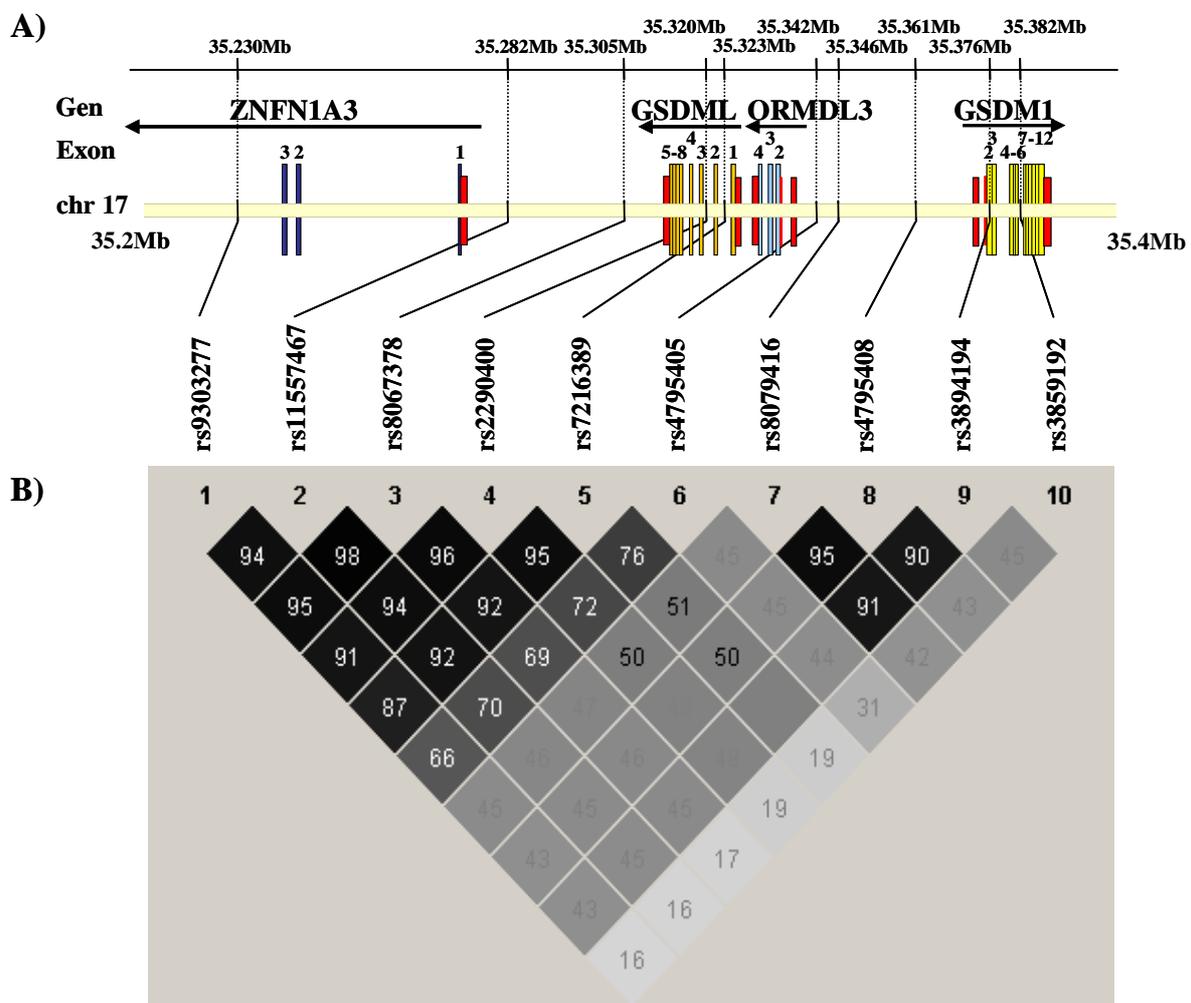


Abbildung 27: A) Lage der zehn Hits auf Chromosom 17 und B) deren „linkage disequilibrium“ (berechnet als R^2) in der Population MD3099 aus ISAAC

Da zwischen den SNPs ein hohes LD zu finden ist, ist es möglich, dass die meisten SNPs für den gleichen ursächlichen Locus stehen. Welcher SNP der kausative ist, ist jedoch unklar.

Verschiedene Gene wurden zu Beginn auf dem durch die GWA für Asthma entdeckten Locus auf Chromosom 17 beschrieben. Dazu gehören *GSDML*, *ORMDL3*, *GSDM1* sowie

ZNFNIA3. Jedem dieser Gene lassen sich SNPs der gefundenen „top hits“ zuordnen. So konnten zunächst verschiedene Gene für das ursächliche Signal auf Asthma hypothetisiert werden. Da die Frage nach dem für Asthma kausativen Locus mit Assoziationsstudien nicht zu klären ist, wurden Transkriptionslevel der Gene als Hinweis genutzt, auf den ursächlichen genetischen Einfluss auf Asthma zu schließen.

Werte für die Transkription werden direkt durch Polymorphismen in regulatorischen Elementen modifiziert (Schadt et al. 2003; Morley et al. 2004). Daher wurde Genexpression in B-Zell-abgeleiteten Epstein-Barr-Virus (EBV)-transformierten lymphoblastoiden B-Zelllinien bei den Kindern des MRC-A-Panel gemessen. EBVL-Zellen präsentieren eine B-Zelllinie und sind demzufolge von direkter Relevanz für Asthma. Polymorphismen von 17q21 waren im Rahmen der Genexpressionsanalyse stark und konsistent mit Transkriptionswerten nur eines der genannten Gene, nämlich *ORMDL3*, assoziiert (Moffatt et al. 2007). Dies macht *ORMDL3* zu dem wichtigsten Kandidatengen, das aus der GWA hervorging. Daher wurde den *ORMDL*-Genen nachfolgend hohe Aufmerksamkeit gewidmet. Aber was weiß man von *ORMDL3*?

3.1.2.2 ORMDL

Die *ORMDL*-Gene sind nach den *ORM1* (*Saccharomyces cerevisiae*)-artigen Genen benannt (Hjelmqvist et al. 2002). Bereits 2002 wurden drei Proteinvarianten entdeckt, welche eine Familie konservierter Proteine im endoplasmatischen Retikulum darstellen (Hjelmqvist et al. 2002). Ursprünglich wurde im Rahmen einer Kopplungsstudie für *Retinitis pigmentosa*, eine häufige Form erblicher Netzhautdegeneration, ein Zusammenhang dieser Krankheit mit genetischen Markern auf Chromosom 2 entdeckt. Eine genauere Charakterisierung dieser Region auf Chromosom 2 identifizierte damals *ORMDL1* als neues Gen.

Der „open reading frame“ (ORF) wird bei der Proteinsynthese translatiert und kann daher als Hinweis dienen, für welches Protein die DNA kodiert. In *ORMDL1* wurde ein ORF gefunden, der für 153 Aminosäuren kodiert. Weitere *in silico* Untersuchungen zeigten Homologien von *ORMDL1* mit zwei weiteren Genen. Im humanen Genom wurden homologe Sequenzen zu *ORMDL1* auf Chromosom 12q13.2 und 17q21.1 gefunden. Die vorhergesagten Gene wurden mit *ORMDL2* bzw. mit *ORMDL3* bezeichnet. Solche Sequenzhomologien fand man nicht nur beim Menschen, sondern auch in zahlreichen anderen Organismen. Die *ORMDL*-Proteine sind in Übereinstimmung zu deren vorausgesagter transmembraner Topologie in der Membran des endoplasmatischen Retikulum lokalisiert. Die genomische Struktur der drei kodieren-

den Exons, inklusive der Grenzen von Intron-Exon-Übergängen, ist bei allen drei Genen komplett konserviert, was auf eine wichtige Rolle dieser Gene hindeutet.

Abbildung 28 zeigt, dass *ORMDL3* sowohl in Niere und Leber und besonders stark im fetalen Gewebe der Lunge exprimiert ist. Eine Rolle von *ORMDL3* bei Atemwegserkrankungen kann daher zwar vermutet werden, eine genaue Funktion ist aber bislang für keines der *ORMDL*-Gene bekannt, insbesondere nicht für *ORMDL3*.

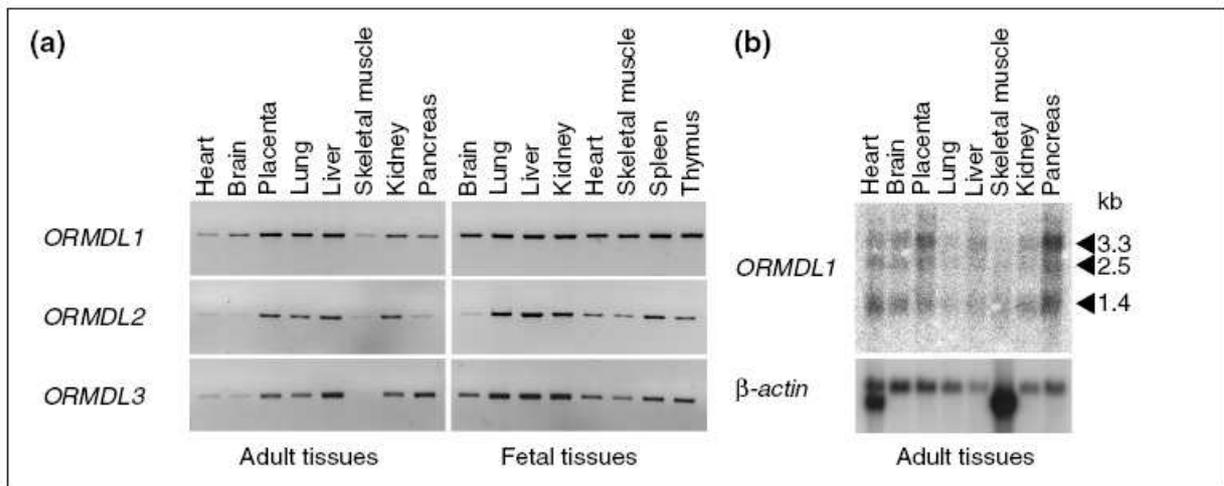


Abbildung 28: Expressionsmuster von *ORMDL1*, *ORMDL2*, und *ORMDL3* in verschiedenen Geweben aus Analysen einer „RealTime-Polymerase chain reaction“ („RealTime-PCR“).

(a) *ORMDL* Expression in Zellgeweben von Erwachsenen und Föten

(b) Northern blot Hybridisierung: *ORMDL1*-Probe mit komplettem ORF gegen mRNA von humanem Gewebe. Die Größen der drei detektierten Transkripte sind gezeigt. Dasselbe Muster wurde bei Verwendung einer *ORMDL1*-Probe mit 3'-UTR gefunden (nach Hjelmqvist (Hjelmqvist et al. 2002)).

3.1.2.3 Replikation des ursprünglichen Befundes

Bereits in der ersten GWA zu Asthma (Moffatt et al. 2007) gab es verschiedene weitere Belege für Assoziationen von SNPs in der identifizierten Region auf Chromosom 17q21 mit Asthma.

Zunächst wurden weitere 27 Marker ausgewählt, welche innerhalb oder anliegend der stark mit Asthma assoziierten Region auf Chromosom 17 lagen. Diese Marker wurden ebenfalls in der aus deutschen und britischen Kindern getesteten Stichprobe untersucht und zeigten ähnlich starke Assoziationen mit Asthma wie die Marker in der ursprünglichen GWA.

Auch innerhalb der Familienstudie entsprechend Abbildung 25 lag der am stärksten mit Asthma assoziierte Marker (rs8067378) in dieser Region auf Chromosom 17 und zeigte beides Mal (also sowohl in der primären Assoziationsstudie als auch in der Familienstudie) dasselbe Risikoallel A. Die Muster der Assoziation waren zudem ähnlich in der deutschen und britischen Stichprobe.

Weiters bestätigten unabhängige Replikationsuntersuchungen die Bedeutung des Locus. In einer britischen Geburtskohorte war der SNP rs3894194 mit Asthma assoziiert. Und in einer

Population von Kindern aus ISAAC, die zuvor nicht in der GWA analysiert worden waren (MD3099 entsprechend Abbildung 24, aber ohne Kinder der GWA), gab es signifikante Assoziationen für acht der zehn „top hits“ (siehe Tabelle 18:). Der SNP rs8067378, bei dem die erste Typisierung in der Replikationskohorte fehlschlug und der daher in Tabelle 18 nicht aufgeführt ist, wurde inzwischen auch untersucht und zeigte ein ähnlich starkes Signal für Asthma wie die anderen Hits (Odds Ratio und 95%-Konfidenzintervall: 1,50 (1,19-1,86); $p=0,0005$). In Tabelle 18 ist auch zu sehen, dass die SNPs, welche dem Gen *GSDMI* zuzuordnen sind (rs3894194, rs3859192) die schwächsten Effekte aufweisen und in der deutschen Replikationskohorte nicht mehr signifikant werden.

In der Zwischenzeit konnte der Effekt des Locus 17q21 in zahlreichen weiteren kaukasischen und nicht-kaukasischen Populationen bestätigt werden (Bisgaard et al. 2008; Bouzigon et al. 2008; Galanter et al. 2008; Hirota et al. 2008; Sleiman et al. 2008). Beeindruckend dabei ist nicht nur die hohe Konsistenz der Assoziation, sondern auch die über die Studien hinweg relativ konstante Größe der Odds Ratio für die Assoziation von Asthma mit SNPs auf Chromosom 17. Zudem wurde in allen Studien dasselbe Risikoallel berichtet.

Tabelle 19 und Tabelle 20 führen einige Assoziationsergebnisse zu den SNPs auf Chromosom 17 in deren Anordnung auf dem Locus auf. Tabelle 19 schließt den ersten Teil des Locus von *ZNFNIA3* bis *ORMDL3* ein, während Tabelle 20 den Teil „upstream“ von *ORMDL3* zeigt. Wie in diesen Tabellen zu erkennen ist, wurden einerseits die zehn „top hits“ der ursprünglichen GWA untersucht, aber auch zahlreiche SNPs, die in dieser Region liegen und in der Studie von Moffatt (Moffatt et al. 2007) noch nicht berücksichtigt worden waren. Die SNPs rs31695721, rs43786501, rs80761311, rs176089251, rs37442461, rs47954021, rs4795405 sind dem Gen *ORMDL3* zuzuordnen. Vor allem die SNPs rs4378650 und rs8076131 von *ORMDL3* fallen auf, da sie häufig und stark als mit Asthma assoziiert nachgewiesen wurden.

Auch Analysen der Genexpression, die unabhängig von der ersten GWA waren, bestätigten die Bedeutung von *ORMDL3*. Polymorphismen von 17q21 waren im Rahmen genomweiter Genexpressionsanalyse nicht nur bei der ersten GWA zu Asthma stark und konsistent mit SNPs auf dem gleichen Chromosom mit Transkriptionslevels von *ORMDL3* assoziiert (Moffatt et al. 2007). Auch in einer genomweiten Expressionsstudien wurden drei „expression quantitative trait nucleotides“ (eQTNs), Loci, die mit der Expression eines Gens assoziiert sind, entdeckt, von denen zwei die Expression von *ORMDL3* beeinflussten. Dies waren der SNP rs7216389 auf Chromosom 17 und ein SNP rs1051740 des Gens *EPHX* auf Chromosom 1. Die beiden SNPs übten einen additiven Effekt auf *ORMDL3* aus (Duan et al. 2008)

Ergebnisse

	Lokation auf Chromosom 17 (in Megabasen)	Primäre GWA Studie -log10 (p) Asthma			eQTL ¹ -log10 (p) ORMDL3	Ref. Allel und Alt. Allel ²		Replikationskohorte	
		MRC-A	deutsch	kombiniert		MRC-A	Ref. Allel	Alt. Allel	Odds Ratio (95%-Konfidenzintervall)
rs9303277	35.230	4,4	5,0	8,8	21,9	C	T	1,41 (1,14-1,76)	2,8
rs11557467	35.282	4,6	5,2	9,1	22,5	G	T	1,45 (1,16-1,82)	3,1
rs8067378	35.305	4,3	5,3	9,0	22,7	-	-	-	-
rs2290400	35.320	4,0	6,4	9,8	22,4	A	G	1,47 (1,18-1,82)	3,2
rs7216389	35.323	4,3	6,4	10,0	22,4	T	C	1,45 (1,17-1,81)	3,1
rs4795405	35.342	4,1	5,3	8,7	14,6	C	T	1,52 (1,20-1,89)	3,6
rs8079416	35.346	2,4	6,5	8,2	10,9	C	T	1,30 (1,05-1,61)	1,9
rs4795408	35.361	2,1	6,1	7,5	10,3	A	G	1,25 (1,01-1,54)	1,4
rs3894194	35.376	1,9	6,6	7,7	11,0	T	C	1,22 (0,98-1,51)	1,2
rs3859192	35.382	2,0	6,0	7,3	3,5	T	C	1,11 (0,89-1,37)	0,4

Tabelle 18: Assoziation von Asthma und ORMDL3 mit SNPs von Chromosom17q21 in kombinierten und Replikationsstichproben SNPs aus dem Illumina 300K Panel. Ergebnisse wurden mit logistischen Regressionsmodellen berechnet. Assoziation mit ORMDL3 Transkriptionsmenge basiert lediglich auf dem Panel der Familien. Replikationsergebnisse sind gezeigt für eine Population von ISAAC II (n=2359).

¹ eQTL=„expression quantitative trait locus“

² Ref. Allel=Referenzallel (entspricht dem Risikoallel); Alt. Allel=alternatives Allel (nach Moffatt (Moffatt et al. 2007))

Ergebnisse

		SNPs (rs_Nummern)																				
		907092	9303277	2060941	11557467	8067378	8069176	2305480	2305479	2290400	4795400	7216389	9303281	7219923	3169572	4378650	8076131	12603332	17608925	3744246	9894164	4795402
	die 10 "top hits":		1		2	3				4	5											
REFERENZ ¹	Replikationskohorte ²	alternatives Allel																				
Moffatt	Deutsch (MD2359)		●		●	●				●	●											
	200 Fälle/2120 Kontrollen		C		G					A	T											
	Odds Ratio		1.41*		1.45*					1.47*	1.45*											
Tavendale	Schottisch										●											
	1054 Fälle/1465 Kontrollen										T											
	(Kinder und junge Erwachsene)																					
	Odds Ratio										1.53*											
Hirota	Japanisch										●											
	545 Fälle/738 Kontrollen										T											
	(Kinder mit atopischem Asthma)																					
	Odds Ratio										1.44*											
Sleiman	Nodamerikanisch		●		●	●				●	●											
	807 Fälle / 2853 Kontrollen		C		G	A				A	T											
	Odds Ratio		1.18*		1.18*	1.19*				1.18*	1.17*											
	Afroamerikanisch																					
	1456 Fälle / 1973 Kontrollen																					
	Odds Ratio		ns		ns	ns				ns	ns											
Galanter	3 Populationen										●					#	#	#	(#)	#	#	
	Mexikanisch										T					C	A	C		C	<5%	
	399 Probanden/798 Eltern										1.26					1.33*	1.38*	1.36*		1.35*		
	Puerto-Ricanisch										T					C	A	C		C	<5%	
	301 Probanden/602 Eltern										1.35*					1.23	1.29*	1.22		1.10		
	Afroamerikanisch										T					C	A	C		C	T	
	261 Fälle/176 Kontrollen										1.21					1.79*	1.20	1.74		1.40	1.25	
Madore	Französisch-Kanadisch				●				#	●	●	#	#		#	#						
	253 Familien /1275 Probanden				G				G	A	T	A	T		C	A						
	p-Wert (IDT) ³				0.01				0.005	0.009		0.014	0.005	0.005		0.017	0.009					
Bouzigon	Französisch		●		●		#	#	#	#	●	#	#	#		▲					▲	▲
	Risikoallel		C		G		G	G	C		C		A	T			A					
	372 Familien /1511 Probanden		0.002		0.004		0.0002	0.0001	0.005		0.0009	ns	0.005	0.009		0.003						
	235 Fälle/1270 Probanden		2.03*		1.95*		2.02*	1.91*	2.03*		1.81*		1.97*	1.90*		1.76*						
Bisgaard	Dänisch		#		#				#	#	●										#	#
	376 Kinder										1.88*											
	316 Fälle /1500 Kontrollen										1.3*											
AG Kabesch	3 Populationen		●		●	●				●	●				▲	#	#		▲	▲		▲
	deutsch		C		G	A				A	T					G	T					
	Dresden (D1940)		1.46*		1.51*	1.50*				1.55*	1.50*					1.56*	1.64*					
	Leipzig (L1165)		1.40*		1.40*	1.49*				1.49*	1.51*					1.53*	1.60*					
	München (M1159)		1.38*		1.40*	1.34*				1.43*	1.50*					1.33*	1.35*					

Tabelle 19: Replikationen von Assoziationen des CHR 17 mit Asthma, Locus Teil 1.

Mit ● markiert sind die als Replikation der „top hits“ untersuchten SNPs. Weitere untersuchte SNPs dieser Region, die nicht zu den 10 „top hits“ zählen, sind mit # markiert, wenn sie mind. einmal zu signifikanten Effekten führten, ansonsten mit ▲ (eingeklammert sind SNPs ohne Genotypisierungsergebnis)

¹ Quellen: Moffatt (Moffatt et al. 2007), Tavendale (Tavendale et al. 2008), Hirota (Hirota et al. 2008), Sleiman (Sleiman et al. 2008), Galanter (Galanter et al. 2008), Madore (Madore et al. 2008), Bouzigon (Bouzigon et al. 2008), Bisgaard (Bisgaard et al. 2008), deutsche Populationen stammen aus Arbeiten der eigenen Arbeitsgruppe

² Fallzahlen der Replikationskohorten

³ falls in Originalpublikation keine Effekte berichtet sind, werden p-Werte angegeben,

⁴ da hier Hazard Ratios berechnet sind, wird nur Effekt für weiter untersuchten SNP berichtet

3.1.2.4 Spezifizierung des ursprünglichen Befundes

Im weiteren Verlauf der Erforschung des Einflusses von Chromosom 17q21 auf Asthma wurde versucht, die Rolle des Locus auf spezifische Formen von Asthma des Kindesalters einzuschränken. Dabei wurden zuerst verschiedene Formen des Schweregrads in Betracht gezogen. Offenbar zeigt sich der Effekt von *ORMDL3* noch deutlicher bei schweren Formen von Asthma im Kindesalter (Bouzigon et al. 2008; Galanter et al. 2008; Tavendale et al. 2008). Als Asthma-Schweregrad wurden dabei verschiedene Parameter wie Symptome, Medikamenteneinnahme, verschlechterte Lungenfunktion oder erhöhte IgE-Werte herangezogen. Die Asthma-Exazerbation definiert über ein Vier-Punkte-Kriterium, das Gebrauch von Medikamenten, Vorliegen täglicher Symptome, Vorliegen nächtlicher Symptome und eine schlechte Lungenfunktion einschloss, verstärkte den Effekt der signifikant mit Asthma assoziierten SNPs rs4378650 und rs1260332 in drei Ethnien (Galanter et al. 2008). Neben Asthma-Exazerbationen wurde das frühe Auftreten von Asthma bei Kindern im Hinblick auf Assoziationen mit Chromosom 17 untersucht. Als Kinder mit frühem Asthma wurden diejenigen definiert, bei denen Asthma bis zum vierten Lebensjahr festgestellt wurde (Bouzigon et al. 2008). Diese Subform von Asthma war besonders stark mit SNPs von Chromosom 17 assoziiert.

Auch die Frage, ob es sich eher um atopisches oder nicht-atopisches Asthma handelt, stellt sich im Kontext nach besonders risikoreichen Subformen von Asthma. Im Rahmen eigener Daten zeichnet sich ab, dass der Effekt von SNPs auf Chromosom 17 auf atopisches Asthma deutlich stärker ausgeprägt ist als deren Effekt auf nicht-atopisches Asthma.

Bei Galanter waren die Effekte auf Asthma innerhalb von Patienten mit einem IgE-Wert über 100 IU/mL innerhalb drei verschiedener Ethnien stärker als die reinen Asthma-Effekte (Galanter et al. 2008). Diese Form des atopischen Asthma ist zwar nicht gleichzusetzen mit dem über Pricktest definiertem atopischem Asthma, aber es lässt vermuten, dass die Assoziationen von Polymorphismen auf Chromosom 17 mit Asthmaformen stärker werden, wenn allergische/atopische Subformen von Asthma Gegenstand der Untersuchung sind. Die bisher einzige Studie, die auf Effekte der Polymorphismen von Chromosom 17 auf nicht-atopisches Asthma hinweist, stammt von Bisgaard (Bisgaard et al. 2008). Hier gab es einen schwachen Effekt von Polymorphismen auf Chromosom 17q21 auf BHR. Dieses Ergebnis beruht jedoch auf einer sehr geringen Fallzahl und war nicht konsistent über verschiedene Zeitpunkte hinweg. Ansonsten wurde bislang kein Effekt von genetischen Determinanten von Chromosom

17 auf BHR gefunden. Zudem konnte in keiner der bisher durchgeführten Studien mit Fokus auf Chromosom 17 ein Zusammenhang der identifizierten „top hits“ mit Lungenfunktion nachgewiesen werden, auch nicht innerhalb der Kinder von ISAAC oder MAGICs.

Ebenso wenig konnte bislang ein klarer Zusammenhang des Locus mit Atopie nachgewiesen werden. In der gepoolten Stichprobe von Kindern aus Dresden, Leipzig und München (DLM4264) zeigte sich eine signifikante Assoziation der Polymorphismen mit Atopie. Interessanterweise zeigen sich einige der Polymorphismen auf Chromosom 17, die unter den „top hits“ waren, auch als Risikofaktor für Atopie (Tabelle 21).

Die Größe des Effektes ist jedoch gering. Zudem blieb der Effekt nicht bestehen, wenn man die Analysen für Asthma adjustierte, so dass hier eher davon auszugehen ist, dass der Atopie-Effekt in Zusammenhang mit dem Asthma-Effekt steht.

Gen	SNP	Odds Ratio (95%-Konfidenzintervall)	p-Wert (Chi- Quadrat-Test)
<i>ZNFNIA3</i>	rs9303277	1,08 (0,97-1,19)	0,1499
<i>ZNFNIA3</i>	rs11557467	1,09 (0,98-1,21)	0,1010
<i>GSDML</i>	rs8067378	1,10 (0,99-1,21)	0,0793
<i>GSDML</i>	rs2290400	1,12 (1,01-1,24)	0,0298
<i>GSDML</i>	rs7216389	1,14 (1,03-1,26)	0,0137
<i>ORMDL3</i>	rs4795405	1,12 (1,01-1,24)	0,0302
<i>ORMDL3</i>	rs8079416	1,11 (1,00-1,23)	0,0488
	rs4795408	1,12 (1,01-1,24)	0,0277
<i>GSDMI</i>	rs3894194	1,11 (1,00-1,23)	0,0436
<i>GSDMI</i>	rs3859192	1,12 (1,01-1,24)	0,0325

Tabelle 21: Atopieeffekte in der Querschnittpopulation aus Kindern von Dresden, Leipzig und München (DLM 4264)

3.1.2.5 Spezifizierung des ORMDL-Effektes

Die SNPs von Chromosom 17 aus der GWA wurden zusammen mit „tagging SNPs“ von *ORMDL3* in der Querschnittspopulation von Kindern aus Dresden, Leipzig und München (DLM4264) vertiefend im Hinblick auf atopisches und nicht-atopisches Asthma untersucht. Tabelle 22 zeigt die Allelfrequenzen der „top hits“ sowie der „tagging SNPs“ von *ORMDL3*.

	genotypi- sierter SNP	Nicht- Risiko- allel	Risiko- allel	AF ¹ Risiko- allel	Call- rate	„bins“ ² <i>ORMDL3</i>	p (HWE) ³
„top hits“	rs9303277	T	C	0,50	93,9%		0,8384
	rs11557467	T	G	0,51	93,5%		0,8660
	rs8067378	G	A	0,50	96,2%		0,4900
	rs2290400	G	A	0,51	93,3%		0,5501
	rs7216389	C	T	0,50	94,0%		0,5372
<i>ORMDL3</i> „tagging SNPs“	rs3169572¹	G	A	0,05	95,8%	„bin“ 1	0,1897
	rs4378650¹	A	G	0,52	96,0%	„bin“ 2	0,3467
	rs8076131¹	C	T	0,55	96,2%	„bin“ 3	0,9797
	rs17608925¹	C	T	0,91	95,6%	„bin“ 4	0,0360
	rs3744246¹	T	C	0,83	95,8%	„bin“ 5	0,0531
	rs4795402¹	A	C	0,78	95,9%	„bin“ 6	0,1193
„top hits“	rs4795405	T	C	0,56	93,6%	„bin“ 3	0,6601
	rs8079416	T	C	0,45	94,6%		0,3991
	rs4795408	G	A	0,45	93,9%		0,8737
	rs3894194	C	T	0,45	94,8%		0,8060
	rs3859192	C	T	0,45	93,8%		0,6761

Tabelle 22: SNP-Beschreibung der Polymorphismen auf Chromosom 17 in DLM4264

¹ AF=Allelfrequenz des Risikoallels. Risikoallel ist immer das Allel, welches das Risiko für Asthma erhöht.

² Fett gedruckte Buchstaben repräsentieren identifizierte „tagging SNPs“ für „bins“ in der interessierenden Region auf Chromosom 17 (SNPs beziehen sich auf eine Region von Chromosom 17, die 20kB um *ORMDL3* aufspannt) – LD-„bins“ basieren auf einem $R^2 > 0.85$.

³ HWE=Hardy Weinberg Equilibrium, p-Wert des χ^2 -Test auf Abweichung vom HWE.

Promotor-SNPs gelten als wahrscheinliche Ursache für eine Veränderung in der Genexpression. Der SNP rs8076131, der in der gemutmaßten Promotorregion von *ORMDL3* liegt, zeigte den stärksten Effekt auf atopisches Asthma in der Studienpopulation DLM4264 (OR=1,71, 95%-KI=1,35-2,18; p=0,00000869), und dieser Effekt war homogen über drei Studienzentren. Zudem konnten Haplotypen, welche für „tagging SNPs“ der *ORMDL3*-Region geschätzt wurden (rs3169572, rs4378650, rs8076131, rs17608925, rs3744246, rs4795402) in Haplotyp-trendregressionen das Risiko für Asthma und atopisches Asthma beeinflussen. Die Effektstärken waren mit einer Odds Ratio von ≥ 2 deutlich und übertrafen die Effekte auf Einzel-SNP-Ebene. Dies stärkt weiter die Bedeutung von *ORMDL3* für Asthma und gibt einen ersten Hinweis auf einen möglichen kausalen Locus.

Da *ORMDL3* nur eines von drei Mitgliedern einer hoch konservierten Genfamilie mit bislang unbekannter biologischer Funktion ist (Hjelmqvist et al. 2002), kann auch eine Rolle der auf Grund von Sequenzhomologien verwandten *ORMDL1* und *ORMDL2* bei der Entwicklung von Asthma vermutet werden. Dies sollte vor allem dann gelten, wenn *ORMDL3* wirklich einen funktionellen Beitrag auf das Asthmasignal auf Chromosom 17q21 leistet.

Neben den Untersuchungen des Einflusses von SNPs von *ORMDL3* auf Asthma wurden daher auch für *ORMDL2* und *ORMDL1* Assoziationsstudien mit deren „tagging SNPs“ in derselben Population von deutschen Kindern (DLM4264) durchgeführt.

Dabei konnten für *ORMDL2* und *ORMDL1* Assoziationen mit atopischem Asthma für zwei seltene SNPs gezeigt werden. Der Polymorphismus rs7954619 von *ORMDL2* zeigte einen protektiven Effekt in der Querschnittspopulation DLM4264 (Odds Ratio und 95%-Konfidenzintervall 0,39 (0,20-0,76); p=0,0039), der homogen in Dresden, Leipzig und München zu finden war. Einen konsistenten Effekt auf atopisches Asthma zeigte in der gleichen Population auch rs12990184, ein Polymorphismus von *ORMDL1* (Odds Ratio und 95%-Konfidenzintervall 1,81 (1,03-3,18); p=0,0361).

Die Frage, ob die Effekte unabhängig voneinander Bestand haben, klärte eine wechselseitig adjustierte Analyse für SNPs aus *ORMDL1*, 2 und 3 zur Berechnung der Effekte für atopisches Asthma (Tabelle 23).

Gen	SNP	Risiko- allel	Wechselseitig adjustierte Effekte OR (95% K.I.) ¹	Wechselseitig und für Kova- riaten adjustierte Effekte ² OR (95% K.I.) ¹
<i>ORMDL 3</i>	rs8076131	T	1,68 (1,32-2,14) p<0,0001	1,70 (1,33-2,17) p<0,0001
<i>ORMDL 2</i>	rs7954619	C	2,45 (1,25-4,79) p=0,0087	2,41 (1,23-4,71) p=0,0102
<i>ORMDL 1</i>	rs12990184	C	1,82 (1,03-3,21) p=0,0394	1,97 (1,12-3,49) p=0,0196

Tabelle 23: Wechselseitig adjustierte Analyse für Effekte von drei SNPs, die stellvertretend für die *ORMDL*-GENE stehen, auf atopisches Asthma .

¹ Effekte beziehen sich auf das Risikoallel, Odds Ratios (OR) und 95% Konfidenzintervalle (95% K.I.) sowie p-Werte des Wald-Test sind angegeben.

² zusätzlich adjustiert für Geschlecht und Rauchen

Da jeder der drei SNPs aus den unterschiedlichen Genen einen signifikanten Zusammenhang (p≤0,05) mit atopischem Asthma zeigte, sind voneinander unabhängige Mechanismen der hier untersuchten Polymorphismen zu vermuten. Welche Mechanismen dies sein könnten, ist jedoch noch zu klären.

Erste Hinweise könnten die Untersuchungen auf dem SNP rs8076131 liefern. Weitere Untersuchungen auf diesem SNP zeigten, dass die Promotoraktivität in Anwesenheit des C-Allels signifikant erhöht gegenüber Promotoraktivität bei Präsenz des T-Allels war. „Electrophoretic Mobility Shift Assays“ (EMSAs) mit einem nuklearen Extrakt einer T-Zelllinie zeigten zudem, dass bei Vorhandensein des C-Allels die Bindung der Transkriptionsfaktoren USF1 und 2 begünstigt war.

3.2 *ADAM33 als Vertreter des „positional cloning“*

ADAM33 ist das erste Gen, welches durch „positional cloning“ identifiziert wurde. Im *ADAM33*-Gen wurden bislang zahlreiche (>55) Polymorphismen identifiziert und im Zusammenhang mit Asthma bronchiale untersucht. Genetische Untersuchungen haben gezeigt, dass sich vor allen Dingen in der 3'-Region des Gens, in der die transmembranösen und zytoplasmatischen Domänen des Proteins kodiert werden, die Polymorphismen häufen.

Die ersten Replikationsstudien ergaben jedoch inkonsistente Befunde im Hinblick auf deren Assoziation mit Asthma (Howard et al. 2003; Jongepier et al. 2004; Raby et al. 2004; Werner et al. 2004; Blakey et al. 2005). Die eigene durchgeführte Replikationsstudie an Kindern aus ISAAC und MAS widerspiegelte ebenfalls dieses inkonsistente Bild (Schedel et al. 2006).

In unserer Analyse wurden zehn SNPs in einer Fall-Kontroll-Studie von 1872 deutschen Kindern (MDL1872 entsprechend Abbildung 24) und einer Population von 824 deutschen Kindern aus einer Kohortenstudie (MAS) untersucht. Für die Analysen der Kohorte wurden die Ergebnisse herangezogen, welche die Kinder im Alter von 10 Jahren beschreiben. Die zehn ausgewählten Polymorphismen begründen sich auf zuvor gezeigten Assoziationen und deren Lage im Gen ist Abbildung 29 zu entnehmen.

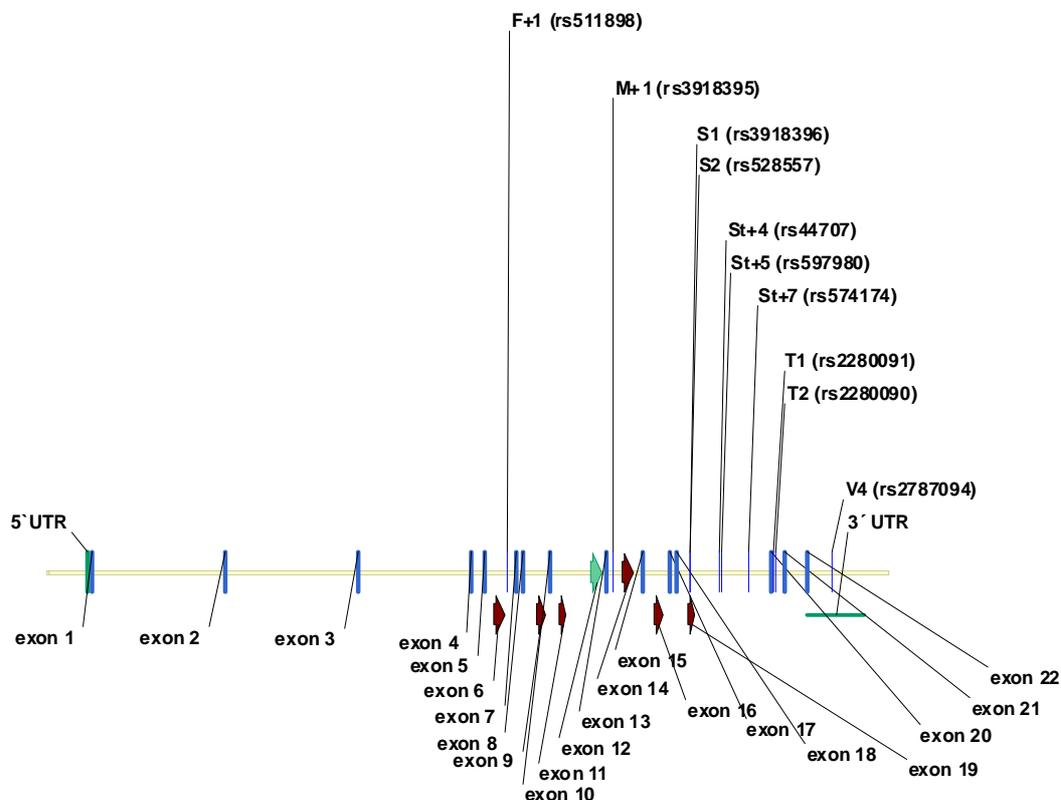


Abbildung 29: Lage der untersuchten SNPs in *ADAM33*

Auf Einzel-SNP-Ebene wurde keine Assoziation mit Asthma in MDL1872 oder MAS gefunden (Tabelle 24). Ebenso zeigten sich keinerlei signifikanten Ergebnisse bezüglich BHR oder dem Phänotyp Asthma mit gleichzeitigem Auftreten von BHR. Einige Effekte für Asthma+BHR fielen jedoch etwas stärker aus, obgleich sie nicht signifikant wurden. Insbesondere ist der SNP S2 zu beachten, der für den kombinierten Phänotyp die höchste Odds Ratio aufweist.

			Asthma	Asthma	Asthma	Asthma und BHR
SNP ¹	rs-Nummern Basentausch	Allelfrequenz ²	Fall-Kontroll-Studie MDL1872 (N=1872)	Kohortenstudie (N=824) zu bestimmtem Zeitpunkt	Studien gepoolt	Fall-Kontroll-Studie MDL1872
<i>F+I</i>	rs511898 G/A	0,38 0,36	0,92 (0,72-1,17) p=0,502	0,94 (0,58-1,52) p=0,795	0,92 (0,74-1,15) p=0,478	1,61 (0,90-2,88) p=0,105
<i>M+I</i>	rs3918395 G/T	0,15 0,13	0,86 (0,65-1,13) p=0,269	0,94 (0,53-1,65) p=0,819	0,89 (0,70-1,14) p=0,351	1,46 (0,84-2,54) p=0,177
<i>S1</i>	rs3918396 G/A	0,09 0,09	1,23 (0,90-1,69) p=0,201	1,16 (0,64-2,10) p=0,635	1,20 (0,90-1,58) p=0,210	1,10 (0,54-2,25) p=0,789
<i>S2</i>	rs528557 G/C	0,28 0,28	0,99 (0,78-1,26) p=0,949	0,97 (0,60-1,56) p=0,893	0,99 (0,80-1,22) p=0,908	1,72 (0,99-3,00) p=0,052
<i>ST+4</i>	rs44707 A/C	0,41 0,41	1,04 (0,81-1,34) p=0,766	1,31 (0,79-2,18) p=0,292	1,10 (0,87-1,37) p=0,432	1,22 (0,69-2,18) p=0,490
<i>ST+5</i>	rs597980 C/T	0,44 0,45	0,94 (0,73-1,22) p=0,649	0,80 (0,48-1,33) p=0,384	0,89 (0,71-1,12) p=0,338	0,91 (0,51-1,60) p=0,731
<i>ST+7</i>	rs574174 G/A	0,19 0,19	1,11 (0,86-1,42) p=0,428	1,16 (0,71-1,89) p=0,549	1,10 (0,88-1,38) p=0,384	0,96 (0,55-1,70) p=0,901
<i>T1</i>	rs2280091 T/C	0,16 0,14	0,96 (0,74-1,25) p=0,762	0,92 (0,53-1,60) p=0,764	0,97 (0,76-1,23) p=0,804	1,50 (0,87-2,59) p=0,141
<i>T2</i>	rs2280090 C/T	0,16 0,14	0,95 (0,73-1,24) p=0,711	0,88 (0,50-1,55) p=0,666	0,96 (0,76-1,22) p=0,731	1,51 (0,88-2,61) p=0,134
<i>V4</i>	rs2787094 C/G	0,22 0,23	1,22 (0,95-1,55) p=0,115	0,82 (0,51-1,34) p=0,439	1,11 (0,90-1,38) p=0,329	1,31 (0,76-2,26) p=0,332

Tabelle 24: Assoziation von SNPs aus ADAM33 mit Asthma (für MDL1872 auch Asthma+BHR)

¹ Nomenklatur der SNPs in Anlehnung an die Publikation von van Eerdewegh (Van Eerdewegh et al. 2002)

² Allelfrequenz jeweils für Fall-Kontroll-Studie (MDL1872) und die Kohortenstudie (MAS)

Innerhalb der Fall-Kontroll-Studie von Kindern aus München, Dresden und Leipzig war das Risiko, nicht-atopisches Asthma zu entwickeln, bei Trägern des polymorphen Allels S1 (OR 1,53, 95%-K.I. 1,01–2,31; p = 0,042) und Trägern des polymorphen G-Allels des SNPs V4 (OR 1,44, 95%-K.I. 1,03–2,01; p = 0,031) erhöht. Reduziert war dagegen das Risiko für die Entwicklung von nicht atopischen Asthma für Träger des polymorphen T-Allels von M+1 (OR = 0,60, 95%-K.I. 0,40 – 0,91; p = 0,016).

Dieses Ergebnis deutet auf die Wichtigkeit der Differenzierung verschiedener Formen von Asthma hin und bestätigt die Vermutung, dass *ADAM33* eher für nicht-atopische Formen von Asthma relevant sein dürfte. Die Häufigkeit von Haplotypen wurde innerhalb der Kinder geschätzt, für welche ausreichende Phaseninformation vorlag (n=1802 in MDL1872; n=782 in MAS). Die geschätzten Haplotypen wurden ebenfalls auf Zusammenhänge mit Asthma und BHR untersucht. Tabelle 25 zeigt die Ergebnisse der Haplotypanalysen mit Asthma für aus 8 SNPs (F+1, S1, S2, ST+4, ST+5, ST+7, T1 und V4) gebildete Haplotypen. Die SNPs M+1, T1 und T2 waren in hohem LD, so dass die Hinzunahme von M+1 und T2 zum Haplotyp keine Änderung der Resultate bewirkte.

Ein häufiger Haplotyp, H4 (G-G-G-C-C-G-T-G), zeigte eine schwache aber nicht signifikante Assoziation mit Asthma in der Fall-Kontroll-Population. Diese Assoziation wurde in einer gepoolten Analyse signifikant (OR 1,28; 95%CI 1,02–1,60, p = 0,033). Keine Assoziation wurde hingegen für BHR mit Haplotypen von *ADAM33* gefunden.

	Haplotyp	Popu- lation ¹	Haplotyp-Häufigkeiten		χ^2 -Test	
			Asthmatiker	Kontrollen	OR und 95%-K.I. ²	p-Wert
H1	G-G-G-A-T-G-T-C	alle	30,35%	32,21%	0,92(0,78-1,08)	0,293
		FK	30,90%	31,56%	0,97(0,81-1,15)	0,735
		Koh	30,03%	34,02%	0,83(0,58-1,19)	0,317
H2	G-G-G-C-C-G-T-C	alle	17,02%	16,69%	1,02(0,84-1,25)	0,804
		FK	16,77%	16,85%	0,99(0,80-1,24)	0,939
		Koh	16,93%	16,32%	1,04(0,67-1,63)	0,847
H3	A-G-C-A-C-G-C-C	alle	12,98%	14,08%	0,91(0,73-1,13)	0,402
		FK	13,60%	14,85%	0,90(0,71-1,15)	0,406
		Koh	10,33%	12,04%	0,84(0,49-1,45)	0,531
H4	G-G-G-C-C-G-T-G	alle	12,96%	10,44%	1,28(1,02-1,60)	0,033
		FK	13,15%	10,65%	1,27(0,99-1,64)	0,063
		Koh	11,43%	9,85%	1,18(0,70-2,00)	0,528
H5	A-A-C-A-T-A-T-G	alle	8,35%	7,69%	1,09(0,83-1,43)	0,517
		FK	8,62%	7,16%	1,22(0,90-1,66)	0,191
		Koh	7,18%	9,07%	0,78(0,41-1,46)	0,430
H6	A-G-G-C-C-A-T-C	alle	6,70%	6,54%	1,03(0,76-1,38)	0,867
		FK	6,17%	6,67%	0,92(0,65-1,30)	0,638
		Koh	8,99%	6,19%	1,50(0,82-2,72)	0,184

Tabelle 25: Assoziation von Haplotypen in *ADAM33* mit Asthma

¹ alle = Fall-Kontroll-Studie und Kohortenstudie gepoolt, FK = Fall-Kontroll-Studie (n=358 Fälle/n = 1198 Kontrollen), Koh = Kohortenstudie (n= 82 Fälle/n = 464 Kontrollen); Kinder ohne Haplotypinformation wurden ausgeschlossen.

² OR=Odds Ratio, 95%-K.I.=95%-iges Konfidenzintervall. Odds Ratios wurden für einen Haplotyp versus alle anderen berechnet.

Somit gab es zwar einen schwachen Hinweis für eine Assoziation eines Haplotypen von *ADAM33* mit Asthma. Dieser konnte jedoch nur in der gepoolten Analyse das Signifikanzniveau von $p=0,05$ unterschreiten und wurde nicht signifikant, wenn man die Populationen getrennt betrachtete. Ebenso wenig konnte auf Einzel-SNP-Ebene ein konsistenter Effekt gefunden werden, obgleich der SNP S2 Hinweise auf einen möglichen Zusammenhang mit Asthma oder dem kombinierten Phänotyp Asthma+BHR gab.

Auch Untersuchungen der Lungenfunktion zeigten nur wenig überzeugende Hinweise für einen Einfluss von *ADAM33* auf Asthmaphänotypen. In der Gruppe der kranken Kinder war für Träger des polymorphen Allels von S2, T1 und T2 der Lungenfunktionsparameter „forced vital capacity“ (FVC) und für Träger des polymorphen Allels von S2 und M+1 der Lungenfunktionsparameter FEV_1 erhöht. Die Unterschiede waren aber an der Grenze zur klinischen Relevanz und würden einer Korrektur für multiples Testen nicht standhalten. Somit waren auch die Befunde unserer Studie zu *ADAM33* recht inkonsistent.

3.3 TLR als Vertreter der Kandidatengene

Die Toll-like Rezeptoren stellen eine Studie dar, welche dem klassischen Kandidatengenansatz folgte. Die Polymorphismen dieser Studie wurden innerhalb des Programms Innate Immunity systematisch aus 259 zuvor in den zehn menschlichen TLRs identifizierten Polymorphismen ausgewählt (National Heart 2008).

Ausgewählt wurden dabei a) Polymorphismen mit einer $MAF > 0,03$, die zu Aminosäureänderungen führten und dadurch eine Änderung der Proteinstruktur und Funktion bewirkten ($n=9$) oder b) SNPs mit einer $MAF > 0,1$, die zu einer Veränderung des Transkriptionsbindungsverhaltens führten ($n=9$), so wie es von dem Programm FastSNP (Yuan et al. 2006) vorhergesagt wurde. Zusätzlich wurden c) drei TLR-SNPs aufgenommen, von denen bereits bevor unsere Studie begann beschrieben worden war, dass sie eine Rolle bei Asthma spielen ($n=3$).

Der Ansatz ist damit eher auf bestimmte SNPs fokussiert, die eine begründete Hypothese liefern, mit Asthma oder Atopie assoziiert zu sein und steht im Gegensatz zu einem Ansatz, der sich auf „tagging SNPs“ beschränkt. Außerdem steht er selbstverständlich im Gegensatz zu den hypothesenfreien genomweiten Studien. Insgesamt wurden 21 SNPs untersucht, welche das Spektrum aller zehn im Menschen vorkommenden *TLR*-Gene überdecken. Sämtliche ausgewählten SNPs wurden auf Assoziationen mit Asthma und den Varianten von Asthma (atopisches, nicht-atopisches Asthma) getestet (Tabelle 26). Diese Analysen wurden in der Fall-Kontroll-Population MDL1872 durchgeführt.

Ergebnisse

Gen	SNP	rs-Nummer	Position relativ zum ATG ¹		Allele	Lokation	Δ AA ²	Δ TF ³	IE ⁴	MAF ⁵	Asthma (n = 369)		Atopisches Asthma (n = 171)		Nicht-atopisches Asthma (n = 171)	
			OR (95%-K.I.) ⁶	p-Wert							OR (95%-K.I.) ⁶	p-Wert	OR (95%-K.I.) ⁶	p-Wert		
TLR1	a	rs5743594	-2299	C/T	Intron			✓	0,19	0,84 (0,65-1,08)	0,184	0,99 (0,70-1,39)	0,935	0,78 (0,55-1,11)	0,174	
	b	rs5743595	-2192	T/C	Intron			✓	0,17	0,81 (0,63-1,05)	0,119	0,54 (0,37-0,81)	0,002 §	1,07 (0,76-1,50)	0,708	
	c	rs4833095	743	A/G	KDS	✓			0,23	0,82 (0,64-1,05)	0,110	0,59 (0,41-0,83)	0,003 §	1,03 (0,74-1,43)	0,848	
TLR2	a	rs4696480	-16934	T/A	Promotor				0,50	1,10 (0,84-1,44)	0,504	1,31 (0,89-1,94)	0,173	0,84 (0,59-1,19)	0,322	
	b	rs1898830	-15607	A/G	Promotor		✓		0,35	1,04 (0,82-1,32)	0,750	0,87 (0,63-1,20)	0,389	1,27 (0,91-1,77)	0,161	
	c	rs3804099	597	T/C	KDS				0,43	0,83 (0,65-1,06)	0,133	0,98 (0,70-1,38)	0,906	0,69 (0,50-0,96)	0,027	
TLR3	a	rs3775291	6301	C/T	KDS	✓			0,29	0,85 (0,67-1,09)	0,201	0,87 (0,62-1,21)	0,410	0,80 (0,57-1,12)	0,199	
TLR4	a	rs2737190	-6687	A/G	Promotor		✓		0,32	1,32 (1,04-1,68)	0,025	1,25 (0,89-1,74)	0,195	1,41 (1,01-1,96)	0,045	
	b	rs10759932	-5724	T/C	Promotor		✓		0,13	1,07 (0,81-1,42)	0,621	1,12 (0,76-1,64)	0,574	1,15 (0,79-1,67)	0,467	
	c	rs4986791	4735	C/T	KDS	✓			0,06	0,85 (0,58-1,25)	0,397	0,65 (0,36-1,18)	0,154	0,97 (0,58-1,62)	0,913	
TLR5	a	rs5744168	1174	C/T	KDS	✓			0,05	1,11 (0,75-1,62)	0,605	0,89 (0,51-1,56)	0,684	1,18 (0,71-1,98)	0,525	
	b	rs2072493	1775	A/G	KDS	✓			0,14	1,08 (0,82-1,41)	0,588	0,99 (0,68-1,44)	0,966	1,07 (0,74-1,54)	0,737	
	c	rs5744174	1846	T/C	KDS	✓			0,43	0,79 (0,62-1,01)	0,061	0,85 (0,60-1,19)	0,342	0,75 (0,53-1,04)	0,085	
TLR6	a	rs5743789	-2078	T/A	Promotor		✓		0,18	0,78 (0,61-1,01)	0,060	0,54 (0,37-0,79)	0,002 §	1,02 (0,72-1,42)	0,927	
	b	rs5743810	745	C/T	KDS	✓			0,40	1,20 (0,93-1,53)	0,154	1,79 (1,24-2,58)	0,002 §	0,83 (0,60-1,16)	0,276	
TLR7	a	rs179008	17962	A/T	KDS	✓			0,23	^f 0,81 (0,55-1,20)	0,299	1,01 (0,56-1,82)	0,982	0,74 (0,45-1,22)	0,244	
										^m 0,99 (0,68-1,46)	0,976	0,92 (0,56-1,51)	0,751	1,40 (0,84-2,35)	0,200	
TLR8	a	rs3761624	-4824	A/G	Promotor		✓		0,25	^f 0,94 (0,64-1,37)	0,748	0,92 (0,51-1,67)	0,786	0,96 (0,59-1,55)	0,865	
										^m 1,16 (0,81-1,66)	0,421	0,99 (0,62-1,58)	0,962	1,33 (0,80-2,22)	0,269	
TLR9	a	rs187084	-2871	T/C	Promotor		✓		0,42	1,33 (1,03-1,71)	0,030	1,35 (0,94-1,92)	0,101	1,18 (0,83-1,66)	0,353	
	b	rs5743836	-2622	T/C	Promotor		✓		0,14	1,08 (0,82-1,42)	0,600	1,04 (0,71-1,52)	0,831	1,16 (0,80-1,68)	0,440	
TLR10	a	rs11096956	1032	G/T	KDS				0,19	0,90 (0,70-1,16)	0,408	0,75 (0,53-1,08)	0,121	1,07 (0,76-1,51)	0,682	
	b	rs4129009	2323	A/G	KDS	✓			0,16	0,79 (0,61-1,03)	0,080	0,58 (0,39-0,86)	0,006	1,01 (0,71-1,43)	0,965	

Tabelle 26: Ergebnisse der Assoziationsstudien für 21 SNPs aus der Familie der Toll-like Rezeptoren in MDL1872

¹ Basierend auf März 2006: human reference sequence (NCBI build 36.1)

² SNP führt zu einer Aminosäureänderung (AA=„amino acid“)

³ SNP verändert das Transkriptionsfaktorbindungsverhalten, basierend auf einer Analyse mit FastSNP (TF=„transcription factor“)

⁴ SNP verändert die Sequenz eines intronischen Regulationselements basierend auf einer Analyse mit FastSNP (IE=„Intronic enhancer“)

⁵ MAF=„minor allele frequency“

⁶ OR (und 95%K.I.)=Odds Ratio und 95%-Konfidenzintervall

[§] signifikant nach Korrektur für multiples Testen (FDR-Korrektur nach Benjamini (Benjamini et al. 2001))

SNPs aus *TLR4* und *TLR9* zeigten ein erhöhtes Risiko für Asthma (*TLR4* insbesondere für nicht-atopisches Asthma) in der nicht adjustierten Analyse. Dies blieb jedoch nach Korrektur für multiples Testen nicht signifikant. Dagegen zeigten SNPs in *TLR1*, *TLR6* und *TLR10* starke überwiegend protektive Effekte für atopisches Asthma. Lediglich ein SNP innerhalb des *TLR1-6-10*-Clusters verfehlte für atopisches Asthma das Signifikanzniveau nach Korrektur für multiples Testen. Eine vertiefende Analyse zeigte auch die Homogenität dieser Effekte in allen drei Städten.

Die Frage, ob dieser Effekt eher ein auf Asthma zu beziehender oder ein reiner Atopieeffekt ist, konnte im Rahmen der Fall-Kontroll-Studie nicht ausreichend geklärt werden, da dort auf Grund der systematischen Selektion von Asthmatikern die Atopievariable nur in einer durch das Studiendesign verzerrten Form vorlag und dementsprechend nicht ausgewertet werden konnte. Daher wurden die signifikanten SNPs aus *TLR1*, *TLR6* und *TLR10* noch einmal in einer Querschnittspopulation aus Dresdner und Münchner Kindern, nämlich in MD3099, untersucht. Hier wurde ein signifikanter Effekt für Atopie über all diese SNPs von *TLR1*, *TLR6* und *TLR10* festgestellt, der auch nach Korrektur für multiples Testen Bestand hatte. Allerdings konnte gezeigt werden, dass die Effekte für atopisches Asthma, die ebenfalls nach Korrektur für multiples Testen signifikant blieben, über alle SNPs deutlich höher ausfielen als für Atopie ohne das gleichzeitige Auftreten von Asthma (Tabelle 27).

<i>Gen</i>		Asthma (n = 272 ¹)	Nicht-atopisches Asthma (n = 133 ¹)	Atopisches Asthma (n = 124 ¹)	Atopie (n = 777 ¹)	Atopie ohne Asthma (n = 645 ¹)
<i>TLR1</i>	b	0,92 (0,70-1,21) p=0,561	1,15 (0,79-1,67) p=0,461	0,60 (0,38-0,93) p=0,021*	0,78 (0,65-0,94) p=0,009*	0,84 (0,68-1,02) p=0,075
	c	0,94 (0,73-1,21) p=0,635	1,08 (0,76-1,54) p=0,664	0,61 (0,41-0,91) p=0,014*	0,75 (0,63-0,89) p=0,001*	0,79 (0,66-0,95) p=0,012
<i>TLR6</i>	a	0,94 (0,72-1,24) p=0,664	1,16 (0,80-1,67) p=0,437	0,64 (0,42-0,98) p=0,038*	0,80 (0,67-0,96) p=0,015*	0,84 (0,69-1,03) p=0,087
	b	1,11 (0,85-1,44) p=0,457	0,80 (0,56-1,15) p=0,228	1,76 (1,16-2,67) p=0,007*	1,24 (1,04-1,47) p=0,017*	1,13 (0,94-1,36) p=0,202
<i>TLR10</i>	b	0,90 (0,68-1,19) p=0,467	1,09 (0,75-1,59) p=0,647	0,60 (0,38-0,95) p=0,026*	0,74 (0,62-0,90) p=0,002*	0,79 (0,64-0,96) p=0,019

Tabelle 27: Assoziation von *TLR1*, *TLR6* und *TLR10* Polymorphismen mit Asthma, Atopie und Subformen von Asthma und Atopie in einer Querschnittspopulation aus Dresdner und Münchner Kindern (MD3099). Odds Ratio und 95%-Konfidenzintervalle für das dominante Modell werden gezeigt

¹ Anzahl der jeweiligen Fälle im Phänotyp

* signifikant nach Korrektur für multiples Testen

Die Studie wurde durch funktionelle Untersuchungen komplettiert. Quantitative Realtime-PCR zeigte, dass Träger der selteneren, protektiven Allele in *TLR1*, *TLR6* und *TLR10* erhöhte

mRNA Levels der entsprechenden *TLRs* aufwiesen. Dies wurde mit Durchflusszytometrie auf Proteinebene bestätigt. Außerdem führte die Stimulation von „peripheral blood mononuclear cells“ (PBMCs) - gewonnen aus Trägern der selteneren Allele - mit spezifischen TLR2/1 und TLR2/6 Liganden ex vivo zu einer erhöhten inflammatorischen Antwort sowie zu erhöhter Expression typischer T_h1 -Zytokine (IL-12 und $IFN-\gamma$) und verminderter T_h2 -Antwort eines T_h2 -Zytokins (IL-4), verglichen mit Trägern der Wildtypallele (Kormann et al. 2008) (s. Abbildung 30).

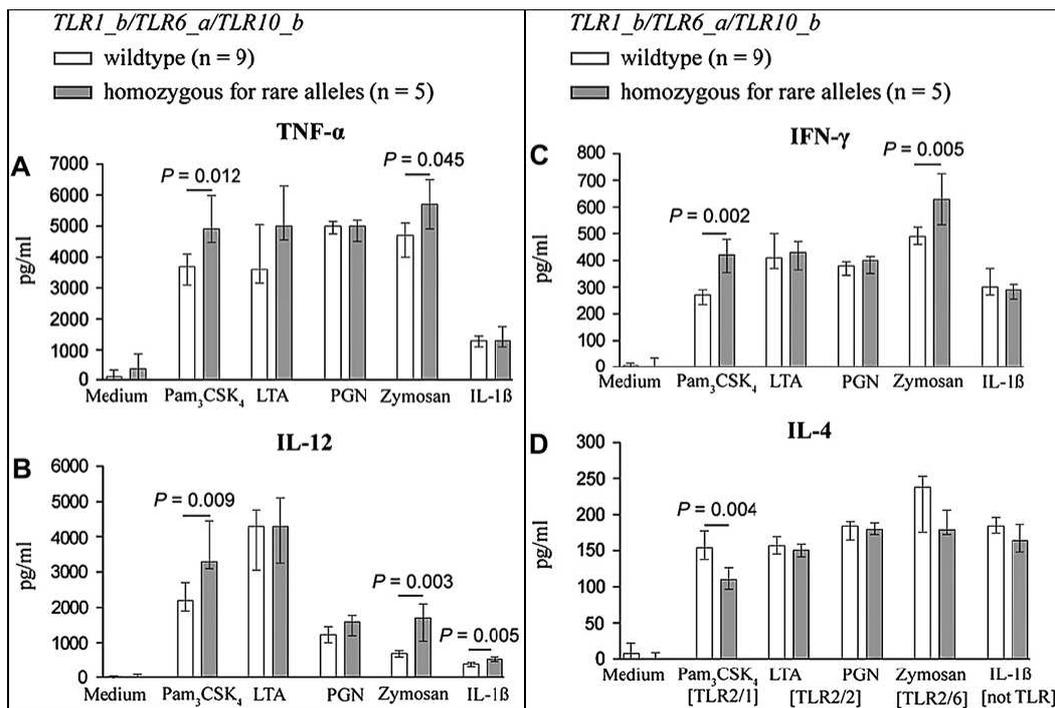


Abbildung 30: TLR1, TLR6 und TLR10 SNPs erhöhen proinflammatorische und T_h1 -assoziierte Zytokinlevels und reduzieren T_h2 -assoziierte IL-4-Expression nach TLR2/1 und TLR2/6 Stimulation

1 Ausschüttung von TNF- α (A), IL-12 (B), $IFN-\gamma$ (C) und IL-4 (D) wurde in Zellkulturüberständen mittels ELISA in Triplikaten gemessen

2 Säulen repräsentieren Mediane und Interquartilsabschnitt zwischen den verschiedenen Spendern

LTA=„Lipoteichoic acid“; PGN=Peptidoglykan.

Diese Effekte wurden nach Stimulation mit ausschließlich TLR2-spezifischen Liganden (LTA und PGN) nicht beobachtet. Dies deutet darauf hin, dass diese Effekte durch die Heterodimerisierung bedingt sind und *TLR1* sowie *TLR6* die T-Zell abhängige Zytokinexpression nach Stimulation regulieren. Leider konnte ein solcher Versuch für *TLR10* nicht durchgeführt werden, da hierfür bislang keine Liganden bekannt waren.

Die Heterodimerisierung von *TLR1*, *TLR6* und *TLR10* mit *TLR2* legt die Vermutung nahe, dass Polymorphismen von *TLR2* die Effekte in *TLR1*, *6* und *10* modifizieren könnten.

Solche Interaktionen ließen sich für alle drei Polymorphismen von *TLR2* nachweisen. Dabei modifizierten die SNPs aus *TLR1*, *TLR6* oder *TLR10* den Effekt der *TLR2*-SNPs immer der-

art, dass stärkere dominante protektive Effekte der *TLR2*-SNPs gefunden wurden, wenn die SNPs von *TLR1*, *TLR6* und *TLR10* für das seltene Allel stratifiziert wurden.

rs4129009(*TLR10*) und rs4833095(*TLR1*) modifizierten den Effekt von rs3804099(*TLR2*) auf atopisches und nicht-atopisches Asthma. rs5743595(*TLR1*) und rs5743789(*TLR6*) modifizierten den Effekt von rs4696480(*TLR2*) für atopisches Asthma. Und rs5743594(*TLR1*) modifizierte den Effekt von rs1898830(*TLR2*) für Asthma (Kormann 2007).

4 Diskussion

4.1. Zusammenfassung wesentlicher Ergebnisse

Drei methodische Annäherungen zur Identifizierung krankheitsrelevanter Gene haben die Forschung der letzten Jahrzehnte bestimmt: der Kandidatengenansatz, das „positional cloning“ und genomweite Assoziationsstudien. Im Rahmen der vorliegenden Arbeit wurden drei Studien behandelt, in denen Gene mit je einer der methodischen Annäherungen in Verbindung gebracht werden. Die Ergebnisse für SNPs auf Chromosom 17 und dem dort befindlichen Gen *ORMDL3* sind als Resultat der ersten durchgeführten GWA für Asthma zu verstehen. Resultate für *ADAM33* können mit dem „positional cloning“ in Verbindung gebracht werden; und die Befunde einer Studie, welcher der Kandidatengenansatz zugrunde liegt, repräsentieren Ergebnisse für die *TLRs*.

Den Schwerpunkt der berichteten Ergebnisse stellen die jüngsten Befunde der ersten GWA für Asthma dar. Dabei wurde ein neuer Locus auf Chromosom 17q21 als Kandidatenregion identifiziert, in der Polymorphismen das Risiko für Asthma signifikant erhöhen. Fast alle Replikationsstudien bestätigten diese Effekte, wobei auch die Effektstärken ähnlich den in der Originalpublikation berichteten Odds Ratios von etwa 1,5 entsprachen.

Aufgrund von Assoziationen der „top hits“ der GWA mit erhöhter Genexpression von *ORMDL3* wurden die Untersuchungen auf *ORMDL3* eingeeengt, obgleich nicht auszuschließen ist, dass andere Gene auf Chromosom 17q21 ebenfalls Einfluss auf die Entwicklung von Asthma nehmen können. In einer vertiefenden Analyse zeigte ein SNP, der in der GWA noch nicht untersucht wurde, der Promotor-SNP rs8076131 von *ORMDL3* den stärksten Effekt für atopisches Asthma.

Assoziationstests wurden auch von SNPs in den auf Grund von Sequenzhomologien mit *ORMDL3* verwandten Genen *ORMDL1* und *ORMDL2* durchgeführt. Dabei zeigten sich ebenfalls Assoziationen von Polymorphismen dieser Gene mit atopischem Asthma. Die Effekte von Polymorphismen verschiedener *ORMDL*-Gene waren dabei unabhängig. Dennoch nimmt *ORMDL3* bisher in Anbetracht zahlreicher Replikationen und erster Befunde aus funktionalen Untersuchungen die herausragende Stellung unter den *ORMDLs* ein.

Für *ADAM33* wurden Assoziationen eines Haplotyps mit Asthma sowie verschiedener SNPs mit nicht-atopischem Asthma gefunden. Diese Zusammenhänge waren jedoch nicht so stark und zudem inkonsistent zwischen verschiedenen Populationen.

Aus der Kandidatengenstudie gingen Assoziationen von *TLR1*, *6* und *10* Polymorphismen mit Atopie und atopischem Asthma hervor. Die Stimulation von PBMCs – gewonnen aus Trägern der selteneren, protektiven *TLR1*, *6* und *10* Allele – mit spezifischen TLR2/1 und TLR2/6 Liganden erhöhte die Expression typischer T_h1-Zytokine. Dies gibt erste Hinweise darauf, wie die assoziierten *TLR* SNPs den protektiven Effekt vermitteln könnten.

Die hier vorliegende Arbeit zeigt, dass durch Verwendung verschiedener methodischer Herangehensweisen ein größeres Spektrum genetischer Variationen identifiziert werden kann, die alle einen Einfluss auf Asthma oder bestimmte Formen von Asthma ausüben können. Nachfolgend soll diskutiert werden, wie die Befunde zu *ORMDL3*, *ADAM33* und den *TLRs* im Kontext weiterer Studien zu bewerten sind, die positive Assoziationen von genetischen Determinanten mit Asthma berichten.

4.2 Einordnung der Ergebnisse in Kontext der Asthmaforschung

4.2.1 Gene, deren Assoziation mit Asthma oft repliziert wurde

In einigen Arbeiten wurden Gene im Hinblick darauf untersucht, wie oft ein Zusammenhang mit Asthma und verwandten Phänotypen repliziert werden konnte. Zwei bedeutende Übersichtsarbeiten, welche die Kandidatengene für Asthma mit häufigen positiven Replikationen darstellen, stammen von Ober (Ober et al. 2006) und Vercelli (Vercelli 2008). Abbildung 31 zeigt die Anzahl der positiven Replikationen und die drei Kandidatengene *ADRB2*, *IL4R* und *HLADRBI*, die im Jahr 2008 an der Spitze der Gene lagen, für die positive Assoziationsstudien mit Asthma berichtet wurden (Vercelli 2008). In dieser Aufstellung der über 30 Gene, für die ein Zusammenhang mit Asthma relativ häufig repliziert wurde, findet sich auch *ADAM33*, jedoch nicht *ORMDL3* oder *TLRs*. Stattdessen sind zahlreiche weitere in der vorliegenden Arbeit nicht betrachtete Gene aufgeführt (Abbildung 31).

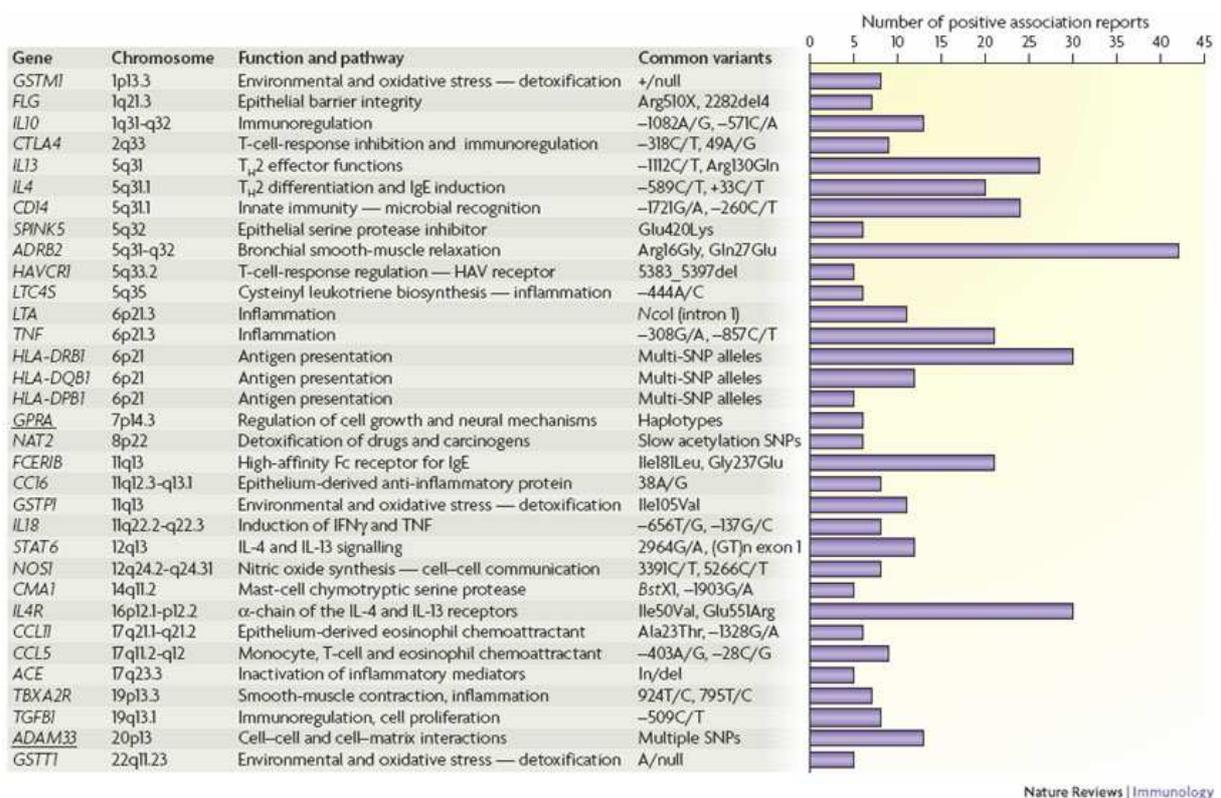


Abbildung 31: Überblick über die am häufigsten replizierten Gene (nach Vercelli (Vercelli 2008))

4.2.2 Varianten der Definition von Replikation

In den Literaturrecherchen wurde nicht nur der Zusammenhang von Asthma mit einem spezifischen Polymorphismus, sondern mit einer beliebigen Variante innerhalb eines Gens als die zu replizierende Einheit herangezogen. So benutzte Vercelli (Vercelli 2008) folgende Definition, um eine Replikation zu definieren: „association of any asthma and asthma-related phenotypes with any variant or combinations of variants within the same gene“. Diesem Ansatz stehen strengere Definitionen gegenüber. Die Verschiebung des Interesses vom SNP zum Haplotyp, in deren Kontext gefordert wurde, dass eher das Gen als der SNP die Einheit der Replikation ist (Neale et al. 2004), bietet einen Erklärungsansatz, warum Replikation nicht ausschließlich auf denselben SNP bezogen wird.

Dies kompliziert die Definition von Replikation, denn verschiedene Entitäten können zur Suche nach Replikationen herangezogen werden. Am besten lässt sich Replikation in unterschiedlichen Variationen der Strenge definieren. Eine Replikationsmatrix (Tabelle 28) beurteilt dabei, wie genau Replikationsstudien im Hinblick auf Genotyp und Phänotyp tatsächlich das ursprüngliche Ergebnis einer Assoziationsstudie wieder zeigen. Noch weiter kompliziert sich die Replikationsdefinition, wenn diese Matrix weiter aufgefächert wird, je nachdem welche genetischen Modelle unterschiedlichen Untersuchungen zugrunde gelegt sind.

Definition des Replikationskriteriums im Hinblick auf den untersuchten Genotyp	Definition des Replikationskriteriums in Hinblick auf den untersuchten Phänotyp	
	streng	milde
streng	gleicher SNP, gleicher Phänotyp	gleicher SNP, anderer Phänotyp
milde	anderer SNP, gleicher Phänotyp	anderer SNP, anderer Phänotyp

Tabelle 28: Replikationsmatrix, um den Grad der Evidenz einer Replikation einzuordnen
In Anlehnung an Holloway (Holloway et al. 2007)

Ob man die Replikationen in einem weiteren oder strengeren Sinne auffassen sollte, ist nicht geklärt und wird noch diskutiert (Vercelli 2008). Geht es darum, einen kausalen Locus zu identifizieren, so dürften engere Definitionen zielführender sein. Die Definition der zu betrachtenden Einheit wirkt sich zwangsläufig auf die Ergebnisse aus. *ADAM33* ist ein sehr gutes Beispiel, wie abhängig Replikationen von deren Definition sind.

4.2.3 *ORMDLs* im Kontext der Replikationsstudien

Starke Assoziation mit Asthma von Polymorphismen des Gens *ORMDL3* und einiger benachbarter SNPs sind das Resultat der GWA, einer hypothesenfreien, statistisch geleiteten Annä-

herung an die Identifizierung von Asthmagenen. Eine Vielzahl von Replikationen dort gefundener Zusammenhänge mit Asthma zeigt, dass GWAs Kandidatengene ausfindig machen können, die konsistent mit Krankheiten assoziiert sind. Einige der durch die GWA identifizierten SNPs genügten dabei sogar dem strengsten Replikationskriterium (Replikation einer Assoziation des gleichen SNPs mit dem gleichen Phänotyp). Zu nennen ist hierbei unter anderem der „top hit“ rs7261389 aus der GWA, dessen Assoziation mit Asthma bereits in sechs Publikationen bestätigt wurde (Bisgaard et al. 2008; Galanter et al. 2008; Hirota et al. 2008; Sleiman et al. 2008; Tavendale et al. 2008; Wu et al. 2008).

Letztgenannter SNP wird nicht dem Gen *ORMDL3*, sondern dem Gen *GSDML* zugeordnet. Jedoch belegen weitere Resultate die herausragende Stellung von *ORMDL3* bei der Entwicklung von Asthma. So konnte auch der Effekt des funktionell relevanten SNP rs8076131 von *ORMDL3* auf Asthma bis jetzt dreimal (Bouzigon et al. 2008; Galanter et al. 2008; Madore et al. 2008) repliziert werden. Beide SNPs (rs7261389, rs8076131) zeigten ebenso wie die meisten mit Asthma assoziierten SNPs auf dem Locus 17q21 zudem eine hohe Konsistenz des Effekts über die verschiedenen Studienzentren (Dresden, Leipzig, München).

SNP rs8076131, der in der Promotorregion von *ORMDL3* liegt, zeigte dabei den stärksten Effekt auf atopisches Asthma in der gepoolten Studienpopulation (DLM4264). Auch dieser Effekt war homogen über die Studienzentren von Kindern aus Dresden, München und Leipzig. Polymorphismen von 17q21 waren zudem nicht nur im Rahmen globaler Genexpressionanalyse bei der ersten GWA zu Asthma stark und konsistent mit SNPs auf dem gleichen Chromosom mit Transkriptionsleveln von *ORMDL3* assoziiert (Moffatt et al. 2007). Der SNP rs7261389 beeinflusste die Expression von *ORMDL3* auch in einer anderweitig durchgeführten genomweiten Expressionsstudie (Duan et al. 2008).

Weitere Untersuchungen eines Promotor-SNPs von *ORMDL3* zeigten, dass in Anwesenheit des C-Allels des SNPs rs8076131 die Promotoraktivität signifikant erhöht ist. Dafür könnte eine stärkere Bindung der Transkriptionsfaktoren USF1 und 2 verantwortlich sein. USF1 reguliert Gene des Glukose- und Lipidmetabolismus und ist einbezogen in Vorgänge, die man als Gründe für Diabetes, eine andere komplexe proinflammatorische Erkrankung, vermutet (Holzapfel et al. 2008). Alternatives Bindungsverhalten der Transkriptionsfaktoren USF1/2 könnte eine Erklärung für die unterschiedlichen Expressionslevels von *ORMDL3* sein, auch wenn damit nicht geklärt ist, wie *ORMDL3* Asthma beeinflusst.

Die Ergebnisse für *ORMDL2* und *ORMDL1* sind deutlich vorsichtiger zu interpretieren. Zwar wurden auch für Polymorphismen dieser Gene Assoziationen mit atopischem Asthma gezeigt, doch fehlt es hier noch an unabhängigen Replikationen und funktionellen Belegen.

Keines der in der ersten genomweiten Studien für Asthma identifizierten Gene auf Chromosom 17 wurde bisher in Übersichtsarbeiten, die sich häufig berichteten positiven Assoziationen von Genen mit Asthma widmen, aufgeführt. Dies dürfte eindeutig darin begründet sein, dass zu diesen Genen zum entsprechenden Zeitpunkt nur begrenzt weitere Studien durchgeführt wurden. Im Rahmen der vielen durchgeführten Assoziationsstudien und rapider Fortschritte, die sich nicht zuletzt in den Ergebnissen der GWAs äußern, ist es sehr schwierig, solche Überblicksarbeiten aktuell zu halten.

ORMDL3 ist ein Gen, das aus einer genomweiten Studie zu Asthma hervorgeht. Mit *CHI3L1* (Chitinase 3-like 1) fehlt in der Übersichtarbeit von Vercelli ein weiteres Gen, das aus einer genomweiten Studie hervorgeht. Da dieses Gen auch im Hinblick auf Asthma untersucht wurde, ist es von hoher Bedeutung für die gegenwärtige Diskussion möglicher Asthmagene.

Ober et al. betrachteten „human cartilage glycoprotein“ (YKL)-40, ein Glykoprotein, als möglichen Marker für Asthma, weil sich die Werte von YKL-40 in der Lunge und im Blutkreislauf von Patienten mit schwerem Asthma erhöht zeigten (Chupp et al. 2007). Sie stellten daher die Hypothese auf, dass *CHI3L1* als das Gen, welches für YKL-40 kodiert, eine wichtige Rolle bei der Entwicklung von Asthma spielen müsse und führten eine genomweite Studie an einer Population von Hutterern, einer tief religiösen Bevölkerungsgemeinschaft, die eher abgeschottet von ihrer Umwelt lebt, durch (Ober et al. 2008). Tatsächlich war ein *CHI3L1* Promotor-SNP (rs4950928) mit erhöhten Werten von YKL-40 im Serum assoziiert. Dieser SNP zeigte auch einen Zusammenhang mit Asthma, BHR und Lungenfunktion. Derselbe Zusammenhang des SNPs mit Asthma konnte in der gleichen Publikation in zwei Fall-Kontroll-Studien nachgewiesen werden. Ober et al. folgerten, dass *CHI3L1* ein Kandidatengen für Asthma, BHR und reduzierte Lungenfunktion ist und erhöhte YKL-40-Werte als Marker für Asthma und verschlechterte Lungenfunktion fungieren können.

4.2.4 *ADAM33* im Kontext der Replikationsstudien

Der Zusammenhang von *ADAM33* mit Asthma zeigt sich sehr konsistent, wenn das Gen als die zu replizierende Einheit herangezogen wird. Jedoch zeigt kein *ADAM33* SNP eine konsistente Assoziation mit Asthma oder einem anderen Phänotyp in den veröffentlichten Replikationsstudien. Dies ist auch aus Tabelle 29 ersichtlich, die einen Überblick über Ergebnisse auf SNP-Ebene in Studien, die vor 2006 publiziert waren, gibt. Auf Einzel-SNP-Ebene ist *ADAM33* damit auch ein passendes Beispiel für anfänglich starke Befunde mit deutlichen Inkonsistenzen in nachfolgenden Studien (Bosse et al. 2007).

Studie	Studienpopulation	N ² Fälle/Kontrollen, Familien	ADAM33 SNPs und berichtete Assoziationen mit Asthma, BHR oder Atopie ¹																
			F+1	I1	L-1	M+1	Q-1	S1	S2	ST+1	ST+4	ST+5	ST+7	T1	T2	T+1	V-1	V1	V4
(Van Eer-dewegh et al. 2002)	US/UK kombiniert (FK)	130/217	neg	neg	neg	neg	A	A	neg	neg	A	neg	A	neg	neg	neg	A	neg	A
	UK (FK)	(nicht berichtet)	A	neg	neg	neg	A	A	A	neg	A	neg	neg	neg	neg	neg	A	neg	A
	US (FK)	(nicht berichtet)	neg	A	A	A	neg	neg	neg	neg	neg	neg	neg	A	A	A	neg	neg	neg
	US/UK (FA)	460fa (1840)						A						B			neg	neg	neg
(Lind et al. 2003)	Mexikanisch (FK)	190/160						neg						neg	neg		neg	neg	neg
	Puerto Ricanisch (FK)	183/165						neg						neg	neg		neg	neg	neg
	Mexik./Puerto R. (FA)	583fa (1749) ³						neg						neg	neg		neg	neg	neg
(Raby et al. 2004)	Nichtspanisch Weiß (FA)	474 (1462) ⁴				neg			neg		neg	neg		neg	neg	neg	neg		neg
	Spanisch (FA)	47 (149) ⁴				neg			neg		neg	neg		A	neg	A	neg		neg
	Afroamerikanisch (FA)	66 (203) ⁴				neg			neg		neg	neg		neg	neg	neg	neg		neg
(Werner et al. 2004)	Deutsch (FK)	48/499	neg	neg		neg	neg	neg	neg	neg	neg	B	A	neg	neg	neg	neg		neg
	Deutsch (FA)	171fa (732)	A,B, AB	neg		neg	neg	neg	neg	B, AB	neg	A	A	neg	neg	neg	neg		neg
(Howard et al. 2003)	Afroamerikanisch (FK)	161/265 ⁵						neg	A		AT		neg	neg	neg		neg		AT
	US Weiß (FK)	220/229 ⁵						neg	A ⁶ ,AT		neg		A	A, AT	A,AT		neg		neg
	US Spanisch (FK)	113/127 ⁵						neg	A		AT		neg	A ⁶ ,AT	A,AT		AT		neg
	Holländer Weiß (FK)	180/133 ⁵						AT	AT		neg		A,AT	neg	neg		AT		A
(Blakey et al. 2005)	Isländisch (FK)	348/262	neg			neg	neg	neg	neg		neg	neg	neg	neg	neg	neg	neg		neg
	UK (Fa)	60fa (240)	neg			neg	neg	neg			neg	neg	neg	neg	neg	neg	neg	neg	neg

Tabelle 29: Widersprüchliche Befunde von Replikationsstudien für SNPs von ADAM33

¹ Assoziation mit Asthma (=A), BHR (=B), Asthma + BHR (=AB) oder Atopie (=AT) in verschiedenen Studien; SNPs, die untersucht wurden, aber nicht signifikant assoziiert waren=neg

² in Fall-Kontroll-Studien die Anzahl von Fällen und Kontrollen, bei Familienstudien die Anzahl von Familien und Individuen

³ 265 Mexikanische Familien und 318 Puerto Ricanische Familien, keine Assoziation weder in einer einzelnen noch in der gepoolten Population

⁴ in der Studie von Raby et al. wurden zusätzliche 8 SNPs untersucht: G1, I1, KL+3, N1, S+1, T+2, V-2, V3. Keiner dieser zusätzlichen SNPs zeigte dort eine Assoziation mit Asthma.

⁵ N bedeutet hier die maximale Anzahl erfolgreich genotypisierter Personen.

⁶ Hier wurde nur ein Trend gefunden ($0,05 \leq p \leq 0,06$)

Auf Haplotypebene lassen sich unterschiedliche Studien meist nicht vergleichen, da unterschiedliche SNPs in die Untersuchungen einbezogen wurden. Dies zeigt einen weiteren Nachteil der weichen Definition von Replikation. Bei einer Replikation für Assoziationen mit denselben SNPs lassen sich auch exakt dieselben Haplotypen untersuchen. Auch Veröffentlichungen jüngerer Zeit bleiben widersprüchlich und behaupten zum Teil die Bedeutung von *ADAM33* bezüglich Asthma (Lee et al. 2006) oder können wiederum keine Assoziation mit Asthma nachweisen (Wang, P. et al. 2006; Hersh et al. 2007). Macht man ein strenges Replikationskriterium zum Maßstab, so ist die Rolle von *ADAM33* sehr zweifelhaft, denn keine Assoziation eines SNPs von *ADAM33* mit einem asthmarelevanten Phänotyp lässt sich konsistent finden.

In weiterem Sinne bleibt *ADAM33* dennoch ein interessanter Kandidat für Asthma, da sich immer wieder Assoziationen mit Asthma zeigen. Die genaue Lokalisation und Funktion eines Locus, der diese Zusammenhänge bedingt, ist aber ungeklärt. Unklar bleibt auch, ob nicht vielleicht eines der benachbarten Gene für die Assoziation mit Asthma verantwortlich gemacht werden kann, da *ADAM33* in hohem LD mit benachbarten Genen ist (Wjst 2007). Die Rolle von *ADAM33* in der Asthmaentwicklung bleibt daher im Gesamtbild kontrovers und weitere genetische Studien werden benötigt, bevor das Gen einhellige Akzeptanz als wichtiger genetischer Regulator von Asthma oder verwandter Phänotypen bekommt (Kere et al. 2004).

Eine systematische Untersuchung der Rolle von *ADAM33* in Form einer Sequenzierung von *ADAM33* und umgebender Regionen mit nachfolgenden Assoziationsstudien von „tagging SNPs“ sowie funktionelle Untersuchungen für bestimmte SNPs, die in Assoziationsstudien häufig repliziert wurden, könnte helfen, die Rolle von *ADAM33* zu präzisieren. Auch eine Einengung des Replikationskriteriums sowie die Untersuchung identischer Haplotypen und Phänotypen könnten hilfreiche Informationen liefern, selbst wenn sich dabei herausstellen sollte, dass *ADAM33* eben nicht bedeutend für die Entwicklung von Asthma ist. Die Mechanismen von *ADAM33* sind sehr komplex und beziehen viele verschiedene Möglichkeiten ein (Cakebread et al. 2004). Neuere Untersuchungen zeigen *ADAM33* auch als Kandidatengen für atopische Dermatitis (Matsusue et al. 2008) oder Psoriasis (Siroux et al. 2008; Li, Y. et al. 2009), so dass es vielleicht durch *ADAM33* bedingte Mechanismen gibt, die sich in verschiedenen Phänotypen äußern.

Insofern müssen die Pathomechanismen, mit denen *ADAM33* assoziiert ist, nicht auf Lungenfunktion oder Umbauprozesse der Atemwege beschränkt sein, wie anfänglich vermutet wurde. Der Krankheitsmechanismus, welcher aus der genetischen Variation, die in den Genen des

„positional cloning“ identifiziert wurde, resultiert, ist weitgehend unbekannt (Laitinen 2007). Ebenso sind die Interaktionen der kodierten Proteine und deren biologische Relevanz in Signalwegen noch wenig verstanden. Dies gilt im Moment für die meisten Gene, welche durch „positional cloning“ identifiziert wurden, obwohl es vereinzelt Hypothesen über deren Funktion gibt.

ADAM33 ist nur ein Beispiel für ein Gen, welches durch „positional cloning“ identifiziert wurde. Nach der Entdeckung von *ADAM33* wurden weitere positionelle Kandidatengene für Asthma oder verwandte Phänotypen entdeckt. Dazu gehören *PHF11* (Zhang, Y. et al. 2003), *DPP10* (Allen et al. 2003), *NPSRI* (Laitinen et al. 2004), auch bekannt unter dem Namen *GPR154* oder *GPRA*, *CYFIP2* (Noguchi et al. 2005), *HLA-G* (Nicolae et al. 2005), *IRAKM* (Balaci et al. 2007) und *COL29A1* (Soderhall et al. 2007). Eine Replikationsstudie, die Assoziationen mit Asthma für fünf Gene, die aus dem „positional cloning“ hervorgehen, in zwei Kohorten untersuchte, liefert Unterstützung für die Bedeutung der genetischen Variation in *GPR154* (*NPSRI*) und *PHF11* für Asthma (Hersh et al. 2007), nicht für *ADAM33*. Am stärksten waren dabei die Bestätigungen für *GPRA*. SNPs von *GPRA* waren in fast allen größeren Replikationsstudien mit Asthma assoziiert, jedoch bleiben die kausale Variante sowie die dafür verantwortlichen Mechanismen noch aufzudecken (Bosse et al. 2007).

Herausforderungen bei Studien, welche Gene untersuchen, die durch „positional cloning“ identifiziert werden, bestehen somit in zweierlei Hinsicht. Einerseits sollte der für einen Phänotyp hypothetisierte Locus weiter eingegrenzt werden. Andererseits ist eine Funktion des Gens oder bestimmter Varianten des Locus zu ergründen.

4.2.5 TLRs im Kontext der Replikationsstudien

Assoziationen von *TLR1*, 6 und 10 Polymorphismen mit Atopie und atopischem Asthma, die Korrekturverfahren für multiples Testen standhielten und homogen über verschiedene Studienzentren beobachtet wurden, gingen aus einer Kandidatengenstudie hervor (Kormann et al. 2008). Bislang gibt es keine Studie, welche diese Ergebnisse von 2008 exakt zu replizieren versuchte. Jedoch gibt es einige Belege, welche Polymorphismen in den Genen *TLR1*, 6 und 10 mit der Entstehung von Asthma in Verbindung bringen. So war der SNP rs5743810 (*TLR6*), der einen deutlichen Zusammenhang nicht nur mit atopischem Asthma, sondern auch mit Asthma zeigte, bereits in anderen Studien mit Asthma assoziiert (Tantisira et al. 2004; Hoffjan et al. 2005). Eine weitere Studie, in der 93 Kandidatengene auf Assoziationen mit Asthma in vier Populationen untersucht wurden, zeigte Evidenz für den Zusammenhang von *TLR10* und atopischem Asthma (Daley et al. 2009). Daneben war ein Haplotyp, der SNPs des Locus auf Chromosom 4, der *TLR6* und *TLR10* einschließt, mit Asthma assoziiert (Puthothu

et al. 2006). Eigene Daten belegen, dass Haplotypen über den gesamten Locus, also mit Einbeziehung der SNPs von *TLR1*, *6* und *10*, nicht nur mit atopischem Asthma, sondern auch mit Asthma assoziiert waren.

Interessanterweise wurden *TLRs* bislang nicht als Hits in einer der genomweiten Studien entdeckt. Auffällig dabei ist, dass keiner der assoziierten SNPs der *TLRs* aus der Publikation 2008 (Kormann et al. 2008) in die Analysen der GWA zu Asthma überhaupt einging. Daher wurden als Annäherung Assoziationen mit Asthma für die auf dem Illumina Chip vorhandenen SNPs durchgeführt, die in Regionen der *TLRs* lagen. Hierbei wurde unter Annahme des dominanten Modells eine signifikante Assoziation für einen SNP von *TLR6* (rs7147) mit Asthma detektiert. Daneben wurden weitere Trends ($0,05 < p < 0,1$) für Asthma für die SNPs rs337609, rs337623, rs9992667 (für *TLR1*), rs337609, rs337623, rs9992667 (für *TLR6*) sowie rs1962053 und rs337609 (für *TLR10*) entdeckt (unveröffentlichte Daten).

Hier zeigt sich zwar eine weniger strenge Form der Replikation, da Assoziation nur auf Genebene und mit einem anderen Phänotyp nachgewiesen wird. Dennoch lassen sich aus den GWA-Daten Hinweise entnehmen, die für eine Bestätigung des Zusammenhangs des *TLR1-6-10*-Clusters mit Asthma sprechen. Dafür spricht weiter, dass Asthma in der GWA-Population eher atopischem Asthma entspricht, da laut subjektiver Einschätzung der Kliniker der Großteil der dort untersuchten Kinder atopisch war. Allerdings ist diese vage Definition nicht ausreichend, um sie mit der genauen Atopie-Messung mittels Prick-Test vergleichen zu können. Um zu den häufig replizierten Asthmagenen gezählt werden zu können, bedarf es noch weiterer Bestätigungen der Assoziation von *TLR1*, *6* oder *10* mit atopischem Asthma, die auch hier um so glaubhafter ausfallen, desto strenger die Replikationskriterien sind.

SNPs von *TLR1*, *6* und *10* liegen alle auf Chromosom 4p4 und zeigen ein hohes Kopplungsungleichgewicht. Während dies einen gemeinsamen Effekt, der von einem Locus ausgelöst wird, vermuten lässt, weisen die unabhängigen Effekte von SNPs in *TLR1* und *TLR6* auf die Zytokinexpression in den in vitro Studien eher auf separate Mechanismen hin. Auch hier sind jedoch weitere Studien auf funktioneller Ebene gefordert. Außerdem steht eine systematische Untersuchung der gesamten Region, z.B. in Form einer Untersuchung mittels „tagging SNPs“ noch aus. Der *TLR6* SNP rs5743810 ist zudem ein Kandidat für weitere funktionelle Studien, da dieser konsistent mit Asthma assoziierte SNP zu einer Aminosäureänderung (Ser/Pro) führt.

Neben dem *TLR1-6-10*-Cluster bleiben auch andere *TLRs* bei der Untersuchung von Zusammenhängen genetischer Determinanten und Asthma von Interesse. Dies geht aus Studien hervor, die *TLR2*, *4* und *9* einbezogen haben, also jene *TLRs*, welche im Rahmen der eigenen

Arbeit zumindest in der nicht für multiples Testen adjustierten Analyse zu signifikanten Ergebnissen führten. Belege von Assoziationen dieser Gene mit atopischen Erkrankungen geben verschiedene Studien, in denen z.B. eine Assoziation von *TLR4* mit Heuschnupfen und Atopie (Senthilselvan et al. 2008) oder die Assoziation des Polymorphismus rs5743836 von *TLR9* mit dem Auftreten von Asthma gezeigt wurde (Lazarus et al. 2003). Es wurde auch gezeigt, wie dabei die Umwelt genetische Assoziation modifizierte (Eder et al. 2004) oder wie die genetische Variation den Umwelteffekt modifizierte (Werner et al. 2003; Bieli et al. 2007) (sogenannte Gen-Umwelt-Interaktion). Exposition mit Endotoxin schützt Atopiker vor Asthma oder Giemen und erhöht das Risiko für nicht atopisches Asthma oder Giemen (Braun-Fahrlander et al. 2002; Douwes et al. 2002; Smit, L. A. et al. 2008). Dies belegt die Wichtigkeit, getrennte Analysen für atopisches und nicht-atopisches Asthma durchzuführen und begründet es, die *TLRs* in Interaktionsanalysen mit der Umwelt einzubeziehen. Häufig wird auch *CD14* als weiteres Gen in der Gruppe der Gene, die für Mustererkennungsrezeptoren kodieren, in die Analysen einbezogen.

In einer Population von Erwachsenen wurden SNPs aus *TRL2*, 4 und 9 zusammen mit *CD14* nicht nur auf ihre Assoziation mit Asthma, sondern auch auf die Modifikation des Umwelteinflusses betrachtet (Smit, L. A. et al. 2008). Dabei wurde die Umweltvariable als Leben auf dem Land im Alter von weniger als einem Jahr definiert.

Ein SNP von *TLR2* (rs3804099) modifizierte bei nicht atopischen Asthmatikern den Umwelteffekt, so dass für nicht-atopisches Asthma dieser einmal protektiv (TT) und einmal ein Risiko (Träger des C-Allels: CT+CC) bedeutete (Smit, L. A. et al. 2008). Der dort assoziierte SNP rs3804099 modifizierte ebenso wie ein weiterer SNP von *TLR2* (rs4696480) innerhalb unserer Population die Effekte von *TLR1*, 6 und 10 Polymorphismen. *TLR2*, *TLR4* und *TLR9* Varianten zeigten für die Haupteffekte bei uns zwar nur Borderline-Assoziationen, möglicherweise liegt deren bedeutsame Rolle aber in der Effektmodifikation anderer SNPs oder umweltbezogener Variablen. Eine Gen-Umwelt-Interaktion wurde auch in einer Studie von Eder (Eder et al. 2004) gezeigt. Bauernkinder, die ein T-Allel des SNPs rs4696480 tragen, sind dabei signifikant weniger gegenüber denen, welchen dieses Allel fehlte, gefährdet, eine Asthmad Diagnose zu erhalten oder akute Asthmasymptome, bzw. Atopie zu entwickeln.

So bleibt die Rolle von *TLR2* für verschiedene Varianten der Interaktionsanalysen von Interesse: einerseits für die Interaktion mit *TLR1*, *TLR6* und *TLR10*, da es durch die Heterodimerisierung auch eine plausible Hypothese gibt, inwieweit SNPs dieser Gene biologisch interagieren könnten, andererseits für die Interaktion mit Umwelteinflüssen.

Die Einbeziehung von Effektmodifikationen in die Auswertungen wurde zu Beginn genomweiter Assoziationsstudien noch als zu komplex erachtet und hatte nicht die gleiche Bedeutung wie die Identifizierung sogenannter Hits. Inzwischen lassen sich Interaktionsanalysen aber in verschiedener Weise genomweit durchführen, wobei hier logistische Regressionen oder die Verwendung von „case only“-Studien vorgeschlagen wurden (Murcray et al. 2009). Gerade im Rahmen des Paradigmas, dass Interaktionen einen wichtigen Beitrag zur Erklärung komplexer Krankheiten leisten können, sind diese Ansätze von hohem Interesse und könnten bereits bekannte Effektmodifikationen von *TLRs* bestätigen. Hingewiesen sei noch auf eine Studie, die *TLR7* und *TLR8* als neue Kandidatengene zeigt (Moller-Larsen et al. 2008). Dieser Befund konnte in unseren Daten jedoch nicht bestätigt werden. So sind die *TLRs* Kandidatengene für Asthma mit hoher Evidenz. Für Polymorphismen von *TLR1*, *6* und *10* wurden sogar bereits Einflüsse auf die Zytokinexpression nachgewiesen. Funktionelle Studien, welche Asthma plausibel begründen, untermauern die Bedeutsamkeit eines Gens.

Die *TLRs* sind nur ein Beispiel für Gene aus einer Kandidatengenstudie, aus der Assoziationen mit Asthma hervorgehen. Es gibt jedoch zahlreiche andere Gene, die in Kandidatengenstudien untersucht wurden. In diesen Arbeiten zu verschiedenen Genen wurden dabei entweder keine Assoziationen mit Asthma gefunden oder solche Assoziationen entdeckt. Am häufigsten wurden Assoziationen für die Gene *HLADRB1*, *ADRB2* und *IL4R* berichtet. Daher soll auf deren Funktion kurz eingegangen werden.

Die klinische Komponente von *HLADRB1*, das im Haupthistokompatibilitätskomplex auf dem kurzen Arm von Chromosom 6 lokalisiert ist, liegt in der Förderung der Allergien auf die Hausstaubmilbe begründet. Das Zytokincluster mit *ADRB2* und *IL4R* liegt auf Chromosom 5q31-33. Der adrenerge β_2 -Rezeptor hat dabei vor allem eine Bedeutung für die Pharmakokinetik, da β_2 -Agonisten als Möglichkeit für die Behandlung der Bronchokonstriktion betrachtet wurden. Hingegen gehört der *IL-4*-Rezeptor (*IL4-R*) neben *IL-4* und *IL-13* zu den Schlüsselgenen der T_H2 -Signalkaskade. Auch *IL-4* und *IL-13* gehören den Genen an, für die in Kandidatengenstudien positive Assoziationen häufig repliziert wurden. *IL-4* und *IL-4R* sind Proteine, von denen nachweislich bekannt ist, dass sie die IgE-Produktion induzieren. Für *IL4-Rezeptor* wurden auch nichtsynonyme SNPs in Exon 12 nachgewiesen (Wenzel et al. 2007). Ein Beispiel für eine strenge Replikation aus dem Kontext der Kandidatengene ist der Zusammenhang von *IL13* und Gesamt-IgE. Dabei zeigte sich ein bestimmter Polymorphismus konsistent mit exakt demselben Phänotyp assoziiert. Diese Beispiele belegen, dass ein Nachweis funktioneller Wirkungsweisen für einen SNP und das Auffinden einer Assoziation dieses SNPs mit einem allergierelevanten Phänotyp hoch korreliert sind.

4.2.6 Methodologische Überlegungen zu GWAs

4.2.6.1 Publikationsbias

Bislang gibt es wenige Untersuchungen, welche die Überlappung von Resultaten, die aus den verschiedenen Ansätzen zur Identifizierung von Asthmagenen entstehen, zeigt. Wenn ein Gen jedoch einen tatsächlich kausalen Zusammenhang mit Asthmaentstehung aufweist, wäre zu hoffen, dass jede der Herangehensweisen auf dieses Gen stößt. Ansonsten wären methodologische Schwächen bei der Suche nach Asthmagenen zu ergründen.

Wie berichtet ließ sich keines der in der ersten genomweiten Studien für Asthma identifizierten Gene auf Chromosom 17 in der Übersichtsarbeit wiederfinden, welche Gene aufführt, deren Assoziation mit Asthma relevanten Phänotypen häufig repliziert wurde. Der oben aufgeführte Überblick (Abbildung 31) wurde 2008 publiziert, baut auf einer zwei Jahre zuvor durchgeführten Publikation auf (Ober et al. 2006) und gibt auch explizit die Kandidaten an, die im Zeitraum zwischen diesen beiden Studien neu untersucht wurden. Dabei haben im Zeitraum von 2006 bis 2008 neue Gene an Relevanz gewonnen, wobei vor allem Filaggrin zu nennen ist. Filaggrin scheint jedoch eine stärkere Rolle bei atopischer Dermatitis als bei Asthma auszuüben (Moffatt 2008). Mit dem Wissen von 2009 dürfte mit ziemlicher Sicherheit *ORMDL3* auch in Überblicksarbeiten aufgeführt sein.

Die Rolle von Genen, die aus solchen Listen als besonders bedeutsam hervorgehen, soll zumindest durch zwei weitere Einwände relativiert werden. Zunächst gibt es einen Publikationsbias zugunsten positiver Berichte, da Negativbefunde seltener veröffentlicht werden. Insofern wäre die Wertigkeit solcher Literaturrecherchen noch höher, wenn negative Befunde in geeigneter Form den positiven Berichten entgegengestellt werden könnten. Maße in Abhängigkeit von Studien, die sowohl Berichte positiver Assoziationen als auch deren Fehlen einschließen, könnten ein anderes Bild vermitteln. Hier sind z.B. einige der Negativstudien zu *ADAM33* zu nennen. Außerdem gibt es ein Ungleichgewicht bezüglich überhaupt durchgeführter Studien. Gene, die seit längerem als Kandidaten gelten, wurden häufiger untersucht und weisen somit eine höhere Wahrscheinlichkeit auf, bei den häufig replizierten Genen vorne zu stehen.

Gerade durch die Möglichkeiten der GWAs ist davon auszugehen, dass in nächster Zeit noch zahlreiche weitere Gene detektiert werden können, die hier nicht aufgeführt sind.

GWAs müssen sich hingegen hinsichtlich anderer Fragen kritisch prüfen lassen. Die GWA basiert auf der Idee, durch Einbeziehung von SNPs, welche das gesamte Genom repräsentieren, auch genomweit auf Assoziation von genetischer Variation mit einem Phänotyp zu testen

und dabei die für eine Krankheit relevanten Effekte zu finden. Damit einher gehen die Fragen, wie komplett die Stichprobe untersuchter SNPs das Genom tatsächlich beschreibt und ob wirklich alle relevanten Effekte durch die GWA gefunden werden.

4.2.6.2 „Coverage“

Im günstigsten Fall kann man bei GWAs den kausativen Locus auf Grund einer entdeckten, hohen Assoziation mit dem untersuchten Phänotyp direkt finden. Dies wäre zum Beispiel der Fall, wenn der auf dem Chip befindliche SNP rs7216389, der „top hit“ der ersten genomweiten GWA zu Asthma sich auch als ursächlich für Asthma herauskristallieren würde. Auch wenn bei GWAs inzwischen bis zu einer Million SNPs gleichzeitig typisiert werden können, so gibt es immer noch Regionen, die durch die Chips nicht erfasst werden.

Gewöhnlich wird jeder SNP auf den Chips sorgfältig von Wissenschaftlern ausgewählt, wobei die Auswahl darauf beruht, ein möglichst hohes „coverage“, über das Genom zu erhalten. Das „coverage“, definiert als der Anteil von häufigen SNPs, die durch die SNPs auf dem Chip abgedeckt sind (Barrett et al. 2006), variiert über das Genom (Li, M. et al. 2008).

Da Chips eher im Hinblick auf ein hohes „coverage“ erstellt sind als im Hinblick auf die Einbeziehung bekannter Effekte aus anderen Studien, kommt es vor, dass Loci, die aus Kandidatengenstudien hervorgingen, auf den Chips gar nicht oder nur teilweise zu finden sind.

So zeigte sich, dass die Ergebnisse aus der Kandidatengenstudie der *TLRs* mit den Ergebnissen der ersten GWA für Asthma nicht direkt abzugleichen waren, da keiner der 21 SNPs, die in der durchgeführten Kandidatengenstudie analysiert wurden, in die Analysen der GWA einging. Auch die in der Kandidatengenstudie untersuchten SNPs von *ADAM33* fehlten auf dem Chip.

Gewöhnlich sollten die Marker dicht genug gesetzt sein, um zumindest durch LD auch die Allelzustände der Marker aufdecken zu können, die nicht typisiert werden. Daher wurden eigene Maße entwickelt, welche das „coverage“ beschreiben (Jorgenson et al. 2006). Das „coverage“ ist ein Maß dafür, wie gut die genotypisierten SNPs alle Varianten im Genom reflektieren und ist bestimmt durch den Grad von LD zwischen dem Set von SNPs, welche genotypisiert werden und den SNPs, die nicht typisiert werden. Häufig werden „tagging SNPs“ verwendet, um kosteneffizient vorgehen zu können, ohne Information zu verlieren. Hierbei sind verschiedene Szenarien denkbar. So kann es auch bei GWAs geschehen, dass man nur die für einen Phänotyp relevante Region streift und eine nachfolgende Feinkartierung notwendig ist, Varianten zu finden, denen man eher eine kausale Rolle für den Phänotyp zuschreibt. So besteht zum Beispiel hohes LD zwischen Varianten von *ORMDL3*, die nicht in

der GWA untersucht wurden und hoch mit Asthma assoziierten SNPs auf Chromosom 17. Wenn jedoch im Vorfeld einer Studie die Korrelationsstruktur durch die gewählten „tagging SNPs“ nicht adäquat beschrieben ist, was auch als „Wirksamkeit“ der „tagging SNPs“ bezeichnet wird, so ist das „coverage“ inadäquat und die Studien haben zu wenig Power. Simulationsstudien zeigen, dass die Wirksamkeit der „tagging SNPs“ sogar noch bei einer Ausgangstichprobe von 100 Individuen verzerrt sein kann, wenn sie nicht richtig berechnet wird (Iles 2008b). Dies relativiert die früher getroffene Aussage, in der etwa 25 Individuen als ausreichend betrachtet wurden, um geeignete „tagging SNPs“ zu bestimmen (Thompson et al. 2003). Und es stellt sich die Frage, ob ein Ansatz, der auf „tagging SNPs“ basiert, für GWAs der beste ist. Falls ein kausaler Locus in einer Region liegt, die nicht adäquat abgedeckt ist, dürften auch GWAs scheitern, diesen zu finden. Es bleibt damit offen, ob GWAs bis zu einer vollständigen Genotypisierung sämtlicher genetischer Variation weiter entwickelt werden müssen, um einen kausalen Krankheitslocus sicher entdecken zu können.

Ein typisches Design bei komplexeren Studien wie GWAs ist das stufenweise Vorgehen, in denen bestimmte Ergebnisse einer ersten Auswertungsphase in einem zweiten Schritt vertieft werden. Auch dieses Vorgehen schließt die Gefahr nicht aus, wichtige Varianten zu übersehen, da der Fokus auf bestimmte meist hochsignifikante Effekte und die dazu gehörigen Regionen gerichtet ist und damit möglicherweise wichtige andere Regionen übersehen werden. Inzwischen sind zwar einige zweistufige Verfahren entwickelt worden, die eine gute Power besitzen (z.B. (Millstein et al. 2006; Murphy et al. 2008; Murcay et al. 2009)) und auch bei GWAs zum Einsatz kommen können (Murcay et al. 2009). Oft genügen diese aber nicht, um alles aufzudecken, was für die Krankheit relevant sein könnte. So ist die Region auf Chromosom 17 sehr genau untersucht worden, während andere „top hits“ bislang kaum weiterverfolgt wurden. Selbstverständlich ist es naheliegend, sich zuerst mit den zumindest im statistischen Sinne überzeugendsten Effekten auseinander zu setzen, aber auch andere Hits bleiben von Interesse. Wie aus anderen Kontexten bekannt, müssen nicht immer die „top hits“ diejenigen sein, die in GWAs nach mehreren Replikationsrunden bestehen bleiben (Hunter et al. 2007; Hafler et al. 2007). In einer unabhängigen Studie konnte beispielsweise die Assoziation von zwei der anderen „top hits“ (rs4512342 in *NRG1* auf Chromosom 8 sowie rs10924993 in *EROILB* auf Chromosom 1) für Asthma zwar nicht repliziert werden (Tavendale et al. 2008). Dies genügt jedoch nicht, bereits darauf zu schließen, dass in der GWA hoch mit Asthma assoziierte Regionen außerhalb von Chromosom 17 für die Entwicklung von Asthma keine Rol-

le spielen. Vertiefende Untersuchungen von weiteren assoziierten Regionen mit der GWA sollten daher unbedingt noch folgen.

Zu bedenken sind bei vertiefenden Analysen neben der Untersuchung einfacher SNP-Phänotyp-Assoziationen auch die Betrachtung von möglichen Kombinationen verschiedener Effekte, Multilocuseffekten oder Effektmodifikationen. Dadurch könnten SNPs von Bedeutung werden, die in der Einzel-SNP-Analyse nur wenig Relevanz für Asthma zeigen.

4.2.6.3 „Common disease – common variant“-Hypothese

Eine GWA kann bislang nur dann erfolgreich sein, wenn die sogenannte „common-disease-common-variant“ (CDCV)-Hypothese gilt. Diese populäre Hypothese nimmt an, dass das Meiste des genetischen Risikos für gängige komplexe Krankheiten durch eine kleine bis mittlere Anzahl von Krankheitsloci, die häufige Varianten haben, verursacht wird (Ziegler et al. 2008). Dies bedeutet, dass viele seltene Varianten bisher noch nicht im Rahmen genomweiter Studien entdeckt werden können (Pearson et al. 2008). Hierbei ist bemerkenswert, dass es nur wenig Evidenz für die Richtigkeit der CDCV-Hypothese gibt.

Zwar gibt es neben Asthma auch andere Krankheiten wie z.B. Alzheimer, für welche häufige Allele als Kandidaten gefunden wurden (Iles 2008a). Genauso gibt es aber Hinweise, dass seltene Allele häufige Störungen wie z.B. geringe Plasma „high density lipoprotein“ (HDL)-Cholesterinwerte bedingen können (Cohen, J. C. et al. 2004).

Das Beispiel von Filaggrin, das mit einer Odds Ratio >3 für atopische Dermatitis zu den am häufigsten replizierten Genen für atopische Dermatitis zählt (Weidinger et al. 2008), belegt die Wichtigkeit, seltene Varianten von den Überlegungen nicht auszuschließen. Dabei zeigen die SNPs von Filaggrin, R501X und s2282del, jeweils eine „minor allele frequency“ von 0,98% oder 2,37% in der Querschnittspopulation aus Münchner und Dresdner Kindern.

Bedenkt man, dass atopische Dermatitis noch eine höhere Prävalenz als Asthma besitzt und dennoch durch diese relativ seltenen Varianten bedingt sein dürfte, ist es um so mehr begründet, auch Polymorphismen mit seltenen Varianten als kausale Loci für Asthma heranzuziehen. Auch Effekte von *TLR2* auf Asthma oder Atopie konnten nicht nur für häufige Varianten gefunden werden. Auch die seltene Variante des SNPs 753Q von *TLR2* erhöhte das Risiko für die Entwicklung von Atopie bei Kindern (Kormann et al. 2009). Einen Hinweis auf mögliche Relevanz niedrig frequenter Polymorphismen gibt darüber hinaus ein SNP von *ORMDL2* (rs12308277), dessen seltenes Allel einen protektiven Effekt auf Asthma ausübt. Dieser SNP hat eine MAF von 1,86%. Die MAF des SNP von *ORMDL1* (rs12990184), der das Risiko für atopisches Asthma modifizierte, lag bei 2,58%.

Diese Zusammenhänge bedürfen natürlich noch einer Validierung durch unabhängige Replikationen. Wenn sich diese Befunde aber bestätigen, so sollte die CDCV-Hypothese stets mit der nötigen Vorsicht vertreten werden, da sich die Frage stellt, ob tatsächlich frequente SNPs die Haupteffekte ausmachen oder GWAs nicht doch seltene SNPs einbeziehen sollten. Bei Einbeziehung niedrig frequenter SNPs in GWAs ließen sich auch Vergleiche anstellen, inwieweit hoch oder niedrig frequente Polymorphismen Assoziationen aufdecken.

4.2.6.4 Allelfrequenz, Power und Effektgrößen

Iles et al. (Iles 2008a) zeigten in einem Überblick über 54 Studien für 22 Krankheiten die Verteilung der Allelfrequenz von 43 mit der Krankheit assoziierten SNPs, deren Befunde fast alle repliziert waren. Die Allelfrequenz dieser in GWAs gefundenen Risikoallele war etwa normal verteilt mit einem Mittelwert von 0,43. In der Mehrheit waren diese Polymorphismen somit hochfrequent und nur drei SNPs dieser Kandidatengene hatten eine $MAF < 0,01$. Was auf den ersten Blick nahelegt, dass tatsächlich die meisten der mit häufigen Krankheiten assoziierten Kandidatengene auch häufige Allelfrequenzen aufweisen, ist bei genauerer Betrachtung ein Powerproblem.

Simulationsstudien für eine häufige Krankheit zeigen, dass die meisten der bisher durchgeführten Studien lediglich die Power besitzen, um häufige Allele als Krankheitsloci zu detektieren, es sei denn ein SNP mit seltener Allelfrequenz besitzt einen sehr starken Effekt. Auch in einer Simulationsstudie zur Aufdeckung von Gen-Umwelt-Interaktionen war die Power bei häufigen Expositionen und häufiger Allelfrequenz am höchsten (Murcay et al. 2009). Einige Autoren vermuten, dass die Power der GWAs noch viel geringer ist als allgemein angenommen (Wang, W. Y. et al. 2005; Jorgenson et al. 2006). Effekte mit häufigen Allelen haben eine höhere Wahrscheinlichkeit, detektiert zu werden. Es ist daher zu befürchten, dass man mit den bisher verwendeten Studiengrößen Effekte mit seltenen Allelen selbst dann übersehen könnte, wenn es sich bei diesen um kausative Loci handelt (Iles 2008a).

Es gibt somit für Assoziationen von Phänotypen mit selteneren Varianten, die kleine Effekte haben, eine geringere Chance, durch GWAs detektiert zu werden, vorausgesetzt natürlich, dass sie frequent genug sind, um auf dem Chip zu sein. Dies ist bemerkenswert, denn bekanntermaßen sind die Effekte in genetischen Assoziationsstudien meist sehr klein. Ein Überblick eigener Auswertungen für Publikationen der Arbeitsgruppe Kabesch zeigt, dass bereits Effekte mit einer $OR > 1,5$ als hoch angesehen werden können (Tabelle 30). Eine Ausnahme bildet lediglich Filaggrin.

Diskussion

Publikation	Gen	Population ¹	Phänotyp	rs-Nummer	genetisches Modell	Odds Ratio	Odds Ratio (positiv)
(Kormann et al. 2009)	<i>TLR2</i>	MD3099	spezifisches IgE (>0.7, RAST 2)	rs7543708	dominant	1,54	1,54
		MD3099	Skin-Prick-Test	rs7543708	dominant	1,52	1,52
(Pinto et al. 2009)	<i>IL15</i>	MD3099	Skin-Prick-Test	rs10833	allelisch	0,88	1,13
(Suttner et al. 2009)	<i>HLX1</i>	MD3099	Asthma	rs2738751	dominant	1,44	1,44
		MD3099	Asthma	rs12141189	dominant	0,73	1,37
		MD3099	nicht-atopisches Asthma	rs2738751	dominant	1,45	1,45
		MD3099	nicht-atopisches Asthma	rs12141189	dominant	0,67	1,49
(Weidinger et al. 2008)	<i>FCERIA²</i>	MD3099	Gesamt-Ige>66.Perzentil	rs2251746	linear	0,84	1,19
		MD3099	Gesamt-Ige>66.Perzentil	rs2427837	linear	0,84	1,19
(Kormann et al. 2008)	<i>TLR1</i>	MDL1872	atopisches Asthma	rs5743595	dominant	0,54	1,85
		MDL1872	atopisches Asthma	rs4833095	dominant	0,59	1,69
	<i>TLR6</i>	MDL1872	atopisches Asthma	rs5743810	dominant	1,79	1,79
		MDL1872	atopisches Asthma	rs5743789	dominant	0,54	1,85
	<i>TLR10</i>	MDL1872	atopisches Asthma	rs4129009	dominant	0,58	1,72
	<i>TLR4</i>	MDL1872	Dr. Diagnose Asthma	rs2737190	dominant	1,32	1,32
	<i>TLR9</i>	MDL1872	Dr. Diagnose Asthma	rs187084	dominant	1,33	1,33
(Weidinger et al. 2008)	<i>Filaggrin³</i>	MD3099	Atopische Dermatitis	R501x §	linear	3,65	3,65
		MD3099	Heuschnupfen	R501x §	linear	3,13	3,13
		MD3099	Dr. Diagnose Asthma	R501x §	linear	2,02	2,02
		MD3099	Aktuell sichtbares Ekzem	R501x §	linear	2,31	2,31
(Schedel et al. 2008)	<i>IRF1</i>	Dresden	spezifisches IgE	rs2070721	dominant	1,25	1,25
		Dresden	Skin-Prick-Test	rs17622656	dominant	0,80	1,25
		München	spezifisches IgE	rs2706384	dominant	1,32	1,32
		München	spezifisches IgE	rs2070721	dominant	1,33	1,33
		München	spezifisches IgE	rs10035166	dominant	1,34	1,34
		München	Skin-Prick-Test	rs2706384	dominant	1,46	1,46
		München	Skin-Prick-Test	rs2070721	dominant	1,41	1,41
		München	Skin-Prick-Test	rs10035166	dominant	1,40	1,40
München	Skin-Prick-Test	rs17622656	dominant	0,67	1,49		

-Tabelle wird fortgesetzt-

Publikation	Gen	Population ¹	Phänotyp	rs-Nummer	genetisches Modell	Odds Ratio	Odds Ratio (positiv)
(Pinto et al. 2007)	<i>STAT1</i>	Dresden	Skin-Prick-Test	G-306A §	allelisch	0,76	1,32
		München	Skin-Prick-Test	rs3771300	allelisch	0,80	1,25
		MD3099	spezifisches IgE	rs3771300	allelisch	0,88	1,14
		MD3099	Skin-Prick-Test	rs3771300	allelisch	0,88	1,14
(Moffatt et al. 2007)	<i>Chromosom 17 (inklusive ORMDL3)</i>	MD (ohne GWA)	Dr. Diagnose Asthma	rs9303277	linear	1,41	1,41
		MD (ohne GWA)	Dr. Diagnose Asthma	rs11557467	linear	1,45	1,45
		MD (ohne GWA)	Dr. Diagnose Asthma	rs2290400	linear	1,47	1,47
		MD (ohne GWA)	Dr. Diagnose Asthma	rs7216389	linear	1,45	1,45
		MD (ohne GWA)	Dr. Diagnose Asthma	rs4795405	linear	1,52	1,52
		MD (ohne GWA)	Dr. Diagnose Asthma	rs8079416	linear	1,30	1,30
		MD (ohne GWA)	Dr. Diagnose Asthma	rs4795408	linear	1,25	1,25
(Depner et al. 2007)	<i>CX3CR1</i>	Dresden	Skin-Prick-Test	rs3732379	dominant	0,77	1,30
		Dresden	Skin-Prick-Test	rs3732378	dominant	0,71	1,41
		Dresden	Gesamt-IgE>66.Perzentil	rs3732378	dominant	0,81	1,23
		Dresden	Gesamt-IgE>90.Perzentil	rs3732378	dominant	0,69	1,45
		Dresden	spezifisches IgE	rs3732379	dominant	0,73	1,37
		Dresden	spezifisches IgE	rs3732378	dominant	0,70	1,43
(Kabesch et al. 2007)	<i>GMCSF, IL5</i>	MD3099	Dr. Diagnose Heuschnupfen	rs25882	allelisch	1,25	1,25
		MD3099	Skin-Prick-Test	rs2069812	allelisch	0,83	1,20
		MAS (alle deutschen)	Dr. Diagnose Heuschnupfen	T3085G §	allelisch	0,70	1,43
(Schedel et al. 2006)	<i>ADAM33</i> ⁴						

Tabelle 30: Überblick über Odds Ratios signifikanter Effekte auf Einzel-SNP-Ebene innerhalb eigener Auswertungen für Publikationen der Arbeitsgruppe Kabesch. Dargestellt sind in Veröffentlichungen zwischen 2006 und 2009 tabellarisch berichtete Ergebnisse signifikanter Effekte, die in Form von Odds Ratios berichtet wurden. Zu Vergleichszwecken wurde für alle Effekte < 1 zusätzlich der Kehrwert der Odds Ratio berechnet.

¹ MD3099= Querschnittspopulation von 3099 Kindern aus München und Dresden, nur München oder nur Dresden beschreibt die jeweiligen Teilpopulationen von 1159 (M1159) oder 1940 (D1940) Kindern, MDL1872=Fall-Kontroll_Population von 1872 Kindern aus München, Dresden und Leipzig , MD (ohne GWA)=Querschnittspopulation von 2359 Kinder aus München und Dresden ohne Kinder, die in der GWA analysiert wurden, MAS=824 deutsche Kinder der mulizentrischen Allergiestudie mit Kindern aus 5 Zentren.

² In der Originalpublikation wird die Assoziation mit Gesamt-IgE berichtet. Für eine Vergleichbarkeit ist hier die Odds Ratio für eine dichotomisierte IgE-Variable berechnet.

³ Aus Gründen der Vergleichbarkeit wurden hier Beispiele für einen SNP aufgeführt. In der Originalpublikation wurden Effekte für eine Kombination der SNP R501x und 2282del4 berichtet. Weitere Kombinationen von Phänotypen und stratifizierte Analysen sind hier nicht dargestellt.

⁴ Für ADAM33 wurden keine Tabellen mit positiven Assoziationen auf Einzel-SNP-Ebene publiziert. Effekte mit nicht atopischem Asthma für S1 (OR 1,53), V4 (OR 1,44) und M+1 (OR 0,60) sind im Range berichteter Werte.

[§] rs-Nummer nicht verfügbar

Sieht man von den Effekten bei Filaggrin ab, dessen Ausnahmestellung für atopische Dermatitis gut etabliert ist, so zeigt sich keine Odds Ratio über 2 und die Mehrheit der gefundenen Assoziationen reflektiert in Relation kleinere Effekte. Eine $OR > 1,5$ zeigen lediglich der *ORMDL3*-SNP rs4795405 und die SNPs einiger *TLRs*. Auch ein Überblick über Effekte von Genen, die aus 54 Studien, die entweder selbst GWAs waren oder aus GWAs folgten und über 22 Krankheiten berücksichtigten (Iles 2008a) verweist auf kleinere Effekte bei GWAs. In diesem Überblick lagen die meisten Odds Ratios der Assoziationsstudien unter 1,5. Nur acht Odds Ratios waren über 1,5 und nur eine über 2. Die Odds Ratios genomweiter Studien liegen typischerweise zwischen 1,2 und 1,3 (Pearson et al. 2008). In diesem Kontext ist der Effekt von *ORMDL3* und den anderen Genen auf dem Chromosom17-Lokus eher über dem Schnitt und somit beachtlich. Möchte man kleine Effekte auch für niedrig frequente SNPs aufdecken, kommt man leicht zu unpraktikabel großen Fallzahlen. So benötigen Allele mit einer $MAF < 0,1$ Stichprobengrößen von mehr als 10000 Fällen und Kontrollen (oder 10000 Familien), um eine für genetische Effekte übliche Odds Ratio von 1,3 oder kleiner aufzudecken (Wang, W. Y. et al. 2005).

Dies wird weiter verschärft, wenn man bedenkt, dass kausale Varianten lediglich über einen „tagging SNP“ zu entdecken sind. Der Effekt des „coverage“ auf die Power wird am besten verstanden, wenn man sich den Faktor von $1/R^2$ bewusst macht, um den eine Stichprobe erhöht werden muss, wenn eine nicht typisierte Variante durch einen genotypisierten SNP aufgedeckt werden soll. Ebenso wird das Powerproblem verschärft, wenn Kombinationen von Effekten oder deren Interaktion betrachtet werden sollen. Bereits bei populationsbasierten Studien hat man das Problem einer reduzierten Power, insbesondere, wenn man sich nicht-additiven Gen-Umwelt-Interaktionen zuwenden möchte (Wong et al. 2003; Cordell et al. 2005). Dieser Befund einer reduzierten Power für stratifizierte Analysen lässt sich ebenfalls auf GWAs übertragen.

Es wurde gezeigt, dass durch geeignete Imputationsverfahren die Power einer Studie erhöht werden kann (Marchini et al. 2007; Servin et al. 2007). Jedoch ist die Qualität der Imputation davon abhängig, wie gut die LD-Informationen, auf welcher die Imputationen basieren, Variation innerhalb der Studienpopulation dokumentiert. Gibt es gute Informationen über LD in der zu schätzenden Region, dann kann auch die Power durch Imputation erhöht werden. Ist dies LD-Information nicht zugänglich, ist kaum davon auszugehen, dass eine Verbesserung der Power zu erreichen ist (Anderson et al. 2008).

Wie bei allen Imputationsverfahren stellt sich auch hier die Frage nach der Genauigkeit. Die Genauigkeit hängt vom Dichtegrad der SNPs sowie von der Übereinstimmung der LD-Muster

zwischen HapMap und realen Daten ab. Besonders vorsichtig gilt es bei niedrig frequenten SNPs zu sein, da deren Schätzungen dazu tendieren weniger genau zu sein (Anderson et al. 2008). Insofern sind Imputationsverfahren unter gewissen Voraussetzungen durchaus eine geeignete Methode, „coverage“ und Power zu erhöhen, aber eben nur dann, wenn bereits ausreichend valide Information zu den gesuchten Daten vorliegt.

4.2.6.5 Kombination der Methoden zur Identifizierung von Asthmagenen

Sämtliche Methoden zur Identifizierung von Asthmagenen zeigen Stärken und Schwächen. Das hohe Potential falsch positiver Resultate, die Forderung nach großen Stichproben und mögliche Verzerrungen entsprechend der Fall-Kontroll-Selektion sind mögliche Beschränkungen der GWAs. Bereits Kandidatengenstudien verlangen nach ausreichend großen Populationen, einer genauen Definition der Phänotypen, ausreichenden Qualitätskontrollen der Genotypen, statistischen Korrekturverfahren sowie Replikationen. Kandidatengenstudien verlangen zudem die Präsentation einer plausiblen biologischen Hypothese. Um Effekte durch eine Kandidatengenstudie aufzudecken, gilt es mindestens, Pathomechanismen im Ansatz zu kennen. Im Vergleich zum genomweiten Ansatz sind sie damit einerseits effektiv, da man nach Aufdecken einer Assoziation bereits den dahinter liegenden Pathomechanismus kennt oder zumindest zu kennen glaubt. Andererseits sind sie auch durch das existierende Wissen begrenzt, da der vermutete Mechanismus gelegentlich nur auf Hypothesen beruht und nicht unbedingt zuverlässig nachgewiesen ist. Zudem werden bei Kandidatengenstudien nur Loci untersucht für die überhaupt eine plausible Hypothese besteht, warum sie etwas mit der Krankheit zu tun haben könnten. Dies wird noch mehr zur Schwäche, wenn man bedenkt, dass die meisten komplexen Merkmale als polygen bedingt vermutet werden. Dass es vermutlich noch zahlreiche unbekannte pathophysiologische Mechanismen gibt, zeigen auch die Entdeckungen neuer Kandidatengene in genomweiten Studien. Genomweite Studien benötigen keine Vorannahmen und decken daher alles auf, was im statistischen Sinne relevant ist. Diese „blinde“ Vorgehensweise ist aber auch ein Hauptnachteil der genomweiten Studien, der sowohl für GWAs als auch das „positional cloning“ gilt. Denn bei diesen Ansätzen fehlt im Vorfeld die Information über die Genfunktion. Biologische Funktionen können daher erst im Anschluss an diese Studien ergründet werden. Eine Kombination verschiedener Ansätze galt bereits vor den GWAs als eine gute Vorgehensweise (Blumenthal 2005) und soll für die aktuelle Situation vorgeschlagen werden (Abbildung 32).

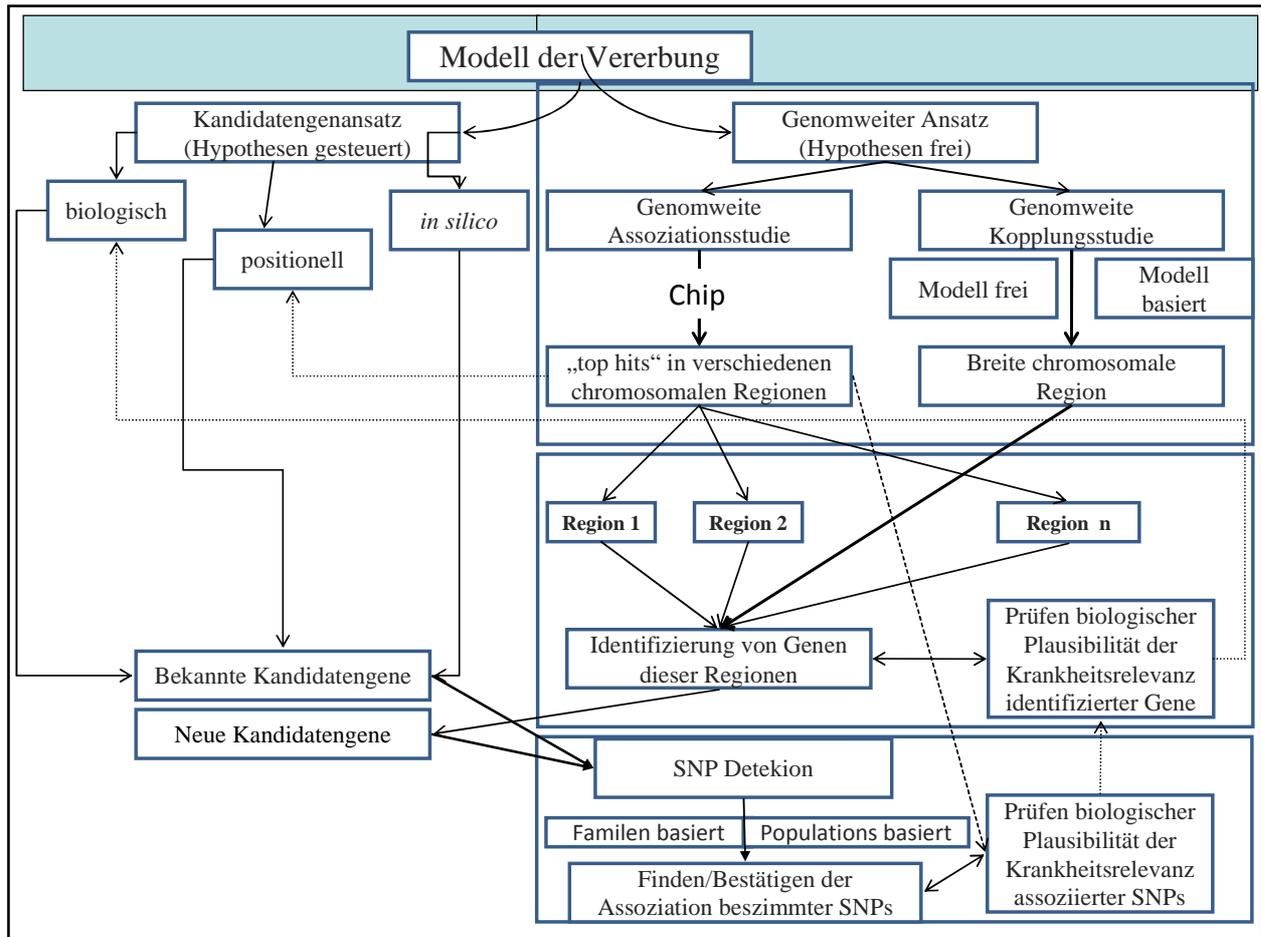


Abbildung 32: Kombination von Kandidatengenansatz mit genomweiten Ansätzen.
 Nicht dargestellt sind hier die genomweiten Expressionsstudien, die einen Erklärungsansatz für Relevanz von SNPs bieten.

GWAs repräsentieren eine Methode, die es in kurzer Zeit ermöglicht, viele mit einer Krankheit assoziierte genetische Varianten aufzudecken. Auf Grund der Vielzahl einbezogener SNPs ist anzunehmen, dass dabei auch neue Loci entdeckt werden, die im Kontext der Krankheit eine Rolle spielen. Kopplungsstudien sind vom Design eher dafür geeignet, seltene Varianten mit großen Effekten aufzudecken, während die GWAs eher Effekte von häufigen Varianten anzeigen (Iles 2008a). Von familienbasierten Kopplungsstudien erhofft man sich vor allem die Aufdeckung klarer Effekte, die gleichbedeutend mit substantiellen genetischen Effekten für Assoziationsstudien sind (Wang, W. Y. et al. 2003). Dennoch ist nicht auszuschließen, dass das teure Ressourcen-intensive Vorgehen des „positional cloning“ irgendwann komplett von den GWAs abgelöst wird.

Eine Kombination von Kopplungs- und Assoziationsstudien galt bislang als effektive Strategie, welche die Präzision für die Identifizierung eines krankheitsrelevanten Locus verbessern kann (Lou et al. 2006). Beide Verfahren führen zur Identifizierung von Regionen, die in Zusammenhang mit einem Phänotyp stehen, wobei einer Phase der Grobkartierung eine Feinkar-

tierung folgt. Bei der Feinkartierung können hoch verdichtete Marker in der gleichen Region und alternative analytische Methoden wie Analysen von LD die Spanne des Vertrauensintervalls weiter verkleinern, wobei diese reduzierten Regionen immer noch mehrere Gene enthalten können (Ron et al. 2007).

Ist eine Region identifiziert, die in Zusammenhang mit einer Krankheit gebracht werden kann, so bilden Gene in dieser Region neue positionelle Kandidatengene. Werden für bestimmte Gene dieser Region auch plausible biologische Erklärungsmuster gefunden, inwieweit sie in das Krankheitsgeschehen einbezogen sein können, so entstehen neue biologische Kandidatengene. Biologische Plausibilität kann es auch für Gene geben, die nicht innerhalb einer identifizierten Region liegen, aber durch biologische Charakteristika mit Genen der identifizierten Region verwandt sind. So sind *ORMDL1* und *ORMDL2* neue Kandidatengene auf Grund ihrer Ähnlichkeit zu *ORMDL3*.

Für die Fortführung hypothesenbasierter Kandidatengenstudien sprechen zumindest zwei Gründe: Einerseits ist bislang ungeklärt, inwieweit sich wirklich alle Effekte durch genomweite Studien aufdecken lassen. Wenn es tatsächlich die seltenen Varianten wären, welche die häufig vorkommenden Krankheiten erklären, so hätten die meisten GWAs (zu) wenig Power um die entscheidenden Varianten aufzudecken. Kandidatengenstudien bleiben somit gerade für seltene Varianten essentiell. In weiteren Kandidatengenstudien sollten traditionelle Methoden mit Daten aus dem „fine mapping“ kombiniert werden. Zudem sollten alle Ressourcen verwendet werden. Dazu gehören Ideen aus Tiermodellen, Hinweise aus Literatur und Bioinformatik, aber auch durch Hochdurchsatz genomweiter Studien gewonnene Ideen inklusive sequenzbasierter und Genexpressionsdaten (Zhu et al. 2007). Funktionelle Studien, die zwar teuer, aber unabdingbar sind, wenn man das Gesamtbild verstehen will, sollten erst durchgeführt werden, sobald es dazu genügend Belege durch andere Untersuchungen gibt.

4.3 Funktionen der Gene

4.3.1 Klassifikation von Kandidatengen

Das Ziel der (genomweiten) Assoziationsstudien bleibt die volle Beschreibung der Bau- und Funktionsweise aller Kandidaten für größere biomedizinische Phänotypen und die darauf basierende Umsetzung der Funde in die klinische Praxis. Gerade im Hinblick auf die Funktion von neuen Kandidatengen, die durch GWAs identifiziert werden, sind noch viele Fragen offen. So ist es schwierig, die Funktion der Gene auf dem identifizierten Locus auf Chromosom 17q21 mit bekannten Mechanismen abzugleichen.

Verschiedene Versuche wurden unternommen, Kandidatengene zu gruppieren. Eine Einteilung, in der vier Gruppen von Genen definiert werden, stammt von Moffatt et al. (Moffatt 2008): Die erste Gruppe von Genen ist in das Auslösen der Immunantwort und die Differenzierung CD4-positiver T-Helfer Zellen involviert.

Diese Gruppe schließt sowohl Gene, die für Mustererkennungsrezeptoren kodieren (*CD14*, *TLR2*, *TLR4*, *TLR6*, *TLR10*, *NOD1* und *NOD2*) wie auch immunregulatorische Zytokine (*IL10* and *TGF β 1*), den Transkriptionsfaktor *STAT3*, den Prostaglandinrezeptor *PTGER2* und Moleküle, die in der Antigenpräsentation eine Rolle spielen (*HLA-DR*, *HLA-DQ* und *HLA-DP* Allele), ein.

Die zweite Gruppe umfasst Gene, welche die T_H2-Zelldifferenzierung und T_H2-Zellen-Effektorfunktionen regulieren (*GATA3*, *TBX21*, *IL4*, *IL13*, *IL4RA*, *FCER1B*, *IL5*, *IL5RA*, *STAT6* and *IL12B*). Die dritte Gruppe ist in Epithelzellen exprimiert und schließt Chemokine (*CCL5*, *CCL11*, *CCL24* and *CCL26*), antimikrobielle Peptide (*DEFB1*), *CC16* und Faktoren ein, die für das Aufrechterhalten der Integrität der epithelialen Zellbarriere verantwortlich gemacht werden (*SPINK5* and *FLG*). Die vierte Gruppe enthält Kandidatengene, welche durch „positional cloning“ entdeckt wurden (*ADAM33*, *DPP10*, *PHF11*, *GPRA*, *HLA-G*, *IRAKM* und *COL29A1*). Diese Gene sind im Epithel und/oder der glatten Muskulatur exprimiert. Alternative Klassifikationen verwenden ähnliche Gruppen, wobei eine Gruppe von Genen, die mit Lungenfunktion, Umbauprozessen der Atemwege und Schwere der Krankheit verknüpft sind, zu ergänzen ist (Vercelli 2008). Während *ORMDL3* hier (noch) nicht zuzuordnen ist, lassen sich die Toll-like Rezeptoren als wichtige Regulatoren der Immunantwort wieder finden. Auch *ADAM33* lässt sich durch seine Rolle in Zusammenhang mit der Lungenfunktion und auf das „airway remodelling“ einordnen.

Ein etwas differenzierteres Schema zeigt Abbildung 33 (Bosse et al. 2007).

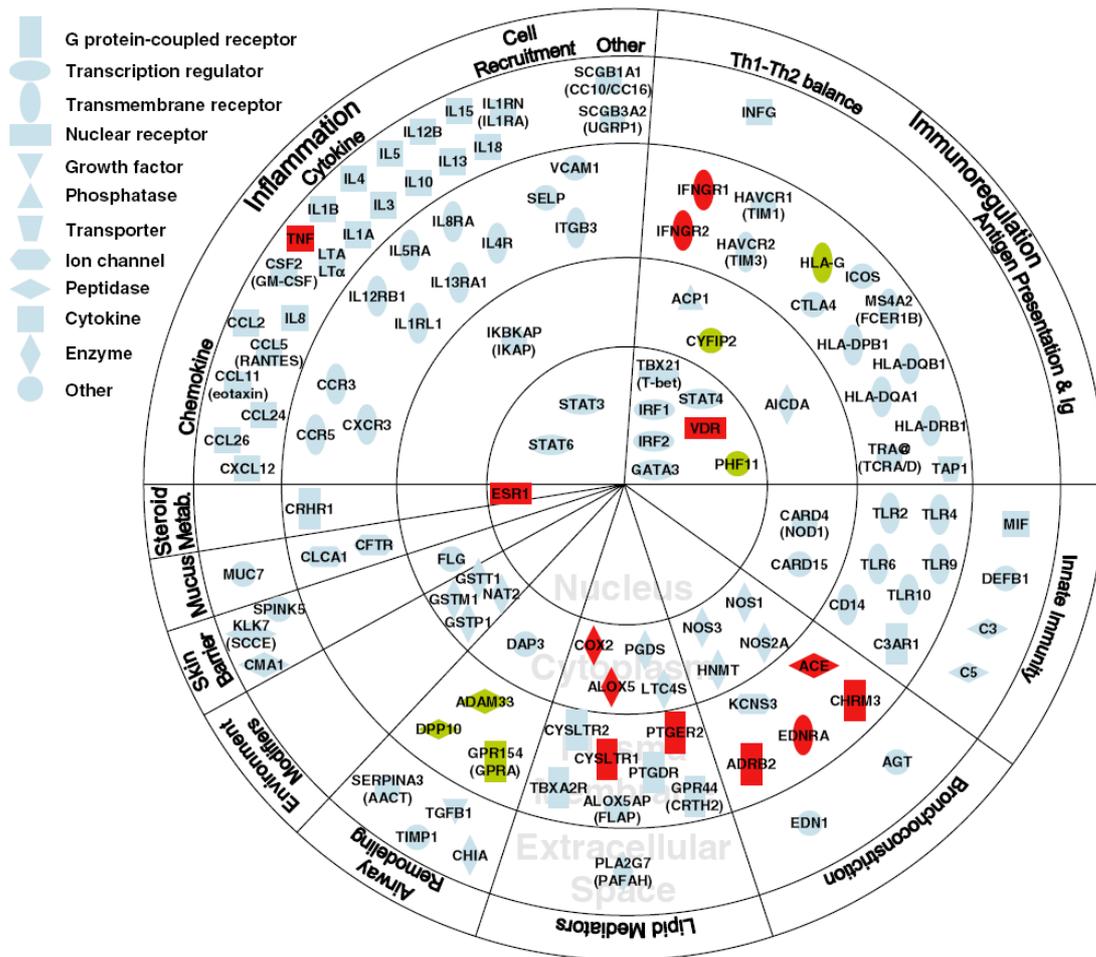


Abbildung 33: Klassifikation von Kandidatengenen nach funktionellen Mechanismen.

Gene, die mit Asthma oder atopischen Phänotypen assoziiert wurden, können nach Funktion und subzellulärer Lokalisation entsprechend dem Ingenuity System (<http://www.ingenuity.com>) kategorisiert werden. Die Kreise bedeuten (von innen nach außen): Wirkung im Nucleus (z.B. Transkriptionsfaktoren), im Zytoplasma (also in der Zelle, außerhalb des Nucleus), in der Plasmamembran (viele Rezeptoren), und außerhalb der Zelle (extrazellulär, z.B. Zytokine).

Alle berichteten Gene waren 2006 zumindest einmal als assoziiert berichtet. Genaue Referenzen finden sich bei Ober (Ober et al. 2006).

Gene sind mit den offiziellen Symbolen bezeichnet, alternative Namen stehen in Klammern.

Gene, welche durch „positional cloning“ identifiziert wurden, sind grün markiert.

Rote Knoten indizieren, dass ein „Food and Drug Administration“ (FDA)-anerkanntes Medikament gegen das Genprodukt wirkt (FDA ist die US-amerikanische Behörde für Lebensmittel- und Arzneimittelsicherheit)

nach Bosse (Bosse et al. 2007)

Man sieht dort, dass die meisten Kandidatengene der Vermittlung von Entzündung und Immunregulation zugeordnet werden. Interessanterweise lassen sich auch hier Gene finden, die nicht primär in der Lunge aktiv sind. Vermutlich existieren demnach auch Pathomechanismen für Lungenerkrankungen, die nicht spezifisch in der Lunge selbst begründet sind. Dazu gehören zum Beispiel Barrierefunktionen der Haut. Obwohl die normale Funktion und die Rolle dieser Gene bei der Krankheit noch weiter beleuchtet werden muss, gibt es auch eine Übereinstimmung darin, dass zahlreiche Gene Einfluss auf das Epithel nehmen (Moffatt 2008).

Jüngste Erfolge bei der Genidentifizierung belegen den großen Einfluss, den der epitheliale Barrieredefekt auf die Ätiologie von atopischer Dermatitis hat. Sie stellen einen Meilenstein bei der Aufdeckung der genetischen Ursachen komplexer Erkrankungen dar und ermöglichen eine neue Sicht auf die molekularen Mechanismen, die zur Krankheitsentstehung führen. Darüber hinaus können sie wegweisend für die Entwicklung neuer Behandlungs- und Präventionsstrategien sein.

4.3.2 Bedeutung der Gene auf Chromosom 17

Wie gezeigt ist die Einteilung von Genen in Gruppen nach funktionellen Mechanismen für eine erste Charakterisierung hilfreich. Genauer untergliederte Abbildungen zeigen allerdings eine große Vielfalt an möglichen Mechanismen und es bleibt unklar, welche Einflüsse dabei stärker oder schwächer sind. Eine hohe Wahrscheinlichkeit besteht auch für ein Zusammenspiel mehrerer Faktoren. Außerdem ist unklar, ob die Mechanismen ubiquitär oder lokal ausgeübt werden. Vor allem aber stellt sich im Kontext dieser Arbeit die Frage, wie Gene, die im Rahmen einer GWA auf Chromosom 17 identifiziert wurden, hier einzuordnen sind.

Verschiedene Gene wurden zu Beginn auf dem durch die GWA für Asthma entdeckten Locus auf Chromosom 17 beschrieben. Dazu gehören *GSDM1*, *ORMDL3*, *GSDML* sowie *ZNFN1A3*, denen später noch *ZBP2* hinzugefügt wurde.

Während für Polymorphismen in der Region von *GSDM1* und für das im Chromosom weiter aufwärts gelegene Gen *TOP2A* in Replikationsstudien die Asthmaeffekte kaum repliziert werden konnten, sind die Assoziationsergebnisse für drei verbleibende Gene (*ORMDL3*, *GSDML* und *ZNFN1A3*) sehr stark und konsistent.

Vor allem die eigenen Befunde für den potentiellen Promotor-SNP rs8076131, aber auch Replikationen mit dem benachbarten SNP rs4378650 (Galanter et al. 2008; Madore et al. 2008; Wu et al. 2008), der in recht hohem LD zu rs8076131 steht, belegen die Bedeutung von *ORMDL3*. Zudem lassen genomweite Studien der Genexpression eine Regulierung der Expression von *ORMDL3* durch SNPs in potentiell cis-regulatorischen (also auf dem gleichen Chromosom befindlichen) Elementen vermuten. Die eigentliche Funktion von Genen der *ORMDL*-Familie ist zwar ungeklärt, aber die starke Verbreitung in verschiedenen Geweben spricht für einen zentral wirkenden Mechanismus, den es jedoch noch zu ergründen gilt.

Der anfängliche „top hit“ rs7216389 wurde vielfach als Asthma relevant bestätigt und gehört zu dem Gen *GSDML*. Ähnlich wie bei *ORMDL3* ist auch die Funktion der *GSDML*-Gene bislang weitgehend unklar. *GSDML* gehört zu einer Proteinfamilie, welche durch eine konservierte Gasdermindomäne definiert ist und aus fünf Mitgliedern besteht (*DFNA5*, *DFNA5L*, *MLZE*, *GSDM* und *GSDML*), wobei die Funktion von *GSDML* am wenigsten bekannt ist

(Sun, Q. et al. 2008). Diese Proteinfamilie wird insbesondere mit Krebserkrankungen in Verbindung gebracht (Carl-McGrath et al. 2008) und hohe Expression von *GSDML* ist mit Tumorprogression assoziiert (Sun, Q. et al. 2008). Dieser Einfluss des Gens auf verschiedene Krankheiten spricht für einen breiter gefächerten Krankheitsmechanismus. Dafür würde auch sprechen, dass neben der Assoziation von Krebs mit Loci auf Chromosom 17 auch für Morbus Crohn Assoziationen mit Loci auf Chromosom 17 (17q11 and 17q23) nachgewiesen wurden (Mathew 2008), die in einer kanadischen und deutschen Population Gültigkeit hatten (Raelson et al. 2007).

Als weiteres spannendes Kandidateng für Asthma auf dem Locus 17q21 stellt sich *ZNFNIA3* dar. Mit rs9303277 und rs11557467 sind zwei der hoch assoziierten SNPs auf Aiolos (*ZNFNIA3/IKZF3*) lokalisiert, und auch hier belegen Replikationsstudien deren Bedeutung (Bisgaard et al. 2008; Bouzigon et al. 2008; Madore et al. 2008; Sleiman et al. 2008). Aiolos kodiert für ein Zinkfinger-DNA-bindendes Protein, das in reifen B-Zellen exprimiert ist und homolog zu Ikaros ist. Transkriptionsfaktor Ikaros ist das Gründungsmitglied einer Familie von Zink-Finger-DNA-Bindungsproteinen, die für die normale Lymphozyten-Entwicklung notwendig sind.

T-Zell abhängige B-Zell Reaktionen, einschließlich der Bildung von Keimzentren und erhöhter IgG- und IgE-Werte wurden bei nicht immunisierten Aiolos-defizienten Mäusen detektiert (Wang, J. H. et al. 1998). Somit ist *ZNFNIA3* ein interessantes Kandidateng (Wjst 2008). Zinkfinger sind häufige Proteine beim Menschen, weshalb auch hier ein Pathomechanismus vermutet werden kann, der an verschiedenen Stellen ansetzt.

Die ubiquitäre Genexpression, die Assoziation desselben Gens mit verschiedenen Krankheiten und die Tatsache, dass dieses Gen für ein Protein einer häufig vorkommenden Proteinfamilie kodiert, sprechen für einen Pathomechanismus, der nicht sehr spezifisch sein muss.

Wenn aus GWAs identifizierte Gene in Zusammenhang mit unspezifischen Mechanismen zu bringen sind, so könnte man für Gene, die aus Studien des „positional cloning“ entstammen, Ähnliches vermuten.

Eine Schnittmenge mit der GWA besteht darin, dass *PHF11* und *ZNFNIA3* beide einer Familie von Zinkfingern entstammen. Freilich ist letzterer Befund nicht allzu überraschend, da geschätzt wird, dass das menschliche Genom mindestens 700 Zinkfinger enthält (Schmidt et al. 2004). Dennoch gibt es frappierende Hinweise darauf, dass Gene aus Familien der Zinkfinger zu wichtigen Kandidaten für Asthma zählen. *PHF11* ist ein Mitglied der Familie von pflanzlichen Homeodomänen, welche nach der Präsenz von einem oder mehreren Zinkfingern des „plant homeodomain“ (PHD)-Typs benannt ist. Proteine mit einem PHD-Finger sind häu-

fig im Nukleus aktiv oder in die Regulation der Transkription einbezogen (Bienz 2006). *PHF11* bindet Zinkionen (<http://www.uniprot.org/uniprot/Q9UIL8>, 15.2.2009). Da *PHF11* sowohl in die T- als auch B-Zell-Expression einbezogen ist, ist die Hypothese, dass *PHF11* die Transkription von T- oder B-Zellen reguliert, recht attraktiv. Zudem wird in einer funktionellen Studie gefolgert, dass *PHF11* ein Regulator der Zytokinenexpression vom T_H1 -Typ ist. *PHF11* ist in T_H1 - höher als in T_H2 -Zellen exprimiert. Daher könnte eine Reduktion von *PHF11*-Expressionslevels zu den starken T_H2 -Reaktionen beitragen, welche allergische Patienten charakterisieren (Clarke et al. 2008).

Auch *ZNFNIA3* spielt als Transkriptionsfaktor eine wichtige Rolle bei der Differenzierung von Lymphozyten (<http://www.uniprot.org/uniprot/Q9UKT9>, 27.2.2009).

Beide Gene sind somit in plausible biologische Mechanismen bei der Entwicklung von Asthma einbezogen, und beide Gene haben in weiterem Sinne etwas mit Zink zu tun. In diesem Falle würde die CDCV-Hypothese doch greifen. Ein Mechanismus, in dem Zink involviert ist, könnte ein sehr verbreiteter Mechanismus sein, der sich auf das Immunsystem auswirkt, an verschiedenen Stellen ansetzt und somit auch durch verschiedene Gene vermittelt wird. Für einen solchen verbreiteten Mechanismus sollten dann auch häufige Variationen kodieren und dementsprechend sollten die GWAs dann diese Assoziationen auch detektieren können. Wachsende Evidenz verbindet Asthma und Zinkdefizienz (Devirgiliis et al. 2007). Beispielfhaft seien hier die Erhöhung des kombinierten Risikos von Atopie, BHR und Allergie typischer Symptome von Individuen mit geringer Zinkeinnahme im Vergleich zu solchen mit hoher Zinkeinnahme (Soutar et al. 1997) erwähnt. Zudem zeigt sich eine negative Assoziation von mütterlicher Zinkeinnahme während der Schwangerschaft mit Kurzatmigkeit und Asthma in einer Follow-up-Studie bei 5-jährigen Kindern (Devereux et al. 2006). Möglicherweise steht die Zinkdefizienz hier nur als Marker für ein geschwächtes Immunsystem. Bekanntermaßen sind Zinkspiegel auch bei Kindern mit atopischer Dermatitis deutlich erniedrigt.

Wenn sich Zinkdefizienz als ein Marker für Asthma bestätigen würde, so wären Einflüsse auf andere Krankheitsbilder nicht überraschend. Hier ließe sich auch wieder *ADAM33* einordnen, denn in seiner Eigenschaft als Metalloproteasen besitzt es ebenfalls eine Zinkbindungsstelle. Interessanter Weise gibt es Belege, dass diese Gene auch für atopische Dermatitis relevant sein könnten, denn ein Zusammenhang mit atopischer Dermatitis wurde sowohl für Polymorphismen von *PHF11* (Jang et al. 2005) als auch für einen Polymorphismus von *ADAM33* (rs2853209) gezeigt (Matsusue et al. 2008). Die Reduktion der Expression von *PHF11*, welche bei Genotypen gesehen wird, die mit atopischer Dermatitis assoziiert sind, könnte zu den starken T_H2 -Reaktionen beitragen, die viele allergische Individuen charakterisieren (Clarke et

al. 2008). Verfolgt man diesen Gedanken weiter, so ist auch das Zinkfingerprotein Gfi1 zu nennen, das unter anderem in der Lunge exprimiert ist und eine Rolle bei der Makrophagen-abhängigen Zytokinproduktion und bei der Regulation der Toll-like-Rezeptor-Signalwege spielt (Moroy et al. 2008). Mit ZCCHC11 scheint ein weiteres Zinkfingerprotein in die Regulation der TLR-Signalwege einbezogen (Minoda et al. 2006). Auch wenn im Mosaik der komplexen Pathogenese von Asthma Zink-vermittelte Mechanismen nur ein Baustein von vielen sind, könnte es sich lohnen, solche Wirkungsweisen weiter zu untersuchen, um spezifische Mechanismen aufzudecken, die in Prävention und Therapie von Immundefekten eine Rolle spielen.

Die mögliche Rolle der TLRs wurde nicht nur für Asthma, sondern auch für rheumatoide Arthritis, Morbus Crohn, ulcerative colitis, Psoriasis, COPD und Atherosklerose (Andreacos et al. 2004) untersucht. Zudem wurden TLRs auf Zusammenhang mit Krebs getestet (Romagne 2007). Polymorphismen von *TLR1* und *TLR10* übten dabei einen protektiven Effekt auf das Risiko von Prostatakrebs aus (Stevens et al. 2008). Einen solchen protektiven Effekt zeigte interessanterweise *TLR1* SNP rs5743595, für den wir den protektiven Effekt auf Atopie und atopisches Asthma zeigen konnten. Dies kann als weiterer Hinweis auf grundlegende Mechanismen des Immunsystems gelten, welche zumindest einige der Kandidatengene für Asthma reflektieren könnten. Allgemeine Mechanismen, insbesondere breitere Immundefekte, die sich dann eben auch bei Asthma, bzw. einer Komponente von Asthma manifestieren, könnten dadurch gefunden werden. Dies gilt vermutlich noch stärker bei hypothesenfrei durchgeführten Studien (GWAs), da sie aus medizinischer Sicht sehr unspezifisch jegliche Assoziationen aufdecken.

4.4 Rückschlüsse auf weiteres Vorgehen bei GWAs

So ist zu vermuten, dass im Rahmen von Assoziationsstudien gelegentlich Mechanismen identifiziert werden, die sich in verschiedenen Pathologien äußern. Auch Asthma ist eher ein Syndrom als ein einheitlich definiertes Krankheitsbild und viele Genvarianten mit kleinen Effekten können vermutlich das komplexe Bild größtenteils besser erklären als einige wenige mit durchschlagenden Effekten (von Mutius, E. 2009). Daher ist anzunehmen, dass auch breitere Erklärungsansätze nicht alle Varianten eines Krankheitsbildes erklären werden. Viele komplexe Erkrankungen werden als Resultat eines Zusammenspiels der kombinierten Effekte von Genen, Umweltfaktoren und deren Interaktionen betrachtet (Murcay et al. 2009). Auch für Asthma besteht ein Konsens über dieses Zusammenspiel (Koppelman 2006; Guerra et al. 2008; Martinez 2008; von Mutius, E. 2009).

Die Variation in Sequenzen der pathophysiologisch relevanten Gene wirkt sich auf spezifische Manifestationen des Phänotyps aus und nimmt dabei gleichzeitige Einflüsse von biologischen Systemen und Umweltfaktoren auf (Guerra et al. 2008).

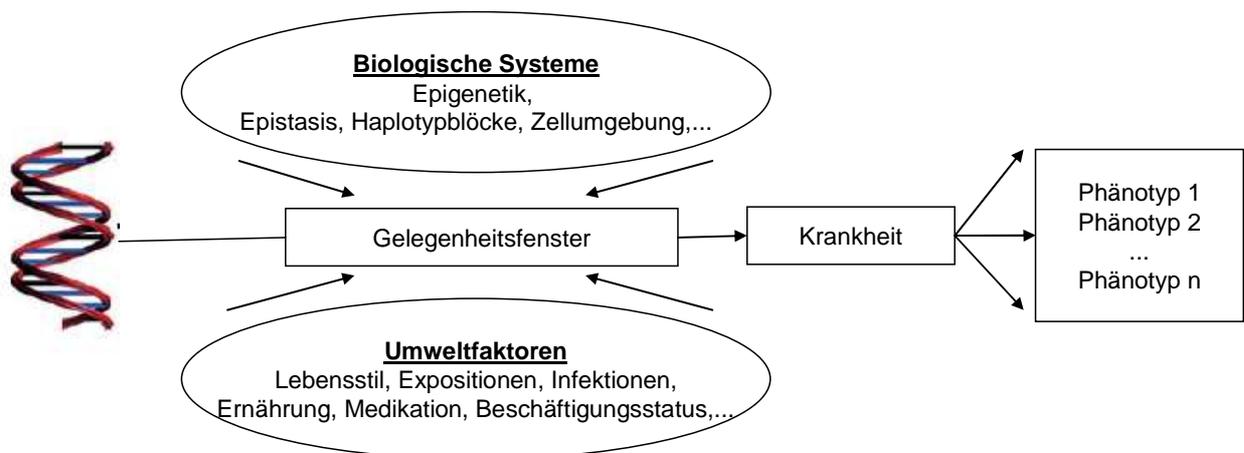


Abbildung 34: Integrierter Ansatz zur Untersuchung von Genotyp-Phänotyp-Zusammenhängen.
 Die Variation in der Sequenz wirkt sich auf spezifische Manifestationen des Phänotyps aus und nimmt dabei gleichzeitige Einflüsse von biologischen Systemen und Umweltfaktoren auf
 (Skizze in Anlehnung an Guerra (Guerra et al. 2008))

Hier ist jedoch die methodologische Seite noch nicht berücksichtigt. Dadurch wird die Komplexität, die sich im Zusammenhang mit der Untersuchung genetischer Determinanten von Asthma ergibt, noch weiter ausdifferenziert. Das Finden einer Assoziation zwischen genetischen Varianten und Asthma oder einem Asthma-verwandten Phänotyp ist kein einfacher Weg, sondern stets beeinflusst durch die Komplexität der Krankheit und methodologischen Fragen. Die Effekte scheinen oft davon abzuhängen, in welchem Kontext man Asthma untersucht (Eder et al. 2006) und welche Methoden man dabei verwendet. Genetische Assoziationsstudien versuchen eine lineare Beziehung zwischen genetischen Varianten und

Asthma aufzustellen. Jedoch ist die Genotyp-Phänotyp-Beziehung durch zwei Kategorien beeinflusst: die Komplexität von Asthma und methodologische Merkmale (Bosse et al. 2007). Letztlich ist auch dies nur ein Modell und die zehn Basisfragen an die GWA, die es zu stellen gilt (Pearson et al. 2008), nur Leitlinien jeder Studie.

1. Sind die Fälle klar definiert, so dass sie mit Fällen verglichen werden können, die typischer Weise in der klinischen Praxis gesehen werden?
2. Sind Fälle und Kontrollen in wichtigen Charakteristika, die sich auf die Krankheit oder die genetische Variation beziehen könnten, vergleichbar?
3. War die Studie von ausreichender Stichprobengröße, um auch kleine Effekte aufdecken zu können?
4. Deckte die Genotypisierungsplattform hinreichend dicht alle Regionen ab, um den Großteil der Varianz in der studierten Population zu erfassen?
5. Wurden geeignete Qualitätskontrollen bezüglich der Genotypisierung durchgeführt (inklusive visueller Inspektion der „cluster plots“ und Replikation eines Teils der Genotypisierungsergebnisse auf einer unabhängigen Plattform)
6. Entdeckte die Studie zuverlässig Assoziationen mit zuvor berichteten und replizierten Varianten (bekannten positiven Befunden)?
7. Wurden stringente Korrekturen angewandt, um für Tausende durchgeführter Tests zu korrigieren?
8. Wurden die Ergebnisse in unabhängigen Populationen repliziert?
9. Waren die Replikationsstichproben bezüglich des geographischen Ursprungs und der Definition des Phänotyps vergleichbar? (und falls nicht: erstrecken sich diese Differenzen auf die Anwendbarkeit der Befunde?)
10. Wurde Evidenz für eine funktionelle Rolle des identifizierten Polymorphismus geliefert?

5 Zusammenfassung

Im Rahmen dieser Arbeit wurden die wesentlichen Methoden zur Identifizierung von krankheitsrelevanten Genen vorgestellt, die GWA, das „positional cloning“ und der Kandidatengenansatz. Ergebnisse aus Studien der eigenen Arbeitsgruppe, an deren Auswertungen ich maßgeblich beteiligt war, wurden zu Genen berichtet, die aus je einem Ansatz hervorgingen. Die Arbeiten führten zu der Entdeckung ganz neuer Loci, die eine Rolle bei der Entwicklung von Asthma im Kindesalter haben können.

Der jüngste Ansatz, die GWA, ist erst vor wenigen Jahren durch den Einsatz neuer Technologien möglich geworden. Mit diesem Ansatz wurde die erste GWA zu Asthma durchgeführt. Dabei wurde ein Locus auf Chromosom 17q21 entdeckt, der hoch mit Asthma assoziiert war. Auf diesem Locus lagen Gene, die zuvor nicht mit Asthma in Verbindung gebracht wurden. Von diesen scheint vor allem *ORMDL3* eine bedeutende Rolle zu spielen. Dies ist durch Befunde in Genexpressionsstudien belegt, bei denen die *ORMDL3* Expression durch SNPs auf Chromosom 17 reguliert wurde. Neben *ORMDL3* liegen andere Gene wie *GSDML* oder *ZNFNIA3* auf diesem Locus, für die eine relevante Funktion bei der Entwicklung auf Asthma ebenfalls in Diskussion ist. Das modernere, statistisch geleitete Suchen der GWA führte zur Aufdeckung neuer Gene, die bei der Entwicklung von Asthma eine Rolle spielen können und bedeutet somit einen großen Fortschritt. Dennoch bleiben viele Fragen, die vor allem in der Ergründung der Funktion neuer krankheitsrelevanter Loci und methodischer Einwände bestehen. Ob GWAs die kausalen Loci aufdecken können, ist derzeit nicht klar. Dies könnte u.a. daran scheitern, dass nicht alle SNPs des Genoms derzeit in einem Chip vereint werden können. Zudem scheinen bestimmte Konstellationen der Allelfrequenz das Auffinden von Assoziationen zu begünstigen. Dies könnte zu einer Benachteiligung seltener SNPs führen.

Das „positional cloning“ ging der GWA als genomweitem Ansatz voraus. Durch diese Methode wurde das Gen *ADAM33* entdeckt. Eine Assoziationsstudie der eigenen Arbeitsgruppe zu diesem Gen lieferte Ergebnisse, deren Interpretation sich schwierig gestaltet. Die dort aufgedeckten Effekte waren schwach und zum Teil inkonsistent. Dies steht aber in Übereinstimmung mit der Gesamtheit von Studien zu *ADAM33*, in denen kein *ADAM33* SNP eine konsistente Assoziation mit Asthma oder einem anderen Phänotyp aufwies. An *ADAM33* zeigte sich zudem, dass auch die Definition von Replikation Auswirkungen auf die Wahrnehmung der Bedeutung eines Gens haben kann. Der Zusammenhang von *ADAM33* mit Asthma war

über viele Studien konsistent, wenn statt eines SNPs das Gen als die zu replizierende Einheit herangezogen wurde. Zwar gibt es Hinweise, dass das Gen in Zusammenhang mit dem „airway remodelling“ steht, jedoch ist auch der Pathomechanismus, welcher Effekte von *ADAM33* auf Asthma erklären soll, nicht endgültig ergründet.

Den genomweiten Ansätzen ist es gemein, dass sie Pathomechanismen erst im Anschluss an eine Assoziationsstudie zu finden versuchen, insbesondere nach der Entdeckung neuer Gene. Solche Pathomechanismen können auf verschiedene Krankheitsbilder einwirken. So können sie auch auf verschiedene atopische Erkrankungen wirken oder aber auch nur eine Facette der Asthmadefinition betreffen. Auch generelle Einflüsse auf das Immunsystem könnten sich symptomatisch als Asthma äußern. In diesem Kontext wurde als weitere Hypothese der Einfluss von Zinkfingern vorgestellt. Daneben ist anzunehmen, dass auch hinter den Beobachtungen zu *ORMDL3* Varianten ein (gemeinsamer) Mechanismus versteckt ist. Dafür sprechen Assoziationen in drei verschiedenen *ORMDL*-Genen.

Als dritte Studie wurde eine reine Kandidatengenstudie beschrieben. Bei dieser war die Rolle von Toll-like Rezeptoren auf die Entwicklung von Asthma Gegenstand, da mikrobielle Exposition die Modulation der T-Effektor-Zellen zu Gunsten einer T_H1 -Antwort durch TLRs vermittelt. Stimulation von TLRs könnte daher vor Asthma und atopischen Erkrankungen schützen. Tatsächlich wurden protektive Effekte von Polymorphismen in *TLR1*, *TLR6* und *TLR10* festgestellt, die auch nach Korrektur für multiples Testen Bestand hatten. Es konnte zudem gezeigt werden, dass die Effekte für atopisches Asthma höher ausfielen als für Atopie ohne das gleichzeitige Auftreten von Asthma.

Genomweite Studien und Kandidatengenstudien führten zu der Identifizierung unterschiedlicher Loci. Derzeit ist nicht geklärt, welcher der beste Ansatz bei der Suche nach Asthmagenen ist oder ob es überhaupt einen solchen gibt. Die Methoden sollten daher nicht gegeneinander verglichen werden, sondern besser zu einem multimethodalen Ansatz verarbeitet werden. Um die ganze Komplexität von Asthma zu erfassen, sollten nicht nur Kombinationen von Methoden bei der Identifizierung von Genen, sondern auch die Analyse von Umweltfaktoren, Gen-Umweltfaktoren und die Verwendung verschiedener statistischer Methoden berücksichtigt werden.

Trotz einiger methodischer Einwände ist die GWA eine innovative Methode der Gegenwart. Voraussichtlich wird es in den nächsten Jahren gelingen, sämtliche Loci zu erfassen, so dass sich Probleme des „coverage“ erübrigen. Mit den großen Datenmengen gehen aber auch Probleme einher, welche die Frage nach der richtigen Auswertungsstrategie stellen. So wie sich die Komplexität der asthmatischen Erkrankung bereits in der Definition zeigt, bedarf es auch eines komplexen Vorgehens, um die verschiedenen Erscheinungsformen des Bildes klarer auszuleuchten. Dazu gehört das Verfolgen verschiedener Hypothesen ebenso wie das Verwenden neuer Methoden. Derzeit erscheint daher eine Kombination der verschiedenen Ansätze eine gute Strategie, die Suche nach den relevanten Asthmagenen fortzusetzen, da so die Vorteile verschiedener Vorgehensweisen zum Tragen kommen. Neue Loci können zu neuen Hypothesen führen, die im Rahmen klassischer, Theorie geleiteter Studien weiter geführt werden.

6 Literaturverzeichnis

- (1987). "Standardization of spirometry--1987 update. Official statement of American Thoracic Society." *Respir Care* **32**(11): 1039-60.
- (2003). European Lung White Book. *ERS*.
- (2003). "The International HapMap Project." *Nature* **426**(6968): 789-96.
- (2004). "Finishing the euchromatic sequence of the human genome." *Nature* **431**(7011): 931-45.
2. National Center for Biotechnology Information, N. and L. o. Medicine. (2007). "Database of Single Nucleotide Polymorphisms." Accessed February 14, 2008.
- Akey, J., L. Jin and M. Xiong (2001). "Haplotypes vs single marker linkage disequilibrium tests: what do we gain?" *Eur J Hum Genet* **9**(4): 291-300.
- Akira, S., K. Takeda and T. Kaisho (2001). "Toll-like receptors: critical proteins linking innate and acquired immunity." *Nat Immunol* **2**(8): 675-80.
- Alexopoulou, L., A. C. Holt, R. Medzhitov and R. A. Flavell (2001). "Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3." *Nature* **413**(6857): 732-8.
- Allen, M., A. Heinzmann, E. Noguchi, G. Abecasis, J. Broxholme, C. P. Ponting, et al. (2003). "Positional cloning of a novel gene influencing asthma from chromosome 2q14." *Nat Genet* **35**(3): 258-63.
- Anderson, C. A., F. H. Pettersson, J. C. Barrett, J. J. Zhuang, J. Ragoussis, L. R. Cardon, et al. (2008). "Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms." *Am J Hum Genet* **83**(1): 112-9.
- Andreaskos, E., B. Foxwell and M. Feldmann (2004). "Is targeting Toll-like receptors and their signaling pathway a useful therapeutic approach to modulating cytokine-driven inflammation?" *Immunol Rev* **202**: 250-65.
- Ardlie, K. G., L. Kruglyak and M. Seielstad (2002). "Patterns of linkage disequilibrium in the human genome." *Nat Rev Genet* **3**(4): 299-309.
- Armitage, P. (1955). "Tests for Linear Trend in proportions and Frequencies." *Biometrics* **11**: 375-386.
- Balaci, L., M. C. Spada, N. Olla, G. Sole, L. Loddo, F. Anedda, et al. (2007). "IRAK-M is involved in the pathogenesis of early-onset persistent asthma." *Am J Hum Genet* **80**(6): 1103-14.
- Balding, D. J. (2006). "A tutorial on statistical methods for population association studies." *Nat Rev Genet* **7**(10): 781-91.
- Barbato, A., G. Turato, S. Baraldo, E. Bazzan, F. Calabrese, C. Panizzolo, et al. (2006). "Epithelial damage and angiogenesis in the airways of children with asthma." *Am J Respir Crit Care Med* **174**(9): 975-81.
- Barnette, T., P. A. Gourraud and A. Cambon-Thomsen (2005). "Strategies in analysis of the genetic component of multifactorial diseases; biostatistical aspects." *Transpl Immunol* **14**(3-4): 255-66.
- Barrett, J. C. and L. R. Cardon (2006). "Evaluating coverage of genome-wide association studies." *Nat Genet* **38**(6): 659-62.
- Barrett, J. C., B. Fry, J. Maller and M. J. Daly (2005). "Haploview: analysis and visualization of LD and haplotype maps." *Bioinformatics* **21**(2): 263-5.
- Benjamini, Y., D. Drai, G. Elmer, N. Kafkafi and I. Golani (2001). "Controlling the false discovery rate in behavior genetics research." *Behav Brain Res* **125**(1-2): 279-84.
- Benjamini, Y. and Y. Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *J R Stat Soc Ser B* **57**: 289-300.

- Bentley, A. M., S. R. Durham and A. B. Kay (1994). "Comparison of the immunopathology of extrinsic, intrinsic and occupational asthma." J Investig Allergol Clin Immunol **4**(5): 222-32.
- Bergmann, R. L., K. E. Bergmann, S. Lau-Schadensdorf, W. Luck, A. Dannemann, C. P. Bauer, et al. (1994). "Atopic diseases in infancy. The German multicenter atopy study (MAS-90)." Pediatr Allergy Immunol **5**(6 Suppl): 19-25.
- Bieli, C., W. Eder, R. Frei, C. Braun-Fahrlander, W. Klimecki, M. Waser, et al. (2007). "A polymorphism in CD14 modifies the effect of farm milk consumption on allergic diseases and CD14 gene expression." J Allergy Clin Immunol **120**(6): 1308-15.
- Bienz, M. (2006). "The PHD finger, a nuclear protein-interaction domain." Trends Biochem Sci **31**(1): 35-40.
- Bisgaard, H., K. Bonnelykke, P. M. Sleiman, M. Brasholt, B. Chawes, E. Kreiner-Moller, et al. (2008). "ORMDL3 Associated Gene Variants are Associated with Asthma and Exacerbations but not Atopy in Early Childhood." Am J Respir Crit Care Med.
- Blakey, J., E. Halapi, U. S. Bjornsdottir, A. Wheatley, S. Kristinsson, R. Upmanyu, et al. (2005). "Contribution of ADAM33 polymorphisms to the population risk of asthma." Thorax **60**(4): 274-6.
- Blumenthal, M. N. (2005). "The role of genetics in the development of asthma and atopy." Curr Opin Allergy Clin Immunol **5**(2): 141-5.
- Boddeker, I. and A. Ziegler (2000). "[Association and linkage studies for the analysis of candidate genes]." Dtsch Med Wochenschr **125**(25-26): 810-5.
- Bosse, Y. and T. J. Hudson (2007). "Toward a comprehensive set of asthma susceptibility genes." Annu Rev Med **58**: 171-84.
- Bousquet, J., P. K. Jeffery, W. W. Busse, M. Johnson and A. M. Vignola (2000). "Asthma. From bronchoconstriction to airways inflammation and remodeling." Am J Respir Crit Care Med **161**(5): 1720-45.
- Bouzigon, E., E. Corda, H. Aschard, M. H. Dizier, A. Boland, J. Bousquet, et al. (2008). "Effect of 17q21 variants and smoking exposure in early-onset asthma." N Engl J Med **359**(19): 1985-94.
- Braun-Fahrlander, C., J. Riedler, U. Herz, W. Eder, M. Waser, L. Grize, et al. (2002). "Environmental exposure to endotoxin and its relation to asthma in school-age children." N Engl J Med **347**(12): 869-77.
- Burton, P. R., M. D. Tobin and J. L. Hopper (2005). "Key concepts in genetic epidemiology." Lancet **366**(9489): 941-51.
- Busselmaier, W. and G. Tarivedian (2007). Humangenetik. Heidelberg.
- Cakebread, J. A., H. M. Haitchi, J. W. Holloway, R. M. Powell, T. Keith, D. E. Davies, et al. (2004). "The role of ADAM33 in the pathogenesis of asthma." Springer Semin Immunopathol **25**(3-4): 361-75.
- Cardon, L. R. and G. R. Abecasis (2003). "Using haplotype blocks to map human complex trait loci." Trends Genet **19**(3): 135-40.
- Carl-McGrath, S., R. Schneider-Stock, M. Ebert and C. Rocken (2008). "Differential expression and localisation of gasdermin-like (GSDML), a novel member of the cancer-associated GSDMDC protein family, in neoplastic and non-neoplastic gastric, hepatic, and colon tissues." Pathology **40**(1): 13-24.
- Carlson, C. S., M. A. Eberle, M. J. Rieder, J. D. Smith, L. Kruglyak and D. A. Nickerson (2003). "Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans." Nat Genet **33**(4): 518-21.
- Carlson, C. S., M. A. Eberle, M. J. Rieder, Q. Yi, L. Kruglyak and D. A. Nickerson (2004). "Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium." Am J Hum Genet **74**(1): 106-20.
- Carroll, W. (2005). "Asthma genetics: pitfalls and triumphs." Paediatr Respir Rev **6**(1): 68-74.

- Chanock, S. J., T. Manolio, M. Boehnke, E. Boerwinkle, D. J. Hunter, G. Thomas, et al. (2007). "Replicating genotype-phenotype associations." Nature **447**(7145): 655-60.
- Cheung, V. G. and R. S. Spielman (2002). "The genetics of variation in gene expression." Nat Genet **32 Suppl**: 522-5.
- Chudley, A. E. (1998). "Genetic landmarks through philately--Gregor Johann Mendel (1822-1884)." Clin Genet **54**(2): 121-3.
- Chupp, G. L., C. G. Lee, N. Jarjour, Y. M. Shim, C. T. Holm, S. He, et al. (2007). "A chitinase-like protein in the lung and circulation of patients with severe asthma." N Engl J Med **357**(20): 2016-27.
- Clark, A. G. (1990). "Inference of haplotypes from PCR-amplified samples of diploid populations." Mol Biol Evol **7**(2): 111-22.
- Clark, A. G. (2004). "The role of haplotypes in candidate gene studies." Genet Epidemiol **27**(4): 321-33.
- Clarke, E., N. Rahman, N. Page, M. S. Rolph, G. J. Stewart and G. J. Jones (2008). "Functional characterization of the atopy-associated gene PHF11." J Allergy Clin Immunol **121**(5): 1148-1154 e3.
- Cohen, J. C., R. S. Kiss, A. Pertsemlidis, Y. L. Marcel, R. McPherson and H. H. Hobbs (2004). "Multiple rare alleles contribute to low plasma levels of HDL cholesterol." Science **305**(5685): 869-72.
- Cohen, L., X. E, J. Tarsi, T. Ramkumar, T. K. Horiuchi, R. Cochran, et al. (2007). "Epithelial cell proliferation contributes to airway remodeling in severe asthma." Am J Respir Crit Care Med **176**(2): 138-45.
- Consortium, T. I. H. (2005). "A haplotype map of the human genome." Nature **437**(7063): 1299-320.
- Consortium, W. T. C. C. (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**(7145): 661-78.
- Cordell, H. J. (2002). "Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans." Hum Mol Genet **11**(20): 2463-8.
- Cordell, H. J. and D. G. Clayton (2005). "Genetic association studies." Lancet **366**(9491): 1121-31.
- Cousin, E., E. Genin, S. Mace, S. Ricard, C. Chansac, M. del Zompo, et al. (2003). "Association studies in candidate genes: strategies to select SNPs to be tested." Hum Hered **56**(4): 151-9.
- Crawford, D. C. and D. A. Nickerson (2005). "Definition and clinical importance of haplotypes." Annu Rev Med **56**: 303-20.
- Cromie, G. A. and G. R. Smith (2007). "Branching out: meiotic recombination and its regulation." Trends Cell Biol **17**(9): 448-55.
- Daley, D., M. Lemire, L. Akhabir, M. Chan-Yeung, J. Q. He, T. McDonald, et al. (2009). "Analyses of associations with asthma in four asthma population samples from Canada and Australia." Hum Genet.
- Daly, M. J., J. D. Rioux, S. F. Schaffner, T. J. Hudson and E. S. Lander (2001). "High-resolution haplotype structure in the human genome." Nat Genet **29**(2): 229-32.
- Dempfle, A., A. Scherag, R. Hein, L. Beckmann, J. Chang-Claude and H. Schafer (2008). "Gene-environment interactions for complex traits: definitions, methodological requirements and challenges." Eur J Hum Genet **16**(10): 1164-72.
- Depner, M., M. S. Kormann, N. Klopp, T. Illig, C. Vogelberg, S. K. Weiland, et al. (2007). "CX3CR1 polymorphisms are associated with atopy but not asthma in German children." Int Arch Allergy Immunol **144**(1): 91-4.
- Devereux, G., S. W. Turner, L. C. Craig, G. McNeill, S. Martindale, P. J. Harbour, et al. (2006). "Low maternal vitamin E intake during pregnancy is associated with asthma in 5-year-old children." Am J Respir Crit Care Med **174**(5): 499-507.

- Devirgiliis, C., P. D. Zalewski, G. Perozzi and C. Murgia (2007). "Zinc fluxes and zinc transporter genes in chronic diseases." Mutat Res **622**(1-2): 84-93.
- Devlin, B., S. A. Bacanu and K. Roeder (2004). "Genomic Control to the extreme." Nat Genet **36**(11): 1129-30; author reply 1131.
- Devlin, B. and N. Risch (1995). "A comparison of linkage disequilibrium measures for fine-scale mapping." Genomics **29**(2): 311-22.
- Devlin, B. and K. Roeder (1999). "Genomic control for association studies." Biometrics **55**(4): 997-1004.
- Douwes, J., N. Pearce and D. Heederik (2002). "Does environmental endotoxin exposure prevent asthma?" Thorax **57**(1): 86-90.
- Du, F. X., A. C. Clutter and M. M. Lohuis (2007). "Characterizing linkage disequilibrium in pig populations." Int J Biol Sci **3**(3): 166-78.
- Duan, S., R. S. Huang, W. Zhang, W. K. Bleibel, C. A. Roe, T. A. Clark, et al. (2008). "Genetic architecture of transcript-level variation in humans." Am J Hum Genet **82**(5): 1101-13.
- Dudbridge, F. (2003). "Pedigree disequilibrium tests for multilocus haplotypes." Genet Epidemiol **25**(2): 115-21.
- Eder, W., M. J. Ege and E. von Mutius (2006). "The asthma epidemic." N Engl J Med **355**(21): 2226-35.
- Eder, W., W. Klimecki, L. Yu, E. von Mutius, J. Riedler, C. Braun-Fahrlander, et al. (2004). "Toll-like receptor 2 as a major gene for asthma in children of European farmers." J Allergy Clin Immunol **113**(3): 482-8.
- Edfors-Lubs, M. L. (1971). "Allergy in 7000 twin pairs." Acta Allergol **26**(4): 249-85.
- Editorial (1999). "Freely associating." Nat Genet **22**(1): 1-2.
- Ennis, S., N. Maniatis and A. Collins (2001). "Allelic association and disease mapping." Brief Bioinform **2**(4): 375-87.
- Excoffier, L., G. Laval and D. Balding (2003). "Gametic phase estimation over large genomic regions using an adaptive window approach." Hum Genomics **1**(1): 7-19.
- Excoffier, L. and M. Slatkin (1995). "Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population." Mol Biol Evol **12**(5): 921-7.
- Feijen, M., J. Gerritsen and D. S. Postma (2000). "Genetics of allergic disease." Br Med Bull **56**(4): 894-907.
- Frayling, T. M. and M. I. McCarthy (2007). "Genetic studies of diabetes following the advent of the genome-wide association study: where do we go from here?" Diabetologia **50**(11): 2229-33.
- Frazer, K. A., D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve, R. A. Gibbs, et al. (2007). "A second generation human haplotype map of over 3.1 million SNPs." Nature **449**(7164): 851-61.
- Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, et al. (2002). "The structure of haplotype blocks in the human genome." Science **296**(5576): 2225-9.
- Galanter, J., S. Choudhry, C. Eng, S. Nazario, J. R. Rodriguez-Santana, J. Casal, et al. (2008). "ORMDL3 gene is associated with asthma in three ethnically diverse populations." Am J Respir Crit Care Med **177**(11): 1194-200.
- Garlisi, C. G., J. Zou, K. E. Devito, F. Tian, F. X. Zhu, J. Liu, et al. (2003). "Human ADAM33: protein maturation and localization." Biochem Biophys Res Commun **301**(1): 35-43.
- Griffin, T. J. and L. M. Smith (2000). "Genetic identification by mass spectrometric analysis of single-nucleotide polymorphisms: ternary encoding of genotypes." Anal Chem **72**(14): 3298-302.
- Guerra, S. and F. D. Martinez (2008). "Asthma genetics: from linear to multifactorial approaches." Annu Rev Med **59**: 327-41.

- Guillot, L., S. Medjane, K. Le-Barillec, V. Balloy, C. Danel, M. Chignard, et al. (2004). "Response of human pulmonary epithelial cells to lipopolysaccharide involves Toll-like receptor 4 (TLR4)-dependent signaling pathways: evidence for an intracellular compartmentalization of TLR4." J Biol Chem **279**(4): 2712-8.
- Hafler, D. A., A. Compston, S. Sawcer, E. S. Lander, M. J. Daly, P. L. De Jager, et al. (2007). "Risk alleles for multiple sclerosis identified by a genomewide study." N Engl J Med **357**(9): 851-62.
- Halperin, E. and D. A. Stephan (2009). "SNP imputation in association studies." Nat Biotechnol **27**(4): 349-51.
- Hartl, D. L. C., Andrew. G. Clark (1997). Principles of population genetics. Sunderland, Massachusetts, Sinauer associates, Inc.
- Hartung, D. M. and D. Touchette (2009). "Overview of clinical research design." Am J Health Syst Pharm **66**(4): 398-408.
- Hastbacka, J., A. de la Chapelle, I. Kaitila, P. Sistonen, A. Weaver and E. Lander (1992). "Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland." Nat Genet **2**(3): 204-11.
- Hawley, M. E. and K. K. Kidd (1995). "HAPLO: a program using the EM algorithm to estimate the frequencies of multi-site haplotypes." J Hered **86**(5): 409-11.
- Hayashi, F., K. D. Smith, A. Ozinsky, T. R. Hawn, E. C. Yi, D. R. Goodlett, et al. (2001). "The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5." Nature **410**(6832): 1099-103.
- Heil, F., H. Hemmi, H. Hochrein, F. Ampenberger, C. Kirschning, S. Akira, et al. (2004). "Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8." Science **303**(5663): 1526-9.
- Hemmi, H., O. Takeuchi, T. Kawai, T. Kaisho, S. Sato, H. Sanjo, et al. (2000). "A Toll-like receptor recognizes bacterial DNA." Nature **408**(6813): 740-5.
- Hersh, C. P., B. A. Raby, M. E. Soto-Quiros, A. J. Murphy, L. Avila, J. Lasky-Su, et al. (2007). "Comprehensive testing of positionally cloned asthma genes in two populations." Am J Respir Crit Care Med **176**(9): 849-57.
- Hirota, T., M. Harada, M. Sakashita, S. Doi, A. Miyatake, K. Fujita, et al. (2008). "Genetic polymorphism regulating ORM1-like 3 (*Saccharomyces cerevisiae*) expression is associated with childhood atopic asthma in a Japanese population." J Allergy Clin Immunol **121**(3): 769-70.
- Hirschhorn, J. N. and M. J. Daly (2005). "Genome-wide association studies for common diseases and complex traits." Nat Rev Genet **6**(2): 95-108.
- Hjelmqvist, L., M. Tuson, G. Marfany, E. Herrero, S. Balcells and R. Gonzalez-Duarte (2002). "ORMDL proteins are a conserved new family of endoplasmic reticulum membrane proteins." Genome Biol **3**(6): RESEARCH0027.
- Hochberg, Y. and Y. Benjamini (1990). "More powerful procedures for multiple significance testing." Stat Med **9**(7): 811-8.
- Hoffjan, S., S. Stemmler, Q. Parwez, E. Petrasch-Parwez, U. Arinir, G. Rohde, et al. (2005). "Evaluation of the toll-like receptor 6 Ser249Pro polymorphism in patients with asthma, atopic dermatitis and chronic obstructive pulmonary disease." BMC Med Genet **6**: 34.
- Hoggart, C. J., T. G. Clark, M. De Iorio, J. C. Whittaker and D. J. Balding (2008). "Genome-wide significance for dense SNP and resequencing data." Genet Epidemiol **32**(2): 179-85.
- Holgate, S. T., Y. Yang, H. M. Haitchi, R. M. Powell, J. W. Holloway, H. Yoshisue, et al. (2006). "The genetics of asthma: ADAM33 as an example of a susceptibility gene." Proc Am Thorac Soc **3**(5): 440-3.

- Holloway, J. W. and G. H. Koppelman (2007). "Identifying novel genes contributing to asthma pathogenesis." Curr Opin Allergy Clin Immunol **7**(1): 69-74.
- Holzappel, C., J. Baumert, H. Grallert, A. M. Muller, B. Thorand, N. Khuseyinova, et al. (2008). "Genetic variants in the USF1 gene are associated with low-density lipoprotein cholesterol levels and incident type 2 diabetes mellitus in women: results from the MONICA/KORA Augsburg case-cohort study, 1984-2002." Eur J Endocrinol **159**(4): 407-16.
- Howard, T. D., D. S. Postma, H. Jongepier, W. C. Moore, G. H. Koppelman, S. L. Zheng, et al. (2003). "Association of a disintegrin and metalloprotease 33 (ADAM33) gene with asthma in ethnically diverse populations." J Allergy Clin Immunol **112**(4): 717-22.
- Hughes, A. L. and H. Piontkivska (2008). "Functional diversification of the toll-like receptor gene family." Immunogenetics **60**(5): 249-56.
- Humbert, M., S. R. Durham, S. Ying, P. Kimmitt, J. Barkans, B. Assoufi, et al. (1996). "IL-4 and IL-5 mRNA and protein in bronchial biopsies from patients with atopic and non-atopic asthma: evidence against "intrinsic" asthma being a distinct immunopathologic entity." Am J Respir Crit Care Med **154**(5): 1497-504.
- Hunter, D. J., P. Kraft, K. B. Jacobs, D. G. Cox, M. Yeager, S. E. Hankinson, et al. (2007). "A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer." Nat Genet **39**(7): 870-4.
- Iles, M. M. (2008a). "What can genome-wide association studies tell us about the genetics of common disease?" PLoS Genet **4**(2): e33.
- Iles, M. M. (2008b). "Quantification and correction of bias in tagging SNPs caused by insufficient sample size and marker density by means of haplotype-dropping." Genet Epidemiol **32**(1): 20-8.
- Illi, S., E. von Mutius, S. Lau, R. Nickel, C. Gruber, B. Niggemann, et al. (2004). "The natural course of atopic dermatitis from birth to age 7 years and the association with asthma." J Allergy Clin Immunol **113**(5): 925-31.
- Illi, S., E. von Mutius, S. Lau, B. Niggemann, C. Gruber and U. Wahn (2006). "Perennial allergen sensitisation early in life and chronic asthma in children: a birth cohort study." Lancet **368**(9537): 763-70.
- Janeway, C. A., Jr. and R. Medzhitov (2002). "Innate immune recognition." Annu Rev Immunol **20**: 197-216.
- Jang, N., G. Stewart and G. Jones (2005). "Polymorphisms within the PHF11 gene at chromosome 13q14 are associated with childhood atopic dermatitis." Genes Immun **6**(3): 262-4.
- Jones, M. (2008). "Understanding of the molecular mechanisms of allergy." Methods Mol Med **138**: 1-15.
- Jongepier, H., H. M. Boezen, A. Dijkstra, T. D. Howard, J. M. Vonk, G. H. Koppelman, et al. (2004). "Polymorphisms of the ADAM33 gene are associated with accelerated lung function decline in asthma." Clin Exp Allergy **34**(5): 757-60.
- Jorde, L. B. (1995). "Linkage disequilibrium as a gene-mapping tool." Am J Hum Genet **56**(1): 11-4.
- Jorde, L. B. (2000). "Linkage disequilibrium and the search for complex disease genes." Genome Res **10**(10): 1435-44.
- Jorgenson, E. and J. S. Witte (2006). "Coverage and power in genomewide association studies." Am J Hum Genet **78**(5): 884-8.
- Judson, R., J. C. Stephens and A. Windemuth (2000). "The predictive power of haplotypes in clinical response." Pharmacogenomics **1**(1): 15-26.
- Kabesch, M. (2004). "Candidate genes and the genetic epidemiology of asthma." Paediatr Respir Rev **5 Suppl A**: S23-5.

- Kabesch, M., M. Depner, I. Dahmen, S. K. Weiland, C. Vogelberg, B. Niggemann, et al. (2007). "Polymorphisms in eosinophil pathway genes, asthma and atopy." *Allergy* **62**(4): 423-8.
- Karas, M. and F. Hillenkamp (1988). "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons." *Anal Chem* **60**(20): 2299-301.
- Kay, A. B. (2001). "Allergy and allergic diseases. First of two parts." *N Engl J Med* **344**(1): 30-7.
- Kere, J. and T. Laitinen (2004). "Positionally cloned susceptibility genes in allergy and asthma." *Curr Opin Immunol* **16**(6): 689-94.
- Koppelman, G. H. (2006). "Gene by environment interaction in asthma." *Curr Allergy Asthma Rep* **6**(2): 103-11.
- Kormann, M. S. (2007). Toll-like receptor heterodimer variants influence Th1/Th2 signaling and protect from the development of childhood asthma. *Institut für Experimentelle Genetik, GSF-Forschungszentrum für Umwelt und Gesundheit, Neuherberg*. Munich, Ludwig-Maximilians University.
- Kormann, M. S., M. Depner, D. Hartl, N. Klopp, T. Illig, J. Adamski, et al. (2008). "Toll-like receptor heterodimer variants protect from childhood asthma." *J Allergy Clin Immunol* **122**(1): 86-92, 92 e1-8.
- Kormann, M. S., R. Ferstl, M. Depner, N. Klopp, S. Spiller, T. Illig, et al. (2009). "Rare TLR2 mutations reduce TLR2 receptor function and can increase atopy risk." *Allergy*.
- Krishnan, J., K. Selvarajoo, M. Tsuchiya, G. Lee and S. Choi (2007). "Toll-like receptor signal transduction." *Exp Mol Med* **39**(4): 421-38.
- Kruglyak, L. (1999). "Prospects for whole-genome linkage disequilibrium mapping of common disease genes." *Nat Genet* **22**(2): 139-44.
- Kruglyak, L. and D. A. Nickerson (2001). "Variation is the spice of life." *Nat Genet* **27**(3): 234-6.
- Kulig, M., R. Bergmann, U. Klettke, V. Wahn, U. Tacke and U. Wahn (1999). "Natural course of sensitization to food and inhalant allergens during the first 6 years of life." *J Allergy Clin Immunol* **103**(6): 1173-9.
- Laitinen, T. (2007). "Gene mapping in asthma-related traits." *Methods Mol Biol* **376**: 213-34.
- Laitinen, T., A. Polvi, P. Rydman, J. Vendelin, V. Pulkkinen, P. Salmikangas, et al. (2004). "Characterization of a common susceptibility locus for asthma-related traits." *Science* **304**(5668): 300-4.
- Lazarus, R., W. T. Klimecki, B. A. Raby, D. Vercelli, L. J. Palmer, D. J. Kwiatkowski, et al. (2003). "Single-nucleotide polymorphisms in the Toll-like receptor 9 gene (TLR9): frequencies, pairwise linkage disequilibrium, and haplotypes in three U.S. ethnic groups and exploratory case-control disease association studies." *Genomics* **81**(1): 85-91.
- Lee, J. Y., S. W. Park, H. K. Chang, H. Y. Kim, T. Rhim, J. H. Lee, et al. (2006). "A disintegrin and metalloproteinase 33 protein in patients with asthma: Relevance to airflow limitation." *Am J Respir Crit Care Med* **173**(7): 729-35.
- Lemanske, R. F., Jr. and W. W. Busse (2003). "6. Asthma." *J Allergy Clin Immunol* **111**(2 Suppl): S502-19.
- Li, M. and C. Li (2008). "Assessing departure from Hardy-Weinberg equilibrium in the presence of disease association." *Genet Epidemiol* **32**(7): 589-99.
- Li, M., C. Li and W. Guan (2008). "Evaluation of coverage variation of SNP chips for genome-wide association studies." *Eur J Hum Genet* **16**(5): 635-43.
- Li, Y., W. Liao, M. Chang, S. J. Schrodi, N. Bui, J. J. Catanese, et al. (2009). "Further genetic evidence for three psoriasis-risk genes: ADAM33, CDKAL1, and PTPN22." *J Invest Dermatol* **129**(3): 629-34.

- Lin, D. Y. (2006). "Evaluating statistical significance in two-stage genomewide association studies." *Am J Hum Genet* **78**(3): 505-9.
- Lind, D. L., S. Choudhry, N. Ung, E. Ziv, P. C. Avila, K. Salari, et al. (2003). "ADAM33 is not associated with asthma in Puerto Rican or Mexican populations." *Am J Respir Crit Care Med* **168**(11): 1312-6.
- Liu, X., T. H. Beaty, P. Deindl, S. K. Huang, S. Lau, C. Sommerfeld, et al. (2003). "Associations between total serum IgE levels and the 6 potentially functional variants within the genes IL4, IL13, and IL4RA in German children: the German Multicenter Atopy Study." *J Allergy Clin Immunol* **112**(2): 382-8.
- Lou, X. Y., J. Z. Ma, M. C. Yang, J. Zhu, P. Y. Liu, H. W. Deng, et al. (2006). "Improvement of mapping accuracy by unifying linkage and association analysis." *Genetics* **172**(1): 647-61.
- Madore, A. M., K. Tremblay, T. J. Hudson and C. Laprise (2008). "Replication of an association between 17q21 SNPs and asthma in a French-Canadian familial collection." *Hum Genet* **123**(1): 93-5.
- Marchini, J., B. Howie, S. Myers, G. McVean and P. Donnelly (2007). "A new multipoint method for genome-wide association studies by imputation of genotypes." *Nat Genet* **39**(7): 906-13.
- Martinez, F. D. (2008). "Gene-environment interaction in complex diseases: asthma as an illustrative case." *Novartis Found Symp* **293**: 184-92; discussion 192-7.
- Mathew, C. G. (2008). "New links to the pathogenesis of Crohn disease provided by genome-wide association scans." *Nat Rev Genet* **9**(1): 9-14.
- Matsusue, A., C. Kiyohara, K. Tanaka, S. Sasaki and Y. Miyake (2008). "ADAM33 genetic polymorphisms and risk of atopic dermatitis among Japanese children." *Clin Biochem.*
- McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. Ioannidis, et al. (2008). "Genome-wide association studies for complex traits: consensus, uncertainty and challenges." *Nat Rev Genet* **9**(5): 356-69.
- Medzhitov, R. and C. A. Janeway, Jr. (1998). "Innate immune recognition and control of adaptive immune responses." *Semin Immunol* **10**(5): 351-3.
- Millstein, J., D. V. Conti, F. D. Gilliland and W. J. Gauderman (2006). "A testing framework for identifying susceptibility genes in the presence of epistasis." *Am J Hum Genet* **78**(1): 15-27.
- Minoda, Y., K. Saeki, D. Aki, H. Takaki, T. Sanada, K. Koga, et al. (2006). "A novel Zinc finger protein, ZCCHC11, interacts with TIFA and modulates TLR signaling." *Biochem Biophys Res Commun* **344**(3): 1023-30.
- Moffatt, M. F. (2008). "Genes in asthma: new genes and new ways." *Curr Opin Allergy Clin Immunol* **8**(5): 411-7.
- Moffatt, M. F., M. Kabesch, L. Liang, A. L. Dixon, D. Strachan, S. Heath, et al. (2007). "Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma." *Nature* **448**(7152): 470-3.
- Moller-Larsen, S., M. Nyegaard, A. Haagerup, J. Vestbo, T. A. Kruse and A. D. Borlum (2008). "Association analysis identifies TLR7 and TLR8 as novel risk genes in asthma and related disorders." *Thorax* **63**(12): 1064-9.
- Morley, M., C. M. Molony, T. M. Weber, J. L. Devlin, K. G. Ewens, R. S. Spielman, et al. (2004). "Genetic analysis of genome-wide variation in human gene expression." *Nature* **430**(7001): 743-7.
- Moroy, T., H. Zeng, J. Jin, K. W. Schmid, A. Carpintero and E. Gulbins (2008). "The zinc finger protein and transcriptional repressor Gfi1 as a regulator of the innate immune response." *Immunobiology* **213**(3-4): 341-52.
- Mosmann, T. R. and S. Sad (1996). "The expanding universe of T-cell subsets: Th1, Th2 and more." *Immunol Today* **17**(3): 138-46.

- Murcay, C. E., J. P. Lewinger and W. J. Gauderman (2009). "Gene-environment interaction in genome-wide association studies." *Am J Epidemiol* **169**(2): 219-26.
- Murphy, A., S. T. Weiss and C. Lange (2008). "Screening and replication using the same data set: testing strategies for family-based studies in which all probands are affected." *PLoS Genet* **4**(9): e1000197.
- Musani, S. K., D. Shriner, N. Liu, R. Feng, C. S. Coffey, N. Yi, et al. (2007). "Detection of gene x gene interactions in genome-wide association studies of human population data." *Hum Hered* **63**(2): 67-84.
- Nakao, Y., K. Funami, S. Kikkawa, M. Taniguchi, M. Nishiguchi, Y. Fukumori, et al. (2005). "Surface-expressed TLR6 participates in the recognition of diacylated lipopeptide and peptidoglycan in human cells." *J Immunol* **174**(3): 1566-73.
- National Heart, L., and Blood Institute Program in Genomic Applications. (2008). "Innate Immunity PGA (Program for Genomic Applications)." Retrieved Accessed May 30, 2008., from <http://innateimmunity.net>.
- Neale, B. M. and P. C. Sham (2004). "The future of association studies: gene-based analysis and replication." *Am J Hum Genet* **75**(3): 353-62.
- Nelson, M. R., S. L. Kardia, R. E. Ferrell and C. F. Sing (2001). "A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation." *Genome Res* **11**(3): 458-70.
- Ngoc, P. L., D. R. Gold, A. O. Tzianabos, S. T. Weiss and J. C. Celedon (2005). "Cytokines, allergy, and asthma." *Curr Opin Allergy Clin Immunol* **5**(2): 161-6.
- Nicolae, D., N. J. Cox, L. A. Lester, D. Schneider, Z. Tan, C. Billstrand, et al. (2005). "Fine mapping and positional candidate studies identify HLA-G as an asthma susceptibility gene on chromosome 6p21." *Am J Hum Genet* **76**(2): 349-57.
- Nicolai, T., E. V. Mutius, P. Reitmeir and M. Wjst (1993). "Reactivity to cold-air hyperventilation in normal and in asthmatic children in a survey of 5,697 schoolchildren in southern Bavaria." *Am Rev Respir Dis* **147**(3): 565-72.
- Niggemann, B., S. Illi, C. Madloch, K. Volkel, S. Lau, R. Bergmann, et al. (2001). "Histamine challenges discriminate between symptomatic and asymptomatic children. MAS-Study Group. Multicentre Allergy Study." *Eur Respir J* **17**(2): 246-53.
- Niu, T., Z. S. Qin, X. Xu and J. S. Liu (2002). "Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms." *Am J Hum Genet* **70**(1): 157-69.
- Noguchi, E., Y. Yokouchi, J. Zhang, K. Shibuya, A. Shibuya, M. Bannai, et al. (2005). "Positional identification of an asthma susceptibility gene on human chromosome 5q33." *Am J Respir Crit Care Med* **172**(2): 183-8.
- Ober, C. and S. Hoffjan (2006). "Asthma genetics 2006: the long and winding road to gene discovery." *Genes Immun* **7**(2): 95-100.
- Ober, C., Z. Tan, Y. Sun, J. D. Possick, L. Pan, R. Nicolae, et al. (2008). "Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function." *N Engl J Med* **358**(16): 1682-91.
- Ozinsky, A., D. M. Underhill, J. D. Fontenot, A. M. Hajjar, K. D. Smith, C. B. Wilson, et al. (2000). "The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors." *Proc Natl Acad Sci U S A* **97**(25): 13766-71.
- Palmer, L. J. and L. R. Cardon (2005). "Shaking the tree: mapping complex disease genes with linkage disequilibrium." *Lancet* **366**(9492): 1223-34.
- Paschou, P., E. Ziv, E. G. Burchard, S. Choudhry, W. Rodriguez-Cintron, M. W. Mahoney, et al. (2007). "PCA-correlated SNPs for structure identification in worldwide human populations." *PLoS Genet* **3**(9): 1672-86.

- Patil, N., A. J. Berno, D. A. Hinds, W. A. Barrett, J. M. Doshi, C. R. Hacker, et al. (2001). "Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21." *Science* **294**(5547): 1719-23.
- Pe'er, I., P. I. de Bakker, J. Maller, R. Yelensky, D. Altshuler and M. J. Daly (2006). "Evaluating and improving power in whole-genome association studies using fixed marker sets." *Nat Genet* **38**(6): 663-7.
- Pearson, T. A. and T. A. Manolio (2008). "How to interpret a genome-wide association study." *Jama* **299**(11): 1335-44.
- Peltonen, L. and V. A. McKusick (2001). "Genomics and medicine. Dissecting human disease in the postgenomic era." *Science* **291**(5507): 1224-9.
- Petes, T. D. (2001). "Meiotic recombination hot spots and cold spots." *Nat Rev Genet* **2**(5): 360-9.
- Phillips, M. S., R. Lawrence, R. Sachidanandam, A. P. Morris, D. J. Balding, M. A. Donaldson, et al. (2003). "Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots." *Nat Genet* **33**(3): 382-7.
- Pinto, L. A., M. Depner, L. Steudemann, N. Klopp, T. Illig, E. von Mutius, et al. (2009). "IL15 gene variants are not associated with asthma and atopy." *Allergy* **64**(4): 643-6.
- Pinto, L. A., L. Steudemann, M. Depner, N. Klopp, T. Illig, S. K. Weiland, et al. (2007). "STAT1 gene variations, IgE regulation and atopy." *Allergy* **62**(12): 1456-61.
- Platts-Mills, T. A. and L. M. Wheatley (1996). "The role of allergy and atopy in asthma." *Curr Opin Pulm Med* **2**(1): 29-34.
- Poltorak, A., X. He, I. Smirnova, M. Y. Liu, C. Van Huffel, X. Du, et al. (1998). "Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene." *Science* **282**(5396): 2085-8.
- Powell, R. M., J. Wicks, J. W. Holloway, S. T. Holgate and D. E. Davies (2004). "The splicing and fate of ADAM33 transcripts in primary human airways fibroblasts." *Am J Respir Cell Mol Biol* **31**(1): 13-21.
- Price, A. L., J. Butler, N. Patterson, C. Capelli, V. L. Pascali, F. Scarnicci, et al. (2008). "Discerning the ancestry of European Americans in genetic association studies." *PLoS Genet* **4**(1): e236.
- Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick and D. Reich (2006). "Principal components analysis corrects for stratification in genome-wide association studies." *Nat Genet* **38**(8): 904-9.
- Primakoff, P. and D. G. Myles (2000). "The ADAM gene family: surface proteins with adhesion and protease activity." *Trends Genet* **16**(2): 83-7.
- Pritchard, J. K. and M. Przeworski (2001). "Linkage disequilibrium in humans: models and data." *Am J Hum Genet* **69**(1): 1-14.
- Puthothu, B. and A. Heinzmann (2006). "Is toll-like receptor 6 or toll-like receptor 10 involved in asthma genetics--or both?" *Allergy* **61**(5): 649-50.
- Raby, B. A., E. K. Silverman, D. J. Kwiatkowski, C. Lange, R. Lazarus and S. T. Weiss (2004). "ADAM33 polymorphisms and phenotype associations in childhood asthma." *J Allergy Clin Immunol* **113**(6): 1071-8.
- Raelson, J. V., R. D. Little, A. Ruether, H. Fournier, B. Paquin, P. Van Eerdewegh, et al. (2007). "Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci." *Proc Natl Acad Sci U S A* **104**(37): 14747-52.
- Ricciardolo, F. L., A. Di Stefano, J. H. van Krieken, J. K. Sont, A. van Schadewijk, K. F. Rabe, et al. (2003). "Proliferation and inflammation in bronchial epithelium after allergen in atopic asthmatics." *Clin Exp Allergy* **33**(7): 905-11.
- Risch, N. and K. Merikangas (1996). "The future of genetic studies of complex human diseases." *Science* **273**(5281): 1516-7.

- Risch, N. J. (2000). "Searching for genetic determinants in the new millennium." Nature **405**(6788): 847-56.
- Robinson, D. S. (2000). "Th-2 cytokines in allergic disease." Br Med Bull **56**(4): 956-68.
- Romagnani, S. (1991). "Human TH1 and TH2 subsets: doubt no more." Immunol Today **12**(8): 256-7.
- Romagnani, S., E. Maggi, F. Liotta, L. Cosmi and F. Annunziato (2009). "Properties and origin of human Th17 cells." Mol Immunol.
- Romagne, F. (2007). "Current and future drugs targeting one class of innate immunity receptors: the Toll-like receptors." Drug Discov Today **12**(1-2): 80-7.
- Romanet-Manent, S., D. Charpin, A. Magnan, A. Lanteaume and D. Vervloet (2002). "Allergic vs nonallergic asthma: what makes the difference?" Allergy **57**(7): 607-13.
- Ron, M. and J. I. Weller (2007). "From QTL to QTN identification in livestock--winning by points rather than knock-out: a review." Anim Genet **38**(5): 429-39.
- Ryckman, K. and S. M. Williams (2008). "Calculation and use of the Hardy-Weinberg model in association studies." Curr Protoc Hum Genet **Chapter 1**: Unit 1 18.
- Sabatti, C., S. Service and N. Freimer (2003). "False discovery rate in linkage and association genome screens for complex disorders." Genetics **164**(2): 829-33.
- Salem, R. M., J. Wessel and N. J. Schork (2005). "A comprehensive literature review of haplotyping software and methods for use with unrelated individuals." Hum Genomics **2**(1): 39-66.
- Sasieni, P. D. (1997). "From genotypes to genes: doubling the sample size." Biometrics **53**(4): 1253-61.
- Satagopan, J. M., E. S. Venkatraman and C. B. Begg (2004). "Two-stage designs for gene-disease association studies with sample size constraints." Biometrics **60**(3): 589-97.
- Satagopan, J. M., D. A. Verbel, E. S. Venkatraman, K. E. Offit and C. B. Begg (2002). "Two-stage designs for gene-disease association studies." Biometrics **58**(1): 163-70.
- Schadt, E. E., S. A. Monks, T. A. Drake, A. J. Lusk, N. Che, V. Colinayo, et al. (2003). "Genetics of gene expression surveyed in maize, mouse and man." Nature **422**(6929): 297-302.
- Schaid, D. J. (1996). "General score tests for associations of genetic markers with disease using cases and their parents." Genet Epidemiol **13**(5): 423-49.
- Schaid, D. J. (2004). "Evaluating associations of haplotypes with traits." Genet Epidemiol **27**(4): 348-64.
- Schaid, D. J., C. M. Rowland, D. E. Tines, R. M. Jacobson and G. A. Poland (2002). "Score tests for association between traits and haplotypes when linkage phase is ambiguous." Am J Hum Genet **70**(2): 425-34.
- Schedel, M., M. Depner, C. Schoen, S. K. Weiland, C. Vogelberg, B. Niggemann, et al. (2006). "The role of polymorphisms in ADAM33, a disintegrin and metalloprotease 33, in childhood asthma and lung function in two German populations." Respir Res **7**: 91.
- Schedel, M., L. A. Pinto, B. Schaub, P. Rosenstiel, D. Cherkasov, L. Cameron, et al. (2008). "IRF-1 gene variations influence IgE regulation and atopy." Am J Respir Crit Care Med **177**(6): 613-21.
- Schmidt, D. and R. Durrett (2004). "Adaptive evolution drives the diversification of zinc-finger binding domains." Mol Biol Evol **21**(12): 2326-39.
- Schneeberger, G. and W. Mueller (2005). "Genetische Variation. Eine Einführung aus naturwissenschaftlicher Sicht." Journal of preventive medicine **2**: 52-57.
- Schoen, C. (2008). Die Rolle von ADAM 33-Polymorphismen bei Asthma bronchiale und Lungenfunktion im Kindesalter. Kinderklinik und Kinderpoliklinik im Dr. von Haunerschen Kinderspital der. München, Ludwig-Maximilians-Universität

- Secrist, H., C. J. Chelen, Y. Wen, J. D. Marshall and D. T. Umetsu (1993). "Allergen immunotherapy decreases interleukin 4 production in CD4+ T cells from allergic individuals." *J Exp Med* **178**(6): 2123-30.
- Senthilselvan, A., D. Rennie, L. Chenard, L. H. Burch, L. Babiuk, D. A. Schwartz, et al. (2008). "Association of polymorphisms of toll-like receptor 4 with a reduced prevalence of hay fever and atopy." *Ann Allergy Asthma Immunol* **100**(5): 463-8.
- Servin, B. and M. Stephens (2007). "Imputation-based analysis of association studies: candidate regions and quantitative traits." *PLoS Genet* **3**(7): e114.
- Setakis, E., H. Stirnadel and D. J. Balding (2006). "Logistic regression protects against population structure in genetic association studies." *Genome Res* **16**(2): 290-6.
- Siroux, V., E. Bouzigon, M. H. Dizier, I. Pin, F. Demenais and F. Kauffmann (2008). "Replication of association between ADAM33 polymorphisms and psoriasis." *PLoS ONE* **3**(6): e2448.
- Skol, A. D., L. J. Scott, G. R. Abecasis and M. Boehnke (2006). "Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies." *Nat Genet* **38**(2): 209-13.
- Sleiman, P. M., K. Annaiah, M. Imielinski, J. P. Bradfield, C. E. Kim, E. C. Frackelton, et al. (2008). "ORMDL3 variants associated with asthma susceptibility in North Americans of European ancestry." *J Allergy Clin Immunol*.
- Smit, J. J., G. Folkerts and F. P. Nijkamp (2004). "Mycobacteria, genes and the 'hygiene hypothesis'." *Curr Opin Allergy Clin Immunol* **4**(1): 57-62.
- Smit, L. A., D. Heederik, G. Doekes, C. Blom, I. van Zweden and I. M. Wouters (2008). "Exposure-response analysis of allergy and respiratory symptoms in endotoxin-exposed adults." *Eur Respir J* **31**(6): 1241-8.
- Smit, L. A., V. Siroux, E. Bouzigon, M. P. Oryszczyn, M. Lathrop, F. Demenais, et al. (2008). "CD14 and Toll-like Receptor Gene Polymorphisms, Country Living, and Asthma in Adults." *Am J Respir Crit Care Med*.
- Soderhall, C., I. Marenholz, T. Kerscher, F. Ruschendorf, J. Esparza-Gordillo, M. Worm, et al. (2007). "Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis." *PLoS Biol* **5**(9): e242.
- Soutar, A., A. Seaton and K. Brown (1997). "Bronchial reactivity and dietary antioxidants." *Thorax* **52**(2): 166-70.
- Stephens, J. C., J. A. Schneider, D. A. Tanguay, J. Choi, T. Acharya, S. E. Stanley, et al. (2001). "Haplotype variation and linkage disequilibrium in 313 human genes." *Science* **293**(5529): 489-93.
- Stephens, M. and P. Donnelly (2003). "A comparison of bayesian methods for haplotype reconstruction from population genotype data." *Am J Hum Genet* **73**(5): 1162-9.
- Stephens, M., N. J. Smith and P. Donnelly (2001). "A new statistical method for haplotype reconstruction from population data." *Am J Hum Genet* **68**(4): 978-89.
- Stevens, V. L., A. W. Hsing, J. T. Talbot, S. L. Zheng, J. Sun, J. Chen, et al. (2008). "Genetic variation in the toll-like receptor gene cluster (TLR10-TLR1-TLR6) and prostate cancer risk." *Int J Cancer* **123**(11): 2644-50.
- Stone, A. L., M. Kroeger and Q. X. Sang (1999). "Structure-function analysis of the ADAM family of disintegrin-like and metalloproteinase-containing proteins (review)." *J Protein Chem* **18**(4): 447-65.
- Strachan, D. P. (1989). "Hay fever, hygiene, and household size." *Bmj* **299**(6710): 1259-60.
- Stumpf, M. P. and D. B. Goldstein (2003). "Demography, recombination hotspot intensity, and the block structure of linkage disequilibrium." *Curr Biol* **13**(1): 1-8.
- Sun, L., R. V. Craiu, A. D. Paterson and S. B. Bull (2006). "Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies." *Genet Epidemiol* **30**(6): 519-30.

- Sun, Q., J. Yang, G. Xing, Q. Sun, L. Zhang and F. He (2008). "Expression of GSDML Associates with Tumor Progression in Uterine Cervix Cancer." Transl Oncol **1**(2): 73-83.
- Suttner, K., I. Ruoss, P. Rosenstiel, M. Depner, L. A. Pinto, M. Schedel, et al. (2009). "HLX1 gene variants influence the development of childhood asthma." J Allergy Clin Immunol **123**(1): 82-88 e6.
- Tabor, H. K., N. J. Risch and R. M. Myers (2002). "Candidate-gene approaches for studying complex genetic traits: practical considerations." Nat Rev Genet **3**(5): 391-7.
- Takeuchi, O., T. Kawai, P. F. Muhlrardt, M. Morr, J. D. Radolf, A. Zychlinsky, et al. (2001). "Discrimination of bacterial lipoproteins by Toll-like receptor 6." Int Immunol **13**(7): 933-40.
- Tantisira, K., W. T. Klimecki, R. Lazarus, L. J. Palmer, B. A. Raby, D. J. Kwiatkowski, et al. (2004). "Toll-like receptor 6 gene (TLR6): single-nucleotide polymorphism frequencies and preliminary association with the diagnosis of asthma." Genes Immun **5**(5): 343-6.
- Tavendale, R., D. F. Macgregor, S. Mukhopadhyay and C. N. Palmer (2008). "A polymorphism controlling ORMDL3 expression is associated with asthma that is poorly controlled by current medications." J Allergy Clin Immunol **121**(4): 860-3.
- Teo, Y. Y. (2008). "Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure." Curr Opin Lipidol **19**(2): 133-43.
- Thomas, D. C., R. W. Haile and D. Duggan (2005). "Recent developments in genomewide association scans: a workshop summary and review." Am J Hum Genet **77**(3): 337-45.
- Thomas, D. C. and J. S. Witte (2002). "Point: population stratification: a problem for case-control studies of candidate-gene associations?" Cancer Epidemiol Biomarkers Prev **11**(6): 505-12.
- Thompson, D., D. Stram, D. Goldgar and J. S. Witte (2003). "Haplotype tagging single nucleotide polymorphisms and association studies." Hum Hered **56**(1-3): 48-55.
- Tian, C., R. M. Plenge, M. Ransom, A. Lee, P. Villoslada, C. Selmi, et al. (2008). "Analysis and application of European genetic substructure using 300 K SNP information." PLoS Genet **4**(1): e4.
- Van Eerdewegh, P., R. D. Little, J. Dupuis, R. G. Del Mastro, K. Falls, J. Simon, et al. (2002). "Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness." Nature **418**(6896): 426-30.
- Vasselon, T., P. A. Detmers, D. Charron and A. Haziot (2004). "TLR2 recognizes a bacterial lipopeptide through direct binding." J Immunol **173**(12): 7401-5.
- Vercelli, D. (2008). "Discovering susceptibility genes for asthma and allergy." Nat Rev Immunol **8**(3): 169-82.
- Vermeulen, S. H., M. Den Heijer, P. Sham and J. Knight (2007). "Application of multi-locus analytical methods to identify interacting loci in case-control studies." Ann Hum Genet **71**(Pt 5): 689-700.
- von Ehrenstein, O. S., E. von Mutius, E. Maier, T. Hirsch, D. Carr, W. Schaal, et al. (2002). "Lung function of school children with low levels of alpha1-antitrypsin and tobacco smoke exposure." Eur Respir J **19**(6): 1099-106.
- von Mutius, E. (1999). Epidemiologie allergischer Erkrankungen im Kindesalter. Pädiatrische Allergologie und Immunologie. U. Wahn, R. Seger & V. Wahn. München Jena, Urban&Fischer: S. 159-172.
- von Mutius, E. (2007). "Allergies, infections and the hygiene hypothesis--the epidemiological evidence." Immunobiology **212**(6): 433-9.
- von Mutius, E. (2009). "Gene-environment interactions in asthma." J Allergy Clin Immunol **123**(1): 3-11.

- von Mutius, E., S. K. Weiland, C. Fritzsche, H. Duhme and U. Keil (1998). "Increasing prevalence of hay fever and atopy among children in Leipzig, East Germany." Lancet **351**(9106): 862-6.
- Wacholder, S., N. Rothman and N. Caporaso (2002). "Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer." Cancer Epidemiol Biomarkers Prev **11**(6): 513-20.
- Wall, J. D. and J. K. Pritchard (2003). "Haplotype blocks and linkage disequilibrium in the human genome." Nat Rev Genet **4**(8): 587-97.
- Wang, J. H., N. Avitahl, A. Cariappa, C. Friedrich, T. Ikeda, A. Renold, et al. (1998). "Aiolos regulates B cell activation and maturation to effector state." Immunity **9**(4): 543-53.
- Wang, N., J. M. Akey, K. Zhang, R. Chakraborty and L. Jin (2002). "Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation." Am J Hum Genet **71**(5): 1227-34.
- Wang, P., Q. J. Liu, J. S. Li, H. C. Li, C. H. Wei, C. H. Guo, et al. (2006). "Lack of association between ADAM33 gene and asthma in a Chinese population." Int J Immunogenet **33**(4): 303-6.
- Wang, W. Y., B. J. Barratt, D. G. Clayton and J. A. Todd (2005). "Genome-wide association studies: theoretical and practical concerns." Nat Rev Genet **6**(2): 109-18.
- Wang, W. Y., H. J. Cordell and J. A. Todd (2003). "Association mapping of complex diseases in linked regions: estimation of genetic effects and feasibility of testing rare variants." Genet Epidemiol **24**(1): 36-43.
- Warner, S. M. and D. A. Knight (2008). "Airway modeling and remodeling in the pathogenesis of asthma." Curr Opin Allergy Clin Immunol **8**(1): 44-8.
- Weidinger, S., C. Gieger, E. Rodriguez, H. Baurecht, M. Mempel, N. Klopp, et al. (2008). "Genome-wide scan on total serum IgE levels identifies FCER1A as novel susceptibility locus." PLoS Genet **4**(8): e1000166.
- Weidinger, S., M. O'Sullivan, T. Illig, H. Baurecht, M. Depner, E. Rodriguez, et al. (2008). "Filaggrin mutations, atopic eczema, hay fever, and asthma in children." J Allergy Clin Immunol **121**(5): 1203-1209 e1.
- Weiland, S. K., E. von Mutius, T. Hirsch, H. Duhme, C. Fritzsche, B. Werner, et al. (1999). "Prevalence of respiratory and atopic disorders among children in the East and West of Germany five years after unification." Eur Respir J **14**(4): 862-70.
- Weiss, K. M. and A. G. Clark (2002). "Linkage disequilibrium and the mapping of complex human traits." Trends Genet **18**(1): 19-24.
- Wellek, S. (2004). "Tests for establishing compatibility of an observed genotype distribution with Hardy-Weinberg equilibrium in the case of a biallelic locus." Biometrics **60**(3): 694-703.
- Wenzel, S. E., S. Balzar, E. Ampleford, G. A. Hawkins, W. W. Busse, W. J. Calhoun, et al. (2007). "IL4R alpha mutations are associated with asthma exacerbations and mast cell/IgE expression." Am J Respir Crit Care Med **175**(6): 570-6.
- Werner, M., N. Herbon, H. Gohlke, J. Altmüller, M. Knapp, J. Heinrich, et al. (2004). "Asthma is associated with single-nucleotide polymorphisms in ADAM33." Clin Exp Allergy **34**(1): 26-31.
- Werner, M., R. Topp, K. Wimmer, K. Richter, W. Bischof, M. Wjst, et al. (2003). "TLR4 gene variants modify endotoxin effects on asthma." J Allergy Clin Immunol **112**(2): 323-30.
- Wjst, M. (2007). "Public data mining shows extended linkage disequilibrium around ADAM33." Allergy **62**(4): 444-6.
- Wjst, M. (2008). "ORMDL3--guilt by association?" Clin Exp Allergy **38**(10): 1579-81.

- Wolfsberg, T. G., P. D. Straight, R. L. Gerena, A. P. Huovila, P. Primakoff, D. G. Myles, et al. (1995). "ADAM, a widely distributed and developmentally regulated gene family encoding membrane proteins with a disintegrin and metalloprotease domain." *Dev Biol* **169**(1): 378-83.
- Wong, M. Y., N. E. Day, J. A. Luan, K. P. Chan and N. J. Wareham (2003). "The detection of gene-environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement?" *Int J Epidemiol* **32**(1): 51-7.
- Wu, H., I. Romieu, J. J. Sienra-Monge, H. Li, B. E. Del Rio-Navarro and S. J. London (2008). "Genetic variation in ORM1-like 3 (ORMDL3) and gasdermin-like (GSDML) and childhood asthma." *Allergy*.
- Xue, X. and D. R. Hoover (2009). "Statistical methods in cancer epidemiological studies." *Methods Mol Biol* **471**: 239-72.
- Yuan, H. Y., J. J. Chiou, W. H. Tseng, C. H. Liu, C. K. Liu, Y. J. Lin, et al. (2006). "FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization." *Nucleic Acids Res* **34**(Web Server issue): W635-41.
- Zang, Y., H. Zhang, Y. Yang and G. Zheng (2007). "Robust genomic control and robust delta centralization tests for case-control association studies." *Hum Hered* **63**(3-4): 187-95.
- Zaykin, D. V., P. H. Westfall, S. S. Young, M. A. Karnoub, M. J. Wagner and M. G. Ehm (2002). "Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals." *Hum Hered* **53**(2): 79-91.
- Zeng, Z. B., T. Wang and W. Zou (2005). "Modeling quantitative trait Loci and interpretation of models." *Genetics* **169**(3): 1711-25.
- Zhang, K., P. Calabrese, M. Nordborg and F. Sun (2002). "Haplotype block structure and its applications to association studies: power and study designs." *Am J Hum Genet* **71**(6): 1386-94.
- Zhang, Y., N. I. Leaves, G. G. Anderson, C. P. Ponting, J. Broxholme, R. Holt, et al. (2003). "Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma." *Nat Genet* **34**(2): 181-6.
- Zhao, J. H., D. Curtis and P. C. Sham (2000). "Model-free analysis and permutation tests for allelic associations." *Hum Hered* **50**(2): 133-9.
- Zhao, Z., N. Timofeev, S. W. Hartley, D. K. Chui, S. Fucharoen, M. H. Steinberg, et al. (2008). "Imputation of Missing Genotypes: An Empirical Evaluation of IMPUTE." *BMC Genet* **9**(1): 85.
- Zheng, G., B. Freidlin and J. L. Gastwirth (2006). "Robust genomic control for association studies." *Am J Hum Genet* **78**(2): 350-6.
- Zheng, G., J. Joo, C. Zhang and N. L. Geller (2007). "Testing association for markers on the X chromosome." *Genet Epidemiol* **31**(8): 834-43.
- Zhu, M. and S. Zhao (2007). "Candidate gene identification approach: progress and challenges." *Int J Biol Sci* **3**(7): 420-7.
- Ziegler, A., I. R. Konig and J. R. Thompson (2008). "Biostatistical aspects of genome-wide association studies." *Biom J* **50**(1): 8-28.
- Zou, J., F. Zhu, J. Liu, W. Wang, R. Zhang, C. G. Garlisi, et al. (2004). "Catalytic activity of human ADAM33." *J Biol Chem* **279**(11): 9818-30.

Danksagung

Sehr herzlich möchte ich mich bei Frau Professor Erika von Mutius bedanken, ohne die diese Arbeit nicht zustande gekommen wäre. Neben der ausgezeichneten fachlichen Beratung schätze ich besonders, dass sie mich in meinen Ideen unterstützt und mich immer wieder neu motiviert hat.

Mein besonderer Dank gilt Herrn Professor Michael Kabesch, der mich in vielen Jahren in meiner wissenschaftlichen Arbeit entscheidend geprägt hat und unter dessen Anleitung die drei Publikationen entstanden sind, die der hier vorgelegten Arbeit zugrunde liegen. Damit verbunden ist auch mein Dank für die Überlassung der Datensätze für die im Rahmen dieser Arbeit durchgeführten Analysen.

Nicht nur Herrn Kabesch selbst, sondern auch allen MitarbeiterInnen der von ihm geführten Arbeitsgruppe gilt mein Dank für die freundliche und kollegiale Unterstützung, insbesondere auch für die Hilfestellungen bei der Einarbeitung in die biologischen Grundlagen der genetischen Epidemiologie. Dazu gehören Dr. Michael Kormann, der mich stets mit wertvollen Anregungen und anspornenden Diskussionen begleitete, Sven Michel, von dem ich detaillierte Informationen zu den genomweiten Analysen bekommen konnte, Dr. Michaela Schedel, Kathrin Suttner, Sonja Zeilinger, Anja Pleiß und Ilona Dahmen.

Außerdem möchte ich mich bei meinen KollegInnen im Statistikerzimmer für konstruktive Kritik und jederzeit hilfreiche Unterstützung bedanken. Dazu gehören Dr. Markus Ege, Dr. Sabina Illi und Dr. Mascha Rochat.

Letztlich bedanke ich mich bei meiner Familie und allen FreundInnen, die mich während dieser Zeit intensiven Arbeitens begleitet haben.

Lebenslauf

Persönliche Daten

Martin Depner
Geburtsdatum: 20.02.1967
Geburtsort: Pforzheim
Staatsangehörigkeit: deutsch
Familienstand: ledig
Adresse: Werneckstraße 11, 80802 München

Schulbildung

1986 Abitur am Reuchlin-Gymnasium, Pforzheim

Zivildienst:

1987-1988 20 Monate Zivildienst beim Deutschen Roten Kreuz in Pforzheim

Studium:

1989-1998 Studium der Psychologie und Mathematik an der Universität Koblenz-Landau und der Freien Universität Berlin
1991 Vordiplom in Psychologie, Universität Koblenz-Landau
1995 Vordiplom in Mathematik, FU Berlin
1998 Diplom in Psychologie, FU Berlin

Berufliche Tätigkeiten

1998 – 2001 Wissenschaftlicher Assistent bei „apollis – Sozialforschung und Demoskopie“ in Bozen/ Südtirol
2001 – 2002 Wissenschaftlicher Mitarbeiter an der Universität Konstanz, Abteilung für empirische Methoden
2003 - 2004 Wissenschaftlicher Mitarbeiter am Zentralinstitut für Seelische Gesundheit, Mannheim
seit 2005 Biostatistiker am Dr. von Haunerschen Kinderspital der LMU München, Arbeitsgruppen Allergogenetik und Pneumologie-Allergologie

