

# Kapitel 3

## Ergebnisse

Die folgenden Abschnitte 3.1 bis 3.3 zeigen zunächst Ergebnisse, die sich aus der Auswertung der Stichprobe deutscher Großfamilien ergaben. Ausgewählte Analysen konnten im Rahmen des Projektes veröffentlicht werden (Bauerfeind et al., 2002; Knoblauch et al., 2002, 2004; Nürnberg et al., 2004). In 3.3 ist die Assoziation der Kandidatengene dargestellt, wobei der Ansatz zur Bemessung des Beitrags einzelner Kandidatenloci in methodischer Hinsicht eine weiterentwickelte Form des in (Knoblauch et al., 2004) publizierten Ansatzes präsentiert. Ergebnisse der Validierungsstudie (Abschnitt 3.4) wurden unter Einbeziehung einer Bevölkerungsstichprobe aus Genf durchgeführt (Publikation im Druck (Bauerfeind et al., 2006)). Durch die aktuelle Veröffentlichung der HapMap-Daten (The International HapMap Consortium, 2005) wurde ein Vergleich mit der genomischen Struktur einer verwandten Bevölkerungsgruppe möglich (Abschnitt 3.5).

### 3.1 Phänotypische Charakterisierung

Als Phänotyp werden Cholesterin-Konzentrationen im Blutplasma betrachtet. Dazu gehören Gesamtcholesterin (TC), HDL-Cholesterin (HDL), LDL-Cholesterin (LDL) und Triglyzeride (TG). Der kombinierte Lipid-Phänotyp LH-Ratio ist ein klinisch relevanter Risikofaktor und bezeichnet das Verhältnis von LDL zu HDL.

Tabelle 3.1 zeigt den Median und die Spannweite für das Alter und die Serum-Lipidkonzentrationen. Wie erwartet sind die Werte für Cholesterin in der elterlichen Generation größer. Männer haben höhere Werte für TC und LDL als Frauen. Die Werte sind in mg/dl angegeben und weichen von der Normalverteilung ab. Für die statistische Analyse wurden die Werte logarithmisch transformiert.

Abbildung 3.1 zeigt die Verteilungsgüte der transformierten standardisierten LDL- und HDL-Werte in einem Histogramm. Die Darstellung am Beispiel LDL und HDL zeigt, dass

	N Alter		TC	HDL	LDL	TG
Großväter	33	63	252 (199 – 353)	174 (127 – 298)	39 (29 – 53)	135 (52 – 203)
Großmütter	31	64	244 (201 – 401)	156 (124 – 323)	51 (21 – 86)	121 (78 – 235)
Väter	232	47	233 (118 – 435)	151 (42 – 353)	46 (25 – 93)	142 (42 – 1615)
Mütter	232	46	231 (137 – 388)	143 (51 – 303)	58 (23 – 109)	119 (34 – 503)
Söhne	245	25	187 (111 – 338)	118 (42 – 257)	51 (24 – 96)	89 (29 – 397)
Töchter	273	26	192 (118 – 385)	115 (43 – 296)	57 (25 – 97)	83 (26 – 292)
Gesamt	1054	36	207 (111 – 435)	127 (42 – 353)	53 (21 – 109)	97 (26 – 1615)

Tabelle 3.1: Darstellung von Median und Spannweite (in Klammern) der Serumlipidwerte: Gesamtcholesterin (TC), LDL-Cholesterin (LDL), HDL-Cholesterin (HDL), LDL/HDL-Ratio und Triglyzeride (TG) in mg/dl und des Alters in Jahren von 1054 Individuen aus 200 Großfamilien; getrennt nach Generation und Geschlecht.

die Transformation (approximativ) normalverteilte Werte liefert. Der Kolmogorov-Smirnov-Test zeigt eine Abweichung von  $N(0,1)$  für Triglyzeride und HDL. Die Verteilung des HDL ist leicht linksschief. Der Hauptfokus der vorliegenden Auswertungen liegt auf LDL- und HDL-Cholesterin; Ergebnisse der TG-Modelle werden angegeben, aber nicht im einzelnen diskutiert. Einflüsse des Alters, des Geschlechts und des BMI auf die Lipid-Phänotypen werden als Kofaktoren berücksichtigt.

## 3.2 Genomische Karten und Kopplungsungleichgewicht

Die Tabelle 3.2 zeigt die Liste untersuchter SNPs mit Identifikationsnummer, Lokalisation (Promotor, Exon, Intron, intergenisch), Allelfrequenz, der Sequenzinformation und genomischer Position. Die letzte Spalte gibt an, ob der jeweilige SNP im HapMap-Projekt (The International HapMap Consortium, 2005) genotypisiert wurde.

SNP	ID	Lokalisation	Allel- frequenz	Basen- tausch	Chrom. Position	HapMap- SNP
<i>APOB</i> : Apolipoprotein B (Chromosom 2)						
1	rs512535	Promotor	45%	G/A	21179434	x
2	rs1367117	Exon/ non-syn	32%	C/T	21175552	x
3	rs520354	Intron	44%	T/C	21171264	x
4	rs679899	Exon/ non-syn	44%	C/T	21162566	

Fortsetzung auf der nächsten Seite

Fortsetzung Tabelle 3.2 von letzter Seite

SNP	ID	Lokalisation	Allel- frequenz	Basen- tausch	Chrom. Position	HapMap- SNP
5	rs673548	Intron	21%	C/T	21149196	x
6	rs693	Exon/ syn	50%	T/C	21143847	x
7	rs1800479	Intron	16%	C/G	21139035	
8	rs1801703	Exon/ non-syn	1%	C/T	21137564	x
9	rs1042034	Exon/ non-syn	22%	T/C	21136933	x
<i>APOE</i> : Apolipoprotein E (Chromosom 19)						
1	rs449647	Promotor	17%	A/T	50100404	
2	rs405509	Promotor	48%	C/A	50100676	x
3	rs440446	Intron	36%	G/C	50101007	
4	rs769450	Intron	41%	C/T	50102284	
5	rs429358	Exon/ non-syn	16%	T/C	50103781	
6	rs7412	Exon/ non-syn	6%	C/T	50103919	
<i>ABCA1</i> : ATP-binding cassette, sub-family A, member 1 (Chromosom 9)						
1	rs2422493	Promotor	45%	G/A	104770550	
2	rs1800977	5'UTR	50%	C/T	104770005	x
3	unbekannt	5'UTR	4%	C/G	104769958	
4	rs2575879	Intron	46%	G/C	104748380	x
5	rs1800978	5'UTR	1%	G/C	104745533	x
6	rs1999429	Intron	3%	G/T	104720341	
7	rs2230806	Exon/ non-syn	26%	C/T	104700422	x
8	rs2274873	Exon/ syn	8%	C/T	104682233	x
9	rs4149313	Exon/ non-syn	13%	T/C	104666308	x
10	unbekannt	Exon/ non-syn	3%	C/G	104659188	
10a	unbekannt	3'UTR	13%	T/G	104653236	
11	rs2066716	Exon/ syn	8%	G/A	104648260	x
11a	rs2777801	3'UTR	7%	G/A	104646432	x
12	rs363717	3'UTR	20%	A/G	104624255	x
<i>CETP</i> : Cholesteryl Ester Transfer Protein (Chromosom 16)						
1	rs4783961	Promotor	47%	C/T	55552395	x
2	rs1800776	Promotor	7%	G/T	55552735	
3	rs1800775	Promotor	49%	G/T	55552737	x
4	rs711752	Intron	45%	C/T	55553712	x
5	rs708272	Intron	45%	C/T	55553789	
5a	rs7205804	Intron	42%	G/A	55562390	x
6	rs1532625	Intron	43%	G/A	55562802	
7	rs158478	Intron	46%	C/A	55565235	
8	rs289718	Intron	29%	T/C	55567433	
9	rs289719	Intron	29%	C/T	55567442	x
10	rs5880	Exon/ non-syn	4%	C/G	55572592	x
11	rs5882	Exon/ non-syn	35%	A/G	55573593	x
12	rs1801706	3'UTR	18%	C/T	55575163	x
<i>LCAT</i> : Lecithin-Cholesteryl-Acyl-Transferase (Chromosom 16)						

Fortsetzung auf der nächsten Seite

Fortsetzung Tabelle 3.2 von letzter Seite

SNP	ID	Lokalisation	Allel- frequenz	Basen- tausch	Chrom. Position	HapMap- SNP
1	rs1109166	Intron	17%	A/G	66534883	x
2	rs4986970	Exon/ non-syn	4%	A/T	66533821	x
3	rs5923	Exon/ syn	4%	G/A	66531454	x
<i>LIPC</i> : Hepatische Triglyzeridlipase (Chromosom 15)						
1	rs723967	5'UTR	46%	G/A	56483592	x
2	rs1800588	Promotor	28%	G/A	56510967	x
2a	rs2070895	Promotor	27%	G/A	56511231	x
3	rs1869144	Intron	39%	A/G	56584592	x
3a	unbekannt	5'UTR	44%	T/C	56620995	
4	rs6078	Exon/ non-syn	4%	G/A	56621285	x
5	rs690	Exon/ syn	49%	T/G	56622033	x
6	rs6082	Exon/ syn	8%	A/G	56625249	x
7	rs6083	Exon/ non-syn	36%	T/C	56625302	x
8	rs6084	Exon/ syn	44%	C/G	56625330	x
9	rs6074	Exon/ syn	13%	G/T	56648255	x
<i>LPL</i> : Hepatische Lipoproteinlipase (Chromosom 8)						
1	unbekannt	Promotor	0%	C/A	19840949	
2	rs1800590	Promotor	1%	A/C	19840951	x
3	rs1031045	Intron	1%	C/T	19845392	x
4	rs253	Intron	47%	C/T	19855697	x
5	rs268	Exon/ non-syn	3%	T/C	19857809	x
6	rs269	Intron	20%	A/C	19857947	
7	rs2075651	Intron	0%	C/A	19857948	
8	rs270	Intron	17%	G/T	19857956	x
9	rs285	Intron	48%	G/A	19859469	x
10	rs320	Intron	29%	A/C	19863357	x
11	rs328	Exon/ non-syn	10%	G/C	19864004	x
12	rs3289	3'UTR	2%	T/C	19867472	x
<i>LDLR</i> : Low Density Lipoprotein Rezeptor (Chromosom 19)						
1	rs2228671	Exon/ syn	9%	C/T	11071912	x
2	rs885765	Intron	8%	T/C	11077700	
3	rs5930	Exon/ syn	40%	G/A	11085265	x
4	rs5925	Exon/ syn	37%	A/G	11091881	
5	rs5927	Exon/ syn	28%	C/T	11094941	x
6	rs2569542	Intron	27%	G/A	11098773	
7	rs1433099	3'UTR	28%	G/A	11103658	x
<i>APOC2</i> : Apolipoprotein C-II (Chromosom 19)						
1	rs5167	Promotor	36%	A/C	50140305	x
2	rs2288911	Promotor	50%	C/A	50141124	x
3	rs5120	Intron	49%	T/A	50143460	
4	rs892101	3'UTR	46%	C/T	50147298	
Gencluster: APO-A-V/-A-IV/-C-III/-A-I (Chromosom 11)						

Fortsetzung auf der nächsten Seite

Fortsetzung Tabelle 3.2 von letzter Seite

SNP	ID	Lokalisation	Allelfrequenz	Basenaustausch	Chrom. Position	HapMap-SNP
<i>APOA5</i>						
1	rs2266788	3'UTR	0%	A/G	116165896	x
2	rs2072560	Intron	7%	G/A	116167036	
3	rs3135506	Exon/ non-syn	8%	C/G	116167617	
4	rs662799	Promotor	7%	T/C	116168917	x
5	rs2266789	intergenisch	0%	C/T	116180026	
6	rs1729408	intergenisch	36%	T/C	116180028	
<i>APOA4</i>						
1	rs675	Exon/ non-syn	22%	A/T	116196885	
2	rs5104	Exon/ non-syn	14%	A/G	116197544	
3	rs5092	Exon/ syn	16%	T/C	116198674	
4	rs2542051	intergenisch	38%	T/G	116202948	
<i>APOC3</i>						
1	rs2854116	Promotor	40%	A/G	116205379	
2	rs4520	Exon/ syn	28%	C/T	116206745	
3	rs5128	3'UTR	9%	G/C	116208850	
<i>APOA1</i>						
1	rs525028	intergenisch	33%	T/C	116210726	
2	rs12718463	3'UTR	2%	T/C	116211553	
3	rs5081	3'UTR	3%	A/T	116211556	
4	rs5070	Intron	34%	G/A	116213230	
5	rs1799837	Intron	1%	C/T	116213463	
6	rs5069	5'UTR	1%	G/A	116213464	
<i>PLTP</i> : Phospholipid Transferprotein (Chromosom 20)						
1	rs2294213	UTR	11%	C/G	43974099	x
<i>SRB1</i> : Scavenger Rezeptor, class B Typ I (Chromosom 12)						
1	rs5888	Intron	49%	G/A	123809628	

Tabelle 3.2: Charakterisierung der SNPs mit Identifikationsnummer (SNP ID), Lokalisation, Allelfrequenz, Basenaustausch und chromosomaler Position. Die SNPs wurden aus der Literatur und öffentlichen Datenbanken (SNP-Konsortium: <http://www.ncbi.nlm.nih.gov/SNP/>) ausgewählt.

SNPs sind in der Tabelle 3.2 separat für die verschiedenen Gene entsprechend ihrer genomischen Position (Angabe in Basenpaaren) auf dem Chromosom angeordnet. 21 SNPs haben eine Allelfrequenz < 5%, 12 SNPs haben eine Frequenz zwischen 5% und 10%, 12 SNPs eine Frequenz zwischen 10% und 20% und 48 SNPs eine Frequenz > 20%. Mehr als ein Drittel befinden sich in Exons, ein Drittel in Introns und die verbleibenden in nicht abgelesenen 5' - und 3' - Regionen der entsprechenden Gene (UTR= „untranslated region“). Die Allelfrequenzen wurden aus unabhängigen „Foundern“ (Eltern, Gründern) berechnet.

Alle untersuchten SNPs befinden sich im Hardy-Weinberg-Gleichgewicht, ein Hinweis

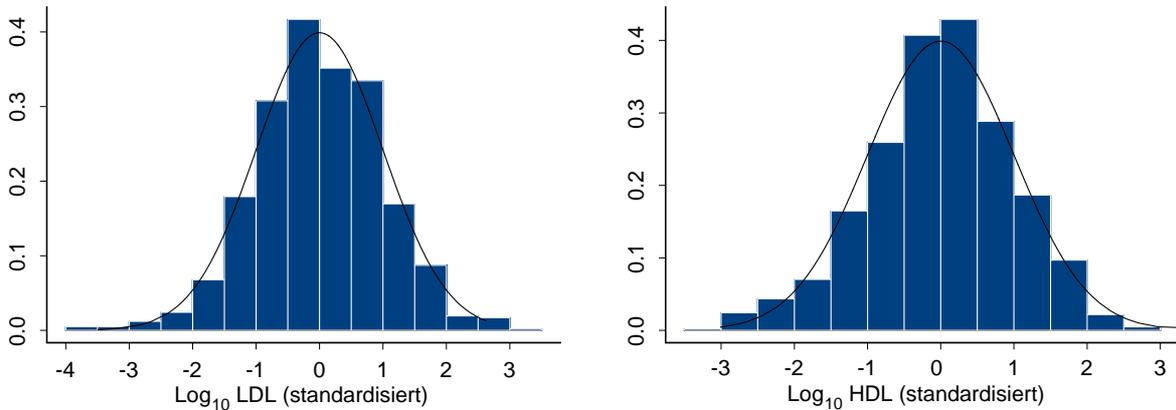


Abbildung 3.1: Darstellung der Verteilung standardisierter  $\log_{10}$ -transformierter LDL- und HDL-Cholesterin-Konzentrationen in einem Histogramm mit Dichtefunktion der Standardnormalverteilung  $N(0, 1)$ .

auf „random-mating“ (zufällige Bildung von Paaren hinsichtlich untersuchter Genotypen) in unserer Studienpopulation.

Die Abbildungen 3.2 und 3.3 zeigen als Beispiel die genomischen Karten des Apoprotein-B-Gens (*APOB*) und des Gens für das Enzym Hepatische Lipase (*LIPC*) mit untersuchten SNP-Positionen.

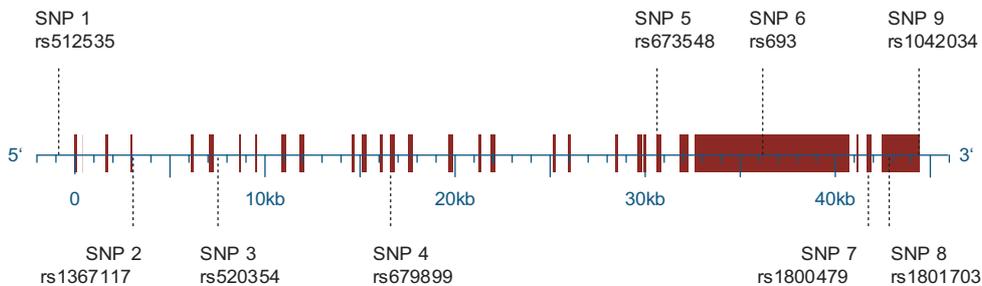


Abbildung 3.2: Genomische Karte des *APOB*-Gens und untersuchte SNP-Marker.

Nur einige Segmente enthalten Informationen, die das Genprodukt kodieren (Exons: rot). Dazwischen liegen die Introns. SNP-Positionen sind in kb angegeben und wurden aus der NCBI-Datenbank abgelesen (<http://www.ncbi.nlm.nih.gov>).

Es wurden 9 SNPs im *APOB*-Gen und 13 SNPs im *LIPC*-Gen für die Hepatische Lipase analysiert. Der durchschnittliche Abstand zwischen zwei SNPs im *APOB* ist mit 5.3 kb gering. Im *LIPC*-Gen ist der mittlere Abstand mit 16.5 kb weitaus höher und dementsprechend auch die Wahrscheinlichkeit des Auftretens von Rekombinationen. Dies zeigt sich unmittelbar in den LD-Werten zwischen den Allelen.

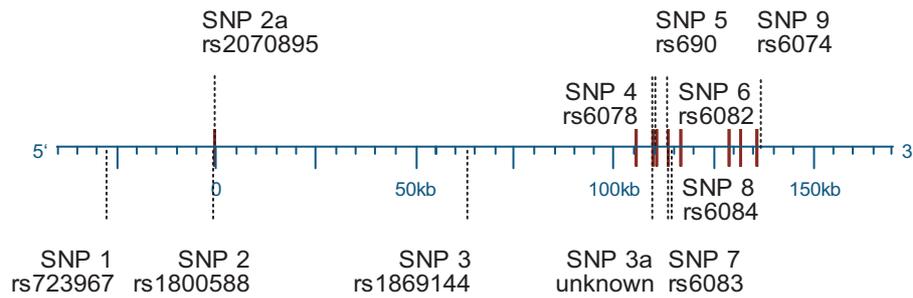


Abbildung 3.3: Genomische Karte des *LIPC*-Gens und untersuchte SNP-Marker.

Die Abbildungen 3.4 und 3.5 zeigen das paarweise LD berechnet als Korrelationskoeffizient  $r$  zwischen den Markerallelen. Die Werte wurden aus den Daten unverwandter Founder berechnet. Das Ergebnis ist eine symmetrische Matrix, die in Form einer oberen Dreiecksmatrix visualisiert wurde. Die Korrelationswerte sind angezeigt, und die Einfärbung der Quadrate symbolisiert die Signifikanz in einem  $\chi^2$ -Unabhängigkeitstest.

Der  $\chi^2$ -Test zeigt ein hohes Maß an Kopplungsungleichgewicht im *APOB*-Gen. Die Ausnahme bildet *APOB*-SNP8 (rs1801703). *APOB*-SNP8 ist ein seltener SNP mit einer Populationsfrequenz von 1%. Alle anderen *APOB*-SNPs haben eine Frequenz  $> 16\%$ . Dies legt die Vermutung nahe, dass es sich im Gegensatz zu den anderen SNPs bei *APOB*-SNP8 um eine neue Mutation handelt, die sich auf einen alten Haplotypblock aufgesetzt hat.

Die Positionen *LIPC*-SNP6-SNP7-SNP8 (rs6082/rs6083/rs6084), *LIPC*-SNP2-SNP2a (rs1800588/rs2070895) und *LIPC*-SNP3a-SNP4 (unbekannt/rs6078) bilden Haplotypblöcke, die anderen SNP-Kombinationen zeigen größtenteils Kopplungsgleichgewicht, d.h. geringen bzw. keinen Zusammenhang.

Genomische Karten und Abbildungen des Kopplungsungleichgewichtes weiterer untersuchter Kandidatengene sind im Anhang dargestellt (Abbildungen A.1 - A.8). Des Weiteren ist die komplette Liste der geschätzten Haplotypen und ihrer Frequenz angegeben (Tabelle A.1).

### 3.3 Einfluss multipler SNPs und SNP-Haplotypen auf die genetische Varianz von HDL- und LDL-Cholesterin

Tabelle 3.3 zeigt die Ergebnisse der Varianzkomponentenzerlegung basierend auf einem gemischten linearen Modell mit Kinship-Struktur unter dem Einfluss von 69 SNPs bzw. 74 SNP-Haplotypen untersuchter Gen-Loci.

	rs512535	rs1367117	rs520354	rs679899	rs673548	rs693	rs1800479	rs1801703	rs1042034
rs512535		0.74	-0.54	-0.52	-0.41	0.19	-0.05	-0.07	-0.41
rs1367117			-0.59	-0.58	-0.32	0.52	-0.23	-0.05	-0.32
rs520354				0.99	0.53	-0.58	0.43	0.09	0.53
rs679899					0.53	-0.58	0.45	0.09	0.52
rs673548						-0.51	-0.24	0.15	0.98
rs693							-0.46	-0.08	-0.52
rs1800479								0.01	-0.25
rs1801703									0.15
rs1042034	r								

Abbildung 3.4: *APOB*-Gen: Paarweises Kopplungsungleichgewicht (Korrelation  $r$ ) und Signifikanz in einem  $\chi^2$ -Unabhängigkeitstest ( $p < 0.01$  = rot;  $p < 0.025$  = orange;  $p < 0.05$  = gelb;  $p < 0.1$  = hellgelb).

Die Auswertung basiert auf 625 genotypisierten Individuen. Für jeden univariaten Phänotyp wurde ein Modell multipler SNPs oder SNP-Haplotypen (Allelfrequenz  $> 5\%$ ) als lineare Einflussvariablen mit einem Modell verglichen, das nur den Interzept, d.h. den globalen Mittelwert, modelliert. Dabei wurden SNPs eines Lokus mit einer mittleren bis hohen paarweisen Korrelation  $|r| > 0.7$  (Kopplungsungleichgewicht) ausgeschlossen. Die Signifikanz  $p(LR)$  ergibt sich aus dem Likelihood-Ratio-Test zwischen beiden Modellen. Dabei wird die Likelihooddifferenz zwischen beiden Modellen in einem  $\chi^2$ -Test mit der theoretischen Verteilung verglichen. Die Anzahl der Freiheitsgrade entspricht der Anzahl eingeschlossener SNPs bzw. SNP-Haplotypen.

Das empirische Signifikanzniveau  $p_{emp}$  wurde aus dem Vergleich der Likelihooddifferenz im originalen versus dem 1000-fach permutierten Modell bestimmt (siehe Methoden).

Die Schätzung der Heritabilität der quantitativen Lipidphänotypen stellt sich als Verhältnis der Schätzungen genetisch bedingter (polygener) Varianz  $v_{poly}$  zur phänotypischen Gesamtvarianz  $v$  im Modell ohne feste Effekte (Interzept-Modell) dar. Die phänotypische Gesamtvarianz wurde auf 1 normiert, so dass  $v_{poly}$  eine direkte Schätzung der Heritabilität

	rs723967	rs1800588	rs2070895	rs1869144	unknown	rs6078	rs690	rs6082	rs6083	rs6084	rsrs6074
rs723967		0.08	0.04	0.00	0.03	-0.07	-0.03	0.02	0.08	0.12	0.02
rs1800588			0.97	-0.12	-0.03	0.07	0.02	0.08	0.04	0.08	-0.04
rs2070895				-0.12	-0.03	0.09	0.03	0.05	0.03	0.07	-0.04
rs1869144					-0.07	-0.02	0.02	-0.04	-0.06	-0.11	0.00
unknown						-0.17	-0.62	-0.02	0.12	0.09	-0.06
rs6078							0.22	0.06	0.12	0.07	-0.04
rs690								0.03	-0.04	-0.11	0.07
rs6082									0.38	0.32	-0.12
rs6083										0.84	0.54
rs6084											0.45
rsrs6074	r										

Abbildung 3.5: *LIPC*-Gen: Paarweises Kopplungsungleichgewicht (Korrelation  $r$ ) und Signifikanz in einem  $\chi^2$ -Unabhängigkeitstest ( $p < 0.01$  = rot;  $p < 0.025$  = orange;  $p < 0.05$  = gelb;  $p < 0.1$  = hellgelb).

liefert. Wie bei der Analyse von Merkmalen mit starker erblicher Komponente zu erwarten ist, zeigt die Schätzung der polygenen Varianz  $v_{poly}$  für alle untersuchten Lipid-Phänotypen einen Wert größer als Null. Für Gesamtcholesterin können 24% der phänotypischen Gesamtvarianz genetischen Faktoren zugeschrieben werden, 11% für Triglyzeride, 28% für LDL, 37% für HDL und 27% für das LH-Ratio.

Der Einschluss genetischer Marker reduziert die polygene Restvarianz  $v_{poly}$  beträchtlich. Multiple SNPs haben einen signifikanten Einfluss auf TC ( $p = 0.02$ ), LDL ( $p = 0.00002$ ) und LH-Ratio ( $p = 0.0001$ ). Die geschätzten Anteile erklärter genetischer Varianz  $v_{locus}/v_{poly.I}$  liegen bei 82% für LDL, 75% für TC und 70% für das LH-Ratio im SNP-Modell.

Multiple Haplotypen beeinflussen LDL nur marginal ( $p = 0.05$ ). Für HDL zeigt das Modell multipler Haplotypen eine höhere Signifikanz ( $p = 0.003$ ) im Vergleich zum SNP-Modell. Der Anteil genetischer Varianz, der durch multiple Haplotypen erklärt werden konnte, wurde mit 71% für LDL, 65% für HDL und 100% für das LH-Ratio in vorliegenden Daten geschätzt.

Modell:	Interzept (I)		Multi-Lokus (M)				$p(LR)$	$p_{emp}$	$\frac{v_{locus}}{v_{poly.I}}$
	Phänotyp	$v_{env.I}$	$v_{poly.I}$	$v_{env.M}$	$v_{poly.M}$	$v_{locus}$			
(A) multiples SNP-Modell ( $q = 69$ )									
TC	0.75	0.24	0.78	0.06	<b>0.18</b>	0.02	0.051	75%	
TG	0.89	0.11	0.87	0.02	<b>0.09</b>	n.s.	n.s.		
LDL	0.71	0.28	0.74	0.05	<b>0.23</b>	0.00002	0.001	82%	
HDL	0.63	0.37	0.66	0.19	<b>0.18</b>	0.08	n.s.	46%	
LH-Ratio	0.72	0.27	0.73	0.08	<b>0.19</b>	0.0001	0.001	70%	
(B) multiples SNP-Haplotyp-Modell ( $q = 74$ )									
TC	0.75	0.24	0.74	0.12	<b>0.12</b>	n.s.	n.s.		
TG	0.89	0.11	0.87	0.00	<b>0.11</b>	n.s.	n.s.		
LDL	0.71	0.28	0.75	0.09	<b>0.19</b>	0.05	0.095	71%	
HDL	0.63	0.37	0.68	0.13	<b>0.24</b>	0.003	0.006	65%	
LH-Ratio	0.72	0.27	0.81	0.00	<b>0.27</b>	0.0009	0.003	100%	

Tabelle 3.3: Zerlegung phänotypischer Gesamtvarianz in Varianzkomponenten:  $v_{env}$  bezeichnet die Komponente umweltbezogener Varianz und  $v_{poly}$  die geschätzte polygene Restvarianz. Im „Interzept Modell“ werden  $v_{env}$  und  $v_{poly}$  als Komponenten der Lipidvarianz  $v$  ( $=1$ ) geschätzt. Im vollen Modell wurden 69 SNPs (A) bzw. SNP-Haplotypen (B) als lineare feste Effekte einbezogen und die Varianzkomponenten simultan aus den Residuen geschätzt. Die Signifikanz ( $p < 0.1$ ) multipler genetischer Marker wurde in einem Likelihood-Ratio-Test (LR-Test) bestimmt.  $p_{emp}$  wurde durch 1000-fache Permutation des Phänotyps im Multi-Lokus-Modell bestimmt.  $v_{locus}$  ergibt sich als  $v_{poly.I} - v_{poly.M}$  und ist ein Schätzer der erklärten genetischen Varianz. ( $v_{env.M} + v_{poly.M} + v_{locus} \neq 1$ , da  $v_{env.I} \neq v_{env.M}$  durch Einfluss von Modellfreiheitsgraden und/oder Interaktionseffekten.) Das Verhältnis  $v_{locus}/v_{poly.I}$  beschreibt den Anteil erklärter genetischer Varianz.

Tabelle 3.4 zeigt die geschätzten Konfidenzintervalle (KI) der Komponenten umweltbedingter und polygener Restvarianz aus  $N = 1000$  (familienbasierten) Bootstrap-Simulationen. Die Breite des KI zeigt die Variabilität der Varianzschätzungen. Die Werte sind durch den Wert 0 nach unten beschränkt. Durch die Standardisierung der phänotypischen Gesamtvarianz können sie maximal den Wert 1 erreichen und sind als relativer Anteil der Gesamtvarianz interpretierbar. Bis auf  $v_{poly.M}$  hat das Konfidenzintervall der Schätzungen eine Breite von durchschnittlich 0.3, d.h. es beträgt 30% der Gesamtvarianz. Dies entspricht den Schwankungen, die in verschiedenen Studienpopulationen bezüglich Heritabilität und Umweltanteil berichtet wurden.

Bis auf Triglyzeride schließen die Konfidenzintervalle der Heritabilitätskomponente  $v_{poly.I}$  die 0 nicht ein. Triglyzeridkonzentrationen sind in hohem Maße ernährungsabhängig und schwanken individuell sehr stark, somit ist dieser Effekt plausibel und es ergibt sich ein vergleichsweise großer Anteil umweltbedingter Varianz  $v_{e.I}$ . Die Heritabilität erreicht für HDL-Cholesterin den größten Wert  $\approx 50\%$ . In zwei Modellen sind die Konfidenzinter-

Phänotyp	Interzept-Modell		Multi-Lokus-Modell	
	$v_{env.I}$	$v_{poly.I}$	$v_{env.M}$	$v_{poly.M}$
(A) SNP-basierte Analyse				
TC	[0.58; 0.89]	[0.11; 0.41]	[0.57; 0.81]	[0.00; 0.17]
TG	[0.73; 1.00]	[0.00; 0.26]	[0.66; 0.85]	[0.00; 0.10]
LDL	[0.56; 0.85]	[0.15; 0.43]	[0.56; 0.76]	[0.00; 0.11]
HDL	[0.46; 0.79]	[0.21; 0.53]	[0.46; 0.78]	[0.00; 0.28]
LH-Ratio	[0.59; 0.91]	[0.10; 0.39]	[0.59; 0.78]	[0.00; 0.13]
(B) Haplotyp-basierte Analyse				
TC	[0.59; 0.94]	[0.06; 0.39]	[0.57; 0.82]	[0.00; 0.17]
TG	[0.70; 1.00]	[0.00; 0.29]	[0.66; 0.83]	[0.00; 0.08]
LDL	[0.55; 0.87]	[0.13; 0.43]	[0.57; 0.80]	[0.00; 0.14]
HDL	[0.48; 0.79]	[0.21; 0.52]	[0.46; 0.77]	[0.00; 0.23]
LH-Ratio	[0.57; 0.87]	[0.13; 0.42]	[0.61; 0.78]	[0.00; 0.07]

Tabelle 3.4: Konfidenzintervalle der Varianzkomponenten aus 1000 Block-Bootstrap-Simulationen.

valle der geschätzten genetischen Komponente  $v_{poly}$  zwischen Interzept- und Multi-Lokus-Assoziationsmodell deutlich differenziert, d.h. es gibt keine Überschneidungen. Dies ist zum einen das SNP-Modell für LDL und das Haplotyp-Modell für den LDL/HDL-Quotienten.

Tabelle 3.5 stellt die Signifikanzen aus dem Vergleich der Likelihoods des vollen Modells versus Modellen mit ausgeschlossenen Gen-Lokus in einem Likelihood-Ratio-Test dar. Beide Modelle wurden im Rahmen des gemischten Modells mit Kinship-Struktur modelliert. Ein p-Wert  $< 0.05$  indiziert, welcher der Gen-Loci unabhängig in der Lage ist, einen Teil der phänotypischen Varianz zu erklären. Der bekannte *APOE*-Effekt auf den Cholesterin-Phänotyp bestätigt sich hier im SNP-Modell mit größerer Signifikanz als im Haplotyp-Modell.

Offenbar ist durch die Schätzung der Haplotypen als Kombination von 6 häufigen *APOE*-SNPs der funktionelle *APOE*- $\epsilon$ -Effekt aus 2 maßgeblichen (hier enthaltenen) SNPs aufgesplittet und wegen der größeren Anzahl von Freiheitsgraden nicht so signifikant. *CETP*-SNPs haben einen Einfluss auf TC, LDL und den LDL/HDL-Quotienten. Durch *CETP*-Haplotypen kann zusätzlich ein signifikanter Einfluss auf HDL-Cholesterin nachgewiesen werden. *LDLR* zeigt einen Einfluss auf LDL und LH-Ratio, der nur durch SNPs nachgewiesen werden konnte. *LIPC*-Effekte können durch die Modellierung von SNP-Haplotypen nachgewiesen werden.

Die getrennte Modellierung für 272 Männer und 348 Frauen (Tabelle 3.6) erlaubt eine differenziertere Betrachtung der Gen-Lokus-Effekte. *APOB*-SNPs zeigen einen Einfluss auf LDL, der zuvor auf Grund von Interaktionen zwischen SNPs und Geschlecht offenbar nicht sichtbar war. Effekte des *APOE* sind weniger signifikant bei separater Betrachtung der

Gen-Lokus (Df)	TC	TG	LDL	HDL	LH-Ratio
(A) SNP-basierte Analyse					
<i>APOB</i> (8)					
<i>APOE</i> (6)	$1 \cdot 10^{-5}$		$1 \cdot 10^{-10}$		$3 \cdot 10^{-8}$
<i>ABCA1</i> (7)					
<i>CETP</i> (9)	0.01		0.006		0.01
<i>LCAT</i> (1)					
<i>LIPC</i> (8)					
<i>LPL</i> (6)					
<i>LDLR</i> (7)			0.004		0.002
<i>APOC2</i> (4)					
Gencluster (13)					
Gen-Lokus (Df)	TC	TG	LDL	HDL	LH-Ratio
(B) Haplotyp-basierte Analyse					
<i>APOB</i> (8)					
<i>APOE</i> (4)	0.03		0.0002		0.0002
<i>ABCA1</i> (8)					
<i>CETP</i> (9)	0.02		0.03	0.01	0.006
<i>LCAT</i> (3)					
<i>LIPC</i> (11)				0.01	0.04
<i>LPL</i> (9)					
<i>LDLR</i> (4)					
<i>APOC2</i> (3)					
Gencluster (15)					

Tabelle 3.5: Signifikanz des partiellen Lokus-Beitrags aus einem Likelihood-Ratio-Test des vollen Modells versus eines Modells mit ausgeschlossenen Gen-Lokus. Das volle Modell enthält multiple Varianten eines Gens als Regressoren, SNPs (A) oder Haplotypen (B). Im reduzierten Modell wurden die Varianten des getesteten Lokus ausgeschlossen. In Klammern ist die Anzahl der Freiheitsgrade angegeben. p-Werte sind angegeben, falls  $p < 0.05$ .

Geschlechter, was auf die geringere statistische Power durch Aufspaltung des Datensatzes deutet. Interessant ist die wesentlich geringere Bedeutung des *APOE* für den LDL/HDL-Quotienten der Männer, wobei hier wahrscheinlich HDL in verschiedener Weise (für Männer anders als für Frauen) durch *APOE* beeinflusst wird. *CETP*-Effekte sind nur signifikant bei Frauen, und diese sind in stärkerem Maße repräsentiert durch Haplotypen. Gleiches gilt für Effekte des Genclusters, die zusätzlich für Frauen nachgewiesen werden konnten. *LIPC*- und *LDLR*-Effekte bestätigen sich mit deutlich höherer Signifikanz nur bei Männern. *LPL*-Effekte sind unterschiedlich für Männer und Frauen und können durch SNPs und SNP-Haplotypen gezeigt werden. *LPL* ist signifikant für TC und LDL bei Frauen und für HDL bei Männern.

Neben der Signifikanz einzelner Gen-Loci ist das Ausmaß ihres Einflusses auf die phänotypische Varianz von besonderem Interesse. Die Signifikanz ist nicht ausschließlich von der Größe erklärbarer genetisch bedingter phänotypischer Varianz abhängig, sondern be-

Lokus	TC	TG	LDL	HDL	LH-Ratio
(A) SNP-basierte Analyse					
<i>APOB</i>			0.04/0.03		
<i>APOE</i>	0.002/0.0005		$3 \cdot 10^{-5}/6 \cdot 10^{-8}$		0.04/1 · 10 <sup>-6</sup>
<i>ABCA1</i>					
<i>CETP</i>	0.02		0.006	0.03	0.01
<i>LCAT</i>					
<i>LIPC</i>					
<i>LPL</i>	0.03		0.04	0.003	
<i>LDLR</i>			$9 \cdot 10^{-5}$	0.006	1 · 10 <sup>-6</sup>
<i>APOC2</i>					
Cluster	0.0002		0.0004		0.01
Gesamt	0.001		$0.006/2 \cdot 10^{-5}$		0.001/0.002
Lokus	TC	TG	LDL	HDL	LH-Ratio
(B) Haplotyp-basierte Analyse					
<i>APOB</i>					0.03
<i>APOE</i>	0.008		0.0004/0.005		0.002/0.0004
<i>ABCA1</i>		0.03			
<i>CETP</i>	0.03		0.007	0.006	0.0001
<i>LCAT</i>					
<i>LIPC</i>				0.0002	
<i>LPL</i>	0.006	0.03		0.02	
<i>LDLR</i>					
<i>APOC2</i>					
Cluster	0.005	0.01	0.02		0.02
Gesamt	0.007	0.04	0.002	0.002/0.03	0.003/0.0003

Tabelle 3.6: Signifikanz des partiellen Lokus-Beitrags getrennt für 272 Männer (schwarz) und 348 Frauen (rot) aus einem Likelihood-Ratio-Test des vollen Modells versus eines Modells mit ausgeschlossenen Gen-Lokus. Das volle Modell enthält multiple Varianten eines Gens als Regressoren, SNPs (A) oder Haplotypen (B). Im reduzierten Modell wurden die Varianten des getesteten Lokus ausgeschlossen. p-Werte sind angegeben, falls  $p < 0.05$ .

dingt durch den Einfluss von Freiheitsgraden, epigenetische Effekte und umweltbedingte Effekte, die sich auf die Schätzung der Varianzkomponenten auswirken und die Modellgüte beeinflussen.

Abbildungen (3.6) und (3.7) zeigen den partiellen Anteil der Gen-Loci an genetisch bedingter LDL- bzw. HDL-Varianz im SNP- versus Haplotyp-basierten Modell für die Gesamtstichprobe. Die gestrichelte Linie markiert den geschätzten genetische bestimmten Varianzanteil.

Für LDL haben die Gen-Loci *APOE*, *CETP* (SNP- und Haplotyp-Modell) und *LDLR* (SNP-Modell) den größten prädiktiven Wert.

Die Gen-Loci *LDLR* (SNP-Modell) sowie *LIPC* und *CETP* (Haplotyp-Modell) haben den größten Einfluss auf HDL.

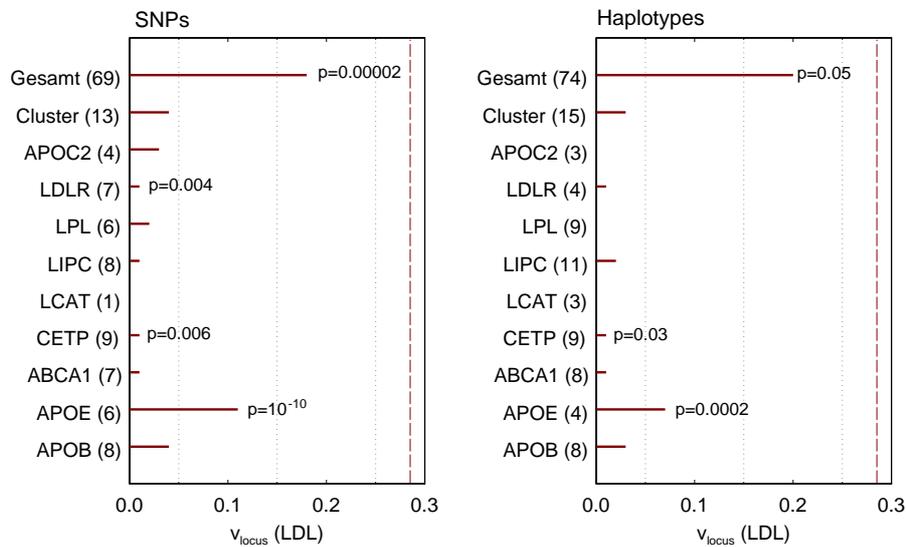


Abbildung 3.6: Partielle Anteile der Gen-Loci an genetisch bedingter LDL-Varianz (rot gestrichelte Linie: 28%) im SNP-/Haplotyp-Modell.

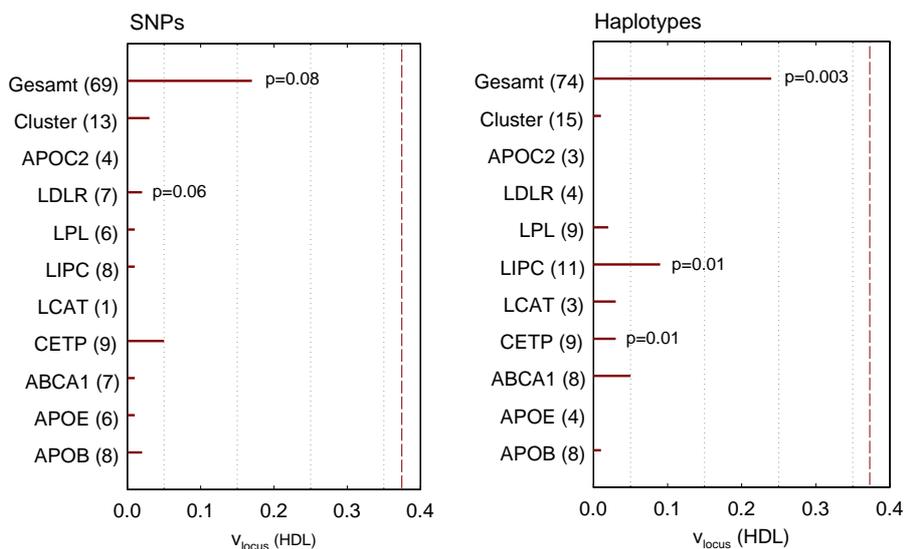


Abbildung 3.7: Partielle Anteile der Gen-Loci an genetisch bedingter HDL-Varianz (rot gestrichelte Linie: 37%) im SNP-/Haplotyp-Modell.

Gen-Lokus	Genf		Berlin	
	Anzahl SNPs	Mittlerer SNP-Abstand (kb)	Anzahl SNPs	Mittlerer SNP-Abstand (kb)
	$q$	$\Delta$	$q$	$\Delta$
<i>APOE</i>	7	0.6	6	0.7
<i>ABCA1</i>	88	1.6	14	11.3
<i>CETP</i>	27	0.8	13	1.9
<i>LCAT</i>	4	1.2	3	1.7
<i>LIPC</i>	37	3.8	11	16.5
<i>LPL</i>	24	1.2	12	2.4
<i>LDLR</i>	19	2.4	7	5.3
<i>APOA1</i>	4	0.6	6	0.5
<i>PLTP</i>	17	0.8	1	
<i>SRB1</i>	17	5.3	1	
Total	244		74	

Tabelle 3.7: Genomische Distanz untersuchter SNPs in Kandidatengenen des Lipoproteinstoffwechsels in Genf und Berlin.  $q$  bezeichnet die Anzahl untersuchter SNPs und  $\Delta$  bezeichnet den mittleren paarweisen Abstand zwischen den SNPs eines Gen-Lokus.

### 3.4 Vergleich genetischer Struktur und Assoziation in zwei europäischen Populationen

In Berlin wurden Großfamilien, die bis zu drei Generationen umfassen, rekrutiert. Dies resultiert in einer sehr viel größeren Spannweite des Alters im Vergleich zur Genfer Stichprobe. Die Cholesterin-Konzentrationen der Individuen mittleren Alters ( $> 34$  Jahre) beider Stichproben sind vergleichbar, der Wilcoxon Rangsummentest zeigte keine signifikanten Unterschiede. Die Werte liegen in klinisch unauffälligem Bereich.

Tabelle 3.7 zeigt die in der gemeinsamen Analyse untersuchten Kandidatengene und genetischen Marker beider Stichproben. Der SNP-Abstand beträgt im Mittel 1-4 kb. Es gibt zwei Ausnahmen in Berlin mit einem paarweisen SNP-Abstand von 14 kb. In Genf befinden sich 163 aus 244 SNPs in Introns und 81 in Exons (davon kodieren 31 nicht-synonym). In Berlin liegen 25 aus 74 SNPs in Introns, 10 im Promotor, 16 in nicht abgelesenen Regionen und 23 in Exons (12 sind nicht-synonym kodierende SNPs).

Gen-Lokus	SNP-Name		Fälle (%)		Kontrollen (%)	
	Berlin	Genf	Berlin	Genf	Berlin	Genf
<i>APOE</i>	rs440446	1b.+69	37.5	37.4	36.3	31.7
	rs429358	4b.152	21.8	13.7*	8.3	7.8
	rs7412	4b.290	4.4	4.8	14.1	15.4
	$\epsilon_4 =$ SNP 4b.152 (hetero- od. homozygot)		21.8	13.7*	7.2	7.8
	$\epsilon_3 =$ „wild Typ“ SNP 4b.152 / 4b.290		74.4	81.5	77.5	76.8

Fortsetzung auf der nächsten Seite

Fortsetzung Tabelle 3.8 von letzter Seite

Gen-Lokus	SNP-Name		Fälle (%)		Kontrollen (%)	
	Berlin	Genf	Berlin	Genf	Berlin	Genf
	$\epsilon_2 = \text{SNP 4b.290 (hetero- od. homozygot)}$		3.8	4.8	13.0	15.4
<i>ABCA1</i>	rs2422493	1b.-588	44.4	47.6	45.2	48.1
	rs2274873	9b.123	9.4	12.6	12.5	8.7
	rs4149313	18b.107	15.1	14.3	10.4	13.8
<i>CETP</i>	rs1800775	629	53.1	55.7	45.2	42.3
	rs7205804	6b.-56	40.7	37.0	47.3	50.0
	rs1532625	7b.+8	40.7	35.2	47.9	49.2
	rs5882	14b.16	24.7	33.5	33.8	32.9
	rs1801706	16b.159	13.8	17.3	21.2	17.2
<i>LCAT</i>	rs4986970	5b.171	7.5	1.9**	2.1	2.2
	rs5923	6b.429	5.9	6.4	3.5	8.1
<i>LIPC</i>	rs1800588	1E.-544	25.3	22.8	29.6	29.1
	rs2070895	1b.-280	25.0	22.6	31.5	28.7
	unknown	3b.-279	47.5	44.1	52.1	43.5
	rs6078	3b.10	3.3	3.8	2.7	3.9
	rs690	4b.9	42.9	39.5	37.8	39.8
	rs6082	5b.17	6.4	3.5	4.7	7.8
	rs6083	5b.70	28.6	30.2	36.3	32.1
	rs6084	5b.98	40.3	46.7	48.6	49.7
	rs6074	9b.49	13.1	12.0	13.0	13.0
<i>LPL</i>	rs269	6b.+73	17.9	17.2	19.9	26.1
	rs270	6b.+82	21.6	14.1*	17.1	12.1
	rs320	8b.+481 (HIND III)	25.3	27.1	29.5	36.9
	rs328	9b.99 (S447X)	9.6	9.2	12.8	16.9
<i>LDLR</i>	rs2228671	LDLR2	8.0	10.6	13.0	14.5
	rs5930	rs5930	44.7	39.7	45.9	37.0
	rs5925	rs5925	44.4	44.1	34.9	45.6
	rs5927	rs5927	28.3	24.2	19.7	25.5
	rs1433099	rs1433099	29.0	25.7	23.3	27.6
<i>APOA1</i>	rs5070	1b.41	34.6	32.6	31.9	30.5
<i>PLTP</i>	rs2294213	1b.+26	13.0	9.6	6.8	13.2
<i>SRB1</i>	rs7205804	8b.41 (A350A)	46.3	48.6	49.3	48.6

Tabelle 3.8: Allelfrequenzen gemeinsamer SNPs in Berlin (N=157) und Genf (N=371), gemessen an Fällen (hohes LDL/niedriges HDL) und Kontrollen (niedriges LDL/hohes HDL). Die deutsche Nomenklatur entspricht den Angaben in der NCBI-Referenz. Die Schweizer Bezeichnungen sind wie bei (Morabia et al., 2003b) definiert (s. Appendix), d.h. die Positionen der Nukleotide werden mit der Anzahl von Basenpaaren links (-) oder rechts (+) relativ zum Start des bezeichneten Exons (b) angegeben. Für einen Vergleich der Allelfrequenzen diploider Genotypen an einzelnen SNP-Positionen zwischen Berlin und Genf wurde ein  $\chi^2$ -Test (bzw. Fishers exakter Test) für Fälle und Kontrollen separat durchgeführt; Signifikanz ist angegeben (\* :  $p < 0.05$ ; \*\* :  $p < 0.01$ ; \*\*\* :  $p < 0.001$ ).

Aus der Gesamtzahl der SNPs wurden 34 SNPs in 10 Genen in jedem Individuum beider Stichproben analysiert. Die Allelfrequenzen der gemeinsamen SNPs beider Studien sind in Tabelle 3.8 dargestellt. Nur 4 von 68 Vergleichen, getrennt nach Geschlecht und Fall-/Kontrollstatus, zeigten einen signifikanten, wenngleich geringen Frequenzunterschied. Wie auch bezüglich LDL- und HDL-Konzentrationen deutet die Übereinstimmung auf eine homogene Struktur beider Stichproben.

Das Kopplungsungleichgewicht wurde durch die Analyse der Korrelationskoeffizienten der SNP-Genotypen für jedes Gen in beiden Stichproben verglichen (Tabelle 3.9).

Gen-Lokus	Berlin/ Genf	N	$Det\{R\}$	Nb. SNPs	Mittlerer SNP- Abstand (kb)	$-\ln(Det\{R\})/q$
CETP	B	145	0.004	5	5.6	1.088 +
	G	338	0.01			0.843 +
LIPC	B	134	0.001	9	17.2	0.729 +
	G	319	0.012			0.493 +
LDLR	B	139	0.11	5	8.0	0.445 +
	G	358	0.13			0.402 +
LPL	B	149	0.37	4	2.0	0.248
	G	348	0.47			0.188
APOE	B	145	0.82	3	1.5	0.066
	G	352	0.84			0.059
ABCA1	B	147	0.92	3	52.1	0.028
	G	361	0.97			0.011
LCAT	B	144	1.00	2	2.4	0.00003
	G	354	0.99			0.001

Tabelle 3.9: Linkage disequilibrium (LD) gemeinsamer SNPs in Lipid-Gen-Loci. Multiples LD wurde als Determinante (Det) der Korrelationsmatrix  $\{R\}$  zwischen SNPs in Berlin und Genf (gemessen an Fällen und Kontrollen) bestimmt. Mit Ausnahme von *LCAT* (Det=1.0) zeigen alle Gen-Loci eine signifikante LD-Struktur ( $p < 0.001$ ). Diese wird als stark (+) bezeichnet, falls  $(-\ln(Det\{R\})/q) > 0.4$ . Ein Vergleich der Korrelationsmatrizen auf der Basis des Determinantenkriteriums (siehe Methoden) zeigt keinen Unterschied zwischen beiden Stichproben.

Bis auf einen Locus (*LCAT*) kann mit dem Permutationstest signifikantes LD nachgewiesen werden. Einige der Gen-Loci zeigen starkes LD, andere mittleres oder sogar schwaches LD trotz geringem genomischen Abstand der SNPs voneinander. Dieser Befund reflektiert die bekannte stochastische Variation von LD auf dem Genom (Gabriel et al., 2002; Nothnagel & Rohde, 2005). *CETP*, *LIPC* und *LDLR* zeigen starkes LD ( $-\log(Det(R))/q > 0.4$ ). Im Gegensatz dazu zeigen SNPs in den Genen *APOE*, *ABCA1* und *LCAT* nur schwachen oder keinen Zusammenhang ( $< 0.1$ ). *LPL* zeigt einen mittleren LD-Wert. Die Korrelationsmatrizen in Genf und Berlin haben eine sehr ähnliche Struktur, unterstützt durch die Ergebnisse statistisch geprüfter Homogenität.

In Tabelle 3.10 sind Odds Ratios dargestellt, soweit sie in mindestens einer der beiden Stichproben statistisch signifikant waren. *APOE- $\epsilon_2$*  zeigt ein starkes „atheroprotektives“ Signal (kleines OR), während *APOE- $\epsilon_4$*  ein „atherogenes“ Signal liefert (hohes OR). Das OR des häufigen *APOE- $\epsilon_3$*  Haplotypen unterscheidet sich nicht signifikant von 1. Die *APOE*-Effekte sind in Bezug auf Signifikanz und Einflussrichtung in beiden Stichproben homogen. Der Vergleich individueller SNP-Effekte zwischen Genf und Berlin (Zweistichproben-t-Test) zeigt in den meisten Fällen Übereinstimmung. Unterschiede gibt es bei *PLTP* mit einem sehr starken gegenläufigen Effekt für die Frauen und mit gemäßigten Effekten in vier weiteren Fällen.

Tabelle 3.11 stellt die Bedeutung der Locus-Beiträge auf der Basis multipler logistischer Regression dar. Trotz eines nicht unerheblichen Anteils nicht-genetischer Hintergrundvarianz stellt sich heraus, dass fünf Gen-Loci (*ABCA1*, *APOE*, *CETP*, *LDLR* und *PLTP*) in beiden Fall-/Kontroll-Gruppen signifikant assoziiert sind. Die Übereinstimmung ist geringer im Fall von *LDLR* und *PLTP*. Dies bestätigt sich auch in kleineren Werten für Somers'  $D_{xy}$  und  $R^2$ . *LIPC* ist in Genf signifikant und hat in der kleineren Berliner Stichprobe einen marginalen Effekt. *SRB1* ist in beiden Gruppen nicht assoziiert. Einen gegensätzlichen Befund zeigt *LPL*: hoch-signifikant in Genf und nicht assoziiert in Berlin. *LCAT* und *APOA1* zeigen in einer der beiden Stichproben signifikante Assoziation.

Gen-Lokus	SNP-Name		Adjustiertes SNP OR (95% KI)		Interaktion: p-Wert ( $OR_w/OR_m$ )	
	Berlin	Genf	Berlin	Genf	Berlin	Genf
<i>APOE</i>	rs429358 <sup>1</sup>	4b.152	3.47** (1.54-7.82)	2.08** (1.17-3.71)		
	rs7412 <sup>1</sup>	4b.290	0.27* (0.10-0.75)	0.24*** (0.12-0.47)		
<i>ABCA1</i>	rs4149313	18b.107			<b>0.047</b> (0.80/4.85**)	(1.36/0.98)
<i>CETP</i>	rs1800775	1b.-629	1.24 (0.59-2.58)	2.01* (1.18-3.45)		
	rs7205804	6b.-56	0.57 (0.28-1.19)	0.45** (0.27-0.75)		
	rs1532625	7b.+8	0.56 (0.27-1.18)	0.45** (0.27-0.74)		
<i>LCAT</i>	rs4986970	5b.171	3.78* (1.01-14.17)	1.07 (0.33-3.41)		
	rs5923	6b.429			(1.98 <sup>+</sup> /1.49)	<b>0.04</b> (0.42/1.71)
<i>LIPC</i>	rs2070895	1b.-280	0.71 (0.38-1.35)	0.60* (0.37-0.97)		
	unknown	3b.-279			(0.65/0.82)	<b>0.03</b> (1.78/0.62)
	rs6078	3b.10			(0.60/2.45)	<b>0.03</b> (0.49/3.59)
	rs6082	5b.17			<b>0.03</b> (7.04 <sup>+</sup> /0.34)	(0.42/0.56)
<i>LPL</i>	rs269	6b.+73	0.92 (0.47-1.81)	0.59* (0.36-0.94)		
	rs320	8b.+481 (HIND III)			<b>0.07</b> (1.21/0.33*)	(0.65/0.65)
	rs328	9b.99 (S447X)	0.65 (0.29-1.47)	0.47** (0.27-0.82)		
<i>LDLR</i>	rs5925	rs5925	2.12* <sup>+</sup> (1.05-4.25)	0.83 (0.50-1.37)		
	rs5927	rs5927			<b>0.05</b> (0.92/3.30*)	(0.84/1.12)
	rs1433099	rs1433099			<b>0.03</b> (0.68/2.53)	(0.87/1.17)
<i>PLTP</i>	rs2294213	1b.+26			<b>0.01</b> (5.37*** <sup>+</sup> /0.44)	<b>0.07</b> (0.22**/0.84)
<i>SRB1</i>	rs7205804	8b.41 (A350A)			(0.93/1.68 <sup>+</sup> )	<b>0.02</b> (1.43/0.37*)

<sup>1</sup> rs7412 (4b.290) entspricht APOε<sub>2</sub>, rs429358 (4b.152) entspricht APOε<sub>4</sub>

Tabelle 3.10: SNP Assoziation mit dem binären Phänotyp. Die Spalten 1-2 enthalten Deutsche und Schweizer SNP-Nomenklatur. Die Spalten 3-4 zeigen Odds Ratios (OR) aus logistischer Regression, adjustiert für Alter und BMI, und das dazugehörige 95%-Konfidenzintervall (in Klammern). Die Spalten 5-6 zeigen Odds Ratios aus geschlechtsspezifischen Modellen ( $OR_w / OR_m$  in Klammern) falls in der logistischen Regression ein signifikanter SNP-Geschlechts-Interaktionseffekt nachgewiesen werden konnte. Sterne (\* :  $p < 0.05$ ; \*\* :  $p < 0.01$ ; \*\*\* :  $p < 0.001$ ) beziehen sich auf die Ablehnung der Nullhypothese  $OR=1$  in einem Wald- $\chi^2$ -Test. Kreuze (<sup>+</sup> :  $p < 0.05$ ; <sup>++</sup> :  $p < 0.01$ ; <sup>+++</sup> :  $p < 0.001$ ) beziehen sich auf einen t-Test der Nullhypothese  $OR_{Berlin} = OR_{Genf}$  in unabhängigen Stichproben.

Gen-Lokus	Berlin (Fälle/Kontrollen: 82/75)					Genf (Fälle/Kontrollen: 186/185)				
	$R^2$ (%)	Df	p (LR)	Klassif. Rate	$D_{xy}$	$R^2$ (%)	Df	p (LR)	Klassif. Rate	$D_{xy}$
<i>APOE</i>	18.0	3	0.00003	62.9***	0.35	10.0	3	0.000002	60.5***	0.29
<i>ABCA1</i>	12.9	7	0.0013	66.0***	0.35	12.9	33	0.00019	58.0**	0.33
<i>CETP</i>	11.9	6	0.0028	62.5***	0.30	5.1	6	0.0027	52.9	0.20
<i>LDLR</i>	8.5	4	0.012	56.8*	0.24	2.8	4	0.03	57.1**	0.14
<i>PLTP</i>	7.3	3	0.012	61.0**	0.23	2.7	4	0.04	44.6	0.11
<i>LCAT</i>	5.0	1	0.03	38.6	0.09	1.8	3	0.08	55.4*	0.10
<i>LIPC</i>	8.6	12	0.06	56.0	0.22	4.3	9	0.024	56.8**	0.18
<i>LPL</i>	3.3	5	0.22	51.3	0.12	6.4	8	0.0013	57.9**	0.21
<i>APOA1</i>	n.s.					2.2	1	0.02	54.8*	0.09
<i>SRB1</i>	n.s.					1.4	3	0.12	55.4*	0.08

Tabelle 3.11: Qualitative (Fall-/Kontroll-) Assoziation mit SNPs und/oder SNP-Geschlecht-Interaktionstermen einzelner Loci (separat) als unabhängige Variablen. SNPs werden durch stepweise logistische Regression in Genf und Berlin unabhängig selektiert. Die Signifikanz wird in einem Likelihood-Ratio-Test lokusweise bestimmt.  $R^2$  bezeichnet das verallgemeinerte Bestimmtheitsmaß (Nagelkerke, 1991). Somers  $D_{xy}$  charakterisiert die Korrelation zwischen beobachtetem Fall-/Kontrollstatus und vorhergesagten Wahrscheinlichkeiten. Unverzerrte Schätzungen für  $R^2$  und  $D_{xy}$  wurden in einem Bootstrap-Verfahren bestimmt (Harrell, 1999). Die Rate korrekter Klassifikation wurde durch 10-fache-Kreuzvalidierung (Efron & Tibshirani, 1998) geschätzt und gegen die Binomialverteilung getestet (\* :  $p < 0.05$ ; \*\* :  $p < 0.01$ ; \*\*\* :  $p < 0.001$ ). n.s.: kein SNP wurde in der Vorauswahl selektiert

### 3.5 Populationsgenetischer Vergleich mit HapMap-Daten

Von 100 untersuchten SNPs in Berlin sind 56 Positionen ebenfalls im Rahmen des HapMap-Projektes untersucht worden (Tabelle 3.2). Zum Vergleich wurden die Daten der CEPH-Population (Einwohner von Utah mit Vorfahren aus Nord- und Westeuropa) herangezogen (<http://www.hapmap.org>).

In Abbildung 3.8 sind die Allelfrequenzen der 56 gemeinsamen SNPs gegenübergestellt. Mit Ausnahme eines SNPs im *LPL*-Gen (rs690) stimmen die Frequenzen gut überein.

Die Werte paarweisen Linkage Disequilibriums zwischen SNPs wurden als quadrierte Korrelationskoeffizienten ( $r^2$ ) nach Pearson berechnet und im Verhältnis zur genomischen Distanz dargestellt.

Die Abbildung 3.9 zeigt paarweises LD in der Berliner Stichprobe und in der Stichprobe aus Utah (HapMap). Der ähnliche Abfall allelischen Zusammenhangs mit steigender genomischer Distanz legt eine große Übereinstimmung beider Stichproben aus populationshistorischer Sicht nahe.

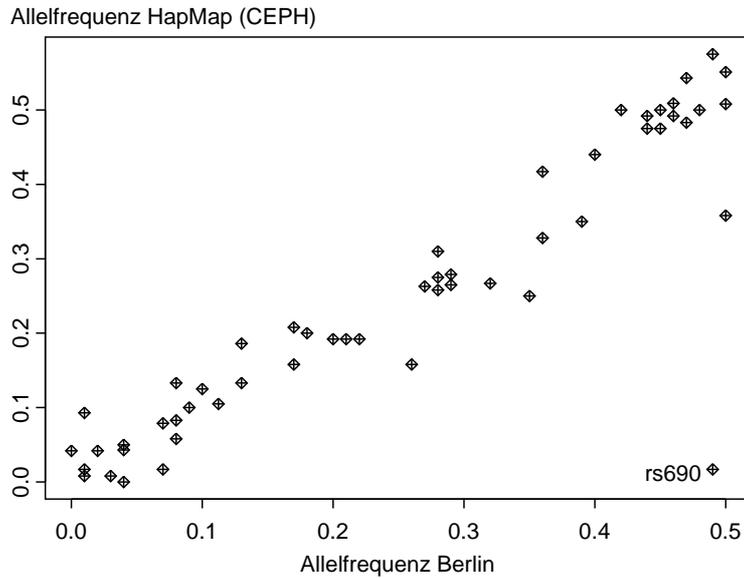


Abbildung 3.8: Vergleich der Allelfrequenzen häufiger SNPs der Berliner Stichprobe und der HapMap an 56 identischen SNP-Positionen. Der Korrelationskoeffizient nach Bravis-Pearson ergibt  $r = 0.91$ . Einen Ausreißer stellt der SNP rs690 dar mit einer Frequenz von 0.49 in Berlin und 0.02 in den Daten der HapMap.

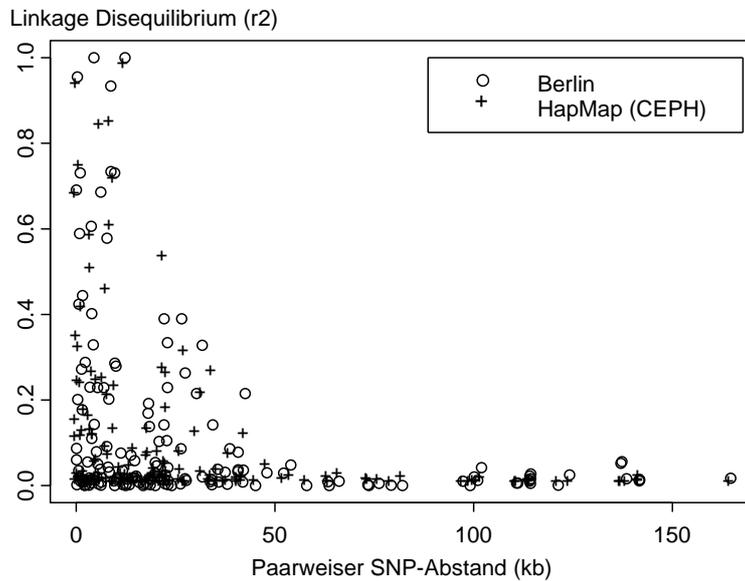


Abbildung 3.9: Linkage Disequilibrium von 159 SNP-Paaren in Kandidatengenomen des Lipoproteinstoffwechsels berechnet als quadrierter Korrelationskoeffizient ( $r^2$ ) in einer Vierfeldertafel. Dargestellt sind (vergleichsweise) die Werte der Berliner Stichprobe und der HapMap im Verhältnis zur genomischen Distanz (in kb) innerhalb der Gene.

