

UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Zentrum für Innere Medizin, III. Medizinische Klinik und Poliklinik

Direktor: Prof. Dr. med. Rolf A. K. Stahl

**Genetische Varianten im KLF14-Gen und deren Einfluss auf Dyslipidämien sowie
Adipositas und Diabetes mellitus Typ 2**

Dissertation

zur Erlangung des Grades eines Doktors der Medizin
an der Medizinischen Fakultät der Universität Hamburg

vorgelegt von:

Anna Franziska Schlegel
aus Dresden

Hamburg 2015

Angenommen von der

Medizinischen Fakultät der Universität Hamburg am: 05.02.2016

Veröffentlicht mit Genehmigung der

Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der Vorsitzende: Prof. Dr. med. F.U. Beil

Prüfungsausschuss, zweiter Gutachter: Prof. Dr. med. E. Windler

Inhaltsverzeichnis

1 Einleitung	1
1.1 Lipide im Blut	1
1.1.1 Klassifikation der Lipoproteine	1
1.1.2 Stoffwechsel der Lipoproteine	2
1.1.3 Einflussfaktoren auf Lipid- und Lipoproteinprofile im Blut	4
1.2 Genetik der Dyslipidämien	6
1.2.1 Monogenetische Erkrankungen des Lipidmetabolismus	6
1.2.2 Dyslipidämien als komplexe Erkrankungen	8
1.2.3 Genomweite Assoziationsstudien	10
1.2.4 Resequenzierungsstudien	11
1.3 Der Transkriptionsfaktor Krüppel-like Faktor 14	13
1.3.1 Die genetische Variante S58P im KLF14-Gen	16
1.3.2 Die genetische Variante P47Q im KLF14-Gen	17
1.3.3 Die genetische Variante A39A im KLF14-Gen	17
1.4 Fragestellung	18
2 Material und Methoden	19
2.1 Chemikalien	19
2.2 Geräte und Arbeitsmittel	19
2.3 Software	19
2.4 Methoden	20
2.4.1 Polymerasekettenreaktion (PCR)	20
2.4.2 Restriktionsendonukleasenverdau	21
2.4.3 Kapillargelelektrophorese	21
2.4.4 Sequenzierung	21
2.4.4.1 Aufreinigung des PCR-Produkts	21
2.4.4.2 Sequenzierung	21
2.5 Datenverwaltung und statistische Auswertung	23
2.6 Berechnung eines Hardy-Weinberg-Gleichgewichts	23
2.7 Kopplungsungleichgewicht	25
3 Ergebnisse	27
3.1 Patientenkollektiv	27
3.2 Genotypisierung des SNP S58P	27

3.3 Hardy-Weinberg-Gleichgewicht.....	29
3.4 Die genetische Variante S58P des KLF14-Gens und Dyslipidämien	29
3.5 Zusammenhang zwischen dem SNP S58P und den Serumlipidwerten.....	32
3.6 Zusammenhang zwischen der genetischen Variante S58P und den Parametern Adipositas und Diabetes mellitus Typ 2 als Komponenten des metabolischen Syndroms	33
3.7 Kopplungsungleichgewicht für die SNPs S58P, P47Q und A39A.....	38
3.8 Auffälliges Bandenmuster in der Genotypisierung	43
4 Diskussion	46
4.1 Einordnung der Ergebnisse in die aktuelle Literatur	46
4.2 Bewertung des Studienkollektivs	49
4.3 Möglichkeiten zur Bewertung der Relevanz der Variante S58P und der identifizierten Deletion.....	50
4.4 Ausblick und neue Ansätze der Forschung	51
4.5 Bewertung der CDCV- und CDRV-Hypothese.....	53
4.6 Bedeutung und Nutzen genetischer Untersuchungen	55
5 Zusammenfassung.....	57
6 Abkürzungsverzeichnis.....	58
7 Tabellenverzeichnis	60
8 Abbildungsverzeichnis	60
9 Formelverzeichnis	61
10 Literaturverzeichnis	62
11 Danksagung	70
12 Lebenslauf.....	71
13 Eidesstattliche Versicherung.....	71
Anhang 1.....	73
A 1.1 Polymerasekettenreaktion	73
A 1.2 Restriktionsendonukleasenverdau	73
A 1.3 Kapillargelelektrophorese	74
A 1.4 Aufreinigung des PCR-Produktes.....	75
A 1.5 Sequenzierung mit der Kettenabbruchmethode nach Sanger	75

Anhang 2.....	77
A 2.1 Verwendete Begriffe der deskriptiven Statistik.....	77
A 2.1.1 Arithmetisches Mittel $\bar{x}$	77
A 2.1.2 Median $\tilde{x}$	77
A 2.1.3 Varianz und Standardabweichung.....	78
A 2.2 Verwendete Testverfahren der induktiven Statistik.....	78
A 2.2.1 Der Chi ² -Vierfeldertest als Grundlage	78
A 2.2.2 Fishers exakter Test.....	79
A.2.2.3 Mann-Whitney-U-Test	80

1 Einleitung

1.1 Lipide im Blut

Lipide sind hydrophobe Stoffe, welche für den Transport im wässrigen Milieu des Plasmas Transportproteine benötigen. An diese binden sie entweder als Einzelmolekül oder bilden mit diesen Makromoleküle, sogenannte Lipoproteine. Die Lipoproteine bestehen aus einem hydrophoben Kern, in welchem sich Triglyzeride und Cholesterinester befinden. Ummantelt wird der hydrophobe Kernbereich von einer hydrophilen Oberfläche aus Phospholipiden, freiem Cholesterin und Proteinen, welche als Apolipoproteine bezeichnet werden. Die Lipoprotein-Partikel werden in der Leber und im Darm synthetisiert und dienen dem Transport der Lipide vom Darm zur Leber und von dieser weiter zu den Zellen (Lund-Katz u. Phillips, 2010).

1.1.1 Klassifikation der Lipoproteine

Die Lipoproteine können nach verschiedenen Kriterien eingeteilt werden. Hierfür wird zum Beispiel die Dichtebestimmung mittels Auftrennung in einer präparativen Ultrazentrifuge genutzt. Dabei unterscheidet man vier nach ihrer Dichte unterschiedliche Lipoproteinfraktionen. Die Hauptklassen sind dabei Chylomikronen, VLDL (very low density lipoprotein, Lipoprotein sehr geringer Dichte), LDL (low density lipoprotein, Lipoprotein geringer Dichte) und HDL (high density lipoprotein, Lipoprotein hoher Dichte). Als weitere Möglichkeiten zur Klassifizierung der Lipoproteinfraktionen dienen die Bestimmung des chemischen Lipid-Protein-Verhältnisses, die Bestimmung der vorhandenen Apolipoproteine oder die Nutzung ihrer unterschiedlichen Mobilität in der Elektrophorese (Heinrich et al., 2007). Die Tabelle 1 gibt einen Überblick über die Lipoproteine und deren Eigenschaften.

Tabelle 1: Eigenschaften der Lipoproteine. Modifiziert nach Heinrich et al., 2007.

Lipoprotein	Chylomikronen	VLDL	LDL	HDL
Dichte (g/ml)	< 0,93	0,93 – 1,006	1,019 - 1,063	1,063 – 1,21
Durchmesser (nm)	75 - 1200	30 - 80	18 - 25	5 - 12
Verhalten bei der Elektrophorese	keine Mobilität	prä- β -Fraktion	β -Fraktion	α -Fraktion
Chemisches Lipid-Protein-Verhältnis:				
Triglyzeride (%)	86	55	6	4
Cholesterinester (%)	5	19	50	19
Phospholipide (%)	7	18	22	34
Protein (%)	2	8	22	42
Apolipoproteine	ApoB48, ApoC2 und C3, ApoE, ApoA1 und A4	ApoB100, ApoC1-C3, ApoE, ApoA5	ApoB100, ApoE	ApoA1, A2, A4 und A5, ApoE, ApoC1-C3

1.1.2 Stoffwechsel der Lipoproteine

Die mit der Nahrung zugeführten Triglyzeride werden nach Abspaltung von Mono- und Diglyceriden über Fettsäure-Transporter (fatty acid transporter, FA) in die Mukosazellen des proximalen Darmes aufgenommen. Intrazellulär erfolgt die Synthese der Chylomikronen aus Triglyzeriden, unverestertem Cholesterin, Cholesterinestern und den Apolipoproteinen A1 (ApoA1) und A4 (ApoA4) sowie B48 (ApoB48). Diese werden mittels Exocytose in die Lymphe sezerniert. Über den Ductus thoracicus gelangen die Chylomikronen in den Kreislauf. Im Blut nehmen sie von den HDL-Partikeln Apolipoprotein C2 (ApoC2) auf. Dieses ist besonders wichtig, da es ein Kofaktor für die lipolytische Lipoproteinlipase (LPL) ist. Die Lipoproteinlipase ist ein endothelständiges Enzym und katalysiert die Spaltung der Triglyzeride in freie Fettsäuren und Glycerin. Die Fettsäuren können dadurch von den peripheren Zellen aufgenommen und zur Energiegewinnung genutzt werden. Zurück bleiben sogenannte Remnants der Chylomikronen. Während der Hydrolyse der Triglyzeride werden ApoA1, ApoA4 und ApoC2 auf die HDL-Partikel übertragen und die Remnants nehmen Apolipoprotein E (ApoE) von den discoidalen HDL-Vorstufen auf. Die ApoE-beladenen Remnants werden von der Leber aufgenommen und abgebaut. Dazu bindet ApoE an den LDL-related-Rezeptor (LRP) unter Vermittlung der LPL. Über den LRP werden neben den Remnants auch andere ApoE-reiche Lipoproteine aufgenommen.

Die in der Leber gebildeten triglyzeridreichen VLDL transportieren die endogen synthetisierten Lipide auf ähnliche Weise zu den extrahepatischen Geweben. Nach Sezernierung der Apolipoprotein B100 (ApoB100)-reichen Lipoproteine in das Blut werden ihnen aus der HDL-Fraktion die Faktoren ApoC2 und ApoE übertragen. Die LPL wird durch ihren Kofaktor ApoC2 wieder aktiviert und spaltet aus den Triglyzeriden freie Fettsäuren ab. Aus den VLDL entstehen dabei IDL (intermediate density lipoprotein, Lipoprotein mittlerer Dichte). Die IDL werden zur Leber transportiert und über den LDL-Rezeptor aufgenommen. Der Rest der VLDL wird in LDL umgewandelt.

Das ApoB100 der LDL dient als Ligand mit dem LDL-Rezeptor. Nach Bindung von ApoB100 an den LDL-Rezeptor wird LDL in die Zelle aufgenommen. In nicht-hepatischen Zellen dient das Cholesterin als Baustein von Zellmembranen und zur Herstellung von Steroidhormonen, in der Leber werden aus Cholesterin die Gallensäuren hergestellt. Mehr als die Hälfte des zirkulierenden LDL wird über die Leber verstoffwechselt, welche den größten Anteil der LDL-Rezeptoren exprimiert. Ein geringerer Teil wird über periphere Rezeptoren aufgenommen. Ein vom LDL-Rezeptor unabhängiger Abbau des LDL erfolgt über den Scavenger-Pfad. Die Scavenger-Rezeptoren (SR) sind membranständig und werden von verschiedensten Zellen und Geweben wie Makrophagen, glatten Muskelzellen und Epithelzellen exprimiert. Liganden für diesen Rezeptor sind modifizierte LDL. Im Gegensatz zu den LDL-Rezeptoren, welche bei einem LDL-Cholesterinspiegel >200 mg/dl gesättigt sind, werden SR des Typs A (SR-A) in Abhängigkeit vom Vorhandensein von Liganden vermehrt exprimiert. Sie sind somit bei hohen LDL-Cholesterinwerten von Bedeutung.

Eine wichtige Funktion der HDL ist der reverse Cholesterintransport. Über HDL wird Cholesterin aus der Peripherie zur Leber transportiert und über die Gallensäuren ausgeschieden. Dadurch kommt der HDL-Fraktion eine atheroprotektive Funktion zu. In der Leber und der Darmmukosa werden ApoA1 und ApoA2 sezerniert und zusätzlich bei der Hydrolyse triglyzeridreicher Lipoproteine (Chylomikronen und VLDL) freigesetzt. Die ApoA-Moleküle nehmen freies Cholesterin und Phospholipide auf, welche über die aktiven, ATP-bindenden Kassettentransporter A1 und G1 (ATP-binding cassette A1 und G1, ABCA1 und ABCG1) aus der Zelle geschleust werden. Dabei entstehen die prä-HDL-Partikel als Vorstufen reifer HDL-Moleküle. Das Enzym Lecithin-Cholesterin-Acyl-Transferase (LCAT) vermittelt die Veresterung des Cholesterins, welches dann in den Kern des HDL-Moleküls wandert. Die mit Cholesterin beladenen HDL-Partikel werden über den Scavenger-Rezeptor B1 (SR-B1) in die Leber aufge-

nommen. Eine weitere Möglichkeit, HDL-Cholesterin zur Leber zu transportieren, wird über das Cholesterinester-Transferprotein (CETP) reguliert. Dieses überträgt im Austausch gegen Triglyzeride die Cholesterinester der HDL-Partikel auf VLDL und LDL. Die ApoB-haltigen Lipoproteine werden dann mittels des LDL-Rezeptors der Leber aufgenommen und Cholesterin eliminiert (Heinrich et al., 2007; Hegele, 2009; Havel, 1995; Beisiegel et al., 1989; Schwandt, 2007; Verdier et al., 2013; Stephen et al., 2010; Moore u. Freeman, 2006; Brown et al., 1979). Einen Überblick über die wichtigsten Rezeptoren und Enzyme des Lipoproteinstoffwechsel zeigt Tabelle 2

Tabelle 2: Wichtige Rezeptoren und Enzyme des Lipoproteinmetabolismus und deren Funktion.

Fettsäure-Transporter (fatty acid transporter, FA)	Aufnahme der intestinalen Triglyzeride in die Mukosazellen des proximalen Darms
LDL-Rezeptor	Aufnahme von ApoB- und ApoE-haltigen Lipoproteinen (ubiquitär)
LDL-related Rezeptor (LRP)	Aufnahme der Remnants und anderer ApoE-haltiger Lipoproteine (leberspezifisch)
Scavenger-Rezeptor Typ A (SR-A)	Aufnahme von LDL-Cholesterin (Makrophagen, glatte Muskelzellen)
Scavenger Rezeptor Typ B1 (SR-B1)	Aufnahme von HDL-Cholesterin in die Leber
ATP-bindendes Kassettenprotein A1 und G1 (ATP-binding cassette A1 und G1, ABCA1 und ABCG1)	Aktive Cholesterintransporter aus der Zelle ins Blut
Lipoproteinlipase (LPL)	Enzym zur Spaltung von Triglyzeriden in freie Fettsäuren und Glycerin, Kofaktor ist ApoC2
Lecithin-Cholesterin-Acyl-Transferase (LCAT)	Bildung von Cholesterinestern aus überschüssigem Cholesterin und Lecithin
Cholesterinester-Transportprotein (CETP)	Austausch von Cholesterinestern des HDL-Cholesterins mit Triglyzeriden ApoB-haltiger Lipoproteine

1.1.3 Einflussfaktoren auf Lipid- und Lipoproteinprofile im Blut

Die einzelnen Lipid- und Lipoproteinfraktionen beeinflussen sich gegenseitig. So besteht zwischen Triglyzeriden und HDL-Cholesterin eine inverse Korrelation. Erhöhte Triglyzeridspiegel entstehen meist durch einen Überschuss an Chylomikronen-Remnants und VLDL. Bei erhöhten Triglyzeridwerten sinken die HDL-Cholesterinwerte. Wiederum ist HDL-Cholesterin positiv mit seinem

Hauptapolipoprotein A1 assoziiert. Daneben werden die Laborwerte auch von individuell beeinflussbaren sowie nicht-beeinflussbaren Umweltfaktoren bestimmt. Beeinflussbare Faktoren wie Rauchen, Bewegungsmangel und die Einnahme von Medikamenten können zu Veränderungen der Blutfettwerte führen. So steigt der Triglyzeridspiegel mit der Anzahl an gerauchten Zigaretten und führt ebenfalls zu erniedrigten HDL-Cholesterinwerten. Bei Frauen ist dieser Zusammenhang ausgeprägter als bei Männern. Bei den nicht-beeinflussbaren Faktoren ist neben dem Alter und dem Geschlecht vor allem die genetische Prädisposition zu erwähnen (Patsch et al., 1992; Hegele, 2009; Hegele, 2001; Brinton et al., 1994; Phillips et al., 1981). In der ILSE-Studie (Interdisziplinäre Langzeit-Studie des Erwachsenenalters über die Bedingungen zufriedenen und gesunden Alterns) wurden 239 Frauen und 262 Männer auf geschlechts- und altersabhängige Veränderungen hämatologischer Labormesswerte untersucht. Ein typischer Anstieg der Werte für Triglyzeride, Gesamtcholesterin und LDL-Cholesterin mit Höchstwerten in den mittleren Dezennien und ein Abfall der Messwerte ab dem 80. Lebensjahr wurde beobachtet. Eine altersabhängige Abnahme des HDL-Cholesterinspiegels zeigte sich bei Frauen. Generell weisen jüngere Frauen ein günstigeres Lipidprofil im Vergleich zu gleichaltrigen Männern auf, welches sich aber im höheren Alter dem der Männer angleicht (Martin et al., 1999). Jedoch leiden nicht alle Menschen, welche denselben nicht-genetischen Faktoren ausgesetzt sind oder einen ungesunden Lebensstil pflegen, unter erhöhten Lipidwerten oder erkranken an einer Fettstoffwechselstörung. Familien- und Zwillingsstudien zeigen, dass genetische Faktoren einen größeren Einfluss auf Lipidwerte haben als Umwelteinflüsse (Rao et al., 1982; Perusse et al., 1989). In einer großen longitudinalen Studie wurden die Anteile genetischer Effekte bzw. von Umwelteinflüssen auf den Lipidstatus von 965 Zwillingen und deren Geschwister zwischen 12 und 16 Jahren untersucht. Die Resultate der Studie zeigten keinen umweltbedingten Effekt auf die Lipidwerte, mit Ausnahme von HDL. Hinsichtlich des genetischen Einflusses wurden verschiedene Gene identifiziert, welche Gesamtcholesterin, LDL- und HDL-Cholesterin während der Adoleszenz mit regulieren. Des Weiteren lassen die Ergebnisse der Studie alters- und entwicklungsabhängige Einflüsse der einzelnen Gene vermuten. Somit bildet die genetische Prädisposition die Grundlage des individuellen Lipidprofils. Sie unterliegt jedoch Schwankungen durch verschiedene Gene zu unterschiedlichen Zeitpunkten des Lebens. (Middelberg et al., 2007). Die Gesamtspanne der genetischen Prädisposition wird zwischen 28% und 78% geschätzt (Heller et al., 1993). Die Erforschung genetischer Veränderungen ist deshalb von entscheidender Bedeutung.

1.2 Genetik der Dyslipidämien

1.2.1 Monogenetische Erkrankungen des Lipidmetabolismus

Dyslipidämien können entweder durch Veränderungen eines bestimmten Gens mit hoher Effektstärke oder durch die Summe vieler Varianten mit jeweils geringer Einflussstärke entstehen. Die erste Möglichkeit bezeichnet eine Gruppe von monogenetischen Dyslipidämien. Diese unterliegen typischen Vererbungsmustern, welche den Mendel'schen Vererbungsgesetzen folgen. Hierbei wird ein dominanter und ein rezessiver Erbgang unterschieden. Führt das Vorhandensein eines krankheitsverursachenden Allels zur Manifestation liegt ein dominanter Erbgang vor. Zur phänotypischen Ausprägung einer rezessiv vererbten Erkrankung müssen zwei krankheitsverursachende Allele vorliegen. Die ursächlichen Gene monogenetischer Dyslipidämien wurden zumeist in Studien mit Patienten im Extrembereich der Lipidwertverteilung identifiziert. Als Extrembereich werden Werte unterhalb der 5. Perzentile für Alter und Geschlecht und oberhalb der 95. Perzentile angenommen. Eine andere Methode zur Identifizierung sind Familienstudien, in welchen Dyslipidämien gehäuft vorkommen (Kuivenhoven u. Hegele, 2014). Eine Auswahl an monogenetischen Dyslipidämien ist in Tabelle 3 dargestellt. Der klinische Verdacht auf eine monogenetische Erkrankung kann bei extremen Abweichungen des Lipidprofils von den Referenzbereichen, einer familiären Belastung mit Dyslipidämien, einem früheren Manifestationsalter der Erkrankung, als zu erwarten wäre, eventuell frühzeitiger Arteriosklerose, typischen Begleiterscheinungen der Dyslipidämie-Syndrome und dem relativen Fehlen von sekundären Faktoren, welche mit Fettstoffwechselstörungen assoziiert sind, gestellt werden. Sekundäre Faktoren beinhalten zum Beispiel Adipositas, Rauchen, Medikamente und metabolische Begleiterkrankungen. Da es sich bei monogenetischen Erkrankungen um Gene mit einer hohen Penetranz handelt, kommt es auch ohne diese sekundären Faktoren zur phänotypischen Ausprägung (Hegele et al., 2015).

Tabelle 3: Monogenetische Dyslipidämien, geordnet nach betroffenem Lipoprotein. OMIM (Online Mendelian Inheritance in Man) ist ein frei zugänglicher Online-Katalog menschlicher Gene und genetischer Erkrankungen mit Fokussierung auf die molekulare Beziehung zwischen genetischer Veränderung und phänotypischer Ausprägung. Der Katalog wird regelmäßig aktualisiert (McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), Updated 27. April 2015).

Lipoprotein	Phänotyp	Gen	Erkrankung	OMIM-Nummer	Quelle
Triglyzeride	Erhöhung	LPL	Hyperlipoproteinämie	23860	(Brunzell, 2011)
		ApoA5	Familiäre Hypertriglyzeridämie	145750	(Calandro et al., 2006)
LDL-Cholesterin	Erhöhung	LDLR	Familiäre Hypercholesterinämie	143890	(Austin et al., 2004)
		ApoB	Familiärer ApoB-Defekt	144010	(Austin et al., 2004)
		PCSK9	Hypercholesterinämie	603776	(Abifadel et al., 2003)
	Erniedrigung	ApoB	Hypobetalipoproteinämie	615558	(Tarugi u. Aversa, 2011)
		PCSK9	Hypobetalipoproteinämie	607786	(Tarugi u. Aversa, 2011)
HDL-Cholesterin	Erhöhung	CETP	Hyperalpha-lipoproteinämie	118470	(Saito, 1984)
		LIPC	Hepatische Lipase Defekt	614025	(Hegele, 2001)
	Erniedrigung	ABCA1	Tangier-Erkrankung	205400	(Remaley et al., 1999)
		LCAT	Familiärer LCAT-Defekt (Fischaugen-Syndrom)	136120	(Rader, 2009)

Nichtsdestotrotz erklärt die Gruppe der monogenetischen Dyslipidämien nur einen geringen Teil der phänotypischen Varianz. Einige Patienten besitzen keine genetischen Varianten in den bereits bekannten Genen oder zeigen keine typischen Vererbungsmuster nach Mendel. Es wird deshalb vermutet, dass zum einen noch unbekannte Gene an

der Regulation des Lipidmetabolismus beteiligt sind und zum anderen multiple Gene in einem komplexen Zusammenspiel auf den Lipidhaushalt wirken (Hegele, 2009). Eine exakte Zahl der beteiligten Gene und deren Beziehung zueinander ist bisher noch nicht geklärt.

1.2.2 Dyslipidämien als komplexe Erkrankungen

Als komplexe Erkrankung wird jeder Phänotyp beschrieben, welcher keiner klassischen dominanten oder rezessiven Vererbung nach Mendel folgt. Die Komplexität wird durch die fehlende Verknüpfung von Genotyp und Phänotyp verursacht. Dabei kann sowohl ein Genotyp für mehrere Phänotypen verantwortlich sein als auch ein Phänotyp durch mehrere Genotypen entstehen (Lander u. Schork, 1994). Neben Dyslipidämien werden auch Erkrankungen, wie die koronare Herzkrankheit, Diabetes mellitus sowie das metabolische Syndrom, aber auch onkologische Erkrankungen zum Spektrum der komplexen Krankheiten gezählt. Mit welcher Häufigkeit multiple Gene oder genetische Varianten auftreten müssen, ist eine grundlegende Frage in der Betrachtung einer Krankheit als komplexes Geschehen. Die Common-Disease-Common-Variant (CDCV)-Hypothese postuliert, dass die genetische Veranlagung für eine komplexe Krankheit durch häufig vorkommende genetische Varianten, meist mit einer Frequenz in der Population zwischen 1-5 % angegeben, entsteht. Der Einfluss der Varianten ist jedoch gering. Durch Akkumulation mehrerer häufiger genetischer Varianten und deren additiver Effekte kommt es schließlich zur Manifestation der Erkrankung. Der Terminus genetische Varianten beinhaltet vornehmlich Einzelnukleotidpolymorphismen (single nucleotide polymorphism, SNP) sowie Insertionen, Deletionen, aber auch Kopienzahlvarianten (copy number variants, CNV). Bei SNPs handelt es sich um Bereiche in der DNA-Sequenz, welche sich um ein einzelnes Basenpaar zwischen zwei Individuen unterscheiden.

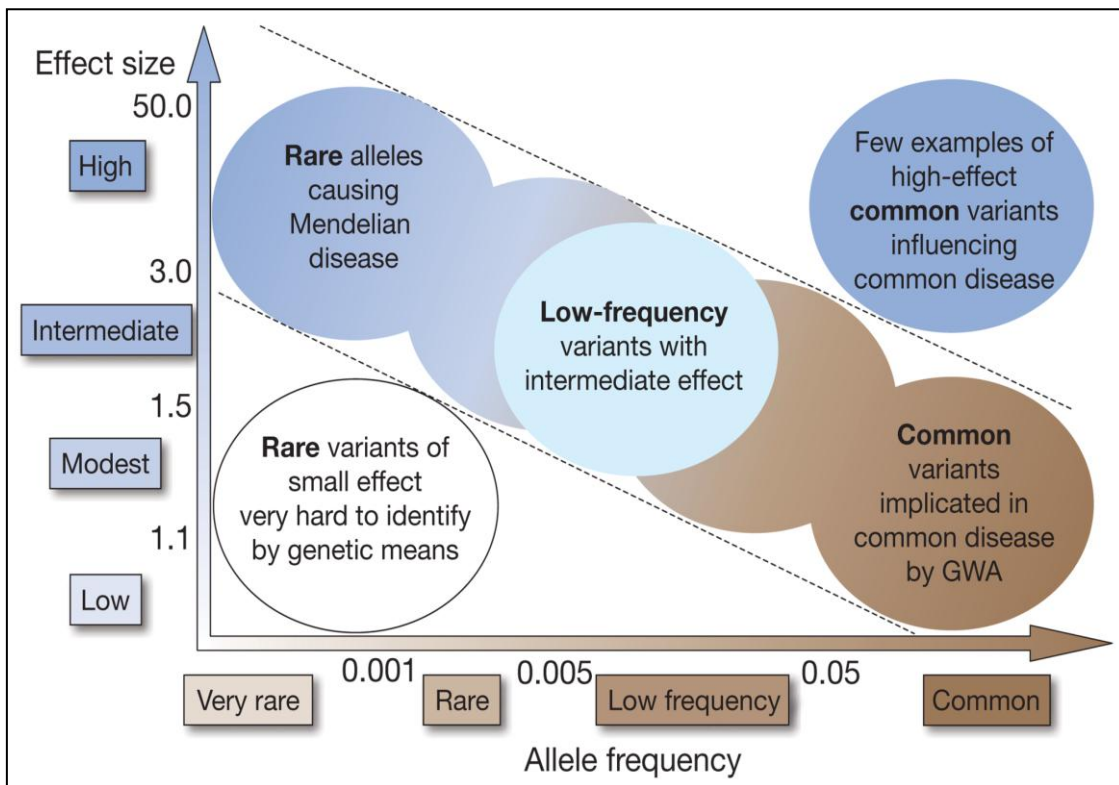


Abbildung 1: Zusammenhang zwischen Allelhäufigkeit und Effektgröße genetischer Varianten. Die X-Achse zeigt die Allelhäufigkeit und definiert die identifizierten Varianten als sehr selten (very rare), selten (rare), niedrig frequent (low frequency) und häufig (common). Die Y-Achse beschreibt die Effektgröße mit den Graden niedrig (low), moderat (modest), mäßig (intermediate) und hoch (high) als Maß für die Penetranz. Übernommen mit freundlicher Genehmigung von Macmillan Publishers Ltd.: Nature, Manolio et al., 2009, copyright (2009).

Die Common-Disease-Rare-Variant (CDRV)-These geht von multiplen, seltenen genetischen Varianten aus, welche in der Population mit einer relativ geringen Frequenz auftreten, aber eine hohe Effektgröße haben (Evans et al., 2011; Schork et al., 2009; Manolio et al., 2009; Manolio et al., 2008). Die Abbildung 1 stellt den Zusammenhang zwischen der Effektgröße einer genetischen Variante und deren Allelfrequenz dar. Monogenetische Erkrankungen nach Mendel besitzen eine geringe Frequenz mit hoher Penetranz. Demgegenüber besitzen häufig vorkommende Varianten meist nur einen geringen Einfluss auf den Phänotyp. Die nach der CDRV-Hypothese ursächlichen seltenen Varianten mit einer Frequenz $< 0,05\%$ besitzen mäßige bis mittlere Effektgrößen. Die Abbildung 1 verdeutlicht auch, dass die beiden Hypothesen sich nicht grundsätzlich ausschließen, da es auch seltene Varianten mit nur geringen Effekten auf den Phänotyp und ebenso häufige Varianten mit hoher Effektstärke gibt. Die genetische Komponente für eine Erkrankung ist das Resultat sowohl seltener als auch häufiger genetischer Varianten mit unterschiedlichen Effektgrößen. Jedoch ist zu berücksichtigen, dass die Effektgröße einiger Gene durch die gegebene Stoffwechsellage im Organismus moduliert

wird. Als Beispiel seien der L162V-Polymorphismus im PPAR α -Gen und der SNP rs3741297 im ApoA5-Gen genannt. Das Risikoallel V des L162V-Polymorphismus hat nur bei Patienten mit Diabetes mellitus Typ 2 einen senkenden Einfluss auf den Body-Mass-Index (BMI) (Evans et al., 2001). Ebenso wird der triglyzeridsteigernde Effekt des T-Allels der genetischen Variante rs3741297 im ApoA5-Gen durch ein erhöhtes Hüft-Tailen-Verhältnis verstärkt (Wu et al., 2013). Die beiden grundlegenden Strategien zur Identifizierung genetischer Varianten werden nachfolgend vorgestellt.

1.2.3 Genomweite Assoziationsstudien

Die Grundlage von genomweiten Assoziationsstudien (genome wide association studies, GWAS) ist die CDCV-Hypothese. Hierbei werden in groß angelegten Studien meist mehrere tausend Patienten nach genetischen Varianten in ihrem Genom untersucht und deren Korrelation mit Erkrankungen oder Merkmalen überprüft. Die Studien sind durch die großen Fallzahlen und durch das Ziel, das Genom zu untersuchen, kosten- und zeitaufwendig. Realisiert werden konnten sie erst durch moderne Hochdurchsatzverfahren in der Sequenzierung und durch den Einsatz von DNA-Arrays. Die Studien werden ohne vorhergehende Vermutungen über physiologische und biochemische Prozesse bzw. damit verbundene Krankheiten durchgeführt. Die identifizierten genetischen Varianten besitzen meist eine geringe Effektstärke und sind mit den untersuchten Merkmalen assoziiert, aber nicht ursächlich für diese (Khor u. Li-Meng Goh, 2010). Durch die immer größer werdende Zahl an GWAS wurde 2008 vom National Human Genome Research Institute (NHGRI) ein Online-Katalog entwickelt. In diesem werden alle Ergebnisse der GWAS zu SNP-assoziierten Phänotypen gesammelt. Seit 2010 beteiligt sich das Europäische Bioinformatik-Institut an diesem Projekt. Die SNPs sind zumeist häufig auftretend mit einer Frequenz des selteneren Allels über 5 % (minor allele frequency, MAF), mäßigen Effektgrößen (odds ratio 1,33) und unterscheiden sich nicht zwischen den untersuchten Populationen (Hindorff et al., 2009). Die schwer überschaubare Anzahl an GWAS birgt einen enormen Wissensschatz an Informationen über die Genetik komplexer Krankheiten. Allein für den Suchbegriff „Lipid“ sind im Online-Katalog, abgerufen im April 2015, über 100 Studien verfügbar und mehr als 1200 assoziierte SNPs gelistet (Burdett et al., 2014). Jedoch erklären die bislang identifizierten Varianten nur einen geringen Anteil der genetischen Komponente komplexer Erkrankungen. Etwa 90-95 % bleiben weiterhin ungeklärt (Schork et al., 2009). Ähnliche Zahlen finden sich auch für Serumlipidwerte. In einer der größten bisher durchgeführten GWA-Studien von Teslovich et al. (2010) wurden 95 bisher unbekannte

Genloci im Zusammenhang mit Triglyzeriden, LDL-Cholesterin, HDL-Cholesterin und Gesamtcholesterin identifiziert. Diese können jedoch nur 10-12 % der Gesamtvarianz und nur 25-30 % der genetischen Varianz in den einzelnen Lipidfraktionen erklären (Teslovich et al., 2010). Trotz der Bemühungen mit weiteren GWA-Studien und Meta-Analysen, neue Genloci zu identifizieren, wird angenommen, dass seltene Varianten mit moderater Penetranz einen beachtlichen Anteil an der fehlenden genetischen Komponente haben. Diese können ab einer MAF unter 0,5 % nicht mehr ausreichend in GWA-Analysen detektiert werden. Für eine Identifizierung durch GWAS wären dann große Effektgrößen, wie bei monogenetischen Erkrankungen, nötig (Manolio et al., 2009; McCarthy et al., 2008).

1.2.4 Resequenzierungsstudien

Die CDRV-Hypothese formuliert als Ursache komplexer Erkrankungen das Vorhandensein multipler seltener Varianten. Durch die Resequenzierung von Genen von Patienten mit Phänotypen im Extrembereich der Merkmalsausprägung, meist als Werte oberhalb der 95. Perzentile und unterhalb der 5. Perzentile für Alter und Geschlecht definiert, können seltene Varianten identifiziert werden. Dabei wird eine Vorauswahl an sogenannten Kandidatengenen, für welche ein Zusammenhang mit dem untersuchten Merkmal vermutet wird, getroffen. Für die Auswahl der Kandidatengene werden vor allem die bereits assoziierten Genloci aus Sequenzierungen der GWAS verwendet. Die identifizierten niedrigfrequenten und seltenen Varianten besitzen meist eine höhere Effektstärke als häufige Varianten, befinden sich jedoch selbst im moderaten bis mäßigen Stärkebereich (siehe Abbildung 1). Dies liegt nicht zuletzt am ausgewählten Studiendesign. Die Limitierung auf Kandidatengene, die vorhergehende Einschätzung der Auswirkung der Genvariante auf das Genprodukt sowie die geringe Wahrscheinlichkeit, dass zwei seltene Varianten mit vergleichbaren Effekten in benachbarten Genbereichen liegen, erhöhen das Potential, seltene Varianten mit hohen Effektstärken zu finden (Khor u. Li-Meng Goh, 2010). Die Mehrheit der seltenen Varianten führt zu einer Veränderung der Aminosäuresequenz und folglich zu einer veränderten Protein-Protein-Interaktion. Sind die Genvarianten im Promoterbereich lokalisiert, kann es auch zu Veränderungen der Genexpression kommen. Daraus ergibt sich eine Notwendigkeit für Resequenzierungsstudien, um das multigenetische Zusammenspiel komplexer Krankheiten zu verstehen (Bodmer u. Bonilla, 2008). Für einige bekannte Lipid-Gene wurden bereits Resequenzierungen durchgeführt und seltene Varianten gefunden, welche Frequenzunterschiede zwischen Individuen mit und

Tabelle 4: Überblick über Resequenzierungsstudien und deren Resultate.

Referenz	Gene	Phänotyp	Resultat
Johansen et al., 2010	ApoA5, LPL, ApoB, GCKR	Erhöhung der Triglyzeride	Multiple seltene missense und nonsense Varianten
Evans et al., 2011	LPL	Erhöhung der Triglyzeride	Multiple seltene missense und nonsense Varianten, sowie Deletionen und eine Insertion
Evans et al., 2011	ApoA5	Erhöhung der Triglyzeride	Multiple seltene missense und nonsense Mutationen sowie Insertionen und Deletionen, z. T. mit Verschiebung des Leserasters
Cohen et al., 2004	ABCA1, ApoA1, LCAT	Erniedrigung des HDL-Cholesterins	Multiple seltene Varianten
Romeo et al., 2007	ANGPTL4	Erniedrigung der Triglyzeride und Erhöhung des HDL-Cholesterins	Multiple synonyme und nicht-synonyme Mutationen
Khovidhunkit et al., 2012	LIPC, CETP	Erhöhung des HDL-Cholesterins	Multiple seltene missense Mutationen, z. T. mit Verschiebung des Leserasters, und eine Deletion
Kotowski et al., 2006	PCSK9	Erniedrigung des LDL-Cholesterins	Mehrere seltene nonsense Mutationen
Cohen et al., 2006	NPC1L1	Erniedrigung des LDL-Cholesterins	Multiple seltene Varianten
Edmondson et al., 2009	LIPG	Erhöhung des HDL-Cholesterins	Multiple seltene loss-of function Varianten

ohne einen bestimmten Phänotyp aufweisen. Einen Überblick über bisher durchgeführte Resequenzierungsstudien im Bereich der Dyslipidämien gibt Tabelle 4. Da davon ausgegangen wird, dass niedrigfrequente bis seltene Varianten eine höhere Effektstärke

auf den Phänotyp haben, ist deren Identifizierung sinnvoll zur Entwicklung neuer Screening- und Präventionsmethoden.

1.3 Der Transkriptionsfaktor Krüppel-like Faktor 14

Das Krüppel-like Faktor 14 (KLF14)-Gen wird auf Chromosom 7q32 kodiert und besitzt keine Introns. Es besteht aus 1383 Basenpaaren. Das Protein ist ein Transkriptionsfaktor und gehört zu einer Gruppe von Krüppel-like Faktoren (KLF) mit derzeit 17 bekannten Vertretern bei Säugetieren. Diese greifen in ihrer Funktion als Transkriptionsfaktoren sowohl aktivierend als auch reprimierend in diverse Zellfunktionen ein. Sie steuern sowohl die Proliferation und Differenzierung von Geweben als auch die Apoptose von Zellen mit. Die KLF-Mitglieder sind beteiligt an der Physiologie und Pathophysiologie von Erkrankungen des respiratorischen und gastrointestinalen Traktes sowie metabolischen Störungen und auch onkologischen Erkrankungen (McConnell u. Yang, 2010). Weiterhin wirken sie modulierend in der Entstehung der Atherosklerose und damit verbundenen kardiovaskulären Erkrankungen. Glatte Muskelzellen können in Abhängigkeit von KLF4 in Makrophagen-ähnliche Zellen konvertieren. Ein Verlust von KLF4 in diesen Zellen führt zu einer Reduktion der Atherosklerose und zu einer erhöhten Plaquestabilität (Shankman et al., 2015; Rosenfeld, 2015). Allen KLF-Proteinen sind drei Zink-Finger-Strukturen, über welche sie die DNA binden können, gemein. Die Bindung der DNA-Strukturen erfolgt über GC-reiche Sequenzen oder CACCC-Box-Elemente. Die Zink-Finger-Strukturen befinden sich nahe dem Carboxy-terminalen Ende des Proteins und besitzen ein Zink-Atom, welches von zwei Cystein- und Histidin-Derivaten vom Typ C_2H_2 umgeben ist. Die Zink-Finger-Strukturen selbst sind durch eine Linker-Sequenz miteinander verbunden. Diese Sequenzen werden als Krüppel-Links bezeichnet (Turner u. Crossley, 1999). Die Zink-Finger-Strukturen werden auch für die Aufnahme des Transkriptionsfaktors in den Zellkern benötigt. Nukleäre Signale sind für die KLF1, -4, -6, -8, -9 und -11 nachgewiesen (Mehta et al., 2009; Pandya u. Townes, 2002; Rodríguez et al., 2010; Pandya et al., 2004; Fan et al., 2012; Knoedler u. Denver, 2014).

Durch die GWA-Studie mit insgesamt 17296 Frauen der Women's Genome Health Study wurde der KLF14-Genlocus im Zusammenhang mit dem Lipoproteinmetabolismus assoziiert. In der Studie sind 43 Genloci identifiziert worden. Darunter waren zehn bislang unbekannte Genloci, zu welchen auch KLF14 gehört. Hierbei zeigte sich eine genomweite Signifikanz für die Lipoproteinfraktionen HDL-Cholesterin, LDL-Cholesterin und Triglyzeride (Chasman et al., 2009). 2011 veröffentlichten Teslovich et

al. eine große Meta-Analyse von 46 GWA-Studien mit Individuen europäischer Herkunft. Dabei wurden mehr als 100 000 Individuen auf häufige Varianten und deren Assoziation zu Triglyzeriden, HDL-Cholesterin, LDL-Cholesterin oder Gesamtcholesterin untersucht. Es wurden SNPs mit einer Frequenz des selteneren Allels $> 1\%$ in die Analyse eingeschlossen. 95 Genloci zeigten eine signifikante Assoziation mit mindestens einem der oben genannten Serumlipidwerte. Zum einen konnten 36 bereits bekannte Genvarianten bestätigt und zum anderen 59 bisher nicht publizierte Genloci identifiziert werden. Eine zusätzliche Assoziation mit einem weiteren Lipidwert zeigten 21 der 36 bereits bekannten Genloci. Von den 59 neuen Varianten sind 39 mit Gesamtcholesterin, 22 mit LDL-Cholesterin, 16 mit Triglyzeriden und 31 mit HDL-Cholesterin assoziiert. Der SNP rs1562398 wurde, neben der von Chasman et al. identifizierten Variante, als neue genetische Variante des KLF14-Gens für HDL-Cholesterin präsentiert. Dieser zeigt geschlechtsspezifische phänotypische Ausprägungen. Der für beide Geschlechter mit HDL-Cholesterin assoziierte SNP ist signifikant mit niedrigeren Triglyzeridwerten bei Frauen assoziiert, nicht jedoch bei Männern. Die klinische Relevanz der identifizierten Varianten wurde ebenfalls untersucht. Hier ist KLF14 signifikant mit einem verminderten Risiko einer koronaren Herzkrankheit assoziiert (Teslovich et al., 2010).

Des Weiteren konnte KLF14 in der DIAGRAM+ Meta-Analyse mit Diabetes mellitus Typ 2 assoziiert werden. Dabei zeigte sich ein direkter Effekt auf die Wirkung von Insulin, welcher nicht durch Adipositas beeinflusst zu sein scheint. Eine ebenfalls reduzierte Insulinsensitivität wurde mit dem KLF14-Locus und zwei weiteren Loci demonstriert (Voight et al., 2010).

Als Transkriptionsfaktor dirigiert KLF14 ein ganzes Netzwerk von Genen durch seine trans-regulatorische Funktion. Dabei wird die Expression des KLF14-Gens selbst durch cis-regulatorische Mechanismen kontrolliert. Unter einer cis-Regulation versteht man eine Kontrolle der Expression durch auf demselben DNA-Strang liegende Bereiche. Diese DNA-Abschnitte können dabei die Transkription sowohl verstärken (Enhancer) als auch hemmen (Silencer). Als trans-Element wird ein DNA-Abschnitt bezeichnet, welcher auf nicht auf demselben DNA-Strang liegende DNA-Bereiche Einfluss nehmen kann. Das als trans-Element kodierte Protein reguliert dann die Expression des auf einem anderen DNA-Strang liegenden Gens (Abbildung 2).

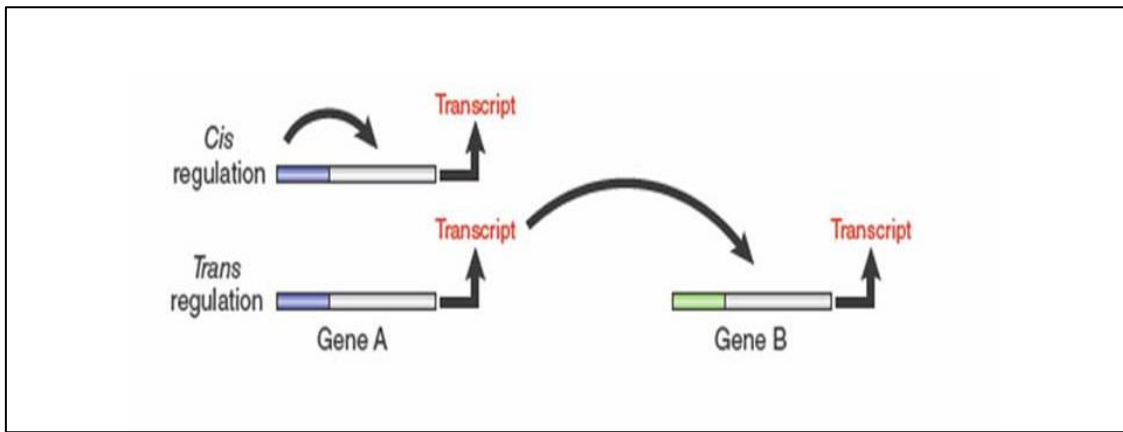


Abbildung 2: Darstellung der cis- und trans-Funktion von Genen. Das cis-Element und das dadurch kontrollierte Gen befinden sich auf dem gleichen DNA-Strang. Bei einer trans-Regulation können das trans-Element auf Gen A (Gene A) und das dadurch kontrollierte Gen B (Gene B) auf verschiedenen DNA-Strängen liegen. Übernommen mit freundlicher Genehmigung von Macmillan Publishers Ltd.: Nature Genetics, Civelek u. Lusis, 2011, copyright (2011).

Die genetische Variante rs4731702 mit dem Risikoallel C zeigt bei Vererbung mütterlicherseits eine starke Assoziation mit der Expression von KLF14 im subkutanen Fettgewebe sowie Diabetes mellitus Typ 2 (Kong et al., 2009). Dies belegt zum einen die cis-regulatorische Wirkung der Variante auf die Transkription von KLF14 und zum anderen, dass die Vererbung des KLF14-Gens dem epigenetischen Effekt des Imprinting unterliegt und nur durch das maternale Allel exprimiert wird. Die trans-regulatorische Funktion des KLF14-Transkriptionsfaktors wurde von Small et al. (2011) näher untersucht. Für insgesamt zehn Gene konnte eine trans-assoziierte Expression durch den rs4731702-Polymorphismus in subkutanem Fettgewebe nachgewiesen werden. Die trans-assoziierten Gene sind mit einer Reihe metabolischer Phänotypen assoziiert. So zeigen sechs der zehn Gene eine signifikante Assoziation mit dem BMI und den HDL-Cholesterinwerten, jeweils fünf mit Triglyzerid- und Nüchterninsulinwerten, vier mit einem Index für Insulinsensitivität und jeweils zwei mit Nüchtern glukosewerten und Adiponectin (Small et al., 2011). Der Transkriptionsfaktor KLF14 beeinflusst somit über die trans-Regulation von Genen eine Reihe von Stoffwechselwegen (Abbildung 3). Die assoziierten Parameter beinhalten alle Faktoren des metabolischen Syndroms und lassen eine entscheidende Rolle von KLF14 in der Genese des metabolischen Syndroms vermuten (Civelek u. Lusis, 2011).

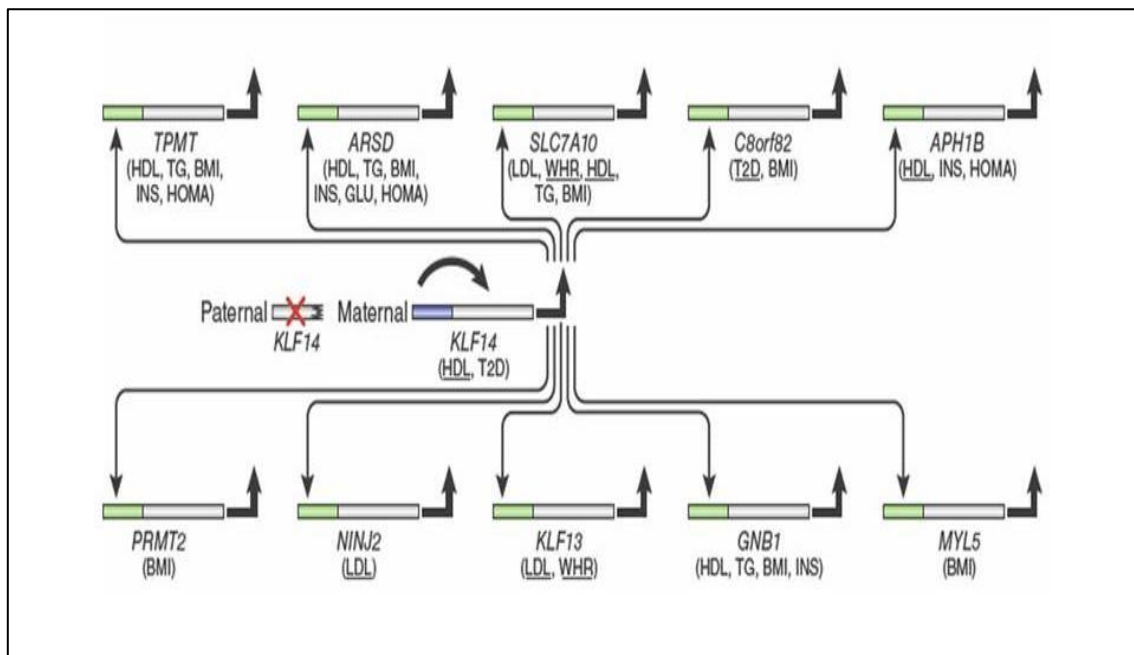


Abbildung 3: Darstellung der cis- und trans-regulatorischen Funktion des KLF14-Gens. Die cis-Funktion auf die Expression des KLF14-Gens wird durch das mütterliche (maternale) Allel bestimmt. Das väterlicherseits (paternal) vererbte Allel hat keine Auswirkungen. KLF14 selbst wirkt durch trans-regulatorische Mechanismen auf die Expression von Genen, welche mit den Hauptfaktoren des metabolischen Syndroms assoziiert sind. Für die untersuchten Parameter sind SNPs mit genomweiter Signifikanz in den trans-regulierten Genen bekannt. Die verwendeten Abkürzungen sind im Abkürzungsverzeichnis gelistet. Mit freundlicher Genehmigung übernommen von Macmillan Publishers Ltd.: Nature Genetics, Civelek u. Lusis, 2011, copyright (2011).

In einer Studie mit chinesischen Probanden wurden 79 mit dem Lipidmetabolismus und 13 mit koronarer Herzkrankheit assoziierte Genloci auf deren Risiko für arteriosklerotisch bedingte kardiovaskuläre Erkrankungen, inklusive Schlaganfall und Myokardinfarkt, untersucht (Chen et al., 2012). Das seltenere T-Allel des SNP rs4731702 war signifikant mit einem reduzierten Risiko für Schlaganfall und Myokardinfarkt und erhöhten HDL-Cholesterinwerten assoziiert. Der erstmals identifizierte Polymorphismus rs111400400 im KLF14-Gen ist ebenfalls mit einem niedrigeren Myokardinfarktrisiko assoziiert. Für das seltenere G-Allel wurde eine Odds Ratio von 0,84 pro Allel zur Reduzierung des Risikos berechnet. Im Folgenden wird der SNP rs111400400 als S58P bezeichnet.

1.3.1 Die genetische Variante S58P im KLF14-Gen

Die hier dargestellte Sequenz des reversen DNA-Stranges ist der webbasierten Datenbank Ensembl entnommen (Cunningham et al., 2015). In der ersten Zeile ist die fortlaufende Zahl des jeweiligen Codons, in der zweiten Zeile die dazugehörige Gensequenz und in der letzten Zeile die entsprechend kodierte Aminosäure dargestellt. Die Gense-

quenz wird durch den Ambiguity Code verschlüsselt und die Aminosäuren durch den Ein-Buchstaben-Code. Rot markiert wurden die Variante und die möglichen Aminosäuren.

Der S58P-Polymorphismus ist eine Punktmutation im Exonbereich des KLF14-Gens. In der Nomenklatur bezieht sich der erste Buchstabe auf die Aminosäure der häufigeren Variante, die Zahl auf die Position des Codons und der letzte Buchstabe auf die Aminosäure der selteneren Variante. Für die Variante S58P bedeutet dies, dass die Aminosäure Serin gegen Prolin im Codon 58 ausgetauscht wird (siehe Abbildung 4). Nachfolgend ist ein Ausschnitt der KLF14-Gensequenz dargestellt, welcher die Punktmutation einschließt. Es handelt sich bei der genetischen Variante um eine missense-Mutation. Bei dieser wird durch Punktmutation eine Aminosäure ersetzt, jedoch verändert sich die Länge der Gensequenz dadurch nicht.

52	53	54	55	56	57	58	59	60	61	62	63	64	65
CTG	CCG	GGT	CCG	GGG	CCA	RCG	GGG	CCC	GCG	TCG	GTC	CCC	CAG
L	P	G	P	G	P	S/P	G	P	A	S	V	P	Q

Abbildung 4: Ausschnitt der KLF14-Gensequenz im Bereich des S58P-Polymorphismus. Modifiziert nach Cunningham et al., 2015.

1.3.2 Die genetische Variante P47Q im KLF14-Gen

Die genetische Variante P47Q ist ebenfalls eine missense Mutation (siehe Abbildung 5). Durch Punktmutation im Codon 47 wird die Aminosäure Prolin durch Glutamin ersetzt.

41	42	43	44	45	46	47	48	49	50	51	52	53	54
TCG	GAG	GTG	GGT	GCG	GCG	CMG	CCG	GAG	TCC	GCT	CTG	CCG	GGT
S	E	V	G	A	A	P/Q	P	E	S	A	L	P	G

Abbildung 5: Ausschnitt der KLF14-Gensequenz im Bereich des P47Q-Polymorphismus. Modifiziert nach Cunningham et al., 2015.

1.3.3 Die genetische Variante A39A im KLF14-Gen

Diese Punktmutation betrifft das Codon 39 der Gensequenz für den Transkriptionsfaktor KLF14 (siehe Abbildung 6). Da es sich um die dritte Base des Basentriplets handelt, wird die bestehende Aminosäure Alanin nicht verändert. Der Polymorphismus ist somit eine synonyme oder stille Mutation.

33	34	35	36	37	38	39	40	41	42	43	44	45	46
GAG	GGC	GCG	GGT	GGA	GCC	GCH	GGC	TCG	GAG	GTG	GGT	GCG	GCG
E	G	A	G	G	A	A/A	G	S	E	V	G	A	A

Abbildung 6: Ausschnitt der KLF14-Gensequenz im Bereich des A39A-Polymorphismus. Modifiziert nach Cunningham et al., 2015.

1.4 Fragestellung

Dyslipidämien allein oder als Teil des metabolischen Syndroms sind Risikofaktoren für Arteriosklerose und damit verbundene Folgeerkrankungen wie Myokardinfarkt, Schlaganfall oder koronare Herzerkrankungen. Die präzise genetische Ursache zu kennen ist die Voraussetzung für neue Therapiemöglichkeiten und dient der Entwicklung frühzeitiger Screening-, Präventions- und Diagnostikmethoden. Dem KLF14-Gen kommt als Transkriptionsfaktor in der Regulation eines ganzen Netzwerkes von Stoffwechselwegen eine besondere Bedeutung zu. Varianten in diesem Gen sind mit dem Lipidmetabolismus assoziiert und zeigen ihre klinische Relevanz durch eine Verminderung des Myokard- und Schlaganfallrisikos. In der vorliegenden Arbeit wurde die genetische Variante S58P des KLF14-Gens auf einen möglichen Zusammenhang mit Dyslipidämien und den metabolischen Erkrankungen Diabetes mellitus Typ 2 und Adipositas untersucht. Dazu wurden Patienten der Lipidambulanz des Universitätsklinikums Hamburg-Eppendorf genotypisiert. Das Ziel ist es, bei verschiedenen Dyslipidämieformen Frequenzunterschiede des SNP S58P zu identifizieren. Die Patienten wurden dafür in die Kategorien Hypertriglyzeridämie, gemischte Hyperlipidämie, Hypercholesterinämie, familiäre Hypercholesterinämie oder Normolipidämie eingeteilt. Eine weitere Einteilung erfolgte nach Perzentilen für HDL-Cholesterin. Außerdem wird ein Zusammenhang zwischen den Genotypen und den durchschnittlichen Serumlipidwerten sowie den Parametern Adipositas und Diabetes mellitus Typ 2 untersucht. Das Ziel ist die Analyse eines Zusammenhanges zwischen der genetischen Variante S58P und Veränderungen des Lipidhaushaltes sowie Bestandteilen des metabolischen Syndroms. Im Genbereich von KLF14 sind weiterhin die Varianten P47Q und A39A bekannt. Ein Subkollektiv wurde sequenziert und auf ein mögliches Kopplungsungleichgewicht zwischen den SNPs S58P, P47Q und A39A im KLF14-Gen getestet.

2 Material und Methoden

2.1 Chemikalien

- Aqua dest.
- Betaine Solution 5M (Sigma Life Science)
- HotStar TaqTM MasterMix 0,85ml (QIAGEN)
- Vorwärtsprimer S58PFP: 5'-GAT CGG GTC GGA GAA GCC GGA CG-3'
(Eurofins MWG Operon)
- Rückwärtsprimer S58PRP: 5'-GCG AGC ACC CGG CCT CCT GCT TCT-3'
(Eurofins MWG Operon)
- Restriktionsendonuklease BccI 10.000 U/ml #R0704L
(New England BioLabs, Inc.)
- QX DNA Size Marker (QIAGEN)
- Alignment Marker 15-500bp (QIAGEN)
- QIAquick PCR Purification Kit (250) (QIAGEN)
- Exonuklease I (20 U/μl) (New England Biolabs)
- Shrimp Alkaline Phosphatase (SAP) (1 U/μl) (Roche)

2.2 Geräte und Arbeitsmittel

- Pipetten, verschiedene Größen (Eppendorf)
- Pipettenspitzen mit Filter Typ Biosphere Fil (Sarstedt)
- Reaktionsgefäße, verschiedene Größen (Sarstedt)
- Reaktionsgefäße 0,2 ml (ThermowellTM TubeStripes)
- T3000 Thermocycler (Biometra)
- SEPATECH Biofuge 13 (Heraeus SEPATECH)
- CleneCab (Herolab GmbH Laborgeräte)
- QIAxcel-System (QIAGEN) mit dem QIAxcel DNA Screening Kit (QIAGEN)
und der BioCalculator Software
- Inkubator (Heraeus)
- Vortexer Typ REAX1 (Heidolph Instruments GmbH)

2.3 Software

- Finch TV (Finch TV Version 1.4.0, Copyright© 2004-2006, Geospiza Inc.)
- STATISTICA (Version 10), StatSoft, Inc. (2011)
- Microsoft Office Excel 2007 (Microsoft Corp., Bothell, Washington)

2.4 Methoden

Nachfolgend ist die Durchführung der verwendeten Methoden beschrieben. Für ergänzende Beschreibungen des Prinzips der genutzten Verfahren siehe Anhang 1.

2.4.1 Polymerasekettenreaktion (PCR)

Es wurde zuerst ein Master-Mix für 25 Proben hergestellt. Um eine Kontamination mit fremder DNA zu vermeiden, erfolgte dies unter einer abgeschirmten sterilen Werkbank (clene cab). Für den Master-Mix wurden benötigt:

Master-Mix: 75 µl Aqua dest.
100 µl Betaine Solution
250 µl HotStarTaq Master Mix Kit
25 µl Vorwärtsprimer S58PFP
25 µl Rückwärtsprimer S58PRP

Der Master-Mix wurde mit dem Vortex gut durchmischt. Jeweils 19 µl des angesetzten Master-Mix wurden in 24 Reaktionsgefäße pipettiert. Danach wurde jeweils 1 µl der Patienten-DNA in 23 Reaktionsgefäße pipettiert. Das letzte Reaktionsgefäß dient als Negativkontrolle für eine eventuelle Kontamination. In dieses wurde 1 µl Aqua dest. pipettiert. Zur Durchführung der einzelnen Reaktionsschritte der PCR wurden die Reaktionsgefäße im Thermocycler platziert. Nach einmaligem Aufheizen auf 96 °C für 15 min. zu Beginn war das Programm folgendermaßen eingestellt:

1 Zyklus:	Denaturierung	96 °C/ 1 min.
	Annealing	60 °C/ 1 min.
	Elongation	72 °C/ 1 min.
33 Zyklen:	Denaturierung	96 °C/ 1 min.
	Annealing	60 °C/ 1 min.
	Elongation	72 °C/ 1 min.
1 Zyklus:	Denaturierung	96 °C/ 1 min.
	Annealing	60 °C/ 1 min.
	Elongation	72 °C/ 10 min.

Nach Beendigung der PCR wurden die Reaktionsgefäße bis zur weiteren Verwendung bei 4 °C im Kühlschrank gelagert.

2.4.2 Restriktionsendonukleasenverdau

Der folgende Versuchsaufbau dient dem Restriktionsverdau von 20 PCR-Reaktionsgefäßen. Zur Fragmentierung von 10 µl PCR-Produkt sind 2 µl BccI in einer 1:10-Verdünnung nötig. In einem 0,5-ml-Eppendorf-Reaktionsgefäß wurde ein Ansatz für 21 PCR-Produkte hergestellt. Dazu wurden 4,1 µl BccI-Enzym pipettiert und 36,9 µl Aqua dest. hinzugefügt. Es wurden dann jeweils 2 µl des Restriktionsenzym auf 20 Reaktionsgefäße verteilt und anschließend wurden jeweils 10 µl des PCR-Produkts hinzugefügt. Die Inkubation erfolgte für drei Stunden bei einer Temperatur von 37 °C. Bis zur weiteren Analyse mittels Kapillargelelektrophorese wurden die Restriktionsverdaus bei 4 °C im Kühlschrank gelagert.

2.4.3 Kapillargelelektrophorese

Die Kapillargelelektrophorese wurde mit Hilfe des QIAxcel-System (QIAGEN) durchgeführt. Zur Auswertung wurde die Screening Cartridge verwendet. Die Einstellung der Auflösungsmethode war AH320. In einer Reihe liefen jeweils 12 Proben aus dem Restriktionsverdau gleichzeitig. Der verwendete Alignmentmarker hatte eine Basenspannbreite von 15-500 bp. Die entstehenden Bandenmuster durch den Restriktionsverdau mit BccI für die verschiedenen Genotypen sind in Tabelle 5 zusammengefasst.

Tabelle 5: Bandenmuster nach Restriktionsverdau mit BccI.

Genotyp	AA	AG	GG
Bandenmuster (bp)		406	406
	229	229	
	177	177	

2.4.4 Sequenzierung

Die DNA-Sequenzierung erfolgt nach dem Prinzip des Kettenabbruchs bzw. nach der Didesoxymethode nach Sanger. Vor der Sequenzierung wurde das PCR-Produkt aufgereinigt.

2.4.4.1 Aufreinigung des PCR-Produkts

Die Aufreinigung des PCR-Produkts erfolgte entsprechend dem Protokoll des QIAquick PCR-Purification Kit. Es wurde jeweils 100 µl Puffer PB zu 20 µl des PCR-Produkts pipettiert und gemischt. Das Gemisch wurde dann in eine QIAquick-Säule mit dazugehörigem 2-ml-Sammelgefäß pipettiert. Zur Bindung der DNA mit dem PB-Puffer er-

folgte die Zentrifugation bei 13000 rpm für 1 min.. Der abzentrifugierte Durchfluss wurde verworfen und die QIAquick-Säule wieder in das zugehörige 2-ml-Sammelröhrchen platziert. Anschließend erfolgte die Aufreinigung. Dafür wurden je 750 µl Puffer PE in die QIAquick-Säulen pipettiert und nochmals für 1 min. bei 13000 rpm zentrifugiert. Nach Abgießen des entstandenen Durchflusses wurde mit denselben Einstellungen wiederholt zentrifugiert. Danach wurden die QIAquick-Säulen in ein neues 1,5-ml-Reaktionsgefäß platziert. Zur Elution der amplifizierten DNA wurden 50 µl Puffer EB auf die Membran der QIAquick-Säule pipettiert und bei 13000 rpm für 1 min. zentrifugiert. Zur Beurteilung der erfolgreichen Aufreinigung wurde eine Kapillargel-elektrophorese mit dem QIAxel-System durchgeführt.

Als weitere Methode zur Aufreinigung wurde der Restriktionsverdau mit den Enzymen Shrimp Alkaline Phosphatase (SAP) und Exonuklease I (ExoI) genutzt. Hierbei muss das aufzureinigende PCR-Produkt nicht in ein anderes Gefäß pipettiert werden und die Gefahr einer Verunreinigung wird minimiert. Die ExoI entfernt enzymatisch überschüssige Primer, während die SAP Nukleotide entfernt. Es wurde ein Master-Mix für die Aufreinigung von fünf Proben hergestellt.

Master-Mix: 0,6 µl Exo I
 3,0 µl SAP
 2,4 µl Aqua dest.

Jeweils 1 µl des Master-Mixes wurde zu 10 µl des PCR-Produkts pipettiert und vermischt. Danach erfolgte die Inkubation im Thermocycler für 30 min. bei 37 °C und danach bei 80 °C für weitere 20 min.. Zur Beendigung der Reaktion wurde das Gemisch abgekühlt auf 4 °C und im Kühlschrank bis zur weiteren Verwendung gelagert.

2.4.4.2 Sequenzierung

Unter der clene cab wurde ein Master-Mix für fünf Proben angesetzt. Hierfür wurden Aqua dest. und der Rückwärtsprimer S58PRP wie folgt pipettiert:

Master-Mix: 25 µl Aqua dest.
 5 µl Primer S58PRP

Es wurden dann jeweils 6 µl des Mixes auf vier Reaktionsgefäße pipettiert. Die Zugabe des aufgereinigten PCR-Produkts erfolgte außerhalb der clene cab. Jeweils 1 µl des gereinigten PCR-Produktes wurde zum Master-Mix pipettiert und gemischt. Die Proben wurden zur Sequenzierung an die Firma Seqlab (Seqlab-Sequence Laboratories Göttingen GmbH, Hannah-Vogt-Str.1, 37085 Göttingen) versandt und konnten am Folgetag

auf deren Website unter „<http://www.seqlab.de>“ abgerufen werden. Die Elektropherogramme wurden mit Hilfe des Programms FinchTV visualisiert und ausgewertet.

2.5 Datenverwaltung und statistische Auswertung

Die statistischen Verfahren lassen sich in zwei große Bereiche unterteilen. Die deskriptive Statistik beschäftigt sich mit der Strukturierung und Zusammenfassung von Ergebnissen. Mit der induktiven Statistik werden die im Rahmen von Studien und Forschungsprojekten gestellten Hypothesen überprüft. Diese ermöglicht mittels geeigneter Testverfahren den Schluss von der untersuchten Stichprobe hin zur Grundgesamtheit (Eckstein, 2014). Die für die vorliegende Arbeit berechneten deskriptiven Werte und durchgeführte Testverfahren erfolgten mit der Computer-Software STATISTICA (Version 10). Mit Hilfe des Programms wurden die Variablen in einer Datentabelle gesammelt und verarbeitet. Zudem wurden wichtige Ergebnisse grafisch veranschaulicht. Zur Erstellung von Balkendiagrammen wurde Microsoft Office Excel 2007 genutzt. In der vorliegenden Arbeit wurde untersucht, ob das seltenere Allel G der genetischen Variante S58P im Genlocus KLF14 überzufällig häufig, d. h. statistisch signifikant, mit verschiedenen Phänotypen auftritt. Dabei geben die Häufigkeiten p_1 und p_2 den jeweiligen Anteil der Allele an der Grundgesamtheit eines untersuchten Phänotyps an. Die Nullhypothese H_0 formuliert dementsprechend, dass die Verteilung der beiden Allele A und G unabhängig von den untersuchten Phänotypen ist ($p_1 = p_2$). Als komplementäre Aussage postuliert die Alternativhypothese H_1 , dass die Allelverteilung und der entsprechende Phänotyp voneinander abhängig, d. h. assoziiert sind.

Es wurden verschiedene statistische Testverfahren genutzt. Die Allelfrequenz wurde durch Genzählung für den jeweiligen Phänotyp bestimmt und mittels Fishers exakten Test verglichen. Der χ^2 -Test kam bei der Berechnung des Hardy-Weinberg-Gleichgewichts zum Einsatz. Auf den U-Test nach Mann und Whitney wurde bei der Testung der Effekte der Genotypen auf die Serumlipidwerte zurückgegriffen. Als Signifikanzniveau für alle Tests wird $\alpha = 0,05$ festgelegt. Somit gelten p-Werte $\leq 0,05$ als statistisch signifikant. Als Tendenz wird $p \leq 0,1$ definiert. Im Anhang 2 werden die genutzten statistischen Mittel näher beschrieben.

2.6 Berechnung eines Hardy-Weinberg-Gleichgewichts

Das Hardy-Weinberg Gesetz wurde 1908 von dem britischen Mathematiker G. H. Hardy (Hardy, 1908) veröffentlicht und unabhängig davon im selben Jahr vom deutschen

Arzt W. Weinberg (Weinberg, 1908). Es beschreibt eine Grundregel der Populationsgenetik, bei der sich in einer idealen Population die Allelfrequenzen von einer Generation zur anderen nicht ändern und damit auch die Frequenzen der Genotypen unverändert bleiben. Dies wird auch als Hardy-Weinberg-Gleichgewicht oder Hardy-Weinberg-Equilibrium bezeichnet. Zum Erreichen des Gleichgewichts müssen verschiedene Bedingungen erfüllt werden. Die Voraussetzungen für ein Hardy-Weinberg-Gleichgewicht sind (Mayo, 2008):

- ein diploider Organismus
- keine geschlechtsabhängigen Genfrequenzen
- eine sehr große (unendliche) Population
- zufällige Paarungen
- keine Einwirkungen durch Selektion, Mutation oder Migration.

Unter der Annahme, dass es sich um einen biallelischen Locus handelt mit den Allelen A und a mit den zugehörigen Allelhäufigkeiten p und q, gelten in der Standardpopulation, bei unabhängiger Vererbung der Allele, folgende Gleichungen für die Häufigkeiten der Genotypen:

Formel 1: Berechnung der Häufigkeit der Genotypen in einer idealen Population.

$$P(AA) = p^2$$

$$H(Aa) = 2pq$$

$$Q(aa) = q^2$$

sowie

$$P + H + Q = 1$$

$$(p + q)^2 = p^2 + 2pq + q^2 = 1$$

Die in einer Population erhobenen Genotyphäufigkeiten sollten auf die Gültigkeit des Hardy-Weinberg-Equilibrium getestet werden. Eine Abweichung ist ein Marker für Veränderungen der Populationsstruktur, kann zufallsbedingt oder Folge einer Störung des gewählten Verfahrens der Genotypisierung sein (Wigginton et al., 2005). Zunächst wird die Allelhäufigkeit berechnet, indem die Anzahl beobachteter Allele durch die Gesamtanzahl der Allele 2N geteilt wird. Hier ist N die Anzahl der Probanden.

Danach werden die zu erwartenden Genotyphäufigkeiten P_E , H_E und Q_E unter Hardy-Weinberg-Gleichgewicht wie folgt berechnet:

Formel 2: Berechnung der erwarteten Häufigkeit der Genotypen in einer idealen Population.

$$\begin{aligned}P_E(AA) &= p^2 \times N \\H_E(Aa) &= 2pq \times N \\Q_E(aa) &= q^2 \times N\end{aligned}$$

Mit Hilfe eines Chi²-Test mit einem Freiheitsgrad werden dann die beobachteten (O_i) und erwarteten Genotyphäufigkeiten (E_i) für alle drei Genotypen verglichen.

Formel 3: Berechnung der Prüfgröße χ^2 im Hardy-Weinberg-Gleichgewicht.

$$\chi^2 = \sum_{i=1}^3 \frac{(E_i - O_i)^2}{O_i}$$

Der ermittelte χ^2 -Wert wird mit dem tabellarischen Wert für die χ^2 -Verteilung für einen Freiheitsgrad bei einem Signifikanzniveau von 5 % verglichen. Wenn $\chi^2 < \chi^2_{5\%}$, liegt ein Hardy-Weinberg-Gleichgewicht in der untersuchten Population vor (Bickeböller u. Fischer, 2007).

2.7 Kopplungsungleichgewicht

Das Kopplungsungleichgewicht (linkage disequilibrium, LD) bezeichnet die nicht zufällige häufigere gemeinsame Vererbung von zwei Allelen an zwei verschiedenen Genorten (Murken et al., 2006). Die Wahrscheinlichkeit, dass zwei Genloci zusammen vererbt werden, ist größer, wenn diese räumlich nahe auf dem Chromosom lokalisiert sind. Durch den Bezug der Definition auf die Kopplung von Allelen wird das LD auch als allelische Assoziation bezeichnet. Eine weitere Definitionsmöglichkeit ist die Beschreibung des LD als eine statistische Assoziation zwischen Sequenzvarianten verschiedener Positionen entlang eines Chromosoms (Weiss u. Clark, 2002). Zur Ermittlung des Maßes der Assoziation der untersuchten Varianten wurde 1964 von Lewontin (Lewontin, 1964) der Disequilibriumskoeffizient D eingeführt. Zur Veranschaulichung der Berechnung sei Tabelle 6 gezeigt. Dargestellt ist der Genlocus 1 mit den Allelen A_1 und a_1 und den dazugehörigen Allelfrequenzen p_1 und q_1 . Das gleiche gilt für den Genlocus 2.

Tabelle 6: Schema der möglichen 2-Genlocus-Genotypen.

		Genlocus 2		Frequenz
Genlocus1		B ₂	b ₂	
	A ₁	A ₁ B ₂	A ₁ b ₂	p ₁
	a ₁	a ₁ B ₂	a ₁ b ₂	q ₁
Frequenz		p ₂	q ₂	1

Zur Berechnung von D wird die Differenz zwischen beobachteter und erwarteter Häufigkeit der 2-Genlocus-Genotypen berechnet. Die am häufigsten verwendete Formel ist nachfolgend gezeigt:

Formel 4: Berechnung des Disequilibriumskoeffizienten D .

$$D = f(A_1B_2) f(a_1b_2) - f(A_1b_2) f(a_1B_2)$$

Der Wertebereich für D ist $-0,25 \leq D \leq 0,25$. Bei vollständiger Kopplung und dem Fall, dass p_1 und p_2 jeweils 0,5 sind, ist $D \pm 0,25$.

3 Ergebnisse

3.1 Patientenkollektiv

In der vorliegenden Arbeit wurde die DNA von 1325 Patienten der III. Medizinischen Klinik und Poliklinik des Universitätsklinikums Hamburg-Eppendorf untersucht. Dabei wurden Probanden zwischen den Serennummern 2004 bis 19296 ausgewählt. Davon wurden 1324 Patienten für die statistische Auswertung genutzt. Ein Patient wurde aufgrund einer Deletionsmutation im Bereich der genetischen Variante S58P ausgeschlossen. Die Patienten wurden zufällig hinsichtlich Geschlecht, Alter, ethnischer Herkunft oder Laborwerte ausgewählt. Die Patienten des untersuchten Kollektivs sind nicht miteinander verwandt und spiegeln die Bevölkerung im Großraumgebiet Hamburg wider. Ein positives Votum der Ethikkommission der Ärztekammer Hamburg liegt vor.

In der Tabelle 7 sind die klinischen Parameter der untersuchten Patientengruppe dargestellt.

Tabelle 7: Darstellung der klinischen Parameter des Gesamtkollektivs.

Parameter	n	arithm. Mittel (±Standardabweichung)	Median
Männer/Frauen	730/594		
Alter (Jahre)	943	47 (±15)	47
BMI (kg/m²)	837	27,17(±4,21)	27
Gesamtcholesterin (mg/dl)	810	259,24 (±76,61)	251
LDL (mg/dl)	708	165,55(±61,58)	158
HDL (mg/dl)	807	50,87(±16,39)	48
Triglyzeride (mg/dl)	809	292,92(±524,79)	172
Manifester Diabetes mellitus Typ 2	155		
Raucher (aktiv + Ex-Raucher)	764 (188 + 576)		

3.2 Genotypisierung des SNP S58P

Insgesamt wurden 1325 Patienten für den SNP S58P genotypisiert. Eine Typisierung konnte nicht eindeutig zugeordnet werden. Diese wurde sequenziert und die Ergebnisse in Kapitel 3.8 zusammengefasst. Die Probe wurde aus der statistischen Auswertung ausgeschlossen. Es wurden 734 Patienten mit dem häufig homozygoten Genotyp AA, 66 Patienten mit dem selten homozygoten Genotyp GG und 524 Patienten mit einem

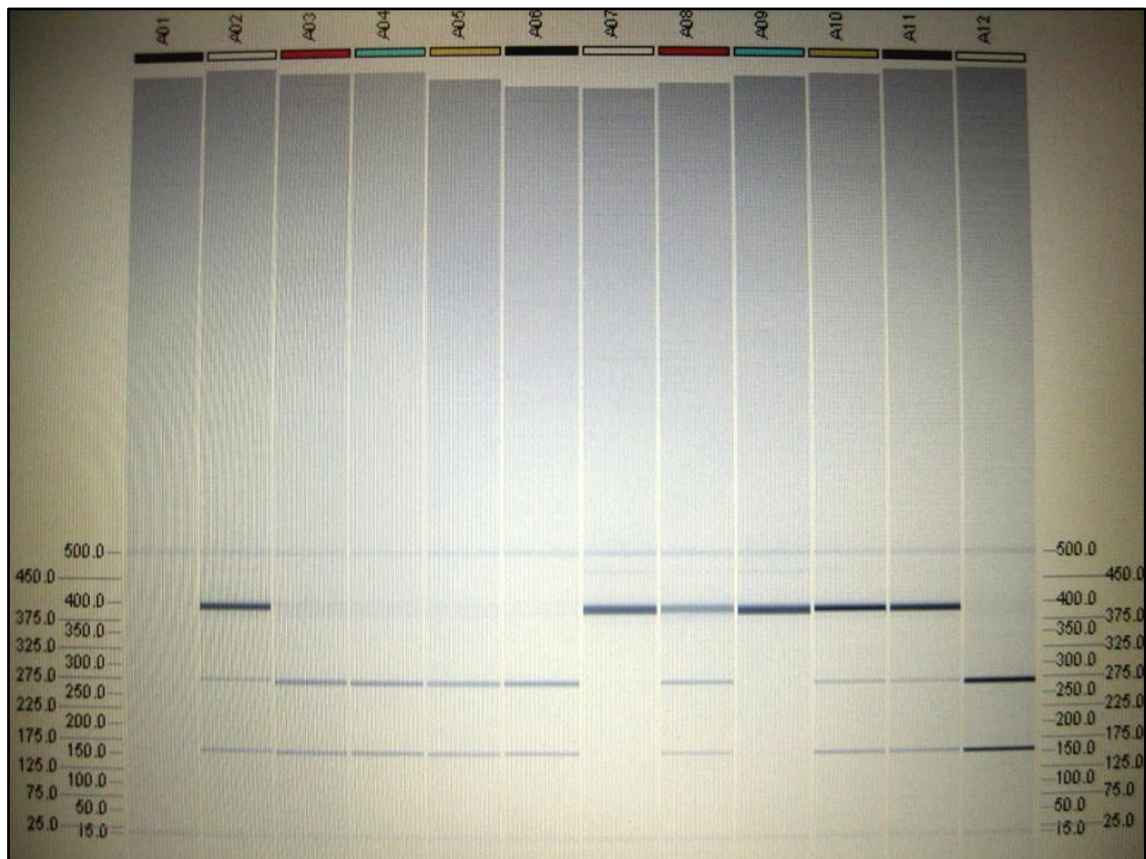


Abbildung 7: Darstellung des Elektropherogramms vom 20.06.2012 für den SNP S58P. Die Proben A02, A08, A10 und A11 zeigen das Bandenmuster des heterozygoten Genotyps AG. Die Proben A03-A06 zeigen das Bandenmuster des häufig homozygoten Genotyps AA und die Proben A07 und A09 das Bandenmuster des selten homozygoten Genotyps GG.

heterozygoten Genotyp AG ermittelt. Die Abbildung 7 zeigt beispielhaft das Ergebnis einer Genotypisierung von elf Patienten. Die homozygoten GG-Genotypen wurden wiederholt, um einen inkompletten Restriktionsverdau auszuschließen. In Tabelle 8 findet sich eine detaillierte Auflistung der Genotypverteilung und der Frequenz des selteneren G-Allels im Kollektiv sowie eine separate Analyse für Männer und Frauen. Die G-Allelfrequenz lag im gesamten Kollektiv bei 0,25 und entspricht damit ähnlichen Werten der dbSNP-Datenbank (Sherry et al., 2001), welche eine G-Allelfrequenz von 0,26 angibt. Es finden sich keine signifikanten Unterschiede in der Verteilung der G-Allelfrequenz zwischen Männern und Frauen ($p = 0,2$).

Tabelle 8: Darstellung der Genotypverteilung und der Frequenz des G-Allels.

	Genotyp			
	AA	AG	GG	G-Allelfrequenz
Gesamtkollektiv	734	524	66	0,25
Männer	395	295	40	0,26
Frauen	339	229	26	0,24

3.3 Hardy-Weinberg-Gleichgewicht

Es wurde das Hardy-Weinberg-Gleichgewicht (HWG) für den SNP S58P berechnet. Zunächst wurden die erwarteten Genotyphäufigkeiten für das Gesamtkollektiv ermittelt. Die Ergebnisse sind der Tabelle 9 zu entnehmen. Der χ^2 -Wert für das Gesamtkollektiv beträgt 5,06 und ist damit größer als $\chi^2_{5\%}$. Die Abweichung der untersuchten Population vom Hardy-Weinberg-Equilibrium deutet auf eine Verletzung der Bedingungen für das Erreichen des Gleichgewichts hin. Bei der Betrachtung der Ergebnisse fällt auf, dass die homozygoten G-Allel-Träger in der Studie geringfügig unterrepräsentiert sind.

Tabelle 9: Darstellung der Verteilung des Genotyps S58P in der Studie und nach Hardy-Weinberg-Gesetz.

Genotyp des SNP S58P	Verteilung des Genotyps	
	in der Studie (in %)	nach Hardy-Weinberg-Gesetz
AA	734 (56,78 %)	749 (56,57 %)
AG	524 (38,52 %)	494 (37,31 %)
GG	66 (4,7 %)	81 (6,11 %)

Als Ursachen sind zunächst Populationsstratifikationen zu nennen. Unter dem Begriff Populationsstratifikation versteht man eine Zusammensetzung der untersuchten Population aus verschiedenen Subpopulationen. Diese unterscheiden sich hinsichtlich der Allelhäufigkeiten oder es liegen ethnische bzw. geografische Gegebenheiten vor, wodurch selektiv wirkende Vor- und Nachteile für Genotypen entstehen. Die hier dargestellte Population besteht aus Patienten der Lipidambulanz des Universitätsklinikums Hamburg-Eppendorf. Diese benötigen einen erhöhten Behandlungsbedarf aufgrund schwer einstellbarer Lipidprofile, weshalb sie in der Regel von externen Arztpraxen an die Universitätsklinik überwiesen wurden. Dadurch ist eine Selektion der Probanden für besonders schwere Dyslipidämien denkbar.

3.4 Die genetische Variante S58P des KLF14-Gens und Dyslipidämien

Zur Untersuchung der Beziehung des S58P-Polymorphismus und der Dyslipidämien wurden die Patienten in fünf Gruppen eingeteilt. Die Gruppen wurden definiert als Hypertriglyzeridämie (Gesamtcholesterin $\leq$ 200 mg/dl, Triglyzeride $>$ 200 mg/dl), gemischte Hyperlipidämie (Gesamtcholesterin $>$ 200 mg/dl, Triglyzeride $>$ 200 mg/dl), Hypercholesterinämie (Gesamtcholesterin $>$ 200 mg/dl, Triglyzeride $\leq$ 200 mg/dl) und Normolipidämie (Gesamtcholesterin $\leq$ 200 mg/dl, Triglyzeride $\leq$ 200 mg/dl). Als letzte

Gruppe wurde gemäß der neuen EAS-Guidelines der Dutch Lipid Clinic Network (Nordestgaard et al., 2013) die familiäre Hypercholesterinämie (FH) mit LDL-Cholesterin ≥ 330 mg/dl definiert. Eine zusammenfassende Darstellung der Verteilung des SNP S58P in den Dyslipidämiegruppen findet sich in Tabelle 10.

Tabelle 10: Darstellung der Verteilung der S58P-Genotypen nach Dyslipidämie. Aufteilung der Studiengruppe nach Gesamtkollektiv sowie Männern und Frauen getrennt.

	n	AA	AG	GG	G-Frequenz
Gesamt					
Hypertriglyzeridämie	57	31	23	3	0,25
Gemischte Hyperlipidämie	369	200	155	14	0,25
Hypercholesterinämie	547	305	214	28	0,25
Normolipidämie	36	20	11	5	0,29
Familiäre Hypercholesterinämie	125	72	45	8	0,24
Männer					
Hypertriglyzeridämie	45	22	20	3	0,29
Gemischte Hyperlipidämie	259	132	114	13	0,27
Hypercholesterinämie	215	126	78	11	0,23
Normolipidämie	73	40	28	5	0,26
Familiäre Hypercholesterinämie	18	11	4	3	0,28
Frauen					
Hypertriglyzeridämie	12	9	3	0	0,54
Gemischte Hyperlipidämie	110	68	41	1	0,19
Hypercholesterinämie	332	179	136	17	0,26
Normolipidämie	52	32	17	3	0,22
Familiäre Hypercholesterinämie	18	9	7	2	0,31

Im Gesamtkollektiv war die G-Allelfrequenz am höchsten in der Normolipidämie-Gruppe (0,29) und etwa gleich häufig in den restlichen Gruppen. Bei Auftrennung des Kollektivs nach Geschlecht ist das G-Allel am häufigsten bei männlichen Patienten mit Hypertriglyzeridämie (0,29) und bei Frauen mit FH vertreten (0,31). Es konnte kein statistisch signifikanter Unterschied der Frequenz zwischen den Dyslipidämie-Gruppen nachgewiesen werden. Die geschlechtsabhängige Auswertung zeigte eine starke Tendenz zu einem häufigeren Auftreten des G-Allels bei weiblichen Patienten mit Hypercholesterinämie im Vergleich zu einer gemischten Hyperlipidämie (0,19 zu 0,26

mit $p = 0,06$). Da KLF14 mehrfach mit HDL-Cholesterin assoziiert wurde, erfolgte eine Untersuchung für dieses Lipoprotein. HDL-Cholesterin wird sowohl geschlechtsabhängig als auch vom Alter mit moduliert, weshalb eine Einteilung nach Perzentilengruppen sinnvoll ist. Es wurden zwei Extrembereiche mit HDL-Werten $< 5.$ Perzentile und oberhalb der 95. Perzentile definiert. Als Normbereich wurde die Gruppe mit HDL-Werten zwischen der 5. und 95. Perzentile festgelegt.

Tabelle 11: Verteilung des S58P-Polymorphismus nach HDL-Perzentilengruppen. Als Extrembereiche wurden Perzentilenwerte $< 5.$ Perzentile und $> 95.$ Perzentile für Alter und Geschlecht definiert. Aufteilung der Studiengruppe nach Gesamtkollektiv, Männern und Frauen.

	n	AA	AG	GG	G-Frequenz
Gesamt					
HDL-Perzentile					
< 5. Perzentile	51	25	23	3	0,28
5.-95. Perzentile	909	503	360	46	0,25
> 95. Perzentile	28	14	10	4	0,32
Männer					
HDL-Perzentile					
< 5. Perzentile	36	18	15	3	0,29
5.-95. Perzentile	492	264	198	30	0,26
> 95. Perzentile	1	1	0	0	0
Frauen					
HDL-Perzentile					
< 5. Perzentile	15	7	8	0	0,27
5.-95. Perzentile	417	239	162	16	0,23
> 95. Perzentile	27	13	10	4	0,33

Das G-Allel ist am häufigsten bei der Patientengruppe mit HDL-Werten oberhalb der 95. Perzentile für Alter und Geschlecht aufgetreten (0,32) und etwa gleich häufig im unteren Extrembereich und im Normbereich (respektive 0,28 und 0,25, siehe Tabelle 11). Es wurden keine signifikanten Unterschiede zwischen den jeweiligen Perzentilengruppen gefunden. Bei separater Analyse der Geschlechter ist das G-Allel bei den Männern in der Gruppe mit HDL-Werten oberhalb der 95. Perzentile gar nicht aufgetreten und bei den Frauen mit einer Frequenz von 0,33 am häufigsten gewesen. Im Gegensatz dazu war der S58P-Polymorphismus bei Männern im Bereich unterhalb der 5. Perzentile am häufigsten (siehe dazu auch Abbildung 8). Jedoch war dieser Zusammenhang statistisch nicht signifikant. Die G-Allelfrequenz zeigte innerhalb der Geschlechtergruppen keine signifikanten Unterschiede. Einzige Ausnahme ist die Tendenz

zu einem häufigeren Auftreten des G-Allels bei Frauen mit HDL-Werten oberhalb der 95. Perzentile gegenüber der Normalgruppe ($p = 0,1$).

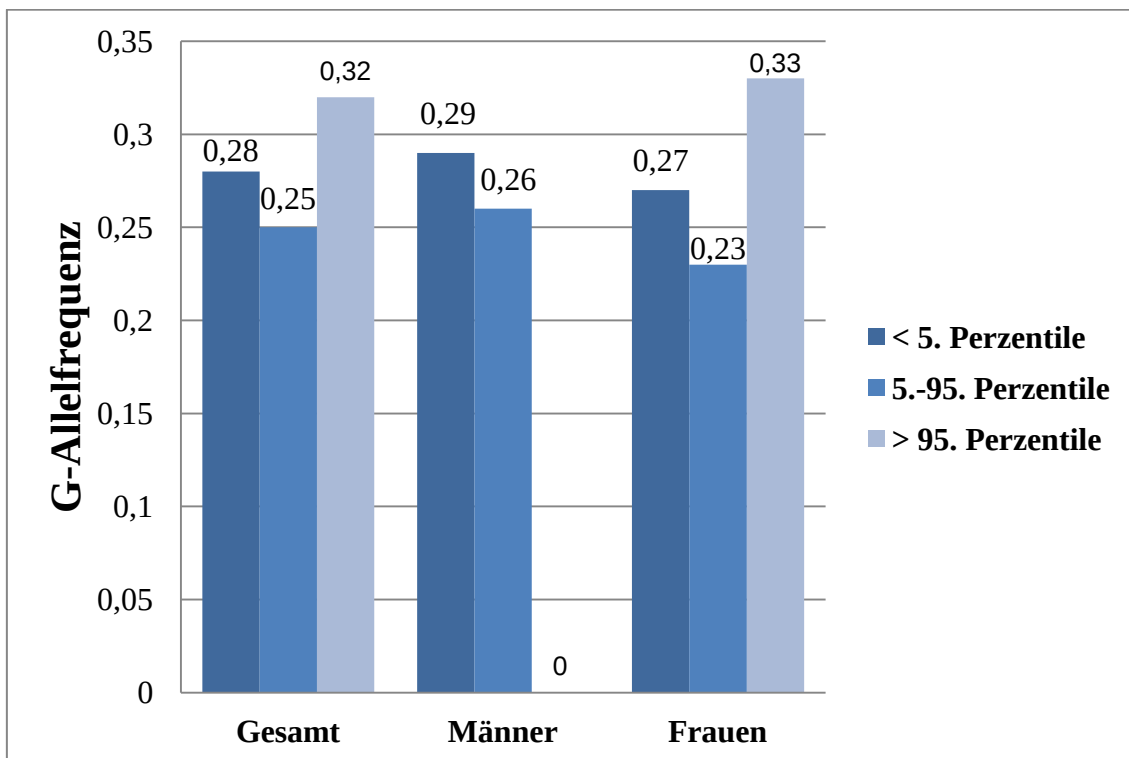


Abbildung 8: Darstellung der G-Allelfrequenz in den HDL-Perzentilengruppen im Gesamtkollektiv sowie getrennt nach Männern und Frauen.

3.5 Zusammenhang zwischen dem SNP S58P und den Serumlipidwerten

Die Untersuchung des S58P-Polymorphismus im Zusammenhang mit den durchschnittlichen Lipidwerten ist in Tabelle 12 für das Gesamtkollektiv sowie als getrennte Betrachtung der Geschlechter aufgeführt. Der durchschnittliche Triglyzeridwert für Träger des untersuchten Polymorphismus liegt bei 276,3 mg/dl. Es wurde kein statistisch signifikanter Unterschied zu Nicht-Trägern mit einem mittleren Triglyzeridwert von 254,6 mg/dl gefunden ($p = 0,49$). Ebenso gab es keine signifikanten Unterschiede für Gesamtcholesterin und LDL-Cholesterin. Die HDL-Werte bei Nicht-Trägern des G-Allels (51,4 mg/dl) waren signifikant höher gegenüber Trägern (49,3 mg/dl) der Variante ($p = 0,03$). Bei Auftrennung der Lipidwerte nach Geschlecht zeigten sich bei den Männern höhere Lipidwerte für Triglyzeride, Gesamtcholesterin und LDL-Cholesterin bei den Trägern des Polymorphismus. Eine Tendenz zu signifikant höheren Werten war jedoch nur bei Gesamtcholesterin zu zeigen mit $p = 0,09$. Außerdem waren die HDL-Werte männlicher Träger signifikant niedriger (41,9 mg/dl zu 44,1 mg/dl, p -Wert = 0,03). Bei den Frauen waren die durchschnittlichen Lipidwerte bei Nicht-Trägern höher bis auf LDL-Cholesterin, welches mit 179,7 mg/dl niedriger war als bei Trägern des SNP S58P mit

185,2 mg/dl. Es konnte kein statistisch signifikanter Unterschied aufgrund des S58P-Polymorphismus bei Frauen gezeigt werden

Tabelle 12: Darstellung der durchschnittlichen Lipidwerte nach Genotyp des S58P-Polymorphismus. Es wurde der Mittelwert sowie die Standardabweichung ($\pm$) angegeben. Fett gedruckte p-Werte sind auf dem 5%-Niveau statistisch signifikant zwischen G-Allelträgern und Nicht-Trägern.

	Genotyp				P-Wert
	AA	AG	GG	AG + GG	
Gesamtkollektiv					
Triglyzeride (mg/dl)	254,6 $\pm$ 425,2	283,0 $\pm$ 515,4	224,6 $\pm$ 256,0	276,3 $\pm$ 493,0	0,49
Gesamtcholesterin (mg/dl)	263,6 $\pm$ 81,8	265,1 $\pm$ 71,1	250,8 $\pm$ 70,5	263,5 $\pm$ 71,1	0,26
LDL (mg/dl)	171,5 $\pm$ 68,5	176,0 $\pm$ 65,5	166,5 $\pm$ 74,1	175,0 $\pm$ 66,5	0,11
HDL (mg/dl)	51,4 $\pm$ 36	49,4 $\pm$ 15,6	48,8 $\pm$ 16,9	49,3 $\pm$ 15,7	0,03
Männer					
Triglyzeride (mg/dl)	287 $\pm$ 351,2	350,3 $\pm$ 535,9	289,6 $\pm$ 311,3	342,8 $\pm$ 513,4	0,11
Gesamtcholesterin (mg/dl)	249,6 $\pm$ 68,3	255,5 $\pm$ 69,6	249,2 $\pm$ 72,1	254,7 $\pm$ 69,8	0,09
LDL (mg/dl)	163,5 $\pm$ 66,3	169,9 $\pm$ 59,4	167,47 $\pm$ 80,5	165,24 $\pm$ 62,3	0,24
HDL (mg/dl)	44,1 $\pm$ 11,1	42,3 $\pm$ 10,0	39,5 $\pm$ 10,7	41,9 $\pm$ 10,1	0,03
Frauen					
Triglyzeride (mg/dl)	218,4 $\pm$ 493,2	201,7 $\pm$ 478,4	125,7 $\pm$ 55,7	194,0 $\pm$ 454,4	0,23
Gesamtcholesterin (mg/dl)	279,2 $\pm$ 92,3	276,7 $\pm$ 71,4	253,3 $\pm$ 69,7	274,3 $\pm$ 71,4	0,96
LDL (mg/dl)	179,7 $\pm$ 69,9	187,5 $\pm$ 69,5	165,2 $\pm$ 66,7	185,2 $\pm$ 69,4	0,25
HDL (mg/dl)	59,6 $\pm$ 16,9	57,9 $\pm$ 16,8	63 $\pm$ 14,6	58,4 $\pm$ 16,6	0,65

3.6 Zusammenhang zwischen der genetischen Variante S58P und den Parametern Adipositas und Diabetes mellitus Typ 2 als Komponenten des metabolischen Syndroms

Zur Untersuchung des Parameters Adipositas wurden die Patienten anhand des BMI in eine Adipositasgruppe und eine Nicht-Adipositasgruppe eingeteilt. In Anlehnung an die Leitlinie der Deutschen Adipositas-Gesellschaft (Deutsche Adipositas-Gesellschaft, 2014) wurde ein BMI ≥ 30 kg/m² als manifeste Adipositas definiert. In der nicht adipösen Gruppe sind 954 Patienten mit einer G-Allelfrequenz von 0,26 (AA 518, AG 382, GG 54). Diese ist im Vergleich zu 0,23 (AA 148, AG 98, GG 9) in der

Adipositasgruppe höher, jedoch nicht signifikant (siehe Abbildung 9). Bei getrennter Betrachtung der Geschlechter ist das G-Allel mit 0,26 (AA 84, AG 69, GG 7) in der männlichen Adipositasgruppe gleich der Frequenz bei Männern mit einem BMI <30kg/m² (AA 270, AG 203, GG 31). Das G-Allel war bei nicht erkrankten Frauen mit 0,25 (AA 248, AG 179, GG 23) zu 0,17 (AA 64, AG 29, GG 2) bei Adipositas-Patientinnen signifikant häufiger aufgetreten (p = 0,02). Es erfolgte ebenfalls eine Auswertung des mittleren BMI-Wertes für Träger und Nicht-Träger der Variante, inklusive geschlechtsspezifischer Aufschlüsselung (Tabelle 13).

Tabelle 13: Darstellung des mittleren BMI (kg/m²) für den S58P-Polymorphismus im Gesamtkollektiv und separat für Männer und Frauen.

	AA	AG	GG	AG + GG	p-Wert
Gesamtkollektiv					
BMI (kg/m²)	27,2±4,7	26,8±4,0	26,1±3,8	26,7±4,0	0,26
Männer					
BMI (kg/m²)	27,8±4,1	27,7±3,7	27,1±3,4	27,6±3,6	0,96
Frauen					
BMI (kg/m²)	26,4±5,2	25,5±4,1	24,5±3,8	25,4±4,1	0,08

Im Gesamtkollektiv war der BMI-Mittelwert für Träger des G-Allels niedriger (26,7kg/m²) als für Nicht-Träger (27,2 kg/m²), jedoch nicht signifikant. In der männlichen Untergruppe variierten die mittleren BMI-Werte nicht. Es zeigte sich eine Tendenz zu einem niedrigeren BMI für Träger des Polymorphismus bei weiblichen Patienten (25,4 kg/m² zu 26,4 kg/m², p-Wert = 0,08). Dies passt zu der signifikant höheren G-Allelfrequenz bei nicht-adipösen Frauen mit einem BMI < 30 kg/m². Die Spannweite der BMI-Werte für die einzelnen Genotypen ist in Abbildung 11 dargestellt.

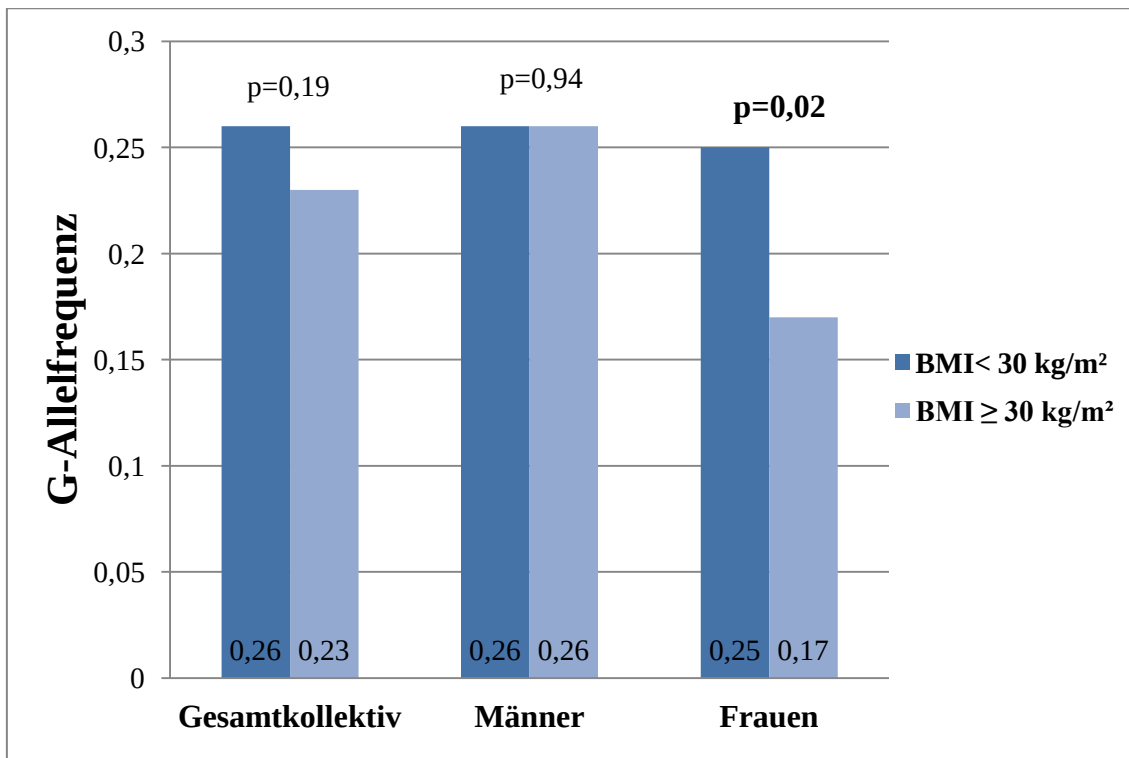


Abbildung 9: Darstellung der G-Allelfrequenz nach BMI-Gruppe. Ein BMI $\geq 30 \text{ kg/m}^2$ gilt als manifeste Adipositas. Die Studiengruppe wurde links als Gesamtkollektiv betrachtet. In der Mitte sind Männer und rechts Frauen getrennt dargestellt. Signifikante p-Werte auf dem 5%-Niveau sind fett gedruckt hervorgehoben.

Als weiterer Parameter des metabolischen Syndroms wurde der Zusammenhang des SNP S58P mit einer Erkrankung an Diabetes mellitus Typ 2 untersucht. Die Patienten der Lipidambulanz wurden nach den WHO-Kriterien diagnostiziert (Alberti u. Zimmet, 1998). Aus dem Patientenkollektiv konnten 155 Probanden (AA 91, AG 58, GG 6) mit Diabetes mellitus Typ 2 klassifiziert werden, bei 768 Probanden (AA 418, AG 305, GG 45) konnte eine Erkrankung ausgeschlossen werden. Die restlichen Patienten konnten nicht eindeutig zugeordnet werden und wurden deshalb aus der Analyse ausgeschlossen. Generell war das G-Allel bei nicht erkrankten Individuen häufiger, auch bei getrennter Betrachtung von Männern und Frauen. Es zeigte sich jedoch kein statistisch signifikanter Unterschied aufgrund der Diabetes-Erkrankung (siehe Abbildung 10). Die Untergruppe der Männer zeigte ähnliche Frequenzen des G-Allels wie das Gesamtkollektiv. Die Frequenz des G-Allels ist bei gesunden Männer mit einer Häufigkeit von 0,27 (AA 219, AG 161, GG 28) höher als bei an Diabetes erkrankten mit 0,24 (AA 50, AG 38, GG 3). In der weiblichen Untergruppe war die Differenz der Frequenz zwischen an Diabetes erkrankten Frauen mit 0,2 (AA 41, AG 20, GG 3) und gesunden Patientinnen mit 0,25 (AA 199, AG 144, GG 17) am größten, jedoch nicht signifikant. Es konnten keine

statistisch signifikanten Unterschiede zwischen den Geschlechtern nachgewiesen werden.

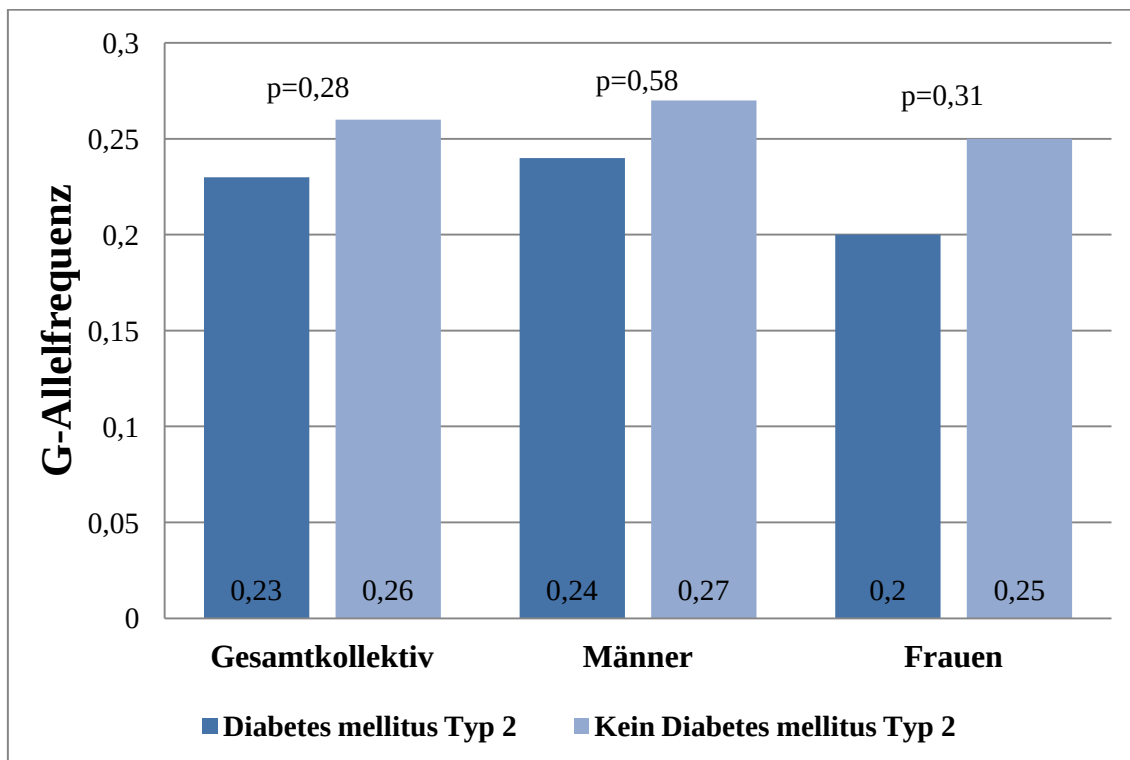
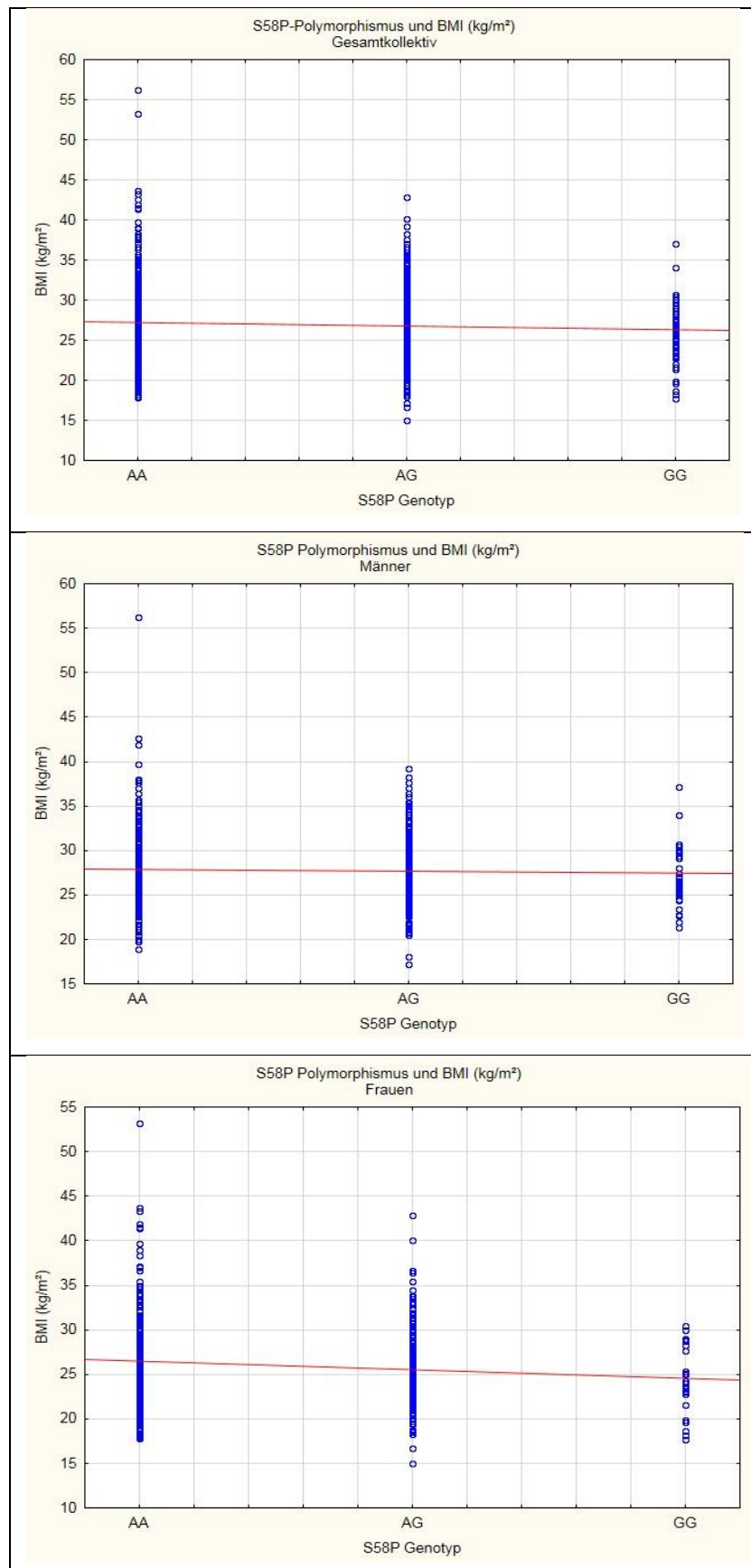


Abbildung 10: Darstellung der Frequenz des G-Allels des S58P-Polymorphismus und Diabetes mellitus Typ 2. Dargestellt sind die Frequenzen im Gesamtkollektiv sowie in der männlichen und weiblichen Untergruppe.

Abbildung 11: Darstellung der Genotypen des SNP S58P im KLF14-Gen und BMI (kg/m^2) im Gesamtkollektiv (oben), Männer (Mitte) und Frauen (unten). Ein Kreis ist jeweils ein Patientenfall.



3.7 Kopplungsungleichgewicht für die SNPs S58P, P47Q und A39A

Für die Darstellung des Kopplungsungleichgewichts der SNPs S58P, P47Q und A39A wurden insgesamt 15 Proben sequenziert. Die Proben wurden anhand des Genotyps für den S58P-Polymorphismus ausgewählt. Jeweils fünf Proben entsprechend den Genotypen AA, AG und GG wurden sequenziert. Nachfolgend sind in Abbildung 12 bis Abbildung 14 die Sequenzmuster von drei Patienten dargestellt. Es war jeweils der häufig homozygote Genotyp AA des SNP S58P mit dem homozygoten Genotyp TT (SNP P47Q) und TT im SNP A39A kombiniert. Analog dazu lag der homozygote GG-Genotyp im SNP S58P mit den homozygoten Genotypen GG bzw. AA im SNP P47Q und A39A vor. Patienten mit dem heterozygoten Genotyp AG des SNP S58P zeigten auch in den anderen untersuchten SNPs einen heterozygoten Genotyp mit den Allelen TG (SNP P47Q) und TA (SNP A39A). Die Ergebnisse sind in Tabelle 14 zusammengefasst.

Tabelle 14: Darstellung des Kopplungsungleichgewichts der SNPs S58P, P47Q, A39A.

Serennummer	SNP S58P	SNP P47Q	SNP A39A
12103	AA	TT	TT
18760	AA	TT	TT
18943	AA	TT	TT
18971	AA	TT	TT
17084	AA	TT	TT
18698	AG	TG	TA
18734	AG	TG	TA
12045	AG	TG	TA
12378	AG	TG	TA
13716	AG	TG	TA
12578	GG	GG	AA
14522	GG	GG	AA
12380	GG	GG	AA
9484	GG	GG	AA
10044	GG	GG	AA

Unter Verwendung der Formel 4 ergibt sich für die Kopplungsanalyse von S58P und P47Q:

$$D = (0,5 \times 0,5) - 0 = 0,25$$

Für die Kopplungsanalyse von S58P mit A39A berechnet sich D wie folgt:

$$D = (0,5 \times 0,5) - 0 = 0,25$$

Der Disequilibriumskoeffizient D für die SNPs P47Q und A39A berechnet sich ebenfalls als:

$$D = (0,5 \times 0,5) - 0 = 0,25.$$

Somit ist für alle drei untersuchten Polymorphismen eine vollständige Kopplung nachgewiesen.

Abbildung 12: Darstellung der Sequenzierung der Serenprobe 18943 für die SNPs S58P, A39A und P47Q auf dem reversen Strang der DNA. Es wurden in Klammern jeweils die Bereiche markiert, in denen die Varianten zu finden sind. In den vergrößerten Bildausschnitten wurden die Polymorphismen umkreist.

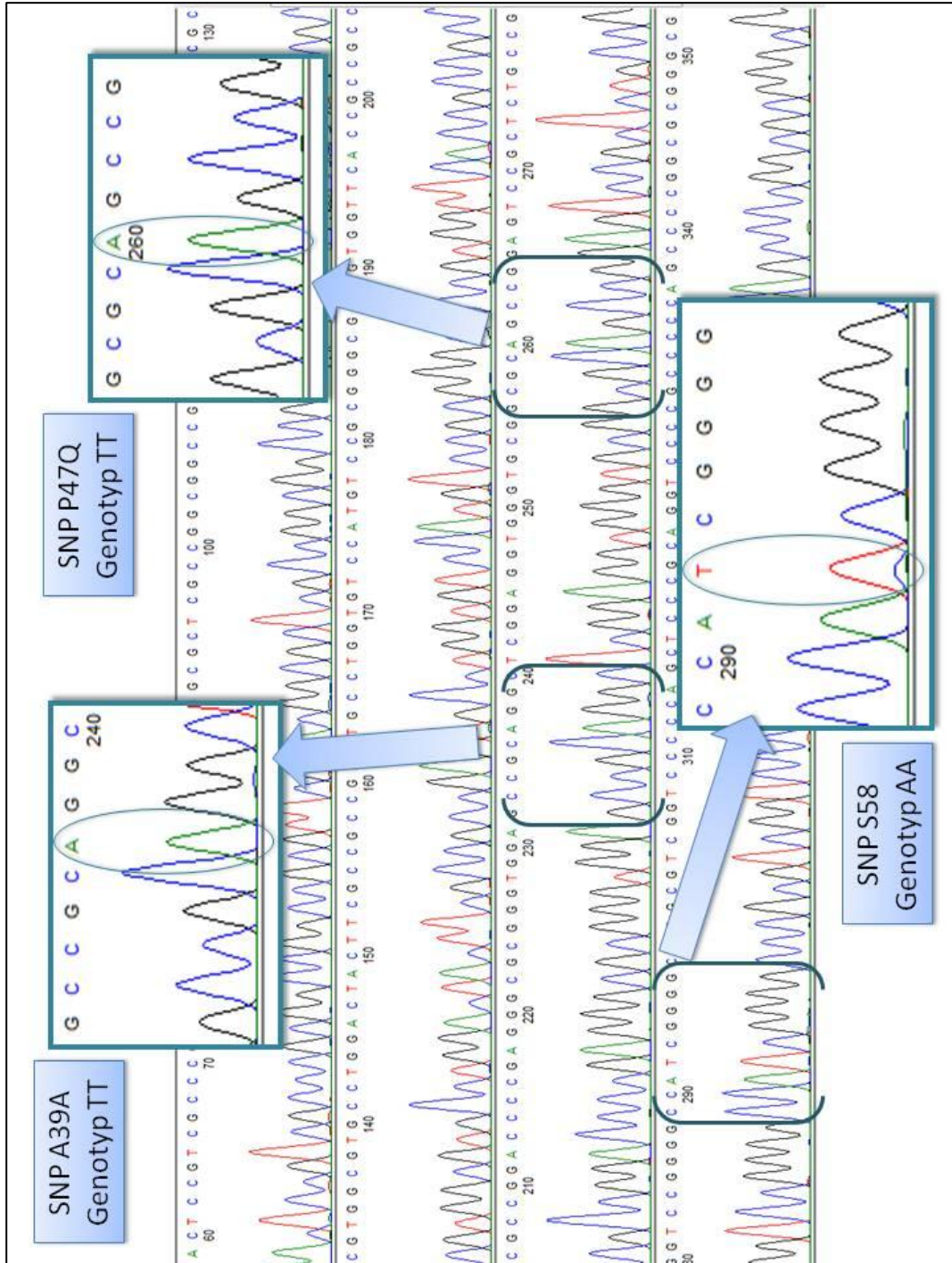
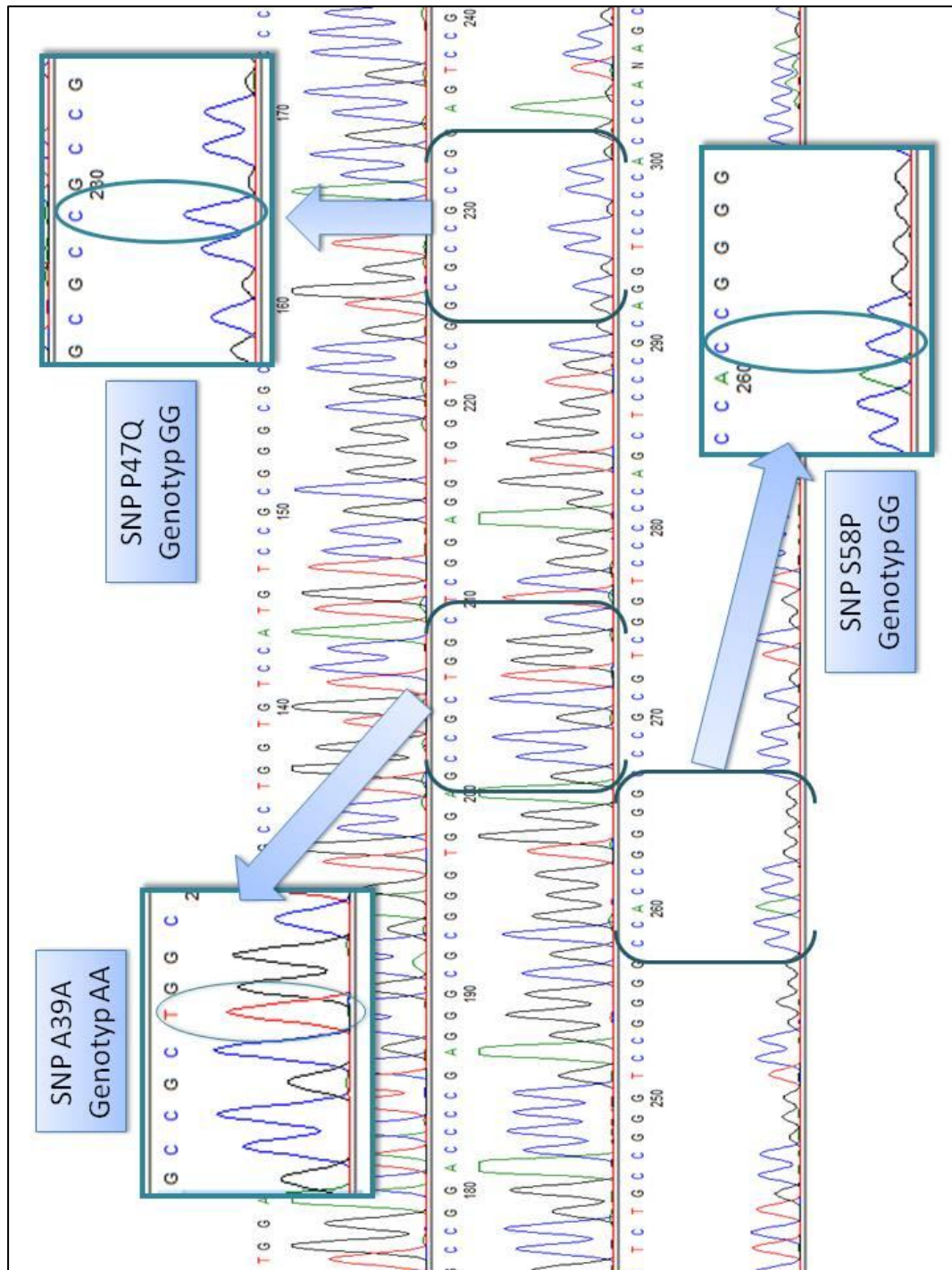


Abbildung 14: Darstellung der Sequenzierung der Serenprobe 12380 für die SNPs S58P, A39A und P47Q auf dem reversen Strang der DNA. Es wurden in Klammern jeweils die Bereiche markiert, in denen die Varianten zu finden sind. In den vergrößerten Bildausschnitten wurden die Polymorphismen umkreist.



3.8 Auffälliges Bandenmuster in der Genotypisierung

Bei der Durchführung der Genotypisierung fand sich eine auffällige Probe. Diese unterscheidet sich von den anderen Proben durch eine doppelte Bande in dem Elektropherogramm (siehe Abbildung 15). Die DNA-Serenprobe wurde anschließend sequenziert und eine Deletionsmutation im kodierenden Bereich des Gens KLF14 im Sinne einer seltenen Variante (rare variant) identifiziert.

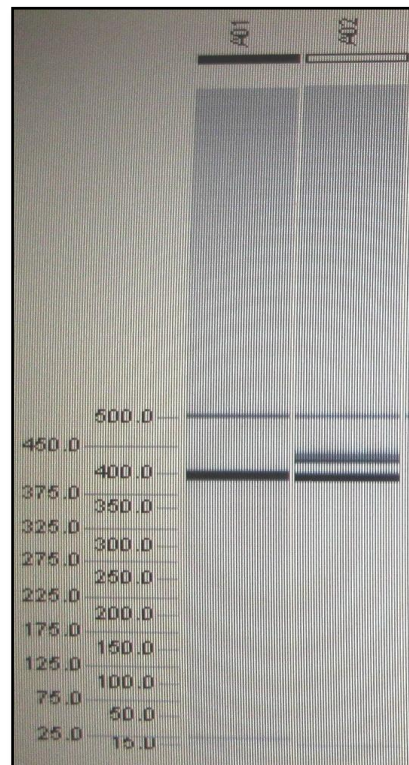


Abbildung 15: Auffälliges Bandenmuster mit doppelter Bande im Elektropherogramm des Patienten mit der Seriennummer 11151 in der Linie A02.

Der deletierte Abschnitt liest sich wie folgt:

3'-ATCGGGGCC-5'

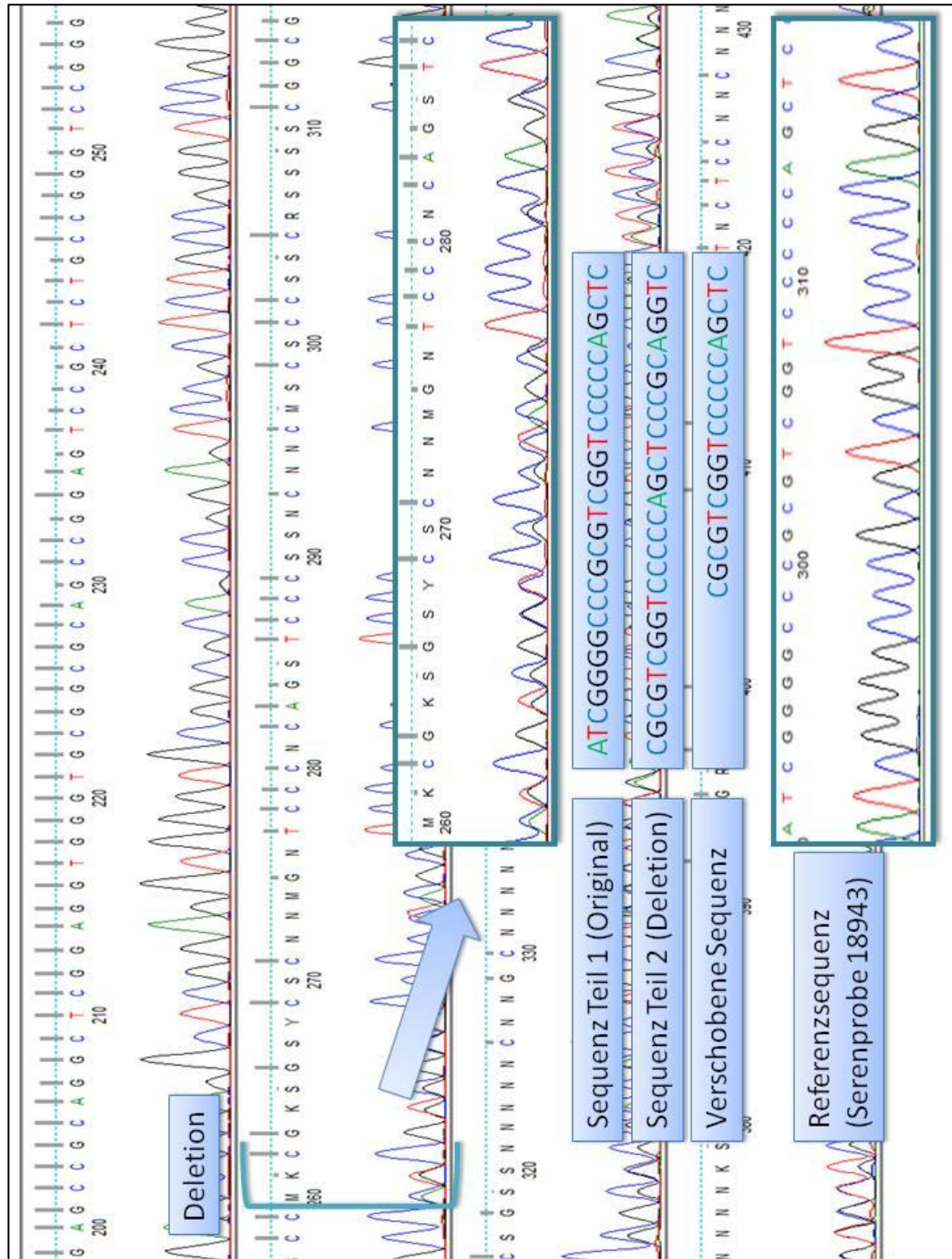
Die Deletion liegt im Codonbereich 55 bis 58. Die kodierte Aminosäure des Codon 55 bleibt unverändert. Die durch die Codons 56 bis 58 verschlüsselten Aminosäuren Serin, Glycin und Prolin gehen hingegen verloren. Da es sich bei den deletierten Basen um ein Vielfaches von 3 handelt, wird das Leseraster nach der Deletion nicht verändert. Jedoch ist die kodierte Proteinsequenz dadurch um drei Aminosäuren verkürzt. Ein Schema der Deletionsmutation ist in Abbildung 16 und das Ergebnis der Sequenzierung in Abbildung 17 dargestellt.

Codon	53	54	55	56	57	58	59	60	
Original-sequenz	CCG	GGG	CCA	TCG	GGG	CCC	GCG	TCG	
Aminosäuren-sequenz	Pro	Gly	Pro	Ser	Gly	Pro	Ala	Ser	
Deletion	CCG	GGG	CC				C	GCG	TCG
Neues Leseraster	CCG	GGG	CCC	GCG	TCG	GTC	CCC	CAG	
Neue Protein-sequenz	Pro	Gly	Pro	Ala	Ser	Val	Pro	Gln	

Abbildung 16: Schema der Deletionsmutation.

Die seltene Variante liegt im Bereich des untersuchten SNP S58P, wodurch dieser entfällt und nicht in die Untersuchung des SNP S58P eingeht. Bei den Laborwerten zeigten sich erhöhte Triglyzerid- und Gesamtcholesterinwerte im Sinne einer gemischten Hyperlipidämie. Das HDL-Cholesterin war mit 73 mg/dl im Bereich der 90.-95. Perzentile für Alter und Geschlecht. Der BMI war mit 19 kg/m² unauffällig. Eine statistische Auswertung ist durch das einmalige Vorliegen der Variante nicht sinnvoll. Da durch die Deletion die Aminosäure Prolin verloren geht, kann eine funktionelle Bedeutung vermutet werden. Prolin-Motive sind wichtig zur regelrechten Proteinfaltung und damit auch für die Funktion der entstandenen Proteine.

Abbildung 17: Darstellung der Sequenzierung der Serenprobe 11151. Der Bereich der Deletion ist hervorgehoben sowie als Referenzsequenz die Serenprobe 18943 abgebildet.



4 Diskussion

4.1 Einordnung der Ergebnisse in die aktuelle Literatur

In der von Teslovich veröffentlichten Studie zu 95 Genloci und deren Einfluss auf die Blutlipidwerte (Teslovich et al., 2010) wurde eine Assoziation zwischen dem Genlocus KLF14 und HDL-Cholesterin mit einer Effektstärke von +0,59 mg/dl für das seltenere Allel des SNP rs4731702 publiziert. Für diesen SNP konnten Chasman et al. (2009) ähnliche Ergebnisse in ihrer genomweiten Studie mit signifikant erhöhten Werten für HDL-Cholesterin und erniedrigten Triglyzerid- und LDL-Cholesterinwerten zeigen (Chasman et al., 2009). Weiterhin wird vermutet, dass das KLF14-Gen bei der Pathogenese des metabolischen Syndroms von Bedeutung ist (Civelek u. Lusis, 2011). Auf Grundlage der veröffentlichten Ergebnisse in den verschiedenen GWAS wurde die Fragestellung der vorliegenden Arbeit generiert. Als Hypothese wird ein Einfluss des SNP S58P in der Pathogenese von Dyslipidämien sowie auf die Komponenten Adipositas und gestörte Insulinempfindlichkeit bzw. Diabetes mellitus Typ 2 des metabolischen Syndroms vermutet.

Für die genetische Variante S58P wurde in unserer Studie eine Frequenz des G-Allels von 0,25 erhoben. Diese Frequenz ist ähnlich wie die im Rahmen des 1000 Genome-Projektes publizierte Frequenz von 0,26 in der dbSNP Datenbank (Sherry et al., 2001), jedoch größer als 0,22 in der von Chen et al. untersuchten chinesischen Population (Chen et al., 2012).

Die Resultate dieser Arbeit weichen teilweise von den bereits publizierten Ergebnissen zum Genlocus KLF14 ab. Träger des Polymorphismus zeigen im Gesamtkollektiv und in der Untergruppe der Männer signifikant niedrigere HDL-Cholesterinwerte. Diese Ergebnisse stimmen mit den bisher veröffentlichten Ergebnissen, bei denen Träger genetischer Varianten im KLF14-Gen höhere HDL-Cholesterinwerte besitzen, nicht überein. Als mögliche Faktoren, welche zu abweichenden Ergebnissen führen, sind zum einen HDL-Cholesterin beeinflussende Faktoren im Patientenkollektiv überprüft als auch zum anderen statistische Überlegungen angestellt worden. Komorbiditäten wie Diabetes mellitus können modulierend auf den Lipidstoffwechsel wirken und zu einem niedrigeren HDL-Cholesterinlevel beitragen (Kreisberg, 1998). Daneben wirken individuelle Umweltfaktoren wie das Rauchverhalten auf den HDL-Cholesterinspiegel. Eine Untersuchung von 1504 aktiven Rauchern zeigte, dass die HDL-Cholesterinwerte ein Jahr nach Beenden des Rauchens signifikant um 2,4 mg/dl angestiegen sind (Gepner et al., 2011). In der hier untersuchten Population haben 11,7 % der Individuen einen

Diabetes mellitus Typ 2 und 57,7 % eine positive Raucheranamnese. Eine erneute statistische Auswertung der mittleren HDL-Cholesterinwerte, unter Ausschluss von Patienten mit Diabetes mellitus Typ 2 und einer positiven Raucheranamnese, wurde durchgeführt. Der mittlere HDL-Wert für das Gesamtkollektiv beträgt 50,86 mg/dl. Träger des G-Allels der genetischen Variante S58P haben mit 49,8 mg/dl einen niedrigeren HDL-Cholesterinwert als Nicht-Träger (51,7 mg/dl). Bei Männern und Frauen haben ebenfalls G-Allel-Träger geringere HDL-Cholesterinwerte (42,0 mg/dl und 59,1 mg/dl zu 43,6 mg/dl und 60,4 mg/dl). Die Ergebnisse sind nicht signifikant. Bei weiteren statistischen Überlegungen zum Ergebnis wurde berücksichtigt, dass es sich beim Mittelwert um ein für Extremwerte anfälliges Lagemaß handelt. Als Alternative zum arithmetischen Mittelwert bietet sich der Median als robuste statistische Größe an (Weiß, 2013). Eine zweite Auswertung der HDL-Cholesterinwerte anhand des Median wurde daraufhin durchgeführt. Der Median im Gesamtkollektiv beträgt 48 mg/dl. Die Patienten wurden dann in zwei Gruppen unterhalb und oberhalb des Median eingeteilt. Eine niedrigere Frequenz des G-Allels oberhalb des HDL-Median, 0,26 (AA 315, AG 245, GG 34) zu 0,23 (AA 312, AG 206, GG 24), ist auch hier tendenziell signifikant ($p = 0,1$). Bei alleiniger Betrachtung der Männer liegt der Median für den HDL-Cholesterinwert bei 42 mg/dl. Die Tendenz ($p = 0,1$) zu einer niedrigeren G-Frequenz oberhalb des Median mit 0,24 (AA 169, AG 114, GG 13) zu 0,28 (AA 164, AG 132, GG 22) ist auch hier zu sehen. Frauen zeigen bei einem HDL-Median von 56 mg/dl eine erhöhte G-Allelfrequenz von 0,25 (AA 143, AG 98, GG 14) oberhalb des Median zu 0,23 (AA 151, AG 107, GG 9), jedoch nicht signifikant ($p=0,6$).

Abschließend ist zu den Ergebnissen der Serumlipidwerte zu bemerken, dass es sich bei den GWA-Studien um ein heterogenes Kollektiv handelt, welches vornehmlich gesunde Individuen oder Patienten mit Phänotypen einer kardiovaskulären Erkrankung einschließt. Im Gegensatz dazu handelt es sich in der vorliegenden Arbeit um eine dyslipidämische Population. In dieser trifft eine Vielzahl von genetischen Faktoren aufeinander, sodass ein von den GWA-Studien abweichendes Ergebnis durchaus möglich ist. Eine vermutlich geringe Effektstärke des S58P-Polymorphismus kann somit durch die Kombination anderer häufigerer genetischer Varianten überlagert werden.

Lipidwerte verändern sich nicht nur in Abhängigkeit von genetischen Faktoren, sondern werden auch durch nicht beeinflussbare Komponenten wie Alter und Geschlecht be-

stimmt (Henze et al., 1981, National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III), 2002). Deshalb erfolgte eine nach Perzentilen geordnete Auswertung der HDL-Cholesterinwerte. Im Gesamtkollektiv und für die Untergruppe der Männer zeigten sich keine signifikanten Unterschiede innerhalb der Perzentilengruppen. Bei Frauen mit HDL-Cholesterinwerten oberhalb der 95. Perzentile konnte die Tendenz ($p = 0,1$) zu einem häufigeren Auftreten der genetischen Variante S58P im Vergleich zu dem definierten Kollektiv im Normbereich (5.-95. Perzentile) festgestellt werden. In der vorliegenden Studie zeichnete sich außerdem für Frauen eine starke Tendenz ($p = 0,06$) zu einer erhöhten G-Allelfrequenz bei Hypercholesterinämie im Vergleich zu einer gemischter Hyperlipidämie ab. Diese ist neben hohen Cholesterinwerten durch ebenfalls erhöhte Triglyzeridwerte definiert. Eine tendenzielle Signifikanz zu erhöhten LDL-Cholesterinwerten und - bei männlichen Individuen zudem - Triglyzeridwerten war zu sehen. Diese Trends passen zu den Ergebnissen von Teslovich et al. (2010) und Chasman et al (2009). Letzterer wies neben der Assoziation des KLF14-Locus zu HDL-Cholesterin ebenfalls eine Assoziation mit Triglyzeriden und LDL-Cholesterin nach (Chasman et al., 2009).

Im Hinblick auf das metabolische Syndrom sind die tendenziell höheren HDL-Perzentilen weiblicher Träger des SNP S58P schlüssig mit den Ergebnissen zu der Variablen Adipositas. Das G-Allel in der Nicht-Adipositas-Gruppe war signifikant erhöht und stimmt mit der starken Tendenz zu niedrigeren BMI-Mittelwerten bei Frauen mit dem S58P-Polymorphismus überein. Außerdem wurde die Komponente Diabetes mellitus untersucht. In verschiedenen Studien wurde bereits ein Zusammenhang zwischen dem Transkriptionsfaktor KLF14 und einer Diabetes-mellitus-Erkrankung publiziert. Der SNP rs972283 zeigte in einer Meta-Studie von GWAS mit Individuen europäischer Abstammung ein assoziiertes Risiko für höhere Nüchterninsulinwerte, welches vermutlich durch einen primären Effekt auf die Insulinwirkung bedingt ist (Voight et al., 2010). Außerdem zeigte eine weltweite Meta-Analyse von fünf GWA-Studien, welche zwischen 2008 und 2013 veröffentlicht wurden, ein konstant erhöhtes signifikantes relatives Risiko von 4,18 % für das Risikoallel G des SNP rs972283 (Wang et al., 2014). In der hier untersuchten Population konnte sich keine Assoziation mit Diabetes mellitus nachweisen lassen, welche die bisherigen Daten unterstützt oder widerlegt.

Ein geschlechtsspezifischer Effekt kann bei signifikant erniedrigten HDL-Cholesterinwerten männlicher Individuen und bei einer Tendenz zu erhöhten HDL-

Perzentilen weiblicher G-Allel-Träger in derselben Studienpopulation vermutet werden. Bereits für den häufiger untersuchten SNP rs4731702 wurde eine geschlechtsspezifische und/oder ethnische Assoziation zum Lipidstoffwechsel vermutet. Bei Mulao-Chinesen besitzen die Träger des T-Allels des SNP rs4731702 höhere LDL-Cholesterin- und ApoA1-Werte und männliche Mulao-Chinesen haben zusätzlich erhöhte HDL-Cholesterinwerte im Vergleich zu Trägern des C-Allels. Im Gegensatz dazu haben bei der ethnischen Gruppe der Han-Chinesen T-Allel-Träger erniedrigte HDL-Cholesterinwerte sowie ein niedrigeres ApoA1-zu-ApoB-Verhältnis. Dabei zeigten männliche Han-Chinesen als Träger des T-Allels ein erniedrigtes ApoA1-zu-ApoB-Verhältnis und ApoA1-Werte, wohingegen weibliche Träger erhöhte Triglyzerid-, LDL-Cholesterin-, ApoA1- und ApoB-Level hatten (Huang et al., 2013).

Weiterhin wurde für die Varianten S58P, P47Q und A39A ein mögliches Kopplungsungleichgewicht untersucht. Die gemeinsame Vererbung von Sequenzvarianten ist stark von der räumlichen Nähe der betrachteten Loci zueinander auf dem Chromosom begünstigt. Bei den untersuchten Polymorphismen handelt es sich um drei SNPs, welche im selben Genabschnitt liegen und 23 bzw. 31 Basenpaare voneinander entfernt sind. Das häufigere gemeinsame Auftreten von SNPs auf einem Chromosom wird auch zur Erstellung von Kopplungskarten genutzt. Dabei kann ein einziger SNP als sogenannter „tagging SNP“ genotypisiert werden, welcher mit weiteren für die Studie interessanten SNPs einen hohen LD-Koeffizienten aufweist. Somit können Kosten für aufwändige Genotypisierungen eingespart werden. Der in der hier durchgeführten Studie berechnete LD-Koeffizient von jeweils $D = 0,25$ zeigt eine vollständige Kopplung der drei SNPs, wodurch sie sich als tagging-SNPs eignen. Durch die vollständige Kopplung ist eine weiterführende Untersuchung der Varianten P47Q und A39A nicht sinnvoll.

4.2 Bewertung des Studienkollektivs

Zur Bearbeitung der Fragestellung wurden insgesamt 1325 Patienten aus der Lipidambulanz des Universitätsklinikums Hamburg-Eppendorf genotypisiert. Ein Patient wurde aus der statistischen Auswertung ausgeschlossen, da er eine seltene Deletionsmutation im Bereich des S58P-Polymorphismus aufwies. Es konnte auf eine bereits vorhandene Datenbank mit Lipidwerten und klinischen Daten von Patienten der universitären Lipidambulanz zurückgegriffen werden, sodass es möglich war, eine große Fallzahl in die Studie einzuschließen. Auf der anderen Seite ist zu berücksichtigen, dass es sich bei der Universitätsklinik nicht um die primäre Anlaufstelle zur Behandlung von Dyslipidämien handelt. Komplikationslose Krankheitsverläufe mit gutem The-

rapieansprechen werden vorrangig in niedergelassenen Arztpraxen behandelt und nur selten in die Lipidambulanz der Universitätsklinik überwiesen. Somit handelt es sich bei dem Patientenkollektiv der Datenbank vor allem um schwerwiegende Erkrankungen des Lipidmetabolismus. Dadurch kann es zu Selektionsmechanismen bei der Auswahl der Patienten für diese Studie kommen, was eine geringere Repräsentativität der Population im Vergleich zu einer zufällig ausgewählten Kohorte aus der Gesamtbevölkerung zur Folge hat.

Für die Berechnung des HWG (Hardy-Weinberg-Gleichgewicht) werden die beobachteten Genotypfrequenzen mit den in einer idealen Population zu erwartenden Genotyphäufigkeiten verglichen. Im untersuchten Patientenkollektiv wich die Verteilung der Genotypen geringfügig vom HWG ab. Ein systematischer Effekt in der Genotypverteilung sollte hier vermutet werden. Als mögliche Ursachen kommen Typisierungsfehler, Mischungen von Subpopulationen oder andere Verletzungen der Voraussetzungen in Frage. Fehler in der Genotypisierung können durch die sorgfältigen und präzisen Verfahren in unserer Studie ausgeschlossen werden. In einer nicht isolierten Population, wie sie hier vorliegt, können neue Allele durch Migration eingetragen werden und in einer Verteilung der Genotypen, die nicht dem HWG gerecht werden, resultieren. Als weitere Möglichkeit ist die Mischung verschiedener Subpopulationen mit einem selektiven Vor- bzw. Nachteil für bestimmte Genotypen und deshalb unterschiedlichen Allelhäufigkeiten zu berücksichtigen. Ein Genotyp, der zu einem schlechteren Lipidprofil führt, ist ein selektiver Nachteil für die Population. Dadurch steigt das arteriosklerotische Risiko und die damit verbundene Möglichkeit, einen Myokardinfarkt oder Schlaganfall zu erleiden, innerhalb der Gruppe. Unterschiedliche Genfrequenzen bei Männern und Frauen führen ebenfalls zur Missachtung der Voraussetzungen für ein HWG. Diese sind in der Studienpopulation in etwa gleich groß und können als Fehlerquelle ausgeschlossen werden. Die Stichprobe wurde nicht aus der Gesamtheit aller an Dyslipidämien erkrankter Patienten erhoben, sondern aus einem mit einem schwer therapierbaren Lipidstatus vorhandenen Kollektiv an der Universitätsklinik. Dadurch ist eine ungewollte Selektion im Patientenkollektiv am wahrscheinlichsten.

4.3 Möglichkeiten zur Bewertung der Relevanz der Variante S58P und der identifizierten Deletion

In der Untersuchung genetischer Varianten des Erbguts ist deren funktionelle Relevanz eine der entscheidenden Fragen. Der Vielzahl an identifizierten Varianten ihre Funktion zuzuordnen und ihren Anteil an der Entwicklung eines Phänotyps zu entschlüsseln, wird

eine der Hauptaufgaben in der Zukunft. Nur ein geringer Teil der assoziierten Varianten wird letztlich einen Einfluss auf die Funktion eines Proteins haben (Manolio et al., 2009). Einer Schätzung zu Folge sind 20 % der Varianten schädlich, weitere 53 % besitzen moderate schädliche Effekte und immerhin 20% der Varianten verhalten sich neutral (Kryukov et al., 2007). Die Auswirkungen des Austausches von Aminosäuren auf Veränderungen in der Proteinstruktur und damit einhergehende Funktionsveränderungen müssten für jede Variante mit in-vitro-Experimenten erforscht werden. Die in-silico-Analyse mit der PolyPhen (Polymorphism-Phenotyping)-Software der Harvard-Universität, Boston, ist eine preisgünstige und schnelle Möglichkeit, um eine fragliche funktionelle Auswirkung im menschlichen Organismus beurteilen zu können (Adzhubei et al., 2010). Die Einstufung der Varianten erfolgt als gutartig, möglicherweise schädlich oder wahrscheinlich schädlich. Der S58P-Polymorphismus wird als funktionell gutartig eingestuft. Dies wird durch die bei Chen et al. (2012) identifizierte Assoziation zwischen S58P und dem Risiko, einen Myokardinfarkt zu erleiden, unterstützt. Das relative Risiko ist um 0,84 pro Allel für Träger des G-Allels reduziert (Chen et al., 2012). In theoretischen Überlegungen zu der S58P-Variante und der Deletionsmutation ist der Verlust der protektiven Aminosäure Prolin zu bedenken. Für die korrekte Konformation der Proteine ist die Anwesenheit von Prolin in der Aminosäuresequenz entscheidend. Prolin-Motive sind auch für strukturelle und biologische Funktionen von Proteinen des Immunsystems oder der Gerinnungskaskade wichtig. Die neuroendokrine oder vasoaktive Funktion verschiedener Hormone wie Bradykinin und Angiotensin wird ebenfalls durch prolinreiche Strukturen bestimmt (Vanhoof et al., 1995).

4.4 Ausblick und neue Ansätze der Forschung

Durch GWAS und Resequenzierungsstudien werden neue Informationen gewonnen, welche zum besseren Verständnis der Physiologie und Pathophysiologie von Dyslipidämien im Sinne einer komplexen Krankheit nützlich sind. Das Wissen um die Effekte genetischer Veränderungen bei der Entstehung von Erkrankungen des Lipidmetabolismus eröffnet neue Möglichkeiten in der Diagnostik und Therapie und führt so zu neuen Medikamenten oder Screeningmethoden. HDL-Cholesterin wird seit Längerem als modulierbarer Risikofaktor für kardiovaskuläre Erkrankungen diskutiert. Die Assoziation des KLF14-Gens sowohl mit HDL-Cholesterin als auch einem reduzierten Risiko für koronare Herzerkrankungen macht es zu einem geeigneten Ziel neuer therapeutischer Ansätze. Zu beachten ist jedoch die inverse Korrelation zwischen Triglyzeriden und HDL-Cholesterin. Da Varianten im KLF14-Gen mit beiden

Lipidparametern assoziiert sind, könnte in weiterführenden Forschungsarbeiten die Genotypisierung von Patienten speziell mit hohen Triglyzerid- und gleichzeitig hohen HDL-Cholesterinwerten bzw. niedrigen Triglyzerid- und HDL-Cholesterinwerten von Interesse sein. Die Verteilung der Genotypen in diesen beiden Extremgruppen kann zu neuen Informationen hinsichtlich der Funktion des Transkriptionsfaktors KLF14 als auch des Zusammenspiels der Lipidfraktionen führen.

Ein weiterer Ansatz in der Erforschung von genetischen Komponenten sind CNVs (copy number variants). Hierbei handelt es sich um unterschiedliche basenpaarlange, strukturelle Variationen von DNA-Abschnitten. Diese können bei unterschiedlichen Individuen dupliziert oder deletiert vorliegen. Häufig vorkommende CNVs ($MAF \geq 5\%$) zeigen moderate Effektgrößen und erklären einen weiteren Anteil der fehlenden genetischen Komponente. Einige bereits identifizierte seltene CNVs, ähnlich seltener SNPs, scheinen höhere Effektstärken zu haben (Manolio et al., 2009). Erste Assoziationen zwischen CNVs und Psychosen, Schizophrenie oder Autismus (Gazzellone et al., 2014; Giaroli et al., 2014) sind bereits veröffentlicht worden. In zukünftigen Untersuchungen zum Einfluss von KLF14 auf den Lipidmetabolismus könnten CNVs eine größere Bedeutung zukommen.

Weiterhin ist sowohl die Bedeutung der gegenseitigen Beeinflussungen der Gene untereinander als auch die Wirkung von Umweltfaktoren auf die Entwicklung von Dyslipidämien nur wenig untersucht. Umweltfaktoren lassen sich in Studien nur schwer berücksichtigen und können die Effekte von ursächlichen Genen verschleiern. Zudem können sie erhebliche Auswirkungen auf den Schweregrad eines Phänotyps haben. Durch Zwillingsstudien können diese verifiziert werden. Dabei beziehen sie sich auf die Entwicklung von Zwillingen, welche nach der Geburt an unterschiedlichen Orten aufgewachsen und dabei unterschiedlichen Umwelteinflüssen ausgesetzt sind. Als Umweltfaktoren zählen aber ebenso Einwirkungen während der Schwangerschaft, welche bei der Ausprägung unterschiedlicher Phänotypen modulierend wirken können. So ist zum Beispiel ein niedriges Geburtsgewicht und ein erhöhtes Plazentagewicht mit der Entwicklung eines Hypertonus im Erwachsenenalter assoziiert (Barker et al., 1990). Die Datenerhebung zu Einflüssen und möglichen Komplikationen in der Schwangerschaft ist meist nicht möglich oder beruht auf subjektiven Aussagen der Patienten. Die Evaluation ist schwierig und bedarf eines großen Aufwands in der Versuchsplanung.

Zu den Mechanismen der Gen-Gen-Interaktion gehört die Epistase. Erstmals führte Bateson 1909 den Begriff ein. Er beschrieb damit einen Effekt, bei welchem die

phänotypische Manifestation einer genetischen Variante an Locus 1 verhindert wird durch eine zweite genetische Variante an Locus 2 (Bateson, 1909). Eine weitere statistische Definition der Epistase veröffentlichte Fisher 1918. Dieser formulierte die Epistase als eine Abweichung der additiven Effektstärke durch Kombination von an verschiedenen Loci liegenden Allelen auf den Phänotyp (Fisher, 1918). Vor allem im Hinblick auf die noch offenen Anteile der genetischen Komponenten komplexer Krankheiten wird die Epistase an Bedeutung gewinnen. Dadurch lassen sich unterschiedliche Schweregrade bei Patienten mit denselben genetischen Veränderungen erklären. Jedoch können so auch die durch Assoziationsstudien gewonnenen Ergebnisse fälschlicherweise einem anderen Gen zugeschrieben bzw. nicht entdeckt werden. Ebenfalls können durch epigenetische Prozesse, wie zum Beispiel Imprinting-Funktionen, die Schweregrade von Krankheiten beeinflusst werden. Das KLF14-Gen unterliegt maternalen Imprinting-Effekten. Der SNP rs4731702 im Bereich des KLF14-Gens besitzt ein erhöhtes Risiko für Diabetes mellitus Typ 2 bei mütterlicher Vererbung. Zudem korreliert das Diabetes-assoziierte Risikoallel dieses SNPs mit einer geringeren Expression des KLF14-Transkriptionsfaktors im Fettgewebe (Kong et al., 2009). Für den SNP rs4731702 wurde neben seiner cis-regulatorischen Aufgabe eine trans-Funktion auf ein Netzwerk von Genen nachgewiesen, welche mit dem Lipidmetabolismus assoziiert sind. Die durch trans-Regulation erklärbare Varianz der Expression dieser Gene wird mit 3-7,8 %, dementsprechend 6-25% der Vererbung, angegeben (Small et al., 2011). Die Ergebnisse zu Gen-Gen Interaktionen sind eine sinnvolle Erweiterung der GWA- und Resequenzierungsstudien und bergen ein großes Potential zur Aufklärung der genetischen Ursachen von Krankheiten.

4.5 Bewertung der CDCV- und CDRV-Hypothese

Als zentrale Thesen im Rahmen genetischer Untersuchungen dienen die CDCV- und die CDRV-Hypothesen. Die CDCV-Hypothese sieht als Ursache für komplexe Krankheiten, wie Diabetes mellitus, Herz-Kreislauf-Erkrankungen und Adipositas, einen additiven Effekt von häufig vorkommenden Varianten mit einer hohen Frequenz in der Bevölkerung und einer jeweils geringen Effektgröße der einzelnen Varianten. Im Gegensatz dazu geht die CDRV-Hypothese davon aus, dass komplexe Krankheiten durch seltene Varianten mit jeweils geringer Frequenz, aber dafür hoher Effektgröße, entstehen. Auf Basis der CDCV-Hypothese wurden zahlreiche GWA-Studien durchgeführt. Bereits bekannte Gene wie CETP, LIPC, LIPG oder LPL (Keebler et al., 2009; Knoblauch et al., 2004; Rip et al., 2006; Guerra et al., 1997, und Padmaja et al., 2007)

und ABCA1 (Frikke-Schmidt et al., 2004) wurden mehrfach in GWAS mit HDL-Cholesterin assoziiert. Neu identifizierte Genloci mit einem Einfluss auf den HDL-Cholesterinspiegel sind GALNT2 (Kathiresan et al., 2008), MVK-MMAB (Willer et al., 2008) und KLF14 (Teslovich et al., 2010). Trotz der enormen Menge an bislang nachgewiesenen Genloci mit häufigen genetischen Varianten ist eine vollständige Klärung der Pathophysiologie bisher nicht gelungen. Nur etwa 5-10 % der erblichen Komponente komplexer Erkrankungen lassen sich durch den aktuellen Wissensstand erklären (Schork et al., 2009). Der große Nachteil von GWAS ist, dass Effekte einzelner seltener Varianten, welche eine hohe Effektgröße haben, durch die große Anzahl an untersuchten Patienten nicht detektiert werden. Am Beispiel der Körpergröße kann dieser Informationsverlust veranschaulicht werden. Zwei Patienten werden als Träger derselben häufigen Variante A für das Merkmal Körpergröße detektiert und für dieses eingruppiert. Zusätzlich ist einer der beiden Patienten Träger einer seltenen Variante B mit großem Effekt auf die Körpergröße. Durch GWAS werden nun weitere Träger der häufigeren genetischen Varianten A detektiert und die Ergebnisse über mehrere tausend Patienten gemittelt. Dadurch mindert sich der Effekt der seltenen Variante B (Maher, 2008). Somit gewinnen bei der Entschlüsselung der Genetik komplexer Erkrankungen die CDRV-Hypothese und die damit verbundenen Resequenzierungsstudien an Bedeutung. Als potentielle Kandidatengene eignen sich vor allem die in GWAS bereits identifizierten Genbereiche zur Sequenzierung. Ein Beispiel für die Generierung neuer Informationen hinsichtlich der genetischen Variabilität ist die von Johansen et al. (2010) durchgeführte Vergleichsstudie, welche eine GWA-Analyse und eine Resequenzierungsarbeit in derselben Population beinhaltet. Als Kandidatengene wurden APOA5, GCKR, LPL und APOB ausgewählt. Die Patienten sind in zwei Kohorten eingeteilt gewesen: eine GWAS-Kohorte mit 463 Hypertriglyzeridämie-(HTG)-Patienten und 1197 Kontrollpatienten sowie eine Sequenzierungskohorte mit 438 erkrankten Individuen und 327 gesunden Kontrollpatienten. Seltene Varianten, definiert mit einer Frequenz des selteneren Allels $< 1\%$, waren bei HTG-Patienten signifikant gehäuft. Insgesamt konnten in der Sequenzierungskohorte 154 seltene Varianten bei HTG-Patienten und 53 seltene Varianten in der Kontrollgruppe entdeckt werden. Davon sind 47 seltene Varianten ausschließlich bei erkrankten HTG-Individuen aufgetreten und neun seltene Varianten in der Kontrollgruppe. In der Gesamtpopulation, bestehend aus der GWAS- und der Sequenzierungskohorte, ließ sich für 41,6 % der Varianz an Triglyzeridwerten eine ursächliche Erklärung durch klinische Faktoren, häufige und seltene genetische Varianten finden. 19,7 % konnten durch klinische Faktoren, 20,8 % durch häufige gene-

tische Varianten und 1,1 % durch seltene Varianten in den HTG-assoziierten Loci erklärt werden (Johansen et al., 2010). Bei dieser Studie handelt es sich speziell um die Resequenzierung HTG-assoziiierter Genloci. Die Menge an neu identifizierten genetischen Varianten allein in dieser Publikation stützt den Stellenwert von Resequenzierungen und das Potential von seltenen genetischen Varianten. Auch die in der vorliegenden Arbeit identifizierte Deletionsmutation konnte durch die gewählte Art der Genotypisierung gefunden werden. Diese würde in GWAS vermutlich nicht identifiziert werden bzw. hätte durch die statistische Auswertung kein signifikantes Niveau erreicht. Die beiden zentralen Hypothesen sollten nicht als sich gegenseitig ausschließende Thesen betrachtet werden. Vielmehr haben sie jeweils einen relativen Anteil im komplexen Zusammenspiel des Lipidmetabolismus und ergänzen sich gegenseitig. Die Anzahl und die Art der Varianten sowie deren Kombination schwankt dabei nicht nur zwischen einzelnen Individuen, sondern auch zwischen unterschiedlichen Populationen (Evans et al., 2011).

4.6 Bedeutung und Nutzen genetischer Untersuchungen

Ein wichtiger Punkt einer genetischen Diagnostik ist ihr lebenslanger Bestand. Die einmalige Untersuchung für einen genetischen Marker ist ausreichend im Gegensatz zu Messwerten wie Lipidwerten oder Blutdruckmessungen. Diese unterliegen ständigen Schwankungen und werden zum Beispiel durch die Tageszeit, die Ernährung, Medikamente oder psychosoziale Faktoren beeinflusst. Die Diagnosesicherung ist dadurch oft erschwert und meist erst mit der Manifestation der Erkrankung möglich. Genetische Untersuchungen, wie bei der familiären Hypercholesterinämie mit LDL-Rezeptordefekt, können als sinnvolle Sekundärprävention genutzt werden, um betroffene Angehörige zu identifizieren. Erleichtert wird die Untersuchung großer DNA-Mengen durch das Verfahren der Sequenzierung der nächsten Generation (next generation sequencing). Die in der vorliegenden Arbeit verwendeten Verfahren zur Sequenzierung sind im Moment noch kostenintensiv. Mit Hilfe von Hochdurchsatzverfahren kann bald in kurzer Zeit eine enorme Menge an Daten kostengünstig produziert werden. Projekte wie das \$1000-Genome-Technology-Programm des NHGRI haben bereits die Kosten für die Sequenzierung eines menschlichen Genoms von circa 10 Millionen US-Dollar im Jahr 2007 auf heute circa 5000 US-Dollar sinken lassen (Hayden, 2014). Dadurch wird es auch möglich, Sequenzierungen im Intronbereich durchzuführen. Genetische Veränderungen in diesen Bereichen können zu veränderten Spleißprodukten führen und damit Einfluss auf den Phänotyp haben. Außerdem werden durch die sinkenden Kosten Projekte wie

personalisierte Medizin und Gen-Scores als diagnostisches Hilfsmittel realisierbar. In Zukunft soll die Erstellung von individuellen genetischen Profilen, in der Zusammenschau mit klinischen Daten und der Patientenanamnese, die Interpretation des patienteneigenen Risikoprofils erleichtern (Wilson u. Nicholls, 2015). Die Nutzung von prädiktiven Gen-Scores zur Prävention wird derzeit noch kontrovers diskutiert. In der Studie von Kathiresan et al. (2008) zeigte sich ein genetischer Score aus neun SNPs, welche mit HDL- und LDL-Cholesterin assoziiert sind, als unabhängiger Risikomarker für kardiovaskuläre Erkrankungen (Kathiresan et al., 2008). Dagegen konnte in einer Follow-up-Studie mit mehr als 19000 gesunden weißen Frauen keine signifikante Assoziation eines genetischen Risikoscores mit der Inzidenz für kardiovaskuläre Ereignisse nachgewiesen werden. Dabei wurde ein genetischer Risikoscore aus 101 SNPs, welche in GWAS in unterschiedlichsten Populationen identifiziert wurden, erstellt. Ein nach Altersangleichung vorhandenes Risiko-Verhältnis von 1,015 pro Risikoallel für kardiovaskuläre Ereignisse konnte ermittelt werden. Jedoch konnte nach Anpassung für bereits etablierte klinische Risikofaktoren, wie arterieller Hypertonus oder positive Familienanamnese, keine Verbesserung der Risikoeinschätzung oder Assoziation mit kardiovaskulären Erkrankungen gezeigt werden (Paynter et al., 2010). In der Berechnung der meisten genetischen Scores werden zur Zeit noch vorrangig häufige Varianten genutzt, da sie durch die Vielzahl an GWA-Studien verhältnismäßig häufiger publiziert wurden. Die Identifizierung seltener Varianten und die Entschlüsselung von Gen-Gen- und Gen-Umwelt-Interaktionen lassen eine Verbesserung der Vorhersagekraft dieser Scores erwarten.

5 Zusammenfassung

Der Lipidstoffwechsel wird durch diverse Rezeptoren und Enzyme reguliert. Diese werden auf verschiedenen Genen kodiert. Bei der Mehrheit der Dyslipidämie-Patienten besteht keine monogenetische Störung mit klassischer Vererbung nach Mendel, sondern es können eine Vielzahl von genetischen Varianten ursächlich sein. Dabei ist noch nicht vollständig geklärt, welchen Effekt die einzelnen Varianten haben und wie diese zusammenwirken. Der KLF14-Genlocus kodiert für einen Transkriptionsfaktor und wurde mit verschiedenen Stoffwechselwegen assoziiert, unter anderem mit dem Lipidmetabolismus. In der vorliegenden Arbeit ist die genetische Variante S58P des KLF14-Genlocus auf einen Zusammenhang mit Dyslipidämien untersucht wurden. Hierbei zeigten Träger des selteneren G-Allels einen signifikant niedrigeren HDL-Cholesterinwert, welcher auch in der männlichen Subpopulation bestätigt wurde. Dieses Ergebnis spiegelt nicht die bisher veröffentlichten Ergebnisse zu genetischen Varianten im KLF14-Gen wider. Da es sich in der untersuchten Population um schwer dyslipidämische Patienten handelt, sind abweichende Ergebnisse durchaus möglich. Die Patienten können Träger zusätzlicher genetischer Varianten mit größeren Effekten als der S58P-Polymorphismus sein. Patienten mit Dyslipidämien leiden meist auch an anderen Stoffwechselstörungen, weshalb die Arbeit auf die Parameter Adipositas und Diabetes mellitus Typ 2 erweitert wurde. Trägerinnen des G-Allels sind signifikant häufiger in der Nicht-Adipositas Gruppe mit einem BMI $< 30 \text{ kg/m}^2$ vorhanden. Dieses Ergebnis wird bestätigt durch die Tendenz zu einem durchschnittlich niedrigeren BMI-Wert und ist schlüssig mit tendenziell höheren HDL-Perzentilen bei Frauen verknüpft. Ein Zusammenhang mit Diabetes mellitus Typ 2 konnte nicht gezeigt werden. Durch die auseinander divergierenden Werte bei Männern und Frauen wird ein geschlechtsspezifischer Effekt der Variante vermutet. Für die Variante rs4731702 sind bereits ähnliche Vermutungen publiziert worden. Außerdem konnte in der Arbeit eine vollständige Kopplung der SNPs S58P, P47Q und A39A nachgewiesen werden, welche für nachfolgende Studien genutzt werden kann. In der weiteren Forschung müssen auch die Interaktionen zwischen den Genen, epigenetische Modulationen und der Einfluss von Umweltfaktoren berücksichtigt werden. Hierfür ist die Entwicklung effizienter und leistungsfähiger neuer Technologien notwendig. Das Verständnis der genetischen Ursachen komplexer Erkrankungen ist die Grundlage für Zukunftsvisionen wie die personalisierte Medizin.

6 Abkürzungsverzeichnis

ABCA1	ATP-bindendes Kassettenprotein A1
ABCG1	ATP-bindendes Kassettenprotein G1
ANGPTL4	Angiopoietin-like 4
APH1B	Anterior pharynx defective 1b homolog
Apo	Apolipoprotein
ARSD	Arylsulfatase D
Aqua dest.	destilliertes Wasser
BMI	Body-Mass-Index
bp	Basenpaare
C8orf82	Chromosome 8 open reading frame 82
CDCV-Hypothese	Common-disease-common-variant-Hypothese
CDRV-Hypothese	Common-disease-rare-variant-Hypothese
CETP	Cholesterinester-Transportprotein
CNV	Kopienzahlvarianten (copy number variants)
dbSNP	NCBI database of genetic variation
<i>D</i>	Disequilibriumskoeffizient nach Lewontin
DIAGRAM+	Diabetes Genetics Replication and Meta-analysis (Stufe 2)
dl	Deziliter
EAS-Guidelines	European Atherosclerosis Society Guidelines
et al.	und andere
ExoI	Exonuklease I
<i>f</i>	Frequenz
FA	Fettsäure-Transporter (fatty acid transporter)
FH	Familiäre Hypercholesterinämie
g	Gramm
GCKR	Glucokinase (Hexokinase 4) Regulator
GLU	Glukosespiegel
GNB1	Guanine nucleotide-binding protein, beta polypeptide1
GWA, GWAS	Genomweite Assoziationsstudie/n
H ₀	Nullhypothese
H ₁	Alternativhypothese
HDL	Lipoprotein hoher Dichte (high density lipoprotein)
HOMA	Homeostasis Model Assessment (Index für Insulinsensitivität)
HTG	Hypertriglyzeridämie
HWG	Hardy-Weinberg-Gleichgewicht
IDL	Lipoprotein mittlerer Dichte (intermediate density lipoprotein)
ILSE	Interdisziplinäre Langzeit-Studie des Erwachsenenalters über die Bedingungen zufriedenen und gesunden Alterns
INS	Insulinwerte
kg	Kilogramm
KLF13	Krüppel-like Faktor 13
KLF14	Krüppel-like Faktor 14
LCAT	Lecithin-Cholesterin-Acyl-Transferase
LD	Disequilibriumskoeffizient (linkage disequilibrium)
LDL	Lipoprotein geringer Dichte (low density lipoprotein)
LIPC	Hepatische Triglyzeridlipase
LIPG	Endotheliale Lipase
LPL	Lipoproteinlipase

LRP	LDL-related Rezeptor
MAF	Frequenz des selteneren Allels (minor allele frequency)
m ²	Quadratmeter
mg	Milligramm
min.	Minute
ml	Milliliter
MVK-MMAB	Mevalonate kinase and methylmalonic aciduria (cobalamin deficiency) cb1B type protein
MYL5	Myosin light chain 5
n	Anzahl
NHGRI	National Human Genome Research Institute
NINJ2	Nerve injury induced protein 2
nm	Nanometer
NPC1L1	Niemann-Pick C1-like protein 1
OMIM	Online Mendelian Inheritance in Man
p	Allelhäufigkeit oder kurzer Arm des Chromosoms
PCR	Polymerasekettenreaktion
PCSK9	Proteinkonvertase Subtilisin/Kexin Typ 9
PolyPhen-Software	Polymorphism-Phenotyping Software
PPAR α	Peroxisome proliferator-activated receptor α
PRMT2	Protein-Arginin-Methyltransferase 2
q	Allelhäufigkeit oder langer Arm des Chromosoms
rpm	Umdrehungen pro Minute (rounds per minute)
rs	Reference SNP cluster identification
SAP	Shrimp Alkaline Phosphatase
SLC7A10	Solute carrier family 7 (neutral amino acid transporter light chain, asc system), member 10
SNP	Einzelnukleotidpolymorphismen (single nucleotide polymorphism)
SR-A	Scavenger Rezeptor Typ A
SR-B1	Scavenger Rezeptor Typ B1
TG	Triglyzeride
TPMT	Thiopurin-S-Methyltransferase
T2D	Diabetes mellitus Typ 2
U	Unit
u.	und
WHO	World Health Organization
WHR	Hüft-Taillen-Verhältnis (waist hip ratio)
VLDL	Lipoprotein sehr geringer Dichte (very low density lipoprotein)
z.T.	zum Teil
α	Signifikanzniveau, Fehler 1. Art
μ l	Mikroliter
°C	Grad Celsius
\$	Dollar

7 Tabellenverzeichnis

Tabelle 1: Eigenschaften der Lipoproteine.....	2
Tabelle 2: Wichtige Rezeptoren und Enzyme des Lipoproteinmetabolismus und deren Funktion.....	4
Tabelle 3: Monogenetische Dyslipidämien.	7
Tabelle 4: Überblick über Resequenzierungsstudien.	12
Tabelle 5: Bandenmuster nach Restriktionsverdau mit BccI.	21
Tabelle 6: Schema der möglichen 2-Genlocus-Genotypen.	26
Tabelle 7: Darstellung der klinischen Parameter des Gesamtkollektivs.	27
Tabelle 8: Darstellung der Genotypverteilung und der Frequenz des G-Allels.	28
Tabelle 9: Darstellung der Verteilung des Genotyps S58P in der Studie und nach Hardy-Weinberg-Gesetz.	29
Tabelle 10: Darstellung der Verteilung der S58P-Genotypen nach Dyslipidämie.....	30
Tabelle 11: Verteilung des S58P-Polymorphismus nach HDL-Perzentilengruppen.....	31
Tabelle 12: Darstellung der durchschnittlichen Lipidwerte nach Genotyp des S58P-Polymorphismus..	33
Tabelle 13: Darstellung des mittleren BMI (kg/m ²) für den S58P-Polymorphismus.....	34
Tabelle 14: Darstellung des Kopplungsungleichgewichts.....	38
Tabelle 15: Vierfeldertafel des Chi ² -Tests..	79

8 Abbildungsverzeichnis

Abbildung 1: Zusammenhang zwischen Allelhäufigkeit und Effektgröße genetischer Varianten	9
Abbildung 2: Darstellung der cis- und trans- Funktion von Genen.	15
Abbildung 3: Darstellung der cis- und trans-regulatorischen Funktion des KLF14-Gens	16
Abbildung 4: Ausschnitt der KLF14-Gensequenz im Bereich des S58P-Polymorphismus	17
Abbildung 5: Ausschnitt der KLF14-Gensequenz im Bereich des P47Q-Polymorphismus	17
Abbildung 6: Ausschnitt der KLF14-Gensequenz im Bereich des A39A-Polymorphismus	18
Abbildung 7: Darstellung des Elektropherogramms vom 20.06.2012 für den SNP S58P..	28
Abbildung 8: Darstellung der G-Allelfrequenz in den HDL-Perzentilengruppen	32
Abbildung 9: Darstellung der G-Allelfrequenz nach BMI-Gruppe.....	35
Abbildung 10: Darstellung der Frequenz des G-Allels des S58P-Polymorphismus und Diabetes mellitus Typ 2.	36
Abbildung 11: Darstellung der Genotypen des SNP S58P im KLF14-Gen und BMI (kg/m ²).....	37
Abbildung 12: Darstellung der Sequenzierung der Serenprobe 18943.	40
Abbildung 13: Darstellung der Sequenzierung der Serenprobe 13716	41
Abbildung 14: Darstellung der Sequenzierung der Serenprobe 12380	42

Abbildung 15: Auffälliges Bandenmuster mit doppelter Bande im Elektropherogramm des Patienten mit der Serennummer 11151	43
Abbildung 16: Schema der Deletionsmutation.....	44
Abbildung 17: Darstellung der Sequenzierung der Serenprobe 11151	45
Abbildung 18: Schematische Darstellung des Prinzips einer Kapillarelektrophorese.	75

9 Formelverzeichnis

Formel 1: Berechnung der Häufigkeit der Genotypen in einer idealen Population	24
Formel 2: Berechnung der erwarteten Häufigkeit der Genotypen in einer idealen Population.....	25
Formel 3: Berechnung der Prüfgröße χ^2 im Hardy-Weinberg-Gleichgewicht	25
Formel 4: Berechnung des Disequilibriumskoeffizienten D	26
Formel 5: Berechnung des arithmetischen Mittels	77
Formel 6: Berechnung des Medians bei ungeraden n.....	77
Formel 7: Berechnung des Medians bei geraden n.....	77
Formel 8: Berechnung der Varianz	78
Formel 9: Berechnung der Standardabweichung.....	78
Formel 10: Berechnung der Prüfgröße χ^2 im Chi ² -Test.....	79
Formel 11: Berechnung der Prüfgröße U1 im Mann-Whitney-U-Test	80
Formel 12: Berechnung der Prüfgröße U2 im Mann-Whitney-U-Test	80

10 Literaturverzeichnis

Abifadel, M., et al., 2003. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet*, 34(2): 154-156.

Adzhubei, I., et al., 2010. PolyPhen-2 prediction of functional effects of human nsSNPs. [Online] Available at: <http://genetics.bwh.harvard.edu/pph2/> [Zugriff am 13. März 2015].

Alberti, K. G., Zimmet, P. Z., 1998. Definition, diagnosis and classification of diabetes mellitus and its complications. I. Diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med*, 15(7): 539-553.

Austin, M. A., Hutter, C. M., Zimmern, R. L., Humphries, S. E., 2004. Genetic causes of monogenic heterozygous familial hypercholesterolemia: a HuGE prevalence review. *Am J Epidemiol*, 160(5): 407-420.

Barker, D. J., Bull, A. R., Osmond, C., Simmonds, S. J., 1990. Fetal and placental size and risk of hypertension in adult life. *BMJ*, 301(6746): 259–262.

Bateson, W., 1909. *Mendel's Principles of Heredity*. Cambridge: Cambridge University Press.

Beisiegel, U., et al., 1989. The LDL-receptor-related Protein, LRP, is an apolipoprotein E-binding protein. *Nature*, 341(6238): 162-164.

Bickeböller, H., Fischer, C., 2007. *Statistik und ihre Anwendungen: Einführung in die Genetische Epidemiologie*, 1. Auflage, Springer-Verlag, Berlin, Heidelberg, 76-85.

Bodmer, W., Bonilla, C., 2008. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet*, 40(6): 695-701.

Brinton, E. A., Eisenberg, S., Breslow, J. L., 1994. Human HDL Cholesterol Levels are determined by ApoA-I fractional catabolic rate, which correlates inversely with estimates of HDL particle size. Effects of gender, hepatic and lipoprotein lipases, triglyceride and insulin levels, and body fat distribution. *Arterioscler Thromb*, 14(5): 707-720.

Brown, M. S., et al., 1979. Reversible accumulation of cholesteryl esters in macrophages incubated with acetylated lipoproteins. *J Cell Biol*, 82(3): 597-613.

Brunzell, J. D., 1999. Familial Lipoprotein Lipase Deficiency. [Updated 2011 Dec 15]. In: Pagon, R. A., Adam, M. P., Ardinger, H. H., et al., editors. *GeneReviews®* [Online]. Seattle (WA): University of Washington, Seattle; 1993-2013. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK1308/> (Zugriff am 20. April 2013)

Burdett, T., et al., 2014. The NHGRI-EBI Catalog of published genome-wide association studies. [Online] Available at: www.ebi.ac.uk/gwas [Zugriff am 14. April 2015].

- Calandro, S., Oliva, C. P., Tarugi, P., Bertolini, S., 2006. APOA5 and triglyceride metabolism, lesson from human APOA5 deficiency. *Curr Opin Lipidol*, 17(2): 122-127.
- Chasman, D. I., et al., 2009. Forty-Three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet*. 2009; 5(11): e1000730. doi: 10.1371/journal.pgen.1000730. Epub 2009 Nov 20.
- Chen, X., et al., 2012. Genome-wide association study validation identifies novel loci for atherosclerotic cardiovascular disease. *J Thromb Haemost*, 10(8): 1508-1514.
- Civelek, M., Lusis, A. J., 2011. Conducting the metabolic syndrome orchestra. *Nat Genet*, 43(6): 506-508.
- Cohen, J. C., et al., 2004. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science*, 305(5685): 869-872.
- Cohen, J. C., et al., 2006. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci U S A*, 103(6): 1810–1815.
- 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): 2463-2468.
- Cunningham, F., et al., 2015. e!Ensembl Version 79. [Online] Available at: <http://www.ensembl.org/index.html> [Zugriff am 19. März 2015].
- Deutsche Adipositas-Gesellschaft, 2014. Adipositas - Prävention und Therapie. AWMF-Online. Das Portal der wissenschaftlichen Medizin. [Online] Available at: <http://www.awmf.org/leitlinien/detail/ll/050-001.html> [Zugriff am 11. Februar 2015].
- Eckstein, P. P., 2014. Repetitorium Statistik: Deskriptive Statistik - Stochastik - Induktive Statistik. 8. Auflage, Springer Gabler Verlag, Wiesbaden.
- Edmondson, A. C., et al., 2009. Loss-of-function variants in endothelial lipase are a cause of elevated HDL cholesterol in humans. *J Clin Invest*, 119(4): 1042-1050.
- Evans, D., Aberle, J., Beil, F. U., 2011. Resequencing the Apolipoprotein A5 (APOA5) gene in patients with various forms of hypertriglyceridemia. *Atherosclerosis*, 219(2): 715-720.
- Evans, D., Aberle, J., Beil, F. U., 2011. The relative importance of common and rare genetic variants in the development of hypertriglyceridemia. *Expert Rev Cardiovasc Ther*, 9(5): 637-644.
- Evans, D., Arzer, J., Aberle, J., Beil, F. U., 2011. Rare variants in the lipoprotein lipase (LPL) gene are common in hypertriglyceridemia but rare in Type III hyperlipidemia. *Artherosclerosis*, 214(2): 386–390.

- Evans, D., et al., 2001. A polymorphism, L162V, in the peroxisome proliferator-activated receptor α (PPAR α) gene is associated with lower body mass index in patients with non-insulin-dependent diabetes mellitus. *J Mol Med (Berl)*, 79(4): 198-204.
- Fan, Y., et al., 2012. Krüppel-like factor-11, a transcription factor involved in diabetes mellitus, suppresses endothelial cell activation via the nuclear factor- κ B signaling pathway. *Arterioscler Thromb Vasc Biol*, 32(12): 2981-2988.
- Fisher, R. A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edin*, 52: 399-433.
- Frikke-Schmidt, R., Nordestgaard, B. G., Jensen, G. B., Tybjaerg-Hansen, A., 2004. Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *J Clin Invest*, 114(9): 1343–1353.
- Gazzellone, et al., 2014. Copy number variation in Han Chinese individuals with autism spectrum disorder. *J Neurodev Disord*. 2014; 6(1): 34. doi: 10.1186/1866-1955-6-34. Epub 2014 Aug 23.
- Gepner, A. D., et al., 2011. Effects of smoking and smoking cessation on lipids and lipoproteins: Outcomes from a Randomized Clinical Trial. *Am Heart J*, 161(1): 145–151.
- Giaroli, G., et al., 2014. Does rare matter? Copy number variants at 16p11.2 and the risk of psychosis: A systematic review of literature and meta-analysis. *Schizophr Res*, 159(2-3): 340-346.
- Guerra, R., Wang, J., Grundy, S. M., Cohen, J. C., 1997. A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. *Proc Natl Acad Sci U S A*, 94(9): 4532–4537.
- Hardy, G. H., 1908. Mendelian proportions in a mixed population. *Science*, 28(706): 49-50.
- Havel, R. J., 1995. Chylomicron remnants: hepatic receptors and metabolism. *Curr Opin Lipidol*, 6(5): 312-316.
- Hayden, E. C., 2014. Technology: The \$1,000 genome. *Nature*, 507(7492): 294–295.
- Hegele, R. A., 2001. Monogenic dyslipidemia: Window on determinants of plasma lipoprotein metabolism. *Am J Hum Genet*, 69(6): 1161-1177.
- Hegele, R. A., 2009. Plasma lipoproteins: genetic influences and clinical implications. *Nat Rev Genet*, 10(2): 109-121.
- Hegele, R. A., et al., 2015. Targeted next-generation sequencing in monogenic dyslipidemias. *Curr Opin Lipidol*, 26(2): 103-113.
- Heinrich, P. C., Müller, M., Graeve, L., 2007. Löffler/Petrides Biochemie und Pathobiochemie. 8. Auflage, Springer-Verlag ,Heidelberg, 300-306.

- Heller, D. A., et al., 1993. Genetic and environmental influences on serum lipid levels in twins. *N Engl J Med*, 328(16): 1150-1156.
- Henze, K., Wallmüller-Strycker, A., Bauer, M., 1981. Cholesterin und Triglyceride im Serum einer Münchner Bevölkerungsgruppe: Beziehungen zu Alter und Geschlecht. *J Clin Chem Clin Biochem*, 19(10): 1013-1019.
- Hindorff, L. A., et al., 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*, 106(23): 9362–9367.
- Huang, P., et al., 2013. Association of the KLF14 rs4731702 SNP and serum lipid levels in the Guangxi Mulao and Han Populations. *Biomed Res Int*. 2013; 2013:231515. doi: 10.1155/2013/231515. Epub 2013 Sep 30.
- Johansen, C. T., et al., 2010. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet*, 42(8): 684-688.
- Kathiresan, S., et al., 2008. Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med*, 358(12): 1240-1249.
- Kathiresan, S., Melander, O., Guiducci, C., Surt, A., 2008. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet*, 40(2): 189-197.
- Keebler, M. E., Sanders, C. L., Surti, A., Guiducci, C., 2009. Association of blood lipids with common DNA sequence variants at nineteen genetic loci in the multiethnic United States National Health and Nutrition Examination Survey III. *Circ Cardiovasc Genet*, 2(3): 238-243.
- Khor, C., Li-Meng Goh, D., 2010. Strategies for identifying the genetic basis of dyslipidemia: genome-wide association studies vs. the resequencing of extremes. *Curr Opin Lipidol*, 21(2): 123-127.
- Khovidhunkit, W., et al., 2012. Resequencing CETP, LIPC and LIPG Genes in Thai Subjects with Hyperalphalipoproteinemia. *Am J Cardiol*, 110(1): 62-66.
- Knoblauch, H., et al., 2004. Haplotypes and SNPs in 13 lipid-relevant genes explain most of the genetic variance in high-density lipoprotein and low-density lipoprotein cholesterol. *Hum Mol Genet*, 13(10): 993-1004.
- Knoedler, J. R., Denver, R. J., 2014. Krüppel-like factors are effectors of nuclear receptor signaling. *Gen Comp Endocrinol*, 203: 49-59.
- Kong, A., et al., 2009. Parental origin of sequence variants associated with complex diseases. *Nature*, 462(7275): 868-874.
- Kotowski, I. K., et al., 2006. A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am J Hum Genet*, 78(3): 410-422.

- Kreisberg, R., 1998. Diabetic dyslipidemia. *Am J Cardiol*, 82(12A): 67U–73U.
- Kryukov, G. V., Pennacchio, L. A., Sunyaev, S. R., 2007. Most rare missense alleles are deleterious in Humans: Implications for complex disease and association studies. *Am J Hum Genet*, 80(4): 727-739.
- Kuckartz, U., 2013. Statistik - Eine verständliche Einführung. 2. Auflage, Springer, Wiesbaden.
- Kuivenhoven, J. A., Hegele, R. A., 2014. Mining the genome for lipid genes. *Biochim Biophys Acta*, 1842(10): 1993-2009.
- Lander, E. S., Schork, N. J., 1994. Genetic dissection of complex traits. *Science*, 265(5181): 2037-2048.
- Lewontin, R. C., 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, 49(1): 49-67.
- Lund-Katz, S., Phillips, M. C., 2010. High density lipoprotein structure–Function and role in reverse cholesterol transport. *Subcell Biochem*, 51: 183-227.
- Maher, B., 2008. Personal Genomes: The case of the missing heritability. *Nature*, 456(7218): 18-21.
- Manolio, T. A., Brooks, L. D., Collins, F. S., 2008. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest*, 118(5): 1590-1605.
- Manolio, T. A., et al., 2009. Finding the missing heritability of complex diseases. *Nature*, 461(7265): 747-753.
- Martin, H., et al., 1999. Klinische Labordiagnostik und Altern. Teil 1: Ergebnisse einer Datenauswertung klinisch-chemischer Laborwerte in einer Alternsstudie. *Z Gerontol Geriatr*, 32(2): 89-96.
- Mayo, O., 2008. A century of Hardy-Weinberg equilibrium. *Twin Res Hum Genet*, 11(3):249-256.
- McCarthy, M. I., et al., 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*, 9(5): 356-69.
- McConnell, B. B., Yang, V. W., 2010. Mammalian Krüppel-Like Factors in health and diseases. *Physiol Rev*, 90(4): 1337–1381.
- McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), Updated 27. April 2015. Online Mendelian Inheritance in Man, OMIM®. [Online] Available at: <http://omim.org/> [Zugriff am 28. April 2015].
- Mehta, T. S., et al., 2009. A unique sequence in the N-terminal regulatory region controls the nuclear localization of KLF8 by cooperating with the C-terminal zinc-fingers. *Cell Res*, 19(9): 1098-1109.

Middelberg, R. P., Martin, N. G., Whitfield, J. B., 2007. A longitudinal genetic study of plasma lipids in adolescent twins. *Twin Res Hum Genet*, 10(1): 127-135.

Moore, K. J., Freeman, M. W., 2006. Scavenger receptors in atherosclerosis beyond lipid uptake. *Arterioscler Thromb Vasc Biol*, 26(8): 1702-1711.

Murken, J., Grimm, T., Holinski-Feder, E., 2006. Taschenlehrbuch Humangenetik. 7. Auflage, Georg Thieme-Verlag, Stuttgart, 348.

National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III), 2002. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation*, 106(25): 3143–3421.

Nordestgaard, B. G., et al., 2013. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease. *Eur Heart J*, 34(45): 3478-3490a.

Padmaja, N., Ravindra-Kumar, M., Soya, S. S., Adithan, C., 2007. Common variants of cholesteryl ester transfer protein gene and their association with lipid parameters in healthy volunteers of TAMILIAN population. *Clin Chim Acta*, 375(1-2): 140-146.

Pandya, A. Y., et al., 2004. Nuclear localization of KLF4 is associated with an aggressive phenotype in early-stage breast cancer. *Clin Cancer Res*, 10(8): 2709-2719.

Pandya, K., Townes, T. M., 2002. Basic residues within the Kruppel zinc finger DNA binding domains are the critical nuclear localization determinants of EKLF/KLF-1. *J Biol Chem*, 277(18): 16304-16312.

Patsch, W., et al., 1992. The relation of high density lipoprotein cholesterol and its subfractions to Apolipoprotein A-I and fasting triglycerides: The role of environmental factors (The Atherosclerosis Risk in Communities (ARIC) Study). *Am J Epidemiol*, 136(5): 546-557.

Paynter, N. P., et al., 2010. Association between a literature-based genetic risk score and cardiovascular events in 19,313 Women. *JAMA*, 303(7): 631-637.

Perusse, L., et al., 1989. Genetic and environmental determinants of serum lipids and lipoproteins in french canadian families. *Arteriosclerosis*, 9(3): 308-318.

Phillips, N. R., Havel, R. J., Kane, J. P., 1981. Levels and interrelationships of serum and lipoprotein cholesterol and triglycerides. Association with adiposity and the consumption of ethanol, tobacco, and beverages containing caffeine. *Arteriosclerosis*, 1(1): 13-24.

QIAgen, 01/2011. QIAxcel® DNA Handbook. QIAgen, Hilden.

QIAgen, 2008. QIAquick Spin Handbook. Qiagen, Hilden.

- Rader, D. J., 2009. Lecithin - Cholesterol - Acyltransferase and atherosclerosis: Another high-density lipoprotein story that doesn't quite follow the script. *Circulation*, 120(7): 549–552.
- Rao, D. C., et al., 1982. The Cincinnati Lipid Research Clinic family study: cultural and biological determinants of lipids and lipoprotein concentrations. *Am J Hum Genet*, 34(6): 888-903.
- Remaley, A. T., et al., 1999. Human ATP-binding cassette transporter 1 (ABC1): Genomic organization and identification of the genetic defect in the original Tangier disease kindred. *Proc Natl Acad Sci U S A*, 96(22): 12685-12690.
- Rip, J., et al., 2006. Lipoprotein Lipase S447X: A naturally occurring gain-of-function mutation. *Arterioscler Thromb Vasc Biol*, 26(6): 1236-1245.
- Rodríguez, E., et al., 2010. Nucleo-cytoplasmic localization domains regulate Krüppel-Like Factor 6 (KLF6) protein stability and tumor suppressor function. *PLoS ONE* 5(9): e12639. doi: 10.1371/journal.pone.0012639.
- Romeo, S., et al., 2007. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet*, 39(4): 513-516.
- Rosenfeld, M. E., 2015. Converting smooth muscle cells to macrophage-like cells with KLF4 in atherosclerotic plaques. *Nat Med*, 21(6): 549-551.
- Saito, F., 1984. A pedigree of homozygous familial hyperalphalipoproteinemia. *Metabolism*, 33(7): 629-633.
- Schork, N. J., Murray, S. S., Frazer, K. A., Topol, E. J., 2009. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev*, 19(3): 212-219.
- Schwandt, P., Parhofer, K. G., 2006. *Handbuch der Fettstoffwechselstörungen: Dyslipoproteinämien und Atherosklerose: Diagnostik, Therapie und Prävention*. 3. Auflage, Schattauer Verlag, Stuttgart, 3-48.
- Shankman, L. S., et al., 2015. KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. *Nat Med*, 21(6): 628-637.
- Sherry, S. T., et al., 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*, 29(1): 308-311.
- Small, K. S., et al., 2011. Identification of an imprinted master trans-regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet*, 43(6): 561-564.
- Stephen, S. L., et al., 2010. Scavenger receptors and their potential as therapeutic targets in the treatment of cardiovascular disease. *Int J Hypertens*. 2010; 2010:646929. doi: 10.4061/2010/646929.

- Tarugi, P., Aversa, M., 2011. Hypobetalipoproteinemia: genetics, biochemistry, and clinical spectrum. *Adv Clin Chem*, 54: 81-107.
- Teslovich, T. M., et al., 2010. Biological, clinical, and population relevance of 95 loci for blood lipids. *Nature*, 466(7307): 707-713.
- Turner, J., Crossley, M., 1999. Mammalian Krüppel-like transcription factors: more than just a pretty finger. *Trends Biochem Sci*, 24(6): 236-240.
- Vanhoof, G., et al., 1995. Proline motifs in peptides and their biological processing. *FASEB J*, 9(9): 736-744.
- Verdier, C., et al., 2013. Targeting high-density lipoproteins: update on a promising therapy. *Arch Cardiovasc Dis*, 106(11): 601-611.
- Voight, B. F., et al., 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet*, 42(7): 579-589.
- Wang, J., et al., 2014. Association of KCNQ1 and KLF14 polymorphisms and risk of type 2 diabetes mellitus: A global meta-analysis. *Hum Immunol*, 75(4): 342-347.
- Weinberg, W., 1908. Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg*, 64: 369-382.
- Weiß, C., 2013. *Basiswissen Medizinische Statistik*. 6. Auflage, Springer-Verlag, Berlin, Heidelberg.
- Weiss, K. M., Clark, A. G., 2002. Linkage disequilibrium and the mapping of complex human traits. *Trends Genet*, 18(1): 19-24.
- Wigginton, J. E., Cutler, D. J., Abecasi, G. R., 2005. A note on exact tests of Hardy-Weinberg Equilibrium. *Am J Hum Genet*, 76(5): 887–893.
- Willer, C. J., et al., 2008. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet*, 40(2): 161-169.
- Wilson, B. J., Nicholls, S. G., 2015. The human genome project, and recent advances in personalized genomics. *Risk Manag Healthc Policy*, 8: 9-20.
- Wu, Y., et al., 2013. Genetic association with lipids in Filipinos: waist circumference modifies an APOA5 effect on triglyceride levels. *J Lipid Res*, 54(11): 3198–3205.

11 Danksagung

Abschließend möchte ich all denen Dank sagen, ohne deren Hilfe und Unterstützung diese Arbeit nicht möglich gewesen wäre.

Zunächst gilt mein Dank Herrn Prof. Dr. med. Beil für die Möglichkeit der Promotion in der III. Medizinischen Klinik und Poliklinik, Sektion Endokrinologie und Diabetologie.

Außerdem möchte ich mich bei Herrn Dr. David Evans für die Bereitstellung des Themas und für die stets freundliche und fachliche Unterstützung bedanken.

Für die Einarbeitung, Unterstützung und Hilfe bei technischen Fragen im Labor sei Christine und Anja gedankt.

Mein besonderer Dank gilt meiner lieben Familie, welche mich während des Studiums und der Promotion fortwährend motiviert und unterstützt hat.

12 Lebenslauf

Entfällt aus datenschutzrechtlichen Gründen.

13 Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift:

Anhang 1

A 1.1 Polymerasekettenreaktion

Die Polymerasekettenreaktion (PCR) dient der Vervielfältigung (Amplifizierung) einer spezifischen DNA-Sequenz in vitro. Für die Amplifizierung sind bereits kleinste Mengen DNA, sogenanntes Template, ausreichend. Insgesamt umfasst die PCR etwa 20-40 Zyklen. Jeder Zyklus besteht aus drei Schritten: der Denaturierung des DNA-Doppelstranges, der Hybridisierung der Primer (Annealing) und der DNA-Synthese durch eine Polymerase (Elongation). Alle Schritte laufen bei unterschiedlichen Temperaturen ab. Zunächst wird die DNA-Doppelhelix durch Erhitzen auf 95 °C in Einzelstränge getrennt. Danach wird die Reaktionskette abgekühlt, um die Hybridisierung der Primer zu ermöglichen. Jeder Primer benötigt für das Annealing eine andere Temperatur. Für die in dieser Arbeit gewählten Primer ist eine Temperatur bei 60 °C günstig. Die Primer lagern sich dabei spezifisch an ihren komplementären DNA-Einzelstrang an. Im nächsten Schritt der Kettenreaktion erfolgt die DNA-Synthese. Hierfür wird die Temperatur auf 72 °C erhöht, da bei dieser Temperatur die hitzestabile HotStarTaq-Polymerase optimal arbeitet. Die Polymerase beginnt mit der Elongation am freien 3'-OH-Ende unter Anlagerung von Desoxynukleotiden. Die Anheftung erfolgt komplementär zur Sequenz am DNA-Einzelstrang. Die Replikationsrichtung ist dabei vom 3'-OH-Ende zum 5'-OH-Ende verlaufend. Anschließend beginnt ein neuer Zyklus mit der Erhöhung der Temperatur auf 95 °C und der erneuten Auftrennung der entstandenen DNA-Doppelstränge.

A 1.2 Restriktionsendonukleasenverdau

Restriktionsendonukleasen sind Enzyme, welche spezifisch doppelsträngige DNA-Sequenzen erkennen und an diesen Stellen die Phosphodiesterbindung des DNA-Doppelstranges hydrolysieren und somit zerschneiden. Zur Erkennung dient den Restriktasen eine Basensequenz von vier, sechs oder acht Nukleotiden. Die Erkennungssequenz besteht meist aus einem Palindrom, d. h. die Erkennungsstellen sind auf den komplementären DNA-Strängen spiegelverkehrt zueinander. Durch eine genetische Variante im Bereich der Erkennungssequenz kann sich die Schnittstelle eines Restriktionsenzym verändert. Die ursprüngliche Schnittstelle wird dann entweder nicht mehr erkannt oder eine weitere Schnittstelle entsteht. Das zur Analyse der genetischen Variante S58P genutzte Restriktionsenzym ist BccI. Das Enzym wird aus Escherichia coli Bakterien, welche mit Bacteroides caccae infiziert wurden, gewonnen. Das Temperatur-

optimum liegt bei 37 °C mit einer Konzentration von 10.000 U/ml. Zur Erkennung dient folgende Sequenz:

5'...CCATC(N)4...3'

3'...GGTAG(N)5...5'

Die Fragmentierung des PCR-Produkts dient der Gewinnung eines definierten Längenabschnitts zur weiteren Analyse in der Gelelektrophorese.

A 1.3 Kapillargelelektrophorese

Die Gelelektrophorese dient dem Nachweis der DNA-Fragmente. Durch das Vorhandensein von Phosphatresten in der Primärstruktur der DNA ist diese negativ geladen und somit ein Polyanion. Im elektrischen Feld ist das DNA-Fragment dadurch mobil und bewegt sich in Richtung der Anode. Die Kapillargelelektrophorese wird hier mit Hilfe des QIAxcel-System (QIAgen) und dem QIAxcel-DNA-Screening Kit (QIAgen) durchgeführt. Dabei erfolgt die größenabhängige Auftrennung der DNA-Fragmente aus dem Restriktionsverdau in einem Kapillarröhrchen. Insgesamt sind zwölf Kapillarröhrchen in einer vorgefertigten Gelkartusche angeordnet und die Proben-Lösungen werden automatisch aufgenommen. Durch das Anlegen einer Spannung wandern die Proben in Abhängigkeit von Ladung und Zeit in dem Kapillarröhrchen zum positiv geladenen Ende und nehmen dabei Fluoreszenzmaterial auf. Ethidiumbromid dient dabei als Fluoreszenzmittel. Ein Detektor registriert und misst das Fluoreszenzsignal und konvertiert das Emissionssignal in ein elektronisches Signal. Die dazugehörige BioCalculator-Software berechnet die Größe des DNA-Fragments (siehe auch Abbildung 18). Der QX DNA Size Marker wird dabei als Referenz verwendet. Die erste und letzte Bande des Gels wird durch den QX Alignment Marker gebildet, welcher in allen zwölf Röhrchen zur einheitlichen Darstellung der Wanderungszeit dient (QIAgen, 01/2011).

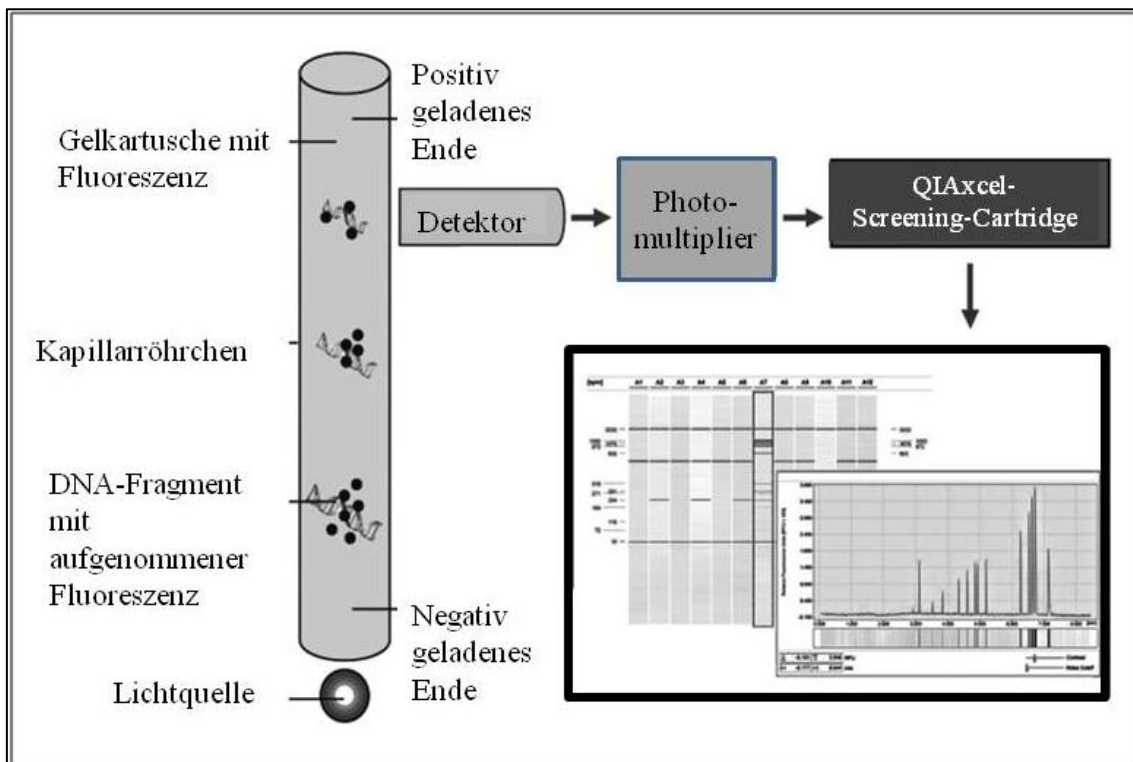


Abbildung 18: Schematische Darstellung des Prinzips einer Kapillarelektrophorese. Die DNA-Fragmente werden in einer Gel-befüllten Kapillare durch Anlegen eines elektrischen Feldes nach Größe sortiert. Während die Fragmente zum positiv geladenen Ende der Kapillare wandern, nehmen sie aus dem Gel Fluoreszenz auf. Bei Erreichen des Detektor wird das Fluoreszenzsignal erfasst. Die Daten werden mit der Software in ein Elektropherogramm und in ein Gelbild umgewandelt. Vgl. QIAXcel DNA Handbook (QIAGEN, 01/2011).

A 1.4 Aufreinigung des PCR-Produktes

Das PCR-Produkt wurde nach dem Protokoll des QIAquick PCR-Purification Kit aufgereinigt. Eine im Kit enthaltene Silikat-Säule absorbiert das PCR-Produkt und Verunreinigungen wie Primer und überschüssige Nukleotide werden ausgewaschen. Die Bindung der DNA an die Säule ist pH-abhängig, sodass ein pH-Wert von $< 7,5$ vorhanden sein sollte. Unterstützt wird die Absorption durch PB-Puffer, welcher eine hohe Salzkonzentration im Gemisch schafft. Danach erfolgt das Auswaschen mit Hilfe des ethanolhaltigen PE-Puffers. Für die Elution der gereinigten DNA-Amplifikate sind wiederum niedrige Salzkonzentrationen von Vorteil, welche durch EB-Puffer geschaffen werden (QIAGEN, 2008).

A 1.5 Sequenzierung mit der Kettenabbruchmethode nach Sanger

Für die Sequenzierung nach Sanger wird das aufgereinigte PCR-Produkt zunächst mit dem Primer hybridisiert. Es wird dabei der gleiche Primer wie für die PCR-Reaktion

genutzt. Danach werden DNA-Polymerase sowie alle vier Nukleotidbausteine und zusätzlich auch spezielle farbmarkierte Nukleotide beigemischt. Die Markierungen sind dabei spezifisch für die Nukleotide Adenin, Cytosin, Guanin und Thymin. Mit Hilfe der DNA-Polymerase erfolgt die enzymatische Synthese des komplementären DNA-Stranges. Es werden dann per Zufall sowohl die Nukleotide in die zu verlängernde Kette eingebaut als auch die farbmarkierten Nukleotide. Durch den Einbau der farbmarkierten Nukleotide bricht die Kette an dieser Stelle ab (Kettenabbruch), da kein freies 3'-OH-Ende zur Verfügung steht, um ein weiteres Nukleotid anzufügen. So entsteht eine Vielzahl an Syntheseprodukten unterschiedlicher Länge. In der Gelelektrophorese werden die synthetisierten DNA-Stränge aufgetragen. Anhand des farblich markierten Bandenmusters lässt sich abschließend die Basensequenz bestimmen.

Anhang 2

A 2.1 Verwendete Begriffe der deskriptiven Statistik

A 2.1.1 Arithmetisches Mittel $\bar{x}$

Das arithmetische Mittel ist das wohl am häufigsten verwendete Lagemaß. Eine Berechnung ist nur möglich, wenn eine Differenz zwischen zwei Merkmalsausprägungen definiert ist. Daraus ergibt sich, dass die vorhandenen Daten quantitativer Art sein müssen. Zur Ermittlung des Mittelwertes werden alle Stichprobenwerte addiert und durch die Gesamtprobenanzahl dividiert (Weiß, 2013). Es gilt folgende Formel:

Formel 5: Berechnung des arithmetischen Mittels

$$\bar{x} = \frac{1}{n} \times \sum_{i=1}^n x_i$$

n ...Stichprobenumfang
 x ...Stichprobenwert
 i ...Laufindex

A 2.1.2 Median $\tilde{x}$

Das zweite bekannte Lagemaß ist der Median. Dieser teilt die Stichprobe in zwei Hälften. In jeder Gruppe befinden sich jeweils Werte, die entweder maximal so groß wie der Median sind oder mindestens so groß. Dafür müssen die Daten mit dem Stichprobenumfang n mindestens ordinalskaliert sein, um eine Rangfolge mit $x_{(1)} \leq \dots \leq x_{(n)}$ festlegen zu können (Weiß, 2013).

Es gilt folgende Formel:

Für ungerade n :

Formel 6: Berechnung des Medians bei ungeraden n

$$\tilde{x} = x_{\frac{(n+1)}{2}}$$

n ...Stichprobenumfang
 x ...Stichprobenwert

Für gerade n :

Formel 7: Berechnung des Medians bei geraden n

$$\tilde{x} = \frac{x_{\frac{(n+1)}{2}} + x_{\frac{n}{2}}}{2}$$

n ...Stichprobenumfang
 x ...Stichprobenwert

A 2.1.3 Varianz und Standardabweichung

Die Eigenschaften Varianz und Standardabweichung gehören zu den Streuungsmaßen. Diese helfen, die Variabilität der erhobenen Daten zu quantifizieren. Die Varianz ist das zum arithmetischen Mittel zugehörige Streumaß. Sie liefert Informationen darüber, wie weit die einzelnen Stichprobenwerte vom Mittelwert entfernt liegen. Zum besseren Vergleichen der Streuung mit dem Mittelwert wird aus der Varianz die Wurzel gezogen und als Standardabweichung bezeichnet.

Es gelten folgende Formeln:

Formel 8: Berechnung der Varianz

$$\text{Varianz} = \text{Var} = \sum_{i=1}^n (x_i - \bar{x})^2 \times \frac{1}{n}$$

n ... Stichprobenumfang

x ... Stichprobenwert

i ... Laufindex

$\bar{x}$... arithmetisches Mittel

Formel 9: Berechnung der Standardabweichung

$$\text{Standardabweichung} = s = \sqrt{\text{Var}}$$

Eine Berechnung der Streuungsmaße ist nur für qualitative Merkmale möglich, was aus den Bedingungen für die Berechnung des arithmetischen Mittel hervorgeht (Kuckartz, 2013).

A 2.2 Verwendete Testverfahren der induktiven Statistik

A 2.2.1 Der Chi²-Vierfeldertest als Grundlage

Zur Analyse von Häufigkeiten werden oftmals Chi²-Tests verwendet, welche sich bei jedem Skalenniveau berechnen lassen. Im Vierfeldertest (siehe Tabelle 15) wird geprüft, ob ein Zusammenhang zwischen zwei Alternativmerkmalen (A_1 , A_2 , B_1 und B_2) besteht. Er wird auch als Chi²-Unabhängigkeitstest bezeichnet. Die Idee des Chi²-Tests besteht im Vergleichen empirisch beobachteter Häufigkeiten mit den erwarteten Häufigkeiten. Je näher die beobachteten Häufigkeiten bei den erwarteten Häufigkeiten liegen, desto kleiner wird die ermittelte Prüfgröße und würde genau Null sein bei identischen Häufigkeiten.

Tabelle 15: Vierfeldertafel des Chi²-Tests. Modifiziert nach Weiß, 2013.

Merkmal	A ₁	A ₂	Randsummen
B ₁	a	b	a+b
B ₂	c	d	c+d
Randsummen	a+c	b+d	n=a+b+c+d

Die Formel zur Berechnung lautet:

Formel 10: Berechnung der Prüfgröße χ^2 im Chi²-Test

$$\chi^2 = \frac{n \times (ad - bc)^2}{(a + b)(a + c)(c + d)(b + d)}$$

χ^2 ...Prüfgröße
a,b,c,d...Häufigkeit der
jeweiligen Phänotyp-
ausprägung
n...Stichprobenumfang

Bei der Vierfeldertafel hat die Prüfgröße χ^2 einen Freiheitsgrad. Die Beibehaltung der Nullhypothese auf dem α -Niveau erfolgt, wenn die Prüfgröße innerhalb des Intervalls $[0, \chi^2_{1;l-\alpha}]$ liegt. Wenn als Ergebnis des Chi²-Tests eine Prüfgröße $\chi^2 > \chi^2_{1;l-\alpha}$ ermittelt wird, ist die Alternativhypothese anzunehmen (Weiß, 2013). Für $\alpha = 5\%$ ist

$\chi^2_{1;0,95} = 3,841$. In Abhängigkeit von der Prüfgröße χ^2 kann ein entsprechender p-Wert angegeben werden. Dieser enthält Informationen über die Wahrscheinlichkeit, dass das Testergebnis vorliegt, wenn in Wahrheit die Nullhypothese zutrifft. Die Alternativhypothese kann angenommen werden, wenn der p-Wert kleiner als das gewählte Signifikanzniveau α ist. Die Voraussetzungen für die Anwendung des Chi²-Tests besagen, dass die erwartete Häufigkeit mindestens 5 ergeben muss und keine der beobachteten Häufigkeiten 0 ist. Der Chi²-Test wurde deshalb nur zur Berechnung des Hardy-Weinberg-Gleichgewichts genutzt.

A 2.2.2 Fishers exakter Test

Der exakte Test nach Fisher kann auch für kleine Stichproben mit einer erwarteten Häufigkeit kleiner 5 angewendet werden und liefert zuverlässige Ergebnisse (Kuckartz, 2013). Der Test wird als „exakt“ bezeichnet, da der p-Wert als Prüfgröße direkt berechnet wird. Das Berechnen des p-Wertes erfolgt in einem mehrschrittigen Rechenverfahren und ist sehr aufwändig. Mit der Software STATISTICA können die p-Werte zuverlässig berechnet werden. Da die Voraussetzungen des Chi²-Vierfeldertests nicht für alle Berechnungen erfüllt werden konnten, wurde als Alternative für das Vergleichen der

Allelhäufigkeiten der exakte Test nach Fisher verwendet. Er wurde für alle Vergleiche dieser Art genutzt, um eine einheitliche Darstellung der Ergebnisse zu ermöglichen.

A.2.2.3 Mann-Whitney-U-Test

Bei der Testung der erhobenen Daten zeigten nicht alle Werte eine Normalverteilung. Deshalb wurde bei der Auswertung der Effekte der Genotypen auf die Serumlipidwerte mit dem nichtparametrischen Mann-Whitney-U-Test gearbeitet. Dieser setzt nur voraus, dass die abhängige Variable mindestens ordinalskaliert ist. Die Werte aus den zwei zu vergleichenden Stichproben werden zunächst in aufsteigender Reihe geordnet und mit Rängen versehen. Danach werden die Rangzahlen separat in jeder Stichprobe addiert und die Summe als R_1 bzw. R_2 bezeichnet. Zur Berechnung der Prüfgröße U werden dann U_1 und U_2 für die jeweilige Stichprobe berechnet, und der kleinere der beiden Werte wird als Testgröße verwendet.

Die Formel zur Berechnung von U_1 und U_2 lautet:

Formel 11: Berechnung der Prüfgröße U_1 im Mann-Whitney-U-Test

$$U_1 = n_1 \times n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

n_1 ...Stichprobenumfang
Gruppe 1
 n_2 ...Stichprobenumfang
Gruppe 2
 R_1 ...Rangsumme
Gruppe 1

Formel 12: Berechnung der Prüfgröße U_2 im Mann-Whitney-U-Test

$$U_2 = n_1 \times n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

n_1 ...Stichprobenumfang
Gruppe 1
 n_2 ...Stichprobenumfang
Gruppe 2
 R_2 ...Rangsumme
Gruppe 2

Die Prüfgröße U wird wieder mit dem kritischen Wert für den Stichprobenumfang verglichen bzw. der entsprechende p-Wert ermittelt (Weiß, 2013).